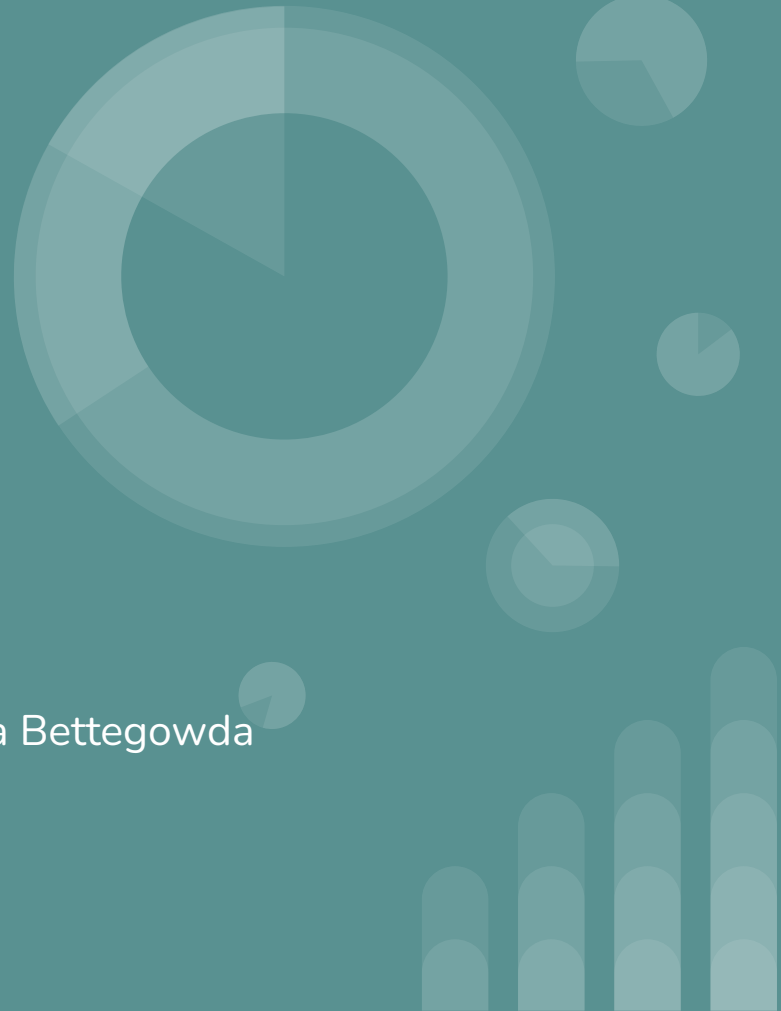


Heart Disease Analysis

Kethan Poduri | Carly Hubert | Oliver Hess | Anya Bettegowda





Topic/Motivation

Research Question: How does heart disease presence and magnitude vary based on the age, sex, cholesterol, blood sugar, and blood pressure? Furthermore, what variables are the best predictor of whether or not an individual will get heart disease?

- Heart Disease can come in many different shapes and forms, such as coronary artery disease, heart failure, and heart valve disorders
- Heart disease affects hundreds of thousands of people across the world.
- According to the CDC, heart disease is responsible for around 1 every 5 deaths
- Many of our group members have family affected by heart disease
- We wanted to figure out the best predictors of whether or not an individual will contract heart disease



Data Introduction

- Data initially collected at the **Hungarian Institute of Cardiology** and **Cleveland Clinic Foundation**
- The information was **donated to UC Irvine's Machine Learning Repository** (1988)
- Other relevant variables measure:
 - Type of chest pain
 - Resting electrocardiographic results
 - Maximum heart rate achieved
 - Exercise-induced angine, T/F
 - ST depression by exercise relative to rest
 - Slope of peak exercise ST segment
 - Number of major blood vessels colored by fluoroscopy

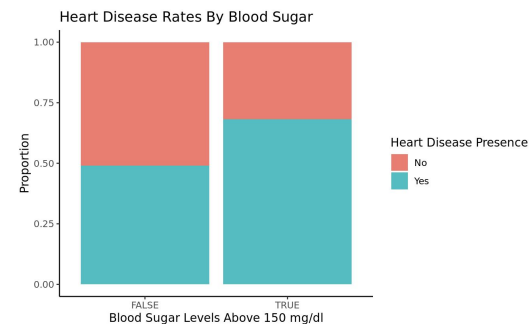
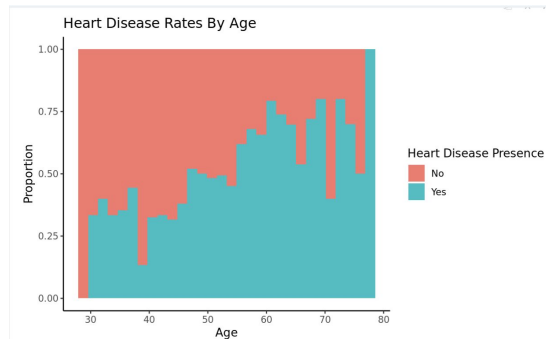
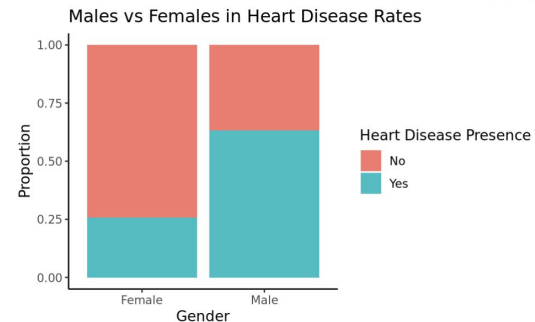
Relevant Variable	Description
<i>age</i>	The age of the patient in years
<i>sex</i>	The sex of the patient, male/female
<i>num</i>	Presence of heart disease, 0 means no heart disease, 1-4 is classified as heart disease. We masked to 1 or 0.
<i>trestbps</i>	Resting blood pressure (at hospital admission in mmHg)
<i>chol</i>	Serum cholesterol in mg/dl
<i>lbs</i>	If fasting blood sugar > 120 mg/dl



