Some Internet Data

Contents

Introduction

Through this project, we seek to understand the relationship between internet access and inequality on a global scale before examining the factors in detail using American Census Data. We take this dual-pronged approach so that we can get a picture of Internet Inequality globally while using reliable US Census data to do the quantitaive analysis. We will use data from the United Nations to compare selected socio-economic indexes to international communication measurements.

Questions

> What economic indicators (race, occupation, community poverty rate) are most strongly correlated with internet access rates? Can we build a model that accurately predicts said rates?
>
> Are internet access rates a stronger predictor of poverty rates than other forms of social investment (ie roads, schools, hospitals)?
>
> Do these effects extend across internet technologies (cell phones and broadband internet)? If not, which type of infrastructure investment is better.

Motivation

We are interested in this problem as data scientists because our field is a mixed bag. On one hand, big data can be used to influence elections, spread hateful propaganda, and be used to track every purchase and decision we make. These political consequences are well known. However, the Internet has a history of advancing economies, and those without the internet tend to be left behind. To speak about this in particular, we need to investigate the ways in which internet access influences occupational outlook while controlling for other confounding factors like geography, race, and infrastructure investment more generally.

Literature Review

Data

American Community Survey

Annual Survey of State Finances

World Bank Data

IEE MAC Address Blocks

List of Internet Exchange Points

Methodology

First we will examine the problem on a global scale using chloropleth maps that will inform our future choices.

We will build several models for predicting poverty rate, using both the generalized logistic model and the generalized linear model. In this way, we'll see how things like internet access and infrastructure investment influence poverty rates. The American Community Survey includes internet access rates, poverty, race, industry, language, occupation, place of birth, and familial origin. Using this data alone, we should be able to see if race or occupation is a better indicator of aggregate povery than internet access rates.

Hypothesis

Pew Research says that 20% of teens are unable to finish their homework due to the digital divide. The end result of this is likely low-skill careers and lower incomes. In fact, the internet tends to raise the tide for all, as a breadth study (also by Pew) showed that per capita income and access rates are highly correlated. We'd like to investigate the relationship between technology and the economy and see if we can build models resilient to the particle type of device. Previous work has used infrastructure invesment to build logistic models for poverty using satellite images of infrastructure. It is also well known that poverty and broadband access rates are highly correlated. However, it is unknown if there is an underlying causal factor or if internet can, *by itself*, lift people out of poverty. The McKinsey Global Institute did a massive study on the economic potential of internet investment in China that will inform our approach in this matter. Finally, the Internet Society, a global organization that builds internet infrastructure (mostly in the developing world), has compiled a list of internet penetration rates and other such metrics by country across the world. However, due to data collection limitations and the quality of data sources across continents, it would be impossible to investigate these things wtih respect to more generic features like race and infrastructure. Since the United States has a non-uniform income distribution across states, this should allow us to draw from a breadth of circumstances. Due to the multiplicative of effects in education, business opportunities, and spending opportunities available on the Internet, we suspect that governmental investment in digital infrastructure will have at least as much affect as road or school spending. Additionally, we suspect that this multiplier is reduced for cellular infrastructure relative to fixed (broadband) infrastructure because of the productivity gains associated with PCs over smartphones. This research will reveal to governments (both local and national) what kinds of infrastructure investment yields the most economic gains in the digital age. To our knowledge, this particular question has not been answered.

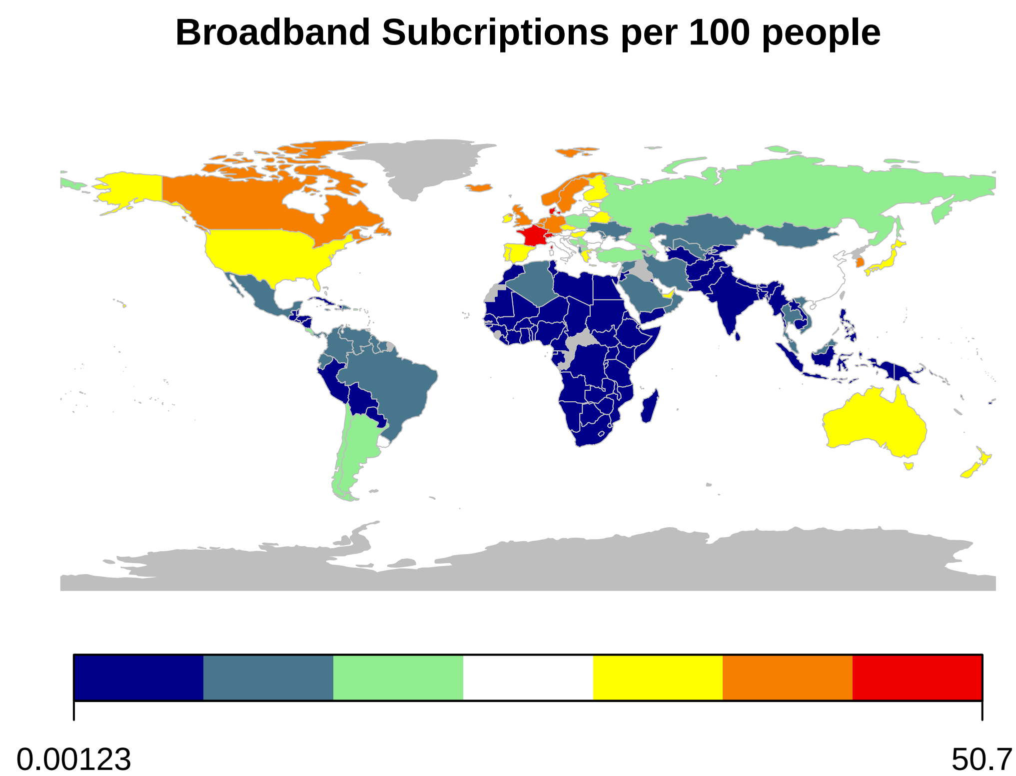
Executive Summary

TODO

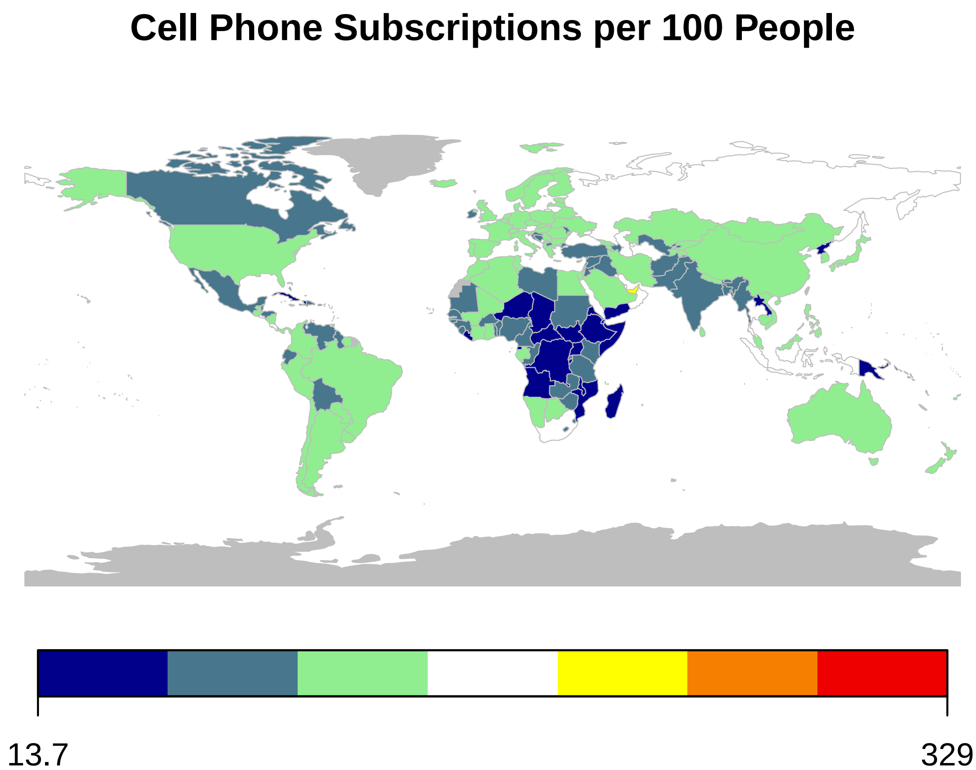
```
##  [1] "tidycensus"   "rworldmap"    "sp"           "xlsx"
##  [5] "forcats"      "stringr"      "dplyr"        "purrr"
##  [9] "readr"        "tibble"       "tidyverse"    "tidyr"
## [13] "reshape2"     "readxl"       "randomForest" "nnet"
## [17] "ModelMetrics" "knitr"        "imputeTS"     "forecast"
## [21] "fastDummies"  "e1071"        "corrplot"     "caTools"
## [25] "caret"        "ggplot2"      "lattice"      "stats"
## [29] "graphics"     "grDevices"    "utils"        "datasets"
## [33] "methods"      "base"
```

Internet Inequality– A Global Persepective

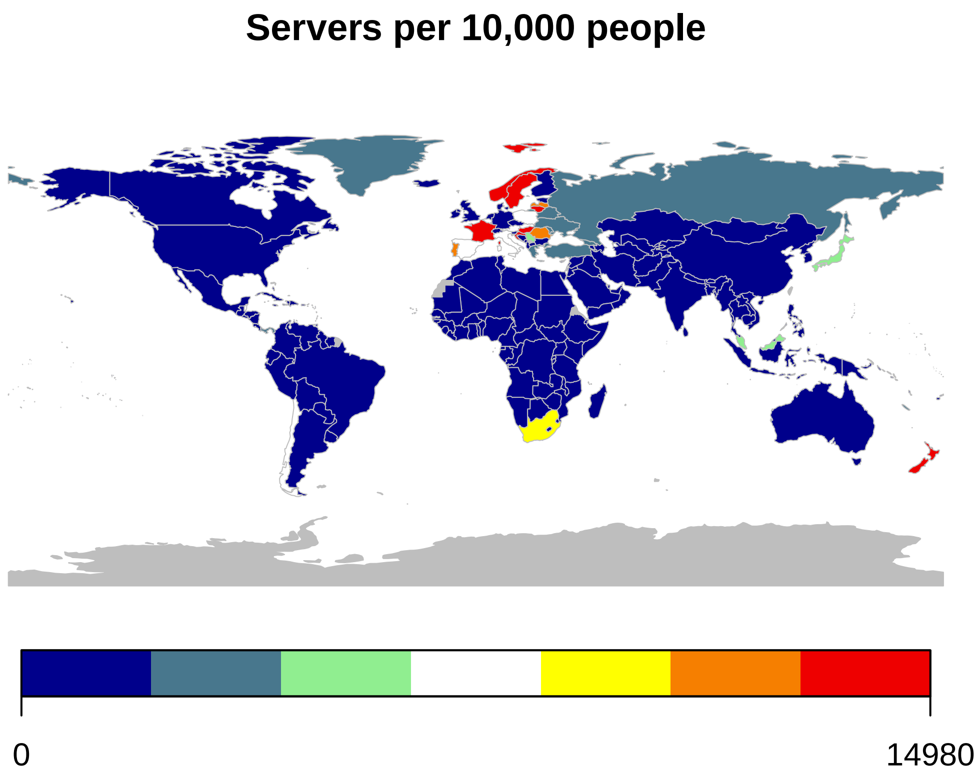
```
## 215 codes from your data successfully matched countries in the map
## 66 codes from your data failed to match with a country code in the map
## 28 codes from the map weren't represented in your data
```



Immediately, we can see that broadband subscription rates are higher in strongly developed places like North America, Western Europe, and Australia. Conversely, poor countries across South America, Africa, and South Asia have significantly lower broadband access rates. Please note that countries in grey have unknown values.

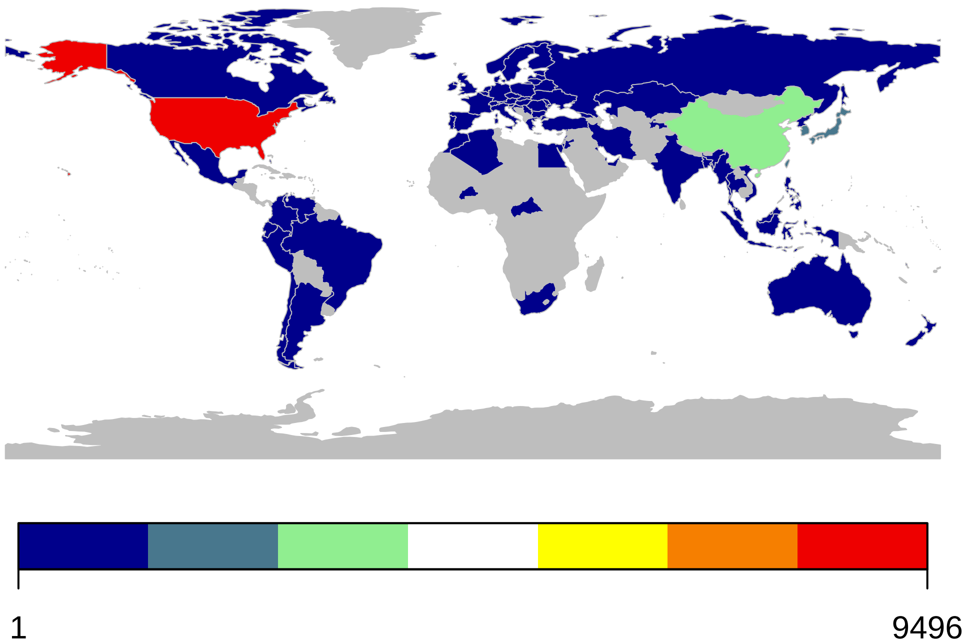


However, the number of cell phone subscriptions per 100 people is much more uniform. That is due to the lower cost of wireless network deployment compared to the capital-intesive processes of digging trenches to lay copper or fiber.

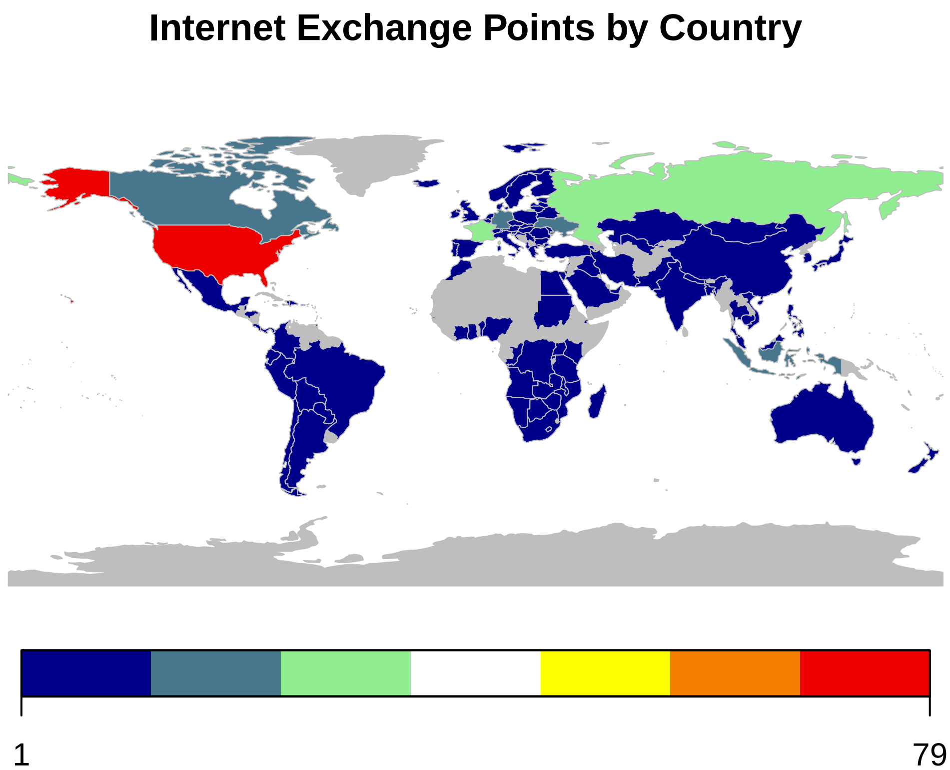


When we look at the number of servers available in each country, we find that Western Europe has the highest per capita server load. Countries like New Zealand and South Africa are also high because they are conveniently located for undersea cables that compose the back-bone of the internet. TODO Source

