

# Data Analysis and Visualization of Heart Disease

**Syed Irteza Haseem**

Rutgers University, Newark, NJ 07102, USA

## **Abstract:**

Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias) and heart defects you're born with (congenital heart defects), among others.

I chose this topic since heart disease predominantly are becoming abundant globally. Heart strokes is one of the biggest causes of mortality among the population of the world. But it is difficult to identify heart disease because of many factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate etc. Therefore, data analysis of heart disease is regarded as the most important subjects in the section of clinical data analysis. In this project I will be performing exploratory data analysis and visualization to analyze and understand whether a person is suffering from heart strokes or not, using the dataset based on the given information about each individual.

## **Outline of the Report:**

Graph 1: Analysis of Heart disease by Age Group.

Graph 2: Analysis of variation of age for each target class

Graph 3: Distribution of Age with Sex for each target class.

Graph 4: People suffering from heart disease across the globe.

Graph 5: Factors contributing to people who are suffering from heart disease.

Graph 6: Analysis of factor variation among the age groups.

## **Introduction:**

### *Goal:*

- To understand and visualize the data collected from people whether they are suffering from the disease or not.
- To analyze the causes of the heart diseases

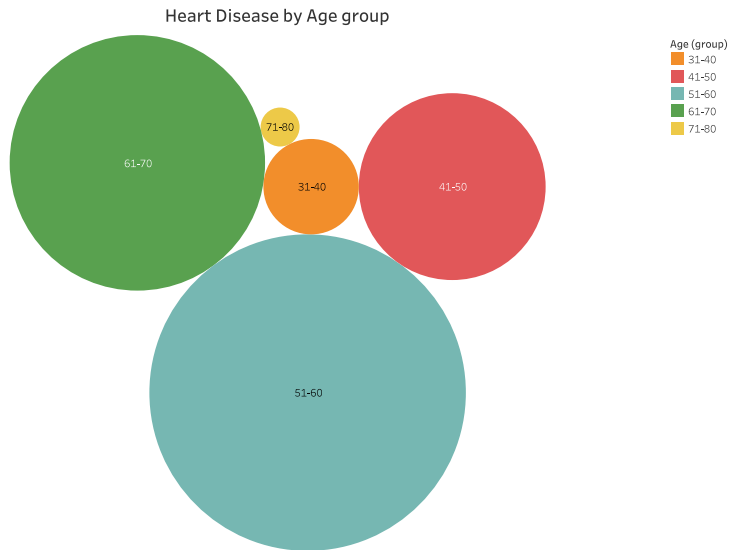
### *Dataset:*

For my final project I explored a lot of datasets for analysis on heart disease on google and I decided to use the heart strokes prediction dataset from Kaggle. I did so since the dataset size is manageable, and I have the correct variable types I wanted for my visualization so as to get a clear output for the goal of this project. The dataset consists of 303 individual's data. There are 15 columns in the dataset

S.No	Attribute Name	Description	Range of Values
1	Age	Age of the person in years	29 to 79
2	Sex	Gender of the person [1: Male, 0: Female]	0, 1
3	Cp	Chest pain type [1-Typical Type 1 Angina 2- Atypical Type Angina 3-Non-angina pain 4-Asymptomatic)	1, 2, 3, 4
4	Trestbps	Resting Blood Pressure in mm Hg	94 to 200
5	Chol	Serum cholesterol in mg/dl	126 to 564
6	Fbs	Fasting Blood Sugar in mg/dl	0, 1
7	Restecg	Resting Electrocardiographic Results	0, 1, 2
8	Thalach	Maximum Heart Rate Achieved	71 to 202
9	Exang	Exercise Induced Angina	0, 1
10	OldPeak	ST depression induced by exercise relative to rest	1 to 3
11	Slope	Slope of the Peak Exercise ST segment	1, 2, 3
12	Ca	Number of major vessels colored by fluoroscopy	0 to 3
13	Thal	3 – Normal, 6 – Fixed Defect, 7 – Reversible Defect	3, 6, 7
14	Target	Class Attribute	0 or 1
15	Location	Data collected from various parts of the world	String

