To Save a Heart

Henry Oluka, Paul Parkin, and Ryan Woodward

Abstract—Sudden cardiac arrests are among the leading causes of death each year. To improve the health outcomes of heart disease patients, signs leading up to a potential cardiac arrest need to be presented in such a way that they are easy to understand. By performing data analysis we have found correlative evidence between certain factors such as age, heart defects, exercise induced angina, and resting heart rates. An effective visualisation would visualize these factors and their relationships in a manner that improves readability for average citizens. Our project goal is to determine indicators that can be monitored and used to determine when health complications (as a result of heart disease) arises, and possibly contribute to further research and discoveries on this topic. This study promise to provide Physicians with another tool for quickly visualizing diagnostic test data when caring for their patients. We believe that such a visualisation can aid a physician's judgment and decision-making especially in cardiac arrest cases and potentially improve health care delivery and patient health outcomes.

Index Terms—Health, heart disease, heart and hypertension, biology

1 Introduction

About 2.4 million Canadian adults aged 20 years and above are living with heart disease; the second leading cause of death in Canada [1]. This makes heart diseases a threat to Canadians and a financial burden to the healthcare system in Canada. Not all heart diseases are fatal. However, they can lead to hospitalization, cardiac arrests and other complications. Cardiac arrests can occur as a result of various underlying factors including the presence of heart disease.

This project is extrinsically motivated by our team-member's loss of a family member to heart complications. The focus is on determining patterns in heart diseases that can help physicians effectively care for patients presenting patterns of a fatal complication of a heart disease. These patterns can also help physicians prevent fatal heart well in advance of the event.

The approach we used to conduct this research analysis was methodological. We used a modified version of the prominent SDLC methodology and incorporate the best practices of modern DevOps (which is continuous integration and continuous delivery) to tackle this project. The phases of our approach are: Phase 1 (Requirements Collection & Analysis), Phase 2 (Design), Phase 3 (Development - Continuous integration and Continuous delivery), Phase 4 (Final Presentation and Report).

For data analysis and visualization we used MS Excel and Microsoft Power Bi, respectively. We hope to see how selective and associative data can be interactively visualized to model the sequence of interactive activities a physician would perform, when diagnosing patients.

2 Motivation, Data, and Data questions

Our motivation into heart disease attributes is the goal of determining patterns within fatal heart attacks through analysis and visualization of research data from 303 patients.

- *Ryan Woodward is a UVic Computer Science, Software Engineering undergraduate student. Email: rnw@uvic.ca.*
- *Henry Oluka is a UVic Computer Science, Health Information Science and Statistics undergraduate student. Email: holuka@uvic.ca.*
- *Paul Parkin is a UVic Computer Science undergraduate student. Email: paulparkin@uvic.ca.*

2.1 Data

The research data we use in this study can be downloaded publicly from https://www.kaggle.com/ronitf/heart-disease-uci as a .csv file. There are 303 data cases and 14 data dimensions in the relational dataset. It is collected from two cardiology institutes and two universities. The data dimensions in the dataset describes patients by age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting cardio graphic maximum heart rate achieved, exercise induced angina, old-peak, slope peak exercise, number of major vessels, and thal.

2.2 Data Questions

Through the analysis and visualization of the research data, we intend to answer three data questions.

2.2.1 Are there emerging patterns for patients who experience cardiac issues at a younger age?

This question will require the grouping of the values for the age data dimension into 4 categories, one of which will be the youngest age category. The youngest age category will be visualized in comparison to the other categories to identify any patterns or relationship with the remaining data dimensions.

2.2.2 Are there warning signs that a cardiac arrest is imminent given the pain described and the underlying symptoms?

This question requires analysing the dataset for data dimension values that indicate signs of a cardiac arrest. The question requires the "human-in-the-middle" in making connections between the pain and the symptoms described by a cardiac arrest patient, and the data dimension values that correspond to a cardiac arrest patient's health status.

2.2.3 Is there a threshold at which a patient needs invasive surgery, or can surgery be avoided?

This question requires the "human-in-the-middle" to make decisive judgements on what benchmarks should be ascribed to each data dimension. These benchmarks will further group values for a given dimension into categories (mild, worse, and severe) depending on the severity of the health status the value represents. This should allow the identification of thresholds (per data dimension) that indicate a patient needs invasive surgery.

3 Related work

Jae Duk Seo has performed some basic data exploration using the python programming language interfacing with pandas dataframes and numpy algorithms [2]. Using data exploration Seo has been able to separate categorical and quantitative value. He has also been able to provide a variance/covariance matrix which observes correlative relationships between the attributes [4]. Simple visualizations have been provided, however they are limited to bar graphs, box plot charts, and x-y scatter plots. Using these limited visualizations provides the opportunity to discover which values could be selective and associative.

