Energy Distribution in North Carolina

https:
//github.com/meganlundequam/LundequamVanasse_EDA_FinalProject

Megan Lundequam, Casey Slaught, Sam Vanasse

RMarkdown Cheat Sheet: https://www.rstudio.com/wp-content/uploads/2015/02/rmarkd
own-cheatsheet.pdf

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Awareness of environmental injustice has been prevalent in North Carolina for decades. In fact, Warren County, a county just north of Durham, is widely recognized as the birthplace of today's environmental justice movement(_______). In the late 1980s, the National Association for the Advancement of Colored People and others staged a protest against the state of North Carolina's decision to site a hazardous waste landfill in the predominantly colored county(_______). Although the protests failed to prevent the siting of the disposal facility, the event shed light on environmental injustice trends with the EPA describing the movement in its 1994 Environmental Equity Draft as "the watershed event that led to the environmental equity movement of the 1980's" (_______).

In North Carolina and across the US, a common expression of environmental injustice is the disproportionate biased siting of locally unwanted land uses sited next to low income and minority communities. An example of such a land use is energy production. The largest source of energy in the U.S. is fossil fuels and the burning of fossil fuels at power plants created emissions that can lead to respiratory and cardiovascular problems and increase the possibility of health issues ranging from cancer to immune system damage (_______). Minority, low-income, and indigenous populations frequently bear a disproportionate burden of these environmental harms and adverse health outcomes as a result of these siting trends (_______).

Due to a growing awareness of this trend and actions to combat environmental injustice across the U.S., it is necessary to examine the present day correlation of energy generators and minority, low income communities. This analysis explores the current (2020) distribution of utilities in relation to the distribution of low income communities to investigate to what extent this correlation still exists. This analysis further explores if that trend is true for different energy types. We were interested to see if there was any difference in the distribution of fossil fuel based utilities versus that of renewable energy sources in relation to minority and low income communities.

1.1 Research Questions:

This study examines the following key questions:

1.1.1 Question 1: Are energy production utilities in North Carolina located closer to low-income communities?

1.1.2 Question 2: Does the location of energy production utilities vary according to energy type? i.e., do different types of utilities have differing degrees of correlation with low income and minority communities?

2 Dataset Information

The data used for this analysis falls into two categories, energy related data and demographic data. We derived energy related data from the US Energy Information Association for the year 2020 (EIA-860 2020). The complete set of data files contains generator-level specific information for each year about existing and planned generators and associated environmental equipment at electric power plants with 1 megawatt or greater of combined nameplate capacity. The data sets utilized for this analysis include the plant-level data (2___Plant_Y2020.xlsx) and generator-level data (3_1_GeneratorY2020.xlsx) (where only data contained in the "operable" tab was analyzed). Geographic coordinates were obtained from the plant-level data set and combined with generator-level variables including Utility ID, Utility Name, Plant Code, Plant Name, State, County, Technology, Energy Source 1 (which represents the predominant energy that fuels the generator), and Nameplate Capacity (MW) (which represents the generator's maximum generation capacity in megawatts).

We derived demographic data from the US Census Bureau which we obtained through the tidycensus R package (API key = a8cad28557bae7c89aae6ea747549dd4816c6fbd) and specifically looked at the variables representing median household income, total population, the percent of the population with college degrees, the percent of the population at or below the poverty line, the percent of the population that is considered a 'minority' and the percent of the population that has received no schooling. For this analysis, we aggregated each of these variables by county.

3 Exploratory Analysis

In our exploratory analysis, we first examined the distribution of income levels across the state. Figure 1 shows there is a marked difference in median household income across counties in North Carolina.

