

Bank Credit's Customer Dashboard

MinhChau

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ forcats   1.0.0   ✓ readr     2.1.5
## ✓ ggplot2   3.5.1   ✓ stringr  1.5.1
## ✓ lubridate 1.9.4   ✓ tibble   3.2.1
## ✓ purrr     1.0.4   ✓ tidyr    1.3.1
## — Conflicts ————— tidyverse_conflicts() —
## * dplyr::filter() masks stats::filter()
## * dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
## # A tibble: 6 × 12
##   `Loan Purpose` Checking Savings `Months Customer` `Months Employed` Gender
##   <chr>          <dbl>   <dbl>          <dbl>          <dbl> <chr>
## 1 Small Appliance      0     739            13            12 M
## 2 Furniture            0    1230            25             0 M
## 3 New Car              0     389            19           119 M
## 4 Furniture          638     347            13            14 M
## 5 Education          963    4754            40            45 M
## 6 Furniture        2827      0             11            13 M
## # i 6 more variables: `Marital Status` <chr>, Age <dbl>, Housing <chr>,
## #   Years <dbl>, Job <chr>, `Credit Risk` <chr>
```

1. Customer Profile Dashboard for Credit Risk Analysis

- i. View the frequency distributions for each of the following customer demographic variables in a chart and table (i.e. one table and chart per variable): `Credit Risk` and `Total`. `Total` is the sum of `Checking` and `Savings`. (hint: You will need to create the variable `Total` in the dataframe).
- ii. View the relationship between `Total` and `Months Employed` in one chart, and the relationship between `Total` and `Age` in another chart.
- iii. Provide a description of any interesting patterns observed from the charts. (You may type your answer in the space below)
- iv. Frequency distributions for `Housing`, `Job` and `Months Employed` on your own.

```
BD$Total <- BD$Checking+BD$Savings
BD <- BD %>%
  mutate(Total = Checking+Savings)

