Home Activity

15F-8069

#Load Dataset

```
data("airquality")
```

#Data Exploration

```
str(airquality)
```

```
## 'data.frame':    153 obs. of  6 variables:
##  $ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
##  $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
##  $ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
##  $ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...
##  $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ Day    : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
head(airquality, n=3)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
```

```
tail(airquality, n=3)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 151    14     191 14.3   75     9  28
## 152    18     131  8.0   76     9  29
## 153    20     223 11.5   68     9  30
```

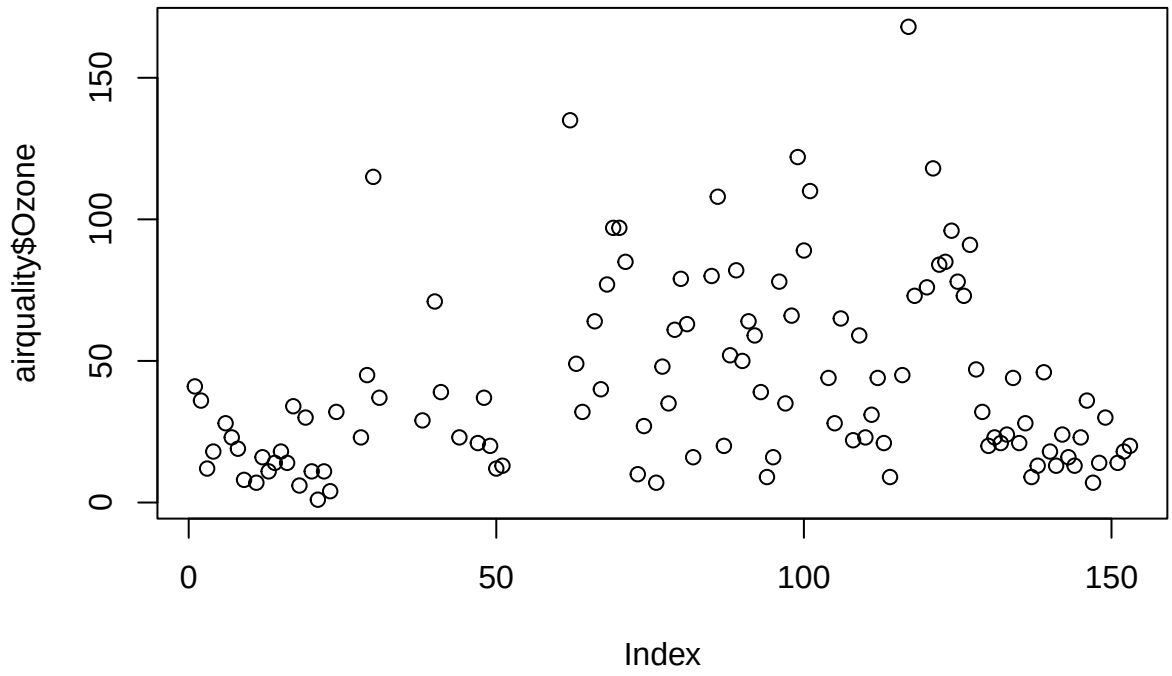
```
summary(airquality)
```

```
##      Ozone           Solar.R           Wind             Temp      
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00  
##  1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00  
##  Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
##  Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88  
##  3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
##  NA's   :37       NA's   :7
```

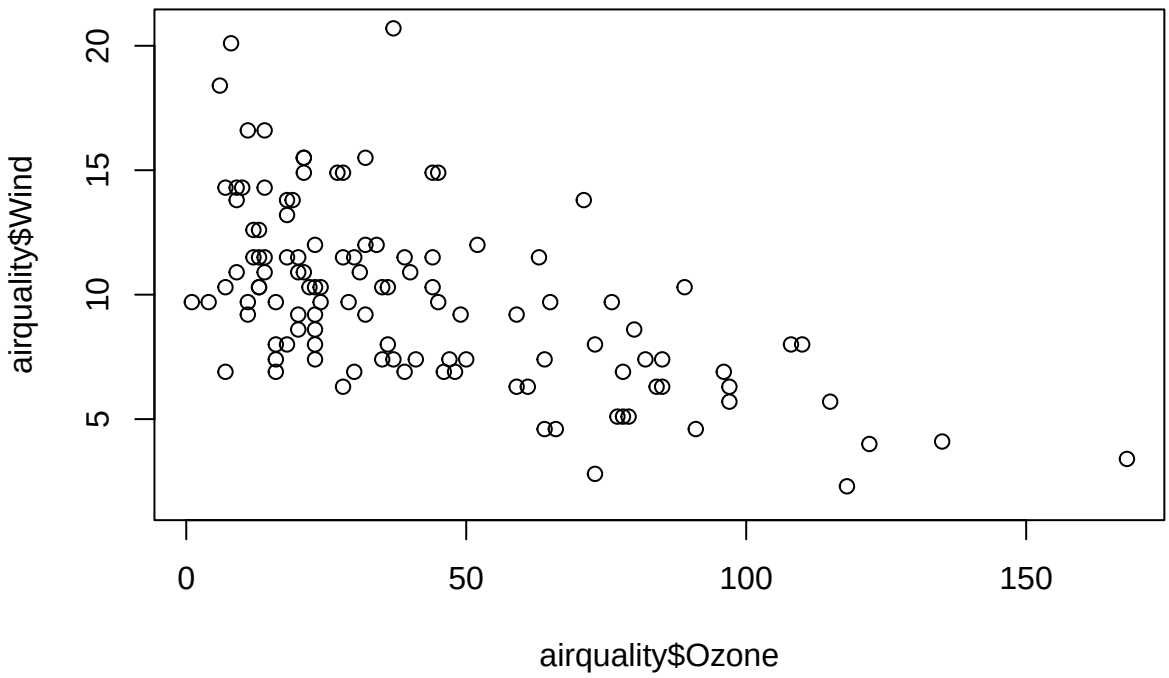
```
##      Month            Day
##  Min.   :5.000   Min.   : 1.0
##  1st Qu.:6.000   1st Qu.: 8.0
##  Median :7.000   Median :16.0
##  Mean   :6.993   Mean   :15.8
##  3rd Qu.:8.000   3rd Qu.:23.0
##  Max.   :9.000   Max.   :31.0
##
```

#Getting Started with Basic Plots

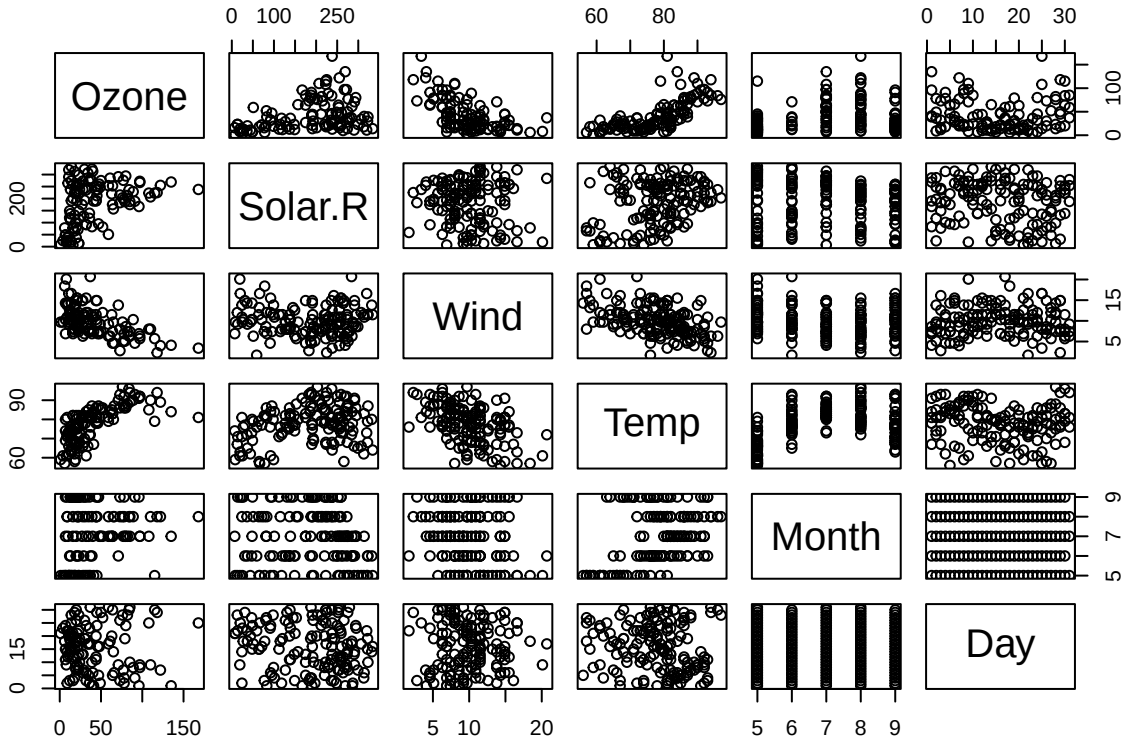
```
plot(airquality$Ozone)
```




```
plot(airquality$Ozone, airquality$Wind)
```

