Analysis of the Brazilian Immigrant Electorate

John Bestavros, Jenna Peters, Wiley Hunt, Jerry Zhang

December 4, 2021

1 Introduction

Our project seeks to understand the composition of the Brazilian immigrant electorate and analyze the trends in popularity of political parties among this group. By analyzing electorate data collected for all Brazilian Presidential elections from 1998—2018, we hope to understand changing demographics and voting patterns of the Brazilian immigrant electorate. In addition, we analyzed the differences in these demographic and voting patterns across the communities they live in.

Our analysis has been broken into two sections. We first analyzed the electorate as a whole. We then ranked the countries represented in the electorate by immigrant population and selected the top ten countries. We did a similar, but deeper analysis on this section of the data set. The purpose of the top ten analysis was to uncover demographic or political differences in Brazilian immigrant communities.

2 Data and Methodology

2.1 Overview of the Data

For this project, we used two separate data sets. The first was the Brazilian electorate data set (provided by the client), which contains demographic data on Brazilian immigrant voters for the years 1998, 2002, 2006, 2010, 2014, and 2018. The second data set is the election data set, and it contains voting records for Brazilian immigrant voters for the same years.

2.2 Data Cleaning

At the beginning of the project, we focused on reconciling pre-2010 data sets with post-2010 data sets as there were missing, incomplete, or inconsistent values across the data sets:

- Prior to 2010, the municipality code (location of voter) feature only identified the country of the voter. However, after 2010, the codes were updated to identify the region within a country of the voter. In order to compare data from all years, we developed a mapping function to reconcile the codes for the pre-2010 and post-2010 data sets.

- From 2002 until 2010, there is an attribute "educational attainment code" numerically representing education level along with an "educational description". However in 2010 and 2018, this attribute is no longer available and we are left with just an education description. We therefore developed a function to map this attribute to 2010 and 2018 electorate data sets.

- We noticed that the "age group" categories were different between the 2010 and 2014/2018 data sets (essentially, the 2010 data set bucketed age groups in intervals which were inconsistent with those used in 2014/2018). We therefore developed a mapping function to map the 2014/2018 age group descriptions to those used in 2010.

- In order to compare country names across data sets (where different conventions of accents or capitalization were being used), we created a function to strip all accents and capitalize all country names.

2.3   Methodology

After cleaning the data set, we looked at how the composition of the electorate populations had changed over time (at an aggregate level) as well as voting patterns among the aggregate data set. Then, to get more granular, we performed deeper analysis on the 10 most populous countries (in terms of the 2018 Brazilian immigrant electorate data set).

Our project primarily focused on uncovering patterns and therefore relied heavily on filtering the data according to specific features of interest (age, sex, education) and then visually displaying plots to analyze any changes/trends in these features over the course of the six election years that were provided.

3   Analysis of the Aggregate Electorate

3.1   Aggregate Electorate Size and Features

We began our analysis by looking at the features and composition of the overall electorate and how it has changed over the years in our data set.

For all data sets since 1998, the United States has remained the country with the highest number of Brazilian Immigrant voters. Italy and Portugal have also been among the top 5 countries every year in the data set. Meanwhile (as shown in our Top 10 Countries analysis
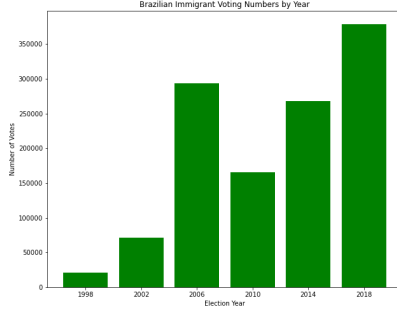
Figure 1: Brazilian Immigrant Electorate Size by Year
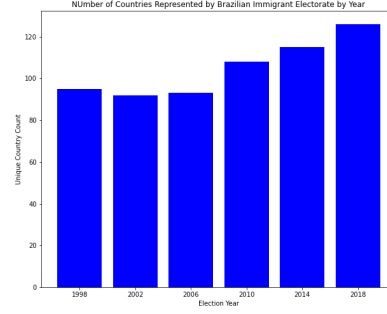

Figure 2: Countries Represented by the Brazilian Immigrant Electorate by Year

below), in 2010, the number of Brazilian immigrant voters in Japan increased 20x, moving it into 4th place among countries with the most Brazilian immigrant voters.

Meanwhile, the top growing countries for Brazilian immigrant voters (as measured by change in population from the prior election year) has varied each year. However, the UAE was the top growing country in both 2014 and 2018.

3.2 Analysis of Sex in the Electorate

After analyzing the size of the electorate, we looked into the specific demographic features of the overall Brazilian immigrant population.

As seen in Figure 3, the majority of electorate is comprised of women. The female proportion of the electorate steadily increased from 1998 to 2006 (reaching a peak of 60.8% of the electorate in 2006), but declined slightly between 2010 and 2018. As of 2018, women make up 58.4% of the electorate (which is below 1998 levels in which women comprised 59.4% of the electorate population).
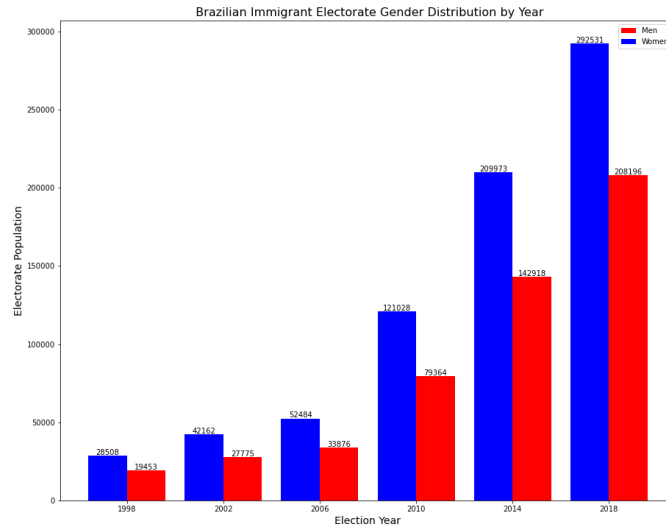
Figure 3: Brazilian Immigrant Electorate Sex Distribution by Year

3.3 Analysis of Education in the Electorate

Next, we looked into the education levels of the electorate. We can see in Figure 4 that the individuals with a complete college education and complete high school education are the most prevalent (followed by individuals with an incomplete college education) across all years. While not shown here, we found that this trend is the same for both men and women.

