Loan Approval Project

Benjamin Garcia

2025-05-08

Data Prep

```
library(readr)
loan_data <- read_csv("loan_data.csv")
```

```
## Rows: 24000 Columns: 7
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (3): Text, Employment_Status, Approval
## dbl (4): Income, Credit_Score, Loan_Amount, DTI_Ratio
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Set binary values

```
loan_data$Employment_Status <- as.factor(ifelse(loan_data$Employment_Status == 'employed', 1,
0))
loan_data$Approval <- as.factor(ifelse(loan_data$Approval == 'Approved' , 1, 0))
```

For employment status, 1 if employed, 0 if not employed. For approval, 1 if approved, 0 if not.

```
table(loan_data$Approval)
```

```
##
##     0     1
## 20067  3933
```

```
(2*3933)/20067
```

```
## [1] 0.3919868
```

balance classes

```
set.seed(8675309)
ind <- sample(1:20067, 3933, replace = F)
```

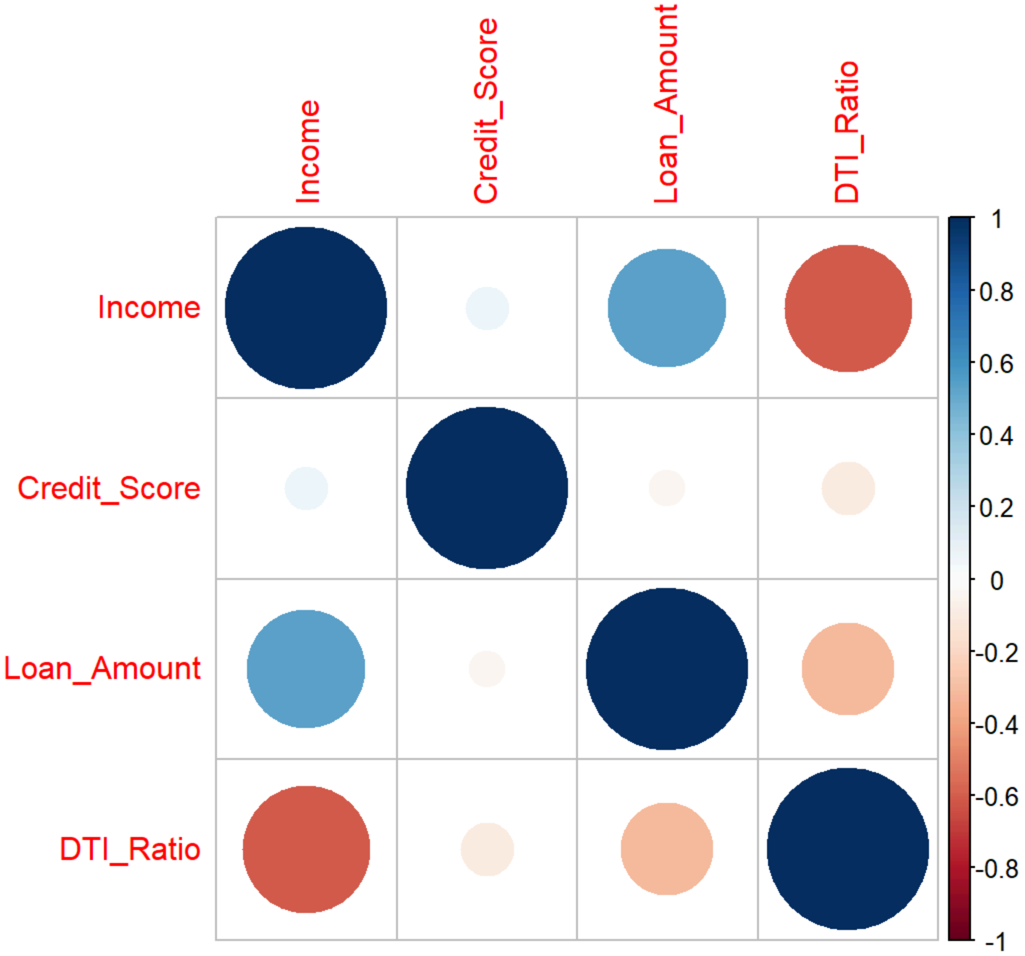
```
not_approved <- loan_data[loan_data$Approval == 0,]
not_approved <- not_approved[ind,]
approved <- loan_data[loan_data$Approval == 1,]
loan2 <- rbind(not_approved, approved)
```

scale numeric data

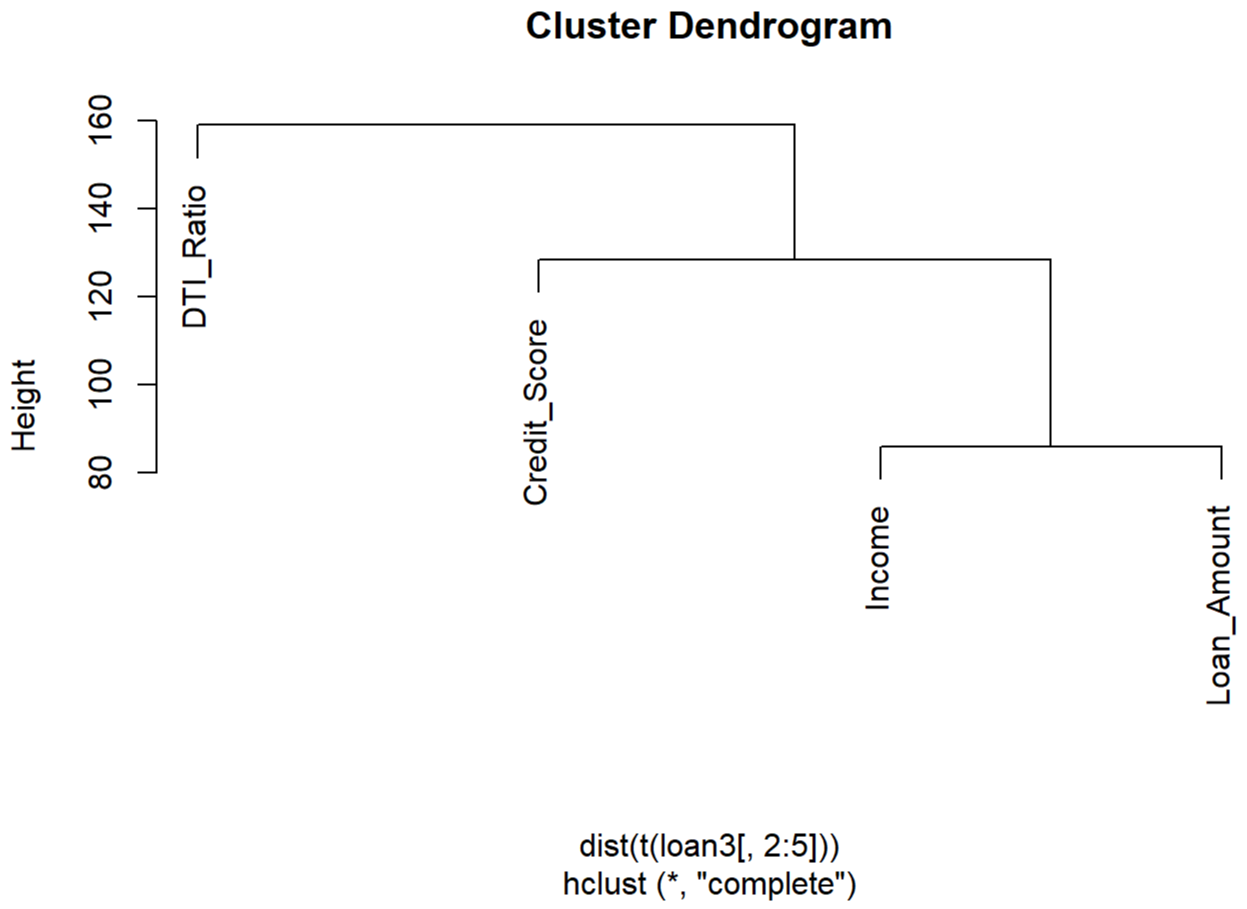
```
loan3 <- cbind(loan2[,1],scale(loan2[,2:5]),loan2[,6:7])
```

Visualizations and Variable Analysis

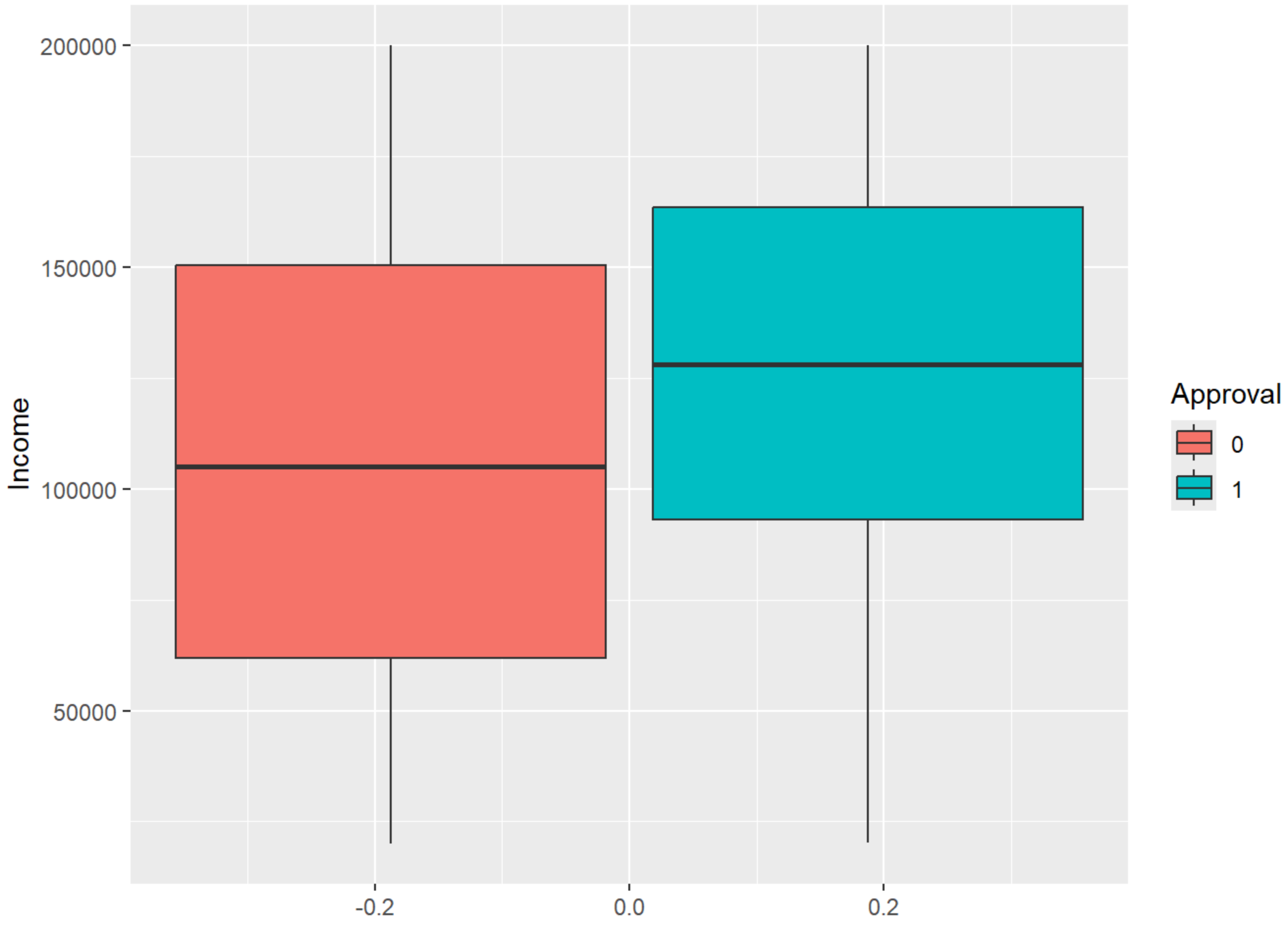
```
library(ggplot2)
corrplot::corrplot(cor(loan3[,2:5]))
```



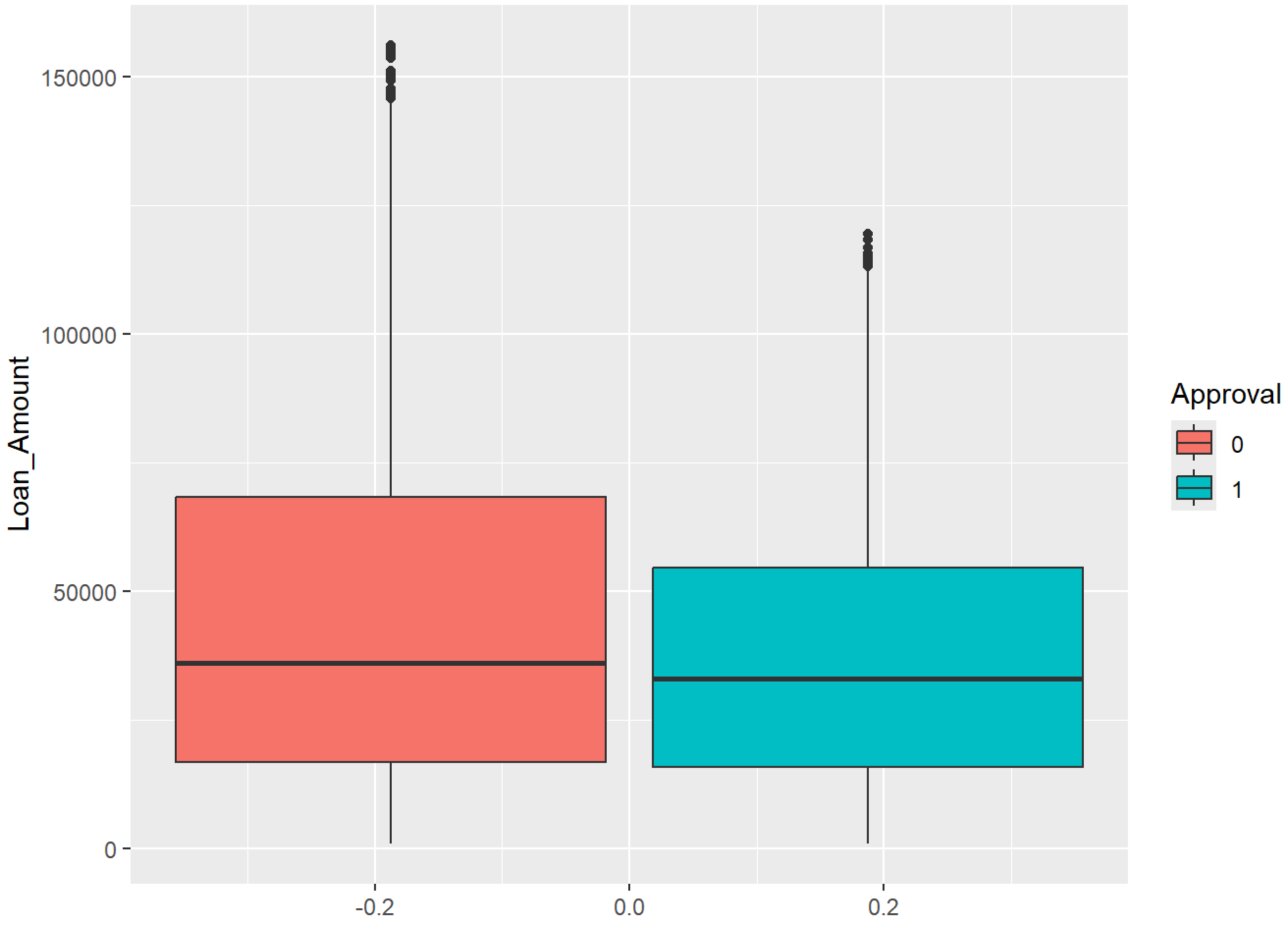
```
plot(hclust(dist(t(loan3[,2:5]))))
```



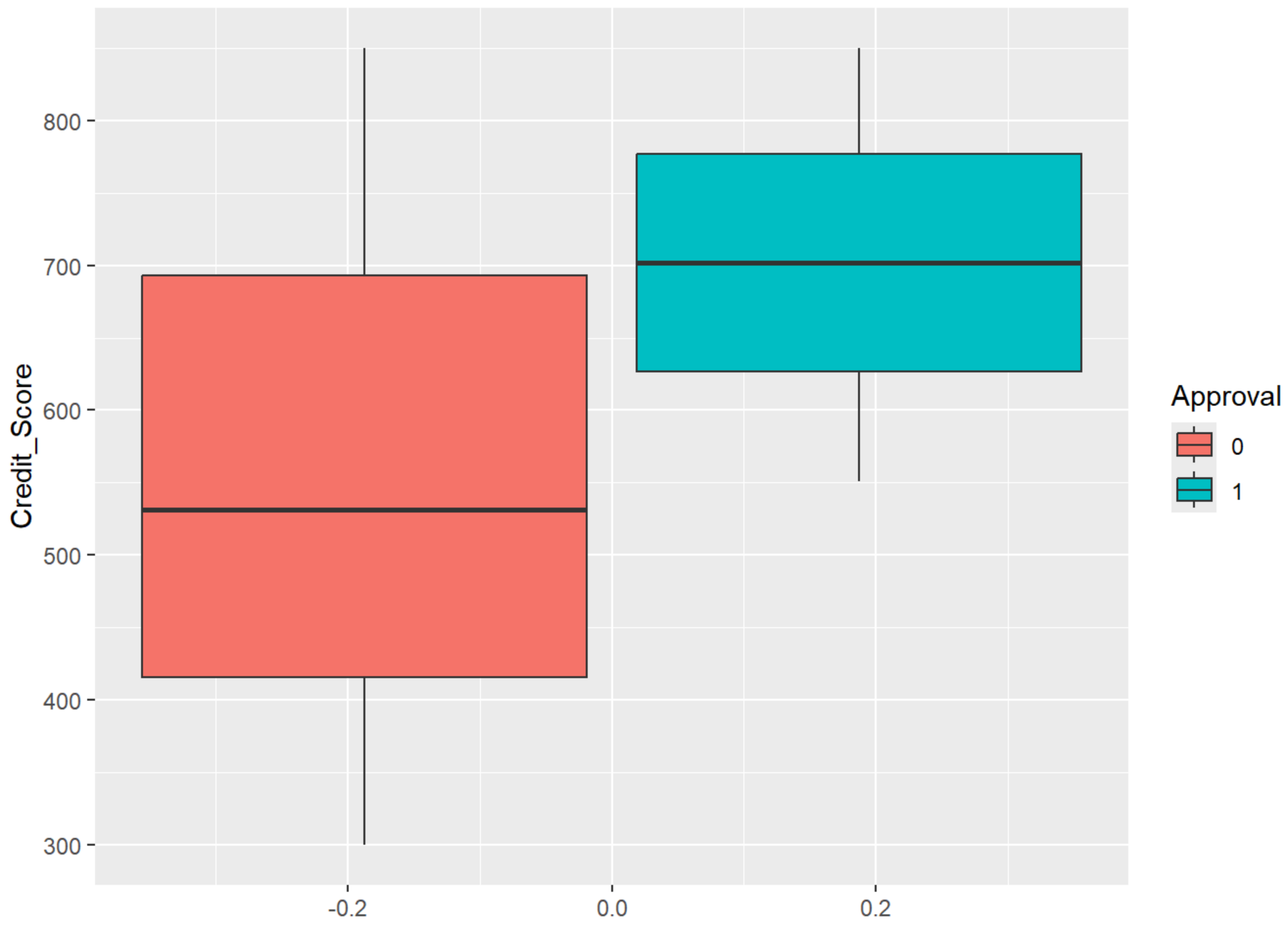
```
ggplot(loan2, aes(fill = Approval, y = Income)) + geom_boxplot()
```



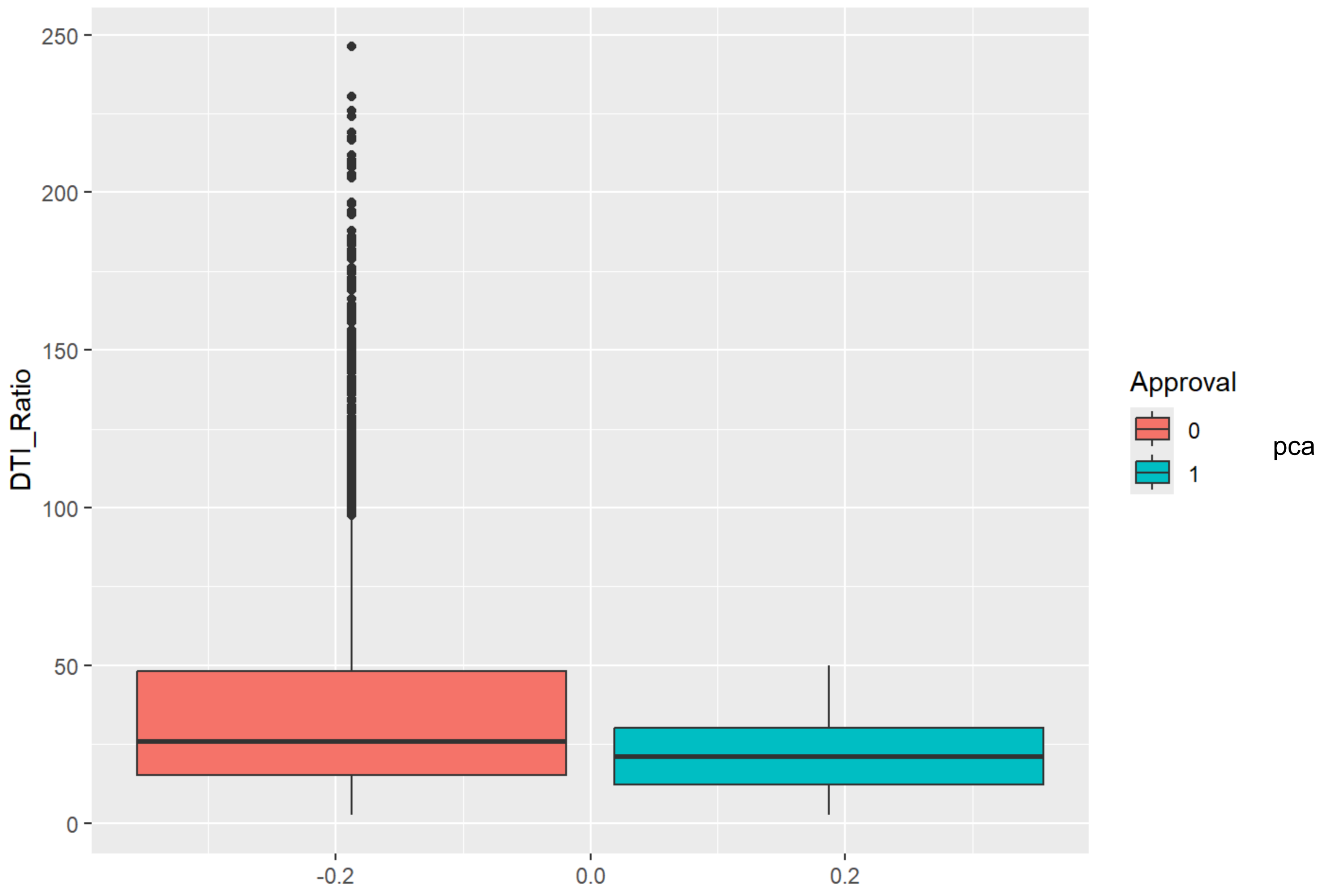
```
ggplot(loan2, aes(fill = Approval, y = Loan_Amount)) + geom_boxplot()
```



