Boxplots and Variations

- Real name *box-and-whisker plots*.
- Draw a box from the lower quartile to the upper quartile.
- Extend a whisker from the ends of the box to the furthest observation which is no more than 1.5 times inter-quartile range from the box.
- Mark any observations beyond this as “outliers.”

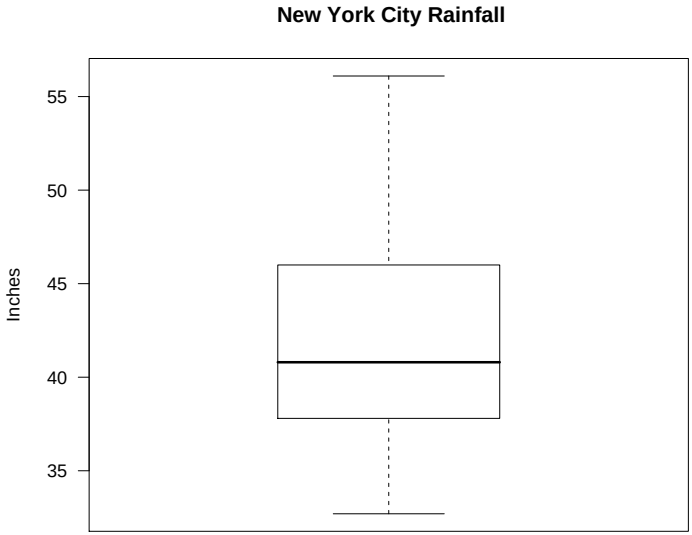
Producing Boxplots With R

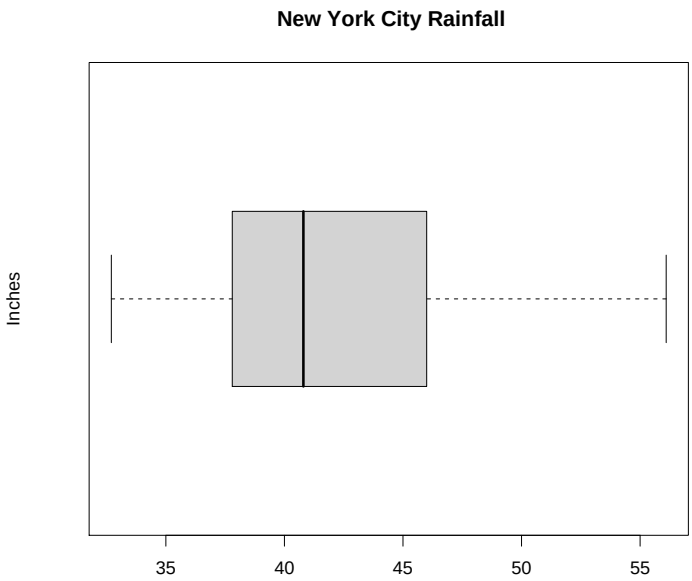
A single boxplot, vertically aligned.

```
> boxplot(rain.nyc,
          main = "New York City Rainfall",
          ylab = "Inches")
```

The basic call can be heavily customized.

```
> boxplot(rain.nyc,
          col = "lightgray",
          horizontal = TRUE,
          main = "New York City Rainfall",
          xlab = "Inches")
```


New York City Rainfall
Inches
55
50
45
40
35


New York City Rainfall
Inches
35
40
45
50
55

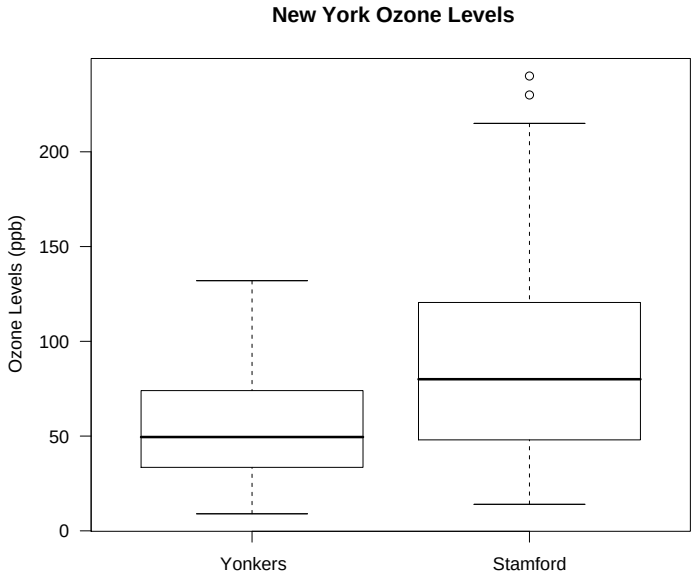
Comparing Samples with Boxplots

- It is possible to compare two or more samples with boxplots.
- By producing the plots on the same scale we are able to make direct comparisons of:
  - medians
  - quartiles
  - inter-quartile ranges
- The comparison of medians using boxplots can be regarded as the graphical equivalent of two-sample t-tests and one-way analysis of variance.

The New York Ozone Data

- The ozone levels in Yonkers and Stamford can be compared with boxplots.
- The two samples are passed to boxplot as separate arguments.
- Labels can be provided to label the two samples.

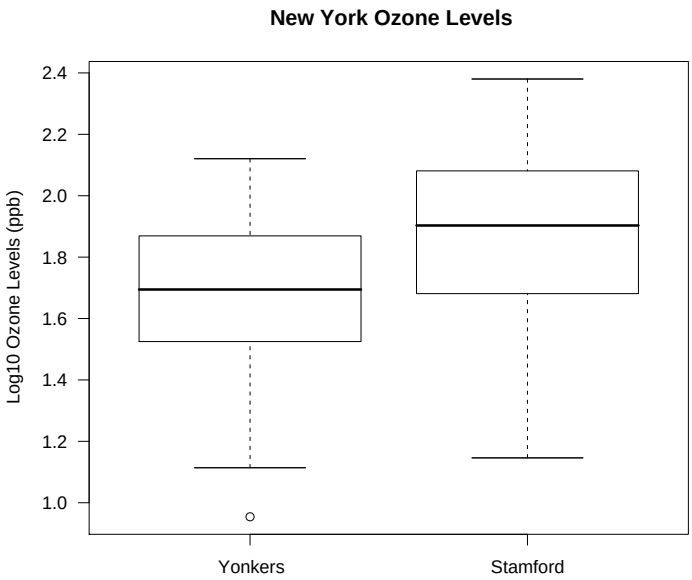
```
> boxplot(yonkers, stamford,
          names = c("Yonkers", "Stamford"),
          main = "New York Ozone Levels",
          ylab = "Ozone Levels (ppb)")
```


New York Ozone Levels
Ozone Levels (ppb)
0
50
100
150
200
Yonkers
Stamford

Transformations

- The median ozone level in Stamford is higher than that in Yonkers.
- The spread of the values in Stamford is also larger than the spread in Yonkers.
- When there is a difference in data spreads it is common to transform the values so that the speads are equal.
- This makes it possible to compare the medians in the absence of any other differences.

```
> boxplot(log10(yonkers), log10(stamford),
          names = c("Yonkers", "Stamford"),
          main = "New York Ozone Levels",
          ylab = "Log10 Ozone Levels (ppb)")
```


New York Ozone Levels
Log10 Ozone Levels (ppb)
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
Yonkers
Stamford