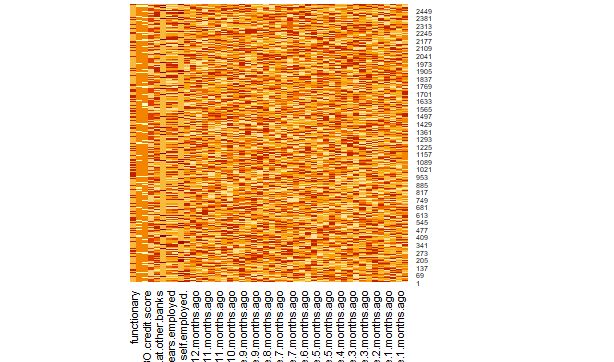
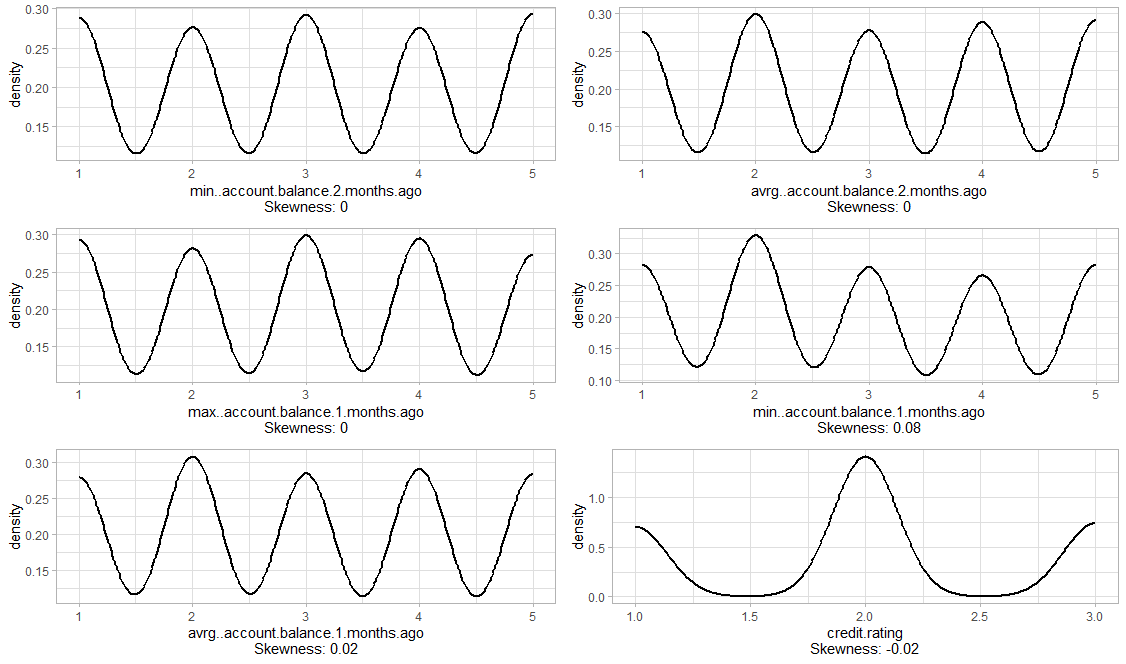
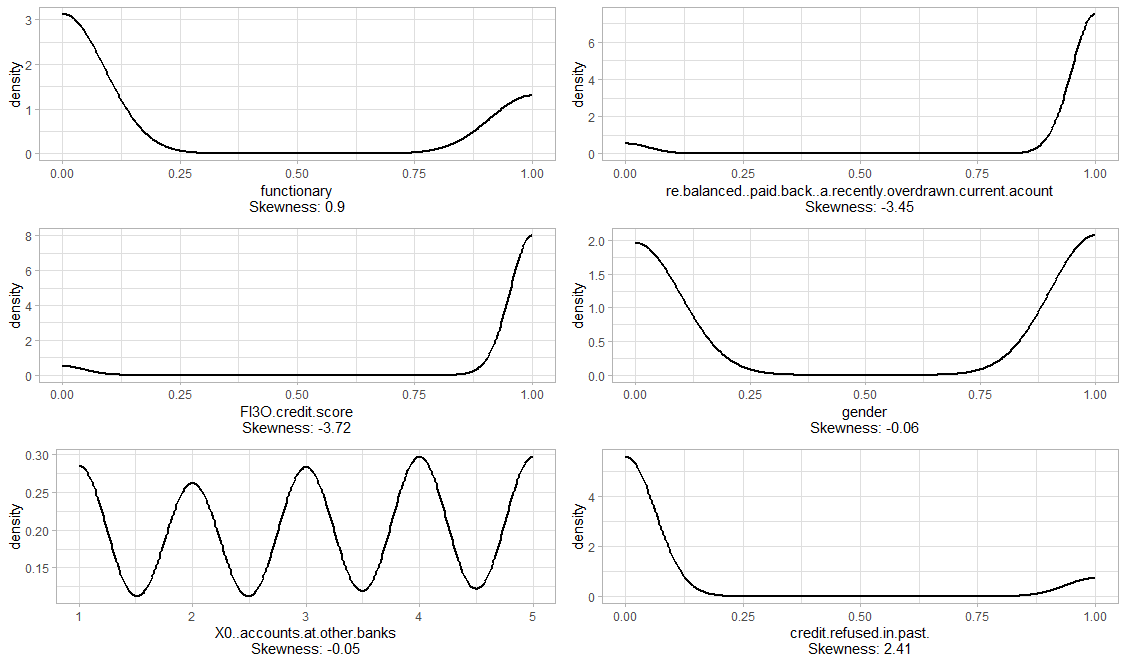
# Question1:

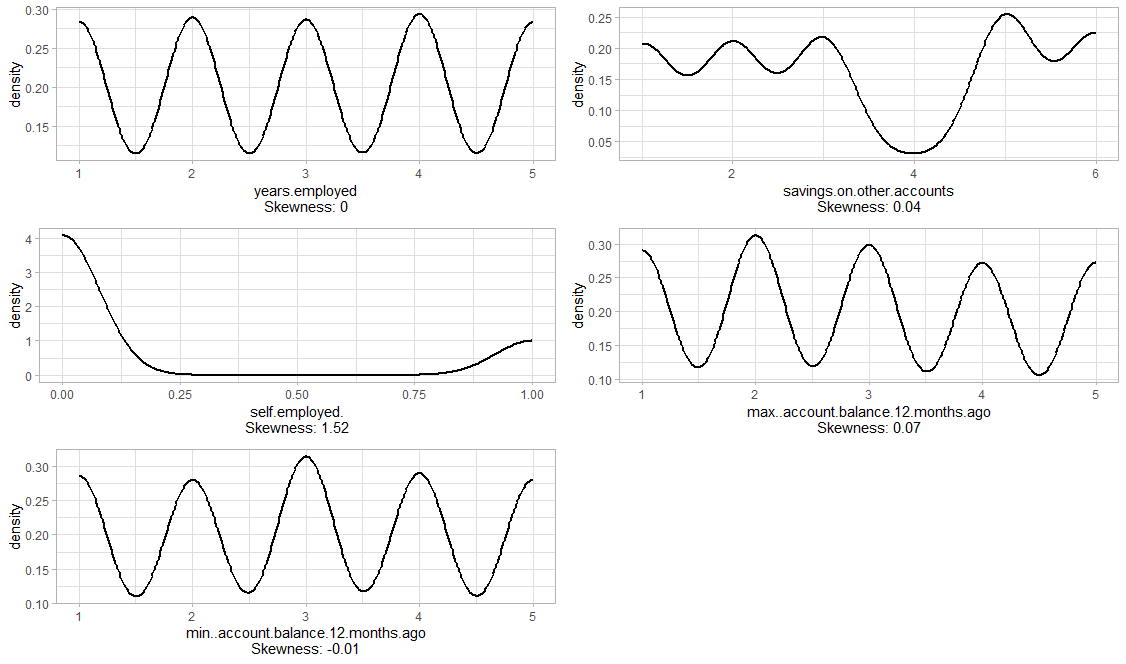
Our Dataset consisted of 2500 obs. of 46 variables and One 47-th Variable is Target Variable showing the rate of the Customer by number 1,2,3 or 0 where two separate test and train dataset generated. The Summary of the data showed that there are no missing values in the Dataset. After the importing of the Data into R, the statistical parameters analyzed. According to Heatmap, it seems that one modelling difficulty arise form highly unbalanced and dimensionality nature of the Dataset. It is also interesting that there is a correlation between some of those variables.

