Heart Disease Diagnostic Analysis and Visualization

Table of Contents

Problem Statement

As it is rightly said, 'Health is Wealth'. We have realized this fact in the pandemic time after witnessing the brute effects of Covid-19 on people of all age groups. Apart from this, another major contributor to the death rate is heart related diseases.

Located in the chest region of the body, the heart beats at around 80 times per minute. Even though it is just the size of an average human fist, it is the strongest muscle which continuously to pump blood to the body organs, even at rest.

Heart diseases have been known to take a major toll of people's lives. As a layman, we may feel that the common factors for heart related diseases are cardiac arrest or blockages. But the dataset under analysis describes multiple different medical parameters associated with the heart and their typical values. We will be analyzing the relationships between them and study the implications of changes in those parameters. In this project, we will be incorporating most trending and powerful BI tool namely Tableau.

Objective and Benefits

- Objective:
  - The dataset contains the records for the patients and their medical parameters details and the target variable whether they will suffer from heart disease or not.
  - The aim of this project is to use the given data and perform ETL and data analysis to infer key metrics and patterns in the dataset. In addition to this, different visualizations are developed to depict meaningful relationships.

- Benefits:
  - The data analysis will reveal some common and unique patterns in the dataset related to the medical parameters.
  - Data visualizations will enhance the understanding of the effect of the high or low of these features on the chances of heart rate and give a better chance of prediction.

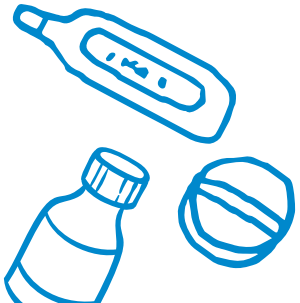
Steps Followed

Data Extraction

Data Exporting

Data Analysis

01 02 03 04 05 06



Data Preprocessing

Data Loading and Modification

Deployment

Steps Followed contd...

1) Data Extraction: This step involves extracting the data from different sources relevant to the problem statement or obtaining data from the client.

2) Data Preprocessing: Once the raw data is obtained, we need to ensure that the data is free from errors. We perform Exploratory Data Analysis followed by Data Cleaning which involves imputing missing values, removing duplicates, finding anomalies or outliers and treating them.

3) Data Exporting: The preprocessed data is exported to a .csv file to be used for analysis.

4) Data Loading and Modification: The preprocessed data in .csv file is loaded into the Tableau Desktop for analysis purpose and modified for simplicity purpose.

5) Data Analysis: Once the data is loaded, we perform the data analysis using Tableau features and store the visualizations in Tableau worksheets.

6) Deployment: The prepared visualizations are deployed on the Tableau Online Software where they will be available publicly.

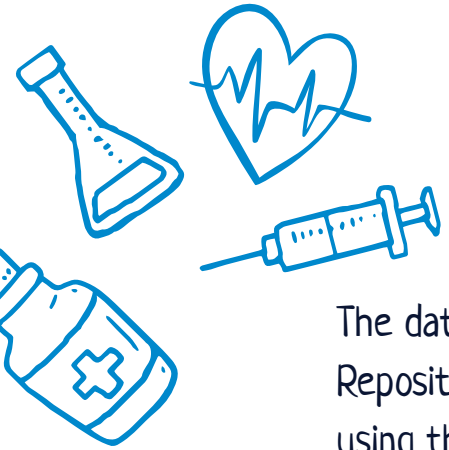
Data Sharing Agreement

Dataset File Name: 'heart_disease_dataset.csv'

Number of columns: 14

Column Names

Column Data Types

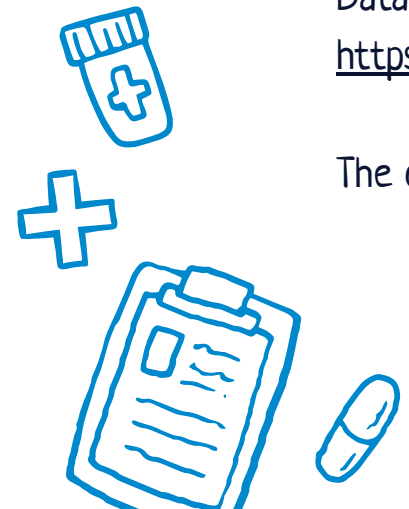


Data Extraction

The dataset used for analysis is the Heart Disease dataset provided by the UCI Repository. It actually contains 76 attributes out of which only 14 are used. We will be using the Cleveland dataset.

