Stats 10 Lab Submission
Name: Preyasi Gaur
Section: 4A

Exercise 1

a) flint <- read.csv('/Users/preyasigaur/Desktop/flint.csv')

b) Input: mean(flint$Pb >= 15)
Output:

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

c) Input: mean(flint$Cu[flint$Region == "North"])
Output:

```
> mean(flint$Cu[flint$Region == "North"])
[1] 44.6424
```

d) Input: mean(flint$Cu[flint$Pb >= 15])

```
> mean(flint$Cu[flint$Pb >= 15])
[1] 305.8333
```

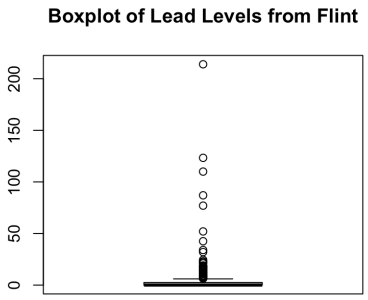
e) Input: mean(flint$Pb)
Output:

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

Input: mean(flint$Cu)
Output:

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

f) Input: boxplot(flint$Pb, main = "Boxplot of Lead Levels from Flint")
Output:

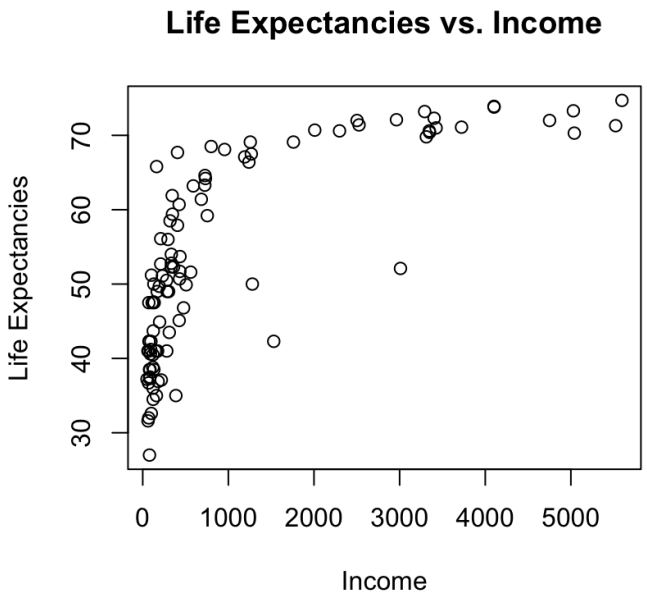


g) No, the mean does not seem to be a good measure for the center of the data. The data is skewed, and thus in this case if skewed datasets the median is a better measure of central tendency.
median(flint$Pb)

```
> median(flint$Pb)
[1] 0
```

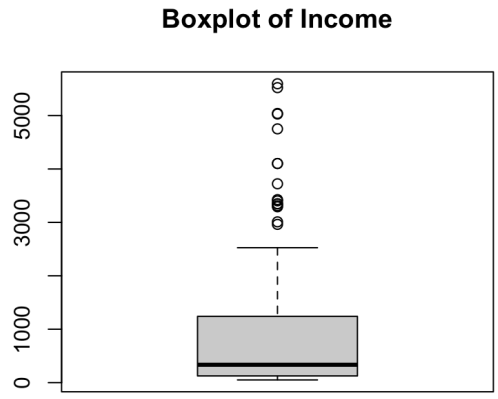
Exercise 2

a) Input: plot (life$Life ~ life$Income, xlab = "Income", ylab = "Life Expectancies", main = "Life Expectancies vs. Income")
Output:

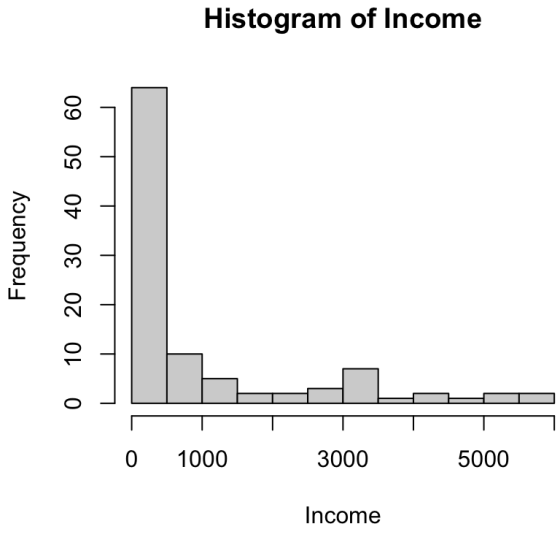


When income is low(0 - 1000), there is a strong positive correlation between the Life Expectancy and Income. When the income is high, the positive correlation becomes weaker. Also, since it is a scatter plot, we cannot identify any causal relationship between them.

b) Input: boxplot(life$Income, main = "Boxplot of Income")
Output:



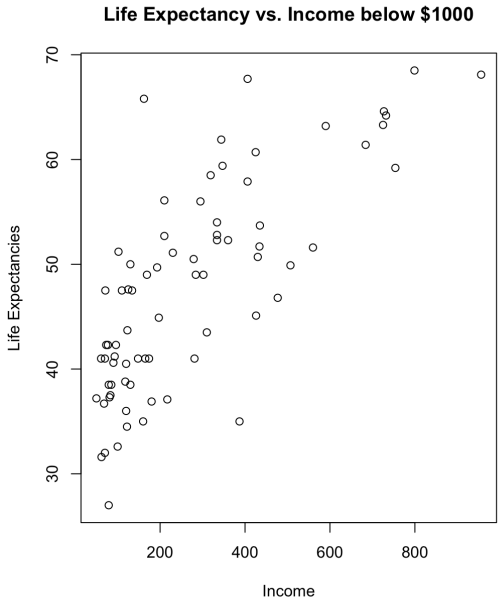
Input: hist(life$Income, xlab = "Income", main = "Histogram of Income")
Output:



Yes, there are outliers present, as is clear from the histogram and the boxplot.

c) Input: income_below_1000 <- life[life$Income < 1000,]
income_above_1000 <- life[life$Income >= 1000,]

d) Input: plot(income_below_1000$Life ~ income_below_1000$Income, xlab = "Income", ylab = "Life Expectancies", main = "Life Expectancy vs. Income below $1000")
Output:



e) Input: cor(income_below_1000$Life, income_below_1000$Income)
Output:

```
> cor(income_below_1000$Life, income_below_1000$Income)
[1] 0.752886
```

Exercise 3

maas <- read.table("http://www.stat.ucla.edu/~nchristo/statistics12/soil.txt", header = TRUE)

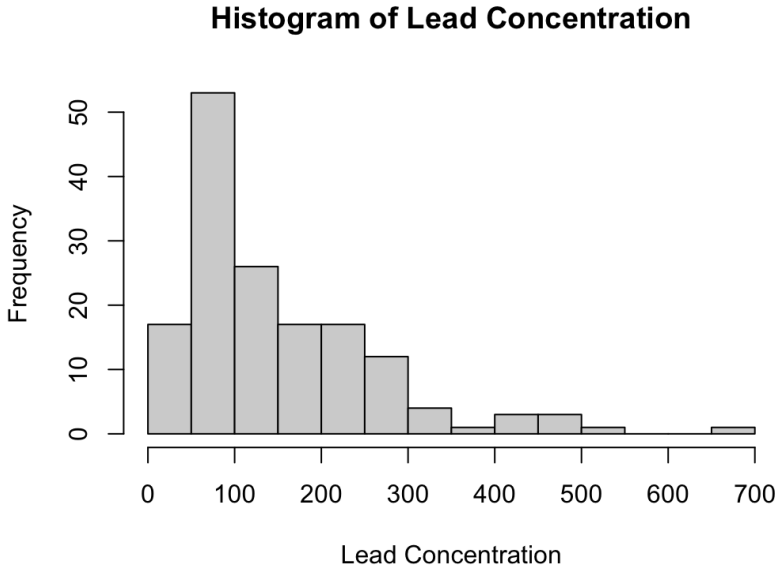
a) Input: summary(maas$lead)
Output:

```
summary(maas$lead)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 37.0    72.5   123.0   153.4   207.0   654.0
```

