

A brief Introduction to R

Song Wang

January 20, 2016

1. Download and install R and Rstudio

Visit <http://cran.us.r-project.org>

<https://www.rstudio.com/products/rstudio/download/>

2. play with R

- arithmetic operation

```
1+2
```

```
## [1] 3
```

```
3*4
```

```
## [1] 12
```

```
3/6
```

```
## [1] 0.5
```

```
3>5
```

```
## [1] FALSE
```

- use easy in-built functions

```
x <- c(12, 11, 16, 20) # c standards for ?  
?c  
sum(x)
```

```
## [1] 59
```

```
max(x)
```

```
## [1] 20
```

```
mean(x)
```

```
## [1] 14.75
```

```
median(x)
```

```
## [1] 14
```

```
sd(x)
```

```
## [1] 4.112988
```

```
var(x)
```

```
## [1] 16.91667
```

```
?mean
```

3. Get Data into R

- use import Dataset button in the top right corner of Rstudio (discouraged) click the button, select the file you want to import.
- use read.table command

```
rm(list=ls())  
mtcars <- read.table(file = "mtcars.csv", sep=";", header = T)  
# mtcars <- read.csv(file = "mtcars.csv")  
?read.csv  
?read.table  
str(mtcars)
```

```
## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : int 6 6 4 6 8 6 8 4 4 6 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : int 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...  
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...  
## $ qsec: num 16.5 17 18.6 19.4 17 ...  
## $ vs : int 0 0 1 1 0 1 0 1 1 1 ...  
## $ am : int 1 1 1 0 0 0 0 0 0 0 ...  
## $ gear: int 4 4 4 3 3 3 3 4 4 4 ...  
## $ carb: int 4 4 1 1 2 1 4 2 2 4 ...
```

```
?mtcars  
mean(mtcars$mpg)
```

