Stat 140 - Quiz 1 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 Weight Height   BMI BPSysAve BPDiaAve Diabetes
## 1 51624   male  34   87.4  164.7 32.22      113       85       No
## 2 51625   male   4   17.0  105.4 15.30       NA       NA       No
## 3 51630 female  49   86.7  168.4 30.57      112       75       No
## 4 51638   male   9   29.8  133.1 16.82       86       47       No
## 5 51646   male   8   35.2  130.6 20.64      107       37       No
## 6 51647 female  45   75.7  166.7 27.24      118       64       No
```

1. What is each observational unit in this data set?

2. For each of the following variables, is that variable categorical or quantitative? If it is categorical, is it ordinal or nominal?

- `Gender`
- `Height`
- `Diabetes`

3. The following command counts how many observational units are in each combination of levels of the gender and diabetes variables.

```
tally(Diabetes ~ Gender, data = NHANES)
```

```
##         Gender
## Diabetes female male
##      No    3088 3013
##      Yes    269  283
```

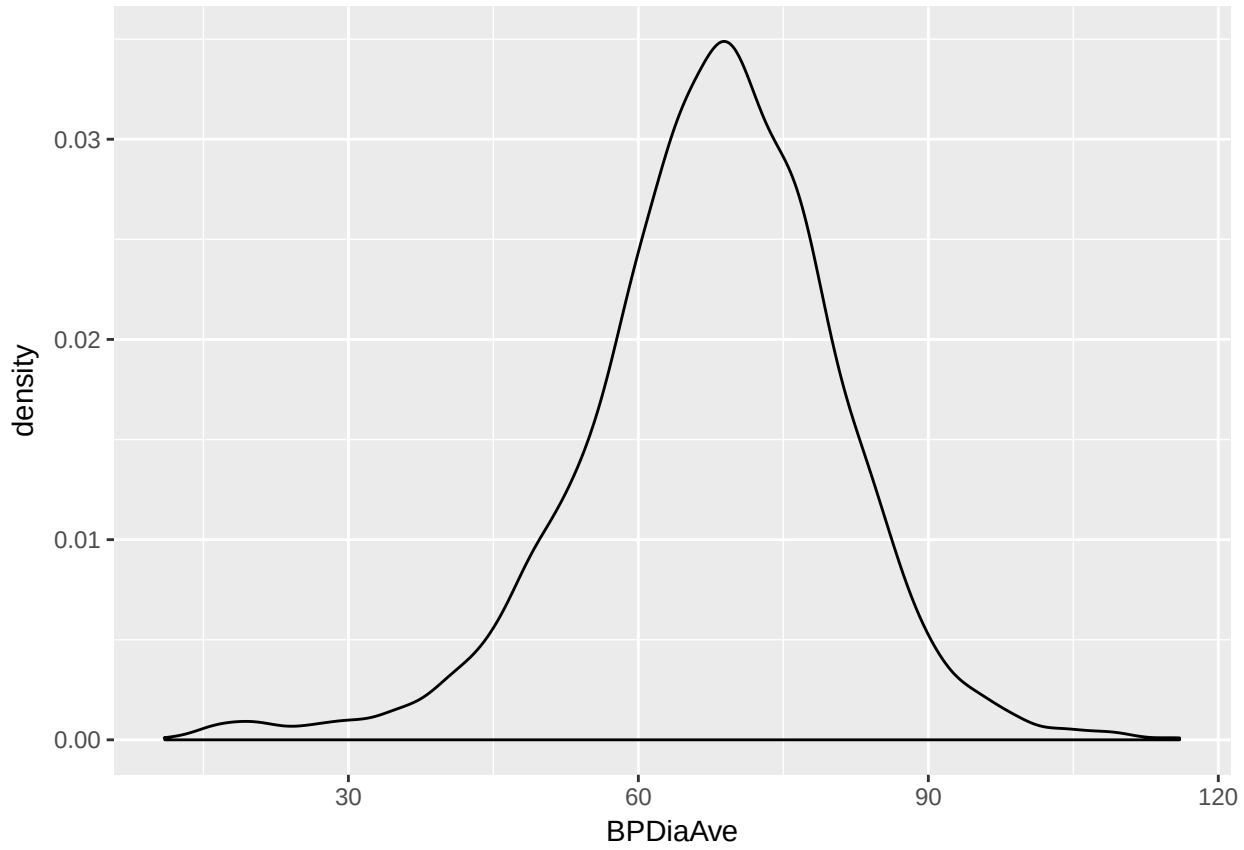
a. Calculate the joint distribution of Diabetes and Gender

b. Calculate the marginal distribution of Diabetes

c. Calculate the conditional distribution of Diabetes given that the subject's Gender is male

4. Here is a plot of the study participants' blood pressure measurements:

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