3.4 Analysis of Voting Patterns

Finally, we looked at voting patterns among the electorate as a whole. Below, we can see the number of votes for each party over the 6 election years in the data set. Overall, the PSDB and PT have received the most votes in the last 6 elections, with the PSL coming in 3rd.

Next, we zoomed in on the last 4 elections (2006-2018). Looking at these last four elections, we can see in Figure 5 that while the PSDB and PT were the top two parties in 2006-2014, the PSL emerged as the overwhelming top party in 2018.
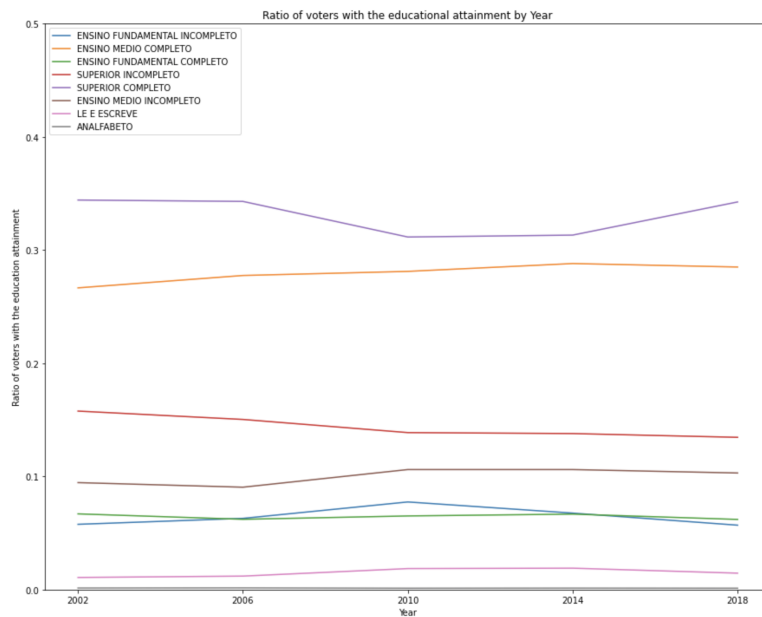
4

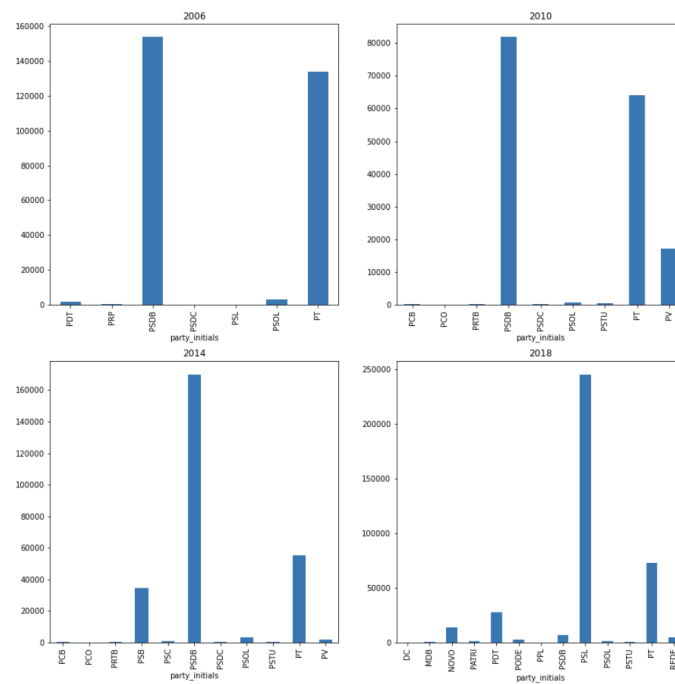Figure 4: Brazilian Electorate Education Distribution by Year

Figure 5: Votes by Party by Year

5

4 Analysis of the Top 10 Countries

The ten countries with the highest Brazilian immigrant population (as of 2018) are as follows (listed in order of highest to lowest population):

1. The United States

2. Japan

3. Portugal

4. The United Kingdom

5. Italy

6. Germany

7. Canada

8. Spain

9. Switzerland

10. France

4.1 Electorate Size and Growth

In Figure 6 we plot the number of Brazilian immigrants within each of these top 10 countries during the period 1998-2018. As noted above, the United States has the largest number of Brazilian immigrants for all years analyzed, while Japan has reached second place in the past 2 election cycles (after surpassing Portugal in 2014). Meanwhile (as expected), all 10 countries have higher Brazilian immigrant populations than the data set average (the dashed line).

