Hate Crimes in the Age of Trump: An Exercise in Fact-Checking

Pawin Jiravanon, Brooke McWherter,
Varad Satam, Roberta Weiner

STAT 512

Introduction

According to William Gamson (1989), facts "have no intrinsic meaning…they take on meaning by being embedded in a frame or storyline that organizes them and gives them coherence, selecting certain ones to emphasize while ignoring others." By prioritizing one storyline over another (in other words by choosing a *frame*, or field of meaning), communicators inherently shape news consumers' interpretation of data. To frame necessarily means "setting a specific train of thought in motion," (Nisbet 2015) in order to meet some communication goal, which often means persuading consumers to behave in a certain way (such as encouraging critical evaluation of our engagement with a given risk or problem) or influencing news consumers' acceptance of various cultural or social narratives. While framing is essentially unavoidable in the current media landscape (due to the fact that news is nearly always presented as a story), it can also be problematic or misleading. In some cases, presenting a fact through radically different frames can contribute to very different interpretations of the same material (Tversky & Kahneman 1981).

Another challenge facing consumers of news today is the fact that the source and volume of information to which they are exposed have undergone a dramatic shift. Since the advent of social media platforms such as Facebook and Twitter, there has been an explosion in the number of news sources to which consumers of media have access. Unfortunately, a large number of these new sources of news are unverified by official channels, and may present false or misleading information without any real institutional or professional oversight, either to generate viewer traffic, or to (overtly or covertly) promote a specific ideology via a biased presentation of information (Allcot & Gentzkow 2017) . So, as a whole, news consumers must be more wary of not only the intentions of their news sources but also the veracity of the information presented. This contributes to an increasing sense of cynicism among consumers of news, and an erosion of faith in not only new media but also in *conventional* news sources, even though those sources were previously relied upon as 'watchdogs' protecting the public from government and corporate abuses" (Marchi, 2012).

Today, savvy consumers of news must be their own "watchdogs," protecting themselves from implausible or misleading claims and misinformation promoted as factual. Fact-checking and do-it-yourself analysis of the data presented by any sort of source helps prevent the spread of misinformation and is also empowering to the analysts. Some emerging news sources, however, have endeavored to address the problem by adding a layer of transparency to their presentation of news, in order to renew faith in their messages' credibility. For example, fivethirtyeight.com, a news platform produced by statistician Nate Silver, focuses on analyses of raw data, then presents several plausible storylines about a given social problem or issue, and provides the dataset from which the storylines were drawn. This allows readers to perform their own analysis, and to choose which storyline presented they believe to be most correct or applicable. However, the assumptions for the analysis performed are generally vague and not explicitly noted in the public-facing news articles. Other atypical news sources, such as the Southern Poverty Law Center, have clearly persuasive communication goals. The Southern Poverty Law Center states their purpose as "fighting hate and bigotry and…seeking justice for the most vulnerable members of our society," and frequently present facts in the context of social justice or injustice narratives. While the persuasive or ideological intent is overtly stated, which makes intentional framing slightly more ethically acceptable, but warrants further investigation.

We ultimately decided to fact-check the claims from two contrasting articles. First, we selected a fivethirtyeight.org article, "Higher rate of hate crimes are tied to income inequality." (Majumder, 2017) For an alternate perspective, we analyzed the Southern Poverty Law Center's article "Rise in hate crimes tied to the 2016 presidential election." (Johnson, 2018). These articles present separate causal theories for the causes of an increase in hate crimes following the 2016 American presidential election. To add additional levels of richness to our analysis of these claims, we also decided to draw from Colin Woodard's idea of cultural regions, or "eleven rival regional cultures," regions defined by "contrasting ideals of the distinct European colonial cultures that first took root," (Woodard, 2018) in different regions of the United States. The eleven regions defined by Woodard were also grouped by similarity of the inhabitants' average worldview along dimensions such as moralism, individualism, and traditionalism (Elazar, 1984); as well as egalitarianism and hierarchy (Ellis 1993). We therefore used Woodard's regions, which offered the finest granularity of regional breakdowns, to act as a proxy for worldview, and added it to our analysis, in order to suss out whether worldview (a factor not explicitly addressed by either the SPLC or FiveThirtyEight articles) also had some correlation to the rate of hate crimes. Our dataset was comprised of three different datasets that were merged: one drawn from the fivethirtyeight.com repository (Mehta, 2016); one from the FBI's website (FBI, 2016); and the last from Woodard's book (Woodard, 2012). From the three sources from which we drew data, we tested three research questions.

Research Questions

First, checking the claims of the FiveThirtyEight and SPLC's claims we ask;

Q1*: Are high shares of Trump voters the single best predictor of high hate crime rates when compared to other socioeconomic factors?*

- H_0: There are no relationships between the share of Trump voters or socioeconomic factors and hate crime occurrence
- H_1: There is a relationship between share of Trump voters and hate crime occurrence, and it is the most significant.
- H_2: There is a relationship between share of Trump voters and hate crime occurrence, and but it is not the most significant.

Examining whether our proxy for worldview (cultural region) has some influence on the rate of hate crimes we ask;

Q2: *Are there differences between cultural regions and their hate crime rates?*

- H_0: No differences exist between the rates of hate crimes in different cultural regions.
- H_1: There is a difference among rates of hate crimes between different US cultural regions.

Sussing out the potentially overlapping sources of causation among the proposed explanatory variables we analyzed,

Q3: *What are the factors correlated with rate of Trump voters in a state following the 2016 elections.*

- H_0: No relationship exists between the rate of Trump voters and socioeconomic conditions in a state.
- H_1: Some relationship exists between the rate of Trump voters and the socioeconomic conditions in a state.

Methods

Given the various questions our research had, we first developed a data analysis plan to maintain consistency and then ran three tests to test for our hypotheses in our three research questions.

Data Analysis Plan (DAP)

Prior to any analyses being conducted a DAP was developed with a protocol for transformations and treatment of variables in the dataset. If outliers were found that overly skew the data (greater than 4 times the mean – as indicated significant by Bonferroni Test) and were representative of the data then it would be replaced with the next lowest score. All empty cells would be imputed with the mean of that variable to reduce variable loss. If nonlinearity was determined in the data we were to log transform the variable first before conducting any advanced transformations to reduce unnecessary data manipulation.

Software

For all analyses in this research, RStudio 1.1463 was used.

Data

The dataset contained 51 variables, a review of the data determined that one variable was not a state (“District of Columbia”) and was removed reducing our dataset to 50 variables.

Sources

The data used for this analysis was a result of three data sources that were merged by the shared variable “state” (n=50).

All socioeconomic data, the percent).age of trump supporters, and hate crime data were obtained from fivethirtyeight.com (https://github.com/fivethirtyeight/data/tree/master/hate-crimes).The dataset consists of socioeconomic indicators including poverty, the Gini index, share of non-white, two collections of hate crime one through the Southern Poverty Law Center (SPLC) and the other from the FBI and all data was organized by state (Appendix B, table1). All the data was collected within 2009-2015, with a majority collected in 2015. A division of the hate crime data into categories of hate crime motivations in 2016, was utilized, it included motivations of hate crimes including race, religion, sexuality, ethnicity, disability, and gender (FBI, 2016) (Appendix B, table2). The last two data sets were obtained from a publication on the cultural regions of the United States (Woodard, 2012) (Appendix 1, table3) and the US Census Bureau's definitions of geographic regions in the United States (Appendix B, table4). Both sets were merged with the primary data set from fivethirtyeight.com using the common variable 'state'.

Defining the Response Variable

The main predictor variable of interest was the average hate crimes per 100,000 people collected by the Southern Poverty Law Center. This variable was selected over the FBI collection of hate crimes, due to the SPLC's more holistic approach in collecting reports from both media and self-reports in contrast to the FBI's collection of voluntary submission by agencies (Majumder, 2017).

A Welch's Two-Sample T-test was conducted to understand the differences of these two variables with the prediction that based on our understanding of the sources the SPLC would contain a higher number of hate crime incidents than that of the FBI. Given the variable collection rates of the two variables, the SPLC values were multiplied to represent a yearly estimate of the data compared to the yearly measurement of the FBI. The T-Test confirmed our predictions that the two variables were significantly different (t=6.77, df=47.9, p < 1.66-08) and a mean comparison confirmed that SPLC contained higher rates of hate crimes on average than the FBI data (SPLC; M = 13.87, FBI, M = 2.37).

Further Exploratory Analyses

Using the missing value test by column in R empty cells were located within the data frame. They were then imputed and the empty cells filled with the mean of that column per the DAP protocol. Scatterplot Matrices were run on hate crime with the socioeconomic data, geographic regions and cultural regions (Appendix B, tables 5-7). Preliminary Analyses with the socioeconomic data indicated a few variables that were highly correlated within the matrix and some potential trends with the hate crime, reviewing the scatterplot matrices in relation to the response variable indicated a dispersion of data located at the low end of the scale, while some trends appeared both positive and negative and some linear and nonlinear (Appendix B, Figure 1). In the regional scatterplots, there appeared to be a few relationships present as well as potential outliers in the cultural region data, but no relationships presented themselves in the geographic region data (Appendix B, Figure 2).

Primary Analyses

Test 1: Predictors of Hate Crimes

Developing a Model

A multiple linear regression analysis was determined as the best test to be used to indicate the best predictors of hate crimes based on socioeconomic factors and trump voters. Interaction terms were not included initially due to the scope of the question. The first model for the analysis contained no transformations or removal of the variables of interest. Predictor variables included are found in Appendix B, table 1.

All assumption tests were run on the same model before transformations were conducted, afterward, the assumptions were re-run and at the end, models were selected through a stepwise AIC process (Appendix B, table 6).

Testing Assumptions of Model 1

```
    Model1<lm (hate~+metropop+noncit+shrnonwhite+poverty+trump+gini+unemployed+
education, data=finalproj_new)
```

Residuals were plotted to observe the presence of any outliers, the visual plot indicated observations 37, 23, and 45 as potential outliers (Appendix B, Figure 3). A Bonferroni test was conducted to look for these outliers, the test indicated a single observation as a potential outlier but the adjusted p-value was not significant (observation = 37, r-student = 3.13, $p < .003$, Bonferroni $p > .162$) and the observation was kept. Additionally, an influence plot using Cook's distance was conducted to determine the strength of various variables on the model; this confirmed the presence of 37 on the upper end and 45 on the lower end as outliers but they are under 4 times the mean, and so are kept.

A Breusch-Pagan test (NCV) test was run to test for heteroscedasticity ($X^2 = 10.53$, $p < .001$). The test indicated that heteroscedasticity was present and transformations were needed, prompting further tests. A variance inflation factor test was run to test multicollinearity, variables that exceeded 5 were considered highly correlated and removed since the high correlation indicated a redundancy in the data (James et al. 2014). Two variables met the standard and were removed for the second model iteration (education = 5.57 and medhhincome = 5.30).

Linearity was tested using Ceres plots, Ceres was selected, as it was less prone to leakage of non-linearity (Wetzel, 1995), the plots indicated potential nonlinearities (Appendix B, Figure 5). An Anderson Darling test (NIST, 2012) was then used to test the residuals for normality ($A = 0.25$, $p > .74$), the test was non-significant indicating that the residuals did not follow a normal distribution. Shapiro tests were then run on individual variables testing for linearity and log transformations were done on non-linear variables per the DAP protocol. The Shapiro tests indicated non-normality in the share of non-citizens ($p < .01$), poverty ($p < .03$), and hate crimes ($p < .001$).

Lastly, a Durbin Watson Test was used to test for auto correlated errors (independence) (Racine and Hyndman, 2002). The test was non-significant (Autocorrelation =0.03, D-W = 1.93, $p > .82$), indicating an independence of the errors.

Testing Model 2

```
Model2<lm(loghate2~+metropop+lognoncit2+shrnonwhite+logpoverty+trump+gini+unemployed
, data=finalproj_new)
```

A second model was created with the following changes; education and median household income were removed from the model, and variables hate crime, non-citizenship share, and poverty was log transformed for normality. The second model maintained the outlier as it did not meet assumptions for imputation post transformation of the variables (Appendix B, Figure 6), the NCV test was not significant ((X^2 = 3.41, p < .06), indicating a failure to reject the null of heteroscedasticity. The variance inflation factor test did not indicate any redundant values and the Anderson Darling Normality Test was significant (A = 1.25, p > .003), indicating that the model still failed normality. Additionally the Durbin Watson Test remained non-significant (Autocorrelation =0.02, D-W = 1.94, p > 0.91).

A BoxCox transformation was applied and an arcsine transformation was suggested for the regressor variable. The hate crime variable was then transformed to an arcsine and this was used in place of the log transform.

Testing Model 3

Model 3 met all assumptions for the linear regression model, the NCV test was significant (p < 6.366e-06), there was no multicollinearity, the outlier was still present, but did not meet standards from the DAP for removal, and a re-ran of the Anderson Darling Test was not significant indicating normality (A = 0.52, p > .18).

```
Model3<- lm((arcsine)hate3~+metropop+lognoncit2+shrnonwhite+logpoverty+
trump+gini+unemployed, data=finalproj_new)
```

Prior to final selection, a stepwise AIC was run to confirm the fit of model 2, the AIC indicated that the non-significant predictors could be removed without effect to the model's fit or model p value. (See Appendix B) and the lowest AIC model was selected. An ANOVA was ran to determine changes, the ANOVA test indicated that the reduced model did not result in a significant reduction of sum of square residuals and was selected for analysis (p > .69)

```
Model4 < - lm(hate3~shrnonwhite+trump)
```

Interactions in Test 1

While our hypothesis looked at what single best predictor predicted hate crimes, the exploratory nature of this research and an attempt to build a holistic understanding of the issue led to a separate interaction model to be developed that could contribute to the objectives of this paper. All interaction terms were applied to a new model with transformed variables maintained, a backward AIC was ran and the smallest AIC value model was selected. An ANOVA was ran comparing AIC model with a secondary model with removed insignificant interactions, the ANOVA was significant, the AIC model was kept (p >.38).

```
Interaction Model <- lm(hate3 ~ metropop + lognoncit2 + shrnonwhite + logpoverty + trump +
gini + unemployed + metropop:logpoverty +lognoncit2:trump + shrnonwhite:logpoverty +
shrnonwhite:trump + logpoverty:gini +logpoverty:unemployed + trump:unemployed + gini:unemployed)
```

Test 2: Hate Crime Occurrence by Region

Developing a Model

Given the categorical nature of the predictor variable, a one-way analysis of variance was run. The model was developed using the SPLC hate crime rates as the regressor variable and cultural region as the predictor variable.

