Stats 10 Lab 1 Submission
Name: Enmin Zhou
UID: 104756697

**Exercise 1**

1) a)

```
> flint <- read.csv("~/Desktop/stats10/lab2/flint.csv")
```

b)

```
> mean(flint$Pb >= 15)
[1] 0.04436229
```

c)

```
> region <- flint$Region
> copper <- flint$Cu
> is_north <- region == "North"
> north_cu <- copper[is_north]
> mean(north_cu)
[1] 44.6424
```

d)

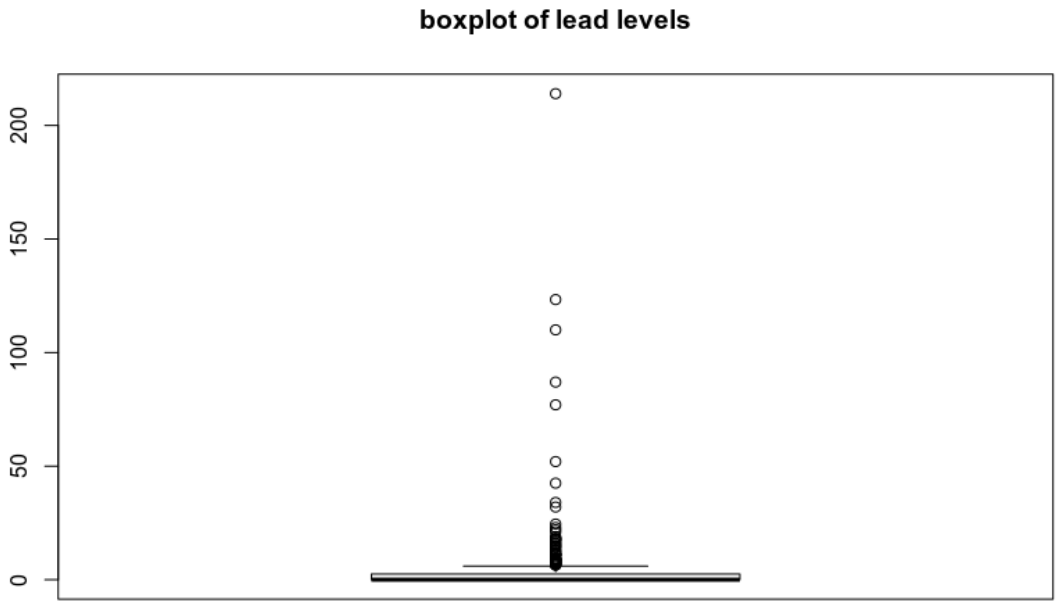
```
> pb <- flint$Pb
> is_danger <- pb >= 15
> danger_cu <- copper[is_danger]
> mean(danger_cu)
[1] 305.8333
```

