Portfolio 4
Bella Wu
Stat 479

Personal Reflection

This portfolio is revised upon the first portfolio about heart disease data exploration. I choose to extend the first portfolio because of its theme and aesthetic values. I want to explore more in this area, and moreover, I am quite satisfied with my visualization styles in the first version, so I continue to apply the same color encoding and aesthetic setup to the final version. However, there are also some problems regarding my first portfolio that needs to be improved.

As shown in Figure 1, the first version contains duplicated and unconcise code that is difficult to read and takes up too much space. In the final version, I compress the code by combining values into a list, making the code more readable.

```
levels(df$`Heart Disease`)[levels(df$`Heart Disease`)
                           =="yes"] <- "Have Heart Disease"
levels(df$`Heart Disease`)[levels(df$`Heart Disease`)
                           =="no"] <- "Not Have Heart Disease"
levels(df$thalassemia)[levels(df$thalassemia)
                       =="normal"] <- "Normal"
levels(df$thalassemia)[levels(df$thalassemia)
                       =="fixed_defect"] <- "Fixed Defect"
levels(df$thalassemia)[levels(df$thalassemia)
                       =="reversible_defect"] <- "Reversible Defect"
levels(df$`Chest Pain`)[levels(df$`Chest Pain`)
                        =="typical"] = "Typical"
levels(df$`Chest Pain`)[levels(df$`Chest Pain`)
                        =="atypical"] = "Atypical"
levels(df$`Chest Pain`)[levels(df$`Chest Pain`)
                        =="non_anginal"] = "Non-anginal"
levels(df$`Chest Pain`)[levels(df$`Chest Pain`)
                        =="asymptomatic"] = "Asymptomatic"
```

```
levels(df$`Heart Disease`) = c("Have Heart Disease", "Not Have Heart Disease")
levels(df$thalassemia) = c("Normal", "Fixed Defect", "Reversible Defect")
levels(df$`Chest Pain`) = c("Typical", "Atypical", "Non-anginal", "Asymptomatic")
levels(df$resting_electrocardiograph) = c("Normal", "ST-T Wave Abnormal", "Ventricular
Hypertrophy")
```

Figure 1. Comparison of codes in the previous version (left) and the final version (right).

Furthermore, the first version failed to achieve its aim in identifying the population that is susceptible to heart disease. To accomplish the goal, I incorporated additional visualization in the final version as demonstrated in Figure 2. The combination between violin plot and box plot allows readers to understand the distribution of data and get an overview about the quantiles and medians. Different colors in patients with heart disease and without heart disease make a contrast in population that is susceptible to the disease. It seems like patients with the disease are at the age of around 57, max heart rate of 145, and cholesterol of 250. Comparatively, the majority of normal people have a higher maximum heart rate, lower cholesterol, and old peak at level 0.

Figure 2: Violin plots of age, max heart rate, cholesterol, and old peak, colored by heart disease.

There are some problems in the first portfolio regarding the functionality since the plots failed to tell an interesting story. For the distribution of resting blood pressure, if being faceted by Thalassemia, there barely any different in patients with heart disease and without it. However, if being faceted by chest pain, readers could draw more insightful information from it. In the case of typical chest pain, patients tend to have a lower resting blood pressure than normal people. Under the circumstance of atypical chest pain, the distribution of resting blood pressure for patients is more scattered than normal people.

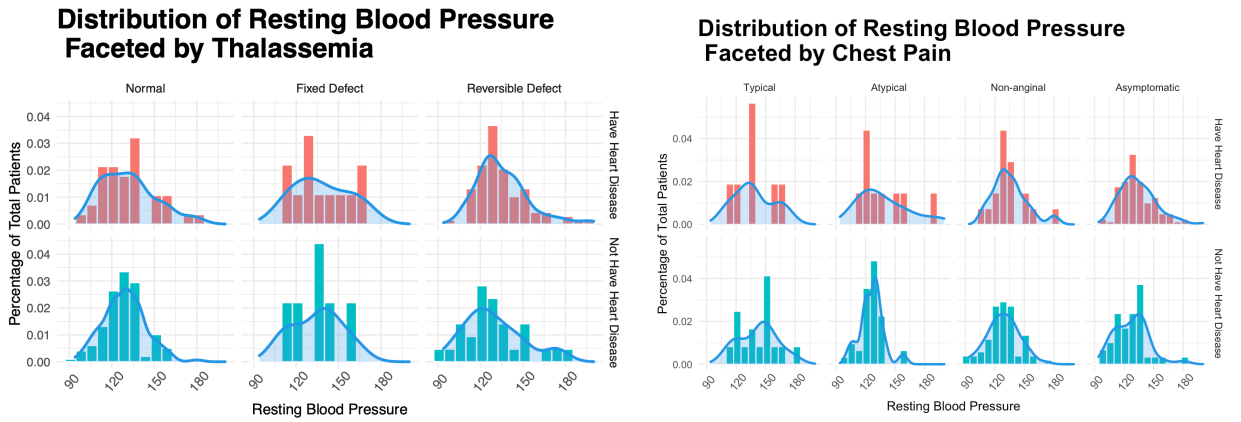

Figure 3. Distribution of resting blood pressure faceted by Thalassemia (left -first version) and Chest Pain (right -final version)

The first portfolio contains one visualization with great information density. In Figure 4 (left), the points have differing colors and sizes and the plot is faceted by heart disease. Readers have to compare between two facets, point size and point color and this is simply too much information contained in one plot. To improve this, I remove the point size in the final version. Instead of coloring by chest pain, I choose to color by thalassemia since chest pain is already analyzed in Figure 3 (right).

