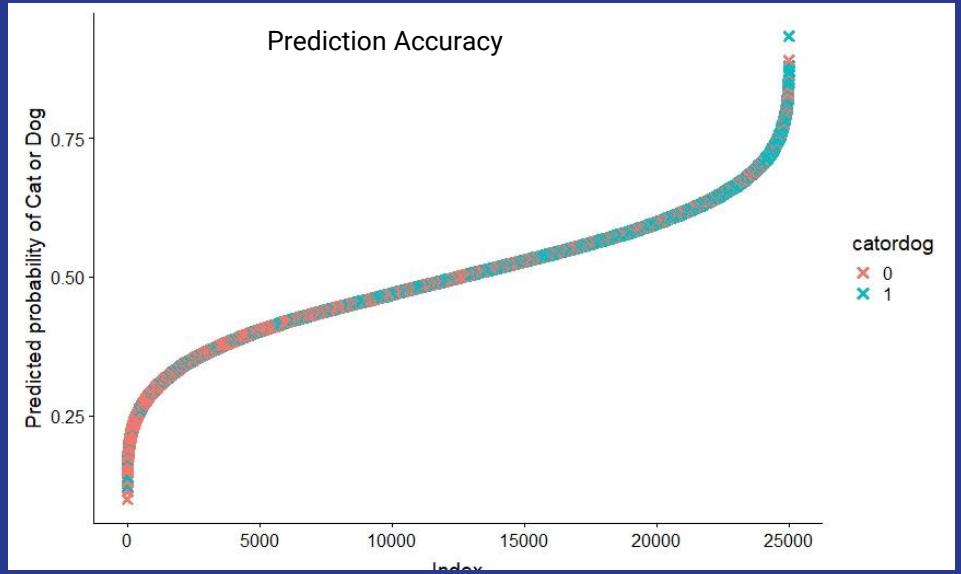
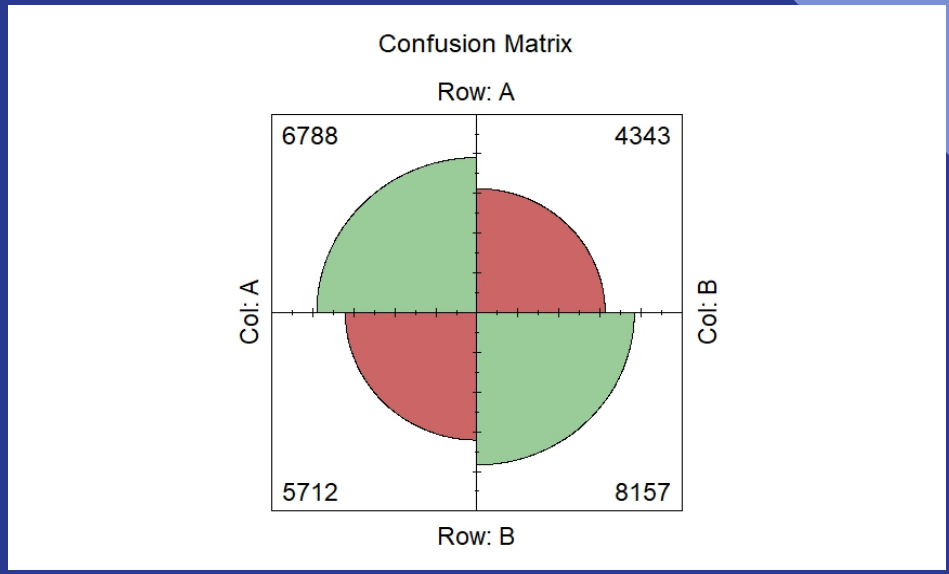
CIS 490: Sectional Project 1 Presentation

By: Blake Simmons, Daniel Gomes, and John Gomes

Cats vs Dogs - Logistic Regression





Using whole dataset

```
V14         -0.2195757  0.1347969  -1.629 0.103326
V15          0.1330690  0.1362515   0.977 0.328746
V16         -0.0694667  0.1337163  -0.520 0.603407
V17          0.0971324  0.1355961   0.716 0.473784
V18         -0.0078639  0.1382705  -0.057 0.954646
V19         -0.3262506  0.1387319  -2.352 0.018690 *
 [ reached getOption("max.print") -- omitted 606 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 34657  on 24999  degrees of freedom
Residual deviance: 33289  on 24374  degrees of freedom
AIC: 34541

Number of Fisher Scoring iterations: 4

> pR2(catsvsdogs.glm)
         llh       llhNull            G2      McFadden          r2ML          r2CU
-1.664425e+04 -1.732868e+04  1.368851e+03  3.949670e-02  5.328203e-02  7.104271e-02
```