e)

```
> mean(flint$Cu)
[1] 54.58102
> mean(flint$Pb)
[1] 3.383272
```

f)

```
> boxplot(flint$Pb, main="mean of lead levels")
```

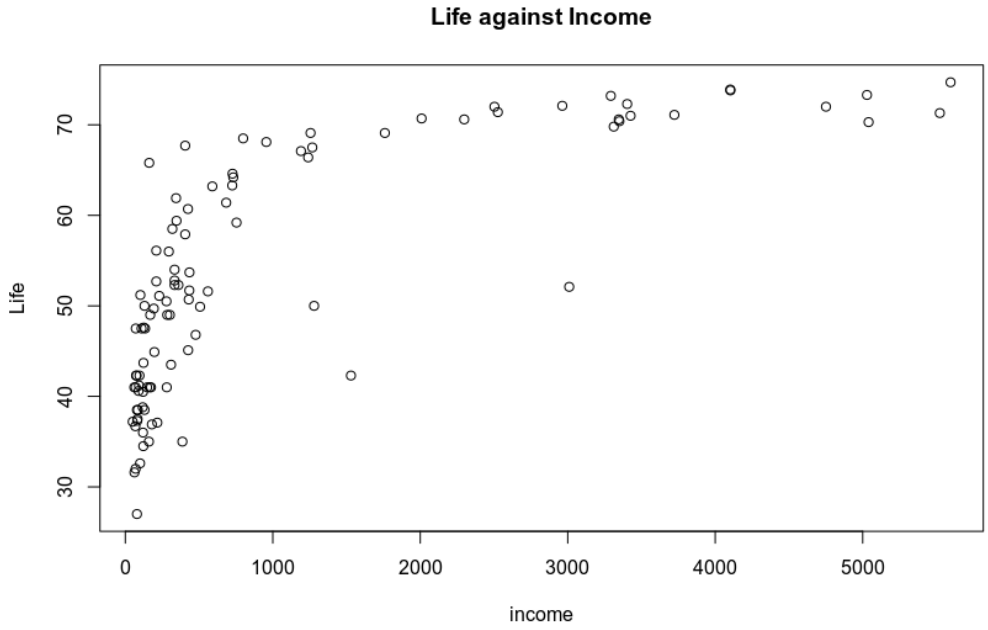


g)No.  I think the median is 0. The IQR is 2.5. The Q1 is 0 and Q3 is 2.5.

**Exercise 2**

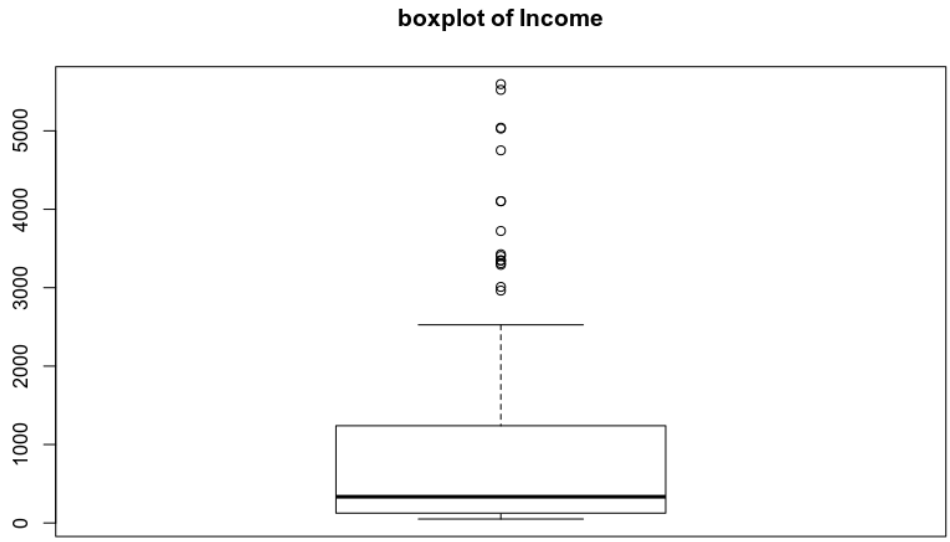
2) a)

```
> life<-read.table("http://www.stat.ucla.edu/~nchristo/statistics12/countries_life.txt", header=TRUE)
> plot(life$Income, life$Life, xlab="income", ylab="Life", main="Life against Income")
```

