Stat 140 - Quiz 2 Sample

What's Your Name? _____

Which section are you in? _____

This is a sample quiz. For the real quiz, I will use a different data set, but will pick roughly 2-3 of the questions that are below and adapt them to the new data set with minimal modification.

Below are the first few rows of a data frame named `NHANES`. NHANES stands for "National Health and Nutrition Examination Surveys", and the data frame contains information about the health of randomly sampled Americans.

```
##       ID Gender Age MaritalStatus Poverty Weight Height Testosterone
## 1 62163   male  14          <NA>    4.07   49.4  168.9       274.95
## 2 62172 female  43  NeverMarried    2.02   98.6  172.0        47.53
## 3 62174   male  80       Married    4.30   95.8  168.1       642.82
## 4 62174   male  80       Married    4.30   95.8  168.1       642.82
## 5 62176 female  34       Married    5.00   68.7  171.6        21.11
## 6 62178   male  80       Widowed    0.05   85.9  173.5       562.78
##   BPDiaAve
## 1       37
## 2       72
## 3       39
## 4       39
## 5       69
## 6       72
```

1. Poverty and Marital Status

The `Poverty` variable in this data set is described as follows: "A ratio of family income to poverty guidelines. Smaller numbers indicate more poverty."

Here are some side-by-side box plots summarizing the distribution of this poverty index within each level of the `MaritalStatus` variable.

```
ggplot() +
  geom_boxplot(data = NHANES, mapping = aes(x = MaritalStatus, y = Poverty))
```
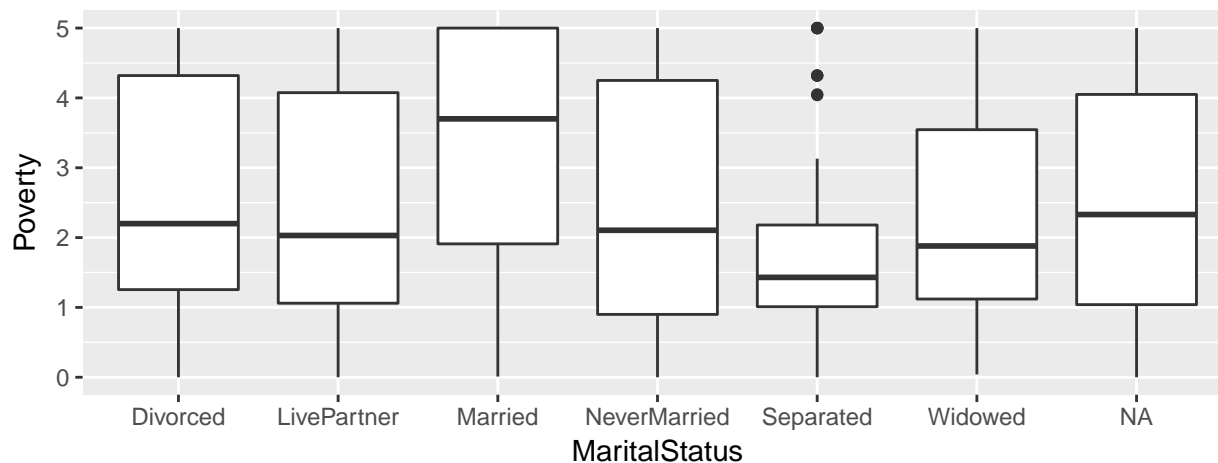
a. For which marital status is the median of the poverty index lowest (indicating more poverty).

b. The median poverty index for individuals in the "NeverMarried" group is about **2**. What is the interpretation of this number, in terms of the proportion of individuals in this group who have a poverty index of less than **2**?

c. For which marital status is the **75th** percentile of the poverty index largest?

d. The **75the** percentile for the "Widowed" group is about **3.5**. What is the interpretation of this number, in terms of the proporiton of individuals in this group who have a poverty index of less than **4**?

e. For which marital status is the inter-quartile range (**IQR**) of the poverty index largest?

f. Based on the plot, what is the IQR for the "Separated" group, approximately? What is the interpretation of this value?

