R Visualization Capabilities Introduction

Pramod Verma, Data For Excellence (DFE), GPE, PSA International

24 September 2019

Use of this document

This file is created for the sole use of PSA Data Analytics Technical Workshop participants for demonstration and learning about the R visualization capabilities.

Scripting language used

This document is created using R Markdown, a scripting language available as open source from R Foundation.

Loading all the required packages

```
#install.packages("dplyr",repos = "http://cran.us.r-project.org")
library(dplyr)

#install.packages("ggcorplot",repos = "http://cran.us.r-project.org")
library(ggcorrplot)

#install.packages("ggplot2",repos = "http://cran.us.r-project.org")
library(ggplot2)
```

May need to load more libraries/packages depending on local computer/server

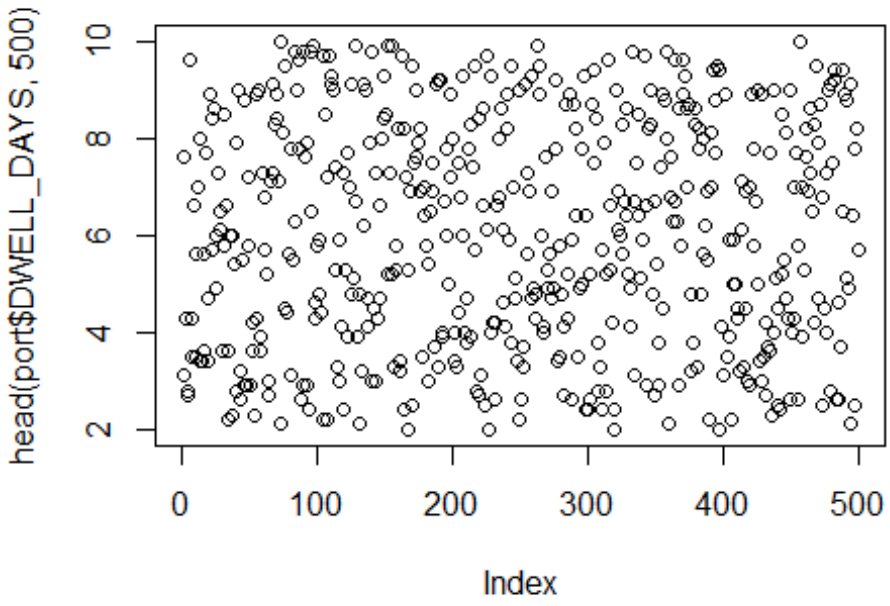
Loading the file into R data-frame

```
#Reading the csv file
port = read.csv('D:/Data Analytics Workshop/Data Analytics Technical Workshop Singapore/Data/port.csv')
```

Visualization using ggplot library

Scatter Plot

```
plot(head(port$DWELL_DAYS,500))
```



Plotting of categorical and continouous variable

```
# specify dataset and mapping
table(port$COMMODITY)
```

```
##
##              Adhesive               Biscuit              Chemicals
##                    28                    24                   172
##       Confectionaries                 Dairy                 Diary
##                    45                    35                   189
##                 Drink                  Eggs                 Films
##                   118                    24                    22
##               Flowers                Fruits                 Glass
##                  2651                  3612                    22
##             Hydroxide            Instrument                 Juice
##                    28                    26                    28
##                  Meat                   Nut                Others
##                  1834                    26                    23
##               Pharma               Poultry Processed/Cooked Food
##                   118                   738                   128
##             Sea Food               Tobacco             Vegetables
##                 35329                    19                  4710
##              Waffles
##                    31
```

```
#Display as data frame
as.data.frame(table(port$COMMODITY))
```

```
##                    Var1  Freq
## 1              Adhesive    28
## 2               Biscuit    24
## 3             Chemicals   172
## 4       Confectionaries    45
## 5                 Dairy    35
## 6                 Diary   189
## 7                 Drink   118
## 8                  Eggs    24
## 9                 Films    22
## 10              Flowers  2651
## 11               Fruits  3612
## 12                Glass    22
## 13            Hydroxide    28
## 14           Instrument    26
## 15                Juice    28
## 16                 Meat  1834
## 17                  Nut    26
## 18               Others    23
## 19               Pharma   118
## 20              Poultry   738
## 21 Processed/Cooked Food   128
## 22             Sea Food 35329
## 23              Tobacco    19
## 24           Vegetables  4710
## 25              Waffles    31
```