Dataset source:
https://archive.ics.uci.edu/ml/datasets/Heart+Disease

The dataset is available is a .csv file - 'heart_disease_dataset.csv'



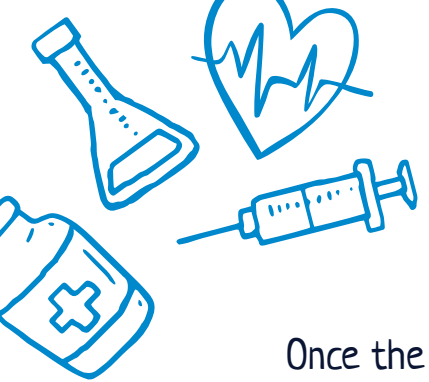
Data Preprocessing

After Exploratory Data Analysis carried out on the dataset we have certain observations with the dataset.

1) There is no column in the dataset with missing values.

2) There are a few columns which actually contain categorical values but have been incorrectly labeled as numeric. As a part of data preprocessing we will convert them to categorical values.

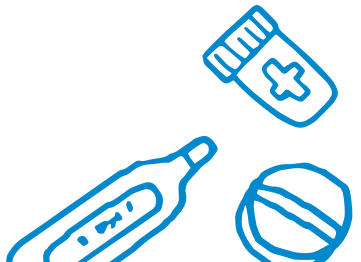
3) There are a few columns which have unusual values / outliers. We will impute these values with the median / mode value obtained from the remaining values of the columns.



Data Exporting

Once the data has been cleaned in the data preprocessing stage, we will export the cleaned dataset into a new file with .csv format.

The new dataset file has name - 'preprocessed_heart_disease_dataset.csv'



Dataset Loading and Modification

The exported .csv dataset file – 'preprocessed_heart_disease_dataset.csv' will be imported into Tableau Public Desktop. Since this a .csv file, we will choose the 'Microsoft Excel' file option when prompted to import dataset into Tableau.

Since the dataset contains many categorical columns which store the categories in the form of integers we will convert these numbers into meaningful phrases which will be understandable to the viewer and also easy to understand the terms used in the visualizations.

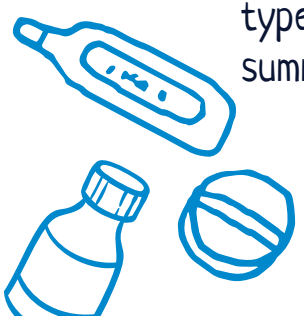
These phrases are called as 'Aliases' and will be provided to the values available in the categorical columns as part of data modification.

Data Analysis

Once the data has been loaded into the Tableau Desktop software, we perform the analysis for the various medical parameters provided in the dataset and study the relationship between them.

Based on these patterns, we try to draw approximate inferences about the data provided on the basis of visualizations created.

We have made use of different aspects of Tableau like different charts, labeling, aliases, filtering, and actions based on user choice. We have created separate worksheets for each type of visualization which contains the chart and a caption as well which contains the summary of analyses drawn.