Testing Assumptions of Model 1

```
Model1 <- aov(hatecrime2 ~ cultural_regions, data = data=finalproj_new)
```

Initial residuals plots indicated non-normal data. An Anderson Darling normality test was performed to test our visual assumption, the test was significant, indicating non-normal data (A=1.67, p<0.002).A Bonferonni test was run to check the significance of the potential outlier in the scatterplot, the test indicated that 37 was significant as an outlier (observations = 37, r-student = 4.08, p < 0.002, Bonferonni p = 0.009).

As a result, a log transform was conducted on hate crime and a second model was generated.

Testing Model 2

```
Model2 <- aov(loghatecrime2 ~ cultural_regions, data =   data=finalproj_new)
```

Assumption tests were re-run on the second model, the results indicated that the outlier was no longer significant (Bonferroni, observations=37, r-student = 2.583, p > 0.013, Bonferonni p = 0.638), normality could be assumed (Anderson Darling, A=0.29, p > 0.61), and the errors were independent (Durbin-Watson, Autocorrelation =0.02, D-W = 1.94, p > 0.80) (Appendix B, Figure 8). Model 2 met assumptions and was selected for analysis.

Test 3: Predictors of High Share of Trump voters using socio-economic data

Developing a Model

An initial model was generated with all potential socio-economic predictor variables from the fivethirtyeight.com dataset.

Testing Assumptions of Model 1

```
Model 1<-lm(trump~medhhincome+metropop+noncit2+shrnonwhite+education+
poverty+hatcrime+gini)
```

A residual plot was run to check for normality and errors (Appendix B, Figure 9. A dip in the residual line indicated a need for transformation and points indicated a presence of outliers. A Bonferonni test was run to check the significance of the potential outliers (observations = 45, r-student = -5.25, p < 5.27e-06, Bonferonni p < 0.0002), the test indicated that the outlier was significant.

A NCV test was ran to test for heteroscedasticity (X^2 = 0.08, p < 0.78). The test indicated that heteroscedasticity was present and transformations were needed, prompting further tests. A variance inflation factor test was run to test multicollinearity, variables that exceeded 5 were considered highly correlated and removed since the high correlation indicated a redundancy in the data (James et al. 2014). Only one variable met the standard and was removed for the second model iteration (education = 5.93). Given the similarities in the data and assumption checks from question 1, log transformations were applied to the same variables (hate crime, non-citizenship share, and poverty), Shapiro tests were used to check the assumption and results were the same as the variable tests in question 1.

Testing Model 2

Model 2<-trump ~ medhhincome + metropop + log(noncit) + shrnonwhite + log(poverty) + log(hatecrime) + gini

A residual plot was obtained to observe changes in the second model(Appendix B, Figure 10). The second model indicated no significant outliers, the NCV test was not significant (X^2 = 8.72e-06, $p < 1.0$), indicating a failure to reject the null of heteroscedasticity. The variance inflation factor test did not indicate any redundant values.

Model 2 met all assumptions for the linear regression model, a backward AIC step function was run to test the selection, current selected model was not determined to be the best fit, and the lowest AIC scored model was selected and ran for analysis(Appendix B, Table 6).

Final Model<- trump ~ medhhincome + metropop + shrnonwhite + log(poverty) +log(hatecrime)

Results

Overall, Trump voters were highly related to hate crime occurred as predicted, however the relationship was negative, however when viewed, however, worldview did not have a significant effect on the occurrence of hate crimes in the United States. Lastly, it was found that high shares of trump voters were significantly correlated with multiple socio-economic factors.

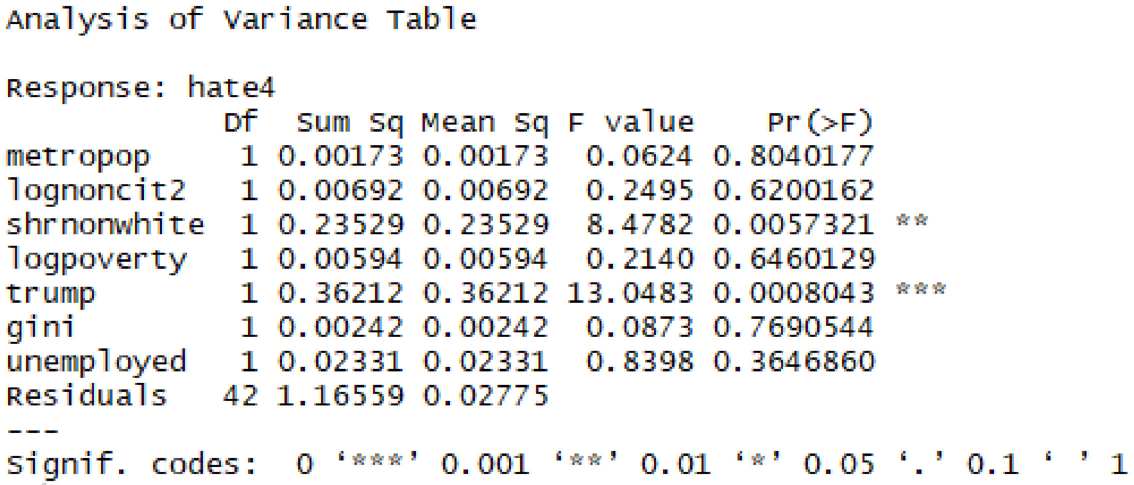
Predictors of Hate Crime

The final model used for the first analysis was determined after using a backward stepwise deletion (Table 7).

Iteration	Model	Justification
1	hatmod1 <- lm(hatecrime2 ~ medhhincome + metropop + noncit2 + shrnonwhite + education + poverty + trump + gini + unemployed)	Initial Model
2	hatmod2 <- lm(loghate2 ~ + metropop + lognoncit2 + shrnonwhite + logpoverty + trump + gini + unemployed, data = finalproj_new)	Post Variable Transformation for Linearity
3	hatmod3 <- lm(hate3 ~ + metropop + lognoncit2 + shrnonwhite + logpoverty + trump + gini + unemployed, data = finalproj_new)	Regressor Variable Transformation Re-selected, due to failure of normality
4	hatmod4 < - lm(hate3~shrnonwhite+trump)	Lowest AIC Score

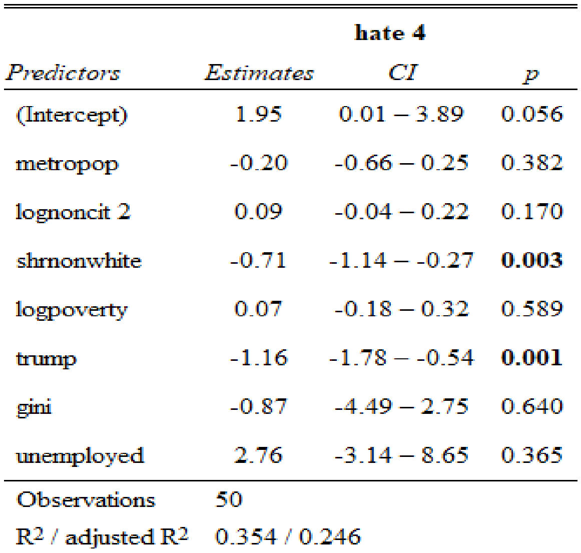
Table 7: Indicates all models used in determination of final model and justifications for changes. See Methods - Test1 for methodology, and Appendix B, Table 5 for AIC scores.

A multiple regression analysis (type 2) was used to test whether trump voters were significant predictors of hate crimes in relation to other socio-economic factors. The results of the regression indicated that two predictors explained 25% of the variance (Adjusted R2 = 0.25, F (7, 42) = 3.28, p < .007), a share of nonwhite population (F (1, 42) = 8.48, p <.006) and share of trump voters (F (1, 42) = 13.05, p < 0.0001) (table 8 and 9).



```
Analysis of Variance Table

Response: hate4
            Df  Sum Sq Mean Sq F value    Pr(>F)    
metropop     1 0.00173 0.00173  0.0624 0.8040177    
lognoncit2   1 0.00692 0.00692  0.2495 0.6200162    
shrnonwhite  1 0.23529 0.23529  8.4782 0.0057321 ** 
logpoverty   1 0.00594 0.00594  0.2140 0.6460129    
trump        1 0.36212 0.36212 13.0483 0.0008043 ***
gini         1 0.00242 0.00242  0.0873 0.7690544    
unemployed   1 0.02331 0.02331  0.8398 0.3646860    
Residuals   42 1.16559 0.02775                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

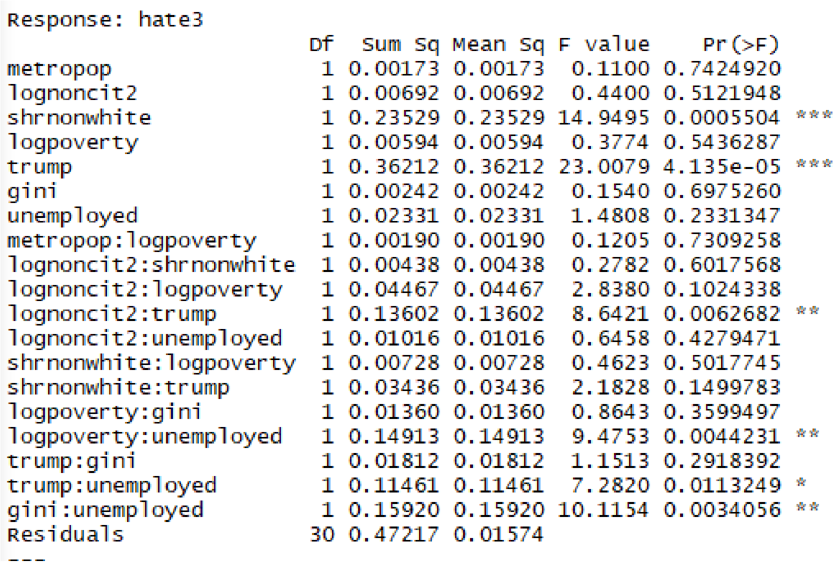
Table 8: R Type 2 ANOVA output table for the final regression model. Results indicate Trump as the most significant predictor variable.



	hate 4		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.95	0.01 – 3.89	0.056
metropop	-0.20	-0.66 – 0.25	0.382
lognoncit 2	0.09	-0.04 – 0.22	0.170
shrnonwhite	-0.71	-1.14 – -0.27	**0.003**
logpoverty	0.07	-0.18 – 0.32	0.589
trump	-1.16	-1.78 – -0.54	**0.001**
gini	-0.87	-4.49 – 2.75	0.640
unemployed	2.76	-3.14 – 8.65	0.365
Observations	50		
R^2 / adjusted R^2	0.354 / 0.246		

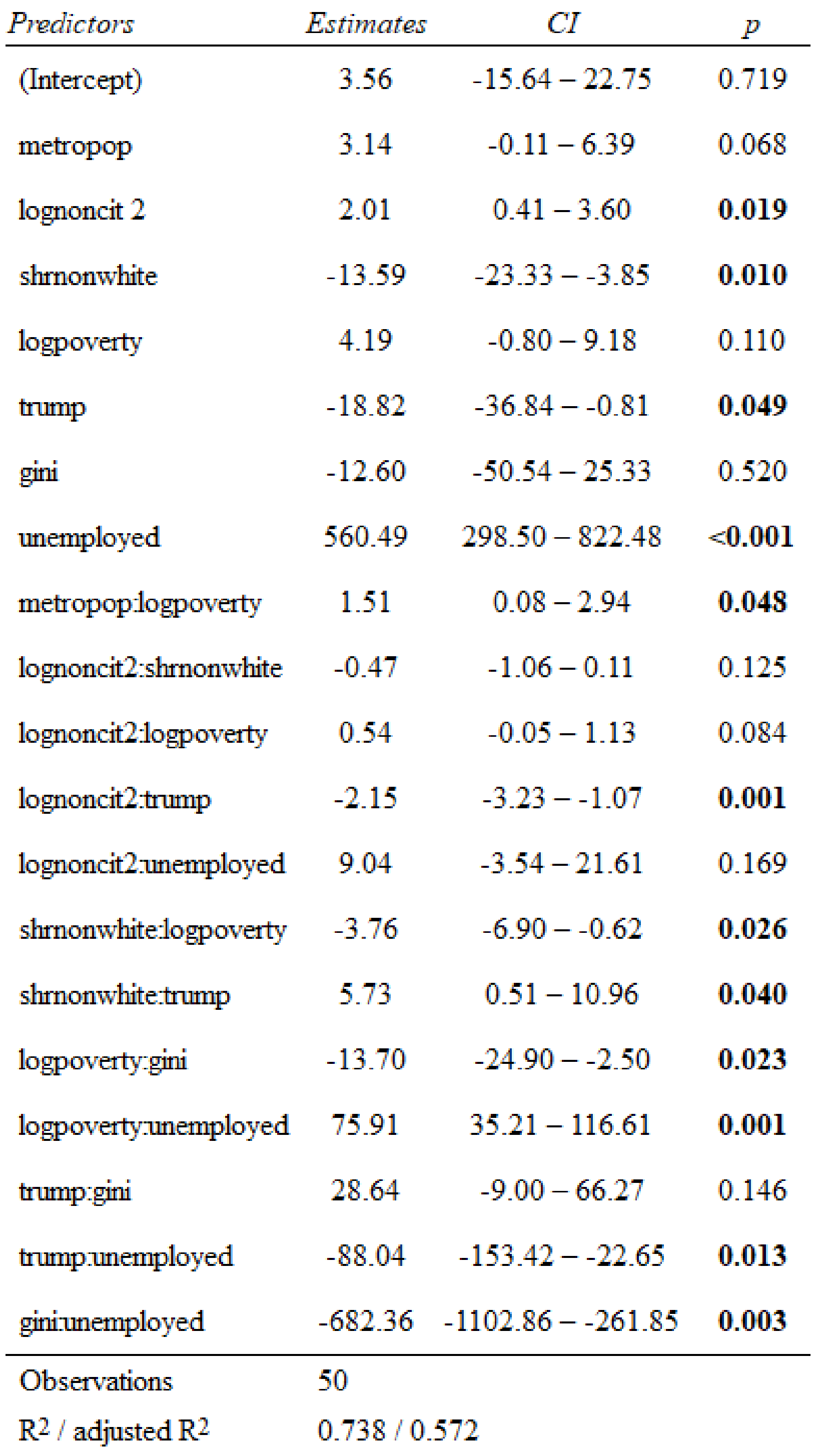
Table 9: Parameter estimates for final model of question 1. Trump is labeled as the most influential parameter.

While not in the main scope of the question, a separate interaction model was developed and linear regression (type 2) analysis was run and analyzed to test for potential interactions occurring within the model. The model was significant (F (1, 30) = 23, p < 4.135e-05) and the interaction model confirmed Trump as the main predictor, though as a negative predictor (Adjusted R2 = 0.57, F (15, 30) = 4.45, p < .0001). The results indicated that while Trump was maintained as the highest predictor it was closely followed by share of nonwhite participants (F(1,30)=14.95, p > 0.001) and in particular (log)noncitizens/trump(F(1,30) = 8.64, p < .006), and gini/unemployed (F(1,30) = 10.11, p < .003)(table 10 and 11).