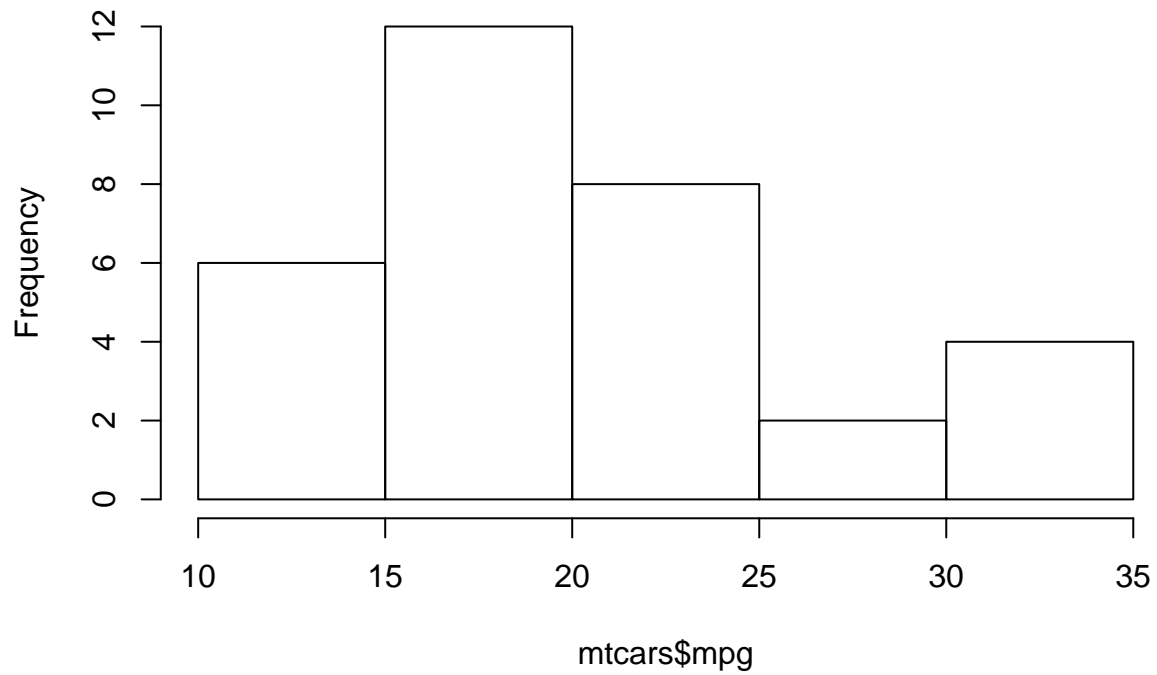
```
## [1] 20.09062
```

4. Visualize your data

```
# for numeric variable (1 or 2 variables)
```

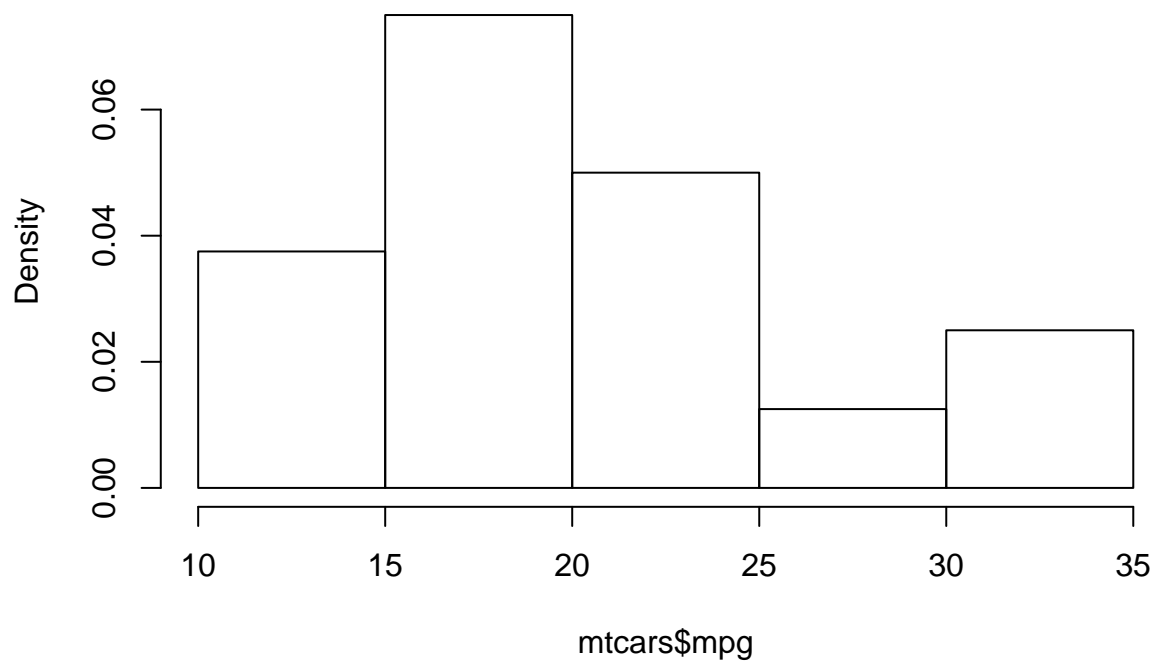
```
hist(mtcars$mpg)
```

Histogram of mtcars\$mpg



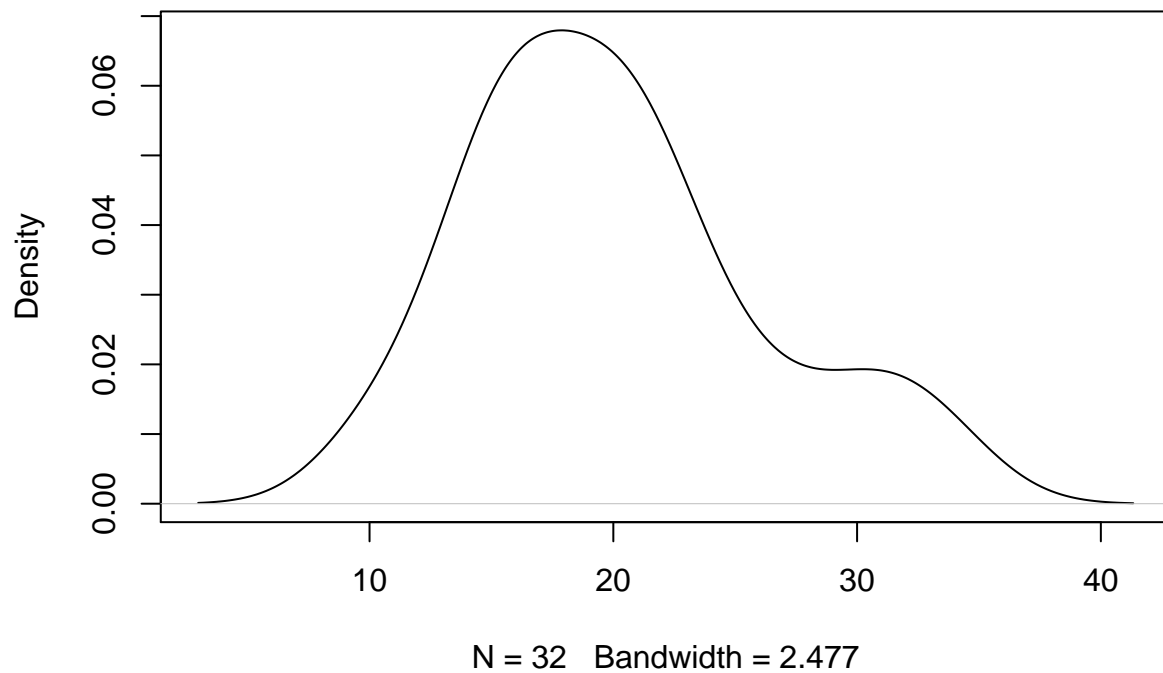
```
hist(mtcars$mpg, freq = FALSE) #density histogram
```