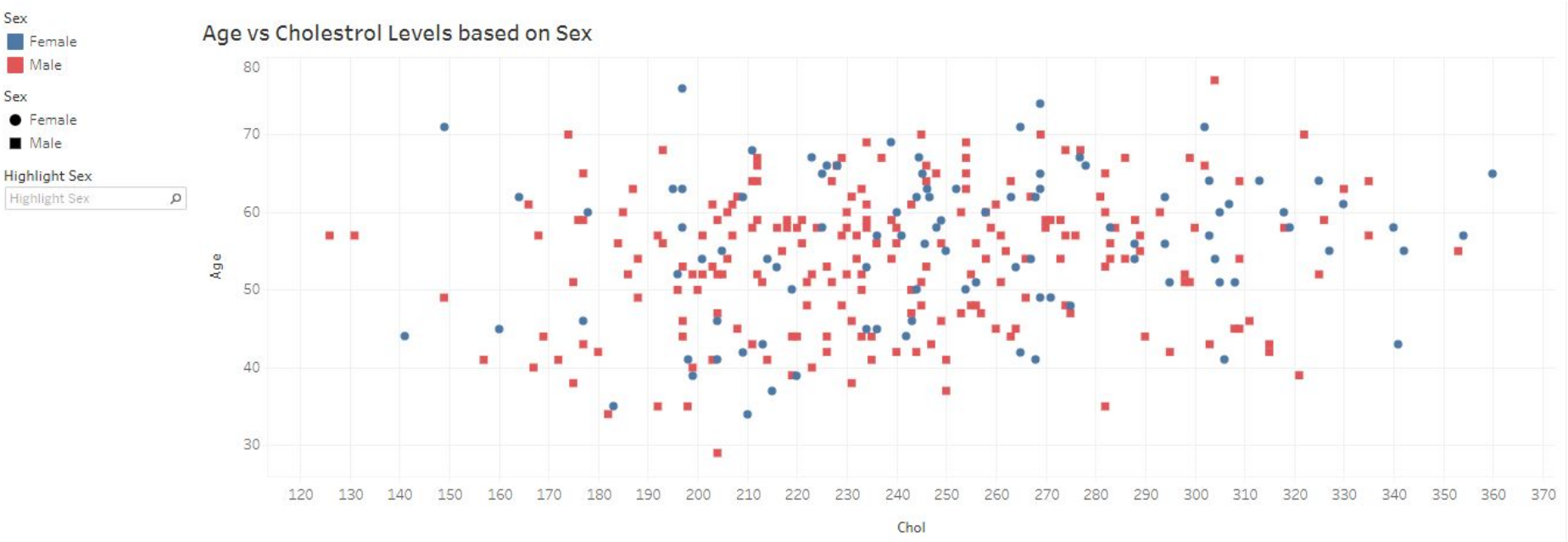
Deployment

All the different worksheets that have been created are compiled together into a Tableau workbook. Each worksheet is named based on the type of visualization performed in the chart.

When we save all the worksheets on Tableau Desktop software, it connects to the Tableau Public Software via personal email id and credentials. All these sheets are uploaded onto the Tableau Public Software on personal profile and this viz. is visible to public.

The link for the worksheets is at :
https://public.tableau.com/views/HeartDiseaseDiagnosticAnalysis/AgeandChol?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Visualizations



