A Logistic Regression Model for Heart Disease Dataset

Abolfazl Joukar

1 Dataset Description

Data was drawn from the UCI machine learning repository. The data collection took place at 4 different medical clinics, and the Cleveland clinic database is used here. There are 303 patients in the data set and 14 measurements provided for each patient: 5 measurements of heart activity, 8 demographics/risk factors, and the response variable (heart disease). The objectives of this project are to build a model that accurately predicts if a patient has heart disease or not based on their demographics and to determine the most significant risk factors for heart disease. The 14 variables used for analysis are:

- Age
- Sex
  value 1: Male
  value 0 : Female
- Chest Pain Type (cp)
  value 1: typical angina
  value 2: atypical angina
  value 3: non-angina pain
  value 4: asymptomatic
- Resting blood pressure (trestbps)
- Serum cholestoral (chol)
- Fasting Blood Sugar (fbs)
  value 0: fbs $\leq$ 120 mg/d
  value 1: fbs $>$ 120 mg/dl
- Resting electrocardiographic results (restecg)
  value 0: normal
  value 1: 1 having ST-T wave abnormality
  value 2: showing probable or definite left ventricular hypertrophy
- Maximum heart rate achieved (thalach)
- Exercise induced angina (exang)
  value 1: yes
  value 0: no

- ST depression induced by exercise relative to rest (oldpeak)
- The slope of the peak exercise ST segment (slope)
  value 1: unsloping
  value 2: flat
  value 3: downsloping
- Number of major vessels colored by floursopy (ca)
  value 0 - 3
- thal
  value 3: normal
  value 6: fixed defect
  value 7: reversible defect
- Diagnosis of heart disease (heart disease)
  value 0: heart disease absence
  value 1: heart disease present

2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) helps us in analyzing our data in detail and discover trends, patterns and other crucial information through statistical summaries and graphical representations. This understanding will help us in optimal training of the Logistic Regression model for better prediction accuracy. Firstly we will check if our dataset consists of any null values in any columns. There are 6 missing values in the values in the varaibles "ca" and "thal". So, 297 observations are used for analysis. Now we will visualize our heart disease variable using a barplot. We can observe from the Figure 1(b) that the number of instances of patients with heart disease is slightly less than those without heart disease.

In Figure 1(a), the histograms for age faceted on the presence and absence of heart disease have different distribution shapes, suggesting that age does have a relationship with heart disease. The distribution of the presence of heart disease is left skewed while the distribution of the absence of heart disease appears more normally distributed. These graphics suggests that there are more older people with heart disease than younger people with heart disease. The bar plot in Figure 1(c), shows that the proportion of males with heart disease is much higher than females with heart disease showing a relation between gender and heart disease. In the following, we use a correlation matrix to assess the relationships between the features. This matrix is a powerful tool for summarizing a large dataset, offering insights into patterns and potential dependencies among variables. The Figure 2 shows that variables "thal" and "ca" have a high correlation with the heart disease variable, so these must be having strong relation with this variable . On the other hand, the variable fbs has 0 correlation indicating it doesnt have any relationship with the heart disease variable. In addition, the variable "thalach" has a negative correlation with the heart disease variable and indicating it does have an inverse relationship with the heart disease variable. The table 1 is the summary table of numeric variables in the dataset.We deduce for numeric variables from Table 1 that:

- For the "age" variable we can see that mean age is 54.54 and the median age is 56.0 which means that the age distribution is skewed little bit to the left.

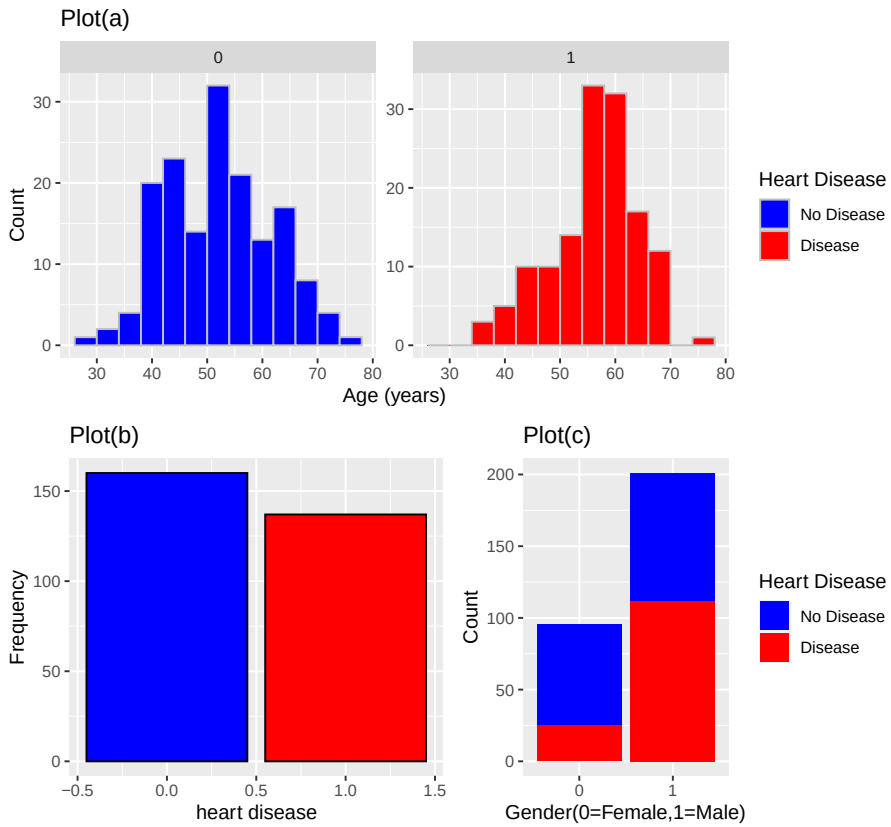


Figure 1: Prevelance of Heart Disease Across Age(Plot a), Distribution of heart disease variable(Plot b) and Distribution of Gender by Heart Disease Status(Plot c).

- The "trestbps" variable has mean resting blood pressure of 131.7 and median resting blood pressure of 130.0. Since the median is little less, the distribution is skewed to right by very small amount.
- For the variable "chol" mean serum cholesterol level is 247.4 and median serum cholesterol level is 243.0. The median cholesterol level is less than mean which means that the distribution of cholesterol level is skewed to right.
- The variable "thalach" has mean maxmium heart rate achieved of 149.6 and median heart rate achieve of 153.0. The median is higher than mean in this case. Hence, the distribution of maximum heart rate achieved is skewed to the left.
- The variable "oldpeak" has mean ST depression induced by exercise of 1.056 and median of 0.8. Since the median is less than distribution is skewed to right.

