Scatterplots & Line Plots
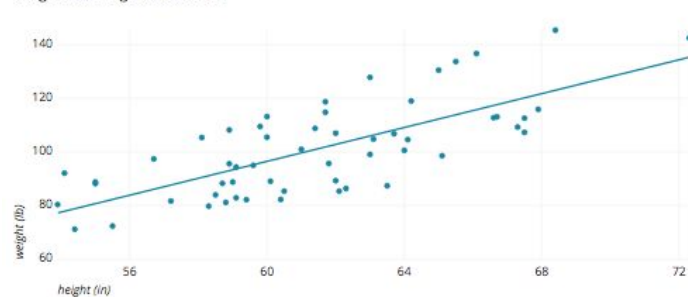
Kady Lazerte, Krista Ruggieri, and Tiffany Nguyen

Scatter Plot

- the values of two variables are plotted along two axes
- Used to reveal a pattern, if any of the resulting points revealing any correlation present
- Displays the relationship between two continuous variables
- Each value in a data set is represented by a point


Weight and Height of Children


Weight vs. Exercise
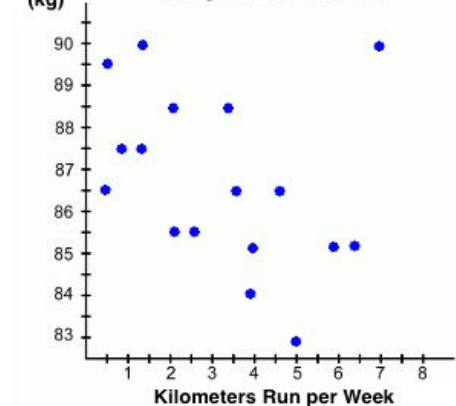
Line Plots

- a series of data points called 'markers' connected by straight line segments.
- similar to a scatter plot except that the measurement points are ordered (typically by their x-axis value) and joined with straight line segments
- often used to visualize a trend in data over intervals of time – a time series – thus the line is often drawn chronologically.

