Name: Elias Fedai
Date: 07/22/19
Title: Analysis of Heart Disease

Section 1

- **Introduction**
  A human can live without a brain, but cannot live without a heart. The human heart is one of the most if not the most vital organ in the human body. It is essential that a healthy heart is needed to maintain an overall healthy body. Unfortunately, for millions of people heart disease has caused serious medical conditions some of which are irreversible and life threatening.

- **Research questions**
  1) Can cholesterol level be used to determine if a person has heart disease?
  2) Can gender be used either alone or in tandem with any other variables to determine if a person has heart disease?
  3) Do people with genetic disorders such as Beta thalassemia exhibit higher cases of heart disease than those with no disorder?
  4) What are the outcomes of ST readings on predicting heart diseases?
  5) Do individuals with higher levels of blood glucose display higher rates of heart disease?
  6) Can resting ECG in tandem with age be used to predict if a person has heart disease?

- **Approach**
  My approach is to determine if there are sufficient findings to answer my research questions while using the various techniques in R that I have learned throughout this course. The use of regression analysis will be used heavily.

- **How your approach addresses (fully or partially) the problem**.
  My approach only addresses the problem partially. First, the data sets are not too extensive or long enough to fully address the problem. Second, the complexity of heart disease and its diagnosis are an extremely multi-layer situation alongside extensive clinical presentation.

- **Data**
  When reviewing all the datasets sources one thing that caught my attention and must be kept as side note was the age of the datasets, some of which were older.
  1) https://archive.ics.uci.edu/ml/datasets/Heart+Disease (VA)
  2) https://www.kaggle.com/ronitf/heart-disease-uci
  3) https://healthdata.gov/dataset/national-health-interview-survey-nhis-national-cardiovascular-disease-surveillance-data

- **Required Packages**
  For my research I will use many different packages some of the immediate necessary ones are ggplots, car, caret, QuantPsyc, blorr, and some others I will able to name as I proceed further into the research.

- **Plots and Table Needs**
  The need for visualization is vital in my research. I will be using various different types of tables and plots such as histograms, various plots (i.e. scatter, box), and tables as needed for each of the respective research questions.

- **Questions for future steps**
  1) Is the data clean?
  2) Are there other variables in my data that can obscure my analysis or just don't make sense?
  3) How do all the variables in my data correlate to each other?
  4) What are the takeaways from the visualizations?