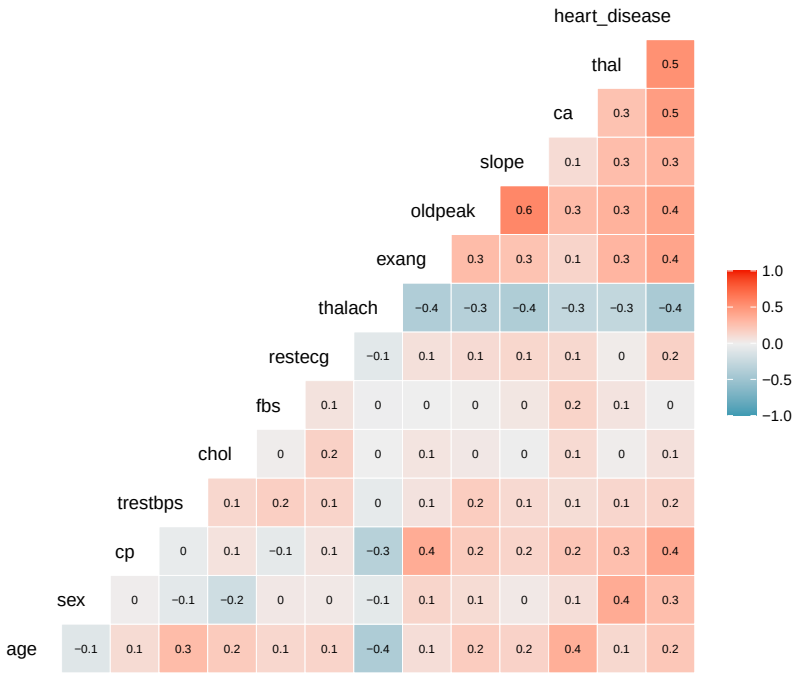


Figure 2: Correlation matrix visualization.

3 Model Setup

When dealing with categorical data from a target variable, logistic regression is often used to model the data. There are different types of logistic regression that can be used, such as simple logistic regression and multiple logistic regression. In the case of the data set that is used throughout this project, multiple logistic regression is used, as the outcome variable represents binary responses (indication of heart disease in a patient). The full model is given as follow:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1\text{age} + \beta_2\text{sex} + \beta_3\text{cp} + \beta_4\text{trestbps} + \beta_5\text{chol} + \beta_6\text{fbs} + \beta_7\text{restecg} + \beta_8\text{thalach} + \beta_9\text{exang} + \beta_{10}\text{oldpeak} + \beta_{11}\text{slope} + \beta_{12}\text{ca} + \beta_{13}\text{thal}.$$

The results of full model are given in Table 2. From Table 2, the variables found to be significant at the 0.05 level by the logistic regression model are: sex, Chest pain type (cp), Resting blood pressure (trestbps), Maximum heart rate achieved (thalach), Exercise induced angina (exang), Number of major vessels colored by floursopy (ca) and thal.

Table 1: Summary of Dataset.

Variables	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
age	29.00	48.00	56.00	54.54	61.00	77.00
sex	0.0000	0.0000	1.0000	0.6768	1.0000	1.0000
cp	1.000	3.000	3.000	3.158	4.000	4.000
trestbps	94.0	120.0	130.0	131.7	140.0	200.0
chol	126.0	211.0	243.0	247.4	276.0	564.0
fbs	0.0000	0.0000	0.0000	0.1448	0.0000	1.0000
restecg	0.0000	0.0000	1.0000	0.9966	2.0000	2.0000
thalach	71.0	133.0	153.0	149.6	166.0	202.0
exang	0.0000	0.0000	0.0000	0.3266	1.0000	1.0000
oldpeak	0.000	0.000	0.800	1.056	1.600	6.200
slope	1.000	1.000	2.000	1.603	2.000	3.000
ca	0.0000	0.0000	0.0000	0.6768	1.0000	3.0000
thal	3.000	3.000	3.000	4.731	7.000	7.000
heart disease	0.0000	0.0000	0.0000	0.4613	1.0000	1.0000

Table 2: Summary of the results for the logistic regression model(full model) for Heart Disease dataset.

Coefficients	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	−7.372042	2.879476	−2.560	0.01046
age	−0.014164	0.023970	−0.591	0.55459
sex	1.312073	0.488474	2.686	0.00723
cp	0.575898	0.191197	3.012	0.00259
trestbps	0.024044	0.010730	2.241	0.02504
chol	0.004995	0.003774	1.324	0.18561
fbs	−1.021918	0.555330	−1.840	0.06574
restecg	0.245153	0.185005	1.325	0.18513
thalach	−0.020665	0.010225	−2.021	0.04327
exang	0.926104	0.413343	2.241	0.02506
oldpeak	0.247386	0.211832	1.168	0.24287
slope	0.570009	0.363085	1.570	0.11644
ca	1.267719	0.265384	4.777	0.000002
thal	0.343936	0.100361	3.427	0.00061

3.1 Outliers & Unusual Points

Regression models are very sensitive to be aected by outliers. In a logistic model, observations whose values diverge from their expected intervals and result in a model peculiarity may be outliers. Influential values are extreme individual data points that can alter the quality of the logistic regression model. A failure to detect influential cases can have severe distortion on the validity of the inferences drawn from such modeling. We will use standardized residuals to detect outliers, it is suggested the range of a standardized residual to be $(-3, +3)$, if a standardized residual is out of this range, it is an outlier. In the table 3, we show 5 observations (1.7% of the dataset) were removed.
The models performance is re-evaluated using the new dataset after removing these observations. Examining the Akaike information criterion(AIC) indicates an improvement in the models performance for without Outliers (AIC=198.44) compared to the model with Outliers (AIC=232.69).

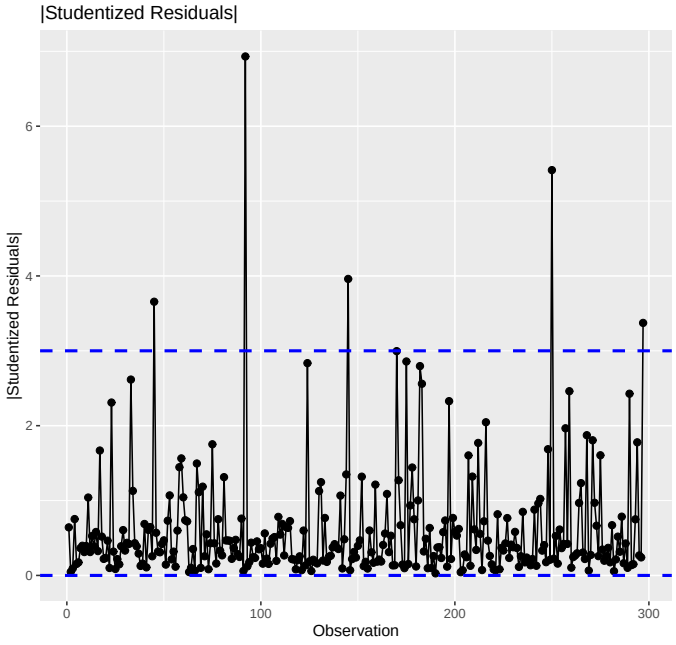


Figure 3: Outliers and Unusual Points based on Studentized Residuals.

Table 3: Outliers and Unusual Points based on Studentized Residuals.

