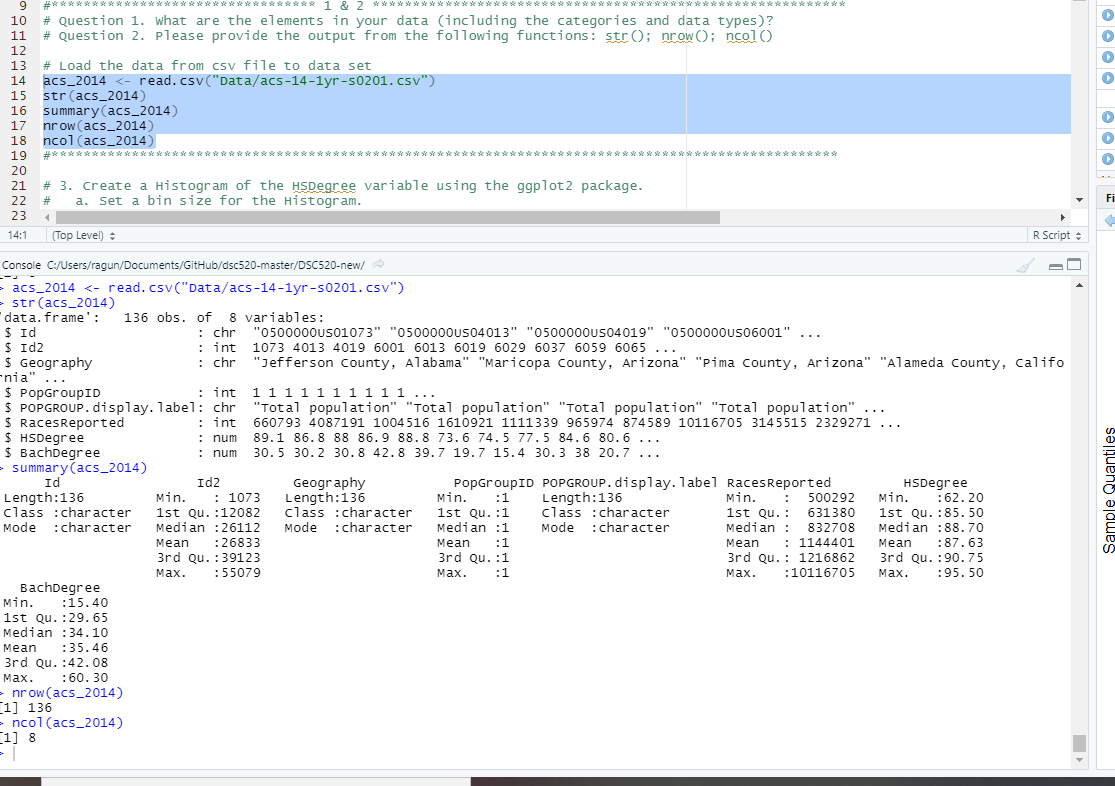
For this assignment, you will need to load and activate the ggplot2 package. For this deliverable, you should provide the following:

1. **What are the elements in your data (including the categories and data types)?**
2. Please provide the output from the following functions: str(); nrow(); ncol()

