Regression_lab1

2229027 Jimin Lee

The dataset record.txt contains running records obtained from athletes from different countries in various types of athletics events (sprints and middle-distance). We have data about 55 countries (observations) and 6 records (variables): 100 meters, 200 meters, 400 meters, 800 meters, 1500 meters and 3000 meters.

Load the dataset record.txt in R, using the function read.table

```
# prepare data
setwd('/Users/jimin/Desktop/ㅈ ㅣㅁ ㅣㄴ/ewha/2023-2/Regression/')
record <- read.table("record.txt" ,head=TRUE)
head(record, n=10)
```

```
##           m100  m200  m400  m800 m1500 m3000
## argentin 11.61 22.94 54.50 129.0 265.8 587.4
## australi 11.20 22.35 51.08 118.8 247.8 544.8
## austria  11.43 23.09 50.62 119.4 253.2 560.4
## belgium  11.41 23.04 52.00 120.0 248.4 532.8
## bermuda  11.46 23.05 53.30 129.6 274.8 588.6
## brazil   11.31 23.17 52.80 126.0 269.4 586.2
## burma    12.14 24.47 55.00 130.8 267.0 570.6
## canada   11.00 22.25 50.06 120.0 243.6 528.6
## chile    12.00 24.52 54.90 123.0 253.8 562.2
## china    11.95 24.41 54.97 124.8 259.8 558.6
```

Produce summaries of the variable m800, including

Numerical summaries: average, standard deviation, median and quartiles, maximum and minimum, interquartile difference

```
# numerical summaries
print(summary(record$m800))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   113.4   120.0   123.0   124.6   129.0   139.8
```

```
cat("sd : ", sd(record$m800), "\n")
```

```
## sd :  6.493447
```

```
cat("IQR : ", IQR(record$m800))
```

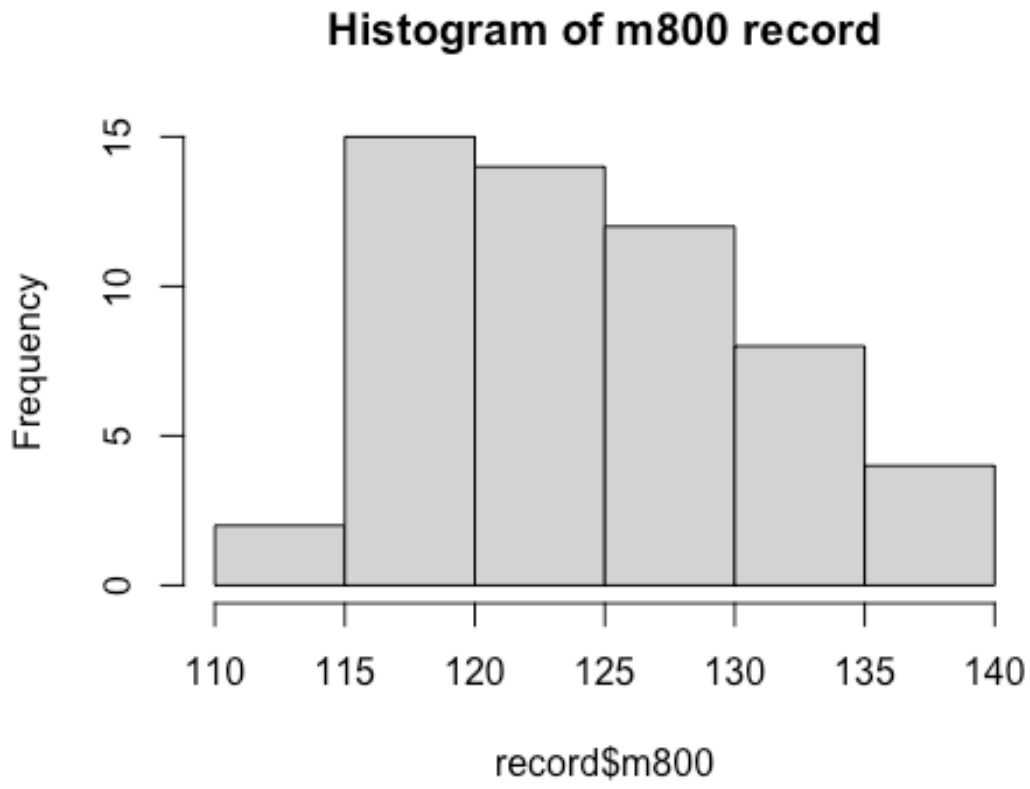
```
## IQR :  9
```

Graphical summaries: histogram and boxplot

What can you observe about the variable distribution?

