

# Lecture- November 15th, 2017

Wendy Rummerfield

11/13/2017

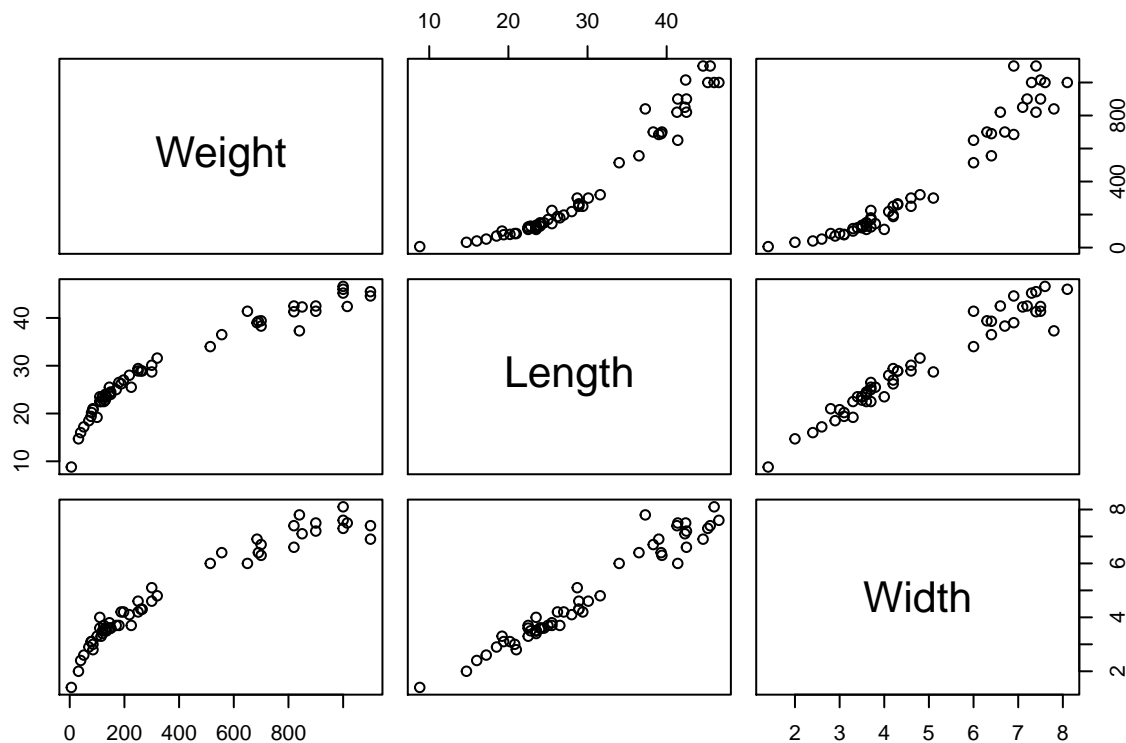
We have information on perch- a type of fish- caught in a lake in Finland. For each of the 56 fish caught we have data on their weight (in grams), length (in cm), and width (in cm). Create a model using the variables collected to predict the weight of perch causing at Lake Laengelmavesi in Finland.

First, let's calculate the correlation between each of the terms using a correlation matrix as well as a scatterplot matrix.

```
cor(Perch[, c(2:4)])
```

```
##           Weight   Length   Width
## Weight 1.0000000 0.9595061 0.9642244
## Length 0.9595061 1.0000000 0.9751074
## Width  0.9642244 0.9751074 1.0000000
```

```
plot(Perch[, c(2:4)])
```



It appears that both width and length are highly correlated with weight; however, when we look at the scatterplot matrix the relationships do not appear to be linear.

```
p1 <- ggplot(Perch) + geom_point(aes(x = Length, y = Weight)) +
  labs(x = "Length", y = "Weight", title = "Scatterplot: Weight vs Length")
```

```
p2 <- ggplot(Perch) + geom_point(aes(x = Length^2, y = Weight)) +
  labs(x = "Length^2", y = "Weight", title = "Scatterplot: Weight vs Length^2")
```

```
p3 <- ggplot(Perch) + geom_point(aes(x = Width, y = Weight)) +
```

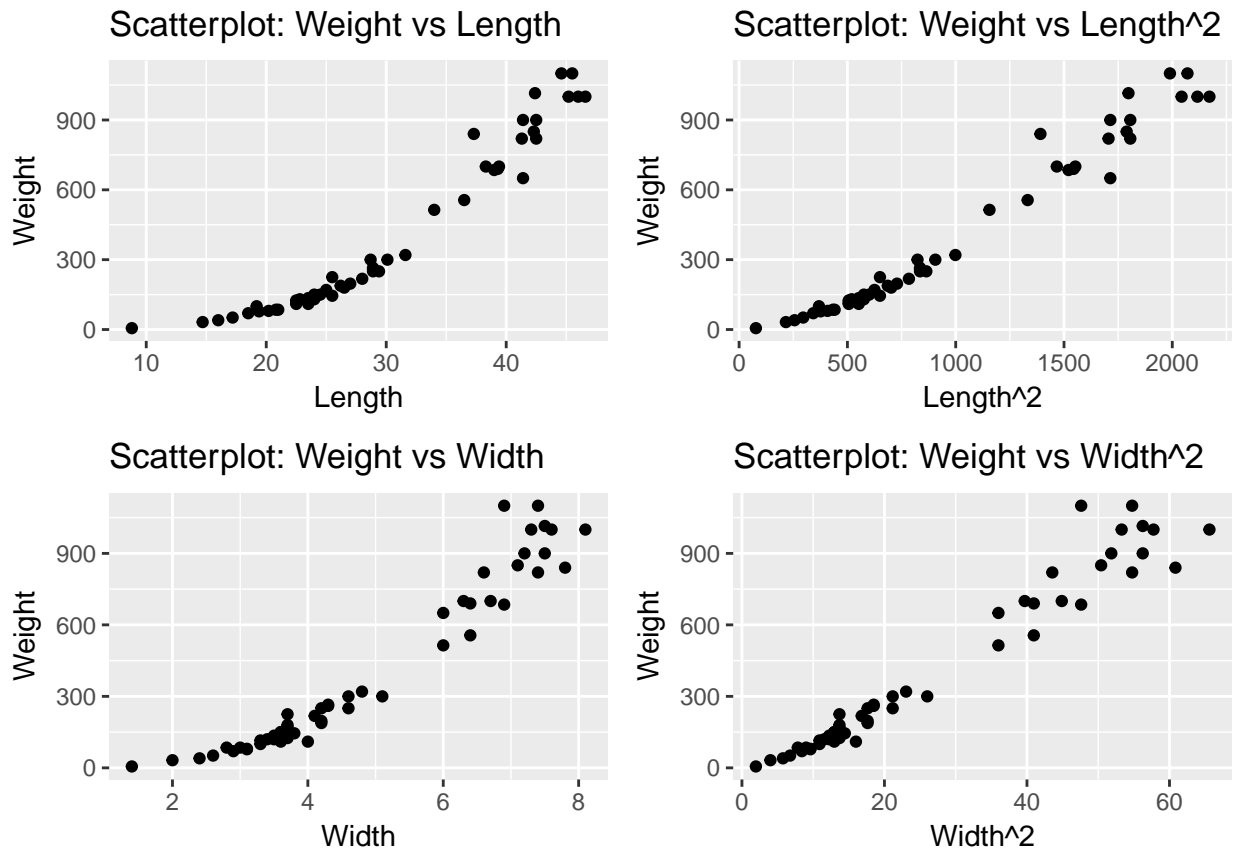
```

labs(x = "Width", y = "Weight", title = "Scatterplot: Weight vs Width")

p4 <- ggplot(Perch) + geom_point(aes(x = Width^2, y = Weight)) +
  labs(x = "Width^2", y = "Weight", title = "Scatterplot: Weight vs Width^2")

grid.arrange(p1,p2,p3,p4)

```



Squaring length and width does make each relationship with height more linear, but there is still some curvature. This may indicate the need to have both predictors in the model and possibly even an interaction between the two.

