Kelvin Nguyen &

CIS 3920

Term Project Report

Anti-Lgbtq Hatecrimes and Anti-Lgbtq Legislation in 2021

##### **Problem**

Does the number of anti-lgbtq hate crimes correlate to number of anti-lgbtq legislation passed within the same year?

What contributes to such high lgbtq hate crimes?

##### **Introduction**

Within the past couple of years, there has been a huge anti-lgbtq movement mainly from right-wing / religious groups. In addition, with the Supreme Court’s overturning of Roe v.Wade, Supreme Court Justice Samuel Alito referred to overturning Obergefell v. Hodges (legalizing gay marriage). In addition, in 2022 and 2023, there has been a huge wave of not only anti-lgbtq legislation being passed, but huge turnovers of gender-affirming care. An alarming example was on April 19th of 2023, Florida’s senate passed bill CS/SB 254 to the house with 82 YEAS and 31 NAYS. This law grants the state of Florida temporary emergency custody over a child that had recently been subjected to sex-reassignment prescriptions or procedures, thus removing custody of a transgender child from their parents. Out of concern, we thought it would be interesting to see if there is a correlation between the number of hate crime incidents that occur and are correlated respective to one’s state’s government / legislation. Thus we would need a total number of laws per state and hate crimes per state that pertain to anti-gender/ non-gender conforming/ gender identity related crimes.

##### **Project Description/ Abstract**

##### 

We decided to do this project because it has been a politically stigmatizing topic even though it shouldn’t be political. This project would steer a path towards human rights and experiment to see if these laws contribute to such discrimination and eventually towards an increased crime rate towards these groups of people. We trained regression, support vector machines, and random forest models in order to test this hypothesis.

##### **Dataset**

The datasets we used are from the FBI’s 2021 hate crime database and the ACLU’s 2021 anti-lgbtq legislation tracker as well as Ballotpedia’s state level government majorities. Such contained information such as:

State Incidents Crime Type

Bill Number Bill Type Date

Incident Bias Victim Type Percent Democratic

Percent Republican Senate Majority House Majority

##### **Literature Survey**

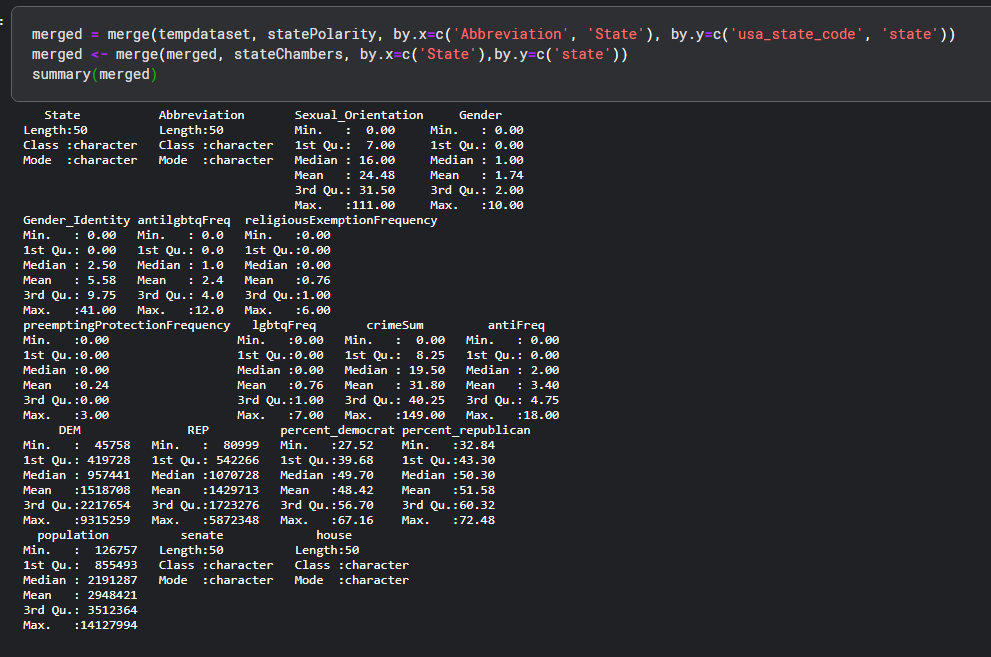
Although there are many companies that keep track of anti-lgbtq legislation, yet anyone has tested the hypothesis of whether or not these laws actually contribute to increased hatecrimes. Given that it would seemingly be obvious, it can be argued that most of those who ideologically align as anti-lgbtq would not keep up with all legislation being passed so there is a high possibility of no correlation, which companies would be hesitant to enact upon. Yet, here our goal was to do this, even if it would be either ‘obvious’ or common knowledge purely for analytical purposes. Most activism companies such as the ACLU make sure to spread awareness on these types of bills in order for citizens to use their voices towards their local congressmen and women. Most of these bills do in fact restrict the freedom of those who have sought gender-affirming care or went through gender-reassignment procedures. For instance, many transgender people are unable to compete in sports or use public restrooms or even be allowed to express themselves if they are a minor. Such behavior could lead to conversion therapy which is incredibly mentally taxing upon the youth to dismiss their expression. However our project does not consider this into our scope, we purely are just testing for a relationship between the number of legislation and the number of anti-lgtbq hate crimes in the 50 states of America.

##### **Project Methodology**

1. We analyzed the data provided within all 14 FBI tables, using mainly table 13’s hate crime per bias motivation per state.
2. We grouped said data by state.
3. We used webscraping to retrieve the ACLU's anti-lgbtq legislation tracker from the following website: <https://www.aclu.org/legislation-affecting-lgbtq-rights-across-country-2021>.
4. Again, we grouped by state and made a count per state, then aggregated both data frames into 1 big one.
5. Then we took the Ballotpedia’s federal election and state government statistics into the dataframe:
   1. Number voted for democrats
   2. Number voted for republicans
   3. Percent democrats
   4. Percent republican
   5. Senate majority (categorical of rep or dem)
   6. House majority (categorical of rep or dem)
6. In Regression we used libraries:
   1. Ggplot2
   2. tidyverse
7. In SVM we used Libraries:
   1. e1071
   2. caTools
8. In Random Forest we used libraries:
   1. randomForest
   2. Datasets
   3. Tidyverse
   4. Corrplot
   5. Caret
   6. Grid
   7. grid extra
   8. caTools
   9. Rpart
   10. Rpart.plot
   11. GGally
   12. dyplr

##### 

##### **Analysis and Results**

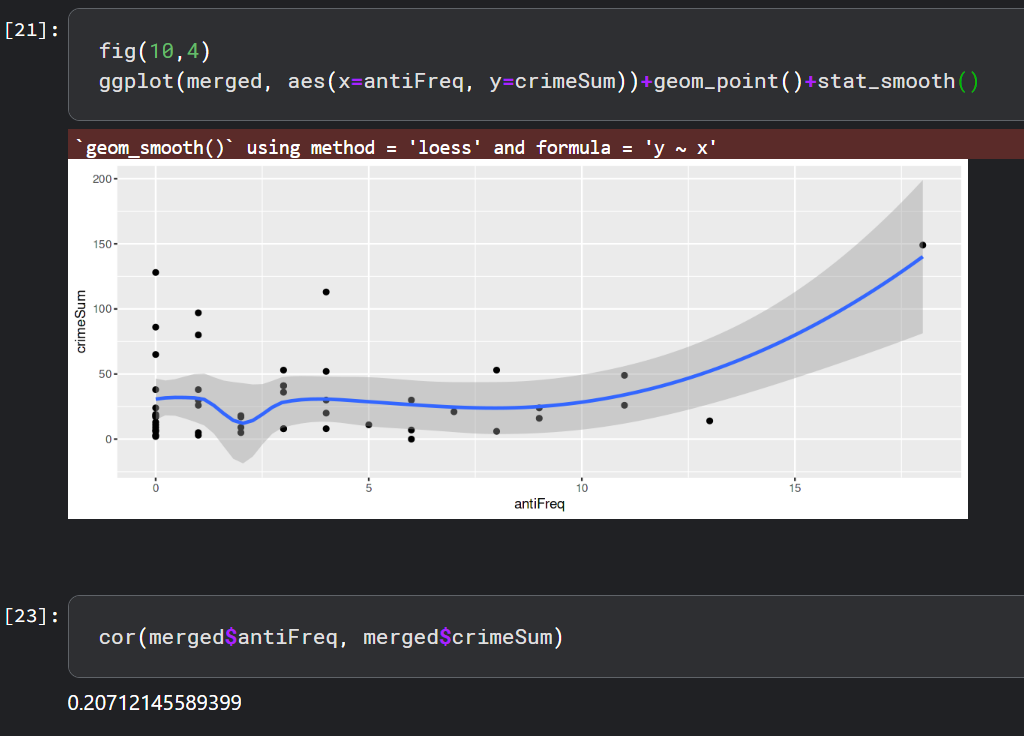


This is the summary of the dataset when fully aggregated named merged. This includes hatecrime data per state, total laws (and types), population, # and % of either democrat or republican per state, and senate/house majority.

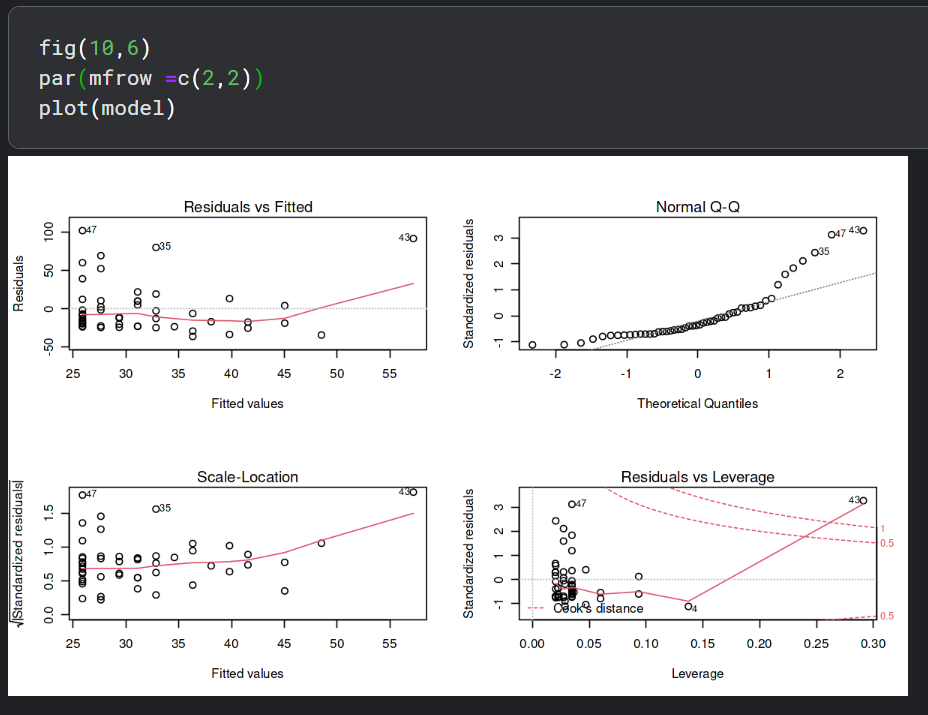
Total anti-lgbtq hatecrimes equate to 1590, and 170 anti-lgbtq laws passed.

There is an average of 31.8 anti-lgbtq hatecrimes per state and 3.4 anti-lgbtq legislation passed.

#### **Regression:**

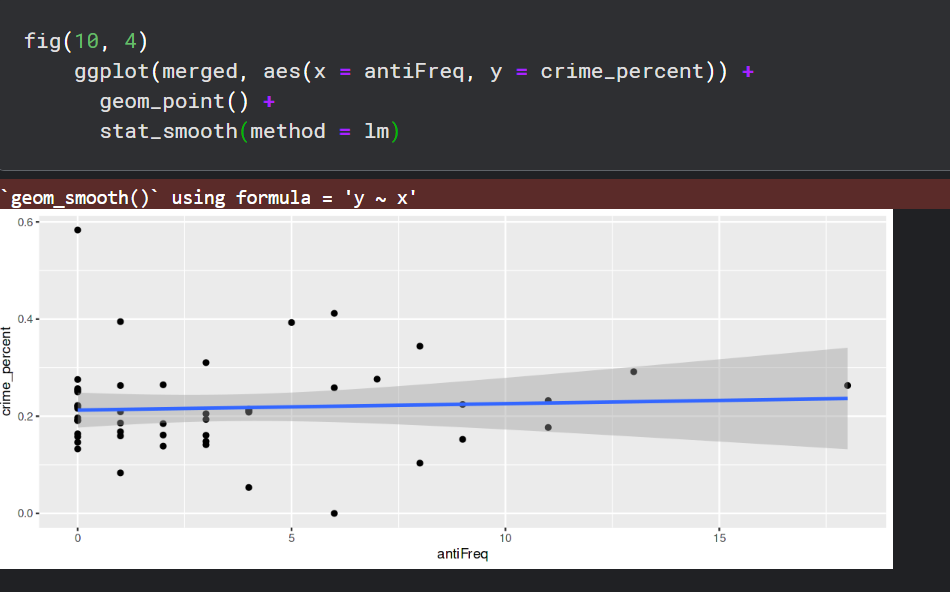


In the image above, it is evident that linear regression was not successful given that there was a very high value of p of 0.14 and low correlation coefficient of .207. Therefore it is safe to assume that there is no linear correlation between the number of anti-lgbtq laws passed and anti-lgbtq hatecrimes. Below are the residuals and Q-Q plot.