Now, if we look at the growth rates of these top 10 countries in Figure 7, we can see that while all 10 countries have shown positive growth over the course of the years analyzed, the highest spikes occur in 2010, with the most significant growth occurring in Japan, where the population increased by roughly 20x when compared to the population in 2006.
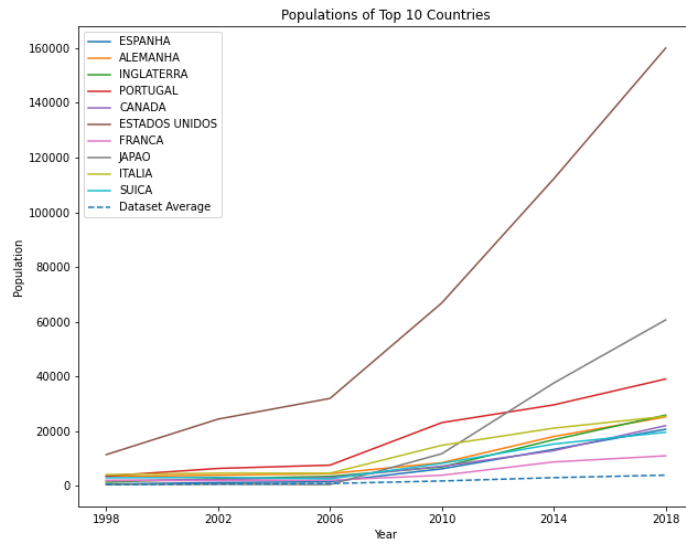
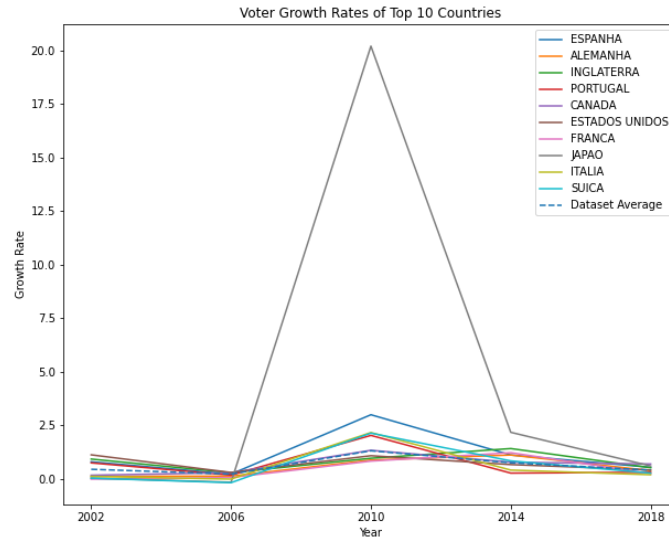Figure 6: Population of the Top Ten Countries

Figure 7: Population Growth of the Top Ten Countries

4.2 Analysis of Sex in the Electorate

Next, we began diving into the demographic features of these top 10 countries, beginning with sex.
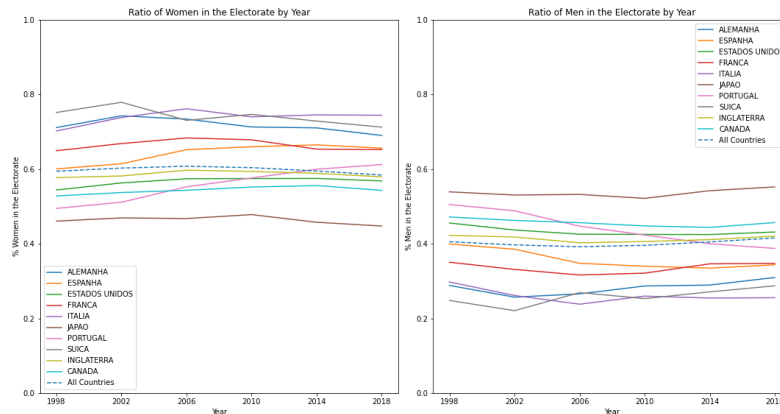

Figure 8: Top Ten Countries Sex Ratios

Looking at Figure 8, we can see that for most of the top ten countries, women make up the majority of the Brazilian electorate. This trend has not changed in the period between 1998 - 2018. There doesn't appear to be a trend upward, or downward

If we compare these top 10 countries to the overall data set (where the sex ratio has remained relatively constant at 60% women and 40% men), we can see that the proportion of women is higher in Switzerland, Italy, Germany, France, and Spain than the data set average (across all years). In Germany, Italy, and Switzerland, Brazilian immigrants are predominantly female, with women making up 70% or more of the Brazilian Immigrant population over all years.

Meanwhile, in Japan, Canada, the US, and England, the proportion of men has remained higher than the data set average across all years. In fact, Japan is the only country where the electorate population is majority male. Portugal is the one exception in our data set where the electorate population has shifted from having more male voters to more female voters (in comparison to the data set average) over the years analyzed.

4.3 Analysis of Education in the Electorate

We then analyzed the education-level demographics of the electorate in these countries.

When looking at the general trends over all years, we see that the ratio of members of the electorate who have a completed college education has actually dropped, especially from 2006 to 2010. The numbers went up again from 2014 to 2018, but they did not return to pre-2006 levels. Likewise, the ratios of members of the electorate who have not finished
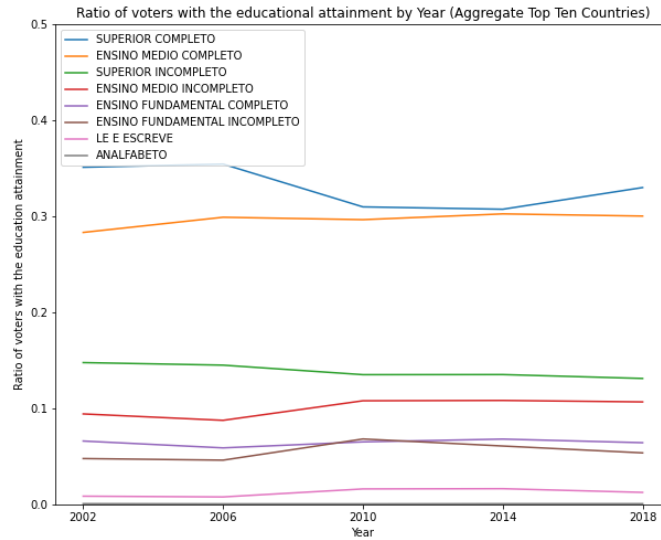
Figure 9: Top Ten Countries Proportion of Electorate by Education Level

high school or elementary school actually increased from 2006 to 2010, before dropping a
little from 2014 to 2018. The other educational attainments have remained relatively stable.
Once again, we cannot interpret trends for literate and illiterate individuals because the
numbers are too insignificant.

