Team Project Deliverable 2

Joan Cortes - 100438579
Jose Antonio Jijon Vorbeck - 100438045
Didier Dirks - 100443386

22-01-2021

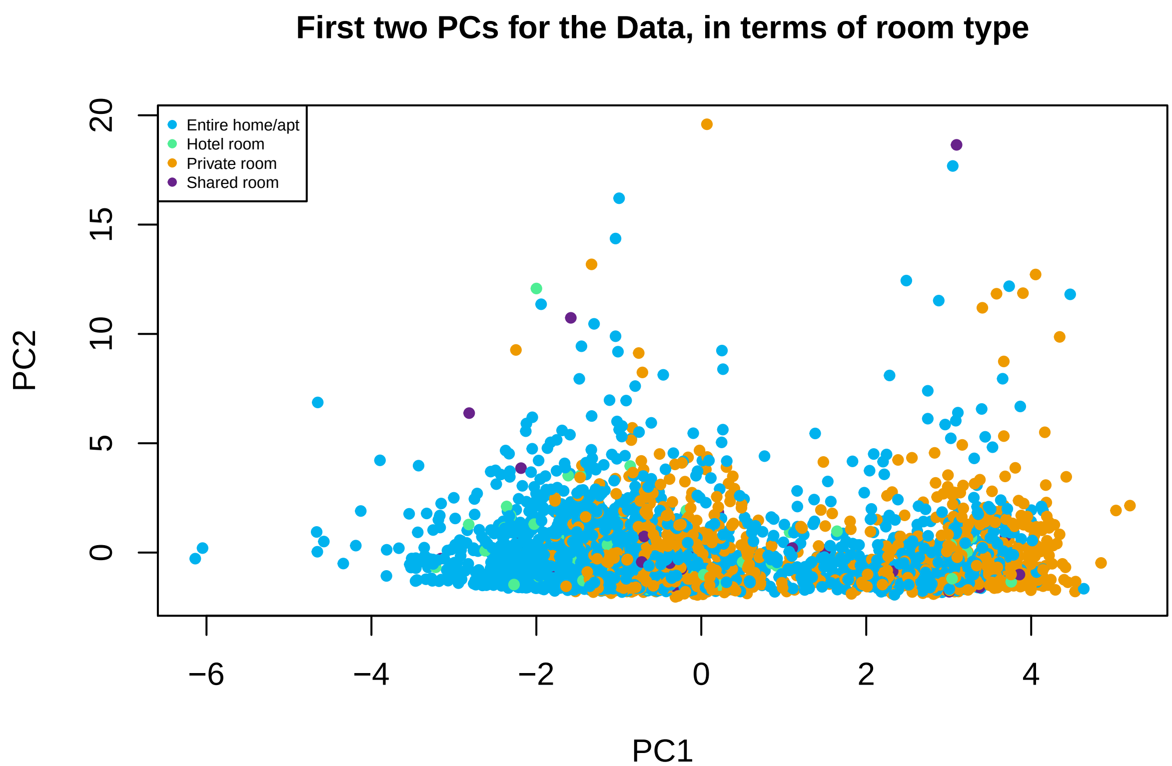
This report will be divided into two main sections. The first part will be devoted to unsupervised learning using several methodologies to make clusters learned in class. The second part of the report will be devoted to a supervised learning using methods like KNN, bayes theorems and logistic regressions.

The data that we will be using throughout the analysis is the same as the one for the previous step of the project, it consist of AirBnB listings along with many characteristics about their location, owner info and services included. In the first part of the project we have already done the preprocessing of the data (imputation of missing data, selection of meaningful attributes, removal of non realistic data etc.). We will take this already modified data and use it for the rest of the project.

For computation purposes, we had to take only a smaller part of the original data, since it contained more than 20k observations, and our personal computers where taking more than 30 mins to compute the silhouette of each of the examples. With a smaller dataset, we kept the proportions of the groups, and tried to stay with the same overall characteristics, by splitting the data randomly.

The categorical variable of interest that we have chosen for making the true clusters is "room_type". This variable contains 4 classes, which are: 'Entire home/apt', 'Hotel room', 'Private Room' and 'Shared room'. Most of the instances are divided into entire homes or private rooms, and the other two categories combined add up to only 3.5% of the whole dataset.

We have performed the principal component analysis for the selected dataset, and plotted the first two principal components in terms of room_type. Below we see the graph.



We can see that ther is not a very strong separation when we choose the grouping criterion to be the room_type, maybe because there are many other characteristics that will have a greater impact on the separation of the groups. This is nevertheless very interesting, since we will see, throught the report, how the different algorithms and methods chosen will make different groups out of the same data, according to its different criterions.

1. Cluster analysis (unsupervised classification).

For that you have to skip the categorical variable of interest for carrying out supervised classification. It is important to note that the groups obtained with the clustering may not be those that define the categorical variable of interest. Obtain conclusions from the analysis.

The unsupervised cluster analysis can be done with several methodologies. In this report we will present the following ones:

- Partitional Clustering
  - K-means
  - K-medoids
- Hierarchical Clustering
  - Agglomerative
  - Divisive
- Model Based Clustering

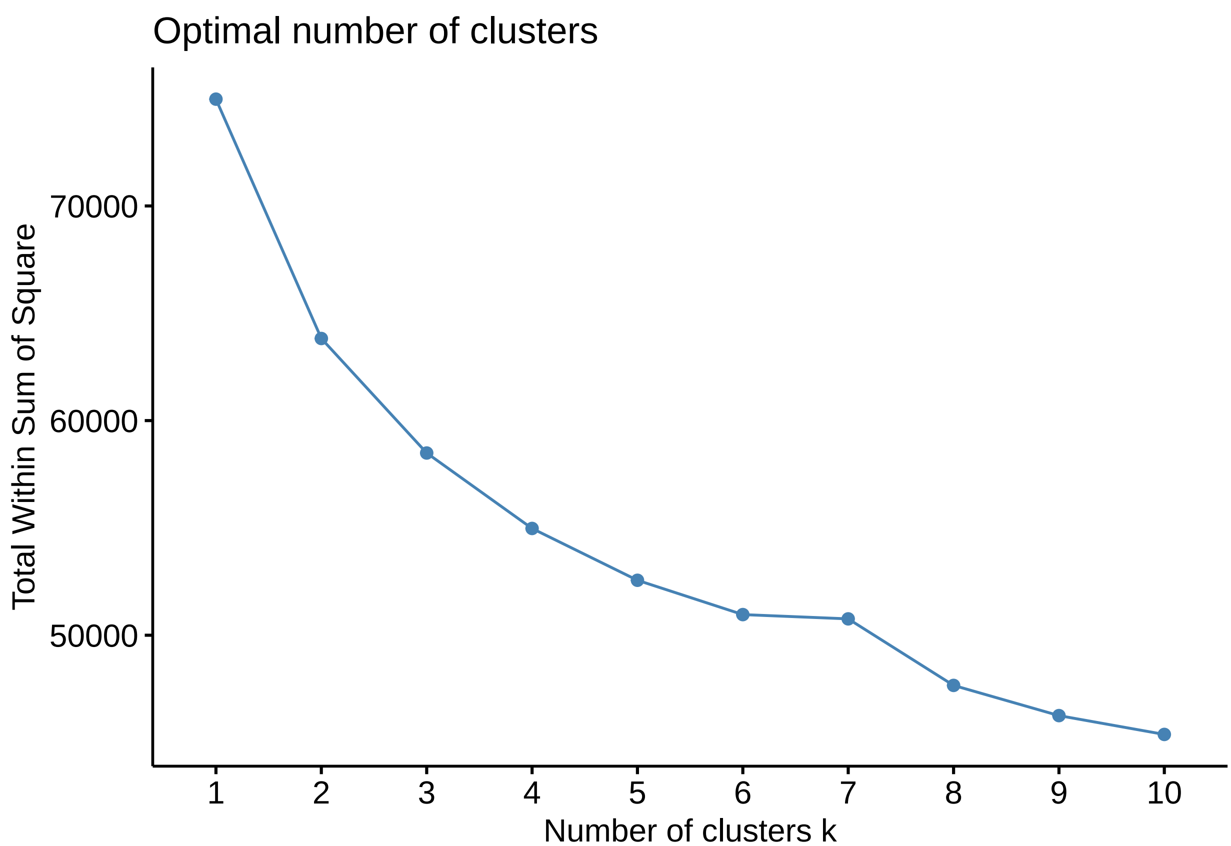
Partitional Clustering

Clustering refers to a wide range of methods for finding subgroups (clusters) in a dataset. The aim of these techniques is to group observations so the instances assigned to a particular group are quite homogenous, whereas the observations in the other groups are as different as possible from each other.

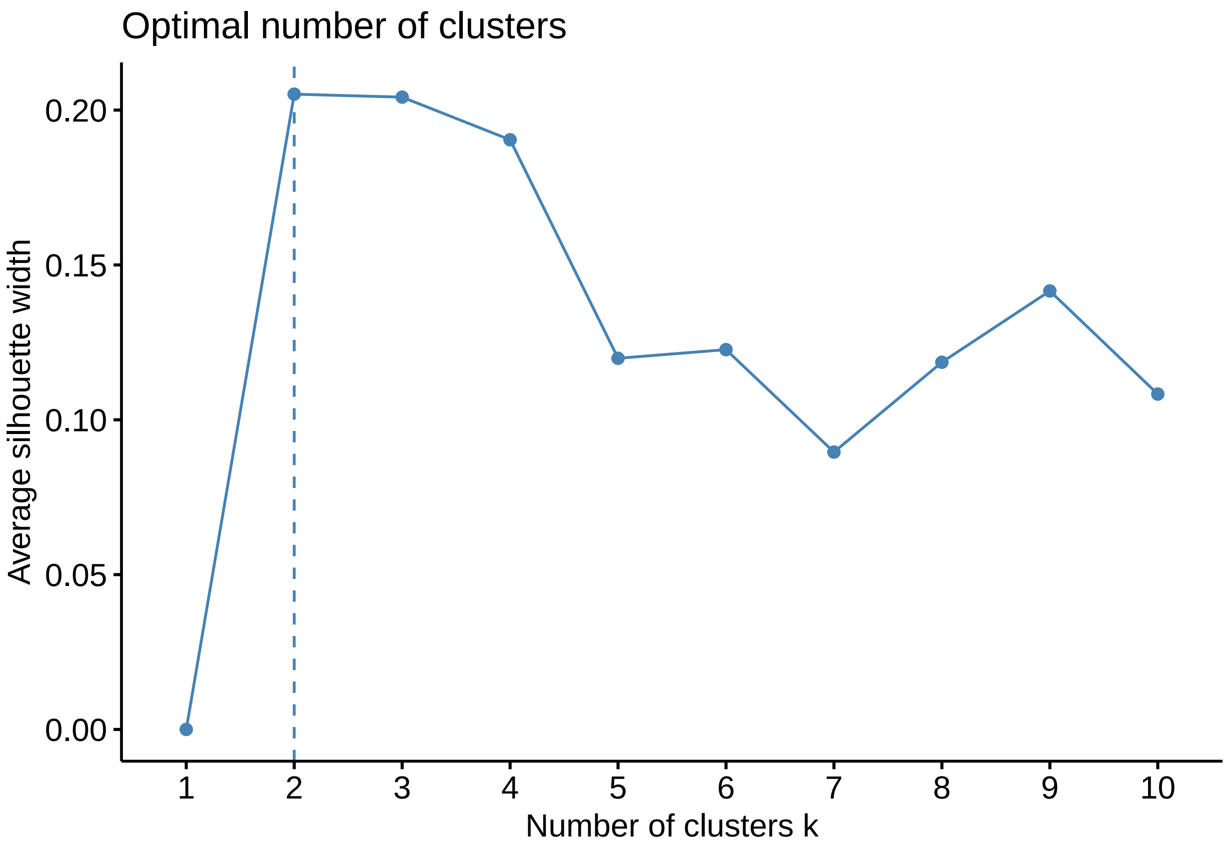
In this section we are going to focus on the family of methods belonging to Partitional clustering. The idea of Partitional clustering is to start with a particular partition and exchange observations until an optimal clustering structure is reached. We are going to employ the most standard algorithms belonging to Partitional clustering such us K-means, PAM, and CLARA.

We start by assessing the existence of subgroups in our dataset with the use of the K-means algorithm. For every specific number of groups selected, the algorithm will try to minimize the within-cluster variation. This is to say, the difference between each observation belonging to a specific group and the sample mean associated to that specific group. The algorithm then stops when stabilizes.

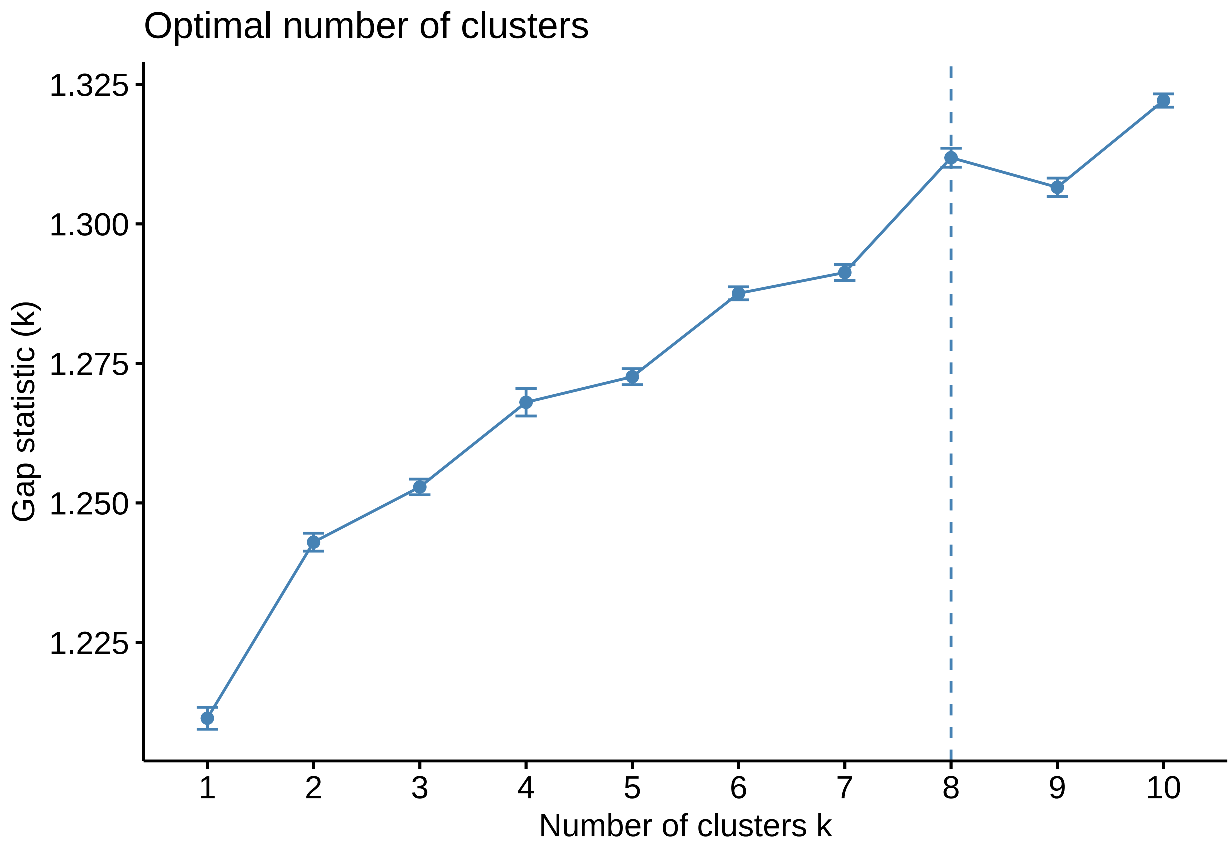
K-means We start by assessing the optimal K by conducting k-means for different number of K.



The results exhibited by the elbow graph are unclear. The within-cluster variation decreases smoothly as K increases (natural result). In order to assess the optimal number of K we are going to carry out a silhouette analysis. This technique provides more information regarding the goodness of fit of the clustering structure. This method delivers information not just about the distances within a given group but with respect to other groups. Therefore, well defined groups will yield better results than groups with week boundaries.

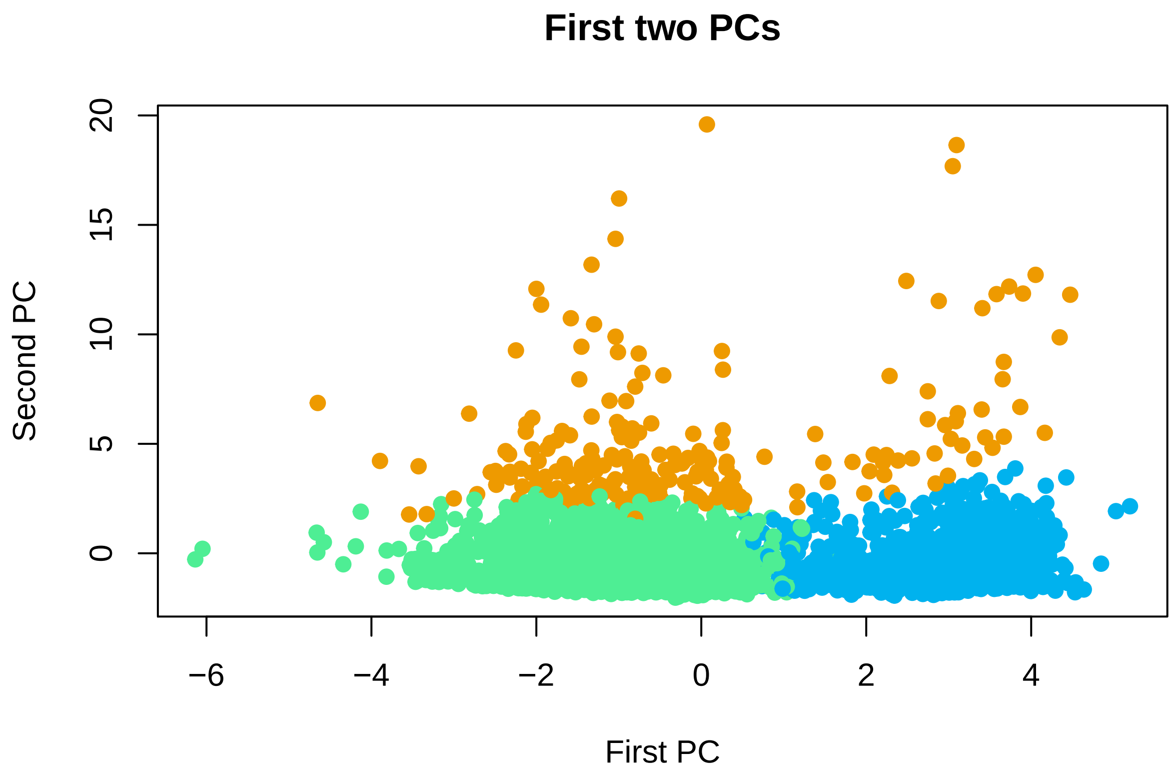


The results suggest that the optimal number of K is equal to two. Notwithstanding, the average silhouette for K 3 is really close to the one obtained for 2. Once again, the results are unclear. In order to get more insights about the optimal number of K we are going to employ the Gap Statistic. This technique compares the within-cluster variation with the expected value under assumption of K=1.



The results provided by the Gap-statistic suggest k=8. However, the results are unclear. There is not a well-defined local maximum.

Due to the fact that the previous results were unclear, we are going to conduct the K-means for K=3. Since the silhouette for k=3 was really close to the optimal and the rest of techniques assessed suggest a larger number of k.

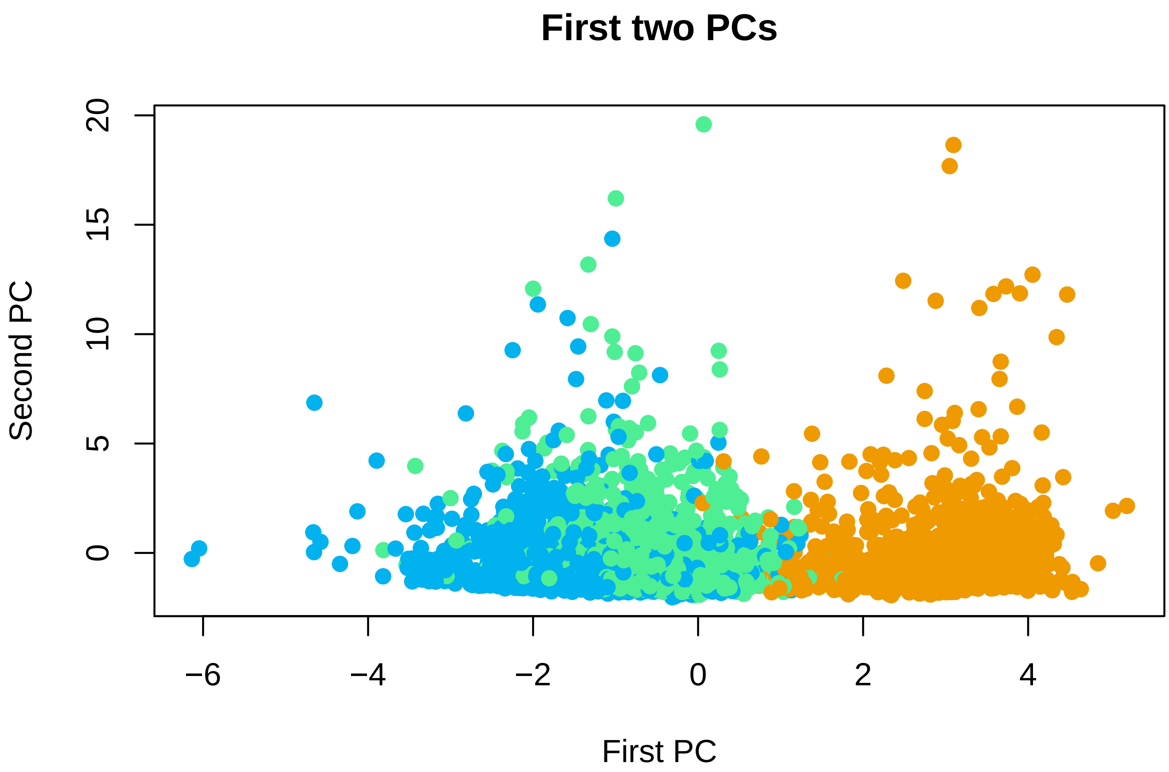


The plot above shows the distribution of each cluster. It can be seen three well differentiated groups. Two streams of data composed by a different groups each and on top of both streams a third cluster. Unlike the original clustering structure, where let's recall, we have two streams of data composed mainly by two groups (Private rooms and Entire Home) with a well-defined separation with each stream. In our case, each stream is assigned to a particular group.

```
## Silhouette of 3125 units in 3 clusters from silhouette.default(x = kmeans_X$cluster, dist = dist(sma
##  Cluster sizes and average silhouette widths:
##        854        2043         228
##  0.00852227  0.05132138 -0.14799621
## Individual silhouette widths:
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.2269297  0.0001056  0.0226093  0.0250830  0.0783749  0.0982671
```

The average silhouette obtained for the previous case is really close to 0. This results suggest that the partitions obtained are week.

Due to the unclear results obtained so far we are going to carry out a clustering structure analysis with the use of PAM and CLARA. PAM and CLARA unlike K-means use the most central point of a group instead of the sample mean. Despite being computational more costly than K-means it allows us to take into consideration categorical variables as well with the use of Gower distances.



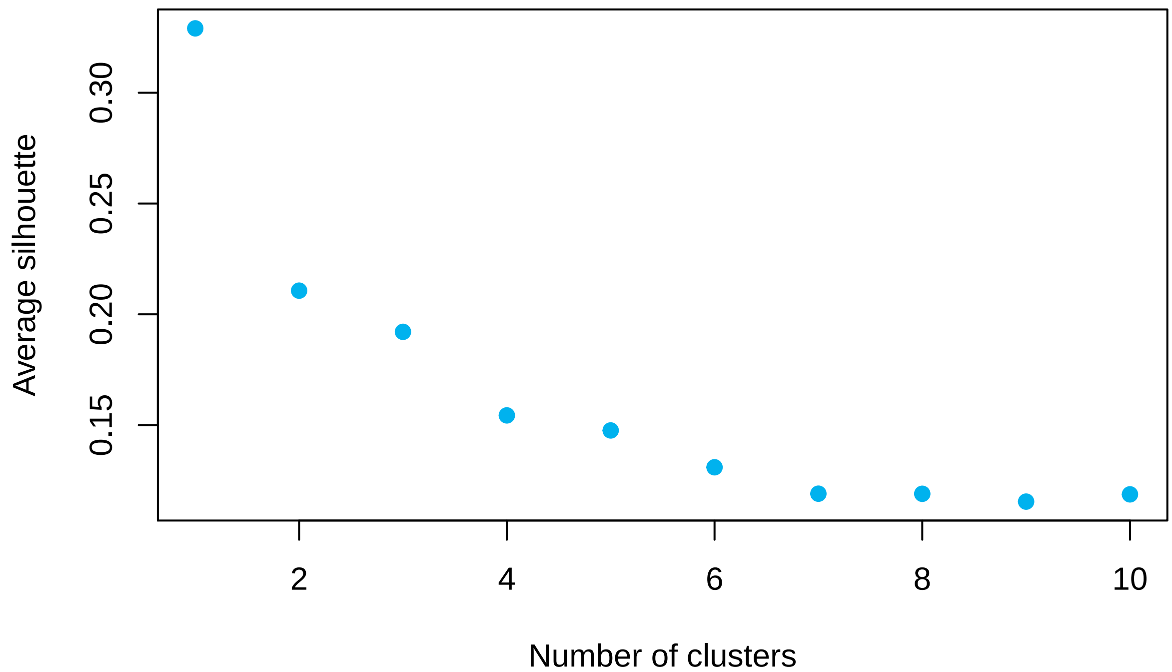
PAM

The results obtained with PAM differ from the ones obtained with K-means. As it can been depicted above, one of the streams of data is assigned to one group, whereas the other stream of data is assigned to the other two groups. The distribution of the clusters within this second stream of data is closer to the original distribution that the one indicated by the K-means algorithm.

```
## Silhouette of 3125 units in 3 clusters from silhouette.default(x = pam_data$cluster, dist = dist(smal
##  Cluster sizes and average silhouette widths:
##        1162        1072         891
##  0.19716402 -0.10062072  0.06103673
## Individual silhouette widths:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.29361 -0.05617  0.07916  0.05620  0.20494  0.33168
```

As it can been seen in the figure above, the average silhouette obtained for PAM with k=3 is really close to 0 as in previous methods.

As it has been previously stated, PAM also works with mixed data. For this reason, we are going to assess the existence of different clusters with the use of the quantitative variables employed so far in addition to some categorical variables such as: property type, Centro, host_is_superhost, and cancelation_policy.



