Table of Contents

Project Name: Heart Failure Prediction with JMP

Prepared by: Group 8

Nov. 2021

Purpose of This Report

Heart failure has become a major problem for human health across the US and the globe. According to CDC data, heart failure has responsible for around 6.2 million occurrences among American adults, 379,800 American deaths in 2018 (13.4%), and cost the nation a fortune to its health care system of around $30.7 million. It's so common for the system to assume that men are much more prone to heart disease, which is true in many cases. However, the more attention on men will mean a less focus on women. According to American College of Cardiology, heart disease has become the leading killer for women (American College of Cardiology, 2017). Moreover, the health of the elders is much more likely to be threatened by heart disease. According to AHA Journals, heart disease, especially ischemic heart disease, makes the diagnosis in the elderly more difficult (AHA Journals). Our team uses the dataset that is collected from various sources to explore the how heart disease occurrences/symptoms differentiate among men and women, and the young and the elders, while developing a robust heart failure prediction formula to help improve heart disease diagnosis. Besides, our team provides constructive diagnosis advice that can be quick tests to improve heart failure prediction, including for the elderly.

After comparing several models' performances, our team chooses the logistic regression model as the best fit. In addition to the general model performance evaluation and interpretation, our team goes in-depth to interpret the relationships between the key predictors and the variable of interest. We do this when we are investigating the anomalies observed by trimming the predictors with other connected variables, a fundamental technique used in the models of the tree family (decision tree, boosted tree etc.). Our team also is aware of the limitations of the dataset and of our model. Finally, we provide constructive suggestions on reducing heart failure occurrences as well as improving heart disease diagnosis. Please note that throughout this report, we use heart failure and heart disease interchangeably.

An Introduction to Our Data

This dataset was created by combining different datasets already available but not combined before. In this dataset, five heart datasets are combined over 11 common features which makes it the largest heart disease dataset available so far for research purposes. The five datasets used for the curation are:

- Cleveland: 303 observations
- Hungarian: 294 observations
- Switzerland: 123 observations
- Long Beach VA: 200 observations
- Stalog (Heart) Data Set: 270 observations

This dataset contains records of physical factors (of or relating to human body) which we will study to predict how these factors affect the heart failure in humans. All predictors and the variable of interest are described and summarized as follows:

- Continuous variables:
  - Age: age of the patient [years]
  - RestingBP: resting blood pressure [mg Hg]
  - Cholesterol: serum cholesterol [mg/dl]
  - MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
  - Oldpeak: oldpeak = ST [Numeric value measured in depression]
- Categorical variables:
  - Sex: sex of the patient [M: Male, F: Female]
  - ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
  - FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
  - RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
  - ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
  - ST_Slope: the slope of ST segment [Up: up-sloping, Flat: flat, Down: down-sloping]
- Variable of Interest: HeartDisease: output class [1: heart disease, 0: Normal]

For all variables, we use Missing Value Analysis to explore missing values if any. Figure 1.1 tells good news that we don't have any missing values for this dataset. For continuous variables, we use Explore Outliers → Robust Fit Outliers to explore outliers. Figure 1.2 shows that we only have 3 outliers for the Oldpeak variable and 1 outlier for the RestingBP variable. The number of outliers is acceptably small, but our team has taken note of them in case the outliers have significant impacts on the model.

Missing Columns

Show only columns with missing

Close

Select columns and choose an action.

Select Rows | Color Cells | Exclude Rows | Color Rows

Column	Number Missing
Age	0
Sex	0
ChestPainType	0
RestingBP	0
Cholesterol	0
FastingBS	0
RestingECG	0
MaxHR	0
ExerciseAngina	0
Oldpeak	0
ST_Slope	0
HeartDisease	0

Figure 1.1

Robust Estimates and Outliers

Column	Huber Center	Huber Spread	Huber N Outliers
Age	53.529286	9.7617846	0
RestingBP	132.04022	17.527749	1
Cholesterol	198.03517	113.19474	0
FastingBS	0.2331155	0.4459343	0
MaxHR	136.97176	26.294455	0
Oldpeak	0.8549732	1.0181907	3

Figure 1.2

For all the categorical variables, the number of elements for each of them is acceptable (less than five). We preserve all the categorical variable as they are important, based on our research (see more details in Our Model Interpretation Part Two). For continuous variables, we perform Principal Components Analysis to see if the number of predictors can be reduced. Figure 1.3 shows the result of the principal components analysis in which all continuous variables should be retained as we aim to explain at least 95% of the data with our model.

Correlations

	Age	RestingBP	Cholesterol	MaxHR	Oldpeak
Age	1.0000	0.2544	-0.0953	-0.3820	0.2586
RestingBP	0.2544	1.0000	0.1009	-0.1121	0.1648
Cholesterol	-0.0953	0.1009	1.0000	0.2358	0.0501
MaxHR	-0.3820	-0.1121	0.2358	1.0000	-0.1607
Oldpeak	0.2586	0.1648	0.0501	-0.1607	1.0000

Principal Components / Factor Analysis

Principal Components: on Correlations

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	1.7071	34.142		34.142
2	1.1940	23.880		58.022
3	0.8314	16.627		74.649
4	0.6914	13.827		88.477
5	0.5762	11.523		100.000

Figure 1.3

Our Modeling Process

Up till now, we have completed the first three steps of SEMMA. Our data is independently **S**ampled and randomly selected. We have **E**xplored missing values and outliers. The data is already clean, so we do not need to **M**odify. From this section on, we will focus on **M**odel and **A**ssess. Before the modeling process, we preprocess the data by randomly partitioning the dataset with a rule of training : validation = 60% : 40%. We consider the following options from which we pick up a best model with the best performance in classification of the data:

- Neural Network
- Decision Tree
- Bootstrap Forest
- Boosted Tree
- Logistic Regression
- Naive Bayes
- K Nearest Neighbors (KNN)

Figure 2.1 and Figure 2.2 shows the performance of these models in terms of various measurements.

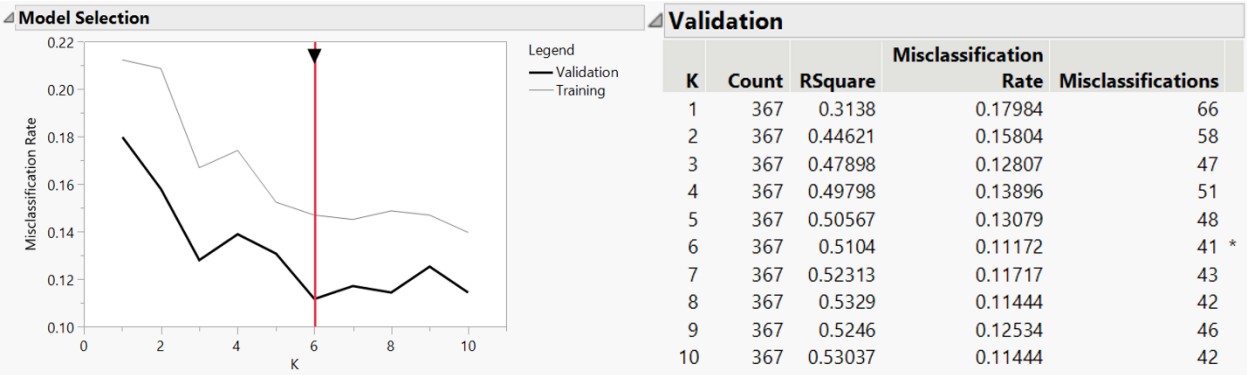
Model Comparison

Predictors

Measures of Fit for HeartDisease

Validation	Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N	AUC
Training	Neural Model NTanH(3)		0.5484	0.7088	0.3104	0.3062	0.1967	0.1216	551	0.9388
Training	Partition		0.4820	0.6485	0.3561	0.3350	0.2271	0.1597	551	0.9116
Training	Bootstrap Forest		0.7127	0.8361	0.1975	0.2284	0.1572	0.0544	551	0.9902
Training	Boosted Tree		0.7245	0.8441	0.1894	0.2309	0.1390	0.0762	551	0.9837
Training	Fit Ordinal Logistic		0.4907	0.6567	0.3501	0.3290	0.2152	0.1416	551	0.9218
Training	Naive Bayes		0.2354	0.3701	0.5256	0.3544	0.1680	0.1615	551	0.9154
Training	K Nearest Neighbors		.	.	.	.	.	0.1270	551	.
Validation	Neural Model NTanH(3)		0.5669	0.7246	0.2978	0.3007	0.1967	0.1253	367	0.9452
Validation	Partition		0.4477	0.6153	0.3797	0.3455	0.2462	0.1608	367	0.9048
Validation	Bootstrap Forest		0.5023	0.6675	0.3422	0.3221	0.2253	0.1471	367	0.9298
Validation	Boosted Tree		0.5280	0.6908	0.3245	0.3128	0.1874	0.1281	367	0.9325
Validation	Fit Ordinal Logistic		0.5656	0.7234	0.2987	0.2974	0.1986	0.1063	367	0.9455
Validation	Naive Bayes		0.3828	0.5477	0.4243	0.3171	0.1440	0.1253	367	0.9346
Validation	K Nearest Neighbors		.	.	.	.	.	0.1117	367	.

Figure 2.1 – Models Performances



Validation

K	Count	RSquare	Misclassification Rate	Misclassifications	
1	367	0.3138	0.17984	66	
2	367	0.44621	0.15804	58	
3	367	0.47898	0.12807	47	
4	367	0.49798	0.13896	51	
5	367	0.50567	0.13079	48	
6	367	0.5104	0.11172	41	*
7	367	0.52313	0.11717	43	
8	367	0.5329	0.11444	42	
9	367	0.5246	0.12534	46	
10	367	0.53037	0.11444	42	

Figure 2.2 – KNN Performance

We choose the best model according to the model's performance in the validation dataset. Logistic stands out of all with a very good RSquare (0.7234) and the lowest misclassification rate (0.1063), to which we put more weights when evaluating the accuracy of classifying elements of zero and one of a model.

Our Model Interpretation

Part One: A Bird's View

1. Effect Summary and Parameter Estimates

To see whether the dependent variables have statistical significance overall in predicting the independent variable, heart failure rate, we go through the *effect summary* on the logistic regression model performance page. In the rankings of p values of all the variables, those with statistical significance are the slope of ST segment (ST_Slope), chest pain type (ChestPainType), fasting blood sugar test (FastingBS), and exercise-induced angina (ExerciseAngina), given a confidence level of 95%. *Parameter Estimates* provides details of how each element within/among the predicting variables contributes to the predicted variable per incremental change in each element of the predicting variables from the benchmarking one. Noticing the intercept of the logistic regression model is zero, we can see that the model predicts the influence of these elements of the predicting variables on non-heart failure probability. A logistic regression model uses benchmarking elements from each predicting variables as starting points, and for each incremental change of these elements to a non-benchmarking one, the model records the change in the non-heart failure rate as wells as the corresponding p value of the change (i.e., the significance level).