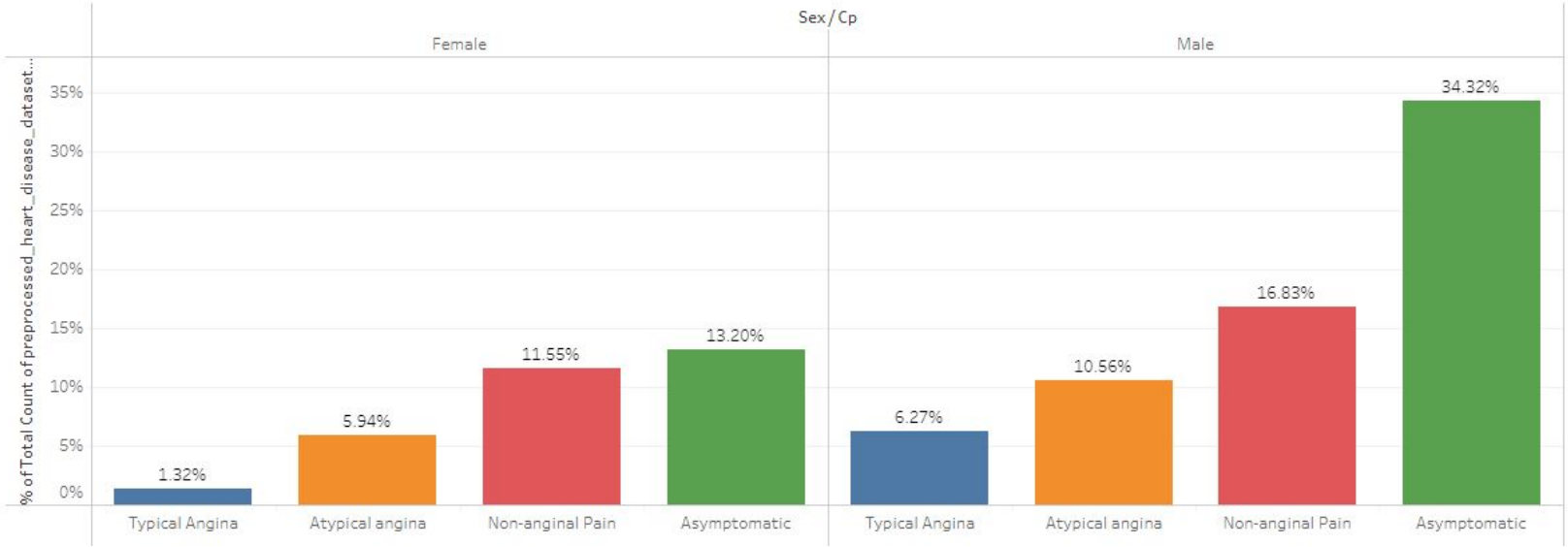
Highlight Cp

Highlight Cp

Cp

- Typical Angina
- Atypical angina
- Non-anginal Pain
- Asymptomatic

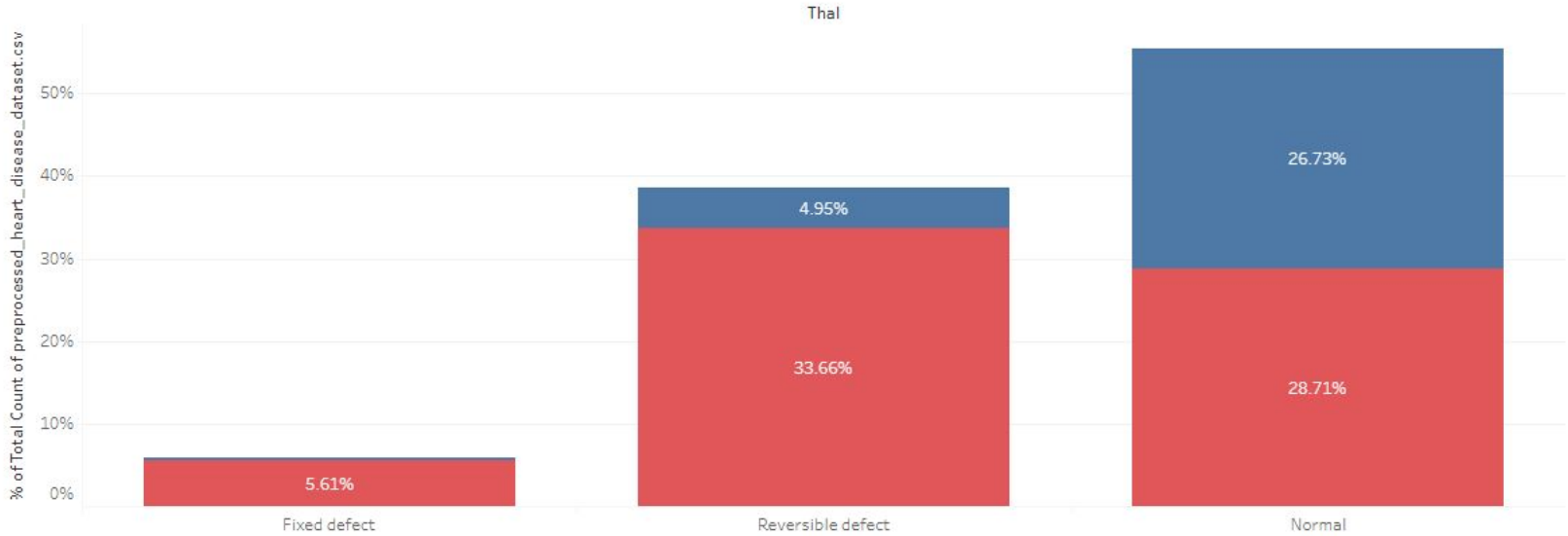
Variation of Chest Pain vs Sex

