Homework 3 - Intro to R

Matt Isaac

September 18, 2017

```
athletes <-  read.table("athletes2016.txt")
```

1. Completed

```
cnames <-  c("Country", "NumAthletes", "NumGold", "NumSilver", "NumBronze", "NumTotal")
colnames(athletes, do.NULL = FALSE)
colnames(athletes) <- cnames
```

2. Completed

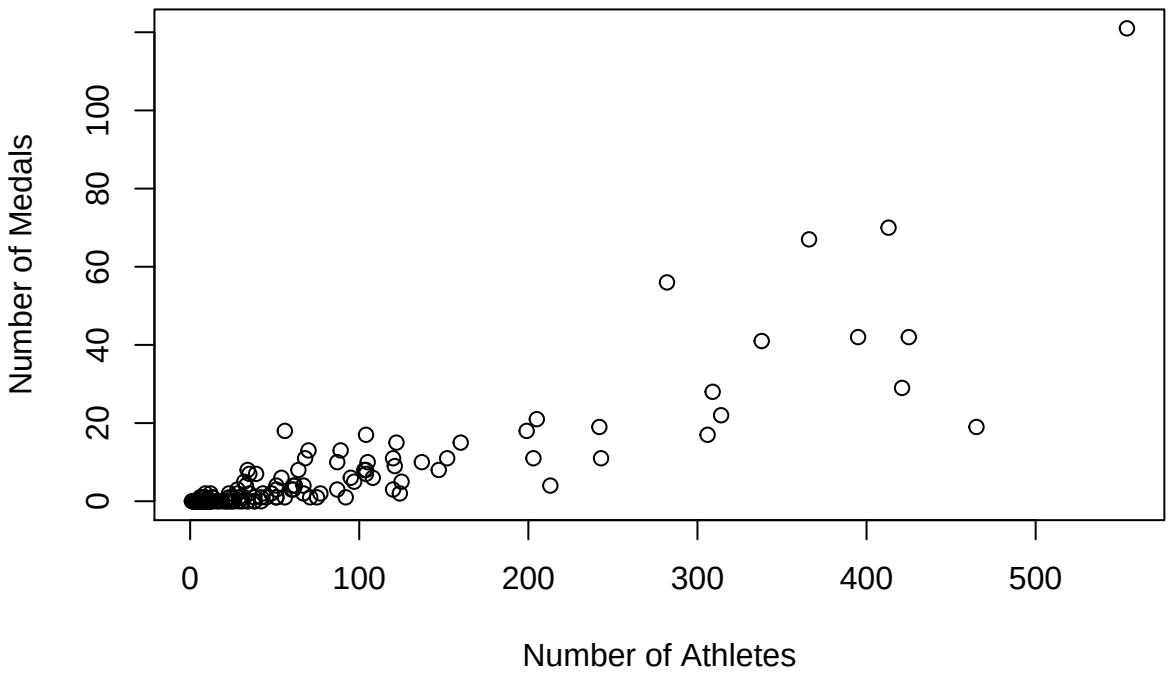
```
sorted_by_athletes <- athletes[order(-athletes$NumAthletes),]
most_athletes_15 <- sorted_by_athletes$Country[1:15]
```

3. The 15 countries that sent the most athletes to the 2016 Olympics are (in descending order): United_States, Brazil, Germany, Australia, China, France, Great_Britain, Japan, Canada, Italy, Spain, Russia, Poland, Netherlands, Argentina

```
sorted_by_medals <- athletes[order(-athletes$NumTotal),]
most_medals_15 <- sorted_by_medals$Country[1:15]
```

4. The 15 countries that won the most medals in the 2016 Olympics are (in descending order): United_States, China, Great_Britain, Russia, France, Germany, Japan, Australia, Italy, Canada, South_Korea, Brazil, Netherlands, Azerbaijan, New_Zealand

```
plot(athletes$NumAthletes, athletes$NumTotal, xlab = 'Number of Athletes',
     ylab = 'Number of Medals')
```