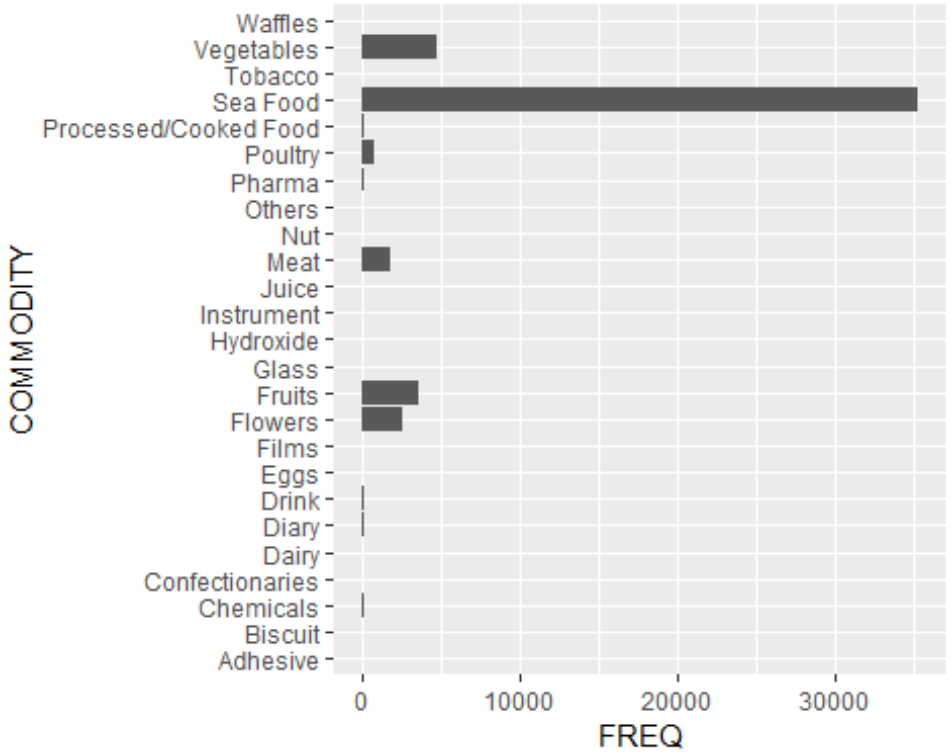
```
# Arrange the data frame
arrange(as.data.frame(table(port$COMMODITY)),desc(Freq))
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

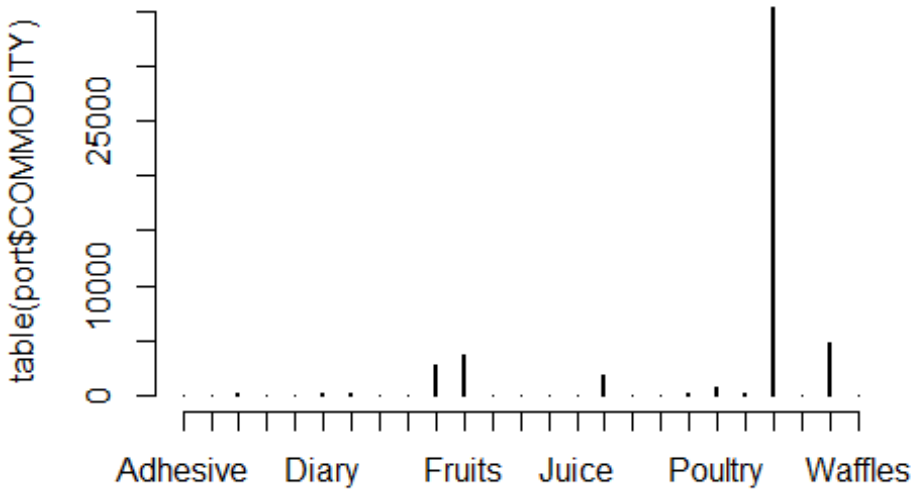
```
##                    Var1  Freq
## 1              Sea Food 35329
## 2            Vegetables  4710
## 3                Fruits  3612
## 4               Flowers  2651
## 5                  Meat  1834
## 6               Poultry   738
## 7                 Diary   189
## 8             Chemicals   172
## 9  Processed/Cooked Food   128
## 10                Drink   118
## 11               Pharma   118
## 12      Confectionaries    45
## 13                Dairy    35
## 14               Waffles    31
## 15             Adhesive    28
## 16            Hydroxide    28
## 17                Juice    28
```

