Heart Disease Prediction

Bhavani Chalamalla & Emily Wood

MA5790 - Predictive Modeling

https://www.news-medical.net/health/Structure-and-Function-of-the-Heart.aspx

Table Of Contents

01

Data Exploration

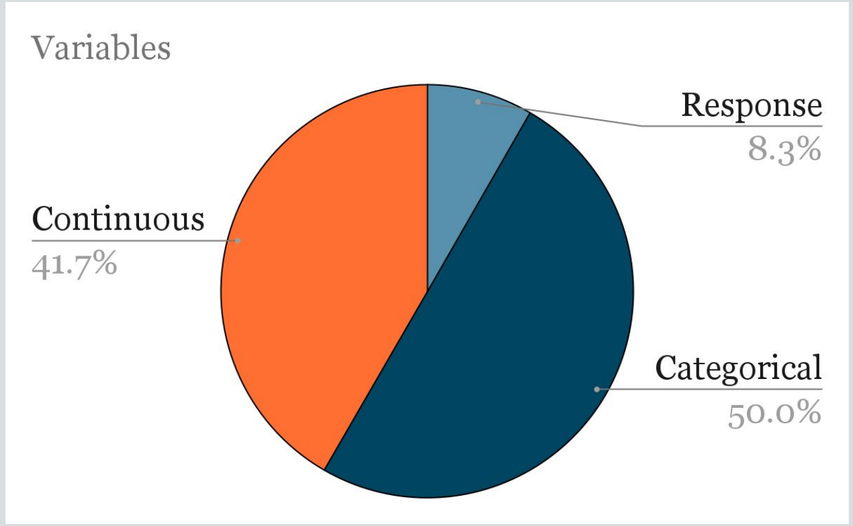
Goal

Predict heart disease of individuals based on various medical and clinical attributes. Cardiovascular diseases are a leading cause of death globally, and early detection and management are crucial for individuals at risk.

https://www.foundationhealth.org/our_community/fhp_healthbreak/7_heart_symptoms_you_shouldnt_ignore

Our Dataset

- 12 columns
- 918 observations



	A	B	C	D	E	F	G	H	I	J	K	L
1	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
2	40	M	ATA	140	289	0	Normal	172	N	0	Up	0
3	49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
4	37	M	ATA	130	283	0	ST	98	N	0	Up	0
5	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
6	54	M	NAP	150	195	0	Normal	122	N	0	Up	0
7	39	M	NAP	120	339	0	Normal	170	N	0	Up	0

Heart Disease Dataset

Categorical

1. **Sex**: sex of the patient [M: Male, F: Female]
2. **ChestPainType**: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
3. **FastingBS**: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
4. **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]
5. **ExerciseAngina**: exercise-induced angina [Y: Yes, N: No]
6. **ST_Slope**: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]

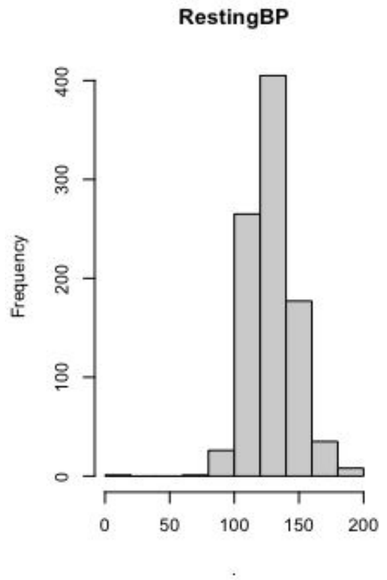
Continuous

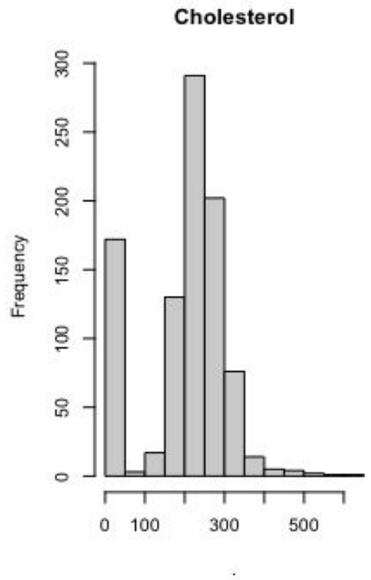
1. **Age**: age of the patient [years]
2. **RestingBP**: resting blood pressure [mm Hg]
3. **Cholesterol**: serum cholesterol [mm/dl]
4. **MaxHR**: maximum heart rate achieved [Numeric value between 60 and 202]
5. **Oldpeak**: oldpeak = ST [Numeric value measured in depression]

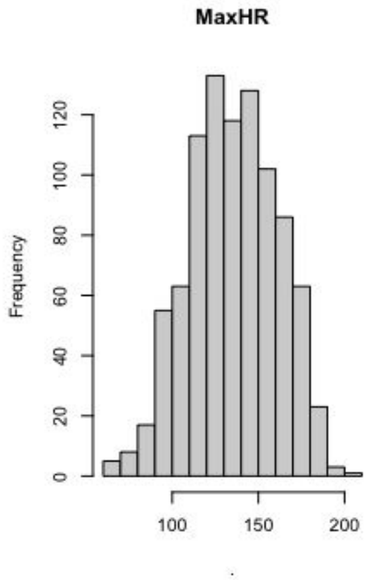
Response Variable

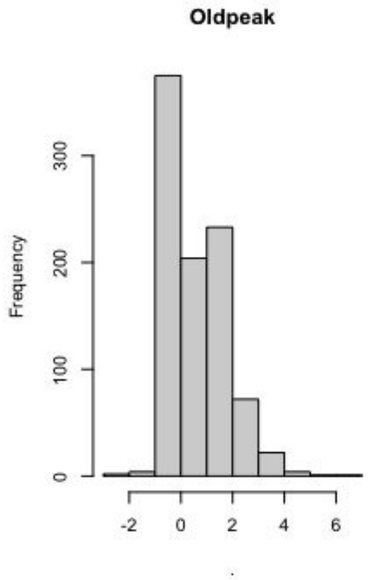
- **HeartDisease**: output class [1: heart disease, 0: Normal]

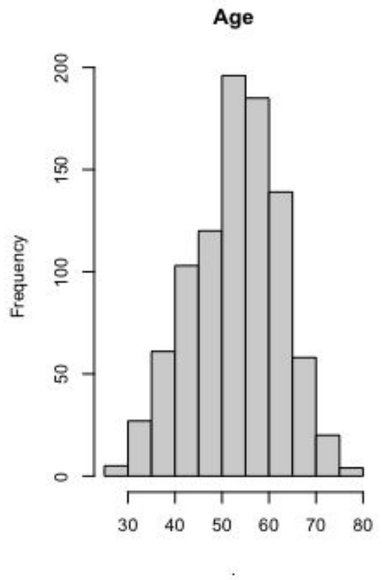
Continuous Data

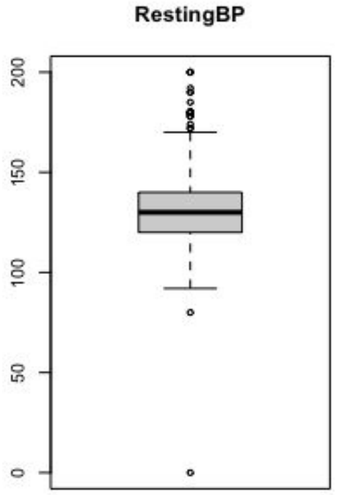


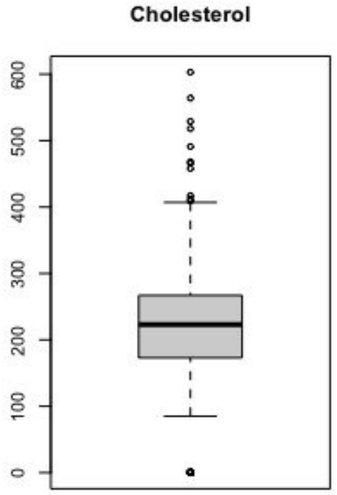


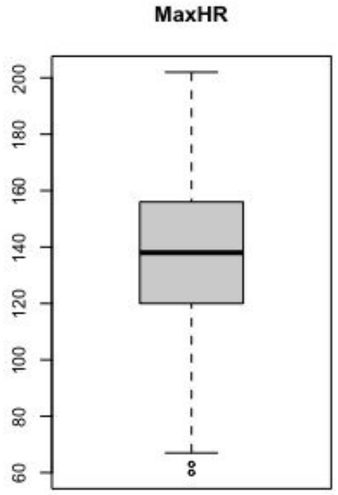


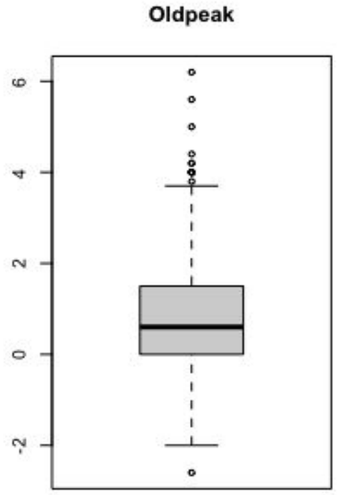


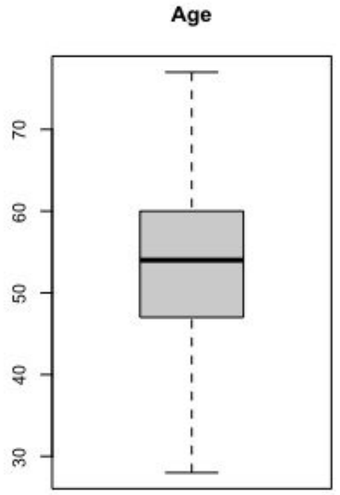






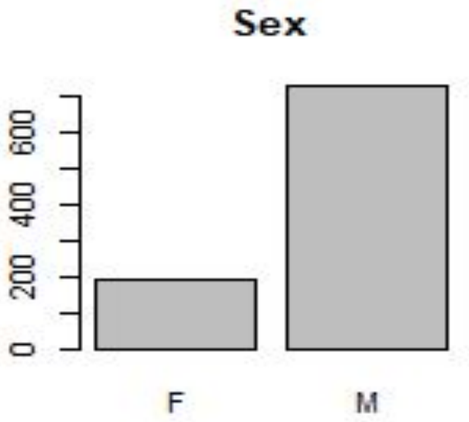


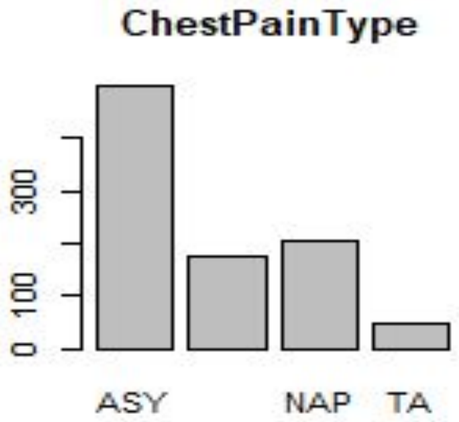


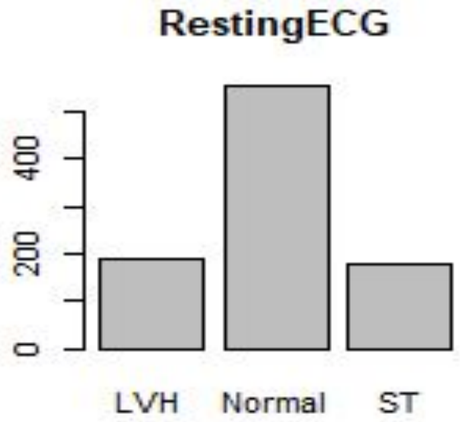


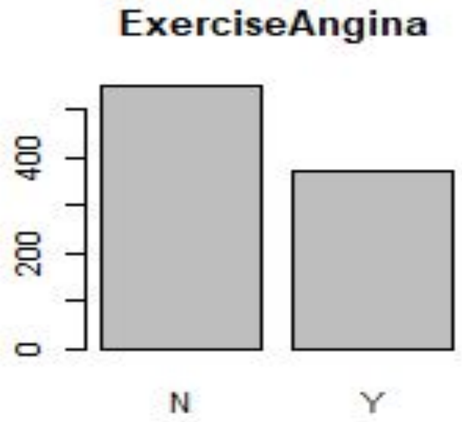
Skew:	0.1792520	-0.6080941	-0.1438880	1.0195317	-0.1952932

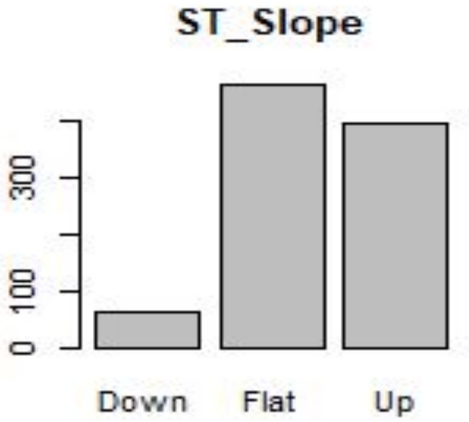
Categorical Data

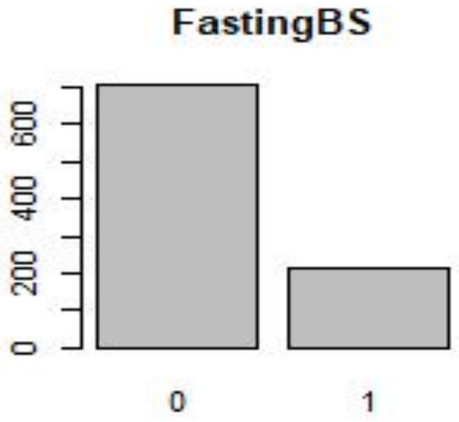


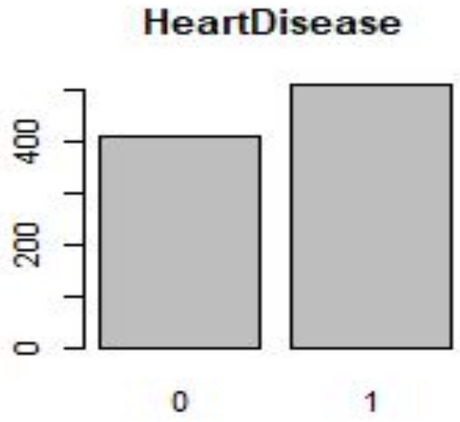












02

Data Preprocessing/ Spending

https://deepnote.com/@a_mas/Data-Preprocessing-Tools-4943e322-768c-4961-b30f-c0e8f63bf0ec

Missing Values and duplicates

- No missing values
- No duplicate observations

```
> cat("Number of missing values in the dataset:", missing_values, "\n")
Number of missing values in the dataset: 0
```

```
> duplicate_rows <- heart_data[duplicated(heart_data), ]
> print(duplicate_rows)
# A tibble: 0 × 12
# i 12 variables: Age <dbl>, Sex <chr>, ChestPainType <chr>, RestingBP <dbl>, Cholesterol <dbl>,
#   FastingBS <dbl>, RestingECG <chr>, MaxHR <dbl>, ExerciseAngina <chr>, Oldpeak <dbl>,
#   ST_Slope <chr>, HeartDisease <dbl>
```

Near- zero variance for Categorical

```
> nearzero_var <- nearZeroVar(heart, saveMetrics = TRUE)
> # Display the results
> print(nearzero_var)
                  freqRatio percentUnique zeroVar   nzv
SexM               3.756477     0.2178649   FALSE FALSE
ChestPainTypeATA   4.306358     0.2178649   FALSE FALSE
ChestPainTypeNAP   3.522167     0.2178649   FALSE FALSE
ChestPainTypeTA   18.956522     0.2178649   FALSE FALSE
RestingECGNormal   1.508197     0.2178649   FALSE FALSE
RestingECGST       4.157303     0.2178649   FALSE FALSE
ExerciseAnginaY    1.474394     0.2178649   FALSE FALSE
ST_SlopeFlat       1.004367     0.2178649   FALSE FALSE
ST_SlopeUp         1.324051     0.2178649   FALSE FALSE
Age                1.214286     5.4466231   FALSE FALSE
RestingBP          1.118644     7.2984749   FALSE FALSE
Cholesterol       15.636364    24.1830065   FALSE FALSE
FastingBS          3.289720     0.2178649   FALSE FALSE
MaxHR              1.048780    12.9629630   FALSE FALSE
Oldpeak            4.279070     5.7734205   FALSE FALSE
```

Analysing near zero variance

The output we received indicates that there are no near-zero variance predictors among the categorical variables in our dataset.

freqRatio: This is the frequency ratio, and it checks the ratio of the most common category to the second most common category.

percentUnique: This metric checks the percentage of unique values in a variable. If this percentage is low (typically less than 1), it can indicate near-zero variance. However, in our case, there are no categorical predictors with a low percentage of unique values.

