Data Analysis of Factors That Could Predict Heart Attack

Yijie Nie

College of Arts and Sciences, Boston University, Boston, Massachusetts 02215, U.S.A

yjnie_27@163.com

Keywords: Heart attack, Prediction, Data analysis

Abstract: Cardiovascular diseases have been a global focus for decades of years. Heart attack, one of the possible signs of cardiovascular diseases, is also a concern. This study aims to show whether there are risk factors that could predict heart attack. Specifically, it investigates whether there is a relationship between variables and the possibility of having a heart attack. To test the hypothesis that five variables-age, gender, cholesterol value, resting electrocardiographic results, and the chest pain types-could predict heart attack, a dataset containing 303 subjects collected from four databases was analyzed by Tableau. The results showed that part of the hypothesis was true-gender, resting electrocardiographic results, and the chest pain types correlate with the probability of heart attack.

1. Introduction

Cardiovascular diseases are one of the top ten causes of death around the world. Heart attack belongs to cardiovascular diseases and is worth studying. Both causes and treatments of heart attack are important to research. However, optimally, people should take steps ahead of time and nip heart attacks in the bud. Therefore, it is necessary to pay attention to indicators that relate to the probability of heart attack. To find the factors, the writer analyzes variables in the dataset by using the data visualization tool-Tableau.

The original dataset contains 76 attributes. It was collected by Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D., University Hospital, Zurich, Switzerland: William Steinbrunn, M.D., University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D., and V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D. in 1988. It is denoted on UCI Machine Learning Repository on the first of July in 1988 [1]. This report will use a subset of 14 attributions and charts to show findings [2].

2. The Analysis Procedure and Graphs of the Dataset

2.1 Age is the First to Concern

The first variable people focus on is age. Figure 1 uses box-and-whisker plots that show output-less or more chance of heart attack–on the x-axis, and age plotted on the y-axis. Each dot represents a patient with a specific age. There are two outliers in the left box-and-whisker plot, which should be omitted.

The writer cares about the relationship between age and the probability of having a heart attack. The graph shows that the range of age is wider for patients who have a higher chance of a heart attack. The difference between the maximum and minimum age values of the right graph is larger than that of the left graph. Besides, the upper quartile of the box-and-whisker plot related to age and more chance of heart attack is at about 60 years old. The possibility of having a heart attack is higher for people who are younger than 60 years old. Therefore, not only the old need to pay much attention to heart health but also young people.

  DOI: 10.25236/icbcme.2021.040

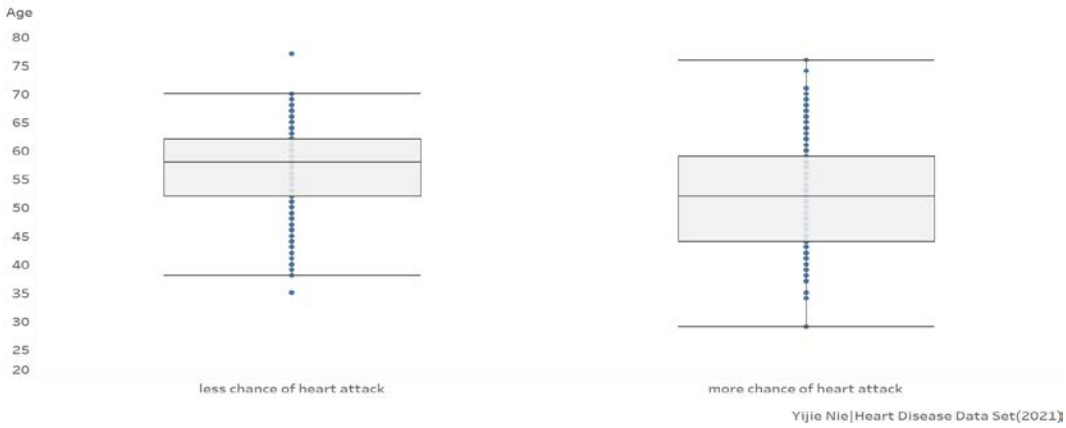

Fig.1 Box-and-Whisker Plots That Show the Relationship between Age and Heart Attack

2.2 Gender and the Possibility of Heart Attack

Another variable people often take care of is gender. In Figure 2, each gender as a whole is represented by a pie chart. The blue color means the patient is less likely to have a heart attack, and the orange color means the patient is more likely to have a heart attack.