2. Blood Pressure

Here is a density plot representing the blood pressures of the subjects in this study, as well as the 25th percentile (also known as the first quartile), 50th percentile (also known as the median), and 75the percentile (also known as the third quartile).

```
NHANES_orig <- NHANES
NHANES <- filter(NHANES, BPDiaAve > 0)
```

```
ggplot(data = NHANES, mapping = aes(x = BPDiaAve)) + geom_density()
```
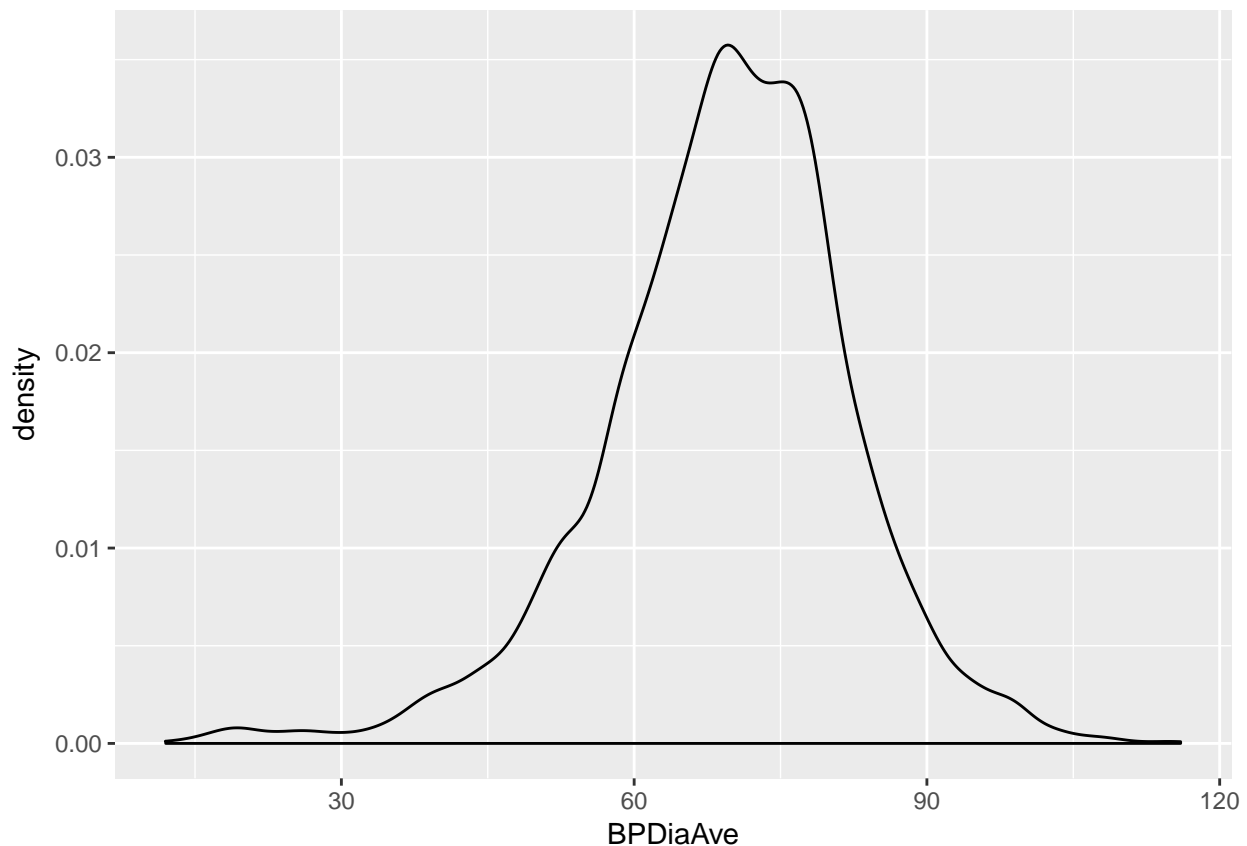


```
NHANES %>%
  summarize(
    q25 = quantile(BPDiaAve, probs = 0.25),
    median = median(BPDiaAve),
    q75 = quantile(BPDiaAve, probs = 0.75))
```

```
##   q25 median q75
## 1  62     70  77
```

a. What does the area under the density curve between 30 and 60 represent?

b. What is the approximate area under the density curve to the left of 62? (Don't calculate this based on the plot – answer based on what you know about the interpretation of area

3

under the curve and the definition of a percentile or quantile.)

c. What is the approximate area under the density curve to the left of 77? (Don't calculate this based on the plot – answer based on what you know about the interpretation of area under the curve and the definition of a percentile or quantile.)

