Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXIV ciclo

Comparing Different Approaches for Clustering Categorical Data

Laura Anderlucci

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Gennaio 2012

Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXIV ciclo

Comparing Different Approaches for Clustering Categorical Data

Laura Anderlucci

Coordinatore:
Prof.ssa Daniela Cocchi

Tutor:
Prof.ssa Angela Montanari

Co-Tutor:
Prof. Christian Hennig

Settore Disciplinare: SECS-S/01
Settore Concorsuale: 13/D1

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Gennaio 2012

Abstract

There are different ways to do cluster analysis of categorical data in the literature and the choice among them is strongly related to the aim of the researcher, if we do not take into account time and economical constraints.

Main approaches for clustering are usually distinguished into *model-based* and *distance-based* methods: the former assume that objects belonging to the same class are similar in the sense that their observed values come from the same probability distribution, whose parameters are unknown and need to be estimated; the latter evaluate distances among objects by a defined dissimilarity measure and, basing on it, allocate units to the closest group.

In clustering, one may be interested in the classification of similar objects into groups, and one may be interested in finding observations that come from the same true homogeneous distribution.

But do both of these aims lead to the same clustering? And how good are clustering methods designed to fulfil one of these aims in terms of the other?

In order to answer, two approaches, namely a latent class model (mixture of multinomial distributions) and a partition around medoids one, are evaluated and compared by Adjusted Rand Index, Average Silhouette Width and Pearson-Gamma indexes in a fairly wide simulation study. Simulation outcomes are plotted in bi-dimensional graphs via Multidimensional Scaling; size of points is proportional to the number of points that overlap and different colours are used according to the cluster membership.

Contents

Chapter 1

Introduction

1.1 Cluster Analysis

A cluster can be defined as a group of the same or similar elements gathered or occurring closely together. How to find and/or how to identify homogenous groups in a multivariate context is the aim of Cluster Analysis. Indeed, Kaufman and Rousseeuw ([44]) defined Cluster Analysis as *the art of finding groups in data*.

There are different ways to do cluster analysis of categorical data in the literature and the choice among them is strongly related to the aim of the researcher, if we do not take into account time and economical constraints.

Main approaches for clustering are usually distinguished into *model-based* and *distance-based* methods: the former assume that objects belonging to the same class are similar in the sense that their observed values come from the same probability distribution, whose parameters are unknown and need to be estimated; the latter evaluate distances among objects by a defined dissimilarity measure and, basing on it, allocate units to the closest group.

1.1.1 Model-based clustering: Latent Class Analysis

As evoked by its name, a model-based clustering approach postulates the existence of a true statistical model for the population under study. In this direction a very well known method is the Latent Class Analysis (LCA): it assumes that data is generated by a mixture of underlying probability distributions. Each cluster is represented by a single component of the mixture (i.e. *latent class*), thus it is described by a probability distribution whose parameters and size are unknown quantities to be estimated. More precisely, when focusing on categorical variables only, the underlying model is a mixture of multinomial distributions.

By way of illustration, consider a case-control study in which the relationship between exposure to a potential risk-factor and occurrence of a disease is investigated. In particular the exposure is evaluated by several, say p, empirical measures $X_1, \ldots, X_p$; each test X_i will classify some true risk factor positives as negative (false negative) and/or some true risk factor negatives as positive (false positive). In this field, the goodness of the classification is usually quantified in terms of *sensitivity* and *specificity*: the former is the proportion of truly exposed individuals who are correctly classified as exposed, the latter is the proportion of truly not exposed individuals who are correctly classified as not exposed. Sensitivity and specificity may be different across the measures and may also vary between the study groups (i.e. cases and controls). While sensitivity and specificity refer to the probability of a positive or negative test given true exposure status, *predicted values* reflect the probability of true exposure status conditional on test results [25]. In this example, predicted values are the main interest; in other words, given the observed test results, the aim is to assign individuals to the true exposure status.

Latent Class Analysis can be used to estimate the latent distribution of true exposure in the study groups; the basic idea would be to conceive both study groups as comprising an unknown mixture of truly exposed and truly unexposed individuals. The observed association between the measures $X_1, \ldots, X_p$ would be assumed to be solely due to their dependence on the unknown true exposure status; what is expected is that after an appropriate decomposition of the mixture, local independence among the observed variables in each mixture component is found.

An exhaustive description of the Latent Class Analysis method is given in Chapter 2.

1.1.2 Distance-based clustering: Partition Around Medoids

Distance-based methods are probably the most intuitive approach to clustering: the idea is to form groups so that objects in the same group are similar to each other, whereas objects in different groups are as dissimilar as possible. Of course there are many methods that try to achieve this aim.

The approach that is briefly presented here (but that will be fully described in the following) is Partition Around Medoids (PAM).

Partition Around Medoids (developed by Kaufman and Rousseeuw, 1990 [44]) is based on the search for k *representative objects* among the units of the data set. As their name suggests, these objects should be somehow representative of the structure of the data; they are called medoids. After finding a set of k representative objects, units are assigned to the nearest medoid, outlining k clusters. Crucial is the choice of proximity measure to be used: it defines how two units can be considered similar.

By way of illustration, consider a marketing research study where a sample of customers of a certain product have been asked to answer to a questionnaire about their satisfaction and their personal habits, with multiple choice items. The aim is to identify group of customers with similar motivations.

Given the responses to the questionnaire, Partition Around Medoids can be used to identify homogenous groups of customers according to specific features (e.g. geographic differences, personality differences, demographic differences, use of product differences, psychographic differences, gender differences etc.) thus improving the market knowledge and allowing for a targeted advertising campaign.

An exhaustive description of the Partition Around Medoids procedure is given in Chapter 3.

1.2 Motivation

In clustering, one may be interested in the classification of similar objects into groups, and one may be interested in finding observations that come from the same true homogeneous distribution.

But do both of these aims lead to the same clustering? And how good are clustering methods designed to fulfil one of these aims in terms of the other?

Researchers do not often think to these questions, thus the choice between the two approaches is sometimes not very well justified.

In order to answer, two approaches, namely a **latent class model** (mixture of multinomial distributions) and a **partition around medoids** one, are evaluated and compared in a fairly wide simulation study. The study would serve as a basis to understand similarities and differences in terms of classification of the two approaches and to detect, if any, different roles played by data features.

1.3 The study

Simulations consisted in generating several data sets from different parameterizations (according to specific data features), then the two clustering methods were applied and finally the obtained classifications were compared.

For each parameterization 2000 different data sets were generated with the LatentGold® software and the true classification of units was recorded. To do so, we fixed the parameter values according to a simulation scheme and, by telling LatentGold® the number of variables, the number of categories and the number of latent classes, we generated the 2000 data set for each parameterization. A full list of the parameter values we adopted is in the Appendix .

Then we performed the clustering according to a model-based approach with the same commercial software and with an open-source software (using an EM algorithm, implemented as a function `lcmixed` in the R-package `fpc`), with the aim of comparing results, precision and time with LatentGold®; we also performed the clustering according to a distance-based method using `pam` function, contained in the R-package `cluster` (dissimilarity measure = *manhattan*).

LatentGold®, developed by Vermunt ([73]), is currently the leader software for Latent Class Analysis. To find the Maximum Likelihood (ML) estimates for the model parameters, LatentGold® uses both the EM and the Newton-Raphson algorithm. In practice, the estimation process starts with a number of EM iterations. When close enough to the final solution, the program switches to Newton-Raphson. According to Vermunt ([75]), "this is a way to exploit the advantages of both algorithms; that is, the stability of EM even when it is far away from the optimum and the speed of Newton-Raphson when it is close to the optimum".

The algorithm developed for PAM consists of two phases: a BUILD phase (where an initial clustering is obtained by successive selection of representative points until k objects have been found) and a SWAP phase (where it is attempted to improve a set of representative objects and also to improve the clustering yielded by this set). Since all the potential swaps are considered, the results of the algorithm do not depend on the order of the objects in the input file (unless there are some ties among the distances between objects).

Once all the models have run, in order to compare the obtained classifications we use three indexes: the Adjust Rand Index, the Average Silhouette Width and the Pearson Gamma.

The *Adjusted Rand Index* (ARI) is a measure of the similarity between two data clusterings. In this context, the ARI is used to compare the classifications yielded by a model-based and a distance-based clustering approach with what is recorded as 'true' cluster membership.

The *Average Silhouette Width* index (ASW) is a measure of tradeoff between similarity of observations in the same cluster and dissimilarity of observations in different clusters. In the definition of ASW, the dissimilarities of observations from other observations of the same cluster are compared with dissimilarities from observations of the nearest other cluster, which emphasises separation between the cluster and their neighbouring clusters ("gaps" between clusters).

The *Pearson Gamma* (PG) index is the Pearson correlation $\rho(\mathbf{d}, \mathbf{m})$ between the vector $\mathbf{d}$ of pairwise dissimilarities and the binary vector $\mathbf{m}$ that is 0 for every pair of observations in the same cluster and 1 for every pair of observations in different clusters. PG emphasises a good approximation of the dissimilarity structure by the clustering in the sense that observations in different clusters should strongly be correlated with large dissimilarity.

Latent Class Clustering is by definition aimed to recover the 'true' classification, since it is a model-based clustering method. Therefore, we expect it to perform better than PAM in terms of Adjusted Rand Index. Whereas, since PAM has a distance-based approach, we expect it to perform better than LatentGold® in terms of Average Silhouette Width and Pearson Gamma values.

A full description of the three indexes is in Chapter 4.

1.4 Previous Results

An important result in comparing different approaches to the clustering of categorical data was previously obtained by Celeux and Govaert in 1991 ([14]).

In their ‘Discrete Data and Latent Class Model’ they showed that a well-known clustering criterion for discrete data, the information criterion, is closely related to the Classification Maximum Likelihood (CML) criterion for the latent class model.

In particular, in the CML method the mixing proportions and the parameter vector are estimated so that a likelihood function is maximized. The authors showed that, by using a standard Lagrangian manipulation, the parameter vector of the k^{th} mixture component can be viewed as a “center” of cluster k. Using this expression, the maximization of the CML criterion is equivalent to the maximization of the classical information criterion.

Focusing on binary data, they considered a clustering criterion where the information to be minimized was the Manhattan distance between an object and its cluster representation (which is similar to the idea behind PAM algorithm). They showed that this criterion is directly related to a Bernoulli mixture (i.e. the latent class model for binary data): maximizing the CML criterion leads to minimizing the information criterion, even though there are some degenerating configurations. For example when the size of any of the clusters tends to zero.

In an application with empirical data, they compared the results of the CML with those obtained with the EM algorithm: CML estimates show an important bias for the mixing proportions, i.e. the information criterion tends to provide equal-sized clusters. As pointed out by Bryant and Williamson ([11]), the more rare a component is, the more CML's bias tends to be serious. Nevertheless the difference between Bernoulli probability estimates with both methods is not so marked.

For further references see [14].

Chapter 2

Latent Class Clustering

The subject of clustering is concerned with the investigation of the relationships within a set of 'objects' in order to establish whether or not the data can validly be summarized and better interpreted by a small number of classes (or clusters) of similar objects.

In this section we focus on a model-based approach, presenting the Latent Class Clustering (LCC) method.

2.1 The method

A milestone in the literature of the latent class models with categorical variables is one of the papers Goodman published in 1974 ([27]), which presents a relatively simple method for calculating the maximum likelihood estimate of the frequencies in the p-way contingency table expected under the model (where p indicates the number of manifest polytomous variables), and for determining whether the parameters in the estimated model are identifiable.

He firstly considered a p-way contingency table which cross-classifies a sample of n individuals with respect to p manifest polytomous variables. The observed relationships - if any - among the p variables can be somehow explained by a K-class latent structure if there is some latent polytomous variable K, so that each of the n individuals is in only one of the K classes with respect to this variable, and within the k^{th} latent class the manifest variables are mutually independent.

The model is described by equation 2.1:

$$f(x) = \sum_{k=1}^{K} p_k f(x, a_k), \tag{2.1}$$

with $\sum_{k=1}^{K} p_k = 1$, i.e. the mixing proportions sum to 1. The probability mass function $f(x, a_k)$ describes a multinomial distribution with parameters $a_k = (a_k^{jl},\ l = 1, \ldots, m_j,\ j = 1, \ldots, p)$:

$$f(x, a_k) = \prod_{j=1}^{p} \prod_{l=1}^{m_j} (a_k^{jl})^{x^{jl}}, \tag{2.2}$$

with $\sum_{l=1}^{m_j} a_k^{jl} = 1$. The generic polytomous variable j ($j = 1, \ldots, p$) consists of m_j categories, and $m = \sum_{j=1}^{p} m_j$ indicates the total number of levels.

Example

To illustrate the method, Goodman analyzed data contained in Table 2.1, a 2^4 contingency table presented earlier by Stouffer and Toby [64], which cross-classified 216 respondents with respect to whether they tend towards universalistic values (+) or particularistic values (-) when confronted by each of four different situations of role conflict.

Table 2.1: Observed cross-classification of 216 respondents with respect to whether they tend toward universalistic (+) or particularistic (-) values in four situations of role conflict (A, B, C, D).

A	B	C	D	Observed frequency	A	B	C	D	Observed frequency
+	+	+	+	42	-	+	+	+	1
+	+	+	-	23	-	+	+	-	4
+	+	-	+	6	-	+	-	+	1
+	+	-	-	25	-	+	-	-	6
+	-	+	+	6	-	-	+	+	2
+	-	+	-	24	-	-	+	-	9
+	-	-	+	7	-	-	-	+	2
+	-	-	-	38	-	-	-	-	20

The idea is to determine whether a latent structure can explain the observed relationships among the four binary variables and hence allows for a meaningful clustering of the data.

Let π_{abcd} denote the probability that an individual will be at level (a,b,c,d) with respect to the joint variable (A,B,C,D) $(a = 1,\ldots,m_A;\ b = 1,\ldots,m_B;\ c = 1,\ldots,m_c;\ d = 1,\ldots,m_D)$. Suppose that there is a latent polytomous variable K consisting of K classes, that can explain the relationships among the manifest variables (A, B, C, D). This means that π_{abcd} can be expressed as follows:

$$\pi_{abcd} = \sum_{k=1}^{K} \pi_{abcdk}, \tag{2.3}$$

where

$$\pi_{abcdk} = \pi_k \pi_{ak} \pi_{bk} \pi_{ck} \pi_{dk} \tag{2.4}$$

denotes the probability of an individual will be at level (a, b, c, d, k) with respect to the joint variable (A, B, C, D, K). The π_k is the probability that an individual will be at level k with respect to variable K; moreover, π_{ak} is the conditional probability that an individual will be at level a with respect to variable A, given that he is at level k with respect to variable K, and finally π_{bk}, π_{ck} and π_{dk} denote similar conditional probabilities. Formula (2.3) avers that the individuals can be classified into K mutually exclusive and exhaustive latent classes, and the product of the single probabilities in (2.4) is the result of the hypothesis of local independence within each latent class.

From these premises it is straightforward to see that:

$$\sum_{k=1}^{K} \pi_k = 1;\quad \sum_{a=1}^{m_A} \pi_{ak} = 1;\quad \sum_{b=1}^{m_B} \pi_{bk} = 1;\quad \sum_{c=1}^{m_C} \pi_{ck} = 1;\quad \sum_{d=1}^{m_D} \pi_{dk} = 1; \tag{2.5}$$

$$\pi_k = \sum_{a,b,c,d} \pi_{abcdk} \tag{2.6}$$

$$\pi_k \pi_{ak} = \sum_{b,c,d} \pi_{abcdk}. \tag{2.7}$$

Formulas similar to (2.7) can be obtained for the other variables, i.e. π_k multiplied by π_{bk}, π_{ck} and π_{dk}.

Furthermore, from the law of total probability we have that the conditional probability $\pi_{k|abcd}$ that an individual is in latent class k, given that he was at level (a, b, c, d) with respect to the joint variable (A, B, C, D) is equal to:

$$\pi_{k|abcd} = \frac{\pi_{abcdk}}{\pi_{abcd}}. \tag{2.8}$$

Using expression (2.8), π_k and $\pi_k \pi_{ak}$, in (2.6) and (2.7) respectively, can be rewritten as

$$\pi_k = \sum_{a,b,c,d} \pi_{abcd}\pi_{k|abcd}, \tag{2.9}$$

$$\pi_{ak} = \frac{\sum_{b,c,d} \pi_{abcd}\pi_{k|abcd}}{\pi_k}. \tag{2.10}$$

Formulas similar to (2.10) can be obtained for the other variables π_{bk}, π_{ck} and π_{dk}.

2.1.1 Description of the Algorithm

In order to estimate the parameters of equation (2.1) from the observed data, Goodman sketched a simple algorithm. Using the notation of the example, equation (2.1) becomes:

$$\begin{aligned} f(x) &= \sum_{k=1}^{K} \pi_k f(x, a_k) = \\ &= \sum_{k=1}^{K} \pi_k \left(\prod_{a=1}^{m_A} (\pi_{ak})^{x_a} \prod_{b=1}^{m_B} (\pi_{bk})^{x_b} \prod_{c=1}^{m_C} (\pi_{ck})^{x_c} \prod_{d=1}^{m_D} (\pi_{dk})^{x_d} \right). \end{aligned} \tag{2.11}$$

Let p_{abcd} indicate the observed proportion of individuals at level (a, b, c, d) and let π denote the vector of parameters $(\pi_k, \pi_{ak}, \pi_{bk}, \pi_{ck}, \pi_{dk})$ in the latent class model; finally let $\hat{\pi}$ denote the corresponding maximum likelihood estimate of the vector. To calculate $\hat{\pi}$, the algorithm is organized in the following steps:

1. Start with an initial trial value for $\hat{\pi}$,

$$\hat{\pi}(0) = \{\hat{\pi}_k(0), \hat{\pi}_{ak}(0), \hat{\pi}_{bk}(0), \hat{\pi}_{ck}(0), \hat{\pi}_{dk}(0)\};$$

2. Substitute the components of $\hat{\pi}(0)$ into the corresponding terms on the right-hand side of formula (2.4) to obtain a trial value for $\hat{\pi}_{abcdk}$;

3. Use (2.3) to obtain a trial value for $\hat{\pi}_{abcd}$, replacing the terms on the right-hand side of (2.3) by the corresponding trial values found at the previous step;

4. Obtain a trial value for $\hat{\pi}_{k|abcd}$ by calculating $\hat{\pi}_{k|abcd} = \frac{\hat{\pi}_{abcdk}}{\hat{\pi}_{abcd}}$;

5. Similarly, obtain a new trial value for $\hat{\pi}_k$ by calculating:

$$\hat{\pi}_k = \sum_{a,b,c,d} p_{abcd} \hat{\pi}_{k|abcd};$$

6. By using the following expressions obtain new trial values for $\hat{\pi}_{ak}, \hat{\pi}_{bk}, \hat{\pi}_{ck}$, and $\hat{\pi}_{dk}$:

$$\begin{aligned}
\hat{\pi}_{ak} &= \frac{\sum_{bcd} p_{abcd} \hat{\pi}_{k|abcd}}{\hat{\pi}_k}, \\
\hat{\pi}_{bk} &= \frac{\sum_{acd} p_{abcd} \hat{\pi}_{k|abcd}}{\hat{\pi}_k}, \\
\hat{\pi}_{ck} &= \frac{\sum_{abd} p_{abcd} \hat{\pi}_{k|abcd}}{\hat{\pi}_k}, \\
\hat{\pi}_{dk} &= \frac{\sum_{abc} p_{abcd} \hat{\pi}_{k|abcd}}{\hat{\pi}_k};
\end{aligned}$$

7. Repeat the procedure from step 2 to obtain the next trial value for $\hat{\pi}$.

In this iterative procedure a latent class is deleted if the corresponding estimate tends to zero. The procedure converges to a solution to the system of equations and to a corresponding likelihood. By trying various initial trial values for $\hat{\pi}$ it is possible to compare the solutions obtained by the corresponding likelihood values.

2.1.2 Example results

Picture 3.1 is a two-dimensional Multi Dimensional Scaling (MDS) representation of the example considered. Size of points is proportional to the number of units which overlap.

By applying the model described in the previous section to the data in Table 2.1, we concluded that the underlying latent structure that better accounts for the association between the manifest variables is described by

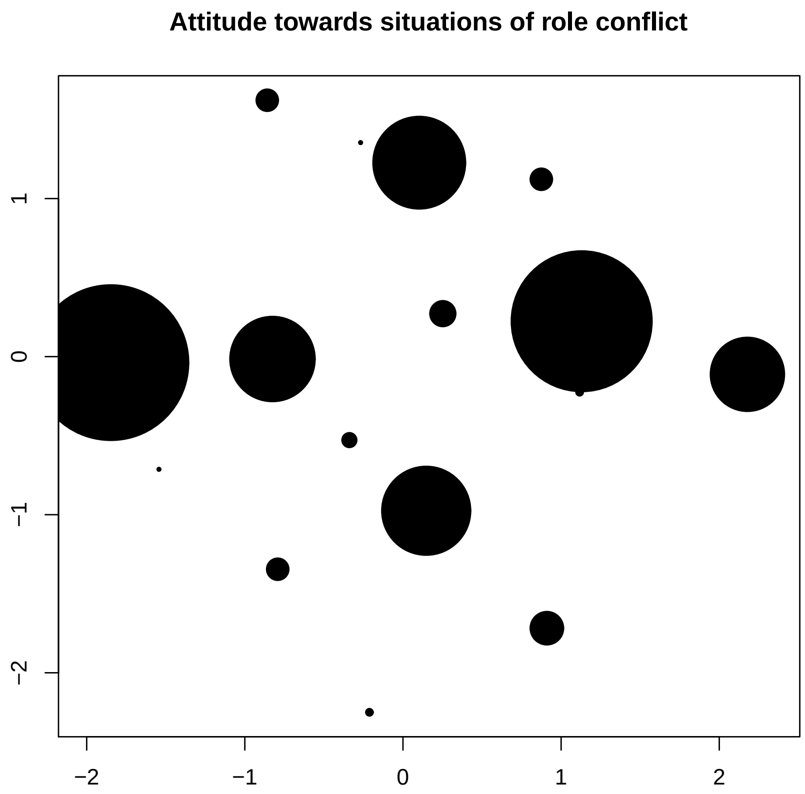


Figure 2.1: Multi Dimensional Scaling of Stouffer and Toby (1951) dataset, included in Goodman ([27])

two latent classes. This finding arises from comparing the values of some goodness-of-fit test performed on different models (i.e. models with different number of latent classes and/or with some parameter restrictions).

Table 2.2 contains the parameter estimates, for a model with a two-class latent structure.

From Table 2.2 it is clear that, with respect to the joint manifest variable (A, B, C, D), the modal levels are (+,+,+,+) and (+,-,-,-) for latent class 1 and 2, respectively. Furthermore, the second latent class is modal, since $\hat{\pi}_2$ is much larger than $\hat{\pi}_1$. Thus, most individuals (i.e. those in cluster 2) tend to be 'intrinsically' particularistic, except for situation A, whereas individuals in cluster 1 tend to be 'intrinsically' universalistic.

Latent Class Analysis yields a probabilistic clustering approach. Al-

Table 2.2: Estimated parameters in the latent structure applied to Table 2.1

Class Latent	$\hat{\pi}_k$	$\hat{\pi}^A_{1k}$	$\hat{\pi}^B_{1k}$	$\hat{\pi}^C_{1k}$	$\hat{\pi}^D_{1k}$
1	0.279	0.993	0.940	0.927	0.769
2	0.721	0.714	0.330	0.354	0.132

Table 2.3: Classification of units from Goodman's dataset according to Latent Class Clustering

A	B	C	D	n_i	Clust	A	B	C	D	n_i	Clust
+	+	+	+	42	1	-	+	+	+	1	2
+	+	+	-	23	1	-	+	+	-	4	2
+	+	-	+	6	1	-	+	-	+	1	2
+	+	-	-	25	2	-	+	-	-	6	2
+	-	+	+	6	2	-	-	+	+	2	2
+	-	+	-	24	2	-	-	+	-	9	2
+	-	-	+	7	2	-	-	-	+	2	2
+	-	-	-	38	2	-	-	-	-	20	2

though each object is assumed to belong to one class, it is taken into account that there is uncertainty about a unit's class membership. For each individual his posterior class-membership probabilities are computed from the estimated model parameters and his observed score ([51]); units are thus assigned to the class with the highest posterior probability. Classification of units is in Table 2.3.

In Figure 2.2, which is analogue to Figure 3.1, we used different colours to distinguish cluster membership. Clusters look well separated and of different size.

2.2 Identifiability

So far we have presented how to estimate the set of parameters π of a Latent Class Model, but we have not considered whether vector $\hat{\pi}$ is uniquely determined. If it is so, we say it is identifiable; if $\hat{\pi}$ is identifiable within some neighbourhood of π then it is locally identifiable. In his paper ([27]), Goodman gave a useful sufficient condition for local identifiability.

In a latent class model, the number of parameters to estimate is equal to:

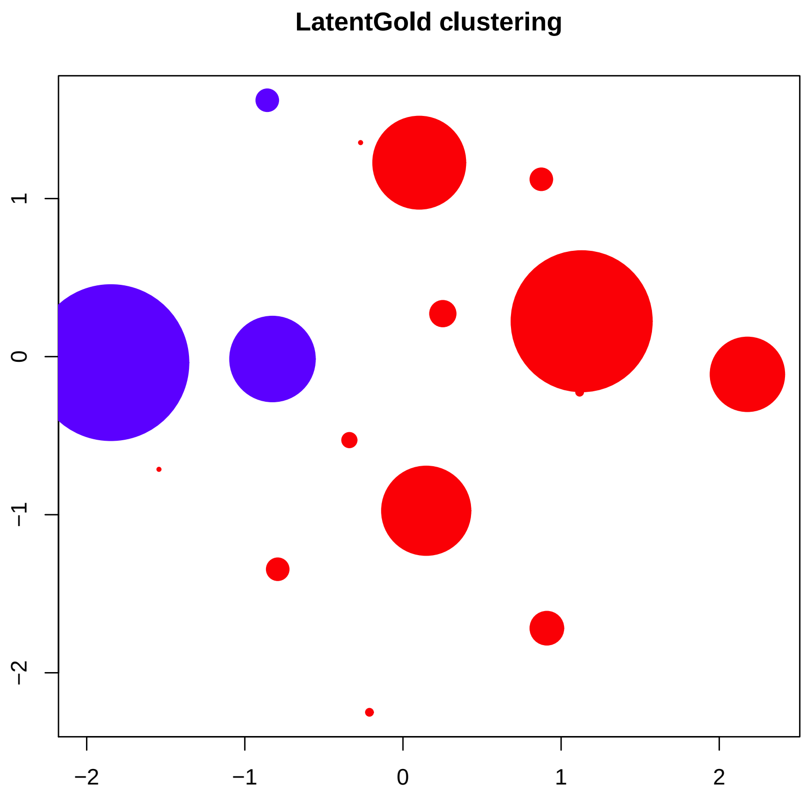


Figure 2.2: Data from Stouffer and Toby (1951) according to Latent Class Clustering

$$
\begin{aligned}
& \underbrace{K-1}_{\sum_{k=1}^{K}\pi_k=1} + \left(\underbrace{m_1-1}_{\sum_{l=1}^{m_1}\pi_{1k}=1} + \ldots + \underbrace{m_p-1}_{\sum_{l=1}^{m_p}\pi_{pk}=1} \right) K \\
= \quad & K-1+\left(m_1+\ldots+m_p-\underbrace{p}_{'1'\times p} \right) K \\
= \quad & \left(\sum_{j=1}^{p} m_j-(p-1) \right) K-1.
\end{aligned}
$$

This set of parameters can be called ‘basic set’.

The distributions resulting from the model lie in a space of dimension $\prod_{j=1}^{p} m_j - 1$, since all the joint probabilities sum to 1. When

$$\prod_{j=1}^{p} m_j - 1 < \left(\sum_{j=1}^{p} m_j - (p-1) \right) K - 1$$
$$\prod_{j=1}^{p} m_j < \left(\sum_{j=1}^{p} m_j - (p-1) \right) K, \tag{2.12}$$

the number of parameters in the basic set exceeds the corresponding number of joint probabilities, hence the parameters will not be identifiable.

If condition (2.12) is not verified, i.e. the number of parameters in the basic set does not exceed the corresponding number of joint probabilities, for each joint probability the derivative with respect to the parameters in the basic set has to be calculated. A matrix consisting of $\prod_{j=1}^{p} m_j - 1$ rows and $\left(\sum_{j=1}^{p} m_j - (p-1) \right) K - 1$ columns is obtained. By extension of a standard result about Jacobian, the parameters in the model will be locally identifiable if the rank of the matrix is equal to the number of columns, i.e. to the number of parameters in the basic set.

Notice that this condition only refers to local identifiability. A stronger and easier result is proposed by Allman et al. ([3]) and it is outlined in the following section.

2.2.1 Background

The study of identifiability asks whether one may, in principle, recover the parameters of the distribution of some observed variables. Although identification problem is not a problem of statistical inference in a strict sense, non-identifiable parameters cannot be consistently estimated, thus identifiability becomes a prerequisite of parametric statistical inference [3].

The classical definition of identifiability requires that for any two different values $\pi \neq \pi'$ in the parameter space, the corresponding probability distributions are different. In many cases, this map will not be strictly injective. In the Latent Class Analysis for instance, the latent classes can be freely relabelled without changing the distribution of the observations (i.e. "label swapping"). In the following we will refer to generic identifiability, which means that the set of points for which identifiability does not hold has measure zero. In other words, when the parameters of a latent class model are generically identifiable any observed data set has probability one of being drawn from a distribution with identifiable parameters.

2.2.2 Parameter identifiability of finite mixtures of finite measure products

The work of Allman et al. shows that it is possible to derive some identifiability results for latent class models, by extending a fundamental algebraic result of Kruskal ([46]) on 3-ways tables.

To do so, they observed that p categorical variables can be clumped into 3 agglomerate variables, so that Kruskal's result can be applied. Here the Theorem follows:

Theorem 2.2.2.1. *Consider the latent class model with K latent classes and p categorical variables x_j, ($j = 1, \dots, p$), with number of categories m_j. Suppose there exists a tripartition of the set $S = \{1, \dots, p\}$ into three disjoint nonempty subsets S_1, S_2, S_3, such that if $\nu_h = \prod_{j \in S_h} m_j$ then*

$$\min(K, \nu_1) + \min(K, \nu_2) + \min(K, \nu_3) \geq 2K + 2. \qquad (2.13)$$

Then model parameters are generically identifiable, up to label swapping.

Let consider the special case of finite mixture of p Bernoulli products with K components. In order to obtain the strongest identifiability result, they chose a tripartition that maximized the left-hand side of inequality 2.13. This yielded the following Corollary.

Corollary 2.2.2.2. *Parameters of the finite mixture of p different Bernoulli products with K components are generically identifiable, up to label swapping, provided*

$$p \geq 2\lceil \log_2 K \rceil + 1,$$

where $\lceil x \rceil$ is the smallest integer at least as large as x.

For the more general model with nominal variables with same number of categories $m_j = m > 2$, the lower bound on the number of variates needed in order to generically identify the parameters, up to label swapping, is

$$p \geq 2\lceil \log_m K \rceil + 1.$$

Despite its simple appearance, condition (2.13) is not easy to verify in an exact automatic procedure. So far, the only way to do this is to consider all the possible tripartition of the set of variables. Nevertheless, with reasonable large numbers the procedure is timing acceptable.

Table 2.4 contains a summary of identifiable/nonidentifiable models for some specific situations, according to condition 2.13.

The first column contains the number of latent classes considered (from 2 to 5), whereas the second one contains the lower bound at the right-hand side of inequality (2.13), i.e. $2K+2$, with K number of latent classes.

Only a selection of cases are in Table 2.4. In particular we include some 'border-line' situations: for each number of latent classes, we show the smallest number of categories for each variable needed in order to have identifiability of parameters, for a given number of manifest variables.

By way of illustration, consider the case of a model with 3 latent classes and 4 manifest variables. If at least one of the variables has more than two categories then the parameters are generically identifiable. Instead, if the considered variables are all binary there are no sufficient conditions to claim identifiability.

Table 2.4: General Identifiability - Summary

No. of latent classes	Lower bound	No. of items	No. of categories	Identifiability
2	6	3	any	Identifiable
3	8	3	2,2,3	Non-Identifiable
			2,3,3	Identifiable
		4	2,2,2,2	Non-Identifiable
			2,2,2,3	Identifiable
4	10	3	2,3,4	Non-Identifiable
			2,4,4	Identifiable
			3,3,4	Identifiable
		4	2,2,2,3	Non-Identifiable
			2,2,3,3	Identifiable
		5	2,2,2,2,2	Identifiable
5	12	3	3,4,4	Non-Identifiable
			4,4,4	Identifiable
			3,4,5	Identifiable
			2,5,5	Identifiable
		4	2,2,4,4	Identifiable
			2,3,3,4	Identifiable
		5	2,2,2,2,3	Non-Identifiable
			2,2,2,2,4	Identifiable
			2,2,2,3,3	Identifiable
		6	2,2,2,2,2,2	Identifiable

Chapter 3

Partitioning Around Medoids

Clustering a set of n objects into k groups is usually motivated by the aim of identifying internally homogeneous groups, which allow a summary of the information.

Main approaches for clustering are usually distinguished into *model-based* and *distance-based* methods (but there are more): the former assume that objects belonging to the same class are similar in the sense that their observed values come from the same probability distribution, whose parameters are unknown and need to be estimated; the latter evaluate distances among objects by a defined dissimilarity measure and, basing on it, allocate units to the closest group. In other words, they aim to partition the observations in such a way that objects within the same group are similar to each other, whereas objects in different groups are as dissimilar as possible.

Hence, a partition of a set of objects is considered "good" if objects of the same cluster are close or related to each other, whereas objects of different clusters are far apart or very different.

3.1 The method

In this section we focus on the distance-based approach, presenting a particular algorithm: the *partitioning around medoids* (PAM, developed by L. Kaufman and P. J. Rousseeuw, [45]).

The idea of the partition around medoids approach is to find k representative objects, which should represent special features or aspects of the data. Specifically, they are those units for which the average dissimilarity to all the objects of the same cluster is minimal. Each of them is called the **medoid**[1] of the cluster. After finding the set of medoids, each object of the data set is assigned to the nearest medoid. Note that it is similar to the k-means algorithm, but here the centers are members of the data set and not the cluster means. The aim is usually to uncover a structure that is already present in the data, but sometimes it is used to impose a new structure.

In the following we indicate a set of n observation with X

$$X = \{x_1, x_2, \ldots, x_n\}$$

and the dissimilarity between objects x_i and x_j with $d(i, j)$.

3.1.1 Dissimilarity definition

Since PAM is a distance-based approach to clustering, the choice of the dissimilarity measure is quite a central aspect to consider, because it is supposed to reflect what is taken as 'similar'.

A popular distance measure between two objects x_i and x_j on p variables is the *Euclidean* one:

$$\begin{aligned} d_E(i,j) &= \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{ip} - x_{jp})^2} \\ &= \sqrt{\sum_{l=1}^{p} (x_{il} - x_{jl})^2}. \end{aligned} \tag{3.1}$$

It corresponds to the true geometrical distance between the points of coordinates $(x_{i1}, x_{i2}, \ldots, x_{ip})$ and $(x_{j1}, x_{j2}, \ldots, x_{jp})$.

According to its formula (3.1), the Euclidean distance tends to give the variables with larger summand more weight because of the squares: it means that two observations are treated as less similar if there is a very large dissimilarity on one variable and small dissimilarities on the others than if there is about the same (a little bit larger) dissimilarity on all variables.

Another well-known metric is the *Manhattan* (or *city block* or L_1) dis-

[1] In the cluster analysis literature they are sometimes called *centrotypes*.

tance, defined by:

$$\begin{aligned} d_M(i,j) &= |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \ldots + |x_{ip} - x_{jp}| \\ &= \sum_{l=1}^{p} |x_{il} - x_{jl}| \end{aligned} \tag{3.2}$$

The use of the Manhattan distance is advised in those situations where, for example, a difference of 1 in the first variable and of 3 in the second variable is treated as a difference of 2 in both the first and the second ones.

Since we are in the context of categorical variables and we do not have any prior knowledge about the variables it makes sense to choose the Manhattan distance as a measure of dissimilarity, so that what matters is the number of disagreements.

Generally, different values of a nominal variable should not carry numerical information, unless there are interpretative reasons that can justify it. Therefore, when dealing with categorical variables it would be better to replace them with binary indicator variables for all their values. Let m_j denote the number of categories of variable j; technically only $m_j - 1$ binary variables would be needed to represent all information, but in terms of dissimilarity definition, leaving one of the categories out would lead to asymmetric treatment of the categories ([41]).

3.2 Description of the Algorithm

In the original version of the PAM algorithm, developed by Kaufman and Rousseeuw (1987), the *sum* of the dissimilarities of objects to their closest representative object was minimized (rather than the *average* dissimilarity).

The algorithm developed for PAM consists of two phases:

1. a BUILD phase, where an initial clustering is obtained by successive selection of representative points until k objects have been found;

2. a SWAP phase, where it is attempted to improve a set of representative objects and also to improve the clustering yielded by this set.

The algorithm is completely deterministic: the first object to be selected is the one for which the sum of dissimilarities to all other objects is as small as possible. Following this heuristic principle, at each step another object is selected, according to the highest decrease in an objective function. In order to find this object, the following steps are carried out:

1. It considers an object i which has not yet been selected.
2. It considers a non selected object j and calculates the dissimilarity with all the previously selected objects; it then indicates with D_j the dissimilarity with the most similar one and with $d(j, i)$ its dissimilarity with object i. It finally computes the difference between $D_j - d(j, i)$.
3. If this difference is positive, object j will contribute to the decision to select object i. Therefore it calculates
$$C_{ji} = max(D_j - d(j, i), 0).$$
4. It then calculates the total gain obtained by selecting object i:
$$\sum_j C_{ji}.$$
5. It finally chooses the not yet selected object i which yields
$$\max_i \sum_j C_{ji}.$$

This process continues until k objects have been found. At the end of the *build* phase, the algorithm attempts to improve the *value of the clustering*, which is defined as the sum of dissimilarities between each object and the most similar representative object.

During this phase (the so called 'SWAP' phase), the process considers all the pairs of objects (i, h), where i is an object that has been selected and object h has not, in order to determine what effect is obtained on the value of a clustering if a swap is carried out, namely if object i is no longer selected but object h is.

Operatively, in order to evaluate the effect of a swap between i and h the algorithm:

1. firstly considers a nonselected object j and calculates its contribution C_{jih} to the swap; two situations show up:

a. If j is more distant from both i and h than from one of the other representative objects, C_{jih} is zero;

b. If j is not further from i than from any other selected representative object (i.e. $d(j,i) = D_j$) then:

(i) either j is closer to h than to the second closest representative object

$$d(j,h) < E_j$$

where E_j is the dissimilarity between j and the second most similar representative object. In this case the contribution of object j to the swap between objects i and h is

$$C_{jih} = d(j,h) - d(j,i);$$

(ii) or, alternatively, j is at least as distant from h as from the second closest representative object

$$d(j,h) \geq E_j.$$

In this case the contribution of object j to the swap is

$$C_{jih} = E_j - D_j.$$

In situation (i) the contribution C_{jih} can be either positive or negative, since it depends on the relative position of objects j, h and i. Obviously, if the contribution is positive it means that j is closer to i than to h and so the swap is not favorable from the point of view of object j. Differently, in situation (ii) the contribution is always positive because it can never be advantageous to replace i by a point h which is further from j than the second closest representative object.

c. If j is further away from i than from at least one of the other representative objects but closer to h than to any representative object, the contribution of j to the swap is

$$C_{jih} = d(j,h) - D_j.$$

2. Then, it calculates the total result of a swap by adding the contribu-

tions C_{jih}:

$$T_{ih} = \sum_j C_{jih}.$$

Once it calculates the result T_{ih}, the algorithm has to decide whether to carry out the swap. In order to do this

3. It selects the pair (i, h) which

$$\min_{i,h} T_{ih}.$$

If the minimum is negative then the swap is carried out and the algorithm returns to step 1. Whereas, if the minimum T_{ih} is positive or 0 it means that carrying out a swap will not improve the value of the clustering and, hence, the algorithm stops.

Since all the potential swaps are considered, the results of the algorithm do not depend on the order of the objects in the input file (unless there are some ties among the distances between objects).

The original algorithm was written in Fortran, but currently a version for the R software exists: `pam` function is available in the `cluster` R-package.

3.3 Example

In order to fully understand this approach, let's consider again the example from Goodman's paper that we presented at page 8. Table 2.1 cross-classifies 216 respondents with respect to whether they tend towards universalistic values (+) or particularistic values (—) when confronted by each of four different situations of role conflict, labelled as 'A', 'B', 'C', 'D'.

Picture 3.1 is a two-dimensional Multi Dimensional Scaling (MDS) representation of the example considered. Size of points is proportional to the number of units which overlap.

The first thing to do in order to classify the units is to compute a dissimilarity matrix for all the observations; we considered the 'full' one, i.e. the one that includes ties, so that observations are weighted according to their frequency.

Furthermore, since the aim is to recover a separation between 'universalistic' and 'particularistic', the number of clusters we are interested in is

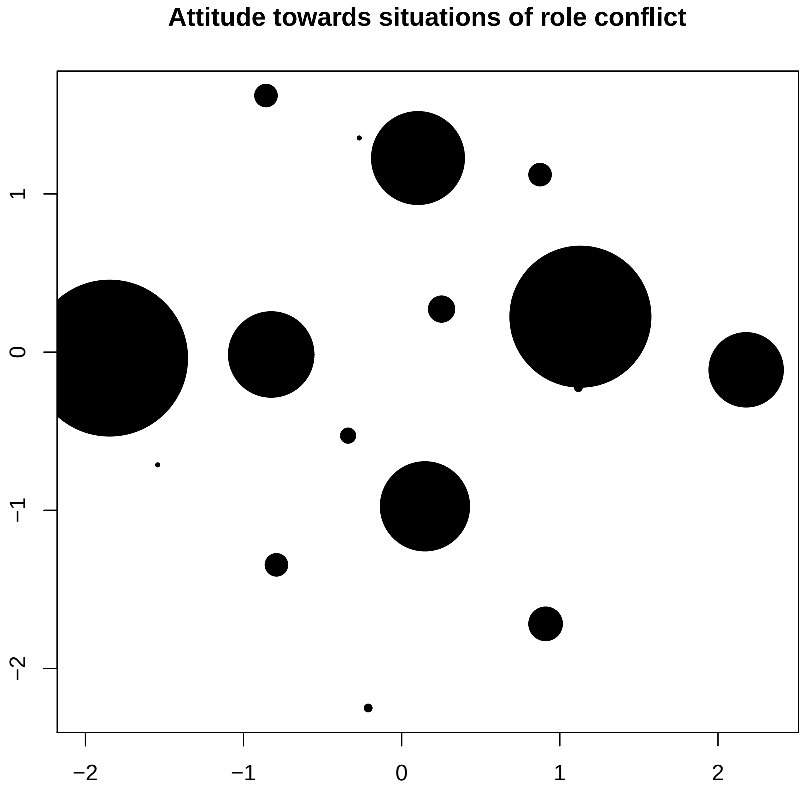


Figure 3.1: Multidimensional scaling of Stouffer and Toby (1951) dataset, included in Goodman ([27])

two.

Given all this information, by simply applying the `pam` function to the dissimilarity matrix and by fixing the number of cluster to two, the algorithm produces the classification, according to a distance-based approach.

In Figure 3.2, which is analogue to Figure 3.1, we used different colours to distinguish cluster membership. Clusters look well separated and of similar size.

Whereas, colours in Figure 2.2, page 14, indicate the cluster memberships assigned by LCC approach.

In order to underline differences between the two clustering methods, Table 3.1 shows the classification yielded by LCC and PAM.

Clusters obtained with LCC approach have size respectively 71 and 145,

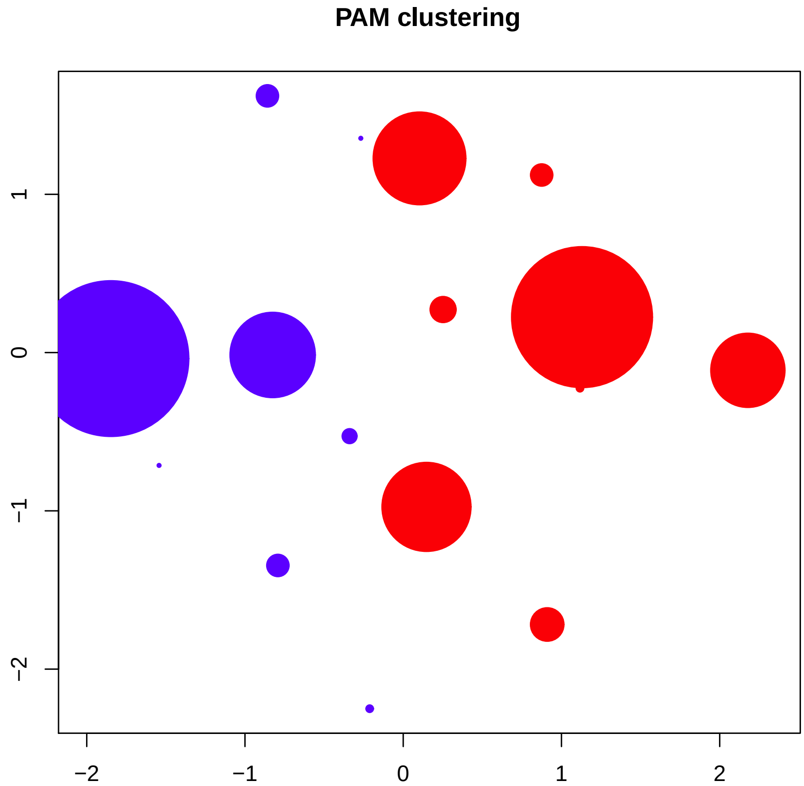


Figure 3.2: Data from Stouffer and Toby (1951) according to PAM clustering

whereas clusters yielded by PAM have size respectively equal to 85 and 131.

Table 3.2 contains the values of the Average Silhouette Width (ASW) and Pearson Gamma (PG) indexes (they will be fully described in Section 4.2). They can both be interpreted as a measure of the clustering quality; they can take values from -1 to 1: 1 indicates a good clustering, whereas -1 indicates a poor one.

In this case, PAM produced a better clustering than Latent Class Clustering in terms of similarity of observations belonging to the same cluster and dissimilarity of observations from different clusters, because both its ASW and PG are a bit higher than those relative to LCC.

Table 3.1: Classification of units from Goodman's dataset according to LCC and PAM

A	B	C	D	n_i	LCC	PAM
+	+	+	+	42	+	1
+	+	+	-	23	+	1
+	+	-	+	6	+	1
+	+	-	-	25	2	2
+	-	+	+	6	2	1
+	-	+	-	24	2	2
+	-	-	+	7	2	2
+	-	-	-	38	2	2
-	+	+	+	1	2	1
-	+	+	-	4	2	1
-	+	-	+	1	2	1
-	+	-	-	6	2	2
-	-	+	+	2	2	1
-	-	+	-	9	2	2
-	-	-	+	2	2	2
-	-	-	-	20	2	2

Table 3.2: Distance based statistics of LCC and PAM clustering of Goodman's dataset

Method	Average Silhouette Width	Pearson Gamma
Latent Class Clustering (LCC)	0.446	0.488
Partition Around Medoids (PAM)	0.493	0.585

Chapter 4

Simulations

In **clustering**, one may be interested in *the classification of similar objects into groups*, and one may be interested in *finding observations that come from the same true homogeneous distribution*.

In this framework, the main question is then: do both of these aims lead to the same clustering? And how good are clustering methods designed to fulfil one of these aims in terms of the other one?

In order to answer, two approaches, namely a **latent class model** (mixture of multinomial distributions) and a **partition around medoids** one, are evaluated and compared in a fairly wide simulation study.

4.1 Description of the study

The study would serve as a basis to understand similarities and differences in terms of classification performances of the two approaches and to detect, if any, different roles played by data features.

Basically, simulations consisted of generating several data sets from different parameterizations. Then we applied the two clustering methods and finally we compared the obtained classifications.

In particular, we have examined the impact of the following aspects:

- *number of latent classes* (2/3/5): we generated data from models with 2 and 5 latent classes, and in a few cases from 3 latent classes (namely when the too small number of variables and levels would not have allowed for 5 identified classes);

- *number of observed variables* (4/12) and *number of their categories* (2/4/8): data has been generated from models with small and large number of variables; the variables considered each time were respectively only binary, only 4-levels, only 8-levels variables and with a different number of categories;

- *entity of mixing proportions* (extremely different/equal): data sets were generated according to models that have allowed for different mixing proportions and for clusters supposed to have about the same size;

- *expected cluster separation* (clear/unclear): parameters values have been chosen with the idea of having, on one hand, a situation where clusters do not have a clear characterization (hence one would expect to have overlapped clusters) and, on the other hand, a situation where clusters have an evident characterization (therefore one would expect to have clearly separated clusters)

- *number of units for each data set* (small samples/big samples): for each of the previous framework we generated data sets with a small number of units, typically one hundred (but in a few cases two hundred or five hundred, depending on the sample size needed in order to estimate the model), and a big number of units, namely one thousand.

From the combination of all these specific features we obtain 128 settings, which we call ‘patterns’. These are schematized in Tables 4.1, 4.2, 4.3 and 4.4.

For each pattern 2000 different data sets were generated with the Latent Gold® software and the true classification of units has been recorded. Then we estimated the model according to a model-based approach with the same (commercial) software and with a distance-based method (using `pam` function, contained in the R-package `cluster`, dissimilarity measure = *manhattan*). We also estimated the model, again according to a maximum likelihood approach, with an open-source software (using an EM algorithm, implemented as a function `lcmixed` in the R-package `fpc`), with the aim of comparing results, precision and time with Latent Gold.

Latent Class analysis yields a probabilistic clustering approach. Although each object is assumed to belong to one class, it is taken into account

Table 4.1: Simulations with binary variables only - Summary

No. binary variables	No. clusters	Mixing Proportions	Cluster separation	No. units
4	2 cl	Extreme	Clear	100
				1000
			Unclear	100
				1000
		Equal	Clear	100
				1000
			Unclear	100
				1000
	3 cl	Extreme	Clear	100
				1000
			Unclear	100
				1000
		Equal	Clear	100
				1000
			Unclear	100
				1000
12	2 cl	Extreme	Clear	100
				1000
			Unclear	100
				1000
		Equal	Clear	100
				1000
			Unclear	100
				1000
	5 cl	Extreme	Clear	100
				1000
			Unclear	100
				1000
		Equal	Clear	100
				1000
			Unclear	100
				1000

Table 4.2: Simulations with 4-level variables only - Summary

No. of categories	No. of variables	No. clusters	Mixing Proportions	Cluster separation	No. units
4	4	2 cl	Extreme	Clear	100
					1000
				Unclear	100
					1000
			Equal	Clear	100
					1000
				Unclear	100
					1000
		5 cl	Extreme	Clear	100
					1000
				Unclear	100
					1000
			Equal	Clear	100
					1000
				Unclear	100
					1000
	12	2 cl	Extreme	Clear	100
					1000
				Unclear	100
					1000
			Equal	Clear	100
					1000
				Unclear	100
					1000
		5 cl	Extreme	Clear	200
					1000
				Unclear	200
					1000
			Equal	Clear	200
					1000
				Unclear	200
					1000

Table 4.3: Simulations with 8-level variables only - Summary

No. of categories	No. of variables	No. clusters	Mixing Proportions	Cluster separation	No. units
8	4	2 cl	Extreme	Clear	100 1000
				Unclear	100 1000
			Equal	Clear	100 1000
				Unclear	100 1000
		5 cl	Extreme	Clear	200 1000
				Unclear	200 1000
			Equal	Clear	200 1000
				Unclear	200 1000
	12	2 cl	Extreme	Clear	200 1000
				Unclear	200 1000
			Equal	Clear	200 1000
				Unclear	200 1000
		5 cl	Extreme	Clear	500 1000
				Unclear	500 1000
			Equal	Clear	500 1000
				Unclear	500 1000

Table 4.4: Simulations with mixed no. of level variables - Summary

No. of categories	No. of variables	No. clusters	Mixing Proportions	Cluster separation	No. units
2 33 4	4	2 cl	Extreme	Clear	100
					1000
				Unclear	100
					1000
			Equal	Clear	100
					1000
				Unclear	100
					1000
		5 cl	Extreme	Clear	100
					1000
				Unclear	100
					1000
			Equal	Clear	100
					1000
				Unclear	100
					1000
222 333 4444 88	12	2 cl	Extreme	Clear	100
					1000
				Unclear	100
					1000
			Equal	Clear	100
					1000
				Unclear	100
					1000
		5 cl	Extreme	Clear	200
					1000
				Unclear	200
					1000
			Equal	Clear	200
					1000
				Unclear	200
					1000

that there is uncertainty about a unit's class membership. For each individual the posterior class-membership probabilities are computed from the estimated model parameters and his observed score ([51]); units are thus assigned to the class with highest posterior probability.

In order to find ML estimates for the model parameters, Latent GOLD uses both EM and Newton-Raphson algorithms: the estimation process starts with 250 EM iterations. When close enough to the final solution, the program switches to Newton-Raphson, carrying on for other 50 iterations. To avoid local maxima, each process has been started from 20 different sets.

This is a way to exploit the advantages of both algorithms; that is, the stability of EM even when it is far away from the optimum and the speed of Newton-Raphson when it is close to the optimum [75]. The exact algorithm implemented in Latent GOLD works as follows. The program starts with EM until either the maximum number of EM iterations (Iteration Limits EM) or the EM convergence criterion (EM Tolerance) is reached. Then, the program switches to NR iterations which stop when the maximum number of NR iterations (Iteration Limits Newton-Raphson) or the overall converge criterion (Tolerance) is reached. The convergence criterion that is used is the sum of the absolute relative changes in the parameters. The program also stops iterating when the change in the log-posterior is negligible, i.e., smaller than 10^{-12}. The program reports the iteration process in the Iteration Detail output file listing. Thus, it can easily be checked whether the maximum number of iterations is reached without convergence. In addition, a warning is given if one of the elements of the gradient is larger than 10^{-3}. It should be noted that sometimes it is more efficient to use only the EM algorithm, which is accomplished by setting Iteration Limits Newton- Raphson = 0 in the Technical Tab. This is, for instance, the case in models with many parameters.

When using the open source software, data sets are processed through `lcmixed`, a R function contained in `fpc` package. It allows to fit a latent class mixture model, with both continuous and categorical variables. In particular, categorical ones are modelled within components by independent multinomial distributions. The fit is by maximum likelihood estimation computed with the EM-algorithm. Also in this case, 20 sets are used as starting points, in order to avoid local maxima.

4.2 Measures of comparison

Once all the models have run, in order to compare the obtained classifications we use three indexes: the Adjusted Rand Index, the Average Silhouette Width and the Pearson Gamma [43].

Adjusted Rand Index

The *Rand Index* is a measure of the similarity between two data clusterings.

Given a set of n elements $S = \{O_1, \ldots, O_n\}$ and two partitions of S to compare, $U = \{u_1, \ldots, u_R\}$ and $V = \{v_1, \ldots, v_C\}$, the following is defined:

- a, the number of pairs of elements in S that are in the same set in U and in the same set in V;
- b, the number of pairs of elements in S that are in different sets in U and in different sets in V;
- c, the number of pairs of elements in S that are in the same set in U and in different sets in V;
- d, the number of pairs of elements in S that are in different sets in U and in the same set in V;

The **Rand index**, R, is:

$$R = \frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{n}{2}}$$

Intuitively, $a+b$ can be considered as the number of agreements between U and V and $c+d$ as the number of disagreements between U and V.

The **Adjusted Rand Index** (**ARI**) is the corrected-for-chance version of the Rand index:

$$ARI = \frac{Index - Expected\ Index}{Maximum\ Index - Expected\ Index}.$$

In this context, the ARI is used to compare the classifications yielded by a model-based and a distance-based clustering approach with respect to what is recorded as ‘true’ cluster membership. The uncorrected version has a value between 0 and 1, with 0 indicating that the two data clusters do not agree on any pair of points and 1 indicating that the data clusters are exactly the same.

Average Silhouette Width

For a partition of n units into k clusters $C_1, \ldots, C_k$, suppose object i has been assigned to cluster C_h. We indicate with $a(i)$ the average dissimilarity of i to all other objects of cluster C_h:

$$a(i,h) = a(i) = \frac{1}{|C_h| - 1} \sum_{j \in C_h} d(i,j)$$

This expression makes sense only when C_h contains other objects other than i. Let consider now any cluster C_l different from C_h and define the average dissimilarity of i to all objects of C_l

$$d(i, C_l) = \frac{1}{|C_l|} \sum_{j \in C_l} d(i,j)$$

After computing $d(i, C_l)$ for all clusters C_l different from C_h, we select the smallest of those:

$$b(i) = \min_{i \notin C_l} d(i, C_l)$$

The cluster for which this minimum is obtained is call *neighbour* of object i; this is like the second-best choice for object i.

The silhouette $s(i)$ is obtained by combining $a(i)$ and $b(i)$ as follows:

$$\begin{aligned} s(i) &= 1 - \frac{a(i)}{b(i)} && \text{if } a(i) < b(i) \\ &= 0 && \text{if } a(i) = b(i) \qquad (4.1) \\ &= \frac{b(i)}{a(i)} - 1 && \text{if } a(i) > b(i) \end{aligned}$$

$$(4.2)$$

This can be rearranged in one formula

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)}.$$

And it can be easily seen that

$$-1 \leq s(i) \leq 1$$

for each object i.

When $s(i)$ is at its largest (that is, close to 1), this implies that the 'within' dissimilarity $a(i)$ is much smaller than the smallest 'between' dissimilarity $b(i)$. Therefore, we can say that i is well classified: the second best choice is not nearly as close as the actual choice.

When $s(i)$ is about zero, then $a(i)$ and $b(i)$ are approximately equal and so it is not clear whether i should have been assigned to C_h or to C_l, it lies equally far away from both.

The worst situation takes place when $s(i)$ is close to -1, when $a(i)$ is actually much larger than $b(i)$, and hence i lies on average closer to C_l than to C_h; therefore it would have seemed better to assign object i to its neighbour.

The **silhouette** $s(i)$ hence measures how well unit i has been classified. By computing the average of the $s(i)$, calculated for all the observations $i = 1, \ldots, n$, we obtain the so called **average silhouette width** (**ASW**):

$$\bar{s}(i) = \frac{1}{n} \sum_{i=1}^{n} s(i, k). \tag{4.3}$$

If k is not fixed and needs to be estimated, the ASW estimate k_{ASW} is obtained by maximizing equation 4.3. Its expression leads to a clustering that emphasises the separation between the clusters and their neighbouring clusters.

For further references see ([43], [44]).

Pearson Gamma

The **Pearson Gamma** (**PG**) index is the Pearson correlation $\rho(\mathbf{d}, \mathbf{m})$ between the vector $\mathbf{d}$ of pairwise dissimilarities and the binary vector $\mathbf{m}$ that is 0 for every pair of observations in the same cluster and 1 for every pair of observations in different clusters.

PG emphasises a good approximation of the dissimilarity structure by the clustering in the sense that observations in different clusters should be strongly correlated with large dissimilarity.

For further details see Halkidi, Batistakis and Vazirgiannis ([34]) and Hennig ([40]).

Comments

It is worth to notice that both the Average Silhouette Width and the Pearson Gamma are usually used to estimate the number of clusters. Since here the number of latent classes is assumed to be fixed and known, these indexes are used to compare the quality of the clustering.

Furthermore, since Latent Class Clustering is by definition aimed to recover the 'true' classification we expect it to perform better than PAM in terms of Adjusted Rand Index; whereas, since PAM is a distance-based approach, we expect it to perform better than Latent Gold in terms of Average Silhouette Width and Pearson Gamma.

Chapter 5

Visualization

Visualization is a key feature in clustering and it is a very useful tool in understanding data structure. Data display shows how units are located in a specific space and some considerations may spring from such a graphical representation. For example, in an exploratory phase it may give some insights in determining the appropriate number of clusters, or it may help to understand how clusters look like and which clustering method is the best in order to identify them. Of course, information coming from a graphical representation should be integrated with some theoretical information when available, so that a complete set of information is used.

Data display is not uniquely intended for an exploratory use; a common and interesting use of the graphical representations is the plot of the classification obtained from a clustering method. Visualization of the results can help in understanding and in interpreting the outcome, as well as it helps to detect uncertainty and unexpected allocations.

On the other hand, it is not always very easy to produce meaningful representations, in particular when dealing with categorical data. Indeed, by definition categorical data does not lie onto an Euclidean space and thus its representation is not straightforward.

A statistical tool that proved to produce effective representations of categorical data is the MultiDimensional Scaling.

5.1 Multidimensional Scaling

Multidimensional scaling (MDS) is a set of related statistical techniques often used for exploring similarities or dissimilarities in data. It actually concerns the problem of constructing a configuration of n points in the Euclidean space using information about the distances between the n objects.

Starting with a distance matrix $\mathbf{D}$, the aim of MDS is to find points $P_1, \ldots, P_n$ in k dimensions such that if $\hat{d}_{rs}$ denotes the Euclidean distance between P_r and P_s, then $\hat{\mathbf{D}}$ is "similar" in some sense to D. The points P_r are unknown and usually the dimension k is also unknown; in practise it is usually limited to 1,2 or 3 in order to being able to visualize the data.

The configuration produced by any MDS method is indeterminate with respect to translation, rotation, and reflection. In general, if $P_1, \ldots, P_n$ with coordinates $\mathbf{x}_i^{'} = (p_{i1}, \ldots, p_{1k})$, $i = 1, \ldots, n$ represents an MDS solution in k dimensions, then

$$\mathbf{y}_i = \mathbf{A}\mathbf{p}_i + \mathbf{b}, \quad i = 1, \ldots, n,$$

is also a solution, where $\mathbf{A}$ is an orthogonal matrix and $\mathbf{b}$ is any vector.

Two main types of solution can be distinguished: *non-metric* and *metric* methods of multidimensional scaling. The former use only the rank order of the distances

$$d_{r_1,s_1} < d_{r_2,s_2} < \ldots < d_{r_m,s_m}, \quad m = \frac{n(n-1)}{2},$$

where $(r_1, s_1), \ldots, (r_1, s_1)$ denotes all pairs of subscripts of r and s, $r < s$.

The rank orders are invariant under monotone increasing transformations f of the d_{rs}. Therefore the configurations which arise from non-metric scaling are indeterminate not only with respect to translation, rotation, and reflection, but also with respect to uniform expansion or contraction.

Differently, the metric methods are the solutions which try to obtain P_i directly from the given distances. These methods derive P_r such that, in some sense, the new distances $\hat{d}_{rs}$ between points P_r and P_s are as close to the original d_{rs} as possible.

For further details on the method see Mardia, Kent, Bibby (1979, [54]).

5.2 Graphical representations of the simulation results

In general, the purpose of MDS is to provide a “picture” which can be used to give meaningful interpretation of the data.

In this context data are simulated so there is not a proper interpretation to derive. Nevertheless, we are interested in comparing two clustering methods and in understanding possible differences, therefore a visualization of the obtained classifications is useful in this sense.

We selected one data set for each pattern and we compute the MDS, by using the function `cmdscale` (contained in the library `MASS` of the R statistical software). In order to identify differences in the allocation of the units, we plot the data by using different colours, according to the cluster memberships. In particular, for each data set we computed four different plots, one for each clustering method: the ‘true’, the LatentGold®, the PAM and the lcmixed outcomes.

Size of points is proportional to the number of points that overlap; when units of different clusters overlap the surface of the circles is divided into sectors of the corresponding colour and of width proportional to the points belonging to the corresponding cluster.

In this section we present only a selection of cases.

Figure 5.1 refers to one of the simplest cases: there are four binary variables and two clusters of different size. According to the parametrization, clusters were supposed to overlap; in fact, the true clustering reveals that some overlapping points do belong to different clusters. Furthermore it can be seen that the clustering yielded by LatentGold® and `lcmixed` looks the same. Finally, with respect to the model-based clustering, PAM has assigned a larger number of points to the ‘blue’ cluster.

In Figure 5.2 the number of clusters is three, and the true clustering shows that there are many overlapping points that have been assigned to different groups. Despite the parametrization, in the other situations clusters look well separated; again PAM tends to produce clusters of the same size, allocating more units to cluster ‘green’.

The dataset in Figure 5.3 has four variables, and there are three clusters of about the same size that are supposed to overlap. Model-based clustering looks clearer and tidier than PAM clustering; this can be due to the fact

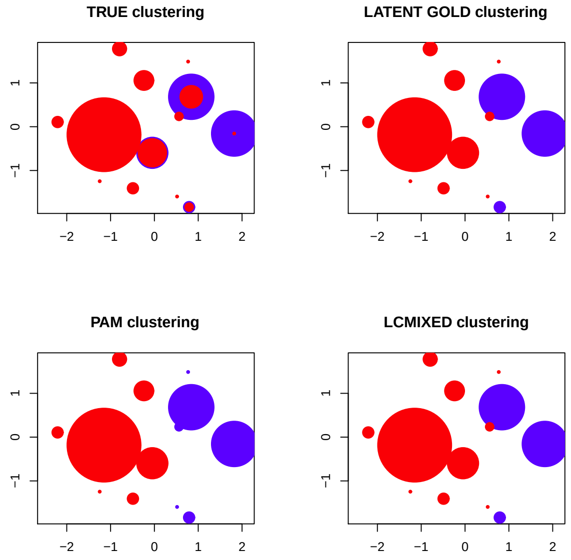


Figure 5.1: 4 binary variables, 2 clusters, different mixing proportion, unclear separation - 100 units

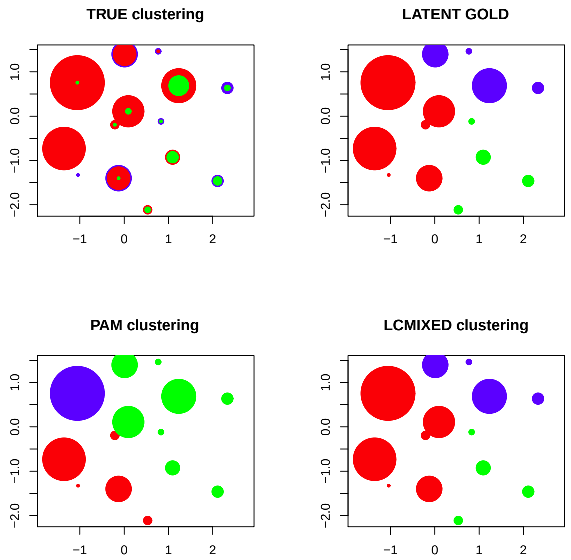


Figure 5.2: 4 binary variables, 3 clusters, different mixing proportion, unclear separation - 100 units

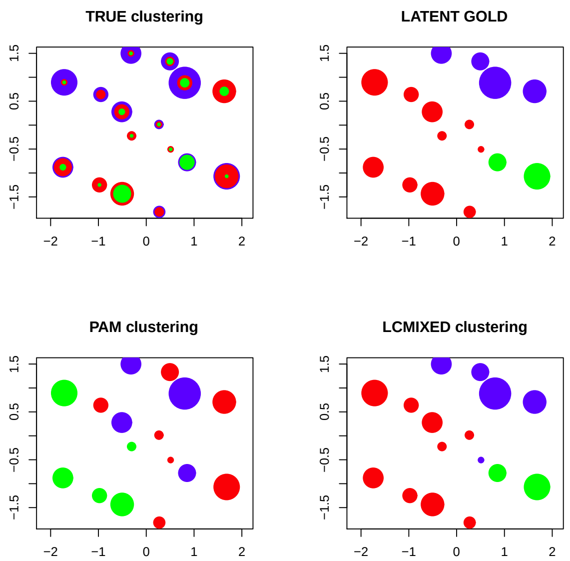


Figure 5.3: 4 binary variables, 3 clusters, equal mixing proportion, unclear separation - 100 units

that here we are considering only two dimensions.

Figure 5.4 considers a dataset with twelve binary variables. According to the parametrization, the two clusters are supposed to be clearly separated and of about the same size; the four plots are indeed very similar.

Figure 5.5 represents a dataset with twelve binary variables; there are five clusters of different size which are expected to be partially overlapped. In this framework, the cloud of points is actually quite chaotic; in the model-based clustering the five groups appear more delineated than they are in the distance-based one.

The dataset in Figure 5.6 has 12 binary variables and there are 5 well separated clusters. Indeed, across the four clustering outcomes the classification of points is very similar.

Figure 5.7 refers to a dataset with twelve 4-level variables; there are two clusters of different size that are not very well separated. In these representations, indeed, it is quite difficult to identify the two groups. This may be due to the fact that we are considering only two dimensions; it is possible that some orthogonal transformations can improve the global visualization of the clustering.

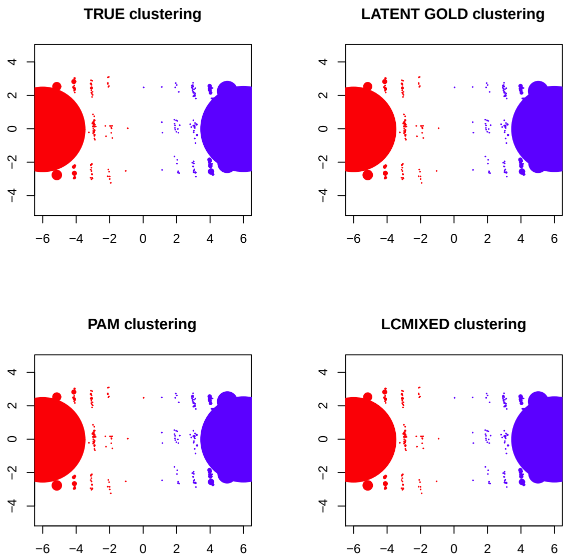


Figure 5.4: 12 binary variables, 2 clusters, equal mixing proportion, clear separation - 1000 units

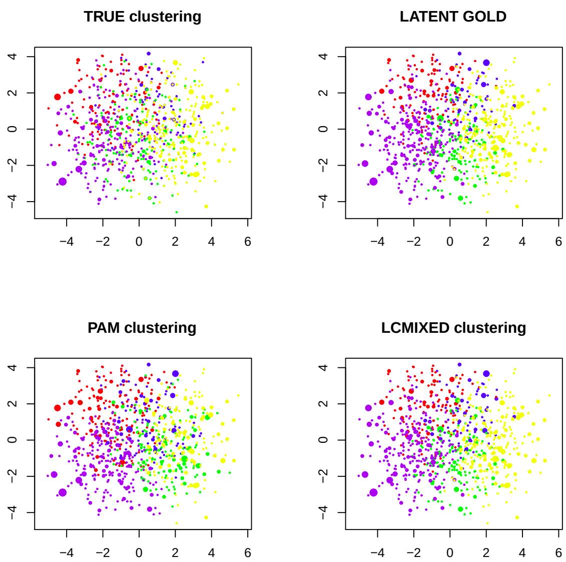


Figure 5.5: 12 binary variables, 5 clusters, different mixing proportion, unclear separation - 1000 units

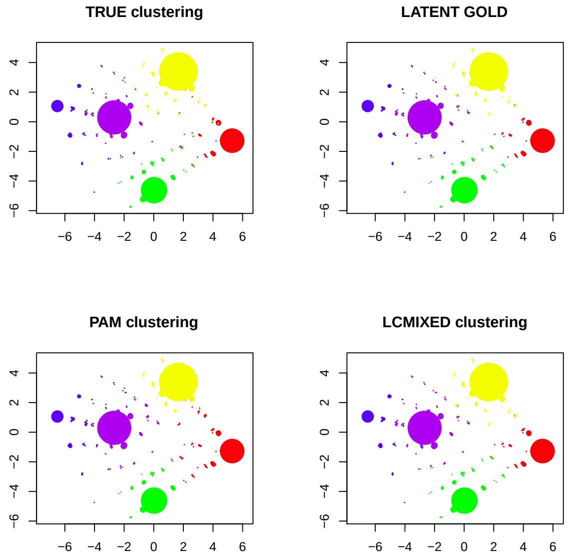


Figure 5.6: 12 binary variables, 5 clusters, different mixing proportion, clear separation - 1000 units

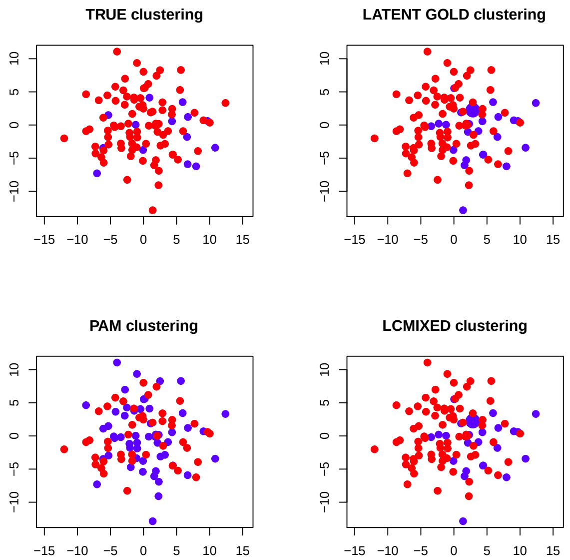


Figure 5.7: 12 4-level variables, 2 clusters, different mixing proportion, unclear separation - 100 units

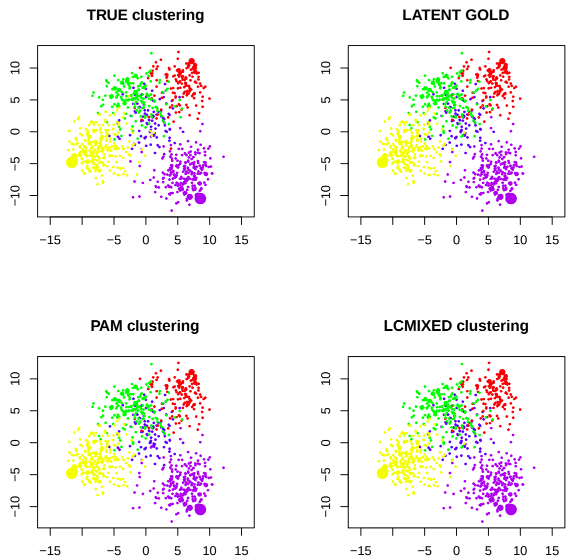


Figure 5.8: 12 4-level variables, 5 clusters, different mixing proportion, clear separation - 1000 units

Figure 5.8 refers to a dataset with twelve 4-level variables; there are five clusters that are expected to be separated. Indeed the classification is similar across the different clustering outcomes and the groups appear well defined.

