Problem Set 1

ENGRD 2700

Cort Breuer

09/08/19

Question 1

```
quartet <- read.csv("Data/Quartet.csv")
```

Part A

X1 Sample Mean $= \frac{10+8+13+9+11+14+6+4+12+7+5}{11} = 9$

X1 Sample Median $= 4, 5, 6, 7, 8, \mathbf{9}, 10, 11, 12, 13, 14 = 9$

X1 Sample Standard Deviation $= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{10}[(4-9)^2 + (5-9)^2 + ...(13-9)^2 + (14-9)^2]} = \sqrt{11} = 3.32$

```
colNames <- c("X1", "X2", "X3", "X4", "Y1", "Y2", "Y3", "Y4")

quartetMean <- apply(quartet, 2, mean)
quartetMedian <- apply(quartet, 2, median)
quartetSD <- apply(quartet, 2, sd)

summaryTable <- tibble(colNames, quartetMean, quartetMedian, quartetSD) %>% rename(Column = colNames, "

kable(summaryTable) %>% kable_styling(bootstrap_options = c("striped", "hover"))
```

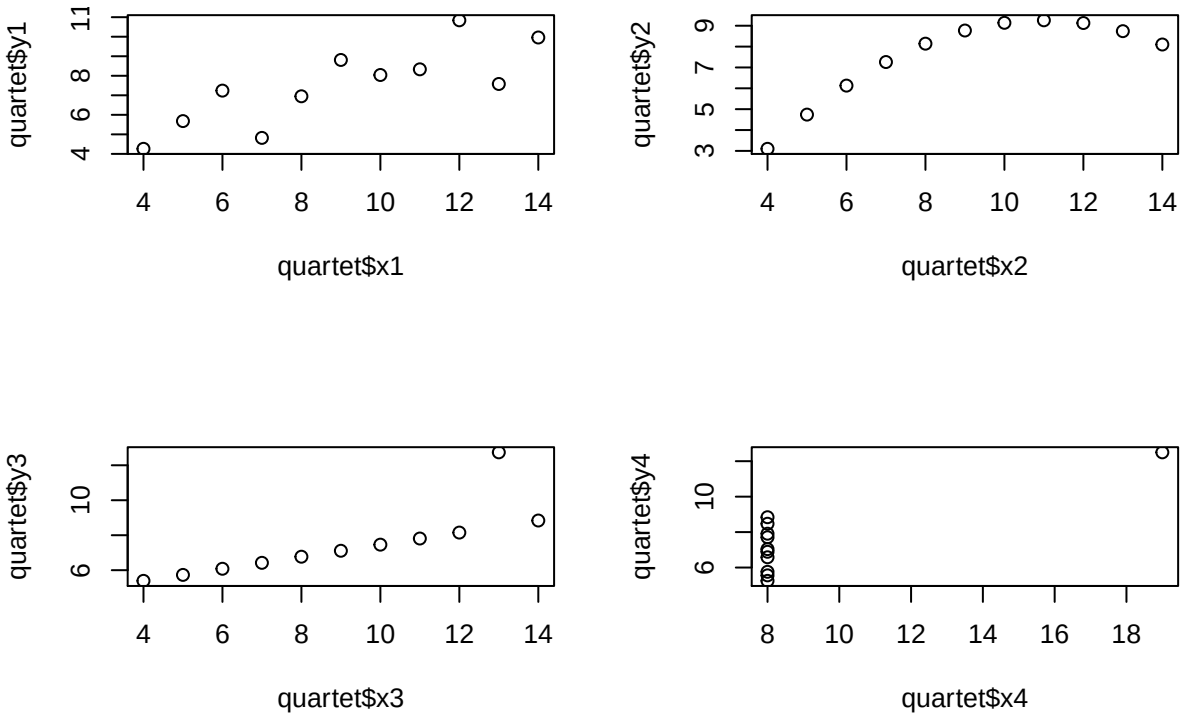
Column	Sample Mean	Sample Median	Sample Standard Deviation
X1	9.0	9.00	3.32
X2	7.5	7.58	2.03
X3	9.0	9.00	3.32
X4	7.5	8.14	2.03
Y1	9.0	9.00	3.32
Y2	7.5	7.11	2.03
Y3	9.0	8.00	3.32
Y4	7.5	7.04	2.03