Using 80% training, 20% testing

```
V194         0.1908143  0.1315291   1.451 0.146853
V195         0.1382606  0.1359386   1.017 0.309115
V196         0.1798571  0.1392049   1.292 0.196346
V197        -0.1582244  0.1455677  -1.087 0.277060
V198         0.2424607  0.1509698   1.606 0.108269
V199         0.3572198  0.1601102   2.231 0.025675 *
 [ reached getOption("max.print") -- omitted 426 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27726  on 19999  degrees of freedom
Residual deviance: 26501  on 19374  degrees of freedom
AIC: 27753

Number of Fisher Scoring iterations: 4

> pR2(catsvsdogs.glm)
         llh       llhNull            G2      McFadden          r2ML          r2CU
-1.325052e+04 -1.386291e+04  1.224778e+03  4.417465e-02  5.940151e-02  7.920210e-02
```

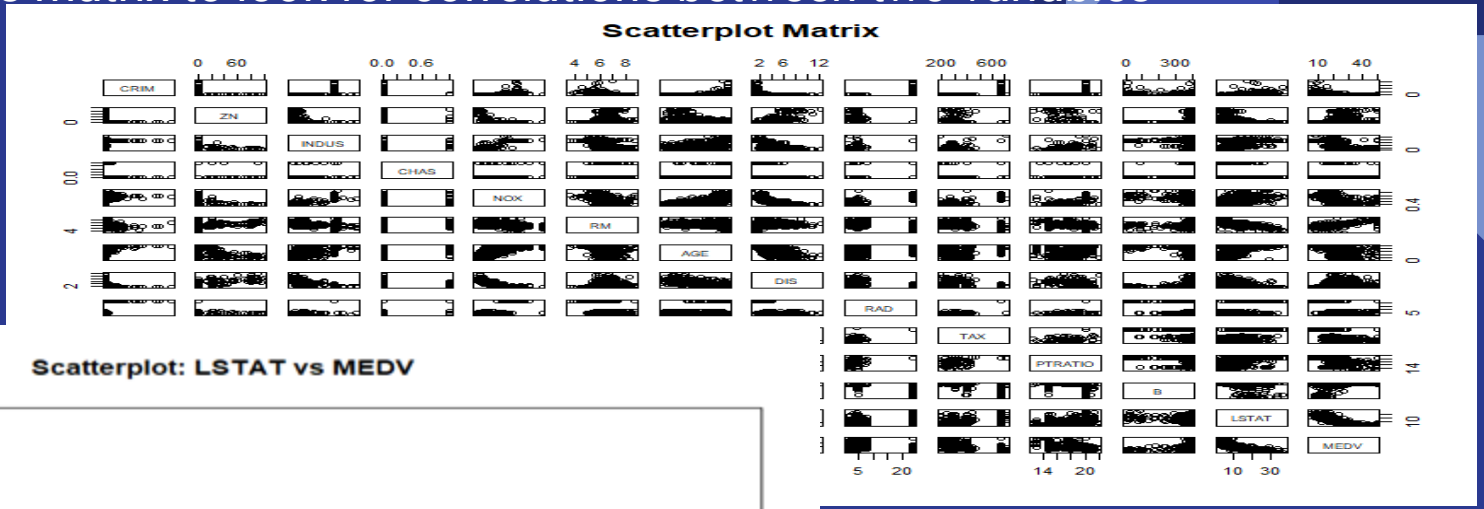
Boston Housing - Linear Regressions

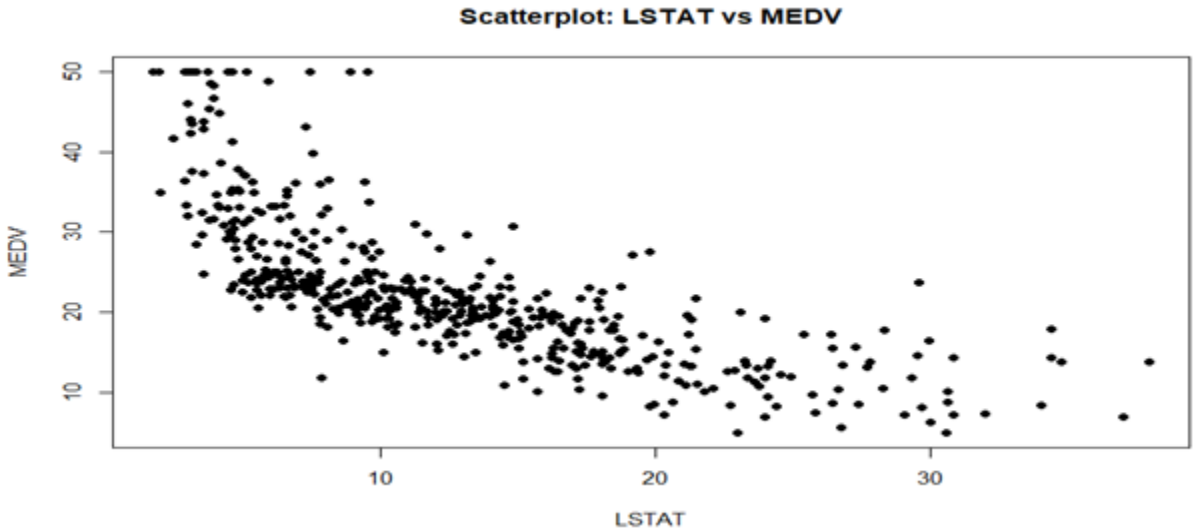
Data Description

Mean		Variance		Median	
CRIM:	3.61	**CRIM:**	7.39e+01	**CRIM:**	0.26
ZN:	11.36	**ZN:**	5.44e+02	**ZN:**	0.00
INDUS:	11.14	**INDUS:**	4.71e+01	**INDUS:**	9.69
CHAS:	0.069	**CHAS:**	6.43e-02	**CHAS:**	0.00
NOX:	0.55	**NOX:**	1.34e-02	**NOX:**	0.54
RM:	6.28	**RM:**	4.94e-01	**RM:**	6.21
AGE:	68.57	**AGE:**	7.92e+02	**AGE:**	77.50
DIS:	18.45	**DIS:**	4.43+00	**DIS:**	3.21
RAD:	9.55	**RAD:**	7.58e+01	**RAD:**	5.00
TAX:	408.24	**TAX:**	2.84e+04	**TAX:**	330.00
PTRATIO:	18.45	**PTRATIO:**	4.69e+00	**PTRATIO:**	19.05
B:	356.67	**B:**	8.33e+03	**B:**	391.44
LSTAT:	12.65	**LSTAT:**	5.09e+01	**LSTAT:**	11.36
MEDV:	22.53	**MEDV:**	8.46e+01	**MEDV:**	21.20

