Model-selection-backwork

Exploratory Analysis

```
# read data
cdi_df = read_csv("./cdi.csv") %>%
  select(-id) %>%
  select(-cty) %>%
  select(-state) %>%
  relocate(crimes)

# fit regression using all predictors
mult.fit = lm(crimes ~ ., data = cdi_df)

step(mult.fit, direction = 'backward')
```

```
## Start:  AIC=8865.77
## crimes ~ area + pop + pop18 + pop65 + docs + beds + hsgrad +
##     bagrad + poverty + unemp + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - pop65     1 3.8881e+06 2.3261e+11 8863.8
## - unemp     1 3.2146e+07 2.3264e+11 8863.8
## - bagrad    1 7.7720e+07 2.3268e+11 8863.9
## - pop18     1 1.4795e+08 2.3275e+11 8864.0
## - hsgrad    1 3.7187e+08 2.3298e+11 8864.5
## - region    1 6.6550e+08 2.3327e+11 8865.0
## <none>                   2.3260e+11 8865.8
## - beds      1 1.2164e+09 2.3382e+11 8866.1
## - docs      1 1.4817e+09 2.3409e+11 8866.6
## - poverty   1 2.9923e+09 2.3560e+11 8869.4
## - pcincome  1 6.0893e+09 2.3869e+11 8875.1
## - area      1 7.0732e+09 2.3968e+11 8876.9
## - totalinc  1 3.0507e+10 2.6311e+11 8918.0
## - pop       1 7.2961e+10 3.0557e+11 8983.8
##
## Step:  AIC=8863.77
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + bagrad +
##     poverty + unemp + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - unemp     1 3.0922e+07 2.3264e+11 8861.8
## - bagrad    1 7.9111e+07 2.3269e+11 8861.9
## - pop18     1 1.6527e+08 2.3277e+11 8862.1
## - hsgrad    1 3.8763e+08 2.3300e+11 8862.5
## - region    1 6.6170e+08 2.3327e+11 8863.0
```

```
## <none>                  2.3261e+11 8863.8
## - beds      1 1.2871e+09 2.3390e+11 8864.2
## - docs      1 1.4796e+09 2.3409e+11 8864.6
## - poverty   1 3.0262e+09 2.3563e+11 8867.5
## - pcincome  1 6.0969e+09 2.3871e+11 8873.2
## - area      1 7.0971e+09 2.3971e+11 8875.0
## - totalinc  1 3.0600e+10 2.6321e+11 8916.2
## - pop       1 7.4342e+10 3.0695e+11 8983.8
##
## Step:  AIC=8861.83
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + bagrad +
##     poverty + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - bagrad    1 1.1046e+08 2.3275e+11 8860.0
## - pop18     1 1.6702e+08 2.3281e+11 8860.1
## - hsgrad    1 3.6423e+08 2.3300e+11 8860.5
## - region    1 7.4425e+08 2.3338e+11 8861.2
## <none>                  2.3264e+11 8861.8
## - beds      1 1.4088e+09 2.3405e+11 8862.5
## - docs      1 1.4861e+09 2.3413e+11 8862.6
## - poverty   1 3.0992e+09 2.3574e+11 8865.7
## - pcincome  1 6.1584e+09 2.3880e+11 8871.3
## - area      1 7.3697e+09 2.4001e+11 8873.6
## - totalinc  1 3.0573e+10 2.6321e+11 8914.2
## - pop       1 7.4823e+10 3.0746e+11 8982.5
##
## Step:  AIC=8860.04
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + poverty +
##     pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - hsgrad    1 2.5445e+08 2.3300e+11 8858.5
## - pop18     1 5.5080e+08 2.3330e+11 8859.1
## - region    1 8.4589e+08 2.3360e+11 8859.6
## <none>                  2.3275e+11 8860.0
## - beds      1 1.3067e+09 2.3406e+11 8860.5
## - docs      1 1.3831e+09 2.3413e+11 8860.6
## - poverty   1 4.1542e+09 2.3690e+11 8865.8
## - area      1 7.5698e+09 2.4032e+11 8872.1
## - pcincome  1 1.3566e+10 2.4632e+11 8883.0
## - totalinc  1 3.1712e+10 2.6446e+11 8914.2
## - pop       1 7.6620e+10 3.0937e+11 8983.3
##
## Step:  AIC=8858.52
## crimes ~ area + pop + pop18 + docs + beds + poverty + pcincome +
##     totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - pop18     1 3.5255e+08 2.3336e+11 8857.2
## - region    1 6.6233e+08 2.3367e+11 8857.8
## <none>                  2.3300e+11 8858.5
## - beds      1 1.3147e+09 2.3432e+11 8859.0
## - docs      1 1.4252e+09 2.3443e+11 8859.2
```

```
