Stat 291 - Recitation 9

Orçun Oltulu

24 / 12 / 2021

Basic Plotting:

Exercise 1:

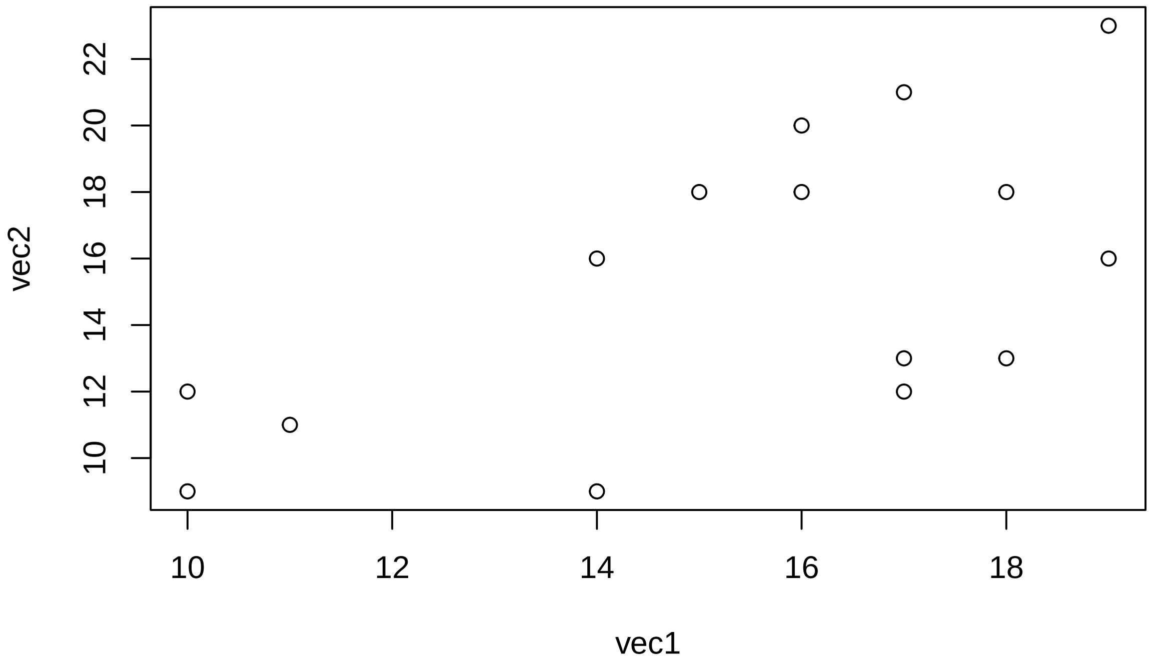
Create following vectors;

```
set.seed(291)
vec1 <- floor(runif(15, 10, 20))
vec2 <- floor(vec1 + runif(15, -5, 5))
```

Part A.

Plot vec1 vs vec2 without any specific option.

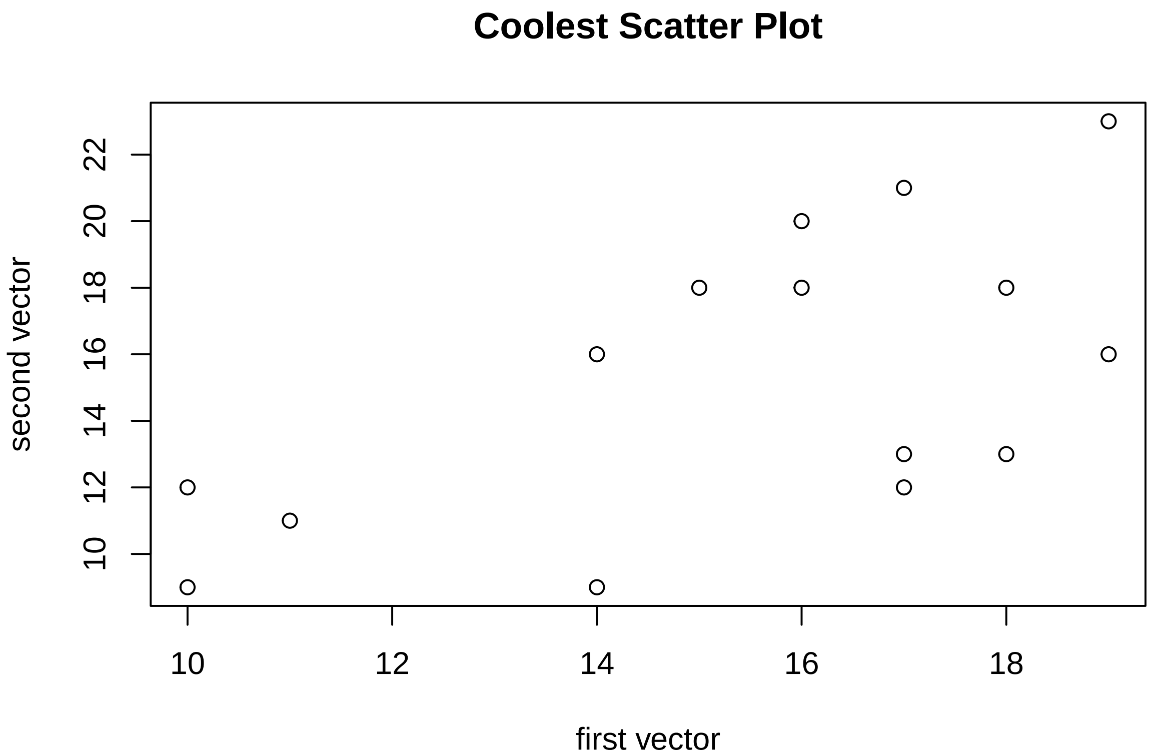
```
plot(vec1,vec2)
```



Part B.