Scatterplot Matrix

We Created this matrix to look for correlations between two variables





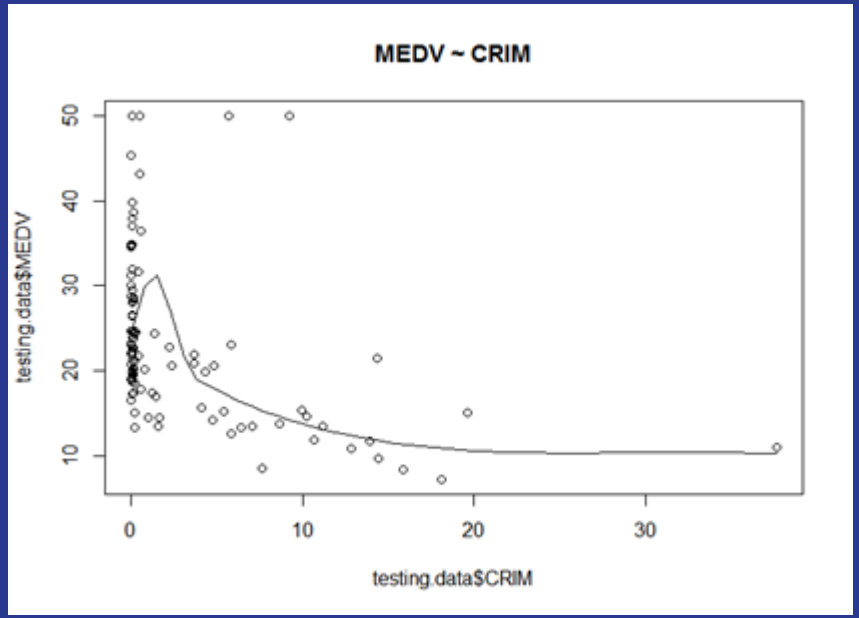
Multiple Linear Regression

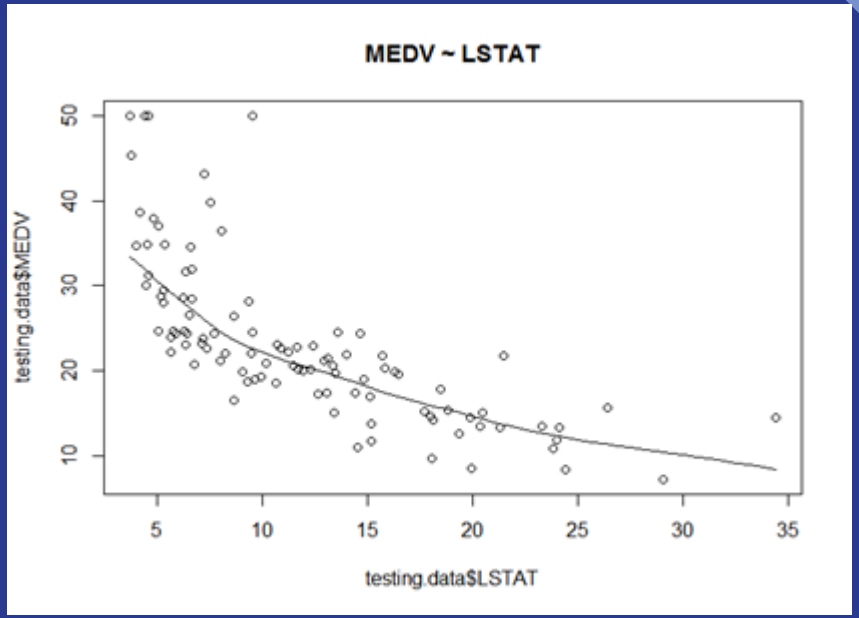
```
Call:
lm(formula = frml, data = training.data)

Residuals:
     Min       1Q   Median       3Q      Max
-14.6682  -2.8641  -0.6328   1.7493  26.6253

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.704860   5.493485   7.046 8.41e-12 ***
CRIM         -0.101555   0.033725  -3.011  0.00277 **
ZN            0.039328   0.015274   2.575  0.01040 *
INDUS         0.035499   0.067568   0.525  0.59961
CHASTRUE      2.435641   0.910881   2.674  0.00781 **
NOX         -17.757531   4.182255  -4.246 2.72e-05 ***
RM            3.691904   0.447782   8.245 2.55e-15 ***
AGE          -0.002027   0.014481  -0.140  0.88875
DIS          -1.332287   0.221490  -6.015 4.14e-09 ***
RAD           0.295501   0.073359   4.028 6.76e-05 ***
TAX          -0.011194   0.004184  -2.676  0.00777 **
PTRATIO      -1.072440   0.148331  -7.230 2.57e-12 ***
B             0.008018   0.003053   2.626  0.00898 **
LSTAT        -0.498635   0.055112  -9.048  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.687 on 390 degrees of freedom