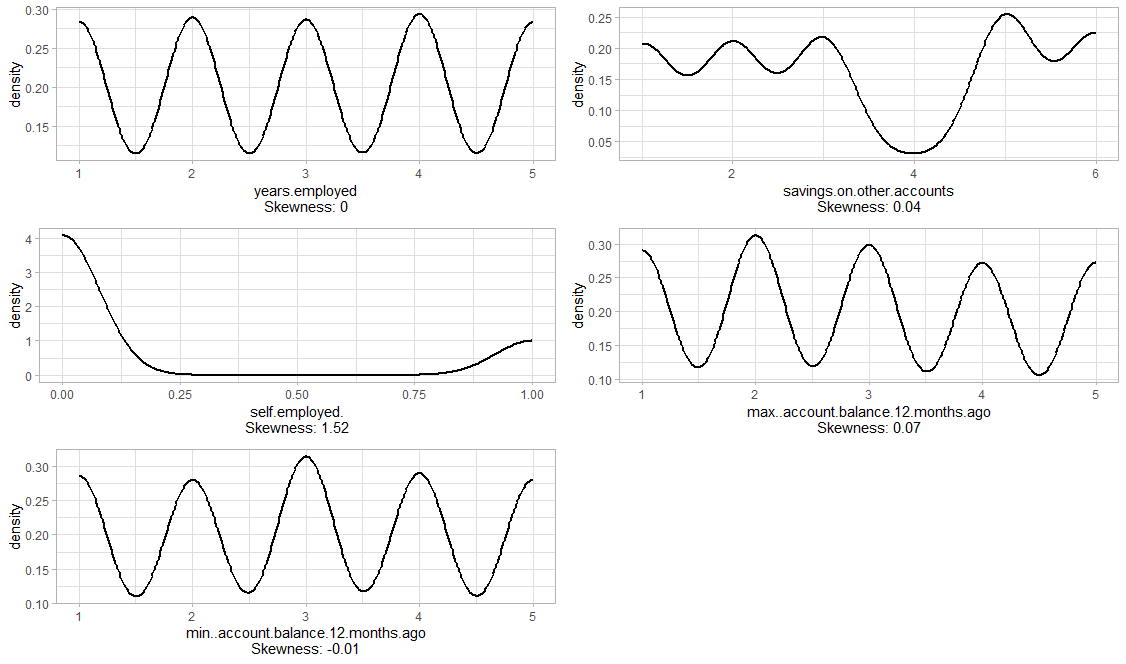


Banks and Credit Institutes are evaluating their Customer according to a Ranking Framework. One shortcoming for the evaluation of the Customers’ Creditworthiness and Financial Strength is Scoring System like FI30 in this case. According to density graph, confirmed score are in majority:

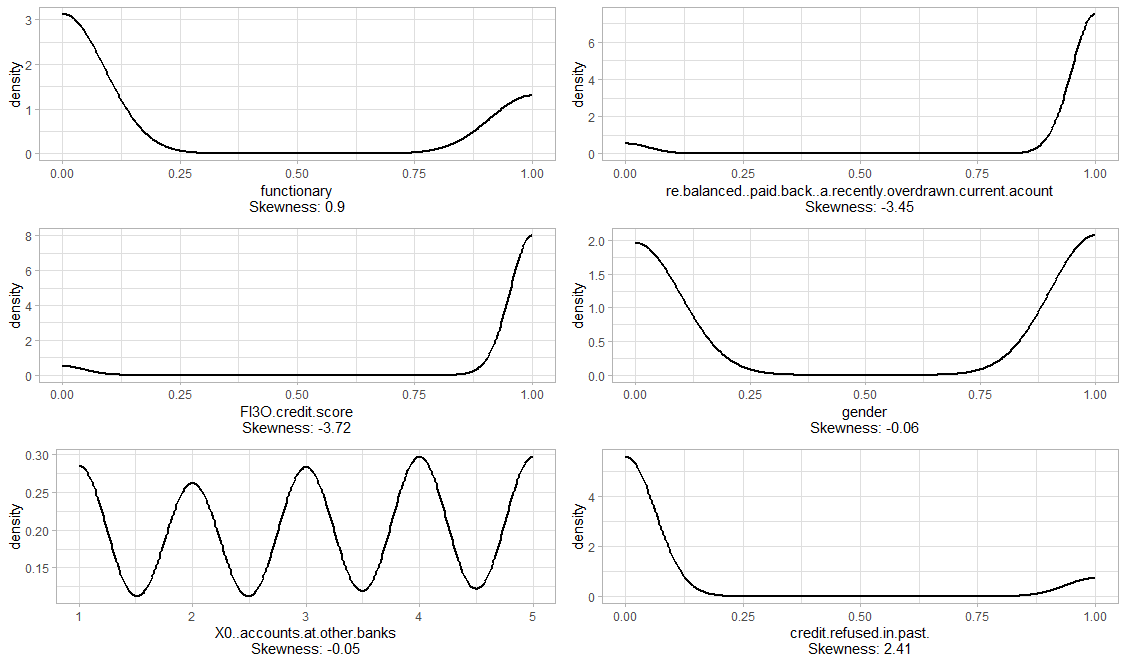


They have also collected the Customer Employment history and saving accounts. Interestingly the self-employed are in higher proportion while their saving accounts are either less than 1000, 2500, 7500 or more than 10000 and 20000 respectively. The years employed also showed no significant difference other than between 5 and 10.