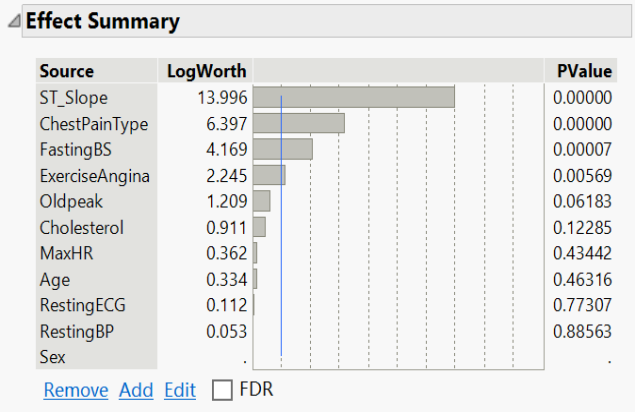


Figure 3.1

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	0.43477527	1.6853522	0.07	0.7964
Age	-0.0116734	0.015925	0.54	0.4635
Sex[F]	0.58993026	0.1734273	11.57	0.0007*
ChestPainType[ASY]	-0.9990148	0.2284674	19.12	<.0001*
ChestPainType[ATA]	0.61955363	0.2851755	4.72	0.0298*
ChestPainType[NAP]	0.58698531	0.2549065	5.30	0.0213*
RestingBP	-0.0010095	0.0070157	0.02	0.8856
Cholesterol	0.00196395	0.001278	2.36	0.1244
FastingBS	-1.3128792	0.3411926	14.81	0.0001*
RestingECG[LVH]	-0.1695769	0.2377944	0.51	0.4758
RestingECG[Normal]	0.05565011	0.1855823	0.09	0.7643
MaxHR	0.00514411	0.0065775	0.61	0.4342
ExerciseAngina[N]	0.43280932	0.1562961	7.67	0.0056*
Oldpeak	-0.2746584	0.1483459	3.43	0.0641
ST_Slope[Down]	-0.0528706	0.3716878	0.02	0.8869
ST_Slope[Flat]	-1.1301324	0.2319066	23.75	<.0001*

Figure 3.2

We summarize the how the model works in predicting heart failure rate with these predicting variable, either continuous or categorical, in the following table:

a	b	c	d	e	F
Columns	Benchmarking Element	Non-Benchmarking Element	Incremental Effect Estimated (b→c)	Relationship between c and Non-heart Failure Rate**	Relationship between c and Heart Failure Rate**
Age	0	>0	-0.01	NOB	NOB
Sex	Male	Female	0.59	P	N
ChestPainType	TA	ASY	-0.10	N	P
		ATA	0.62	P	N
		NAP	0.59	P	N
RestingBP	minimum	>minimum	-0.01	NOB	NOB
Cholesterol	minimum	>minimum	0.00	NOB	NOB
FastingBS	minimum	>minimum	-1.31	N	P
RestingECG	ST-T-wave abnormality	LVH	-0.17	N0B	NOB
		Normal	0.06	NOB	NOB
MaxHR	minimum	>minimum	0.01	NOB	NOB
ExerciseAngina	Y	N	0.43	P	N
Oldpeak	minimum	>minimum	-0.27	NOB	NOB
ST_Slope	UP	Down	-0.05	NOB	NOB
		Flat	-1.13	N	P

**P: the relationship is positive when p value < 0.05
**N: the relationship is negative when p value < 0.05
**NOB: no obvious relationship because p value > 0.05

Table 3.1

Again, *parameter estimates* provide detailed interpretation of the logistic regression model, in which the predicting variables indicate that, on average, female has lower heart failure rate, people with ASY chest pain type have higher heart failure rate, and the same goes with people with high fastingBS and people with flat ST_Slope, as compared with respective benchmarking element.

2. Prediction Profiler

Prediction profiler provides an interactive platform for us to see how each variable has an impact on the predicted heart failure rate (Figure 3.3). The contour shows how the probability of heart failure changes for different values of the predictors. We summarize the effect of each predictor as follows:

Figure 3.3

- Age: age has a positive correlation with heart failure rate -- as age increases, the probability of having a heart disease increases.
- Sex: males are more prone to heart failure as when we move the profiler towards females, the probability of having a heart disease decreases while the probability of heart failure increases when we move the profiler towards males.
- ChestPainType: an ASY chest pain has the highest heart failure rate. A TA angina follows at the second place in terms of heart failure rate.
- RestingBP: blood pressure when at rest has a positive but poor correlation with heart failure rate.
- Cholesterol: cholesterol seems to move in an opposite direction to the trajectory of heart failure rate.
- FastingBS: when the fasting blood sugar rate is greater than 120 mg/dl, the risk of heart failure is pretty high, compared to a healthy fasting blood sugar rate.
- RestingECG: heart failure rate is slightly higher when the RestingECG result is LVH, compared with the other types.
- MaxHR: it seems MaxHR is moving in a different direction to the trend of heart failure rate.
- ExerciseAngina: people who experience exercise-induced angina have higher chance of getting heart failure.
- Oldpeak: Oldpeak moves in the same direction as that of heart failure rate. As Oldpeak level grows, heart failure rate increases even faster.
- ST_Slope: people with flat ST_Slope have the greatest heart failure rate, compared with people with other ST_Slope types. A Down ST_Slope ranks the second as an attribute to heart failure.

3. Overall Strength-of-Fit

Before looking at single predictors, we first evaluate the overall strength-of-fit of our multivariate logistic regression model and try to answer the question, are the predictors doing better than a simple naïve model, the majority deciding the category? -LogLikelihood, from the Whole Model Test panel (Figure 3.4), is the statistic measures for the overall fitness of a logistic model. This is an estimate similar to a summary of errors. A lower -LogLikelihood will indicate a better model. In the "-LogLikelihood" column of Whole Model Test, we can see that a "Full" model, the target model, has a much lower -LogLikelihood than the "Reduced" model, the naïve one. The p value of the difference is so small that the improvement in classification accuracy by using the logistic model is statistically significant. Fit Details (Figure 3.5) provides more measures of the fit: misclassification is low and R square is decent for the validation dataset.

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	185.85388	15	371.7078	<.0001*
Full	192.90535			
Reduced	378.75923			

Figure 3.4

Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.4907	0.5656	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.6567	0.7234	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3501	0.2987	∑ -Log(ρ[j])/n
RASE	0.3290	0.2974	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2152	0.1986	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1416	0.1063	∑ (ρ[j]≠ρMax)/n
N	551	367	n

Figure 3.5

4. Lift Curve and Confusion Matrix

A lift curve is a visual that aids performance/effectiveness evaluation of the model. We usually interpret the curve with a rule in mind that the greater the area between the baseline and the curve, the better does the model perform. Generally, we have lift curves above the baseline 1.0 and we can see that the model is effective in classifying non-heart failure. When predicting non-heart failure, the model has a maximum lift of over 2.2 when the portion of the data arrives at around 10%. For predicting heart failure, the best lift of the model is slightly over 1.8 when the data portion is around 10%. Confusion Matrix in Figure 3.7 has proven the better performance in predicting zero: accuracy rates are 84.8% and 93.1% for predicting zero and one respectively.

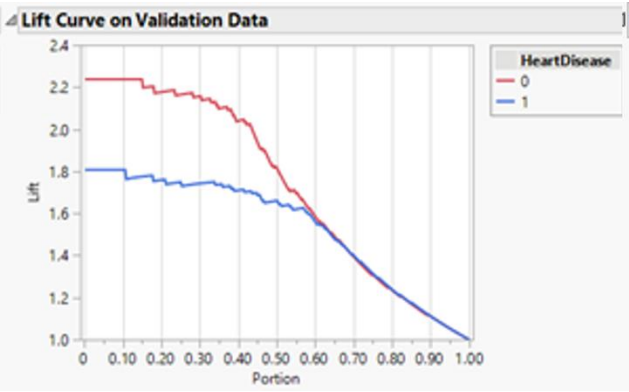


Figure 3.6

Confusion Matrix

Training

Actual HeartDisease	Predicted Count 0	1
0	202	44
1	34	271

Actual HeartDisease	Predicted Rate 0	1
0	0.821	0.179
1	0.111	0.889

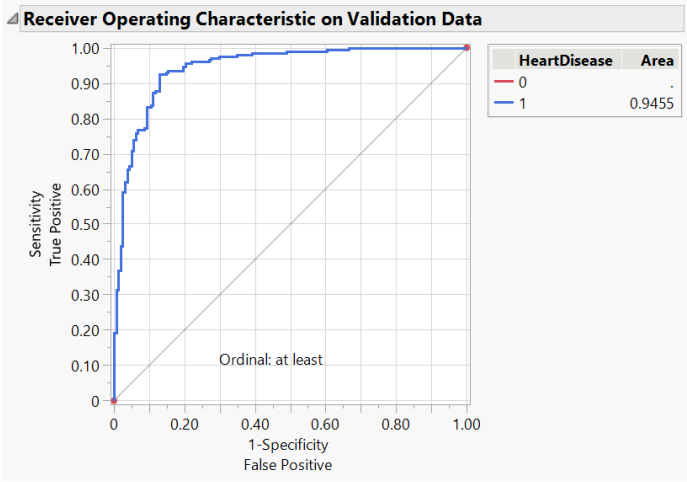
Validation

Actual HeartDisease	Predicted Count 0	1
0	139	25
1	14	189

Actual HeartDisease	Predicted Rate 0	1
0	0.848	0.152
1	0.069	0.931

Figure 3.7

5. Receiver Operating Frequency Curve (ROC)



We use ROC as a tool to summarize the model's performance in classifying with or without heart disease (Figure 3.8). Using the numerous trade-off points between true positive (type II error) and false positive (type I error) error rates, we can predict how accurately the model classifies an event(sensitivity), with the cut-off value of 0.5. The farther the ROC is away from the baseline (the one between (0, 0) and (1, 1)), the greater the area below the ROC down to the point (1, 0), the higher the accuracy of the model in predicting heart failure. The area of 0.9455 means an excellent accuracy of the diagnostic test.

Figure 3.8

Part Two: In-depth Model Interpretation

1. Age and Sex

While evaluating the logistic regression model performance and interpretating the model itself on a high level, our team is intrigued by the underlying insightful variable relationships with which we can explore more business intelligence to explain the occurrence of heart failures. In this section, we focus on how key variables, namely ST_Slope/Oldpeak, ChestPainType, FastingBS, ExerciseAngna, and cholesterol will interact with heart failure rate, when stratified into different gender groups and age groups—we use a very popular method when performing statistical analysis. By investigating heart failure occurrences of these different groups, we believe intellectual variable relationship can be mined to guide us in providing constructive medical suggestions to prevent heart disease.

Even though the variable age and sex have no obvious relationship with the occurrence of heart failure (Table 3.1), they can help the model interpretation go deeper when stratifying other variables into different groups. For the variable age, we divide the year line into different age groups with equal interval of ten years, i.e., [0, 10), [10, 20) … [70, 80). For gender groups, of course, we have male and female. Knowing male has much higher chance of getting heart disease on average than female, we are interested in how heart failure rate differentiates among male and female of different age groups. Looking at the frequencies of records of heart failure, we can see that, overall, male has much more records than female. For each the two gender groups, the frequency distribution is very similar to a normal one. However, when looking at the number of women of age group [70, 80) and of men of age group [20, 30), we have very few records to estimate the population parameter of interest (only 6 and 4 respectively for male and female of these age groups). Hence, we are not confident that the data of these groups can represent the population of such age groups.

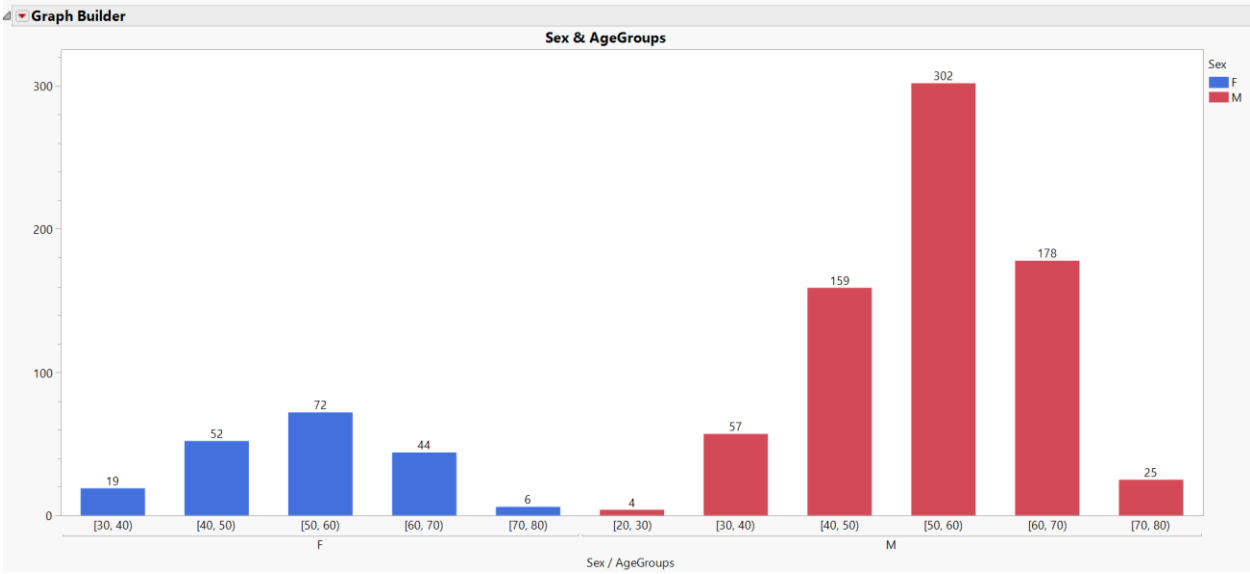


Figure 3.9

Using Graph Builder, we can see how the probability of heart failure -- the mean of heart failure parameters in (0,1) -- fluctuates along the age line for male and female. Overall, the older the person is, the higher chance of getting heart failure disease the person has; Male, on average, has higher chance of getting heart disease than female. For male, the heart failure rate increases from 0.3 to around 0.7 when they are getting older. When it comes to female, heart failure rate remains low at around 0.1 by the age of about but increases rapidly afterwards. The curve (the black one) then arrives at a peak of around 0.6 at the age of 62 but plummets to zero out of our expectation (the red curve). We investigate the outliers from 70 years old to 80 years old on a model level and realize that the six records of the age group [70, 80) are all with no heart failure occurrences (Figure 3.9), which is not representative to the population at large.

