2016 United Kingdom Referendum (Brexit) Analysis

AVERY GUETHING, Marquette University, United States

In 2016, the United Kingdom elected to leave the European Union. With a close result of 52% to 48% and a large turnout, the vote started the process to change the future relationship of the two. This article takes a look at the official voting data of the U.K region and whether we can predict if the area voted to leave based on location size by using linear regression. It also analyzes the possibility of voter fraud that can be found in the presented data set.

CCS Concepts: • **General and reference → Cross-computing tools and techniques**; *Reliability*; *Estimation*.

Additional Key Words and Phrases: United Kingdom, European Union, Brexit, Linear Regression, Referendum

ACM Reference Format:
Avery Guething. 2019. 2016 United Kingdom Referendum (Brexit) Analysis. 1, 1 (December 2019), 9 pages. https://doi.org/NaN

1 INTRODUCTION

The United Kingdom consists of the following countries: England, Scotland, Whales, and Northern Ireland. Before 2016, discussions about leaving the European Union rose due to higher immigration and economic problems throughout the continent. [5]

In 2016, a referendum was held, and the U.K officially voted to leave the European Union. The name of the referendum soon became "Brexit" as a play on words of "Britain" and "Exit." The official data was released by the *Electoral Commission*, who is authorized by the government to release voting information as well as information about referendums in the country. The data itself contains electorate size, whether or not they voted to leave, and information about ballots that were not counted.It should be recognized that this process is still being continued today as an official leave plan has not been agreed upon by the United Kingdom.

The goal of this analysis is to see whether or not there is relationship between voting size and the location size of places in the U.K. The second goal is determine if there any suspicious outliers that can be seen when visualizing all reasons of invalid ballots. The approach is to use linear regression and scatter plots.

2 LITERARY REVIEW

As noted in the introduction, the history between the United Kingdom and the European Union is complex. In order to be able to understand the data analyzed, one needs to understand the connection between the variables, the history of the two, as well as the election process. The following will go over the past, present, and future process of the United Kingdom leaving the European Union (otherwise known as Brexit).

Author's address: Avery Guething, avery.guething@marquette.edu, Marquette University, Milwaukee, Wisconsin, United States.

XXXX-XXXX/2019/12 $NA
https://doi.org/NaN

2.1 Prior to the Vote

The European Union, which was formally called the European Economic Community (EEC), was founded during the 1950s as a way for countries to encourage peace. This is because "Nations that traded together, it was believed, would be less likely to go to war with each other."[8] Though the number of countries was significantly smaller, the idea was to provide a better way of communication and an easier way to set economic boundaries. After back and forth debate, the United Kingdom was officially accepted into the European Union in 1973. However, the following years would be proved troublesome for the two.

At the time that the U.K was accepted, there were only nine countries in the EEC, and the United Kingdom was struggling economically. There was inflation on many types of goods, taxes that were unable to be paid, and widespread strikes.[4] In 1975, there was the first referendum to leave the European Union; however, the majority decided to stay. The rest of the history between the two continued to see tensions as the European Union grew, and economic unrest was seen many times from the 1980-2000s. The main goal of the E.U also grew from just economic sanctions but to acknowledge political agreements, foreign policies, and even made a single currency. [8]

2.2 Brexit in 2016 and Voting Procedures

In 2013, the current Prime Minister of the United Kingdom, David Cameron, was determined to bring the topic of the relationship between the two to the people. This was to conclude all previous tensions and negotiations that went back and forth for the past 40 years. The voting options were broad, and in 2016, the people had the vote to either "stay" or "remain." The main issues at hand were more targeted at policies of immigration due to a large number of refugees that needed new homes. [6]. A final vote was held in 2016, and with a turnout of 72%, the United Kingdom chose to leave the European Union. However, the processing of leaving the E.U. would be a long process, and not everyone wanted to leave. The majority of Scotland voted to remain as well as Northern Ireland. This lead to debates about other countries trying to leave the United Kingdom. [7]

The term "electorate" refers to the voter, and the registration process is similar to the United States. A voter has to be over the age of 16 to register, but cannot vote in most elections until 18. They also have to be a citizen to the country that they are in or be a "European Union citizen who is resident in the U.K." [1] There is currently a register of voters that individual members of the government can access and this helps provide more information about locations' decisions.