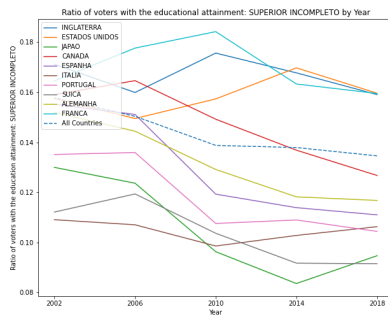We then zoomed in into these education levels and analyzed the trends at a country
level.

Figure 14: Ratio of Electorate with
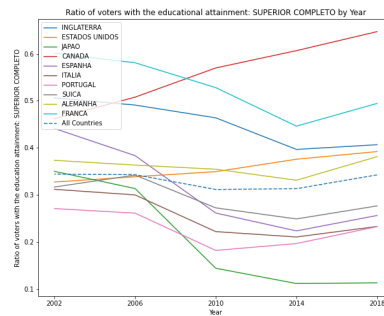Incomplete College Education by Year

Figure 15: Ratio of Electorate with
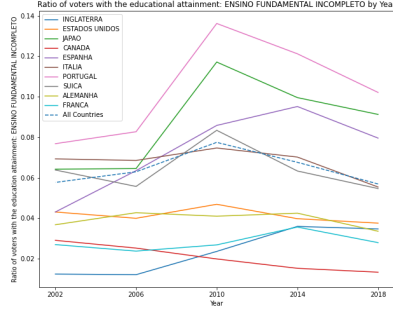Complete College Education by Year

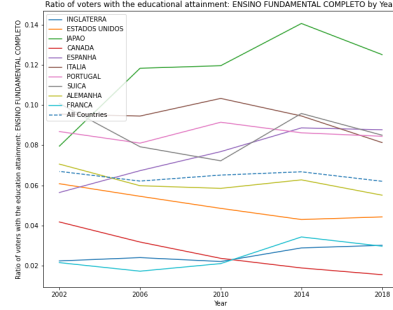Figure 10: Ratio of Electorate with Incomplete Fundamental Education by Year

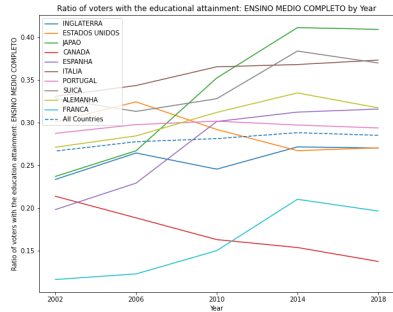Figure 11: Ratio of Electorate with Complete Fundamental Education by Year

Figure 12: Ratio of Electorate with Incomplete Secondary Education by Year
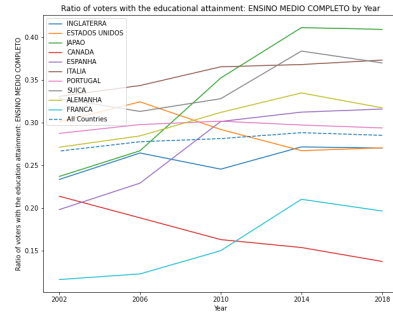

Figure 13: Ratio of Electorate with Complete Secondary Education by Year

As can be seen in Figures 10-15, in the majority of the countries we see an upward trend in the ratio of high school educated and incompletely high school educated individuals beginning in 2006. There seems to be a downward trend in the ratio of individuals who have completed or have not completed college beginning in 2006, which is likely just due to the ratio for high school education level ratios increasing. In 2010, we can see a peak for individuals with complete or incomplete elementary education, then a trend downwards. We cannot interpret trends in literate and illiterate voters as their ratios are inconsequentially small.

We then conducted further analysis by breaking these education levels down by sex. While not shown (in consideration of space), we found that for most countries except for the United States and Canada, where the ratios increased, the ratios of college educated men and women both went down. Similar but opposite patterns appear for high school education, where the ratios of high school educated men and women both increased, except

for in the US and Canada, where they decreased. For incomplete college education, the trend is universally downwards for men. For women, the trends are similar except for in the US where the numbers of incompletely college educated men increased. For incomplete high school education, the trends for both men and women are going up as a whole, but significant increases occurred in the countries of Japan and Spain. Similarly, the exceptions to this were once again Canada and the US, where the numbers decreased slightly.