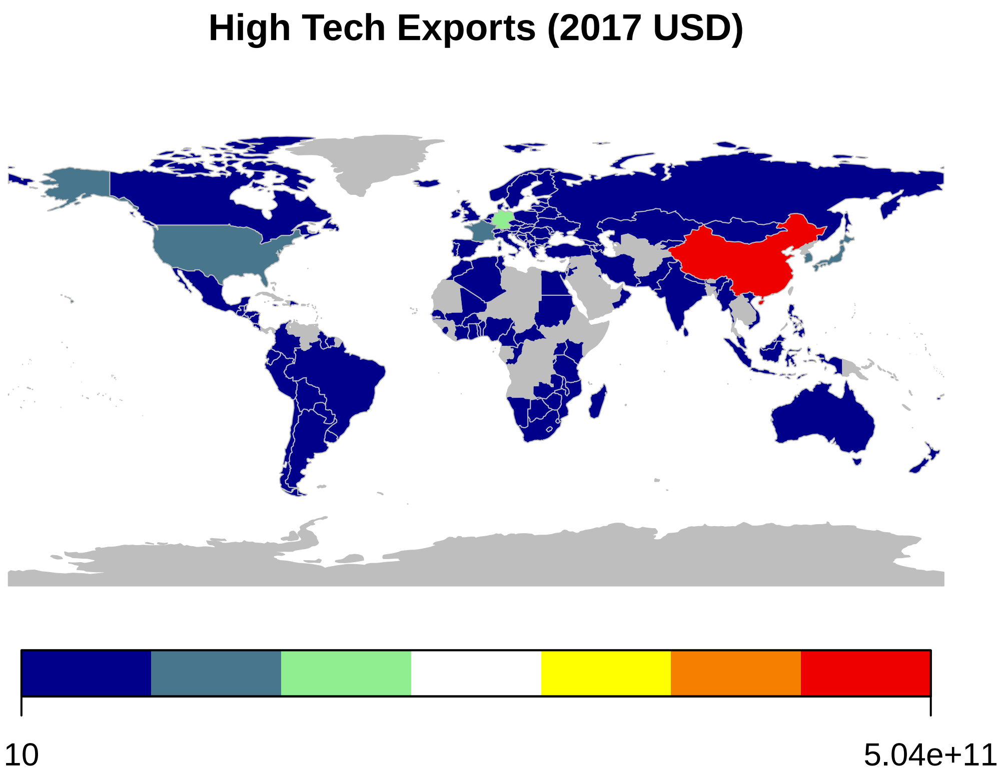
Mac Addresses Blocks Assigned per Country



The IEEE is a global organization that manages technological standards, publishes and circulates literature about electronics and and electrical engineering. In addition, they allocate MAC addresses which are the physical address of every bluetooth/wifi radio, ethernet port, and fiber cable on the internet. As we can see, a relatively small number of countries have original electronics manufacturers, with the US registering more than twice the number of devices as the next country (China).



The undersea cables mentioned earlier wind up at one of 600 buildings around the world where network operators connect their computers to their peers and create what we think of as the 'inter' net. These 600 buildings are not even distributed, with most countries only have a single access point to the Internet. In addition, regimes known for censorship (ie Egypt, Turkey, and China) have relatively few internet exchange points, allowing for centralized control and censorship. TODO: Source



The net result of the modern Internet infrastructure is a centralized model with a few players making all of the profits. Above we see the total amount of high tech exports as measured in 2017 USD. Three countries account for the bulk of the profit here, seeming to indicate that a centralized Internet infrastructure does not raise the standards for everybody. It is apparent that today's paradigm encourages consumption over creation.

Correlation between Various Technology Indicators and Poverty Rates

Data Initialization and Preprocessing

```
## # A tibble: 6 x 28
##   Percent.Limited~ Percent.With.Br~ Percent.Occupie~ Percent.In.Serv~
##              <dbl>            <dbl>            <dbl>            <dbl>
## 1              1               82.2             85               68.3
## 2              2.1             88.7             79.5             58
## 3              4               87.7             82.3             68.2
## 4              1.5             77.1             86.6             64.2
## 5              9               90.2             82.4             65.7
## 6              2.7             90.9             83.4             70.1
## # ... with 24 more variables: Percent.STEM.education <dbl>,
## #   Percent.Computer.in.Household <dbl>,
## #   Percent.Smartphone.In.Household <dbl>,
## #   Percent.No.of.Internet.Subscriptions <dbl>, Employment.Rate <dbl>,
```

```
## #   Median.Annual.Income <dbl>, Percent.in.Public.School <dbl>,
## #   Percent.Foreign.Born <dbl>, White <dbl>, Black <dbl>, Native <dbl>,
## #   Asian <dbl>, Other <dbl>, Two.or.more <dbl>, Population <dbl>,
## #   Median.Age <dbl>, Percent.male <dbl>,
## #   Percent.Less.Than.High.School <dbl>, Percent.High.School <dbl>,
## #   Percent.Some.College <dbl>, Percent.Bachelors <dbl>,
## #   Population.Graduate <dbl>, Percent.Below.Poverty.Line <dbl>,
## #   Median.Monthly.Housing.Costs <dbl>
```

Fixing Missing Data

Below we impute the missing values using a monotone cubic approximator (known as a Stineman interpolation). It has a tendency to perform well on linear as well as higher-order data vectors.

```
##    Percent.Limited.English.Households Percent.With.Broadband
## 1                                 1.0                   82.2
## 2                                 2.1                   88.7
## 3                                 4.0                   87.7
## 4                                 1.5                   77.1
## 5                                 9.0                   90.2
## 6                                 2.7                   90.9
## 7                                 5.3                   89.6
## 8                                 2.4                   89.3
## 9                                 3.6                   84.1
## 10                                6.9                   85.6
## 11                                2.8                   85.6
## 12                                5.5                   88.3
## 13                                2.1                   88.3
## 14                                4.5                   88.0
## 15                                1.7                   85.1
## 16                                1.9                   86.6
## 17                                2.4                   87.2
## 18                                1.2                   83.3
## 19                                2.1                   80.6
## 20                                0.8                   87.0
## 21                                3.2                   90.7
## 22                                6.1                   90.7
## 23                                1.6                   87.0
## 24                                2.4                   90.0
## 25                                0.7                   78.3
## 26                                1.2                   84.9
## 27                                0.4                   84.8
## 28                                3.0                   88.8
## 29                                5.8                   86.4
## 30                                1.3                   91.8
## 31                                7.2                   90.0
## 32                                5.5                   79.2
## 33                                8.2                   86.9
## 34                                2.4                   85.4
## 35                                0.8                   84.5
## 36                                1.3                   87.3
## 37                                1.9                   83.6
## 38                                2.5                   90.5
```

```
## 39                               2.5                   86.3
## 40                              68.2                   65.8
## 41                               6.3                   90.1
## 42                               1.3                   83.1
## 43                               1.1                   84.8
## 44                               1.6                   83.3
## 45                               7.9                   85.7
## 46                               2.4                   91.0
## 47                               0.5                   85.4
## 48                               2.7                   88.2
## 49                               3.8                   92.1
## 50                               0.4                   81.6
## 51                               1.5                   87.4
## 52                               1.5                   87.8
##    Percent.Occupied.Single.Family.Homes Percent.In.Service.Industry
## 1                                  85.0                        68.3
## 2                                  79.5                        58.0
## 3                                  82.3                        68.2
## 4                                  86.6                        64.2
## 5                                  82.4                        65.7
## 6                                  83.4                        70.1
## 7                                  83.8                        67.2
## 8                                  76.3                        69.9
## 9                                  25.7                        66.4
## 10                                 73.7                        71.9
## 11                                 86.0                        67.3
## 12                                 72.8                        72.5
## 13                                 90.4                        65.2
## 14                                 79.6                        68.3
## 15                                 92.0                        69.0
## 16                                 90.5                        67.6
## 17                                 92.2                        65.4
## 18                                 85.2                        67.6
## 19                                 83.8                        67.4
## 20                                 83.5                        61.6
## 21                                 71.0                        65.3
## 22                                 77.2                        66.7
## 23                                 88.8                        71.5
## 24                                 85.8                        68.7
## 25                                 83.0                        66.8
## 26                                 89.5                        67.9
## 27                                 84.7                        60.7
## 28                                 93.1                        65.0
## 29                                 84.2                        83.5
## 30                                 83.2                        64.8
## 31                                 78.1                        70.8
## 32                                 78.2                        64.8
## 33                                 69.4                        66.4
## 34                                 82.7                        67.2
## 35                                 83.0                        58.6
## 36                                 89.6                        70.4
## 37                                 88.8                        64.4
## 38                                 84.1                        66.3
## 39                                 75.6                        69.1
```

```
## 40                                    81.1                          54.9
## 41                                    83.0                          70.2
## 42                                    80.2                          70.2
## 43                                    86.5                          65.2
## 44                                    87.0                          70.6
## 45                                    88.7                          69.5
## 46                                    87.5                          67.7
## 47                                    83.3                          60.2
## 48                                    79.7                          65.4
## 49                                    84.5                          66.2
## 50                                    83.4                          67.7
## 51                                    88.8                          68.5
## 52                                    81.3                          62.0
##    Percent.STEM.education Percent.Computer.in.Household
## 1                    28.8                          86.1
## 2                    39.9                          94.2
## 3                    32.5                          92.2
## 4                    29.7                          86.3
## 5                    41.1                          93.5
## 6                    39.3                          94.2
## 7                    36.1                          90.9
## 8                    34.5                          92.9
## 9                    49.4                          91.9
## 10                   31.7                          91.9
## 11                   32.7                          90.9
## 12                   34.8                          91.3
## 13                   33.2                          91.3
## 14                   32.6                          90.3
## 15                   29.0                          88.8
## 16                   30.1                          89.5
## 17                   29.7                          90.7
## 18                   29.2                          86.3
## 19                   28.1                          86.0
## 20                   34.6                          89.7
## 21                   41.1                          93.0
## 22                   40.3                          91.2
## 23                   33.5                          90.0
## 24                   34.4                          92.1
## 25                   25.5                          84.6
## 26                   29.1                          89.7
## 27                   34.6                          88.7
## 28                   28.3                          90.8
## 29                   32.6                          93.3
## 30                   38.1                          93.7
## 31                   37.7                          92.6
## 32                   36.1                          86.6
## 33                   34.3                          90.1
## 34                   34.3                          89.4
## 35                   28.4                          90.5
## 36                   30.7                          89.6
## 37                   28.1                          89.2
## 38                   40.1                          93.2
## 39                   33.6                          88.0
## 40                   24.6                          69.2
```

```
## 41                         33.9                             89.7
## 42                         30.9                             88.8
## 43                         28.2                             88.4
## 44                         29.8                             87.4
## 45                         35.1                             91.8
## 46                         32.4                             95.5
## 47                         38.6                             89.5
## 48                         40.3                             91.8
## 49                         42.1                             94.2
## 50                         28.2                             84.6
## 51                         32.3                             89.5
## 52                         35.5                             91.7
##    Percent.Smartphone.In.Household Percent.No.of.Internet.Subscriptions
## 1                             77.3                                 78.5
## 2                             87.6                                 86.4
## 3                             83.3                                 86.0
## 4                             78.2                                 73.3
## 5                             86.8                                 87.9
## 6                             86.1                                 88.6
## 7                             81.1                                 85.9
## 8                             83.3                                 86.5
## 9                             86.5                                 82.8
## 10                            82.8                                 83.4
## 11                            83.5                                 82.9
## 12                            83.1                                 84.8
## 13                            79.8                                 83.1
## 14                            82.0                                 84.0
## 15                            78.7                                 81.6
## 16                            78.1                                 82.4
## 17                            80.8                                 83.3
## 18                            76.5                                 79.2
## 19                            78.6                                 75.8
## 20                            73.8                                 82.9
## 21                            84.4                                 87.9
## 22                            81.3                                 87.0
## 23                            79.5                                 83.1
## 24                            81.9                                 86.4
## 25                            77.0                                 73.7
## 26                            79.9                                 81.6
## 27                            75.9                                 82.0
## 28                            80.3                                 84.9
## 29                            85.4                                 83.6
## 30                            80.8                                 88.9
## 31                            83.9                                 87.0
## 32                            76.7                                 77.0
## 33                            80.4                                 83.7
## 34                            80.4                                 82.0
## 35                            80.2                                 81.6
## 36                            78.8                                 83.6
## 37                            81.3                                 80.0
## 38                            83.1                                 87.1
## 39                            75.8                                 82.0
## 40                            61.7                                 62.0
## 41                            79.5                                 85.8
```

```
## 42                        80.5                                  79.5
## 43                        75.0                                  80.9
## 44                        78.8                                  79.6
## 45                        85.7                                  83.5
## 46                        88.9                                  88.1
## 47                        72.9                                  82.0
## 48                        83.4                                  85.2
## 49                        85.5                                  89.4
## 50                        69.7                                  76.3
## 51                        77.8                                  83.9
## 52                        80.5                                  84.0
##    Employment.Rate Median.Annual.Income Percent.in.Public.School
## 1             52.9                48123                     86.7
## 2             60.2                73181                     88.3
## 3             56.1                56581                     88.8
## 4             54.6                45869                     88.6
## 5             59.5                71805                     86.0
## 6             64.6                69117                     88.4
## 7             61.4                74168                     80.0
## 8             55.8                62852                     84.4
## 9             65.3                82372                     58.1
## 10            54.9                52594                     82.1
## 11            59.1                56183                     85.9
## 12            59.0                77765                     77.9
## 13            59.2                52225                     86.3
## 14            60.8                62992                     81.7
## 15            60.4                54181                     83.6
## 16            64.8                58570                     84.4
## 17            63.0                56422                     85.8
## 18            55.6                48375                     83.9
## 19            54.7                46145                     79.9
## 20            60.3                56277                     82.4
## 21            63.8                80776                     81.5
## 22            63.6                77385                     72.8
## 23            57.8                54909                     87.2
## 24            66.9                68388                     84.7
## 25            52.2                43529                     87.1
## 26            59.8                53578                     81.2
## 27            61.3                53386                     84.6
## 28            67.4                59970                     82.0
## 29            59.6                58003                     88.5
## 30            65.1                73381                     79.2
## 31            61.8                80088                     81.7
## 32            52.6                46744                     90.5
## 33            59.6                64894                     76.8
## 34            58.0                52752                     85.1
## 35            67.9                61843                     90.6
## 36            59.6                54021                     82.1
## 37            57.2                50051                     88.8
## 38            59.1                60212                     84.9
## 39            59.0                59195                     76.4
## 40            36.2                19343                     65.4
## 41            61.3                63870                     73.9
## 42            56.1                50570                     86.2
```

```
## 43            65.1              56521                    87.8
## 44            58.1              51340                    82.8
## 45            60.7              59206                    89.2
## 46            66.0              68358                    85.9
## 47            62.8              57513                    80.5
## 48            61.1              71535                    83.3
## 49            60.6              70979                    85.0
## 50            48.8              43469                    88.8
## 51            63.9              59305                    84.1
## 52            61.9              60434                    91.1
##    Percent.Foreign.Born     White       Black      Native       Asian
## 1            0.03592806 0.6795672 0.268212279 0.005165601 0.013725430
## 2            0.09528630 0.6418751 0.029882603 0.148619550 0.066754979
## 3            0.15183776 0.7756852 0.043848797 0.045319522 0.033139688
## 4            0.04890648 0.7633469 0.152571382 0.006360927 0.015938932
## 5            0.36884416 0.5863362 0.057463564 0.007938431 0.145821777
## 6            0.10892523 0.8418700 0.040918441 0.010128489 0.031968268
## 7            0.17339947 0.7588833 0.106064238 0.003140307 0.045684391
## 8            0.11406967 0.6882806 0.218679147 0.002836978 0.040470342
## 9            0.17202826 0.4100108 0.458548760 0.002435257 0.041066498
## 10           0.26373535 0.7514303 0.161763405 0.003032491 0.028024961
## 11           0.11384398 0.5870873 0.315652639 0.003533767 0.039399757
## 12           0.22854662 0.2501601 0.016382051 0.001864048 0.382223801
## 13           0.06259426 0.9004440 0.006608839 0.012574092 0.013461717
## 14           0.16704535 0.7123805 0.142085825 0.002336349 0.054285795
## 15           0.05561369 0.8366656 0.093592025 0.001859508 0.022254395
## 16           0.05596809 0.8999765 0.034038728 0.003430385 0.025976321
## 17           0.07393627 0.8452149 0.057397851 0.007354993 0.029466658
## 18           0.03945879 0.8694445 0.081190313 0.001745997 0.014565615
## 19           0.04260093 0.6171248 0.324771744 0.005091013 0.018106313
## 20           0.09528630 0.9443854 0.012255344 0.006991505 0.011345850
## 21           0.18108750 0.5489732 0.298516220 0.003127965 0.064659047
## 22           0.20277076 0.7853576 0.077536594 0.002200787 0.065995473
## 23           0.07601724 0.7844340 0.137981539 0.005366325 0.030921741
## 24           0.09552226 0.8267511 0.064758027 0.010903227 0.048989296
## 25           0.02251864 0.5817121 0.380125666 0.004606414 0.009280520
## 26           0.04390081 0.8200819 0.114235110 0.004182034 0.020367604
## 27           0.09528630 0.8858203 0.004412214 0.061788132 0.007011946
## 28           0.08067055 0.8734508 0.046327333 0.008134574 0.024539654
## 29           0.24813241 0.6460420 0.091576527 0.012644932 0.085063937
## 30           0.06602505 0.9307087 0.016688325 0.001329317 0.026558782
## 31           0.29566761 0.6787119 0.135200992 0.001981313 0.098411729
## 32           0.10330335 0.7576470 0.021278501 0.095728113 0.014092918
## 33           0.29658212 0.6309187 0.158039243 0.003832307 0.087391210
## 34           0.08782462 0.6878669 0.214785263 0.011671772 0.028780292
## 35           0.09528630 0.8659281 0.030836929 0.055210996 0.016811117
## 36           0.04746414 0.8131394 0.123629929 0.002034548 0.022088227
## 37           0.06039497 0.7218494 0.072939690 0.076929652 0.022057237
## 38           0.10945679 0.8441784 0.018767609 0.011512088 0.043849583
## 39           0.07503244 0.8074232 0.112223408 0.001836940 0.034718419
## 40           0.02816030 0.6624545 0.120958223 0.002377459 0.001625326
## 41           0.16109298 0.8175029 0.062908217 0.005265944 0.035916005
## 42           0.05110673 0.6727649 0.270160293 0.002986843 0.015162302
## 43           0.09528630 0.8469700 0.019532786 0.087479561 0.012389814
```

```
## 44           0.05460764 0.7774871 0.167164633 0.002179427 0.017736046
## 45           0.20699555 0.7394255 0.120942585 0.004773006 0.048057496
## 46           0.09505033 0.8565945 0.012144110 0.010540219 0.024402023
## 47           0.09528630 0.9416474 0.013244460 0.003776114 0.018070831
## 48           0.14341525 0.6746972 0.192164954 0.003163157 0.064436211
## 49           0.16706926 0.7537881 0.036672080 0.012562953 0.085360240
## 50           0.09528630 0.9275626 0.039645743 0.001342617 0.007562820
## 51           0.05230243 0.8527634 0.063900627 0.008499896 0.027526610
## 52           0.09528630 0.9121359 0.009953134 0.024069807 0.008340885
##          Other Two.or.more Population Median.Age Percent.male
## 1  0.013807486  0.01726038    4874747       38.9        46.90
## 2  0.015246115  0.08069127     739795       34.5        54.50
## 3  0.061692751  0.03014907    7016270       37.7        49.40
## 4  0.030586374  0.02612540    3004279       38.1        48.55
## 5  0.149527149  0.03870770   39536653       36.5        49.40
## 6  0.039090954  0.02884387    5607154       36.8        50.60
## 7  0.053334221  0.02553325    3588184       40.9        47.65
## 8  0.019098924  0.02813380     961939       40.1        46.90
## 9  0.057806367  0.02568259     693972       34.0        45.10
## 10 0.028689503  0.02217709   20984400       42.0        47.80
## 11 0.026064160  0.02399932   10429379       36.8        47.30
## 12 0.015044083  0.22636455    1427538       39.2        50.35
## 13 0.034417567  0.02706846    1716943       36.3        50.50
## 14 0.062327337  0.02093255   12802023       38.0        48.35
## 15 0.018858322  0.02334442    6666818       37.7        48.60
## 16 0.013179532  0.01977740    3145711       38.3        49.35
## 17 0.024504630  0.03194681    2913123       36.7        49.75
## 18 0.009469064  0.02107252    4454189       38.9        48.55
## 19 0.016135915  0.01701800    4684333       36.8        47.80
## 20 0.001961963  0.02163624    1335907       44.6        48.05
## 21 0.050445154  0.02984876    6052177       38.7        47.10
## 22 0.038170249  0.02591934    6859819       39.5        47.25
## 23 0.011435901  0.02671729    9962311       39.8        48.55
## 24 0.019856701  0.02528850    5576606       37.9        49.75
## 25 0.010359572  0.01213130    2984100       37.5        47.05
## 26 0.013288063  0.02456730    6113532       38.5        48.25
## 27 0.007413662  0.03001353    1050493       40.0        50.20
## 28 0.020219512  0.02419488    1920076       36.5        49.90
## 29 0.111316431  0.03898482    2998039       38.0        50.50
## 30 0.003364624  0.01943707    1342795       43.2        49.15
## 31 0.059781510  0.01962514    9005644       39.8        47.70
## 32 0.077131514  0.02430857    2088070       37.7        49.05
## 33 0.089113932  0.02264865   19849399       38.7        47.20
## 34 0.030091345  0.02335941   10273419       38.8        47.40
## 35 0.010922791  0.01822627     755393       35.4        52.90
## 36 0.009499246  0.02686830   11658609       39.3        48.00
## 37 0.027349458  0.07325743    3930864       36.6        49.15
## 38 0.029822274  0.04303660    4142776       39.3        49.15
## 39 0.018976166  0.02188007   12805537       40.8        48.00
## 40 0.159899820  0.04754168    3337177       41.4        45.60
## 41 0.044640675  0.02486790    1059639       39.5        46.95
## 42 0.016312297  0.01949598    5024369       39.4        47.10
## 43 0.006493297  0.02458300     869666       36.9        50.85
## 44 0.013290681  0.01964284    6715984       38.6        47.50
```

