Demographic

Michael Hall

9/27/2020

Prepare for analyses

```
Loading required package: xts

Loading required package: zoo


Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric


Attaching package: 'PerformanceAnalytics'

The following object is masked from 'package:graphics':

    legend


Attaching package: 'olsrr'

The following object is masked from 'package:datasets':

    rivers
```

Read in the data from my working directory

The head of the data set providing a table of observations with Physician

```
Warning: package 'knitr' was built under R version 3.6.3
```

Table 1: Demographic Data

Y	X1	X2	X3	X4	X5
23677	8863164	184230	0	0	0
15153	5105067	110928	0	1	0
7553	2818199	55003	0	0	1
5905	2498016	48931	0	0	0
6062	2410556	58818	0	0	0

The number of active physicians (Y) is regressed against

total population (X1), total personal income (X2), and

geographic area (X3,X4,X5)

If X3 = 1 then the observation is located in the North East region and 0 otherwise.

If X4 = 1 then the observation is located in the North Central Region and 0 otherwise.

If X5 = 1 then the observation is located in the Southern region and 0 otherwise.

Fitting a first order regression model to determine if geographic effects are present at the significance level of 10%.

The Model is regressed against the number of physicians.

```
model <- lm(Y ~ X1 + X2 + X3 + X4 + X5, data = Region)
summary(model)
```

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5, data = Region)

Residuals:
    Min      1Q  Median      3Q     Max
-1866.8  -207.7   -81.5    72.4  3721.7

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -2.075e+02  7.028e+01  -2.952  0.00332 **
X1           5.515e-04  2.835e-04   1.945  0.05243 .
X2           1.070e-01  1.325e-02   8.073  6.8e-15 ***
X3           1.490e+02  8.683e+01   1.716  0.08685 .
X4           1.455e+02  8.515e+01   1.709  0.08817 .
X5           1.912e+02  8.003e+01   2.389  0.01731 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 566.1 on 434 degrees of freedom
Multiple R-squared:  0.9011,    Adjusted R-squared:  0.8999
F-statistic: 790.7 on 5 and 434 DF,  p-value: < 2.2e-16
```

The overall model is significant and each regressor is significant at a 10% significance level.

The Southern region is most statistically significant with a p-value of 1.731%

The Northeast region is statistically significant at p-value 8.685%

North Central region is statistcally significant at p-value 8.817%

The intercept of the model is close to zero

If a randomly selected State and City are in the southern indexed region holding all other predictors constant, the number of predicted physicians increase by approximately 191.

If a randomly selected State and City are in the Northeast region holding all other predictors constant, the number of predicted physicians increases by approximately 149.

If a randomly selected State and City are in the North Central region holding all other predictors constant, the number of predicted physicians increases by approximately 145.5.

The coefficient for X1(Total Population) is significant but close to zero, however, because it is associated with population, the interpretation is perhaps more important than its relative proximity to zero. According the the model, for every 100,000 people in a population there is approximately 55.15 physicians

The coefficient for X2(Total Personal Income) is significant. According to the model, for every additional $1000 in personal income the number of predicted physicians increases by 107.

A public safety official wishes to predict the rate of serious crimes.

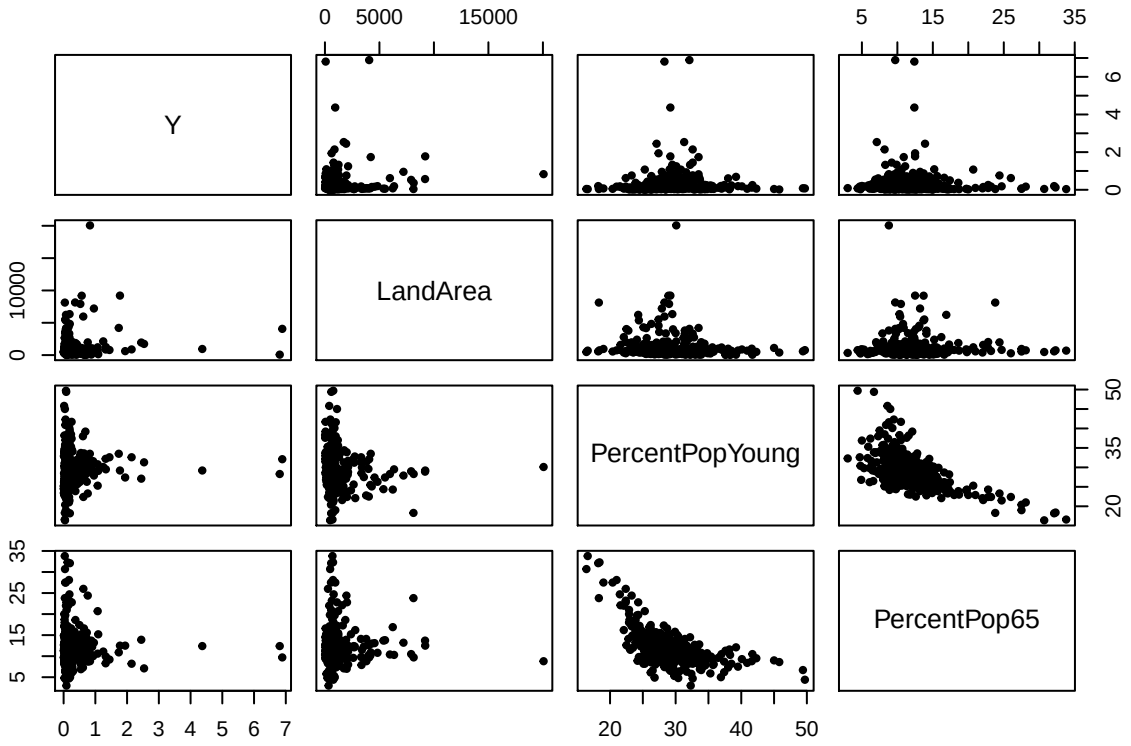
Y:Serious Crimes per 100,000 population