```
ggplot(loan2, aes(fill = Approval, y = Credit_Score)) + geom_boxplot()
```



```
ggplot(loan2, aes(fill = Approval, y = DTI_Ratio)) + geom_boxplot()
```



```
data_pca <- princomp(loan3[,2:5])
```

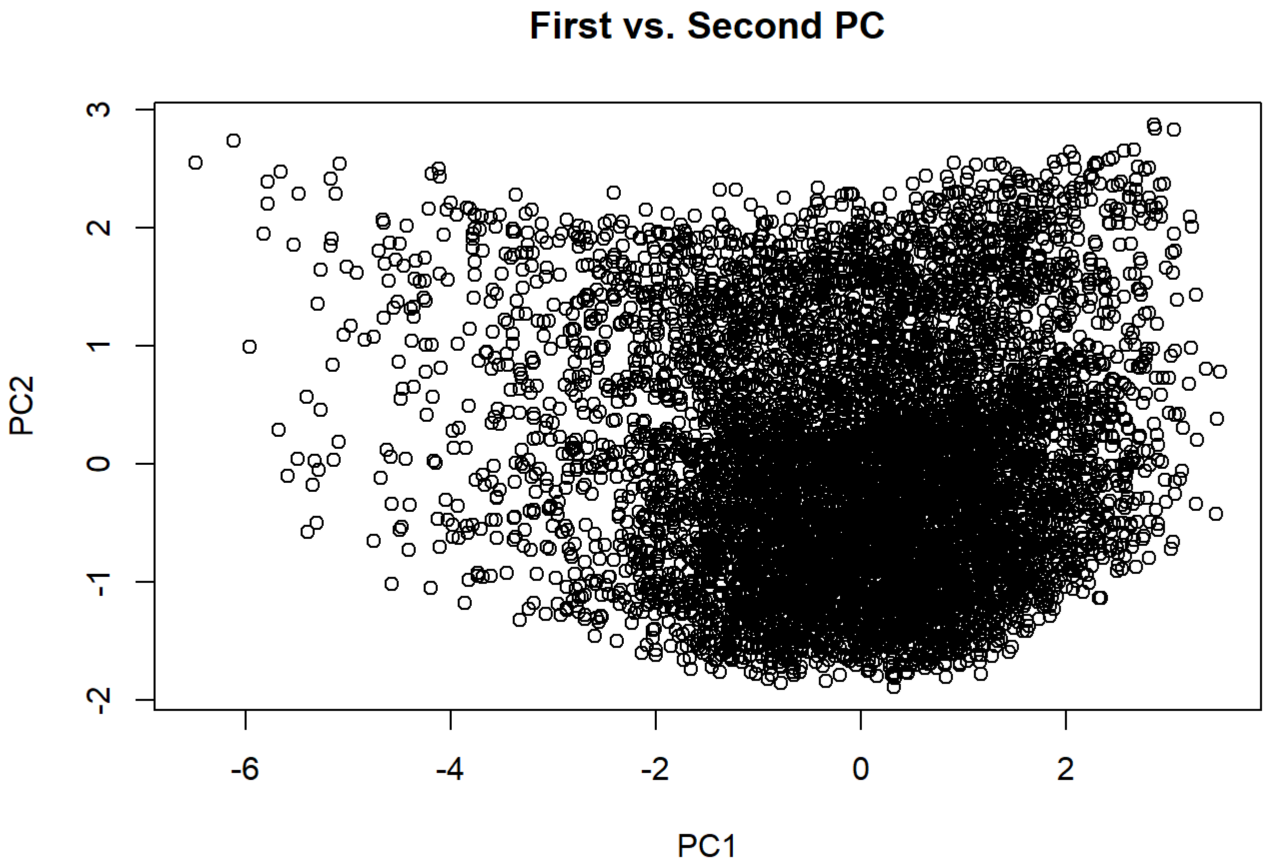
```
print(summary(data_pca))
```

```
## Importance of components:
##                           Comp.1    Comp.2    Comp.3     Comp.4
## Standard deviation     1.4096277 1.0146459 0.8069985 0.57592389
## Proportion of Variance 0.4968257 0.2574093 0.1628323 0.08293262
## Cumulative Proportion  0.4968257 0.7542350 0.9170674 1.00000000
```

```
print(summary(prcomp(loan3[,2:5])))
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4
## Standard deviation     1.4097 1.0147 0.8070 0.57596
## Proportion of Variance 0.4968 0.2574 0.1628 0.08293
## Cumulative Proportion  0.4968 0.7542 0.9171 1.00000
```

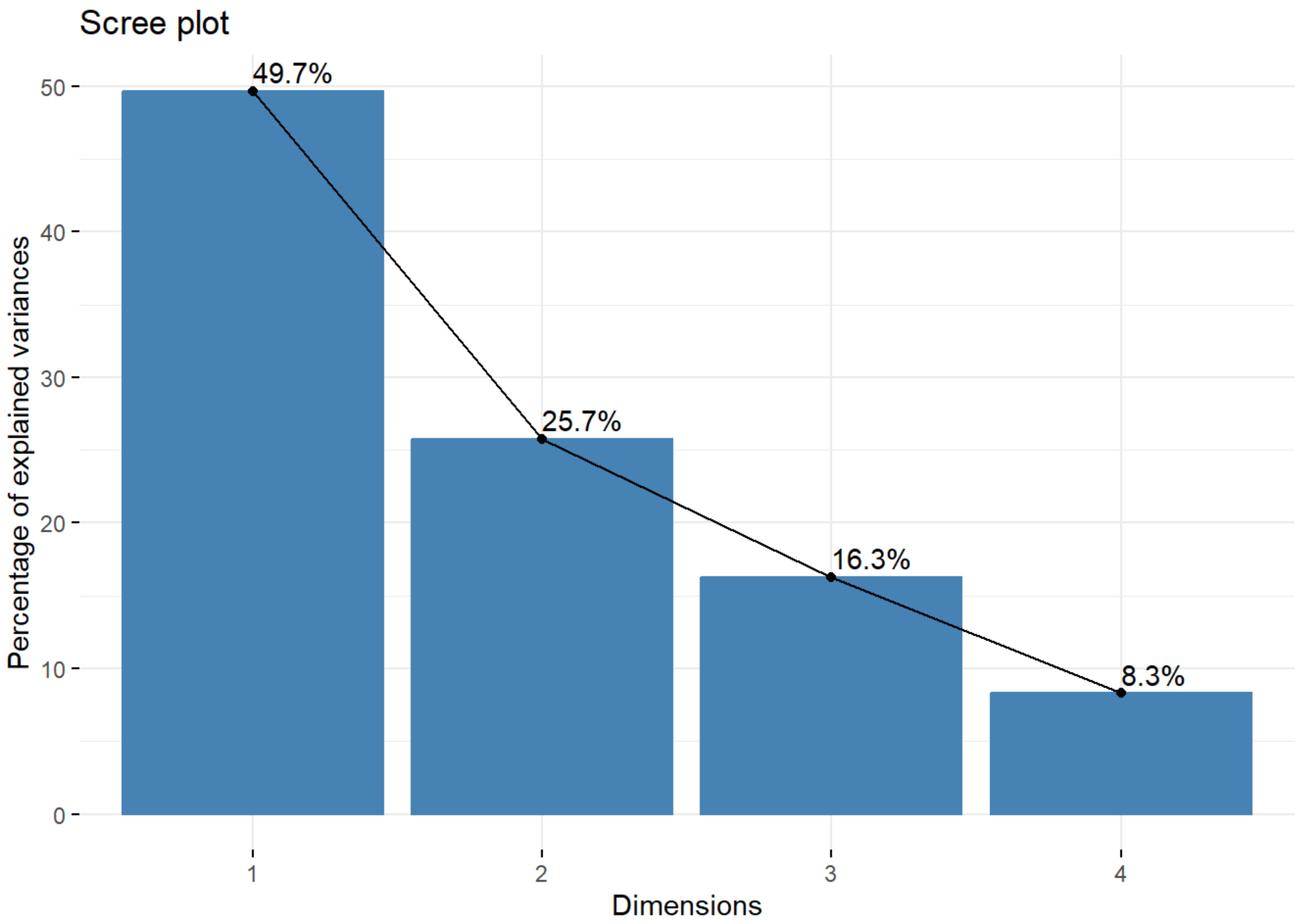
```
plot(data_pca$scores[,1], data_pca$scores[,2], xlab = "PC1", ylab = "PC2", main = "First vs. Sec
ond PC")
```



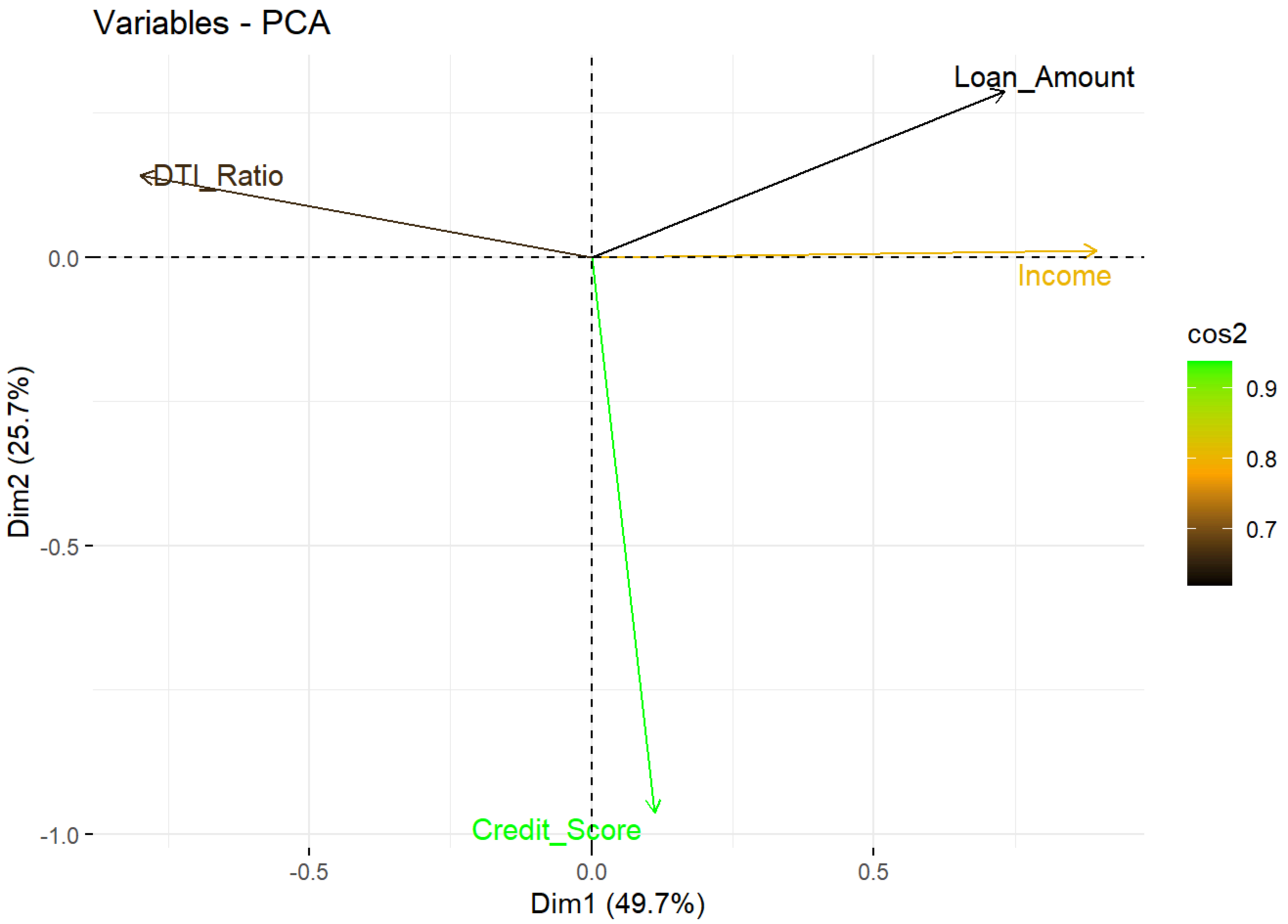
```
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

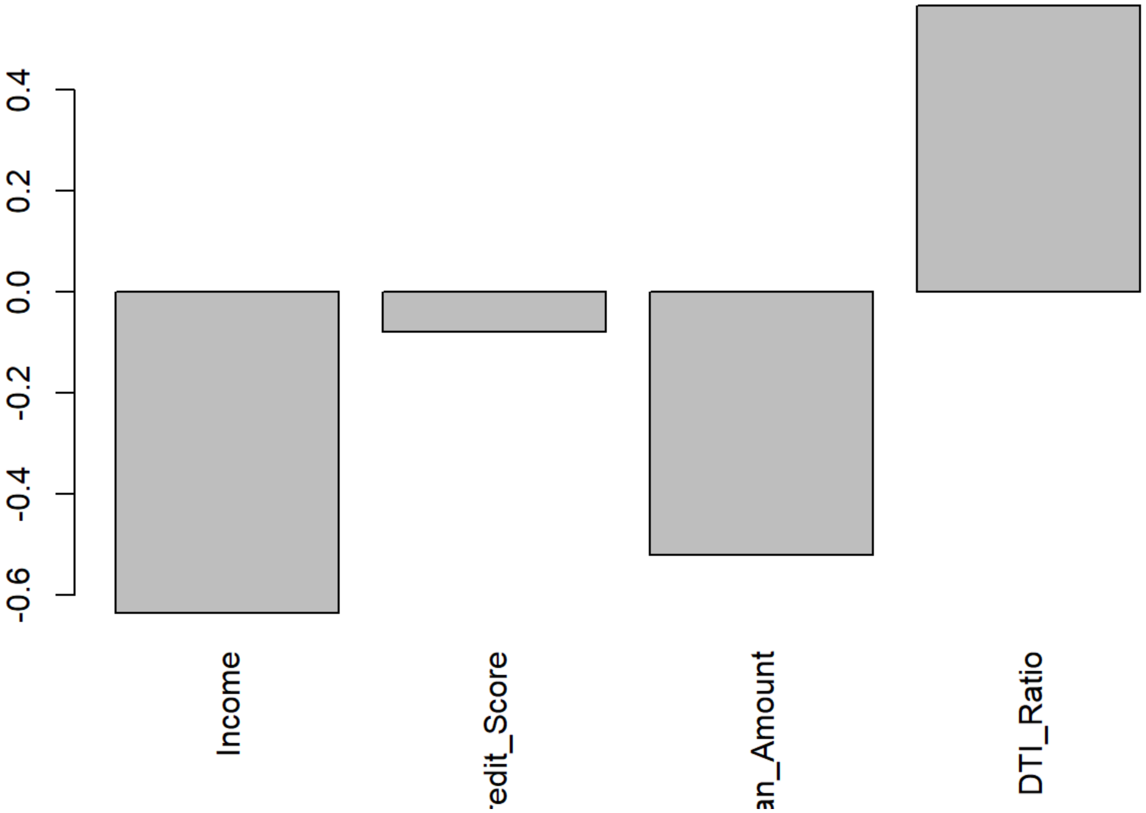
```
par(mfrow = c(1,1))
fviz_eig(data_pca, addlabels = TRUE)
```