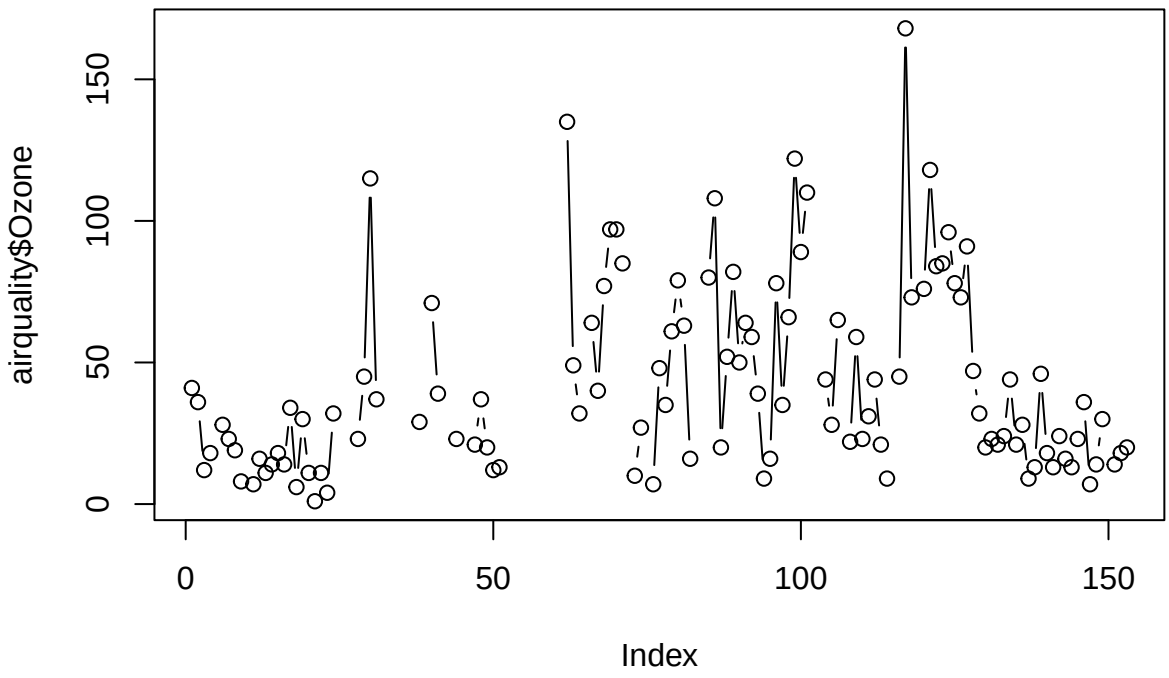



```
plot(airquality)
```

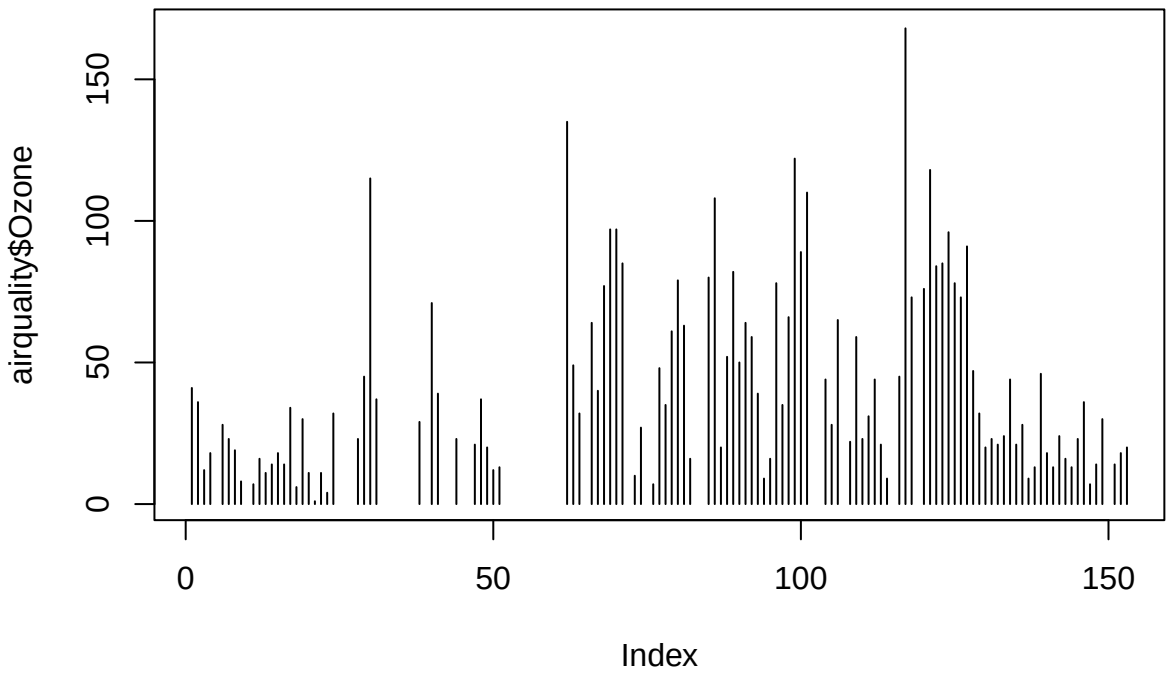


#Using arguments with the plot() function

```
# points and lines
plot(airquality$Ozone, type= "b")
```

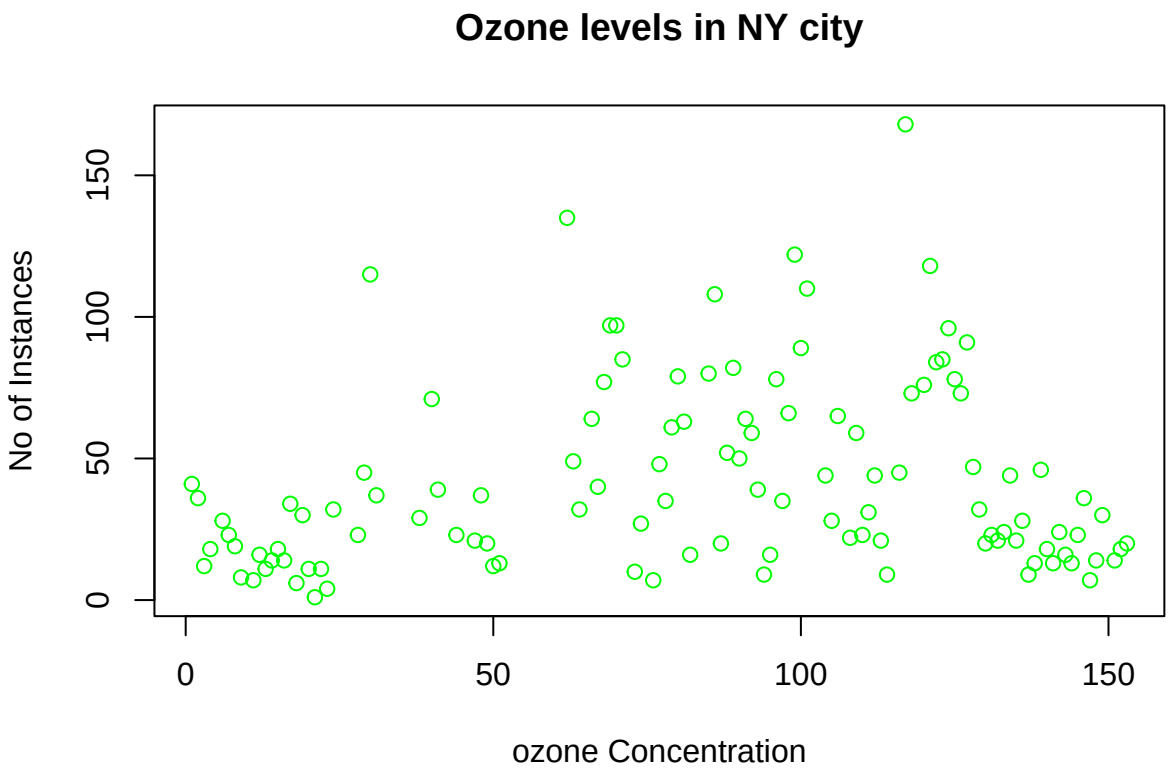



```
# high density vertical lines.
plot(airquality$Ozone, type= "h")
```



