Module5: Daniel Bihnam Problem Set

## Background

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient. Data source: [https://www.kaggle.com/fedesoriano/stroke-prediction-dataset.](https://www.kaggle.com/fedesoriano/stroke-prediction-dataset) We will assume that the sample is representative of adults.

## Research Question

Suppose you want to know if the average adult BMI is different between patients that do not suffer a stroke versus those that suffer a stroke.

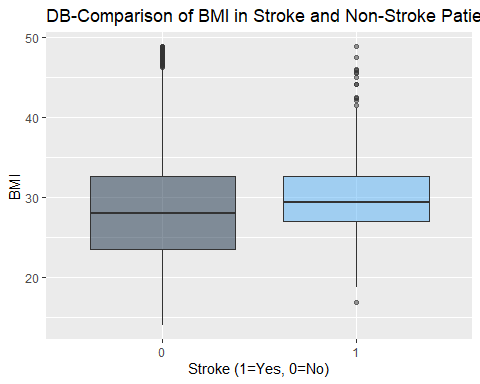
## Solution

Read in the data and include all libraries needed.

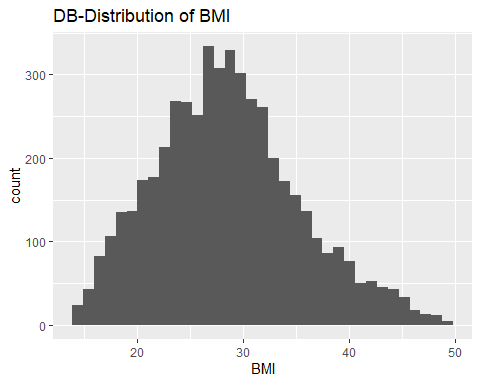
brainstroke<- read.csv("https://raw.githubusercontent.com/jenbroatch/LSC540/main/DataSets/brainstroke.csv")  
  
library(ggplot2)

First, we will visualize our data here.

ggplot(brainstroke, aes(x=as.factor(stroke), y=bmi, fill=stroke)) +   
 geom\_boxplot(alpha=0.5) +  
 ggtitle("DB-Comparison of BMI in Stroke and Non-Stroke Patients") +  
 ylab("BMI") +   
 xlab("Stroke (1=Yes, 0=No)") +  
 theme(legend.position='none')



ggplot(data=brainstroke,aes(x=bmi))+  
 geom\_histogram(bins=35)+  
 ggtitle('DB-Distribution of BMI')+  
 xlab('BMI')



Our visualization shows a slight increase in the average BMI of stroke patients compared to non-stroke patients. We can also see that the variable BMI follows a fairly normal distribution across all samples.

t.test(bmi~stroke,data=brainstroke)

Welch Two Sample t-test  
  
data: bmi by stroke  
t = -4.7669, df = 286.11, p-value = 2.982e-06  
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
95 percent confidence interval:  
 -2.510728 -1.043263  
sample estimates:  
mean in group 0 mean in group 1   
 28.40970 30.18669

We are 95% confident that the true difference in BMI between stroke and non-stroke patients lies between 2.51 and 1.04. Our reported p-value was 2.9e-6, and so our p-value<0.05. This helps us conclude that these results hold statistical significance. Thus, we can reject the null hypothesis as we have displayed there is a correlation between having a stroke and higher BMI.