```
## 18            Instrument    26
## 19                   Nut    26
## 20               Biscuit    24
## 21                  Eggs    24
## 22                Others    23
## 23                 Films    22
## 24                 Glass    22
## 25               Tobacco    19
```

```
#plot the dataframe
df = arrange(as.data.frame(table(port$COMMODITY)),desc(Freq))
names(df) = c('COMMODITY', 'FREQ')
g = ggplot(df, aes(x=COMMODITY, y=FREQ))
g+ geom_col() +coord_flip()
```

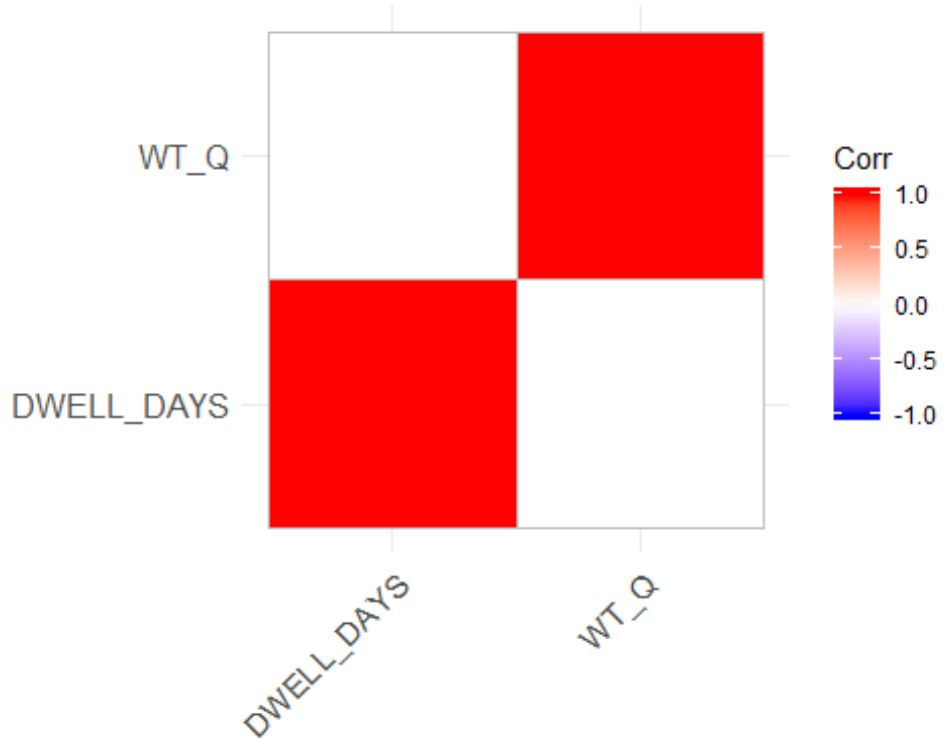



