Sta 108 Final Project

Daniel Cayapan

2023-06-11

```r
# First off, I need to set the directory to my downloads, and import the dataset of Demographic.txt to
# get the data
filepath <- '/Users/danie/Downloads/'
dat <- read.table(paste0(filepath,'Demographic.txt'))
head(dat)
```

```
##   V1          V2 V3   V4      V5   V6   V7    V8    V9    V10  V11  V12  V13
## 1  1 Los_Angeles CA 4060 8863164 32.1  9.7 23677 27700 688936 70.0 22.3 11.6
## 2  2        Cook IL  946 5105067 29.2 12.4 15153 21550 436936 73.4 22.8 11.1
## 3  3      Harris TX 1729 2818199 31.3  7.1  7553 12449 253526 74.9 25.4 12.5
## 4  4   San_Diego CA 4205 2498016 33.5 10.9  5905  6179 173821 81.9 25.3  8.1
## 5  5      Orange CA  790 2410556 32.6  9.2  6062  6369 144524 81.2 27.8  5.2
## 6  6       Kings NY   71 2300664 28.3 12.4  4861  8942 680966 63.7 16.6 19.5
##   V14   V15    V16 V17
## 1 8.0 20786 184230   4
## 2 7.2 21729 110928   2
## 3 5.7 19517  55003   3
## 4 6.1 19588  48931   4
## 5 4.8 24400  58818   4
## 6 9.5 16803  38658   1
```

```r
n = nrow(dat)
n
```

