L02_ExploratoryVis

Anya Chaliotis

5/18/2019

Exploratory Visualizations

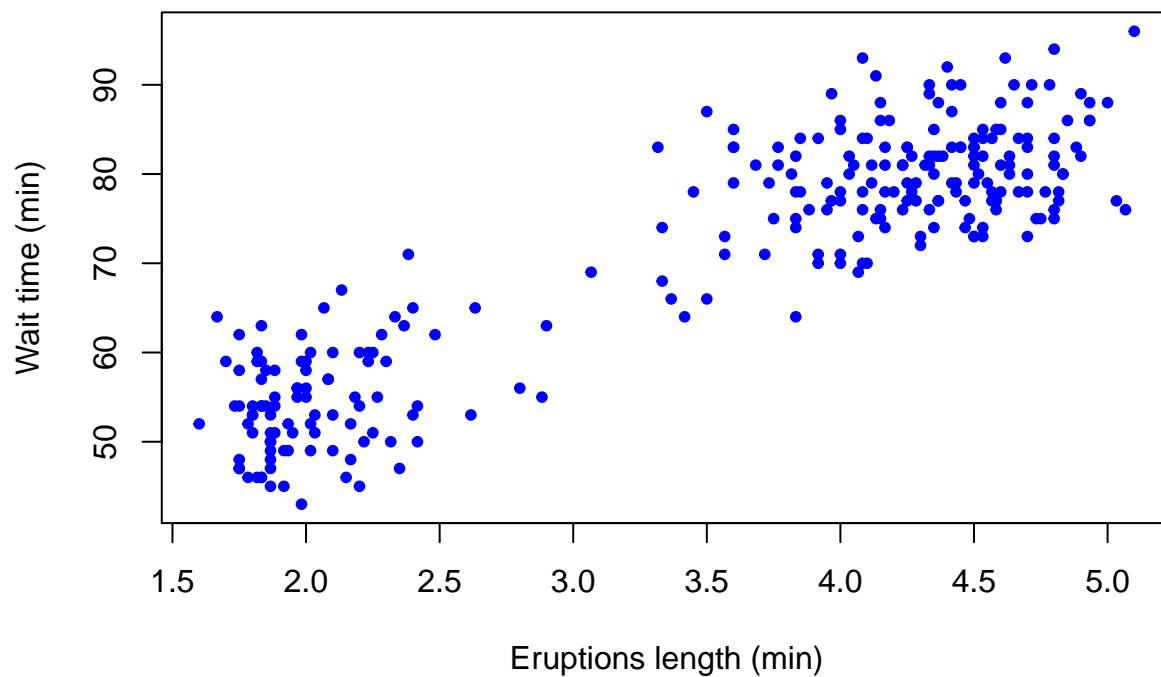Used to:
a. Vizualize shape and spread
b. Find patterns in your data
c. Locate outliers
d. summarize the main characteristics

```
library(datasets)
head(faithful)
```

```
