CoreLogic

Chris Eshleman

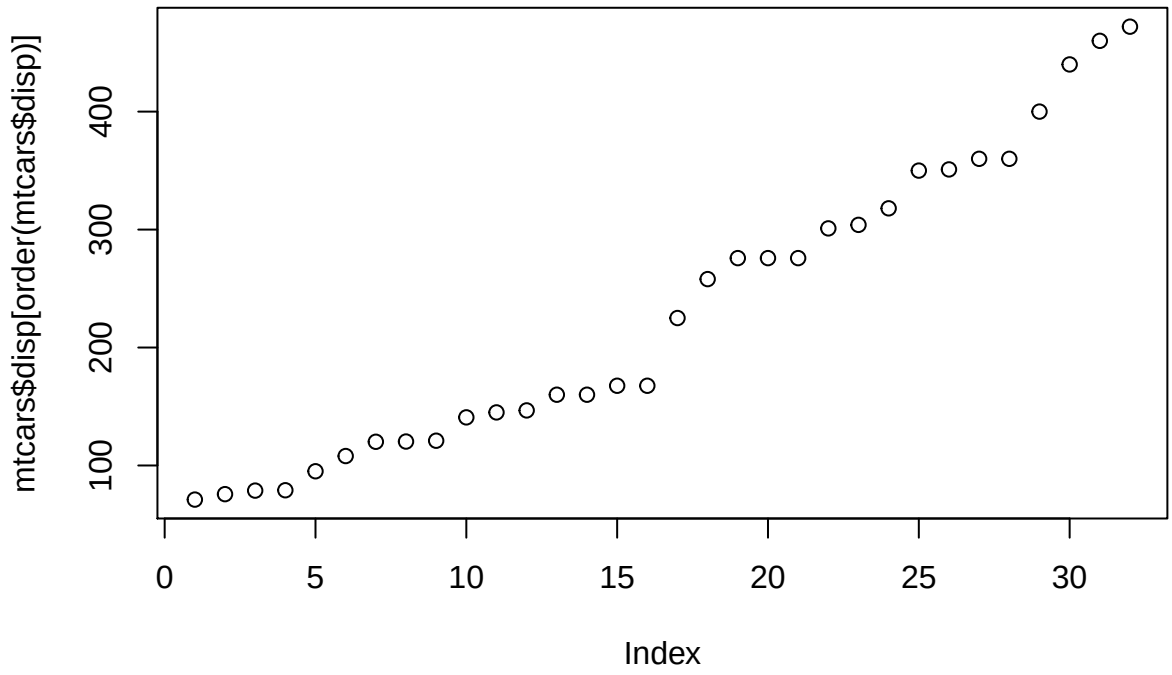
TBD

```
##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##     rivers
```

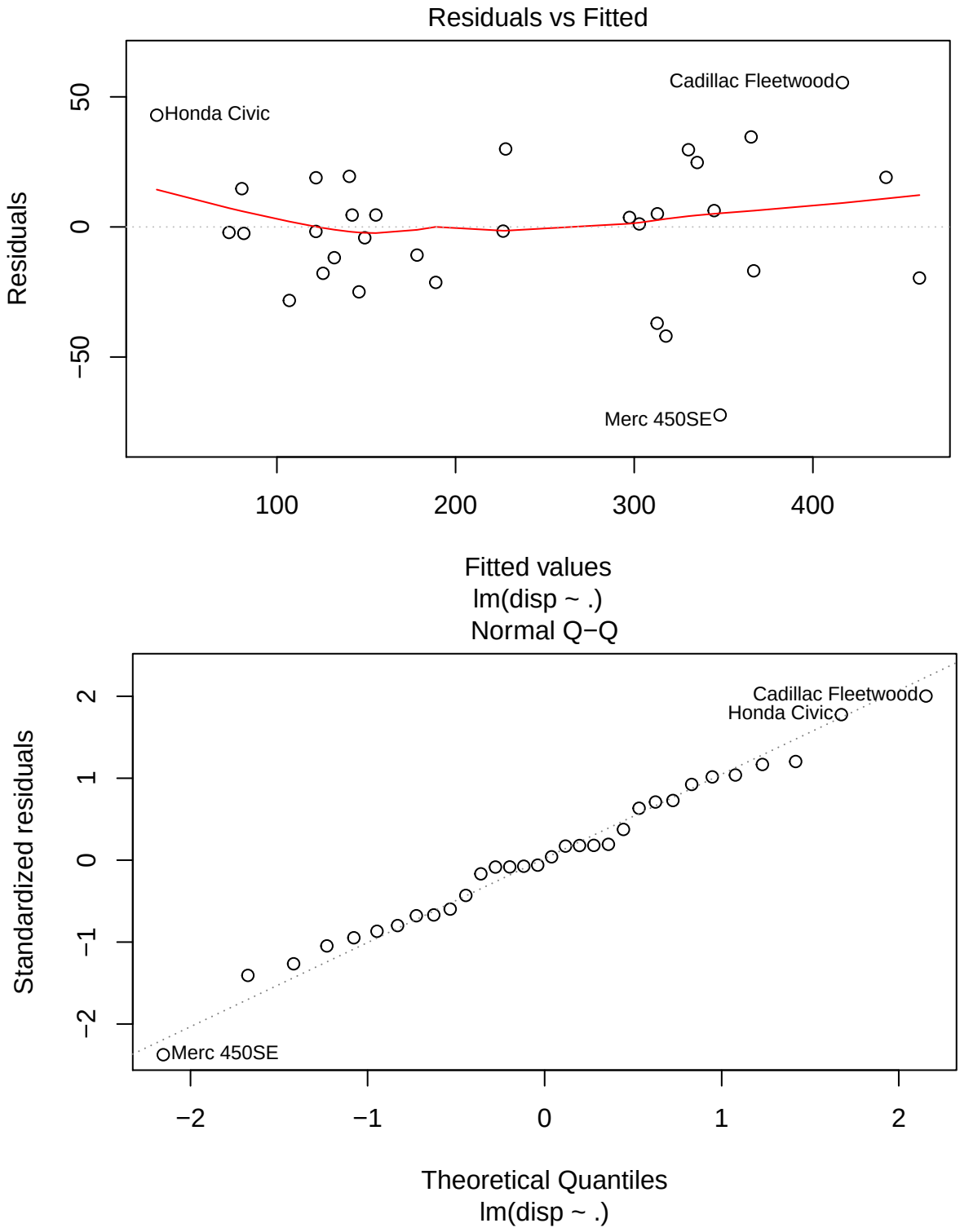
- Check for outliers or influential observations.
- Check index and other covariates for necessary transformations. Visual for the predictors and also a whatchamacallit for the index itself.
- Subset selection of inputs on index. Correlation analysis from there. Then machine learning on leftovers, consider adding to the mix. Does the best-scoring input really deviate strongly from the index?
- Treat the index as continuous. Technically, an index from 1-100 isn't a pure continuous (quantitative) variable but it falls closer to that than other variable types.
- Potential follow-up: check variation of crime statistics across the Conrex-non Conrex divide. (Just use MSA designations.)