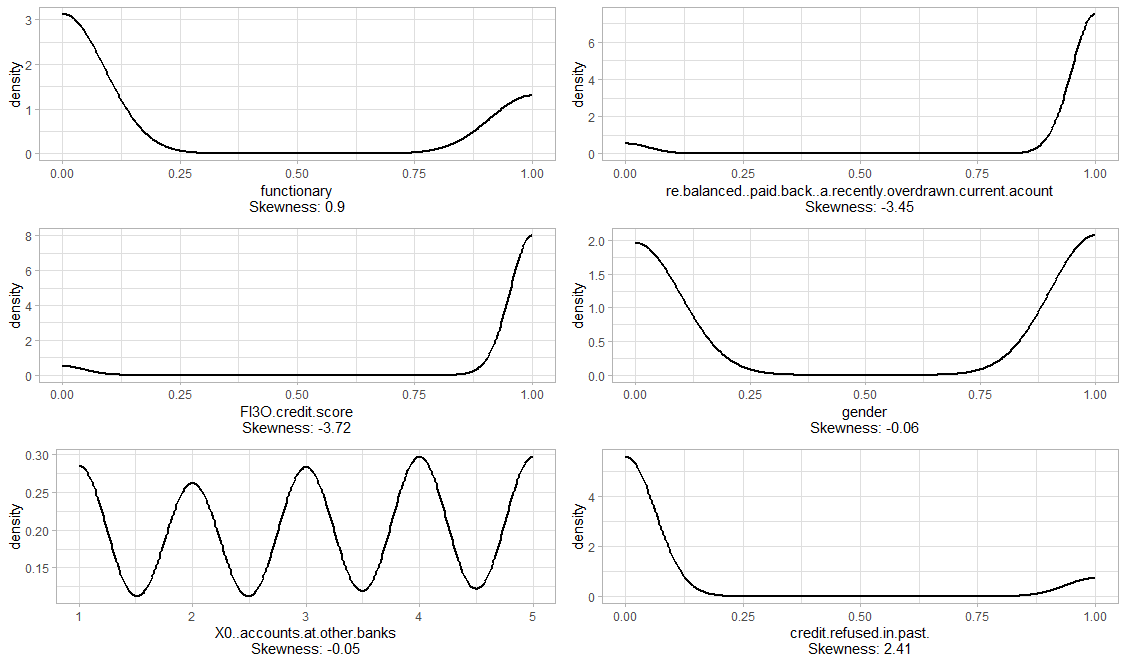
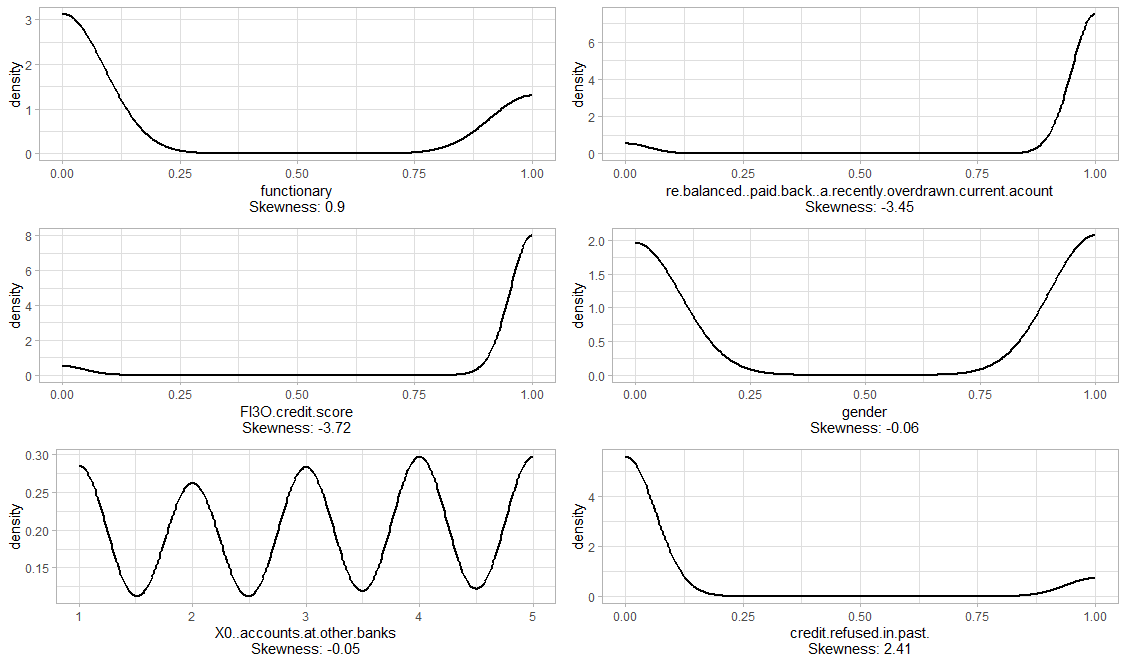




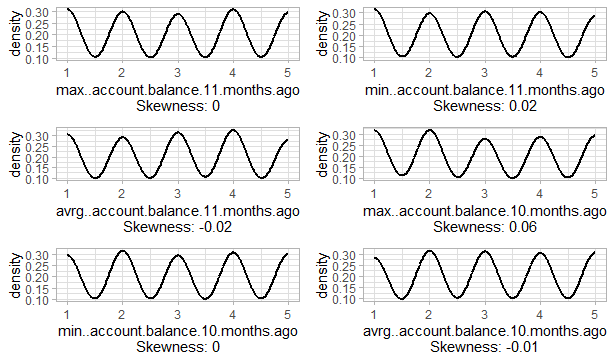
It is clear that for the Customer with Credit Contracts with other Banks, the Credit Amount can be transferred. Therefore, the Banks require the Information related to the accounts that Customer holds in other banks.

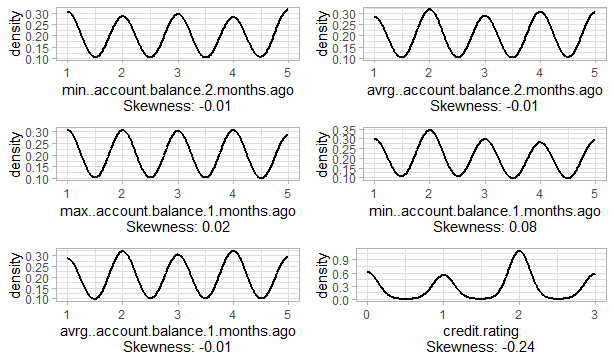


The rebalancing or paid back of a recently overdrawn current account considers as a decisive criteria for whom take a Line of Credit or other kind of Debt such as Schlecks. The Interesting fact here is that the saving accounts can be correlated to this variable since the Customers are rebalancing their current account for a Debt. The refused credit in past are also not in Majority.

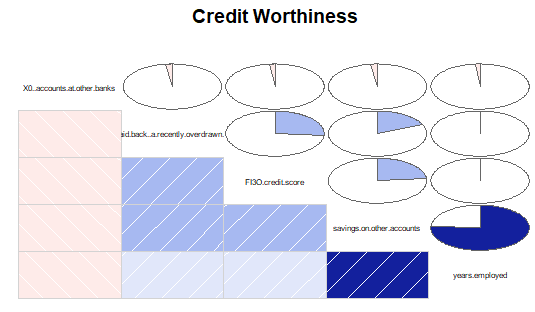


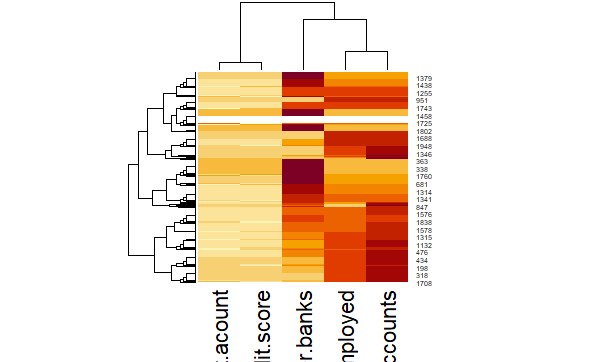
We have further used density plots to show the level of past balance.





