#read the dataset  
file=readRDS("loan\_data\_1.rds")

#create frame for dataset  
credit\_frame=data.frame(file)

str(credit\_frame)

## 'data.frame': 29092 obs. of 8 variables:  
## $ loan\_status : int 0 0 0 0 0 0 1 0 1 0 ...  
## $ loan\_amnt : int 5000 2400 10000 5000 3000 12000 9000 3000 10000 1000 ...  
## $ int\_rate : num 10.6 NA 13.5 NA NA ...  
## $ grade : Factor w/ 7 levels "A","B","C","D",..: 2 3 3 1 5 2 3 2 2 4 ...  
## $ emp\_length : int 10 25 13 3 9 11 0 3 3 0 ...  
## $ home\_ownership: Factor w/ 4 levels "MORTGAGE","OTHER",..: 4 4 4 4 4 3 4 4 4 4 ...  
## $ annual\_inc : num 24000 12252 49200 36000 48000 ...  
## $ age : int 33 31 24 39 24 28 22 22 28 22 ...

head(credit\_frame)

## loan\_status loan\_amnt int\_rate grade emp\_length home\_ownership  
## 1 0 5000 10.65 B 10 RENT  
## 2 0 2400 NA C 25 RENT  
## 3 0 10000 13.49 C 13 RENT  
## 4 0 5000 NA A 3 RENT  
## 5 0 3000 NA E 9 RENT  
## 6 0 12000 12.69 B 11 OWN  
## annual\_inc age  
## 1 24000 33  
## 2 12252 31  
## 3 49200 24  
## 4 36000 39  
## 5 48000 24  
## 6 75000 28

sapply(credit\_frame,function(x) sum(is.na(x)))

## loan\_status loan\_amnt int\_rate grade emp\_length   
## 0 0 2776 0 809   
## home\_ownership annual\_inc age   
## 0 0 0

library(gmodels)

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

CrossTable(credit\_frame$loan\_status)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 29092   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 25865 | 3227 |   
## | 0.889 | 0.111 |   
## |-----------|-----------|  
##   
##   
##   
##

CrossTable(credit\_frame$home\_ownership)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 29092   
##   
##   
## | MORTGAGE | OTHER | OWN | RENT |   
## |-----------|-----------|-----------|-----------|  
## | 12002 | 97 | 2301 | 14692 |   
## | 0.413 | 0.003 | 0.079 | 0.505 |   
## |-----------|-----------|-----------|-----------|  
##   
##   
##   
##

CrossTable(credit\_frame$grade)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 29092   
##   
##   
## | A | B | C | D | E |   
## |-----------|-----------|-----------|-----------|-----------|  
## | 9649 | 9329 | 5748 | 3231 | 868 |   
## | 0.332 | 0.321 | 0.198 | 0.111 | 0.030 |   
## |-----------|-----------|-----------|-----------|-----------|  
##   
##   
## | F | G |   
## |-----------|-----------|  
## | 211 | 56 |   
## | 0.007 | 0.002 |   
## |-----------|-----------|  
##   
##   
##   
##

print(summary(credit\_frame$loan\_amnt))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 500 5000 8000 9594 12250 35000

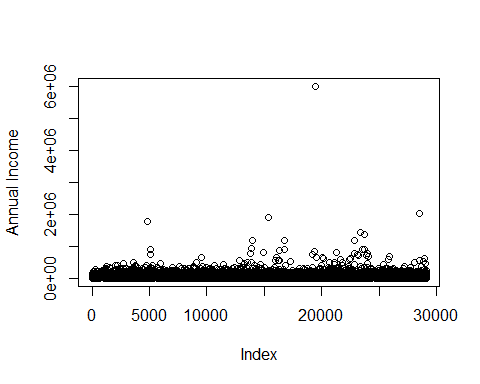
print(summary(credit\_frame$int\_rate))

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 5.42 7.90 10.99 11.00 13.47 23.22 2776

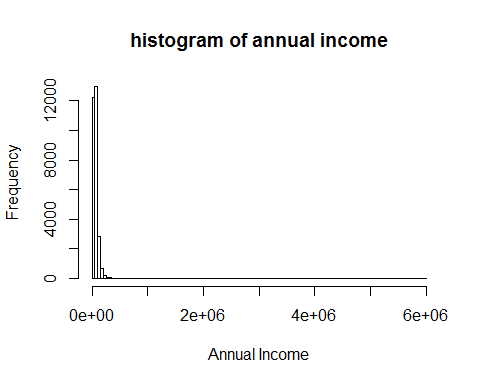
print(summary(credit\_frame$emp\_length))

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.000 2.000 4.000 6.145 8.000 62.000 809

plot(credit\_frame$annual\_inc,ylab="Annual Income")



hist(credit\_frame$annual\_inc,sqrt(nrow(credit\_frame)),xlab = "Annual Income",main = "histogram of annual income")

 From scatter plot extrem value is removed

index\_outlier\_annua\_inc<-which(credit\_frame$annual\_inc>3000000)  
 index\_outlier\_annua\_inc

## [1] 19486

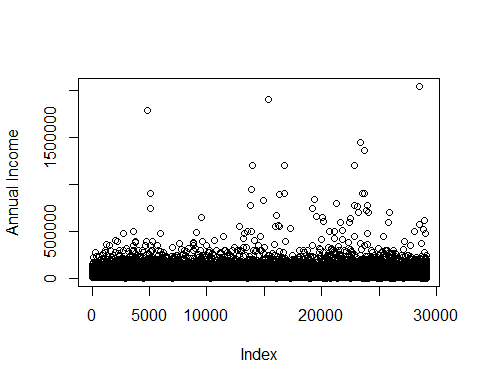
nrow(credit\_frame)

## [1] 29092

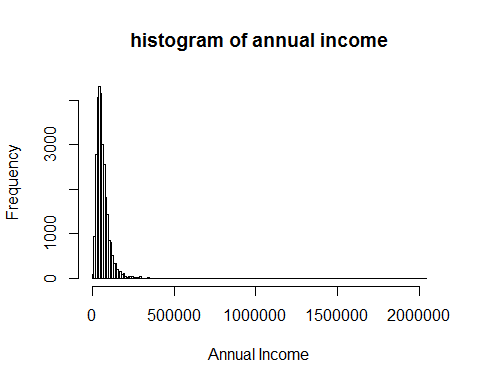
credit\_frame<-credit\_frame[-index\_outlier\_annua\_inc,]  
 nrow(credit\_frame)

## [1] 29091

plot(credit\_frame$annual\_inc,ylab="Annual Income")

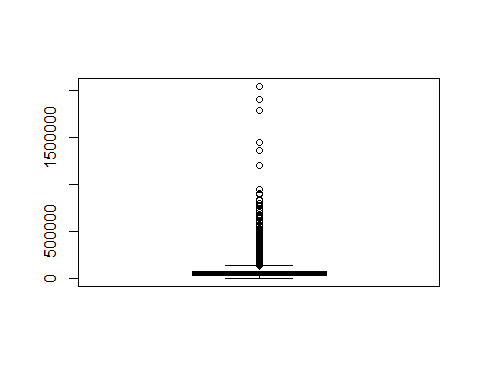


hist(credit\_frame$annual\_inc,sqrt(nrow(credit\_frame)),xlab = "Annual Income",main = "histogram of annual income")



still not look normanl

boxplot(credit\_frame$annual\_inc)



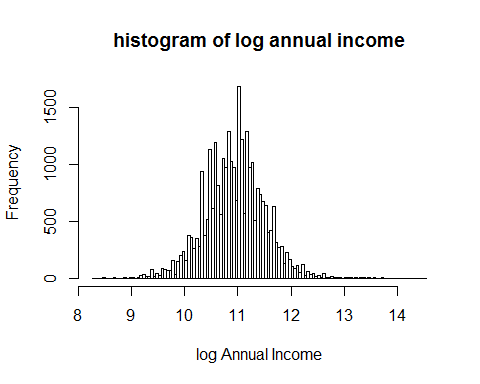
summary(credit\_frame$annual\_inc)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4000 40000 56400 66965 80000 2039784

temp\_vector<-c("log\_annual\_income")  
 credit\_frame[,temp\_vector]<-NA

credit\_frame$log\_annual\_income<-log(credit\_frame$annual\_inc)

hist(credit\_frame$log\_annual\_inc,sqrt(nrow(credit\_frame)),xlab = "log Annual Income",main = "histogram of log annual income")

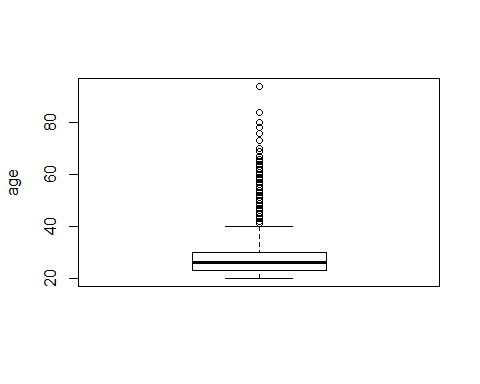


summary(credit\_frame$log\_annual\_income)