To test this create the following models to predict weight:

1. Length and width
2. Length up to the second-order and width
3. Quadratic model with length and width
4. Complete model

```

mod1 <- lm(Weight ~ Length + Width, data = Perch)
mod2 <- lm(Weight ~ Length + I(Length^2) + Width, data = Perch)
mod3 <- lm(Weight ~ Length + I(Length^2) + Width + I(Width^2), data = Perch)
mod4 <- lm(Weight ~ Length*Width + I(Length^2) + I(Width^2), data = Perch)

summary(mod1)

```

```

##
## Call:
## lm(formula = Weight ~ Length + Width, data = Perch)

```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -113.86  -59.02  -23.29   30.93  299.85
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -578.758     43.667  -13.254 < 2e-16 ***
## Length       14.307       5.659   2.528 0.014475 *
## Width        113.500     30.265   3.750 0.000439 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 88.68 on 53 degrees of freedom
## Multiple R-squared:  0.9373, Adjusted R-squared:  0.9349
## F-statistic: 396.1 on 2 and 53 DF,  p-value: < 2.2e-16
```

```
summary(mod2)
```

```
##
## Call:
## lm(formula = Weight ~ Length + I(Length^2) + Width, data = Perch)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -110.59  -20.75    2.33   10.32  159.38
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 147.12090    61.25958   2.402  0.0199 *
## Length      -34.71805     4.78840  -7.250 1.97e-09 ***
## I(Length^2)  0.86134     0.06794  12.679 < 2e-16 ***
## Width        91.09772    15.20858   5.990 2.00e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.26 on 52 degrees of freedom
## Multiple R-squared:  0.9847, Adjusted R-squared:  0.9838
## F-statistic: 1114 on 3 and 52 DF,  p-value: < 2.2e-16
```

```
summary(mod3)
```

```
##
## Call:
## lm(formula = Weight ~ Length + I(Length^2) + Width + I(Width^2),
##     data = Perch)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -129.605  -12.121    1.783    9.553   170.034
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 138.0015    60.4435   2.283 0.026621 *
## Length      -15.2436    12.4688  -1.223 0.227124
```

```
## I(Length^2)  0.6065      0.1652   3.672 0.000577 ***
## Width      -31.0365    73.9416  -0.420 0.676436
## I(Width^2)  10.0718     5.9717   1.687 0.097793 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.49 on 51 degrees of freedom
## Multiple R-squared:  0.9855, Adjusted R-squared:  0.9843
## F-statistic: 865.5 on 4 and 51 DF,  p-value: < 2.2e-16
```

```
summary(mod4)
```

```
##
## Call:
## lm(formula = Weight ~ Length * Width + I(Length^2) + I(Width^2),
##     data = Perch)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -117.175  -11.904    2.822   11.556  157.596
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  156.3486    61.4152   2.546  0.0140 *
## Length      -25.0007    14.2729  -1.752  0.0860 .
## Width        20.9772    82.5877   0.254  0.8005
## I(Length^2)   1.5719     0.7244   2.170  0.0348 *
## I(Width^2)   34.4058    18.7455   1.835  0.0724 .
## Length:Width -9.7763     7.1455  -1.368  0.1774
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.13 on 50 degrees of freedom
## Multiple R-squared:  0.986, Adjusted R-squared:  0.9846
## F-statistic: 704.6 on 5 and 50 DF,  p-value: < 2.2e-16
```

```
anova(mod3, mod4)
```

```
## Analysis of Variance Table
##
## Model 1: Weight ~ Length + I(Length^2) + Width + I(Width^2)
## Model 2: Weight ~ Length * Width + I(Length^2) + I(Width^2)
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      51 96482
## 2      50 93000  1    3481.8 1.8719 0.1774
```

## Automatic selection

To perform automatic selection we first need to load the *leaps* package and the *HH* package.

```
# If you haven't already installed the packages, use the following two lines of code
#install.packages("leaps")
#install.packages("HH")
library(leaps)
```

```
library(HH)
```

The `regsubsets` function will show us the best model for each size possible. By including the argument `nbest = 2`, R will output the best two models for each size where the first model of a specific size is the “better” of the two.

```
Best <- regsubsets(Weight ~ Length*Width + I(Length^2) + I(Width^2), data = Perch, nbest = 2)
summary(Best)
```

```
## Subset selection object
## Call: regsubsets.formula(Weight ~ Length * Width + I(Length^2) + I(Width^2),
##   data = Perch, nbest = 2)
## 5 Variables (and intercept)
##           Forced in Forced out
## Length      FALSE      FALSE
## Width        FALSE      FALSE
## I(Length^2)  FALSE      FALSE
## I(Width^2)   FALSE      FALSE
## Length:Width FALSE      FALSE
## 2 subsets of each size up to 5
## Selection Algorithm: exhaustive
##           Length Width I(Length^2) I(Width^2) Length:Width
## 1 ( 1 ) " " " " " " " " " " " "
## 1 ( 2 ) " " " " " " " " " " " "
## 2 ( 1 ) " " " * " " " " " " " "
## 2 ( 2 ) " * " " " " " " " " " "
## 3 ( 1 ) " * " " " " * " " " " "
## 3 ( 2 ) " * " " " " * " " " " "
## 4 ( 1 ) " * " " " " * " " " " "
## 4 ( 2 ) " * " " * " " * " " " "
## 5 ( 1 ) " * " " * " " * " " * " " "
```

Just using the `summary` function, we only see the variables that were selected for each size model. From the `HH` package, we can use the `summaryHH()` function to get much more information including but not limited to Mallows’s  $C_p$ , AIC, and Adjusted  $R^2$ .

```
summaryHH(Best)
```

```
##           model p   rsq   rss adjr2   cp   bic stderr
## 1             L: 2 0.978 147441 0.977 27.27 -205  52.3
## 2             I(L 2 0.967 221059 0.966 66.85 -183  64.0
## 3             W-L: 3 0.984 104154 0.984  6.00 -221  44.3
## 4             Ln-L: 3 0.979 137020 0.979 23.67 -205  50.8
## 5           Ln-I(L-I(W 4 0.985  96815 0.985  4.05 -221  43.1
## 6           Ln-I(L-L: 4 0.985  99270 0.984  5.37 -219  43.7
## 7         Ln-I(L-I(W-L: 5 0.986  93120 0.985  4.06 -219  42.7
## 8           Ln-W-I(L-I(W 5 0.985  96482 0.984  5.87 -217  43.5
## 9         Ln-W-I(L-I(W-L: 6 0.986  93000 0.985  6.00 -215  43.1
##
## Model variables with abbreviations
##
##                                     model
## L:                                     Length:Width
## I(L                                     I(Length^2)
## W-L:                                    Width-Length:Width
## Ln-L:                                   Length-Length:Width
## Ln-I(L-I(W                               Length-I(Length^2)-I(Width^2)
```

```

## Ln-I(L-L:                Length-I(Length^2)-Length:Width
## Ln-I(L-I(W-L:          Length-I(Length^2)-I(Width^2)-Length:Width
## Ln-W-I(L-I(W          Length-Width-I(Length^2)-I(Width^2)
## Ln-W-I(L-I(W-L: Length-Width-I(Length^2)-I(Width^2)-Length:Width
##
## model with largest adjr2
## 7
##
## Number of observations
## 56

```

As you've learned, it is also good to look at adjusted  $R^2$  values when performing model selection. Since nearly all of the models above have an adjusted  $R^2$  greater than .98, it is best to look at Mallows's  $C_p$ . Models 5, 7 and 9 have  $C_p \approx p + 1$ . While clearly the model with all 5 variables has the most accurate  $C_p$ ; however, our nested F test from above revealed that the interaction between length and width does not provide significantly more information given all of the other predictors are in the model. Let's run another nested F test to determine if the 5th or 7th model is better than the other.

```

mod5th <- lm(Weight ~ Length + I(Length^2) + I(Width^2), data = Perch)
mod7th <- lm(Weight ~ Length + I(Length^2) + I(Width^2) + Length:Width, data = Perch)
anova(mod5th, mod7th)

```

```

## Analysis of Variance Table
##
## Model 1: Weight ~ Length + I(Length^2) + I(Width^2)
## Model 2: Weight ~ Length + I(Length^2) + I(Width^2) + Length:Width
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      52 96815
## 2      51 93120  1    3695.1 2.0237 0.1609

```

While the test indicates that we should fail to reject the null hypothesis that the coefficient for the interaction is 0, we still shouldn't choose this model. If you are going to include higher order terms and/ or interaction, you should always keep the main effect in the model as well. With this information, and the nested F test we performed earlier, it appears that model 8 from the best subsets output is the more appropriate model.

This is a great example of when automatic variable selection methods may not be the most useful means for deciding on the optimal model. Let's look at a new dataset.

This data set contains County Demographic Information (CDI). Researchers would like to conduct an exploratory observational study with these data to see which variables help predict the number of active physicians (Physicians) in a county.