The result of the above analysis revealed that accounts on other banks, paid back debt(with consideration of savings on other accounts), Credit Score are interesting variables for the Clustering of the whole 2600 Customers. The years employed can also support the Target Rating in a non-directive manner. We have used a HeatMap and a Dendograph to show this.

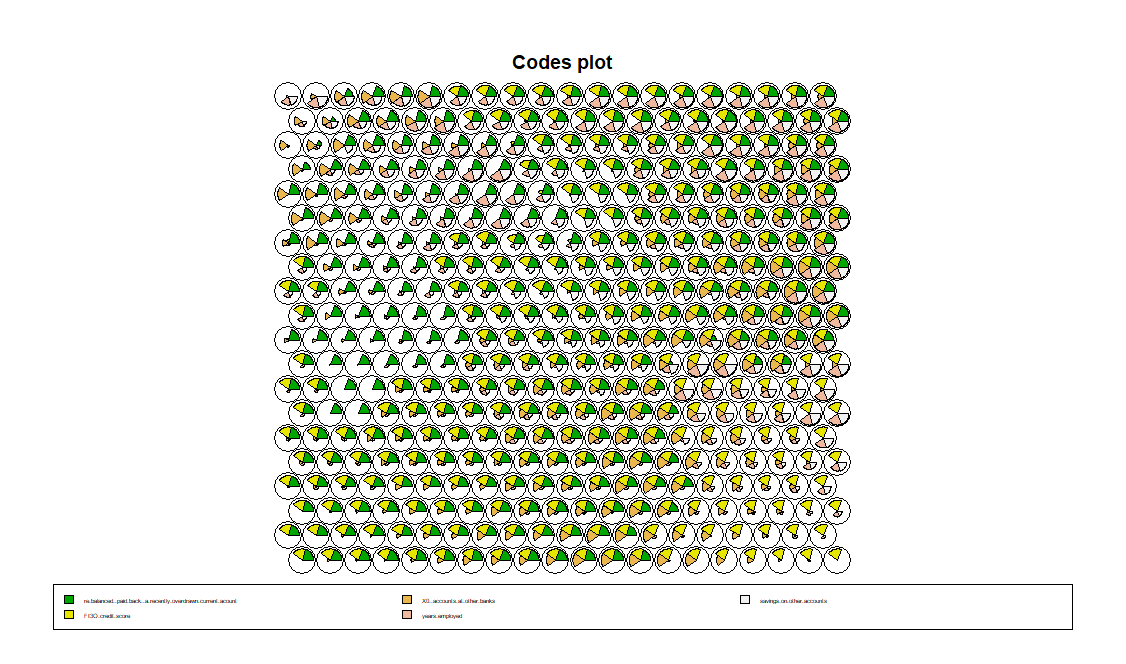
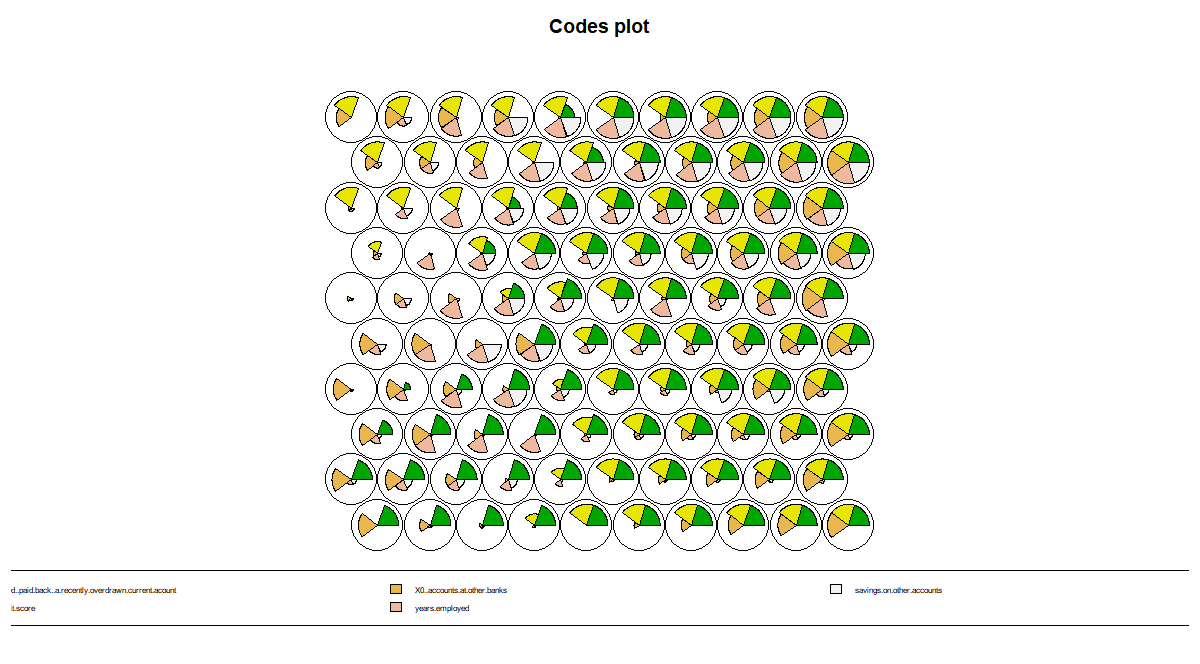
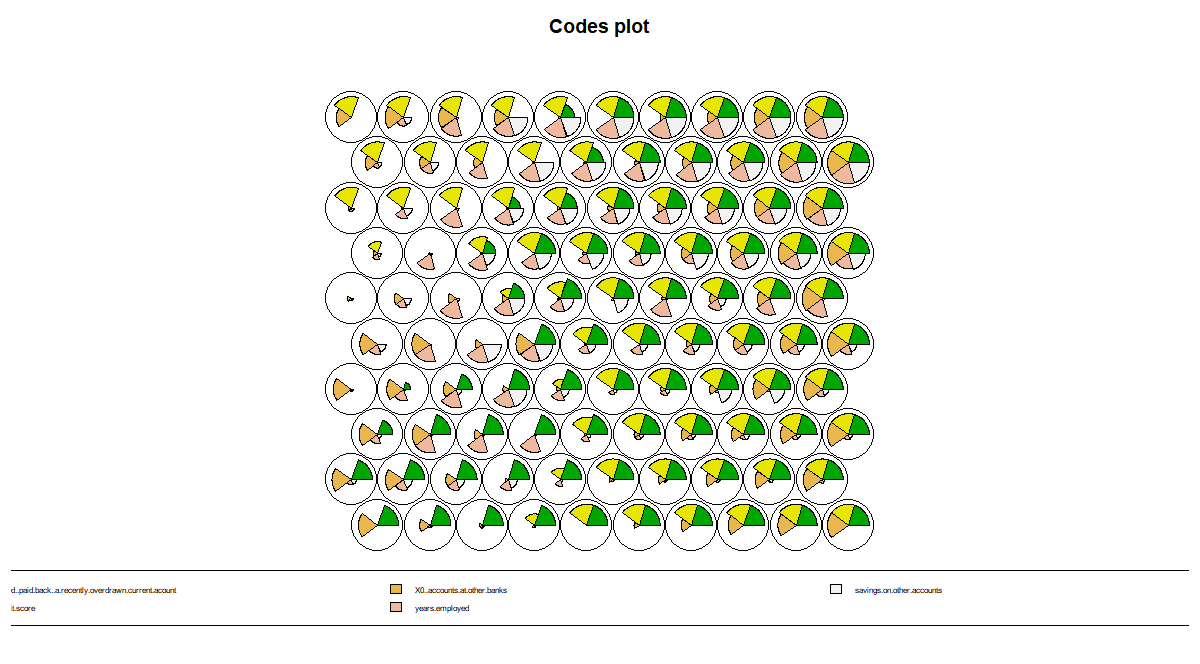
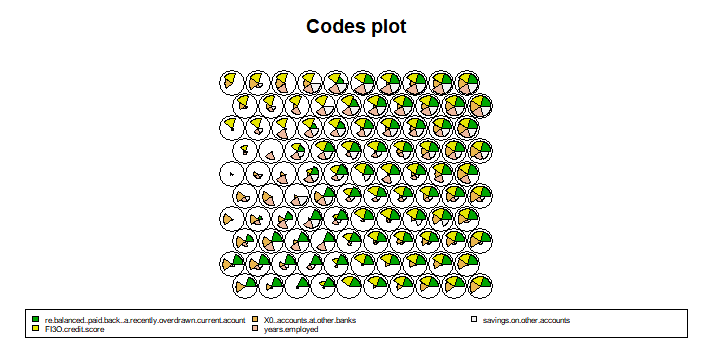