Histogram of mtcars\$mpg

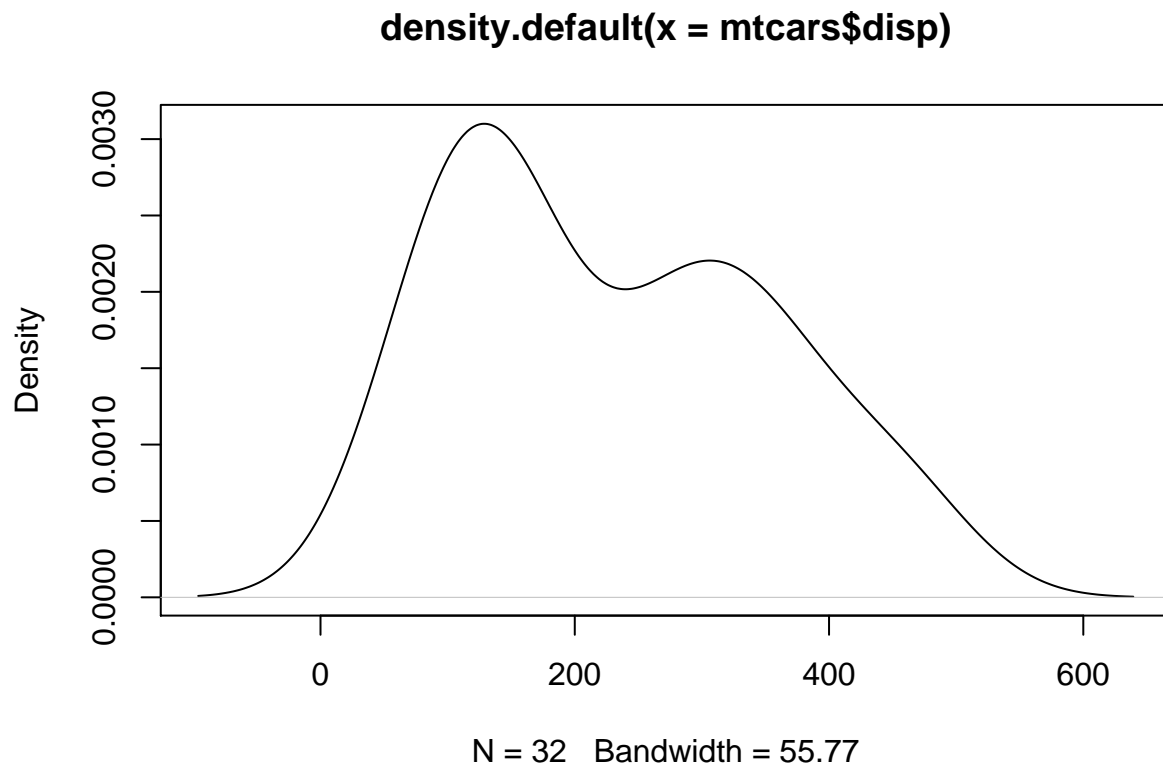


```
plot(density(mtcars$mpg))
```

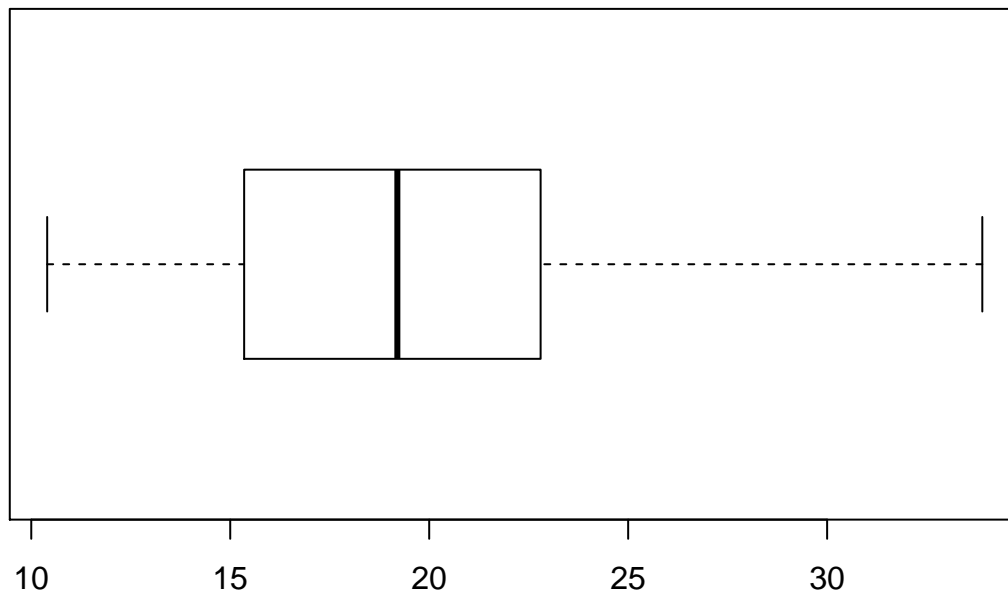
density.default(x = mtcars\$mpg)