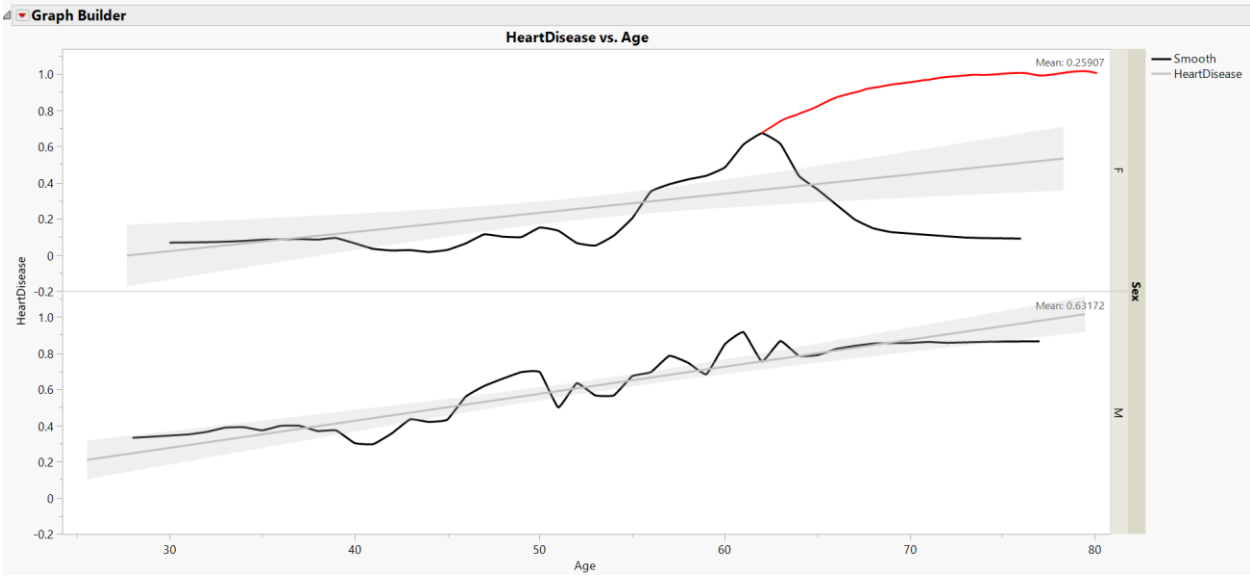


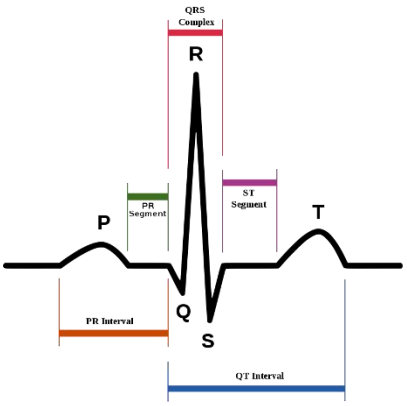
Figure 3.10

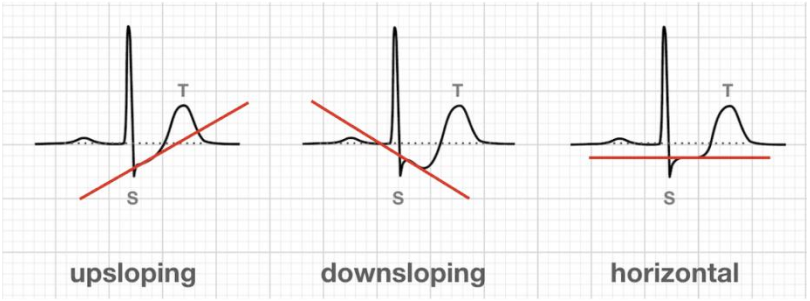
Confused by the huge discrepancy of what the data we have tells and what we have seen in the real world, we take further steps to do some research in how ages influence male and female on heart failures. The topic of age versus heart failure is not new and there are many reliable external sources our team can refer to. It seems to be common for people to assume that men are more likely to experience heart failure than women, though in many cases, it is true, but things are more complicated than that. According to Dr, Jen Tan's research, men are more at risk of heart disease up until the age of 75, while women are more likely to develop problems after the age of 75 (Jen Tan, 2017). This is in line with what we are expecting of the trajectory of women's heart failure rate over the ages (the red line).

The statistics have proven the limitations of our dataset for women from age 70 to 80. The bias of the model in predicting heart failure rate of women of this age group has alerted our team to our poor attention in women's heart failure data, which is in much smaller scale than men's (M : F=725 : 193). [1]Given that age has much less contribution to the model (Figure 3.1), as compared to other kay variables, we take no further actions to increase our sample size for women in this case.

2. ST_Slope and Oldpeak

By obtaining an understanding of how heart failure rate differentiates among male and female of different age groups, our team now can focus on investigating how other key variables will have impacts on heart disease in detail. The first box to tick is ST_Slope, the slope of ST segment[2]. Walking through the rankings of variables in effect summary, we can see that ST_Slope stands at the first place, meaning ST_Slope is the most effective predictor of heart disease. Robert et al. at their research paper, The ST Segment/Heart Rate Slope as a Predictor of Coronary Artery Disease, have examined that ST/heart rate slope is a more accurate ECG criterion [3]for diagnosing significant coronary artery disease (CAD) (Robert et al., 2004). A website called Life in the Fast Lane, has presented an easily understanding category of the types of ST_Slope (Figure 3.12) (Ed Burns & Robert Buttner, 2021).





[1] A note from the editor: considering the much higher male proportion of data, we interpret males' contribution to a variable of interest as more significant to women's when and only when the contribution > M/(M+F)=80%, viso versa for women.

[2] In electrocardiography, the ST segment connects the QRS complex and the T wave and has a duration of 0.005 to 0.150 sec (5 to 150 ms) (Figure 3.11). The normal ST segment has a slight upward concavity; Flat, downsloping, or depressed ST segments may indicate coronary ischemia; ST depression (Figure 3.16) may be associated with subendocardial myocardial infarction, hypokalemia, or digitalis toxicity.

[3] ECG criterion is the process of producing an electrocardiogram in Electrocardiography.

Figure 3.11 **Figure 3.12**

Our predictors profiler indicates that people with a flat ST_Slope type have much higher heart failure rate whilst those with an upsloping ST segment are healthier (Figure 3.3). Using a mosaic plot, we can see the heart failure proportion for each ST_Slope type (Figure 3.13). The red map covers most part of the "Down" and "Flat" ST_Slope columns, both of which have very close high heart failure rates. For people with a flat ST_Slope type, the proportion of people diagnosed with heart disease is extremely high (82.8%). This means that down and flat ST_Slopes are very good indicator for heart failure diagnostic purposes.

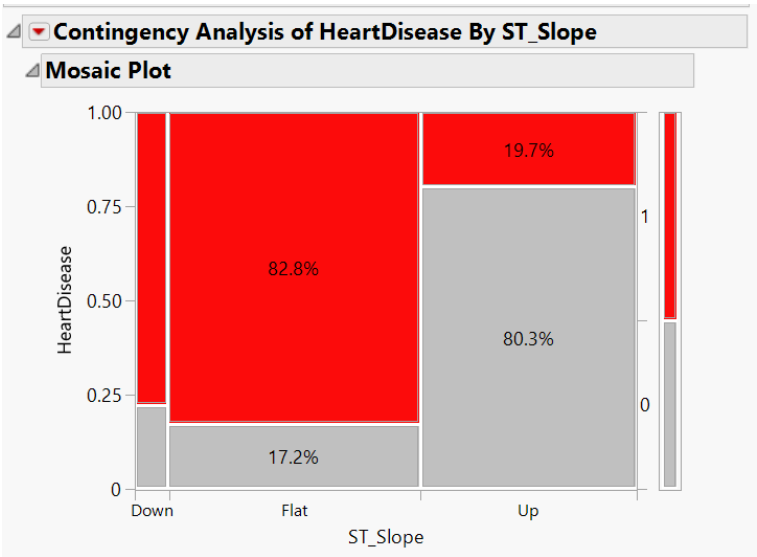
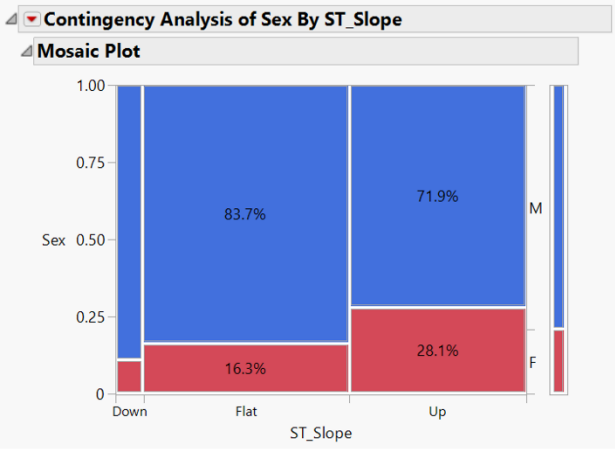


Figure 3.13

Also, our team is interested in the demographic features of these people of different ST_Slope types. Figure 3.14 and Figure 3.15 visualize that male (given 83.7%>M/(M+F) =80%) and older person are more like develop down and flat ST_Slopes, and hence more prone to heart disease. These insights are in line with what we have learned in section Age and Sex.



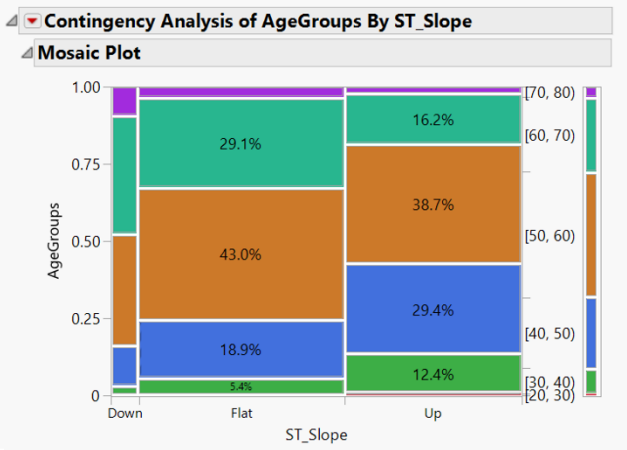


Figure 3.14 **Figure 3.15**

Apart from ST_Slope, our model incorporates Oldpeak, ST depression [4]to facilitate heart failure diagnosis. Figure 3.16 presents a clear picture of what a ST depression looks like on an electrocardiogram. Combining Figure 3.12 and Figure 3.16, we can see that there is no obvious relationship between ST depression and ST_Slope types, but to see whether heart failure diagnosis can be improved, our team liaise ST depression with the types of ST_Slope.