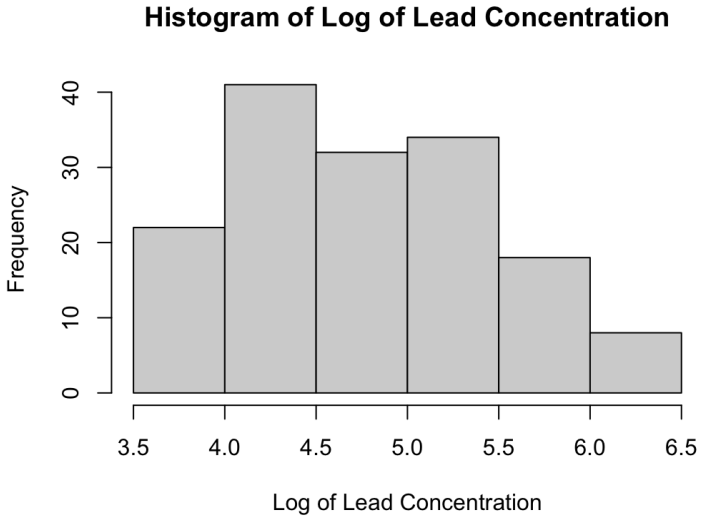
Input: summary(maas$zinc)
Output:

```
summary(maas$zinc)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
113.0   198.0   326.0   469.7   674.5  1839.0
```

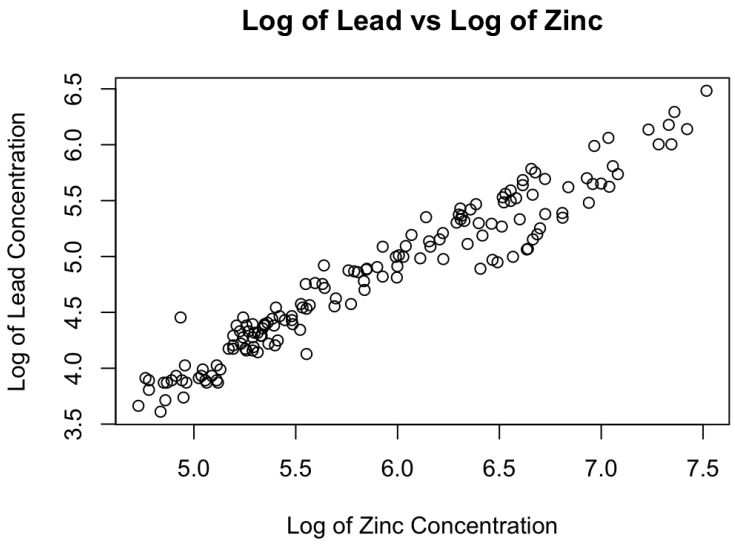
b) Input: hist(maas$lead, xlab = "Lead Concentration", main = "Histogram of Lead Concentration")
Output:



Input: hist(log(maas$lead), xlab = "Log of Lead Concentration", main = "Histogram of Log of Lead Concentration")
Output:

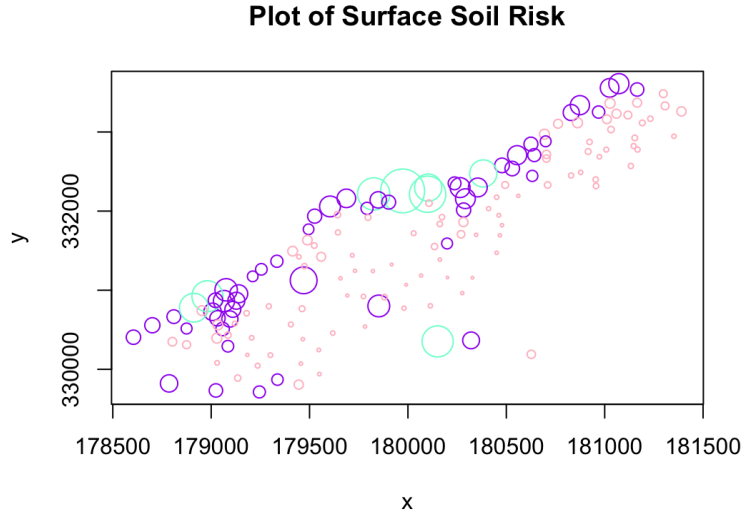


c) plot(log(maas$lead) ~ log(maas$zinc),
xlab = "Log of Zinc Concentration",
ylab = "Log of Lead Concentration",
main = "Log of Lead vs Log of Zinc")



There is a strong positive linear trend between the two variables - the log of load and the log of zinc.

