Data Analysis: Cardiovascular Disease (CVD)

By Jordan M. Flood

Table of contents

Causes of CVDs

Lifestyle Factors

Diet: High intake of saturated fats, trans fats, and salt can lead to high cholesterol and high blood pressure, respectively, both of which are major risk factors for CVD.

Physical Inactivity: Lack of exercise can contribute to obesity and poor heart health.

Smoking: Tobacco smoke contributes to the buildup of fatty substances in the arteries, leading to atherosclerosis.

Genetic Factors:

Heredity: A family history of CVD increases one's risk, pointing to the genetic components that influence cardiovascular health.

Some individuals may have a genetic predisposition to hypertension, diabetes, and obesity, all of which are significant risk factors for CVDs.

Environmental Factors

Air Pollution: Long-term exposure to air pollutants is linked to cardiovascular risk.

Stress: Chronic stress may indirectly influence heart health through the unhealthy behaviors it encourages, like poor diet and smoking.
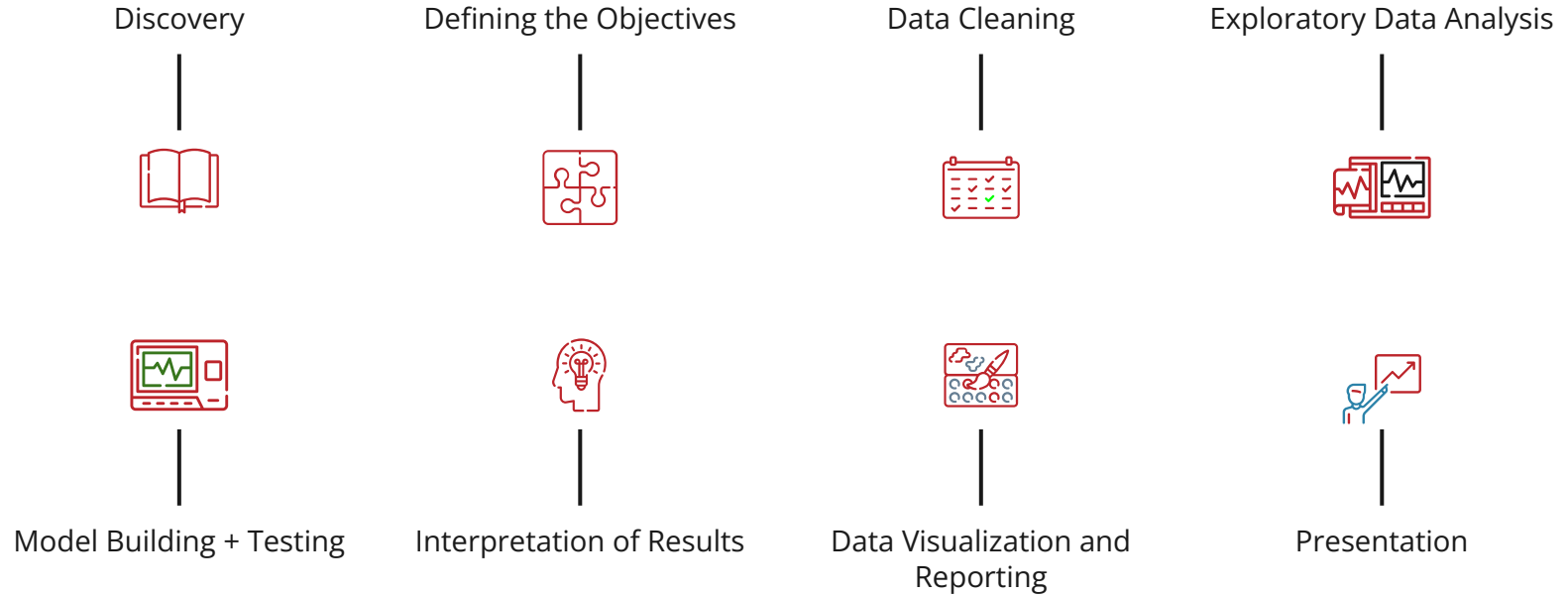
The Data

In a dataset of 70,000 individuals I conducted thorough Analysis to discover what are the factors involved with CVD and whether or not I can build a Model to accurately predict if someone has CVD.

