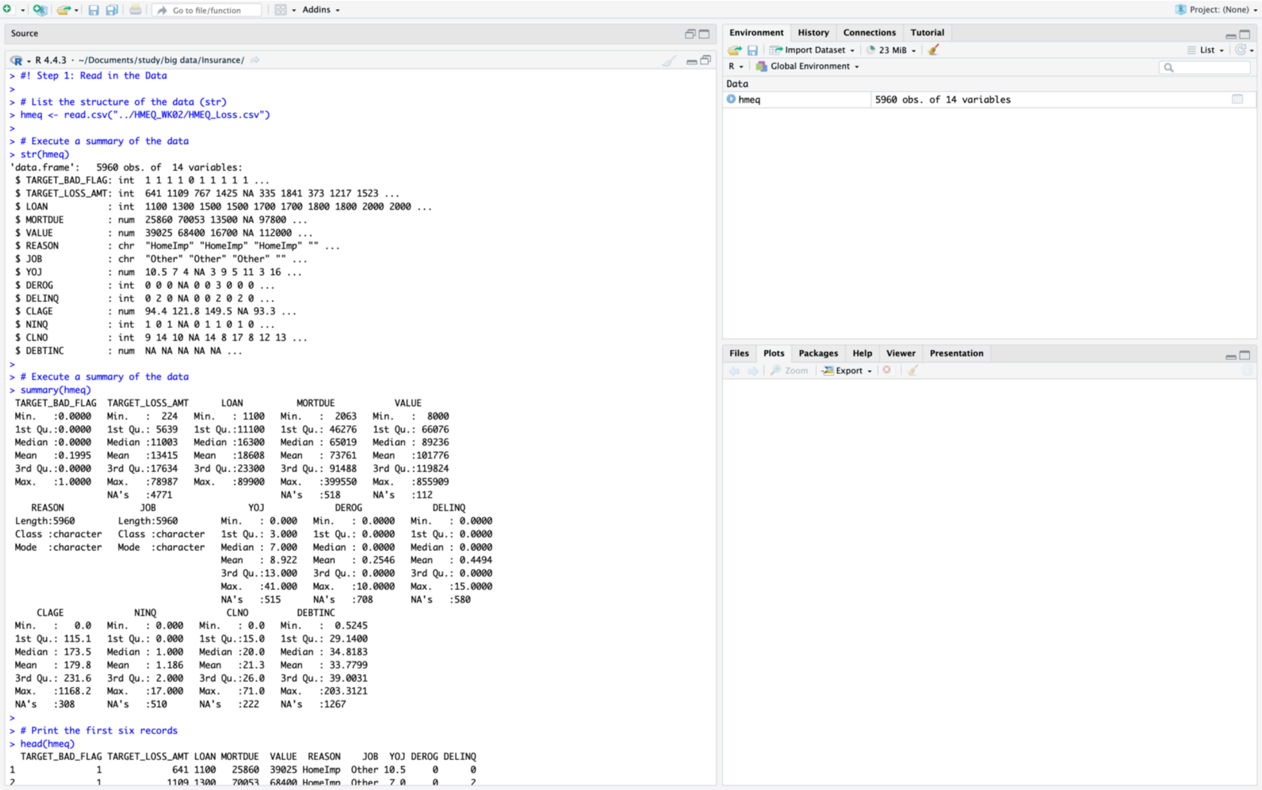
Shao Yun Gao–Week 2: R Output

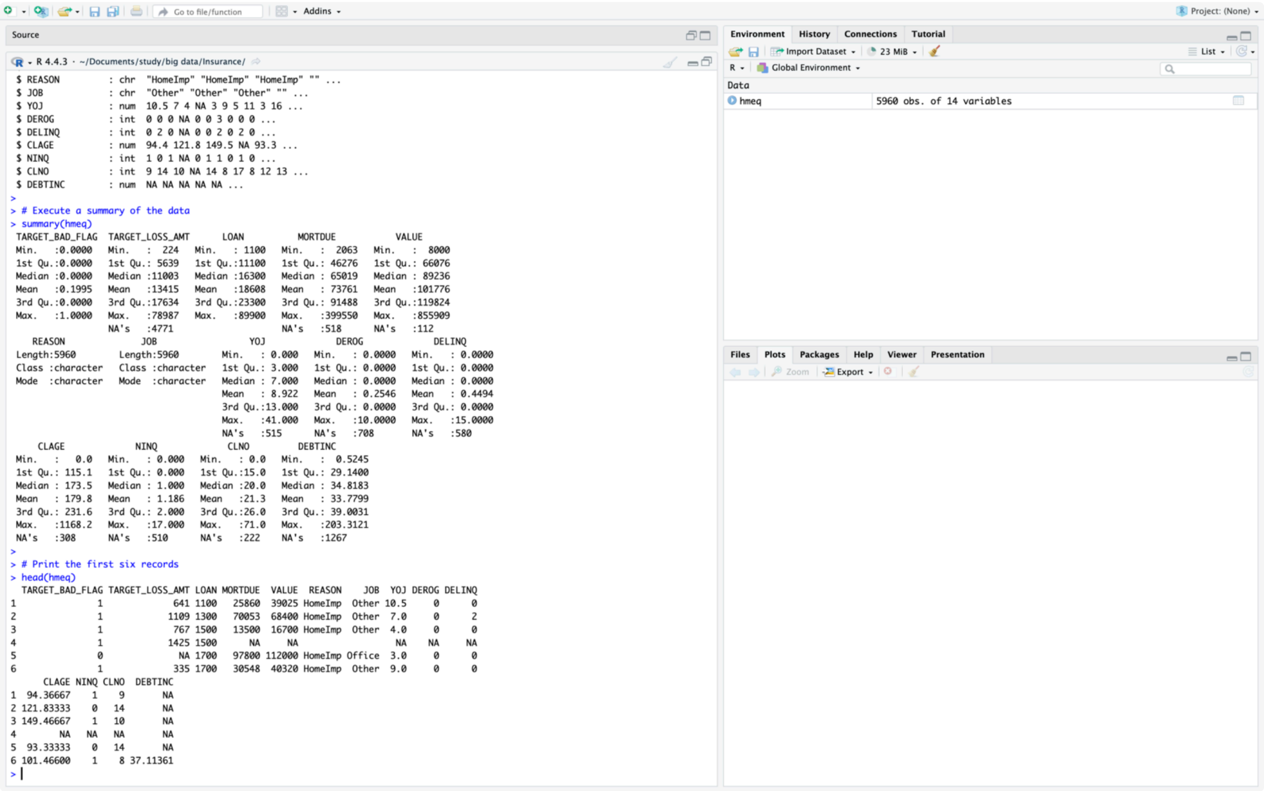
Step 1: Read in the Data

```R
#! Step 1: Read in the Data

# List the structure of the data (str)
hmeq <- read.csv("../HMEQ_WK02/HMEQ_Loss.csv")

# Execute a summary of the data
str(hmeq)

# Execute a summary of the data
summary(hmeq)

# Print the first six records
head(hmeq)
```



```
> #! Step 1: Read in the Data
>
> # List the structure of the data (str)
> hmeq <- read.csv("../HMEQ_WK02/HMEQ_Loss.csv")
>
> # Execute a summary of the data
> str(hmeq)
'data.frame':	5960 obs. of  14 variables:
 $ TARGET_BAD_FLAG: int  1 1 1 1 0 1 1 1 1 1 ...
 $ TARGET_LOSS_AMT: int  641 1109 767 1425 NA 335 1841 373 1217 1523 ...
 $ LOAN           : int  1100 1300 1500 1500 1700 1700 1800 1800 2000 2000 ...
 $ MORTDUE        : num  25860 70053 13500 NA 97800 ...
 $ VALUE          : num  39025 68400 16700 NA 112000 ...
 $ REASON         : chr  "HomeImp" "HomeImp" "HomeImp" "" ...
 $ JOB            : chr  "Other" "Other" "Other" "" ...
 $ YOJ            : num  10.5 7 4 NA 3 9 5 11 3 16 ...
 $ DEROG          : int  0 0 0 NA 0 0 3 0 0 0 ...
 $ DELINQ         : int  0 2 0 NA 0 0 2 0 2 0 ...
 $ CLAGE          : num  94.4 121.8 149.5 NA 93.3 ...
 $ NINQ           : int  1 0 1 NA 0 1 1 0 1 0 ...
 $ CLNO           : int  9 14 10 NA 14 8 17 8 12 13 ...
 $ DEBTINC        : num  NA NA NA NA NA ...
>
> # Execute a summary of the data
> summary(hmeq)
 TARGET_BAD_FLAG  TARGET_LOSS_AMT      LOAN          MORTDUE           VALUE       
 Min.   :0.0000   Min.   :  224   Min.   : 1100   Min.   :  2063   Min.   :  8000  
 1st Qu.:0.0000   1st Qu.: 5639   1st Qu.:11100   1st Qu.: 46276   1st Qu.: 66076  
 Median :0.0000   Median :11003   Median :16300   Median : 65019   Median : 89236  
 Mean   :0.1995   Mean   :13415   Mean   :18608   Mean   : 73761   Mean   :101776  