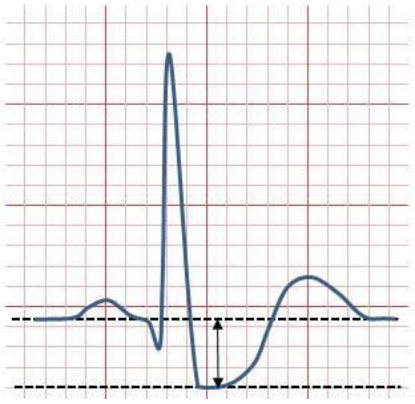


Figure 3.16 – ST depression

Dr. Araz Rawshani on his paper published on ECG & ECHO Learning suggests that ST segment depression less than 0.5 mm is acceptable but a level of 0.5 mm or more is considered pathological (Dr. Araz Rawshani, 2018). We stratify the Oldpeak variable into nineteen sub-groups each of which has equal interval of 0.5 mm. We notice that there are negative ST depressions (Figure 3.17 to Figure 3.22), which probably the results of poor skin contact of the electrode. However, given the high heart failure rates among these negative depression groups, we consider them pathological ST depression results. Our team summarizes our findings when walking through each of the mosaic plots by ST_Slope types from Figure 3.17 to Figure 3.22 as follows:

- ST depression is a highly positively related to heart failure rate when the level is extremely high (over 1.5 mm) or abnormally negative.
- When the ST_Slope is down (unhealthy), combining the ST depression variable helps improve the heart disease diagnosis when the ST depression is high (over or equal to 1.5 mm) – given the average heart failure rate is greater than 77.8% in Figure 3.17.
- When the ST_Slope is flat (unhealthy), combining the ST depression variable helps improve the heart disease diagnosis when the ST depression is high (over or equal to 2 mm) – given the average heart failure rate is greater than 82.5% in Figure 3.19.
- When the ST_Slope is up (healthy), combining the ST depression variable would help little to improve the heart disease diagnosis unless the depression level is highly pathological[5].

[4] ST depression refers to a finding on an electrocardiogram, wherein the trace in the ST segment is abnormally low below the baseline (Figure 3.16).

[5] Depressed but upsloping ST segment generally rules out ischemia (a kind of heart failure) as a cause – Wikipedia.

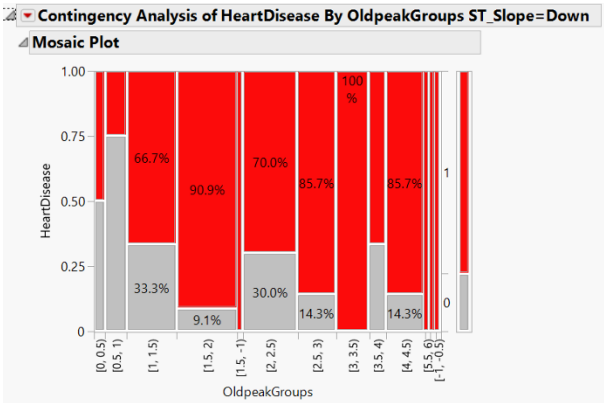


Figure 3.17 – Down ST_Slope

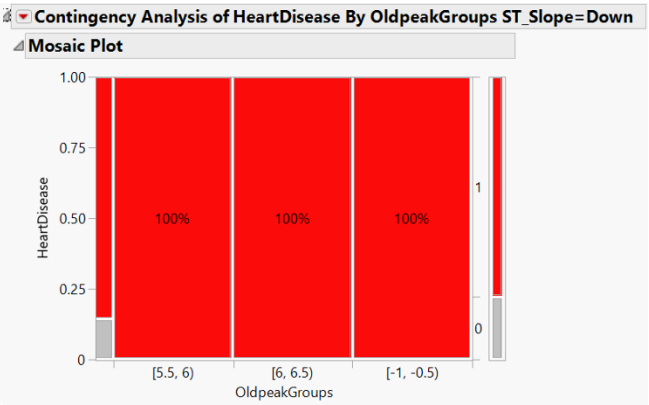


Figure 3.18 – Down ST_Slope (Zoomed in)

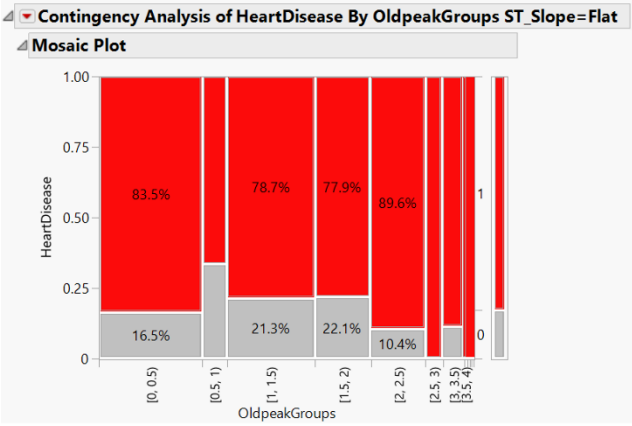


Figure 3.19 – Flat ST_Slope

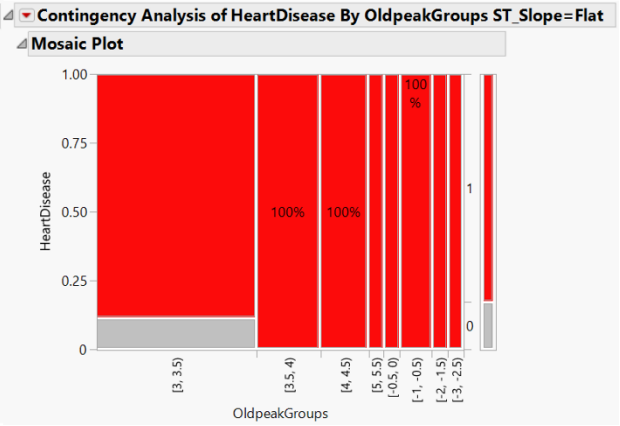


Figure 3.20 – Flat ST_Slope (Zoomed in)

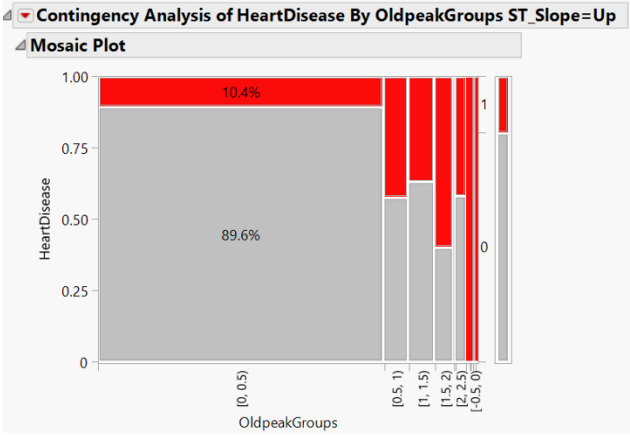


Figure 3.21 – Up ST_Slope

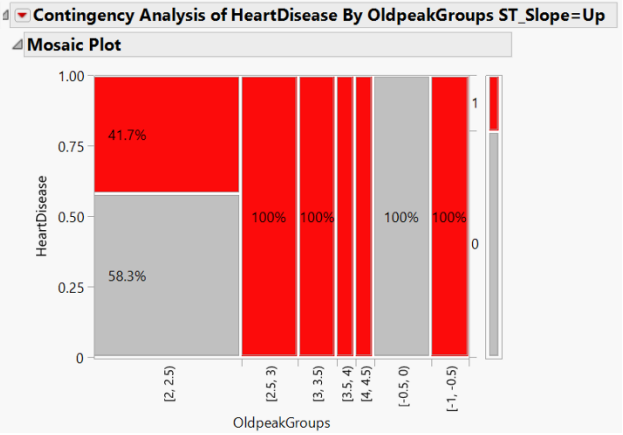


Figure 3.22 – Up ST_Slope (Zoomed in)

Therefore, a combination of ST_Slope and Oldpeak variables can be a very accurate measurement for heart failure rate when the ST_Slope is down or flat, given that the heart failure rate is strongly positively related to high ST depression level in these cases. For diagnosis referring to the ST depression level only, the heart failure rate is high only when the depression level is far away from the healthy

range. It's possible for healthcare workers to develop a matrix of ST_Slope versus Oldpeak to better guide the heart disease diagnosis.

3. ChestPainType and ExerciseAngina

Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen-rich blood (American Heart Association). Our dataset categorizes angina symptoms into typical angina [6](TA), atypical angina [7](ATA), non-typical angina pain (NAP), and asymptomatic angina[8] (ASY). The predictor profilers (Figure 3.3) shows that the people who develop chest pain types of TA and ASY are much more prone to heart disease, as is the case in Figure 3.23 and Figure 3.24.

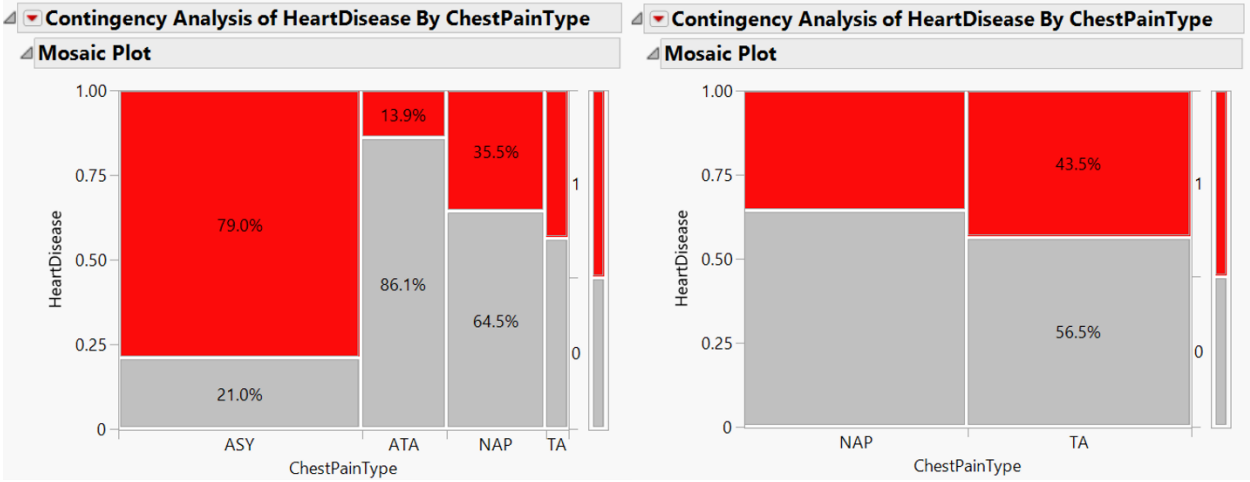
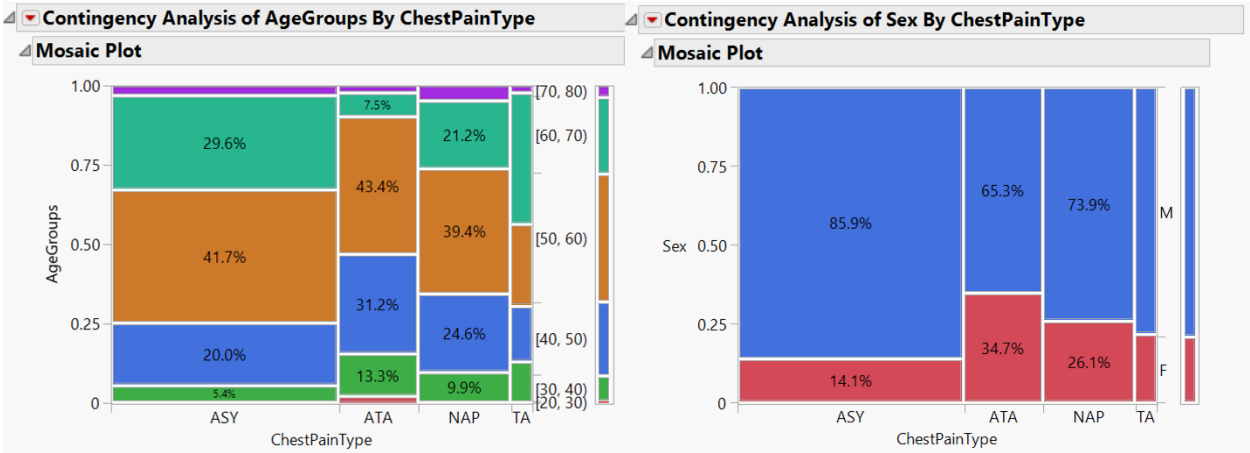


Figure 3.23 **Figure 2.24**

For more details, we stratify the ChestPainType variable into age and sex as follows:



[6] Typical angina is defined as substernal chest pain precipitated by physical exertion or emotional stress and relieved with rest or nitroglycerin.
[7] Atypical chest pain is any chest pain that doesn't meet criteria for a common or obvious diagnosis.
[8] Silent (asymptomatic) myocardial ischemia (SMI) is defined as a transient alteration in myocardial perfusion in the absence of chest pain or the usual anginal equivalents.