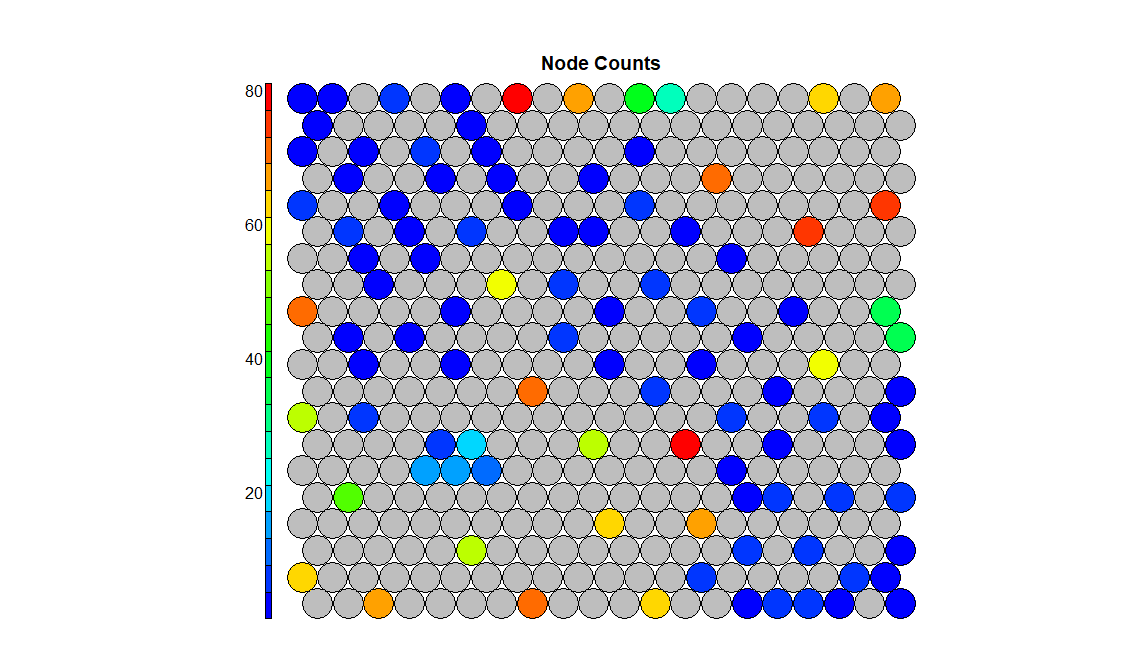
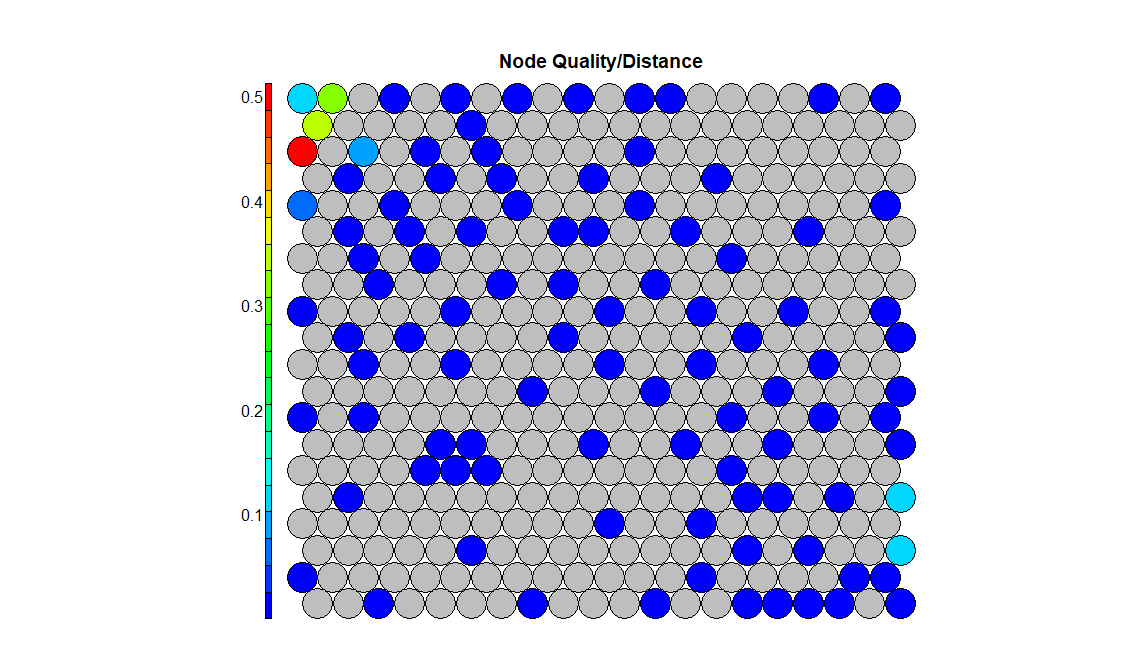
These can help the Creditor to determine the Condition of the Contract like amount, installment Plans or Premium of Credit which would be behold by the bank until the end of the Contract.