Dataset in Figure 5.9 considers four categorical variables with different number of categories. The true clustering shows a high degree of uncertainty: there are many overlapped points that actually belongs to different groups. Despite this situation, the clustering yielded by the two approaches looks clean and tidy, even though there are few differences in the allocation of some units.

Dataset represented in Figure 5.10 has four variables with different number of categories. Groups are in general well defined and separated, even though the true clustering highlights a few overlapped points that belong to different clusters. The model-based and the distance-based clustering do not look much different.

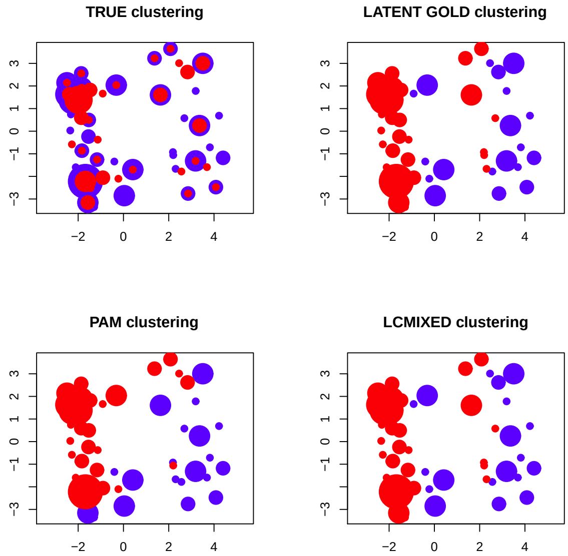


Figure 5.9: 4 mixed-no-level variables, 2 clusters, equal mixing proportion, unclear separation - 100 units

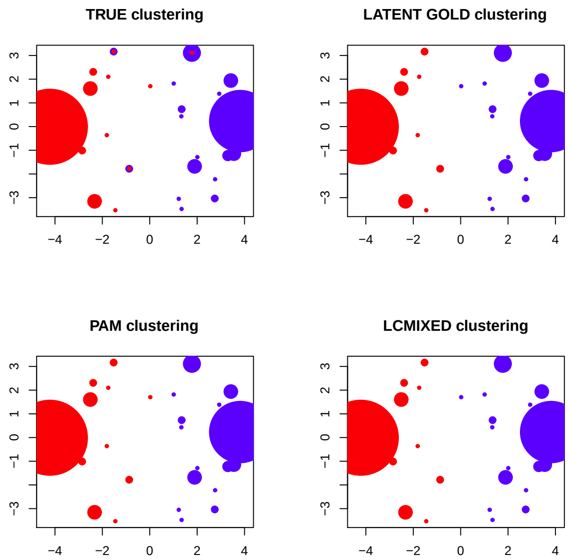


Figure 5.10: 4 mixed-no-level variables, 2 clusters, equal mixing proportion, clear separation - 100 units

Chapter 6

Results

In this chapter we present the outcomes of the simulations we carried out. Since the amount of output is very big, here we will discuss the main results only. For a detailed review see the Appendix A.

6.1 Simulation outcomes

6.1.1 Simulations with binary variables only

Table 6.1 and Table 6.2 contain the average values of the Adjusted Rand Index, Average Silhouette Width and Pearson Gamma indexes (discussed in Section 4.2) for each simulation pattern which involved binary variables only, with respectively expected unclear and clear cluster separation.

A model-based clustering approach is generally aimed to recover the 'true' probability distribution that generated the observed data and it assigns units to the latent class with highest posterior probability. Since observations are generated from a specific probability distribution (more specifically the one assumed by LCC), namely a mixture of binomial distributions, we expected Latent Gold to perform better than PAM in terms of Adjusted Rand Index, i.e. we expect Latent Gold to recover an higher proportion of 'true' memberships with respect to PAM.

By looking at the Table 6.1 we can see that values of the ARI are generally higher for Latent Gold, given the other data features.

Notice that differences between the two approaches in terms of ARI get smaller if clusters are expected to be (according to the parametrization that generated the data) clearly separated. Indeed, from Table 6.2 we can see that their values are really close to each other; nevertheless almost all of these differences are significant, because standard errors (written in brackets) are fairly small (maybe thanks to the fact that the number of n simulations for each pattern is 2000).

On the other hand, since PAM is a distance-based clustering method, we expected it to perform better than LG in terms of Average Silhouette Width and Pearson Gamma indexes.

From Table 6.1 we see that as long as the number of the considered variables is small (i.e. equal to 4) PAM actually behaved better than LG in terms of those dissimilarity criteria, even though differences are generally low.

What is surprising, according to our expectations, is not that LG in many cases performed at least as good as PAM, but rather that in some other cases it was even much better. On one hand this is encouraging, because it means that whenever we are interested in finding the true clustering we often success in getting clusters that are internally homogenous. On the other hand, this is not a very good point for PAM.

When clusters are expected to be clearly separated the two approaches generally yielded similar results, even though there are cases where LG was slightly better.

6.1.2 Simulations with 4-level variables only

When the number of categories for each observed variable increased from two to four, performances of the two approaches did not change so much and we can make considerations similar to those of the previous section. Details are in Table 6.3 and in Table 6.4.

The only difference is that PAM performed a little bit better in terms of ASW and PG when clusters were not expected to overlap.

6.1.3 Simulations with 8-level variables only

When the number of categories for each observed variable increased to eight, from Table 6.5 we can see that both Latent Gold and PAM are less able to find the true clustering, since values of the ARI are lower than those of Table 6.1 or of Table 6.3.

As we observed in Section 6.1.1, PAM shows its better performance in

Table 6.1: Average values (and their standard errors) of ARI, ASW and PG indexes: **binary** variables only and **unclear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.590 (.003)	.440 (.004)	.374 (.001)	.464 (.001)	.490 (.001)	.394 (.002)	.516 (.002)	.583 (.001)
			Big	.620 (.001)	.505 (.002)	.375 (.000)	.443 (.001)	.496 (.001)	.395 (.000)	.478 (.001)	.587 (.001)
		Equal	Small	.112 (.002)	.089 (.002)	.160 (.001)	.341 (.002)	.336 (.001)	.184 (.001)	.373 (.003)	.401 (.001)
			Big	.161 (.002)	.009 (.001)	.161 (.000)	.327 (.002)	.305 (.000)	.184 (.000)	.358 (.003)	.353 (.001)
	3	Different	Small	.150 (.002)	.094 (.002)	.107 (.001)	.370 (.002)	.395 (.001)	.229 (.001)	.505 (.002)	.511 (.001)
			Big	.156 (.003)	.062 (.001)	.128 (.000)	.328 (.002)	.364 (.001)	.231 (.000)	.433 (.002)	.477 (.001)
		Equal	Small	.109 (.001)	.114 (.001)	.042 (.001)	.324 (.002)	.373 (.001)	.163 (.001)	.476 (.002)	.520 (.001)
			Big	.134 (.001)	.120 (.001)	.048 (.000)	.292 (.002)	.369 (.000)	.163 (.000)	.421 (.002)	.513 (.000)
12	2	Different	Small	.224 (.005)	.005 (.001)	.051 (.000)	.071 (.000)	.091 (.000)	.104 (.001)	.145 (.001)	.206 (.001)
			Big	.637 (.001)	.007 (.001)	.075 (.000)	.088 (.000)	.049 (.000)	.190 (.000)	.209 (.000)	.135 (.000)
		Equal	Small	.060 (.002)	.026 (.001)	.032 (.000)	.072 (.000)	.060 (.000)	.091 (.001)	.183 (.001)	.168 (.001)
			Big	.264 (.001)	.025 (.001)	.032 (.000)	.056 (.000)	.050 (.000)	.092 (.000)	.154 (.000)	.137 (.000)
	5	Different	Small	.160 (.001)	.149 (.001)	.018 (.000)	.153 (.001)	.147 (.000)	.229 (.001)	.393 (.001)	.378 (.001)
			Big	.253 (.001)	.177 (.001)	.034 (.000)	.099 (.001)	.128 (.000)	.230 (.000)	.338 (.001)	.339 (.000)
		Equal	Small	.140 (.001)	.137 (.001)	.021 (.000)	.151 (.001)	.147 (.000)	.210 (.001)	.386 (.001)	.376 (.001)
			Big	.212 (.001)	.168 (.001)	.035 (.000)	.105 (.001)	.129 (.000)	.210 (.000)	.332 (.001)	.340 (.000)

Table 6.2: Average values (and their standard errors) of ARI, ASW and PG indexes: **binary** variables only and **clear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.914 (.002)	.819 (.002)	.728 (.001)	.739 (.001)	.731 (.001)	.824 (.001)	.838 (.001)	.839 (.001)
			Big	.935 (.000)	.819 (.001)	.729 (.000)	.739 (.000)	.732 (.000)	.826 (.000)	.838 (.000)	.840 (.000)
		Equal	Small	.898 (.001)	.896 (.001)	.754 (.001)	.763 (.001)	.761 (.001)	.862 (.001)	.871 (.001)	.869 (.001)
			Big	.898 (.000)	.897 (.000)	.754 (.000)	.762 (.000)	.761 (.000)	.862 (.000)	.870 (.000)	.869 (.000)
	3	Different	Small	.544 (.002)	.469 (.003)	.446 (.001)	.556 (.001)	.560 (.001)	.588 (.001)	.668 (.001)	.651 (.001)
			Big	.580 (.001)	.479 (.002)	.450 (.000)	.540 (.001)	.540 (.001)	.589 (.000)	.658 (.000)	.646 (.001)
		Equal	Small	.554 (.002)	.546 (.002)	.469 (.001)	.578 (.001)	.540 (.001)	.598 (.001)	.674 (.001)	.650 (.001)
			Big	.556 (.001)	.546 (.001)	.470 (.000)	.570 (.001)	.535 (.001)	.597 (.000)	.667 (.000)	.648 (.000)
12	2	Different	Small	.980 (.001)	.912 (.002)	.398 (.001)	.398 (.001)	.395 (.001)	.655 (.001)	.656 (.001)	.664 (.001)
			Big	.988 (.001)	.934 (.001)	.399 (.000)	.400 (.000)	.399 (.000)	.658 (.000)	.659 (.000)	.670 (.000)
		Equal	Small	.983 (.001)	.952 (.001)	.411 (.001)	.411 (.001)	.413 (.001)	.739 (.001)	.740 (.001)	.743 (.001)
			Big	.988 (.000)	.961 (.000)	.412 (.000)	.412 (.000)	.414 (.000)	.740 (.000)	.741 (.000)	.745 (.000)
	5	Different	Small	.844 (.001)	.846 (.001)	.461 (.001)	.476 (.001)	.470 (.001)	.643 (.001)	.650 (.000)	.646 (.000)
			Big	.869 (.000)	.844 (.001)	.467 (.000)	.479 (.000)	.476 (.000)	.643 (.000)	.652 (.000)	.646 (.000)
		Equal	Small	.912 (.001)	.922 (.001)	.536 (.001)	.543 (.001)	.542 (.001)	.677 (.000)	.681 (.000)	.680 (.000)
			Big	.922 (.001)	.921 (.000)	.540 (.000)	.547 (.000)	.546 (.000)	.676 (.000)	.681 (.000)	.680 (.000)

Table 6.3: Average values (and their standard errors) of ARI, ASW and PG indexes: **4-levels** variables only and **unclear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.080 (.003)	.009 (.001)	.071 (.001)	.165 (.001)	.160 (.000)	.120 (.001)	.267 (.001)	.272 (.001)
			Big	.232 (.002)	-.003 (.000)	.073 (.000)	.135 (.000)	.142 (.000)	.121 (.000)	.195 (.001)	.230 (.001)
		Equal	Small	.023 (.001)	.012 (.001)	.029 (.000)	.158 (.001)	.152 (.000)	.051 (.001)	.256 (.001)	.264 (.001)
			Big	.033 (.001)	.009 (.000)	.029 (.000)	.104 (.001)	.133 (.000)	.050 (.000)	.146 (.001)	.224 (.002)
	5	Different	Small	.025 (.001)	.023 (.000)	-.053 (.000)	.161 (.001)	.163 (.000)	.056 (.000)	.390 (.001)	.395 (.001)
			Big	.025 (.000)	.021 (.000)	-.026 (.001)	.076 (.000)	.126 (.000)	.056 (.001)	.266 (.001)	.320 (.000)
		Equal	Small	.053 (.001)	.051 (.001)	-.035 (.000)	.166 (.001)	.168 (.000)	.098 (.001)	.399 (.001)	.404 (.001)
			Big	.063 (.000)	.057 (.000)	-.014 (.000)	.090 (.001)	.136 (.000)	.098 (.000)	.294 (.001)	.340 (.000)
12	2	Different	Small	.224 (.005)	.005 (.001)	.051 (.000)	.071 (.000)	.091 (.000)	.104 (.001)	.145 (.001)	.206 (.001)
			Big	.637 (.001)	.007 (.001)	.075 (.000)	.088 (.000)	.049 (.000)	.190 (.000)	.209 (.000)	.135 (.000)
		Equal	Small	.060 (.002)	.026 (.001)	.032 (.000)	.072 (.000)	.060 (.000)	.091 (.001)	.183 (.001)	.168 (.001)
			Big	.264 (.001)	.025 (.001)	.032 (.000)	.056 (.000)	.050 (.000)	.092 (.000)	.154 (.000)	.137 (.000)
	5	Different	Small	.073 (.001)	.040 (.000)	.001 (.000)	.054 (.000)	.045 (.000)	.123 (.000)	.248 (.000)	.224 (.000)
			Big	.159 (.001)	.043 (.000)	.010 (.000)	.032 (.000)	.039 (.000)	.123 (.000)	.189 (.000)	.190 (.000)
		Equal	Small	.063 (.001)	.039 (.000)	.000 (.000)	.054 (.000)	.046 (.000)	.111 (.000)	.248 (.001)	.225 (.000)
			Big	.115 (.001)	.043 (.000)	.006 (.000)	.031 (.000)	.039 (.000)	.110 (.001)	.191 (.000)	.194 (.000)

Table 6.4: Average values (and their standard errors) of ARI, ASW and PG indexes: **4-levels** variables only and **clear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.628 (.004)	.169 (.006)	.257 (.001)	.278 (.001)	.197 (.001)	.424 (.001)	.463 (.001)	.310 (.003)
			Big	.740 (.001)	-.024 (.000)	.258 (.000)	.281 (.000)	.147 (.000)	.425 (.000)	.456 (.000)	.196 (.000)
		Equal	Small	.644 (.002)	.098 (.001)	.200 (.001)	.218 (.002)	.522 (.000)	.209 (.001)	.230 (.002)	.480 (.001)
			Big	.711 (.001)	.093 (.000)	.200 (.000)	.212 (.000)	.507 (.000)	.208 (.000)	.221 (.001)	.458 (.000)
	5	Different	Small	.463 (.002)	.539 (.002)	.245 (.001)	.316 (.001)	.322 (.001)	.517 (.001)	.584 (.001)	.592 (.003)
			Big	.564 (.001)	.554 (.001)	.254 (.000)	.328 (.000)	.330 (.000)	.518 (.000)	.598 (.000)	.597 (.000)
		Equal	Small	.505 (.002)	.604 (.002)	.269 (.001)	.331 (.001)	.342 (.001)	.528 (.001)	.591 (.001)	.606 (.001)
			Big	.603 (.001)	.609 (.001)	.278 (.000)	.350 (.000)	.350 (.000)	.529 (.000)	.610 (.000)	.607 (.000)
12	2	Different	Small	.980 (.001)	.912 (.002)	.398 (.001)	.398 (.001)	.395 (.001)	.655 (.001)	.656 (.001)	.664 (.001)
			Big	.988 (.001)	.934 (.000)	.399 (.000)	.400 (.000)	.399 (.000)	.658 (.000)	.659 (.000)	.670 (.000)
		Equal	Small	.983 (.001)	.952 (.001)	.411 (.001)	.412 (.001)	.413 (.001)	.739 (.001)	.740 (.001)	.743 (.001)
			Big	.988 (.000)	.962 (.000)	.412 (.000)	.412 (.000)	.414 (.000)	.740 (.000)	.741 (.000)	.745 (.000)
	5	Different	Small	.938 (.001)	.941 (.001)	.332 (.000)	.334 (.000)	.333 (.000)	.734 (.000)	.736 (.000)	.736 (.000)
			Big	.958 (.000)	.957 (.000)	.335 (.000)	.338 (.000)	.337 (.000)	.735 (.000)	.739 (.000)	.738 (.000)
		Equal	Small	.929 (.001)	.935 (.001)	.328 (.000)	.330 (.000)	.329 (.000)	.719 (.000)	.721 (.000)	.720 (.000)
			Big	.952 (.001)	.952 (.000)	.330 (.000)	.333 (.000)	.333 (.000)	.720 (.000)	.724 (.000)	.723 (.000)

terms of ASW and PG when the number of variables is fairly small. But not when clusters are supposed to be separated (Table 6.6): in those cases, surprisingly PAM performed at most as well as LG. Values themselves are not low, they are actually very good, but no longer better than those from LG clustering.

When the number of variables is 12, again PAM performed a little bit worse than LG.

6.1.4 Simulations with variables having different number of categories

Lastly, we have considered the case where the variables do not have the same number of categories, and specifically with:

- 4 variables, we had one binary variable, two variables with three categories and one variable with four categories;
- 12 variables, we had three binary variables, three variables with three levels, four variables with four categories and two variables with eight categories

In this framework, again Latent Gold performed better than PAM in finding the true clustering, but the outcome of the latter was not much worse (see Table 6.7). It has to be said that the average performances of the two approaches are much higher if we consider the situations where clusters are supposed to be clearly separated (see Table 6.8).

According to dissimilarity criteria, PAM clustering scored a bit higher values than LG for both ASW and PG as far as we consider a small number of variables; when we include 12 variables the two approaches performed about the same.

When the clusters are expected to be clearly separated (see Table 6.8) there is no particular evidence to prefer one of the two methods in terms of ASW and PG, because values are about the same here, too.

6.1.5 General considerations

Overall, the simulations we carried out tell us that, in terms of recovering the ‘true’ clustering (according to a ‘true’ unknown model), the Latent Class Clustering generally behaves better, yielding better results in terms of ARI,

Table 6.5: Average values (and their standard errors) of ARI, ASW and PG indexes: **8-levels** variables only and **unclear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.016 (.001)	-.001 (.001)	.026 (.000)	.099 (.000)	.106 (.000)	.058 (.001)	.221 (.001)	.229 (.001)
			Big	.032 (.001)	-.003 (.000)	.027 (.000)	.066 (.000)	.095 (.000)	.058 (.000)	.125 (.001)	.204 (.000)
		Equal	Small	.013 (.001)	.006 (.001)	.013 (.000)	.086 (.000)	.076 (.000)	.035 (.000)	.220 (.001)	.184 (.001)
			Big	.024 (.001)	.005 (.000)	.014 (.000)	.052 (.000)	.063 (.000)	.036 (.000)	.110 (.001)	.146 (.000)
	5	Different	Small	.030 (.000)	.024 (.000)	-.016 (.000)	.075 (.000)	.076 (.000)	.072 (.000)	.283 (.000)	.283 (.000)
			Big	.044 (.000)	.007 (.000)	-.005 (.000)	.045 (.000)	.059 (.000)	.071 (.000)	.205 (.000)	.137 (.000)
		Equal	Small	.031 (.000)	.026 (.000)	-.013 (.000)	.075 (.000)	.075 (.000)	.072 (.000)	.283 (.000)	.280 (.000)
			Big	.042 (.000)	.009 (.000)	-.004 (.000)	.046 (.000)	.056 (.000)	.072 (.000)	.207 (.000)	.133 (.000)
12	2	Different	Small	.135 (.004)	.004 (.001)	.037 (.000)	.038 (.000)	.029 (.000)	.140 (.000)	.151 (.001)	.115 (.000)
			Big	.586 (.001)	.002 (.000)	.038 (.000)	.044 (.000)	.026 (.000)	.140 (.000)	.154 (.000)	.101 (.000)
		Equal	Small	.579 (.002)	.105 (.002)	.046 (.000)	.051 (.000)	.030 (.000)	.203 (.000)	.224 (.000)	.129 (.001)
			Big	.710 (.001)	.128 (.002)	.046 (.000)	.051 (.000)	.029 (.000)	.204 (.000)	.224 (.000)	.122 (.000)
	5	Different	Small	.201 (.001)	.050 (.000)	.015 (.000)	.026 (.000)	.019 (.000)	.151 (.000)	.186 (.000)	.151 (.000)
			Big	.350 (.001)	.054 (.000)	.016 (.000)	.026 (.000)	.019 (.000)	.151 (.000)	.185 (.000)	.142 (.000)
		Equal	Small	.137 (.001)	.034 (.000)	.010 (.000)	.024 (.000)	.019 (.000)	.124 (.000)	.175 (.000)	.144 (.000)
			Big	.263 (.001)	.036 (.000)	.012 (.000)	.022 (.000)	.018 (.000)	.124 (.000)	.165 (.000)	.134 (.000)

Table 6.6: Average values (and their standard errors) of ARI, ASW and PG indexes: **8-levels** variables only and **clear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.680 (.004)	.322 (.008)	.294 (.001)	.284 (.001)	.213 (.002)	.481 (.001)	.503 (.002)	.358 (.004)
			Big	.858 (.001)	.061 (.006)	.295 (.000)	.305 (.000)	.145 (.001)	.484 (.000)	.495 (.000)	.206 (.003)
		Equal	Small	.787 (.002)	.802 (.002)	.335 (.001)	.341 (.001)	.343 (.001)	.626 (.001)	.634 (.001)	.639 (.001)
			Big	.848 (.001)	.801 (.001)	.336 (.001)	.348 (.001)	.343 (.001)	.627 (.000)	.650 (.000)	.638 (.000)
	5	Different	Small	.709 (.001)	.768 (.001)	.320 (.001)	.333 (.001)	.344 (.001)	.667 (.001)	.684 (.001)	.699 (.001)
			Big	.777 (.001)	.262 (.001)	.324 (.000)	.346 (.000)	.189 (.000)	.667 (.000)	.698 (.000)	.342 (.000)
		Equal	Small	.704 (.001)	.766 (.001)	.321 (.001)	.334 (.001)	.344 (.001)	.659 (.001)	.674 (.001)	.688 (.001)
			Big	.764 (.000)	.172 (.000)	.324 (.000)	.348 (.000)	.149 (.000)	.659 (.000)	.692 (.000)	.274 (.000)
12	2	Different	Small	.992 (.000)	.984 (.001)	.340 (.000)	.340 (.000)	.339 (.000)	.735 (.001)	.735 (.001)	.735 (.001)
			Big	.996 (.000)	.987 (.000)	.340 (.000)	.340 (.000)	.339 (.000)	.735 (.000)	.735 (.000)	.736 (.000)
		Equal	Small	.995 (.000)	.995 (.000)	.389 (.000)	.389 (.000)	.389 (.000)	.842 (.000)	.842 (.000)	.842 (.000)
			Big	.994 (.001)	.996 (.000)	.389 (.000)	.389 (.000)	.389 (.000)	.842 (.000)	.842 (.000)	.842 (.000)
	5	Different	Small	.990 (.000)	.990 (.000)	.358 (.000)	.358 (.000)	.358 (.000)	.841 (.000)	.841 (.000)	.841 (.000)
			Big	.992 (.000)	.991 (.000)	.358 (.000)	.358 (.000)	.358 (.000)	.841 (.000)	.841 (.000)	.841 (.000)
		Equal	Small	.989 (.000)	.989 (.000)	.357 (.000)	.358 (.000)	.357 (.000)	.835 (.000)	.835 (.000)	.835 (.000)
			Big	.992 (.000)	.991 (.000)	.358 (.000)	.358 (.000)	.358 (.000)	.835 (.000)	.835 (.000)	.835 (.000)

Table 6.7: Average values (and their standard errors) of ARI, ASW and PG indexes: **mixed no.-levels** variables only and **unclear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.052 (.002)	.028 (.001)	.067 (.001)	.214 (.001)	.222 (.001)	.094 (.001)	.284 (.002)	.334 (.001)
			Big	.105 (.002)	.014 (.000)	.160 (.001)	.192 (.000)	.185 (.001)	.182 (.002)	.282 (.001)	.266 (.001)
		Equal	Small	.016 (.001)	.011 (.001)	.028 (.000)	.202 (.001)	.215 (.001)	.040 (.001)	.272 (.002)	.328 (.001)
			Big	.013 (.000)	.003 (.000)	.029 (.000)	.121 (.002)	.190 (.000)	.040 (.000)	.142 (.003)	.287 (.000)
	5	Different	Small	.035 (.001)	.035 (.001)	-.068 (.000)	.217 (.001)	.237 (.001)	.070 (.001)	.443 (.001)	.452 (.001)
			Big	.031 (.000)	.033 (.000)	-.036 (.000)	.109 (.002)	.209 (.000)	.070 (.000)	.337 (.002)	.407 (.000)
		Equal	Small	.040 (.001)	.042 (.001)	-.059 (.000)	.220 (.001)	.236 (.001)	.081 (.001)	.446 (.001)	.452 (.001)
			Big	.035 (.000)	.042 (.000)	-.037 (.000)	.109 (.001)	.206 (.000)	.080 (.000)	.338 (.002)	.404 (.000)
12	2	Different	Small	.121 (.004)	.027 (.001)	.070 (.000)	.084 (.000)	.074 (.000)	.164 (.001)	.200 (.001)	.194 (.001)
			Big	.521 (.001)	.012 (.001)	.071 (.000)	.089 (.000)	.057 (.000)	.165 (.000)	.192 (.001)	.146 (.000)
		Equal	Small	.281 (.003)	.118 (.002)	.071 (.000)	.093 (.000)	.085 (.000)	.194 (.001)	.238 (.001)	.230 (.001)
			Big	.517 (.001)	.176 (.002)	.072 (.000)	.091 (.000)	.082 (.000)	.195 (.000)	.246 (.000)	.223 (.000)
	5	Different	Small	.132 (.001)	.104 (.001)	.023 (.000)	.058 (.000)	.060 (.000)	.202 (.000)	.263 (.000)	.275 (.000)
			Big	.331 (.001)	.128 (.000)	.028 (.000)	.055 (.000)	.054 (.000)	.202 (.000)	.264 (.000)	.251 (.000)
		Equal	Small	.145 (.001)	.111 (.000)	.023 (.000)	.061 (.000)	.061 (.000)	.205 (.000)	.281 (.001)	.287 (.000)
			Big	.319 (.001)	.141 (.001)	.027 (.000)	.055 (.000)	.054 (.000)	.206 (.000)	.273 (.000)	.265 (.000)

Table 6.8: Average values (and their standard errors) of ARI, ASW and PG indexes: **mixed no.-levels** variables only and **clear** cluster separation

No.Var	K	Mixing Prop.	No.obs	ARIlg	ARIpam	ASWtr	ASWlg	ASWpam	PGtr	PGlg	PGpam
4	2	Different	Small	.851 (.002)	.802 (.002)	.570 (.001)	.584 (.001)	.577 (.001)	.725 (.001)	.745 (.001)	.753 (.001)
			Big	.891 (.000)	.806 (.001)	.572 (.000)	.587 (.000)	.578 (.000)	.728 (.000)	.750 (.000)	.754 (.000)
		Equal	Small	.844 (.002)	.856 (.002)	.573 (.001)	.589 (.001)	.588 (.001)	.764 (.001)	.783 (.001)	.784 (.001)
			Big	.859 (.000)	.855 (.001)	.573 (.000)	.590 (.000)	.588 (.000)	.763 (.000)	.787 (.000)	.784 (.000)
	5	Different	Small	.662 (.002)	.689 (.002)	.412 (.001)	.526 (.001)	.525 (.001)	.663 (.001)	.727 (.001)	.697 (.001)
			Big	.681 (.000)	.694 (.001)	.418 (.000)	.512 (.000)	.527 (.000)	.664 (.000)	.727 (.000)	.696 (.000)
		Equal	Small	.666 (.002)	.703 (.002)	.435 (.001)	.499 (.001)	.501 (.001)	.627 (.001)	.679 (.001)	.679 (.001)
			Big	.710 (.001)	.703 (.001)	.440 (.000)	.509 (.000)	.505 (.000)	.627 (.000)	.684 (.000)	.680 (.000)
12	2	Different	Small	.998 (.000)	.998 (.000)	.596 (.000)	.596 (.000)	.596 (.000)	.880 (.001)	.880 (.001)	.880 (.001)
			Big	.999 (.000)	.997 (.000)	.596 (.000)	.596 (.000)	.596 (.000)	.882 (.000)	.882 (.000)	.882 (.000)
		Equal	Small	.998 (.000)	.998 (.000)	.597 (.000)	.597 (.000)	.597 (.000)	.916 (.000)	.916 (.000)	.917 (.000)
			Big	.998 (.001)	.998 (.000)	.596 (.000)	.596 (.000)	.596 (.000)	.916 (.000)	.916 (.000)	.916 (.000)
	5	Different	Small	.982 (.000)	.985 (.000)	.475 (.000)	.476 (.000)	.476 (.000)	.819 (.000)	.819 (.000)	.819 (.000)
			Big	.988 (.000)	.986 (.000)	.477 (.000)	.478 (.000)	.477 (.000)	.819 (.000)	.820 (.000)	.820 (.000)
		Equal	Small	.981 (.000)	.985 (.000)	.476 (.000)	.477 (.000)	.477 (.000)	.811 (.000)	.811 (.000)	.811 (.000)
			Big	.987 (.001)	.986 (.000)	.478 (.000)	.478 (.000)	.478 (.000)	.811 (.000)	.812 (.000)	.812 (.000)

even when the clusters are supposed to overlap. When clusters are expected to be somehow separated, then a PAM approach would not make the results worse.

PAM's performances improve when the mixing proportions of the components of the mixture that generate the data are about the same, i.e. when the clusters have about the same size. Maybe this is due to the fact that in general PAM seems to provide equal-sized clusters (similarly to what the information criterion clustering does, in Celeux [14]).

What is more surprising is that Latent Gold, by trying to put together observations coming from the same distribution, accomplished to get similar observations together and to separate objects that are very different in a way that is not much worse than a distance-based method (as PAM) usually does, and actually sometimes LG works even better.

Of course this does not mean that PAM should not be used anymore, since there are still situations in which it works better than Latent Gold. Moreover, it has to be considered that in this simulation study we imposed the number of classes, according to a parameterization (and so to a probabilistic model); this does not necessarily exclude that if we allowed the number of classes k to vary PAM could have worked better.

Finally, notice that both LG and PAM obtained values of ASW and PG higher than the true clustering. It does not mean that they are better than the 'truth', but sometimes observations coming from different groups are more similar to each other than objects in the same class. The amount of quality they 'lose' by not finding the true class membership they gain in terms of similarity/dissimilarity, which is a good tradeoff.

6.2 Analysis of Variance of the differences between Latent Class Clustering and Partition Around Medoids

In this chapter we have discussed the results of the simulation study and from the considerations we made it results that performances (in terms of quality of clustering) of the two approaches highly depend on data features, even though the direction of the dependence is not always very clear.

In particular, we wonder whether these characteristics may significantly affect the differences between the two approaches and, in case, which are their directions. Hence, in order to improve our understanding of the problem, we think that an analysis of variance on the differences between the indexes we calculated in the Latent Gold and in the PAM clustering outcomes might help to individuate these determinants.

Operatively, we arranged a new data set that contains a summary of the whole simulation study. Each record is a single simulation, thus the database has 256000 rows, since we had 128 patterns times 2000 simulations for each setting. For each row we recorded the value of the following dependent variables:

- the Adjusted Rand Index, for both Latent Gold and PAM clustering, evaluated with respect to the true class membership;
- the Average Silhouette Width index, for the true clustering and for both Latent Gold and PAM;
- the Pearson Gamma index, for the true clustering and for both Latent Gold and PAM.

The factors we included are the data features coded as follows:

- number of variables
  a) 4
  b) 12
- number of categories
  a) binary
  b) 4-levels
  c) 8-levels
  d) mixed number of levels
- number of clusters
  a) small (2)
  b) big (3/5)
- sample size

a) small (100/200/500 units)

b) big (1000 units)

- mixing proportions

a) extremely different

b) equal

- cluster separation

a) unclear

b) clear

6.2.1 Anova on ARI: LG-PAM

Starting from this kind of data we performed an analysis of variance on the difference between the values of the ARI in the Latent Gold and PAM clustering, and we included all the variables we have just described and all the first-order interactions as covariates. Table 6.9 contains the summary of the function.

From Table 6.9 we can see that all the factors are highly significant, and all the interaction terms - other than the number of categories×the sample size - are significant too; this may be partially due to the fact that the number of units is very large.

If we look at the Mean Square column we can see that the factor with the highest effect on the dependent variable is the interaction term of number of variables×the cluster separation; if we concentrate only on the additive effects the more important component is the number of latent classes, followed by the sample size and the entity of the mixing components. The number of variables and the cluster separation taken as additive effects do not affect the outcome more than the other data features.

Since the Anova Table does not give any information about the direction of the dependencies, we plot the mean values of the Adjusted Rand Index for LG and PAM clustering separately of each features (see Figure 6.1).

The blue and the red lines refer respectively to the Latent Gold and to the PAM clustering. From the plots we can say that the number of variables, the number of categories and the cluster separation do not lead to different directions of the mean values of the index; as we have seen from the results

Table 6.9: Analysis of Variance Table - Response: ARI of LG-PAM with interactions

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
No.var	1	7.584	7.584	350.146	0.000
No.cat	3	439.729	146.576	6767.524	0.000
No.cl	1	835.380	835.380	38569.990	0.000
Sample.size	1	528.138	528.138	24384.471	0.000
Mixing.prop	1	167.626	167.626	7739.392	0.000
Separation	1	119.336	119.336	5509.817	0.000
No.var:No.cat	3	219.658	73.219	3380.588	0.000
No.var:No.cl	1	12.988	12.988	599.643	0.000
No.var:Sample.size	1	0.041	0.041	1.893	0.169
No.var:Mixing.prop	1	26.181	26.181	1208.771	0.000
No.var:Separation	1	1559.903	1559.903	72021.677	0.000
No.cat:No.cl	3	338.083	112.694	5203.159	0.000
No.cat:Sample.size	3	144.233	48.078	2219.768	0.000
No.cat:Mixing.prop	3	18.498	6.166	284.685	0.000
No.cat:Separation	3	192.837	64.279	2967.806	0.000
No.cl:Sample.size	1	8.279	8.279	382.229	0.000
No.cl:Mixing.prop	1	79.439	79.439	3667.757	0.000
No.cl:Separation	1	37.644	37.644	1738.030	0.000
Sample.size:Mixing.prop	1	29.433	29.433	1358.959	0.000
Sample.size:Separation	1	11.069	11.069	511.049	0.000
Mixing.prop:Separation	1	1.125	1.125	51.937	0.000
Residuals	255966	5543.917	0.022		

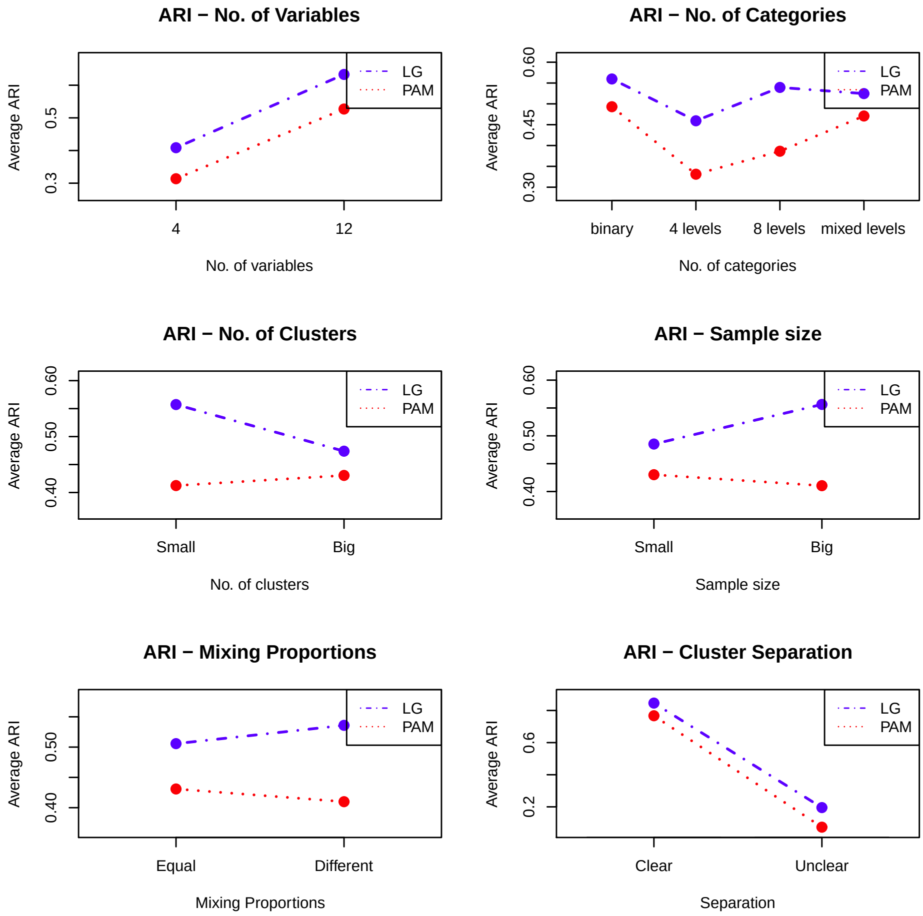


Figure 6.1: Adjusted Rand Index: average values according to data features

of the simulations, when clusters are supposed to be clearly separated both indexes work much better.

Instead, if an increasing number of clusters determinates a decrease in the average value of the ARI for the LG clustering, it leads to an improvement for the ARI of the PAM clustering. Furthermore, LG works better -on average- with big samples, whereas PAM has a higher mean value of ARI with small samples. As we notice from the outcome of the simulations PAM works better when clusters have more or less the same size, whereas LG gives better results on average when the mixing proportions are extremely different.

6.2.2 Anova on ASW: LG-PAM

Since we wanted to investigate further the differences between the two clustering outcomes, we performed an analysis of variance also on the differences between the Average Silhouette Width indexes for the two approaches. A summary of the outcome is shown in Table 6.10.

Table 6.10: Analysis of Variance Table - Response: ASW of LG-PAM with interactions

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
No.var	1	11.023	11.023	4216.653	0.000
No.cat	3	41.782	13.927	5327.727	0.000
No.cl	1	0.016	0.016	6.262	0.012
Sample.size	1	0.039	0.039	14.825	0.000
Mixing.prop	1	15.972	15.972	6109.778	0.000
Separation	1	9.145	9.145	3498.253	0.000
No.var:No.cat	3	34.595	11.532	4411.262	0.000
No.var:No.cl	1	6.111	6.111	2337.678	0.000
No.var:Sample.size	1	0.039	0.039	15.088	0.000
No.var:Mixing.prop	1	9.866	9.866	3773.923	0.000
No.var:Separation	1	27.280	27.280	10435.566	0.000
No.cat:No.cl	3	8.279	2.760	1055.624	0.000
No.cat:Sample.size	3	14.701	4.900	1874.611	0.000
No.cat:Mixing.prop	3	29.615	9.872	3776.204	0.000
No.cat:Separation	3	24.402	8.134	3111.484	0.000
No.cl:Sample.size	1	0.057	0.057	21.824	0.000
No.cl:Mixing.prop	1	18.626	18.626	7124.945	0.000
No.cl:Separation	1	13.985	13.985	5349.890	0.000
Sample.size:Mixing.prop	1	0.380	0.380	145.234	0.000
Sample.size:Separation	1	16.428	16.428	6284.092	0.000
Mixing.prop:Separation	1	12.391	12.391	4739.954	0.000
Residuals	255744	668.551	0.003		

By looking at the p-value column we can see that all the characteristics and the first-order interaction are highly significant. What is more important in determining the differences between values of the ASW in the two approaches is again the interaction between the number of variables and the cluster separation. If we focus on the additive terms, the element that has the highest Mean Square is the mixing proportion term, followed by the number of categories and the number of variables.

Again, in order to visualize the direction of the dependencies we plot the mean values of the ASW separately for each data features. The blue, the red and the black lines refer respectively to the Latent Gold, to the PAM and to the true clustering.

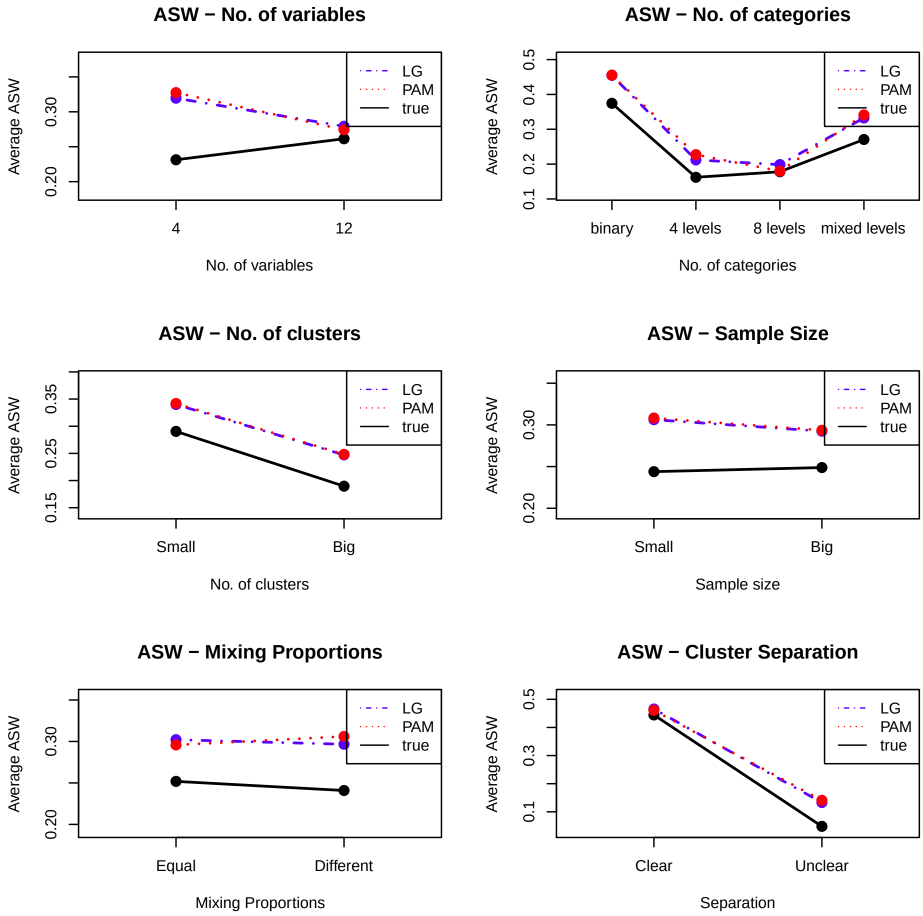


Figure 6.2: Average Silhouette Width: average values according to data features

From Figure 6.2 we can see that, in terms of ASW index, Latent Gold and PAM do not have very big differences in terms of directions: most of the lines overlap or are at least parallel; only in the case of the entity of mixing proportions the blue and the red lines have opposite slope.

Opposite directions are also those of LG and PAM when compared with the true clustering in terms of number of variables: the formers have higher values of ASW when the number of variables is small, whereas the value of the true clustering is higher when the number of variables increases.

6.2.3 Anova on PG: LG-PAM

Finally we performed an analysis of variance on the differences in the PG values of LG and PAM clustering, so that it was possible to say more about the features that affect the outcome.

Table 6.11 contains a summary of the ANOVA output, and we can see that all the data features and all the first-order interaction terms are highly significant.

Table 6.11: Analysis of Variance Table - Response: PG of LG-PAM with interactions

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
No.var	1	5.415	5.415	955.960	0.000
No.cat	3	175.551	58.517	10330.178	0.000
No.cl	1	8.773	8.773	1548.785	0.000
Sample.size	1	9.665	9.665	1706.125	0.000
Mixing.prop	1	23.289	23.289	4111.212	0.000
Separation	1	35.339	35.339	6238.556	0.000
No.var:No.cat	3	75.176	25.059	4423.680	0.000
No.var:No.cl	1	29.710	29.710	5244.792	0.000
No.var:Sample.size	1	1.981	1.981	349.669	0.000
No.var:Mixing.prop	1	22.132	22.132	3907.012	0.000
No.var:Separation	1	147.039	147.039	25957.270	0.000
No.cat:No.cl	3	3.209	1.070	188.828	0.000
No.cat:Sample.size	3	54.277	18.092	3193.922	0.000
No.cat:Mixing.prop	3	31.644	10.548	1862.059	0.000
No.cat:Separation	3	14.892	4.964	876.297	0.000
No.cl:Sample.size	1	7.510	7.510	1325.788	0.000
No.cl:Mixing.prop	1	18.048	18.048	3186.118	0.000
No.cl:Separation	1	3.620	3.620	639.063	0.000
Sample.size:Mixing.prop	1	1.008	1.008	177.916	0.000
Sample.size:Separation	1	32.305	32.305	5702.881	0.000
Mixing.prop:Separation	1	28.451	28.451	5022.632	0.000
Residuals	255744	1448.702	0.006		

By looking at the value of the Mean Square, we can say that the interaction term between the number of variables and the cluster separation is again the most influential factor. Restricting the attention to the additive effects only what affect more the dependent variable is the number of variables, followed by the cluster separation and the entity of mixing proportions.

In order to clarify the direction of the dependence of the Pearson Gamma index from the data features we plot the average values of the PG separately for each characteristic and we used different colours according to clustering outcome it refers to (blue is LG, red is PAM and black is the true one).

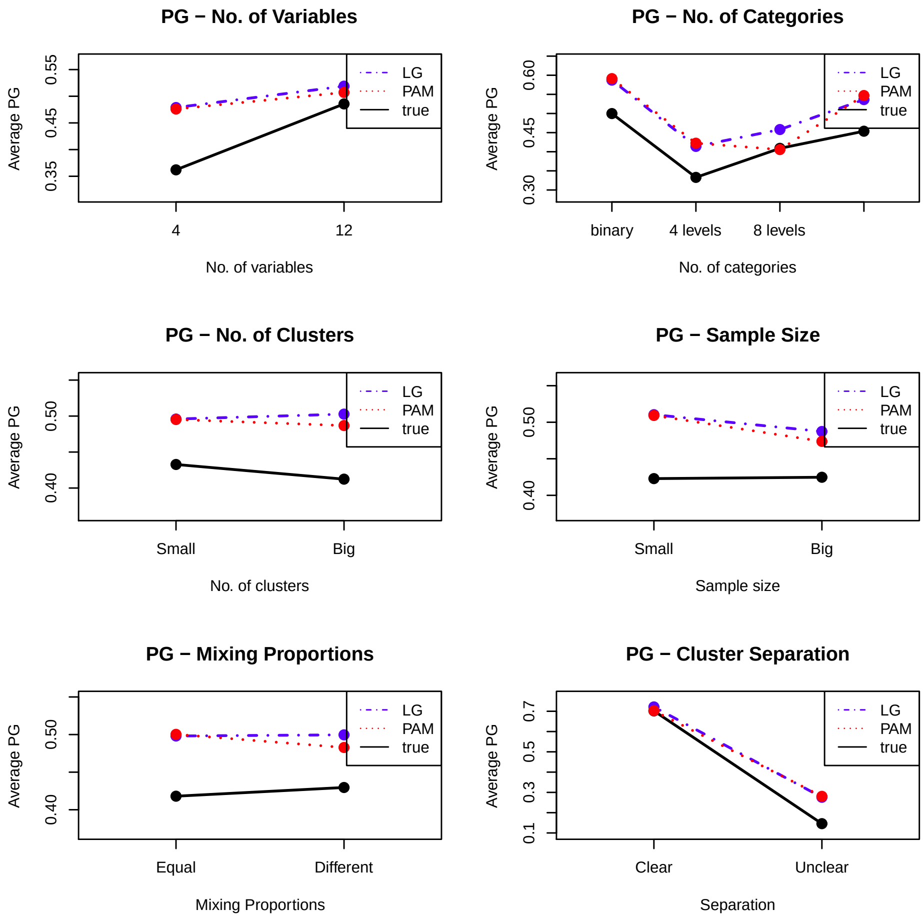


Figure 6.3: Pearson Gamma: average values according to data features

From Figure 6.3 we can see that in terms of number of clusters and entity of the mixing proportions lines for LG and PAM have opposite slopes, even though the differences is smaller than for the other indexes.

It is interesting to see that the bigger differences are with the true clustering values; particularly we see that values of PG for the true clustering are almost independent from the sample size which is not the case for LG and PAM, that actually seem to perform worse when the sample size is big.

By definition PG emphasises good approximation of the dissimilarity structure by the clustering and this is clear from the plot referred to the cluster separation: values of the PG are extremely high when the clusters are supposed to be clearly separated and viceversa.

Chapter 7

Conclusions

The thesis discusses the cluster analysis of categorical data; it focuses on two different approaches, namely a latent class cluster (LCC) analysis and a partition around medoids (PAM), that are considered and evaluated.

LCC and PAM refer to two wider classes of clustering methods, respectively *model-based* and *distance-based* methods. In the practise, the choice between the two approaches is strongly correlated with the aims of the researcher, since they are based on very different assumptions.

The research question that arose was whether both of these approaches lead to the same clustering and how good are clustering methods designed to fulfil one of these aims in terms of the other one. In order to answer, a fairly large simulation study was carried out, with the aim of understanding similarities and differences in terms of classification of the two approaches and with the aim of detecting different roles played by data features. In order to have a fair 'match', the two clustering outcomes were compared according to different criteria, based on the recovery of the true model structure (Adjusted Rand Index, which was expected to favour LCC) and based on dissimilarities (Average Silhouette Width and Pearson Gamma, which were expected to favour PAM).

The simulations were set according to the variation of several data features: the number of latent classes, the number of manifest variables, the number of categories for each variable, variation of class proportion, expected cluster separation and sample size. For all the possible combination of these factors (i.e. 128) we considered 2000 repetitions. Furthermore, we performed an Analysis of Variance on the differences between the the indexes we calculated in the LCC and in the PAM clustering outcomes, using data features as factors, in order to individuate the determinants that affect more the performance of the two approaches.

Overall, the simulations we carried out tell us that, in terms of recovering the 'true' clustering (according to a 'true' unknown model), the Latent Class Clustering generally behaves better, yielding better results in terms of Adjusted Rand Index, even when the clusters are supposed to overlap. When clusters are expected to be somehow separated, then a Partition Around Medoids approach would not make the results worse.

The factors that are more important in making the difference are the number of latent classes, since on average LatentGold® performed better with small number of clusters, whereas the opposite is true for PAM, the sample size, the larger the better is the performance of LatentGold® and the worse the performance of PAM, and the variation of the mixing proportion, since PAM works better when clusters have the same size.

Both approaches, on average, yield better results in terms of ARI when including larger number of variables and when those are binary.

Performance of the two methods in terms of retrieving homogeneous groups is more difficult to evaluate and more considerations need to be done. There is not a method that always outperforms the other one on average, so it is not easy to make general statements. Latent Gold, by trying to put together observations coming from the same distribution, accomplished to get similar observations together and to separate objects that are very different in a way that is not much worse than PAM. What is surprising is that sometimes LatentGold® works even better, in particular when models have a large number of variables and clusters overlap; in fact, in many of these situations PAM failed to put together similar observations.

Notice that the Average Silhouette Width, according to its expression, compares the dissimilarities of observations from other observations of the same cluster with observations of the nearest other cluster, which is not precisely what PAM does. In fact, PAM examines all the pairs of objects, and not only those belonging to the nearest cluster. Thus, by looking at the values of ASW one cannot state that PAM 'failed'; it is possible that the lower value of ASW is partially due to the differences in their formulation (of course if the number of clusters is larger than two).

The factors that have a larger effect in the difference of quality of clus-

Table 7.1: Adjusted Rand Index: highest average values according to data features and to clustering method

Data Feature	LG	PAM
No. of Variables	Large	Large
No. of Categories	2-levels	2-levels
No. of Clusters	Small	Large
Sample Size	Large	Small
Mixing Proportions	Different	Equal
Expected Cluster Separation	Clear	Clear

Table 7.2: Average Silhouette Width: highest average values according to data features and to clustering method

Data Feature	LG	PAM
No. of Variables	Small	Small
No. of Categories	2-levels	2-levels
No. of Clusters	Small	Small
Sample Size	Small	Small
Mixing Proportions	Different	Equal
Expected Cluster Separation	Clear	Clear

tering (according to the Analysis of Variance results) are the variability of mixing proportions, if we consider ASW (as we already said PAM performs better when clusters have about the same size), and the number of categories, if we focus on PG.

Table 7.1 shows the data features that are associated with the highest average values of Adjusted Rand Index in the two approaches. Similarly, Table 7.2 and Table 7.3 show the data features associated with the larger average values of Average Silhouette Width and Pearson Gamma in the two clustering approaches.

These are meant to be just a simplification. Indeed, the Analysis of Variance result shows that all factors and all the interactions have a significant impact on the difference in quality of clustering and this means that whatever is changed the result may look different; thus, make general statements according to these tables can be misleading.

In real situation, when looking for homogenous clustering, one can always perform both latent class and PAM clustering, and then choose the one that gives the best result in term of similarity/dissimiliarity.

Table 7.3: Pearson Gamma: highest average values according to data features and to clustering method

Data Feature	LG	PAM
No. of Variables	Large	Large
No. of Categories	2-levels	2-levels
No. of Clusters	Large	Small
Sample Size	Small	Small
Mixing Proportions	Different	Equal
Expected Cluster Separation	Clear	Clear

Finally, the comparison between the outcomes of LatentGold® and `lcmixed` shows that the two clustering results are generally the same; the small observed differences may be due to different initializations of the EM algorithm. Nevertheless, there is a very important feature that would lead to recommend the use of the commercial software over the open source one and it is the time required to compute the simulations: LatentGold® is much faster than `lcmixed`, precisely about 20 times faster.

Appendix A

Appendix

A.1 4 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	π_t^X	$\pi_{1t}^{\bar{A}X}$	$\pi_{1t}^{\bar{B}X}$	$\pi_{1t}^{\bar{C}X}$	$\pi_{1t}^{\bar{D}X}$
1	0.279	0.993	0.940	0.927	0.769
2	0.721	0.714	0.330	0.354	0.132

A.2 4 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

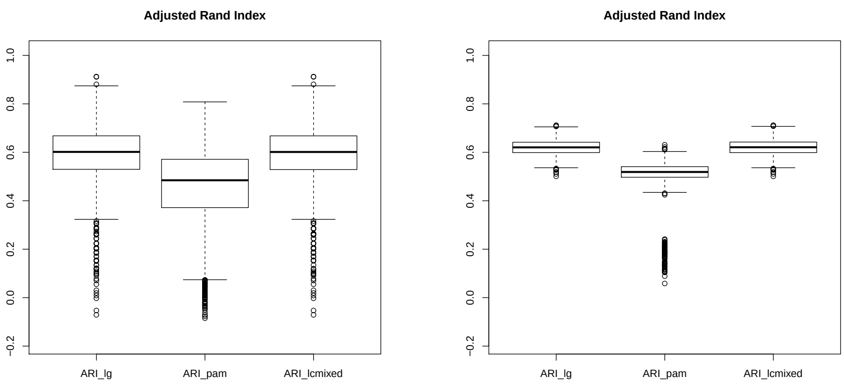
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$
1	0.20	0.90	0.88	0.92	0.10
2	0.80	0.10	0.10	0.15	0.91

A.3 4 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Table A.1: Summary: 4 binary variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.071	-0.085	-0.071
1st Qu.		0.530	0.372	0.529
Median		0.602	0.484	0.601
Mean (se)		0.590 (0.003)	0.440 (0.004)	0.589 (0.003)
3rd Qu.		0.668	0.571	0.668
Max.		0.913	0.808	0.913
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.166	0.250	0.286	0.250
1st Qu.	0.336	0.432	0.460	0.432
Median	0.373	0.468	0.495	0.469
Mean (se)	0.374 (0.001)	0.464 (0.001)	0.490 (0.001)	0.465 (0.001)
3rd Qu.	0.412	0.502	0.526	0.503
Max.	0.580	0.616	0.640	0.616
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.156	0.212	0.320	0.212
1st Qu.	0.345	0.465	0.551	0.466
Median	0.393	0.525	0.591	0.525
Mean (se)	0.394 (0.002)	0.516 (0.002)	0.583 (0.001)	0.517 (0.002)
3rd Qu.	0.444	0.575	0.625	0.576
Max.	0.654	0.727	0.733	0.727

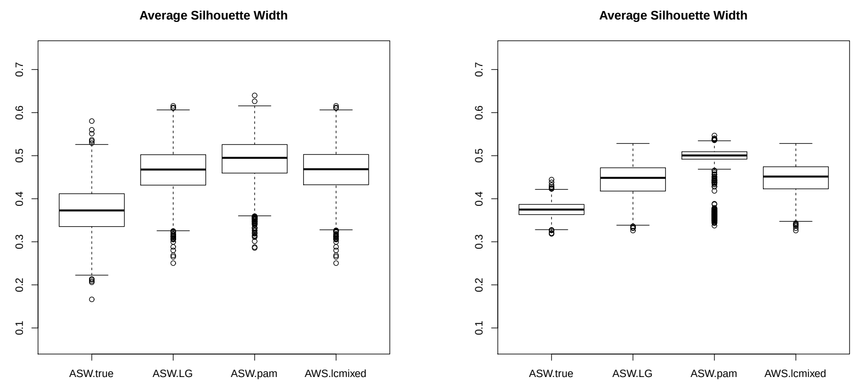


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.1: Adjusted Rand Index: 4 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

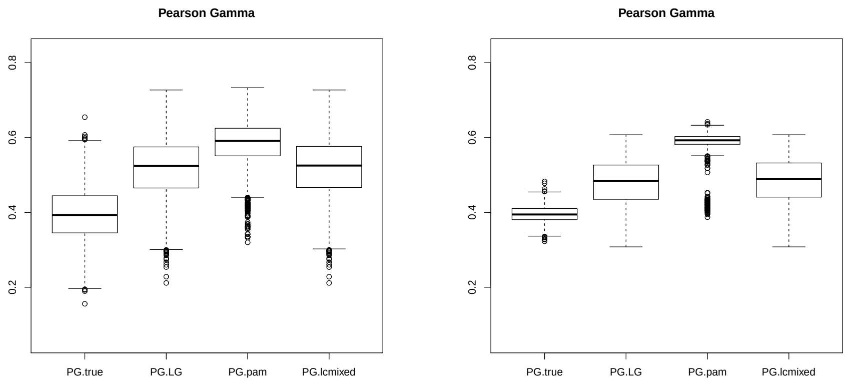
Table A.2: Summary: 4 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.501	0.058	0.501
1st Qu.		0.599	0.497	0.599
Median		0.620	0.519	0.621
Mean (se)		0.620 (0.001)	0.505 (0.002)	0.620 (0.001)
3rd Qu.		0.642	0.541	0.642
Max.		0.712	0.631	0.712
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.318	0.326	0.338	0.326
1st Qu.	0.363	0.4189	0.492	0.423
Median	0.375	0.448	0.500	0.452
Mean (se)	0.375 (0.000)	0.443 (0.001)	0.496 (0.001)	0.446 (0.001)
3rd Qu.	0.387	0.4729	0.509	0.474
Max.	0.444	0.528	0.547	0.528
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.323	0.308	0.387	0.308
1st Qu.	0.381	0.435	0.582	0.441
Median	0.394	0.484	0.593	0.489
Mean (se)	0.395 (0.000)	0.478 (0.001)	0.587 (0.001)	0.483 (0.001)
3rd Qu.	0.410	0.527	0.603	0.532
Max.	0.483	0.607	0.642	0.607



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.2: Average Silhouette Width: 4 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

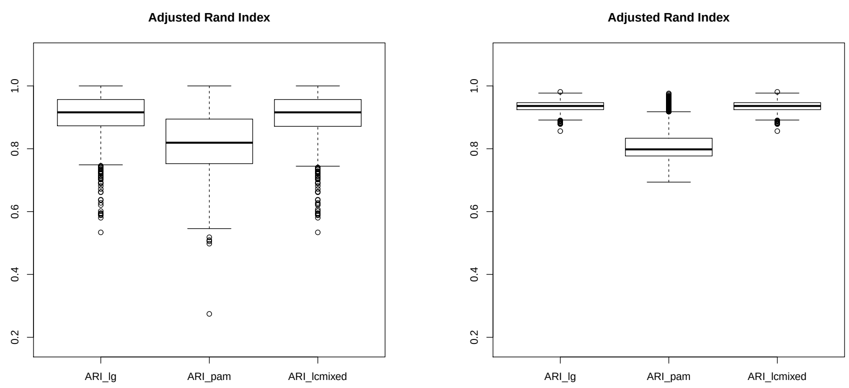
Figure A.3: Pearson Gamma: 4 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

Table A.3: Summary: 4 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.534	0.274	0.534
1st Qu.		0.873	0.753	0.871
Median		0.916	0.819	0.916
Mean (se)		0.914 (0.002)	0.819 (0.002)	0.912 (0.002)
3rd Qu.		0.957	0.894	0.957
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.577	0.607	0.594	0.607
1st Qu.	0.701	0.714	0.705	0.714
Median	0.728	0.740	0.732	0.740
Mean (se)	0.728 (0.001)	0.739 (0.001)	0.731 (0.001)	0.739 (0.001)
3rd Qu.	0.755	0.764	0.758	0.764
Max.	0.850	0.850	0.846	0.850
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.669	0.691	0.615	0.691
1st Qu.	0.801	0.818	0.820	0.818
Median	0.826	0.840	0.840	0.840
Mean (se)	0.824 (0.001)	0.838 (0.001)	0.839 (0.001)	0.838 (0.001)
3rd Qu.	0.850	0.860	0.859	0.860
Max.	0.922	0.922	0.922	0.922

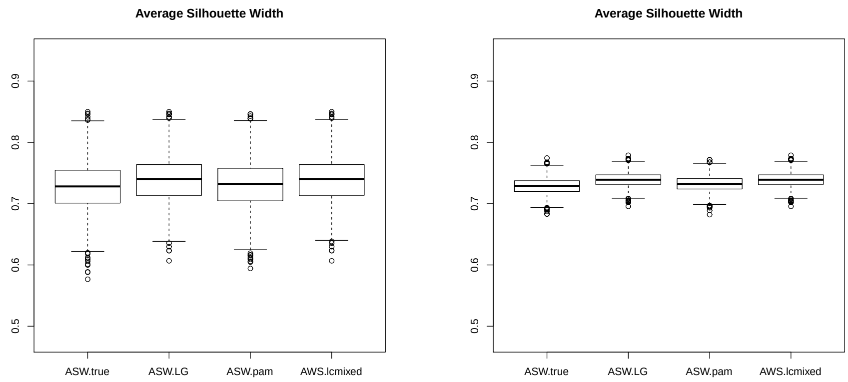
Table A.4: Summary: 4 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.856	0.694	0.856
1st Qu.		0.924	0.777	0.924
Median		0.936	0.798	0.936
Mean (se)		0.935 (0.000)	0.819 (0.001)	0.935 (0.000)
3rd Qu.		0.947	0.833	0.947
Max.		0.981	0.976	0.981
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.683	0.696	0.682	0.696
1st Qu.	0.720	0.732	0.724	0.732
Median	0.729	0.739	0.732	0.739
Mean (se)	0.729 (0.000)	0.739 (0.000)	0.732 (0.000)	0.739 (0.000)
3rd Qu.	0.737	0.747	0.741	0.747
Max.	0.774	0.779	0.772	0.779
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.782	0.800	0.803	0.800
1st Qu.	0.818	0.831	0.834	0.831
Median	0.826	0.838	0.840	0.838
Mean (se)	0.826 (0.000)	0.838 (0.000)	0.840 (0.000)	0.838 (0.000)
3rd Qu.	0.834	0.844	0.846	0.844
Max.	0.863	0.867	0.869	0.867



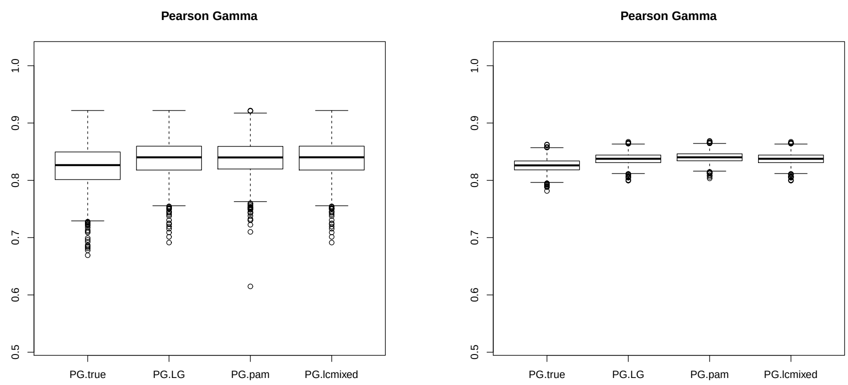
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.4: Adjusted Rand Index: 4 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.5: Average Silhouette Width: 4 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.6: Pearson Gamma: 4 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Table A.5: Summary: 4 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.060	-0.016	-0.037
1st Qu.		0.026	0.004	0.035
Median		0.091	0.030	0.103
Mean (se)		0.112 (0.002)	0.089 (0.002)	0.121 (0.002)
3rd Qu.		0.180	0.168	0.186
Max.		0.512	0.457	0.512
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.036	0.091	0.225	0.080
1st Qu.	0.130	0.277	0.301	0.300
Median	0.158	0.357	0.331	0.373
Mean (se)	0.160 (0.001)	0.341 (0.002)	0.336 (0.001)	0.356 (0.002)
3rd Qu.	0.188	0.420	0.367	0.423
Max.	0.346	0.494	0.490	0.494
NA's		13		2
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.037	-0.013	0.251	-0.003
1st Qu.	0.148	0.262	0.352	0.302
Median	0.181	0.402	0.395	0.434
Mean (se)	0.184 (0.001)	0.373 (0.003)	0.401 (0.001)	0.398 (0.003)
3rd Qu.	0.219	0.507	0.448	0.509
Max.	0.432	0.588	0.602	0.588
NA's		13		2

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$
1	0.55	0.80	0.60	0.30	0.70
2	0.45	0.70	0.50	0.60	0.20

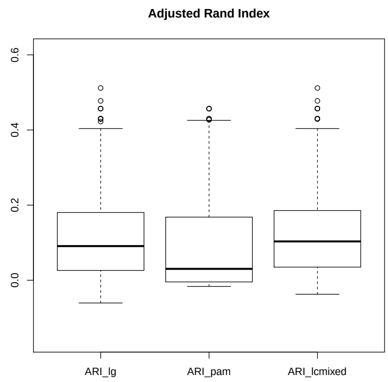
A.4 4 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation

The Model:

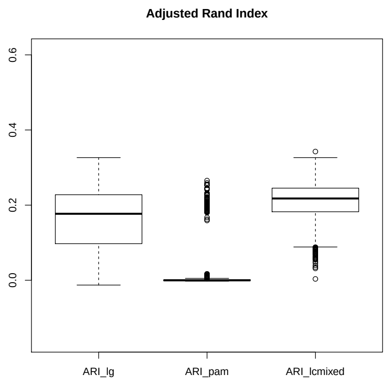
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$
1	0.50	0.10	0.11	0.90	0.92
2	0.50	0.90	0.91	0.08	0.12

Table A.6: Summary: 4 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.013	-0.002	0.003
1st Qu.		0.098	-0.001	0.182
Median		0.177	0.000	0.218
Mean (se)		0.161 (0.002)	0.009 (0.001)	0.210 (0.001)
3rd Qu.		0.228	0.001	0.245
Max.		0.326	0.265	0.343
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.109	0.086	0.274	0.123
1st Qu.	0.151	0.268	0.297	0.350
Median	0.162	0.341	0.303	0.367
Mean (se)	0.161 (0.000)	0.327 (0.002)	0.305 (0.000)	0.378 (0.001)
3rd Qu.	0.171	0.382	0.309	0.423
Max.	0.214	0.447	0.402	0.447
NA's		25		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.122	0.018	0.305	0.062
1st Qu.	0.172	0.268	0.340	0.416
Median	0.185	0.383	0.349	0.444
Mean (se)	0.184 (0.000)	0.358 (0.003)	0.353 (0.001)	0.443 (0.001)
3rd Qu.	0.196	0.456	0.358	0.493
Max.	0.253	0.530	0.496	0.530
NA's		25		

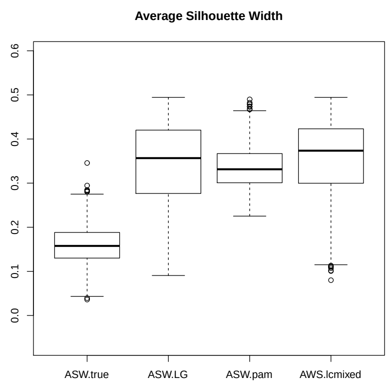


(a) *ARI - 100 units.*

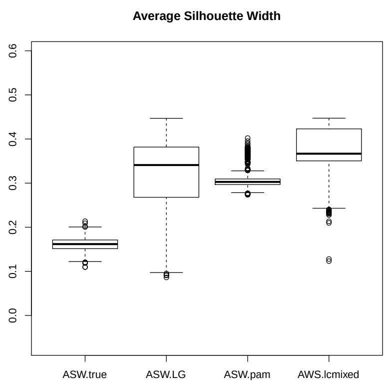


(b) *ARI - 1000 units.*

Figure A.7: Adjusted Rand Index: 4 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation

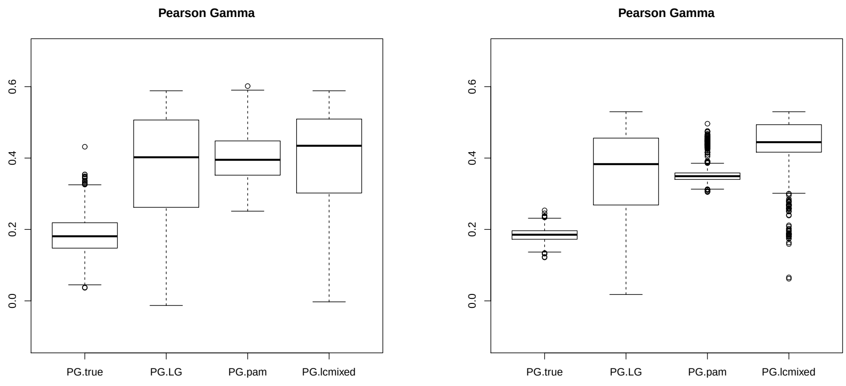


(a) *ASW - 100 units.*



(b) *ASW - 1000 units.*

Figure A.8: Average Silhouette Width: 4 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

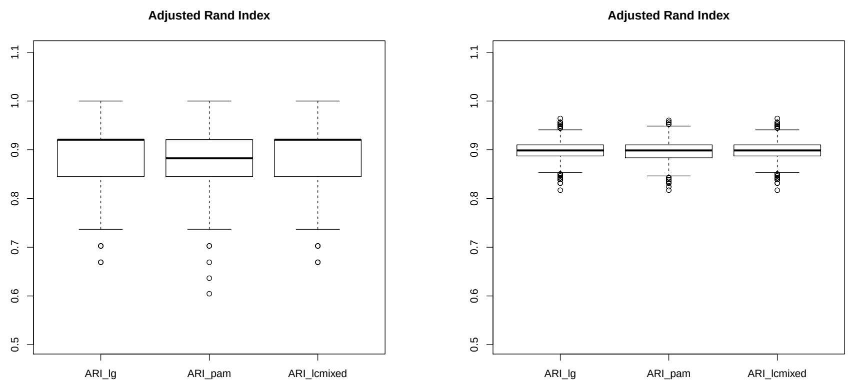
Figure A.9: Pearson Gamma: 4 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation

Table A.7: Summary: 4 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.669	0.604	0.669
1st Qu.		0.845	0.845	0.845
Median		0.921	0.882	0.921
Mean (se)		0.898 (0.001)	0.896 (0.001)	0.898 (0.001)
3rd Qu.		0.921	0.921	0.921
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.612	0.635	0.62 6	0.635
1st Qu.	0.728	0.740	0.736	0.740
Median	0.754	0.764	0.761	0.764
Mean (se)	0.754 (0.001)	0.763 (0.001)	0.761 (0.001)	0.763 (0.001)
3rd Qu.	0.781	0.788	0.786	0.788
Max.	0.881	0.881	0.881	0.881
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.7405	0.756	0.755	0.756
1st Qu.	0.842	0.855	0.852	0.855
Median	0.863	0.873	0.870	0.873
Mean (se)	0.862 (0.001)	0.871 (0.001)	0.869 (0.001)	0.871 (0.001)
3rd Qu.	0.883	0.889	0.887	0.889
Max.	0.948	0.948	0.948	0.948

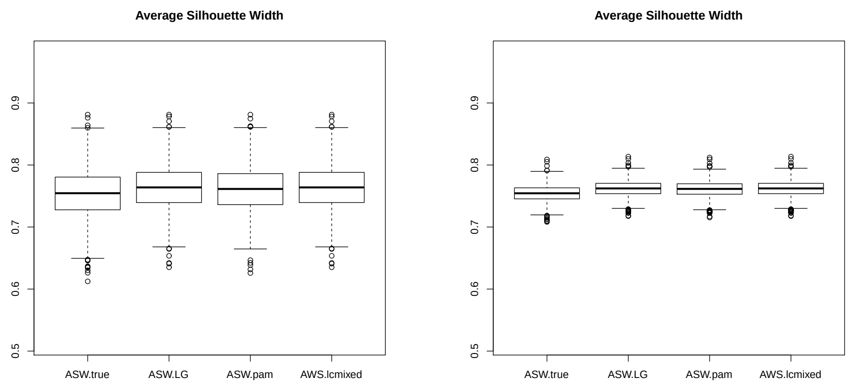
Table A.8: Summary: 4 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.817	0.817	0.817
1st Qu.		0.887	0.884	0.887
Median		0.899	0.899	0.899
Mean (se)		0.898 (0.000)	0.897 (0.000)	0.898 (0.000)
3rd Qu.		0.910	0.910	0.910
Max.		0.964	0.960	0.964
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.708	0.718	0.715	0.718
1st Qu.	0.745	0.754	0.753	0.754
Median	0.754	0.762	0.762	0.762
Mean (se)	0.754 (0.000)	0.762 (0.000)	0.761 (0.000)	0.762 (0.000)
3rd Qu.	0.763	0.770	0.770	0.770
Max.	0.809	0.814	0.812	0.814
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.828	0.836	0.833	0.836
1st Qu.	0.855	0.865	0.863	0.865
Median	0.862	0.871	0.870	0.871
Mean (se)	0.862 (0.000)	0.870 (0.000)	0.869 (0.000)	0.870 (0.000)
3rd Qu.	0.869	0.877	0.876	0.877
Max.	0.902	0.907	0.906	0.907