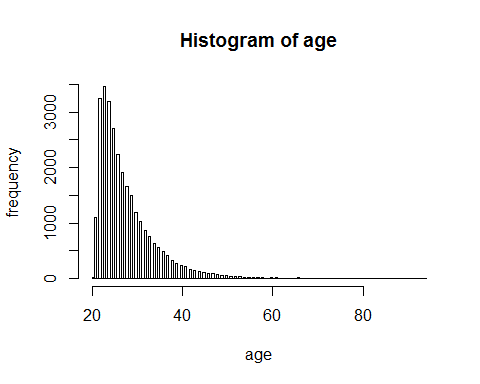
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 8.294 10.597 10.940 10.942 11.290 14.528

Looks normal distribution acepted

boxplot(credit\_frame$age,ylab="age")



hist(credit\_frame$age,sqrt(nrow(credit\_frame)),ylab = "frequency",xlab = "age",main = "Histogram of age")



summary(credit\_frame$age)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 20.0 23.0 26.0 27.7 30.0 94.0

Looks constant mean and median So distribution is quet normal Interest rate has some missing values so fill it with mean

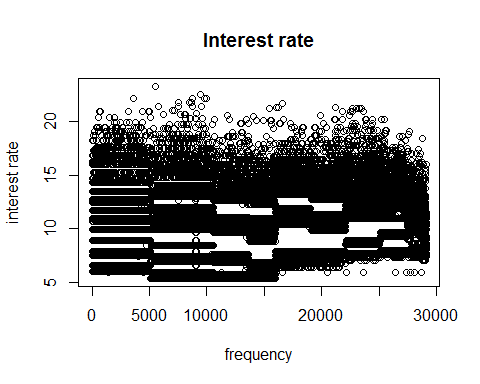
summary(credit\_frame$int\_rate)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 5.42 7.90 10.99 11.00 13.47 23.22 2776

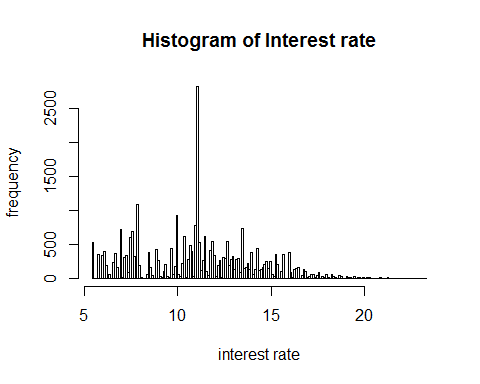
index\_int\_rate\_NA<-which(is.na(credit\_frame$int\_rate))  
credit\_frame$int\_rate[index\_int\_rate\_NA]<-mean(credit\_frame$int\_rate,na.rm = TRUE)  
summary(credit\_frame$int\_rate)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5.42 8.49 11.00 11.00 13.11 23.22

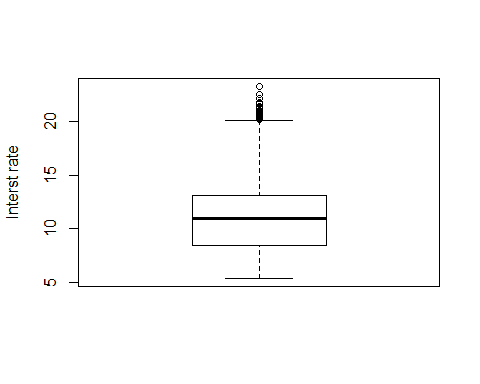
plot(credit\_frame$int\_rate,xlab = "frequency",ylab = "interest rate",main = "Interest rate")



hist(credit\_frame$int\_rate,sqrt(nrow(credit\_frame)),ylab = "frequency",xlab = "interest rate",main = "Histogram of Interest rate")



boxplot(credit\_frame$int\_rate,ylab="Interst rate")



summary(credit\_frame$int\_rate)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5.42 8.49 11.00 11.00 13.11 23.22

From distribution and summary it looks normal and acepted

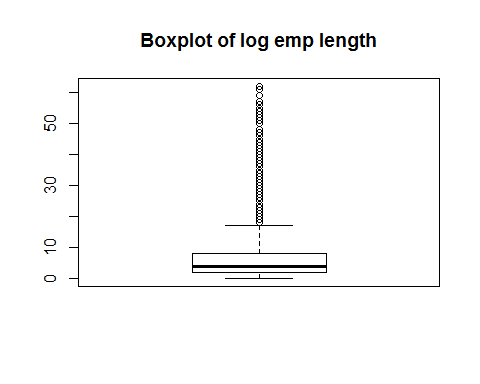
summary(credit\_frame$emp\_length)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.000 2.000 4.000 6.145 8.000 62.000 809

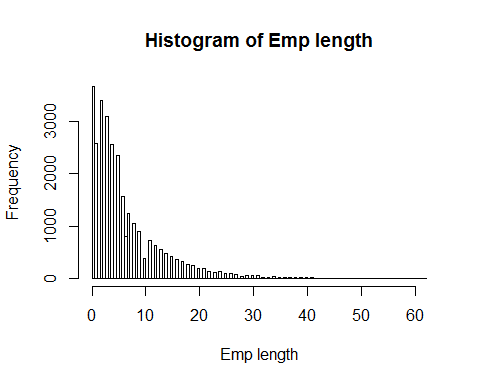
index\_emp\_length\_NA<-which(is.na(credit\_frame$emp\_length))  
credit\_frame$emp\_length[index\_emp\_length\_NA]<-mean(credit\_frame$emp\_length,na.rm = TRUE)  
summary(credit\_frame$emp\_length)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 2.000 4.000 6.145 8.000 62.000

boxplot(credit\_frame$emp\_length,main="Boxplot of log emp length")



hist(credit\_frame$emp\_length,sqrt(nrow(credit\_frame)),ylab = "Frequency",xlab = "Emp length",main = "Histogram of Emp length")

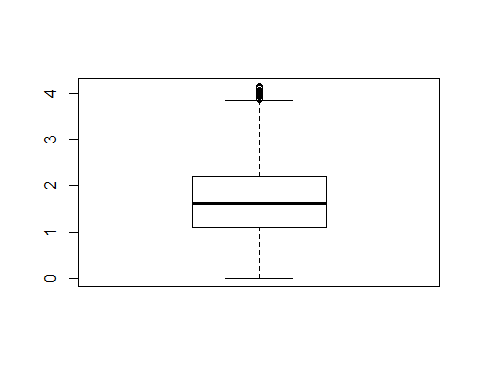
 Emp length has extrem values so preprocess it

credit\_frame$emp\_length<-credit\_frame$emp\_length+1

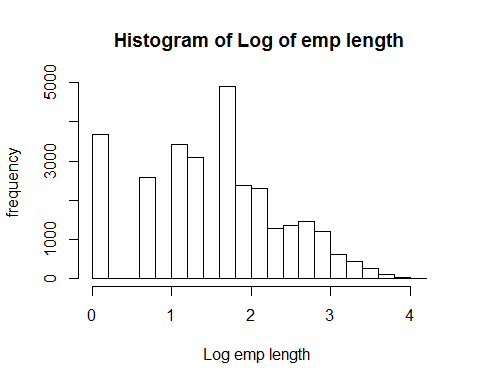
temp\_vector<-c("log\_emp\_length")  
 credit\_frame[,temp\_vector]<-NA

credit\_frame$log\_emp\_length<-log(credit\_frame$emp\_length)

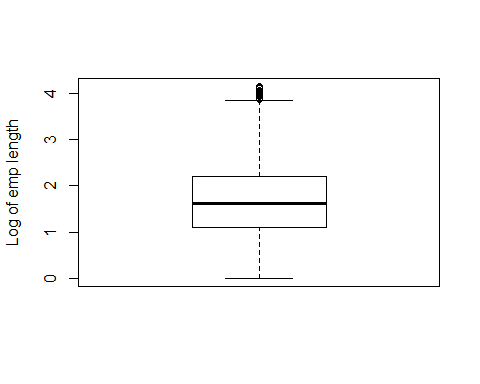
boxplot(credit\_frame$log\_emp\_length)



hist(credit\_frame$log\_emp\_length,ylab = "frequency",xlab ="Log emp length ", main = "Histogram of Log of emp length")



boxplot(credit\_frame$log\_emp\_length,ylab="Log of emp length")