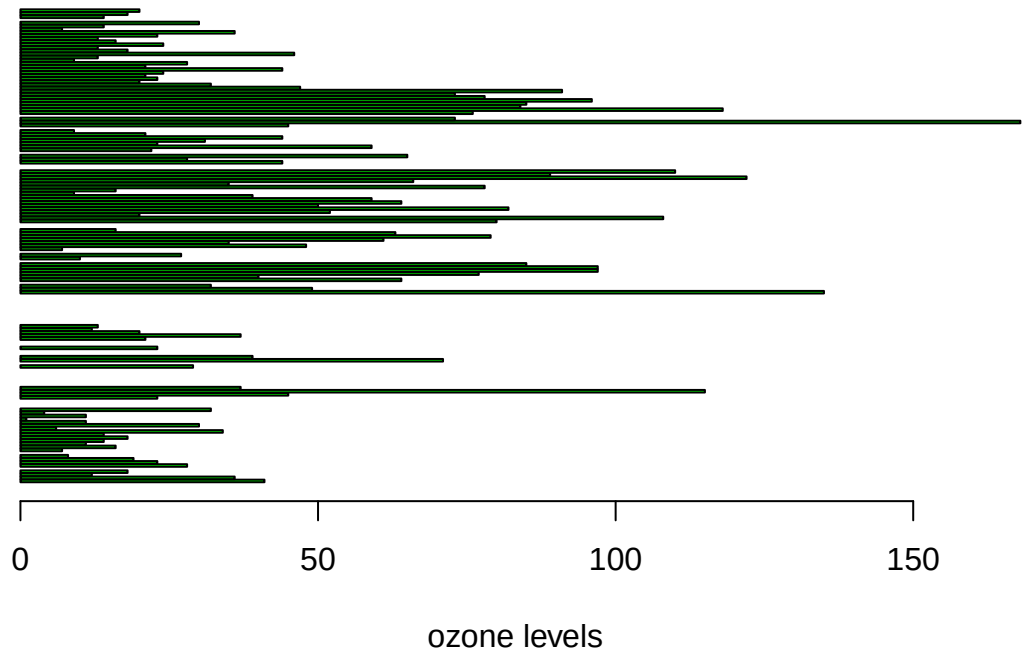
#Labels and Titles

```
plot(airquality$Ozone, xlab = 'ozone Concentration', ylab = 'No of Instances', main = 'Ozone levels in
```

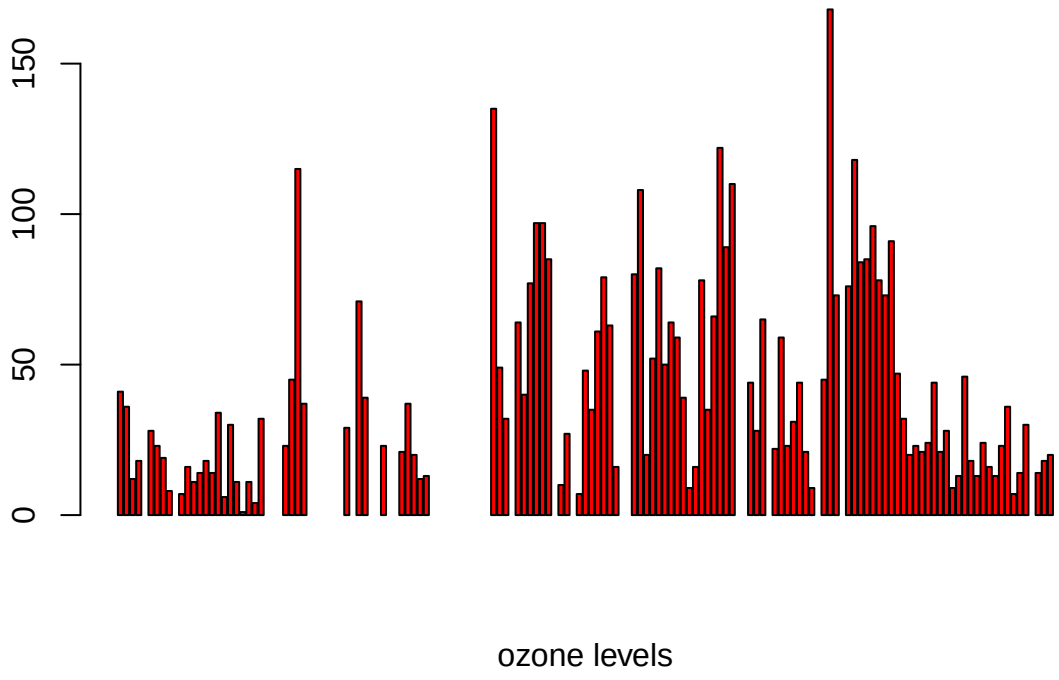


2.Barplot

```
# Horizontal bar plot
barplot(airquality$Ozone, main = 'Ozone Concenteration in air',xlab = 'ozone levels', col= 'green',hori
```

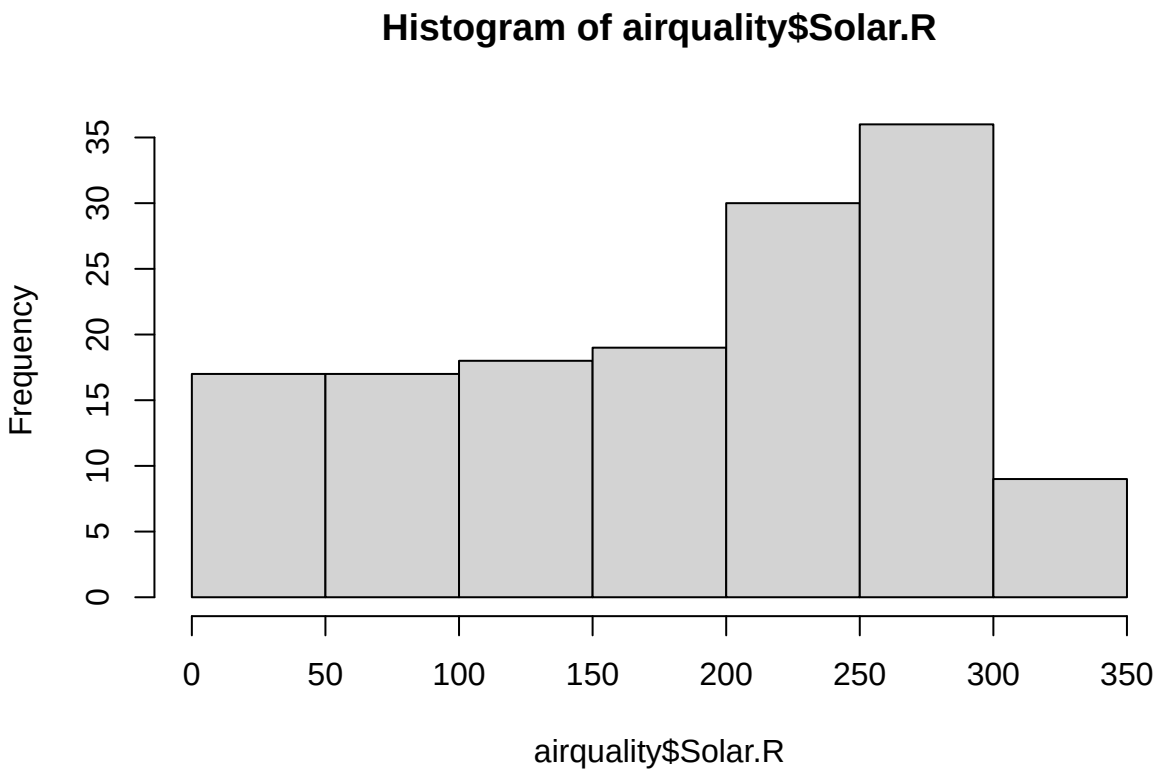



```
# Vertical bar plot
barplot(airquality$Ozone, main = 'Ozone Concenteration in air',xlab = 'ozone levels', col='red',horiz =
```