Add a title and change axes labels.

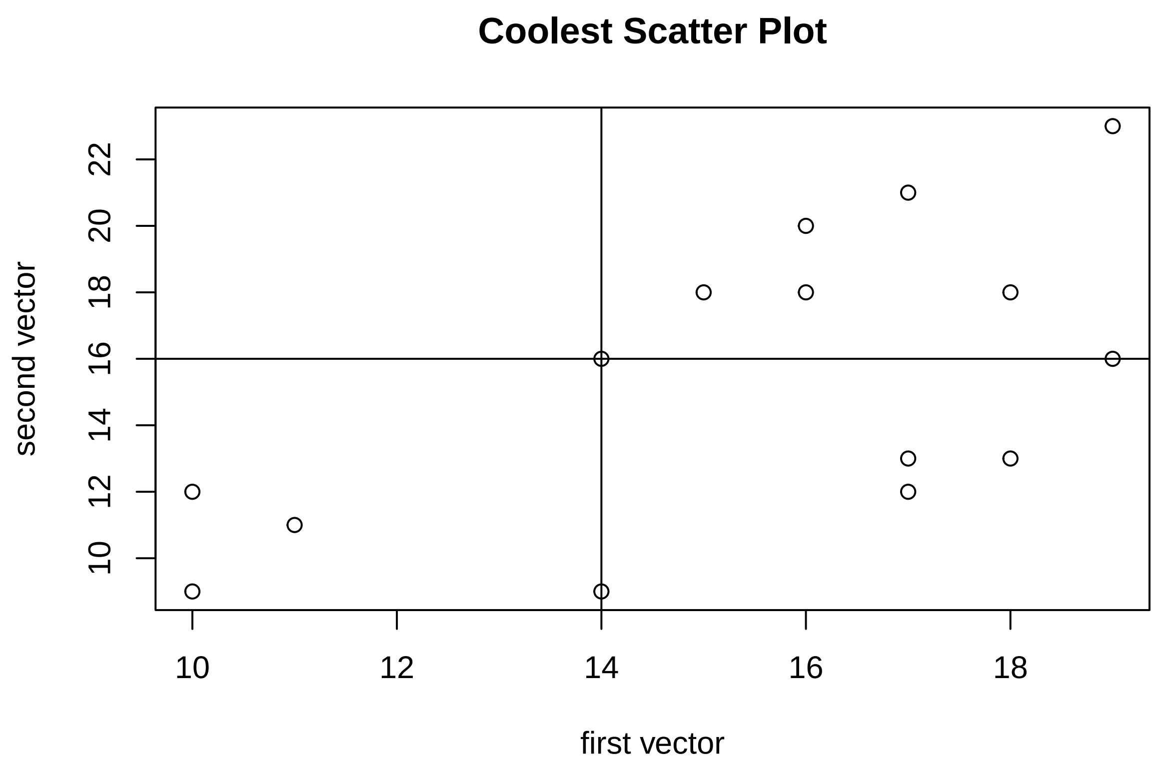
```
plot(vec1, vec2, main = "Coolest Scatter Plot",
     xlab = "first vector", ylab = "second vector")
```



Part C.

Add one horizontal line and one vertical line which intersect at (14,16).