## - area      1 7.5866e+09 2.4059e+11 8870.6
## - poverty   1 7.9501e+09 2.4095e+11 8871.3
## - pcincome  1 1.3431e+10 2.4644e+11 8881.2
## - totalinc  1 3.1475e+10 2.6448e+11 8912.3
## - pop       1 7.6388e+10 3.0939e+11 8981.3
##
## Step:  AIC=8857.19
## crimes ~ area + pop + docs + beds + poverty + pcincome + totalinc +
##     region
##
##            Df  Sum of Sq        RSS    AIC
## - region    1 6.9649e+08 2.3405e+11 8856.5
## <none>                   2.3336e+11 8857.2
## - beds      1 1.1192e+09 2.3448e+11 8857.3
## - docs      1 1.1714e+09 2.3453e+11 8857.4
## - poverty   1 7.8900e+09 2.4125e+11 8869.8
## - area      1 8.0472e+09 2.4140e+11 8870.1
## - pcincome  1 1.3308e+10 2.4666e+11 8879.6
## - totalinc  1 3.2977e+10 2.6633e+11 8913.3
## - pop       1 7.8411e+10 3.1177e+11 8982.6
##
## Step:  AIC=8856.5
## crimes ~ area + pop + docs + beds + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## - beds      1 7.8044e+08 2.3483e+11 8856.0
## - docs      1 9.7325e+08 2.3503e+11 8856.3
## <none>                   2.3405e+11 8856.5
## - area      1 7.3727e+09 2.4143e+11 8868.1
## - poverty   1 9.4412e+09 2.4349e+11 8871.9
## - pcincome  1 1.3301e+10 2.4735e+11 8878.8
## - totalinc  1 3.3807e+10 2.6786e+11 8913.9
## - pop       1 8.0981e+10 3.1503e+11 8985.2
##
## Step:  AIC=8855.96
## crimes ~ area + pop + docs + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## - docs      1 2.7324e+08 2.3511e+11 8854.5
## <none>                   2.3483e+11 8856.0
## - area      1 9.3816e+09 2.4422e+11 8871.2
## - poverty   1 1.1656e+10 2.4649e+11 8875.3
## - pcincome  1 1.5123e+10 2.4996e+11 8881.4
## - totalinc  1 4.9534e+10 2.8437e+11 8938.2
## - pop       1 1.2635e+11 3.6119e+11 9043.4
##
## Step:  AIC=8854.47
## crimes ~ area + pop + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## <none>                   2.3511e+11 8854.5
## - area      1 9.1118e+09 2.4422e+11 8869.2
## - poverty   1 1.2072e+10 2.4718e+11 8874.5
## - pcincome  1 1.4929e+10 2.5004e+11 8879.6
```

```
## - totalinc  1 5.4571e+10 2.8968e+11 8944.3
## - pop       1 1.2651e+11 3.6161e+11 9041.9

##
## Call:
## lm(formula = crimes ~ area + pop + poverty + pcincome + totalinc,
##     data = cdi_df)
##
## Coefficients:
## (Intercept)         area          pop      poverty     pcincome     totalinc
##  -6.389e+04   -3.109e+00    2.496e-01    1.450e+03    2.460e+00   -7.899e+00
```

```
step(mult.fit, direction = 'forward')
```

```
## Start:  AIC=8865.77
## crimes ~ area + pop + pop18 + pop65 + docs + beds + hsgrad +
##     bagrad + poverty + unemp + pcincome + totalinc + region

##
## Call:
## lm(formula = crimes ~ area + pop + pop18 + pop65 + docs + beds +
##     hsgrad + bagrad + poverty + unemp + pcincome + totalinc +
##     region, data = cdi_df)
##
## Coefficients:
## (Intercept)         area          pop        pop18        pop65         docs
##  -5.093e+04   -3.054e+00    2.343e-01    2.211e+02    3.212e+01   -5.189e+00
##        beds       hsgrad       bagrad      poverty        unemp     pcincome
##   3.404e+00   -2.656e+02    1.409e+02    1.143e+03   -1.597e+02    2.335e+00
##    totalinc       region
##  -7.070e+00    1.457e+03
```

```
step(mult.fit, direction = 'both')
```

```
## Start:  AIC=8865.77
## crimes ~ area + pop + pop18 + pop65 + docs + beds + hsgrad +
##     bagrad + poverty + unemp + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - pop65     1 3.8881e+06 2.3261e+11 8863.8
## - unemp     1 3.2146e+07 2.3264e+11 8863.8
## - bagrad    1 7.7720e+07 2.3268e+11 8863.9
## - pop18     1 1.4795e+08 2.3275e+11 8864.0
## - hsgrad    1 3.7187e+08 2.3298e+11 8864.5
## - region    1 6.6550e+08 2.3327e+11 8865.0
## <none>                   2.3260e+11 8865.8
## - beds      1 1.2164e+09 2.3382e+11 8866.1
## - docs      1 1.4817e+09 2.3409e+11 8866.6
## - poverty   1 2.9923e+09 2.3560e+11 8869.4
## - pcincome  1 6.0893e+09 2.3869e+11 8875.1
## - area      1 7.0732e+09 2.3968e+11 8876.9
## - totalinc  1 3.0507e+10 2.6311e+11 8918.0
```