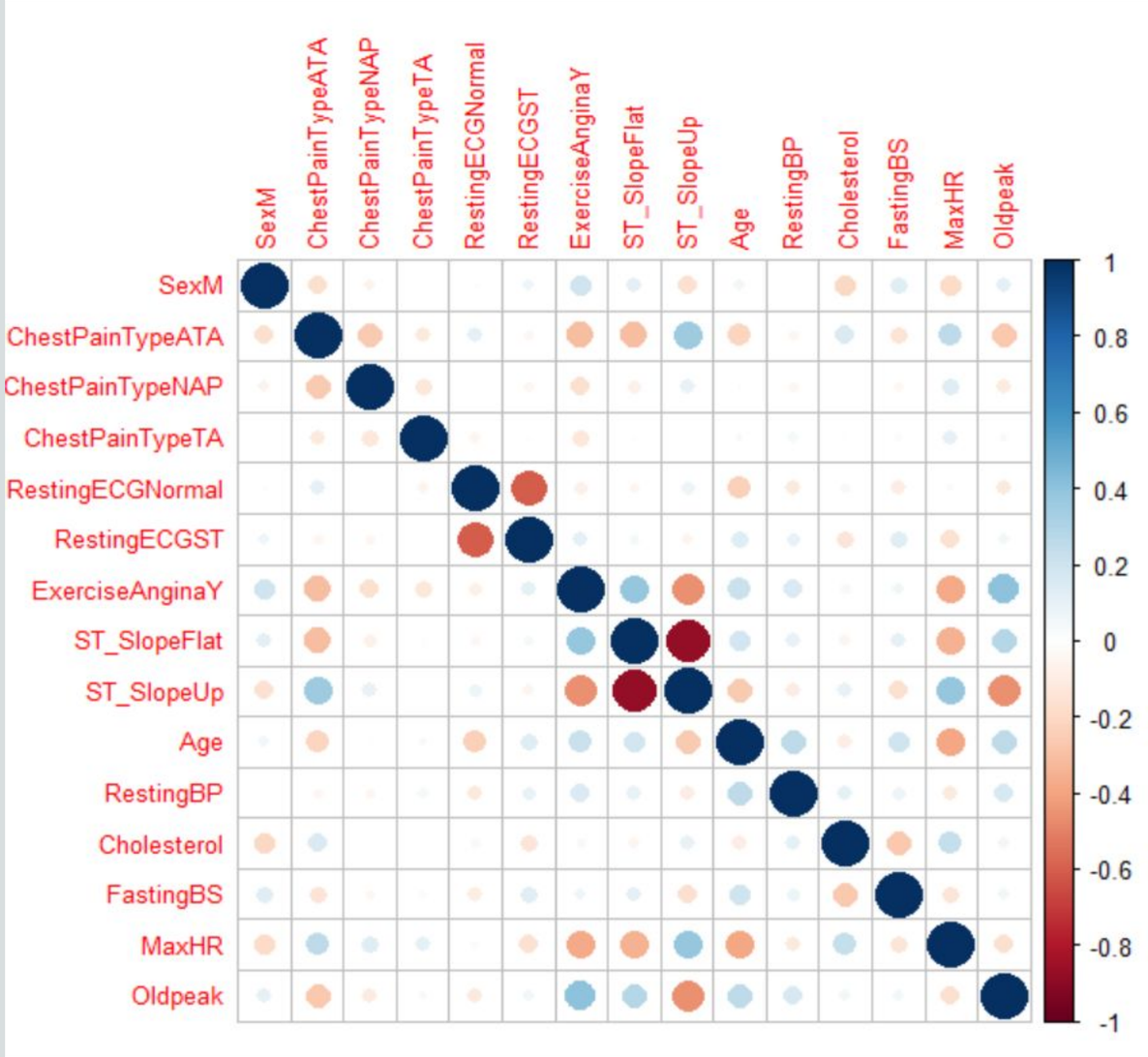
zeroVar: This indicates whether any of the categorical variables have zero variance, meaning that they have only a single unique value. In our case, there are no categorical variables with zero variance.

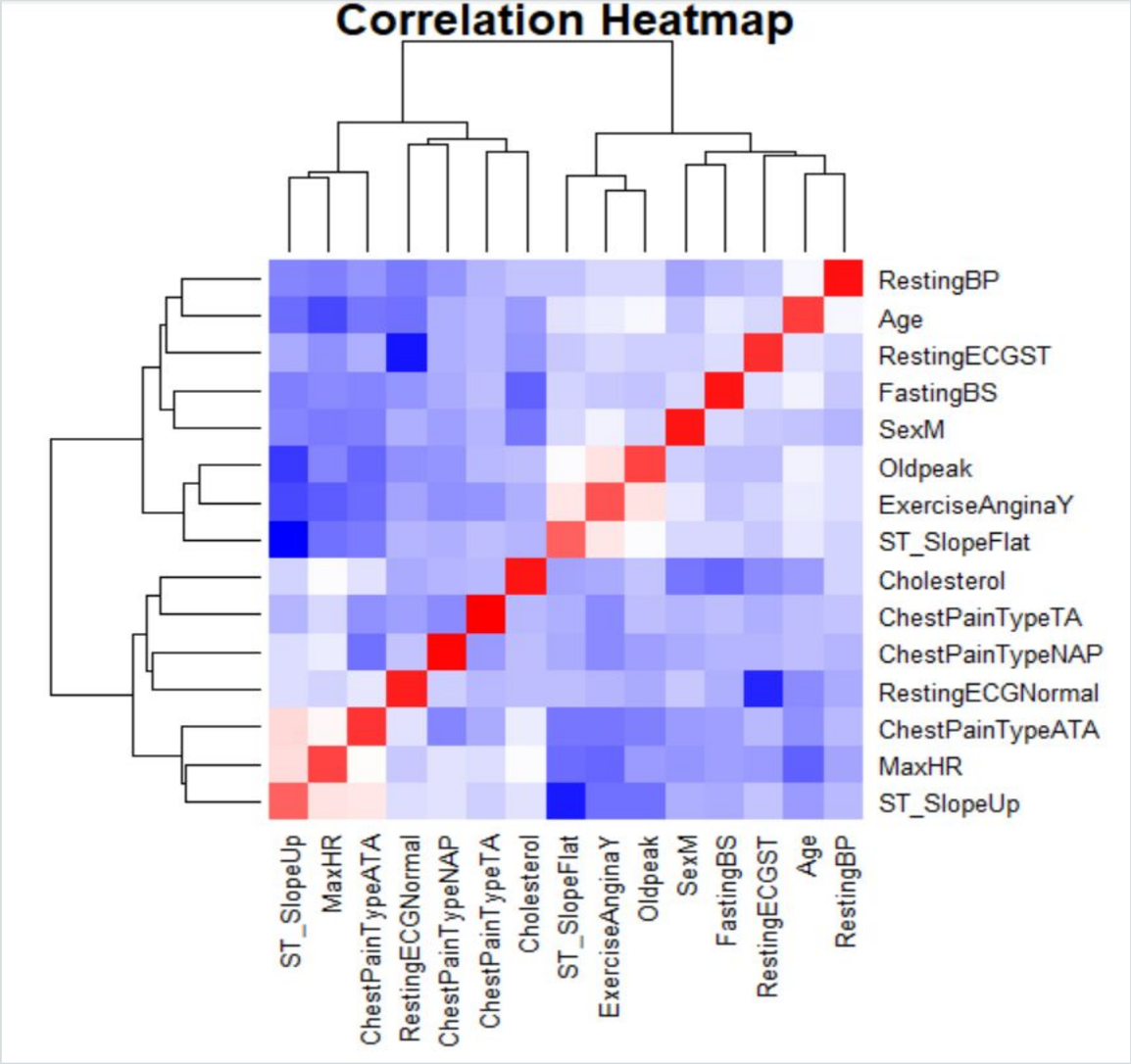
nzv: This is the overall result combining the above metrics. If any of the above checks identify a variable as having near-zero variance, it would be listed here. However, in your case, there are no categorical variables identified as having near-zero variance.

Creating Dummies

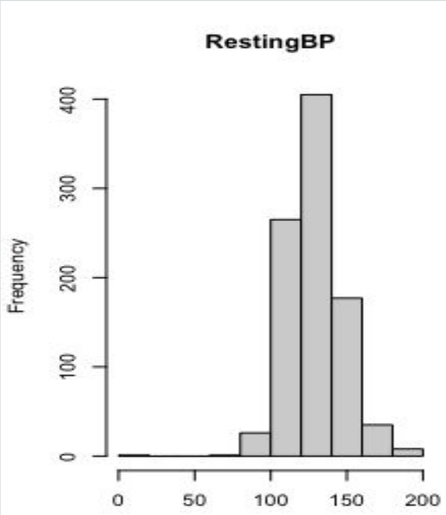
```
> head(heart)
  SexM ChestPainTypeATA ChestPainTypeNAP ChestPainTypeTA RestingECGNormal RestingECGST
1    1                1                0                0                 1            0
2    0                0                1                0                 1            0
3    1                1                0                0                 0            1
4    0                0                0                0                 1            0
5    1                0                1                0                 1            0
6    1                0                1                0                 1            0
  ExerciseAnginaY ST_SlopeFlat ST_SlopeUp Age RestingBP Cholesterol FastingBS MaxHR Oldpeak
1               0             0          1  40       140         289         0   172     0.0
2               0             1          0  49       160         180         0   156     1.0
3               0             0          1  37       130         283         0    98     0.0
4               1             1          0  48       138         214         0   108     1.5
5               0             0          1  54       150         195         0   122     0.0
6               0             0          1  39       120         339         0   170     0.0
> dim(heart)
[1] 918  15
```

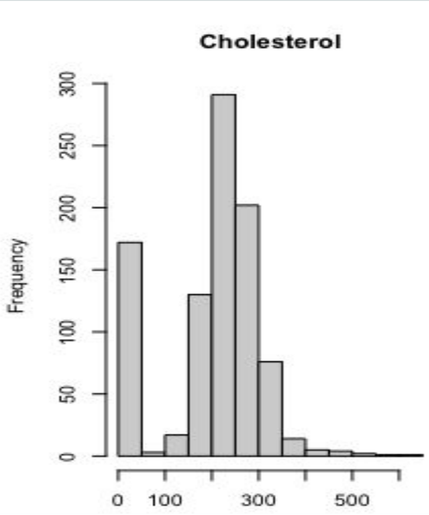
Correlation Plot/ Heat Map

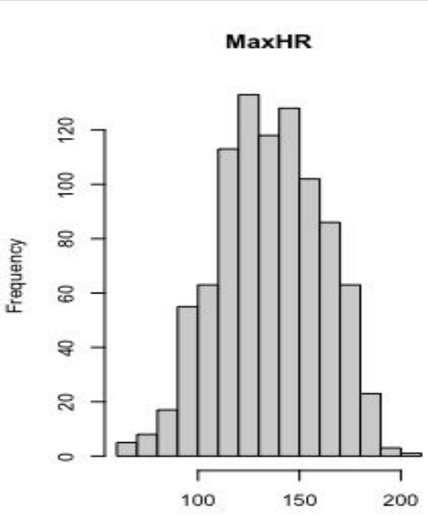


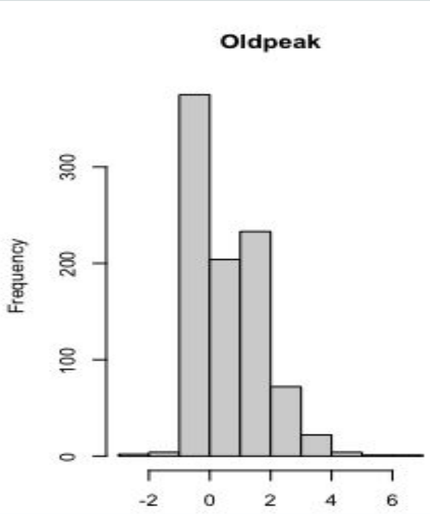


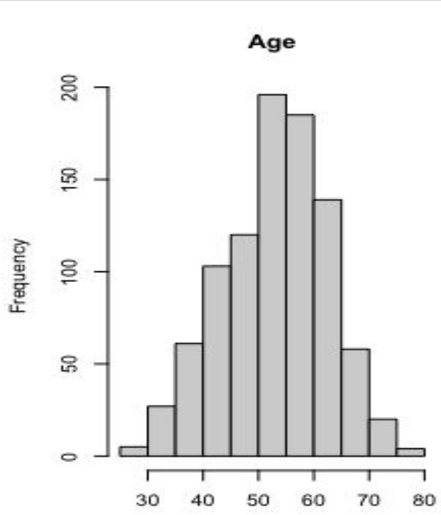
Checking for Transformations using Histograms and Boxplots



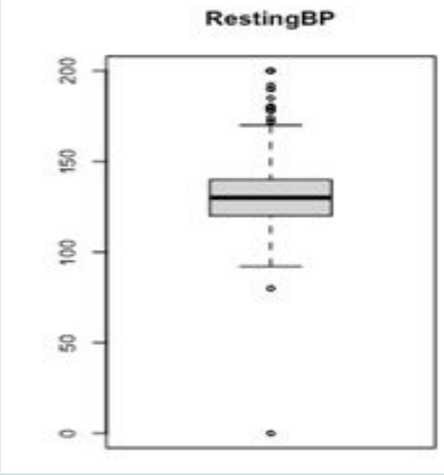


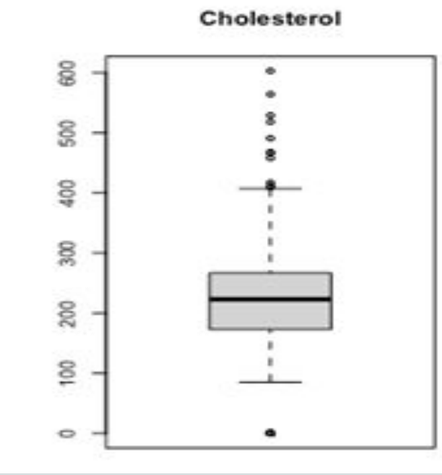


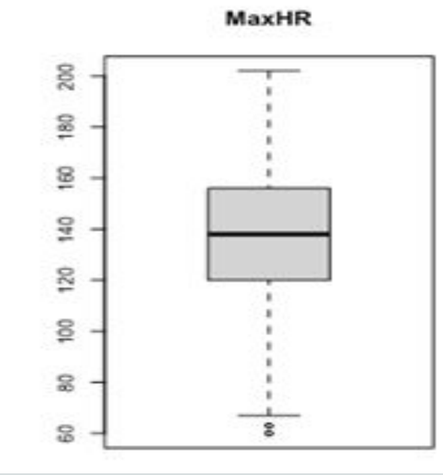


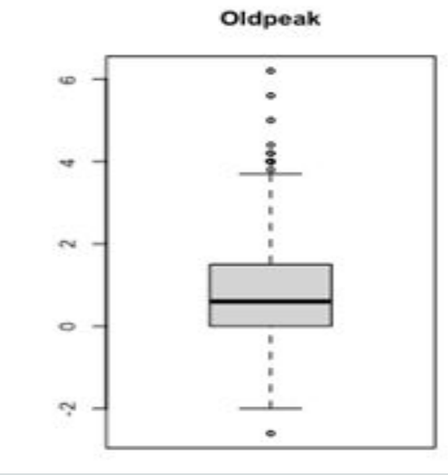


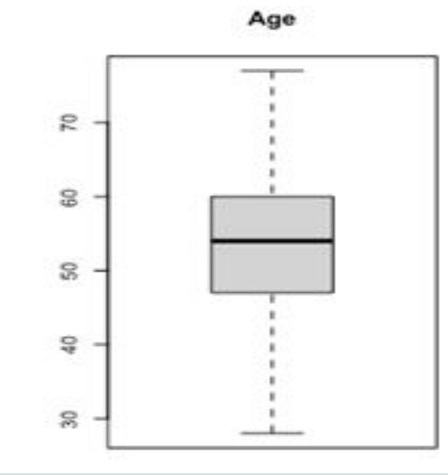
Skew:	0.1792520	-0.6080941	-0.1438880	1.0195317	-0.1952932