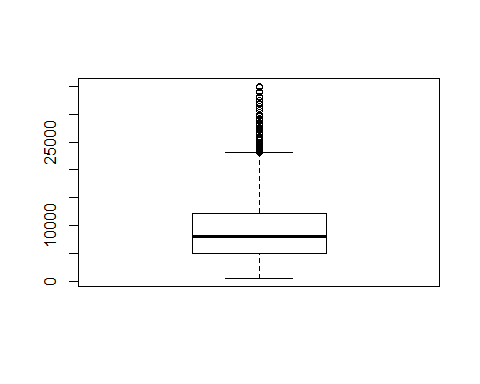


summary(credit\_frame$log\_emp\_length)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 1.099 1.609 1.591 2.197 4.143

Looks normal because normal mean and median

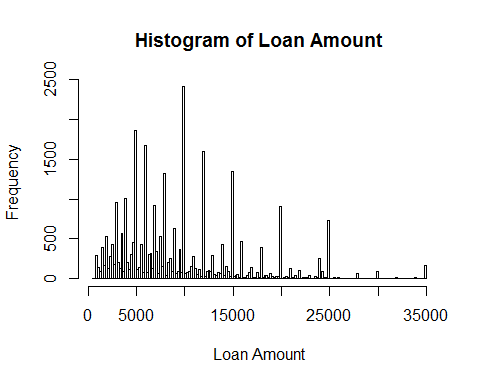
boxplot(credit\_frame$loan\_amnt)



summary(credit\_frame$loan\_amnt)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 500 5000 8000 9594 12250 35000

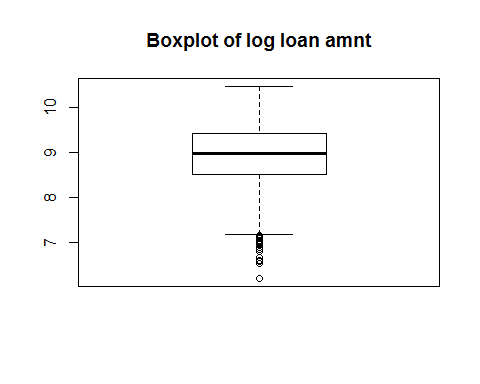
hist(credit\_frame$loan\_amnt,sqrt(nrow(credit\_frame)),ylab = "Frequency",xlab = "Loan Amount", main = "Histogram of Loan Amount")



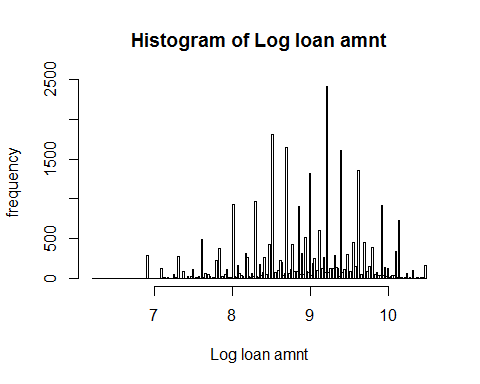
temp\_vector<-c("log\_loan\_amnt")  
 credit\_frame[,temp\_vector]<-NA

credit\_frame$log\_loan\_amnt<-log(credit\_frame$loan\_amnt)

boxplot(credit\_frame$log\_loan\_amnt,main="Boxplot of log loan amnt")



hist(credit\_frame$log\_loan\_amnt,sqrt(nrow(credit\_frame)),ylab = "frequency",xlab = "Log loan amnt",main = "Histogram of Log loan amnt")



summary(credit\_frame$log\_loan\_amnt)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 6.215 8.517 8.987 8.942 9.413 10.463

lets create some new features

CrossTable(credit\_frame$grade,credit\_frame$loan\_status,prop.r = TRUE,prop.c = FALSE,prop.t = FALSE,prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 29091   
##   
##   
## | credit\_frame$loan\_status   
## credit\_frame$grade | 0 | 1 | Row Total |   
## -------------------|-----------|-----------|-----------|  
## A | 9084 | 565 | 9649 |   
## | 0.941 | 0.059 | 0.332 |   
## -------------------|-----------|-----------|-----------|  
## B | 8344 | 985 | 9329 |   
## | 0.894 | 0.106 | 0.321 |   
## -------------------|-----------|-----------|-----------|  
## C | 4903 | 844 | 5747 |   
## | 0.853 | 0.147 | 0.198 |   
## -------------------|-----------|-----------|-----------|  
## D | 2651 | 580 | 3231 |   
## | 0.820 | 0.180 | 0.111 |   
## -------------------|-----------|-----------|-----------|  
## E | 692 | 176 | 868 |   
## | 0.797 | 0.203 | 0.030 |   
## -------------------|-----------|-----------|-----------|  
## F | 155 | 56 | 211 |   
## | 0.735 | 0.265 | 0.007 |   
## -------------------|-----------|-----------|-----------|  
## G | 35 | 21 | 56 |   
## | 0.625 | 0.375 | 0.002 |   
## -------------------|-----------|-----------|-----------|  
## Column Total | 25864 | 3227 | 29091 |   
## -------------------|-----------|-----------|-----------|  
##   
##

Accordingly from above observation the loan status depends on the grade of the customer

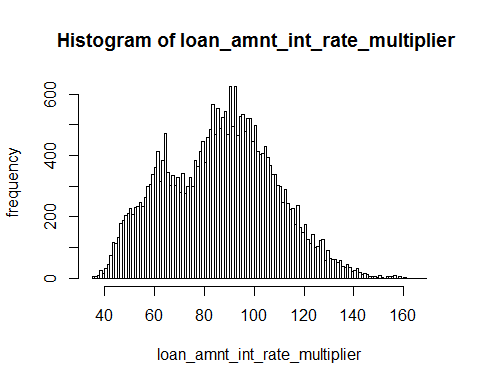
temp\_vector<-c("grade\_multiplier")  
 credit\_frame[,temp\_vector]<-NA

credit\_frame<-within(credit\_frame,grade\_multiplier[grade=='A']<-(0.941))  
credit\_frame<-within(credit\_frame,grade\_multiplier[grade=='B']<-(0.894))  
credit\_frame<-within(credit\_frame,grade\_multiplier[grade=='C']<-(0.853))  
credit\_frame<-within(credit\_frame,grade\_multiplier[grade=='D']<-(0.820))  
credit\_frame<-within(credit\_frame,grade\_multiplier[grade=='E']<-(0.797))  
credit\_frame<-within(credit\_frame,grade\_multiplier[grade=='F']<-(0.735))  
credit\_frame<-within(credit\_frame,grade\_multiplier[grade=='G']<-(0.625))

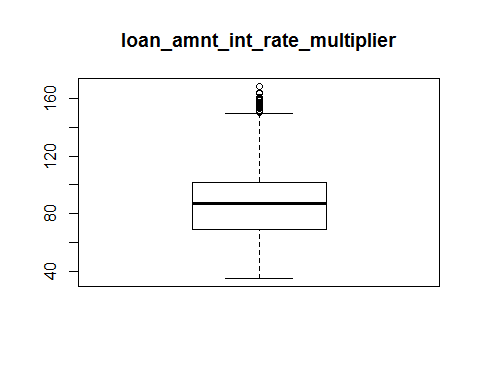
temp\_vector<-c("loan\_amnt\_int\_rate\_multiplier")  
 credit\_frame[,temp\_vector]<-NA

credit\_frame$loan\_amnt\_int\_rate\_multiplier<-((credit\_frame$int\_rate)\*(credit\_frame$log\_loan\_amnt)\*(credit\_frame$grade\_multiplier))

hist(credit\_frame$loan\_amnt\_int\_rate\_multiplier,sqrt(nrow(credit\_frame)),ylab = "frequency",xlab = "loan\_amnt\_int\_rate\_multiplier",main = "Histogram of loan\_amnt\_int\_rate\_multiplier ")



boxplot(credit\_frame$loan\_amnt\_int\_rate\_multiplier,main="loan\_amnt\_int\_rate\_multiplier")



summary(credit\_frame$loan\_amnt\_int\_rate\_multiplier)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 35.23 69.34 87.35 86.42 101.56 168.33

index\_train <- sample(1:nrow(credit\_frame), 2 / 3 \* nrow(credit\_frame))

training\_set<-credit\_frame[index\_train,]  
 test\_set<-credit\_frame[-index\_train,]

x\_test=test\_set[,2:13]  
 y\_test=test\_set[,1]

library(rpart)  
 dt\_model<-rpart(loan\_status~loan\_amnt\_int\_rate\_multiplier+grade+log\_loan\_amnt+int\_rate+log\_emp\_length+log\_annual\_income+age+annual\_inc+home\_ownership+emp\_length+loan\_amnt, method = 'class',data = training\_set, parms = list(prior = c(0.7, 0.3)),control = rpart.control(cp = 0.001) )

summary(dt\_model)