 3rd Qu.:0.0000   3rd Qu.:17634   3rd Qu.:23300   3rd Qu.: 91488   3rd Qu.:119824  
 Max.   :1.0000   Max.   :78987   Max.   :89900   Max.   :399550   Max.   :855909  
                  NA's   :4771                    NA's   :518      NA's   :112     
    REASON              JOB                 YOJ             DEROG             DELINQ       
 Length:5960        Length:5960        Min.   : 0.000   Min.   : 0.0000   Min.   : 0.0000  
 Class :character   Class :character   1st Qu.: 3.000   1st Qu.: 0.0000   1st Qu.: 0.0000  
 Mode  :character   Mode  :character   Median : 7.000   Median : 0.0000   Median : 0.0000  
                                       Mean   : 8.922   Mean   : 0.2546   Mean   : 0.4494  
                                       3rd Qu.:13.000   3rd Qu.: 0.0000   3rd Qu.: 0.0000  
                                       Max.   :41.000   Max.   :10.0000   Max.   :15.0000  
                                       NA's   :515      NA's   :708       NA's   :580      
     CLAGE             NINQ             CLNO         DEBTINC        
 Min.   :   0.0   Min.   : 0.000   Min.   : 0.0   Min.   :  0.5245  
 1st Qu.: 115.1   1st Qu.: 0.000   1st Qu.:15.0   1st Qu.: 29.1400  
 Median : 173.5   Median : 1.000   Median :20.0   Median : 34.8183  
 Mean   : 179.8   Mean   : 1.186   Mean   :21.3   Mean   : 33.7799  
 3rd Qu.: 231.6   3rd Qu.: 2.000   3rd Qu.:26.0   3rd Qu.: 39.0031  
 Max.   :1168.2   Max.   :17.000   Max.   :71.0   Max.   :203.3121  
 NA's   :308      NA's   :510      NA's   :222    NA's   :1267      
>
> # Print the first six records
> head(hmeq)
  TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN MORTDUE  VALUE  REASON   JOB  YOJ DEROG DELINQ
1               1             641 1100   25860  39025 HomeImp Other 10.5     0      0
2               1            1109 1300   70053  68400 HomeImp Other  7.0     0      2
```