Id	| Studentized Residuals |
45	3.655321
92	6.931208
145	3.959994
250	5.413198
297	3.372509

3.2 Multicollinearity

Multicollinearity is a statistical problem occurs when explanatories are highly correlated to each other. The existence of multicollinearity inflates the variances of the estimated parameter, and misleads the relationships between predictors and response variable. To assess multicollinearity, we use the Variance Inflation Factor (VIF). For each independent variable, we calculate the VIF, and if the VIF value exceeds 10, multicollinearity is present, and we need to address this issue. Therefore, based on the VIF values in Table 4, it is concluded that there is no multicollinearity for predictor variables.

Table 4: Multicollinearity for predictor variables.

Coefficients	VIF
age	1.467726
sex	1.616824
cp	1.181558
trestbps	1.291021
chol	1.310279
fbs	1.193257
restecg	1.074654
thalach	1.360148
exang	1.162967
oldpeak	1.375606
slope	1.570193
ca	1.410375
thal	1.161345

4 Model and Variable Selection

One of the most important statistical issues is the problem of choosing important variables in a regression model with a large number of predictor variables. Statistical models will not do well with massive irrelevant variables. Therefore, we ought to select a better model that only keep only useful and important variables. In this section, we use stepwise method to select important variables. It is observed from R output and Table 5, that the model with important variables (Best

Table 5: Results of fitting several logistic regression models to the heart disease dataset.

Model	Omitted predictor variables	Deviance G^2	df	AIC	Models Compared	Deviance Difference	P value
(1)	Full model	170.4385	278	198.4385	-	-	-
(2)	"Age"	170.9628	279	196.9628	(2)-(1)	$0.5243(df = 1)$	0.4690
(3)	"Age" & "Restecg"	171.3890	280	195.3890	(3)-(2)	$0.4262(df = 1)$	0.5139
(4)	"Age", "Restecg" & "Oldpeak"	172.6207	281	194.6207	(4)-(3)	$1.2317(df = 1)$	0.2671
(5)	"Age", "Restecg", "Oldpeak" & "Chol"	175.5397	282	195.5397	(5)-(4)	$2.9190(df = 1)$	0.0875

model) is as follow:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 \text{sex} + \beta_2 \text{cp} + \beta_3 \text{trestbps} + \beta_4 \text{chol} + \beta_5 \text{fbs} + \beta_6 \text{thalach} + \beta_7 \text{exang} + \beta_8 \text{slope} + \beta_9 \text{ca} + \beta_{10} \text{thal}.$$

5 Confusion matrix, Odds ratio and ROC curve for Models

In this section, we compute the confusion matrix and odds ratio for the full and the best models and evaluate models accuracy. The confusion matrix a tabular representation of actual vs predicted values. This helps us to find the accuracy of the model and avoid overfitting. The predictive power of the logistic regression model can also be tested using the receiver operating characteristic (ROC) technique. This technique determines the diagnostic ability and accuracy rate of the model. The area under the curve (AUC) of the sensitivity-specificity graph measures the models ability to distinguish between classes. The computed AUC value reflects our models binary classification ability and the returned value is between 0 and 1, with a perfect model having an AUC score of 1. The OR indicates how well each predictor, x value, affects the outcome variable. The odds ratio should range between 0 and infinity and given that the $OR > 1$, the variable is positively associated with the outcome. The higher the OR value, the higher it positively influences the outcome variable.

Listed in Table 6 are the odds ratio for the 13 independent variables for the full model and for the 10 independent variables for the best model. In the full model the variables with $OR < 1$ are Age, Fbs and Thalach. In Section 4, Age is the first variable removed from the full model. The OR value shows that Age does not positively influence the outcome variable. This indicates that the decision to remove Age from the first model is a good decision. The ROC curve for the best and full models are given in Figures 4. Additional information about the full model are:

- the model fits the dataset with an accuracy of 86.64%,
- 141 patients who are predicted to not have heart disease and 112 patients who were predicted to have heart disease were predicted correctly. So 253 out of the 292 patients are correctly classified,
- 17 of the patients were predicted to have heart disease when they do not have heart diseases.Also, 22 of the patients were predicted to not have heart disease, when they do have heart disease. So, 39 out of the 292 patients are misdiagnosed.
- The area under the curve (AUC) of the full model is 94.7%. This indicates that there is 94.7% chance that the model can distinguish between classes (0 and 1). Thus, the full model can discriminate the outcomes well.

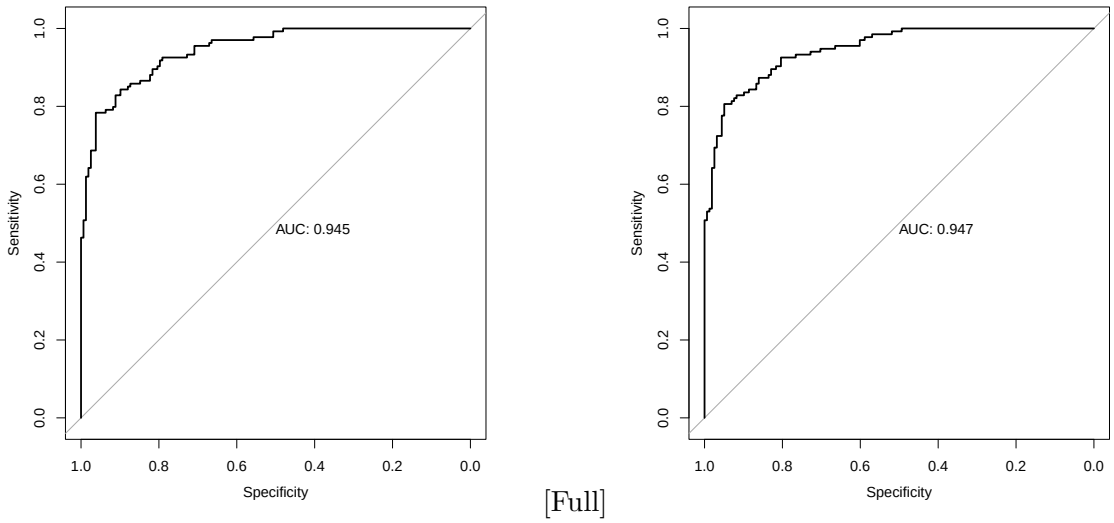
Now that the full model has been analyzed, we will compare the interpretations of the full model to the best model. The best model includes the independent variables "Sex", "CP", "Trestbps", "Chol", "Fbs", "Thalach", "Exang", "Slope", "CA" and "Thal" against the dependent variable "Heart disease". Additional information about this model are:

- The OR values for the best model are listed in Table 6. Comparing the OR values of the best model to the OR values of the full model in Table 6, the OR values from the best model

shows improvement. This indicates that the best model contains predictor values that have a higher chance of influencing the outcome.

- the model fits the dataset with an accuracy of 86.64%,
- 142 patients who are predicted to not have heart disease and 111 patients who were predicted to have heart disease were predicted correctly. So 253 out of the 292 patients are correctly classified,
- 16 of the patients were predicted to have heart disease when they do not have heart diseases.Also, 23 of the patients were predicted to not have heart disease, when they do have heart disease. So, 39 out of the 292 patients are misdiagnosed.
- The area under the curve (AUC) of the best model is 94.5%. This indicates that there is 94.5% chance that the model can distinguish between classes (0 and 1). Thus, the best model can discriminate the outcomes well.