4.4 Analysis of Sex and Education in the Electorate

Given that the majority of the electorate is High School or College educated, we decided to break down these two education attainment levels by sex.
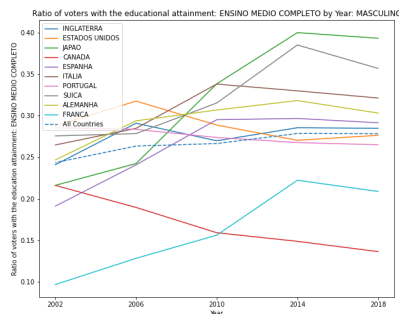

Figure 16: Ratio of Men with Complete High School Education by Year
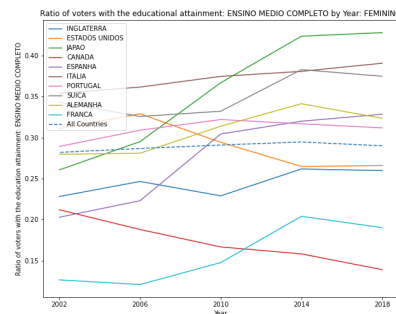

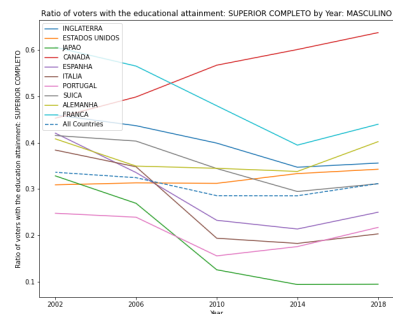Figure 17: Ratio of Women with Complete High School Education by Year

Figure 18: Ratio of Men with Complete College Education by Year
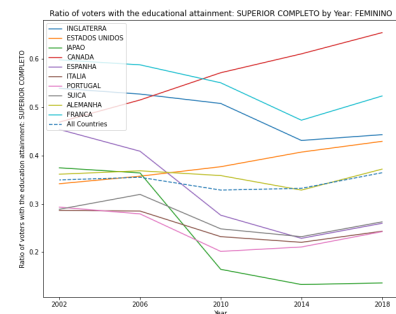

Figure 19: Ratio of Women with Complete College Education by Year

The trends and ratios appear not significantly different across sex in Figures 16-19.

11

However, France and Germany do show that the proportion of women with a complete college education is higher than the proportion of men with a complete college education.

4.5 Analysis of Age in the Electorate

The final demographic feature for the top 10 countries that we analyzed was the age distribution of the electorate.
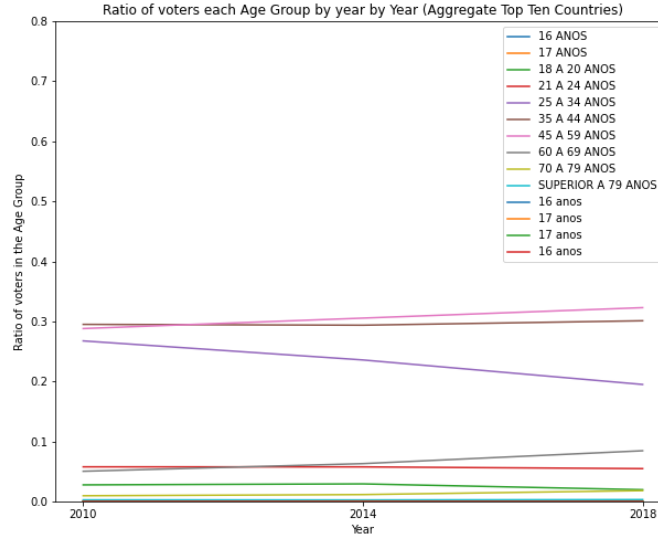

Figure 20: Ratio of Electorate in Each Age Range by Year: Aggregate of the Top Ten

In Figure 20, we plot the overall age group distribution in the top 10 countries. From this plot, we can see that the majority of the populations in these countries fall into an age range of 25-59, with all other age groups combined comprising less than 25% of the overall population.

We then decided to dive deeper into the three age groups in this range: 25-34, 35-44, and 45-59.

As seen in Figure 21, individuals aged 25-34 represent a significant proportion of the electorate.However, we can actually see a steady decline in this group over the course of the years analyzed. While not shown here, this trend is especially prominent among women in this age group, as the ratio of female individuals aged 25-34 has declined every year. The peak percentage of individuals aged 25 to 34 in 2010 was in Spain, with 35% of all Brazilian immigrants in the country in that age window. However, by 2018, that had dropped to 27.5% with France now having the highest proportion of the electorate population in this
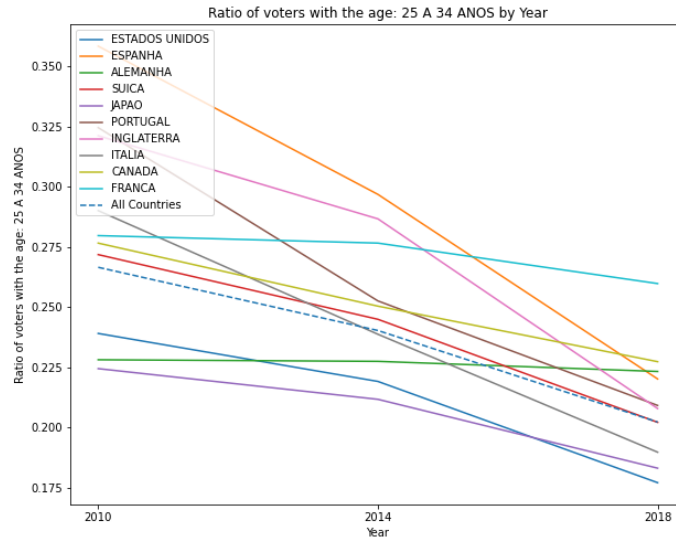
Figure 21: Ratio of Voters in the Age Range 25-34 by Year

age category (among the Top 10 countries).

As seen in Figure 22, the proportion of Brazilian immigrants aged 35-49 has comprised at least 24% of the population in each of our top 10 countries across the years analyzed. However, only England, Canada, France, the US, and Spain have seen growth in this category between 2010 and 2018, while growth in other countries has either remained flat or negative. Japan presents an interesting case in which the proportion of Brazilian immigrant electorate in this age group has declined from roughly 30% to 24%, with an especially sharp drop among women.