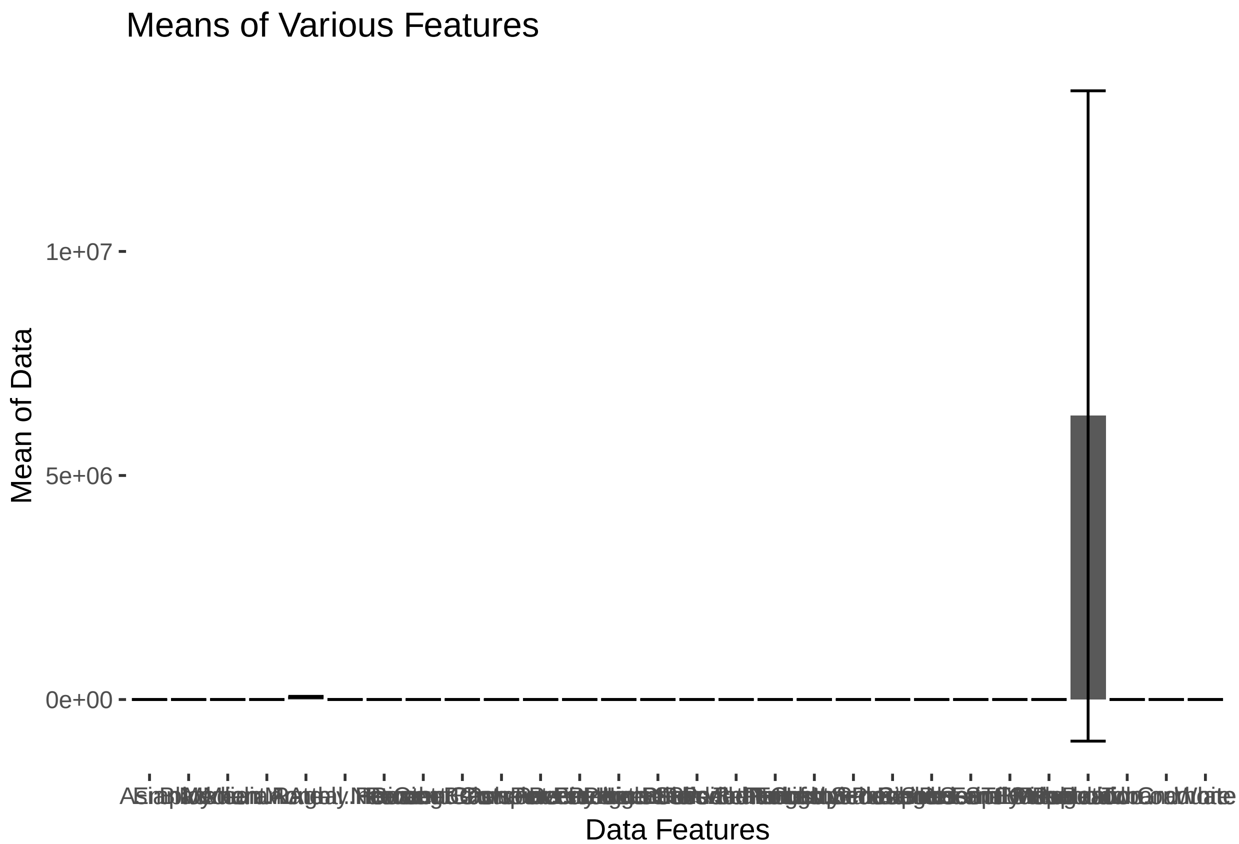
```
## 45 0.059491893  0.02073101   28304596       34.7        49.35
## 46 0.058011827  0.02427565    3101833       31.0        50.80
## 47 0.002231996  0.01966947     623657       42.6        48.75
## 48 0.025915051  0.03535139    8470020       38.2        48.35
## 49 0.045908425  0.05350685    7405743       37.7        49.90
## 50 0.006887106  0.01549682    1815857       42.4        48.85
## 51 0.021873414  0.02117580    5795483       39.5        49.40
## 52 0.017263492  0.02434427     579315       37.5        52.15
##    Percent.Less.Than.High.School Percent.High.School Percent.Some.College
## 1                            9.4                31.1                 21.4
## 2                            5.4                27.6                 26.4
## 3                            7.4                24.1                 25.0
## 4                            8.7                34.0                 22.1
## 5                            7.5                20.8                 21.1
## 6                            5.0                21.3                 20.9
## 7                            5.5                27.1                 16.5
## 8                            6.0                32.4                 19.0
## 9                            5.5                17.2                 12.5
## 10                           6.9                28.8                 19.9
## 11                           8.4                28.1                 20.2
## 12                           4.3                28.1                 20.5
## 13                           5.9                28.2                 26.3
## 14                           6.0                26.1                 20.6
## 15                           7.7                32.7                 20.2
## 16                           4.8                30.5                 21.0
## 17                           5.5                25.8                 22.7
## 18                           8.2                33.0                 21.3
## 19                          10.0                34.0                 21.4
## 20                           5.1                30.9                 19.0
## 21                           6.1                24.5                 18.9
## 22                           4.9                24.3                 15.5
## 23                           6.3                28.9                 23.4
## 24                           3.9                24.8                 20.9
## 25                          10.8                30.4                 22.0
## 26                           7.2                30.8                 22.0
## 27                           4.8                28.1                 23.5
## 28                           4.8                26.3                 23.1
## 29                           8.0                28.7                 25.1
## 30                           5.1                28.0                 17.9
## 31                           5.3                27.2                 16.3
## 32                           8.1                26.6                 24.0
## 33                           7.3                26.3                 15.4
## 34                           7.7                25.8                 21.3
## 35                           4.3                26.4                 22.4
## 36                           7.0                33.3                 20.2
## 37                           7.9                31.1                 23.3
## 38                           5.5                23.2                 25.2
## 39                           6.3                35.0                 15.8
## 40                           8.1                27.9                 12.2
## 41                           6.6                29.9                 16.9
## 42                           8.6                29.5                 20.3
## 43                           5.4                30.8                 22.0
## 44                           7.8                32.4                 20.8
## 45                           8.2                25.1                 21.7
```

```
## 46                            5.1                 22.3                   25.7
## 47                            5.2                 29.0                   16.8
## 48                            6.1                 24.2                   19.0
## 49                            5.0                 22.1                   23.6
## 50                            8.4                 41.2                   18.6
## 51                            5.1                 30.7                   20.3
## 52                            5.1                 29.6                   25.3
##    Percent.Bachelors Population.Graduate Percent.Below.Poverty.Line
## 1               16.0                 9.6                       29.1
## 2               18.0                10.8                       21.0
## 3               18.3                11.0                       24.6
## 4               15.0                 8.4                       30.7
## 5               21.1                12.6                       25.5
## 6               26.0                15.2                       21.0
## 7               21.4                17.3                       20.1
## 8               18.0                13.5                       26.7
## 9               23.9                33.4                       19.4
## 10              18.9                10.8                       24.9
## 11              19.0                11.9                       26.0
## 12              21.7                11.2                       23.4
## 13              18.2                 8.5                       27.8
## 14              21.0                13.4                       23.4
## 15              17.0                 9.8                       26.5
## 16              19.4                 9.5                       23.6
## 17              21.2                12.6                       25.2
## 18              14.0                 9.9                       31.0
## 19              15.5                 8.3                       31.8
## 20              19.9                12.1                       25.0
## 21              21.3                18.3                       20.1
## 22              23.9                19.5                       21.2
## 23              17.6                11.5                       26.2
## 24              23.5                12.5                       20.7
## 25              13.5                 8.3                       34.1
## 26              17.9                11.1                       25.2
## 27              21.7                10.6                       24.8
## 28              20.9                10.8                       21.8
## 29              16.5                 8.4                       23.1
## 30              22.6                14.3                       17.5
## 31              24.2                15.6                       21.2
## 32              15.2                11.8                       28.8
## 33              20.2                15.8                       24.5
## 34              20.1                11.2                       25.8
## 35              21.8                 9.0                       22.7
## 36              17.3                10.6                       25.4
## 37              16.9                 8.6                       27.9
## 38              21.0                12.7                       25.8
## 39              18.9                12.5                       25.6
## 40              18.3                 7.4                       56.6
## 41              20.3                13.1                       22.7
## 42              17.6                10.4                       26.9
## 43              19.1                 9.0                       22.9
## 44              17.2                10.1                       26.5
## 45              19.3                10.3                       24.9
## 46              22.8                11.8                       26.4
```

```
## 47             22.5                 15.8                           22.2
## 48             22.0                 16.7                           22.7
## 49             22.2                 13.3                           22.5
## 50             12.2                  8.0                           33.7
## 51             19.8                 10.6                           23.2
## 52             17.4                 10.3                           23.0
##    Median.Monthly.Housing.Costs
## 1                           734
## 2                          1285
## 3                          1015
## 4                           691
## 5                          1567
## 6                          1300
## 7                          1390
## 8                          1126
## 9                          1641
## 10                         1047
## 11                          980
## 12                         1585
## 13                          880
## 14                         1081
## 15                          815
## 16                          825
## 17                          858
## 18                          743
## 19                          781
## 20                          884
## 21                         1456
## 22                         1464
## 23                          862
## 24                         1070
## 25                          672
## 26                          837
## 27                          816
## 28                          887
## 29                         1089
## 30                         1280
## 31                         1545
## 32                          782
## 33                         1291
## 34                          878
## 35                          799
## 36                          833
## 37                          770
## 38                         1164
## 39                          947
## 40                          329
## 41                         1190
## 42                          820
## 43                          774
## 44                          820
## 45                         1009
## 46                         1122
## 47                         1088
```

```
## 48                             1237
## 49                             1319
## 50                              585
## 51                              913
## 52                              890
```

Below we can see the mean and standard deviation for each data vector. As we can see, our data occurs across many orders of magnitude. For the best fit, the data should be centered and scaled.

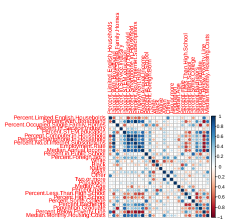


These plots confirm the non-normality of most of our data. For the non-linear models, we must center and scale them. Additionally, we will need other data transformations (discussed below). TODO: Density Plots

```
#ggplot(data = gather(training), mapping = aes(x = value)) +
#  geom_histogram(aes(y=..density..,), colour="black", fill="white")+
#  geom_density(alpha=.2, fill="lightgrey")+
#  facet_wrap(~key, ncol = 1, scales = 'free')
```

Correlation Plot of Predictors

We can see significant covariance in the data. Additionally, many data points have near-zero variance. Excluding these confounding variables will improve our model.



Data Pre-processing

For preprocessing of data, we remove near zero predictors, fill in missing values with KNN method, and transform predictors using the Yeo-Johnson transformation method. We also center and scale the data. Additionally, we remove the covariant terms and the ones with near zero variance here, since they will not improve our models.