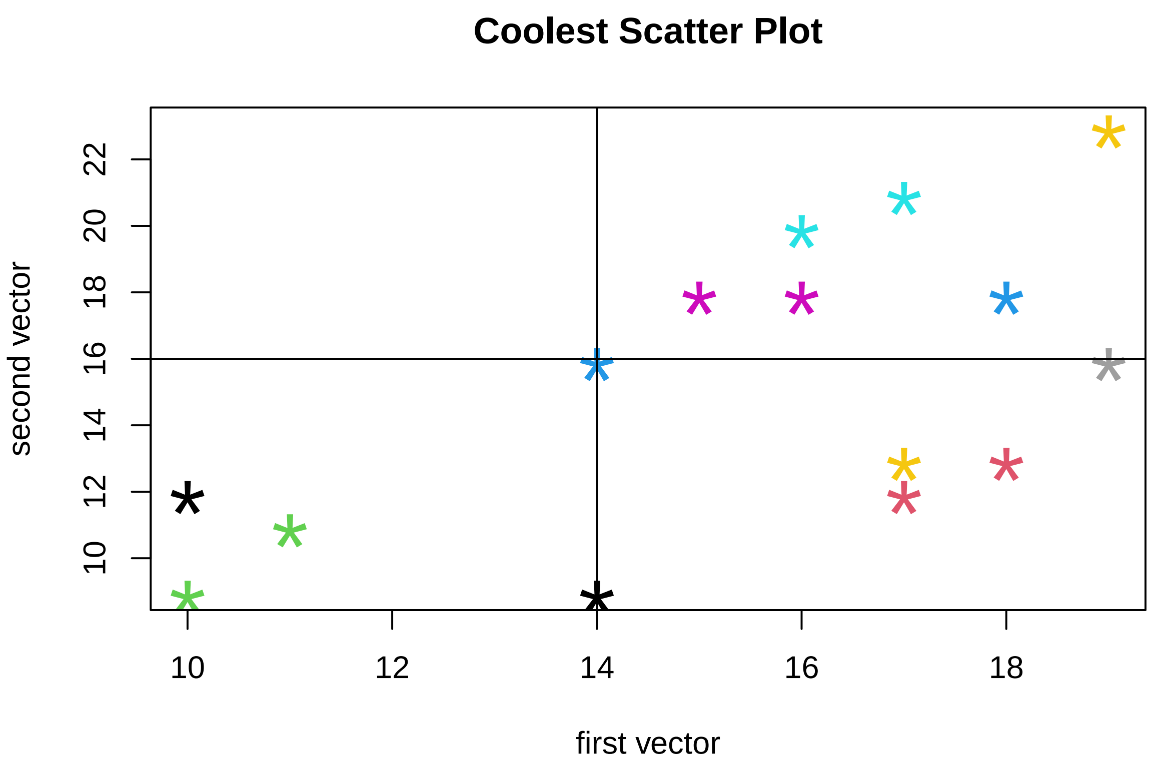
```
plot(vec1, vec2, main = "Coolest Scatter Plot",
     xlab = "first vector", ylab = "second vector")
abline(h = 16, v = 14)
```



Part D.

Now, change the point character, character expansion and color.

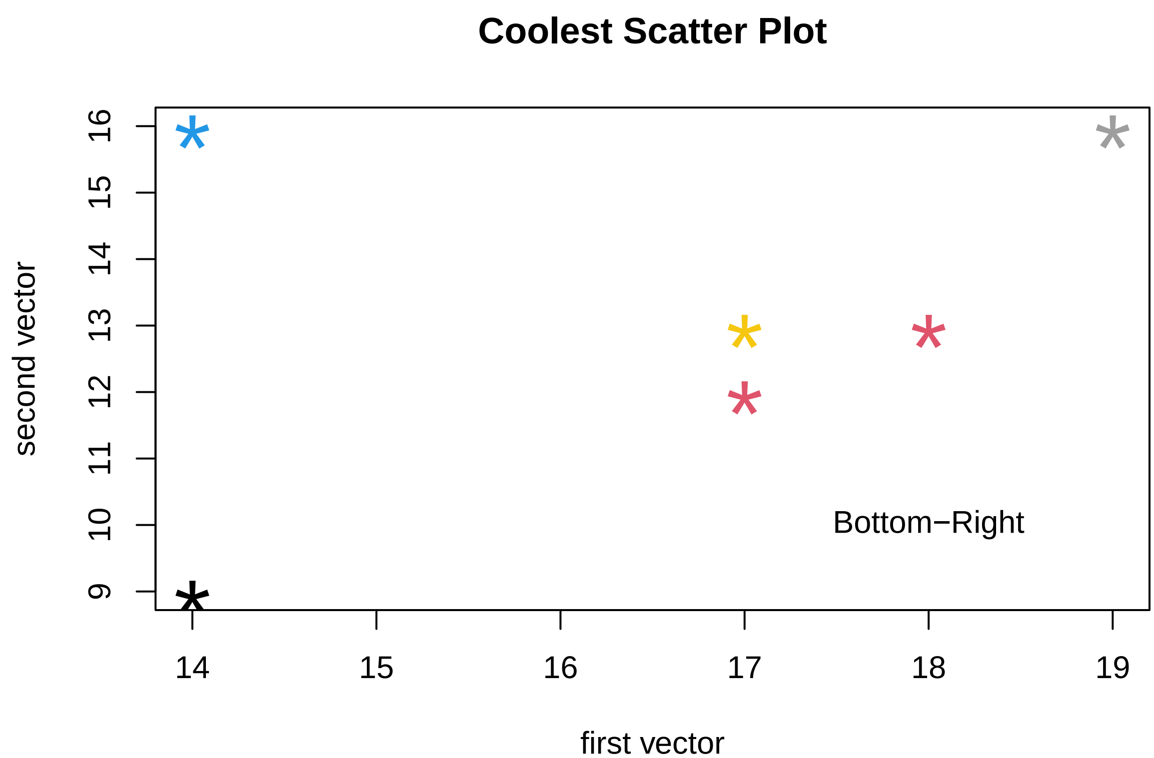
```
plot(vec1, vec2,
     main = "Coolest Scatter Plot",
     xlab = "first vector",
     ylab = "second vector",
     col = c(1:length(vec1)),
     cex = 3,
     pch = "*")
abline(h = 16, v = 14)
```



Part E.