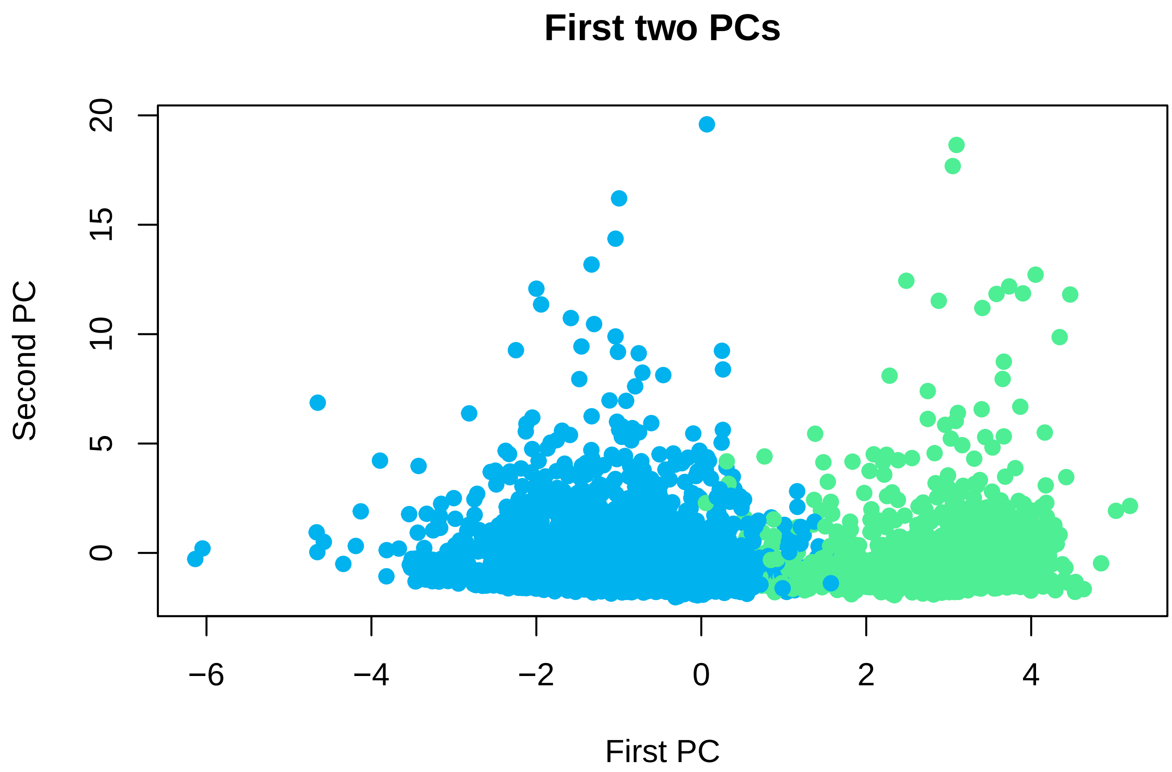
By assessing the average silhouette yielded by the different number of clusters with PAM using mixed data, we observe that the optimal amount of K suggested is equal to 2.

```
## Silhouette of 3125 units in 2 clusters from silhouette.default(x = pam_X_Gower_mat$cluster, dist = X_
##  Cluster sizes and average silhouette widths:
##      2173       952
## 0.3345737 0.3164106
## Individual silhouette widths:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.3185  0.2903  0.3556  0.3290  0.4066  0.5054
```

When conducting PAM with K=2, the average silhouette yielded by the clustering structure defined is 0.32.

CLARA

Although Clara is similar to PAM but employed for large data sets, we are going to carry out the analysis with it as well so we can compare results.



When conducting Clara with k=2, we observe that each stream of data is assigned to a different group.

```
## Silhouette of 3125 units in 2 clusters from silhouette.default(x = clara_X$cluster, dist = dist(smal
##  Cluster sizes and average silhouette widths:
##      2224       901
## 0.1503541 0.1308600
## Individual silhouette widths:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.10785  0.09055  0.15970  0.14473  0.19828  0.24179
```

Furthermore, the average silhouette exhibited by this clustering structure is around 0.15.

Hierarchical Clustering

Hierarchical clustering is a different approach to K-means clustering for identifying and discovering different groups within a dataset. However, in contrast to k-means, hierarchical clustering will create a hierarchy of clusters and therefore does not require us to choose the number of clusters (k) in advance. Besides this, hierarchical clustering has another added advantage over k-means clustering: its results can be easily visualized using a tree-based graph called a *dendrogram*.

Hierarchical clustering can be divided into two main types:

1. **Agglomerative clustering:** This method works in a bottom-up manner. This means that each observation is at the start put in a single-element cluster. This means that at step 1 the number of clusters equals the number of observations. At each step of the algorithm, the two clusters that are the most similar (least distance between them) are combined into a new bigger cluster. This procedure is

iterated until all points are a member of just one single big cluster. The result is a tree which can be displayed using a *dendrogram*.

2. **Divisive hierarchical clustering:** DIANA (DIvise ANAlysis) works in a top-down way and can therefore be seen as the reverse of **agglomerative clustering**. It begins with the root, in which all observations are included in a single cluster. At each step of the algorithm, the current cluster is split into two clusters that are considered most heterogeneous. The process is iterated until every observations has its own cluster. The number of clusters will then equal the number of instances.

Both of these two methods of hierarchical clustering will be performed on the dataset below.

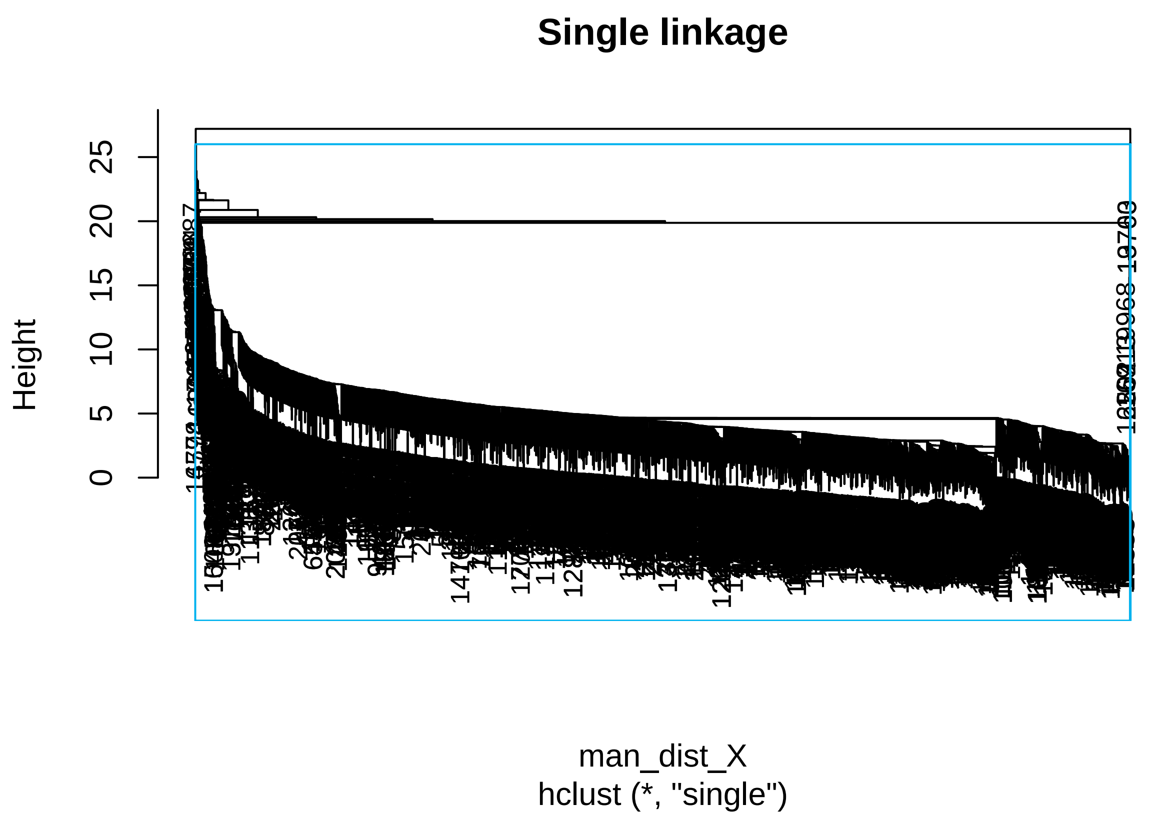
Agglomerative Method

Similar to k-means, the similarity of observations are measured using distance measures (*Euclidean* or *Manhattan* distance). The Euclidean distance is most commonly used, but that is not a factor with a strong influence in this method. On the other hand, an important question in hierarchical clustering is: How can the dissimilarity between two clusters of observations be measured? A number of different cluster *linkage methods* have been developed to answer this question. The most common methods are:

1. **Complete linkage clustering:** All pairwise distances between the elements in cluster 1 and the elements in cluster 2 are computed, and the **largest** value of these distances is taken as the distance between the two clusters. This method tends to create more compact clusters.
2. **Single linkage clustering:** All pairwise distances between the elements in cluster 1 and the elements in cluster 2 are computed, and the **smallest** value of these distances is taken as the linkage criterion. This methods tends to create long, "loose" clusters.
3. **Average linkage clustering:** All pairwise distances between the elements in cluster 1 and the elements in cluster 2 are computed, and the **average** value of these distances is taken as the linkage criterion. This method can create both compact or less compact clusters. There is not a general rule of thumb for this.
4. **Ward's minimum variance method:** This method minimizes the total within-cluster variance. At each step the pair of clusters with the smallest between-cluster distance is merged. This method tends to produce more compact and equal clusters.

All of these 4 methods will be performed on the dataset below.

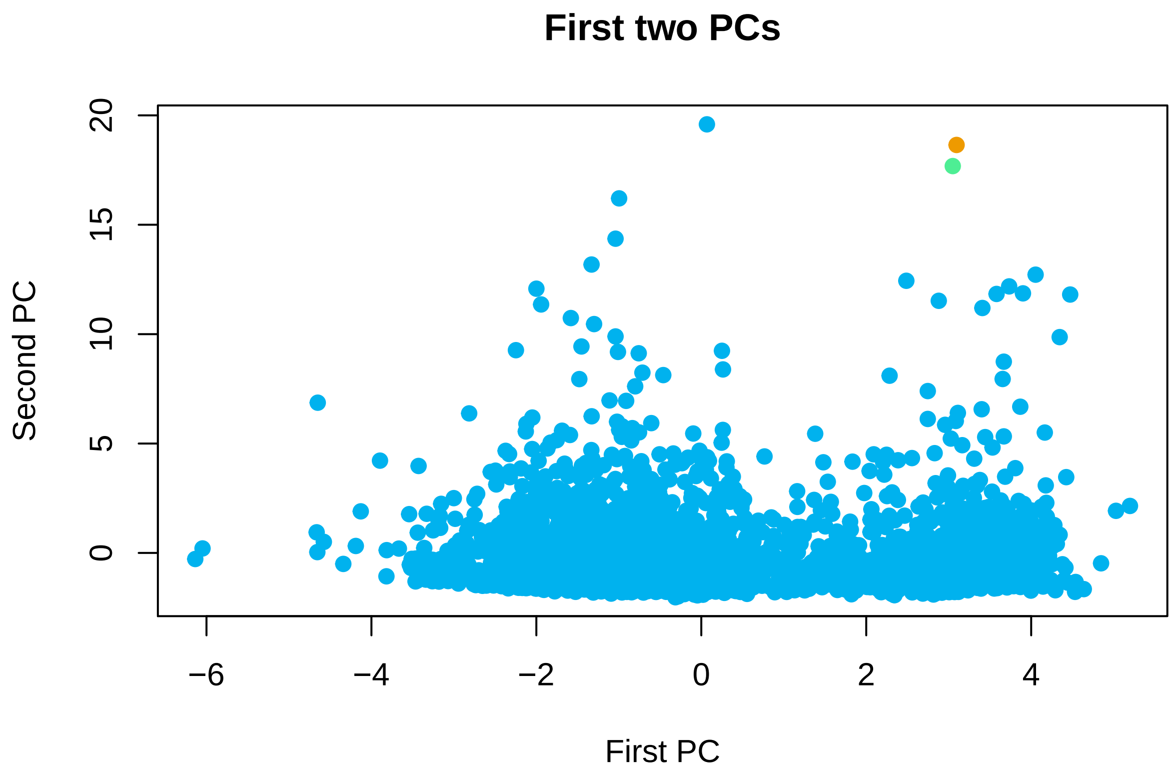
```
K = 3 # based on silhouette

# Compute distances
man_dist_X <- daisy(scale(X),metric="manhattan")

# Single linkage
single_X <- hclust(man_dist_X,method="single")

# Plot dendrogram
plot(single_X,main="Single linkage",cex=0.8)
rect.hclust(single_X,k=K,border=color_1)
```



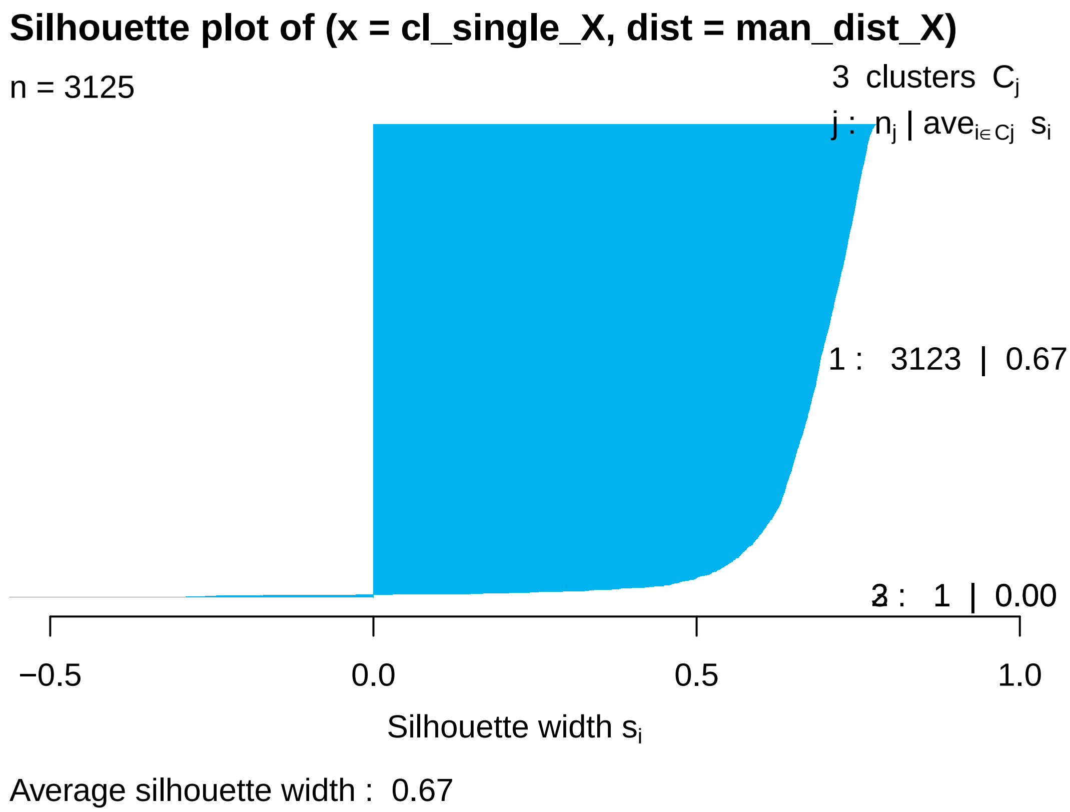

```
cl_single_X <- cutree(single_X,K)
table(cl_single_X)
```

```
## cl_single_X
##    1    2    3
## 3123    1    1
```

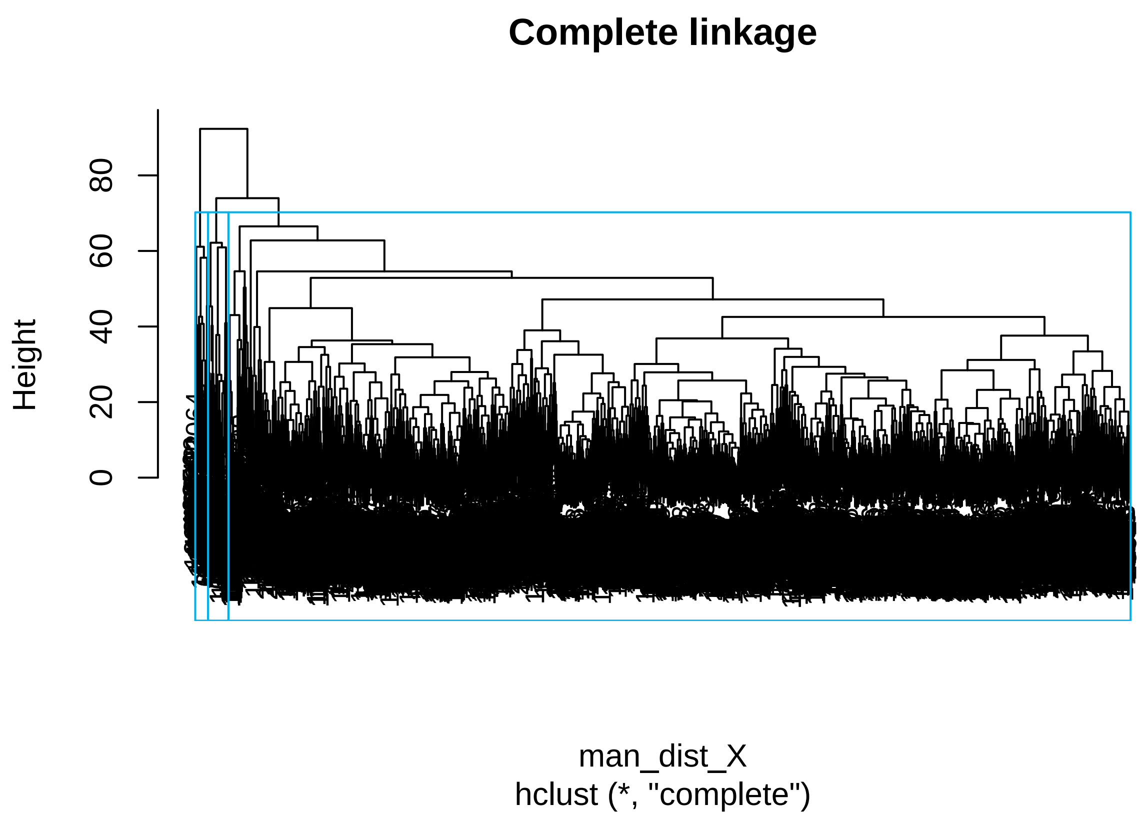
```
# Plot of the first two PCs with the five clusters
colors_single_X <- c(color_1,color_2,color_3,color_4,color_5)[cl_single_X]
plot(X_pcs$x[,1:2],pch=19,col=colors_single_X,main="First two PCs",xlab="First PC",ylab="Second PC")
```




```r
# Silhouette
sil_single_X <- silhouette(cl_single_X,man_dist_X)
plot(sil_single_X,col=color_1)
```



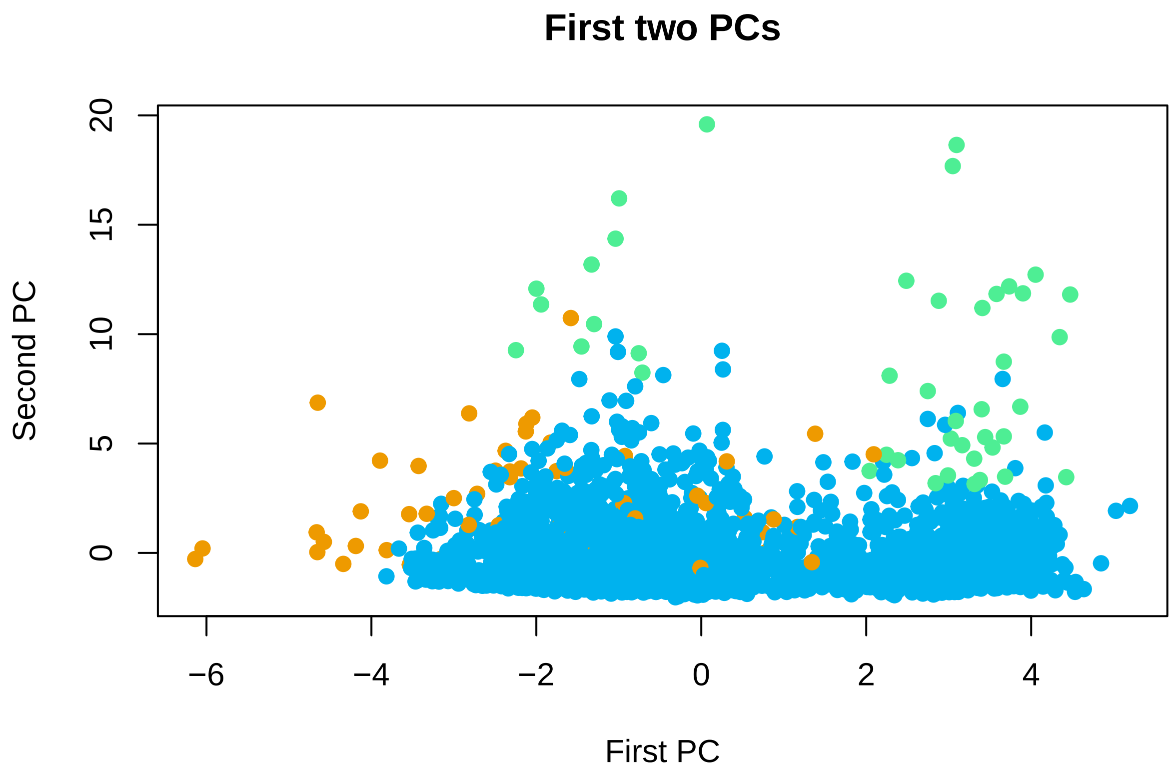

```r
# Complete linkage
complete_X <- hclust(man_dist_X,method="complete")

# Plot dendrogram
plot(complete_X,main="Complete linkage",cex=0.8)
rect.hclust(complete_X,k=K,border=color_1)
```



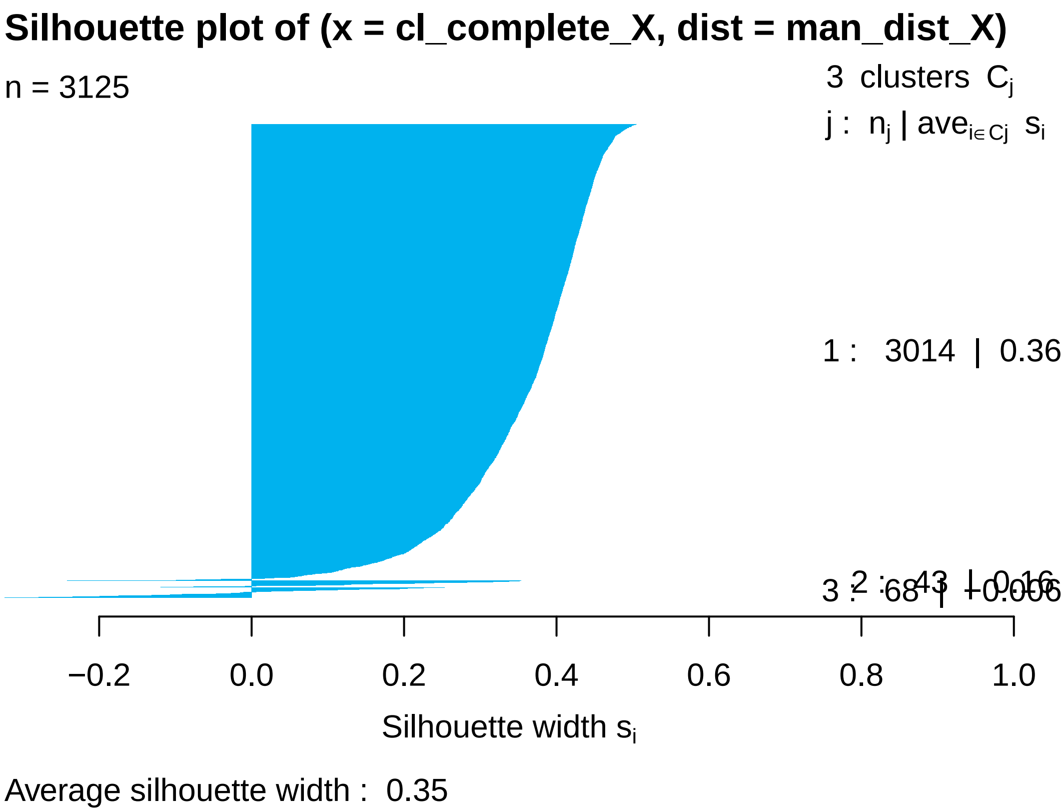

```
cl_complete_X <- cutree(complete_X,K)
table(cl_complete_X)
```

```
## cl_complete_X
##    1    2    3
## 3014   43   68
```

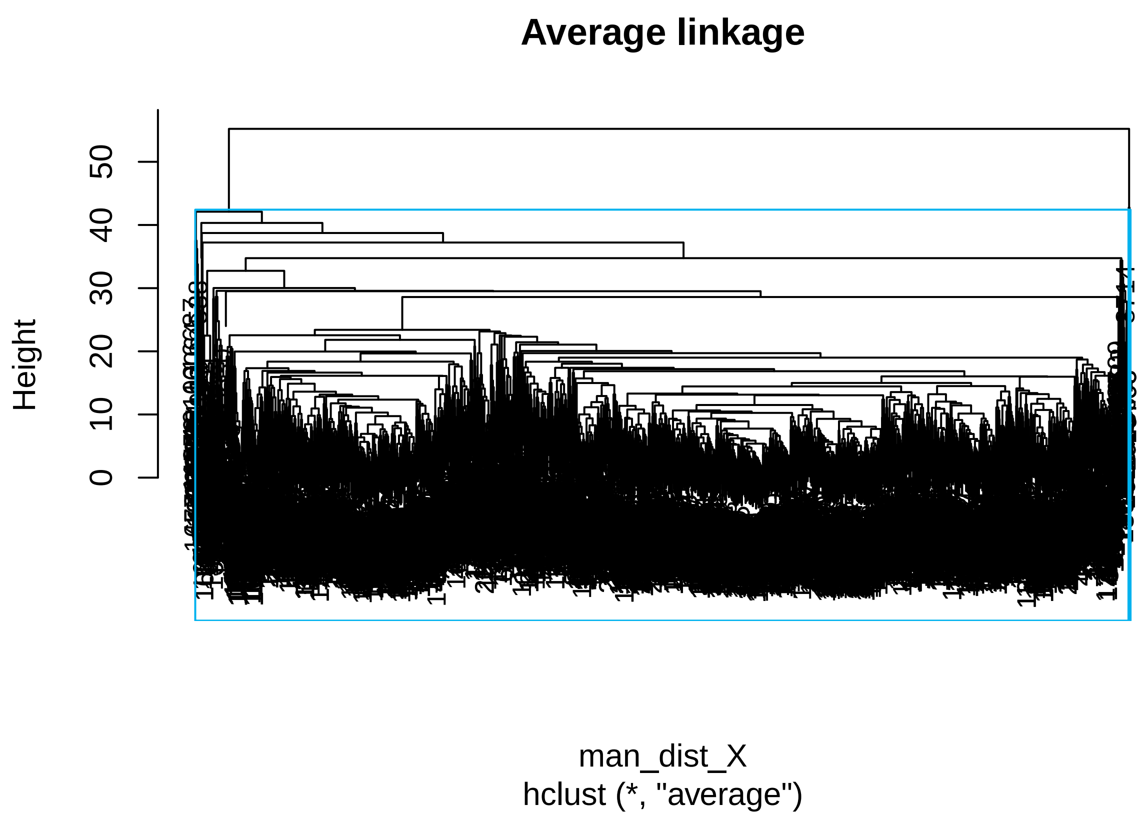
```
# Plot of the first two PCs with the five clusters
colors_complete_X <- c(color_1,color_2,color_3,color_4,color_5)[cl_complete_X]
plot(X_pcs$x[,1:2],pch=19,col=colors_complete_X,main="First two PCs",xlab="First PC",ylab="Second PC")
```




```r
# Silhouette
sil_complete_X <- silhouette(cl_complete_X,man_dist_X)
plot(sil_complete_X,col=color_1)
```