Data
hmeq	5960 obs. of 14 variables



```
 $ REASON        : chr  "HomeImp" "HomeImp" "HomeImp" "" ...
 $ JOB           : chr  "Other" "Other" "Other" "" ...
 $ YOJ           : num  10.5 7 4 NA 3 9 5 11 3 16 ...
 $ DEROG         : int  0 0 0 NA 0 0 3 0 0 0 ...
 $ DELINQ        : int  0 2 0 NA 0 0 2 0 2 0 ...
 $ CLAGE         : num  94.4 121.8 149.5 NA 93.3 ...
 $ NINQ          : int  1 0 1 NA 0 1 1 0 1 0 ...
 $ CLNO          : int  9 14 10 NA 14 8 17 8 12 13 ...
 $ DEBTINC       : num  NA NA NA NA NA ...
>
> # Execute a summary of the data
> summary(hmeq)
 TARGET_BAD_FLAG  TARGET_LOSS_AMT      LOAN          MORTDUE           VALUE
 Min.   :0.0000   Min.   :  224   Min.   : 1100   Min.   :  2063   Min.   :  8000
 1st Qu.:0.0000   1st Qu.: 5639   1st Qu.:11100   1st Qu.: 46276   1st Qu.: 66076
 Median :0.0000   Median :11003   Median :16300   Median : 65019   Median : 89236
 Mean   :0.1995   Mean   :13415   Mean   :18608   Mean   : 73761   Mean   :101776
 3rd Qu.:0.0000   3rd Qu.:17634   3rd Qu.:23300   3rd Qu.: 91488   3rd Qu.:119824
 Max.   :1.0000   Max.   :78987   Max.   :89900   Max.   :399550   Max.   :855909
                  NA's   :4771                    NA's   :518      NA's   :112
   REASON              JOB                 YOJ             DEROG             DELINQ
 Length:5960        Length:5960        Min.   : 0.000   Min.   : 0.0000   Min.   : 0.0000
 Class :character   Class :character   1st Qu.: 3.000   1st Qu.: 0.0000   1st Qu.: 0.0000
 Mode  :character   Mode  :character   Median : 7.000   Median : 0.0000   Median : 0.0000
                                       Mean   : 8.922   Mean   : 0.2546   Mean   : 0.4494
                                       3rd Qu.:13.000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
                                       Max.   :41.000   Max.   :10.0000   Max.   :15.0000
                                       NA's   :515      NA's   :708       NA's   :580
     CLAGE            NINQ             CLNO          DEBTINC
 Min.   :   0.0   Min.   : 0.000   Min.   : 0.0   Min.   :  0.5245
 1st Qu.: 115.1   1st Qu.: 0.000   1st Qu.:15.0   1st Qu.: 29.1400
 Median : 173.5   Median : 1.000   Median :20.0   Median : 34.8183
 Mean   : 179.8   Mean   : 1.186   Mean   :21.3   Mean   : 33.7799
 3rd Qu.: 231.6   3rd Qu.: 2.000   3rd Qu.:26.0   3rd Qu.: 39.0031
 Max.   :1168.2   Max.   :17.000   Max.   :71.0   Max.   :203.3121
 NA's   :308      NA's   :510      NA's   :222    NA's   :1267
>
> # Print the first six records
> head(hmeq)
  TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN MORTDUE  VALUE  REASON    JOB  YOJ DEROG DELINQ
1               1             641 1100   25860  39025 HomeImp  Other 10.5     0      0
2               1            1109 1300   70053  68400 HomeImp  Other  7.0     0      2
3               1             767 1500   13500  16700 HomeImp  Other  4.0     0      0
4               1            1425 1500      NA     NA                  NA    NA     NA
5               0              NA 1700   97800 112000 HomeImp Office  3.0     0      0
6               1             335 1700   30548  40320 HomeImp  Other  9.0     0      0
      CLAGE NINQ CLNO  DEBTINC
1  94.36667    1    9       NA
2 121.83333    0   14       NA
3 149.46667    1   10       NA
4        NA   NA   NA       NA
5  93.33333    0   14       NA
6 101.46600    1    8 37.11361
> |
```