Figure 1. Distribution of Median Family Income Across NC Counties

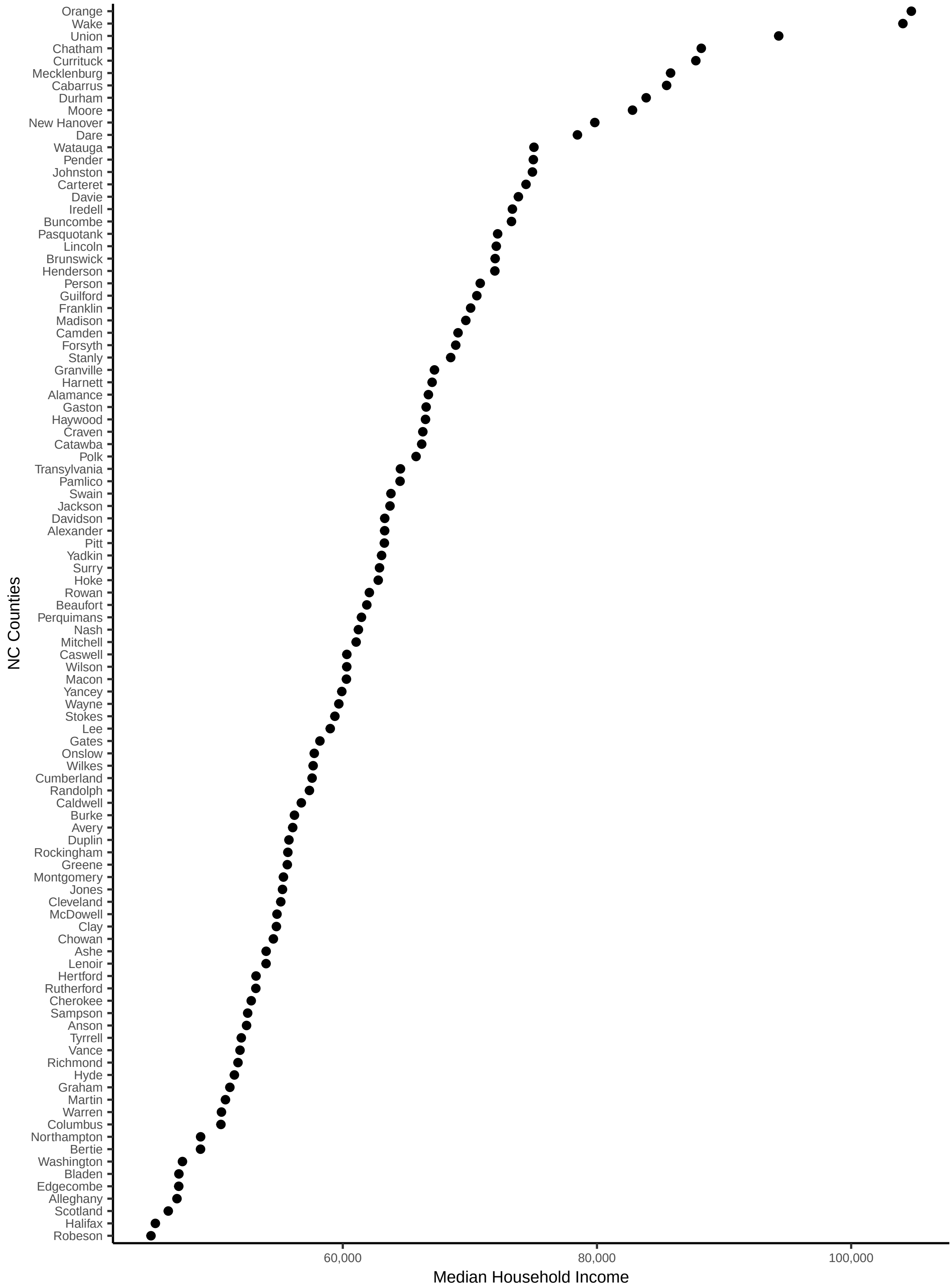


Figure 2 shows the different types of generators across North Carolina, defined by fuel type. The counties are coded based on the median family income in that county. Note that there does appear to be some overlap between generator location and lower income counties, however a clear pattern is not discernible based on this visual alone. Energy generators are diverse in type and widespread throughout the state. Also of note is that generators powered by solar appear to be the most abundant, despite widespread knowledge that the predominant energy sources in the US are fossil fuels.

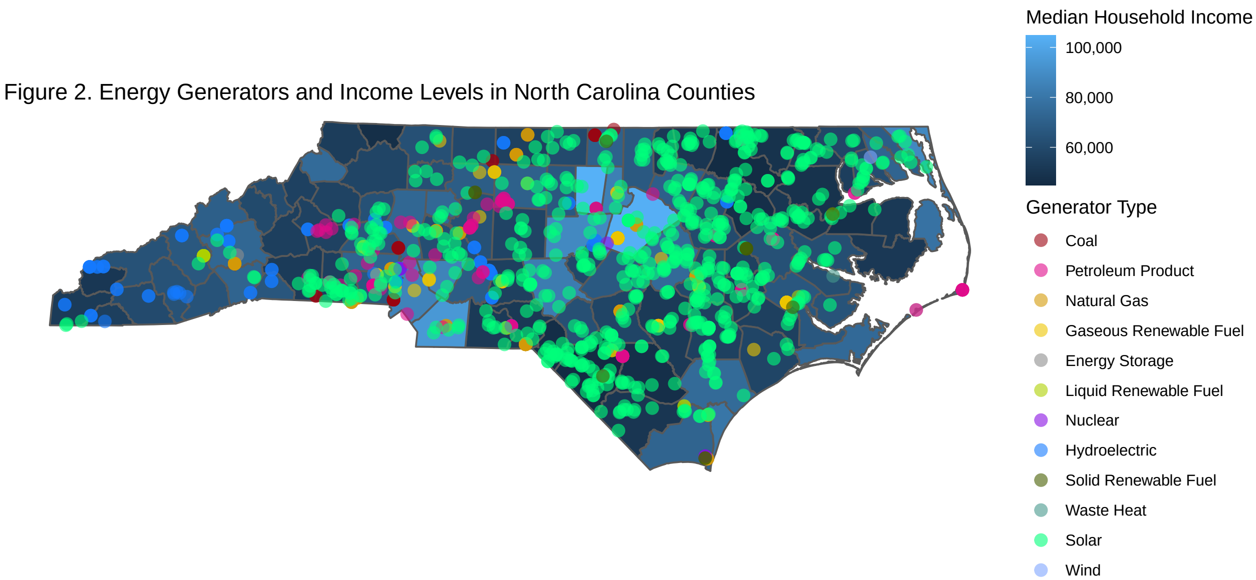


Figure 2. Energy Generators and Income Levels in North Carolina Counties

Figures 3 and 4 provide the context for interpreting the above figure. Figure 3 provides a breakdown of the number of individual generators based on fuel type, coded according to whether or not that fuel type is considered a renewable fuel, a fossil fuel, or other, for the state of North Carolina. While the renewable-fueled generators are far greater in number, Figure 4 shows that the majority of energy production comes from fossil fuel sources.