We have used the variables 2,3,5,7 and 8 for further evaluation. We used the Matrix of the above variables into a matrix for training of the Self Organizing Map with using of the Kohonen Method:

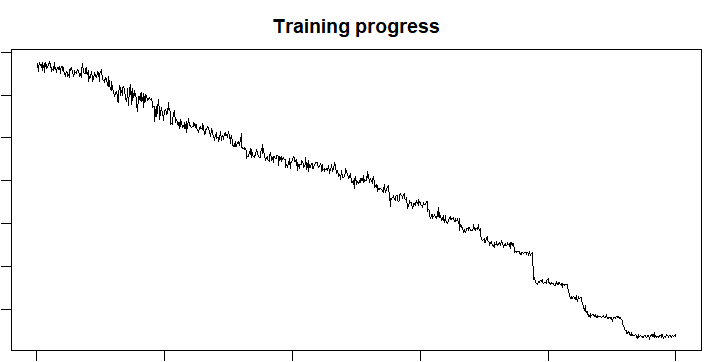
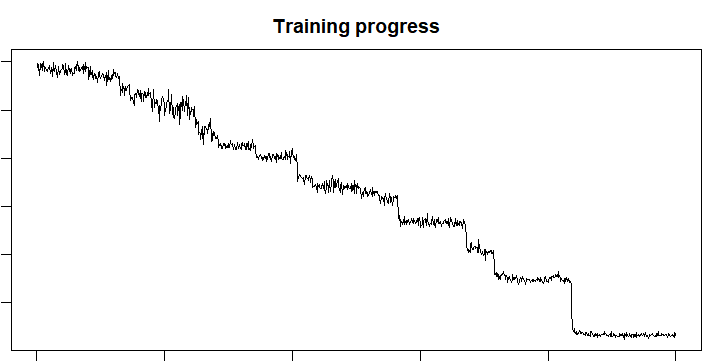
There are two dimensions associated with a Kohonen Self Organizing Map. Each Node is therefore can be assigned with unique X and Y which is called Neuron n. Inside of each Cluster, there are slices where bigger radios slices shows the variables dominate in more proportion of the population of that Neuron.

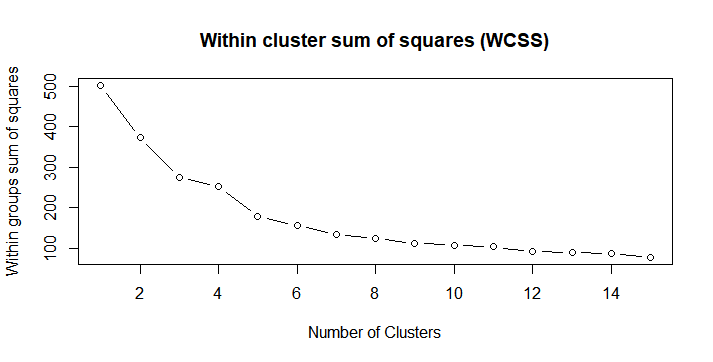


The similar Nodes are mapped near each other like the following Figure.

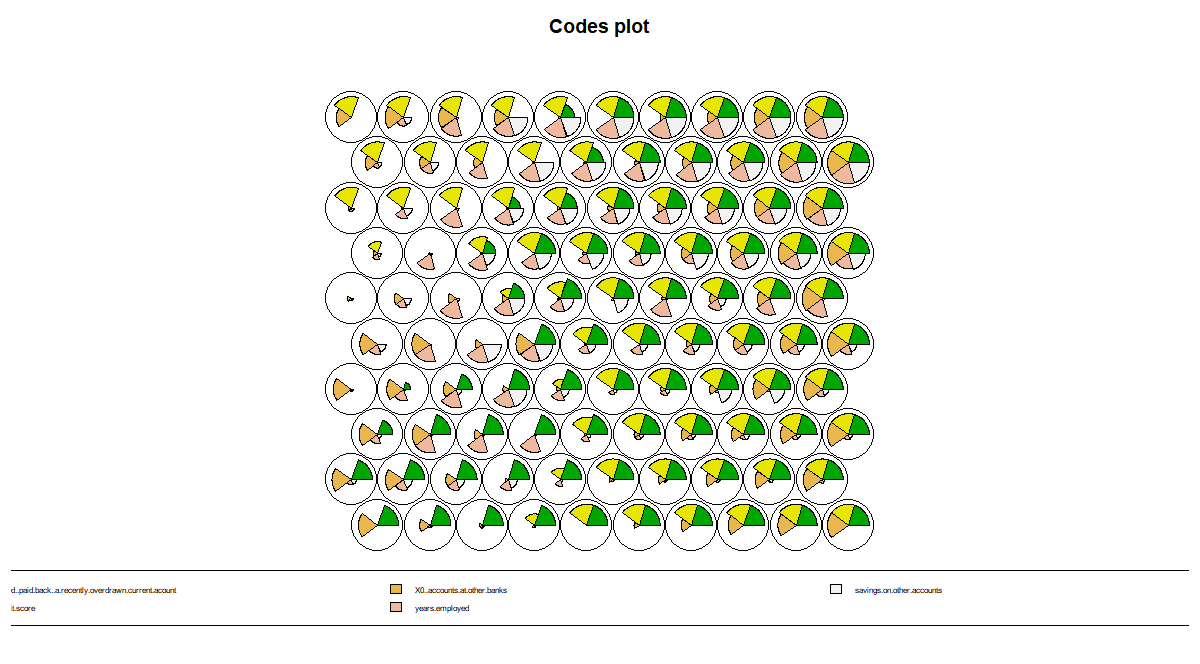
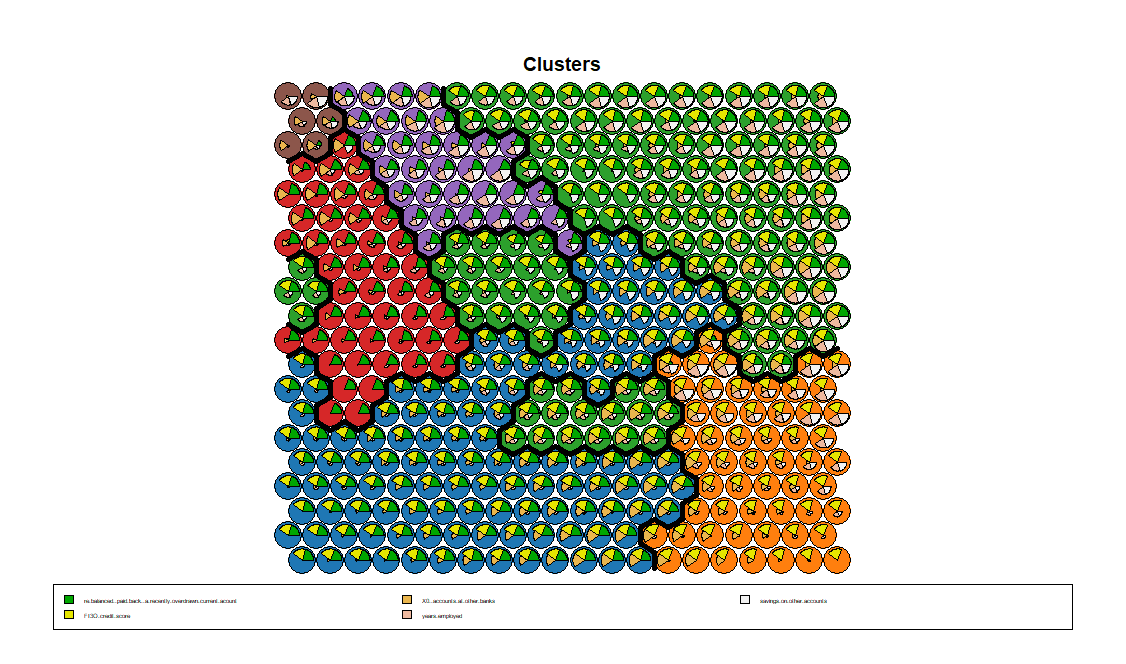