So, the best model with fewer predictor variables has the same performance as the full model.The summary of the best model fit is given in Table 8.



[Best] [Full]

Figure 4: ROC Curve for the best and the full models.

Table 6: Odds ratio for the Models.

Predictor Vaiables	Full Model	Best Model
age	0.981	-
sex	6.162	7.334
cp	1.867	1.859
trestbps	1.033	1.032
chol	1.006	1.007
fbs	0.265	0.254
restecg	1.145	-
thalach	0.971	0.973
exang	3.118	3.455
oldpeak	1.264	-
slope	2.225	2.767
ca	5.131	5.082
thal	1.497	1.485

Table 7: Confusion Matrix.

	Predicted Status	Observed Status	
		Heart disease absence(0)	Heart disease present(1)
Best model	Heart disease absence(0)	142(0.486)	23(0.079)
	Heart disease present(1)	16(0.055)	111(0.380)
Full model	Heart disease absence(0)	141(0.483)	22(0.075)
	Heart disease present(1)	17(0.058)	112(0.384)

Table 8: Summary of the results for the best model for Heart Disease dataset.

Coefficients	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	−10.077019	2.840953	−3.547	0.000390
sex	1.992461	0.557716	3.573	0.000354
cp	0.619796	0.205724	3.013	0.002589
trestbps	0.031864	0.011137	2.861	0.004222
chol	0.006943	0.004020	1.727	0.084160
fbs	−1.369869	0.610527	−2.244	0.024849
thalach	−0.027074	0.010607	−2.552	0.010700
exang	1.239848	0.451873	2.744	0.006073
slope	1.017624	0.345229	2.948	0.003202
ca	1.625693	0.290580	5.595	0.00000002
thal	0.395130	0.106760	3.701	0.000215

6 Conclusions

Creating models that can aide in the medical industry for the prediction of a disease is crucial. This project demonstrated the prediction of a patient having heart disease based on numerous variables that influence the outcome, using the logistic regression. The importance of the model is that doctors can use the model to aide in their clinical decision making. The model can identify that if the patient is not yet diagnosed with heart disease, their health factors put them more at risk to developing heart disease in the future. In this project, we used the logistic regression to assess the variables to understand their significance and to obtain an accurate prediction model. The data set originally contained 303 instances and was reduced to 297 instances as the instances with missing values were removed. There are also 14 measurements provided for each patient. Initially, we fit full model with all 13 independent variables to dataset. In section 3, we explored outliers & unusual points and multicollinearity for the full model. In section 4, we use stepwise method to select important variables and introduce best model. The best model included the variables "Sex", "CP", "Trestbps", "Chol", "Fbs", "Thalach", "Exang", "Slope", "CA" and "Thal". Thus, the model removed the variables "Age", "Restecg", and "Oldpeak". In Section 5, the full model was compared to the best regressed model. Based on the odds ratio, confusion matrix, accuracy and AUC, the best model with fewer predictor variables has well performance the same as the full model.

Appendix

Appendix A: Heart Disease data set

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	heart_disease
1	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
2	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
3	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
4	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
5	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
6	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
7	62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
8	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
9	63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
10	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
11	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
12	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
13	56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
14	44	1	2	120	263	0	0	173	0	0.0	1	0	7	0
15	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
16	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
17	48	1	2	110	229	0	0	168	0	1.0	3	0	7	1
18	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
19	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
20	49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
21	64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
22	58	0	1	150	283	1	2	162	0	1.0	1	0	3	0
23	58	1	2	120	284	0	2	160	0	1.8	2	0	3	1
24	58	1	3	132	224	0	2	173	0	3.2	1	2	7	1
25	60	1	4	130	206	0	2	132	1	2.4	2	2	7	1
26	50	0	3	120	219	0	0	158	0	1.6	2	0	3	0
27	58	0	3	120	340	0	0	172	0	0.0	1	0	3	0
28	66	0	1	150	226	0	0	114	0	2.6	3	0	3	0
29	43	1	4	150	247	0	0	171	0	1.5	1	0	3	0
30	40	1	4	110	167	0	2	114	1	2.0	2	0	7	1
31	69	0	1	140	239	0	0	151	0	1.8	1	2	3	0
32	60	1	4	117	230	1	0	160	1	1.4	1	2	7	1
33	64	1	3	140	335	0	0	158	0	0.0	1	0	3	1
34	59	1	4	135	234	0	0	161	0	0.5	2	0	7	0
35	44	1	3	130	233	0	0	179	1	0.4	1	0	3	0
36	42	1	4	140	226	0	0	178	0	0.0	1	0	3	0
37	43	1	4	120	177	0	2	120	1	2.5	2	0	7	1
38	57	1	4	150	276	0	2	112	1	0.6	2	1	6	1
39	55	1	4	132	353	0	0	132	1	1.2	2	1	7	1
40	61	1	3	150	243	1	0	137	1	1.0	2	0	3	0
41	65	0	4	150	225	0	2	114	0	1.0	2	3	7	1
42	40	1	1	140	199	0	0	178	1	1.4	1	0	7	0
43	71	0	2	160	302	0	0	162	0	0.4	1	2	3	0
44	59	1	3	150	212	1	0	157	0	1.6	1	0	3	0
45	61	0	4	130	330	0	2	169	0	0.0	1	0	3	1
46	58	1	3	112	230	0	2	165	0	2.5	2	1	7	1
47	51	1	3	110	175	0	0	123	0	0.6	1	0	3	0
48	50	1	4	150	243	0	2	128	0	2.6	2	0	7	1
49	65	0	3	140	417	1	2	157	0	0.8	1	1	3	0
50	53	1	3	130	197	1	2	152	0	1.2	3	0	3	0
51	41	0	2	105	198	0	0	168	0	0.0	1	1	3	0
52	65	1	4	120	177	0	0	140	0	0.4	1	0	7	0

