Programming for Data Science
Digital Assignment 5

Name: Tejas Rahul Rokade

Reg No: 20BDS0033

Course Code: CSE3046

Code

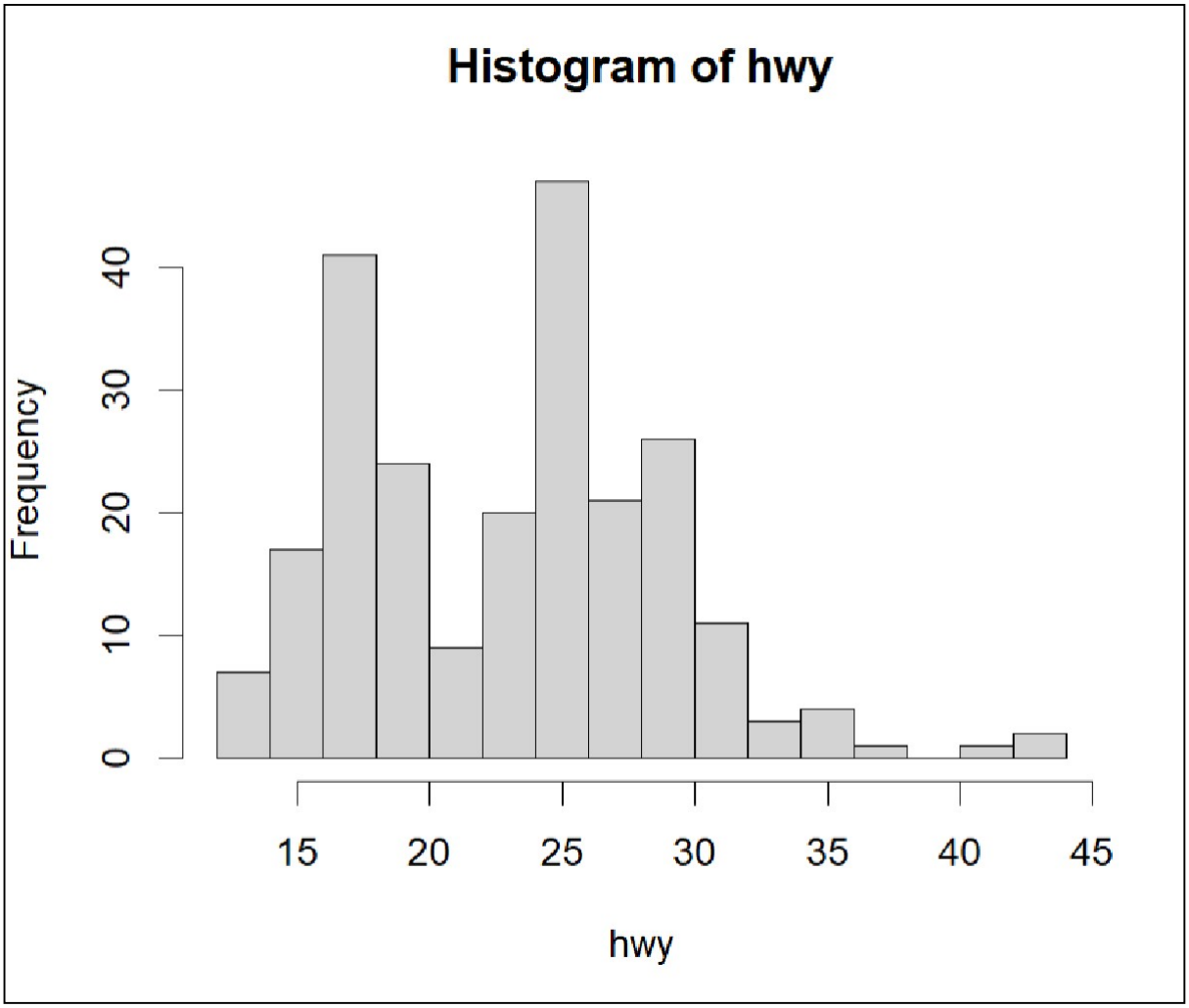
```r
dat <- ggplot2::mpg
summary(dat$hwy)
min(dat$hwy)
max(dat$hwy)
range(dat$hwy)
#---histogram--
hist(dat$hwy,
     xlab = "hwy",
     main = "Histogram of hwy",
     breaks = sqrt(nrow(dat))
) # set number of bins
#---2nd way
library(ggplot2)
ggplot(dat) +
  aes(x = hwy) +
  geom_histogram(bins = 30L, fill = "#0c4c8a") +
  theme_minimal()
#--boxplot
boxplot(dat$hwy,
        ylab = "hwy"
)
ggplot(dat) +
  aes(x = "", y = hwy) +
  geom_boxplot(fill = "#0c4c8a") +
  theme_minimal()
boxplot.stats(dat$hwy)$out
out <- boxplot.stats(dat$hwy)$out
out_ind <- which(dat$hwy %in% c(out))
out_ind
dat[out_ind, ]
```

