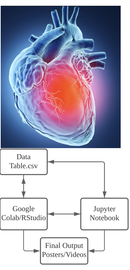
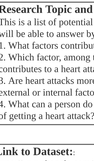
Tingyue Cui
DigHum 100
Dr. Adam Anderson
Jun 27th, 2021

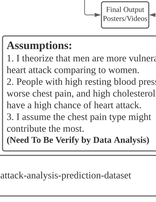
Introduction:

Heart disease is the leading factor of human death across different racial groups. It's important to know more about heart disease, specifically heart attack. Heart attack, also known as myocardial infarction, happens when the muscle of the heart didn't get enough blood. It's usually caused by the blockage of one or more coronary arteries. It could be fatal without immediate medical support. According to CDC, 1 out of 5 heart attacks is silent, which means the patients would not be aware of it when it kicks in. I listed a couple of potential questions below. I hope I can answer by analyzing the dataset.



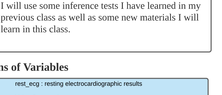


Project Work Flow:



Dataset:

I found this dataset on Kaggle, the link is cited below for reference. It uploaded as a CSV file and updated 2 months ago. This dataset includes 303 patients and some of their physical characteristics. These variables are age, sex, exercise-induced angina, number of major vessels, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, and the chance of heart attack. Some variables have specific units. A snapshot of the dataset loaded in RStudio is shown below.



Research Topic and Arguments:

This is a list of potential questions that I hope I will be able to answer by exploring the data.
1. What factors contribute to the heart attack?
2. Which factor, among these variables, contributes to a heart attack the most?
3. Are heart attacks more likely to be caused by external or internal factors?
4. What can a person do to reduce their chance of getting a heart attack?

Assumptions:

1. I theorize that men are more vulnerable to heart attack comparing to women.
2. People with high resting blood pressure, worse chest pain, and high cholesterol might have a high chance of heart attack.
3. I assume the chest pain type might contribute the most.
(Need To Be Verify by Data Analysis)

Tools and Methods:

Tools:
I will use Google Colab, Jupyter Notebook, and Rstudio to analyze the dataset. Because I'm not so familiar with Colab and Jupyter Notebook, there might be a minor difference.
Methods:
I will use some inference tests I have learned in my previous class as well as some new materials I will learn in this class.

Link to Dataset:

https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset

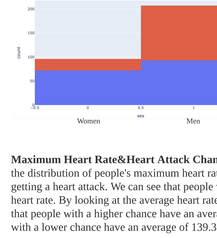
Dataset Exploratory

Descriptions of Variables

Age : Age of the patient	rest_ecg : resting electrocardiographic results
Sex : Sex of the patient (1 = Male; 0 = Female)	Value 0: normal
exang: exercise induced angina (1 = yes; 0 = no)	Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
ca: number of major vessels (0-3)	Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
cp : Chest Pain type chest pain type	thalach : maximum heart rate achieved
Value 1: typical angina consists(1) Substernal chest pain or discomfort that is (2) Provoked by exertion or emotional stress and (3) relieved by rest or nitroglycerine (or both)	slope: the slope of the peak exercise ST segment
Value 2: atypical angina: when when 2 out of 3 criteria of typical angina are present	-- Value 1: up-sloping
Value 3: non-anginal pain: heart pain but patients don't have heart disease	-- Value 2: flat
Value 4: asymptomatic: no significant pain or feelings	-- Value 3: down-sloping
trtbps : resting blood pressure (in mm Hg)	thal: 3 = normal; 6 = fixed defect; 7 = reversible defect
chol : cholesterol in mg/dl fetched via BMI sensor	old peak = ST depression induced by exercise relative to rest
fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)	bs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
	target : 0= less chance of heart attack 1= more chance of heart attack

Gender Distribution:

In this dataset, 31.7 % of them are female participants; 68.3% of them are male participants. There are a total of 303 participants. 96 of them are female; 207 of them are male.(1=Male; 0= Female)

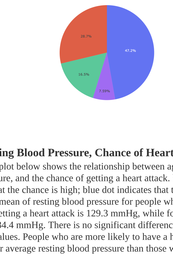


Age Distribution:

Among these 303 participants, the age ranges from 29 to 77 years old. The average age is 54.37 years old. The median age is 55 years old. The distribution looks approximately skewed to the left with more elders.

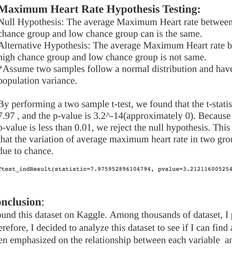
Chance of Heart Attack:

Among these 303 participants, 45.5% of them have low chance of heart attach; 54.5% of them have high chance of heart attach.



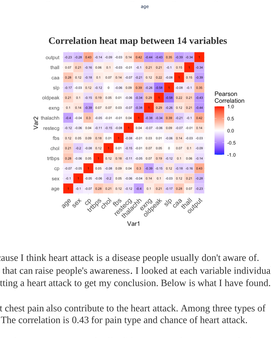
Gender and Heart Attack Chance:

Among 96 female participants, 72 of them have a high chance of heart attack. Among 207 male participants, 93 of them have a high chance heart attack. Just by looking at the relationship between gender and heart attack chance, female is more prone to have a high chance.



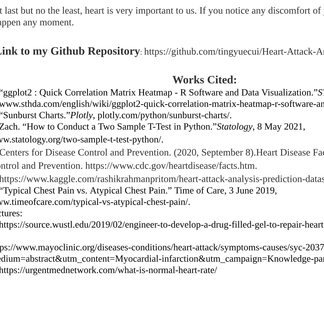
Chest Pain Type and Heart Attack Chance:

By looking at the pie chart above, we could tell that people with a higher chance of heart attack seems to have more type 2 chest pain. The distribution of type 1 and type 3 in high chance group are larger than people in low chance.