Figure 3.25 **Figure 3.26**

For both ASY and TA ChestPainType groups, most of the people are seniors and hence age seems to be a key factor for people developing such symptoms (Figure 3.25). Men are the gender group that shares the highest percentage of all chest pain types (Figure 3.26).

American Heart Association outlines the major risks factors for heart disease, and age becomes one of the critical drivers of heart failure occurrences: the risk increases for men after 45 years of age and for women after 55 years of age (American Heart Association). Thus, older people are much more likely to develop typical angina, a chest pain type directly links to heart disease. However, older people should be more aware of the ASY chest pain type as most of the chest pain types are diagnosed as ASY which leads to the highest heart failure rate according to the predictors profiler (Figure 3.3). In the report, The Danger of Silent Heart Attacks, Harvard Health Publishing reveals a truth that a silent heart attack, known as a silent myocardial infarction (SMI), has account for nearly half of heart attacks (45%) and strike more men than women (Harvard Health Publishing) – as is shown by our data visualization in Figure 3.26 (given 85.9%>M/(M+F) =80%). Therefore, for male seniors, if experiencing "silent" symptoms like extreme chest pain and pressure; stabbing pain in the arm, neck, or jaw; sudden shortness of breath; sweating, and dizziness (Harvard Health Publishing), they should be very careful about heart failures. However, this does not mean a great relief for women, as study has shown that even though it's true that occurrences of silent heart attacks are much more in men, the death rate of the disease is higher in women (Richard N. Fogoros M.D., 2021)

Similar with the approach taken in section ST_Slope and Oldpeak, a combination of ChestPainType and ExerciseAngina is used as a parameter to see if we can develop a quick but efficient diagnosis tool. According to Medical News Today, chest pain (angina) during exercise may be the result of heart attack or less serious ailments such as muscle strains and asthma (Jenna Fletcher, 2018). Referring to effect summary, we can see that ExerciseAngina is not that significant as the other key variables, but useful when combined with ChestPainType as a compository indicator. After rearranging the HeartDisease vs ChestPainType mosaic plot by ExerciseAngina (Figure 3.27 and Figure 3.28), we can see that the heart failure rate increases from 79% to 90.2% when the group of people developing ASY chest pain are known to have experienced exercise-induced angina – experiencing exercise-induced angina even adds to the heart failure risk. However, the heart failure rate increases just slightly (from 43.5% to 50%) for the TA chest pain group, given a "Y" ExerciseAngina is known. Another easy-to-use tool here is an ASY chest pain symptom plus a "Y" exercise-induced angina, for us to predict heart failure.

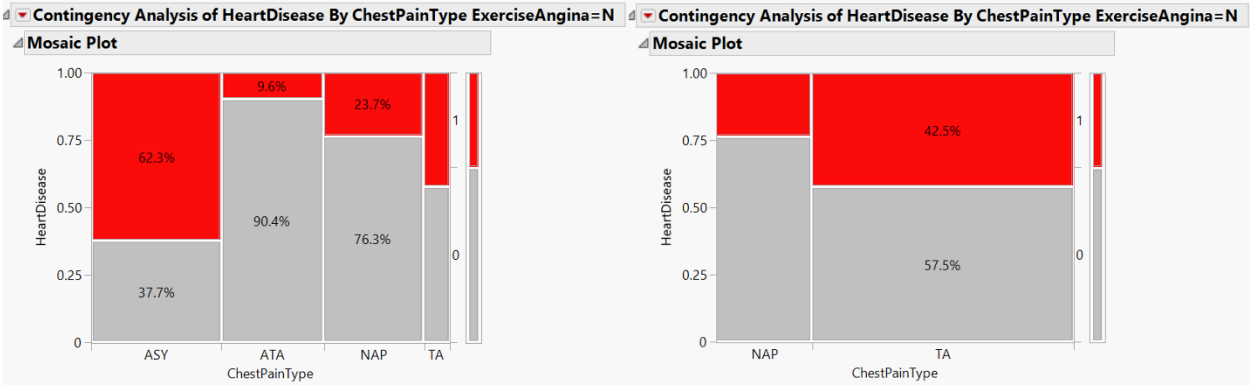


Figure 3.27 – No ExerciseAngina

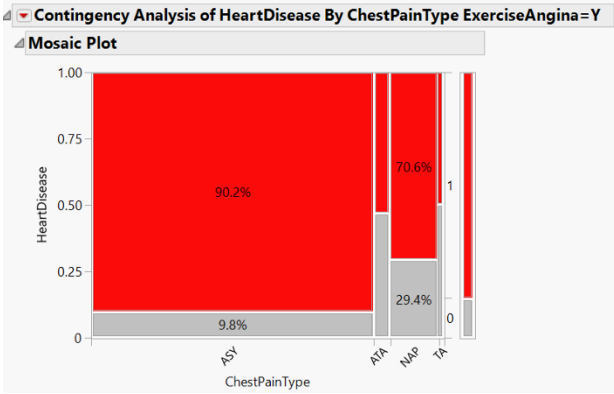


Figure 3.28 – No ExerciseAngina (Zoomed in)

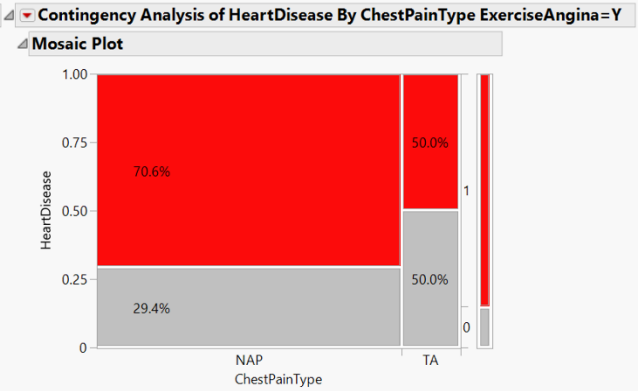


Figure 3.29 – With ExerciseAngina

Figure 3.30 – With ExerciseAngina (Zoomed in)

4. FastingBS

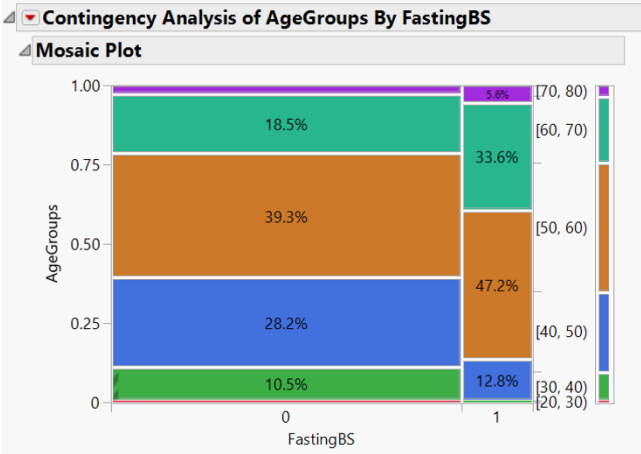
According to a website called Diabetes in Control, the normal range of a person's fasting blood sugar is below 100 mg/dl; a person with pre-diabetes has a fasting blood sugar rate between 110 mg/dl and 125 mg/d; and a person with a level greater than the threshold 125 mg/dl will be diagnosed as having diabetes (Diabetes in Control, 2002). In our dataset, FastingBS is higher if it is greater than 120 mg/dl. Based on the research, Let's just say people with a FastingBS of higher than 120 mg/dl have diabetes. Our model indicates that people with diabetes have much higher heart failure rate (Figure 3.3). CDC lists three conditions of people with diabetes that raise the risk for heart disease (CDC, 2021):

- High blood pressure, which increases the chance of getting heat disease.
- Too much LDL ("bad") cholesterol[9], which can damage your artery walls.
- High triglycerides (a type of fat in your blood) and low HDL ("good") cholesterol [10]or high LDL cholesterol, hardening your arteries.

Figure 3.31 and Figure 3.32 provide a clear picture of the demographic features of people having diabetes: older people and men (given 84.8% > M/(F+M) = 80%) have higher rate of diabetes.

[9] Low-density lipoprotein (LDL) is one of the five major groups of lipoproteins which transport all fat molecules around the body in the extracellular water. LDL has been associated with the progression of atherosclerosis and blockage of the artery lumen, because it can carry cholesterol into smaller vessels.

[10] High-density lipoprotein (HDL) is one of the five major groups of lipoproteins. Unlike the larger lipoprotein particles, which deliver fat molecules to cells, HDL particles remove fat molecules from cells. Increasing concentrations of HDL particles are associated with decreasing accumulation of atherosclerosis within the walls of arteries, reducing the risk of sudden plaque ruptures, cardiovascular disease, stroke and other vascular diseases.



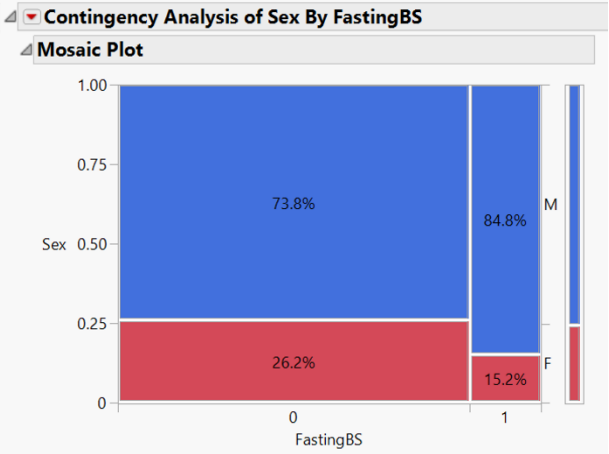
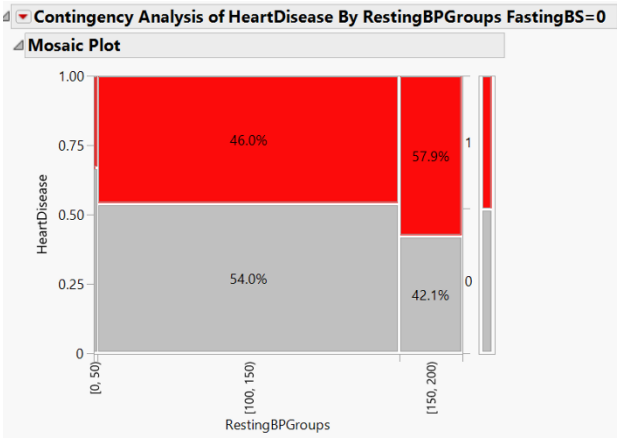


Figure 3.31 **Figure 3.32**

For a better view of how blood pressure (RestingBP) interacts with diabetes occurrences, we divide RestingBP into four intervals and develop yet another mosaic plot, grouped by FastingBS (Figure 3.33 and Figure 3.34). Per CDC guideline, blood pressure is normal when it is less than 80 mmHg in a diastolic blood pressure, a measurement of the pressure in your arteries when your heart rests between beats (Division for Heart Disease and Stroke Prevention). In our case, therefore, a RestingBP of greater than 100 mmHg is considered high. Overall, people with both diabetes and high blood pressure have much higher heart failure rate than those with high pressure but have no diabetes, meaning having both high blood pressure and diabetes can greatly increase your risk for heart disease (CDC).



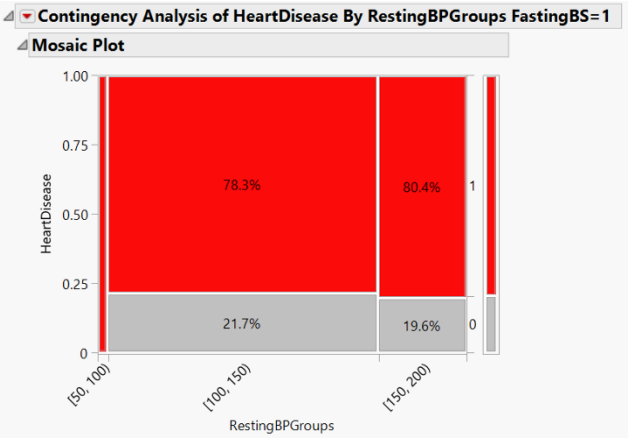


Figure 3.33 **Figure 3.34**

When comparing RestingBP with Cholesteroal, our team cannot tell the bad and good ones in the Cholesterol variable. Figure 3.35 seems to indicate that a high blood pressure will lead to low cholesterol level. When excluding the anomaly of the cholesterol group [0, 50), the story is a totally different one: people with high blood pressure tend to have high cholesterol level (Figure 3.36). The reason why we exclude the anomaly will be explained in detail in the coming section. We are not sure about the type of high cholesterol level of people having diabetes, but our best guess is that it's the high LDL ("bad") cholesterol level contributes to the heart failures of these people.