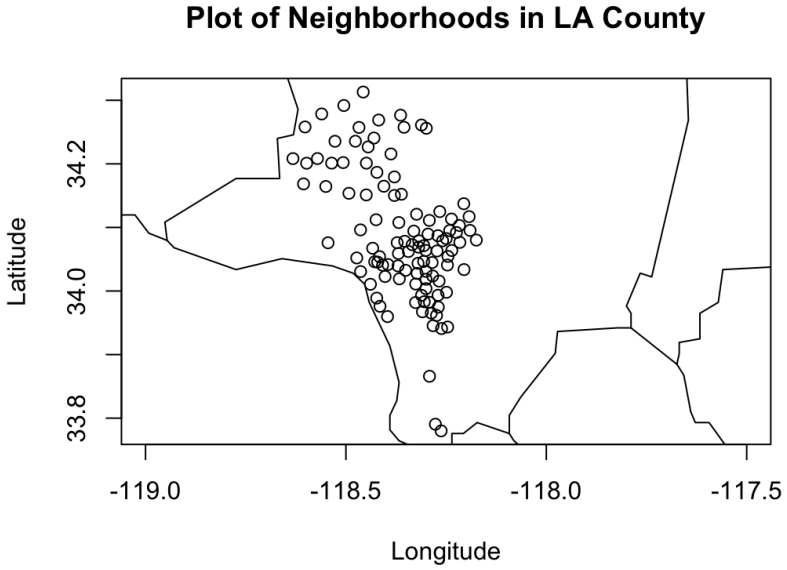
d) plot(maas$x, maas$y, xlab ="x", ylab = "y", main = "Plot of Surface Soil Risk", "n")
points(maas$x, maas$y, col = soil_color[as.numeric(soil_level)], cex = maas$lead / mean(maas$lead))



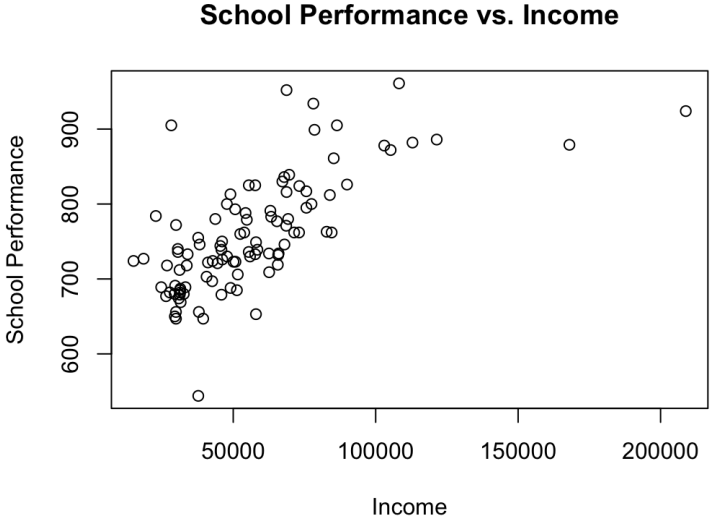
Exercise 4

LA <- read.table("http://www.stat.ucla.edu/~nchristo/statistics12/la_data.txt", header = TRUE)

a) library(maps)
plot(LA$Longitude, LA$Latitude, xlab = "Longitude", ylab = "Latitude", main = "Plot of Neighborhoods in LA County", xlim = c(-119, -117.5))
map("county", "california", add = TRUE)



b) We can see the correlation between Income and School Performance by plotting them in a scatter plot.



As is clear, there is a moderately positive linear relationship between income and school performance. But there are some outliers too.