Topic 4.2 Techniques for Choosing Predictors

Load needed packages.

```
library(Stat2Data)
library(leaps)
library(HH)
```

EXAMPLE 4.2 First-year GPA

Load **FirstYearGPA**data from Stat2Data package and look at the structure of the data.

```
data(FirstYearGPA)
str(FirstYearGPA)
```

```
## 'data.frame':    219 obs. of  10 variables:
##  $ GPA          : num  3.06 4.15 3.41 3.21 3.48 2.95 3.6 2.87 3.67 3.49 ...
##  $ HSGPA        : num  3.83 4 3.7 3.51 3.83 3.25 3.79 3.6 3.36 3.7 ...
##  $ SATV         : int  680 740 640 740 610 600 710 390 630 680 ...
##  $ SATM         : int  770 720 570 700 610 570 630 570 560 670 ...
##  $ Male         : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ HU           : num  3 9 16 22 30.5 18 5 10 8.5 16 ...
##  $ SS           : num  9 3 13 0 1.5 3 19 0 15.5 12 ...
##  $ FirstGen     : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ White        : int  1 1 0 1 1 1 1 0 1 1 ...
##  $ CollegeBound: int  1 1 1 1 1 1 1 0 1 1 ...
```

FIGURE 4.4 Scatterplot matrix for first-year GPA data

For the quantitative variables

```
pairs(FirstYearGPA[,c(1,2,3,4,6,7)], pch=16)
```

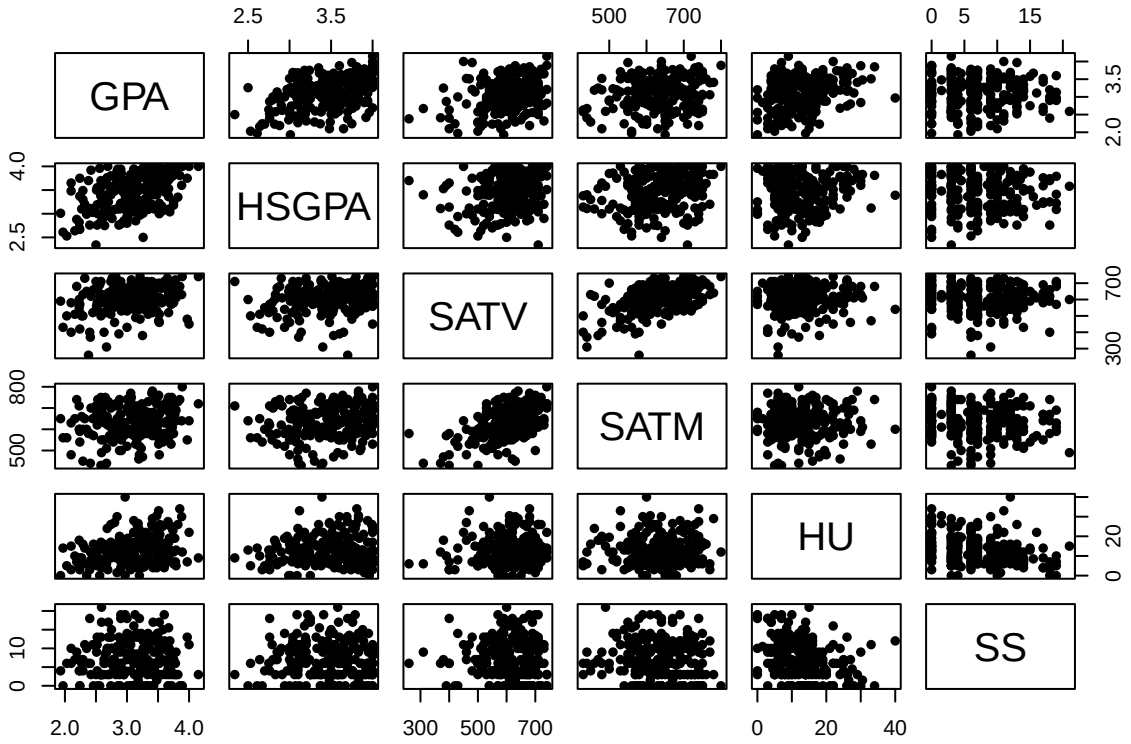
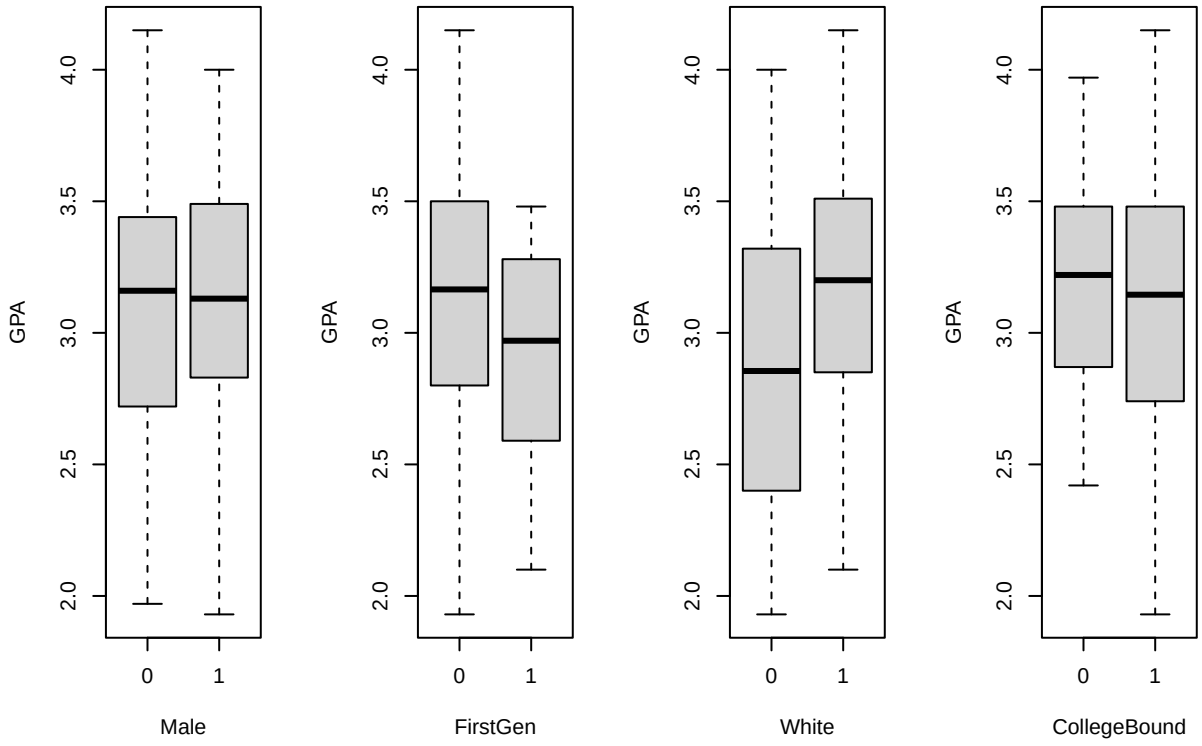