The Process

Discovery

Defining the Objectives

Data Cleaning

Exploratory Data Analysis

Model Building + Testing

Interpretation of Results

Data Visualization and Reporting

Presentation

17.9 Million

Deaths Caused by CVD every year*

33 seconds

Death caused by CVD**

60%

The Increase of deaths over 30 years

Understanding the numbers
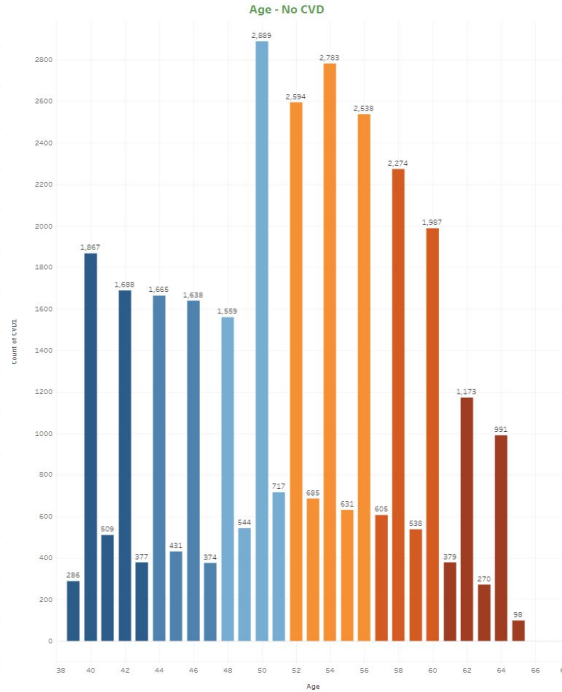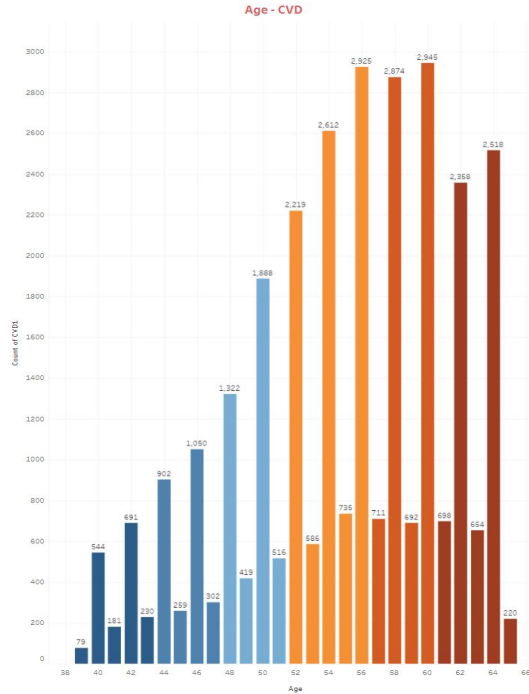
Age

Of course, as we get older the likelihood of developing these conditions increases but can we mitigate the possibility?
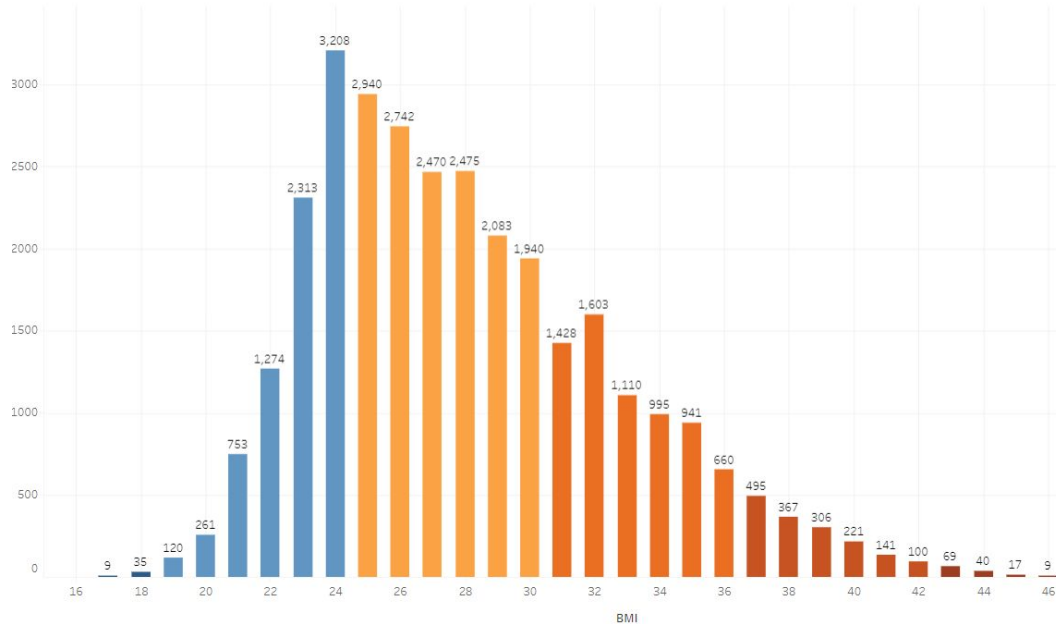
We can look at some other factors to see what increases the risk of developing Cardiovascular Disease.