Part B

Part C

```
par(mfrow=c(2,2))

plot(quartet$x1, quartet$y1)
plot(quartet$x2, quartet$y2)
plot(quartet$x3, quartet$y3)
plot(quartet$x4, quartet$y4)
```



Part D

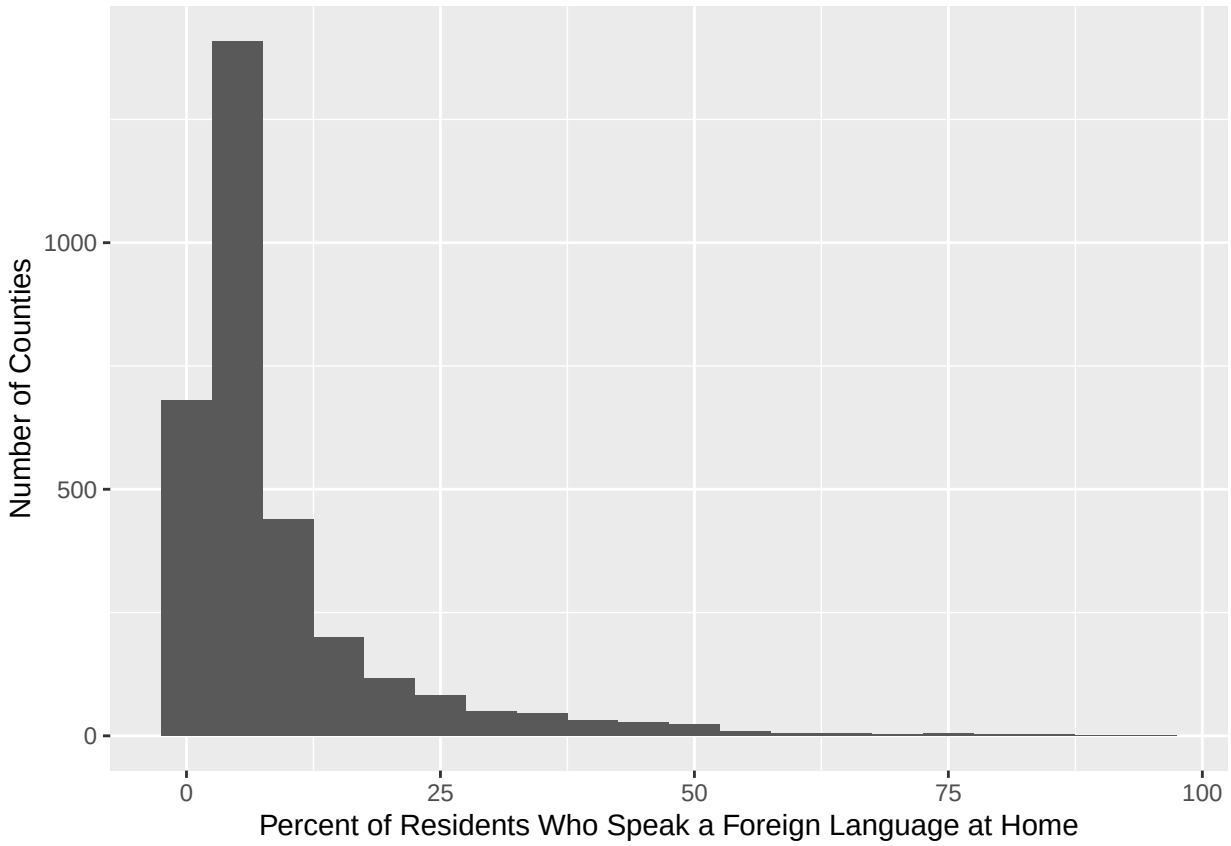
Part E

Question 2

```
countyData <- read.csv("Data/CountyData.csv")
countyData <- as_tibble(countyData)
```

Part A

```
ggplot(data = countyData) + geom_histogram(mapping = aes(foreign_spoken_at_home), binwidth = 5) + labs(
```