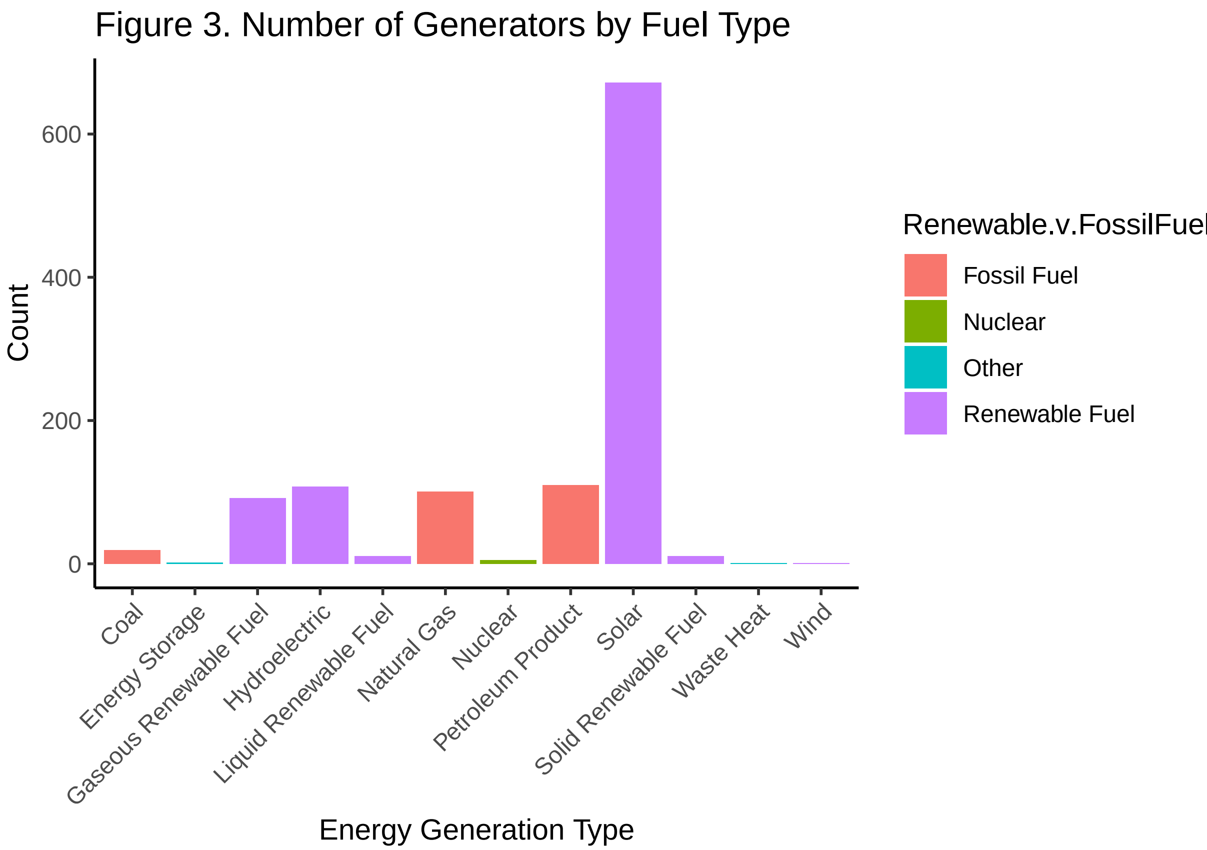


Figure 4 provides a visual representation of the aggregated energy production potential of the individual generators, using the nameplate capacity for each generator, again based on fuel type and coded according to renewable versus fossil fuel categorization. This shows that the largest contributor to energy production in North Carolina is natural gas, followed by coal.

