mda14jx_week2

August 21, 2018

```
In [90]: options(jupyter.plot_mimetypes ='image/png')
```

1 Exercise 1

```
In [91]: names<- c('Bob','Claire','Luisa','Matt','Marta','Mike')
         score<- c(34,82,59,72,50,100)

         game_cards<- data.frame(names,score,stringsAsFactors=FALSE)
         game_cards
```

```
In [6]: game_cards$names
```

```
In [92]: names<- c('Bob','Claire','Luisa','Matt','Marta','Mike')
         score1<- c(34,82,59,72,50,100)
         score2<- c(64,82,36,48,29,85)

         game_cards<- data.frame(names,score1,score2,stringsAsFactors=FALSE)
         game_cards
```

```
In [8]: game_cards$score1
```

```
In [9]: game_cards$score2
```

The additional field for the second match score is created by adding score2<- c(, , , , ,), with values of the same length as previous. score2 was also added to the data.frame. Each field is accessed seperately by game_cards$fieldname fieldname:score1/ score2

2 Ecercise 2

```
In [93]: names(game_cards)<- c("names","match1","match2")
         game_cards
```

```
In [18]: dim(game_cards)
```

```
In [29]: min(game_cards$match1)
         min(game_cards$match2)
```

The minimum scores from match 1 and match 2 were 34 and 29 respectively.

In [30]:
```
max(game_cards$match1)
max(game_cards$match2)
```

The maximum scores from mathc 1 and match 2 were 100 and 85 respectively.

In [28]:
```
min(game_cards[,2:3])
```

The minimum score from both matches was 29.

In [32]:
```
max(game_cards[,2:3])
```

The maximum score from both mathes was 100

In [50]:
```
which.min(game_cards$match1)
which.min(game_cards$match2)
```

In [51]:
```
which.max(game_cards$match1)
which.max(game_cards$match2)
```

In [52]:
```
names[which.min(game_cards$match1)]
names[which.min(game_cards$match2)]
```

In [53]:
```
names[which.max(game_cards$match1)]
names[which.max(game_cards$match2)]
```

In [71]:
```
x<- c(min(game_cards$match1))
y<- c(names[which.min(game_cards$match1)])
z<- c(y,as.character(x))
print(z)

a<- c(min(game_cards$match2))
b<- c(names[which.min(game_cards$match2)])
c<- c(b,as.character(a))
print(c)
```

```
[1] "Bob" "34"
[1] "Marta" "29"
```

The minimum score from match 1 was 34, which was scored by Bob. The minimum score from match 2 was 29, which was scored by Marta.

In [75]:
```
d<- c(max(game_cards$match1))
e<- c(names[which.max(game_cards$match1)])
f<- c(e,as.character(d))
print(f)

g<- c(max(game_cards$match2))
h<- c(names[which.max(game_cards$match2)])
i<- c(h,as.character(g))
print(i)
```

```
[1] "Mike" "100"
[1] "Mike" "85"
```

The maximum score from match 1 was 100, which was scored by Mike. The maximum score from match 2 was 85, which was also scored by Mike.

```
In [94]: game_cards[order(game_cards$match1),]
```

```
In [95]: game_cards[order(game_cards$match2),]
```

The function order() arranges the sequence of numbers into an ascending order. order(game_cards$score) rearranges the scores of the matches on the game cards into a sequential order. The output is the ordered sequence of numbers.

```
In [3]: ?plot
```

The command plot() is used for generic X-Y graph plotting of R objects, through the use of command plot(x,y,...). x = the coordinates of points in the plot. y = the y co-ordinates in the plot. ... = arguments to be passed to the methods, such as graphical parameters. "type" = the type of plot that should be drawn, eg. "p" for points, "l" for lines, "b" for both etc. To add an overall title to the plot, add "main", to add a subtitle fot the plot, add "sub", to add a title for the x and y axis, add "xlab" and "ylab" respectively.

```
In [5]: par(mfrow=c(1,2))
        barplot(game_cards$match1, names=game_cards$names)
        barplot(game_cards$match2, names=game_cards$names)


        Error in barplot(game_cards$match1, names = game_cards$names): object 'game_cards' not
    Traceback:


        1. barplot(game_cards$match1, names = game_cards$names)
```

3 Exercise 4

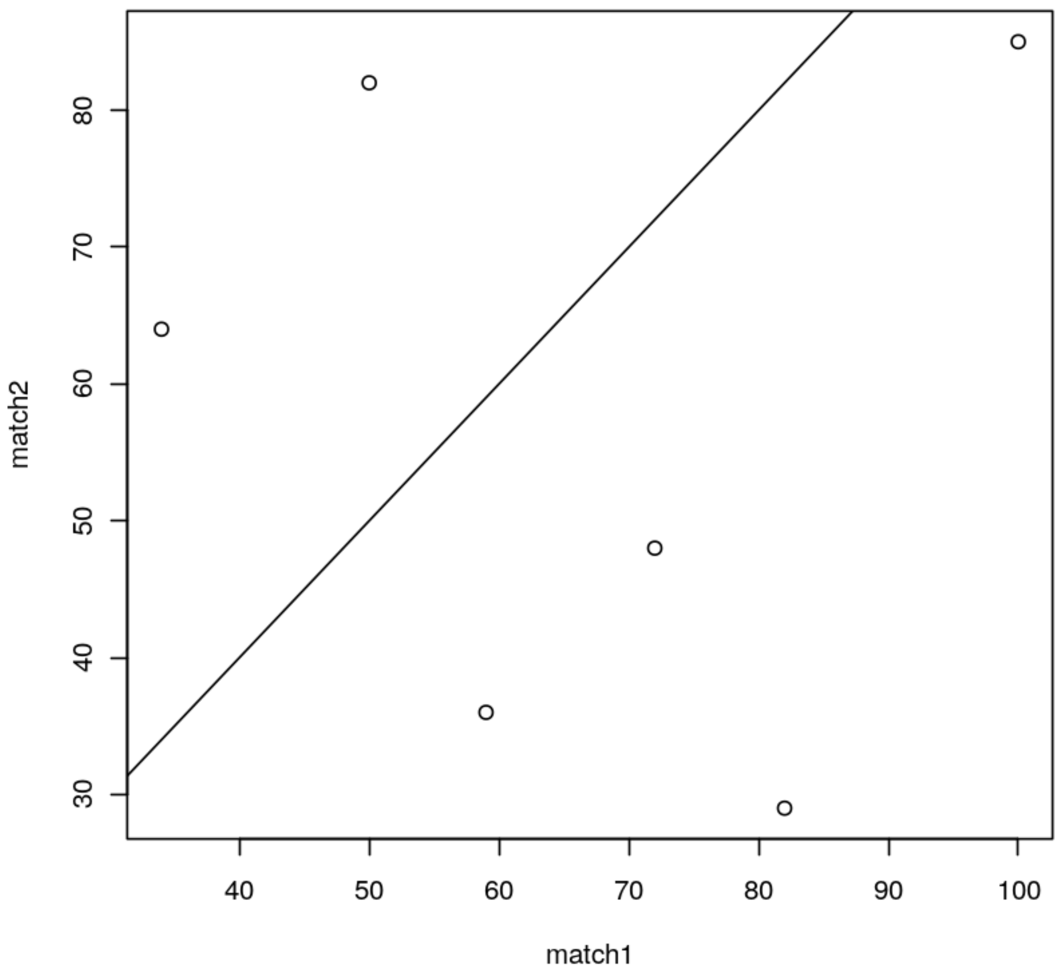
```
In [7]: ?par
```

The par() command is used to set or query graphucal parameters. Parameters can be set by specifying them as arguments to par in tag = value form, or by passing them as a list of tagged values.