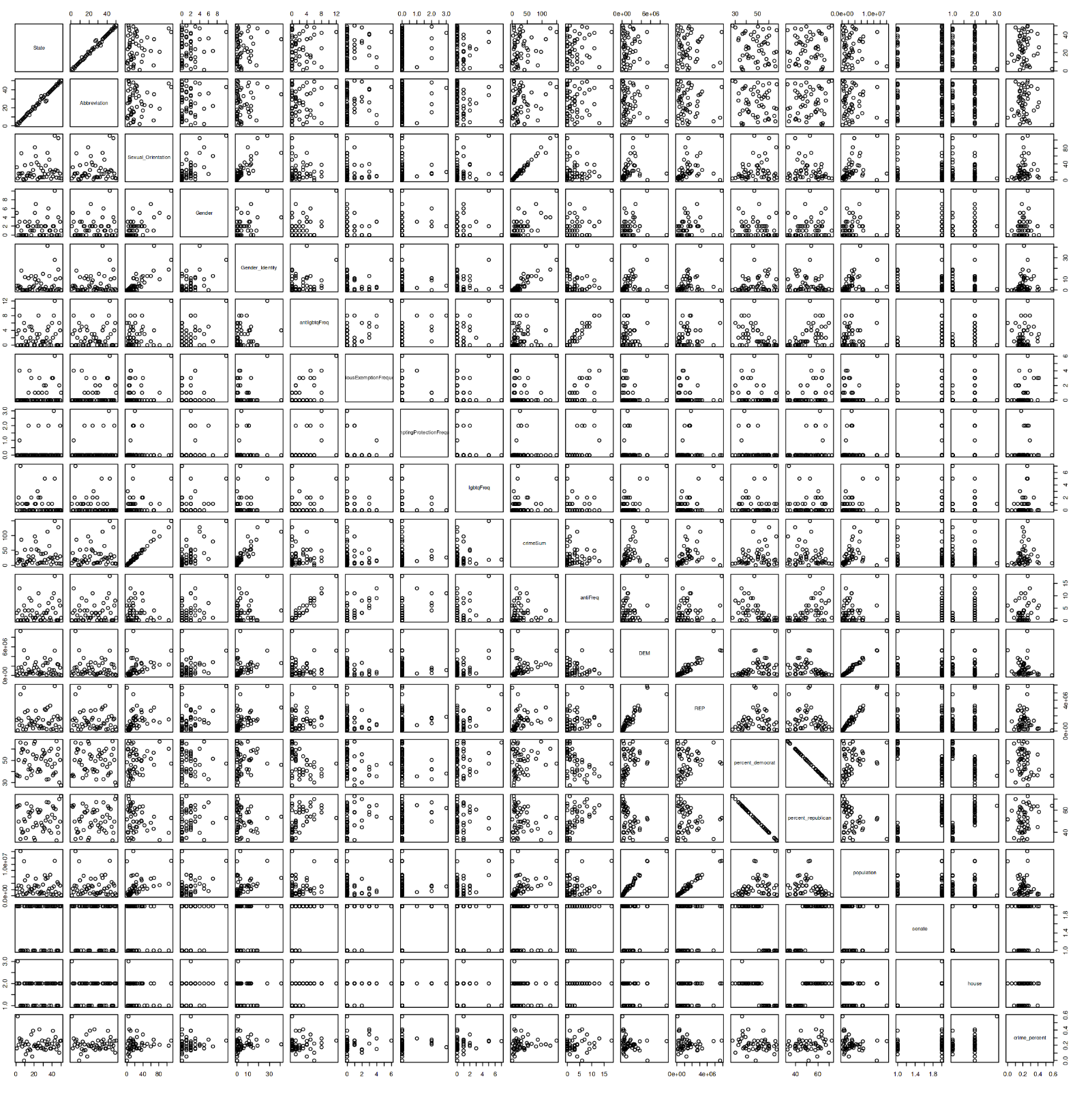


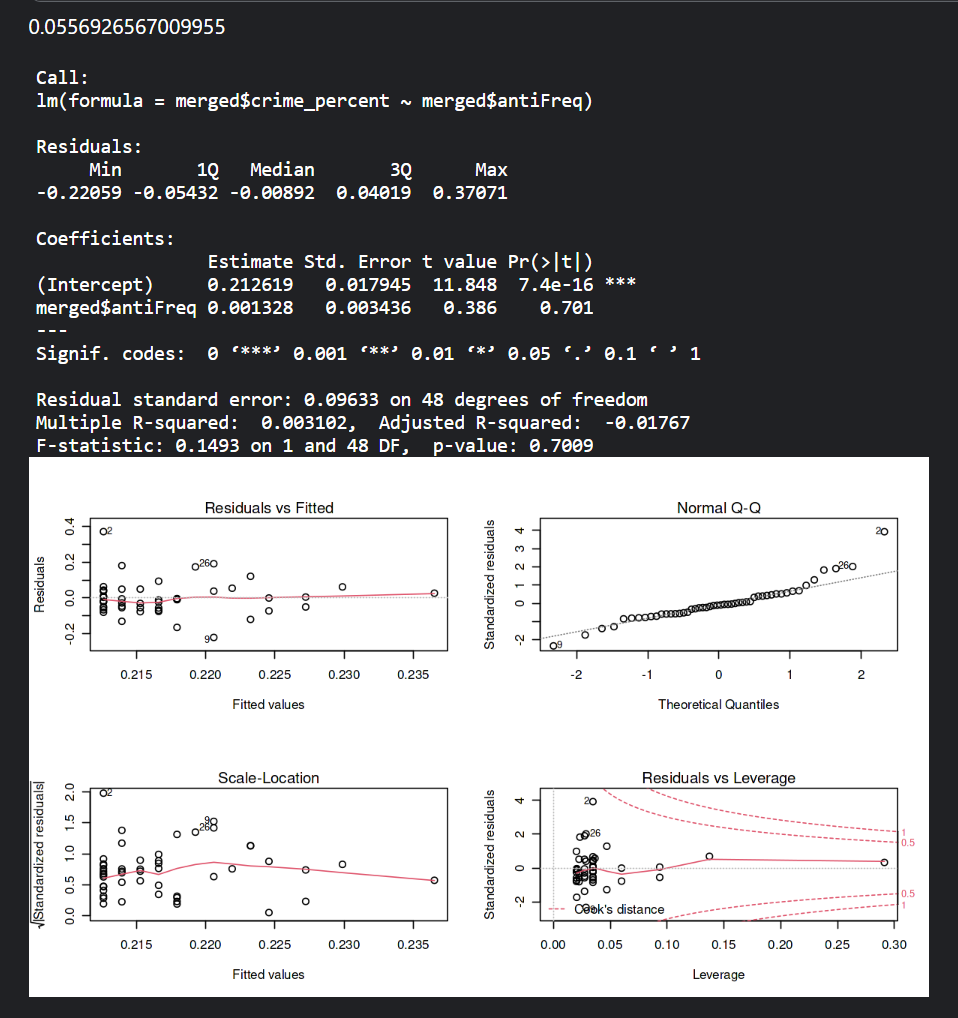
As we can see here, simple linear regression is too sensitive to the data here to outliers. In addition, the Q-Q plot has data points that somewhat fall on the line but there is too much data that simply do not fit into this distribution with high outliers. Considering this model had an RSE of 33.29 on 48 degrees of freedom, this model performed poorly.

So we thought of a solution using lgbtq crime as a percentage of all hate crimes per state and modeled it.



It performed poorly too. The correlation coefficient came out to be .055 and here is the summary of the model. Therefore it was concluded that there is no linear correlation for this dataset.

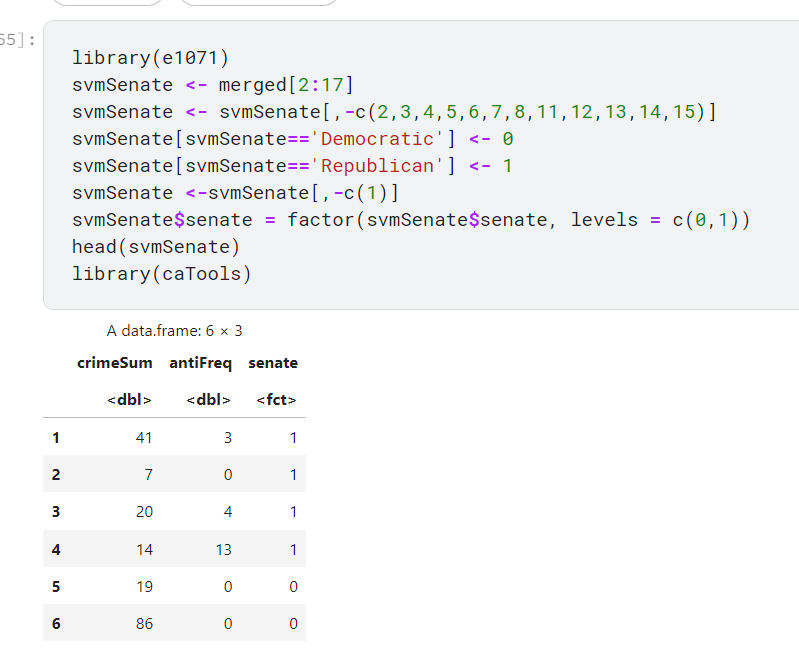




#### **Support Vector machines:**

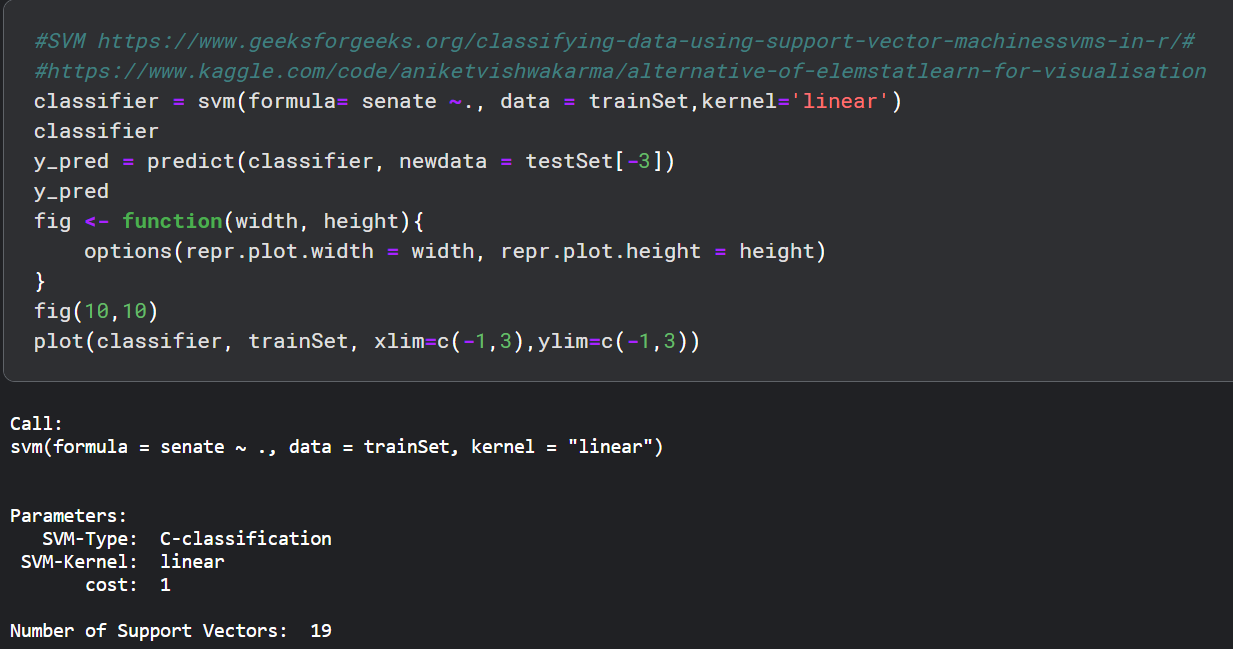
First I labeled each state based on their senate majority. 0=democrat 1= republican

Using hate crimes as response variable and total anti-lgbtq as predictor variable:

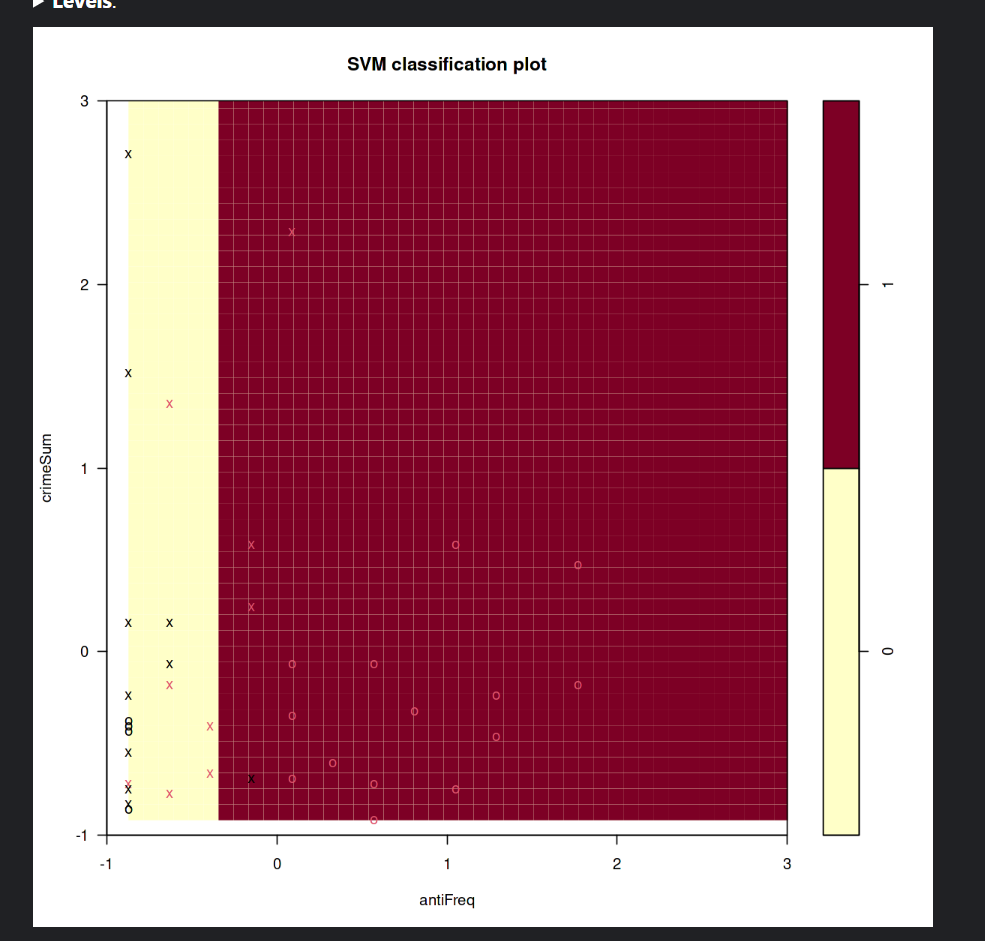


Now we split the data with a .75, .25 ratio

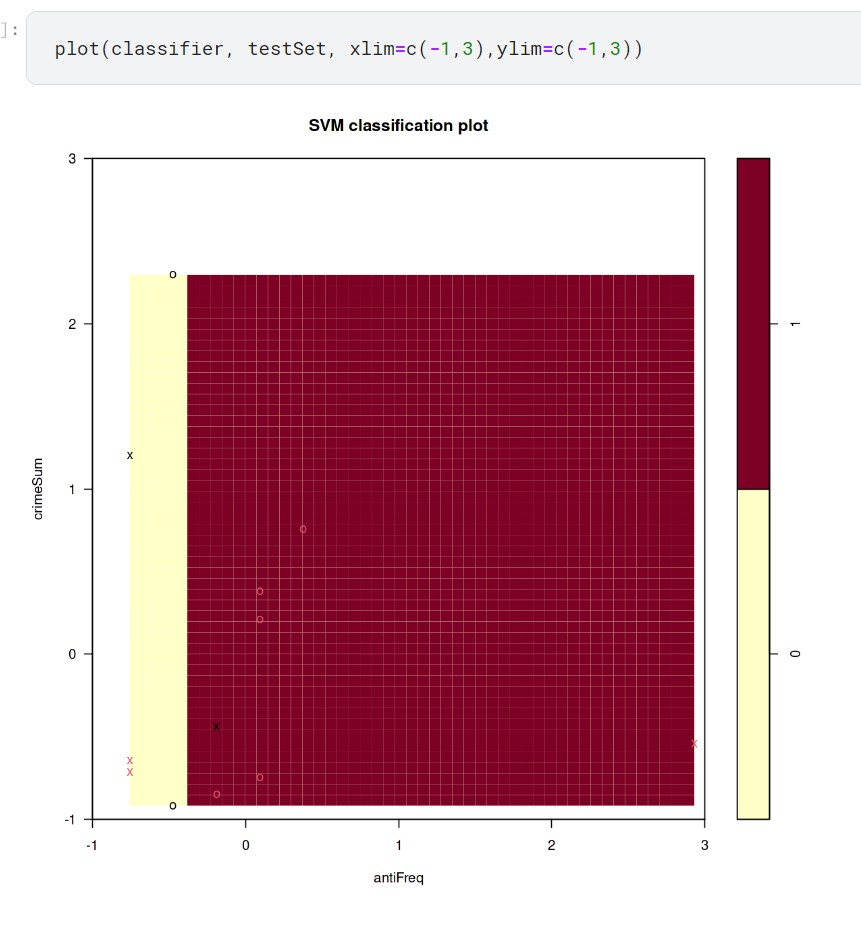




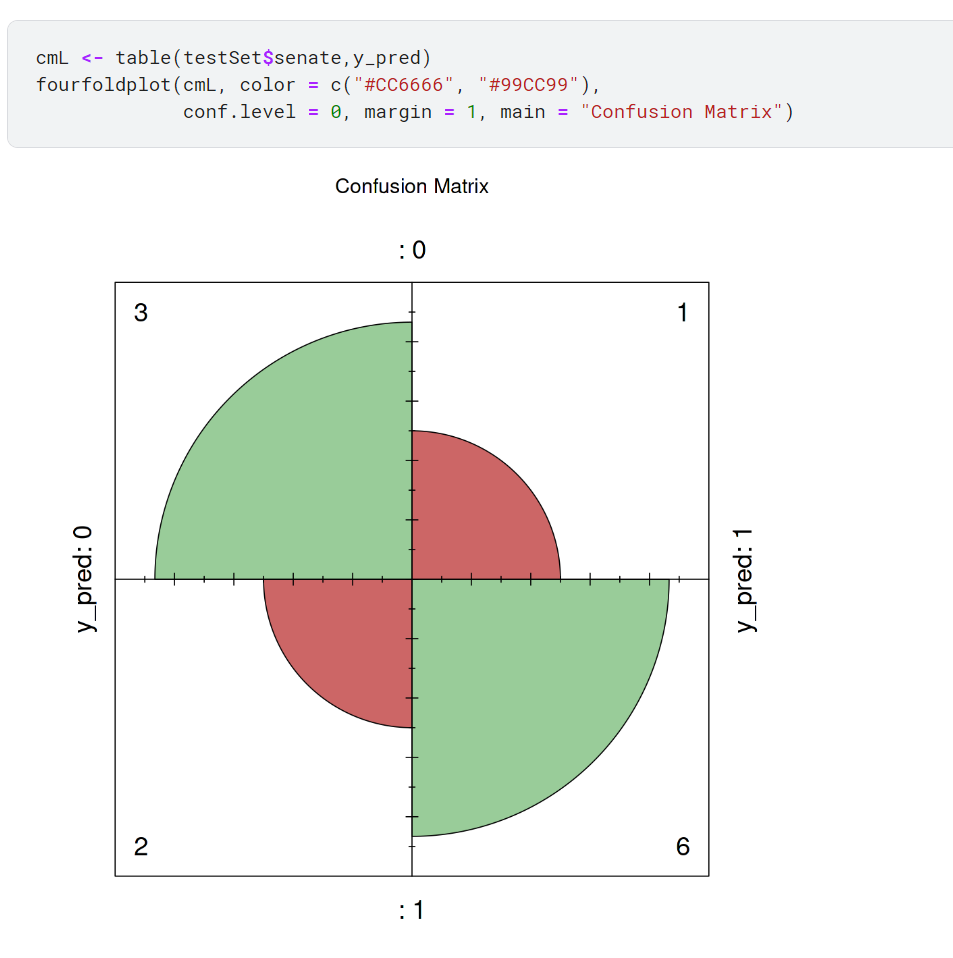
Train set using strictly 0= democrat and 1= republican, crimeSum as response variable.



Test set results:



Confusion matrix for test set:



In the confusion matrix above:

remeber that 0 is Democrat and 1 is Republican majority Senate

- 2 false democratic

- 1 false republican

- 3 correct for democratic

- 6 correct for republican

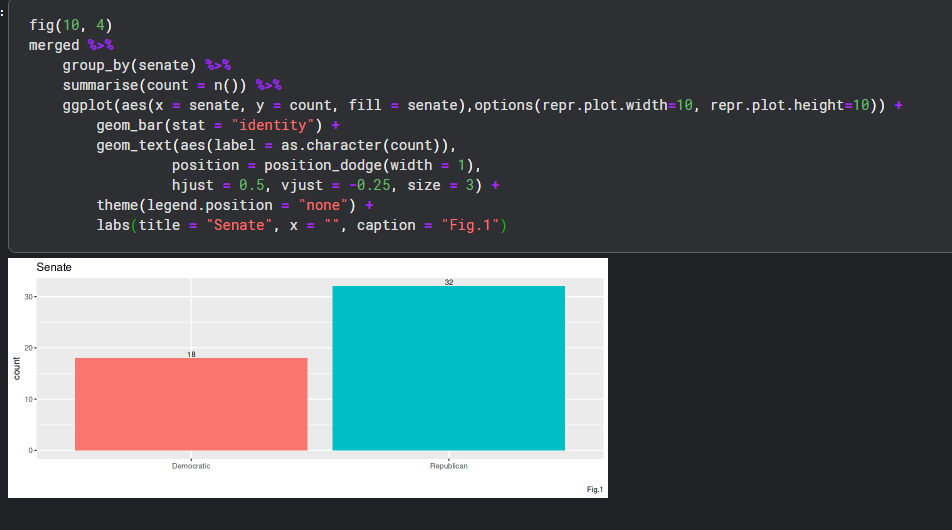
- 75% accuracy rate

- 8.33% false republican rate

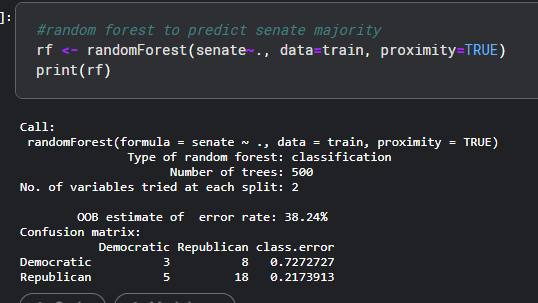
- 16.67% false democratic

I will discuss these results in the next section of the paper, but I still need ot present the results for the random forest model.

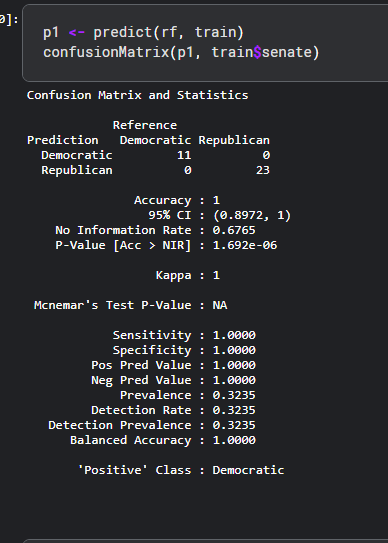
#### **Random Forest, plot of actual DEM vs REP senate counts**



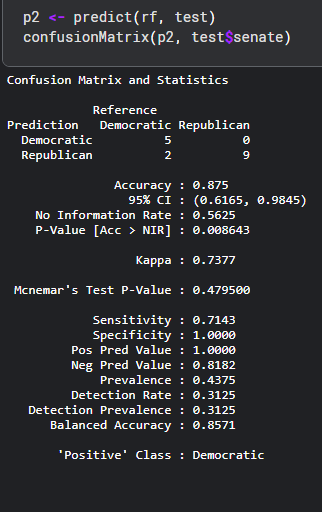
Dataset used to predict, removed Democratic count for 2020 election, Republican count, to avoid better predictions.



