Workshop

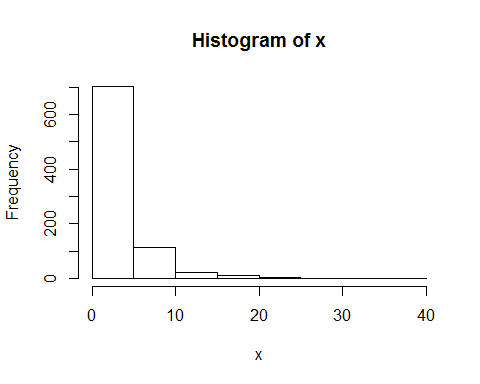
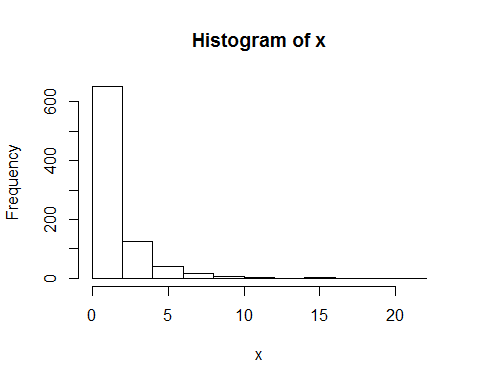
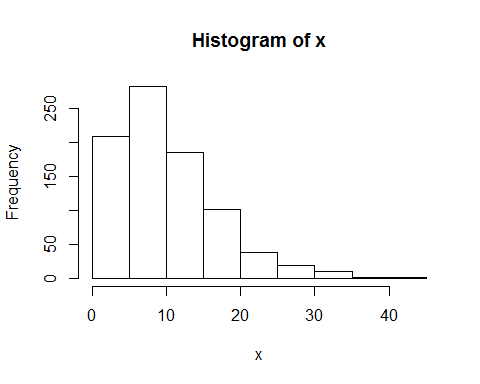
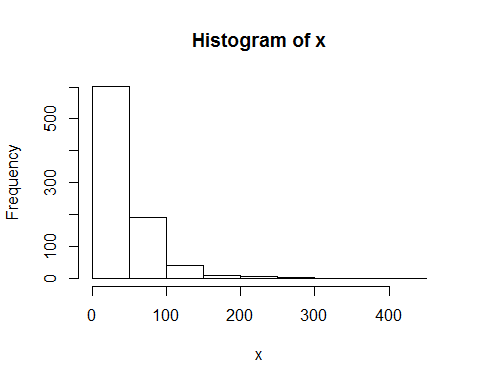
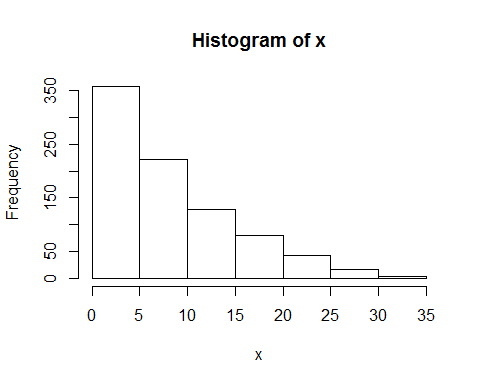
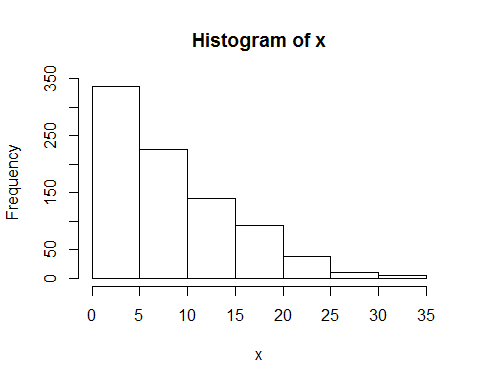
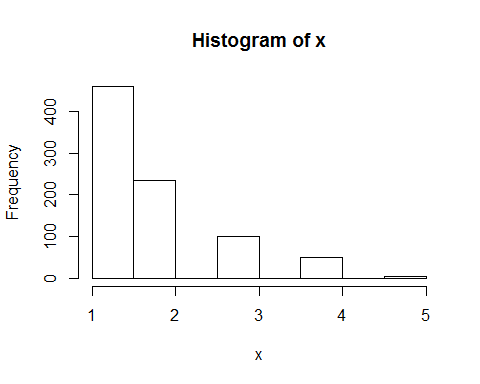
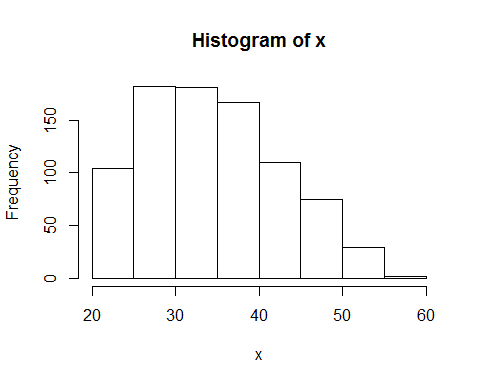
library(readxl)  
BankBayesLoan <- read\_excel("D:/Desktop/4th term/Workshop/BankBayesLoan.xlsx")  
View(BankBayesLoan)  
data<-BankBayesLoan  
#data exploration  
summary(data)