FIGURE 4.5 GPA versus categorical predictors

For each categorical predictor 0 = no, 1 = yes

```
par(mfrow=c(1,4))     #puts all four plots side-by-side
boxplot(GPA~Male,data=FirstYearGPA,xlab="Male")
boxplot(GPA~FirstGen,data=FirstYearGPA,xlab="FirstGen")
boxplot(GPA~White,data=FirstYearGPA,xlab="White")
boxplot(GPA~CollegeBound,data=FirstYearGPA,xlab="CollegeBound")
```



BEST SUBSETS

Using the leaps package, run best subsets for predicting GPA. The regsubsets( ) function finds the best model of each size (or use nbest = _____ to show more models of each size). The command below shows the best two models of each size.

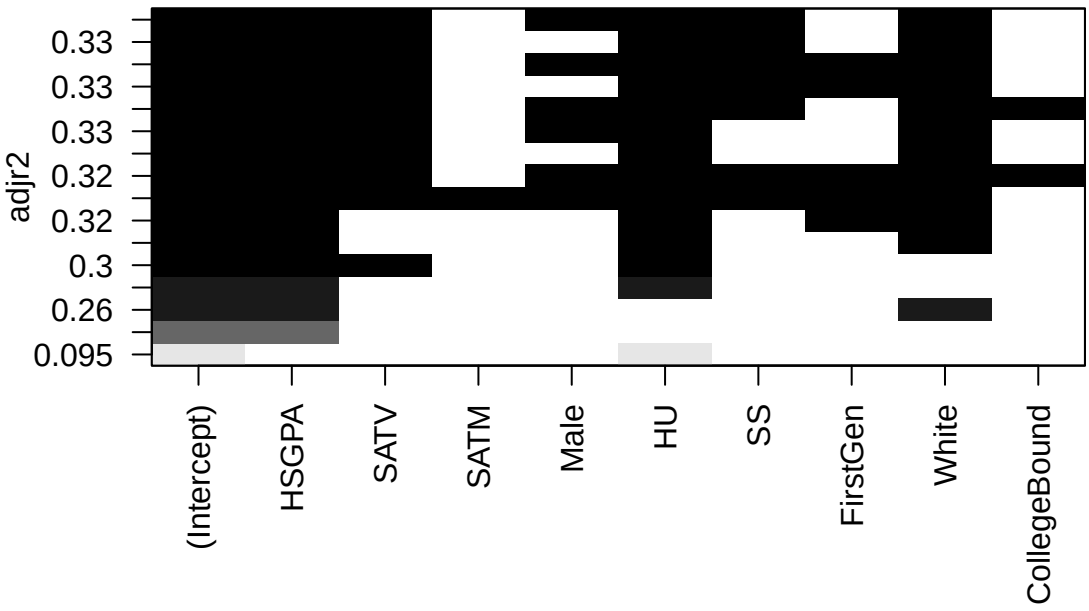
```
all=regsubsets(GPA~.,nbest=2,data=FirstYearGPA)
summary(all)
```

```
## Subset selection object
## Call: regsubsets.formula(GPA ~ ., nbest = 2, data = FirstYearGPA)
## 9 Variables  (and intercept)
##              Forced in Forced out
## HSGPA            FALSE      FALSE
## SATV             FALSE      FALSE
## SATM             FALSE      FALSE
## Male             FALSE      FALSE
## HU               FALSE      FALSE
## SS               FALSE      FALSE
## FirstGen         FALSE      FALSE
## White            FALSE      FALSE
## CollegeBound     FALSE      FALSE
## 2 subsets of each size up to 8
```

```
## Selection Algorithm: exhaustive
##          HSGPA SATV SATM Male HU  SS  FirstGen White CollegeBound
## 1  ( 1 ) "*"   " "  " "  " "  " " " " " " "       " "   " "
## 1  ( 2 ) " "   " "  " "  " "  "*" " " " " "       " "   " "
## 2  ( 1 ) "*"   " "  " "  " "  "*" " " " " "       " "   " "
## 2  ( 2 ) "*"   " "  " "  " "  " " " " " " "       "*"   " "
## 3  ( 1 ) "*"   " "  " "  " "  "*" " " " " "       "*"   " "
## 3  ( 2 ) "*"   "*"  " "  " "  "*" " " " " "       " "   " "
## 4  ( 1 ) "*"   "*"  " "  " "  "*" " " " " "       "*"   " "
## 4  ( 2 ) "*"   " "  " "  " "  "*" " " " " "*"     "*"   " "
## 5  ( 1 ) "*"   "*"  " "  " "  "*" "*" " " "       "*"   " "
## 5  ( 2 ) "*"   "*"  " "  "*"  "*" " " " " "       "*"   " "
## 6  ( 1 ) "*"   "*"  " "  "*"  "*" "*" " " "       "*"   " "
## 6  ( 2 ) "*"   "*"  " "  " "  "*" "*" "*" "       "*"   " "
## 7  ( 1 ) "*"   "*"  " "  "*"  "*" "*" "*" "       "*"   " "
## 7  ( 2 ) "*"   "*"  " "  "*"  "*" "*" " " "       "*"   "*"
## 8  ( 1 ) "*"   "*"  " "  "*"  "*" "*" "*" "       "*"   "*"
## 8  ( 2 ) "*"   "*"  "*"  "*"  "*" "*" "*" "       "*"   " "
```

Here's an option to create a plot showing the variables in each model

```
plot(all,scale="adjr2")
```