Money Held By Branch

Evolution of Bitcoin price
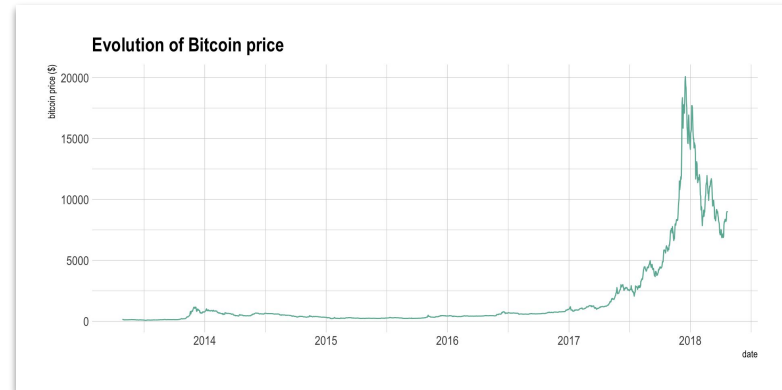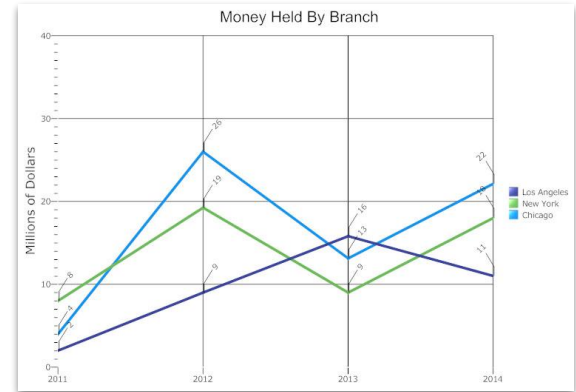
Line Plots Using Base Graphics

```
plot(AP, xlab="Year", type = "o", pch = 20,
     ylab = "Passenger numbers (1000's)",
     main="Air Passenger numbers from 1949 to 1961")
```
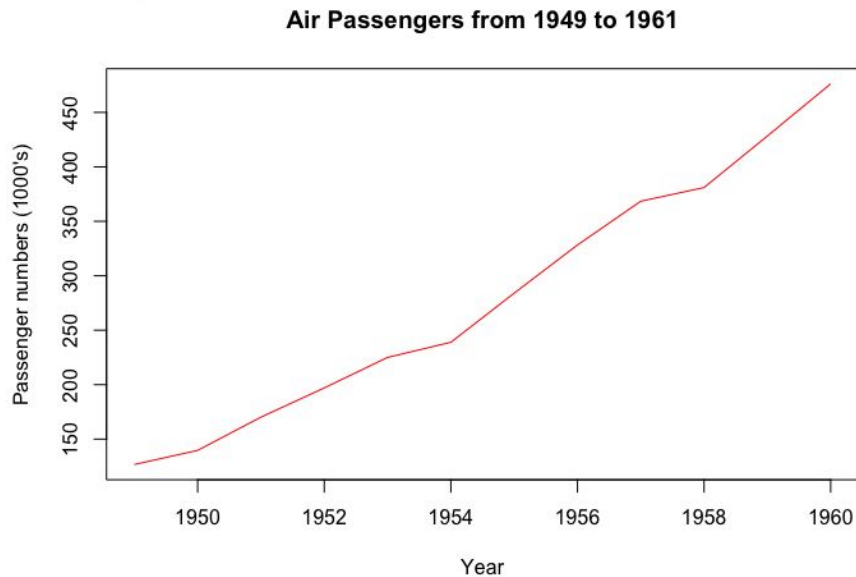
Air Passenger numbers from 1949 to 1961
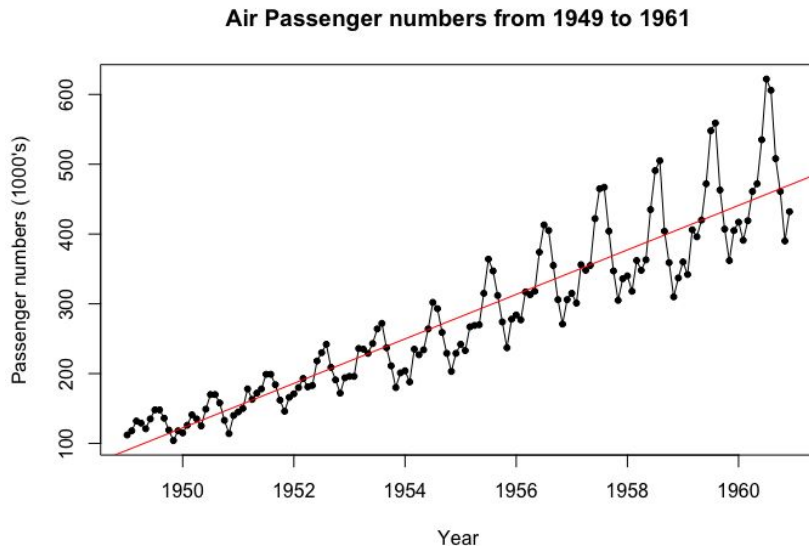
Since "Air Passenger" is a time series data set, Base Graphics automatically does a line graph
- Use abline function to add a line that best fits

```
abline(reg=lm(AirPassengers~time(AirPassengers)), col = "red")
```
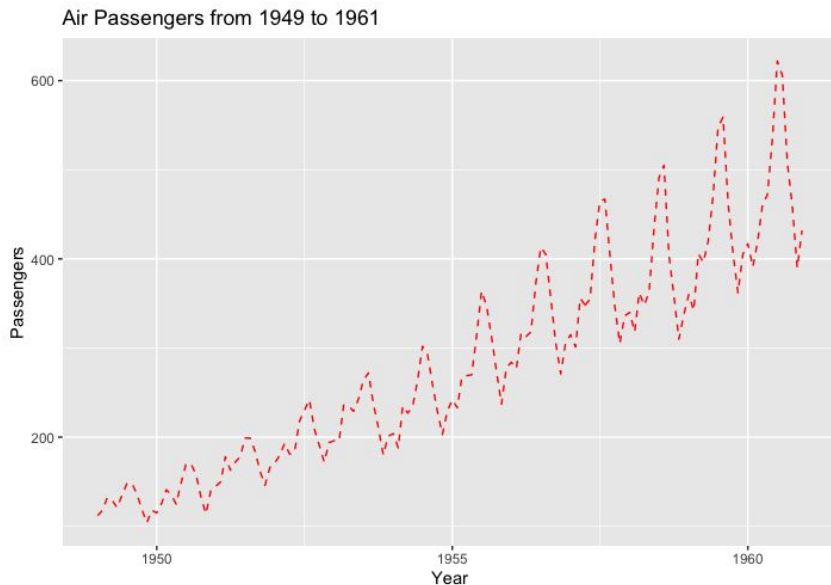
Line Plots Using Base Graphics

```
plot(aggregate(AirPassengers,FUN=mean),
    main="Air Passengers from 1949 to 1961",
    xlab="Year", ylab = "Passenger numbers (1000's)",
    col = "red")
```