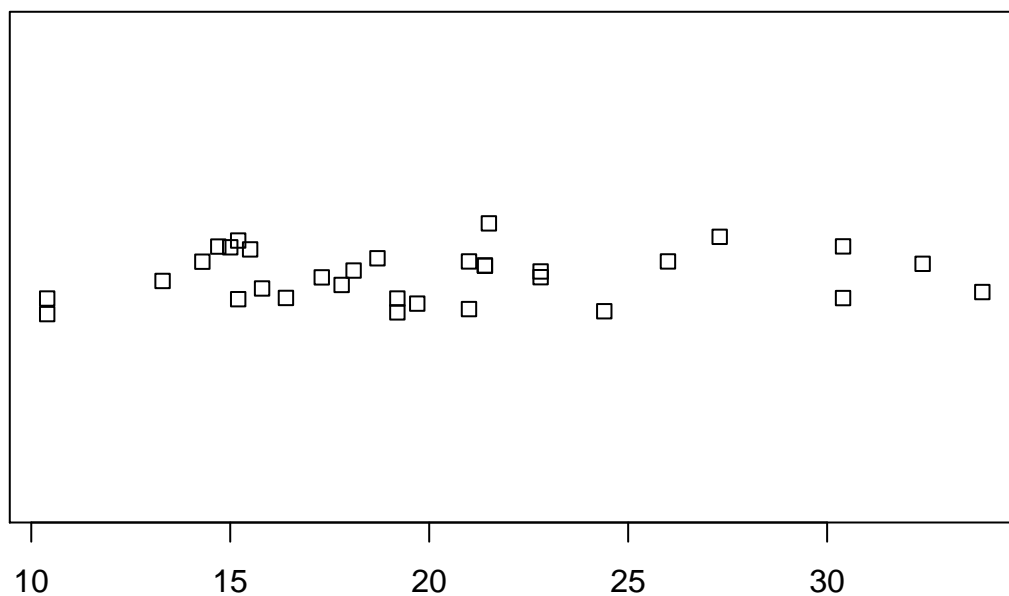
```
plot(density(mtcars$disp))
```



```
boxplot(mtcars$mpg, horizontal = T)
```



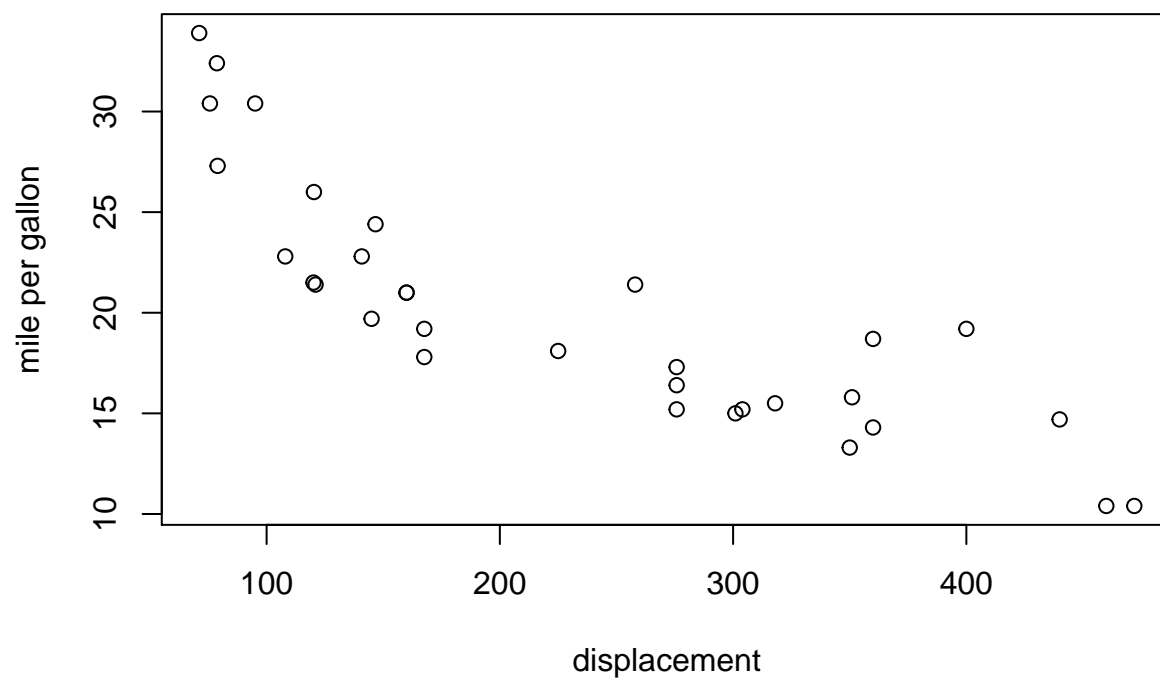
```
stripchart(mtcars$mpg, method = 'jitter')
```



```
# two numerical variables
```