## age ed employ address   
## Min. :20.00 Min. :1.000 Min. : 0.000 Min. : 0.000   
## 1st Qu.:29.00 1st Qu.:1.000 1st Qu.: 3.000 1st Qu.: 3.000   
## Median :34.00 Median :1.000 Median : 7.000 Median : 7.000   
## Mean :35.03 Mean :1.711 Mean : 8.566 Mean : 8.372   
## 3rd Qu.:41.00 3rd Qu.:2.000 3rd Qu.:13.000 3rd Qu.:12.000   
## Max. :56.00 Max. :5.000 Max. :33.000 Max. :34.000   
##   
## income debtinc creddebt otherdebt   
## Min. : 13.00 Min. : 0.10 Min. : 0.0117 Min. : 0.04558   
## 1st Qu.: 24.00 1st Qu.: 5.10 1st Qu.: 0.3822 1st Qu.: 1.04594   
## Median : 35.00 Median : 8.70 Median : 0.8851 Median : 2.00324   
## Mean : 46.68 Mean :10.17 Mean : 1.5768 Mean : 3.07879   
## 3rd Qu.: 55.75 3rd Qu.:13.80 3rd Qu.: 1.8984 3rd Qu.: 3.90300   
## Max. :446.00 Max. :41.30 Max. :20.5613 Max. :35.19750   
##   
## default   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.2614   
## 3rd Qu.:1.0000   
## Max. :1.0000   
## NA's :150

lapply(data, function(x) shapiro.test(x))

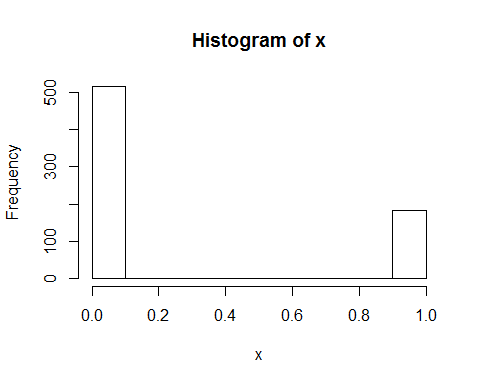
## $age  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.97566, p-value = 1.043e-10  
##   
##   
## $ed  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.74925, p-value < 2.2e-16  
##   
##   
## $employ  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.93136, p-value < 2.2e-16  
##   
##   
## $address  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.91629, p-value < 2.2e-16  
##   
##   
## $income  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.67239, p-value < 2.2e-16  
##   
##   
## $debtinc  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.92262, p-value < 2.2e-16  
##   
##   
## $creddebt  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.63099, p-value < 2.2e-16  
##   
##   
## $otherdebt  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.70257, p-value < 2.2e-16  
##   
##   
## $default  
##   
## Shapiro-Wilk normality test  
##   
## data: x  
## W = 0.54789, p-value < 2.2e-16

lapply(data, function(x) hist(x))