4 Exercise 5

```
In [96]: match1<- c(34,82,59,72,50,100)
         match2<-c(64,29,36,48,82,85)
         plot(match1,match2)
         abline(0,1)
```

Out[96]:

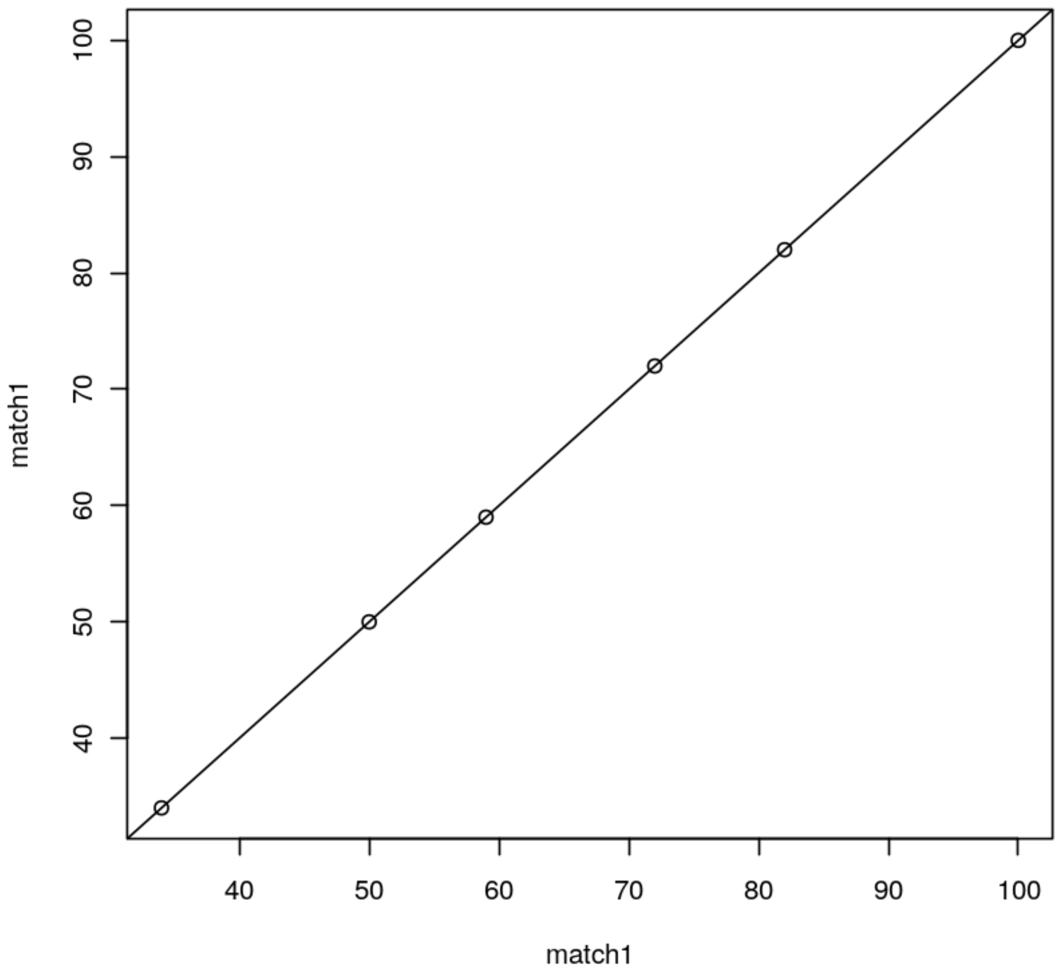


Scatter plots are used to plot data points on a horizontal and vertical axis to show how much one variable is affected by another. It uses cartesian co-ordinates to display values for typically two variables of a set of data. A scatter plot can be used either when one continuous variable that is under the control of the experimenter and the other depends on it or when both continuous variables are independent. If a parameter exists that is systematically incremented and/or decremented by the other, it is called the control parameter or independent variable and is customarily plotted along the horizontal axis. The measured or dependent variable is customarily plotted along the vertical axis. If no dependent variable exists, either type of variable can be plotted on either axis and a scatter plot will illustrate only the degree of correlation (not causation) between two variables.

```
In [97]: match1<- c(34,82,59,72,50,100)
```

```
plot(match1,match1)
abline(0,1)
```

Out[97]:



By plotting the values of a variable against itself, the resulting scatter plot shows the points falling along a straight line, with the line of best fit travelling directly through all points.

5 Exercise 6

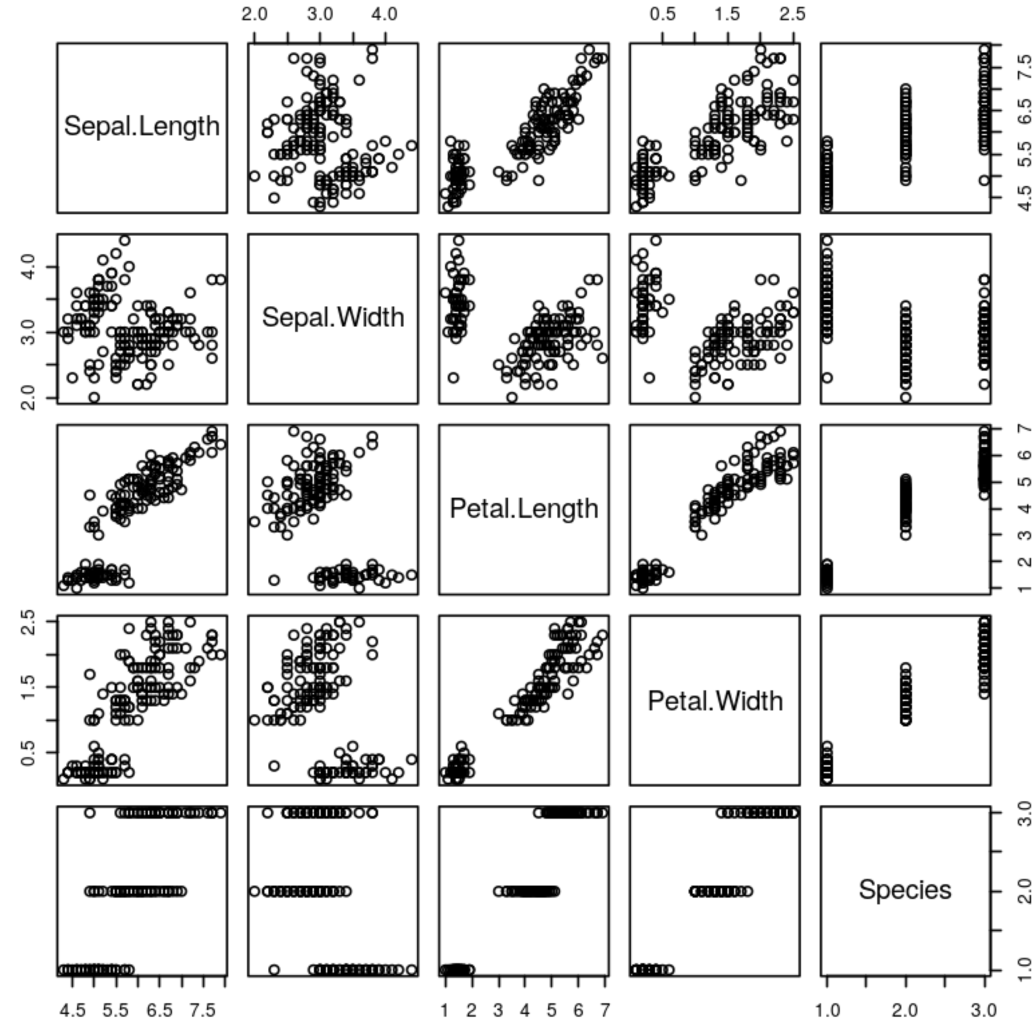
In [98]:
```
data(iris)
?iris
iris
```

In [16]: summary(iris)

Out[16]:
```
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width   
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100  
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300  
 Median :5.800   Median :3.000   Median :4.350   Median :1.300  
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199  
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800  
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500  
       Species  
 setosa    :50  
 versicolor:50  
 virginica :50  
```