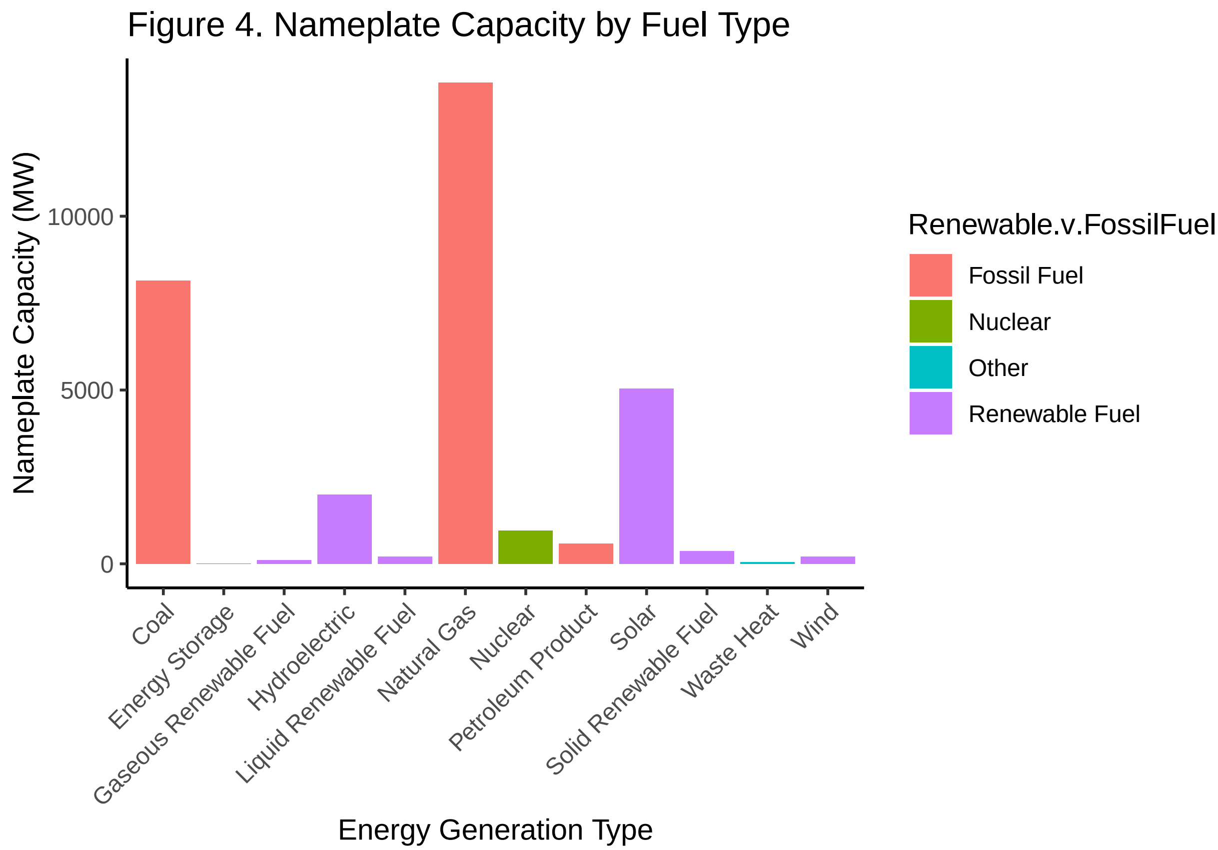


Figure 4. Nameplate Capacity by Fuel Type
Nameplate Capacity (MW)
10000
5000
0
Coal
Energy Storage
Gaseous Renewable Fuel
Hydroelectric
Liquid Renewable Fuel
Natural Gas
Nuclear
Petroleum Product
Solar
Solid Renewable Fuel
Waste Heat
Wind
Energy Generation Type
Renewable.v.FossilFuel
Fossil Fuel
Nuclear
Other
Renewable Fuel

4 Analysis

```
## Start:  AIC=144.69
## log(total.gen) ~ log(MedianHouseholdIncome) + Per.CollegeDeg +
##     Per.PovertyMobility + Per.MinorityMobility
##
##                             Df Sum of Sq    RSS    AIC
## - Per.CollegeDeg             1    0.0823 368.08 142.72
## - Per.PovertyMobility        1    1.2538 369.25 143.11
## - log(MedianHouseholdIncome)  1    2.3039 370.30 143.46
## - Per.MinorityMobility       1    4.2926 372.29 144.11
## <none>                                   367.99 144.69
##
## Step:  AIC=142.72
## log(total.gen) ~ log(MedianHouseholdIncome) + Per.PovertyMobility +
##     Per.MinorityMobility
##
##                             Df Sum of Sq    RSS    AIC
## - Per.PovertyMobility        1    1.1715 369.25 141.11
## - log(MedianHouseholdIncome)  1    2.4121 370.49 141.52
## - Per.MinorityMobility       1    4.2338 372.31 142.12
## <none>                                   368.08 142.72
##
## Step:  AIC=141.11
## log(total.gen) ~ log(MedianHouseholdIncome) + Per.MinorityMobility
##
##                             Df Sum of Sq    RSS    AIC
## - log(MedianHouseholdIncome)  1    1.3217 370.57 139.54
## <none>                                   369.25 141.11
## - Per.MinorityMobility       1   11.2341 380.48 142.76
##
## Step:  AIC=139.54
## log(total.gen) ~ Per.MinorityMobility
##
##                        Df Sum of Sq    RSS    AIC
## <none>                             370.57 139.54
## - Per.MinorityMobility  1    10.006 380.57 140.79

##
## Call:
## lm(formula = log(total.gen) ~ Per.MinorityMobility, data = All.Energy)
##
## Coefficients:
##          (Intercept)  Per.MinorityMobility
##                4.284                 1.782
```

```
## Start:  AIC=78.6
## log(energy.type.gen) ~ log(MedianHouseholdIncome) + Per.CollegeDeg +
##     Per.PovertyMobility + Per.MinorityMobility
##
##                              Df Sum of Sq    RSS    AIC
## - Per.MinorityMobility        1    2.3237 244.36 76.945
## - log(MedianHouseholdIncome)  1   10.3121 252.35 78.103
## - Per.PovertyMobility         1   10.9343 252.97 78.191
## - Per.CollegeDeg              1   11.9746 254.01 78.339
## <none>                                    242.04 78.601
##
## Step:  AIC=76.94
## log(energy.type.gen) ~ log(MedianHouseholdIncome) + Per.CollegeDeg +
##     Per.PovertyMobility
##
##                              Df Sum of Sq    RSS    AIC
## - log(MedianHouseholdIncome)  1    8.5517 252.91 76.183
## - Per.PovertyMobility         1    8.7072 253.07 76.205
## - Per.CollegeDeg              1   13.3764 257.74 76.863
## <none>                                    244.36 76.945
##
## Step:  AIC=76.18
## log(energy.type.gen) ~ Per.CollegeDeg + Per.PovertyMobility
##
##                       Df Sum of Sq    RSS    AIC
## - Per.PovertyMobility  1    0.5056 253.42 74.255
## - Per.CollegeDeg       1    5.4153 258.33 74.946
## <none>                             252.91 76.183
##
## Step:  AIC=74.25
## log(energy.type.gen) ~ Per.CollegeDeg
##
##                  Df Sum of Sq    RSS    AIC
## - Per.CollegeDeg  1     5.962 259.38 73.092
## <none>                        253.42 74.255
##
## Step:  AIC=73.09
## log(energy.type.gen) ~ 1

##
## Call:
## lm(formula = log(energy.type.gen) ~ 1, data = FossilFuel.Energy)
##
## Coefficients:
## (Intercept)
```

