ACADGILD

SESSION 7: Basic Statistics

Assignment 3

PROBLEM STATEMENT

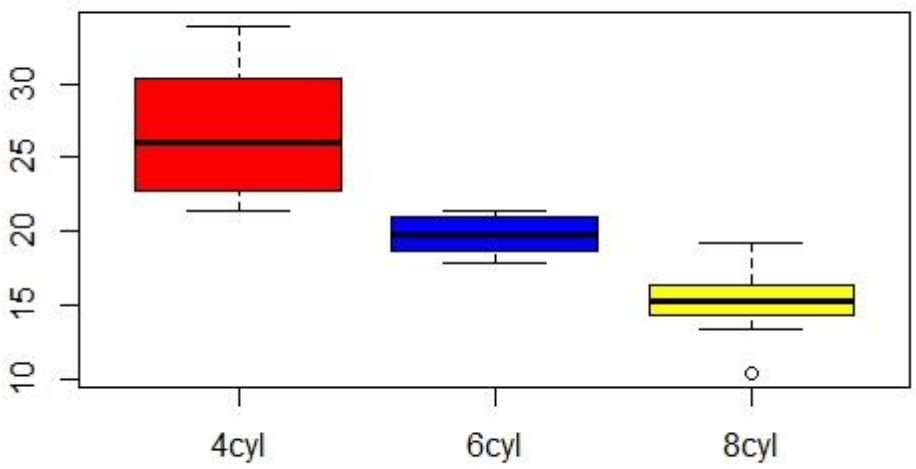
1. Create a **box and whisker plot** by class using **mtcars** dataset.