For more info, click here

Understanding the numbers

BMI - CVD

Body-Mass Index (BMI)

This chart illustrates the relationship between Body Mass Index (BMI) and the incidence of cardiovascular diseases (CVD). As the data shows, higher BMI values are associated with a greater count of CVD cases. This trend underscores the importance of managing body weight as a preventive measure against cardiovascular conditions. The graph highlights how incremental increases in BMI correspond to significant rises in CVD occurrences, suggesting that even moderate reductions in BMI could potentially lead to a decrease in the prevalence of CVDs. Utilizing this data can be instrumental in public health initiatives aimed at reducing the burden of cardiovascular diseases through targeted interventions focused on weight management.
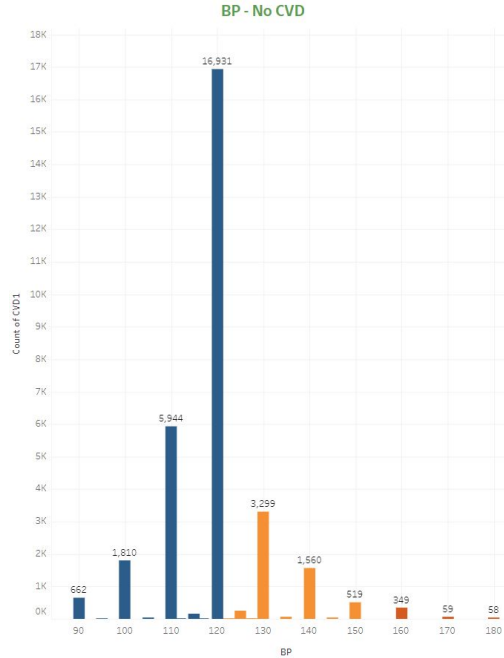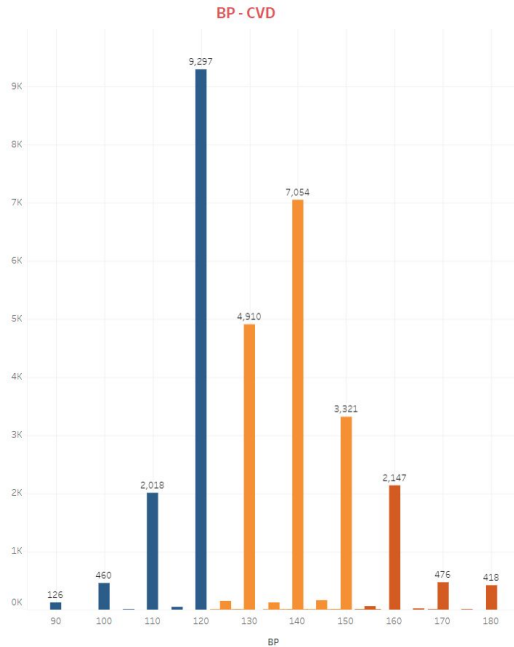
For more info, click here

Understanding the numbers

BP - CVD

BP - No CVD

Blood Pressure (BP)