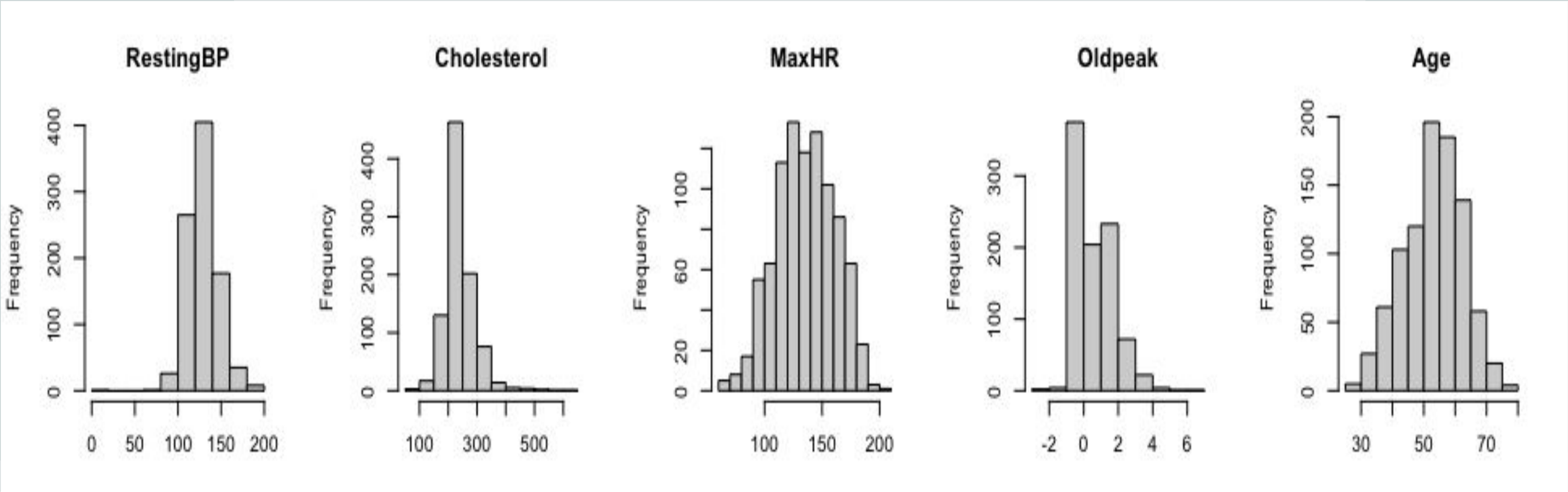






Mean Imputation on Cholesterol

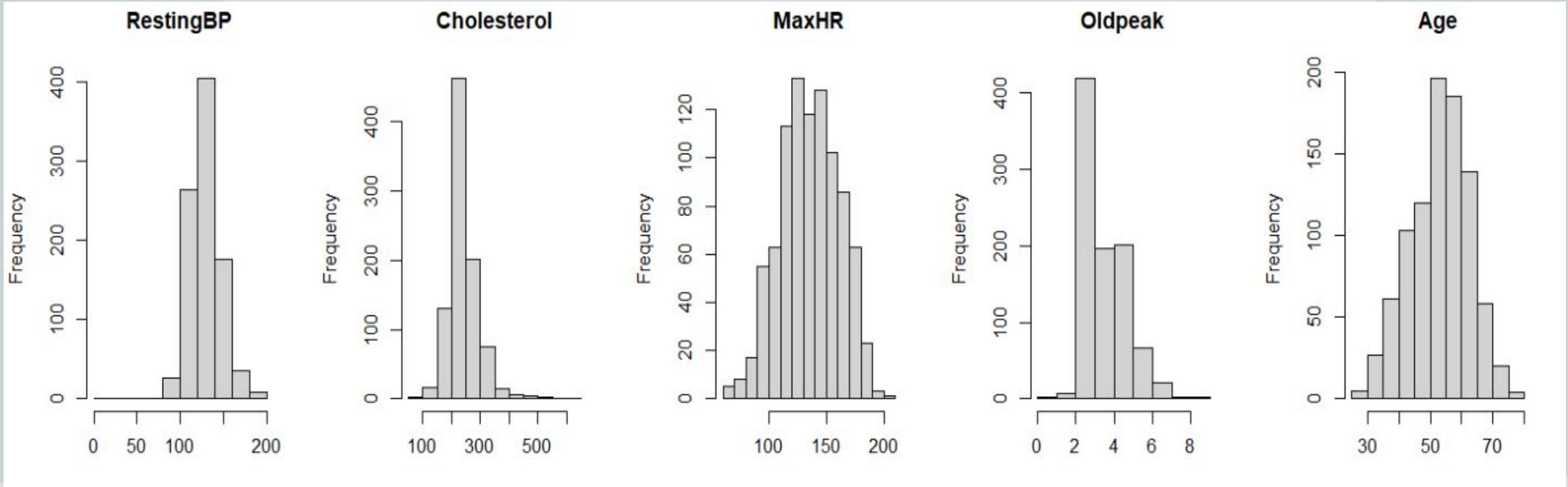
- It is impossible for a person to have zero cholesterol. Even if a person doesn't take it through the diet, it is automatically produced by the body.
- So we've used mean imputation to replace all the zeros in the Cholesterol variable with the mean of that variable.
- Then we tried using centre and scale, Boxcox



Skew	0.1792520	1.3711514	-0.1438880	1.0195317	-0.1952932

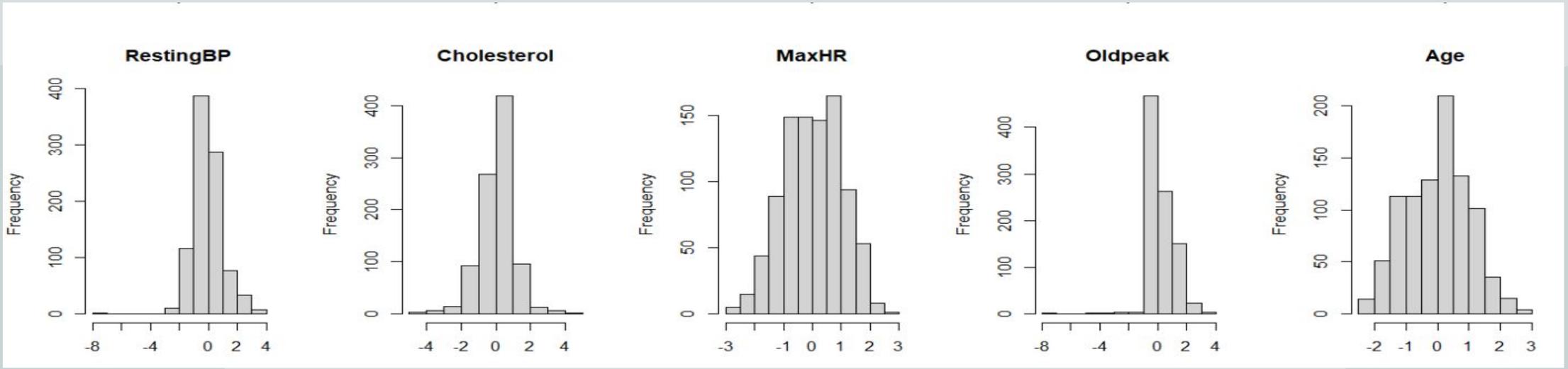
Addition of a constant to Oldpeak Variable

- Oldpeak column has 368 zero values and also few negative values. It is possible for Oldpeak variable to have zeros.
- But as it is heavily right skewed we need to perform transformations to bring it to normal distribution.
- So we added a constant of 2.7 to each value of Oldpeak Variable.
- Now we can do Boxcox transformation.

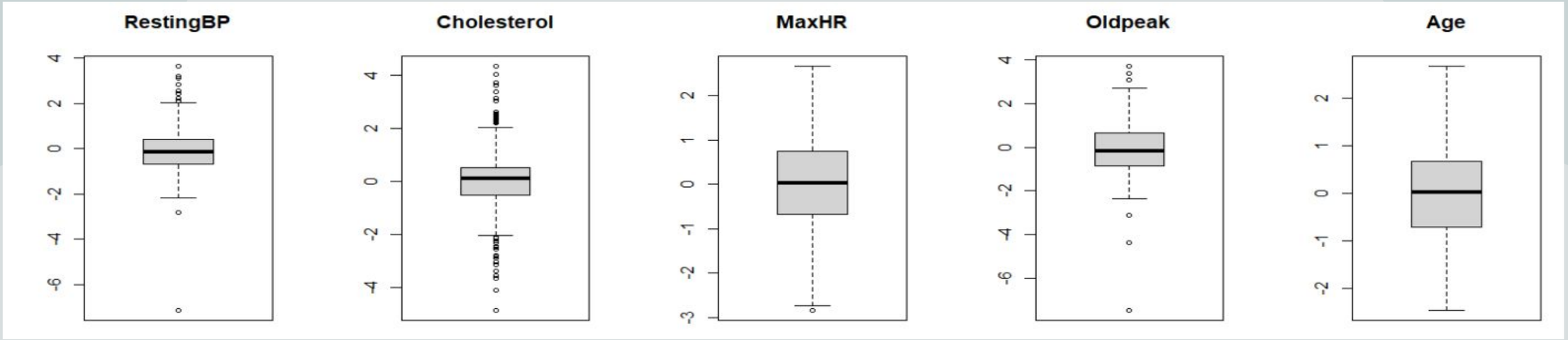


Skew	0.1795453	1.3711514	-0.1441234	1.0211999	-0.1956127

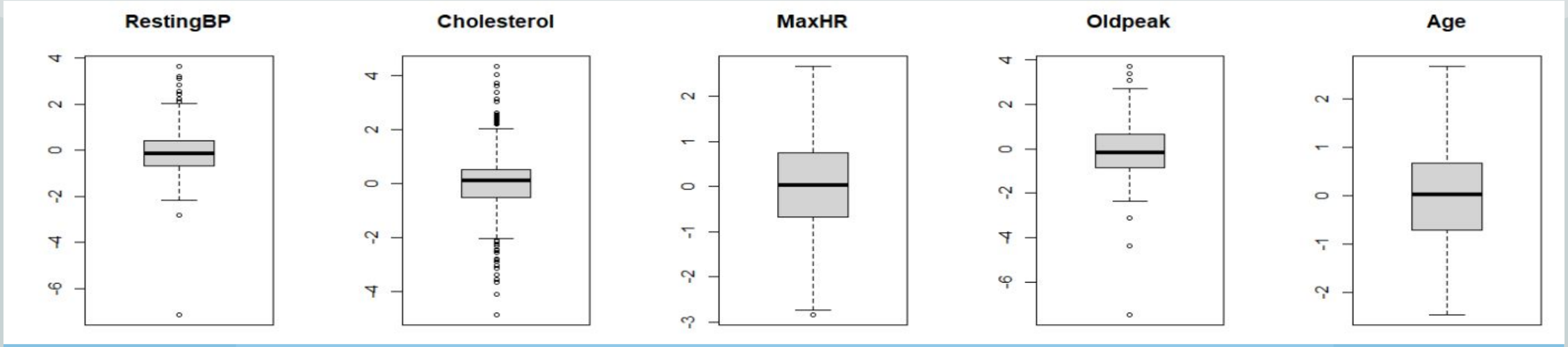
Center & Scale, BoxCox Transformation



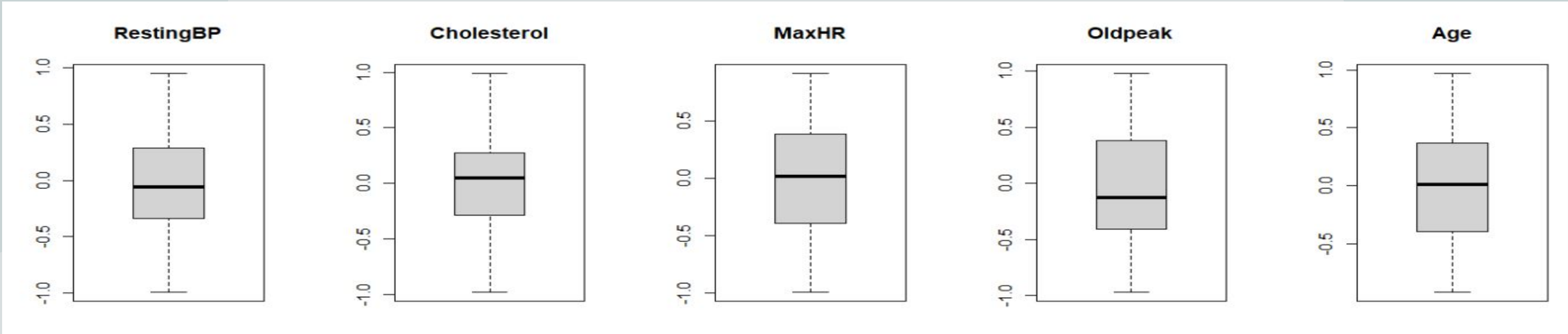
	RestingBP	Cholesterol	MaxHR	Oldpeak	Age
Skewness	0.17954532	-0.17759682	-0.05508755	0.06651038	-0.02521824
BoxCox Lambda		0.1	1.2	0.3	1.4



Before & After Spatial Sign for Outliers

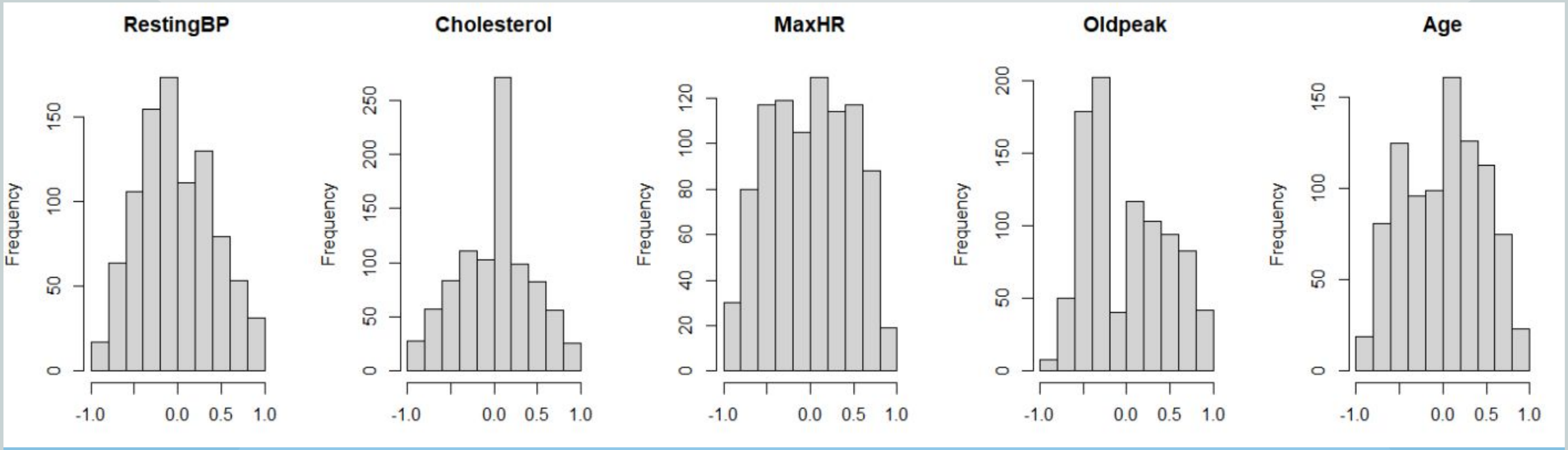


Skew:	0.17925203	-0.17759682	-0.05508755	0.06651038	-0.02521824



Skew:	0.16871322	-0.06582053	-0.02827240	0.31156797	-0.02879606

Histograms after Spatial Sign



Skew:	0.16871322	-0.06582053	-0.02827240	0.31156797	-0.02879606

- Now all our data is transformed to normal distribution

Predictors & Sample Size after Pre-Processing

- 1 Response Variable
- 15 Predictors

- 918 Observations
  - 735 training observations
  - 183 testing observations

```
> names(heart)
 [1] "SexM"
 [2] "ChestPainTypeATA"
 [3] "ChestPainTypeNAP"
 [4] "ChestPainTypeTA"
 [5] "RestingECGNormal"
 [6] "RestingECGST"
 [7] "ExerciseAnginaY"
 [8] "ST_SlopeFlat"
 [9] "ST_SlopeUp"
[10] "FastingBS"
[11] "RestingBP"
[12] "Cholesterol"
[13] "MaxHR"
[14] "Oldpeak"
[15] "Age"
```