2.3 Future of Brexit

Many different outcomes could happen from Brexit. As of right now, the United Kingdom Parliament is in discussions of creating a proper exit plan that will help organize trade plans as well as immigration policies. However, many proposals have been negated, and the deadline for them to leave has been pushed back to January 2020. If no deal is reached between the two, there could be another referendum, and the people would get the opportunity to vote again, or they could leave without other precautions put into place. [3]

If a no-deal Brexit goes through, it is expected to cause significant economic rifts and create more issues for travel as U.K citizens would have a new and different process for immigration through the E.U. However, if the deal ends up in the United Kingdom's favor, Kallum Pickering, a senior U.K economist, writes, "An orderly Brexit … would significantly reduce economic uncertainty by dramatically narrowing the range of likely scenarios for the near-term economic and political outlook." [2] As the discussion continue, it also could make implications for the voters and their representation. If it does not go through, the voters' decisions would have been overlooked and also put into question the democracy of the region.

Table 1. Linear Regression for Electorate versus Decision Margin

Variable	Coefficient	Standard Error	P-Value > |t|
Const	-9675.6843	1712.507	0.000
Electorate	0.0522	0.011	0.000
R-squared Value:	0.0583		

3 DATA ANALYSIS

3.1 Data Cleaning and Data Organization

When initially searching, I found two sets of data that could be combined together to help determine possible relationships. The initial data set provided information about the ballots received, such as whether they were accepted or the reason they were denied, as well as the dates the ballots were counted. The other data set had the two percentages of each decision per voting decision. I combined both of them to help provide the most accurate information about voting areas and their decisions.

With these two data sets, I combined them both using *cat* command in Mac's terminal as well as deleting extra columns that appeared from combing them using python. The data was then imported into a pandas data frame. After this, the data needed to be in tidy format so unnecessary header columns were deleted. Additionally, I went through the combined data set and filled in NA values with 0 as deleting certain rows would provide incorrect results, but leaving blanks could cause error with calculations. I, also, deleted empty, unnamed columns as well as vague columns that would not provide any necessary information for my research.

4 RELATIONSHIP BETWEEN ELECTORATES AND DECISION MARGIN

For the first, original problem, I decided to take a look at the relationship between electorates and the decision the location made in terms of "remaining" or "leaving." I created another data frame with column copies of the original (named df) as well as an additional column where I counted the decision margin by taking the number of "remain votes" minus the number of "stay votes". The negative values in this column indicated that the voting area in that row decided to leave the European Union.

The variable "electorates" was picked because it shows the number of people who voted. The larger this number is, the larger the voting area is. For example, the Isles of Scilly had 1,799 electorates compared to the city of Birmingham, which had 227,251 electorates.

4.1 Linear Regression for this Relationship

I elected to use linear regression on this because my outcome was continuous and could have an infinite number of possible values. I originally plotted the values on a regression-plot using the python library seaborn. Figure 1 shows plotted values in blue, a base orange line at 0, and the solid regression line in blue. This function creates the linear fit, but does not provide this equation, so I needed to calculate the line on my own. Figure 2 shows the residuals found in figure 1. There were approximately 380 residuals. With using scikit-learn's python library, I used their linear regression equation to help predict the best line of fit for my regression. I determined the following equation:**y = -9675.7 + 0.052185x**. This equation was then confirmed by using putting it back into Figure 1 and the two lines appeared to be overlapping. After this, I used the statsmodel library to provide Regression results from the two columns. Table 1 was indicates the result of figure 1 by using the regression line that was calculated.

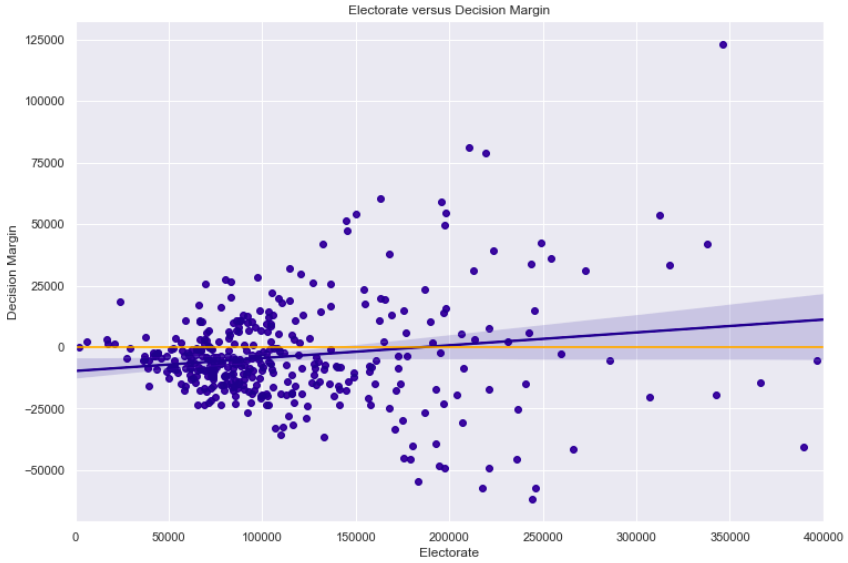

Fig. 1. Electorate versus Decision Margin. Below 0 indicates voting to leave while above indicates voting to remain.