Air Passengers from 1949 to 1961

Line Plot Using ggplot2

```
autoplot(AirPassengers,
         main = "Air Passengers from 1949 to 1961",
         xlab = "Year",
         ylab = "Passengers",
         ts.colour = 'red', ts.linetype = 'dashed')
```
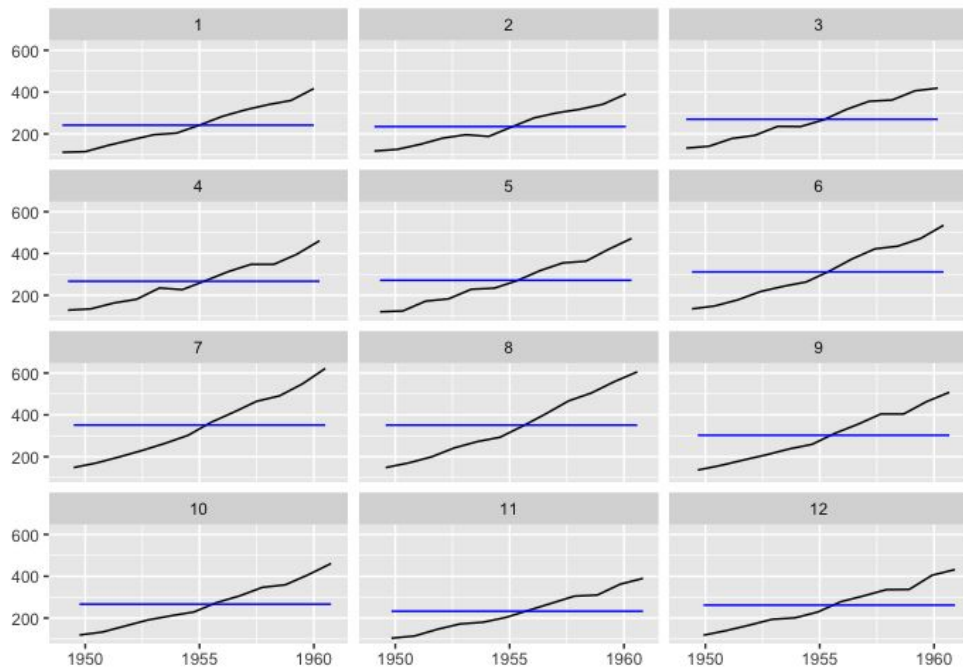
Use an extension of ggplot2 known as ggfortify
- Allows you to easily plot time series datasets