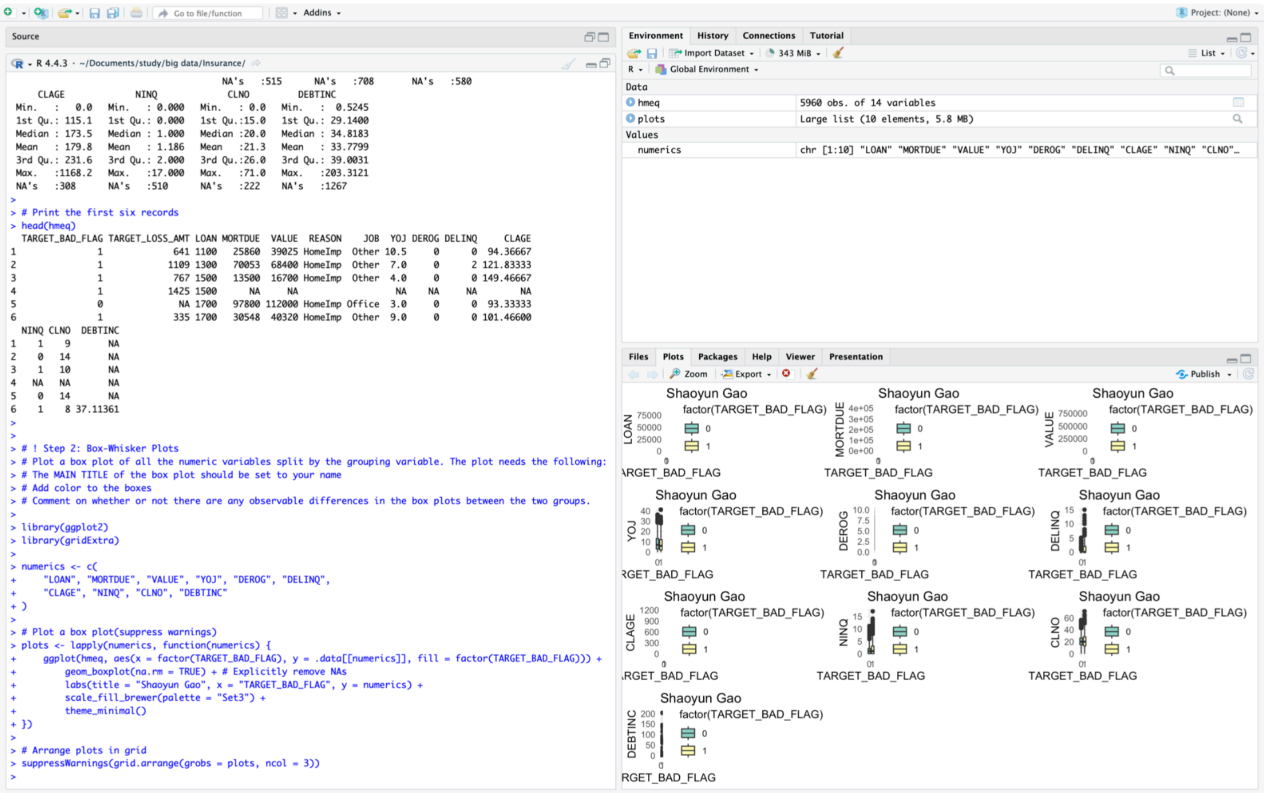
hmeq   5960 obs. of 14 variables

Step 2: Box–Whisker Plots

R

```r
# ! Step 2: Box-Whisker Plots
# Plot a box plot of all the numeric variables split by the grouping variable. The plot needs the following:
# The MAIN TITLE of the box plot should be set to your name
# Add color to the boxes
# Comment on whether or not there are any observable differences in the box plots between the two groups.

library(ggplot2)
library(gridExtra)

numerics <- c(
  "LOAN", "MORTDUE", "VALUE", "YOJ", "DEROG", "DELINQ",
  "CLAGE", "NINQ", "CLNO", "DEBTINC"
)

# Plot a box plot(suppress warnings)
plots <- lapply(numerics, function(numerics) {
  ggplot(hmeq, aes(x = factor(TARGET_BAD_FLAG), y = .data[[numerics]], fill = factor(TARGET_BAD_FLAG))) +
      geom_boxplot(na.rm = TRUE) + # Explicitly remove NAs
      labs(title = "Shaoyun Gao", x = "TARGET_BAD_FLAG", y = numerics) +
      scale_fill_brewer(palette = "Set3") +
      theme_minimal()
})
```

```
# Arrange plots in grid
suppressWarnings(grid.arrange(grobs = plots, ncol = 3))
```



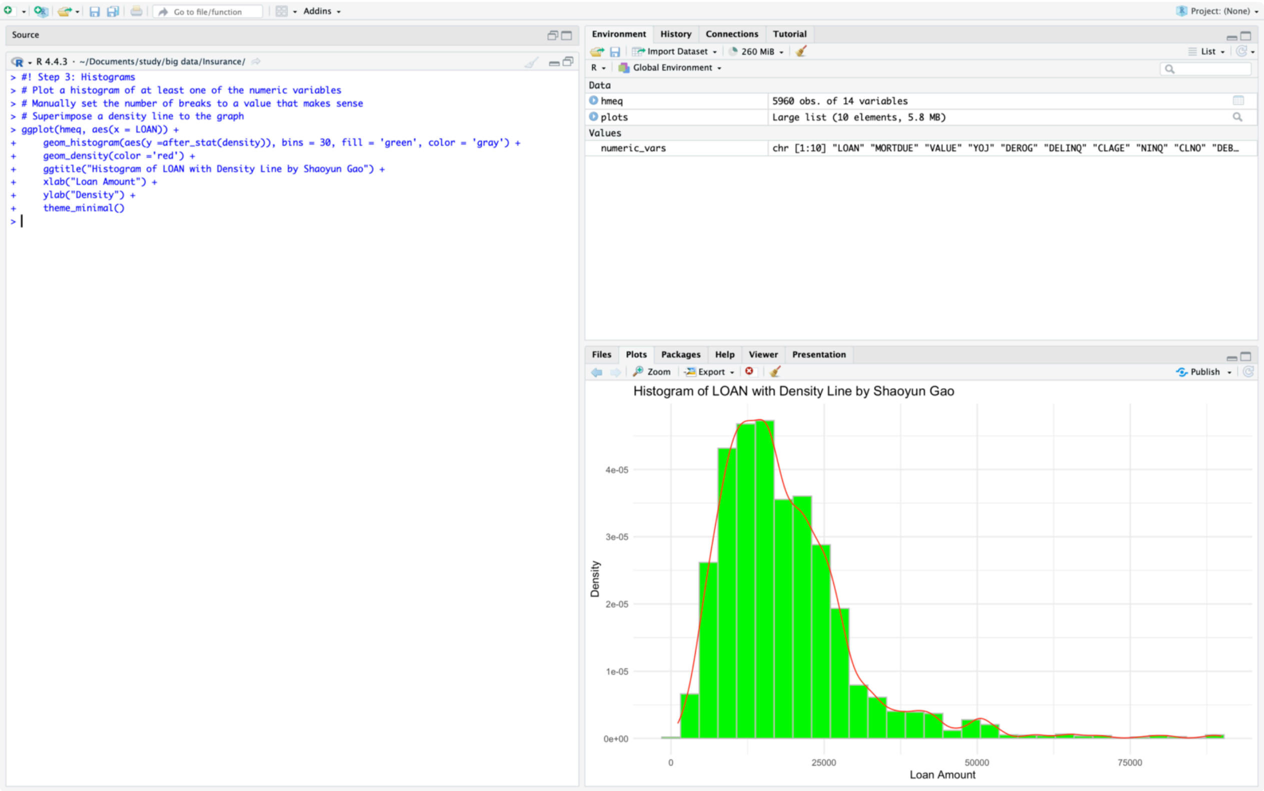
```
                                        NA's   :515      NA's   :708       NA's   :580  
     CLAGE             NINQ             CLNO          DEBTINC       
 Min.   :   0.0   Min.   : 0.000   Min.   : 0.0   Min.   :  0.5245  
 1st Qu.: 115.1   1st Qu.: 0.000   1st Qu.:15.0   1st Qu.: 29.1400  
 Median : 173.5   Median : 1.000   Median :20.0   Median : 34.8183  
 Mean   : 179.8   Mean   : 1.186   Mean   :21.3   Mean   : 33.7799  
 3rd Qu.: 231.6   3rd Qu.: 2.000   3rd Qu.:26.0   3rd Qu.: 39.0031  
 Max.   :1168.2   Max.   :17.000   Max.   :71.0   Max.   :203.3121  
 NA's   :308      NA's   :510      NA's   :222    NA's   :1267      
>
> # Print the first six records
> head(hmeq)
  TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN MORTDUE  VALUE  REASON    JOB  YOJ DEROG DELINQ     CLAGE
1               1             641 1100   25860  39025 HomeImp  Other 10.5     0      0  94.36667
2               1            1109 1300   70053  68400 HomeImp  Other  7.0     0      2 121.83333
3               1             767 1500   13500  16700 HomeImp  Other  4.0     0      0 149.46667
4               1            1425 1500      NA     NA                  NA    NA     NA        NA
5               0              NA 1700   97800 112000 HomeImp Office  3.0     0      0  93.33333
6               1             335 1700   30548  40320 HomeImp  Other  9.0     0      0 101.46600
  NINQ CLNO  DEBTINC
1    1    9       NA
2    0   14       NA
3    1   10       NA
4   NA   NA       NA
5    0   14       NA
6    1    8 37.11361
>
>
> # ! Step 2: Box-Whisker Plots
> # Plot a box plot of all the numeric variables split by the grouping variable. The plot needs the following:
> # The MAIN TITLE of the box plot should be set to your name
> # Add color to the boxes
> # Comment on whether or not there are any observable differences in the box plots between the two groups.
>
> library(ggplot2)
> library(gridExtra)
>
> numerics <- c(
+     "LOAN", "MORTDUE", "VALUE", "YOJ", "DEROG", "DELINQ",
+     "CLAGE", "NINQ", "CLNO", "DEBTINC"
+ )
>
> # Plot a box plot(suppress warnings)
> plots <- lapply(numerics, function(numerics) {
+     ggplot(hmeq, aes(x = factor(TARGET_BAD_FLAG), y = .data[[numerics]], fill = factor(TARGET_BAD_FLAG))) +
+         geom_boxplot(na.rm = TRUE) + # Explicitly remove NAs
+         labs(title = "Shaoyun Gao", x = "TARGET_BAD_FLAG", y = numerics) +
+         scale_fill_brewer(palette = "Set3") +
+         theme_minimal()
+ })
>
> # Arrange plots in grid
> suppressWarnings(grid.arrange(grobs = plots, ncol = 3))
>
```

```
Data
hmeq        5960 obs. of 14 variables
plots       Large list (10 elements, 5.8 MB)
Values
numerics    chr [1:10] "LOAN" "MORTDUE" "VALUE" "YOJ" "DEROG" "DELINQ" "CLAGE" "NINQ" "CLNO"…
```