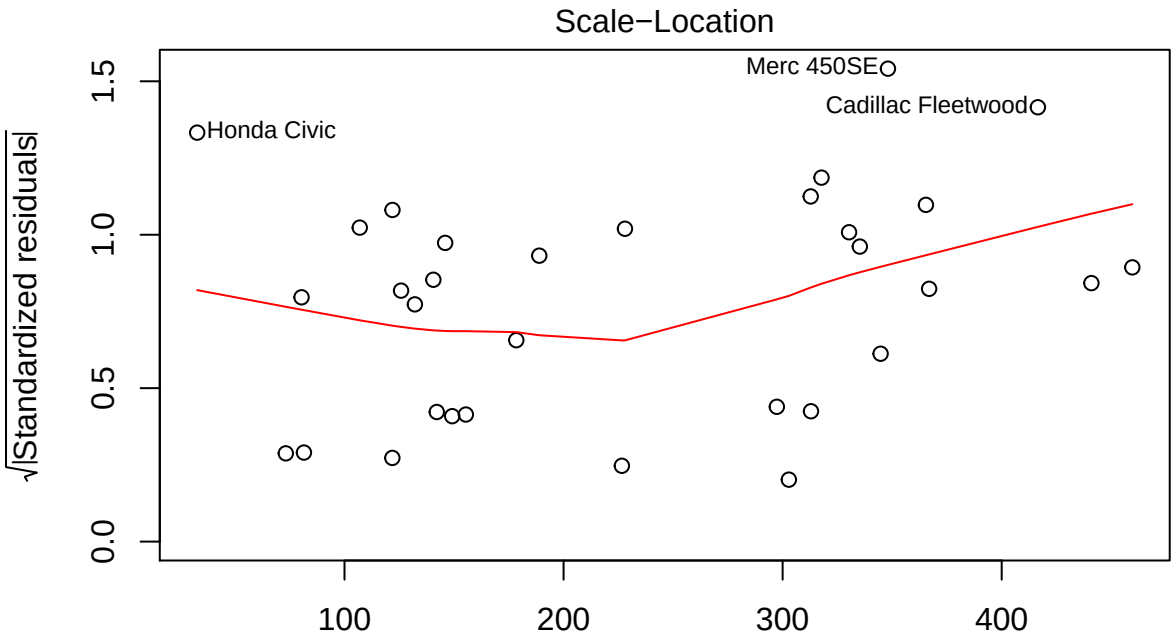
Index

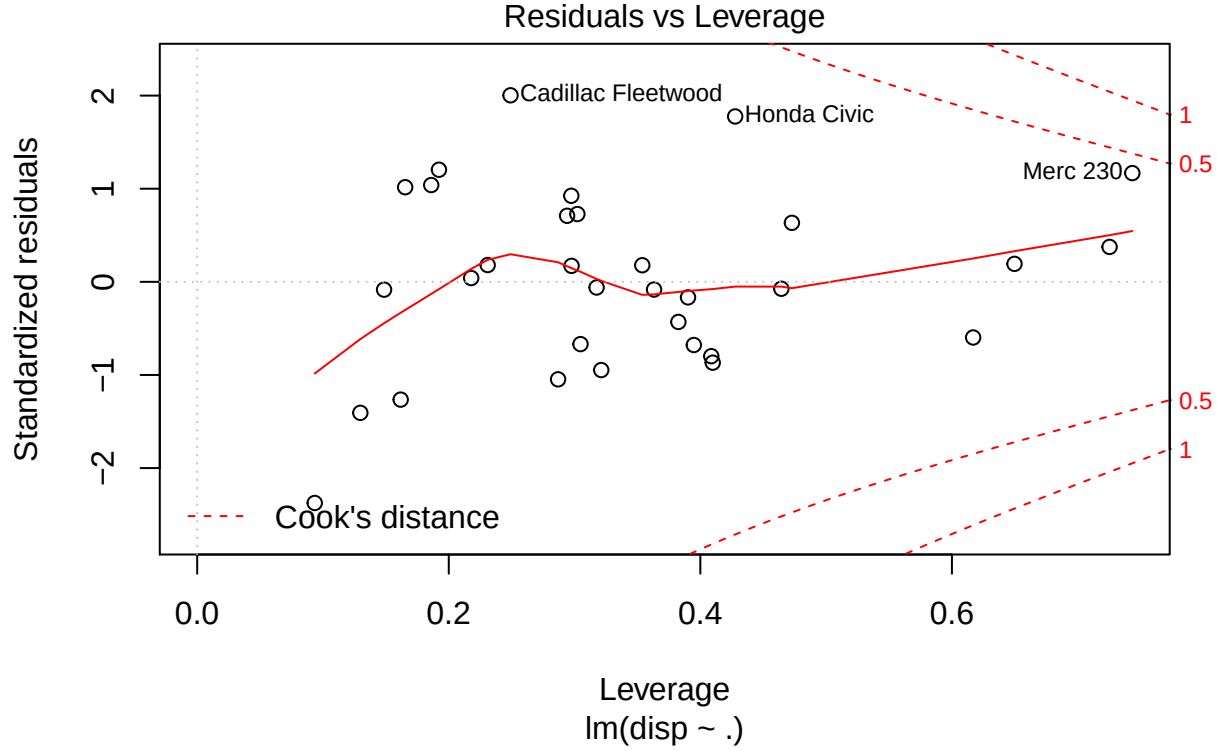