```
## - pop       1 7.2961e+10 3.0557e+11 8983.8
##
## Step:  AIC=8863.77
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + bagrad +
##     poverty + unemp + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - unemp     1 3.0922e+07 2.3264e+11 8861.8
## - bagrad    1 7.9111e+07 2.3269e+11 8861.9
## - pop18     1 1.6527e+08 2.3277e+11 8862.1
## - hsgrad    1 3.8763e+08 2.3300e+11 8862.5
## - region    1 6.6170e+08 2.3327e+11 8863.0
## <none>                   2.3261e+11 8863.8
## - beds      1 1.2871e+09 2.3390e+11 8864.2
## - docs      1 1.4796e+09 2.3409e+11 8864.6
## + pop65     1 3.8881e+06 2.3260e+11 8865.8
## - poverty   1 3.0262e+09 2.3563e+11 8867.5
## - pcincome  1 6.0969e+09 2.3871e+11 8873.2
## - area      1 7.0971e+09 2.3971e+11 8875.0
## - totalinc  1 3.0600e+10 2.6321e+11 8916.2
## - pop       1 7.4342e+10 3.0695e+11 8983.8
##
## Step:  AIC=8861.83
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + bagrad +
##     poverty + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - bagrad    1 1.1046e+08 2.3275e+11 8860.0
## - pop18     1 1.6702e+08 2.3281e+11 8860.1
## - hsgrad    1 3.6423e+08 2.3300e+11 8860.5
## - region    1 7.4425e+08 2.3338e+11 8861.2
## <none>                   2.3264e+11 8861.8
## - beds      1 1.4088e+09 2.3405e+11 8862.5
## - docs      1 1.4861e+09 2.3413e+11 8862.6
## + unemp     1 3.0922e+07 2.3261e+11 8863.8
## + pop65     1 2.6644e+06 2.3264e+11 8863.8
## - poverty   1 3.0992e+09 2.3574e+11 8865.7
## - pcincome  1 6.1584e+09 2.3880e+11 8871.3
## - area      1 7.3697e+09 2.4001e+11 8873.6
## - totalinc  1 3.0573e+10 2.6321e+11 8914.2
## - pop       1 7.4823e+10 3.0746e+11 8982.5
##
## Step:  AIC=8860.04
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + poverty +
##     pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - hsgrad    1 2.5445e+08 2.3300e+11 8858.5
## - pop18     1 5.5080e+08 2.3330e+11 8859.1
## - region    1 8.4589e+08 2.3360e+11 8859.6
## <none>                   2.3275e+11 8860.0
## - beds      1 1.3067e+09 2.3406e+11 8860.5
## - docs      1 1.3831e+09 2.3413e+11 8860.6
## + bagrad    1 1.1046e+08 2.3264e+11 8861.8
```

```
## + unemp     1 6.2267e+07 2.3269e+11 8861.9
## + pop65     1 3.5133e+06 2.3275e+11 8862.0
## - poverty   1 4.1542e+09 2.3690e+11 8865.8
## - area      1 7.5698e+09 2.4032e+11 8872.1
## - pcincome  1 1.3566e+10 2.4632e+11 8883.0
## - totalinc  1 3.1712e+10 2.6446e+11 8914.2
## - pop       1 7.6620e+10 3.0937e+11 8983.3
##
## Step:  AIC=8858.52
## crimes ~ area + pop + pop18 + docs + beds + poverty + pcincome +
##     totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - pop18     1 3.5255e+08 2.3336e+11 8857.2
## - region    1 6.6233e+08 2.3367e+11 8857.8
## <none>                   2.3300e+11 8858.5
## - beds      1 1.3147e+09 2.3432e+11 8859.0
## - docs      1 1.4252e+09 2.3443e+11 8859.2
## + hsgrad    1 2.5445e+08 2.3275e+11 8860.0
## + pop65     1 1.8070e+07 2.3299e+11 8860.5
## + unemp     1 4.9993e+06 2.3300e+11 8860.5
## + bagrad    1 6.8036e+05 2.3300e+11 8860.5
## - area      1 7.5866e+09 2.4059e+11 8870.6
## - poverty   1 7.9501e+09 2.4095e+11 8871.3
## - pcincome  1 1.3431e+10 2.4644e+11 8881.2
## - totalinc  1 3.1475e+10 2.6448e+11 8912.3
## - pop       1 7.6388e+10 3.0939e+11 8981.3
##
## Step:  AIC=8857.19
## crimes ~ area + pop + docs + beds + poverty + pcincome + totalinc +
##     region
##
##            Df  Sum of Sq        RSS    AIC
## - region    1 6.9649e+08 2.3405e+11 8856.5