```
Response: hate3
                        Df  Sum Sq Mean Sq F value    Pr(>F)    
metropop                 1 0.00173 0.00173  0.1100 0.7424920    
lognoncit2               1 0.00692 0.00692  0.4400 0.5121948    
shrnonwhite              1 0.23529 0.23529 14.9495 0.0005504 ***
logpoverty               1 0.00594 0.00594  0.3774 0.5436287    
trump                    1 0.36212 0.36212 23.0079 4.135e-05 ***
gini                     1 0.00242 0.00242  0.1540 0.6975260    
unemployed               1 0.02331 0.02331  1.4808 0.2331347    
metropop:logpoverty      1 0.00190 0.00190  0.1205 0.7309258    
lognoncit2:shrnonwhite   1 0.00438 0.00438  0.2782 0.6017568    
lognoncit2:logpoverty    1 0.04467 0.04467  2.8380 0.1024338    
lognoncit2:trump         1 0.13602 0.13602  8.6421 0.0062682 ** 
lognoncit2:unemployed    1 0.01016 0.01016  0.6458 0.4279471    
shrnonwhite:logpoverty   1 0.00728 0.00728  0.4623 0.5017745    
shrnonwhite:trump        1 0.03436 0.03436  2.1828 0.1499783    
logpoverty:gini          1 0.01360 0.01360  0.8643 0.3599497    
logpoverty:unemployed    1 0.14913 0.14913  9.4753 0.0044231 ** 
trump:gini               1 0.01812 0.01812  1.1513 0.2918392    
trump:unemployed         1 0.11461 0.11461  7.2820 0.0113249 *  
gini:unemployed          1 0.15920 0.15920 10.1154 0.0034056 ** 
Residuals               30 0.47217 0.01574                      
---
```

Table 10: Type 2 ANOVA output table for the final regression model with interaction terms. Results indicate Trump as the most significant predictor variable.



Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.56	-15.64 – 22.75	0.719
metropop	3.14	-0.11 – 6.39	0.068
lognoncit 2	2.01	0.41 – 3.60	**0.019**
shrnonwhite	-13.59	-23.33 – -3.85	**0.010**
logpoverty	4.19	-0.80 – 9.18	0.110
trump	-18.82	-36.84 – -0.81	**0.049**
gini	-12.60	-50.54 – 25.33	0.520
unemployed	560.49	298.50 – 822.48	**<0.001**
metropop:logpoverty	1.51	0.08 – 2.94	**0.048**
lognoncit2:shrnonwhite	-0.47	-1.06 – 0.11	0.125
lognoncit2:logpoverty	0.54	-0.05 – 1.13	0.084
lognoncit2:trump	-2.15	-3.23 – -1.07	**0.001**
lognoncit2:unemployed	9.04	-3.54 – 21.61	0.169
shrnonwhite:logpoverty	-3.76	-6.90 – -0.62	**0.026**
shrnonwhite:trump	5.73	0.51 – 10.96	**0.040**
logpoverty:gini	-13.70	-24.90 – -2.50	**0.023**
logpoverty:unemployed	75.91	35.21 – 116.61	**0.001**
trump:gini	28.64	-9.00 – 66.27	0.146
trump:unemployed	-88.04	-153.42 – -22.65	**0.013**
gini:unemployed	-682.36	-1102.86 – -261.85	**0.003**
Observations	50		
R^2 / adjusted R^2	0.738 / 0.572		

Table 11: Parameter estimates for the interaction model of question 1. Parameter estimates indicate unemployment as most influential variable. Trump remains significant, however the relationship is negative.

These results indicate that our first hypothesis was partially supported, the multiple linear regression indicated that Trump was a significant predictor, the relationship was a negative one, indicating that crime rates increase as shares of trump voters increase.

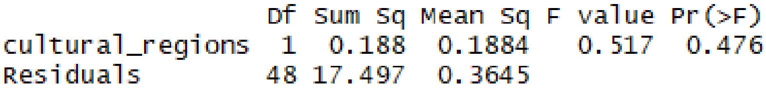
Hate Crime Occurrence by Region

A one way ANOVA was used to calculate differences in hate crime occurrence by cultural region. The final model was determined based on variable transformations for linearity (Table 10).

Iteration	Model	Justification
2	ModelCultSPLC1 <- aov(hatecrime2 ~ cultural_regions, data = cultregions.new2)	Initial Model
3	ModelCultSPLC2 <- aov(log(hatecrime2) ~ cultural_regions, data = cultregions.new2)	Post Variable Transformation for normality

Table 12: Indicates all models used in determination of final model and justifications. See Methods - Test2 for methodology.

The ANOVA results indicated that there was not a significant difference of cultural region by hate crime occurrence (F (1, 47) = 1.25, p > .27) (Table 11, and 12).



```
                 Df Sum Sq Mean Sq F value Pr(>F)
cultural_regions  1  0.188  0.1884   0.517  0.476
Residuals        48 17.497  0.3645
```

Table 13: Indicates one-way ANOVA results on cultural region and hate crime occurrence, test indicates no significant difference.

A Boxplot of the cultural regions demonstrates the unequal variances of the model which could be affecting the results (Figure 11). A parameter estimate table was developed to look at the strength of cultural regions as a parameter estimate, the table indicated a weak influence on the regressor variable (Table 12).

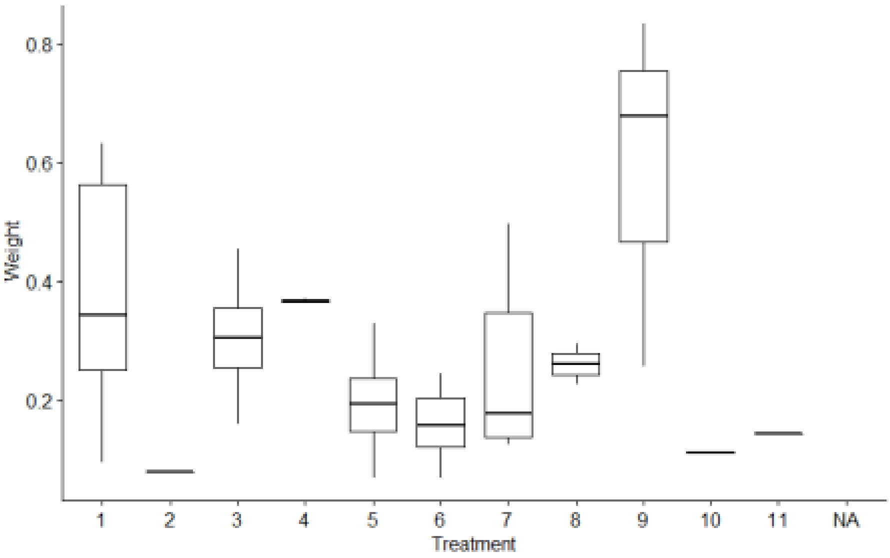
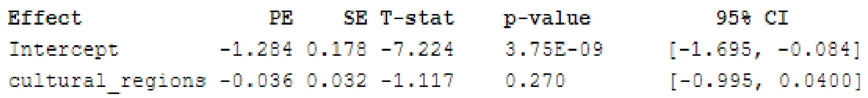


Figure 11: Indicates the weights of each of the 11 cultural regions. Boxplot indicates high variability across the regions. Cultural Region 2 (Yankee) was only represented by 1 state, 4 (tidewater) by 3 states and 10 (New France) was not represented. Cultural regions 1(Yankeedom) and 9 (left coast) had the greatest weights



```
Effect               PE    SE T-stat    p-value         95% CI
Intercept        -1.284 0.178 -7.224    3.75E-09     [-1.695, -0.084]
cultural_regions -0.036 0.032 -1.117    0.270        [-0.995, 0.0400]
```

Table 14. shows cultural_regions does not have a strong influence on log (hatecrime2).

Our second hypothesis was not supported, cultural region did not differ in their occurrence of hate crimes.

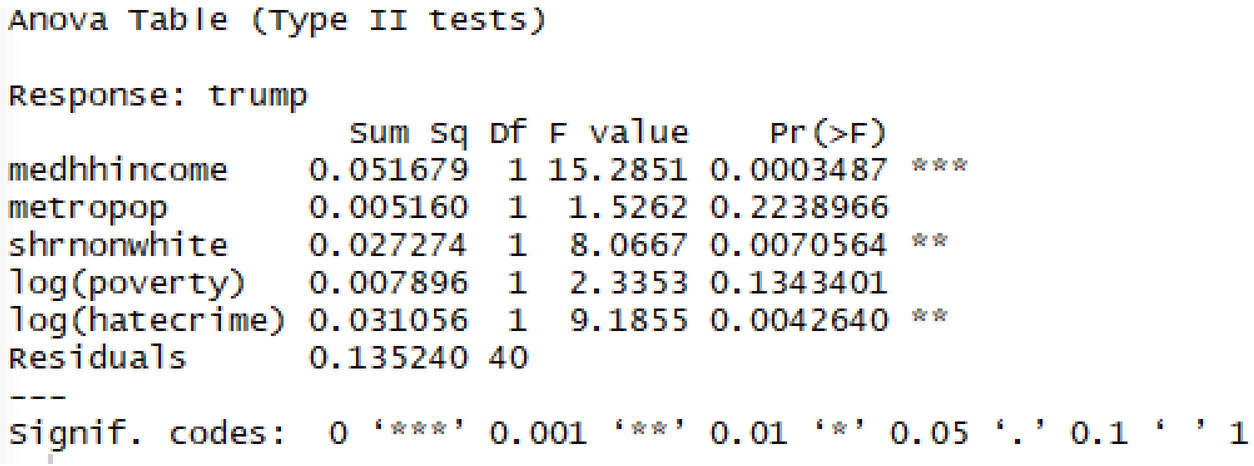
Predictors of High Share of Trump voters using socio-economic data

In determining predictors of high shares of Trump voters, a multiple linear regression was selected as the best test, the ultimate model of analysis was selected using backward stepwise elimination (Table 13).

Iteration	Model	Justification
1	m1 <- lm(trump ~ medhhincome + metropop + noncit2 + shrnonwhite + education + poverty + hatecrime + gini)	Initial Model
2	m2 <- lm(trump ~ medhhincome + metropop + log(noncit) + shrnonwhite + education + log(poverty) + log(hatecrime) + gini)	Transformed Variables for linearity
3	m3 <- lm(trump ~ medhhincome + metropop + shrnonwhite + log(poverty) + log(hatecrime))	Lowest AIC Score

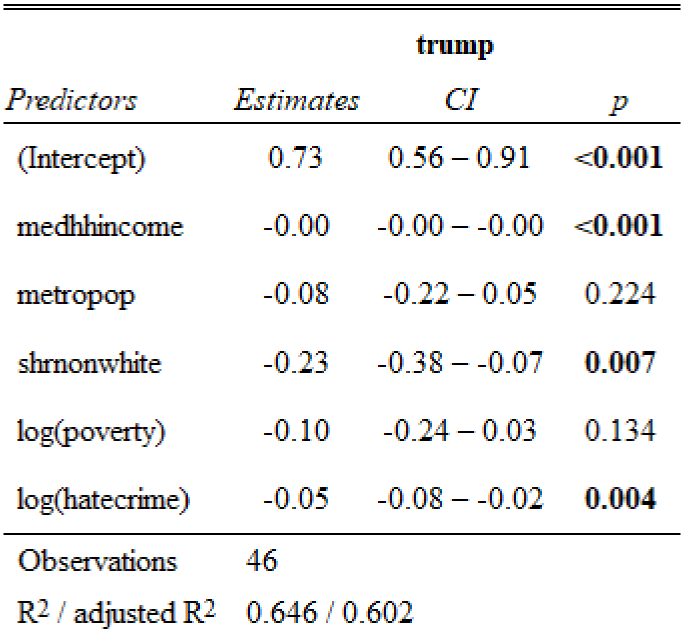
Table 15: Demonstrates model selection changes and justification for ultimate model. See Methods - Test3 for methodology, and Appendix B, Table 6 for AIC scores.

The multiple linear regression (type 2) Analysis indicated that there were multiple significant predictors of share of trump voters (Adjusted R2= 0.60, F (5, 40) =14.6, p < 3.803e-08). The most significant predictor was median household income (F (1, 40) = 15.28, p <0.0003) followed by (log) hate crime occurrence (F (1, 40) =9.18, p < 0.004) and the share of nonwhite individuals (F (1, 40) = 1.52, p < 0.007) (Table 14).