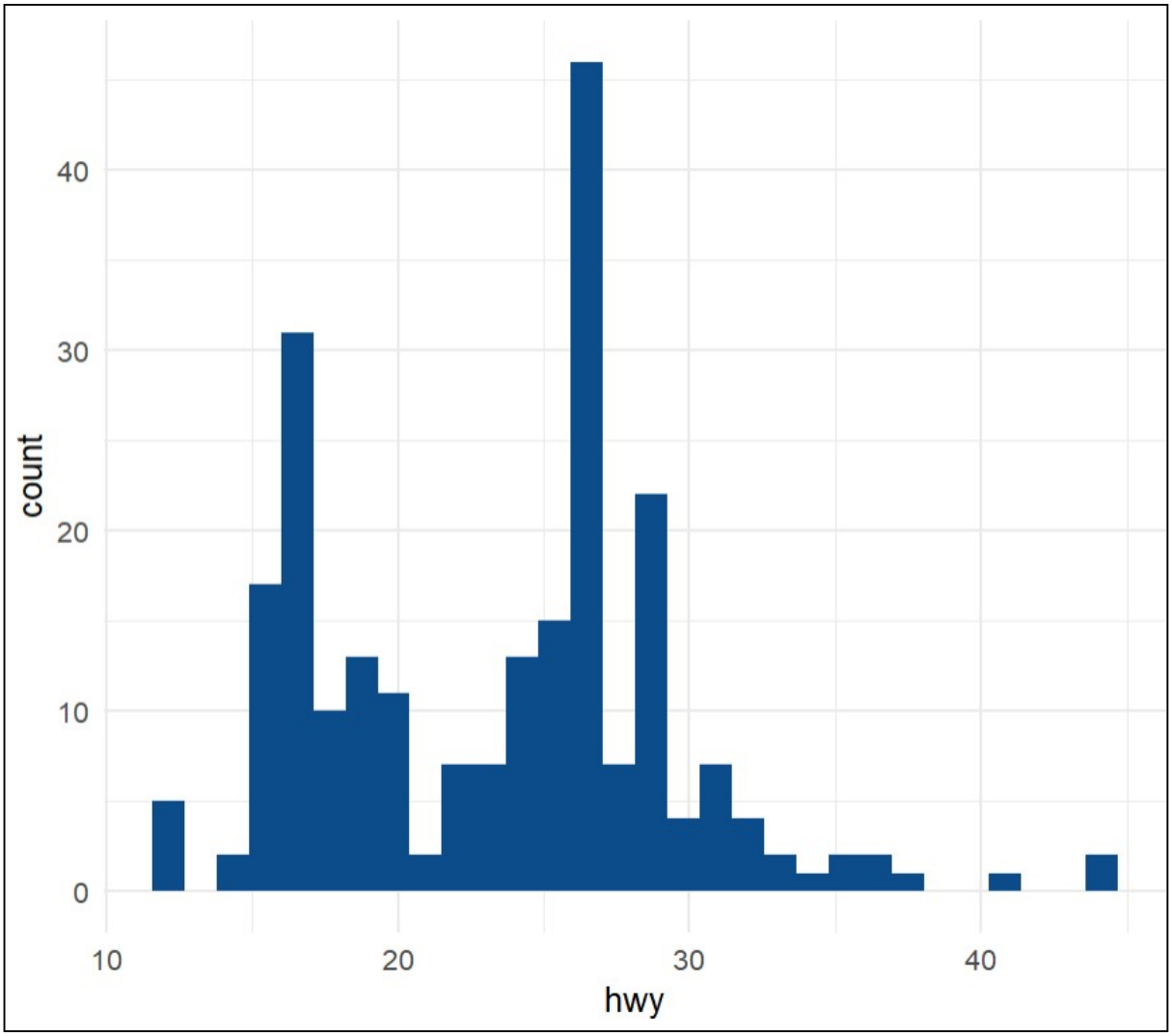
```r
dat[out_ind, ]
boxplot(dat$hwy,
        ylab = "hwy",
        main = "Boxplot of highway miles per gallon"
)
mtext(paste("Outliers: ", paste(out, collapse = ", ")))
```