53	44	1	4	112	290	0	2	153	0	0.0	1	1	3	1
54	44	1	2	130	219	0	2	188	0	0.0	1	0	3	0
55	60	1	4	130	253	0	0	144	1	1.4	1	1	7	1
56	54	1	4	124	266	0	2	109	1	2.2	2	1	7	1
57	50	1	3	140	233	0	0	163	0	0.6	2	1	7	1
58	41	1	4	110	172	0	2	158	0	0.0	1	0	7	1
59	54	1	3	125	273	0	2	152	0	0.5	3	1	3	0
60	51	1	1	125	213	0	2	125	1	1.4	1	1	3	0
61	51	0	4	130	305	0	0	142	1	1.2	2	0	7	1
62	46	0	3	142	177	0	2	160	1	1.4	3	0	3	0
63	58	1	4	128	216	0	2	131	1	2.2	2	3	7	1
64	54	0	3	135	304	1	0	170	0	0.0	1	0	3	0
65	54	1	4	120	188	0	0	113	0	1.4	2	1	7	1
66	60	1	4	145	282	0	2	142	1	2.8	2	2	7	1
67	60	1	3	140	185	0	2	155	0	3.0	2	0	3	1
68	54	1	3	150	232	0	2	165	0	1.6	1	0	7	0
69	59	1	4	170	326	0	2	140	1	3.4	3	0	7	1
70	46	1	3	150	231	0	0	147	0	3.6	2	0	3	1
71	65	0	3	155	269	0	0	148	0	0.8	1	0	3	0
72	67	1	4	125	254	1	0	163	0	0.2	2	2	7	1
73	62	1	4	120	267	0	0	99	1	1.8	2	2	7	1
74	65	1	4	110	248	0	2	158	0	0.6	1	2	6	1
75	44	1	4	110	197	0	2	177	0	0.0	1	1	3	1
76	65	0	3	160	360	0	2	151	0	0.8	1	0	3	0
77	60	1	4	125	258	0	2	141	1	2.8	2	1	7	1
78	51	0	3	140	308	0	2	142	0	1.5	1	1	3	0
79	48	1	2	130	245	0	2	180	0	0.2	2	0	3	0
80	58	1	4	150	270	0	2	111	1	0.8	1	0	7	1
81	45	1	4	104	208	0	2	148	1	3.0	2	0	3	0
82	53	0	4	130	264	0	2	143	0	0.4	2	0	3	0
83	39	1	3	140	321	0	2	182	0	0.0	1	0	3	0
84	68	1	3	180	274	1	2	150	1	1.6	2	0	7	1
85	52	1	2	120	325	0	0	172	0	0.2	1	0	3	0
86	44	1	3	140	235	0	2	180	0	0.0	1	0	3	0
87	47	1	3	138	257	0	2	156	0	0.0	1	0	3	0
88	53	0	4	138	234	0	2	160	0	0.0	1	0	3	0
89	51	0	3	130	256	0	2	149	0	0.5	1	0	3	0
90	66	1	4	120	302	0	2	151	0	0.4	2	0	3	0
91	62	0	4	160	164	0	2	145	0	6.2	3	3	7	1
92	62	1	3	130	231	0	0	146	0	1.8	2	3	7	0
93	44	0	3	108	141	0	0	175	0	0.6	2	0	3	0
94	63	0	3	135	252	0	2	172	0	0.0	1	0	3	0
95	52	1	4	128	255	0	0	161	1	0.0	1	1	7	1
96	59	1	4	110	239	0	2	142	1	1.2	2	1	7	1
97	60	0	4	150	258	0	2	157	0	2.6	2	2	7	1
98	52	1	2	134	201	0	0	158	0	0.8	1	1	3	0
99	48	1	4	122	222	0	2	186	0	0.0	1	0	3	0
100	45	1	4	115	260	0	2	185	0	0.0	1	0	3	0
101	34	1	1	118	182	0	2	174	0	0.0	1	0	3	0
102	57	0	4	128	303	0	2	159	0	0.0	1	1	3	0
103	71	0	3	110	265	1	2	130	0	0.0	1	1	3	0
104	49	1	3	120	188	0	0	139	0	2.0	2	3	7	1
105	54	1	2	108	309	0	0	156	0	0.0	1	0	7	0
106	59	1	4	140	177	0	0	162	1	0.0	1	1	7	1
107	57	1	3	128	229	0	2	150	0	0.4	2	1	7	1
108	61	1	4	120	260	0	0	140	1	3.6	2	1	7	1

109	39	1	4	118	219	0	0	140	0	1.2	2	0	7	1
110	61	0	4	145	307	0	2	146	1	1.0	2	0	7	1
111	56	1	4	125	249	1	2	144	1	1.2	2	1	3	1
112	52	1	1	118	186	0	2	190	0	0.0	2	0	6	0
113	43	0	4	132	341	1	2	136	1	3.0	2	0	7	1
114	62	0	3	130	263	0	0	97	0	1.2	2	1	7	1
115	41	1	2	135	203	0	0	132	0	0.0	2	0	6	0
116	58	1	3	140	211	1	2	165	0	0.0	1	0	3	0
117	35	0	4	138	183	0	0	182	0	1.4	1	0	3	0
118	63	1	4	130	330	1	2	132	1	1.8	1	3	7	1
119	65	1	4	135	254	0	2	127	0	2.8	2	1	7	1
120	48	1	4	130	256	1	2	150	1	0.0	1	2	7	1
121	63	0	4	150	407	0	2	154	0	4.0	2	3	7	1
122	51	1	3	100	222	0	0	143	1	1.2	2	0	3	0
123	55	1	4	140	217	0	0	111	1	5.6	3	0	7	1
124	65	1	1	138	282	1	2	174	0	1.4	2	1	3	1
125	45	0	2	130	234	0	2	175	0	0.6	2	0	3	0
126	56	0	4	200	288	1	2	133	1	4.0	3	2	7	1
127	54	1	4	110	239	0	0	126	1	2.8	2	1	7	1
128	44	1	2	120	220	0	0	170	0	0.0	1	0	3	0
129	62	0	4	124	209	0	0	163	0	0.0	1	0	3	0
130	54	1	3	120	258	0	2	147	0	0.4	2	0	7	0
131	51	1	3	94	227	0	0	154	1	0.0	1	1	7	0
132	29	1	2	130	204	0	2	202	0	0.0	1	0	3	0
133	51	1	4	140	261	0	2	186	1	0.0	1	0	3	0
134	43	0	3	122	213	0	0	165	0	0.2	2	0	3	0
135	55	0	2	135	250	0	2	161	0	1.4	2	0	3	0
136	70	1	4	145	174	0	0	125	1	2.6	3	0	7	1
137	62	1	2	120	281	0	2	103	0	1.4	2	1	7	1
138	35	1	4	120	198	0	0	130	1	1.6	2	0	7	1
139	51	1	3	125	245	1	2	166	0	2.4	2	0	3	0
140	59	1	2	140	221	0	0	164	1	0.0	1	0	3	0
141	59	1	1	170	288	0	2	159	0	0.2	2	0	7	1
142	52	1	2	128	205	1	0	184	0	0.0	1	0	3	0
143	64	1	3	125	309	0	0	131	1	1.8	2	0	7	1
144	58	1	3	105	240	0	2	154	1	0.6	2	0	7	0
145	47	1	3	108	243	0	0	152	0	0.0	1	0	3	1
146	57	1	4	165	289	1	2	124	0	1.0	2	3	7	1
147	41	1	3	112	250	0	0	179	0	0.0	1	0	3	0
148	45	1	2	128	308	0	2	170	0	0.0	1	0	3	0
149	60	0	3	102	318	0	0	160	0	0.0	1	1	3	0
150	52	1	1	152	298	1	0	178	0	1.2	2	0	7	0
151	42	0	4	102	265	0	2	122	0	0.6	2	0	3	0
152	67	0	3	115	564	0	2	160	0	1.6	2	0	7	0
153	55	1	4	160	289	0	2	145	1	0.8	2	1	7	1
154	64	1	4	120	246	0	2	96	1	2.2	3	1	3	1
155	70	1	4	130	322	0	2	109	0	2.4	2	3	3	1
156	51	1	4	140	299	0	0	173	1	1.6	1	0	7	1
157	58	1	4	125	300	0	2	171	0	0.0	1	2	7	1
158	60	1	4	140	293	0	2	170	0	1.2	2	2	7	1
159	68	1	3	118	277	0	0	151	0	1.0	1	1	7	0
160	46	1	2	101	197	1	0	156	0	0.0	1	0	7	0
161	77	1	4	125	304	0	2	162	1	0.0	1	3	3	1
162	54	0	3	110	214	0	0	158	0	1.6	2	0	3	0
163	58	0	4	100	248	0	2	122	0	1.0	2	0	3	0
164	48	1	3	124	255	1	0	175	0	0.0	1	2	3	0