```

Best = regsubsets(Num_physicians ~ Location + Population_1990 +
                  Pct_Age18_to_34 + Pct_65_or_over + Num_hospital_beds +
                  Num_serious_crimes + Pct_High_Sch_grads + Pct_Bachelors +
                  Pct_below_poverty + Pct_unemployed + Per_cap_1990income +
                  Total_personal_income + Region_num, nvmax = 15, data = CDI)
summary(Best)

```

```

## Subset selection object
## Call: regsubsets.formula(Num_physicians ~ Location + Population_1990 +
##   Pct_Age18_to_34 + Pct_65_or_over + Num_hospital_beds + Num_serious_crimes +
##   Pct_High_Sch_grads + Pct_Bachelors + Pct_below_poverty +
##   Pct_unemployed + Per_cap_1990income + Total_personal_income +
##   Region_num, nvmax = 15, data = CDI)
## 15 Variables (and intercept)
##               Forced in Forced out

```

```

## LocationWest          FALSE    FALSE
## Population_1990      FALSE    FALSE
## Pct_Age18_to_34     FALSE    FALSE
## Pct_65_or_over      FALSE    FALSE
## Num_hospital_beds   FALSE    FALSE
## Num_serious_crimes  FALSE    FALSE
## Pct_High_Sch_grads  FALSE    FALSE
## Pct_Bachelors       FALSE    FALSE
## Pct_below_poverty   FALSE    FALSE
## Pct_unemployed      FALSE    FALSE
## Per_cap_1990income  FALSE    FALSE
## Total_personal_income FALSE    FALSE
## Region_num2         FALSE    FALSE
## Region_num3         FALSE    FALSE
## Region_num4         FALSE    FALSE
## 1 subsets of each size up to 15
## Selection Algorithm: exhaustive
##      LocationWest Population_1990 Pct_Age18_to_34 Pct_65_or_over
## 1 ( 1 ) " " " " " " " "
## 2 ( 1 ) " " " " " " " "
## 3 ( 1 ) " " "*" " " " " "
## 4 ( 1 ) " " "*" "*" " " " "
## 5 ( 1 ) " " "*" " " " " "
## 6 ( 1 ) " " "*" " " " " "
## 7 ( 1 ) " " "*" " " " " "
## 8 ( 1 ) "*" "*" " " " " " "
## 9 ( 1 ) "*" "*" " " " " " "
## 10 ( 1 ) "*" "*" "*" " " " " "
## 11 ( 1 ) "*" "*" "*" " " " " "
## 12 ( 1 ) "*" "*" "*" " " " " "
## 13 ( 1 ) "*" "*" "*" " " " " "
## 14 ( 1 ) "*" "*" "*" " " " " "
## 15 ( 1 ) "*" "*" "*" "*" " " " "
##      Num_hospital_beds Num_serious_crimes Pct_High_Sch_grads
## 1 ( 1 ) "*" " " " "
## 2 ( 1 ) "*" " " " "
## 3 ( 1 ) "*" " " " "
## 4 ( 1 ) "*" " " " "
## 5 ( 1 ) "*" " " " "
## 6 ( 1 ) "*" " " " "
## 7 ( 1 ) "*" " " "*"
## 8 ( 1 ) "*" " " "*"
## 9 ( 1 ) "*" "*" "*"
## 10 ( 1 ) "*" "*" "*"
## 11 ( 1 ) "*" "*" "*"
## 12 ( 1 ) "*" "*" "*"
## 13 ( 1 ) "*" "*" "*"
## 14 ( 1 ) "*" "*" "*"
## 15 ( 1 ) "*" "*" "*"
##      Pct_Bachelors Pct_below_poverty Pct_unemployed
## 1 ( 1 ) " " " " " "
## 2 ( 1 ) " " " " " "
## 3 ( 1 ) " " " " " "
## 4 ( 1 ) " " " " " "

```

```

## 5 ( 1 ) "*"          " "          " "
## 6 ( 1 ) "*"          " "          " "
## 7 ( 1 ) "*"          " "          " "
## 8 ( 1 ) "*"          " "          " "
## 9 ( 1 ) "*"          " "          " "
## 10 ( 1 ) "*"         " "          " "
## 11 ( 1 ) "*"         " "          "*"
## 12 ( 1 ) "*"         " "          "*"
## 13 ( 1 ) "*"         " "          "*"
## 14 ( 1 ) "*"         "*"         "*"
## 15 ( 1 ) "*"         "*"         "*"
##          Per_cap_1990income Total_personal_income Region_num2 Region_num3
## 1 ( 1 ) " "          " "          " "          " "
## 2 ( 1 ) " "          "*"         " "          " "
## 3 ( 1 ) " "          "*"         " "          " "
## 4 ( 1 ) " "          "*"         " "          " "
## 5 ( 1 ) "*"         "*"         " "          " "
## 6 ( 1 ) "*"         "*"         " "          " "
## 7 ( 1 ) "*"         "*"         " "          " "
## 8 ( 1 ) "*"         "*"         " "          " "
## 9 ( 1 ) "*"         "*"         " "          " "
## 10 ( 1 ) "*"        "*"         " "          " "
## 11 ( 1 ) "*"        "*"         " "          " "
## 12 ( 1 ) "*"        "*"         " "          "*"
## 13 ( 1 ) "*"        "*"         "*"         "*"
## 14 ( 1 ) "*"        "*"         "*"         "*"
## 15 ( 1 ) "*"        "*"         "*"         "*"
##          Region_num4
## 1 ( 1 ) " "
## 2 ( 1 ) " "
## 3 ( 1 ) " "
## 4 ( 1 ) " "
## 5 ( 1 ) " "
## 6 ( 1 ) "*"
## 7 ( 1 ) "*"
## 8 ( 1 ) "*"
## 9 ( 1 ) "*"
## 10 ( 1 ) "*"
## 11 ( 1 ) "*"
## 12 ( 1 ) "*"
## 13 ( 1 ) "*"
## 14 ( 1 ) "*"
## 15 ( 1 ) "*"

```

summaryHH(Best)

```

##          model  p  rsq
## 1          Nm_h_  2 0.903
## 2          Nm_h_-T  3 0.948
## 3          P_1-Nm_h_-T  4 0.955
## 4          P_1-P_A-Nm_h_-T  5 0.958
## 5          P_1-Nm_h_-P_B-P__1-T  6 0.960
## 6          P_1-Nm_h_-P_B-P__1-T-R_4  7 0.961
## 7          P_1-Nm_h_-P_H-P_B-P__1-T-R_4  8 0.962
## 8          L-P_1-Nm_h_-P_H-P_B-P__1-T-R_4  9 0.962

```



```

## 9          L-P_1-Nm_h_-Nm_s_-P_H-P_B-P__1-T-R_4 10 0.962
## 10         L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-P__1-T-R_4 11 0.962
## 11         L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-P__1-T-R_4 12 0.962
## 12         L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-P__1-T-R_3-R_4 13 0.962
## 13         L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-P__1-T-R_2-R_3-R_4 14 0.962
## 14         L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-Pc_-P__1-T-R_2-R_3-R_4 15 0.962
## 15 L-P_1-P_A-P_6-Nm_h_-Nm_s_-P_H-P_B-Pc_-Pc_-P__1-T-R_2-R_3-R_4 16 0.962
##          rss adjr2      cp      bic stderr
## 1  1.36e+08 0.903 652.80 -1016    557
## 2  7.37e+07 0.947 156.78 -1279    411
## 3  6.29e+07 0.955  72.05 -1343    380
## 4  5.92e+07 0.957  44.72 -1363    369
## 5  5.65e+07 0.959  24.64 -1378    361
## 6  5.51e+07 0.960  15.25 -1383    357
## 7  5.40e+07 0.961   8.99 -1385    354
## 8  5.36e+07 0.961   7.32 -1383    353
## 9  5.32e+07 0.961   6.54 -1380    352
## 10 5.30e+07 0.961   7.10 -1375    352
## 11 5.30e+07 0.961   8.84 -1369    352
## 12 5.30e+07 0.961  10.49 -1364    352
## 13 5.29e+07 0.961  12.11 -1358    352
## 14 5.29e+07 0.961  14.01 -1352    353
## 15 5.29e+07 0.961  16.00 -1346    353

```

```
## Model variables with abbreviations
```

```
##
```

```
## Nm_h_
```

```
## Nm_h_-T
```

```
## P_1-Nm_h_-T
```

```
## P_1-P_A-Nm_h_-T
```

```
## P_1-Nm_h_-P_B-P__1-T
```

```
## P_1-Nm_h_-P_B-P__1-T-R_4
```

```
## P_1-Nm_h_-P_H-P_B-P__1-T-R_4
```

```
## L-P_1-Nm_h_-P_H-P_B-P__1-T-R_4
```

```
## L-P_1-Nm_h_-Nm_s_-P_H-P_B-P__1-T-R_4
```

```
## L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-P__1-T-R_4
```

```
## L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-P__1-T-R_4
```

```
## L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-P__1-T-R_3-R_4
```

```
## L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-P__1-T-R_2-R_3-R_4
```

```
## L-P_1-P_A-Nm_h_-Nm_s_-P_H-P_B-Pc_-Pc_-P__1-T-R_2-R_3-R_4
```

```
## L-P_1-P_A-P_6-Nm_h_-Nm_s_-P_H-P_B-Pc_-Pc_-P__1-T-R_2-R_3-R_4 LocationWest-Population_1990-Pct_Age18
```

