Session_3

Antoine Mayerowitz

24/09/2018

Chaper 2 - Working with data

```
data("mpg")
dim(mpg)
```

```
## [1] 234  11
```

```
nrow(mpg)
```

```
## [1] 234
```

```
ncol(mpg)
```

```
## [1] 11
```

```
names(mpg)
```

```
##  [1] "manufacturer" "model"        "displ"        "year"
##  [5] "cyl"          "trans"        "drv"          "cty"
##  [9] "hwy"          "fl"           "class"
```

```
head(mpg)
```

```
## # A tibble: 6 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 audi         a4      1.8  1999     4 auto~ f        18    29 p     comp~
## 2 audi         a4      1.8  1999     4 manu~ f        21    29 p     comp~
## 3 audi         a4      2    2008     4 manu~ f        20    31 p     comp~
## 4 audi         a4      2    2008     4 auto~ f        21    30 p     comp~
## 5 audi         a4      2.8  1999     6 auto~ f        16    26 p     comp~
## 6 audi         a4      2.8  1999     6 manu~ f        18    26 p     comp~
```

```
tail(mpg)
```

```
## # A tibble: 6 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   pass~   1.8  1999     4 auto~ f        18    29 p     mids~
## 2 volkswagen   pass~   2    2008     4 auto~ f        19    28 p     mids~
## 3 volkswagen   pass~   2    2008     4 manu~ f        21    29 p     mids~
## 4 volkswagen   pass~   2.8  1999     6 auto~ f        16    26 p     mids~
## 5 volkswagen   pass~   2.8  1999     6 manu~ f        18    26 p     mids~
## 6 volkswagen   pass~   3.6  2008     6 auto~ f        17    26 p     mids~
```

```
str(mpg)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    234 obs. of  11 variables:
##  $ manufacturer: chr  "audi" "audi" "audi" "audi" ...
##  $ model       : chr  "a4" "a4" "a4" "a4" ...
##  $ displ       : num  1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : int  1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
##  $ cyl         : int  4 4 4 4 6 6 6 4 4 4 ...
##  $ trans       : chr  "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
##  $ drv         : chr  "f" "f" "f" "f" ...
##  $ cty         : int  18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : int  29 29 31 30 26 26 27 26 25 28 ...
##  $ fl          : chr  "p" "p" "p" "p" ...
##  $ class       : chr  "compact" "compact" "compact" "compact" ...
```

Summary statistics

```
# central tendency
x = runif(10)
mean(x)
```

```
## [1] 0.3774358
```

```
sum(x) / length(x)
```

```
## [1] 0.3774358
```

```
median(x)
```

```
## [1] 0.2202851
```

```
# Spread
var(x)
```

```
## [1] 0.1330643
```

```
sd(x)
```

```
## [1] 0.3647798
```

```
myVar  = sum((x-mean(x))^2) / (length(x) - 1)
mySd = sqrt(myVar)

# Misc.
range(x)
```

```
## [1] 0.03178333 0.95961800
```

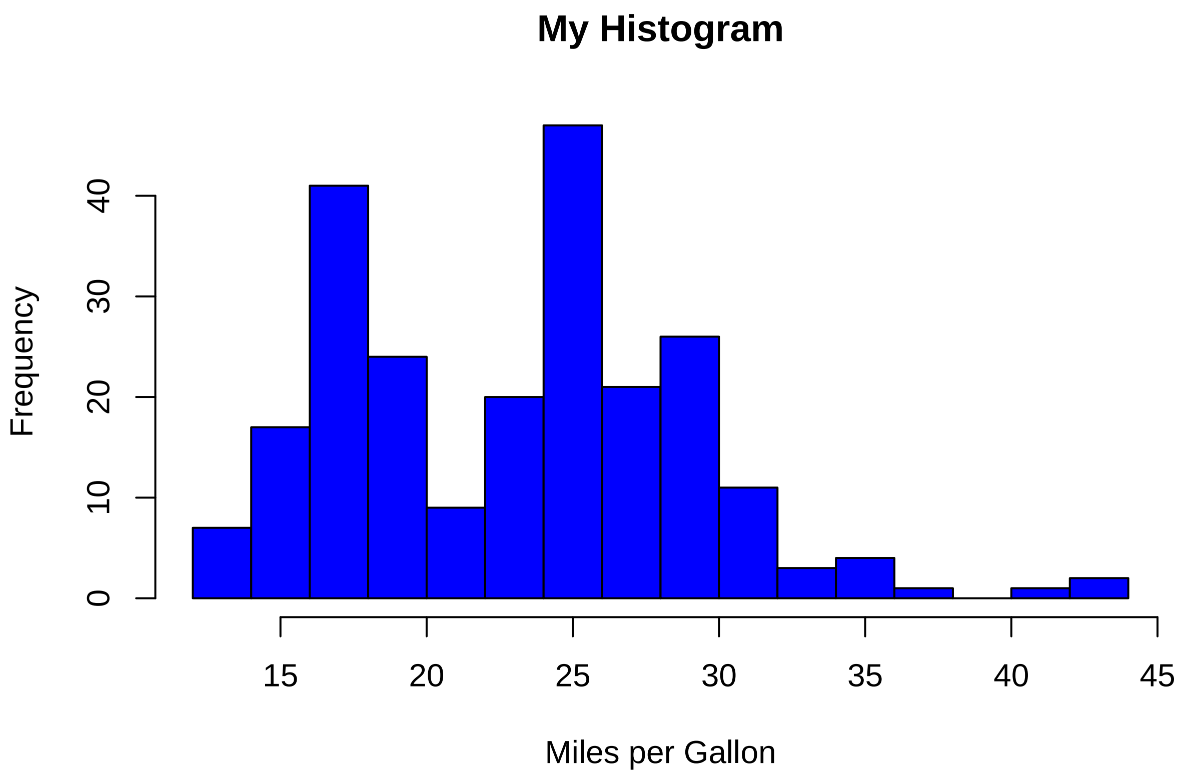
```
table(mpg$drv, mpg$class)
```

```
##    
##     2seater compact midsize minivan pickup subcompact suv
##   4       0      12       3       0     33          4  51
##   f       0      35      38      11      0         22   0
##   r       5       0       0       0      0          9  11
```