The HH package has a nicer function for displaying the regsubsets results.

```
summaryHH(all)
```

```
##                        model p   rsq  rss  adjr2    cp   bic stderr
## 1                         HS 2 0.200 37.8 0.1960 42.18 -38.0  0.417
## 2                         HU 2 0.099 42.6 0.0949 74.54 -12.1  0.443
## 3                      HS-HU 3 0.270 34.5 0.2630 21.68 -52.7  0.400
## 4                       HS-W 3 0.268 34.6 0.2613 22.21 -52.2  0.400
## 5                    HS-HU-W 4 0.323 32.0 0.3136  6.53 -63.9  0.386
## 6                 HS-SATV-HU 4 0.308 32.7 0.2983 11.41 -59.0  0.390
## 7               HS-SATV-HU-W 5 0.337 31.3 0.3251  3.90 -63.2  0.382
## 8                  HS-HU-F-W 5 0.330 31.7 0.3171  6.43 -60.6  0.385
## 9            HS-SATV-HU-SS-W 6 0.344 31.0 0.3283  3.89 -59.9  0.381
## 10            HS-SATV-M-HU-W 6 0.341 31.1 0.3256  4.76 -59.0  0.382
## 11         HS-SATV-M-HU-SS-W 7 0.347 30.8 0.3285  4.85 -55.6  0.381
## 12         HS-SATV-HU-SS-F-W 7 0.346 30.9 0.3278  5.08 -55.4  0.382
## 13       HS-SATV-M-HU-SS-F-W 8 0.349 30.7 0.3278  6.08 -51.0  0.382
## 14       HS-SATV-M-HU-SS-W-C 8 0.347 30.8 0.3256  6.77 -50.3  0.382
## 15    HS-SATV-M-HU-SS-F-W-C 9 0.350 30.7 0.3247  8.04 -45.7  0.383
## 16 HS-SATV-SATM-M-HU-SS-F-W 9 0.349 30.7 0.3247  8.05 -45.7  0.383
##
## Model variables with abbreviations
##                                                                    model
## HS                                                                 HSGPA
## HU                                                                    HU
## HS-HU                                                           HSGPA-HU
## HS-W                                                         HSGPA-White
## HS-HU-W                                                   HSGPA-HU-White
## HS-SATV-HU                                                 HSGPA-SATV-HU
## HS-SATV-HU-W                                            HSGPA-SATV-HU-White
## HS-HU-F-W                                             HSGPA-HU-FirstGen-White
## HS-SATV-HU-SS-W                                        HSGPA-SATV-HU-SS-White
## HS-SATV-M-HU-W                                       HSGPA-SATV-Male-HU-White
## HS-SATV-M-HU-SS-W                                 HSGPA-SATV-Male-HU-SS-White
## HS-SATV-HU-SS-F-W                              HSGPA-SATV-HU-SS-FirstGen-White
## HS-SATV-M-HU-SS-F-W                         HSGPA-SATV-Male-HU-SS-FirstGen-White
## HS-SATV-M-HU-SS-W-C                      HSGPA-SATV-Male-HU-SS-White-CollegeBound
## HS-SATV-M-HU-SS-F-W-C    HSGPA-SATV-Male-HU-SS-FirstGen-White-CollegeBound
## HS-SATV-SATM-M-HU-SS-F-W         HSGPA-SATV-SATM-Male-HU-SS-FirstGen-White
##
## model with largest adjr2
## 11
##
## Number of observations
## 219
```

Regression output for the six predictor model (the model with the best adjusted R^2).

```
mod6=lm(GPA~HSGPA+SATV+Male+HU+SS+White,data=FirstYearGPA)
summary(mod6)
```

```
##
## Call:
```

```
## lm(formula = GPA ~ HSGPA + SATV + Male + HU + SS + White, data = FirstYearGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06228 -0.26731  0.05287  0.27230  0.85843
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5466634  0.2835072   1.928   0.0552 .
## HSGPA       0.4829491  0.0714659   6.758 1.33e-10 ***
## SATV        0.0006945  0.0003449   2.013   0.0453 *
## Male        0.0541049  0.0526937   1.027   0.3057
## HU          0.0167958  0.0038181   4.399 1.72e-05 ***
## SS          0.0075702  0.0054421   1.391   0.1657
## White       0.2045215  0.0685954   2.982   0.0032 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3814 on 212 degrees of freedom
## Multiple R-squared:  0.347,  Adjusted R-squared:  0.3285
## F-statistic: 18.78 on 6 and 212 DF,  p-value: < 2.2e-16
```

Perhaps we don't need Male or SS in this model?

Mallow's Cp is given in the summaryHH( ) and SHowSubsets( ) output (see alternative solutions below). It's also stored in the summary of the regsubsets object, but you need to be careful to track which model corresponds to each Cp value.