Step 3: Histograms

R

```
#! Step 3: Histograms
# Plot a histogram of at least one of the numeric variables
# Manually set the number of breaks to a value that makes sense
# Superimpose a density line to the graph
ggplot(hmeq, aes(x = LOAN)) +
  geom_histogram(aes(y =after_stat(density)), bins = 30, fill = 'green', color = 'gray') +
  geom_density(color ='red') +
  ggtitle("Histogram of LOAN with Density Line by Shaoyun Gao") +
  xlab("Loan Amount") +
  ylab("Density") +
  theme_minimal()
```



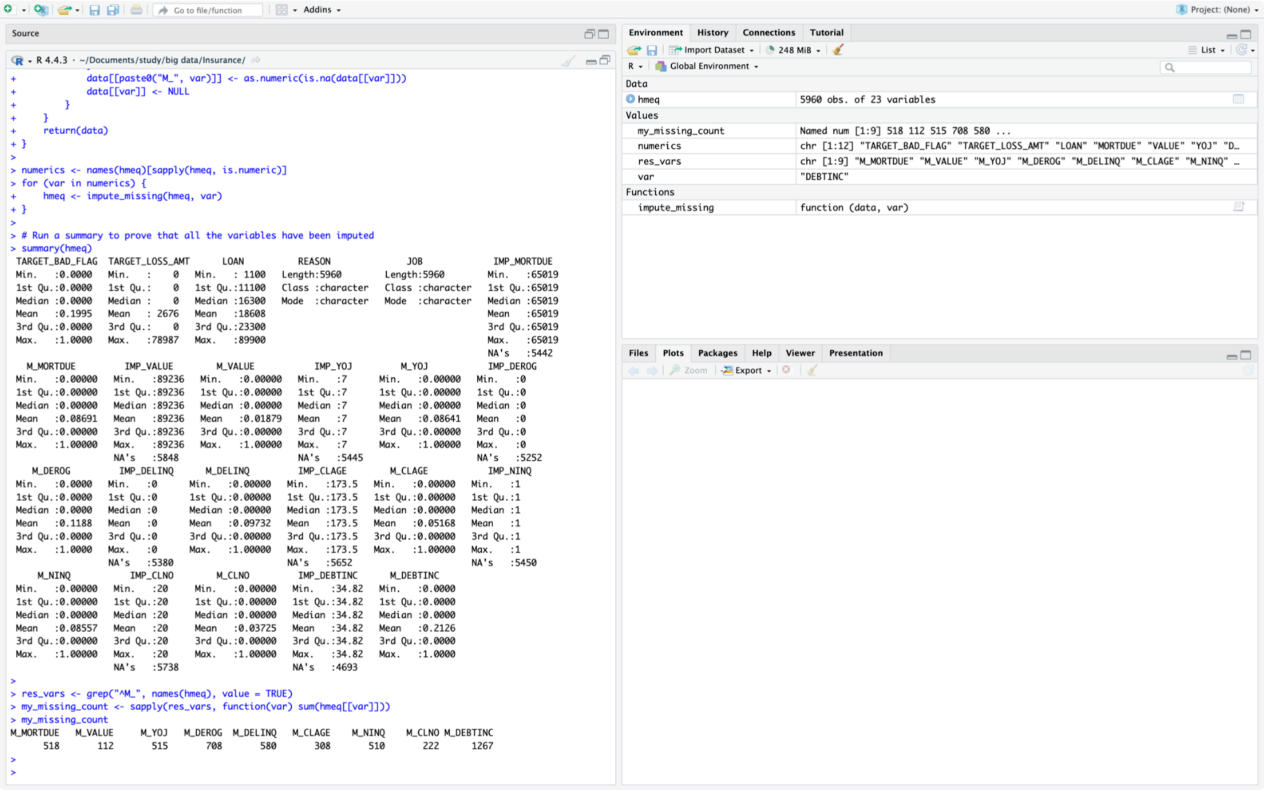
```
> #! Step 3: Histograms
> # Plot a histogram of at least one of the numeric variables
> # Manually set the number of breaks to a value that makes sense
> # Superimpose a density line to the graph
> ggplot(hmeq, aes(x = LOAN)) +
+     geom_histogram(aes(y =after_stat(density)), bins = 30, fill = 'green', color = 'gray') +
+     geom_density(color ='red') +
+     ggtitle("Histogram of LOAN with Density Line by Shaoyun Gao") +
+     xlab("Loan Amount") +
+     ylab("Density") +
+     theme_minimal()
> |
```

```
Data
hmeq          5960 obs. of 14 variables
plots         Large list (10 elements, 5.8 MB)
Values
numeric_vars  chr [1:10] "LOAN" "MORTDUE" "VALUE" "YOJ" "DEROG" "DELINQ" "CLAGE" "NINQ" "CLNO" "DEB…
```