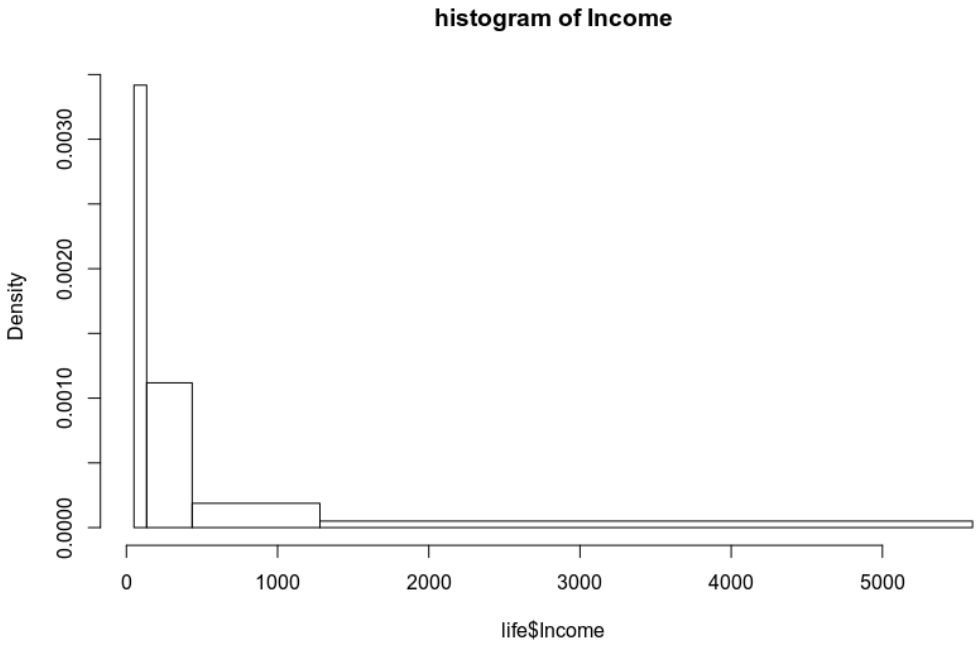



b)

```
> boxplot(life$Income, main="boxplot of Income")
```




```
> histogram::histogram(life$Income, main="histogram of Income")
```

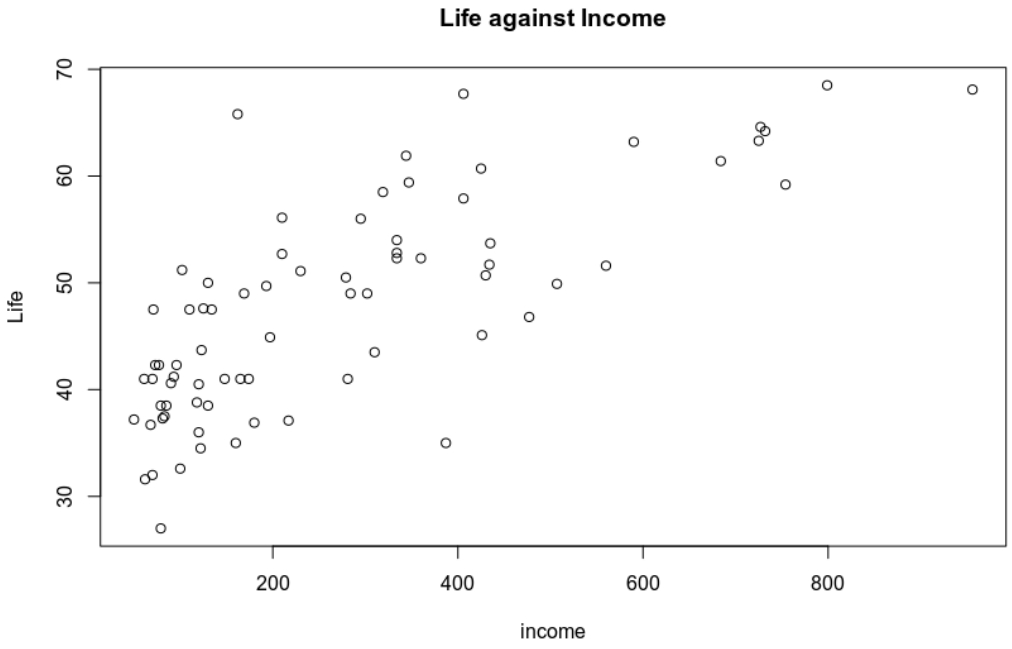



c)

```
> income <- life$Income
> is_above <- income >= 1000
> is_below <- income < 1000
> above_1000 <- life[is_above,]
> below_1000 <- life[is_below,]
```

d)

```
> plot(below_1000$Income, below_1000$Life, xlab="income", ylab="Life", main="Life against Income")
```




```
> cor(below_1000$Income, below_1000$Life)
```

```
[1] 0.752886
```