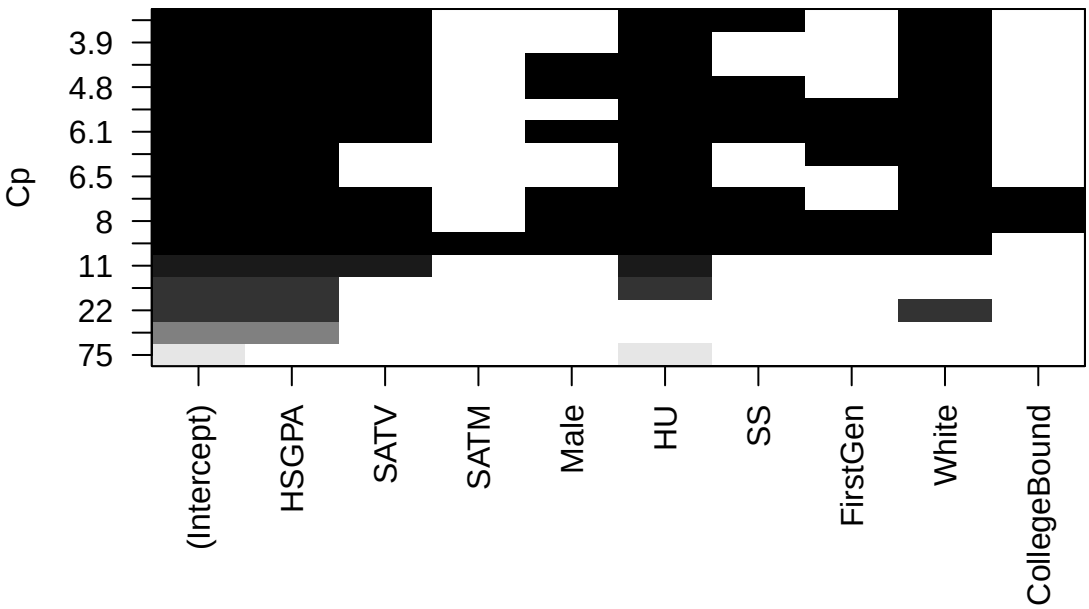
```
summary(all)$cp
```

```
##  [1] 42.180513 74.541330 21.682884 22.210037  6.530981 11.406836  3.900456
##  [8]  6.425605  3.892373  4.764152  4.848791  5.080066  6.080830  6.765744
## [15]  8.036280  8.045716
```

In the ouptut above, the smallest Cp is for the ninth model (3.89), which is the first five-predictor model.

You can also modify the plot of the regsubsets results to sort by Cp.

```
plot(all,scale="Cp")
```



Smallest Cp has HSGPA,SATV,HU,SS, and White as predictors

To calculate Cp for any particular model, we can use extractAIC since Cp is equivalent to AIC for regression models. Here is Cp for the six-predictor model (4.85).

```
full=lm(GPA~.,data=FirstYearGPA)    #model with all predictors in the pool
MSE=(summary(full)$sigma)^2         #Get out the MSE for full model
extractAIC(lm(GPA~HSGPA+SATV+Male+HU+SS+White,data=FirstYearGPA),scale=MSE)
```

```
## [1] 7.000000 4.848791
```

BACKWARD ELIMINATION (with the GPA data)

First, we'll use "brute force" by fitting each model. Code to do this more automaticaly using the step() function appears later.

Full model with nine predictors

```
mod9=lm(GPA~.,data=FirstYearGPA)
summary(mod9)
```

```
##
## Call:
## lm(formula = GPA ~ ., data = FirstYearGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.07412 -0.25827  0.05384  0.27675  0.85761
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5268983  0.3487584   1.511  0.13235
## HSGPA         0.4932945  0.0745553   6.616 3.03e-10 ***
## SATV          0.0005919  0.0003945   1.501  0.13498
## SATM          0.0000847  0.0004447   0.190  0.84912
## Male          0.0482478  0.0570277   0.846  0.39850
## HU            0.0161874  0.0039723   4.075 6.53e-05 ***
## SS            0.0073370  0.0055635   1.319  0.18869
## FirstGen     -0.0743417  0.0887490  -0.838  0.40318
## White         0.1962316  0.0700182   2.803  0.00555 **
## CollegeBound  0.0214530  0.1003350   0.214  0.83090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3834 on 209 degrees of freedom
## Multiple R-squared:  0.3496, Adjusted R-squared:  0.3216
## F-statistic: 12.48 on 9 and 209 DF,  p-value: 8.674e-16
```