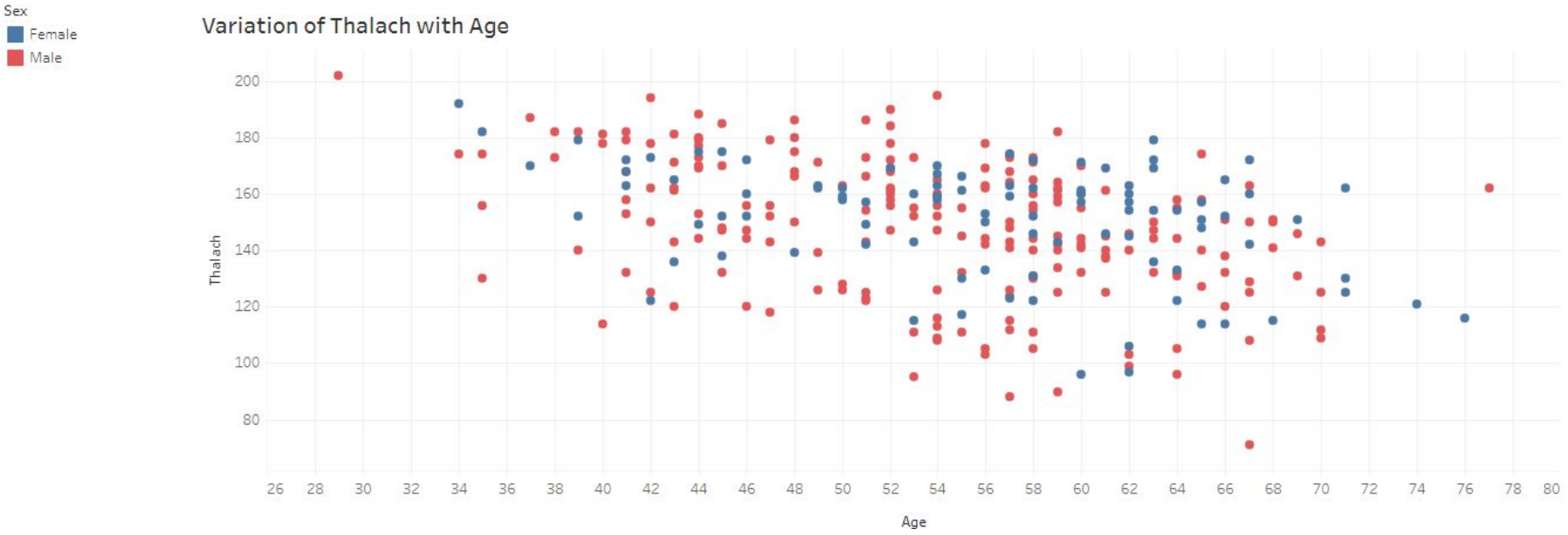


Sex
Female
Male

Highlight Thal
Highlight Thal

Variation of Thal with Sex

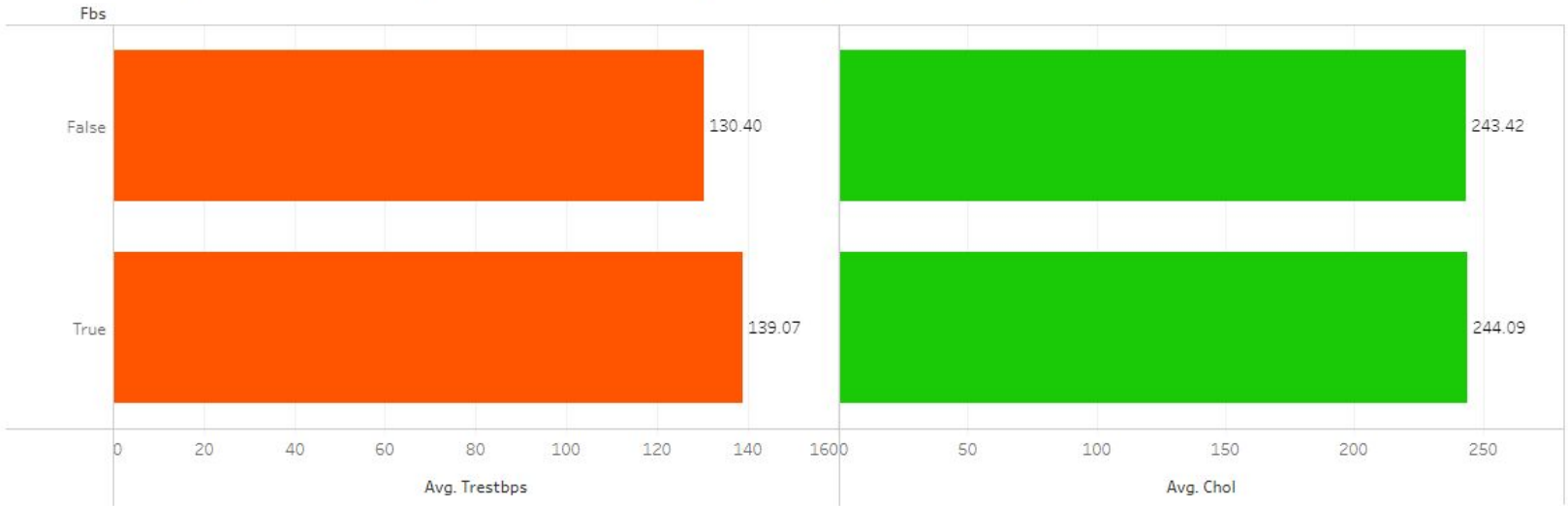



Sex
Female
Male
Variation of Thalach with Age
Thalach
200
180
160
