Lab 2 Solutions

1. Use the sample of following observations of variable x to find the values below.

10, 2, 15, 6, 4, 9, 12, 11, 3, 0, 12, 10, 9, 7, 11, 10, 8, 5, 10, 6

```
x=c(10, 2, 15, 6, 4, 9, 12, 11, 3, 0, 12, 10, 9, 7, 11, 10, 8, 5, 10, 6)
```

a. n (number of observations)

```
length(x)
```

```
## [1] 20
```

b. sum of all the observations in y

```
sum(x)
```

```
## [1] 160
```

c. mean

```
mean(x)
```

```
## [1] 8
```

d. median

```
median(x)
```

```
## [1] 9
```

e. mode

```
#sort(x): by default increasing
#table(x): create a contingency table that displays the frequency distribution of the variable.
#sort(-table(x))
names(sort(-table(x)))[1]
```

```
## [1] "10"
```

f. five number summary - Min, Q1, M, Q3, Max

```
min(x)
```

```
## [1] 0
```

```
max(x)
```

```
## [1] 15
```

```
median(x)
```

```
## [1] 9
```

```r
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    5.75    9.00    8.00   10.25   15.00
```

g. s^2 (variance)

```r
var(x)
```

```
## [1] 14.52632
```

Note that this is unbiased sample variance.

h. s (standard deviation)

```r
sd(x)
```

```
## [1] 3.81134
```

Note that this is unbiased sample standard deviation.

Read page 22 of Session 1b – Introduction to RStudio, and learn the function rnorm().

2. Create a vector y of random normal variables. Let y be of length 10, with the same mean as x, and standard deviation 1.

```r
y = rnorm(10,mean(x),1)
```

a. Calculate the covariance and correlation between x and y. Can you do that? Why or why not?

```r
#cor(x,y)
```

Error in cor(x, y) : incompatible dimensions. x and y should have the same length.

b. Now change your y so that it has the same length as x. The mean and standard deviation stay the same. Calculate the covariance and correlation between x and y again.

```r
y=rnorm(20,mean(x),1)
cov(x,y)
```

```
## [1] 0.7781274
```

```r
cor(x,y)
```

```
## [1] 0.2269831
```

c. Repeat b. several times. Did you get the same result every time? If not, why? What can you do make your result repeatable?

No, because rnorm() is a random function. We need to set a seed to get repeatable results.

d. Now pick an integer, say, 63. Run set.seed(63) before the rnorm function. Repeat the two functions for several times and check whether you get the same results from rnorm every time.

Yes.