165	57	1	4	132	207	0	0	168	1	0.0	1	0	7	0
166	54	0	2	132	288	1	2	159	1	0.0	1	1	3	0
167	35	1	4	126	282	0	2	156	1	0.0	1	0	7	1
168	45	0	2	112	160	0	0	138	0	0.0	2	0	3	0
169	70	1	3	160	269	0	0	112	1	2.9	2	1	7	1
170	53	1	4	142	226	0	2	111	1	0.0	1	0	7	0
171	59	0	4	174	249	0	0	143	1	0.0	2	0	3	1
172	62	0	4	140	394	0	2	157	0	1.2	2	0	3	0
173	64	1	4	145	212	0	2	132	0	2.0	2	2	6	1
174	57	1	4	152	274	0	0	88	1	1.2	2	1	7	1
175	52	1	4	108	233	1	0	147	0	0.1	1	3	7	0
176	56	1	4	132	184	0	2	105	1	2.1	2	1	6	1
177	43	1	3	130	315	0	0	162	0	1.9	1	1	3	0
178	53	1	3	130	246	1	2	173	0	0.0	1	3	3	0
179	48	1	4	124	274	0	2	166	0	0.5	2	0	7	1
180	56	0	4	134	409	0	2	150	1	1.9	2	2	7	1
181	42	1	1	148	244	0	2	178	0	0.8	1	2	3	0
182	59	1	1	178	270	0	2	145	0	4.2	3	0	7	0
183	60	0	4	158	305	0	2	161	0	0.0	1	0	3	1
184	63	0	2	140	195	0	0	179	0	0.0	1	2	3	0
185	42	1	3	120	240	1	0	194	0	0.8	3	0	7	0
186	66	1	2	160	246	0	0	120	1	0.0	2	3	6	1
187	54	1	2	192	283	0	2	195	0	0.0	1	1	7	1
188	69	1	3	140	254	0	2	146	0	2.0	2	3	7	1
189	50	1	3	129	196	0	0	163	0	0.0	1	0	3	0
190	51	1	4	140	298	0	0	122	1	4.2	2	3	7	1
191	62	0	4	138	294	1	0	106	0	1.9	2	3	3	1
192	68	0	3	120	211	0	2	115	0	1.5	2	0	3	0
193	67	1	4	100	299	0	2	125	1	0.9	2	2	3	1
194	69	1	1	160	234	1	2	131	0	0.1	2	1	3	0
195	45	0	4	138	236	0	2	152	1	0.2	2	0	3	0
196	50	0	2	120	244	0	0	162	0	1.1	1	0	3	0
197	59	1	1	160	273	0	2	125	0	0.0	1	0	3	1
198	50	0	4	110	254	0	2	159	0	0.0	1	0	3	0
199	64	0	4	180	325	0	0	154	1	0.0	1	0	3	0
200	57	1	3	150	126	1	0	173	0	0.2	1	1	7	0
201	64	0	3	140	313	0	0	133	0	0.2	1	0	7	0
202	43	1	4	110	211	0	0	161	0	0.0	1	0	7	0
203	45	1	4	142	309	0	2	147	1	0.0	2	3	7	1
204	58	1	4	128	259	0	2	130	1	3.0	2	2	7	1
205	50	1	4	144	200	0	2	126	1	0.9	2	0	7	1
206	55	1	2	130	262	0	0	155	0	0.0	1	0	3	0
207	62	0	4	150	244	0	0	154	1	1.4	2	0	3	1
208	37	0	3	120	215	0	0	170	0	0.0	1	0	3	0
209	38	1	1	120	231	0	0	182	1	3.8	2	0	7	1
210	41	1	3	130	214	0	2	168	0	2.0	2	0	3	0
211	66	0	4	178	228	1	0	165	1	1.0	2	2	7	1
212	52	1	4	112	230	0	0	160	0	0.0	1	1	3	1
213	56	1	1	120	193	0	2	162	0	1.9	2	0	7	0
214	46	0	2	105	204	0	0	172	0	0.0	1	0	3	0
215	46	0	4	138	243	0	2	152	1	0.0	2	0	3	0
216	64	0	4	130	303	0	0	122	0	2.0	2	2	3	0
217	59	1	4	138	271	0	2	182	0	0.0	1	0	3	0
218	41	0	3	112	268	0	2	172	1	0.0	1	0	3	0
219	54	0	3	108	267	0	2	167	0	0.0	1	0	3	0
220	39	0	3	94	199	0	0	179	0	0.0	1	0	3	0