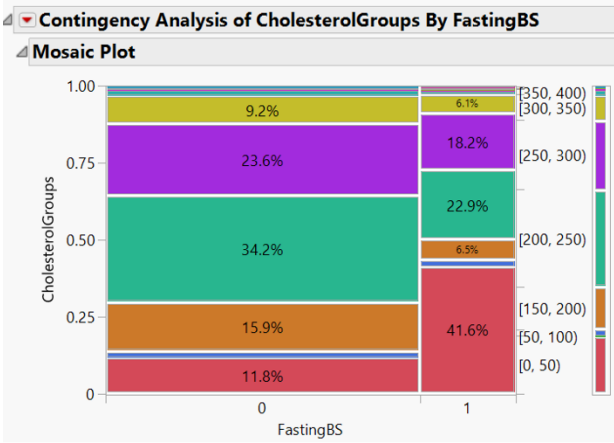


Figure 3.35

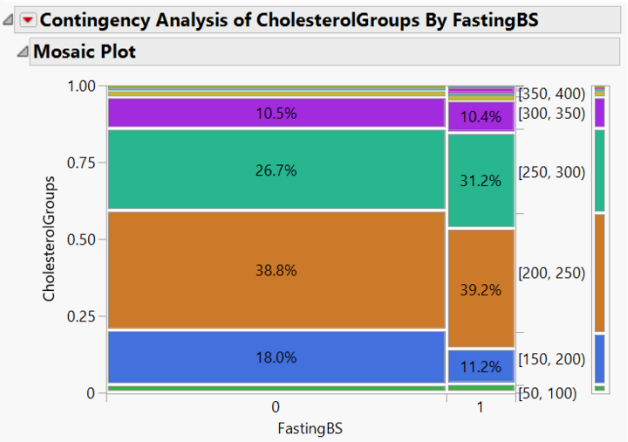


Figure 3.36

5. Cholesterol

Talking about staying healthy, people are always cautious about their cholesterol level. A doctor would always advise their patients to take a healthy diet, do regular exercise, and even sometimes receive medication treatment to reduce cholesterol level. Though your body needs cholesterol to build healthy cells, a high cholesterol level will harm your health, especially by increasing the chances of getting a heart attack or stroke. A report by Mayo Clinic shows that, with high cholesterol, human body develops fatty deposits in blood vessels and these deposits grow, making it difficult for enough blood to flow through arteries (Mayo Clinic, 2021). Typically, a high blood cholesterol level increases the chance of getting heart disease (Francisco Lopez-Jimenez, 2020), but studies also find that total cholesterol, a measure of cholesterol level, is a poor predictor of heart disease and heart attack (Everyday Health, 2012): total cholesterol levels of people who suffer from heart disease and those do not are almost the same. In light of the medical research, we investigate how cholesterol level will lead to the occurrences of heart failure by considering the participation of the key variables like ChestPainType, ST_Slope, and FastingBS and the associative variables age and sex.

The heart disease by cholesterol curve presents a clear relationship between heart disease and cholesterol level within our sample pool (Figure 3.37). The lowest point of the curve is where cholesterol level is around 200 mg/dl, meaning maintaining a cholesterol level of 200 mg/dl may be the optimal level to reduce heart disease. According to Medical News Today, a health newsletter provider, suggests that total cholesterol level of less than 200 mg/dl be desirable for adults (Jenna Fletcher, 2020). From this perspective, it looks reasonable for people getting higher chance of heart failure as the cholesterol level increases from 200 mg/dl. However, the study will make our data for those with cholesterol level of less than 200 mg/dl a soft spot where, abnormally, people with very low cholesterol level also have worryingly high heart failure rate. According to Francisco Lopez-Jimenez, M.D., the association between a very low total cholesterol level and some health problems only exists in very rare cases (Francisco Lopez-Jimenez, 2020).

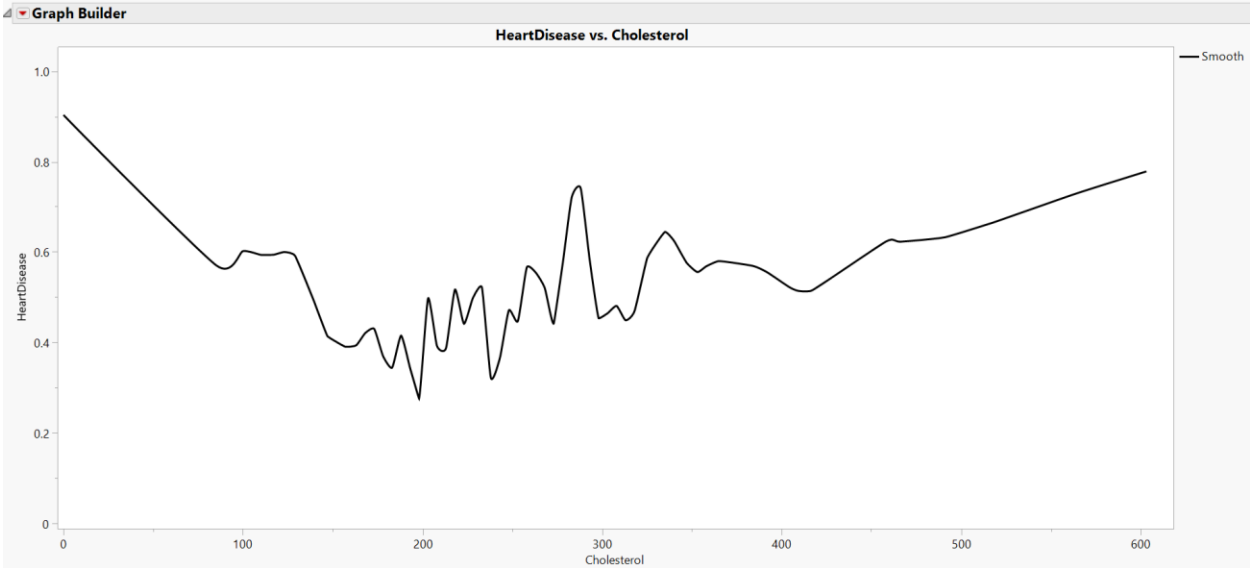


Figure 3.37

Using a heat map, we have a full picture of the distribution of frequencies of each cholesterol level of different gender groups. Figure 3.38 shows that the density is high for the groups with cholesterol level from 150 mg/dl to 300 mg/dl and between 40 to 70 years old (for male, the density is so high that a red spot appears at the center of the graph). For both male and female, we can see people of different ages have extremely low cholesterol level (the blocks at the left end).

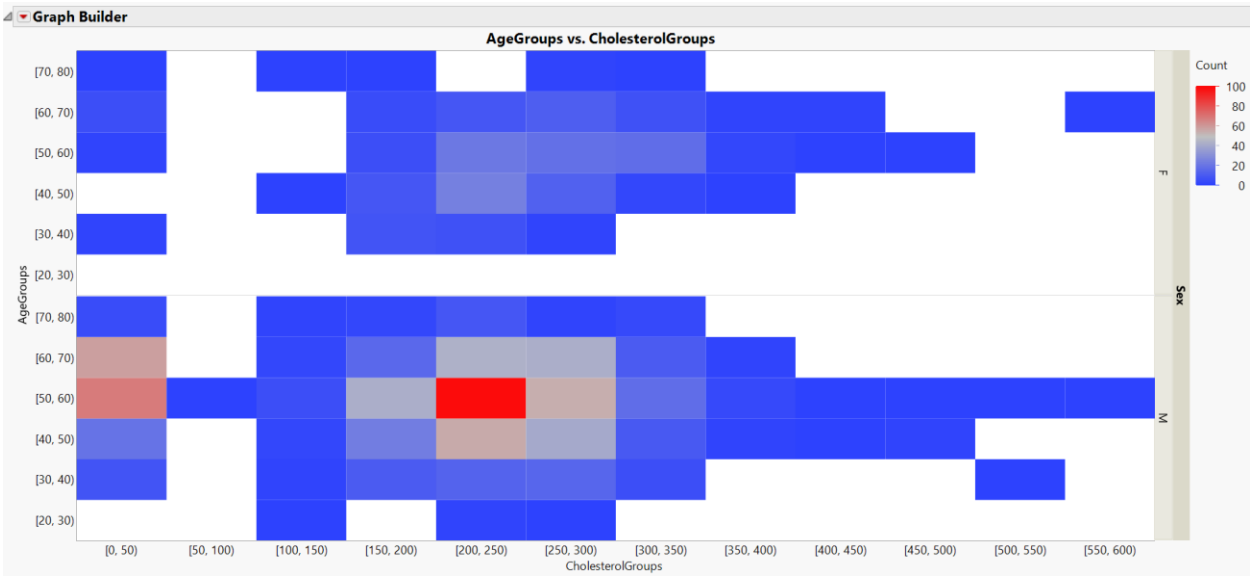


Figure 3.38

Turning the page back to the distribution of cholesterol levels (Figure 3.39), we can see we have a "slump" at the left side of the distribution where a great number of people have extremely low cholesterol level (from 0 mg/dl to 50 mg/dl). The existence of such group in our sample pool has attracted our team and we are very curious about the demographic features and health status of such group of people. We divide our data into thirteen sub-groups each of which has equal cholesterol level interval of 50 mg/dl, i.e., [0, 50), [50, 100) … [600, 650). First, we investigate the demographic features of these groups of people.

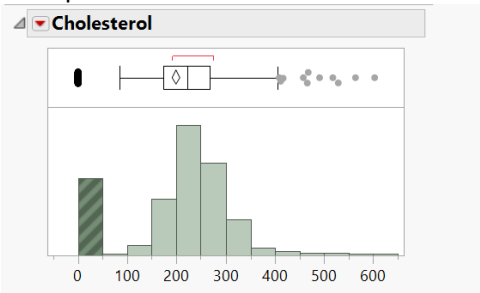


Figure 3.39

A mosaic plot of CholesterolGroups by AgeGroups provides a visual of age group components for each cholesterol group (Figure 3.40). For people with cholesterol level of [0, 50), they are mostly seniors of more than 50 years old (with colors of orange, green, and purple). Using x axis adjusting tool, we zoom in to see the age groups of the people with high cholesterol level (Figure 3.41). We can see that all the people with extremely high level of cholesterol are people of more than 50 years old. Old ages of the people with extremely low and extremely high cholesterol level may somehow link to the high heart failure rates of these people at the two ends of the cholesterol axis (Figure 3.38). Old age is already a big problem for people's health. A newsletter published by Medical New Today has shown that cholesterol levels tend to increase with age (Jenna Fletcher, 2020). However, it can be arbitrary to attribute the high cholesterol level to just being older. Factors that are of your control like inactivity, obesity, and an unhealthy diet, or even those out of your control like genetic makeup can lead to harmful cholesterol level (Mayo Clinic, 2021). We need to investigate further to get more intelligence.