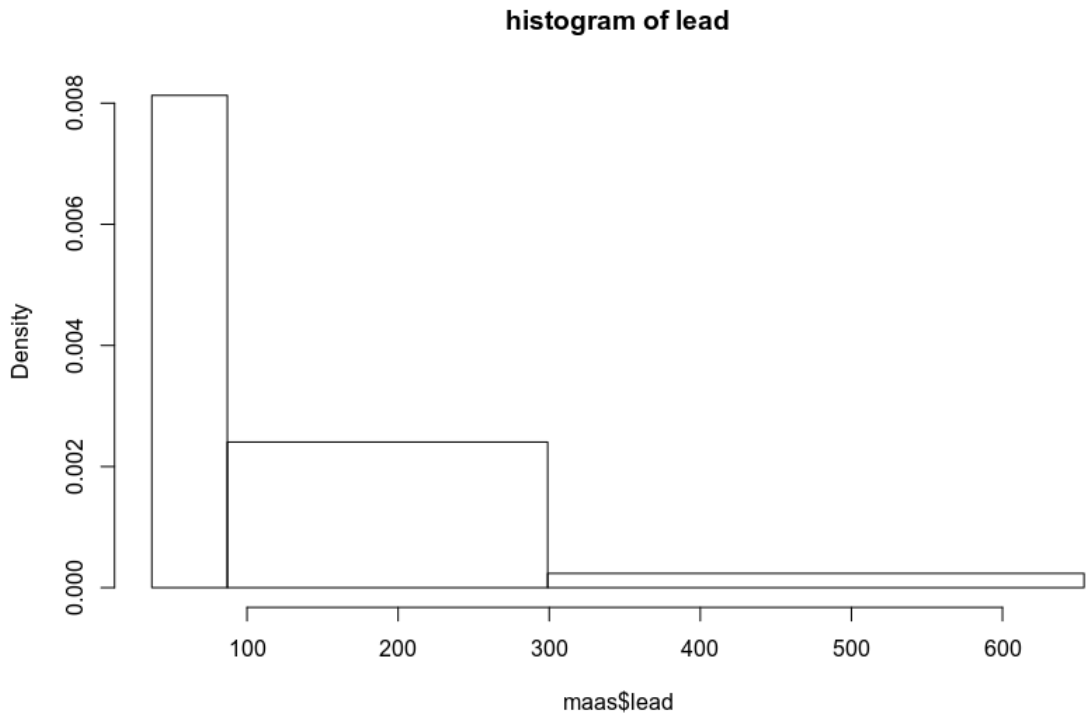
## Exercise 3

### 3) a)

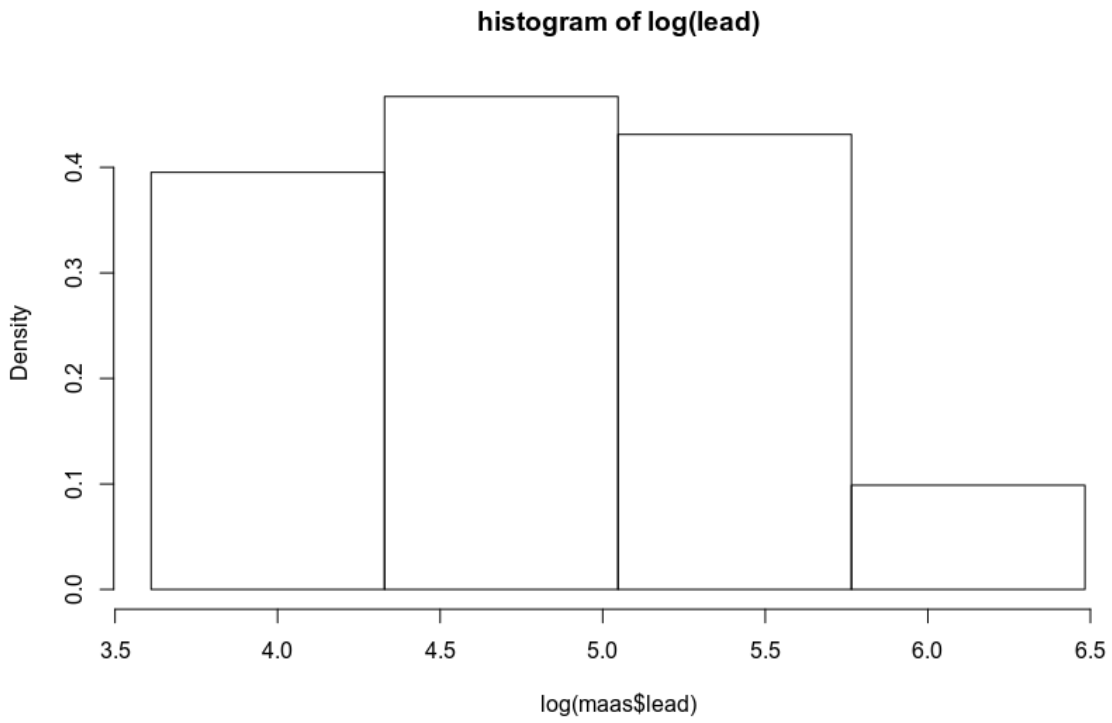
```
> maas<-read.table("http://www.stat.ucla.edu/~nchristo/statistics12/soil.txt", header=TRUE)
> summary(maas$lead)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    37.0    72.5   123.0   153.4   207.0   654.0