### Analysis of Visualizations:

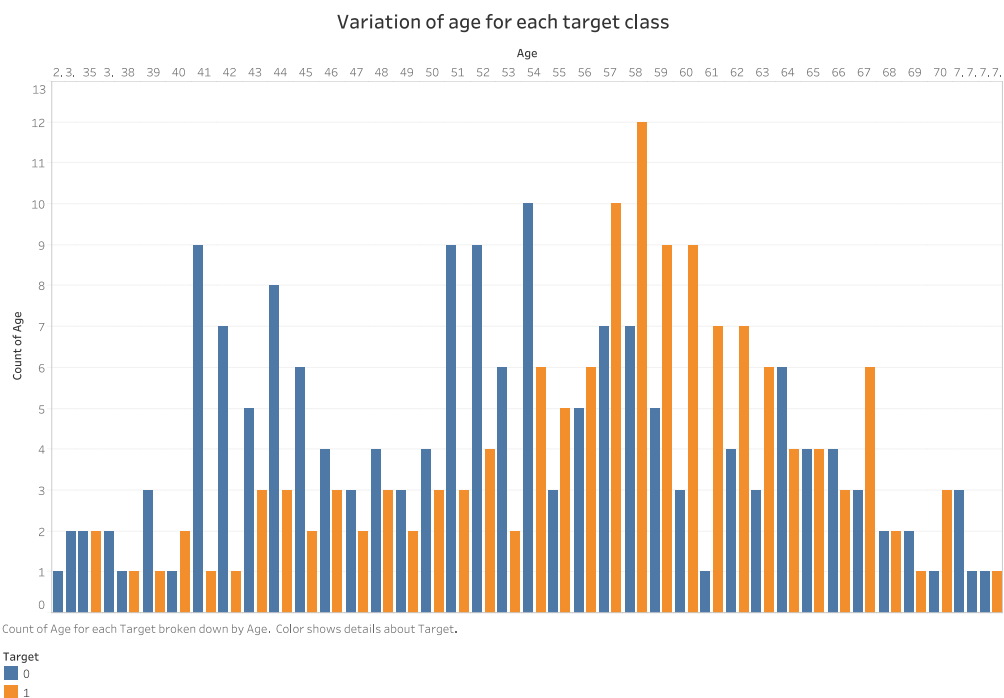
To do this first we are going to analyze our dataset's variables to gain insights about the information collected and perform exploratory data analysis.



Age (group). Color shows details about Age (group). Size shows count of Target. The marks are labeled by Age (group). The data is filtered on Target, which keeps 1.

First I created age groups in tableau to help us understand what age groups are more prone to heart diseases. Now based on the data I have filtered the age group on target=1 which denotes the number of people who are suffering currently from heart diseases. As we can clearly see a large portion of the population in the age of 50's and 60's fall under this category.

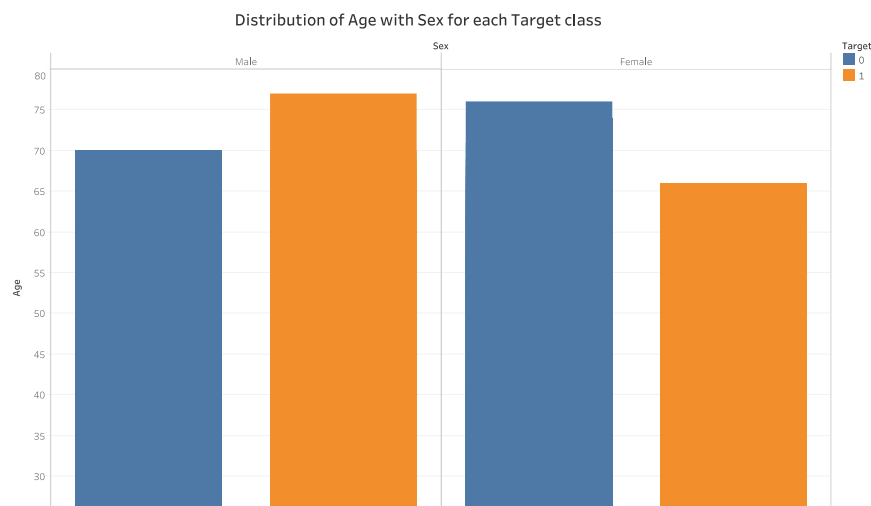
Now, let us look at the number of people age-wise who are suffering from the strokes or not. Here, target = 1 implies that the person is suffering from heart strokes and target = 0 implies the person is not suffering.



Count of Age for each Target broken down by Age. Color shows details about Target.

From above we can infer that most people who are suffering are of the age of 58, followed by 57. Majorly, people belonging to the age group 50+ are suffering from the strokes. Therefore, this analysis compared to the above pie chart is more distinct and clear to understand the exact age number of people suffering from the heart disease.