```
#Another simple way of plotting
plot(table(port$COMMODITY))
```

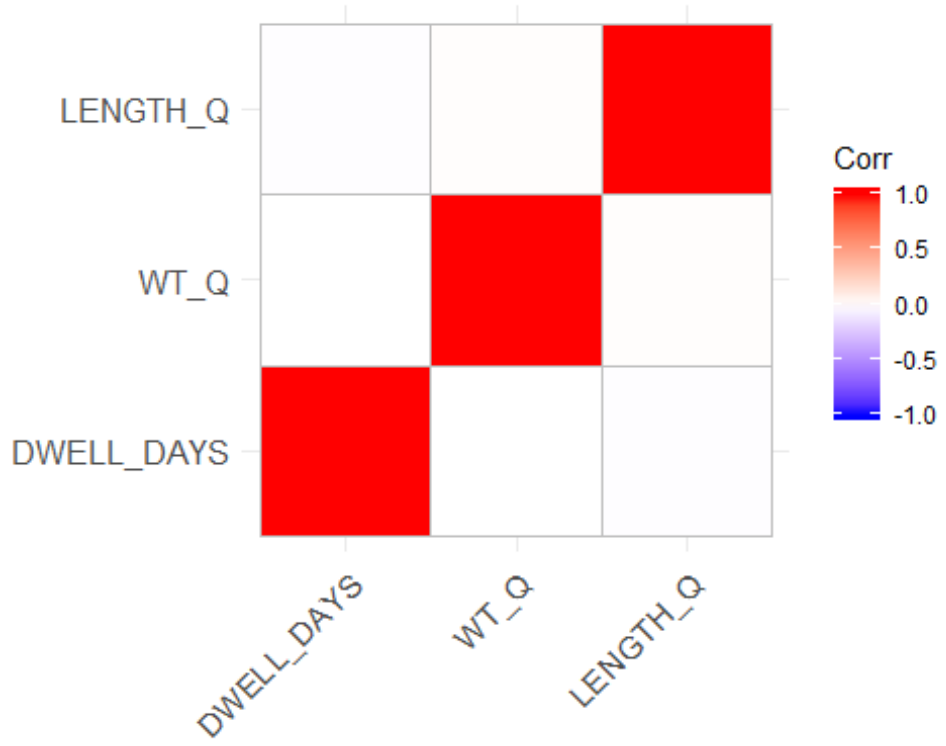


Visualization of co-relation among variables

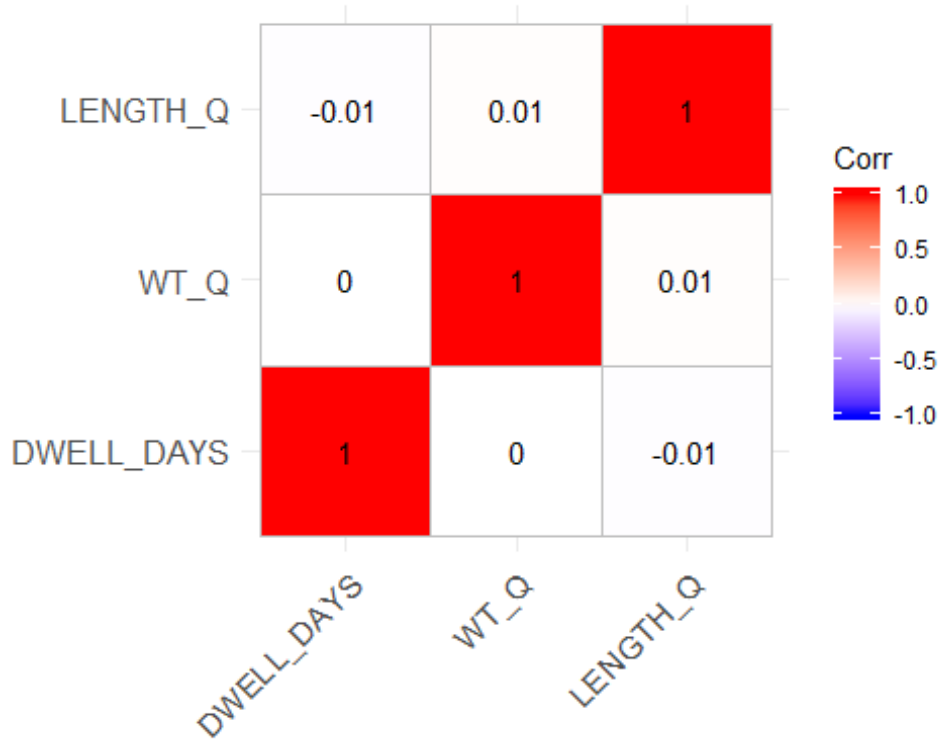
```
#Plot the corelation
ggcorrplot(cor(port[,c('DWELL_DAYS', 'WT_Q')]))
```