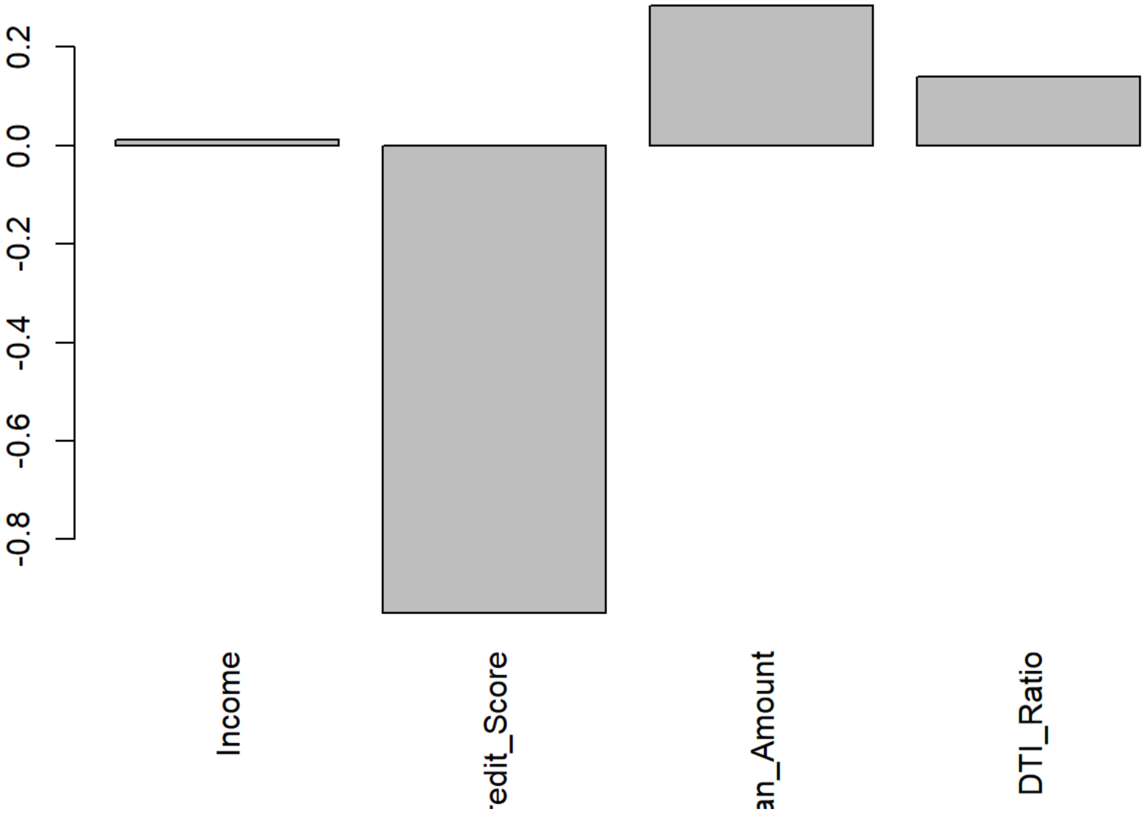
```
#fviz_pca_var(data_pca, col.var = "black")

#fviz_cos2(data_pca, choice = "var", axes = 1:2)

fviz_pca_var(data_pca, col.var = "cos2",
             gradient.cols = c("black", "orange", "green"),
             repel = TRUE)
```



```
par(las = 3)
#barplot(data_pca$ $rotation[,"PC1"])
barplot(prcomp(loan3[,2:5])$rotation[,"PC1"])
```



```
barplot(prcomp(loan3[,2:5])$rotation[,"PC2"])
```



##

Supervised Learning

logistic regression

```
set.seed(8675309)
ind <- sample(1:7866, 7866*.7, replace = F)
train.df <- loan3[ind,]
holdout.df <- loan3[-ind,]
```

```
logmod1 <- glm(formula = Approval ~., family = binomial(link = "logit"), data = train.df[,2:7])
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(logmod1)
```

```
##
## Call:
## glm(formula = Approval ~ ., family = binomial(link = "logit"),
##     data = train.df[, 2:7])
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -23.10740  310.56057  -0.074    0.941
## Income               0.93040    0.09225  10.086   <2e-16 ***
## Credit_Score         2.64796    0.09902  26.740   <2e-16 ***
## Loan_Amount         -1.54380    0.08786 -17.572   <2e-16 ***
## DTI_Ratio           -1.71979    0.11999 -14.332   <2e-16 ***
## Employment_Status1  24.25803  310.56059   0.078    0.938
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7632.8  on 5505  degrees of freedom
## Residual deviance: 2005.7  on 5500  degrees of freedom
## AIC: 2017.7
##
## Number of Fisher Scoring iterations: 19
```

```
step_mod_both <- MASS::stepAIC(
  object = logmod1,
  direction = "both"
)
```

```
## Start:  AIC=2017.68
## Approval ~ Income + Credit_Score + Loan_Amount + DTI_Ratio +
##     Employment_Status
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
##                     Df Deviance    AIC
## <none>                   2005.7 2017.7
## - Income             1   2117.4 2127.4
## - DTI_Ratio          1   2308.2 2318.2
## - Loan_Amount        1   2428.2 2438.2
## - Credit_Score       1   3665.6 3675.6
## - Employment_Status  1   5120.3 5130.3
```

```
step_mod_both
```

```
##
## Call:  glm(formula = Approval ~ Income + Credit_Score + Loan_Amount +
##     DTI_Ratio + Employment_Status, family = binomial(link = "logit"),
##     data = train.df[, 2:7])
##
## Coefficients:
##        (Intercept)              Income        Credit_Score         Loan_Amount
##           -23.1074              0.9304              2.6480             -1.5438
##          DTI_Ratio  Employment_Status1
##            -1.7198             24.2580
##
## Degrees of Freedom: 5505 Total (i.e. Null);  5500 Residual
## Null Deviance:       7633
## Residual Deviance: 2006  AIC: 2018
```

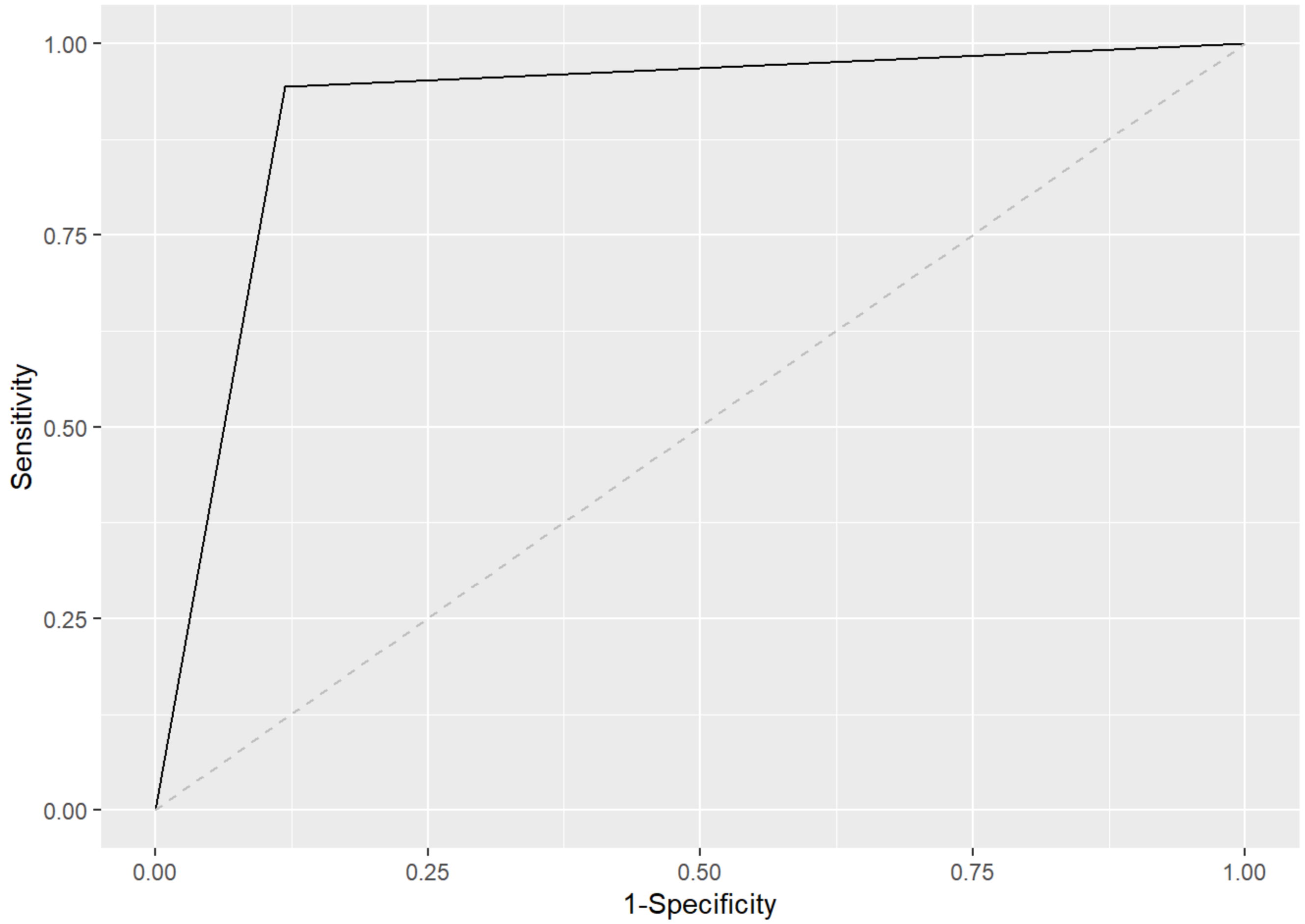
leave all variables in

```
pred <- predict(step_mod_both, holdout.df[,2:7])

prob.predictions <- 1 / (1 + exp(-pred))

caret::confusionMatrix(factor(ifelse(prob.predictions > .5, 1, 0)), factor(holdout.df$Approval
))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 1050   65
##          1  142 1103
##                                           
##                Accuracy : 0.9123          
##                  95% CI : (0.9001, 0.9234)
##     No Information Rate : 0.5051          
##     P-Value [Acc > NIR] : < 2.2e-16       
##                                           
##                   Kappa : 0.8247          
##                                           
##  Mcnemar's Test P-Value : 1.275e-07       
##                                           
##             Sensitivity : 0.8809          
##             Specificity : 0.9443          
##          Pos Pred Value : 0.9417          
##          Neg Pred Value : 0.8859          
##              Prevalence : 0.5051          
##          Detection Rate : 0.4449          
##    Detection Prevalence : 0.4725          
##       Balanced Accuracy : 0.9126          
##                                           
##        'Positive' Class : 0               
## 
```

```r
# ROC curve
library(ROCR)
predob <- prediction(ifelse(prob.predictions > .5, 1, 0), holdout.df$Approval )
perf <- performance(predob, "tpr", "fpr")
perf.df <- data.frame(tpr = perf@x.values[[1]],
                      fpr = perf@y.values[[1]])
ggplot2::ggplot(perf.df, aes(x = tpr, y = fpr))+
  geom_line()+
  geom_segment(aes(x=0, y=0, xend=1, yend=1), color = "gray", linetype = "dashed")+
  labs(x = "1-Specificity", y = "Sensitivity")
```

```
## Warning in geom_segment(aes(x = 0, y = 0, xend = 1, yend = 1), color = "gray", : All aestheti
cs have length 1, but the data has 3 rows.
## ℹ Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```



```
performance(predob, measure = "auc")@y.values[[1]]
```

```
## [1] 0.9126109
```

random forest

```
library(randomForest)
```

```
## randomForest 4.7-1.2
```

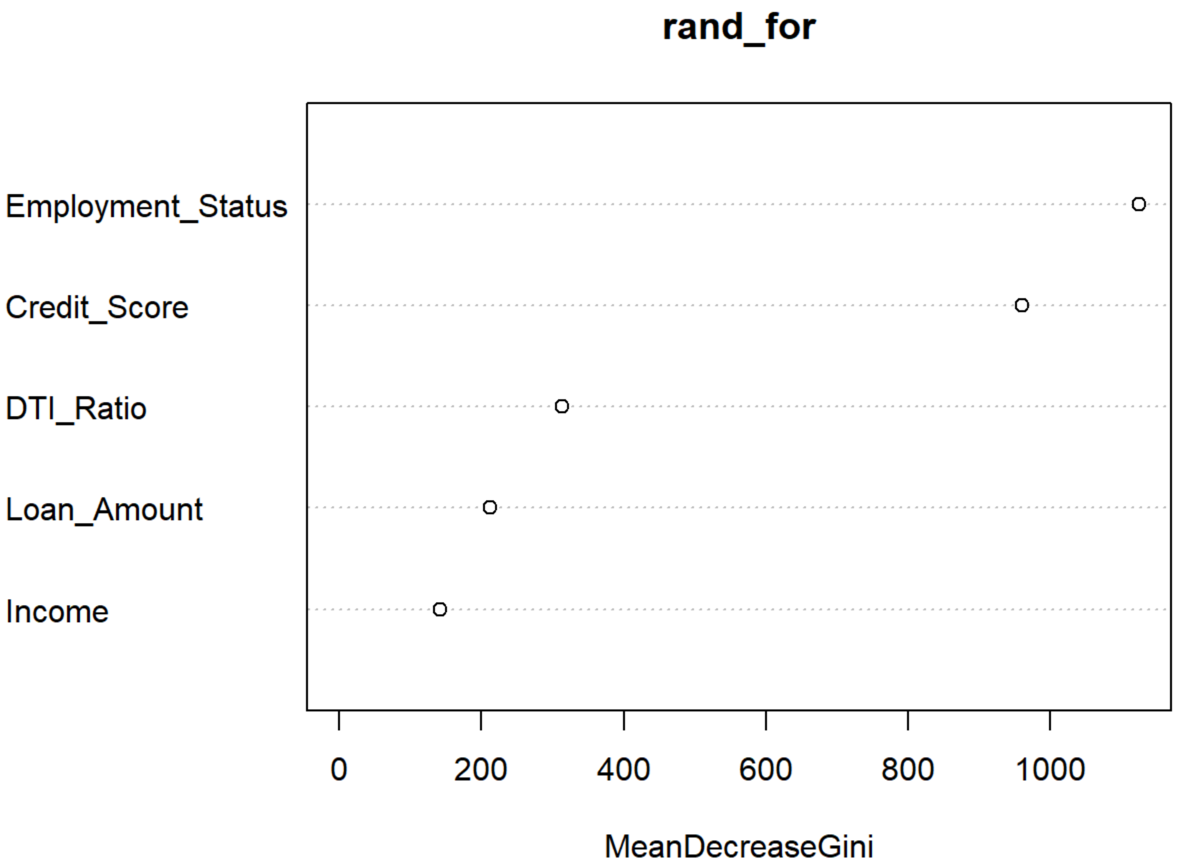
```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
rand_for <- randomForest(Approval ~ ., data = train.df[,2:7])
```

```
varImpPlot(rand_for)
```



```
rf.pred <- predict(rand_for, holdout.df[,2:7], type = "class")
caret::confusionMatrix(as.factor(rf.pred), as.factor(holdout.df$Approval))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 1175    4
##          1   17 1164
##                                           
##                Accuracy : 0.9911          
##                  95% CI : (0.9864, 0.9945)
##     No Information Rate : 0.5051          
##     P-Value [Acc > NIR] : < 2.2e-16       
##                                           
##                   Kappa : 0.9822          
##                                           
##  Mcnemar's Test P-Value : 0.008829        
##                                           
##             Sensitivity : 0.9857          
##             Specificity : 0.9966          
##          Pos Pred Value : 0.9966          
##          Neg Pred Value : 0.9856          
##              Prevalence : 0.5051          
##          Detection Rate : 0.4979          
##    Detection Prevalence : 0.4996          
##       Balanced Accuracy : 0.9912          
##                                           
##        'Positive' Class : 0               
## 
```

```
rownames(holdout.df) <- 1:nrow(holdout.df)
```

```
#prob_bt2 <- predict(rand_for, holdout.df[,2:7])
#predob2 <- prediction(prob_bt2, holdout.df$Approval )
#perf2 <- performance(predob2, "tpr", "fpr")
#perf.df2 <- data.frame(tpr = perf2@x.values[[1]],
#                      fpr = perf2@y.values[[1]])
```

Unsupervised Learning

bayesian

```
library(Rgraphviz)
```

```
## Loading required package: graph
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: generics
```

```
##
## Attaching package: 'generics'
```

```
## The following objects are masked from 'package:base':
##
##     as.difftime, as.factor, as.ordered, intersect, is.element, setdiff,
##     setequal, union
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following object is masked from 'package:randomForest':
##
##     combine
```

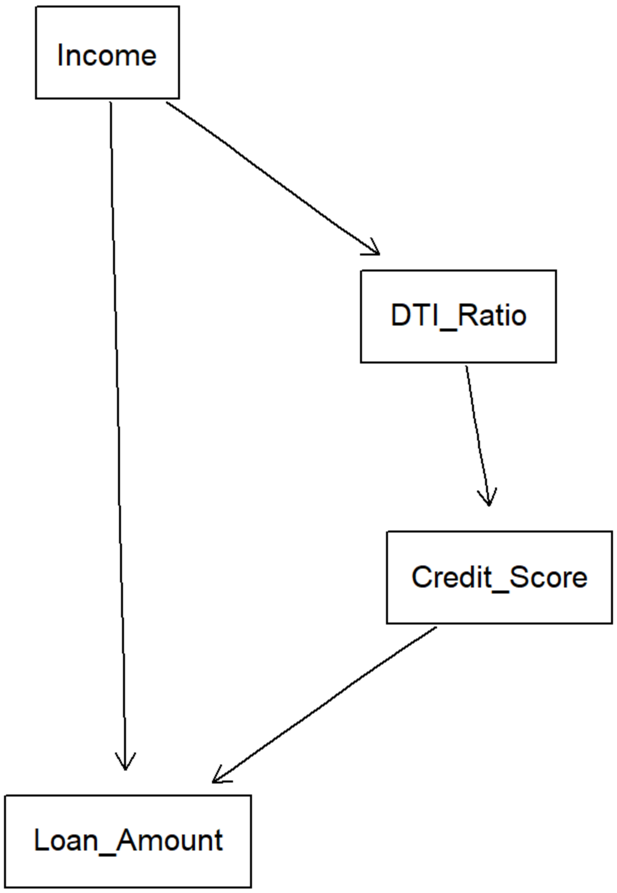
```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, is.unsorted, lapply, Map, mapply, match, mget,
##     order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##     rbind, Reduce, rownames, sapply, saveRDS, table, tapply, unique,
##     unsplit, which.max, which.min
```

```
## Loading required package: grid
```

```
library(bnlearn)
model_hc <- bnlearn::hc(loan3[,2:5])
```

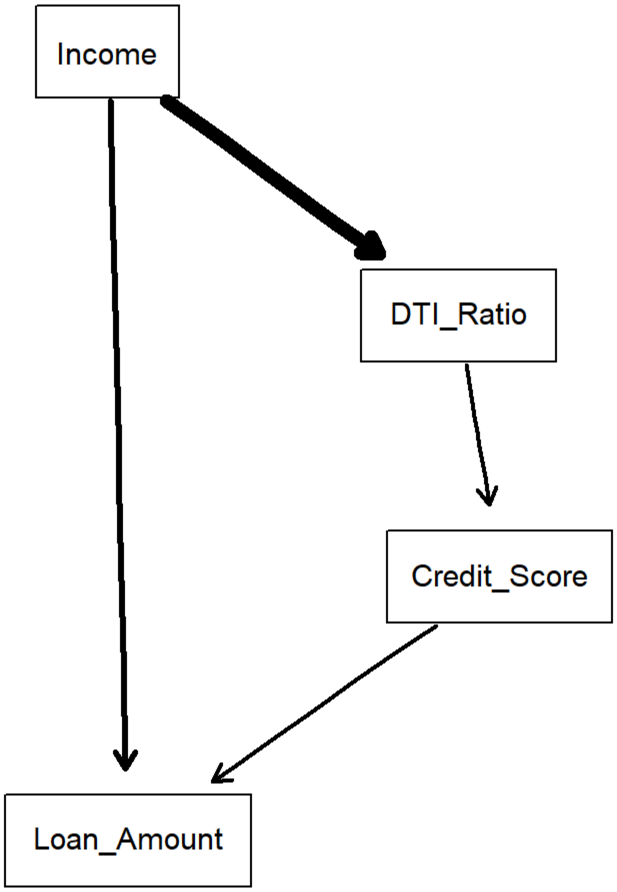
```
graphviz.plot(model_hc)
```



