HW4

Calvin Makelky, A01872013

September 9, 2016

1

```
olympics = read.table("athletes2016.txt")
```

2

```
cnames = c("Country", "Athletes", "Golds", "Silvers", "Bronzes", "Medals")
colnames(olympics) = cnames
names(olympics)
```

```
## [1] "Country"  "Athletes" "Golds"    "Silvers"  "Bronzes"  "Medals"
```

3

```
ath_count = olympics[c("Country","Athletes")]
ath_count = ath_count[order(ath_count$Athletes, decreasing=TRUE),]
head(ath_count, 15)
```

```
##           Country Athletes
## 198 United_States      554
## 27         Brazil      465
## 71        Germany      425
## 11      Australia      421
## 41          China      413
## 67         France      395
## 73  Great_Britain      366
## 96          Japan      338
## 35         Canada      314
## 93          Italy      309
## 176         Spain      306
## 155        Russia      282
## 149        Poland      243
## 133   Netherlands      242
## 8       Argentina      213
```

4

```r
medal_count = olympics[c("Country", "Medals")]
medal_count = medal_count[order(medal_count$Medals, decreasing=TRUE),]
head(medal_count, 15)
```

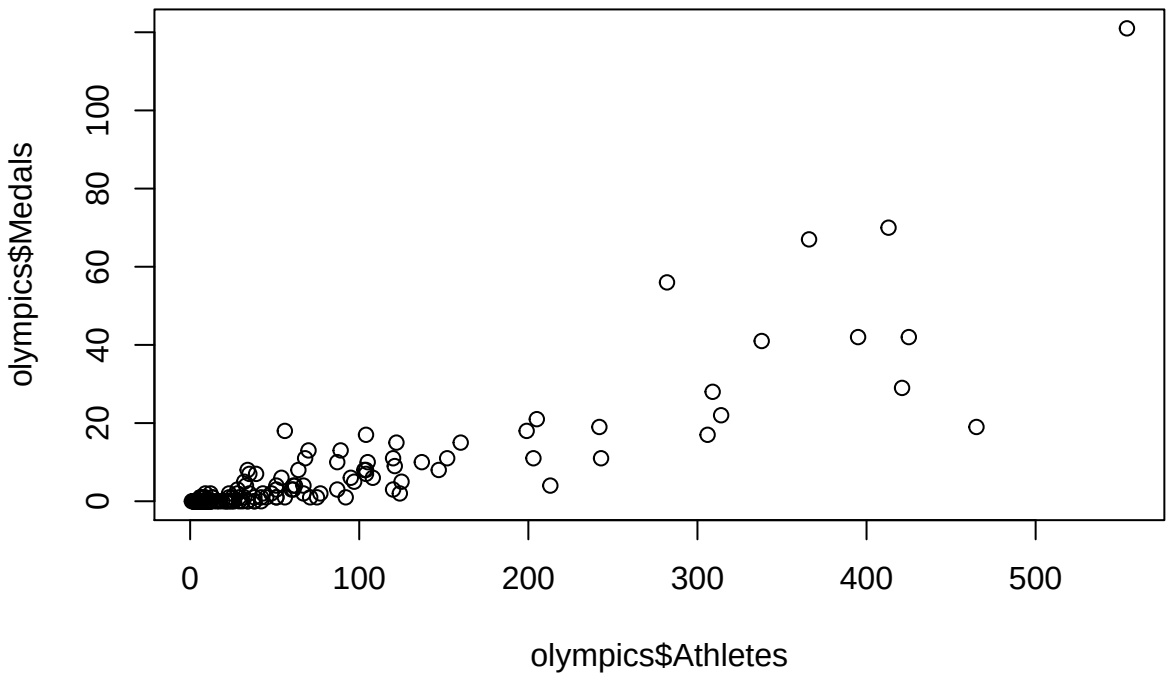
```
##           Country Medals
## 198 United_States    121
## 41          China     70
## 73  Great_Britain     67
## 155        Russia     56
## 67         France     42
## 71        Germany     42
## 96          Japan     41
## 11      Australia     29
## 93          Italy     28
## 35         Canada     22
## 174   South_Korea     21
## 27         Brazil     19
## 133   Netherlands     19
## 13     Azerbaijan     18
## 134   New_Zealand     18
```

5

```r
cor(olympics$Athletes, olympics$Medals)
```

```
## [1] 0.855311
```

```r
plot(olympics$Athletes, olympics$Medals)
```



The correlation between medals and the number of athletes a country has is quite high judging by the high r correlation coefficient. Countries with more athletes tend to win more medals.