```
ggcorrplot(cor(port[,c('DWELL_DAYS', 'WT_Q', 'LENGTH_Q')]))
```

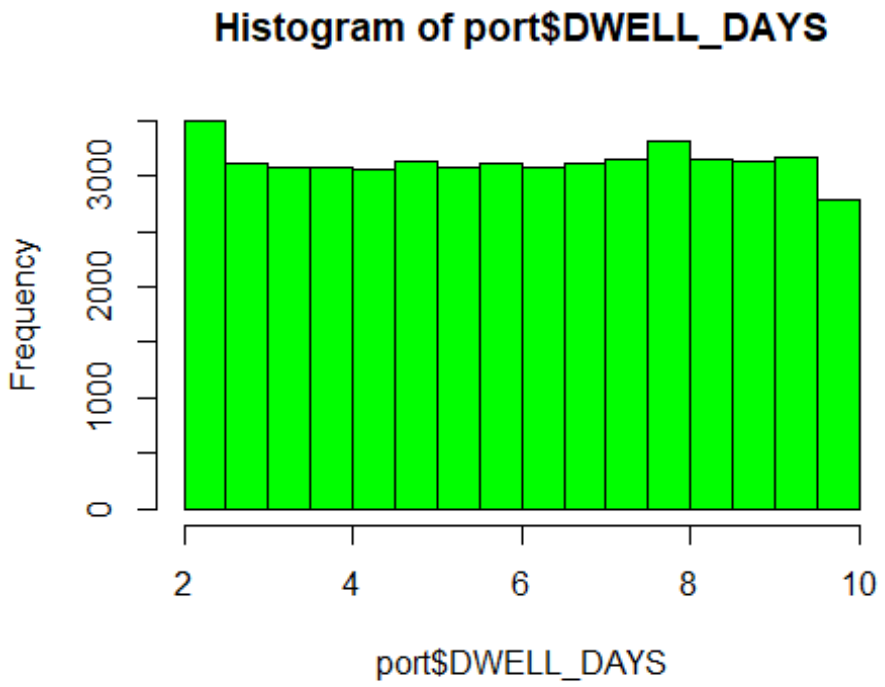



