Using R to compute descriptive statistics

- To assign the specified numbers $x_1, x_2, \ldots, x_n$, to a variable, say x, type
  x<-c($x_1, x_2, \ldots, x_n$)

  Example: x<-c(1,4,6,5) assigns the numbers 1, 4, 6 and 5 to the vector x.

- To generate a sequence of numbers starting from 1 and ending at n with a jump of size k try `seq(1,n,by=k)`

  Example: `seq(1,10,by=1)` generates: 1 2 3 4 5 6 7 8 9 10.

- To compute sample mean for the data x, type `mean(x)`

- To compute sample variance for the data x, type `var(x)`

- To compute sample standard deviation for the data x, type `sd(x)`

- To compute the first quantile for the data x, type
  `quantile(X,0.25)`

- To compute the second quantile for the data x, type
  `quantile(X,0.50)`

- To compute the third quantile for the data x, type
  `quantile(X,0.75)`

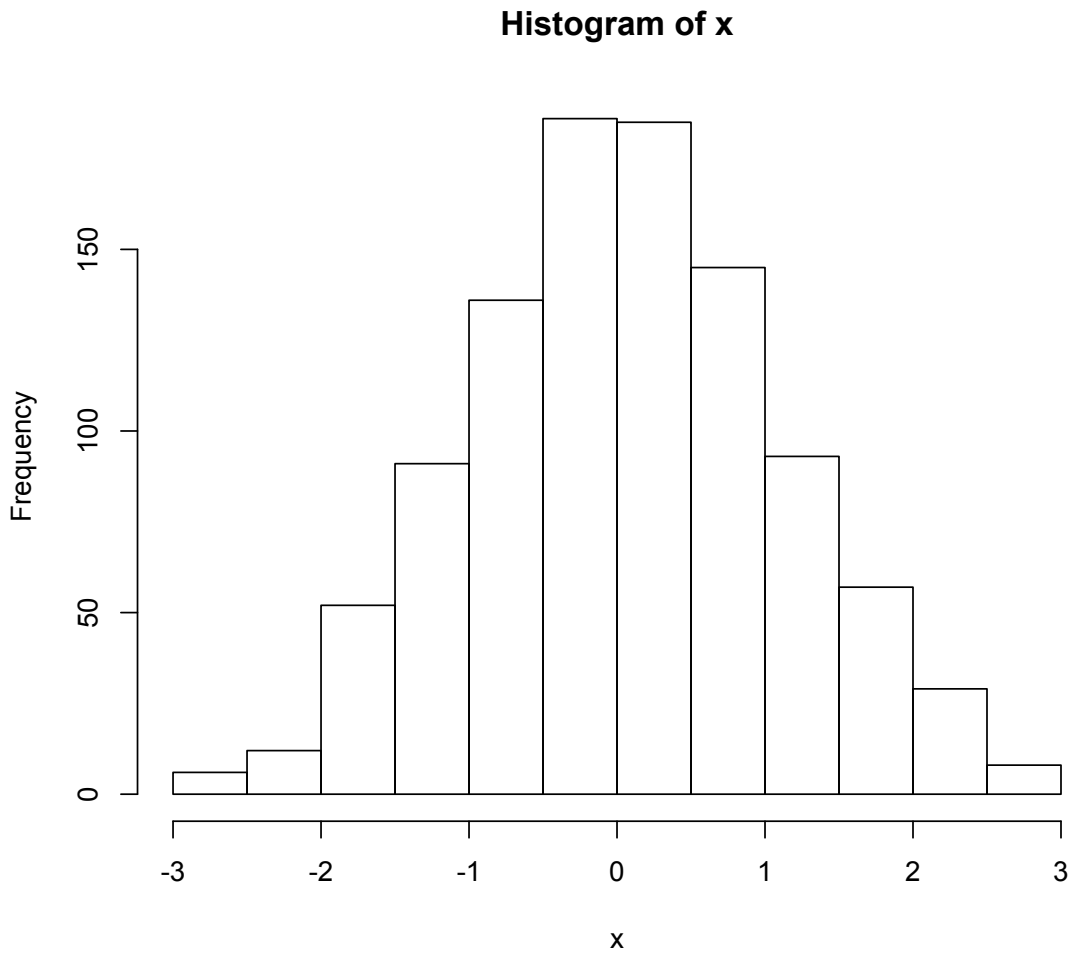
- To get summary for the data x try `summary(x)`
  Exercise: generate 100 samples from a Normal distribution with mean 2 and variance 5. Then compute sample mean, variance, quantiles and standard deviation. Observe how close the sample mean is to 2! Also try summary of x.