```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```



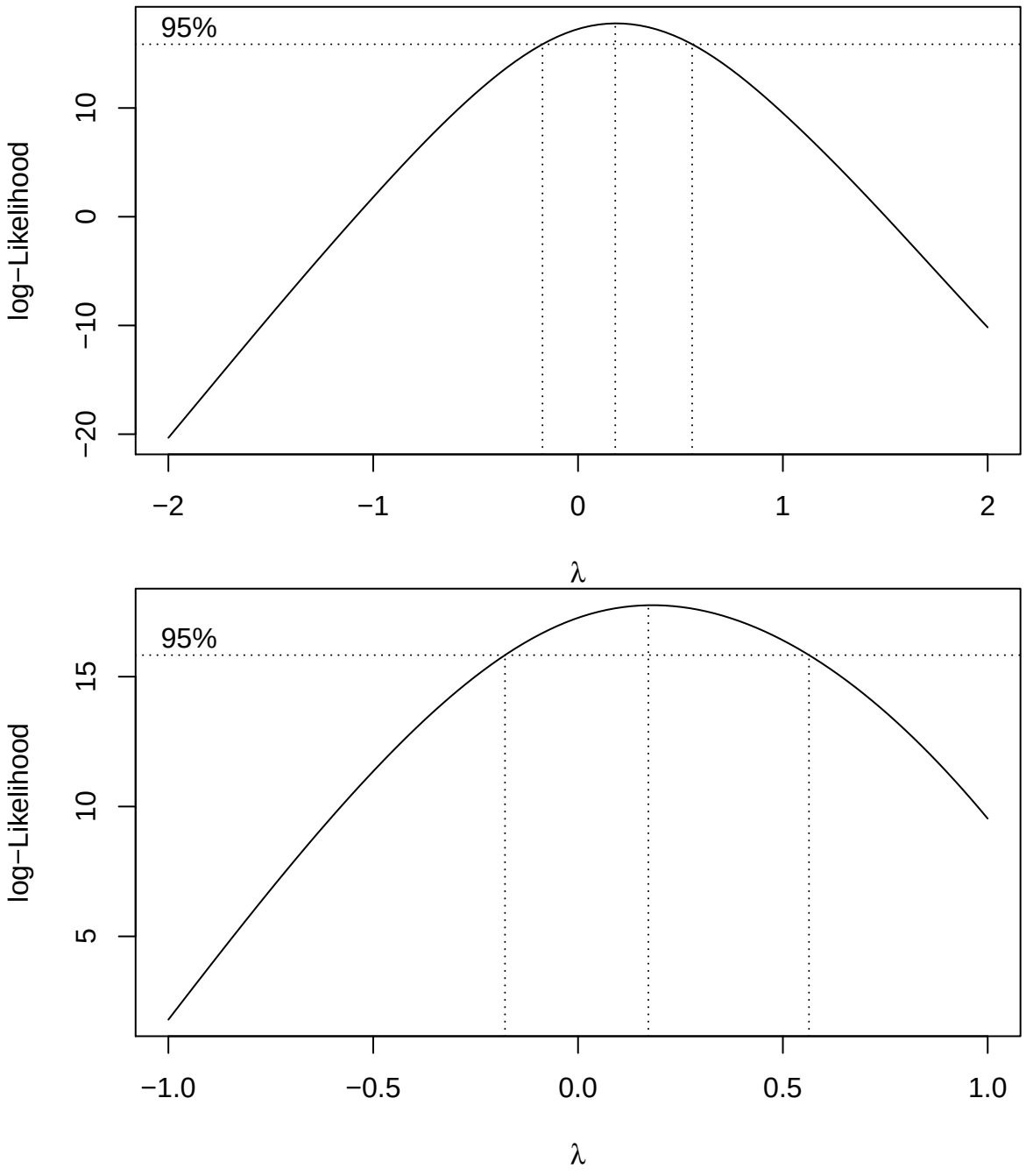