```
par(mfrow=c(3,3))
Training<-read.csv("C:/Users/micha/OneDrive/Documents/Demographic/Demographic/Training.csv")
attach(Training)
```

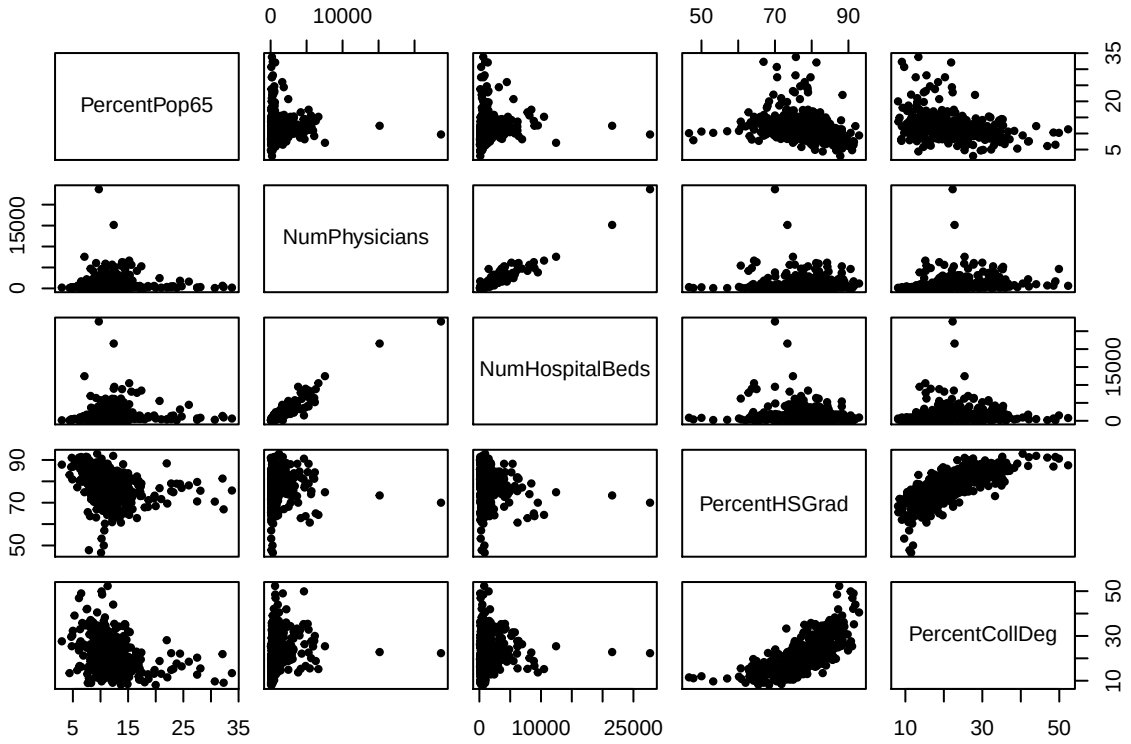
```
The following object is masked from Region:

    Y
```

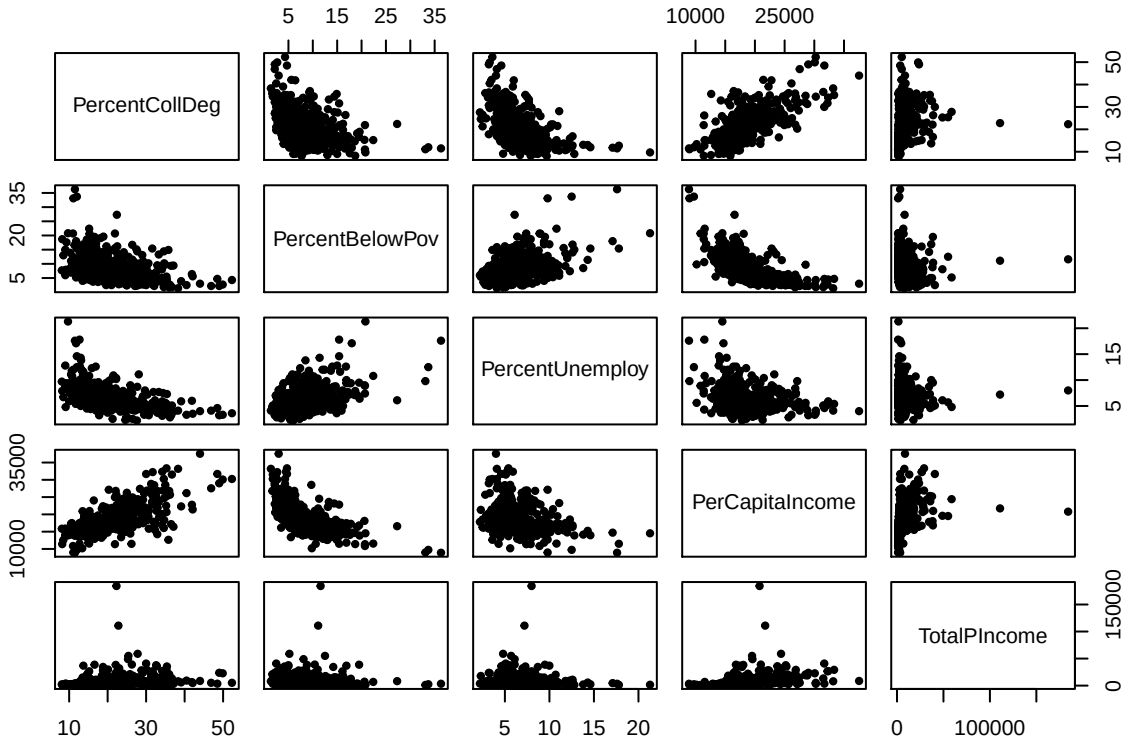
```
pairs(Training[1:4],col="black",pch=20)
```