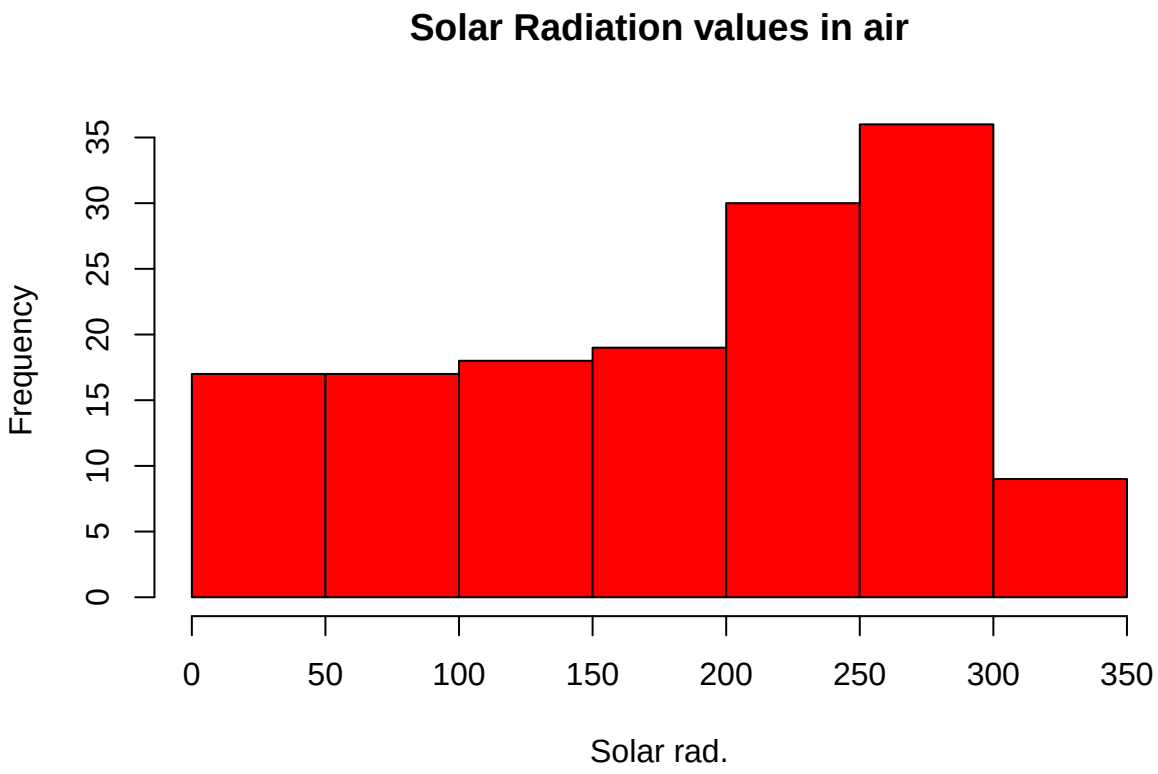


3.Histogram

```
hist(airquality$Solar.R)
```

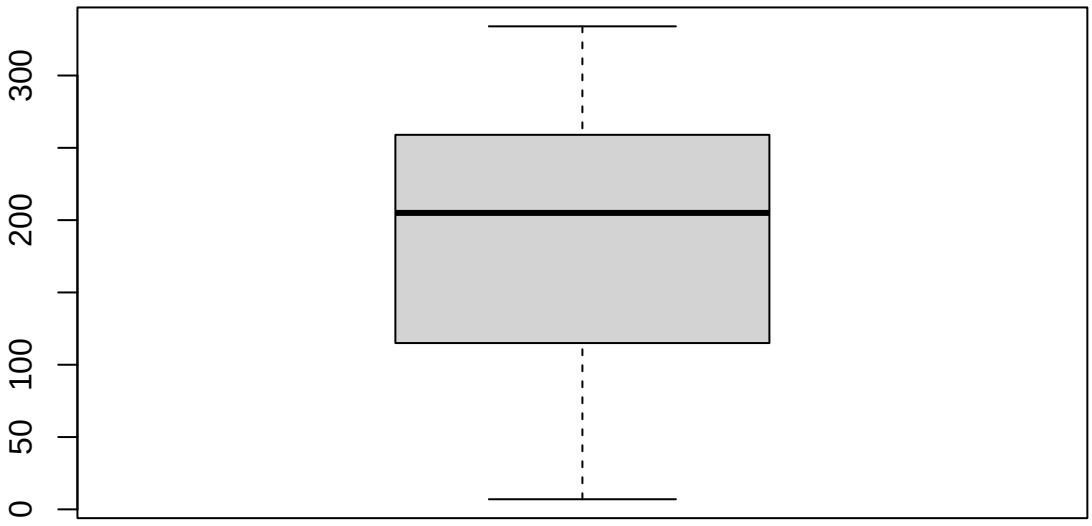



```
#coloured histogram
hist(airquality$Solar.R, main = 'Solar Radiation values in air',xlab = 'Solar rad.', col='red')
```

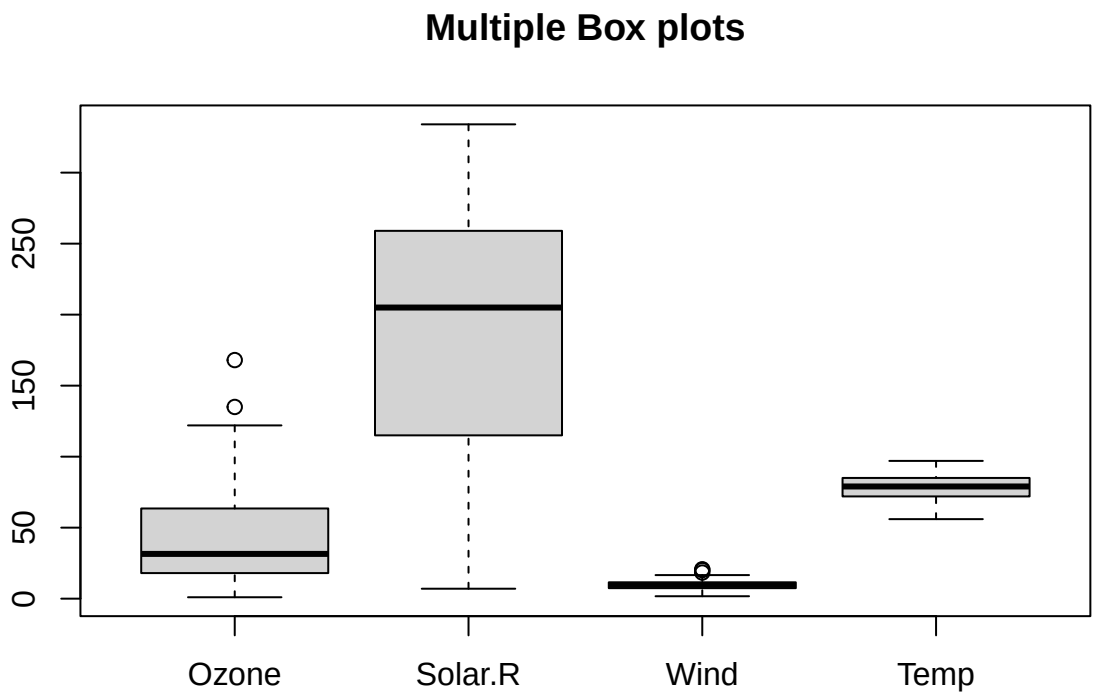


4.Boxplot

```
#Single box plot
boxplot(airquality$Solar.R)
```

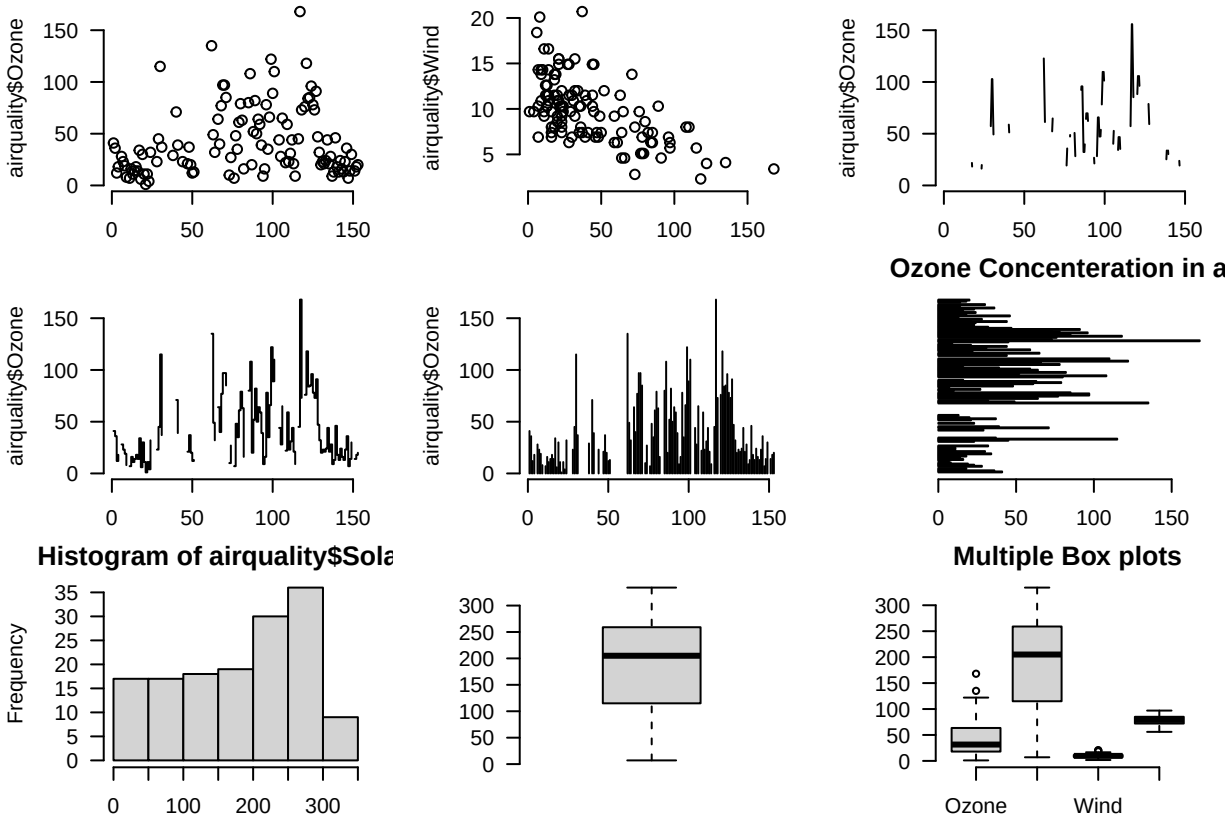



```
# Multiple box plots
boxplot(airquality[,0:4], main='Multiple Box plots')
```



#5. Grid of Charts

```
par(mfrow=c(3,3), mar=c(2,5,2,1), las=1, bty="n")
plot(airquality$Ozone)
plot(airquality$Ozone, airquality$Wind)
plot(airquality$Ozone, type= "c")
plot(airquality$Ozone, type= "s")
plot(airquality$Ozone, type= "h")
barplot(airquality$Ozone, main = 'Ozone Concenteration in air',xlab = 'ozone levels', col='green',horiz
hist(airquality$Solar.R)
boxplot(airquality$Solar.R)
boxplot(airquality[,0:4], main='Multiple Box plots')
```