Data Splitting

Stratified sampling split

80%

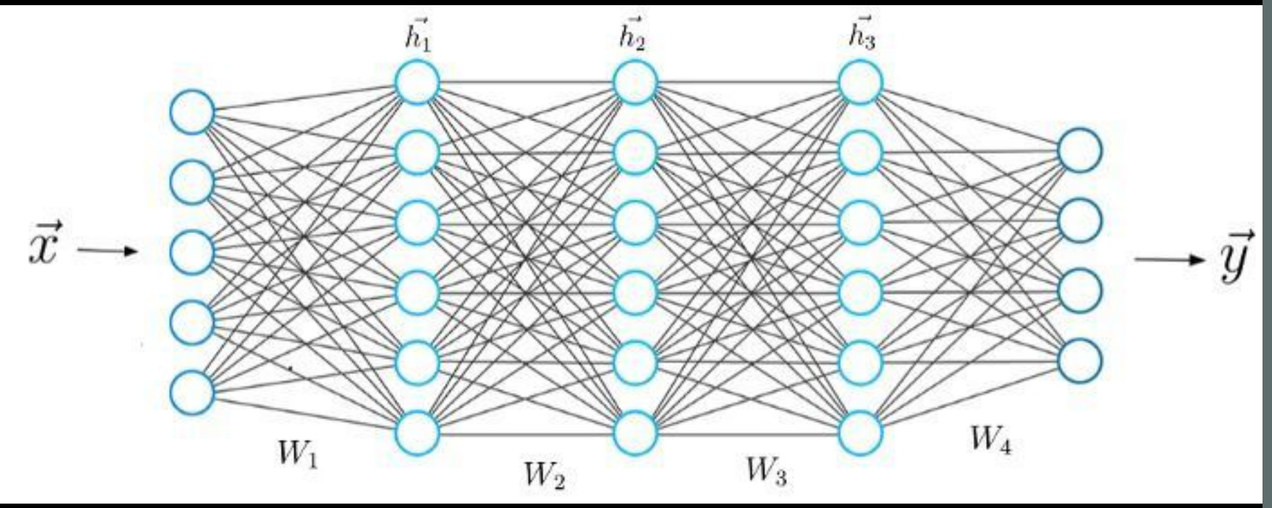
Training

20%

Testing

Resampling

10 - fold cross validation



03 Models

How neural network models in Machine Learning work?

Classification Models Performed

Linear Models

- Logistic Regression
- Linear Discriminant Analysis (LDA)
- Partial Least Squares Discriminant Analysis (PLSDA)
- Penalized Model
- Nearest Shrunken Centroids

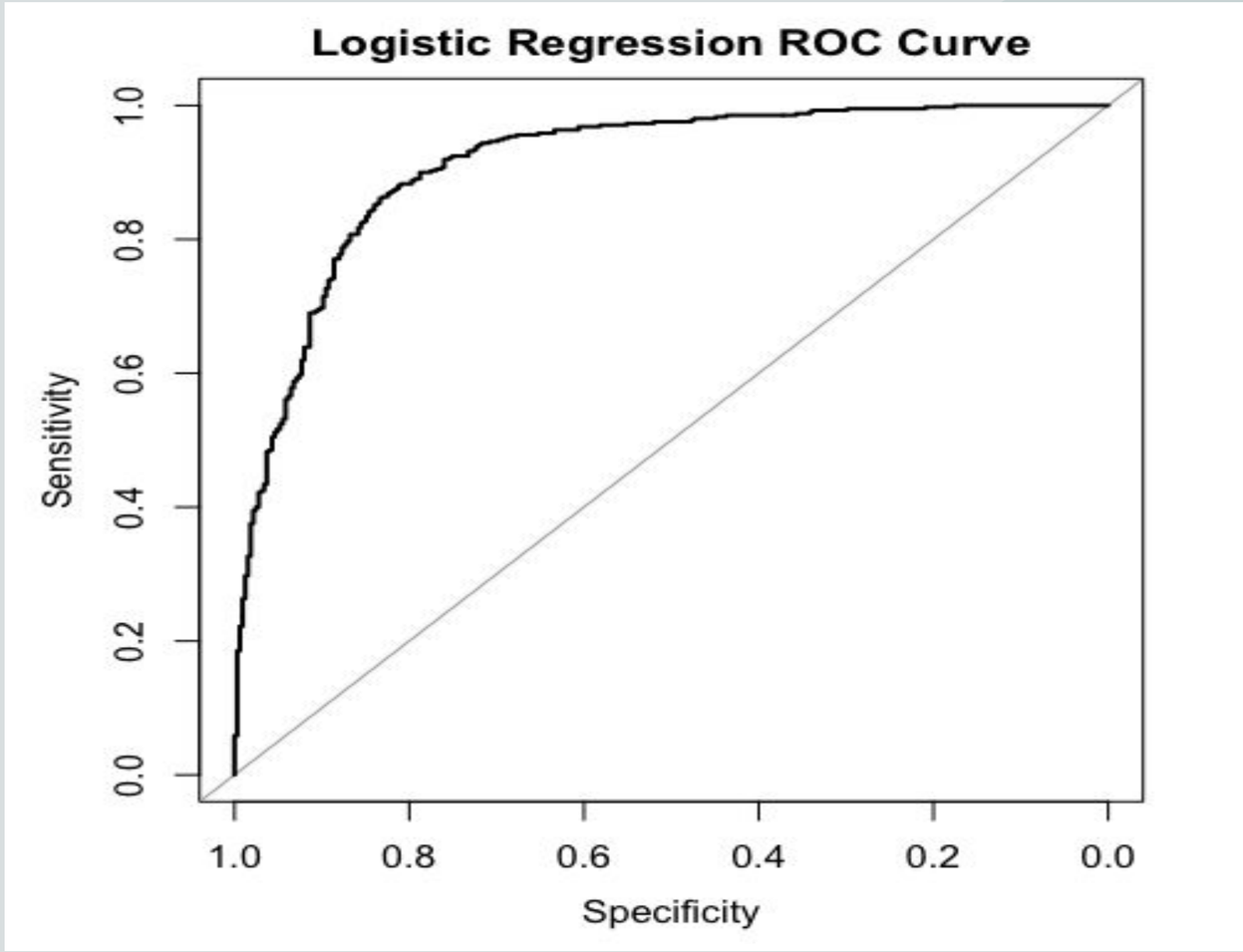
Nonlinear Models

- MARS
- Nonlinear Discriminant Analysis
- Neural Networks
- Flexible Discriminant Analysis
- Support Vector Machines (SVM)
- K - Nearest Neighbors (KNN)
- Naive Bayes

Linear Models

Logistic Regression

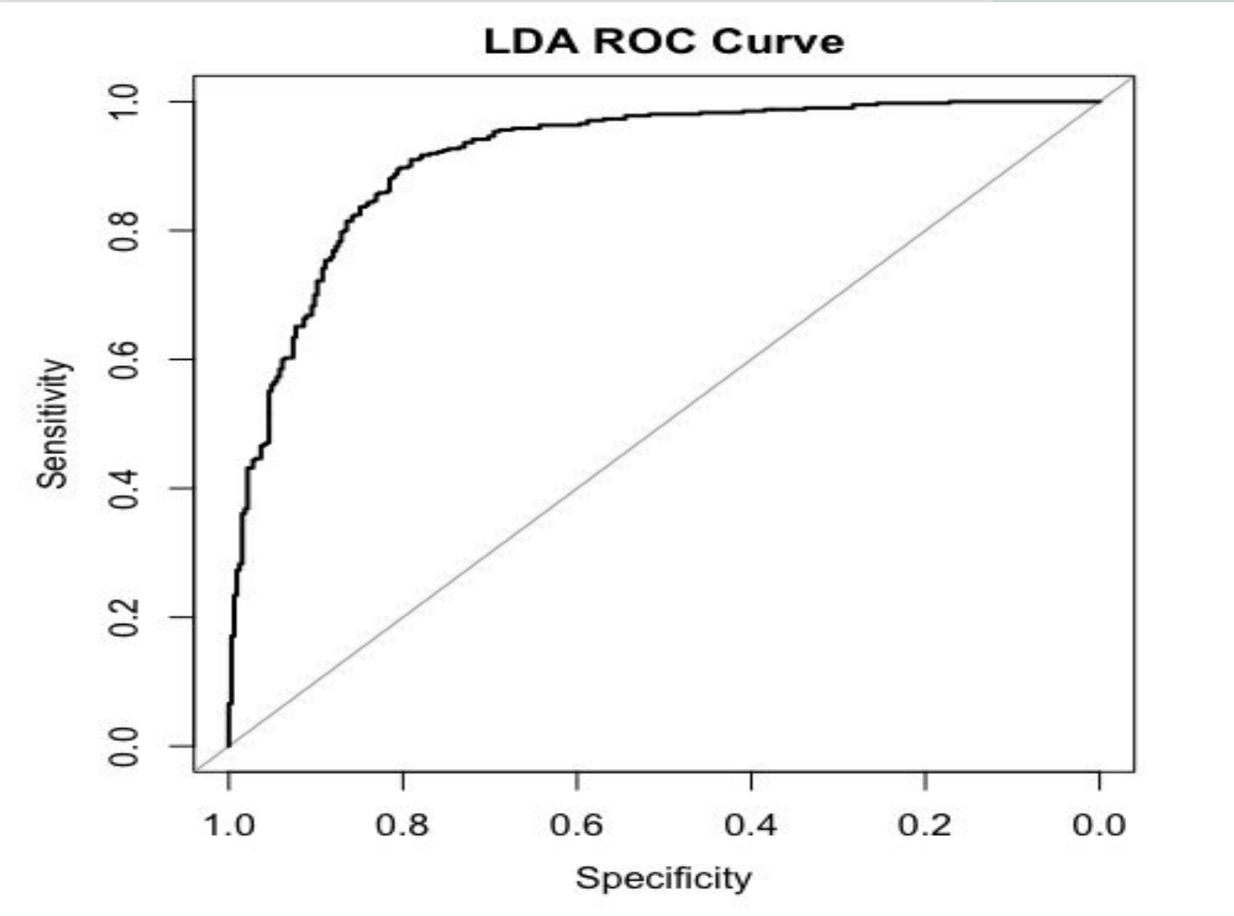
```
> modelLR
Generalized Linear Model

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results:

  ROC        Sens    Spec
  0.9168953  0.8125  0.8756098
```



Area Under the Curve: **0.9168953**

Linear Discriminant Analysis (LDA)

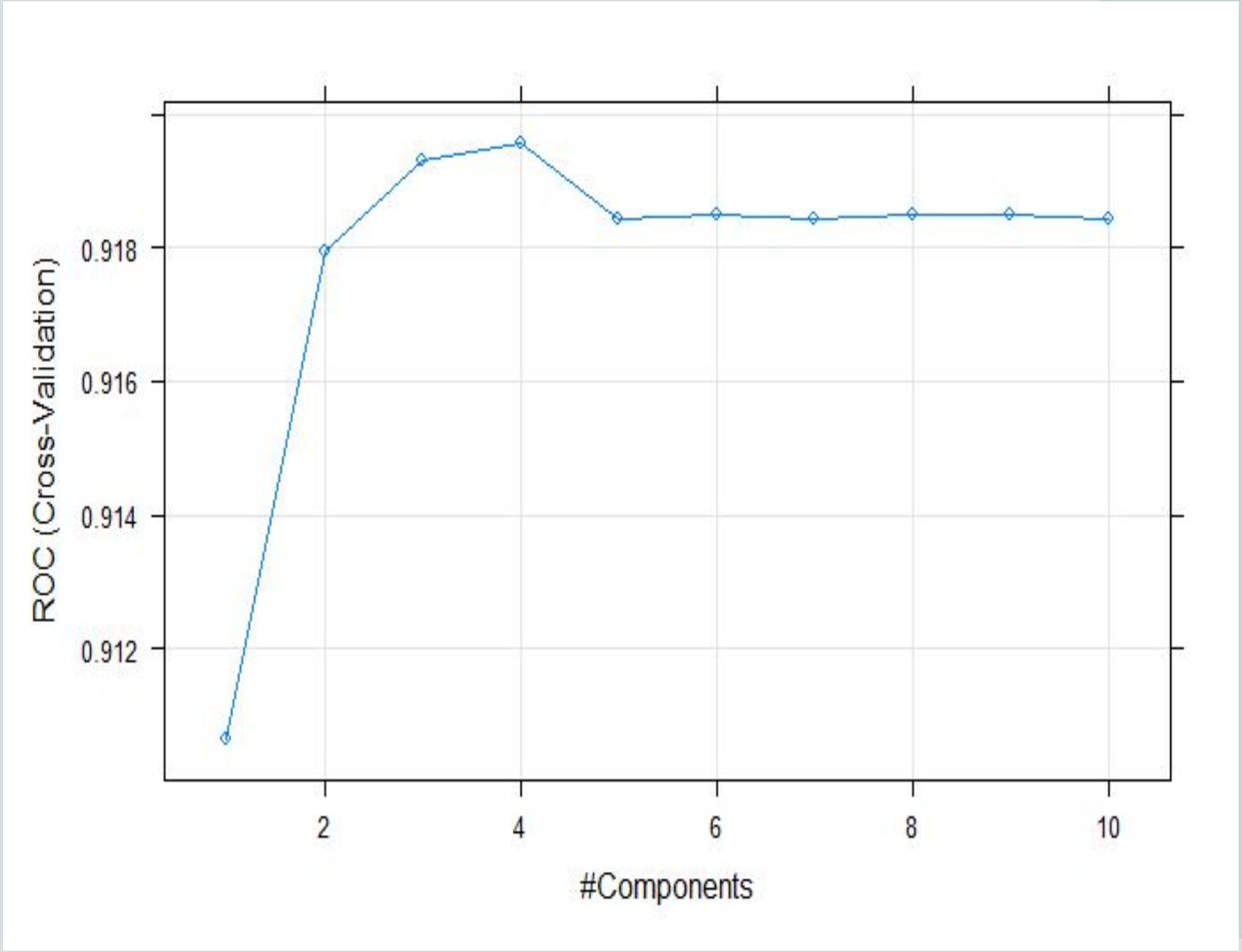
```
> modelLDA
Linear Discriminant Analysis

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results:

  ROC        Sens       Spec
  0.918429   0.8094697  0.8829268
```



Area Under the Curve: **0.918429**

Partial Least Squares Discriminant Analysis (PLSDA)

```
> plsFit2
Partial Least Squares

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:

  ncomp  ROC        Sens       Spec
   1     0.9106523  0.7910038  0.8829268
   2     0.9179601  0.8125000  0.8853659
   3     0.9193251  0.8064394  0.8780488
   4     0.9195607  0.8156250  0.8829268
   5     0.9184266  0.8125000  0.8829268
   6     0.9185029  0.8125000  0.8829268
   7     0.9184243  0.8125000  0.8829268
   8     0.9185029  0.8094697  0.8829268
   9     0.9185029  0.8094697  0.8829268
  10     0.9184290  0.8094697  0.8829268

ROC was used to select the optimal model using the largest value.
The final value used for the model was ncomp = 4.
```