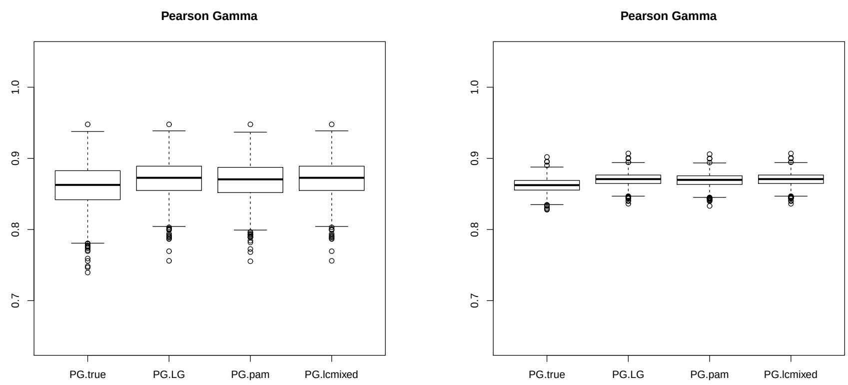
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.10: Adjusted Rand Index: 4 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.11: Average Silhouette Width: 4 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.12: Pearson Gamma: 4 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation

Table A.9: Summary: 4 binary variables - 3 clusters, mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.091	-0.059	-0.065
1st Qu.		0.071	0.032	0.069
Median		0.140	0.087	0.132
Mean (se)		0.150 (0.002)	0.094 (0.002)	0.144 (0.002)
3rd Qu.		0.218	0.145	0.203
Max.		0.496	0.413	0.563
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.118	0.072	0.277	0.169
1st Qu.	0.076	0.312	0.354	0.349
Median	0.105	0.386	0.392	0.403
Mean (se)	0.107 (0.001)	0.370 (0.002)	0.395 (0.001)	0.388 (0.001)
3rd Qu.	0.138	0.429	0.437	0.435
Max.	0.257	0.533	0.568	0.533
NA's		2		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.073	0.083	0.397	0.166
1st Qu.	0.188	0.467	0.471	0.487
Median	0.226	0.5250	0.510	0.534
Mean (se)	0.229 (0.001)	0.505 (0.002)	0.511 (0.001)	0.520 (0.001)
3rd Qu.	0.267	0.560	0.549	0.564
Max.	0.441	0.656	0.656	0.656
NA's		2		

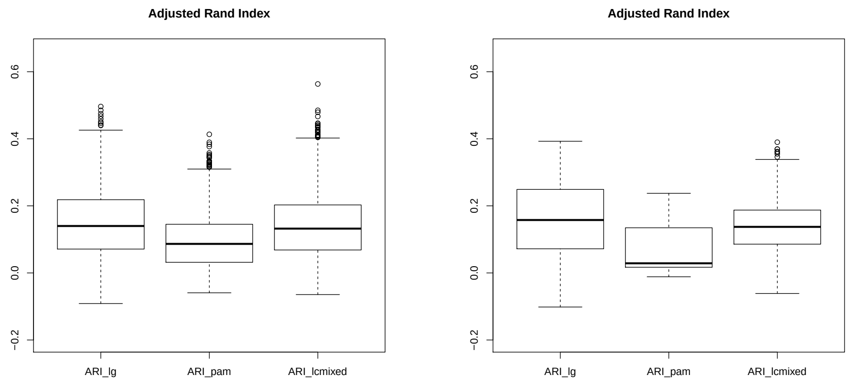
A.5 4 binary variables - 3 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$
1	0.11	0.40	0.50	0.60	0.50
2	0.63	0.70	0.10	0.30	0.40
3	0.26	0.30	0.20	0.80	0.40

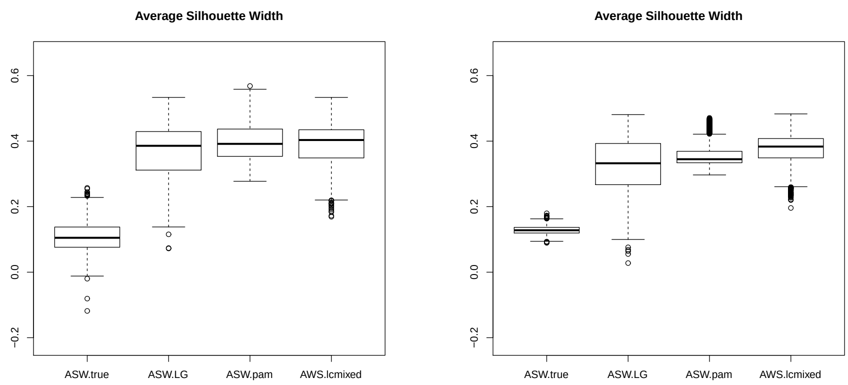
Table A.10: Summary: 4 binary variables - 3 clusters, mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.102	-0.012	-0.061
1st Qu.		0.072	0.017	0.086
Median		0.158	0.029	0.137
Mean (se)		0.156 (0.003)	0.062 (0.001)	0.139 (0.002)
3rd Qu.		0.249	0.135	0.187
Max.		0.393	0.237	0.390
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.090	0.028	0.297	0.196
1st Qu.	0.119	0.267	0.334	0.349
Median	0.128	0.332	0.345	0.383
Mean (se)	0.128 (0.000)	0.328 (0.002)	0.364 (0.001)	0.374 (0.001)
3rd Qu.	0.137	0.393	0.369	0.408
Max.	0.180	0.481	0.471	0.483
NA's		4		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.175	0.020	0.417	0.190
1st Qu.	0.219	0.371	0.446	0.469
Median	0.230	0.460	0.457	0.504
Mean (se)	0.231 (0.000)	0.433 (0.002)	0.477 (0.001)	0.492 (0.001)
3rd Qu.	0.242	0.519	0.490	0.531
Max.	0.299	0.610	0.578	0.608
NA's		4		



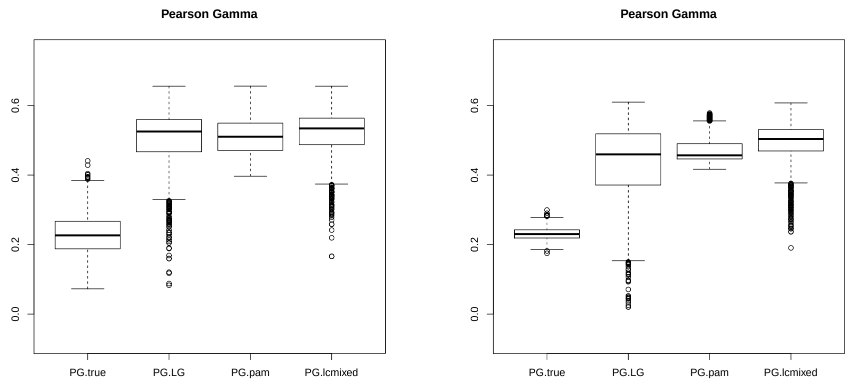
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.13: Adjusted Rand Index: 4 binary variables - 3 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.14: Average Silhouette Width: 4 binary variables - 3 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.15: Pearson Gamma: 4 binary variables - 3 clusters, mixing proportions extremely different - Unclear cluster separation

A.6 4 binary variables - 3 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$
1	0.10	0.20	0.10	0.15	0.09
2	0.30	0.80	0.90	0.10	0.20
3	0.60	0.90	0.80	0.85	0.91

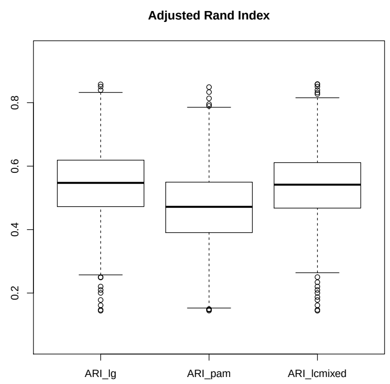
A.7 4 binary variables - 3 clusters, equal mixing proportions - Unclear cluster separation

The Model:

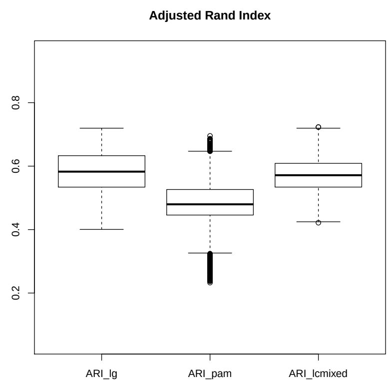
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$
1	0.33	0.80	0.30	0.50	0.20
2	0.33	0.20	0.70	0.50	0.60
3	0.33	0.60	0.50	0.55	0.45

Table A.11: Summary: 4 binary variables - 3 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.144	0.145	0.144
1st Qu.		0.472	0.391	0.468
Median		0.547	0.472	0.542
Mean (se)		0.544 (0.002)	0.469 (0.003)	0.540 (0.002)
3rd Qu.		0.619	0.550	0.611
Max.		0.858	0.849	0.859
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.200	0.348	0.384	0.351
1st Qu.	0.407	0.527	0.532	0.528
Median	0.446	0.558	0.561	0.559
Mean (se)	0.446 (0.001)	0.556 (0.001)	0.560 (0.001)	0.557 (0.001)
3rd Qu.	0.486	0.589	0.587	0.590
Max.	0.634	0.748	0.738	0.748
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.334	0.374	0.450	0.452
1st Qu.	0.555	0.643	0.622	0.643
Median	0.588	0.671	0.657	0.670
Mean (se)	0.588 (0.001)	0.668 (0.001)	0.651 (0.001)	0.667 (0.001)
3rd Qu.	0.624	0.698	0.684	0.696
Max.	0.753	0.822	0.807	0.822



(a) *ARI - 100 units.*

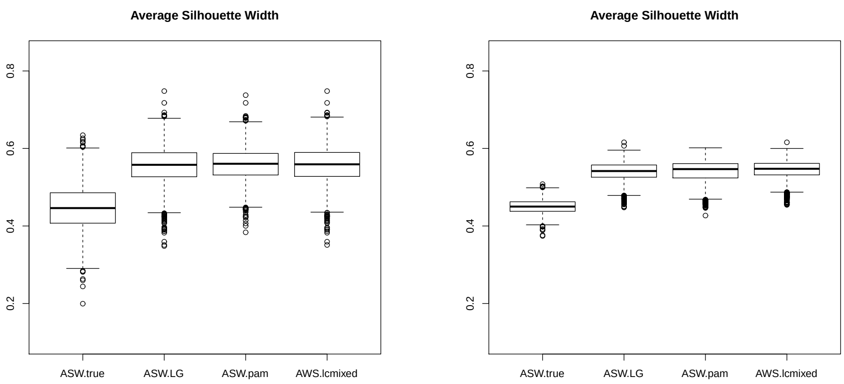


(b) *ARI - 1000 units.*

Figure A.16: Adjusted Rand Index: 4 binary variables - 3 clusters, mixing proportions extremely different - Clear cluster separation

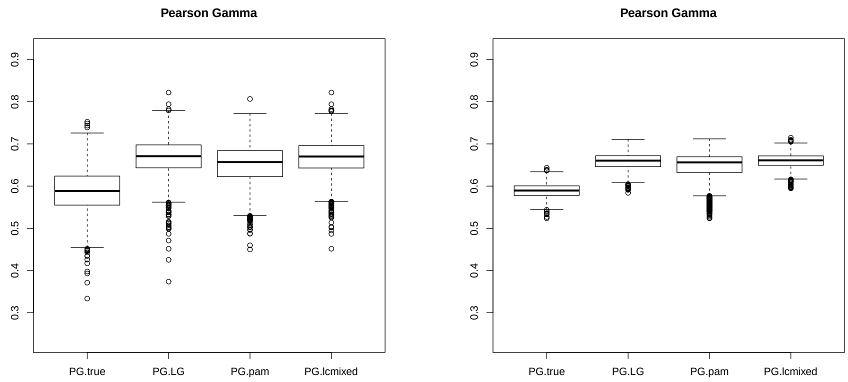
Table A.12: Summary: 4 binary variables - 3 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.400	0.233	0.422
1st Qu.		0.534	0.446	0.534
Median		0.583	0.480	0.571
Mean (se)		0.580 (0.001)	0.479 (0.002)	0.572 (0.001)
3rd Qu.		0.633	0.526	0.609
Max.		0.720	0.695	0.723
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.374	0.448	0.427	0.454
1st Qu.	0.438	0.526	0.524	0.532
Median	0.450	0.542	0.547	0.548
Mean (se)	0.450 (0.000)	0.540 (0.001)	0.540 (0.001)	0.544 (0.001)
3rd Qu.	0.462	0.557	0.561	0.562
Max.	0.508	0.616	0.602	0.616
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.524	0.584	0.524	0.595
1st Qu.	0.578	0.646	0.632	0.649
Median	0.590	0.660	0.656	0.661
Mean (se)	0.589 (0.000)	0.658 (0.000)	0.646 (0.001)	0.660 (0.000)
3rd Qu.	0.601	0.672	0.670	0.672
Max.	0.644	0.711	0.712	0.715



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.17: Average Silhouette Width: 4 binary variables - 3 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

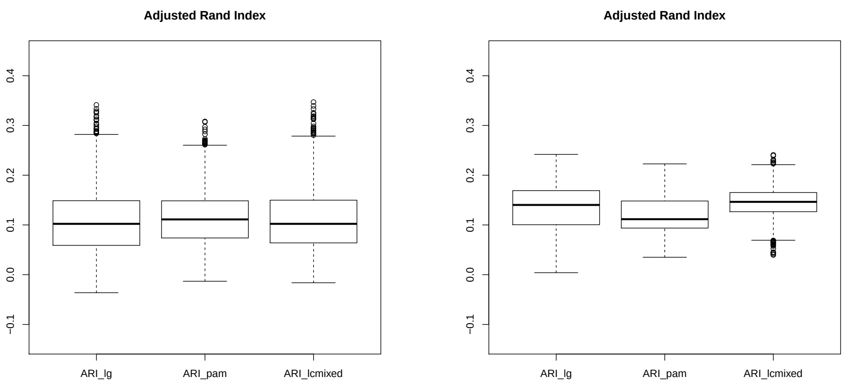
Figure A.18: Pearson Gamma: 4 binary variables - 3 clusters, mixing proportions extremely different - Clear cluster separation

Table A.13: Summary: 4 binary variables - 3 clusters, equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.036	-0.013	-0.016
1st Qu.		0.059	0.074	0.064
Median		0.102	0.111	0.102
Mean (se)		0.109 (0.001)	0.114 (0.001)	0.110 (0.001)
3rd Qu.		0.149	0.148	0.150
Max.		0.341	0.308	0.347
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.041	-0.020	0.264	0.129
1st Qu.	0.019	0.274	0.347	0.307
Median	0.040	0.340	0.374	0.355
Mean (se)	0.042 (0.001)	0.324 (0.002)	0.373 (0.001)	0.342 (0.001)
3rd Qu.	0.062	0.381	0.398	0.385
Max.	0.167	0.473	0.492	0.484
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.046	0.065	0.394	0.100
1st Qu.	0.129	0.440	0.495	0.475
Median	0.160	0.508	0.524	0.515
Mean (se)	0.163 (0.001)	0.476 (0.002)	0.520 (0.001)	0.496 (0.002)
3rd Qu.	0.192	0.537	0.548	0.541
Max.	0.341	0.635	0.628	0.640

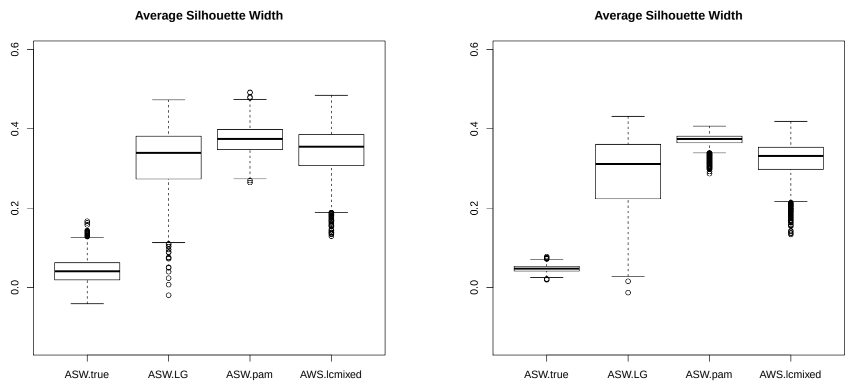
Table A.14: Summary: 4 binary variables - 3 clusters, equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.004	0.035	0.040
1st Qu.		0.100	0.094	0.127
Median		0.140	0.112	0.146
Mean (se)		0.134 (0.001)	0.120 (0.001)	0.145 (0.001)
3rd Qu.		0.169	0.148	0.165
Max.		0.242	0.223	0.241
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.019	-0.013	0.287	0.134
1st Qu.	0.041	0.223	0.364	0.298
Median	0.047	0.311	0.374	0.332
Mean (se)	0.048 (0.000)	0.292 (0.002)	0.369 (0.000)	0.321 (0.001)
3rd Qu.	0.053	0.361	0.381	0.353
Max.	0.078	0.431	0.407	0.419
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.122	0.058	0.427	0.145
1st Qu.	0.153	0.370	0.500	0.459
Median	0.162	0.456	0.518	0.491
Mean (se)	0.163 (0.000)	0.421 (0.002)	0.513 (0.000)	0.475 (0.001)
3rd Qu.	0.172	0.495	0.529	0.512
Max.	0.208	0.575	0.555	0.568



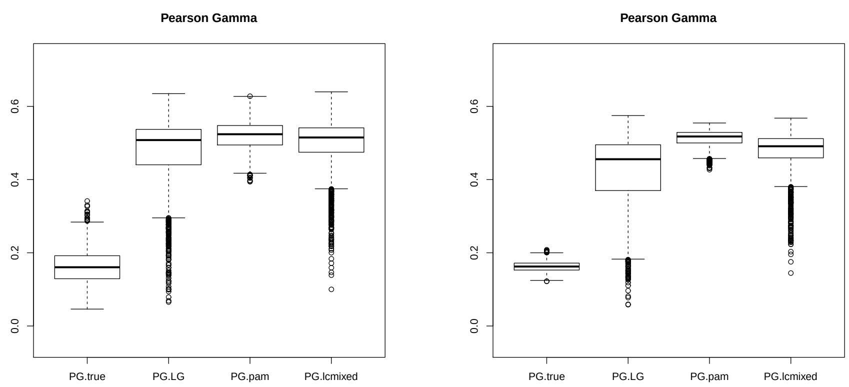
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.19: Adjusted Rand Index: 4 binary variables - 3 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.20: Average Silhouette Width: 4 binary variables - 3 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.21: Pearson Gamma: 4 binary variables - 3 clusters, equal mixing proportions - Unclear cluster separation

Table A.15: Summary: 4 binary variables - 3 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.264	0.284	0.299
1st Qu.		0.493	0.490	0.494
Median		0.552	0.546	0.554
Mean (se)		0.554 (0.002)	0.546 (0.002)	0.555 (0.002)
3rd Qu.		0.615	0.602	0.616
Max.		0.852	0.861	0.852
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.302	0.377	0.339	0.389
1st Qu.	0.431	0.549	0.504	0.549
Median	0.469	0.580	0.542	0.580
Mean (se)	0.469 (0.001)	0.578 (0.001)	0.540 (0.001)	0.579 (0.001)
3rd Qu.	0.507	0.611	0.5772	0.611
Max.	0.722	0.764	0.7313	0.764
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.469	0.470	0.536	0.477
1st Qu.	0.569	0.655	0.625	0.656
Median	0.601	0.677	0.651	0.677
Mean (se)	0.598 (0.001)	0.674 (0.001)	0.650 (0.001)	0.674 (0.001)
3rd Qu.	0.627	0.696	0.675	0.696
Max.	0.755	0.784	0.757	0.784

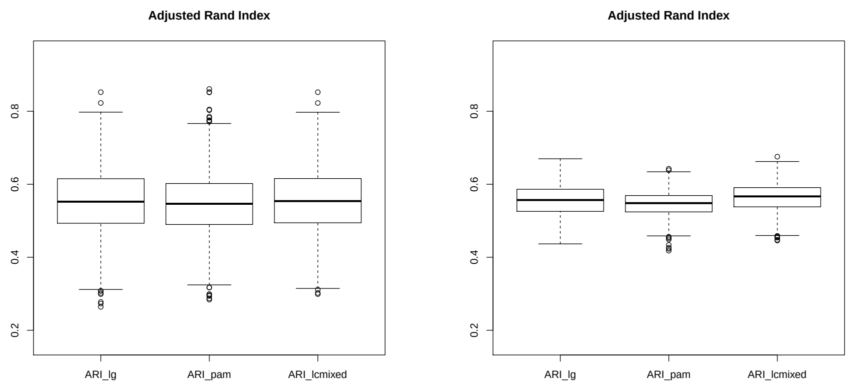
A.8 4 binary variables - 3 clusters, equal mixing proportions - Clear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$
1	0.33	0.90	0.80	0.85	0.92
2	0.33	0.90	0.20	0.10	0.90
3	0.33	0.10	0.15	0.15	0.10

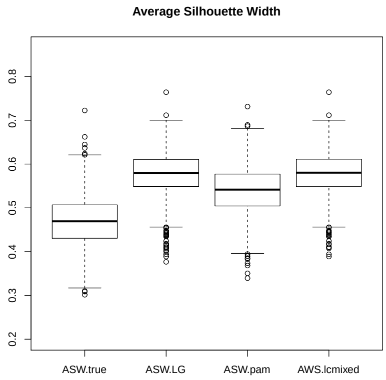
Table A.16: Summary: 4 binary variables - 3 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.437	0.418	0.446
1st Qu.		0.526	0.524	0.538
Median		0.557	0.548	0.567
Mean (se)		0.556 (0.001)	0.546 (0.001)	0.564 (0.001)
3rd Qu.		0.586	0.569	0.591
Max.		0.670	0.642	0.676
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.394	0.469	0.417	0.467
1st Qu.	0.458	0.553	0.513	0.553
Median	0.471	0.572	0.538	0.570
Mean (se)	0.470 (0.000)	0.570 (0.001)	0.535 (0.001)	0.568 (0.001)
3rd Qu.	0.482	0.591	0.559	0.589
Max.	0.533	0.642	0.616	0.632
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.546	0.604	0.593	0.604
1st Qu.	0.588	0.656	0.640	0.655
Median	0.597	0.668	0.648	0.666
Mean (se)	0.597 (0.000)	0.667 (0.000)	0.648 (0.000)	0.666 (0.000)
3rd Qu.	0.606	0.679	0.656	0.677
Max.	0.639	0.706	0.688	0.706

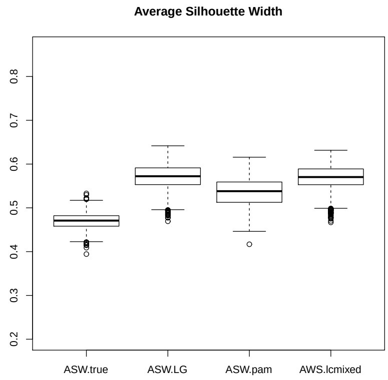


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.22: Adjusted Rand Index: 4 binary variables - 3 clusters, equal mixing proportions - Clear cluster separation

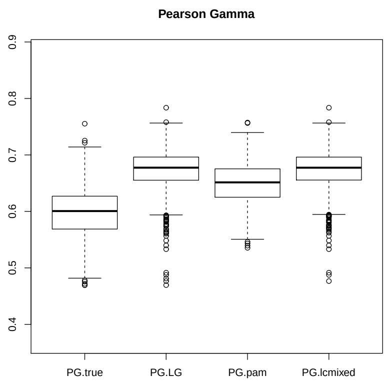


(a) *ASW - 100 units.*

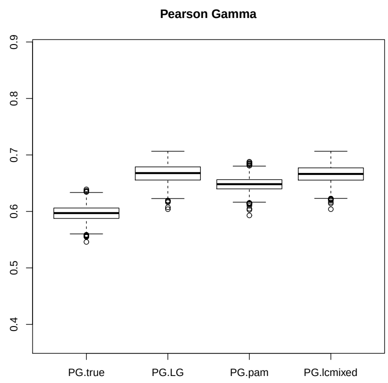


(b) *ASW - 1000 units.*

Figure A.23: Average Silhouette Width:4 binary variables - 3 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 100 units.*



(b) *PG - 1000 units.*

Figure A.24: Pearson Gamma: 4 binary variables - 3 clusters, equal mixing proportions - Clear cluster separation

Table A.17: Summary: 12 binary variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.112	-0.092	-0.110
1st Qu.		0.400	0.016	0.367
Median		0.562	0.130	0.557
Mean (se)		0.516 (0.005)	0.216 (0.005)	0.502 (0.005)
3rd Qu.		0.678	0.391	0.674
Max.		0.955	0.903	0.944
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.125	-0.034	0.098	-0.034
1st Qu.	0.210	0.232	0.155	0.227
Median	0.233	0.260	0.181	0.258
Mean (se)	0.233 (0.001)	0.253 (0.001)	0.194 (0.001)	0.251 (0.001)
3rd Qu.	0.254	0.284	0.232	0.282
Max.	0.355	0.368	0.361	0.368
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.183	-0.120	0.130	-0.120
1st Qu.	0.341	0.365	0.250	0.368
Median	0.380	0.419	0.309	0.422
Mean (se)	0.378 (0.001)	0.404 (0.002)	0.333 (0.002)	0.406 (0.002)
3rd Qu.	0.418	0.466	0.416	0.466
Max.	0.563	0.594	0.574	0.594

A.9 12 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

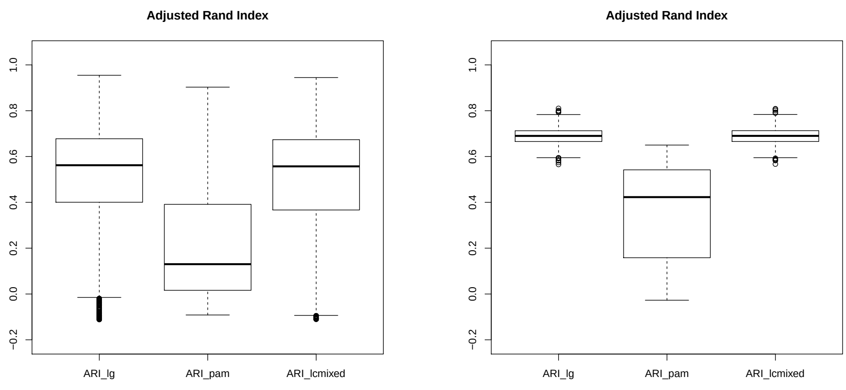
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.85	0.20	0.20	0.30	0.60	0.40	0.30
2	0.15	0.70	0.60	0.40	0.60	0.50	0.80

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.85	0.90	0.60	0.70	0.70	0.80	0.35
2	0.15	0.60	0.70	0.30	0.40	0.20	0.15

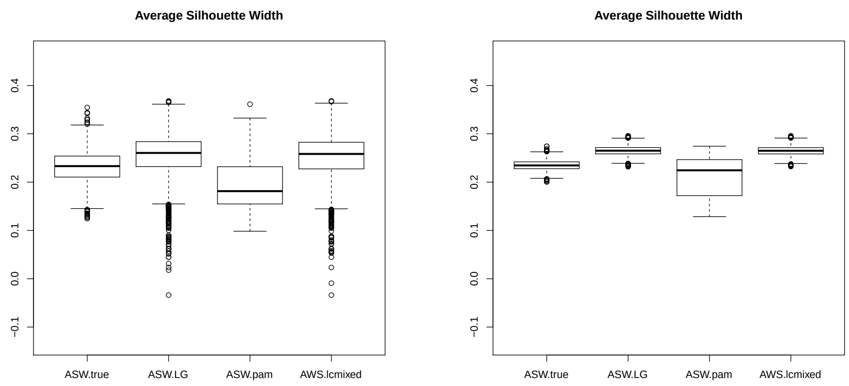
Table A.18: Summary: 12 binary variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.566	-0.027	0.567
1st Qu.		0.666	0.158	0.666
Median		0.690	0.423	0.690
Mean (se)		0.689 (0.001)	0.351 (0.005)	0.689 (0.001)
3rd Qu.		0.713	0.542	0.713
Max.		0.810	0.650	0.808
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.200	0.232	0.129	0.232
1st Qu.	0.228	0.258	0.172	0.258
Median	0.234	0.265	0.224	0.265
Mean (se)	0.235 (0.000)	0.265 (0.000)	0.210 (0.001)	0.265 (0.000)
3rd Qu.	0.242	0.272	0.247	0.272
Max.	0.274	0.296	0.274	0.296
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.322	0.346	0.201	0.352
1st Qu.	0.369	0.406	0.302	0.408
Median	0.381	0.420	0.400	0.422
Mean (se)	0.381 (0.000)	0.420 (0.000)	0.367 (0.002)	0.421 (0.000)
3rd Qu.	0.392	0.434	0.434	0.436
Max.	0.453	0.478	0.485	0.479



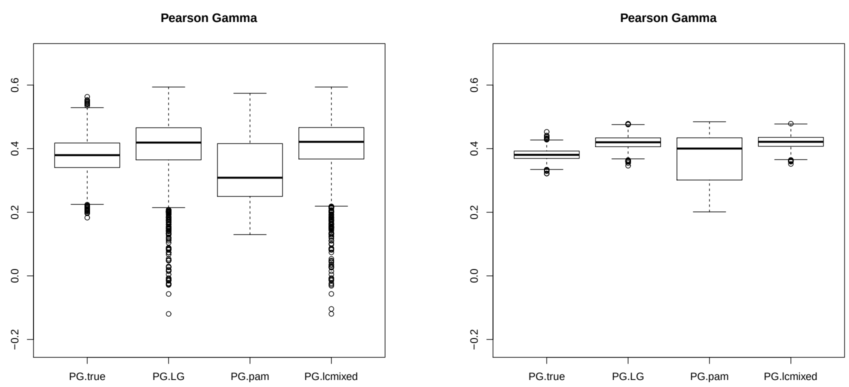
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.25: Adjusted Rand Index: 12 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.26: Average Silhouette Width: 12 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.27: Pearson Gamma: 12 binary variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

Table A.19: Summary: 12 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.655	0.075	0.655
1st Qu.		1.000	1.000	1.000
Median		1.000	1.000	1.000
Mean (se)		0.998 (0.000)	0.997 (0.001)	0.998 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.703	0.699	0.401	0.699
1st Qu.	0.760	0.760	0.760	0.760
Median	0.775	0.775	0.775	0.775
Mean (se)	0.774 (0.000)	0.774 (0.000)	0.774 (0.001)	0.774 (0.000)
3rd Qu.	0.789	0.789	0.789	0.789
Max.	0.847	0.847	0.847	0.847
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.692	0.687	0.329	0.687
1st Qu.	0.897	0.898	0.898	0.898
Median	0.914	0.914	0.914	0.914
Mean (se)	0.910 (0.001)	0.910 (0.001)	0.910 (0.001)	0.910 (0.001)
3rd Qu.	0.926	0.926	0.926	0.926
Max.	0.962	0.962	0.962	0.962

A.10 12 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

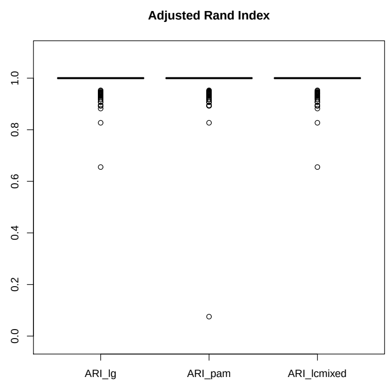
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.10	0.80	0.90	0.90	0.10	0.92	0.88
2	0.90	0.10	0.08	0.11	0.90	0.12	0.09

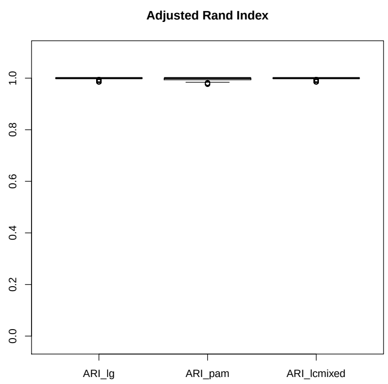
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.10	0.12	0.08	0.91	0.88	0.90	0.11
2	0.90	0.92	0.90	0.11	0.10	0.09	0.09

Table A.20: Summary: 12 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.986	0.976	0.986
1st Qu.		1.000	0.994	1.000
Median		1.000	1.000	1.000
Mean (se)		0.999 (0.000)	0.997 (0.000)	0.999 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.747	0.747	0.746	0.747
1st Qu.	0.770	0.770	0.770	0.770
Median	0.775	0.775	0.775	0.775
Mean (se)	0.775 (0.000)	0.775 (0.000)	0.774 (0.000)	0.775 (0.000)
3rd Qu.	0.779	0.779	0.779	0.779
Max.	0.796	0.796	0.795	0.796
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.890	0.890	0.890	0.890
1st Qu.	0.909	0.909	0.909	0.909
Median	0.913	0.913	0.913	0.913
Mean (se)	0.913 (0.000)	0.913 (0.000)	0.913 (0.000)	0.913 (0.000)
3rd Qu.	0.917	0.917	0.918	0.917
Max.	0.932	0.932	0.932	0.932

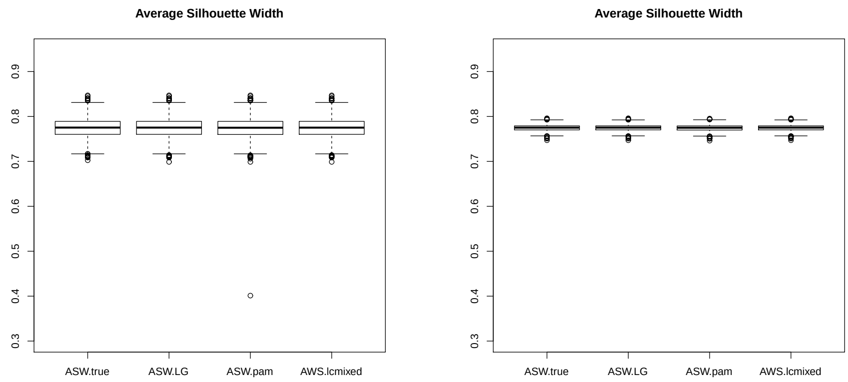


(a) *ARI - 100 units.*



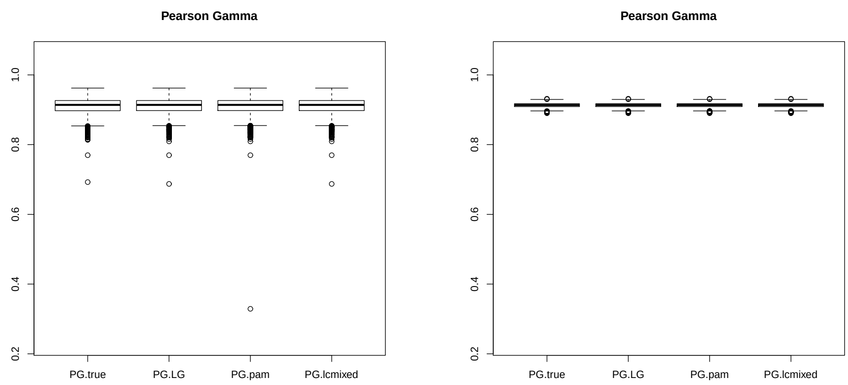
(b) *ARI - 1000 units.*

Figure A.28: Adjusted Rand Index: 12 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.29: Average Silhouette Width: 12 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.30: Pearson Gamma: 12 binary variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Table A.21: Summary: 12 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.058	-0.026	-0.058
1st Qu.		0.043	0.010	0.050
Median		0.136	0.058	0.151
Mean (se)		0.154 (0.003)	0.087 (0.002)	0.165 (0.003)
3rd Qu.		0.242	0.136	0.263
Max.		0.636	0.573	0.636
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.029	0.013	0.097	0.001
1st Qu.	0.085	0.147	0.140	0.151
Median	0.100	0.173	0.155	0.174
Mean (se)	0.101 (0.001)	0.167 (0.001)	0.156 (0.000)	0.169 (0.001)
3rd Qu.	0.117	0.192	0.170	0.192
Max.	0.194	0.270	0.234	0.270
NA's		2		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.048	-0.062	0.145	-0.074
1st Qu.	0.145	0.195	0.238	0.209
Median	0.171	0.271	0.263	0.281
Mean (se)	0.174 (0.001)	0.253 (0.002)	0.267 (0.001)	0.261 (0.002)
3rd Qu.	0.202	0.318	0.293	0.321
Max.	0.335	0.446	0.418	0.446
NA's		2		

A.11 12 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation

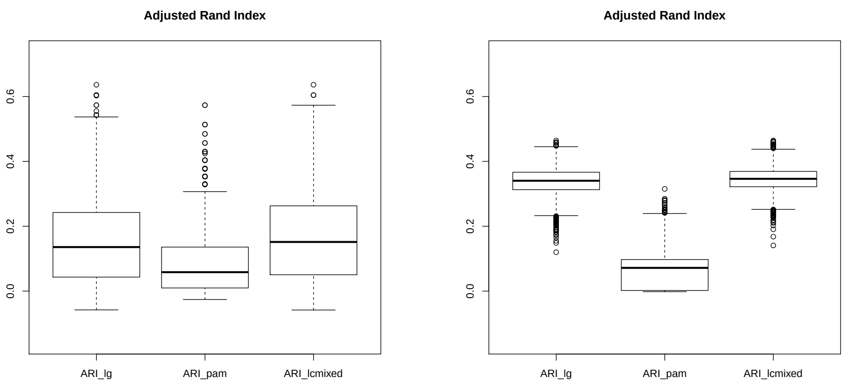
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.55	0.70	0.50	0.30	0.80	0.20	0.90
2	0.45	0.60	0.60	0.40	0.70	0.10	0.80

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.55	0.40	0.20	0.50	0.30	0.60	0.40
2	0.45	0.90	0.50	0.50	0.60	0.30	0.40

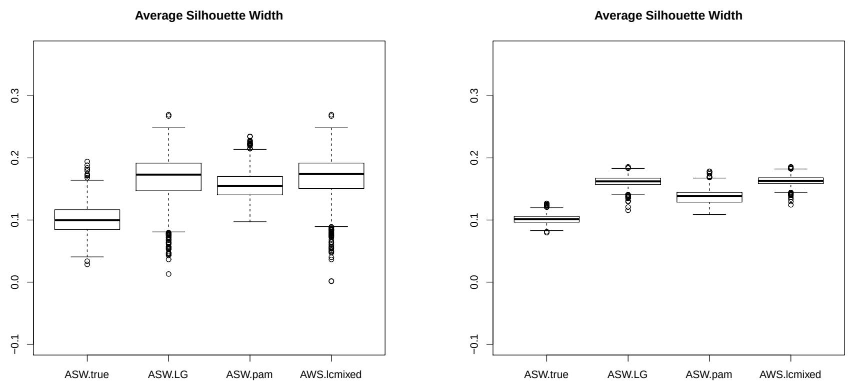
Table A.22: Summary: 12 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.120	-0.002	0.141
1st Qu.		0.313	0.002	0.322
Median		0.340	0.072	0.346
Mean (se)		0.338 (0.001)	0.068 (0.001)	0.345 (0.001)
3rd Qu.		0.367	0.097	0.369
Max.		0.465	0.315	0.464
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.080	0.116	0.109	0.124
1st Qu.	0.097	0.157	0.129	0.158
Median	0.101	0.162	0.138	0.163
Mean (se)	0.101 (0.000)	0.162 (0.000)	0.138 (0.000)	0.163 (0.000)
3rd Qu.	0.106	0.168	0.145	0.168
Max.	0.127	0.186	0.178	0.186
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.133	0.139	0.176	0.162
1st Qu.	0.165	0.264	0.216	0.270
Median	0.173	0.276	0.232	0.279
Mean (se)	0.173 (0.000)	0.274 (0.000)	0.231 (0.000)	0.278 (0.000)
3rd Qu.	0.182	0.286	0.244	0.288
Max.	0.220	0.322	0.306	0.322



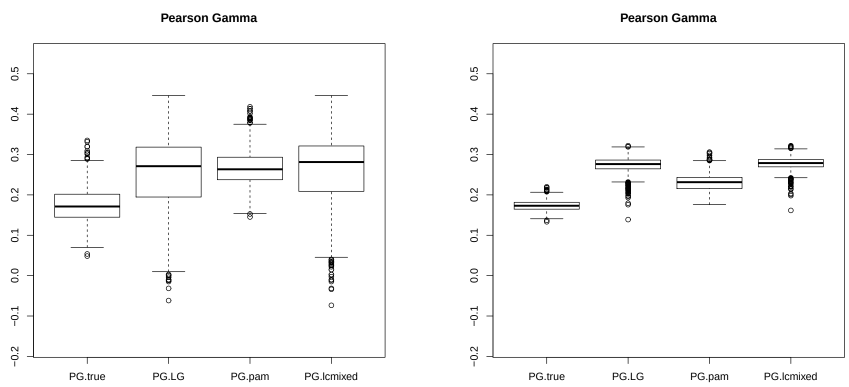
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.31: Adjusted Rand Index: 12 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.32: Average Silhouette Width: 12 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.33: Pearson Gamma: 12 binary variables - 2 clusters, equal mixing proportions - Unclear cluster separation

Table A.23: Summary: 12 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.984	0.984	0.984
1st Qu.		0.996	0.996	0.996
Median		1.000	1.000	1.000
Mean (se)		0.998 (0.000)	0.997 (0.000)	0.998 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.703	0.703	0.703	0.703
1st Qu.	0.724	0.724	0.724	0.724
Median	0.729	0.729	0.729	0.729
Mean (se)	0.729 (0.000)	0.729 (0.000)	0.729 (0.000)	0.729 (0.000)
3rd Qu.	0.734	0.734	0.734	0.734
Max.	0.751	0.751	0.751	0.751
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.917	0.917	0.917	0.917
1st Qu.	0.927	0.927	0.927	0.927
Median	0.929	0.929	0.929	0.929
Mean (se)	0.929 (0.000)	0.929 (0.000)	0.929 (0.000)	0.929 (0.000)
3rd Qu.	0.932	0.932	0.932	0.932
Max.	0.941	0.941	0.941	0.941

A.12 12 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation

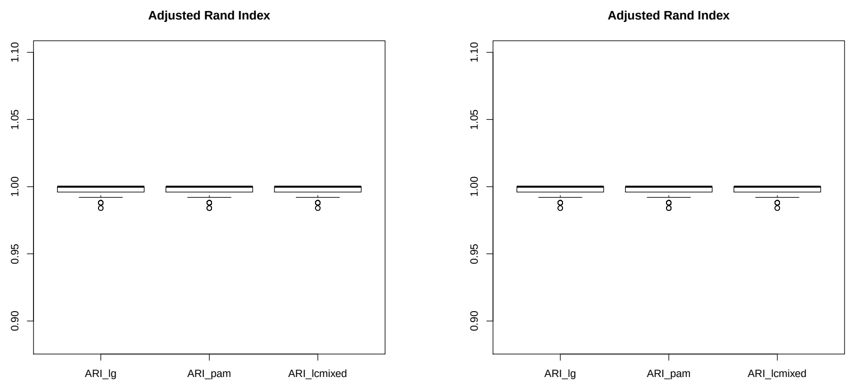
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.50	0.10	0.02	0.10	0.15	0.08	0.10
2	0.50	0.80	0.88	0.90	0.91	0.90	0.92

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.50	0.90	0.80	0.91	0.92	0.89	0.90
2	0.50	0.10	0.12	0.20	0.13	0.09	0.10

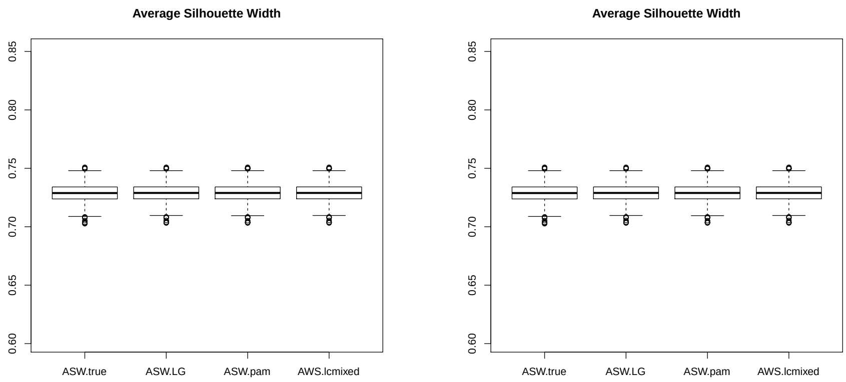
Table A.24: Summary: 12 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.984	0.984	0.984
1st Qu.		0.996	0.996	0.996
Median		1.000	1.000	1.000
Mean (se)		0.998 (0.000)	0.997 (0.000)	0.998 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.703	0.703	0.703	0.703
1st Qu.	0.724	0.724	0.724	0.724
Median	0.729	0.729	0.729	0.729
Mean (se)	0.729 (0.000)	0.729 (0.000)	0.729 (0.000)	0.729 (0.000)
3rd Qu.	0.734	0.734	0.734	0.734
Max.	0.751	0.751	0.751	0.751
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.917	0.917	0.917	0.917
1st Qu.	0.927	0.927	0.927	0.927
Median	0.929	0.929	0.929	0.929
Mean (se)	0.929 (0.000)	0.929 (0.000)	0.929 (0.000)	0.929 (0.000)
3rd Qu.	0.932	0.932	0.932	0.932
Max.	0.941	0.941	0.941	0.941



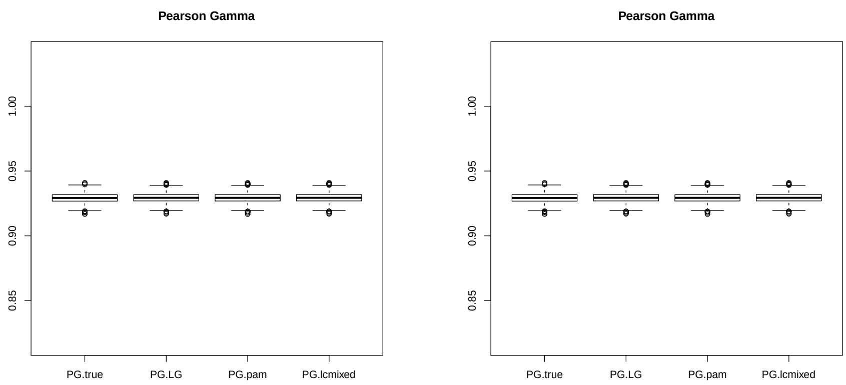
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.34: Adjusted Rand Index: 12 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation



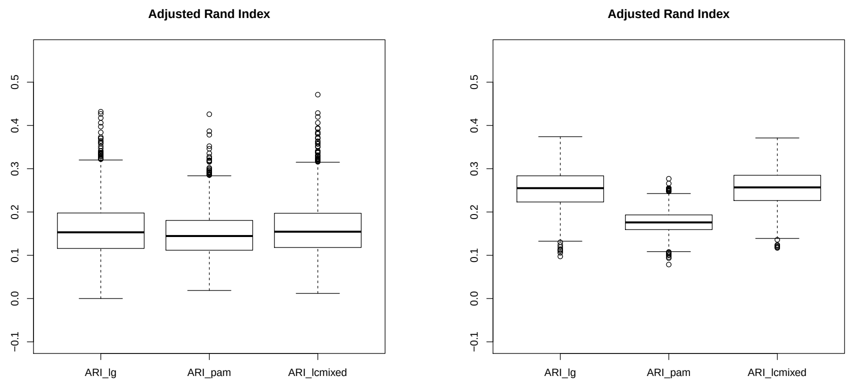
(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.35: Average Silhouette Width: 12 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.36: Pearson Gamma: 12 binary variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.37: Adjusted Rand Index: 12 binary variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

A.13 12 binary variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.10	0.20	0.30	0.80	0.60	0.30	0.50
2	0.15	0.40	0.50	0.60	0.20	0.70	0.30
3	0.20	0.30	0.60	0.40	0.20	0.50	0.70
4	0.25	0.70	0.60	0.30	0.40	0.20	0.20
5	0.30	0.20	0.20	0.50	0.70	0.40	0.80

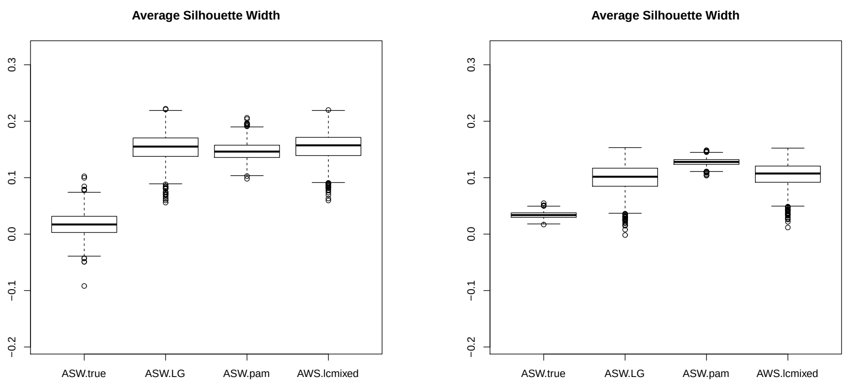
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.10	0.90	0.40	0.20	0.30	0.40	0.20
2	0.15	0.10	0.20	0.20	0.30	0.40	0.50
3	0.20	0.30	0.60	0.20	0.60	0.40	0.60
4	0.25	0.40	0.30	0.20	0.60	0.50	0.30
5	0.30	0.80	0.60	0.20	0.60	0.50	0.70

Table A.25: Summary: 12 binary variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.019	0.012
1st Qu.		0.116	0.112	0.118
Median		0.153	0.144	0.154
Mean (se)		0.160 (0.001)	0.149 (0.001)	0.161 (0.001)
3rd Qu.		0.198	0.180	0.197
Max.		0.432	0.426	0.471
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.092	0.056	0.098	0.060
1st Qu.	0.003	0.138	0.136	0.139
Median	0.017	0.155	0.146	0.157
Mean (se)	0.018 (0.000)	0.153 (0.001)	0.147 (0.000)	0.155 (0.001)
3rd Qu.	0.032	0.170	0.158	0.171
Max.	0.102	0.222	0.206	0.220
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.103	0.217	0.304	0.227
1st Qu.	0.204	0.373	0.361	0.375
Median	0.229	0.397	0.377	0.399
Mean (se)	0.229 (0.001)	0.393 (0.001)	0.378 (0.001)	0.395 (0.001)
3rd Qu.	0.254	0.417	0.394	0.418
Max.	0.380	0.540	0.463	0.499
NA's		1		

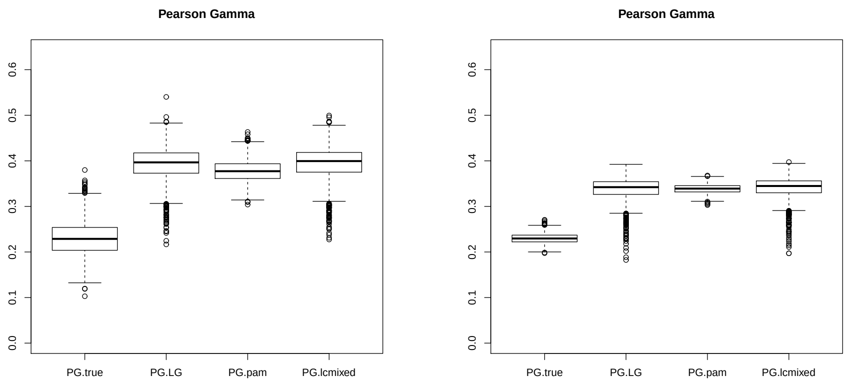
Table A.26: Summary: 12 binary variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.097	0.079	0.117
1st Qu.		0.223	0.159	0.226
Median		0.255	0.176	0.257
Mean (se)		0.253 (0.001)	0.177 (0.001)	0.255 (0.001)
3rd Qu.		0.284	0.193	0.285
Max.		0.374	0.277	0.371
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.017	-0.002	0.104	0.012
1st Qu.	0.030	0.085	0.124	0.092
Median	0.034	0.102	0.128	0.107
Mean (se)	0.034 (0.000)	0.099 (0.001)	0.128 (0.000)	0.105 (0.000)
3rd Qu.	0.038	0.117	0.132	0.120
Max.	0.055	0.153	0.149	0.152
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.198	0.182	0.303	0.197
1st Qu.	0.222	0.326	0.332	0.330
Median	0.230	0.342	0.339	0.345
Mean (se)	0.230 (0.000)	0.338 (0.001)	0.339 (0.000)	0.341 (0.001)
3rd Qu.	0.237	0.354	0.346	0.356
Max.	0.271	0.392	0.368	0.397



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.38: Average Silhouette Width: 12 binary variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.39: Pearson Gamma: 12 binary variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

A.14 12 binary variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

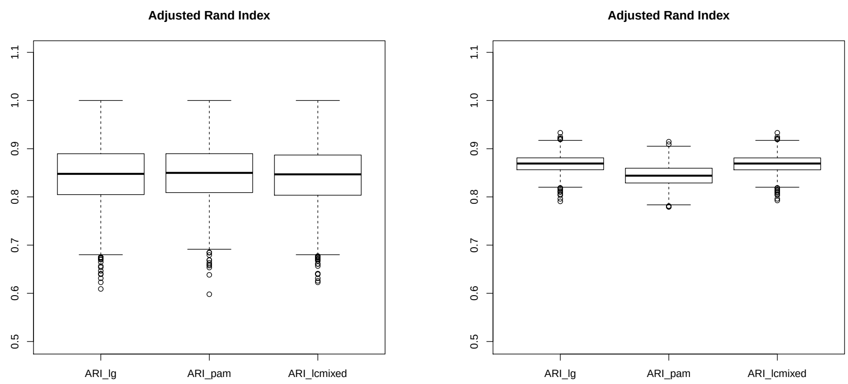
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.10	0.10	0.11	0.09	0.12	0.12	0.10
2	0.15	0.90	0.91	0.92	0.89	0.90	0.88
3	0.20	0.09	0.10	0.11	0.10	0.88	0.90
4	0.25	0.10	0.09	0.10	0.12	0.10	0.11
5	0.30	0.90	0.91	0.92	0.89	0.11	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.10	0.10	0.11	0.08	0.12	0.10	0.10
2	0.15	0.90	0.92	0.89	0.80	0.80	0.93
3	0.20	0.91	0.89	0.90	0.87	0.92	0.90
4	0.25	0.12	0.08	0.90	0.88	0.90	0.91
5	0.30	0.12	0.09	0.89	0.91	0.90	0.90

Table A.27: Summary: 12 binary variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.609	0.598	0.623
1st Qu.		0.805	0.809	0.804
Median		0.848	0.850	0.847
Mean (se)		0.844 (0.001)	0.846 (0.001)	0.842 (0.001)
3rd Qu.		0.890	0.890	0.887
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.341	0.359	0.355	0.356
1st Qu.	0.435	0.452	0.446	0.451
Median	0.460	0.476	0.470	0.476
Mean (se)	0.461 (0.001)	0.476 (0.001)	0.470 (0.001)	0.476 (0.001)
3rd Qu.	0.487	0.500	0.494	0.499
Max.	0.595	0.591	0.596	0.592
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.556	0.569	0.577	0.569
1st Qu.	0.627	0.635	0.631	0.634
Median	0.643	0.651	0.646	0.650
Mean (se)	0.643 (0.001)	0.650 (0.000)	0.646 (0.000)	0.650 (0.000)
3rd Qu.	0.658	0.664	0.660	0.664
Max.	0.719	0.727	0.720	0.727

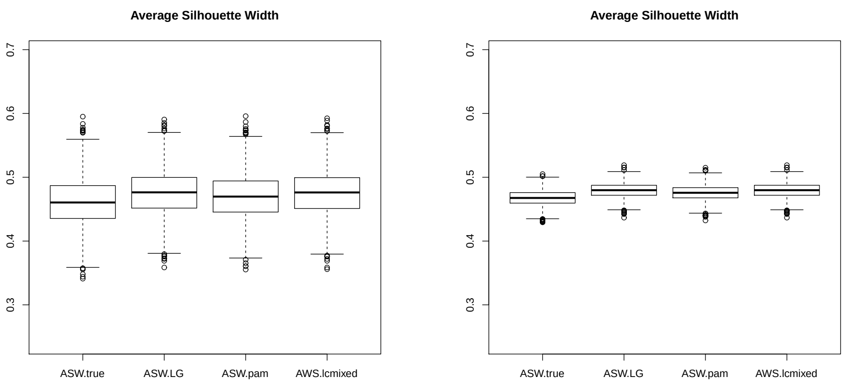


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.40: Adjusted Rand Index: 12 binary variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

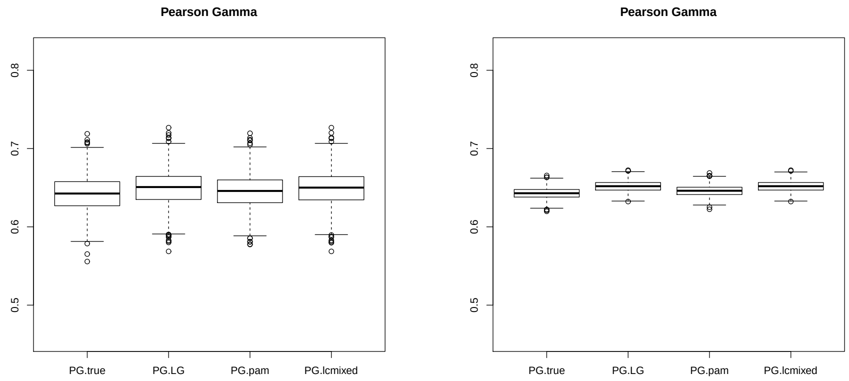
Table A.28: Summary: 12 binary variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.791	0.779	0.793
1st Qu.		0.856	0.829	0.856
Median		0.869	0.844	0.869
Mean (se)		0.869 (0.000)	0.844 (0.001)	0.869 (0.000)
3rd Qu.		0.881	0.860	0.881
Max.		0.933	0.915	0.933
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.429	0.437	0.432	0.437
1st Qu.	0.460	0.472	0.468	0.472
Median	0.468	0.480	0.476	0.480
Mean (se)	0.467 (0.000)	0.479 (0.000)	0.476 (0.000)	0.479 (0.000)
3rd Qu.	0.476	0.487	0.484	0.487
Max.	0.505	0.519	0.515	0.519
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.620	0.632	0.622	0.632
1st Qu.	0.638	0.647	0.641	0.647
Median	0.643	0.652	0.646	0.652
Mean (se)	0.643 (0.000)	0.652 (0.000)	0.646 (0.000)	0.652 (0.000)
3rd Qu.	0.648	0.657	0.651	0.657
Max.	0.666	0.672	0.669	0.672



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.41: Average Silhouette Width: 12 binary variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.42: Pearson Gamma: 12 binary variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

A.15 12 binary variables - 5 clusters, equal mixing proportions - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.20	0.30	0.20	0.50	0.40	0.70	0.10
2	0.20	0.30	0.60	0.70	0.70	0.80	0.50
3	0.20	0.30	0.20	0.80	0.30	0.60	0.40
4	0.20	0.30	0.60	0.40	0.20	0.30	0.70
5	0.20	0.30	0.20	0.40	0.80	0.50	0.70

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.20	0.80	0.60	0.40	0.30	0.20	0.20
2	0.20	0.30	0.40	0.20	0.50	0.60	0.70
3	0.20	0.80	0.50	0.20	0.50	0.40	0.40
4	0.20	0.30	0.50	0.20	0.60	0.70	0.80
5	0.20	0.80	0.30	0.40	0.60	0.60	0.70

A.16 12 binary variables - 5 clusters, equal mixing proportions - Clear cluster separation

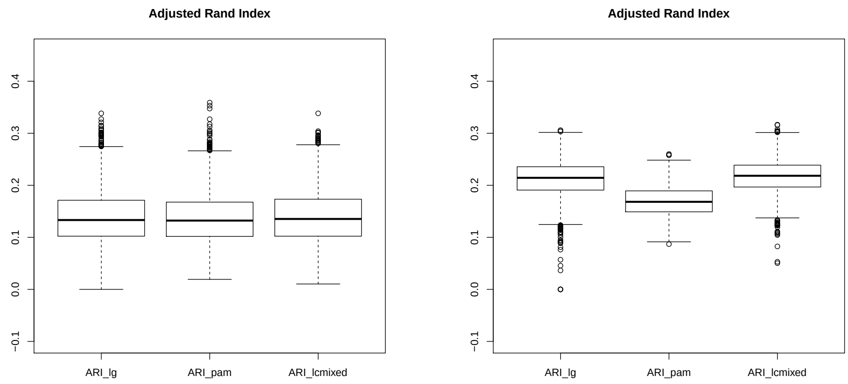
The Model:

Table A.29: Summary: 12 binary variables - 5 clusters, equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.019	0.010
1st Qu.		0.102	0.102	0.102
Median		0.133	0.132	0.135
Mean (se)		0.140 (0.001)	0.137 (0.001)	0.140 (0.001)
3rd Qu.		0.171	0.168	0.173
Max.		0.338	0.359	0.338
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.035	0.052	0.097	0.061
1st Qu.	0.009	0.134	0.135	0.138
Median	0.020	0.153	0.146	0.154
Mean (se)	0.021 (0.000)	0.151 (0.001)	0.147 (0.000)	0.152 (0.001)
3rd Qu.	0.034	0.169	0.158	0.169
Max.	0.088	0.238	0.212	0.238
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.085	0.181	0.305	0.222
1st Qu.	0.188	0.365	0.359	0.367
Median	0.209	0.390	0.375	0.391
Mean (se)	0.210 (0.001)	0.386 (0.001)	0.376 (0.001)	0.388 (0.001)
3rd Qu.	0.232	0.412	0.392	0.414
Max.	0.335	0.528	0.472	0.494
NA's		1		

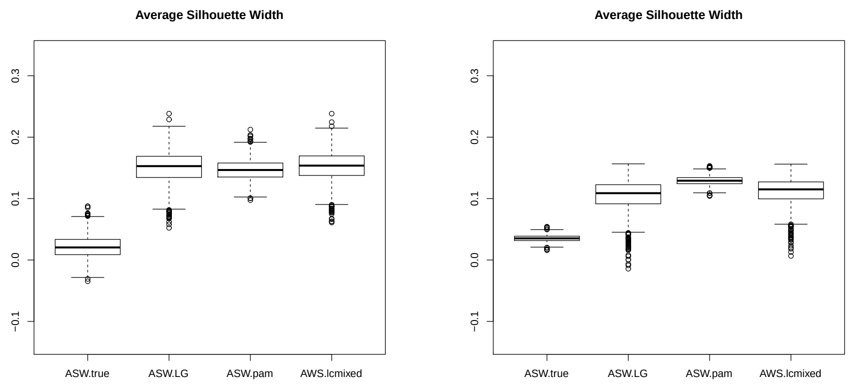
Table A.30: Summary: 12 binary variables - 5 clusters, equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.087	0.050
1st Qu.		0.191	0.149	0.197
Median		0.214	0.168	0.218
Mean (se)		0.212 (0.001)	0.168 (0.001)	0.217 (0.001)
3rd Qu.		0.236	0.189	0.239
Max.		0.306	0.260	0.316
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.016	-0.014	0.104	0.007
1st Qu.	0.032	0.009	0.124	0.100
Median	0.035	0.109	0.129	0.115
Mean (se)	0.035 (0.000)	0.105 (0.001)	0.129 (0.000)	0.112 (0.000)
3rd Qu.	0.039	0.123	0.134	0.127
Max.	0.054	0.157	0.153	0.156
NA's		5		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.167	0.103	0.298	0.120
1st Qu.	0.204	0.319	0.332	0.327
Median	0.210	0.338	0.341	0.342
Mean (se)	0.210 (0.000)	0.332 (0.001)	0.340 (0.000)	0.338 (0.001)
3rd Qu.	0.217	0.352	0.348	0.355
Max.	0.240	0.397	0.381	0.397
NA's		5		



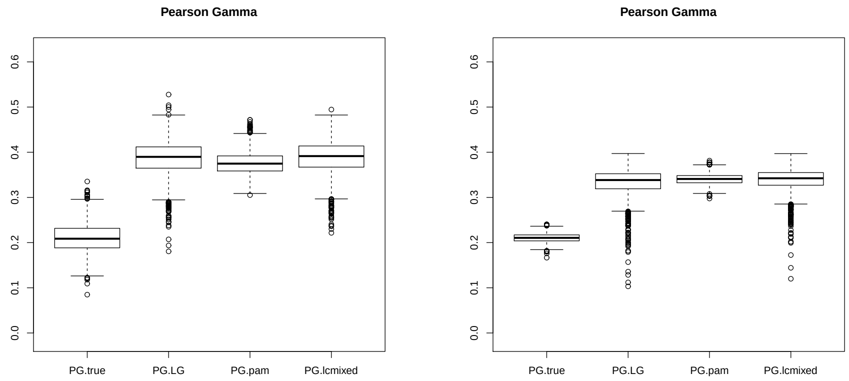
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.43: Adjusted Rand Index: 12 binary variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.44: Average Silhouette Width: 12 binary variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.45: Pearson Gamma: 12 binary variables - 5 clusters, equal mixing proportions - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}X}$	$\hat{\pi}_{1t}^{\bar{B}X}$	$\hat{\pi}_{1t}^{\bar{C}X}$	$\hat{\pi}_{1t}^{\bar{D}X}$	$\hat{\pi}_{1t}^{\bar{E}X}$	$\hat{\pi}_{1t}^{\bar{F}X}$
1	0.20	0.90	0.88	0.10	0.11	0.91	0.90
2	0.20	0.10	0.09	0.91	0.89	0.11	0.11
3	0.20	0.90	0.90	0.91	0.88	0.10	0.12
4	0.20	0.10	0.09	0.11	0.08	0.89	0.90
5	0.20	0.91	0.10	0.92	0.09	0.88	0.12

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}X}$	$\hat{\pi}_{1t}^{\bar{H}X}$	$\hat{\pi}_{1t}^{\bar{I}X}$	$\hat{\pi}_{1t}^{\bar{L}X}$	$\hat{\pi}_{1t}^{\bar{M}X}$	$\hat{\pi}_{1t}^{\bar{N}X}$
1	0.20	0.09	0.12	0.92	0.89	0.10	0.12
2	0.20	0.90	0.90	0.10	0.12	0.88	0.91
3	0.20	0.12	0.10	0.90	0.10	0.90	0.11
4	0.20	0.91	0.88	0.10	0.91	0.11	0.89
5	0.20	0.90	0.10	0.91	0.11	0.88	0.12

A.17 4 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

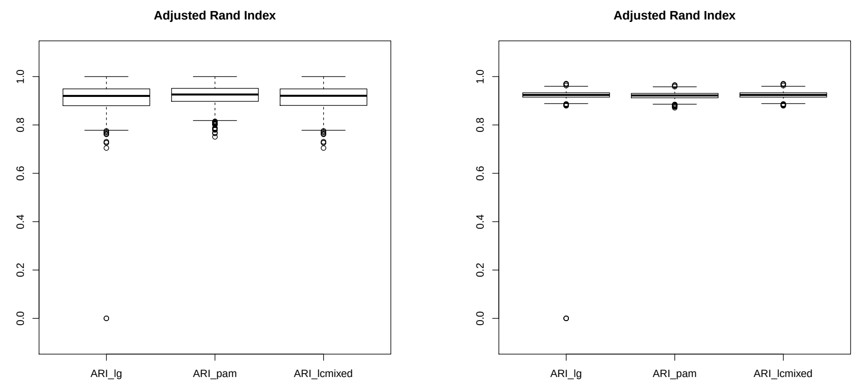
The Model:

Table A.31: Summary: 12 binary variables - 5 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.751	0.705
1st Qu.		0.880	0.898	0.881
Median		0.920	0.926	0.921
Mean (se)		0.912 (0.001)	0.922 (0.001)	0.913 (0.001)
3rd Qu.		0.949	0.951	0.949
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.402	0.416	0.415	0.416
1st Qu.	0.512	0.519	0.518	0.519
Median	0.537	0.543	0.543	0.544
Mean (se)	0.536 (0.001)	0.543 (0.001)	0.542 (0.001)	0.543 (0.001)
3rd Qu.	0.560	0.566	0.565	0.566
Max.	0.680	0.680	0.680	0.680
NA's		2		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.614	0.620	0.622	0.625
1st Qu.	0.666	0.670	0.669	0.670
Median	0.677	0.681	0.680	0.681
Mean (se)	0.677 (0.000)	0.681 (0.000)	0.680 (0.000)	0.681 (0.000)
3rd Qu.	0.689	0.692	0.691	0.692
Max.	0.734	0.734	0.734	0.734
NA's		2		

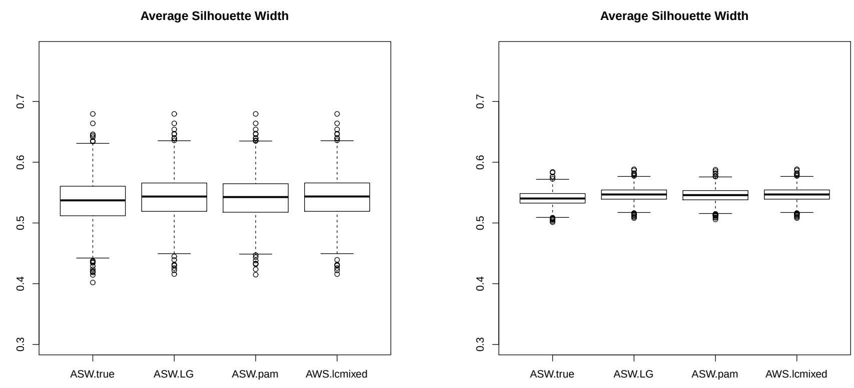
Table A.32: Summary: 12 binary variables - 5 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.871	0.879
1st Qu.		0.915	0.912	0.915
Median		0.924	0.922	0.924
Mean (se)		0.922 (0.001)	0.921 (0.000)	0.924 (0.000)
3rd Qu.		0.933	0.931	0.933
Max.		0.971	0.965	0.971
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.501	0.508	0.506	0.508
1st Qu.	0.533	0.539	0.538	0.539
Median	0.540	0.547	0.546	0.547
Mean (se)	0.540 (0.000)	0.547 (0.000)	0.546 (0.000)	0.547 (0.000)
3rd Qu.	0.548	0.554	0.553	0.554
Max.	0.584	0.588	0.588	0.588
NA's		4		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.658	0.665	0.665	0.665
1st Qu.	0.673	0.677	0.677	0.677
Median	0.677	0.681	0.680	0.681
Mean (se)	0.676 (0.000)	0.681 (0.000)	0.680 (0.000)	0.681 (0.000)
3rd Qu.	0.680	0.684	0.683	0.684
Max.	0.695	0.697	0.696	0.697
NA's		4		



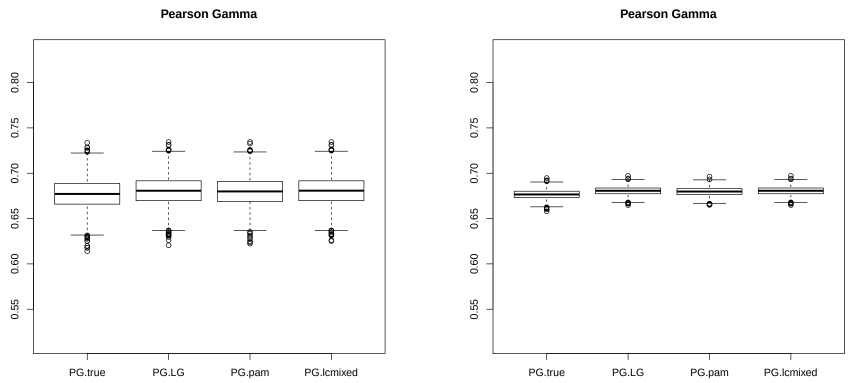
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.46: Adjusted Rand Index: 12 binary variables - 5 clusters, equal mixing proportions - Clear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.47: Average Silhouette Width: 12 binary variables - 5 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.48: Pearson Gamma: 12 binary variables - 5 clusters, equal mixing proportions - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.20	0.40	0.40	0.10	0.10	0.20	0.40	0.30	0.10
2	0.80	0.20	0.30	0.20	0.30	0.40	0.10	0.20	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.20	0.20	0.30	0.30	0.20	0.30	0.30	0.30	0.10
2	0.80	0.10	0.40	0.40	0.10	0.50	0.10	0.20	0.20

A.18 4 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

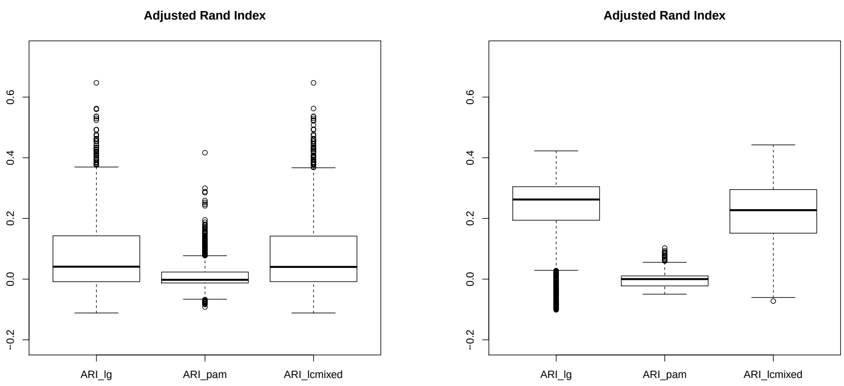
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.20	0.10	0.10	0.40	0.40	0.10	0.10	0.10	0.70
2	0.80	0.40	0.40	0.10	0.10	0.70	0.10	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.20	0.10	0.40	0.10	0.40	0.40	0.40	0.10	0.10
2	0.80	0.40	0.10	0.40	0.10	0.10	0.10	0.40	0.40

Table A.33: Summary: 4 4-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.112	-0.092	-0.112
1st Qu.		-0.008	-0.013	-0.008
Median		0.041	-0.002	0.040
Mean (se)		0.080 (0.003)	0.009 (0.001)	0.078 (0.003)
3rd Qu.		0.143	0.023	0.142
Max.		0.647	0.417	0.647
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.002	0.074	0.101	0.078
1st Qu.	0.055	0.149	0.146	0.149
Median	0.070	0.164	0.160	0.164
Mean (se)	0.071 (0.001)	0.165 (0.001)	0.160 (0.000)	0.165 (0.001)
3rd Qu.	0.087	0.181	0.173	0.181
Max.	0.148	0.250	0.235	0.252
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.005	0.032	0.161	0.064
1st Qu.	0.091	0.232	0.245	0.233
Median	0.118	0.269	0.271	0.270
Mean (se)	0.120 (0.001)	0.267 (0.001)	0.272 (0.001)	0.268 (0.001)
3rd Qu.	0.147	0.306	0.297	0.306
Max.	0.262	0.434	0.425	0.434