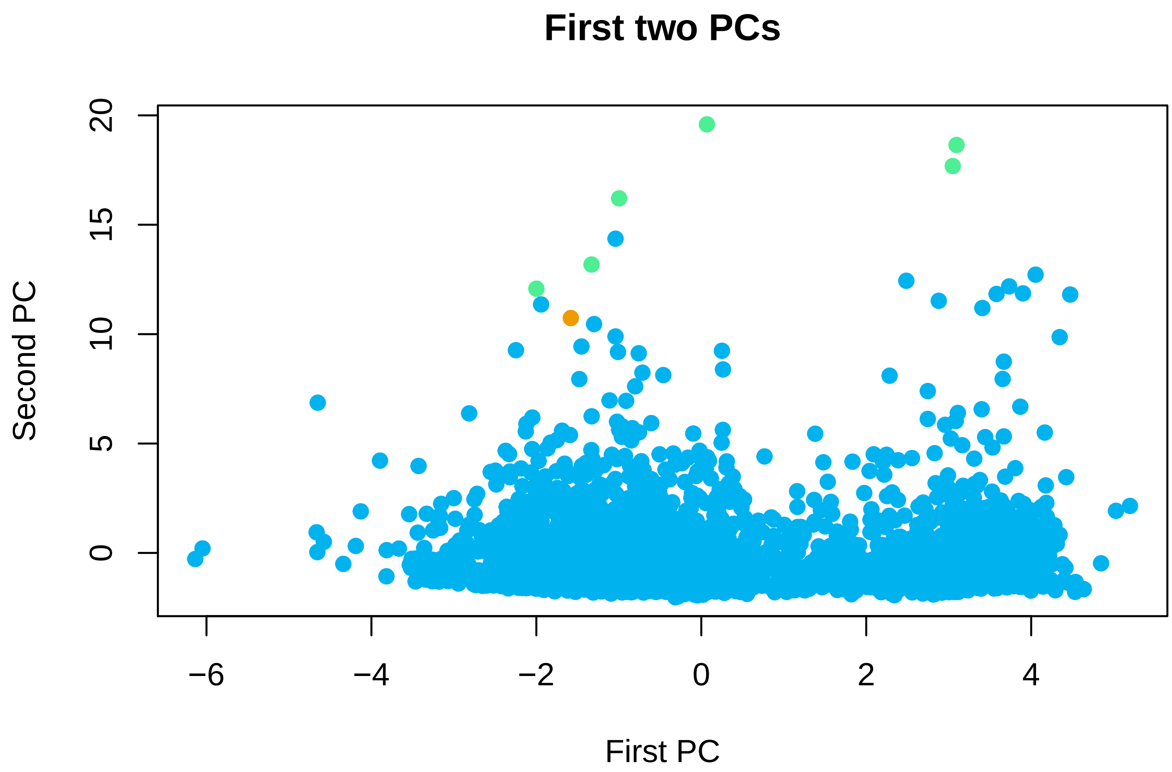

```r
# Average linkage
average_X <- hclust(man_dist_X,method="average")

# Plot dendogram
plot(average_X,main="Average linkage",cex=0.8)
rect.hclust(average_X,k=K,border=color_1)
```



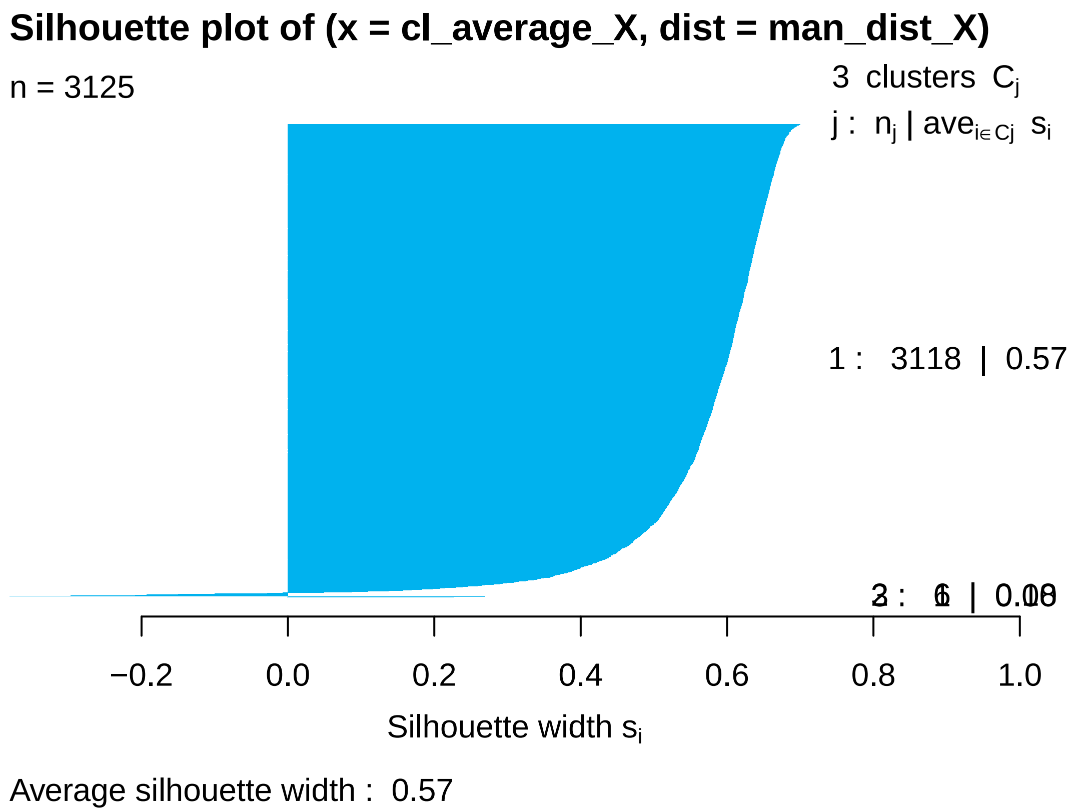

```r
cl_average_X <- cutree(average_X,K)
table(cl_average_X)
```

```
## cl_average_X
##    1    2    3
## 3118    6    1
```

```r
# Plot of the first two PCs with the five clusters
colors_average_X <- c(color_1,color_2,color_3,color_4,color_5)[cl_average_X]
plot(X_pcs$x[,1:2],pch=19,col=colors_average_X,main="First two PCs",xlab="First PC",ylab="Second PC")
```



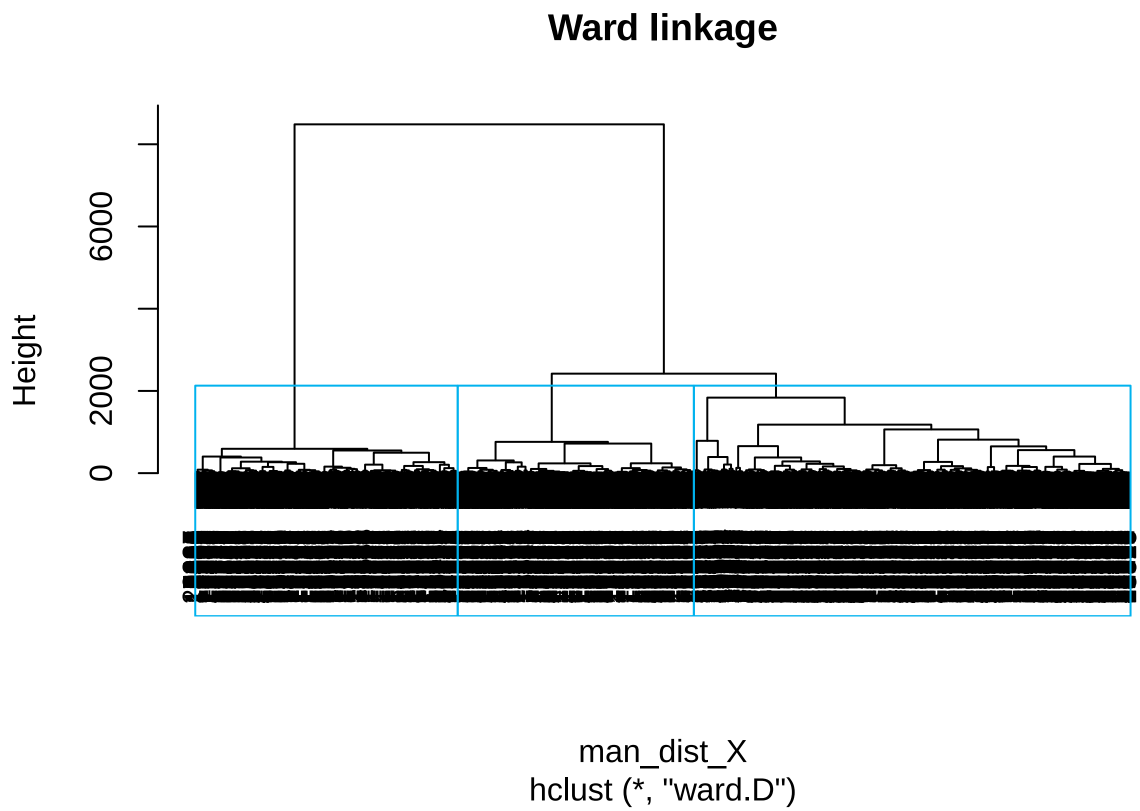

```r
# Silhouette
sil_average_X <- silhouette(cl_average_X,man_dist_X)
plot(sil_average_X,col=color_1)
```




```
# Ward linkage
ward_X <- hclust(man_dist_X,method="ward")
```

```
## The "ward" method has been renamed to "ward.D"; note new "ward.D2"
```

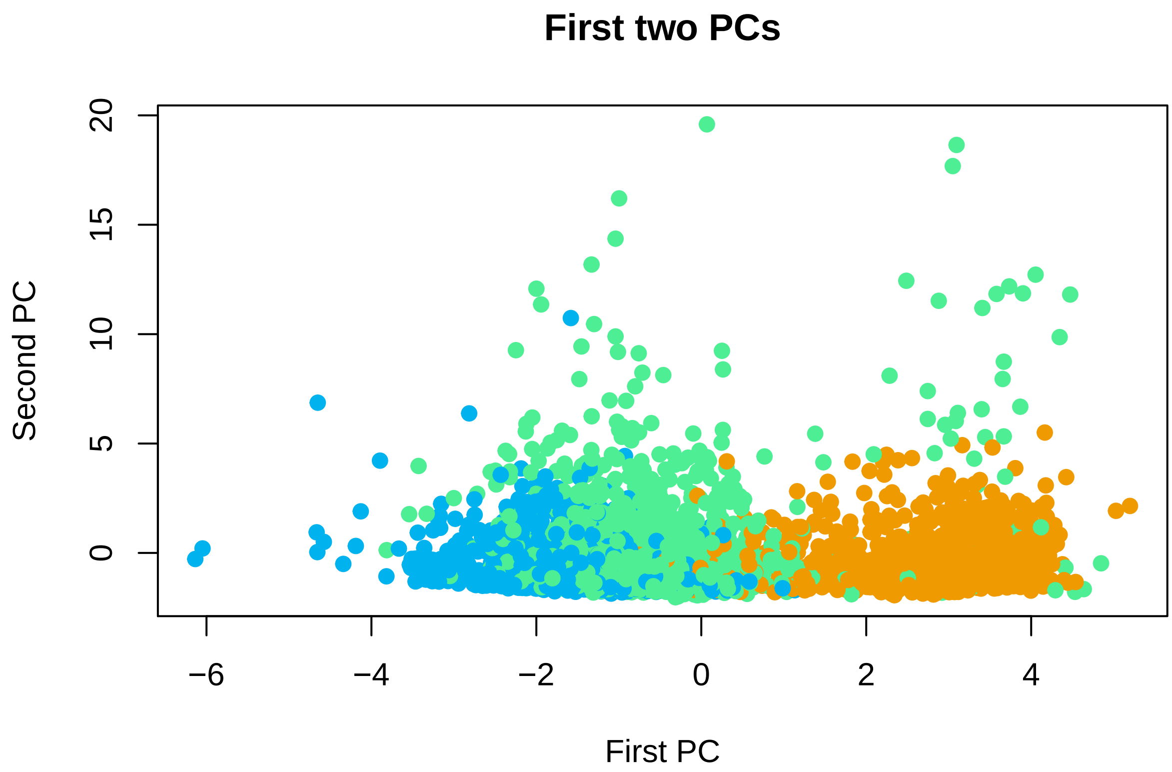
```
# Plot dendrogram
plot(ward_X,main="Ward linkage",cex=0.8)
rect.hclust(ward_X,k=K,border=color_1)
```



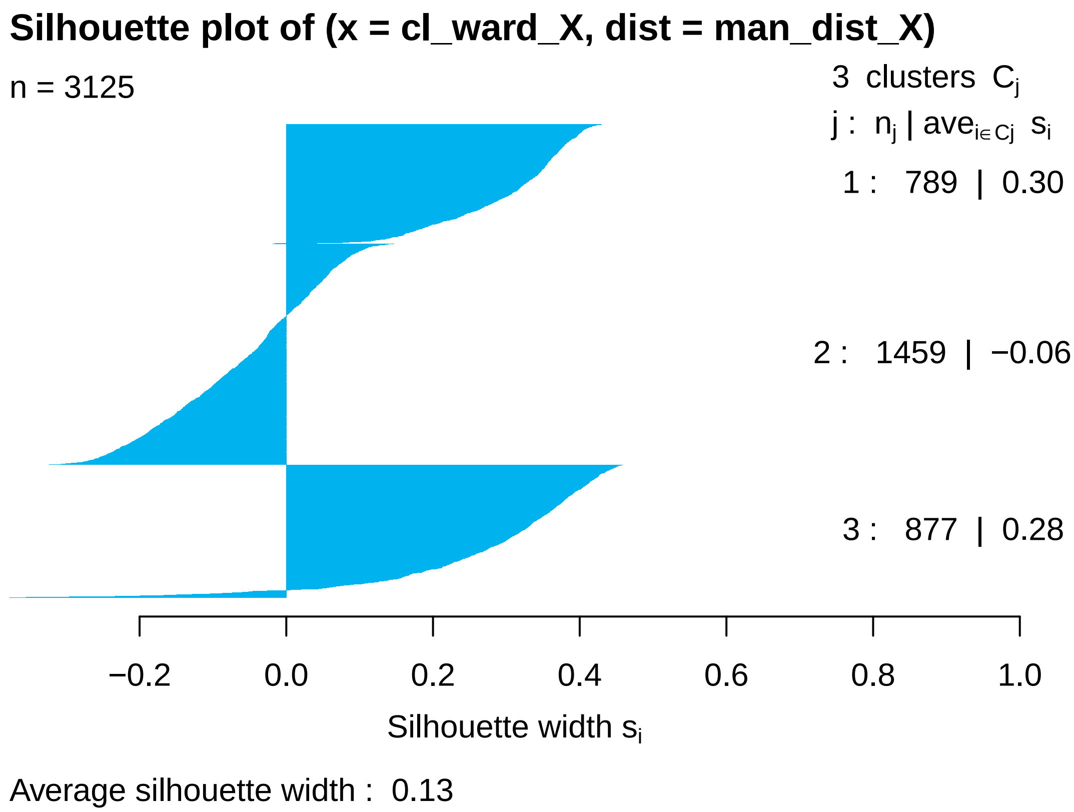

```
cl_ward_X <- cutree(ward_X,K)
table(cl_ward_X)
```

```
## cl_ward_X
##    1    2    3
##  789 1459  877
```

```
# Plot of the first two PCs with the five clusters
colors_ward_X <- c(color_1,color_2,color_3,color_4,color_5)[cl_ward_X]
plot(X_pcs$x[,1:2],pch=19,col=colors_ward_X,main="First two PCs",xlab="First PC",ylab="Second PC")
```




```r
# Silhouette
sil_ward_X <- silhouette(cl_ward_X,man_dist_X)
plot(sil_ward_X,col=color_1)
```




```
# This solution is probably the best one among the agglomerative hierarchical clustering methods
```

Now rises the question: which of the agglomerative hierarchical clustering methods works the best (gives the best results)? Based on the plots shown above, the conclusion could be made that the last method (*Ward Method*) yields the best results. To further ground this conclusion the *Agglomerative Coefficient* (AC) could be used.

```
# methods to assess
m <- c( "average", "single", "complete", "ward")
names(m) <- c( "average", "single", "complete", "ward")

# function to compute coefficient
ac <- function(x) {
  agnes(scale(X), method = x)$ac
}

# get agglomerative coefficient for each linkage method
purrr::map_dbl(m, ac)
```

```
##   average    single  complete      ward
## 0.8937960 0.8637393 0.9240589 0.9836633
```

The AC shows the strength of the generated clustering structure. The closer the value is to 1, the more balanced the clustering structure is. The *complete* and *Ward's* linkage methods generally yield higher AC values. Values closer to 0 imply less well-formed clusters, as can be seen above in the dendrogram of the *single linkage*. Important is to note that the AC tends to become larger as n increases, so it should only be used to compare the same dataset (with an equal n) and should therefore not be used across data sets of very different sizes.

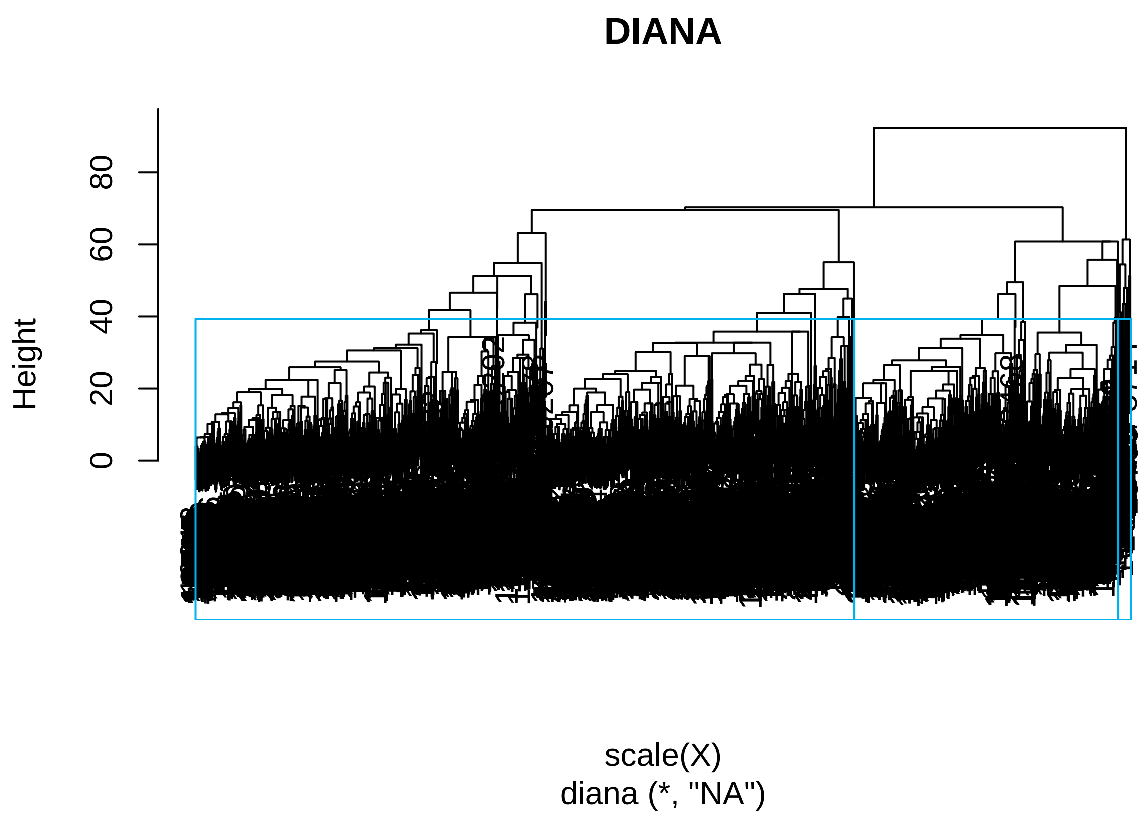
Here we see that Ward's method has the highest AC which means that it has the strongest clustering structure of the four methods assessed.

Divisive Method (DIANA)

For DIANA, clusters are divided based on the maximum average dissimilarity which is very similar to the mean or average linkage clustering method outlined above.

```r
# Divisive hierarchical clustering
diana_X <- diana(scale(X),metric="manhattan")
```

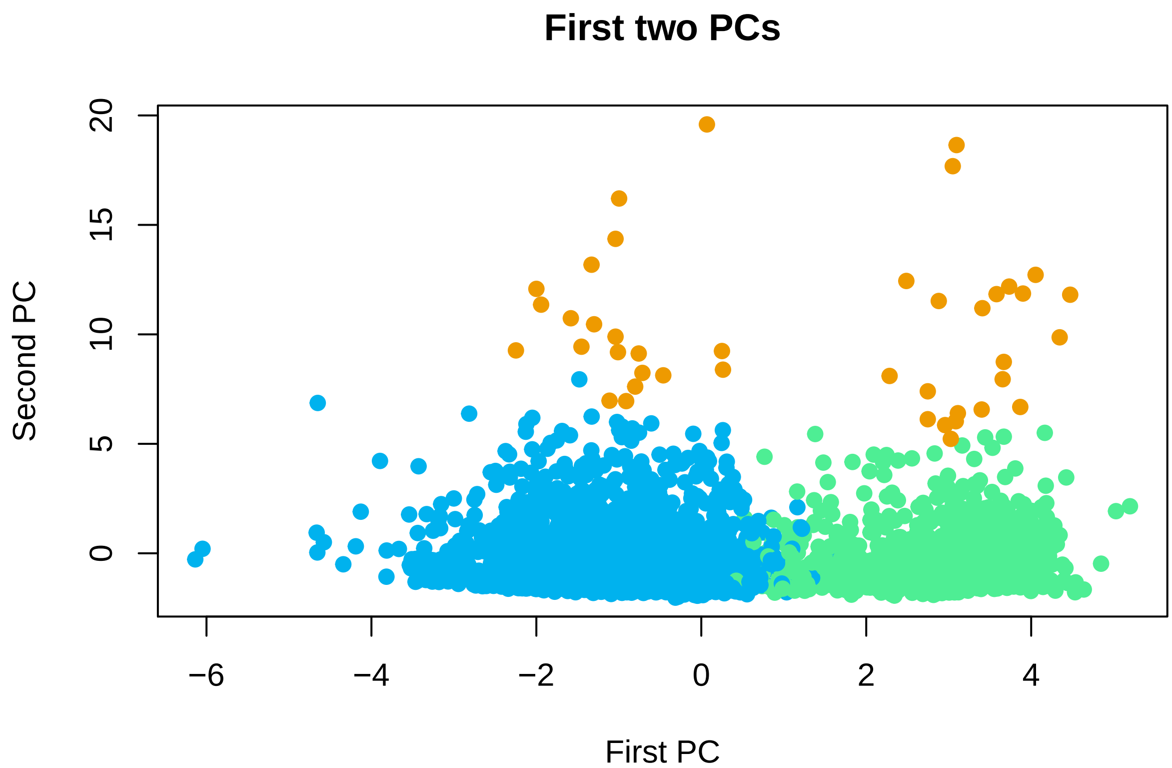
```r
# Plot dendrogram of the solution
pltree(diana_X,main="DIANA")
rect.hclust(diana_X,k=K,border=color_1)
```



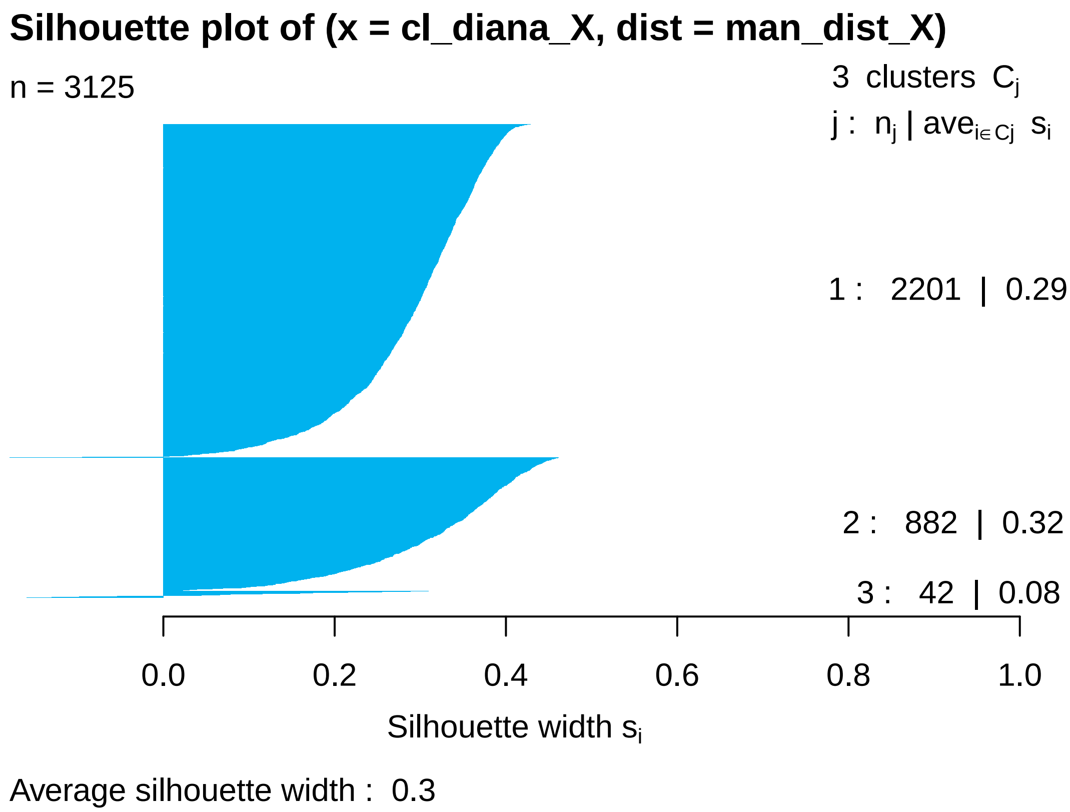

```r
cl_diana_X <- cutree(diana_X,K)
table(cl_diana_X)
```

```
## cl_diana_X
##    1    2    3
## 2201  882   42
```

```r
# Plot of the first two PCs with the five clusters
colors_diana_X <- c(color_1,color_2,color_3,color_4,color_5)[cl_diana_X]
plot(X_pcs$x[,1:2],pch=19,col=colors_diana_X,main="First two PCs",xlab="First PC",ylab="Second PC")
```




```
# Silhouette
sil_diana_X <- silhouette(cl_diana_X,man_dist_X)
plot(sil_diana_X,col=color_1)
```




```r
# Divisive Coeffectient
diana_X$dc
```

```
## [1] 0.9308197
```

The best method of the Hierarchical Clustering analysis seems to be the last one: the *DIANA* method. As can be seen above is that the divisive coeffectient of 0.9308197 is lower than the aglomerative coefficient of the *Ward Method*. This can be explained, however, by the fact that the *Ward Method* creates more equal clusters, which drives up the score. Reality is, though, that the data is imbalanced and therefore the clusters do not need to be equal.

Model Based Clustering

This method, in contrast to traditional methods such as the partitional or hierarchical methods that compute clusters based on the data, tries to allocate a mixture of distributions to the data by allocating a measure of probability and uncertainty to the cluster assignments.

Model-based clustering tries to provide a *soft-assignment*, which means that observations have a probability of belonging to a specific cluster. This method has also the benefit of automatically identifying the optimal number of clusters.

We start by computing the BIC for the different types of covariance matrices that describe the distributions. The function mclust is able to work with up to 14 different covariance matrix configurations, and makes the computations up to the number of selected possible groups, and then selects the best value along with the best covariance.