As seen in Figure 23, the proportion of the electorate aged 45-59 has comprised at least 15% of the population in each of our top 10 countries (and is as high as 35% in some countries) across the years analyzed. While for most countries we have seen an upward trend among this age group, Canada has actually seen a steady decline.
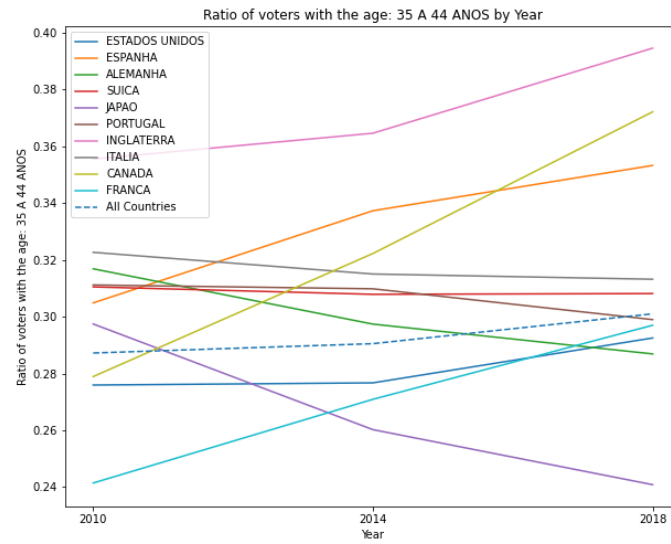
13

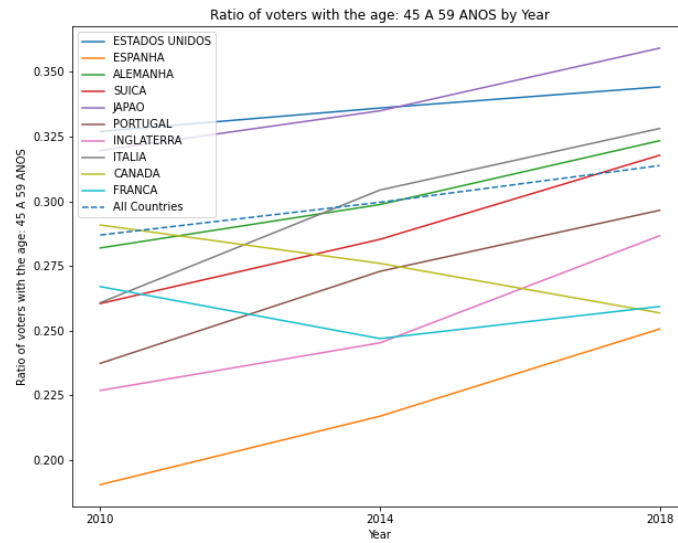Figure 22: Ratio of Voters in the Age Range 35-44 by Year

Figure 23: Ratio of Voters in the Age Range 45-59 by Year

4.6   Analysis of Voting Patterns

Finally, we looked into how voting patterns have changed in the top 10 countries as these demographic trends have changed over time to answer the question *Who do these Countries Vote for?*

First, we looked into which political parties these countries have voted for (in comparison to the electorate in general). Below, we can see the proportion of votes received by each party among the top 10 countries (and the entire data set) for the 2006, 2010, 2014, and 2018 elections.
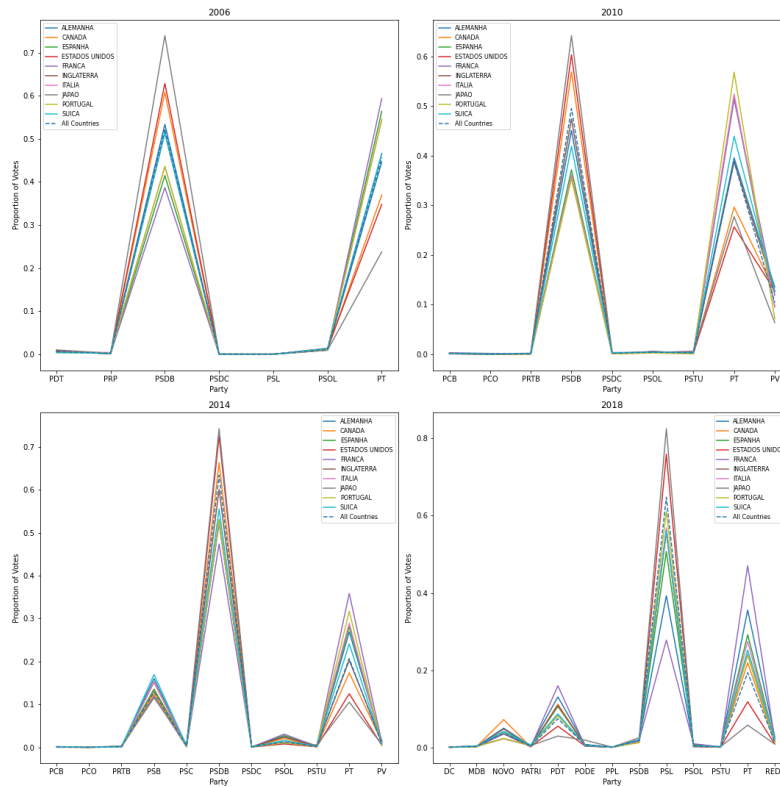