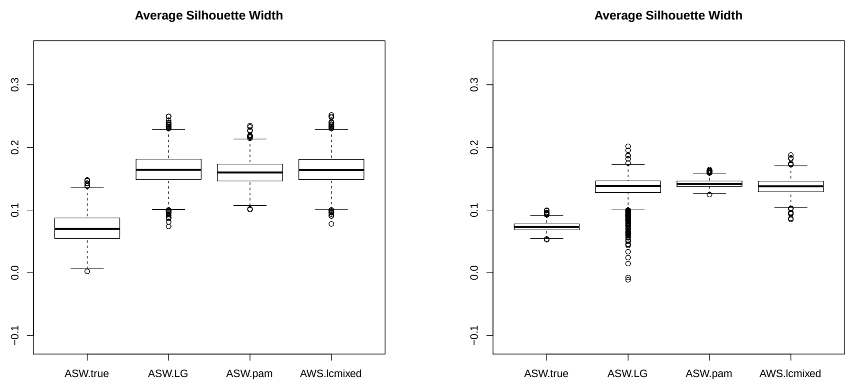


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.49: Adjusted Rand Index: 4 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

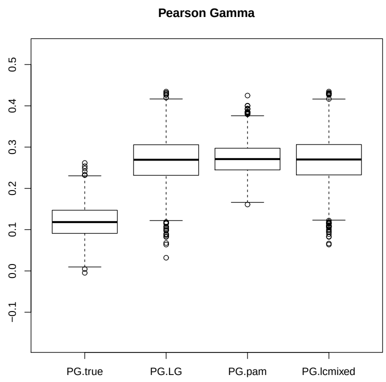
Table A.34: Summary: 4 4-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.101	-0.050	-0.072
1st Qu.		0.194	-0.022	0.152
Median		0.262	0.000	0.227
Mean (se)		0.232 (0.002)	-0.003 (0.000)	0.219 (0.002)
3rd Qu.		0.305	0.010	0.295
Max.		0.423	0.102	0.442
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.053	-0.011	0.125	0.085
1st Qu.	0.068	0.128	0.138	0.129
Median	0.073	0.138	0.142	0.138
Mean (se)	0.073 (0.000)	0.135 (0.000)	0.142 (0.000)	0.138 (0.000)
3rd Qu.	0.078	0.146	0.146	0.146
Max.	0.100	0.202	0.165	0.188
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.083	-0.069	0.190	0.079
1st Qu.	0.113	0.166	0.220	0.208
Median	0.121	0.197	0.230	0.230
Mean (se)	0.121 (0.000)	0.195 (0.001)	0.230 (0.000)	0.228 (0.001)
3rd Qu.	0.130	0.231	0.240	0.249
Max.	0.170	0.358	0.279	0.338

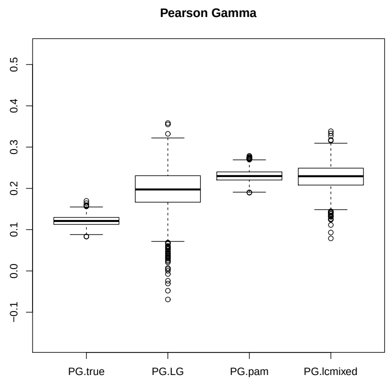


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.50: Average Silhouette Width: 4 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.*



(b) *PG - 1000 units.*

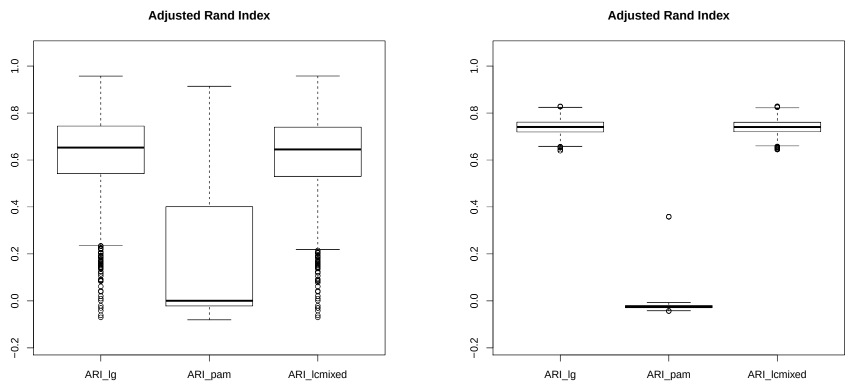
Figure A.51: Pearson Gamma: 4 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

Table A.35: Summary: 4 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.070	-0.081	-0.070
1st Qu.		0.542	-0.021	0.531
Median		0.653	0.001	0.645
Mean (se)		0.628 (0.004)	0.169 (0.006)	0.621 (0.004)
3rd Qu.		0.745	0.401	0.740
Max.		0.957	0.914	0.958
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.145	0.124	0.088	0.124
1st Qu.	0.235	0.258	0.154	0.257
Median	0.257	0.279	0.178	0.279
Mean (se)	0.257 (0.001)	0.278 (0.001)	0.197 (0.001)	0.277 (0.001)
3rd Qu.	0.279	0.300	0.244	0.300
Max.	0.353	0.366	0.361	0.366
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.221	0.130	0.095	0.130
1st Qu.	0.388	0.429	0.210	0.429
Median	0.424	0.467	0.259	0.468
Mean (sd)	0.424 (0.001)	0.463 (0.001)	0.310 (0.003)	0.464 (0.001)
3rd Qu.	0.460	0.503	0.435	0.504
Max.	0.596	0.633	0.607	0.633

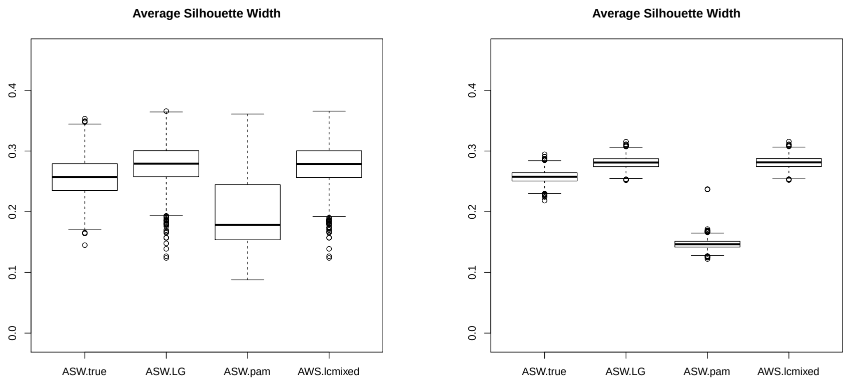
Table A.36: Summary: 4 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.639	-0.044	0.644
1st Qu.		0.719	-0.029	0.720
Median		0.740	-0.024	0.740
Mean (se)		0.740 (0.001)	-0.024 (0.000)	0.740 (0.001)
3rd Qu.		0.761	-0.020	0.761
Max.		0.829	0.360	0.829
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.218	0.252	0.122	0.252
1st Qu.	0.251	0.274	0.142	0.274
Median	0.258	0.281	0.146	0.281
Mean (sd)	0.258 (0.000)	0.281 (0.000)	0.147 (0.000)	0.281 (0.000)
3rd Qu.	0.264	0.287	0.151	0.288
Max.	0.295	0.315	0.237	0.316
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.366	0.376	0.158	0.376
1st Qu.	0.414	0.441	0.187	0.443
Median	0.425	0.456	0.195	0.458
Mean (se)	0.425 (0.000)	0.456 (0.000)	0.196 (0.000)	0.458 (0.000)
3rd Qu.	0.436	0.472	0.204	0.473
Max.	0.495	0.532	0.429	0.534



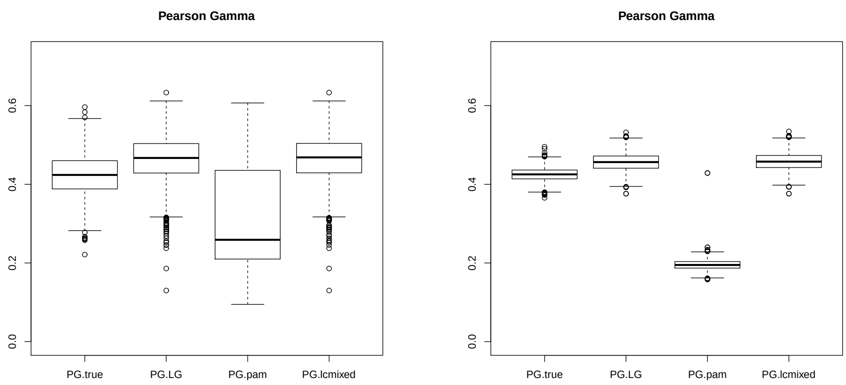
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.52: Adjusted Rand Index: 4 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.53: Average Silhouette Width: 4 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.54: Pearson Gamma: 4 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Table A.37: Summary: 4 4-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.033	-0.023	-0.033
1st Qu.		-0.006	-0.007	-0.006
Median		0.006	0.000	0.007
Mean (se)		0.023 (0.001)	0.012 (0.001)	0.023 (0.001)
3rd Qu.		0.035	0.022	0.039
Max.		0.305	0.242	0.261
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.007	0.077	0.102	0.048
1st Qu.	0.019	0.142	0.138	0.142
Median	0.028	0.158	0.151	0.157
Mean (se)	0.029 (0.000)	0.158 (0.001)	0.152 (0.000)	0.158 (0.001)
3rd Qu.	0.039	0.173	0.164	0.173
Max.	0.095	0.242	0.237	0.242
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.012	0.052	0.144	0.008
1st Qu.	0.032	0.220	0.238	0.221
Median	0.049	0.259	0.263	0.260
Mean (se)	0.051 (0.001)	0.256 (0.001)	0.264 (0.001)	0.257 (0.001)
3rd Qu.	0.068	0.295	0.289	0.296
Max.	0.164	0.452	0.430	0.452

A.19 4 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

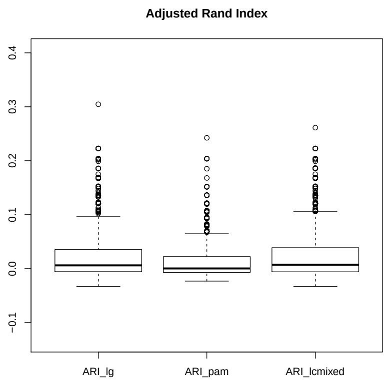
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$
1	0.50	0.20	0.20	0.30	0.30	0.10	0.50	0.20	0.20
2	0.50	0.20	0.10	0.40	0.30	0.20	0.40	0.10	0.30

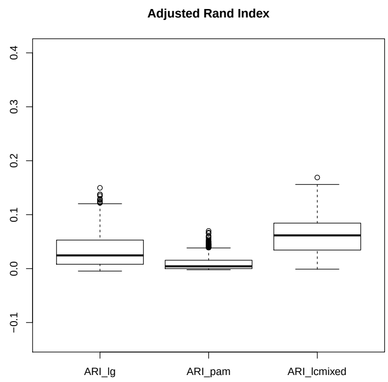
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.50	0.20	0.30	0.40	0.10	0.30	0.30	0.20	0.20
2	0.50	0.30	0.30	0.20	0.20	0.20	0.30	0.40	0.10

Table A.38: Summary: 4 4-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.005	-0.002	-0.001
1st Qu.		0.008	0.000	0.034
Median		0.024	0.004	0.062
Mean (se)		0.033 (0.001)	0.009 (0.000)	0.060 (0.001)
3rd Qu.		0.053	0.015	0.084
Max.		0.150	0.070	0.169
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.017	-0.039	0.115	0.033
1st Qu.	0.027	0.085	0.129	0.115
Median	0.029	0.106	0.133	0.125
Mean (se)	0.029 (0.000)	0.104 (0.001)	0.133 (0.000)	0.126 (0.000)
3rd Qu.	0.032	0.125	0.137	0.136
Max.	0.045	0.204	0.157	0.204
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.029	-0.034	0.184	0.003
1st Qu.	0.045	0.091	0.215	0.192
Median	0.050	0.143	0.225	0.216
Mean (se)	0.050 (0.000)	0.146 (0.071)	0.224 (0.002)	0.211 (0.000)
3rd Qu.	0.055	0.202	0.235	0.237
Max.	0.079	0.351	0.271	0.351

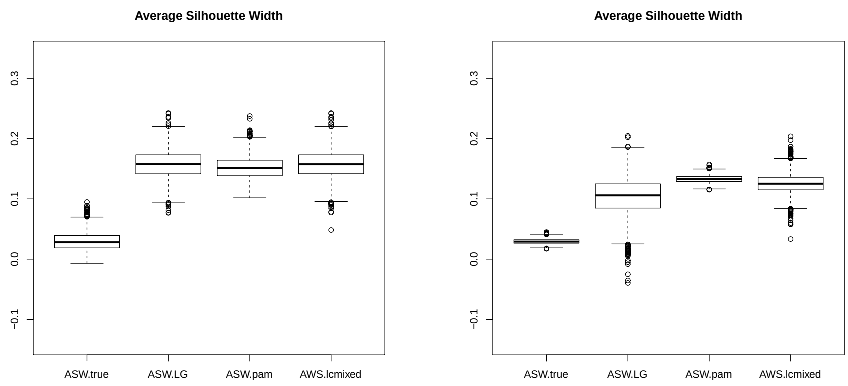


(a) *ARI - 100 units.*



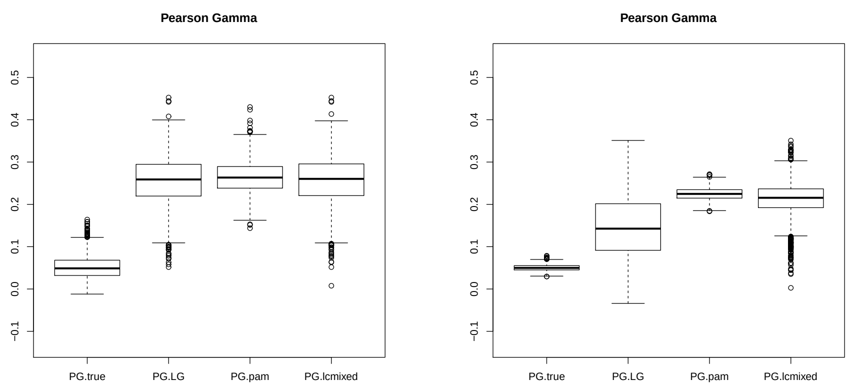
(b) *ARI - 1000 units.*

Figure A.55: Adjusted Rand Index: 4 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.56: Average Silhouette Width: 4 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.57: Pearson Gamma: 4 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

Table A.39: Summary: 4 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.136	-0.024	0.152
1st Qu.		0.573	0.056	0.573
Median		0.636	0.093	0.636
Mean (se)		0.644 (0.002)	0.098 (0.001)	0.644 (0.002)
3rd Qu.		0.737	0.136	0.737
Max.		0.921	0.378	0.921
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.053	0.020	0.486	0.020
1st Qu.	0.165	0.168	0.508	0.169
Median	0.199	0.216	0.520	0.216
Mean (se)	0.200 (0.001)	0.218 (0.002)	0.522 (0.000)	0.218 (0.002)
3rd Qu.	0.233	0.266	0.534	0.266
Max.	0.381	0.488	0.598	0.488
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.050	0.017	0.427	0.017
1st Qu.	0.169	0.172	0.459	0.172
Median	0.206	0.225	0.476	0.225
Mean (se)	0.209 (0.001)	0.230 (0.002)	0.480 (0.001)	0.230 (0.002)
3rd Qu.	0.246	0.283	0.499	0.285
Max.	0.430	0.559	0.591	0.559

A.20 4 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

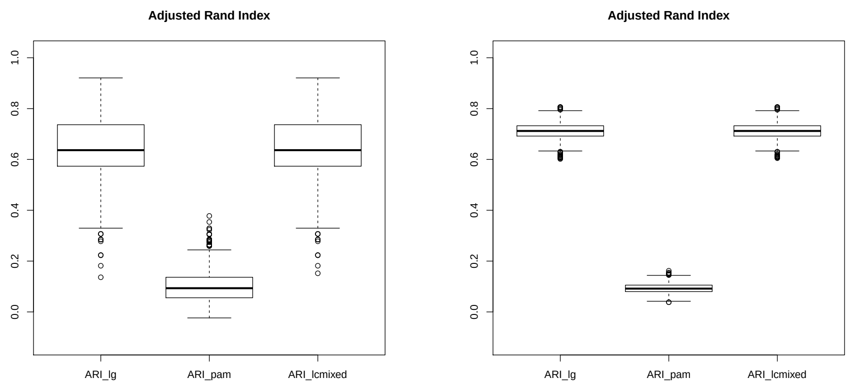
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$
1	0.50	0.10	0.40	0.40	0.10	0.10	0.10	0.10	0.70
2	0.50	0.40	0.10	0.10	0.40	0.10	0.70	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.50	0.40	0.10	0.40	0.10	0.10	0.10	0.40	0.40
2	0.50	0.10	0.40	0.10	0.40	0.40	0.40	0.10	0.10

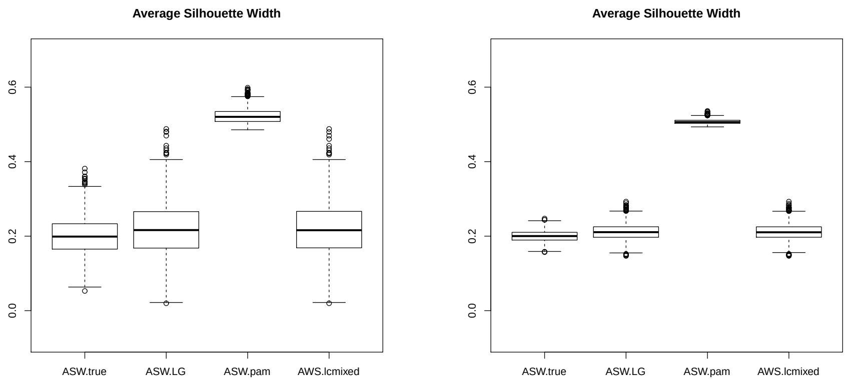
Table A.40: Summary: 4 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.602	0.038	0.605
1st Qu.		0.692	0.080	0.692
Median		0.712	0.092	0.712
Mean (se)		0.711 (0.001)	0.093 (0.000)	0.711 (0.001)
3rd Qu.		0.732	0.106	0.732
Max.		0.806	0.162	0.806
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.157	0.147	0.493	0.147
1st Qu.	0.190	0.197	0.503	0.197
Median	0.200	0.211	0.506	0.210
Mean (se)	0.200 (0.000)	0.212 (0.000)	0.507 (0.000)	0.212 (0.000)
3rd Qu.	0.210	0.225	0.511	0.225
Max.	0.247	0.293	0.536	0.293
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.160	0.149	0.437	0.149
1st Qu.	0.196	0.204	0.451	0.204
Median	0.208	0.220	0.456	0.220
Mean (se)	0.208 (0.000)	0.221 (0.001)	0.458 (0.000)	0.221 (0.001)
3rd Qu.	0.220	0.237	0.463	0.236
Max.	0.262	0.317	0.500	0.317



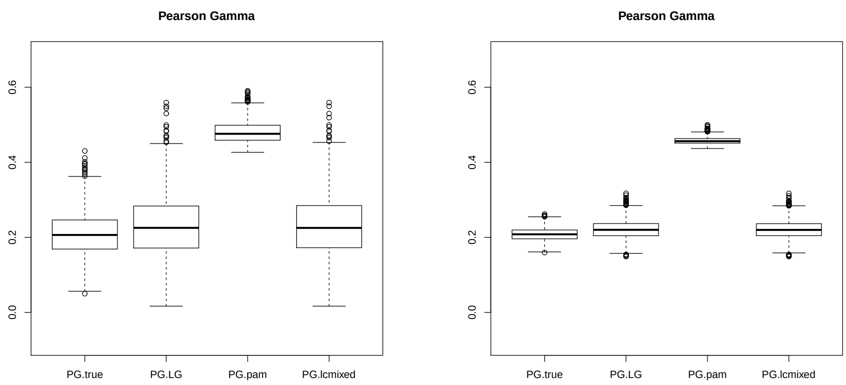
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.58: Adjusted Rand Index: 4 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation



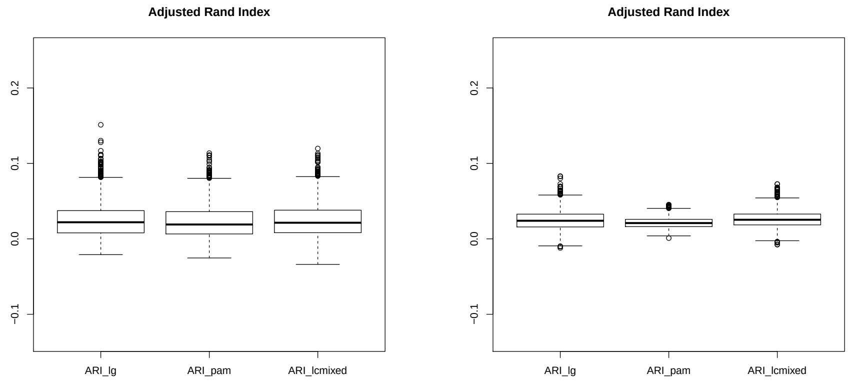
(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.59: Average Silhouette Width: 4 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.60: Pearson Gamma: 4 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.61: Adjusted Rand Index: 4 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

A.21 4 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

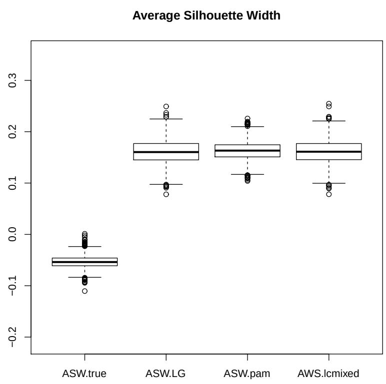
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.10	0.20	0.10	0.30	0.40	0.40	0.20	0.20	0.20
2	0.15	0.20	0.20	0.10	0.50	0.30	0.10	0.30	0.30
3	0.20	0.40	0.30	0.20	0.10	0.30	0.30	0.20	0.20
4	0.25	0.30	0.20	0.20	0.30	0.40	0.10	0.20	0.30
5	0.30	0.20	0.30	0.20	0.30	0.30	0.20	0.30	0.30

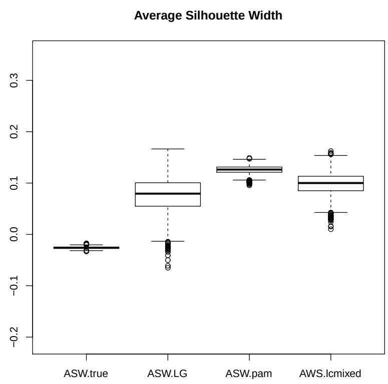
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.10	0.20	0.10	0.50	0.20	0.30	0.50	0.10	0.10
2	0.15	0.40	0.30	0.10	0.20	0.20	0.20	0.20	0.40
3	0.20	0.20	0.20	0.30	0.30	0.30	0.10	0.30	0.30
4	0.25	0.20	0.30	0.20	0.30	0.40	0.30	0.10	0.20
5	0.30	0.30	0.30	0.10	0.30	0.40	0.20	0.30	0.10

Table A.41: Summary: 4 4-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.021	-0.025	-0.034
1st Qu.		0.008	0.007	0.008
Median		0.022	0.019	0.021
Mean (se)		0.025 (0.001)	0.023 (0.000)	0.024 (0.001)
3rd Qu.		0.037	0.036	0.038
Max.		0.151	0.114	0.120
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.110	0.078	0.104	0.078
1st Qu.	-0.061	0.145	0.151	0.146
Median	-0.054	0.160	0.163	0.161
Mean (se)	-0.053 (0.000)	0.161 (0.001)	0.163 (0.000)	0.161 (0.001)
3rd Qu.	-0.046	0.177	0.174	0.177
Max.	0.001	0.249	0.226	0.255
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.008	0.262	0.318	0.260
1st Qu.	0.039	0.368	0.378	0.369
Median	0.055	0.389	0.396	0.391
Mean (se)	0.056 (0.000)	0.390 (0.001)	0.395 (0.001)	0.391 (0.001)
3rd Qu.	0.070	0.411	0.411	0.412
Max.	0.150	0.501	0.479	0.501



(a) *ASW - 100 units.*

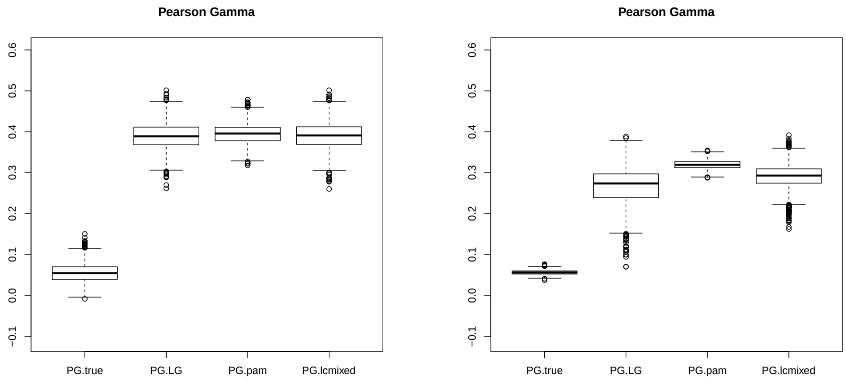


(b) *ASW - 1000 units.*

Figure A.62: Average Silhouette Width: 4 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

Table A.42: Summary: 4 4-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.012	0.001	-0.008
1st Qu.		0.016	0.016	0.019
Median		0.024	0.021	0.025
Mean (se)		0.025 (0.000)	0.021 (0.000)	0.026 (0.000)
3rd Qu.		0.033	0.026	0.033
Max.		0.083	0.045	0.073
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.033	-0.065	0.095	0.010
1st Qu.	-0.027	0.055	0.121	0.085
Median	-0.026	0.079	0.126	0.100
Mean (se)	-0.026 (0.001)	0.076 (0.000)	0.126 (0.000)	0.098 (0.000)
3rd Qu.	-0.025	0.101	0.131	0.113
Max.	-0.017	0.166	0.149	0.162
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.037	0.070	0.288	0.163
1st Qu.	0.052	0.239	0.312	0.275
Median	0.056	0.274	0.319	0.293
Mean (se)	0.056 (0.001)	0.266 (0.001)	0.320 (0.000)	0.291 (0.001)
3rd Qu.	0.060	0.297	0.328	0.309
Max.	0.076	0.389	0.355	0.392
NA's		1		



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.63: Pearson Gamma: 4 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

A.22 4 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.10	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
2	0.15	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70
3	0.20	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
4	0.25	0.10	0.10	0.10	0.70	0.10	0.10	0.70	0.10
5	0.30	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10

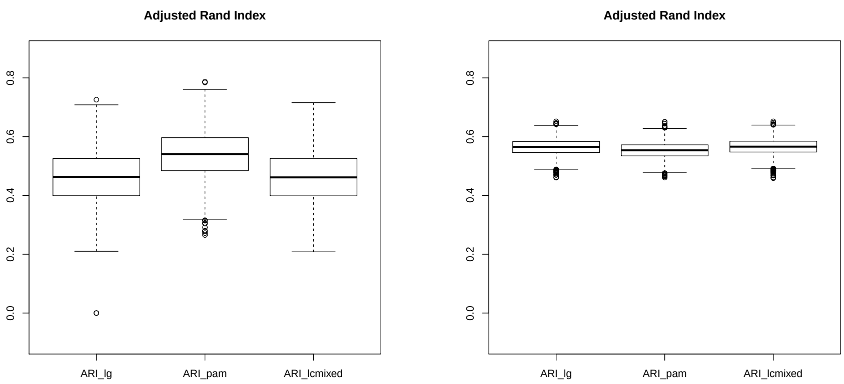
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.10	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
2	0.15	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10
3	0.20	0.10	0.70	0.10	0.10	0.10	0.70	0.10	0.10
4	0.25	0.10	0.10	0.10	0.70	0.10	0.10	0.70	0.10
5	0.30	0.10	0.10	0.10	0.70	0.10	0.10	0.10	0.70

Table A.43: Summary: 4 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.265	0.208
1st Qu.		0.399	0.484	0.399
Median		0.463	0.540	0.462
Mean (se)		0.463 (0.002)	0.539 (0.002)	0.462 (0.002)
3rd Qu.		0.526	0.597	0.526
Max.		0.726	0.787	0.716
	ASW true	**ASW lg**	**ASW pam**	**ASW lcmixed**
Min.	0.098	0.170	0.214	0.173
1st Qu.	0.216	0.290	0.297	0.289
Median	0.244	0.316	0.322	0.316
Mean (se)	0.245 (0.001)	0.316 (0.001)	0.322 (0.001)	0.316 (0.001)
3rd Qu.	0.271	0.342	0.345	0.342
Max.	0.382	0.456	0.457	0.456
NA's		2		
	PG true	**PG lg**	**PG pam**	**PG lcmixed**
Min.	0.368	0.434	0.459	0.414
1st Qu.	0.487	0.556	0.568	0.557
Median	0.518	0.586	0.593	0.586
Mean (sd)	0.517 (0.001)	0.584 (0.001)	0.592 (0.003)	0.584 (0.001)
3rd Qu.	0.548	0.614	0.616	0.614
Max.	0.638	0.715	0.720	0.706
NA's		2		

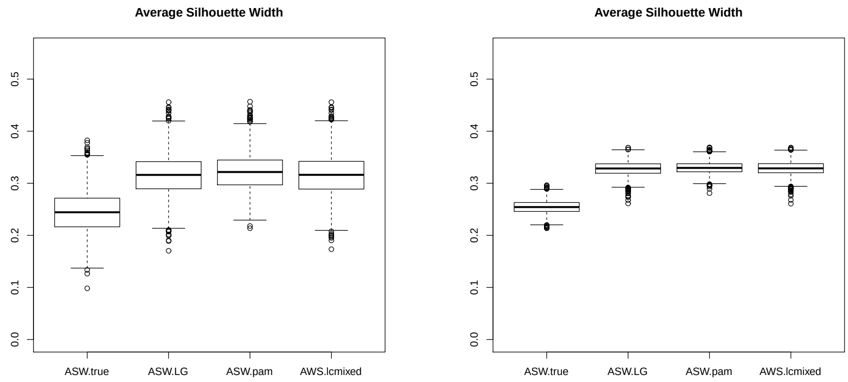
Table A.44: Summary: 4 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.461	0.461	0.459
1st Qu.		0.546	0.535	0.548
Median		0.565	0.553	0.566
Mean (se)		0.564 (0.001)	0.554 (0.001)	0.565 (0.001)
3rd Qu.		0.584	0.572	0.585
Max.		0.652	0.651	0.652
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.214	0.261	0.281	0.261
1st Qu.	0.246	0.319	0.322	0.320
Median	0.254	0.328	0.330	0.329
Mean (se)	0.254 (0.000)	0.328 (0.000)	0.330 (0.000)	0.328 (0.000)
3rd Qu.	0.263	0.337	0.338	0.338
Max.	0.296	0.368	0.369	0.368
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.473	0.513	0.552	0.509
1st Qu.	0.509	0.589	0.589	0.590
Median	0.518	0.598	0.597	0.599
Mean (se)	0.518 (0.000)	0.598 (0.000)	0.597 (0.000)	0.598 (0.000)
3rd Qu.	0.528	0.608	0.605	0.608
Max.	0.562	0.637	0.633	0.637



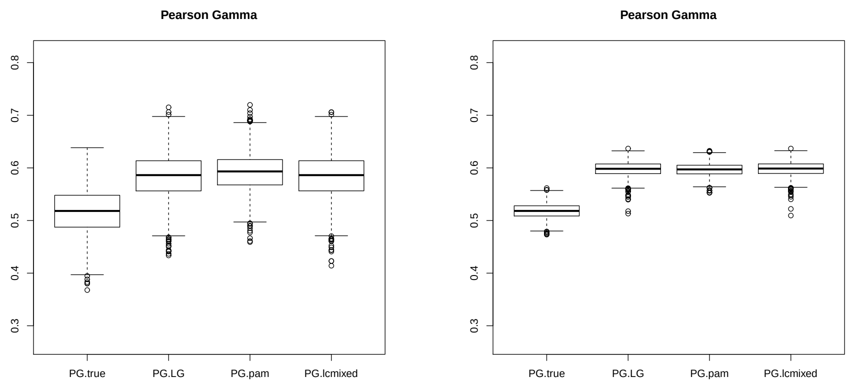
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.64: Adjusted Rand Index: 4 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



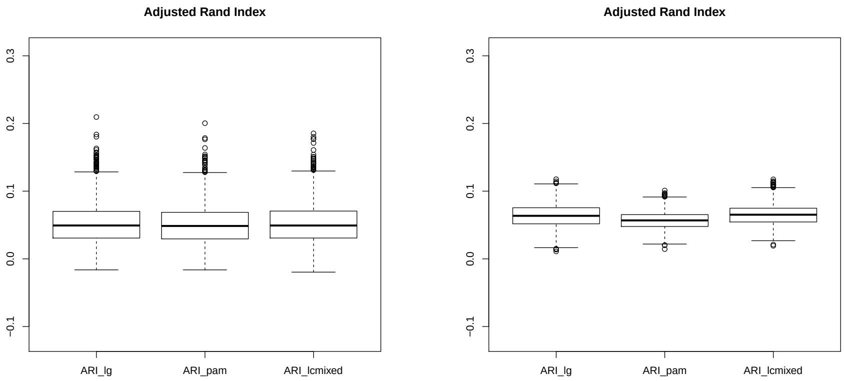
(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.65: Average Silhouette Width: 4 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.66: Pearson Gamma: 4 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.67: Adjusted Rand Index: 4 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

A.23 4 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.20	0.20	0.50	0.10	0.20	0.30	0.10	0.10	0.50
2	0.20	0.20	0.40	0.30	0.10	0.20	0.30	0.40	0.10
3	0.20	0.30	0.30	0.20	0.20	0.40	0.10	0.30	0.20
4	0.20	0.30	0.10	0.40	0.20	0.20	0.40	0.30	0.10
5	0.20	0.40	0.10	0.40	0.10	0.20	0.30	0.30	0.20

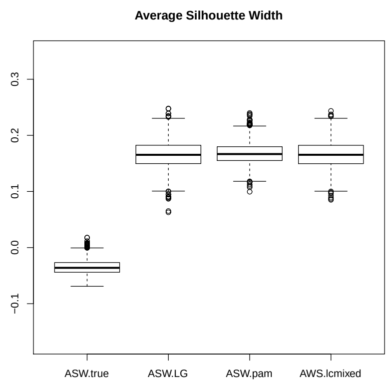
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.20	0.20	0.20	0.30	0.30	0.30	0.20	0.10	0.40
2	0.20	0.40	0.20	0.20	0.20	0.20	0.50	0.10	0.20
3	0.20	0.10	0.30	0.30	0.30	0.40	0.10	0.20	0.30
4	0.20	0.20	0.50	0.10	0.20	0.30	0.30	0.20	0.20
5	0.20	0.30	0.40	0.20	0.10	0.20	0.40	0.20	0.20

A.24 4 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

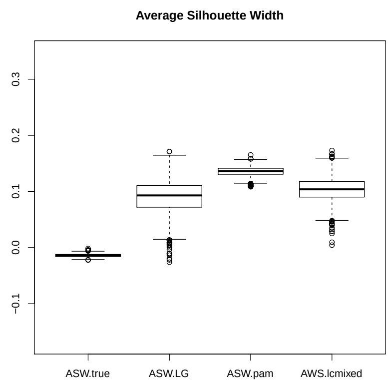
The Model:

Table A.45: Summary: 4 4-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.016	-0.016	-0.020
1st Qu.		0.031	0.030	0.031
Median		0.049	0.049	0.049
Mean (se)		0.053 (0.001)	0.051 (0.001)	0.052 (0.001)
3rd Qu.		0.070	0.069	0.071
Max.		0.210	0.200	0.185
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.069	0.063	0.100	0.085
1st Qu.	-0.044	0.150	0.155	0.150
Median	-0.036	0.165	0.167	0.165
Mean (se)	-0.035 (0.000)	0.166 (0.001)	0.168 (0.000)	0.166 (0.001)
3rd Qu.	-0.027	0.182	0.180	0.182
Max.	0.018	0.248	0.240	0.244
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.024	0.200	0.318	0.278
1st Qu.	0.080	0.376	0.386	0.376
Median	0.096	0.398	0.403	0.398
Mean (se)	0.098 (0.001)	0.399 (0.001)	0.404 (0.001)	0.399 (0.001)
3rd Qu.	0.115	0.421	0.421	0.421
Max.	0.195	0.518	0.507	0.514



(a) *ASW - 100 units.*

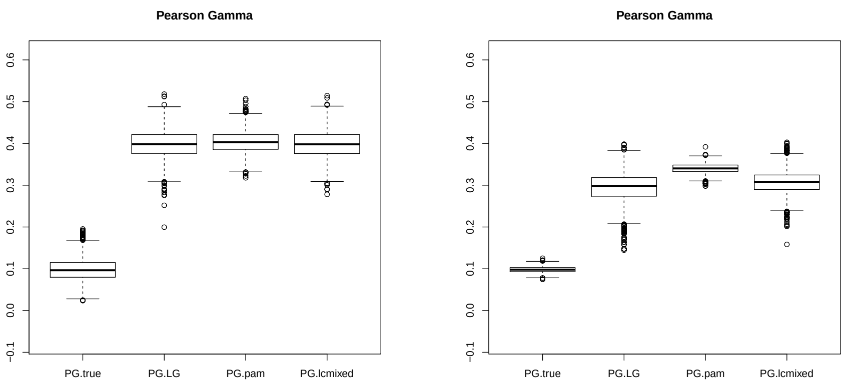


(b) *ASW - 1000 units.*

Figure A.68: Average Silhouette Width: 4 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

Table A.46: Summary: 4 4-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.011	0.014	0.019
1st Qu.		0.052	0.048	0.055
Median		0.064	0.057	0.065
Mean (se)		0.063 (0.000)	0.057 (0.000)	0.065 (0.000)
3rd Qu.		0.075	0.065	0.075
Max.		0.118	0.101	0.118
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.022	-0.026	0.108	0.004
1st Qu.	-0.016	0.072	0.131	0.090
Median	-0.014	0.093	0.136	0.104
Mean (se)	-0.014 (0.000)	0.090 (0.001)	0.136 (0.000)	0.104 (0.000)
3rd Qu.	-0.012	0.111	0.141	0.118
Max.	-0.002	0.171	0.165	0.173
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.075	0.144	0.298	0.158
1st Qu.	0.093	0.274	0.333	0.290
Median	0.098	0.298	0.340	0.308
Mean (se)	0.098 (0.000)	0.294 (0.001)	0.340 (0.000)	0.307 (0.001)
3rd Qu.	0.103	0.318	0.348	0.324
Max.	0.125	0.398	0.392	0.403



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.69: Pearson Gamma: 4 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

Table A.47: Summary: 4 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 100 units

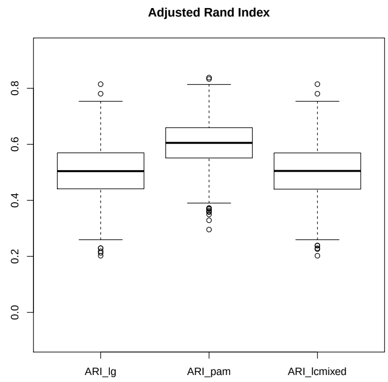
		ARI lg	ARI pam	ARI lcmixed
Min.		0.202	0.296	0.202
1st Qu.		0.441	0.551	0.440
Median		0.504	0.605	0.505
Mean (se)		0.505 (0.002)	0.604 (0.002)	0.505 (0.002)
3rd Qu.		0.570	0.660	0.569
Max.		0.815	0.838	0.815
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.135	0.191	0.219	0.191
1st Qu.	0.241	0.306	0.317	0.306
Median	0.269	0.332	0.342	0.332
Mean (se)	0.269 (0.001)	0.331 (0.001)	0.342 (0.001)	0.332 (0.001)
3rd Qu.	0.297	0.360	0.366	0.359
Max.	0.408	0.505	0.514	0.505
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.361	0.424	0.471	0.425
1st Qu.	0.498	0.564	0.582	0.564
Median	0.528	0.593	0.606	0.593
Mean (se)	0.528 (0.001)	0.591 (0.001)	0.606 (0.001)	0.591 (0.001)
3rd Qu.	0.559	0.621	0.630	0.621
Max.	0.659	0.729	0.741	0.729

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.20	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10
2	0.20	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
3	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70
4	0.20	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10
5	0.20	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10

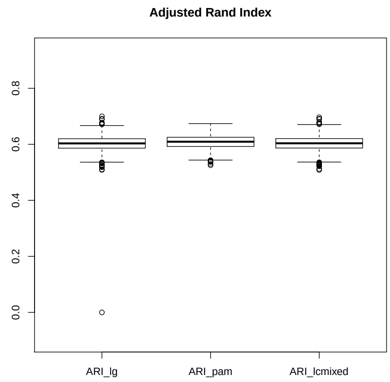
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
2	0.20	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
3	0.20	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10
4	0.20	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
5	0.20	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10

Table A.48: Summary: 4 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.525	0.508
1st Qu.		0.586	0.592	0.587
Median		0.603	0.609	0.604
Mean (se)		0.603 (0.001)	0.609 (0.001)	0.604 (0.001)
3rd Qu.		0.620	0.625	0.621
Max.		0.670	0.674	0.697
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.221	0.313	0.314	0.314
1st Qu.	0.270	0.343	0.342	0.343
Median	0.278	0.350	0.350	0.350
Mean (sd)	0.278 (0.000)	0.350 (0.000)	0.350 (0.000)	0.350 (0.000)
3rd Qu.	0.287	0.358	0.357	0.358
Max.	0.319	0.384	0.382	0.384
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.466	0.567	0.568	0.566
1st Qu.	0.520	0.602	0.600	0.603
Median	0.529	0.610	0.607	0.610
Mean (se)	0.529 (0.000)	0.610 (0.000)	0.607 (0.000)	0.610 (0.000)
3rd Qu.	0.539	0.618	0.615	0.618
Max.	0.574	0.642	0.639	0.642
NA's		1		

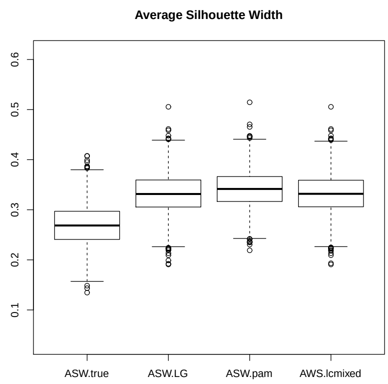


(a) *ARI - 100 units.*

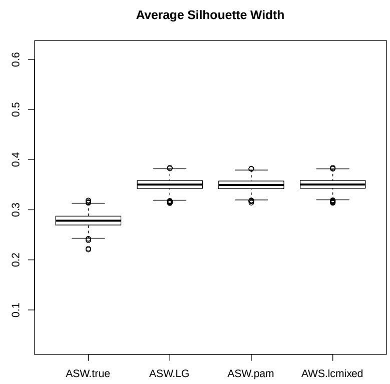


(b) *ARI - 1000 units.*

Figure A.70: Adjusted Rand Index: 4 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

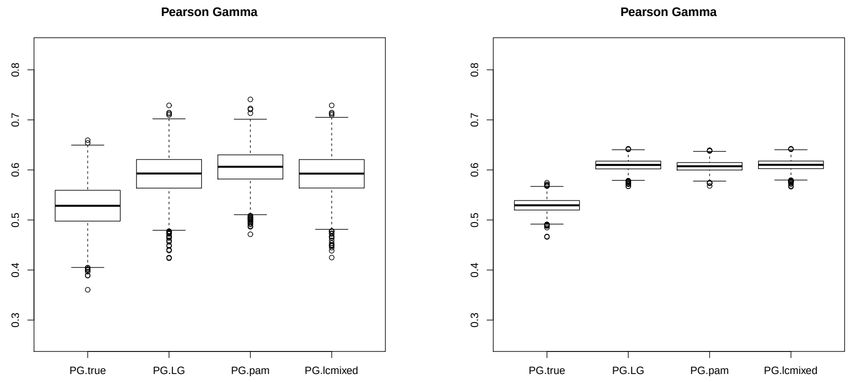


(a) *ASW - 100 units.*



(b) *ASW - 1000 units.*

Figure A.71: Average Silhouette Width: 4 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.72: Pearson Gamma: 4 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

A.25 12 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$
1	0.15	0.20	0.30	0.40	0.10	0.30	0.10	0.20	0.40
2	0.85	0.30	0.20	0.30	0.20	0.20	0.40	0.30	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.15	0.20	0.30	0.30	0.20	0.20	0.20	0.40	0.20
2	0.85	0.10	0.20	0.20	0.50	0.30	0.40	0.10	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1X}$	$\hat{\pi}_{1t}^{\bar{E}_2X}$	$\hat{\pi}_{1t}^{\bar{E}_3X}$	$\hat{\pi}_{1t}^{\bar{E}_4X}$	$\hat{\pi}_{1t}^{\bar{F}_1X}$	$\hat{\pi}_{1t}^{\bar{F}_2X}$	$\hat{\pi}_{1t}^{\bar{F}_3X}$	$\hat{\pi}_{1t}^{\bar{F}_4X}$
1	0.15	0.10	0.50	0.10	0.30	0.20	0.20	0.20	0.40
2	0.85	0.20	0.30	0.20	0.30	0.10	0.30	0.10	0.50

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1X}$	$\hat{\pi}_{1t}^{\bar{G}_2X}$	$\hat{\pi}_{1t}^{\bar{G}_3X}$	$\hat{\pi}_{1t}^{\bar{G}_4X}$	$\hat{\pi}_{1t}^{\bar{H}_1X}$	$\hat{\pi}_{1t}^{\bar{H}_2X}$	$\hat{\pi}_{1t}^{\bar{H}_3X}$	$\hat{\pi}_{1t}^{\bar{H}_4X}$
1	0.15	0.20	0.10	0.30	0.40	0.20	0.40	0.10	0.30
2	0.85	0.30	0.20	0.10	0.40	0.40	0.10	0.20	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1X}$	$\hat{\pi}_{1t}^{\bar{I}_2X}$	$\hat{\pi}_{1t}^{\bar{I}_3X}$	$\hat{\pi}_{1t}^{\bar{I}_4X}$	$\hat{\pi}_{1t}^{\bar{L}_1X}$	$\hat{\pi}_{1t}^{\bar{L}_2X}$	$\hat{\pi}_{1t}^{\bar{L}_3X}$	$\hat{\pi}_{1t}^{\bar{L}_4X}$
1	0.15	0.20	0.30	0.10	0.40	0.30	0.30	0.10	0.30
2	0.85	0.20	0.20	0.50	0.10	0.20	0.40	0.10	0.30

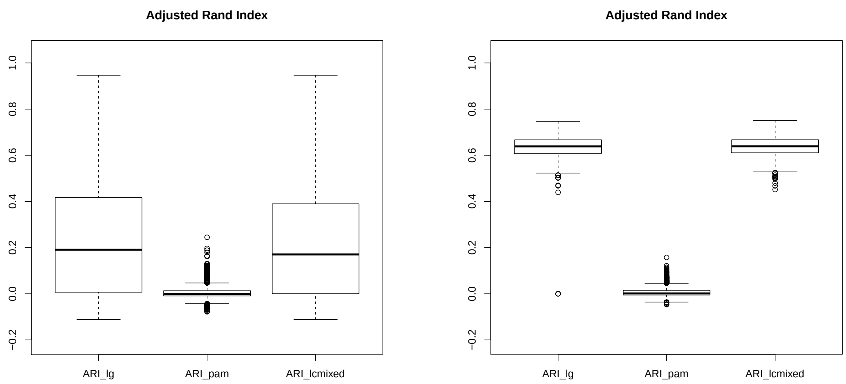
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1X}$	$\hat{\pi}_{1t}^{\bar{M}_2X}$	$\hat{\pi}_{1t}^{\bar{M}_3X}$	$\hat{\pi}_{1t}^{\bar{M}_4X}$	$\hat{\pi}_{1t}^{\bar{N}_1X}$	$\hat{\pi}_{1t}^{\bar{N}_2X}$	$\hat{\pi}_{1t}^{\bar{N}_3X}$	$\hat{\pi}_{1t}^{\bar{N}_4X}$
1	0.15	0.20	0.10	0.50	0.20	0.30	0.10	0.10	0.50
2	0.85	0.20	0.30	0.10	0.40	0.20	0.20	0.30	0.30

A.26 12 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Table A.49: Summary: 12 4-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.112	-0.078	-0.112
1st Qu.		0.007	-0.010	0.000
Median		0.191	-0.002	0.170
Mean (se)		0.224 (0.005)	0.005 (0.001)	0.210 (0.005)
3rd Qu.		0.417	0.013	0.390
Max.		0.947	0.245	0.947
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.019	0.004	0.054	0.008
1st Qu.	0.040	0.057	0.082	0.057
Median	0.051	0.070	0.091	0.069
Mean (se)	0.051 (0.000)	0.071 (0.000)	0.091 (0.000)	0.070 (0.000)
3rd Qu.	0.062	0.084	0.100	0.084
Max.	0.119	0.150	0.130	0.139
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.033	-0.062	0.096	-0.062
1st Qu.	0.078	0.109	0.183	0.109
Median	0.104	0.145	0.206	0.144
Mean (se)	0.104 (0.001)	0.145 (0.001)	0.206 (0.001)	0.144 (0.001)
3rd Qu.	0.131	0.180	0.229	0.180
Max.	0.276	0.313	0.303	0.313

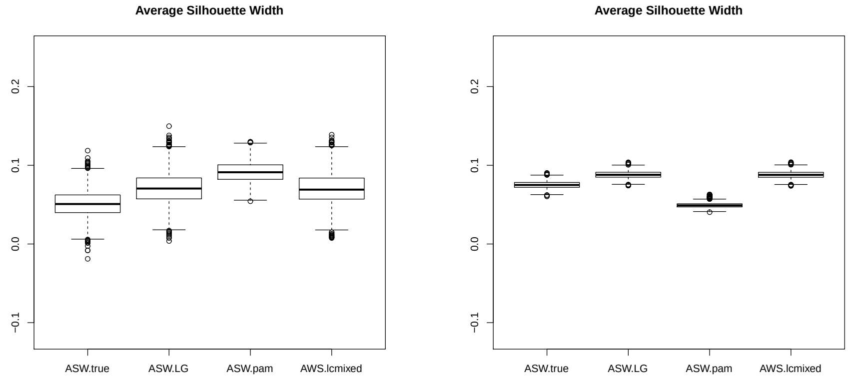


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.73: Adjusted Rand Index: 12 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

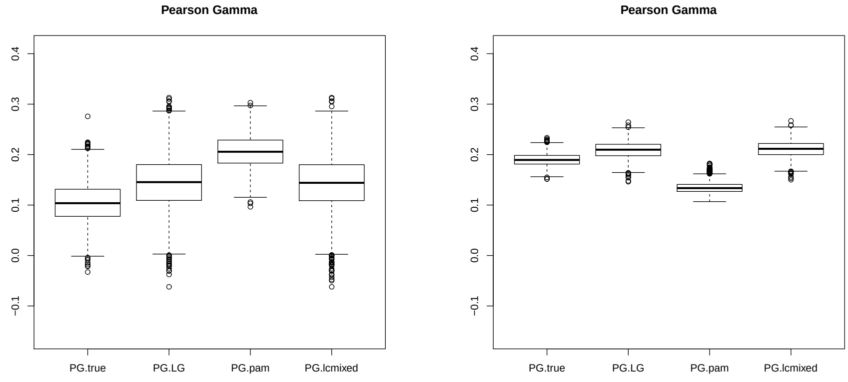
Table A.50: Summary: 4 4-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	-0.047	0.452
1st Qu.		0.608	-0.005	0.611
Median		0.639	0.001	0.639
Mean (se)		0.637 (0.001)	0.007 (0.001)	0.638 (0.001)
3rd Qu.		0.667	0.015	0.667
Max.		0.746	0.157	0.751
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.060	0.074	0.040	0.074
1st Qu.	0.072	0.085	0.047	0.085
Median	0.075	0.088	0.049	0.088
Mean (se)	0.075 (0.000)	0.088 (0.000)	0.049 (0.000)	0.088 (0.000)
3rd Qu.	0.078	0.091	0.051	0.091
Max.	0.090	0.104	0.063	0.104
NA's		3		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.152	0.146	0.107	0.150
1st Qu.	0.181	0.198	0.127	0.200
Median	0.189	0.210	0.134	0.212
Mean (se)	0.190 (0.000)	0.209 (0.000)	0.135 (0.000)	0.211 (0.000)
3rd Qu.	0.198	0.220	0.141	0.222
Max.	0.233	0.264	0.182	0.267
NA's		3		



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.74: Average Silhouette Width: 12 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.75: Pearson Gamma: 12 4-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.15	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10
2	0.85	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.15	0.10	0.70	0.10	0.10	0.70	0.10	0.10	0.10
2	0.85	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_4 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_4 X}$
1	0.15	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
2	0.85	0.10	0.70	0.10	0.10	0.70	0.10	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.15	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
2	0.85	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.15	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
2	0.85	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10

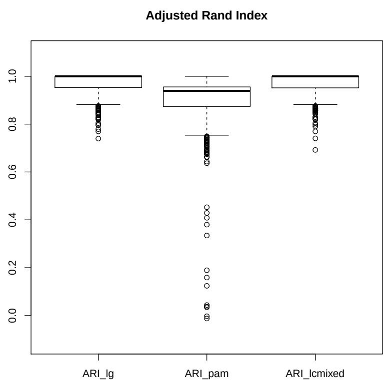
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$
1	0.15	0.10	0.10	0.10	0.70	0.10	0.10	0.70	0.10
2	0.85	0.10	0.10	0.70	0.10	0.10	0.10	0.10	0.70

A.27 12 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

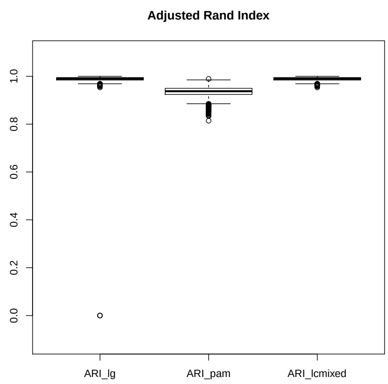
The Model:

Table A.51: Summary: 12 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.740	-0.012	0.692
1st Qu.		0.953	0.874	0.952
Median		1.000	0.939	1.000
Mean (se)		0.980 (0.001)	0.912 (0.002)	0.979 (0.001)
3rd Qu.		1.000	0.956	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.298	0.298	0.065	0.286
1st Qu.	0.378	0.378	0.375	0.378
Median	0.398	0.398	0.396	0.398
Mean (se)	0.398 (0.001)	0.398 (0.001)	0.395 (0.001)	0.398 (0.001)
3rd Qu.	0.418	0.418	0.417	0.418
Max.	0.493	0.493	0.493	0.493
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.439	0.439	0.081	0.439
1st Qu.	0.624	0.626	0.638	0.626
Median	0.658	0.660	0.668	0.660
Mean (se)	0.655 (0.001)	0.656 (0.001)	0.664 (0.001)	0.657 (0.001)
3rd Qu.	0.692	0.692	0.698	0.692
Max.	0.779	0.779	0.779	0.779



(a) *ARI - 100 units.*

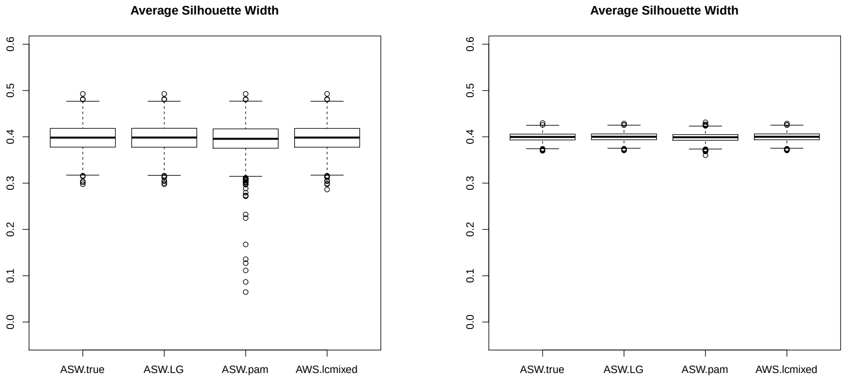


(b) *ARI - 1000 units.*

Figure A.76: Adjusted Rand Index: 12 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

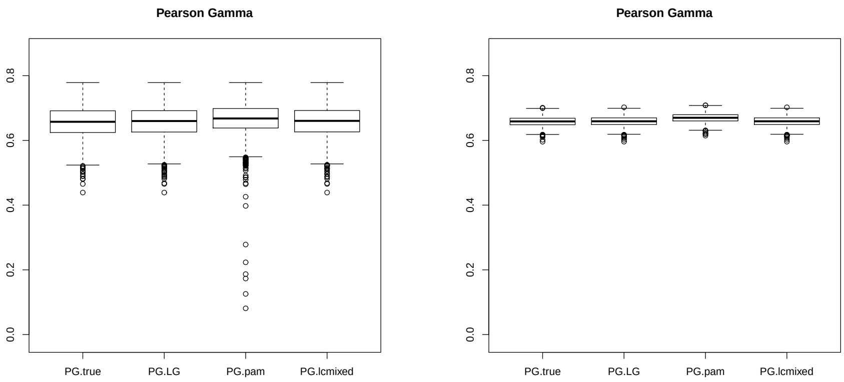
Table A.52: Summary: 12 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.814	0.953
1st Qu.		0.984	0.924	0.984
Median		0.990	0.938	0.990
Mean (se)		0.988 (0.001)	0.934 (0.000)	0.989 (0.000)
3rd Qu.		0.995	0.950	0.995
Max.		1.000	0.989	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.370	0.371	0.360	0.371
1st Qu.	0.393	0.394	0.393	0.394
Median	0.400	0.400	0.399	0.400
Mean (sd)	0.399 (0.000)	0.400 (0.000)	0.399 (0.000)	0.400 (0.000)
3rd Qu.	0.406	0.406	0.405	0.406
Max.	0.430	0.429	0.431	0.429
NA's		3		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.596	0.596	0.615	0.596
1st Qu.	0.648	0.649	0.660	0.649
Median	0.659	0.659	0.670	0.659
Mean (se)	0.658 (0.000)	0.659 (0.000)	0.670 (0.000)	0.659 (0.000)
3rd Qu.	0.669	0.670	0.680	0.670
Max.	0.701	0.702	0.709	0.702
NA's		3		



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.77: Average Silhouette Width: 12 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.78: Pearson Gamma: 12 4-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.50	0.20	0.10	0.20	0.50	0.40	0.10	0.20	0.30
2	0.50	0.30	0.20	0.20	0.30	0.30	0.10	0.40	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.50	0.20	0.20	0.30	0.30	0.10	0.20	0.50	0.20
2	0.50	0.30	0.20	0.20	0.30	0.20	0.10	0.40	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_4 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_4 X}$
1	0.50	0.30	0.10	0.20	0.40	0.10	0.30	0.20	0.40
2	0.50	0.20	0.20	0.40	0.20	0.20	0.40	0.10	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.50	0.50	0.10	0.20	0.20	0.30	0.20	0.20	0.30
2	0.50	0.40	0.20	0.20	0.20	0.40	0.10	0.20	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.50	0.20	0.40	0.10	0.30	0.10	0.30	0.30	0.30
2	0.50	0.30	0.50	0.10	0.10	0.20	0.30	0.20	0.30

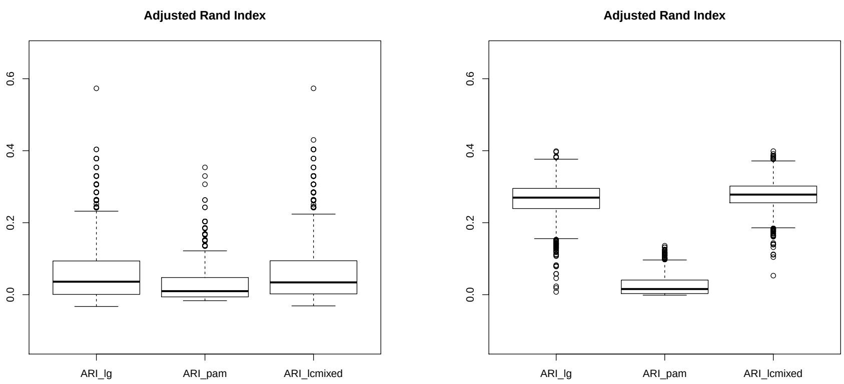
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$
1	0.50	0.20	0.30	0.10	0.40	0.50	0.10	0.20	0.20
2	0.50	0.40	0.10	0.20	0.30	0.40	0.20	0.30	0.10

A.28 12 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

The Model:

Table A.53: Summary: 12 4-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.033	-0.017	-0.031
1st Qu.		0.001	-0.006	0.002
Median		0.036	0.010	0.034
Mean (se)		0.060 (0.002)	0.026 (0.001)	0.062 (0.002)
3rd Qu.		0.094	0.048	0.094
Max.		0.573	0.354	0.573
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.007	0.033	0.029	0.033
1st Qu.	0.025	0.065	0.053	0.065
Median	0.032	0.072	0.059	0.072
Mean (se)	0.032 (0.000)	0.072 (0.000)	0.060 (0.000)	0.073 (0.000)
3rd Qu.	0.038	0.079	0.065	0.080
Max.	0.072	0.112	0.103	0.111
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.009	0.015	0.077	0.022
1st Qu.	0.071	0.154	0.149	0.157
Median	0.090	0.187	0.167	0.189
Mean (se)	0.091 (0.001)	0.183 (0.001)	0.168 (0.001)	0.185 (0.001)
3rd Qu.	0.109	0.216	0.186	0.217
Max.	0.210	0.314	0.289	0.314

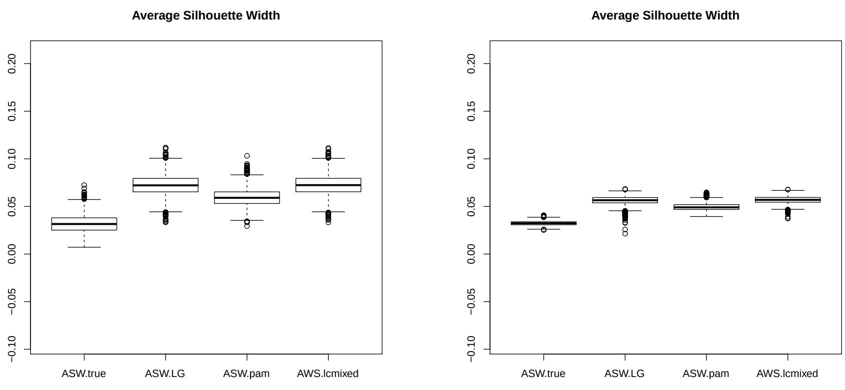


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.79: Adjusted Rand Index: 12 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

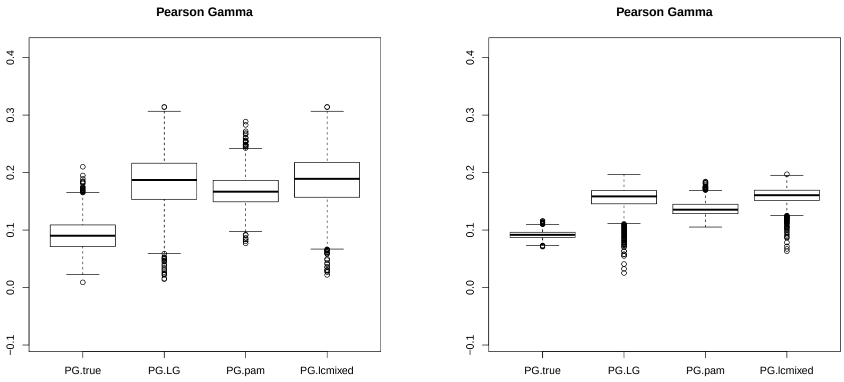
Table A.54: Summary: 12 4-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.008	-0.001	0.053
1st Qu.		0.239	0.003	0.255
Median		0.270	0.016	0.278
Mean (se)		0.264 (0.001)	0.025 (0.001)	0.277 (0.001)
3rd Qu.		0.295	0.041	0.302
Max.		0.399	0.136	0.399
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.025	0.021	0.039	0.037
1st Qu.	0.031	0.054	0.047	0.054
Median	0.032	0.056	0.049	0.057
Mean (se)	0.032 (0.000)	0.056 (0.000)	0.050 (0.000)	0.057 (0.000)
3rd Qu.	0.034	0.059	0.052	0.059
Max.	0.041	0.068	0.065	0.068
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.071	0.025	0.105	0.063
1st Qu.	0.087	0.146	0.129	0.152
Median	0.092	0.158	0.135	0.161
Mean (se)	0.092 (0.000)	0.154 (0.000)	0.137 (0.000)	0.159 (0.000)
3rd Qu.	0.096	0.169	0.145	0.169
Max.	0.116	0.197	0.184	0.197



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.80: Average Silhouette Width: 12 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.81: Pearson Gamma: 12 4-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.50	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
2	0.50	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.50	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10
2	0.50	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_4 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_4 X}$
1	0.50	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
2	0.50	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.50	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
2	0.50	0.10	0.10	0.70	0.10	0.10	0.10	0.10	0.70

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.50	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
2	0.50	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$
1	0.50	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70
2	0.50	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10

Table A.55: Summary: 12 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.845	0.636	0.845
1st Qu.		0.960	0.921	0.960
Median		1.000	0.960	1.000
Mean (se)		0.983 (0.001)	0.952 (0.001)	0.983 (0.001)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.329	0.329	0.324	0.329
1st Qu.	0.393	0.394	0.395	0.394
Median	0.411	0.411	0.413	0.411
Mean (se)	0.411 (0.001)	0.412 (0.001)	0.413 (0.001)	0.412 (0.001)
3rd Qu.	0.429	0.429	0.431	0.429
Max.	0.494	0.494	0.494	0.494
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.639	0.618	0.635	0.618
1st Qu.	0.720	0.720	0.724	0.720
Median	0.740	0.741	0.744	0.741
Mean (se)	0.739 (0.001)	0.740 (0.001)	0.743 (0.001)	0.740 (0.001)
3rd Qu.	0.761	0.762	0.764	0.762
Max.	0.831	0.831	0.831	0.831

Table A.56: Summary: 12 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

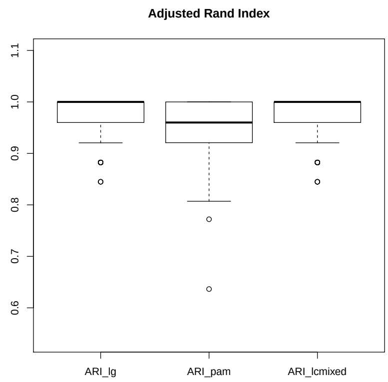
		ARI lg	ARI pam	ARI lcmixed
Min.		0.956	0.918	0.956
1st Qu.		0.984	0.952	0.984
Median		0.988	0.960	0.988
Mean (se)		0.988 (0.000)	0.962 (0.000)	0.988 (0.000)
3rd Qu.		0.992	0.968	0.992
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.383	0.384	0.386	0.384
1st Qu.	0.406	0.407	0.409	0.407
Median	0.412	0.413	0.415	0.413
Mean (se)	0.412 (0.000)	0.412 (0.000)	0.414 (0.000)	0.412 (0.000)
3rd Qu.	0.418	0.418	0.420	0.418
Max.	0.438	0.438	0.439	0.438
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.703	0.704	0.714	0.704
1st Qu.	0.734	0.735	0.739	0.735
Median	0.740	0.741	0.745	0.741
Mean (se)	0.740 (0.000)	0.741 (0.000)	0.745 (0.000)	0.741 (0.000)
3rd Qu.	0.746	0.747	0.751	0.747
Max.	0.768	0.769	0.773	0.769

A.29 12 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

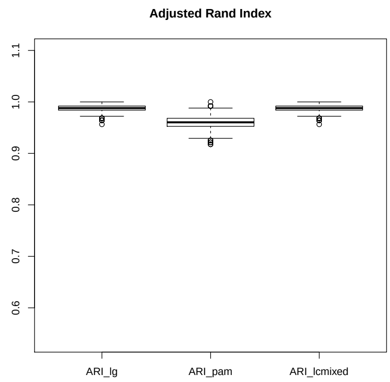
The Model:

A.30 12 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

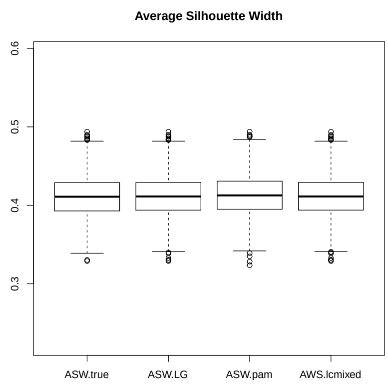


(a) *ARI - 100 units.*

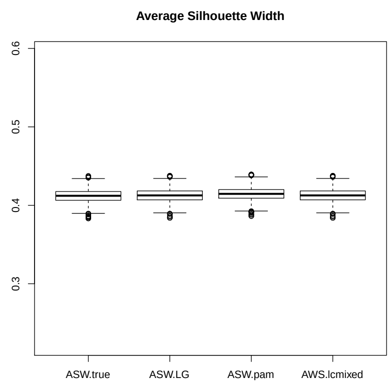


(b) *ARI - 1000 units.*

Figure A.82: Adjusted Rand Index: 12 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

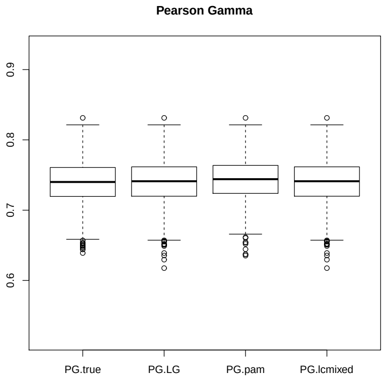


(a) *ASW - 100 units.*

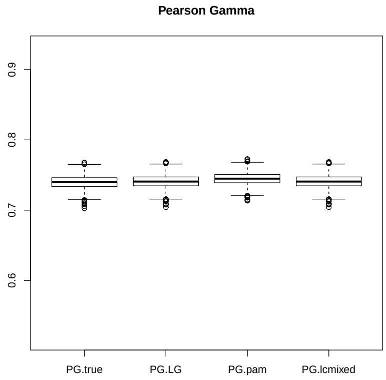


(b) *ASW - 1000 units.*

Figure A.83: Average Silhouette Width: 12 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

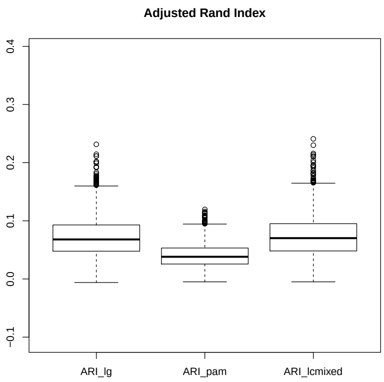


(a) *PG - 100 units.*

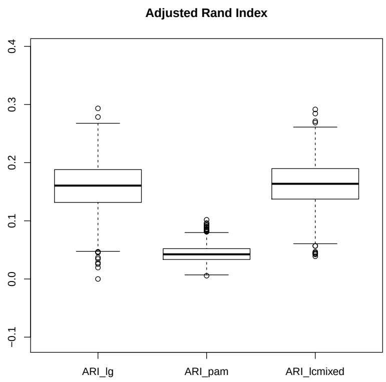


(b) *PG - 1000 units.*

Figure A.84: Pearson Gamma: 12 4-level variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *ARI - 200 units.*



(b) *ARI - 1000 units.*

Figure A.85: Adjusted Rand Index: 12 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$
1	0.10	0.20	0.10	0.30	0.40	0.30	0.10	0.40	0.20
2	0.15	0.30	0.20	0.20	0.30	0.20	0.20	0.30	0.30
3	0.20	0.10	0.20	0.40	0.30	0.10	0.30	0.20	0.40
4	0.25	0.20	0.30	0.10	0.40	0.30	0.30	0.30	0.10
5	0.30	0.30	0.20	0.30	0.20	0.20	0.10	0.10	0.60

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.10	0.20	0.10	0.20	0.50	0.30	0.20	0.10	0.40
2	0.15	0.30	0.20	0.10	0.40	0.20	0.30	0.20	0.30
3	0.20	0.10	0.30	0.20	0.40	0.10	0.20	0.30	0.40
4	0.25	0.20	0.20	0.30	0.30	0.30	0.10	0.20	0.40
5	0.30	0.20	0.10	0.30	0.40	0.20	0.30	0.20	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_4 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_4 X}$
1	0.10	0.20	0.30	0.30	0.20	0.30	0.20	0.20	0.30
2	0.15	0.50	0.10	0.10	0.30	0.40	0.10	0.10	0.40
3	0.20	0.30	0.20	0.40	0.10	0.30	0.30	0.30	0.10
4	0.25	0.30	0.10	0.20	0.40	0.30	0.40	0.20	0.10
5	0.30	0.20	0.30	0.30	0.20	0.50	0.10	0.10	0.30