##   eruptions waiting
## 1     3.600      79
## 2     1.800      54
## 3     3.333      74
## 4     2.283      62
## 5     4.533      85
## 6     2.883      55
```

```
plot(faithful, main='Old Faithful Eruptions', xlab='Eruptions length (min)', ylab='Wait time (min)', pc
```
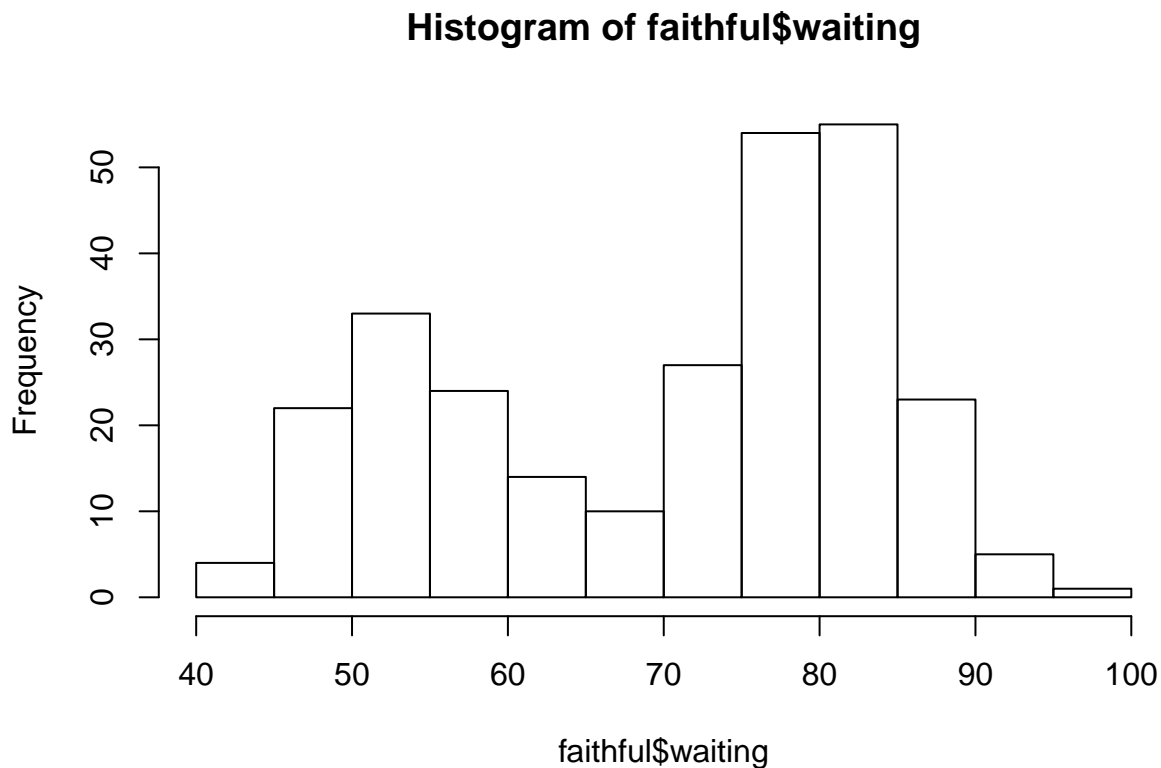
Old Faithful Eruptions