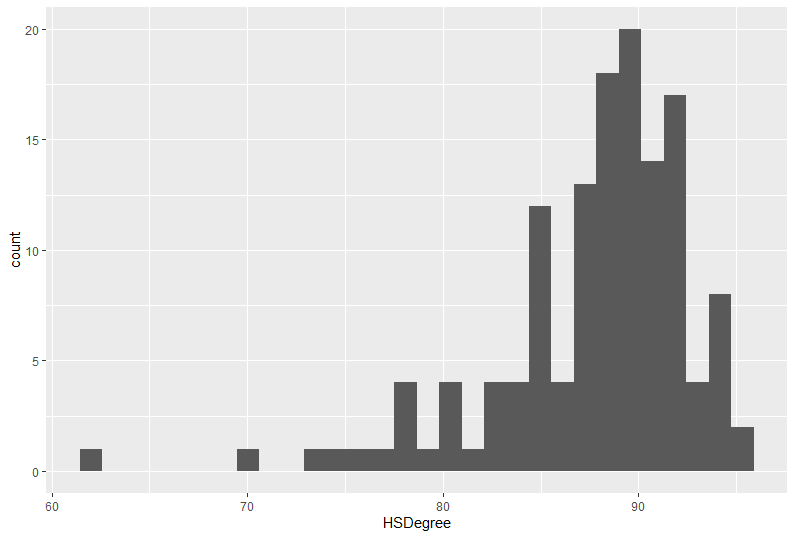
**Answer for Question 1 & 2 :**

8 elements in our data which including Id, Geography, Population Group Id, Label, Races Reported id, HS and bachelor’s degree. Please find the below details for data type and values.

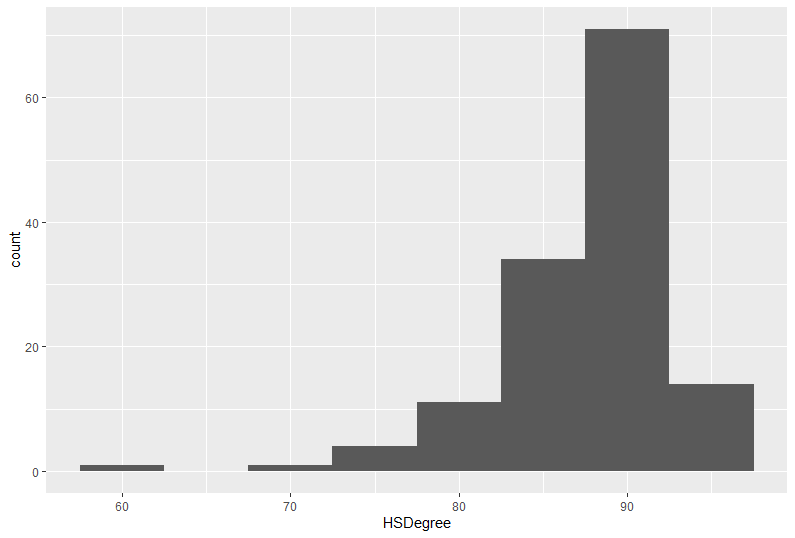


1. **Create a Histogram of the HSDegree variable using the ggplot2 package.**

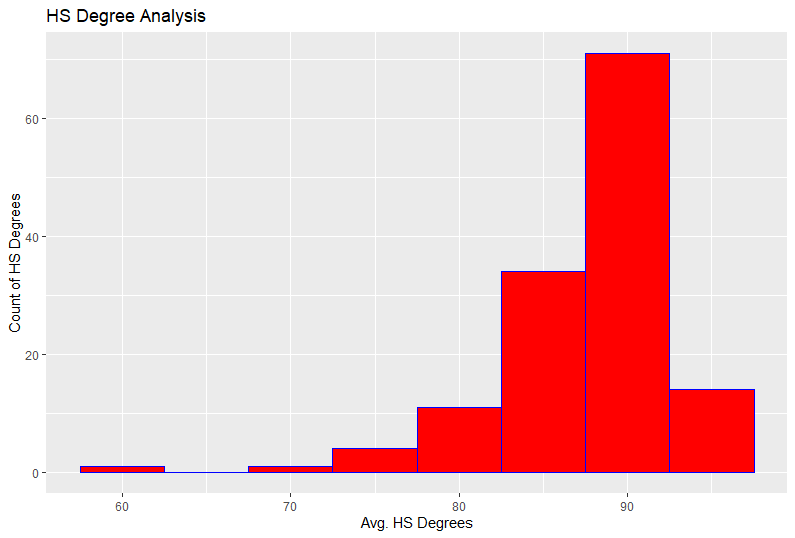
|  |
| --- |
| library(ggplot2)  ggplot(acs\_2014, aes(x = HSDegree)) + geom\_histogram()  # a. Set a bin size for the Histogram  ggplot(acs\_2014, aes(x = HSDegree)) + geom\_histogram(binwidth = 5)  # b. Include a Title and appropriate X/Y axis labels on your Histogram Plot.  ggplot(acs\_2014, aes(x = HSDegree)) + geom\_histogram(binwidth = 5, color="blue", fill="red") +  labs(title = "HS Degree Analysis", x = "Avg. HS Degrees", y="Count of HS Degrees") |