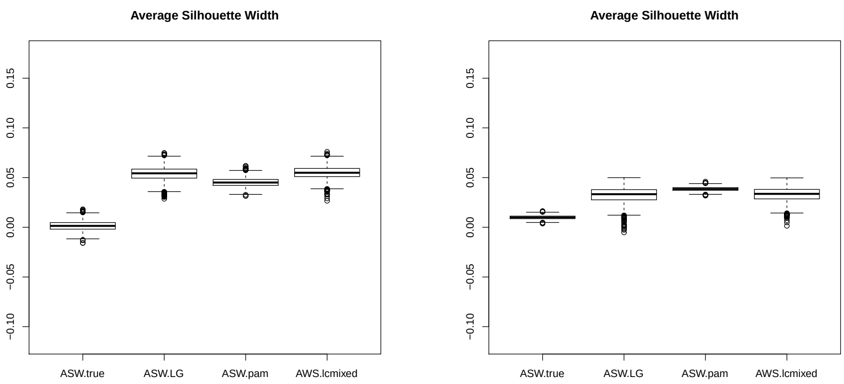
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.10	0.30	0.20	0.10	0.40	0.20	0.20	0.40	0.20
2	0.15	0.20	0.30	0.20	0.30	0.30	0.20	0.20	0.30
3	0.20	0.40	0.10	0.10	0.40	0.20	0.30	0.30	0.20
4	0.25	0.30	0.10	0.10	0.50	0.40	0.10	0.10	0.40
5	0.30	0.30	0.30	0.10	0.30	0.20	0.30	0.20	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.10	0.30	0.10	0.20	0.40	0.10	0.30	0.20	0.40
2	0.15	0.20	0.20	0.30	0.30	0.20	0.10	0.20	0.50
3	0.20	0.40	0.10	0.20	0.30	0.20	0.30	0.20	0.30
4	0.25	0.20	0.30	0.30	0.20	0.30	0.20	0.30	0.20
5	0.30	0.30	0.30	0.20	0.20	0.10	0.40	0.30	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$
1	0.10	0.20	0.40	0.10	0.30	0.30	0.20	0.30	0.20
2	0.15	0.30	0.30	0.10	0.30	0.20	0.20	0.30	0.30
3	0.20	0.40	0.20	0.20	0.20	0.40	0.30	0.20	0.10
4	0.25	0.30	0.40	0.20	0.10	0.30	0.10	0.40	0.20
5	0.30	0.40	0.10	0.20	0.30	0.20	0.30	0.40	0.10

Table A.57: Summary: 12 4-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.006	-0.005	-0.005
1st Qu.		0.048	0.026	0.048
Median		0.068	0.038	0.070
Mean (se)		0.073 (0.001)	0.040 (0.000)	0.074 (0.001)
3rd Qu.		0.093	0.053	0.095
Max.		0.232	0.120	0.241
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.016	0.029	0.032	0.027
1st Qu.	-0.002	0.049	0.042	0.051
Median	0.001	0.054	0.045	0.055
Mean (se)	0.001 (0.000)	0.054 (0.000)	0.045 (0.000)	0.055 (0.000)
3rd Qu.	0.005	0.059	0.048	0.059
Max.	0.018	0.075	0.062	0.076
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.062	0.157	0.177	0.163
1st Qu.	0.112	0.236	0.214	0.239
Median	0.122	0.249	0.223	0.251
Mean (se)	0.123 (0.000)	0.248 (0.000)	0.224 (0.000)	0.250 (0.000)
3rd Qu.	0.134	0.260	0.232	0.262
Max.	0.172	0.309	0.268	0.310

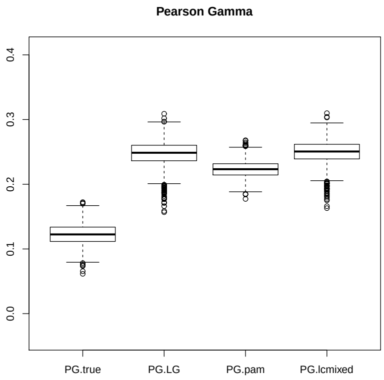


(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

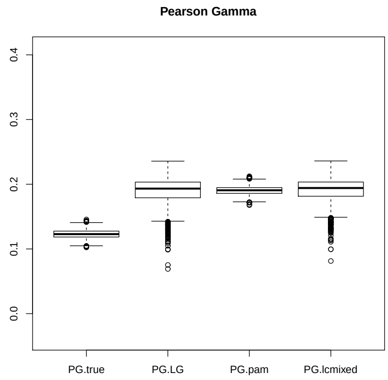
Figure A.86: Average Silhouette Width: 12 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

Table A.58: Summary: 12 4-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.006	0.039
1st Qu.		0.132	0.034	0.137
Median		0.161	0.042	0.164
Mean (se)		0.159 (0.001)	0.043 (0.000)	0.163 (0.001)
3rd Qu.		0.188	0.052	0.190
Max.		0.293	0.102	0.292
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.004	-0.005	0.032	0.001
1st Qu.	0.009	0.028	0.037	0.029
Median	0.010	0.033	0.039	0.034
Mean (se)	0.010 (0.000)	0.032 (0.000)	0.039 (0.000)	0.033 (0.000)
3rd Qu.	0.011	0.038	0.040	0.038
Max.	0.016	0.050	0.046	0.050
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.102	0.069	0.168	0.081
1st Qu.	0.119	0.179	0.186	0.182
Median	0.123	0.193	0.191	0.194
Mean (se)	0.123 (0.000)	0.189 (0.000)	0.190 (0.000)	0.191 (0.000)
3rd Qu.	0.128	0.203	0.195	0.203
Max.	0.146	0.236	0.212	0.236
NA's		1		

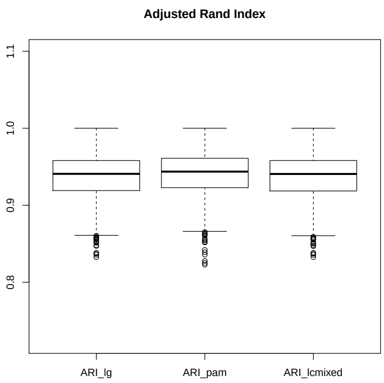


(a) *PG - 200 units.*

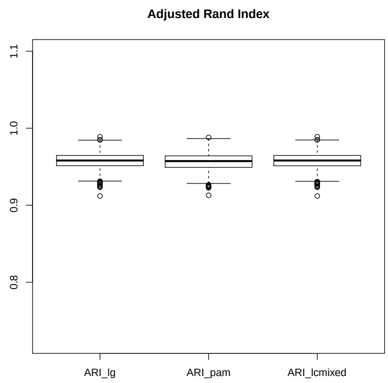


(b) *PG - 1000 units.*

Figure A.87: Pearson Gamma: 12 4-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *ARI - 200 units.*



(b) *ARI - 1000 units.*

Figure A.88: Adjusted Rand Index: 12 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$
1	0.10	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10
2	0.15	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
3	0.20	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
4	0.25	0.10	0.10	0.70	0.10	0.10	0.10	0.10	0.70
5	0.30	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.10	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
2	0.15	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
3	0.20	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
4	0.25	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
5	0.30	0.10	0.10	0.10	0.70	0.10	0.10	0.70	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1X}$	$\hat{\pi}_{1t}^{\bar{E}_2X}$	$\hat{\pi}_{1t}^{\bar{E}_3X}$	$\hat{\pi}_{1t}^{\bar{E}_4X}$	$\hat{\pi}_{1t}^{\bar{F}_1X}$	$\hat{\pi}_{1t}^{\bar{F}_2X}$	$\hat{\pi}_{1t}^{\bar{F}_3X}$	$\hat{\pi}_{1t}^{\bar{F}_4X}$
1	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
2	0.15	0.10	0.70	0.10	0.10	0.70	0.10	0.10	0.10
3	0.20	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
4	0.25	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
5	0.30	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10

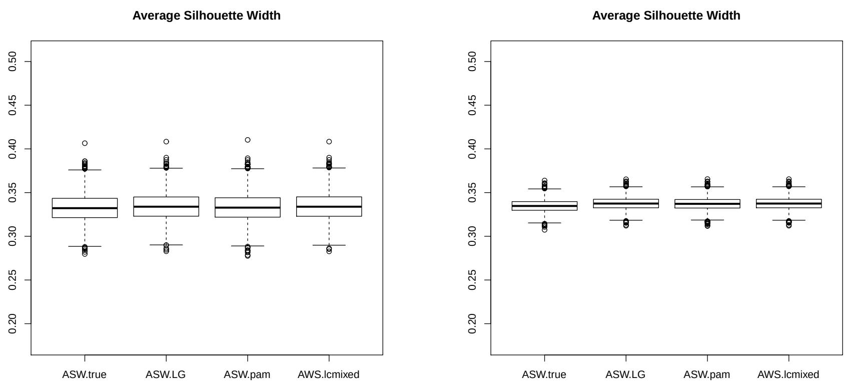
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1X}$	$\hat{\pi}_{1t}^{\bar{G}_2X}$	$\hat{\pi}_{1t}^{\bar{G}_3X}$	$\hat{\pi}_{1t}^{\bar{G}_4X}$	$\hat{\pi}_{1t}^{\bar{H}_1X}$	$\hat{\pi}_{1t}^{\bar{H}_2X}$	$\hat{\pi}_{1t}^{\bar{H}_3X}$	$\hat{\pi}_{1t}^{\bar{H}_4X}$
1	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10
2	0.15	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10
3	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70
4	0.25	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
5	0.30	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1X}$	$\hat{\pi}_{1t}^{\bar{I}_2X}$	$\hat{\pi}_{1t}^{\bar{I}_3X}$	$\hat{\pi}_{1t}^{\bar{I}_4X}$	$\hat{\pi}_{1t}^{\bar{L}_1X}$	$\hat{\pi}_{1t}^{\bar{L}_2X}$	$\hat{\pi}_{1t}^{\bar{L}_3X}$	$\hat{\pi}_{1t}^{\bar{L}_4X}$
1	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
2	0.15	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
3	0.20	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10
4	0.25	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10
5	0.30	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1X}$	$\hat{\pi}_{1t}^{\bar{M}_2X}$	$\hat{\pi}_{1t}^{\bar{M}_3X}$	$\hat{\pi}_{1t}^{\bar{M}_4X}$	$\hat{\pi}_{1t}^{\bar{N}_1X}$	$\hat{\pi}_{1t}^{\bar{N}_2X}$	$\hat{\pi}_{1t}^{\bar{N}_3X}$	$\hat{\pi}_{1t}^{\bar{N}_4X}$
1	0.10	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
2	0.15	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
3	0.20	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10
4	0.25	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
5	0.30	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10

Table A.59: Summary: 12 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.833	0.823	0.833
1st Qu.		0.919	0.923	0.919
Median		0.941	0.944	0.941
Mean (se)		0.938 (0.001)	0.941 (0.001)	0.938 (0.001)
3rd Qu.		0.958	0.961	0.958
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.280	0.283	0.278	0.283
1st Qu.	0.321	0.323	0.322	0.323
Median	0.332	0.334	0.333	0.334
Mean (se)	0.332 (0.000)	0.334 (0.000)	0.333 (0.000)	0.334 (0.000)
3rd Qu.	0.343	0.345	0.344	0.345
Max.	0.406	0.408	0.410	0.408
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.676	0.677	0.667	0.677
1st Qu.	0.722	0.724	0.724	0.724
Median	0.734	0.736	0.736	0.736
Mean (se)	0.734 (0.000)	0.736 (0.000)	0.736 (0.000)	0.736 (0.000)
3rd Qu.	0.746	0.748	0.747	0.748
Max.	0.795	0.796	0.798	0.796

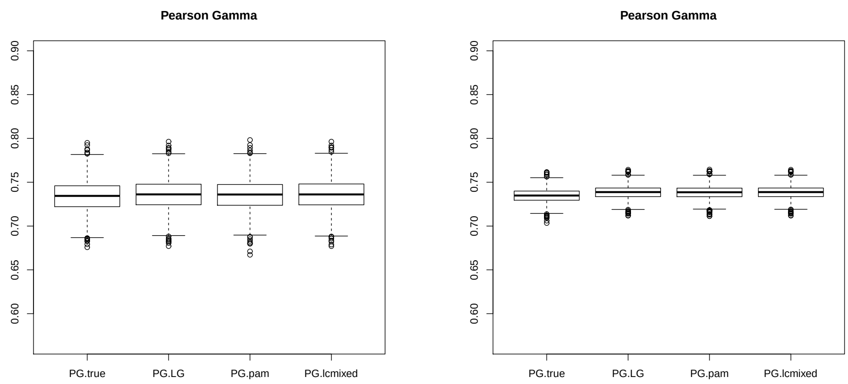


(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.89: Average Silhouette Width: 12 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

Table A.60: Summary: 12 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.912	0.913	0.912
1st Qu.		0.951	0.949	0.951
Median		0.958	0.957	0.958
Mean (se)		0.958 (0.000)	0.957 (0.000)	0.958 (0.000)
3rd Qu.		0.965	0.964	0.965
Max.		0.989	0.988	0.989
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.307	0.312	0.312	0.312
1st Qu.	0.330	0.333	0.332	0.333
Median	0.335	0.338	0.337	0.338
Mean (se)	0.335 (0.000)	0.338 (0.000)	0.337 (0.000)	0.338 (0.000)
3rd Qu.	0.340	0.342	0.342	0.342
Max.	0.364	0.365	0.366	0.365
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.703	0.712	0.711	0.712
1st Qu.	0.730	0.734	0.734	0.734
Median	0.735	0.739	0.738	0.739
Mean (se)	0.735 (0.000)	0.739 (0.000)	0.738 (0.000)	0.739 (0.000)
3rd Qu.	0.740	0.743	0.743	0.743
Max.	0.762	0.764	0.764	0.764



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.90: Pearson Gamma: 12 4-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

A.31 12 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$
1	0.20	0.20	0.40	0.10	0.30	0.40	0.10	0.10	0.40
2	0.20	0.30	0.30	0.20	0.20	0.30	0.10	0.20	0.40
3	0.20	0.40	0.30	0.10	0.20	0.30	0.20	0.20	0.30
4	0.20	0.30	0.20	0.30	0.20	0.20	0.30	0.20	0.30
5	0.20	0.30	0.10	0.10	0.50	0.20	0.40	0.20	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.20	0.30	0.20	0.20	0.30	0.10	0.30	0.30	0.30
2	0.20	0.20	0.30	0.30	0.20	0.10	0.40	0.20	0.30
3	0.20	0.40	0.20	0.20	0.20	0.20	0.20	0.20	0.40
4	0.20	0.50	0.10	0.20	0.20	0.20	0.30	0.30	0.20
5	0.20	0.30	0.40	0.20	0.10	0.10	0.40	0.20	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1X}$	$\hat{\pi}_{1t}^{\bar{E}_2X}$	$\hat{\pi}_{1t}^{\bar{E}_3X}$	$\hat{\pi}_{1t}^{\bar{E}_4X}$	$\hat{\pi}_{1t}^{\bar{F}_1X}$	$\hat{\pi}_{1t}^{\bar{F}_2X}$	$\hat{\pi}_{1t}^{\bar{F}_3X}$	$\hat{\pi}_{1t}^{\bar{F}_4X}$
1	0.20	0.20	0.20	0.50	0.10	0.20	0.30	0.40	0.10
2	0.20	0.30	0.20	0.40	0.10	0.10	0.50	0.20	0.20
3	0.20	0.30	0.30	0.20	0.20	0.30	0.20	0.30	0.20
4	0.20	0.10	0.40	0.30	0.20	0.20	0.40	0.10	0.30
5	0.20	0.10	0.20	0.30	0.40	0.30	0.20	0.20	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1X}$	$\hat{\pi}_{1t}^{\bar{G}_2X}$	$\hat{\pi}_{1t}^{\bar{G}_3X}$	$\hat{\pi}_{1t}^{\bar{G}_4X}$	$\hat{\pi}_{1t}^{\bar{H}_1X}$	$\hat{\pi}_{1t}^{\bar{H}_2X}$	$\hat{\pi}_{1t}^{\bar{H}_3X}$	$\hat{\pi}_{1t}^{\bar{H}_4X}$
1	0.20	0.20	0.30	0.20	0.30	0.30	0.20	0.40	0.10
2	0.20	0.30	0.30	0.20	0.20	0.20	0.10	0.50	0.20
3	0.20	0.30	0.20	0.40	0.10	0.30	0.30	0.20	0.20
4	0.20	0.10	0.40	0.20	0.30	0.40	0.20	0.30	0.10
5	0.20	0.20	0.30	0.10	0.40	0.40	0.20	0.10	0.30

Table A.61: Summary: 12 4-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 200 units

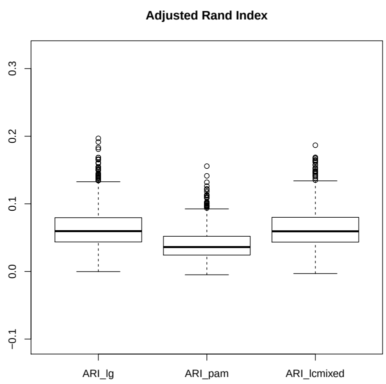
		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	-0.005	-0.003
1st Qu.		0.044	0.024	0.043
Median		0.060	0.036	0.059
Mean (se)		0.063 (0.001)	0.039 (0.000)	0.063 (0.001)
3rd Qu.		0.079	0.052	0.080
Max.		0.197	0.156	0.187
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.015	0.024	0.031	0.028
1st Qu.	-0.004	0.049	0.042	0.051
Median	0.000	0.054	0.045	0.055
Mean (se)	0.000 (0.000)	0.054 (0.000)	0.046 (0.000)	0.055 (0.000)
3rd Qu.	0.003	0.059	0.049	0.060
Max.	0.016	0.075	0.063	0.080
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.063	0.136	0.181	0.156
1st Qu.	0.101	0.236	0.216	0.240
Median	0.110	0.250	0.225	0.252
Mean (se)	0.111 (0.000)	0.248 (0.000)	0.225 (0.000)	0.251 (0.000)
3rd Qu.	0.121	0.262	0.234	0.264
Max.	0.161	0.309	0.272	0.319

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1X}$	$\hat{\pi}_{1t}^{\bar{I}_2X}$	$\hat{\pi}_{1t}^{\bar{I}_3X}$	$\hat{\pi}_{1t}^{\bar{I}_4X}$	$\hat{\pi}_{1t}^{\bar{L}_1X}$	$\hat{\pi}_{1t}^{\bar{L}_2X}$	$\hat{\pi}_{1t}^{\bar{L}_3X}$	$\hat{\pi}_{1t}^{\bar{L}_4X}$
1	0.20	0.10	0.20	0.40	0.30	0.20	0.20	0.10	0.50
2	0.20	0.10	0.30	0.30	0.30	0.10	0.10	0.20	0.60
3	0.20	0.20	0.20	0.20	0.40	0.20	0.20	0.30	0.30
4	0.20	0.20	0.10	0.50	0.20	0.30	0.10	0.20	0.40
5	0.20	0.20	0.20	0.30	0.30	0.30	0.20	0.20	0.30

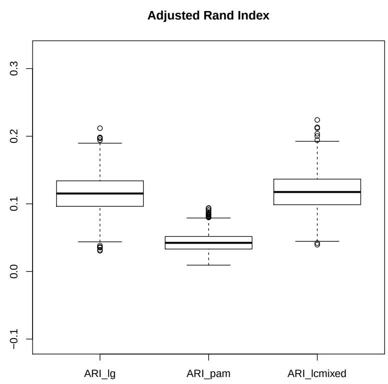
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1X}$	$\hat{\pi}_{1t}^{\bar{M}_2X}$	$\hat{\pi}_{1t}^{\bar{M}_3X}$	$\hat{\pi}_{1t}^{\bar{M}_4X}$	$\hat{\pi}_{1t}^{\bar{N}_1X}$	$\hat{\pi}_{1t}^{\bar{N}_2X}$	$\hat{\pi}_{1t}^{\bar{N}_3X}$	$\hat{\pi}_{1t}^{\bar{N}_4X}$
1	0.20	0.30	0.10	0.20	0.40	0.50	0.10	0.10	0.30
2	0.20	0.30	0.20	0.10	0.40	0.40	0.20	0.10	0.30
3	0.20	0.40	0.10	0.30	0.20	0.50	0.10	0.20	0.20
4	0.20	0.30	0.20	0.30	0.20	0.40	0.10	0.10	0.40
5	0.20	0.20	0.20	0.40	0.20	0.40	0.40	0.10	0.10

Table A.62: Summary: 12 4-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.031	0.009	0.040
1st Qu.		0.096	0.033	0.099
Median		0.115	0.042	0.118
Mean (se)		0.115 (0.001)	0.043 (0.000)	0.118 (0.001)
3rd Qu.		0.134	0.052	0.137
Max.		0.212	0.094	0.224
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.000	-0.001	0.033	0.001
1st Qu.	0.005	0.026	0.038	0.027
Median	0.006	0.032	0.039	0.033
Mean (se)	0.006 (0.000)	0.031 (0.000)	0.039 (0.000)	0.032 (0.000)
3rd Qu.	0.007	0.038	0.041	0.038
Max.	0.013	0.052	0.046	0.052
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.088	0.073	0.171	0.092
1st Qu.	0.106	0.179	0.189	0.182
Median	0.110	0.195	0.193	0.197
Mean (se)	0.110 (0.001)	0.191 (0.000)	0.194 (0.000)	0.194 (0.000)
3rd Qu.	0.114	0.207	0.198	0.208
Max.	0.132	0.246	0.226	0.247

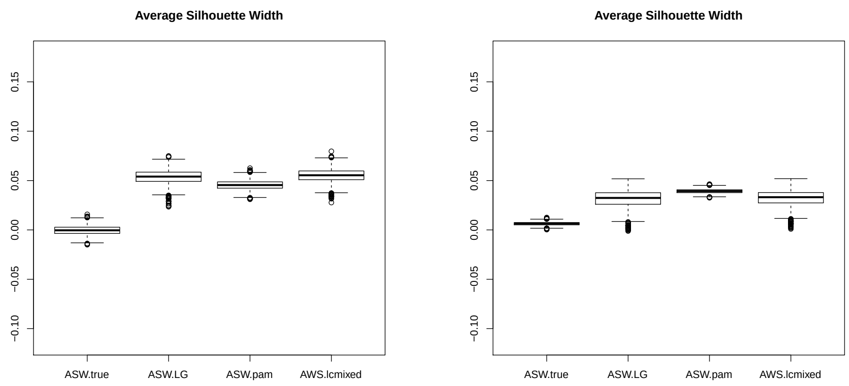


(a) *ARI - 200 units.*



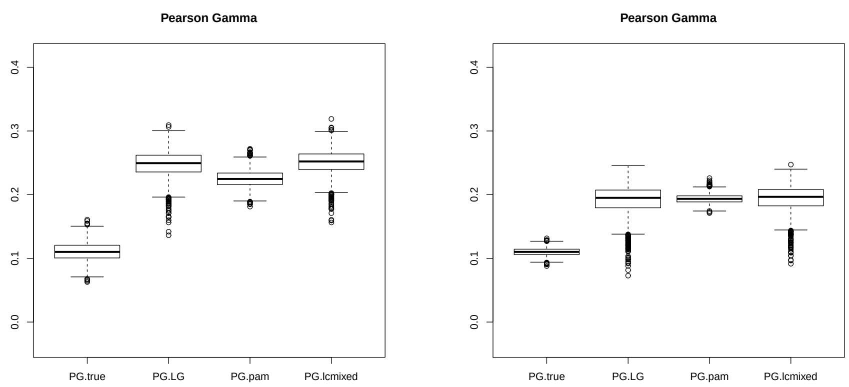
(b) *ARI - 1000 units.*

Figure A.91: Adjusted Rand Index: 12 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.92: Average Silhouette Width: 12 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.93: Pearson Gamma: 12 4-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

A.32 12 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

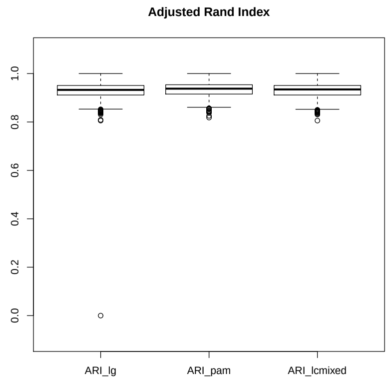
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$
1	0.20	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
2	0.20	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
3	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
4	0.20	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
5	0.20	0.10	0.10	0.10	0.70	0.10	0.10	0.10	0.70

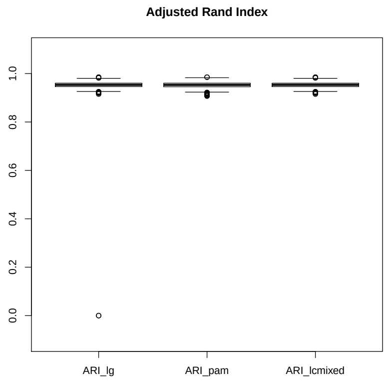
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.20	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10
2	0.20	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10
3	0.20	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
4	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70
5	0.20	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{E}_1X}$	$\hat{\pi}_{1t}^{\bar{E}_2X}$	$\hat{\pi}_{1t}^{\bar{E}_3X}$	$\hat{\pi}_{1t}^{\bar{E}_4X}$	$\hat{\pi}_{1t}^{\bar{F}_1X}$	$\hat{\pi}_{1t}^{\bar{F}_2X}$	$\hat{\pi}_{1t}^{\bar{F}_3X}$	$\hat{\pi}_{1t}^{\bar{F}_4X}$
1	0.20	0.10	0.70	0.10	0.10	0.10	0.70	0.10	0.10
2	0.20	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10
3	0.20	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10
4	0.20	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
5	0.20	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1X}$	$\hat{\pi}_{1t}^{\bar{G}_2X}$	$\hat{\pi}_{1t}^{\bar{G}_3X}$	$\hat{\pi}_{1t}^{\bar{G}_4X}$	$\hat{\pi}_{1t}^{\bar{H}_1X}$	$\hat{\pi}_{1t}^{\bar{H}_2X}$	$\hat{\pi}_{1t}^{\bar{H}_3X}$	$\hat{\pi}_{1t}^{\bar{H}_4X}$
1	0.20	0.10	0.70	0.10	0.10	0.10	0.70	0.10	0.10
2	0.20	0.10	0.10	0.10	0.70	0.10	0.10	0.10	0.70
3	0.20	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10
4	0.20	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10
5	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10



(a) *ARI - 200 units.*



(b) *ARI - 1000 units.*

Figure A.94: Adjusted Rand Index: 12 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.20	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10
2	0.20	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
3	0.20	0.10	0.70	0.10	0.10	0.70	0.10	0.10	0.10
4	0.20	0.10	0.10	0.10	0.70	0.10	0.10	0.10	0.70
5	0.20	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$
1	0.20	0.10	0.10	0.70	0.10	0.10	0.10	0.70	0.10
2	0.20	0.10	0.10	0.70	0.10	0.10	0.10	0.10	0.70
3	0.20	0.70	0.10	0.10	0.10	0.10	0.70	0.10	0.10
4	0.20	0.10	0.70	0.10	0.10	0.70	0.10	0.10	0.10
5	0.20	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10

A.33 4 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

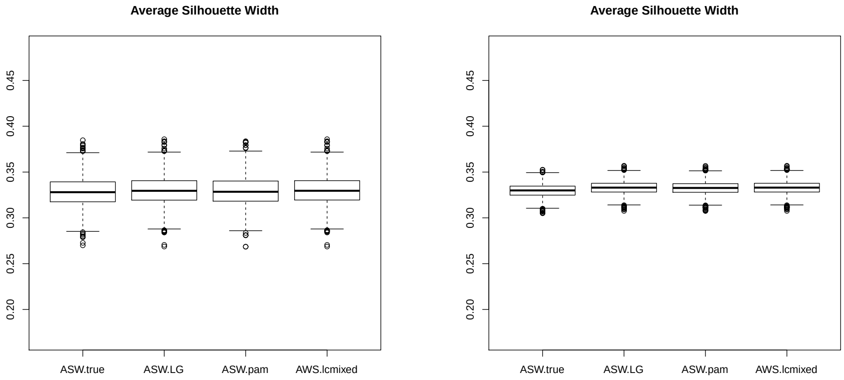
The Model:

Table A.63: Summary: 12 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.818	0.806
1st Qu.		0.911	0.915	0.912
Median		0.933	0.938	0.934
Mean (se)		0.929 (0.001)	0.935 (0.001)	0.930 (0.001)
3rd Qu.		0.951	0.954	0.951
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.270	0.269	0.268	0.269
1st Qu.	0.318	0.319	0.318	0.320
Median	0.328	0.330	0.329	0.330
Mean (se)	0.328 (0.000)	0.330 (0.000)	0.329 (0.000)	0.330 (0.000)
3rd Qu.	0.339	0.341	0.340	0.341
Max.	0.385	0.386	0.384	0.386
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.660	0.654	0.653	0.654
1st Qu.	0.709	0.711	0.710	0.711
Median	0.720	0.721	0.721	0.721
Mean (se)	0.719 (0.000)	0.721 (0.000)	0.720 (0.000)	0.721 (0.000)
3rd Qu.	0.730	0.732	0.732	0.732
Max.	0.772	0.774	0.774	0.774
NA's		1		

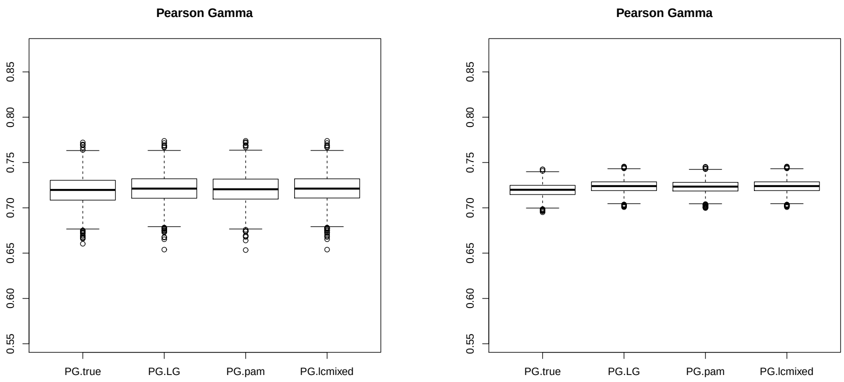
Table A.64: Summary: 12 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.907	0.915
1st Qu.		0.946	0.945	0.946
Median		0.953	0.953	0.953
Mean (se)		0.952 (0.001)	0.952 (0.000)	0.953 (0.000)
3rd Qu.		0.960	0.960	0.960
Max.		0.985	0.985	0.985
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.305	0.308	0.308	0.308
1st Qu.	0.325	0.328	0.328	0.328
Median	0.330	0.333	0.333	0.333
Mean (se)	0.330 (0.000)	0.333 (0.000)	0.333 (0.000)	0.333 (0.000)
3rd Qu.	0.335	0.338	0.337	0.338
Max.	0.353	0.357	0.357	0.357
NA's		2		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.695	0.701	0.700	0.701
1st Qu.	0.715	0.719	0.719	0.719
Median	0.720	0.724	0.723	0.724
Mean (se)	0.720 (0.000)	0.724 (0.000)	0.723 (0.000)	0.724 (0.000)
3rd Qu.	0.725	0.729	0.728	0.729
Max.	0.742	0.746	0.745	0.746
NA's		2		



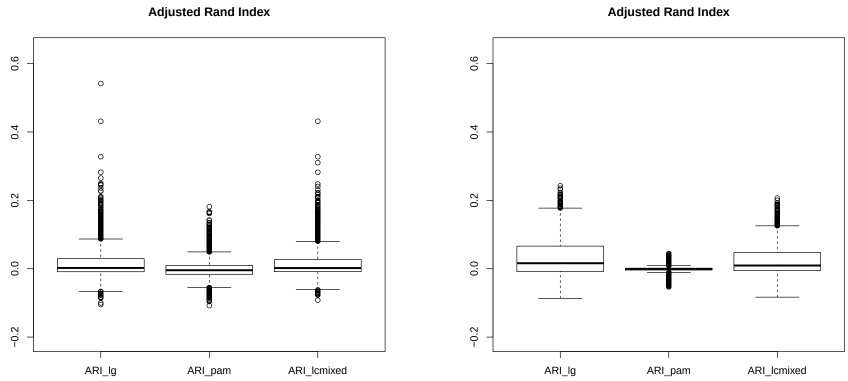
(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.95: Average Silhouette Width: 12 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.96: Pearson Gamma: 12 4-level variables - 5 clusters, equal mixing proportions - Clear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.97: Adjusted Rand Index: 4 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{A}_5X}$	$\hat{\pi}_{1t}^{\bar{A}_6X}$	$\hat{\pi}_{1t}^{\bar{A}_7X}$	$\hat{\pi}_{1t}^{\bar{A}_8X}$
1	0.85	0.05	0.15	0.20	0.40	0.05	0.05	0.05	0.05
2	0.15	0.05	0.15	0.15	0.30	0.10	0.05	0.05	0.15

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_5X}$	$\hat{\pi}_{1t}^{\bar{B}_6X}$	$\hat{\pi}_{1t}^{\bar{B}_7X}$	$\hat{\pi}_{1t}^{\bar{B}_8X}$
1	0.85	0.05	0.10	0.05	0.20	0.10	0.15	0.05	0.30
2	0.15	0.05	0.05	0.10	0.15	0.15	0.20	0.05	0.25

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{C}_5X}$	$\hat{\pi}_{1t}^{\bar{C}_6X}$	$\hat{\pi}_{1t}^{\bar{C}_7X}$	$\hat{\pi}_{1t}^{\bar{C}_8X}$
1	0.85	0.05	0.20	0.10	0.05	0.10	0.15	0.05	0.30
2	0.15	0.20	0.15	0.05	0.10	0.05	0.10	0.10	0.25

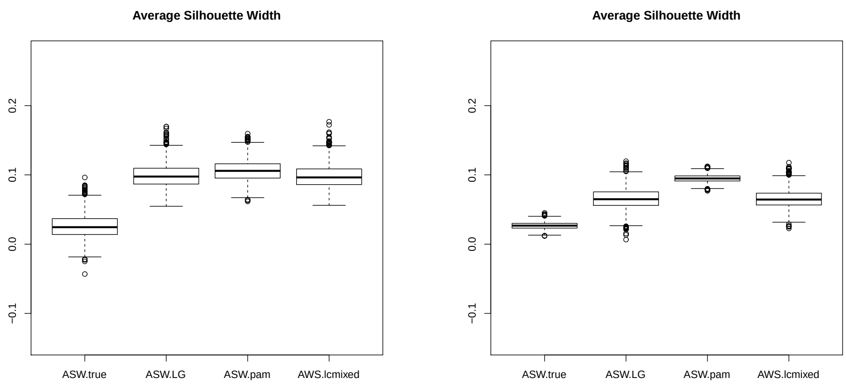
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_5X}$	$\hat{\pi}_{1t}^{\bar{D}_6X}$	$\hat{\pi}_{1t}^{\bar{D}_7X}$	$\hat{\pi}_{1t}^{\bar{D}_8X}$
1	0.85	0.10	0.20	0.05	0.15	0.05	0.05	0.20	0.20
2	0.15	0.15	0.15	0.10	0.10	0.10	0.05	0.10	0.25

A.34 4 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Table A.65: Summary: 4 8-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.105	-0.108	-0.092
1st Qu.		-0.009	-0.016	-0.008
Median		0.002	-0.004	0.002
Mean (se)		0.016 (0.001)	-0.001 (0.001)	0.016 (0.001)
3rd Qu.		0.030	0.010	0.027
Max.		0.542	0.182	0.431
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.043	0.055	0.062	0.056
1st Qu.	0.014	0.087	0.095	0.086
Median	0.025	0.098	0.106	0.096
Mean (se)	0.026 (0.000)	0.099 (0.000)	0.106 (0.000)	0.098 (0.000)
3rd Qu.	0.037	0.110	0.116	0.109
Max.	0.096	0.170	0.160	0.177
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.069	0.086	0.088	0.097
1st Qu.	0.032	0.191	0.203	0.189
Median	0.057	0.218	0.229	0.217
Mean (se)	0.058 (0.001)	0.221 (0.001)	0.229 (0.001)	0.220 (0.001)
3rd Qu.	0.083	0.249	0.256	0.246
Max.	0.196	0.392	0.376	0.425

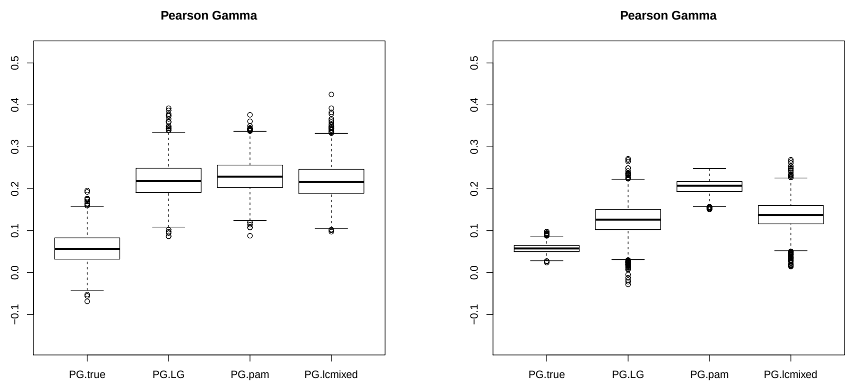


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.98: Average Silhouette Width: 4 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

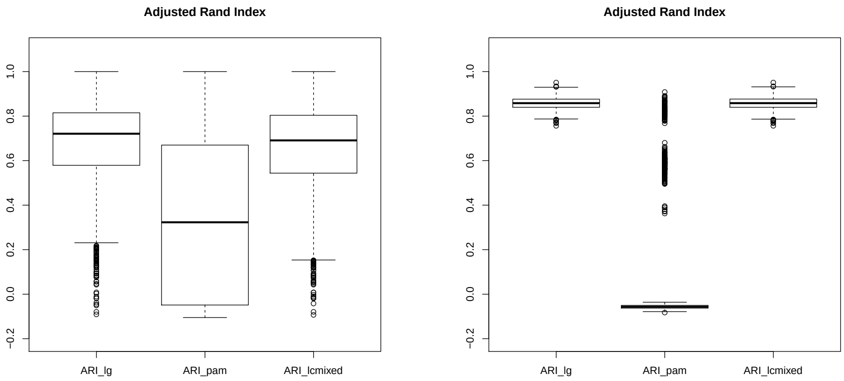
Table A.66: Summary: 4 8-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.087	-0.053	-0.083
1st Qu.		-0.008	-0.004	-0.005
Median		0.016	-0.001	0.009
Mean (se)		0.032 (0.001)	-0.003 (0.000)	0.024 (0.001)
3rd Qu.		0.066	0.001	0.047
Max.		0.242	0.044	0.207
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.012	0.007	0.077	0.023
1st Qu.	0.023	0.056	0.091	0.057
Median	0.027	0.065	0.095	0.064
Mean (se)	0.027 (0.000)	0.066 (0.000)	0.095 (0.000)	0.066 (0.000)
3rd Qu.	0.030	0.075	0.099	0.074
Max.	0.045	0.120	0.112	0.118
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.024	-0.028	0.151	0.014
1st Qu.	0.050	0.102	0.193	0.116
Median	0.058	0.126	0.207	0.137
Mean (se)	0.058 (0.000)	0.125 (0.001)	0.204 (0.000)	0.138 (0.001)
3rd Qu.	0.065	0.151	0.217	0.160
Max.	0.098	0.271	0.248	0.269



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.99: Pearson Gamma: 4 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.100: Adjusted Rand Index: 4 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{A}_5X}$	$\hat{\pi}_{1t}^{\bar{A}_6X}$	$\hat{\pi}_{1t}^{\bar{A}_7X}$	$\hat{\pi}_{1t}^{\bar{A}_8X}$
1	0.85	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_5X}$	$\hat{\pi}_{1t}^{\bar{B}_6X}$	$\hat{\pi}_{1t}^{\bar{B}_7X}$	$\hat{\pi}_{1t}^{\bar{B}_8X}$
1	0.85	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{C}_5X}$	$\hat{\pi}_{1t}^{\bar{C}_6X}$	$\hat{\pi}_{1t}^{\bar{C}_7X}$	$\hat{\pi}_{1t}^{\bar{C}_8X}$
1	0.85	0.05	0.05	0.35	0.35	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.35	0.35

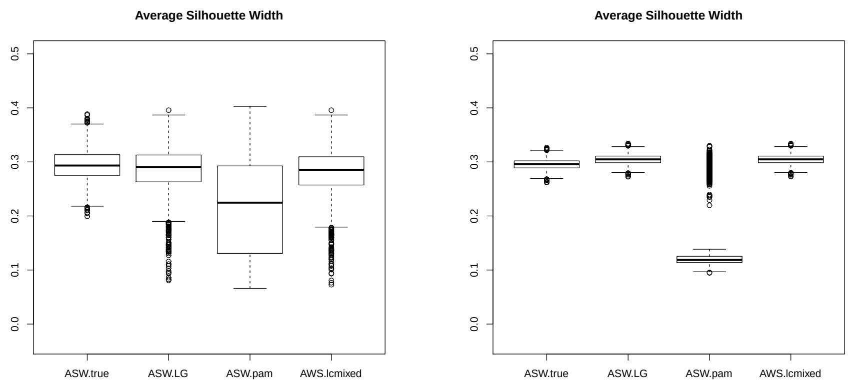
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_5X}$	$\hat{\pi}_{1t}^{\bar{D}_6X}$	$\hat{\pi}_{1t}^{\bar{D}_7X}$	$\hat{\pi}_{1t}^{\bar{D}_8X}$
1	0.85	0.05	0.35	0.05	0.05	0.35	0.05	0.05	0.05
2	0.15	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05

A.35 4 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Table A.67: Summary: 4 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.091	-0.105	-0.093
1st Qu.		0.579	-0.049	0.544
Median		0.721	0.323	0.691
Mean (se)		0.680 (0.004)	0.322 (0.008)	0.656 (0.004)
3rd Qu.		0.815	0.669	0.804
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.199	0.081	0.066	0.073
1st Qu.	0.275	0.263	0.131	0.257
Median	0.293	0.291	0.225	0.286
Mean (se)	0.294 (0.001)	0.284 (0.001)	0.213 (0.002)	0.280 (0.001)
3rd Qu.	0.313	0.313	0.293	0.310
Max.	0.388	0.396	0.403	0.396
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.284	0.076	0.038	0.060
1st Qu.	0.446	0.471	0.164	0.469
Median	0.484	0.510	0.441	0.509
Mean (se)	0.481 (0.001)	0.503 (0.002)	0.358 (0.004)	0.501 (0.002)
3rd Qu.	0.518	0.547	0.543	0.547
Max.	0.655	0.680	0.691	0.680

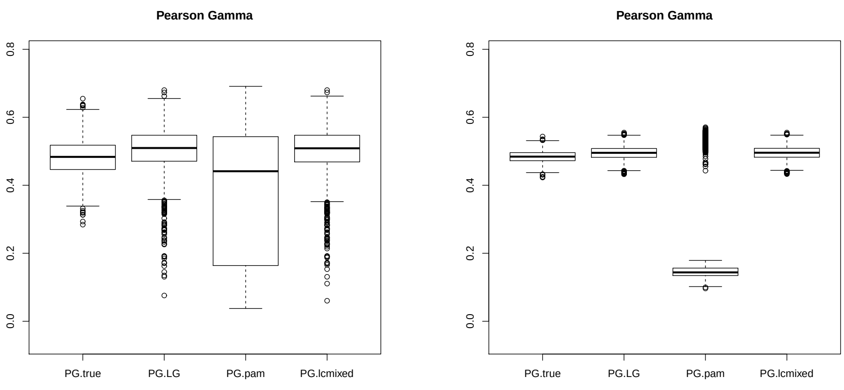


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.101: Average Silhouette Width: 4 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

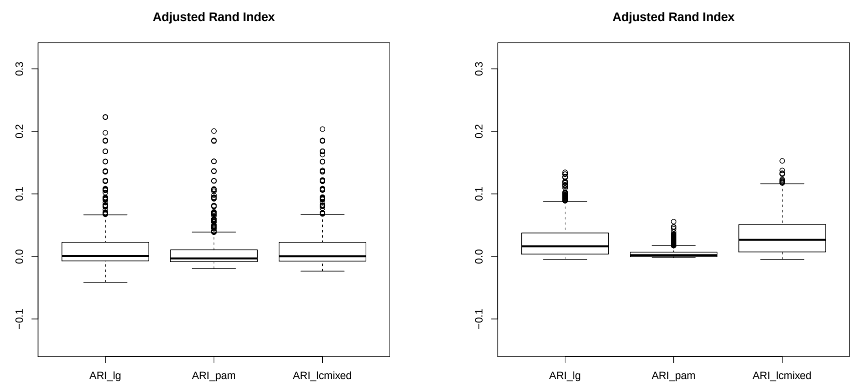
Table A.68: Summary: 4 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.756	-0.082	0.756
1st Qu.		0.840	-0.062	0.840
Median		0.858	-0.057	0.858
Mean (se)		0.858 (0.001)	0.061 (0.006)	0.858 (0.001)
3rd Qu.		0.876	-0.050	0.877
Max.		0.951	0.908	0.951
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.262	0.272	0.095	0.273
1st Qu.	0.289	0.298	0.114	0.298
Median	0.296	0.305	0.119	0.305
Mean (se)	0.295 (0.000)	0.305 (0.000)	0.145 (0.001)	0.305 (0.000)
3rd Qu.	0.302	0.311	0.125	0.311
Max.	0.326	0.334	0.330	0.334
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.423	0.432	0.096	0.433
1st Qu.	0.472	0.482	0.134	0.483
Median	0.484	0.496	0.144	0.496
Mean (se)	0.484 (0.000)	0.495 (0.000)	0.206 (0.003)	0.495 (0.000)
3rd Qu.	0.496	0.508	0.156	0.509
Max.	0.543	0.555	0.571	0.555



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.102: Pearson Gamma: 4 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.103: Adjusted Rand Index: 4 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{A}_3 X}$	$\hat{\pi}_{1t}^{\bar{A}_4 X}$	$\hat{\pi}_{1t}^{\bar{A}_5 X}$	$\hat{\pi}_{1t}^{\bar{A}_6 X}$	$\hat{\pi}_{1t}^{\bar{A}_7 X}$	$\hat{\pi}_{1t}^{\bar{A}_8 X}$
1	0.50	0.10	0.05	0.15	0.10	0.20	0.10	0.15	0.15
2	0.50	0.15	0.10	0.10	0.05	0.15	0.15	0.10	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{B}_4 X}$	$\hat{\pi}_{1t}^{\bar{B}_5 X}$	$\hat{\pi}_{1t}^{\bar{B}_6 X}$	$\hat{\pi}_{1t}^{\bar{B}_7 X}$	$\hat{\pi}_{1t}^{\bar{B}_8 X}$
1	0.50	0.15	0.15	0.10	0.05	0.05	0.10	0.20	0.20
2	0.50	0.10	0.20	0.15	0.10	0.10	0.05	0.15	0.15

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_4 X}$	$\hat{\pi}_{1t}^{\bar{C}_5 X}$	$\hat{\pi}_{1t}^{\bar{C}_6 X}$	$\hat{\pi}_{1t}^{\bar{C}_7 X}$	$\hat{\pi}_{1t}^{\bar{C}_8 X}$
1	0.50	0.20	0.10	0.15	0.10	0.10	0.15	0.15	0.05
2	0.50	0.15	0.15	0.10	0.20	0.05	0.05	0.20	0.10

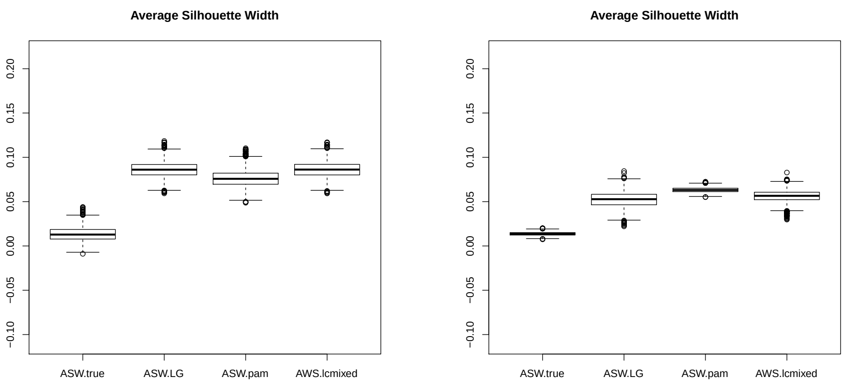
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$	$\hat{\pi}_{1t}^{\bar{D}_5 X}$	$\hat{\pi}_{1t}^{\bar{D}_6 X}$	$\hat{\pi}_{1t}^{\bar{D}_7 X}$	$\hat{\pi}_{1t}^{\bar{D}_8 X}$
1	0.50	0.10	0.15	0.05	0.20	0.15	0.15	0.10	0.10
2	0.50	0.05	0.10	0.10	0.15	0.10	0.20	0.20	0.10

A.36 4 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

The Model:

Table A.69: Summary: 4 8-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.041	-0.019	-0.024
1st Qu.		-0.007	-0.008	-0.008
Median		0.001	-0.003	0.000
Mean (se)		0.013 (0.001)	0.006 (0.001)	0.013 (0.001)
3rd Qu.		0.023	0.011	0.023
Max.		0.223	0.200	0.204
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.009	0.059	0.049	0.059
1st Qu.	0.008	0.080	0.070	0.080
Median	0.013	0.086	0.076	0.086
Mean (se)	0.013 (0.000)	0.086 (0.000)	0.076 (0.000)	0.086 (0.000)
3rd Qu.	0.019	0.092	0.082	0.092
Max.	0.044	0.118	0.110	0.117
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.025	0.113	0.111	0.122
1st Qu.	0.020	0.202	0.166	0.203
Median	0.034	0.221	0.183	0.221
Mean (se)	0.035 (0.000)	0.220 (0.001)	0.184 (0.001)	0.221 (0.001)
3rd Qu.	0.049	0.238	0.200	0.239
Max.	0.113	0.308	0.272	0.308

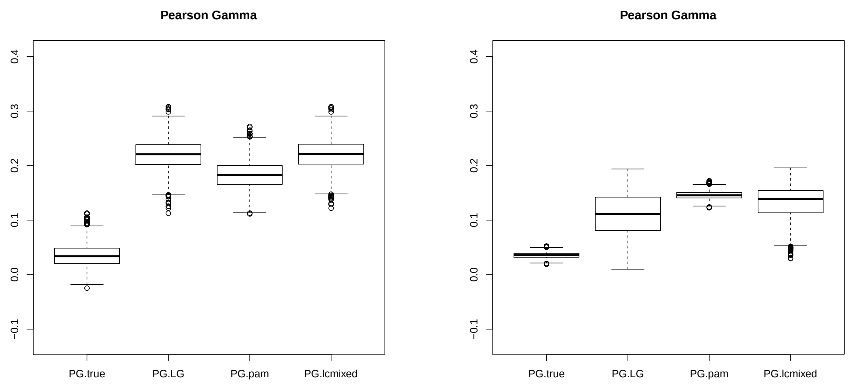


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.104: Average Silhouette Width: 4 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

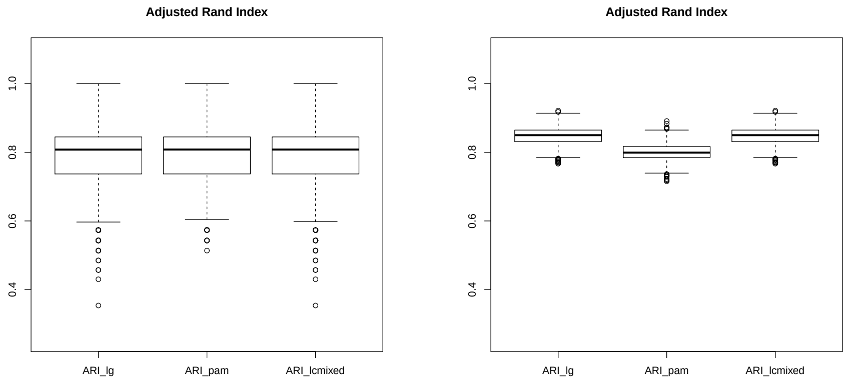
Table A.70: Summary: 4 8-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.005	-0.002	-0.005
1st Qu.		0.004	0.000	0.007
Median		0.016	0.002	0.027
Mean (se)		0.024 (0.001)	0.005 (0.000)	0.032 (0.001)
3rd Qu.		0.038	0.007	0.051
Max.		0.135	0.056	0.153
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.007	0.022	0.055	0.030
1st Qu.	0.012	0.046	0.061	0.052
Median	0.014	0.053	0.063	0.057
Mean (se)	0.014 (0.000)	0.052 (0.000)	0.063 (0.000)	0.056 (0.000)
3rd Qu.	0.015	0.058	0.065	0.061
Max.	0.020	0.084	0.072	0.083
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.019	0.010	0.123	0.029
1st Qu.	0.032	0.081	0.141	0.113
Median	0.036	0.111	0.146	0.139
Mean (se)	0.036 (0.000)	0.110 (0.001)	0.146 (0.000)	0.132 (0.001)
3rd Qu.	0.039	0.142	0.151	0.154
Max.	0.052	0.194	0.172	0.196



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.105: Pearson Gamma: 4 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.106: Adjusted Rand Index: 4 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{A}_3X}$	$\hat{\pi}_{1t}^{\bar{A}_4X}$	$\hat{\pi}_{1t}^{\bar{A}_5X}$	$\hat{\pi}_{1t}^{\bar{A}_6X}$	$\hat{\pi}_{1t}^{\bar{A}_7X}$	$\hat{\pi}_{1t}^{\bar{A}_8X}$
1	0.50	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.65	0.65	0.05	0.05	0.05

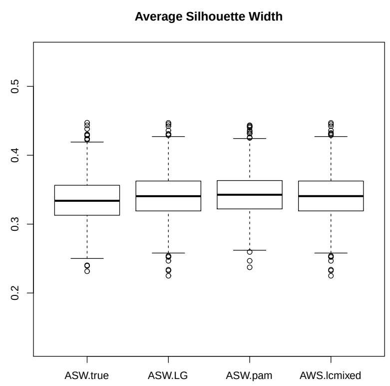
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{B}_4X}$	$\hat{\pi}_{1t}^{\bar{B}_5X}$	$\hat{\pi}_{1t}^{\bar{B}_6X}$	$\hat{\pi}_{1t}^{\bar{B}_7X}$	$\hat{\pi}_{1t}^{\bar{B}_8X}$
1	0.50	0.05	0.35	0.05	0.05	0.05	0.05	0.05	0.35
2	0.50	0.05	0.05	0.05	0.35	0.35	0.05	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_4X}$	$\hat{\pi}_{1t}^{\bar{C}_5X}$	$\hat{\pi}_{1t}^{\bar{C}_6X}$	$\hat{\pi}_{1t}^{\bar{C}_7X}$	$\hat{\pi}_{1t}^{\bar{C}_8X}$
1	0.50	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

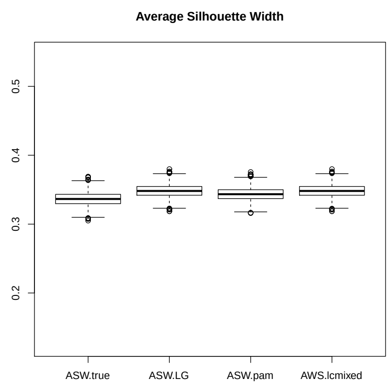
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$	$\hat{\pi}_{1t}^{\bar{D}_5X}$	$\hat{\pi}_{1t}^{\bar{D}_6X}$	$\hat{\pi}_{1t}^{\bar{D}_7X}$	$\hat{\pi}_{1t}^{\bar{D}_8X}$
1	0.50	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Table A.71: Summary: 4 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.354	0.514	0.354
1st Qu.		0.737	0.737	0.737
Median		0.808	0.808	0.808
Mean (se)		0.787 (0.002)	0.802 (0.002)	0.788 (0.002)
3rd Qu.		0.845	0.845	0.845
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.232	0.225	0.237	0.225
1st Qu.	0.313	0.319	0.322	0.319
Median	0.334	0.341	0.343	0.341
Mean (se)	0.335 (0.001)	0.341 (0.001)	0.343 (0.001)	0.341 (0.001)
3rd Qu.	0.356	0.363	0.363	0.363
Max.	0.447	0.447	0.444	0.447
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.482	0.432	0.495	0.432
1st Qu.	0.597	0.605	0.613	0.606
Median	0.627	0.637	0.639	0.637
Mean (se)	0.626 (0.001)	0.634 (0.001)	0.639 (0.001)	0.635 (0.001)
3rd Qu.	0.654	0.666	0.667	0.666
Max.	0.760	0.760	0.760	0.760



(a) *ASW - 100 units.*

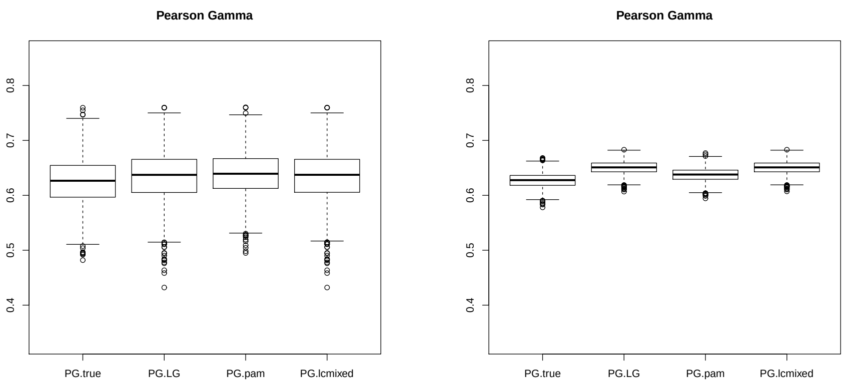


(b) *ASW - 1000 units.*

Figure A.107: Average Silhouette Width: 4 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

Table A.72: Summary: 4 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.767	0.715	0.767
1st Qu.		0.832	0.785	0.832
Median		0.850	0.799	0.850
Mean (se)		0.848 (0.001)	0.800 (0.001)	0.848 (0.001)
3rd Qu.		0.865	0.817	0.865
Max.		0.922	0.891	0.922
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.305	0.319	0.316	0.319
1st Qu.	0.330	0.342	0.337	0.342
Median	0.336	0.348	0.343	0.348
Mean (se)	0.336 (0.001)	0.348 (0.001)	0.343 (0.001)	0.348 (0.001)
3rd Qu.	0.343	0.355	0.350	0.355
Max.	0.369	0.380	0.376	0.380
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.578	0.607	0.594	0.607
1st Qu.	0.618	0.643	0.629	0.643
Median	0.628	0.651	0.638	0.651
Mean (se)	0.627 (0.000)	0.650 (0.000)	0.638 (0.000)	0.650 (0.000)
3rd Qu.	0.636	0.659	0.646	0.659
Max.	0.668	0.683	0.677	0.683



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.108: Pearson Gamma: 4 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

A.37 4 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1X}$	$\pi_t^{\bar{A}_2X}$	$\pi_t^{\bar{A}_3X}$	$\pi_t^{\bar{A}_4X}$	$\pi_t^{\bar{A}_5X}$	$\pi_t^{\bar{A}_6X}$	$\pi_t^{\bar{A}_7X}$	$\pi_t^{\bar{A}_8X}$
1	0.10	0.10	0.05	0.15	0.20	0.05	0.15	0.10	0.20
2	0.15	0.15	0.10	0.25	0.10	0.10	0.10	0.10	0.10
3	0.20	0.05	0.05	0.05	0.25	0.15	0.10	0.15	0.20
4	0.25	0.20	0.05	0.10	0.10	0.10	0.20	0.30	0.10
5	0.30	0.05	0.20	0.10	0.15	0.10	0.05	0.10	0.10

Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.10	0.05	0.10	0.15	0.20	0.20	0.10	0.10	0.10
2	0.15	0.10	0.15	0.05	0.15	0.25	0.05	0.15	0.10
3	0.20	0.15	0.20	0.10	0.10	0.20	0.10	0.05	0.10
4	0.25	0.05	0.10	0.15	0.15	0.15	0.05	0.15	0.20
5	0.30	0.15	0.15	0.10	0.10	0.15	0.05	0.05	0.25

Latent class	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.10	0.15	0.10	0.05	0.20	0.15	0.05	0.10	0.20
2	0.15	0.10	0.15	0.05	0.10	0.20	0.10	0.15	0.15
3	0.20	0.05	0.20	0.10	0.15	0.15	0.10	0.15	0.10
4	0.25	0.05	0.10	0.20	0.05	0.10	0.20	0.20	0.10
5	0.30	0.20	0.10	0.10	0.20	0.10	0.05	0.05	0.20

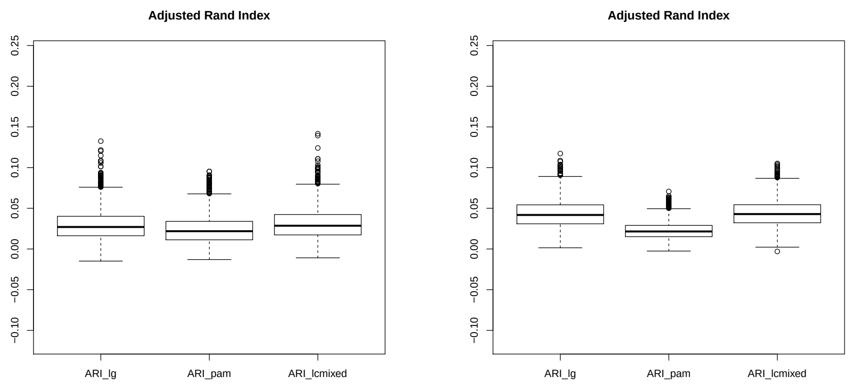
Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.10	0.20	0.05	0.10	0.05	0.15	0.15	0.25	0.05
2	0.15	0.15	0.10	0.15	0.10	0.20	0.10	0.05	0.15
3	0.20	0.10	0.20	0.05	0.15	0.10	0.05	0.15	0.20
4	0.25	0.10	0.15	0.20	0.10	0.15	0.15	0.05	0.10
5	0.30	0.05	0.10	0.25	0.10	0.05	0.20	0.10	0.15

Table A.73: Summary: 4 8-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.015	-0.013	-0.011
1st Qu.		0.016	0.011	0.017
Median		0.027	0.022	0.029
Mean (se)		0.030 (0.000)	0.024 (0.000)	0.031 (0.000)
3rd Qu.		0.040	0.034	0.042
Max.		0.133	0.095	0.142
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.031	0.047	0.050	0.050
1st Qu.	-0.020	0.069	0.072	0.070
Median	-0.017	0.074	0.076	0.074
Mean (se)	-0.016 (0.000)	0.075 (0.000)	0.076 (0.000)	0.075 (0.000)
3rd Qu.	-0.013	0.079	0.081	0.080
Max.	0.006	0.109	0.107	0.103
NA's		3		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.028	0.215	0.228	0.219
1st Qu.	0.062	0.272	0.272	0.273
Median	0.070	0.283	0.283	0.283
Mean (se)	0.072 (0.000)	0.283 (0.000)	0.283 (0.000)	0.283 (0.000)
3rd Qu.	0.081	0.293	0.293	0.294
Max.	0.127	0.344	0.348	0.344
NA's		3		

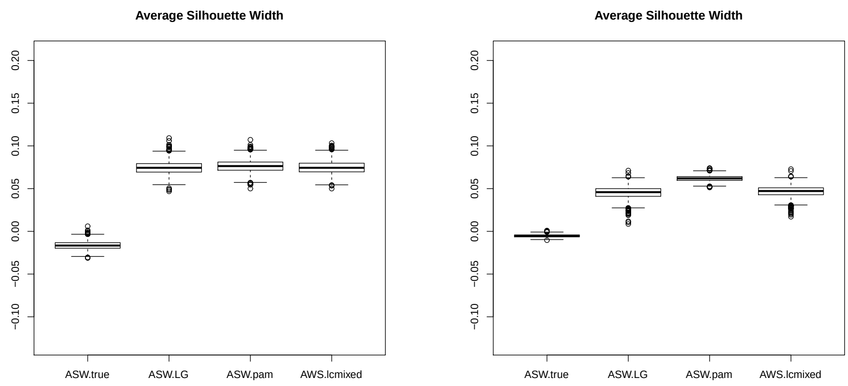
Table A.74: Summary: 4 8-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.001	-0.003	-0.003
1st Qu.		0.031	0.015	0.032
Median		0.042	0.022	0.043
Mean (se)		0.044 (0.000)	0.023 (0.000)	0.044 (0.000)
3rd Qu.		0.054	0.029	0.054
Max.		0.117	0.071	0.105
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.010	0.009	0.051	0.017
1st Qu.	-0.006	0.041	0.060	0.043
Median	-0.005	0.046	0.062	0.047
Mean (se)	-0.005 (0.000)	0.045 (0.000)	0.062 (0.000)	0.047 (0.000)
3rd Qu.	-0.004	0.050	0.064	0.051
Max.	0.001	0.071	0.074	0.073
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.053	0.120	0.215	0.144
1st Qu.	0.067	0.196	0.233	0.200
Median	0.071	0.207	0.238	0.209
Mean (se)	0.071 (0.000)	0.205 (0.000)	0.238 (0.000)	0.208 (0.000)
3rd Qu.	0.075	0.216	0.244	0.217
Max.	0.092	0.250	0.269	0.256



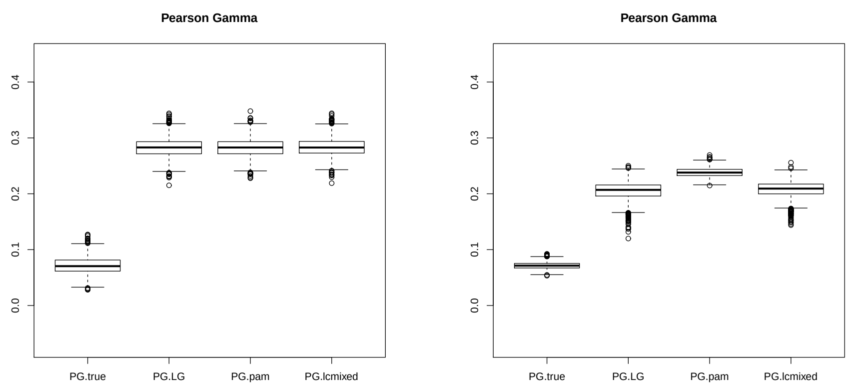
(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

Figure A.109: Adjusted Rand Index: 4 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.110: Average Silhouette Width: 4 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.111: Pearson Gamma: 4 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

A.38 4 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1 X}$	$\pi_t^{\bar{A}_2 X}$	$\pi_t^{\bar{A}_3 X}$	$\pi_t^{\bar{A}_4 X}$	$\pi_t^{\bar{A}_5 X}$	$\pi_t^{\bar{A}_6 X}$	$\pi_t^{\bar{A}_7 X}$	$\pi_t^{\bar{A}_8 X}$
1	0.10	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
2	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
4	0.25	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

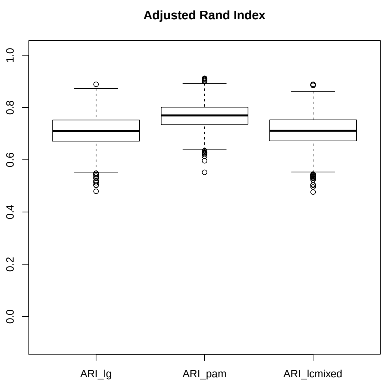
Latent class	π_t^X	$\pi_t^{\bar{B}_1 X}$	$\pi_t^{\bar{B}_2 X}$	$\pi_t^{\bar{B}_3 X}$	$\pi_t^{\bar{B}_4 X}$	$\pi_t^{\bar{B}_5 X}$	$\pi_t^{\bar{B}_6 X}$	$\pi_t^{\bar{B}_7 X}$	$\pi_t^{\bar{B}_8 X}$
1	0.10	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
4	0.25	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{C}_1 X}$	$\pi_t^{\bar{C}_2 X}$	$\pi_t^{\bar{C}_3 X}$	$\pi_t^{\bar{C}_4 X}$	$\pi_t^{\bar{C}_5 X}$	$\pi_t^{\bar{C}_6 X}$	$\pi_t^{\bar{C}_7 X}$	$\pi_t^{\bar{C}_8 X}$
1	0.10	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
4	0.25	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

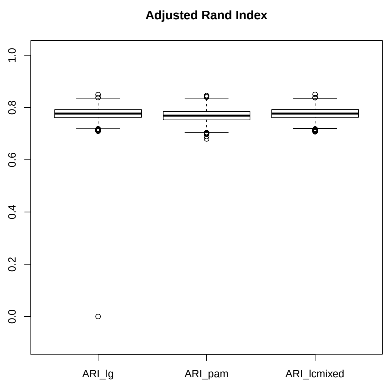
Latent class	π_t^X	$\pi_t^{\bar{D}_1 X}$	$\pi_t^{\bar{D}_2 X}$	$\pi_t^{\bar{D}_3 X}$	$\pi_t^{\bar{D}_4 X}$	$\pi_t^{\bar{D}_5 X}$	$\pi_t^{\bar{D}_6 X}$	$\pi_t^{\bar{D}_7 X}$	$\pi_t^{\bar{D}_8 X}$
1	0.10	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
3	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
4	0.25	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Table A.75: Summary: 4 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.479	0.552	0.476
1st Qu.		0.671	0.736	0.672
Median		0.710	0.770	0.711
Mean (se)		0.709 (0.001)	0.768 (0.001)	0.710 (0.001)
3rd Qu.		0.752	0.801	0.753
Max.		0.889	0.912	0.889
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.234	0.222	0.263	0.218
1st Qu.	0.303	0.315	0.328	0.315
Median	0.320	0.334	0.344	0.333
Mean (se)	0.320 (0.001)	0.333 (0.001)	0.344 (0.001)	0.333 (0.001)
3rd Qu.	0.337	0.351	0.360	0.351
Max.	0.408	0.418	0.420	0.418
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.565	0.520	0.601	0.529
1st Qu.	0.647	0.665	0.683	0.666
Median	0.666	0.685	0.700	0.685
Mean (se)	0.667 (0.001)	0.684 (0.001)	0.699 (0.001)	0.684 (0.001)
3rd Qu.	0.687	0.706	0.717	0.705
Max.	0.757	0.773	0.780	0.773



(a) *ARI - 200 units.*

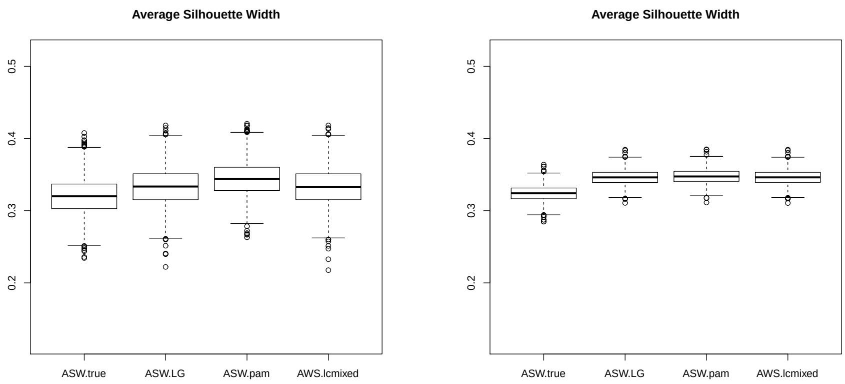


(b) *ARI - 1000 units.*

Figure A.112: Adjusted Rand Index: 4 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

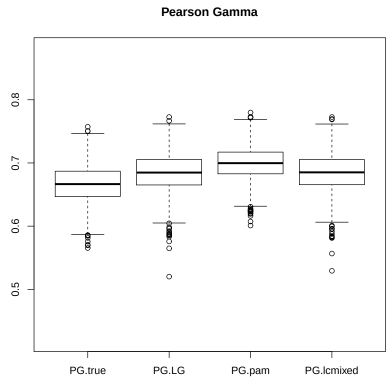
Table A.76: Summary: 4 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.183	0.223
1st Qu.		0.762	0.247	0.327
Median		0.777	0.263	0.342
Mean (se)		0.777 (0.001)	0.262 (0.001)	0.337 (0.001)
3rd Qu.		0.792	0.277	0.354
Max.		0.849	0.340	0.393
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.285	0.311	0.145	0.169
1st Qu.	0.317	0.339	0.183	0.185
Median	0.324	0.346	0.189	0.191
Mean (se)	0.324 (0.000)	0.346 (0.000)	0.189 (0.000)	0.191 (0.000)
3rd Qu.	0.331	0.353	0.196	0.196
Max.	0.364	0.385	0.225	0.234
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.621	0.661	0.275	0.295
1st Qu.	0.659	0.691	0.329	0.380
Median	0.667	0.698	0.342	0.389
Mean (se)	0.667 (0.000)	0.698 (0.000)	0.342 (0.000)	0.385 (0.000)
3rd Qu.	0.676	0.706	0.356	0.397
Max.	0.712	0.737	0.405	0.424

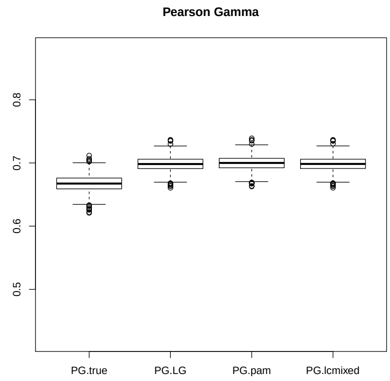


(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.113: Average Silhouette Width: 4 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 200 units.*



(b) *PG - 1000 units.*

Figure A.114: Pearson Gamma: 4 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

A.39 4 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1 X}$	$\pi_t^{\bar{A}_2 X}$	$\pi_t^{\bar{A}_3 X}$	$\pi_t^{\bar{A}_4 X}$	$\pi_t^{\bar{A}_5 X}$	$\pi_t^{\bar{A}_6 X}$	$\pi_t^{\bar{A}_7 X}$	$\pi_t^{\bar{A}_8 X}$
1	0.20	0.10	0.20	0.15	0.05	0.20	0.10	0.15	0.05
2	0.20	0.15	0.15	0.20	0.10	0.15	0.15	0.05	0.05
3	0.20	0.05	0.10	0.15	0.15	0.15	0.10	0.10	0.20
4	0.20	0.20	0.05	0.10	0.10	0.10	0.05	0.30	0.10
5	0.20	0.15	0.15	0.10	0.20	0.10	0.05	0.10	0.15

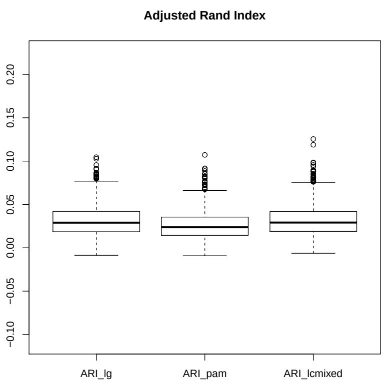
Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.20	0.15	0.05	0.20	0.10	0.25	0.05	0.10	0.10
2	0.20	0.20	0.10	0.15	0.15	0.10	0.10	0.15	0.05
3	0.20	0.10	0.15	0.10	0.25	0.20	0.10	0.05	0.15
4	0.20	0.05	0.20	0.05	0.10	0.15	0.25	0.10	0.10
5	0.20	0.10	0.05	0.15	0.05	0.15	0.05	0.20	0.25