## <none>                   2.3336e+11 8857.2
## - beds      1 1.1192e+09 2.3448e+11 8857.3
## - docs      1 1.1714e+09 2.3453e+11 8857.4
## + pop18     1 3.5255e+08 2.3300e+11 8858.5
## + bagrad    1 1.3798e+08 2.3322e+11 8858.9
## + pop65     1 6.8842e+07 2.3329e+11 8859.1
## + unemp     1 6.2992e+07 2.3329e+11 8859.1
## + hsgrad    1 5.6206e+07 2.3330e+11 8859.1
## - poverty   1 7.8900e+09 2.4125e+11 8869.8
## - area      1 8.0472e+09 2.4140e+11 8870.1
## - pcincome  1 1.3308e+10 2.4666e+11 8879.6
## - totalinc  1 3.2977e+10 2.6633e+11 8913.3
## - pop       1 7.8411e+10 3.1177e+11 8982.6
##
## Step:  AIC=8856.5
## crimes ~ area + pop + docs + beds + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## - beds      1 7.8044e+08 2.3483e+11 8856.0
## - docs      1 9.7325e+08 2.3503e+11 8856.3
```

```
## <none>                  2.3405e+11 8856.5
## + region    1 6.9649e+08 2.3336e+11 8857.2
## + pop18     1 3.8671e+08 2.3367e+11 8857.8
## + bagrad    1 3.0727e+08 2.3375e+11 8857.9
## + unemp     1 2.3537e+08 2.3382e+11 8858.1
## + pop65     1 1.3754e+08 2.3392e+11 8858.2
## + hsgrad    1 9.8089e+04 2.3405e+11 8858.5
## - area      1 7.3727e+09 2.4143e+11 8868.1
## - poverty   1 9.4412e+09 2.4349e+11 8871.9
## - pcincome  1 1.3301e+10 2.4735e+11 8878.8
## - totalinc  1 3.3807e+10 2.6786e+11 8913.9
## - pop       1 8.0981e+10 3.1503e+11 8985.2
##
## Step:  AIC=8855.96
## crimes ~ area + pop + docs + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## - docs      1 2.7324e+08 2.3511e+11 8854.5
## <none>                  2.3483e+11 8856.0
## + beds      1 7.8044e+08 2.3405e+11 8856.5
## + region    1 3.5777e+08 2.3448e+11 8857.3
## + unemp     1 2.2880e+08 2.3460e+11 8857.5
## + pop18     1 1.9526e+08 2.3464e+11 8857.6
## + bagrad    1 7.2185e+07 2.3476e+11 8857.8
## + hsgrad    1 1.7566e+07 2.3482e+11 8857.9
## + pop65     1 1.1459e+07 2.3482e+11 8857.9
## - area      1 9.3816e+09 2.4422e+11 8871.2
## - poverty   1 1.1656e+10 2.4649e+11 8875.3
## - pcincome  1 1.5123e+10 2.4996e+11 8881.4
## - totalinc  1 4.9534e+10 2.8437e+11 8938.2
## - pop       1 1.2635e+11 3.6119e+11 9043.4
##
## Step:  AIC=8854.47
## crimes ~ area + pop + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## <none>                  2.3511e+11 8854.5
## + region    1 3.9479e+08 2.3471e+11 8855.7
## + docs      1 2.7324e+08 2.3483e+11 8856.0
## + unemp     1 1.3334e+08 2.3497e+11 8856.2
## + pop18     1 1.2327e+08 2.3498e+11 8856.2
## + beds      1 8.0433e+07 2.3503e+11 8856.3
## + bagrad    1 3.3627e+07 2.3507e+11 8856.4
## + hsgrad    1 3.0775e+07 2.3508e+11 8856.4
## + pop65     1 1.9005e+07 2.3509e+11 8856.4
## - area      1 9.1118e+09 2.4422e+11 8869.2
## - poverty   1 1.2072e+10 2.4718e+11 8874.5
## - pcincome  1 1.4929e+10 2.5004e+11 8879.6
## - totalinc  1 5.4571e+10 2.8968e+11 8944.3
## - pop       1 1.2651e+11 3.6161e+11 9041.9

##
## Call:
## lm(formula = crimes ~ area + pop + poverty + pcincome + totalinc,
```

```
##     data = cdi_df)
##
## Coefficients:
## (Intercept)         area          pop      poverty     pcincome     totalinc
##  -6.389e+04   -3.109e+00    2.496e-01    1.450e+03    2.460e+00   -7.899e+00
```

```
#forward
forward = lm(formula = crimes ~ area + pop + pop18 + pop65 + docs + beds +
    hsgrad + bagrad + poverty + unemp + pcincome + totalinc +
    region, data = cdi_df)
stepf = step(forward)
```

```