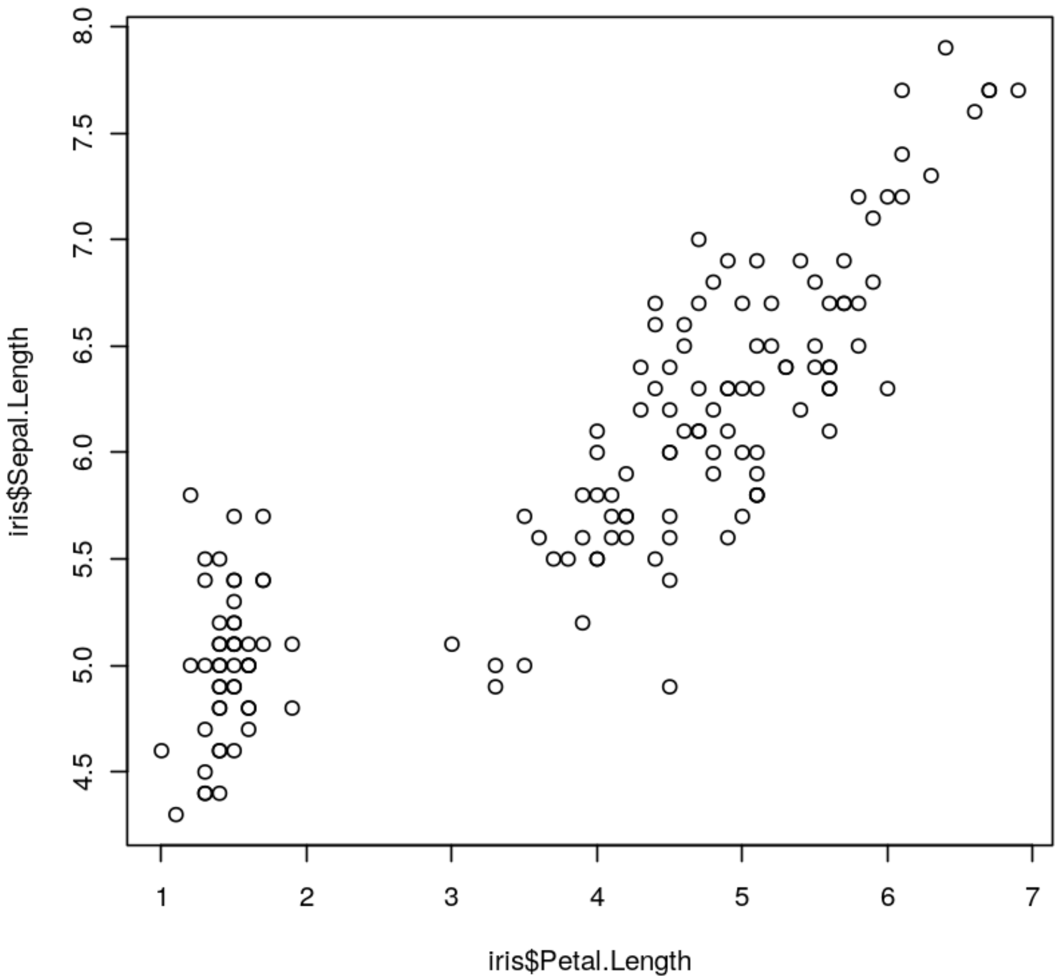
In [99]: plot(iris)

Out[99]:



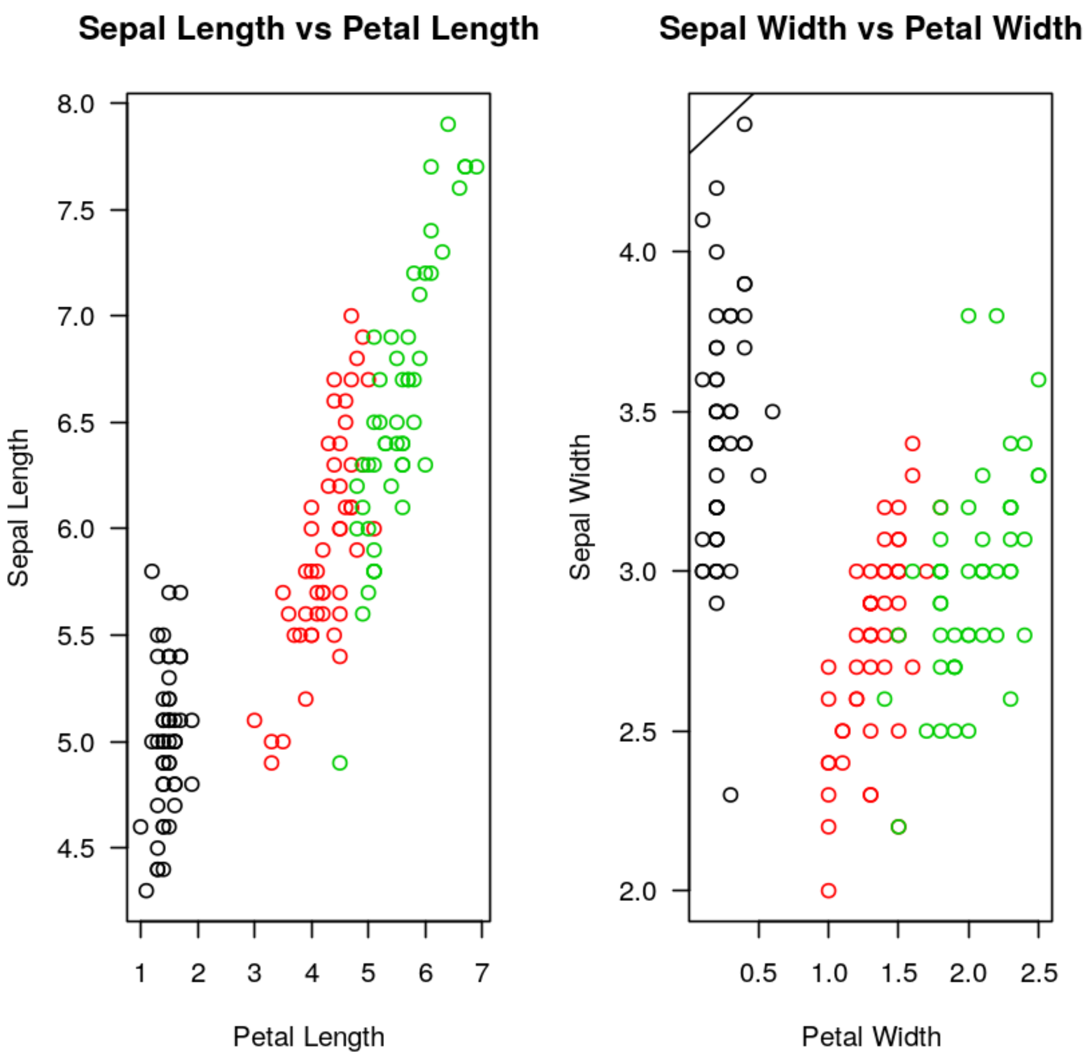

```
In [100]: plot(iris$Sepal.Length~iris$Petal.Length)
```

Out[100]:



In [101]:
```
par(mfrow=c(1,2))
plot(iris$Sepal.Length~iris$Petal.Length, xlab="Petal Length",ylab="Sepal Length", ma
plot(iris$Sepal.Width~iris$Petal.Width, xlab="Petal Width", ylab="Sepal Width", main=
reg1<- lm(iris$Sepal.Length~iris$Petal.Length)
reg2<- lm(iris$Sepal.Width~iris$Petal.Width)
abline(reg1,reg2)
```