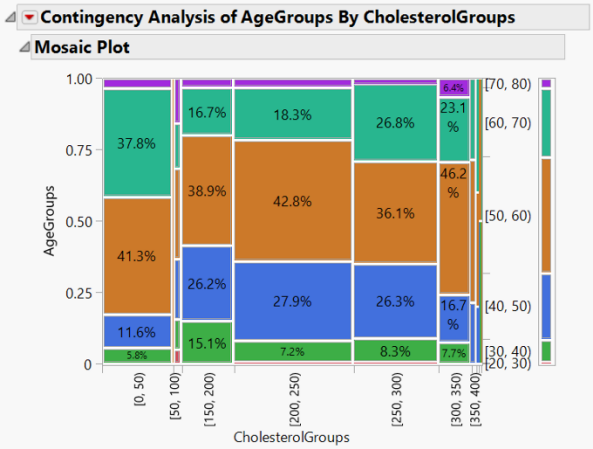


Figure 3.40

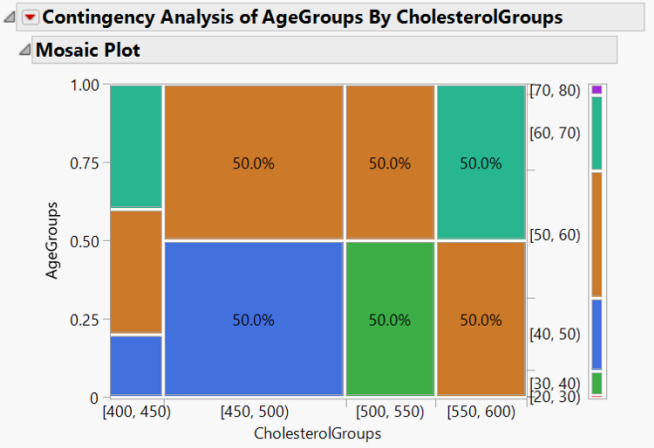
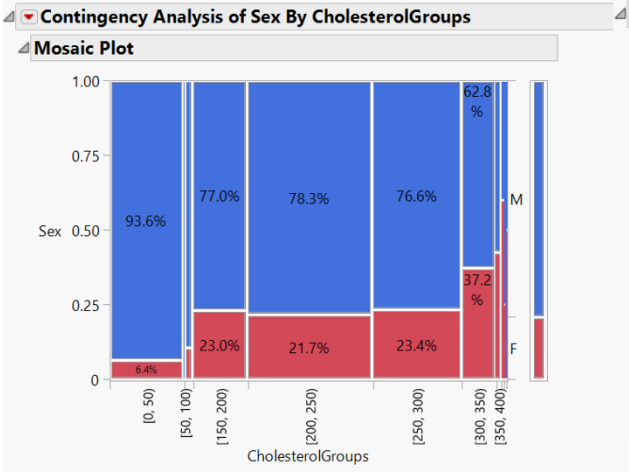


Figure 3.41

Second, we explore the gender groups for each cholesterol group. In Figure 3.42, we can see that the proportion of women seems to increase as the cholesterol level grows from 0 to 450 mg/dl. When zooming in for a clearer picture, male seems to have the largest share of the high cholesterol group of [550, 600) and [600, 650). High proportion of males at both ends of cholesterol may explain why we see high heart failure rates at both sides in the curve of Figure 3.37 since, as we conclude in section Age and Sex, male has higher chance of getting heart failure on average. However, we also find the cholesterol

level trend among women interesting: the proportion of women tend to be high on the right side of the cholesterol axis (given the proportion of women is greater than F/(M+F) = 20% when cholesterol level > 150 mg/dl). Medical News Today reports that women aren't immune to high cholesterol: a woman's cholesterol often increases when she goes through menopause (starting between ages of 40 and 58 years old) (Jenna Fletcher, Yvette Brazier, 2021). The surge in cholesterol level of women of this age or older helps explain why the heart failure rate increases so rapidly when women are getting older (Figure 3.10).



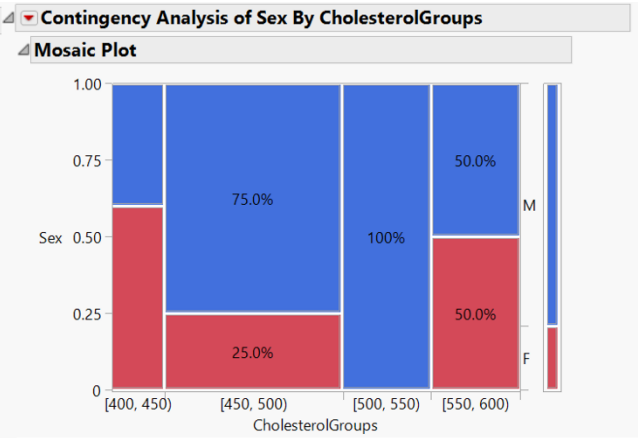
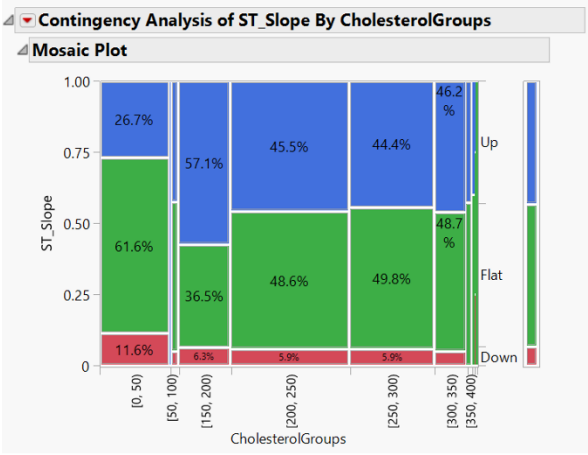


Figure 3.42 **Figure 3.43**

Third, we come to the first key variable, ST_Slope. As the mosaic plot shows (Figure 3.44 and Figure 3.45), most of the ST_Slope types for people with extremely low and extremely high cholesterol level are flat, meaning they are exposed to higher danger of heart disease (as is explained in the section ST_Slope and Olpeak.



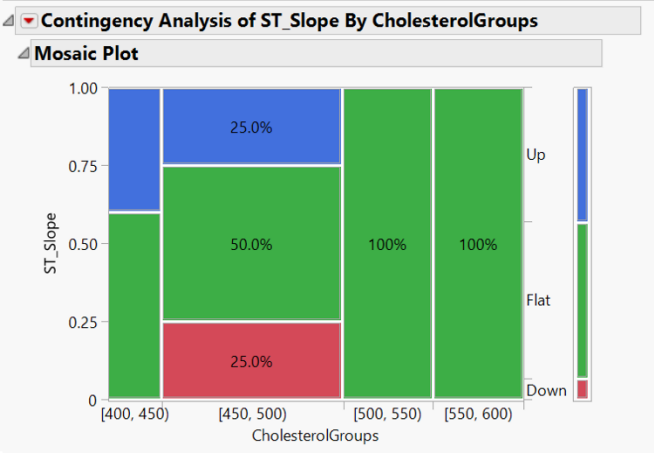
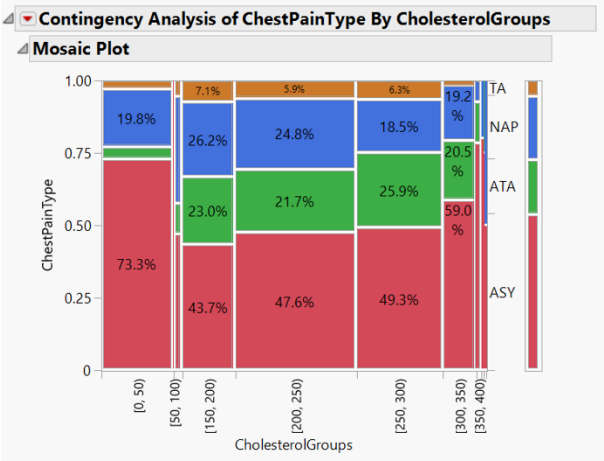


Figure 3.44 **Figure 3.45**

Forth, we explore the relationship between chest pain type and cholesterol level. We can see that the proportion of asymptomatic chest pain type remains high for both low and high cholesterol level group. What we can see in the plot is aligned with our understanding that people with asymptomatic chest pain type will also in high danger of getting heart disease (Section ChestPainType).



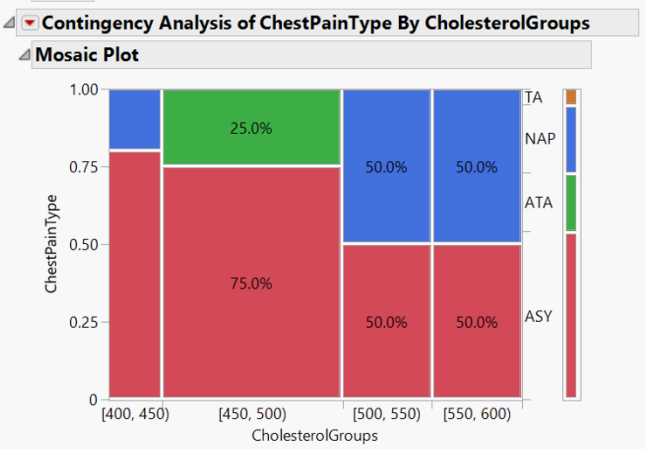


Figure 4.46 **Figure 4.47**

The last key variable with which we want to compare cholesterol level is FastingBS. Again, high FastingBS rates are polarized at both sides of the cholesterol axis (Figure 3.37), meaning these people are have higher diabetes rates with either low cholesterol level or high cholesterol level. Hence, they are more prone to heart disease, as is explained in section FastingBS.

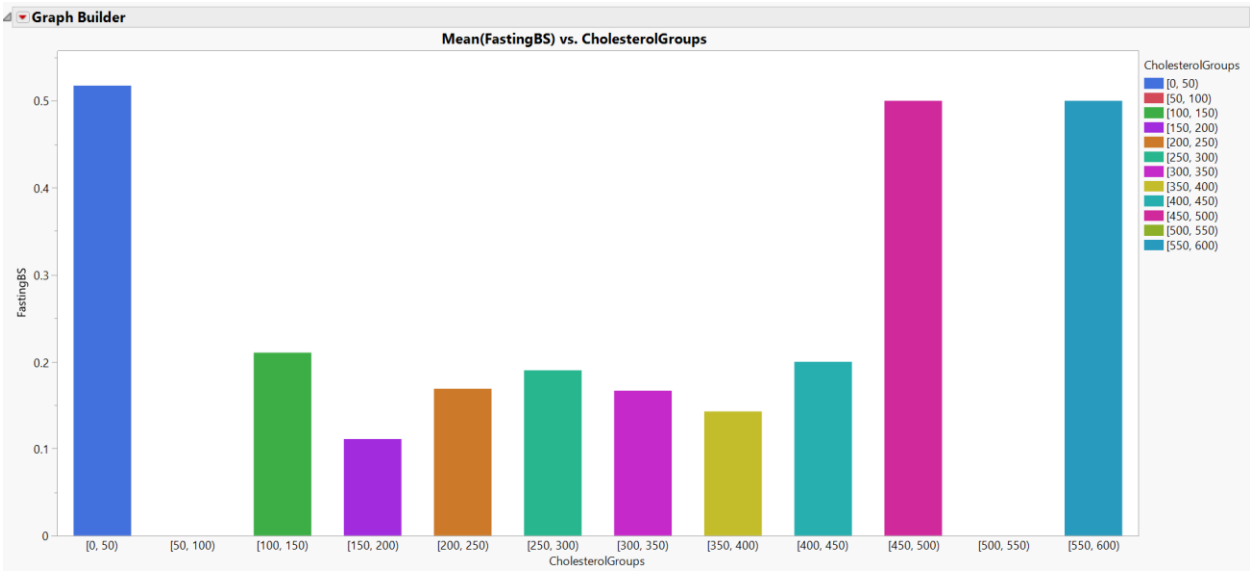


Figure 4.48

By the end of this section, our data analysis has shown that high cholesterol level tends to increase the chance of heart failure but other variables like age, gender, and FastingBS will also have impacts on the probability of getting heart disease. In this section, we investigate the rare extremely low cholesterol level that appears among the group exposing to high heart failure rate. We find that most of such group of people are males, seniors, and many of them have develop symptoms of flat ST slope and asymptomatic chest pain, and have unhealthy blood sugar rate. However, why does the rare case happen? Maybe this is one of the symptoms developed by diabetes which may lead to very low HDL ("good") cholesterol, based on what we've learned from section FastingBS. Anyway, this is a problem of medicine and not possible for us to investigate further just with JMP. Our team realizes that total

cholesterol level is not a good predictor of heart failure, but it is an easy measure for people to use and an interesting variable to explore. For another key variable ExerciseAngina, we find no obvious relationship to explain the anomaly in cholesterol levels.