This chart shows how blood pressure levels are linked to the number of cardiovascular disease (CVD) cases. As blood pressure increases, so does the likelihood of developing CVD. The data underscores the importance of controlling blood pressure to lower the risk of heart-related illnesses. Maintaining healthy blood pressure is crucial for preventing heart diseases.
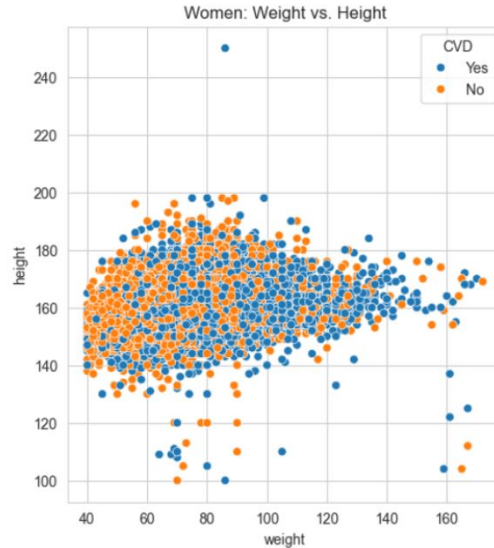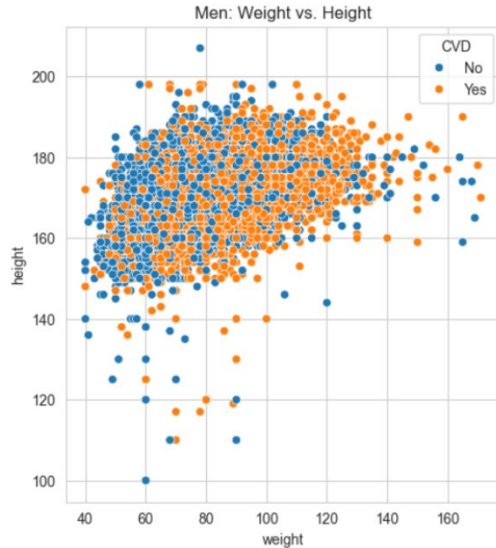
For more info, click here

Understanding the numbers

Men: Weight vs. Height

Women: Weight vs. Height

Weight vs. Height
In men and women

Here we can see the differences between Men and Women based on their Height and Weight on a scatterplot showing the presence of cardiovascular disease (CVD).

As weight increases, so does the likelihood of developing CVD. This ties in with the fact that since 1990 obesity has more than doubled worldwide for adults and more than quadrupled for Adolescents - (WHO)

The data underscores the importance of controlling a healthy weight to lower the risk of heart-related illnesses and premature death.

For more info, click here

Understanding the numbers

Active - CVD

Not Active - CVD
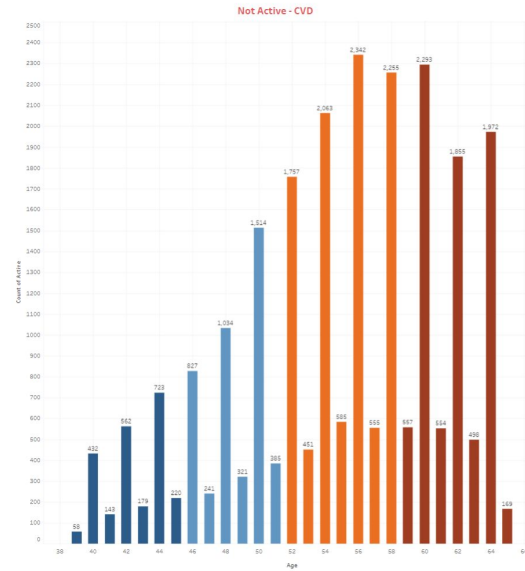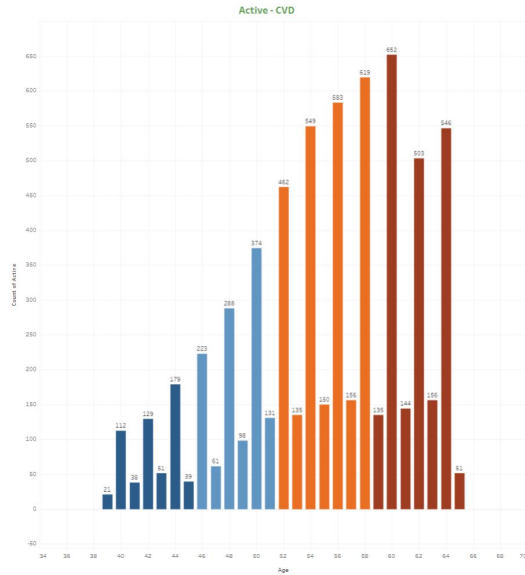
Physical Activity

This chart shows how those who have CVD while active are far less than those who are inactive.