```
score(x = model_hc, data = loan_data[,2:5], type = 'bic-g')
```

```
## [1] -840875
```

```
#graphviz.plot(model_hc)
strength_loan <- arc.strength(
x = model_hc,
data = loan3[,2:5])
strength.plot(x = model_hc, strength = strength_loan)
```



```
bn_loan <- bn.fit(
x = model_hc,
data = loan3[,2:5]
)
```

```
bn_loan_pred <- predict(
object = bn_loan,
data = loan3[,2:5],
node = colnames(loan3)[4])
```

```
colnames(loan3)[4]
```

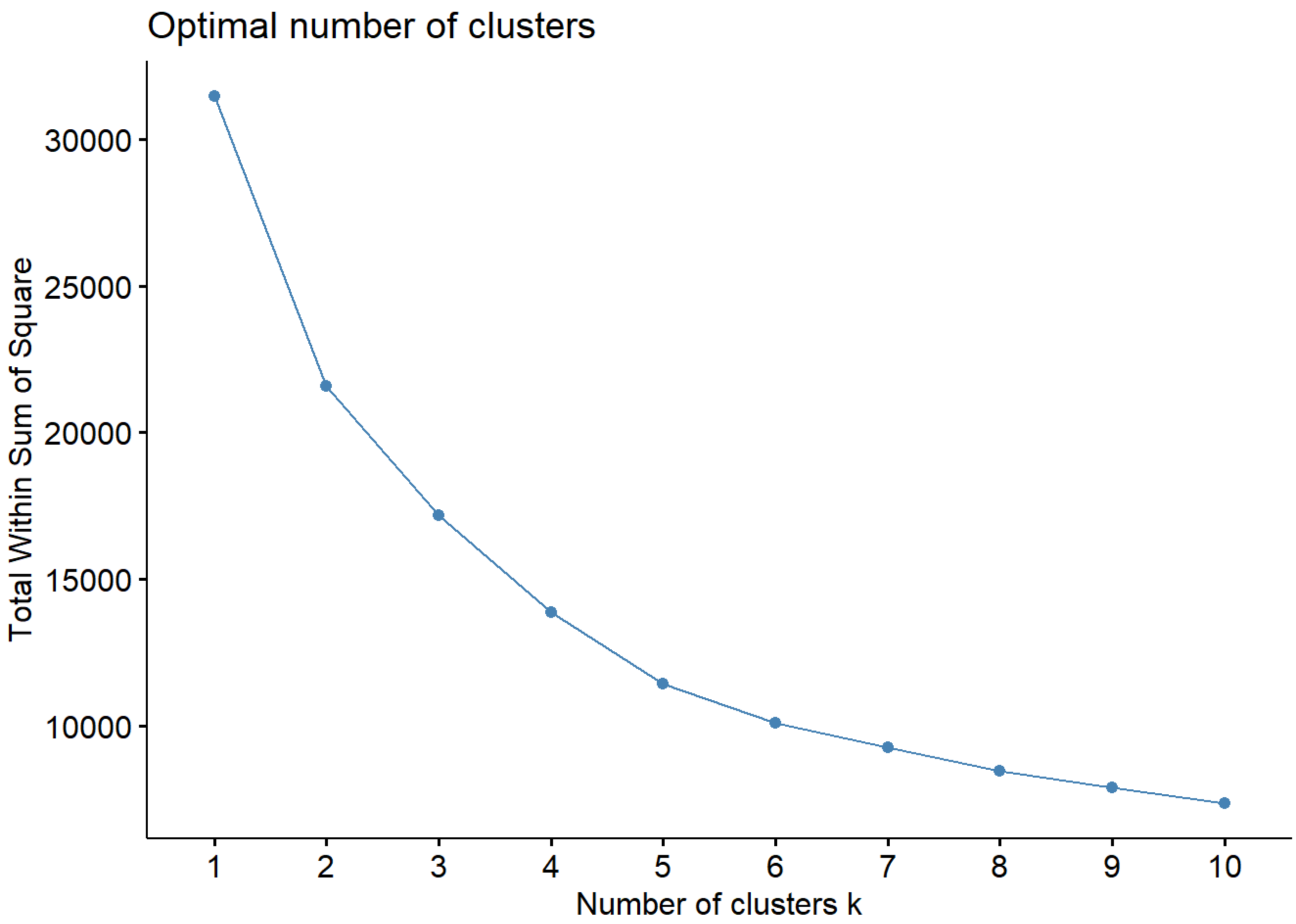
```
## [1] "Loan_Amount"
```

```
mean((bn_loan_pred - loan3$Loan_Amount )^2)
```

```
## [1] 0.7102918
```

clustering

```
library(cluster)
library(factoextra)

#create plot of number of clusters vs total within sum of squares
fviz_nbclust(loan3[,2:5], kmeans, method = "wss")
```



```
kmeans_loan <- kmeans(loan3[,2:5], centers = 2)