#Visualisation libraries in R

```
library(lattice)

#Loading the dataset
attach(mtcars)

# Exploring the dataset
head(mtcars)
```

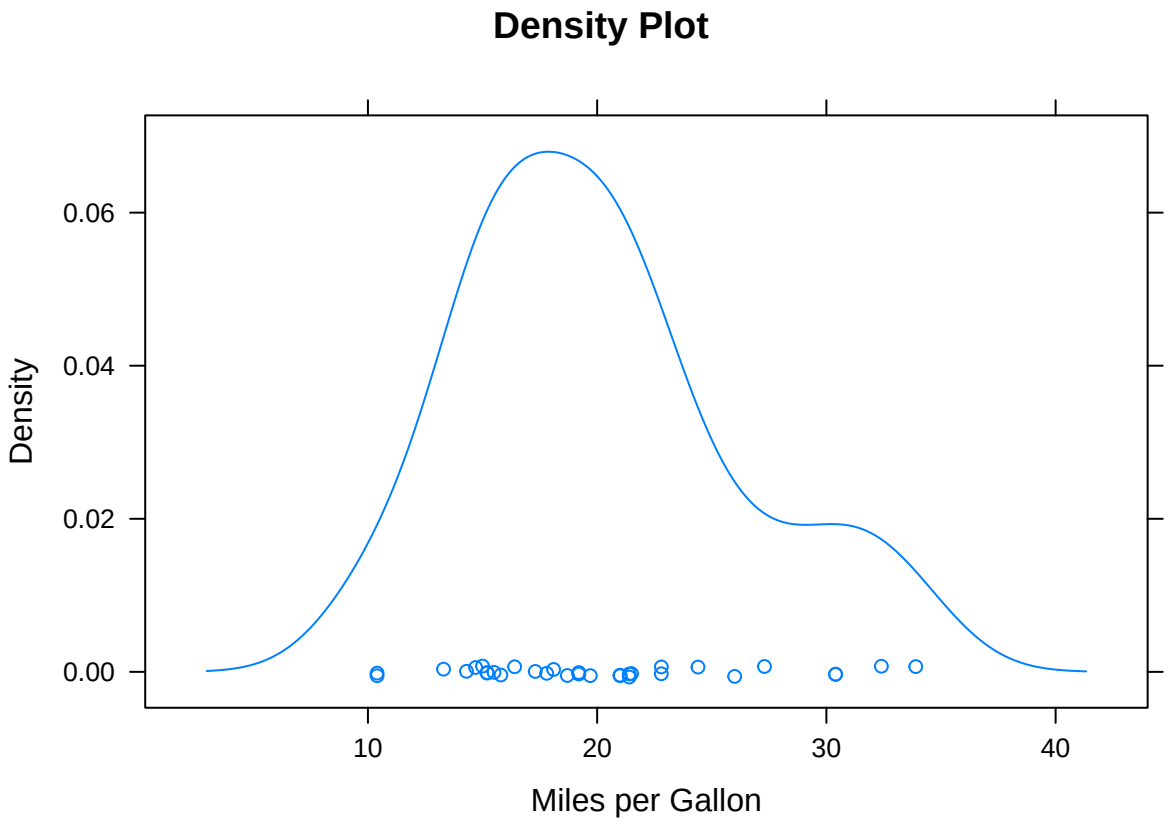
```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
gear_factor<-factor(gear,levels=c(3,4,5),
                    labels=c("3gears","4gears","5gears"))
cyl_factor <-factor(cyl,levels=c(4,6,8),
                    labels=c("4cyl","6cyl","8cyl"))
```

We can use the lattice package to create some basic plots in R.

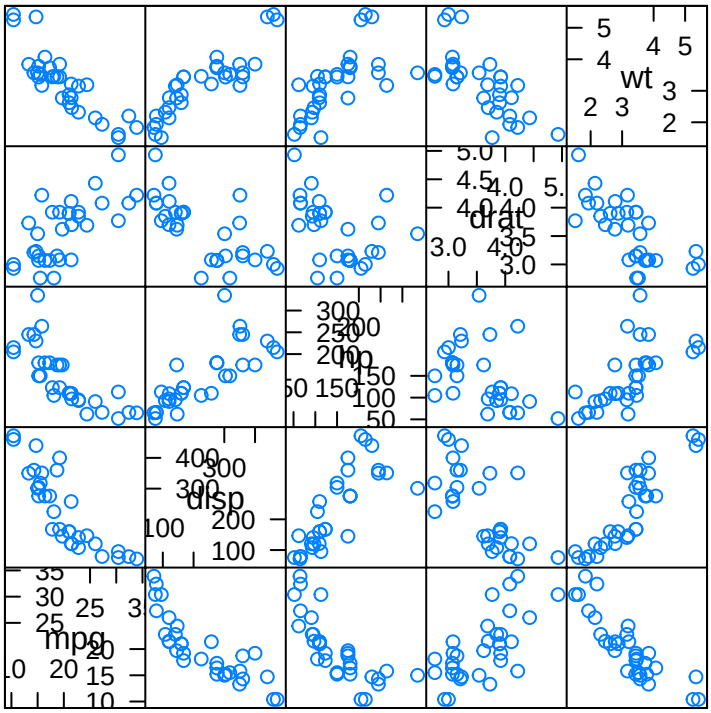
Kernel density plots

```
densityplot(~mpg, main="Density Plot",  xlab="Miles per Gallon")
```



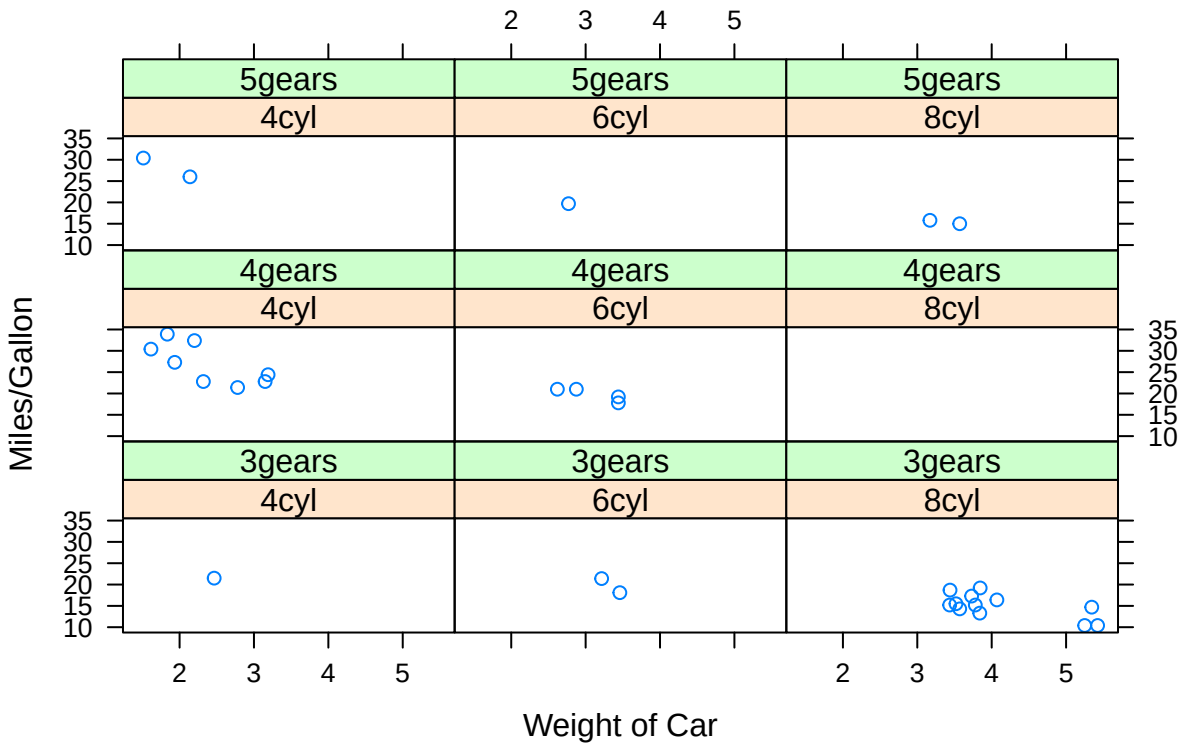
#scatterplot matrix

```
splom(mtcars[c(1,3,4,5,6)], main="MTCARS Data")
```



#Scatterplots depicting a combination of two factors

```
xyplot(mpg~wt|cyl_factor*gear_factor,
       main="Scatterplots : Cylinders and Gears",
       ylab="Miles/Gallon", xlab="Weight of Car")
```



#ggplot2

```
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'mtcars':
##
##     mpg
```

```
#Loading the dataset
attach(mtcars)
```

```
## The following object is masked from package:ggplot2:
##
##     mpg
```