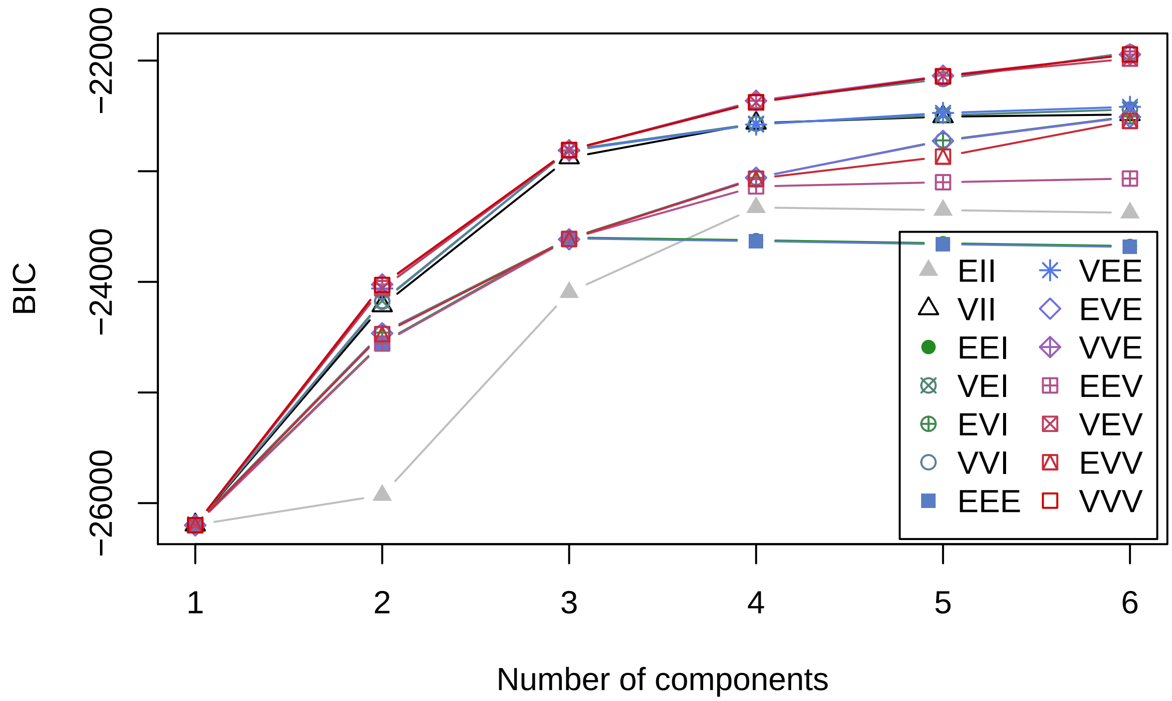
At the end, this method will select the top 3 models that have a better fit to the data.

```r
# compute the BIC for up to 6 groups
BIC_X <- mclustBIC(Z,G=1:6)
BIC_X
```

```
## Bayesian Information Criterion (BIC):
##         EII       VII       EEI       VEI       EVI       VVI       EEE
## 1 -26197.24 -26197.24 -26191.93 -26191.93 -26191.93 -26191.93 -26199.98
## 2 -25928.67 -24215.31 -24550.06 -24172.62 -24458.99 -24180.68 -24556.54
## 3 -24095.06 -22877.76 -23599.08 -22807.74 -23607.79 -22809.34 -23607.07
## 4 -23327.40 -22564.89 -23623.50 -22570.19 -23059.04 -22365.52 -23631.34
## 5 -23351.57 -22506.64 -23651.35 -22494.21 -22721.32 -22167.06 -23659.35
## 6 -23375.72 -22487.55 -23675.94 -22441.53 -22504.76 -21925.50 -23684.07
##         VEE       EVE       VVE       EEV       VEV       EVV       VVV
## 1 -26199.98 -26199.98 -26199.98 -26199.98 -26199.98 -26199.98 -26199.98
## 2 -24059.95 -24464.98 -24023.47 -24557.68 -24057.05 -24472.26 -24028.31
## 3 -22815.24 -23615.57 -22810.75 -23612.17 -22812.11 -23612.58 -22807.59
## 4 -22578.00 -23058.84 -22363.20 -23137.39 -22378.89 -23067.53 -22376.25
## 5 -22473.19 -22726.07 -22136.90 -23099.58 -22140.52 -22867.84 -22143.88
## 6 -22418.05 -22509.84 -21945.16 -23066.10 -21982.22 -22545.54 -21944.77
##
## Top 3 models based on the BIC criterion:
##     VVI,6     VVV,6     VVE,6
## -21925.50 -21944.77 -21945.16
```

We can see that the chosen models are 2 VVE (ellipsoidal with equal orientation), and one VII (diagonal, varying volume and shape). From the VVE, there is one with 4 and other with 5 clusters, and one VII with 5 clusters. This means that the clusters obtained with the model should be between 4-5 in general terms.

Below we plot the graph of the results with the different covariance shapes and the resulting BIC scores. Keep in mind that these results are displayed as the negative of the BIC, thus we select the maximum possible value. We see that the graph is strognly increasing, until it reaches around 5 groups, where the BIC value starts to decrease again.

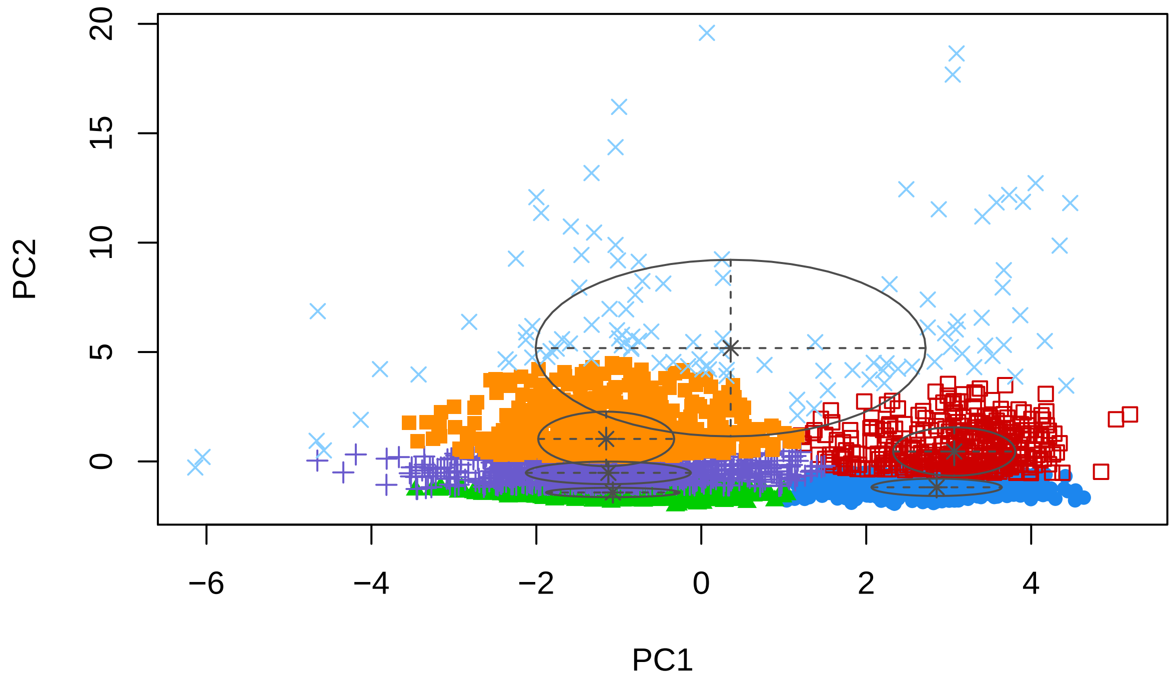


Now we will run the mclust function with the optimal solution obtained in the part above.

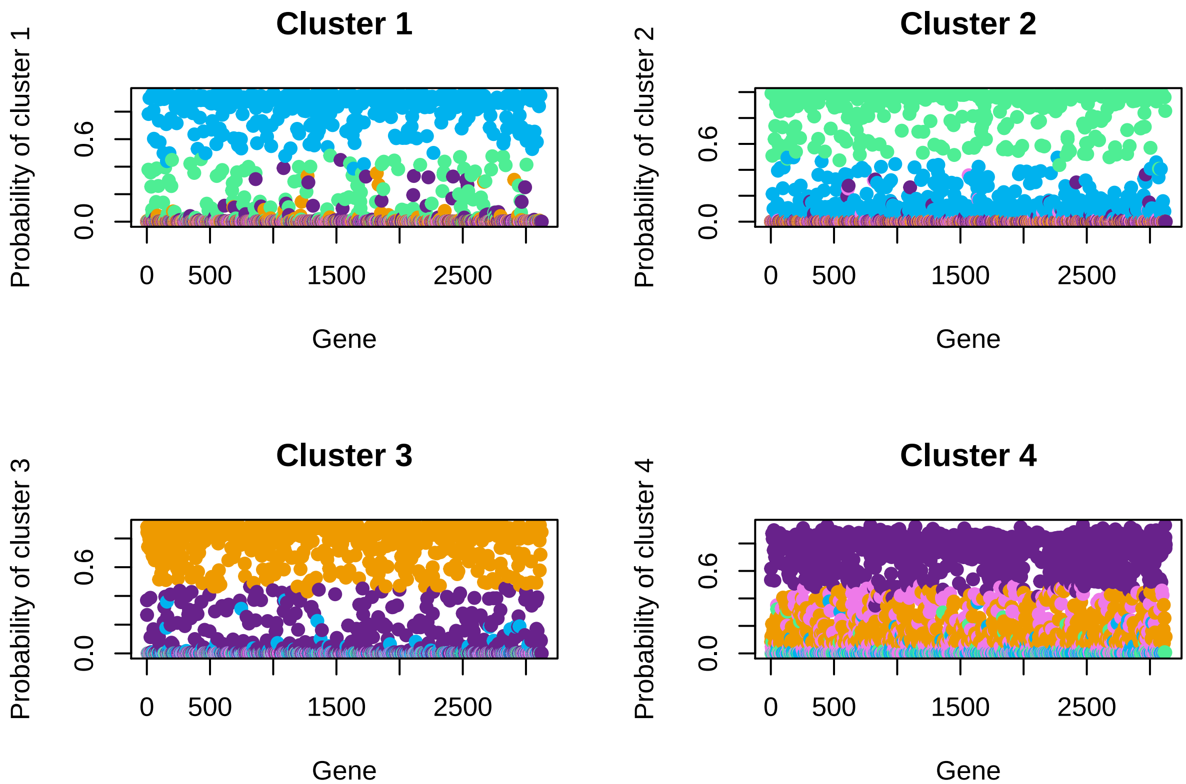
```
Mclust_X <- Mclust(Z,x=BIC_X)
summary(Mclust_X)
```

```
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## ----------------------------------------------------
##
## Mclust VVI (diagonal, varying volume and shape) model with 6 components:
##
##  log-likelihood    n df       BIC       ICL
##       -10846.07 3125 29 -21925.5 -23327.27
##
## Clustering table:
##   1   2   3   4   5   6
## 477 344 631 912 651 110
```

And finally, we can plot the classification according to this model-based clustering approach. And we plot also the probabilities of the data belonging to each of the different clusters.



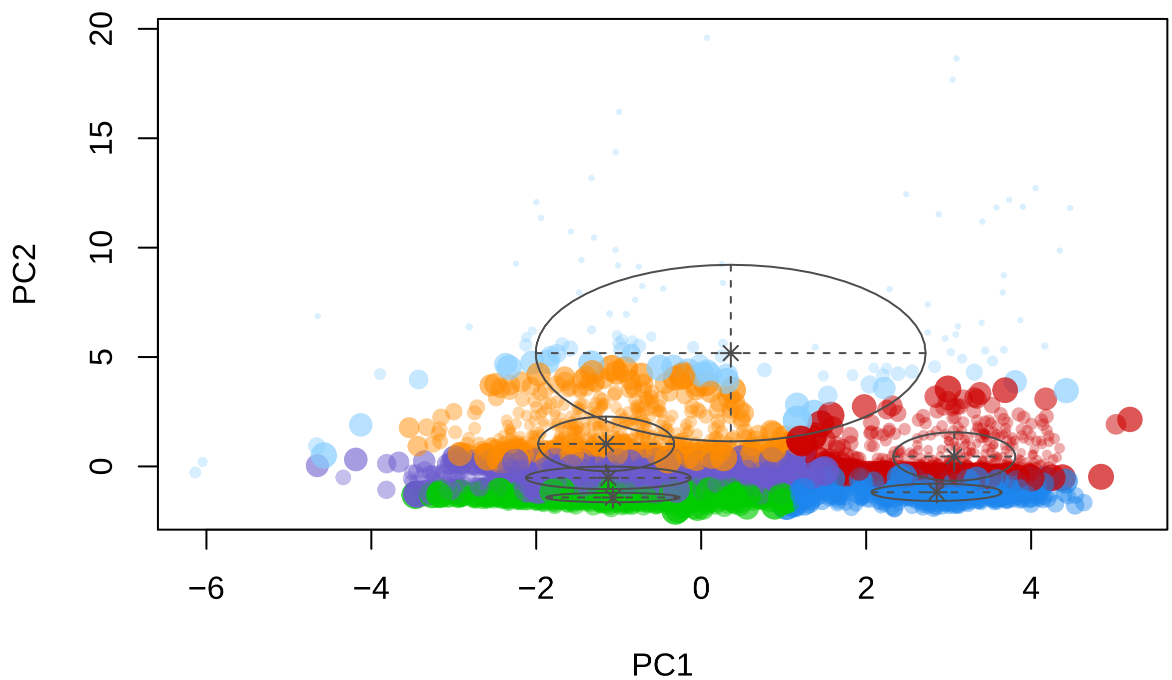
PC2
20
15
10
5
0
−6
−4
−2
0
2
4
PC1



From the probability graphs from each of the clusters, we see that there are some uncertainties in the cluster #1 and #4, since we see that there are more point that stay around the midline and less that go to the extremes (0,1).

To finish, we plot the uncertainty of the points that were assigned in a similar manner as the classification plot shown above. The bigger the points are, the less certain its allocation to its current cluster is.

```
par(mfrow=c(1,1))
plot(Mclust_X,what="uncertainty")
```



To conclude, we can see that the model-based approach is a very interesting one, since it does not base the clusters on the data itself, but rather builds dome possible mixture of models based on the structure of the covariance matrices of each of the distributions. The name of the model chosen is given by the characteristics and shape of the distribution, some might be ellipsiodal, other are spherical and some can also be diagonal. The package used "mclust" has a total of 14 possible multivariate mixtures.

We have seen that the multivariate mixture that better fits our case is the VVV, 5. This consists of a multivariate mixture with 5 ellipsoidal shapes with varying volume, shape and orientation, thus it does not really help much on the clarification of the distribution, but it gives us a general understanding that the data has indeed some different clusters that must be considered.

2. Perform supervised classification. Obtain conclusions from the analysis.

This part of the report will show three different methods for performing supervised learning. For this type of learning, we will need a response variable, this will be 'room_type', just as in the previous case.

The three different methods to be explained are the following:

- K-nearest-neighbours KNN
- Methods based on the Bayes Theorem
  - Linear Discriminant Analysis (LDA)
  - Quadratic Discriminant Analysis (QDA)
  - Naive Bayes
- Logistic Regression

KNN K-nearest-neighbours is a relatively simple algorithm in which each new observation is predicted based on the K nearest neighbours of this new observation. KNN is a memory based algorithm, since it has to store all the training data for making future comparissons, and cannot be described in any closed-form. The Knn algortihm has shown to be very useful, but it can also be somehow computationally inefficient.

We will show below the process for classifying instances according to the 'room_type' attribute. This attribute, as discussed above, has 4 possible categories.

We must first split the data into training a testing datasets, we have chosen to do a 70-30 strategy, which is commonly used. Below we can see the division of the groups.

```
## [1] "The training set has 2187.000000 observations"
```

```
## [1] "The testing set has 938.000000 observations"
```

```
## [1] "Proportions in the Training set"
```

```
## [1] "Entire home/apt"
## [1] 0.5953361
## [1] "Private room"
## [1] 0.3689986
## [1] "Hotel room"
## [1] 0.02652035
## [1] "Shared room"
## [1] 0.009144947
```

```
## [1] "-------"
```

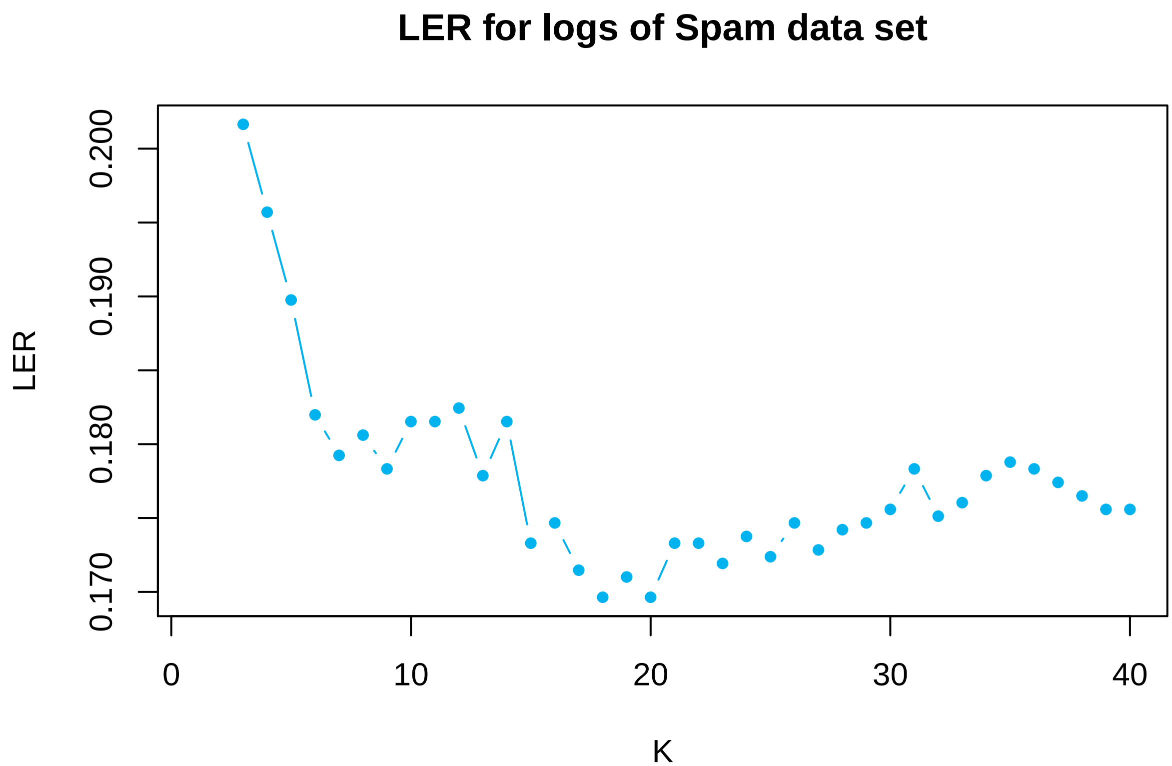
```
## [1] "Proportions in the Testing set"
```

```
## [1] "Entire home/apt"
## [1] 0.6481876
## [1] "Private room"
## [1] 0.3144989
## [1] "Hotel room"
## [1] 0.02452026
## [1] "Shared room"
## [1] 0.01279318
```

We can see that the proportions of the groups in the testing and training datasets are very similar. We must scale the data before we perform any computations. In this case, we will use the library 'class'.

```
stan_X_train <- scale(X_train)
stan_X_test <- scale(X_test)
```

```
LER <- matrix(NA,nrow=40,ncol=1)
for (i in 3 : 40){
  knn_output <- knn.cv(stan_X_train,Y_train,k=i)
  LER[i] <- 1 - mean(knn_output==Y_train)
}
plot(1:40,LER,pch=20,col=color_1,type="b",xlab="K",ylab="LER",main="LER for logs of Spam data set")
```




```
K <- which.min(LER)
K
```

```
## [1] 18
```

```
knn_Y_test <- knn(stan_X_train,stan_X_test,Y_train,k=K,prob=TRUE)
```

```
# Confusion table
table(Y_test,knn_Y_test)
```

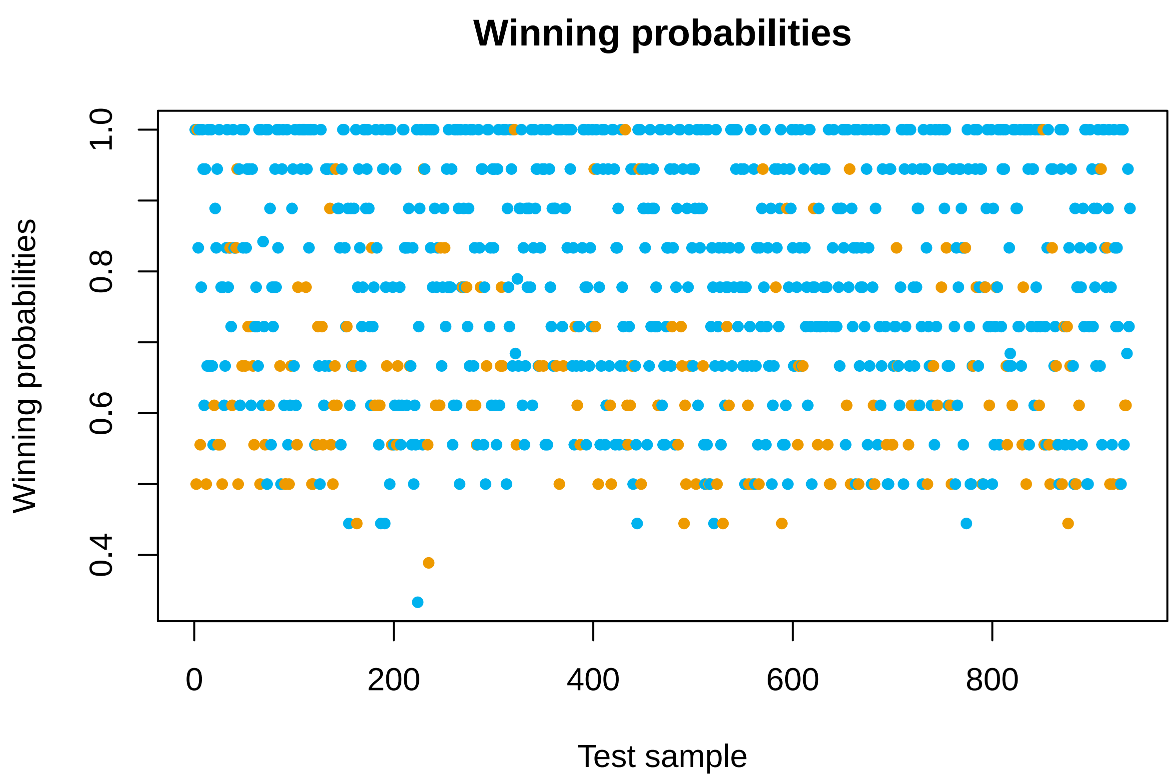
```
##                    knn_Y_test
## Y_test              Entire home/apt Hotel room Private room Shared room
##   Entire home/apt               511          0           97           0
##   Hotel room                     15          0            8           0
##   Private room                   53          0          242           0
##   Shared room                     6          0            6           0
```

```
# Obtain the Test Error Rate (TER)
prob_knn_Y_test <- attributes(knn_Y_test)$prob

knn_TER <- mean(Y_test!=knn_Y_test)
knn_TER
```

```
## [1] 0.1972281
```

```
# Make a plot of the probabilities of the winner group
# In blue, good classifications, in red, wrong classifications

colors_errors <- c(color_3,color_1)[1*(Y_test==knn_Y_test)+1]
```

```r
plot(1:n_test,prob_knn_Y_test,col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Winning probabi
     main="Winning probabilities")
```