Latent class	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.20	0.10	0.10	0.20	0.05	0.10	0.15	0.20	0.10
2	0.20	0.20	0.05	0.10	0.15	0.20	0.10	0.15	0.05
3	0.20	0.15	0.20	0.05	0.20	0.05	0.10	0.10	0.15
4	0.20	0.05	0.15	0.10	0.30	0.10	0.05	0.05	0.20
5	0.20	0.10	0.10	0.15	0.15	0.10	0.20	0.10	0.10

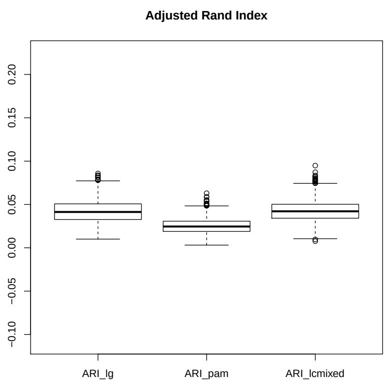
Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.20	0.15	0.20	0.05	0.10	0.20	0.05	0.20	0.05
2	0.20	0.10	0.15	0.10	0.15	0.10	0.20	0.10	0.10
3	0.20	0.20	0.10	0.15	0.05	0.10	0.15	0.05	0.20
4	0.20	0.05	0.10	0.20	0.10	0.15	0.10	0.10	0.20
5	0.20	0.10	0.05	0.15	0.20	0.15	0.05	0.15	0.15

Table A.77: Summary: 4 8-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.009	-0.009	-0.006
1st Qu.		0.019	0.014	0.019
Median		0.029	0.024	0.029
Mean (se)		0.031 (0.000)	0.026 (0.000)	0.032 (0.000)
3rd Qu.		0.042	0.035	0.042
Max.		0.104	0.107	0.126
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.025	0.048	0.053	0.049
1st Qu.	-0.017	0.069	0.070	0.070
Median	-0.014	0.074	0.075	0.075
Mean (se)	-0.013 (0.000)	0.075 (0.000)	0.075 (0.000)	0.075 (0.000)
3rd Qu.	-0.010	0.079	0.080	0.080
Max.	0.009	0.114	0.101	0.107
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.028	0.229	0.227	0.224
1st Qu.	0.063	0.272	0.269	0.274
Median	0.071	0.283	0.280	0.284
Mean (se)	0.072 (0.000)	0.283 (0.000)	0.280 (0.000)	0.285 (0.000)
3rd Qu.	0.080	0.293	0.290	0.295
Max.	0.123	0.351	0.330	0.349



(a) *ARI - 200 units.*

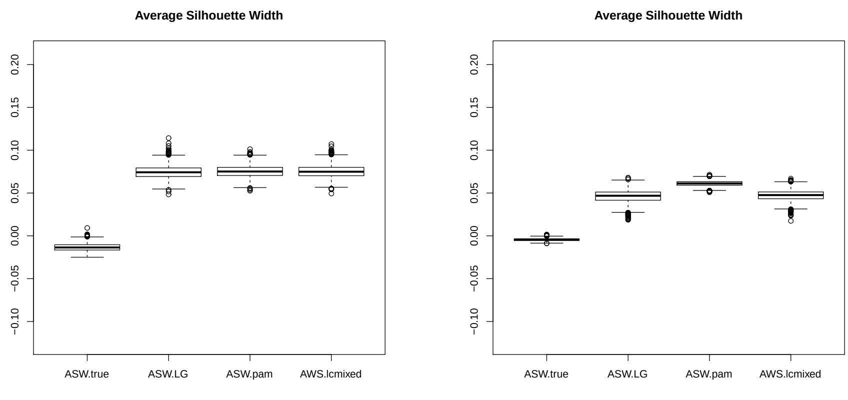


(b) *ARI - 1000 units.*

Figure A.115: Adjusted Rand Index: 4 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

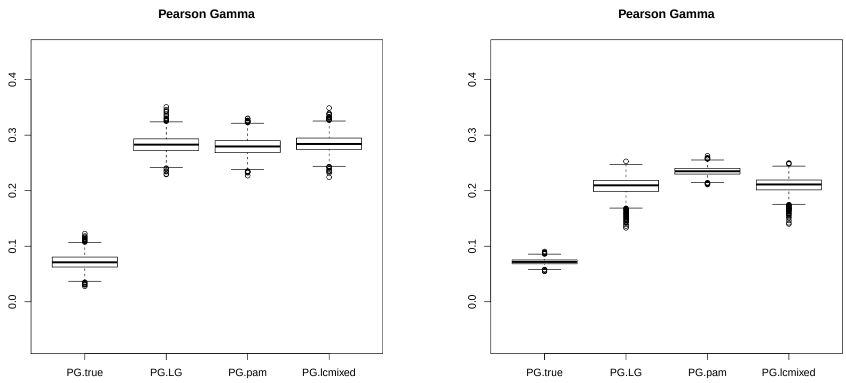
Table A.78: Summary: 4 8-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.010	0.003	0.008
1st Qu.		0.033	0.019	0.034
Median		0.041	0.025	0.042
Mean (se)		0.042 (0.000)	0.025 (0.000)	0.043 (0.000)
3rd Qu.		0.051	0.031	0.050
Max.		0.086	0.063	0.095
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.009	0.019	0.051	0.017
1st Qu.	-0.005	0.042	0.059	0.043
Median	-0.004	0.047	0.061	0.048
Mean (se)	-0.004 (0.000)	0.046 (0.000)	0.061 (0.000)	0.047 (0.000)
3rd Qu.	-0.003	0.051	0.063	0.051
Max.	0.002	0.068	0.071	0.067
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.055	0.133	0.211	0.140
1st Qu.	0.068	0.199	0.230	0.202
Median	0.072	0.210	0.235	0.211
Mean (se)	0.072 (0.000)	0.207 (0.000)	0.235 (0.000)	0.210 (0.000)
3rd Qu.	0.075	0.219	0.240	0.219
Max.	0.091	0.253	0.263	0.250



(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.116: Average Silhouette Width: 4 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

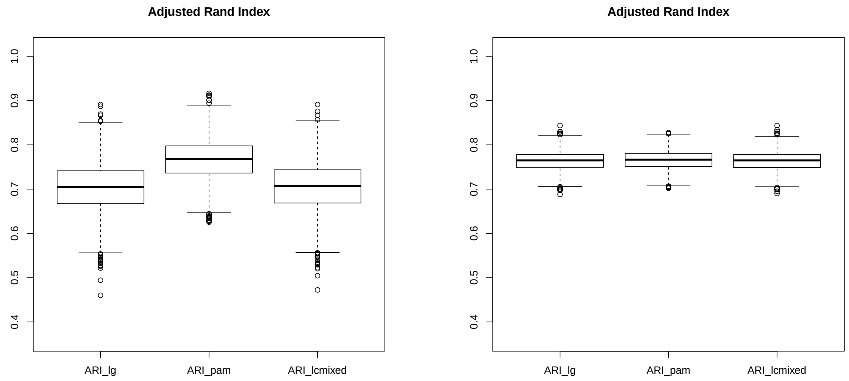
Figure A.117: Pearson Gamma: 4 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

A.40 4 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1 X}$	$\pi_t^{\bar{A}_2 X}$	$\pi_t^{\bar{A}_3 X}$	$\pi_t^{\bar{A}_4 X}$	$\pi_t^{\bar{A}_5 X}$	$\pi_t^{\bar{A}_6 X}$	$\pi_t^{\bar{A}_7 X}$	$\pi_t^{\bar{A}_8 X}$
1	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

Latent class	π_t^X	$\pi_t^{\bar{B}_1 X}$	$\pi_t^{\bar{B}_2 X}$	$\pi_t^{\bar{B}_3 X}$	$\pi_t^{\bar{B}_4 X}$	$\pi_t^{\bar{B}_5 X}$	$\pi_t^{\bar{B}_6 X}$	$\pi_t^{\bar{B}_7 X}$	$\pi_t^{\bar{B}_8 X}$
1	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
4	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
5	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05



(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

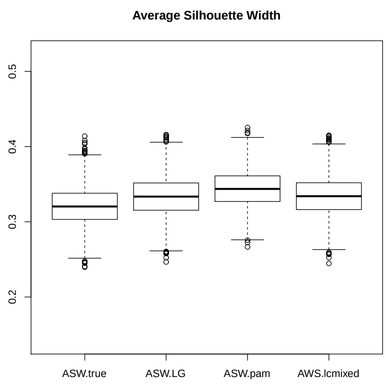
Figure A.118: Adjusted Rand Index: 4 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

Latent class	π_t^X	$\pi_t^{\bar{C}_1 X}$	$\pi_t^{\bar{C}_2 X}$	$\pi_t^{\bar{C}_3 X}$	$\pi_t^{\bar{C}_4 X}$	$\pi_t^{\bar{C}_5 X}$	$\pi_t^{\bar{C}_6 X}$	$\pi_t^{\bar{C}_7 X}$	$\pi_t^{\bar{C}_8 X}$
1	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
2	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
5	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

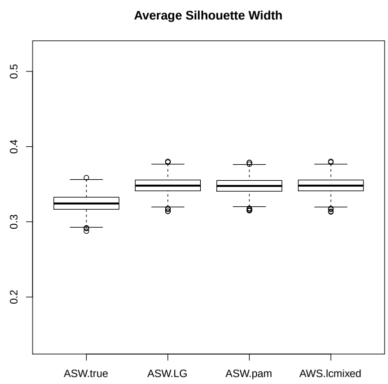
Latent class	π_t^X	$\pi_t^{\bar{D}_1 X}$	$\pi_t^{\bar{D}_2 X}$	$\pi_t^{\bar{D}_3 X}$	$\pi_t^{\bar{D}_4 X}$	$\pi_t^{\bar{D}_5 X}$	$\pi_t^{\bar{D}_6 X}$	$\pi_t^{\bar{D}_7 X}$	$\pi_t^{\bar{D}_8 X}$
1	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
2	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
5	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Table A.79: Summary: 4 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.460	0.625	0.472
1st Qu.		0.667	0.736	0.669
Median		0.705	0.768	0.707
Mean (se)		0.704 (0.001)	0.766 (0.001)	0.705 (0.001)
3rd Qu.		0.741	0.798	0.744
Max.		0.891	0.916	0.891
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.240	0.247	0.267	0.245
1st Qu.	0.303	0.315	0.327	0.316
Median	0.320	0.334	0.344	0.334
Mean (se)	0.321 (0.001)	0.334 (0.001)	0.344 (0.001)	0.334 (0.001)
3rd Qu.	0.338	0.352	0.361	0.352
Max.	0.414	0.416	0.425	0.415
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.558	0.542	0.603	0.557
1st Qu.	0.640	0.655	0.672	0.655
Median	0.659	0.675	0.689	0.676
Mean (se)	0.659 (0.001)	0.674 (0.001)	0.688 (0.001)	0.675 (0.001)
3rd Qu.	0.678	0.695	0.706	0.696
Max.	0.745	0.763	0.764	0.759



(a) *ASW - 200 units.*

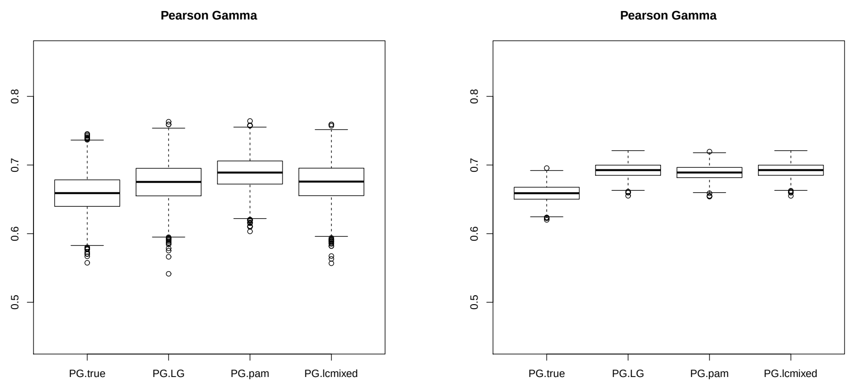


(b) *ASW - 1000 units.*

Figure A.119: Average Silhouette Width: 4 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

Table A.80: Summary: 4 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.688	0.702	0.690
1st Qu.		0.749	0.751	0.749
Median		0.765	0.767	0.765
Mean (se)		0.764 (0.000)	0.766 (0.000)	0.764 (0.000)
3rd Qu.		0.778	0.781	0.778
Max.		0.844	0.828	0.844
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.288	0.314	0.315	0.313
1st Qu.	0.317	0.341	0.341	0.341
Median	0.324	0.348	0.348	0.348
Mean (se)	0.324 (0.000)	0.348 (0.000)	0.348 (0.000)	0.348 (0.000)
3rd Qu.	0.333	0.356	0.355	0.356
Max.	0.359	0.380	0.379	0.380
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.620	0.655	0.654	0.655
1st Qu.	0.650	0.685	0.682	0.685
Median	0.659	0.693	0.689	0.693
Mean (se)	0.659 (0.000)	0.692 (0.000)	0.696 (0.000)	0.692 (0.000)
3rd Qu.	0.668	0.700	0.696	0.700
Max.	0.696	0.721	0.720	0.721



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.120: Pearson Gamma: 4 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

A.41 12 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1X}$	$\pi_t^{\bar{A}_2X}$	$\pi_t^{\bar{A}_3X}$	$\pi_t^{\bar{A}_4X}$	$\pi_t^{\bar{A}_5X}$	$\pi_t^{\bar{A}_6X}$	$\pi_t^{\bar{A}_7X}$	$\pi_t^{\bar{A}_8X}$
1	0.85	0.05	0.15	0.10	0.20	0.30	0.05	0.10	0.05
2	0.15	0.10	0.20	0.15	0.10	0.20	0.10	0.05	0.10

Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.85	0.10	0.15	0.05	0.20	0.10	0.20	0.15	0.05
2	0.15	0.20	0.10	0.15	0.15	0.15	0.15	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.85	0.20	0.05	0.10	0.05	0.15	0.20	0.15	0.10
2	0.15	0.15	0.10	0.05	0.10	0.20	0.10	0.20	0.10

Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.85	0.10	0.15	0.05	0.20	0.15	0.05	0.20	0.10
2	0.15	0.15	0.10	0.10	0.15	0.20	0.10	0.05	0.15

Latent class	π_t^X	$\pi_t^{\bar{E}_1X}$	$\pi_t^{\bar{E}_2X}$	$\pi_t^{\bar{E}_3X}$	$\pi_t^{\bar{E}_4X}$	$\pi_t^{\bar{E}_5X}$	$\pi_t^{\bar{E}_6X}$	$\pi_t^{\bar{E}_7X}$	$\pi_t^{\bar{E}_8X}$
1	0.85	0.10	0.20	0.05	0.15	0.10	0.20	0.05	0.15
2	0.15	0.05	0.10	0.10	0.05	0.20	0.10	0.10	0.30

Latent class	π_t^X	$\pi_t^{\bar{F}_1X}$	$\pi_t^{\bar{F}_2X}$	$\pi_t^{\bar{F}_3X}$	$\pi_t^{\bar{F}_4X}$	$\pi_t^{\bar{F}_5X}$	$\pi_t^{\bar{F}_6X}$	$\pi_t^{\bar{F}_7X}$	$\pi_t^{\bar{F}_8X}$
1	0.85	0.20	0.10	0.05	0.15	0.10	0.05	0.25	0.10
2	0.15	0.15	0.15	0.10	0.10	0.20	0.20	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{G}_1X}$	$\pi_t^{\bar{G}_2X}$	$\pi_t^{\bar{G}_3X}$	$\pi_t^{\bar{G}_4X}$	$\pi_t^{\bar{G}_5X}$	$\pi_t^{\bar{G}_6X}$	$\pi_t^{\bar{G}_7X}$	$\pi_t^{\bar{G}_8X}$
1	0.85	0.10	0.20	0.05	0.10	0.15	0.10	0.20	0.10
2	0.15	0.05	0.15	0.20	0.05	0.10	0.05	0.10	0.30

Latent class	π_t^X	$\pi_t^{\bar{H}_1X}$	$\pi_t^{\bar{H}_2X}$	$\pi_t^{\bar{H}_3X}$	$\pi_t^{\bar{H}_4X}$	$\pi_t^{\bar{H}_5X}$	$\pi_t^{\bar{H}_6X}$	$\pi_t^{\bar{H}_7X}$	$\pi_t^{\bar{H}_8X}$
1	0.85	0.20	0.10	0.05	0.15	0.10	0.20	0.05	0.15
2	0.15	0.10	0.05	0.15	0.05	0.05	0.10	0.20	0.30

Latent class	π_t^X	$\pi_t^{\bar{I}_1X}$	$\pi_t^{\bar{I}_2X}$	$\pi_t^{\bar{I}_3X}$	$\pi_t^{\bar{I}_4X}$	$\pi_t^{\bar{I}_5X}$	$\pi_t^{\bar{I}_6X}$	$\pi_t^{\bar{I}_7X}$	$\pi_t^{\bar{I}_8X}$
1	0.85	0.05	0.10	0.25	0.10	0.05	0.05	0.10	0.30
2	0.15	0.10	0.05	0.15	0.30	0.10	0.10	0.05	0.15

Latent class	π_t^X	$\pi_t^{\bar{L}_1X}$	$\pi_t^{\bar{L}_2X}$	$\pi_t^{\bar{L}_3X}$	$\pi_t^{\bar{L}_4X}$	$\pi_t^{\bar{L}_5X}$	$\pi_t^{\bar{L}_6X}$	$\pi_t^{\bar{L}_7X}$	$\pi_t^{\bar{L}_8X}$
1	0.85	0.05	0.10	0.20	0.15	0.05	0.05	0.25	0.15
2	0.15	0.10	0.05	0.10	0.25	0.10	0.20	0.05	0.15

Latent class	π_t^X	$\pi_t^{\bar{M}_1X}$	$\pi_t^{\bar{M}_2X}$	$\pi_t^{\bar{M}_3X}$	$\pi_t^{\bar{M}_4X}$	$\pi_t^{\bar{M}_5X}$	$\pi_t^{\bar{M}_6X}$	$\pi_t^{\bar{M}_7X}$	$\pi_t^{\bar{M}_8X}$
1	0.85	0.10	0.05	0.15	0.10	0.30	0.05	0.05	0.20
2	0.15	0.15	0.15	0.10	0.05	0.25	0.10	0.15	0.05

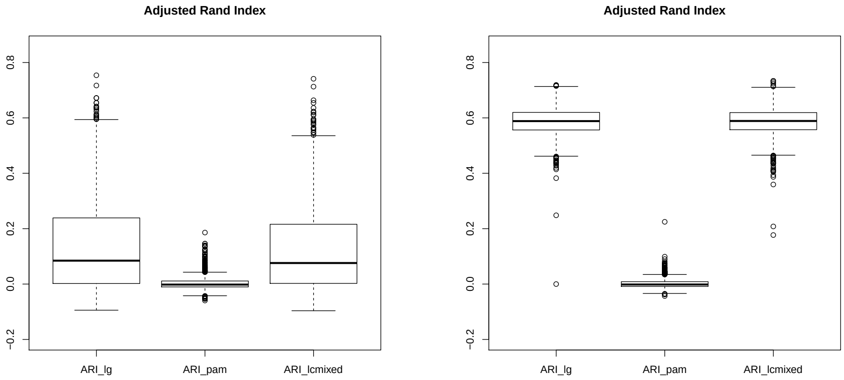
Latent class	π_t^X	$\pi_t^{\bar{N}_1X}$	$\pi_t^{\bar{N}_2X}$	$\pi_t^{\bar{N}_3X}$	$\pi_t^{\bar{N}_4X}$	$\pi_t^{\bar{N}_5X}$	$\pi_t^{\bar{N}_6X}$	$\pi_t^{\bar{N}_7X}$	$\pi_t^{\bar{N}_8X}$
1	0.85	0.05	0.10	0.30	0.05	0.20	0.10	0.15	0.05
2	0.15	0.10	0.05	0.15	0.05	0.10	0.20	0.20	0.15

Table A.81: Summary: 12 8-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.094	-0.059	-0.096
1st Qu.		0.002	-0.010	0.002
Median		0.085	-0.002	0.076
Mean (se)		0.135 (0.004)	0.004 (0.001)	0.123 (0.003)
3rd Qu.		0.239	0.011	0.216
Max.		0.754	0.186	0.741
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.018	0.021	0.017	0.022
1st Qu.	0.034	0.034	0.027	0.033
Median	0.037	0.037	0.029	0.037
Mean (se)	0.037 (0.000)	0.038 (0.000)	0.029 (0.000)	0.038 (0.000)
3rd Qu.	0.041	0.042	0.031	0.042
Max.	0.055	0.058	0.042	0.059
NA's		3		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.060	0.043	0.066	0.047
1st Qu.	0.125	0.132	0.104	0.132
Median	0.139	0.151	0.115	0.151
Mean (se)	0.140 (0.000)	0.151 (0.001)	0.115 (0.000)	0.152 (0.001)
3rd Qu.	0.155	0.173	0.125	0.173
Max.	0.203	0.241	0.175	0.237
NA's		3		

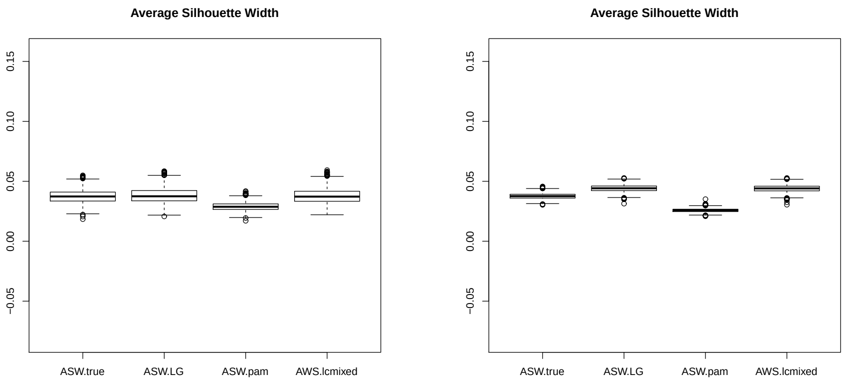
Table A.82: Summary: 12 8-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	-0.043	0.177
1st Qu.		0.557	-0.009	0.557
Median		0.588	-0.002	0.589
Mean (se)		0.586 (0.001)	0.002 (0.000)	0.586 (0.001)
3rd Qu.		0.620	0.009	0.619
Max.		0.719	0.225	0.734
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.030	0.031	0.021	0.030
1st Qu.	0.036	0.042	0.025	0.042
Median	0.038	0.044	0.026	0.044
Mean (se)	0.038 (0.000)	0.044 (0.000)	0.026 (0.000)	0.044 (0.000)
3rd Qu.	0.039	0.046	0.027	0.046
Max.	0.046	0.053	0.035	0.053
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.110	0.107	0.077	0.110
1st Qu.	0.133	0.145	0.096	0.148
Median	0.140	0.155	0.101	0.157
Mean (se)	0.140 (0.000)	0.154 (0.000)	0.101 (0.000)	0.156 (0.000)
3rd Qu.	0.146	0.164	0.106	0.166
Max.	0.169	0.194	0.148	0.195
NA's		1		



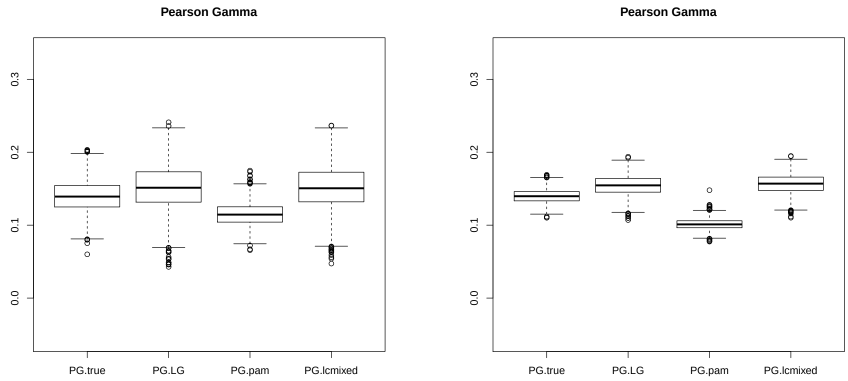
(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

Figure A.121: Average Silhouette Width: 12 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.122: Average Silhouette Width: 12 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.123: Pearson Gamma: 12 8-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

A.42 12 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1X}$	$\pi_t^{\bar{A}_2X}$	$\pi_t^{\bar{A}_3X}$	$\pi_t^{\bar{A}_4X}$	$\pi_t^{\bar{A}_5X}$	$\pi_t^{\bar{A}_6X}$	$\pi_t^{\bar{A}_7X}$	$\pi_t^{\bar{A}_8X}$
1	0.85	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
2	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.85	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.15	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.85	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.85	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.15	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{E}_1X}$	$\pi_t^{\bar{E}_2X}$	$\pi_t^{\bar{E}_3X}$	$\pi_t^{\bar{E}_4X}$	$\pi_t^{\bar{E}_5X}$	$\pi_t^{\bar{E}_6X}$	$\pi_t^{\bar{E}_7X}$	$\pi_t^{\bar{E}_8X}$
1	0.85	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{F}_1X}$	$\pi_t^{\bar{F}_2X}$	$\pi_t^{\bar{F}_3X}$	$\pi_t^{\bar{F}_4X}$	$\pi_t^{\bar{F}_5X}$	$\pi_t^{\bar{F}_6X}$	$\pi_t^{\bar{F}_7X}$	$\pi_t^{\bar{F}_8X}$
1	0.85	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	π_t^X	$\pi_t^{\bar{G}_1X}$	$\pi_t^{\bar{G}_2X}$	$\pi_t^{\bar{G}_3X}$	$\pi_t^{\bar{G}_4X}$	$\pi_t^{\bar{G}_5X}$	$\pi_t^{\bar{G}_6X}$	$\pi_t^{\bar{G}_7X}$	$\pi_t^{\bar{G}_8X}$
1	0.85	0.05	0.05	0.05	0.35	0.35	0.05	0.05	0.05
2	0.15	0.05	0.05	0.35	0.05	0.05	0.35	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{H}_1X}$	$\pi_t^{\bar{H}_2X}$	$\pi_t^{\bar{H}_3X}$	$\pi_t^{\bar{H}_4X}$	$\pi_t^{\bar{H}_5X}$	$\pi_t^{\bar{H}_6X}$	$\pi_t^{\bar{H}_7X}$	$\pi_t^{\bar{H}_8X}$
1	0.85	0.05	0.35	0.35	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.35	0.35	0.05

Latent class	π_t^X	$\pi_t^{\bar{I}_1X}$	$\pi_t^{\bar{I}_2X}$	$\pi_t^{\bar{I}_3X}$	$\pi_t^{\bar{I}_4X}$	$\pi_t^{\bar{I}_5X}$	$\pi_t^{\bar{I}_6X}$	$\pi_t^{\bar{I}_7X}$	$\pi_t^{\bar{I}_8X}$
1	0.85	0.35	0.05	0.05	0.05	0.05	0.05	0.05	0.35
2	0.15	0.05	0.05	0.05	0.05	0.05	0.35	0.35	0.05

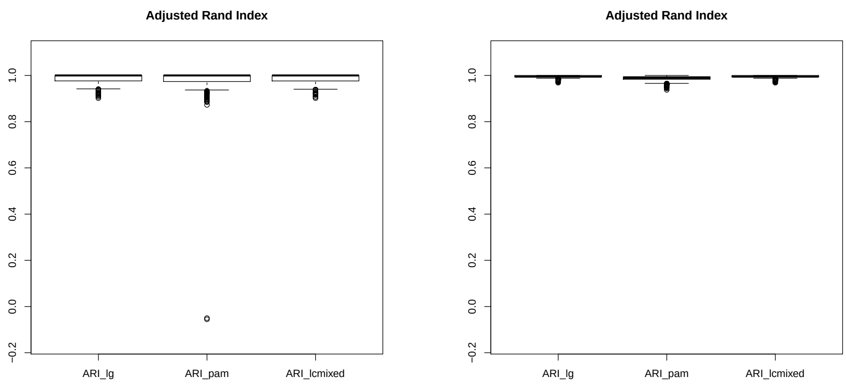
Latent class	π_t^X	$\pi_t^{\bar{L}_1X}$	$\pi_t^{\bar{L}_2X}$	$\pi_t^{\bar{L}_3X}$	$\pi_t^{\bar{L}_4X}$	$\pi_t^{\bar{L}_5X}$	$\pi_t^{\bar{L}_6X}$	$\pi_t^{\bar{L}_7X}$	$\pi_t^{\bar{L}_8X}$
1	0.85	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{M}_1X}$	$\pi_t^{\bar{M}_2X}$	$\pi_t^{\bar{M}_3X}$	$\pi_t^{\bar{M}_4X}$	$\pi_t^{\bar{M}_5X}$	$\pi_t^{\bar{M}_6X}$	$\pi_t^{\bar{M}_7X}$	$\pi_t^{\bar{M}_8X}$
1	0.85	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{N}_1X}$	$\pi_t^{\bar{N}_2X}$	$\pi_t^{\bar{N}_3X}$	$\pi_t^{\bar{N}_4X}$	$\pi_t^{\bar{N}_5X}$	$\pi_t^{\bar{N}_6X}$	$\pi_t^{\bar{N}_7X}$	$\pi_t^{\bar{N}_8X}$
1	0.85	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05

Table A.83: Summary: 12 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.902	-0.055	0.902
1st Qu.		0.976	0.974	0.976
Median		1.000	1.000	1.000
Mean (se)		0.992 (0.000)	0.984 (0.001)	0.991 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.298	0.293	0.027	0.293
1st Qu.	0.331	0.330	0.329	0.330
Median	0.340	0.340	0.339	0.340
Mean (se)	0.340 (0.000)	0.340 (0.000)	0.339 (0.000)	0.340 (0.000)
3rd Qu.	0.349	0.349	0.348	0.349
Max.	0.388	0.388	0.388	0.388
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.629	0.629	0.036	0.629
1st Qu.	0.718	0.719	0.720	0.719
Median	0.736	0.736	0.736	0.736
Mean(se)	0.735 (0.001)	0.735 (0.001)	0.734 (0.001)	0.735 (0.001)
3rd Qu.	0.753	0.753	0.754	0.753
Max.	0.803	0.808	0.809	0.808

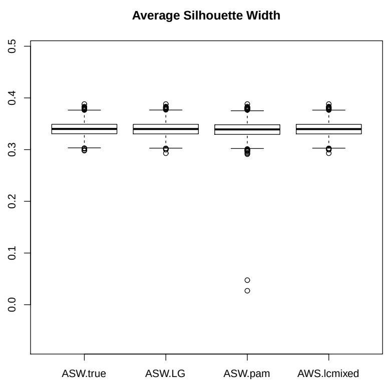


(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

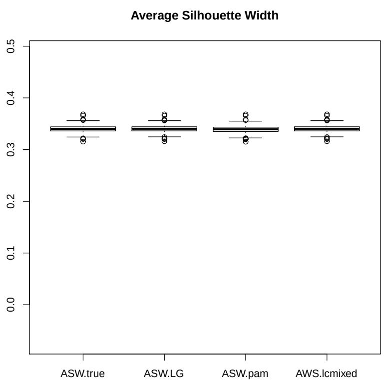
Figure A.124: Adjusted Rand Index: 12 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Table A.84: Summary: 12 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.969	0.938	0.969
1st Qu.		0.994	0.983	0.994
Median		0.995	0.989	0.995
Mean (se)		0.996 (0.000)	0.987 (0.000)	0.996 (0.000)
3rd Qu.		1.000	0.995	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.315	0.316	0.315	0.316
1st Qu.	0.336	0.336	0.335	0.336
Median	0.340	0.340	0.339	0.340
Mean (se)	0.340 (0.000)	0.340 (0.000)	0.339 (0.000)	0.340 (0.000)
3rd Qu.	0.344	0.344	0.343	0.344
Max.	0.368	0.368	0.368	0.368
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.696	0.695	0.696	0.695
1st Qu.	0.728	0.728	0.729	0.728
Median	0.735	0.735	0.736	0.735
Mean (se)	0.735 (0.000)	0.735 (0.000)	0.736 (0.000)	0.735 (0.000)
3rd Qu.	0.742	0.742	0.743	0.742
Max.	0.769	0.769	0.769	0.769

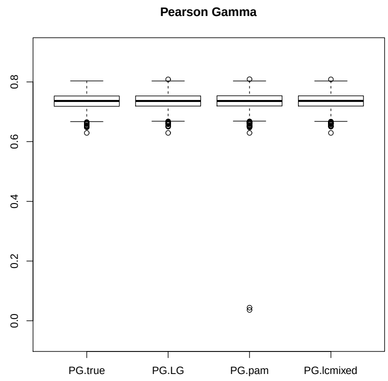


(a) *ASW - 200 units.*

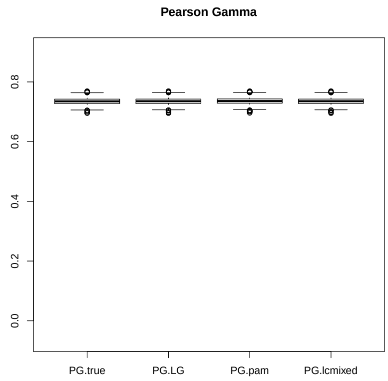


(b) *ASW - 1000 units.*

Figure A.125: Average Silhouette Width: 12 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 200 units.*



(b) *PG - 1000 units.*

Figure A.126: Pearson Gamma: 12 8-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

A.43 12 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1X}$	$\pi_t^{\bar{A}_2X}$	$\pi_t^{\bar{A}_3X}$	$\pi_t^{\bar{A}_4X}$	$\pi_t^{\bar{A}_5X}$	$\pi_t^{\bar{A}_6X}$	$\pi_t^{\bar{A}_7X}$	$\pi_t^{\bar{A}_8X}$
1	0.50	0.20	0.05	0.15	0.10	0.15	0.05	0.20	0.10
2	0.50	0.10	0.20	0.05	0.15	0.10	0.10	0.05	0.25

Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.50	0.10	0.20	0.20	0.05	0.15	0.05	0.10	0.15
2	0.50	0.15	0.10	0.15	0.10	0.20	0.10	0.15	0.05

Latent class	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.50	0.15	0.05	0.10	0.20	0.10	0.30	0.05	0.05
2	0.50	0.20	0.10	0.05	0.15	0.20	0.10	0.10	0.10

Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.50	0.10	0.20	0.05	0.15	0.10	0.15	0.20	0.05
2	0.50	0.10	0.05	0.20	0.05	0.10	0.20	0.10	0.20

Latent class	π_t^X	$\pi_t^{\bar{E}_1X}$	$\pi_t^{\bar{E}_2X}$	$\pi_t^{\bar{E}_3X}$	$\pi_t^{\bar{E}_4X}$	$\pi_t^{\bar{E}_5X}$	$\pi_t^{\bar{E}_6X}$	$\pi_t^{\bar{E}_7X}$	$\pi_t^{\bar{E}_8X}$
1	0.50	0.15	0.10	0.05	0.20	0.10	0.20	0.05	0.15
2	0.50	0.10	0.30	0.10	0.05	0.05	0.10	0.20	0.10

Latent class	π_t^X	$\pi_t^{\bar{F}_1X}$	$\pi_t^{\bar{F}_2X}$	$\pi_t^{\bar{F}_3X}$	$\pi_t^{\bar{F}_4X}$	$\pi_t^{\bar{F}_5X}$	$\pi_t^{\bar{F}_6X}$	$\pi_t^{\bar{F}_7X}$	$\pi_t^{\bar{F}_8X}$
1	0.50	0.10	0.15	0.25	0.05	0.05	0.10	0.15	0.15
2	0.50	0.20	0.10	0.10	0.20	0.10	0.05	0.10	0.15

Latent class	π_t^X	$\pi_t^{\bar{G}_1X}$	$\pi_t^{\bar{G}_2X}$	$\pi_t^{\bar{G}_3X}$	$\pi_t^{\bar{G}_4X}$	$\pi_t^{\bar{G}_5X}$	$\pi_t^{\bar{G}_6X}$	$\pi_t^{\bar{G}_7X}$	$\pi_t^{\bar{G}_8X}$
1	0.50	0.15	0.10	0.15	0.05	0.15	0.20	0.05	0.15
2	0.50	0.20	0.05	0.10	0.15	0.20	0.10	0.15	0.05

Latent class	π_t^X	$\pi_t^{\bar{H}_1X}$	$\pi_t^{\bar{H}_2X}$	$\pi_t^{\bar{H}_3X}$	$\pi_t^{\bar{H}_4X}$	$\pi_t^{\bar{H}_5X}$	$\pi_t^{\bar{H}_6X}$	$\pi_t^{\bar{H}_7X}$	$\pi_t^{\bar{H}_8X}$
1	0.50	0.05	0.25	0.10	0.15	0.15	0.20	0.05	0.05
2	0.50	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20

Latent class	π_t^X	$\pi_t^{\bar{I}_1X}$	$\pi_t^{\bar{I}_2X}$	$\pi_t^{\bar{I}_3X}$	$\pi_t^{\bar{I}_4X}$	$\pi_t^{\bar{I}_5X}$	$\pi_t^{\bar{I}_6X}$	$\pi_t^{\bar{I}_7X}$	$\pi_t^{\bar{I}_8X}$
1	0.50	0.20	0.05	0.10	0.15	0.10	0.20	0.05	0.15
2	0.50	0.15	0.10	0.15	0.20	0.05	0.15	0.15	0.05

Latent class	π_t^X	$\pi_t^{\bar{L}_1X}$	$\pi_t^{\bar{L}_2X}$	$\pi_t^{\bar{L}_3X}$	$\pi_t^{\bar{L}_4X}$	$\pi_t^{\bar{L}_5X}$	$\pi_t^{\bar{L}_6X}$	$\pi_t^{\bar{L}_7X}$	$\pi_t^{\bar{L}_8X}$
1	0.50	0.10	0.15	0.05	0.10	0.30	0.05	0.05	0.20
2	0.50	0.20	0.20	0.10	0.05	0.05	0.10	0.20	0.10

Latent class	π_t^X	$\pi_t^{\bar{M}_1X}$	$\pi_t^{\bar{M}_2X}$	$\pi_t^{\bar{M}_3X}$	$\pi_t^{\bar{M}_4X}$	$\pi_t^{\bar{M}_5X}$	$\pi_t^{\bar{M}_6X}$	$\pi_t^{\bar{M}_7X}$	$\pi_t^{\bar{M}_8X}$
1	0.50	0.05	0.30	0.15	0.10	0.05	0.15	0.05	0.15
2	0.50	0.20	0.10	0.05	0.15	0.20	0.10	0.15	0.05

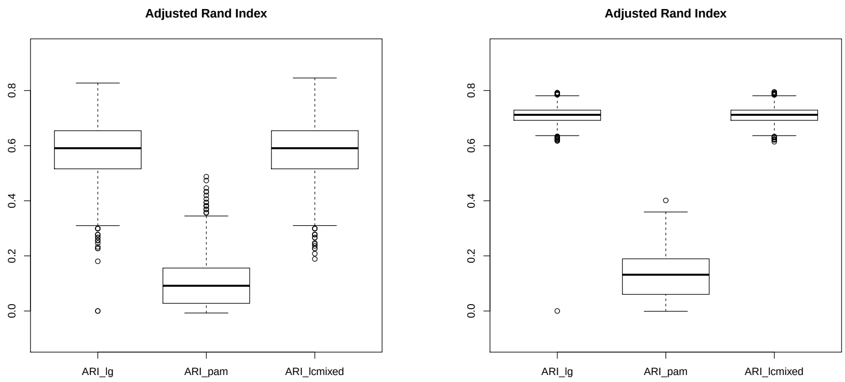
Latent class	π_t^X	$\pi_t^{\bar{N}_1X}$	$\pi_t^{\bar{N}_2X}$	$\pi_t^{\bar{N}_3X}$	$\pi_t^{\bar{N}_4X}$	$\pi_t^{\bar{N}_5X}$	$\pi_t^{\bar{N}_6X}$	$\pi_t^{\bar{N}_7X}$	$\pi_t^{\bar{N}_8X}$
1	0.50	0.30	0.05	0.15	0.10	0.05	0.10	0.05	0.20
2	0.50	0.05	0.15	0.05	0.20	0.15	0.05	0.25	0.10

Table A.85: Summary: 12 8-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	-0.007	0.188
1st Qu.		0.516	0.028	0.516
Median		0.591	0.092	0.591
Mean (se)		0.579 (0.002)	0.105 (0.002)	0.581 (0.002)
3rd Qu.		0.654	0.156	0.654
Max.		0.827	0.487	0.846
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.032	0.036	0.018	0.034
1st Qu.	0.042	0.048	0.026	0.048
Median	0.046	0.051	0.030	0.051
Mean (se)	0.046 (0.000)	0.051 (0.000)	0.030 (0.000)	0.051 (0.000)
3rd Qu.	0.049	0.054	0.034	0.054
Max.	0.061	0.066	0.057	0.066
NA's		2		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.138	0.132	0.072	0.116
1st Qu.	0.189	0.210	0.110	0.211
Median	0.202	0.224	0.126	0.225
Mean (se)	0.203 (0.000)	0.224 (0.000)	0.129 (0.001)	0.224 (0.000)
3rd Qu.	0.216	0.238	0.145	0.238
Max.	0.272	0.293	0.251	0.293
NA's		2		

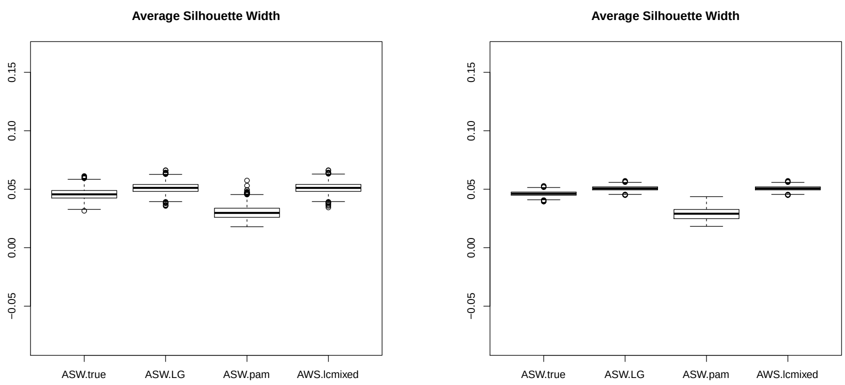
Table A.86: Summary: 12 8-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	-0.001	0.614
1st Qu.		0.692	0.060	0.692
Median		0.712	0.132	0.712
Mean (se)		0.710 (0.001)	0.128 (0.002)	0.711 (0.001)
3rd Qu.		0.729	0.189	0.729
Max.		0.792	0.401	0.796
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.039	0.045	0.018	0.045
1st Qu.	0.045	0.049	0.025	0.049
Median	0.046	0.051	0.029	0.051
Mean (se)	0.046 (0.000)	0.051 (0.000)	0.029 (0.000)	0.051 (0.000)
3rd Qu.	0.048	0.052	0.033	0.052
Max.	0.053	0.057	0.044	0.057
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.175	0.198	0.077	0.199
1st Qu.	0.198	0.218	0.104	0.218
Median	0.204	0.224	0.123	0.224
Mean (se)	0.204 (0.000)	0.224 (0.000)	0.122 (0.000)	0.224 (0.000)
3rd Qu.	0.210	0.230	0.139	0.230
Max.	0.233	0.253	0.189	0.253
NA's		1		



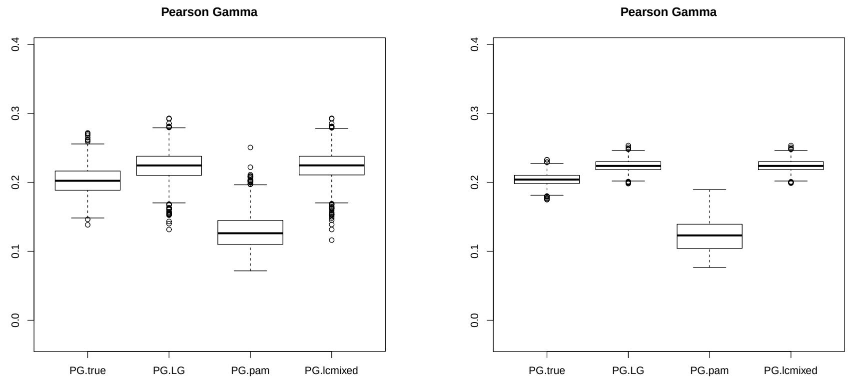
(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

Figure A.127: Adjusted Rand Index: 12 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.128: Average Silhouette Width: 12 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.129: Pearson Gamma: 12 8-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

A.44 12 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1X}$	$\pi_t^{\bar{A}_2X}$	$\pi_t^{\bar{A}_3X}$	$\pi_t^{\bar{A}_4X}$	$\pi_t^{\bar{A}_5X}$	$\pi_t^{\bar{A}_6X}$	$\pi_t^{\bar{A}_7X}$	$\pi_t^{\bar{A}_8X}$
1	0.50	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
2	0.50	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.50	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
2	0.50	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

Latent class	π_t^X	$\pi_t^{\bar{E}_1X}$	$\pi_t^{\bar{E}_2X}$	$\pi_t^{\bar{E}_3X}$	$\pi_t^{\bar{E}_4X}$	$\pi_t^{\bar{E}_5X}$	$\pi_t^{\bar{E}_6X}$	$\pi_t^{\bar{E}_7X}$	$\pi_t^{\bar{E}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.50	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{F}_1X}$	$\pi_t^{\bar{F}_2X}$	$\pi_t^{\bar{F}_3X}$	$\pi_t^{\bar{F}_4X}$	$\pi_t^{\bar{F}_5X}$	$\pi_t^{\bar{F}_6X}$	$\pi_t^{\bar{F}_7X}$	$\pi_t^{\bar{F}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	π_t^X	$\pi_t^{\bar{G}_1X}$	$\pi_t^{\bar{G}_2X}$	$\pi_t^{\bar{G}_3X}$	$\pi_t^{\bar{G}_4X}$	$\pi_t^{\bar{G}_5X}$	$\pi_t^{\bar{G}_6X}$	$\pi_t^{\bar{G}_7X}$	$\pi_t^{\bar{G}_8X}$
1	0.50	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.50	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{H}_1X}$	$\pi_t^{\bar{H}_2X}$	$\pi_t^{\bar{H}_3X}$	$\pi_t^{\bar{H}_4X}$	$\pi_t^{\bar{H}_5X}$	$\pi_t^{\bar{H}_6X}$	$\pi_t^{\bar{H}_7X}$	$\pi_t^{\bar{H}_8X}$
1	0.50	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{I}_1X}$	$\pi_t^{\bar{I}_2X}$	$\pi_t^{\bar{I}_3X}$	$\pi_t^{\bar{I}_4X}$	$\pi_t^{\bar{I}_5X}$	$\pi_t^{\bar{I}_6X}$	$\pi_t^{\bar{I}_7X}$	$\pi_t^{\bar{I}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

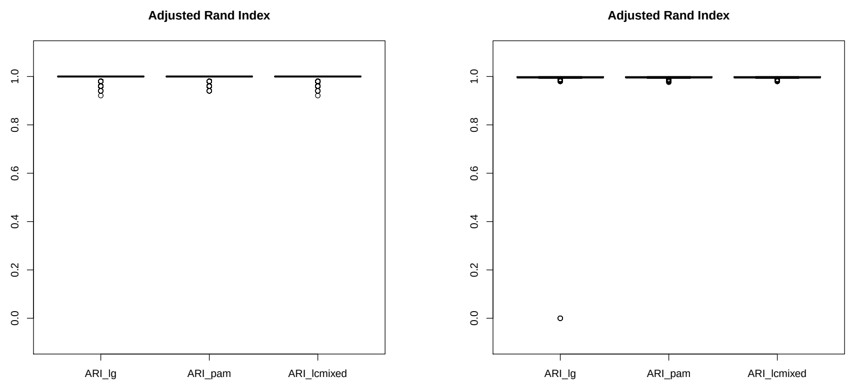
Latent class	π_t^X	$\pi_t^{\bar{L}_1X}$	$\pi_t^{\bar{L}_2X}$	$\pi_t^{\bar{L}_3X}$	$\pi_t^{\bar{L}_4X}$	$\pi_t^{\bar{L}_5X}$	$\pi_t^{\bar{L}_6X}$	$\pi_t^{\bar{L}_7X}$	$\pi_t^{\bar{L}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
2	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

Latent class	π_t^X	$\pi_t^{\bar{M}_1X}$	$\pi_t^{\bar{M}_2X}$	$\pi_t^{\bar{M}_3X}$	$\pi_t^{\bar{M}_4X}$	$\pi_t^{\bar{M}_5X}$	$\pi_t^{\bar{M}_6X}$	$\pi_t^{\bar{M}_7X}$	$\pi_t^{\bar{M}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	π_t^X	$\pi_t^{\bar{N}_1X}$	$\pi_t^{\bar{N}_2X}$	$\pi_t^{\bar{N}_3X}$	$\pi_t^{\bar{N}_4X}$	$\pi_t^{\bar{N}_5X}$	$\pi_t^{\bar{N}_6X}$	$\pi_t^{\bar{N}_7X}$	$\pi_t^{\bar{N}_8X}$
1	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.50	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Table A.87: Summary: 12 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.921	0.941	0.921
1st Qu.		1.000	1.000	1.000
Median		1.000	1.000	1.000
Mean (se)		0.995 (0.000)	0.995 (0.000)	0.995 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.348	0.348	0.347	0.348
1st Qu.	0.380	0.380	0.380	0.380
Median	0.389	0.389	0.389	0.389
Mean (se)	0.389 (0.000)	0.389 (0.000)	0.389 (0.000)	0.389 (0.000)
3rd Qu.	0.398	0.398	0.398	0.398
Max.	0.436	0.436	0.436	0.436
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.803	0.797	0.803	0.797
1st Qu.	0.834	0.834	0.834	0.834
Median	0.843	0.843	0.843	0.843
Mean (se)	0.842 (0.000)	0.842 (0.000)	0.842 (0.000)	0.842 (0.000)
3rd Qu.	0.850	0.850	0.850	0.850
Max.	0.875	0.875	0.875	0.875

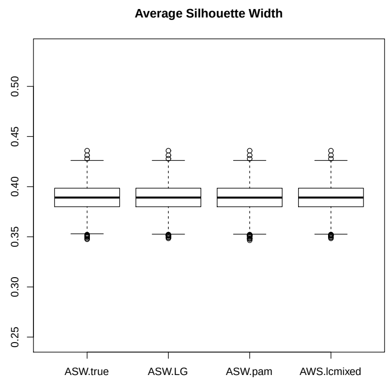


(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

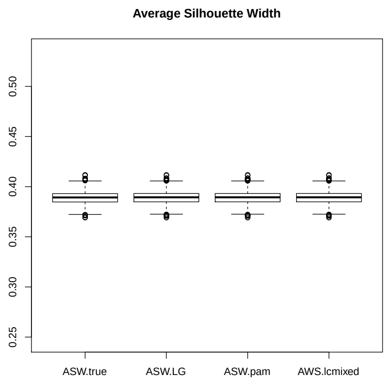
Figure A.130: Adjusted Rand Index: 12 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

Table A.88: Summary: 12 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.976	0.980
1st Qu.		0.996	0.996	0.996
Median		0.996	0.996	0.996
Mean (se)		0.994 (0.001)	0.994 (0.000)	0.996 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.369	0.369	0.369	0.369
1st Qu.	0.385	0.385	0.385	0.385
Median	0.389	0.389	0.389	0.389
Mean (se)	0.389 (0.000)	0.389 (0.000)	0.389 (0.000)	0.389 (0.000)
3rd Qu.	0.393	0.393	0.393	0.393
Max.	0.412	0.412	0.412	0.412
NA's		4		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.822	0.824	0.824	0.824
1st Qu.	0.838	0.838	0.838	0.838
Median	0.842	0.842	0.842	0.842
Mean (se)	0.842 (0.000)	0.842 (0.000)	0.842 (0.000)	0.842 (0.000)
3rd Qu.	0.846	0.846	0.846	0.846
Max.	0.860	0.860	0.860	0.860
NA's		4		

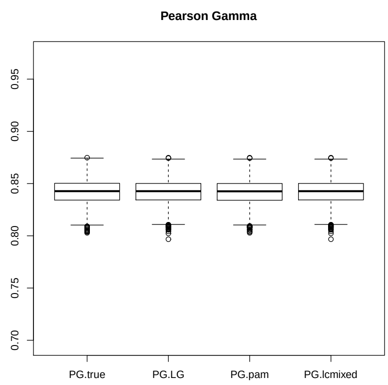


(a) *ASW - 200 units.*

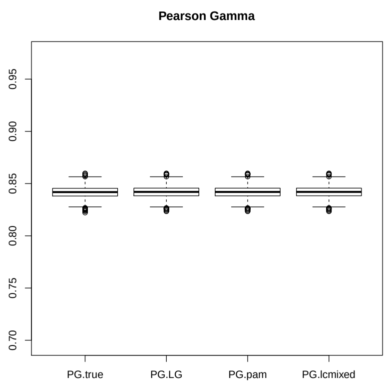


(b) *ASW - 1000 units.*

Figure A.131: Average Silhouette Width: 12 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 200 units.*



(b) *PG - 1000 units.*

Figure A.132: Pearson Gamma: 12 8-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

A.45 12 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1X}$	$\pi_t^{\bar{A}_2X}$	$\pi_t^{\bar{A}_3X}$	$\pi_t^{\bar{A}_4X}$	$\pi_t^{\bar{A}_5X}$	$\pi_t^{\bar{A}_6X}$	$\pi_t^{\bar{A}_7X}$	$\pi_t^{\bar{A}_8X}$
1	0.10	0.05	0.10	0.20	0.15	0.10	0.20	0.15	0.05
2	0.15	0.10	0.20	0.05	0.20	0.05	0.15	0.10	0.15
3	0.20	0.20	0.15	0.10	0.15	0.20	0.05	0.05	0.10
4	0.25	0.15	0.05	0.15	0.05	0.10	0.10	0.20	0.20
5	0.30	0.10	0.20	0.05	0.10	0.15	0.20	0.15	0.05

Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.10	0.15	0.30	0.05	0.10	0.15	0.05	0.10	0.10
2	0.15	0.10	0.05	0.20	0.15	0.10	0.20	0.05	0.15
3	0.20	0.20	0.10	0.10	0.05	0.05	0.10	0.30	0.10
4	0.25	0.05	0.15	0.10	0.20	0.20	0.15	0.10	0.05
5	0.30	0.15	0.15	0.05	0.15	0.10	0.20	0.15	0.05

Latent class	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.10	0.10	0.05	0.20	0.15	0.05	0.20	0.15	0.10
2	0.15	0.05	0.10	0.05	0.20	0.30	0.05	0.10	0.15
3	0.20	0.20	0.20	0.10	0.05	0.15	0.10	0.15	0.05
4	0.25	0.15	0.10	0.15	0.10	0.05	0.25	0.05	0.15
5	0.30	0.25	0.05	0.10	0.20	0.05	0.15	0.10	0.10

Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.10	0.20	0.05	0.10	0.05	0.15	0.15	0.10	0.20
2	0.15	0.05	0.10	0.05	0.20	0.10	0.15	0.20	0.15
3	0.20	0.15	0.20	0.20	0.10	0.10	0.05	0.15	0.05
4	0.25	0.20	0.15	0.10	0.15	0.05	0.20	0.05	0.10
5	0.30	0.05	0.20	0.05	0.10	0.20	0.10	0.15	0.15

Latent class	π_t^X	$\pi_t^{\bar{E}_1 X}$	$\pi_t^{\bar{E}_2 X}$	$\pi_t^{\bar{E}_3 X}$	$\pi_t^{\bar{E}_4 X}$	$\pi_t^{\bar{E}_5 X}$	$\pi_t^{\bar{E}_6 X}$	$\pi_t^{\bar{E}_7 X}$	$\pi_t^{\bar{E}_8 X}$
1	0.10	0.15	0.20	0.05	0.15	0.10	0.05	0.20	0.10
2	0.15	0.30	0.05	0.10	0.05	0.20	0.10	0.05	0.15
3	0.20	0.05	0.10	0.20	0.30	0.05	0.15	0.10	0.05
4	0.25	0.15	0.05	0.20	0.10	0.15	0.20	0.05	0.10
5	0.30	0.20	0.15	0.05	0.05	0.10	0.15	0.10	0.20

Latent class	π_t^X	$\pi_t^{\bar{F}_1 X}$	$\pi_t^{\bar{F}_2 X}$	$\pi_t^{\bar{F}_3 X}$	$\pi_t^{\bar{F}_4 X}$	$\pi_t^{\bar{F}_5 X}$	$\pi_t^{\bar{F}_6 X}$	$\pi_t^{\bar{F}_7 X}$	$\pi_t^{\bar{F}_8 X}$
1	0.10	0.20	0.05	0.10	0.15	0.10	0.05	0.20	0.15
2	0.15	0.10	0.15	0.20	0.05	0.05	0.10	0.15	0.20
3	0.20	0.05	0.20	0.05	0.10	0.15	0.20	0.10	0.15
4	0.25	0.05	0.20	0.15	0.10	0.20	0.15	0.05	0.10
5	0.30	0.15	0.10	0.05	0.20	0.15	0.20	0.10	0.05

Latent class	π_t^X	$\pi_t^{\bar{G}_1 X}$	$\pi_t^{\bar{G}_2 X}$	$\pi_t^{\bar{G}_3 X}$	$\pi_t^{\bar{G}_4 X}$	$\pi_t^{\bar{G}_5 X}$	$\pi_t^{\bar{G}_6 X}$	$\pi_t^{\bar{G}_7 X}$	$\pi_t^{\bar{G}_8 X}$
1	0.10	0.10	0.05	0.20	0.15	0.05	0.15	0.10	0.20
2	0.15	0.20	0.15	0.15	0.05	0.20	0.10	0.05	0.10
3	0.20	0.15	0.10	0.05	0.10	0.15	0.20	0.20	0.05
4	0.25	0.05	0.20	0.10	0.10	0.15	0.05	0.15	0.20
5	0.30	0.10	0.15	0.05	0.20	0.10	0.20	0.05	0.15

Latent class	π_t^X	$\pi_t^{\bar{H}_1 X}$	$\pi_t^{\bar{H}_2 X}$	$\pi_t^{\bar{H}_3 X}$	$\pi_t^{\bar{H}_4 X}$	$\pi_t^{\bar{H}_5 X}$	$\pi_t^{\bar{H}_6 X}$	$\pi_t^{\bar{H}_7 X}$	$\pi_t^{\bar{H}_8 X}$
1	0.10	0.15	0.15	0.05	0.20	0.10	0.05	0.10	0.20
2	0.15	0.20	0.05	0.10	0.10	0.15	0.20	0.15	0.05
3	0.20	0.05	0.10	0.30	0.05	0.20	0.10	0.05	0.15
4	0.25	0.30	0.20	0.05	0.10	0.05	0.15	0.05	0.10
5	0.30	0.10	0.10	0.25	0.15	0.05	0.10	0.20	0.05

Latent class	π_t^X	$\pi_t^{\bar{I}_1 X}$	$\pi_t^{\bar{I}_2 X}$	$\pi_t^{\bar{I}_3 X}$	$\pi_t^{\bar{I}_4 X}$	$\pi_t^{\bar{I}_5 X}$	$\pi_t^{\bar{I}_6 X}$	$\pi_t^{\bar{I}_7 X}$	$\pi_t^{\bar{I}_8 X}$
1	0.10	0.20	0.15	0.10	0.05	0.10	0.20	0.15	0.05
2	0.15	0.10	0.20	0.20	0.15	0.15	0.05	0.05	0.10
3	0.20	0.05	0.10	0.20	0.10	0.05	0.15	0.20	0.15
4	0.25	0.15	0.05	0.10	0.05	0.20	0.10	0.15	0.20
5	0.30	0.30	0.05	0.15	0.10	0.10	0.20	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{L}_1 X}$	$\pi_t^{\bar{L}_2 X}$	$\pi_t^{\bar{L}_3 X}$	$\pi_t^{\bar{L}_4 X}$	$\pi_t^{\bar{L}_5 X}$	$\pi_t^{\bar{L}_6 X}$	$\pi_t^{\bar{L}_7 X}$	$\pi_t^{\bar{L}_8 X}$
1	0.10	0.05	0.10	0.05	0.20	0.15	0.10	0.20	0.15
2	0.15	0.15	0.20	0.10	0.10	0.05	0.15	0.05	0.20
3	0.20	0.20	0.05	0.15	0.15	0.20	0.05	0.10	0.10
4	0.25	0.10	0.30	0.05	0.15	0.05	0.10	0.05	0.20
5	0.30	0.05	0.10	0.20	0.10	0.20	0.05	0.20	0.10

Latent class	π_t^X	$\pi_t^{\bar{M}_1 X}$	$\pi_t^{\bar{M}_2 X}$	$\pi_t^{\bar{M}_3 X}$	$\pi_t^{\bar{M}_4 X}$	$\pi_t^{\bar{M}_5 X}$	$\pi_t^{\bar{M}_6 X}$	$\pi_t^{\bar{M}_7 X}$	$\pi_t^{\bar{M}_8 X}$
1	0.10	0.10	0.20	0.05	0.15	0.20	0.05	0.15	0.10
2	0.15	0.05	0.10	0.20	0.10	0.15	0.20	0.05	0.15
3	0.20	0.20	0.15	0.10	0.05	0.10	0.15	0.05	0.20
4	0.25	0.15	0.05	0.15	0.20	0.05	0.10	0.20	0.10
5	0.30	0.10	0.20	0.20	0.05	0.15	0.15	0.10	0.05

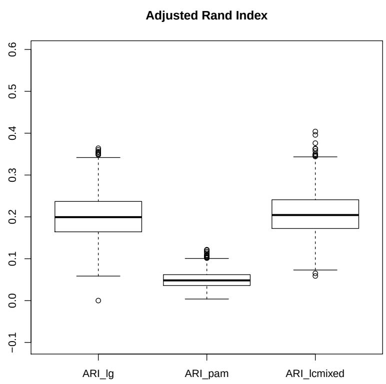
Latent class	π_t^X	$\pi_t^{\bar{N}_1 X}$	$\pi_t^{\bar{N}_2 X}$	$\pi_t^{\bar{N}_3 X}$	$\pi_t^{\bar{N}_4 X}$	$\pi_t^{\bar{N}_5 X}$	$\pi_t^{\bar{N}_6 X}$	$\pi_t^{\bar{N}_7 X}$	$\pi_t^{\bar{N}_8 X}$
1	0.10	0.20	0.05	0.10	0.20	0.15	0.10	0.15	0.05
2	0.15	0.10	0.15	0.05	0.05	0.10	0.20	0.15	0.20
3	0.20	0.15	0.10	0.20	0.10	0.05	0.05	0.20	0.15
4	0.25	0.05	0.30	0.10	0.15	0.10	0.15	0.05	0.10
5	0.30	0.10	0.20	0.15	0.05	0.05	0.30	0.05	0.10

Table A.89: Summary: 12 8-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 500 units

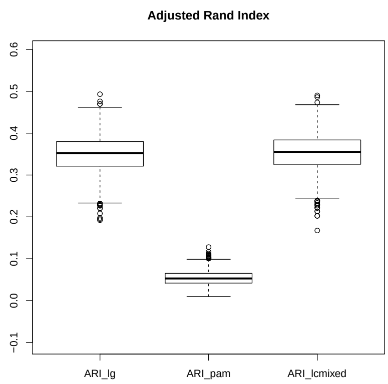
		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.004	0.059
1st Qu.		0.164	0.036	0.172
Median		0.199	0.048	0.204
Mean (se)		0.201 (0.001)	0.050 (0.000)	0.208 (0.001)
3rd Qu.		0.237	0.062	0.241
Max.		0.364	0.122	0.404
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.008	0.016	0.014	0.018
1st Qu.	0.013	0.025	0.018	0.025
Median	0.015	0.026	0.019	0.027
Mean (se)	0.015 (0.000)	0.026 (0.000)	0.019 (0.000)	0.027 (0.000)
3rd Qu.	0.016	0.027	0.021	0.028
Max.	0.021	0.034	0.024	0.036
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.125	0.128	0.123	0.140
1st Qu.	0.145	0.180	0.145	0.184
Median	0.151	0.187	0.151	0.190
Mean (se)	0.151 (0.000)	0.186 (0.000)	0.151 (0.000)	0.190 (0.000)
3rd Qu.	0.157	0.193	0.156	0.196
Max.	0.180	0.219	0.176	0.229
NA's		1		

Table A.90: Summary: 12 8-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.192	0.010	0.167
1st Qu.		0.321	0.042	0.326
Median		0.352	0.053	0.355
Mean (se)		0.350 (0.001)	0.054 (0.000)	0.353 (0.001)
3rd Qu.		0.380	0.065	0.384
Max.		0.493	0.128	0.490
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.012	0.017	0.015	0.016
1st Qu.	0.015	0.024	0.018	0.025
Median	0.016	0.026	0.019	0.026
Mean (se)	0.016 (0.000)	0.026 (0.000)	0.019 (0.000)	0.026 (0.000)
3rd Qu.	0.017	0.027	0.019	0.027
Max.	0.021	0.031	0.022	0.031
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.131	0.148	0.123	0.130
1st Qu.	0.147	0.180	0.138	0.181
Median	0.151	0.185	0.142	0.186
Mean (se)	0.151 (0.000)	0.185 (0.000)	0.142 (0.000)	0.186 (0.000)
3rd Qu.	0.155	0.191	0.146	0.191
Max.	0.172	0.208	0.161	0.209

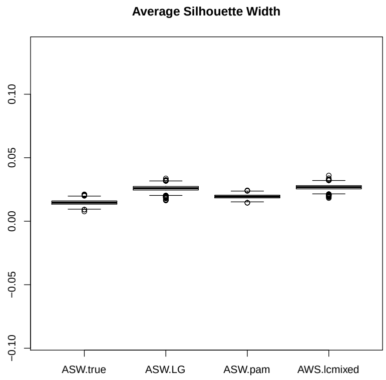


(a) *ARI - 500 units.*

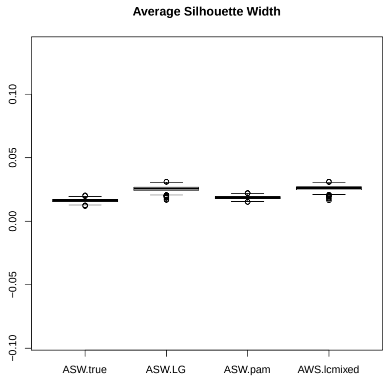


(b) *ARI - 1000 units.*

Figure A.133: Adjusted Rand Index: 12 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

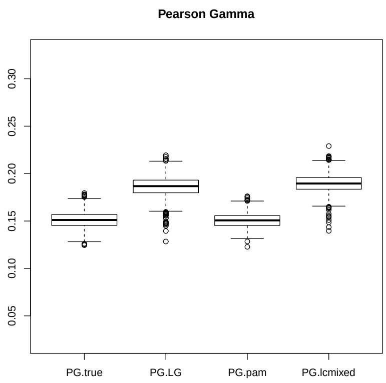


(a) *ASW - 500 units.*

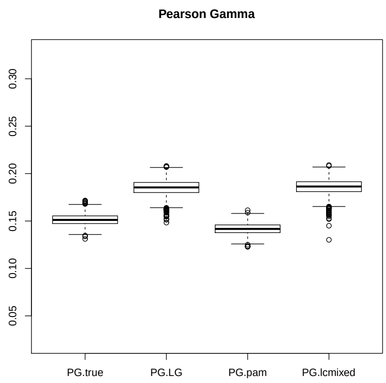


(b) *ASW - 1000 units.*

Figure A.134: Average Silhouette Width: 12 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 500 units.*



(b) *PG - 1000 units.*

Figure A.135: Pearson Gamma: 12 8-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

A.46 12 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1X}$	$\pi_t^{\bar{A}_2X}$	$\pi_t^{\bar{A}_3X}$	$\pi_t^{\bar{A}_4X}$	$\pi_t^{\bar{A}_5X}$	$\pi_t^{\bar{A}_6X}$	$\pi_t^{\bar{A}_7X}$	$\pi_t^{\bar{A}_8X}$
1	0.10	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
4	0.25	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{B}_1X}$	$\pi_t^{\bar{B}_2X}$	$\pi_t^{\bar{B}_3X}$	$\pi_t^{\bar{B}_4X}$	$\pi_t^{\bar{B}_5X}$	$\pi_t^{\bar{B}_6X}$	$\pi_t^{\bar{B}_7X}$	$\pi_t^{\bar{B}_8X}$
1	0.10	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
2	0.15	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
4	0.25	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent	π_t^X	$\pi_t^{\bar{C}_1X}$	$\pi_t^{\bar{C}_2X}$	$\pi_t^{\bar{C}_3X}$	$\pi_t^{\bar{C}_4X}$	$\pi_t^{\bar{C}_5X}$	$\pi_t^{\bar{C}_6X}$	$\pi_t^{\bar{C}_7X}$	$\pi_t^{\bar{C}_8X}$
1	0.10	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
4	0.25	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{D}_1X}$	$\pi_t^{\bar{D}_2X}$	$\pi_t^{\bar{D}_3X}$	$\pi_t^{\bar{D}_4X}$	$\pi_t^{\bar{D}_5X}$	$\pi_t^{\bar{D}_6X}$	$\pi_t^{\bar{D}_7X}$	$\pi_t^{\bar{D}_8X}$
1	0.10	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
3	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
4	0.25	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Latent class	π_t^{X}	$\pi_t^{\bar{E}_1 X}$	$\pi_t^{\bar{E}_2 X}$	$\pi_t^{\bar{E}_3 X}$	$\pi_t^{\bar{E}_4 X}$	$\pi_t^{\bar{E}_5 X}$	$\pi_t^{\bar{E}_6 X}$	$\pi_t^{\bar{E}_7 X}$	$\pi_t^{\bar{E}_8 X}$
1	0.10	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.15	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
4	0.25	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
5	0.30	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^{X}	$\pi_t^{\bar{F}_1 X}$	$\pi_t^{\bar{F}_2 X}$	$\pi_t^{\bar{F}_3 X}$	$\pi_t^{\bar{F}_4 X}$	$\pi_t^{\bar{F}_5 X}$	$\pi_t^{\bar{F}_6 X}$	$\pi_t^{\bar{F}_7 X}$	$\pi_t^{\bar{F}_8 X}$
1	0.10	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
4	0.25	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^{X}	$\pi_t^{\bar{G}_1 X}$	$\pi_t^{\bar{G}_2 X}$	$\pi_t^{\bar{G}_3 X}$	$\pi_t^{\bar{G}_4 X}$	$\pi_t^{\bar{G}_5 X}$	$\pi_t^{\bar{G}_6 X}$	$\pi_t^{\bar{G}_7 X}$	$\pi_t^{\bar{G}_8 X}$
1	0.10	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
4	0.25	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

Latent class	π_t^{X}	$\pi_t^{\bar{H}_1 X}$	$\pi_t^{\bar{H}_2 X}$	$\pi_t^{\bar{H}_3 X}$	$\pi_t^{\bar{H}_4 X}$	$\pi_t^{\bar{H}_5 X}$	$\pi_t^{\bar{H}_6 X}$	$\pi_t^{\bar{H}_7 X}$	$\pi_t^{\bar{H}_8 X}$
1	0.10	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
2	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
4	0.25	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
5	0.30	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{I}_1X}$	$\pi_t^{\bar{I}_2X}$	$\pi_t^{\bar{I}_3X}$	$\pi_t^{\bar{I}_4X}$	$\pi_t^{\bar{I}_5X}$	$\pi_t^{\bar{I}_6X}$	$\pi_t^{\bar{I}_7X}$	$\pi_t^{\bar{I}_8X}$
1	0.10	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
2	0.15	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
4	0.25	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

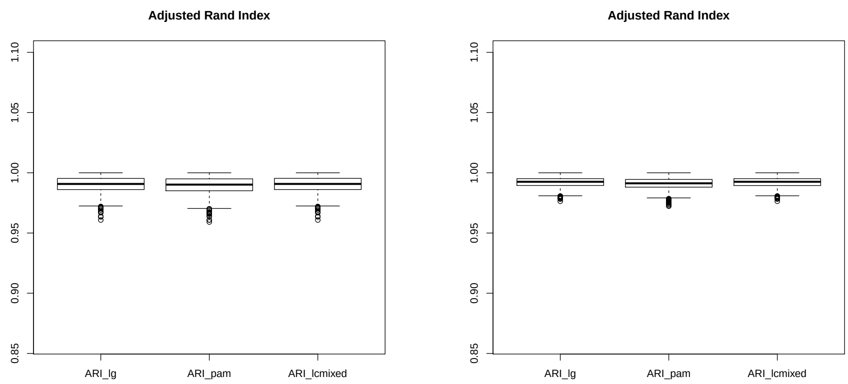
Latent class	π_t^X	$\pi_t^{\bar{L}_1X}$	$\pi_t^{\bar{L}_2X}$	$\pi_t^{\bar{L}_3X}$	$\pi_t^{\bar{L}_4X}$	$\pi_t^{\bar{L}_5X}$	$\pi_t^{\bar{L}_6X}$	$\pi_t^{\bar{L}_7X}$	$\pi_t^{\bar{L}_8X}$
1	0.10	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.15	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
4	0.25	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{M}_1X}$	$\pi_t^{\bar{M}_2X}$	$\pi_t^{\bar{M}_3X}$	$\pi_t^{\bar{M}_4X}$	$\pi_t^{\bar{M}_5X}$	$\pi_t^{\bar{M}_6X}$	$\pi_t^{\bar{M}_7X}$	$\pi_t^{\bar{M}_8X}$
1	0.10	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
4	0.25	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
5	0.30	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{N}_1X}$	$\pi_t^{\bar{N}_2X}$	$\pi_t^{\bar{N}_3X}$	$\pi_t^{\bar{N}_4X}$	$\pi_t^{\bar{N}_5X}$	$\pi_t^{\bar{N}_6X}$	$\pi_t^{\bar{N}_7X}$	$\pi_t^{\bar{N}_8X}$
1	0.10	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
4	0.25	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
5	0.30	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05

Table A.91: Summary: 12 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 500 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.961	0.959	0.961
1st Qu.		0.986	0.985	0.986
Median		0.991	0.990	0.991
Mean (se)		0.990 (0.000)	0.990 (0.000)	0.990 (0.000)
3rd Qu.		0.995	0.995	0.995
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.325	0.325	0.324	0.325
1st Qu.	0.352	0.352	0.351	0.352
Median	0.358	0.358	0.358	0.358
Mean (se)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)
3rd Qu.	0.364	0.364	0.364	0.364
Max.	0.386	0.387	0.387	0.387
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.816	0.816	0.816	0.816
1st Qu.	0.836	0.836	0.836	0.836
Median	0.841	0.841	0.841	0.841
Mean (se)	0.841 (0.000)	0.841 (0.000)	0.841 (0.000)	0.841 (0.000)
3rd Qu.	0.846	0.846	0.846	0.846
Max.	0.863	0.863	0.863	0.863