```
Anova Table (Type II tests)

Response: trump
                 Sum Sq Df F value    Pr(>F)    
medhhincome    0.051679  1 15.2851 0.0003487 ***
metropop       0.005160  1  1.5262 0.2238966    
shrnonwhite    0.027274  1  8.0667 0.0070564 ** 
log(poverty)   0.007896  1  2.3353 0.1343401    
log(hatecrime) 0.031056  1  9.1855 0.0042640 ** 
Residuals      0.135240 40                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 16: Type 2 ANOVA output table for the final regression model with interaction terms. Results indicate



	trump		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.73	0.56 – 0.91	**<0.001**
medhhincome	-0.00	-0.00 – -0.00	**<0.001**
metropop	-0.08	-0.22 – 0.05	0.224
shrnonwhite	-0.23	-0.38 – -0.07	**0.007**
log(poverty)	-0.10	-0.24 – 0.03	0.134
log(hatecrime)	-0.05	-0.08 – -0.02	**0.004**
Observations	46		
R^2 / adjusted R^2	0.646 / 0.602		

Table 17: Parameter estimate demonstrates that household income, share of non white and hate crime all are significant parameters with a negative relationship to Trump share in U.S. states.

Our hypothesis was supported in that there were significant socio-economic predictors of Trump voters.

Discussion

As stated in the beginning, the objective of this project was to fact-check the claims from two contrasting articles. The Results section shows the statistical behavior of the response with changes in the variables.

Hate crimes as a response was explained by predictors like poverty, unemployment, the number of people who have voted for trump and the population of non-white citizens. These observations are in line with literature on hate crimes which suggest that majority of these variables are socio-economic factors. For instance, poverty has a positive relationship with the response, meaning that as poverty increases hate crime increases. So same relation is observed between unemployment and hate crimes. This can be explained by the fact that most hate crimes are driven by envy. (Gale, 2002). Studies conducted shows that hate crimes are most significantly explained by socio-economic factors. So there is evidence to support the result of the model that as unemployment increases, hate crimes do increase. Variables like metropop, shrnonwhite, trump and gini have a negative effect on the response. This implies that as the population of metropolitan region reduces, the number of hate crimes increases. This can be logically supported by the fact that metropolitan regions generally have a more diverse community of people. A high share of metro population would allow more exposure to a diverse community thereby developing affinity towards people from different backgrounds. Hence, hate crimes are low in metro cities with high populations. The primary targets, however, vary widely by metropolitan area, tending to correlate with local demographics. (Hauslohner, 2018). Gini had a negative effect on hate crimes. As the index value decreased, hate crime incidents increased. These results support and are in sync with studies that explain the relationship between hate crimes and socio-economic factors. This supports out alternative hypothesis that a relationship between Trump voters and hate crime incident does exists.

As far as cultural regions are concerned, the ANOVA tests (Table 13) shows that there is no significant difference between hate crime rates amongst the cultural regions. Even though certain states, eg Texas, have a high crime rate was observed, they can't be explained by cultural regions. This is because of the inherent definition of these regions. Several states like Texas fall in multiple regions. The geographical boundaries of these regions are not as the same as the state boundaries. The data set used in the project deals with crimes reported in states. Hence, the results obtained supports the null hypothesis that we cannot explain hate crime based on cultural regions.

As far the factors explaining trump voters as a response are concerned, these factors are also socio-economical. Factors like metro population, income and poverty explain the response significantly. High share of trump voters is present where income and metro population has low value. This can be explained by claims made during the political campaigns during the Presidential election of 2016. A promise for more jobs would ensure more financial support and stability for individuals. Hence, as income is low, the share of trump voters is high. Similarly, if the population is low in metro regions, more jobs would imply less unemployment and poverty. Hence, as population of metro regions increase, the share of trump supporters decreased. Thus, there is evidence to accept the hypothesis that there is indeed a relationship between population share of trump voters and socio-economic factors.

Because of the complex nature of hate crimes, a model with standard socio-economic factors is not enough to analyze the subject. Models proposed by Becker (Becker, 1968) which incorporates envy demands further consideration of the topic. An interesting observation was that on the removal of an outlier, the adjusted R square increased by 10%. The significant change was probably because of the small size of the data set. Using the SPLC data gave negative relations with some variables. However, after running the model using the FBI dataset certain socio-economic factors showed positive relations. This can be explained by the inconsistency between SPLC and FBI data. The method in which the SPLC collected data is different from FBI. FBI collects data throughout the year whereas SPLC data was collected over a number of days after 2016 elections.

These limitations in our analyses could be addressed through incorporating more data, both on valuation variables and in sample size. Despite empirical difficulties, the results and models were able to strongly explain a part of a complex variable like hate crimes.

Resources

Allcot, H., and Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives,* 3(2); 211-236

Elazer, D.J. (1966). American Federalism: A View from the States. New York.

Ellis, R. (1993). American Political Cultures. Oxford University Press, New York.

Federal Bureau of Investigations (FBI). (2016). 2016 Hate Crime Statistics. Retrieved from: https://ucr.fbi.gov/hate-crime/2016/topic-pages/jurisdiction

Gamson, W. 1989. News as Framing. *The American Behavioral Scientist.* 33(2): 157-161.

Gareth, J., Witten, D., Hastie, T., and Tibshirani, R. (2014). An Introduction to Statistical Learning: With Applications in R. *Springer Publishing Company, Incorporated.*

Johnson, D. (2018). Report: Rise in Hate Crimes Tied to 2016 Election. *SPLC Online.* https://www.splcenter.org/hatewatch/2018/03/01/report-rise-hate-violence-tied-2016-presidential-election

Majumder, M. (2017). Higher Rates of Hate Crimes are Tied to Income Inequality. Retrieved from: https://fivethirtyeight.com/features/higher-rates-of-hate-crimes-are-tied-to-income-inequality/

Marchi, R. (2012). With Facebook, Blogs, and Fake News, Teens Reject Journalistic “Objectivity.” *Journal of Communication Inquiry*. 36(3); 246-262.

Mehta, Dhrumil. (2016) Hate Crimes Dataset. *Fivethirtyeight.com; Github.* (https://github.com/fivethirtyeight/data/tree/master/hate-crimes

Nisbet, M. (2015). Framing, the Media, and Risk Communication in Policy Debates. *The SAGE Handbook of Risk Communication.* Sage Publications, In

NIST. (2012).NIST/SEMATECH e-Handbook of Statistical Methods, 1.3.5.14, Retrieved from https://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm

Racine J. and Hyndman, R., (2002), Using R To Teach Econometrics. *Journal of Applied Econometrics* 17, 175–189.

US Census Bureau, (2016). "Geography: Distributions."https://www.census.gov/geo/reference/webatlas/divisions.html

Tversky, A., and Kahneman, D., (1981). The Framing of Decisions and the Psychology of Choice. *Science.* 211(4481). 453-458.

Wetzel, N. (1995). Graphical Data Modeling Methods Using CERES plots. *Journal of Statistical Computation and Simulation,* 54(1-3), 37-44.

Woodard, C. (2012). American Nations: A History of the Eleven Rival Regional Cultures of North America. *Penguin Group Books, USA.*
; https://www.washingtonpost.com/blogs/govbeat/wp/2013/11/08/which-of-the-11-american-nations-do-you-live-in/?utm_term=.d6e62ce022ed

Woodard, C. (2018). Opinion: The Maps That Show That City vs Country Is Not Our Political Fault Line. *New York Times.* https://www.nytimes.com/2018/07/30/opinion/urban-rural-united-states-regions-midterms.html

Abigail Hauslohner (2018). Hate Crime Rates Are Still On The Rise

https://www.washingtonpost.com/news/post-nation/wp/2018/05/11/hate-crime-rates-are-still-on-the-rise/?utm_term=.4162f88c13c1

Lewis R. Gale, Will Carrington Heath and Rand W. Ressler (2002). An Economic Analysis of Hate Crime. Eastern Economic Journal Vol. 28, No. 2 (Spring, 2002), pp. 203-216

Becker, G. S. (1968) Crime and Punishment: An Economic Approach. *Journal of Political Economy*, March/April 1968, 169-217