Histograms

Show the underlying shape of continuous data. Allows to inspect data for outliers and overall spread.

```
hist(faithful$waiting)
```

Histogram of faithful$waiting


```
# Breaks (aka bins)
hist(faithful$waiting, plot=FALSE)$breaks
```

```
##  [1]  40  45  50  55  60  65  70  75  80  85  90  95 100
```
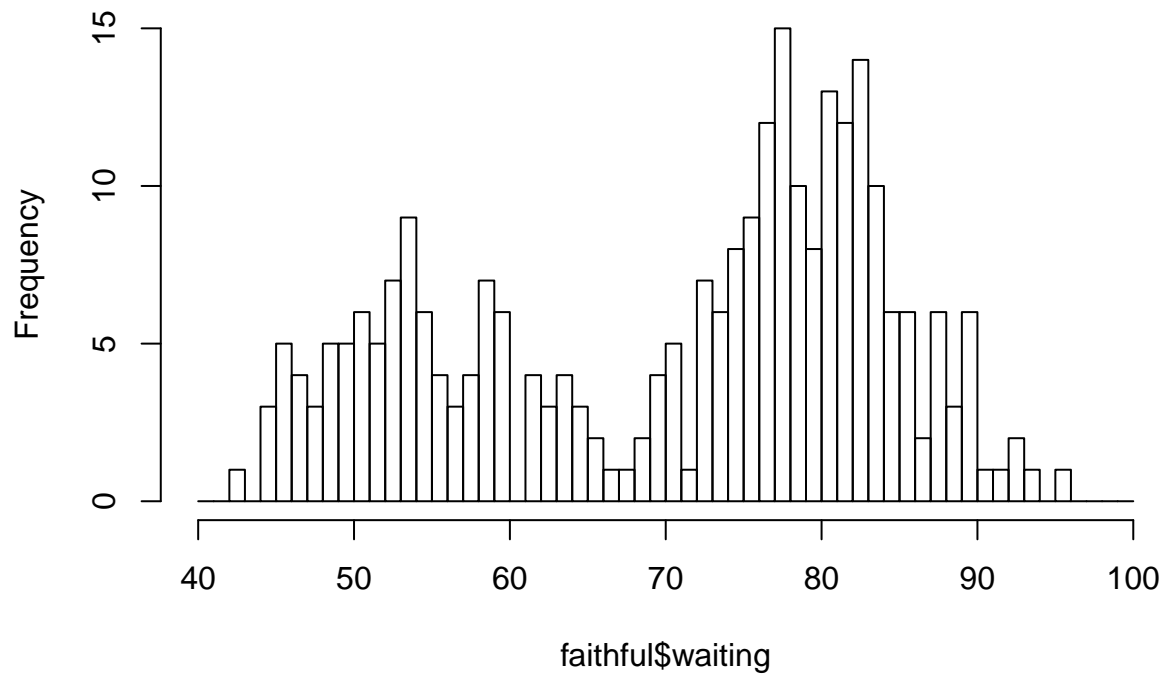
```
# how many counts in each bin
hist(faithful$waiting, plot=FALSE)$counts
```

```
##  [1]  4 22 33 24 14 10 27 54 55 23  5  1
```

```
#change size of bins
hist(faithful$waiting, main='Histogram', breaks = seq(from=40,to=100,by=1))
```