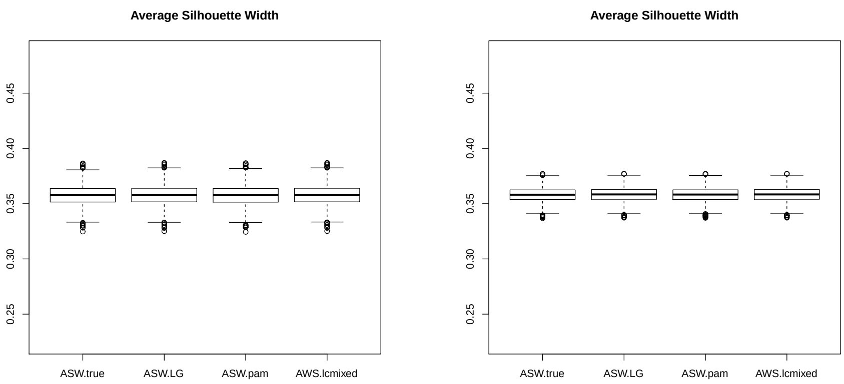


(a) *ARI - 500 units.* (b) *ARI - 1000 units.*

Figure A.136: Adjusted Rand Index: 12 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

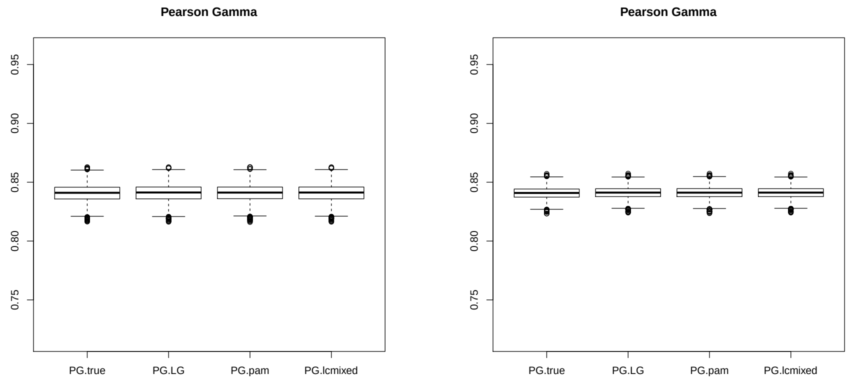
Table A.92: Summary: 12 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.976	0.972	0.976
1st Qu.		0.989	0.988	0.989
Median		0.993	0.991	0.993
Mean (se)		0.992 (0.000)	0.991 (0.000)	0.992 (0.000)
3rd Qu.		0.995	0.994	0.995
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.337	0.337	0.337	0.337
1st Qu.	0.354	0.354	0.354	0.354
Median	0.358	0.358	0.358	0.358
Mean (se)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)
3rd Qu.	0.362	0.363	0.362	0.363
Max.	0.377	0.377	0.377	0.377
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.824	0.824	0.824	0.824
1st Qu.	0.838	0.838	0.838	0.838
Median	0.841	0.841	0.841	0.841
Mean (se)	0.841 (0.000)	0.841 (0.000)	0.841 (0.000)	0.841 (0.000)
3rd Qu.	0.844	0.844	0.845	0.844
Max.	0.857	0.857	0.857	0.857



(a) *ASW - 500 units.* (b) *ASW - 1000 units.*

Figure A.137: Average Silhouette Width: 12 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 500 units.* (b) *PG - 1000 units.*

Figure A.138: Pearson Gamma: 12 8-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

A.47 12 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1 X}$	$\pi_t^{\bar{A}_2 X}$	$\pi_t^{\bar{A}_3 X}$	$\pi_t^{\bar{A}_4 X}$	$\pi_t^{\bar{A}_5 X}$	$\pi_t^{\bar{A}_6 X}$	$\pi_t^{\bar{A}_7 X}$	$\pi_t^{\bar{A}_8 X}$
1	0.20	0.10	0.05	0.15	0.05	0.10	0.20	0.20	0.15
2	0.20	0.20	0.10	0.05	0.20	0.15	0.15	0.05	0.10
3	0.20	0.15	0.20	0.10	0.15	0.05	0.05	0.10	0.20
4	0.20	0.05	0.15	0.20	0.10	0.20	0.10	0.15	0.05
5	0.20	0.10	0.05	0.15	0.05	0.15	0.20	0.10	0.20

Latent class	π_t^X	$\pi_t^{\bar{B}_1 X}$	$\pi_t^{\bar{B}_2 X}$	$\pi_t^{\bar{B}_3 X}$	$\pi_t^{\bar{B}_4 X}$	$\pi_t^{\bar{B}_5 X}$	$\pi_t^{\bar{B}_6 X}$	$\pi_t^{\bar{B}_7 X}$	$\pi_t^{\bar{B}_8 X}$
1	0.20	0.05	0.10	0.20	0.05	0.10	0.20	0.15	0.15
2	0.20	0.15	0.05	0.10	0.15	0.20	0.10	0.20	0.05
3	0.20	0.10	0.20	0.05	0.10	0.05	0.15	0.15	0.20
4	0.20	0.20	0.15	0.15	0.20	0.10	0.05	0.05	0.10
5	0.20	0.05	0.10	0.20	0.15	0.15	0.10	0.20	0.05

Latent class	π_t^X	$\pi_t^{\bar{C}_1 X}$	$\pi_t^{\bar{C}_2 X}$	$\pi_t^{\bar{C}_3 X}$	$\pi_t^{\bar{C}_4 X}$	$\pi_t^{\bar{C}_5 X}$	$\pi_t^{\bar{C}_6 X}$	$\pi_t^{\bar{C}_7 X}$	$\pi_t^{\bar{C}_8 X}$
1	0.20	0.10	0.05	0.20	0.15	0.10	0.05	0.15	0.20
2	0.20	0.20	0.10	0.10	0.05	0.15	0.15	0.20	0.05
3	0.20	0.05	0.15	0.05	0.10	0.20	0.10	0.20	0.15
4	0.20	0.15	0.20	0.15	0.20	0.05	0.10	0.05	0.10
5	0.20	0.10	0.05	0.10	0.15	0.15	0.20	0.05	0.20

Latent class	π_t^X	$\pi_t^{\bar{D}_1 X}$	$\pi_t^{\bar{D}_2 X}$	$\pi_t^{\bar{D}_3 X}$	$\pi_t^{\bar{D}_4 X}$	$\pi_t^{\bar{D}_5 X}$	$\pi_t^{\bar{D}_6 X}$	$\pi_t^{\bar{D}_7 X}$	$\pi_t^{\bar{D}_8 X}$
1	0.20	0.10	0.15	0.05	0.10	0.20	0.15	0.05	0.10
2	0.20	0.20	0.05	0.10	0.15	0.10	0.05	0.20	0.15
3	0.20	0.15	0.20	0.15	0.05	0.05	0.20	0.10	0.10
4	0.20	0.05	0.10	0.20	0.15	0.10	0.15	0.05	0.20
5	0.20	0.05	0.15	0.05	0.20	0.20	0.10	0.10	0.15

Latent class	π_t^X	$\pi_t^{\bar{E}_1X}$	$\pi_t^{\bar{E}_2X}$	$\pi_t^{\bar{E}_3X}$	$\pi_t^{\bar{E}_4X}$	$\pi_t^{\bar{E}_5X}$	$\pi_t^{\bar{E}_6X}$	$\pi_t^{\bar{E}_7X}$	$\pi_t^{\bar{E}_8X}$
1	0.20	0.20	0.05	0.10	0.15	0.20	0.05	0.05	0.10
2	0.20	0.05	0.15	0.20	0.05	0.10	0.20	0.20	0.15
3	0.20	0.15	0.20	0.05	0.20	0.15	0.10	0.10	0.10
4	0.20	0.10	0.15	0.20	0.10	0.05	0.05	0.05	0.05
5	0.20	0.15	0.10	0.15	0.05	0.20	0.10	0.10	0.20

Latent class	π_t^X	$\pi_t^{\bar{F}_1X}$	$\pi_t^{\bar{F}_2X}$	$\pi_t^{\bar{F}_3X}$	$\pi_t^{\bar{F}_4X}$	$\pi_t^{\bar{F}_5X}$	$\pi_t^{\bar{F}_6X}$	$\pi_t^{\bar{F}_7X}$	$\pi_t^{\bar{F}_8X}$
1	0.20	0.15	0.20	0.05	0.10	0.05	0.15	0.20	0.10
2	0.20	0.20	0.15	0.05	0.15	0.20	0.10	0.10	0.05
3	0.20	0.10	0.05	0.10	0.20	0.15	0.20	0.05	0.15
4	0.20	0.05	0.10	0.15	0.05	0.20	0.10	0.15	0.20
5	0.20	0.15	0.20	0.20	0.10	0.05	0.05	0.10	0.15

Latent class	π_t^X	$\pi_t^{\bar{G}_1X}$	$\pi_t^{\bar{G}_2X}$	$\pi_t^{\bar{G}_3X}$	$\pi_t^{\bar{G}_4X}$	$\pi_t^{\bar{G}_5X}$	$\pi_t^{\bar{G}_6X}$	$\pi_t^{\bar{G}_7X}$	$\pi_t^{\bar{G}_8X}$
1	0.20	0.05	0.10	0.05	0.20	0.15	0.20	0.15	0.10
2	0.20	0.10	0.20	0.10	0.15	0.05	0.15	0.20	0.05
3	0.20	0.20	0.05	0.15	0.05	0.20	0.10	0.10	0.15
4	0.20	0.15	0.10	0.20	0.10	0.15	0.20	0.05	0.05
5	0.20	0.10	0.15	0.05	0.20	0.10	0.05	0.15	0.20

Latent class	π_t^X	$\pi_t^{\bar{H}_1X}$	$\pi_t^{\bar{H}_2X}$	$\pi_t^{\bar{H}_3X}$	$\pi_t^{\bar{H}_4X}$	$\pi_t^{\bar{H}_5X}$	$\pi_t^{\bar{H}_6X}$	$\pi_t^{\bar{H}_7X}$	$\pi_t^{\bar{H}_8X}$
1	0.20	0.10	0.05	0.20	0.15	0.05	0.20	0.10	0.15
2	0.20	0.05	0.20	0.15	0.10	0.15	0.10	0.05	0.20
3	0.20	0.20	0.10	0.10	0.05	0.15	0.15	0.20	0.05
4	0.20	0.15	0.05	0.05	0.20	0.10	0.20	0.15	0.10
5	0.20	0.05	0.15	0.10	0.15	0.20	0.05	0.10	0.20

Latent class	π_t^X	$\pi_t^{\bar{I}_1 X}$	$\pi_t^{\bar{I}_2 X}$	$\pi_t^{\bar{I}_3 X}$	$\pi_t^{\bar{I}_4 X}$	$\pi_t^{\bar{I}_5 X}$	$\pi_t^{\bar{I}_6 X}$	$\pi_t^{\bar{I}_7 X}$	$\pi_t^{\bar{I}_8 X}$
1	0.20	0.20	0.05	0.20	0.10	0.15	0.05	0.10	0.15
2	0.20	0.10	0.20	0.05	0.15	0.05	0.10	0.15	0.20
3	0.20	0.15	0.10	0.15	0.20	0.10	0.05	0.20	0.05
4	0.20	0.05	0.15	0.10	0.05	0.20	0.20	0.10	0.15
5	0.20	0.10	0.05	0.20	0.10	0.15	0.15	0.20	0.05

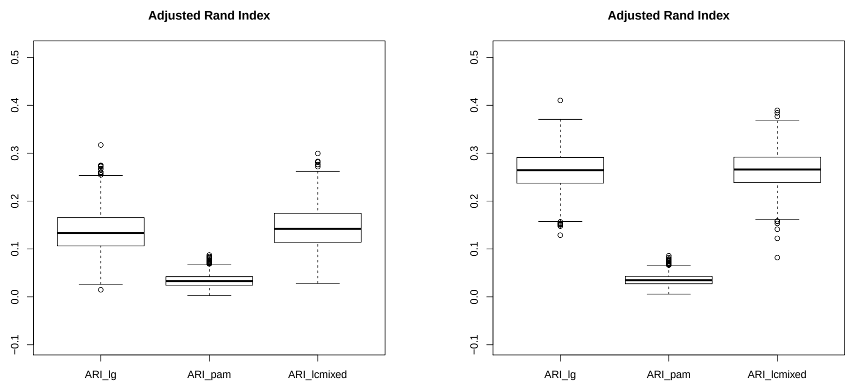
Latent class	π_t^X	$\pi_t^{\bar{L}_1 X}$	$\pi_t^{\bar{L}_2 X}$	$\pi_t^{\bar{L}_3 X}$	$\pi_t^{\bar{L}_4 X}$	$\pi_t^{\bar{L}_5 X}$	$\pi_t^{\bar{L}_6 X}$	$\pi_t^{\bar{L}_7 X}$	$\pi_t^{\bar{L}_8 X}$
1	0.20	0.15	0.05	0.10	0.20	0.10	0.05	0.15	0.20
2	0.20	0.20	0.10	0.05	0.15	0.20	0.15	0.10	0.05
3	0.20	0.10	0.20	0.15	0.10	0.05	0.20	0.05	0.15
4	0.20	0.05	0.15	0.20	0.20	0.10	0.05	0.15	0.10
5	0.20	0.10	0.05	0.15	0.05	0.20	0.10	0.20	0.15

Latent class	π_t^X	$\pi_t^{\bar{M}_1 X}$	$\pi_t^{\bar{M}_2 X}$	$\pi_t^{\bar{M}_3 X}$	$\pi_t^{\bar{M}_4 X}$	$\pi_t^{\bar{M}_5 X}$	$\pi_t^{\bar{M}_6 X}$	$\pi_t^{\bar{M}_7 X}$	$\pi_t^{\bar{M}_8 X}$
1	0.20	0.10	0.05	0.20	0.15	0.10	0.15	0.20	0.05
2	0.20	0.20	0.10	0.15	0.05	0.20	0.10	0.05	0.15
3	0.20	0.05	0.20	0.10	0.10	0.15	0.20	0.15	0.05
4	0.20	0.15	0.15	0.05	0.20	0.10	0.05	0.10	0.20
5	0.20	0.10	0.05	0.20	0.05	0.20	0.15	0.10	0.15

Latent class	π_t^X	$\pi_t^{\bar{N}_1 X}$	$\pi_t^{\bar{N}_2 X}$	$\pi_t^{\bar{N}_3 X}$	$\pi_t^{\bar{N}_4 X}$	$\pi_t^{\bar{N}_5 X}$	$\pi_t^{\bar{N}_6 X}$	$\pi_t^{\bar{N}_7 X}$	$\pi_t^{\bar{N}_8 X}$
1	0.20	0.05	0.05	0.20	0.15	0.05	0.15	0.20	0.10
2	0.20	0.05	0.20	0.15	0.10	0.15	0.20	0.10	0.05
3	0.20	0.05	0.10	0.05	0.20	0.20	0.05	0.15	0.10
4	0.20	0.05	0.15	0.20	0.05	0.10	0.15	0.05	0.20
5	0.20	0.05	0.05	0.10	0.15	0.05	0.10	0.20	0.15

Table A.93: Summary: 12 8-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 500 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.015	0.003	0.028
1st Qu.		0.106	0.024	0.114
Median		0.133	0.033	0.142
Mean (sde)		0.137 (0.001)	0.034 (0.000)	0.145 (0.001)
3rd Qu.		0.165	0.042	0.174
Max.		0.317	0.087	0.299
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.006	0.013	0.015	0.017
1st Qu.	0.009	0.023	0.018	0.024
Median	0.010	0.024	0.019	0.025
Mean (se)	0.010 (0.000)	0.024 (0.000)	0.019 (0.000)	0.025 (0.000)
3rd Qu.	0.012	0.026	0.019	0.026
Max.	0.016	0.030	0.025	0.031
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.100	0.124	0.125	0.137
1st Qu.	0.119	0.170	0.140	0.174
Median	0.124	0.176	0.144	0.179
Mean (se)	0.124 (0.000)	0.175 (0.000)	0.144 (0.000)	0.179 (0.000)
3rd Qu.	0.129	0.182	0.148	0.184
Max.	0.145	0.200	0.169	0.205

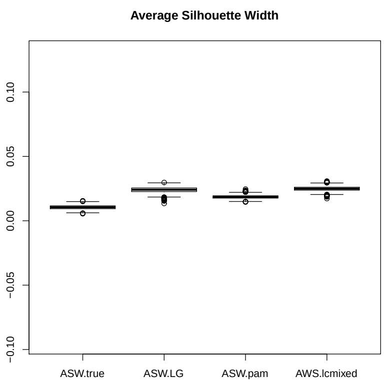


(a) *ARI - 500 units.* (b) *ARI - 1000 units.*

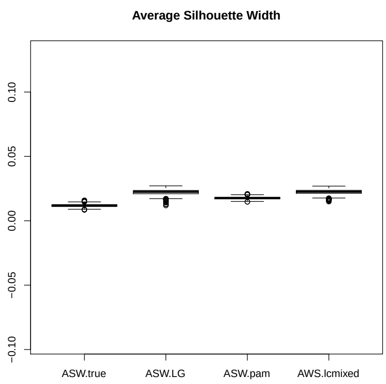
Figure A.139: Adjusted Rand Index: 12 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

Table A.94: Summary: 12 8-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.129	0.006	0.082
1st Qu.		0.238	0.027	0.239
Median		0.264	0.035	0.266
Mean (se)		0.263 (0.001)	0.036 (0.000)	0.265 (0.001)
3rd Qu.		0.291	0.043	0.292
Max.		0.410	0.086	0.389
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.008	0.012	0.015	0.015
1st Qu.	0.011	0.021	0.017	0.021
Median	0.012	0.022	0.018	0.023
Mean (se)	0.012 (0.000)	0.022 (0.000)	0.018 (0.000)	0.022 (0.000)
3rd Qu.	0.013	0.024	0.018	0.024
Max.	0.016	0.027	0.021	0.027
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.108	0.122	0.120	0.132
1st Qu.	0.120	0.161	0.130	0.162
Median	0.124	0.166	0.133	0.167
Mean (se)	0.124 (0.000)	0.165 (0.000)	0.134 (0.000)	0.166 (0.000)
3rd Qu.	0.127	0.170	0.137	0.171
Max.	0.143	0.184	0.154	0.184

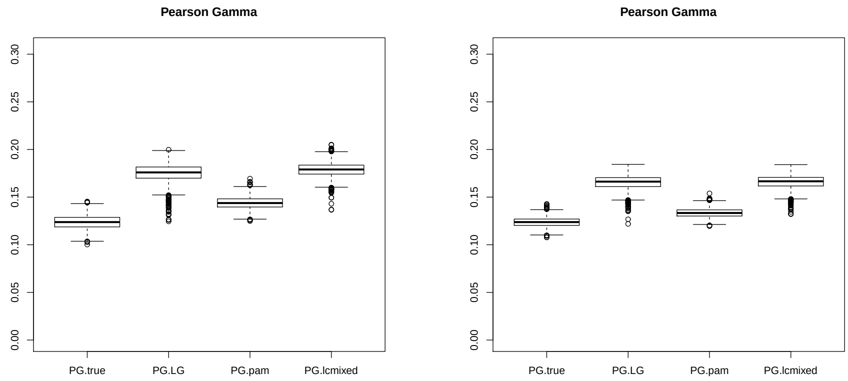


(a) *ASW - 500 units.*



(b) *ASW - 1000 units.*

Figure A.140: Average Silhouette Width: 12 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 500 units.* (b) *PG - 1000 units.*

Figure A.141: Pearson Gamma: 12 8-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

A.48 12 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

The Model:

Latent class	π_t^X	$\pi_t^{\bar{A}_1 X}$	$\pi_t^{\bar{A}_2 X}$	$\pi_t^{\bar{A}_3 X}$	$\pi_t^{\bar{A}_4 X}$	$\pi_t^{\bar{A}_5 X}$	$\pi_t^{\bar{A}_6 X}$	$\pi_t^{\bar{A}_7 X}$	$\pi_t^{\bar{A}_8 X}$
1	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
2	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
4	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	π_t^X	$\pi_t^{\bar{B}_1 X}$	$\pi_t^{\bar{B}_2 X}$	$\pi_t^{\bar{B}_3 X}$	$\pi_t^{\bar{B}_4 X}$	$\pi_t^{\bar{B}_5 X}$	$\pi_t^{\bar{B}_6 X}$	$\pi_t^{\bar{B}_7 X}$	$\pi_t^{\bar{B}_8 X}$
1	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent	π_t^X	$\pi_t^{\bar{C}_1 X}$	$\pi_t^{\bar{C}_2 X}$	$\pi_t^{\bar{C}_3 X}$	$\pi_t^{\bar{C}_4 X}$	$\pi_t^{\bar{C}_5 X}$	$\pi_t^{\bar{C}_6 X}$	$\pi_t^{\bar{C}_7 X}$	$\pi_t^{\bar{C}_8 X}$
1	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	π_t^X	$\pi_t^{\bar{D}_1 X}$	$\pi_t^{\bar{D}_2 X}$	$\pi_t^{\bar{D}_3 X}$	$\pi_t^{\bar{D}_4 X}$	$\pi_t^{\bar{D}_5 X}$	$\pi_t^{\bar{D}_6 X}$	$\pi_t^{\bar{D}_7 X}$	$\pi_t^{\bar{D}_8 X}$
1	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
2	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	π_t^X	$\pi_t^{\bar{E}_1X}$	$\pi_t^{\bar{E}_2X}$	$\pi_t^{\bar{E}_3X}$	$\pi_t^{\bar{E}_4X}$	$\pi_t^{\bar{E}_5X}$	$\pi_t^{\bar{E}_6X}$	$\pi_t^{\bar{E}_7X}$	$\pi_t^{\bar{E}_8X}$
1	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
4	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
5	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{F}_1X}$	$\pi_t^{\bar{F}_2X}$	$\pi_t^{\bar{F}_3X}$	$\pi_t^{\bar{F}_4X}$	$\pi_t^{\bar{F}_5X}$	$\pi_t^{\bar{F}_6X}$	$\pi_t^{\bar{F}_7X}$	$\pi_t^{\bar{F}_8X}$
1	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
2	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
3	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
4	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{G}_1X}$	$\pi_t^{\bar{G}_2X}$	$\pi_t^{\bar{G}_3X}$	$\pi_t^{\bar{G}_4X}$	$\pi_t^{\bar{G}_5X}$	$\pi_t^{\bar{G}_6X}$	$\pi_t^{\bar{G}_7X}$	$\pi_t^{\bar{G}_8X}$
1	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
4	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{H}_1X}$	$\pi_t^{\bar{H}_2X}$	$\pi_t^{\bar{H}_3X}$	$\pi_t^{\bar{H}_4X}$	$\pi_t^{\bar{H}_5X}$	$\pi_t^{\bar{H}_6X}$	$\pi_t^{\bar{H}_7X}$	$\pi_t^{\bar{H}_8X}$
1	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
2	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
3	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
5	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{I}_1 X}$	$\pi_t^{\bar{I}_2 X}$	$\pi_t^{\bar{I}_3 X}$	$\pi_t^{\bar{I}_4 X}$	$\pi_t^{\bar{I}_5 X}$	$\pi_t^{\bar{I}_6 X}$	$\pi_t^{\bar{I}_7 X}$	$\pi_t^{\bar{I}_8 X}$
1	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
3	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
4	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

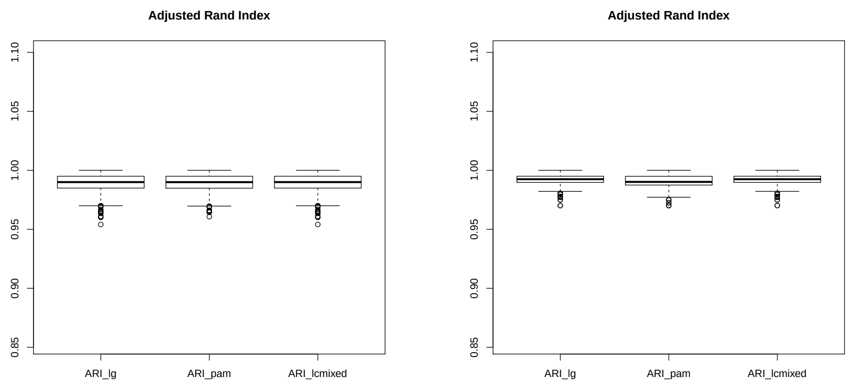
Latent class	π_t^X	$\pi_t^{\bar{L}_1 X}$	$\pi_t^{\bar{L}_2 X}$	$\pi_t^{\bar{L}_3 X}$	$\pi_t^{\bar{L}_4 X}$	$\pi_t^{\bar{L}_5 X}$	$\pi_t^{\bar{L}_6 X}$	$\pi_t^{\bar{L}_7 X}$	$\pi_t^{\bar{L}_8 X}$
1	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
2	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	π_t^X	$\pi_t^{\bar{M}_1 X}$	$\pi_t^{\bar{M}_2 X}$	$\pi_t^{\bar{M}_3 X}$	$\pi_t^{\bar{M}_4 X}$	$\pi_t^{\bar{M}_5 X}$	$\pi_t^{\bar{M}_6 X}$	$\pi_t^{\bar{M}_7 X}$	$\pi_t^{\bar{M}_8 X}$
1	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
2	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
3	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Latent class	π_t^X	$\pi_t^{\bar{N}_1 X}$	$\pi_t^{\bar{N}_2 X}$	$\pi_t^{\bar{N}_3 X}$	$\pi_t^{\bar{N}_4 X}$	$\pi_t^{\bar{N}_5 X}$	$\pi_t^{\bar{N}_6 X}$	$\pi_t^{\bar{N}_7 X}$	$\pi_t^{\bar{N}_8 X}$
1	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
3	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
5	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05

Table A.95: Summary: 12 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 500 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.954	0.961	0.954
1st Qu.		0.985	0.985	0.985
Median		0.990	0.990	0.990
Mean (se)		0.989 (0.000)	0.989 (0.000)	0.989 (0.000)
3rd Qu.		0.995	0.995	0.995
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.324	0.323	0.324	0.323
1st Qu.	0.351	0.352	0.351	0.352
Median	0.358	0.358	0.358	0.358
Mean (se)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)
3rd Qu.	0.364	0.364	0.364	0.364
Max.	0.389	0.390	0.389	0.390
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.807	0.808	0.807	0.808
1st Qu.	0.830	0.830	0.830	0.830
Median	0.835	0.835	0.835	0.835
Mean (se)	0.835 (0.000)	0.835 (0.000)	0.835 (0.000)	0.835 (0.000)
3rd Qu.	0.840	0.840	0.840	0.840
Max.	0.860	0.862	0.861	0.862

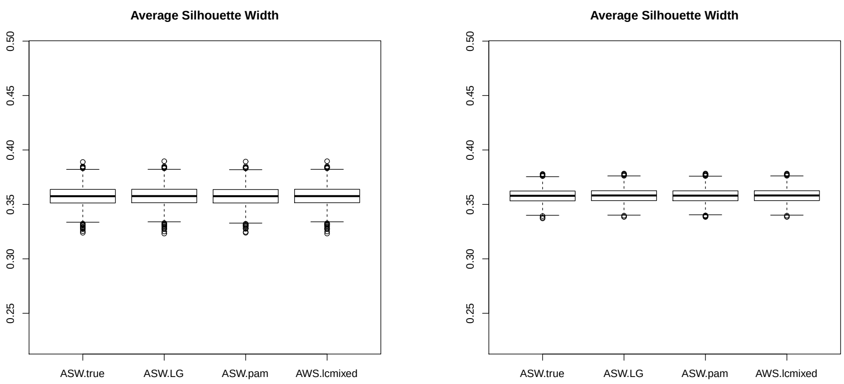


(a) *ARI - 500 units.* (b) *ARI - 1000 units.*

Figure A.142: Adjusted Rand Index: 12 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

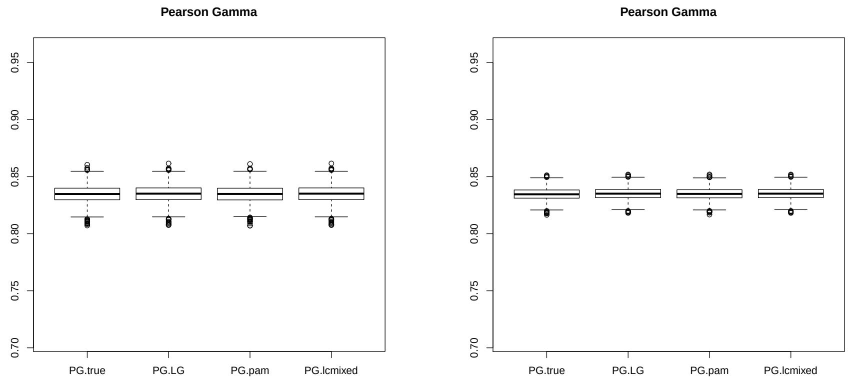
Table A.96: Summary: 12 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.970	0.970	0.970
1st Qu.		0.990	0.988	0.990
Median		0.992	0.990	0.992
Mean (se)		0.992 (0.000)	0.991 (0.000)	0.992 (0.000)
3rd Qu.		0.995	0.995	0.995
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.337	0.339	0.338	0.339
1st Qu.	0.353	0.354	0.353	0.354
Median	0.358	0.358	0.358	0.358
Mean (se)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)	0.358 (0.000)
3rd Qu.	0.362	0.363	0.362	0.363
Max.	0.378	0.378	0.378	0.378
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.816	0.818	0.817	0.818
1st Qu.	0.831	0.832	0.832	0.832
Median	0.835	0.835	0.835	0.835
Mean (se)	0.835 (0.000)	0.835 (0.000)	0.835 (0.000)	0.835 (0.000)
3rd Qu.	0.838	0.839	0.839	0.839
Max.	0.851	0.852	0.852	0.852



(a) *ASW - 500 units.* (b) *ASW - 1000 units.*

Figure A.143: Average Silhouette Width: 12 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 500 units.* (b) *PG - 1000 units.*

Figure A.144: Pearson Gamma: 12 8-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

A.49 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$
1	0.20	0.60	0.40	0.30	0.40	0.30	0.20	0.30	0.50
2	0.80	0.70	0.30	0.20	0.50	0.30	0.30	0.50	0.20

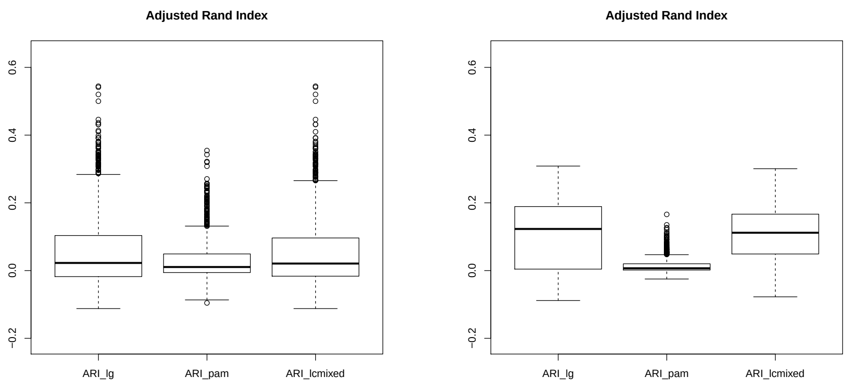
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.20	0.20	0.10	0.30	0.40
2	0.80	0.30	0.20	0.40	0.10

A.50 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Table A.97: Summary: 4 mixed-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.112	-0.096	-0.112
1st Qu.		-0.018	-0.006	-0.017
Median		0.023	0.011	0.021
Mean (se)		0.052 (0.002)	0.028 (0.001)	0.049 (0.002)
3rd Qu.		0.104	0.049	0.096
Max.		0.544	0.354	0.544
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.019	0.035	0.140	0.027
1st Qu.	0.045	0.186	0.198	0.187
Median	0.065	0.212	0.219	0.213
Mean (se)	0.067 (0.001)	0.214 (0.001)	0.222 (0.001)	0.216 (0.001)
3rd Qu.	0.086	0.242	0.241	0.243
Max.	0.179	0.381	0.363	0.374
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.020	-0.064	0.175	-0.068
1st Qu.	0.062	0.228	0.288	0.239
Median	0.091	0.286	0.329	0.291
Mean (se)	0.094 (0.001)	0.284 (0.002)	0.334 (0.001)	0.292 (0.002)
3rd Qu.	0.122	0.344	0.372	0.349
Max.	0.265	0.588	0.546	0.586

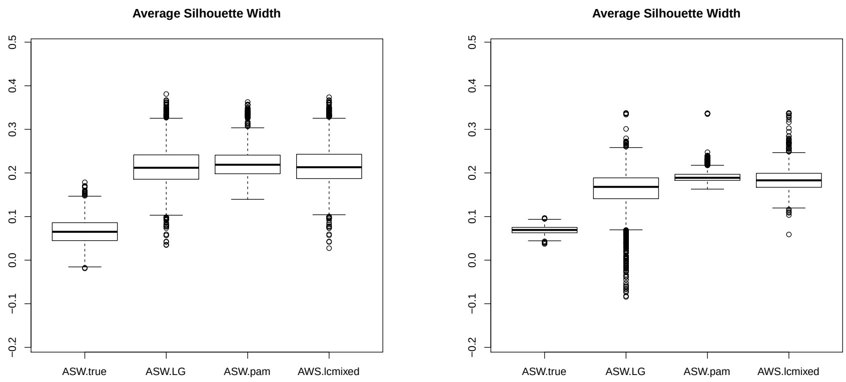


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.145: Adjusted Rand Index: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

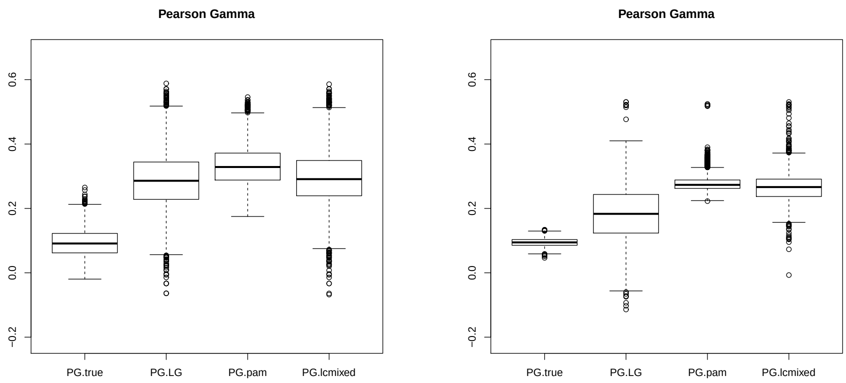
Table A.98: Summary: 4 mixed-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.088	-0.025	-0.078
1st Qu.		0.004	0.002	0.049
Median		0.123	0.007	0.112
Mean (se)		0.105 (0.002)	0.014 (0.000)	0.109 (0.002)
3rd Qu.		0.189	0.020	0.167
Max.		0.309	0.166	0.301
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.037	-0.084	0.163	0.059
1st Qu.	0.063	0.141	0.183	0.167
Median	0.069	0.168	0.189	0.183
Mean (se)	0.069 (0.000)	0.160 (0.001)	0.192 (0.000)	0.185 (0.001)
3rd Qu.	0.075	0.189	0.197	0.199
Max.	0.097	0.338	0.337	0.338
NA's		29		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.046	-0.114	0.223	-0.007
1st Qu.	0.085	0.123	0.262	0.237
Median	0.094	0.183	0.273	0.266
Mean (se)	0.095 (0.000)	0.182 (0.002)	0.282 (0.001)	0.266 (0.001)
3rd Qu.	0.103	0.244	0.288	0.291
Max.	0.134	0.531	0.524	0.531
NA's		29		



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.146: Average Silhouette Width: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.147: Pearson Gamma: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

Table A.99: Summary: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.257	0.230	0.257
1st Qu.		0.804	0.740	0.801
Median		0.867	0.802	0.866
Mean (se)		0.851 (0.002)	0.802 (0.002)	0.850 (0.002)
3rd Qu.		0.913	0.871	0.913
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.444	0.364	0.433	0.364
1st Qu.	0.540	0.557	0.549	0.557
Median	0.571	0.585	0.577	0.586
Mean (se)	0.570 (0.001)	0.584 (0.001)	0.577 (0.001)	0.584 (0.001)
3rd Qu.	0.599	0.612	0.606	0.612
Max.	0.723	0.721	0.721	0.721
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.545	0.480	0.458	0.480
1st Qu.	0.694	0.718	0.728	0.718
Median	0.728	0.748	0.754	0.749
Mean (se)	0.725 (0.001)	0.745 (0.001)	0.753 (0.001)	0.746 (0.001)
3rd Qu.	0.760	0.778	0.782	0.778
Max.	0.874	0.874	0.871	0.874

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$
1	0.20	0.90	0.10	0.10	0.80	0.10	0.80	0.10	0.10
2	0.80	0.10	0.90	0.80	0.10	0.10	0.10	0.80	0.10

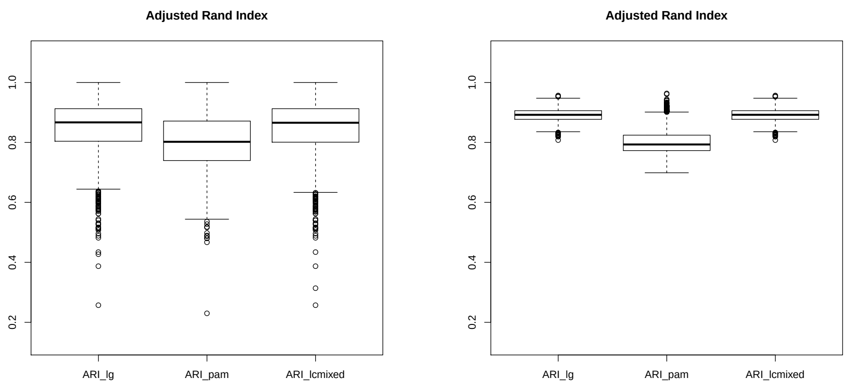
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.20	0.10	0.10	0.10	0.70
2	0.80	0.10	0.10	0.70	0.10

A.51 4 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

The Model:

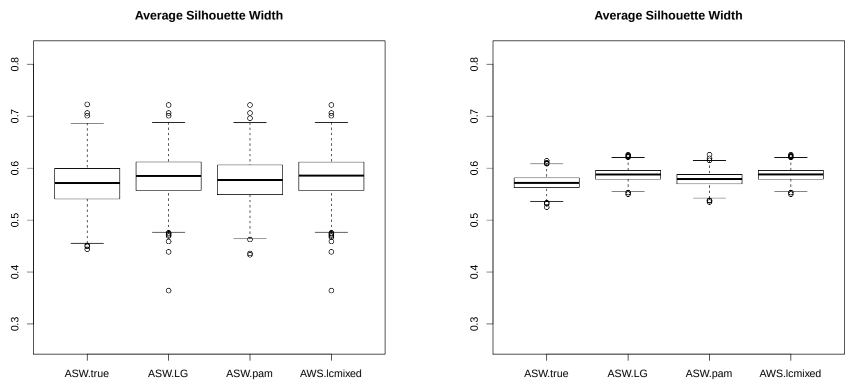
Table A.100: Summary: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.808	0.699	0.808
1st Qu.		0.877	0.773	0.877
Median		0.892	0.793	0.892
Mean (se)		0.891 (0.000)	0.806 (0.001)	0.891 (0.000)
3rd Qu.		0.906	0.824	0.906
Max.		0.957	0.963	0.957
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.525	0.550	0.535	0.550
1st Qu.	0.563	0.579	0.570	0.579
Median	0.572	0.588	0.579	0.588
Mean (se)	0.572 (0.000)	0.587 (0.000)	0.578 (0.000)	0.587 (0.000)
3rd Qu.	0.581	0.596	0.588	0.596
Max.	0.614	0.626	0.626	0.626
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.670	0.687	0.712	0.687
1st Qu.	0.718	0.741	0.747	0.742
Median	0.728	0.751	0.755	0.751
Mean (se)	0.728 (0.000)	0.750 (0.000)	0.754 (0.000)	0.750 (0.000)
3rd Qu.	0.738	0.760	0.762	0.760
Max.	0.770	0.794	0.794	0.794



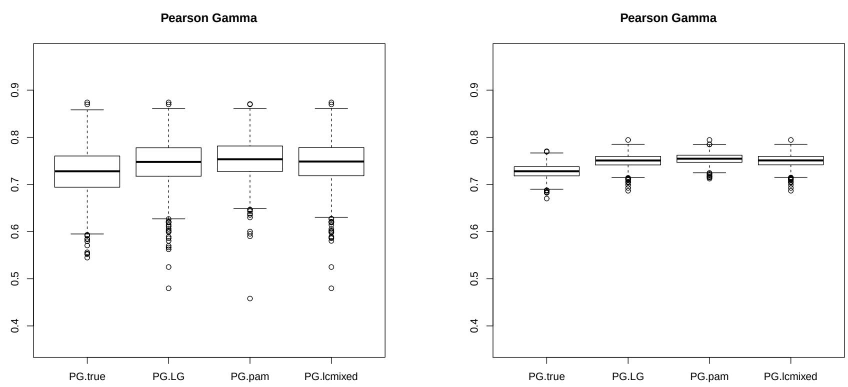
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.148: Adjusted Rand Index: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.149: Average Silhouette Width: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.150: Pearson Gamma: 4 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Table A.101: Summary: 4 mixed-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.037	-0.021	-0.033
1st Qu.		-0.006	-0.008	-0.006
Median		0.004	0.000	0.004
Mean (se)		0.016 (0.001)	0.011 (0.001)	0.017 (0.001)
3rd Qu.		0.024	0.019	0.026
Max.		0.263	0.263	0.285
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.012	0.026	0.134	0.034
1st Qu.	0.015	0.173	0.196	0.178
Median	0.026	0.198	0.214	0.201
Mean (se)	0.028 (0.000)	0.202 (0.001)	0.215 (0.001)	0.206 (0.001)
3rd Qu.	0.039	0.224	0.232	0.228
Max.	0.105	0.374	0.313	0.374
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.017	-0.030	0.178	-0.003
1st Qu.	0.020	0.208	0.297	0.225
Median	0.038	0.271	0.328	0.283
Mean (se)	0.040 (0.001)	0.272 (0.002)	0.328 (0.001)	0.285 (0.002)
3rd Qu.	0.056	0.322	0.360	0.332
Max.	0.158	0.595	0.504	0.595

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$
1	0.50	0.65	0.35	0.40	0.40	0.20	0.35	0.25	0.40
2	0.50	0.55	0.45	0.30	0.60	0.10	0.25	0.40	0.35

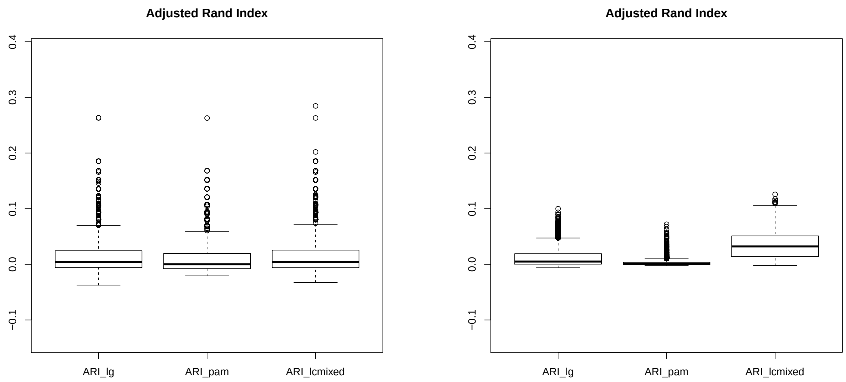
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.50	0.20	0.30	0.25	0.25
2	0.50	0.30	0.15	0.25	0.30

A.52 4 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

The Model:

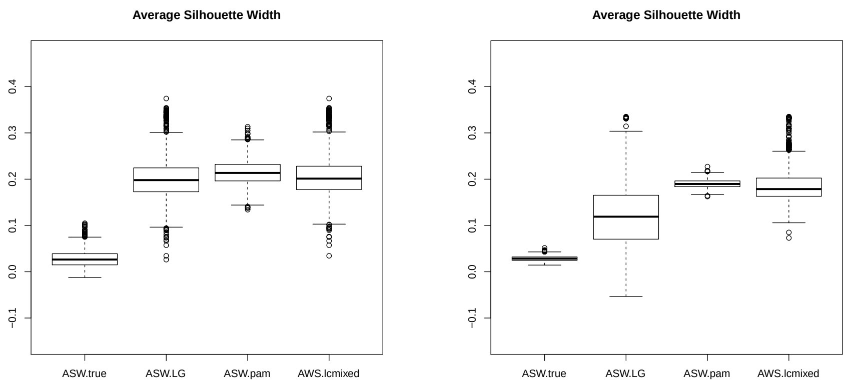
Table A.102: Summary: 4 mixed-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.006	-0.002	-0.002
1st Qu.		0.000	-0.001	0.014
Median		0.005	0.001	0.032
Mean (se)		0.013 (0.000)	0.003 (0.000)	0.034 (0.001)
3rd Qu.		0.019	0.004	0.051
Max.		0.100	0.072	0.126
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.014	-0.053	0.163	0.073
1st Qu.	0.025	0.070	0.184	0.163
Median	0.028	0.119	0.190	0.179
Mean (se)	0.029 (0.000)	0.121 (0.002)	0.190 (0.000)	0.188 (0.001)
3rd Qu.	0.032	0.165	0.196	0.202
Max.	0.052	0.335	0.227	0.335
NA's		94		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.020	-0.045	0.230	0.053
1st Qu.	0.035	0.057	0.277	0.236
Median	0.040	0.111	0.287	0.265
Mean (se)	0.040 (0.000)	0.142 (0.003)	0.287 (0.000)	0.277 (0.001)
3rd Qu.	0.045	0.217	0.297	0.303
Max.	0.075	0.531	0.339	0.531
NA's		94		



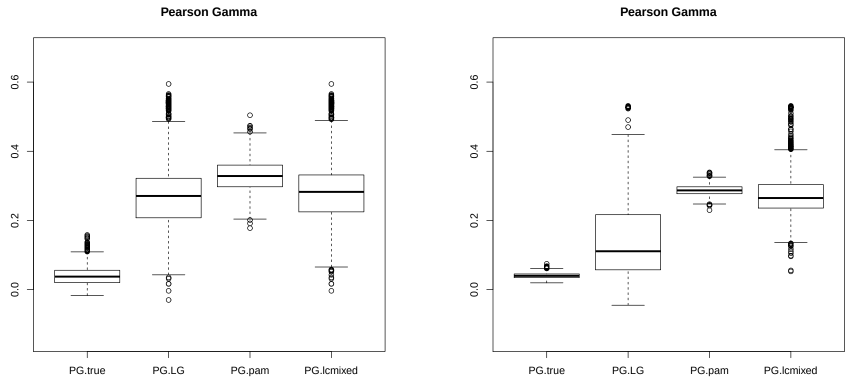
(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.151: Adjusted Rand Index: 4 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.152: Average Silhouette Width: 4 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.153: Pearson Gamma: 4 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1X}$	$\hat{\pi}_{1t}^{\bar{A}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_1X}$	$\hat{\pi}_{1t}^{\bar{B}_2X}$	$\hat{\pi}_{1t}^{\bar{B}_3X}$	$\hat{\pi}_{1t}^{\bar{C}_1X}$	$\hat{\pi}_{1t}^{\bar{C}_2X}$	$\hat{\pi}_{1t}^{\bar{C}_3X}$
1	0.50	0.90	0.10	0.80	0.10	0.10	0.10	0.10	0.80
2	0.50	0.10	0.90	0.10	0.80	0.10	0.80	0.10	0.10

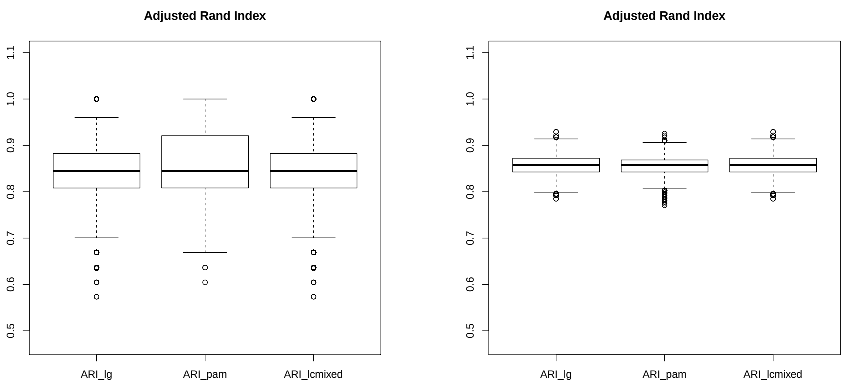
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1X}$	$\hat{\pi}_{1t}^{\bar{D}_2X}$	$\hat{\pi}_{1t}^{\bar{D}_3X}$	$\hat{\pi}_{1t}^{\bar{D}_4X}$
1	0.50	0.10	0.70	0.10	0.10
2	0.50	0.70	0.10	0.10	0.10

A.53 4 4-mixed variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Table A.103: Summary: 4 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.573	0.604	0.573
1st Qu.		0.808	0.808	0.808
Median		0.845	0.845	0.845
Mean (se)		0.844 (0.002)	0.856 (0.002)	0.845 (0.002)
3rd Qu.		0.882	0.921	0.882
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.416	0.448	0.465	0.452
1st Qu.	0.544	0.562	0.561	0.563
Median	0.573	0.589	0.588	0.589
Mean (se)	0.573 (0.001)	0.589 (0.001)	0.588 (0.001)	0.589 (0.001)
3rd Qu.	0.602	0.616	0.615	0.616
Max.	0.704	0.703	0.707	0.703
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.586	0.626	0.668	0.642
1st Qu.	0.736	0.761	0.762	0.761
Median	0.766	0.785	0.786	0.785
Mean (se)	0.764 (0.001)	0.783 (0.001)	0.784 (0.001)	0.784 (0.001)
3rd Qu.	0.791	0.808	0.808	0.808
Max.	0.869	0.881	0.881	0.881

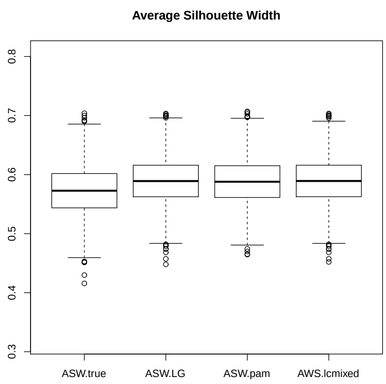


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

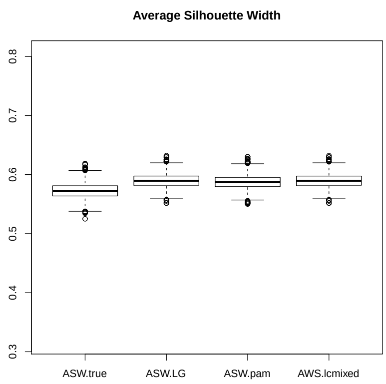
Figure A.154: Adjusted Rand Index: 4 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

Table A.104: Summary: 4 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.785	0.771	0.785
1st Qu.		0.843	0.843	0.843
Median		0.857	0.857	0.857
Mean (se)		0.859 (0.000)	0.855 (0.001)	0.859 (0.000)
3rd Qu.		0.872	0.868	0.872
Max.		0.929	0.925	0.929
	ASW true	**ASW lg**	**ASW pam**	**ASW lcmixed**
Min.	0.525	0.552	0.550	0.552
1st Qu.	0.564	0.582	0.580	0.582
Median	0.572	0.590	0.587	0.590
Mean (se)	0.573 (0.000)	0.590 (0.000)	0.588 (0.000)	0.590 (0.000)
3rd Qu.	0.581	0.598	0.595	0.598
Max.	0.619	0.632	0.630	0.632
	PG true	**PG lg**	**PG pam**	**PG lcmixed**
Min.	0.711	0.750	0.746	0.750
1st Qu.	0.755	0.780	0.777	0.780
Median	0.763	0.787	0.784	0.787
Mean (se)	0.763 (0.000)	0.787 (0.000)	0.784 (0.000)	0.787 (0.000)
3rd Qu.	0.771	0.794	0.790	0.794
Max.	0.806	0.820	0.817	0.820

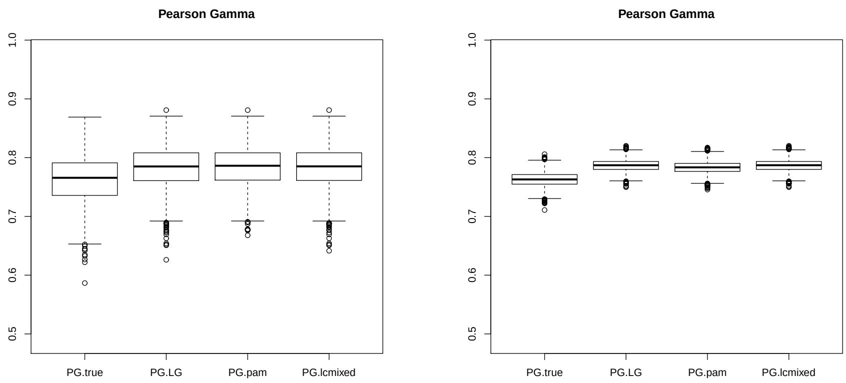


(a) *ASW - 100 units.*



(b) *ASW - 1000 units.*

Figure A.155: Average Silhouette Width: 4 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.156: Pearson Gamma: 4 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$
1	0.10	0.55	0.45	0.20	0.50	0.30	0.10	0.30	0.60
2	0.15	0.60	0.40	0.40	0.40	0.20	0.20	0.40	0.40
3	0.20	0.35	0.65	0.30	0.40	0.30	0.30	0.30	0.40
4	0.25	0.70	0.30	0.10	0.50	0.40	0.30	0.20	0.50
5	0.30	0.40	0.60	0.20	0.30	0.50	0.20	0.50	0.30

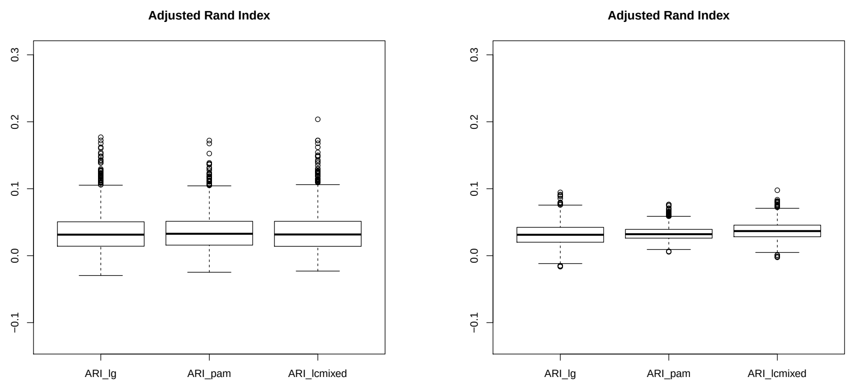
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.10	0.20	0.10	0.30	0.40
2	0.15	0.30	0.20	0.30	0.20
3	0.20	0.40	0.30	0.20	0.10
4	0.25	0.20	0.40	0.10	0.30
5	0.30	0.30	0.20	0.20	0.30

A.54 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Table A.105: Summary: 4 mixed-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.030	-0.025	-0.023
1st Qu.		0.014	0.016	0.014
Median		0.031	0.033	0.032
Mean (se)		0.035 (0.001)	0.035 (0.001)	0.036 (0.001)
3rd Qu.		0.051	0.051	0.051
Max.		0.177	0.172	0.204
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.148	0.030	0.146	0.047
1st Qu.	-0.079	0.190	0.220	0.192
Median	-0.069	0.221	0.237	0.221
Mean (se)	-0.068 (0.000)	0.217 (0.001)	0.237 (0.001)	0.218 (0.001)
3rd Qu.	-0.058	0.247	0.254	0.247
Max.	0.001	0.340	0.322	0.326
	PG true	PG lg	PG pam	PG lcmixed
Min.	-0.010	0.254	0.342	0.258
1st Qu.	0.049	0.412	0.435	0.411
Median	0.068	0.450	0.452	0.452
Mean (se)	0.070 (0.001)	0.443 (0.001)	0.452 (0.001)	0.444 (0.001)
3rd Qu.	0.089	0.481	0.470	0.481
Max.	0.202	0.566	0.544	0.575

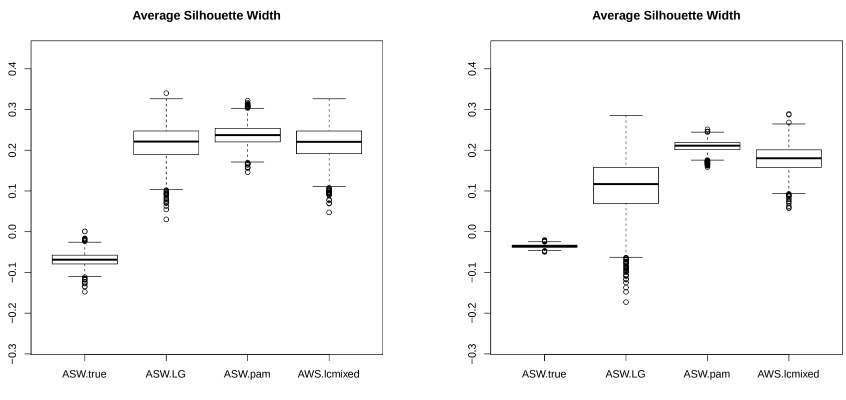


(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.157: Adjusted Rand Index: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

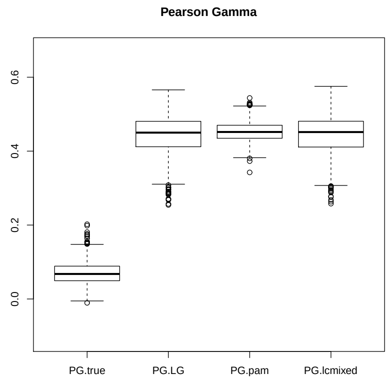
Table A.106: Summary: 4 mixed-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.016	0.006	-0.002
1st Qu.		0.020	0.026	0.028
Median		0.031	0.032	0.037
Mean (se)		0.031 (0.000)	0.033 (0.000)	0.038 (0.000)
3rd Qu.		0.042	0.039	0.046
Max.		0.095	0.077	0.098
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.050	-0.173	0.159	0.058
1st Qu.	-0.039	0.069	0.202	0.158
Median	-0.036	0.117	0.211	0.180
Mean (se)	-0.036 (0.000)	0.109 (0.002)	0.209 (0.000)	0.178 (0.001)
3rd Qu.	-0.033	0.158	0.219	0.201
Max.	-0.020	0.286	0.251	0.289
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.046	0.066	0.364	0.250
1st Qu.	0.064	0.287	0.398	0.358
Median	0.069	0.341	0.406	0.390
Mean (se)	0.070 (0.000)	0.337 (0.002)	0.407 (0.000)	0.390 (0.001)
3rd Qu.	0.075	0.397	0.415	0.423
Max.	0.100	0.506	0.472	0.497

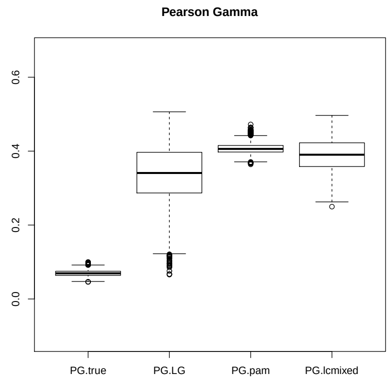


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.158: Average Silhouette Width: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.*



(b) *PG - 1000 units.*

Figure A.159: Pearson Gamma: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$
1	0.10	0.90	0.10	0.90	0.05	0.05	0.05	0.90	0.05
2	0.15	0.10	0.90	0.05	0.90	0.05	0.05	0.05	0.90
3	0.20	0.10	0.90	0.05	0.05	0.90	0.05	0.05	0.90
4	0.25	0.90	0.10	0.90	0.05	0.05	0.05	0.90	0.05
5	0.30	0.90	0.10	0.05	0.05	0.90	0.90	0.05	0.05

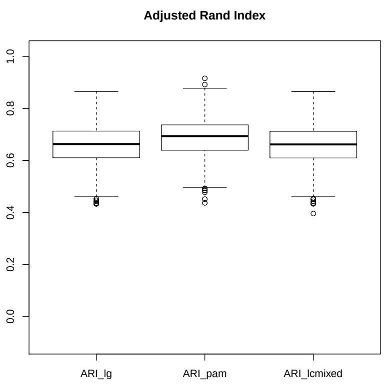
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.10	0.10	0.70	0.10	0.10
2	0.15	0.10	0.10	0.10	0.70
3	0.20	0.70	0.10	0.10	0.10
4	0.25	0.70	0.10	0.10	0.10
5	0.30	0.10	0.10	0.70	0.10

A.55 4 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

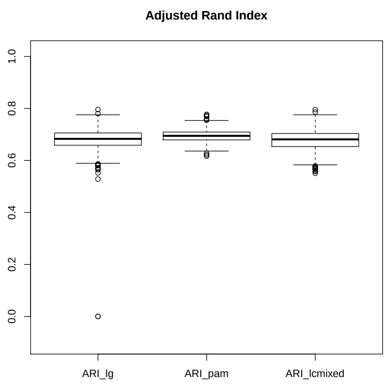
The Model:

Table A.107: Summary: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.433	0.437	0.396
1st Qu.		0.610	0.640	0.610
Median		0.663	0.693	0.662
Mean (se)		0.662 (0.002)	0.689 (0.002)	0.662 (0.002)
3rd Qu.		0.713	0.737	0.712
Max.		0.866	0.916	0.865
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.238	0.351	0.393	0.334
1st Qu.	0.373	0.491	0.494	0.491
Median	0.412	0.526	0.524	0.526
Mean (se)	0.412 (0.001)	0.526 (0.001)	0.525 (0.001)	0.526 (0.001)
3rd Qu.	0.449	0.561	0.553	0.561
Max.	0.593	0.702	0.691	0.702
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.537	0.574	0.598	0.547
1st Qu.	0.639	0.702	0.676	0.702
Median	0.664	0.728	0.696	0.728
Mean (se)	0.663 (0.001)	0.727 (0.001)	0.697 (0.001)	0.727 (0.001)
3rd Qu.	0.687	0.754	0.717	0.754
Max.	0.798	0.831	0.803	0.831



(a) *ARI - 100 units.*

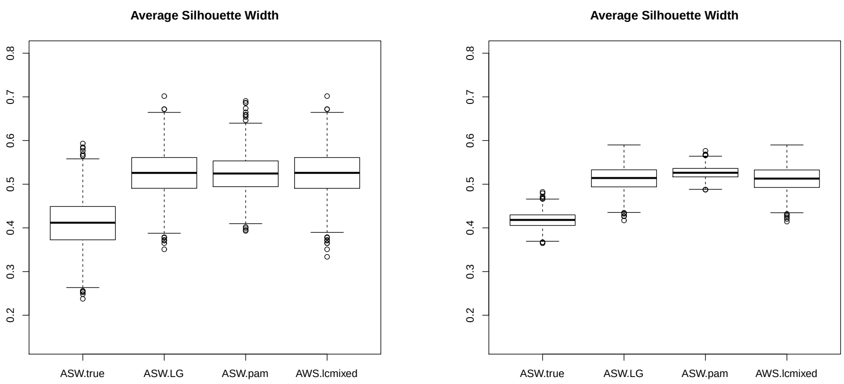


(b) *ARI - 1000 units.*

Figure A.160: Adjusted Rand Index: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

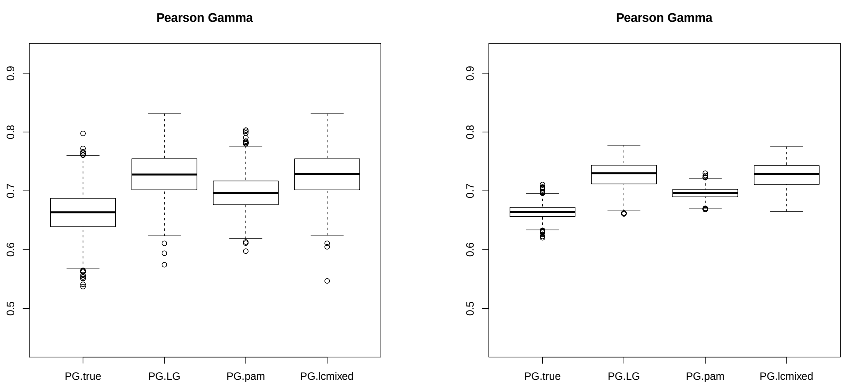
Table A.108: Summary: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.617	0.551
1st Qu.		0.659	0.679	0.654
Median		0.683	0.695	0.681
Mean (se)		0.681 (0.000)	0.694 (0.001)	0.678 (0.000)
3rd Qu.		0.706	0.709	0.704
Max.		0.796	0.777	0.796
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.365	0.417	0.487	0.414
1st Qu.	0.406	0.494	0.517	0.493
Median	0.418	0.514	0.526	0.513
Mean (se)	0.418 (0.000)	0.512 (0.000)	0.527 (0.000)	0.512 (0.000)
3rd Qu.	0.430	0.533	0.536	0.533
Max.	0.614	0.590	0.577	0.590
NA's		2		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.620	0.661	0.668	0.665
1st Qu.	0.656	0.712	0.690	0.711
Median	0.664	0.730	0.696	0.729
Mean (se)	0.664 (0.000)	0.727 (0.000)	0.696 (0.000)	0.726 (0.000)
3rd Qu.	0.672	0.744	0.703	0.743
Max.	0.711	0.778	0.730	0.775
NA's		2		



(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.161: Average Silhouette Width: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.162: Pearson Gamma: 4 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

Table A.109: Summary: 4 mixed-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 100 units

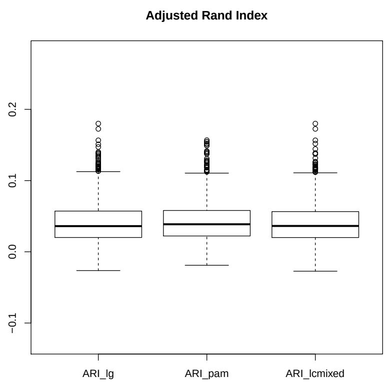
		ARI lg	ARI pam	ARI lcmixed
Min.		-0.026	-0.019	-0.027
1st Qu.		0.020	0.022	0.020
Median		0.036	0.039	0.036
Mean (se)		0.040 (0.001)	0.042 (0.001)	0.040 (0.001)
3rd Qu.		0.057	0.058	0.056
Max.		0.180	0.157	0.180
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.096	0.071	0.154	0.071
1st Qu.	-0.069	0.193	0.220	0.193
Median	-0.061	0.222	0.236	0.222
Mean (se)	-0.059 (0.000)	0.220 (0.001)	0.236 (0.001)	0.220 (0.001)
3rd Qu.	-0.050	0.249	0.253	0.249
Max.	0.011	0.335	0.335	0.331
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.005	0.272	0.360	0.241
1st Qu.	0.060	0.414	0.434	0.415
Median	0.079	0.451	0.452	0.453
Mean (se)	0.081 (0.001)	0.446 (0.001)	0.452 (0.001)	0.447 (0.001)
3rd Qu.	0.099	0.482	0.470	0.482
Max.	0.201	0.570	0.538	0.570

Latent class	$\hat{\pi}^{X}_{t}$	$\hat{\pi}^{\bar{A}_1 X}_{1t}$	$\hat{\pi}^{\bar{A}_2 X}_{1t}$	$\hat{\pi}^{\bar{B}_1 X}_{1t}$	$\hat{\pi}^{\bar{B}_2 X}_{1t}$	$\hat{\pi}^{\bar{B}_3 X}_{1t}$	$\hat{\pi}^{\bar{C}_1 X}_{1t}$	$\hat{\pi}^{\bar{C}_2 X}_{1t}$	$\hat{\pi}^{\bar{C}_3 X}_{1t}$
1	0.20	0.60	0.40	0.20	0.10	0.70	0.30	0.20	0.50
2	0.20	0.45	0.55	0.40	0.20	0.40	0.50	0.10	0.40
3	0.20	0.70	0.30	0.30	0.30	0.40	0.40	0.30	0.30
4	0.20	0.35	0.65	0.20	0.30	0.50	0.20	0.60	0.20
5	0.20	0.30	0.70	0.40	0.30	0.30	0.30	0.40	0.30

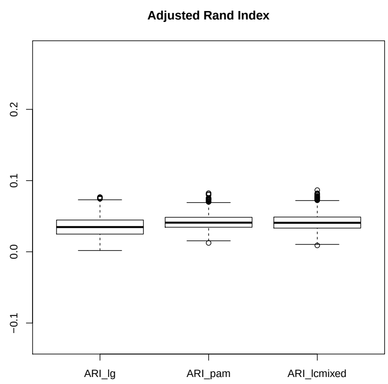
Latent class	$\hat{\pi}^{X}_{t}$	$\hat{\pi}^{\bar{D}_1 X}_{1t}$	$\hat{\pi}^{\bar{D}_2 X}_{1t}$	$\hat{\pi}^{\bar{D}_3 X}_{1t}$	$\hat{\pi}^{\bar{D}_4 X}_{1t}$
1	0.20	0.20	0.40	0.30	0.10
2	0.20	0.30	0.30	0.20	0.20
3	0.20	0.30	0.40	0.10	0.20
4	0.20	0.40	0.20	0.20	0.20
5	0.20	0.20	0.20	0.30	0.30

Table A.110: Summary: 4 mixed-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.002	0.012	0.009
1st Qu.		0.025	0.034	0.033
Median		0.035	0.041	0.041
Mean (se)		0.035 (0.000)	0.042 (0.000)	0.041 (0.000)
3rd Qu.		0.045	0.048	0.049
Max.		0.077	0.082	0.087
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.047	-0.155	0.148	0.036
1st Qu.	-0.039	0.069	0.198	0.159
Median	-0.037	0.116	0.207	0.179
Mean (se)	-0.037 (0.000)	0.109 (0.001)	0.206 (0.000)	0.178 (0.001)
3rd Qu.	-0.035	0.158	0.215	0.200
Max.	-0.025	0.277	0.237	0.281
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.056	0.048	0.365	0.209
1st Qu.	0.075	0.292	0.395	0.360
Median	0.080	0.343	0.404	0.390
Mean (se)	0.080 (0.000)	0.338 (0.002)	0.404 (0.000)	0.390 (0.001)
3rd Qu.	0.085	0.392	0.413	0.420
Max.	0.117	0.506	0.463	0.501

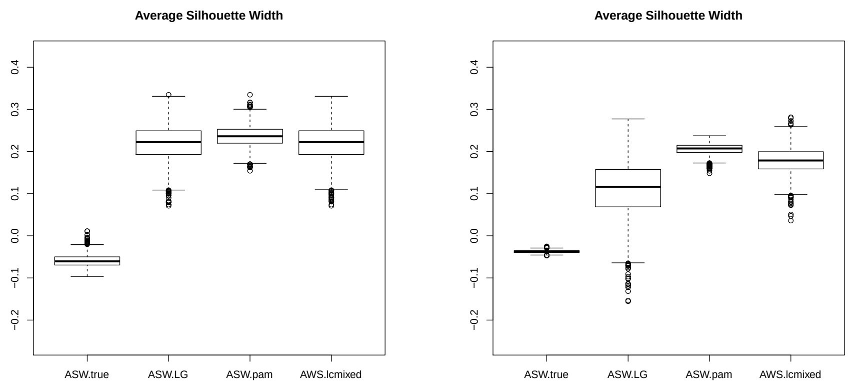


(a) *ARI - 100 units.*



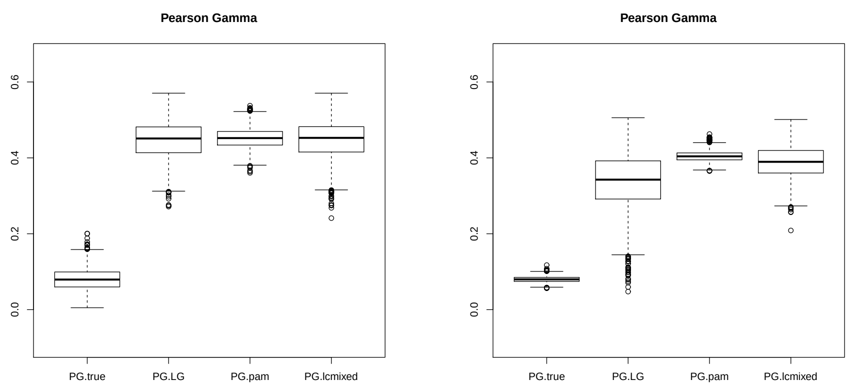
(b) *ARI - 1000 units.*

Figure A.163: Adjusted Rand Index: 4 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



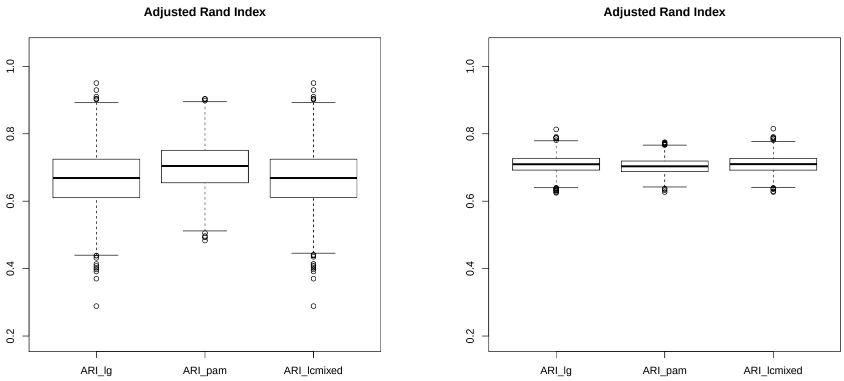
(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.164: Average Silhouette Width: 4 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.165: Pearson Gamma: 4 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.166: Adjusted Rand Index: 4 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