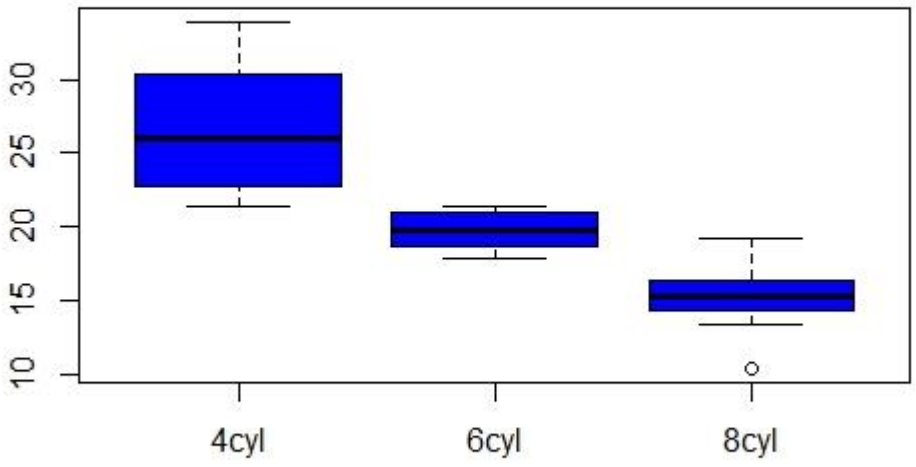
SOLUTION

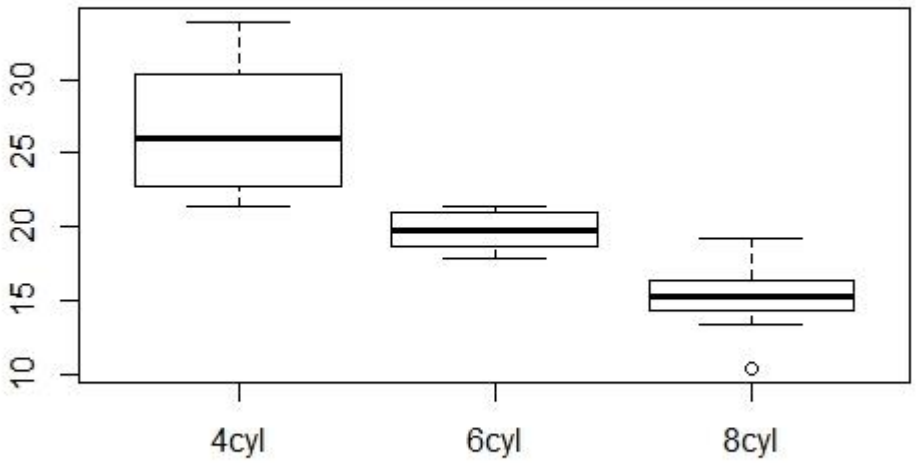
```
summary(cars)
boxplot(mtcars$mpg)
boxplot(mtcars$mpg, horizontal = TRUE)
boxplot(mtcars$mpg, col = 'blue')
boxplot(mtcars$mpg, border = 'red')
boxplot(mtcars$mpg, range = 0)
boxplot(mtcars$mpg, range = 1)
boxplot(mtcars$mpg, range = 1, outline = FALSE)
boxplot(mtcars$mpg ~ mtcars$cyl)
mpg_split <- split(mtcars$mpg, mtcars$cyl)
mpg_split
mpg_4 <- mpg_split$`4`
mpg_6 <- mpg_split$`6`
mpg_8 <- mpg_split$`8`
boxplot(mpg_4, mpg_6, mpg_8)
boxplot(mtcars$mpg ~ mtcars$cyl, col = 'blue')
boxplot(mtcars$mpg ~ mtcars$cyl,col = c('red', 'blue', 'yellow'))
boxplot(mtcars$mpg ~ mtcars$cyl, range = 1, outline = TRUE,horizontal =
TRUE, col = c('red', 'blue', 'yellow'), main = 'Miles Per Gallon by Cylinders',ylab
= 'Number of Cylinders', xlab = 'Miles Per Gallon',names = c('Four', 'Six',
'Eight'))

library(ggplot2)
library(xtable)
head(mtcars)
mtcars$cyl <- factor(mtcars$cyl)
```

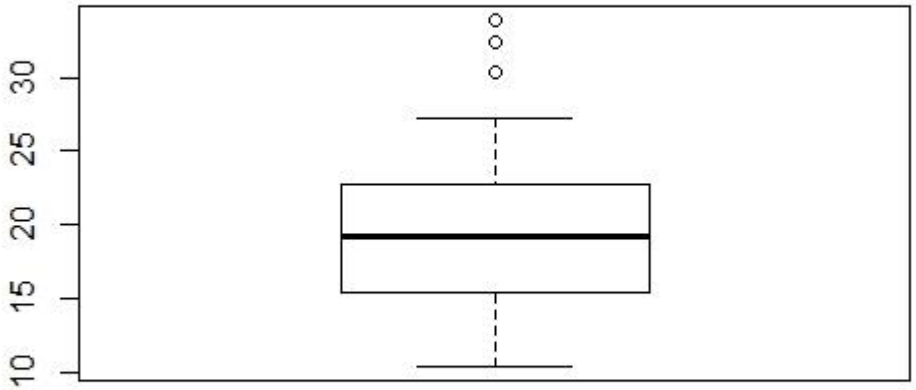
```
mtcars$labels <- row.names(mtcars)
summary(mtcars)
library(gridExtra)
library(ggplot2)
library(ggthemes)
library(tufte)
p <- ggplot(data = mtcars, aes(x = cyl, y = mpg, fill = cyl))

p <- p + geom_boxplot() +ggtitle("Car Milage Data") +labs(x = "Number of
Cylinders", y = "Miles Per Gallon") +scale_fill_discrete(name = "Cylinders")

p

p <- ggplot(mtcars, aes(x = wt, y = mpg)) +geom_point() +ggtitle("Cars")
p2 <- ggplot(mtcars, aes(x = wt, y = mpg, colour = factor(gear))) +geom_point()
+ggtitle("Cars")

p3 <- p2 + facet_wrap(~ am)p + geom_rangeframe() +theme_tufte() +
scale_x_continuous(breaks = extended_range_breaks()(mtcars$wt))
+scale_y_continuous(breaks = extended_range_breaks()(mtcars$mpg))

p4 <- ggplot(mtcars, aes(factor(cyl), mpg))p4 + theme_tufte(ticks=FALSE) +
geom_tufteboxplot()p4 + theme_tufte(ticks=FALSE)
+geom_tufteboxplot(median.type = "line")p4 + theme_tufte(ticks=FALSE)
+geom_tufteboxplot(median.type = "line", whisker.type = 'point', hoffset = 0)p4
+ theme_tufte(ticks=FALSE) +geom_tufteboxplot(median.type = "line",
whisker.type = 'line', hoffset = 0, width = 3)
```

