Final Project

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death in the United States, with one person dying of the disease every 34 seconds (Centers for Disease Control and Prevention, 2022). Previous research has highlighted male gender, old age, obesity, abnormal cholesterol and fasting blood glucose as important predictors for CVDs, among many others (Damen et al., 2016).

As CVD incidence continues to soar, the need for an effective predictive model cannot be overstated. Such a model could enable doctors to take preventive measures, treat patients early, or encourage individuals at high risk to adopt lifestyle changes. In this report, we aim to construct a predictive model based on the "Heart Failure Prediction Dataset" by assessing the 11 possible predictors (including chest pain type, resting blood pressure, cholesterol levels, maximum heart rate, resting ECG measurements, etc.). Our research question is, which combination of these factors is most effective in predicting CVDs? In answering this question, we seek to enhance the scientific community's understanding of CVDs and their associated risk factors, ultimately contributing to better prevention, early identification, and management of this critical disease for both individuals and populations.

Data Description

This "Heart Failure Prediction Dataset" includes 11 characteristics that can be utilized to anticipate the potential risk of a CVD; it is called "Heart Failure Prediction Dataset" because it is not uncommon for a CVD to lead to heart failure (Velagaleti et al., 2007). The dataset was formed by merging five heart-related datasets (the Cleveland, Hungarian, Switzerland, Long Beach VA, and Stalog Heart Datasets) based on their 11 common features.This dataset is currently the largest available heart disease dataset for research purposes, consisting of 918 observations.

Source: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

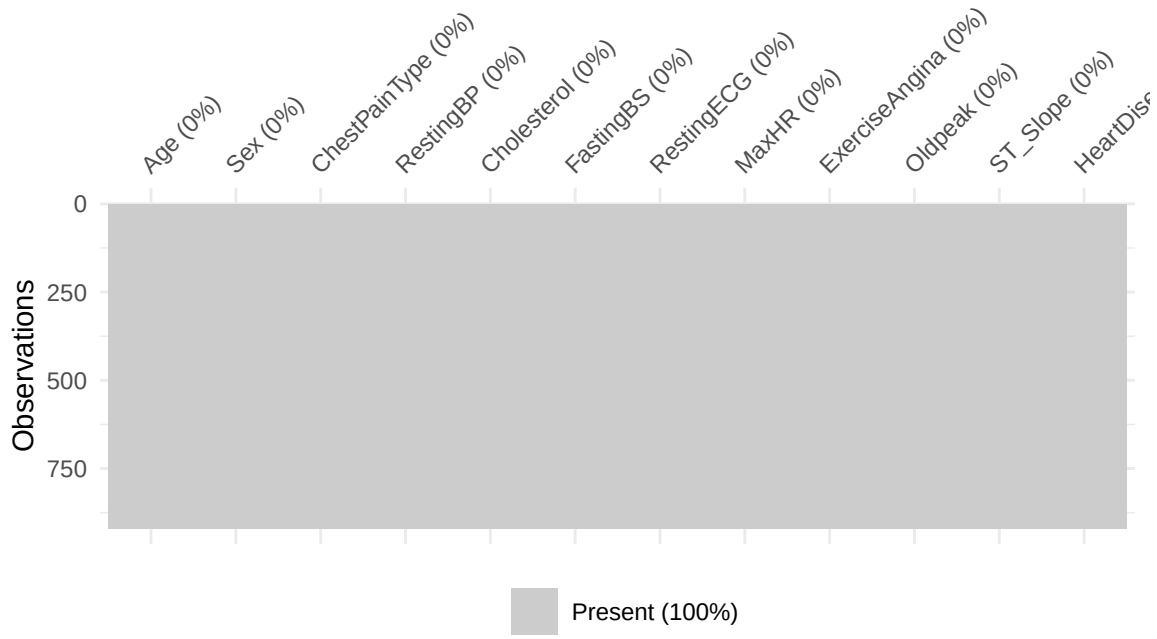
Data Dictionary: Age: The age of the patient in years. Sex: The gender of the patient, identified as Male (M) or Female (F). ChestPainType: The type of chest pain experienced by the

patient, classified as Typical Angina (TA), Atypical Angina (ATA), Non-Anginal Pain (NAP), or Asymptomatic (ASY). RestingBP: The resting blood pressure of the patient in mm Hg. Cholesterol: The level of serum cholesterol in mm/dL. FastingBS: The level of fasting blood sugar, indicated as 1 if FastingBS $>$ 120 mg/dL and 0 otherwise. RestingECG: The results of the patient's resting electrocardiogram, classified as Normal, having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of $>$ 0.05 mV), or showing probable or definite left ventricular hypertrophy by Estes' criteria (LVH). MaxHR: The maximum heart rate achieved by the patient, measured in beats per minute (bpm) and ranging from 60 to 202. ExerciseAngina: Whether the patient experienced exercise-induced angina, identified as Yes (Y) or No (N). Oldpeak: The degree of ST depression induced by exercise relative to rest, measured in depression. ST_Slope: The slope of the peak exercise ST segment, classified as Up (upsloping), Flat (flat), or Down (downsloping). HeartDisease: The output class indicating the presence (1) or absence (0) of heart disease.

Age, RestingBP, Cholesterol, MaxHR, and Oldpeak are continuous variables. Sex, ChestPainType, FastingBS, RestingECG, ExerciseAngina, ST_Slope, and HeartDisease are categorical variables.