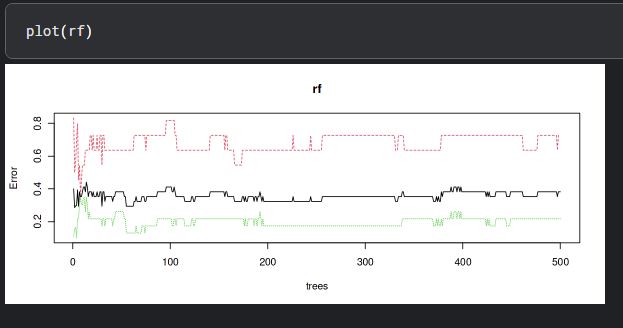
Prediction for train: 100% accuracy

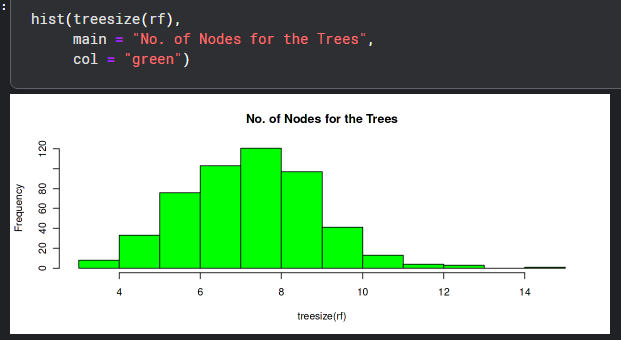


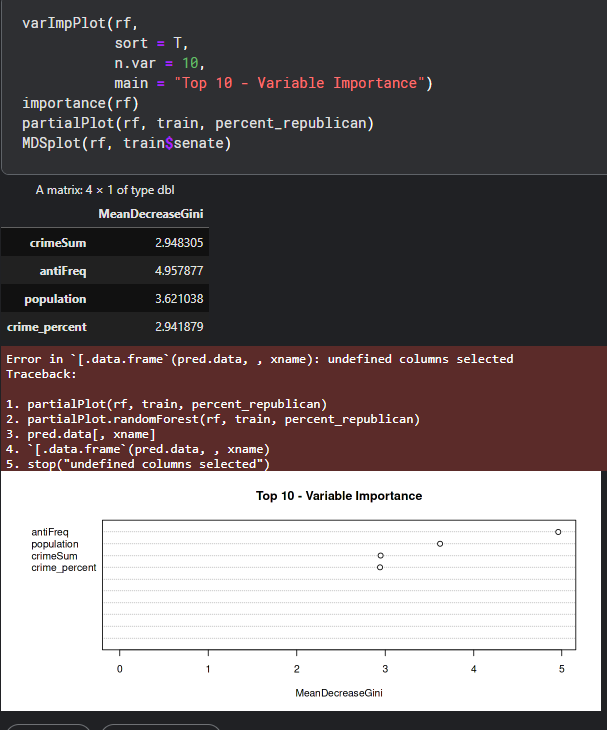
Prediction for test 87.5% accuracy rate



Random forest graph







As we can see from the results, Anti-Freq laws had the most importance when splitting the data into democratic/republican majority states.

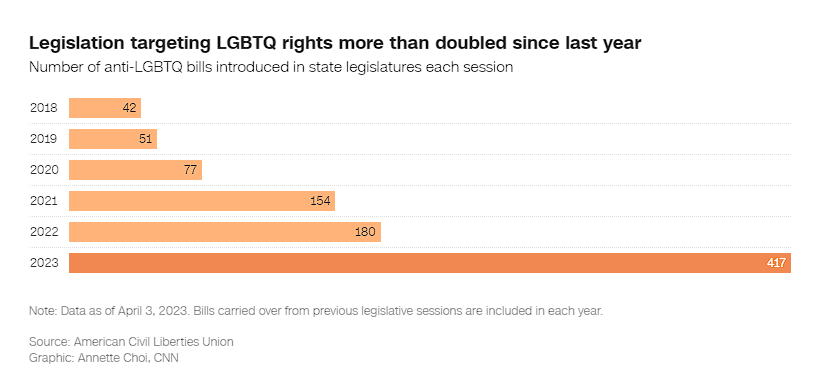
##### **Summary:**

Out of the three models, the only one with success was random forest in order to classify senate majority of the state, which contributes to the state’s laws. It was able to have a 100% training set prediction rate and 87.5% testing set prediction rate. Else, regression and SVM models performed very poor. In the random forest, predictors included total anti-lgbtq hate crimes and total anti-lgbtq legislation grouped by state, ass well as population and number of presidential democratic and republican voter count, and lastly lgbtq hate crime percentage compared to cumulative hate crimes. This is considered a success, being that the most influential variable according to the GINI index would be antiFreq which was the count of anti-lgbtq laws.

##### 

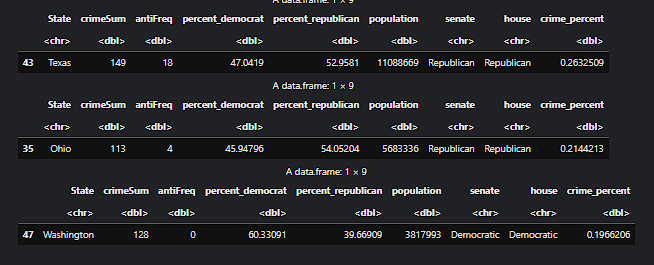
##### **Struggles and Roadblocks:**

The main priorities I had with this project was assessing the years 2022 and 2023, for having back to back record breaking anti-lgbtq legislation. Prior to 2021, we experienced a pandemic which significantly decreased hatecrimes for that year due to quarantine. In addition, just within the first quarter of 2023, there has been over 400 anti-lgbtq legislation, where in 2021 there was 158, and in 2022 there was 180, according to the [ACLU](https://www.cnn.com/2023/04/06/politics/anti-lgbtq-plus-state-bill-rights-dg/index.html)’s data depicted by CNN below.



However, when I requested datasets from the FBI for either 2022 or whatever was recorded of 2023 hatecrimes, they were unable to give me the information. In addition, a major roadblock encountered was the ACLU’s data for 2021, stating that they were unable to give me datasets for that too, therefore a significant portion of my efforts on this project included data scraping off of their [public archives](https://www.aclu.org/legislation-affecting-lgbtq-rights-across-country-2021).

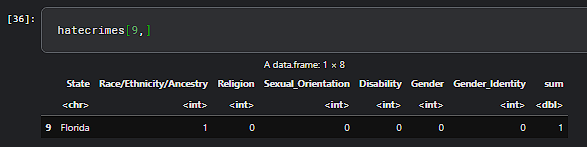
Also, it is important to mention that there is a potential huge margin of error for hatecrimes based on how state agencies report them. You can find more information at this [link](https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/hate-crime). There could be a huge number of crimes that go unreported each year due to intimidation or overlooking certain aspects. There is potential that one’s own ideological beliefs or tolerance for the acknowledgement of other’s beliefs may have led to either overreported or underreported crimes. For instance when running the regression for total crime sums I ran into many outliers.



Some major outliers included Ohio and Washington. Both of which had very little anti-lgbtq laws passed that year yet a significantly high number of anti-lgbtq hatecrimes.

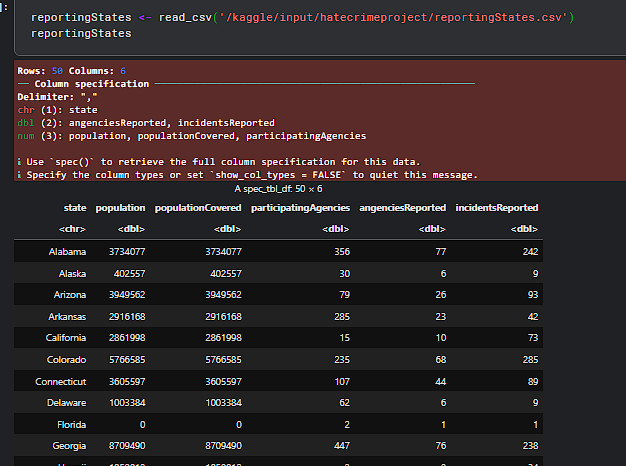
Texas as well had the second most hatecrimes at 566 (Washington at first), yet Texas was also an outlier due to how high the laws were and how high the crime was.

But the most interesting state was Florida.





Florida had no anti-lgbtq laws and no hatecrimes of any type reported. For such large population state:



The FBI database actually shows us that only 2 agencies reported for Florida whereas you could see other states having significantly more. Therefore Florida created a huge problem for all my models and such poor reporting. Perhaps these issues must be addressed in the future.

##### **Conclusion:**

To conclude: there is no direct relationship as of 2021’s lgbtq hatecrime’s and anti-lgbtq legislation. Therefore my original hypothesis was wrong and H0 correct. However, in the experiment with running random forest, it is at least relevant in the context that there is a correlation between increased rate of anti-lgbtq laws and increased number of lgbtq hate crimes. Although I was not able to experiment with 2022’s data, this could potentially attribute to future experiments. I definitely will return to this project, in an expected two years when the data is publicized by both the ACLU and FBI. Regardless of these results, it is clear that many will propose this project to be too politically aligned. However, in the context of human rights, anti-lgbtq and specifically this new wave of anti-transgenderism within American government creates no benefit towards anyone in any form. People should be free to express themselves in a form that does not harm others. In addition, this project shows the flaws within Florida’s hate crime reporting, which should be addressed for the benefit of those harmed. Without the necessary data, more problems will build on without knowing its root. Perhaps in the future a potential project could be using politicians’ themselves as a source of crime rate or legislation passed. For instance, many of those in congress are heavily present on media outlets such as Twitter. Therefore, their activity could determine potential laws to be passed or a trend in certain legislation and how it affects crime rate.

##### **Contributions:**

Though many corporations and media outlets heavily follow anti-lgbtq legislation, work has yet to be done on the relationship between hate crimes and lgbtq legislation passed.

**References:**

<https://www.aclu.org/legislation-affecting-lgbtq-rights-across-country-2021>

<https://www.aclu.org/legislative-attacks-on-lgbtq-rights>

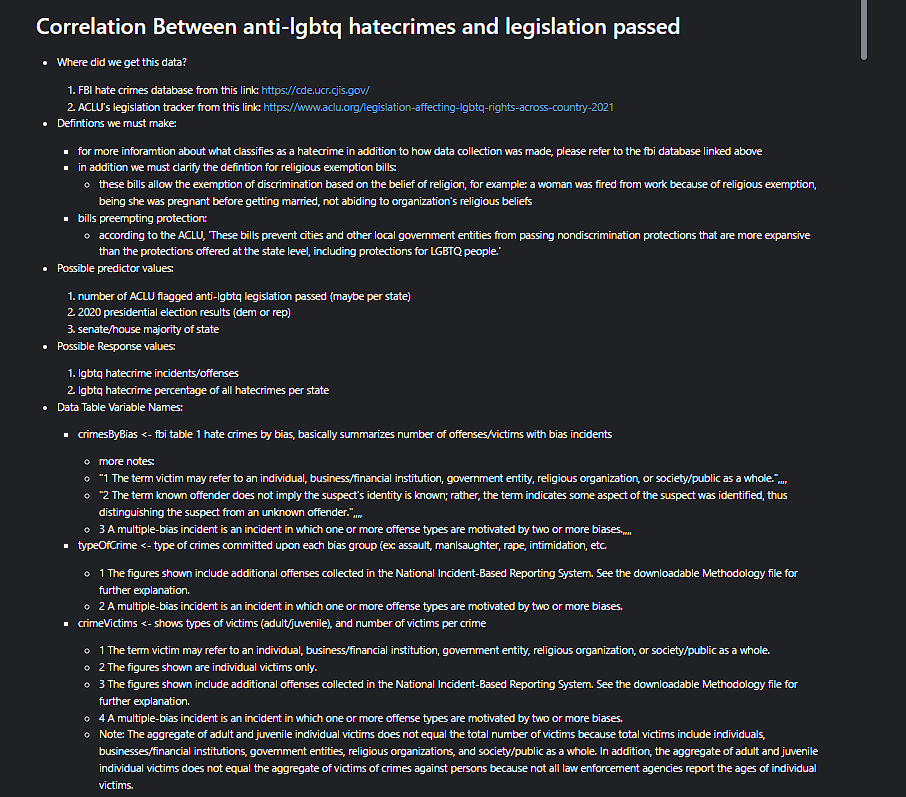
<https://www.cnn.com/2023/04/06/politics/anti-lgbtq-plus-state-bill-rights-dg/index.html>