Weakest (highest) P-value is for SATM so drop it, using the update function.

```
mod8=update(mod9,.~.-SATM)
summary(mod8)
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + SATV + Male + HU + SS + FirstGen +
##     White + CollegeBound, data = FirstYearGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.07160 -0.26357  0.05167  0.27469  0.85550
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5551540  0.3149111   1.763  0.07937 .
## HSGPA         0.4950161  0.0738354   6.704 1.84e-10 ***
## SATV          0.0006245  0.0003548   1.760  0.07988 .
## Male          0.0522103  0.0529758   0.986  0.32549
## HU            0.0160823  0.0039247   4.098 5.96e-05 ***
## SS            0.0071772  0.0054873   1.308  0.19231
## FirstGen     -0.0755918  0.0883026  -0.856  0.39294
## White         0.1974200  0.0695794   2.837  0.00499 **
## CollegeBound  0.0211753  0.1000939   0.212  0.83266
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3825 on 210 degrees of freedom
## Multiple R-squared:  0.3495, Adjusted R-squared:  0.3247
## F-statistic: 14.11 on 8 and 210 DF,  p-value: 2.253e-16
```

CollegeBound should be dropped next, so use the update function again.

```
mod7=update(mod8,.~.-CollegeBound)
summary(mod7)
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + SATV + Male + HU + SS + FirstGen +
##     White, data = FirstYearGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06911 -0.26259  0.05236  0.26954  0.84134
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.5824756  0.2865599   2.033  0.04334 *
## HSGPA        0.4919452  0.0722304   6.811 9.94e-11 ***
## SATV         0.0006315  0.0003524   1.792  0.07458 .
## Male         0.0529590  0.0527377   1.004  0.31643
## HU           0.0160503  0.0039129   4.102 5.85e-05 ***
## SS           0.0071224  0.0054687   1.302  0.19420
## FirstGen    -0.0772533  0.0877533  -0.880  0.37967
## White        0.1963878  0.0692509   2.836  0.00501 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3816 on 211 degrees of freedom
## Multiple R-squared:  0.3494, Adjusted R-squared:  0.3278
## F-statistic: 16.19 on 7 and 211 DF,  p-value: < 2.2e-16
```

Now, drop FirstGen.

```
mod6=update(mod7,.~.-FirstGen)
summary(mod6)
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + SATV + Male + HU + SS + White, data = FirstYearGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06228 -0.26731  0.05287  0.27230  0.85843
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 0.5466634  0.2835072   1.928   0.0552 .
## HSGPA       0.4829491  0.0714659   6.758 1.33e-10 ***
## SATV        0.0006945  0.0003449   2.013   0.0453 *
## Male        0.0541049  0.0526937   1.027   0.3057
## HU          0.0167958  0.0038181   4.399 1.72e-05 ***
## SS          0.0075702  0.0054421   1.391   0.1657
## White       0.2045215  0.0685954   2.982   0.0032 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3814 on 212 degrees of freedom
## Multiple R-squared:  0.347,  Adjusted R-squared:  0.3285
## F-statistic: 18.78 on 6 and 212 DF,  p-value: < 2.2e-16
```

Next, drop Male.

```
mod5=update(mod6,.~.-Male)
summary(mod5)
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + SATV + HU + SS + White, data = FirstYearGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08660 -0.25827  0.04326  0.25822  0.87954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5684876  0.2827454   2.011  0.04563 *
## HSGPA       0.4739983  0.0709413   6.682 2.03e-10 ***
## SATV        0.0007481  0.0003410   2.194  0.02932 *
## HU          0.0167447  0.0038183   4.385 1.82e-05 ***
## SS          0.0077474  0.0054401   1.424  0.15587
## White       0.2060408  0.0685881   3.004  0.00298 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3815 on 213 degrees of freedom
## Multiple R-squared:  0.3437, Adjusted R-squared:  0.3283
## F-statistic: 22.31 on 5 and 213 DF,  p-value: < 2.2e-16
```

Next, drop SS.

```
mod4=update(mod5,.~.-SS)
summary(mod4)
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + SATV + HU + White, data = FirstYearGPA)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -1.06370 -0.26286  0.02436  0.27338  0.87190
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6409767  0.2787933   2.299  0.02246 *
## HSGPA       0.4761952  0.0710947   6.698 1.83e-10 ***
## SATV        0.0007372  0.0003417   2.157  0.03209 *
## HU          0.0150566  0.0036383   4.138 5.03e-05 ***
## White       0.2121164  0.0686196   3.091  0.00226 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3824 on 214 degrees of freedom
## Multiple R-squared:  0.3375, Adjusted R-squared:  0.3251
## F-statistic: 27.25 on 4 and 214 DF,  p-value: < 2.2e-16
```

The weakest predictor now is SATV, but it's P-value is less than 0.05, so we stop, and the backward elimination model is GPA~HSGPA+SATV+HU+White.