Appendix A: R Code

```
library(readxl)
finalproj <- read_excel("Merged Dataset (aka Frankenstein's Monster).xlsx")

##################  Library  #################
library("car")
library("gvlma") #for global assumption test
library("caret") #for influence plot
library("Hmisc") #for impute function
library("dplyr")
library("ggpubr")
library("ggplot2")
library("nortest") # for normality test
library("Rcpp")
library("rlang")
library("alr4")


##############################################################################
################  Exploratory Analyses of Whole Data Set    ##################
##############################################################################

####################  Check Assumption on SPLC DATA    #######################

##Since SPLC measures the number of hate crimes committed between November 8-
15, 2016, data was transformed to be more comparable to the FBI data ( a
yearly measure)

## 365/8 = 45.625
#SPLC data multiplied for yearly estimate based on above calculation#
hate_crimes_per_100k_splc_ADJ <-finalproj$hate_crimes_per_100k_splc*45.625

#Welches Two sample t test#
t.test(hate_crimes_per_100k_splc_ADJ, finalproj$avg_hatecrimes_per_100k_fbi)


#Remove Non-State Variable "District of Columbia" #
finalproj_new<-finalproj[-9,]


#Relabel Variables#
head(finalproj_new)
hatecrime<-finalproj_new$hate_crimes_per_100k_splc
medhhincome<-finalproj_new$median_household_income
metropop<-finalproj_new$share_population_in_metro_areas
```

```
noncit<-finalproj_new$share_non_citizen
shrnonwhite<-finalproj_new$share_non_white
education<-finalproj_new$share_population_with_high_school_degree
poverty<-finalproj_new$share_white_poverty
trump<-finalproj_new$share_voters_voted_trump
gini<-finalproj_new$gini_index
unemployed<-finalproj_new$share_unemployed_seasonal

#motivations
racemot<-finalproj_new$race_motivation_2016
religmot<-finalproj_new$religion_motivation_2016
sexmot<-finalproj_new$sexuality_motivation_2016
ethnimot<-finalproj_new$ethnicity_motivation_2016
disabmot<-finalproj_new$disability_motivation_2016
gendermot<-finalproj_new$gender_motivation_2016
percentotal_racemot<-finalproj_new$pct_race_motivation


#cultural regions
cultural_regions<-finalproj_new$cultural_region_categorical
yankee_cultr<-finalproj_new$cultural_region_yankee
newneth_cultr<-finalproj_new$cultural_region_newneth
midland_cultr<-finalproj_new$cultural_region_midland
tidewater_cultr<-finalproj_new$cultural_region_tidewater
appalachia_cultr<-finalproj_new$cultural_region_appalachia
deepsth_cultr<-finalproj_new$cultural_region_deepsouth
farwest_cultr<-finalproj_new$cultural_region_farwest
norte_cultr<-finalproj_new$cultural_region_norte
left_cultr<-finalproj_new$cultural_region_left
newfrance_cultr<-finalproj_new$cultural_region_newfrance
firstnation_cultr<-finalproj_new$cultural_region_firstnation

#Dataframe of Interest#

hatedf.new <-data.frame(hatecrime, medhhincome, metropop,
noncit,shrnonwhite,education,poverty,trump,gini,unemployed)

cultregions.new <-
      data.frame(hatecrime,cultural_regions,yankee_cultr,newneth_cultr,midlan
d_cultr,tidewater_cultr,appalachia_cultr,deepsth_cultr,farwest_cultr,norte_cu
ltr,left_cultr,newfrance_cultr,firstnation_cultr)

###############################  Check NA    ################################

colSums(is.na(hatedf.new))
colSums(is.na(cultregions.new))

#hatecrime = 4 and noncit = 3
hatecrime2<-impute(hatecrime, fun = mean)
noncit2<-impute(noncit,fun=mean)

# modify the dfs with new cells#
hatedf.new2<-
      data.frame(hatecrime2,medhhincome,metropop,noncit2,shrnonwhite,educatio
n,poverty,trump,gini,unemployed)
```

```
cultregions.new2<-
data.frame(hatecrime2,cultural_regions,yankee_cultr,newneth_cultr,midland_cul
tr,tidewater_cultr,appalachia_cultr,deepsth_cultr,farwest_cultr,norte_cultr,l
eft_cultr,newfrance_cultr,firstnation_cultr)

####################### Scatterplot Matrices ################

pairs(hatedf.new2,panel=panel.car,pch="16",col="blue")
pairs(cultregions.new2,panel=panel.car,pch="16",col="blue")


#############################################################################
#####################Q1:PredictorsofHateCrimeLinearRegression################
#############################################################################

#Formation of Model#
hatmod1<-
lm(hatecrime2~medhhincome+metropop+noncit2+shrnonwhite+education+poverty+trum
p+gini+unemployed)

##################Test Assumptions of Model 1#################

#plot residuals#
plot(hatmod1)

#Test for outliers#
outlierTest(hatmod1) # Bonferonni p-value for most extreme obs
qqPlot(hatmod1, main="QQ Plot") #qq plot for studentized residuals
leveragePlots(hatmod1) # leverage plots

# Influence Plot #
influencePlot(hatmod1,id.method="identify", main="Influence Plot",
              sub="Circle size is proportial to Cook's Distance" )

#Test for homoscedasticity#
# non-constant error variance test
ncvTest(hatmod1) # p <0.003

#test for multicollinearity#
vif(hatmod1) #variance inflation

##Education =5.8 and medhhincome = 5.6##

#Evaluate Nonlinearity#
ceresPlots(hatmod1)
resid1<-resid(hatmod1)
ad.test(resid1) #specific to normal distribution#  #non-linear P>0.737

# Test for Autocorrelated Errors#
durbinWatsonTest(hatmod1)  ##errors are independent (p > 0.918 )
```

```
#################Transformations of Model 1################
shapiro.test(hatecrime2) # (p < 0.001)
shapiro.test(gini) # (p > 0.50) normal
shapiro.test(metropop) # (p < 0.22) normal
shapiro.test(noncit2) # (p < 0.01)
shapiro.test(shrnonwhite) # # (p > 0.07) normal
shapiro.test(poverty)# (p < 0.03)
shapiro.test(trump) #(p > 0.61) normal
shapiro.test(unemployed) # (p > 0.83) normal

##Transform Non Linear Variables ##
loghate2<-log(hatecrime2)
lognoncit2<-log(noncit2)
logpoverty<-log(poverty)


hatmod2<-lm(loghate2~+metropop+lognoncit2+shrnonwhite+
            logpoverty+trump+gini+unemployed, data=finalproj_new)


######################Test Assumptions of Model 2####################
#plot residuals#
plot(hatmod2)

#Test for outliers#
outlierTest(hatmod2) # Bonferonni p-value for most extreme obs  (p <0.02)
qqPlot(hatmod2, main="QQ Plot") #qq plot for studentized resid
leveragePlots(hatmod2) # leverage plots
influencePlot(hatmod2,id.method="identify", main="Influence Plot",
              sub="Circle size is proportial to Cook's Distance")   #indicate
that the outlier no longer exceeds 4

#Test for homoscedasticity#
#non-constant error variance test#
ncvTest(hatmod2)  # p <0.065

#test for multicollinearity#
vif(hatmod2) #variance inflation  - no variables beyond 3 on VIF#

#Evaluate Nonlinearity#
crPlots(hatmod2)
ceresPlots(hatmod2)
resid3<-resid(hatmod2)
ad.test(resid3) #specific to normal distribution# ## non-linear - p > 0.002##
ad.test(hatmod2)

#Re-Evalualte Transformation#
BoxCoxTrans(hatecrime2) #- imputed 4 values to 0.22, lambda = 0 for
transformations#
hate3<-asin(hatecrime2)
```

```
################################ Model 3 ######################################
hatmod3<-lm(hate3~+metropop+lognoncit2+shrnonwhite+
              logpoverty+trump+gini+unemployed, data=finalproj_new)

##################### Test Assumptions of Model 3 #############################
#obtain residuals and do darling test for normality
resid4<-resid(hatmod3)
ad.test(resid4) # p >0.1788, normal

#Test for homoscedasticity#
#non-constant error variance test
ncvTest(hatmod3) #still significant p < 6.3662e-06

#test for multicollinearity#
vif(hatmod3) #same as model 2 , no values beyond 3

#Test for outliers#
outlierTest(hatmod3) # Bonferonni p-value for most extreme obs  (p <0.03)
qqPlot(hatmod3, main="QQ Plot") #qq plot for studentized resid
influencePlot(hatmod3,id.method="identify", main="Influence Plot",
              sub="Circle size is proportial to Cook's Distance")   #indicate
that the outlier no longer exceeds 4

#37 is more noticable however normality is maintained and it does not exceed
4 influence chart, it is left alone.

leveragePlots(hatmod3) # leverage plots

# Influence Plot #
influencePlot(hatmod3,id.method="identify", main="Influence Plot",
              sub="Circle size is proportial to Cook's Distance"
)   #indicate that the outlier no longer exceeds 4


# Test for Autocorrelated Errors#
durbinWatsonTest(hatmod2)
##errors are independent (p > 0.87)

##################  Stepwise Check of Model 3  #####################

hatmod3<-lm(hate3~+metropop+lognoncit2+shrnonwhite+
              logpoverty+trump+gini+unemployed, data=finalproj_new)

summary(hatmod3)
plot(hatmod3)

#Conduct Backward Stepwise#
step(hatmod3)#indicates a simplified model is possible with an AIC of -178.4
vs -170 of Model 3

hatmodAIC<-lm(hate3~shrnonwhite+trump)
summary(hatmodAIC)
plot(hatmodAIC)
```

```
#obtain ANOVA table for Results
anova(hatmod3,hatmodAIC, test="Chisq")  # p > 0.689 - reduction in SSres not
sig.
anova(hatmodAIC)

#calculate confidence intervals for parameters
confint(hatmodAIC)

#confirm two models using AIC and BIC
AIC(hatmod3) #-28
AIC(hatmodAIC) #-34

BIC(hatmod3) # -10
BIC(hatmodAIC) #-26  # results same for BIC, AIC used for report

#################  Part B – Interactions  #########################
##Check interactions - not in question#
hatmod4<-lm(hate3~+metropop+lognoncit2+shrnonwhite+
            logpoverty+trump+gini+unemployed+metropop:lognoncit2+
            metropop:shrnonwhite+metropop:logpoverty+metropop:trump+
            metropop:gini+metropop:unemployed+lognoncit2:shrnonwhite+
            lognoncit2:logpoverty+lognoncit2:trump+lognoncit2:gini+
            lognoncit2:gini+lognoncit2:unemployed+shrnonwhite:logpoverty+
            shrnonwhite:trump+shrnonwhite:gini+shrnonwhite:unemployed+
            logpoverty:trump+logpoverty:gini+logpoverty:unemployed+trump:gin+
            trump:unemployed+gini:unemployed,data=finalproj_new)

step(hatmod4)

#lowest AIC model
hatmod5<-lm (hate3 ~ metropop + lognoncit2 + shrnonwhite + logpoverty + trump
+ gini + unemployed + metropop:logpoverty + lognoncit2:shrnonwhite +
  lognoncit2: logpoverty + lognoncit2: trump + lognoncit2: unemployed +
  shrnonwhite:logpoverty + shrnonwhite:trump + logpoverty:gini +
  logpoverty:unemployed + trump:gini + trump:unemployed + gini:unemployed)

Anova(hatmod5, type="II")

plot_model(hatmod5)

#remove non-significant interaction terms#

hatmod6<- lm(hate3 ~ metropop + lognoncit2 + shrnonwhite + logpoverty + trump
+  gini + unemployed + metropop:logpoverty +lognoncit2:trump +
shrnonwhite:logpoverty + shrnonwhite:trump + logpoverty:gini +
logpoverty:unemployed + trump:unemployed + gini:unemployed)

Anova(hatmod6, type="II")
summary(hatmod6)

#obtain ANOVA table for Results
anova(hatmod5,hatmod6) #significant change in reduced model, old one kept
anova(hatmod5)
summary(hatmod5)
```

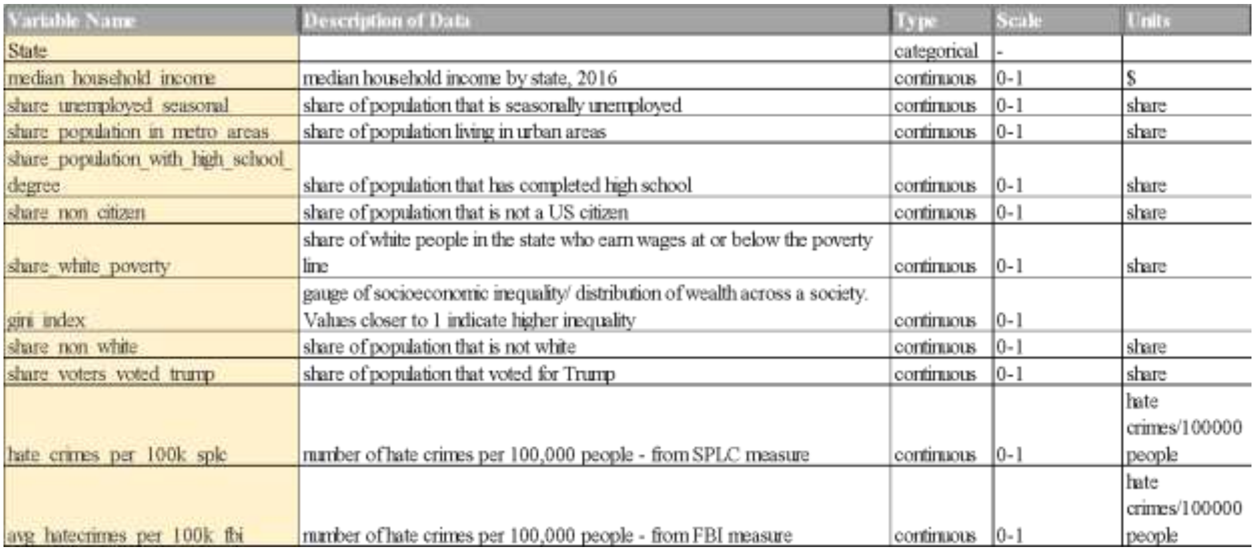
```
#parameter estimates#
tab_model(hatmod3)
tab_model(hatmod5)


#############################################################################
#######Q2: Do Cultural Regions Differ by their Rate of Hate Crime ###########
#############################################################################

finalproj_new[18,20] <- 10  # Set the region for Louisiana to New France
State <- finalproj_new$state

# Make data frames
cultregions.new <- data.frame(hatecrime, State, cultural_regions,
yankee_cultr, newneth_cultr, midland_cultr, tidewater_cultr,
appalachia_cultr, deepsth_cultr, farwest_cultr, norte_cultr, newfrance_cultr,
firstnation_cultr)
cultregions.new2 <- data.frame(hatecrime, State, cultural_regions)
cultregions.new3 <- data.frame(log(hatecrime), State, cultural_regions)

# Check NA
colSums(is.na(cultregions.new))
hatecrime2 <- impute(hatecrime, fun = mean)

#################Model 1#################
ModelCultSPLC1 <- aov(hatecrime2 ~ cultural_regions, data =
cultregions.new2)                         # Create the first model
drop1(ModelCultSPLC1, ~., test = "F")     # Make it type 3
summary(ModelCultSPLC1)                   # Get information
par(mfrow=c(2,2))                         # initiate 4 charts in 1 panel
plot(ModelCultSPLC1)                      # Create plots
# QQ Plot shows data is not normal and needs to be transformed.

##################  Test Assumptions of Model 1  #########################
outlierTest(ModelCultSPLC1)          # Checking outliers in first model -
significant outlier #37, p<0.01
ad.test(residuals(ModelCultSPLC1))   # Test for normality in first model -
Not Sig - normal p > 0.78

##################  Model 2  #################
ModelCultSPLC2 <- aov(log(hatecrime2) ~ cultural_regions, data =
cultregions.new2)                    # Create the revised model with log
transformed regressor
drop1(ModelCultSPLC2, ~., test = "F")     # Make it type 3
summary(ModelCultSPLC2)                   # Get information
par(mfrow=c(2,2))                         # initiate 4 charts in 1 panel
plot(ModelCultSPLC2)                      # Create plots

# Create a box plot
ggboxplot(cultregions.new3, x = "cultural_regions", y = "hatecrime", ylab =
"Weight", xlab = "Treatment")
```

```
##################  Test Assumptions of Model 2  ######################

outlierTest(ModelCultSPLC2)          # Bonferroni p-value for most extreme
observations - outlier no longer sig #37, p > 0.90
ad.test(residuals(ModelCultSPLC2))   # Test for normality - Not sig - normal
p >0.78
durbinWatsonTest(ModelCultSPLC2)     # Test for autocorrelated errors -errors
are indepdendent- p > 0.96

#Final Model
ModelCultSPLC2 <- aov(log(hatecrime2) ~ cultural_regions, data =
cultregions.new2)



summary(ModelCultSPLC2)

#Calculate CI
confint(ModelCultSPLC2)

##############################################################################
####################  Q3: What Predicts Trump Voter Presence?  ##############
##############################################################################
attach(finalproj)
m1<-
lm(trump~medhhincome+metropop+noncit2+shrnonwhite+education+poverty+hatecrime
+gini)

#plot residuals
plot(m1)

##################Test Assumptions of Model 1##################
#Test for outliers
outlierTest(m1)  #Outlier found of significance, obs. 45, p < 0.0003
qqPlot(m1, main="QQ Plot") #qq plot for studentized residuals

#Test for homoscedasticity#
# non-constant error variance test
ncvTest(m1) # p >0.78 - Variance is not constant

#test for multicollinearity#
vif(m1) #variance inflation - one variable over 5 = education

##################  Model 1 Transformations  #################

#Maintained from Model Transformations in Q2
loghate2<-log(hate_crimes_per_100k_splc)

m2<-
lm(trump~medhhincome+metropop+log(noncit)+shrnonwhite+education+log(poverty)+
log(hatecrime)+gini)
```

```
##################  Test Assumptions of Model 2  #################

#plot residuals
plot(m2)

#Test for homoscedasticity#
# non-constant error variance test
ncvTest(m2) # p > 1.0 - Variance is Constant

#test for multicollinearity#
vif(m2) #variance inflation # no variables are pass 5, though medhhincome
comes close (4.9)

################## Stepwise Check of Model 2  #################
step(m2)

############Model 3 derived from 2 where all assumptions were
met#############

m3<-lm(trump ~ medhhincome + metropop + shrnonwhite + log(poverty) +
log(hatecrime))
summary(m3)

####NCV and VIF check on m3######
ncvTest(m3)

vif(m3)

anova(m3)
confint(m3)
```