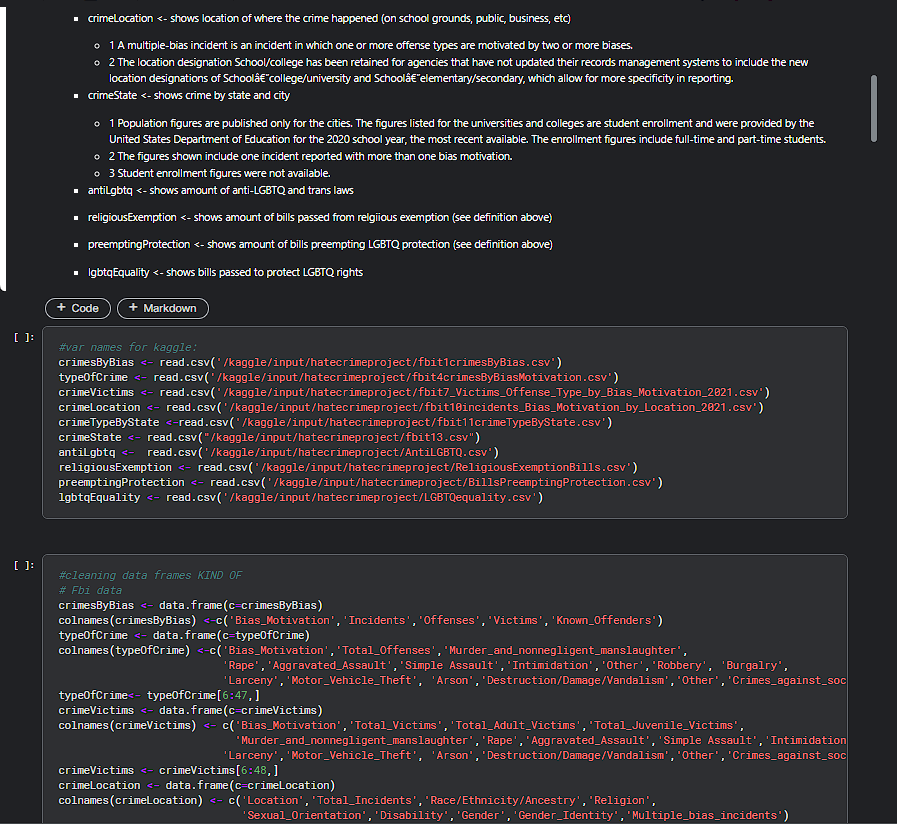
<https://www.r-bloggers.com/2021/04/random-forest-in-r/>

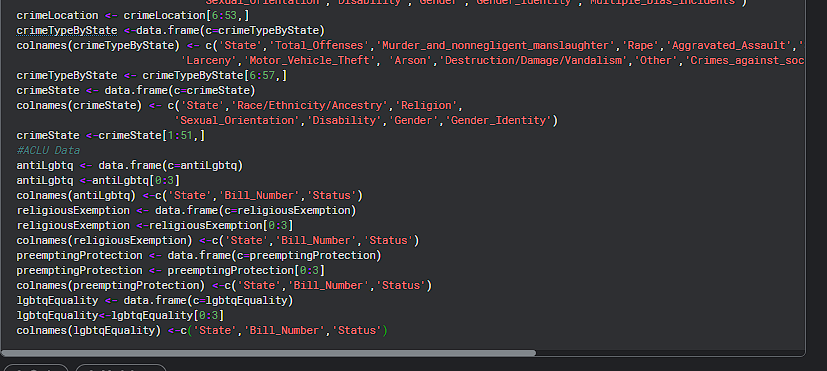
<https://www.kaggle.com/datasets/paultimothymooney/percent-voting-for-democratic-party-by-state>

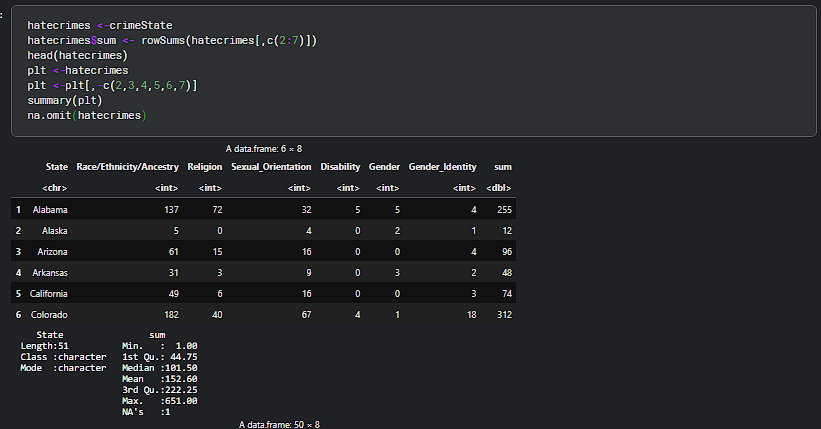
##### **Appendix:**

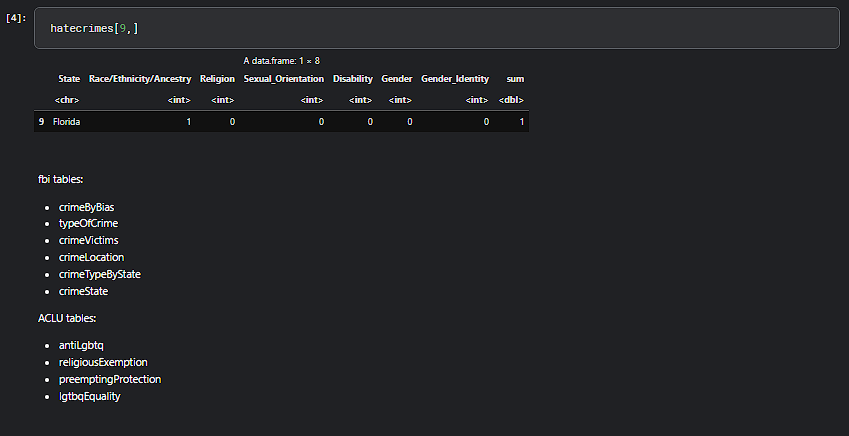
[**https://www.kaggle.com/code/kelvinnn25/cis3920project**](https://www.kaggle.com/code/kelvinnn25/cis3920project)