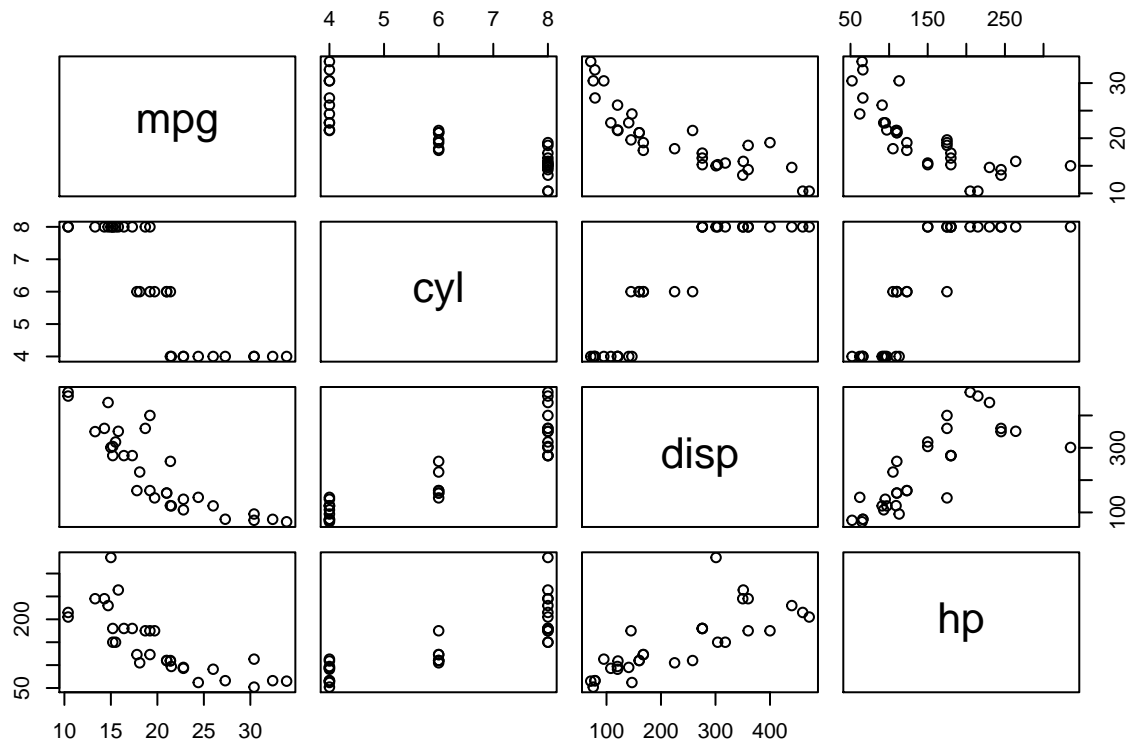
```
plot(mtcars$disp, mtcars$mpg, xlab="displacement", ylab = "mile per gallon", main = "mpg vs. disp") #
```

mpg vs. disp

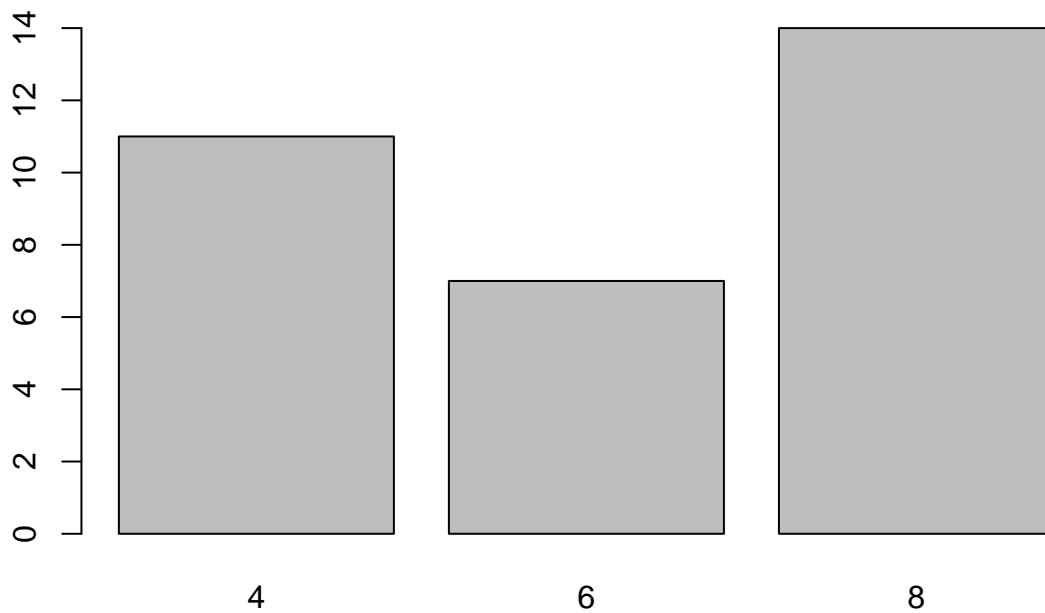


```
#abline(lm( mtcars$mpg ~ mtcars$disp ), col="red")
```