```
ggcorrplot(cor(port[,c('DWELL_DAYS', 'WT_Q', 'LENGTH_Q')]),lab=TRUE)
```

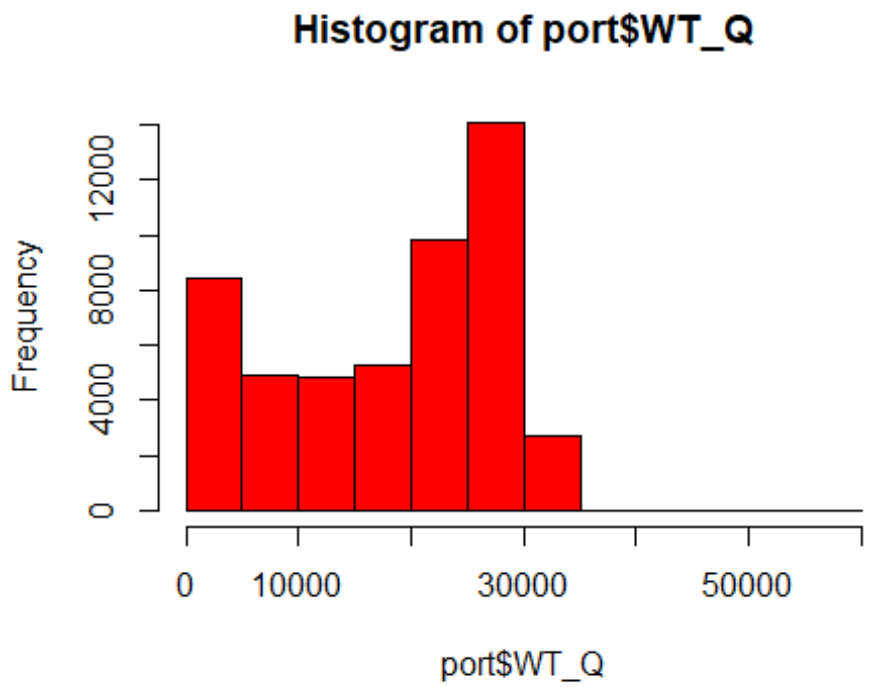


Looking at the histogram of dwelldays and weight

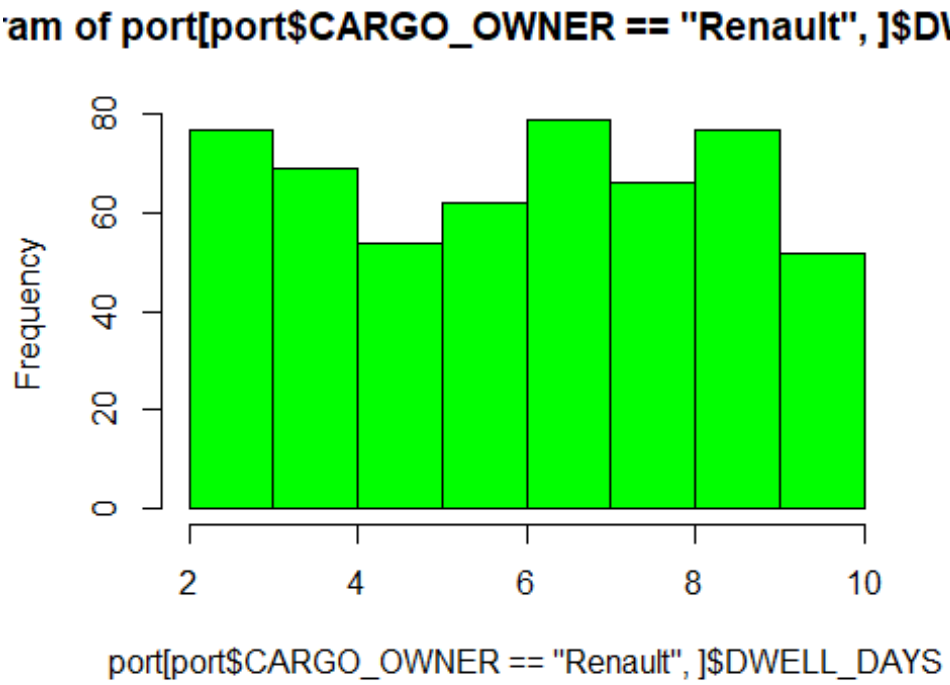
```
#plotting the histogram
hist(port$DWELL_DAYS, col='green')
```



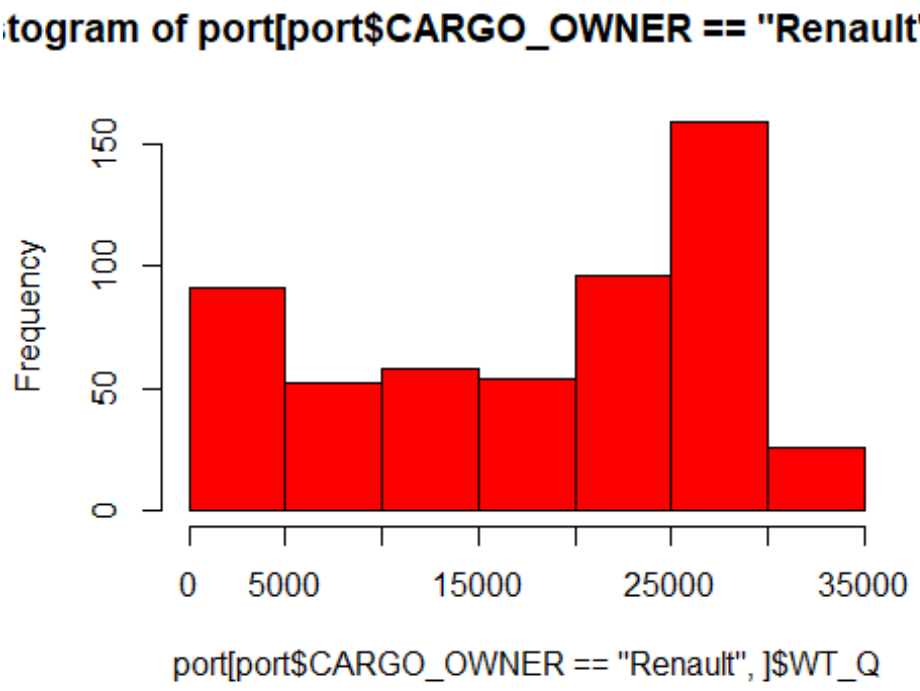