Histogram

Quantile-Quantile Plots

Quantiles of the sample are plotted against quantiles of a proposed theoretical distribution. If the points fall on a straight line, this indicates that the quantiles of the sample data are consistent with the theoretical distribution quantiles.

```r
x <- rnorm(300)
qqnorm(x, pch=16, cex=.5, col='blue', main='Q-Q Plot of Normal')
qqline(x)
```
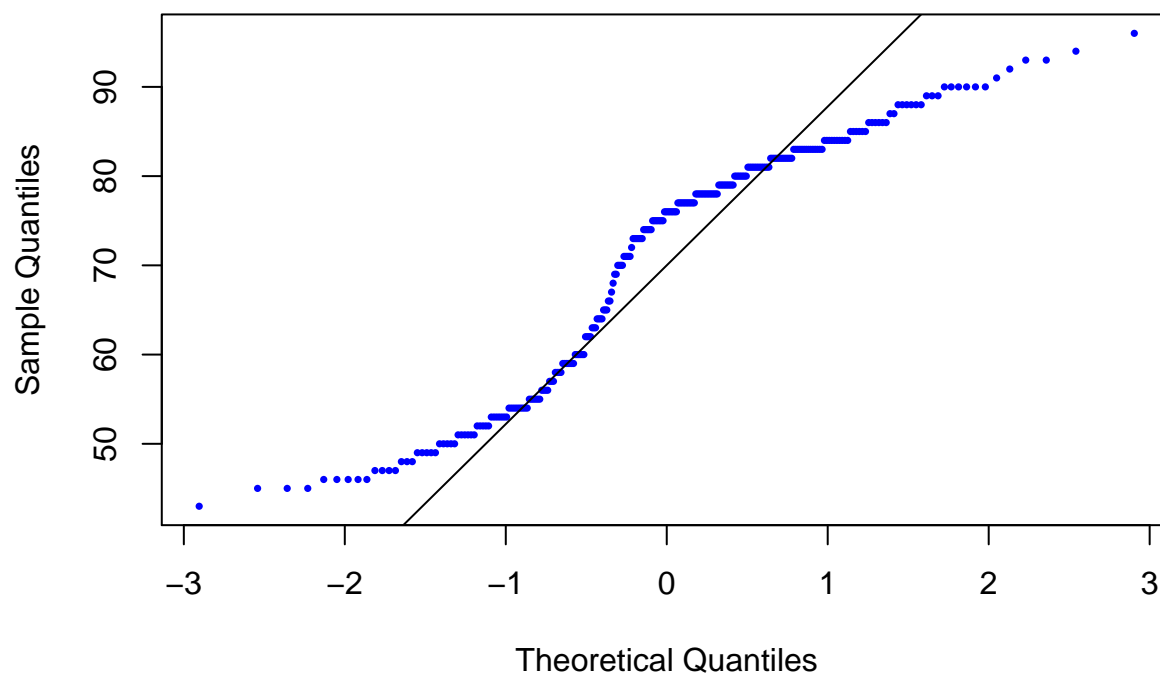
Q–Q Plot of Normal
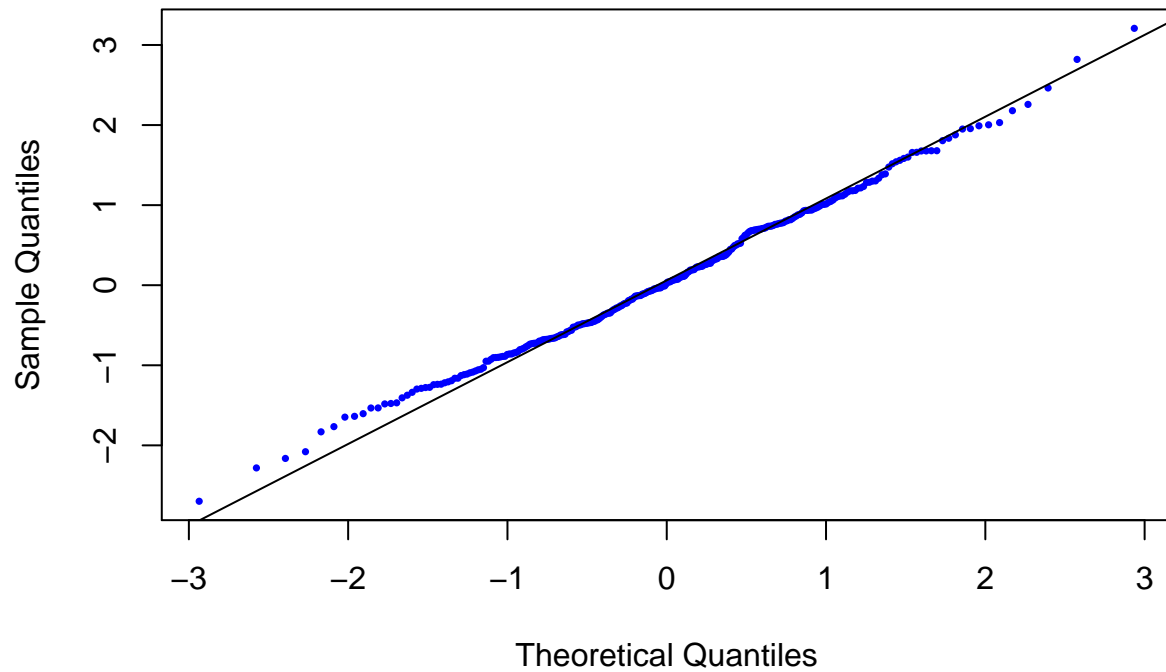
Q-Q Plot of Faithful