```
##
```

```
## model with largest adjr2
```

```
## 10
```

```
##
```

```
## Number of observations
```

```
## 440
```

```
Base <- lm(Num_physicians ~ 1, data = CDI)
```

```
Full <- lm(Num_physicians ~ Location + Population_1990 +
```

```
  Pct_Age18_to_34 + Pct_65_or_over + Num_hospital_beds +
```

```
  Num_serious_crimes + Pct_High_Sch_grads + Pct_Bachelors +
```

```
  Pct_below_poverty + Pct_unemployed + Per_cap_1990income +
```

```
  Total_personal_income + Region_num, data = CDI)
```

```
MSE <- (summary(Full)$sigma)^2
step(Full, scale = MSE, direction = "backward") # Backward Elimination
```

```
## Start: AIC=16
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Pct_65_or_over + Num_hospital_beds + Num_serious_crimes +
##   Pct_High_Sch_grads + Pct_Bachelors + Pct_below_poverty +
##   Pct_unemployed + Per_cap_1990income + Total_personal_income +
##   Region_num
##
##           Df Sum of Sq      RSS      Cp
## - Pct_65_or_over      1      1214 52909130 14.010
## - Pct_below_poverty    1     12540 52920455 14.101
## - Pct_unemployed       1     72264 52980179 14.579
## - Pct_Age18_to_34      1    137143 53045058 15.099
## <none>                                52907915 16.000
## - Location             1    324340 53232255 16.599
## - Num_serious_crimes   1    342597 53250512 16.745
## - Pct_High_Sch_grads   1    599909 53507824 18.808
## - Per_cap_1990income   1    848855 53756770 20.803
## - Region_num           3    2109381 55017296 26.904
## - Pct_Bachelors        1    2349330 55257245 32.827
## - Population_1990      1    4224111 57132027 47.852
## - Total_personal_income 1  14355133 67263048 129.041
## - Num_hospital_beds    1  51826830 104734746 429.336
##
## Step: AIC=14.01
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Num_hospital_beds + Num_serious_crimes + Pct_High_Sch_grads +
##   Pct_Bachelors + Pct_below_poverty + Pct_unemployed + Per_cap_1990income +
##   Total_personal_income + Region_num
##
##           Df Sum of Sq      RSS      Cp
## - Pct_below_poverty    1     11952 52921082 12.105
## - Pct_unemployed       1     71416 52980546 12.582
## - Pct_Age18_to_34      1    165349 53074479 13.335
## <none>                                52909130 14.010
## - Location             1    338770 53247900 14.725
## - Num_serious_crimes   1    343334 53252464 14.761
## - Pct_High_Sch_grads   1    603286 53512415 16.844
## - Per_cap_1990income   1    856850 53765980 18.877
## - Region_num           3    2139061 55048190 25.152
## - Pct_Bachelors        1    2349693 55258823 30.840
## - Population_1990      1    4302820 57211949 46.492
## - Total_personal_income 1  14435703 67344833 127.696
## - Num_hospital_beds    1  56006526 108915656 460.842
##
## Step: AIC=12.11
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Num_hospital_beds + Num_serious_crimes + Pct_High_Sch_grads +
##   Pct_Bachelors + Pct_unemployed + Per_cap_1990income + Total_personal_income +
##   Region_num
##
##           Df Sum of Sq      RSS      Cp
```

```

## - Pct_unemployed      1      60656  52981738  10.592
## - Pct_Age18_to_34    1      160349  53081431  11.390
## <none>                52921082  12.105
## - Location            1      327387  53248470  12.729
## - Num_serious_crimes  1      333852  53254934  12.781
## - Pct_High_Sch_grads  1     1006406  53927488  18.171
## - Per_cap_1990income  1     1216705  54137788  19.856
## - Region_num          3     2128004  55049086  23.159
## - Pct_Bachelors       1     2980964  55902046  33.995
## - Population_1990     1     4544652  57465734  46.526
## - Total_personal_income 1    14601836  67522918 127.124
## - Num_hospital_beds   1     68889489 121810571 562.181
##
## Step: AIC=10.59
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Num_hospital_beds + Num_serious_crimes + Pct_High_Sch_grads +
##   Pct_Bachelors + Per_cap_1990income + Total_personal_income +
##   Region_num
##
##           Df Sum of Sq      RSS      Cp
## - Pct_Age18_to_34      1    165268  53147006   9.9161
## <none>                  52981738  10.5916
## - Location              1    324048  53305786  11.1885
## - Num_serious_crimes    1    337129  53318867  11.2933
## - Pct_High_Sch_grads    1    985879  53967616  16.4924
## - Per_cap_1990income    1   1243294  54225032  18.5553
## - Region_num            3   2070716  55052454  21.1862
## - Pct_Bachelors         1   3091816  56073554  33.3692
## - Population_1990       1   4635327  57617065  45.7388
## - Total_personal_income  1  14714000  67695737 126.5085
## - Num_hospital_beds     1  70783644 123765382 575.8463
##
## Step: AIC=9.92
## Num_physicians ~ Location + Population_1990 + Num_hospital_beds +
##   Num_serious_crimes + Pct_High_Sch_grads + Pct_Bachelors +
##   Per_cap_1990income + Total_personal_income + Region_num
##
##           Df Sum of Sq      RSS      Cp
## <none>                  53147006   9.9161
## - Num_serious_crimes    1    328045  53475052  10.5450
## - Location              1    345999  53493006  10.6889
## - Pct_High_Sch_grads    1   1046800  54193807  16.3050
## - Region_num            3   2002845  55149852  19.9667
## - Per_cap_1990income    1   2276521  55423527  26.1599
## - Population_1990       1   4676274  57823281  45.3914
## - Pct_Bachelors         1   6443121  59590127  59.5507
## - Total_personal_income  1  14855821  68002827 126.9695
## - Num_hospital_beds     1  70793939 123940945 575.2533
##
## Call:
## lm(formula = Num_physicians ~ Location + Population_1990 + Num_hospital_beds +
##   Num_serious_crimes + Pct_High_Sch_grads + Pct_Bachelors +
##   Per_cap_1990income + Total_personal_income + Region_num,

```

```
## data = CDI)
##
## Coefficients:
## (Intercept) LocationWest Population_1990
## 842.677027 -91.128259 -0.001930
## Num_hospital_beds Num_serious_crimes Pct_High_Sch_grads
## 0.510014 -0.001161 -11.585523
## Pct_Bachelors Per_cap_1990income Total_personal_income
## 29.616127 -0.034947 0.142759
## Region_num2 Region_num3 Region_num4
## -28.870258 -38.635995 222.690899
```

```
step(Base, scale = MSE, direction = "forward") # Forward Selection
```

```
## Start: AIC=10831.23
## Num_physicians ~ 1
##
## Call:
## lm(formula = Num_physicians ~ 1, data = CDI)
##
## Coefficients:
## (Intercept)
## 988
```

```
step(Base, scope = list(upper = Full), scale = MSE) # Stepwise Regression
```

```
## Start: AIC=10831.23
## Num_physicians ~ 1
##
## Df Sum of Sq RSS Cp
## + Num_hospital_beds 1 1270342254 135864045 652.80
## + Total_personal_income 1 1264058045 142148254 703.17
## + Population_1990 1 1243181164 163025135 870.47
## + Num_serious_crimes 1 946593047 459613252 3247.31
## + Per_cap_1990income 1 140537806 1265668493 9706.97
## + Pct_Bachelors 1 78828952 1327377347 10201.50
## + Pct_Age18_to_34 1 20147995 1386058304 10671.77
## + Region_num 3 17468148 1388738151 10697.24
## + Pct_below_poverty 1 5784372 1400421927 10786.87
## + Location 1 4405498 1401800801 10797.92
## + Pct_unemployed 1 3588467 1402617832 10804.47
## <none> 1406206299 10831.23
## + Pct_High_Sch_grads 1 25377 1406180922 10833.03
## + Pct_65_or_over 1 13764 1406192535 10833.12
##
## Step: AIC=652.8
## Num_physicians ~ Num_hospital_beds
##
## Df Sum of Sq RSS Cp
## + Total_personal_income 1 62144628 73719417 156.78
## + Population_1990 1 37164568 98699477 356.97
## + Pct_Bachelors 1 28367391 107496654 427.47
## + Per_cap_1990income 1 25074801 110789244 453.86
## + Pct_High_Sch_grads 1 14851917 121012128 535.78
## + Pct_below_poverty 1 14523310 121340735 538.42
```

