The course professor will setup groups (2 students per group) for this assignment. Data for this assignment originated from the [2014 American Community Survey](http://content.bellevue.edu/cst/dsc/520/id/resources/acs-14-1yr-s0201.csv), which is conducted monthly by the Census Bureau and is designed to show how communities are changing. Through asking questions of a sample of the population, it produces national data on more than 35 categories of information, such as education, income, housing, and employment.

For this assignment, you will need to load and activate the ggplot2 package. Students are encouraged to complete the assigned reading and DataCamp exercises before starting their work.

Each student should first produce their own version of the following deliverables then share with their team members. Team members will then collaborate on their approach and insights, refining their work before submitting for a grade.

The following guidelines describe minimum deliverables needed for the assignment submission.

1. Document the elements in your data including the categories and data types.

a. The following are the variable types:

1. Categorical and Nominal: Id, Geography

2. Continuous and Ratio: RacesReported, HSDegree, BachDegree

3. Continuous and Interval: Id2

4. Categorical: PopGroupID, POPGROUP.display.label (While I think that these categories are nominal, the data set only includes a single value for these variables.)

1. Report the output from at least the following functions: str(); nrow(); ncol()

> str(survey)

'data.frame': 136 obs. of 8 variables:

$ Id : Factor w/ 136 levels "0500000US01073",..: 1 2 3 4 5 6 7 8 9 10 ...

$ Id2 : int 1073 4013 4019 6001 6013 6019 6029 6037 6059 6065 ...

$ Geography : Factor w/ 136 levels "Alameda County, California",..: 56 70 98 1 20 43 62 68 92 106 ...

$ PopGroupID : int 1 1 1 1 1 1 1 1 1 1 ...

$ POPGROUP.display.label: Factor w/ 1 level "Total population": 1 1 1 1 1 1 1 1 1 1 ...

$ RacesReported : int 660793 4087191 1004516 1610921 1111339 965974 874589 10116705 3145515 2329271 ...

$ HSDegree : num 89.1 86.8 88 86.9 88.8 73.6 74.5 77.5 84.6 80.6 ...

$ BachDegree : num 30.5 30.2 30.8 42.8 39.7 19.7 15.4 30.3 38 20.7 ...

> nrow(survey)

[1] 136

> ncol(survey)

[1] 8

1. Create a Histogram of the HSDegree variable using the ggplot2 package. Set a bin size for the Histogram.
2. Include a Title and appropriate X/Y axis labels on your Histogram Plot.
3. Answer the following questions based on the Histogram produced:
   1. Based on what you see in this histogram, is the data distribution unimodal?
   2. Is it approximately symmetrical?
   3. Is it approximately bell-shaped?
   4. Is it approximately normal?
   5. If not normal, is the distribution skewed? If so, in which direction? Provide specifics in your responses.
   6. Include a normal curve to the Histogram that you plotted.
   7. Explain whether a normal distribution can accurately be used as a model for this data.
4. Create a Probability Plot of the HSDegree variable.
5. Answer the following questions based on the Probability Plot:
   1. Based on what you see in this probability plot, is the distribution approximately normal? Explain how you know.
   2. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.
6. Now that you have looked at this data visually for normality, you will now quantify normality with numbers using the stat.desc() function. Include a screen capture of the results produced.
7. In several sentences provide an explanation of the result produced for skew, kurtosis, and z-scores. In addition, explain how a change in the sample size may change your explanation?
8. Repeat steps 3 through 9 for the BachDegree variable.