Faithful dataset doesn't appear to be normally distributed. It falls pretty far away from the normal distribution line.

```
qqnorm(faithful$waiting, pch=16, cex=.5,col='blue', main='Q-Q Plot for Waiting Time')
qqline(faithful$waiting)
```
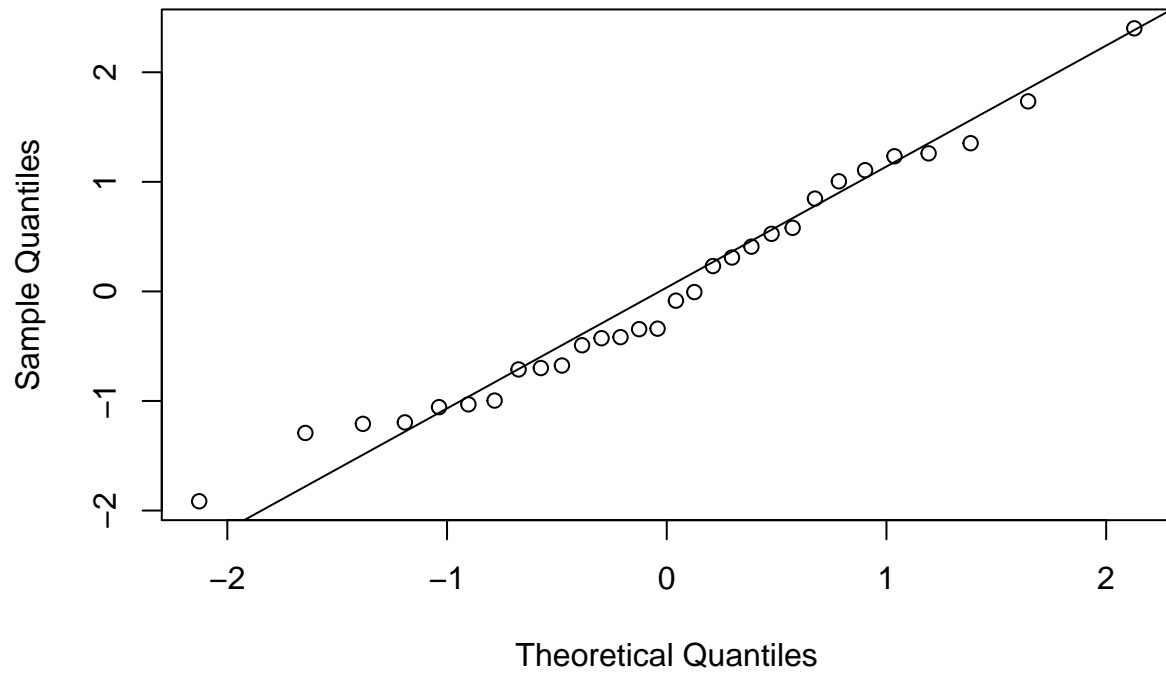
Q–Q Plot for Waiting Time

Variability in Q-Q Plots

Smaller sample size of data that we know is normally distributed. Still some variability in the data, eventhough we're simulating it from a normal distribution.
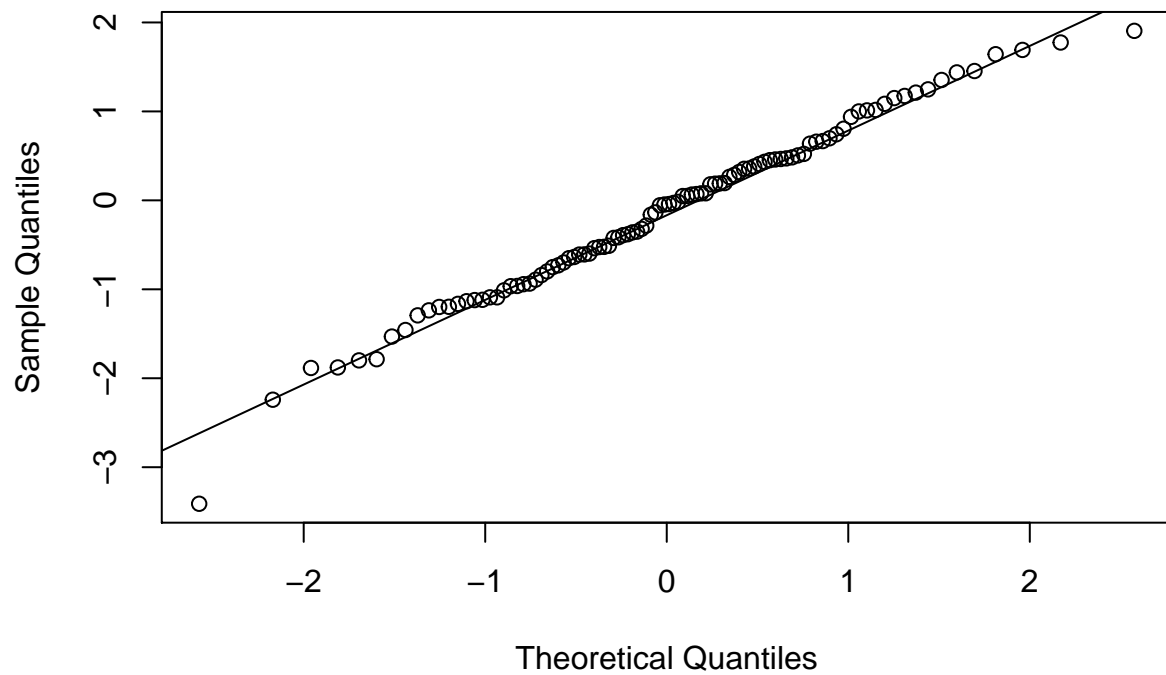
```
x = rnorm(30); qqnorm(x); qqline(x)
```