3. In this question, always set seed to be 100.

a. Create z of random normal variables of length 20, mean 8 and SD 1. Plot the scatter plot of x and z. Add the main label, x-label and z-label.
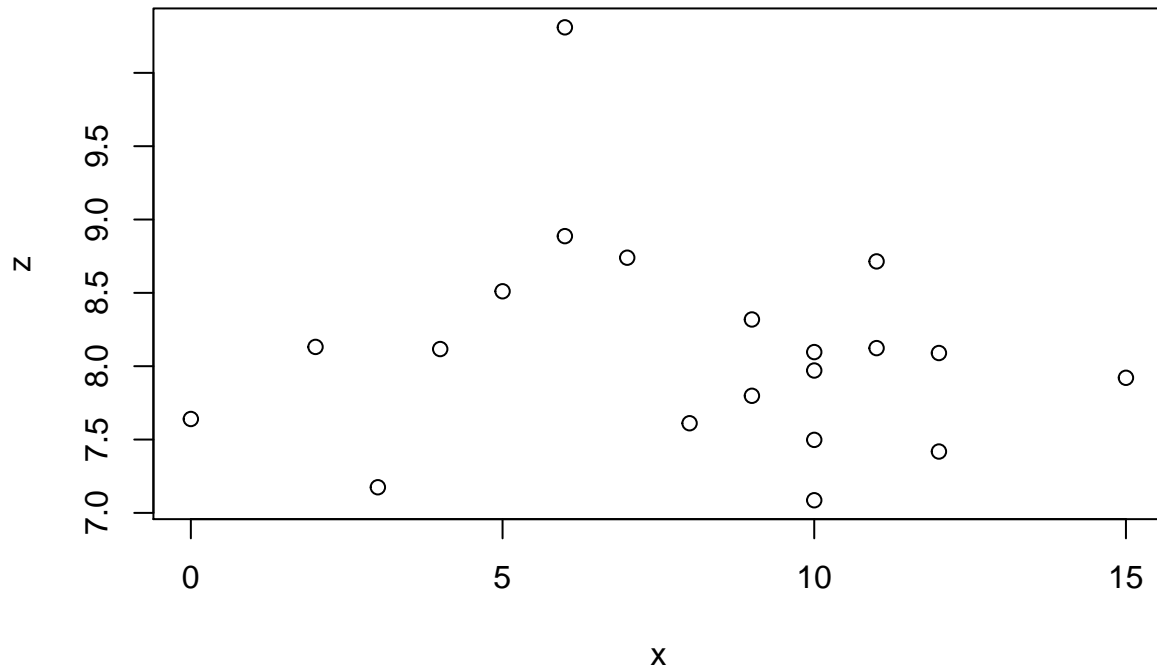
```
set.seed(100)
z = rnorm(20,8,1)
plot(z~x, main = "Scatter plot of x and z", xlab = "x", ylab= "z")
```
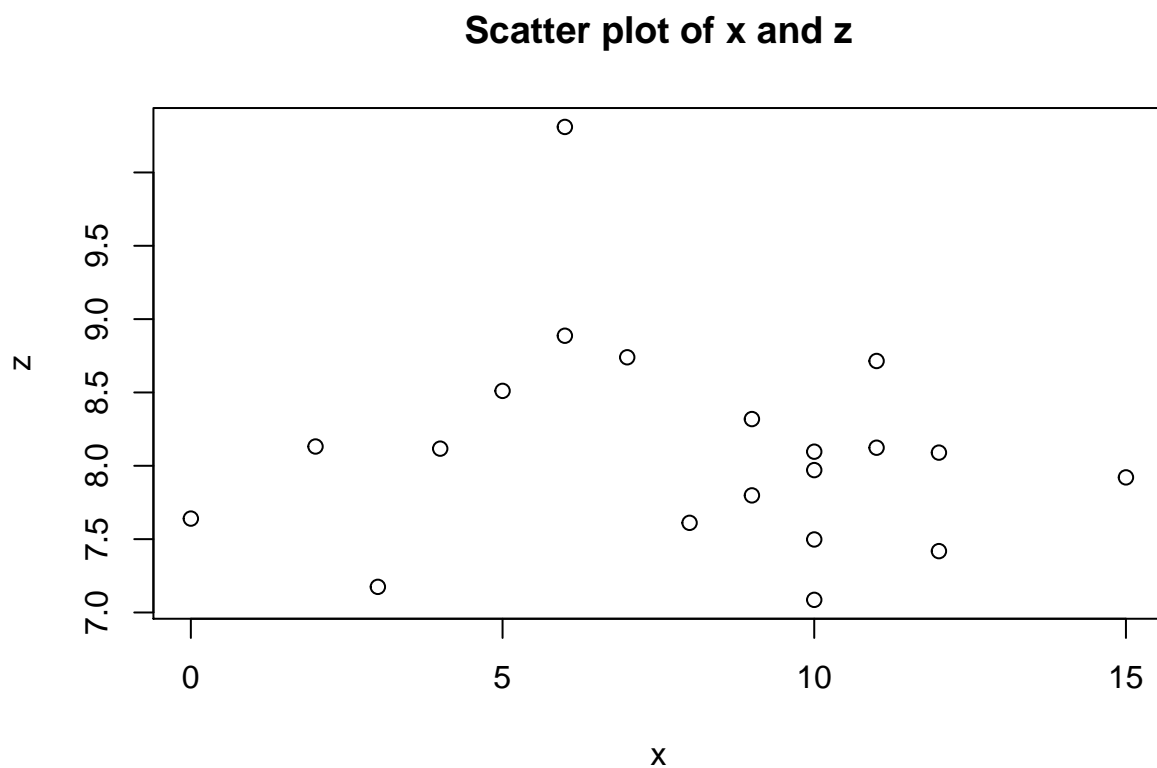

Note that the above plot is the same as

```
plot(x,z,main = "Scatter plot of x and z", xlab = "x", ylab= "z")
```