## Start:  AIC=8865.77
## crimes ~ area + pop + pop18 + pop65 + docs + beds + hsgrad +
##     bagrad + poverty + unemp + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - pop65     1 3.8881e+06 2.3261e+11 8863.8
## - unemp     1 3.2146e+07 2.3264e+11 8863.8
## - bagrad    1 7.7720e+07 2.3268e+11 8863.9
## - pop18     1 1.4795e+08 2.3275e+11 8864.0
## - hsgrad    1 3.7187e+08 2.3298e+11 8864.5
## - region    1 6.6550e+08 2.3327e+11 8865.0
## <none>                   2.3260e+11 8865.8
## - beds      1 1.2164e+09 2.3382e+11 8866.1
## - docs      1 1.4817e+09 2.3409e+11 8866.6
## - poverty   1 2.9923e+09 2.3560e+11 8869.4
## - pcincome  1 6.0893e+09 2.3869e+11 8875.1
## - area      1 7.0732e+09 2.3968e+11 8876.9
## - totalinc  1 3.0507e+10 2.6311e+11 8918.0
## - pop       1 7.2961e+10 3.0557e+11 8983.8
##
## Step:  AIC=8863.77
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + bagrad +
##     poverty + unemp + pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - unemp     1 3.0922e+07 2.3264e+11 8861.8
## - bagrad    1 7.9111e+07 2.3269e+11 8861.9
## - pop18     1 1.6527e+08 2.3277e+11 8862.1
## - hsgrad    1 3.8763e+08 2.3300e+11 8862.5
## - region    1 6.6170e+08 2.3327e+11 8863.0
## <none>                   2.3261e+11 8863.8
## - beds      1 1.2871e+09 2.3390e+11 8864.2
## - docs      1 1.4796e+09 2.3409e+11 8864.6
## - poverty   1 3.0262e+09 2.3563e+11 8867.5
## - pcincome  1 6.0969e+09 2.3871e+11 8873.2
## - area      1 7.0971e+09 2.3971e+11 8875.0
## - totalinc  1 3.0600e+10 2.6321e+11 8916.2
## - pop       1 7.4342e+10 3.0695e+11 8983.8
##
## Step:  AIC=8861.83
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + bagrad +
##     poverty + pcincome + totalinc + region
```

```
##
##            Df  Sum of Sq        RSS    AIC
## - bagrad    1 1.1046e+08 2.3275e+11 8860.0
## - pop18     1 1.6702e+08 2.3281e+11 8860.1
## - hsgrad    1 3.6423e+08 2.3300e+11 8860.5
## - region    1 7.4425e+08 2.3338e+11 8861.2
## <none>                   2.3264e+11 8861.8
## - beds      1 1.4088e+09 2.3405e+11 8862.5
## - docs      1 1.4861e+09 2.3413e+11 8862.6
## - poverty   1 3.0992e+09 2.3574e+11 8865.7
## - pcincome  1 6.1584e+09 2.3880e+11 8871.3
## - area      1 7.3697e+09 2.4001e+11 8873.6
## - totalinc  1 3.0573e+10 2.6321e+11 8914.2
## - pop       1 7.4823e+10 3.0746e+11 8982.5
##
## Step:  AIC=8860.04
## crimes ~ area + pop + pop18 + docs + beds + hsgrad + poverty +
##     pcincome + totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - hsgrad    1 2.5445e+08 2.3300e+11 8858.5
## - pop18     1 5.5080e+08 2.3330e+11 8859.1
## - region    1 8.4589e+08 2.3360e+11 8859.6
## <none>                   2.3275e+11 8860.0
## - beds      1 1.3067e+09 2.3406e+11 8860.5
## - docs      1 1.3831e+09 2.3413e+11 8860.6
## - poverty   1 4.1542e+09 2.3690e+11 8865.8
## - area      1 7.5698e+09 2.4032e+11 8872.1
## - pcincome  1 1.3566e+10 2.4632e+11 8883.0
## - totalinc  1 3.1712e+10 2.6446e+11 8914.2
## - pop       1 7.6620e+10 3.0937e+11 8983.3
##
## Step:  AIC=8858.52
## crimes ~ area + pop + pop18 + docs + beds + poverty + pcincome +
##     totalinc + region
##
##            Df  Sum of Sq        RSS    AIC
## - pop18     1 3.5255e+08 2.3336e+11 8857.2
## - region    1 6.6233e+08 2.3367e+11 8857.8
## <none>                   2.3300e+11 8858.5
## - beds      1 1.3147e+09 2.3432e+11 8859.0
## - docs      1 1.4252e+09 2.3443e+11 8859.2
## - area      1 7.5866e+09 2.4059e+11 8870.6
## - poverty   1 7.9501e+09 2.4095e+11 8871.3
## - pcincome  1 1.3431e+10 2.4644e+11 8881.2
## - totalinc  1 3.1475e+10 2.6448e+11 8912.3
## - pop       1 7.6388e+10 3.0939e+11 8981.3
##
## Step:  AIC=8857.19
## crimes ~ area + pop + docs + beds + poverty + pcincome + totalinc +
##     region
##
##            Df  Sum of Sq        RSS    AIC
## - region    1 6.9649e+08 2.3405e+11 8856.5
```

```