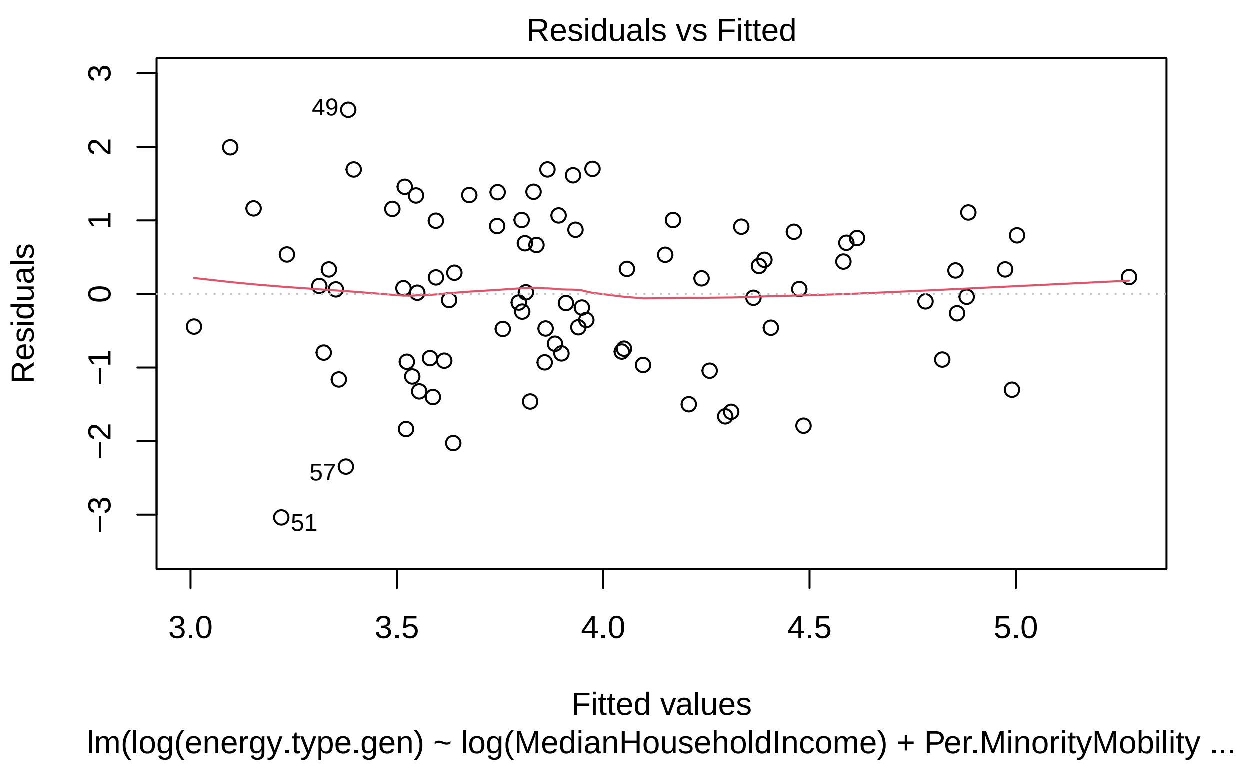
```
##      4.039

## Start:  AIC=21.31
## log(energy.type.gen) ~ log(MedianHouseholdIncome) + Per.CollegeDeg +
##     Per.PovertyMobility + Per.MinorityMobility
##
##                              Df Sum of Sq     RSS    AIC
## - Per.CollegeDeg              1    0.0091  99.089 19.320
## - Per.PovertyMobility         1    1.1825 100.262 20.344
## <none>                                     99.080 21.311
## - log(MedianHouseholdIncome)  1    4.8122 103.892 23.438
## - Per.MinorityMobility        1    8.5580 107.638 26.519
##
## Step:  AIC=19.32
## log(energy.type.gen) ~ log(MedianHouseholdIncome) + Per.PovertyMobility +
##     Per.MinorityMobility
##
##                              Df Sum of Sq     RSS    AIC
## - Per.PovertyMobility         1    1.1987 100.288 18.366
## <none>                                     99.089 19.320
## - log(MedianHouseholdIncome)  1    5.4167 104.506 21.950
## - Per.MinorityMobility        1    8.5820 107.671 24.546
##
## Step:  AIC=18.37
## log(energy.type.gen) ~ log(MedianHouseholdIncome) + Per.MinorityMobility
##
##                              Df Sum of Sq    RSS    AIC
## <none>                                    100.29 18.366
## - log(MedianHouseholdIncome)  1    5.6144 105.90 21.105
## - Per.MinorityMobility        1    7.9372 108.22 22.992

##
## Call:
## lm(formula = log(energy.type.gen) ~ log(MedianHouseholdIncome) +
##     Per.MinorityMobility, data = Renewable.Energy)
##
## Coefficients:
##                (Intercept)  log(MedianHouseholdIncome)
##                     19.981                      -1.506
##       Per.MinorityMobility
##                      1.972

##
## Call:
## lm(formula = log(energy.type.gen) ~ log(MedianHouseholdIncome) +
##     Per.MinorityMobility, data = Renewable.Energy)
```