Source: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

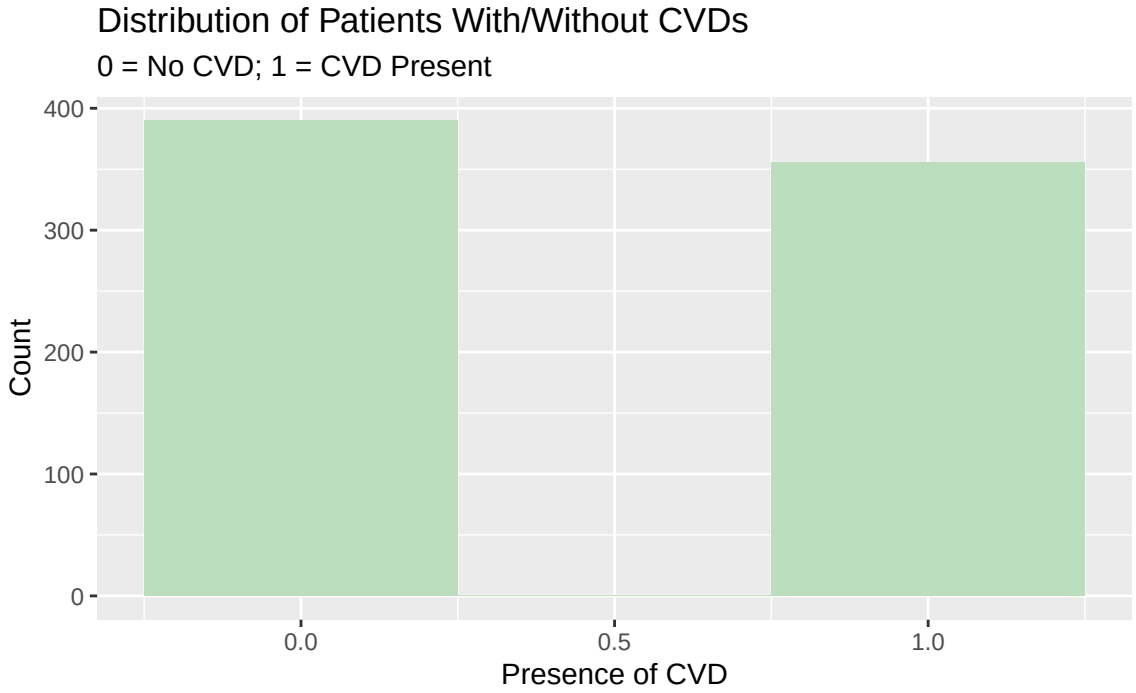
EDA



At first glance, there seems to be no missingness in the dataset. However, upon closer examination, we discovered that 173 out of the 919 observations had a serum cholesterol level of 0 mm/dl. Since this is physiologically impossible, we concluded that these observations were actually missing cholesterol data. This meant that our dataset had missing values after all.

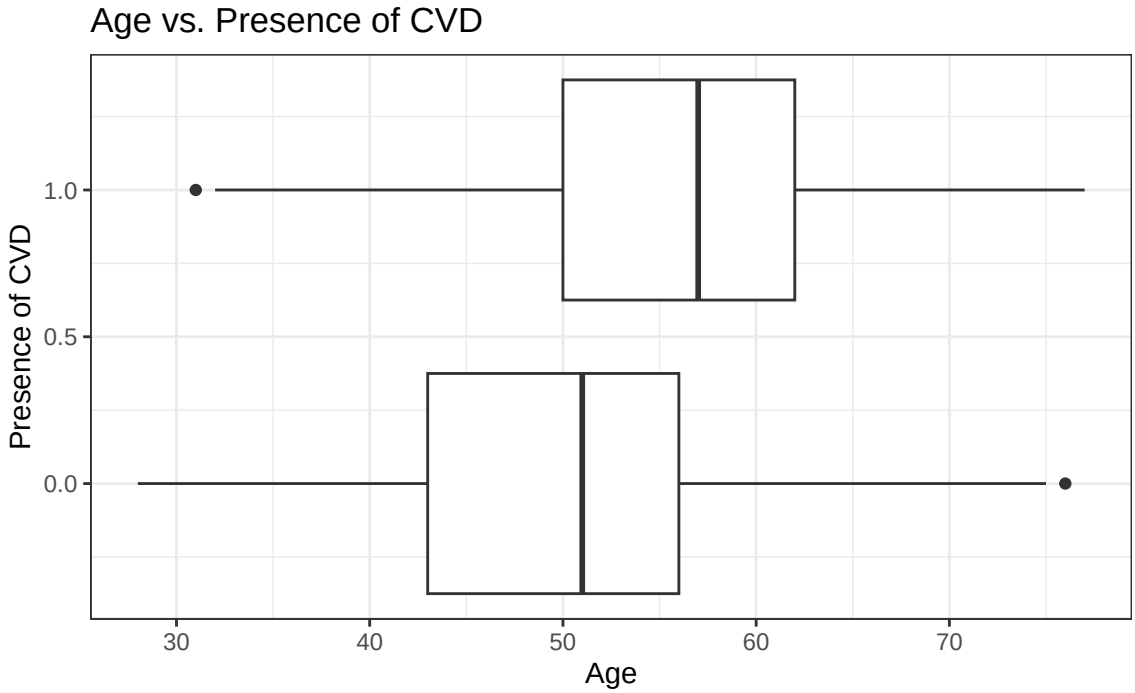
We decide to use complete case analysis because it is relatively easy to implement, and it can reduce bias in the estimates of the variable of interest when the missing data are missing completely at random or missing at random. Although a complete case analysis in this case would lead to an approximately 20% sample loss, we believe that it is better than imputation because our current knowledge does not allow accurate prediction of cholesterol levels of these observations.