> summary(maas$zinc)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   113.0   198.0   326.0   469.7   674.5  1839.0
```

### b)

```
> histogram(maas$lead, main="histogram of lead")
```

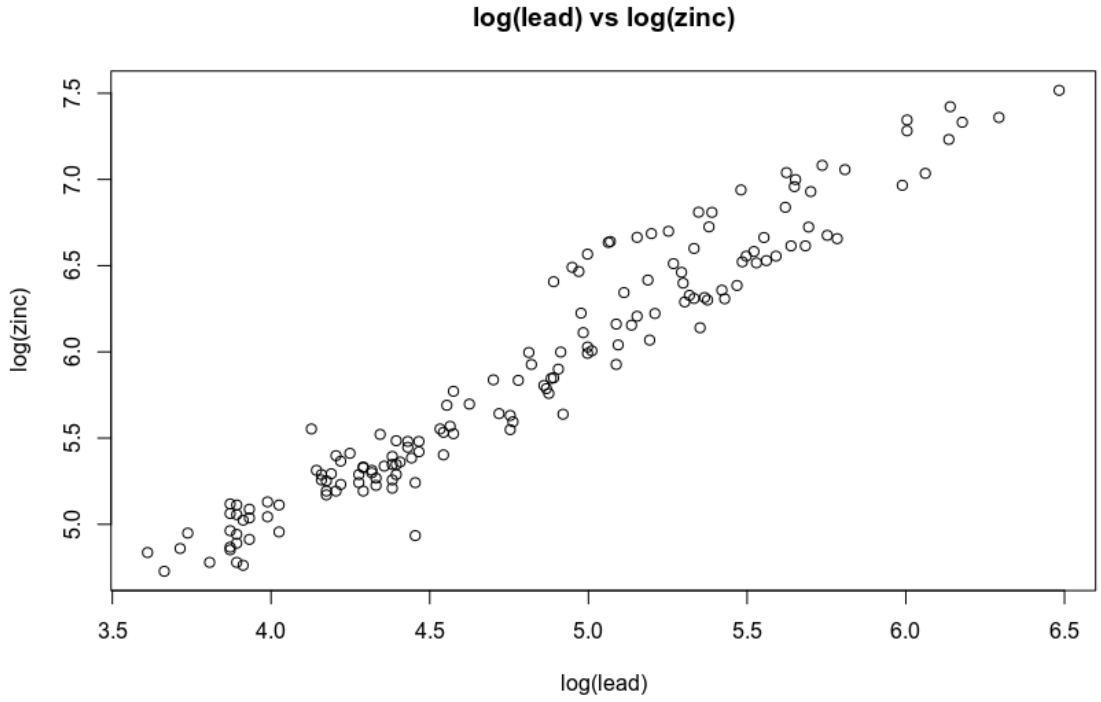



```
> histogram(log(maas$lead), main="histogram of log(lead)")
```




c)