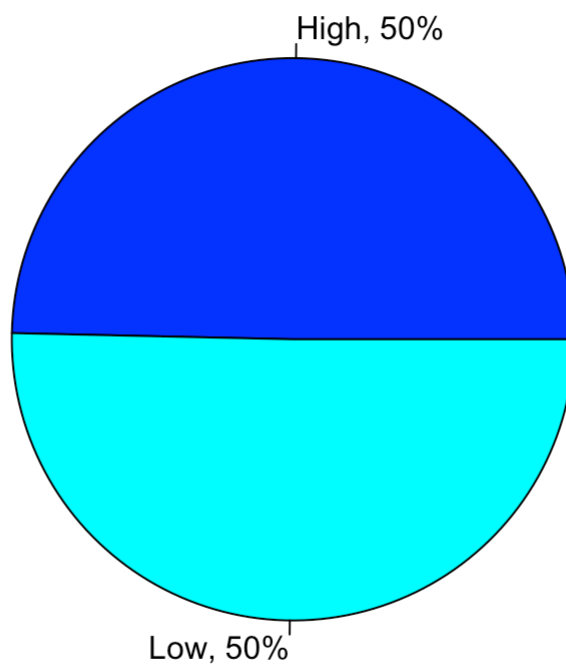
crFreq <- BD%>%count(`Credit Risk`)
kable(crFreq, caption = "Frequency of Bank Customers by Credit Risk")
```

Frequency of Bank Customers by Credit Risk

Credit Risk	n
High	211
Low	214

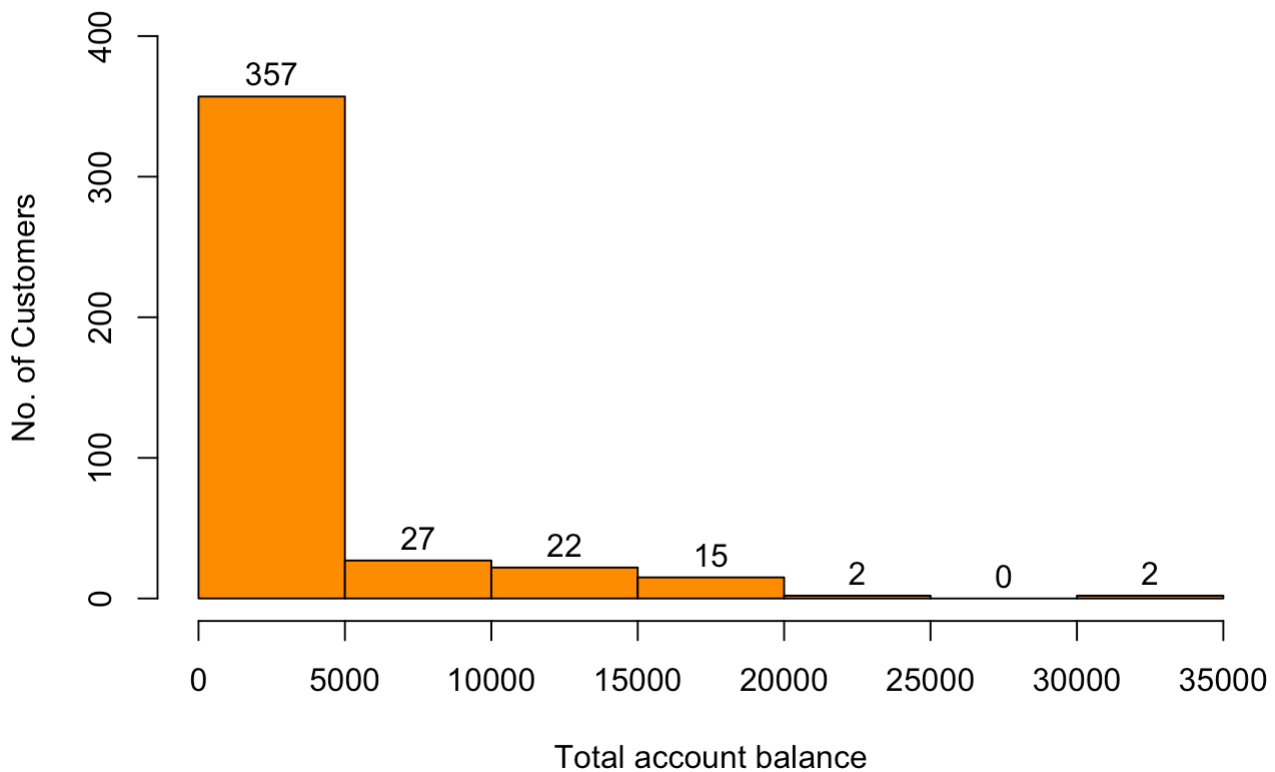
```
slice.cr <- crFreq$n
cr.piepercent <- 100*round(crFreq$n/sum(crFreq$n),2)
label<-crFreq`Credit Risk`
label<-paste(label, ",", sep="")
label<-paste(label,cr.piepercent)
label<-paste(label, "%", sep="")
pie(slice.cr,
     labels=label,
     col=c("blue", "cyan"),
     radius=1,
     main="Customer Credit Risk")
```

Customer Credit Risk