Figure 24: Votes for Each Party by Year, Broken Down by Country

One interesting pattern that emerges across all years is that the general shape of the voting patterns among the top 10 countries is roughly the same as the aggregate data set (ie, the most popular parties among the top 10 countries are also the most popular parties among the aggregate data set). As seen in Figure 24, the most popular parties appeared to be to PT, PSDB, and the PSL, with the PSL emerging only in 2018.
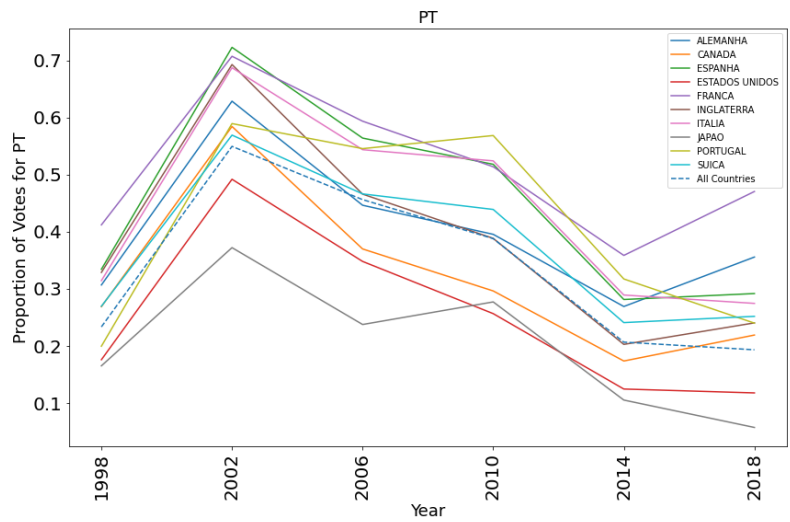
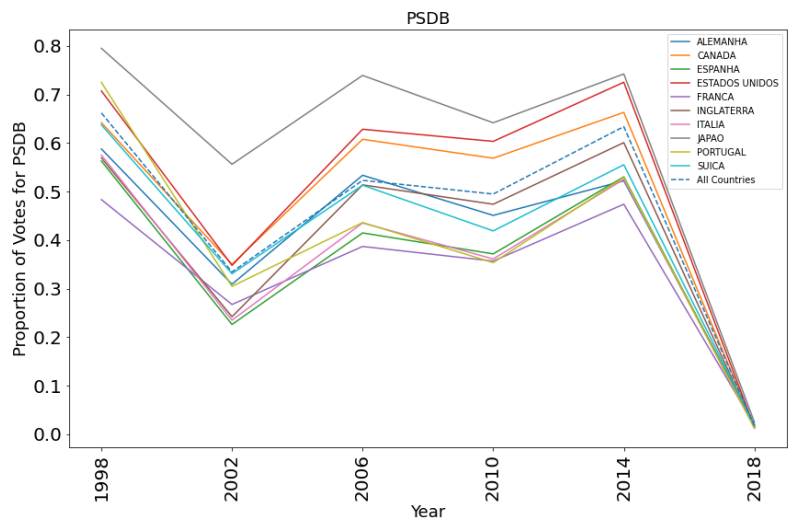Figure 25: Votes for the PT by Year, Broken Down by Country

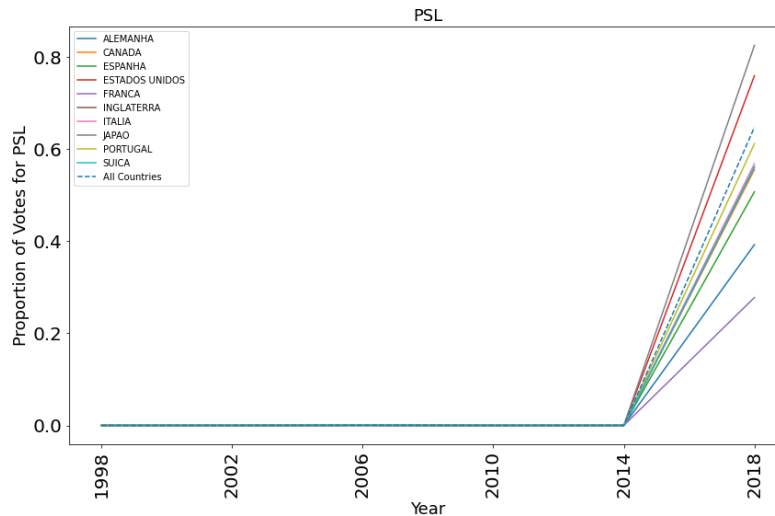Figure 26: Votes for the PSDB, Broken Down by Country

16

Figure 27: Votes for the PSL, Broken Down by Country

Seeing that the PT, PSDB, and the PSL were the most popular parties, we next tracked the proportion of votes that went to each party over the span of our data set to get a clearer picture of the trends in popularity.

Overall the top ten countries move similarly to the aggregate. You can see in Figure 25, that the PT peaked in popularity in 2002, and has fallen in popularity since. Some countries saw an increase in popularity in 2018, but the peak remains in 2002. In Figure 26, you can see that beginning in 2002, the PSBD grew in popularity, but hit a sharp decline in 2018. In Figure 27, the PSL only emerged in 2018, but immediately shot up in popularity.

However there are some notable divergences from the aggregate in some countries. The PSL is extremely popular in the US and Japan, more so than when compared to its popularity among the aggregate electorate. Almost 80% of voters in the US and Japan voted for the PSL in 2018. However, the PSL appears to not be as popular compared to the aggregate electorate in European countries, most notably in France and Germany. Naturally then, the PT is more popular in European countries compared to the US, Japan, and the aggregate.