Step 4: Impute "Fix" all the numeric variables that have missing values

R

```
# ! Step 4: Impute "Fix" all the numeric variables that have missing values
# For the missing Target variables, simply set the missing values to zero
# For the remaining numeric variables with missing values, create two new variables. One variable will have a name beginning with IMP_ and it will contained the imputed value. The second value will have a name beginning with M_ and it will contain a 1 if the record was imputed and a zero if it was not.
# You may impute with any method that makes sense. The median or mean value will be useful in most cases.
# Push yourself! Try one complex imputation like the one described in the lectures.
# Delete the original variable after it has been imputed.

impute_missing <- function(data, var) {
  if (any(is.na(data[[var]]))) {
    if (var == "TARGET_LOSS_AMT") {
      data[[var]][is.na(data[[var]])] <- 0
    } else {
      if (var %in% c("INCOME", "HOME_VAL")) {
        job_groups <- split(data, data$JOB)
        for (group in names(job_groups)) {
          median_value <- median(job_groups[[group]][[var]], na.rm = TRUE)
          data[[paste0("IMP_", var)]][data$JOB == group & is.na(data[[var]])] <- median_value
        }
      } else {
        median_value <- median(data[[var]], na.rm = TRUE)
        data[[paste0("IMP_", var)]][is.na(data[[var]])] <- median_value
      }
      data[[paste0("M_", var)]] <- as.numeric(is.na(data[[var]]))
      data[[var]] <- NULL
    }
  }
  return(data)
}

numerics <- names(hmeq)[sapply(hmeq, is.numeric)]
for (var in numerics) {
  hmeq <- impute_missing(hmeq, var)
}

# Run a summary to prove that all the variables have been imputed
summary(hmeq)

res_vars <- grep("^M_", names(hmeq), value = TRUE)
my_missing_count <- sapply(res_vars, function(var) sum(hmeq[[var]]))
my_missing_count
```



```
+             data[[paste0("M_", var)]] <- as.numeric(is.na(data[[var]]))
+             data[[var]] <- NULL
+         }
+     }
+     return(data)
+ }
>
> numerics <- names(hmeq)[sapply(hmeq, is.numeric)]
> for (var in numerics) {
+     hmeq <- impute_missing(hmeq, var)
+ }
>
> # Run a summary to prove that all the variables have been imputed
> summary(hmeq)
 TARGET_BAD_FLAG  TARGET_LOSS_AMT      LOAN           REASON              JOB             IMP_MORTDUE
 Min.   :0.0000   Min.   :    0   Min.   : 1100   Length:5960        Length:5960        Min.   :65019
 1st Qu.:0.0000   1st Qu.:    0   1st Qu.:11100   Class :character   Class :character   1st Qu.:65019
 Median :0.0000   Median :    0   Median :16300   Mode  :character   Mode  :character   Median :65019
 Mean   :0.1995   Mean   : 2676   Mean   :18608                                         Mean   :65019
 3rd Qu.:0.0000   3rd Qu.:    0   3rd Qu.:23300                                         3rd Qu.:65019
 Max.   :1.0000   Max.   :78987   Max.   :89900                                         Max.   :65019
                                                                                        NA's   :5442
   M_MORTDUE          IMP_VALUE        M_VALUE            IMP_YOJ          M_YOJ            IMP_DEROG
 Min.   :0.00000   Min.   :89236   Min.   :0.00000   Min.   :7       Min.   :0.00000   Min.   :0
 1st Qu.:0.00000   1st Qu.:89236   1st Qu.:0.00000   1st Qu.:7       1st Qu.:0.00000   1st Qu.:0
 Median :0.00000   Median :89236   Median :0.00000   Median :7       Median :0.00000   Median :0
 Mean   :0.08691   Mean   :89236   Mean   :0.01879   Mean   :7       Mean   :0.08641   Mean   :0
 3rd Qu.:0.00000   3rd Qu.:89236   3rd Qu.:0.00000   3rd Qu.:7       3rd Qu.:0.00000   3rd Qu.:0
 Max.   :1.00000   Max.   :89236   Max.   :1.00000   Max.   :7       Max.   :1.00000   Max.   :0
                   NA's   :5848                      NA's   :5445                      NA's   :5252
    M_DEROG         IMP_DELINQ       M_DELINQ          IMP_CLAGE         M_CLAGE           IMP_NINQ
 Min.   :0.0000   Min.   :0       Min.   :0.00000   Min.   :173.5   Min.   :0.00000   Min.   :1
 1st Qu.:0.0000   1st Qu.:0       1st Qu.:0.00000   1st Qu.:173.5   1st Qu.:0.00000   1st Qu.:1
 Median :0.0000   Median :0       Median :0.00000   Median :173.5   Median :0.00000   Median :1
 Mean   :0.1188   Mean   :0       Mean   :0.09732   Mean   :173.5   Mean   :0.05168   Mean   :1
 3rd Qu.:0.0000   3rd Qu.:0       3rd Qu.:0.00000   3rd Qu.:173.5   3rd Qu.:0.00000   3rd Qu.:1
 Max.   :1.0000   Max.   :0       Max.   :1.00000   Max.   :173.5   Max.   :1.00000   Max.   :1
                  NA's   :5380                      NA's   :5652                      NA's   :5450
     M_NINQ          IMP_CLNO        M_CLNO          IMP_DEBTINC       M_DEBTINC
 Min.   :0.00000   Min.   :20      Min.   :0.00000   Min.   :34.82   Min.   :0.0000
 1st Qu.:0.00000   1st Qu.:20      1st Qu.:0.00000   1st Qu.:34.82   1st Qu.:0.0000
 Median :0.00000   Median :20      Median :0.00000   Median :34.82   Median :0.0000
 Mean   :0.08557   Mean   :20      Mean   :0.03725   Mean   :34.82   Mean   :0.2126
 3rd Qu.:0.00000   3rd Qu.:20      3rd Qu.:0.00000   3rd Qu.:34.82   3rd Qu.:0.0000
 Max.   :1.00000   Max.   :20      Max.   :1.00000   Max.   :34.82   Max.   :1.0000
                   NA's   :5738                      NA's   :4693
>
> res_vars <- grep("^M_", names(hmeq), value = TRUE)
> my_missing_count <- sapply(res_vars, function(var) sum(hmeq[[var]]))
> my_missing_count
M_MORTDUE   M_VALUE     M_YOJ   M_DEROG  M_DELINQ   M_CLAGE    M_NINQ    M_CLNO M_DEBTINC
      518       112       515       708       580       308       510       222      1267
>
>
```