FORWARD SELECTION (using the GPA data)

We'll do this with the step function. We need to specify the full model to give a pool of predictors and a starting point (none) that has just the intercept.

```
full=lm(GPA~.,data=FirstYearGPA)
none=lm(GPA~1,data=FirstYearGPA)
step(none,scope=list(upper=full),direction="forward")
```

```
## Start:  AIC=-333.94
## GPA ~ 1
##
##                Df Sum of Sq    RSS     AIC
## + HSGPA         1    9.4329 37.801 -380.73
## + HU            1    4.6765 42.557 -354.77
## + SATV          1    4.3741 42.859 -353.22
## + White         1    3.7501 43.483 -350.06
## + SATM          1    1.7840 45.450 -340.37
## + FirstGen      1    1.1580 46.076 -337.37
## <none>                      47.234 -333.94
## + CollegeBound  1    0.1876 47.046 -332.81
## + Male          1    0.1319 47.102 -332.55
## + SS            1    0.0006 47.233 -331.94
##
## Step:  AIC=-380.73
## GPA ~ HSGPA
##
##                Df Sum of Sq    RSS     AIC
## + HU            1    3.3067 34.494 -398.78
## + White         1    3.2292 34.571 -398.28
## + SATV          1    2.1861 35.615 -391.77
```

```
## + FirstGen      1    1.6278 36.173 -388.37
## + SATM          1    0.7683 37.032 -383.22
## + Male          1    0.4138 37.387 -381.14
## <none>                      37.801 -380.73
## + CollegeBound  1    0.0342 37.766 -378.93
## + SS            1    0.0008 37.800 -378.73
##
## Step:  AIC=-398.78
## GPA ~ HSGPA + HU
##
##                Df Sum of Sq    RSS     AIC
## + White         1   2.52100 31.973 -413.40
## + SATV          1   1.80435 32.690 -408.54
## + SATM          1   0.86034 33.634 -402.31
## + FirstGen      1   0.80022 33.694 -401.92
## + Male          1   0.43380 34.060 -399.55
## + SS            1   0.37935 34.115 -399.20
## <none>                      34.494 -398.78
## + CollegeBound  1   0.03905 34.455 -397.02
##
## Step:  AIC=-413.4
## GPA ~ HSGPA + HU + White
##
##                Df Sum of Sq    RSS     AIC
## + SATV          1   0.68060 31.292 -416.11
## + FirstGen      1   0.30945 31.663 -413.53
## <none>                      31.973 -413.40
## + SATM          1   0.28236 31.691 -413.34
## + Male          1   0.27919 31.694 -413.32
## + SS            1   0.27526 31.698 -413.29
## + CollegeBound  1   0.04854 31.924 -411.73
##
## Step:  AIC=-416.11
## GPA ~ HSGPA + HU + White + SATV
##
##                Df Sum of Sq    RSS     AIC
## + SS            1  0.295150 30.997 -416.18
## <none>                      31.292 -416.11
## + Male          1  0.167015 31.125 -415.28
## + FirstGen      1  0.156003 31.136 -415.20
## + SATM          1  0.026915 31.265 -414.30
## + CollegeBound  1  0.013720 31.279 -414.20
##
## Step:  AIC=-416.18
## GPA ~ HSGPA + HU + White + SATV + SS
##
##                Df Sum of Sq    RSS     AIC
## <none>                      30.997 -416.18
## + Male          1  0.153387 30.844 -415.27
## + FirstGen      1  0.119394 30.878 -415.03
## + SATM          1  0.054109 30.943 -414.57
## + CollegeBound  1  0.018808 30.978 -414.32


##
```

```
## Call:
## lm(formula = GPA ~ HSGPA + HU + White + SATV + SS, data = FirstYearGPA)
##
## Coefficients:
## (Intercept)        HSGPA           HU        White         SATV           SS
##   0.5684876    0.4739983    0.0167447    0.2060408    0.0007481    0.0077474
```