```

## + Region_num          3    9838681 126025364  579.96
## + Pct_unemployed      1    4676645 131187400  617.33
## + Pct_65_or_over      1    4076892 131787153  622.13
## + Pct_Age18_to_34     1    3375694 132488351  627.75
## + Location            1    1374580 134489465  643.79
## <none>                1          135864045  652.80
## + Num_serious_crimes  1     193905 135670140  653.25
## - Num_hospital_beds   1 1270342254 1406206299 10831.23
##
## Step: AIC=156.78
## Num_physicians ~ Num_hospital_beds + Total_personal_income
##
##              Df Sum of Sq      RSS      Cp
## + Population_1990  1 10822467 62896949  72.051
## + Pct_Bachelors    1  9343406 64376011  83.904
## + Num_serious_crimes  1  4645644 69073773 121.552
## + Per_cap_1990income  1  3903752 69815665 127.497
## + Pct_Age18_to_34   1  3116689 70602728 133.805
## + Pct_unemployed    1  2033967 71685450 142.482
## + Pct_High_Sch_grads  1  1627501 72091916 145.739
## + Pct_65_or_over    1   539110 73180307 154.461
## + Region_num        3   953674 72765743 155.139
## <none>              1          73719417 156.782
## + Pct_below_poverty  1    51707 73667710 158.367
## + Location          1    36574 73682843 158.489
## - Total_personal_income  1 62144628 135864045 652.804
## - Num_hospital_beds   1 68428837 142148254 703.165
##
## Step: AIC=72.05
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990
##
##              Df Sum of Sq      RSS      Cp
## + Pct_Age18_to_34    1  3659641 59237308  44.723
## + Pct_Bachelors      1  3570313 59326637  45.439
## + Region_num         3  2550764 60346186  57.610
## + Pct_65_or_over     1  1474156 61422794  62.238
## + Location           1   894039 62002911  66.887
## + Pct_below_poverty  1   699920 62197030  68.442
## + Pct_unemployed     1   494919 62402031  70.085
## <none>              1          62896949  72.051
## + Num_serious_crimes  1   230467 62666482  72.204
## + Pct_High_Sch_grads  1   227409 62669541  72.229
## + Per_cap_1990income  1   107668 62789281  73.189
## - Population_1990    1 10822467 73719417 156.782
## - Total_personal_income  1 35802527 98699477 356.970
## - Num_hospital_beds   1 78070132 140967081 695.699
##
## Step: AIC=44.72
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Pct_Age18_to_34
##
##              Df Sum of Sq      RSS      Cp
## + Region_num        3   2754623 56482685  28.648

```

```

## + Pct_Bachelors      1  1047538  58189771  38.328
## + Location           1    728337  58508971  40.886
## + Pct_below_poverty  1    650134  58587174  41.513
## + Num_serious_crimes 1    292780  58944528  44.377
## <none>                59237308  44.723
## + Per_cap_1990income 1     68262  59169046  46.176
## + Pct_unemployed     1     19434  59217874  46.568
## + Pct_High_Sch_grads 1      8234  59229074  46.657
## + Pct_65_or_over     1         96  59237212  46.722
## - Pct_Age18_to_34    1  3659641  62896949  72.051
## - Population_1990    1 11365419  70602728 133.805
## - Total_personal_income 1 36620180  95857489 336.195
## - Num_hospital_beds  1  78076866 137314175 668.425
##
## Step: AIC=28.65
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Pct_Age18_to_34 + Region_num
##
##           Df Sum of Sq      RSS      Cp
## + Pct_Bachelors      1    625629  55857056  25.634
## + Num_serious_crimes  1    289425  56193260  28.328
## + Pct_below_poverty  1    276747  56205938  28.430
## <none>                56482685  28.648
## + Location           1     86114  56396571  29.958
## + Per_cap_1990income  1     61317  56421368  30.157
## + Pct_unemployed     1     53633  56429052  30.218
## + Pct_High_Sch_grads  1     27655  56455031  30.426
## + Pct_65_or_over     1      1728  56480958  30.634
## - Region_num         3    2754623  59237308  44.723
## - Pct_Age18_to_34    1    3863501  60346186  57.610
## - Population_1990    1 12913781  69396466 130.138
## - Total_personal_income 1 36696594  93179279 320.732
## - Num_hospital_beds  1  79328605 135811290 662.381
##
## Step: AIC=25.63
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Pct_Age18_to_34 + Region_num + Pct_Bachelors
##
##           Df Sum of Sq      RSS      Cp
## + Per_cap_1990income  1   1364121  54492935  16.702
## + Pct_High_Sch_grads  1    934995  54922062  20.141
## + Pct_below_poverty  1    786481  55070575  21.331
## + Num_serious_crimes  1    398991  55458066  24.437
## <none>                55857056  25.634
## + Location           1    124939  55732117  26.633
## + Pct_unemployed     1     20060  55836996  27.473
## + Pct_65_or_over     1     18115  55838941  27.489
## - Pct_Bachelors      1    625629  56482685  28.648
## - Pct_Age18_to_34    1   1512066  57369122  35.752
## - Region_num         3    2332714  58189771  38.328
## - Population_1990    1   7486385  63343441  83.629
## - Total_personal_income 1 21073262  76930318 192.514
## - Num_hospital_beds  1  79102272 134959328 657.554
##

```

```

## Step: AIC=16.7
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Pct_Age18_to_34 + Region_num + Pct_Bachelors +
##   Per_cap_1990income
##
##           Df Sum of Sq      RSS      Cp
## + Pct_High_Sch_grads      1    836005 53656930 12.002
## + Location                  1    298738 54194197 16.308
## - Pct_Age18_to_34          1    236497 54729433 16.598
## <none>                      54492935 16.702
## + Num_serious_crimes       1    239489 54253446 16.783
## + Pct_below_poverty         1    221836 54271099 16.924
## + Pct_unemployed            1     28917 54464018 18.471
## + Pct_65_or_over            1     17747 54475189 18.560
## - Per_cap_1990income        1   1364121 55857056 25.634
## - Region_num                 3    1907057 56399993 25.985
## - Pct_Bachelors             1    1928433 56421368 30.157
## - Population_1990           1    8748039 63240975 84.808
## - Total_personal_income     1   20212562 74705497 176.684
## - Num_hospital_beds         1   80291518 134784454 658.152
##
## Step: AIC=12
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Pct_Age18_to_34 + Region_num + Pct_Bachelors +
##   Per_cap_1990income + Pct_High_Sch_grads
##
##           Df Sum of Sq      RSS      Cp
## + Num_serious_crimes       1    351144 53305786 11.188
## + Location                  1    338064 53318867 11.293
## - Pct_Age18_to_34          1    177801 53834731 11.427
## <none>                      53656930 12.002
## + Pct_unemployed            1     60498 53596433 13.518
## + Pct_below_poverty         1     22413 53634517 13.823
## + Pct_65_or_over            1     18188 53638742 13.857
## - Pct_High_Sch_grads        1    836005 54492935 16.702
## - Per_cap_1990income        1   1265131 54922062 20.141
## - Region_num                 3   1998565 55655495 22.019
## - Pct_Bachelors             1   2762340 56419270 32.140
## - Population_1990           1   8046345 61703275 74.485
## - Total_personal_income     1   19562745 73219675 166.777
## - Num_hospital_beds         1   70399618 124056548 574.180
##
## Step: AIC=11.19
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Pct_Age18_to_34 + Region_num + Pct_Bachelors +
##   Per_cap_1990income + Pct_High_Sch_grads + Num_serious_crimes
##
##           Df Sum of Sq      RSS      Cp
## + Location                  1    324048 52981738 10.592
## - Pct_Age18_to_34          1    187220 53493006 10.689
## <none>                      53305786 11.188
## - Num_serious_crimes       1    351144 53656930 12.002
## + Pct_unemployed            1     57317 53248470 12.729
## + Pct_65_or_over            1     14007 53291779 13.076

```