```
h.tot <- hist(BD$Total,
              main="Histogram of customer total amount balance",
              xlab="Total account balance",
              ylab="No. of Customers",
              col=c("darkorange"),
              ylim=c(0,400),
              labels=TRUE)
```

Histogram of customer total amount balance



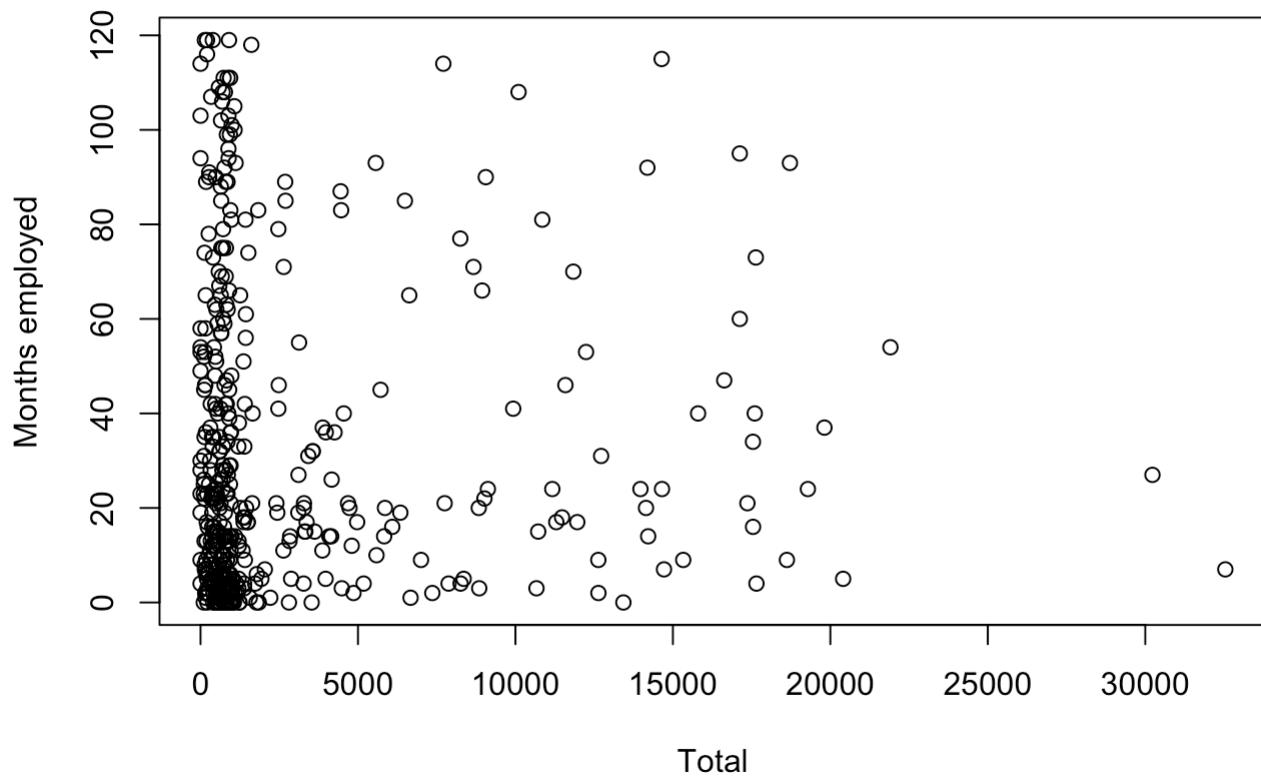
```
ab.Group <- cut(BD$Total, h.tot$breaks, include.lowest = TRUE, dig.lab=5)
t.emp<- table(ab.Group)