```
pairs(Training[4:8],col="black",pch=20)
```




```
pairs(Training[8:12],col="black",pch=20)
```




```
cor(Training)
```

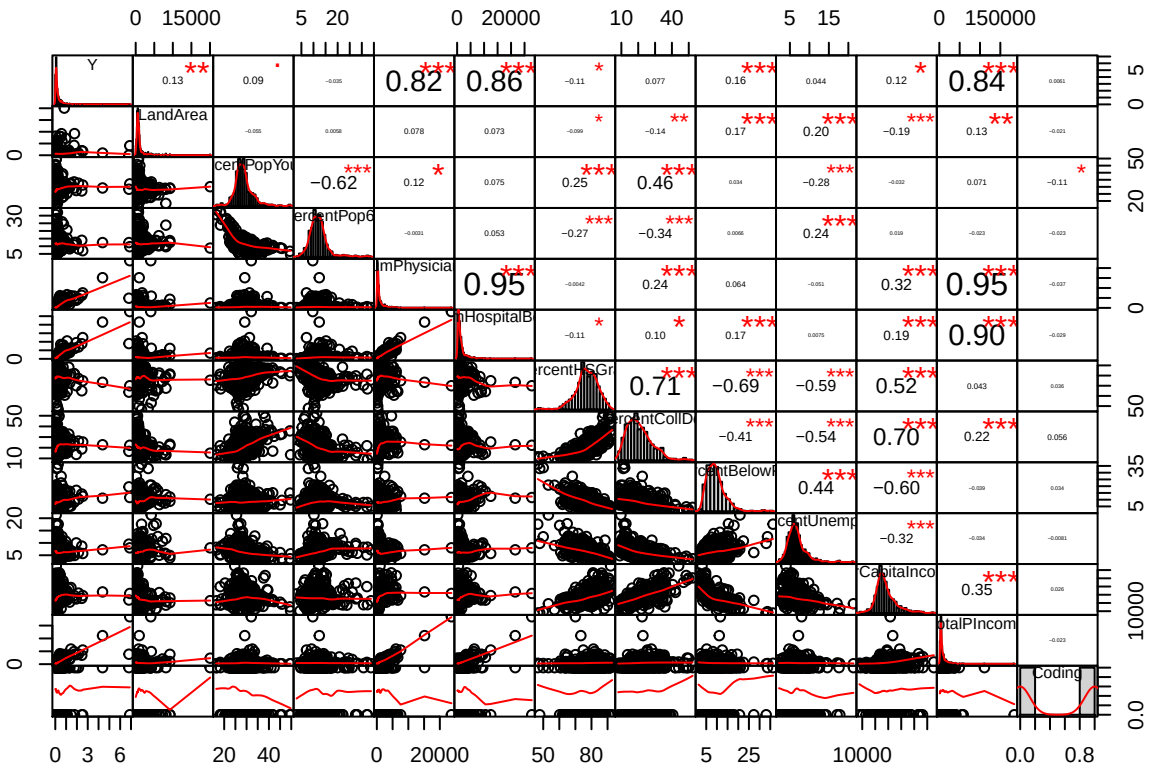
```
                      Y     LandArea PercentPopYoung PercentPop65
Y                1.000000000  0.129475371      0.08994063 -0.035290324
LandArea         0.129475371  1.000000000     -0.05487812  0.005770871
PercentPopYoung  0.089940629 -0.054878125      1.00000000 -0.616309639
PercentPop65    -0.035290324  0.005770871     -0.61630964  1.000000000
NumPhysicians    0.820459477  0.078074657      0.11969924 -0.003128630
NumHospitalBeds  0.856849883  0.073047270      0.07453191  0.053278417
PercentHSGrad   -0.106328401 -0.098598111      0.25058429 -0.268251758
PercentCollDeg   0.077076521 -0.137237736      0.45609703 -0.339228765
PercentBelowPov  0.164405659  0.171343348      0.03397551  0.006578474
PercentUnemploy  0.043556752  0.199209277     -0.27852706  0.236309411
PerCapitaIncome  0.117539139 -0.187715132     -0.03164843  0.018590706
TotalPIncome     0.843098049  0.127074261      0.07116151 -0.022733151
Coding           0.006070083 -0.020806227     -0.10993607 -0.023079872
                NumPhysicians NumHospitalBeds PercentHSGrad PercentCollDeg
Y                0.820459477     0.856849883  -0.106328401     0.07707652
LandArea         0.078074657     0.073047270  -0.098598111    -0.13723774
PercentPopYoung  0.119699240     0.074531907   0.250584290     0.45609703
PercentPop65    -0.003128630     0.053278417  -0.268251758    -0.33922877
NumPhysicians    1.000000000     0.950464395  -0.004248085     0.23676547
NumHospitalBeds  0.950464395     1.000000000  -0.111916382     0.10042653
PercentHSGrad   -0.004248085    -0.111916382   1.000000000     0.70778672
PercentCollDeg   0.236765466     0.100426534   0.707786723     1.00000000
```

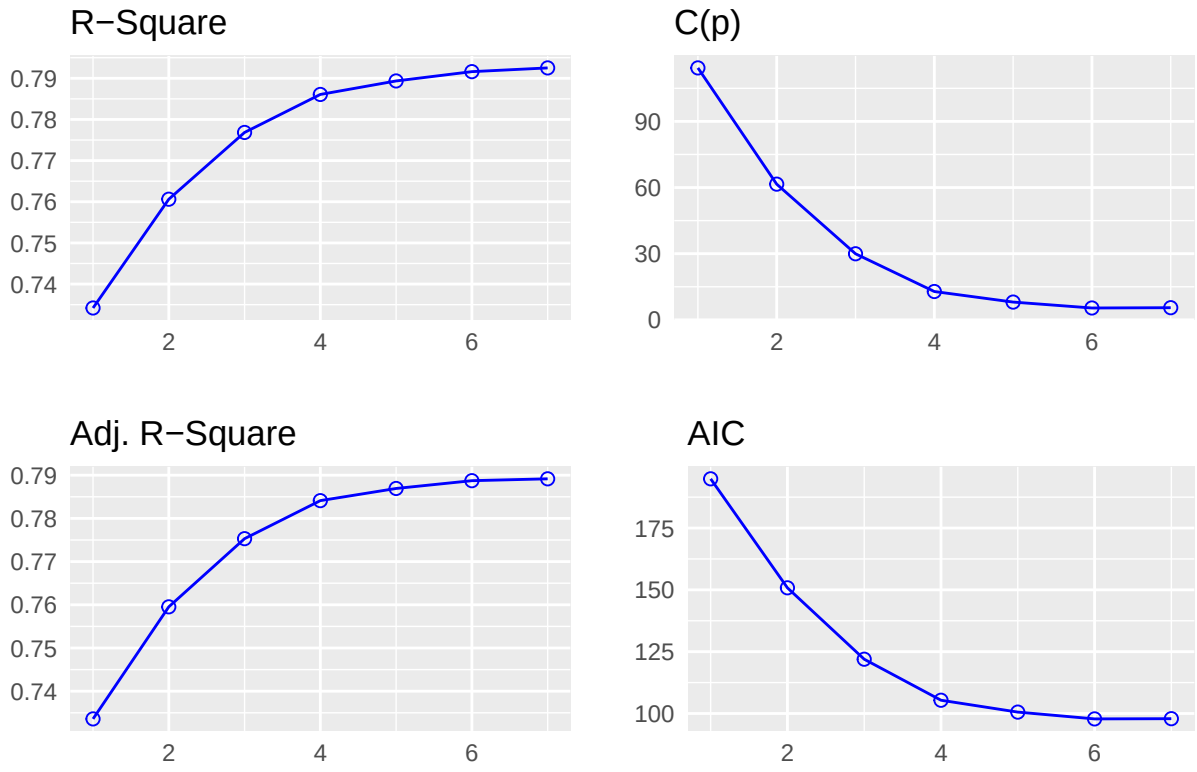
```
PercentBelowPov   0.064136254     0.172793840  -0.691750483    -0.40842385
