Introduction

Whether you exercise a lot, just a bit or not at all, lifestyle choices play a pivotal role in shaping our overall health. Understanding the interplay between exercise habits and the risk of heart attacks can be an important factor in managing your health and preventing heart attacks (Exercise induced angina).

In this analysis of the dataset from Kaggle, I aim to unravel the complex relationship between exercise patterns and heart attack susceptibility, factoring in the nuances of Age, Sex, Cholesterol Level and Resting Blood Pressure as critical variables that may influence this association.

While the benefits of regular physical activity on cardiovascular health are well established, the extent to which exercise mitigates the risk of heart attacks is still under debate. My goal with this project is to make that link a little less blurry.

While the analysis will mention the binary relationship between exercise and heart health, the key to my analysis is the consideration of demographic factors such as age and sex, acknowledging that these variables may introduce nuanced dynamics to the correlation between heart attacks and exercise.

Setting out on this exploration of heart health in relation to exercise, the goal is to uncover insights on the data behind, and foster a deeper understanding of, the intricate web connecting exercise, age, sex, and heart attack risk.

This project aims to shed light on the multifaceted nature of cardiovascular health and to see what can potentially be done differently on an individual level for increased personal well being.

Descriptive Statistic

	count	mean	std	min	25%	50%	75%	max
age	303.0	54.366337	9.082101	29.0	47.5	55.0	61.0	77.0
cholesterol_level	303.0	246.264026	51.830751	126.0	211.0	240.0	274.5	564.0
resting_blood_pressure	303.0	131.623762	17.538143	94.0	120.0	130.0	140.0	200.0
exercise_induced_angina	303.0	0.326733	0.469794	0.0	0.0	0.0	1.0	1.0

- Age: The average age of individuals in the dataset is approximately 54.37 years, with a standard deviation of about 9.08 years. This indicates a moderate spread around the mean age. The youngest individual is 29 years old, and the oldest is 77 years old. The interquartile range (IQR), which is the range between the 25th and 75th percentile, is

from about 47.5 to 61 years, suggesting that half of the individuals are within this age range.

- Cholesterol Level: The average cholesterol level is roughly 246.26 mg/dL, with a standard deviation of 51.83 mg/dL, indicating a relatively wide variability in cholesterol levels among individuals. The cholesterol levels range from 126 to 564 mg/dL, which is quite a broad range. The IQR is from 211 to 274.5 mg/dL, showing that the middle 50% of individuals have cholesterol levels within this range.
- Resting Blood Pressure: The mean resting blood pressure is 131.62 mmHg, with a standard deviation of 17.54 mmHg, which suggests a moderate variation in blood pressure readings among the population. The minimum blood pressure recorded is 94 mmHg, and the maximum is 200 mmHg. The IQR is from 120 to 140 mmHg, meaning that half of the observed blood pressures lie within this range.
- Exercise Induced Angina: This binary variable has values of 0 or 1, where 0 likely indicates the absence of exercise-induced angina and 1 indicates its presence. On average, about 32.67% of individuals have exercise-induced angina. Since this is a binary variable, the standard deviation isn't particularly informative about the spread but does indicate variability in the presence of this condition within the group. The min and max values (0 and 1) confirm that it's a binary variable.

The count value of 303 for each variable indicates that there are no missing values for these variables in the dataset.

In summary, the table offers a quick statistical snapshot of the data. It suggests that the sample includes a middle-aged adult population with a broad range of cholesterol levels and resting blood pressure readings. Additionally, a significant minority of the population experiences exercise-induced angina. The spread of the data for age and resting blood pressure is moderately wide, whereas cholesterol levels show a wider range, which could be due to various dietary and genetic factors among the individuals. The presence of exercise-induced angina in about a third of the population could be a point of interest for further investigation, especially in relation to other cardiovascular health indicators.

Exercise Induced Angina by Sex and Age, controlled for Resting Blood Pressure

The graph (shown in the appendix) provided is a scatter plot with regression lines, illustrating the relationship between age and resting blood pressure, differentiated by sex and marked for the presence of exercise-induced angina. The points are color-coded: blue represents male and red represents female. Additionally, the shape of the points indicates whether exercise-induced angina is present (triangles) or not (circles).