kable(t.emp, caption = "Frequency distribution by Total Account Balance")
```

Frequency distribution by Total Account Balance

ab.Group	Freq
[0,5000]	357
(5000,10000]	27
(10000,15000]	22
(15000,20000]	15
(20000,25000]	2
(25000,30000]	0
(30000,35000]	2

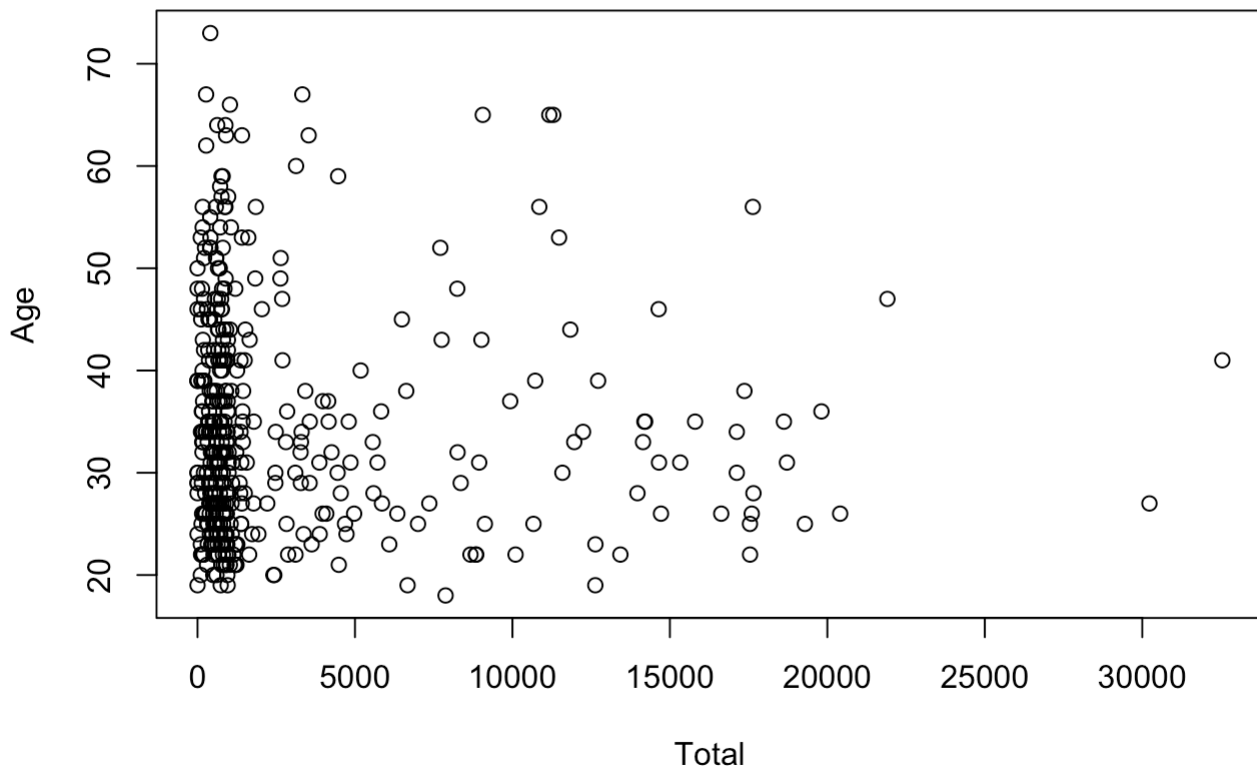
```
plot(BD$Total,  
     BD$`Months Employed`,  
     main="Scatterplot of months memployed to total",  
     ylab="Months employed", xlab="Total")
```

Scatterplot of months memployed to total