## $age  
## $breaks  
## [1] 20 25 30 35 40 45 50 55 60  
##   
## $counts  
## [1] 104 182 181 167 110 75 29 2  
##   
## $density  
## [1] 0.0244705882 0.0428235294 0.0425882353 0.0392941176 0.0258823529  
## [6] 0.0176470588 0.0068235294 0.0004705882  
##   
## $mids  
## [1] 22.5 27.5 32.5 37.5 42.5 47.5 52.5 57.5  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $ed  
## $breaks  
## [1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0  
##   
## $counts  
## [1] 460 235 0 101 0 49 0 5  
##   
## $density  
## [1] 1.08235294 0.55294118 0.00000000 0.23764706 0.00000000 0.11529412  
## [7] 0.00000000 0.01176471  
##   
## $mids  
## [1] 1.25 1.75 2.25 2.75 3.25 3.75 4.25 4.75  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $employ  
## $breaks  
## [1] 0 5 10 15 20 25 30 35  
##   
## $counts  
## [1] 337 226 141 93 38 10 5  
##   
## $density  
## [1] 0.079294118 0.053176471 0.033176471 0.021882353 0.008941176 0.002352941  
## [7] 0.001176471  
##   
## $mids  
## [1] 2.5 7.5 12.5 17.5 22.5 27.5 32.5  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $address  
## $breaks  
## [1] 0 5 10 15 20 25 30 35  
##   
## $counts  
## [1] 358 222 128 80 43 16 3  
##   
## $density  
## [1] 0.0842352941 0.0522352941 0.0301176471 0.0188235294 0.0101176471  
## [6] 0.0037647059 0.0007058824  
##   
## $mids  
## [1] 2.5 7.5 12.5 17.5 22.5 27.5 32.5  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $income  
## $breaks  
## [1] 0 50 100 150 200 250 300 350 400 450  
##   
## $counts  
## [1] 602 190 40 8 5 3 1 0 1  
##   
## $density  
## [1] 1.416471e-02 4.470588e-03 9.411765e-04 1.882353e-04 1.176471e-04  
## [6] 7.058824e-05 2.352941e-05 0.000000e+00 2.352941e-05  
##   
## $mids  
## [1] 25 75 125 175 225 275 325 375 425  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $debtinc  
## $breaks  
## [1] 0 5 10 15 20 25 30 35 40 45  
##   
## $counts  
## [1] 209 283 186 102 38 19 10 2 1  
##   
## $density  
## [1] 0.0491764706 0.0665882353 0.0437647059 0.0240000000 0.0089411765  
## [6] 0.0044705882 0.0023529412 0.0004705882 0.0002352941  
##   
## $mids  
## [1] 2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5 42.5  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $creddebt  
## $breaks  
## [1] 0 2 4 6 8 10 12 14 16 18 20 22  
##   
## $counts  
## [1] 652 124 40 18 7 2 1 4 1 0 1  
##   
## $density  
## [1] 0.3835294118 0.0729411765 0.0235294118 0.0105882353 0.0041176471  
## [6] 0.0011764706 0.0005882353 0.0023529412 0.0005882353 0.0000000000  
## [11] 0.0005882353  
##   
## $mids  
## [1] 1 3 5 7 9 11 13 15 17 19 21  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $otherdebt  
## $breaks  
## [1] 0 5 10 15 20 25 30 35 40  
##   
## $counts  
## [1] 703 113 20 10 2 1 0 1  
##   
## $density  
## [1] 0.1654117647 0.0265882353 0.0047058824 0.0023529412 0.0004705882  
## [6] 0.0002352941 0.0000000000 0.0002352941  
##   
## $mids  
## [1] 2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"  
##   
## $default  
## $breaks  
## [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0  
##   
## $counts  
## [1] 517 0 0 0 0 0 0 0 0 183  
##   
## $density  
## [1] 7.385714 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  
## [8] 0.000000 0.000000 2.614286  
##   
## $mids  
## [1] 0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95  
##   
## $xname  
## [1] "x"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"

library("ggplot2", lib.loc="E:/R-3.4.1/R-3.4.2/library")

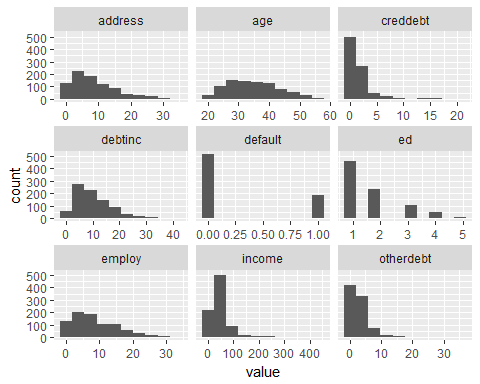


library("tidyr", lib.loc="E:/R-3.4.1/R-3.4.2/library")