Plots

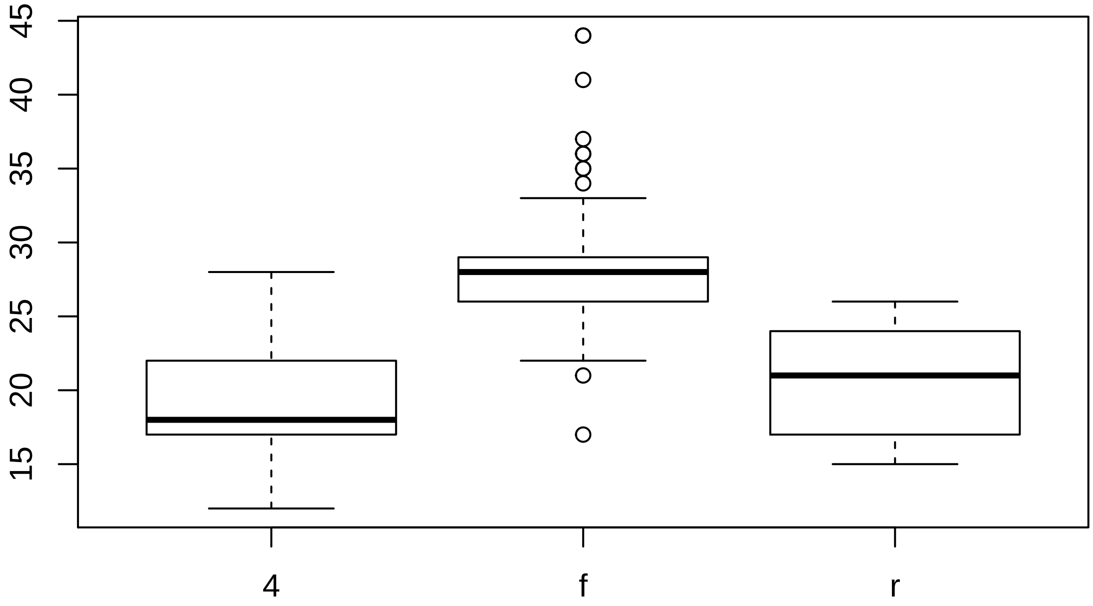
Histogram

```
hist(mpg$hwy, xlab = "Miles per Gallon", main = "My Histogram", breaks =12, col = "blue")
```