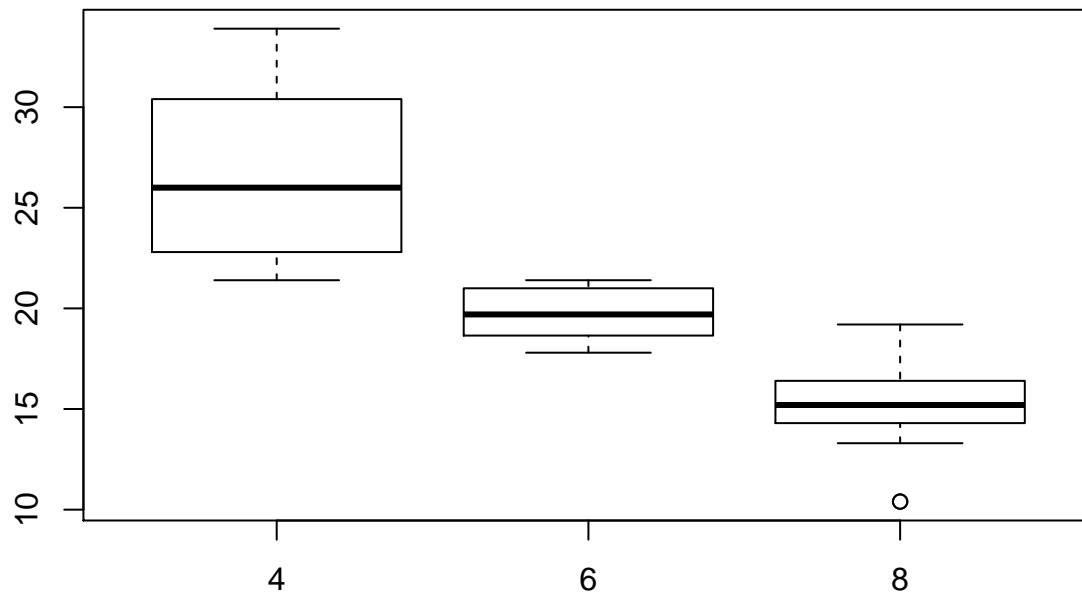
```
pairs(mtcars[,1:4]) # pairwise plot
```



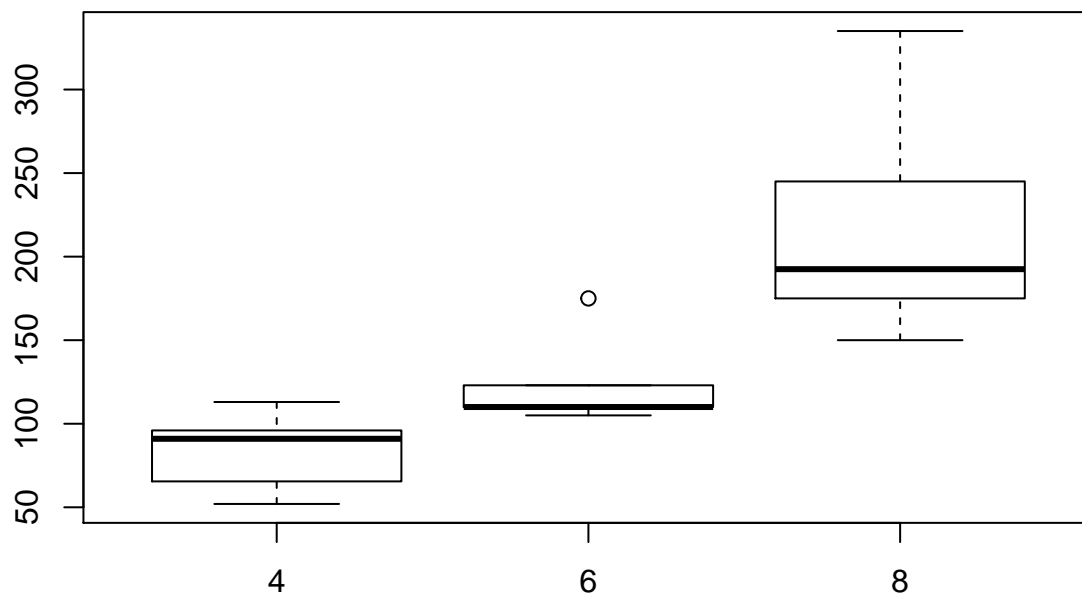
```
# For categorical data
counts = table(mtcars$cyl); barplot(counts)
```



```
boxplot(mtcars$mpg ~ mtcars$cyl)
```