```

## + Pct_below_poverty      1      7921  53297865  13.125
## - Pct_High_Sch_grads     1     947660  54253446  16.783
## - Per_cap_1990income     1    1076897  54382683  17.819
## - Region_num             3    1937940  55243726  20.719
## - Pct_Bachelors          1    2852866  56158652  32.051
## - Population_1990        1    4605002  57910788  46.093
## - Total_personal_income  1   14674685  67980471 126.790
## - Num_hospital_beds      1   70631430 123937216 575.223
##
## Step: AIC=10.59
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Pct_Age18_to_34 + Region_num + Pct_Bachelors +
##   Per_cap_1990income + Pct_High_Sch_grads + Num_serious_crimes +
##   Location
##
##           Df Sum of Sq      RSS      Cp
## - Pct_Age18_to_34      1    165268  53147006  9.9161
## <none>                                52981738 10.5916
## - Location              1    324048  53305786 11.1885
## - Num_serious_crimes    1    337129  53318867 11.2933
## + Pct_unemployed        1     60656  52921082 12.1055
## + Pct_below_poverty     1     1192  52980546 12.5821
## + Pct_65_or_over         1      273  52981465 12.5894
## - Pct_High_Sch_grads    1    985879  53967616 16.4924
## - Per_cap_1990income     1   1243294  54225032 18.5553
## - Region_num            3   2070716  55052454 21.1862
## - Pct_Bachelors          1   3091816  56073554 33.3692
## - Population_1990        1   4635327  57617065 45.7388
## - Total_personal_income  1  14714000  67695737 126.5085
## - Num_hospital_beds      1  70783644 123765382 575.8463
##
## Step: AIC=9.92
## Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Region_num + Pct_Bachelors + Per_cap_1990income +
##   Pct_High_Sch_grads + Num_serious_crimes + Location
##
##           Df Sum of Sq      RSS      Cp
## <none>                                53147006  9.9161
## - Num_serious_crimes    1    328045  53475052 10.5450
## + Pct_Age18_to_34        1    165268  52981738 10.5916
## - Location              1    345999  53493006 10.6889
## + Pct_unemployed         1     65575  53081431 11.3905
## + Pct_65_or_over         1     34493  53112513 11.6396
## + Pct_below_poverty     1      44    53146963 11.9157
## - Pct_High_Sch_grads    1   1046800  54193807 16.3050
## - Region_num            3   2002845  55149852 19.9667
## - Per_cap_1990income     1   2276521  55423527 26.1599
## - Population_1990        1   4676274  57823281 45.3914
## - Pct_Bachelors          1   6443121  59590127 59.5507
## - Total_personal_income  1  14855821  68002827 126.9695
## - Num_hospital_beds      1  70793939 123940945 575.2533
##
## Call:

```



```
## lm(formula = Num_physicians ~ Num_hospital_beds + Total_personal_income +
##   Population_1990 + Region_num + Pct_Bachelors + Per_cap_1990income +
##   Pct_High_Sch_grads + Num_serious_crimes + Location, data = CDI)
##
## Coefficients:
##   (Intercept)      Num_hospital_beds  Total_personal_income
##   842.677027          0.510014          0.142759
##   Population_1990      Region_num2      Region_num3
##   -0.001930          -28.870258          -38.635995
##   Region_num4      Pct_Bachelors      Per_cap_1990income
##   222.690899          29.616127          -0.034947
##   Pct_High_Sch_grads  Num_serious_crimes      LocationWest
##   -11.585523          -0.001161          -91.128259
```

You can also choose to keep certain variables in your model by creating a new *lower* model like *Pop* as illustrated below.

```
Pop <- lm(Num_physicians ~ Population_1990, data = CDI)
# Backward Elimination
step(Full, scope = list(upper = Full, lower = Pop), scale = MSE, direction = "backward")
```

```
## Start: AIC=16
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Pct_65_or_over + Num_hospital_beds + Num_serious_crimes +
##   Pct_High_Sch_grads + Pct_Bachelors + Pct_below_poverty +
##   Pct_unemployed + Per_cap_1990income + Total_personal_income +
##   Region_num
##
##           Df Sum of Sq      RSS      Cp
## - Pct_65_or_over      1      1214 52909130 14.010
## - Pct_below_poverty    1     12540 52920455 14.101
## - Pct_unemployed       1     72264 52980179 14.579
## - Pct_Age18_to_34     1    137143 53045058 15.099
## <none>                                52907915 16.000
## - Location             1    324340 53232255 16.599
## - Num_serious_crimes   1    342597 53250512 16.745
## - Pct_High_Sch_grads   1    599909 53507824 18.808
## - Per_cap_1990income   1    848855 53756770 20.803
## - Region_num           3    2109381 55017296 26.904
## - Pct_Bachelors        1    2349330 55257245 32.827
## - Total_personal_income 1  14355133 67263048 129.041
## - Num_hospital_beds    1   51826830 104734746 429.336
##
## Step: AIC=14.01
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Num_hospital_beds + Num_serious_crimes + Pct_High_Sch_grads +
##   Pct_Bachelors + Pct_below_poverty + Pct_unemployed + Per_cap_1990income +
##   Total_personal_income + Region_num
##
##           Df Sum of Sq      RSS      Cp
## - Pct_below_poverty    1     11952 52921082 12.105
## - Pct_unemployed       1     71416 52980546 12.582
## - Pct_Age18_to_34     1    165349 53074479 13.335
## <none>                                52909130 14.010
## - Location             1    338770 53247900 14.725
```

```

## - Num_serious_crimes      1    343334  53252464  14.761
## - Pct_High_Sch_grads      1    603286  53512415  16.844
## - Per_cap_1990income      1    856850  53765980  18.877
## - Region_num              3    2139061  55048190  25.152
## - Pct_Bachelors           1    2349693  55258823  30.840
## - Total_personal_income   1   14435703  67344833  127.696
## - Num_hospital_beds       1   56006526 108915656  460.842
##
## Step:  AIC=12.11
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Num_hospital_beds + Num_serious_crimes + Pct_High_Sch_grads +
##   Pct_Bachelors + Pct_unemployed + Per_cap_1990income + Total_personal_income +
##   Region_num
##
##              Df Sum of Sq      RSS      Cp
## - Pct_unemployed      1      60656  52981738  10.592
## - Pct_Age18_to_34     1     160349  53081431  11.390
## <none>                                52921082  12.105
## - Location            1     327387  53248470  12.729
## - Num_serious_crimes  1     333852  53254934  12.781
## - Pct_High_Sch_grads  1    1006406  53927488  18.171
## - Per_cap_1990income  1    1216705  54137788  19.856
## - Region_num          3    2128004  55049086  23.159
## - Pct_Bachelors       1    2980964  55902046  33.995
## - Total_personal_income 1   14601836  67522918  127.124
## - Num_hospital_beds   1   68889489 121810571  562.181
##
## Step:  AIC=10.59
## Num_physicians ~ Location + Population_1990 + Pct_Age18_to_34 +
##   Num_hospital_beds + Num_serious_crimes + Pct_High_Sch_grads +
##   Pct_Bachelors + Per_cap_1990income + Total_personal_income +
##   Region_num
##
##              Df Sum of Sq      RSS      Cp
## - Pct_Age18_to_34     1     165268  53147006   9.9161
## <none>                                52981738  10.5916
## - Location            1     324048  53305786  11.1885
## - Num_serious_crimes  1     337129  53318867  11.2933
## - Pct_High_Sch_grads  1     985879  53967616  16.4924
## - Per_cap_1990income  1    1243294  54225032  18.5553
## - Region_num          3    2070716  55052454  21.1862
## - Pct_Bachelors       1    3091816  56073554  33.3692
## - Total_personal_income 1   14714000  67695737  126.5085
## - Num_hospital_beds   1   70783644 123765382  575.8463
##
## Step:  AIC=9.92
## Num_physicians ~ Location + Population_1990 + Num_hospital_beds +
##   Num_serious_crimes + Pct_High_Sch_grads + Pct_Bachelors +
##   Per_cap_1990income + Total_personal_income + Region_num
##
##              Df Sum of Sq      RSS      Cp
## <none>                                53147006   9.9161
## - Num_serious_crimes  1     328045  53475052  10.5450
## - Location            1     345999  53493006  10.6889

```

```

## - Pct_High_Sch_grads      1  1046800  54193807  16.3050
## - Region_num              3   2002845  55149852  19.9667
## - Per_cap_1990income      1   2276521  55423527  26.1599
## - Pct_Bachelors           1   6443121  59590127  59.5507
## - Total_personal_income   1  14855821  68002827 126.9695
## - Num_hospital_beds       1  70793939 123940945 575.2533

##
## Call:
## lm(formula = Num_physicians ~ Location + Population_1990 + Num_hospital_beds +
##     Num_serious_crimes + Pct_High_Sch_grads + Pct_Bachelors +
##     Per_cap_1990income + Total_personal_income + Region_num,
##     data = CDI)
##
## Coefficients:
##             (Intercept)              LocationWest          Population_1990
##             842.677027                -91.128259                -0.001930
##     Num_hospital_beds      Num_serious_crimes      Pct_High_Sch_grads
##             0.510014                -0.001161                -11.585523
##             Pct_Bachelors      Per_cap_1990income      Total_personal_income
##             29.616127                -0.034947                 0.142759
##             Region_num2          Region_num3          Region_num4
##             -28.870258             -38.635995             222.690899