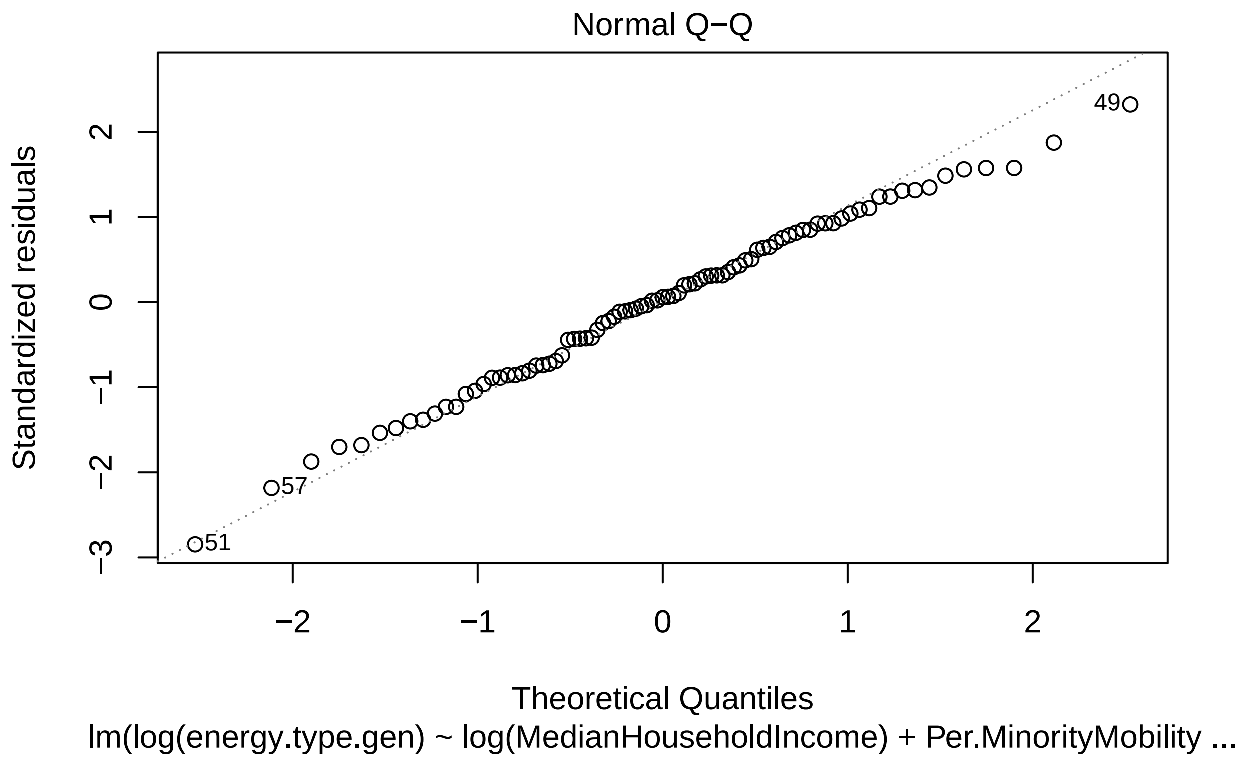
```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.03756 -0.80244  0.06227  0.82101  2.50399
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 19.9808     7.7644   2.573   0.0118 *
## log(MedianHouseholdIncome)  -1.5056     0.6943  -2.169   0.0329 *
## Per.MinorityMobility         1.9725     0.7650   2.578   0.0117 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.093 on 84 degrees of freedom
## Multiple R-squared:  0.1832, Adjusted R-squared:  0.1638
## F-statistic: 9.421 on 2 and 84 DF,  p-value: 0.0002034
```

