```
# Assignment: ASSIGNMENT 3 - second part
# Name: Singhal, Sarika
# Date: 2021-06-22

install.packages("ggplot2")
## Load the ggplot2 package
library(ggplot2)
theme_set(theme_minimal())

## Load the `data/acs-14-1yr-s0201.csv` to
community_Survey <- read.csv("acs-14-1yr-s0201.csv")

## What are the elements in your data (including the categories and data types)?
categories <- unique(community_Survey)
numberOfCategories <- length(categories)
print(numberOfCategories)

categoriesId <- unique(community_Survey$Id)
numberOfCategoriesId <- length(categoriesId)
print(numberOfCategoriesId)

categoriesId2 <- unique(community_Survey$Id2)
numberOfCategoriesId2 <- length(categoriesId2)
print(numberOfCategoriesId2)

categoriesGeography <- unique(community_Survey$Geography)
numberOfCategoriesGeography <- length(categoriesGeography)
print(numberOfCategoriesGeography)

categoriesPopGroupID <- unique(community_Survey$PopGroupID)
numberOfCategoriesPopGroupID <- length(categoriesPopGroupID)
print(numberOfCategoriesPopGroupID)

categoriesRacesReported <- unique(community_Survey$RacesReported)
numberOfCategoriesRacesReported <- length(categoriesRacesReported)
print(numberOfCategoriesRacesReported)

categoriesHSDegree <- unique(community_Survey$HSDegree)
numberOfCategoriesHSDegree <- length(categoriesHSDegree)
print(numberOfCategoriesHSDegree)

categoriesBachDegree <- unique(community_Survey$BachDegree)
numberOfCategoriesBachDegree <- length(categoriesBachDegree)
print(numberOfCategoriesBachDegree)

typeof(community_Survey$Id)
typeof(community_Survey$Id2)
typeof(community_Survey$Geography)
typeof(community_Survey$PopGroupID)
typeof(community_Survey$POPGROUP.display.label)
typeof(community_Survey$RacesReported)
typeof(community_Survey$HSDegree)
typeof(community_Survey$BachDegree)
```

```
##Please provide the output from the following functions: str(); nrow(); ncol()

str(community_Survey)
nrow(community_Survey)
ncol(community_Survey)

##Create a Histogram of the HSDegree variable using the ggplot2 package.

##1. Set a bin size for the Histogram.
qplot(community_Survey$HSDegree, geom="histogram",bins = 30)

##2Include a Title and appropriate X/Y axis labels on your Histogram Plot.

qplot(community_Survey$HSDegree, geom="histogram",bins = 30)+ ggtitle('HS Degree Plot') +
xlab('HS Degree')


##Answer the following questions based on the Histogram produced:
##Based on what you see in this histogram, is the data distribution unimodal? - Yes
## Is it approximately symmetrical? - No
##Is it approximately bell-shaped? - No
##Is it approximately normal? - No
##If not normal, is the distribution skewed? If so, in which direction? - Left
mean<-mean(community_Survey$HSDegree)
sd<-sd(community_Survey$HSDegree)
mean
sd
##Include a normal curve to the Histogram that you plotted.

y<-dnorm(community_Survey$HSDegree, mean, sd)
plot(community_Survey$HSDegree,y)
install.packages("moments")
library(moments)
skewness(community_Survey$HSDegree)
#-1.69
kurtosis(community_Survey$HSDegree)
#7.4
##Explain whether a normal distribution can accurately be used as a model for this data.

#Looking at the skewness and kurtosis numbers the distribution is higly skewed.

#5. Create a Probability Plot of the HSDegree variable.


mean <- mean(community_Survey$HSDegree)
sd <- sd(community_Survey$HSDegree)
n <- nrow(community_Survey)
p <- (1 : n) / n - 0.5 / n
ggplot(community_Survey) + geom_point(aes(x = p, y = sort(pnorm(HSDegree, mean, sd))))
+ggtitle('HS Degree Probability Plot')

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

#A straight, diagonal line means that you have normally distributed data. If the line is skewed to the left or right, it means that you do not have normally distributed data.

#In this case this data is not exactly normally distributed

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

#Distribution skeweded in right as the curve starts to move in right direction.

#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.

```
install.packages("stat.decs")
library(stat.decs)
stat.desc(community_Survey$HSDegree, basic=TRUE, desc=TRUE, norm=FALSE, p=0.95)

install.packages("pastecs")
library(pastecs)

hs0<-read.table("acs-14-1yr-s0201.csv", sep=",", header=T)
head(hs0)
attach(hs0)
scores<-cbind(HSDegree)
stat.desc(scores,basic=F)
stat.desc(scores,desc=F)
```