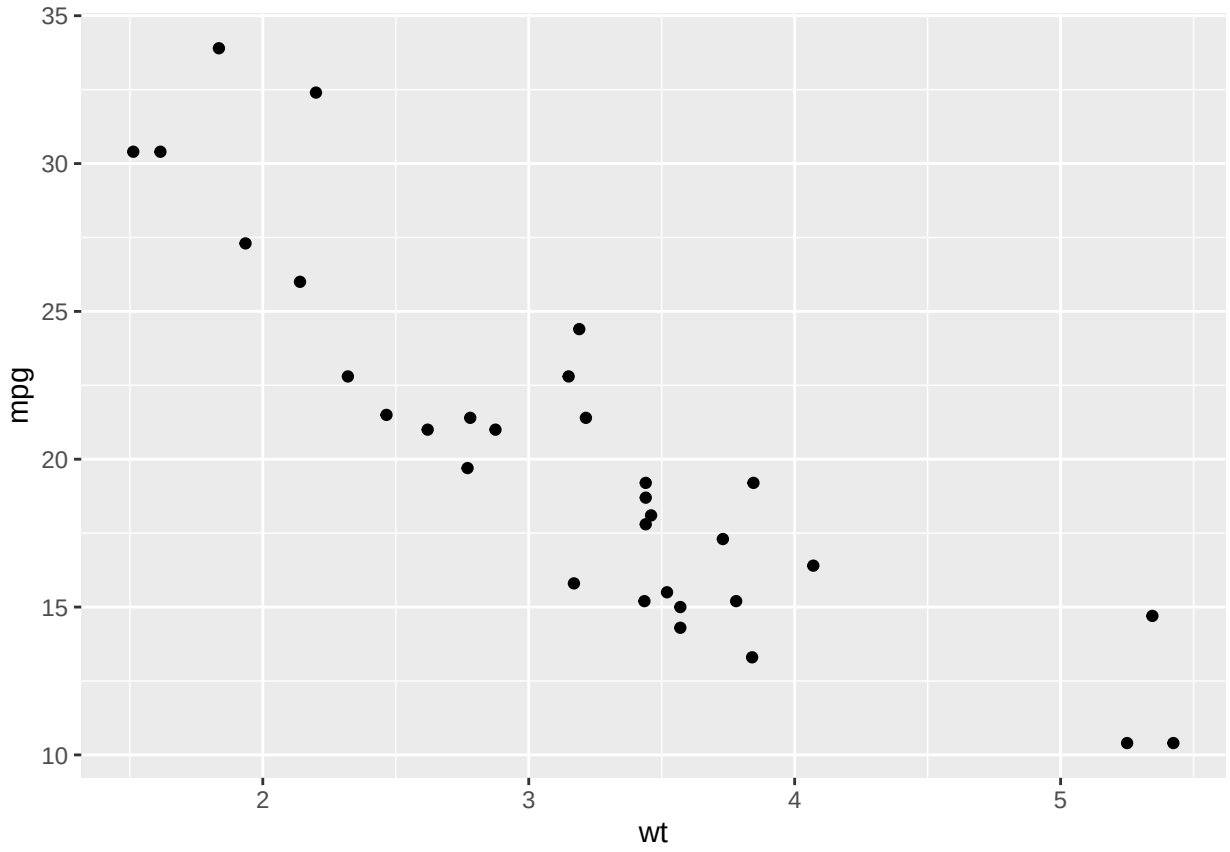
```
## The following objects are masked from mtcars (pos = 4):
##
##     am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt
```

```
# create factors with value labels
mtcars$gear <- factor(mtcars$gear,levels=c(3,4,5),
                      labels=c("3gears", "4gears", "5gears"))
mtcars$am <- factor(mtcars$am,levels=c(0,1),
                    labels=c("Automatic","Manual"))
mtcars$cyl <- factor(mtcars$cyl,levels=c(4,6,8),
                     labels=c("4cyl","6cyl","8cyl"))
```

#create a few plots to understand the capability of ggplot2

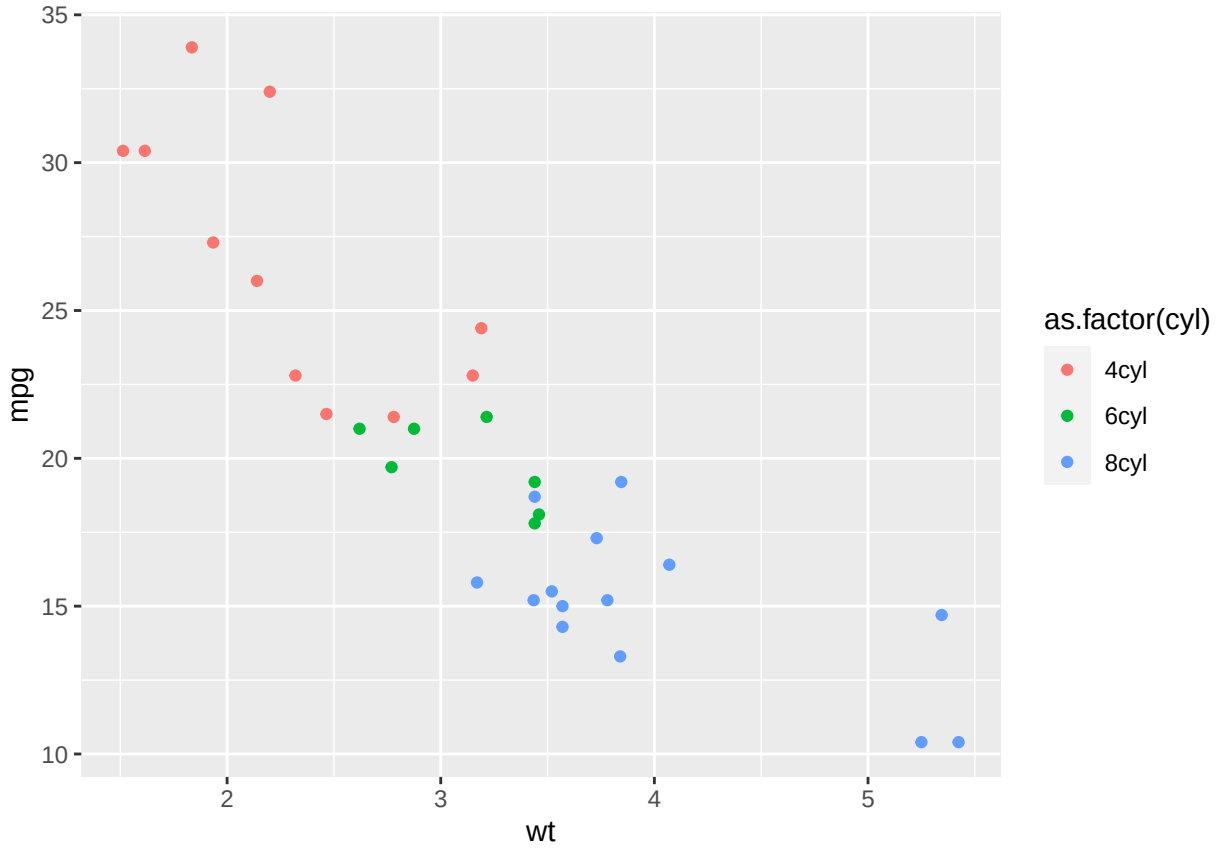
#Scatter Plots

```
ggplot(data = mtcars, mapping = aes(x = wt, y = mpg)) + geom_point()
```



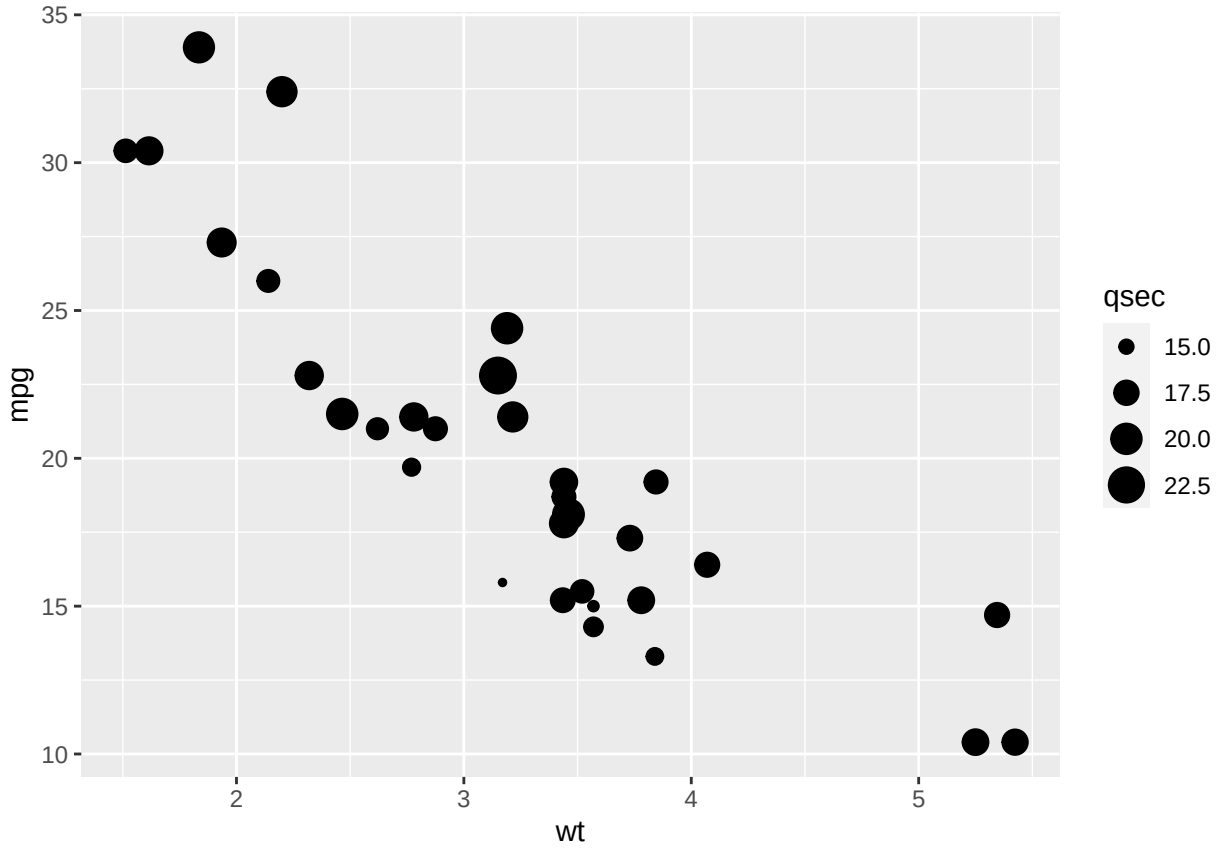
#Styling scatter plots by factor

```
ggplot(data = mtcars, mapping = aes(x = wt, y = mpg, color = as.factor(cyl))) + geom_point()
```