```
##
## Call:
## lm(formula = disp ~ ., data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -72.28 -17.11  -0.23  18.95  55.48
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -5.812    228.061  -0.025  0.97991
## mpg            1.940      2.598   0.747  0.46349
## cyl           15.389     12.152   1.266  0.21924
## hp             0.665      0.226   2.942  0.00778 **
## drat           8.812     19.739   0.446  0.65987
## wt            86.711     16.113   5.382 2.45e-05 ***
## qsec         -12.974      8.623  -1.505  0.14730
## vs           -12.115     25.258  -0.480  0.63643
## am            -7.914     25.618  -0.309  0.76044
## gear           5.127     18.058   0.284  0.77927
## carb         -30.107      7.551  -3.987  0.00067 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31.96 on 21 degrees of freedom
## Multiple R-squared:  0.9549, Adjusted R-squared:  0.9335
## F-statistic: 44.51 on 10 and 21 DF,  p-value: 7.255e-12
```

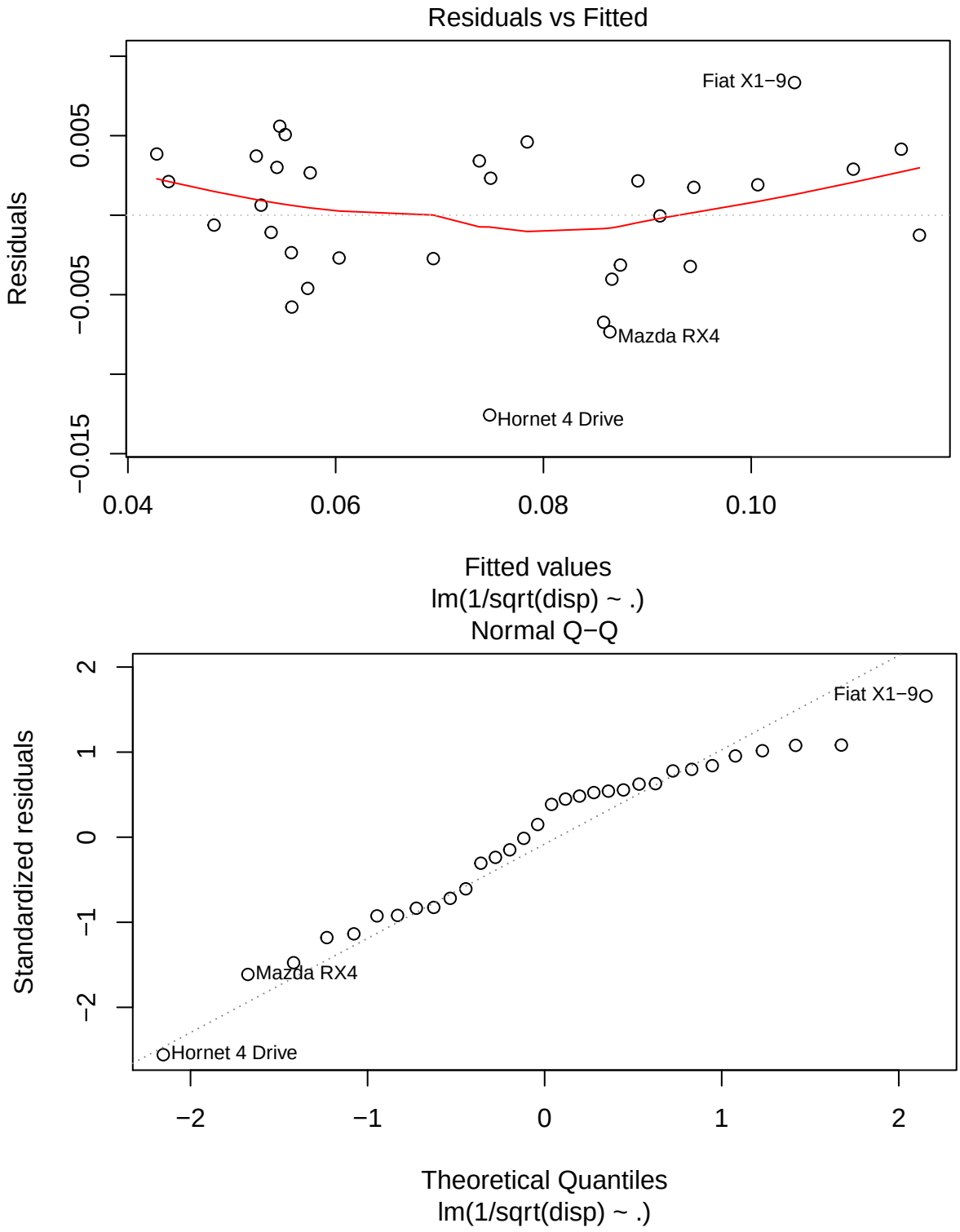

Residuals vs Fitted
Cadillac Fleetwood
Honda Civic
Merc 450SE
Residuals
50
0
−50
100
200
300
400
Fitted values
lm(disp ~ .)
Normal Q−Q
Cadillac Fleetwood
Honda Civic
Merc 450SE
Standardized residuals
2
1
0
−1
−2
−2
−1
0
1
2
Theoretical Quantiles
lm(disp ~ .)

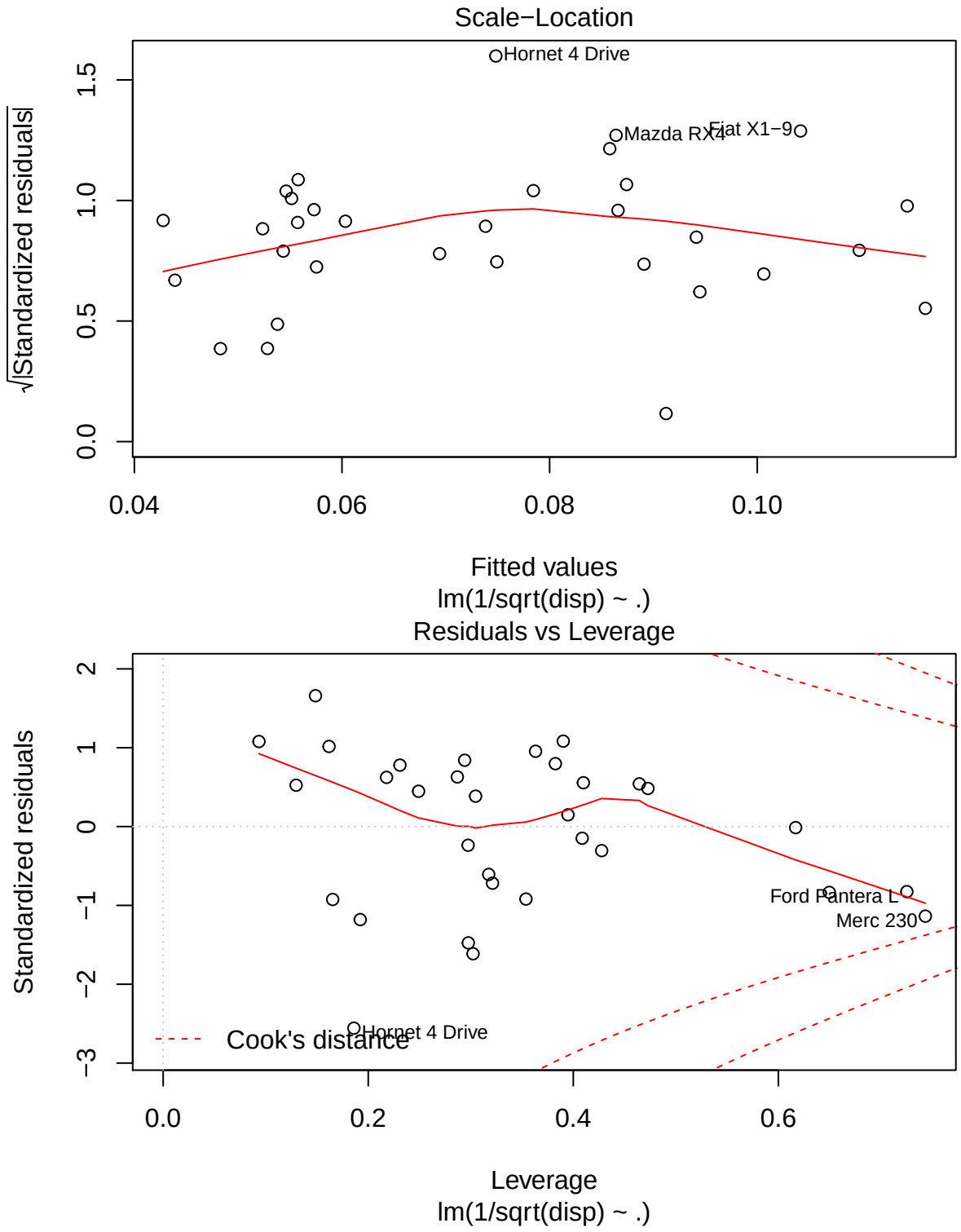




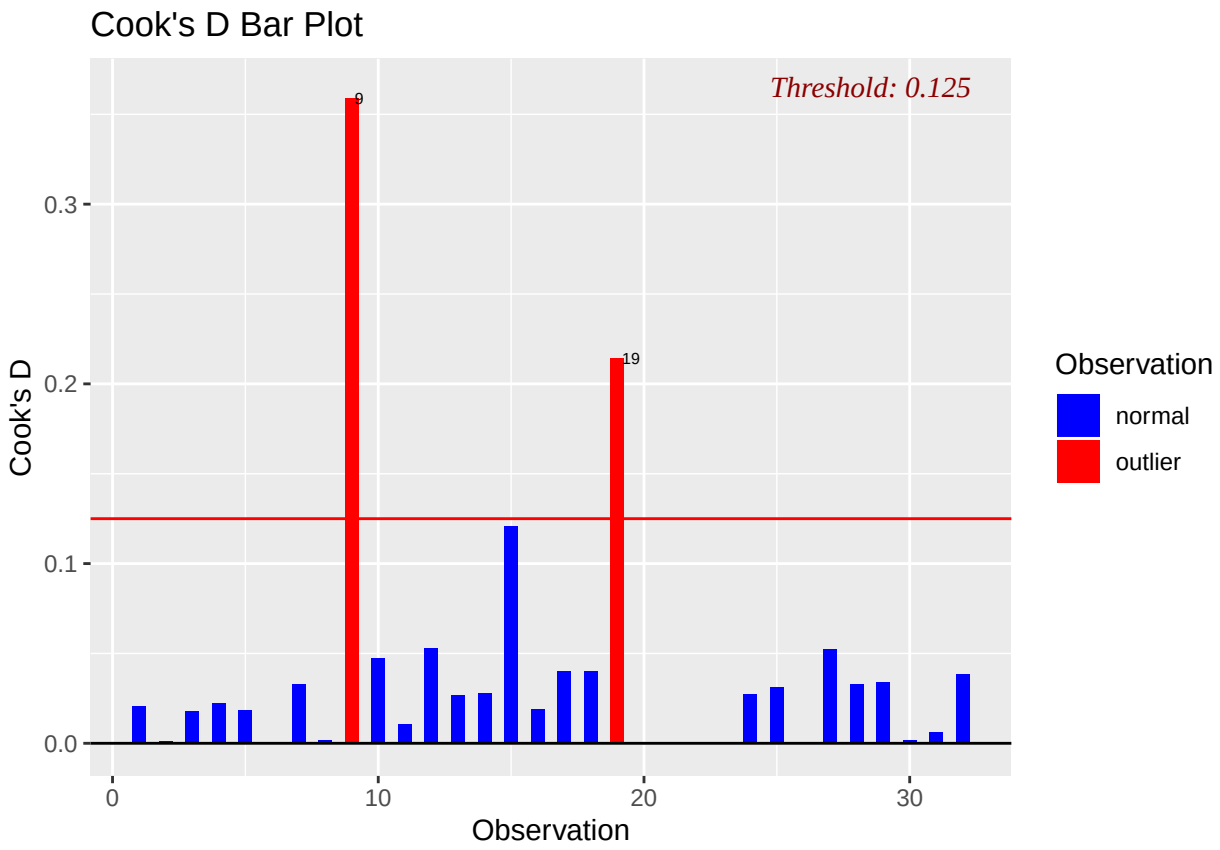

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:olsrr':
##
##     cement
```

