Project Proposal

due October 11, 2021 by 11:59 PM

Shelby Brown, Katie Lam, Kaeden Hill

10/9/2021

Load Packages

```
library(tidyverse)
```

Load Data

```
heart <- readr::read_csv("heart.csv")
```

Introduction and Data, including Research Questions

(The introduction should introduce your general research question and your data (where it came from, how it was collected, what are the cases, what are the variables, etc.). Your research questions should be clearly specified. The motivation for your research question should be clear, with citations to relevant literature as appropriate.)

For this data visualiztion, we will be analyzing chest pain types and their relation to other physiological factors. We will be looking for an association between factors such as blood pressure, cholesterol, and exercise and the type of chest pain a patient experiences. We also plan on using this dataset to look for a possible association between chest pain type and whether that patient experiences heart disease. The dataset we have chosen is the Heart Failure Prediction Dataset. We retrieved it from Kaggle, and it is a compiled dataset from five sets with common variables. These sources are the Hungarian Institute of Cardiology. Budapest, University Hospital, Zurich, Switzerland, University Hospital, Basel, Switzerland, the V.A. Medical Center, Long Beach, and the Cleveland Clinic Foundation. The variables of interest include the age of the patient, gender, their resting blood pressure (mm Hg), their serum cholesterol (mm/d), and whether they have heart disease. We will also be using the categorical variables compiled on the chest pain type (typical angina, atypical angina, non-anginal pain, or asymptomatic), and whether the angina was exercise induced (yes or no).

Research Question: Can factors such as blood pressure, cholesterol, and exercise affect the chest pain type a patient experiences? Is a certain chest pain type more associated with heart disease?

fedesoriano. (September 2021). Heart Failure Prediction Dataset. Retrieved 10/09/2021 from https://www.kaggle.com/fedesoriano/heart-failure-prediction.

Glimpse

(Please use `glimpse` for your data, uploaded into the data folder, here.)

```r
glimpse(heart)
```

```
## Rows: 918
## Columns: 12
## $ Age            <dbl> 40, 49, 37, 48, 54, 39, 45, 54, 37, 48, 37, 58, 39, 49,~
## $ Sex            <chr> "M", "F", "M", "F", "M", "M", "F", "M", "M", "F", "F", ~
## $ ChestPainType  <chr> "ATA", "NAP", "ATA", "ASY", "NAP", "NAP", "ATA", "ATA",~
## $ RestingBP      <dbl> 140, 160, 130, 138, 150, 120, 130, 110, 140, 120, 130, ~
## $ Cholesterol    <dbl> 289, 180, 283, 214, 195, 339, 237, 208, 207, 284, 211, ~
## $ FastingBS      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ RestingECG     <chr> "Normal", "Normal", "ST", "Normal", "Normal", "Normal",~
## $ MaxHR          <dbl> 172, 156, 98, 108, 122, 170, 170, 142, 130, 120, 142, 9~
## $ ExerciseAngina <chr> "N", "N", "N", "Y", "N", "N", "N", "N", "Y", "N", "N", ~
## $ Oldpeak        <dbl> 0.0, 1.0, 0.0, 1.5, 0.0, 0.0, 0.0, 0.0, 1.5, 0.0, 0.0, ~
## $ ST_Slope       <chr> "Up", "Flat", "Up", "Flat", "Up", "Up", "Up", "Up", "Fl~
## $ HeartDisease   <dbl> 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1~
```

Data Analysis Plan

(Specify the outcome (response, Y) and predictor (explanatory, X) variables you will use to answer your question, as well as the comparison groups you will use, if applicable. You may include very preliminary exploratory data analysis, including some summary statistics and visualizations, along with some explanation on how they help you learn more about your data. Note the statistical method(s) that you believe will be useful in answering your question(s). What results from these specific statistical methods are needed to support your hypothesized answer?)