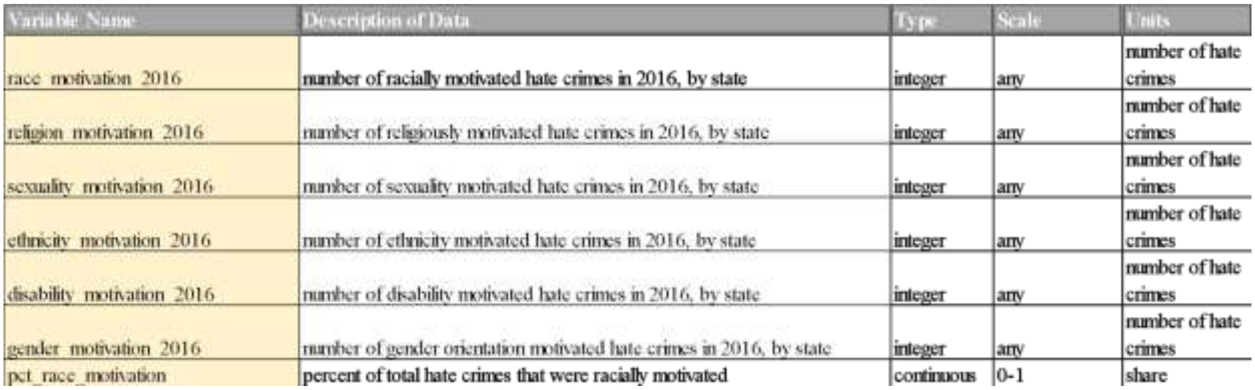
Appendix B: Output

Sources: Variables of Interest



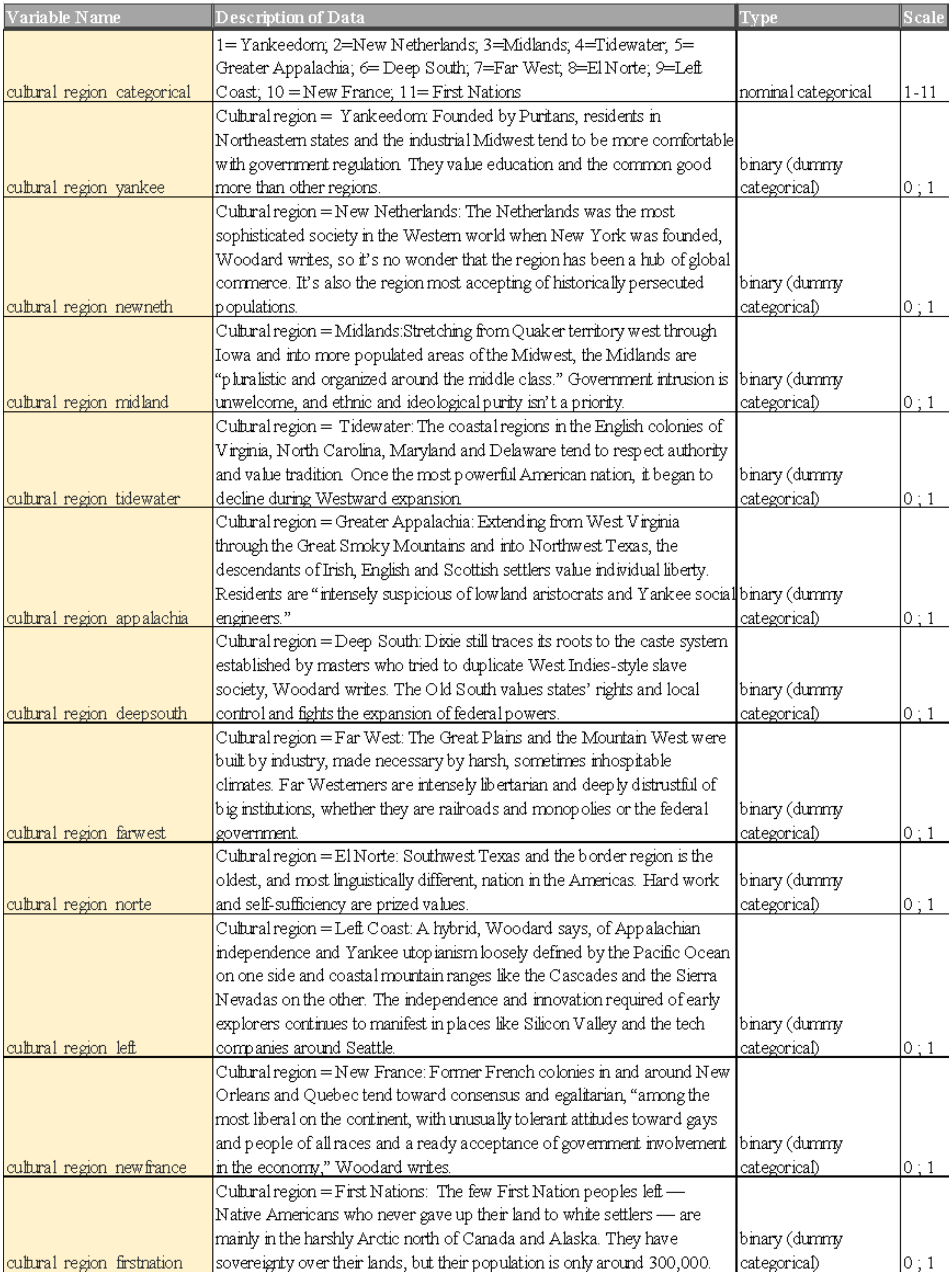
Variable Name	Description of Data	Type	Scale	Units
State		categorical	-	
median_household_income	median household income by state, 2016	continuous	0-1	$
share_unemployed_seasonal	share of population that is seasonally unemployed	continuous	0-1	share
share_population_in_metro_areas	share of population living in urban areas	continuous	0-1	share
share_population_with_high_school_degree	share of population that has completed high school	continuous	0-1	share
share_non_citizen	share of population that is not a US citizen	continuous	0-1	share
share_white_poverty	share of white people in the state who earn wages at or below the poverty line	continuous	0-1	share
gini_index	gauge of socioeconomic inequality/ distribution of wealth across a society. Values closer to 1 indicate higher inequality	continuous	0-1	
share_non_white	share of population that is not white	continuous	0-1	share
share_voters_voted_trump	share of population that voted for Trump	continuous	0-1	share
hate_crimes_per_100k_splc	number of hate crimes per 100,000 people - from SPLC measure	continuous	0-1	hate crimes/100000 people
avg_hatecrimes_per_100k_fbi	number of hate crimes per 100,000 people - from FBI measure	continuous	0-1	hate crimes/100000 people

Table 1: Data obtained from fivethirtyeight.com, the table includes the variable name, its description, type, scale and units of measure.



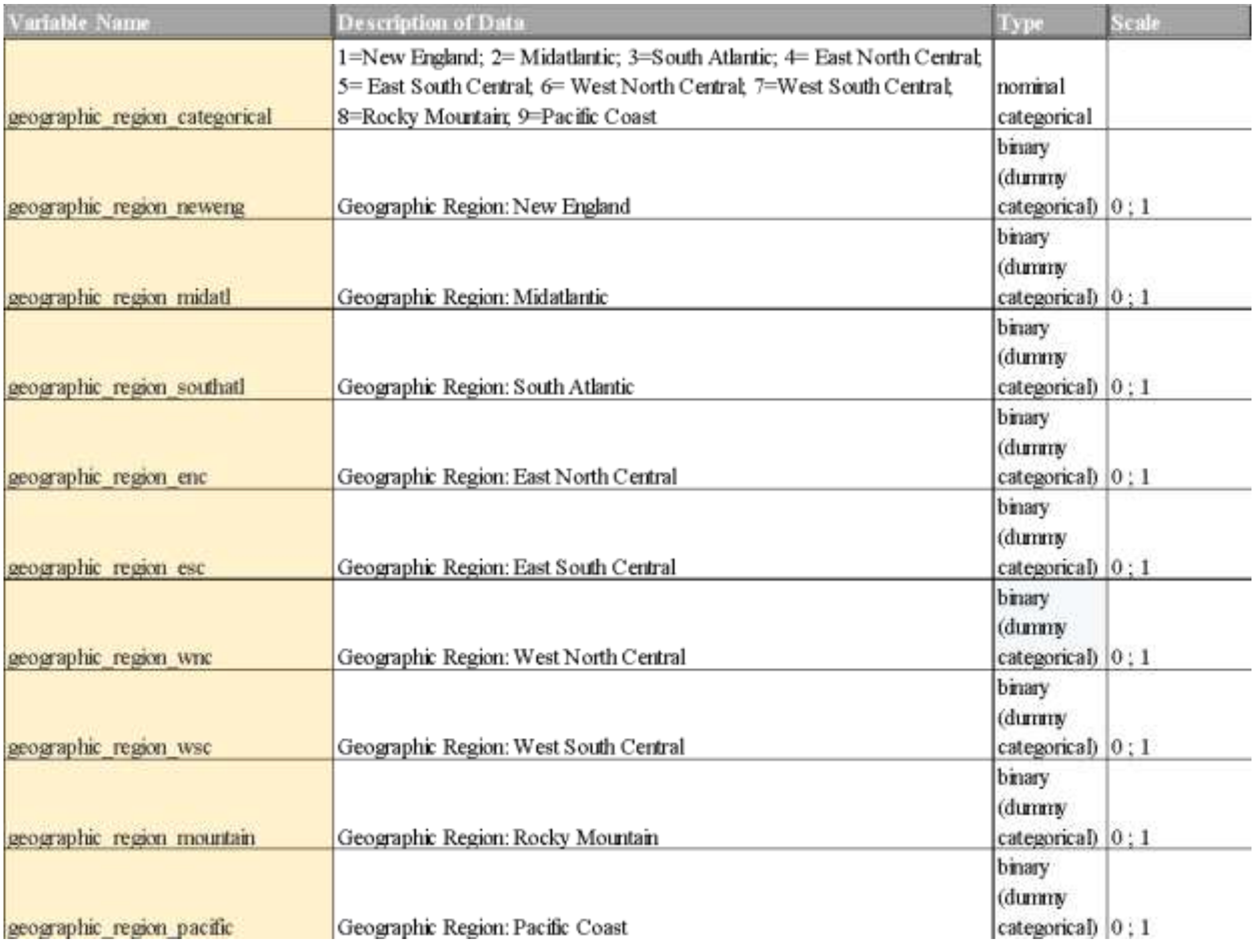
Variable Name	Description of Data	Type	Scale	Units
race_motivation_2016	number of racially motivated hate crimes in 2016, by state	integer	any	number of hate crimes
religion_motivation_2016	number of religiously motivated hate crimes in 2016, by state	integer	any	number of hate crimes
sexuality_motivation_2016	number of sexuality motivated hate crimes in 2016, by state	integer	any	number of hate crimes
ethnicity_motivation_2016	number of ethnicity motivated hate crimes in 2016, by state	integer	any	number of hate crimes
disability_motivation_2016	number of disability motivated hate crimes in 2016, by state	integer	any	number of hate crimes
gender_motivation_2016	number of gender orientation motivated hate crimes in 2016, by state	integer	any	number of hate crimes
pct_race_motivation	percent of total hate crimes that were racially motivated	continuous	0-1	share

Table 2: Division of hate crime data by motivation, data obtained from the FBI, the table includes the variable name, its description, type, scale and units of measure.



Variable Name	Description of Data	Type	Scale
cultural region categorical	1= Yankeedom; 2=New Netherlands; 3=Midlands; 4=Tidewater; 5= Greater Appalachia; 6= Deep South; 7=Far West; 8=El Norte; 9=Left Coast; 10 = New France; 11= First Nations	nominal categorical	1-11
cultural region yankee	Cultural region = Yankeedom: Founded by Puritans, residents in Northeastern states and the industrial Midwest tend to be more comfortable with government regulation. They value education and the common good more than other regions.	binary (dummy categorical)	0 ; 1
cultural region newneth	Cultural region = New Netherlands: The Netherlands was the most sophisticated society in the Western world when New York was founded, Woodard writes, so it's no wonder that the region has been a hub of global commerce. It's also the region most accepting of historically persecuted populations.	binary (dummy categorical)	0 ; 1
cultural region midland	Cultural region = Midlands:Stretching from Quaker territory west through Iowa and into more populated areas of the Midwest, the Midlands are "pluralistic and organized around the middle class." Government intrusion is unwelcome, and ethnic and ideological purity isn't a priority.	binary (dummy categorical)	0 ; 1
cultural region tidewater	Cultural region = Tidewater: The coastal regions in the English colonies of Virginia, North Carolina, Maryland and Delaware tend to respect authority and value tradition. Once the most powerful American nation, it began to decline during Westward expansion.	binary (dummy categorical)	0 ; 1
cultural region appalachia	Cultural region = Greater Appalachia: Extending from West Virginia through the Great Smoky Mountains and into Northwest Texas, the descendants of Irish, English and Scottish settlers value individual liberty. Residents are "intensely suspicious of lowland aristocrats and Yankee social engineers."	binary (dummy categorical)	0 ; 1
cultural region deepsouth	Cultural region = Deep South: Dixie still traces its roots to the caste system established by masters who tried to duplicate West Indies-style slave society, Woodard writes. The Old South values states' rights and local control and fights the expansion of federal powers.	binary (dummy categorical)	0 ; 1
cultural region farwest	Cultural region = Far West: The Great Plains and the Mountain West were built by industry, made necessary by harsh, sometimes inhospitable climates. Far Westerners are intensely libertarian and deeply distrustful of big institutions, whether they are railroads and monopolies or the federal government.	binary (dummy categorical)	0 ; 1
cultural region norte	Cultural region = El Norte: Southwest Texas and the border region is the oldest, and most linguistically different, nation in the Americas. Hard work and self-sufficiency are prized values.	binary (dummy categorical)	0 ; 1
cultural region left	Cultural region = Left Coast: A hybrid, Woodard says, of Appalachian independence and Yankee utopianism loosely defined by the Pacific Ocean on one side and coastal mountain ranges like the Cascades and the Sierra Nevadas on the other. The independence and innovation required of early explorers continues to manifest in places like Silicon Valley and the tech companies around Seattle.	binary (dummy categorical)	0 ; 1
cultural region newfrance	Cultural region = New France: Former French colonies in and around New Orleans and Quebec tend toward consensus and egalitarian, "among the most liberal on the continent, with unusually tolerant attitudes toward gays and people of all races and a ready acceptance of government involvement in the economy," Woodard writes.	binary (dummy categorical)	0 ; 1
cultural region firstnation	Cultural region = First Nations: The few First Nation peoples left — Native Americans who never gave up their land to white settlers — are mainly in the harshly Arctic north of Canada and Alaska. They have sovereignty over their lands, but their population is only around 300,000.	binary (dummy categorical)	0 ; 1

Table3: Cultural Regions found within the United States, data obtained from Woodard, 2012. Contains variables used, their description, type, and scale.