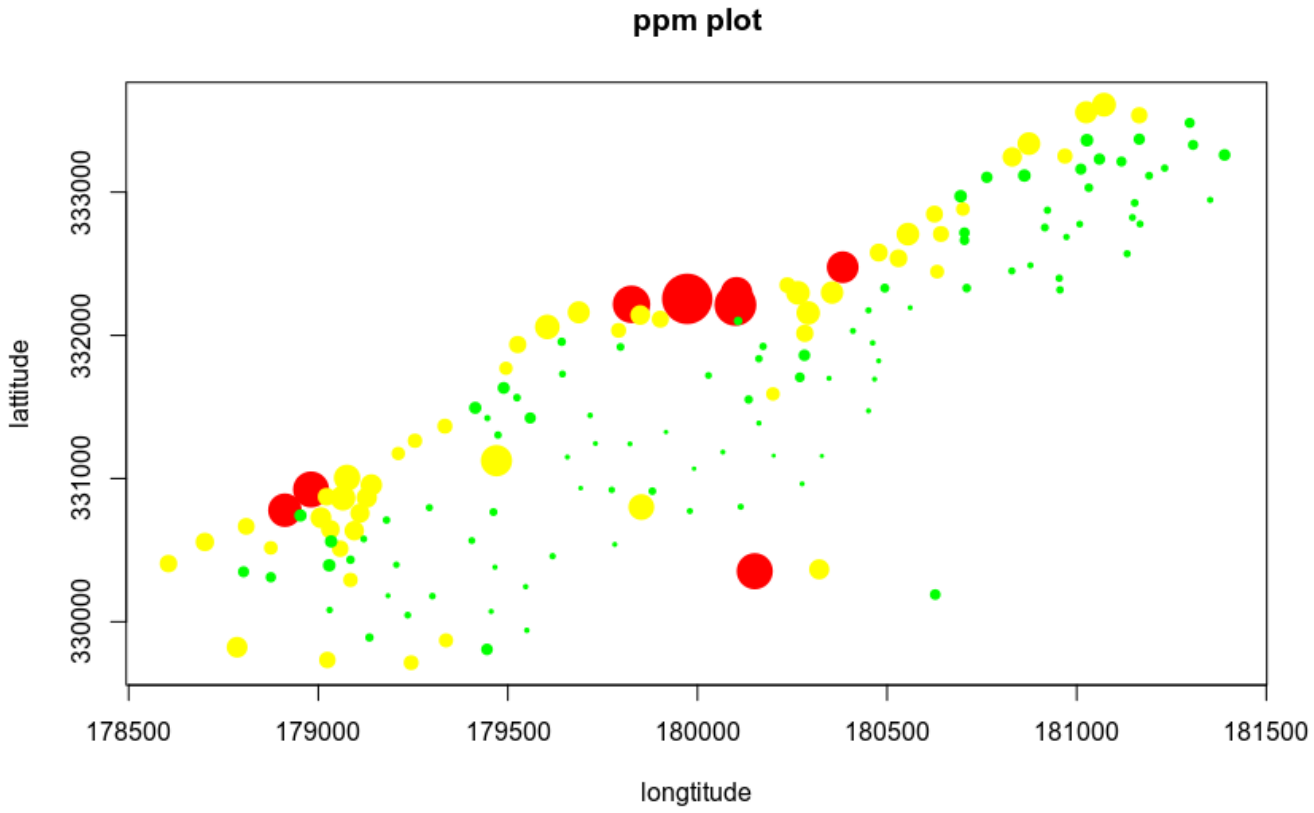
```
> plot(log(maas$lead), log(maas$zinc), xlab="log(lead)", ylab="log(zinc)", main="log(lead) vs
log(zinc)")
```