221	53	1	4	123	282	0	0	95	1	2.0	2	2	7	1
222	63	0	4	108	269	0	0	169	1	1.8	2	2	3	1
223	34	0	2	118	210	0	0	192	0	0.7	1	0	3	0
224	47	1	4	112	204	0	0	143	0	0.1	1	0	3	0
225	67	0	3	152	277	0	0	172	0	0.0	1	1	3	0
226	54	1	4	110	206	0	2	108	1	0.0	2	1	3	1
227	66	1	4	112	212	0	2	132	1	0.1	1	1	3	1
228	52	0	3	136	196	0	2	169	0	0.1	2	0	3	0
229	55	0	4	180	327	0	1	117	1	3.4	2	0	3	1
230	49	1	3	118	149	0	2	126	0	0.8	1	3	3	1
231	74	0	2	120	269	0	2	121	1	0.2	1	1	3	0
232	54	0	3	160	201	0	0	163	0	0.0	1	1	3	0
233	54	1	4	122	286	0	2	116	1	3.2	2	2	3	1
234	56	1	4	130	283	1	2	103	1	1.6	3	0	7	1
235	46	1	4	120	249	0	2	144	0	0.8	1	0	7	1
236	49	0	2	134	271	0	0	162	0	0.0	2	0	3	0
237	42	1	2	120	295	0	0	162	0	0.0	1	0	3	0
238	41	1	2	110	235	0	0	153	0	0.0	1	0	3	0
239	41	0	2	126	306	0	0	163	0	0.0	1	0	3	0
240	49	0	4	130	269	0	0	163	0	0.0	1	0	3	0
241	61	1	1	134	234	0	0	145	0	2.6	2	2	3	1
242	60	0	3	120	178	1	0	96	0	0.0	1	0	3	0
243	67	1	4	120	237	0	0	71	0	1.0	2	0	3	1
244	58	1	4	100	234	0	0	156	0	0.1	1	1	7	1
245	47	1	4	110	275	0	2	118	1	1.0	2	1	3	1
246	52	1	4	125	212	0	0	168	0	1.0	1	2	7	1
247	62	1	2	128	208	1	2	140	0	0.0	1	0	3	0
248	57	1	4	110	201	0	0	126	1	1.5	2	0	6	0
249	58	1	4	146	218	0	0	105	0	2.0	2	1	7	1
250	64	1	4	128	263	0	0	105	1	0.2	2	1	7	0
251	51	0	3	120	295	0	2	157	0	0.6	1	0	3	0
252	43	1	4	115	303	0	0	181	0	1.2	2	0	3	0
253	42	0	3	120	209	0	0	173	0	0.0	2	0	3	0
254	67	0	4	106	223	0	0	142	0	0.3	1	2	3	0
255	76	0	3	140	197	0	1	116	0	1.1	2	0	3	0
256	70	1	2	156	245	0	2	143	0	0.0	1	0	3	0
257	57	1	2	124	261	0	0	141	0	0.3	1	0	7	1
258	44	0	3	118	242	0	0	149	0	0.3	2	1	3	0
259	58	0	2	136	319	1	2	152	0	0.0	1	2	3	1
260	60	0	1	150	240	0	0	171	0	0.9	1	0	3	0
261	44	1	3	120	226	0	0	169	0	0.0	1	0	3	0
262	61	1	4	138	166	0	2	125	1	3.6	2	1	3	1
263	42	1	4	136	315	0	0	125	1	1.8	2	0	6	1
264	59	1	3	126	218	1	0	134	0	2.2	2	1	6	1
265	40	1	4	152	223	0	0	181	0	0.0	1	0	7	1
266	42	1	3	130	180	0	0	150	0	0.0	1	0	3	0
267	61	1	4	140	207	0	2	138	1	1.9	1	1	7	1
268	66	1	4	160	228	0	2	138	0	2.3	1	0	6	0
269	46	1	4	140	311	0	0	120	1	1.8	2	2	7	1
270	71	0	4	112	149	0	0	125	0	1.6	2	0	3	0
271	59	1	1	134	204	0	0	162	0	0.8	1	2	3	1
272	64	1	1	170	227	0	2	155	0	0.6	2	0	7	0
273	66	0	3	146	278	0	2	152	0	0.0	2	1	3	0
274	39	0	3	138	220	0	0	152	0	0.0	2	0	3	0
275	57	1	2	154	232	0	2	164	0	0.0	1	1	3	1
276	58	0	4	130	197	0	0	131	0	0.6	2	0	3	0

```
277  57   1  4      110  335   0       0     143     1     3.0     2  1    7            1
278  47   1  3      130  253   0       0     179     0     0.0     1  0    3            0
279  55   0  4      128  205   0       1     130     1     2.0     2  1    7            1
280  35   1  2      122  192   0       0     174     0     0.0     1  0    3            0
281  61   1  4      148  203   0       0     161     0     0.0     1  1    7            1
282  58   1  4      114  318   0       1     140     0     4.4     3  3    6            1
283  58   0  4      170  225   1       2     146     1     2.8     2  2    6            1
284  56   1  2      130  221   0       2     163     0     0.0     1  0    7            0
285  56   1  2      120  240   0       0     169     0     0.0     3  0    3            0
286  67   1  3      152  212   0       2     150     0     0.8     2  0    7            1
287  55   0  2      132  342   0       0     166     0     1.2     1  0    3            0
288  44   1  4      120  169   0       0     144     1     2.8     3  0    6            1
289  63   1  4      140  187   0       2     144     1     4.0     1  2    7            1
290  63   0  4      124  197   0       0     136     1     0.0     2  0    3            1
291  41   1  2      120  157   0       0     182     0     0.0     1  0    3            0
292  59   1  4      164  176   1       2      90     0     1.0     2  2    6            1
293  57   0  4      140  241   0       0     123     1     0.2     2  0    7            1
294  45   1  1      110  264   0       0     132     0     1.2     2  0    7            1
295  68   1  4      144  193   1       0     141     0     3.4     2  2    7            1
296  57   1  4      130  131   0       0     115     1     1.2     2  1    7            1
297  57   0  2      130  236   0       2     174     0     0.0     2  1    3            1
```