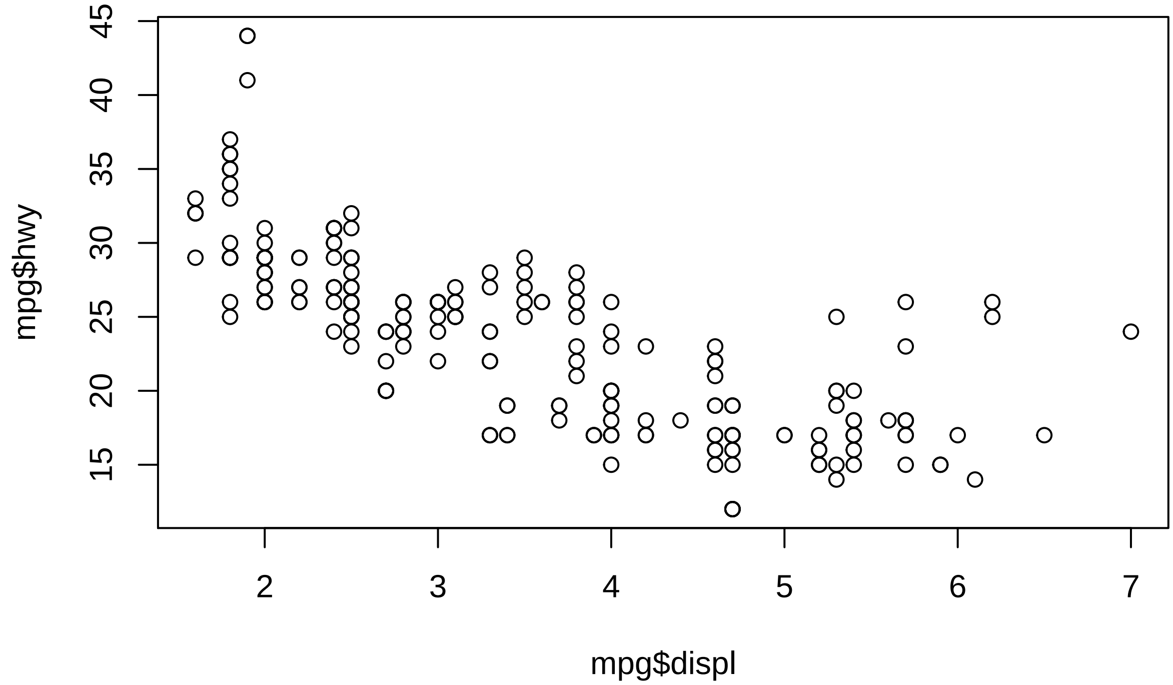
Boxplots

```r
boxplot(hwy ~ drv, data = mpg)
```

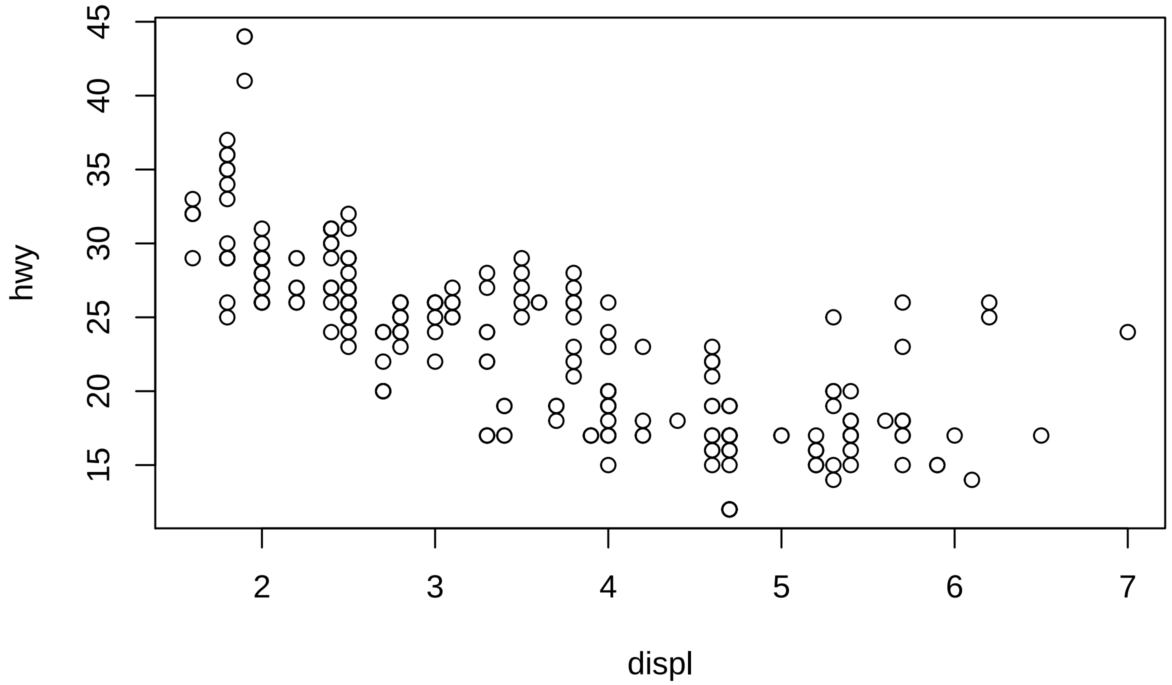


### Scatter plot

```r
plot(mpg$displ, mpg$hwy)
```



```r
plot( hwy ~ displ, data = mpg)
```



###

TUTORIAL `runTutorial('chapter2')` `runTutorial('correlation')`

Tydiverse ??

Dplyr

```r
data = mpg %>%
  filter(hwy > 30) %>%
  mutate(Test = hwy / cty) %>%
  select(manufacturer, Test, hwy)
```