Part B

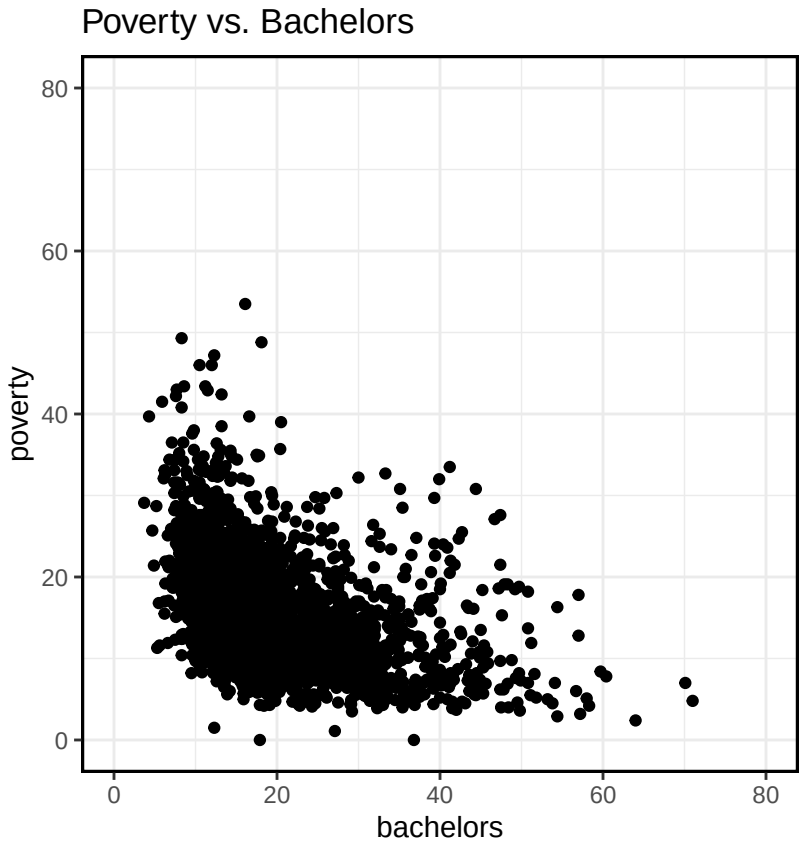
```
median(countyData$fed_spending, na.rm = TRUE)
```

```
## [1] 214994
```

Part C

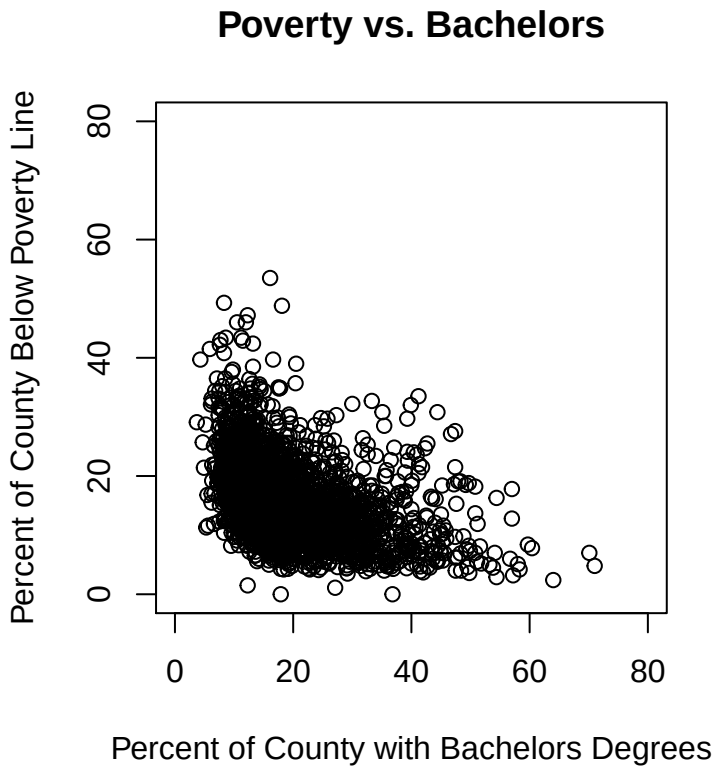
```
theme_set(theme_bw())

ggplot(data = countyData) + geom_point(mapping = aes(x = bachelors, y = poverty)) + xlim(0, 80) + ylim(
```




```
par(pty = "s")

plot(countyData$bachelors, countyData$poverty, xlim = c(0, 80), ylim = c(0, 80), main = "Poverty vs. Ba
```



Part D

Question 3

Part A

Part B

Part C

Question 4

Part A

Part B

$$z_i = \frac{x_i - \overline{x}}{s_x}$$