**a. Set a bin size for the Histogram.**



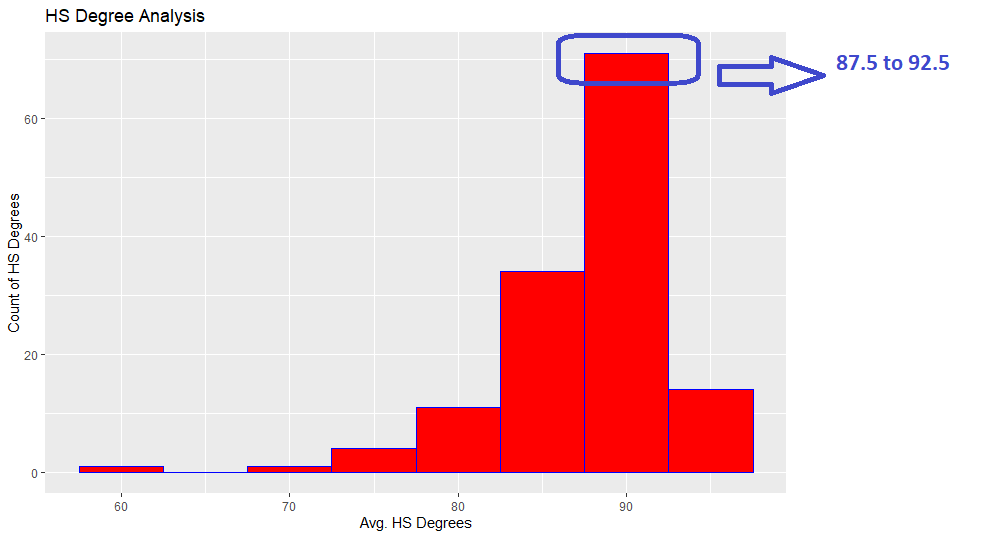
**b. Include a Title and appropriate X/Y axis labels on your Histogram Plot.**



4. Answer the following questions based on the Histogram produced:

a. Based on what you see in this histogram, is the data distribution unimodal?

Answer : Yes



b. Is it approximately symmetrical?

Answer : No, It is not symmetrical

c. Is it approximately bell-shaped?

Answer : No, it’s not bell Shaped

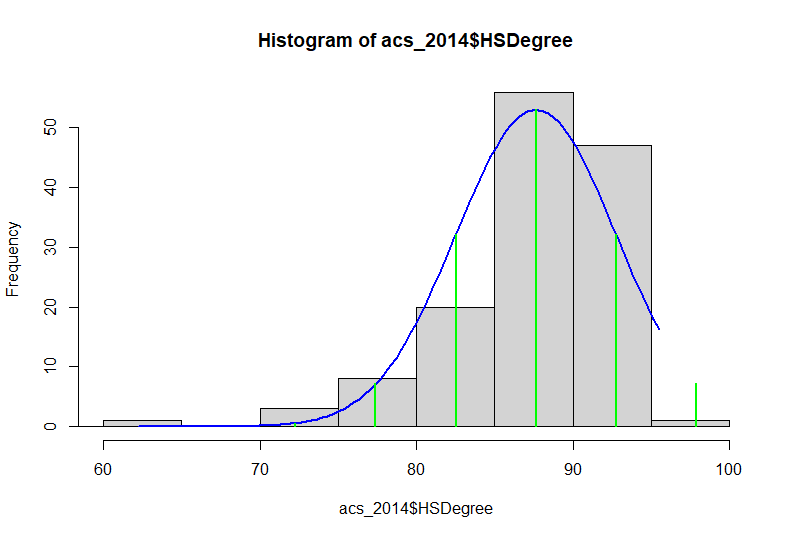
d. Is it approximately normal?

Answer : No, It’s not symmetrical. Hence not normal too

e. If not normal, is the distribution skewed? If so, in which direction?

