Introduction

This project aims to help patients prevent the heart disease in advance and increase the accuracy of doctors’ diagnosis by analyzing the relationship between characters of patients and whether a man have a heart disease or not.

According to the WHO, cardiovascular diseases (CVDs) are the leading cause of death globally, taking an estimated 17.9 million lives each year, and heart disease is the number one killer of human health. The high death rate is since the symptoms of cardiac disease are not always visible and detectable, and it is difficult to draw adequate attention to them. If it is possible to extract the physical measurement indicators related to the human body and analyze the impact of different characteristics on heart disease through data mining, machine learning, and modeling algorithm which can help patients prevent heart disease in advance and improve the accuracy of doctors' diagnosis.

In this project, we mainly focus on detecting relationship between different characteristics of real patients and heart disease with the statistical method and our second goal is to visualize

statistical result and make it more explainable.

The dataset comes a heart disease study on the Kaggle website provided by UCI Machine Learning Repository, which recorded 300 patients from Cleveland and some characteristics related to heart disease. As you can see, the variables are all in categorical and numerical form, including 303 sample, 13 variables and 5 numerical data.

We mainly use descriptive and Inferential Statistics to showcase our results. Descriptive Statistics includes Indicator statistics such as Range, mean, median, mode, standard deviation to make a comparison between indicators for patients and normal indicators. We also conduct shiny to display data including scatter, linear, pie, bar, and box line charts for each indicator. Inferential Statistics is also divided into variable correlation analysis and machine learning models. On the one hand we use ANOVA, Linear Regression Analysis, and Chi-square test to analysis relationship between different variables; On the other hand, we build three models to make future prediction and examine accuracy of these models.