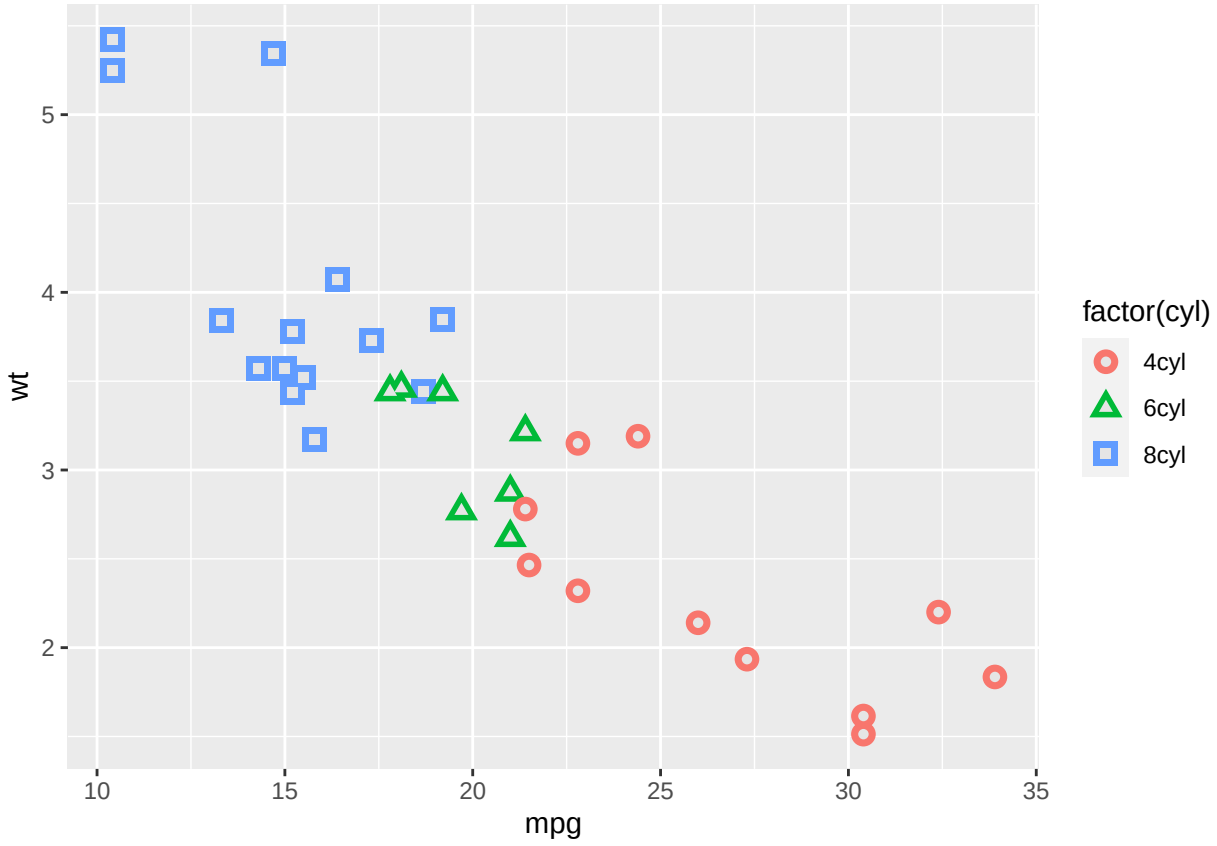
#Styling scatter plots by size

```
ggplot(data = mtcars, mapping = aes(x = wt, y = mpg, size = qsec)) + geom_point()
```



#Different symbols for different sizes

```
p  <-  ggplot(mtcars,aes(mpg, wt, shape  =  factor(cyl)))
p + geom_point(aes(colour  =  factor(cyl)), size  =  4) + geom_point(colour  =  "grey90", size  =  1.5)
```



3. Plotly

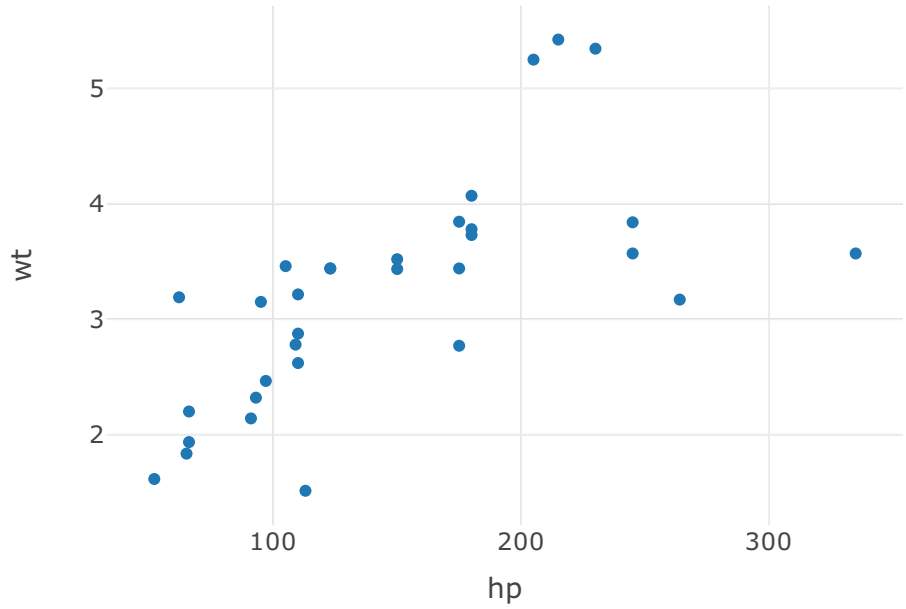
```
library(plotly)
```

```
##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

#Basic Scatter Plot

```
p <- plot_ly(data = mtcars, x = ~hp, y = ~wt)
p
```

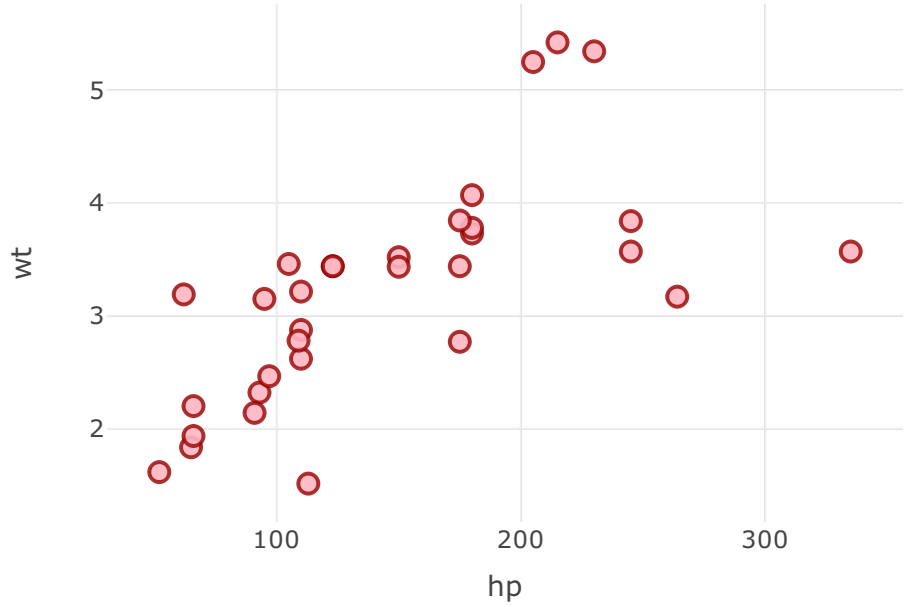
```
## No trace type specified:
##   Based on info supplied, a 'scatter' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter

## No scatter mode specifed:
##   Setting the mode to markers
##   Read more about this attribute -> https://plotly.com/r/reference/#scatter-mode
```



#Styled Scatter Plot

```
p <- plot_ly(data = mtcars, x = ~hp, y = ~wt, marker = list(size = 10, color = 'rgba(255, 182, 193, .9)
p
```

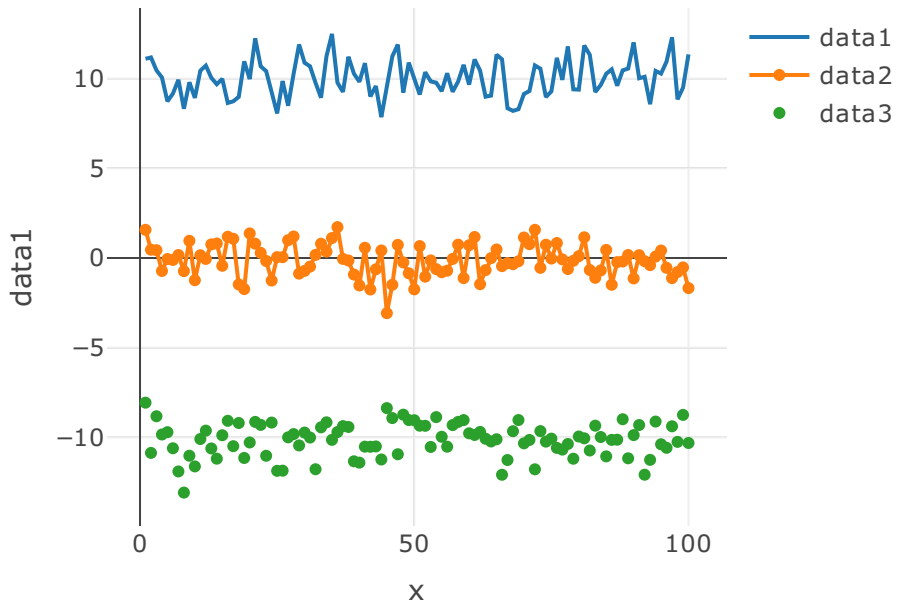
```
## No trace type specified:
##   Based on info supplied, a 'scatter' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter

## No scatter mode specifed:
##   Setting the mode to markers
##   Read more about this attribute -> https://plotly.com/r/reference/#scatter-mode
```



#Markers and Lines

```
data1 <- rnorm(100, mean = 10)
data2 <- rnorm(100, mean = 0)
data3 <- rnorm(100, mean = -10)
x <- c(1:100)
data <- data.frame(x, data1, data2, data3)
p <- plot_ly(data, x = ~x)%>%

  add_trace(y = ~data1, name = 'data1',mode = 'lines')%>%
  add_trace(y = ~data2, name = 'data2', mode = 'lines+markers')%>%
  add_trace(y = ~data3, name = 'data3', mode = 'markers')
p
```

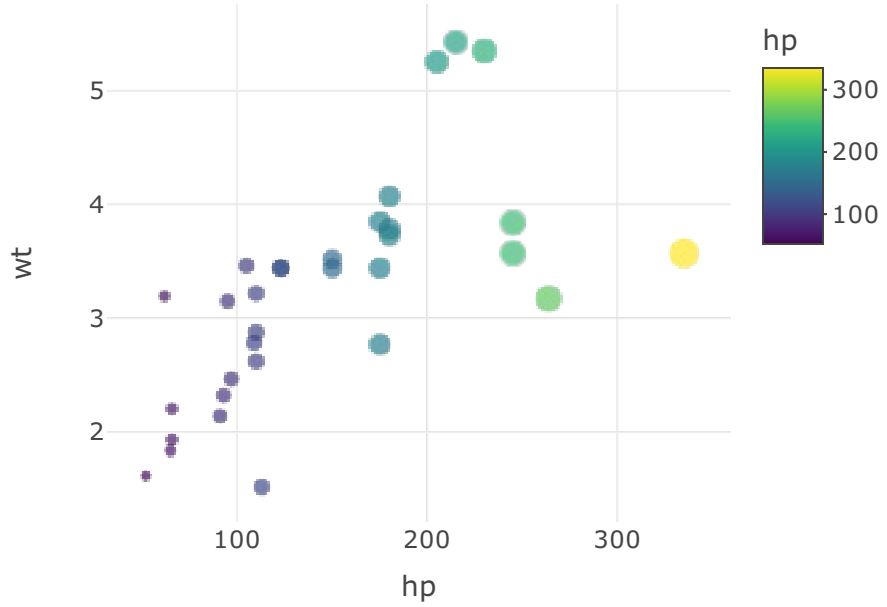
```
## No trace type specified:
##   Based on info supplied, a 'scatter' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter
## No trace type specified:
##   Based on info supplied, a 'scatter' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter
## No trace type specified:
##   Based on info supplied, a 'scatter' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter
```



#Adding Color and Size Mapping

```
p <- plot_ly(data = mtcars, x =~hp, y = ~wt,color = ~hp, size = ~hp )
p
```

```
## No trace type specified:
##   Based on info supplied, a 'scatter' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter

## No scatter mode specifed:
##   Setting the mode to markers
##   Read more about this attribute -> https://plotly.com/r/reference/#scatter-mode

## Warning: 'line.width' does not currently support multiple values.
```



#Visualising Geographical data in R

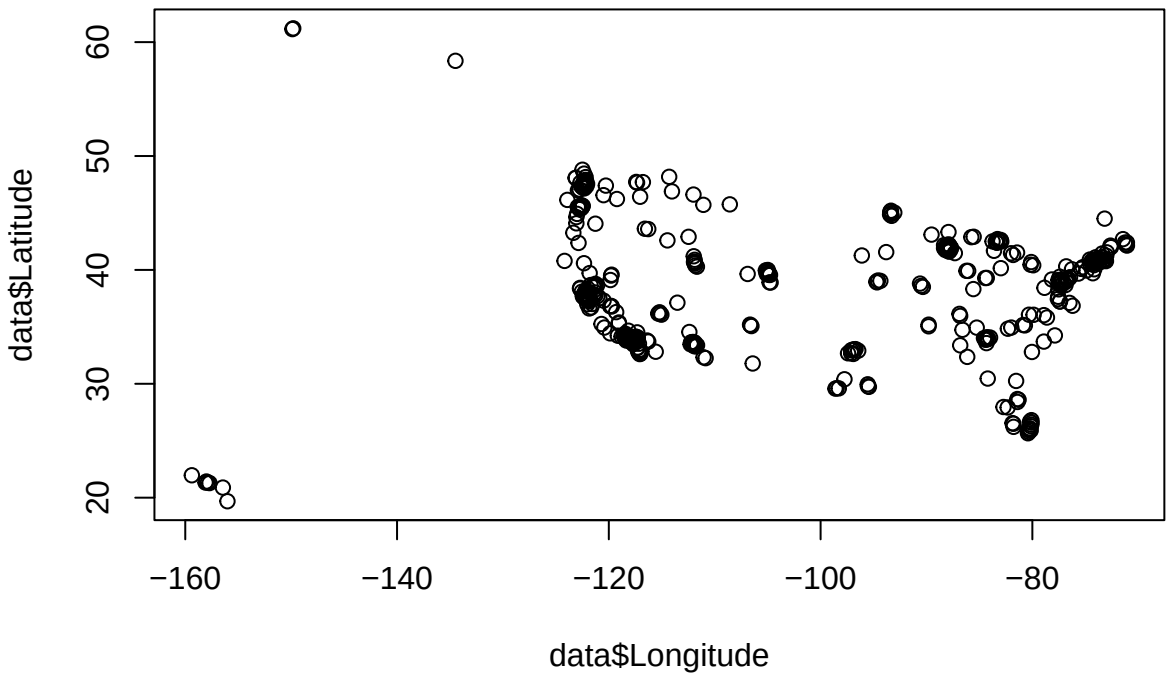
Geographical maps

```
data <- read.csv('ABC_locations.csv', sep=",")
head(data)
```

```
##                     Address       City    State   Zip.Code Latitude  Longitude
## 1  1205 N. Memorial Parkway Huntsville Alabama 35801-5930 34.74309  -86.60096
## 2      3650 Galleria Circle     Hoover Alabama 35244-2346 33.37765  -86.81242
## 3    8251 Eastchase Parkway Montgomery Alabama      36117 32.36389  -86.15088
## 4 5225 Commercial Boulevard     Juneau  Alaska 99801-7210 58.35920 -134.48300
## 5      330 West Dimond Blvd  Anchorage  Alaska 99515-1950 61.14327 -149.88422
## 6          4125 DeBarr Road  Anchorage  Alaska 99508-3115 61.21081 -149.80434
```

#plot() function

```
plot(data$Longitude,data$Latitude)
```



#map() function

```
library(maps)
```

#Using the map() function to plot a base map of the US

```
map(database="state")
```