Answer : Yes, the distribution is skewed. Negative Skewed

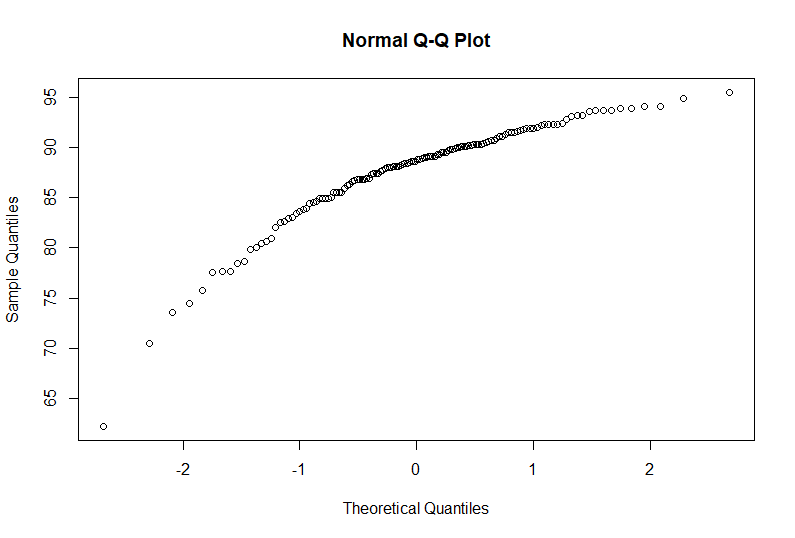
f. Include a normal curve to the Histogram that you plotted.



g. Explain whether a normal distribution can accurately be used as a model for this data.

Answer : As per the graph, we can not say that normal distribution since its negatively skewed. We can see the graph and histogram are not in sync.

1. Create a Probability Plot of the HSDegree variable.



6. Answer the following questions based on the Probability Plot:

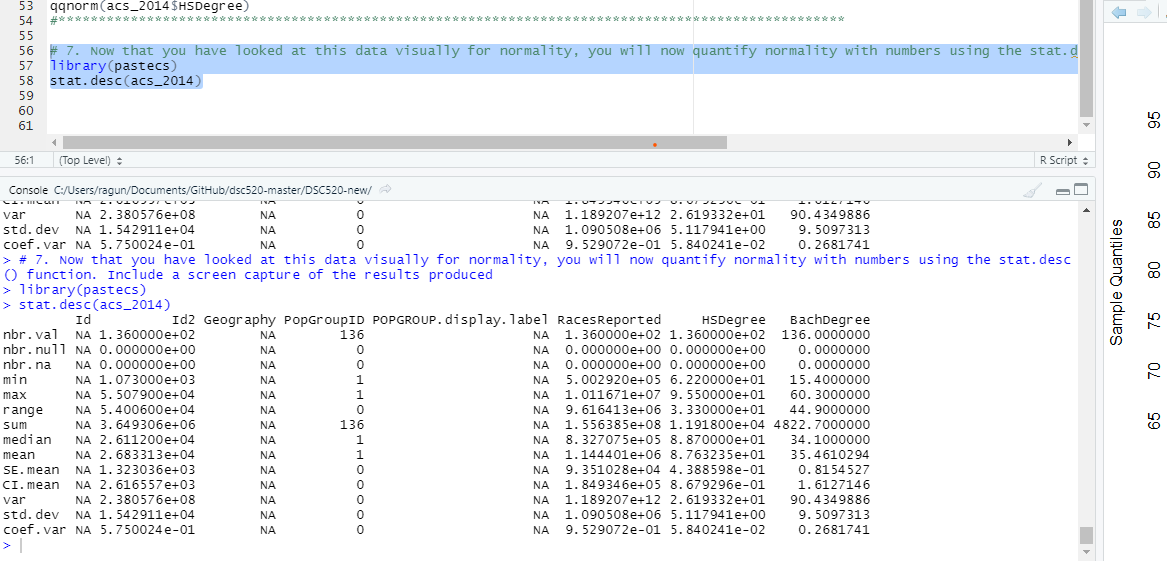
a. Based on what you see in this probability plot, is the distribution approximately normal? Explain how you know.

Answer : No, It is not normal

b. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.

Answer : Negative Skewed

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.



8. 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?

Answer : This data is negative Skewed. As we know, the sample size increases, the margin of error started decreases. ( inverse because the two moves in opposite directions )