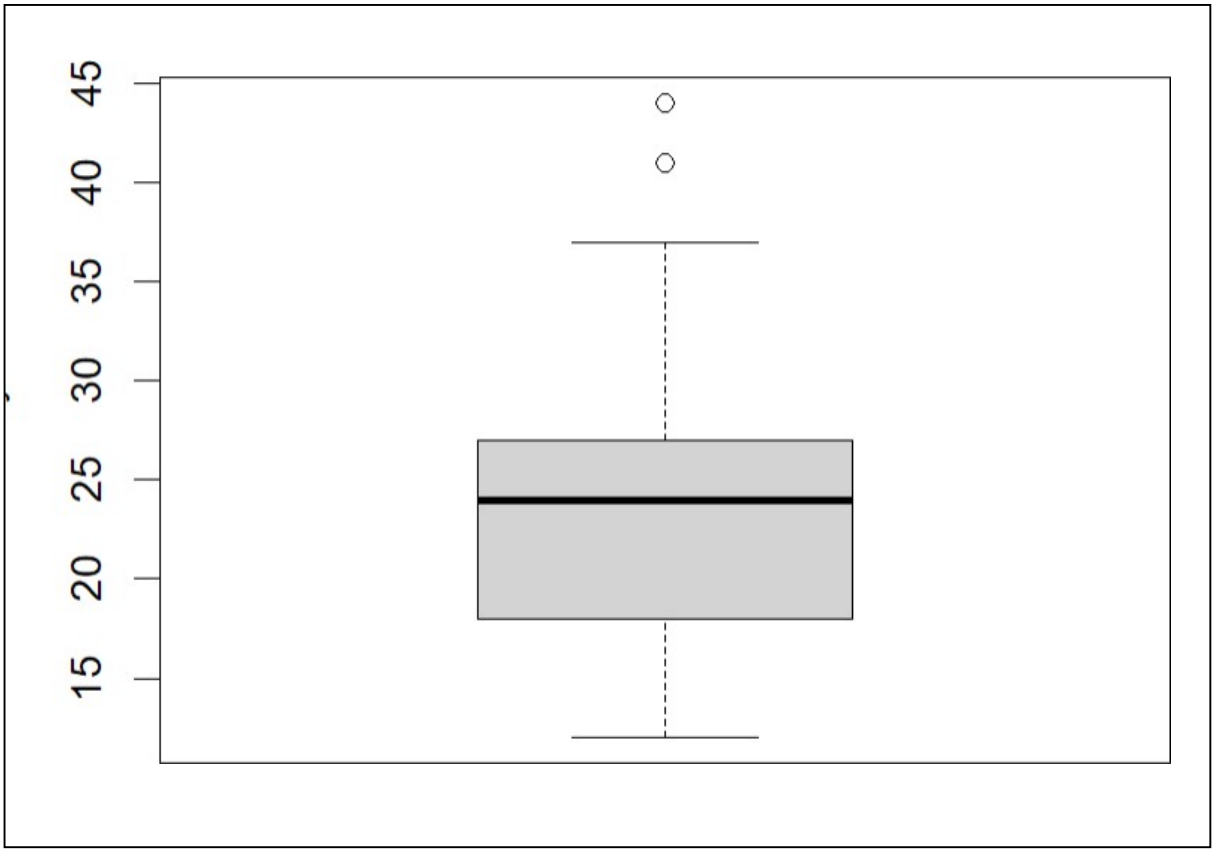
Output

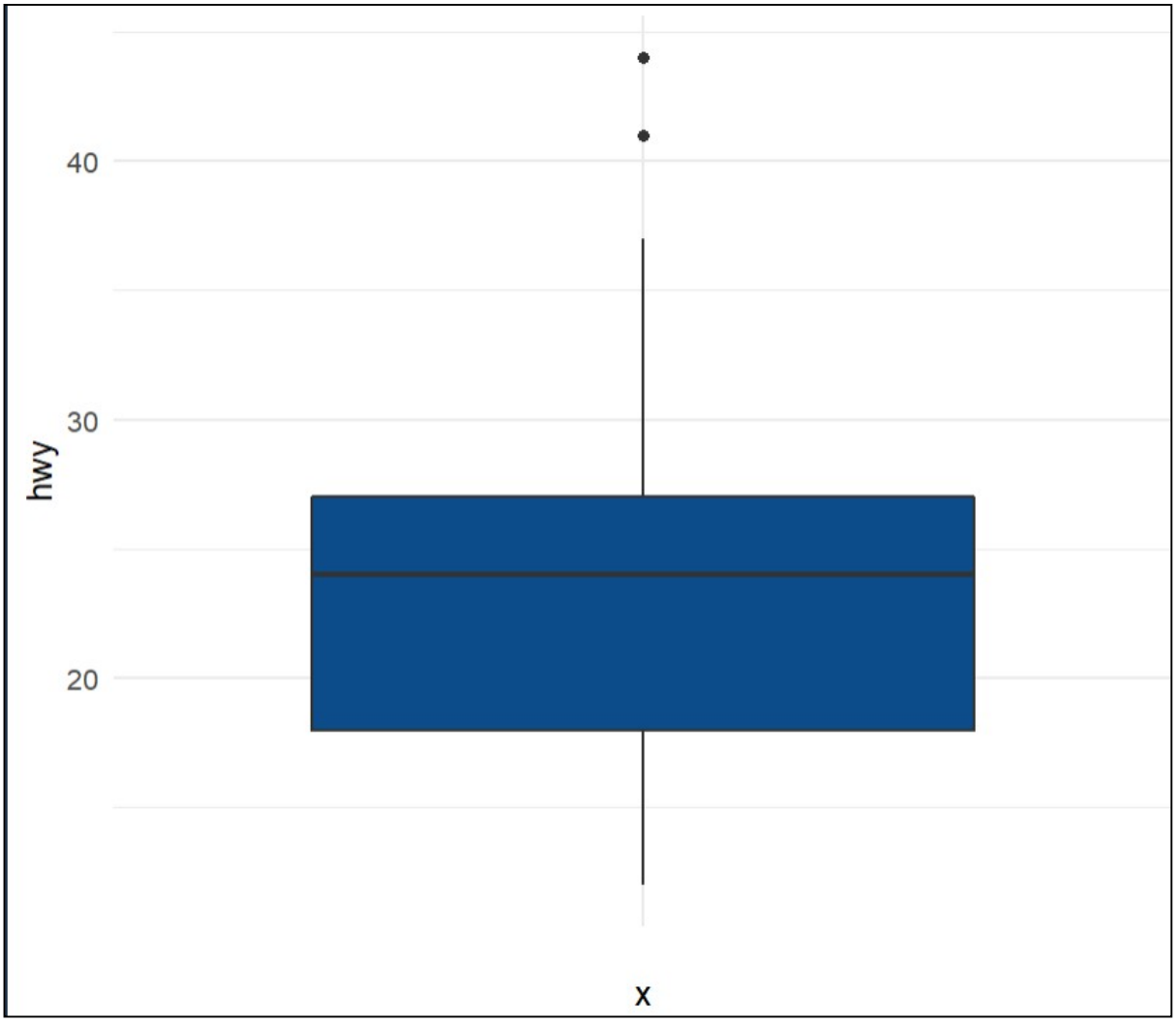
```
> dat <- ggplot2::mpg
> summary(dat$hwy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.00   18.00   24.00   23.44   27.00   44.00
> min(dat$hwy)
[1] 12
> max(dat$hwy)
[1] 44
> range(dat$hwy)
[1] 12 44
> #---histogram--
> hist(dat$hwy,
+      xlab = "hwy",
+      main = "Histogram of hwy",
+      breaks = sqrt(nrow(dat))
+ ) # set number of bins
> #---2nd way
> library(ggplot2)
> ggplot(dat) +
+   aes(x = hwy) +
+   geom_histogram(bins = 30L, fill = "#0c4c8a") +
+   theme_minimal()
> #--boxplot
> boxplot(dat$hwy,
+         ylab = "hwy"
+ )
> ggplot(dat) +
+   aes(x = "", y = hwy) +
+   geom_boxplot(fill = "#0c4c8a") +
```