PercentUnemploy  -0.050516116     0.007523992  -0.593595788    -0.54090691
PerCapitaIncome   0.316134625     0.194808180   0.522996133     0.69536186
TotalPIncome      0.948110571     0.902061545   0.043355729     0.22223013
Coding           -0.036802957    -0.028671853   0.036293926     0.05555620
                PercentBelowPov PercentUnemploy PerCapitaIncome TotalPIncome
Y                  0.164405659     0.043556752      0.11753914   0.84309805
LandArea           0.171343348     0.199209277     -0.18771513   0.12707426
PercentPopYoung    0.033975512    -0.278527058     -0.03164843   0.07116151
PercentPop65       0.006578474     0.236309411      0.01859071  -0.02273315
NumPhysicians      0.064136254    -0.050516116      0.31613462   0.94811057
NumHospitalBeds    0.172793840     0.007523992      0.19480818   0.90206155
PercentHSGrad     -0.691750483    -0.593595788      0.52299613   0.04335573
PercentCollDeg    -0.408423848    -0.540906913      0.69536186   0.22223013
PercentBelowPov    1.000000000     0.436947236     -0.60172504  -0.03873934
PercentUnemploy    0.436947236     1.000000000     -0.32214439  -0.03387633
PerCapitaIncome   -0.601725039    -0.322144395      1.00000000   0.34768161
TotalPIncome      -0.038739339    -0.033876330      0.34768161   1.00000000
Coding             0.034251353    -0.008077726      0.02626549  -0.02339432
                      Coding
Y                 0.006070083
LandArea         -0.020806227
PercentPopYoung  -0.109936072
PercentPop65     -0.023079872
NumPhysicians    -0.036802957
NumHospitalBeds  -0.028671853
PercentHSGrad     0.036293926
PercentCollDeg    0.055556197
PercentBelowPov   0.034251353
PercentUnemploy  -0.008077726
PerCapitaIncome   0.026265493
TotalPIncome     -0.023394324
Coding            1.000000000
```

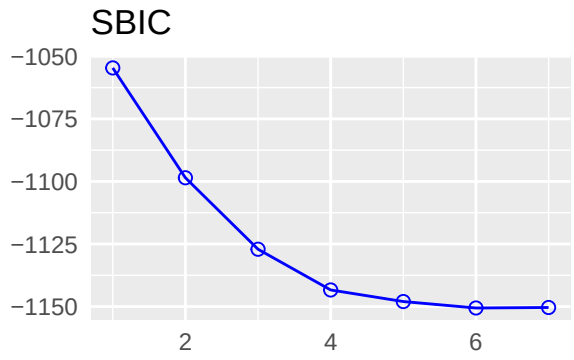
```
VariableVector<-c("Number of Serious Crimes per 100,000 population","LandArea","PercentYoungPopulatoin"