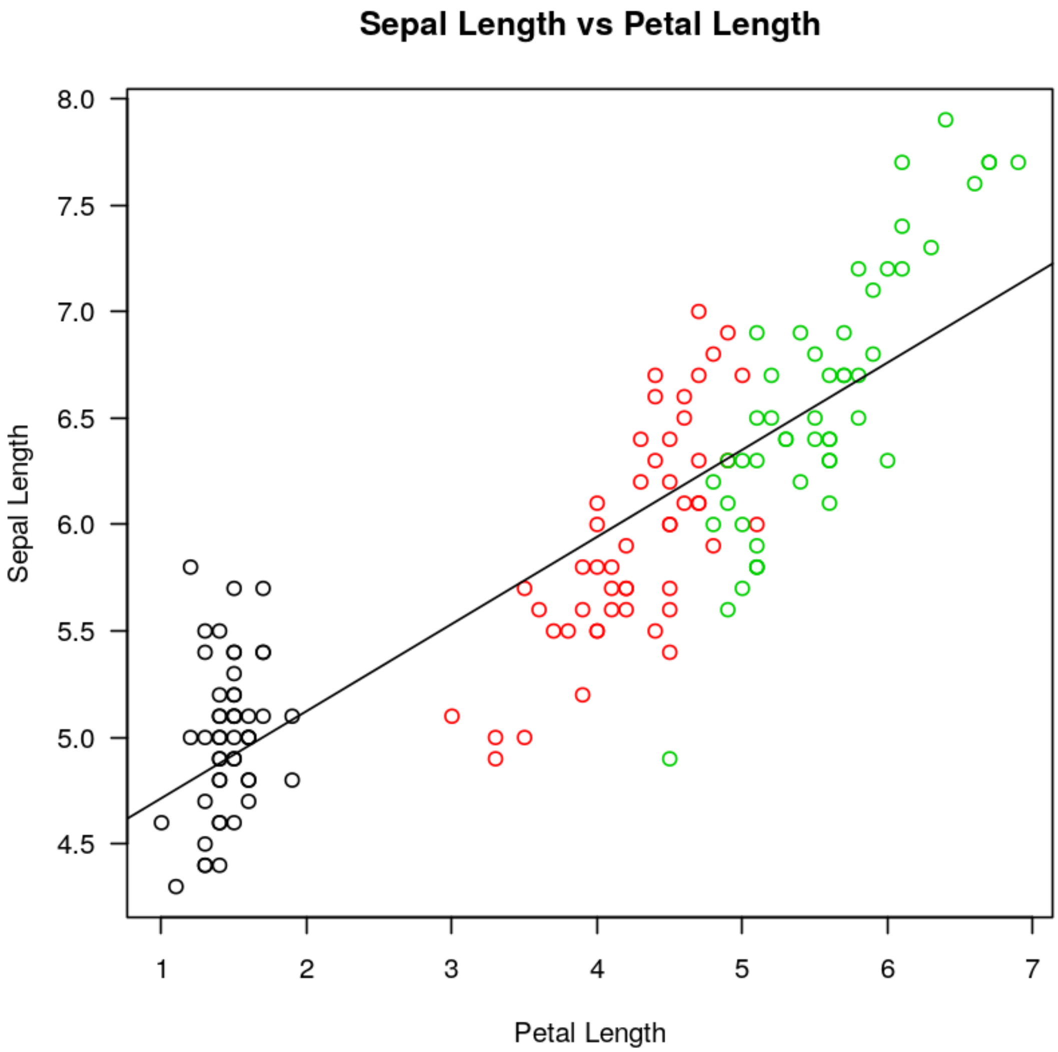
Out[101]:



6 Exercise 7

In [102]:
```
plot(iris$Sepal.Length~iris$Petal.Length, xlab="Petal Length",ylab="Sepal Length", ma
reg1<-lm(iris$Sepal.Length~iris$Petal.Length)
abline(reg1)
plot(iris$Sepal.Width~iris$Petal.Width, xlab="Petal Width", ylab="Sepal Width", main=
reg2<-lm(iris$Sepal.Width~iris$Petal.Width)
abline(reg2)
```

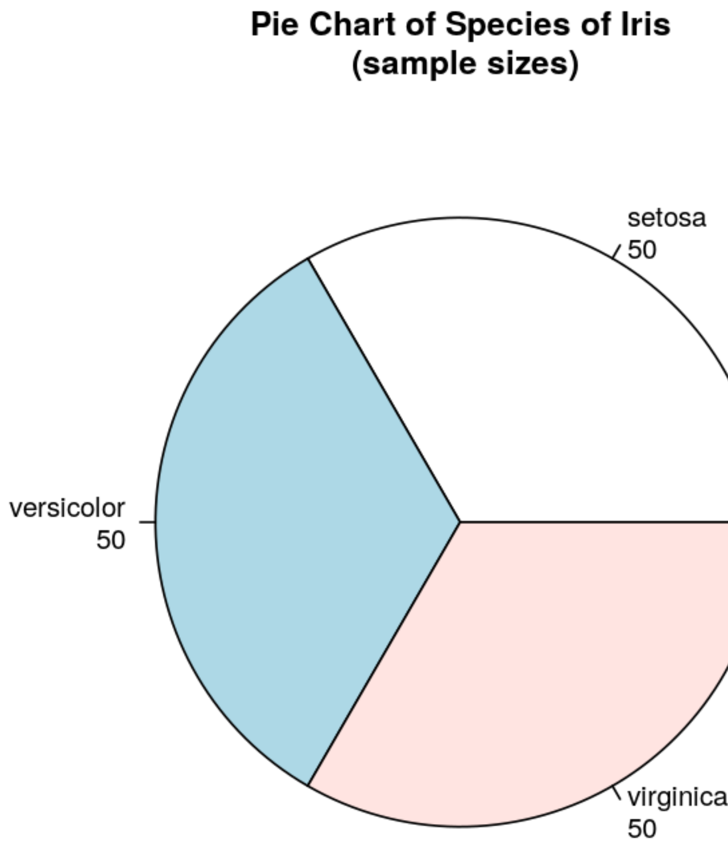
Out[102]:


Sepal Length vs Petal Length
Sepal Length
Petal Length
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
1
2
3
4
5
6
7

Out[102]:


```
In [103]: iris_table <- table(iris$Species)
          lbls <- paste(names(iris_table), "\n", iris_table, sep="")
          pie(iris_table, labels = lbls,
                    main="Pie Chart of Species of Iris\n (sample sizes)")
```

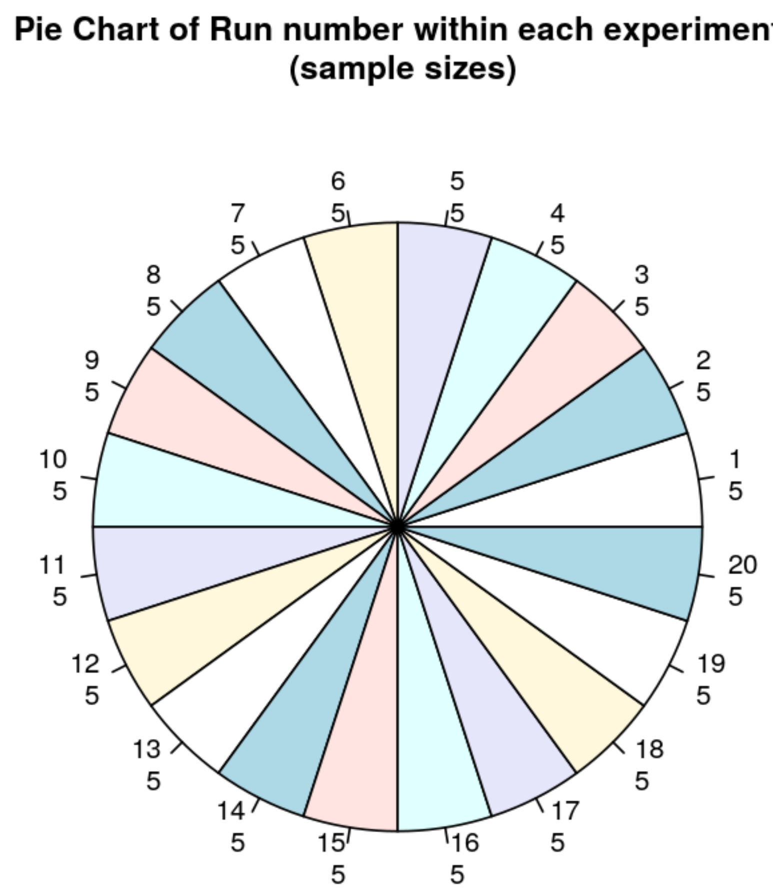
Out[103]:



7 Exercise 8

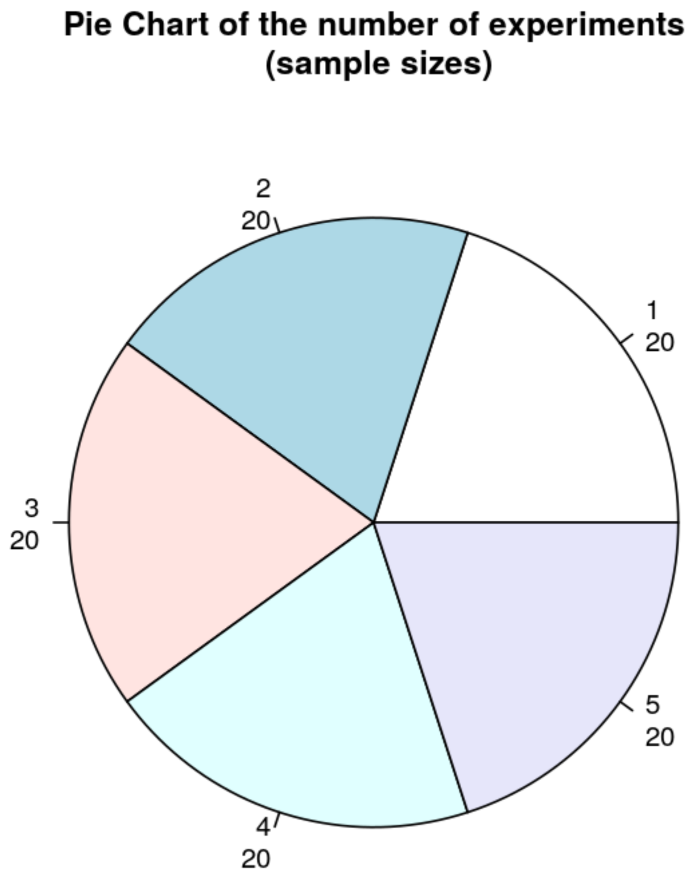
```
In [104]: data(morley)
          ?morley
          morley

In [105]: morley_table <- table(morley$Run)
          lbls <- paste(names(morley_table), "\n", morley_table, sep="")
          pie(morley_table, labels = lbls,
                    main="Pie Chart of Run number within each experiment\n (sample sizes)")
```