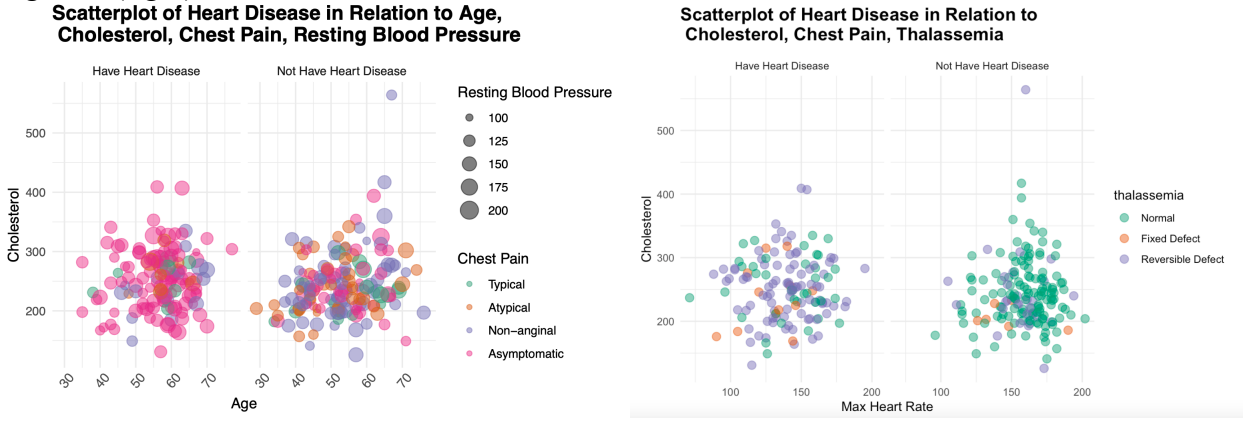

Figure 4. Left (first version): Scatterplot between age and cholesterol in relation to resting blood pressure and chest pain. Right (final version): Scatterplot between max heart rate and cholesterol in relation to thalassemia.

Resting electrocardiogram is one paramount features used by doctors in the diagnosis of heart disease, but the first portfolio does not include this feature. Therefore, in the final version, I also explore this feature in relation to cholesterol. As shown in the plot below, normal people do not

have abnormal ST-T wave. Nevertheless, for patients with abnormal ST-T wave, their cholesterol levels centered at around either 200 or 350, indicating that abnormally high or low cholesterol tends to correlated with abnormal ST-T wave in heart disease patients.

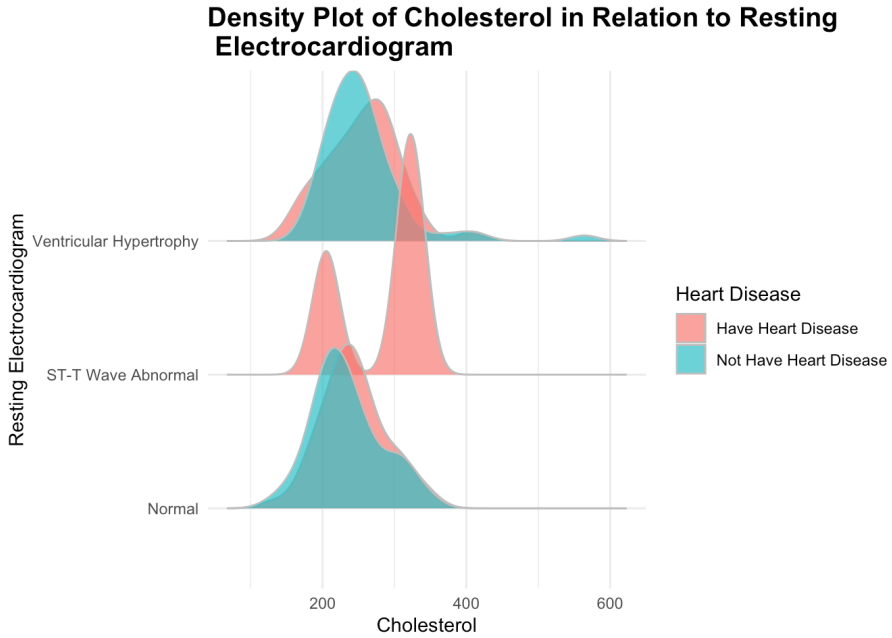

Figure 5: Density plot of cholesterol in relation to resting electrocardiogram.

Short Description

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.