Air Passengers from 1949 to 1961

Line Plot Using ggplot2

ggfreqplot(AirPassengers)
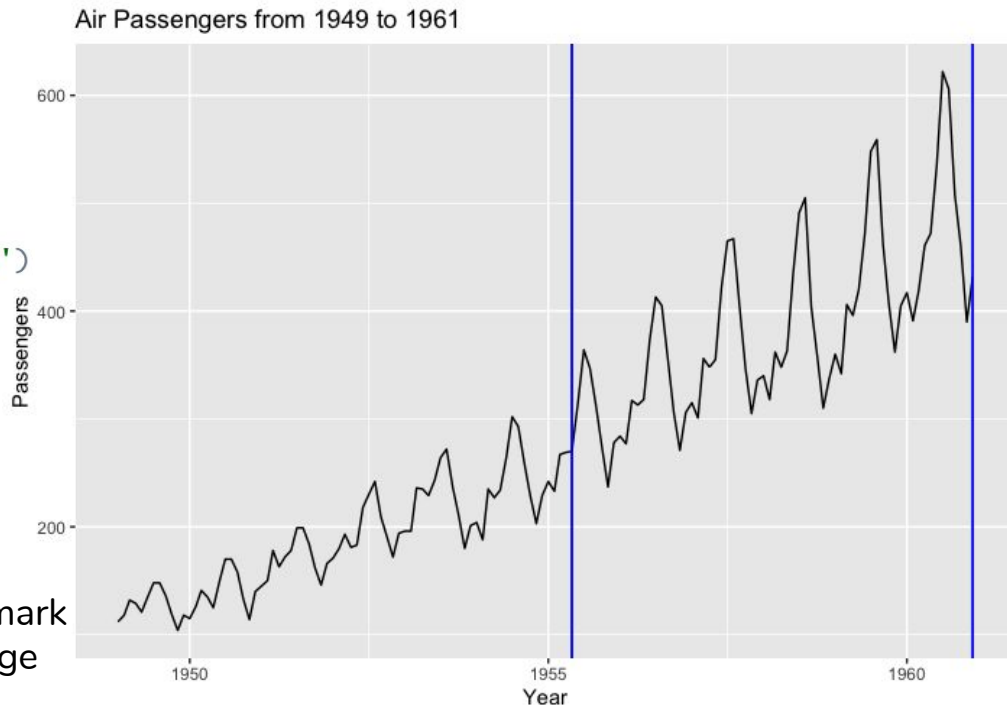
Line Plot Using ggplot2

```
autoplot(cpt.meanvar(AirPassengers),
        main = "Air Passengers from 1949 to 1961",
        xlab = "Year",
        ylab = "Passengers",
        cpt.colour = 'blue', cpt.linetype = 'solid')
```

Using another package, changepoint, ggplot2 will mark the points of greatest change

Air Passengers from 1949 to 1961
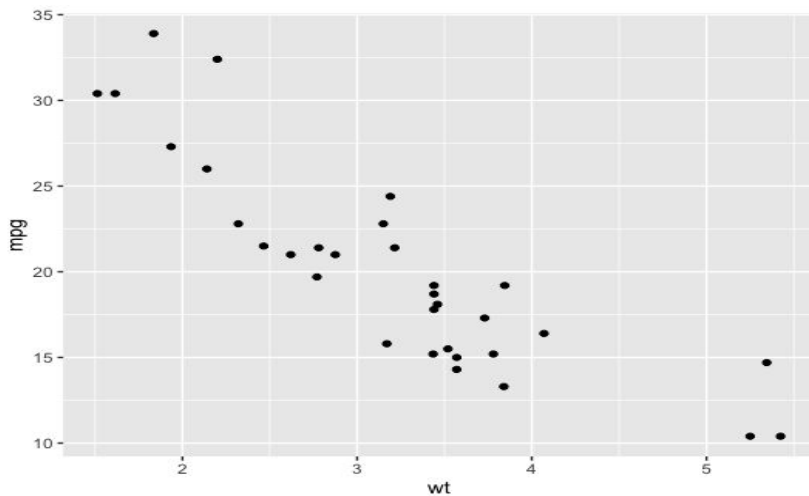
Scatter Plot using Basic Graphics

```
plot(mtcars$wt,mtcars$mpg, main= "Weight vs. Miles per  gallon",
     xlab= "weight",ylab="Miles per gallon")
abline(lm(mpg ~wt, data=mtcars), col= 'red')
```