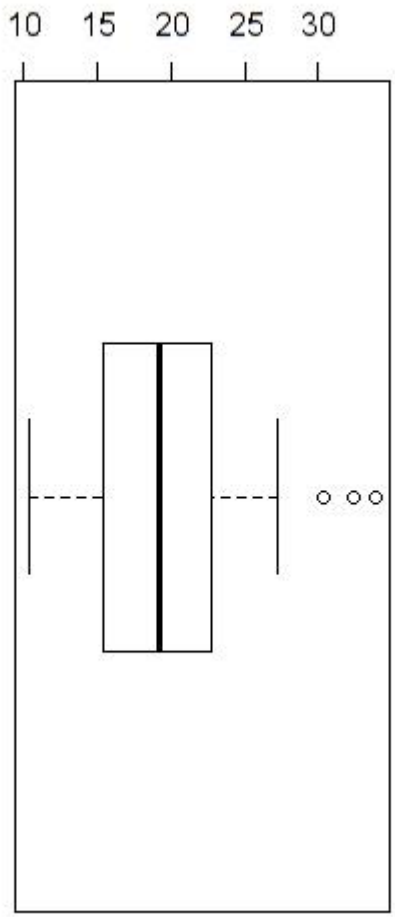
```
> summary(cars)
      X1                  mpg              cyl              disp
 Length:32          Min.   :10.40   Min.   :4.000   Min.   : 71.1
 Class :character   1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8
 Mode  :character   Median :19.20   Median :6.000   Median :196.3
                    Mean   :20.09   Mean   :6.188   Mean   :230.7
                    3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0
                    Max.   :33.90   Max.   :8.000   Max.   :472.0
       hp             drat             wt             qsec             vs
 Min.   : 52.0   Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
 1st Qu.: 96.5   1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
 Median :123.0   Median :3.695   Median :3.325   Median :17.71   Median :0.0000
 Mean   :146.7   Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
 3rd Qu.:180.0   3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
 Max.   :335.0   Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
       am              gear            carb
 Min.   :0.0000   Min.   :3.000   Min.   :1.000
 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
 Median :0.0000   Median :4.000   Median :2.000
 Mean   :0.4062   Mean   :3.688   Mean   :2.812
 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
 Max.   :1.0000   Max.   :5.000   Max.   :8.000
> boxplot(mtcars$mpg)
> boxplot(mtcars$mpg, horizontal = TRUE)
> boxplot(mtcars$mpg, col = 'blue')
> boxplot(mtcars$mpg, border = 'red')
> boxplot(mtcars$mpg, range = 0)
> boxplot(mtcars$mpg, range = 1)
> boxplot(mtcars$mpg, range = 1, outline = FALSE)
> boxplot(mtcars$mpg ~ mtcars$cyl)
> mpg_split <- split(mtcars$mpg, mtcars$cyl)
> mpg_split
$`4`
 [1] 22.8 24.4 22.8 32.4 30.4 33.9 21.5 27.3 26.0 30.4 21.4

$`6`
[1] 21.0 21.0 21.4 18.1 19.2 17.8 19.7

$`8`
 [1] 18.7 14.3 16.4 17.3 15.2 10.4 10.4 14.7 15.5 15.2 13.3 19.2 15.8 15.0

> mpg_4 <- mpg_split$`4`
> mpg_6 <- mpg_split$`6`
> mpg_8 <- mpg_split$`8`
> boxplot(mpg_4, mpg_6, mpg_8)
> boxplot(mtcars$mpg ~ mtcars$cyl, col = 'blue')
> boxplot(mtcars$mpg ~ mtcars$cyl,col = c('red', 'blue', 'yellow'))
```