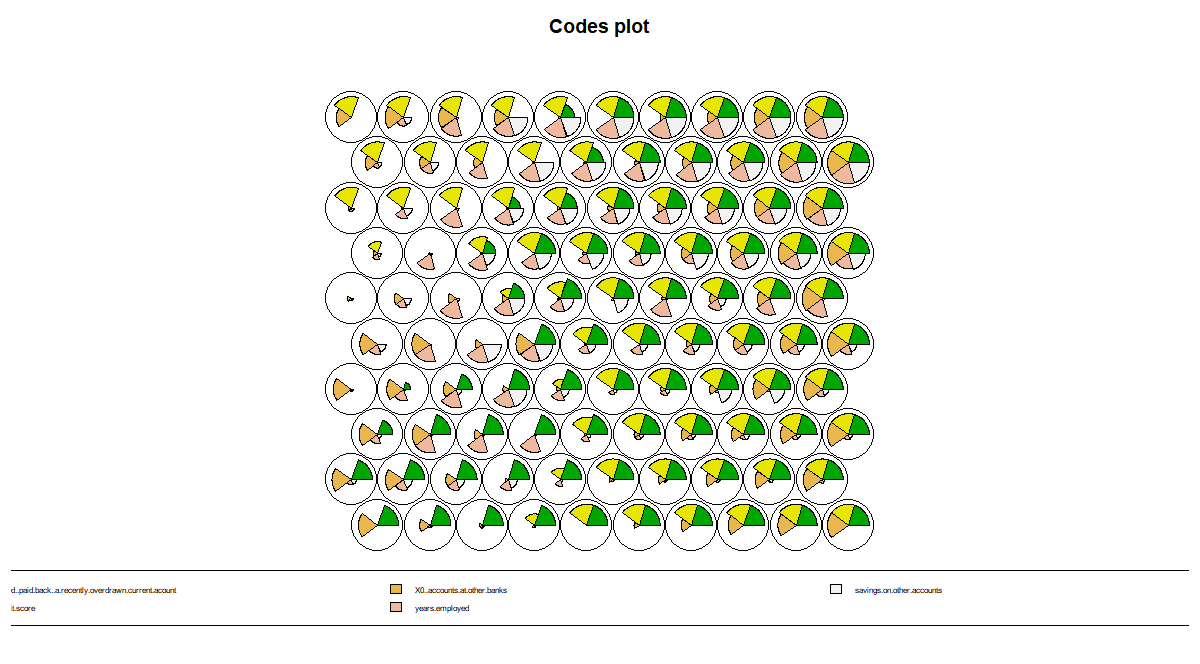
By training Neural Networks of the Kohonen, two other declining parameters are eta and sigma. Eta is the learning rate and sigma is the Neighborhood size. We have chosen the initial and final eta to be 0.2 and 0.01. The iterations showed a different decline pattern so it saves times.

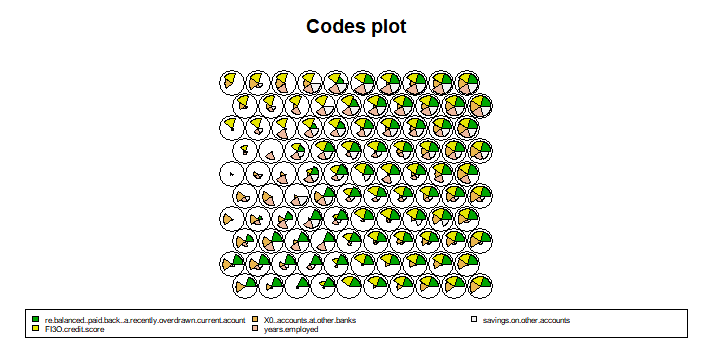


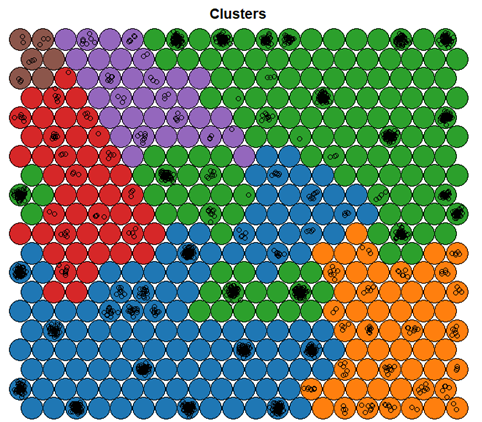


In the Following graph, six Clusters can be recognized for the Xdim. Ydim = 20. In Brown, Purple Orange and Red Partitions the Rebalanced repaid account and credit FI30 are not always associated. Among these three regions. the credit FI30 is in veto condition in Orange and rebalanced is also in veto condition in Brown, Purple and Red. In Green partition, all possibilities are between variables and in Blue accounts at other Banks.





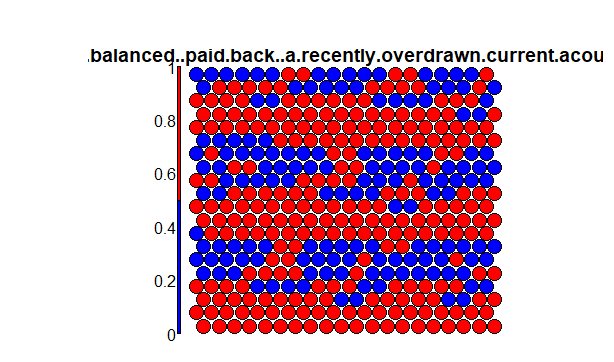
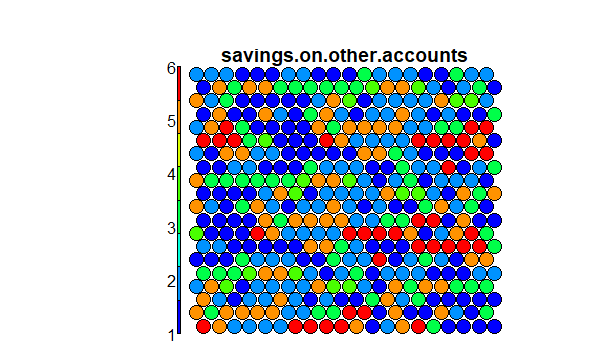
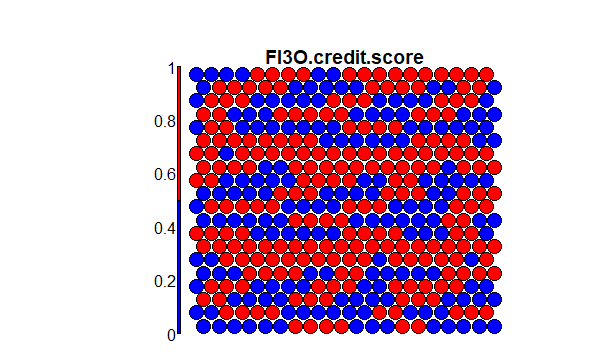
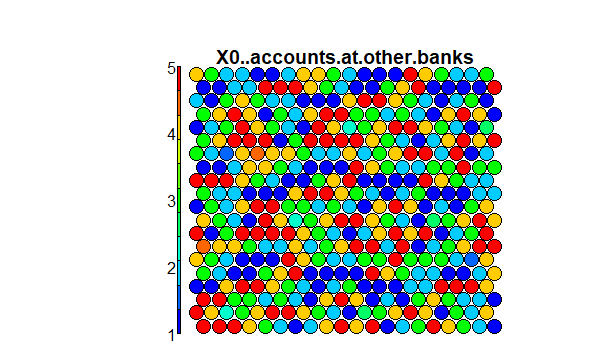
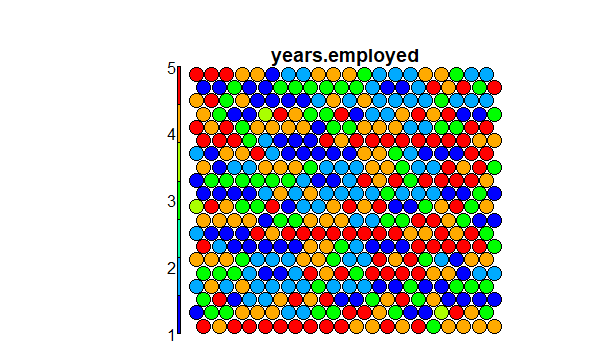