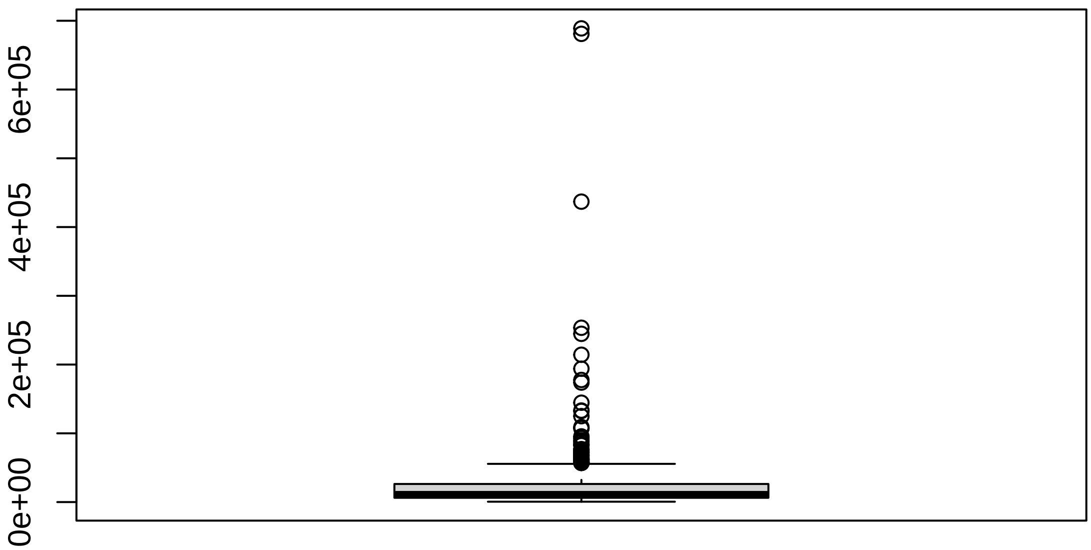
```
## [1] 440
```

```r
depend = dat[,10]
depend
```

```
##   [1] 688936 436936 253526 173821 144524 680966 177593 193978 244725 214258
##  [11] 109148 124959  77009  83110  73150  35825  50186  66723  43203 107338
##  [21] 107386 133098  95494 132495  50964  84305  71753  42595  55306  82680
##  [31]  73681  37118  57208  76142  51926  62344  51032  89895  44374  67032
##  [41]  28521  30202  52903  51243  61004  29708  75595  34754  52786  87355
##  [51]  54469  58610  71234  41048  43780   7099  46789  20335  52577  68586
##  [61]  30548  34312  30235  62004  57051  32419  68808  55604  30473  93025
##  [71]  61760  14830  64393  57045  34627  54002  41980  58216  60961  36665
##  [81]  22302  40581  18924  56950  36318  16894  22349  26228  57999  45237
##  [91]  22023  39496  34814  31553  54238  30299  30574  41179  41280  29926
## [101]  23249  37466  20323  38194  24247  26434  41625  27582  27717  23453
```

```
## [111] 14846 17379 33136 21826 26006 13086 12147 26006 37290 34071
## [121] 23686 25234 64103 13034 23412 27587 18556 26712 41592 30409
## [131]  9491 25736 25461  7445 18313 17466 20344 20042 16432 17870
## [141] 15238 17119 42404 28212 21756 11292 12827 12843 22422 18442
## [151]  4982 20153 20504 16721 19489 12630 10975  8935 19842 28190
## [161] 15077 13850 28606 12254 17518 24101 23532 19367 16091 17337
## [171] 12509 12855 19801  9936 10953 25247 20372  9864 21554  7194
## [181] 19369 13181 17625 10605 14563 14380 25194 18842  9087  7807
## [191] 18831 28841 19674  9234 10666 22091  6452 10637  7295 13816
## [201]  8308 21677  6635 10706 11563  9460  9746 13707 14825  7315
## [211] 17198  9433 17378 14124 25167 15306 14509  9437  4368 18586
## [221]  8103 18732  9001  5481 16414  7785  7435 16190 16916 17388
## [231] 13551 10246  9701  4526  7518  5247 14860  8101 23363  8692
## [241] 10865  8996 17918 16486 15139  6279  6140  9057  7441  8921
## [251]  4088  4854 12483  9469  6735  8939 15419 15477 18218  5281
## [261] 11454  8587 11929 11865 11110  9512 12194  3593 11508  8904
## [271]  3128 12526 11776  6399  9814  7043  4701  8405 10599  6027
## [281]  3187 18902 12229  7882  7302 13845 12181  9191  7573  5297
## [291] 11085  2119  7099  9426  3420  8427 11712 12377  4939  5114
## [301]  9842 12701  7505  8630  6364  7154 10131  7336  4749  2547
## [311]  7170  5365  6568  3612  7525  4447 12459  5153  6236  3409
## [321]  2769 10605  6925  5178 10650  4860  6170  5724  7643  3862
## [331]  9071  5221  9810  8850  8376  6726 12202  7901  4739  6101
## [341]  9525  3174  5373  6141  3196  5949  4014  1799  1380  6616
## [351]  7070  8634  5662  7977  5152  5625  4136  3140  3430  8640
## [361] 10727   563  8203  3582  3155  2777  6835  7226 11892  3325
## [371]  1064  6785  5737 14643  4101  3826  6072  6830  6103  6275
## [381]  4901  4184  3760  3655  4777  8421  2919  3879  6122  3759
## [391]  6249  5456  3851  4849  5979  8537  5846  3741  3064  5056
## [401]  5233  3838  4734  8042  1657  1699   898  4152  4603  2068
## [411]  9665  3952  6021  2982  5414  7546  2940  4310  3435  4560
## [421]  4020  4318  9181  4682  3651  4433   625  3285  3401  7138
## [431]  2689  1397  4449  2377  5279  4414  5081  6537  7130  4693
```

```
boxplot(depend)
```




```r
# The fixed dependent variable is 10. Total # of Serious Crimes, which is what is displayed in the
# boxplot generated
```

```r
datS = dat[dat[,17]==3,]
nS = nrow(datS) # I chose geographic region South (3=S)
datS
```

```
##      V1                  V2 V3   V4      V5   V6   V7   V8    V9    V10  V11
## 3     3              Harris TX 1729 2818199 31.3  7.1 7553 12449 253526 74.9
## 9     9                Dade FL 1945 1937094 27.1 13.9 6274  8840 244725 65.0
## 10   10              Dallas TX  880 1852810 32.6  8.2 4718  6934 214258 77.1
## 21   21             Broward FL 1209 1255488 25.3 20.7 2456  5543 107386 76.8
## 22   22               Bexar TX 1247 1185394 29.5  9.9 3062  4086 133098 72.7
## 24   24             Tarrant TX  864 1170103 32.2  8.3 1677  3672 132495 79.9
## 34   34          Palm_Beach FL 1974  863518 23.3 24.4 1833  3164  76142 78.8
## 36   36            Pinellas FL  280  851659 22.4 26.0 1620  4458  62344 78.1
## 38   38        Hillsborough FL 1051  834054 29.4 12.2 2012  3068  89895 75.6
## 40   40              Shelby TN  755  826330 29.4 10.4 2489  4918  67032 75.1
## 42   42      Fairfax_County VA  396  818584 29.2  6.5 1694   135  30202 91.4
## 48   48          Montgomery MD  495  757027 28.6 10.2 4635  1507  34754 90.6
## 50   50      Baltimore_City MD   81  736014 30.0 13.7 5444  6203  87355 60.7
## 51   51     Prince_George's MD  486  729268 33.7  6.9 1253  1322  54469 83.2
## 57   57           Baltimore MD  599  692134 27.8 14.0 1269   641  46789 78.4
## 59   59              Orange FL  908  677491 33.2 10.6 1367  2929  52577 78.8
## 60   60               Duval FL  774  672971 30.7 10.7 1538  2623  68586 76.9
```

```
## 66   66             Jefferson KY  385  664937 27.0 13.4 2171  3559  32419 74.1
## 68   68             Jefferson AL 1113  651525 26.7 14.0 2532  4602  55604 73.8
## 70   70                Fulton GA  529  648951 31.6 10.0 3368  5757  93025 77.8
## 73   73  District_of_Columbia DC   61  606900 33.6 12.8 3674  4262  64393 73.1
## 74   74              Oklahoma OK  709  599611 28.4 12.1 1922  3487  57045 79.1
## 76   76               El_Paso TX 1013  591610 29.5  8.1  795  1650  54002 63.7
## 79   79                Travis TX  989  576407 38.0  7.3 1254  1392  60961 83.4
## 84   84               De_Kalb GA  268  545837 32.6  8.5 1036   922  56950 83.9
## 89   89           Mecklenburg NC  527  511433 32.0  9.4 1114  2021  57999 81.6
## 90   90              Davidson TN  502  510784 32.1 11.6 2293  3847  45237 75.9
## 92   92                 Tulsa OK  570  503341 28.2 11.5 1158  2512  39496 81.7
## 95   95               Orleans LA  181  496938 28.3 13.0 2500  4018  54238 68.1
## 107 107             Jefferson LA  306  448306 28.1 10.2 1237  1648  41625 76.0
## 108 108                  Cobb GA  340  447745 31.7  6.3  622   983  27582 85.8
## 114 114          Anne_Arundel MD  416  427239 29.6  8.8  616   617  21826 81.1
## 115 115                  Wake NC  834  423380 34.5  7.8  761  1199  26006 85.4
## 119 119                  Polk FL 1875  405382 23.9 18.6  559  1288  37290 68.0
## 121 121               Brevard FL 1019  398978 26.2 16.6  563  1085  23686 82.3
## 125 125         VA_Beach_City VA  248  393069 35.3  5.9  679   530  23412 88.0
## 128 128               Hidalgo TX 1569  383545 26.4 10.1  311   860  26712 46.6
## 129 129      East_Baton_Rouge LA  456  380105 31.5  9.2  841  1876  41592 80.5
## 130 130                Mobile AL 1233  378643 26.7 11.8  850  1898  30409 70.1
## 132 132               Volusia FL 1106  370712 24.3 22.8  495  1349  