## Call:  
## rpart(formula = loan\_status ~ loan\_amnt\_int\_rate\_multiplier +   
## grade + log\_loan\_amnt + int\_rate + log\_emp\_length + log\_annual\_income +   
## age + annual\_inc + home\_ownership + emp\_length + loan\_amnt,   
## data = training\_set, method = "class", parms = list(prior = c(0.7,   
## 0.3)), control = rpart.control(cp = 0.001))  
## n= 19394   
##   
## CP nsplit rel error xerror xstd  
## 1 0.002246273 0 1.0000000 1.000000 0.02028301  
## 2 0.002008868 8 0.9784841 1.010646 0.01979192  
## 3 0.001969818 11 0.9720816 1.009917 0.01971457  
## 4 0.001718216 14 0.9661721 1.008143 0.01969392  
## 5 0.001402052 18 0.9583664 1.007108 0.01960433  
## 6 0.001356169 19 0.9569644 1.010995 0.01954934  
## 7 0.001295995 21 0.9542520 1.011593 0.01955369  
## 8 0.001249173 25 0.9490681 1.011737 0.01952665  
## 9 0.001083130 31 0.9413241 1.009676 0.01943937  
## 10 0.001083130 32 0.9402409 1.014445 0.01940805  
## 11 0.001022016 33 0.9391578 1.015192 0.01943207  
## 12 0.001000000 35 0.9371138 1.017021 0.01933571  
##   
## Variable importance  
## int\_rate grade   
## 26 25   
## loan\_amnt\_int\_rate\_multiplier log\_annual\_income   
## 22 7   
## annual\_inc log\_loan\_amnt   
## 7 4   
## loan\_amnt home\_ownership   
## 4 1   
## log\_emp\_length emp\_length   
## 1 1   
##   
## Node number 1: 19394 observations, complexity param=0.002246273  
## predicted class=0 expected loss=0.3 P(node) =1  
## class counts: 17234 2160  
## probabilities: 0.700 0.300   
## left son=2 (12678 obs) right son=3 (6716 obs)  
## Primary splits:  
## grade splits as LLRRRRR, improve=277.71310, (0 missing)  
## int\_rate < 11.18 to the left, improve=239.32370, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 73.73915 to the left, improve=180.00460, (0 missing)  
## log\_annual\_income < 11.04171 to the right, improve= 84.82864, (0 missing)  
## annual\_inc < 62424 to the right, improve= 84.82864, (0 missing)  
## Surrogate splits:  
## int\_rate < 12.705 to the left, agree=0.935, adj=0.812, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 99.62825 to the left, agree=0.835, adj=0.524, (0 split)  
## log\_annual\_income < 8.496782 to the right, agree=0.654, adj=0.001, (0 split)  
## annual\_inc < 4900 to the right, agree=0.654, adj=0.001, (0 split)  
## log\_loan\_amnt < 10.39663 to the left, agree=0.654, adj=0.000, (0 split)  
##   
## Node number 2: 12678 observations, complexity param=0.001249173  
## predicted class=0 expected loss=0.2332994 P(node) =0.6167565  
## class counts: 11642 1036  
## probabilities: 0.767 0.233   
## left son=4 (4835 obs) right son=5 (7843 obs)  
## Primary splits:  
## int\_rate < 8.435 to the left, improve=73.17616, (0 missing)  
## grade splits as LR-----, improve=70.45706, (0 missing)  
## log\_annual\_income < 10.83153 to the right, improve=64.80708, (0 missing)  
## annual\_inc < 50591 to the right, improve=64.80708, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 73.73915 to the left, improve=42.97257, (0 missing)  
## Surrogate splits:  
## loan\_amnt\_int\_rate\_multiplier < 69.85782 to the left, agree=0.941, adj=0.846, (0 split)  
## grade splits as LR-----, agree=0.870, adj=0.659, (0 split)  
## log\_annual\_income < 13.87997 to the right, agree=0.619, adj=0.000, (0 split)  
## annual\_inc < 1074000 to the right, agree=0.619, adj=0.000, (0 split)  
##   
## Node number 3: 6716 observations, complexity param=0.002246273  
## predicted class=0 expected loss=0.4073418 P(node) =0.3832435  
## class counts: 5592 1124  
## probabilities: 0.593 0.407   
## left son=6 (2528 obs) right son=7 (4188 obs)  
## Primary splits:  
## log\_annual\_income < 11.09832 to the right, improve=27.678330, (0 missing)  
## annual\_inc < 66060 to the right, improve=27.678330, (0 missing)  
## int\_rate < 17.77 to the left, improve=16.146940, (0 missing)  
## grade splits as --LLLRR, improve=15.725440, (0 missing)  
## log\_loan\_amnt < 8.529613 to the right, improve= 9.121034, (0 missing)  
## Surrogate splits:  
## annual\_inc < 66060 to the right, agree=1.000, adj=1.000, (0 split)  
## log\_loan\_amnt < 9.58534 to the right, agree=0.710, adj=0.229, (0 split)  
## loan\_amnt < 14550 to the right, agree=0.710, adj=0.229, (0 split)  
## home\_ownership splits as LRRR, agree=0.679, adj=0.147, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 115.7983 to the right, agree=0.656, adj=0.087, (0 split)  
##   
## Node number 4: 4835 observations  
## predicted class=0 expected loss=0.1593784 P(node) =0.221346  
## class counts: 4581 254  
## probabilities: 0.841 0.159   
##   
## Node number 5: 7843 observations, complexity param=0.001249173  
## predicted class=0 expected loss=0.2746794 P(node) =0.3954105  
## class counts: 7061 782  
## probabilities: 0.725 0.275   
## left son=10 (4449 obs) right son=11 (3394 obs)  
## Primary splits:  
## log\_annual\_income < 10.84876 to the right, improve=38.88012, (0 missing)  
## annual\_inc < 51470.5 to the right, improve=38.88012, (0 missing)  
## log\_loan\_amnt < 8.181696 to the right, improve=25.45843, (0 missing)  
## loan\_amnt < 3575 to the right, improve=25.45843, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 68.2134 to the right, improve=25.27186, (0 missing)  
## Surrogate splits:  
## annual\_inc < 51470.5 to the right, agree=1.000, adj=1.000, (0 split)  
## log\_loan\_amnt < 9.155091 to the right, agree=0.636, adj=0.160, (0 split)  
## loan\_amnt < 9462.5 to the right, agree=0.636, adj=0.160, (0 split)  
## home\_ownership splits as LLRR, agree=0.626, adj=0.135, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 76.81985 to the right, agree=0.604, adj=0.085, (0 split)  
##   
## Node number 6: 2528 observations, complexity param=0.002246273  
## predicted class=0 expected loss=0.3492898 P(node) =0.1363879  
## class counts: 2185 343  
## probabilities: 0.651 0.349   
## left son=12 (2431 obs) right son=13 (97 obs)  
## Primary splits:  
## int\_rate < 18.275 to the left, improve=14.675520, (0 missing)  
## grade splits as --LLLRR, improve=14.069250, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 116.9874 to the left, improve=11.471920, (0 missing)  
## log\_annual\_income < 11.1191 to the left, improve= 3.919195, (0 missing)  
## annual\_inc < 67447 to the left, improve= 3.919195, (0 missing)  
## Surrogate splits:  
## loan\_amnt\_int\_rate\_multiplier < 151.1022 to the left, agree=0.969, adj=0.196, (0 split)  
## grade splits as --LLLRR, agree=0.969, adj=0.186, (0 split)  
##   
## Node number 7: 4188 observations, complexity param=0.002246273  
## predicted class=0 expected loss=0.4394156 P(node) =0.2468557  
## class counts: 3407 781  
## probabilities: 0.561 0.439   
## left son=14 (3905 obs) right son=15 (283 obs)  
## Primary splits:  
## int\_rate < 16.795 to the left, improve=8.421488, (0 missing)  
## grade splits as --LRRRR, improve=7.543036, (0 missing)  
## log\_annual\_income < 10.25924 to the right, improve=6.299200, (0 missing)  
## annual\_inc < 28545 to the right, improve=6.299200, (0 missing)  
## log\_loan\_amnt < 8.529613 to the right, improve=5.038794, (0 missing)  
## Surrogate splits:  
## loan\_amnt\_int\_rate\_multiplier < 132.2566 to the left, agree=0.953, adj=0.300, (0 split)  
## grade splits as --LLLRR, agree=0.946, adj=0.194, (0 split)  
##   
## Node number 10: 4449 observations  
## predicted class=0 expected loss=0.2286973 P(node) =0.2155931  
## class counts: 4094 355  
## probabilities: 0.771 0.229   
##   
## Node number 11: 3394 observations, complexity param=0.001249173  
## predicted class=0 expected loss=0.32981 P(node) =0.1798173  
## class counts: 2967 427  
## probabilities: 0.670 0.330   
## left son=22 (3200 obs) right son=23 (194 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 68.22236 to the right, improve=19.30124, (0 missing)  
## int\_rate < 11.2 to the left, improve=17.43447, (0 missing)  
## log\_loan\_amnt < 8.181696 to the right, improve=14.90016, (0 missing)  
## loan\_amnt < 3575 to the right, improve=14.90016, (0 missing)  
## log\_annual\_income < 10.33936 to the right, improve=11.68743, (0 missing)  
## Surrogate splits:  
## log\_loan\_amnt < 6.93215 to the right, agree=0.951, adj=0.149, (0 split)  
## loan\_amnt < 1025 to the right, agree=0.951, adj=0.149, (0 split)  
## log\_annual\_income < 8.981163 to the right, agree=0.943, adj=0.005, (0 split)  
## annual\_inc < 7952.02 to the right, agree=0.943, adj=0.005, (0 split)  
##   
## Node number 12: 2431 observations  
## predicted class=0 expected loss=0.3373263 P(node) =0.1296964  
## class counts: 2116 315  
## probabilities: 0.663 0.337   
##   
## Node number 13: 97 observations  
## predicted class=1 expected loss=0.4188305 P(node) =0.006691488  
## class counts: 69 28  
## probabilities: 0.419 0.581   
##   
## Node number 14: 3905 observations, complexity param=0.002246273  
## predicted class=0 expected loss=0.4309329 P(node) =0.2281871  
## class counts: 3197 708  
## probabilities: 0.569 0.431   
## left son=28 (2908 obs) right son=29 (997 obs)  
## Primary splits:  
## log\_emp\_length < 1.956167 to the left, improve=7.377983, (0 missing)  
## emp\_length < 7.072537 to the left, improve=7.377983, (0 missing)  
## log\_annual\_income < 9.847733 to the right, improve=6.865091, (0 missing)  
## annual\_inc < 18915.6 to the right, improve=6.865091, (0 missing)  
## int\_rate < 12.51 to the right, improve=4.567045, (0 missing)  
## Surrogate splits:  
## emp\_length < 7.072537 to the left, agree=1, adj=1, (0 split)  
##   
## Node number 15: 283 observations, complexity param=0.002246273  
## predicted class=1 expected loss=0.4568997 P(node) =0.01866854  
## class counts: 210 73  
## probabilities: 0.457 0.543   
## left son=30 (187 obs) right son=31 (96 obs)  
## Primary splits:  
## log\_annual\_income < 10.59161 to the right, improve=5.460304, (0 missing)  
## annual\_inc < 39800 to the right, improve=5.460304, (0 missing)  
## log\_emp\_length < 2.022932 to the right, improve=3.862336, (0 missing)  
## emp\_length < 7.572537 to the right, improve=3.862336, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 141.4427 to the right, improve=3.732127, (0 missing)  
## Surrogate splits:  
## annual\_inc < 39800 to the right, agree=1.000, adj=1.000, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 117.2659 to the right, agree=0.724, adj=0.188, (0 split)  
## log\_loan\_amnt < 7.928365 to the right, agree=0.724, adj=0.188, (0 split)  
## loan\_amnt < 2775 to the right, agree=0.724, adj=0.188, (0 split)  
## age < 21.5 to the right, agree=0.668, adj=0.021, (0 split)  
##   
## Node number 22: 3200 observations  
## predicted class=0 expected loss=0.3154324 P(node) =0.1673188  
## class counts: 2820 380  
## probabilities: 0.685 0.315   
##   
## Node number 23: 194 observations, complexity param=0.001249173  
## predicted class=1 expected loss=0.4777165 P(node) =0.01249853  
## class counts: 147 47  
## probabilities: 0.478 0.522   
## left son=46 (116 obs) right son=47 (78 obs)  
## Primary splits:  
## age < 24.5 to the right, improve=4.836441, (0 missing)  
## home\_ownership splits as LLLR, improve=3.610448, (0 missing)  
## log\_annual\_income < 10.03888 to the right, improve=3.054623, (0 missing)  
## annual\_inc < 22900 to the right, improve=3.054623, (0 missing)  
## grade splits as RL-----, improve=2.769010, (0 missing)  
## Surrogate splits:  
## loan\_amnt\_int\_rate\_multiplier < 56.70491 to the right, agree=0.613, adj=0.038, (0 split)  
## log\_loan\_amnt < 8.026779 to the left, agree=0.608, adj=0.026, (0 split)  
## loan\_amnt < 3062.5 to the left, agree=0.608, adj=0.026, (0 split)  
## home\_ownership splits as LRLL, agree=0.603, adj=0.013, (0 split)  
##   
## Node number 28: 2908 observations, complexity param=0.001969818  
## predicted class=0 expected loss=0.4134482 P(node) =0.1669563  
## class counts: 2411 497  
## probabilities: 0.587 0.413   
## left son=56 (2766 obs) right son=57 (142 obs)  
## Primary splits:  
## log\_annual\_income < 9.847733 to the right, improve=4.957201, (0 missing)  
## annual\_inc < 18915.6 to the right, improve=4.957201, (0 missing)  
## int\_rate < 16.2 to the left, improve=4.901051, (0 missing)  
## age < 56 to the right, improve=4.054437, (0 missing)  
## home\_ownership splits as LRLL, improve=3.347302, (0 missing)  
## Surrogate splits:  
## annual\_inc < 18915.6 to the right, agree=1, adj=1, (0 split)  
##   
## Node number 29: 997 observations, complexity param=0.002246273  
## predicted class=0 expected loss=0.4786079 P(node) =0.06123082  
## class counts: 786 211  
## probabilities: 0.521 0.479   
## left son=58 (701 obs) right son=59 (296 obs)  
## Primary splits:  
## log\_annual\_income < 10.5796 to the right, improve=7.342232, (0 missing)  
## annual\_inc < 39324.5 to the right, improve=7.342232, (0 missing)  
## log\_emp\_length < 2.022932 to the right, improve=6.563638, (0 missing)  
## emp\_length < 7.572537 to the right, improve=6.563638, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 90.67624 to the right, improve=4.471188, (0 missing)  
## Surrogate splits:  
## annual\_inc < 39324.5 to the right, agree=1.000, adj=1.000, (0 split)  
## log\_emp\_length < 2.022932 to the right, agree=0.743, adj=0.135, (0 split)  
## emp\_length < 7.572537 to the right, agree=0.743, adj=0.135, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 63.58772 to the right, agree=0.704, adj=0.003, (0 split)  
## log\_loan\_amnt < 7.130098 to the right, agree=0.704, adj=0.003, (0 split)  
##   
## Node number 30: 187 observations, complexity param=0.002246273  
## predicted class=0 expected loss=0.4739804 P(node) =0.01142804  
## class counts: 148 39  
## probabilities: 0.526 0.474   
## left son=60 (172 obs) right son=61 (15 obs)  
## Primary splits:  
## int\_rate < 20.01 to the left, improve=6.403223, (0 missing)  
## log\_emp\_length < 2.740319 to the right, improve=6.004504, (0 missing)  
## emp\_length < 15.5 to the right, improve=6.004504, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 141.4427 to the right, improve=3.965956, (0 missing)  
## grade splits as ---LLLR, improve=3.545293, (0 missing)  
## Surrogate splits:  
## grade splits as ---LLLR, agree=0.947, adj=0.333, (0 split)  
##   
## Node number 31: 96 observations  
## predicted class=1 expected loss=0.3478044 P(node) =0.0072405  
## class counts: 62 34  
## probabilities: 0.348 0.652   
##   
## Node number 46: 116 observations, complexity param=0.001249173  
## predicted class=0 expected loss=0.4304841 P(node) =0.006775318  
## class counts: 95 21  
## probabilities: 0.570 0.430   
## left son=92 (49 obs) right son=93 (67 obs)  
## Primary splits:  
## home\_ownership splits as L-LR, improve=5.556195, (0 missing)  
## log\_loan\_amnt < 6.998916 to the left, improve=4.833509, (0 missing)  
## loan\_amnt < 1100 to the left, improve=4.833509, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 59.76406 to the left, improve=4.116279, (0 missing)  
## int\_rate < 8.61 to the right, improve=3.475406, (0 missing)  
## Surrogate splits:  
## log\_loan\_amnt < 7.958296 to the right, agree=0.655, adj=0.184, (0 split)  
## loan\_amnt < 2862.5 to the right, agree=0.655, adj=0.184, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 65.55488 to the right, agree=0.629, adj=0.122, (0 split)  
## log\_emp\_length < 2.191013 to the right, agree=0.621, adj=0.102, (0 split)  
## age < 35.5 to the right, agree=0.621, adj=0.102, (0 split)  
##   
## Node number 47: 78 observations  
## predicted class=1 expected loss=0.3690415 P(node) =0.005723215  
## class counts: 52 26  
## probabilities: 0.369 0.631   
##   
## Node number 56: 2766 observations  
## predicted class=0 expected loss=0.406764 P(node) =0.1577491  
## class counts: 2304 462  
## probabilities: 0.593 0.407   
##   
## Node