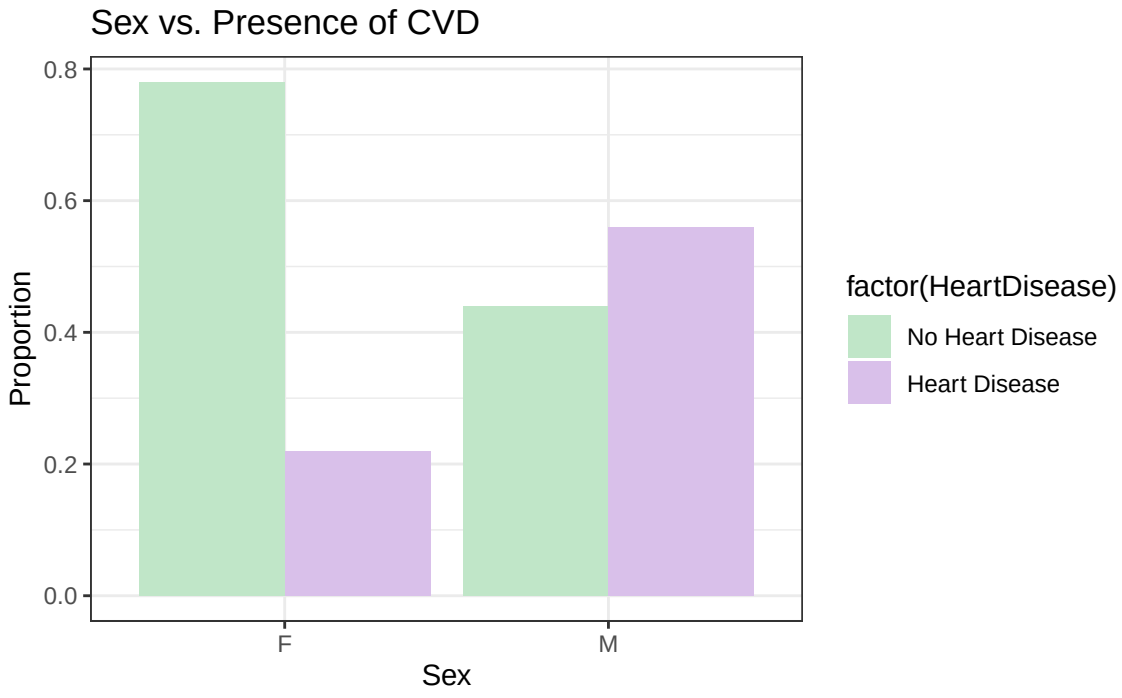
The distribution of CVD in our data appears to be relatively evenly distributed between patients with and without the disease. This suggests that there are sufficient data available for both categories of the response variable to perform further statistical analysis and develop a predictive model.



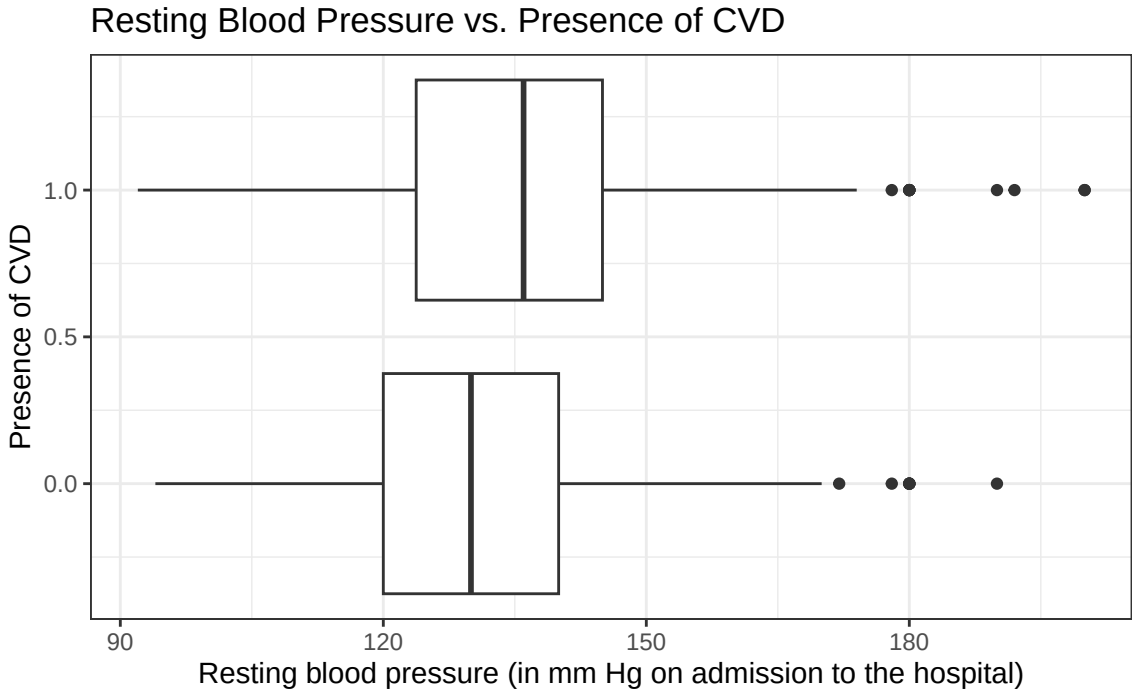
Previous research has identified age, resting blood pressure, serum cholesterol, the presence of exercise-induced angina, and maximum heart rate as indicators of the presence of heart disease (Hajar, 2017). In addition, we suppose that sex should be an important predictor of CVDs because men are at a higher risk of developing CVDs than women (Maas, 2010). This could be due to hormonal and physiological differences as well as lifestyle factors such as smoking and high alcohol consumption in men, on average. As such, we will examine these predictors

mentioned above in our EDA. However, we will still consider all of the remaining variables in when constructing our own model in the later sections.