```

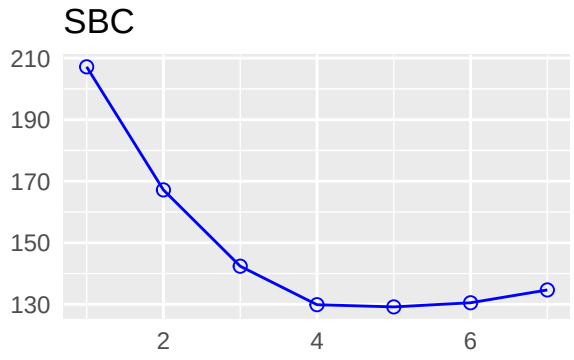
Table 2: Demographic Data

x
Number of Serious Crimes per 100,000 population
LandArea
PercentYoungPopulatoin
PercentOLDPopulation
NumberofPhysicians
NumberofHospitalBeds
PercentHSGraduation
PercentBachelorsDegree
PercentBelowPoverty
PercentUnemployed
PerCapitaIncome
TotalPersonalIncome


Y
LandArea
Coding


R−Square
0.79
0.78
0.77
0.76
0.75
0.74
2
4
6
C(p)
90
60
30
0
2
4
6
Adj. R−Square
0.79
0.78
0.77
0.76
0.75
0.74
2
4
6
AIC
175
150
125
100
2
4
6





Selection Summary

Step	Variable Entered	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	NumHospitalBeds	0.7342	0.7336	114.2475	194.9098	0.3006
2	TotalPIncome	0.7606	0.7595	61.5365	150.8300	0.2856
3	PerCapitaIncome	0.7768	0.7753	29.9587	121.9571	0.2761
4	NumPhysicians	0.7861	0.7841	12.8438	105.3634	0.2706
5	PercentCollDeg	0.7894	0.7869	8.0539	100.5649	0.2688
6	PercentBelowPov	0.7916	0.7887	5.3752	97.8183	0.2677
7	PercentPop65	0.7925	0.7892	5.5000	97.9015	0.2674

```
Call:
lm(formula = Y ~ NumHospitalBeds + TotalPIncome + PerCapitaIncome +
    NumPhysicians + PercentCollDeg + PercentBelowPov, data = Training)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8810 -0.0682 -0.0033  0.0532  4.7587

Coefficients:
```

```
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.729e-02  1.050e-01   0.545 0.585729
NumHospitalBeds  1.722e-04  2.029e-05   8.484 3.47e-16 ***
TotalPIncome     3.473e-05  3.349e-06  10.370  < 2e-16 ***
PerCapitaIncome -1.854e-05  5.357e-06  -3.460 0.000593 ***
NumPhysicians   -1.738e-04  3.363e-05  -5.169 3.61e-07 ***
PercentCollDeg   6.252e-03  2.424e-03   2.580 0.010219 *
PercentBelowPov  8.418e-03  3.884e-03   2.167 0.030772 *
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 0.2677 on 433 degrees of freedom
Multiple R-squared:  0.7916,    Adjusted R-squared:  0.7887
F-statistic: 274.1 on 6 and 433 DF,  p-value: < 2.2e-16
```