Now, zoom in to the bottom right rectangular area i.e. only plot that one and wright "Bottom-Right" somewhere in the plot.

```
plot(vec1, vec2,
     main = "Coolest Scatter Plot",
     xlab = "first vector",
     ylab = "second vector",
     col = c(1:length(vec1)),
     cex = 3,
     pch = "*",
     xlim = c(14,max(vec1)),
     ylim = c(min(vec2),16))
text(x = 18, y = 10, labels = "Bottom-Right")
```



Exercise 2:

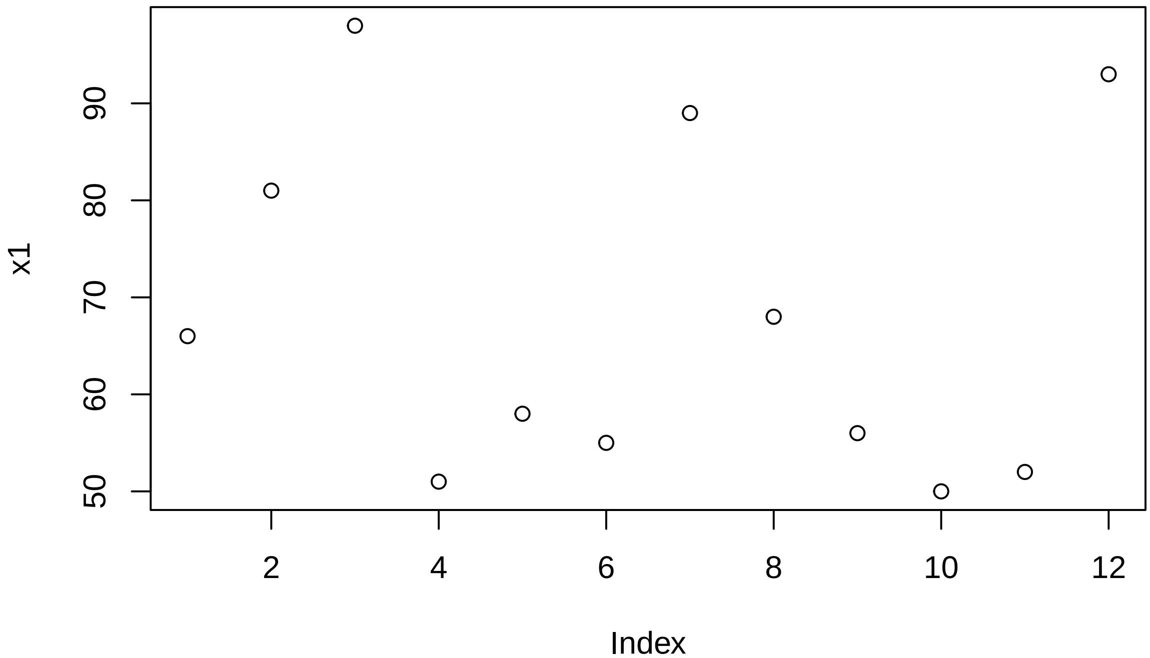
Assume X1 and X2 are prices for certain stocks. Create following vectors;

```
set.seed(291)
x1 <- sample(50:100, size = 12)
x2 <- sample(40:90, size = 12)
```

Part A.

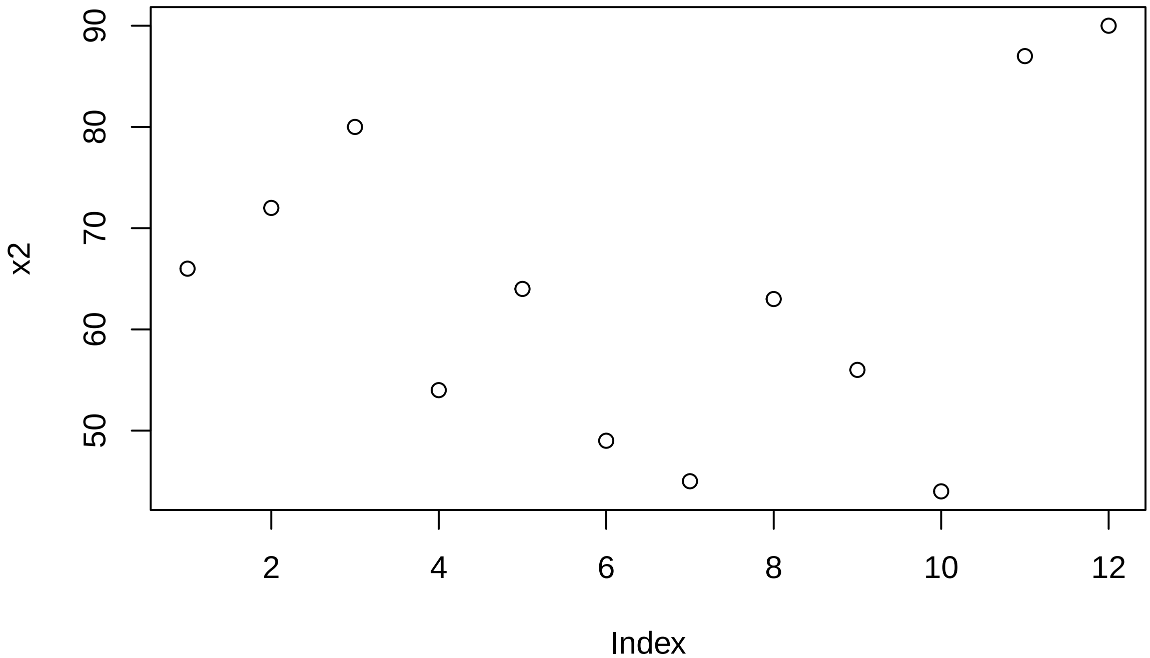
Plot x1.

```
plot(x1)
```



Plot x2.