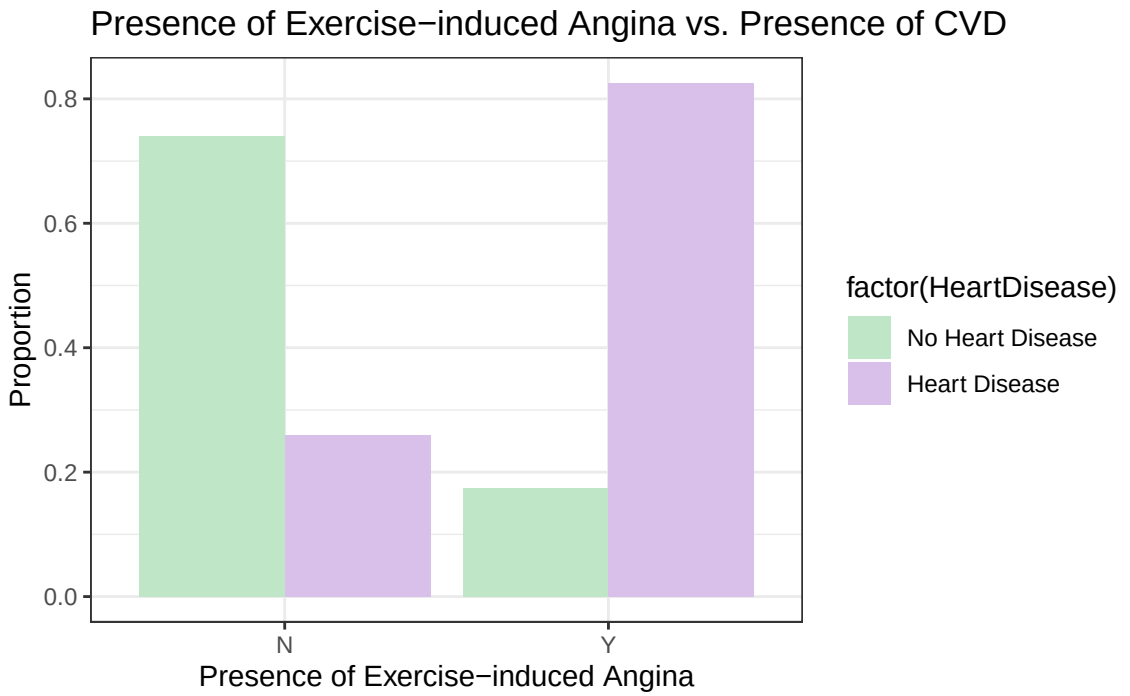




There seems to be a sex difference in the prevalence of CVDs. There is a nearly even split between men who do/do not have a CVD; about 45% of the men do not have a CVD, while about 55% of them do. However, just over 20% of the women in this dataset have a CVD.



It seems that patients diagnosed with CVD have higher resting blood pressure, which is consistent with the aforementioned literature findings.



Over 80% of patients with exercise-induced angina have CVD, whereas only around 25% of patients without exercise-induced angina have CVD. Therefore, exercise-induced angina might be an important predictor for CVDs.

Methodology

We use a logistic regression model because we are trying to predict the log-odds of getting a CVD, a binary response variable. Since all predictors are associated with HeartDisease based on literature as established in the EDA, and the there are only 11 predictors in the dataset, we plan to start with an all subset selection method. An all subset selection takes all possible combinations of variables and predicts combination of variables leads to the "best" result. The results are evaluated based on Mallow's Cp values. If adding new parameters doesn't meaningfully decrease squared errors, the overall Cp score will be higher, hence the model with the lowest Cp score through an all subset selection will be chosen as the "best" model.

All subset selection

```
 [1] 255.73724 142.24583  93.94848  78.45767  42.99584  33.63583  23.26726
 [8]  17.63977  11.64749  10.81918  10.67854  10.93762  12.19263  14.09560
[15]  16.00000
```

The model with the lowest Cp score included 8 predictors: Age, Sex, ChestPainType, RestingBP, FastingBS, ExerciseAngina, Oldpeak, ST_slope. After applying stepwise selection in both directions to the dataset, the same 8 predictors were selected. This confirms that the 8 predictors selected by the all subset selection approach are meaningful.

did not choose: cholesterol, resting ecg, max hr (need to maybe say why didn't include, because EDA mean is reversed? Maybe because some people with heart disease took medicine or are older has lower maxHR, while others have higher HR as expected)

All stepwise

```
Call:
lm(formula = HeartDisease ~ ST_Slope + ChestPainType + Sex +
    ExerciseAngina + Oldpeak + Age + FastingBS + RestingBP, data = data_new)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01279 -0.12978  0.00683  0.15934  1.01115

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.0377513  0.1295794   0.291 0.770876
ST_SlopeFlat      0.1650154  0.0560800   2.943 0.003358 **
ST_SlopeUp       -0.2338131  0.0633769  -3.689 0.000242 ***
ChestPainTypeATA -0.2165109  0.0357433  -6.057 2.21e-09 ***
ChestPainTypeNAP -0.2204843  0.0329386  -6.694 4.32e-11 ***
ChestPainTypeTA  -0.2105406  0.0556890  -3.781 0.000169 ***
SexM              0.1804987  0.0286867   6.292 5.38e-10 ***
ExerciseAnginaY   0.1512692  0.0317113   4.770 2.22e-06 ***
Oldpeak           0.0525977  0.0150353   3.498 0.000497 ***
Age               0.0035397  0.0013990   2.530 0.011613 *
FastingBS         0.0504047  0.0336769   1.497 0.134897
RestingBP         0.0010714  0.0007317   1.464 0.143510
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3265 on 734 degrees of freedom
Multiple R-squared:  0.5795,    Adjusted R-squared:  0.5732
F-statistic: 91.96 on 11 and 734 DF,  p-value: < 2.2e-16
```

Since our outcome is binary, we decided to fit a logistic model using the 8 predictors selected by the all subset selection approach and confirmed by the stepwise selection approach.