Variable Name	Description of Data	Type	Scale
geographic_region_categorical	1=New England; 2= Midatlantic; 3=South Atlantic; 4= East North Central; 5= East South Central; 6= West North Central; 7=West South Central; 8=Rocky Mountain; 9=Pacific Coast	nominal categorical	
geographic_region_neweng	Geographic Region: New England	binary (dummy categorical)	0 ; 1
geographic_region_midatl	Geographic Region: Midatlantic	binary (dummy categorical)	0 ; 1
geographic_region_southatl	Geographic Region: South Atlantic	binary (dummy categorical)	0 ; 1
geographic_region_enc	Geographic Region: East North Central	binary (dummy categorical)	0 ; 1
geographic_region_esc	Geographic Region: East South Central	binary (dummy categorical)	0 ; 1
geographic_region_wnc	Geographic Region: West North Central	binary (dummy categorical)	0 ; 1
geographic_region_wsc	Geographic Region: West South Central	binary (dummy categorical)	0 ; 1
geographic_region_mountain	Geographic Region: Rocky Mountain	binary (dummy categorical)	0 ; 1
geographic_region_pacific	Geographic Region: Pacific Coast	binary (dummy categorical)	0 ; 1

Table 4: Geographic Regions of the United States, data obtained from the US Census Bureaus, 2016. Contains main geographic regions of the USA their descriptions, type, and scales.

Exploratory Analyses

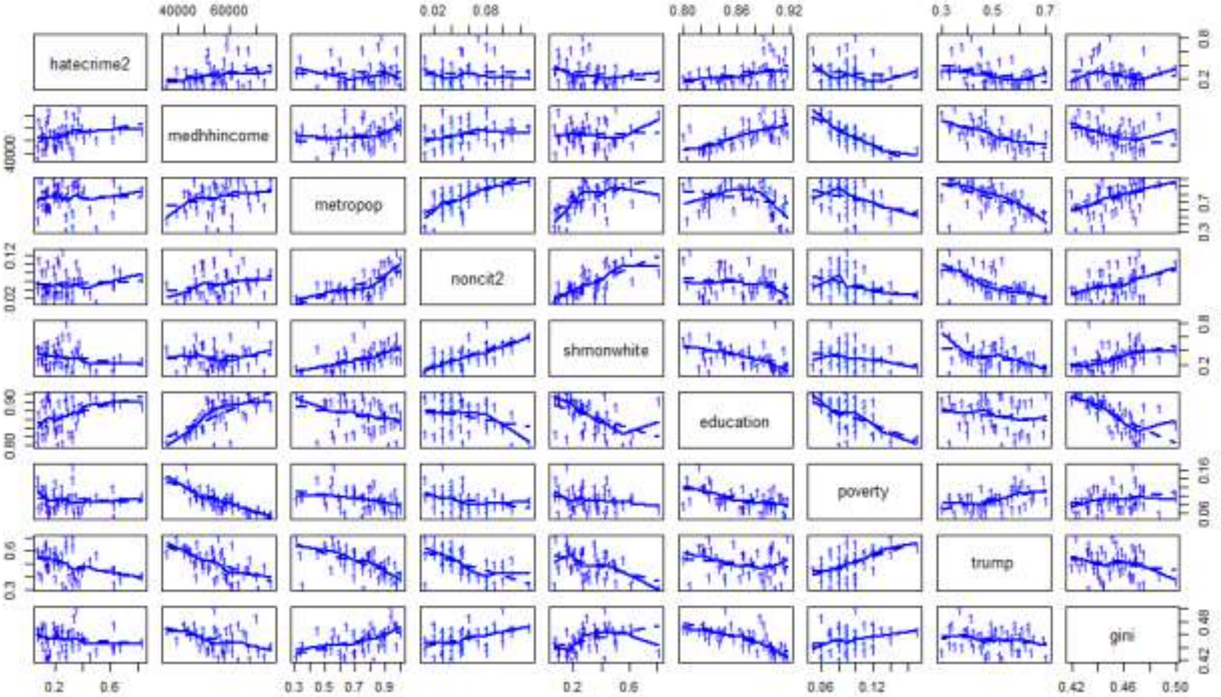


Figure 1: Scatterplot Matrix, with smoothed loess line, hate crime data (top left) in relation to socio-economic factors and trump voters.

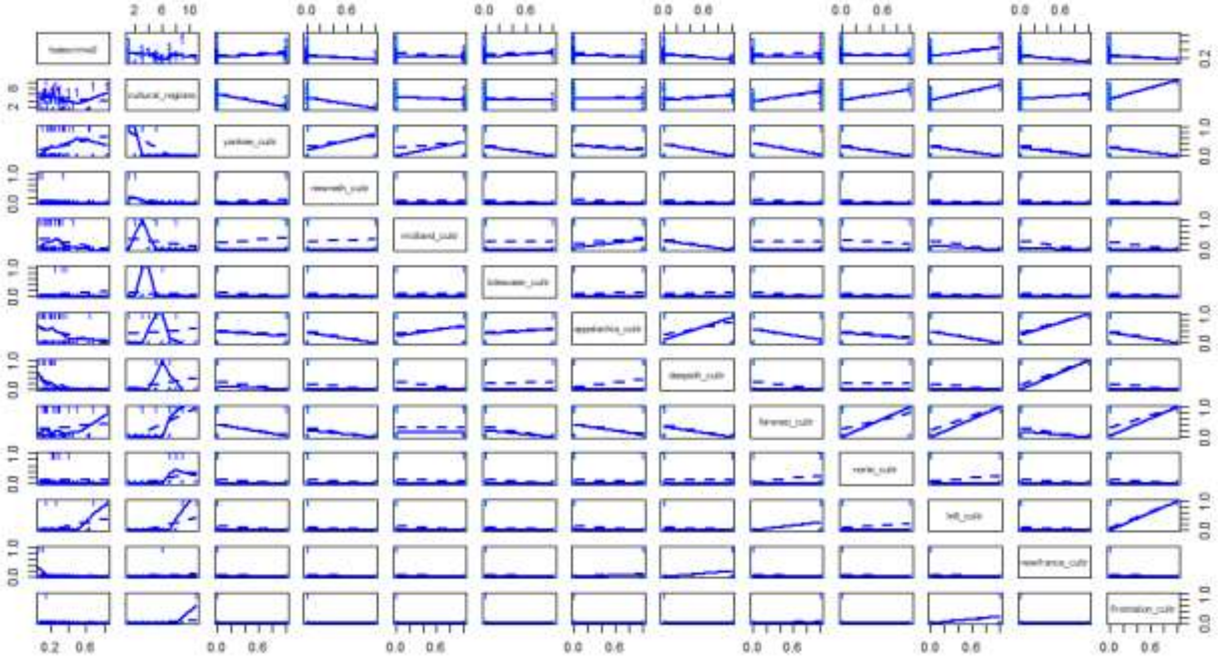


Figure 2: Scatterplot Matrix with smoothed loess line, hate crime data (top left) in relation to cultural regions in the United States.

Methods: Model Design and Assumption Testing

Test 1: Predictors of Hate Crimes - Model 1

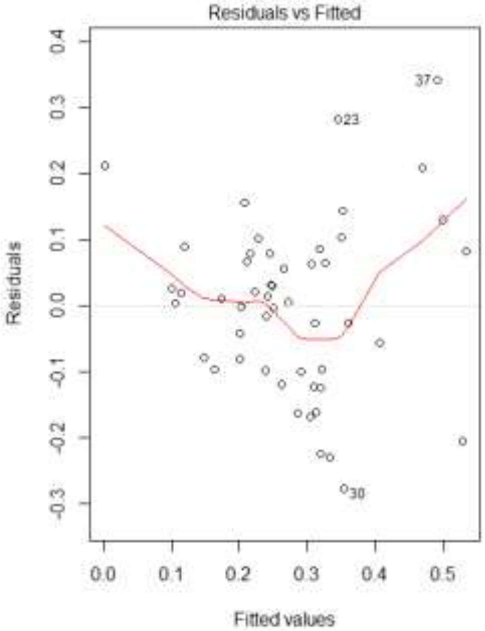


Figure 3: Residual plot of Model 1 (left) and QQplot (right) for Question 1, testing for normality and outliers. Bonferonni Tests confirmed observed outliers were significant (obs 37, r-student = 3.13, p < 0.003, Bonferonni p > 0.162), and the NCV test confirmed heteroscedasticity (X^2 = 0.08, p < 0.78).

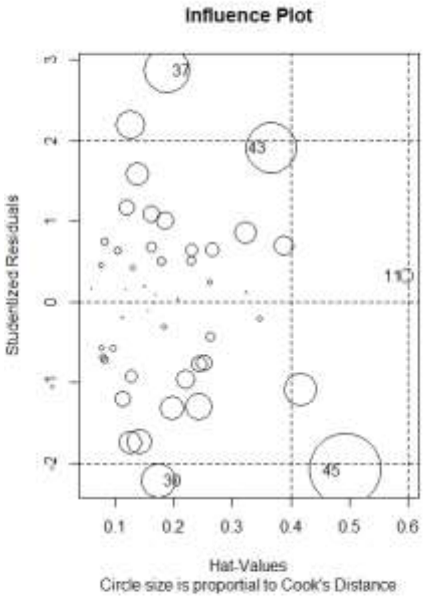


Figure 4: Influence Plot, uses Cook's distance to determine the strength of variables on the model. Variables exceeding 4 are removed per data protocol.

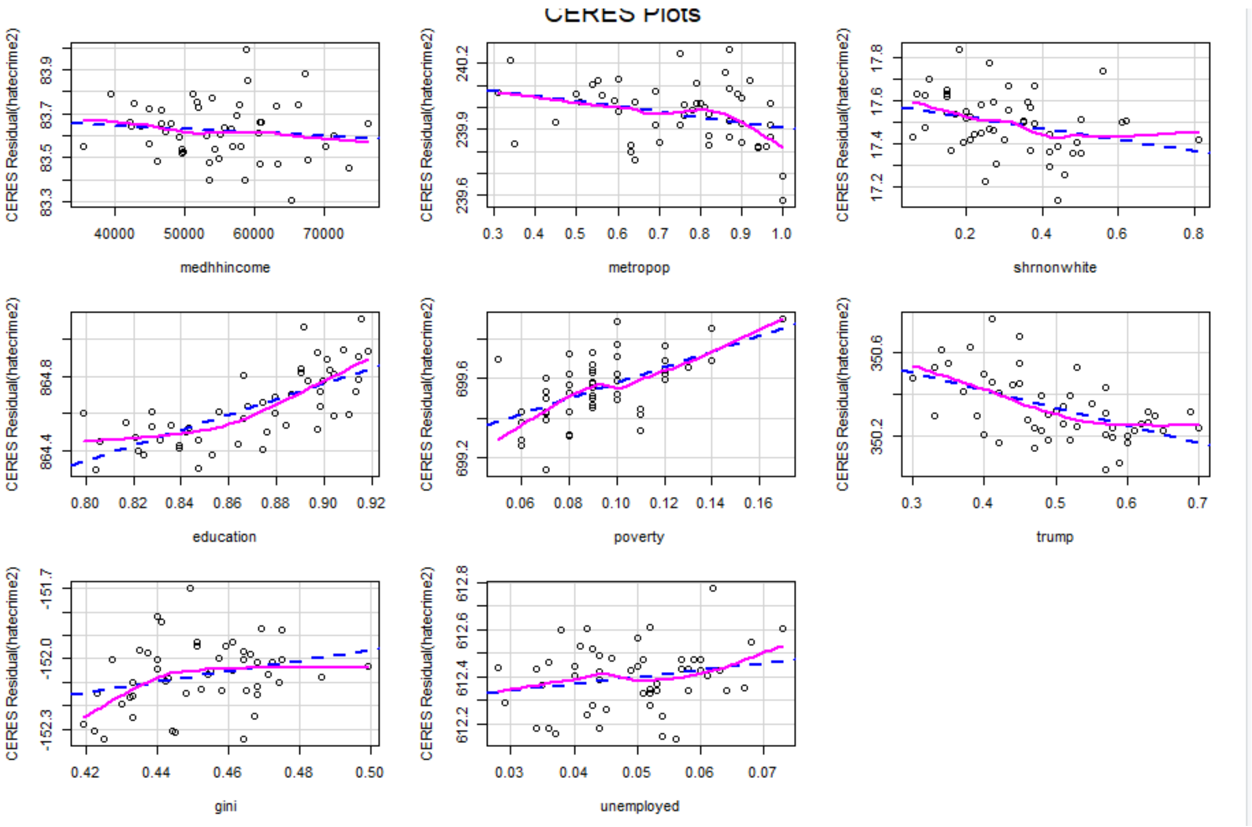


Figure 5: Ceres Plot, demonstrate linearity. Anderson Darling Test was run to confirm visual results and was non-significant, indicating distribution was not normally distributed ((A = 0.25, p > 0.74). Shapiro tests confirmed nonlinear variables in the model as visualized in the Ceres plots.