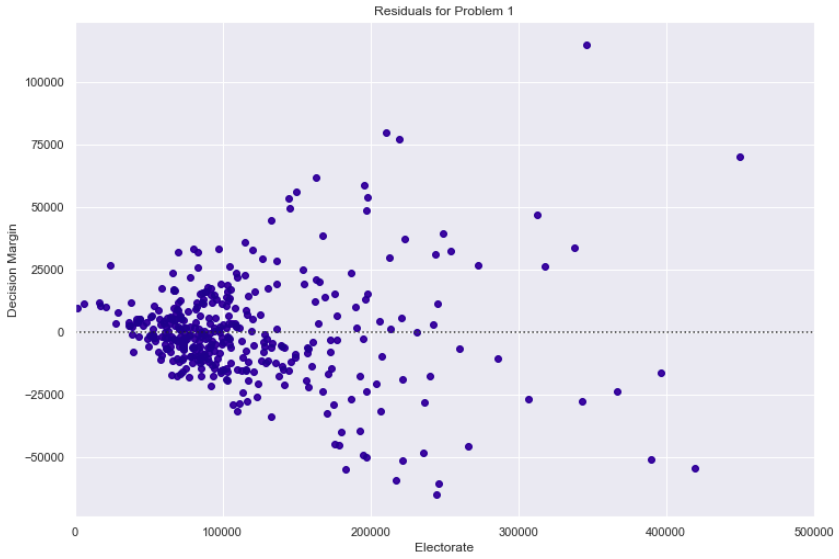

Fig. 2. Residuals from Figure 1

5 OUTLIERS ANALYSIS AND VISUALIZATIONS

For the next half, I wanted to explore the outliers that were found from "Invalid Votes" from "Electorates". The goal was to find any suspicious outliers that continuously had a higher number of invalid votes in different categories provided by the data.

Table 2. Linear Regression Results for Invalid Votes Count versus Electorate

Variable	Coefficient	Standard Error	P-Value > |t|
Const	3.202e+04	4186.256	0.000
Invalid Votes Count	1351.2670	47.004	0.000
R-squared Value:	0.685		

I created another data frame that was smaller and easier to analyze. It contained copies of all reasons for invalid votes (i:e: Missing DOB or mismatched signature), electorates, and a column where I calculated the number of invalid votes for that row by subtracting "Valid Votes"from "Votes Cast."

5.1 Linear Regression

I used linear regression as I had an infinite number of possibilities and the relationship grew in a linear line. Figure 3 shows the results of comparing "Invalid Votes Count" to "Electorates" and figure 4 shows the residuals of the regression I found. Both were done using the seaborn library.

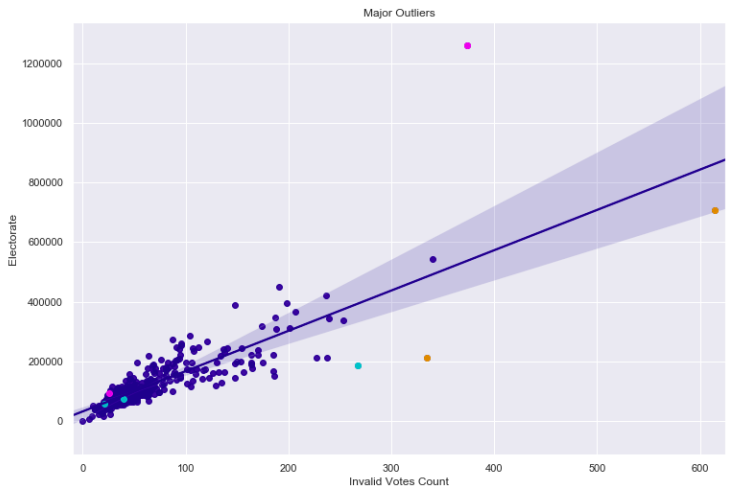

Fig. 3. Relationship between Invalid Votes versus Electorates using Regplot

Again, with using scikit-learn's python library, I used their linear regression equation to help predict the best line of fit for my regression. I determined the following equation:**y = 3.2024e+04 + 1351.3x**. This equation was then confirmed by using putting it back into Figure 3 and the two lines appeared to be overlapping. After this, I used the statsmodel library to provide Regression results from the Electorates and Number of Invalid Votes. Table 2 was indicates the result of figure 3 by using the regression line that was calculated. The highlighted points in Figure 3 are used in subsection 5.2.