For the second part of our Election Analysis, we asked: *what percentage of the population votes?* In Figures 28 and 29 we analyze the percentage of the electorate that voted in our top 10 countries for the years 2010 to 2018 (analyzing rounds one and two of the Presidential Election). Our estimations for the voter turnout in these years are based on the number of votes cast in each country divided by the number of eligible voters (members of the electorate older than 18) for that country.

Figures 28 and 29, show that overall, the voter turnout in each country is pretty steady
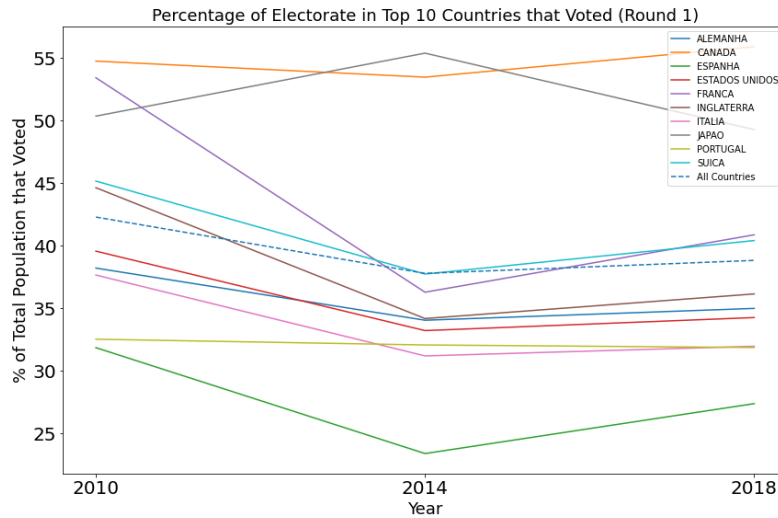
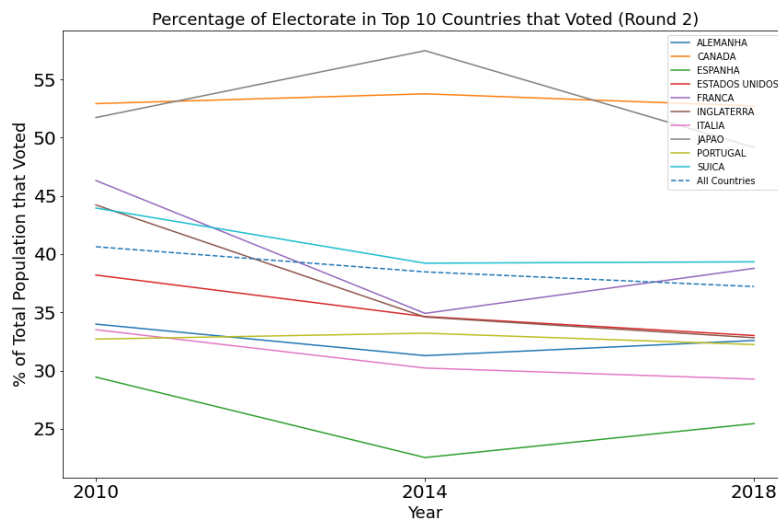Figure 28: Voter Turnout for Round One of the Presidential Election by Year

Figure 29: Voter Turnout for Round Two of the Presidential Election by Year

across the span of our data set, save for France. The voter turnout in each country has a wide range. 4/10 countries were above the data set average in 2018, but Canada and Japan are far above the data set average. Voter turnout in Canada in Japan is around 50-55%, while the next two countries with highest turnout rates (France and Switzerland) only had around 40%. Of 6 countries below the dataset average, most hover in the 33-37% range.

The voter turnout in Spain is especially low, with under 30% of the electorate voting in 2010 and 2018, and under 25% turnout in 2014.

5 Conclusion

5.1 Findings in The Aggregate Electorate

- The population and number of countries represented in the electorate has increased every year.

- Women make up the majority of the electorate (a trend which has continued across all years).

- The predominant education level categories among the Brazilian immigrant electorate are "complete high school" and "complete college".

- Country-level growth rates within the electorate have varied by year, but Japan had the largest growth of any country (across all years analyzed) in 2010.

- The most popular parties are the PSL, PT, and PSDB. The PSL exploded in popularity in 2018 when the PSDB took a significant dive in popularity.

5.2 Findings in The Top Ten Countries

- The United States has the most Brazilian immigrants (and Japan currently has the second most).

- In terms of sex, Germany, Italy, and Switzerland have the most significantly female populations (compared to the rest of the Top 10), with women making up over 70% of the Brazilian immigrant electorate population over all years. However Japan is the only country in the top 10 with a majority male Brazilian immigrant population.

- In terms of education, the majority of Brazilian immigrants in the top 10 countries are either high school or college educated.

- In terms of age, the majority of the Brazilian immigrants in the top 10 countries are between 25 and 59 years old.

- The voting patterns of the top 10 countries (parties the electorate votes for) roughly mirrors the overall data set. However the important key difference would be that in European countries, the PSL is less popular, and the PT is more popular when compared to the aggregate. For the US and Japan, the PSL is more popular and the PT is less popular. There appears to be a difference in the popularity of parties across countries.

- An interesting observation is the Germany and France have the most predominantly female electorate, and at the same time favored the PSL significantly less than the data set average, and the PT significantly more than the data set average.

- Voter turnout is the highest in Canada and Japan, and the lowest in Spain. The rest of the countries vary about the data set average which sits around 40%, but has been declining since 2010.

6 Links to Source Code and Project Documentation

- Code Documentation: `https://github.com/BU-Spark/ds-boston-bpda-br-election/blob/team1-final-deliverable/TEAM1/Code_Description.md`

- Source Code: `https://github.com/BU-Spark/ds-boston-bpda-br-election/tree/team1-final-deliverable/TEAM1/src`