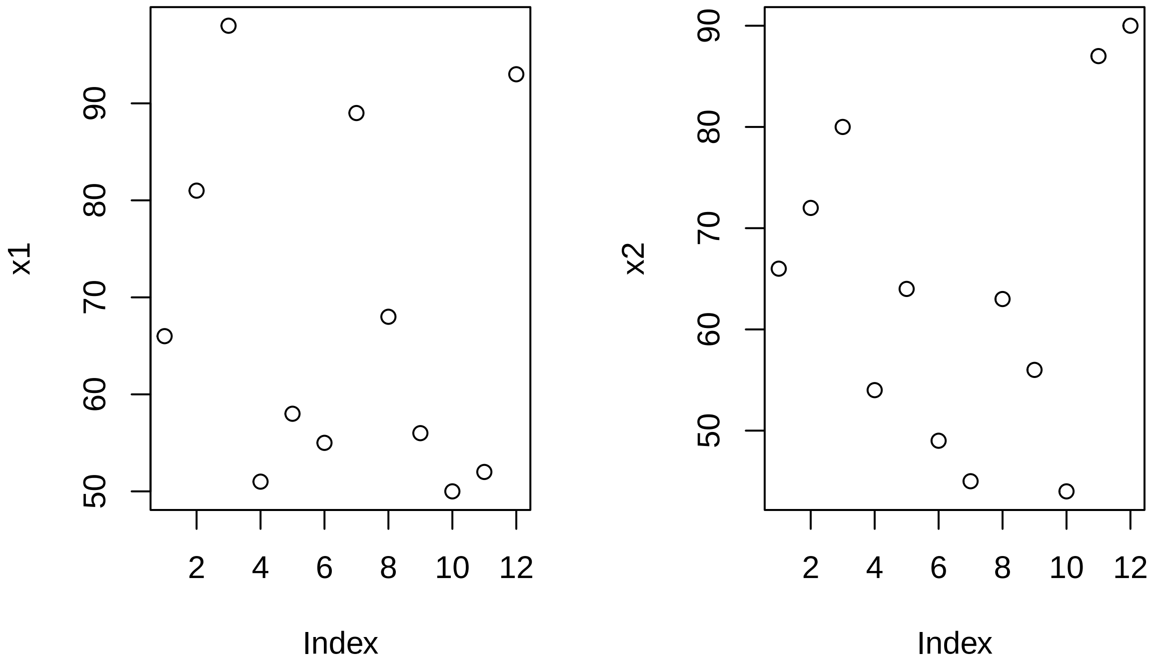
```
plot(x2)
```



Part B.

Use par() function to combine these two plots in the same window.

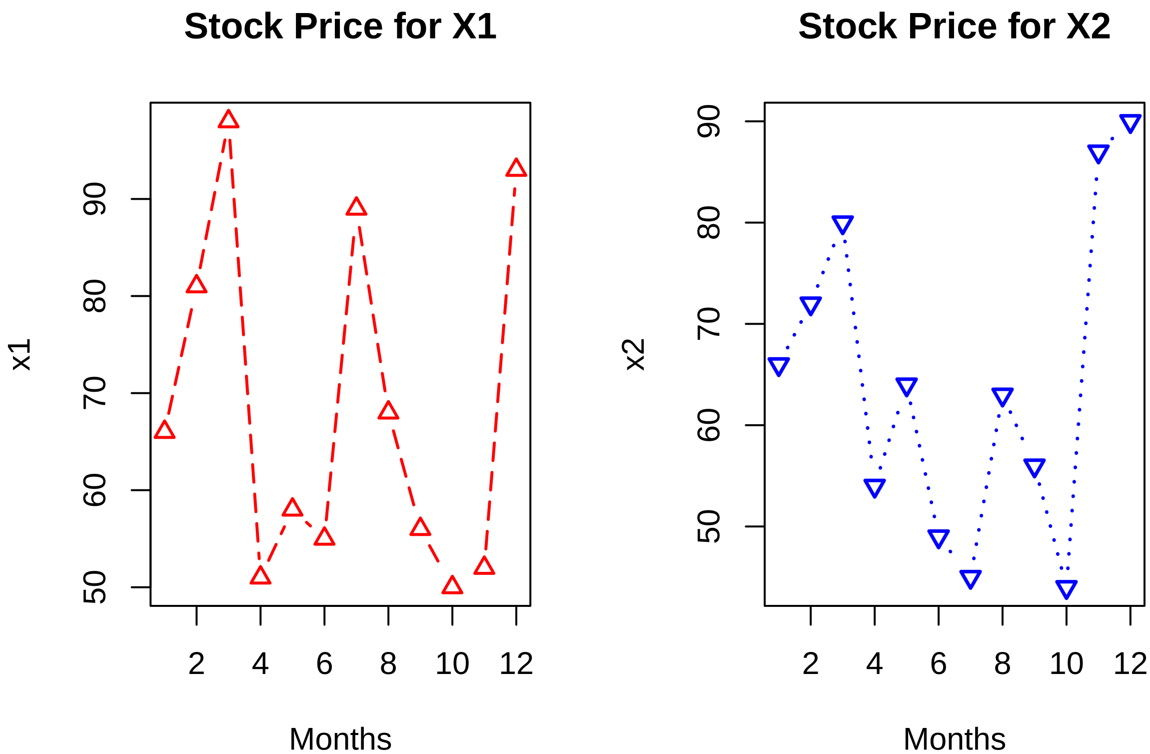
```
par(mfrow = c(1,2))
plot(x1)
plot(x2)
```



Part C.