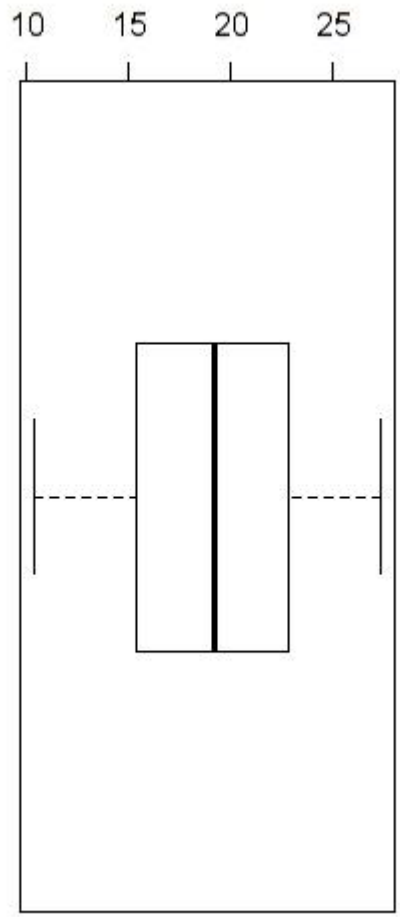

10
15
20
25
30
4cyl
6cyl
8cyl

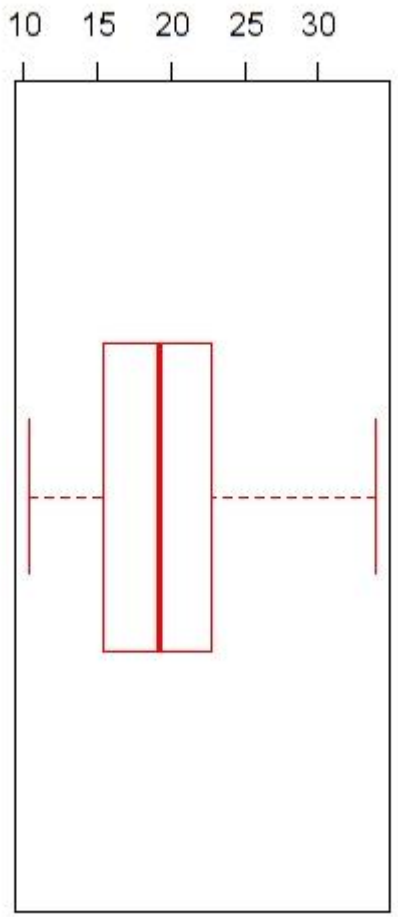

10
15
20
25
30
4cyl
6cyl
8cyl

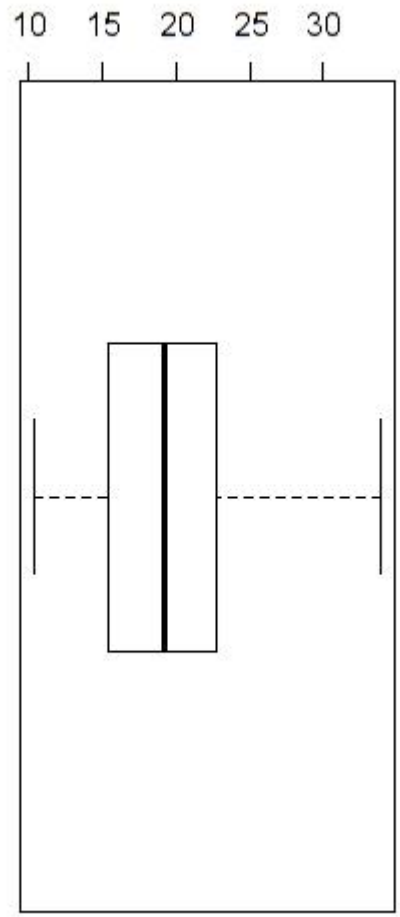

10
15
20
25
30
4cyl
6cyl
8cyl

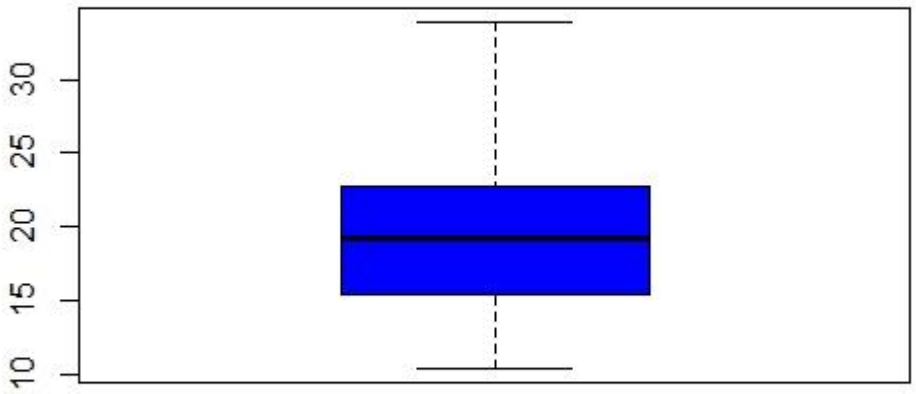

10
15
20
25
30

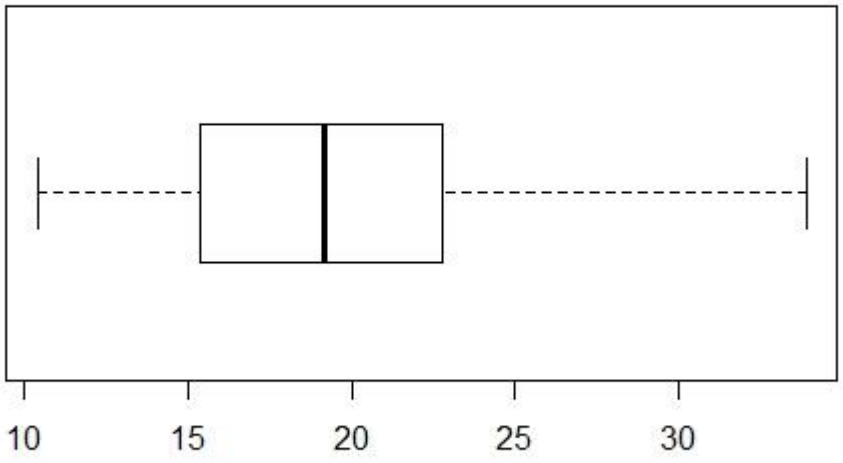

10
15
20
25
30

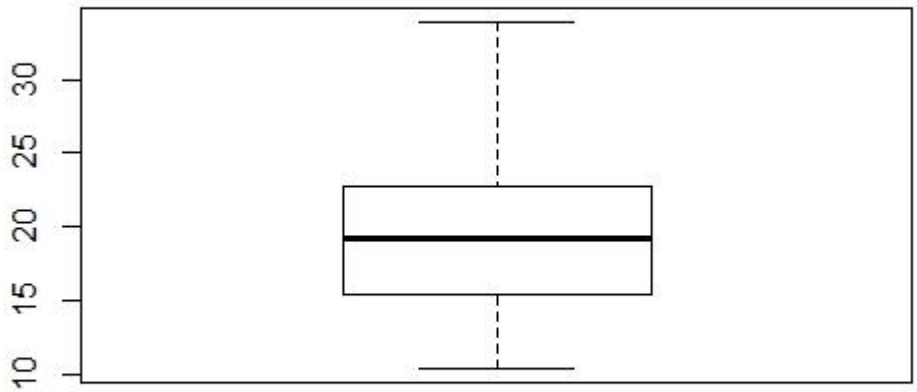

10
15
20
25

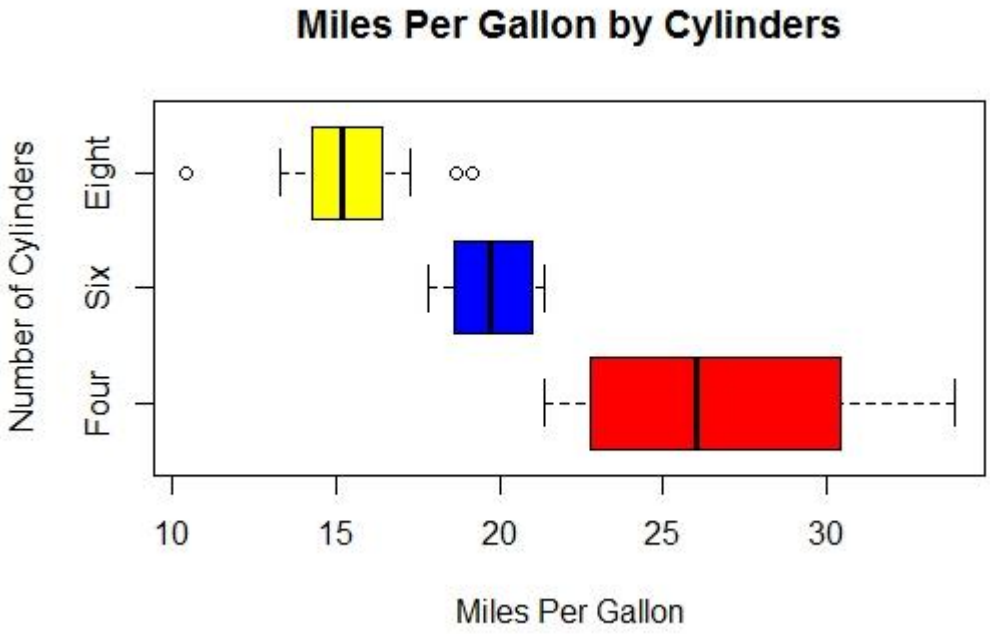

10
15
20
25
30


10
15
20
25
30


10
15
20
25
30

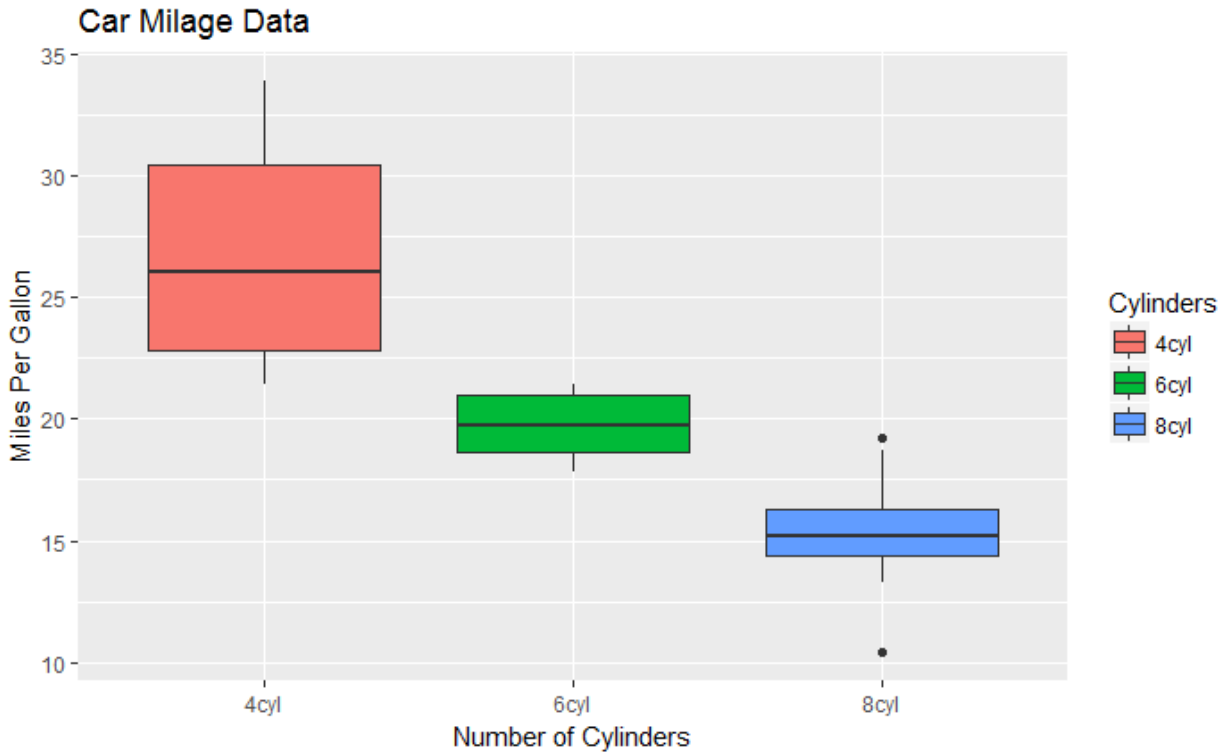

10
15
20
25
30

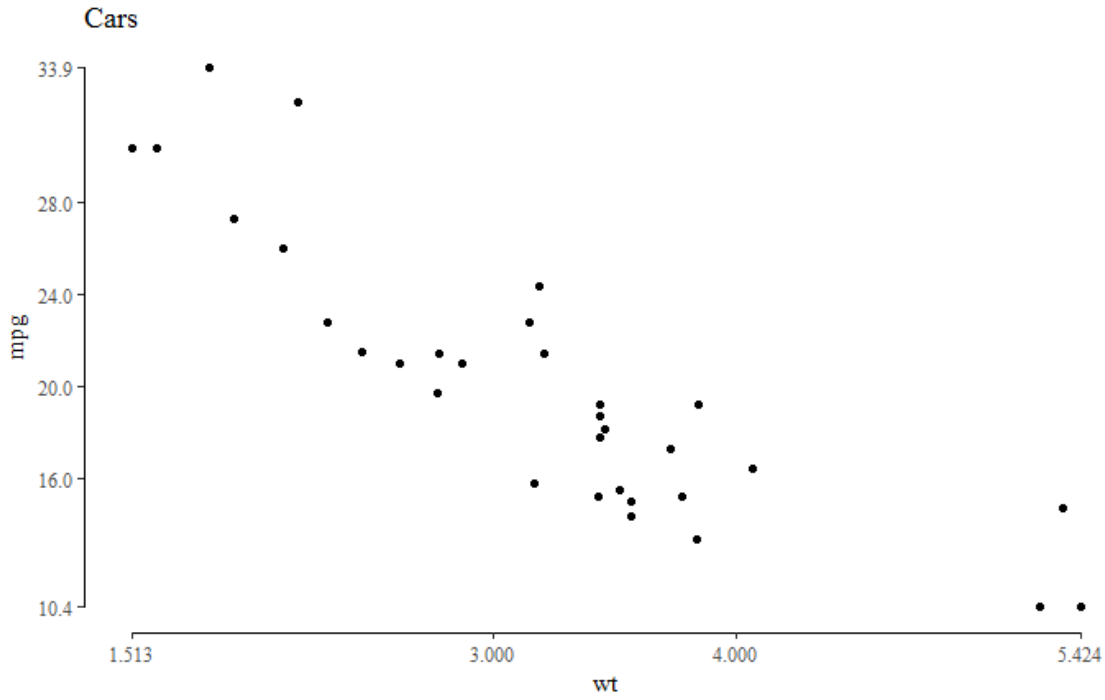




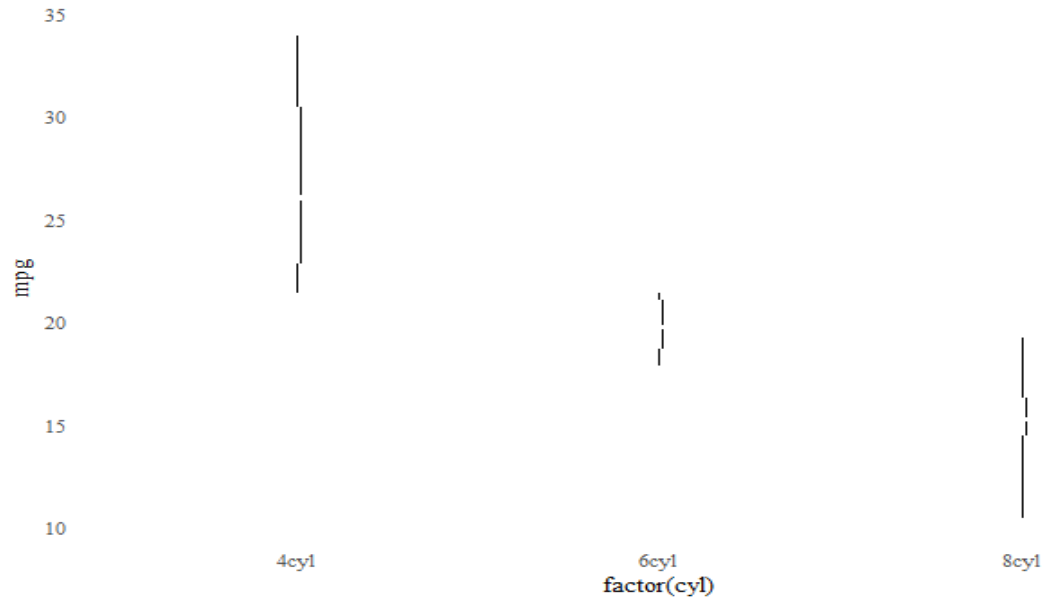

```
> library(ggplot2)
> library(xtable)
> head(mtcars)
# A tibble: 6 x 12
  X1                mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear
carb
  <chr>           <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
<dbl>