# Question2:

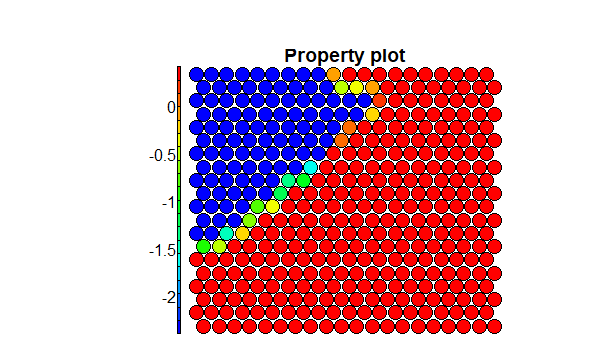
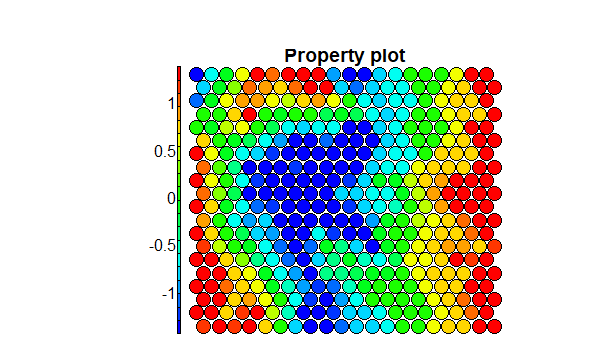
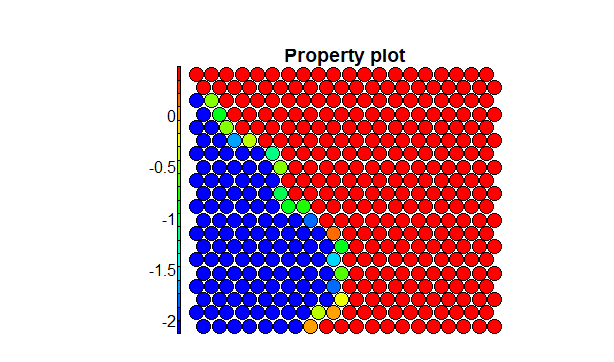
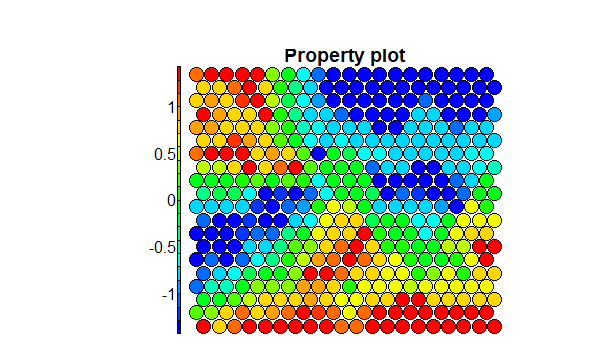
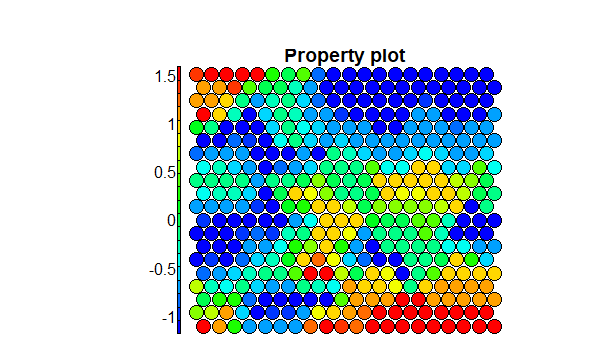
1. For performing of the prediction, the dataset was firstly cleansed. The proportion of the dataset, which their credit are not still assessed removed from the training dataset. We have changed the dataset and remove the zero credit.rating field from it.

From the raw dataset only the variables, which selected from the clustering considered:



*dataset<-data\_raw[c(2,3,5,7,8,46)]*

Their mapping on the normalized state is also shown in the following:



The new training with 1962 records is now defined as training.

*training<-subset(dataset, !is.na(dataset$credit.rating))*

The ratings is now imported from the dataset as the factor

*ratings<-as.factor(dataset[, c(6)])*

supersom is the multiplication of the variables X and Y in SOM below:

*credit.supersom <- supersom(lapply(training,function(x) subset(x, !0)), somgrid(20, 20, "hexagonal"))*

*credit.som.prediction <-predict(credit.supersom, newdata = as.list(training, function(x) subset(training, is.na(dataset$credit.rating))))*

*table(training$credit.rating, credit.som.prediction$prediction$credit.rating)*

The prediction result showed 83 false predicted values with the precision of 95.8%.



1. We have defined a neural network. This resulted in different weighting which was not necessarily

*neural\_net = neuralnet(training$credit.rating ~ re.balanced..paid.back..a.recently.overdrawn.current.acount*

*+ FI3O.credit.score+X0..accounts.at.other.banks+savings.on.other.accounts+years.employed, data = training, hidden = 1, linear.output = TRUE*





By training of different nets, we see that four layers of red, orange, purple and brown from SOM are mutually affect the weight of hidden linear network. The hidden layer more than one caused the computing to be divergence.

All is said and done; best strategy is to reduce the variables to reduce the variable saving on other accounts:

The same procedure defined the testing or new observations:

*testing<-subset(dataset, is.na(dataset$credit.rating))*

The test partition result of the remaining 538 records can be seen from

*credit.som.test.prediction <-predict(credit.supersom, newdata = as.list(test, function(x) subset(testing, !is.na(dataset$credit.rating))))*



By Changing of the number of X and Y dimension we have gained other result as well 98.4%