Keep par() function, obtain line plots and add x labels and titles to each plot. Assign line types, line widths and colors for each plot.

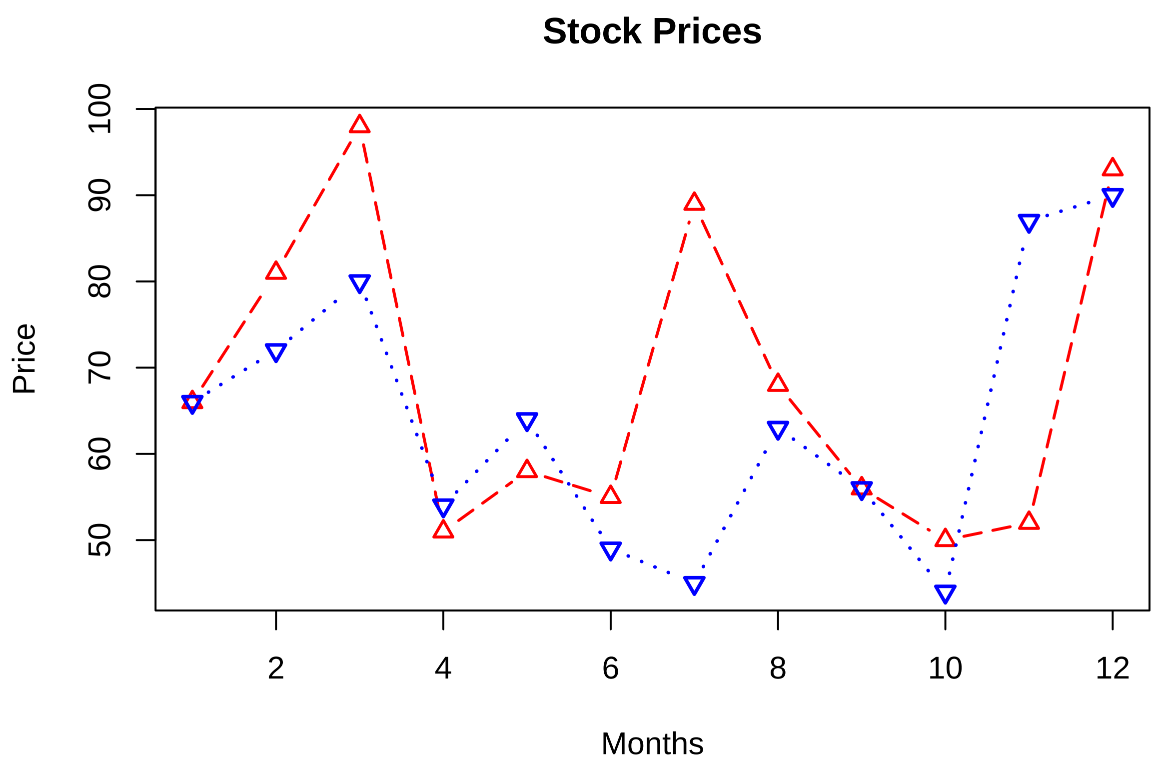
```
par(mfrow = c(1,2))
plot(x1, main = "Stock Price for X1", xlab = "Months",
     type = "b", pch = 2, lwd = 1.5, lty = 5, col = "Red")
plot(x2, main = "Stock Price for X2", xlab = "Months",
     type = "b", pch = 6, lwd = 1.8, lty = 3, col = "Blue")
```



Part D.

Now, plot X1 and X2 in a same plot.