Analyzing the chart:

- **Age and Resting Blood Pressure Relationship**: There is a general upward trend in resting blood pressure with age, as indicated by the regression lines. This suggests that older individuals tend to have higher resting blood pressure.
- **Sex Differentiation**: The regression lines for males and females appear to have different slopes, with the male (blue) line being steeper than the female (red) line. This could imply that the increase in blood pressure with age is more pronounced in males compared to females in this dataset.
- **Exercise-Induced Angina**: The scatter points are differentiated by shape based on the presence of exercise-induced angina. Triangles, which represent individuals with exercise-induced angina, are present across the age spectrum for both sexes. However, without a clear pattern distinguishing the triangles from circles, it's difficult to infer a strong relationship between age, sex, and the presence of angina from this graph alone.
- **Confidence Interval**: The shaded area around the regression lines represents the confidence interval for the linear regression estimate. A wider shaded area would indicate more uncertainty in the estimate. In this graph, the confidence intervals seem moderately wide, suggesting some variability in the blood pressure readings that isn't explained solely by age.
- **Population Distribution**: The distribution of points shows a relatively even spread across age groups, with a concentration of data points in the middle age range. Both younger and older ages have fewer data points, which is typical in many biomedical datasets.
- **Outliers**: There seem to be a few potential outliers, especially in the higher blood pressure readings for given ages, which appear as points far from the general cluster of data.

In summary, the graph indicates a positive correlation between age and resting blood pressure, with possible sex differences in this relationship. The presence of exercise-induced angina is noted but does not display a distinct pattern in relation to age or resting blood pressure. The regression analysis, while informative, shows that there is considerable variation in blood pressure that may be influenced by factors not included in the graph, such as lifestyle or genetic predispositions.

Exercise Induced Angina by Sex and Age, controlled for Cholesterol Level

The graph (shown in the appendix) presented is a scatter plot with overlaid regression lines, depicting the relationship between age and cholesterol level, and it's color-coded by sex. Points are also marked to show the presence of exercise-induced angina.

Here's a detailed analysis of the graph:

- **Age and Cholesterol Level Relationship**: The plot indicates a slight positive correlation between age and cholesterol level, as shown by the upward trend in the regression lines.

This suggests that as age increases, there's a tendency for cholesterol levels to rise as well.

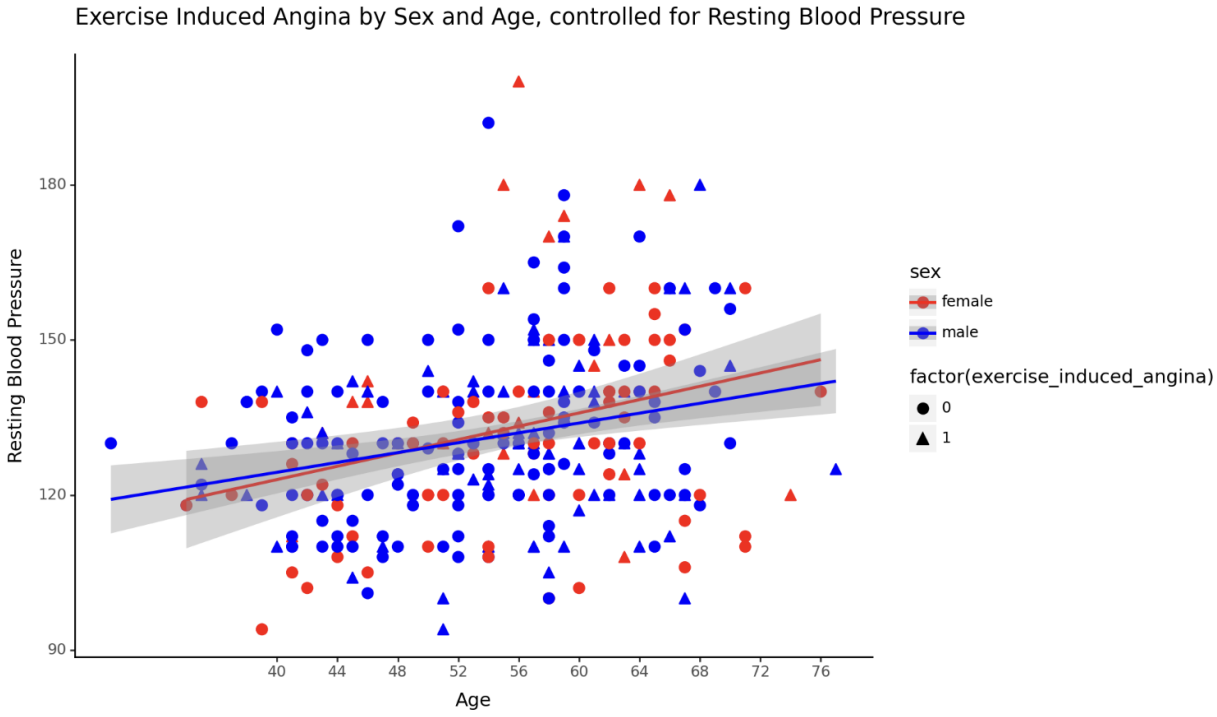
- **Sex Differentiation**: The plot differentiates between male (blue) and female (red) individuals. The two regression lines suggest there might be different patterns in how cholesterol levels change with age between sexes. However, the slopes of the regression lines for both sexes seem relatively flat, indicating that age may not be a strong predictor of cholesterol level in this dataset.
- **Exercise-Induced Angina**: The distribution of triangles among the circles doesn't show a clear pattern that relates the presence of exercise-induced angina with cholesterol level and age.
- **Confidence Interval**: The relatively wide confidence intervals, especially at higher ages, suggest there is considerable uncertainty about the true relationship between age and cholesterol level. This variability implies other factors may influence cholesterol levels in addition to age.
- **Data Distribution**: The spread of points across age groups is even, with a relatively high density of data points in the middle age range. The concentration of points around the middle of the graph reflects the median cholesterol levels and the most common ages in the dataset.
- **Outliers**: There are some apparent outliers, especially in cholesterol levels, where some individuals have significantly higher levels than others in the same age group. These outliers could be due to individual differences in diet, genetics, or other health conditions.