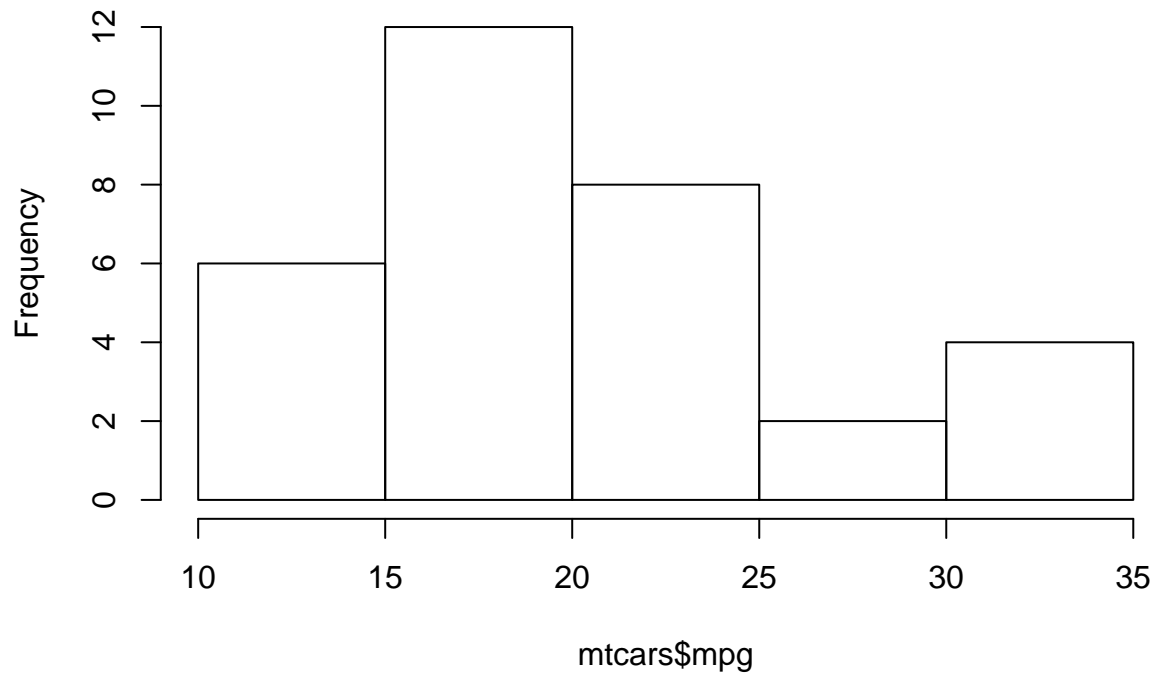
```
boxplot(mtcars$hp ~ mtcars$cyl)
```



- Checking Normality

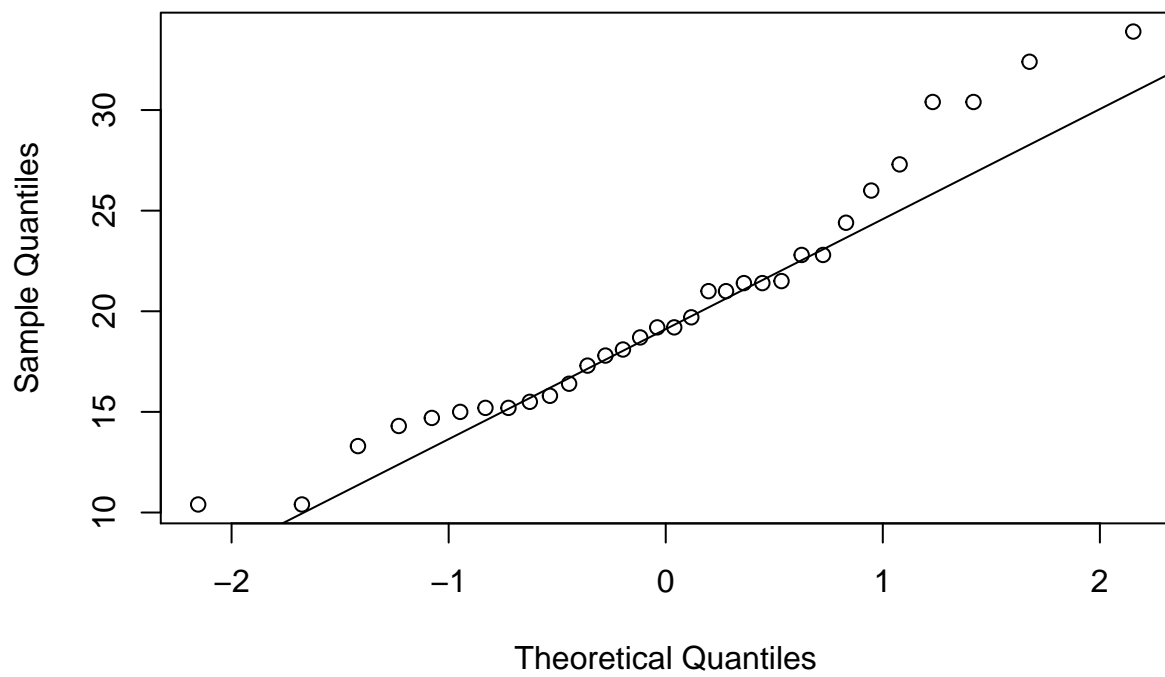
```
hist(mtcars$mpg)
```


Histogram of mtcars\$mpg

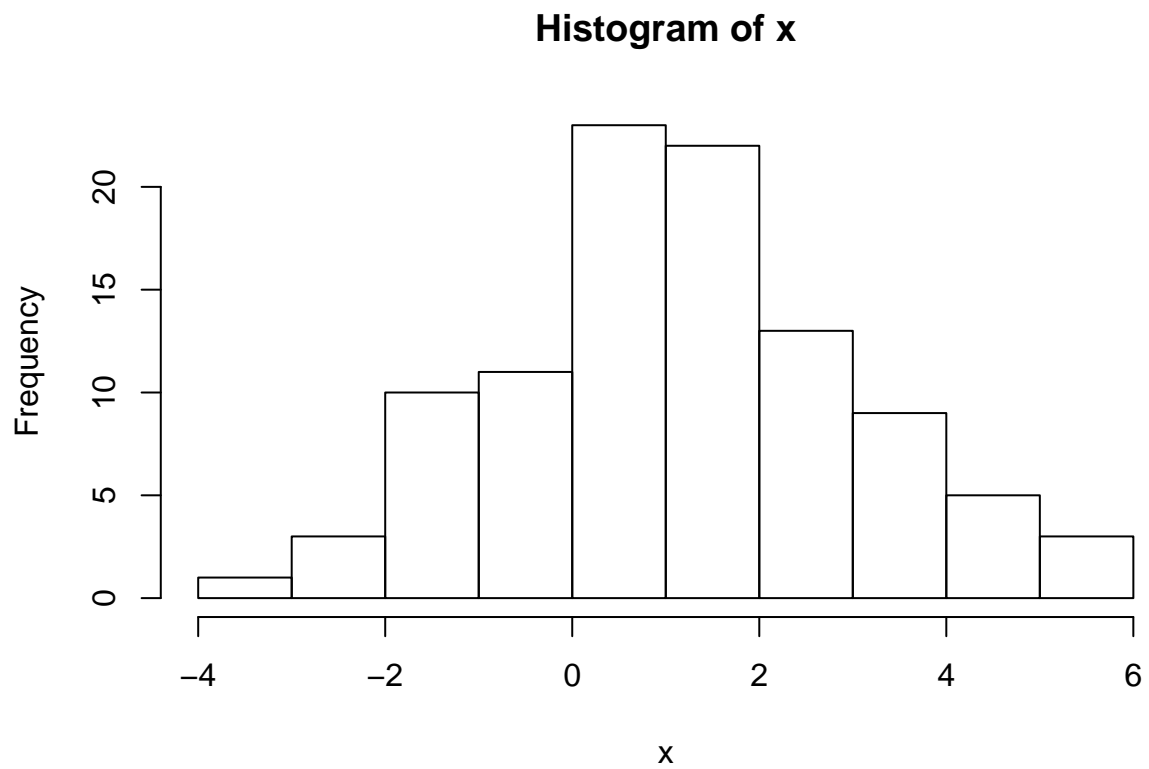


```
qqnorm(mtcars$mpg)  
qqline(mtcars$mpg)
```

Normal Q-Q Plot