Model attempts

```
# A tibble: 12 x 5
   term             estimate std.error statistic       p.value
   <chr>               <dbl>     <dbl>     <dbl>         <dbl>
 1 (Intercept)       -4.95     1.26       -3.93  0.0000851
 2 Age                0.0320   0.0135      2.37  0.0177
 3 SexM               1.79     0.307       5.84  0.00000000529
 4 RestingBP          0.0125   0.00722     1.73  0.0837
 5 FastingBS          0.323    0.326       0.989 0.323
 6 ChestPainTypeATA  -1.70     0.349      -4.87  0.00000113
 7 ChestPainTypeNAP  -1.60     0.296      -5.39  0.0000000715
 8 ChestPainTypeTA   -1.64     0.472      -3.47  0.000517
 9 ExerciseAnginaY    0.888    0.262       3.39  0.000712
10 Oldpeak            0.412    0.140       2.95  0.00319
11 ST_SlopeFlat       1.30     0.517       2.51  0.0120
12 ST_SlopeUp        -1.22     0.559      -2.19  0.0286
```

Using a $\alpha = 0.05$ significance level, the p-values of RestingBP and FastingBS are respectively 0.084 and 0.32, which is larger than 0.05, suggesting the these 2 predictors might not be statistically significant. Hence, we considered a second logistic model with the 6 predictors left.

```
# A tibble: 10 x 5
   term             estimate std.error statistic       p.value
   <chr>               <dbl>     <dbl>     <dbl>         <dbl>
 1 (Intercept)       -3.60      0.947      -3.80 0.000143
 2 Age                0.0389    0.0132      2.96 0.00310
 3 SexM               1.79      0.300       5.98 0.00000000222
 4 ChestPainTypeATA  -1.66      0.344      -4.81 0.00000154
 5 ChestPainTypeNAP  -1.58      0.296      -5.34 0.0000000938
 6 ChestPainTypeTA   -1.51      0.463      -3.26 0.00111
 7 ExerciseAnginaY    0.935     0.261       3.59 0.000331
 8 Oldpeak            0.421     0.138       3.06 0.00223
 9 ST_SlopeFlat       1.24      0.510       2.43 0.0152
10 ST_SlopeUp        -1.27      0.552      -2.31 0.0208
```

#exercise angina Exercise-induced angina refers to chest pain or discomfort that occurs during physical activity, such as exercise, and is usually a result of reduced blood flow to the heart. It can be a symptom of underlying heart disease, such as coronary artery disease, which can be caused by the buildup of plaque in the arteries that supply the heart with blood. Therefore, the presence or absence of exercise-induced angina can be a useful diagnostic tool in identifying individuals at risk for or already suffering from heart disease.

#old peak The variable "Oldpeak" refers to the ST segment depression induced by exercise relative to rest. It is a numeric variable that measures the difference in the ST segment of an electrocardiogram (ECG) before and after exercise. A positive value of Oldpeak indicates that the ST segment is depressed during exercise compared to at rest, which can be a sign of ischemia (insufficient blood flow) to the heart. The greater the magnitude of the depression, the more severe the ischemia. Therefore, Oldpeak is a useful predictor of the likelihood of heart disease and can be used in models to predict the risk of heart failure.

#st slope The variable "ST_Slope" makes sense to include, as the ST segment exercise test, is one of the most widely used tests to screen for CVDs. Either depression (horizontal or downsloping) in the slope often suggests presence of CAD and warrants further management (Lim, 2016).

Finally, we fitted a logistic model with all predictors included in order to provide a baseline of comparing the models.