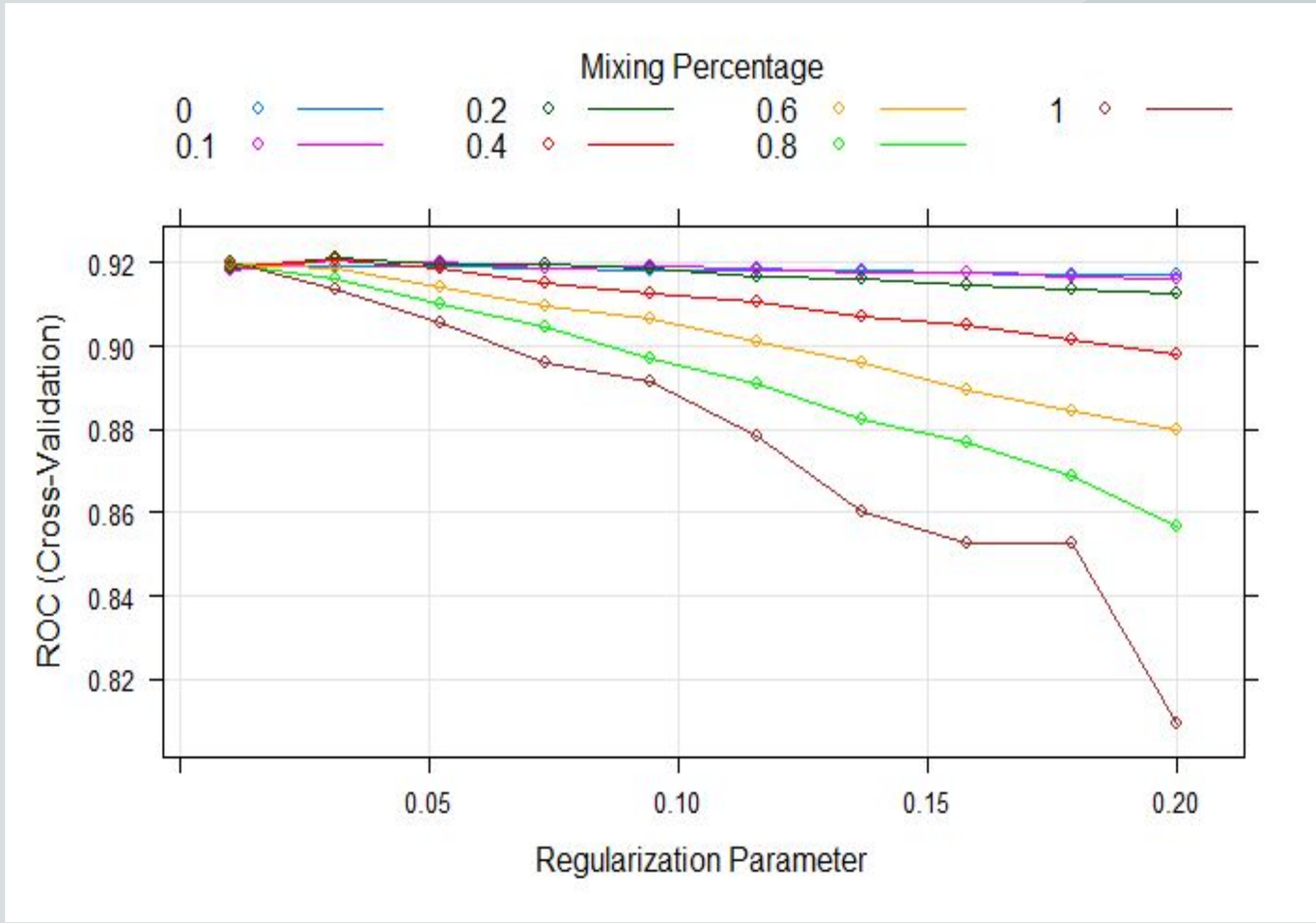


Area Under the Curve: **0.9195607**

Penalized Models

```
> glmnTuned
glmnet

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:
```

```
ROC was used to select the optimal model using the largest value.
The final values used for the model were alpha = 0.2 and lambda = 0.03111111.
```



Area Under the Curve: **0.9209765**

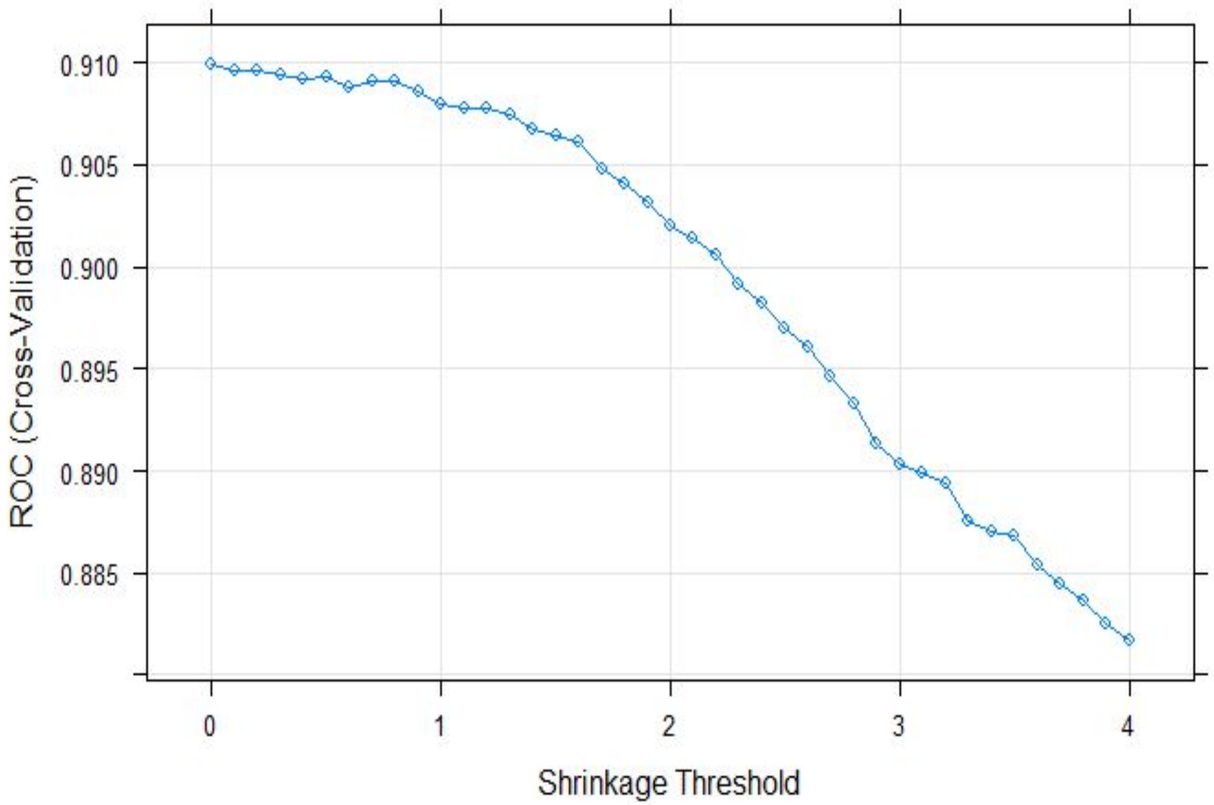
Penalized Models Output

alpha	lambda	ROC	Sens	Spec
0.0	0.01000000	0.9193159	0.8001894	0.8829268
0.0	0.03111111	0.9193921	0.8001894	0.8829268
0.0	0.05222222	0.9191727	0.7971591	0.8829268
0.0	0.07333333	0.9185722	0.7941288	0.8878049
0.0	0.09444444	0.9180340	0.7972538	0.8878049
0.0	0.11555556	0.9179462	0.7972538	0.8878049
0.0	0.13666667	0.9180155	0.7972538	0.8975610
0.0	0.15777778	0.9177892	0.7941288	0.9024390
0.0	0.17888889	0.9173342	0.7910038	0.9024390
0.0	0.20000000	0.9172579	0.7910038	0.9024390
0.1	0.01000000	0.9183273	0.8094697	0.8731707
0.1	0.03111111	0.9208957	0.8032197	0.8853659
0.1	0.05222222	0.9202213	0.8002841	0.8853659
0.1	0.07333333	0.9186207	0.7972538	0.8926829
0.1	0.09444444	0.9189948	0.7972538	0.8951220
0.1	0.11555556	0.9186091	0.7910038	0.9000000
0.1	0.13666667	0.9176367	0.7910038	0.8975610
0.1	0.15777778	0.9174127	0.7940341	0.9000000
0.1	0.17888889	0.9165188	0.7940341	0.8975610
0.1	0.20000000	0.9160731	0.7909091	0.8975610
0.2	0.01000000	0.9187754	0.8094697	0.8731707
0.2	0.03111111	0.9209765	0.8032197	0.8853659
0.2	0.05222222	0.9197524	0.8001894	0.8878049
0.2	0.07333333	0.9196831	0.8001894	0.8975610
0.2	0.09444444	0.9186946	0.7971591	0.8975610
0.2	0.11555556	0.9165235	0.7940341	0.9024390
0.2	0.13666667	0.9161539	0.7970644	0.9000000
0.2	0.15777778	0.9147312	0.7940341	0.8975610
0.2	0.17888889	0.9138350	0.7940341	0.8951220
0.2	0.20000000	0.9124838	0.7847538	0.8951220
0.4	0.01000000	0.9191542	0.8093750	0.8731707
0.4	0.03111111	0.9207409	0.8062500	0.8878049
0.4	0.05222222	0.9187015	0.8093750	0.8902439
0.4	0.07333333	0.9149390	0.8124053	0.9000000
0.4	0.09444444	0.9128557	0.8032197	0.8926829
0.4	0.11555556	0.9103728	0.7939394	0.8902439
0.4	0.13666667	0.9072247	0.7877841	0.8902439
0.4	0.15777778	0.9052291	0.7692235	0.8902439
0.4	0.17888889	0.9016399	0.7538826	0.8878049
0.4	0.20000000	0.8978220	0.7537879	0.8853659
0.6	0.01000000	0.9194637	0.8062500	0.8731707
0.6	0.03111111	0.9186992	0.8062500	0.8878049
0.6	0.05222222	0.9140336	0.8124053	0.8951220
0.6	0.07333333	0.9097676	0.8092803	0.8902439
0.6	0.09444444	0.9065479	0.7877841	0.8853659
0.6	0.11555556	0.9012518	0.7569129	0.8804878
0.6	0.13666667	0.8961013	0.7537879	0.8780488
0.6	0.15777778	0.8892900	0.7538826	0.8731707
0.6	0.17888889	0.8845505	0.7600379	0.8634146
0.6	0.20000000	0.8798434	0.7631629	0.8439024
0.8	0.01000000	0.9198356	0.8062500	0.8780488
0.8	0.03111111	0.9159807	0.8062500	0.8926829
0.8	0.05222222	0.9101210	0.8093750	0.8853659
0.8	0.07333333	0.9045986	0.7846591	0.8804878
0.8	0.09444444	0.8971914	0.7599432	0.8780488
0.8	0.11555556	0.8907959	0.7600379	0.8707317
0.8	0.13666667	0.8825504	0.7631629	0.8439024
0.8	0.15777778	0.8767611	0.7631629	0.8414634
0.8	0.17888889	0.8689579	0.7631629	0.8414634
0.8	0.20000000	0.8565849	0.7631629	0.8414634
1.0	0.01000000	0.9202120	0.8062500	0.8780488
1.0	0.03111111	0.9135047	0.8093750	0.8878049
1.0	0.05222222	0.9056241	0.8000000	0.8756098
1.0	0.07333333	0.8960574	0.7569129	0.8756098
1.0	0.09444444	0.8914900	0.7631629	0.8439024
1.0	0.11555556	0.8785315	0.7631629	0.8414634
1.0	0.13666667	0.8601869	0.7631629	0.8414634
1.0	0.15777778	0.8528097	0.7631629	0.8414634
1.0	0.17888889	0.8526250	0.7631629	0.8414634
1.0	0.20000000	0.8091590	0.7631629	0.8414634

Nearest Shrunken Centroids

```
> nscTuned
Nearest Shrunken Centroids

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:
```

```
ROC was used to select the optimal model using the largest value.
The final value used for the model was threshold = 0.
```



Area Under the Curve: **0.9098901**

Nearest Shrunken Centroids Output

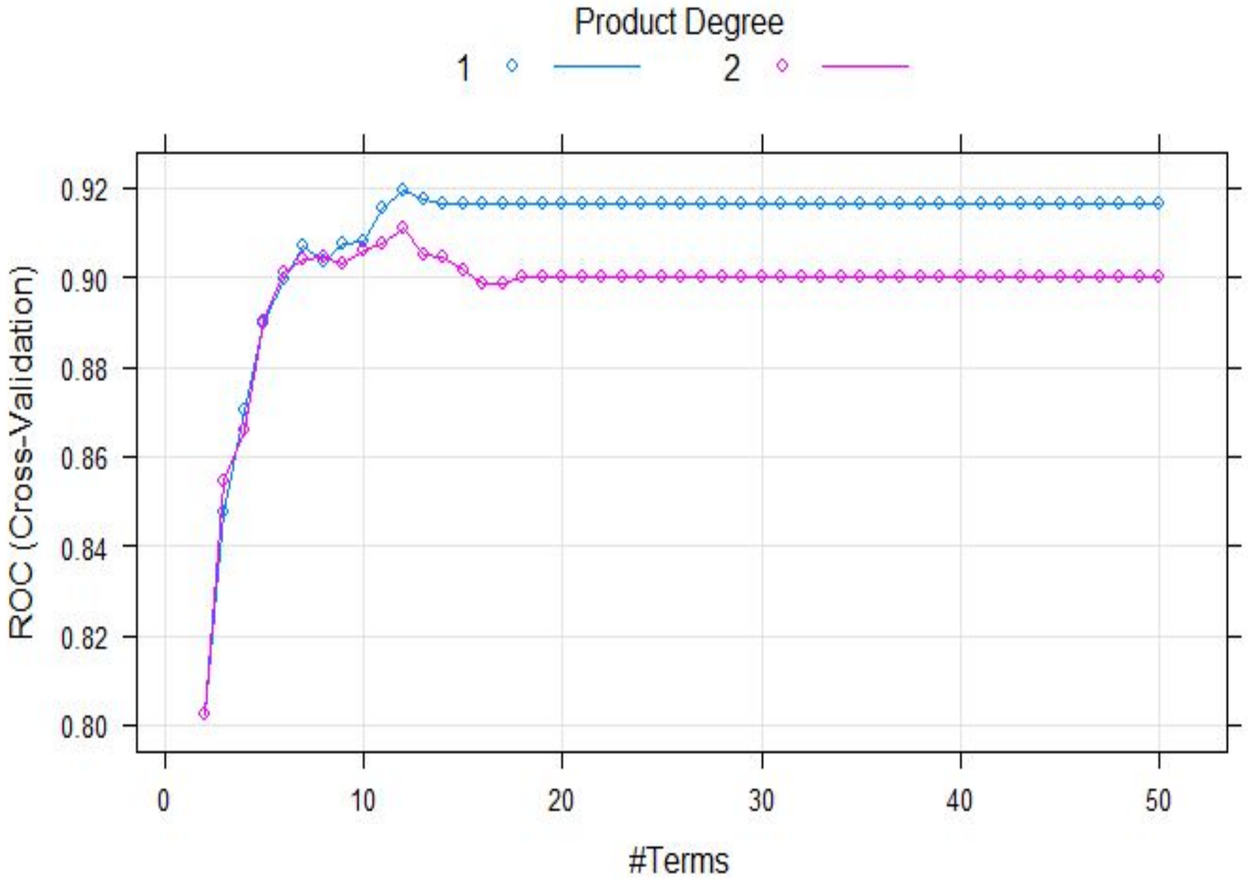
threshold	ROC	Sens	Spec
0.0	0.9098901	0.7693182	0.8902439
0.1	0.9096614	0.7693182	0.8902439
0.2	0.9095875	0.7693182	0.8902439
0.3	0.9094374	0.7723485	0.8902439
0.4	0.9092110	0.7723485	0.8951220
0.5	0.9092826	0.7753788	0.8951220
0.6	0.9087606	0.7723485	0.8951220
0.7	0.9090609	0.7723485	0.8951220
0.8	0.9090655	0.7723485	0.8951220
0.9	0.9086220	0.7753788	0.8951220
1.0	0.9079522	0.7753788	0.8951220
1.1	0.9078113	0.7753788	0.8951220
1.2	0.9078160	0.7722538	0.8975610
1.3	0.9074464	0.7691288	0.8975610
1.4	0.9066935	0.7691288	0.8975610
1.5	0.9063955	0.7692235	0.8951220
1.6	0.9061738	0.7692235	0.8951220
1.7	0.9047533	0.7661932	0.8951220
1.8	0.9040858	0.7599432	0.8926829
1.9	0.9031135	0.7599432	0.8926829
2.0	0.9020649	0.7569129	0.8902439
2.1	0.9013950	0.7569129	0.8902439
2.2	0.9005774	0.7569129	0.8902439
2.3	0.8991547	0.7537879	0.8902439
2.4	0.8982608	0.7537879	0.8902439
2.5	0.8969836	0.7537879	0.8878049
2.6	0.8960204	0.7538826	0.8878049
2.7	0.8946762	0.7538826	0.8878049
2.8	0.8932603	0.7538826	0.8878049
2.9	0.8913872	0.7538826	0.8853659
3.0	0.8903501	0.7538826	0.8853659
3.1	0.8898998	0.7507576	0.8853659
3.2	0.8893708	0.7507576	0.8853659
3.3	0.8875785	0.7507576	0.8853659
3.4	0.8870496	0.7507576	0.8878049
3.5	0.8868279	0.7507576	0.8878049
3.6	0.8853312	0.7507576	0.8878049
3.7	0.8844258	0.7507576	0.8878049
3.8	0.8836775	0.7507576	0.8878049
3.9	0.8824788	0.7507576	0.8878049
4.0	0.8817235	0.7507576	0.8853659