25736 75.4
## 142 142              Gwinnett GA  433  352910 32.6  4.7  271   439  17119 86.7
## 143 143               Pulaski AR  771  349660 28.5 11.5 1510  2785  42404 79.0
## 144 144              Guilford NC  650  347420 30.4 11.9  676  1188  28212 76.1
## 149 149                  Knox TN  509  335749 30.0 12.7  984  2178  22422 74.6
## 150 150                   Lee FL  804  335113 21.5 24.7  509  1202  18442 76.9
## 153 153            Greenville SC  792  320167 28.2 11.9  650  1358  20504 71.6
## 160 160            Charleston SC  917  295039 34.1 10.1 1357  1956  28190 75.5
## 163 163                Nueces TX  836  291145 27.3 10.1  584  1406  28606 68.9
## 165 165              Seminole FL  308  287529 27.9 10.3  357   352  17518 84.6
## 166 166              Richland SC  757  285720 34.7  9.5  999  1207  24101 79.4
## 167 167              Hamilton TN  543  285536 26.3 13.5  738  1573  23532 72.5
## 171 171                 Pasco FL  745  281131 18.4 32.3  308   941  12509 66.9
## 173 173              Sarasota FL  572  277776 18.2 32.1  631  1363  19801 81.3
## 176 176            Cumberland NC  653  274566 37.4  6.2  291   586  25247 80.3
## 177 177                Denton TX  889  273525 36.9  5.0  216   458  20372 86.8
## 179 179               Forsyth NC  410  265878 29.2 12.3 1194  1609  21554 77.6
## 183 183                Collin TX  848  264036 29.8  5.3  282   571  17625 88.3
## 186 186              Escambia FL  664  262798 29.2 11.9  522  1584  14380 76.2
## 187 187          Norfolk_City VA   54  261229 41.7 10.5 1101  1471  25194 72.7
## 188 188               Cameron TX  906  260120 25.9 10.6  270   825  18842 50.0
## 192 192                 Hinds MS  869  254441 29.5 11.2 1076  2118  28841 75.2
## 196 196                 Caddo LA  882  248253 25.2 13.3  898  1868  22091 73.4
## 202 202             Jefferson TX  904  239397 25.8 14.0  449  1724  21677 74.4
## 203 203               Madison AL  805  238912 31.4  8.9  399   933   6635 80.2
## 211 211           Spartanburg SC  811  226800 27.0 12.6  375   832  17198 63.0
## 212 212             Fort_Bend TX  875  225421 26.8  4.9  231   301   9433 80.9
## 213 213               Fayette KY  285  225366 34.9  9.9 1248  1851  17378 80.2
## 217 217               Lubbock TX  900  222636 34.1  9.8  655  1562  14509 74.2
## 220 220             Galveston TX  399  217399 26.9 10.5  950  1592  18586 75.8
## 222 222               Chatham GA  440  216935 28.5 12.8  494  1112  18732 73.7
## 223 223 Prince_William_County VA  338  215686 32.3  3.0  196   153   9001 87.8
```

```
## 229 229            Manatee FL  741  211707 21.0 28.1  322   855  16916 75.6
## 230 230         Montgomery AL  790  209085 28.4 11.6  419  1102  17388 75.3
## 232 232            Kanawha WV  903  207619 23.9 15.7  569  1342  10246 72.4
## 237 237             Marion FL 1579  194833 21.6 22.1  235   451  14860 69.6
## 239 239               Leon FL  667  192493 38.5  8.2  413   823  23363 84.9
## 240 240           Brazoria TX 1387  191707 28.7  7.8  156   318   8692 75.5
## 241 241               Bell TX 1059  191088 34.6  8.8  513   572  10865 79.1
## 243 243           Richmond GA  324  189719 31.1 10.0 1032  1787  17918 70.9
## 244 244           McLennan TX 1042  189123 30.0 13.6  301   560  16486 71.6
## 248 248             Howard MD  252  187328 29.8  6.1  695   208   9057 91.1
## 254 254         Montgomery TX 1044  182201 25.5  8.6  125   340   9469 75.5
## 255 255            Harford MD  440  182132 28.8  8.3  247   333   6735 81.6
## 257 257            Clayton GA  143  182052 32.4  5.8  191   346  15419 77.2
## 258 258             Durham NC  291  181835 33.7 10.7 1944  1496  15477 78.9
## 259 259            Alachua FL  874  181596 40.1  9.3 1180  1096  18218 82.7
## 261 261           Muscogee GA  216  179278 30.6 10.8  360  1168  11454 71.5
## 264 264             Gaston NC  357  175093 27.3 12.1  142   368  11865 60.9
## 266 266           Buncombe NC  656  174821 24.8 16.1  469   725   9512 74.5
## 267 267          Cleveland OK  536  174253 33.9  6.7  217   319  12194 83.9
## 272 272    Arlington_County VA   26  170936 37.6 11.3  615   781  12526 87.5
## 273 273   Newport_News_City VA   68  170045 33.9  9.3  354   836  11776 79.3
## 274 274          Calcasieu LA 1071  168134 26.6 10.9  248   845   6399 70.3
## 275 275          Lexington SC  701  167611 27.8  8.9  145   259   9814 77.3
## 276 276           Harrison MS  581  165365 30.0 10.8  313   764   7043 74.7
## 279 279          Lafayette LA  270  164762 31.1  8.3  361  1018  10599 73.3
## 293 293               Lake FL  953  152104 19.0 27.5  167   664   7099 70.6
## 294 294            Collier FL 2026  152099 22.4 22.8  282   431   9426 79.0
## 296 296      Chesapeake_City VA  341  151976 28.6  8.4  212   210   8427 77.1
## 297 297              Smith TX  929  151309 26.2 13.7  349   795  11712 75.7
## 298 298         Tuscaloosa AL 1325  150522 33.3 11.4  299   731  12377 69.6
## 299 299          Frederick MD  663  150208 28.7  9.4  172   241   4939 80.4
## 301 301          St._Lucie FL  573  150171 22.9 21.0  176   425   9842 71.7
## 302 302               Bibb GA  250  149967 27.5 12.9  438  1010  12701 68.2
## 303 303             Onslow NC  767  149838 49.7  4.4  104   133   7505 83.0
## 315 315           Anderson SC  718  145196 25.5 13.6  199   456   7525 64.0
## 316 316        St._Tammany LA  854  144508 24.2  8.9  282   512   4447 76.9
## 317 317              Horry SC 1134  144053 28.2 12.7  175   505  12459 74.3
## 318 318           Okaloosa FL  936  143776 30.8  9.3  178   482   5153 83.8
## 319 319           Sullivan TN  413  143596 24.6 14.3  377   982   6236 66.8
## 322 322           Ouachita LA  611  142191 28.3 11.2  268  1043  10605 71.6
## 323 323             Kenton KY  163  142031 28.3 11.5  263   733   6925 74.4
## 328 328         Williamson TX 1124  139551 29.1  7.6   88   185   5724 81.4
## 335 335       Hampton_City VA   52  133793 33.0  9.6  163   251   8376 79.7
## 337 337               Webb TX 3357  133239 28.5  7.9  107   382  12202 47.8
## 340 340            Rapides LA 1323  131556 26.8 12.0  246   768   6101 69.0
## 341 341               York SC  683  131497 28.7 10.6  121   276   9525 67.5
## 352 352                Bay FL  764  126994 27.7 12.0  178   478   8634 74.7
## 353 353           Davidson NC  552  126677 26.9 12.0   78   221   5662 64.2
## 359 359            Carroll MD  449  123372 26.6 10.2  142   123   3430 78.5
## 361 361            Wichita TX  628  122378 29.5 12.8  243   457  10727 75.1
## 363 363             Brazos TX  586  121862 49.4  6.7  170   279   8203 79.8
## 367 367              Aiken SC 1073  120940 26.7 11.4  128   191   6835 70.7
## 369 369        New_Hanover NC  199  120284 29.0 12.5  297   554  11892 78.1
## 372 372             Taylor TX  916  119655 30.7 12.0  204   467   6785 75.4
```