number 57: 142 observations, complexity param=0.001969818  
## predicted class=1 expected loss=0.4720299 P(node) =0.009207171  
## class counts: 107 35  
## probabilities: 0.472 0.528   
## left son=114 (117 obs) right son=115 (25 obs)  
## Primary splits:  
## log\_loan\_amnt < 8.496782 to the left, improve=4.244725, (0 missing)  
## loan\_amnt < 4900 to the left, improve=4.244725, (0 missing)  
## log\_emp\_length < 1.700599 to the right, improve=3.172113, (0 missing)  
## emp\_length < 5.5 to the right, improve=3.172113, (0 missing)  
## age < 22.5 to the right, improve=2.598261, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 4900 to the left, agree=1.000, adj=1.00, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 109.3237 to the left, agree=0.852, adj=0.16, (0 split)  
## log\_annual\_income < 9.819875 to the left, agree=0.845, adj=0.12, (0 split)  
## annual\_inc < 18396 to the left, agree=0.845, adj=0.12, (0 split)  
## home\_ownership splits as LRLL, agree=0.831, adj=0.04, (0 split)  
##   
## Node number 58: 701 observations, complexity param=0.001718216  
## predicted class=0 expected loss=0.4400496 P(node) =0.04134635  
## class counts: 570 131  
## probabilities: 0.560 0.440   
## left son=116 (361 obs) right son=117 (340 obs)  
## Primary splits:  
## log\_loan\_amnt < 8.998045 to the right, improve=5.335836, (0 missing)  
## loan\_amnt < 8087.5 to the right, improve=5.335836, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 89.32745 to the right, improve=4.270797, (0 missing)  
## log\_annual\_income < 10.68034 to the left, improve=3.357251, (0 missing)  
## annual\_inc < 43492.5 to the left, improve=3.357251, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 8087.5 to the right, agree=1.000, adj=1.000, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 104.5094 to the right, agree=0.722, adj=0.426, (0 split)  
## log\_annual\_income < 10.78704 to the right, agree=0.546, adj=0.065, (0 split)  
## annual\_inc < 48389.8 to the right, agree=0.546, adj=0.065, (0 split)  
## int\_rate < 13.05 to the right, agree=0.542, adj=0.056, (0 split)  
##   
## Node number 59: 296 observations, complexity param=0.002008868  
## predicted class=1 expected loss=0.4412165 P(node) =0.01988447  
## class counts: 216 80  
## probabilities: 0.441 0.559   
## left son=118 (192 obs) right son=119 (104 obs)  
## Primary splits:  
## log\_loan\_amnt < 8.984062 to the left, improve=4.550100, (0 missing)  
## loan\_amnt < 7975 to the left, improve=4.550100, (0 missing)  
## log\_emp\_length < 3.113268 to the left, improve=1.957785, (0 missing)  
## emp\_length < 22.5 to the left, improve=1.957785, (0 missing)  
## log\_annual\_income < 10.55706 to the left, improve=1.857807, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 7975 to the left, agree=1.000, adj=1.000, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 109.7825 to the left, agree=0.750, adj=0.288, (0 split)  
## grade splits as --LLR--, agree=0.662, adj=0.038, (0 split)  
## log\_annual\_income < 10.43406 to the left, agree=0.652, adj=0.010, (0 split)  
## annual\_inc < 33998 to the left, agree=0.652, adj=0.010, (0 split)  
##   
## Node number 60: 172 observations, complexity param=0.001356169  
## predicted class=0 expected loss=0.4291562 P(node) =0.01003261  
## class counts: 141 31  
## probabilities: 0.571 0.429   
## left son=120 (18 obs) right son=121 (154 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 141.4028 to the right, improve=5.633441, (0 missing)  
## log\_emp\_length < 2.740319 to the right, improve=4.633830, (0 missing)  
## emp\_length < 15.5 to the right, improve=4.633830, (0 missing)  
## log\_annual\_income < 10.95615 to the left, improve=3.720656, (0 missing)  
## annual\_inc < 57306.4 to the left, improve=3.720656, (0 missing)  
## Surrogate splits:  
## log\_loan\_amnt < 10.27446 to the right, agree=0.907, adj=0.111, (0 split)  
## loan\_amnt < 29000 to the right, agree=0.907, adj=0.111, (0 split)  
##   
## Node number 61: 15 observations  
## predicted class=1 expected loss=0.2037516 P(node) =0.001395433  
## class counts: 7 8  
## probabilities: 0.204 0.796   
##   
## Node number 92: 49 observations  
## predicted class=0 expected loss=0.2330998 P(node) =0.002383338  
## class counts: 45 4  
## probabilities: 0.767 0.233   
##   
## Node number 93: 67 observations, complexity param=0.001249173  
## predicted class=1 expected loss=0.462404 P(node) =0.00439198  
## class counts: 50 17  
## probabilities: 0.462 0.538   
## left son=186 (21 obs) right son=187 (46 obs)  
## Primary splits:  
## log\_loan\_amnt < 7.278724 to the left, improve=3.984300, (0 missing)  
## loan\_amnt < 1450 to the left, improve=3.984300, (0 missing)  
## int\_rate < 8.61 to the right, improve=3.408730, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 67.92667 to the left, improve=2.479759, (0 missing)  
## age < 30.5 to the left, improve=1.880807, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 1450 to the left, agree=1.000, adj=1.000, (0 split)  
## int\_rate < 9.635 to the right, agree=0.761, adj=0.238, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 58.15038 to the left, agree=0.746, adj=0.190, (0 split)  
##   
## Node number 114: 117 observations, complexity param=0.001969818  
## predicted class=0 expected loss=0.4687738 P(node) =0.00711075  
## class counts: 93 24  
## probabilities: 0.531 0.469   
## left son=228 (84 obs) right son=229 (33 obs)  
## Primary splits:  
## log\_loan\_amnt < 7.481457 to the right, improve=5.620671, (0 missing)  
## loan\_amnt < 1775 to the right, improve=5.620671, (0 missing)  
## log\_emp\_length < 1.497866 to the right, improve=3.674375, (0 missing)  
## emp\_length < 4.5 to the right, improve=3.674375, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 101.1905 to the right, improve=2.599802, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 1775 to the right, agree=1.000, adj=1.000, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 82.54801 to the right, agree=0.795, adj=0.273, (0 split)  
##   
## Node number 115: 25 observations  
## predicted class=1 expected loss=0.2712448 P(node) =0.002096421  
## class counts: 14 11  
## probabilities: 0.271 0.729   
##   
## Node number 116: 361 observations, complexity param=0.001718216  
## predicted class=0 expected loss=0.3806539 P(node) =0.02006781  
## class counts: 306 55  
## probabilities: 0.619 0.381   
## left son=232 (300 obs) right son=233 (61 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 122.1986 to the left, improve=7.208961, (0 missing)  
## log\_loan\_amnt < 9.609938 to the left, improve=6.104287, (0 missing)  
## loan\_amnt < 14912.5 to the left, improve=6.104287, (0 missing)  
## log\_annual\_income < 10.87008 to the left, improve=4.874700, (0 missing)  
## annual\_inc < 52580 to the left, improve=4.874700, (0 missing)  
## Surrogate splits:  
## int\_rate < 15.35 to the left, agree=0.934, adj=0.607, (0 split)  
## grade splits as --LLRL-, agree=0.850, adj=0.115, (0 split)  
## log\_loan\_amnt < 9.958195 to the left, agree=0.837, adj=0.033, (0 split)  
## loan\_amnt < 21125 to the left, agree=0.837, adj=0.033, (0 split)  
##   
## Node number 117: 340 observations, complexity param=0.001718216  
## predicted class=0 expected loss=0.4960657 P(node) =0.02127854  
## class counts: 264 76  
## probabilities: 0.504 0.496   
## left son=234 (59 obs) right son=235 (281 obs)  
## Primary splits:  
## log\_loan\_amnt < 7.88071 to the left, improve=4.576275, (0 missing)  
## loan\_amnt < 2650 to the left, improve=4.576275, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 118.1625 to the right, improve=4.356086, (0 missing)  
## int\_rate < 16.045 to the right, improve=3.982175, (0 missing)  
## log\_annual\_income < 10.62133 to the left, improve=2.505748, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 2650 to the