Multiple R-squared:  0.745,     Adjusted R-squared:  0.7365
F-statistic: 87.65 on 13 and 390 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = frml, data = testing.data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.7645 -2.6535 -0.2508  1.8671 24.2126

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.436e+01  1.459e+01   1.670 0.098517 .
CRIM        -3.163e-01  1.497e-01  -2.112 0.037480 *
ZN           8.145e-02  3.387e-02   2.405 0.018260 *
INDUS        6.427e-02  1.677e-01   0.383 0.702455
CHASTRUE     5.416e+00  2.605e+00   2.079 0.040538 *
NOX         -1.774e+01  9.568e+00  -1.854 0.067036 .
RM           4.984e+00  1.227e+00   4.062 0.000105 ***
AGE          1.631e-04  3.329e-02   0.005 0.996101
DIS         -1.913e+00  5.017e-01  -3.814 0.000254 ***
RAD          3.937e-01  1.744e-01   2.257 0.026475 *
TAX         -1.721e-02  9.546e-03  -1.803 0.074762 .
PTRATIO     -4.790e-01  2.899e-01  -1.653 0.101983
B            9.325e-03  6.482e-03   1.439 0.153802
LSTAT       -6.402e-01  1.309e-01  -4.892 4.48e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.943 on 88 degrees of freedom
Multiple R-squared:  0.7629,     Adjusted R-squared:  0.7279
F-statistic: 21.78 on 13 and 88 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