Highlights from EDA

Key takeaways from our visualizations:

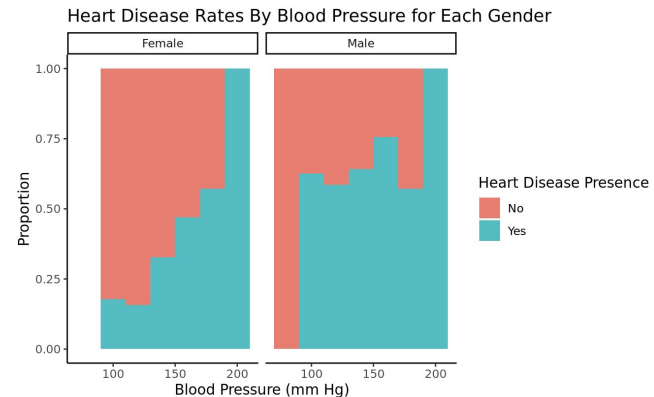
- The rate of heart disease is **higher for men than for women**
- Heart disease rates increase **as age increases**
- **People with fasting blood sugar above 150 mg/dl are more likely to have heart disease** than those with fasting blood sugar below 150 (mg/dl)





Highlights from EDA cont.

- The proportion of heart disease rates increases **as cholesterol increases**
- The proportion of heart disease rates increases **as blood pressure increases**



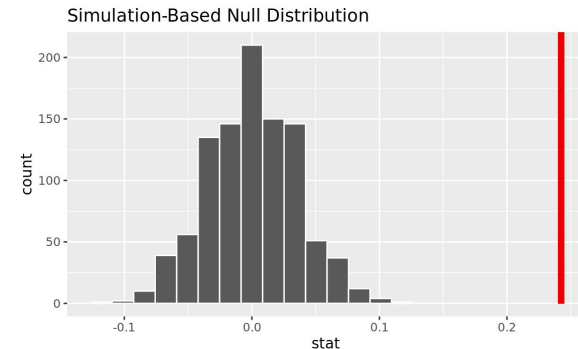
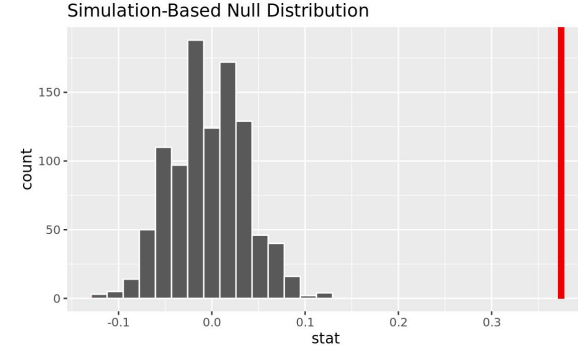


Inference/Modeling - Logistic Regression

- Logistic regressions and AIC
- **Best model:**
 - Additive of cholesterol and sex
 - AIC = 919.4695
 - $\log(\hat{p}/(1 - \hat{p})) = -2.76 + .006 * chol + 1.6 * sex$
- **Second best model:**
 - Interactive of cholesterol and age
 - AIC = 926
 - $\log(\hat{p}/(1 - \hat{p})) = -9.54 + .024 * chol + .162 * age - .0004 * chol * age$

Other Analysis - Hypothesis Testing

- **Hypothesis test 1:** Do males and females have the same rates of heart disease?
 - $H_0 : \pi_m - \pi_f = 0$ $H_a : \pi_m - \pi_f \neq 0$
 - P-value close to 0 - reject the null hypothesis, strong evidence to conclude that the proportion of men and women who get heart disease are not the same
- **Hypothesis test 2:** Do people over 50 have higher rates of heart disease than those 50 and under?
 - $H_0 : \pi_{over} - \pi_{under} = 0$ $H_a : \pi_{over} - \pi_{under} > 0$
 - P-value close to 0 - reject the null hypothesis, strong evidence to conclude that the proportion of people over 50 who get heart disease is greater than the proportion of people 50 and under who get heart disease





Conclusions and Future Work

Significant Conclusions:

- The rate of heart disease in men is significantly greater than in women
- **Cholesterol and sex are the two best predictors** of whether or not an individual will have heart disease
- People are significantly **more likely to have a heart disease past age 50**

Future Research:

- **Predict the magnitude** of heart disease
 - Right now it is a binary classification, we want it on a scale from 1-4
- Transition to **specific heart related diseases**
 - Ex: coronary artery disease, heart failure, or heart valve disorders