140
120
100
80
26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80
Age

Highlight Fbs

Highlight Fbs

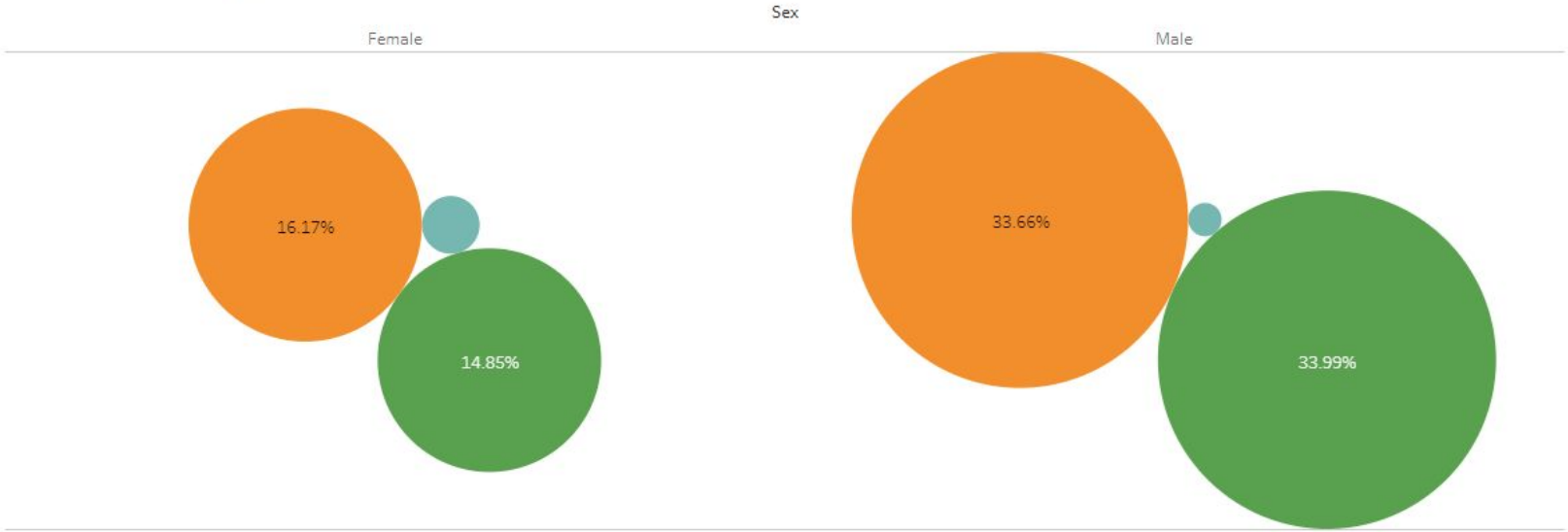
Variation of Avg. Cholestrol and Avg. TrestBps with Fasting Blood Sugar