Scatter plot of x and z

b. Change the SD of z to 10 and obtain a new vector u. Plot the corresponding x and u. Draw the points onto the same plot as in 3a, and change the colour and symbol of the points.

```r
plot(x,z,main = "Scatter plot of x and z", xlab = "x", ylab= "z",ylim=c(-10,40))
#Why do we need to add ylim=c(-10,40)?
set.seed(100)
u = rnorm(20,8,10)
#points function adds points to the existing plot.
points(x,u, col = "blue", pch = "+")
```
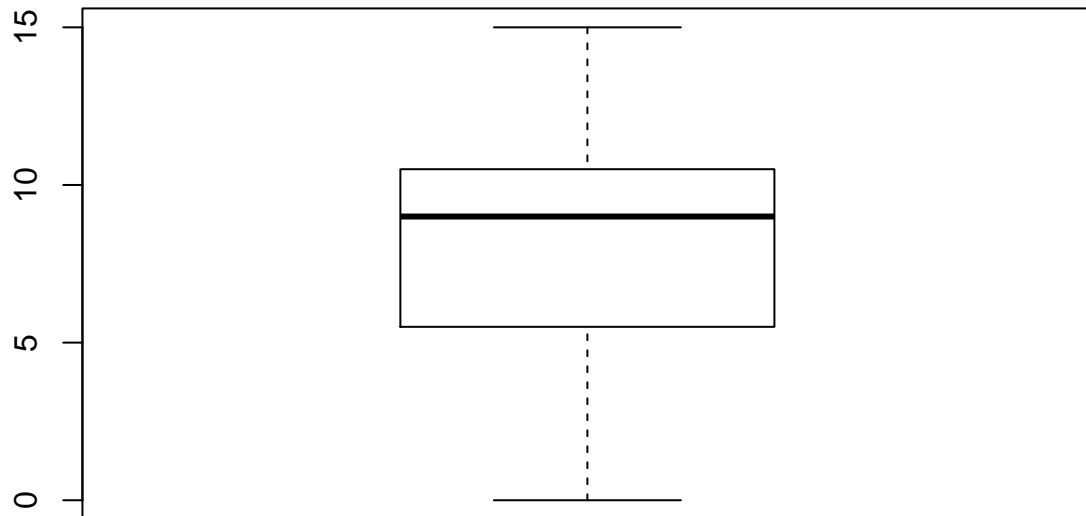
Scatter plot of x and z
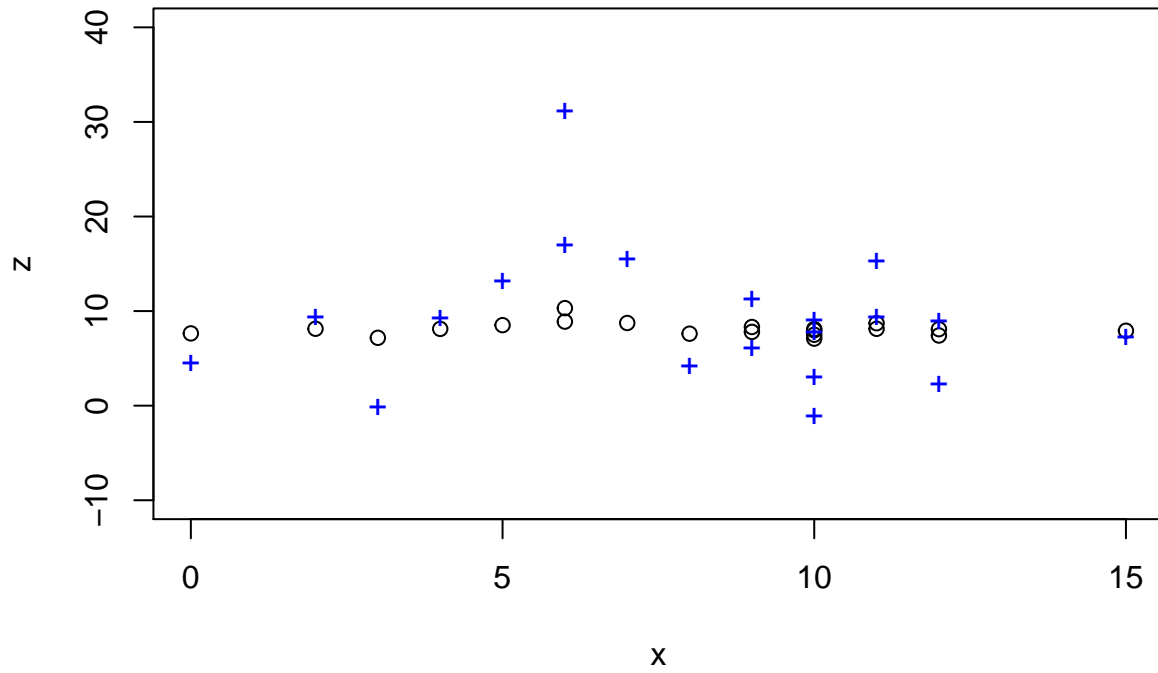
4. Explore yourself how to plot a boxplot in R.
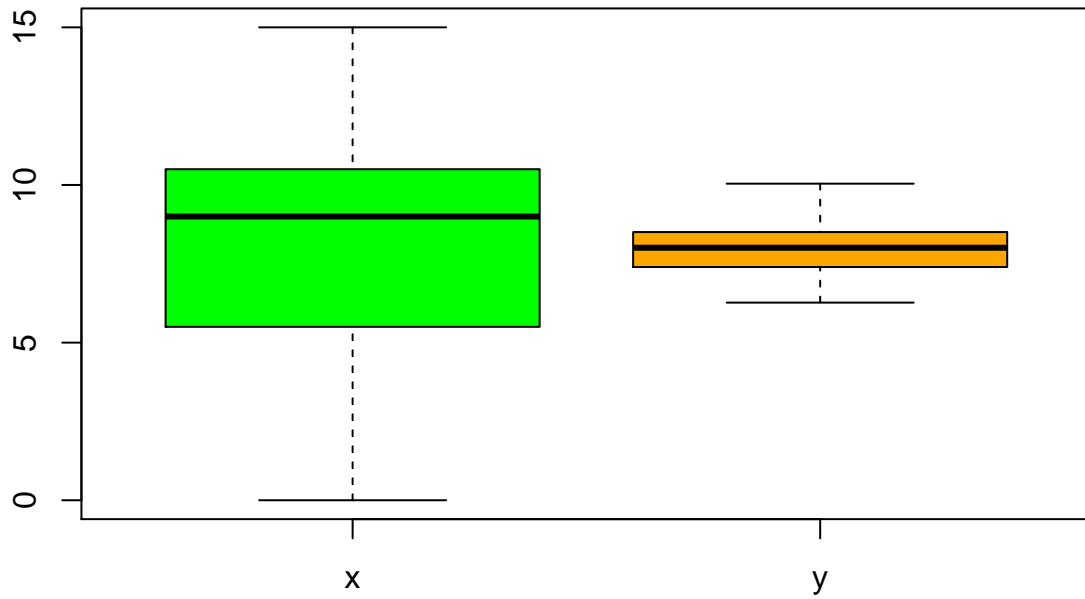
a. Plot vector x in boxplot.

```
boxplot(x)
```