Support Vector Machine

Comparing SVM with Neural Networks (NN), both are non-linear algorithms. A Support Vector Machine with different kernels is comparable to a Neural Network with different layers. One advantage SVMs have over NNs is that NNs need large amounts of data to train, SVMs work with smaller-sized data with less computing power. Finally SVM usually only have 2-3 parameters to tune, they are easy to code, and the results are explainable. On the other hand, SVMs might not beat NNs on the accuracy metric.

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 39 samples
## 24 predictors
##
## No pre-processing
## Resampling: Cross-Validated (3 fold, repeated 3 times)
## Summary of sample sizes: 25, 27, 26, 26, 27, 25, ...
## Resampling results across tuning parameters:
##
##   C      RMSE      Rsquared   MAE
##    0.25  4.951241  0.4549275  2.878613
##    0.50  4.784790  0.4569550  2.712984
##    1.00  4.699560  0.4686361  2.695841
##    2.00  4.619132  0.4881648  2.666214
##    4.00  4.639514  0.4749878  2.689218
##    8.00  4.639514  0.4749878  2.689218
##   16.00  4.639514  0.4749878  2.689218
##   32.00  4.639514  0.4749878  2.689218
##
## Tuning parameter 'sigma' was held constant at a value of 0.04847725
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were sigma = 0.04847725 and C = 2.
```

Model	RMSE.train	RSquared.train	RMSE.test	RSquared.test
Support Vector Machine	2.848699	0.8712312	1.535414	0.5913622

Random Forest Model

Random forests are a modification of bagging that builds a large collection of de-correlated trees [1]. They are considered to belong in the category of non-parametric models since the number of parameters grows with the size of the training set. They are considered to be an improvement to the use of CART (Classification and Regression Tree) models because they do not suffer from some of the problems associated with CART models, such as the fact that CART models are unstable: small changes to the structure of the input data can have large effects on the CART model [2]. Random forests are designed to be low-variance estimators.

Random forests are based on the basic idea of aggregating uncorrelated sets of predictors, since one way to reduce the variance of an estimate is to average several estimates together [2]. A random forest trains a randomly chosen set of input variables over a randomly chosen subset of the data, and aggregates together several such trees to produce an overall estimator. Random forests have proven to be quite successful in a variety of real-world applications and often are seen to generalize very well to unseen real-world data.

```
##    Percent.Limited.English.Households Percent.With.Broadband
## 1                                 1.0                   82.2
## 3                                 4.0                   87.7
## 6                                 2.7                   90.9
## 7                                 5.3                   89.6
## 9                                 3.6                   84.1
## 10                                6.9                   85.6
## 12                                5.5                   88.3
## 13                                2.1                   88.3
## 14                                4.5                   88.0
## 15                                1.7                   85.1
## 17                                2.4                   87.2
## 18                                1.2                   83.3
## 19                                2.1                   80.6
## 22                                6.1                   90.7
## 23                                1.6                   87.0
## 25                                0.7                   78.3
## 26                                1.2                   84.9
## 27                                0.4                   84.8
## 28                                3.0                   88.8
## 29                                5.8                   86.4
## 30                                1.3                   91.8
## 33                                8.2                   86.9
## 35                                0.8                   84.5
## 36                                1.3                   87.3
## 37                                1.9                   83.6
## 38                                2.5                   90.5
## 39                                2.5                   86.3
## 40                               68.2                   65.8
## 41                                6.3                   90.1
## 42                                1.3                   83.1
## 43                                1.1                   84.8
## 44                                1.6                   83.3
## 45                                7.9                   85.7
## 46                                2.4                   91.0
## 47                                0.5                   85.4
## 48                                2.7                   88.2
## 49                                3.8                   92.1
## 51                                1.5                   87.4
## 52                                1.5                   87.8
##    Percent.Occupied.Single.Family.Homes Percent.In.Service.Industry
## 1                                   85.0                        68.3
## 3                                   82.3                        68.2
## 6                                   83.4                        70.1
## 7                                   83.8                        67.2
## 9                                   25.7                        66.4
## 10                                  73.7                        71.9
## 12                                  72.8                        72.5
```

```
## 13                                   90.4                          65.2
## 14                                   79.6                          68.3
## 15                                   92.0                          69.0
## 17                                   92.2                          65.4
## 18                                   85.2                          67.6
## 19                                   83.8                          67.4
## 22                                   77.2                          66.7
## 23                                   88.8                          71.5
## 25                                   83.0                          66.8
## 26                                   89.5                          67.9
## 27                                   84.7                          60.7
## 28                                   93.1                          65.0
## 29                                   84.2                          83.5
## 30                                   83.2                          64.8
## 33                                   69.4                          66.4
## 35                                   83.0                          58.6
## 36                                   89.6                          70.4
## 37                                   88.8                          64.4
## 38                                   84.1                          66.3
## 39                                   75.6                          69.1
## 40                                   81.1                          54.9
## 41                                   83.0                          70.2
## 42                                   80.2                          70.2
## 43                                   86.5                          65.2
## 44                                   87.0                          70.6
## 45                                   88.7                          69.5
## 46                                   87.5                          67.7
## 47                                   83.3                          60.2
## 48                                   79.7                          65.4
## 49                                   84.5                          66.2
## 51                                   88.8                          68.5
## 52                                   81.3                          62.0
##    Percent.STEM.education Percent.Computer.in.Household
## 1                    28.8                          86.1
## 3                    32.5                          92.2
## 6                    39.3                          94.2
## 7                    36.1                          90.9
## 9                    49.4                          91.9
## 10                   31.7                          91.9
## 12                   34.8                          91.3
## 13                   33.2                          91.3
## 14                   32.6                          90.3
## 15                   29.0                          88.8
## 17                   29.7                          90.7
## 18                   29.2                          86.3
## 19                   28.1                          86.0
## 22                   40.3                          91.2
## 23                   33.5                          90.0
## 25                   25.5                          84.6
## 26                   29.1                          89.7
## 27                   34.6                          88.7
## 28                   28.3                          90.8
## 29                   32.6                          93.3
## 30                   38.1                          93.7
```

```
## 33                          34.3                          90.1
## 35                          28.4                          90.5
## 36                          30.7                          89.6
## 37                          28.1                          89.2
## 38                          40.1                          93.2
## 39                          33.6                          88.0
## 40                          24.6                          69.2
## 41                          33.9                          89.7
## 42                          30.9                          88.8
## 43                          28.2                          88.4
## 44                          29.8                          87.4
## 45                          35.1                          91.8
## 46                          32.4                          95.5
## 47                          38.6                          89.5
## 48                          40.3                          91.8
## 49                          42.1                          94.2
## 51                          32.3                          89.5
## 52                          35.5                          91.7
##    Percent.Smartphone.In.Household Percent.No.of.Internet.Subscriptions
## 1                             77.3                                 78.5
## 3                             83.3                                 86.0
## 6                             86.1                                 88.6
## 7                             81.1                                 85.9
## 9                             86.5                                 82.8
## 10                            82.8                                 83.4
## 12                            83.1                                 84.8
## 13                            79.8                                 83.1
## 14                            82.0                                 84.0
## 15                            78.7                                 81.6
## 17                            80.8                                 83.3
## 18                            76.5                                 79.2
## 19                            78.6                                 75.8
## 22                            81.3                                 87.0
## 23                            79.5                                 83.1
## 25                            77.0                                 73.7
## 26                            79.9                                 81.6
## 27                            75.9                                 82.0
## 28                            80.3                                 84.9
## 29                            85.4                                 83.6
## 30                            80.8                                 88.9
## 33                            80.4                                 83.7
## 35                            80.2                                 81.6
## 36                            78.8                                 83.6
## 37                            81.3                                 80.0
## 38                            83.1                                 87.1
## 39                            75.8                                 82.0
## 40                            61.7                                 62.0
## 41                            79.5                                 85.8
## 42                            80.5                                 79.5
## 43                            75.0                                 80.9
## 44                            78.8                                 79.6
## 45                            85.7                                 83.5
## 46                            88.9                                 88.1
## 47                            72.9                                 82.0
```

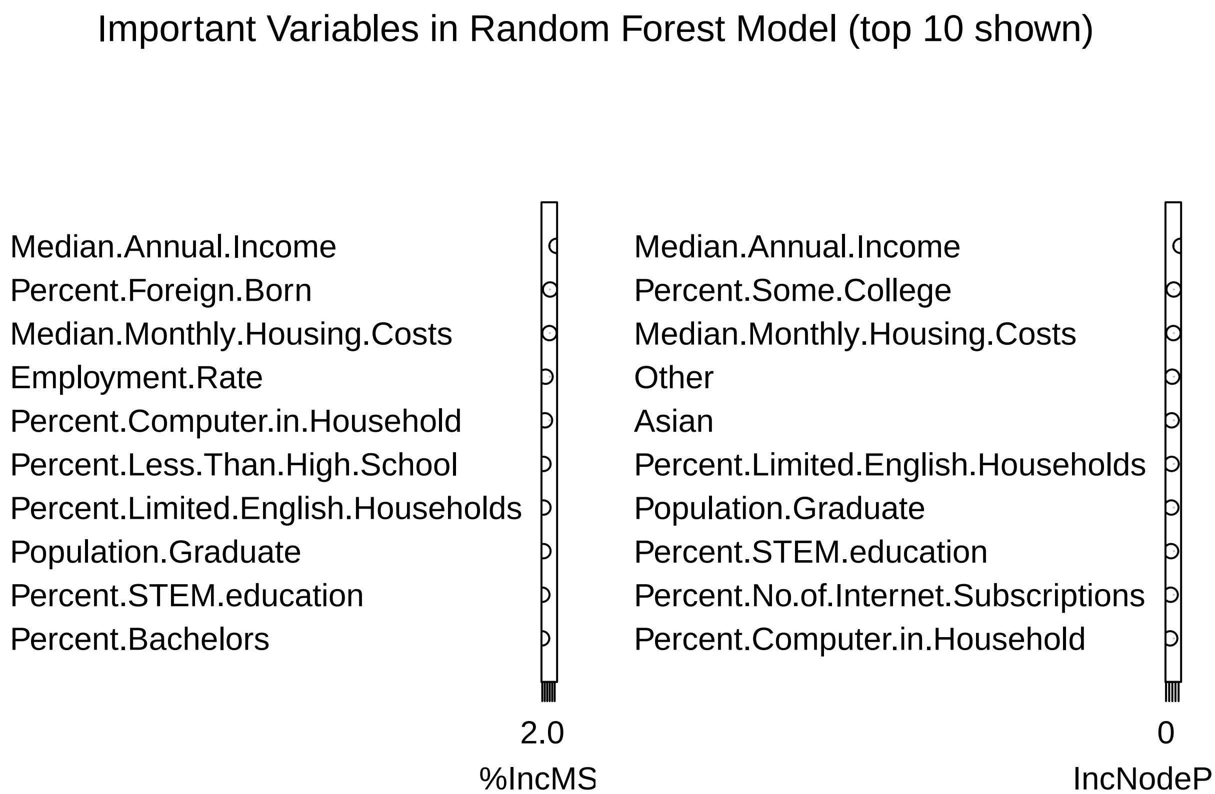
```
## 48                     83.4                                   85.2
## 49                     85.5                                   89.4
## 51                     77.8                                   83.9
## 52                     80.5                                   84.0
##    Employment.Rate Median.Annual.Income Percent.in.Public.School
## 1             52.9                48123                     86.7
## 3             56.1                56581                     88.8
## 6             64.6                69117                     88.4
## 7             61.4                74168                     80.0
## 9             65.3                82372                     58.1
## 10            54.9                52594                     82.1
## 12            59.0                77765                     77.9
## 13            59.2                52225                     86.3
## 14            60.8                62992                     81.7
## 15            60.4                54181                     83.6
## 17            63.0                56422                     85.8
## 18            55.6                48375                     83.9
## 19            54.7                46145                     79.9
## 22            63.6                77385                     72.8
## 23            57.8                54909                     87.2
## 25            52.2                43529                     87.1
## 26            59.8                53578                     81.2
## 27            61.3                53386                     84.6
## 28            67.4                59970                     82.0
## 29            59.6                58003                     88.5
## 30            65.1                73381                     79.2
## 33            59.6                64894                     76.8
## 35            67.9                61843                     90.6
## 36            59.6                54021                     82.1
## 37            57.2                50051                     88.8
## 38            59.1                60212                     84.9
## 39            59.0                59195                     76.4
## 40            36.2                19343                     65.4
## 41            61.3                63870                     73.9
## 42            56.1                50570                     86.2
## 43            65.1                56521                     87.8
## 44            58.1                51340                     82.8
## 45            60.7                59206                     89.2
## 46            66.0                68358                     85.9
## 47            62.8                57513                     80.5
## 48            61.1                71535                     83.3
## 49            60.6                70979                     85.0
## 51            63.9                59305                     84.1
## 52            61.9                60434                     91.1
##    Percent.Foreign.Born     White       Black      Native       Asian
## 1            0.03592806 0.6795672 0.268212279 0.005165601 0.013725430
## 3            0.15183776 0.7756852 0.043848797 0.045319522 0.033139688
## 6            0.10892523 0.8418700 0.040918441 0.010128489 0.031968268
## 7            0.17339947 0.7588833 0.106064238 0.003140307 0.045684391
## 9            0.17202826 0.4100108 0.458548760 0.002435257 0.041066498
## 10           0.26373535 0.7514303 0.161763405 0.003032491 0.028024961
## 12           0.22854662 0.2501601 0.016382051 0.001864048 0.382223801
## 13           0.06259426 0.9004440 0.006608839 0.012574092 0.013461717
## 14           0.16704535 0.7123805 0.142085825 0.002336349 0.054285795
```

```
## 15           0.05561369 0.8366656 0.093592025 0.001859508 0.022254395
## 17           0.07393627 0.8452149 0.057397851 0.007354993 0.029466658
## 18           0.03945879 0.8694445 0.081190313 0.001745997 0.014565615
## 19           0.04260093 0.6171248 0.324771744 0.005091013 0.018106313
## 22           0.20277076 0.7853576 0.077536594 0.002200787 0.065995473
## 23           0.07601724 0.7844340 0.137981539 0.005366325 0.030921741
## 25           0.02251864 0.5817121 0.380125666 0.004606414 0.009280520
## 26           0.04390081 0.8200819 0.114235110 0.004182034 0.020367604
## 27           0.09528630 0.8858203 0.004412214 0.061788132 0.007011946
## 28           0.08067055 0.8734508 0.046327333 0.008134574 0.024539654
## 29           0.24813241 0.6460420 0.091576527 0.012644932 0.085063937
## 30           0.06602505 0.9307087 0.016688325 0.001329317 0.026558782
## 33           0.29658212 0.6309187 0.158039243 0.003832307 0.087391210
## 35           0.09528630 0.8659281 0.030836929 0.055210996 0.016811117
## 36           0.04746414 0.8131394 0.123629929 0.002034548 0.022088227
## 37           0.06039497 0.7218494 0.072939690 0.076929652 0.022057237
## 38           0.10945679 0.8441784 0.018767609 0.011512088 0.043849583
## 39           0.07503244 0.8074232 0.112223408 0.001836940 0.034718419
## 40           0.02816030 0.6624545 0.120958223 0.002377459 0.001625326
## 41           0.16109298 0.8175029 0.062908217 0.005265944 0.035916005
## 42           0.05110673 0.6727649 0.270160293 0.002986843 0.015162302
## 43           0.09528630 0.8469700 0.019532786 0.087479561 0.012389814
## 44           0.05460764 0.7774871 0.167164633 0.002179427 0.017736046
## 45           0.20699555 0.7394255 0.120942585 0.004773006 0.048057496
## 46           0.09505033 0.8565945 0.012144110 0.010540219 0.024402023
## 47           0.09528630 0.9416474 0.013244460 0.003776114 0.018070831
## 48           0.14341525 0.6746972 0.192164954 0.003163157 0.064436211
## 49           0.16706926 0.7537881 0.036672080 0.012562953 0.085360240
## 51           0.05230243 0.8527634 0.063900627 0.008499896 0.027526610
## 52           0.09528630 0.9121359 0.009953134 0.024069807 0.008340885
##          Other Two.or.more Population Median.Age Percent.male
## 1  0.013807486  0.01726038    4874747       38.9        46.90
## 3  0.061692751  0.03014907    7016270       37.7        49.40
## 6  0.039090954  0.02884387    5607154       36.8        50.60
## 7  0.053334221  0.02553325    3588184       40.9        47.65
## 9  0.057806367  0.02568259     693972       34.0        45.10
## 10 0.028689503  0.02217709   20984400       42.0        47.80
## 12 0.015044083  0.22636455    1427538       39.2        50.35
## 13 0.034417567  0.02706846    1716943       36.3        50.50
## 14 0.062327337  0.02093255   12802023       38.0        48.35
## 15 0.018858322  0.02334442    6666818       37.7        48.60
## 17 0.024504630  0.03194681    2913123       36.7        49.75
## 18 0.009469064  0.02107252    4454189       38.9        48.55
## 19 0.016135915  0.01701800    4684333       36.8        47.80
## 22 0.038170249  0.02591934    6859819       39.5        47.25
## 23 0.011435901  0.02671729    9962311       39.8        48.55
## 25 0.010359572  0.01213130    2984100       37.5        47.05
## 26 0.013288063  0.02456730    6113532       38.5        48.25
## 27 0.007413662  0.03001353    1050493       40.0        50.20
## 28 0.020219512  0.02419488    1920076       36.5        49.90
## 29 0.111316431  0.03898482    2998039       38.0        50.50
## 30 0.003364624  0.01943707    1342795       43.2        49.15
## 33 0.089113932  0.02264865   19849399       38.7        47.20
## 35 0.010922791  0.01822627     755393       35.4        52.90
```