STEPWISE REGRESSION

This automated option combines forward and backward, having the same syntax as forward without needing a direction specified. The formula(full) option specifies the pool of predictors. Leaving out a direction gives the default direction = "both."

```
full=lm(GPA~.,data=FirstYearGPA)
none=lm(GPA~1,data=FirstYearGPA)
step(none,scope=formula(full))
```

```
## Start:  AIC=-333.94
## GPA ~ 1
##
##                Df Sum of Sq    RSS     AIC
## + HSGPA         1    9.4329 37.801 -380.73
## + HU            1    4.6765 42.557 -354.77
## + SATV          1    4.3741 42.859 -353.22
## + White         1    3.7501 43.483 -350.06
## + SATM          1    1.7840 45.450 -340.37
## + FirstGen      1    1.1580 46.076 -337.37
## <none>                      47.234 -333.94
## + CollegeBound  1    0.1876 47.046 -332.81
## + Male          1    0.1319 47.102 -332.55
## + SS            1    0.0006 47.233 -331.94
##
## Step:  AIC=-380.73
## GPA ~ HSGPA
##
##                Df Sum of Sq    RSS     AIC
## + HU            1    3.3067 34.494 -398.78
## + White         1    3.2292 34.571 -398.28
## + SATV          1    2.1861 35.615 -391.77
## + FirstGen      1    1.6278 36.173 -388.37
## + SATM          1    0.7683 37.032 -383.22
## + Male          1    0.4138 37.387 -381.14
## <none>                      37.801 -380.73
## + CollegeBound  1    0.0342 37.766 -378.93
## + SS            1    0.0008 37.800 -378.73
## - HSGPA         1    9.4329 47.234 -333.94
##
## Step:  AIC=-398.78
## GPA ~ HSGPA + HU
##
##                Df Sum of Sq    RSS     AIC
## + White         1    2.5210 31.973 -413.40
```

```
## + SATV          1    1.8043 32.690 -408.54
## + SATM          1    0.8603 33.634 -402.31
## + FirstGen      1    0.8002 33.694 -401.92
## + Male          1    0.4338 34.060 -399.55
## + SS            1    0.3793 34.115 -399.20
## <none>                      34.494 -398.78
## + CollegeBound  1    0.0390 34.455 -397.02
## - HU            1    3.3067 37.801 -380.73
## - HSGPA         1    8.0631 42.557 -354.77
##
## Step:  AIC=-413.4
## GPA ~ HSGPA + HU + White
##
##                Df Sum of Sq    RSS     AIC
## + SATV          1    0.6806 31.292 -416.11
## + FirstGen      1    0.3095 31.663 -413.53
## <none>                      31.973 -413.40
## + SATM          1    0.2824 31.691 -413.34
## + Male          1    0.2792 31.694 -413.32
## + SS            1    0.2753 31.698 -413.29
## + CollegeBound  1    0.0485 31.924 -411.73
## - White         1    2.5210 34.494 -398.78
## - HU            1    2.5985 34.571 -398.28
## - HSGPA         1    7.7700 39.743 -367.75
##
## Step:  AIC=-416.11
## GPA ~ HSGPA + HU + White + SATV
##
##                Df Sum of Sq    RSS     AIC
## + SS            1    0.2951 30.997 -416.18
## <none>                      31.292 -416.11
## + Male          1    0.1670 31.125 -415.28
## + FirstGen      1    0.1560 31.136 -415.20
## + SATM          1    0.0269 31.265 -414.30
## + CollegeBound  1    0.0137 31.279 -414.20
## - SATV          1    0.6806 31.973 -413.40
## - White         1    1.3973 32.690 -408.54
## - HU            1    2.5042 33.797 -401.25
## - HSGPA         1    6.5602 37.853 -376.43
##
## Step:  AIC=-416.18
## GPA ~ HSGPA + HU + White + SATV + SS
##
##                Df Sum of Sq    RSS     AIC
## <none>                      30.997 -416.18
## - SS            1    0.2951 31.292 -416.11
## + Male          1    0.1534 30.844 -415.27
## + FirstGen      1    0.1194 30.878 -415.03
## + SATM          1    0.0541 30.943 -414.57
## + CollegeBound  1    0.0188 30.978 -414.32
## - SATV          1    0.7005 31.698 -413.29
## - White         1    1.3133 32.310 -409.10
## - HU            1    2.7987 33.796 -399.25
## - HSGPA         1    6.4968 37.494 -376.51
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + HU + White + SATV + SS, data = FirstYearGPA)
##
## Coefficients:
## (Intercept)        HSGPA           HU        White         SATV           SS
##   0.5684876    0.4739983    0.0167447    0.2060408    0.0007481    0.0077474
```

In this case, stepwise proceeeds same as forward since no variables are dropped at any step.