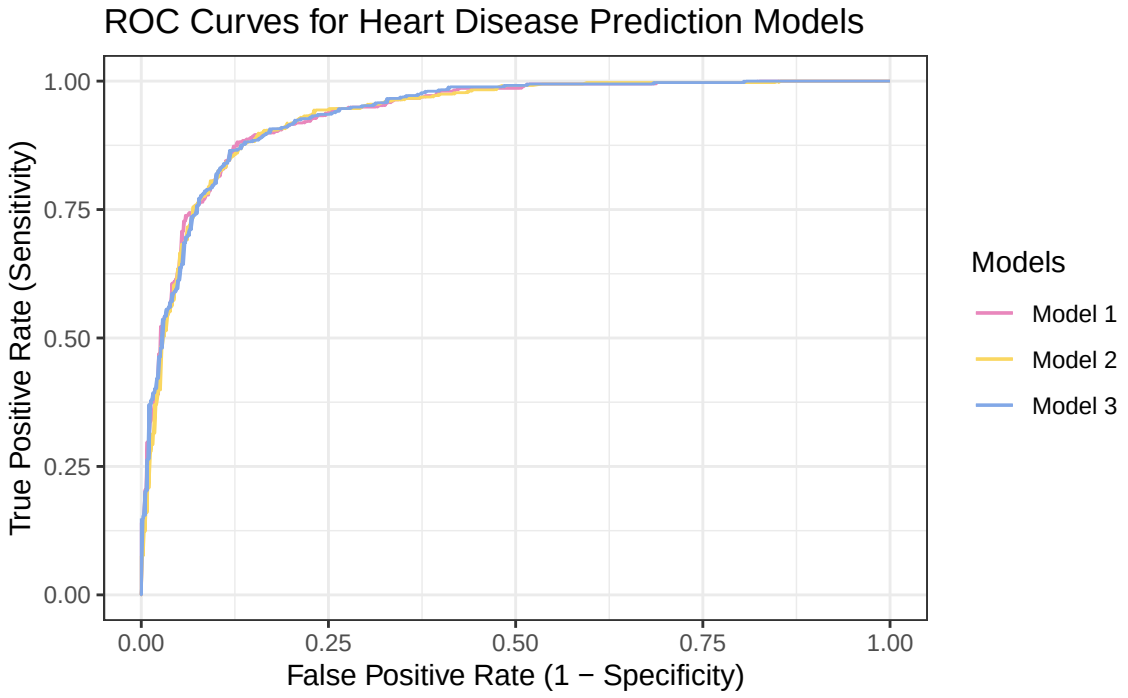
```
# A tibble: 16 x 5
   term              estimate std.error statistic       p.value
   <chr>                <dbl>     <dbl>     <dbl>         <dbl>
 1 (Intercept)      -5.44       1.76       -3.08  0.00204
 2 Age               0.0314     0.0148      2.12  0.0341
 3 SexM              1.87       0.313       5.95  0.00000000264
 4 ChestPainTypeATA -1.67       0.354      -4.72  0.00000235
 5 ChestPainTypeNAP -1.57       0.303      -5.19  0.000000208
 6 ChestPainTypeTA  -1.63       0.484      -3.38  0.000736
 7 RestingBP         0.0118     0.00730     1.61  0.107
 8 Cholesterol       0.00250    0.00198     1.26  0.207
 9 FastingBS         0.292      0.331       0.883 0.377
10 RestingECGNormal -0.230      0.284      -0.809 0.419
11 RestingECGST     -0.175      0.394      -0.443 0.658
12 MaxHR             0.000581   0.00578     0.100 0.920
13 ExerciseAnginaY   0.907      0.267       3.40  0.000682
14 Oldpeak           0.411      0.141       2.92  0.00349
15 ST_SlopeFlat      1.30       0.520       2.51  0.0121
16 ST_SlopeUp       -1.21       0.566      -2.14  0.0324

# A tibble: 1 x 3
```

```
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.934
```

```
# A tibble: 1 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.933
```

```
# A tibble: 1 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.935
```




```
Analysis of Deviance Table

Model 1: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + +Oldpeak +
    ST_Slope
Model 2: HeartDisease ~ Age + Sex + RestingBP + FastingBS + ChestPainType +
    ExerciseAngina + Oldpeak + ST_Slope
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
1       736     490.74
2       734     486.13  2   4.6102  0.09975 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Analysis of Deviance Table

Model 1: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + +Oldpeak +
    ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol +
    FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
    ST_Slope
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       736     490.74
2       730     483.58  6   7.1638    0.306
```