1 Mazda RX4        21       6   160   110  3.9   2.62  16.5     0     1     4
4
2 Mazda RX4 Wag    21       6   160   110  3.9   2.88  17.0     0     1     4
4
3 Datsun 710       22.8     4   108    93  3.85  2.32  18.6     1     1     4
1
```

```
4 Hornet 4 Drive   21.4     6   258   110  3.08  3.22  19.4     1     0     3
1
5 Hornet Sportab~  18.7     8   360   175  3.15  3.44  17.0     0     0     3
2
6 Valiant          18.1     6   225   105  2.76  3.46  20.2     1     0     3
1
> mtcars$cyl <- factor(mtcars$cyl)
> mtcars$labels <- row.names(mtcars)
> summary(mtcars)
      X1                 mpg        cyl         disp             hp
 Length:32          Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0
 Class :character   1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5
 Mode  :character   Median :19.20   8:14   Median :196.3   Median :123.0
                    Mean   :20.09          Mean   :230.7   Mean   :146.7
                    3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0
                    Max.   :33.90          Max.   :472.0   Max.   :335.0
      drat             wt             qsec             vs                am
 Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000   Min.
:0.0000
 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000   1st
Qu.:0.0000
 Median :3.695   Median :3.325   Median :17.71   Median :0.0000   Median
:0.0000
 Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375   Mean
:0.4062
 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000   3rd
Qu.:1.0000
 Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000   Max.
:1.0000
      gear            carb          labels
 Min.   :3.000   Min.   :1.000   Length:32
 1st Qu.:3.000   1st Qu.:2.000   Class :character
 Median :4.000   Median :2.000   Mode  :character