```
plot(BD$Total, BD$Age, main="Scatterplot of Age to Total",ylab="Age", xlab="Total")
```

Scatterplot of Age to Total



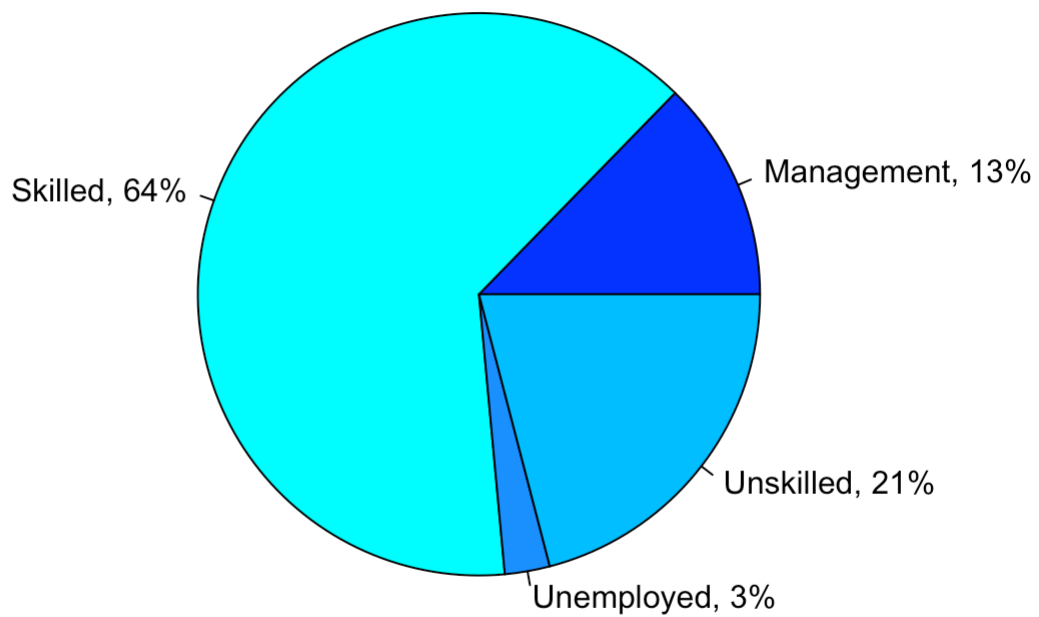
```
JobFreq <- BD%>% count(Job)
kable(JobFreq, caption="Frequency of Bank Customers by Job")
```

Frequency of Bank Customers by Job

Job	n
Management	54
Skilled	271
Unemployed	11
Unskilled	89