Weight vs. Miles per gallon
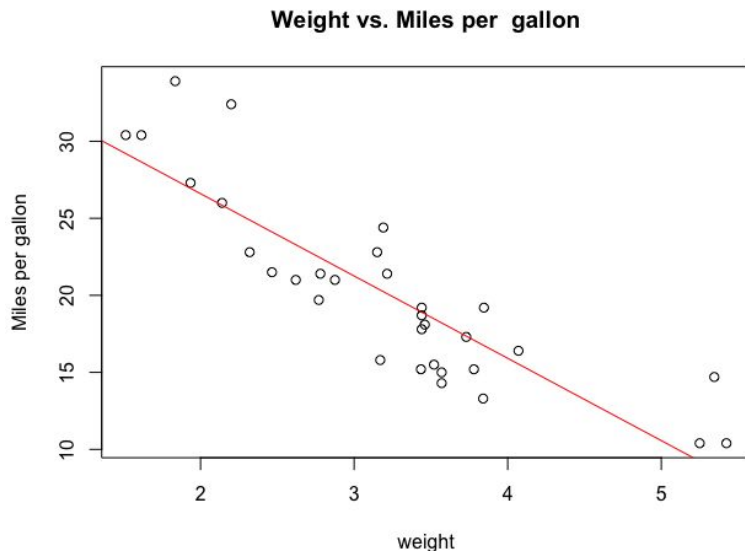
Scatter Plot using ggplot2

```
library(ggplot2)
data("mtcars")
ggplot(mtcars, aes(x=wt, y=mpg)) + geom_point()
```


This is a simple scatter plot of the data "mtcars".

Scatter plot using ggplot2

```r
library(ggplot2)
data("mtcars")
mtcars$cyl <- as.factor(mtcars$cyl)
p <-ggplot(mtcars, aes(x=wt, y=mpg, color=cyl, shape=cyl)) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
p+labs(title = "Relation between Weight and Miles per gallon", x= "Weight", y= "Miles per  gallon")
```


Relation between Weight and Miles per gallon
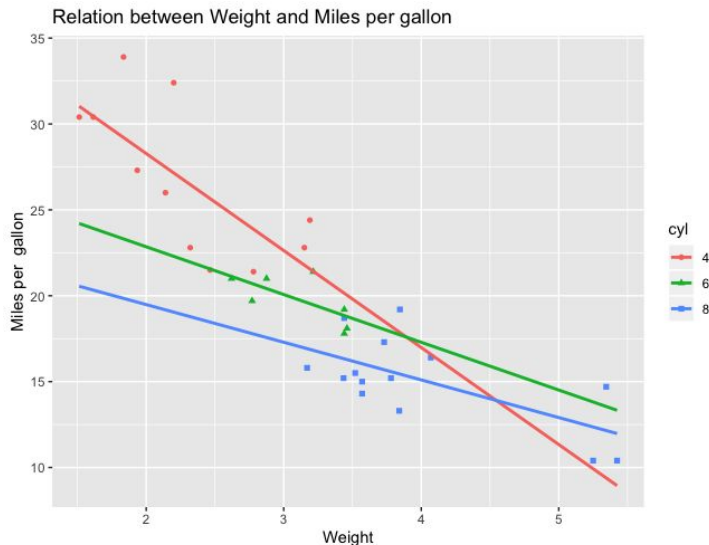
This is a more complex scatter plot that adds different color shapes to represent points, regression lines without confidence intervals, and add a title plus x and y labels.

Citations

- https://chartio.com/images/tutorials/scatter-plot/Scatter-Plot-Weight-and-Height-Scatter-Plot-Trendline.png
- https://d321jvp1es5c6w.cloudfront.net/sites/default/files/imce-user-gen/weight.png
- http://www.sthda.com/english/wiki/ggplot2-scatter-plots-quick-start-guide-r-software-and-data-visualization