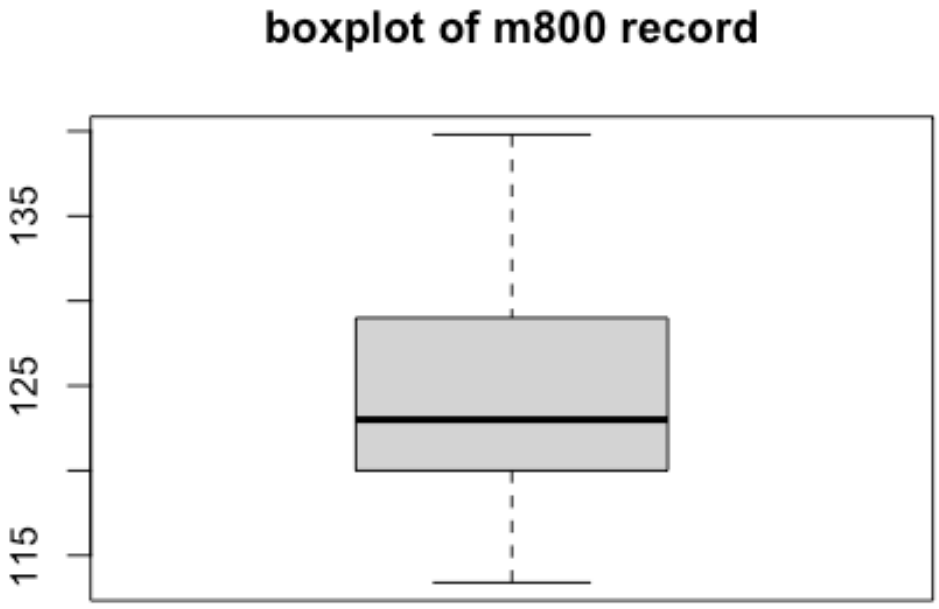
The distribution of m800 record is right-skewed.

```
# graphical summaries
hist(x=record$m800, main='Histogram of m800 record')
```




```
help(hist)

boxplot(record$m800, main='boxplot of m800 record')
```

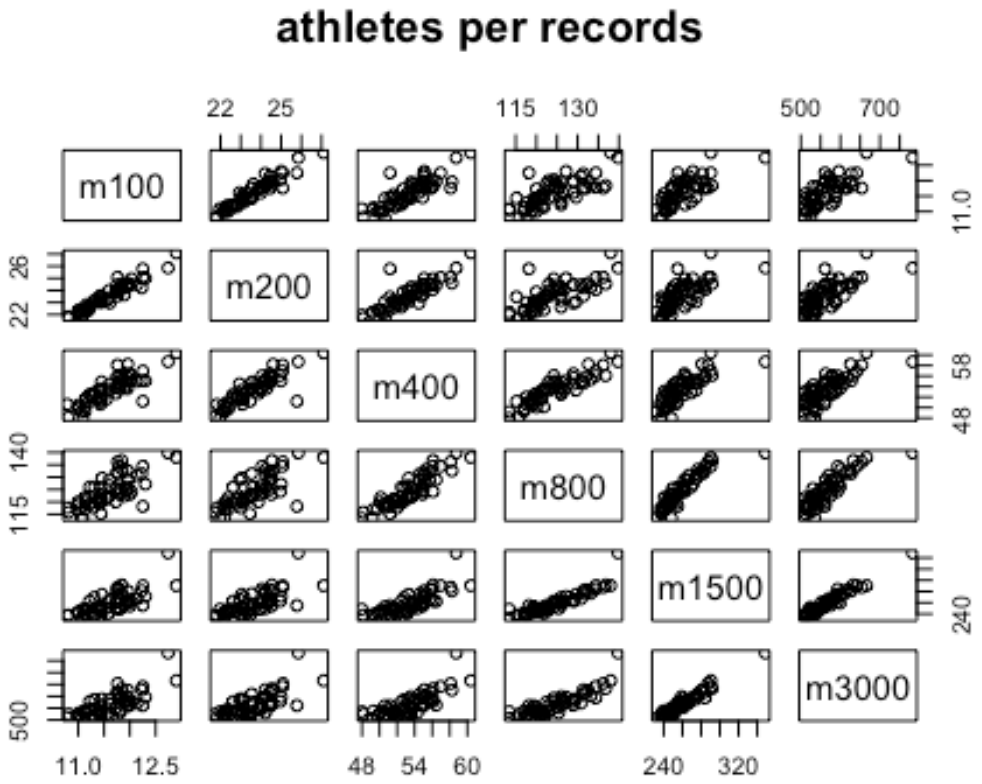


Produce scatter plot between all the variables(m100,m200,m400,m800,m1500,m3000). What can you observe from the scatter plot? Are they correlated?

When x increases, also y increase.

positive correlation.

```
pairs(record, main = "athletes per records") # print all scatters between col
s
```




```r
cor(record) # search the correlation between x,y
```

```
##            m100      m200      m400      m800     m1500     m3000
## m100  1.0000000 0.9527911 0.8346918 0.7276888 0.7283709 0.7416988
## m200  0.9527911 1.0000000 0.8569621 0.7240597 0.6983643 0.7098710
## m400  0.8346918 0.8569621 1.0000000 0.8984052 0.7878417 0.7776369
## m800  0.7276888 0.7240597 0.8984052 1.0000000 0.9016138 0.8635652
## m1500 0.7283709 0.6983643 0.7878417 0.9016138 1.0000000 0.9691690
## m3000 0.7416988 0.7098710 0.7776369 0.8635652 0.9691690 1.0000000
```