```
## 36 0.009499246  0.02686830   11658609       39.3        48.00
## 37 0.027349458  0.07325743    3930864       36.6        49.15
## 38 0.029822274  0.04303660    4142776       39.3        49.15
## 39 0.018976166  0.02188007   12805537       40.8        48.00
## 40 0.159899820  0.04754168    3337177       41.4        45.60
## 41 0.044640675  0.02486790    1059639       39.5        46.95
## 42 0.016312297  0.01949598    5024369       39.4        47.10
## 43 0.006493297  0.02458300     869666       36.9        50.85
## 44 0.013290681  0.01964284    6715984       38.6        47.50
## 45 0.059491893  0.02073101   28304596       34.7        49.35
## 46 0.058011827  0.02427565    3101833       31.0        50.80
## 47 0.002231996  0.01966947     623657       42.6        48.75
## 48 0.025915051  0.03535139    8470020       38.2        48.35
## 49 0.045908425  0.05350685    7405743       37.7        49.90
## 51 0.021873414  0.02117580    5795483       39.5        49.40
## 52 0.017263492  0.02434427     579315       37.5        52.15
##    Percent.Less.Than.High.School Percent.High.School Percent.Some.College
## 1                            9.4                31.1                 21.4
## 3                            7.4                24.1                 25.0
## 6                            5.0                21.3                 20.9
## 7                            5.5                27.1                 16.5
## 9                            5.5                17.2                 12.5
## 10                           6.9                28.8                 19.9
## 12                           4.3                28.1                 20.5
## 13                           5.9                28.2                 26.3
## 14                           6.0                26.1                 20.6
## 15                           7.7                32.7                 20.2
## 17                           5.5                25.8                 22.7
## 18                           8.2                33.0                 21.3
## 19                          10.0                34.0                 21.4
## 22                           4.9                24.3                 15.5
## 23                           6.3                28.9                 23.4
## 25                          10.8                30.4                 22.0
## 26                           7.2                30.8                 22.0
## 27                           4.8                28.1                 23.5
## 28                           4.8                26.3                 23.1
## 29                           8.0                28.7                 25.1
## 30                           5.1                28.0                 17.9
## 33                           7.3                26.3                 15.4
## 35                           4.3                26.4                 22.4
## 36                           7.0                33.3                 20.2
## 37                           7.9                31.1                 23.3
## 38                           5.5                23.2                 25.2
## 39                           6.3                35.0                 15.8
## 40                           8.1                27.9                 12.2
## 41                           6.6                29.9                 16.9
## 42                           8.6                29.5                 20.3
## 43                           5.4                30.8                 22.0
## 44                           7.8                32.4                 20.8
## 45                           8.2                25.1                 21.7
## 46                           5.1                22.3                 25.7
## 47                           5.2                29.0                 16.8
## 48                           6.1                24.2                 19.0
## 49                           5.0                22.1                 23.6
```

```
## 51                          5.1               30.7                 20.3
## 52                          5.1               29.6                 25.3
##    Percent.Bachelors Population.Graduate Percent.Below.Poverty.Line
## 1               16.0                 9.6                       29.1
## 3               18.3                11.0                       24.6
## 6               26.0                15.2                       21.0
## 7               21.4                17.3                       20.1
## 9               23.9                33.4                       19.4
## 10              18.9                10.8                       24.9
## 12              21.7                11.2                       23.4
## 13              18.2                 8.5                       27.8
## 14              21.0                13.4                       23.4
## 15              17.0                 9.8                       26.5
## 17              21.2                12.6                       25.2
## 18              14.0                 9.9                       31.0
## 19              15.5                 8.3                       31.8
## 22              23.9                19.5                       21.2
## 23              17.6                11.5                       26.2
## 25              13.5                 8.3                       34.1
## 26              17.9                11.1                       25.2
## 27              21.7                10.6                       24.8
## 28              20.9                10.8                       21.8
## 29              16.5                 8.4                       23.1
## 30              22.6                14.3                       17.5
## 33              20.2                15.8                       24.5
## 35              21.8                 9.0                       22.7
## 36              17.3                10.6                       25.4
## 37              16.9                 8.6                       27.9
## 38              21.0                12.7                       25.8
## 39              18.9                12.5                       25.6
## 40              18.3                 7.4                       56.6
## 41              20.3                13.1                       22.7
## 42              17.6                10.4                       26.9
## 43              19.1                 9.0                       22.9
## 44              17.2                10.1                       26.5
## 45              19.3                10.3                       24.9
## 46              22.8                11.8                       26.4
## 47              22.5                15.8                       22.2
## 48              22.0                16.7                       22.7
## 49              22.2                13.3                       22.5
## 51              19.8                10.6                       23.2
## 52              17.4                10.3                       23.0
##    Median.Monthly.Housing.Costs
## 1                           734
## 3                          1015
## 6                          1300
## 7                          1390
## 9                          1641
## 10                         1047
## 12                         1585
## 13                          880
## 14                         1081
## 15                          815
## 17                          858
```