## <none>                  2.3336e+11 8857.2
## - beds      1 1.1192e+09 2.3448e+11 8857.3
## - docs      1 1.1714e+09 2.3453e+11 8857.4
## - poverty   1 7.8900e+09 2.4125e+11 8869.8
## - area      1 8.0472e+09 2.4140e+11 8870.1
## - pcincome  1 1.3308e+10 2.4666e+11 8879.6
## - totalinc  1 3.2977e+10 2.6633e+11 8913.3
## - pop       1 7.8411e+10 3.1177e+11 8982.6
##
## Step:  AIC=8856.5
## crimes ~ area + pop + docs + beds + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## - beds      1 7.8044e+08 2.3483e+11 8856.0
## - docs      1 9.7325e+08 2.3503e+11 8856.3
## <none>                  2.3405e+11 8856.5
## - area      1 7.3727e+09 2.4143e+11 8868.1
## - poverty   1 9.4412e+09 2.4349e+11 8871.9
## - pcincome  1 1.3301e+10 2.4735e+11 8878.8
## - totalinc  1 3.3807e+10 2.6786e+11 8913.9
## - pop       1 8.0981e+10 3.1503e+11 8985.2
##
## Step:  AIC=8855.96
## crimes ~ area + pop + docs + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## - docs      1 2.7324e+08 2.3511e+11 8854.5
## <none>                  2.3483e+11 8856.0
## - area      1 9.3816e+09 2.4422e+11 8871.2
## - poverty   1 1.1656e+10 2.4649e+11 8875.3
## - pcincome  1 1.5123e+10 2.4996e+11 8881.4
## - totalinc  1 4.9534e+10 2.8437e+11 8938.2
## - pop       1 1.2635e+11 3.6119e+11 9043.4
##
## Step:  AIC=8854.47
## crimes ~ area + pop + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## <none>                  2.3511e+11 8854.5
## - area      1 9.1118e+09 2.4422e+11 8869.2
## - poverty   1 1.2072e+10 2.4718e+11 8874.5
## - pcincome  1 1.4929e+10 2.5004e+11 8879.6
## - totalinc  1 5.4571e+10 2.8968e+11 8944.3
## - pop       1 1.2651e+11 3.6161e+11 9041.9
```

```
summary(stepf)
```

```
##
## Call:
## lm(formula = crimes ~ area + pop + poverty + pcincome + totalinc,
##     data = cdi_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -77382  -5773    285   5117 406547
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.389e+04  1.023e+04  -6.244 1.02e-09 ***
## area        -3.109e+00  7.580e-01  -4.101 4.91e-05 ***
## pop          2.496e-01  1.633e-02  15.282  < 2e-16 ***
## poverty      1.450e+03  3.071e+02   4.721 3.18e-06 ***
## pcincome     2.460e+00  4.685e-01   5.250 2.39e-07 ***
## totalinc    -7.899e+00  7.870e-01 -10.037  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23270 on 434 degrees of freedom
## Multiple R-squared:  0.8421, Adjusted R-squared:  0.8403
## F-statistic: 462.9 on 5 and 434 DF,  p-value: < 2.2e-16
```

```
#backward
backward = lm(formula = crimes ~ area + pop + poverty + pcincome + totalinc,
    data = cdi_df)
stepb = step(backward)
```

```
## Start:  AIC=8854.47
## crimes ~ area + pop + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## <none>                   2.3511e+11 8854.5
## - area      1 9.1118e+09 2.4422e+11 8869.2
## - poverty   1 1.2072e+10 2.4718e+11 8874.5
## - pcincome  1 1.4929e+10 2.5004e+11 8879.6
## - totalinc  1 5.4571e+10 2.8968e+11 8944.3
## - pop       1 1.2651e+11 3.6161e+11 9041.9
```

```
summary(stepb)
```

```
##
## Call:
## lm(formula = crimes ~ area + pop + poverty + pcincome + totalinc,
##     data = cdi_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -77382  -5773    285   5117 406547
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.389e+04  1.023e+04  -6.244 1.02e-09 ***
## area        -3.109e+00  7.580e-01  -4.101 4.91e-05 ***
## pop          2.496e-01  1.633e-02  15.282  < 2e-16 ***
## poverty      1.450e+03  3.071e+02   4.721 3.18e-06 ***
## pcincome     2.460e+00  4.685e-01   5.250 2.39e-07 ***
## totalinc    -7.899e+00  7.870e-01 -10.037  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23270 on 434 degrees of freedom
## Multiple R-squared:  0.8421, Adjusted R-squared:  0.8403
## F-statistic: 462.9 on 5 and 434 DF,  p-value: < 2.2e-16
```

```
#stepwise
stepwise = lm(formula = crimes ~ area + pop + poverty + pcincome + totalinc,
    data = cdi_df)