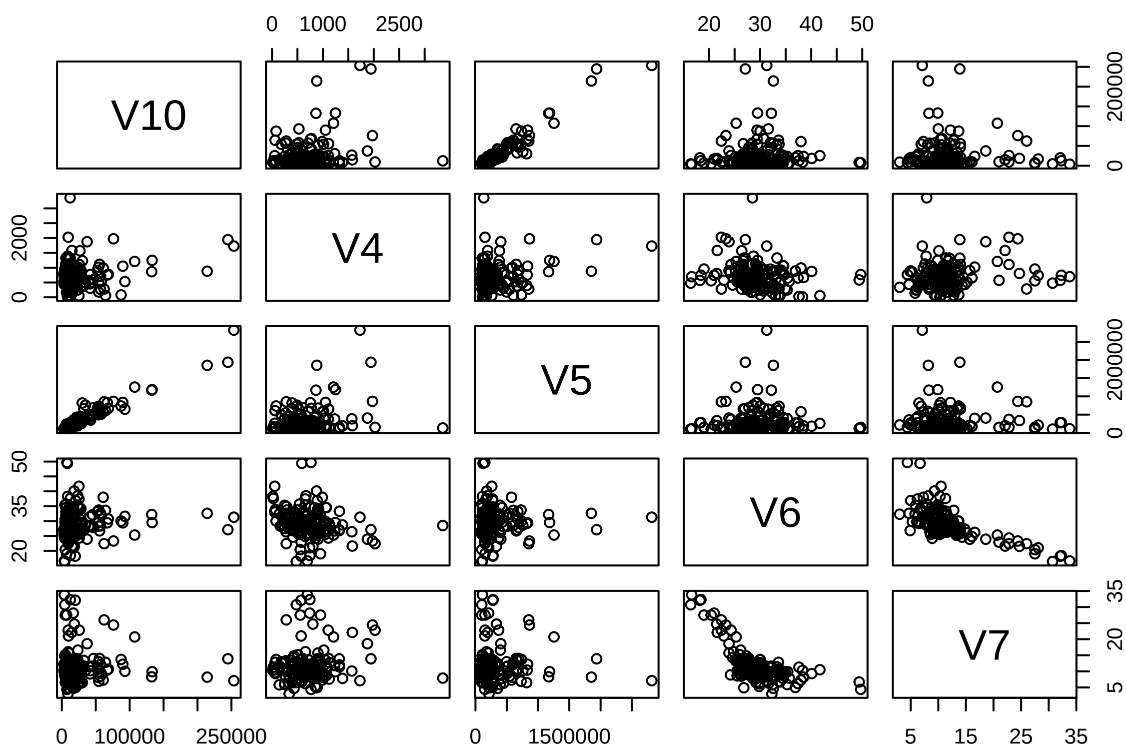
```
## 374 374              Ector TX  901  118934 27.1  9.3  153   389  14643 66.9
## 377 377         Rutherford TN  619  118570 33.1  8.4  133   215   6072 73.9
## 378 378            Catawba NC  400  118412 27.2 11.9  179   464   6830 66.7
## 381 381            Calhoun AL  609  116034 28.8 12.4  133   486   4901 67.4
## 385 385            Jackson MS  727  115243 25.9  9.4  170   346   4777 74.4
## 386 386           Florence SC  799  114344 26.2 11.2  211   731   8421 64.3
## 389 389         Washington AR  950  113409 32.0 11.2  208   651   6122 73.2
## 395 395           Comanche OK 1069  111486 34.5  8.7  127   347   5979 81.1
## 396 396    Alexandria_City VA   15  111183 38.3 10.3  652   662   8537 86.9
## 398 398          Charlotte FL  694  110975 16.6 33.8  183   632   3741 75.7
## 401 401              Rowan NC  511  110605 26.0 15.2  114   244   5233 66.0
## 408 408           Alamance NC  431  108213 27.3 14.8  132   340   4152 67.9
## 409 409               Pitt NC  652  107924 35.4  9.9  496   583   4603 71.0
## 411 411            Osceola FL 1322  107728 27.1 13.9   98   291   9665 73.7
## 416 416            Midland TX  900  106611 26.8  9.0  139   333   7546 76.8
## 417 417           Randolph NC  788  106546 27.1 12.2   69   145   2940 62.0
## 420 420               Clay FL  601  105986 26.3  8.5  164   277   4560 81.2
## 422 422            Robeson NC  949  105179 26.7 10.7   83   281   4318 57.0
## 423 423              Gregg TX  274  104948 26.4 13.3  166   420   9181 75.8
## 424 424              Wayne NC  553  104666 29.7 10.2  113   263   4682 71.2
## 428 428             Sumner TN  529  103281 25.5 10.2   96   259   3285 70.6
## 430 430             Sumter SC  666  102637 31.6  9.4   88   214   7138 69.8
## 435 435            Charles MD  461  101154 29.9  6.5   67   104   5279 81.0
## 436 436           Hernando FL  478  101115 16.4 30.7   98   290   4414 70.5
## 437 437             Martin FL  556  100900 20.4 27.5  193   277   5081 79.7
## 438 438         Montgomery TN  539  100498 35.7  7.9   87   188   6537 77.9
## 440 440             Morgan AL  582  100043 26.3 11.7  122   464   4693 69.4
##      V12  V13  V14   V15   V16 V17
## 3   25.4 12.5  5.7 19517 55003   3
## 9   18.8 14.2  8.7 17823 34525   3
## 10  26.3 10.4  6.1 21001 38911   3
## 21  18.8  7.1  7.4 22355 28066   3
## 22  19.7 16.2  6.7 15508 18383   3
## 24  24.0  8.2  6.6 18825 22027   3
## 34  22.1  6.2  8.4 26798 23141   3
## 36  18.5  6.2  6.2 21610 18404   3
## 38  20.2  9.5  6.0 16876 14075   3
## 40  20.8 14.7  5.4 18430 15229   3
## 42  49.0  2.2  3.2 28999 23738   3
## 48  49.9  2.7  3.3 30081 22772   3
## 50  15.5 17.8  9.4 17263 12706   3
## 51  25.5  3.7  5.0 19568 14270   3
## 57  25.0  3.8  5.7 23470 16244   3
## 59  21.2  7.8  6.6 17879 12113   3
## 60  18.4  9.8  6.6 17662 11886   3
## 66  19.3 10.9  6.3 19140 12727   3
## 68  19.9 12.7  5.3 18624 12134   3
## 70  31.6 15.4  5.3 22819 14808   3
## 73  33.3 13.3  7.7 23603 14325   3
## 74  22.6 11.3  6.0 17741 10638   3
## 76  15.2 22.4 10.8 11545  6830   3
## 79  34.7 10.2  4.8 18340 10571   3
## 84  32.7  7.2  4.4 21005 11465   3
## 89  28.3  7.2  4.8 20942 10710   3
```

```
## 90   24.4 10.0  4.6 19505  9963   3
## 92   23.7 10.0  6.1 19295  9712   3
## 95   22.4 27.3  6.1 16578  8238   3
## 107 18.8 11.4  5.6 17101  7666   3
## 108 33.0  3.8  4.1 21933  9820   3
## 114 24.6  3.0  5.0 22797  9740   3
## 115 35.3  5.5  3.5 20658  8746   3
## 119 12.9  9.4 10.1 14835  6014   3
## 121 20.4  6.3  7.1 17668  7049   3
## 125 25.5  4.3  5.4 17697  6956   3
## 128 11.5 36.3 17.6  8899  3413   3
## 129 27.5 15.1  5.4 17881  6797   3
## 130 15.5 17.5  7.2 14389  5448   3
## 132 14.8  7.9  6.9 15648  5801   3
## 142 29.6  2.9  4.0 19861  7009   3
## 143 23.5 10.5  5.8 18225  6373   3
## 144 24.8  7.3  5.4 20349  7070   3
## 149 23.9 10.2  4.6 17382  5836   3
## 150 16.4  6.1  6.4 18877  6326   3
## 153 21.0  7.8  4.9 17874  5723   3
## 160 22.4 13.8  4.9 16015  4725   3
## 163 17.0 17.2  7.7 15124  4403   3
## 165 26.3  5.3  6.1 17885  5142   3
## 166 28.0 10.1  4.6 17137  4896   3
## 167 19.7 10.2  5.9 18242  5209   3
## 171  9.1  7.9  8.3 13944  3920   3
## 173 21.9  4.6  5.1 24948  6930   3
## 176 16.6 12.1  6.4 12923  3548   3
## 177 32.3  4.5  5.6 17801  4869   3
## 179 24.1  7.8  4.4 20645  5489   3
## 183 39.1  3.9  5.9 22303  5889   3
## 186 18.2 13.3  5.8 15392  4045   3
## 187 16.8 15.1  6.4 16412  4287   3
## 188 12.0 33.7 12.5  9728  2530   3
## 192 26.4 16.7  6.3 16215  4126   3
## 196 18.2 19.1  7.1 16337  4056   3
## 202 15.5 15.5  6.7 17418  4170   3
## 203 30.1  8.4  5.2 18990  4537   3
## 211 14.3  8.8  5.5 15776  3578   3
## 212 30.2  6.9  3.8 18301  4125   3
## 213 30.6 10.2  3.8 19320  4354   3
## 217 23.4 13.2  5.6 15443  3438   3
## 220 19.3 12.5  7.0 17744  3858   3
## 222 18.6 13.6  4.3 17776  3856   3
## 223 27.6  2.3  4.3 20543  4431   3
## 229 15.5  6.8  5.8 17997  3810   3
## 230 24.2 14.1  5.9 17469  3653   3
## 232 17.6 12.3  7.5 17192  3569   3
## 237 11.5 10.8  8.6 13802  2689   3
## 239 37.1  9.4  3.9 16422  3161   3
## 240 15.1  7.7  5.6 17951  3441   3
## 241 17.2 12.2  6.7 13536  2587   3
## 243 17.3 14.8  4.9 15941  3024   3
## 244 16.6 13.9  6.4 14925  2823   3
```

```
## 248 46.9  2.2  4.1 27546  5160   3
## 254 19.4  9.4  5.5 17084  3113   3
## 255 21.5  4.1  6.7 20941  3814   3
## 257 14.7  7.3  5.8 16171  2944   3
## 258 33.4  8.7  3.6 19238  3498   3
## 259 34.6 14.4  4.2 16058  2916   3
## 261 16.6 14.9  4.6 15505  2780   3
## 264 10.8  8.2  6.2 16319  2857   3
## 266 19.1  8.2  4.9 16934  2960   3
## 267 25.9  7.0  4.8 14443  2517   3
## 272 52.3  4.3  3.6 30242  5169   3
## 273 18.4 12.2  6.5 15327  2606   3
## 274 14.7 15.5  7.8 14968  2517   3
## 275 21.0  6.3  4.1 18126  3038   3
## 276 16.3 15.4  6.7 13691  2264   3
## 279 22.5 16.2  5.0 16868  2779   3
## 293 12.7  7.9  8.9 17496  2661   3
## 294 22.3  6.4  7.5 25589  3892   3
## 296 16.9  7.0  6.1 16924  2572   3
## 297 19.8 12.6  6.6 17511  2650   3
## 298 20.0 13.5  5.4 15113  2275   3
## 299 22.0  3.5  6.0 19954  2997   3
## 301 13.1  8.5 13.8 14137  2123   3
## 302 17.0 15.9  4.1 17548  2632   3
## 303 13.4  9.8  5.6 10190  1527   3
## 315 12.9  8.6  6.6 14205  2063   3
## 316 23.1 11.1  6.2 17129  2475   3
## 317 16.0 11.6  7.7 14693  2117   3
## 318 21.0  7.8  6.7 15803  2272   3
## 319 15.6 10.4  4.4 15747  2261   3
## 322 18.9 19.6  6.4 13869  1972   3
## 323 17.0  8.2  5.4 16935  2405   3
## 328 24.6  7.6  3.8 14934  2084   3
## 335 19.1  8.8  6.8 14743  1973   3
## 337 11.1 33.1  9.8  8973  1196   3
## 340 14.6 18.6  7.1 14615  1923   3
## 341 16.9  7.6  5.5 16713  2198   3
## 352 15.7 11.2  8.0 14814  1881   3
## 353 10.0  7.3  5.9 15079  1910   3
## 359 19.6  2.5  5.7 22002  2714   3
## 361 16.5 12.5  6.8 17119  2095   3
## 363 35.8 14.9  3.6 12641  1540   3
## 367 17.2 11.0  4.9 17898  2165   3
## 369 21.2  9.9  5.9 17119  2059   3
## 372 20.7 11.2  5.9 16021  1917   3
## 374 11.4 16.6  6.6 14766  1756   3
## 377 18.7  7.6  5.4 15501  1838   3
## 378 14.2  4.8  6.4 17396  2060   3
## 381 14.2 11.7  7.3 13776  1598   3
## 385 14.4 14.0  7.1 13475  1553   3
## 386 14.8 15.9  5.9 14961  1711   3
## 389 20.0  9.8  3.7 14736  1671   3
## 395 18.4 13.3  6.6 13228  1475   3
## 396 48.5  4.7  4.6 31699  3524   3
```

```
## 398 13.4  5.2  7.2 16362  1816   3
## 401 11.7  6.8  5.4 15691  1736   3
## 408 14.6  6.0  4.5 17306  1873   3
## 409 21.9 15.5  5.5 15852  1711   3
## 411 11.2  6.9  6.7 16451  1772   3
## 416 26.4 11.5  5.4 19345  2062   3
## 417  9.1  6.5  4.8 14721  1568   3
## 420 17.9  5.4  5.9 16029  1699   3
## 422 11.0 20.7  8.9 10849  1141   3
## 423 17.7 13.4  8.2 16775  1761   3
## 424 12.7 11.9  7.3 13350  1397   3
## 428 14.4  7.0  6.8 16514  1706   3
## 430 15.0 16.9  9.4 11803  1211   3
## 435 16.2  3.7  4.9 19317  1954   3
## 436  9.7  7.9  8.2 13919  1407   3
## 437 20.3  5.0  9.8 27125  2737   3
## 438 16.5 10.8  8.0 13169  1323   3
## 440 15.5  9.4  7.1 16458  1647   3
```