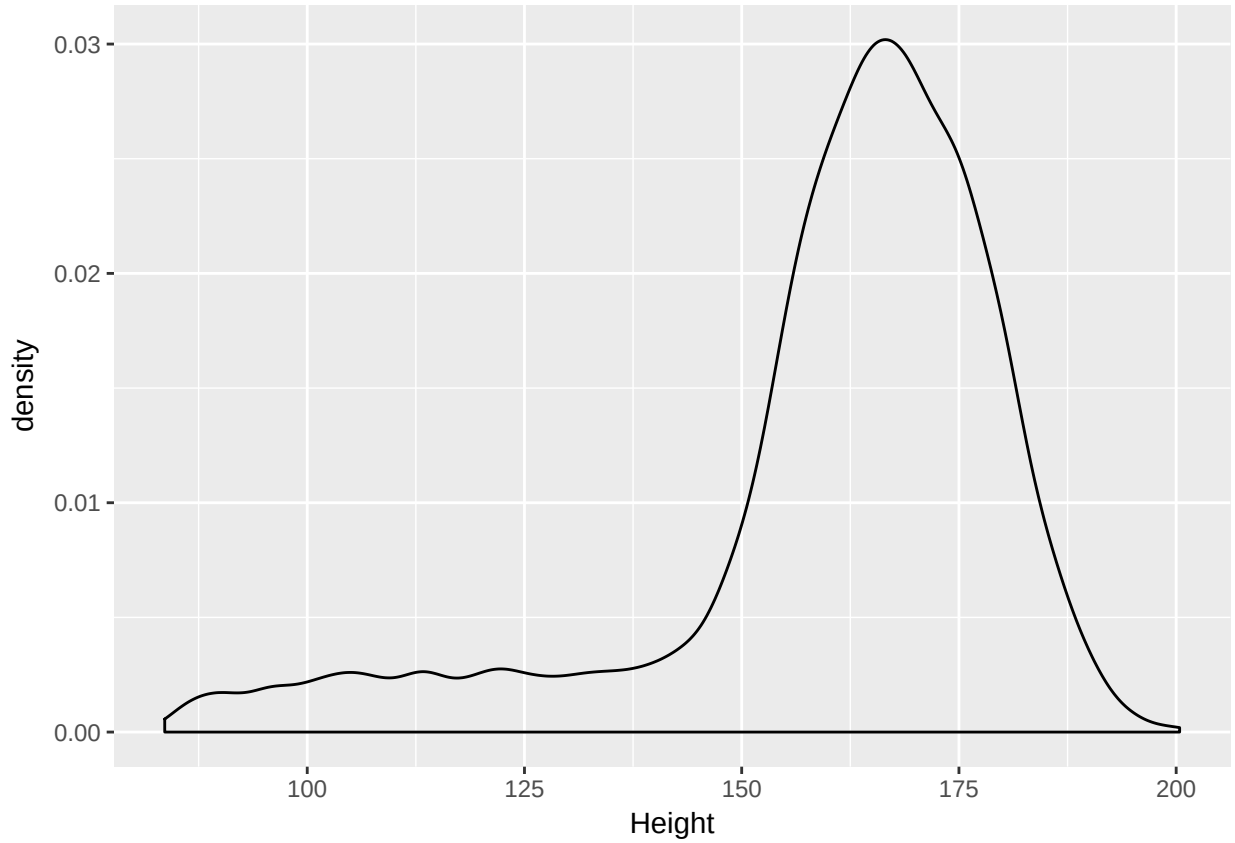
a. What statistics could you use to summarize the center of this distribution?

b. What statistics could you use to summarize the spread of this distribution?

5. Here is a plot of the study participants' heights:

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

```
## Warning: Removed 298 rows containing non-finite values (stat_density).
```



a. What statistics could you use to summarize the center of this distribution?

b. What statistics could you use to summarize the spread of this distribution?

6. For each of the following pairs of variables, circle the type of plot you would make, and write down the type of geometry you would use to make that plot. (more than one answer may be correct – if so, choose one)

6a. `Diabetes` and `Gender`

Bar Plot

Box Plot

Density Plot with groups shown in different colors

Scatter Plot

Geometry type: geom_bar

6b. `Age` and `BPSysAve` (`BPSysAve` is a measure of blood pressure)

Bar Plot

Box Plot

Density Plot with groups shown in different colors

Scatter Plot

Geometry type: geom_point

6c. `Diabetes` and `BPSysAve`

Bar Plot

Box Plot

Density Plot with groups shown in different colors

Scatter Plot

Geometry type: geom_boxplot or geom_density, depending on which option you chose above.

6d. `Weight` and `Height`

Bar Plot

Box Plot

Density Plot with groups shown in different colors

Scatter Plot

Geometry type: geom_point