Methods based on the Bayes Theorem (desc.)

```r
# Linear Discriminant Analysis (LDA)

# Estimate the unknown parameters with the training sample

lda_train <- lda(Y_train ~ .,data=X_train)

# Estimated prior probabilities for the four groups

lda_train$prior
```

```
## Entire home/apt      Hotel room    Private room     Shared room
##     0.595336077     0.026520348     0.368998628     0.009144947
```

```r
# Estimated sample mean vectors

t(lda_train$means)
```

```
##                             Entire home/apt  Hotel room Private room
## host_response_rate               4.444701112   4.5356610   4.34478245
## host_total_listings_count       22.091860215  19.6734138   7.76072739
## accommodates                     4.221430108   3.5182414   1.87088848
```

```
## bathrooms                      0.187345960   0.1712817   0.12836432
## bedrooms                      -0.378789987  -0.6137889  -0.09347087
## beds                           0.641895700   0.6040626  -0.02900099
## price                          4.477114731   4.6329221   3.74196207
## cleaning_fee                   2.543428102  -0.6687550   0.57956951
## guests_included                2.150001536   1.9492759   1.10013259
## extra_people                  -1.832450003  -3.4654436  -3.57560386
## minimum_nights                 2.636176651   2.4320345   2.31574597
## maximum_nights               273.863519201 261.8630690 227.53879430
## availability_30               -0.767170317   0.2470504  -1.72642010
## availability_60                0.701231353   1.6515606  -0.70125882
## availability_90                1.353407520   2.1656795  -0.17503722
## availability_365               3.037985975   4.0163954   1.44810724
## number_of_reviews              1.237821057  -0.2679565  -0.63405979
## review_scores_accuracy         9.502536098   9.4665172   9.60694796
## review_scores_cleanliness      9.361983103   9.5354828   9.39629120
## review_scores_checkin          9.587789555   9.6561724   9.71351549
## review_scores_communication    9.633104455   9.4665172   9.71351549
## review_scores_location         9.701460829   9.8630690   9.61066543
## review_scores_value            9.146929339   9.2596207   9.31078934
## reviews_per_month             -0.001768022  -0.4578986  -0.39967057
##                              Shared room
## host_response_rate          4.381705e+00
## host_total_listings_count   7.651000e+00
## accommodates                3.451000e+00
## bathrooms                   5.382275e-01
## bedrooms                    9.995003e-04
## beds                        7.422745e-01
## price                       3.209264e+00
## cleaning_fee                2.453346e-01
## guests_included             1.001000e+00
## extra_people               -2.914277e+00
## minimum_nights              2.751000e+00
## maximum_nights              2.563010e+02
## availability_30            -1.791647e+00
## availability_60            -5.806842e-01
## availability_90            -2.893422e-01
## availability_365            1.150029e+00
## number_of_reviews          -9.255777e-01
## review_scores_accuracy      9.301000e+00
## review_scores_cleanliness   9.551000e+00
## review_scores_checkin       9.601000e+00
## review_scores_communication  9.651000e+00
## review_scores_location      9.601000e+00
## review_scores_value         9.051000e+00
## reviews_per_month          -8.906210e-01
```

```r
# The function does not return the estimated covariance matrix

##########################################################################################################
# Classify the observations in the test sample

lda_test <- predict(lda_train, newdata = X_test)
```

```r
# The vector of classifications made can be found here

lda_Y_test <- lda_test$class
lda_Y_test
```

```
##   [1] Entire home/apt Entire home/apt Entire home/apt Private room   
##   [5] Entire home/apt Private room    Entire home/apt Entire home/apt
##   [9] Entire home/apt Private room    Entire home/apt Entire home/apt
##  [13] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
##  [17] Private room    Entire home/apt Private room    Entire home/apt
##  [21] Entire home/apt Private room    Private room    Entire home/apt
##  [25] Entire home/apt Entire home/apt Private room    Private room   
##  [29] Private room    Entire home/apt Entire home/apt Private room   
##  [33] Entire home/apt Private room    Entire home/apt Entire home/apt
##  [37] Private room    Private room    Entire home/apt Entire home/apt
##  [41] Entire home/apt Private room    Entire home/apt Entire home/apt
##  [45] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
##  [49] Private room    Entire home/apt Private room    Entire home/apt
##  [53] Entire home/apt Entire home/apt Private room    Entire home/apt
##  [57] Entire home/apt Entire home/apt Private room    Entire home/apt
##  [61] Entire home/apt Private room    Private room    Entire home/apt
##  [65] Entire home/apt Private room    Entire home/apt Entire home/apt
##  [69] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
##  [73] Entire home/apt Private room    Private room    Private room   
##  [77] Entire home/apt Entire home/apt Private room    Entire home/apt
##  [81] Entire home/apt Entire home/apt Private room    Entire home/apt
##  [85] Entire home/apt Private room    Entire home/apt Entire home/apt
##  [89] Entire home/apt Private room    Entire home/apt Private room   
##  [93] Entire home/apt Private room    Entire home/apt Entire home/apt
##  [97] Private room    Entire home/apt Private room    Private room   
## [101] Entire home/apt Private room    Entire home/apt Entire home/apt
## [105] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [109] Entire home/apt Entire home/apt Entire home/apt Private room   
## [113] Entire home/apt Entire home/apt Private room    Entire home/apt
## [117] Entire home/apt Private room    Private room    Entire home/apt
## [121] Private room    Entire home/apt Private room    Private room   
## [125] Private room    Entire home/apt Entire home/apt Entire home/apt
## [129] Private room    Private room    Entire home/apt Entire home/apt
## [133] Entire home/apt Entire home/apt Private room    Entire home/apt
## [137] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [141] Private room    Private room    Entire home/apt Private room   
## [145] Private room    Entire home/apt Private room    Entire home/apt
## [149] Private room    Entire home/apt Entire home/apt Entire home/apt
## [153] Private room    Private room    Entire home/apt Entire home/apt
## [157] Private room    Private room    Hotel room      Entire home/apt
## [161] Private room    Entire home/apt Private room    Entire home/apt
## [165] Entire home/apt Private room    Entire home/apt Private room   
## [169] Entire home/apt Private room    Entire home/apt Private room   
## [173] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [177] Private room    Entire home/apt Entire home/apt Private room   
## [181] Entire home/apt Entire home/apt Private room    Private room   
## [185] Entire home/apt Private room    Private room    Entire home/apt
## [189] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [193] Entire home/apt Entire home/apt Entire home/apt Private room   
```

```
## [197] Entire home/apt Private room    Entire home/apt Entire home/apt
## [201] Entire home/apt Entire home/apt Private room    Private room
## [205] Private room    Private room    Private room    Entire home/apt
## [209] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [213] Private room    Private room    Entire home/apt Entire home/apt
## [217] Private room    Entire home/apt Private room    Entire home/apt
## [221] Entire home/apt Entire home/apt Entire home/apt Hotel room
## [225] Private room    Private room    Entire home/apt Entire home/apt
## [229] Private room    Private room    Entire home/apt Entire home/apt
## [233] Entire home/apt Entire home/apt Shared room     Entire home/apt
## [237] Entire home/apt Entire home/apt Private room    Entire home/apt
## [241] Entire home/apt Entire home/apt Private room    Entire home/apt
## [245] Entire home/apt Entire home/apt Entire home/apt Private room
## [249] Entire home/apt Entire home/apt Private room    Entire home/apt
## [253] Private room    Entire home/apt Entire home/apt Entire home/apt
## [257] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [261] Entire home/apt Entire home/apt Entire home/apt Private room
## [265] Private room    Private room    Entire home/apt Private room
## [269] Entire home/apt Private room    Entire home/apt Entire home/apt
## [273] Entire home/apt Private room    Entire home/apt Entire home/apt
## [277] Entire home/apt Private room    Entire home/apt Entire home/apt
## [281] Entire home/apt Private room    Private room    Private room
## [285] Entire home/apt Entire home/apt Private room    Entire home/apt
## [289] Entire home/apt Private room    Private room    Entire home/apt
## [293] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [297] Entire home/apt Private room    Entire home/apt Entire home/apt
## [301] Entire home/apt Entire home/apt Private room    Entire home/apt
## [305] Entire home/apt Private room    Entire home/apt Private room
## [309] Shared room     Entire home/apt Entire home/apt Entire home/apt
## [313] Private room    Entire home/apt Entire home/apt Private room
## [317] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [321] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [325] Private room    Entire home/apt Entire home/apt Entire home/apt
## [329] Entire home/apt Private room    Entire home/apt Entire home/apt
## [333] Private room    Entire home/apt Entire home/apt Private room
## [337] Private room    Entire home/apt Entire home/apt Entire home/apt
## [341] Entire home/apt Private room    Private room    Entire home/apt
## [345] Private room    Private room    Entire home/apt Entire home/apt
## [349] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [353] Private room    Private room    Entire home/apt Private room
## [357] Entire home/apt Private room    Entire home/apt Private room
## [361] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [365] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [369] Private room    Entire home/apt Private room    Entire home/apt
## [373] Entire home/apt Private room    Entire home/apt Entire home/apt
## [377] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [381] Entire home/apt Private room    Entire home/apt Entire home/apt
## [385] Private room    Private room    Private room    Private room
## [389] Private room    Entire home/apt Entire home/apt Entire home/apt
## [393] Private room    Entire home/apt Entire home/apt Private room
## [397] Private room    Private room    Entire home/apt Private room
## [401] Private room    Private room    Entire home/apt Entire home/apt
## [405] Entire home/apt Private room    Entire home/apt Private room
## [409] Entire home/apt Private room    Entire home/apt Entire home/apt
```

```
## [413] Private room    Entire home/apt Private room    Private room
## [417] Entire home/apt Private room    Entire home/apt Entire home/apt
## [421] Entire home/apt Private room    Entire home/apt Entire home/apt
## [425] Entire home/apt Entire home/apt Private room    Entire home/apt
## [429] Private room    Entire home/apt Private room    Entire home/apt
## [433] Private room    Private room    Private room    Private room
## [437] Private room    Entire home/apt Entire home/apt Private room
## [441] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [445] Entire home/apt Private room    Entire home/apt Entire home/apt
## [449] Entire home/apt Entire home/apt Entire home/apt Private room
## [453] Entire home/apt Private room    Entire home/apt Entire home/apt
## [457] Entire home/apt Entire home/apt Entire home/apt Private room
## [461] Private room    Private room    Private room    Private room
## [465] Private room    Entire home/apt Private room    Entire home/apt
## [469] Entire home/apt Entire home/apt Entire home/apt Private room
## [473] Private room    Private room    Private room    Entire home/apt
## [477] Private room    Entire home/apt Entire home/apt Entire home/apt
## [481] Entire home/apt Entire home/apt Private room    Entire home/apt
## [485] Entire home/apt Entire home/apt Entire home/apt Private room
## [489] Shared room     Entire home/apt Private room    Private room
## [493] Entire home/apt Private room    Entire home/apt Entire home/apt
## [497] Entire home/apt Private room    Entire home/apt Entire home/apt
## [501] Entire home/apt Entire home/apt Private room    Entire home/apt
## [505] Entire home/apt Private room    Private room    Entire home/apt
## [509] Entire home/apt Entire home/apt Entire home/apt Private room
## [513] Entire home/apt Entire home/apt Entire home/apt Private room
## [517] Entire home/apt Entire home/apt Private room    Private room
## [521] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [525] Entire home/apt Private room    Entire home/apt Entire home/apt
## [529] Entire home/apt Entire home/apt Private room    Entire home/apt
## [533] Private room    Private room    Entire home/apt Hotel room
## [537] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [541] Private room    Entire home/apt Entire home/apt Entire home/apt
## [545] Private room    Private room    Private room    Entire home/apt
## [549] Private room    Entire home/apt Entire home/apt Entire home/apt
## [553] Entire home/apt Private room    Private room    Entire home/apt
## [557] Private room    Entire home/apt Private room    Private room
## [561] Entire home/apt Entire home/apt Entire home/apt Private room
## [565] Private room    Entire home/apt Entire home/apt Entire home/apt
## [569] Entire home/apt Private room    Entire home/apt Private room
## [573] Private room    Entire home/apt Entire home/apt Entire home/apt
## [577] Entire home/apt Private room    Hotel room      Private room
## [581] Private room    Entire home/apt Private room    Private room
## [585] Entire home/apt Entire home/apt Private room    Entire home/apt
## [589] Entire home/apt Entire home/apt Entire home/apt Private room
## [593] Entire home/apt Entire home/apt Private room    Entire home/apt
## [597] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [601] Private room    Entire home/apt Entire home/apt Private room
## [605] Private room    Private room    Private room    Entire home/apt
## [609] Entire home/apt Private room    Entire home/apt Entire home/apt
## [613] Private room    Private room    Private room    Entire home/apt
## [617] Entire home/apt Private room    Private room    Private room
## [621] Private room    Entire home/apt Entire home/apt Entire home/apt
## [625] Hotel room      Entire home/apt Private room    Private room
```

```
## [629] Entire home/apt Entire home/apt Private room    Entire home/apt
## [633] Entire home/apt Private room    Private room    Entire home/apt
## [637] Entire home/apt Entire home/apt Entire home/apt Private room
## [641] Private room    Private room    Entire home/apt Entire home/apt
## [645] Private room    Private room    Entire home/apt Private room
## [649] Private room    Entire home/apt Entire home/apt Entire home/apt
## [653] Entire home/apt Private room    Entire home/apt Private room
## [657] Entire home/apt Private room    Entire home/apt Entire home/apt
## [661] Private room    Private room    Private room    Entire home/apt
## [665] Entire home/apt Private room    Private room    Private room
## [669] Entire home/apt Private room    Entire home/apt Private room
## [673] Entire home/apt Entire home/apt Entire home/apt Private room
## [677] Private room    Entire home/apt Private room    Private room
## [681] Private room    Hotel room      Private room    Entire home/apt
## [685] Private room    Entire home/apt Entire home/apt Private room
## [689] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [693] Private room    Entire home/apt Private room    Entire home/apt
## [697] Entire home/apt Entire home/apt Private room    Entire home/apt
## [701] Private room    Private room    Entire home/apt Private room
## [705] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [709] Entire home/apt Entire home/apt Private room    Entire home/apt
## [713] Private room    Entire home/apt Entire home/apt Private room
## [717] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [721] Entire home/apt Entire home/apt Private room    Private room
## [725] Entire home/apt Entire home/apt Private room    Entire home/apt
## [729] Entire home/apt Entire home/apt Entire home/apt Private room
## [733] Entire home/apt Private room    Private room    Entire home/apt
## [737] Private room    Entire home/apt Entire home/apt Entire home/apt
## [741] Entire home/apt Private room    Entire home/apt Entire home/apt
## [745] Private room    Private room    Entire home/apt Entire home/apt
## [749] Entire home/apt Entire home/apt Entire home/apt Private room
## [753] Entire home/apt Entire home/apt Private room    Private room
## [757] Private room    Private room    Private room    Entire home/apt
## [761] Entire home/apt Entire home/apt Private room    Entire home/apt
## [765] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [769] Private room    Private room    Private room    Entire home/apt
## [773] Entire home/apt Private room    Entire home/apt Private room
## [777] Entire home/apt Entire home/apt Private room    Entire home/apt
## [781] Private room    Entire home/apt Entire home/apt Entire home/apt
## [785] Entire home/apt Entire home/apt Private room    Entire home/apt
## [789] Entire home/apt Private room    Private room    Private room
## [793] Private room    Entire home/apt Entire home/apt Private room
## [797] Entire home/apt Private room    Entire home/apt Entire home/apt
## [801] Private room    Entire home/apt Entire home/apt Entire home/apt
## [805] Entire home/apt Entire home/apt Private room    Entire home/apt
## [809] Private room    Entire home/apt Entire home/apt Entire home/apt
## [813] Entire home/apt Entire home/apt Entire home/apt Private room
## [817] Entire home/apt Private room    Entire home/apt Entire home/apt
## [821] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [825] Entire home/apt Entire home/apt Private room    Entire home/apt
## [829] Entire home/apt Private room    Private room    Entire home/apt
## [833] Entire home/apt Private room    Entire home/apt Entire home/apt
## [837] Entire home/apt Private room    Entire home/apt Entire home/apt
## [841] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
```

```
## [845] Entire home/apt Entire home/apt Private room    Entire home/apt
## [849] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [853] Private room    Private room    Entire home/apt Entire home/apt
## [857] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [861] Entire home/apt Private room    Entire home/apt Entire home/apt
## [865] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [869] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [873] Entire home/apt Private room    Private room    Shared room
## [877] Entire home/apt Private room    Entire home/apt Entire home/apt
## [881] Entire home/apt Entire home/apt Entire home/apt Private room
## [885] Private room    Entire home/apt Entire home/apt Private room
## [889] Private room    Entire home/apt Entire home/apt Private room
## [893] Private room    Entire home/apt Private room    Private room
## [897] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [901] Private room    Entire home/apt Entire home/apt Private room
## [905] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [909] Entire home/apt Private room    Entire home/apt Entire home/apt
## [913] Entire home/apt Entire home/apt Private room    Private room
## [917] Entire home/apt Entire home/apt Private room    Entire home/apt
## [921] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [925] Private room    Private room    Private room    Entire home/apt
## [929] Private room    Entire home/apt Entire home/apt Entire home/apt
## [933] Entire home/apt Private room    Entire home/apt Entire home/apt
## [937] Entire home/apt Private room
## Levels: Entire home/apt Hotel room Private room Shared room
```

```r
# Number of properties classified in each group

table(lda_Y_test)
```

```
## lda_Y_test
## Entire home/apt      Hotel room    Private room     Shared room
##             598               6             330               4
```

```r
# Contingency table with good and bad classifications

table(Y_test,lda_Y_test)
```

```
##                    lda_Y_test
## Y_test              Entire home/apt Hotel room Private room Shared room
##   Entire home/apt               533          1           73           1
##   Hotel room                     16          2            5           0
##   Private room                   47          3          245           0
##   Shared room                     2          0            7           3
```

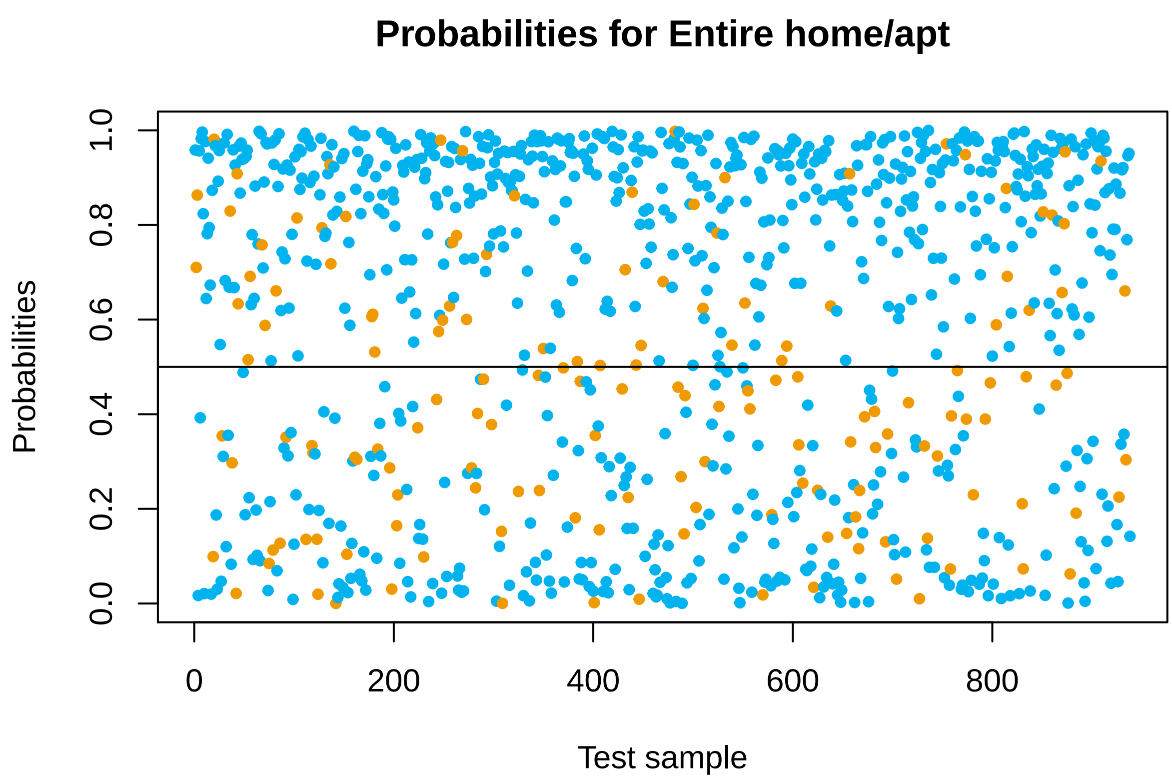
```r
# Test Error Rate (TER)

lda_TER <- mean(Y_test!=lda_Y_test)
lda_TER
```

```
## [1] 0.1652452
```

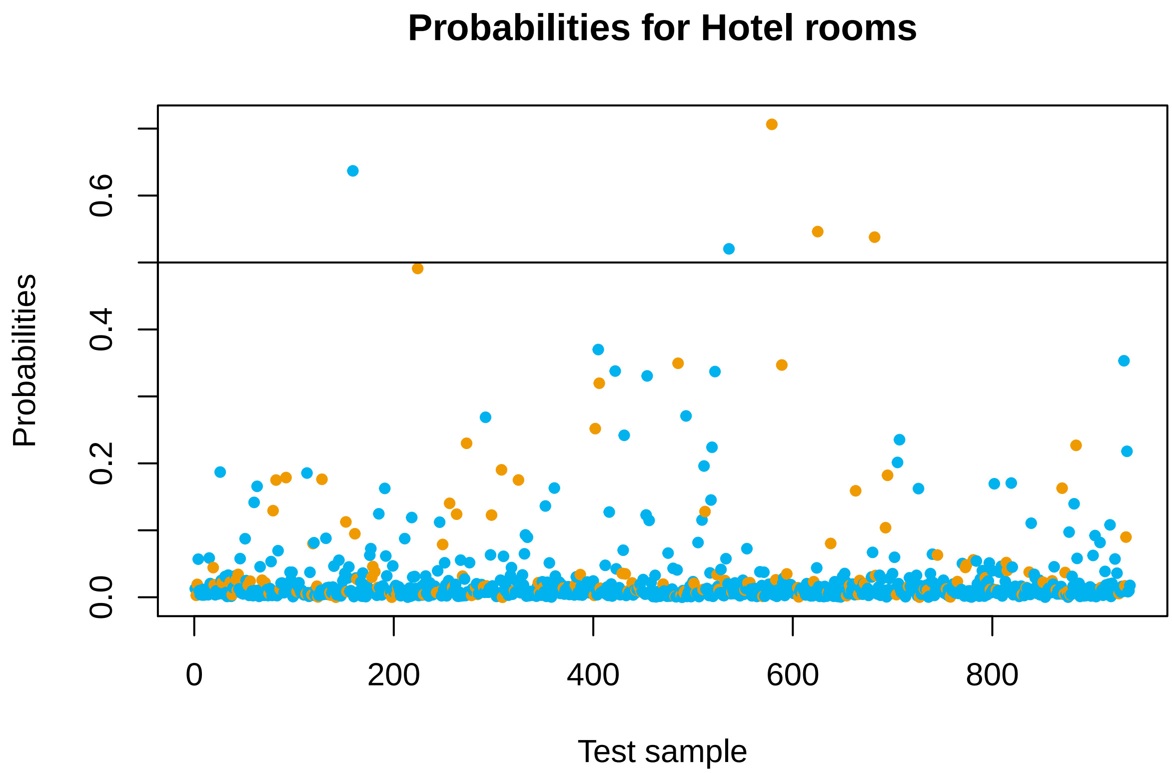
```r
# Posterior probabilities of the classifications made with the test sample

prob_lda_Y_test <- lda_test$posterior
head(prob_lda_Y_test)
```

```
##        Entire home/apt  Hotel room Private room  Shared room
## 6810       0.95839629 0.012743932   0.02849878 3.610013e-04
## 1416       0.71030384 0.002906008   0.28672623 6.392424e-05
## 20249      0.86307476 0.019215430   0.11763311 7.669757e-05
## 18193      0.01686829 0.056929726   0.91831547 7.886515e-03
## 9928       0.95656104 0.009882895   0.03348248 7.358415e-05
## 16109      0.39243039 0.004043827   0.58351064 2.001514e-02
```

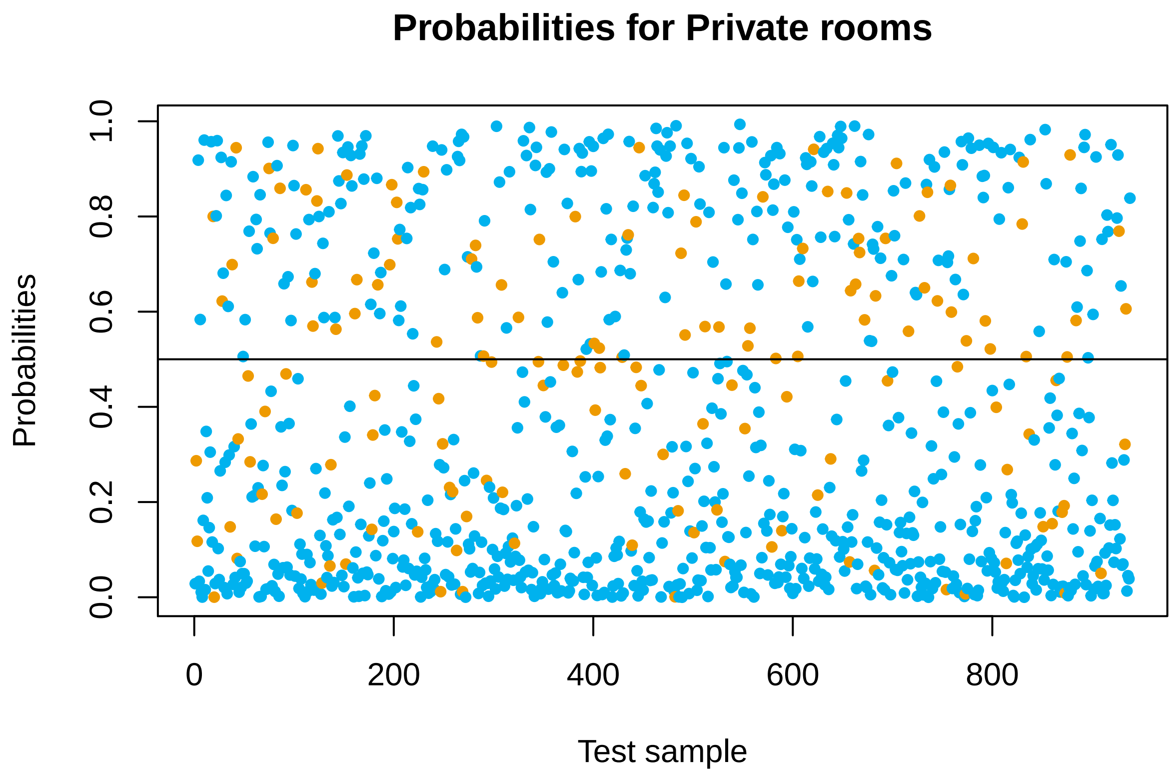
```
# Make a plot of the probabilities of different properties
# In blue, good classifications, in orange, wrong classifications

#### Probabilities for Entire home/apt
colors_errors <- c(color_3,color_1)[1*(Y_test==lda_Y_test)+1]
plot(1:n_test,prob_lda_Y_test[,1],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabilit
     main="Probabilities for Entire home/apt")
abline(h=0.5)
```