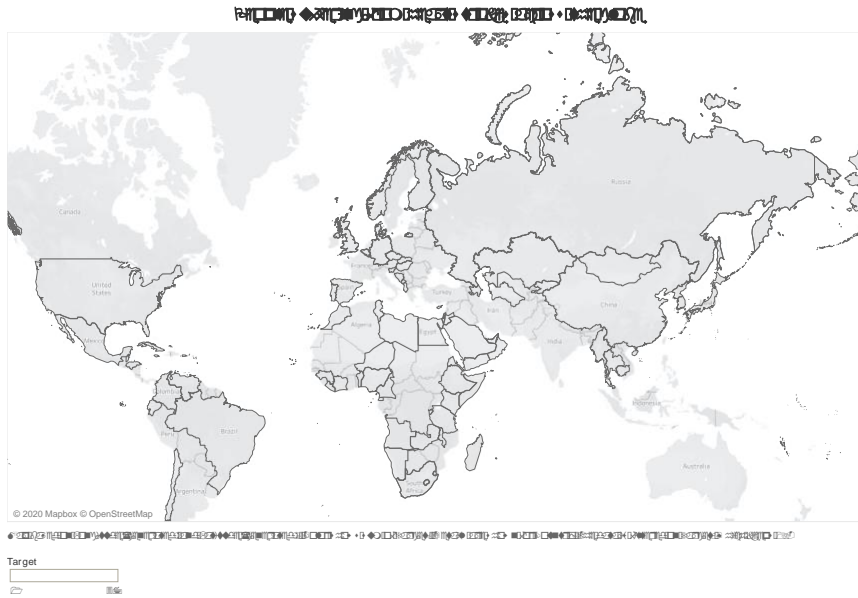
Next, to further enhance our analysis let us look at the distribution of age and gender for each target class.



Age for each Target broken down by Sex. Color shows details about Target. The view is filtered on Age, which ranges from 29 to 77.

From above we can infer that men who are suffering from the strokes are older than females and vice versa.

In the map provided below we can visualize the total number of people suffering from heart strokes by the country. I have put the countries in color so that we can observe the difference between the countries that have the greatest number of people affected with respect to the color they are represented with. For e.g. United States has the highest number of people affected in the world, which is not surprising since In the United States, someone has a heart attack every 40 seconds according to the CDC website. Coronary heart disease is the most common type of heart disease, killing 365,914 people in 2017. One person dies every 37 seconds in the United States from cardiovascular disease. About 647,000 Americans die from heart disease each year that's 1 in every 4 deaths. The most logical reason could be High blood pressure, High cholesterol, Cigarette smoking, Diabetes, Unhealthy diet and physical inactivity, Overweight and obesity etc. Over time, these risk factors cause changes in the heart and blood vessels that can lead to heart attacks, heart failure, and strokes. It is critical to address risk factors early in life to prevent these devastating events.



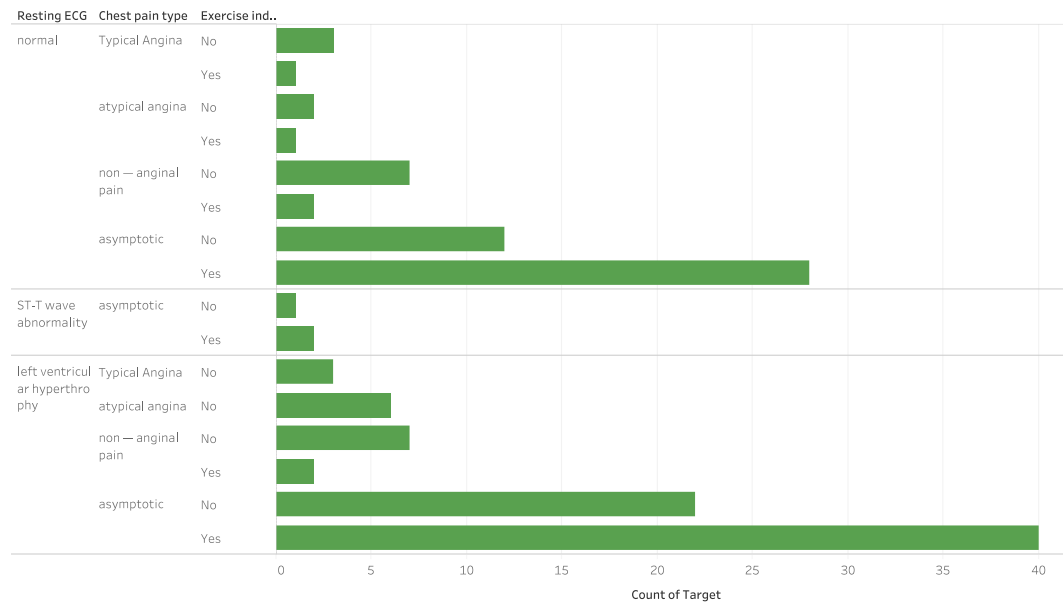
So, what really causes heart diseases?