Maximum Heart Rate&Heart Attack Chance:

In this histogram, it shows the distribution of people's maximum heart rate achieved and the likelihood of getting a heart attack. We can see that people with a high chance have a higher heart rate. By looking at the average heart rate of the two groups, we found that people with a higher chance have an average of 158.47, whereas people with a lower chance have an average of 139.33.



Chest Pain Type Distribution:

Four types of chest pain distribution is shown below.
Value 0: typical angina, 47.2%
Value 1: atypical angina, 28.7%
Value 2: non-angina pain, 16.5%
Value 3: asymptomatic, 7.59%

Age, Cholesterol level, Chance of Heart Attack:

The scatter plot below shows the relationship between age, cholesterol level, and the chance of getting a heart attack. Yellow dot indicates that the chance is high; blue dot indicates that the chance is low. The mean of cholesterol level for people who with a high chance of getting heart attack is 242.2 mg/dL, while for low chance is 250.8 mg/dL. Both values are higher than desirable cholesterol level for adults. People who are more likely to have a heart attack have a lower average cholesterol level than those with low chance.

Age, Resting Blood Pressure, Chance of Heart Attack:

The scatter plot below shows the relationship between age, resting blood pressure, and the chance of getting a heart attack. Yellow dot indicates that the chance is high; blue dot indicates that the chance is low. The mean of resting blood pressure for people who with high chance of getting a heart attack is 129.3 mmHg, while for low chance is 134.4 mmHg. There is no significant difference between these two values. People who are more likely to have a heart attack have a lower average resting blood pressure than those with low chance.

Maximum Heart Rate Hypothesis Testing:

Null Hypothesis: The average Maximum Heart rate between high chance group and low chance group can is the same.
Alternative Hypothesis: The average Maximum Heart rate between high chance group and low chance group is not same.
*Assume two samples follow a normal distribution and have same population variance.

By performing a two sample t-test, we found that the t-statistics is 7.97 , and the p-value is 3.2^-14(approximately 0). Because our p-value is less than 0.01, we reject the null hypothesis. This means that the variation of average maximum heart rate in two groups is not due to chance.

```
Ttest_indResult(statistic=7.975952896104794, pvalue=3.212116005251946e-14)
```

Correlation heat map between 14 variables

Conclusion:

I found this dataset on Kaggle. Among thousands of dataset, I picked this one because I think heart attack is a disease people usually don't aware of. Therefore, I decided to analyze this dataset to see if I can find anything important that can raise people's awareness. I looked at each variable individually. Then emphasized on the relationship between each variable and the chance of getting a heart attack to get my conclusion. Below is what I have found.

In this dataset, women are more vulnerable to heart attack than men. The different chest pain also contribute to the heart attack. Among three types of chest, people with the second type of chest are more likely to have a heart attack. The correlation is 0.43 for pain type and chance of heart attack.

People with higher chance of heart attack are more likely to have a higher maximum heart rate. We specifically did a two sample t-test to verify this observation. In the calculation, the p-value is extremely small. It verified our hypothesis. The correlation is 0.42 which is relatively high among all other factors.

Cholesterol level and resting blood pressure are not significant factors of determining chance of heart attack. By looking at the correlation heat map, we can see that the correlation between cholesterol level and chance of heart attack is -0.09. The correlation between resting blood pressure and chance of heart attack is 0.14.

In order to reduce chance of heart attack, people should try to reduce their maximum heart rate. People can do some moderate exercise such as walking, yoga. They should also decrease the chance of having the second type of chest pain. They should be aware of this pain type.

In this dataset, the heart attack is more likely to be caused by the internal factor, specifically the chest pain type. However, we still need to take the external factors into account. This is because the internal factors sometimes are related to external factors. For example, the blood pressure can be manipulated by eating specific type of food. Therefore, both internal and external can contribute to heart attack.

For future work, people can analyze a larger data to get more accurate results since this dataset only contains 303 patients. Future researchers can also collect more data about external factors such as diet habits and exercise habits to further analyze the correlation between them.

At last but no the least, heart is very important to us. If you notice any discomfort of your heart, make sure to check as early as possible. Heart attack can happen any moment.

Link to my Github Repository: https://github.com/tingyuecui/Heart-Attack-Analysis

Works Cited:

1. "ggplot2 : Quick Correlation Matrix Heatmap - R Software and Data Visualization."*STHDA*, www.sthda.com/english/wiki/ggplot2-quick-correlation-matrix-heatmap-r-software-and-data-visualization.
2. "Sunburst Charts."*Plotly*, plotly.com/python/sunburst-charts/.
3. Zach. "How to Conduct a Two Sample T-Test in Python."*Statology*, 8 May 2021, www.statology.org/two-sample-t-test-python/.
4. Centers for Disease Control and Prevention. (2020, September 8).Heart Disease Facts. Centers for Disease Control and Prevention. https://www.cdc.gov/heartdisease/facts.htm.
5. https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset
6. "Typical Chest Pain vs. Atypical Chest Pain." Time of Care, 3 June 2019, www.timeofcare.com/typical-vs-atypical-chest-pain/.

Pictures:
1. https://source.wustl.edu/2019/02/engineer-to-develop-a-drug-filled-gel-to-repair-heart-after-heart-attack/
2. https://www.mayoclinic.org/diseases-conditions/heart-attack/symptoms-causes/syc-20373106?utm_source=Google&utm_medium=abstract&utm_content=Myocardial-infarction&utm_campaign=Knowledge-panel
3. https://urgentmednetwork.com/what-is-normal-heart-rate/