## Warning: package 'tidyr' was built under R version 3.4.3

ggplot(gather(data), aes(value)) +geom\_histogram(bins = 10) + facet\_wrap(~key, scales = 'free\_x')

## Warning: Removed 150 rows containing non-finite values (stat\_bin).



library("REdaS", lib.loc="E:/R-3.4.1/R-3.4.2/library")

## Warning: package 'REdaS' was built under R version 3.4.4

## Loading required package: grid

bart\_spher(data[1:8])

## Bartlett's Test of Sphericity  
##   
## Call: bart\_spher(x = data[1:8])  
##   
## X2 = 3164.655  
## df = 28  
## p-value < 2.22e-16

cor(data[1:8])

## age ed employ address income  
## age 1.000000000 0.012982860 0.55424133 0.59994931 0.47621808  
## ed 0.012982860 1.000000000 -0.15111705 0.05292718 0.21821936  
## employ 0.554241333 -0.151117051 1.00000000 0.34466405 0.62509257  
## address 0.599949309 0.052927175 0.34466405 1.00000000 0.30834040  
## income 0.476218084 0.218219356 0.62509257 0.30834040 1.00000000  
## debtinc 0.008240009 0.008053386 -0.03362502 -0.03293884 -0.03558488  
## creddebt 0.278907431 0.099079973 0.38174124 0.16164058 0.55151535  
## otherdebt 0.337841384 0.140936800 0.41443082 0.18548072 0.60335168  
## debtinc creddebt otherdebt  
## age 0.008240009 0.27890743 0.3378414  
## ed 0.008053386 0.09907997 0.1409368  
## employ -0.033625023 0.38174124 0.4144308  
## address -0.032938844 0.16164058 0.1854807  
## income -0.035584878 0.55151535 0.6033517  
## debtinc 1.000000000 0.51496502 0.5725475  
## creddebt 0.514965019 1.00000000 0.6449524  
## otherdebt 0.572547500 0.64495242 1.0000000

KMOS(data[1:8])

##   
## Kaiser-Meyer-Olkin Statistics  
##   
## Call: KMOS(x = data[1:8])  
##   
## Measures of Sampling Adequacy (MSA):  
## age ed employ address income debtinc creddebt   
## 0.7502763 0.3299228 0.7673263 0.7046747 0.5863059 0.3661337 0.7340290   
## otherdebt   
## 0.6590004   
##   
## KMO-Criterion: 0.6348444

# For training  
library("bnlearn", lib.loc="E:/R-3.4.1/R-3.4.2/library")

## Warning: package 'bnlearn' was built under R version 3.4.4

##   
## Attaching package: 'bnlearn'

## The following object is masked from 'package:stats':  
##   
## sigma

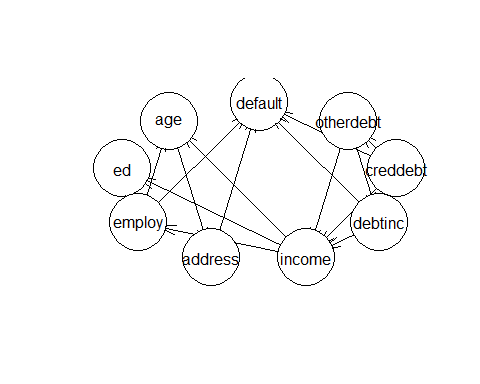
require(caret)

## Loading required package: caret

## Warning: package 'caret' was built under R version 3.4.3

## Loading required package: lattice

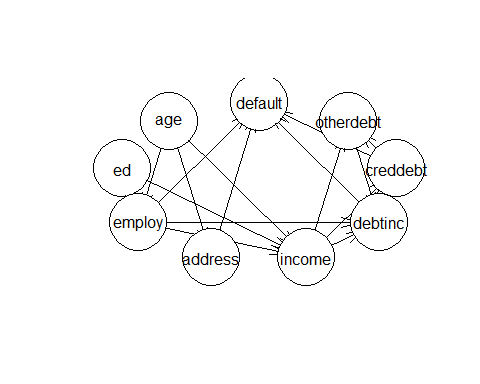
set.seed(123)  
loantrain1<-data[1:600,]  
loantrain2<-data[601:700,]  
  