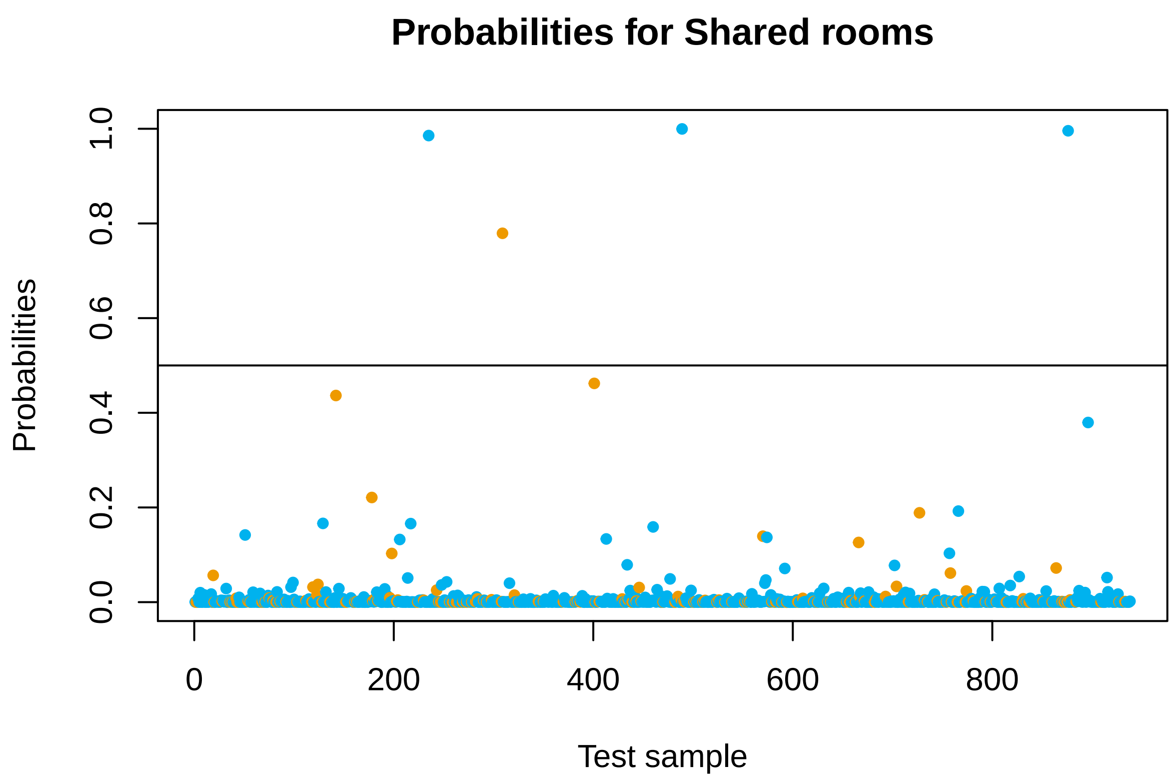

```
#### Probabilities for Hotel rooms
plot(1:n_test,prob_lda_Y_test[,2],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabilit
     main="Probabilities for Hotel rooms")
abline(h=0.5)
```




```r
#### Probabilities for Private rooms
plot(1:n_test,prob_lda_Y_test[,3],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabilit
     main="Probabilities for Private rooms")
abline(h=0.5)
```




```
#### Probabilities for Shared rooms
plot(1:n_test,prob_lda_Y_test[,4],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabilit
     main="Probabilities for Shared rooms")
abline(h=0.5)
```




```
# Quadratic Discriminant Analysis (QDA)

# Obtaining a training data set and test data set with the majority classes

n_qda = nrow(small_X_qda)

# Number of observations in each set

n_train_qda <- floor(0.7*n_qda)
n_test_qda <- n_qda - n_train_qda

# Obtain the indices of the observations in both data sets at random
i_train_qda <- sort(sample(1:n_qda,n_train_qda))
length(i_train_qda)
```

```
## [1] 2108
```

```
i_test_qda <- setdiff(1:n_qda,i_train_qda)

small_X_qda_train <- small_X_qda[i_train_qda,]
Y_train_qda <- y_qda[i_train_qda,]
small_X_qda_test <- small_X_qda[i_test_qda,]
Y_test_qda <- y_qda[i_test_qda,]


# Estimate the unknown parameters with the training sample
```

```r
qda_train <- qda(Y_train_qda ~ .,data=small_X_qda_train)


qda_train$prior
```

```
## Entire home/apt    Private room
##       0.6342505       0.3657495
```

```r
# Estimated sample mean vectors

t(qda_train$means)
```

```
##                              Entire home/apt Private room
## host_response_rate               4.450432798   4.32786698
## host_total_listings_count       22.249391922   7.82253048
## accommodates                     4.158068063   1.85314008
## bathrooms                        0.185925707   0.13239401
## bedrooms                        -0.308845876  -0.12776599
## beds                             0.638244464  -0.03290959
## price                            4.472841571   3.76156275
## cleaning_fee                     2.582554623   0.52023331
## guests_included                  2.098232610   1.08660311
## extra_people                    -1.858815690  -3.50589626
## minimum_nights                   2.542510845   2.30320493
## maximum_nights                 273.305412865 221.32136316
## availability_30                 -0.737188951  -1.76077842
## availability_60                  0.651910538  -0.77838502
## availability_90                  1.327532059  -0.23159558
## availability_365                 3.016291613   1.34308133
## number_of_reviews                1.288623944  -0.61057804
## review_scores_accuracy           9.515584892   9.58336057
## review_scores_cleanliness        9.360760658   9.37583787
## review_scores_checkin            9.630768138   9.68193385
## review_scores_communication      9.656198205   9.67026070
## review_scores_location           9.710798055   9.59243969
## review_scores_value              9.152832461   9.28764073
## reviews_per_month                0.006428501  -0.41980625
```

```r
# The function does not return the estimated covariance matrices

# Classify the observations in the test sample

qda_test <- predict(qda_train,newdata=small_X_qda_test)

# The vector of classifications made can be found here

qda_Y_test <- qda_test$class
qda_Y_test
```

```
##   [1] Entire home/apt Private room    Private room    Entire home/apt
##   [5] Private room    Private room    Private room    Private room
##   [9] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
##  [13] Entire home/apt Private room    Entire home/apt Private room
##  [17] Entire home/apt Entire home/apt Private room    Entire home/apt
##  [21] Entire home/apt Entire home/apt Private room    Private room
```

```
##  [25] Entire home/apt Private room    Entire home/apt Private room   
##  [29] Private room    Private room    Private room    Entire home/apt
##  [33] Entire home/apt Private room    Entire home/apt Entire home/apt
##  [37] Private room    Private room    Private room    Entire home/apt
##  [41] Entire home/apt Private room    Entire home/apt Private room   
##  [45] Private room    Private room    Entire home/apt Entire home/apt
##  [49] Private room    Private room    Private room    Entire home/apt
##  [53] Private room    Private room    Entire home/apt Private room   
##  [57] Private room    Entire home/apt Private room    Entire home/apt
##  [61] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
##  [65] Entire home/apt Entire home/apt Private room    Private room   
##  [69] Private room    Private room    Entire home/apt Private room   
##  [73] Entire home/apt Entire home/apt Private room    Entire home/apt
##  [77] Private room    Entire home/apt Private room    Entire home/apt
##  [81] Private room    Entire home/apt Entire home/apt Entire home/apt
##  [85] Entire home/apt Private room    Entire home/apt Private room   
##  [89] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
##  [93] Private room    Entire home/apt Entire home/apt Private room   
##  [97] Entire home/apt Entire home/apt Entire home/apt Private room   
## [101] Entire home/apt Private room    Private room    Private room   
## [105] Private room    Private room    Entire home/apt Entire home/apt
## [109] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [113] Entire home/apt Entire home/apt Entire home/apt Private room   
## [117] Entire home/apt Private room    Entire home/apt Private room   
## [121] Entire home/apt Private room    Entire home/apt Entire home/apt
## [125] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [129] Private room    Private room    Entire home/apt Entire home/apt
## [133] Entire home/apt Entire home/apt Private room    Entire home/apt
## [137] Entire home/apt Private room    Entire home/apt Private room   
## [141] Entire home/apt Private room    Entire home/apt Private room   
## [145] Entire home/apt Private room    Private room    Private room   
## [149] Entire home/apt Private room    Entire home/apt Private room   
## [153] Entire home/apt Entire home/apt Private room    Entire home/apt
## [157] Entire home/apt Private room    Private room    Entire home/apt
## [161] Entire home/apt Private room    Entire home/apt Private room   
## [165] Entire home/apt Private room    Private room    Private room   
## [169] Entire home/apt Private room    Entire home/apt Entire home/apt
## [173] Private room    Entire home/apt Entire home/apt Private room   
## [177] Entire home/apt Entire home/apt Private room    Entire home/apt
## [181] Entire home/apt Entire home/apt Private room    Entire home/apt
## [185] Entire home/apt Entire home/apt Private room    Entire home/apt
## [189] Entire home/apt Entire home/apt Entire home/apt Private room   
## [193] Private room    Private room    Private room    Entire home/apt
## [197] Entire home/apt Entire home/apt Entire home/apt Private room   
## [201] Private room    Entire home/apt Private room    Entire home/apt
## [205] Entire home/apt Entire home/apt Entire home/apt Private room   
## [209] Entire home/apt Private room    Private room    Private room   
## [213] Private room    Entire home/apt Entire home/apt Entire home/apt
## [217] Entire home/apt Entire home/apt Private room    Entire home/apt
## [221] Entire home/apt Entire home/apt Private room    Entire home/apt
## [225] Entire home/apt Entire home/apt Entire home/apt Private room   
## [229] Entire home/apt Entire home/apt Private room    Entire home/apt
## [233] Private room    Private room    Entire home/apt Private room   
## [237] Private room    Entire home/apt Entire home/apt Entire home/apt
```

```
## [241] Entire home/apt Entire home/apt Private room    Entire home/apt
## [245] Entire home/apt Entire home/apt Private room    Entire home/apt
## [249] Private room    Private room    Entire home/apt Private room
## [253] Private room    Private room    Entire home/apt Entire home/apt
## [257] Private room    Private room    Private room    Private room
## [261] Private room    Private room    Entire home/apt Private room
## [265] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [269] Entire home/apt Private room    Private room    Entire home/apt
## [273] Private room    Entire home/apt Entire home/apt Entire home/apt
## [277] Private room    Private room    Private room    Entire home/apt
## [281] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [285] Private room    Private room    Private room    Private room
## [289] Entire home/apt Private room    Private room    Entire home/apt
## [293] Entire home/apt Private room    Entire home/apt Private room
## [297] Private room    Entire home/apt Private room    Entire home/apt
## [301] Private room    Private room    Private room    Private room
## [305] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [309] Private room    Private room    Entire home/apt Private room
## [313] Entire home/apt Entire home/apt Entire home/apt Private room
## [317] Private room    Entire home/apt Private room    Private room
## [321] Private room    Entire home/apt Private room    Entire home/apt
## [325] Private room    Private room    Entire home/apt Entire home/apt
## [329] Entire home/apt Entire home/apt Private room    Entire home/apt
## [333] Entire home/apt Private room    Private room    Private room
## [337] Private room    Entire home/apt Private room    Private room
## [341] Entire home/apt Entire home/apt Entire home/apt Private room
## [345] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [349] Private room    Private room    Private room    Entire home/apt
## [353] Entire home/apt Private room    Private room    Entire home/apt
## [357] Private room    Private room    Entire home/apt Private room
## [361] Entire home/apt Entire home/apt Private room    Entire home/apt
## [365] Private room    Entire home/apt Private room    Private room
## [369] Private room    Private room    Entire home/apt Entire home/apt
## [373] Private room    Private room    Entire home/apt Entire home/apt
## [377] Entire home/apt Entire home/apt Private room    Private room
## [381] Private room    Entire home/apt Entire home/apt Entire home/apt
## [385] Entire home/apt Entire home/apt Private room    Private room
## [389] Entire home/apt Entire home/apt Entire home/apt Private room
## [393] Private room    Private room    Entire home/apt Entire home/apt
## [397] Entire home/apt Entire home/apt Entire home/apt Private room
## [401] Entire home/apt Private room    Private room    Private room
## [405] Private room    Entire home/apt Entire home/apt Entire home/apt
## [409] Entire home/apt Private room    Entire home/apt Private room
## [413] Entire home/apt Private room    Private room    Entire home/apt
## [417] Private room    Entire home/apt Entire home/apt Entire home/apt
## [421] Entire home/apt Private room    Entire home/apt Entire home/apt
## [425] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [429] Private room    Private room    Entire home/apt Private room
## [433] Private room    Private room    Private room    Entire home/apt
## [437] Entire home/apt Entire home/apt Entire home/apt Private room
## [441] Private room    Private room    Entire home/apt Entire home/apt
## [445] Entire home/apt Private room    Entire home/apt Private room
## [449] Private room    Entire home/apt Entire home/apt Entire home/apt
## [453] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
```

```
## [457] Entire home/apt Private room    Private room    Entire home/apt
## [461] Private room    Private room    Entire home/apt Private room
## [465] Private room    Private room    Private room    Entire home/apt
## [469] Entire home/apt Entire home/apt Private room    Private room
## [473] Entire home/apt Entire home/apt Private room    Entire home/apt
## [477] Private room    Private room    Entire home/apt Entire home/apt
## [481] Private room    Entire home/apt Entire home/apt Private room
## [485] Entire home/apt Private room    Entire home/apt Private room
## [489] Entire home/apt Entire home/apt Private room    Private room
## [493] Private room    Private room    Private room    Private room
## [497] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [501] Entire home/apt Private room    Entire home/apt Entire home/apt
## [505] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [509] Private room    Entire home/apt Entire home/apt Entire home/apt
## [513] Private room    Private room    Private room    Private room
## [517] Private room    Private room    Private room    Private room
## [521] Entire home/apt Private room    Private room    Entire home/apt
## [525] Private room    Private room    Entire home/apt Private room
## [529] Entire home/apt Private room    Entire home/apt Private room
## [533] Entire home/apt Private room    Private room    Entire home/apt
## [537] Entire home/apt Entire home/apt Private room    Private room
## [541] Entire home/apt Entire home/apt Private room    Private room
## [545] Private room    Entire home/apt Private room    Private room
## [549] Private room    Entire home/apt Entire home/apt Entire home/apt
## [553] Private room    Entire home/apt Entire home/apt Entire home/apt
## [557] Private room    Entire home/apt Private room    Entire home/apt
## [561] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [565] Entire home/apt Private room    Entire home/apt Private room
## [569] Entire home/apt Entire home/apt Entire home/apt Private room
## [573] Private room    Private room    Entire home/apt Private room
## [577] Entire home/apt Entire home/apt Private room    Entire home/apt
## [581] Entire home/apt Private room    Entire home/apt Entire home/apt
## [585] Private room    Entire home/apt Entire home/apt Private room
## [589] Private room    Entire home/apt Private room    Entire home/apt
## [593] Entire home/apt Private room    Private room    Entire home/apt
## [597] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [601] Private room    Private room    Entire home/apt Entire home/apt
## [605] Entire home/apt Entire home/apt Entire home/apt Private room
## [609] Entire home/apt Entire home/apt Entire home/apt Private room
## [613] Private room    Entire home/apt Private room    Private room
## [617] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [621] Private room    Private room    Entire home/apt Private room
## [625] Private room    Private room    Private room    Private room
## [629] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [633] Private room    Entire home/apt Private room    Entire home/apt
## [637] Private room    Private room    Entire home/apt Entire home/apt
## [641] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [645] Entire home/apt Entire home/apt Private room    Private room
## [649] Entire home/apt Entire home/apt Private room    Private room
## [653] Private room    Entire home/apt Private room    Private room
## [657] Private room    Entire home/apt Entire home/apt Entire home/apt
## [661] Private room    Entire home/apt Entire home/apt Entire home/apt
## [665] Entire home/apt Private room    Entire home/apt Entire home/apt
## [669] Private room    Entire home/apt Entire home/apt Private room
```

```
## [673] Private room    Private room    Private room    Entire home/apt
## [677] Entire home/apt Entire home/apt Private room    Private room
## [681] Private room    Entire home/apt Private room    Private room
## [685] Private room    Entire home/apt Private room    Private room
## [689] Entire home/apt Entire home/apt Entire home/apt Private room
## [693] Private room    Entire home/apt Entire home/apt Entire home/apt
## [697] Entire home/apt Entire home/apt Private room    Entire home/apt
## [701] Entire home/apt Private room    Entire home/apt Entire home/apt
## [705] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [709] Private room    Entire home/apt Entire home/apt Entire home/apt
## [713] Entire home/apt Private room    Private room    Entire home/apt
## [717] Private room    Private room    Entire home/apt Private room
## [721] Entire home/apt Private room    Private room    Private room
## [725] Private room    Private room    Entire home/apt Private room
## [729] Private room    Entire home/apt Private room    Entire home/apt
## [733] Private room    Entire home/apt Entire home/apt Private room
## [737] Private room    Entire home/apt Entire home/apt Private room
## [741] Entire home/apt Private room    Entire home/apt Private room
## [745] Private room    Entire home/apt Entire home/apt Entire home/apt
## [749] Private room    Private room    Private room    Entire home/apt
## [753] Entire home/apt Entire home/apt Entire home/apt Private room
## [757] Entire home/apt Private room    Private room    Private room
## [761] Private room    Private room    Entire home/apt Private room
## [765] Private room    Entire home/apt Entire home/apt Private room
## [769] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [773] Entire home/apt Entire home/apt Entire home/apt Private room
## [777] Private room    Entire home/apt Private room    Private room
## [781] Private room    Private room    Private room    Entire home/apt
## [785] Private room    Entire home/apt Entire home/apt Entire home/apt
## [789] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [793] Entire home/apt Private room    Private room    Entire home/apt
## [797] Private room    Private room    Private room    Entire home/apt
## [801] Entire home/apt Private room    Entire home/apt Private room
## [805] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [809] Entire home/apt Private room    Entire home/apt Entire home/apt
## [813] Entire home/apt Private room    Private room    Entire home/apt
## [817] Entire home/apt Private room    Entire home/apt Private room
## [821] Private room    Entire home/apt Private room    Private room
## [825] Entire home/apt Private room    Private room    Private room
## [829] Entire home/apt Entire home/apt Private room    Entire home/apt
## [833] Entire home/apt Entire home/apt Entire home/apt Private room
## [837] Entire home/apt Private room    Private room    Entire home/apt
## [841] Entire home/apt Private room    Private room    Private room
## [845] Entire home/apt Private room    Entire home/apt Entire home/apt
## [849] Private room    Private room    Entire home/apt Entire home/apt
## [853] Private room    Entire home/apt Private room    Private room
## [857] Private room    Entire home/apt Entire home/apt Private room
## [861] Private room    Entire home/apt Entire home/apt Entire home/apt
## [865] Private room    Entire home/apt Entire home/apt Entire home/apt
## [869] Private room    Entire home/apt Entire home/apt Entire home/apt
## [873] Private room    Private room    Entire home/apt Entire home/apt
## [877] Entire home/apt Entire home/apt Entire home/apt Entire home/apt
## [881] Entire home/apt Private room    Private room    Entire home/apt
## [885] Private room    Private room    Private room    Private room
```

```
## [889] Entire home/apt Entire home/apt Entire home/apt Private room
## [893] Entire home/apt Entire home/apt Private room    Entire home/apt
## [897] Entire home/apt Private room    Private room    Private room
## [901] Entire home/apt Entire home/apt Entire home/apt Private room
## Levels: Entire home/apt Private room
```

```r
# Number of properties classified in each group

table(qda_Y_test)
```

```
## qda_Y_test
## Entire home/apt    Private room
##             500             404
```

```r
# Contingency table with good and bad classifications

table(Y_test_qda,qda_Y_test)
```

```
##                  qda_Y_test
## Y_test_qda        Entire home/apt Private room
##   Entire home/apt             462          111
##   Private room                 38          293
```

```r
# Test Error Rate (TER)

qda_TER <- mean(Y_test_qda!=qda_Y_test)
qda_TER
```

```
## [1] 0.164823
```

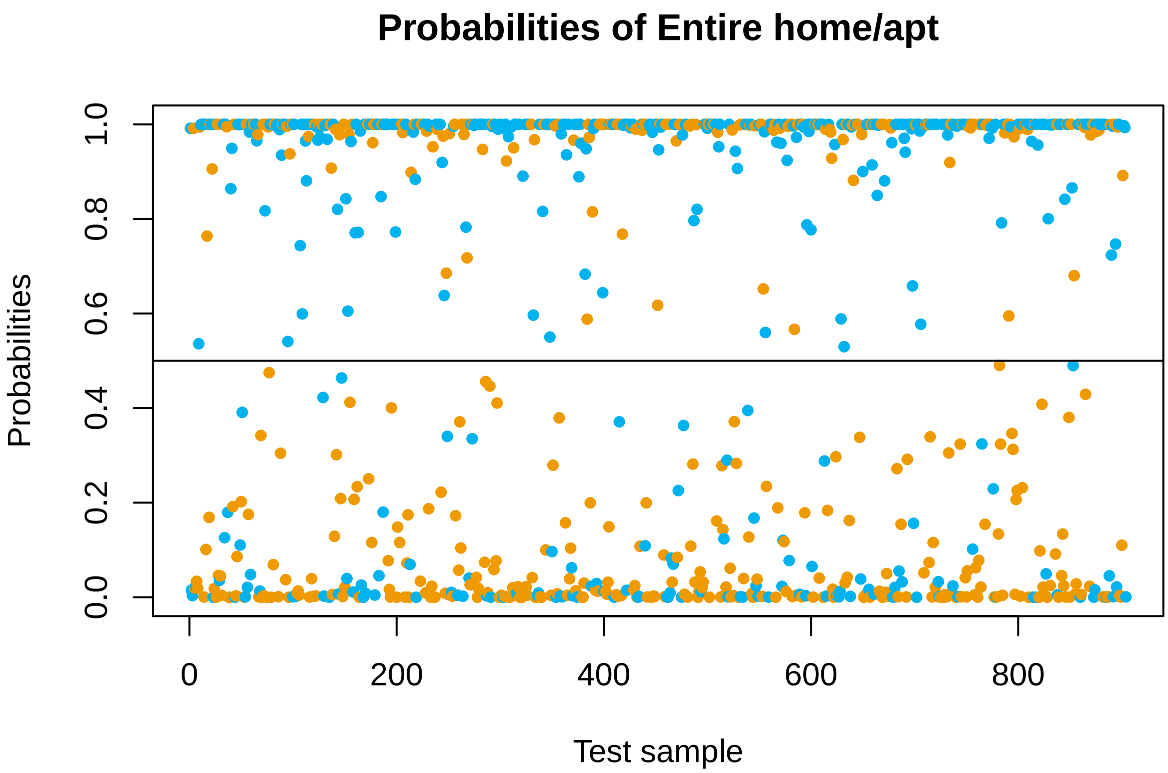
```r
# Posterior probabilities of the classifications made with the test sample

prob_qda_Y_test <- qda_test$posterior
head(prob_qda_Y_test)
```

```
##    Entire home/apt Private room
## 1      0.991740325  0.008259675
## 4      0.015150329  0.984849671
## 7      0.003342939  0.996657061
## 8      0.991459296  0.008540704
## 15     0.019955740  0.980044260
## 18     0.023833993  0.976166007
```