Out[105]:



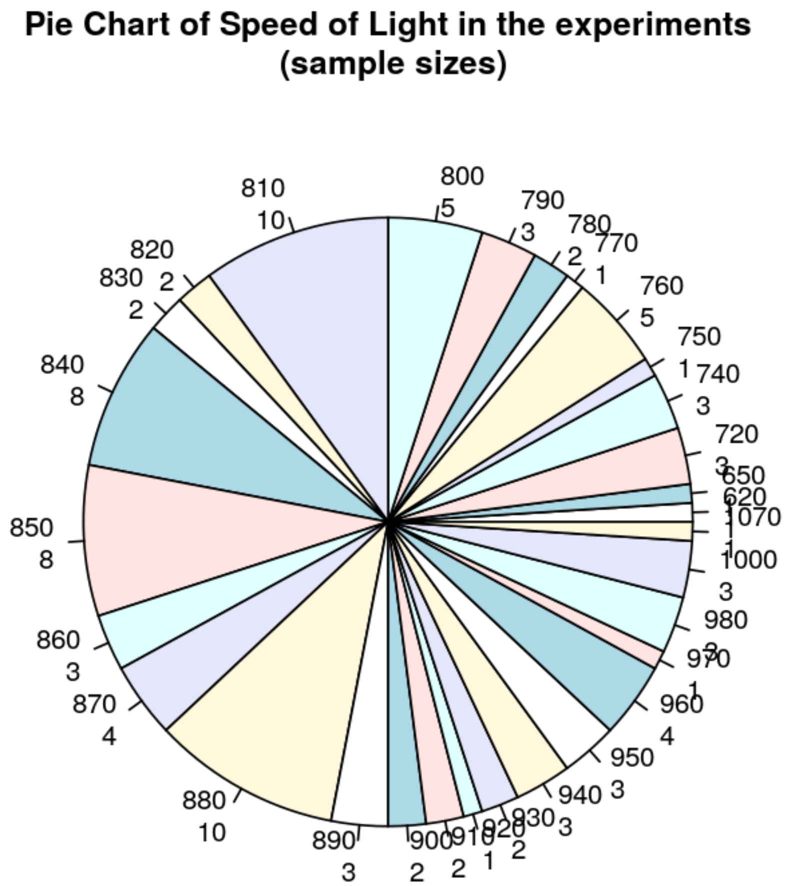

```
In [106]: morley_table <- table(morley$Expt)
          lbls <- paste(names(morley_table), "\n", morley_table, sep="")
          pie(morley_table, labels = lbls,
                    main="Pie Chart of the number of experiments\n (sample sizes)")
```

Out[106]:



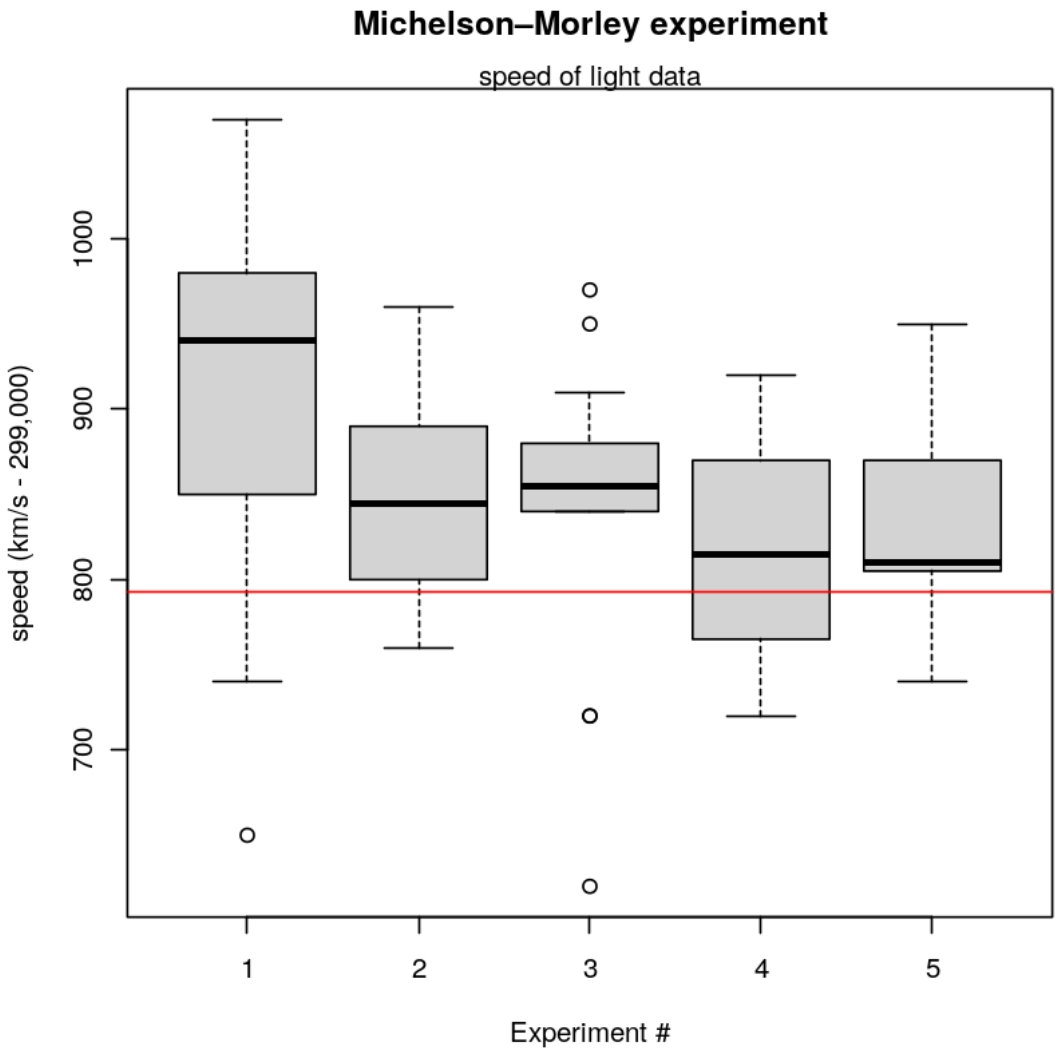
In [107]:
```
morley_table <- table(morley$Speed)
lbls <- paste(names(morley_table), "\n", morley_table, sep="")
pie(morley_table, labels = lbls,
          main="Pie Chart of Speed of Light in the experiments\n (sample sizes)")
```

Out[107]:




```
In [108]: boxplot(morley$Speed ~ morley$Expt,
            col='light grey', xlab='Experiment #',
            ylab="speed (km/s - 299,000)",
            main="MichelsonMorley experiment")
          mtext("speed of light data")

          sol=299792.458-299000
          abline(h=sol, col='red')
```

Out[108]:



8 Exercise 9

```
In [109]: quantile(morley$Speed,prob=0.75)[["75%"]] + 1.5*IQR(morley$Speed)
```

```
In [47]: quantile(morley$Speed,prob=0.25)[["25%"]] - 1.5*IQR(morley$Speed)
```

```
In [48]: quantile(morley$Speed,prob=0.25)
```

```
In [49]: quantile(morley$Speed,prob=0.50)
```

```
In [45]: quantile(morley$Speed,prob=0.75)
```

```
In [46]: IQR(morley$Speed)
```

```
In [50]: mean(morley$Speed)
```

```
In [51]: sd(morley$Speed)
```

```
In [75]: Expt1<- (morley$Speed[morley$Expt==1])
         Expt1
         quantile(Expt1,prob=0.75)[["75%"]] + 1.5*IQR(Expt1)
         quantile(Expt1,prob=0.25)[["25%"]] - 1.5*IQR(Expt1)
         quantile(Expt1,prob=0.25)
         quantile(Expt1,prob=0.50)
         quantile(Expt1,prob=0.75)
         IQR(Expt1)
         mean(Expt1)
         sd(Expt1)
```

```
In [76]: Expt2<- (morley$Speed[morley$Expt==2])
         Expt2
         quantile(Expt2,prob=0.75)[["75%"]] + 1.5*IQR(Expt2)
         quantile(Expt2,prob=0.25)[["25%"]] - 1.5*IQR(Expt2)
         quantile(Expt2,prob=0.25)
         quantile(Expt2,prob=0.50)
         quantile(Expt2,prob=0.75)
         IQR(Expt2)
         mean(Expt2)
         sd(Expt2)
```

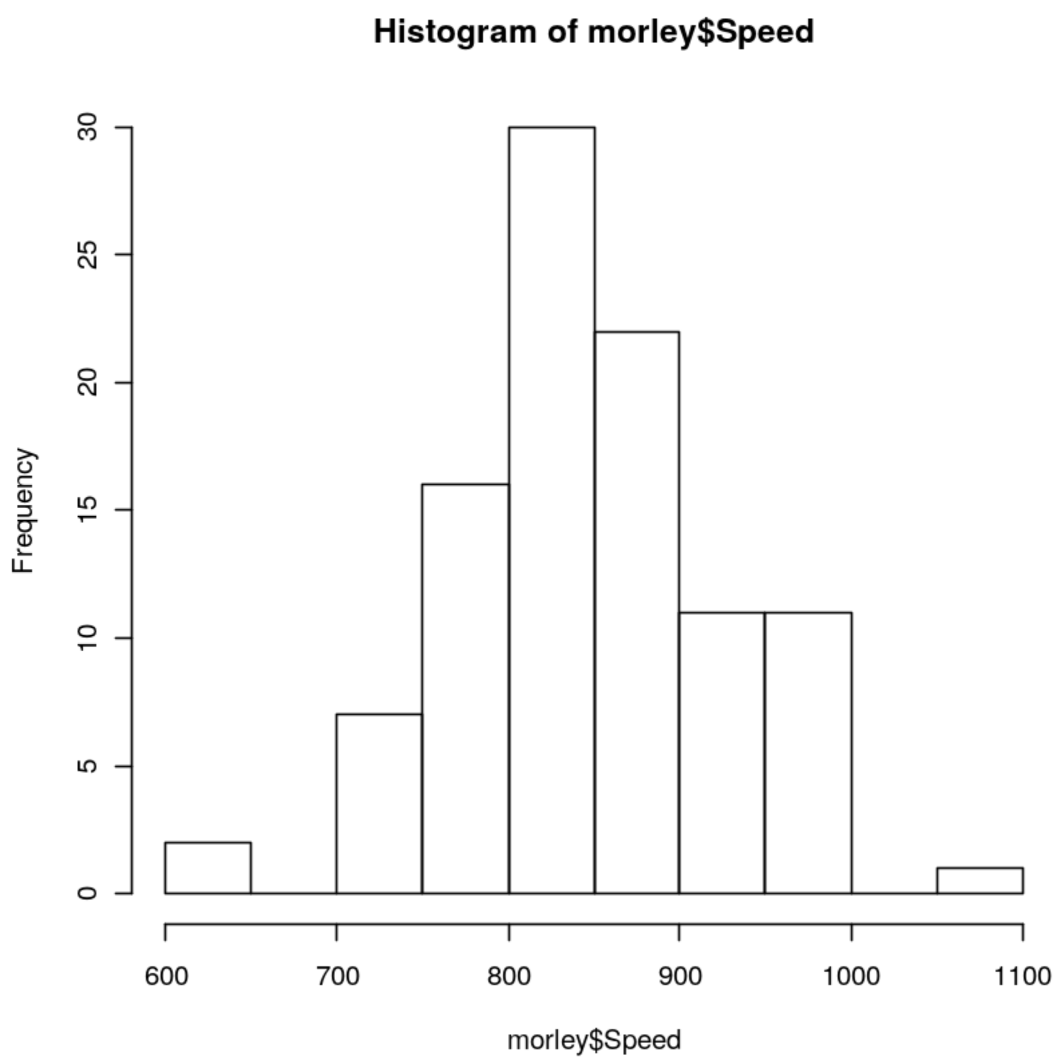
```
In [77]: Expt4<- (morley$Speed[morley$Expt==4])
         Expt4
         quantile(Expt4,prob=0.75)[["75%"]] + 1.5*IQR(Expt4)
         quantile(Expt4,prob=0.25)[["25%"]] - 1.5*IQR(Expt4)
         quantile(Expt4,prob=0.25)
         quantile(Expt4,prob=0.50)
         quantile(Expt4,prob=0.75)
         IQR(Expt4)
         mean(Expt4)
         sd(Expt4)
```

```
In [78]: Expt5<- (morley$Speed[morley$Expt==5])
         Expt5
         quantile(Expt5,prob=0.75)[["75%"]] + 1.5*IQR(Expt5)
         quantile(Expt5,prob=0.25)[["25%"]] - 1.5*IQR(Expt5)
         quantile(Expt5,prob=0.25)
         quantile(Expt5,prob=0.50)
         quantile(Expt5,prob=0.75)
         IQR(Expt5)
         mean(Expt5)
         sd(Expt5)
```

9 Ecercise 10

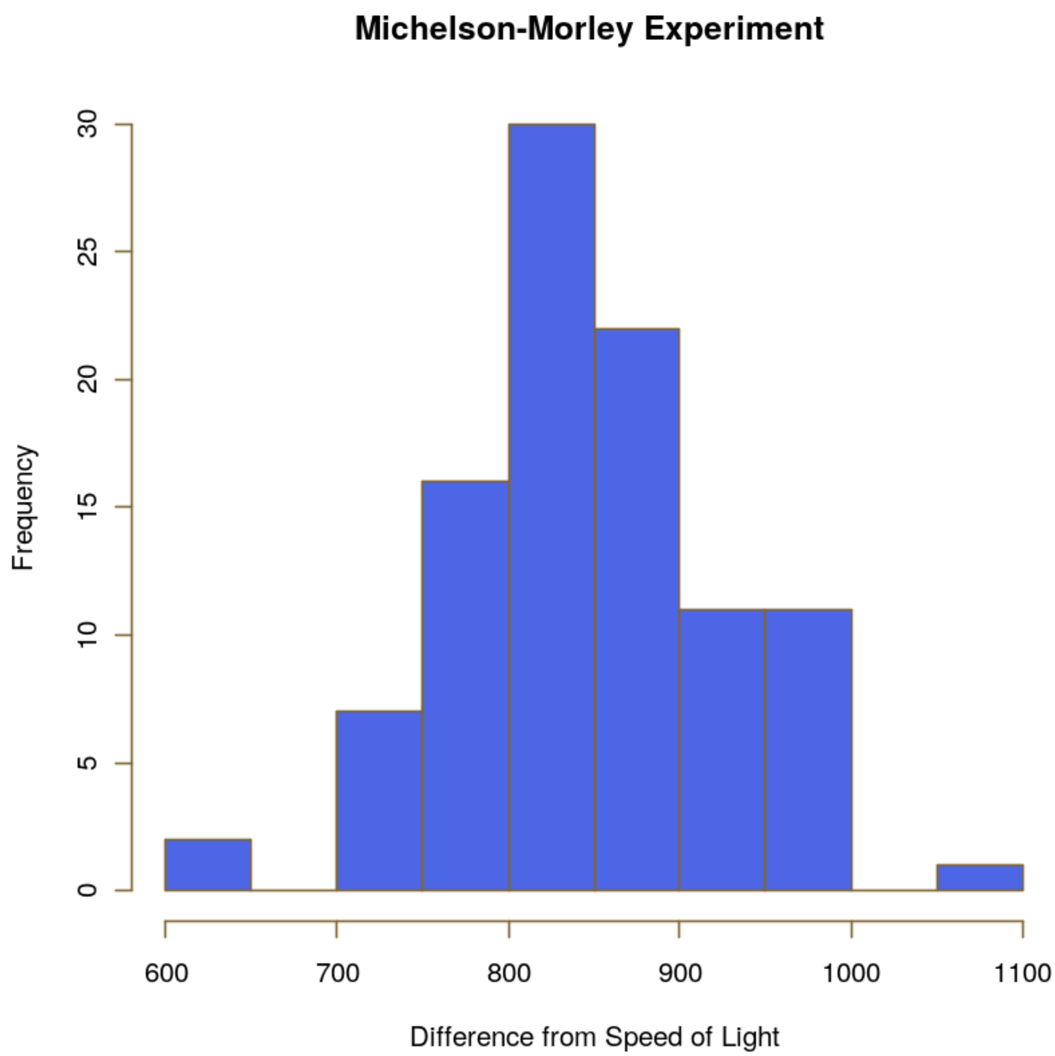
```
In [110]: hist(morley$Speed)
```

Out[110]:



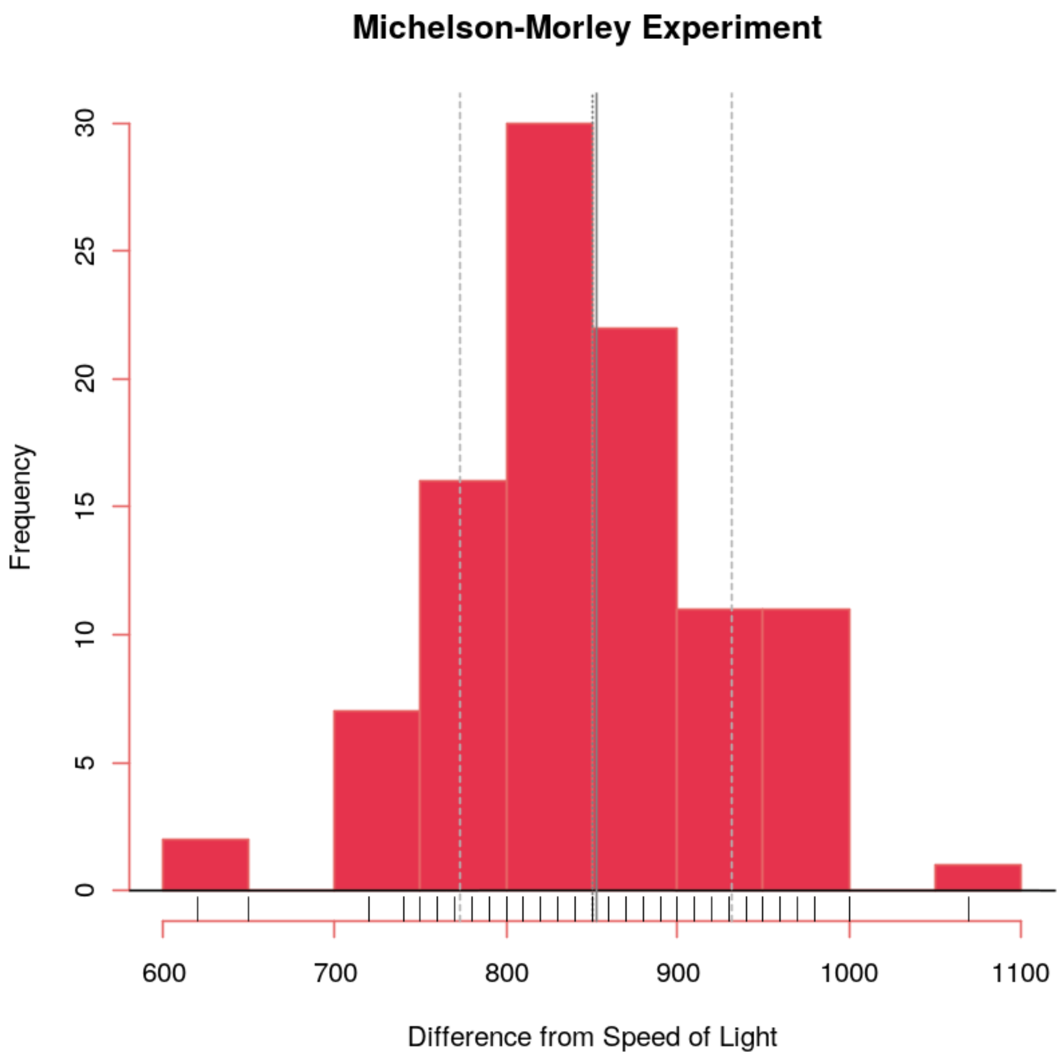

```
In [128]: par(fg=rgb(0.5,0.4,0.2))
          hist(morley$Speed, prob=F,
               col=rgb(0.3,0.4,0.9),
               main='Michelson-Morley Experiment ',
               ylab="Frequency", xlab='Difference from Speed of Light')
          par(fg='black')
```

Out[128]:




```
In [126]: par(fg=rgb(0.9,0.4,0.4))
          hist(morley$Speed, prob=F,
               col=rgb(0.9,0.2,0.3),
               main='Michelson-Morley Experiment ',
               ylab="Frequency", xlab='Difference from Speed of Light')
          par(fg='black')

          lines(density(morley$Speed))
          abline(v=mean(morley$Speed), col=rgb(0.5,0.5,0.5))
          abline(v=median(morley$Speed), lty=3, col=rgb(0.5,0.5,0.5))
          abline(v=mean(morley$Speed)+sd(morley$Speed), lty=2, col=rgb(0.7,0.7,0.7))
          abline(v=mean(morley$Speed)-sd(morley$Speed), lty=2, col=rgb(0.7,0.7,0.7))
          rug(morley$Speed)
```

Out[126]:



In [130]:
```
par(fg=rgb(0.6,0.2,0.3))
hist(morley$Speed[morley$Expt==1], prob=F,
    col=rgb(0.7,0.1,0.3),
    main='Michelson-Morley Experiment 1 ',
     ylab="Frequency", xlab='Difference from Speed of Light')
par(fg='black')

lines(density(morley$Speed[morley$Expt==1]))
abline(v=mean(morley$Speed[morley$Expt==1]), col=rgb(0.5,0.5,0.5))
abline(v=median(morley$Speed[morley$Expt==1]), lty=3, col=rgb(0.5,0.5,0.5))
abline(v=mean(morley$Speed[morley$Expt==1])+sd(morley$Speed[morley$Expt==1]), lty=2,
```

```
abline(v=mean(morley$Speed[morley$Expt==1])-sd(morley$Speed[morley$Expt==1]), lty=2,
rug(morley$Speed[morley$Expt==1])
```

Out[130]:

In [133]:

```
par(fg=rgb(0.6,0.5,0.7))
hist(morley$Speed[morley$Expt==2], prob=F,
    col=rgb(0.3,0.7,0.9),
    main='Michelson-Morley Experiment 2 ',
     ylab="Frequency", xlab='Difference from Speed of Light')
par(fg='black')

lines(density(morley$Speed[morley$Expt==2]))
```

```
abline(v=mean(morley$Speed[morley$Expt==2]), col=rgb(0.5,0.5,0.5))
abline(v=median(morley$Speed[morley$Expt==2]), lty=3, col=rgb(0.5,0.5,0.5))
abline(v=mean(morley$Speed[morley$Expt==2])+sd(morley$Speed[morley$Expt==2]), lty=2,
abline(v=mean(morley$Speed[morley$Expt==2])-sd(morley$Speed[morley$Expt==2]), lty=2,
rug(morley$Speed[morley$Expt==2])
```