kmeans_loan
```

```
## K-means clustering with 2 clusters of sizes 3334, 4532
## 
## Cluster means:
##       Income Credit_Score Loan_Amount  DTI_Ratio
## 1 -0.9507394  -0.08296852  -0.6459580  0.6141009
## 2  0.6994186   0.06103642   0.4752039 -0.4517680
## 
## Clustering vector:
##    [1] 1 2 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 2 1 1 1 2 2 2 1 2 1 1 1 2 2 1 2 2 1 1 1
##   [38] 1 1 2 1 1 2 2 2 1 1 1 1 1 1 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 1 1 1 2 1 1 2 1
##   [75] 2 1 2 2 1 2 1 1 2 2 2 1 1 2 1 2 1 2 1 2 2 2 2 2 2 1 2 1 1 2 2 2 1 1 2 2 2
##  [112] 2 2 1 2 2 2 2 1 2 2 2 2 1 1 2 2 2 1 2 2 1 2 2 1 2 2 1 1 2 1 2 2 1 1 1 1 1
##  [149] 1 1 2 1 1 2 2 1 2 1 1 2 2 1 2 2 1 1 1 2 2 1 1 1 2 2 1 1 2 2 1 1 1 1 1 2 1
##  [186] 2 1 2 1 1 2 1 2 1 1 2 1 1 1 1 2 2 1 2 1 2 1 1 2 1 2 2 1 1 2 1 1 2 1 2 2 1
##  [223] 2 2 1 1 2 2 1 1 1 1 1 1 2 1 2 2 2 1 1 2 2 2 1 1 2 1 2 1 1 1 1 1 1 2 2 2 1
##  [260] 2 1 2 1 1 2 1 2 2 1 1 1 1 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 1 2 1 2 2 2 1 1 1
##  [297] 2 1 1 1 2 1 1 2 1 1 2 2 1 1 2 2 2 1 1 2 2 1 2 1 1 2 1 1 2 1 2 2 2 1 1 1 1
##  [334] 1 2 2 2 1 2 1 2 2 1 2 2 2 1 2 2 1 1 2 1 1 2 1 1 1 2 1 2 1 2 2 1 1 2 1 2 1
##  [371] 1 2 1 2 1 1 2 1 2 2 1 1 1 2 1 2 2 1 2 1 2 1 2 1 2 2 2 1 1 1 2 2 2 2 1 2 2
##  [408] 1 2 1 2 1 1 2 2 2 2 1 1 1 2 1 2 2 1 1 2 2 2 2 2 2 1 2 2 2 1 2 1 1 1 2 2 2
##  [445] 2 2 1 1 1 2 1 2 1 2 1 1 1 2 1 1 2 2 1 2 2 1 2 2 1 1 2 2 1 2 2 1 2 1 1 2 1
##  [482] 1 1 2 1 1 2 2 2 1 2 1 1 2 2 1 1 2 1 2 2 2 1 1 2 1 2 2 2 1 1 2 2 1 1 1 2 2
##  [519] 1 1 1 2 1 2 1 1 1 2 2 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1
##  [556] 1 2 1 1 1 1 1 1 1 1 2 2 2 2 2 2 1 2 1 2 1 1 1 2 2 2 2 1 2 2 1 1 1 2 1 2 1
##  [593] 2 2 2 2 2 1 1 2 2 1 2 1 1 1 2 1 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 1 1 2
##  [630] 2 1 2 1 1 2 1 2 2 2 1 1 2 2 2 1 2 2 1 2 2 2 1 1 2 1 1 2 1 2 2 1 1 2 1 1 2
##  [667] 2 1 1 1 2 2 1 2 1 1 1 2 1 2 2 2 2 1 2 2 2 1 1 2 2 1 2 1 1 1 2 2 2 1 2 2 1
##  [704] 1 1 2 1 1 2 1 1 1 1 2 1 2 1 2 2 2 1 1 2 1 2 2 1 1 1 1 2 1 1 1 1 1 1 2 1 1
##  [741] 2 1 2 1 1 1 2 1 2 2 2 1 2 2 2 1 2 2 1 1 2 1 1 1 1 2 1 1 1 2 2 2 2 1 1 2 1
##  [778] 1 2 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 2 1 2 2 2 1 1 1 1 2 1 2 1 1 1 2 2
##  [815] 1 2 1 2 1 1 2 1 1 1 1 2 2 2 1 1 1 2 2 2 2 2 1 1 2 1 2 2 2 2 1 2 2 2 1 2 2
##  [852] 1 2 1 1 2 2 1 2 1 2 1 1 2 1 1 1 2 2 2 1 2 1 2 2 2 1 2 2 2 1 1 1 1 2 1 2 1
##  [889] 2 2 2 1 2 2 2 2 1 2 2 1 1 1 1 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 1 1 2 2 2 2 2
##  [926] 2 1 1 1 1 1 2 2 1 1 1 2 2 1 2 1 1 1 1 1 2 1 1 1 2 1 2 1 2 2 1 2 2 1 1 1 2
##  [963] 1 2 2 1 2 1 2 1 2 2 1 2 2 1 2 1 1 1 2 1 1 2 2 2 1 1 1 2 2 1 1 1 1 2 2 2 2
## [1000] 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1 1 2 2 2 2 1 1 2 2 2 1 1 1 1 2 1 1 1 1 2 2 2
## [1037] 2 2 2 2 1 1 1 2 2 2 1 1 2 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 2 2 1 1 2 2 2 2
## [1074] 2 2 1 1 1 2 1 1 2 2 2 1 2 1 1 1 1 2 2 1 2 2 2 1 1 1 2 2 1 1 1 2 1 2 2 1 2
## [1111] 2 2 2 1 2 2 2 2 1 2 2 2 1 2 1 1 2 1 2 2 1 1 2 1 2 2 1 1 2 1 1 1 2 2 1 2 1
## [1148] 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 1 2 1 2 2 2 1 1 2 2 1 2 2 1 1 2
## [1185] 1 1 2 2 1 1 1 1 1 1 2 2 1 1 2 1 2 1 1 1 2 2 1 2 1 1 1 1 1 1 1 2 2 2 1 2 1
## [1222] 1 2 2 1 2 1 1 2 2 1 2 1 2 1 1 2 2 2 1 1 1 2 2 2 2 1 2 2 1 2 1 1 1 1 1 1 1
## [1259] 2 2 1 1 1 1 2 2 2 1 2 2 2 2 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 1 2 1 1 1 1 2
## [1296] 2 1 2 1 1 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 1 1 1 2 1 2 1 2 1 2 1 2 1 1 2 2
## [1333] 1 1 1 2 1 1 2 2 1 1 1 2 1 2 1 2 1 1 1 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2
## [1370] 1 1 2 2 1 1 1 1 2 2 2 2 1 2 1 2 1 2 1 1 2 2 1 1 2 2 1 1 2 1 2 2 2 1 2 2 2
## [1407] 2 2 1 1 2 2 1 1 1 2 2 2 2 2 1 1 1 1 1 1 2 1 1 2 1 2 2 1 1 2 2 1 2 2 1 2 1
## [1444] 1 1 2 1 1 2 1 2 1 1 1 2 1 2 2 1 2 2 1 1 2 2 2 2 2 1 2 1 1 2 1 1 1 2 1 2 2
## [1481] 1 1 1 1 2 2 1 2 2 1 2 2 1 2 1 1 2 1 2 1 2 1 2 1 1 1 1 2 2 1 1 1 2 1 2 2 2
## [1518] 2 2 2 1 1 1 1 2 1 1 2 2 2 1 1 1 1 2 1 2 1 2 2 1 2 2 2 2 2 1 1 1 1 2 1 2 2
## [1555] 2 2 1 2 1 1 1 1 1 2 2 2 1 2 1 1 2 1 1 2 2 2 2 2 1 2 1 2 2 1 2 1 1 1 1 1 1
## [1592] 2 2 2 2 1 2 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 2 1 2 2 2 2 2 2 2 1 1 1 1 2 2 2
## [1629] 1 1 2 2 1 1 1 1 2 2 2 2 2 1 2 2 1 2 1 2 1 1 1 1 2 2 2 2 2 1 2 2 2 2 1 1 1
## [1666] 2 1 2 1 2 1 2 1 1 2 1 2 2 1 2 1 1 2 1 1 1 1 1 1 2 2 1 1 1 2 2 2 2 2 1 2 2
## [1703] 1 2 1 1 1 1 2 1 1 1 2 1 2 2 1 2 1 1 1 2 1 2 1 2 1 1 2 1 2 1 2 2 1 2 2 1 2
## [1740] 2 1 1 2 1 1 1 2 2 1 1 1 2 1 2 1 1 2 2 1 2 1 1 2 2 1 2 1 1 1 1 2 2 1 2 2 1
## [1777] 2 1 1 2 2 2 1 2 1 1 1 1 1 2 2 2 2 1 2 2 1 2 2 1 2 1 1 2 2 1 1 1 1 2 1 1 2
## [1814] 2 1 2 1 2 2 2 2 1 2 1 1 1 1 1 2 1 2 1 2 2 2 1 1 1 2 2 2 2 2 2 2 1 2 2 1 2
## [1851] 1 2 2 1 2 2 2 2 1 1 1 1 1 2 1 1 2 1 2 1 2 1 2 2 1 1 2 1 2 2 2 2 2 1 2 1 2
## [1888] 2 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 1 2 1
## [1925] 2 1 1 2 1 1 1 1 2 2 2 2 1 2 1 2 2 1 1 1 2 1 2 2 2 2 2 1 2 1 2 2 2 2 2 1 1
## [1962] 2 1 2 2 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 2 2 1 2 2 1 1 1 1 1 2 1 2 2 2 1 2 2
## [1999] 2 2 2 1 1 2 1 1 1 2 2 2 2 1 1 1 1 2 1 2 2 2 1 1 1 2 2 2 2 1 1 1 1 2 2 2 1
## [2036] 2 2 1 2 1 2 2 2 1 2 1 2 1 1 2 1 2 2 2 1 1 1 1 1 1 2 1 1 1 2 1 2 1 2 1 1 1
## [2073] 1 2 1 2 1 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 1 2 1 2 1
## [2110] 1 2 2 1 1 1 2 1 1 1 1 1 2 1 1 2 2 2 1 2 2 1 2 1 2 2 1 1 2 2 1 1 2 2 2 2 2
## [2147] 2 1 2 1 2 2 2 2 2 2 2 2 1 1 1 2 1 2 1 2 1 1 1 1 2 2 2 2 1 2 1 2 2 2 2 2 1
## [2184] 1 2 1 1 1 1 2 2 1 1 2 1 1 2 1 1 1 1 2 1 1 2 2 2 1 1 2 2 1 2 1 2 2 1 2 2 1
## [2221] 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 1 1 1 2 1 2 2 2 2 1 2 1 1 1 2 2 1 1 2 1
## [2258] 2 1 2 2 2 1 1 1 2 1 2 1 2 2 2 1 1 1 1 1 2 2 1 1 2 2 1 1 1 2 2 2 1 2 1 1 1
## [2295] 2 1 1 2 1 2 1 1 2 1 1 2 1 1 2 1 2 2 1 2 1 2 1 1 1 1 2 1 1 2 1 1 2 2 1 1 1
## [2332] 2 2 2 1 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 1 2 1 2 2 2 1 2 1 1 1
## [2369] 2 2 1 1 1 1 2 2 1 1 1 2 1 2 1 1 1 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 1 2 2 2 2
## [2406] 2 2 1 1 1 2 2 1 2 2 2 1 1 2 1 2 2 1 1 2 1 1 2 2 2 2 1 2 2 1 2 1 2 1 2 2 2
## [2443] 2 1 1 1 2 2 2 1 1 2 1 2 1 2 1 1 1 1 2 1 2 1 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2
## [2480] 1 1 1 2 1 1 1 2 2 2 2 2 2 1 2 1 2 2 1 2 1 1 2 1 2 1 2 1 2 1 2 1 2 1 2 2 2
## [2517] 1 2 1 2 2 2 2 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 2 1 2 2 2 1 2 2 2 2 1 2 2 1
## [2554] 2 1 2 1 1 1 2 1 2 1 2 2 2 2 2 2 1 1 2 1 2 1 2 2 2 2 2 2 1 2 2 2 2 1 2 1 1
## [2591] 1 1 1 2 2 1 2 1 2 1 1 1 1 2 1 1 1 1 2 1 2 2 2 1 1 2 1 2 1 2 1 1 2 1 1 2 2
## [2628] 1 1 2 1 2 1 1 1 2 2 1 2 2 1 1 1 1 2 1 1 1 2 1 1 2 2 1 1 2 1 1 1 2 1 1 1 1
## [2665] 1 1 2 1 1 1 1 2 1 2 2 2 2 2 2 2 2 1 1 1 1 2 2 1 1 1 1 2 2 2 2 2 2 1 2 2 2
## [2702] 2 2 1 1 1 1 2 2 1 1 1 2 2 1 2 1 1 1 2 2 2 2 1 2 2 2 2 2 1 1 1 2 1 2 2 2 1
## [2739] 2 2 1 1 2 2 2 1 2 2 2 2 2 2 1 1 1 2 1 1 2 2 2 2 1 1 2 2 1 2 1 1 1 1 2 1 2
## [2776] 2 1 2 1 1 1 2 1 2 1 1 2 1 1 2 2 1 2 2 2 2 1 2 1 2 1 1 1 2 2 1 2 1 1 2 1 2
## [2813] 1 1 1 2 2 1 2 1 1 2 2 2 1 2 2 1 2 2 1 1 1 2 1 2 1 2 1 2 2 2 1 1 2 2 2 2 2
## [2850] 1 1 2 2 2 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1 2 2 2 2 1 2 1 2 2 2 1 1 2 1 2
## [2887] 1 2 1 2 2 2 1 1 1 2 1 2 2 2 2 1 1 1 2 2 1 2 2 1 2 2 2 2 2 1 1 2 2 1 1 2 2
## [2924] 1 2 2 1 1 2 1 1 1 1 1 1 2 2 2 2 1 2 1 2 1 1 1 1 2 2 2 1 2 1 1 2 1 2 1 1 2
## [2961] 1 2 2 2 1 2 2 2 2 1 1 2 1 1 2 2 2 1 1 2 2 1 1 1 1 2 2 2 1 1 2 1 1 1 1 2 2
## [2998] 2 2 2 2 1 1 2 2 2 2 2 2 2 1 1 1 1 1 2 2 2 2 1 1 2 1 1 1 2 2 1 2 1 1 2 2 2
## [3035] 2 1 1 2 2 2 1 2 2 1 1 2 2 1 2 1 1 2 1 2 2 1 1 1 2 2 2 2 1 1 2 1 1 1 1 2 1
## [3072] 1 1 2 2 1 2 1 2 2 2 1 2 2 2 2 2 1 2 2 2 1 1 2 2 1 2 1 1 2 2 1 2 2 1 2 2 1
## [3109] 2 1 2 2 1 1 1 2 1 1 1 1 2 2 1 1 2 2 1 1 1 1 2 2 1 2 1 1 1 2 2 2 1 2 2 1 2
## [3146] 1 2 2 1 2 2 2 2 2 1 1 1 1 2 1 2 1 2 2 1 2 2 1 1 1 1 2 1 1 2 2 2 2 1 2 2 1
## [3183] 2 1 1 1 1 1 1 1 1 2 1 2 2 2 1 2 1 1 1 2 1 1 2 1 1 1 2 2 1 2 2 2 2 2 2 2 1
## [3220] 1 2 2 1 2 1 1 2 1 2 1 2 1 1 2 2 1 1 1 2 1 1 2 2 2 1 1 2 1 1 2 1 1 1 1 2 1
## [3257] 2 1 2 1 2 2 1 1 1 1 1 1 1 2 2 1 1 2 1 1 2 2 2 1 1 2 1 1 1 1 2 2 1 2 1 2 2
## [3294] 2 1 2 2 2 2 1 1 1 2 1 1 2 2 1 1 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1 1 1 1 2
## [3331] 2 2 2 2 2 2 1 2 2 2 2 1 2 2 1 1 2 1 1 1 1 2 1 1 1 2 2 2 2 2 1 1 1 1 2 2 1
## [3368] 1 1 1 2 2 2 1 2 2 2 1 2 1 2 1 2 2 1 1 2 1 2 2 1 1 1 2 1 1 2 1 2 1 1 2 1 2
## [3405] 1 2 2 2 2 2 2 2 1 2 1 1 2 1 1 1 2 1 2 2 1 1 2 2 2 1 2 2 1 2 2 1 2 1 1 1 2
## [3442] 1 2 1 1 2 1 2 1 2 1 2 1 2 2 1 2 2 1 1 1 2 1 1 2 2 2 2 2 2 2 1 2 2 1 1 2 1
## [3479] 2 2 2 2 2 2 2 2 1 2 2 1 1 2 1 1 1 2 2 2 1 1 2 2 1 1 2 1 1 1 2 1 2 1 2 2 2
## [3516] 1 2 1 1 2 1 2 2 1 1 2 2 1 1 1 2 1 1 2 1 1 2 1 2 2 2 1 1 1 2 2 1 2 1 1 2 1
## [3553] 2 2 1 1 1 1 1 2 1 1 2 1 1 2 2 2 2 2 1 1 2 1 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2
## [3590] 1 2 2 2 2 1 1 1 1 2 2 2 2 1 2 1 2 1 1 2 1 1 2 2 2 1 1 2 2 1 2 1 2 2 1 1 1
## [3627] 1 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 1 1 2 2 1 1 2 1 2 1 1 2 1 2 2 2
## [3664] 2 2 2 2 2 2 1 2 1 1 2 1 1 2 2 1 2 1 2 1 1 2 1 1 1 2 1 2 2 1 2 1 1 2 1 2 2
## [3701] 1 2 2 2 1 1 1 1 1 2 2 2 1 2 1 1 2 2 1 2 2 2 1 1 2 2 1 2 1 1 2 2 2 2 2 1 2
## [3738] 1 2 1 1 2 2 2 1 2 2 1 1 2 1 2 1 2 2 1 2 1 2 2 2 2 1 2 2 2 1 1 1 1 2 1 1 2
## [3775] 1 2 2 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 2 1 1 2 2 1 2 2 1 2 1 2 1 2 1 1
## [3812] 2 1 1 1 1 2 1 2 2 1 2 1 1 2 1 2 2 2 2 1 2 1 1 1 2 1 1 2 2 2 1 2 1 1 2 2 2
## [3849] 1 1 1 2 1 2 1 2 2 2 2 1 1 1 2 1 2 2 2 2 1 2 2 1 2 1 1 2 2 2 1 1 1 2 2 2 1
## [3886] 1 2 2 2 1 2 1 1 1 2 1 2 2 1 1 1 2 1 1 2 2 1 2 2 2 1 1 2 2 1 1 2 2 2 1 2 1
## [3923] 2 1 2 1 2 1 1 2 1 2 1 1 2 2 2 2 1 1 2 1 2 1 2 2 2 1 1 2 1 2 1 1 1 2 2 2 2
## [3960] 1 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 1 2 2 2 1
## [3997] 2 2 2 2 2 2 2 1 2 1 1 1 2 1 2 1 2 1 1 2 2 1 2 1 1 2 2 1 1 2 1 2 2 2 2 2 2
## [4034] 2 1 1 2 2 2 2 2 1 2 2 1 1 1 2 2 2 2 2 2 1 1 2 2 2 1 2 2 2 2 1 1 2 1 2 2 2
## [4071] 1 2 2 2 1 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 1 1 1 2 2 1 2 2 1 2 1 2 1 2 2 2 2
## [4108] 2 2 1 1 2 2 1 1 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 1 1 1 1 2 2 2 2 2 1 1 1 1 1
## [4145] 2 1 2 2 2 2 1 1 2 2 2 1 2 1 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2
## [4182] 1 1 1 2 1 2 2 1 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 2 1 1 2 1 2 1 2 2 2 2
## [4219] 1 1 1 2 1 1 1 2 2 2 2 2 1 2 2 1 2 1 2 1 2 1 2 1 2 2 1 1 1 1 2 2 1 2 2 2 2
## [4256] 2 2 1 2 2 2 2 2 2 1 1 2 1 2 1 2 1 1 1 2 2 1 2 1 2 1 1 1 2 1 2 2 2 2 2 1 2
## [4293] 2 2 2 2 1 1 1 2 2 2 1 1 2 1 2 2 2 2 1 2 2 1 1 2 1 2 1 1 2 1 1 2 2 2 2 2 1
## [4330] 1 2 1 1 2 2 1 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 1 2 2 2 1 2 2 1 1 2 2 2 2 2
## [4367] 2 1 1 2 2 2 1 1 2 2 2 2 2 2 2 2 1 1 1 1 2 2 1 1 2 1 2 2 2 2 2 2 1 2 2 1 2
## [4404] 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 1 2 1 2 2 2 1 2 1 1 2 2 1 1 1 2 2 2 2 2 2 2
## [4441] 2 2 1 2 2 1 2 1 1 1 2 2 2 2 1 1 1 2 1 1 2 2 1 2 2 1 2 1 1 2 1 2 1 2 1 2 1
## [4478] 2 2 1 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 1 2 2
## [4515] 2 1 1 1 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 1 1 2 1 2 2 2 2 1 1 1 2 1 1 1 1 2 2
## [4552] 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1
## [4589] 1 2 2 1 2 2 2 2 1 2 2 2 2 2 1 1 1 2 1 1 2 1 2 1 2 1 2 2 1 2 1 1 2 1 2 2 2
## [4626] 1 2 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 1 1 2 2 2 1 1 1 1 1 2 1 2 1 1 2 2 1 1 2
## [4663] 2 2 2 1 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 2 2 1 2 2 1 1 2 2 2 2 1 1 1 2 2
## [4700] 2 2 1 2 1 1 2 2 1 2 2 1 1 1 2 1 1 2 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2
## [4737] 2 2 2 2 2 2 2 1 1 2 2 1 1 1 1 2 1 2 2 1 2 2 2 1 2 2 1 2 2 2 1 1 2 2 1 1 2
## [4774] 2 2 1 1 1 2 2 1 2 2 2 1 2 1 2 1 1 1 2 2 1 1 1 2 2 2 1 2 2 2 2 1 2 2 2 2 2
## [4811] 2 1 1 1 1 2 1 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 1 2 2 2 2 1 2 1 2 2 1
## [4848] 1 2 2 1 2 1 2 2 1 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [4885] 2 2 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 1 1 1 2 2 2 2 2 2 1 1
## [4922] 2 1 2 2 1 2 2 2 1 1 2 2 1 1 1 2 2 2 1 2 2 2 1 2 2 2 1 1 1 2 2 2 2 2 2 1 2
## [4959] 1 2 1 2 2 1 2 2 1 2 2 1 2 1 1 1 2 2 1 1 2 2 2 2 2 1 2 2 1 2 2 2 1 1 2 1 2
## [4996] 1 2 2 1 1 2 1 2 1 2 2 2 2 1 1 1 1 2 2 2 2 2 1 1 1 2 2 1 2 2 2 1 1 2 1 2 2
## [5033] 2 1 1 1 2 2 2 2 2 1 2 1 2 1 2 2 1 1 1 2 2 2 2 2 1 2 1 2 2 1 2 2 1 2 1 1 2
## [5070] 2 2 1 2 2 2 2 1 2 2 1 2 1 1 2 2 2 1 2 1 2 1 1 2 2 2 1 2 2 2 1 1 1 2 1 1 1
## [5107] 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 1 2 2 2 2 2 1 2 1 2 2 1 2 1 2 2 2 2 2 2
## [5144] 2 2 2 2 2 1 1 1 2 1 2 2 1 2 2 2 2 2 1 2 2 2 2 1 2 2 1 2 2 2 2 2 1 1 2 2 1
## [5181] 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 1 1 2 1 2 2 2 2 2 1 2 2 1 1 1 1 2
## [5218] 2 2 2 1 2 2 1 2 2 1 2 1 1 1 2 1 1 1 1 2 2 2 2 2 1 2 1 2 2 1 1 1 2 1 2 2 2
## [5255] 2 2 2 1 1 1 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 1 2 1 1
## [5292] 1 1 1 2 2 2 2 1 2 1 1 1 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 1
## [5329] 1 2 2 2 2 1 2 1 1 1 1 1 1 2 1 2 1 2 1 2 2 1 2 1 1 1 1 2 2 1 2 2 1 2 2 1 1
## [5366] 2 2 2 2 1 2 2 1 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 1 2 2 2 2
## [5403] 1 2 1 2 2 2 1 2 2 2 1 2 1 2 2 2 1 1 2 2 2 1 2 2 1 2 2 1 2 2 1 1 2 2 1 2 2
## [5440] 2 1 2 2 1 2 1 2 1 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 1 1 2 1 1 2 2 1 2 2 2 2
## [5477] 1 2 2 1 1 2 2 1 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 1 1 2 2 2 2
## [5514] 1 2 1 2 2 1 1 2 2 2 2 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 2 2
## [5551] 2 1 2 2 2 2 2 1 2 2 2 1 2 1 2 2 2 2 1 2 2 1 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2
## [5588] 2 1 2 2 2 2 2 2 1 2 1 2 2 2 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 1 2 2 2 1 2 2 1
## [5625] 2 2 2 1 1 1 2 2 1 2 1 1 1 2 1 2 2 2 2 2 2 1 1 2 1 1 1 2 2 2 2 2 2 2 1 2 2
## [5662] 1 2 2 2 2 1 1 1 1 2 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 1 1 2 2 1 2 1 2 2 2
## [5699] 1 1 1 2 1 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 1
## [5736] 2 1 1 2 2 2 1 2 2 1 1 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [5773] 2 2 2 2 2 2 1 2 1 2 2 2 1 1 1 2 1 1 2 1 2 2 2 2 1 1 2 1 1 1 1 2 1 1 2 1 2
## [5810] 2 2 2 1 1 2 2 1 2 2 2 2 2 2 1 1 1 2 2 1 2 1 2 2 2 1 2 1 1 1 2 2 1 2 2 2 2
## [5847] 1 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 1 1 2 1 2 1
## [5884] 1 2 2 2 1 2 1 2 1 1 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2
## [5921] 1 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2 1 1 2 2 2 2 1 2 2
## [5958] 2 2 1 1 2 2 1 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 1 1 1 1 2 1 2 2 1 2 2 2
## [5995] 1 2 1 2 2 2 2 2 1 1 2 2 1 2 1 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 2 2 2 2 1 2 2
## [6032] 1 1 2 1 1 1 2 2 2 2 1 2 2 2 1 1 1 2 2 1 2 2 1 1 1 2 2 2 2 1 1 2 2 2 2 2 2
## [6069] 1 1 2 1 2 1 2 2 2 2 2 2 1 1 2 1 2 2 1 2 1 1 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1
## [6106] 2 1 2 2 2 1 2 2 2 1 2 2 1 1 2 2 2 2 2 2 1 1 1 1 1 2 1 1 2 2 1 1 1 2 2 2 2
## [6143] 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 1 1 2 2
## [6180] 2 2 2 1 2 2 1 1 2 2 2 2 2 2 1 1 2 1 2 1 2 2 1 1 1 1 2 1 1 2 1 1 2 2 1 1 2
## [6217] 2 2 1 1 1 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 1 2 1 2 2 1 2 2 2 2
## [6254] 2 1 2 1 2 1 1 2 2 1 2 2 1 1 2 2 2 2 1 1 2 1 1 1 2 2 1 1 2 2 1 1 2 2 1 2 2
## [6291] 2 1 1 1 2 2 2 2 1 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 1 2 2 1 1
## [6328] 1 1 1 2 2 1 2 1 2 2 1 1 2 2 1 2 1 1 1 2 2 1 2 1 2 1 2 2 2 2 2 1 1 2 2 2 2
## [6365] 1 1 2 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 1 2 1 2 2 2 2 1 1 2 2 2 1 1 2 2
## [6402] 1 1 2 1 2 1 1 1 2 2 2 2 2 2 1 2 1 1 1 2 1 2 2 2 2 2 1 2 2 2 1 1 2 2 2 2 2
## [6439] 2 2 2 1 2 2 1 1 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 1 2 1 2 2 2 1 2
## [6476] 2 1 1 2 2 1 1 2 2 2 2 2 2 2 1 1 2 2 1 2 2 2 2 1 2 1 2 2 1 2 2 2 2 1 2 2 2
## [6513] 1 2 2 2 2 2 1 1 2 2 1 2 2 2 1 2 1 2 2 2 1 2 1 2 1 1 2 1 1 2 2 1 2 1 2 2 2
## [6550] 2 1 2 1 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 1 1 1 2 2
## [6587] 2 1 2 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1
## [6624] 2 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 1 2 2 2 1 2
## [6661] 1 1 2 2 1 1 2 1 2 1 2 2 1 2 2 2 1 1 2 2 2 2 2 2 1 2 2 1 1 2 2 2 1 2 2 2 1
## [6698] 1 2 1 2 2 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 2 2 1 2 1 1 2 2 1 1 2 2
## [6735] 1 2 1 1 1 1 2 2 2 1 2 2 2 2 2 2 1 1 2 1 2 2 1 2 2 2 2 2 1 2 2 2 1 1 2 2 2
## [6772] 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 1 2 1 2 1 1
## [6809] 1 2 2 2 1 2 2 1 2 1 2 2 2 1 2 1 1 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2
## [6846] 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 1 1 2 2 2 1 2 2 2 2
## [6883] 1 2 2 1 1 2 1 2 1 2 2 1 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2
## [6920] 1 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 1 2 2 2 1 1 1 2 2 2 1 2 1 2 1 2 2 1 2 1 1
## [6957] 2 2 2 2 1 2 1 2 2 2 2 2 1 2 2 1 2 2 1 1 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 1 1
## [6994] 2 2 2 2 1 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 2 1 2 2 2 1 1
## [7031] 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 2 1
## [7068] 2 2 1 2 2 2 2 2 1 2 1 2 1 1 2 1 1 1 1 1 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 2 1
## [7105] 2 1 2 2 2 1 2 1 2 2 2 1 2 2 1 2 2 2 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 2 2 2 2
## [7142] 2 1 2 1 1 1 1 2 1 2 1 2 1 1 2 1 2 1 2 2 1 2 2 1 1 2 1 1 1 1 2 1 2 2 2 2 2
## [7179] 2 1 2 2 2 2 1 2 2 1 2 1 2 2 2 1 1 1 2 2 1 2 1 1 2 1 2 2 1 1 1 2 2 2 1 2 1
## [7216] 1 1 1 1 1 2 1 2 2 1 2 2 1 2 2 2 2 1 2 1 2 2 2 1 1 1 1 2 2 1 2 2 2 2 1 1 2
## [7253] 2 1 2 1 2 1 1 1 1 2 2 2 1 1 2 2 1 2 2 2 2 2 2 1 2 2 2 1 2 1 1 1 2 2 2 2 2
## [7290] 2 2 2 1 1 2 2 1 1 1 2 2 2 1 2 2 2 2 2 1 2 2 2 2 1 2 2 1 2 1 2 2 2 2 2 2 2
## [7327] 1 2 1 1 2 1 2 1 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 1 2 1 1 2 2 1 2 2 1 2 2 1 2
## [7364] 2 1 1 2 1 1 2 2 1 2 1 2 2 1 2 2 2 1 2 2 2 2 1 1 1 2 1 2 2 2 2 1 2 2 2 2 2
## [7401] 2 1 1 2 2 2 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 1 1 2 2 2
## [7438] 2 2 2 1 2 2 2 1 2 1 2 1 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 1 2
## [7475] 1 2 2 2 1 2 1 1 1 1 2 1 1 2 2 1 1 1 2 1 2 1 1 1 2 1 1 2 2 2 2 2 1 2 2 2 2
## [7512] 1 2 2 2 2 1 1 2 1 2 1 1 2 1 1 2 1 2 1 2 2 1 1 1 1 2 2 2 2 2 1 2 2 2 1 1 2
## [7549] 2 1 1 1 1 2 2 2 1 2 2 2 2 2 1 1 1 2 2 2 2 1 2 1 2 2 2 1 1 2 1 2 1 2 2 1 1
## [7586] 2 1 2 2 2 2 2 2 2 2 1 2 2 1 1 1 2 2 2 2 2 1 1 1 2 1 2 2 2 1 2 1 1 2 1 2 1
## [7623] 1 1 1 2 2 2 2 2 2 1 1 1 2 1 2 1 2 2 2 2 1 1 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2
## [7660] 1 2 2 1 2 1 1 1 2 1 1 2 2 1 1 2 2 2 1 1 2 1 2 2 1 1 2 1 1 1 1 2 2 1 1 1 1
## [7697] 2 2 2 1 2 2 2 1 2 2 1 2 2 2 1 1 1 2 1 2 1 2 2 2 2 1 1 1 2 2 2 2 2 2 1 1 1
## [7734] 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1 2 1 2 2 1 1 2 1 2 2 1
## [7771] 1 2 2 2 2 2 1 1 2 2 1 2 1 2 2 1 1 1 2 2 1 2 2 2 1 2 2 1 2 2 2 1 2 1 1 2 2
## [7808] 2 2 2 2 2 2 1 2 2 1 2 1 2 2 2 2 1 2 1 1 2 2 2 1 2 2 1 2 2 2 1 1 2 2 1 1 2
## [7845] 2 2 1 1 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2
## 
## Within cluster sum of squares by cluster:
## [1] 10354.48 11238.24
##  (between_SS / total_SS =  31.4 %)
## 
## Available components:
## 
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
kmeans_loan5 <- kmeans(loan3[,2:5], centers = 5)