Nonlinear Models

Multivariate Adaptive Regression Splines (MARS)

```
> fdaTuned
Flexible Discriminant Analysis

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:
```

```
ROC was used to select the optimal model using the largest value.
The final values used for the model were nprune = 12 and degree = 1.
```



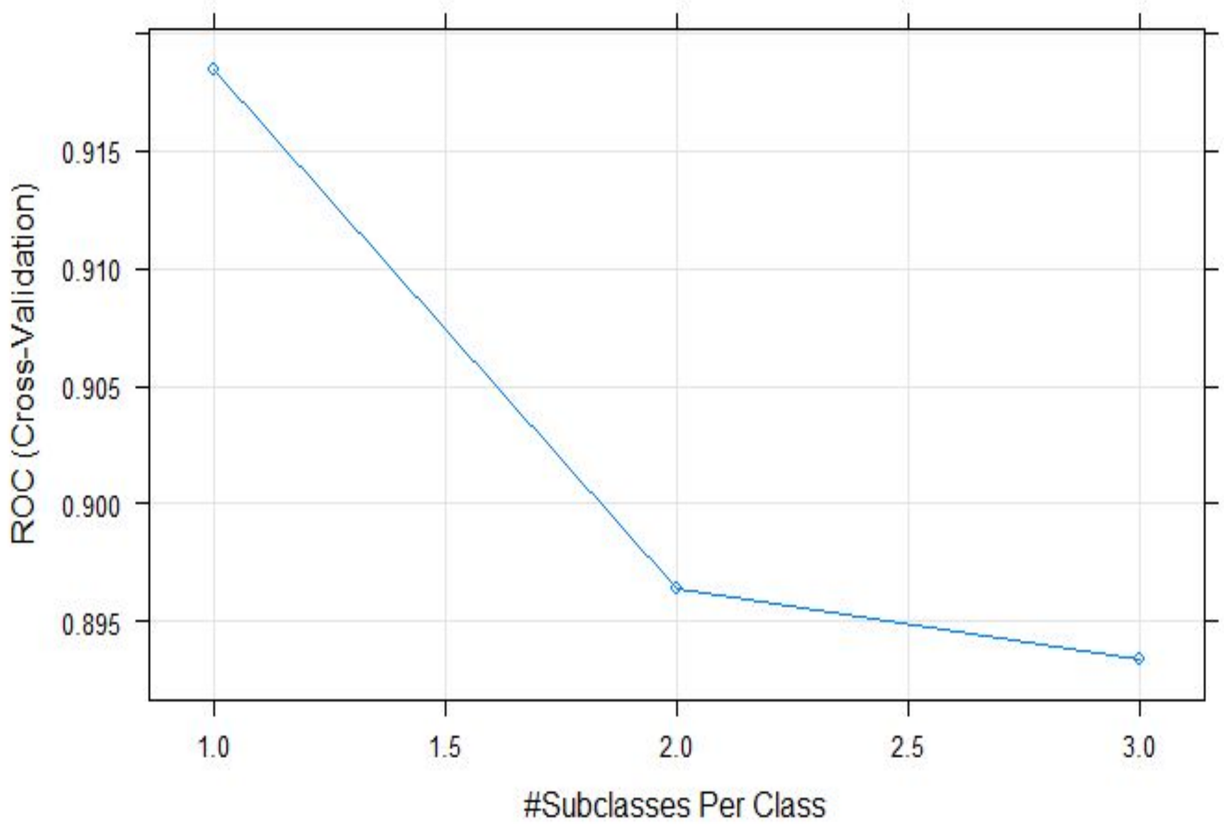
Area Under the Curve: **0.9196727**

MARS Output

degree	nprune	ROC	Sens	Spec
1	2	0.8023131	0.7631629	0.8414634
1	3	0.8477943	0.7412879	0.8609756
1	4	0.8704557	0.7875947	0.8219512
1	5	0.8895025	0.7196970	0.8804878
1	6	0.8997679	0.7877841	0.8609756
1	7	0.9068332	0.7785985	0.8926829
1	8	0.9037532	0.7781250	0.8609756
1	9	0.9075700	0.7908144	0.8731707
1	10	0.9078425	0.7846591	0.8707317
1	11	0.9156458	0.7970644	0.8853659
1	12	0.9196727	0.7970644	0.8878049
1	13	0.9175305	0.8032197	0.8829268
1	14	0.9163756	0.8001894	0.8853659
1	15	0.9163756	0.8001894	0.8853659
1	16	0.9163756	0.8001894	0.8853659
1	17	0.9163756	0.8001894	0.8853659
1	18	0.9163756	0.8001894	0.8853659
1	19	0.9163756	0.8001894	0.8853659
1	20	0.9163756	0.8001894	0.8853659
1	21	0.9163756	0.8001894	0.8853659
1	22	0.9163756	0.8001894	0.8853659
1	23	0.9163756	0.8001894	0.8853659
1	24	0.9163756	0.8001894	0.8853659
1	25	0.9163756	0.8001894	0.8853659
1	26	0.9163756	0.8001894	0.8853659
1	27	0.9163756	0.8001894	0.8853659
1	28	0.9163756	0.8001894	0.8853659
1	29	0.9163756	0.8001894	0.8853659
1	30	0.9163756	0.8001894	0.8853659
1	31	0.9163756	0.8001894	0.8853659
1	32	0.9163756	0.8001894	0.8853659
1	33	0.9163756	0.8001894	0.8853659
1	34	0.9163756	0.8001894	0.8853659
1	35	0.9163756	0.8001894	0.8853659
1	36	0.9163756	0.8001894	0.8853659
1	37	0.9163756	0.8001894	0.8853659
1	38	0.9163756	0.8001894	0.8853659
1	39	0.9163756	0.8001894	0.8853659
1	40	0.9163756	0.8001894	0.8853659
1	41	0.9163756	0.8001894	0.8853659
1	42	0.9163756	0.8001894	0.8853659
1	43	0.9163756	0.8001894	0.8853659
1	44	0.9163756	0.8001894	0.8853659
1	45	0.9163756	0.8001894	0.8853659
1	46	0.9163756	0.8001894	0.8853659
1	47	0.9163756	0.8001894	0.8853659
1	48	0.9163756	0.8001894	0.8853659
1	49	0.9163756	0.8001894	0.8853659
1	50	0.9163756	0.8001894	0.8853659
2	2	0.8023131	0.7631629	0.8414634
2	3	0.8547210	0.7418561	0.8463415
2	4	0.8659137	0.7599432	0.8536585
2	5	0.8902427	0.7785985	0.8634146
2	6	0.9009793	0.8000947	0.8682927
2	7	0.9041597	0.7876894	0.8804878
2	8	0.9044115	0.7910038	0.8804878
2	9	0.9033283	0.7849432	0.8804878
2	10	0.9059266	0.7970644	0.8682927
2	11	0.9073309	0.8062500	0.8707317
2	12	0.9109594	0.8000947	0.8780488
2	13	0.9051852	0.7939394	0.8609756
2	14	0.9045085	0.7909091	0.8634146
2	15	0.9014066	0.7940341	0.8658537
2	16	0.8984433	0.7970644	0.8634146
2	17	0.8985126	0.7970644	0.8585366
2	18	0.9002079	0.7970644	0.8609756
2	19	0.9002079	0.7970644	0.8609756
2	20	0.9002079	0.7970644	0.8609756
2	21	0.9002079	0.7970644	0.8609756
2	22	0.9002079	0.7970644	0.8609756
2	23	0.9002079	0.7970644	0.8609756
2	24	0.9002079	0.7970644	0.8609756
2	25	0.9002079	0.7970644	0.8609756
2	26	0.9002079	0.7970644	0.8609756
2	27	0.9002079	0.7970644	0.8609756
2	28	0.9002079	0.7970644	0.8609756
2	29	0.9002079	0.7970644	0.8609756
2	30	0.9002079	0.7970644	0.8609756
2	31	0.9002079	0.7970644	0.8609756
2	32	0.9002079	0.7970644	0.8609756
2	33	0.9002079	0.7970644	0.8609756
2	34	0.9002079	0.7970644	0.8609756
2	35	0.9002079	0.7970644	0.8609756
2	36	0.9002079	0.7970644	0.8609756
2	37	0.9002079	0.7970644	0.8609756
2	38	0.9002079	0.7970644	0.8609756
2	39	0.9002079	0.7970644	0.8609756
2	40	0.9002079	0.7970644	0.8609756
2	41	0.9002079	0.7970644	0.8609756
2	42	0.9002079	0.7970644	0.8609756
2	43	0.9002079	0.7970644	0.8609756
2	44	0.9002079	0.7970644	0.8609756
2	45	0.9002079	0.7970644	0.8609756
2	46	0.9002079	0.7970644	0.8609756
2	47	0.9002079	0.7970644	0.8609756
2	48	0.9002079	0.7970644	0.8609756
2	49	0.9002079	0.7970644	0.8609756
2	50	0.9002079	0.7970644	0.8609756

Nonlinear Discriminant Analysis

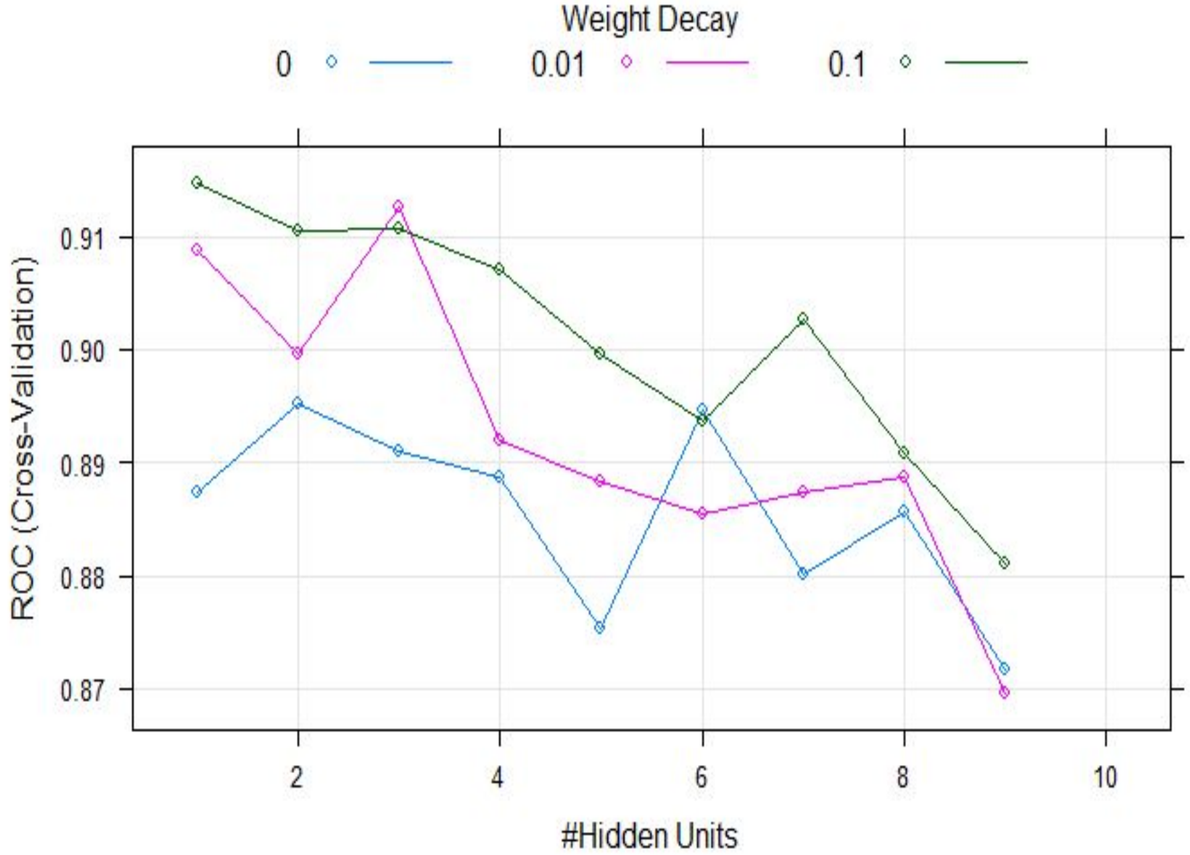
```
> mdaFit
Mixture Discriminant Analysis

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:

  subclasses  ROC        Sens       Spec
  1           0.9184290  0.8094697  0.8829268
  2           0.8963853  0.7972538  0.8658537
  3           0.8934012  0.8064394  0.8512195

ROC was used to select the optimal model using the largest value.
The final value used for the model was subclasses = 1.
```



Area Under the Curve: **0.9184290**

Neural Network

```
> nnetTune
Model Averaged Neural Network

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:

  decay  size  ROC        Sens       Spec
  0.00     1   0.8872967  0.7878788  0.8756098
  0.00     2   0.8952005  0.7970644  0.8780488
  0.00     3   0.8909160  0.7847538  0.8878049
  0.00     4   0.8887022  0.8183712  0.8780488
  0.00     5   0.8752425  0.7697917  0.8536585
  0.00     6   0.8946727  0.7908144  0.8463415
  0.00     7   0.8800259  0.7970644  0.8512195
  0.00     8   0.8857215  0.8398674  0.8390244
  0.00     9   0.8716972  0.7756629  0.8414634
  0.00    10         NaN        NaN        NaN
  0.01     1   0.9088715  0.7848485  0.8878049
  0.01     2   0.8995381  0.7724432  0.8585366
  0.01     3   0.9126547  0.8002841  0.8609756
  0.01     4   0.8918884  0.7849432  0.8487805
  0.01     5   0.8883176  0.7910038  0.8634146
  0.01     6   0.8854351  0.8185606  0.8536585
  0.01     7   0.8872783  0.7939394  0.8463415
  0.01     8   0.8887057  0.7912879  0.8585366
  0.01     9   0.8695376  0.7879735  0.8317073
```



AUC : 0.9147866

Neural Network Output

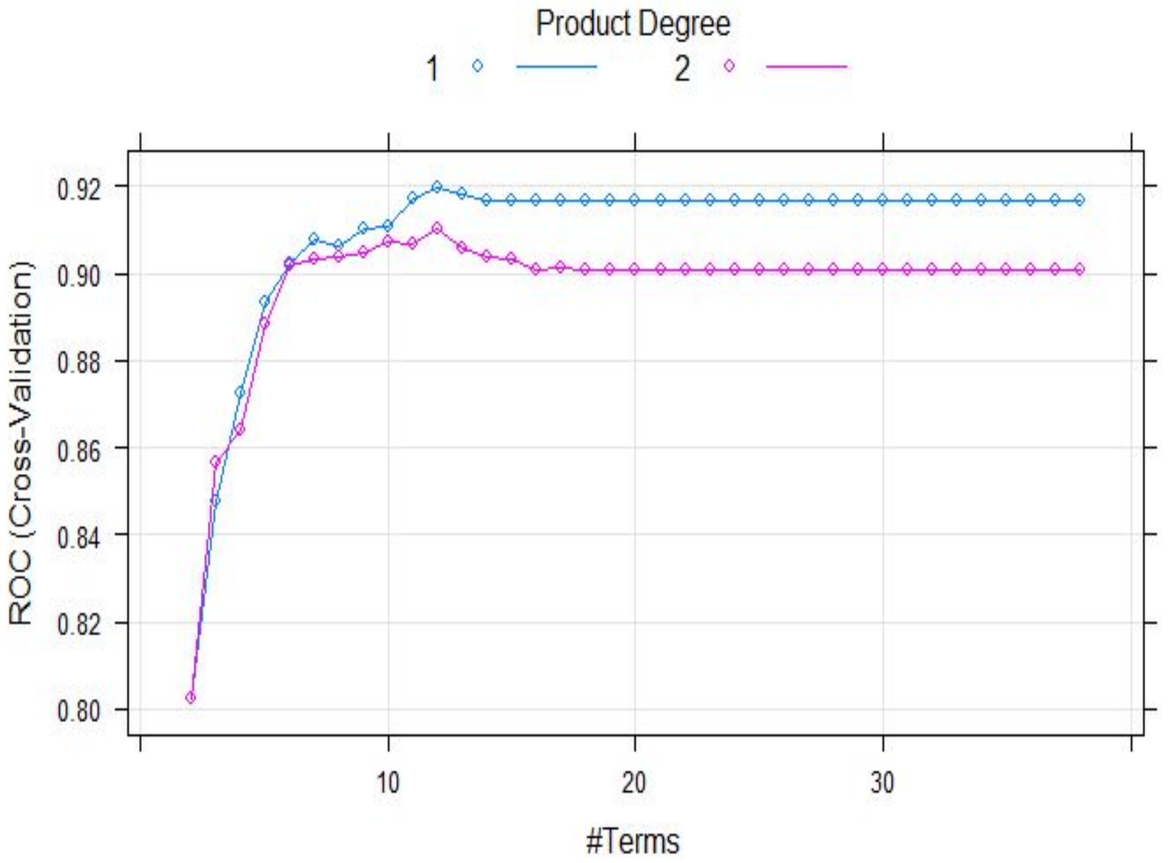
0.01	10	NaN	NaN	NaN
0.10	1	0.9147866	0.8186553	0.8902439
0.10	2	0.9105460	0.8093750	0.8804878
0.10	3	0.9106430	0.8034091	0.8560976
0.10	4	0.9071161	0.7819129	0.8780488
0.10	5	0.8995473	0.7972538	0.8682927
0.10	6	0.8936876	0.7910038	0.8536585
0.10	7	0.9026053	0.8219697	0.8634146
0.10	8	0.8908929	0.7969697	0.8609756
0.10	9	0.8810745	0.7692235	0.8487805
0.10	10	NaN	NaN	NaN

Tuning parameter 'bag' was held constant at a value of TRUE
ROC was used to select the optimal model using the largest value.
The final values used for the model were size = 1, decay = 0.1 and bag = TRUE.