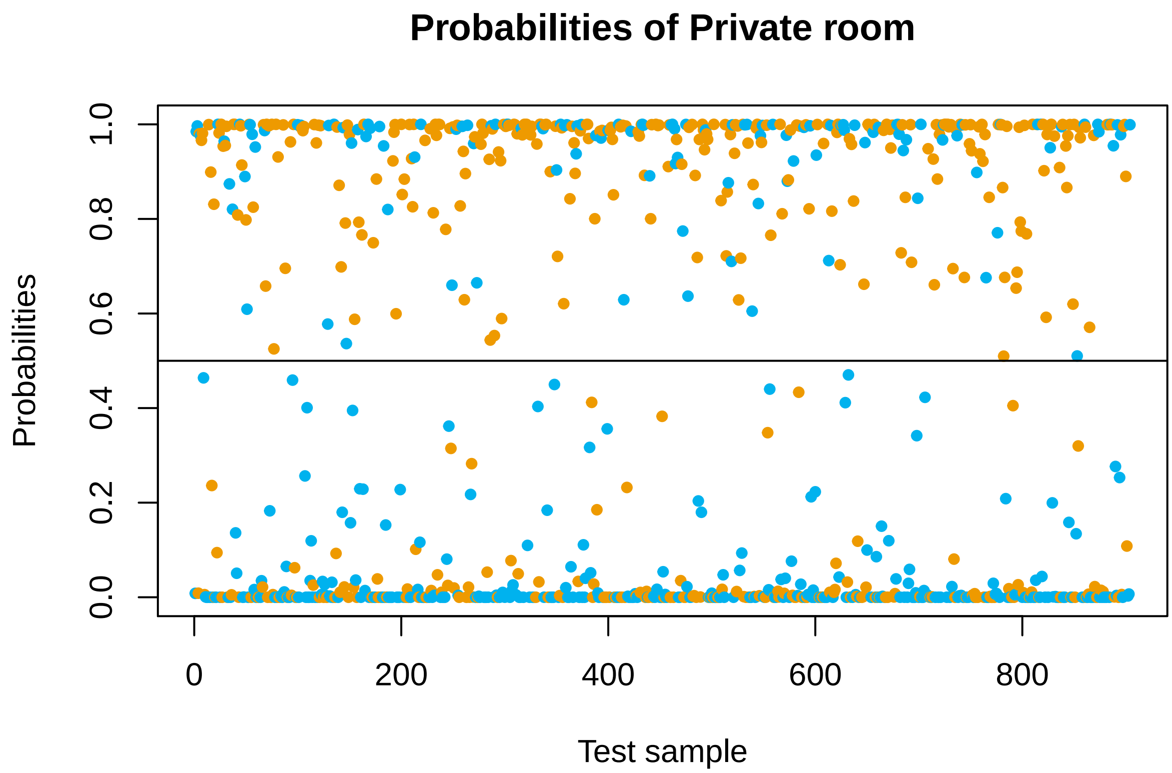
```r
# Make a plot of the probabilities for each type of property
# In blue, good classifications, in orange, wrong classifications

#### Probabilities of Entire home/apt
colors_errors <- c(color_3,color_1)[1*(Y_test==qda_Y_test)+1]
```

```
## Warning in `==.default`(Y_test, qda_Y_test): longer object length is not a
## multiple of shorter object length

## Warning in is.na(e1) | is.na(e2): longer object length is not a multiple of
## shorter object length
```

```r
plot(1:n_test_qda,prob_qda_Y_test[,1],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabi
     main="Probabilities of Entire home/apt")
abline(h=0.5)
```




```
#### "Probabilities of Private room"
plot(1:n_test_qda,prob_qda_Y_test[,2],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probab
     main="Probabilities of Private room")
abline(h=0.5)
```




```r
# Naive Bayes (NB)

#converting dataset into a matrix in order to make use of gaussian_naive_bayes()
small_x_train_nb <- as.matrix(X_train)
small_x_test_nb <- as.matrix(X_test)


nb_train <- gaussian_naive_bayes(small_x_train_nb,Y_train)
# Estimated prior probabilities

nb_train$prior
```

```
## Entire home/apt      Hotel room    Private room     Shared room
##     0.595336077     0.026520348     0.368998628     0.009144947
```

```r
# Estimated sample mean vectors
t(nb_train$params$mu)
```

```
##                             Entire home/apt  Hotel room Private room
## host_response_rate              4.444701112   4.5356610   4.34478245
## host_total_listings_count      22.091860215  19.6734138   7.76072739
## accommodates                    4.221430108   3.5182414   1.87088848
## bathrooms                       0.187345960   0.1712817   0.12836432
## bedrooms                       -0.378789987  -0.6137889  -0.09347087
## beds                            0.641895700   0.6040626  -0.02900099
## price                           4.477114731   4.6329221   3.74196207
## cleaning_fee                    2.543428102  -0.6687550   0.57956951
```

```
## guests_included              2.150001536   1.9492759   1.10013259
## extra_people                -1.832450003  -3.4654436  -3.57560386
## minimum_nights               2.636176651   2.4320345   2.31574597
## maximum_nights             273.863519201 261.8630690 227.53879430
## availability_30             -0.767170317   0.2470504  -1.72642010
## availability_60              0.701231353   1.6515606  -0.70125882
## availability_90              1.353407520   2.1656795  -0.17503722
## availability_365             3.037985975   4.0163954   1.44810724
## number_of_reviews            1.237821057  -0.2679565  -0.63405979
## review_scores_accuracy       9.502536098   9.4665172   9.60694796
## review_scores_cleanliness    9.361983103   9.5354828   9.39629120
## review_scores_checkin        9.587789555   9.6561724   9.71351549
## review_scores_communication  9.633104455   9.4665172   9.71351549
## review_scores_location       9.701460829   9.8630690   9.61066543
## review_scores_value          9.146929339   9.2596207   9.31078934
## reviews_per_month           -0.001768022  -0.4578986  -0.39967057
##                              Shared room
## host_response_rate          4.381705e+00
## host_total_listings_count   7.651000e+00
## accommodates                3.451000e+00
## bathrooms                   5.382275e-01
## bedrooms                    9.995003e-04
## beds                        7.422745e-01
## price                       3.209264e+00
## cleaning_fee                2.453346e-01
## guests_included             1.001000e+00
## extra_people               -2.914277e+00
## minimum_nights              2.751000e+00
## maximum_nights              2.563010e+02
## availability_30            -1.791647e+00
## availability_60            -5.806842e-01
## availability_90            -2.893422e-01
## availability_365            1.150029e+00
## number_of_reviews          -9.255777e-01
## review_scores_accuracy      9.301000e+00
## review_scores_cleanliness   9.551000e+00
## review_scores_checkin       9.601000e+00
## review_scores_communication 9.651000e+00
## review_scores_location      9.601000e+00
## review_scores_value         9.051000e+00
## reviews_per_month          -8.906210e-01
```

```r
# The function does not return the estimated covariance matrices
t(nb_train$params$sd)
```

```
##                            Entire home/apt  Hotel room Private room
## host_response_rate               1.0598628   0.1819237    1.3986215
## host_total_listings_count       69.7850344  18.7937914   25.4459974
## accommodates                     2.0300844   2.4512180    1.0818479
## bathrooms                        0.3354582   0.4390202    0.5324389
## bedrooms                         2.3293731   2.3796408    1.0238866
## beds                             1.0120411   0.5346035    1.2047873
## price                            0.6155694   0.8738927    0.8830698
## cleaning_fee                     3.1067632   5.1490302    4.0840721
## guests_included                  1.6357586   1.6159557    0.6121894
```

```
## extra_people                      4.8034665   4.6255095    4.5303077
## minimum_nights                    2.2046888   2.5278246    2.1638969
## maximum_nights                  145.8316562 152.5146468  168.3923569
## availability_30                   4.1033875   3.9638217    4.6381092
## availability_60                   4.2784733   3.8076985    4.9964698
## availability_90                   4.4392207   4.0100867    5.1873194
## availability_365                  4.4435243   3.8266491    5.3646460
## number_of_reviews                 4.0361912   4.4964866    4.5671372
## review_scores_accuracy            0.9083678   0.9949966    0.7292978
## review_scores_cleanliness         0.9362205   0.6809569    0.8861368
## review_scores_checkin             0.8371159   1.1010637    0.6011812
## review_scores_communication       0.8283867   1.2170646    0.6713783
## review_scores_location            0.6030793   0.3478392    0.7132603
## review_scores_value               0.9361940   0.8897140    0.8516170
## reviews_per_month                 1.3274355   1.1269926    1.3698002
##                                Shared room
## host_response_rate           2.916070e-01
## host_total_listings_count    9.777929e+00
## accommodates                 3.872644e+00
## bathrooms                    6.998007e-01
## bedrooms                     3.533071e-11
## beds                         1.001535e+00
## price                        8.551549e-01
## cleaning_fee                 4.290646e+00
## guests_included              0.000000e+00
## extra_people                 5.065976e+00
## minimum_nights               2.844663e+00
## maximum_nights               1.561639e+02
## availability_30              4.792508e+00
## availability_60              5.308465e+00
## availability_90              5.553690e+00
## availability_365             6.101050e+00
## number_of_reviews            4.190176e+00
## review_scores_accuracy       1.809333e+00
## review_scores_cleanliness    6.863327e-01
## review_scores_checkin        6.805570e-01
## review_scores_communication 6.708204e-01
## review_scores_location       9.403247e-01
## review_scores_value          1.848897e+00
## reviews_per_month            7.939661e-01
```

```r
#########################################################################################################
# Classify the observations in the test sample

nb_test <- predict(nb_train,newdata=small_x_test_nb,type="prob")

# The vector of classifications made can be found here

nb_Y_test <- c("Entire home/apt", "Hotel room", "Private room", "Shared room")[apply(nb_test,1,which.ma

nb_Y_test
```

```
##    [1] "Entire home/apt" "Shared room"     "Entire home/apt" "Shared room"
##    [5] "Entire home/apt" "Shared room"     "Entire home/apt" "Hotel room"
##    [9] "Entire home/apt" "Shared room"     "Entire home/apt" "Shared room"
```

```
##  [13] "Entire home/apt" "Entire home/apt" "Entire home/apt" "Private room"
##  [17] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
##  [21] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
##  [25] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
##  [29] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
##  [33] "Entire home/apt" "Shared room"     "Shared room"     "Entire home/apt"
##  [37] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
##  [41] "Shared room"     "Shared room"     "Shared room"     "Shared room"
##  [45] "Entire home/apt" "Entire home/apt" "Entire home/apt" "Private room"
##  [49] "Shared room"     "Shared room"     "Shared room"     "Shared room"
##  [53] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
##  [57] "Shared room"     "Shared room"     "Shared room"     "Shared room"
##  [61] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
##  [65] "Entire home/apt" "Shared room"     "Entire home/apt" "Shared room"
##  [69] "Entire home/apt" "Entire home/apt" "Shared room"     "Entire home/apt"
##  [73] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
##  [77] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
##  [81] "Entire home/apt" "Hotel room"      "Entire home/apt" "Entire home/apt"
##  [85] "Entire home/apt" "Entire home/apt" "Shared room"     "Shared room"
##  [89] "Shared room"     "Shared room"     "Shared room"     "Hotel room"
##  [93] "Shared room"     "Shared room"     "Private room"    "Shared room"
##  [97] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [101] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [105] "Entire home/apt" "Hotel room"      "Entire home/apt" "Shared room"
## [109] "Entire home/apt" "Entire home/apt" "Entire home/apt" "Shared room"
## [113] "Shared room"     "Entire home/apt" "Shared room"     "Entire home/apt"
## [117] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [121] "Shared room"     "Shared room"     "Entire home/apt" "Private room"
## [125] "Shared room"     "Private room"    "Entire home/apt" "Hotel room"
## [129] "Private room"    "Shared room"     "Shared room"     "Shared room"
## [133] "Entire home/apt" "Entire home/apt" "Shared room"     "Shared room"
## [137] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [141] "Shared room"     "Shared room"     "Private room"    "Shared room"
## [145] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [149] "Shared room"     "Entire home/apt" "Private room"    "Hotel room"
## [153] "Shared room"     "Shared room"     "Shared room"     "Private room"
## [157] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [161] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [165] "Hotel room"      "Shared room"     "Private room"    "Shared room"
## [169] "Entire home/apt" "Entire home/apt" "Shared room"     "Shared room"
## [173] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [177] "Shared room"     "Shared room"     "Hotel room"      "Shared room"
## [181] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [185] "Hotel room"      "Shared room"     "Shared room"     "Entire home/apt"
## [189] "Shared room"     "Entire home/apt" "Private room"    "Entire home/apt"
## [193] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [197] "Entire home/apt" "Shared room"     "Hotel room"      "Shared room"
## [201] "Private room"    "Entire home/apt" "Entire home/apt" "Shared room"
## [205] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [209] "Shared room"     "Entire home/apt" "Hotel room"      "Entire home/apt"
## [213] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [217] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [221] "Private room"    "Shared room"     "Entire home/apt" "Hotel room"
## [225] "Shared room"     "Shared room"     "Hotel room"      "Entire home/apt"
```

```
## [229] "Shared room"      "Shared room"      "Hotel room"       "Entire home/apt"
## [233] "Entire home/apt" "Shared room"      "Shared room"      "Entire home/apt"
## [237] "Entire home/apt" "Entire home/apt" "Shared room"      "Entire home/apt"
## [241] "Shared room"      "Shared room"      "Shared room"      "Hotel room"
## [245] "Shared room"      "Shared room"      "Entire home/apt" "Shared room"
## [249] "Hotel room"       "Shared room"      "Shared room"      "Shared room"
## [253] "Entire home/apt" "Entire home/apt" "Entire home/apt" "Hotel room"
## [257] "Entire home/apt" "Entire home/apt" "Shared room"      "Shared room"
## [261] "Entire home/apt" "Entire home/apt" "Shared room"      "Shared room"
## [265] "Shared room"      "Shared room"      "Hotel room"       "Shared room"
## [269] "Shared room"      "Shared room"      "Shared room"      "Hotel room"
## [273] "Shared room"      "Shared room"      "Entire home/apt" "Shared room"
## [277] "Entire home/apt" "Private room"     "Shared room"      "Private room"
## [281] "Entire home/apt" "Shared room"      "Shared room"      "Shared room"
## [285] "Entire home/apt" "Entire home/apt" "Shared room"      "Shared room"
## [289] "Entire home/apt" "Shared room"      "Shared room"      "Hotel room"
## [293] "Shared room"      "Shared room"      "Entire home/apt" "Private room"
## [297] "Entire home/apt" "Shared room"      "Shared room"      "Shared room"
## [301] "Shared room"      "Shared room"      "Shared room"      "Shared room"
## [305] "Entire home/apt" "Shared room"      "Private room"     "Shared room"
## [309] "Shared room"      "Entire home/apt" "Entire home/apt" "Shared room"
## [313] "Shared room"      "Entire home/apt" "Shared room"      "Shared room"
## [317] "Hotel room"       "Entire home/apt" "Shared room"      "Shared room"
## [321] "Shared room"      "Shared room"      "Shared room"      "Shared room"
## [325] "Shared room"      "Entire home/apt" "Entire home/apt" "Entire home/apt"
## [329] "Entire home/apt" "Shared room"      "Shared room"      "Entire home/apt"
## [333] "Shared room"      "Shared room"      "Shared room"      "Shared room"
## [337] "Shared room"      "Entire home/apt" "Shared room"      "Shared room"
## [341] "Entire home/apt" "Shared room"      "Shared room"      "Hotel room"
## [345] "Shared room"      "Shared room"      "Entire home/apt" "Entire home/apt"
## [349] "Entire home/apt" "Shared room"      "Shared room"      "Shared room"
## [353] "Entire home/apt" "Shared room"      "Entire home/apt" "Shared room"
## [357] "Shared room"      "Shared room"      "Shared room"      "Private room"
## [361] "Hotel room"       "Shared room"      "Shared room"      "Entire home/apt"
## [365] "Entire home/apt" "Shared room"      "Entire home/apt" "Entire home/apt"
## [369] "Shared room"      "Shared room"      "Shared room"      "Shared room"
## [373] "Shared room"      "Shared room"      "Entire home/apt" "Entire home/apt"
## [377] "Shared room"      "Entire home/apt" "Shared room"      "Shared room"
## [381] "Shared room"      "Shared room"      "Private room"     "Shared room"
## [385] "Shared room"      "Shared room"      "Shared room"      "Private room"
## [389] "Shared room"      "Entire home/apt" "Entire home/apt" "Shared room"
## [393] "Shared room"      "Hotel room"       "Entire home/apt" "Shared room"
## [397] "Shared room"      "Shared room"      "Shared room"      "Shared room"
## [401] "Shared room"      "Shared room"      "Shared room"      "Entire home/apt"
## [405] "Hotel room"       "Shared room"      "Shared room"      "Shared room"
## [409] "Shared room"      "Shared room"      "Entire home/apt" "Shared room"
## [413] "Shared room"      "Entire home/apt" "Shared room"      "Hotel room"
## [417] "Shared room"      "Shared room"      "Entire home/apt" "Entire home/apt"
## [421] "Entire home/apt" "Shared room"      "Shared room"      "Entire home/apt"
## [425] "Entire home/apt" "Shared room"      "Shared room"      "Entire home/apt"
## [429] "Shared room"      "Entire home/apt" "Shared room"      "Shared room"
## [433] "Shared room"      "Shared room"      "Entire home/apt" "Private room"
## [437] "Shared room"      "Shared room"      "Shared room"      "Shared room"
## [441] "Entire home/apt" "Shared room"      "Shared room"      "Shared room"
```

```
## [445] "Entire home/apt" "Shared room"     "Entire home/apt" "Private room"
## [449] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [453] "Shared room"     "Private room"    "Shared room"     "Entire home/apt"
## [457] "Shared room"     "Private room"    "Entire home/apt" "Shared room"
## [461] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [465] "Shared room"     "Shared room"     "Entire home/apt" "Entire home/apt"
## [469] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [473] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [477] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [481] "Entire home/apt" "Shared room"     "Shared room"     "Entire home/apt"
## [485] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [489] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [493] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [497] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [501] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [505] "Hotel room"      "Shared room"     "Shared room"     "Entire home/apt"
## [509] "Hotel room"      "Shared room"     "Shared room"     "Shared room"
## [513] "Entire home/apt" "Shared room"     "Entire home/apt" "Shared room"
## [517] "Shared room"     "Hotel room"      "Shared room"     "Shared room"
## [521] "Shared room"     "Hotel room"      "Entire home/apt" "Shared room"
## [525] "Shared room"     "Private room"    "Private room"    "Shared room"
## [529] "Shared room"     "Shared room"     "Shared room"     "Private room"
## [533] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [537] "Shared room"     "Entire home/apt" "Shared room"     "Entire home/apt"
## [541] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [545] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [549] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [553] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [557] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [561] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [565] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [569] "Entire home/apt" "Shared room"     "Shared room"     "Entire home/apt"
## [573] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [577] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [581] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [585] "Entire home/apt" "Shared room"     "Shared room"     "Entire home/apt"
## [589] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [593] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [597] "Entire home/apt" "Entire home/apt" "Entire home/apt" "Entire home/apt"
## [601] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [605] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [609] "Entire home/apt" "Shared room"     "Shared room"     "Entire home/apt"
## [613] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [617] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [621] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [625] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [629] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [633] "Shared room"     "Shared room"     "Private room"    "Entire home/apt"
## [637] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [641] "Entire home/apt" "Private room"    "Shared room"     "Shared room"
## [645] "Shared room"     "Shared room"     "Private room"    "Shared room"
## [649] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [653] "Hotel room"      "Shared room"     "Shared room"     "Shared room"
## [657] "Hotel room"      "Shared room"     "Shared room"     "Entire home/apt"
```

```
## [661] "Shared room"     "Shared room"     "Entire home/apt" "Entire home/apt"
## [665] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [669] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [673] "Entire home/apt" "Entire home/apt" "Shared room"     "Shared room"
## [677] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [681] "Shared room"     "Hotel room"      "Hotel room"      "Shared room"
## [685] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [689] "Shared room"     "Entire home/apt" "Shared room"     "Entire home/apt"
## [693] "Hotel room"      "Shared room"     "Private room"    "Shared room"
## [697] "Entire home/apt" "Entire home/apt" "Shared room"     "Shared room"
## [701] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [705] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [709] "Shared room"     "Entire home/apt" "Shared room"     "Entire home/apt"
## [713] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [717] "Entire home/apt" "Shared room"     "Private room"    "Entire home/apt"
## [721] "Shared room"     "Shared room"     "Shared room"     "Private room"
## [725] "Entire home/apt" "Hotel room"      "Shared room"     "Entire home/apt"
## [729] "Shared room"     "Private room"    "Entire home/apt" "Shared room"
## [733] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [737] "Shared room"     "Shared room"     "Entire home/apt" "Hotel room"
## [741] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [745] "Hotel room"      "Shared room"     "Entire home/apt" "Entire home/apt"
## [749] "Shared room"     "Entire home/apt" "Entire home/apt" "Shared room"
## [753] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [757] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [761] "Shared room"     "Private room"    "Shared room"     "Entire home/apt"
## [765] "Private room"    "Entire home/apt" "Entire home/apt" "Shared room"
## [769] "Shared room"     "Shared room"     "Shared room"     "Hotel room"
## [773] "Hotel room"      "Hotel room"      "Entire home/apt" "Shared room"
## [777] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [781] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [785] "Shared room"     "Private room"    "Shared room"     "Entire home/apt"
## [789] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [793] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [797] "Hotel room"      "Shared room"     "Shared room"     "Entire home/apt"
## [801] "Shared room"     "Hotel room"      "Shared room"     "Shared room"
## [805] "Shared room"     "Entire home/apt" "Shared room"     "Entire home/apt"
## [809] "Shared room"     "Entire home/apt" "Shared room"     "Shared room"
## [813] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [817] "Private room"    "Shared room"     "Shared room"     "Hotel room"
## [821] "Entire home/apt" "Entire home/apt" "Entire home/apt" "Shared room"
## [825] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [829] "Shared room"     "Private room"    "Shared room"     "Hotel room"
## [833] "Shared room"     "Shared room"     "Entire home/apt" "Entire home/apt"
## [837] "Shared room"     "Shared room"     "Entire home/apt" "Entire home/apt"
## [841] "Entire home/apt" "Shared room"     "Entire home/apt" "Entire home/apt"
## [845] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [849] "Shared room"     "Entire home/apt" "Entire home/apt" "Hotel room"
## [853] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [857] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [861] "Entire home/apt" "Shared room"     "Shared room"     "Shared room"
## [865] "Shared room"     "Private room"    "Shared room"     "Entire home/apt"
## [869] "Entire home/apt" "Shared room"     "Entire home/apt" "Shared room"
## [873] "Hotel room"      "Shared room"     "Shared room"     "Shared room"
```

```
## [877] "Hotel room"      "Shared room"     "Shared room"     "Shared room"
## [881] "Shared room"     "Shared room"     "Shared room"     "Entire home/apt"
## [885] "Shared room"     "Shared room"     "Private room"    "Shared room"
## [889] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [893] "Shared room"     "Shared room"     "Shared room"     "Shared room"
## [897] "Shared room"     "Shared room"     "Entire home/apt" "Shared room"
## [901] "Shared room"     "Entire home/apt" "Hotel room"      "Shared room"
## [905] "Shared room"     "Entire home/apt" "Entire home/apt" "Entire home/apt"
## [909] "Entire home/apt" "Shared room"     "Entire home/apt" "Entire home/apt"
## [913] "Entire home/apt" "Hotel room"      "Shared room"     "Shared room"
## [917] "Shared room"     "Hotel room"      "Shared room"     "Shared room"
## [921] "Shared room"     "Entire home/apt" "Hotel room"      "Shared room"
## [925] "Shared room"     "Private room"    "Shared room"     "Entire home/apt"
## [929] "Entire home/apt" "Entire home/apt" "Entire home/apt" "Hotel room"
## [933] "Shared room"     "Shared room"     "Hotel room"      "Entire home/apt"
## [937] "Shared room"     "Shared room"
```

```r
# Number of properties classified in each group

table(nb_Y_test)
```

```
## nb_Y_test
## Entire home/apt      Hotel room    Private room     Shared room
##             257              54              42             585
```

```r
# Contingency table with good and bad classifications

table(Y_test,nb_Y_test)
```

```
##                  nb_Y_test
## Y_test            Entire home/apt Hotel room Private room Shared room
##   Entire home/apt             242         43           31         292
##   Hotel room                    5          1            1          16
##   Private room                 10         10           10         265
##   Shared room                   0          0            0          12
```

```r
# Test Error Rate (TER)

nb_TER <- mean(Y_test!=nb_Y_test)
nb_TER
```

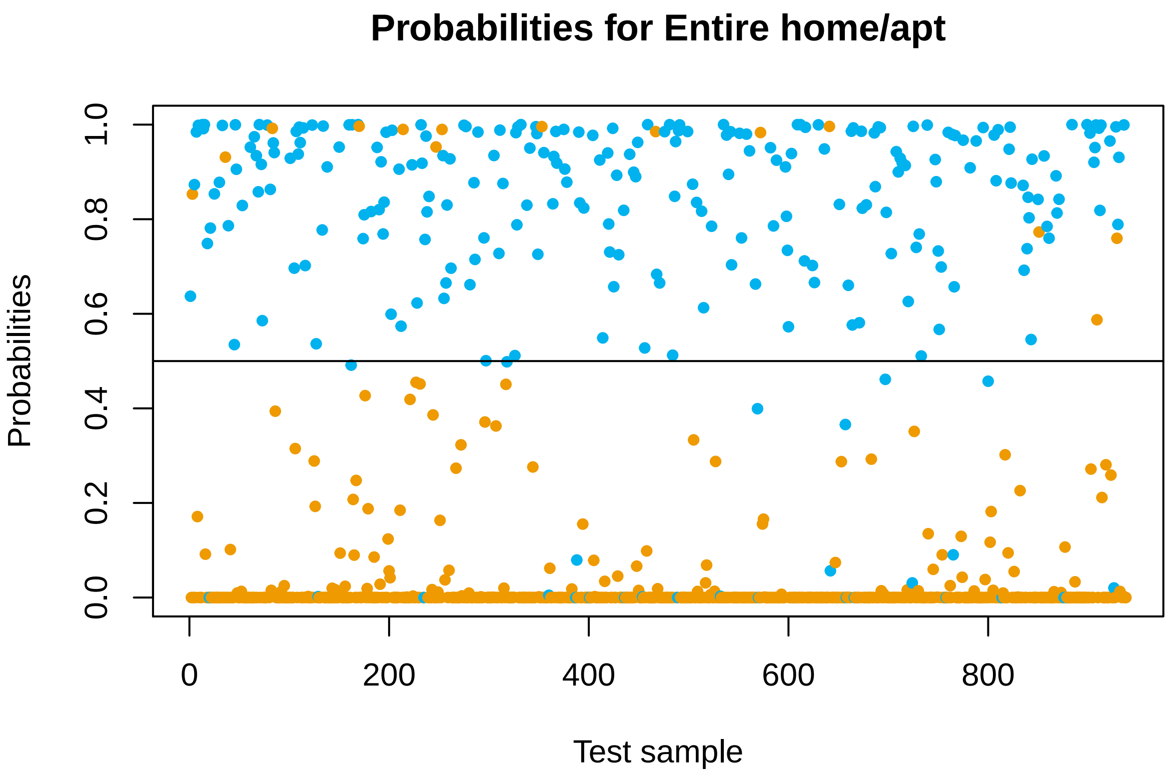
```
## [1] 0.717484
```

```r
# Posterior probabilities of the classifications made with the test sample

prob_nb_Y_test <- nb_test
head(prob_nb_Y_test)
```

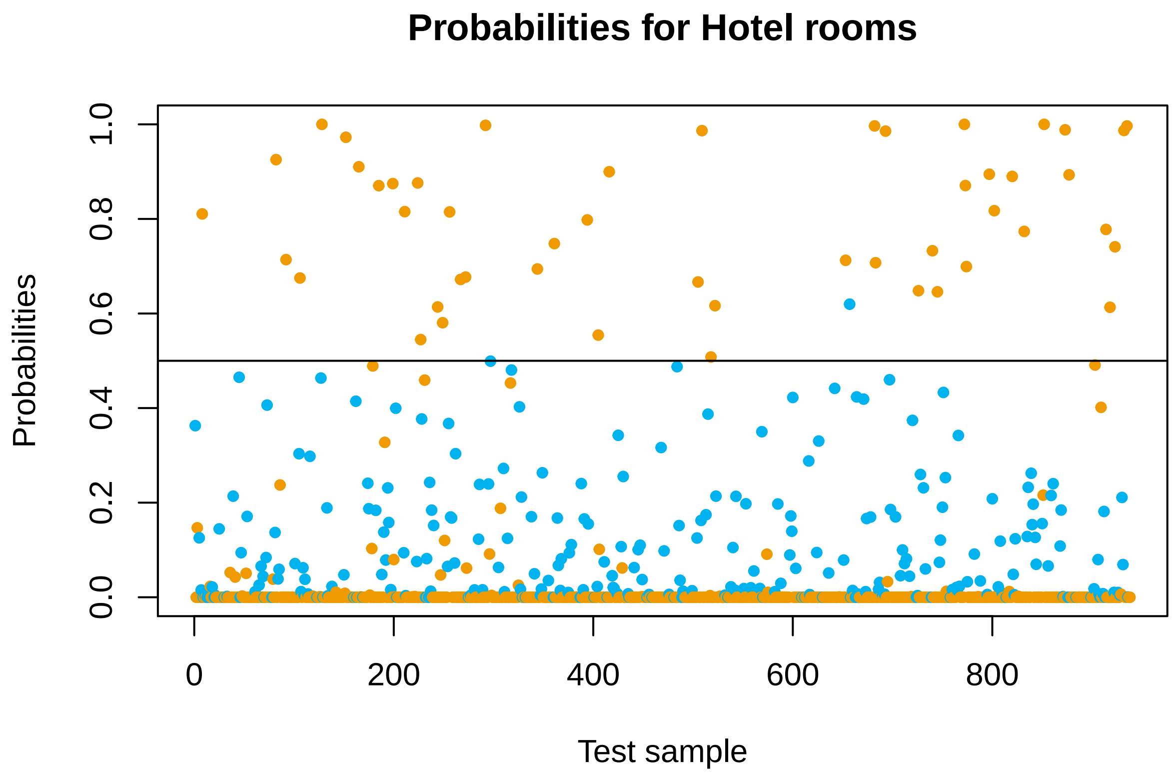
```
##      Entire home/apt   Hotel room Private room  Shared room
## [1,]    6.371445e-01 3.628553e-01 3.190404e-08 1.715908e-07
## [2,]    5.739901e-15 4.085558e-16 1.017387e-12 1.000000e+00
## [3,]    8.531809e-01 1.468180e-01 1.176780e-06 4.810767e-09
## [4,]    1.127921e-12 4.313954e-08 3.839159e-13 1.000000e+00
## [5,]    8.731664e-01 1.256320e-01 1.201633e-03 9.760532e-09
## [6,]    5.886728e-14 2.093702e-14 1.918263e-16 1.000000e+00
```

```r
# Make a plot of the probabilities
```

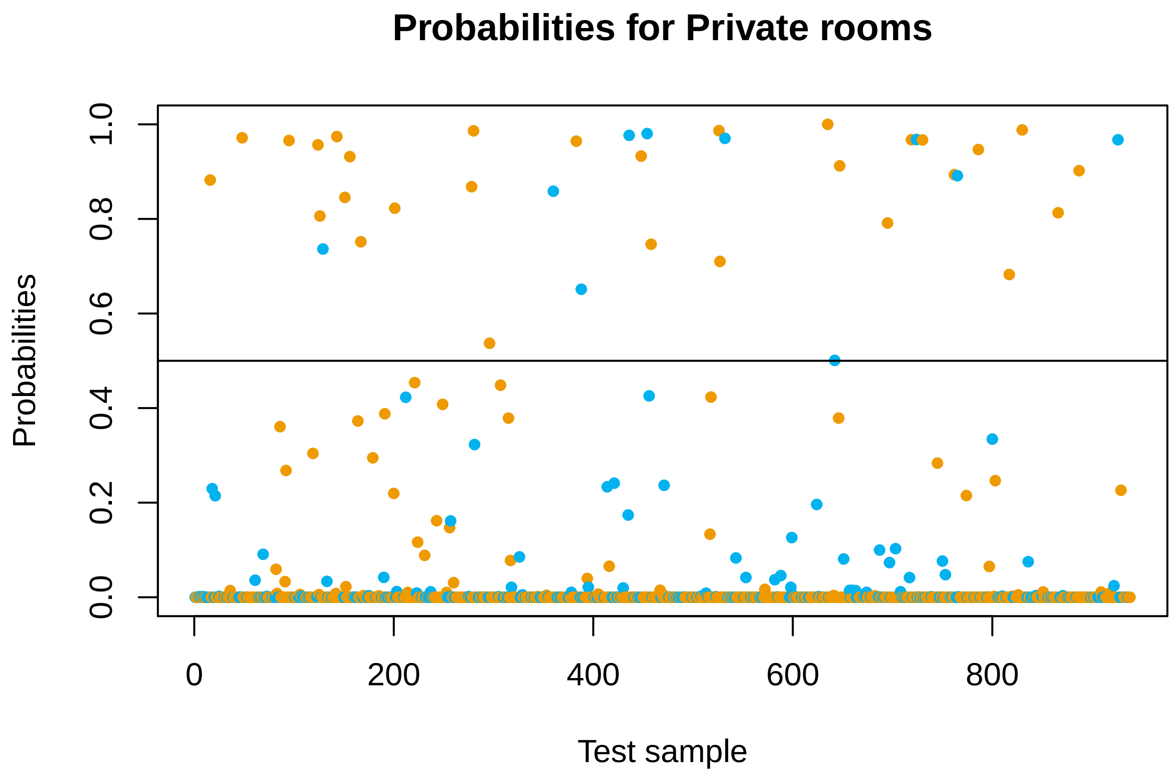
```
# Plot of the probabilities for Entire home/apt
colors_errors <- c(color_3,color_1)[1*(Y_test==nb_Y_test)+1]
plot(1:n_test,prob_nb_Y_test[,1],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabilitie
     main="Probabilities for Entire home/apt")
abline(h=0.5)
```



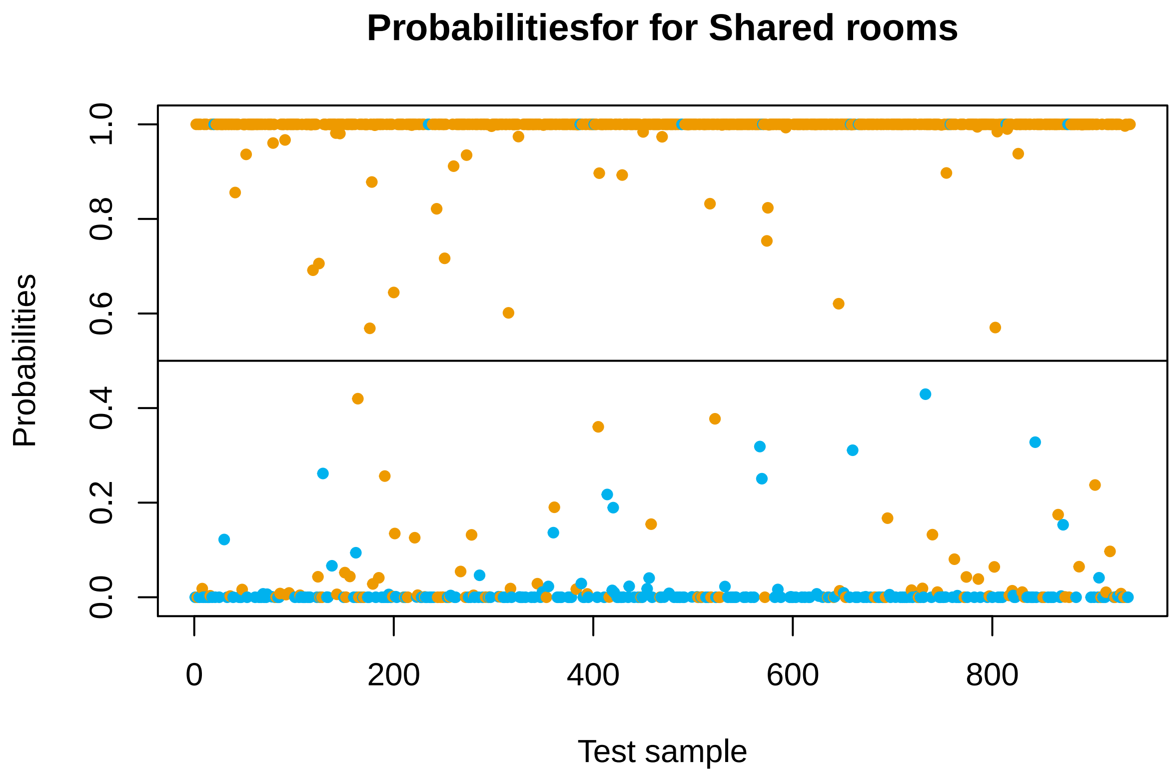

```
# Plot of the probabilities for Hotel rooms
plot(1:n_test,prob_nb_Y_test[,2],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabilitie
     main="Probabilities for Hotel rooms")
abline(h=0.5)
```




```
# Plot of the probabilities for Private rooms
plot(1:n_test,prob_nb_Y_test[,3],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabiliti
     main="Probabilities for Private rooms")
abline(h=0.5)
```




```
#Plot of the probabilities for Shared rooms
plot(1:n_test,prob_nb_Y_test[,4],col=colors_errors,pch=20,type="p",xlab="Test sample",ylab="Probabiliti
     main="Probabilitiesfor for Shared rooms")
abline(h=0.5)
```



Logistic Regression

Linear regression is used to approximate the linear relationship between a continuous response variable and a number of predictor variables. However, when the response variable is categorical linear regression can't be used. However, from a small adaptation to the linear regression method the *logistic regression* method was born. This method is similar to linear regression in many ways, but are used for binary and categorical response variables instead of continuous ones.