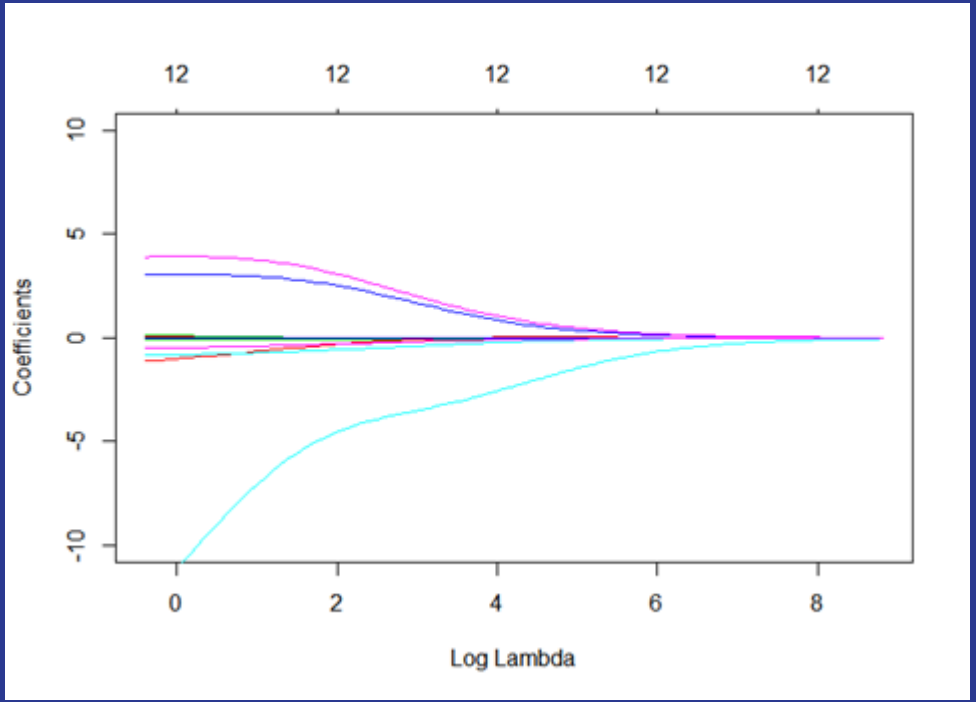
Predictions vs. Testing Data



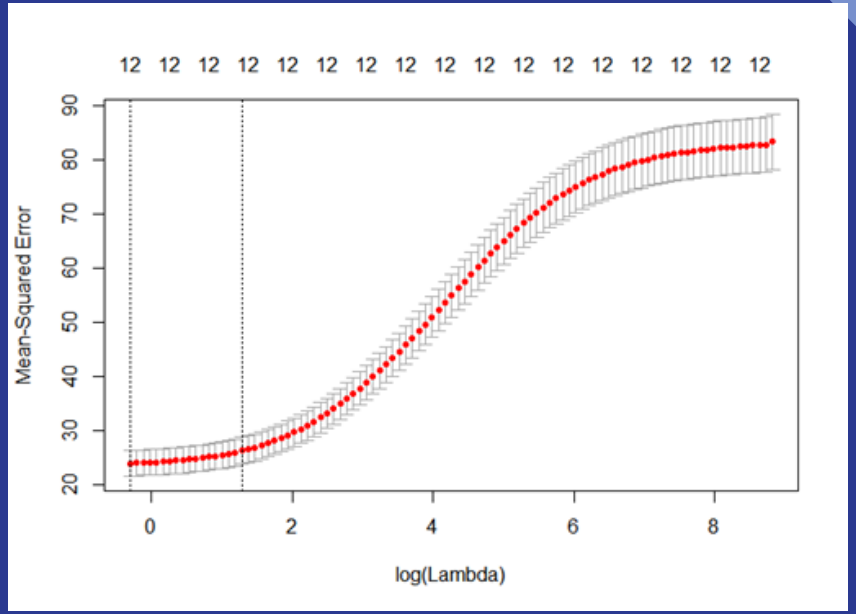


Ridge Regression

Graph of SSE vs lambda



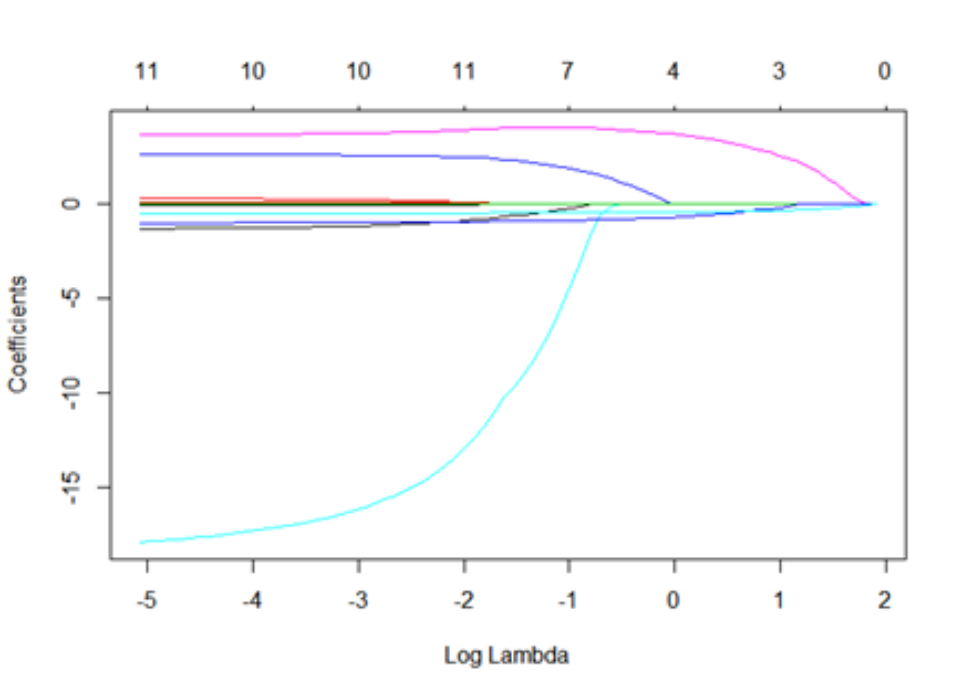
Cross Validation to determine Lambda



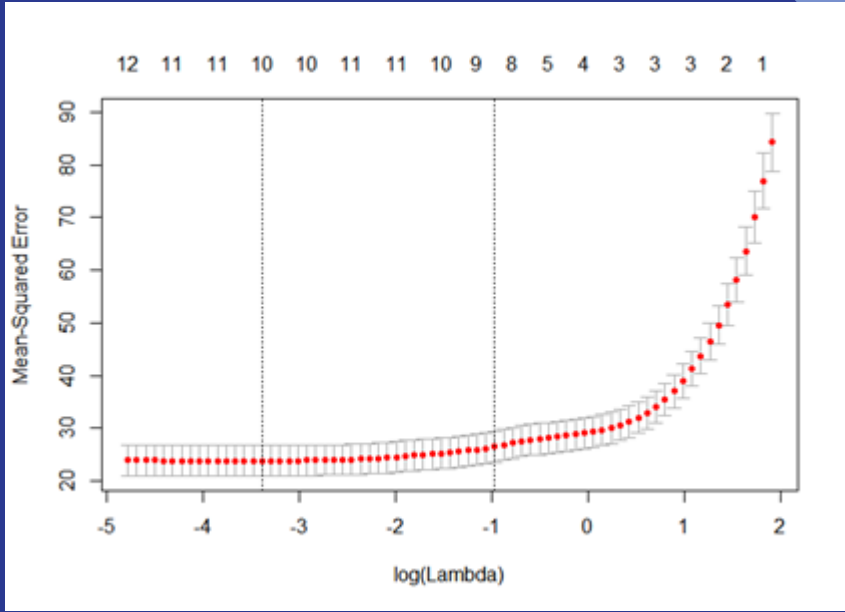
Lambda: 0.619915

Lasso Regression

Graph of SSE vs lambda



Cross Validation to determine Lambda



Lambda: 0.3789258

Ridge Regression

Results

```
> ridge.results
14 x 1 sparse Matrix of class "dgCMatrix"
                       1
(Intercept)  32.262004014
(Intercept)   .
CRIM         -0.099228668
ZN            0.032089861
INDUS        -0.046346825
CHASTRUE      3.044015331
NOX         -12.361675962
RM            3.906808828
AGE          -0.002168196
DIS          -1.101591472
RAD           0.132836240
TAX          -0.005960766
PTRATIO      -0.836374783
LSTAT        -0.489658705
```

Lasso Regression

Results

```
(Intercept) 21.6834491048
(Intercept)  .
CRIM        -0.0361155884
ZN           .
INDUS        .
CHASTRUE     2.0058948525
NOX         -4.4629140052
RM           4.1558518545
AGE          .
DIS         -0.3357429179
RAD          .
TAX         -0.0002192817
PTRATIO     -0.7942328505
LSTAT       -0.5358125716
```

Each member contributed equally.