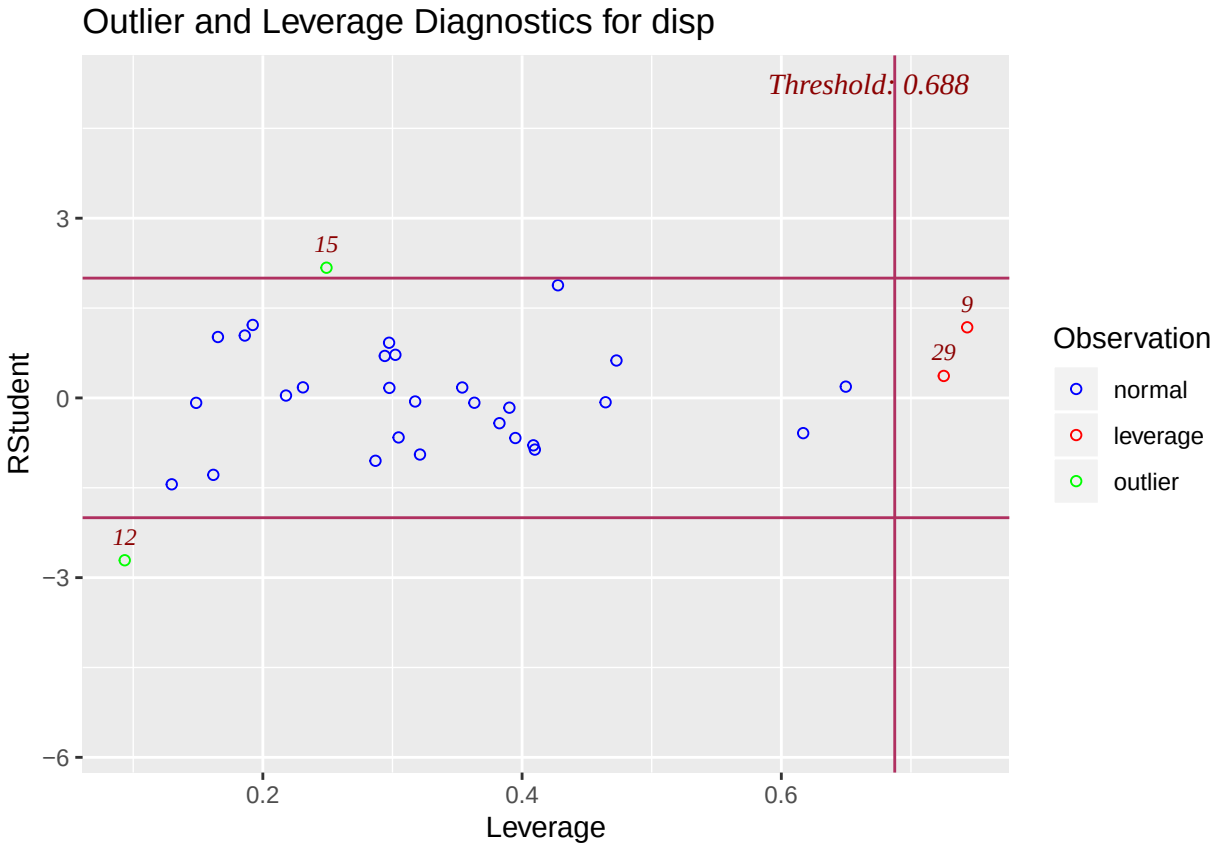

95%
log−Likelihood
10
0
−10
−20
−2
−1
0
1
2
λ
95%
log−Likelihood
15
10
5
−1.0
−0.5
0.0
0.5
1.0
λ