Normal Q–Q Plot

As sample size increases, get closer to a true line but still not exact.

```
x = rnorm(100); qqnorm(x); qqline(x)
```

Normal Q–Q Plot
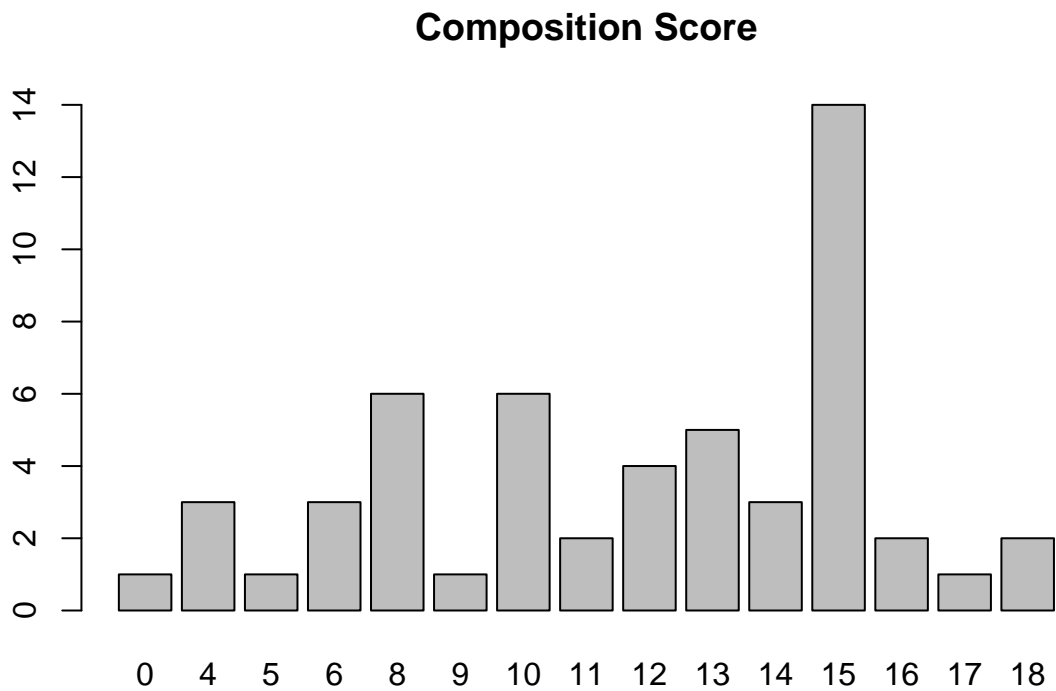
Barplots

```
library(MASS)
data("painters")
head(painters)
```

```
##               Composition Drawing Colour Expression School
## Da Udine               10       8     16          3      A
## Da Vinci               15      16      4         14      A
## Del Piombo              8      13     16          7      A
## Del Sarto              12      16      9          8      A
## Fr. Penni               0      15      8          0      A
## Guilio Romano          15      16      4         14      A
```

Use the table function to create barplots

```
barplot(table(painters$Composition), main='Composition Score')
```

Composition Score


```
barplot(table(painters$School), main='School Score')
```

School Score