Flexible Discriminant Analysis

```
> fdaTuned
Flexible Discriminant Analysis

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:
```

```
ROC was used to select the optimal model using the largest value.
The final values used for the model were degree = 1 and nprune = 12.
```



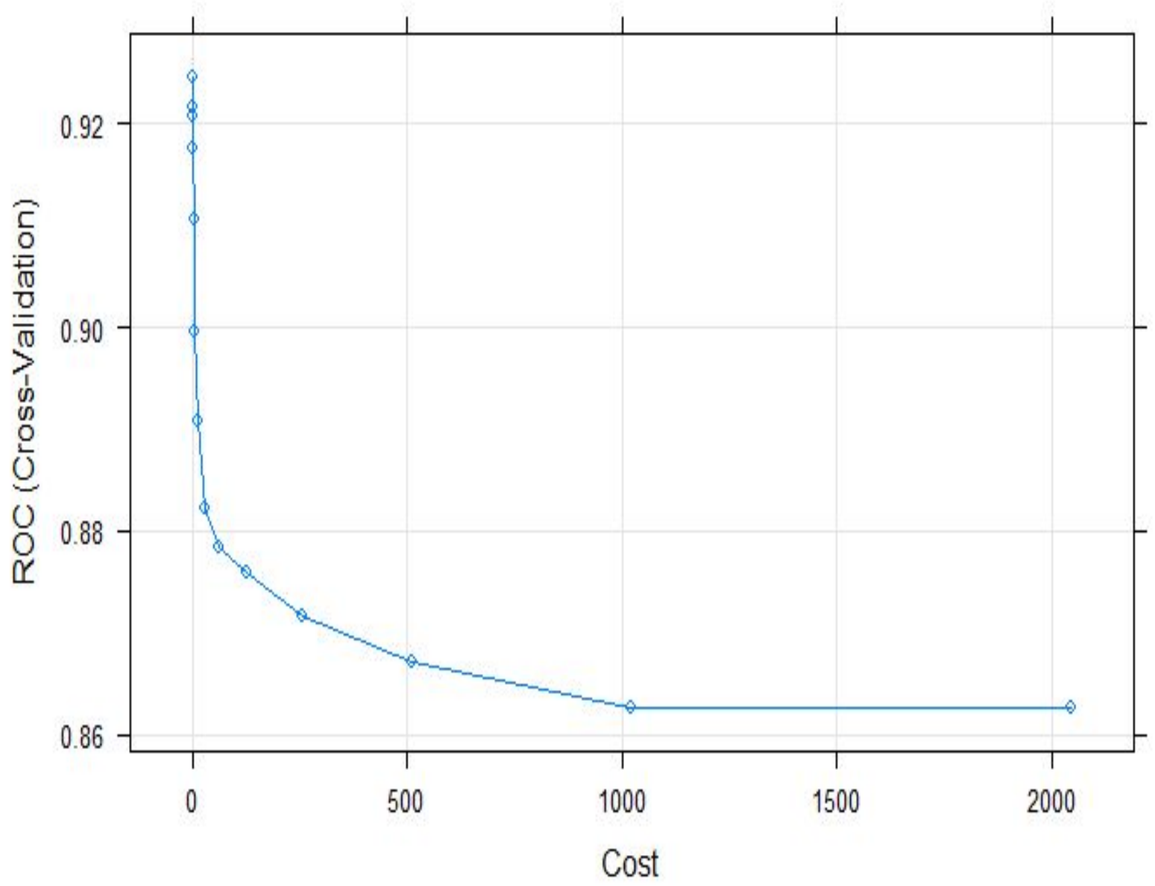
Area Under the Curve: **0.9199545**

Flexible Discriminant Analysis Output

degree	nprune	ROC	Sens	Spec
1	2	0.8023131	0.7631629	0.8414634
1	3	0.8477943	0.7631629	0.8414634
1	4	0.8724582	0.7785985	0.8195122
1	5	0.8935029	0.7936553	0.8292683
1	6	0.9023732	0.7785985	0.8731707
1	7	0.9080527	0.7818182	0.8853659
1	8	0.9061114	0.7813447	0.8634146
1	9	0.9101915	0.7970644	0.8780488
1	10	0.9106372	0.7848485	0.8780488
1	11	0.9173734	0.8032197	0.8853659
1	12	0.9199545	0.8062500	0.8804878
1	13	0.9180271	0.8094697	0.8780488
1	14	0.9166505	0.8125000	0.8780488
1	15	0.9166505	0.8125000	0.8780488
1	16	0.9166505	0.8125000	0.8780488
1	17	0.9166505	0.8125000	0.8780488
1	18	0.9166505	0.8125000	0.8780488
1	19	0.9166505	0.8125000	0.8780488
1	20	0.9166505	0.8125000	0.8780488
1	21	0.9166505	0.8125000	0.8780488
1	22	0.9166505	0.8125000	0.8780488
1	23	0.9166505	0.8125000	0.8780488
1	24	0.9166505	0.8125000	0.8780488
1	25	0.9166505	0.8125000	0.8780488
1	26	0.9166505	0.8125000	0.8780488
1	27	0.9166505	0.8125000	0.8780488
1	28	0.9166505	0.8125000	0.8780488
1	29	0.9166505	0.8125000	0.8780488
1	30	0.9166505	0.8125000	0.8780488
1	31	0.9166505	0.8125000	0.8780488
1	32	0.9166505	0.8125000	0.8780488
1	33	0.9166505	0.8125000	0.8780488
1	34	0.9166505	0.8125000	0.8780488
1	35	0.9166505	0.8125000	0.8780488
1	36	0.9166505	0.8125000	0.8780488
1	37	0.9166505	0.8125000	0.8780488
1	38	0.9166505	0.8125000	0.8780488
2	2	0.8023131	0.7631629	0.8414634
2	3	0.8564948	0.7418561	0.8487805
2	4	0.8640798	0.7356061	0.8560976
2	5	0.8882864	0.7663826	0.8658537
2	6	0.9016653	0.7847538	0.8804878
2	7	0.9033329	0.7694129	0.8926829
2	8	0.9036100	0.7695076	0.8926829
2	9	0.9046263	0.7757576	0.8853659
2	10	0.9074995	0.7849432	0.8853659
2	11	0.9069775	0.7941288	0.8829268
2	12	0.9100979	0.7818182	0.8902439
2	13	0.9057165	0.7818182	0.8731707
2	14	0.9039680	0.7819129	0.8780488
2	15	0.9033190	0.7910038	0.8707317
2	16	0.9008107	0.7910038	0.8707317
2	17	0.9012495	0.7910038	0.8707317
2	18	0.9010879	0.7910038	0.8707317
2	19	0.9010879	0.7910038	0.8707317
2	20	0.9010879	0.7910038	0.8707317
2	21	0.9010879	0.7910038	0.8707317
2	22	0.9010879	0.7910038	0.8707317
2	23	0.9010879	0.7910038	0.8707317
2	24	0.9010879	0.7910038	0.8707317
2	25	0.9010879	0.7910038	0.8707317
2	26	0.9010879	0.7910038	0.8707317
2	27	0.9010879	0.7910038	0.8707317
2	28	0.9010879	0.7910038	0.8707317
2	29	0.9010879	0.7910038	0.8707317
2	30	0.9010879	0.7910038	0.8707317
2	31	0.9010879	0.7910038	0.8707317
2	32	0.9010879	0.7910038	0.8707317
2	33	0.9010879	0.7910038	0.8707317
2	34	0.9010879	0.7910038	0.8707317
2	35	0.9010879	0.7910038	0.8707317
2	36	0.9010879	0.7910038	0.8707317
2	37	0.9010879	0.7910038	0.8707317
2	38	0.9010879	0.7910038	0.8707317

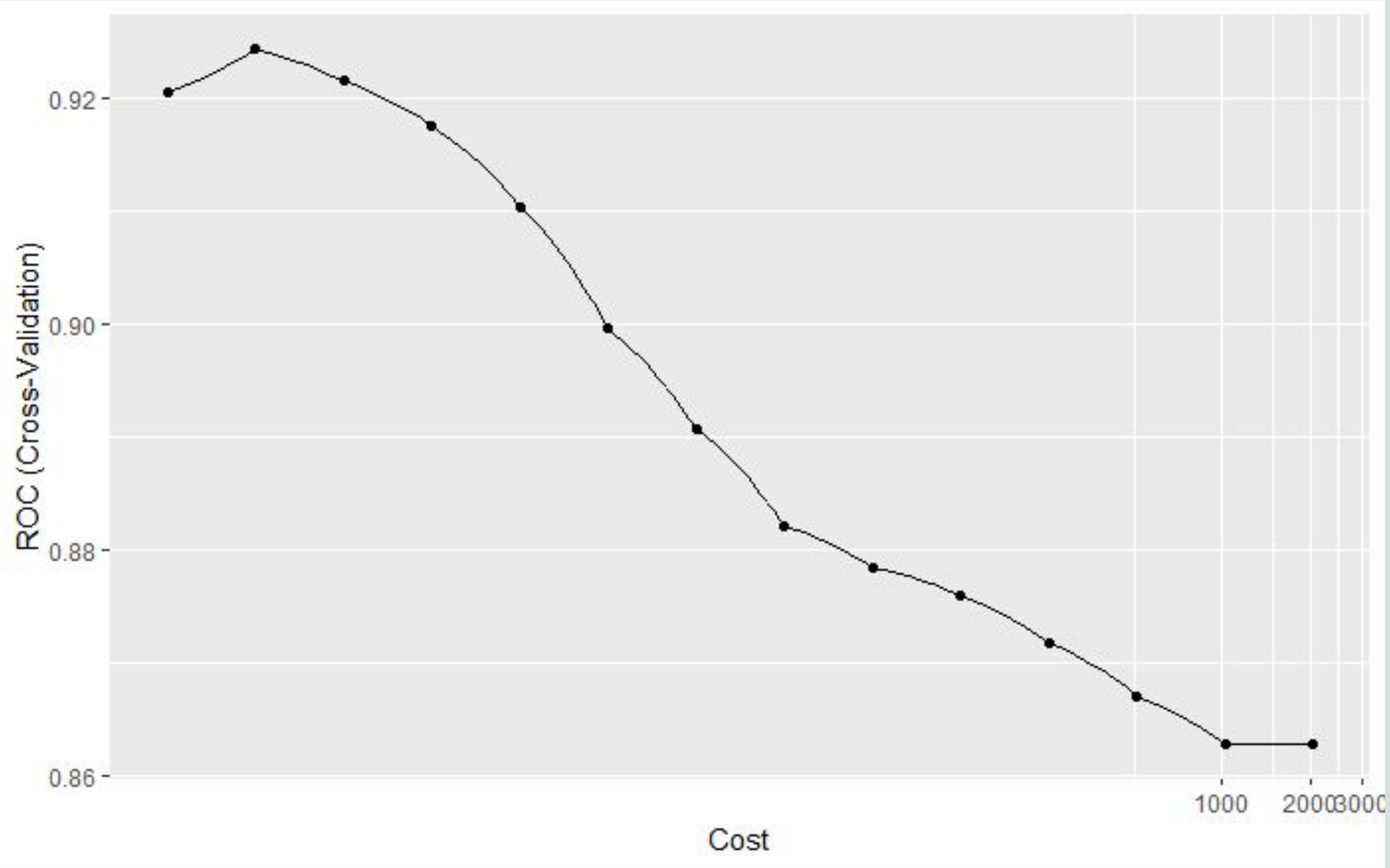
Support Vector Machine

```
> svm_model
Support Vector Machines with Radial Basis Function Kernel

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:

  C        ROC        Sens       Spec
    0.25   0.9206486  0.7971591  0.8878049
    0.50   0.9244157  0.8034091  0.8780488
    1.00   0.9215955  0.7972538  0.8829268
    2.00   0.9176275  0.8032197  0.8731707
    4.00   0.9104675  0.8033144  0.8731707
    8.00   0.8996235  0.7819129  0.8585366
   16.00   0.8907012  0.7849432  0.8634146
   32.00   0.8821739  0.7727273  0.8682927
   64.00   0.8785107  0.7603220  0.8585366
  128.00   0.8759770  0.7568182  0.8390244
  256.00   0.8717803  0.7416667  0.8414634
  512.00   0.8670847  0.7322917  0.8414634
 1024.00   0.8627749  0.7261364  0.8439024
 2048.00   0.8627749  0.7323864  0.8414634

Tuning parameter 'sigma' was held constant at a value of 0.04603174
ROC was used to select the optimal model using the largest value.
The final values used for the model were sigma = 0.04603174 and C = 0.5.
```