 Mean   :3.688   Mean   :2.812
 3rd Qu.:4.000   3rd Qu.:4.000
 Max.   :5.000   Max.   :8.000
> library(gridExtra)
> library(ggplot2)
> library(ggthemes)
> library(tufte)
> p <- ggplot(data = mtcars, aes(x = cyl, y = mpg, fill = cyl))
> p <- p + geom_boxplot() +ggtitle("Car Milage Data") +labs(x = "Number of
Cylinders", y = "Miles Per Gallon") +scale_fill_discrete(name = "Cylinders")
> p
> p <- ggplot(mtcars, aes(x = wt, y = mpg)) +geom_point() +ggtitle("Cars")
> p2 <- ggplot(mtcars, aes(x = wt, y = mpg, colour = factor(gear)))
+geom_point() +ggtitle("Cars")
> p3 <- p2 + facet_wrap(~ am)p + geom_rangeframe() +theme_tufte() +
scale_x_continuous(breaks = extended_range_breaks()(mtcars$wt))
+scale_y_continuous(breaks = extended_range_breaks()(mtcars$mpg))
> p4 <- ggplot(mtcars, aes(factor(cyl), mpg))p4 + theme_tufte(ticks=FALSE) +
geom_tufteboxplot()p4 + theme_tufte(ticks=FALSE)
+geom_tufteboxplot(median.type = "line")p4 + theme_tufte(ticks=FALSE)
+geom_tufteboxplot(median.type = "line", whisker.type = 'point', hoffset =
0)p4 + theme_tufte(ticks=FALSE) +geom_tufteboxplot(median.type = "line",
whisker.type = 'line', hoffset = 0, width = 3)
```






```
p3 <- p2 + facet_wrap(~ am)p + geom_rangeframe() +theme_tufte() +
scale_x_continuous(breaks = extended_range_breaks()(mtcars$wt)) +
scale_y_continuous(breaks = extended_range_breaks()(mtcars$mpg))
```




```
p4 <- ggplot(mtcars, aes(factor(cyl), mpg))p4 + theme_tufte(ticks=FALSE) +
geom_tufteboxplot()p4 + theme_tufte(ticks=FALSE)
+geom_tufteboxplot(median.type = "line")p4 + theme_tufte(ticks=FALSE)
+geom_tufteboxplot(median.type = "line", whisker.type = 'point', hoffset =
0)p4 + theme_tufte(ticks=FALSE) +geom_tufteboxplot(median.type = "line",
whisker.type = 'line', hoffset = 0, width = 3)
```