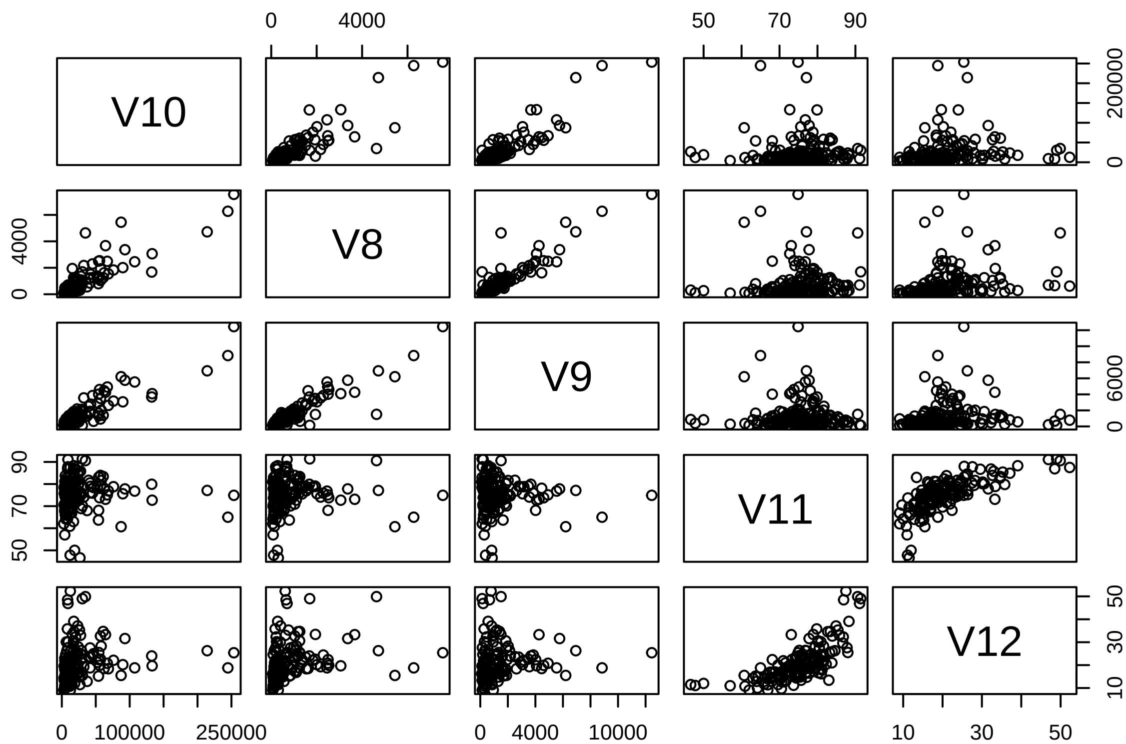
```
# Since the predictors, 1. ID, 2. County, 3. State are categorical variables and have no numeric value,
# I decided to disregard them in my choice for predictors.
pairs(datS[, c(10, 4:7)])
```




```
# Out of 4. Land Area, 5. Total Population,
# 6. % Aged 18-34, 7. % aged >= 65;
```