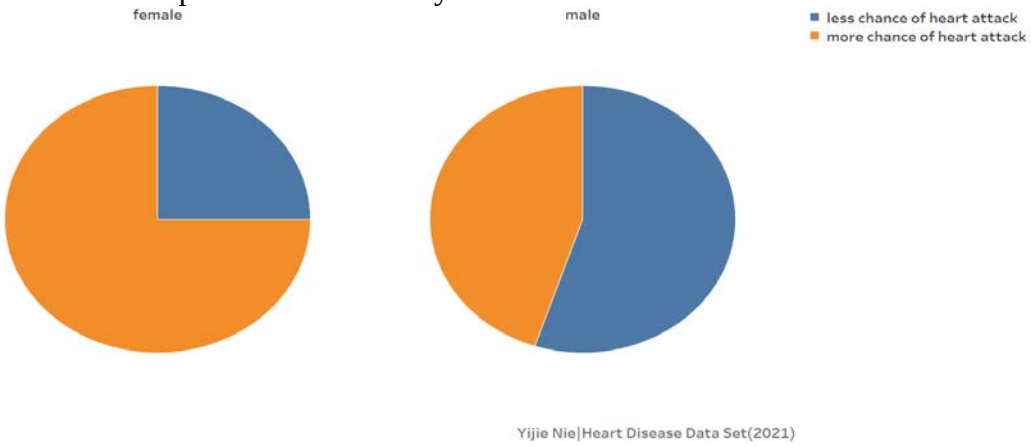

Fig.2 Pie Charts Show Women Are More Likely to Have a Heart Attack

The writer wants to find the relationship between gender and the output-less or more chance of heart attack. The area of the orange plate in the left pie chart is greater than that of the orange plate in the right pie chart, which means the percentage of females who have more chance of heart attack is larger than that of males who have more chance of heart attack. Therefore, females are more likely to experience a heart attack than males.

2.3 Cholesterol Value and the Possibility of Heart Attack

2.3.1 Age and Cholesterol Value

"A heart attack occurs when the flow of blood to the heart is blocked. The blockage is most often a buildup of fat, cholesterol and other substances" [3]. Therefore, there is a link between a heart attack and cholesterol. The writer expects that there is a relationship between cholesterol values and the probability of heart attack. First, Figure 3 looks at the relationship between age and cholesterol values.

Figure 3 shows the scatterplot with age on the x-axis and cholesterol in mg/dl plotted on the y-axis. Each dot represents a subject. The orange color means the subject has desirable cholesterol-cholesterol in mg/dl<200. The blue shows the subject has borderline high cholesterol-200<cholesterol in mg/dl<240. The red color means the subject has high cholesterol-240<cholesterol in mg/dl.

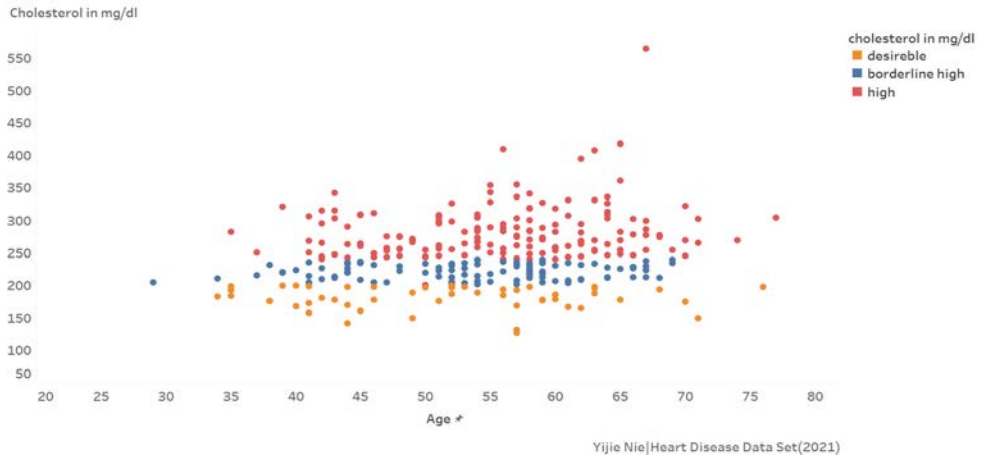

Fig.3 Scatterplot That Shows the Relationship between Age and Cholesterol Values

The writer wants to know the relationship between age and cholesterol value and whether more people have a desirable cholesterol value. As shown by the chart, there is a positive correlation between age and cholesterol value, meaning as age increases, cholesterol value also increases. Besides, from the chart, the number of blue and red dots is greater than that of orange dots, meaning fewer patients have a desirable cholesterol value.

Although cholesterol value increases as age increases, some subjects with large ages have a desirable cholesterol value. This raises the question of what actions people should take to have a low cholesterol value.

2.3.2 Intuitive Performance of the Relationship between Cholesterol Value and the Possibility of Heart Attack

After observing the relationship between age and cholesterol value, the writer would like to know whether the value of cholesterol could predict heart attack.