****

****

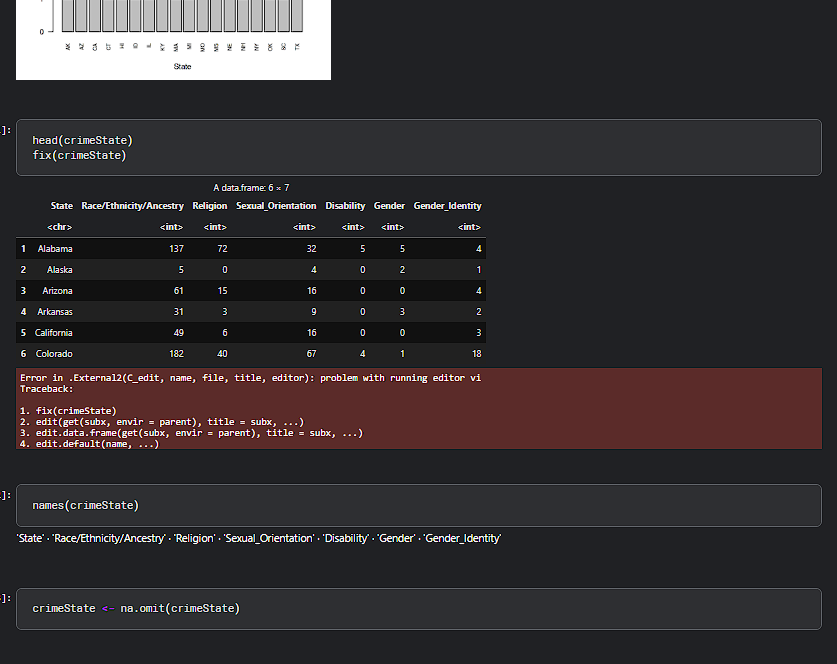
****

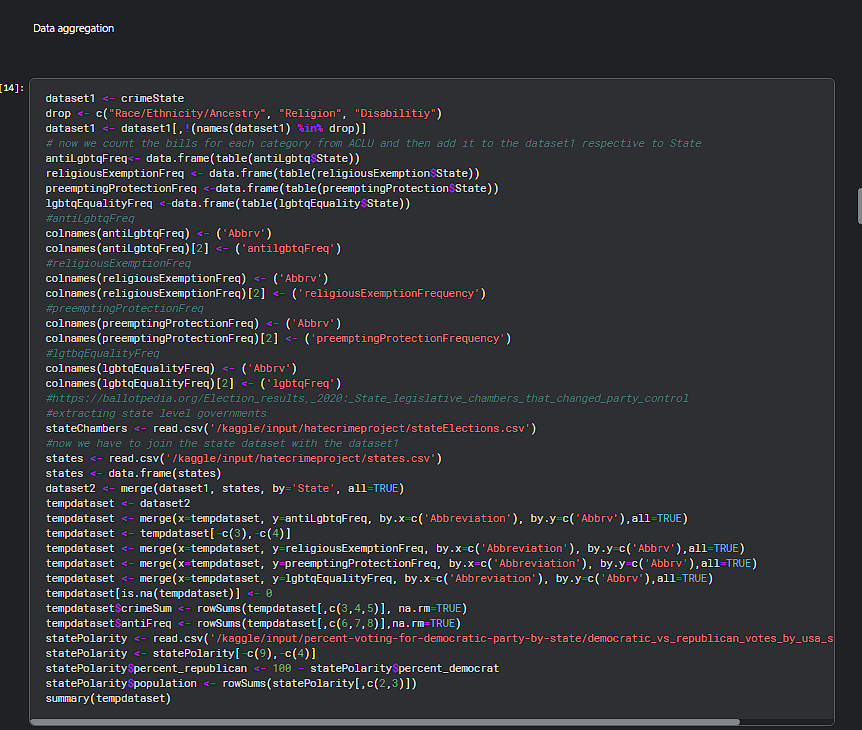
****

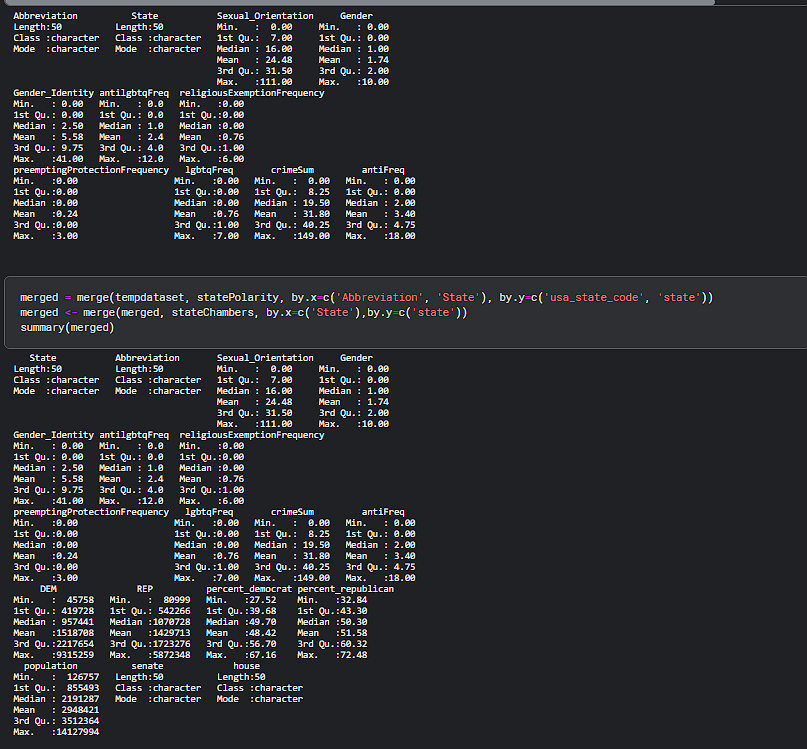
****

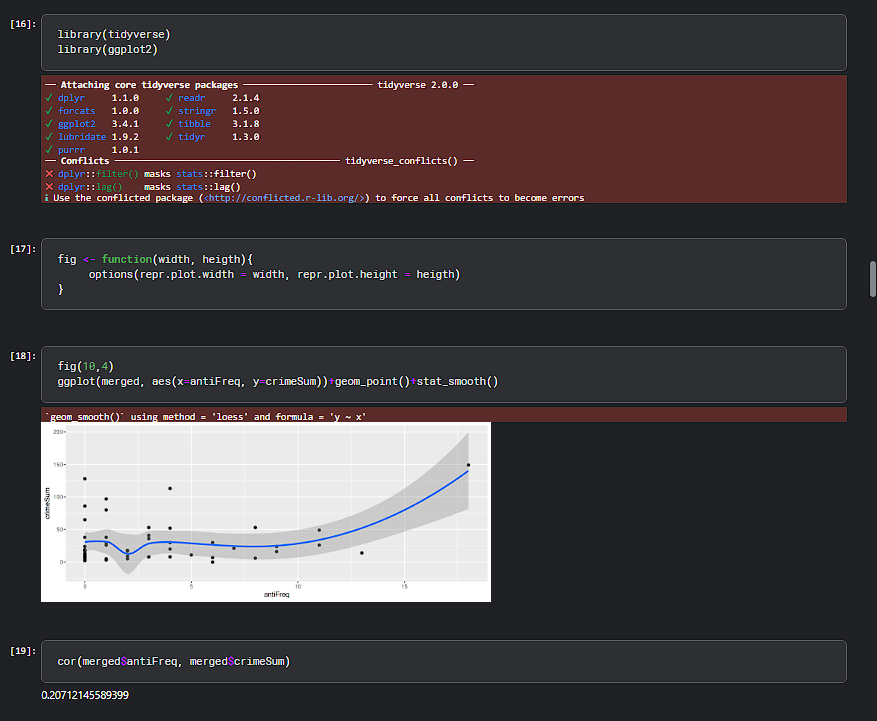
****

****

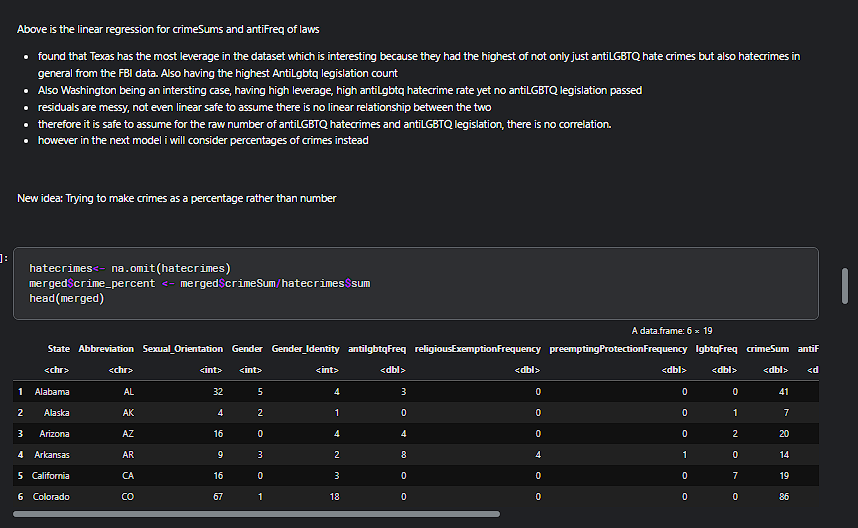
****

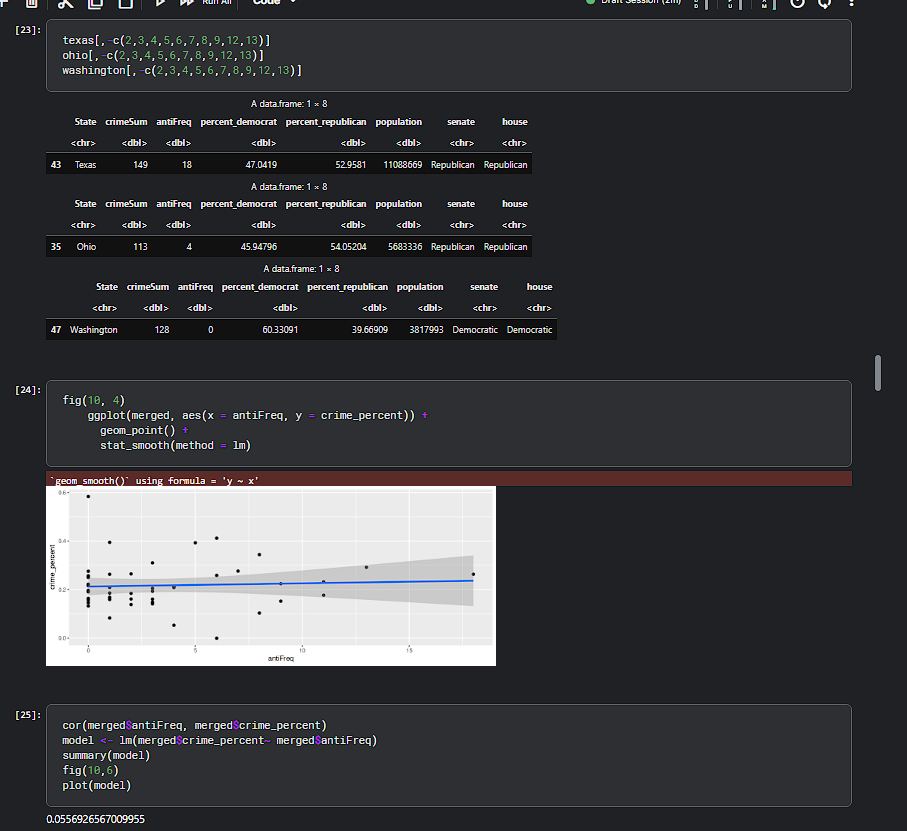
****

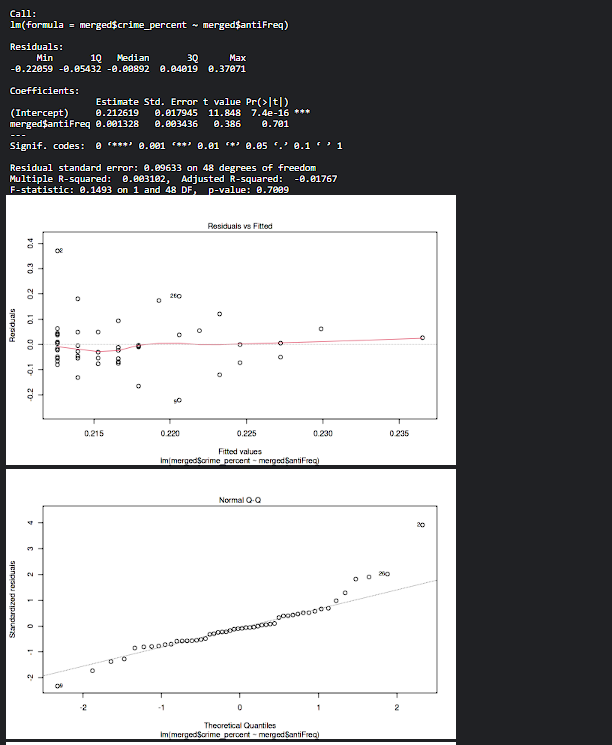
****

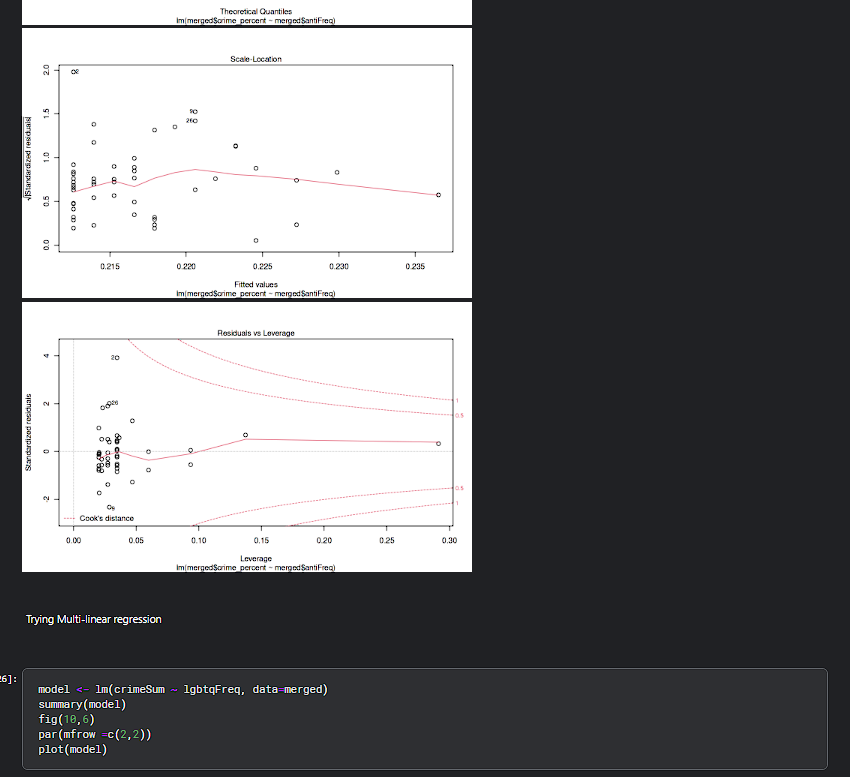
****

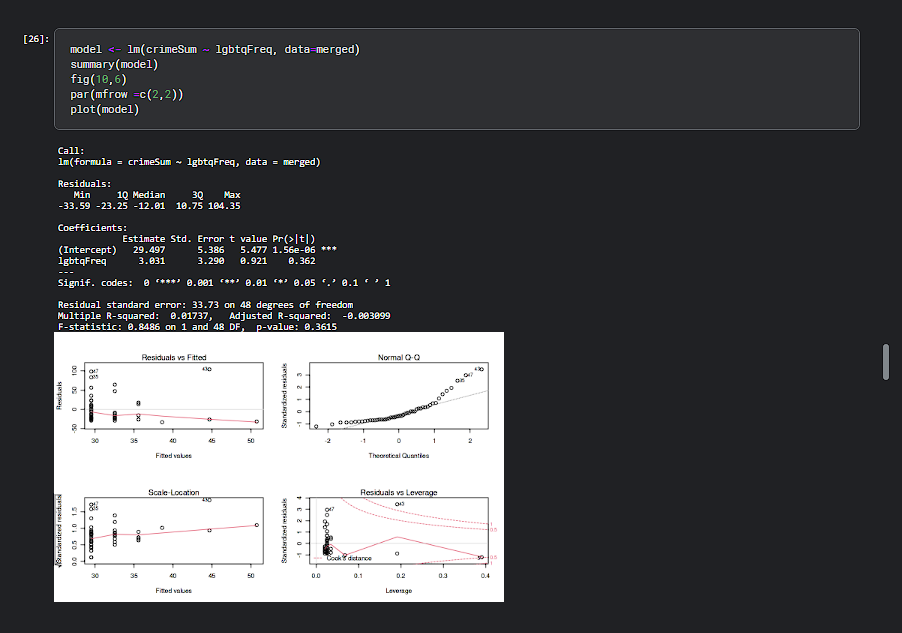
