American Community Survey Exercise

For this exercise, you will use the following dataset, [2014 American Community Survey](http://content.bellevue.edu/cst/dsc/520/id/resources/acs-14-1yr-s0201.csv). This data is maintained by the US Census Bureau and are 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. For this deliverable, you should provide the following:

* + 1. What are the elements in your data (including the categories and data types)?

|  |  |
| --- | --- |
| Categories | Data type |
| Id | Character |
| Id2 | Integer |
| Geography | Character |
| PopGroupID | Character |
| POPGROUP.display-label | Character |
| RacesReported | Character |
| HSDegree | Number |
| BachDegree | Number |

* + 1. Please provide the output from the following functions: str(); nrow(); ncol()

**str()**

> str(surveydata)

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

$ Id : chr "0500000US01073" "0500000US04013" "0500000US04019" "0500000US06001" ...

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

$ Geography : chr "Jefferson County, Alabama" "Maricopa County, Arizona" "Pima County, Arizona" "Alameda County, California" ...

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

$ POPGROUP.display.label: chr "Total population" "Total population" "Total population" "Total population" ...

$ 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()**

> nrow(surveydata)

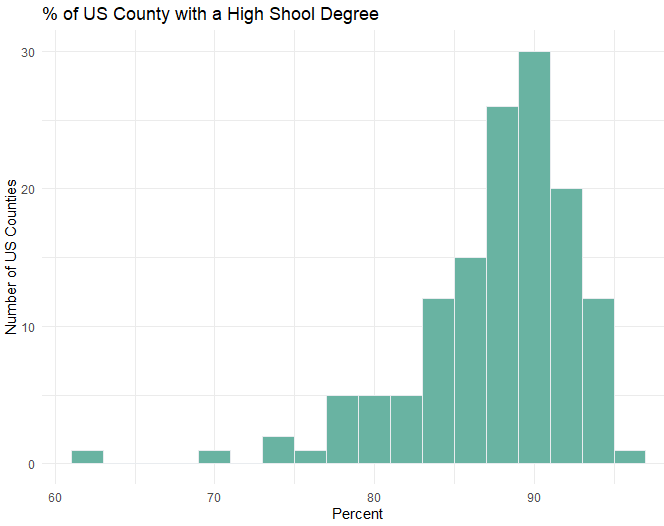
[1] 136

**ncol()**

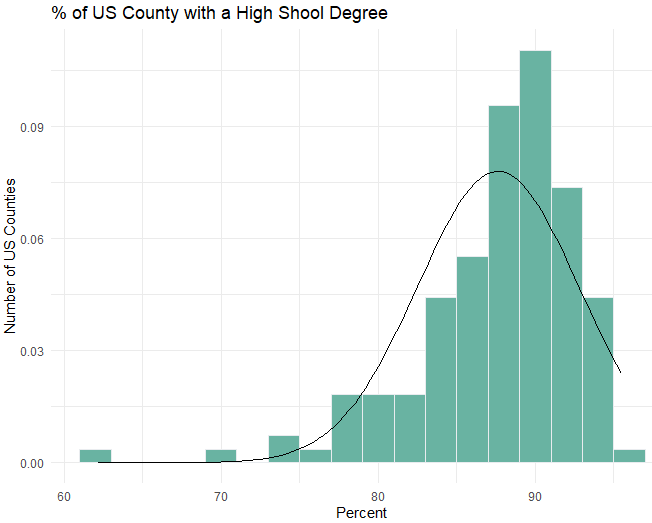
> ncol(surveydata)

[1] 8

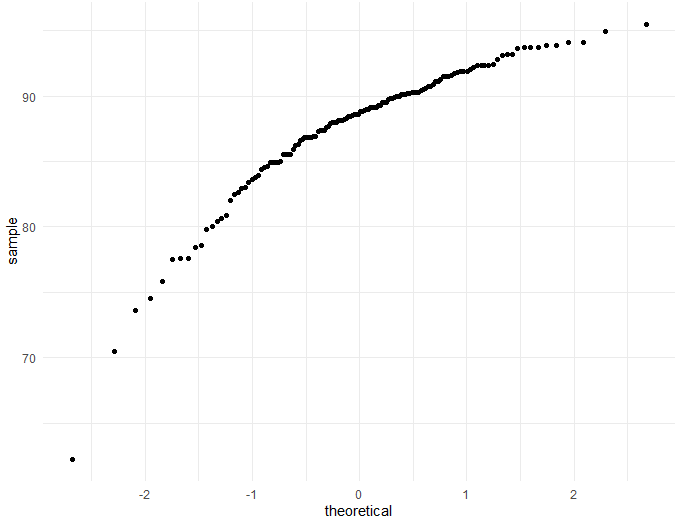
* + 1. Create a Histogram of the HSDegree variable using the ggplot2 package.
       - Set a bin size for the Histogram.
       - Include a Title and appropriate X/Y axis labels on your Histogram Plot.



* + 1. Answer the following questions based on the Histogram produced:
       - Based on what you see in this histogram, is the data distribution unimodal? – Yes the data distribution unimodal (one peak in the distribution)
       - Is it approximately symmetrical? – No. It is negatively skewed.
       - Is it approximately bell-shaped? – Yes with a negative skew.
       - Is it approximately normal? – No. It is negatively skewed.
       - If not normal, is the distribution skewed? If so, in which direction? – No. It is negatively skewed.
       - Include a normal curve to the Histogram that you plotted.



* + - * Explain whether a normal distribution can accurately be used as a model for this data. – A normal distribution cannot be accurately used as a model for this data since it is negatively skewed.
    1. Create a Probability Plot of the HSDegree variable.



* + 1. Answer the following questions based on the Probability Plot:
       - * Based on what you see in this probability plot, is the distribution approximately normal? Explain how you know. – It is not normal because the probability plot shows the curve bending down with more values on the higher end.
         * If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

<http://www.statistics4u.com/fundstat_eng/ee_probability_plot.html>

* + 1. 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.
    2. 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?