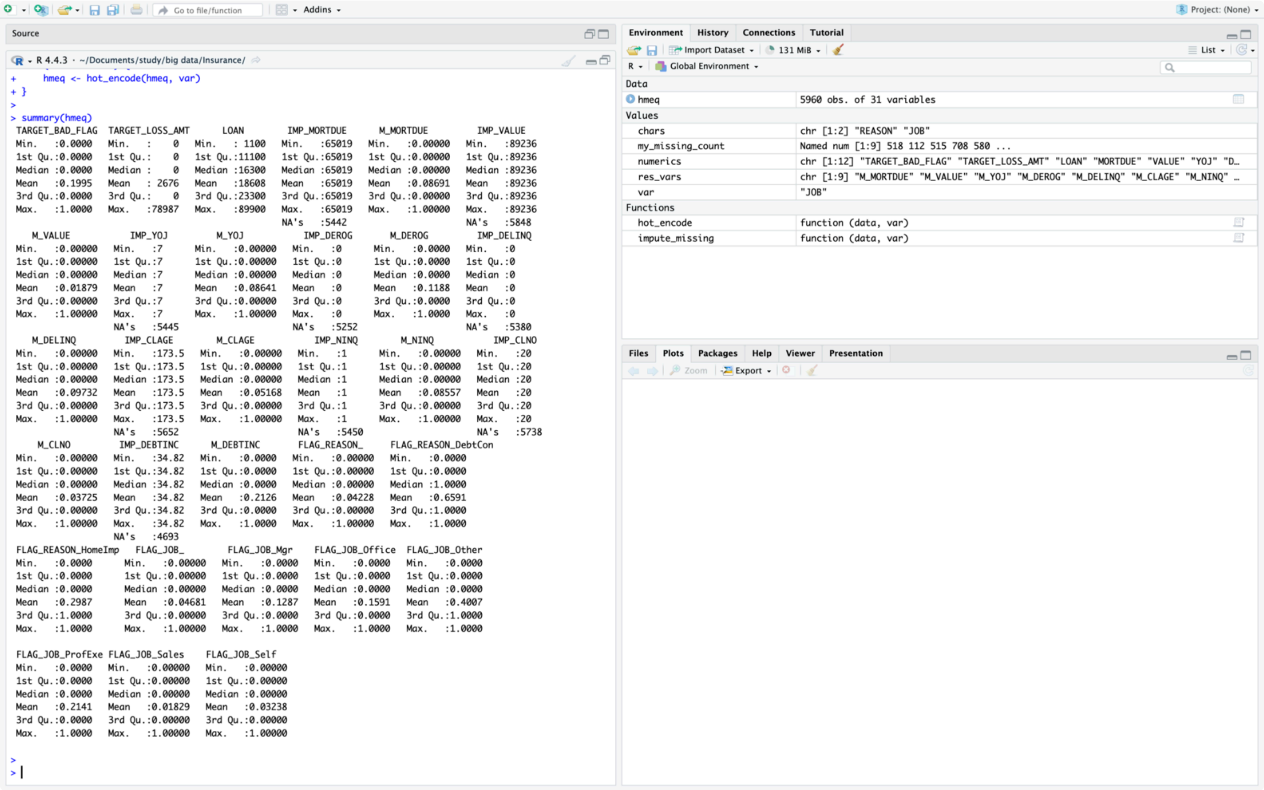
```
Data
hmeq                5960 obs. of 23 variables
Values
  my_missing_count  Named num [1:9] 518 112 515 708 580 ...
  numerics          chr [1:12] "TARGET_BAD_FLAG" "TARGET_LOSS_AMT" "LOAN" "MORTDUE" "VALUE" "YOJ" "D…
  res_vars          chr [1:9] "M_MORTDUE" "M_VALUE" "M_YOJ" "M_DEROG" "M_DELINQ" "M_CLAGE" "M_NINQ" …
  var               "DEBTINC"
Functions
  impute_missing    function (data, var)
```