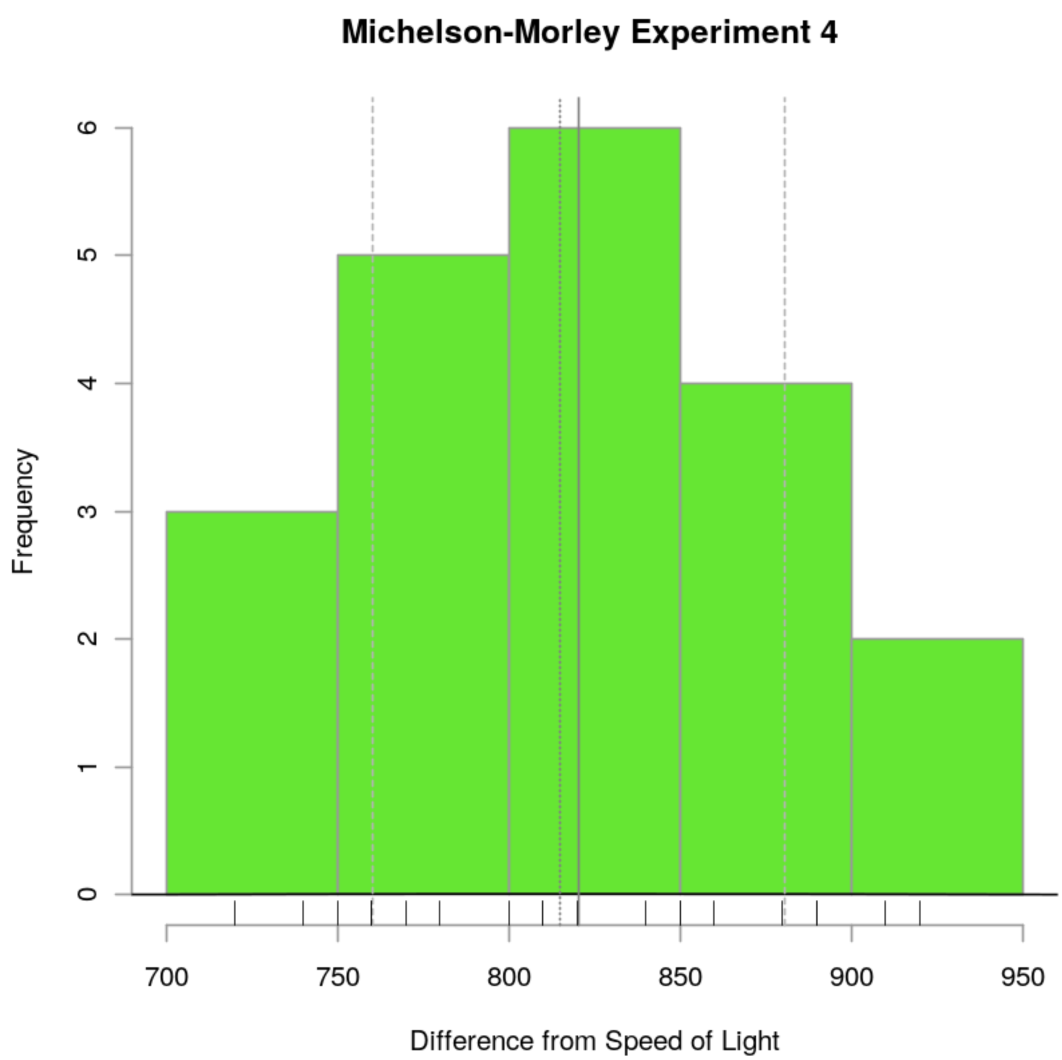
Out[133]:

In [134]:

```
par(fg=rgb(0.6,0.6,0.6))
hist(morley$Speed[morley$Expt==4], prob=F,
    col=rgb(0.4,0.9,0.2),
    main='Michelson-Morley Experiment 4 ',
     ylab="Frequency", xlab='Difference from Speed of Light')
```

```
par(fg='black')

lines(density(morley$Speed[morley$Expt==4]))
abline(v=mean(morley$Speed[morley$Expt==4]), col=rgb(0.5,0.5,0.5))
abline(v=median(morley$Speed[morley$Expt==4]), lty=3, col=rgb(0.5,0.5,0.5))
abline(v=mean(morley$Speed[morley$Expt==4])+sd(morley$Speed[morley$Expt==4]), lty=2,
abline(v=mean(morley$Speed[morley$Expt==4])-sd(morley$Speed[morley$Expt==4]), lty=2,
rug(morley$Speed[morley$Expt==4])
```

Out[134]:




```
In [136]: par(fg=rgb(0.6,0.6,0.6))
          hist(morley$Speed[morley$Expt==5], prob=F,
```

```
     col=rgb(0.7,0.2,0.6),
     main='Michelson-Morley Experiment 5 ',
      ylab="Frequency", xlab='Difference from Speed of Light')
par(fg='black')

lines(density(morley$Speed[morley$Expt==5]))
abline(v=mean(morley$Speed[morley$Expt==5]), col=rgb(0.5,0.5,0.5))
abline(v=median(morley$Speed[morley$Expt==5]), lty=3, col=rgb(0.5,0.5,0.5))
abline(v=mean(morley$Speed[morley$Expt==5])+sd(morley$Speed[morley$Expt==5]), lty=2,
abline(v=mean(morley$Speed[morley$Expt==5])-sd(morley$Speed[morley$Expt==5]), lty=2,
rug(morley$Speed[morley$Expt==5])
```

Out[136]:

10 Exercise 11

```
In [138]: rnorm(morley$Speed)

In [1]: ?rnorm()

In [2]: runif(morley$Speed)

In [3]: rbinom(morley$Speed)
```

```
    Error in rbinom(morley$Speed): argument "size" is missing, with no default
Traceback:


    1. rbinom(morley$Speed)
```

11 Exercise 12

```
In [141]: t.test(morley$Speed[morley$Expt==1], morley$Speed[morley$Expt==2])
```

```
Out[141]:
                  Welch Two Sample t-test

        data:  morley$Speed[morley$Expt == 1] and morley$Speed[morley$Expt == 2]
        t = 1.9516, df = 30.576, p-value = 0.0602
        alternative hypothesis: true difference in means is not equal to 0
        95 percent confidence interval:
          -2.419111 108.419111
        sample estimates:
        mean of x mean of y
              909       856
```

There is not a significant difference between the results of experiment 1 and 2, as p>0.05.

```
In [142]: t.test(morley$Speed[morley$Expt==1], morley$Speed[morley$Expt==4])
```

```
Out[142]:
                  Welch Two Sample t-test

        data:  morley$Speed[morley$Expt == 1] and morley$Speed[morley$Expt == 4]
        t = 3.2739, df = 30.238, p-value = 0.002659
        alternative hypothesis: true difference in means is not equal to 0
        95 percent confidence interval:
          33.31171 143.68829
        sample estimates:
        mean of x mean of y
            909.0     820.5
```

There is a significance difference between the results of experiment 1 and 4, as p<0.05

```
In [144]: t.test(morley$Speed[morley$Expt==1], morley$Speed[morley$Expt==5])
```

```
Out[144]:
                Welch Two Sample t-test

        data:  morley$Speed[morley$Expt == 1] and morley$Speed[morley$Expt == 5]
        t = 2.9346, df = 28.471, p-value = 0.006538
        alternative hypothesis: true difference in means is not equal to 0
        95 percent confidence interval:
          23.44296 131.55704
        sample estimates:
        mean of x mean of y
            909.0     831.5
```

There is a significant difference between the results of experiment 1 and 5, as p<0.05.

```
In [145]: t.test(morley$Speed[morley$Expt==2], morley$Speed[morley$Expt==4])
```

```
Out[145]:
                Welch Two Sample t-test

        data:  morley$Speed[morley$Expt == 2] and morley$Speed[morley$Expt == 4]
        t = 1.8523, df = 37.987, p-value = 0.07176
        alternative hypothesis: true difference in means is not equal to 0
        95 percent confidence interval:
         -3.298237 74.298237
        sample estimates:
        mean of x mean of y
            856.0     820.5
```

There is not a significant difference between the results of experiment 2 and 4 because p>0.05.

```
In [146]: t.test(morley$Speed[morley$Expt==2], morley$Speed[morley$Expt==5])
```

```
Out[146]:
                Welch Two Sample t-test

        data:  morley$Speed[morley$Expt == 2] and morley$Speed[morley$Expt == 5]
        t = 1.3405, df = 37.461, p-value = 0.1882
        alternative hypothesis: true difference in means is not equal to 0
        95 percent confidence interval:
         -12.51683  61.51683
        sample estimates:
        mean of x mean of y
            856.0     831.5
```

There is not a significant difference between the results of experiment 2 and 5, as p>0.05.

In [147]:
```
t.test(morley$Speed[morley$Expt==4], morley$Speed[morley$Expt==5])
```

Out[147]:
```
        Welch Two Sample t-test

data:  morley$Speed[morley$Expt == 4] and morley$Speed[morley$Expt == 5]
t = -0.60808, df = 37.611, p-value = 0.5468
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -47.63309  25.63309
sample estimates:
mean of x mean of y
    820.5     831.5
```

There is not a significant difference between the results of experiment 4 and 5 because p>0.05.