```
## 18                          743
## 19                          781
## 22                         1464
## 23                          862
## 25                          672
## 26                          837
## 27                          816
## 28                          887
## 29                         1089
## 30                         1280
## 33                         1291
## 35                          799
## 36                          833
## 37                          770
## 38                         1164
## 39                          947
## 40                          329
## 41                         1190
## 42                          820
## 43                          774
## 44                          820
## 45                         1009
## 46                         1122
## 47                         1088
## 48                         1237
## 49                         1319
## 51                          913
## 52                          890
```



We now compute RMSE values for the Random Forest model on both the training and the test (withheld) portion of the data set.

Model	RMSE.train	RSquared.train	RMSE.test	RSquared.test
Random Forest	1.732264	0.9750761	1.708613	0.8485353

Neural Network Model

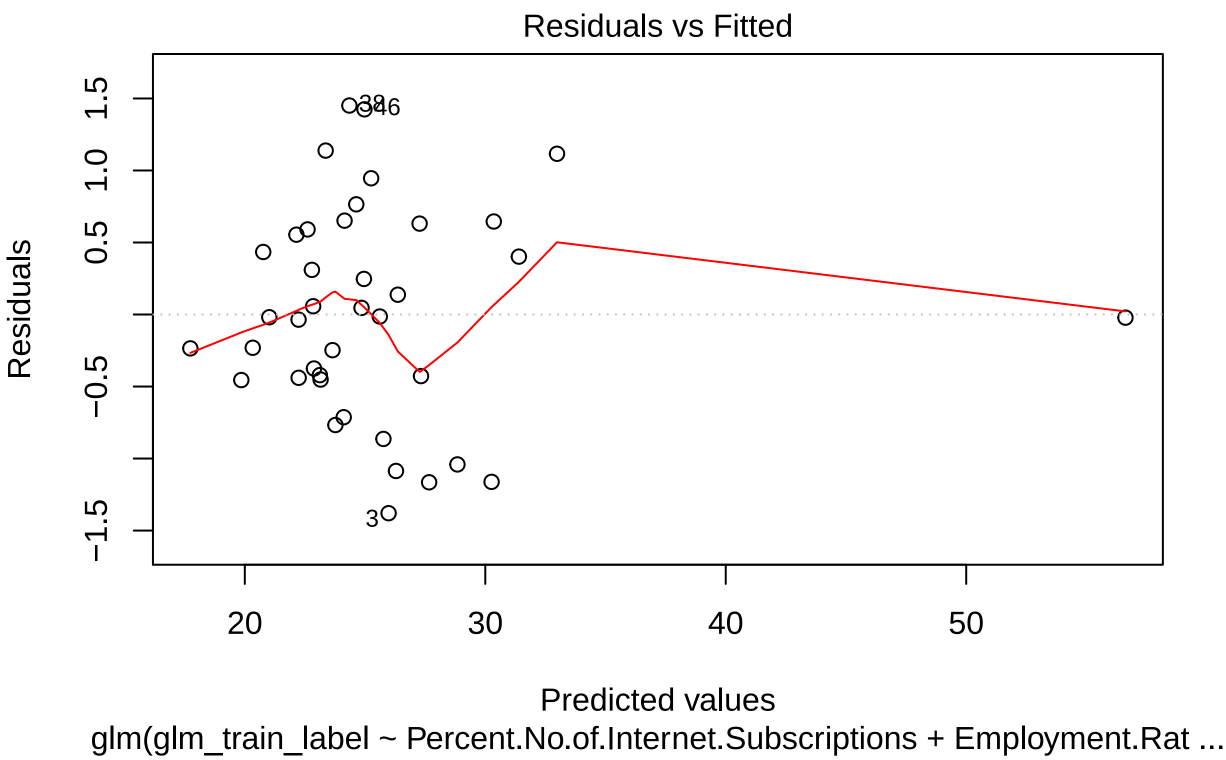
Neural Networks are a powerful nonlinear technique inspired by theories about how the human brain works [5]. Neural Networks can be classifiers (when the output variable is categorical) or regression (when the output variable is numeric). In this problem we use a regression artificial neural network (ANN) using the nnet package in R. Below we build and evaluate a Neural Network model of the regression problem.

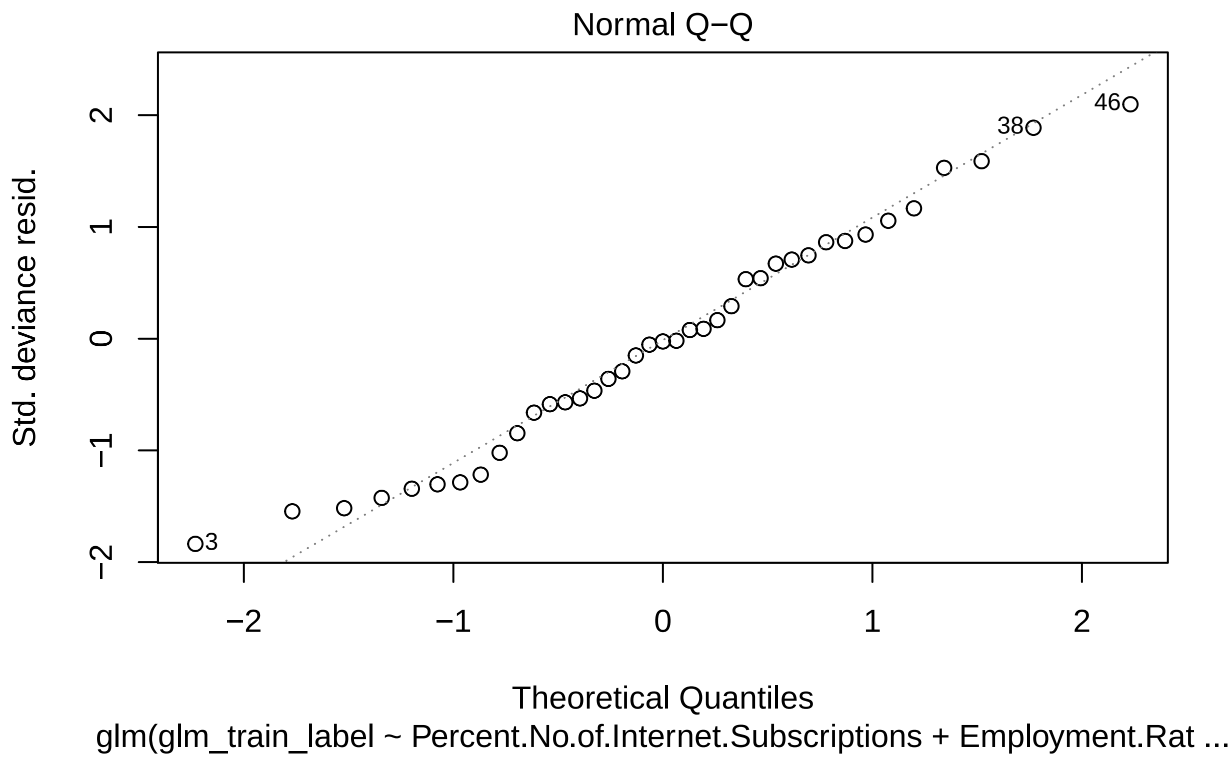
```
## a 27-4-1 network with 117 weights
## options were - linear output units  decay=0.01
```

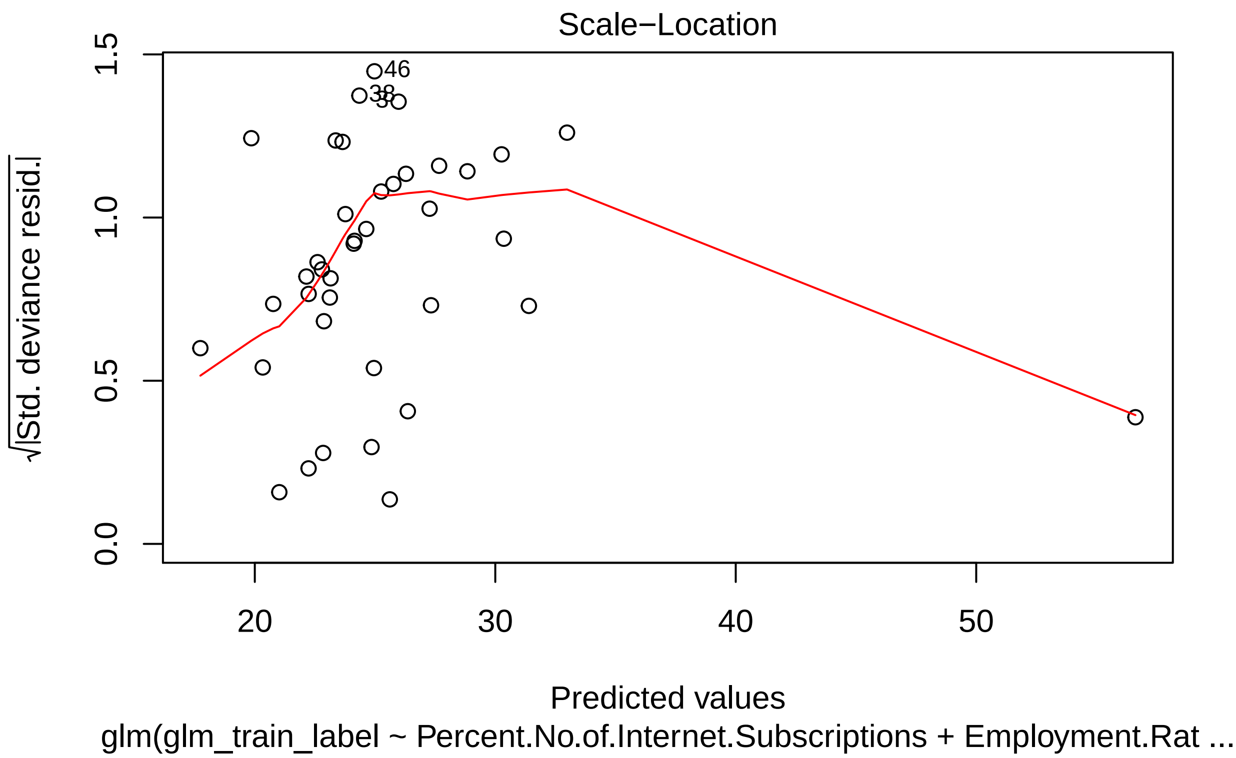
Model	RMSE.train	RSquared.train	RMSE.test	RSquared.test
NeuralNet	5.997393	0.2709523	3.939332	0.3219012

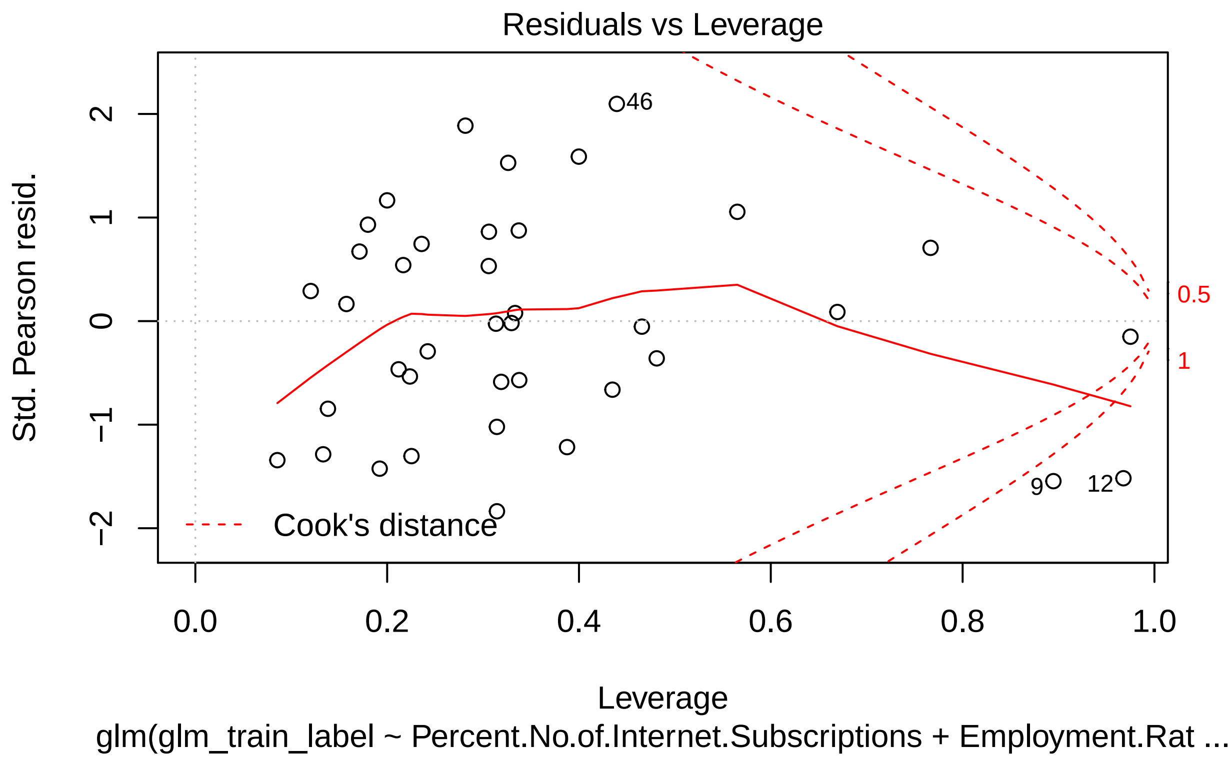
Generalized Linear Model

Because our output value is continuous and our data is numeric, we can use a generalized linear model to compute the pH. These models are generic and assume linearity in response. We will use the "Gaussian" type which assumes normally distributed variables




Normal Q−Q
Std. deviance resid.
Theoretical Quantiles
glm(glm_train_label ~ Percent.No.of.Internet.Subscriptions + Employment.Rat ...
−2
−1
0
1
2
3
38
46

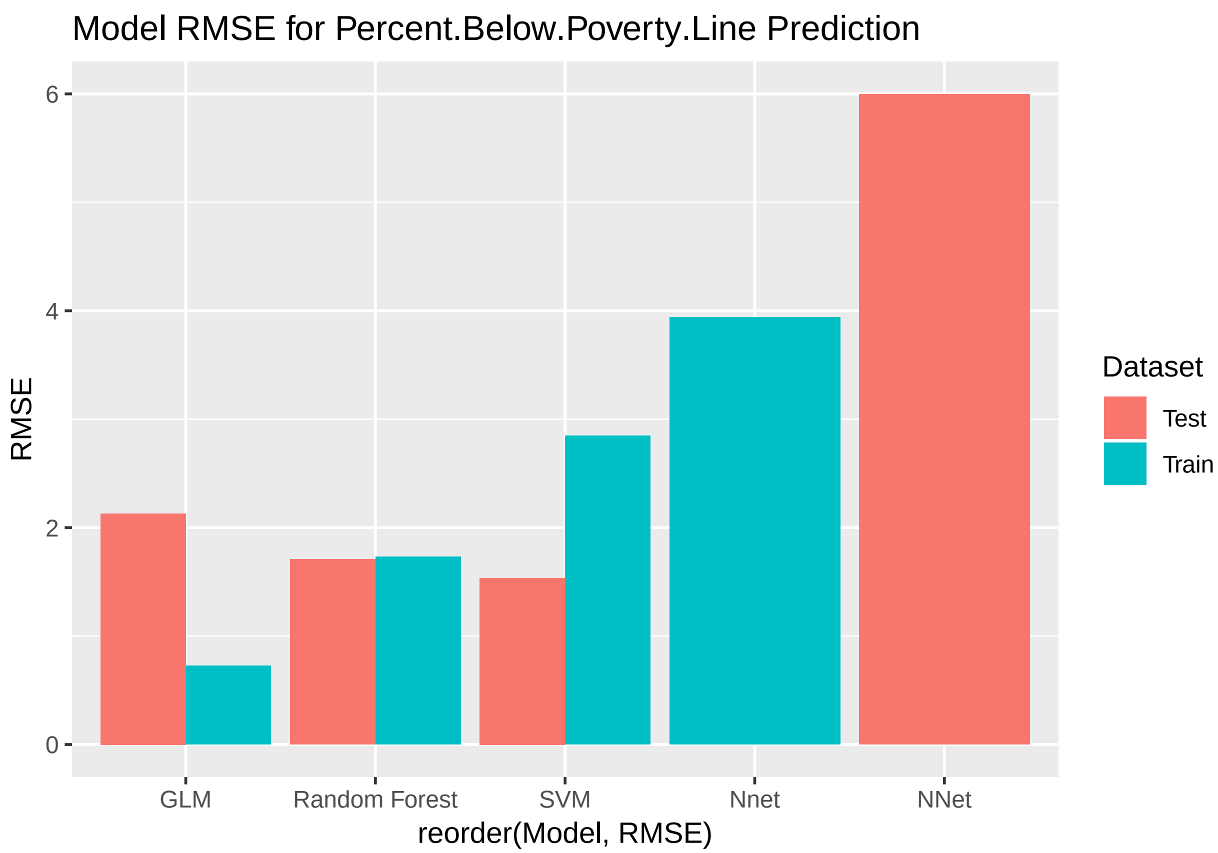

Scale−Location
√|Std. deviance resid.|
0.0
0.5
1.0
1.5
46
38
20
30
40
50
Predicted values
glm(glm_train_label ~ Percent.No.of.Internet.Subscriptions + Employment.Rat ...

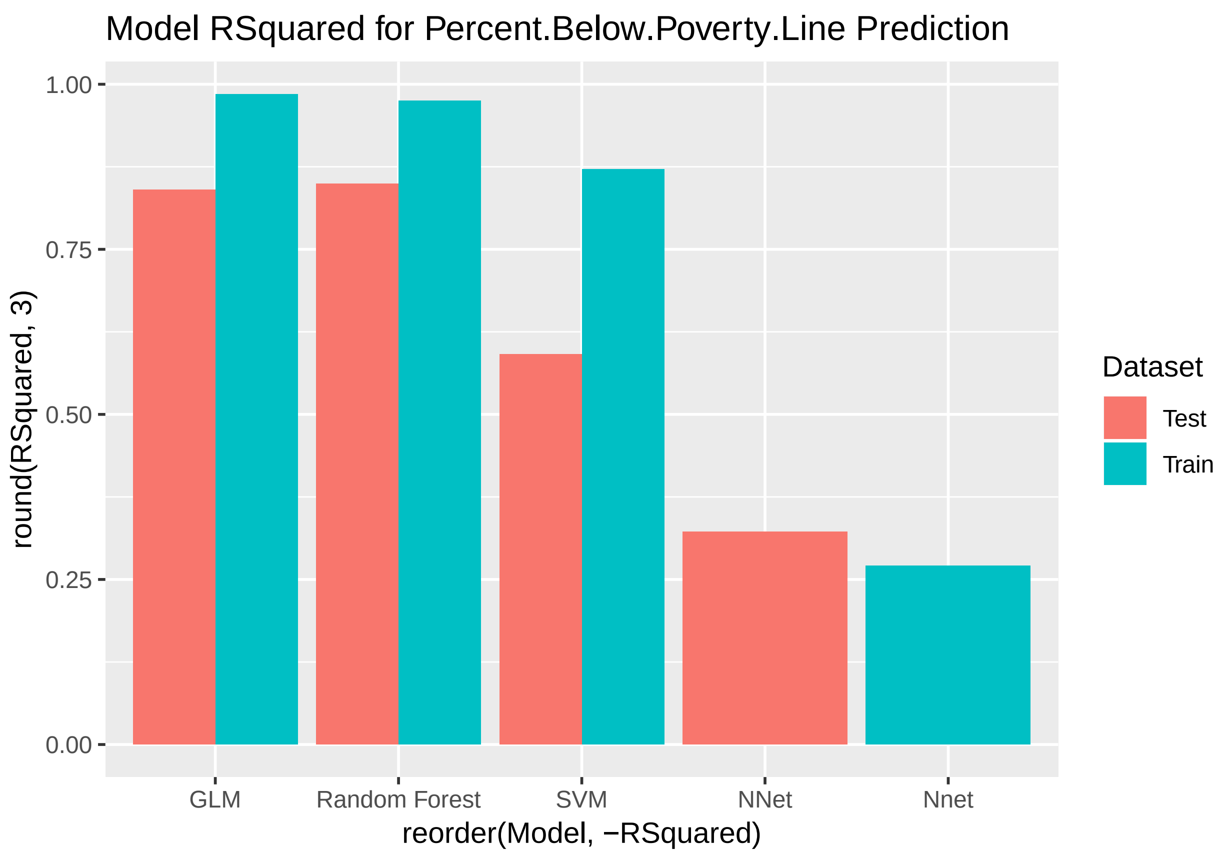


Model	RMSE.train	RSquared.train	RMSE.test	RSquared.test
GLM	0.7263871	0.9853306	2.130617	0.8400044

Conclusion: Social Indicators

Since we see that the Random Forest model produced the lowest RMSE on the withheld training data, we select it as the best model to predict poverty rates.


Model RMSE for Percent.Below.Poverty.Line Prediction
6
4
2
0
RMSE
GLM
Random Forest
SVM
Nnet
NNet
reorder(Model, RMSE)
Dataset
Test
Train



Finance Analysis

```
finances <- read.csv("Cleaned_Finance_Data.csv")

target <- finances$Poverty

finances$X <- NULL
finances <- as.data.frame(cbind(finances$Education,
finances$Hospitals,
finances$Health,
finances$Highways,
finances$Parks.and.recreation,
finances$Police.protection,
finances$Governmental.administration,
finances$Correction,
finances$Natural.resources,
finances$Internet,
finances$Smartphone,
finances$Computer,
finances$Public.welfare,
finances$Debt.at.end.of.fiscal.year))
```

```r
preProcValues <- preProcess(finances, method = c("center", "scale", "YeoJohnson", "nzv", "corr"))
finances <- predict(preProcValues, finances)
finances
```

```
##            V1           V2          V3          V4           V5
## 1   2.0154243  1.995293447 -0.57917326 -0.37607281 -1.318307233
## 2  -0.6921255 -1.031508069 -0.18987493  2.12013592 -0.888184785
## 3  -0.2281786 -1.121100315 -0.91147296 -0.42041083 -0.634179079
## 4   0.3695117  0.429702237 -1.24800727  0.30664543 -0.213462852
## 5  -0.6090109  0.046418106 -0.43794742 -1.46012825 -0.840047742
## 6   1.0948297 -0.099757697 -0.64000412 -0.13220260 -0.022085522
## 7  -0.9449053  0.339346291  0.53420051 -0.50121098 -0.816401521
## 8   1.0882925 -0.954692149  1.65400025  1.18724434  1.262037746
## 9  -0.4758444 -0.734535829  1.12231946  0.93244090 -0.821404045
## 10  0.8877946 -0.294161185  0.19703717 -0.35646583  0.944460422
## 11 -1.1561363  0.209917748  0.88553354 -1.45732010  1.285614034
## 12 -0.8107949 -0.991535512 -1.08969846 -0.34368061 -0.047387939
## 13 -1.1680236 -0.471854888 -0.10885577  0.73422467  0.049957961
## 14  1.3363925 -1.014935148 -0.93683674 -0.17084199 -1.043243411
## 15 -0.1424361  1.326203588 -1.17930795  1.13491365 -1.057064104
## 16  1.0940404  3.233237492 -0.46016990 -0.10652287 -0.682004052
## 17  0.8917389  0.826890681 -0.78441524 -0.05035365  0.466663592
## 18 -0.4304863 -0.786799781 -1.06869688 -0.85446239  3.767803412
## 19 -1.3144301 -0.923366312 -0.86712647  0.21723748 -1.344646827
## 20 -0.2344486 -0.702979219  1.92329024 -0.28642860  0.523166930
## 21 -1.4248763 -0.756198929 -0.16307032 -0.60511580  0.005501044
## 22  0.7007226  0.773483277 -0.30105140 -1.00477241 -0.619671955
## 23  0.3103140 -0.916802633 -1.24044877 -0.37397036  1.399275038
## 24  0.6429537  1.386832267 -0.25164778  0.52263555 -0.301089472
## 25 -0.5041425  0.805776799  2.25272898 -1.11860816 -0.728901260
## 26 -1.1536972 -0.876079299 -0.09488102  0.31266284 -0.645250840
## 27  0.5697854 -0.585889462  0.11453195  0.32947220  0.806983763
## 28 -0.8477898 -0.708391860 -0.92437777 -0.69055746 -1.141784412
## 29 -2.1085096 -0.897297885 -1.16078953 -0.15831497  0.001935987
## 30 -0.7722884 -0.146465831 -0.47360064 -0.91213658 -0.303944381
## 31 -0.3826473  0.741768456 -0.46124763 -1.31907565 -0.197397082
## 32 -1.1655841 -0.272670706  1.52207715 -1.46558175  0.168356395
## 33  0.5733264 -0.002122108 -0.44487985  0.27668494  0.099507823
## 34  0.2880454 -1.007856216  2.03904603  2.74466665  1.236357456
## 35 -0.8343945  0.380088980  0.04985016 -0.88463889 -0.628522046
## 36  0.4373856 -0.789910604  1.24055759  0.86308333  0.045735800
## 37 -0.1606630  0.662527138 -0.13871884 -1.49646860 -0.433248286
## 38 -0.1169696  0.592914237  0.78905788  0.80412199  0.159638481
## 39 -0.7738107 -0.877925567 -0.32377570 -1.13042803  2.024461038
## 40  0.2502619  0.746588843  1.07928378 -0.03654031  0.733306341
## 41 -0.5317806 -0.992683041  0.27111750  2.13379193  2.246217986
## 42  0.3071960 -0.723995446 -0.26216792 -0.61308289  0.053815477
## 43  0.8429188  0.388689759 -0.48839725  0.04272500 -0.824584142
## 44  1.5536902  1.904277355 -0.78565440 -0.29803304 -0.367445381
## 45  3.4744730  0.186611523  0.31131526  0.58172332  0.240693473
## 46 -0.1662693  1.684191745  0.04560163  0.81788535 -0.170889884
## 47  0.5259314  1.311587033  0.62430844 -0.95668449 -0.420411919
## 48 -0.3134040 -0.809046850 -0.48930655  0.43553752  0.171599852
```