Step 5: One Hot Encoding

R

```
# ! Step 5: One Hot Encoding
# For the character / category variables, perform one hot encoding. For this create a
Flag for each categories.
# Delete the original class variable
```

```r
# Run a summary to show that the category variables have been replaced by Flag variables.

hot_encode <- function(data, var) {
  if (is.factor(data[[var]]) || is.character(data[[var]])) {
    levels <- levels(as.factor(data[[var]]))
    for (level in levels) {
      data[[paste0("FLAG_", var, "_", level)]] <- as.numeric(data[[var]] == level)
    }
    data[[var]] <- NULL
  }
  return(data)
}

chars <- names(hmeq)[sapply(hmeq, function(x) is.factor(x) || is.character(x))]
for (var in chars) {
  hmeq <- hot_encode(hmeq, var)
}

summary(hmeq)
```



```
+     hmeq <- hot_encode(hmeq, var)
+ }
>
> summary(hmeq)
 TARGET_BAD_FLAG  TARGET_LOSS_AMT      LOAN        IMP_MORTDUE      M_MORTDUE         IMP_VALUE    
 Min.   :0.0000   Min.   :    0   Min.   : 1100   Min.   :65019   Min.   :0.00000   Min.   :89236  
 1st Qu.:0.0000   1st Qu.:    0   1st Qu.:11100   1st Qu.:65019   1st Qu.:0.00000   1st Qu.:89236  
 Median :0.0000   Median :    0   Median :16300   Median :65019   Median :0.00000   Median :89236  
 Mean   :0.1995   Mean   : 2676   Mean   :18608   Mean   :65019   Mean   :0.08691   Mean   :89236  
 3rd Qu.:0.0000   3rd Qu.:    0   3rd Qu.:23300   3rd Qu.:65019   3rd Qu.:0.00000   3rd Qu.:89236  
 Max.   :1.0000   Max.   :78987   Max.   :89900   Max.   :65019   Max.   :1.00000   Max.   :89236  
                                                  NA's   :5442                      NA's   :5848   
    M_VALUE            IMP_YOJ         M_YOJ            IMP_DEROG        M_DEROG         IMP_DELINQ   
 Min.   :0.00000   Min.   :7      Min.   :0.00000   Min.   :0      Min.   :0.0000   Min.   :0      
 1st Qu.:0.00000   1st Qu.:7      1st Qu.:0.00000   1st Qu.:0      1st Qu.:0.0000   1st Qu.:0      
 Median :0.00000   Median :7      Median :0.00000   Median :0      Median :0.0000   Median :0      
 Mean   :0.01879   Mean   :7      Mean   :0.08641   Mean   :0      Mean   :0.1188   Mean   :0      
 3rd Qu.:0.00000   3rd Qu.:7      3rd Qu.:0.00000   3rd Qu.:0      3rd Qu.:0.0000   3rd Qu.:0      
 Max.   :1.00000   Max.   :7      Max.   :1.00000   Max.   :0      Max.   :1.0000   Max.   :0      
                   NA's   :5445                     NA's   :5252                    NA's   :5380   
    M_DELINQ         IMP_CLAGE         M_CLAGE            IMP_NINQ         M_NINQ            IMP_CLNO    
 Min.   :0.00000   Min.   :173.5   Min.   :0.00000   Min.   :1       Min.   :0.00000   Min.   :20     
 1st Qu.:0.00000   1st Qu.:173.5   1st Qu.:0.00000   1st Qu.:1       1st Qu.:0.00000   1st Qu.:20     
 Median :0.00000   Median :173.5   Median :0.00000   Median :1       Median :0.00000   Median :20     
 Mean   :0.09732   Mean   :173.5   Mean   :0.05168   Mean   :1       Mean   :0.08557   Mean   :20     
 3rd Qu.:0.00000   3rd Qu.:173.5   3rd Qu.:0.00000   3rd Qu.:1       3rd Qu.:0.00000   3rd Qu.:20     
 Max.   :1.00000   Max.   :173.5   Max.   :1.00000   Max.   :1       Max.   :1.00000   Max.   :20     
                   NA's   :5652                      NA's   :5450                      NA's   :5738   
     M_CLNO          IMP_DEBTINC      M_DEBTINC        FLAG_REASON_      FLAG_REASON_DebtCon
 Min.   :0.00000   Min.   :34.82   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000     
 1st Qu.:0.00000   1st Qu.:34.82   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000     
 Median :0.00000   Median :34.82   Median :0.0000   Median :0.00000   Median :1.0000     
 Mean   :0.03725   Mean   :34.82   Mean   :0.2126   Mean   :0.04228   Mean   :0.6591     
 3rd Qu.:0.00000   3rd Qu.:34.82   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:1.0000     
 Max.   :1.00000   Max.   :34.82   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000     
                   NA's   :4693                                                          
 FLAG_REASON_HomeImp   FLAG_JOB_          FLAG_JOB_Mgr     FLAG_JOB_Office  FLAG_JOB_Other  
 Min.   :0.0000      Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:0.0000      1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :0.0000      Median :0.00000   Median :0.0000   Median :0.0000   Median :0.0000  
 Mean   :0.2987      Mean   :0.04681   Mean   :0.1287   Mean   :0.1591   Mean   :0.4007  
 3rd Qu.:1.0000      3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000  
 Max.   :1.0000      Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000  
                                                                                         
 FLAG_JOB_ProfExe FLAG_JOB_Sales    FLAG_JOB_Self    
 Min.   :0.0000   Min.   :0.00000   Min.   :0.00000  
 1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000  
 Median :0.0000   Median :0.00000   Median :0.00000  
 Mean   :0.2141   Mean   :0.01829   Mean   :0.03238  
 3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000  
 Max.   :1.0000   Max.   :1.00000   Max.   :1.00000  

>
> |
```

Data	
hmeq	5960 obs. of 31 variables
Values	
chars	chr [1:2] "REASON" "JOB"
my_missing_count	Named num [1:9] 518 112 515 708 580 ...
numerics	chr [1:12] "TARGET_BAD_FLAG" "TARGET_LOSS_AMT" "LOAN" "MORTDUE" "VALUE" "YOJ" "D…
res_vars	chr [1:9] "M_MORTDUE" "M_VALUE" "M_YOJ" "M_DEROG" "M_DELINQ" "M_CLAGE" "M_NINQ" …
var	"JOB"
Functions	
hot_encode	function (data, var)
impute_missing	function (data, var)