****

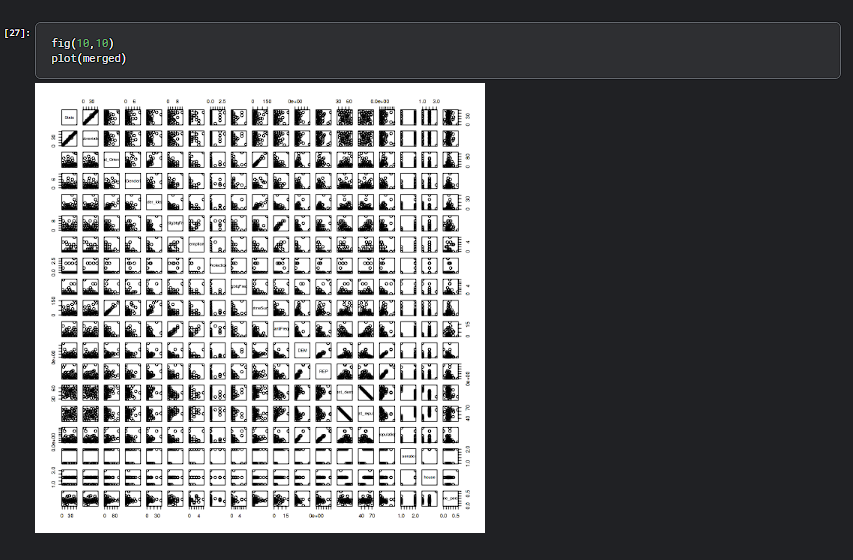
****

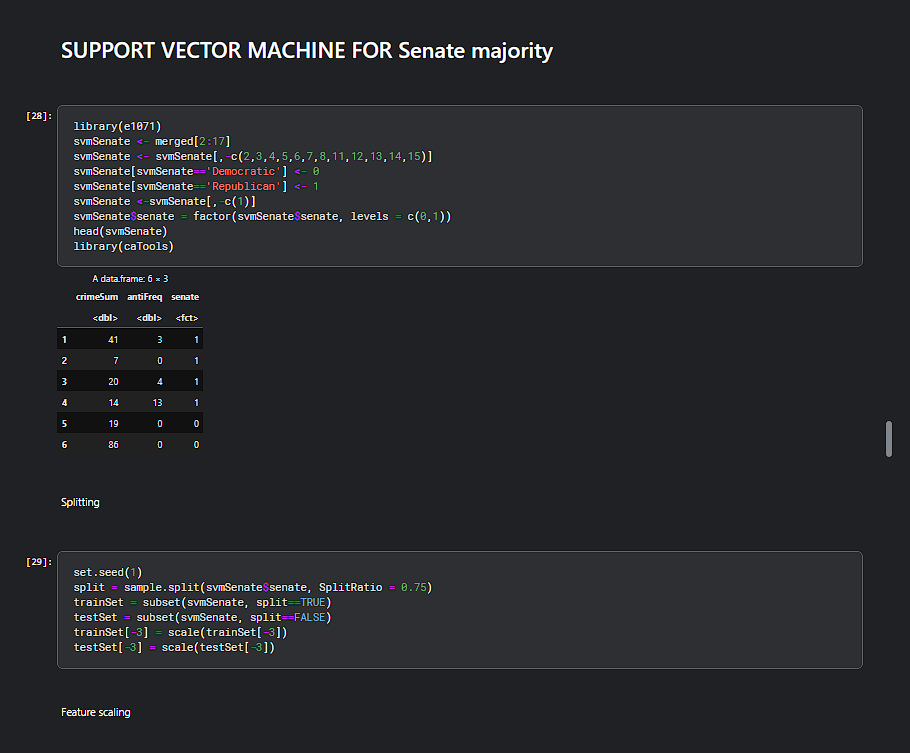
****

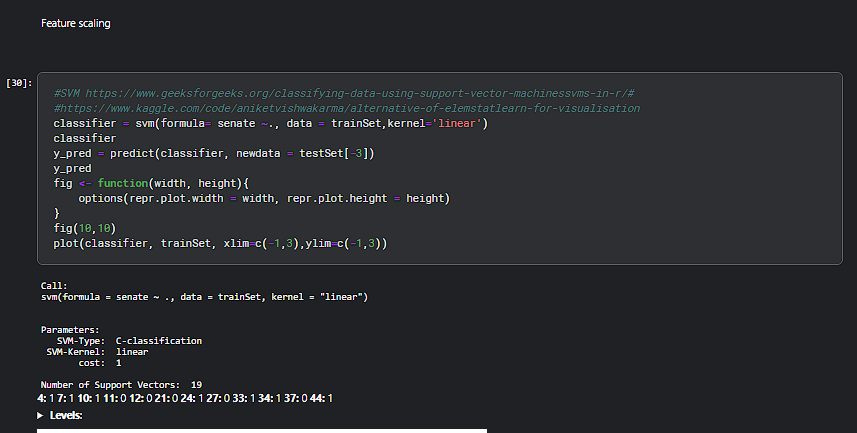
****

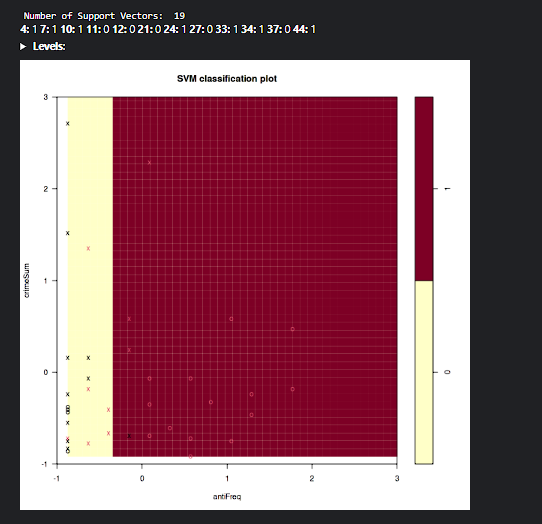
****

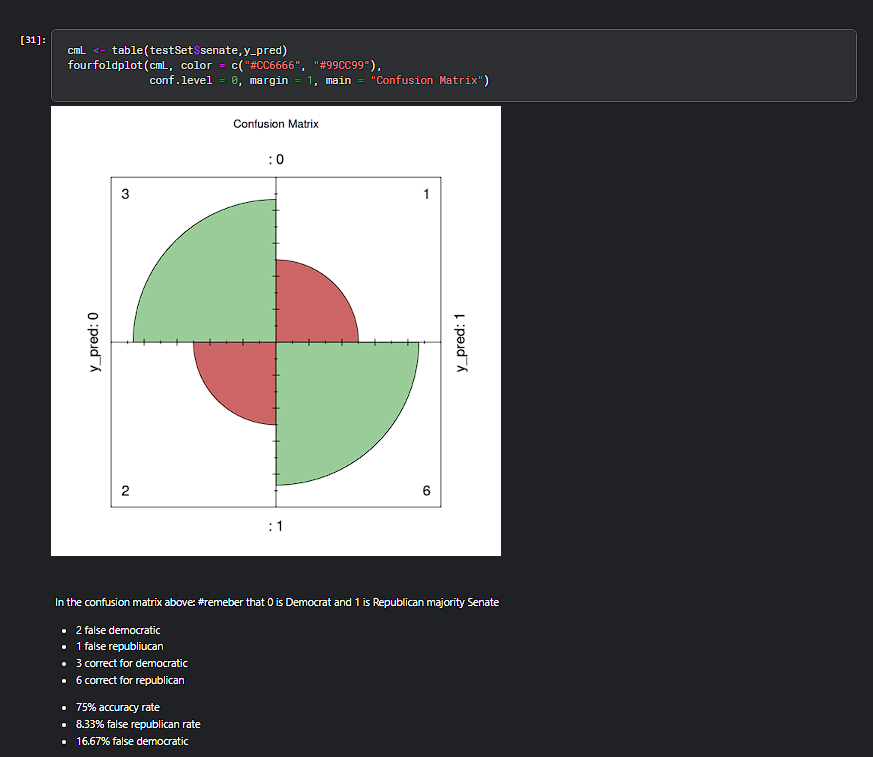
****

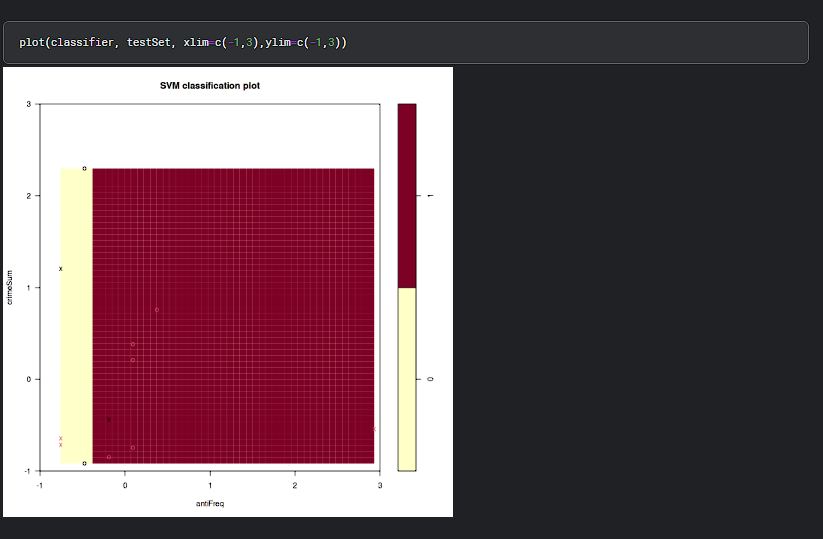
****

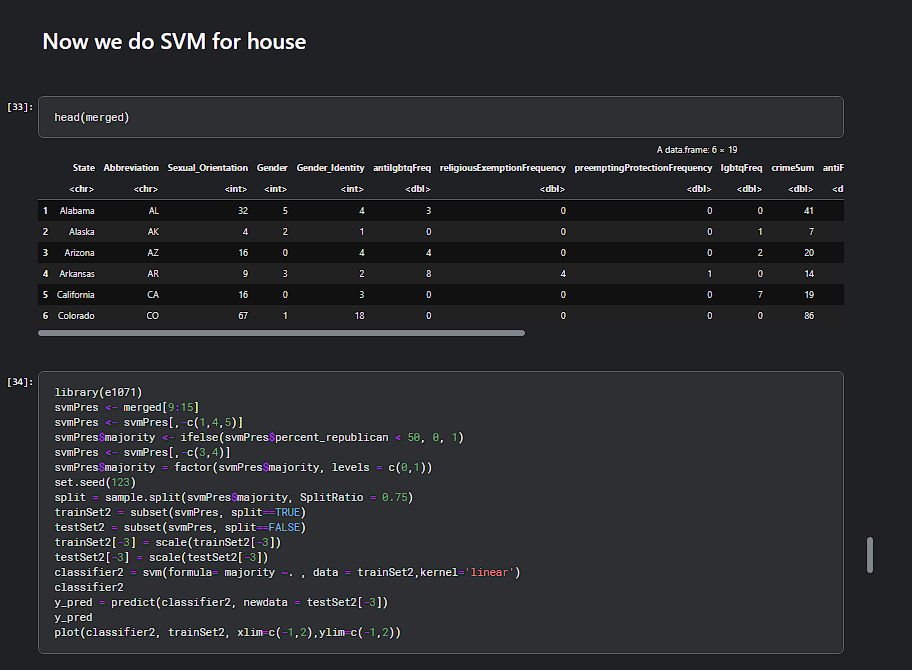
****

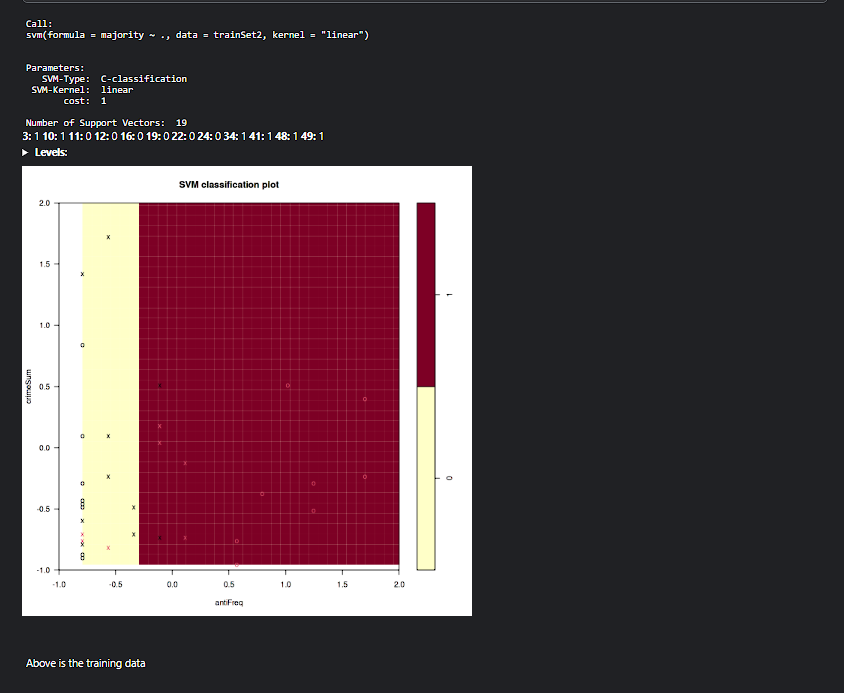
****

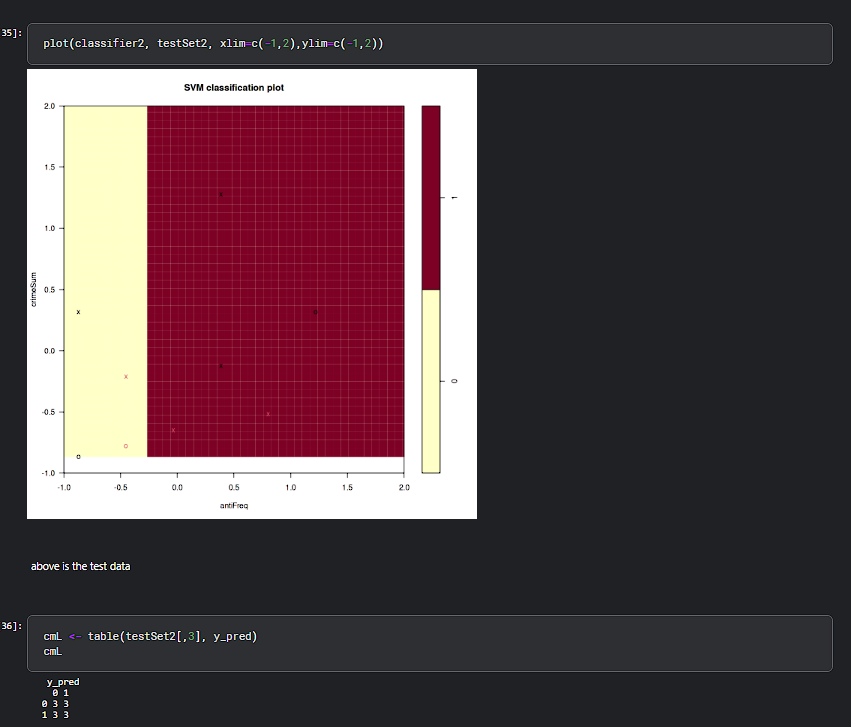