# Forward Selection
step(Pop, scope = list(upper = Full, lower = Pop), scale = MSE, direction = "forward")

## Start:  AIC=870.47
## Num_physicians ~ Population_1990
##
##              Df Sum of Sq      RSS      Cp
## + Num_hospital_beds      1  64325658  98699477  356.97
## + Total_personal_income  1  22058054 140967081  695.70
## + Pct_Bachelors          1  14007395 149017740  760.22
## + Per_cap_1990income     1  13324708 149700428  765.69
## + Pct_unemployed         1   4339094 158686041  837.70
## + Pct_Age18_to_34        1   2995219 160029916  848.47
## + Region_num             3   3042595 159982540  852.09
## + Location                1   2025189 160999946  856.24
## + Pct_below_poverty       1   1134909 161890226  863.38
## + Num_serious_crimes      1   1093534 161931601  863.71
## + Pct_65_or_over          1    822438 162202697  865.88
## <none>                                163025135  870.47
## + Pct_High_Sch_grads     1    207225 162817910  870.81
##
## Step:  AIC=356.97
## Num_physicians ~ Population_1990 + Num_hospital_beds
##
##              Df Sum of Sq      RSS      Cp
## + Total_personal_income  1  35802527 62896949   72.051
## + Pct_Bachelors          1  20313809 78385668  196.177
## + Per_cap_1990income     1  17223333 81476144  220.944
## + Num_serious_crimes      1   8039322 90660154  294.544
## + Pct_High_Sch_grads     1   6465784 92233693  307.154
## + Pct_unemployed         1   4567330 94132147  322.368

```

```

## + Pct_below_poverty      1  3802915 94896562 328.494
## + Pct_Age18_to_34        1  2841988 95857489 336.195
## + Region_num              3  2129500 96569977 345.904
## + Pct_65_or_over         1   621822 98077655 353.987
## <none>                    1  98699477 356.970
## + Location                1   1086 98698391 358.961
##
## Step: AIC=72.05
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income
##
##           Df Sum of Sq      RSS      Cp
## + Pct_Age18_to_34  1  3659641 59237308 44.723
## + Pct_Bachelors    1  3570313 59326637 45.439
## + Region_num       3  2550764 60346186 57.610
## + Pct_65_or_over   1  1474156 61422794 62.238
## + Location         1   894039 62002911 66.887
## + Pct_below_poverty 1   699920 62197030 68.442
## + Pct_unemployed   1   494919 62402031 70.085
## <none>              1  62896949 72.051
## + Num_serious_crimes 1   230467 62666482 72.204
## + Pct_High_Sch_grads 1   227409 62669541 72.229
## + Per_cap_1990income 1   107668 62789281 73.189
##
## Step: AIC=44.72
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34
##
##           Df Sum of Sq      RSS      Cp
## + Region_num       3  2754623 56482685 28.648
## + Pct_Bachelors    1  1047538 58189771 38.328
## + Location         1   728337 58508971 40.886
## + Pct_below_poverty 1   650134 58587174 41.513
## + Num_serious_crimes 1   292780 58944528 44.377
## <none>              1  59237308 44.723
## + Per_cap_1990income 1    68262 59169046 46.176
## + Pct_unemployed   1   19434 59217874 46.568
## + Pct_High_Sch_grads 1    8234 59229074 46.657
## + Pct_65_or_over   1     96 59237212 46.722
##
## Step: AIC=28.65
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num
##
##           Df Sum of Sq      RSS      Cp
## + Pct_Bachelors    1   625629 55857056 25.634
## + Num_serious_crimes 1   289425 56193260 28.328
## + Pct_below_poverty 1   276747 56205938 28.430
## <none>              1  56482685 28.648
## + Location         1    86114 56396571 29.958
## + Per_cap_1990income 1    61317 56421368 30.157
## + Pct_unemployed   1   53633 56429052 30.218
## + Pct_High_Sch_grads 1   27655 56455031 30.426
## + Pct_65_or_over   1    1728 56480958 30.634
##

```

```

## Step: AIC=25.63
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num + Pct_Bachelors
##
##           Df Sum of Sq      RSS      Cp
## + Per_cap_1990income  1  1364121 54492935 16.702
## + Pct_High_Sch_grads  1   934995 54922062 20.141
## + Pct_below_poverty   1   786481 55070575 21.331
## + Num_serious_crimes  1   398991 55458066 24.437
## <none>                                55857056 25.634
## + Location             1   124939 55732117 26.633
## + Pct_unemployed       1    20060 55836996 27.473
## + Pct_65_or_over       1    18115 55838941 27.489
##
## Step: AIC=16.7
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income
##
##           Df Sum of Sq      RSS      Cp
## + Pct_High_Sch_grads  1   836005 53656930 12.002
## + Location             1   298738 54194197 16.308
## <none>                                54492935 16.702
## + Num_serious_crimes  1   239489 54253446 16.783
## + Pct_below_poverty   1   221836 54271099 16.924
## + Pct_unemployed       1    28917 54464018 18.471
## + Pct_65_or_over       1    17747 54475189 18.560
##
## Step: AIC=12
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income +
##   Pct_High_Sch_grads
##
##           Df Sum of Sq      RSS      Cp
## + Num_serious_crimes  1   351144 53305786 11.188
## + Location             1   338064 53318867 11.293
## <none>                                53656930 12.002
## + Pct_unemployed       1    60498 53596433 13.518
## + Pct_below_poverty   1    22413 53634517 13.823
## + Pct_65_or_over       1    18188 53638742 13.857
##
## Step: AIC=11.19
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income +
##   Pct_High_Sch_grads + Num_serious_crimes
##
##           Df Sum of Sq      RSS      Cp
## + Location             1   324048 52981738 10.592
## <none>                                53305786 11.188
## + Pct_unemployed       1    57317 53248470 12.729
## + Pct_65_or_over       1    14007 53291779 13.076
## + Pct_below_poverty   1     7921 53297865 13.125
##
## Step: AIC=10.59
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +

```

```
## Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income +
## Pct_High_Sch_grads + Num_serious_crimes + Location
##
##           Df Sum of Sq      RSS      Cp
## <none>                52981738 10.592
## + Pct_unemployed      1    60656 52921082 12.105
## + Pct_below_poverty   1     1192 52980546 12.582
## + Pct_65_or_over      1      273 52981465 12.589
##
## Call:
## lm(formula = Num_physicians ~ Population_1990 + Num_hospital_beds +
##     Total_personal_income + Pct_Age18_to_34 + Region_num + Pct_Bachelors +
##     Per_cap_1990income + Pct_High_Sch_grads + Num_serious_crimes +
##     Location, data = CDI)
##
## Coefficients:
##           (Intercept)      Population_1990      Num_hospital_beds
##           611.348768                -0.001922                0.509977
## Total_personal_income      Pct_Age18_to_34      Region_num2
##           0.142181                6.430917                -29.365854
##           Region_num3      Region_num4      Pct_Bachelors
##           -33.663588                231.267099                25.944371
## Per_cap_1990income      Pct_High_Sch_grads      Num_serious_crimes
##           -0.029643                -11.269762                -0.001177
##           LocationWest
##           -88.280284
```

```
# Stepwise Regression
step(Pop, scope = list(upper = Full, lower = Pop), scale = MSE)
```

```
## Start: AIC=870.47
## Num_physicians ~ Population_1990
##
##           Df Sum of Sq      RSS      Cp
## + Num_hospital_beds      1  64325658  98699477 356.97
## + Total_personal_income  1  22058054 140967081 695.70
## + Pct_Bachelors          1  14007395 149017740 760.22
## + Per_cap_1990income     1  13324708 149700428 765.69
## + Pct_unemployed         1   4339094 158686041 837.70
## + Pct_Age18_to_34        1   2995219 160029916 848.47
## + Region_num             3   3042595 159982540 852.09
## + Location               1   2025189 160999946 856.24
## + Pct_below_poverty      1   1134909 161890226 863.38
## + Num_serious_crimes     1   1093534 161931601 863.71
## + Pct_65_or_over         1    822438 162202697 865.88
## <none>                    163025135 870.47
## + Pct_High_Sch_grads     1    207225 162817910 870.81
##
## Step: AIC=356.97
## Num_physicians ~ Population_1990 + Num_hospital_beds
##
##           Df Sum of Sq      RSS      Cp
## + Total_personal_income  1  35802527  62896949  72.051
## + Pct_Bachelors         1  20313809  78385668 196.177
```