5.2 Visualizations of different Invalid Vote Reasons

The goal was to look at the outliers from Figure 3 and follow along to see if any voting area continuously had issues with ballots. I made an additional column in my data frame that added "Both mismatched", "Mismatched Signature","No DOB", "No Signature", and "Mismatched DOB."

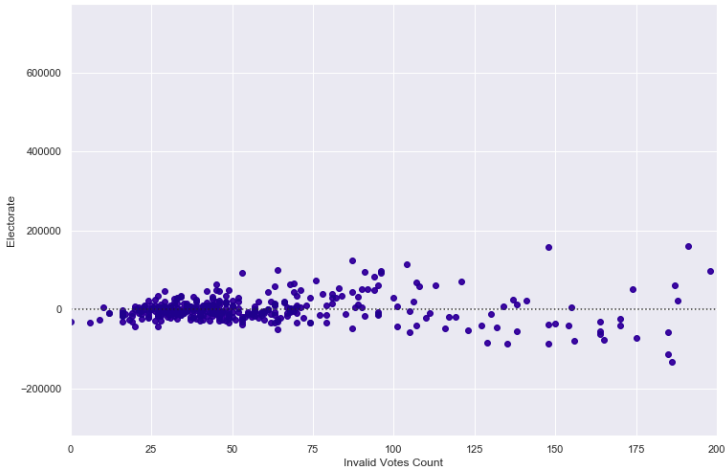

Fig. 4. Residuals from Figure 3

This was to help find starting point outliers that can be analyzed in the regression plot and other scatter plots. This was called "All Added." After sorting my data frame by least to greatest amount of

Fig. 5. Scatter plot of "All Added" versus "Electorate"

combined total of invalid votes reasons and analyzing which had a higher number of votes within each invalid category, I plotted two different reasons in a scatter plot. Figure 5 plots the addition of "All Added versus "Electorates". Figure 6 demonstrates the category "Both Date of Birth and Mismatched Signature" versus Electorate. In figures 4,5, and 6, the orange highlighted points are the cities of Birmingham and Leicester, the pink representing East Ayrshire and Northern Ireland, and the turquoise representing Brent and Bromsgrove.

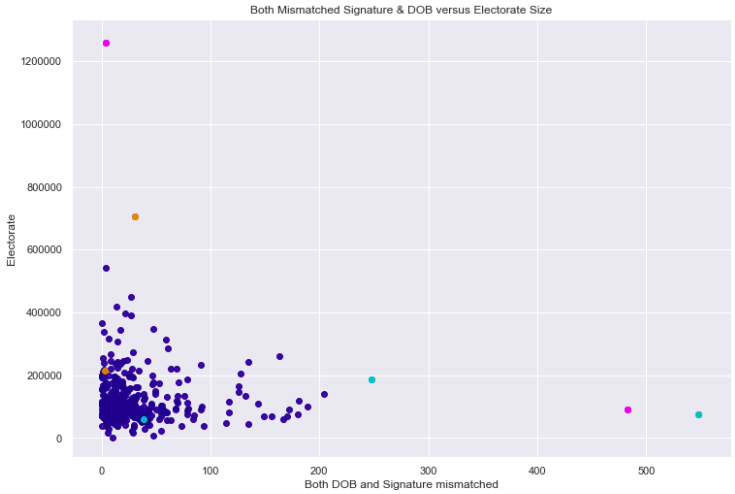

Fig. 6. Scatter pot of "Both Date of Birth and Mismatched Signature versus "Electorate"

6 RESULTS

When analyzing the regression of "Electorates" and "Decision Margin," one should see that the line plotted is not in a linear fit, nor is it any visually detected pattern. Both Figures 1 and 2 show no patterns, and this is the first indication of no relationship. The residuals are may first appear to look balanced, but after 100,000 electorates, the pattern changes dramatically. When looking at the table, one can see that the p-value is low; however, the r-squared value is at around 6%. This shows that a low percentage of the variability of the decision margin can be explained by the electorate size. With this analysis, one can conclude that there is no significant relationship between the two.

When looking at the regression of the "Electorates" versus "Invalid Votes Count," one can see that the line is more in a linear fit than as compared to figure 1. The pink highlighted point is the country of Northern Ireland, which has a very high population in comparison to the other cities in the United Kingdom. This indicated point is more likely to have higher leverage in the data because it is a large population. The residuals are also in a more detectable pattern and can be similar to the "trumpet" pattern. The relationship between the two can be confirmed when looking at the table as the R-squared value is high, and the p-value is low.