6. A Better Model?

Noticing "trimming" some key variables with the connected variables may improve the accuracy of the classification, our team is worried about the collinearity of variables within the logistic model. For logistic regression, a method called maximum likelihood is used to find the estimates that maximize the chance of obtaining the data that we are interested in. Nevertheless, collinearity, a strong correlation among the predictors, can lead to computational difficulties and thus errors in classifying variables. Even though we examine the correlations among continuous variables in Figure 1.3, in which we conclude that there are no obvious correlations among continuous variables, we should also be alerted to the collinearity between the continuous variables and the categorical variables, and among categorical variables (when they are assigned numerous values, for example: Male→1, Female→ 0). In part two, we have proven that combining the ST_Slope (categorical) and the Oldpeak (continuous) variables can improve classification accuracy when only looking at the two variables. This somehow implies that on one hand, there's better classification model available, and on the other, the logistic model stands out just very luckily (for example, the classification process can be unstable when we don't have sufficient records to make it stable). Given that we use the fundamental method called "trimming" in part two, a model of the tree family may do a good job in some cases, for example, when the data size is large. In this way, the model can learn from the complicated connections among the elements of both the continuous and the categorical variables. However, this sometimes can be overfitted, like what we have mentioned, as it needs many records to proceed successfully if the dataset is complex. In Figure 2.1, we can see that the tree family outperforms in the training dataset in which more data is available whilst the logistic regression model performance is just average. Anyway, a logistic one is what data chooses for us and it is a much more efficient model to use when the dataset is not large.

Additionally, we also notice the limitations of our dataset. First, the records for women are in such a small scale, that those for old women become outliers for heart failure of women. This can be improved by involving more women to the research. Second, the variable Cholesterol is not a good predictor. Perhaps it is a good idea to divide it into "good" or "bad" cholesterol. However, this will make it even more difficult to collect data, as defining "good" or "bad" cholesterol may not be recorded in some research institutes.

Our Suggestions to Control Heart Disease

Based on what we've learned from the data, we make the following suggestions for ordinary people:

- Check your blood pressure regularly. Our model indicates that the higher your blood pressure is, the greater risk of heart failure you are in (Figure 3.3). High blood pressure can be so silent so it's important to know before it's too late to control. A healthy blood pressure is 80 mmHg in terms of diastolic blood pressure.
- Be aware of diabetes. The correlation of diabetes and heart failure is so significant. Taking to your doctor and inquire if you are in the danger of having diabetes is a great idea to take a step ahead.

Regularly test your blood sugar rate to make sure it won't be in dangerous trajectory. A healthy blood sugar rate is below 100 mg/dl in terms of fasting blood sugar rate.

- Do not smoke. According to Hopkins Medicine, cigarette smokers are two to four times more likely to get heart disease than nonsmokers (Hopkins Medicine). A special reminder for women is that women of 35 years old or older are much more prone to heart failure when they are taking birth control pills and smoke (Hopkins Medicine). Smoking increases your chances of heart failure in the ways of 1) causing blood pressure rise, 2) causing increase in heart rate, 3) reducing the amount of oxygen to your tissues, and so on and so forth.
- Check your cholesterol level. Based on our data analysis, a high cholesterol level usually links to a high heart failure rate. However, please ask your doctor for some advice if you have diabetes.
- Eat healthy food and keep a healthy weight. According to CDC, having overweight or obesity raises your risk of heart disease (CDC). Men are more likely to live an unhealthy life (Jen Tan, 2017).
- Limit alcohol. Alcohol abuse will weaken your heart, affecting its ability to pump blood and leading to heart failure. According to Jen Tan, 37% of men in the UK regularly exceed the government recommendations for alcohol intake, compared to 28% of women (Jen Tan, 2017).
- Relax and don't be stressed. A webpage on Hopkins Medicine shows that stress can cause the narrowing of the small arteries and a temporary decrease in blood flow to the heart. This may lead to a kind of heart failure called broken heart syndrome (Ilan Wittstein, M.D.). Even though the gap is closing, men are more stressed in the labor market. This helps explain why men have higher heart failure rate, to some extent (Jen Tan, 2017).

Also, we suggest some quick tests for doctors who are diagnosing their patients:

- When predicting heart failure rate with an ECG criterion, combine ST segment Slope with ST depression level.
- When inquiring the patients of any symptoms of angina, ask if they have experienced exercised-induced angina to improve the efficiency in silent heart disease diagnosis.

Conclusion

By the end of the project, we select a logistic model to predict heart failure rate by incorporating various variables from ST_Slope to Sex. First, we evaluate the performance of our model, and we can see a good classification accuracy of the model. Second, we interpret the model in-depth to understand how the key variables like ST_Slope, FastingBS interact with heart failure rate. Some of the variables like ST_Slope and Oldpeak may work together to improve accuracy. For cholesterol, the anomaly appears on the predition profiler is the result of the impacts from other variables that are more important. Third, we justify our model in a critical way, and we conclude that the model may not be the best when dealing with a great amount of collected data but an efficient one if there are not many coming records. Finally, our team can give suggestions regarding personal health and heart disease diagnosis improvement, based on what we've learned from the model.

References

American College of Cardiology (2017). Taking the Risks to Heart: Misdiagnosis of Heart Disease. https://www.acc.org/membership/join-us/benefits/additional-member-only-benefits/acc-and-the-doctors-company/the-doctors-company-updates/2017/02/20/12/55/taking-the-risks-to-heart-misdiagnosis-of-heart-disease

American Heart Association. Angina (Chest Pain). https://www.heart.org/en/health-topics/heart-attack/angina-chest-pain

Centers for Disease Control and Prevention (2021). Diabetes and Your Heart. https://www.cdc.gov/diabetes/library/features/diabetes-and-heart.html

Diabetes in Control (2002). Fasting Blood Sugar Above 90 Puts You At Risk of Heart Disease. https://www.diabetesincontrol.com/fasting-blood-sugar-above-90-puts-you-at-risk-of-heart-disease/

Division for Heart Disease and Stroke Prevention (2020). Heart Failure. National Center for Chronic Disease Prevention and Health Promotion. https://www.cdc.gov/heartdisease/heart_failure.htm

Division for Heart Disease and Stroke Prevention (2021). High Blood Pressure Symptoms and Causes. Centers for Disease Control and Prevention. https://www.cdc.gov/bloodpressure/about.htm

Dr. Araz Rawshani (2018). The ST Segment: Physiology, Normal Appearance, ST Depression & ST Elevation. ECG & ECHO Learning. https://ecgwaves.com/st-segment-normal-abnormal-depression-elevation-causes/

Dr. Jen Tan (2017). Are men more prone to heart disease? A.Vogel. https://www.avogel.co.uk/health/mens-health/are-men-more-prone-to-heart-disease/

Ed Burns & Robert Buttner (2021). The ST Segment. Life in the Fastlane. https://litfl.com/st-segment-ecg-library/

Everday Health (2012). The Single Best Predictor of a Heart Attack. https://www.everydayhealth.com/heart-health-pictures/the-single-best-predictor-of-a-heart-attack.aspx

Francisco Lopez-Jimenez, M.D. (2020). Cholesterol level: Can it be too low? Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/expert-answers/cholesterol-level/faq-20057952

Harvard Health Publishing (2019). The Danger of 'Silent' Heart Attacks. Harvard Health. www.health.harvard.edu/heart-health/the-danger-of-silent-heart-attacks.

Hopkins Medicine. Smoking and Cardiovascular Disease. https://www.hopkinsmedicine.org/health/conditions-and-diseases/smoking-and-cardiovascular-disease

Ilan Shor Wittstein, M.D. Broken Heart Syndrome. Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/conditions-and-diseases/broken-heart-syndrome

Jenna Fletcher (2018). What should my cholesterol level be at my age? Medical News Today. https://www.medicalnewstoday.com/articles/315900

Jenna Fletcher (2020). What should my cholesterol level be at my age? Medical News Today. https://www.medicalnewstoday.com/articles/315900#_noHeaderPrefixedContent

Mayo Clinic (2021). High cholesterol. https://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/symptoms-causes/syc-20350800

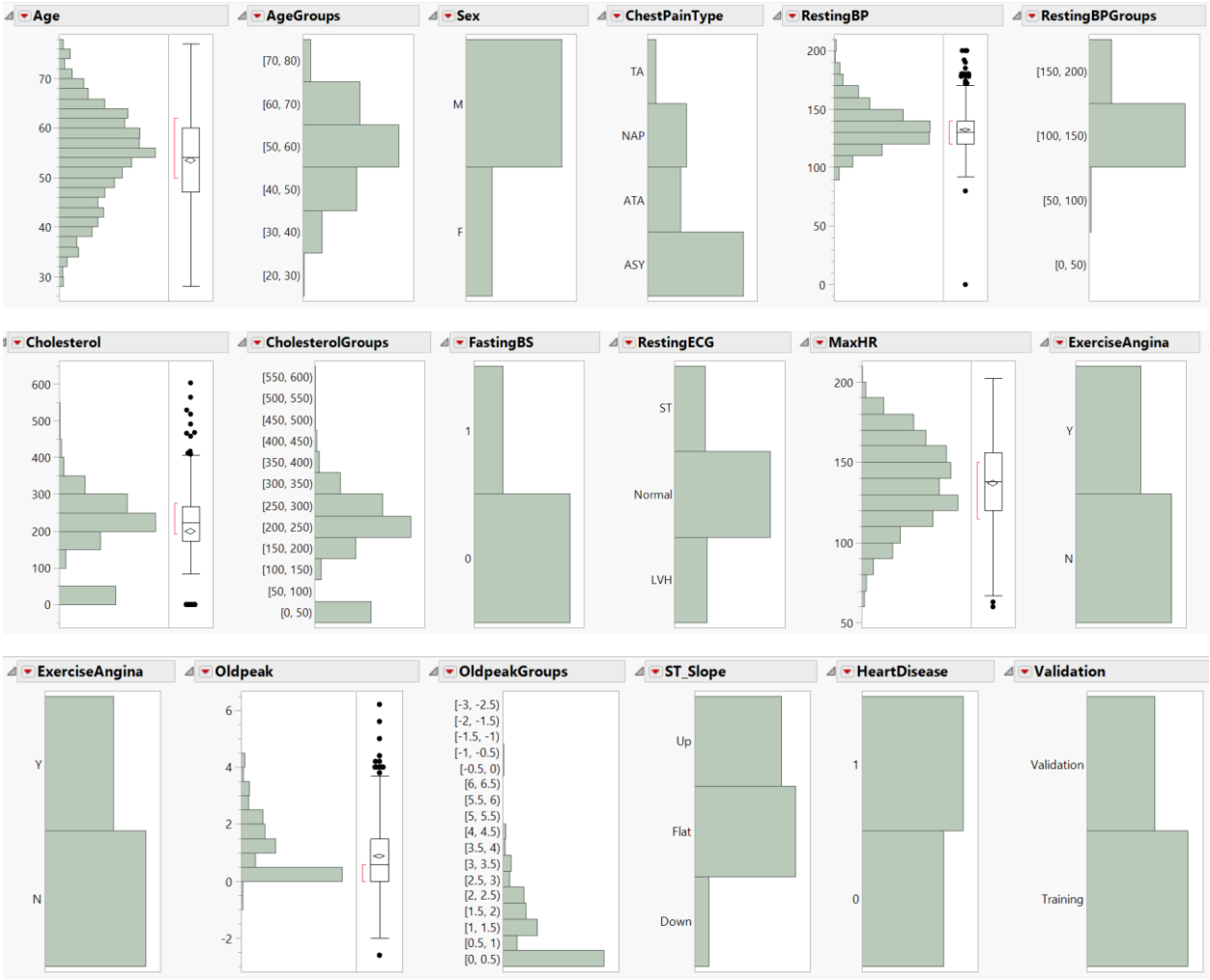
Richard N. Fogoros (2021). What Is a Silent Heart Attack? Verywell Health. https://www.verywellhealth.com/silent-heart-attacks-1746018

Robert et al. (2004). The ST Segment/heart Rate Slope as a Predictor of Coronary Artery Disease. Elsevier. https://www.sciencedirect.com/science/article/abs/pii/0002870386902656

Yvette Brazier (2021). Menopause signs and symptoms, and treatments if you are experiencing them. Medical News Today. https://www.medicalnewstoday.com/articles/155651

Appendix

I. All Variables Distribution



II. Interaction Profiles