****

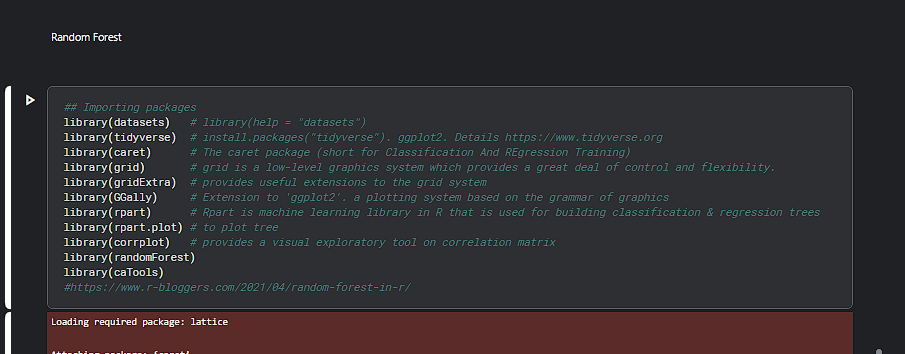
****

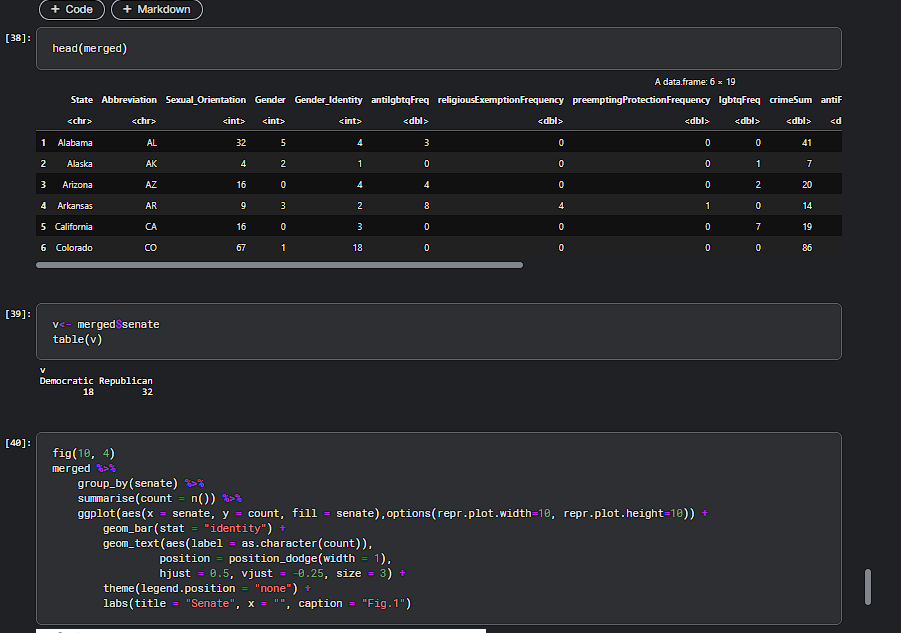
****

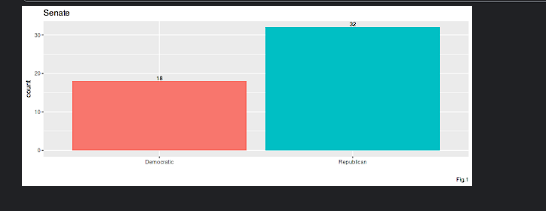
****

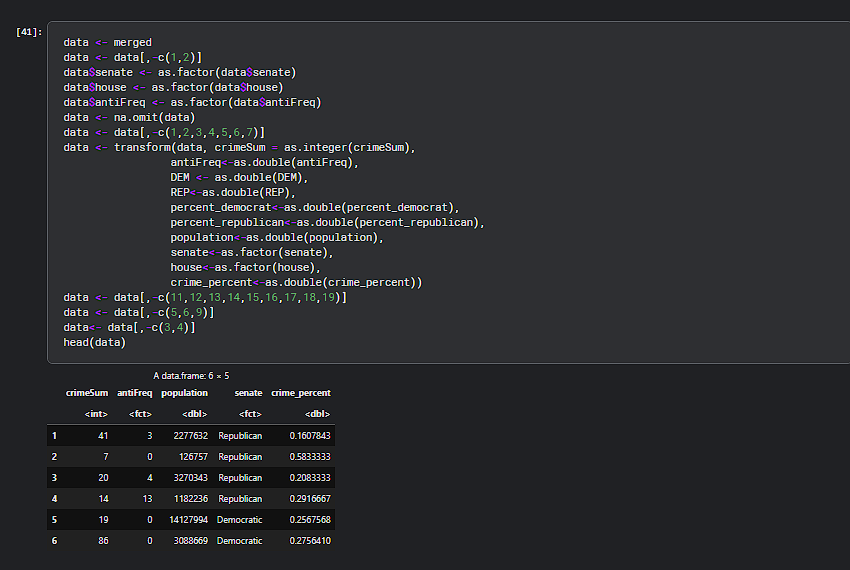
****

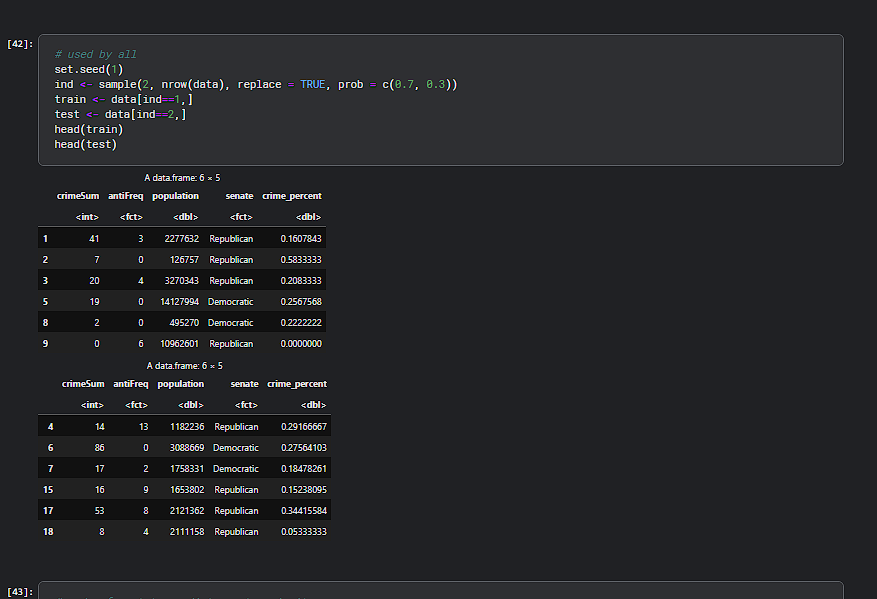
****

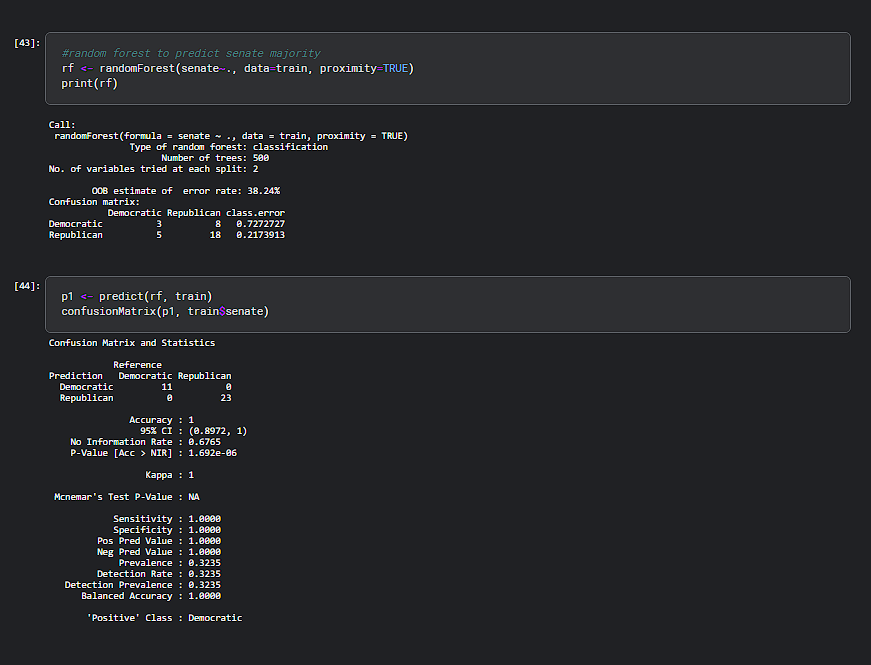
****

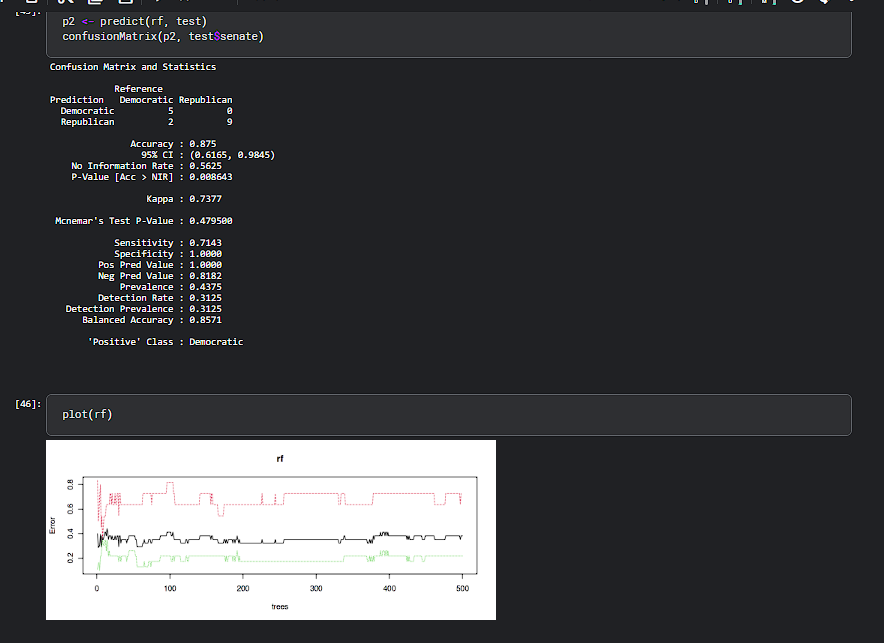
****

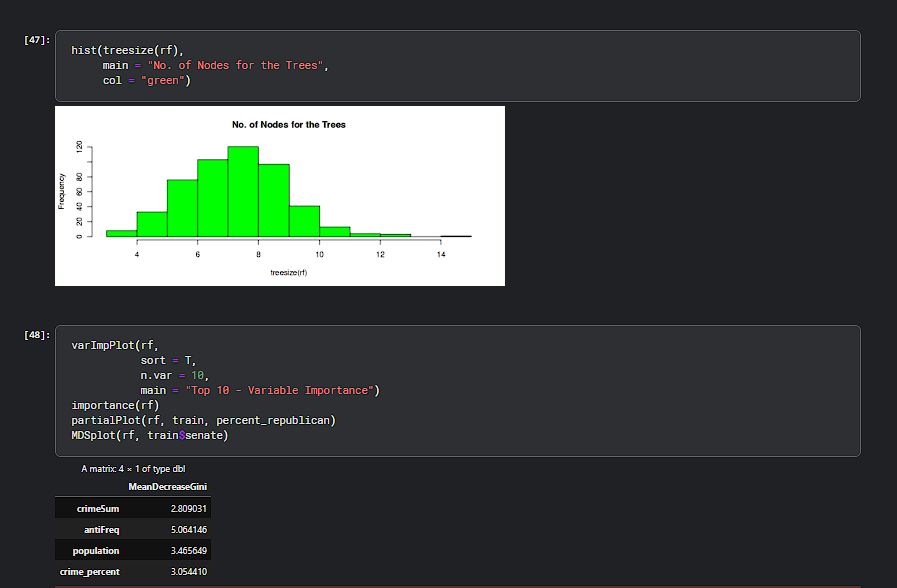
****

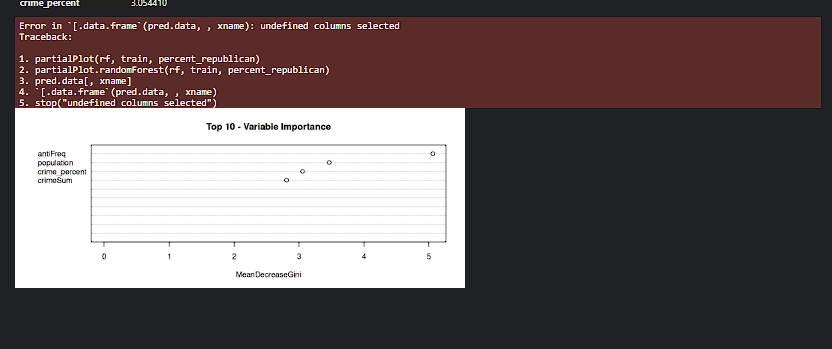
****

****

****

****

****

****