Using R to asses Normality

- To plot histogram for data x type `hist(X)`

Example: The first line of the following generates a random sample of size 100 from a Normal distribution with mean 0 and standard deviation 1, and the second line provides a histogram similar to the one provided below

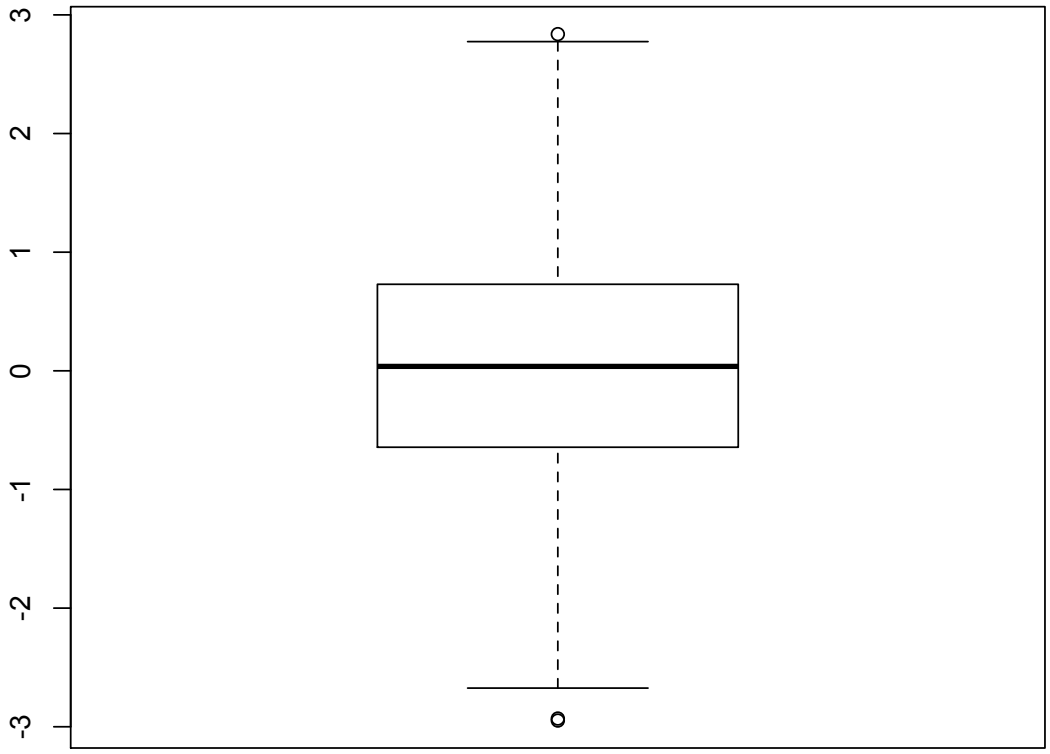
```
x<-rnorm(100,0,1)
hist(x)
```



Exercise: generate 100 samples from a standard Normal distribution. Then look at the resulting histogram. Does it confirm that the data come from a Normal distribution?

• To plot boxplot (or box and whisker plot) for the data x type
`boxplot(x)`

Example: The first line of the following generates a random sample of size 100 from a standard Normal distribution, and the second line provides a boxplot similar to the one provided below
x<-rnorm(100,2,5)
boxplot(x)



Exercise: generate 100 samples from a standard Normal distribution. Then look at the resulting boxplot. Does it confirm that the data come from a Normal distribution?

• To make a stem and leaf plot for the data x type `stem(x)`

Example: The first line of the following generates a random sample of size 100 from a standard Normal distribution, and the second line provides a stem and leaf plot similar to one provided below

```
x<-rnorm(100,0,1)
stem(x)
```

```
The decimal point is at the |

-3 | 0
-2 | 40
-1 | 96333200
-0 | 9998888877666555444333222222111
 0 | 000111112223333344455556666667777888999999
 1 | 012234446677889
 2 | 02
```

Exercise: generate 100 samples from a standard Normal distribution and plot the corresponding stem and leaf plot. Does it confirm that the data come from a Normal distribution?