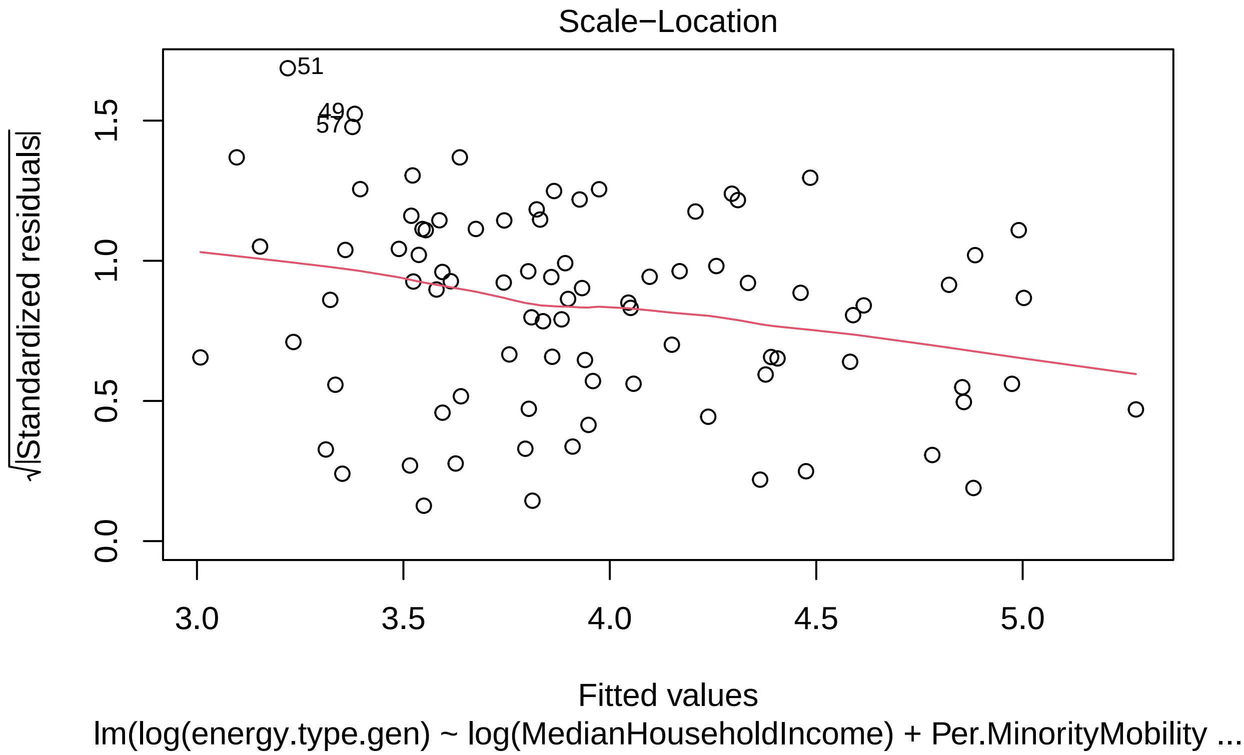
```
##
## Call:
## lm(formula = log(total.gen) ~ log(MedianHouseholdIncome) + Per.CollegeDeg +
##     Per.PovertyMobility + Per.MinorityMobility, data = All.Energy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2015 -1.2280 -0.2956  1.0463  3.2758
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -12.928     20.614  -0.627    0.532
## log(MedianHouseholdIncome)    1.469      1.716   0.856    0.394
## Per.CollegeDeg                1.348      8.333   0.162    0.872
## Per.PovertyMobility           4.992      7.907   0.631    0.529
## Per.MinorityMobility          1.605      1.374   1.168    0.245
##
## Residual standard error: 1.773 on 117 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.03306,    Adjusted R-squared:  4.998e-07
## F-statistic:     1 on 4 and 117 DF,  p-value: 0.4105
```