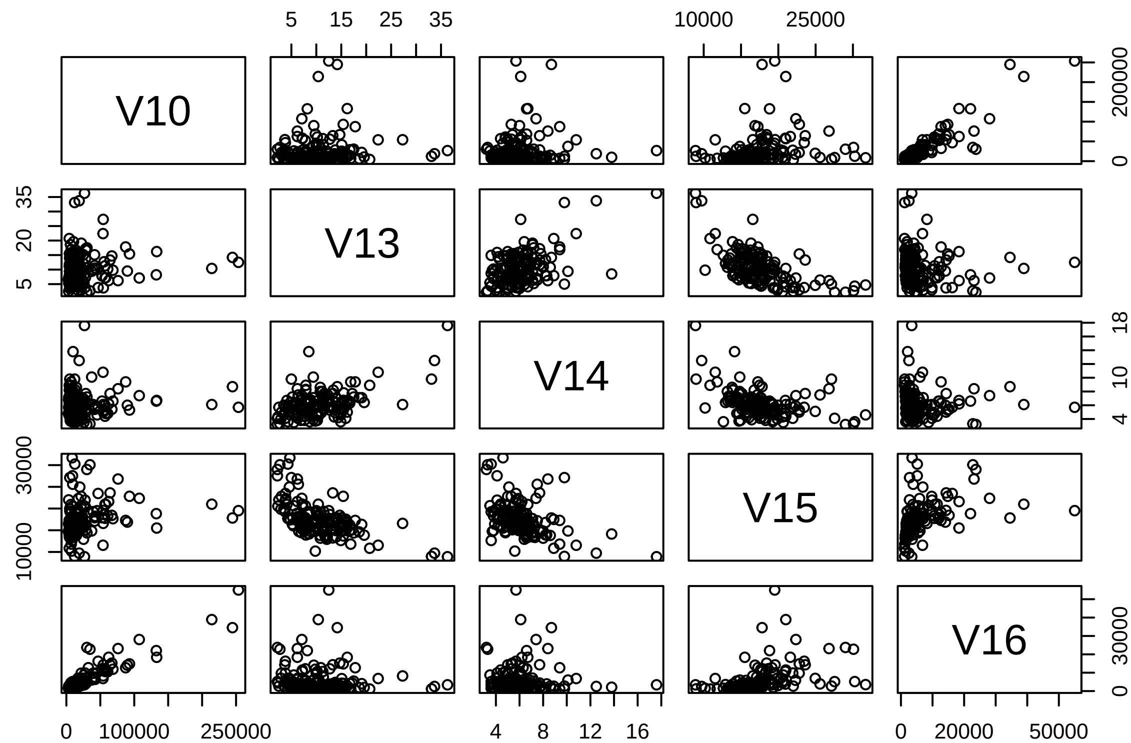
```
# 5. Total Population is the only one that appears to have a positive correlation between
# the dependent variable of 10. Total # of Serious Crimes.
```

```
pairs(datS[, c(10, 8, 9, 11, 12)])
```




```
# Y appears to have positive trend with
# 8. number of active physicians
# 9. number of hospital beds
```

```
pairs(datS[, c(10, 13:16)])
```




```
# Y appears to have positive trend with
# 16. total personal income
```

```
cor(datS[,c(10,5,8,9,16)])
```

```
##           V10       V5        V8        V9        V16
## V10 1.0000000 0.9639008 0.8749421 0.9070338 0.9194097
## V5  0.9639008 1.0000000 0.8767087 0.8917438 0.9786305
## V8  0.8749421 0.8767087 1.0000000 0.9200551 0.8712650
## V9  0.9070338 0.8917438 0.9200551 1.0000000 0.8504703
## V16 0.9194097 0.9786305 0.8712650 0.8504703 1.0000000
```

```
# 5,8,9,16 are highly correlated. let's keep one of them
# Keep 9. # of Hospital Beds
```

```
cor(datS[,-c(1,2,3,5,8,16,17)])
```

```
##             V4          V6          V7          V9         V10         V11
## V4   1.0000000 -0.21150444  0.13162675  0.17008359  0.24968855 -0.33846586
## V6  -0.2115044  1.00000000 -0.70517335  0.01676288  0.07117655  0.35791926
## V7   0.1316268 -0.70517335  1.00000000  0.05909970 -0.03164377 -0.21603557
## V9   0.1700836  0.01676288  0.05909970  1.00000000  0.90703380 -0.05194019
## V10  0.2496885  0.07117655 -0.03164377  0.90703380  1.00000000  0.01086022
## V11 -0.3384659  0.35791926 -0.21603557 -0.05194019  0.01086022  1.00000000
```

```
## V12 -0.2576094  0.45937905 -0.34918479  0.10510286  0.14548531  0.71687822
## V13  0.3250891  0.03855845 -0.10293079  0.18158967  0.09246762 -0.67627212
## V14  0.3762087 -0.37618742  0.32023282  0.04469196  0.06599828 -0.58335458
## V15 -0.2587832 -0.03717299  0.09427310  0.19423004  0.19946126  0.60912447
##            V12         V13         V14         V15
## V4  -0.2576094  0.32508906  0.37620870 -0.25878324
## V6   0.4593791  0.03855845 -0.37618742 -0.03717299
## V7  -0.3491848 -0.10293079  0.32023282  0.09427310
## V9   0.1051029  0.18158967  0.04469196  0.19423004
## V10  0.1454853  0.09246762  0.06599828  0.19946126
## V11  0.7168782 -0.67627212 -0.58335458  0.60912447
## V12  1.0000000 -0.33655343 -0.53208080  0.70531709
## V13 -0.3365534  1.00000000  0.51867164 -0.59223929
## V14 -0.5320808  0.51867164  1.00000000 -0.41549730
## V15  0.7053171 -0.59223929 -0.41549730  1.00000000
```

```
# Other variables that have relatively high correlation with Y:
# 4. land area (r = 0.249)
# 6. percent of population aged 18-34 (r = 0.071)
# 15. per capita income (r = 0.199)
```

```
Y = datS[,10] # This is the response variable in the Southern Region
X4 = datS[,4]
X6 = datS[,6]
X7 = datS[,7]
X9 = datS[,9]
X11 = datS[,11]
X12 = datS[,12]
X13 = datS[,13]
X14 = datS[,14]
X15 = datS[,15]
# All of these are the rest of the variables that I am taking into account
# when choosing my second predictor, b2.
```

```
anova(lm(Y ~ X9 + X4)) # p-value = 0.004504 ** This means 0.001 < p <= 0.01
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value    Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 730.0391 < 2.2e-16 ***
## X4          1 2.0904e+09 2.0904e+09   8.3195  0.004504 **
## Residuals 149 3.7438e+10 2.5126e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(Y ~ X9 + X6))
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value Pr(>F)
```

```
## X9          1 1.8343e+11 1.8343e+11 703.8740 <2e-16 ***
## X6          1 6.9869e+08 6.9869e+08   2.6811 0.1037
## Residuals 149 3.8829e+10 2.6060e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(Y ~ X9 + X7))
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 721.0950 <2e-16 ***
## X7          1 1.6260e+09 1.6260e+09   6.3921 0.0125 *
## Residuals 149 3.7902e+10 2.5438e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(Y ~ X9 + X11))
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq F value  Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 704.829 < 2e-16 ***
## X11         1 7.5132e+08 7.5132e+08   2.887 0.09139 .
## Residuals 149 3.8777e+10 2.6025e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(Y ~ X9 + X12))
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 701.4963 <2e-16 ***
## X12         1 5.6708e+08 5.6708e+08   2.1687  0.143
## Residuals 149 3.8961e+10 2.6148e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(Y ~ X9 + X13))
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value  Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 713.1401 < 2e-16 ***
## X13         1 1.2032e+09 1.2032e+09   4.6779 0.03215 *
## Residuals 149 3.8325e+10 2.5721e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lm(Y ~ X9 + X14))
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 693.9750 <2e-16 ***
## X14         1 1.4483e+08 1.4483e+08   0.5479 0.4603
## Residuals 149 3.9383e+10 2.6432e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lm(Y ~ X9 + X15))
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 693.6374 <2e-16 ***
## X15         1 1.2566e+08 1.2566e+08   0.4752 0.4917
## Residuals 149 3.9402e+10 2.6445e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
xy.lm = lm(Y ~ X9 + X4) # 9. # of hospital beds and 4. land area are the two predictors
# that I chose due to their correlation
# with the response variable of total 10. # of serious crimes
summary(xy.lm)
```

```
##
## Call:
## lm(formula = Y ~ X9 + X4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -39730  -7519   -416   4612  74208
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6245.3530  2556.5420  -2.443   0.0157 *
## X9             20.0222     0.7661  26.135   <2e-16 ***
## X4              8.4789     2.9396   2.884   0.0045 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15850 on 149 degrees of freedom
## Multiple R-squared:  0.8321, Adjusted R-squared:  0.8298
## F-statistic: 369.2 on 2 and 149 DF,  p-value: < 2.2e-16
```

```r
anova(xy.lm)
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df     Sum Sq    Mean Sq  F value    Pr(>F)
## X9          1 1.8343e+11 1.8343e+11 730.0391 < 2.2e-16 ***
## X4          1 2.0904e+09 2.0904e+09   8.3195  0.004504 **
## Residuals 149 3.7438e+10 2.5126e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

My multiple regression model is the following: Yhat = -6245.3530 + 20.0222X9 + 8.4789X4 + Ei