A.56 4 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

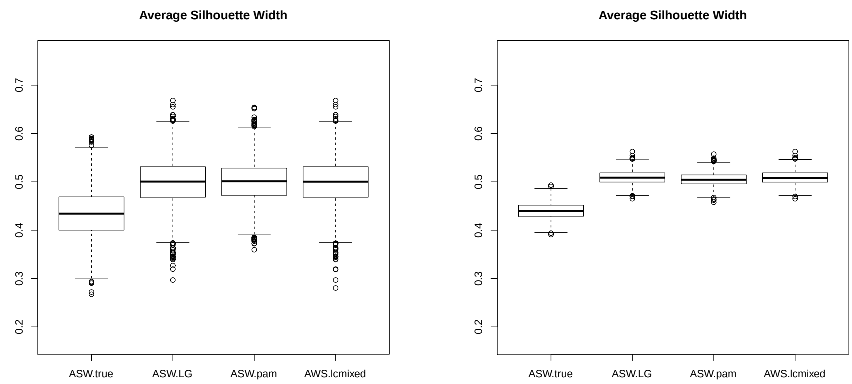
The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_3 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_3 X}$
1	0.20	0.10	0.90	0.80	0.10	0.10	0.05	0.05	0.90
2	0.20	0.10	0.90	0.10	0.10	0.80	0.90	0.05	0.05
3	0.20	0.90	0.10	0.10	0.10	0.80	0.10	0.80	0.10
4	0.20	0.10	0.90	0.10	0.80	0.10	0.05	0.90	0.05
5	0.20	0.90	0.10	0.80	0.10	0.10	0.90	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{D}_4 X}$
1	0.20	0.10	0.10	0.70	0.10
2	0.20	0.10	0.70	0.10	0.10
3	0.20	0.70	0.10	0.10	0.10
4	0.20	0.10	0.10	0.10	0.70
5	0.20	0.10	0.10	0.70	0.10

Table A.111: Summary: 4 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.289	0.483	0.289
1st Qu.		0.610	0.655	0.611
Median		0.669	0.704	0.669
Mean (se)		0.666 (0.002)	0.703 (0.002)	0.666 (0.002)
3rd Qu.		0.725	0.751	0.725
Max.		0.950	0.904	0.950
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.267	0.297	0.360	0.280
1st Qu.	0.400	0.468	0.472	0.468
Median	0.434	0.501	0.501	0. 500
Mean (se)	0.435 (0.001)	0.499 (0.001)	0.501 (0.001)	0.499 (0.001)
3rd Qu.	0.469	0.531	0.528	0.531
Max.	0.592	0.668	0.654	0.668
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.475	0.501	0.574	0.501
1st Qu.	0.598	0.656	0.658	0.656
Median	0.628	0.682	0.680	0.682
Mean (se)	0.627 (0.001)	0.679 (0.001)	0.679 (0.001)	0.679 (0.001)
3rd Qu.	0.656	0.704	0.702	0.704
Max.	0.751	0.785	0.781	0.785

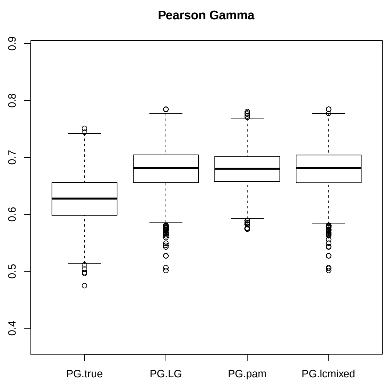


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

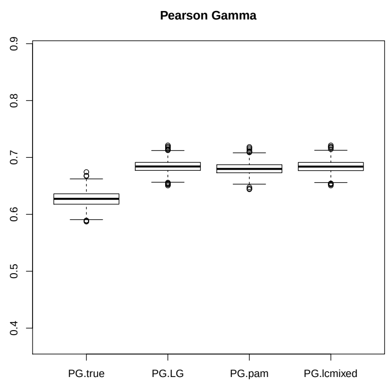
Figure A.167: Average Silhouette Width: 4 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

Table A.112: Summary: 4 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.625	0.627	0.627
1st Qu.		0.692	0.688	0.692
Median		0.710	0.704	0.710
Mean (se)		0.710 (0.001)	0.703 (0.001)	0.710 (0.001)
3rd Qu.		0.727	0.719	0.727
Max.		0.813	0.775	0.815
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.391	0.465	0.458	0.465
1st Qu.	0.429	0.500	0.496	0.500
Median	0.440	0.509	0.504	0.508
Mean (se)	0.440 (0.000)	0.509 (0.000)	0.505 (0.000)	0.509 (0.000)
3rd Qu.	0.452	0.518	0.514	0.519
Max.	0.493	0.563	0.557	0.563
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.587	0.651	0.644	0.651
1st Qu.	0.618	0.677	0.673	0.677
Median	0.627	0.684	0.680	0.684
Mean (se)	0.627 (0.000)	0.684 (0.000)	0.680 (0.000)	0.684 (0.000)
3rd Qu.	0.636	0.691	0.687	0.691
Max.	0.674	0.721	0.718	0.721



(a) *PG - 100 units.*



(b) *PG - 1000 units.*

Figure A.168: Pearson Gamma: 4 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

A.57 12 mix-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.15	0.70	0.30	0.40	0.60	0.65	0.35
2	0.85	0.60	0.40	0.70	0.30	0.30	0.70

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.15	0.30	0.25	0.45	0.20	0.10	0.70	0.30	0.40	0.30
2	0.85	0.40	0.25	0.35	0.40	0.20	0.40	0.50	0.30	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.15	0.20	0.10	0.40	0.30	0.40	0.30	0.10	0.20
2	0.85	0.10	0.40	0.20	0.30	0.10	0.20	0.40	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.15	0.20	0.30	0.20	0.30	0.15	0.35	0.20	0.30
2	0.85	0.30	0.10	0.40	0.20	0.30	0.10	0.35	0.25

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.15	0.10	0.05	0.20	0.15	0.20	0.10	0.05	0.15
2	0.85	0.15	0.10	0.05	0.20	0.15	0.05	0.10	0.20

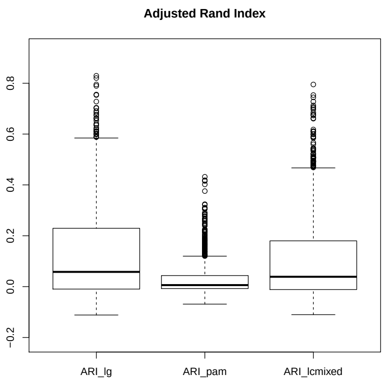
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.15	0.20	0.05	0.10	0.15	0.05	0.15	0.20	0.10
2	0.85	0.15	0.10	0.05	0.20	0.10	0.20	0.15	0.05

A.58 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

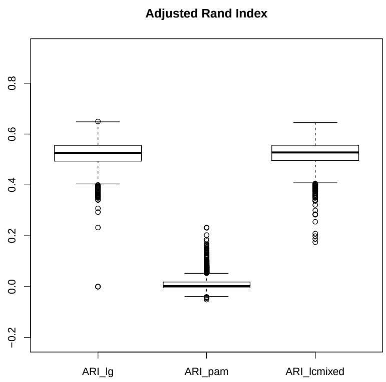
The Model:

Table A.113: Summary: 12 mixed-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.112	-0.069	-0.110
1st Qu.		-0.010	-0.007	-0.011
Median		0.058	0.006	0.039
Mean (se)		0.121 (0.004)	0.027 (0.001)	0.101 (0.004)
3rd Qu.		0.229	0.044	0.180
Max.		0.829	0.432	0.795
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.019	0.029	0.034	0.029
1st Qu.	0.058	0.071	0.065	0.070
Median	0.069	0.081	0.073	0.080
Mean (se)	0.070 (0.000)	0.084 (0.000)	0.074 (0.000)	0.082 (0.000)
3rd Qu.	0.081	0.096	0.082	0.094
Max.	0.132	0.148	0.144	0.148
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.040	0.006	0.077	0.005
1st Qu.	0.135	0.164	0.166	0.164
Median	0.163	0.200	0.190	0.198
Mean (se)	0.164 (0.001)	0.200 (0.001)	0.194 (0.001)	0.200 (0.001)
3rd Qu.	0.191	0.240	0.217	0.240
Max.	0.322	0.377	0.388	0.374



(a) *ARI - 100 units.*

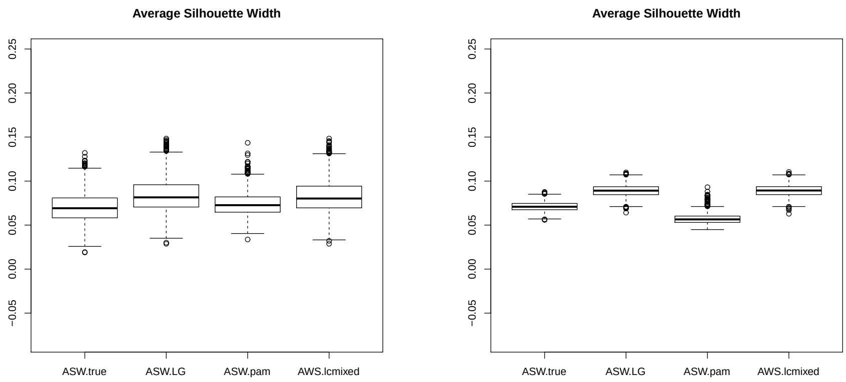


(b) *ARI - 1000 units.*

Figure A.169: Adjusted Rand Index: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation

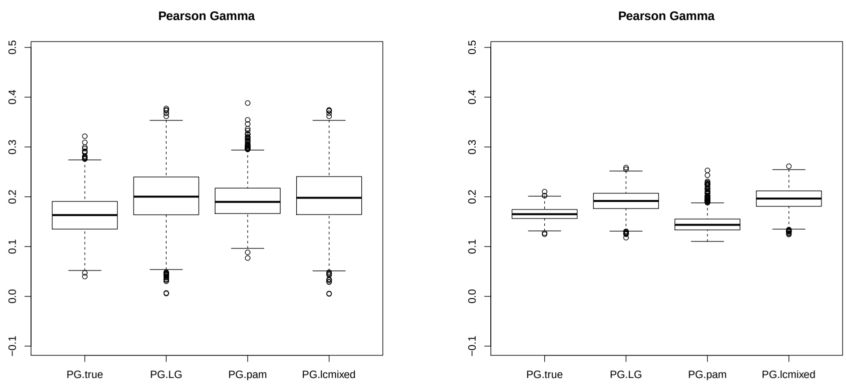
Table A.114: Summary: 12 mixed-level variables - 2 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	-0.051	0.175
1st Qu.		0.494	-0.005	0.496
Median		0.526	0.002	0.528
Mean (se)		0.521 (0.001)	0.012 (0.001)	0.523 (0.001)
3rd Qu.		0.556	0.018	0.556
Max.		0.650	0.233	0.645
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.056	0.064	0.045	0.063
1st Qu.	0.067	0.085	0.053	0.085
Median	0.071	0.089	0.056	0.089
Mean (se)	0.071 (0.000)	0.089 (0.000)	0.057 (0.000)	0.089 (0.000)
3rd Qu.	0.075	0.094	0.060	0.094
Max.	0.088	0.110	0.093	0.110
NA's		4		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.125	0.118	0.110	0.124
1st Qu.	0.156	0.176	0.134	0.181
Median	0.165	0.192	0.144	0.196
Mean (se)	0.165 (0.000)	0.192 (0.001)	0.146 (0.000)	0.196 (0.001)
3rd Qu.	0.174	0.207	0.155	0.212
Max.	0.210	0.259	0.253	0.261
NA's		4		



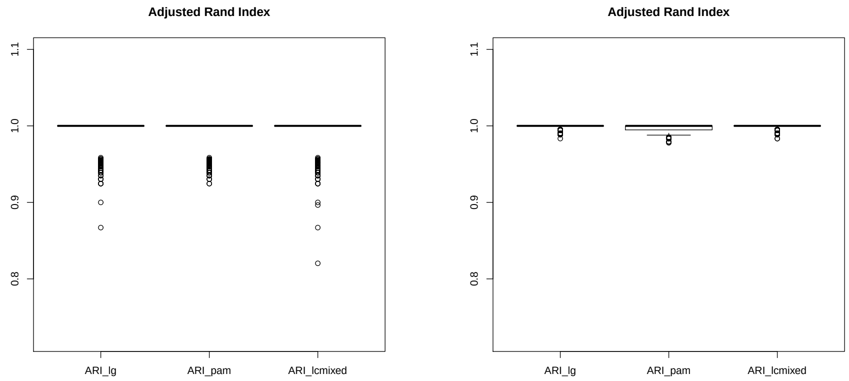
(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.170: Average Silhouette Width: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.171: Pearson Gamma: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.172: Adjusted Rand Index: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.15	0.10	0.90	0.10	0.90	0.90	0.10
2	0.85	0.90	0.10	0.90	0.10	0.10	0.90

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.15	0.90	0.05	0.05	0.05	0.90	0.05	0.05	0.05	0.90
2	0.85	0.05	0.90	0.05	0.05	0.05	0.90	0.90	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.15	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10
2	0.85	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70

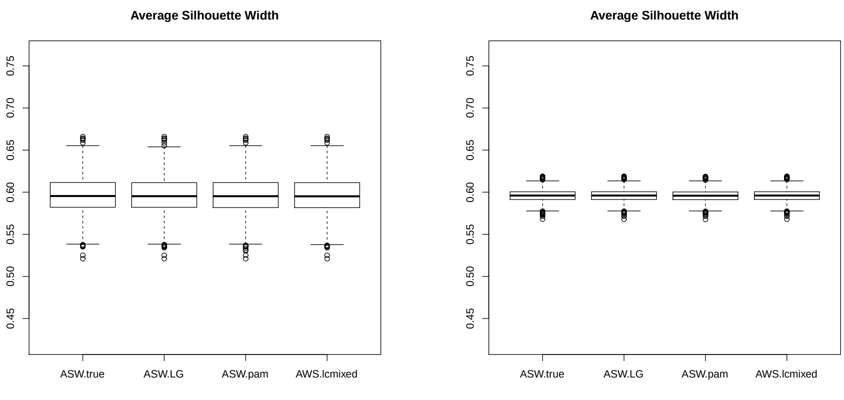
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.15	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70
2	0.85	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.15	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.85	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
2	0.85	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

Table A.115: Summary: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.867	0.924	0.820
1st Qu.		1.000	1.000	1.000
Median		1.000	1.000	1.000
Mean (se)		0.998 (0.000)	0.998 (0.000)	0.997 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.521	0.521	0.521	0.521
1st Qu.	0.582	0.582	0.582	0.582
Median	0.595	0.595	0.595	0.595
Mean (se)	0.596 (0.000)	0.596 (0.000)	0.596 (0.000)	0.596 (0.000)
3rd Qu.	0.612	0.611	0.611	0.611
Max.	0.666	0.666	0.666	0.666
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.766	0.766	0.766	0.766
1st Qu.	0.867	0.867	0.867	0.867
Median	0.883	0.883	0.883	0.883
Mean (se)	0.880 (0.001)	0.880 (0.001)	0.880 (0.001)	0.880 (0.001)
3rd Qu.	0.896	0.897	0.896	0.896
Max.	0.937	0.937	0.937	0.937

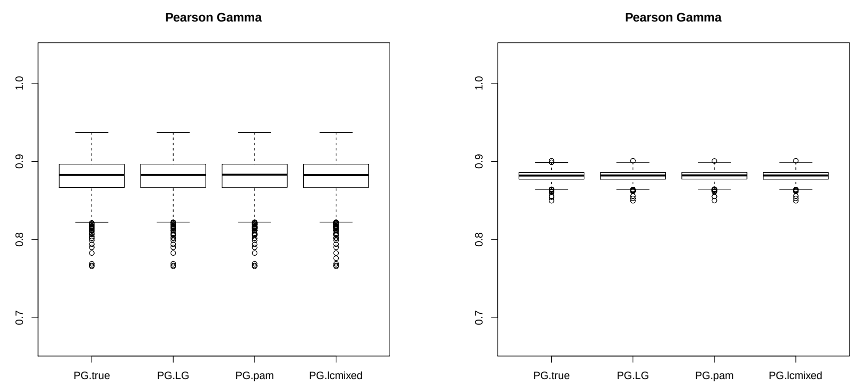


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.173: Average Silhouette Width: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

Table A.116: Summary: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.983	0.978	0.983
1st Qu.		1.000	0.995	1.000
Median		1.000	1.000	1.000
Mean (se)		0.999 (0.000)	0.997 (0.000)	0.999 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.568	0.568	0.568	0.568
1st Qu.	0.591	0.591	0.591	0.591
Median	0.596	0.596	0.596	0.596
Mean (se)	0.596 (0.000)	0.596 (0.000)	0.596 (0.000)	0.596 (0.000)
3rd Qu.	0.600	0.600	0.600	0.600
Max.	0.619	0.619	0.619	0.619
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.850	0.850	0.850	0.850
1st Qu.	0.877	0.877	0.878	0.877
Median	0.882	0.882	0.882	0.882
Mean (se)	0.882 (0.000)	0.882 (0.000)	0.882 (0.000)	0.882 (0.000)
3rd Qu.	0.886	0.886	0.886	0.886
Max.	0.901	0.901	0.900	0.901



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.174: Pearson Gamma: 12 mixed-level variables - 2 clusters, mixing proportions extremely different - Clear cluster separation

A.59 12 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.15	0.60	0.40	0.55	0.45	0.40	0.60
2	0.85	0.25	0.75	0.70	0.30	0.70	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.15	0.40	0.40	0.20	0.10	0.40	0.50	0.30	0.40	0.30
2	0.85	0.30	0.20	0.50	0.30	0.30	0.40	0.20	0.50	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.15	0.30	0.10	0.20	0.40	0.20	0.40	0.10	0.30
2	0.85	0.20	0.30	0.40	0.10	0.10	0.30	0.20	0.40

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.15	0.25	0.10	0.35	0.30	0.10	0.30	0.40	0.20
2	0.85	0.30	0.20	0.15	0.35	0.40	0.30	0.20	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.15	0.20	0.20	0.05	0.10	0.15	0.10	0.05	0.15
2	0.85	0.10	0.05	0.20	0.05	0.20	0.15	0.15	0.10

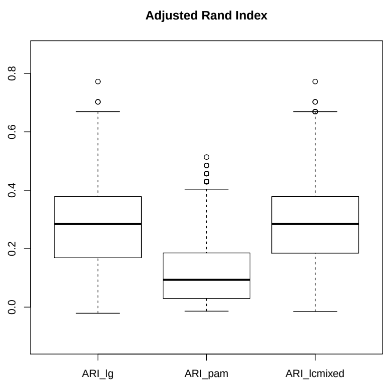
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.15	0.20	0.05	0.10	0.15	0.10	0.15	0.05	0.20
2	0.85	0.15	0.10	0.05	0.20	0.05	0.20	0.20	0.10

A.60 12 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

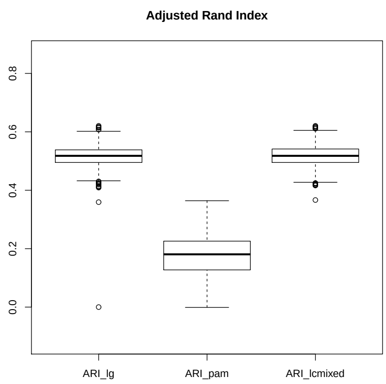
The Model:

Table A.117: Summary: 12 mixed-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		-0.021	-0.014	-0.015
1st Qu.		0.169	0.030	0.185
Median		0.284	0.094	0.285
Mean (se)		0.281 (0.003)	0.118 (0.002)	0.285 (0.003)
3rd Qu.		0.378	0.186	0.378
Max.		0.772	0.514	0.772
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.025	0.046	0.041	0.043
1st Qu.	0.062	0.082	0.073	0.082
Median	0.071	0.093	0.083	0.093
Mean (se)	0.071 (0.000)	0.093 (0.000)	0.085 (0.000)	0.094 (0.000)
3rd Qu.	0.081	0.104	0.097	0.104
Max.	0.124	0.164	0.156	0.164
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.066	0.056	0.112	0.063
1st Qu.	0.167	0.202	0.196	0.203
Median	0.192	0.239	0.226	0.240
Mean (se)	0.194 (0.001)	0.238 (0.001)	0.230 (0.001)	0.240 (0.001)
3rd Qu.	0.219	0.274	0.263	0.276
Max.	0.341	0.446	0.424	0.446



(a) *ARI - 100 units.*

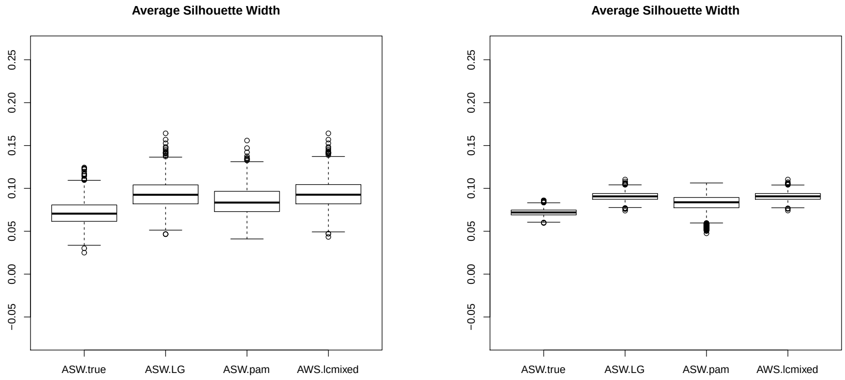


(b) *ARI - 1000 units.*

Figure A.175: Adjusted Rand Index: 12 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation

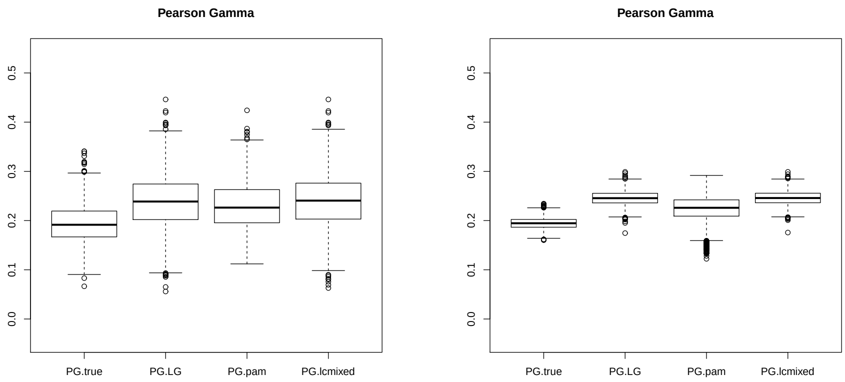
Table A.118: Summary: 12 mixed-level variables - 2 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	-0.001	0.367
1st Qu.		0.495	0.127	0.495
Median		0.518	0.181	0.518
Mean (se)		0.517 (0.001)	0.176 (0.002)	0.519 (0.001)
3rd Qu.		0.538	0.226	0.541
Max.		0.621	0.364	0.621
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.060	0.074	0.048	0.074
1st Qu.	0.069	0.087	0.077	0.087
Median	0.072	0.091	0.084	0.091
Mean (se)	0.072 (0.000)	0.091 (0.000)	0.082 (0.000)	0.091 (0.000)
3rd Qu.	0.075	0.094	0.089	0.094
Max.	0.086	0.110	0.106	0.110
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.160	0.175	0.122	0.176
1st Qu.	0.187	0.236	0.209	0.236
Median	0.195	0.246	0.226	0.246
Mean (se)	0.195 (0.000)	0.246 (0.000)	0.223 (0.001)	0.246 (0.000)
3rd Qu.	0.202	0.255	0.242	0.256
Max.	0.234	0.299	0.292	0.299
NA's		1		



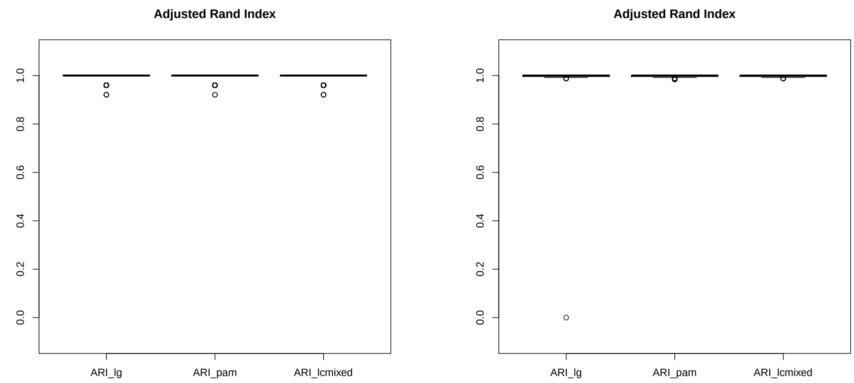
(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.176: Average Silhouette Width: 12 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.177: Pearson Gamma: 12 mixed-level variables - 2 clusters, equal mixing proportions - Unclear cluster separation



(a) *ARI - 100 units.* (b) *ARI - 1000 units.*

Figure A.178: Adjusted Rand Index: 12 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.50	0.90	0.10	0.10	0.90	0.10	0.90
2	0.50	0.10	0.90	0.90	0.10	0.90	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.50	0.05	0.05	0.90	0.05	0.90	0.05	0.90	0.05	0.05
2	0.50	0.90	0.05	0.05	0.05	0.05	0.90	0.05	0.90	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.50	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10
2	0.50	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10

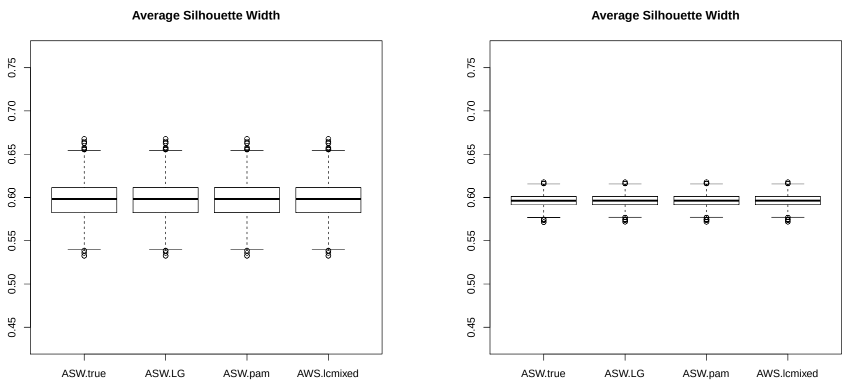
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.50	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
2	0.50	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.50	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
2	0.50	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.50	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
2	0.50	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05

Table A.119: Summary: 12 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 100 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.920	0.921	0.920
1st Qu.		1.000	1.000	1.000
Median		1.000	1.000	1.000
Mean (se)		0.998 (0.000)	0.998 (0.000)	0.998 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.532	0.532	0.532	0.532
1st Qu.	0.582	0.582	0.582	0.582
Median	0.598	0.598	0.598	0.598
Mean (se)	0.597 (0.000)	0.597 (0.000)	0.597 (0.000)	0.597 (0.000)
3rd Qu.	0.611	0.611	0.611	0.611
Max.	0.668	0.668	0.668	0.668
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.876	0.876	0.873	0.876
1st Qu.	0.910	0.910	0.910	0.910
Median	0.917	0.917	0.917	0.917
Mean (se)	0.916 (0.000)	0.916 (0.000)	0.917 (0.000)	0.916 (0.000)
3rd Qu.	0.924	0.924	0.924	0.924
Max.	0.947	0.947	0.947	0.947

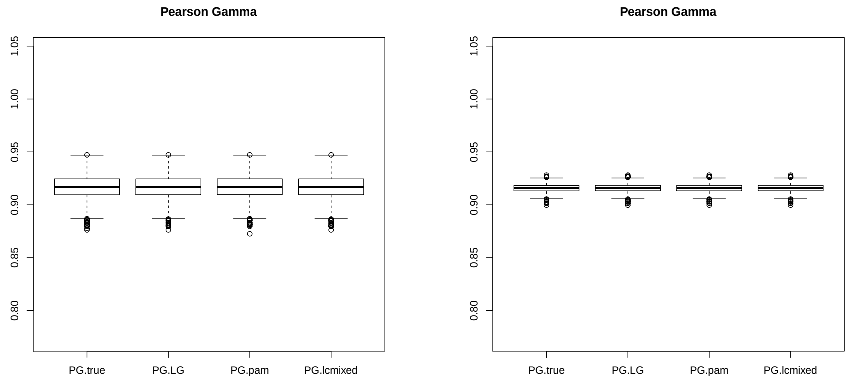


(a) *ASW - 100 units.* (b) *ASW - 1000 units.*

Figure A.179: Average Silhouette Width: 12 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

Table A.120: Summary: 12 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.984	0.988
1st Qu.		0.996	0.996	0.996
Median		1.000	1.000	1.000
Mean (se)		0.998 (0.001)	0.998 (0.000)	0.999 (0.000)
3rd Qu.		1.000	1.000	1.000
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.571	0.572	0.572	0.572
1st Qu.	0.591	0.592	0.592	0.592
Median	0.596	0.596	0.596	0.596
Mean (se)	0.596 (0.000)	0.596 (0.000)	0.596 (0.000)	0.596 (0.000)
3rd Qu.	0.601	0.601	0.601	0.601
Max.	0.618	0.618	0.618	0.618
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.900	0.900	0.900	0.900
1st Qu.	0.913	0.913	0.913	0.913
Median	0.916	0.916	0.916	0.916
Mean (se)	0.916 (0.000)	0.916 (0.000)	0.916 (0.000)	0.916 (0.000)
3rd Qu.	0.918	0.918	0.918	0.918
Max.	0.928	0.928	0.928	0.928
NA's		1		



(a) *PG - 100 units.* (b) *PG - 1000 units.*

Figure A.180: Pearson Gamma: 12 mixed-level variables - 2 clusters, equal mixing proportions - Clear cluster separation

A.61 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.10	0.60	0.40	0.80	0.20	0.30	0.70
2	0.15	0.85	0.15	0.30	0.70	0.60	0.40
3	0.20	0.30	0.70	0.15	0.85	0.20	0.80
4	0.25	0.45	0.55	0.25	0.75	0.40	0.60
5	0.30	0.35	0.65	0.40	0.60	0.70	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.10	0.20	0.35	0.45	0.40	0.20	0.40	0.30	0.40	0.30
2	0.15	0.45	0.15	0.40	0.55	0.15	0.30	0.20	0.30	0.50
3	0.20	0.30	0.20	0.50	0.25	0.45	0.30	0.10	0.20	0.70
4	0.25	0.60	0.30	0.10	0.30	0.50	0.20	0.45	0.15	0.40
5	0.30	0.10	0.70	0.20	0.20	0.30	0.50	0.25	0.25	0.50

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.10	0.40	0.15	0.15	0.30	0.20	0.35	0.15	0.30
2	0.15	0.20	0.40	0.30	0.10	0.10	0.20	0.40	0.30
3	0.20	0.30	0.20	0.15	0.35	0.40	0.20	0.30	0.10
4	0.25	0.15	0.30	0.25	0.30	0.50	0.15	0.15	0.20
5	0.30	0.10	0.20	0.40	0.30	0.30	0.40	0.20	0.10

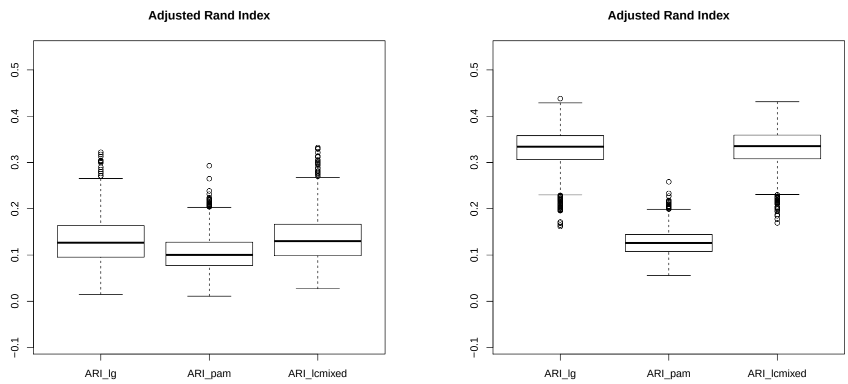
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.10	0.30	0.10	0.20	0.40	0.20	0.50	0.10	0.20
2	0.15	0.10	0.40	0.30	0.20	0.30	0.40	0.20	0.10
3	0.20	0.20	0.10	0.40	0.30	0.45	0.25	0.15	0.15
4	0.25	0.25	0.20	0.35	0.20	0.10	0.35	0.45	0.10
5	0.30	0.30	0.50	0.20	0.10	0.20	0.30	0.25	0.25

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.10	0.10	0.05	0.15	0.10	0.20	0.20	0.15	0.05
2	0.15	0.15	0.20	0.05	0.20	0.10	0.05	0.10	0.15
3	0.20	0.20	0.10	0.10	0.05	0.15	0.05	0.15	0.20
4	0.25	0.05	0.15	0.20	0.10	0.05	0.15	0.20	0.10
5	0.30	0.20	0.10	0.15	0.15	0.20	0.10	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.10	0.20	0.05	0.10	0.15	0.10	0.05	0.20	0.15
2	0.15	0.10	0.15	0.20	0.10	0.05	0.15	0.05	0.20
3	0.20	0.15	0.20	0.15	0.05	0.20	0.10	0.10	0.05
4	0.25	0.05	0.10	0.05	0.20	0.15	0.20	0.15	0.10
5	0.30	0.20	0.05	0.10	0.15	0.10	0.05	0.20	0.15

Table A.121: Summary: 12 mixed-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.015	0.011	0.027
1st Qu.		0.095	0.077	0.098
Median		0.127	0.100	0.130
Mean (se)		0.132 (0.001)	0.104 (0.001)	0.135 (0.001)
3rd Qu.		0.163	0.128	0.166
Max.		0.322	0.293	0.332
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	-0.001	0.013	0.040	0.025
1st Qu.	0.018	0.052	0.056	0.054
Median	0.023	0.058	0.060	0.060
Mean (se)	0.023 (0.000)	0.058 (0.000)	0.060 (0.000)	0.060 (0.000)
3rd Qu.	0.028	0.064	0.065	0.066
Max.	0.049	0.094	0.089	0.093
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.133	0.124	0.221	0.171
1st Qu.	0.187	0.245	0.264	0.250
Median	0.202	0.264	0.275	0.268
Mean (se)	0.202 (0.000)	0.263 (0.000)	0.275 (0.000)	0.268 (0.001)
3rd Qu.	0.216	0.281	0.287	0.286
Max.	0.266	0.354	0.343	0.350

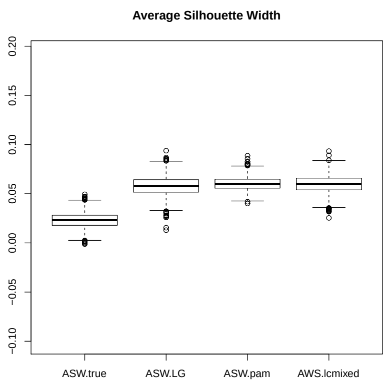


(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

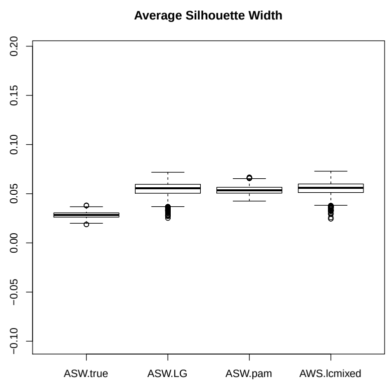
Figure A.181: Adjusted Rand Index: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

Table A.122: Summary: 12 mixed-level variables - 5 clusters - mixing proportions extremely different - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.162	0.056	0.169
1st Qu.		0.307	0.108	0.308
Median		0.334	0.126	0.335
Mean (se)		0.331 (0.001)	0.128 (0.001)	0.332 (0.001)
3rd Qu.		0.358	0.144	0.359
Max.		0.438	0.258	0.431
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.019	0.025	0.043	0.025
1st Qu.	0.026	0.050	0.051	0.051
Median	0.028	0.056	0.053	0.056
Mean (se)	0.028 (0.000)	0.055 (0.000)	0.054 (0.000)	0.055 (0.000)
3rd Qu.	0.031	0.060	0.057	0.060
Max.	0.038	0.072	0.067	0.073
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.175	0.159	0.215	0.167
1st Qu.	0.196	0.254	0.243	0.255
Median	0.202	0.266	0.251	0.267
Mean (se)	0.202 (0.000)	0.264 (0.000)	0.251 (0.000)	0.265 (0.000)
3rd Qu.	0.208	0.277	0.259	0.277
Max.	0.238	0.312	0.287	0.312

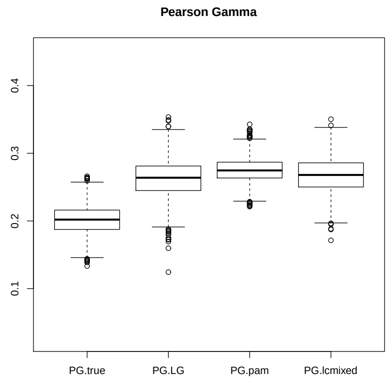


(a) *ASW - 200 units.*

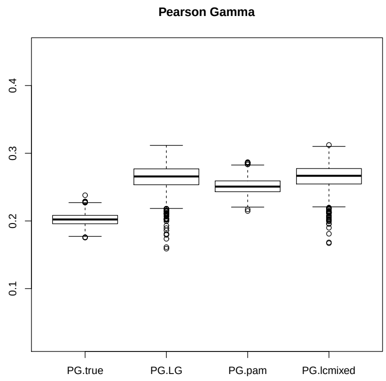


(b) *ASW - 1000 units.*

Figure A.182: Average Silhouette Width: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation



(a) *PG - 200 units.*



(b) *PG - 1000 units.*

Figure A.183: Pearson Gamma: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Unclear cluster separation

A.62 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.10	0.10	0.90	0.90	0.10	0.10	0.90
2	0.15	0.90	0.10	0.90	0.10	0.10	0.90
3	0.20	0.90	0.10	0.10	0.90	0.90	0.10
4	0.25	0.10	0.90	0.90	0.10	0.10	0.90
5	0.30	0.10	0.90	0.10	0.90	0.90	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.10	0.90	0.05	0.05	0.05	0.90	0.05	0.05	0.05	0.90
2	0.15	0.05	0.90	0.05	0.90	0.05	0.05	0.05	0.05	0.90
3	0.20	0.05	0.05	0.90	0.05	0.90	0.05	0.90	0.05	0.05
4	0.25	0.05	0.05	0.90	0.05	0.05	0.90	0.05	0.90	0.05
5	0.30	0.90	0.05	0.05	0.05	0.05	0.90	0.05	0.90	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.10	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
2	0.15	0.10	0.10	0.10	0.70	0.10	0.70	0.10	0.10
3	0.20	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
4	0.25	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
5	0.30	0.70	0.10	0.10	0.10	0.70	0.10	0.10	0.10

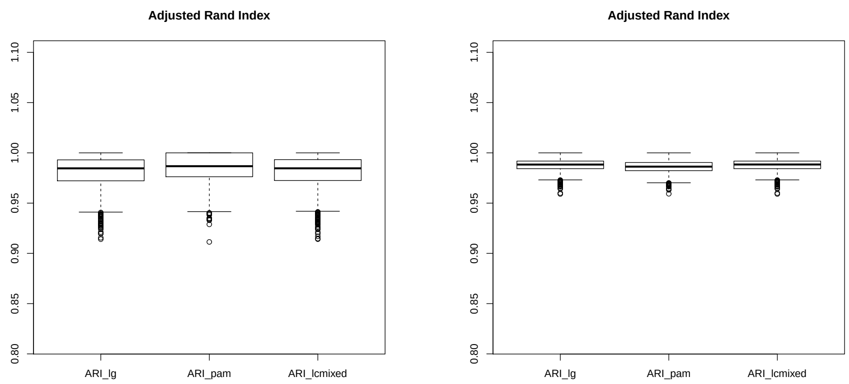
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.10	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10
2	0.15	0.70	0.10	0.10	0.10	0.10	0.10	0.70	0.10
3	0.20	0.10	0.70	0.10	0.10	0.10	0.70	0.10	0.10
4	0.25	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
5	0.30	0.10	0.10	0.10	0.70	0.10	0.10	0.10	0.70

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.10	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
4	0.25	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.10	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
2	0.15	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.05	0.05	0.65	0.05	0.05	0.05
4	0.25	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.30	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05

Table A.123: Summary: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.914	0.911	0.914
1st Qu.		0.972	0.976	0.972
Median		0.985	0.987	0.985
Mean (se)		0.982 (0.000)	0.985 (0.000)	0.982 (0.000)
3rd Qu.		0.993	1.000	0.993
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.400	0.406	0.407	0.406
1st Qu.	0.463	0.464	0.464	0.464
Median	0.475	0.475	0.476	0.475
Mean (se)	0.475 (0.000)	0.476 (0.000)	0.476 (0.000)	0.476 (0.000)
3rd Qu.	0.488	0.488	0.488	0.488
Max.	0.530	0.530	0.530	0.530
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.764	0.770	0.770	0.770
1st Qu.	0.811	0.811	0.811	0.811
Median	0.819	0.819	0.820	0.819
Mean (se)	0.819 (0.000)	0.819 (0.000)	0.819 (0.000)	0.819 (0.000)
3rd Qu.	0.827	0.828	0.828	0.828
Max.	0.857	0.858	0.858	0.858

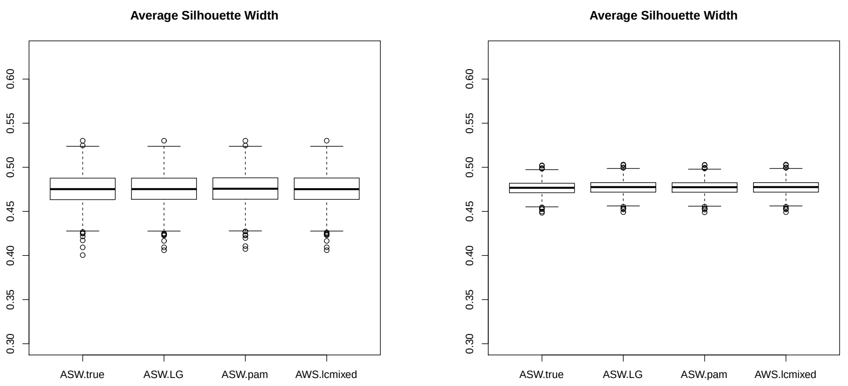


(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

Figure A.184: Adjusted Rand Index: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

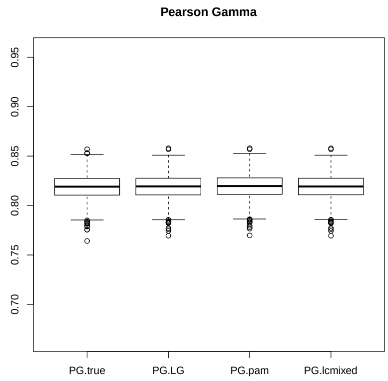
Table A.124: Summary: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.959	0.959	0.959
1st Qu.		0.984	0.982	0.984
Median		0.988	0.986	0.988
Mean (se)		0.988 (0.000)	0.986 (0.000)	0.988 (0.000)
3rd Qu.		0.992	0.990	0.992
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.448	0.449	0.449	0.449
1st Qu.	0.471	0.472	0.472	0.472
Median	0.477	0.478	0.477	0.478
Mean (se)	0.477 (0.000)	0.478 (0.000)	0.477 (0.000)	0.478 (0.000)
3rd Qu.	0.482	0.483	0.482	0.483
Max.	0.502	0.503	0.503	0.503
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.801	0.802	0.802	0.802
1st Qu.	0.815	0.816	0.816	0.816
Median	0.819	0.820	0.820	0.820
Mean (se)	0.819 (0.000)	0.820 (0.000)	0.820 (0.000)	0.820 (0.000)
3rd Qu.	0.822	0.823	0.823	0.823
Max.	0.836	0.837	0.836	0.837

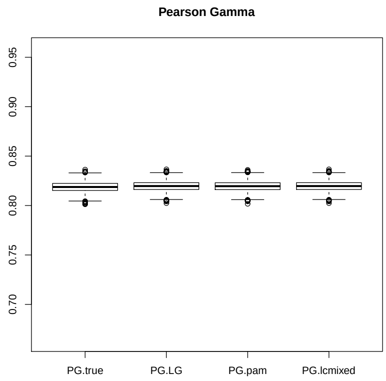


(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.185: Average Silhouette Width: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation



(a) *PG - 200 units.*



(b) *PG - 1000 units.*

Figure A.186: Pearson Gamma: 12 mixed-level variables - 5 clusters, mixing proportions extremely different - Clear cluster separation

A.63 12 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.20	0.30	0.70	0.25	0.75	0.40	0.60
2	0.20	0.80	0.20	0.40	0.60	0.30	0.70
3	0.20	0.55	0.45	0.80	0.20	0.80	0.20
4	0.20	0.60	0.40	0.30	0.70	0.35	0.65
5	0.20	0.25	0.75	0.70	0.30	0.75	0.25

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.20	0.20	0.30	0.50	0.20	0.50	0.30	0.40	0.35	0.25
2	0.20	0.10	0.40	0.50	0.35	0.45	0.20	0.20	0.45	0.35
3	0.20	0.50	0.25	0.25	0.50	0.30	0.20	0.50	0.20	0.30
4	0.20	0.40	0.40	0.20	0.30	0.20	0.50	0.30	0.10	0.60
5	0.20	0.55	0.35	0.10	0.40	0.40	0.20	0.30	0.50	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.20	0.20	0.10	0.40	0.30	0.30	0.20	0.40	0.10
2	0.20	0.40	0.30	0.20	0.10	0.10	0.30	0.20	0.40
3	0.20	0.10	0.40	0.30	0.20	0.20	0.40	0.10	0.30
4	0.20	0.30	0.20	0.10	0.40	0.40	0.10	0.30	0.20
5	0.20	0.20	0.30	0.10	0.40	0.30	0.10	0.40	0.20

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.20	0.40	0.20	0.10	0.30	0.10	0.20	0.30	0.40
2	0.20	0.10	0.20	0.40	0.30	0.40	0.10	0.20	0.30
3	0.20	0.20	0.30	0.40	0.10	0.20	0.40	0.30	0.10
4	0.20	0.30	0.10	0.20	0.40	0.30	0.20	0.40	0.10
5	0.20	0.10	0.40	0.30	0.20	0.20	0.10	0.40	0.30

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.20	0.05	0.15	0.10	0.20	0.10	0.05	0.20	0.15
2	0.20	0.20	0.05	0.20	0.15	0.10	0.15	0.10	0.05
3	0.20	0.10	0.20	0.05	0.05	0.15	0.10	0.15	0.20
4	0.20	0.15	0.10	0.15	0.10	0.20	0.05	0.05	0.20
5	0.20	0.20	0.15	0.10	0.05	0.05	0.20	0.15	0.10

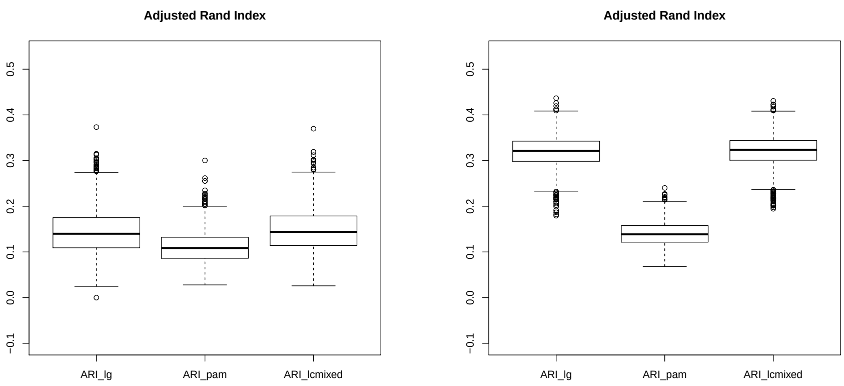
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.20	0.10	0.05	0.20	0.15	0.10	0.15	0.05	0.20
2	0.20	0.20	0.15	0.05	0.10	0.15	0.20	0.10	0.05
3	0.20	0.05	0.10	0.10	0.05	0.20	0.15	0.20	0.15
4	0.20	0.15	0.20	0.15	0.20	0.05	0.05	0.10	0.10
5	0.20	0.10	0.05	0.20	0.05	0.15	0.10	0.20	0.15

Table A.125: Summary: 12 mixed-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.028	0.026
1st Qu.		0.109	0.086	0.114
Median		0.140	0.109	0.144
Mean (se)		0.145 (0.001)	0.111 (0.001)	0.149 (0.001)
3rd Qu.		0.175	0.132	0.179
Max.		0.373	0.300	0.370
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.002	0.023	0.039	0.034
1st Qu.	0.018	0.055	0.057	0.057
Median	0.023	0.061	0.061	0.063
Mean (se)	0.023 (0.000)	0.061 (0.000)	0.061 (0.000)	0.063 (0.000)
3rd Qu.	0.027	0.067	0.066	0.069
Max.	0.051	0.093	0.085	0.091
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.145	0.179	0.229	0.193
1st Qu.	0.192	0.263	0.275	0.267
Median	0.205	0.282	0.287	0.286
Mean (se)	0.205 (0.000)	0.281 (0.001)	0.287 (0.000)	0.285 (0.001)
3rd Qu.	0.218	0.300	0.299	0.303
Max.	0.277	0.370	0.347	0.373
NA's		1		

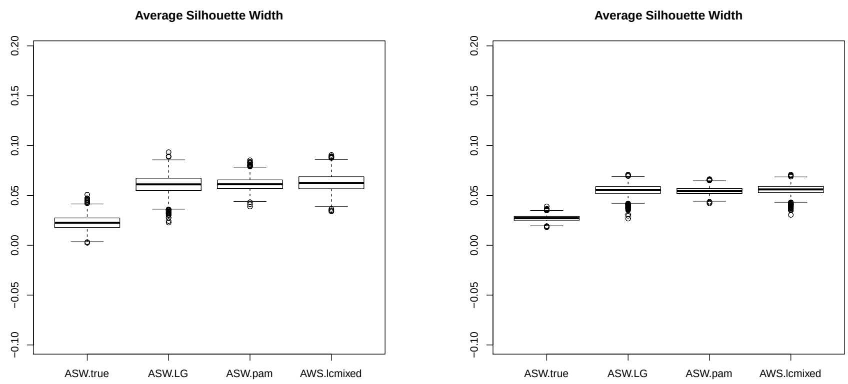
Table A.126: Summary: 12 mixed-level variables - 5 clusters - equal mixing proportions - Unclear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.179	0.068	0.194
1st Qu.		0.298	0.121	0.301
Median		0.321	0.138	0.324
Mean (se)		0.319 (0.001)	0.141 (0.001)	0.321 (0.001)
3rd Qu.		0.342	0.158	0.344
Max.		0.437	0.240	0.431
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.018	0.027	0.042	0.030
1st Qu.	0.025	0.052	0.052	0.053
Median	0.027	0.056	0.055	0.056
Mean (se)	0.027 (0.000)	0.055 (0.000)	0.054 (0.000)	0.056 (0.000)
3rd Qu.	0.029	0.059	0.057	0.059
Max.	0.039	0.071	0.066	0.071
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.178	0.191	0.225	0.206
1st Qu.	0.200	0.265	0.258	0.266
Median	0.206	0.274	0.266	0.274
Mean (se)	0.206 (0.000)	0.273 (0.000)	0.265 (0.000)	0.274 (0.000)
3rd Qu.	0.211	0.282	0.272	0.282
Max.	0.238	0.313	0.297	0.313



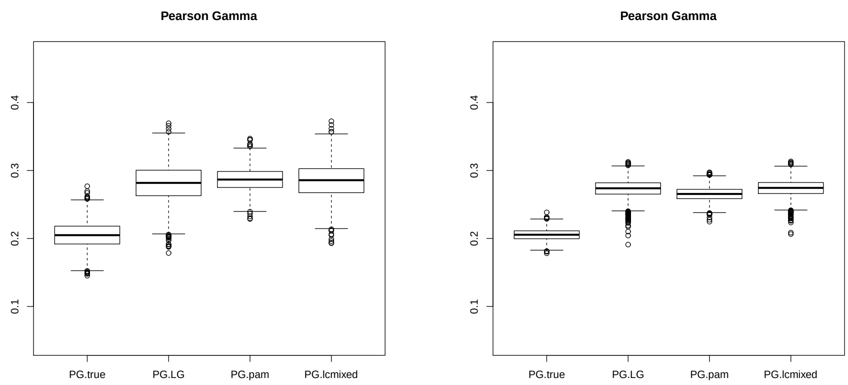
(a) *ARI - 200 units.* (b) *ARI - 1000 units.*

Figure A.187: Adjusted Rand Index: 12 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.188: Average Silhouette Width: 12 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.189: Pearson Gamma: 12 mixed-level variables - 5 clusters, equal mixing proportions - Unclear cluster separation

A.64 12 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

The Model:

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{A}_1 X}$	$\hat{\pi}_{1t}^{\bar{A}_2 X}$	$\hat{\pi}_{1t}^{\bar{B}_1 X}$	$\hat{\pi}_{1t}^{\bar{B}_2 X}$	$\hat{\pi}_{1t}^{\bar{C}_1 X}$	$\hat{\pi}_{1t}^{\bar{C}_2 X}$
1	0.20	0.10	0.90	0.10	0.90	0.90	0.10
2	0.20	0.10	0.90	0.90	0.10	0.10	0.90
3	0.20	0.90	0.10	0.90	0.10	0.10	0.90
4	0.20	0.10	0.90	0.90	0.10	0.90	0.10
5	0.20	0.90	0.10	0.10	0.90	0.10	0.90

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{D}_1 X}$	$\hat{\pi}_{1t}^{\bar{D}_2 X}$	$\hat{\pi}_{1t}^{\bar{D}_3 X}$	$\hat{\pi}_{1t}^{\bar{E}_1 X}$	$\hat{\pi}_{1t}^{\bar{E}_2 X}$	$\hat{\pi}_{1t}^{\bar{E}_3 X}$	$\hat{\pi}_{1t}^{\bar{F}_1 X}$	$\hat{\pi}_{1t}^{\bar{F}_2 X}$	$\hat{\pi}_{1t}^{\bar{F}_3 X}$
1	0.20	0.05	0.05	0.90	0.90	0.05	0.05	0.05	0.90	0.05
2	0.20	0.90	0.05	0.05	0.05	0.05	0.90	0.90	0.05	0.05
3	0.20	0.05	0.05	0.90	0.05	0.90	0.05	0.05	0.05	0.90
4	0.20	0.90	0.05	0.05	0.05	0.90	0.05	0.05	0.90	0.05
5	0.20	0.05	0.90	0.05	0.90	0.05	0.05	0.90	0.05	0.05

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{G}_1 X}$	$\hat{\pi}_{1t}^{\bar{G}_2 X}$	$\hat{\pi}_{1t}^{\bar{G}_3 X}$	$\hat{\pi}_{1t}^{\bar{G}_4 X}$	$\hat{\pi}_{1t}^{\bar{H}_1 X}$	$\hat{\pi}_{1t}^{\bar{H}_2 X}$	$\hat{\pi}_{1t}^{\bar{H}_3 X}$	$\hat{\pi}_{1t}^{\bar{H}_4 X}$
1	0.20	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10
2	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
3	0.20	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
4	0.20	0.10	0.10	0.70	0.10	0.70	0.10	0.10	0.10
5	0.20	0.10	0.70	0.10	0.10	0.10	0.70	0.10	0.10

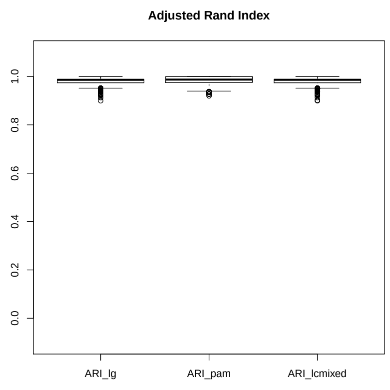
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{I}_1 X}$	$\hat{\pi}_{1t}^{\bar{I}_2 X}$	$\hat{\pi}_{1t}^{\bar{I}_3 X}$	$\hat{\pi}_{1t}^{\bar{I}_4 X}$	$\hat{\pi}_{1t}^{\bar{L}_1 X}$	$\hat{\pi}_{1t}^{\bar{L}_2 X}$	$\hat{\pi}_{1t}^{\bar{L}_3 X}$	$\hat{\pi}_{1t}^{\bar{L}_4 X}$
1	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.70	0.10
2	0.20	0.70	0.10	0.10	0.10	0.10	0.10	0.10	0.70
3	0.20	0.10	0.10	0.70	0.10	0.10	0.70	0.10	0.10
4	0.20	0.10	0.70	0.10	0.10	0.10	0.10	0.10	0.70
5	0.20	0.10	0.10	0.10	0.70	0.70	0.10	0.10	0.10

Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{M}_1 X}$	$\hat{\pi}_{1t}^{\bar{M}_2 X}$	$\hat{\pi}_{1t}^{\bar{M}_3 X}$	$\hat{\pi}_{1t}^{\bar{M}_4 X}$	$\hat{\pi}_{1t}^{\bar{M}_5 X}$	$\hat{\pi}_{1t}^{\bar{M}_6 X}$	$\hat{\pi}_{1t}^{\bar{M}_7 X}$	$\hat{\pi}_{1t}^{\bar{M}_8 X}$
1	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05
3	0.20	0.05	0.05	0.65	0.05	0.05	0.05	0.05	0.05
4	0.20	0.05	0.05	0.05	0.05	0.05	0.65	0.05	0.05
5	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65

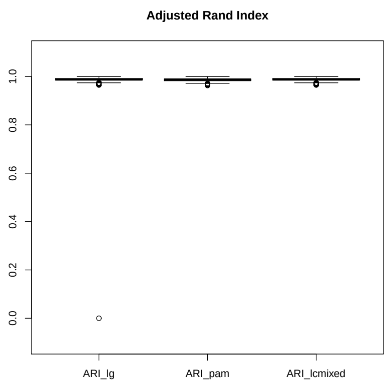
Latent class	$\hat{\pi}_t^X$	$\hat{\pi}_{1t}^{\bar{N}_1 X}$	$\hat{\pi}_{1t}^{\bar{N}_2 X}$	$\hat{\pi}_{1t}^{\bar{N}_3 X}$	$\hat{\pi}_{1t}^{\bar{N}_4 X}$	$\hat{\pi}_{1t}^{\bar{N}_5 X}$	$\hat{\pi}_{1t}^{\bar{N}_6 X}$	$\hat{\pi}_{1t}^{\bar{N}_7 X}$	$\hat{\pi}_{1t}^{\bar{N}_8 X}$
1	0.20	0.05	0.65	0.05	0.05	0.05	0.05	0.05	0.05
2	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.65	0.05
3	0.20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.65
4	0.20	0.65	0.05	0.05	0.05	0.05	0.05	0.05	0.05
5	0.20	0.05	0.05	0.05	0.65	0.05	0.05	0.05	0.05

Table A.127: Summary: 12 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 200 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.900	0.920	0.900
1st Qu.		0.974	0.975	0.974
Median		0.986	0.987	0.987
Mean (se)		0.981 (0.000)	0.985 (0.000)	0.981 (0.000)
3rd Qu.		0.989	1.000	0.989
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.412	0.410	0.412	0.410
1st Qu.	0.464	0.464	0.465	0.464
Median	0.476	0.476	0.477	0.476
Mean (se)	0.476 (0.000)	0.477 (0.000)	0.477 (0.000)	0.477 (0.000)
3rd Qu.	0.489	0.490	0.490	0.490
Max.	0.534	0.534	0.534	0.534
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.762	0.762	0.768	0.765
1st Qu.	0.803	0.803	0.804	0.803
Median	0.811	0.811	0.811	0.811
Mean (se)	0.811 (0.000)	0.811 (0.000)	0.811 (0.000)	0.811 (0.000)
3rd Qu.	0.819	0.819	0.819	0.819
Max.	0.847	0.847	0.847	0.847



(a) *ARI - 200 units.*

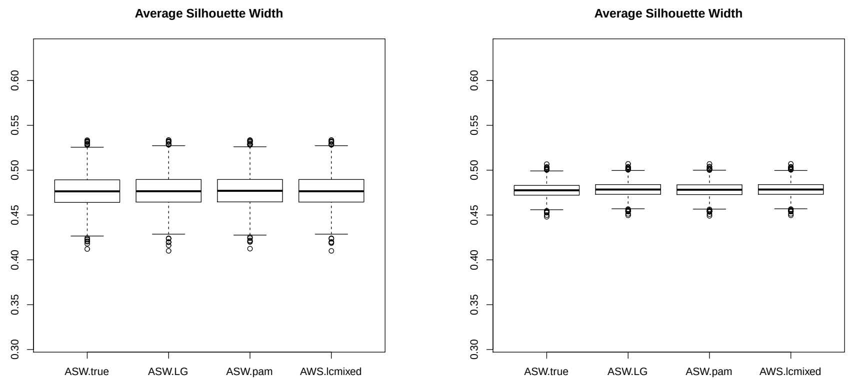


(b) *ARI - 1000 units.*

Figure A.190: Adjusted Rand Index: 12 mix-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

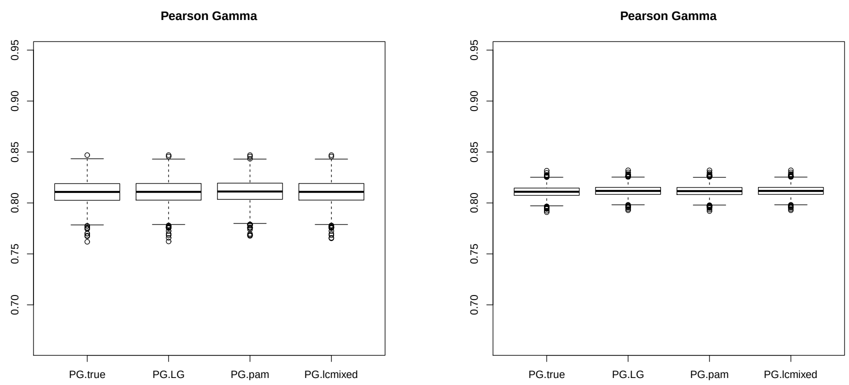
Table A.128: Summary: 12 mixed-level variables - 5 clusters, equal mixing proportions - Clear cluster separation - 1000 units

		ARI lg	ARI pam	ARI lcmixed
Min.		0.000	0.963	0.965
1st Qu.		0.985	0.983	0.985
Median		0.988	0.987	0.988
Mean (se)		0.987 (0.001)	0.986 (0.000)	0.988 (0.000)
3rd Qu.		0.992	0.990	0.992
Max.		1.000	1.000	1.000
	ASW true	ASW lg	ASW pam	ASW lcmixed
Min.	0.448	0.450	0.449	0.450
1st Qu.	0.472	0.473	0.473	0.473
Median	0.478	0.478	0.478	0.478
Mean (se)	0.478 (0.000)	0.478 (0.000)	0.478 (0.000)	0.478 (0.000)
3rd Qu.	0.483	0.484	0.484	0.484
Max.	0.507	0.507	0.507	0.507
NA's		1		
	PG true	PG lg	PG pam	PG lcmixed
Min.	0.791	0.793	0.792	0.793
1st Qu.	0.808	0.809	0.808	0.809
Median	0.811	0.812	0.812	0.812
Mean (se)	0.811 (0.000)	0.812 (0.000)	0.812 (0.000)	0.812 (0.000)
3rd Qu.	0.815	0.815	0.815	0.815
Max.	0.832	0.832	0.832	0.832
NA's		1		



(a) *ASW - 200 units.* (b) *ASW - 1000 units.*

Figure A.191: Average Silhouette Width: 12 mix-level variables - 5 clusters, equal mixing proportions - Clear cluster separation



(a) *PG - 200 units.* (b) *PG - 1000 units.*

Figure A.192: Pearson Gamma: 12 mix-level variables - 5 clusters, equal mixing proportions - Clear cluster separation

Bibliography

[1] Alan Agresti. *Categorical Data Analysis*. Wiley, 2002.

[2] M. Aitkin, B. Francis, and Raynal N. Une étude comparative d'analyses des correspondances ou de classifications et des modèles de variables latentes ou de classe latentes. *Revue de statistique appliquée*, 35(3):53–81, 1987.

[3] Elizabeth S. Allman, Catherine Matias, and John a. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37(6A):3099–3132, December 2009.

[4] Renan Moritz Varnier R Almeida, Antonio Fernando C Infantosi, José Hermógenes R Suassuna, and João Carlos G D Costa. Multiple correspondence analysis in predictive logistic modelling: application to a living-donor kidney transplantation data. *Computer Methods and Programs in Biomedicine*, 95(2):116–28, August 2009.

[5] T. W. Anderson. On estimation of parameters in latent structure analysis. *Psychometrika*, (19):1–10, 1954.

[6] Jeffrey D. Banfield and Adrian E. Raftery. Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, 49(3):803–821, 1993.

[7] Jean-Patrick Baudry, Adrian E Raftery, Gilles Celeux, Kenneth Lo, and Raphaël Gottardo. Combining Mixture Components for Clustering. *Journal of computational and graphical statistics*, 19(2):332–353, June 2010.

[8] Christopher R Bilder and Joshua M Tebbs. An Introduction to Categorical Data Analysis An Introduction to Categorical Data Analysis. *Journal*

of the American Statistical Association, 103(483):1323–1323, September 2008.

[9] Shyam Boriah. Similarity Measures for Categorical Data : A Comparative Evaluation.

[10] Peter Bryant and John a. Williamson. Asymptotic Behaviour of Classification Maximum Likelihood Estimates. *Biometrika*, 65(2):273–281, August 1978.

[11] Peter Bryant and John a. Williamson. On Characterizing Optimization-Based Clustering Methods. *Journal of Classification*, 5:81–84, 1988.

[12] Miguel A. Carreira-Perpinan and S. Renals. Practical identifiability of finite mixtures of multivariate bernoulli distributions. *Neural computation*, 12(1):141–52, January 2000.

[13] Gilles Celeux. Classification et modèles. *Revue de Statistique Appliquée*, 36(4):43–57, 1988.

[14] Gilles Celeux and Gérard Govaert. Clustering Criteria for Discrete Data and Latent Class Models. *Journal of Classification*, 8:157–176, 1991.

[15] Clifford C. Clogg and Leo A. Goodman. Latent Structure Analysis of a Set of Multidimensional Contingency Tables. *Journal of the American Statistical Association*, 79(388):762– 771, 1984.

[16] Trevor F. Cox and Michael A.A. Cox. *Multidimensional Scaling*. Chapman & Hall/CRC, second edition, 2001.

[17] A.P. Dempster, N.M. Laird, and D.B. Rubin.

[18] Wayne S. DeSarbo and William L. Cron. A Maximum Likelihood Methodology for Clusterwise Linear Regression. *Journal of Classification*, 5:249–282, 1988.

[19] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., second edition, 2001.

[20] R. Elmore, P. Hall, and A. Neeman. An application of classical invariant theory to identifiability in nonparametric mixtures. *Ann.Inst.Fourier (Grenoble)*, 55:1–28, 2005.

[21] Ryan Elmore and Shaoli Wang. Identifiability and Estimation in Finite Mixture Models with Multinomial Components. Technical report, Department of Statistics, The Pennsylvania State University, University Park, PA 16802, USA, 2003.

[22] Brian S. Everitt, Sabine Landau, Morven Leese, and Daniel Stahl. *Cluster Analysis*, volume 14. John Wiley & Sons, Ltd, 5th edition, January 2011.

[23] B.S. Everitt and D.J. Hand. *Finite Mixture Distributions*. Chapman and Hall, 1981.

[24] Anton K. Formann. Measuring Change in Latent Subgroups Using Dichotomous Data: Unconditional , Conditional , and Semiparametric Maximum Likelihood Estimation. *Journal of the American Statistical Association*, 89(427):1027– 1034, 1994.

[25] Anton K. Formann and Thomas Kohlmann. Latent Class Analysis in Medical Research. *Statistical Methods in Medical Research*, (5):179– 211, 1996.

[26] H. Friedman and J. Rubin. On some criterion for grouping data. *Journal of the American Statistical Association*, 62, 1967.

[27] Leo a. Goodman. Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika*, 61(2):215, August 1974.

[28] Leo A. Goodman. The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable . Part I-A Modified Latent Structure Approach. *The American Journal of Sociology*, 79(5):1179–1259, 1974.

[29] A.D. Gordon. *Classification - 2nd Edition*. Chapman & Hall, 1999.

[30] J. C. Gower. A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4):857–871, 1971.

[31] John Gower, Sugnet Gardner-Lubbe, and Niël le Roux. *Understanding Biplots*. John Wiley & Sons, Ltd, first edition, 2011.

[32] Mats Gyllenberg, Timo Koski, Edwin Reinlink, and Martin Verlaan. Non-Uniqueness in Probabilistic Numerical Identification of Bacteria. *Journal of Applied Probability*, 31(2):542–548, 1994.

[33] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. Cluster Validity Methods : Part I.

[34] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, (17):107–145, 2001.

[35] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. Clustering validity checking methods. *ACM SIGMOD Record*, 31(3):19, September 2002.

[36] J.A. Hartigan and P.M. Hartigan. The Dip Test of Unimodality. *The Annals of Statistics*, 13(1):70–84, 1985.

[37] Ton Heinen. *Latent class and discrete latent trait models : similarities and differences*. Advanced quantitative techniques in the social sciences. Thousand Oaks, Calif : Sage Publications, 1996.

[38] Christian Hennig. Identifiability of Models for Clusterwise Linear REgression. *Journal of Classification*, 17:273–296, 2000.

[39] Christian Hennig. Tutorial on mixture models (2). (September), 2009.

[40] Christian Hennig. Latent class and dissimilarity based clustering for mixed type variables with application to social stratification. *Philosophy*, 1, 2010.

[41] Christian Hennig and Tim F Liao. How to find an appropriate clustering for mixed type variables with application to socioeconomic stratification. 2010.

[42] Guan-Hua Huang and Karen Bandeen-Roche. Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, 69(1):5–32, March 2004.

[43] Lawrence Hubert and Phipps Arabie. Comparing Partitions. *Journal of Classification*, 2:193–218, 1985.

[44] L. Kaufman and P.J. Rouseeuw. *Finding Groups in Data*, volume 9. Wiley, New York, January 1990.

[45] Leonard Kaufman and Peter J. Rouseeuw. Clustering by Means of Medoids. In Y. Dodge, editor, *Statistical Data Analysis based on the L, Norm*, chapter 2.1, 2.5.5, pages 405–416. Elsevier/North-Holland, Amsterdam, 1987.

[46] J Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18(2):95–138, 1977.

[47] P.F. Lazarsfeld and N.W. Henry. *Latent Structure Analysis*. Boston: Houghton Mifflin, 1968.

[48] Bruce Lindsay, Clifford C. Clogg, and John Grego. Semiparametric Estimation in the Rasch Model and Related Exponential Response Models, Including a Simple Latent Class Model for Item Analysis. *Journal of the American Statistical Association*, 86(413):96– 107, 1991.

[49] Fernando Lourenço, Victor Lobo, and Fernando Bação. Binary-based similarity measures for categorical data and their application in Self-Organizing Maps 1 Introduction. *Measurement*, (Kohonen 2001):1–18, 2004.

[50] A. Madansky. Determinantal methods in latent class analysis. *Psychometrika*, (25):183–198, 1960.

[51] J. Magidson and Vermunt J. Latent Class Modeling as a Probabilistic Extension of K-Means Clustering. pages 3–9, 1990.

[52] Jay Magidson and Jeroen K Vermunt. Latent Class Factor and Cluster Models, Bi-Plots, and Related Graphical Displays. *Sociological Methodology*, 31(2001):223–264, 2001.

[53] Jay Magidson and Jeroen K Vermunt. Latent class models for clustering: A comparison with K-means. *Journal of Marketing*, 20, 2002.

[54] Kanti V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, 1979.

[55] G. McLachlan and D. Peel. *Finite Mixture Model*. John Wiley & Sons, 2000.

[56] M Meila. Comparing clustering-an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, May 2007.

[57] Glenn W. (The Ohio Sate University) Milligan and Martha C. (The Ohio Sate University) Cooper. An esamination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179, 1985.

[58] Bengt O. Muthén and Kerby Shedden. Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics*, 55(2):463–469, 1999.

[59] Karen L Nylund, Tihomir Asparouhov, and Bengt O. Muthén. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling*, 14(4):535–569, 2007.

[60] Adrian E Raftery and Nema Dean. Variable Selection for Model-Based Clustering. *Journal of the American Statistical Association*, 101(473):168–178, March 2006.

[61] B F Rancis and N R Aynal. Evue de statistique appliquée. *Revue*, 1987.

[62] William M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[63] Richard Redner. Note on the Consistency of the Maximum Likelihood Estimate for Nonidentifiable Distributions. *The Annals of Statistics*, 9(1):225–228, 1981.

[64] Samuel A. Stouffer and Jackson Toby. Role Conflict and Personality. *American Journal of Sociology*, 56(5):395–406, 1951.

[65] Henry Teicher. Identifiability of Mixtures. *The Annals of Mathematical Statistics*, 32:244–248, 1961.

[66] Henry Teicher. Identifiability of Finite Mixtures. *The Annals of Mathematical Statistics*, 34:1265–1269, 1963.

[67] Henry Teicher. Identifiability of Mixtures of Product Measures. *The Annals of Mathematical Statistics*, 38(4):1300–1302, 1967.

[68] D. Titterington, A. Smith, and Makov U.

[69] John S Uebersax and William M Grove. A Latent Trait Finite Mixture Model for the Analysis of Rating Agreement. *Biometrics*, 49(3):823–835, 1993.

[70] Stef van Buuren and Willem J. Heiser. Clustering n objects into k groups under optimal scaling of variables. *Psychometrika*, 54(4):699–706, December 1989.

[71] Peter G.M. Van der Heijden, Zvi Gilula, and L. Andries van der Ark. An Extended Study into the Relationship between Correspondence Analysis and Latent Class Analysis. *Sociological Methodology*, 29:147–186, 1999.

[72] W Vanwieringen. On identifiability of certain latent class models. *Statistics & Probability Letters*, 75(3):211–218, December 2005.

[73] Jeroen K Vermunt and Jay Magidson. *Latent GOLD 4.0 User's Guide 1*. 2005.

[74] Jeroen K Vermunt and Jay Magidson. Manual for Latent Gold 4.5 Syntax Module. In *Latent Gold User Guide*. 2005.

[75] Jeroen K Vermunt and Jay Magidson. *Technical Guide for Latent GOLD 4.0: Basic and Advanced 1*. 2005.

[76] Hana Řezanková. Cluster analysis and categorical data. *Statistika*, pages 216–232.

[77] Abraham Wald. Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949.

[78] John Wolfe. Pattern Clustering By Multivariate Mixture Analysis. *Multivariate Behavioral Research*, 5(3):329–350, April 1970.

[79] Sidney J Yakowitz and John D Spragins. On the Identifiability of Finite Mixtures. *The Annals of Mathematical Statistics*, 39(1):209–214, 1968.