Area Under the Curve: **0.9244157**

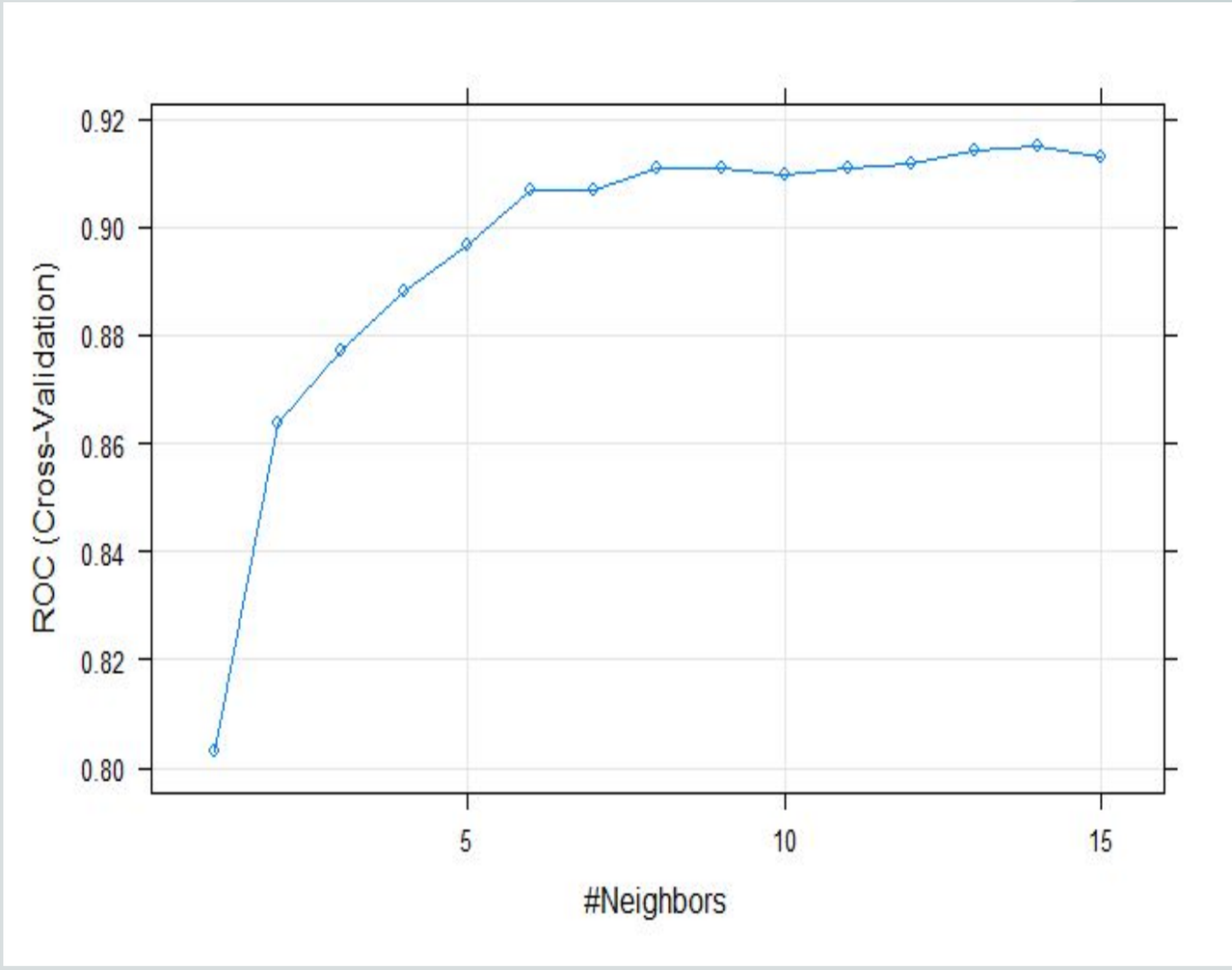
Support Vector Machine



Plotted SVM Model on Log Scale

K - Nearest Neighbors

```
> knnTune
k-Nearest Neighbors

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results across tuning parameters:

  k   ROC        Sens       Spec
   1  0.8032474  0.7723485  0.8341463
   2  0.8637276  0.7720644  0.8170732
   3  0.8769436  0.7850379  0.8658537
   4  0.8879481  0.7819129  0.8756098
   5  0.8967168  0.7940341  0.8853659
   6  0.9068701  0.7879735  0.8902439
   7  0.9066935  0.7881629  0.8829268
   8  0.9108971  0.7974432  0.8756098
   9  0.9107446  0.8002841  0.8878049
  10  0.9096799  0.8035038  0.8926829
  11  0.9108705  0.7972538  0.8829268
  12  0.9115530  0.8157197  0.8878049
  13  0.9142773  0.7972538  0.8902439
  14  0.9147889  0.8001894  0.8902439
  15  0.9130601  0.7940341  0.8926829

ROC was used to select the optimal model using the largest value.
The final value used for the model was k = 14.
```



Area Under the Curve: **0.9147889**

Naive Bayes

```
Naive Bayes

735 samples
 15 predictor
  2 classes: 'No', 'Yes'

Pre-processing: centered (15), scaled (15)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 661, 661, 662, 662, 661, 662, ...
Resampling results:

  ROC       Sens       Spec
  0.913447  0.7569129  0.9146341

Tuning parameter 'fL' was held constant at a value of 2
Tuning parameter 'usekernel' was
 held constant at a value of TRUE
Tuning parameter 'adjust' was held constant at a value of TRUE
```

No Tuning Parameter Plot for Naive Bayes

Area Under the Curve: **0.913447**

Top 2 Models (Considering ROC):
Penalized & Support Vector Machine (SVM)

MODEL	ROC	SENSITIVITY	SPECIFICITY	BEST TUNING PARAMETER
Logistic Regression	0.9168953	0.8125	0.8756098	NA
LDA	0.918429	0.8094697	0.8829268	NA
PLSDA	0.9195607	0.8156250	0.8829268	ncomp = 4
Penalized	0.9209765	0.8032197	0.8853659	alpha = 0.2 lambda = 0.03111111
Nearest Shrunken Centroids	0.9098901	0.7693182	0.8902439	threshold = 0
MARS	0.9196727	0.7970644	0.8878049	nprune = 12 degree = 1
Nonlinear Discriminant Analysis	0.9184290	0.8094697	0.8829268	subclasses = 1
Neural Networks	0.9147866	0.8186553	0.8902439	size = 1 decay = 0.1 bag = T
Flexible Discriminant Analysis	0.9199545	0.8062500	0.8804878	degree = 1 nprune = 12
SVM	0.9244157	0.8034091	0.8780488	sigma = 0.04603174 c = 0.5
KNN	0.9147889	0.8001894	0.8902439	k = 14
Naive Bayes	0.913447	0.7569129	0.9146341	NA

Confusion Matrix for Best Models on Testing Set

Penalized

```
> confusionMatrix(glmnPred,test_response)
Confusion Matrix and Statistics

          Reference
Prediction No Yes
       No  75  10
       Yes 10  88

               Accuracy : 0.8907
                 95% CI : (0.8363, 0.932)
    No Information Rate : 0.5355
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7803

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.8824
            Specificity : 0.8980
         Pos Pred Value : 0.8824
         Neg Pred Value : 0.8980
             Prevalence : 0.4645
         Detection Rate : 0.4098
   Detection Prevalence : 0.4645
      Balanced Accuracy : 0.8902

       'Positive' Class : No
```

Area under the curve: 0.8902

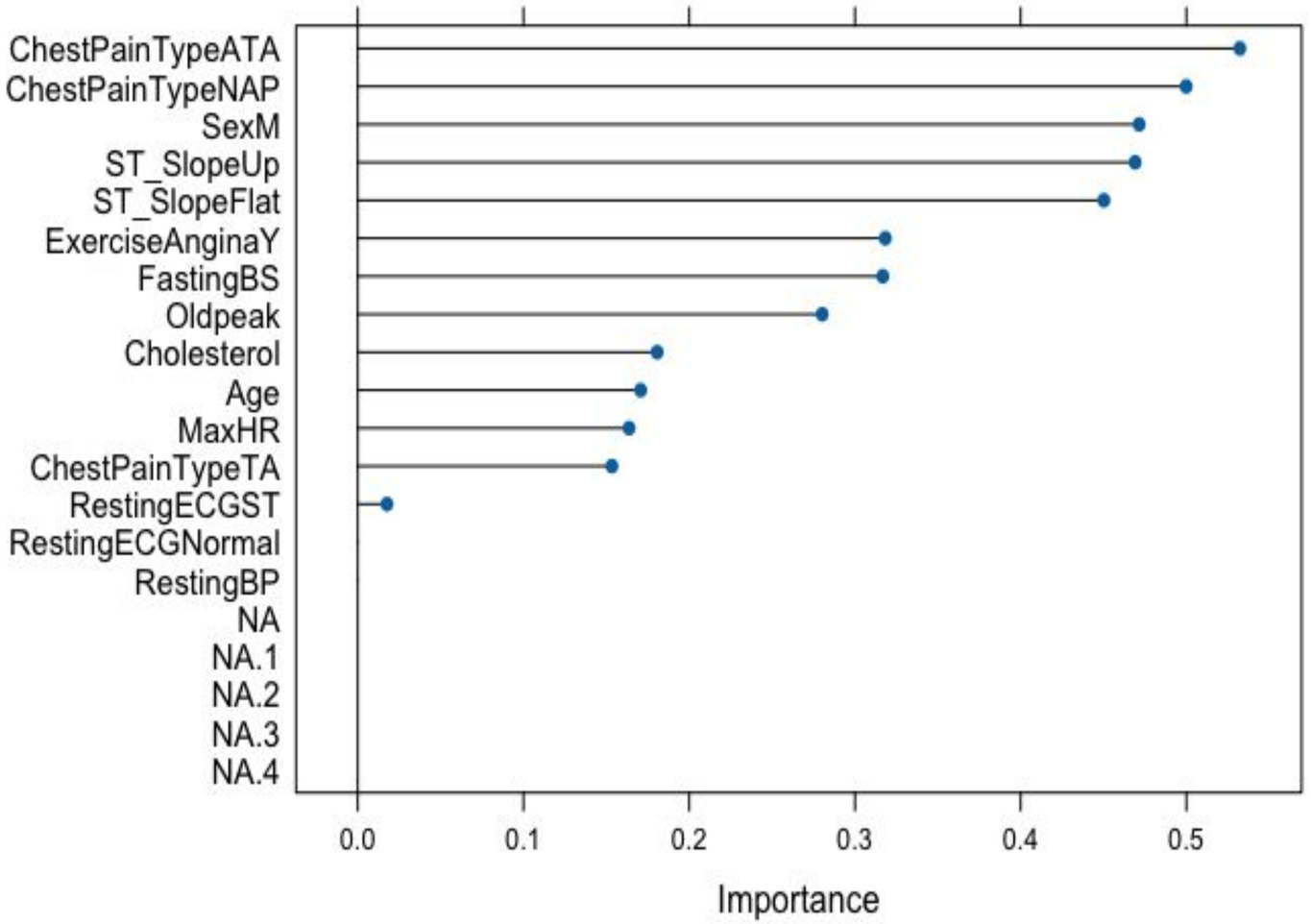
SVM

```
> confusionMatrix(svmRpred,test_response)
Confusion Matrix and Statistics

          Reference
Prediction No Yes
       No  74   9
       Yes 11  89

               Accuracy : 0.8907
                 95% CI : (0.8363, 0.932)
    No Information Rate : 0.5355
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.78

 Mcnemar's Test P-Value : 0.8231

            Sensitivity : 0.8706
            Specificity : 0.9082
         Pos Pred Value : 0.8916
         Neg Pred Value : 0.8900
             Prevalence : 0.4645
         Detection Rate : 0.4044
   Detection Prevalence : 0.4536
      Balanced Accuracy : 0.8894

       'Positive' Class : No
```

Area under the curve: 0.8894

IMPORTANT PARAMS

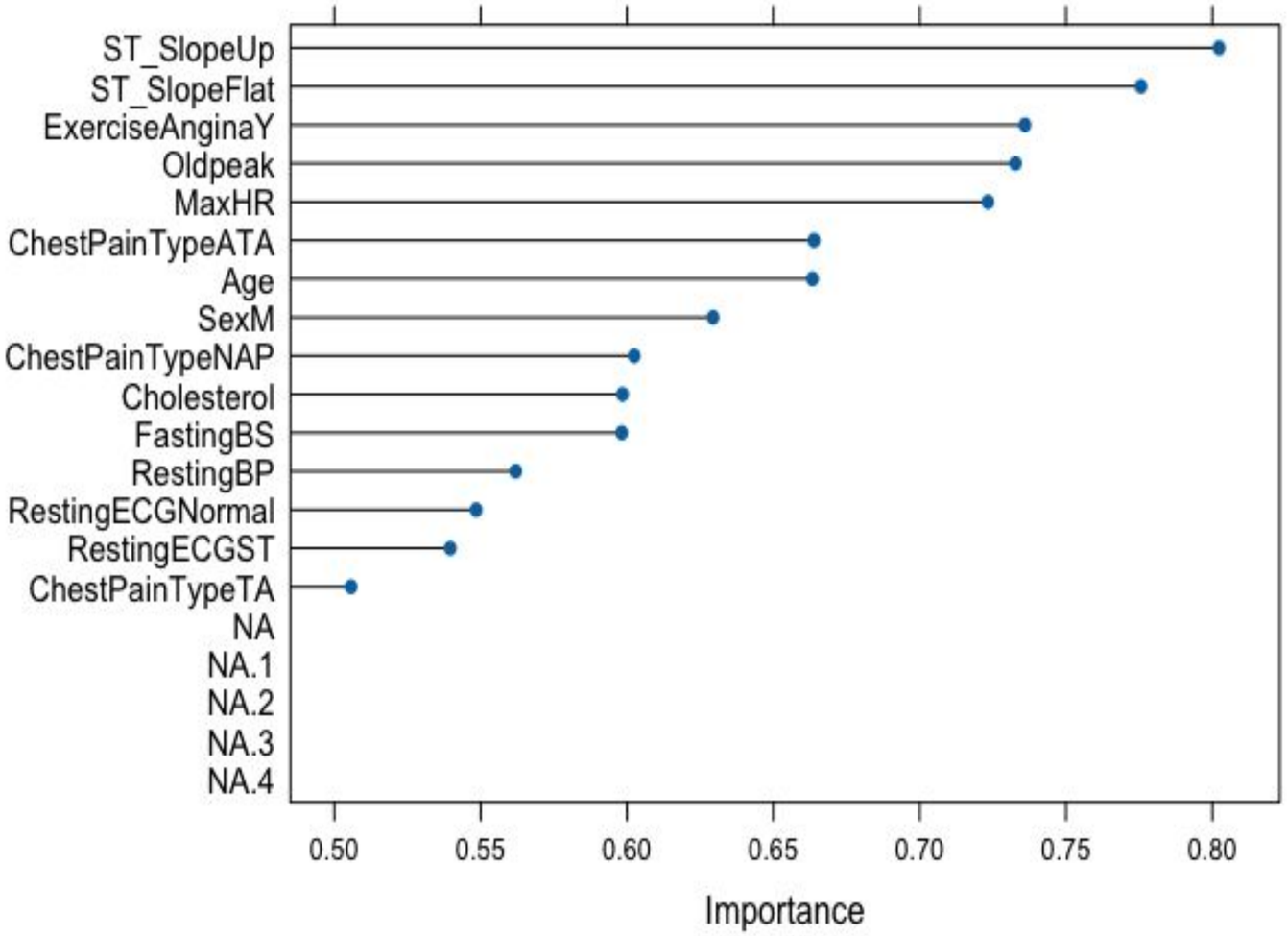
Penalized




```
> varImp(glmnTuned, scale = FALSE)
glmnet variable importance

                  Overall
ChestPainTypeATA  0.53227
ChestPainTypeNAP  0.49991
SexM              0.47152
ST_SlopeUp        0.46910
ST_SlopeFlat      0.45023
ExerciseAnginaY   0.31833
FastingBS         0.31682
Oldpeak           0.28028
Cholesterol       0.18069
Age               0.17069
MaxHR             0.16372
ChestPainTypeTA   0.15335
RestingECGST      0.01764
RestingBP         0.00000
RestingECGNormal  0.00000
```

IMPORTANT PARAMS

SVM




```
> varImp(svm_model, scale = FALSE)
ROC curve variable importance
```

	Importance
ST_SlopeUp	0.8023
ST_SlopeFlat	0.7756
ExerciseAnginaY	0.7360
Oldpeak	0.7327
MaxHR	0.7233
ChestPainTypeATA	0.6639
Age	0.6634
SexM	0.6294
ChestPainTypeNAP	0.6025
Cholesterol	0.5985
FastingBS	0.5982
RestingBP	0.5620
RestingECGNormal	0.5485
RestingECGST	0.5396
ChestPainTypeTA	0.5057

Final Recommendation

```
> confusionMatrix(glmnPred,test_response)
Confusion Matrix and Statistics

          Reference
Prediction No Yes
       No  75  10
       Yes 10  88

               Accuracy : 0.8907
                 95% CI : (0.8363, 0.932)
    No Information Rate : 0.5355
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7803

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.8824
            Specificity : 0.8980
         Pos Pred Value : 0.8824
         Neg Pred Value : 0.8980
             Prevalence : 0.4645
         Detection Rate : 0.4098
   Detection Prevalence : 0.4645
      Balanced Accuracy : 0.8902

       'Positive' Class : No
```

```
Area under the curve: 0.8902
```

Penalized = best model!

According to the area under the curve (AUC) of the receiving operator characteristic (ROC), Penalized (glmnet) is the best model!

AUC of 0.8902

Thank you!

https://www.dreamstime.com/photos-images/thank-you.html