```
slice.job <- JobFreq$n
job.piepercent <- 100*round(JobFreq$n/sum(JobFreq$n),2)
label <- JobFreq$Job
label <- paste(label, ",", sep="")
label <- paste(label, job.piepercent)
label <- paste(label, "%", sep="")
pie(slice.job,
    labels = label,
    col=c("blue", "cyan", "dodgerblue", "deepskyblue"),
    radius=1,
    main="Customer Job")
```

Customer Job



2. Customers' Demographics

- i. see the appropriate chart and table to compare frequency of customers by Credit Risk and Job .
- ii. see a description of any interesting patterns observed from the charts.

```
BDb1 <- BD %>% group_by(`Credit Risk`, Job)%>% tally()
BDb1.spread <- BDb1 %>% spread(key = Job, value=n)
kable(BDb1.spread, caption = "Contringency table for Credit risk and Job")
```

Contringency table for Credit risk and Job

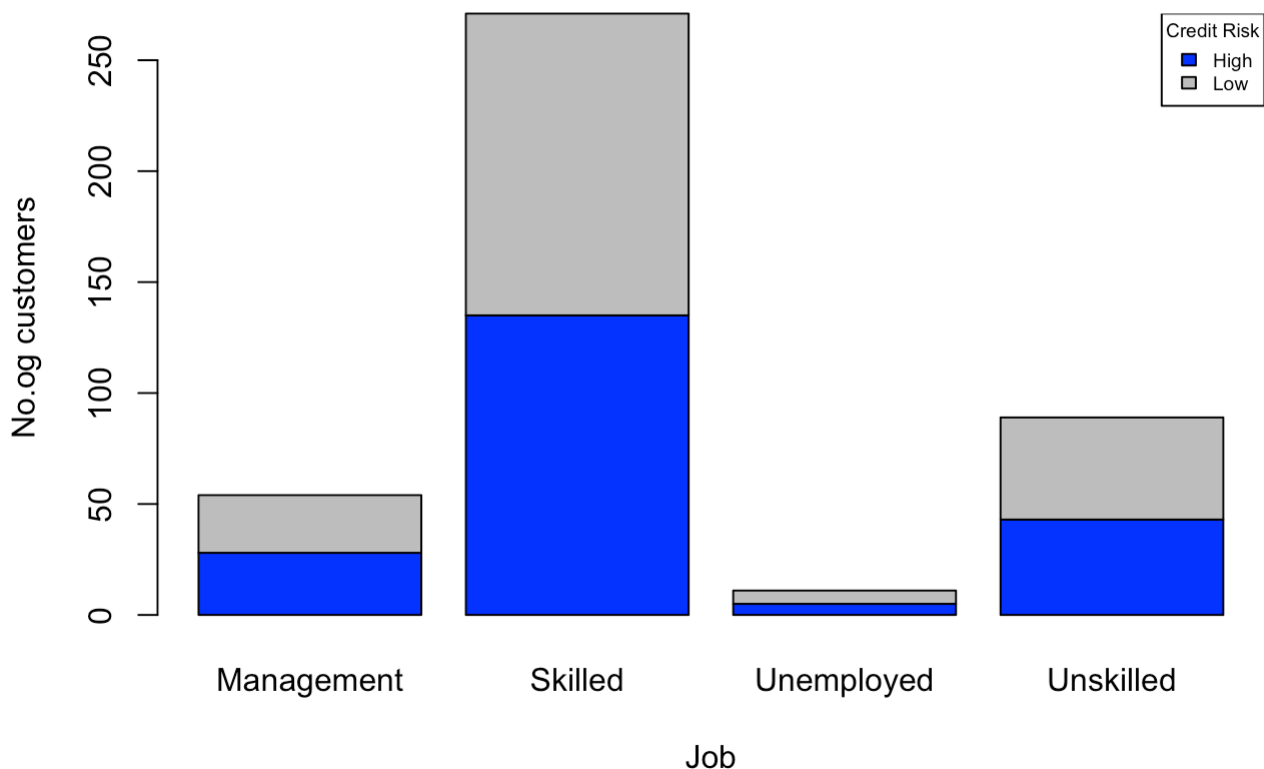
Credit Risk	Management	Skilled	Unemployed	Unskilled
High	28	135	5	43
Low	26	136	6	46

```

barmatrix.BDb1 <- as.matrix(BDb1.spread[,c(2:5)])
barcol <-c("blue", "gray")
barplot(barmatrix.BDb1,
        col=barcol,
        main="Frequency of Customer by Credit Risk and Job",
        ylab="No.og customers",
        xlab="Job")
legend("topright",
       cex=0.6,
       fill=barcol,
       legend=BDb1.spread`Credit Risk`,
       title="Credit Risk")

```

Frequency of Customer by Credit Risk and Job

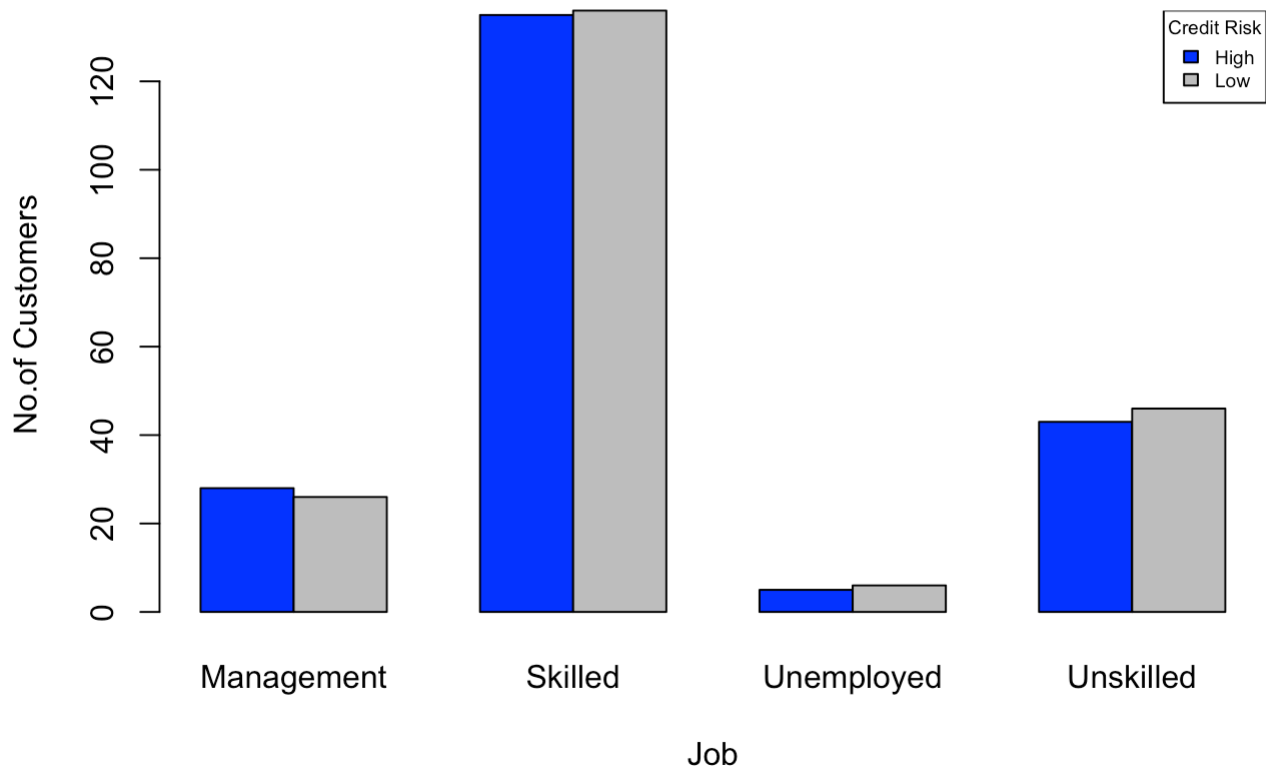