```
## 49 -0.2696155  0.627256610 -1.02830510  0.96583177 -1.466679953
## 50  0.4782341 -1.109041073  2.87805031  2.11644612  0.285150073
##             V6          V7          V8          V9         V11
## 1  -0.59410123 -0.75791548 -0.55676076 -0.36266083 -0.85303176
## 2   1.85240734  3.01156244  1.86329439  3.33086939  1.93771191
## 3  -0.27152845 -0.91904627  0.86681716 -0.63953822  0.70239950
## 4  -0.74938377 -0.40041892 -0.76180983 -0.14406542 -0.63189715
## 5  -0.76385300 -0.61690532  0.23056415  0.29802169  1.70006912
## 6  -0.16191293 -0.03613239  1.83022425 -0.01145203  1.49508438
## 7  -0.02416576  1.33087677 -0.42685611 -0.66134506  0.10971128
## 8   2.47991901  2.11032334  2.62276913 -0.02527469  0.70239950
## 9  -0.58766026 -0.50016523  0.38304152  0.22282186  0.56538944
## 10  0.52853123 -1.09515514  0.59713237 -0.32947755  0.75758616
## 11 -1.36685013  0.71652390 -1.14751869 -0.25803744  0.64743167
## 12 -0.74545340  0.04883803  0.13394956  0.75950823 -0.22825814
## 13 -0.69910206 -1.08535935 -0.90380542 -1.18479151  0.34900434
## 14 -0.36154047 -0.89395061 -0.76116252 -0.41352967 -0.50721092
## 15 -1.02893021 -0.56361856 -1.79416684 -0.13617862 -0.65667686
## 16 -0.63935413 -0.58090104 -0.72608331 -0.19893456  0.03091581
## 17  0.23735205  0.56752107 -0.08596006 -0.04318048 -1.04605564
## 18  0.75169916 -0.22995646 -0.36366934  1.23222813 -0.53225333
## 19  0.54226287  1.08275720 -0.93336875  0.15984422 -1.67322376
## 20  1.86991568  0.23783861  2.08525490 -0.02822486  1.00864341
## 21  1.90759655  0.99508841 -1.15126513 -1.03075005  0.16251022
## 22 -0.32488038 -1.34289675  0.64579864 -0.94625246 -0.30497163
## 23  0.51575115 -0.66490872 -1.60111721  0.07728868  0.32220027
## 24  0.03895392 -0.73984395 -0.39767235 -0.02321569 -0.92580483
## 25 -0.42518496 -1.27216956  0.10817477 -0.56814623 -0.20258070
## 26 -0.90236136  1.70983186  0.61511179  2.45125644 -1.18865140
## 27  0.03265672 -0.39405805  1.63175286  0.81129955 -0.09933847
## 28 -0.92327007 -0.78526135 -1.18934130 -0.73392482  1.29283995
## 29  0.11698448  0.33836887 -0.86858647 -0.74278139  0.03091581
## 30  0.52177298 -0.58264851 -0.51731393 -0.76833128  0.86861734
## 31 -0.12120995 -0.35126565  0.81702959 -0.01823167 -0.99811070
## 32 -0.79353826 -0.22302396 -0.86108737 -1.21195346 -0.07339456
## 33  1.08896125 -0.39514043 -0.05448927 -0.53569999 -0.07339456
## 34 -1.01084604 -0.25695192 -1.30897297  0.74615913 -0.12522899
## 35 -1.03666320 -0.64878324 -0.40806310 -0.97150844 -0.48211583
## 36  0.35262943 -0.20897394  0.18685437 -0.78075641  0.16251022
## 37 -0.86774909  0.86822785 -0.07962493 -0.03510225  0.64743167
## 38  0.96137692 -0.30003835  0.15516900 -0.63507192 -1.21223713
## 39  0.62551370  1.33694707  0.25469614 -0.84846475 -0.30497163
## 40 -0.42404783 -0.93559998 -1.00071121 -0.50426302 -0.04739722
## 41  0.12247398  0.40595219 -0.08167443  2.35835385 -1.39907740
## 42  0.15768827 -0.12776947  1.38836117 -0.26780207 -0.48211583
## 43 -0.03143178 -1.25859696  1.05763510 -0.88252867  1.37918154
## 44 -0.20606572  1.35581764 -0.71134587 -0.33519424  2.33161337
## 45  3.37827778  0.75158468  1.09547739  1.09977781 -1.87406020
## 46  0.21002152  0.43597798  0.92981759 -0.87717955  0.72996546
## 47 -0.15397262 -0.98309541 -0.82563575  0.28147220  1.32156480
## 48 -0.71537867  0.06846889 -0.34529625  0.57980253 -2.55553952
## 49 -1.87422591 -0.64645517  0.51013990 -0.08845452 -0.73070176
## 50 -0.48808431  2.42449933 -0.14570658  2.83360010 -0.04739722
##            V12         V13         V14
```

```
## 1  -1.632990133  0.16697975 -0.78726521
## 2   1.498454015  0.04967626  0.72705235
## 3   0.725257929  1.13408690 -0.42789170
## 4  -1.555670525  0.47230536 -1.49297619
## 5   1.227835384  0.79072250 -0.11370728
## 6   1.498454015 -0.15100982  0.42880185
## 7   0.222680473 -1.82753251  2.08913427
## 8   0.995876559  0.17652525  0.37945612
## 9   0.609278516 -0.02981058 -0.82515543
## 10  0.222680473 -0.48896867 -1.12016955
## 11  0.377319690 -0.96577150  0.86696359
## 12  0.377319690 -0.95381859 -0.95368611
## 13 -0.009278353  0.14496921  1.20645470
## 14 -0.589175417  0.90323962  0.47677853
## 15 -0.318556787 -0.41884121 -1.11903977
## 16  0.145360864 -0.87522511 -0.27532700
## 17 -1.555670525  1.19963376  0.13425251
## 18 -1.671649938  0.73629662  0.45878342
## 19 -0.241237179  0.72280141  0.23545909
## 20  1.034536363  0.24291439  0.83423567
## 21  0.338659886  1.07810869  2.12384409
## 22 -0.125257766 -0.27975161  0.08236603
## 23  0.686598124  0.20706675 -0.57820823
## 24 -2.212887198  0.89890465 -0.22752550
## 25 -0.241237179 -0.40035316  0.30746652
## 26 -0.627835222 -0.21013417 -0.54816080
## 27  0.184020668 -0.37146466 -1.67192948
## 28  1.150515776 -1.12313659 -1.87007921
## 29  1.305154993  0.31446419  1.54114981
## 30  0.879897146 -0.27022332  1.55153681
## 31 -1.439691112  0.28392408 -0.41674187
## 32 -0.086597962  0.77229472  1.01366656
## 33 -0.357216592 -0.85260755 -1.02358326
## 34  0.068041256 -1.27672592 -0.20214919
## 35 -0.279896983  0.03198761 -0.39041950
## 36 -0.434536200 -0.08411378 -0.52849372
## 37  1.111855972 -0.32759027 -0.74973753
## 38 -0.898453852  0.65869411  0.28453876
## 39 -0.241237179  0.62138333  1.82768636
## 40 -0.589175417 -0.41675802  0.26184107
## 41 -0.743814635 -0.91169988  0.78884395
## 42 -1.130412678  1.34967250 -1.60198240
## 43  0.570618711 -0.31616946 -0.53172570
## 44  2.001031470 -1.22247846 -0.41397247
## 45 -0.318556787  4.39155127  0.82754974
## 46  0.570618711 -1.00769855  0.26601056
## 47  1.498454015 -0.84524193  0.60826476
## 48 -2.212887198  0.47946333  0.30136252
## 49 -0.318556787 -0.36081734  0.46539130
## 50  0.531958907 -1.83972360 -2.21896386
```

```
colnames(finances) <- c("Education", "Hospitals", "Health", "Highways", "Parks", "Police",  "Administra

finances$Poverty <- target
```

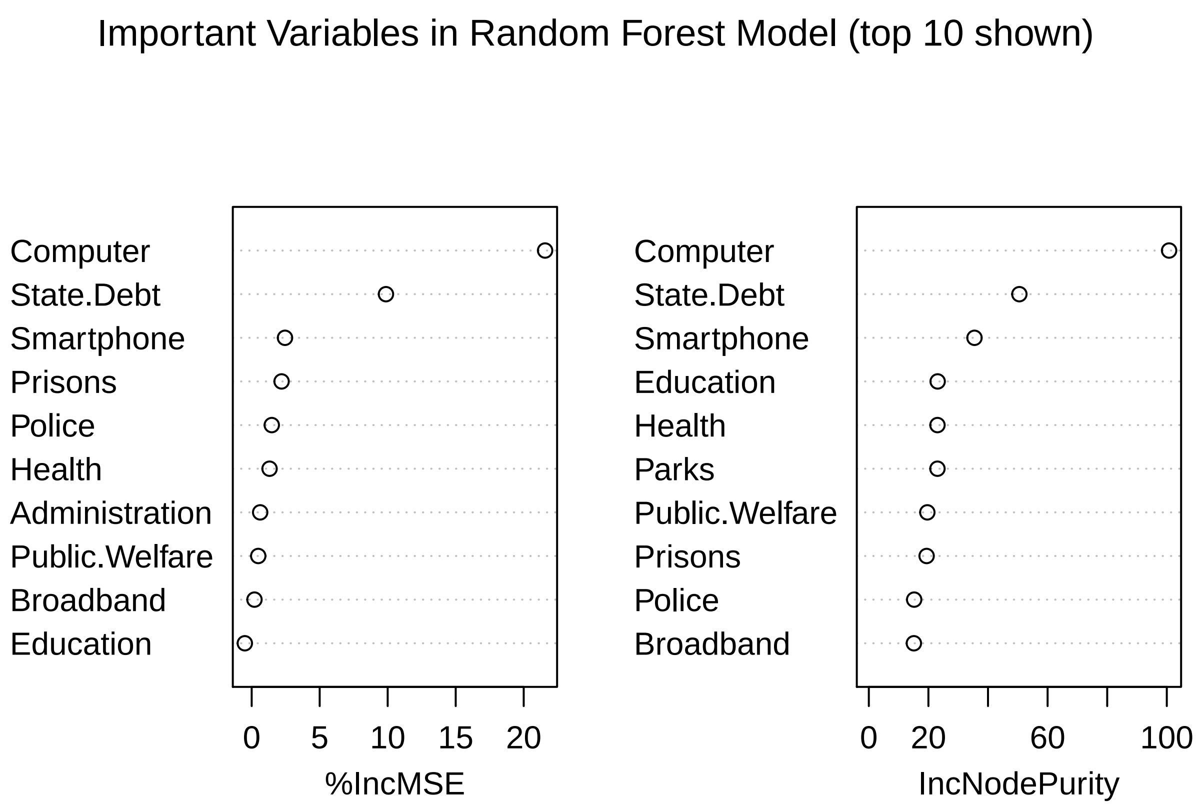
```r
finances
```

```
##     Education    Hospitals      Health    Highways        Parks
## 1   2.0154243  1.995293447 -0.57917326 -0.37607281 -1.318307233
## 2  -0.6921255 -1.031508069 -0.18987493  2.12013592 -0.888184785
## 3  -0.2281786 -1.121100315 -0.91147296 -0.42041083 -0.634179079
## 4   0.3695117  0.429702237 -1.24800727  0.30664543 -0.213462852
## 5  -0.6090109  0.046418106 -0.43794742 -1.46012825 -0.840047742
## 6   1.0948297 -0.099757697 -0.64000412 -0.13220260 -0.022085522
## 7  -0.9449053  0.339346291  0.53420051 -0.50121098 -0.816401521
## 8   1.0882925 -0.954692149  1.65400025  1.18724434  1.262037746
## 9  -0.4758444 -0.734535829  1.12231946  0.93244090 -0.821404045
## 10  0.8877946 -0.294161185  0.19703717 -0.35646583  0.944460422
## 11 -1.1561363  0.209917748  0.88553354 -1.45732010  1.285614034
## 12 -0.8107949 -0.991535512 -1.08969846 -0.34368061 -0.047387939
## 13 -1.1680236 -0.471854888 -0.10885577  0.73422467  0.049957961
## 14  1.3363925 -1.014935148 -0.93683674 -0.17084199 -1.043243411
## 15 -0.1424361  1.326203588 -1.17930795  1.13491365 -1.057064104
## 16  1.0940404  3.233237492 -0.46016990 -0.10652287 -0.682004052
## 17  0.8917389  0.826890681 -0.78441524 -0.05035365  0.466663592
## 18 -0.4304863 -0.786799781 -1.06869688 -0.85446239  3.767803412
## 19 -1.3144301 -0.923366312 -0.86712647  0.21723748 -1.344646827
## 20 -0.2344486 -0.702979219  1.92329024 -0.28642860  0.523166930
## 21 -1.4248763 -0.756198929 -0.16307032 -0.60511580  0.005501044
## 22  0.7007226  0.773483277 -0.30105140 -1.00477241 -0.619671955
## 23  0.3103140 -0.916802633 -1.24044877 -0.37397036  1.399275038
## 24  0.6429537  1.386832267 -0.25164778  0.52263555 -0.301089472
## 25 -0.5041425  0.805776799  2.25272898 -1.11860816 -0.728901260
## 26 -1.1536972 -0.876079299 -0.09488102  0.31266284 -0.645250840
## 27  0.5697854 -0.585889462  0.11453195  0.32947220  0.806983763
## 28 -0.8477898 -0.708391860 -0.92437777 -0.69055746 -1.141784412
## 29 -2.1085096 -0.897297885 -1.16078953 -0.15831497  0.001935987
## 30 -0.7722884 -0.146465831 -0.47360064 -0.91213658 -0.303944381
## 31 -0.3826473  0.741768456 -0.46124763 -1.31907565 -0.197397082
## 32 -1.1655841 -0.272670706  1.52207715 -1.46558175  0.168356395
## 33  0.5733264 -0.002122108 -0.44487985  0.27668494  0.099507823
## 34  0.2880454 -1.007856216  2.03904603  2.74466665  1.236357456
## 35 -0.8343945  0.380088980  0.04985016 -0.88463889 -0.628522046
## 36  0.4373856 -0.789910604  1.24055759  0.86308333  0.045735800
## 37 -0.1606630  0.662527138 -0.13871884 -1.49646860 -0.433248286
## 38 -0.1169696  0.592914237  0.78905788  0.80412199  0.159638481
## 39 -0.7738107 -0.877925567 -0.32377570 -1.13042803  2.024461038
## 40  0.2502619  0.746588843  1.07928378 -0.03654031  0.733306341
## 41 -0.5317806 -0.992683041  0.27111750  2.13379193  2.246217986
## 42  0.3071960 -0.723995446 -0.26216792 -0.61308289  0.053815477
## 43  0.8429188  0.388689759 -0.48839725  0.04272500 -0.824584142
## 44  1.5536902  1.904277355 -0.78565440 -0.29803304 -0.367445381
## 45  3.4744730  0.186611523  0.31131526  0.58172332  0.240693473
## 46 -0.1662693  1.684191745  0.04560163  0.81788535 -0.170889884
## 47  0.5259314  1.311587033  0.62430844 -0.95668449 -0.420411919
## 48 -0.3134040 -0.809046850 -0.48930655  0.43553752  0.171599852
## 49 -0.2696155  0.627256610 -1.02830510  0.96583177 -1.466679953
## 50  0.4782341 -1.109041073  2.87805031  2.11644612  0.285150073
```