The p-values are 0.10 and 0.306 (>0.05 significance level) when comparing model 2 to model 1, and model 2 to model 3. This indicate that there is not enough evidence to reject the null hypothesis that all of the slopes corresponding to the eliminated terms (Cholesterol, RestingECG, MaxHR, RestingBP, and FastingBS) are zero. Therefore, there is not enough evidence to conclude that the model with all predictors is significantly better than the model we created through all subset selection. We would prefer to use the simpler Model 2 (that we came up with through all subset selection and p-value emliniation) over the more complex Model 3 (which includes all of the predictors), as it has fewer variables and possibly has clearer interpretations.

```
Analysis of Deviance Table

Model 1: HeartDisease ~ Age + Sex + RestingBP + FastingBS + ChestPainType +
    ExerciseAngina + Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + +Oldpeak +
    ST_Slope
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       734     486.13
2       736     490.74 -2  -4.6102  0.09975 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

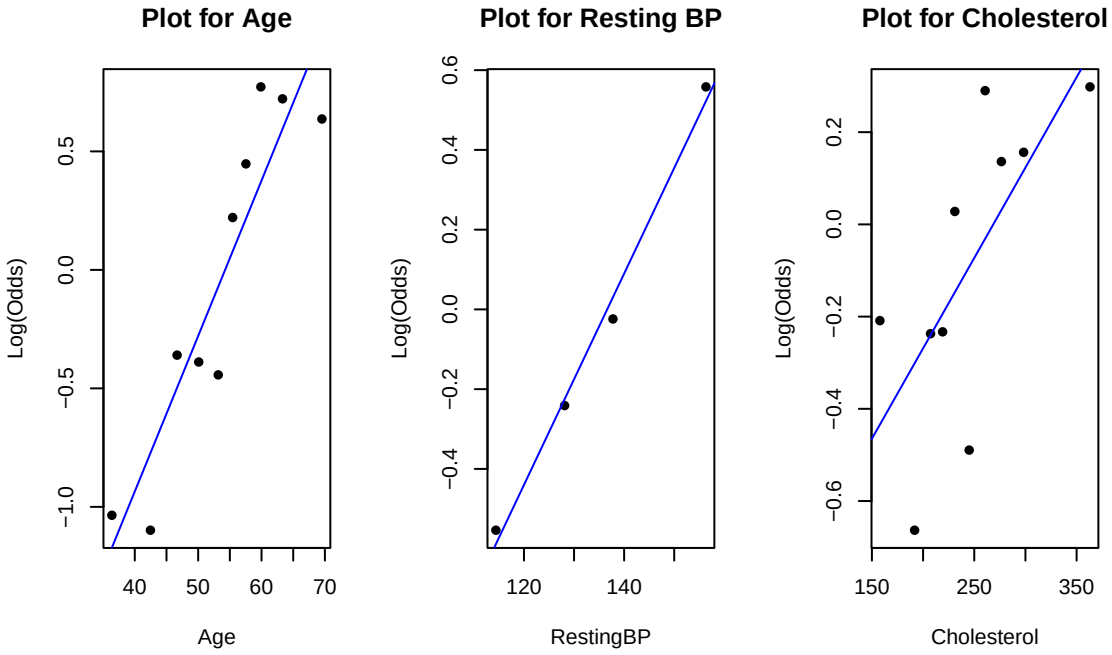
explain this!! mention that the cut off for the predictions is 0.5!

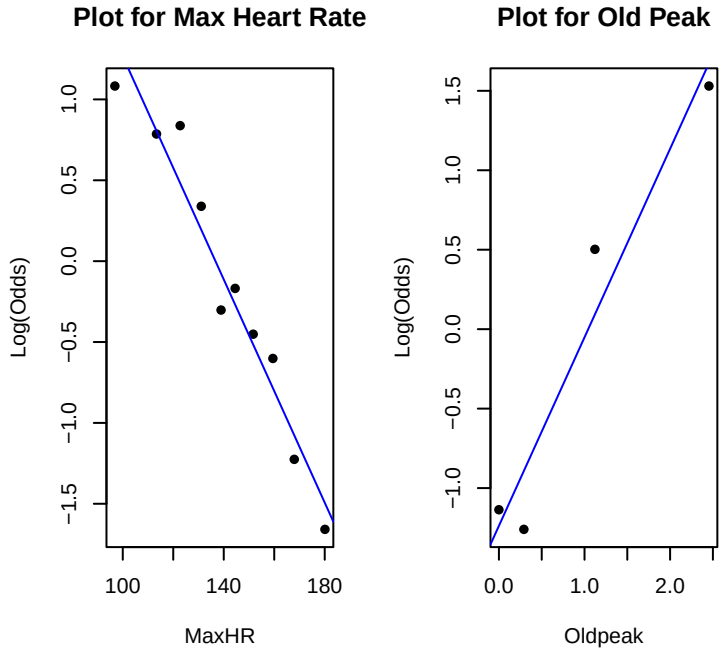
Assumptions

The two assumptions that are important in logistic regression are independence and linearity.

Independence: There might be violations of independence. The dataset is a combination of five datasets from clinics at Cleveland, Hungarian, Switzerland, Long Beach VA, and the Stalog (Heart) Dataset. If, for example, the different institutions had different in measuring Max heart rate, or the Oldpeak that would lead to systematic differences, then knowing one observation might yield information on another observation. However, since the exact data collecting mechanism is unknown, and we are not knowledgable information to make deductions and act upon violations of independence, we assume that independence is not violated.

Linearity: The 5 continuous variables that are involved in all models are Age, RestingBP, Cholesterol, MaxHR, and Oldpeak. As shown in all five plots below, the dots are randomly scattered across the empirical logits plots. Therefore, the linearity assumption is satisfied for the model.





#cholesterol Research suggests that low-density lipoprotein cholesterol (LDL-C, the "bad" cholesterol) is not the best predictor of risk for a major coronary event in people who are generally healthy otherwise.

#resting ecg The variable "RestingECG" may not be the best predictor of heart disease because it is only measuring the resting electrocardiogram results. Many people with heart disease may have a normal resting electrocardiogram, while others without heart disease may show abnormalities in their resting electrocardiogram. In contrast, other variables in the dataset, such as "ExerciseAngina" and "Oldpeak", are more directly related to the physiological response to exercise, which can be a stronger predictor of heart disease. Additionally, "RestingECG" only provides information about the electrical activity of the heart at rest, while "ExerciseAngina" provides information about the heart's response to physical stress, which can be a more sensitive indicator of heart disease.

#max hr Max heart rate is not the best predictor of heart disease because it is highly confounded by age. As we age, our maximum heart rate decreases, so older individuals will have lower maximum heart rates even if they are healthy. Furthermore, medications for CVDs include beta blockers, which naturally slows down a person's heart rate.

Results

#Equation

$log(odds) = -3.60 + 0.039(Age) + 1.792(SexMale) - 1.656(AtypicalAnginalChestPain) - 1.579(NonAnginalChe$

After evaluating and comparing the model performances using the ROC curve, the AUC metric, and an F-test (all three models are nested), we see that all three models achieved very similar performance in terms of AUC score (0.934 for model 1, 0.933 for model 2, and 0.935 for model 3). We selected model 2 as our final model, although it seems to have the lowest AUC score among all three models, because our F-tests show that model 2 is not significantly different from model 1 and model 3. This suggests that there is not sufficient evidence indicating that it is helpful to additionally include the other five predictors Cholesterol, RestingECG, MaxHR, RestingBP, and FastingBS in the model.

The intercept for our model represents a newborn female with no chest pain, no exercise-induced angina, no ST depression (0 degrees) on their ECG upon exercise that remains flat throughout exercise. The intercept value corresponds to log-odds of having a CVD vs. not having a CVD being 0.027 ($e^{-3.6}$) in this population. Consistent with existing scientific literature, we see in our chosen model that as people get older, their log-odds of developing a CVD increases; specifically, their log-odds is expected to increase by $e^{0.039=1.04}$ with every passing year. We found that sex is positively correlated with the log-odds of developing a CVD, in that a male's likelihood of having a CVD is expected to be $e^{1.792} = 6.00$ times that of a female, holding all else constant. As mentioned before, this is consistent with existing scientific literature. We also found that if a patient has exercise angina, their likelihood of having a CVD is $e^{0.935} = 2.55$ times higher than that of a patient who doesn't have exercise angina, when holding all other variables constant. This makes sense because exercise angina has to do with discomfort when heart muscle receives insufficient blood and oxygen during physical activity, which is mostly due to narrowing of arteries and can lead to CVDs and heart failure. With every degree increase in the depression of the ST segment slope on a patient's ECG during exercise, their log-odds of developing a CVD is expected to increase by $e^{0.421} = 1.52$ times. In a similar train of logic, our model also shows that when compared to a decreasing ST slope during exercise, having an flat ST slope is expected to increase a patient's log-odds of developing a CVD by $e^{1.239} = 3.45$ and having an upwards ST slope is expected to decrease their log-odds by $e^{-1.275} = 0.28$.

An unexpected discovery from our model is that people who are asymptomatic for chest pain are predicted to have the highest logodds of developing CVD, while controlling for all other variables. This might be explained through two layers of analysis. First, this dataset is not collected randomly from any population, but rather from people who visit the clinics and are probably suspected to have heart diseases. Hence, there is some sort of self-selection to get these variables documented when a participant does not experience chest pain yet is suspected to have heart disease. Second, how long the patients have been diagnosed with CVD is unknown, therefore some patients might have actively engaged in treatment processes and no longer experience chest pain, while still under the diagnosis of CVD.

Discussion

In this section you'll include a summary of what you have learned about your research question along with statistical arguments supporting your conclusions. In addition, discuss the limitations of your analysis and provide suggestions on ways the analysis could be improved. Any potential issues pertaining to the reliability and validity of your data and appropriateness of the statistical analysis should also be discussed here. Lastly, this section will include ideas for future work.

#Summary Our model identified the main predictors for cardiovascular disease in this dataset and quantified their impact on the likelihood of having the disease. We found that traditional factors such as [cholesterol and blood pressure] lose their predictive power when richer data is available, such as [fluoroscopy and electrocardiogram results, presence of other diseases, chest pain, and exercise tests].

#What we've learned about our research question - newer research is emerging, some of the metrics that we commonly think of as highly predictive actually aren't that good (can use cholesterol as an example) –> After identifying 172 data points with zero cholesterol levels, which is biologically implausible, we opted to exclude them from our study. Nonetheless, we noticed that these eliminated observations had a greater probability of presenting CVD than the average in the complete dataset. Therefore, the exclusion of these data points may have decreased the reliability of our analysis by reducing the sample size from 918 to 746, which could have introduced bias into our results. The validity of our study might be compromised due to the exclusion of this relevant information.

#Limitations and Improvements Our analysis is limited due to the non-representativeness of the dataset, which only includes individuals admitted to hospitals for heart disease-related conditions. As a result, our findings cannot be generalized to the broader population. To address this limitation, future studies should aim to expand the sample size and include a more diverse population to enhance the generalizability of the results. Additionally, we suggest that future research should focus on developing predictive models for heart disease in the general population by collecting data from the general population, as early detection can facilitate lifestyle changes or prompt timely treatment.

#Limitations of our analysis - better ways we couldve done it? –> adding interaction terms Age and Sex: Studies have shown that the relationship between age and heart disease risk may differ by sex, with women generally having a later onset of heart disease compared to men.

Chest Pain Type and Sex: Some studies suggest that there may be differences in the way men and women experience chest pain during a heart attack, which could potentially impact the predictive power of chest pain type in a heart disease model.

Exercise Angina and Oldpeak: Exercise-induced angina and ST depression during exercise are both markers of ischemia (reduced blood flow to the heart), and may therefore be related to each other in predicting heart disease risk.

Oldpeak and ST Slope: The shape of the ST segment during exercise can be indicative of different types of ischemia, and may therefore have a complex relationship with the severity of heart disease risk as measured by the degree of ST depression (Oldpeak).

#talk about how the variables were already narrowed down, making it hard for us It is important to emphasize that the list of 11 variables under consideration in this study has been narrowed down by researchers from a list of 76, involving several previous research studies to confirm this cut. As a result, further reduction of these variables from a model including all 11 variables could be somewhat challenging.