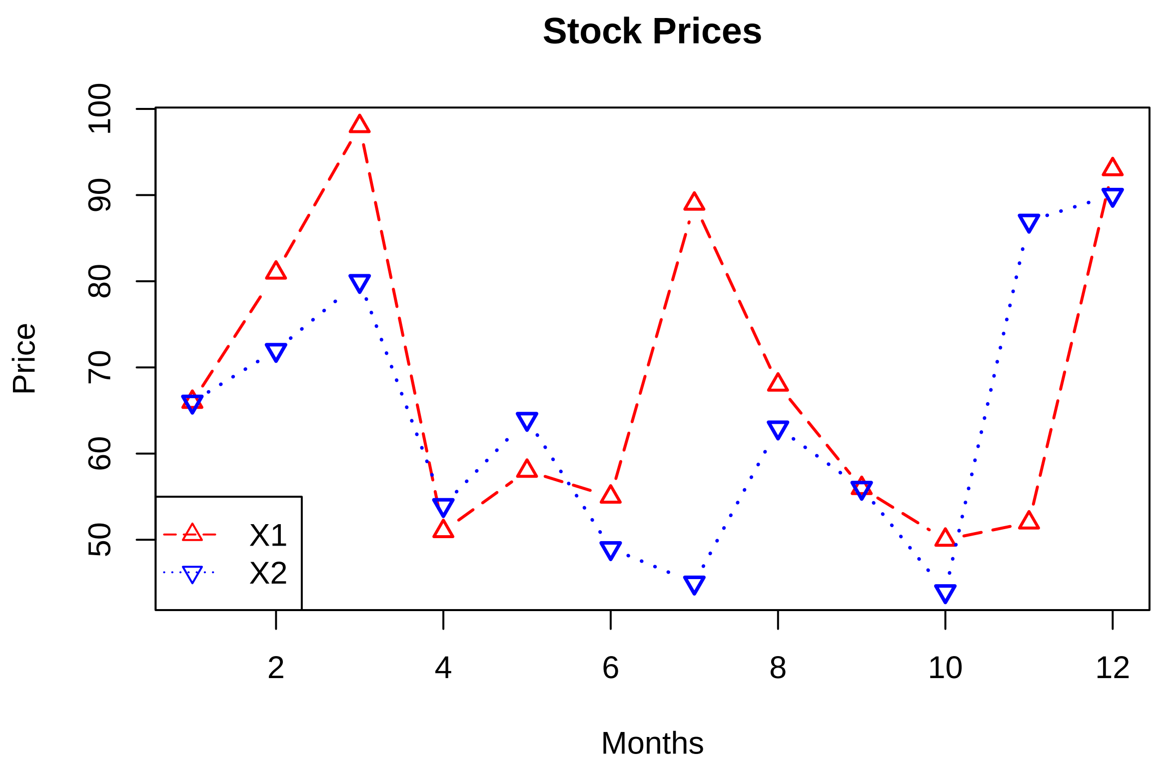
```
plot(x1, main="Stock Prices", xlab = "Months", ylab = "Price",
     type = "b", pch = 2, lwd = 1.5, lty = 5, col = "Red",
     ylim = c(min(x1,x2),max(x1,x2)))
lines(x2, type = "b", pch = 6, lwd = 1.8, lty = 3, col = "Blue")
```



Part E.

Add a legend to bottom left for these two lines.

```
plot(x1, main="Stock Prices", xlab = "Months", ylab = "Price",
     type = "b", pch = 2, lwd = 1.5, lty = 5, col = "Red",
     ylim = c(min(x1,x2),max(x1,x2)))
lines(x2, type = "b", pch = 6, lwd = 1.8, lty = 3, col = "Blue")
legend("bottomleft", legend = c("X1","X2"),
       col = c("Red","Blue"), lty = c(5,3), pch = c(2,6))
```



Summary Statistics:

Exercise 3:

Please load 'ISLR' package and from that package load 'Auto' data set.

```
library(ISLR)
data(Auto)
```

Part A.

Use str() and summary() functions for 'Auto' dataset.

```
str(Auto)
```

```
## 'data.frame':    392 obs. of  9 variables:
##  $ mpg         : num  18 15 18 16 17 15 14 14 14 15 ...
##  $ cylinders   : num  8 8 8 8 8 8 8 8 8 8 ...
##  $ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
##  $ horsepower  : num  130 165 150 150 140 198 220 215 225 190 ...
##  $ weight      : num  3504 3693 3436 3433 3449 ...
##  $ acceleration: num  12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
```

```
##  $ year        : num  70 70 70 70 70 70 70 70 70 70 ...
##  $ origin      : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ name        : Factor w/ 304 levels "amc ambassador brougham",..: 49 36 231 14 161
```

```r
summary(Auto)
```

```
##       mpg          cylinders      displacement     horsepower        weight    
##  Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1613  
##  1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225  
##  Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Median :2804  
##  Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mean   :2978  
##  3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615  
##  Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :5140  
##                                                                                
##   acceleration        year           origin                      name    
##  Min.   : 8.00   Min.   :70.00   Min.   :1.000   amc matador       :  5  
##  1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000   ford pinto        :  5  
##  Median :15.50   Median :76.00   Median :1.000   toyota corolla    :  5  
##  Mean   :15.54   Mean   :75.98   Mean   :1.577   amc gremlin       :  4  
##  3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :  4  
##  Max.   :24.80   Max.   :82.00   Max.   :3.000   chevrolet chevette:  4  
##                                                  (Other)           :365  
```

Part B.

Read data description from help menu and according to this document convert some of the numeric variables to factors. Then, use summary() function again and spot the difference.

```r
Auto$cylinders <- factor(Auto$cylinders)
Auto$origin <- factor(Auto$origin, levels = 1:3,
                      labels = c("American","European","Japanese"))
```

```r
summary(Auto)
```

```
##       mpg          cylinders  displacement     horsepower        weight    
##  Min.   : 9.00   3:  4      Min.   : 68.0   Min.   : 46.0   Min.   :1613  
##  1st Qu.:17.00   4:199      1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225  
##  Median :22.75   5:  3      Median :151.0   Median : 93.5   Median :2804  
##  Mean   :23.45   6: 83      Mean   :194.4   Mean   :104.5   Mean   :2978  
##  3rd Qu.:29.00   8:103      3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615  
##  Max.   :46.60              Max.   :455.0   Max.   :230.0   Max.   :5140  
##                                                                           
##   acceleration        year            origin                     name    
##  Min.   : 8.00   Min.   :70.00   American:245   amc matador       :  5  
##  1st Qu.:13.78   1st Qu.:73.00   European: 68   ford pinto        :  5  
##  Median :15.50   Median :76.00   Japanese: 79   toyota corolla    :  5  
```

```
##  Mean   :15.54   Mean   :75.98                   amc gremlin       :  4
##  3rd Qu.:17.02   3rd Qu.:79.00                   amc hornet        :  4
##  Max.   :24.80   Max.   :82.00                   chevrolet chevette:  4
##                                                  (Other)           :365
```

Part C.