b. Plot vector x and y in boxplot and display the result in one plot. See if you can add a label under each boxplot, and add some colour to each box.

```
data = data.frame(x,y)
boxplot(data,col=c("green","orange"),main="Boxplot for x and y")
```
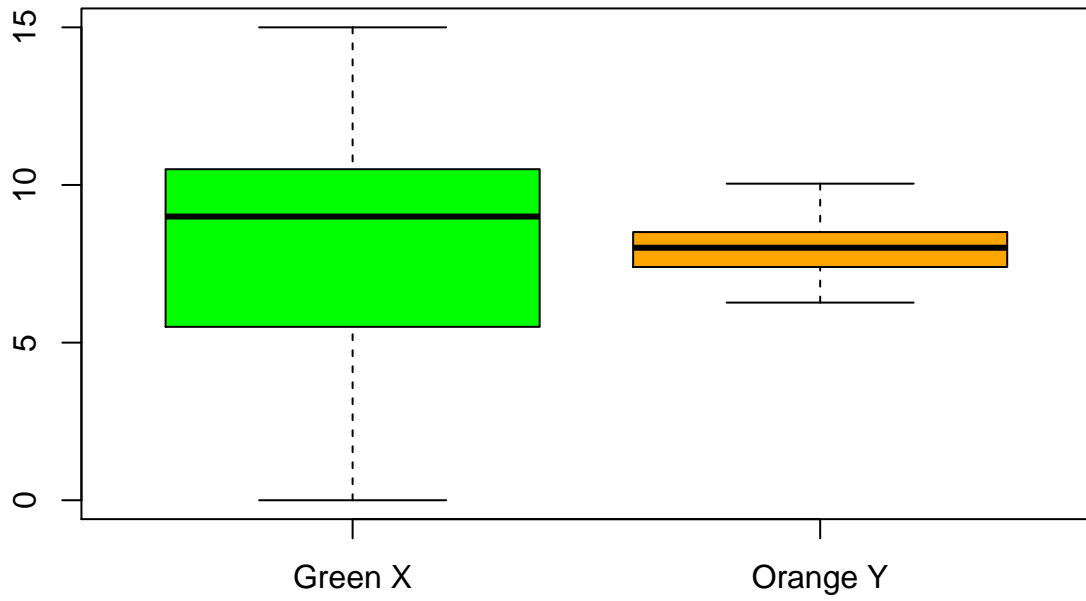
Boxplot for x and y

Or alternatively,

```r
boxplot(x,y,col=c("green","orange"), names=c("Green X", "Orange Y"), main="Boxplot for x and y")
```

Boxplot for x and y

Notice that, if it is a data frame, the boxplot will automatically show their names x and y.