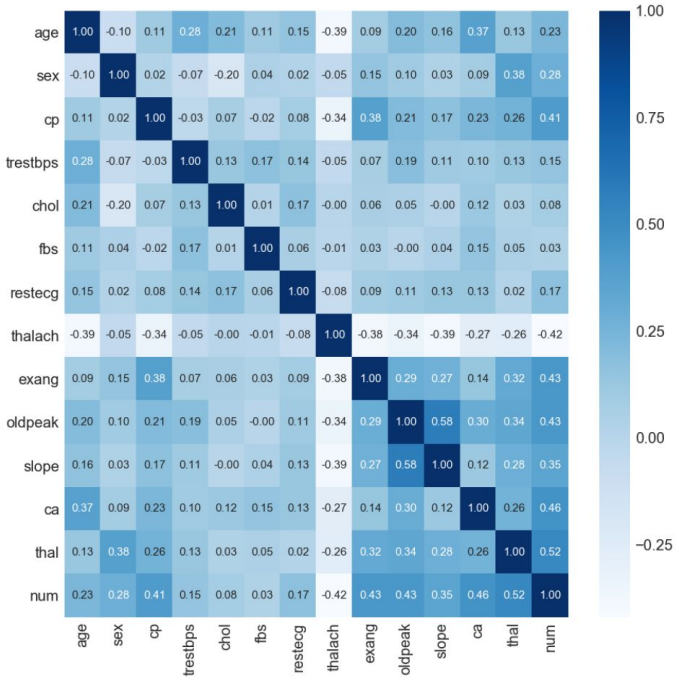

Figure 1: Correlation Matrix

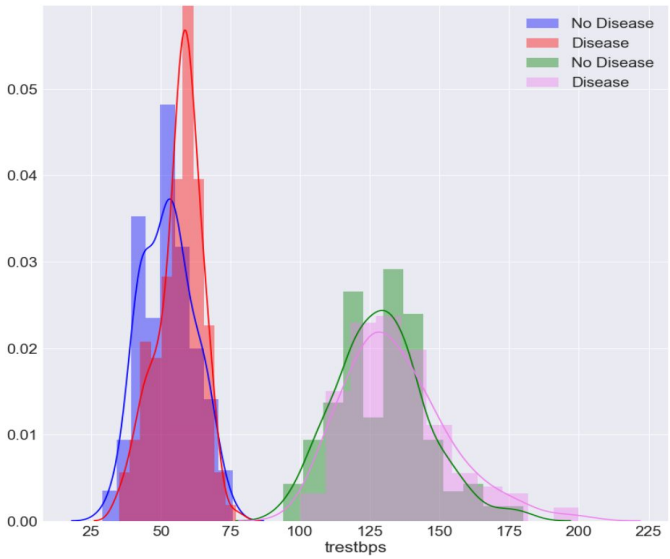

Figure 2: Disease vs Base Heart Rate

The prevalence of blood pressure status in US adults has also been used as a preliminary visualization [3]. The stacked bar graph provides a simple tool to view the percentage of US adults that fall into the three blood pressure categories: Normotension, Prehypertension, and Hypertension. The stacked bar graph visualization provides an easy look into blood pressure, but doesn't address the overall big picture that we are attempting to achieve.

3.1 Alternative Solutions

There exists great potential on expanding upon the existing related visualizations. By examining the correlation matrix it provides us with insight into whether the variables can be considered associative. The data analysis also provides information to categorize the data appropriately as quantitative, nominal, and ordinal. This analysis lends itself to a visualization that focuses on position, color saturation, and color hue. The length of certain data dimensions such as heart rate vs cholesterol will have to be normalized in order to maintain integrity of the data. Alternative visualizations will likely make use of the length function in order to reach a length threshold. This is to say that if the length of these data dimensions were stacked they would exceed the threshold which in turn would indicate that the patient is at a certain risk level. The risk level thresholds could in turn be expanded upon to indicate correlative actions to be taken by the patient or physician.

References

[1] Government of Canada. Heart disease in Canada (2017). https://www.canada.ca/en/public-health/services/publications/diseases-conditions/heart-disease-canada.html Accessed March 2, 2020.

[2] Basic Medical Data Exploration / Visualization - Heart Diseases (2018) https://towardsdatascience.com/basic-medical-data-exploration-visualization-heart-diseases-6ab12bc0a8b7 Accessed February 26, 2020

[3] Vise, Sharlotte. "Prevalence of Blood Pressure Status in US Adults by Age Category: NHANES." *Health Metrics*, 18 June 2019, healthmetrics.heart.org/prevalence-blood-pressure-status-us-adults-age-category-nhanes/.

[4] Mohiteud. "Heart Diseases Visualization." *Kaggle*, Kaggle, 27 Mar. 2019, www.kaggle.com/mohiteud/heart-diseases-visualization.