What this can tell us is that physical activity can lower the risk of having any CVD but not eliminate it and we should still take care of other factors to mitigate the risk as much as we can.
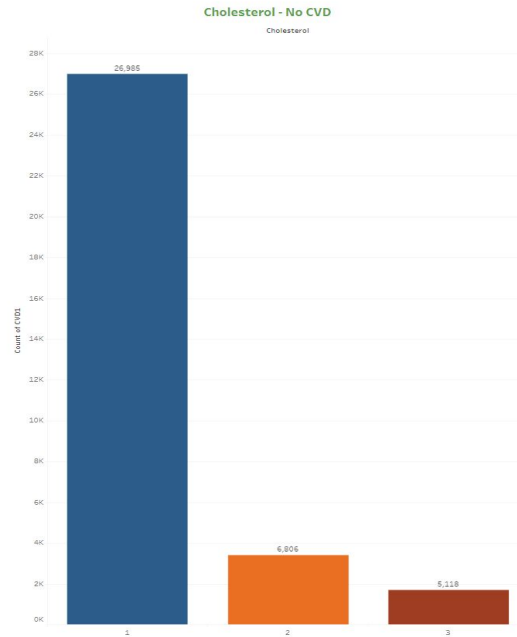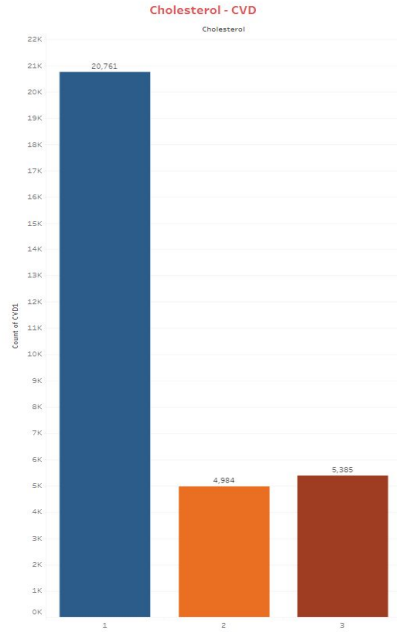
For more info, click here

Understanding the numbers

Cholesterol - CVD

Cholesterol

Cholesterol - No CVD

Cholesterol

Cholesterol

The chart on the right shows a higher count of individuals with no cardiovascular diseases who have high cholesterol levels. This indicates that high cholesterol alone does not necessarily lead to CVD for everyone in the sampled population. This suggests the presence of other mitigating factors that might protect these individuals from developing CVD or variability in the types of cholesterol (such as higher levels of 'good' HDL cholesterol).

These insights could be elaborated further with more specific data on cholesterol levels (such as LDL, HDL, and triglycerides) and other risk factors present in the individuals represented in these charts to provide a deeper understanding of the interplay between cholesterol and cardiovascular health.

For more info, click here

Findings

1	2	3	4
Age	**BMI**	**Blood–Pressure**	**Lifestyle**
• As people age, their heart muscles can thicken, heart valves may stiffen, and the large arteries may become stiffer and less elastic. These changes contribute to increased blood pressure and decreased efficiency in blood circulation. • From exploring and analysing the data, the levels of CVD cases increases noticeably between 50-55 years old	• The higher your Body-Mass Index, the higher your chance of contracting CVDs. • This is largely due to the extra strain your heart goes through to pump blood throughout the body. • Higher BMI is often associated with an increased risk of developing insulin resistance, which can lead to type 2 diabetes.	• Higher body weight then leads to increased blood volume and arterial resistance, both of which require the heart to work harder to pump blood. • This increased workload can raise blood pressure, a major risk factor for CVD as the data has shown. • Elevated blood pressure can damage the inner walls of the arteries, making them more susceptible to the accumulation of fatty plaques. This buildup can narrow and block arteries, leading to heart attacks and strokes.	• **Physical Inactivity**: Often, higher BMI is associated with lower levels of physical activity, which itself is a risk factor for CVD as it leads to poorer cardiovascular fitness and heart health. • **Smoking**: Smoking introduces harmful chemicals that damage blood vessels, increase plaque buildup, thicken blood, and reduce oxygen in the blood, all of which increase CVD risk. • **Alcohol:** Regular heavy drinking can lead to elevated blood pressure, which is a major risk factor for heart disease.