Find mean of 'mpg'.

```
mean(Auto$mpg)
```

```
## [1] 23.44592
```

Part D.

Obtain median 'mpg', remember you have written your median function previous week, you can also use your own function.

```
median(Auto$mpg)
```

```
## [1] 22.75
```

Part E.

Find variance and standard deviation for 'mpg'

```
var(Auto$mpg);sd(Auto$mpg)
```

```
## [1] 60.91814
```

```
## [1] 7.805007
```

Part F.

Find quartiles; Q1 and Q3 for 'mpg'. Obtain IQR.

```
Q1 <- as.numeric(quantile(Auto$mpg,0.25))
Q3 <- as.numeric(quantile(Auto$mpg,0.75))

iqr <- IQR(Auto$mpg)

print(c(Q1 = Q1, Q3 = Q3, IQR = iqr))
```

```
##  Q1  Q3 IQR
##  17  29  12
```

Part G.

Find 5^{th}, 35^{th} and 95^{th} percentiles of 'mpg'.

```r
quantile(Auto$mpg, c(0.05,0.35,0.95))
```

```
##  5% 35% 95%
##  13  19  37
```

Part H.

Create a correlation and covariance matrix for all numeric variables in Auto data set.

Hint: First you have to subset only numeric variables, then use cor() and cov() function.

```r
numeric_variables <- names(Auto)[sapply(Auto, is.numeric)]
numeric_Auto <- Auto[,numeric_variables]
head(numeric_Auto)
```

```
##   mpg displacement horsepower weight acceleration year
## 1  18          307        130   3504         12.0   70
## 2  15          350        165   3693         11.5   70
## 3  18          318        150   3436         11.0   70
## 4  16          304        150   3433         12.0   70
## 5  17          302        140   3449         10.5   70
## 6  15          429        198   4341         10.0   70
```

```r
cor(numeric_Auto)
```

```
##                     mpg displacement horsepower     weight acceleration
## mpg           1.0000000   -0.8051269 -0.7784268 -0.8322442    0.4233285
## displacement -0.8051269    1.0000000  0.8972570  0.9329944   -0.5438005
## horsepower   -0.7784268    0.8972570  1.0000000  0.8645377   -0.6891955
## weight       -0.8322442    0.9329944  0.8645377  1.0000000   -0.4168392
## acceleration  0.4233285   -0.5438005 -0.6891955 -0.4168392    1.0000000
## year          0.5805410   -0.3698552 -0.4163615 -0.3091199    0.2903161
##                    year
## mpg           0.5805410
## displacement -0.3698552
## horsepower   -0.4163615
## weight       -0.3091199
## acceleration  0.2903161
## year          1.0000000
```

```r
cov(numeric_Auto)
```

```
##                      mpg displacement  horsepower      weight acceleration
## mpg            60.918142    -657.5852  -233.85793  -5517.4407     9.115514
## displacement -657.585207   10950.3676  3614.03374  82929.1001  -156.994435
## horsepower   -233.857926    3614.0337  1481.56939  28265.6202   -73.186967
## weight      -5517.440704   82929.1001 28265.62023 721484.7090  -976.815253
```

```
## acceleration     9.115514    -156.9944   -73.18697   -976.8153     7.611331
## year            16.691477    -142.5721   -59.03643   -967.2285     2.950462
##                      year
## mpg             16.691477
## displacement -142.572133
## horsepower    -59.036432
## weight       -967.228457
## acceleration    2.950462
## year           13.569915
```

Part I.

Obtain frequencies of 'cylinders' and 'origin' variables.

```
table(Auto$cylinders)
```

```
##
##   3   4   5   6   8
##   4 199   3  83 103
```

```
table(Auto$origin)
```

```
##
## American European Japanese
##      245       68       79
```

Part J.

Find proportion of observations that fall into each category for both 'cylinders' and 'origin'.

```
table(Auto$cylinders) / nrow(Auto)
```

```
##
##           3           4           5           6           8
## 0.010204082 0.507653061 0.007653061 0.211734694 0.262755102
```

```
table(Auto$origin) / nrow(Auto)
```

```
##
##  American  European  Japanese
## 0.6250000 0.1734694 0.2015306
```

Part K.

Create a contingency table for 'cylinders' VS 'origin' and make comments on it.

```
table(Auto$cylinders, Auto$origin)
```

```
##
```

```
##      American European Japanese
##    3        0        0        4
##    4       69       61       69
##    5        0        3        0
##    6       73        4        6
##    8      103        0        0
```