kmeans_loan5
```

```
## K-means clustering with 5 clusters of sizes 1455, 2370, 450, 2123, 1468
## 
## Cluster means:
##       Income Credit_Score Loan_Amount   DTI_Ratio
## 1 -0.2648569   -1.3358859  -0.2753679 -0.04672331
## 2  0.6494753    0.5640975  -0.1555019 -0.45504568
## 3 -1.6576939   -0.5055340  -0.8761458  3.15643171
## 4 -0.8817393    0.5165137  -0.5765626  0.20934801
## 5  0.9972786   -0.1786549   1.6263676 -0.48937291
## 
## Clustering vector:
##    [1] 1 2 1 1 2 3 3 1 5 3 4 4 4 4 5 1 1 1 1 4 1 5 2 5 1 5 1 3 1 2 5 1 5 5 3 3 3
##   [38] 1 4 5 4 1 5 5 5 3 1 4 3 4 4 1 4 2 5 5 5 5 1 2 5 1 4 2 1 3 3 1 1 5 4 4 1 4
##   [75] 2 1 1 5 4 5 1 3 2 1 1 4 1 2 4 5 4 1 4 1 1 5 5 1 1 3 5 4 1 1 1 1 4 4 2 2 5
##  [112] 5 5 1 5 1 1 1 1 5 5 5 1 3 1 1 2 5 1 1 2 1 1 5 3 1 2 3 4 5 1 5 1 1 3 1 4 1
##  [149] 4 1 5 3 4 5 1 1 5 4 3 1 5 3 5 1 1 3 3 5 1 3 4 1 5 2 3 1 5 1 4 1 3 1 4 5 4
##  [186] 5 1 5 1 4 1 4 5 4 3 5 4 1 4 4 5 1 4 1 3 1 4 3 1 3 1 1 3 4 1 1 3 5 1 5 5 1
##  [223] 1 5 1 3 2 1 3 4 1 4 3 1 5 3 1 5 5 3 4 2 1 5 1 4 5 1 5 3 1 4 3 4 1 2 2 5 3
##  [260] 2 3 5 1 1 5 1 2 5 4 3 3 3 1 2 5 5 5 4 1 1 1 5 2 2 1 5 1 4 1 1 5 2 5 1 4 4
##  [297] 1 3 4 3 1 3 4 5 4 1 1 5 4 3 5 5 4 1 1 5 1 1 1 1 1 5 1 3 5 1 5 2 5 4 4 3 3
##  [334] 4 1 5 5 4 1 4 5 5 1 1 2 5 4 5 1 4 1 5 4 3 5 1 1 4 2 4 5 4 1 5 3 1 5 4 1 4
##  [371] 1 1 4 5 3 1 5 1 5 2 4 1 4 2 4 5 4 4 1 1 5 1 2 3 2 2 5 3 4 4 5 1 5 5 4 5 5
##  [408] 1 1 4 2 3 1 1 5 5 2 1 4 3 5 4 5 5 1 4 2 1 5 5 5 2 1 1 5 1 3 5 1 3 3 5 1 5
##  [445] 1 5 3 4 4 1 4 1 1 1 4 4 4 5 3 1 5 5 4 5 2 1 2 1 1 4 4 1 4 5 1 1 2 4 1 1 1
##  [482] 1 3 5 3 4 5 1 5 4 5 1 1 5 1 3 1 5 4 5 1 2 3 1 5 1 1 1 5 4 1 5 5 4 3 4 5 1
##  [519] 1 4 4 5 4 1 1 1 1 2 5 4 1 1 1 4 4 3 5 2 1 4 3 4 1 1 5 1 4 5 5 4 1 1 1 3 4
##  [556] 1 2 3 1 3 1 1 1 3 4 2 5 5 5 5 5 4 5 3 1 3 1 1 1 1 1 2 4 2 2 4 1 3 5 4 5 1
##  [593] 5 5 5 1 2 4 1 2 1 1 1 3 3 4 5 4 5 1 1 3 2 3 5 5 1 4 5 2 2 1 2 4 2 2 1 3 2
##  [630] 1 1 1 3 4 1 4 2 1 5 4 1 5 2 1 3 1 5 3 2 1 5 4 1 1 3 1 5 1 1 1 3 4 2 3 4 5
##  [667] 2 4 3 3 5 2 1 2 4 4 1 1 4 1 1 1 2 1 1 2 2 4 1 5 1 1 5 4 4 1 5 5 5 1 5 2 4
##  [704] 1 4 1 1 1 5 4 3 4 4 1 1 5 3 1 5 1 3 4 1 1 1 1 4 1 3 3 1 1 4 4 1 4 4 5 4 3
##  [741] 2 3 5 3 3 4 1 4 1 5 1 4 5 2 5 1 5 1 1 3 5 3 1 1 1 1 4 3 3 1 5 5 5 4 1 1 3
##  [778] 4 5 4 1 3 4 1 4 1 5 3 3 5 5 4 3 1 3 4 4 5 3 2 5 2 4 4 3 1 1 4 5 1 1 1 5 2
##  [815] 4 5 3 5 3 1 5 4 4 4 1 5 1 2 4 4 3 1 5 2 2 5 4 4 5 4 2 5 5 1 3 5 5 1 3 2 5
##  [852] 1 1 4 3 5 5 3 5 4 1 1 1 5 4 4 3 5 1 5 4 2 1 2 5 2 4 5 1 1 4 3 3 1 5 3 5 1
##  [889] 2 2 1 1 4 1 5 5 4 2 1 4 1 1 3 5 1 2 1 2 1 2 1 1 2 1 1 5 1 4 3 1 1 5 1 5 5
##  [926] 1 1 1 1 1 4 5 2 3 1 1 1 5 1 5 4 3 4 4 1 5 4 1 3 1 1 2 4 2 1 3 5 2 1 3 4 2
##  [963] 1 2 5 4 5 1 5 4 5 5 4 2 5 4 4 1 4 3 2 1 4 5 2 5 1 4 3 4 1 4 4 1 3 5 5 2 5
## [1000] 1 5 1 4 5 1 5 2 5 5 1 5 2 5 1 3 1 1 2 5 1 3 5 5 5 1 4 4 4 5 1 4 4 1 2 5 5
## [1037] 4 2 5 1 1 1 1 5 5 5 3 3 5 2 3 2 5 5 1 1 3 1 4 1 1 5 1 1 2 5 1 4 1 5 1 5 5
## [1074] 5 5 4 4 1 1 4 1 2 5 1 3 2 1 4 3 3 5 5 3 5 5 5 1 4 1 5 5 1 4 4 2 4 1 5 3 2
## [1111] 5 2 2 3 5 5 5 5 4 2 5 1 4 5 4 1 2 3 1 2 1 4 2 1 1 1 3 1 5 4 1 4 5 5 1 1 1
## [1148] 2 2 2 4 5 1 2 1 1 5 5 3 1 2 4 5 2 2 1 2 1 4 5 1 2 2 5 1 3 2 1 4 5 5 1 3 1
## [1185] 1 1 1 5 2 1 3 3 1 1 2 1 1 1 2 1 2 1 1 3 2 1 1 1 1 1 4 1 1 4 4 5 4 5 1 5 4
## [1222] 3 1 2 1 2 1 4 2 5 1 5 1 5 4 1 5 5 2 3 3 1 2 1 5 5 3 2 2 1 5 4 1 3 1 1 3 4
## [1259] 1 2 4 4 4 3 5 2 2 4 1 5 1 5 4 1 3 4 3 1 4 1 1 4 1 4 1 1 4 4 4 1 1 3 4 3 5
## [1296] 4 4 4 3 4 1 2 2 1 2 4 5 2 2 1 1 1 3 5 2 1 3 1 1 1 1 2 4 5 1 1 3 1 1 4 2 5
## [1333] 4 3 4 5 1 3 4 5 4 1 3 1 4 1 4 5 1 3 4 1 5 5 1 5 1 2 4 5 2 1 2 2 5 5 3 1 5
## [1370] 3 1 2 2 1 4 1 1 2 5 1 1 1 5 1 5 1 5 4 3 2 5 2 4 5 1 1 4 5 4 5 5 5 3 2 5 2
## [1407] 1 4 3 4 2 1 3 1 4 5 2 5 5 5 4 1 3 1 3 4 1 3 1 1 1 2 5 1 1 5 5 4 1 5 4 5 1
## [1444] 4 4 5 4 4 2 4 5 4 1 1 5 4 2 5 1 1 5 1 1 5 2 5 1 2 1 1 1 3 5 1 4 4 2 3 5 1
## [1481] 4 1 3 3 5 1 3 2 2 1 2 1 3 5 3 4 2 4 2 4 2 4 1 1 1 1 4 1 2 1 4 1 1 4 1 1 5
## [1518] 2 5 1 3 4 4 1 2 1 3 2 5 4 1 1 1 3 2 1 1 1 4 5 3 1 2 2 1 2 4 4 1 1 2 1 2 5
## [1555] 5 2 4 5 4 1 4 3 3 1 1 2 3 2 1 1 5 3 4 5 1 5 5 5 3 1 4 1 1 4 1 1 3 1 1 1 1
## [1592] 2 5 1 5 1 5 4 4 2 1 1 1 1 1 1 1 1 4 1 1 3 1 4 5 2 5 5 1 1 5 3 1 1 4 5 5 5
## [1629] 3 1 2 2 4 1 4 1 5 5 1 2 1 1 5 5 3 2 1 1 1 4 4 4 1 2 5 5 5 3 1 1 1 2 4 1 1
## [1666] 2 4 5 4 5 3 1 3 1 2 1 2 2 1 2 4 1 1 1 1 4 4 1 1 5 5 1 4 1 2 5 1 5 2 1 5 5
## [1703] 3 5 1 4 3 3 5 4 1 3 1 1 1 1 4 5 4 4 3 1 4 5 1 1 3 3 1 4 1 1 2 5 3 1 5 3 2
## [1740] 5 1 4 5 1 1 1 2 5 1 4 3 5 1 5 1 1 5 5 3 1 1 3 5 5 1 5 4 3 3 3 5 2 1 5 5 3
## [1777] 5 4 4 1 2 5 1 5 1 1 4 2 4 1 2 1 2 4 5 5 1 5 5 1 1 1 4 2 5 1 1 1 4 5 1 4 1
## [1814] 2 3 1 1 5 2 5 2 4 5 1 1 4 4 1 1 4 2 3 5 5 1 4 4 1 2 5 1 5 2 1 5 4 5 5 1 5
## [1851] 1 1 1 1 1 1 1 2 1 3 1 3 1 1 4 4 1 3 1 3 5 1 2 2 3 1 5 3 5 1 2 1 2 1 2 4 5
## [1888] 2 5 4 1 1 1 3 4 1 2 4 4 1 4 5 1 1 1 3 5 1 2 1 2 2 1 5 4 3 2 4 2 5 5 4 2 3
## [1925] 5 3 4 2 3 3 3 4 5 1 5 1 4 5 3 5 1 1 1 3 1 1 2 5 5 1 2 4 2 1 1 5 1 5 5 1 4
## [1962] 1 1 2 5 3 1 1 5 1 4 4 1 5 4 3 1 1 1 4 2 5 1 5 5 4 4 1 1 1 5 3 5 5 5 1 5 2
## [1999] 5 5 5 4 4 2 1 1 4 1 2 1 2 3 1 1 4 2 4 5 5 5 3 1 1 1 1 2 1 1 4 4 4 1 1 5 4
## [2036] 2 5 4 1 1 2 1 2 3 1 4 5 1 4 5 1 5 2 5 4 1 3 4 4 3 5 4 1 3 5 1 5 4 1 3 3 1
## [2073] 4 5 4 5 3 1 2 1 4 1 1 3 5 2 2 1 2 5 2 5 1 1 4 4 5 5 4 1 2 2 5 2 3 1 4 2 1
## [2110] 3 2 1 4 4 4 2 4 1 3 4 4 5 4 4 2 1 2 1 5 2 1 2 1 5 5 3 1 5 2 4 4 5 5 5 2 5
## [2147] 1 1 5 3 2 1 5 5 1 1 5 2 4 4 4 5 3 2 1 2 1 1 1 4 2 5 1 5 4 5 1 2 1 1 5 5 4
## [2184] 1 1 4 3 4 4 1 1 1 1 1 4 1 1 1 4 1 3 5 1 1 1 1 1 4 4 1 2 4 1 1 1 1 1 1 2 1
## [2221] 4 5 5 5 5 1 1 1 5 3 5 3 5 5 1 1 1 1 3 3 5 1 5 5 4 5 3 5 1 1 4 1 5 3 3 1 1
## [2258] 5 1 1 5 1 4 1 1 5 4 1 3 5 2 5 4 4 4 1 1 2 2 3 1 1 2 1 3 4 1 1 2 4 1 4 4 1
## [2295] 1 3 3 5 4 1 4 1 1 4 1 2 4 4 5 1 5 1 4 5 1 4 4 4 4 1 5 4 1 1 1 4 2 1 1 1 4
## [2332] 2 2 1 4 1 1 1 1 1 2 4 5 5 4 5 5 3 5 2 4 5 5 1 5 5 1 4 2 4 1 5 5 1 2 1 1 4
## [2369] 5 5 4 4 4 4 1 2 3 3 1 1 4 2 4 4 1 1 2 2 5 2 5 5 3 2 4 5 2 5 2 5 3 1 2 5 2
## [2406] 2 5 4 3 1 5 1 1 2 5 5 3 4 2 3 1 5 4 1 5 4 3 1 2 1 2 1 5 1 1 1 2 2 3 5 2 1
## [2443] 5 3 3 3 5 2 5 1 4 2 3 5 3 5 1 1 3 3 1 4 5 2 4 2 1 1 1 1 5 4 1 1 2 1 1 5 5
## [2480] 1 4 4 5 1 1 1 5 5 1 1 5 2 1 5 1 1 5 4 1 1 1 5 3 5 2 5 3 1 4 2 1 2 3 2 5 1
## [2517] 3 5 1 5 1 1 1 3 4 4 4 5 5 4 1 1 1 3 4 1 4 1 4 5 4 1 1 5 1 5 1 5 1 4 5 5 1
## [2554] 4 1 5 3 4 4 5 4 1 3 5 5 5 1 4 5 3 3 5 3 5 4 5 2 5 5 5 1 3 5 5 4 1 3 1 4 1
## [2591] 4 4 4 5 1 4 5 1 1 4 3 3 4 2 1 3 4 1 5 4 5 5 1 4 1 1 4 5 1 1 1 3 4 1 3 5 1
## [2628] 1 1 5 1 2 1 4 3 5 5 3 1 4 4 3 1 3 1 1 1 1 2 4 1 1 1 1 1 2 3 1 4 5 1 1 4 1
## [2665] 2 1 2 3 1 1 4 5 3 5 2 2 1 2 5 1 2 3 1 1 1 5 2 4 4 4 4 5 1 5 1 5 1 4 2 5 5
## [2702] 5 4 1 1 4 1 5 2 4 1 1 1 1 4 1 4 4 4 1 5 5 1 4 5 1 1 2 2 3 4 1 5 4 2 5 2 4
## [2739] 1 5 1 1 5 1 1 3 5 5 5 2 2 5 4 1 1 2 1 1 1 2 1 5 4 1 5 5 3 5 4 3 1 4 1 3 2
## [2776] 5 1 2 3 3 1 2 1 1 1 4 2 4 4 5 5 4 5 5 1 5 1 5 1 2 4 1 3 2 5 4 2 1 4 5 3 5
## [2813] 1 3 1 5 2 4 2 4 4 2 2 2 1 1 5 4 1 2 1 4 1 5 1 2 2 1 1 5 5 2 1 4 2 1 1 1 2
## [2850] 4 4 2 5 5 4 1 3 1 3 1 5 1 4 1 3 4 1 1 4 3 4 4 5 1 5 4 5 4 2 1 1 4 1 1 1 5
## [2887] 1 5 4 2 5 2 4 3 1 2 1 5 2 2 1 4 3 3 5 2 4 5 5 1 5 1 1 1 1 4 1 2 1 1 4 2 5
## [2924] 4 1 5 1 1 5 3 3 4 1 4 3 5 1 2 1 1 5 4 5 3 4 1 1 5 1 2 4 2 4 4 2 1 5 4 1 5
## [2961] 1 1 1 2 1 5 5 2 4 3 1 2 1 3 5 1 5 4 4 5 4 3 4 1 3 5 5 5 1 1 5 1 1 3 1 2 5
## [2998] 2 5 5 5 1 4 2 5 1 5 2 1 1 1 1 1 4 4 5 5 5 2 1 4 5 4 4 1 5 2 4 2 1 3 5 5 1
## [3035] 5 4 1 2 1 5 1 2 1 1 1 2 5 3 1 3 4 1 4 5 5 1 1 1 5 1 5 5 1 1 1 1 1 1 3 5 3
## [3072] 1 1 1 5 3 5 4 2 1 1 3 1 5 1 1 5 3 2 5 1 4 1 1 1 4 1 1 4 5 5 1 1 5 1 2 1 1
## [3109] 2 4 1 5 1 4 3 2 3 1 1 1 5 5 4 1 5 5 4 1 4 1 5 1 1 5 1 4 4 5 5 5 4 1 5 3 5
## [3146] 4 5 5 4 5 1 1 2 5 4 3 1 4 5 3 5 1 5 5 1 2 5 3 3 4 2 5 3 3 2 2 5 5 4 1 5 4
## [3183] 5 1 3 1 4 1 3 4 3 5 1 2 1 5 4 5 1 1 1 2 1 1 5 1 3 3 2 1 3 5 5 5 5 1 2 1 3
## [3220] 1 5 5 1 1 4 4 5 1 1 4 5 4 1 5 5 3 4 4 5 3 3 5 5 5 1 4 4 4 1 5 1 4 1 4 2 3
## [3257] 2 4 4 1 5 2 4 3 4 4 3 3 1 1 5 1 1 5 3 3 5 5 1 1 1 4 1 4 3 1 1 2 3 5 4 2 4
## [3294] 5 4 1 5 5 5 4 1 3 5 3 1 5 2 1 4 1 3 1 5 5 3 5 5 2 4 5 5 2 3 5 2 1 4 3 3 1
## [3331] 5 5 2 2 5 5 1 5 5 5 1 4 1 1 1 1 5 4 4 4 4 5 1 4 1 5 4 1 1 5 4 1 4 3 5 5 3
## [3368] 4 3 3 1 1 5 4 2 2 5 4 5 1 5 3 4 5 1 1 2 1 1 2 1 3 4 1 1 1 1 1 4 1 4 1 3 5
## [3405] 4 5 1 1 1 5 5 5 4 2 1 4 5 4 3 1 1 3 5 5 1 4 1 1 5 4 1 1 4 5 1 3 5 4 1 1 5
## [3442] 4 5 4 1 2 1 2 3 2 1 2 1 5 1 4 1 5 4 4 1 1 4 3 2 1 1 5 4 2 5 3 5 2 1 3 5 1
## [3479] 1 5 2 1 2 1 5 1 1 5 1 1 4 2 1 1 1 2 5 5 4 3 1 5 1 3 5 4 1 4 2 3 2 1 5 1 5
## [3516] 3 1 1 1 1 4 2 2 4 4 1 1 1 3 3 2 4 4 1 4 3 5 1 5 5 5 3 1 1 4 2 1 5 1 4 5 1
## [3553] 5 2 4 4 4 1 4 2 3 4 4 3 1 5 2 5 2 1 3 4 1 3 1 1 3 2 5 5 5 5 5 1 4 1 2 2 4
## [3590] 4 5 1 1 2 4 1 3 1 5 1 1 1 3 2 4 5 4 4 5 3 1 5 2 1 3 3 1 1 3 2 1 1 1 1 1 1
## [3627] 4 1 5 5 3 3 5 5 5 5 3 2 1 5 3 1 1 5 2 3 2 1 4 5 4 4 4 5 3 1 4 4 5 1 5 5 2
## [3664] 5 2 2 5 2 1 4 5 1 4 5 4 3 5 1 4 1 1 5 2 4 1 4 1 3 2 3 5 1 1 5 1 4 5 1 2 2
## [3701] 3 5 1 5 1 4 4 4 1 5 1 4 3 5 1 1 2 5 4 2 1 5 1 1 2 1 3 1 4 1 2 1 1 5 2 1 1
## [3738] 4 5 3 4 5 5 2 1 1 5 3 3 5 4 4 3 5 1 1 5 1 5 5 1 2 4 5 2 2 4 4 4 3 2 3 1 2
## [3775] 4 1 1 1 5 3 1 4 1 4 4 5 1 1 3 1 4 1 1 4 2 5 4 3 1 2 4 1 5 4 2 3 1 3 2 3 1
## [3812] 5 4 3 4 4 2 1 5 2 4 1 1 1 5 4 5 2 5 5 3 5 3 4 1 5 4 1 2 2 1 4 1 1 3 5 5 5
## [3849] 1 4 4 1 1 1 4 2 5 5 5 4 3 1 1 4 1 5 5 2 4 5 2 4 5 4 1 1 5 5 4 4 1 5 1 5 1
## [3886] 3 2 