When analyzing the outliers, one can see that the two turquoise dots are consistently outliers in all major invalid vote cases. These cities are Brent and Bromsgrove. In figure 5, they are seen on the right-hand side even though they have a lower population size, and in figure 6, they are shown to be outside of their cluster for that electorate range. The pink dot on the far right-hand side in figure 6 is East Ayrshire, which only had a high number in "Both DOB and Signature mismatched," so it was not as suspicious as the two other points. It is more likely that Brent and Bromsgrove had more issues with invalid votes than other highlighted outliers.

7 CONCLUSION AND IMPLICATIONS

For my first goal, I wanted to explore the concept of finding a relationship between location size and the decision the voting area made. For example, if rural areas (less population) swayed towards one side over the other. However, I found that the relationship between the decision margin, which explained what side they voted on, and electorates were not explained well with linear regression. The electorate size did not influence their overall decision. This implies that there could be another

method of predicting, or there could be other outside variables, such as media influences. In my final visualization (figure 7), I present a visualization that shows how complex the data was when it came to prediction my first goal. Pink represents voting to stay while Purple represents voting to leave.

In terms of outliers, there were suspicious locations that had an increasing number of invalid votes, and two areas had an increased number of mismatched signatures and dates of birth. This could imply nefarious purpose being done to influence the results, or it could have been done out the innocence of not knowing how to keep voting information updated. The relationship between invalid votes and electorates was confirmed, and this showed that the bigger the size, the more invalid votes. In the future, it would be worth looking into these areas to discuss possible issues with the voters or the process from those locations.

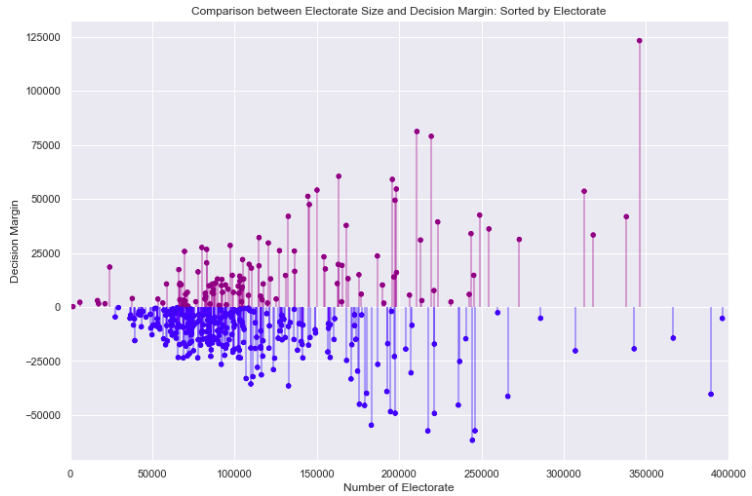

Fig. 7. Final Visualization: Decision Margin versus Electorate. Sorted in least number of Electorates to Greatest.

REFERENCES

[1] 2019. *How do I register to vote?* Retrieved December 3, 2019 from https://www.parliament.uk/get-involved/elections/register/

[2] Silvia Amaro. 2019. *The UK might be close to a deal, but Brexit wrangling could last for a decade.* Retrieved December 4, 2019 from https://www.cnbc.com/2019/10/21/brexit-uncertainty-could-last-for-a-decade-despite-commons-vote.html

[3] Peter Barnes. 2019. *Brexit: What happens now?* Retrieved December 4, 2019 from https://www.bbc.com/news/uk-politics-46393399

[4] Steve Cannane. 2016. *Brexit vote: All too similar to UK's 1975 referendum.* Retrieved December 3, 2019 from https://www.abc.net.au/news/2016-06-17/brexit-the-sequel-plays-out-forty-years-after-first-referendum/7510434

[5] David Frum. 2019. *It's Five Minutes to Midnight in the U.K.* Retrieved December 3, 2019 from https://www.theatlantic.com/ideas/archive/2019/03/brexit-short-history-bad-idea/584524/

[6] Benjamin Mueller. 2019. *What Is Brexit? A Simple Guide to Why It Matters and What Happens Next.* Retrieved December 3, 2019 from https://www.nytimes.com/interactive/2019/world/europe/what-is-brexit.html

[7] CBS News. 2019. *Will Brexit drive Scotland to break from the United Kingdom?* Retrieved December 3, 2019 from https://www.cbsnews.com/news/brexit-news-scotland-independence-gains-momentum-as-united-kingdom-set-to-leave-european-union/

[8] Sarah Pruitt. 2017. *History behind Brexit.* Retrieved December 3, 2019 from https://www.history.com/news/the-history-behind-brexit