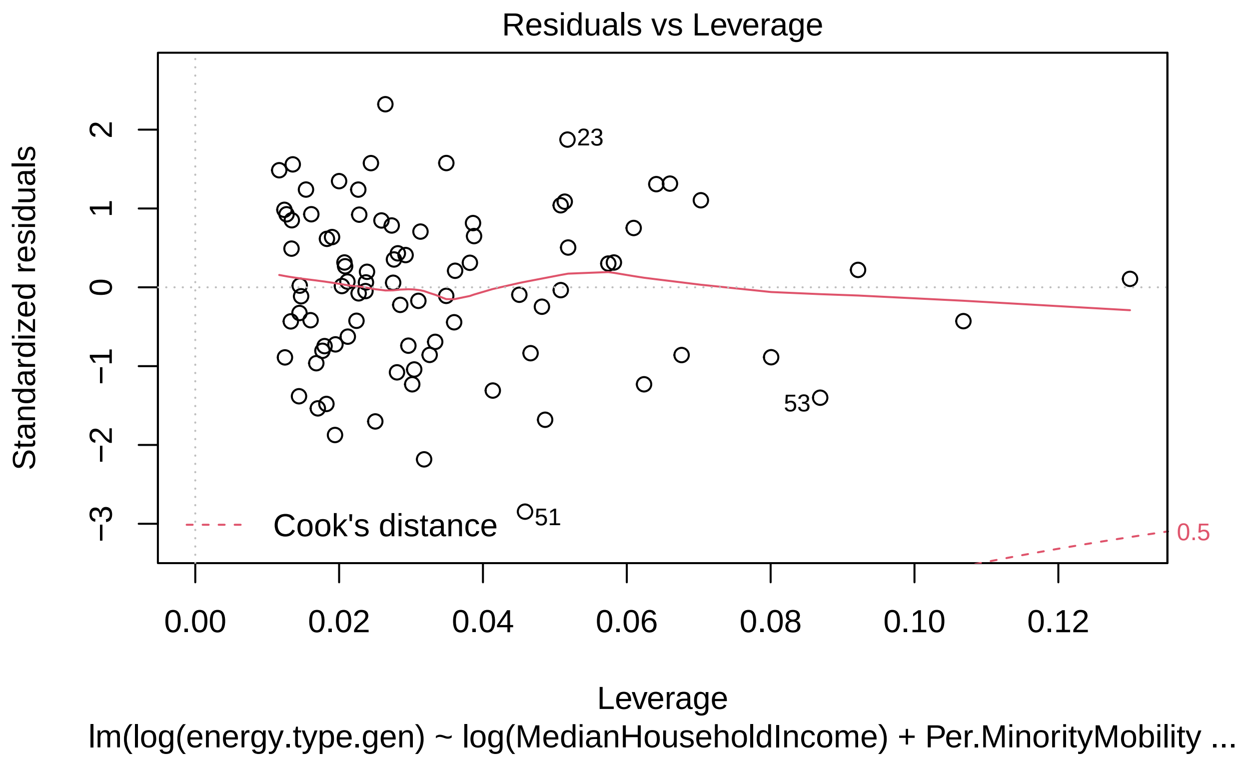
```
##
## Call:
## lm(formula = log(energy.type.gen) ~ log(MedianHouseholdIncome) +
##     Per.CollegeDeg + Per.PovertyMobility + Per.MinorityMobility,
##     data = FossilFuel.Energy)
##
## Residuals:
```

```
##      Min       1Q  Median       3Q      Max
## -4.0798 -2.3940 -0.4693  2.6875  4.1898
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -82.474     72.553  -1.137    0.264
## log(MedianHouseholdIncome)    6.862      5.971   1.149    0.259
## Per.CollegeDeg               34.832     28.126   1.238    0.225
## Per.PovertyMobility          31.882     26.941   1.183    0.246
## Per.MinorityMobility         -2.429      4.452  -0.546    0.589
##
## Residual standard error: 2.794 on 31 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.06686,    Adjusted R-squared:  -0.05354
## F-statistic: 0.5553 on 4 and 31 DF,  p-value: 0.6966
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(energy.type.gen) ~ log(MedianHouseholdIncome) + Per.MinorityMobility ...
49
57
51


Scale−Location
√|Standardized residuals|
Fitted values
lm(log(energy.type.gen) ~ log(MedianHouseholdIncome) + Per.MinorityMobility ...
51
49
57



For this analysis we used the generation site's nameplate capacity as the independent variable demonstrating the size, and assumed impact, to a community. We aggregated these values into three broad categories: fossil fuel, renewable, and total energy generation in each county. We then combined this data with demographic data from the US Census Bureau for the year 2020. These dependent variables include the county's median income, the percent of the county's total population that have received a college or associates degree, and the mobility of the population based on race and poverty status.

Given that both total energy generation within a county and their median income are positively skewed, we used log transformation on those variables. The remaining dependent variables were not log transformed as we already converted them to percentages of each county's total population.

To help narrow the focus of our analysis we utilized Akaike's Information Criterion (AIC) to select our variables for each regression. This exploratory analysis demonstrated that, given our data limitations, focusing on renewable data and using our minority mobility and median income variables will offer the best analysis (AIC=21.31).

The regression analysis (Adj. R-squared = 0.1638) demonstrates with a 1% increase in a county's renewable energy infrastructure, the county's median income decreases by 1.51% (p-value = 0.0329) and the ratio of population that is a minority increases by 1.97% (p-value = 0.0117). This is in line with what we hypothesized, because although renewable energy is often thought of as a luxury resource, the infrastructure still undergoes NIMBY criticism by many wealthier individuals.

5 Summary and Conclusions

Based on this

6 References

<add references here if relevant, otherwise delete this section>