Four types of chest pain can be considered as indicative of heart disease. Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen-rich blood. It may feel like pressure or squeezing in your chest.

The Type 1 angina is caused by reduced blood flow to the heart muscles because of narrowed coronary arteries. Type 2 Angina is a chest pain that occurs during mental or emotional stress. Non-angina chest pain may be caused due to various reasons and may not often be due to actual heart disease. The fourth type, Asymptomatic, may not be a symptom of heart disease. Resting ECG is the resting electrocardiographic result. Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.

Exang is the exercise induced angina which is recorded as 1 if there is pain and 0 if there is no pain. Therefore for our analysis perspective I have filtered the above parameters where the target is 1 which implies that the person is suffering from heart strokes

Factors contributing to people who are suffering from heart strokes

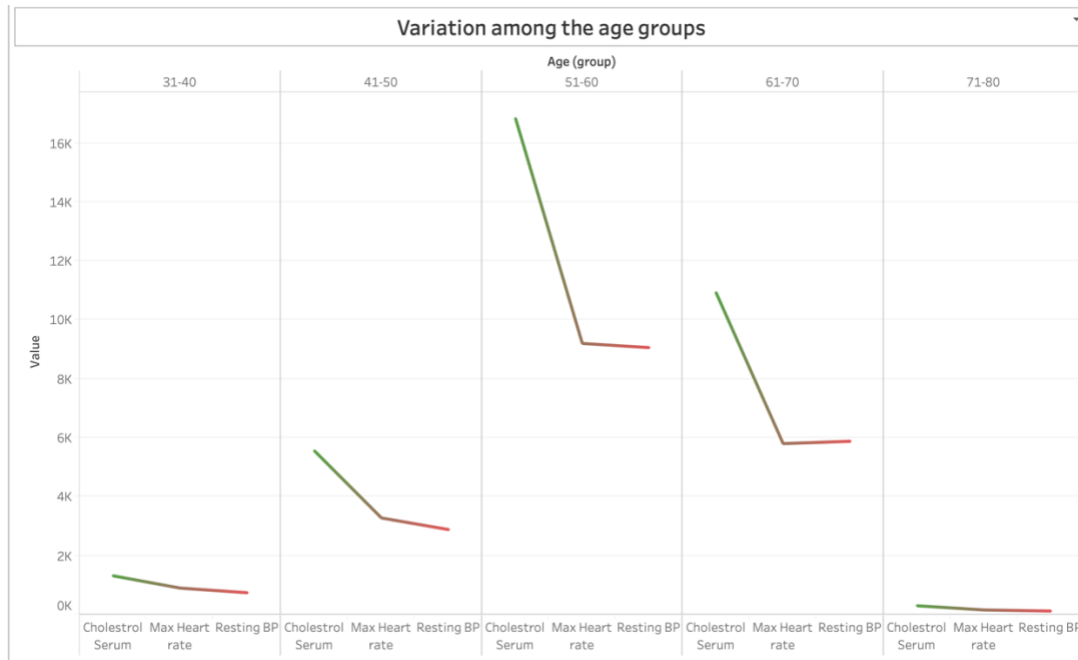


Count of Target for each Exercise induced angina broken down by Resting ECG and Chest pain type. The data is filtered on Target, which keeps 1.

So to further analyze the different factors contributing the heart diseases, I have visualized a line graph among the age groups to get a clear distinction to foresee which age group are more likely to get affected in future based on our data. The below graph is filtered on the target, which is equal to 1, implying that the data shows the people who are suffering from heart disease.

A high level of low-density lipoprotein (LDL) cholesterol which is known as the bad cholesterol is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of a heart attack. The increase in cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.

Therefore, we can analyze that the age group 51-60 who have high cholesterol, high heart rate and high resting blood pressure are the ones who are more prone to the disease.



### Conclusion:

Heart Disease is one of the major concerns for society today. It is becoming abundant and omnipresent across the globe. It is the most urgent threat to our life. It is difficult to determine the odds of getting heart disease based on risk factors. However, based on data analysis and visualization of our dataset we can now predict in future whether or not a person is likely to get affected by this disease based on these factors. This will help not only the people but also the doctors in hospitals to use such information and spread awareness among people the right way to live a healthy lifestyle and practice a healthy diet and exercise.

### References:

<https://www.kaggle.com/ronitf/heart-disease-uci>

<https://towardsdatascience.com/heart-disease-prediction-73468d630cfc>