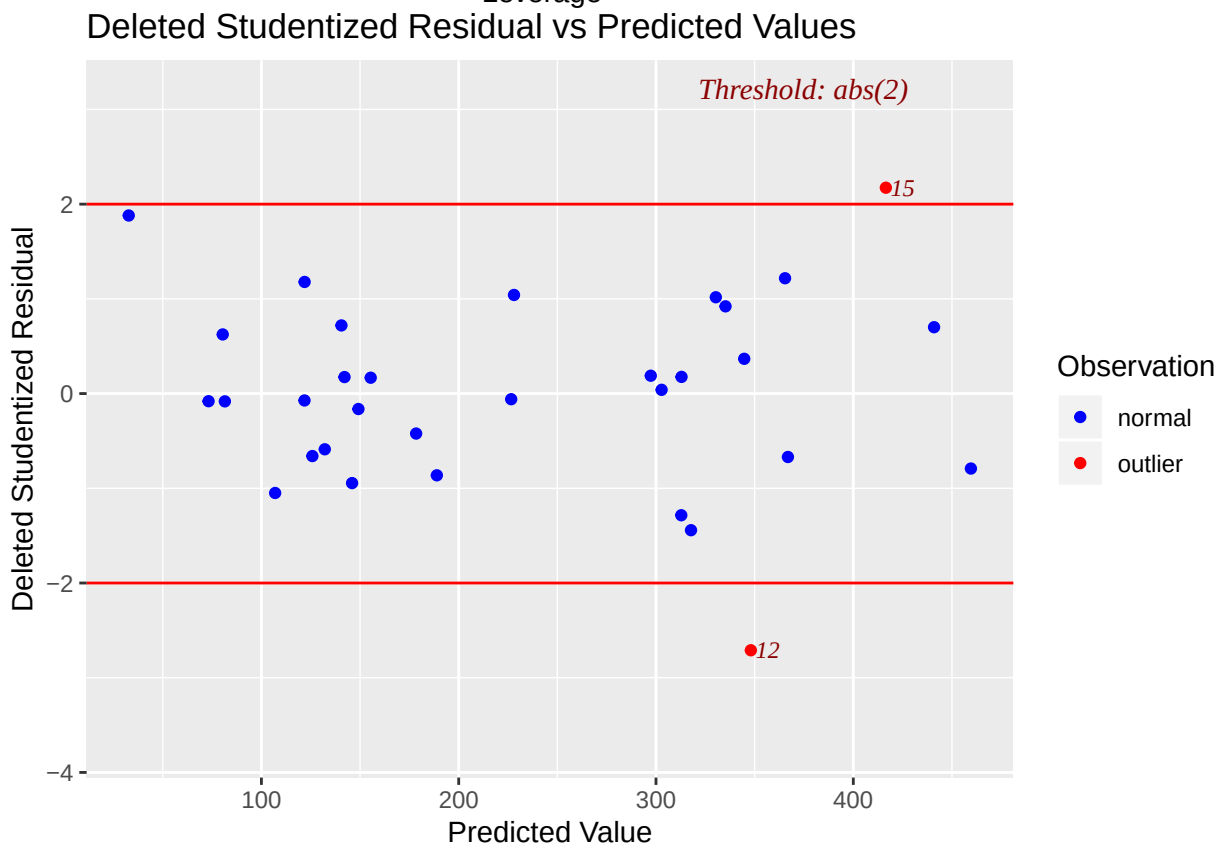

Residuals vs Fitted
Residuals
0.005
−0.005
−0.015
Fiat X1−9
Mazda RX4
Hornet 4 Drive
0.04
0.06
0.08
0.10
Fitted values
lm(1/sqrt(disp) ~ .)
Normal Q−Q
Standardized residuals
2
1
0
−1
−2
Fiat X1−9
Mazda RX4
Hornet 4 Drive
−2
−1
0
1
2
Theoretical Quantiles
lm(1/sqrt(disp) ~ .)


Scale−Location
Hornet 4 Drive
Mazda RX4
Fiat X1−9
√|Standardized residuals|
0.0
0.5
1.0
1.5
0.04
0.06
0.08
0.10
Fitted values
lm(1/sqrt(disp) ~ .)
Residuals vs Leverage
Standardized residuals
2
1
0
−1
−2
−3
1
0.5
Ford Pantera L
Merc 230
0.5
1
Hornet 4 Drive
Cook's distance
0.0
0.2
0.4
0.6
Leverage
lm(1/sqrt(disp) ~ .)