Restecg
Normal
ST-T wave abnormality
Left ventricular hyper...

Highlight Sex
Highlight Sex

Variation of RestEcg with Sex

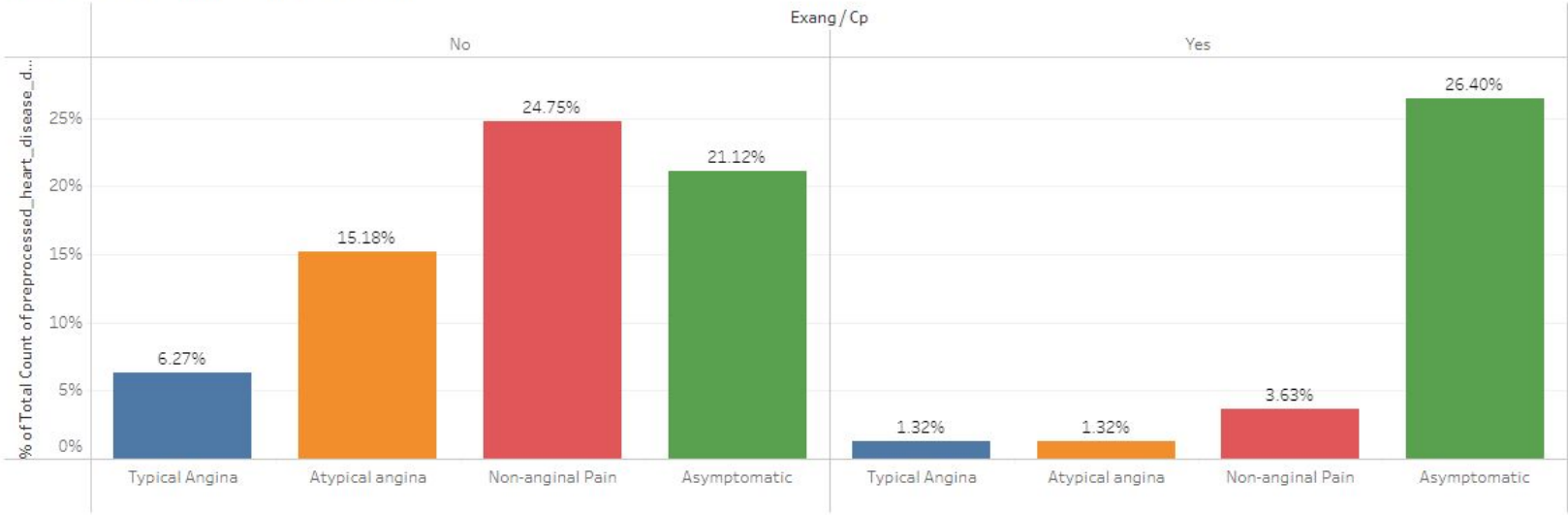


Cp
- Typical Angina
- Atypical angina
- Non-anginal Pain
- Asymptomatic

Highlight Exang

Highlight Exang

Variation if Exang with Chest Pain



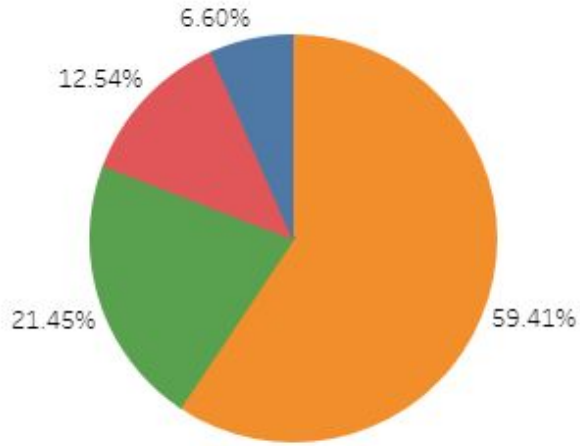