steps = step(stepwise)
```

```
## Start:  AIC=8854.47
## crimes ~ area + pop + poverty + pcincome + totalinc
##
##            Df  Sum of Sq        RSS    AIC
## <none>                   2.3511e+11 8854.5
## - area      1 9.1118e+09 2.4422e+11 8869.2
## - poverty   1 1.2072e+10 2.4718e+11 8874.5
## - pcincome  1 1.4929e+10 2.5004e+11 8879.6
## - totalinc  1 5.4571e+10 2.8968e+11 8944.3
## - pop       1 1.2651e+11 3.6161e+11 9041.9
```

```
summary(steps)
```

```
##
## Call:
## lm(formula = crimes ~ area + pop + poverty + pcincome + totalinc,
##     data = cdi_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -77382  -5773    285   5117 406547
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.389e+04  1.023e+04  -6.244 1.02e-09 ***
## area        -3.109e+00  7.580e-01  -4.101 4.91e-05 ***
## pop          2.496e-01  1.633e-02  15.282  < 2e-16 ***
## poverty      1.450e+03  3.071e+02   4.721 3.18e-06 ***
## pcincome     2.460e+00  4.685e-01   5.250 2.39e-07 ***
## totalinc    -7.899e+00  7.870e-01 -10.037  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23270 on 434 degrees of freedom
## Multiple R-squared:  0.8421, Adjusted R-squared:  0.8403
## F-statistic: 462.9 on 5 and 434 DF,  p-value: < 2.2e-16
```

Call: lm(formula = crimes ~ id + cty + state + area + pop + pop18 + pop65 + docs + beds + bagrad + poverty + unemp + totalinc, data = cdi_df)

Test Based Procedures

```
mat = as.matrix(cdi_df)
# Printing the 2 best models of each size, using the Cp criterion:
leaps(x = mat[,2:14], y = mat[,1], nbest = 2, method = "Cp")
```

```
## $which
##        1     2     3     4     5     6     7     8     9     A     B     C
## 1  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 1  FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
## 2  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## 2  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
## 3   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## 3  FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
## 4  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 4   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 5   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 5  FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 6   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 6   TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 7   TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 7   TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 8   TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 8   TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 9   TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 9   TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
## 10  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE
## 10  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 11  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 11  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## 12  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## 12  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 13  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##        D
## 1  FALSE
## 1  FALSE
## 2  FALSE
## 2  FALSE
## 3  FALSE
## 3  FALSE
## 4  FALSE
## 4  FALSE
## 5  FALSE
## 5  FALSE
## 6   TRUE
## 6  FALSE
## 7  FALSE
## 7   TRUE
## 8   TRUE
## 8  FALSE
## 9   TRUE
## 9   TRUE
## 10  TRUE
```

```