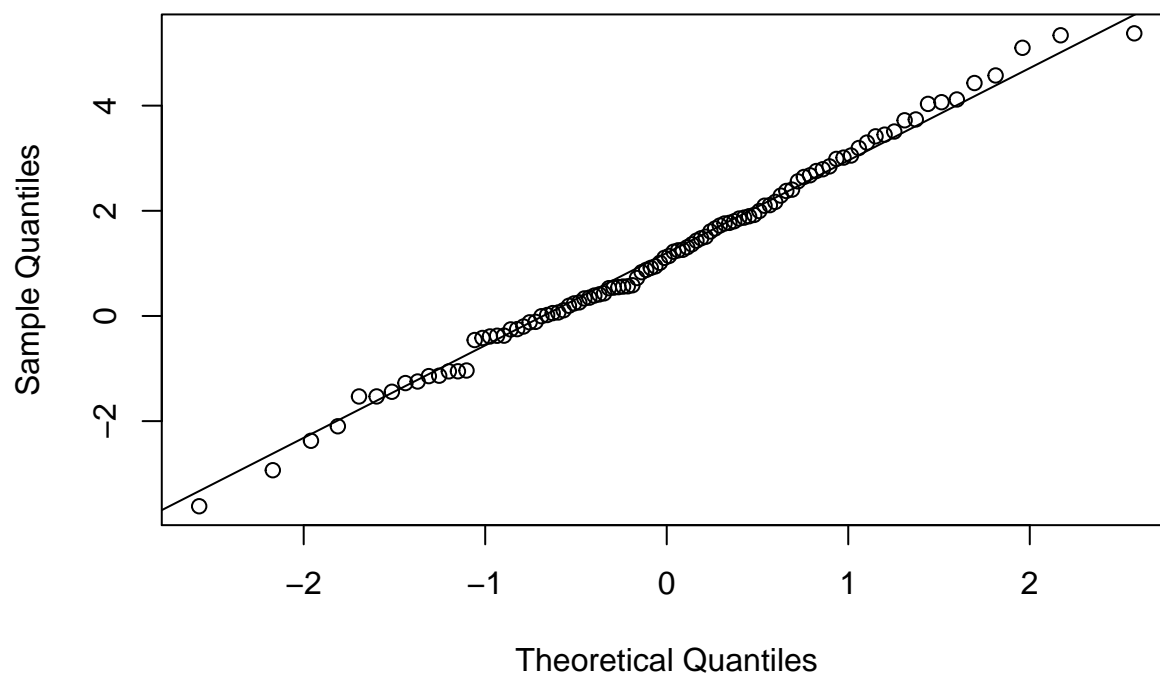


```
set.seed(123)
x <- rnorm(100,1,2)
hist(x)
```



```
qqnorm(x)
qqline(x)
```

Normal Q-Q Plot



5. How to save the plot and use in the report later

- click on the plot (at bottom right corner usually)
- save the plot in .png or .pdf
- copy the plot into the report (in the future project)