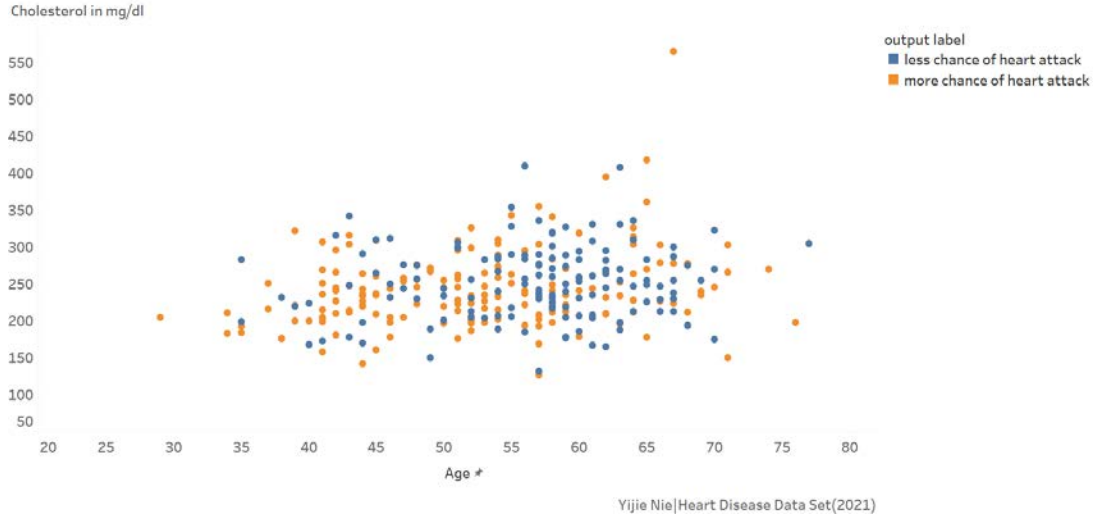

Fig.4 Scatterplot Shows the Relationship between Cholesterol Values and the Possibility of Heart Attack

Figure 4 shows the scatterplot with age on the x-axis and cholesterol in mg/dl plotted on the y-axis. Each dot represents a patient. The blue color means the patient is less likely to have a heart attack, while the orange color means the patient is more likely to have a heart attack.

This graph focuses on the relationship between age and cholesterol values and whether a person with low cholesterol values is less likely to have a heart attack. However, from the chart, orange dots are entangled with blue dots. No evidence shows there is a relationship between cholesterol values and the possibility of having a heart attack. Why are they unrelated? Is it because other factors influence the output, leading to the change of relationship? In addition to this, is it because

this dataset may not be very accurate?

2.4 Resting Electrocardiographic Results and the Possibility of Heart Attack

Resting electrocardiographic results is another factor that the writer would like to examine whether it could predict heart attack. People should lie down to get a resting electrocardiogram (ECG). “No movement is allowed during the test, as electrical impulses generated by other muscles may interfere with those generated by your heart” [4]. The resting ECG relates to the electrical activity of the heart.

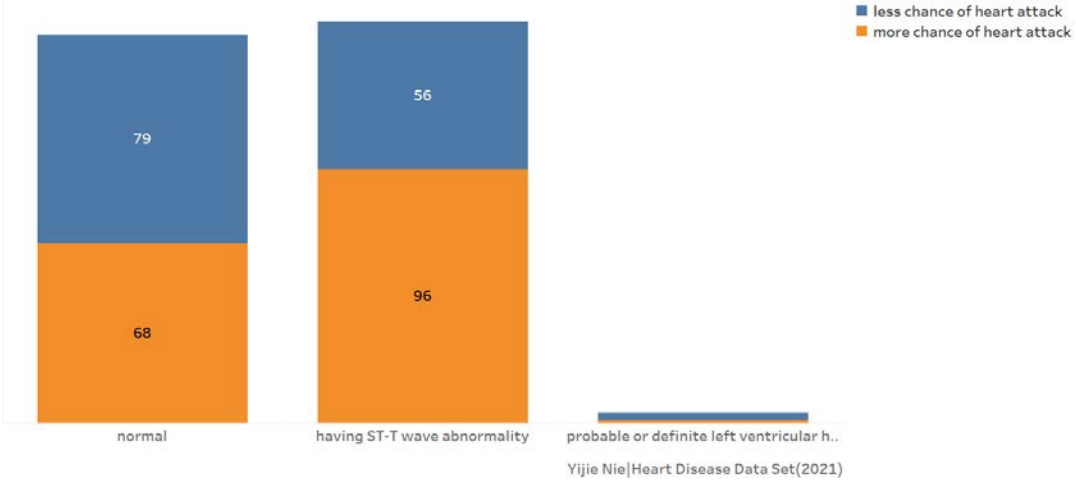

Fig.5 Horizontal Bars That Show the Relationship between Resting Ecg Results and Heart Attack

Figure 5 includes the horizontal bars that show the result of resting ECG on the x-axis and the number of patients on the y-axis. The blue color means the patient is less likely to have a heart attack, while the orange color means the patient is more likely to have a heart attack.