2 5 4 1 3 4 4 1 4 1 5 3 4 4 2 4 1 1 1 4 1 5 2 4 3 5 5 4 3 1 1 1 3 1 1
## [3923] 1 3 5 4 1 4 1 1 1 4 1 4 2 2 2 2 4 4 2 4 2 2 5 2 4 4 4 5 4 2 1 4 4 2 2 2 4
## [3960] 4 5 5 5 2 4 2 4 2 5 2 2 4 2 4 2 2 2 2 2 5 2 4 2 2 5 2 5 4 4 2 2 4 2 1 2 4
## [3997] 2 2 2 2 5 2 5 4 2 4 4 4 5 4 4 4 2 4 4 2 5 4 5 4 4 2 5 4 1 2 4 2 2 2 2 2 2
## [4034] 5 4 4 2 2 2 2 2 4 2 5 4 4 2 5 2 2 2 2 5 4 4 2 2 2 4 2 5 2 4 4 4 2 4 5 2 2
## [4071] 4 2 2 2 4 2 4 2 2 4 5 2 2 2 2 5 2 5 2 4 4 4 2 2 2 4 5 2 4 2 4 5 4 2 2 5 2
## [4108] 2 2 4 4 2 5 4 4 2 4 2 5 2 4 2 5 2 4 2 2 4 5 2 4 4 4 4 2 5 2 2 2 4 4 4 4 4
## [4145] 5 4 2 5 2 2 4 4 2 2 2 4 2 4 2 5 5 4 4 2 2 2 4 2 4 2 2 2 5 2 2 5 4 4 2 2 2
## [4182] 1 4 4 5 4 2 2 4 4 5 2 5 5 4 5 5 2 4 2 2 4 2 5 4 2 5 2 2 4 2 1 2 4 2 2 4 2
## [4219] 4 4 4 2 4 4 2 5 2 2 2 5 4 2 5 4 5 4 2 4 2 4 2 4 5 5 4 4 4 4 2 5 4 5 2 2 2
## [4256] 2 5 2 2 1 5 2 5 2 4 4 5 4 2 4 5 4 4 4 2 2 4 2 4 2 4 4 4 2 4 2 4 5 2 4 4 5
## [4293] 2 5 2 5 4 4 4 2 2 2 4 4 2 4 2 2 2 2 4 2 5 4 4 2 4 2 4 4 2 4 4 2 2 2 1 2 4
## [4330] 1 2 4 4 5 2 4 5 4 5 2 2 5 4 4 2 2 1 2 4 4 2 2 4 2 2 2 4 2 5 4 4 5 2 5 5 2
## [4367] 2 4 4 2 2 2 4 4 2 2 5 5 1 2 2 2 4 4 4 1 2 2 4 4 2 4 2 2 5 5 2 2 4 2 2 4 2
## [4404] 2 5 2 4 2 2 2 2 5 2 4 2 5 5 2 4 5 4 2 2 2 2 2 4 4 2 2 4 4 2 1 5 5 2 2 4 2
## [4441] 2 2 4 2 2 4 2 4 4 4 5 2 5 2 1 4 4 2 4 1 2 5 4 5 5 4 2 1 4 5 1 5 4 2 4 2 4
## [4478] 4 2 4 2 4 4 4 2 2 5 2 2 2 2 2 2 5 2 4 2 2 2 4 2 2 2 2 2 2 2 5 2 2 2 4 5 2
## [4515] 2 4 4 4 2 4 4 4 2 2 2 2 2 2 5 5 4 2 5 4 4 2 4 4 4 2 5 4 4 4 5 4 4 4 4 5 5
## [4552] 5 2 2 4 2 2 5 5 2 5 2 2 4 4 2 5 2 2 2 2 4 2 4 4 2 4 4 2 2 4 4 4 4 4 4 5 4
## [4589] 4 2 2 4 2 2 2 2 4 2 2 2 2 2 4 4 4 2 4 4 2 4 2 4 5 4 2 5 4 2 4 4 2 4 5 2 2
## [4626] 4 2 5 4 2 4 5 4 2 2 2 2 4 2 2 5 2 4 4 2 2 2 4 1 4 4 4 1 4 2 4 4 5 4 1 4 2
## [4663] 2 2 4 4 2 2 2 5 2 2 2 4 2 4 2 4 2 2 5 4 5 2 5 4 2 2 4 4 5 2 2 2 4 4 2 2 2
## [4700] 2 4 4 2 4 4 2 2 4 5 2 4 4 4 2 4 4 5 4 2 4 1 2 2 4 4 2 5 2 5 2 2 5 2 2 2 4
## [4737] 5 5 2 2 2 2 2 4 2 2 5 4 4 4 4 5 4 5 2 4 2 2 2 4 2 2 4 2 5 2 1 4 2 2 4 4 5
## [4774] 2 2 1 4 4 2 2 4 4 2 5 4 2 4 2 4 4 4 2 2 4 4 4 2 2 2 4 5 2 2 2 4 5 2 2 2 5
## [4811] 5 4 4 4 1 2 4 5 2 4 5 5 2 4 5 5 4 2 2 4 2 2 4 5 2 4 4 5 5 2 2 4 2 2 2 2 4
## [4848] 2 2 5 4 2 4 5 5 4 5 4 2 5 2 2 4 4 2 2 4 2 2 2 2 2 1 2 5 2 2 5 5 2 5 4 2 2
## [4885] 2 2 4 4 4 4 4 2 2 5 2 2 5 5 1 2 5 2 5 2 2 2 4 4 4 2 4 4 4 2 4 5 2 2 2 4 4
## [4922] 2 4 2 2 4 5 2 5 4 4 2 2 4 4 4 2 2 2 4 2 2 2 4 2 2 2 4 4 4 2 5 5 2 2 2 4 5
## [4959] 4 2 4 2 2 4 2 1 4 2 2 4 5 4 4 4 2 2 4 4 2 5 2 2 2 4 2 5 4 2 2 2 4 4 2 4 2
## [4996] 4 2 2 4 4 2 4 2 4 2 2 2 5 4 4 4 4 2 2 2 2 2 4 4 4 2 2 4 5 2 2 4 4 2 4 2 2
## [5033] 5 4 4 4 2 4 4 2 2 4 5 1 2 4 2 2 4 2 4 5 5 2 2 2 4 2 4 5 5 4 2 2 4 5 4 4 2
## [5070] 5 2 4 2 4 2 1 4 4 2 4 2 4 4 2 5 2 4 2 4 5 4 2 2 2 2 4 2 2 2 4 4 4 5 4 4 4
## [5107] 2 2 2 5 2 2 2 5 4 2 4 4 5 2 2 5 2 4 2 2 2 2 1 4 2 4 4 2 4 2 4 5 2 2 5 2 2
## [5144] 2 2 5 2 4 4 2 4 2 4 2 5 4 2 2 2 5 2 2 2 5 2 2 4 2 5 4 2 2 2 2 2 2 4 2 2 4
## [5181] 2 2 2 4 4 2 2 2 2 2 5 2 2 4 4 4 5 2 5 4 4 4 2 4 2 2 2 5 2 4 2 2 4 4 4 4 5
## [5218] 2 2 2 4 2 2 4 5 1 4 2 4 4 4 5 4 4 4 4 2 2 2 2 2 4 2 4 2 2 4 4 4 2 4 2 2 2
## [5255] 2 2 2 4 4 2 4 4 2 2 2 2 4 2 5 2 2 2 2 2 5 2 2 2 4 4 4 5 2 2 2 5 2 4 1 4 4
## [5292] 4 4 4 2 2 2 2 4 2 4 4 4 4 2 2 4 2 2 2 4 2 5 2 2 2 4 2 2 5 4 2 2 2 2 4 4 4
## [5329] 4 2 2 2 5 1 5 4 4 1 4 4 4 5 4 5 4 2 4 5 2 4 2 4 4 4 4 2 2 4 2 5 4 2 5 4 2
## [5366] 2 2 2 2 4 5 2 4 2 5 1 5 5 2 4 4 4 2 2 2 1 2 2 2 4 5 2 2 2 2 4 5 4 2 2 5 2
## [5403] 4 5 4 5 2 2 4 5 2 2 4 5 4 2 2 2 4 4 2 5 2 4 5 5 4 2 2 4 5 2 4 4 2 2 4 2 5
## [5440] 2 4 5 5 4 2 4 2 4 2 5 2 2 2 4 4 2 4 4 2 4 2 2 5 5 4 4 2 4 4 2 2 4 2 2 5 2
## [5477] 4 5 5 4 4 2 2 4 2 2 4 2 2 5 2 2 2 4 2 2 2 2 2 2 2 5 4 2 2 2 4 4 4 5 5 2 2
## [5514] 4 2 4 2 4 4 4 5 2 2 2 4 4 1 5 2 2 4 4 4 4 4 4 4 2 4 4 4 1 4 2 2 4 5 4 2 1
## [5551] 2 4 4 2 2 5 2 2 2 2 2 4 2 4 2 2 2 2 4 5 2 1 4 2 5 2 4 2 2 2 4 2 2 2 2 4 2
## [5588] 2 4 2 5 2 5 2 5 4 2 4 2 2 2 4 5 4 4 2 4 2 2 4 4 4 2 4 4 4 4 2 2 2 4 2 2 4
## [5625] 2 2 2 4 2 4 2 2 2 2 4 4 4 2 4 2 2 5 2 2 5 4 4 2 4 4 4 2 2 5 5 2 2 1 1 2 2
## [5662] 4 2 1 2 2 4 4 4 4 1 2 1 4 2 2 2 2 5 2 4 4 2 4 5 2 1 5 4 4 2 2 4 5 1 4 2 2
## [5699] 4 4 4 5 2 4 2 2 5 2 2 2 1 5 2 4 2 4 2 4 4 2 2 2 2 2 5 2 2 2 4 5 4 4 4 2 4
## [5736] 2 4 4 2 2 2 4 5 2 4 4 2 1 4 2 4 4 2 5 5 2 2 2 5 2 2 2 5 2 2 2 2 2 5 5 2 2
## [5773] 2 2 2 2 1 2 4 2 4 2 2 1 4 4 4 2 4 1 2 4 2 2 2 5 4 4 2 4 4 4 4 5 4 4 5 4 2
## [5810] 5 2 2 4 4 2 2 4 5 2 2 2 2 2 4 4 2 5 2 4 2 4 5 2 2 4 2 4 4 4 5 2 4 2 5 1 2
## [5847] 4 2 2 5 2 1 5 4 2 5 2 1 2 2 1 5 4 5 4 2 2 2 4 2 2 5 5 5 4 5 2 4 4 2 4 2 4
## [5884] 4 2 2 2 4 2 4 2 4 4 2 5 5 2 2 4 2 2 5 4 2 5 2 2 2 2 2 2 5 5 5 4 2 4 2 5 2
## [5921] 2 2 2 5 2 2 5 2 2 4 2 5 4 2 4 2 2 2 2 4 2 2 2 2 2 4 2 2 4 4 5 2 2 5 4 2 2
## [5958] 2 2 4 4 2 5 4 2 5 2 2 4 4 2 4 2 2 4 2 4 4 2 2 2 2 4 4 4 4 2 4 2 2 4 2 1 2
## [5995] 4 2 4 2 5 2 2 5 2 4 5 5 4 2 4 2 2 2 2 2 2 4 2 4 4 2 2 2 2 2 2 5 2 2 4 2 2
## [6032] 4 1 2 4 1 4 2 2 5 5 4 5 1 2 4 4 4 4 5 4 5 5 4 4 4 2 2 5 2 4 4 2 2 2 5 4 5
## [6069] 4 4 2 4 5 4 2 2 2 2 2 5 4 4 2 4 2 5 4 2 4 1 2 5 4 2 2 2 2 2 5 2 2 4 2 2 4
## [6106] 5 1 5 2 2 4 5 2 2 4 4 2 4 4 1 2 2 5 2 2 4 4 4 4 4 2 4 4 2 2 4 4 4 2 2 2 2
## [6143] 4 4 4 5 4 2 2 2 4 2 2 2 2 2 2 5 2 2 5 5 5 2 4 4 2 2 2 2 4 4 2 4 5 4 4 5 2
## [6180] 2 2 5 4 2 2 4 4 1 2 2 2 4 2 4 4 2 4 5 4 1 2 4 4 4 4 2 4 4 2 4 4 2 5 4 4 2
## [6217] 2 2 4 1 4 4 5 4 2 2 2 2 2 2 5 2 2 2 2 5 5 4 5 2 5 4 4 4 2 4 2 2 4 2 2 2 2
## [6254] 2 4 2 4 5 4 4 5 2 4 2 2 4 4 5 5 2 2 2 4 5 4 4 4 2 2 4 4 2 5 1 4 2 2 2 2 2
## [6291] 5 4 1 4 5 2 2 2 4 2 2 5 4 4 2 2 4 5 2 2 2 5 2 2 4 4 4 4 4 2 2 5 4 2 5 4 1
## [6328] 4 4 4 1 5 4 1 4 5 5 4 4 2 2 4 5 4 4 4 2 2 4 2 4 2 4 2 2 2 5 5 4 4 2 2 1 2
## [6365] 4 4 2 4 2 2 2 5 5 2 4 2 2 2 2 4 5 2 5 4 4 4 2 4 2 2 1 5 4 4 2 2 2 4 1 2 2
## [6402] 4 4 5 4 2 4 4 4 2 2 2 2 2 2 4 5 4 4 4 5 4 4 5 2 2 1 4 2 2 5 4 4 2 2 2 5 5
## [6439] 2 2 1 4 2 2 4 4 4 5 4 2 2 2 2 2 2 2 5 2 4 2 2 2 2 2 4 5 2 4 5 4 5 2 2 4 2
## [6476] 2 4 4 2 5 4 4 2 2 2 5 2 2 2 4 4 2 2 4 4 2 2 2 4 4 2 2 2 4 2 2 2 5 4 2 2 2
## [6513] 4 2 2 2 2 5 4 4 2 4 4 2 5 2 4 2 4 2 5 5 4 5 4 2 4 4 2 4 4 5 5 4 2 4 2 5 2
## [6550] 2 4 2 1 2 2 2 2 2 2 4 4 4 4 2 2 2 2 2 2 2 2 4 2 4 4 2 2 2 2 4 2 4 4 2 5 2
## [6587] 5 4 5 2 2 5 5 2 2 2 4 4 4 4 2 5 4 5 2 2 2 2 5 2 2 2 2 2 4 2 2 2 5 2 1 2 4
## [6624] 2 2 2 4 2 4 5 2 4 2 4 2 2 4 2 4 2 4 2 2 2 2 2 2 5 2 4 2 4 2 4 1 2 2 5 4 2
## [6661] 4 4 2 5 4 4 2 4 5 4 2 2 4 4 2 2 4 4 2 2 2 2 2 5 4 4 2 4 4 2 2 2 4 2 5 2 2
## [6698] 4 5 4 2 2 4 2 2 2 2 2 2 5 2 4 4 5 2 2 2 4 2 2 4 2 2 2 4 5 1 4 2 5 4 4 2 2
## [6735] 4 2 4 4 4 4 2 2 2 4 2 2 2 2 2 2 4 4 2 4 5 2 4 2 2 2 2 5 4 2 2 2 4 4 2 4 2
## [6772] 2 2 2 4 4 2 2 2 5 2 5 2 1 4 2 4 2 1 2 2 4 2 2 2 2 4 2 5 2 2 2 4 2 4 5 4 4
## [6809] 4 2 2 2 4 2 2 4 2 4 2 2 5 4 2 4 4 2 5 2 5 4 2 4 1 2 2 5 4 4 4 2 2 2 2 2 5
## [6846] 4 2 2 2 2 5 2 5 4 5 2 2 2 2 5 2 4 4 2 2 2 2 2 4 4 2 2 4 4 2 5 2 4 2 2 2 2
## [6883] 4 2 2 4 4 5 4 2 4 2 2 4 5 4 2 4 4 5 2 2 5 4 5 2 5 2 2 1 2 4 2 2 2 5 2 5 2
## [6920] 4 5 2 5 5 2 4 2 4 1 2 2 2 2 2 5 4 2 2 2 4 4 4 2 5 2 4 2 4 5 4 2 2 4 2 4 2
## [6957] 2 5 2 2 4 2 4 2 2 2 2 2 4 2 2 4 5 5 4 4 2 2 5 2 2 2 1 2 5 5 4 4 4 2 5 4 4
## [6994] 2 2 4 2 4 2 2 5 1 5 4 4 2 5 2 2 2 2 2 2 4 4 2 5 2 5 4 2 4 2 2 4 2 2 2 4 4
## [7031] 2 2 2 2 2 2 4 5 4 5 4 1 2 2 2 2 4 2 4 5 2 2 2 2 2 2 4 2 4 5 2 2 2 4 2 2 4
## [7068] 2 5 4 2 2 2 2 5 4 2 4 2 4 4 5 4 4 4 4 4 4 2 2 2 4 2 4 2 4 2 5 5 5 4 2 2 4
## [7105] 2 4 2 2 2 4 2 4 4 5 5 2 2 2 4 2 2 2 4 4 4 2 2 4 4 4 4 5 4 4 4 4 4 2 5 2 5
## [7142] 5 4 2 4 4 4 4 2 4 5 4 2 4 4 2 4 2 4 2 2 4 2 2 4 1 2 4 4 4 1 2 4 2 5 2 2 2
## [7179] 2 4 2 2 5 5 2 5 2 4 2 4 2 2 2 4 4 4 2 2 4 5 4 4 2 4 2 2 4 1 4 2 4 2 4 2 4
## [7216] 4 4 4 4 4 2 4 2 2 4 2 2 4 5 2 2 2 4 5 4 5 5 2 4 4 2 4 5 2 4 2 2 2 2 4 4 2
## [7253] 5 1 2 4 2 4 4 4 4 2 5 2 4 4 5 2 4 2 2 2 5 2 2 4 2 2 5 4 2 4 4 4 2 5 2 2 5
## [7290] 2 2 4 4 1 2 2 4 4 4 2 2 1 4 2 2 2 2 2 4 2 2 2 2 4 2 5 4 5 4 2 5 2 5 2 2 5
## [7327] 4 2 4 4 2 1 2 4 2 4 4 2 2 2 2 2 4 5 5 5 2 2 2 1 2 4 4 2 5 4 2 2 4 5 5 4 2
## [7364] 2 4 4 2 4 4 2 2 4 2 4 2 2 4 2 2 5 4 5 2 5 2 4 4 4 5 4 2 4 2 2 4 2 2 2 2 2
## [7401] 1 4 4 1 2 5 2 2 1 2 2 2 4 4 2 2 2 2 2 4 2 4 2 4 2 4 2 5 2 2 4 5 4 4 2 5 2
## [7438] 2 2 2 4 5 5 2 4 5 4 2 4 2 2 5 2 2 5 4 5 5 4 2 5 4 2 5 5 2 2 2 2 2 2 4 4 5
## [7475] 4 2 5 2 4 5 4 4 4 4 2 4 4 2 2 4 1 4 2 4 2 4 4 4 2 4 4 2 2 2 5 2 4 2 2 2 2
## [7512] 4 2 2 5 2 4 4 2 4 2 4 4 2 4 2 5 4 5 4 2 2 2 4 4 1 2 2 2 5 5 4 2 2 2 4 2 5
## [7549] 2 4 4 4 4 2 2 2 4 2 2 2 2 2 4 4 4 2 5 5 2 4 2 4 5 2 2 4 4 2 4 2 4 2 2 4 4
## [7586] 2 4 2 2 2 2 2 2 5 2 4 5 2 4 4 4 2 2 2 2 2 1 4 4 2 4 2 2 2 4 2 4 4 2 4 2 4
## [7623] 4 4 4 2 2 5 2 2 2 4 4 4 5 4 2 4 2 5 2 5 4 4 5 5 4 2 4 2 2 2 2 2 2 5 5 2 2
## [7660] 4 2 2 4 2 4 4 4 1 4 4 5 5 4 4 2 2 1 4 4 2 4 2 5 2 4 2 4 4 4 4 2 2 4 4 4 4
## [7697] 2 2 2 4 2 2 5 4 2 5 4 2 5 2 4 4 4 2 4 5 4 2 2 2 5 4 4 4 5 5 2 2 5 2 4 4 4
## [7734] 2 2 2 2 2 2 4 5 4 4 5 2 4 2 2 2 2 2 2 2 4 5 5 2 5 4 5 4 5 2 4 4 2 4 2 2 4
## [7771] 4 2 2 2 5 2 2 4 2 2 4 5 4 2 4 4 4 4 2 2 4 5 2 2 4 1 1 4 2 2 2 4 2 4 4 2 2
## [7808] 2 2 2 2 2 2 4 2 2 4 2 4 2 2 2 1 2 2 4 4 2 2 2 4 2 5 4 2 2 2 4 4 2 2 4 2 4
## [7845] 2 2 4 4 2 1 2 5 4 4 2 2 4 2 2 2 2 2 2 5 2 2
## 
## Within cluster sum of squares by cluster:
## [1] 2362.507 2776.912 1357.290 2460.888 2450.813
##  (between_SS / total_SS =  63.7 %)
## 
## Available components:
## 
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
loan5 <- cbind(loan3, kmeans_loan5$cluster)