3. Testosterone and Height

Here's a scatter plot showing a measurement of each subject's testosterone level and their height.

```
NHANES <- filter(NHANES, !is.na(Height), !is.na(Testosterone))
```

```
ggplot(data = NHANES, mapping = aes(x = Testosterone, y = Height)) + geom_point()
```
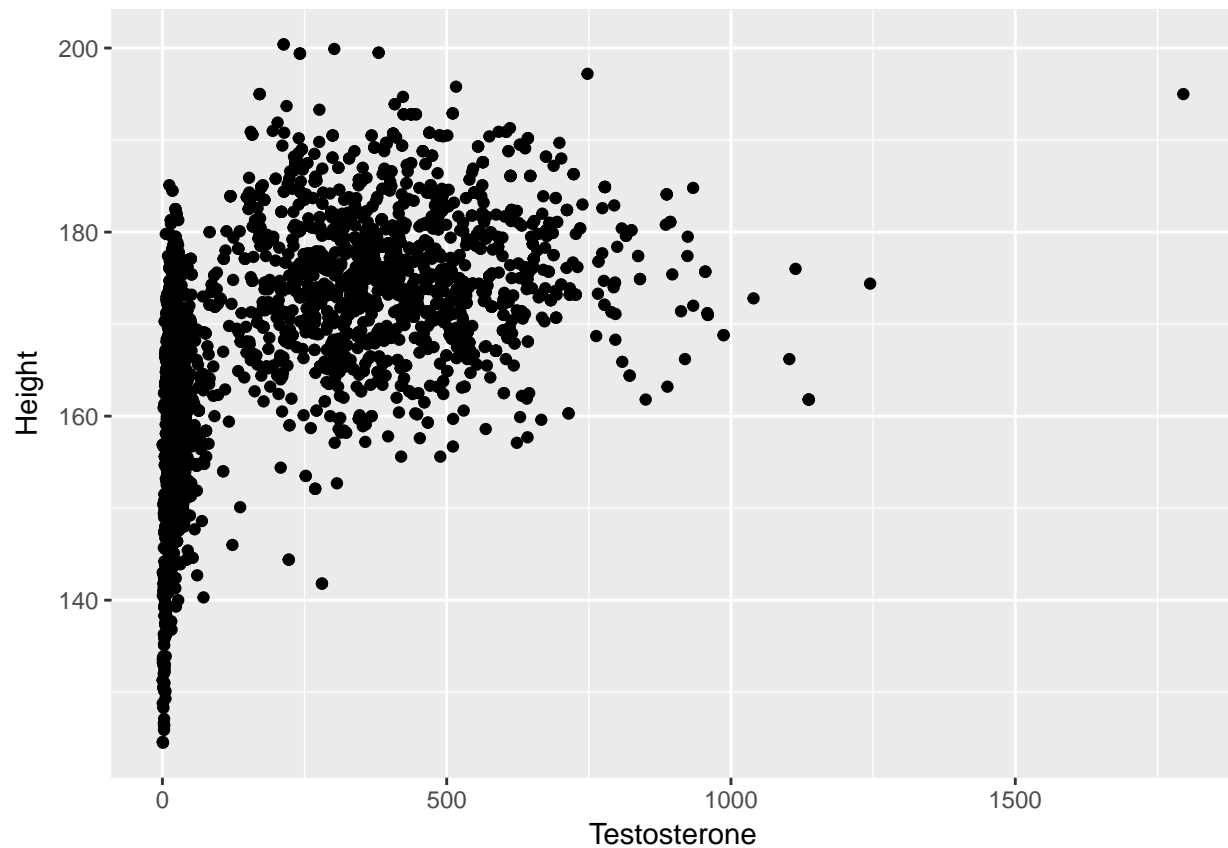

a. Describe the shape of the relationship between Testosterone and Height. Are there any outliers?

4

b. Would it be appropriate to summarize the relationship between Testosterone and Height using the correlation between them?

c. Describe the strength and direction of the relationship between testosterone and height. In case you would like to use it, the correlation cofficient is calculated below. (Even if you don't use the correlation in your answer on this sample quiz, you should know how you would use it.)

```
NHANES %>%
  select(Testosterone, Height) %>%
  cor()
```

```
##              Testosterone    Height
## Testosterone    1.0000000 0.5771714
## Height          0.5771714 1.0000000
```