6

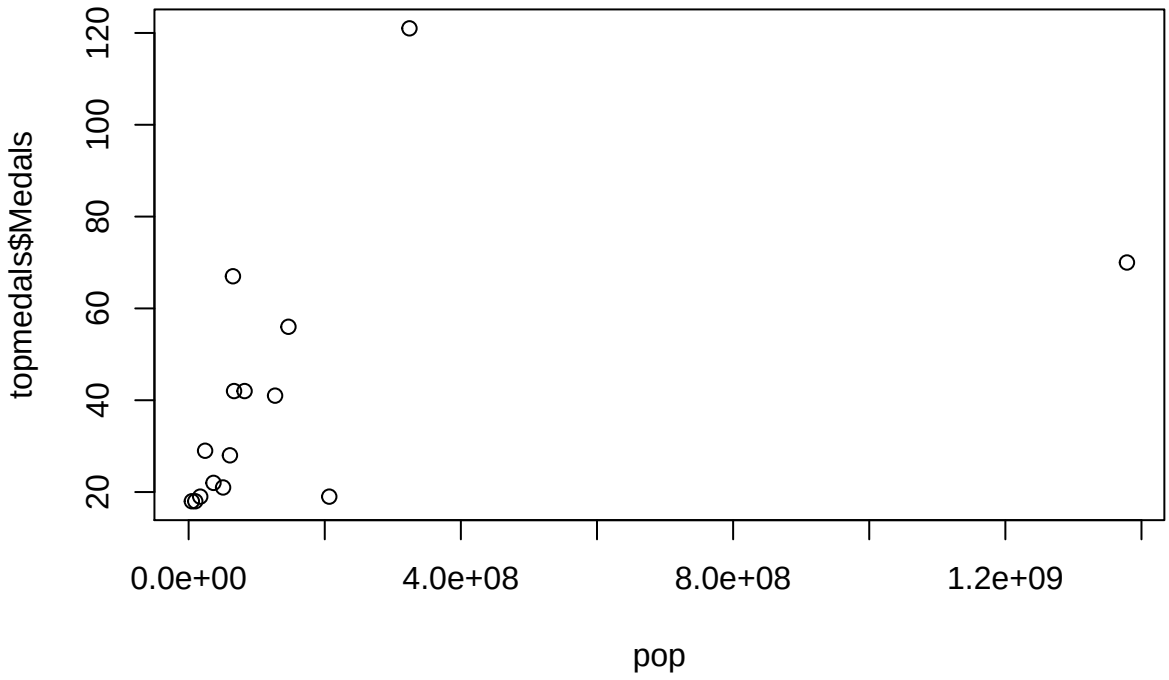
```
pop = c(324437000, 1378650000, 65110000, 146654366,
66736000, 82175700, 127000000, 24183100, 60665551, 36521200,
50617045, 206630000, 17047600, 9755500, 4713310)
```

7

```
topmedals = head(medal_count, 15)
cor(topmedals$Medals, pop)
```

```
## [1] 0.4560032
```

```
plot(pop, topmedals$Medals)
```



The correlation between medals and the country's population is not high at all based off the low r correlation coefficient. There is not strong evidence that countries with more people win more medals because of the population.

8

```
topmedals$pop = pop
topmedals$medal_ratio = topmedals$pop / topmedals$Medals
topmedals = topmedals[order(topmedals$medal_ratio, decreasing=TRUE),]

#best medal to population ratio is New Zealand:
tail(topmedals, 1)
```

```
##         Country Medals     pop medal_ratio
## 134 New_Zealand     18 4713310    261850.6
```

```
#worst medal to population ratio is China:
head(topmedals, 1)
```

```
##    Country Medals        pop medal_ratio
## 41   China     70 1378650000    19695000
```

9

I learned how to read data into a data frame, how to name columns of a data frame, how to order a data frame by a column, how to get the correlation between two columns in a data frame, how to create a vector, how to creae a new column in a data frame, and how to create a new column in a data frame using two columns within the data frame.

From the data I learned that there is a strong correlation between the number of athletes and the number of medals, not a strong correlation between a country's population and its amount of medals, and that of the 15 highest medaling countries in 2016 that China had the worst medal to population ratio while New Zealand had the best.