table(loan5$Approval , loan5$`kmeans_loan5$cluster`)
```

```
## 
##        1    2    3    4    5
##   0 1348  479  450  749  907
##   1  107 1891    0 1374  561
```

```
loan22 <- cbind(loan3, kmeans_loan$cluster)
table(loan22$Approval, loan22$`kmeans_loan$cluster`)
```

```
## 
##        1    2
##   0 1927 2006
##   1 1407 2526
```

Bag of Words

```
loan3$Text <- stringr::str_to_lower(loan3$Text)
```

```
loan3$Text <- qdapRegex::rm_twitter_url(
loan3$Text,
replacement = " ",
clean = TRUE
)
```

```
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "ã",
replacement = "a"
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "š",
replacement = "s"
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "â",
replacement = "a"
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "¿",
replacement = "?"
)
```

```
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = " i'm ",
replacement = " i am "
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "'re ",
replacement = " are "
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "'t ",
replacement = " not "
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "'ve ",
replacement = " have "
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "'ll ",
replacement = " will "
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = " doesn't ",
replacement = " does not "
)
```

```
loan3$Text <- stringr::str_replace_all(
string = loan3$Text ,
pattern = "[:punct:]",
replacement = " "
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "[:digit:]",
replacement = " "
)
loan3$Text <- stringr::str_replace_all(
string = loan3$Text,
pattern = "\\W",
replacement = " "
)
```

```
loan3$Text <- tm::removeWords(
x = loan3$Text,
words = tm::stopwords(kind = "SMART")
)
loan3$Text <- tm::removeWords(
x = loan3$Text,
words = tm::stopwords(kind = "english")
)
loan3$Text <- tm::removeWords(
x = loan3$Text,
words = qdapDictionaries::Top200Words
)
```

```
loan3$Text <- tm::removeWords(
x = loan3$Text,
words = c('loan', 'funds', 'money', 'financial', 'assistance', 'support', 'pay', 'purchase', 'bu
y', 'cover',
          'payment')
)
```

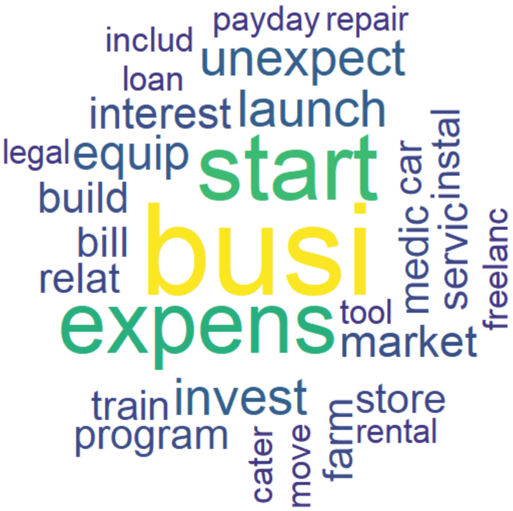
```
loan3$Text <- trimws(stringr::str_replace_all(
string = loan3$Text,
pattern = "\\s+",
replacement = " "
))
```

```
strsplit_text <- strsplit(loan3$Text," ")
dictionary_text <- sort(unique(unlist(strsplit_text)))
strsplit_text <- lapply(
X = strsplit_text,
FUN = tm::stemDocument
)
#strsplit_tweets <- lapply(
# X = strsplit_tweets,
# FUN = tm::stemCompletion,
# dictionary = dictionary_tweets
#)
strsplit_text <- lapply(
X = strsplit_text,
FUN = paste,
collapse = " "
)
loan3$text2 <- unlist(strsplit_text)
```

```
Corpus_text <- tm::VCorpus(tm::VectorSource(loan3$text2))

DocumentTermMatrix_text <- tm::DocumentTermMatrix(Corpus_text)

DocumentTermMatrix_text <- tm::removeSparseTerms(
DocumentTermMatrix_text,
0.995
)

M <- as.matrix(DocumentTermMatrix_text)
dim(M)
```

```
## [1] 7866  183
```

```
term_frequency <- data.frame(
Term = colnames(M),
Frequency = colSums(M),
stringsAsFactors = FALSE
)
term_frequency <- term_frequency[order(term_frequency$Frequency),]

wordcloud::wordcloud(
words = term_frequency$Term,
freq = term_frequency$Frequency,
max.words = 30,
random.order = FALSE,
colors = viridis::viridis(100)
)
```



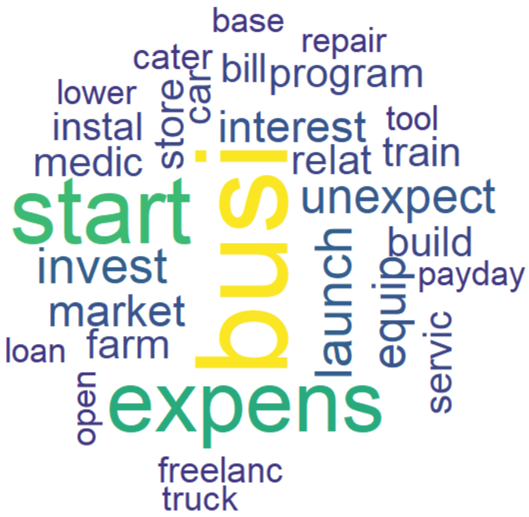
word cloud for all loans

```
approved_wc <- as.data.frame(loan3[loan3$Approval == 1, 8])
```

```
app_Corpus_text <- tm::VCorpus(tm::VectorSource(approved_wc$`loan3[loan3$Approval == 1, 8]`))

app_DocumentTermMatrix_text <- tm::DocumentTermMatrix(app_Corpus_text)

app_DocumentTermMatrix_text <- tm::removeSparseTerms(
app_DocumentTermMatrix_text,
0.995
)

app_M <- as.matrix(app_DocumentTermMatrix_text)
dim(app_M)
```

```
## [1] 3933  183
```

```
app_term_frequency <- data.frame(
Term = colnames(app_M),
Frequency = colSums(app_M),
stringsAsFactors = FALSE
)
app_term_frequency <- app_term_frequency[order(app_term_frequency$Frequency),]

wordcloud::wordcloud(
words = app_term_frequency$Term,
freq = app_term_frequency$Frequency,
max.words = 30,
random.order = FALSE,
colors = viridis::viridis(100)
)
```



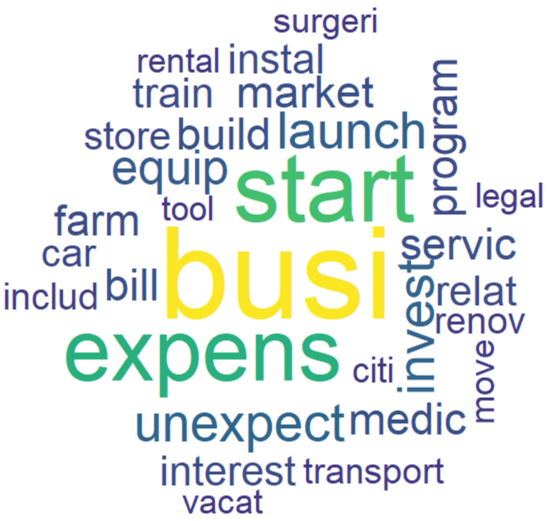
word cloud for loans that were approved

```
napproved_wc <- as.data.frame(loan3[loan3$Approval == 0, 8])
```

```
napp_Corpus_text <- tm::VCorpus(tm::VectorSource(napproved_wc$`loan3[loan3$Approval == 0, 8]`))

napp_DocumentTermMatrix_text <- tm::DocumentTermMatrix(napp_Corpus_text)

napp_DocumentTermMatrix_text <- tm::removeSparseTerms(
napp_DocumentTermMatrix_text,
0.995
)

napp_M <- as.matrix(napp_DocumentTermMatrix_text)
dim(napp_M)
```

```
## [1] 3933  183
```

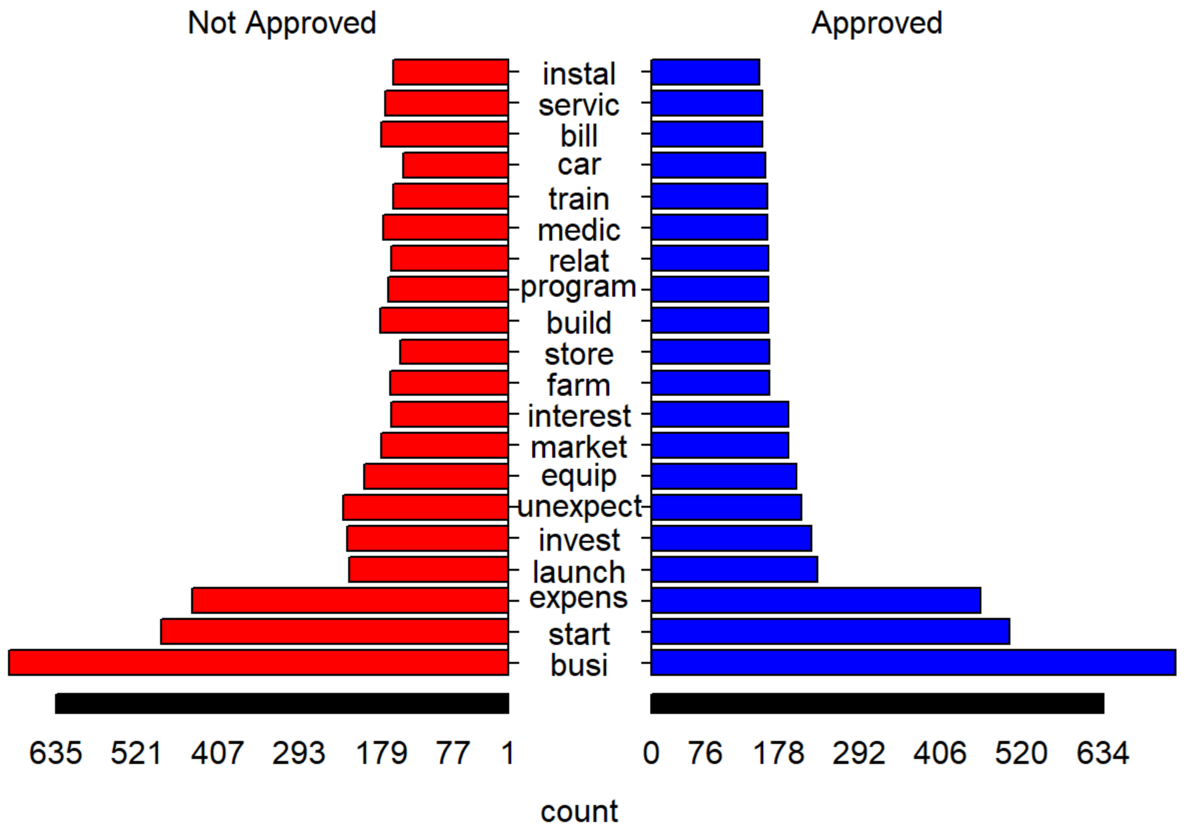
```
napp_term_frequency <- data.frame(
Term = colnames(napp_M),
Frequency = colSums(napp_M),
stringsAsFactors = FALSE
)
napp_term_frequency <- napp_term_frequency[order(napp_term_frequency$Frequency),]

wordcloud::wordcloud(
words = napp_term_frequency$Term,
freq = napp_term_frequency$Frequency,
max.words = 30,
random.order = FALSE,
colors = viridis::viridis(100)
)
```



```
app_term_frequency2 <- app_term_frequency[order(rownames(app_term_frequency)),]
napp_term_frequency2 <- napp_term_frequency[order(rownames(napp_term_frequency)),]
tot_tf <- cbind(app_term_frequency2,napp_term_frequency2$Frequency)
colnames(tot_tf)[2] <- c('approved_tf')
colnames(tot_tf)[3] <- c('not_approved_tf')
```

```
tot_tf <- tot_tf[order(-tot_tf$approved_tf),]
```

```
plotrix::pyramid.plot(
lx = tot_tf[1:20,3],
rx = tot_tf[1:20,2],
labels = rownames(tot_tf)[1:20],
top.labels = c("Not Approved","","Approved"),
lxcol = "red",
rxcol = "blue",
unit="count",
gap = 100
)
```



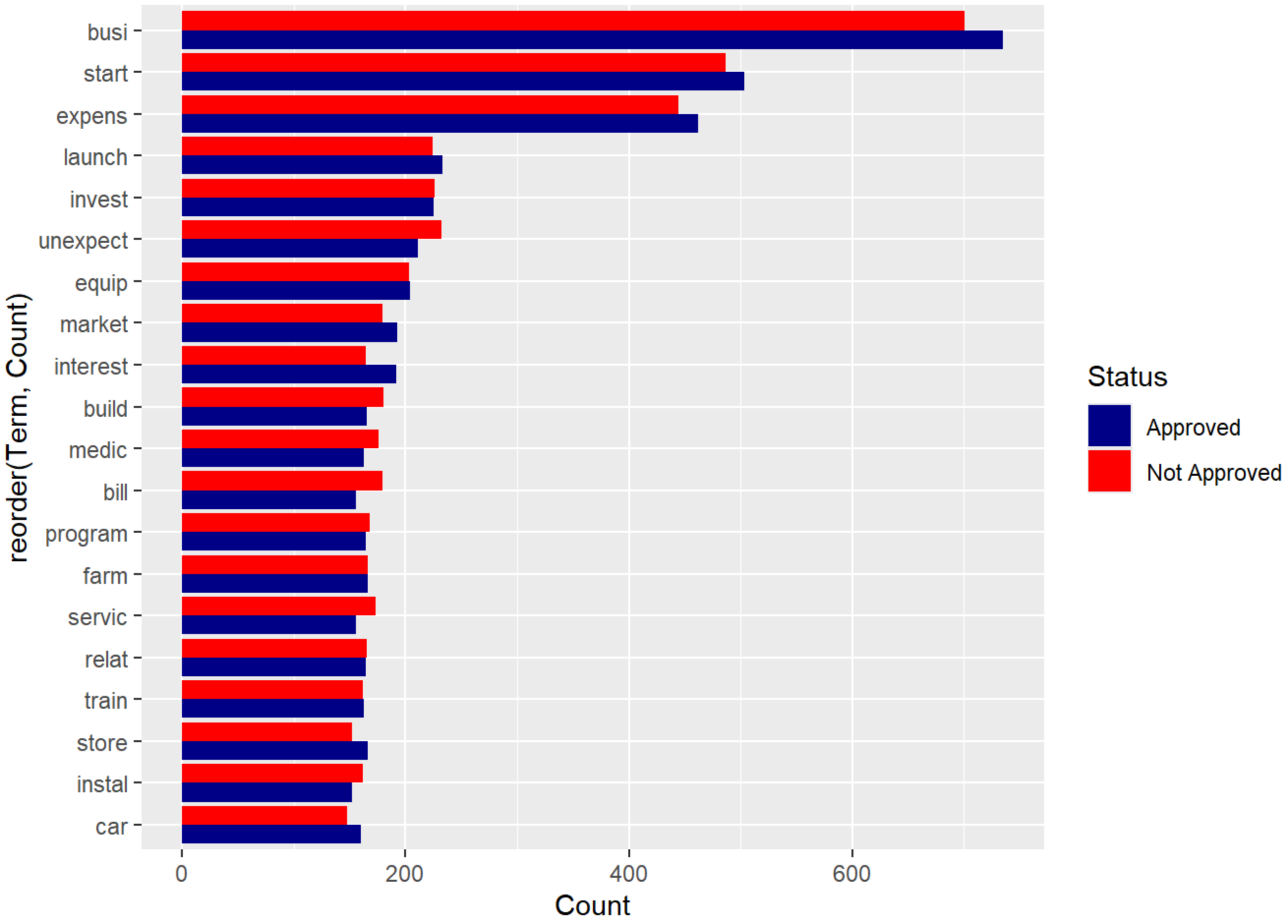
```
## 735 735
```

```
## [1] 5.1 4.1 4.1 2.1
```

```
#rep(tot_tf$Term, 2)
```

```
tim <- tot_tf[1:20,]
total_tf1 <- data.frame(Term = tim$Term, Status = c(rep('Approved', 40)),
                       Count = tim$approved_tf)
total_tf2 <- data.frame(Term = tim$Term, Status = c(rep('Not Approved', 40)),
                       Count = tim$not_approved_tf)

total_tf3 <- rbind(total_tf1, total_tf2)
```

```
library(ggplot2)
ggplot(total_tf3, aes(fill = Status, y=Count, x=reorder(Term, Count))) +
    geom_bar(position="dodge", stat="identity") +
  coord_flip() +
  scale_fill_manual(values=c(
                             "darkblue",
                            "red"))
```


busi
start
expens
launch
invest
unexpect
equip
market
interest
build
medic
bill
program
farm
servic
relat
train
store
instal
car
reorder(Term, Count)
0
200
400
600
Count
Status
Approved
Not Approved