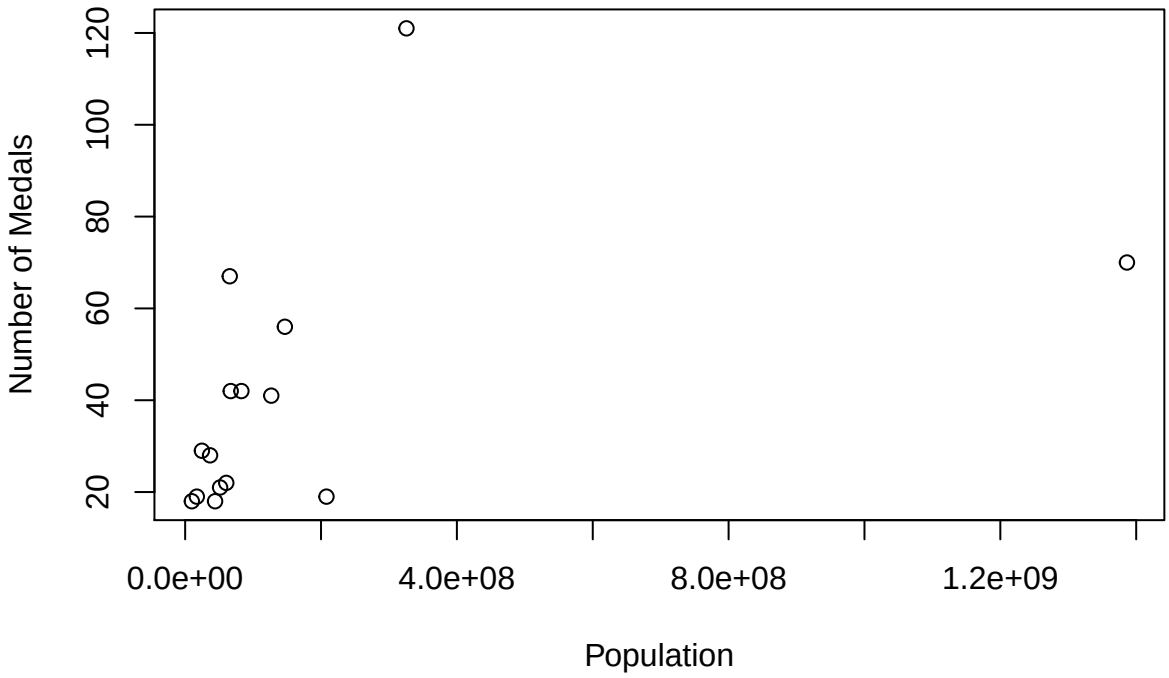

```
ath_cor = cor(athletes$NumAthletes, athletes$NumTotal)
```

5. There appears to be a positive linear relationship between the number of athletes sent by a country and the number of medals won by that country. As we look at the scatter plot, it is clear that (in general) countries with more athletes competing won more medals. Our observations from the plot are confirmed by the correlation coefficient. $r = 0.855$. This value of r suggests a moderately strong positive linear relationship.

```
population = c(325771000, 1386390000, 65648100, 146787329, 67080000,
               82800000, 126750000, 24654000, 36660000,60532325, 51446201,
               208003000, 17155200, 9843031, 44044811)
```

6. Completed

```
plot(population, sorted_by_medals$NumTotal[1:15], xlab = 'Population',
     ylab = 'Number of Medals')
```




```
pop_cor = cor(population, sorted_by_medals$NumTotal[1:15])
```

7. There again appears to be a positive linear relationship between the two variables in question. Less observations make our conclusion less concrete, but there is definitely a linear trend. As population increases, number of medals tends to increase. The plot conveys extremely important information in this case. If we were to look at the correlation coefficient alone, we might be misled. $r = 0.45$. This alone doesn't imply a very strong linear relationship, but we can see that there is one outlier that is bringing the correlation coefficient down.

```
ratios = sorted_by_medals$NumTotal[1:15]/population
ratios_1000000 = ratios * 1000000
```

8. The ratios are as follows. (Note - These ratios have been scaled up by a factor of 1,000,000 for ease in comparison)

Country	Medals to Population * 1,000,000
United States	0.371
China	0.05
Great Britain	1.021
Russia	0.382
France	0.626
Germany	0.507

Country	Medals to Population * 1,000,000
Japan	0.323
Australia	1.176
Italy	0.764
Canada	0.363
South Korea	0.408
Brazil	0.091
Netherlands	1.108
Azerbaijan	1.829
New Zeland	0.409

When we take the ratio, we are in a sense "controlling" for population size. We find that if we take population size into account, the countries with larger populations actually got less medals per population member than the small countries. China (one of the largest populations) had the smallest ratio. And Azerbaijan (one of the smallest populations) had one of the largest ratios. When we analyze the data this way, the correlation isn't as strong, and isn't what I expected.

9. These data tell us that the countries with larger populations generally have more athletes. The more athletes a country has competing, the more medals that country is likely to win. However, we also learned that if we are actually going to do a side-by-side comparison, it would be wise to control for the variability in the number of athletes competing. For having so few athletes, small countries like Azerbaijan and the Netherlands did quite well. By the same token, China and the United States did receive a lot of medals, but this achievement should be recognized with the fact in mind that the large number of competing athletes does give these countries an advantage.