```
[1] 746  12
```

```
Area under the curve: 0.9066
```

```
Area under the curve: 0.9027
```

we also attempted a simple hold-out method, in which 80% of the data was used as a training set and 20%

References: Centers for Disease Control and Prevention, National Center for Health Statistics. About Multiple Cause of Death, 1999–2020. CDC WONDER Online Database website. Atlanta, GA: Centers for Disease Control and Prevention; 2022. Accessed May 1, 2023.

Damen, J. A., Hooft, L., Schuit, E., Debray, T. P., Collins, G. S., Tzoulaki, I., Lassale, C. M., Siontis, G. C., Chiocchia, V., Roberts, C., Schlüssel, M. M., Gerry, S., Black, J. A., Heus, P., van der Schouw, Y. T., Peelen, L. M., & Moons, K. G. (2016). Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ (Clinical research ed.), 353, i2416. https://doi.org/10.1136/bmj.i2416

Downs, J. R., Clearfield, M., Weis, S., Whitney, E., Shapiro, D. R., Beere, P. A., Langendorfer, A., Stein, E. A., Kruyer, W., & Gotto, A. M., Jr (1998). Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. JAMA, 279(20), 1615–1622. https://doi.org/10.1001/jama.279.20.1615

Fedesoriano. (September 2021). Heart Failure Prediction Dataset. Retrieved [May 2, 2023] from https://www.kaggle.com/fedesoriano/heart-failure-prediction.

Lim, Y. C., Teo, S. G., & Poh, K. K. (2016). ST-segment changes with exercise stress. Singapore medical journal, 57(7), 347–353. https://doi.org/10.11622/smedj.2016116

Maas, A. H., & Appelman, Y. E. (2010). Gender differences in coronary heart disease. Netherlands heart journal : monthly journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation, 18(12), 598–602. https://doi.org/10.1007/s12471-010-0841-y

Velagaleti, R. S., & Vasan, R. S. (2007). Heart failure in the twenty-first century: is it a coronary artery disease or hypertension problem?. Cardiology clinics, 25(4), 487–v. https://doi.org/10.1016/j.ccl.2007.08.010