Outliers


Cook's D Bar Plot
Threshold: 0.125
9
19
0.3
0.2
0.1
0.0
Cook's D
0
10
20
30
Observation
Observation
normal
outlier


Outlier and Leverage Diagnostics for disp
Threshold: 0.688
15
9
29
12
3
0
−3
−6
RStudent
0.2
0.4
0.6
Leverage
Observation
normal
leverage
outlier

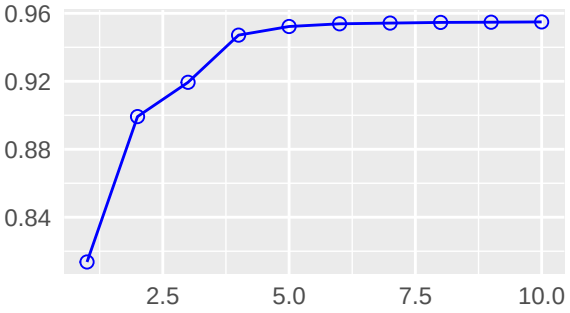

Deleted Studentized Residual vs Predicted Values
Threshold: abs(2)
15
12
2
0
−2
−4
Deleted Studentized Residual
100
200
300
400
Predicted Value
Observation
normal
outlier

Initial selection

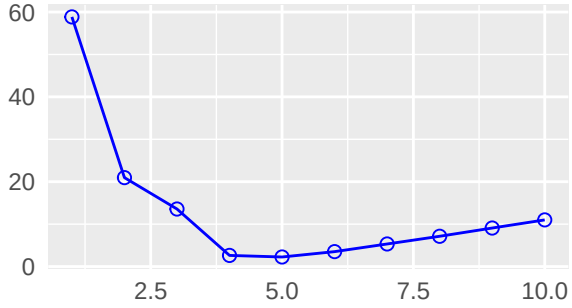
```
##                Best Subsets Regression
## ----------------------------------------------------
## Model Index    Predictors
## ----------------------------------------------------
##      1         cyl
##      2         cyl wt
##      3         hp wt carb
##      4         hp wt qsec carb
##      5         cyl hp wt qsec carb
##      6         mpg cyl hp wt qsec carb
##      7         mpg cyl hp drat wt qsec carb
##      8         mpg cyl hp drat wt qsec vs carb
##      9         mpg cyl hp drat wt qsec vs am carb
##     10         mpg cyl hp drat wt qsec vs am gear carb
## ----------------------------------------------------
##
##                                                    Subsets Regression Summary
## ------------------------------------------------------------------------------------------------------------
##                        Adj.        Pred
## Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC        SBC         MSEP
## ------------------------------------------------------------------------------------------------------------
##   1        0.8137      0.8075      0.7895    58.8481    350.4961    256.3602    354.8933    3155.2944
##   2        0.8992      0.8923        0.88    20.9648    332.8237    239.8967    338.6867    1828.1659
##   3        0.9194      0.9108      0.8934    13.5635    327.6760    235.6352    335.0047    1570.5180
##   4        0.9472      0.9393      0.9283     2.6229    316.1606    228.1650    324.9550    1108.6660
##   5        0.9522      0.9431      0.9312     2.2560    314.9265    228.9849    325.1867    1082.2631
##   6        0.9538      0.9427      0.9289     3.5193    315.8493    231.3814    327.5752    1133.6400
##   7        0.9542      0.9409      0.9256     5.3235    317.5568    234.2545    330.7484    1221.0050
##   8        0.9546      0.9389      0.9199     7.1383    319.2778    237.1867    333.9351    1320.4415
##   9        0.9548      0.9363      0.9159     9.0806    321.1902    240.2012    337.3133    1442.2473
##  10        0.9549      0.9335      0.9076    11.0000    323.0677    243.2306    340.6565    1580.4060
## ------------------------------------------------------------------------------------------------------------
## AIC: Akaike Information Criteria
##  SBIC: Sawa's Bayesian Information Criteria
##  SBC: Schwarz Bayesian Criteria
##  MSEP: Estimated error of prediction, assuming multivariate normality
##  FPE: Final Prediction Error
##  HSP: Hocking's Sp
##  APC: Amemiya Prediction Criteria
```

page 1 of 2

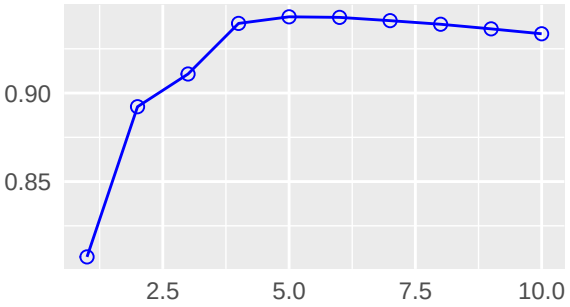
R−Square



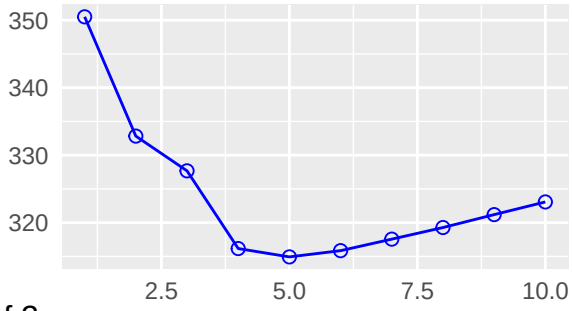
C(p)



Adj. R−Square

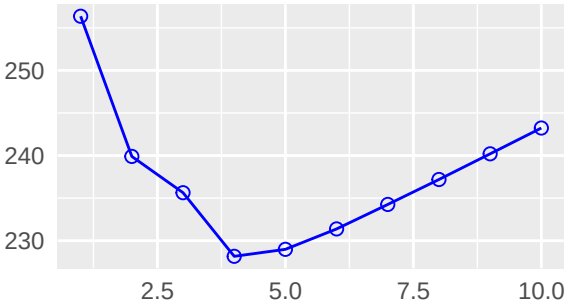


AIC



page 2 of 2

SBIC



SBC