```

barplot(barmatrix.BDb1,
        col=barcol,
        main="Frequency of Customer by Credit risk and Job",
        ylab="No.of Customers",
        xlab="Job",
        beside=TRUE)
legend("topright", cex=0.6, fill=barcol, legend=BDb1.spread`Credit Risk`, title = "C
redit Risk")

```

Frequency of Customer by Credit risk and Job



3. Credit Risk Analysis

- i. View the the Loan Purpose of customers with “High” levels of Credit Risk . Visualize the frequency distribution of Loan Purpose for “High” Credit Risk customers
- ii. The most and least common Loan Purpose types among “High” Credit Risk customers

```
LoanHRFreq<- BD%>%
  filter (`Credit Risk` == "High") %>%
  count(`Loan Purpose`)
kable(LoanHRFreq, caption = "Frequency Distribution for Loan Purpose for High CR Customers")
```

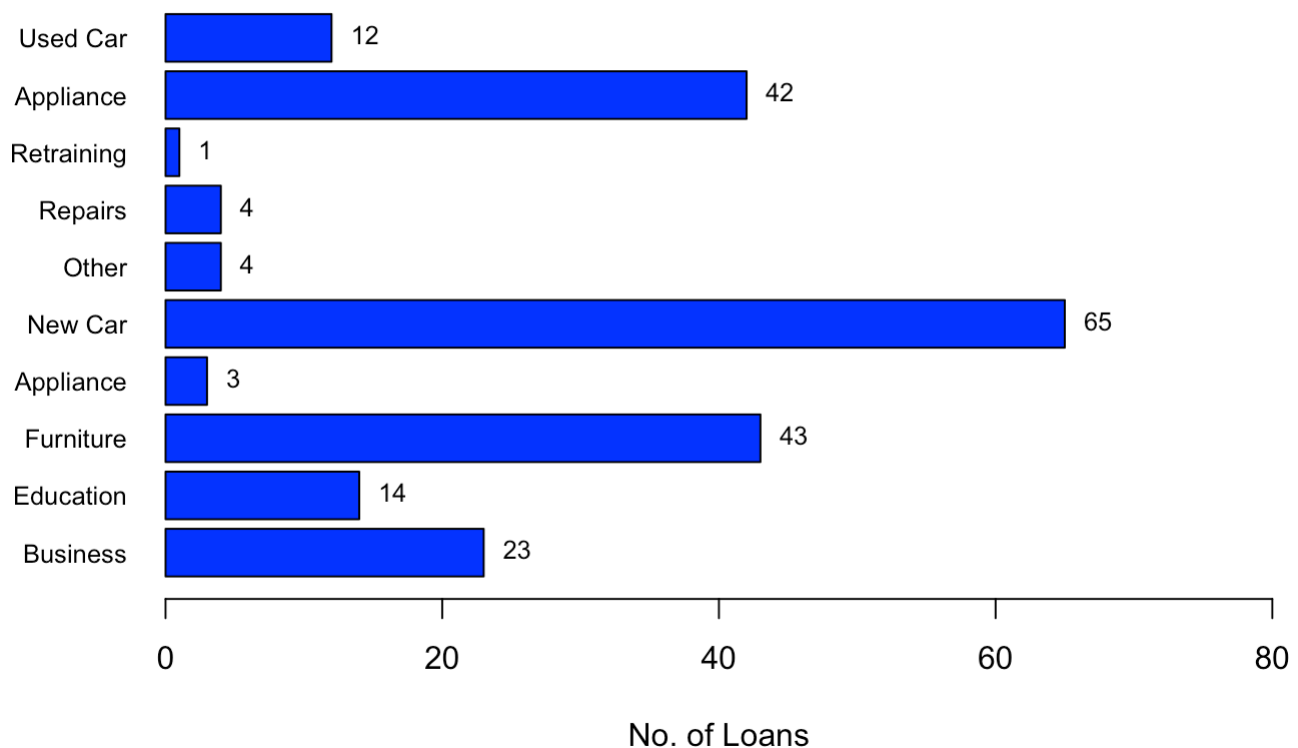
Frequency Distribution for Loan Purpose for High CR Customers

Loan Purpose	n
Business	23
Education	14
Furniture	43
Large Appliance	3
New Car	65
Other	4
Repairs	4
Retraining	1

Loan Purpose	n
Small Appliance	42
Used Car	12

```
LoanHRbar <- LoanHRFreq$n
bploanHR <- barplot(LoanHRbar, names.arg = LoanHRFreq`Loan Purpose`,
  col="blue",
  main="Frequency of Loan Purpose for High CR Customers",
  cex.names = 0.8,
  xlim=c(0,80),
  xlab="No. of Loans",
  horiz=TRUE,
  las=1)
text(x=LoanHRbar, y=bploanHR, col="black", LoanHRFreq$n, cex=0.8, pos=4)
```

Frequency of Loan Purpose for High CR Customers



```
LoanMax <- LoanHRFreq%>% slice_max(n, n=1)
LoanMin <- LoanHRFreq%>% slice_min(n, n=1)
rbind(LoanMax, LoanMin)
```

```
## # A tibble: 2 × 2
##   `Loan Purpose`      n
##   <chr>          <int>
## 1 New Car         65
## 2 Retraining       1
```

Q1.(d) Customer Account Balance Pareto Analyses Show the number and percentage of customers that contribute most, amounting to 80% of the total account balanced with the bank.

```
BD.tot <- BD %>%  
  select(Total)%>%  
  arrange(desc(Total))  
BD.tot$Percentage <-BD.tot$Total/sum(BD.tot$Total)  
BD.tot$Cumulative <- cumsum(BD.tot$Percentage)  
BD.tot$Cumulative.cust<-as.numeric(rownames(BD))/nrow(BD)  
which(BD.tot$Cumulative>0.8)[1]
```

```
## [1] 101
```

```
(which(BD.tot$Cumulative>0.8)[1])/nrow(BD)
```

```
## [1] 0.2376471
```