## 10  TRUE
## 11  TRUE
## 11  TRUE
## 12  TRUE
## 12  TRUE
## 13  TRUE
##
## $label
##  [1] "(Intercept)" "1"           "2"           "3"           "4"
##  [6] "5"           "6"           "7"           "8"           "9"
## [11] "A"           "B"           "C"           "D"
##
## $size
##  [1]  2  2  3  3  4  4  5  5  6  6  7  7  8  8  9  9 10 10 11 11 12 12 13 13 14
##
## $Cp
##  [1] 148.680323 288.819844  47.994680 104.025933  32.345656  36.301818
##  [7]  17.270684  22.692013   2.583062  16.477810   3.860033   4.082634
## [13]   4.653306   5.427401   5.377730   5.945070   6.732057   7.125020
## [19]   8.266044   8.369635  10.063753  10.152007  12.007121  12.058873
## [25]  14.000000
```

```
# Printing the 2 best models of each size, using the adjusted R^2 criterion:
leaps(x = mat[,2:14], y = mat[,1], nbest = 2, method = "adjr2")
```

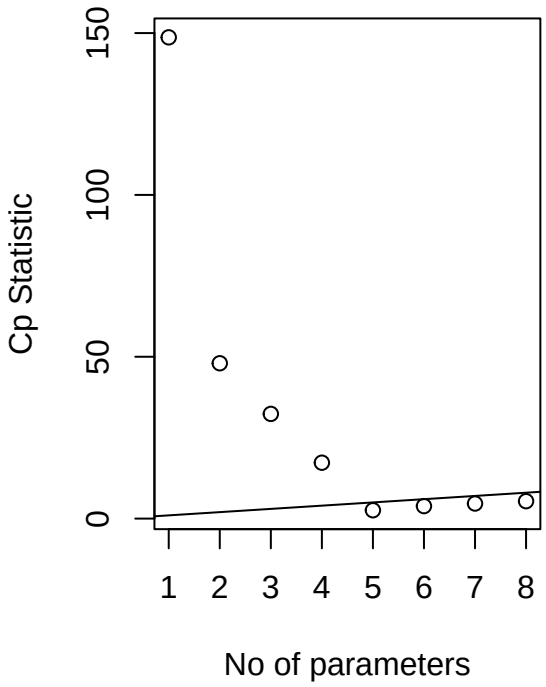
```
## $which
##        1     2     3     4     5     6     7     8     9     A     B     C
## 1  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 1  FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
## 2  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## 2  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
## 3   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## 3  FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
## 4  FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 4   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 5   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 5  FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 6   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 6   TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 7   TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 7   TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 8   TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 8   TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 9   TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 9   TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
## 10  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE
## 10  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 11  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 11  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## 12  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## 12  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 13  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##        D
## 1  FALSE
```

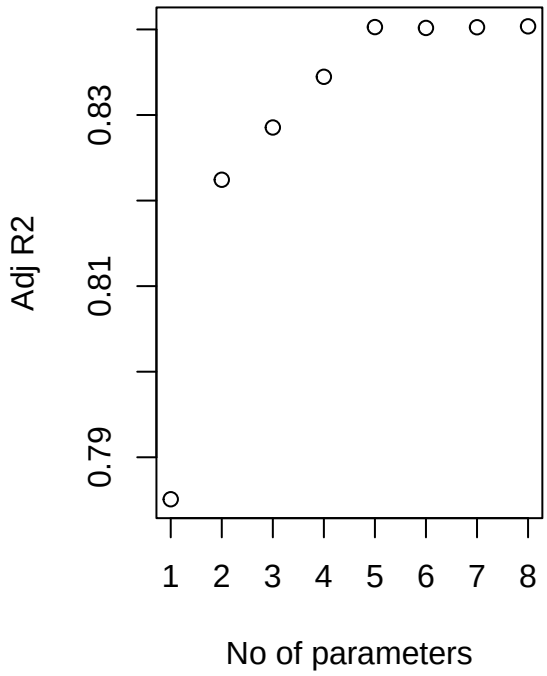
```
## 1  FALSE
## 2  FALSE
## 2  FALSE
## 3  FALSE
## 3  FALSE
## 4  FALSE
## 4  FALSE
## 5  FALSE
## 5  FALSE
## 6   TRUE
## 6  FALSE
## 7  FALSE
## 7   TRUE
## 8   TRUE
## 8  FALSE
## 9   TRUE
## 9   TRUE
## 10  TRUE
## 10  TRUE
## 11  TRUE
## 11  TRUE
## 12  TRUE
## 12  TRUE
## 13  TRUE
##
## $label
##  [1] "(Intercept)" "1"           "2"           "3"           "4"
##  [6] "5"           "6"           "7"           "8"           "9"
## [11] "A"           "B"           "C"           "D"
##
## $size
##  [1]  2  2  3  3  4  4  5  5  6  6  7  7  8  8  9  9 10 10 11 11 12 12 13 13 14
##
## $adjr2
##  [1] 0.7850946 0.7335849 0.8224324 0.8017904 0.8285420 0.8270812 0.8344672
##  [8] 0.8324608 0.8402760 0.8351218 0.8401760 0.8400932 0.8402557 0.8399672
## [15] 0.8403615 0.8401496 0.8402320 0.8400849 0.8400345 0.8399956 0.8397368
## [22] 0.8397036 0.8393828 0.8393633 0.8390085
```

```r
# Function regsubsets() performs a subset selection by identifying the "best" model that contains
# a certain number of predictors. By default "best" is chosen using SSE/RSS (smaller is better)
b = regsubsets(crimes ~ ., data = cdi_df)
rs = summary(b)

# plot of Cp and Adj-R2 as functions of parameters
par(mfrow=c(1,2))

plot(1:8, rs$cp, xlab="No of parameters", ylab="Cp Statistic")
abline(0,1)
#6
plot(1:8, rs$adjr2, xlab="No of parameters", ylab="Adj R2")
```


Cp Statistic
0
50
100
150
1
2
3
4
5
6
7
8
No of parameters


Adj R2
0.79
0.81
0.83
1
2
3
4
5
6
7
8
No of parameters

```
#6
```