```
##         Police Administration     Prisons   Broadband  Smartphone
## 1  -0.59410123    -0.75791548 -0.55676076 -0.36266083 -0.85303176
## 2   1.85240734     3.01156244  1.86329439  3.33086939  1.93771191
## 3  -0.27152845    -0.91904627  0.86681716 -0.63953822  0.70239950
## 4  -0.74938377    -0.40041892 -0.76180983 -0.14406542 -0.63189715
## 5  -0.76385300    -0.61690532  0.23056415  0.29802169  1.70006912
## 6  -0.16191293    -0.03613239  1.83022425 -0.01145203  1.49508438
## 7  -0.02416576     1.33087677 -0.42685611 -0.66134506  0.10971128
## 8   2.47991901     2.11032334  2.62276913 -0.02527469  0.70239950
## 9  -0.58766026    -0.50016523  0.38304152  0.22282186  0.56538944
## 10  0.52853123    -1.09515514  0.59713237 -0.32947755  0.75758616
## 11 -1.36685013     0.71652390 -1.14751869 -0.25803744  0.64743167
## 12 -0.74545340     0.04883803  0.13394956  0.75950823 -0.22825814
## 13 -0.69910206    -1.08535935 -0.90380542 -1.18479151  0.34900434
## 14 -0.36154047    -0.89395061 -0.76116252 -0.41352967 -0.50721092
## 15 -1.02893021    -0.56361856 -1.79416684 -0.13617862 -0.65667686
## 16 -0.63935413    -0.58090104 -0.72608331 -0.19893456  0.03091581
## 17  0.23735205     0.56752107 -0.08596006 -0.04318048 -1.04605564
## 18  0.75169916    -0.22995646 -0.36366934  1.23222813 -0.53225333
## 19  0.54226287     1.08275720 -0.93336875  0.15984422 -1.67322376
## 20  1.86991568     0.23783861  2.08525490 -0.02822486  1.00864341
## 21  1.90759655     0.99508841 -1.15126513 -1.03075005  0.16251022
## 22 -0.32488038    -1.34289675  0.64579864 -0.94625246 -0.30497163
## 23  0.51575115    -0.66490872 -1.60111721  0.07728868  0.32220027
## 24  0.03895392    -0.73984395 -0.39767235 -0.02321569 -0.92580483
## 25 -0.42518496    -1.27216956  0.10817477 -0.56814623 -0.20258070
## 26 -0.90236136     1.70983186  0.61511179  2.45125644 -1.18865140
## 27  0.03265672    -0.39405805  1.63175286  0.81129955 -0.09933847
## 28 -0.92327007    -0.78526135 -1.18934130 -0.73392482  1.29283995
## 29  0.11698448     0.33836887 -0.86858647 -0.74278139  0.03091581
## 30  0.52177298    -0.58264851 -0.51731393 -0.76833128  0.86861734
## 31 -0.12120995    -0.35126565  0.81702959 -0.01823167 -0.99811070
## 32 -0.79353826    -0.22302396 -0.86108737 -1.21195346 -0.07339456
## 33  1.08896125    -0.39514043 -0.05448927 -0.53569999 -0.07339456
## 34 -1.01084604    -0.25695192 -1.30897297  0.74615913 -0.12522899
## 35 -1.03666320    -0.64878324 -0.40806310 -0.97150844 -0.48211583
## 36  0.35262943    -0.20897394  0.18685437 -0.78075641  0.16251022
## 37 -0.86774909     0.86822785 -0.07962493 -0.03510225  0.64743167
## 38  0.96137692    -0.30003835  0.15516900 -0.63507192 -1.21223713
## 39  0.62551370     1.33694707  0.25469614 -0.84846475 -0.30497163
## 40 -0.42404783    -0.93559998 -1.00071121 -0.50426302 -0.04739722
## 41  0.12247398     0.40595219 -0.08167443  2.35835385 -1.39907740
## 42  0.15768827    -0.12776947  1.38836117 -0.26780207 -0.48211583
## 43 -0.03143178    -1.25859696  1.05763510 -0.88252867  1.37918154
## 44 -0.20606572     1.35581764 -0.71134587 -0.33519424  2.33161337
## 45  3.37827778     0.75158468  1.09547739  1.09977781 -1.87406020
## 46  0.21002152     0.43597798  0.92981759 -0.87717955  0.72996546
## 47 -0.15397262    -0.98309541 -0.82563575  0.28147220  1.32156480
## 48 -0.71537867     0.06846889 -0.34529625  0.57980253 -2.55553952
## 49 -1.87422591    -0.64645517  0.51013990 -0.08845452 -0.73070176
## 50 -0.48808431     2.42449933 -0.14570658  2.83360010 -0.04739722
##        Computer Public.Welfare  State.Debt Poverty
## 1  -1.632990133     0.16697975 -0.78726521    29.1
## 2   1.498454015     0.04967626  0.72705235    21.0
```

```
## 3   0.725257929    1.13408690 -0.42789170    24.6
## 4  -1.555670525    0.47230536 -1.49297619    30.7
## 5   1.227835384    0.79072250 -0.11370728    25.5
## 6   1.498454015   -0.15100982  0.42880185    21.0
## 7   0.222680473   -1.82753251  2.08913427    20.1
## 8   0.995876559    0.17652525  0.37945612    26.7
## 9   0.609278516   -0.02981058 -0.82515543    24.9
## 10  0.222680473   -0.48896867 -1.12016955    26.0
## 11  0.377319690   -0.96577150  0.86696359    23.4
## 12  0.377319690   -0.95381859 -0.95368611    27.8
## 13 -0.009278353    0.14496921  1.20645470    23.4
## 14 -0.589175417    0.90323962  0.47677853    26.5
## 15 -0.318556787   -0.41884121 -1.11903977    23.6
## 16  0.145360864   -0.87522511 -0.27532700    25.2
## 17 -1.555670525    1.19963376  0.13425251    31.0
## 18 -1.671649938    0.73629662  0.45878342    31.8
## 19 -0.241237179    0.72280141  0.23545909    25.0
## 20  1.034536363    0.24291439  0.83423567    20.1
## 21  0.338659886    1.07810869  2.12384409    21.2
## 22 -0.125257766   -0.27975161  0.08236603    26.2
## 23  0.686598124    0.20706675 -0.57820823    20.7
## 24 -2.212887198    0.89890465 -0.22752550    34.1
## 25 -0.241237179   -0.40035316  0.30746652    25.2
## 26 -0.627835222   -0.21013417 -0.54816080    24.8
## 27  0.184020668   -0.37146466 -1.67192948    21.8
## 28  1.150515776   -1.12313659 -1.87007921    23.1
## 29  1.305154993    0.31446419  1.54114981    17.5
## 30  0.879897146   -0.27022332  1.55153681    21.2
## 31 -1.439691112    0.28392408 -0.41674187    28.8
## 32 -0.086597962    0.77229472  1.01366656    24.5
## 33 -0.357216592   -0.85260755 -1.02358326    25.8
## 34  0.068041256   -1.27672592 -0.20214919    22.7
## 35 -0.279896983    0.03198761 -0.39041950    25.4
## 36 -0.434536200   -0.08411378 -0.52849372    27.9
## 37  1.111855972   -0.32759027 -0.74973753    25.8
## 38 -0.898453852    0.65869411  0.28453876    25.6
## 39 -0.241237179    0.62138333  1.82768636    22.7
## 40 -0.589175417   -0.41675802  0.26184107    26.9
## 41 -0.743814635   -0.91169988  0.78884395    22.9
## 42 -1.130412678    1.34967250 -1.60198240    26.5
## 43  0.570618711   -0.31616946 -0.53172570    24.9
## 44  2.001031470   -1.22247846 -0.41397247    26.4
## 45 -0.318556787    4.39155127  0.82754974    22.2
## 46  0.570618711   -1.00769855  0.26601056    22.7
## 47  1.498454015   -0.84524193  0.60826476    22.5
## 48 -2.212887198    0.47946333  0.30136252    33.7
## 49 -0.318556787   -0.36081734  0.46539130    23.2
## 50  0.531958907   -1.83972360 -2.21896386    23.0
```

```
## NULL
```




```
##      Education    Hospitals      Health    Highways        Parks
## 1    2.0154243  1.995293447 -0.57917326 -0.37607281 -1.318307233
## 3   -0.2281786 -1.121100315 -0.91147296 -0.42041083 -0.634179079
## 6    1.0948297 -0.099757697 -0.64000412 -0.13220260 -0.022085522
## 7   -0.9449053  0.339346291  0.53420051 -0.50121098 -0.816401521
## 9   -0.4758444 -0.734535829  1.12231946  0.93244090 -0.821404045
## 10   0.8877946 -0.294161185  0.19703717 -0.35646583  0.944460422
## 12 -0.8107949 -0.991535512 -1.08969846 -0.34368061 -0.047387939
## 13 -1.1680236 -0.471854888 -0.10885577  0.73422467  0.049957961
## 14  1.3363925 -1.014935148 -0.93683674 -0.17084199 -1.043243411
## 15 -0.1424361  1.326203588 -1.17930795  1.13491365 -1.057064104
## 17  0.8917389  0.826890681 -0.78441524 -0.05035365  0.466663592
## 18 -0.4304863 -0.786799781 -1.06869688 -0.85446239  3.767803412
## 19 -1.3144301 -0.923366312 -0.86712647  0.21723748 -1.344646827
## 22  0.7007226  0.773483277 -0.30105140 -1.00477241 -0.619671955
## 23  0.3103140 -0.916802633 -1.24044877 -0.37397036  1.399275038
## 25 -0.5041425  0.805776799  2.25272898 -1.11860816 -0.728901260
## 26 -1.1536972 -0.876079299 -0.09488102  0.31266284 -0.645250840
## 27  0.5697854 -0.585889462  0.11453195  0.32947220  0.806983763
## 28 -0.8477898 -0.708391860 -0.92437777 -0.69055746 -1.141784412
## 29 -2.1085096 -0.897297885 -1.16078953 -0.15831497  0.001935987
## 30 -0.7722884 -0.146465831 -0.47360064 -0.91213658 -0.303944381
## 33  0.5733264 -0.002122108 -0.44487985  0.27668494  0.099507823
## 35 -0.8343945  0.380088980  0.04985016 -0.88463889 -0.628522046
## 36  0.4373856 -0.789910604  1.24055759  0.86308333  0.045735800
## 37 -0.1606630  0.662527138 -0.13871884 -1.49646860 -0.433248286
```

```
## 38 -0.1169696  0.592914237  0.78905788  0.80412199  0.159638481
## 39 -0.7738107 -0.877925567 -0.32377570 -1.13042803  2.024461038
## 40  0.2502619  0.746588843  1.07928378 -0.03654031  0.733306341
## 41 -0.5317806 -0.992683041  0.27111750  2.13379193  2.246217986
## 42  0.3071960 -0.723995446 -0.26216792 -0.61308289  0.053815477
## 43  0.8429188  0.388689759 -0.48839725  0.04272500 -0.824584142
## 44  1.5536902  1.904277355 -0.78565440 -0.29803304 -0.367445381
## 45  3.4744730  0.186611523  0.31131526  0.58172332  0.240693473
## 46 -0.1662693  1.684191745  0.04560163  0.81788535 -0.170889884
## 47  0.5259314  1.311587033  0.62430844 -0.95668449 -0.420411919
## 48 -0.3134040 -0.809046850 -0.48930655  0.43553752  0.171599852
## 49 -0.2696155  0.627256610 -1.02830510  0.96583177 -1.466679953
##         Police Administration     Prisons   Broadband  Smartphone
## 1  -0.59410123    -0.75791548 -0.55676076 -0.36266083 -0.85303176
## 3  -0.27152845    -0.91904627  0.86681716 -0.63953822  0.70239950
## 6  -0.16191293    -0.03613239  1.83022425 -0.01145203  1.49508438
## 7  -0.02416576     1.33087677 -0.42685611 -0.66134506  0.10971128
## 9  -0.58766026    -0.50016523  0.38304152  0.22282186  0.56538944
## 10  0.52853123    -1.09515514  0.59713237 -0.32947755  0.75758616
## 12 -0.74545340     0.04883803  0.13394956  0.75950823 -0.22825814
## 13 -0.69910206    -1.08535935 -0.90380542 -1.18479151  0.34900434
## 14 -0.36154047    -0.89395061 -0.76116252 -0.41352967 -0.50721092
## 15 -1.02893021    -0.56361856 -1.79416684 -0.13617862 -0.65667686
## 17  0.23735205     0.56752107 -0.08596006 -0.04318048 -1.04605564
## 18  0.75169916    -0.22995646 -0.36366934  1.23222813 -0.53225333
## 19  0.54226287     1.08275720 -0.93336875  0.15984422 -1.67322376
## 22 -0.32488038    -1.34289675  0.64579864 -0.94625246 -0.30497163
## 23  0.51575115    -0.66490872 -1.60111721  0.07728868  0.32220027
## 25 -0.42518496    -1.27216956  0.10817477 -0.56814623 -0.20258070
## 26 -0.90236136     1.70983186  0.61511179  2.45125644 -1.18865140
## 27  0.03265672    -0.39405805  1.63175286  0.81129955 -0.09933847
## 28 -0.92327007    -0.78526135 -1.18934130 -0.73392482  1.29283995
## 29  0.11698448     0.33836887 -0.86858647 -0.74278139  0.03091581
## 30  0.52177298    -0.58264851 -0.51731393 -0.76833128  0.86861734
## 33  1.08896125    -0.39514043 -0.05448927 -0.53569999 -0.07339456
## 35 -1.03666320    -0.64878324 -0.40806310 -0.97150844 -0.48211583
## 36  0.35262943    -0.20897394  0.18685437 -0.78075641  0.16251022
## 37 -0.86774909     0.86822785 -0.07962493 -0.03510225  0.64743167
## 38  0.96137692    -0.30003835  0.15516900 -0.63507192 -1.21223713
## 39  0.62551370     1.33694707  0.25469614 -0.84846475 -0.30497163
## 40 -0.42404783    -0.93559998 -1.00071121 -0.50426302 -0.04739722
## 41  0.12247398     0.40595219 -0.08167443  2.35835385 -1.39907740
## 42  0.15768827    -0.12776947  1.38836117 -0.26780207 -0.48211583
## 43 -0.03143178    -1.25859696  1.05763510 -0.88252867  1.37918154
## 44 -0.20606572     1.35581764 -0.71134587 -0.33519424  2.33161337
## 45  3.37827778     0.75158468  1.09547739  1.09977781 -1.87406020
## 46  0.21002152     0.43597798  0.92981759 -0.87717955  0.72996546
## 47 -0.15397262    -0.98309541 -0.82563575  0.28147220  1.32156480
## 48 -0.71537867     0.06846889 -0.34529625  0.57980253 -2.55553952
## 49 -1.87422591    -0.64645517  0.51013990 -0.08845452 -0.73070176
##        Computer Public.Welfare  State.Debt Poverty
## 1  -1.632990133     0.16697975 -0.78726521    29.1
## 3   0.725257929     1.13408690 -0.42789170    24.6
## 6   1.498454015    -0.15100982  0.42880185    21.0
```

```
## 7   0.222680473    -1.82753251  2.08913427    20.1
## 9   0.609278516    -0.02981058 -0.82515543    24.9
## 10  0.222680473    -0.48896867 -1.12016955    26.0
## 12  0.377319690    -0.95381859 -0.95368611    27.8
## 13 -0.009278353     0.14496921  1.20645470    23.4
## 14 -0.589175417     0.90323962  0.47677853    26.5
## 15 -0.318556787    -0.41884121 -1.11903977    23.6
## 17 -1.555670525     1.19963376  0.13425251    31.0
## 18 -1.671649938     0.73629662  0.45878342    31.8
## 19 -0.241237179     0.72280141  0.23545909    25.0
## 22 -0.125257766    -0.27975161  0.08236603    26.2
## 23  0.686598124     0.20706675 -0.57820823    20.7
## 25 -0.241237179    -0.40035316  0.30746652    25.2
## 26 -0.627835222    -0.21013417 -0.54816080    24.8
## 27  0.184020668    -0.37146466 -1.67192948    21.8
## 28  1.150515776    -1.12313659 -1.87007921    23.1
## 29  1.305154993     0.31446419  1.54114981    17.5
## 30  0.879897146    -0.27022332  1.55153681    21.2
## 33 -0.357216592    -0.85260755 -1.02358326    25.8
## 35 -0.279896983     0.03198761 -0.39041950    25.4
## 36 -0.434536200    -0.08411378 -0.52849372    27.9
## 37  1.111855972    -0.32759027 -0.74973753    25.8
## 38 -0.898453852     0.65869411  0.28453876    25.6
## 39 -0.241237179     0.62138333  1.82768636    22.7
## 40 -0.589175417    -0.41675802  0.26184107    26.9
## 41 -0.743814635    -0.91169988  0.78884395    22.9
## 42 -1.130412678     1.34967250 -1.60198240    26.5
## 43  0.570618711    -0.31616946 -0.53172570    24.9
## 44  2.001031470    -1.22247846 -0.41397247    26.4
## 45 -0.318556787     4.39155127  0.82754974    22.2
## 46  0.570618711    -1.00769855  0.26601056    22.7
## 47  1.498454015    -0.84524193  0.60826476    22.5
## 48 -2.212887198     0.47946333  0.30136252    33.7
## 49 -0.318556787    -0.36081734  0.46539130    23.2
```

Model	RMSE.train	RSquared.train	RMSE.test	RSquared.test
Random Forest	1.083638	0.9656586	2.220804	0.8002246

Model	RMSE.train	RSquared.train	RMSE.test	RSquared.test
GLM	1.957209	0.6317318	2.005328	0.7493103

Next Steps

There are many ways we can continue improving the model performance, one method could be running more times of cross validation on more folds than 3 times 3-fold we have now for SVM and GBM models. It would take a longer time to compute, but the results would likely be better. Finally, more data would help us build a better model, in particular because the gap between the test and train sets tends to be relatively large across all of the models, with the exception of the Random Forest regressor because it builds its model over 1000 iterations.

We could also use time-series data to examine causal relationships between the data.

References

TO DO: APA format
1. Random Forests. https://uc-r.github.io/random_forests
2. Kevin Murphy (2012). Machine Learning a Probabilistic Perspective.
3. Support Vector Machine. https://uc-r.github.io/svm
4. Support Vector Machines. http://web.mit.edu/6.034/wwwbob/svm.pdf
5. Kuhn et al (2013). Applied Predictive Modeling