Test 1: Predictors of Hate Crimes - Model 2

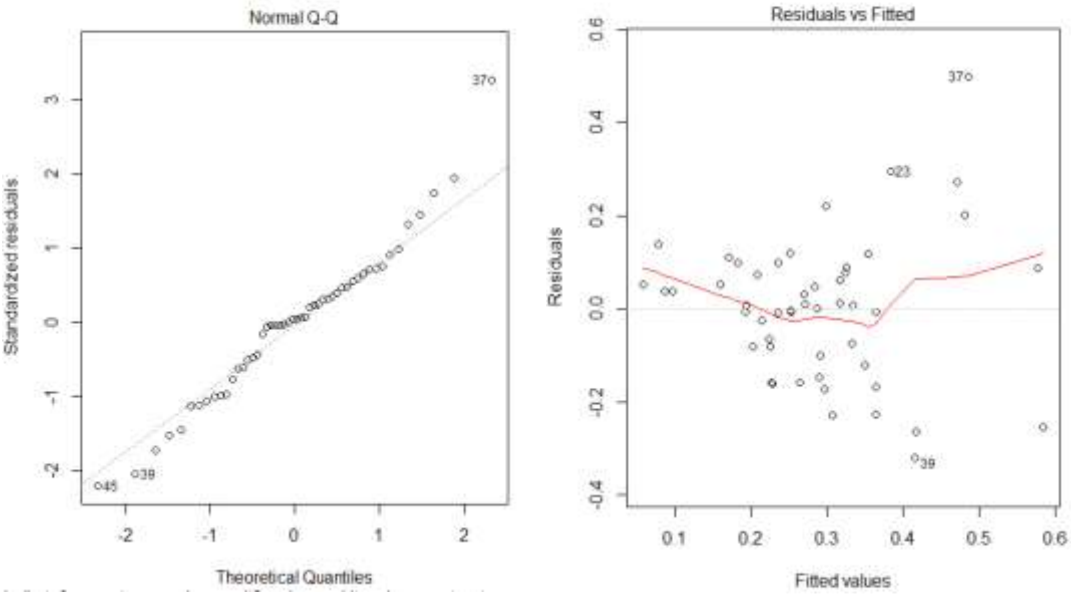


Figure 6: Indicates the residual curve of Model 2, relatively linear, no outliers present.

Model	AIC Score
hate3 ~ +metropop + lognoncit2 + shrnonwhite + logpoverty + trump + gini + unemployed	Start: AIC=-171.94
hate3 ~ metropop + lognoncit2 + shrnonwhite + logpoverty + trump + unemployed	AIC=-173.68
hate3 ~ metropop + lognoncit2 + shrnonwhite + logpoverty + trump	Step: AIC=-175.49
hate3 ~ metropop + lognoncit2 + shrnonwhite + trump	Step: AIC=-176.09
hate3 ~ metropop + shrnonwhite + trump	Step: AIC=-177.13
hate3 ~ shrnonwhite + trump	Step: AIC=-178.41

Table 5: Candidate Models from Stepwise AIC , first row indicated the initial model and as seen from the AIC scores, has the lowest value and was maintained.

Test 2: Hate Crime Occurrence by Region - Model 1

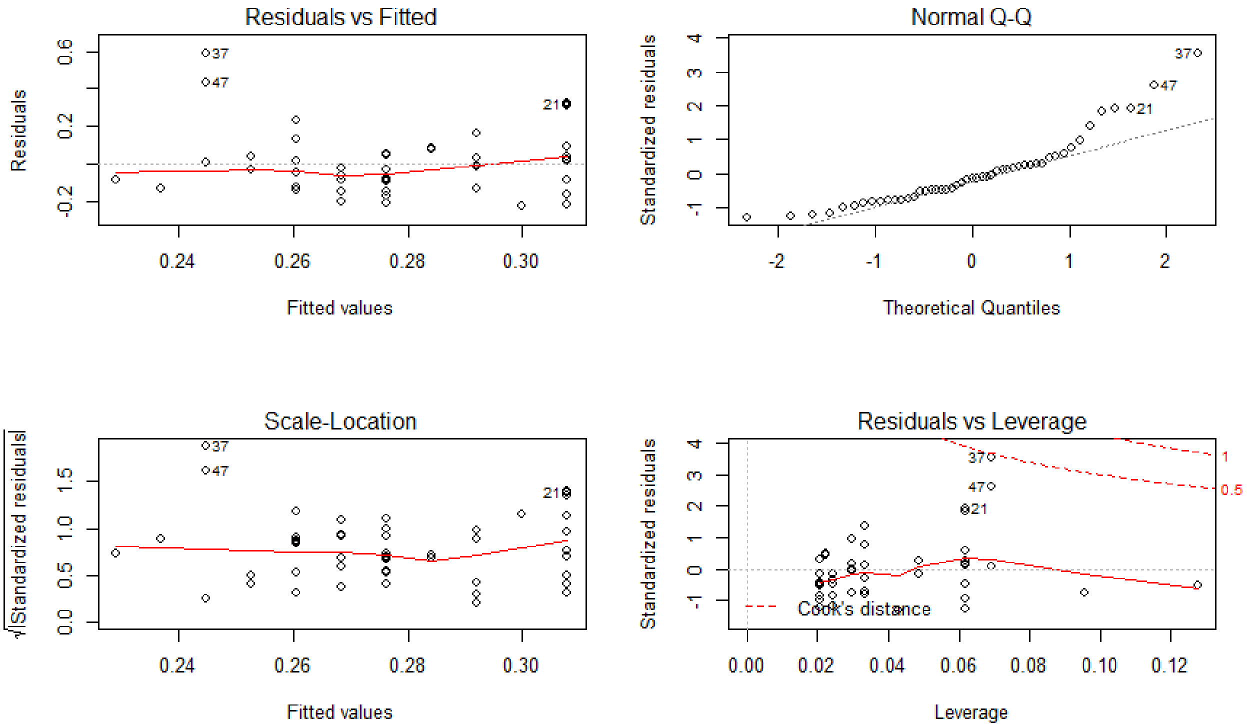


Figure 7. Residual plots for Test 2, model 1. The plot indicates potential non-normality and outliers. Anderson Darling Test confirmed non-normality ((A=1.67, p<0.002), and Bonferroni test indicated a significant outlier (obs = 37, r-student = 4.08, p < 0.002, Bonferonni p = 0.009).

Test 2: Hate Crime Occurrence by Region - Model 2

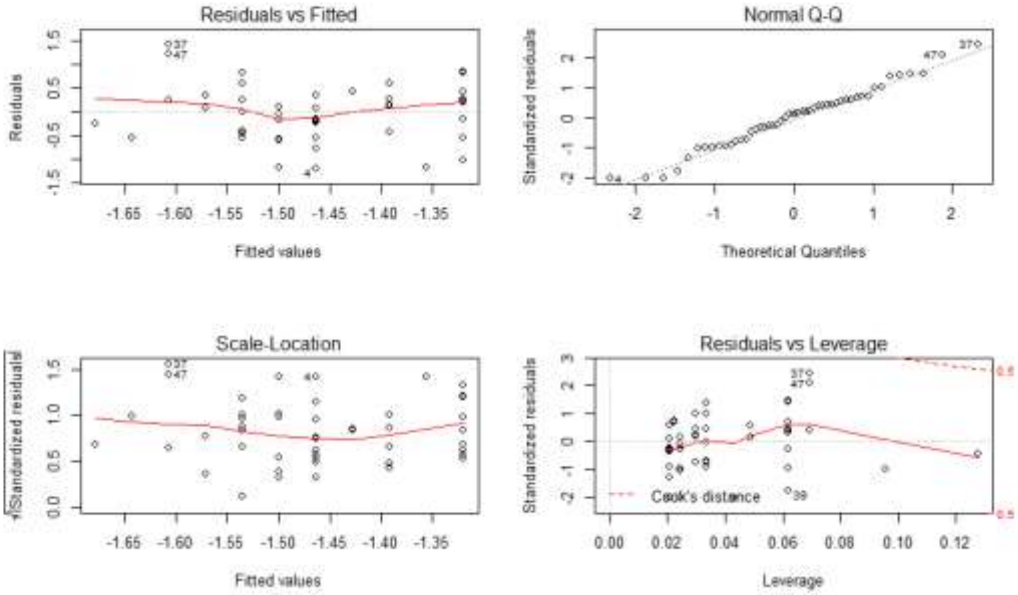


Figure 8: Residual plot for Test 2, model 2. The plot demonstrates a normal curve and reduction of the outlier. Anderson Darling Test confirmed normality (A=0.29, p > 0.61) and the Bonferonni test confirmed the reduction of the outlier obs=37, r-student = 2.583, p > 0.013, Bonferonni p = 0.638).

Test 3: Predictors of High Share of Trump voters using socio-economic data - Model 1

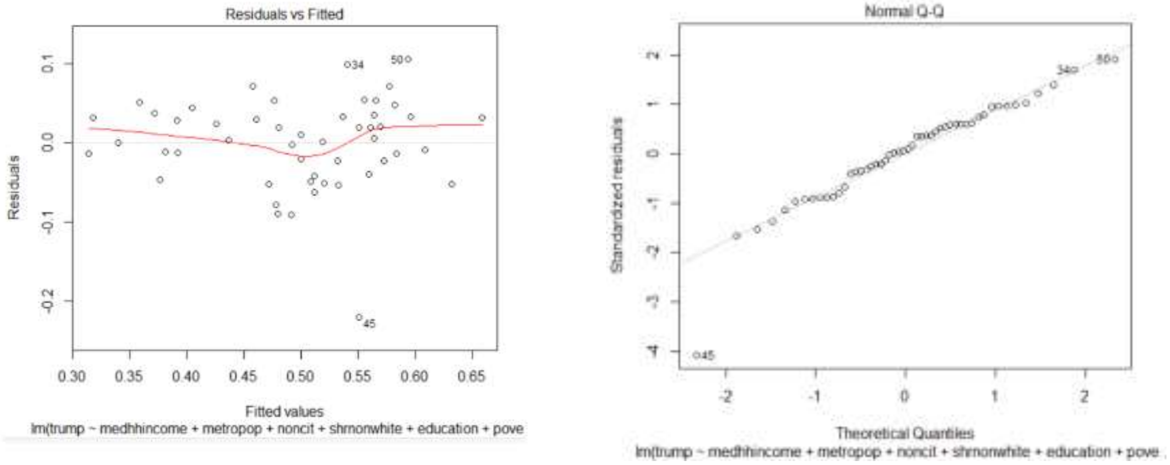


Figure 9: Residual plots , testing for normality and outliers. Bonferonni Tests confirmed obs (obs = 45, r-student = -5.25, p < 5.27e-06, Bonferonni p < 0.0002), and the NCV test confirmed heteroscedasticity ($X^2 = 2.50$, $p < 0.05$).

Test 3: Predictors of High Share of Trump voters using socio-economic data - Model 2

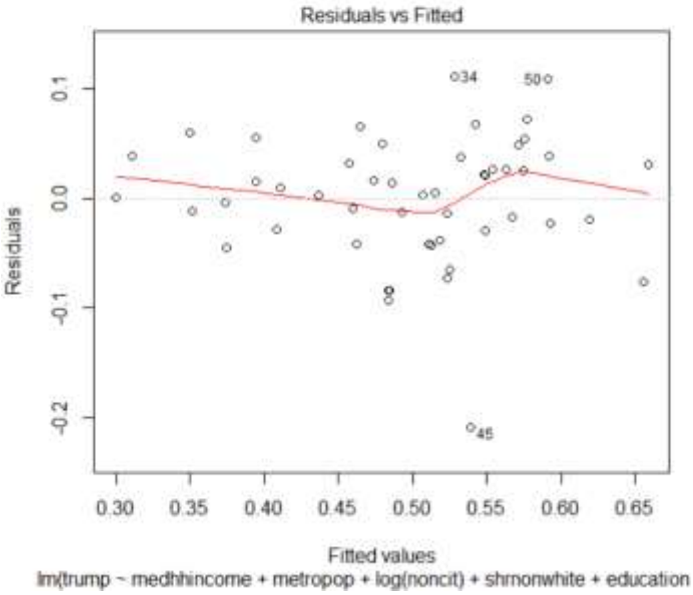


Figure 10: Indicates Model 2 residuals. NCV test was insignificant for heteroscedasiticy ($X^2 =$ 8.72e-06, $p < 1.0$).

Model	AIC Score
trump ~ medhhincome + metropop + log(noncit) + shrnonwhite + education + log(poverty) + log(hatecrime) + gini	Start: AIC=-270.08
trump ~ medhhincome + metropop + shrnonwhite + education + log(poverty) + log(hatecrime) + gini	Step: AIC=-271.91
trump ~ medhhincome + metropop + shrnonwhite + log(poverty) + log(hatecrime) + gini	Step: AIC=-272.17
trump ~ medhhincome + metropop + shrnonwhite + log(poverty) + log(hatecrime)	Step: AIC=-272.48

Table 6: Candidate Models from Stepwise , the lowest value was selected