From the plot, I observe that there is a high positive correlation between log(lead) and log(zinc).

d)

```
> ppm_colors <- c("green", "yellow", "red")
> ppm_levels <- cut(maas$lead, c(0,150,400, 1000))
> plot(maas$x, maas$y,xlab="longtitude", ylab="lattitude", main="ppm plot", "n")
> points(maas$x, maas$y, cex=maas$lead/mean(maas$lead), col=ppm_colors[as.numeric(ppm_levels)],
pch=19)
```



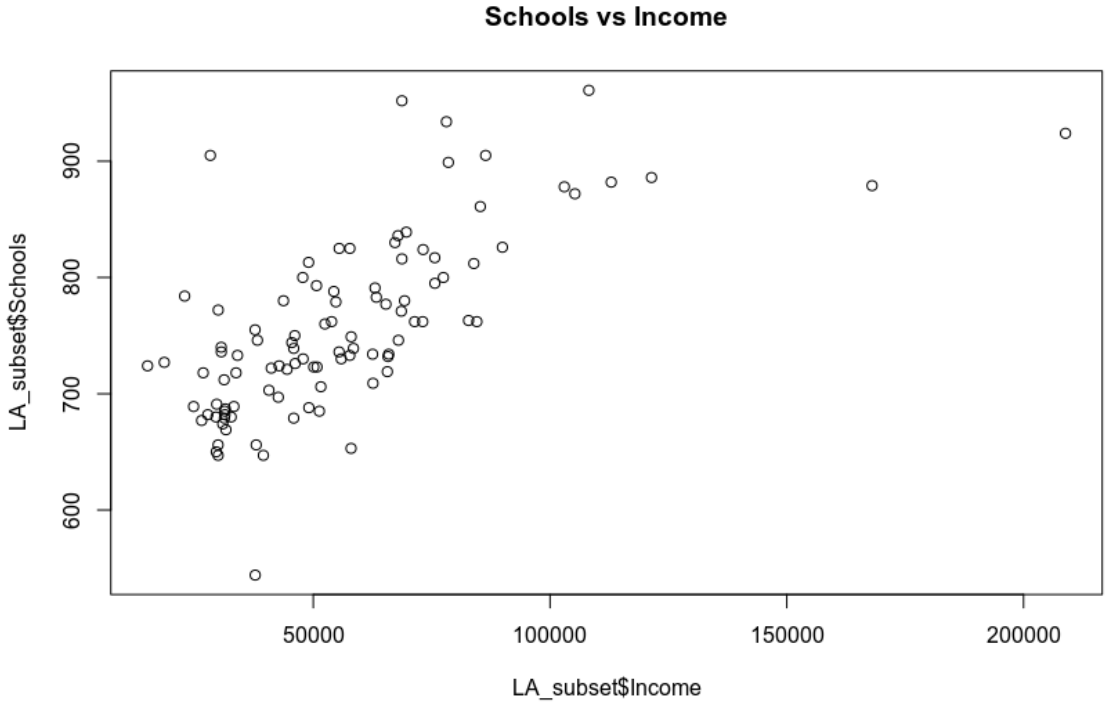

Exercise 4

4) a)

```
> LA <-read.table("http://www.stat.ucla.edu/~nchristo/statistics12/la_data.txt", header=TRUE)
> plot(LA$Longitude, LA$Latitude, xlim=c(-119,-118), ylim=c(33.5,35), xlab="longtitude",
ylab="lattitude", main="LA plot", "n")
> map("county", "california", add = TRUE)
```


b)

```
> zero_school <- LA$Schools != 0
> LA_subset <- LA[zero_school,]
> plot(LA_subset$Income, LA_subset$Schools, main="Schools vs Income")
```




There is a weak positive correlation between the income and schools in LA county.