Appendix B: R Code

```
suppressWarnings({library(dplyr)})
suppressWarnings({library(tidyr)})
suppressWarnings({library(Matrix)})
suppressWarnings({library(corrplot)})
suppressWarnings({library(ggplot2)})
suppressWarnings({library(lmtest)})
suppressWarnings({library(GGally)})
suppressWarnings({library(caTools)})
suppressWarnings({library(lmtest)})
suppressWarnings({library(MASS)})
suppressWarnings({library(pROC)})
suppressWarnings({library(cowplot)})
#------------------------------------------------------------------------
#                         Section 1
#------------------------------------------------------------------------
data_cleveland<-read.csv(file="D:/cleveland.csv",header=F,sep=",")
names<-c("age","sex","cp","trestbps","chol","fbs","restecg","thalach",
         "exang","oldpeak","slope","ca","thal","heart_disease")
colnames(data_cleveland)<-names
str(data_cleveland)
head(data_cleveland)
#------------------------------------------------------------------------
data_cleveland<-data_cleveland%>%
    mutate(heart_disease=case_when(heart_disease==0~"absence",
                           TRUE ~"presence"))
data_cleveland<-data_cleveland%>%
    mutate(heart_disease=case_when(heart_disease=="absence"~0,
                           heart_disease=="presence"~1))
data_cleveland<-data_cleveland%>%drop_na()
#------------------------------------------------------------------------
#                         Section 2
#------------------------------------------------------------------------
```

```
dist_plot<-ggplot(data_cleveland,aes(x=heart_disease))+
geom_bar(fill=c("0"="blue","1"="red"),color="black",stat="count")+
labs(title="Plot(b)",x="heart disease",y="Frequency")
dist_plot
#-------------------------------------------------------------------------
age.plot<-ggplot(data_cleveland,aes(x=age,fill=factor(heart_disease)))+
geom_histogram(binwidth=4,position="dodge",color="grey")+
scale_fill_manual(values=c("0"="blue","1"="red"),labels=c("No Disease","Disease"))+
facet_wrap(~heart_disease,scales="free_y")+
    labs(title="Plot(a)", x = "Age (years)", y = "Count", fill = "Heart Disease")
age.plot
#-------------------------------------------------------------------------
sex_plot<-ggplot(data_cleveland,aes(x=factor(sex),fill=factor(heart_disease)))+geom_bar()+
labs(title="Plot(c)",x="Gender(0=Female,1=Male)",y="Frequency")+
scale_fill_manual(values=c("0"="blue","1"="red"),
                  labels=c("No Disease","Disease"))+
labs(x="Gender(0=Female,1=Male)",y="Count",fill="Heart Disease")
sex_plot
plot_grid(age_plot,plot_grid(dist_plot,sex_plot), ncol = 1)
#-------------------------------------------------------------------------
corr_mat<-cor(data_cleveland,method="pearson")
ggcorr(data_cleveland,label=TRUE,label_size=2.5,hjust=1,layout.exp=2)
#-------------------------------------------------------------------------
summary_table<-summary(data_cleveland)
t(summary_table)
#-------------------------------------------------------------------------
#                        Section 3
#-------------------------------------------------------------------------
heart_full_model<-glm(heart_disease~.,data=data_cleveland,family=binomial(logit))
summary(heart_full_model)
#-------------------------------------------------------------------------
#                        Section (3-1)
#-------------------------------------------------------------------------
aplha<-0.05
r_Stud_Res<-stdres(heart_full_model)
t_range<-3
Data_outliers_Stud_Res<-abs(r_Stud_Res)[abs(r_Stud_Res)>t_range]
id_Data_outliers_Stud_Res<-which(abs(r_Stud_Res)>t_range)
result_Data_outliers_Stud_Res<-data.frame(id_Data_outliers_Stud_Res,
                                  abs(r_Stud_Res)[id_Data_outliers_Stud_Res])
colnames(result_Data_outliers_Stud_Res)<-c("Id","|Studentized Residuals|")
data_outliers_plot<-ggplot(data.frame(1:nrow(data_cleveland),abs(r_Stud_Res)),
    aes(x=1:nrow(data_cleveland),y=abs(r_Stud_Res)))+
    geom_point(size=2)+geom_path()+
    geom_hline(yintercept=0,linetype="dashed",color="blue",size=1)+
    geom_hline(yintercept=3,linetype="dashed",color="blue",size=1)+
    xlab("Observation")+ylab("|Studentized Residuals|")+
    labs(title="|Studentized Residuals|")
data_outliers_plot
#-----------------------------------------------------------------
if(length(Data_outliers_Stud_Res)>0){
cat("\n Outliers and Unusual Points based on Studentized Residuals(r_Stud_Res)\n")
print(result_Data_outliers_Stud_Res)
}else{
cat("\n No outliers in data based on Studentized Residuals(r_Stud_Res)\n")
}
```

```
#-----------------------------------------------------------------------
data_cleveland_new<-data_cleveland[-id_Data_outliers_Stud_Res,]
heart_full_model_new<-glm(heart_disease~.,data=data_cleveland_new,family=binomial(logit))
summary(heart_full_model_new)
#-----------------------------------------------------------------------
#                        Section (3-2)
#-----------------------------------------------------------------------
res_VIF<-matrix(car::vif(heart_full_model_new),13,1)
rownames(res_VIF)<-names[-14]
colnames(res_VIF)<-"VIF"
res_VIF
#-----------------------------------------------------------------------
#                        Section 4
#-----------------------------------------------------------------------
model1<-glm(heart_disease~.,data=data_cleveland_new,family=binomial(logit))
model2<-glm(heart_disease~sex+cp+trestbps+chol+fbs+restecg+thalach+exang+
                 oldpeak+slope+ca+thal,data=data_cleveland_new,family=binomial(logit))
model3<-glm(heart_disease~sex+cp+trestbps+chol+fbs+thalach+exang+oldpeak+slope+ca+thal,
            data=data_cleveland_new,family=binomial(logit))
model4<-glm(heart_disease~sex+cp+trestbps+chol+fbs+thalach+exang+slope+ca+thal,
            data=data_cleveland_new,family=binomial(logit))
model5<-glm(heart_disease~sex+cp+trestbps+fbs+thalach+exang+slope+ca+thal,
            data=data_cleveland_new,family=binomial(logit))
#-----------------------------------------------------------------------
deviance_vec<-c(deviance(model1),deviance(model2),deviance(model3),
                        deviance(model4),deviance(model5))
df_vec<-c(df.residual(model1),df.residual(model2),df.residual(model3),
              df.residual(model4),df.residual(model5))
AIC_vec<-c(AIC(model1),AIC(model2),AIC(model3),AIC(model4),AIC(model5))
Diff_deviance<-c(0,deviance_vec[2]-deviance_vec[1],deviance_vec[3]-deviance_vec[2],
                 deviance_vec[4]-deviance_vec[3],deviance_vec[5]-deviance_vec[4])
Diff_deviance<-round(Diff_deviance,4)
pValue1<-1-pchisq(Diff_deviance[2],1)
pValue2<-1-pchisq(Diff_deviance[3],1)
pValue3<-1-pchisq(Diff_deviance[4],1)
pValue4<-1-pchisq(Diff_deviance[5],1)
pValue_vec<-c(0,pValue1,pValue2,pValue3,pValue4)
pValue_vec<-round(pValue_vec,4)
model_compared<-c("-","2-1","3-2","4-3","5-4")
Table5<-data.frame(deviance_vec,df_vec,AIC_vec,model_compared,Diff_deviance,pValue_vec)
rownames(Table5)<-c("Model 1","Model 2","Model 3","Model 4","Model 5")
#-----------------------------------------------------------------------
best_model<-step(heart_full_model_new)
summary(best_model)
#-----------------------------------------------
nullModel<-glm(heart_disease~1,family=binomial(link=logit),data=data_cleveland_new)
fullModel<-glm(heart_disease~age+sex+cp+trestbps+chol+fbs+restecg+thalach+exang+
                      oldpeak+slope+ca+thal,family=binomial(link=logit),data=data_cleveland_new)
summary(stepAIC(fullModel,direction="backward",
                scope=list(upper=fullModel,lower=nullModel),
                trace=F))
#-----------------------------------------------------------------------
#                        Section 5
#-----------------------------------------------------------------------
models.probs_best<-predict(model4,data_cleveland_new,type="response")
models.pred_best<-rep(0,length(models.probs_best))
```

```
models.pred_best[models.probs_best>0.5] <- 1
Table_model_performance_best<-table(models.pred_best,data_cleveland_new$heart_disease,
                          dnn = c("Predicted Status","Observed Status"))
Table_model_performance_best
prop.table(Table_model_performance_best)
#-----------------------------------------------------------------------
models.probs_full<-predict(model1,data_cleveland_new,type="response")
models.pred_full<-rep(0,length(models.probs_full))
models.pred_full[models.probs_full>0.5] <- 1
Table_model_performance_full<-table(models.pred_full,data_cleveland_new$heart_disease,
                          dnn = c("Predicted Status","Observed Status"))
Table_model_performance_full
prop.table(Table_model_performance_full)
#-----------------------------------------------------------------------
odds_ratio_full<-exp(coef(model1))
round(odds_ratio_full,3)
#-----------------------------------------------------------------------
odds_ratio_best<-exp(coef(model4))
round(odds_ratio_best,3)
#-----------------------------------------------------------------------
ROC_best<-roc(data_cleveland_new$heart_disease~models.probs_best,plot=TRUE,print.auc=TRUE)
as.numeric(ROC_best$auc)
ROC__full<-roc(data_cleveland_new$heart_disease~models.probs_full,plot=TRUE,print.auc=TRUE)
as.numeric(ROC__full$auc)
```

References

[1] Agresti, A. (2013) Categorical Data Analysis. 3rd Edition, John Wiley & Sons Inc., Hoboken.

[2] Hosmer, David W., and Stanley Lemeshow. (2000). Applied Logistic Regression, Second Edition. Wiley- Interscience Publication, New York.

[3] UCI Machine Learning. Heart Disease Data Set. Retrieved from https://archive.ics.uci.edu/ml/datasets/heart+disease.