```r
p = 3 # amount of parameters
b = xy.lm$coefficients
anv = anova(xy.lm)
MSE = anv$`Mean Sq`[p]

X = model.matrix(xy.lm)
XtX = t(X)%*%X

VarCov.b = MSE * solve(XtX)
SE.b = sqrt(diag(VarCov.b))

out = summary(xy.lm)
coef = out$coefficients
SE.b = coef[,2]

alpha = 0.10
lower_int = b - qt(1-alpha/2, df = nS - p) * SE.b
upper_int = b + qt(1-alpha/2, df = nS - p) * SE.b
lower_int
```

```
##   (Intercept)            X9            X4
## -10476.799587     18.754191      3.613403
```

```r
# We are 90% confident that the the true average of total # hospital beds lies within the
# interval of 18.746099 and 21.29832 beds.
# This interval does not include zero,
# so it continues to support the correlation
# with the # of serious crimes
upper_int
```

```
## (Intercept)          X9          X4
## -2013.90651    21.29023    13.34432
```

```r
# We are 90% confident that the the true average of land area lies within the
# interval of 3.582354 and 13.37537 units.
# This interval does not include zero,
# so it continues to support the correlation
# with the # of serious crimes.
```

```r
# H0: B1 = 0
# Ha: B1 != 0
pval.b1 = 2 * pt(q = abs(b[2]/SE.b[2]), df = nS - p, lower.tail = FALSE)
pval.b1
```

```
##           X9
## 1.616588e-57
```

```r
# Since the p-value for B1 is less than significance level of 0.01, we reject the null,
# and conclude that B1, total # of hospital beds has a relationship
# with the dependent variable of # of serious crimes.
pval.b2 = 2 * pt(q = abs(b[3]/SE.b[3]), df = nS - p, lower.tail = FALSE)
pval.b2
```

```
##          X4
## 0.004503946
```

```r
# Since the p-value for B2 is less than significance level of 0.01, we reject the null,
# and conclude that B2, land area, has a relationship
# with the dependent variable of # of serious crimes.
```

```r
# H0: B1 = B2 = 0
# Ha: At least one of B1 or B2 does not equal 0
SSR = sum(anv$`Sum Sq`[1:2])
SSE = anv$`Sum Sq`[3]
F.stat = (SSR/(p-1)) / (SSE/(nS-p))
F.stat # 369.1793
```

```
## [1] 369.1793
```

```r
pval = pf(F.stat, df1 = p-1, df2 = nS-p, lower.tail = FALSE)
pval # 1.858x10^-58
```

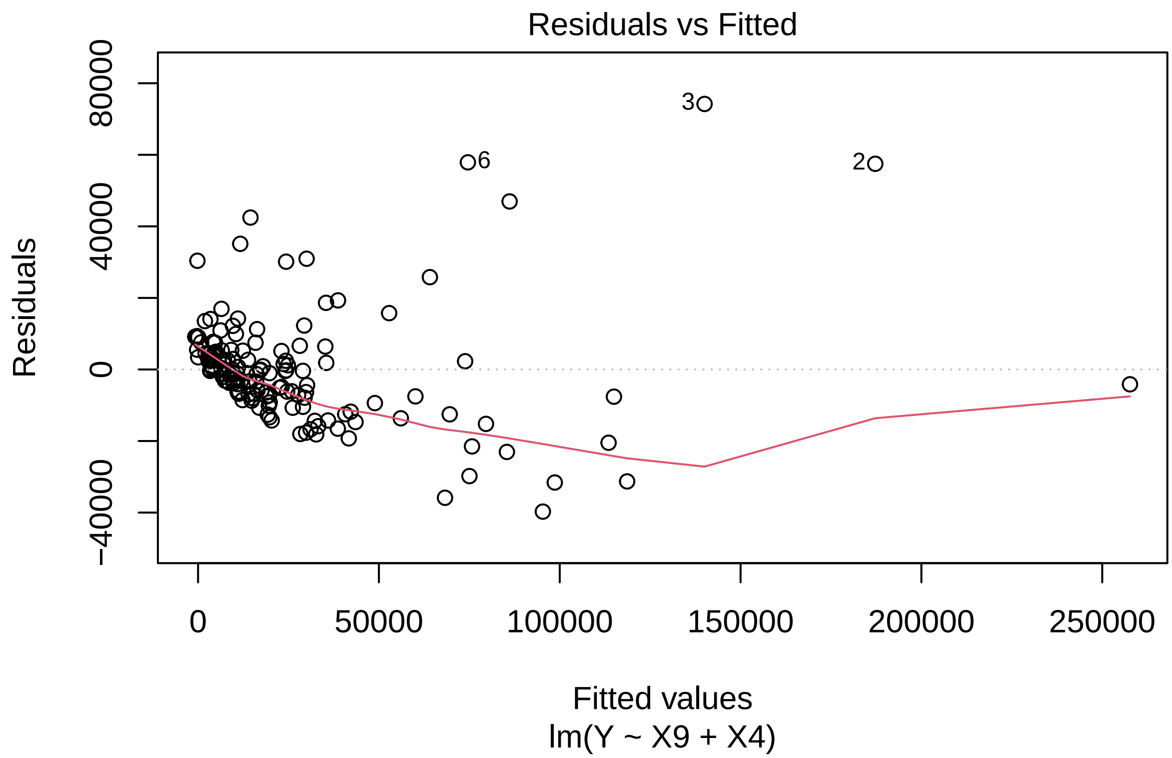
```
## [1] 1.857753e-58
```

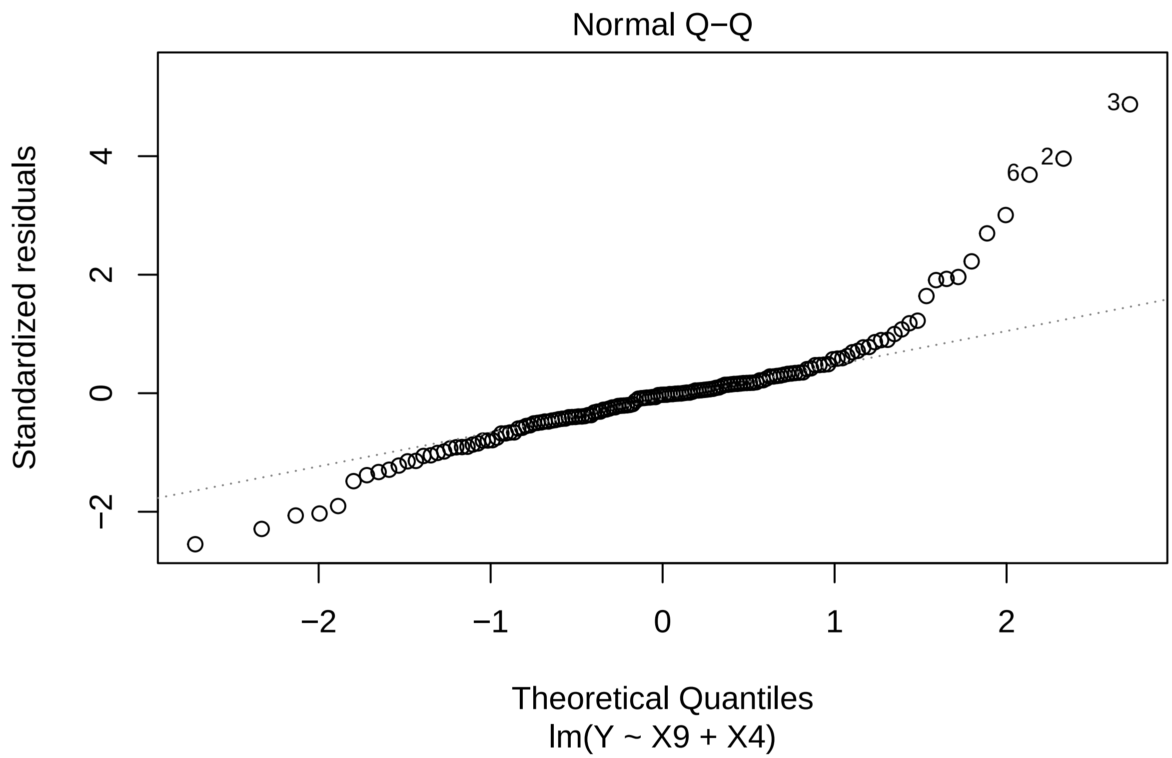
```r
adj.R2 = 1 - (SSE/(nS-p)) / ((SSR+SSE)/(nS-1))
adj.R2 # 0.8298 this is a positive correlation
```

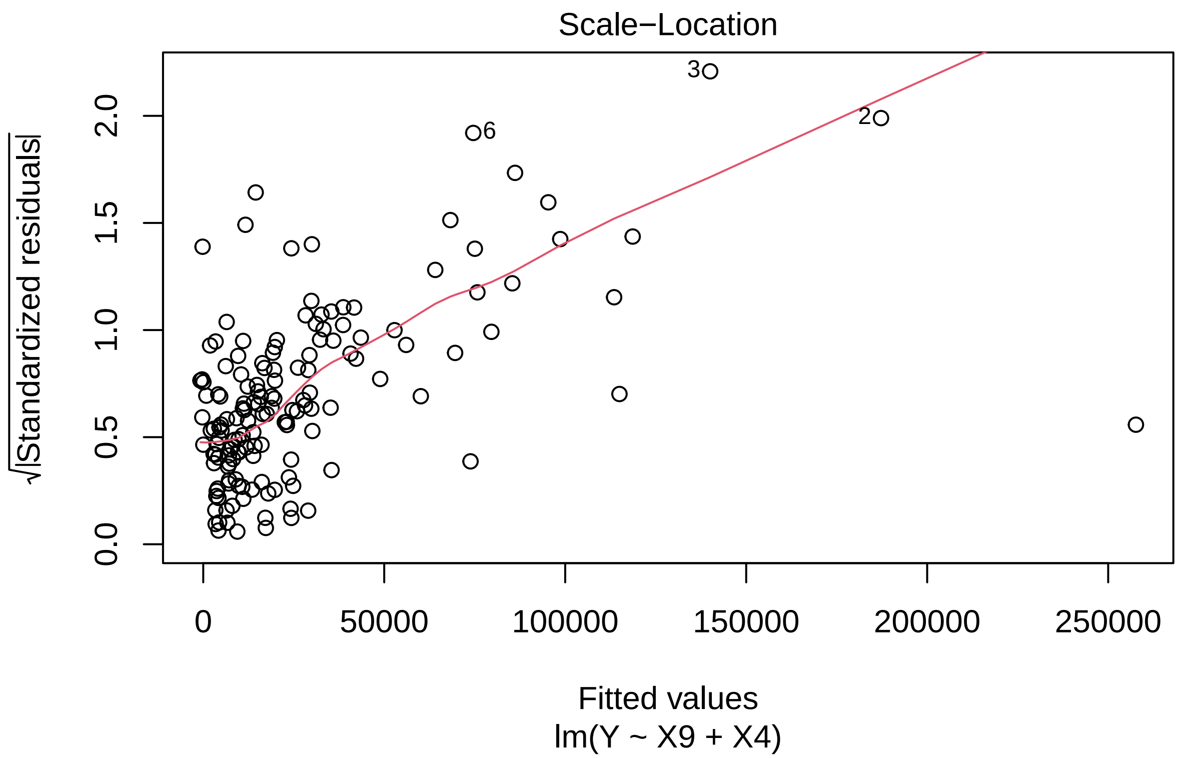
```
## [1] 0.8298321
```