left, agree=1.000, adj=1.000, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 73.7463 to the left, agree=0.847, adj=0.119, (0 split)  
##   
## Node number 118: 192 observations, complexity param=0.002008868  
## predicted class=0 expected loss=0.496683 P(node) =0.01202421  
## class counts: 149 43  
## probabilities: 0.503 0.497   
## left son=236 (149 obs) right son=237 (43 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 105.7363 to the left, improve=5.314651, (0 missing)  
## int\_rate < 13.82 to the left, improve=4.893972, (0 missing)  
## log\_emp\_length < 2.022932 to the right, improve=4.475729, (0 missing)  
## emp\_length < 7.572537 to the right, improve=4.475729, (0 missing)  
## log\_annual\_income < 9.88488 to the right, improve=4.005618, (0 missing)  
## Surrogate splits:  
## int\_rate < 14.935 to the left, agree=0.854, adj=0.349, (0 split)  
## log\_loan\_amnt < 8.867751 to the left, agree=0.786, adj=0.047, (0 split)  
## loan\_amnt < 7100 to the left, agree=0.786, adj=0.047, (0 split)  
##   
## Node number 119: 104 observations, complexity param=0.00108313  
## predicted class=1 expected loss=0.3462184 P(node) =0.007860254  
## class counts: 67 37  
## probabilities: 0.346 0.654   
## left son=238 (8 obs) right son=239 (96 obs)  
## Primary splits:  
## int\_rate < 15.975 to the right, improve=5.619528, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 103.5404 to the right, improve=5.404226, (0 missing)  
## age < 22.5 to the right, improve=2.085322, (0 missing)  
## log\_annual\_income < 10.43667 to the right, improve=1.888251, (0 missing)  
## annual\_inc < 34087 to the right, improve=1.888251, (0 missing)  
## Surrogate splits:  
## loan\_amnt\_int\_rate\_multiplier < 121.0277 to the right, agree=0.942, adj=0.25, (0 split)  
##   
## Node number 120: 18 observations  
## predicted class=0 expected loss=0 P(node) =0.0007311129  
## class counts: 18 0  
## probabilities: 1.000 0.000   
##   
## Node number 121: 154 observations, complexity param=0.001356169  
## predicted class=0 expected loss=0.4628886 P(node) =0.009301494  
## class counts: 123 31  
## probabilities: 0.537 0.463   
## left son=242 (121 obs) right son=243 (33 obs)  
## Primary splits:  
## log\_loan\_amnt < 9.653874 to the left, improve=5.285979, (0 missing)  
## loan\_amnt < 15587.5 to the left, improve=5.285979, (0 missing)  
## log\_annual\_income < 10.95604 to the left, improve=4.455914, (0 missing)  
## annual\_inc < 57300 to the left, improve=4.455914, (0 missing)  
## log\_emp\_length < 1.956167 to the right, improve=4.399029, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 15587.5 to the left, agree=1.000, adj=1.000, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 136.9532 to the left, agree=0.838, adj=0.242, (0 split)  
##   
## Node number 186: 21 observations  
## predicted class=0 expected loss=0.2646743 P(node) =0.001049508  
## class counts: 19 2  
## probabilities: 0.735 0.265   
##   
## Node number 187: 46 observations  
## predicted class=1 expected loss=0.3767089 P(node) =0.003342472  
## class counts: 31 15  
## probabilities: 0.377 0.623   
##   
## Node number 228: 84 observations  
## predicted class=0 expected loss=0.3630198 P(node) =0.004591118  
## class counts: 72 12  
## probabilities: 0.637 0.363   
##   
## Node number 229: 33 observations  
## predicted class=1 expected loss=0.3385277 P(node) =0.002519632  
## class counts: 21 12  
## probabilities: 0.339 0.661   
##   
## Node number 232: 300 observations, complexity param=0.001402052  
## predicted class=0 expected loss=0.3315284 P(node) =0.01591953  
## class counts: 262 38  
## probabilities: 0.668 0.332   
## left son=464 (275 obs) right son=465 (25 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 87.89876 to the right, improve=6.409603, (0 missing)  
## log\_annual\_income < 10.85335 to the left, improve=3.583215, (0 missing)  
## annual\_inc < 51708 to the left, improve=3.583215, (0 missing)  
## int\_rate < 12.7 to the right, improve=3.326097, (0 missing)  
## log\_emp\_length < 3.384247 to the right, improve=2.143566, (0 missing)  
## Surrogate splits:  
## int\_rate < 11.01725 to the right, agree=0.92, adj=0.04, (0 split)  
##   
## Node number 233: 61 observations  
## predicted class=1 expected loss=0.4308211 P(node) =0.004148276  
## class counts: 44 17  
## probabilities: 0.431 0.569   
##   
## Node number 234: 59 observations  
## predicted class=0 expected loss=0.3152138 P(node) =0.003084326  
## class counts: 52 7  
## probabilities: 0.685 0.315   
##   
## Node number 235: 281 observations, complexity param=0.001718216  
## predicted class=1 expected loss=0.4732759 P(node) =0.01819422  
## class counts: 212 69  
## probabilities: 0.473 0.527   
## left son=470 (24 obs) right son=471 (257 obs)  
## Primary splits:  
## log\_annual\_income < 10.62133 to the left, improve=6.981685, (0 missing)  
## annual\_inc < 40999.96 to the left, improve=6.981685, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 118.1625 to the right, improve=4.929086, (0 missing)  
## int\_rate < 16.045 to the right, improve=4.350540, (0 missing)  
## log\_loan\_amnt < 8.541588 to the right, improve=2.778527, (0 missing)  
## Surrogate splits:  
## annual\_inc < 40999.96 to the left, agree=1, adj=1, (0 split)  
##   
## Node number 236: 149 observations, complexity param=0.002008868  
## predicted class=0 expected loss=0.430771 P(node) =0.008705321  
## class counts: 122 27  
## probabilities: 0.569 0.431   
## left son=472 (118 obs) right son=473 (31 obs)  
## Primary splits:  
## log\_annual\_income < 9.889905 to the right, improve=6.123072, (0 missing)  
## annual\_inc < 19732 to the right, improve=6.123072, (0 missing)  
## log\_loan\_amnt < 8.739536 to the right, improve=6.094961, (0 missing)  
## loan\_amnt < 6250 to the right, improve=6.094961, (0 missing)  
## log\_emp\_length < 2.022932 to the right, improve=5.356056, (0 missing)  
## Surrogate splits:  
## annual\_inc < 19732 to the right, agree=1.000, adj=1.000, (0 split)  
## log\_emp\_length < 2.022932 to the right, agree=0.805, adj=0.065, (0 split)  
## emp\_length < 7.572537 to the right, agree=0.805, adj=0.065, (0 split)  
## log\_loan\_amnt < 7.167152 to the right, agree=0.799, adj=0.032, (0 split)  
## loan\_amnt < 1300 to the right, agree=0.799, adj=0.032, (0 split)  
##   
## Node number 237: 43 observations  
## predicted class=1 expected loss=0.3304324 P(node) =0.003318892  
## class counts: 27 16  
## probabilities: 0.330 0.670   
##   
## Node number 238: 8 observations  
## predicted class=0 expected loss=0 P(node) =0.0003249391  
## class counts: 8 0  
## probabilities: 1.000 0.000   
##   
## Node number 239: 96 observations  
## predicted class=1 expected loss=0.318026 P(node) =0.007535315  
## class counts: 59 37  
## probabilities: 0.318 0.682   
##   
## Node number 242: 121 observations  
## predicted class=0 expected loss=0.3891098 P(node) =0.006781862  
## class counts: 102 19  
## probabilities: 0.611 0.389   
##   
## Node number 243: 33 observations  
## predicted class=1 expected loss=0.3385277 P(node) =0.002519632  
## class counts: 21 12  
## probabilities: 0.339 0.661   
##   
## Node number 464: 275 observations  
## predicted class=0 expected loss=0.2951331 P(node) =0.01411793  
## class counts: 245 30  
## probabilities: 0.705 0.295   
##   
## Node number 465: 25 observations  
## predicted class=1 expected loss=0.3832665 P(node) =0.001801607  
## class counts: 17 8  
## probabilities: 0.383 0.617   
##   
## Node number 470: 24 observations  
## predicted class=0 expected loss=0.1294291 P(node) =0.001073089  
## class counts: 23 1  
## probabilities: 0.871 0.129   
##   
## Node number 471: 257 observations, complexity param=0.001295995  
## predicted class=1 expected loss=0.4483749 P(node) =0.01712113  
## class counts: 189 68  
## probabilities: 0.448 0.552   
## left son=942 (9 obs) right son=943 (248 