Is a person with a normal result less likely to have a heart attack? From the chart, the number of patients with normal results is similar to that of patients whose results show ST-T wave abnormality (one is 79+68=147 and another one is 56+96=152). The volume of the orange plate of a person with results showing ST-T wave abnormality is greater than that of the orange plate of a person with normal results. Therefore, people are more likely to have a heart attack if the result of resting electrocardiography shows ST-T wave abnormality. There is little data for the third horizontal bar showing people with probable or definite left ventricular hypertrophy by Estes' criteria. Therefore, the writer prefers to omit it.

2.5 Chest Pain Types and the Possibility of Heart Attack

The last variable is the chest pain types-asymptomatic, typical angina, atypical angina, and non-anginal pain.

In Figure 6, each type of chest pain as a whole is represented by a pie chart. The blue color means the patient is less likely to have a heart attack, and the orange color means the patient is more likely to have a heart attack.

The area of the blue plate of asymptomatic is much larger than the areas of the blue plates of three other types of chest pain, meaning people who have asymptomatic type are the least likely to have a heart attack.

Besides, the areas of the blue plates in the last three pie charts are more than 50% of the total area. Therefore, when a person feels pain near the chest, it would be better to have a medical checkup.

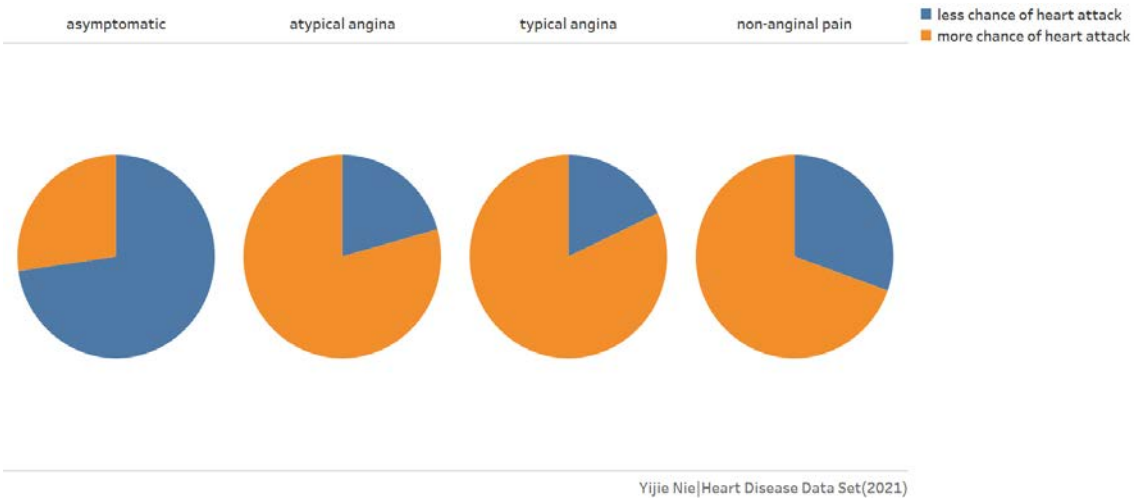

Fig.6 Pie Charts That Show the Relationship between Chest Pain Types and Heart Attack

3. Conclusion

This research mainly focuses on five variables that could predict heart attack. Six charts play an important role when analyzing the data set.

It is surprising to note that the likelihood of having a heart attack is high for people younger than 60 years old. Besides, females are more likely to have a heart attack than males. However, it shows that the cholesterol value does not predict heart attack. More datasets that relate to heart attack should be analyzed to make the conclusions more reliable.

In addition, some results support types of chest pain, and resting electrocardiography results correlate with the probability of heart attack. People should pay much attention to heart attacks and have regular check-ups. Further study is needed to identify more indicators that could predict heart attack.

References

[1] “Heart Disease Data Set.” UCI Machine Learning REPOSITORY: Heart Disease Data Set, [online] Available: http:// archive.ics.uci.edu/ml/datasets/Heart+Disease.

[2] “Heart Attack Analysis & Prediction Dataset.” Kaggle, 22 Mar. 2021, [online] Available: http://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset.

[3] “Heart Attack.” Mayo Clinic, Mayo Foundation for Medical Education and Research, 16 June 2020, [online] Available: http://www.mayoclinic.org/diseases-conditions/heart-attack/symptoms-causes/syc-2073106.

[4] “ECG Test.” ECG Test - Better Health Channel, [online] Available: http://www.betterhealth.vic.gov.au/health/conditionsandtreatments/ecg-test.