```

## + Per_cap_1990income      1 17223333 81476144 220.944
## + Num_serious_crimes      1  8039322 90660154 294.544
## + Pct_High_Sch_grads      1  6465784 92233693 307.154
## + Pct_unemployed          1  4567330 94132147 322.368
## + Pct_below_poverty       1  3802915 94896562 328.494
## + Pct_Age18_to_34         1  2841988 95857489 336.195
## + Region_num               3  2129500 96569977 345.904
## + Pct_65_or_over          1   621822 98077655 353.987
## <none>                     98699477 356.970
## + Location                 1    1086 98698391 358.961
## - Num_hospital_beds       1 64325658 163025135 870.471
##
## Step: AIC=72.05
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income
##
##              Df Sum of Sq      RSS      Cp
## + Pct_Age18_to_34      1  3659641 59237308 44.723
## + Pct_Bachelors        1  3570313 59326637 45.439
## + Region_num           3  2550764 60346186 57.610
## + Pct_65_or_over       1  1474156 61422794 62.238
## + Location             1   894039 62002911 66.887
## + Pct_below_poverty    1   699920 62197030 68.442
## + Pct_unemployed       1   494919 62402031 70.085
## <none>                  62896949 72.051
## + Num_serious_crimes   1   230467 62666482 72.204
## + Pct_High_Sch_grads   1   227409 62669541 72.229
## + Per_cap_1990income    1   107668 62789281 73.189
## - Total_personal_income 1 35802527 98699477 356.970
## - Num_hospital_beds    1 78070132 140967081 695.699
##
## Step: AIC=44.72
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34
##
##              Df Sum of Sq      RSS      Cp
## + Region_num           3  2754623 56482685 28.648
## + Pct_Bachelors        1  1047538 58189771 38.328
## + Location             1   728337 58508971 40.886
## + Pct_below_poverty    1   650134 58587174 41.513
## + Num_serious_crimes   1   292780 58944528 44.377
## <none>                  59237308 44.723
## + Per_cap_1990income    1    68262 59169046 46.176
## + Pct_unemployed       1   19434 59217874 46.568
## + Pct_High_Sch_grads   1    8234 59229074 46.657
## + Pct_65_or_over       1     96 59237212 46.722
## - Pct_Age18_to_34      1  3659641 62896949 72.051
## - Total_personal_income 1 36620180 95857489 336.195
## - Num_hospital_beds    1 78076866 137314175 668.425
##
## Step: AIC=28.65
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num
##
##              Df Sum of Sq      RSS      Cp

```

```

## + Pct_Bachelors          1    625629  55857056  25.634
## + Num_serious_crimes     1    289425  56193260  28.328
## + Pct_below_poverty      1    276747  56205938  28.430
## <none>                   1           56482685  28.648
## + Location                1     86114  56396571  29.958
## + Per_cap_1990income     1     61317  56421368  30.157
## + Pct_unemployed         1     53633  56429052  30.218
## + Pct_High_Sch_grads    1     27655  56455031  30.426
## + Pct_65_or_over        1      1728  56480958  30.634
## - Region_num             3    2754623  59237308  44.723
## - Pct_Age18_to_34       1    3863501  60346186  57.610
## - Total_personal_income  1   36696594  93179279 320.732
## - Num_hospital_beds     1   79328605 135811290 662.381
##
## Step: AIC=25.63
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num + Pct_Bachelors
##
##              Df Sum of Sq      RSS      Cp
## + Per_cap_1990income  1  1364121  54492935  16.702
## + Pct_High_Sch_grads  1   934995  54922062  20.141
## + Pct_below_poverty   1   786481  55070575  21.331
## + Num_serious_crimes  1   398991  55458066  24.437
## <none>                1           55857056  25.634
## + Location            1   124939  55732117  26.633
## + Pct_unemployed      1    20060  55836996  27.473
## + Pct_65_or_over     1   18115  55838941  27.489
## - Pct_Bachelors      1    625629  56482685  28.648
## - Pct_Age18_to_34   1   1512066  57369122  35.752
## - Region_num         3   2332714  58189771  38.328
## - Total_personal_income  1  21073262  76930318 192.514
## - Num_hospital_beds  1   79102272 134959328 657.554
##
## Step: AIC=16.7
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income
##
##              Df Sum of Sq      RSS      Cp
## + Pct_High_Sch_grads  1   836005  53656930  12.002
## + Location            1   298738  54194197  16.308
## - Pct_Age18_to_34    1   236497  54729433  16.598
## <none>                1           54492935  16.702
## + Num_serious_crimes  1   239489  54253446  16.783
## + Pct_below_poverty   1   221836  54271099  16.924
## + Pct_unemployed      1    28917  54464018  18.471
## + Pct_65_or_over     1    17747  54475189  18.560
## - Per_cap_1990income  1  1364121  55857056  25.634
## - Region_num         3   1907057  56399993  25.985
## - Pct_Bachelors      1   1928433  56421368  30.157
## - Total_personal_income  1  20212562  74705497 176.684
## - Num_hospital_beds  1   80291518 134784454 658.152
##
## Step: AIC=12
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +

```



```

##      Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income +
##      Pct_High_Sch_grads
##
##              Df Sum of Sq      RSS      Cp
## + Num_serious_crimes      1   351144 53305786 11.188
## + Location                  1   338064 53318867 11.293
## - Pct_Age18_to_34          1   177801 53834731 11.427
## <none>                      53656930 12.002
## + Pct_unemployed           1    60498 53596433 13.518
## + Pct_below_poverty         1    22413 53634517 13.823
## + Pct_65_or_over            1    18188 53638742 13.857
## - Pct_High_Sch_grads        1   836005 54492935 16.702
## - Per_cap_1990income         1  1265131 54922062 20.141
## - Region_num                 3  1998565 55655495 22.019
## - Pct_Bachelors              1  2762340 56419270 32.140
## - Total_personal_income       1 19562745 73219675 166.777
## - Num_hospital_beds           1 70399618 124056548 574.180
##
## Step:  AIC=11.19
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##      Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income +
##      Pct_High_Sch_grads + Num_serious_crimes
##
##              Df Sum of Sq      RSS      Cp
## + Location                  1   324048 52981738 10.592
## - Pct_Age18_to_34          1   187220 53493006 10.689
## <none>                      53305786 11.188
## - Num_serious_crimes        1   351144 53656930 12.002
## + Pct_unemployed            1    57317 53248470 12.729
## + Pct_65_or_over            1    14007 53291779 13.076
## + Pct_below_poverty         1     7921 53297865 13.125
## - Pct_High_Sch_grads        1   947660 54253446 16.783
## - Per_cap_1990income         1  1076897 54382683 17.819
## - Region_num                 3  1937940 55243726 20.719
## - Pct_Bachelors              1  2852866 56158652 32.051
## - Total_personal_income       1 14674685 67980471 126.790
## - Num_hospital_beds           1 70631430 123937216 575.223
##
## Step:  AIC=10.59
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##      Pct_Age18_to_34 + Region_num + Pct_Bachelors + Per_cap_1990income +
##      Pct_High_Sch_grads + Num_serious_crimes + Location
##
##              Df Sum of Sq      RSS      Cp
## - Pct_Age18_to_34          1   165268 53147006  9.9161
## <none>                      52981738 10.5916
## - Location                  1   324048 53305786 11.1885
## - Num_serious_crimes        1   337129 53318867 11.2933
## + Pct_unemployed            1    60656 52921082 12.1055
## + Pct_below_poverty         1     1192 52980546 12.5821
## + Pct_65_or_over            1     273 52981465 12.5894
## - Pct_High_Sch_grads        1   985879 53967616 16.4924
## - Per_cap_1990income         1  1243294 54225032 18.5553
## - Region_num                 3  2070716 55052454 21.1862

```

```

## - Pct_Bachelors          1  3091816  56073554  33.3692
## - Total_personal_income  1 14714000  67695737 126.5085
## - Num_hospital_beds      1  70783644 123765382 575.8463
##
## Step: AIC=9.92
## Num_physicians ~ Population_1990 + Num_hospital_beds + Total_personal_income +
##   Region_num + Pct_Bachelors + Per_cap_1990income + Pct_High_Sch_grads +
##   Num_serious_crimes + Location
##
##              Df Sum of Sq      RSS      Cp
## <none>
## - Num_serious_crimes      1   328045  53475052 10.5450
## + Pct_Age18_to_34          1   165268  52981738 10.5916
## - Location                  1    345999  53493006 10.6889
## + Pct_unemployed           1     65575  53081431 11.3905
## + Pct_65_or_over           1     34493  53112513 11.6396
## + Pct_below_poverty        1         44  53146963 11.9157
## - Pct_High_Sch_grads       1  1046800  54193807 16.3050
## - Region_num                3  2002845  55149852 19.9667
## - Per_cap_1990income        1  2276521  55423527 26.1599
## - Pct_Bachelors             1   6443121  59590127 59.5507
## - Total_personal_income     1 14855821  68002827 126.9695
## - Num_hospital_beds         1  70793939 123940945 575.2533
##
## Call:
## lm(formula = Num_physicians ~ Population_1990 + Num_hospital_beds +
##   Total_personal_income + Region_num + Pct_Bachelors + Per_cap_1990income +
##   Pct_High_Sch_grads + Num_serious_crimes + Location, data = CDI)
##
## Coefficients:
##   (Intercept)      Population_1990      Num_hospital_beds
##      842.677027          -0.001930           0.510014
## Total_personal_income      Region_num2      Region_num3
##      0.142759          -28.870258          -38.635995
##      Region_num4      Pct_Bachelors      Per_cap_1990income
##      222.690899          29.616127          -0.034947
##      Pct_High_Sch_grads      Num_serious_crimes      LocationWest
##      -11.585523          -0.001161          -91.128259

```