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 118.1625 to the right, improve=4.408724, (0 missing)  
## int\_rate < 16.045 to the right, improve=4.064524, (0 missing)  
## log\_emp\_length < 2.602003 to the left, improve=2.768074, (0 missing)  
## emp\_length < 13.5 to the left, improve=2.768074, (0 missing)  
## log\_loan\_amnt < 8.541588 to the right, improve=2.701600, (0 missing)  
## Surrogate splits:  
## int\_rate < 16.63 to the right, agree=0.977, adj=0.333, (0 split)  
##   
## Node number 472: 118 observations  
## predicted class=0 expected loss=0.3491208 P(node) =0.006365195  
## class counts: 102 16  
## probabilities: 0.651 0.349   
##   
## Node number 473: 31 observations  
## predicted class=1 expected loss=0.3471385 P(node) =0.002340125  
## class counts: 20 11  
## probabilities: 0.347 0.653   
##   
## Node number 942: 9 observations  
## predicted class=0 expected loss=0 P(node) =0.0003655565  
## class counts: 9 0  
## probabilities: 1.000 0.000   
##   
## Node number 943: 248 observations, complexity param=0.001295995  
## predicted class=1 expected loss=0.4363401 P(node) =0.01675557  
## class counts: 180 68  
## probabilities: 0.436 0.564   
## left son=1886 (149 obs) right son=1887 (99 obs)  
## Primary splits:  
## log\_emp\_length < 2.602003 to the left, improve=3.078674, (0 missing)  
## emp\_length < 13.5 to the left, improve=3.078674, (0 missing)  
## log\_annual\_income < 10.79953 to the right, improve=2.242637, (0 missing)  
## annual\_inc < 48998 to the right, improve=2.242637, (0 missing)  
## log\_loan\_amnt < 8.097528 to the right, improve=2.192359, (0 missing)  
## Surrogate splits:  
## emp\_length < 13.5 to the left, agree=1.000, adj=1.000, (0 split)  
## log\_loan\_amnt < 8.318445 to the right, agree=0.625, adj=0.061, (0 split)  
## int\_rate < 14.63 to the left, agree=0.625, adj=0.061, (0 split)  
## loan\_amnt < 4100 to the right, agree=0.625, adj=0.061, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 106.9349 to the left, agree=0.621, adj=0.051, (0 split)  
##   
## Node number 1886: 149 observations, complexity param=0.001295995  
## predicted class=1 expected loss=0.4972734 P(node) =0.009393221  
## class counts: 115 34  
## probabilities: 0.497 0.503   
## left son=3772 (140 obs) right son=3773 (9 obs)  
## Primary splits:  
## log\_loan\_amnt < 8.115712 to the right, improve=3.462290, (0 missing)  
## loan\_amnt < 3350 to the right, improve=3.462290, (0 missing)  
## int\_rate < 15.83 to the right, improve=3.299534, (0 missing)  
## log\_emp\_length < 2.022932 to the right, improve=2.902773, (0 missing)  
## emp\_length < 7.572537 to the right, improve=2.902773, (0 missing)  
## Surrogate splits:  
## loan\_amnt < 3350 to the right, agree=1.000, adj=1.000, (0 split)  
## loan\_amnt\_int\_rate\_multiplier < 75.29871 to the right, agree=0.946, adj=0.111, (0 split)  
##   
## Node number 1887: 99 observations  
## predicted class=1 expected loss=0.3585987 P(node) =0.007362352  
## class counts: 65 34  
## probabilities: 0.359 0.641   
##   
## Node number 3772: 140 observations, complexity param=0.001295995  
## predicted class=0 expected loss=0.4718408 P(node) =0.008536307  
## class counts: 111 29  
## probabilities: 0.528 0.472   
## left son=7544 (100 obs) right son=7545 (40 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 104.5123 to the left, improve=3.852298, (0 missing)  
## log\_loan\_amnt < 8.824245 to the left, improve=2.987728, (0 missing)  
## loan\_amnt < 6800 to the left, improve=2.987728, (0 missing)  
## int\_rate < 15.83 to the right, improve=2.917056, (0 missing)  
## log\_emp\_length < 2.022932 to the right, improve=2.422212, (0 missing)  
## Surrogate splits:  
## int\_rate < 14.63 to the left, agree=0.907, adj=0.675, (0 split)  
## grade splits as --LRL--, agree=0.757, adj=0.150, (0 split)  
## log\_emp\_length < 2.022932 to the right, agree=0.743, adj=0.100, (0 split)  
## emp\_length < 7.572537 to the right, agree=0.743, adj=0.100, (0 split)  
## log\_annual\_income < 11.07828 to the left, agree=0.721, adj=0.025, (0 split)  
##   
## Node number 3773: 9 observations  
## predicted class=1 expected loss=0.1895984 P(node) =0.000856914  
## class counts: 4 5  
## probabilities: 0.190 0.810   
##   
## Node number 7544: 100 observations, complexity param=0.001022016  
## predicted class=0 expected loss=0.3944249 P(node) =0.005634082  
## class counts: 84 16  
## probabilities: 0.606 0.394   
## left son=15088 (38 obs) right son=15089 (62 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 98.05428 to the right, improve=2.978653, (0 missing)  
## age < 27.5 to the left, improve=2.419646, (0 missing)  
## log\_annual\_income < 10.72104 to the right, improve=1.882800, (0 missing)  
## annual\_inc < 45300 to the right, improve=1.882800, (0 missing)  
## home\_ownership splits as L-RR, improve=1.379353, (0 missing)  
## Surrogate splits:  
## int\_rate < 13.015 to the right, agree=0.89, adj=0.711, (0 split)  
## home\_ownership splits as R-LR, agree=0.67, adj=0.132, (0 split)  
## grade splits as --RLR--, agree=0.65, adj=0.079, (0 split)  
## log\_loan\_amnt < 8.940814 to the right, agree=0.64, adj=0.053, (0 split)  
## loan\_amnt < 7637.5 to the right, agree=0.64, adj=0.053, (0 split)  
##   
## Node number 7545: 40 observations, complexity param=0.00108313  
## predicted class=1 expected loss=0.3778719 P(node) =0.002902225  
## class counts: 27 13  
## probabilities: 0.378 0.622   
## left son=15090 (8 obs) right son=15091 (32 obs)  
## Primary splits:  
## int\_rate < 15.83 to the right, improve=5.493225, (0 missing)  
## log\_loan\_amnt < 8.816611 to the left, improve=3.931213, (0 missing)  
## loan\_amnt < 6750 to the left, improve=3.931213, (0 missing)  
## loan\_amnt\_int\_rate\_multiplier < 108.9222 to the right, improve=2.079474, (0 missing)  
## home\_ownership splits as R-LL, improve=1.053526, (0 missing)  
## Surrogate splits:  
## loan\_amnt\_int\_rate\_multiplier < 115.8943 to the right, agree=0.925, adj=0.625, (0 split)  
## age < 32.5 to the right, agree=0.825, adj=0.125, (0 split)  
## home\_ownership splits as R-LR, agree=0.825, adj=0.125, (0 split)  
##   
## Node number 15088: 38 observations  
## predicted class=0 expected loss=0.2266618 P(node) =0.001838275  
## class counts: 35 3  
## probabilities: 0.773 0.227   
##   
## Node number 15089: 62 observations, complexity param=0.001022016  
## predicted class=0 expected loss=0.475671 P(node) =0.003795807  
## class counts: 49 13  
## probabilities: 0.524 0.476   
## left son=30178 (55 obs) right son=30179 (7 obs)  
## Primary splits:  
## loan\_amnt\_int\_rate\_multiplier < 95.81119 to the left, improve=6.657832, (0 missing)  
## int\_rate < 12.895 to the left, improve=3.842823, (0 missing)  
## log\_annual\_income < 10.97505 to the left, improve=1.837302, (0 missing)  
## annual\_inc < 58400 to the left, improve=1.837302, (0 missing)  
## age < 27.5 to the left, improve=1.573368, (0 missing)  
##   
## Node number 15090: 8 observations  
## predicted class=0 expected loss=0 P(node) =0.0003249391  
## class counts: 8 0  
## probabilities: 1.000 0.000   
##   
## Node number 15091: 32 observations  
## predicted class=1 expected loss=0.2994353 P(node) =0.002577286  
## class counts: 19 13  
## probabilities: 0.299 0.701   
##   
## Node number 30178: 55 observations  
## predicted class=0 expected loss=0.367902 P(node) =0.003020128  
## class counts: 47 8  
## probabilities: 0.632 0.368   
##   
## Node number 30179: 7 observations  
## predicted class=1 expected loss=0.1047273 P(node) =0.0007756792  
## class counts: 2 5  
## probabilities: 0.105 0.895

predicted1<-predict(dt\_model,x\_test,type = 'class')

mean(predicted1==y\_test)

## [1] 0.8640817

library(caTools)

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

table(test\_set$loan\_status,predicted1)

## predicted1  
## 0 1  
## 0 8280 350  
## 1 968 99