```
hist(port$WT_Q, col='red')
```




```r
#plotting the histogram for a cargo owner
hist(port[port$CARGO_OWNER == 'Renault',]$DWELL_DAYS, col='green')
```

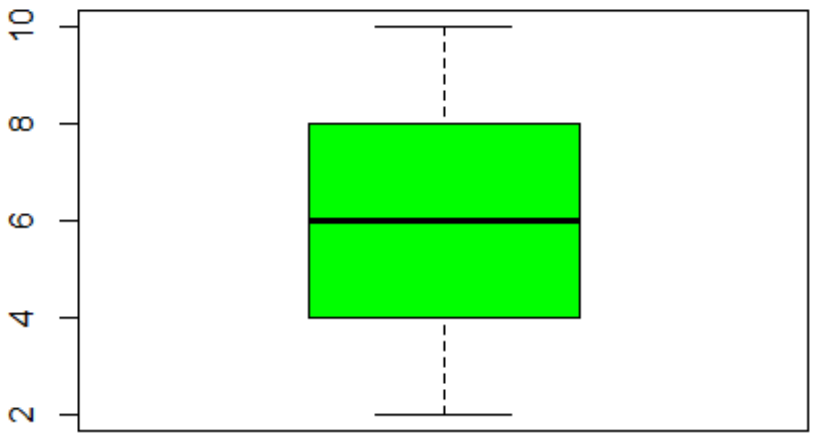