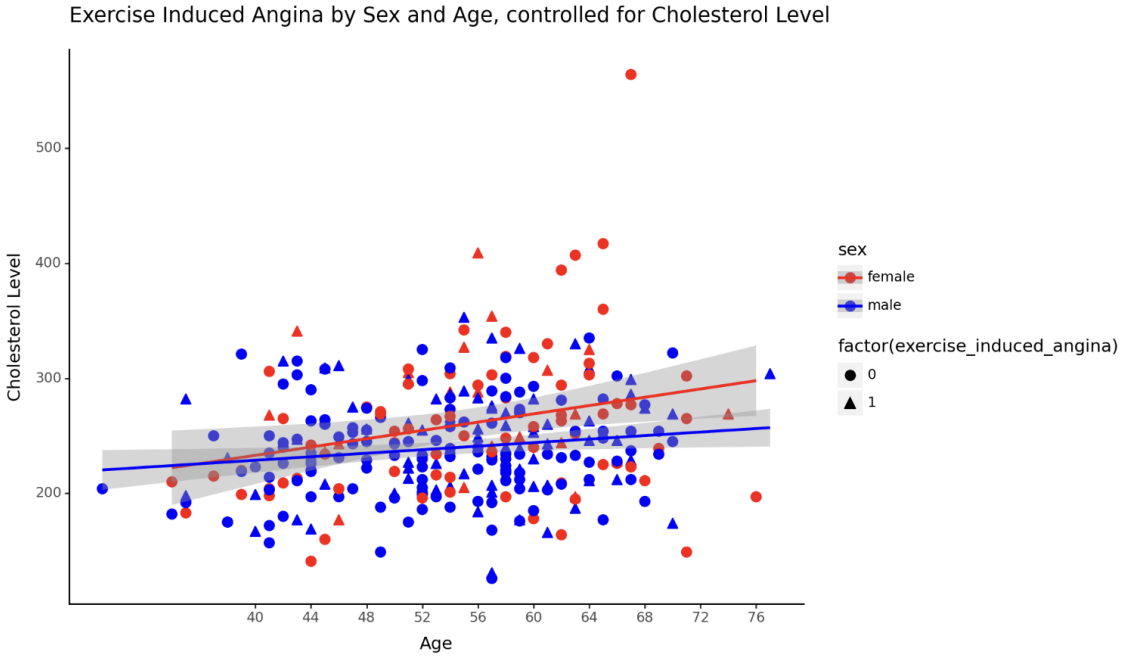
In summary, the graph shows that while there is some increase in cholesterol levels with age, the effect is not pronounced. Both males and females show this trend, but it is not strongly differentiated between the sexes. The presence of exercise-induced angina marked in the data points does not appear to correlate strongly with higher cholesterol levels or with age. The considerable spread of the data points and the wide confidence intervals suggest that other factors may be at play in determining an individual's cholesterol level.

Final thoughts

This investigation not only contributes to the existing body of knowledge on cardiovascular health but also holds practical implications for personalized healthcare strategies. A clearer comprehension of the interrelationships between age, sex, heart health and exercise can guide targeted interventions and preventive measures, thereby promoting a healthier and more informed society. And while we couldn't see a clear correlation between the different variables and heart attack risk, the analysis still provided valuable insights into the relationship between Age, Sex and Cholesterol Levels/ Resting Blood Pressure.

Appendix


Exercise Induced Angina by Sex and Age, controlled for Resting Blood Pressure
Resting Blood Pressure
180
150
120
90
40
44
48
52
56
60
64
68
72
76
Age
sex
female
male
factor(exercise_induced_angina)
0
1


Exercise Induced Angina by Sex and Age, controlled for Cholesterol Level
Cholesterol Level
500
400
300
200
40
44
48
52
56
60
64
68
72
76
Age
sex
female
male
factor(exercise_induced_angina)
0
1