Findings Part 2 – The Model

I took all the data we had on factors like **age**, **BMI**, **blood pressure**, and **lifestyle factors**, and used it to build a machine learning model. This model helps predict whether someone might develop cardiovascular disease (CVD) and so far, it's been able to predict CVD with about 72% accuracy.

This shows that using machine learning can really help us understand and manage heart health better, and it could be a great tool for catching heart issues early on.

```python
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.2f}")

conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", conf_matrix)

print("Classification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy: 0.72
Confusion Matrix:
 [[7158 2112]
 [3066 6342]]
Classification Report:
              precision    recall  f1-score   support

           0       0.70      0.77      0.73      9270
           1       0.75      0.67      0.71      9408

    accuracy                           0.72     18678
   macro avg       0.73      0.72      0.72     18678
weighted avg       0.73      0.72      0.72     18678
```

Results and conclusions

Conclusions

The analysis really highlights how age, BMI, blood pressure, and lifestyle choices all mix together to impact the risk of developing cardiovascular diseases (CVD).

Each factor on its own already ups the risk, but when you combine them—like having a high BMI, high blood pressure, and not-so-great lifestyle habits such as not exercising, smoking, or drinking too much—the risk goes up even more.

Plus, just getting older naturally increases the chances of CVD too. It's clear that these factors are all interconnected and play a big role in heart health.

The comprehensive analysis and subsequent development of a machine learning model have provided valuable insights into the relationship between cardiovascular disease (CVD) and various factors such as age, BMI, blood pressure, and lifestyle choices. Our model's ability to predict CVD occurrence with about 72% accuracy underscores the significant potential of machine learning in enhancing our understanding and management of heart health.

Recommendations and What would I do Different?

Recommendations

Enhanced Public Awareness and Education: It is crucial to continue educating the public about the critical impact of lifestyle choices on heart health and the benefits of early screening and intervention.

Integration into Clinical Practices: Healthcare providers should consider integrating predictive modeling tools into clinical assessments to identify patients at high risk and tailor preventative strategies accordingly.
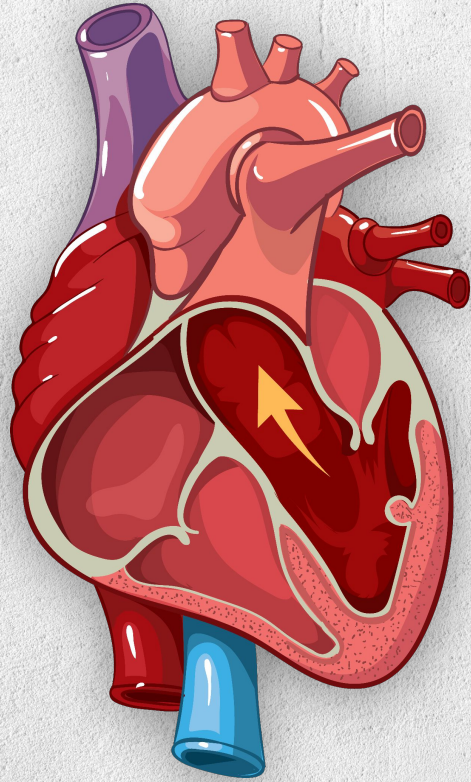
Lifestyle Interventions: Programs focused on reducing risk factors through lifestyle changes such as diet, exercise, and quitting smoking should be expanded and made more accessible to the broader population.

What would I do Different?

Expand Sample Size: To enhance the robustness and generalizability of my model, I would like to redo the analysis with a larger and more diverse sample size.

Incorporate More Data Variables: Including more detailed variables such as specific blood elements (e.g., LDL and HDL cholesterol levels, triglycerides) and genetic markers could provide deeper insights into the mechanisms driving CVD.

Further Resources

https://world-heart-federation.org/resource/world-heart-report-2023/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10809869/
https://www.cdc.gov/heartdisease/facts.htm