# Heart Disease Data Exploration

## 1. Introduction and Data Preparation

As the largest single cause of death on the planet, heart disease is an important and life or death matter. With increasing number of people of being diagnosed of heart disease all over the world, it is paramount to study the symptoms correlates with the disease as well as identify the population that is susceptible to this kind of disease for early prevention. The dataset is collected from this website (https://gmudatamining.com/data/heart_disease.rds). Before creating the plot, the data values of type string are formatted, and the variable names are fixed ahead of time for better layout in legends and labels.

## 2. Figure Description and Analysis

In the Figure 1, the density plot and histogram of the resting blood pressure is shown faceted by different types of Thalassemia. The first row is the patients with heart disease while the second row is the patients without heart disease. The distribution of blood pressure varies among the three types of Thalassemia, and this might imply that Thalassemia correlated with resting blood pressure. For patients with reversible defect Thalassemia, much more patients with heart disease tend to have blood pressure around 130 than patients without heart disease, and the distributions of blood pressure are both skewed to the right.

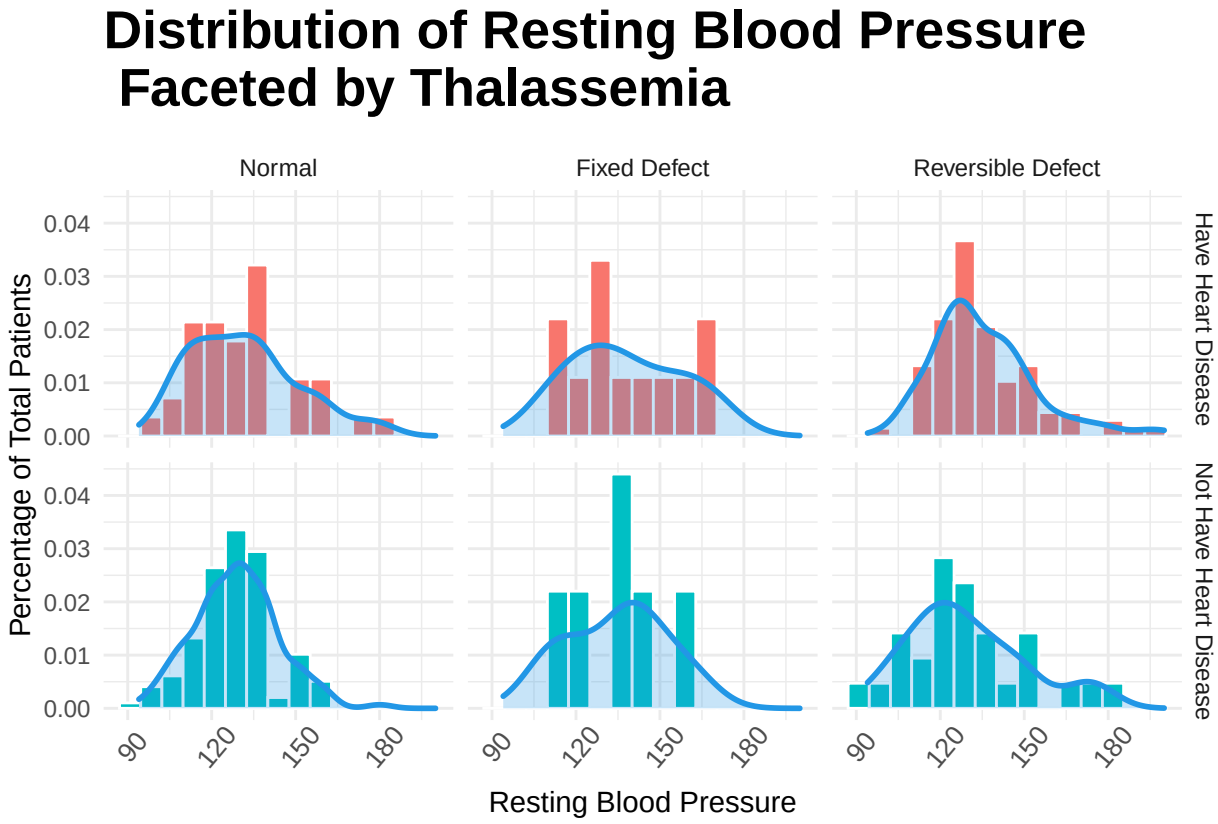


Figure 2 is a scatterplot with age as the x axis and cholesterol as the y axis, faceted by whether the patient

has heart disease. For each data point, the different colors represent different types of chest pain, while different shapes demonstrate different resting blood pressure. According to the plot, it is easy to conclude that most of patients who have heart disease experience asymptomatic chest pain in contrast with patients without heart disease, which is quite unexpected. Intuitively, heart disease should cause uncomfortable feelings in chest where the heart lies, but the asymptomatic chest pain makes heart disease even harder to detect in the early stage and increase the risks of having side effects from heart disease. Comparatively, patients without heart disease experience diverse types of chest pain, and a larger proportion experience atypical and non-anginal chest pain than patients with heart disease. In terms of age, most of patients being diagnosed of the disease are around 60 years old, while the age of patients without the disease are much more scattered. Most people with the disease tend to have cholesterol level of around 200 to 300. There is no apparent relation between resting blood pressure and the disease from the figure.