res <- hc(loantrain1)  
  
res <- hc(loantrain1)  
plot(res)



fitted1 <- bn.fit(res, data = loantrain1);  
  
pred1 <- predict(fitted1, node="default", loantrain1, method="bayes-lw")  
  
pred2 <- predict(fitted1, node="default", loantrain2, method="bayes-lw")  
  
output <- cbind(loantrain2, predicted\_values=round(pred2,digits=0))  
View(output)  
  
confusionMatrix(output$predicted\_values,output$default)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 73 16  
## 1 4 7  
##   
## Accuracy : 0.8   
## 95% CI : (0.7082, 0.8733)  
## No Information Rate : 0.77   
## P-Value [Acc > NIR] : 0.28106   
##   
## Kappa : 0.3089   
## Mcnemar's Test P-Value : 0.01391   
##   
## Sensitivity : 0.9481   
## Specificity : 0.3043   
## Pos Pred Value : 0.8202   
## Neg Pred Value : 0.6364   
## Prevalence : 0.7700   
## Detection Rate : 0.7300   
## Detection Prevalence : 0.8900   
## Balanced Accuracy : 0.6262   
##   
## 'Positive' Class : 0   
##

#for predicting  
data1<-data[1:700,]  
data2<-data[701:850,]  
set.seed(123)  
  
res <- hc(data1)  
plot(res)



fitted1 <- bn.fit(res, data = data1);  
fitted1

##   
## Bayesian network parameters  
##   
## Parameters of node age (Gaussian distribution)  
##   
## Conditional density: age  
## Coefficients:  
## (Intercept)   
## 34.86   
## Standard deviation of the residuals: 7.997342   
##   
## Parameters of node ed (Gaussian distribution)  
##   
## Conditional density: ed  
## Coefficients:  
## (Intercept)   
## 1.722857   
## Standard deviation of the residuals: 0.9282055   
##   
## Parameters of node employ (Gaussian distribution)  
##   
## Conditional density: employ | age + ed  
## Coefficients:  
## (Intercept) age ed   
## -5.2415138 0.4497298 -1.1884309   
## Standard deviation of the residuals: 5.51734   
##   
## Parameters of node address (Gaussian distribution)  
##   
## Conditional density: address | age  
## Coefficients:  
## (Intercept) age   
## -9.4993246 0.5099798   
## Standard deviation of the residuals: 5.476113   
##   
## Parameters of node income (Gaussian distribution)  
##   
## Conditional density: income | age + ed + employ  
## Coefficients:  
## (Intercept) age ed employ   
## -28.8289560 0.7277385 12.7477857 3.2304435   
## Standard deviation of the residuals: 25.73789   
##   
## Parameters of node debtinc (Gaussian distribution)  
##   
## Conditional density: debtinc | employ + income + creddebt + otherdebt  
## Coefficients:  
## (Intercept) employ income creddebt otherdebt   
## 9.82236213 -0.09718114 -0.12775575 1.46303715 1.57162578   
## Standard deviation of the residuals: 3.801818   
##   
## Parameters of node creddebt (Gaussian distribution)  
##   
## Conditional density: creddebt | income  
## Coefficients:  
## (Intercept) income   
## 0.05817742 0.03279229   
## Standard deviation of the residuals: 1.740537   
##   
## Parameters of node otherdebt (Gaussian distribution)  
##   
## Conditional density: otherdebt | income + creddebt  
## Coefficients:  
## (Intercept) income creddebt   
## 0.53330642 0.03303633 0.65552851   
## Standard deviation of the residuals: 2.343804   
##   
## Parameters of node default (Gaussian distribution)  
##   
## Conditional density: default | employ + address + debtinc + creddebt  
## Coefficients:  
## (Intercept) employ address debtinc creddebt   
## 0.272857860 -0.023751563 -0.007297781 0.014818553 0.061910204   
## Standard deviation of the residuals: 0.3734182

pred1 <- predict(fitted1, node="default", data1, method="bayes-lw")  
  
pred2 <- predict(fitted1, node="default", data2[1:8], method="bayes-lw")  
output <- cbind(data2, predicted\_values=round(pred2,digits=0))  
View(output)