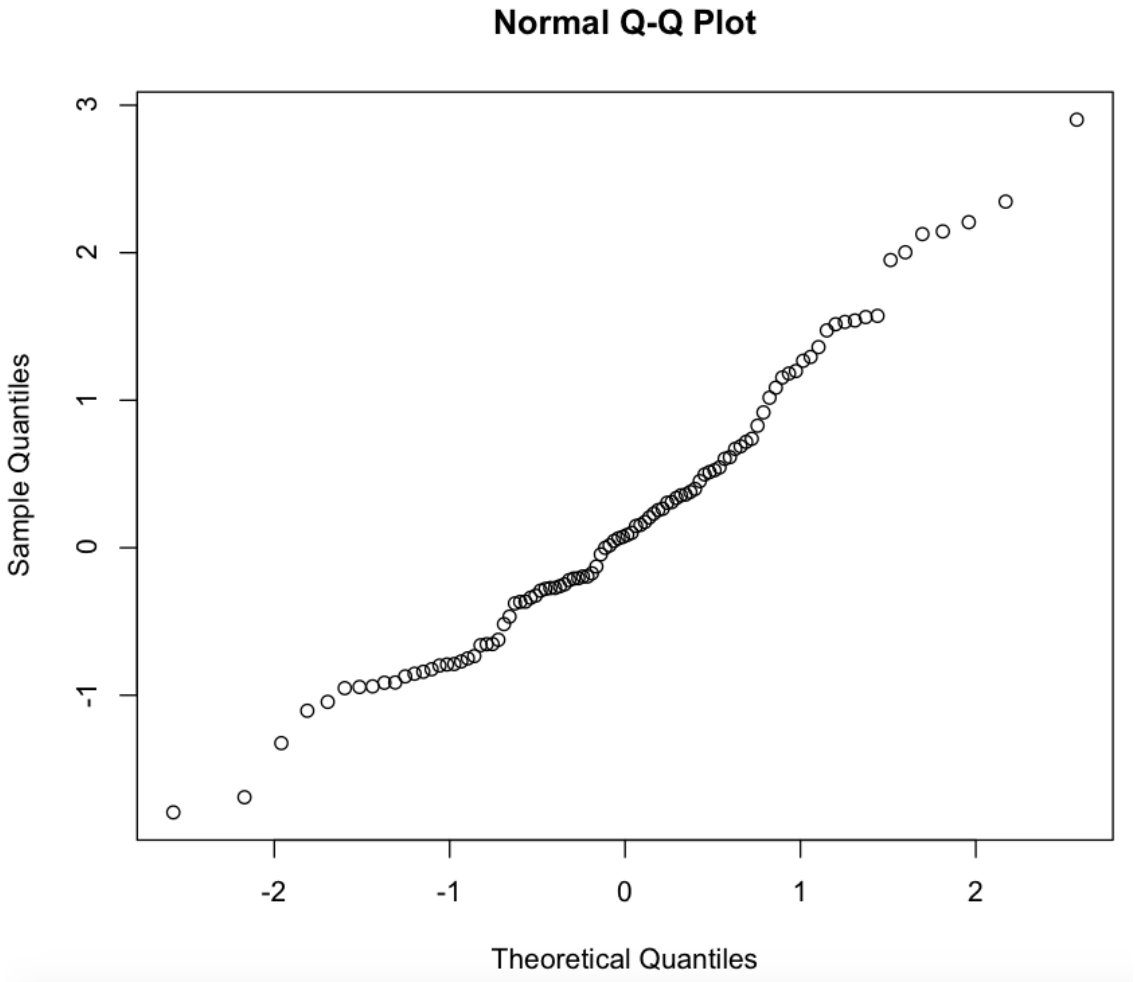
Exercise: For the following data sets (provided in the textbook), construct the stem and leaf plot.

Table 4.2: Car Battery Life

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

• To make a qqplot for the data x type `qqnorm (x)`

Example: The first line of the following generates a random sample of size 100 from a standard Normal distribution, and the second line provides a qqplot similar to the one provided below
x<-rnorm(100,0,1)
qqnorm (x)



Exercise: generate 100 samples from a standard Normal distribution and plot the corresponding qqplot. Does it confirm that the data come from a Normal distribution?