```
hist(port[port$CARGO_OWNER == 'Renault',]$WT_Q, col='red')
```

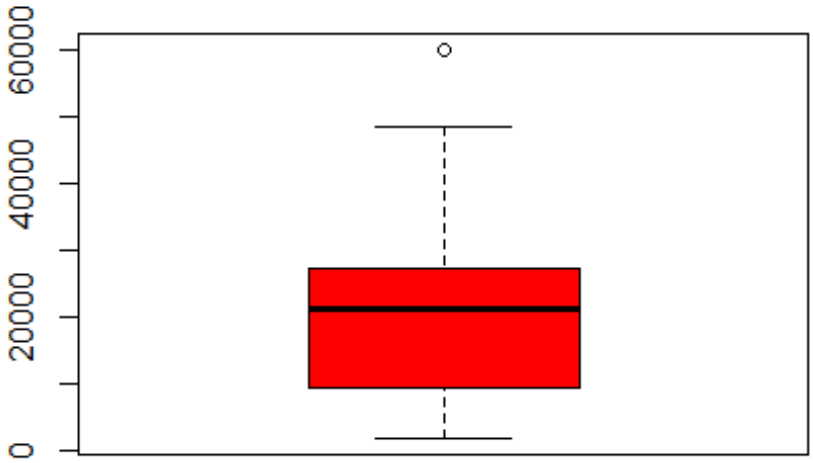


Looking at the boxplot of dwelldays and weight

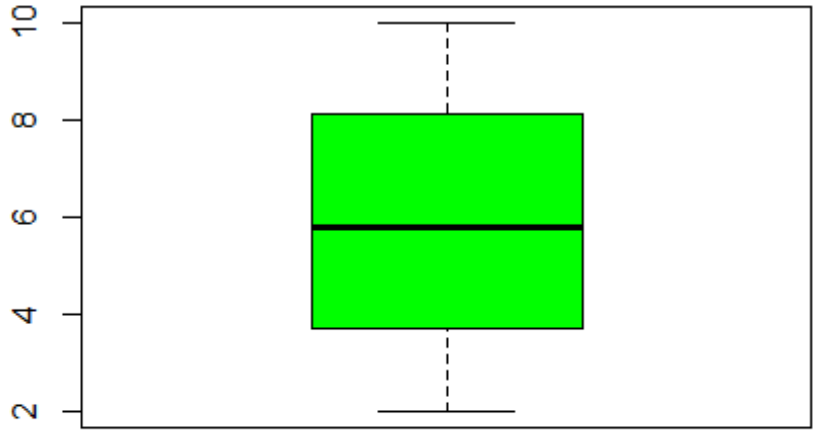
```
#plotting the histogram
boxplot(port$DWELL_DAYS, col='green')
```



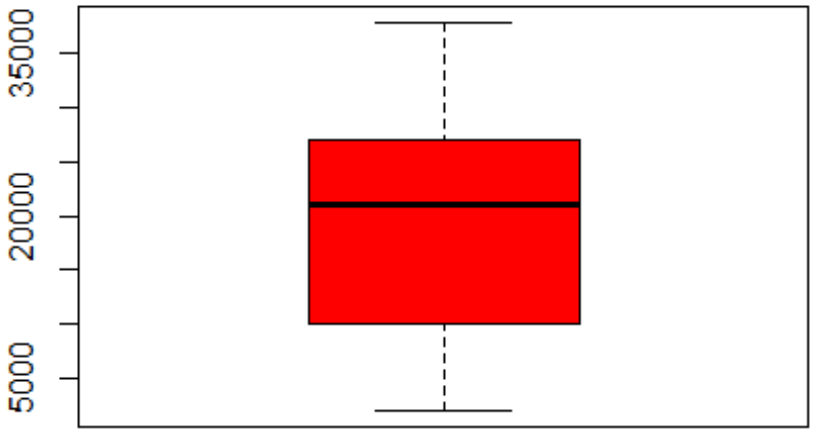

```
boxplot(port$WT_Q, col='red')
```



```r
#plotting the histogram for a cargo owner
boxplot(port[port$CARGO_OWNER == 'Dell',]$DWELL_DAYS, col='green')
```




```r
boxplot(port[port$CARGO_OWNER == 'Dell',]$WT_Q, col='red')
```



End of the Script