```r
# Since the p-value of 1.858x10^-58 is smaller than the significance level of 0.05, we reject H0,
# and conclude that at least one of B1  or B2 has a linear relationship
# with the dependent variable of Total # of Serious Crimes
```

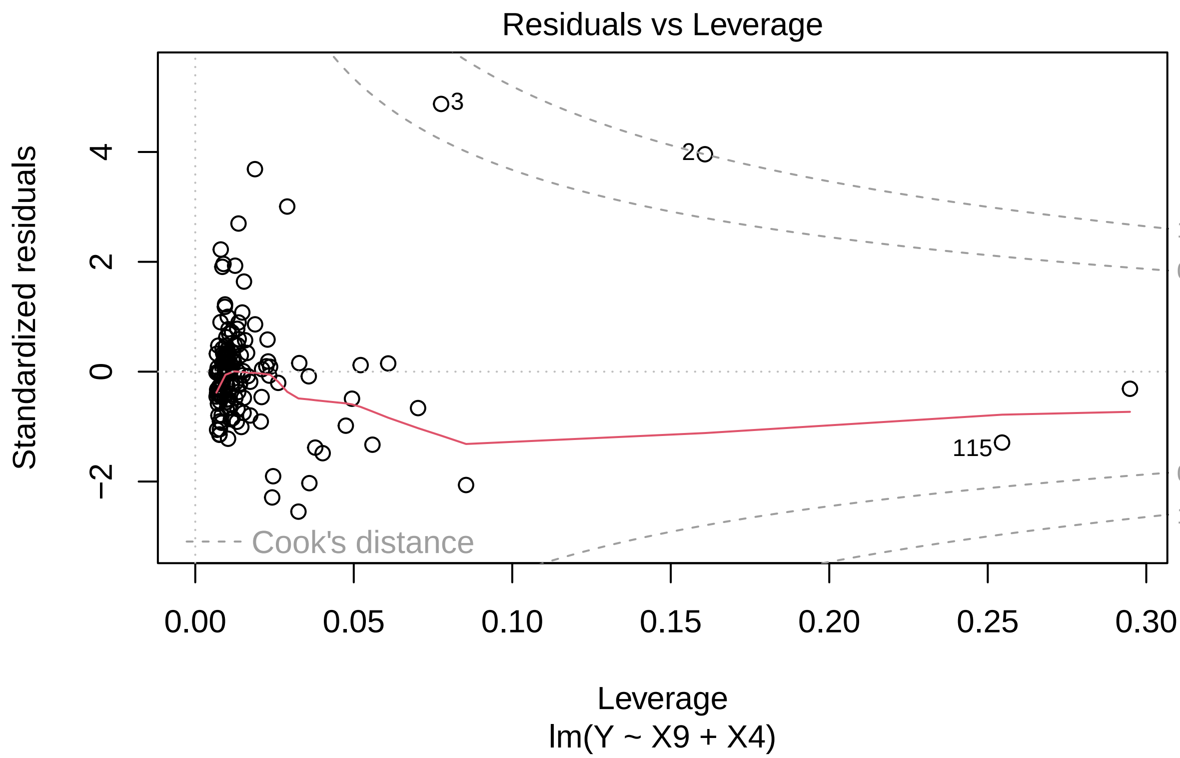
```r
res =xy.lm$residuals
plot(xy.lm)
```


Residuals vs Fitted
Residuals
80000
40000
0
−40000
0
50000
100000
150000
200000
250000
Fitted values
lm(Y ~ X9 + X4)
3
6
2


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Y ~ X9 + X4)
3
2
6
−2
0
2
4
−2
−1
0
1
2



Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
0
50000
100000
150000
200000
250000
Fitted values
lm(Y ~ X9 + X4)
3
2
6




```
# Plot 1 of Residual vs Fitted is good because there is no apparent pattern,
# so, a linear relationship seems like a good model so far, as it checks this criteria.
# Plot 2 of the Normal Q-Q Plot is a bit off the straight line with a couple of outliers
# and high leverage points, but I personally think it is still safe to call this model
# approximately normal and not remove those points,
# as it could make the model less reliable.
# Plot 3 of Fitted Values and Standarized residuals is not really good
# because the residuals are spread evenly,
# so does not really check the box for equal variance.
# Plot 4 of the Cook's Plot makes the model still okay, because all the points are within the line.
```

```
number_models = 2**(p-1)
number_models
```

```
## [1] 4
```

```
# The 4 possible models that I can fit are:
# Yi = B0 + Ei
# Yi = B0 + B1X1 + Ei
# Yi = B0 + B1X1 + B2Xi + Ei
# Yi = B0 + B2Xi + Ei
# However for this project, I am not required to fit all of these models,
# but it is good to see all possibilities, if I were to optimize this project.
```

```r
# My conclusion for this project is that it is important to look at different models
# and apply different tests, in order to provide the best model
# for finding important statistical information
# such as the correlation between multiple predictors and a dependent variable.
```