Restecg
Normal
ST-T wave abnormality
Left ventricular hyper...

Highlight Restecg
Highlight Restecg

Variation of Slope and Restecg with Oldpeak

Ca
0
1
2
3

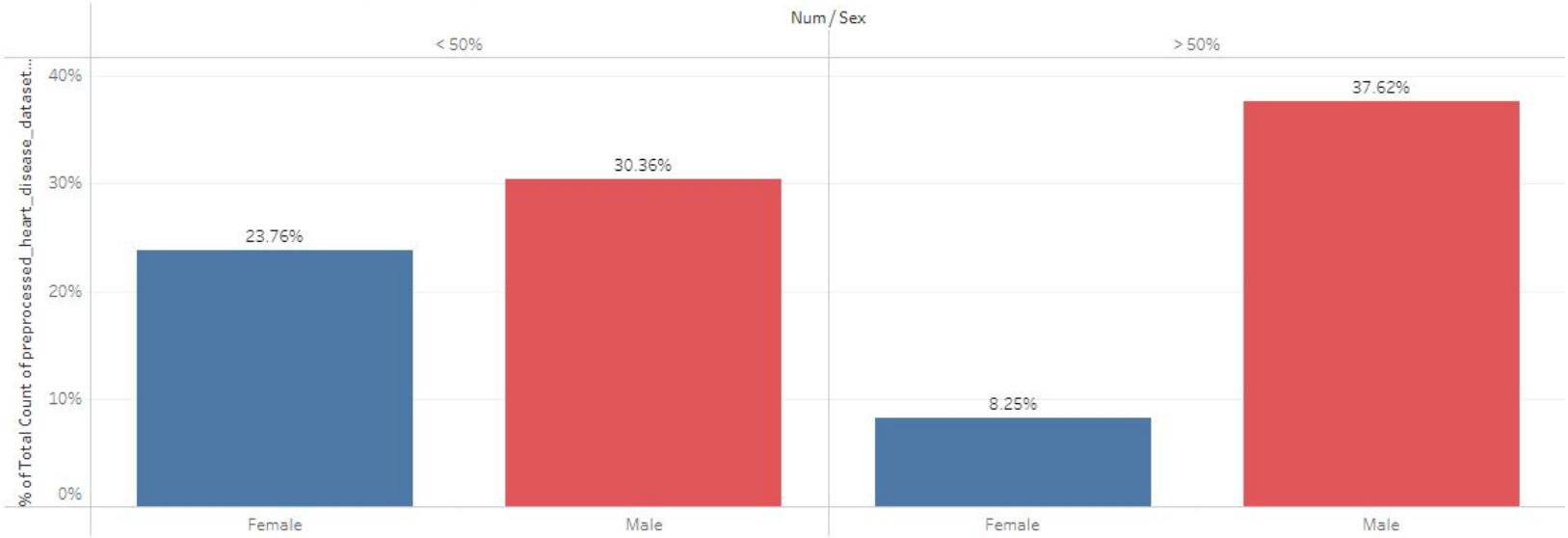
Distribution of Major vessels



Sex
Female
Male

Highlight Num
Highlight Num

Variation of the Num variable against Sex



THANK YOU