Generally the *logistic regression* method is used for binary variables, but can also be used for multinomial problems. Since the response variable in this case is not binary (it can take on 4 different values), a multinomial logistic regression will be performed.

```
train = cbind(X_train, Y_train)
names(train)[25] = "roomtype"
lr_train <- multinom(roomtype ~ .,data=train)
lr_test <- predict(lr_train,newdata=X_test)
```

```
# Number of rooms classified as each type
summary(lr_test)
```

```
## Entire home/apt      Hotel room    Private room     Shared room
##             597               1             336               4
```

```
# Confusion table
table(Y_test,lr_test)
```

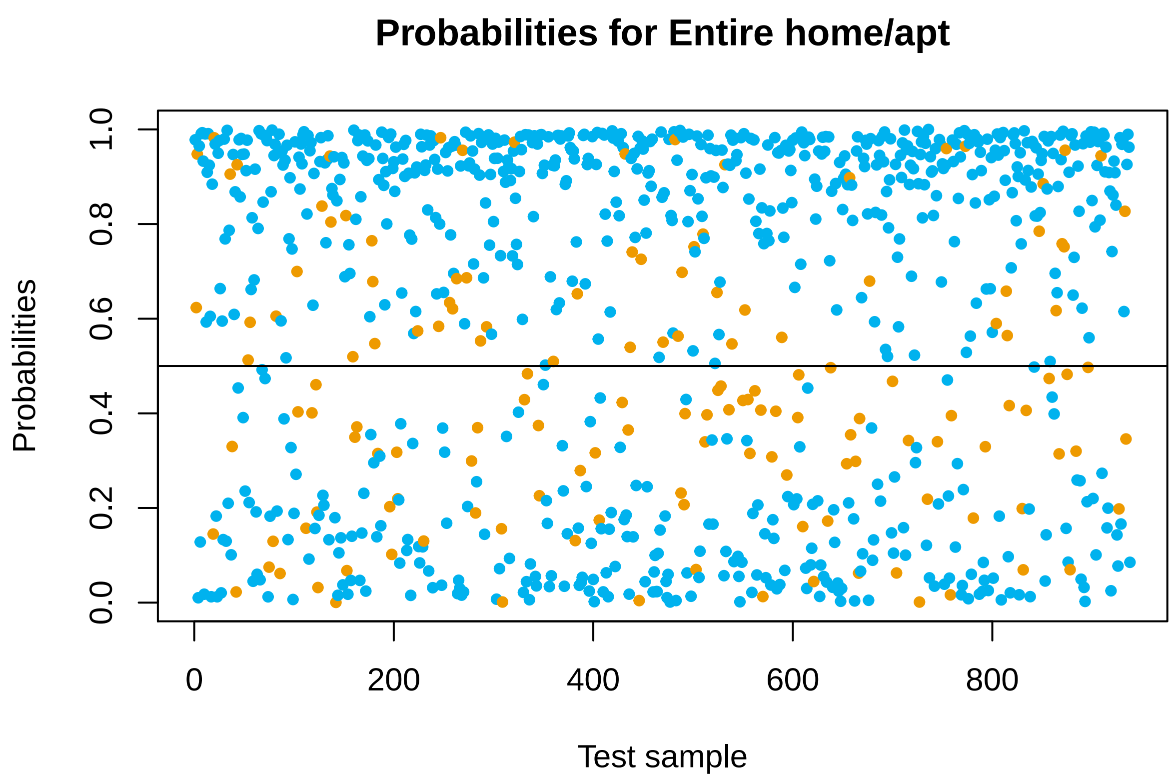
```
##                    lr_test
## Y_test              Entire home/apt Hotel room Private room Shared room
```

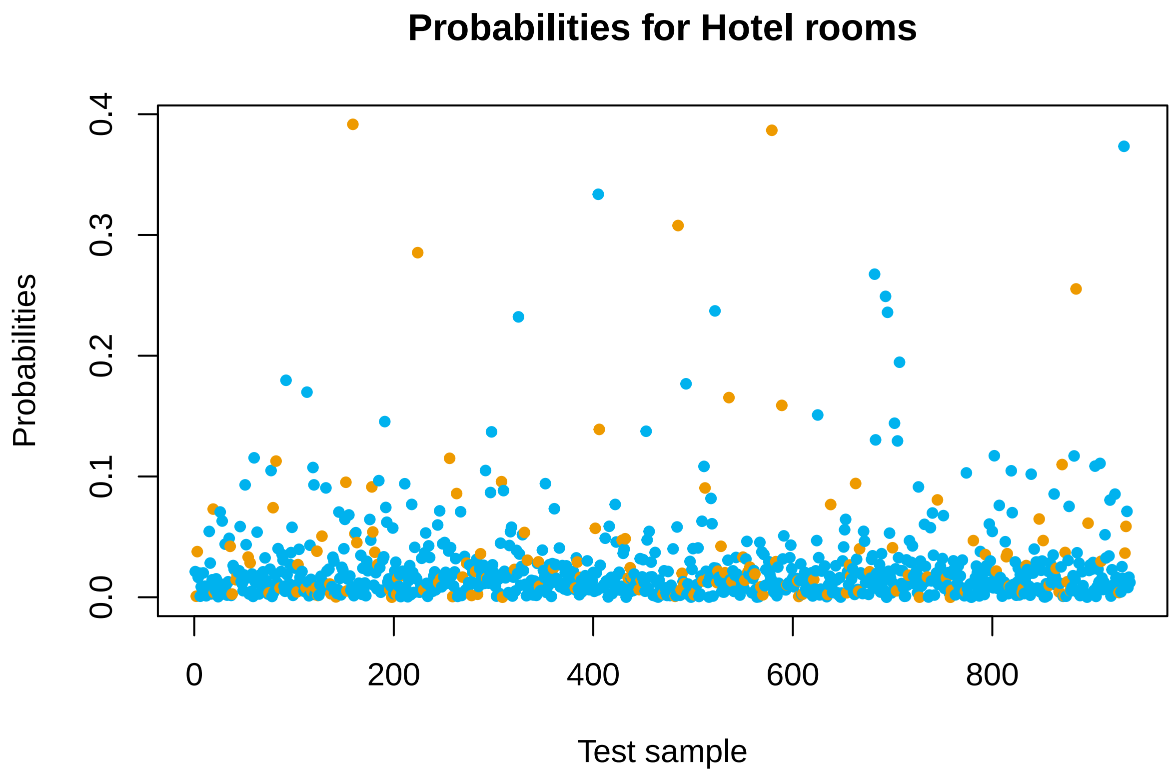
```
##    Entire home/apt              535         0          72          1
##    Hotel room                    16         0           7          0
##    Private room                  43         1         251          0
##    Shared room                    3         0           6          3
```

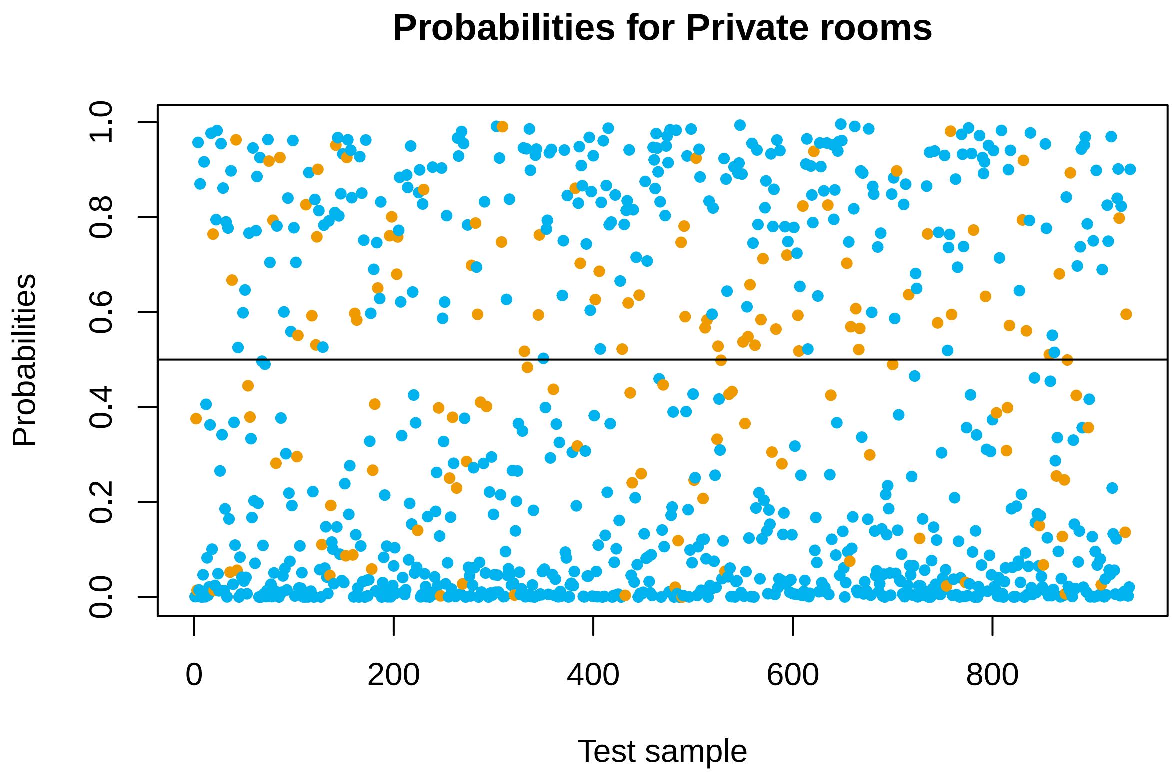
```r
# Obtain the Test Error Rate (TER)
lr_TER <- mean(Y_test!=lr_test)
lr_TER
```

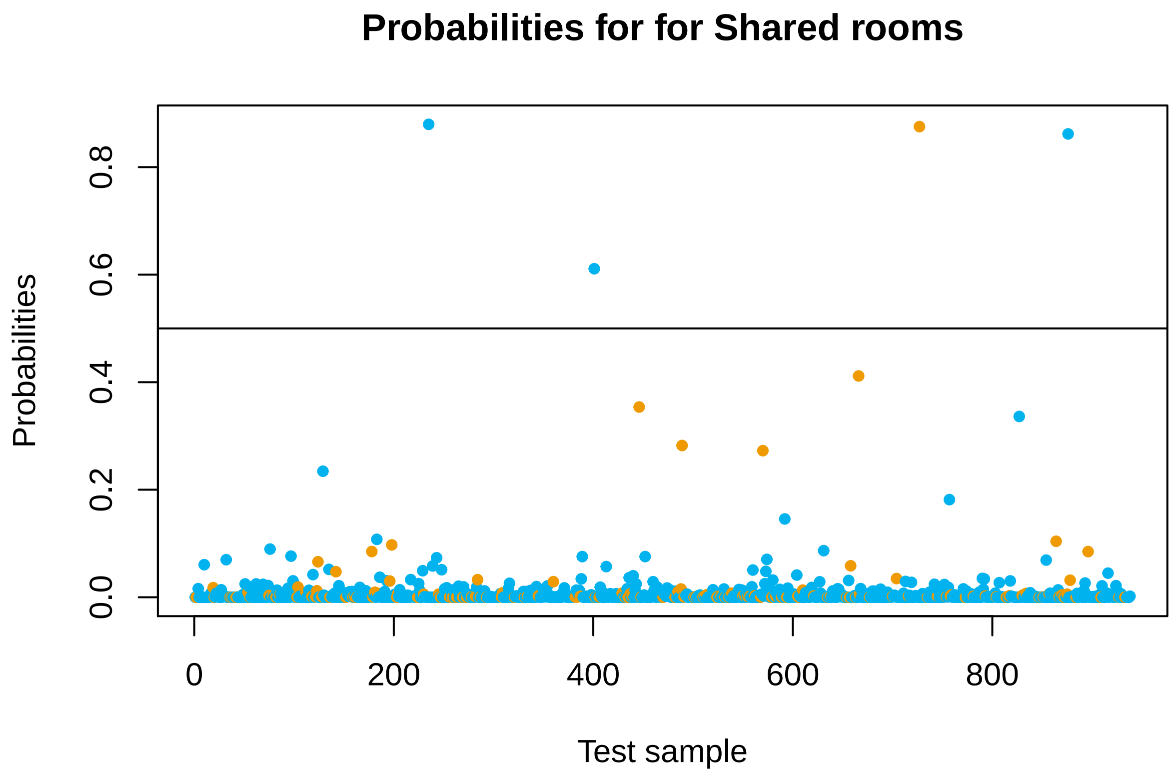
```
## [1] 0.1588486
```

The probabilities for each of the room types is plot below.




Probabilities for Hotel rooms
Probabilities
0.0
0.1
0.2
0.3
0.4
0
200
400
600
800
Test sample


Probabilities for Private rooms
Probabilities
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
Test sample



As we can see above, the method has a prediction error of 0.1588486 which is already not very bad. Let's see if this can be improved by running a backward stepwise algorithm that will select which predictors will be the best at predicting and result in the best predicting score.

```
# Try to improve the test error rate by deleting predictors without discriminatory power
step_lr_train <- step(lr_train,direction="backward",trace=0)
```

```
# Have a look at the variables that have been retained in the model
summary(step_lr_train)$coefnames
```

```
##  [1] "(Intercept)"               "host_total_listings_count"
##  [3] "accommodates"              "bathrooms"
##  [5] "bedrooms"                  "beds"
##  [7] "price"                     "cleaning_fee"
##  [9] "guests_included"           "extra_people"
## [11] "minimum_nights"            "maximum_nights"
## [13] "availability_30"           "number_of_reviews"
## [15] "review_scores_location"    "reviews_per_month"
```

As can be seen above, the algorithm as only selected 16 out of the total of 24 predictors. Let's try to make some predictions with this new model.

```
# Classify the responses in the test data set with this new model
step_lr_test <- predict(step_lr_train,newdata=X_test)

# Number of rooms classified as each type
summary(step_lr_test)
```

```
## Entire home/apt      Hotel room    Private room     Shared room
```

```
##              602                0             331                5
```

```r
# Confusion table
table(Y_test,step_lr_test)
```

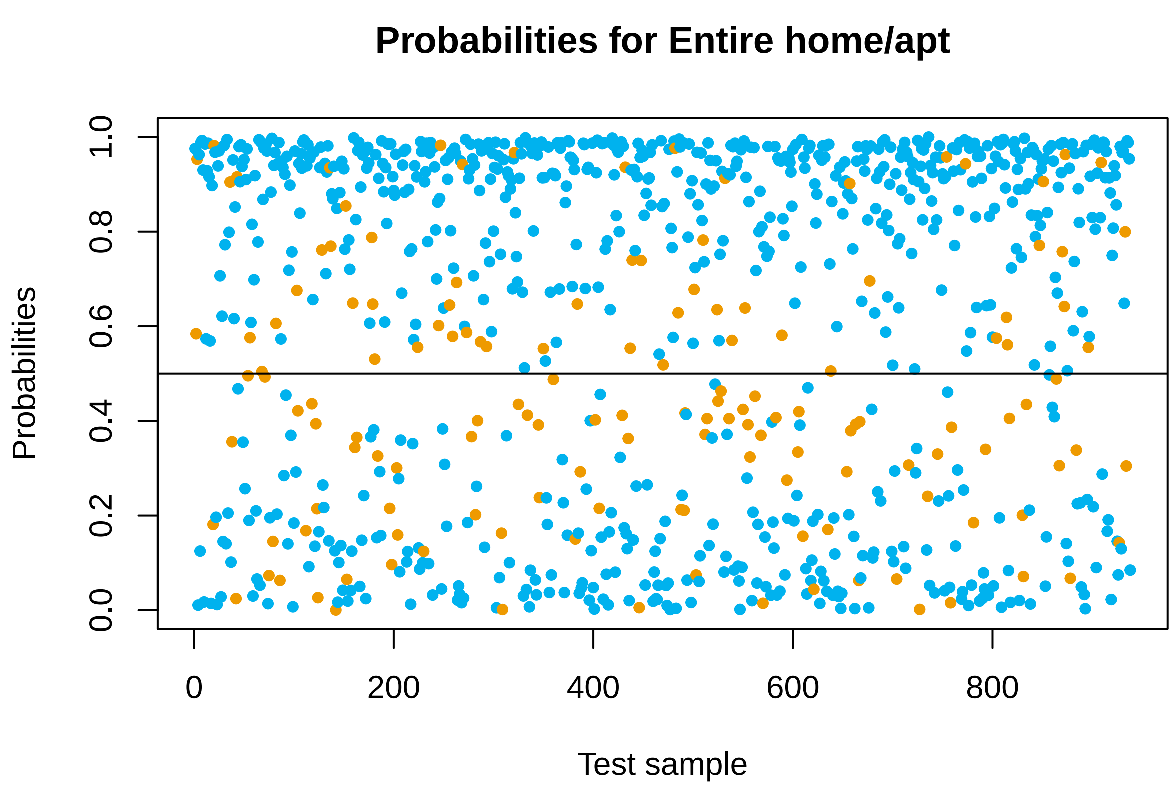
```
##                    step_lr_test
## Y_test              Entire home/apt Hotel room Private room Shared room
##   Entire home/apt               538          0           69           1
##   Hotel room                     16          0            7           0
##   Private room                   46          0          249           0
##   Shared room                     2          0            6           4
```

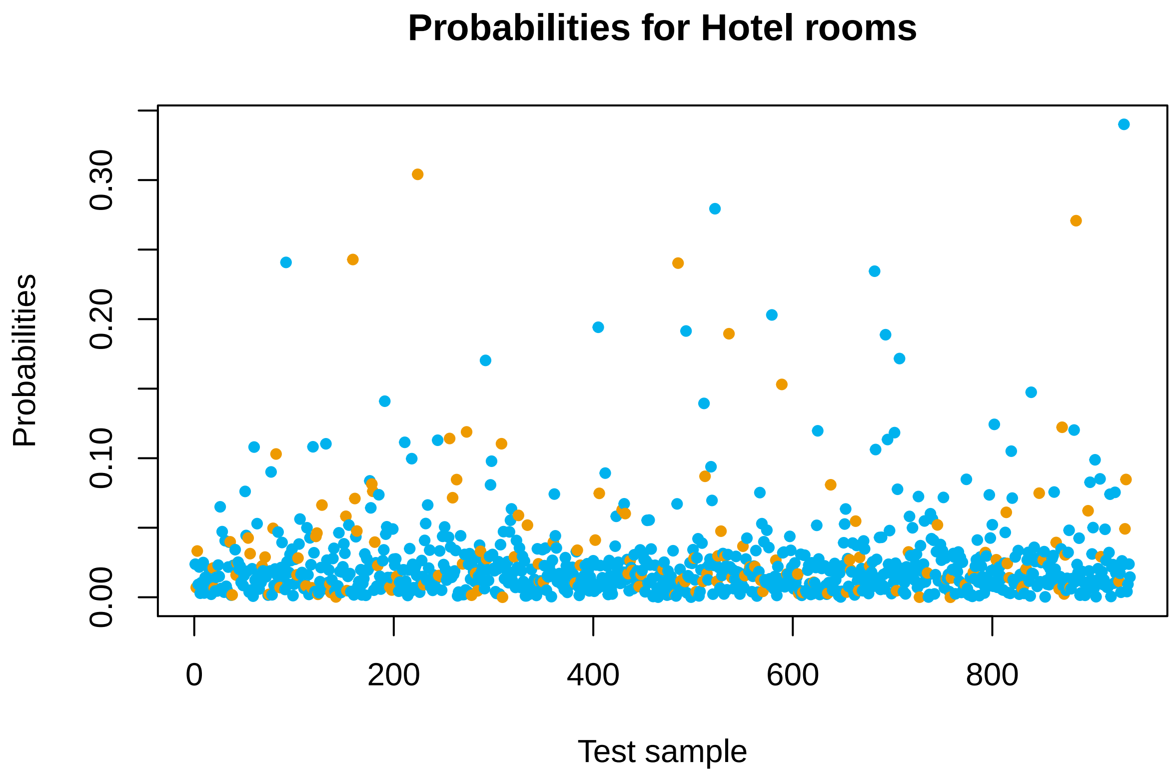
```r
# Obtain the Test Error Rate (TER)
step_lr_TER <- mean(Y_test!=step_lr_test)
step_lr_TER
```

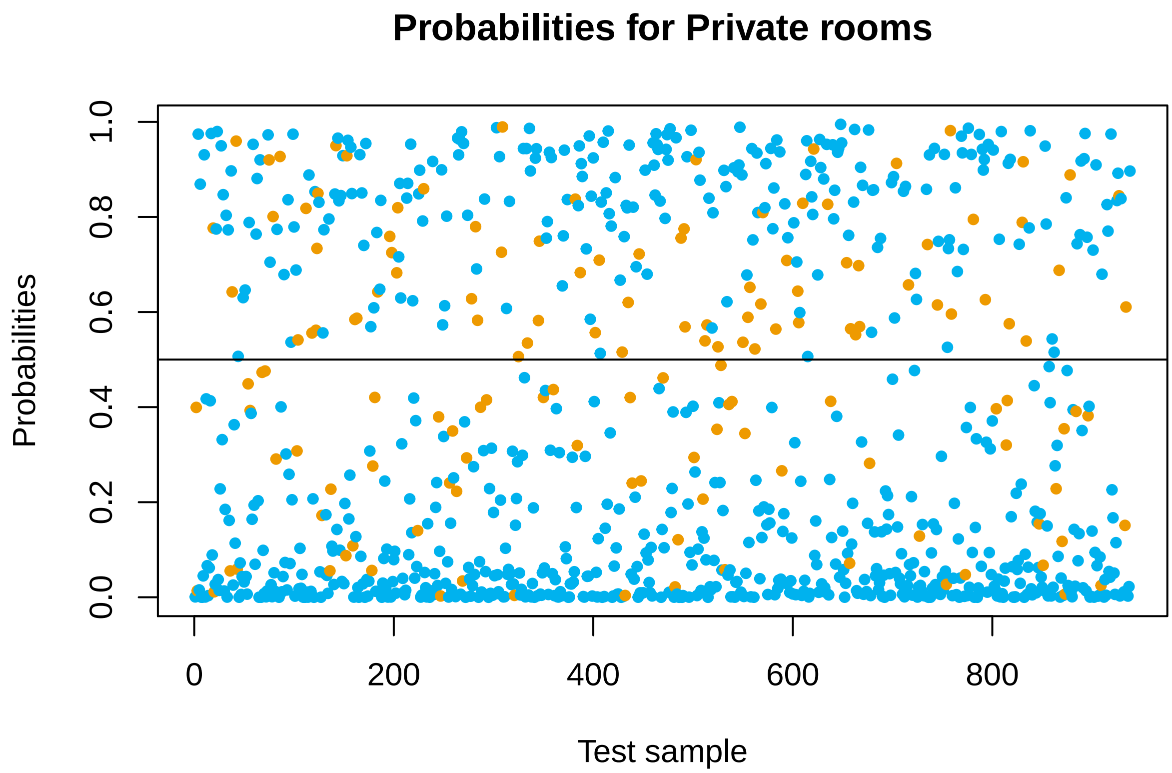
```
## [1] 0.1567164
```

```r
# This model is slightly better than the original one
```

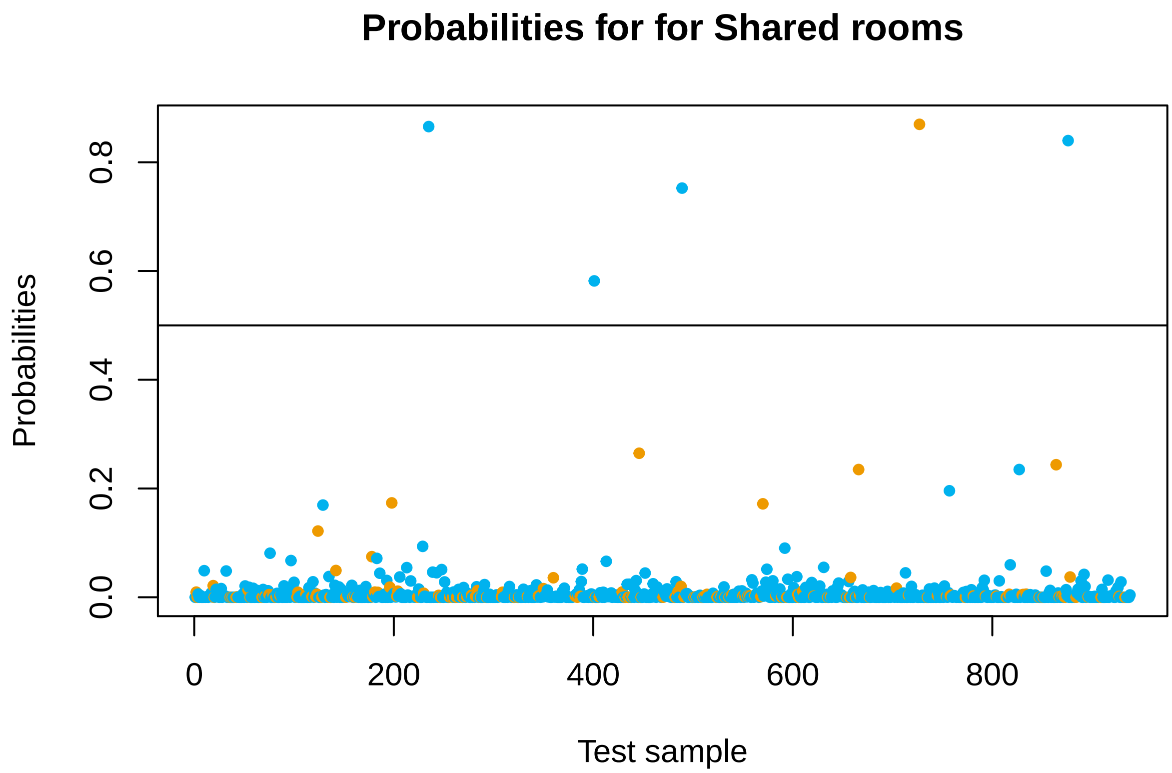
And the probabilities of the room types once again:




Probabilities for Hotel rooms
Probabilities
0.00
0.10
0.20
0.30
0
200
400
600
800
Test sample



Probabilities for Private rooms
Probabilities
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
Test sample



This newer model has an even lower prediction error of 0.1567164.

Conclusion

Important note: This data set is a clear example of imbalanced data set. The number of mistakes for the observations in the small group is large, because the method tends to select the large groups. See the courses on machine learning for more information about solving this problem