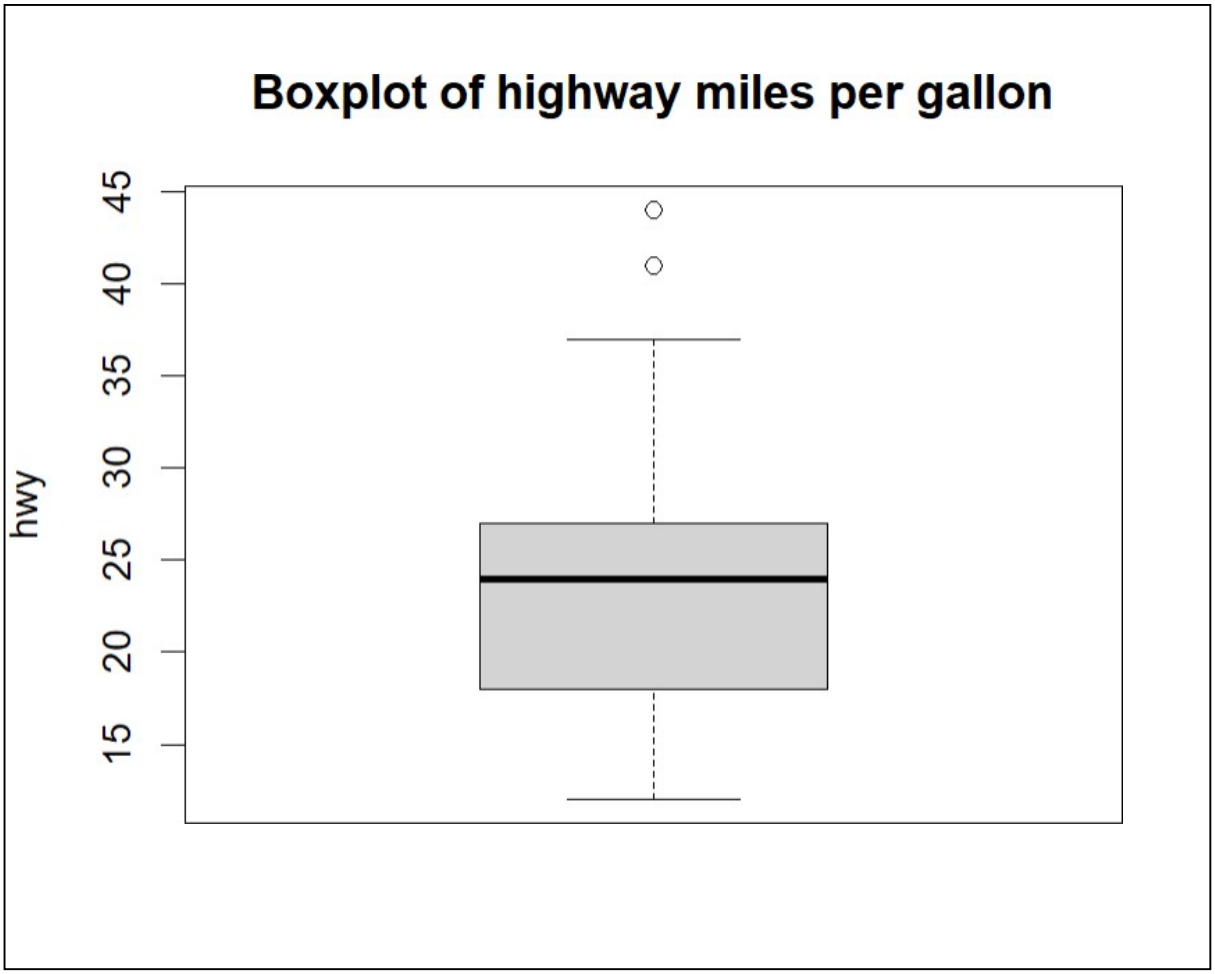
```
+   geom_boxplot(fill = "#0c4c8a") +
+   theme_minimal()
> boxplot.stats(dat$hwy)$out
[1] 44 44 41
> out <- boxplot.stats(dat$hwy)$out
> out_ind <- which(dat$hwy %in% c(out))
> out_ind
[1] 213 222 223
> dat[out_ind, ]
# A tibble: 3 x 11
  manufacturer model      displ  year   cyl trans      drv     cty   hwy fl    class
  <chr>        <chr>      <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
1 volkswagen   jetta        1.9  1999     4 manual(m5) f        33    44 d     compact
2 volkswagen   new beetle   1.9  1999     4 manual(m5) f        35    44 d     subcompact
3 volkswagen   new beetle   1.9  1999     4 auto(l4)   f        29    41 d     subcompact
> boxplot(dat$hwy,
+         ylab = "hwy",
+         main = "Boxplot of highway miles per gallon"
+ )
> mtext(paste("Outliers: ", paste(out, collapse = ", ")))
>
```


Histogram of hwy
Frequency
0
10
20
30
40
15
20
25
30
35
40
45
hwy


40
30
20
10
0
count
10
20
30
40
hwy


45
40
35
30
25
20
15


40
30
20
hwy
x


Boxplot of highway miles per gallon
hwy
15
20
25
30
35
40
45