A visual analytics approach to understand vote results of the Chilean presidential elections of 2017

Cristobal Montt-Volosky

Joaquin Perez-Lapillo

City, University of London

# 1. Introduction

## 1.1. Domain overview and motivation

As in other countries – including the US and UK, last presidential elections of 2017 in Chile surprised most analysts and political experts.

Results of the first round showed a surge of the left, thanks to the rising of a new coalition called Frente Amplio founded the same year. Along with them, left parties as a hole accounted for 55 percent of the first-round votes. On the other hand, right party candidates were led by Sebastian Pinera, a former president of Chile. The surprise came a month later when Pinera obtained almost 55 percent of votes on the second round against Alejandro Guillier – the left party candidate who came second on the first round, becoming president for the second time.

These interesting results of the 2017 first and second round elections motivated experts to come with different hypothesis. Some of them suggested that young voters did influenced the first-round results but not the second, and others proposed a change of preferences from left to right of middle age, less educated adults, following the behaviour seen in the US and UK elections and referendum.

Given this context, our aim will be to use a data-driven visual analytics approach to answer questions regarding changes in demographics in Chile that could be explaining such volatile results in the past presidential elections.

## 1.2. Data

There are two main information sources that will be used to carry this study out. First, electoral data will come from the National Electoral Service (SERVEL). This include a dataset at an individual level regarding their voting location and some brief demographics and another dataset with the actual votes per location for the 2017 first and second rounds. Second, for detailed demographics we will use the 2012 Census dataset available at the National Statistics Office (INE) webpage.

Both sources of information include millions of observations and several columns that make them very time consuming in terms of analysis and processing. Given this, our initial approach was to specialize ourselves in one of the two main databases (electoral and demographics) and merge them lately to continue working over the geographical aspect, computation of correlation measures and visualizations.

We will restring our study to Santiago de Chile, the country’s capital. The reasons behind this decision are mainly two: the city accounts for almost half of the population, being a fair sample of the country’s demographic and vote distributions and the high level of detail that the databases offer, that would make visualizations computationally very expensive.

## 1.3. Research question

We would like to answer the following questions:

1. What are the most relevant demographic variables explaining the results of the first and second rounds of the 2017 presidential election?
2. Which variables explain better the “swing to right” between rounds?
3. How does votes and demographic variables relate in terms of geographic areas?

# 2. Tasks and approach

## 2.1. Analytical tasks

### 2.1.1. Understand geographical patterns of votes of first and second rounds

We will use visualization to identify where the population were most likely to vote right or left for both elections. This will be conducted by creating a hexagon map of Santiago de Chile that would help us compare percentage of votes of equally sized regions, as opposed to a normal choropleth map. In addition, we will plot on the same hexagon map different demographics variables such as level of education and age groups as an exploratory way of identifying correlation between those variables and votes. This analytical task will help us answer research questions 1 and 3.

### 2.1.2. Identify most relevant demographics in terms of correlation with votes

The dataset was previously filtered (from more than 100 attributes) to obtain XX demographic features from which we will identify the most relevant in terms of its correlation against right-vote intensity. This will be done with a combination of visual techniques, i.e. 2D scatterplots and linear regressions and the computation of correlation coefficients that will be also visualized in a correlation matrix. For more specification, we will include in the 2D scatterplots a colour differentiation by four regions of Santiago de Chile: north, south, west, east and centre. This task is key to allow a quantitative measure of the relationship between demographic variables and votes, helping understand research question 1.

### 2.1.3. Identify the relationship between demographics and swing

Similar to the previous analytical task, we will identify most relevant demographic features in terms of correlation with swing-to-right by combining visual and computational techniques such as 2D scatterplots and the calculation of correlation coefficients. This task is key to answer to research question 2.

## 2.2. Approach

### 2.2.1. Left and right vote

The current political spectrum in Chile is clearly defined by four groups: left, centre-left, centre-right and right. For analytical and simplification purposes, we define “left” and “right” vote as a summation of left and centre-left votes and right and centre-right. This assumption reduces the complexity of the problem, allowing us to do meaningful visualizations and analysis.

For the first-round elections, the candidates were previously classified into the two groups matching their political party association to a left-right scale from previous work in the area[[1]](#footnote-1). The eight candidates were classified as follows:

Where:

AG = Alejandro Guillier; BS = Beatriz Sanchez; CG = Carolina Goic; MEO = Marco Enriquez-Ominami; EA = Eduardo Artes; AN = Alejandro Navarro; SP = Sebastian Pinera; JAK = Jose Antonio Kast.

The second-round election only considered the two most voted candidates of the first-round, Sebastian Pinera (right) and Alejandro Gullier (left). Consequently, on the second-round vote analysis we compare just votes of these two candidates.

### 2.2.1. Swing

As defined by political experts, swing is “*the statistical measure by which the switch of voters from one party to another on a national or constituency basis can be judged. It is calculated by adding the rise in one party’s vote to the fall of the other and dividing by two.”* (Comfort, 1996)*.* Using this definition, our calculations of the swing-to-right vote between first and second round elections will be obtained by:

Where:

%Right (or Left)1 (or 2) = percentage of votes obtained by right/left side on the first/second round.

As we kept only two groups because of the aggregation of votes, the swing-to-right calculation can be also simplified to just:

The analysis of swing-to-right will allow us to identify which groups of the population of Santiago de Chile changed their political preferences between the two elections.

# 3. Analytical steps

## 3.1. Geographical analysis of votes and demographics

## 3.2. Correlation analysis of demographics against votes

## 3.2. Explaining “swing to right”

# 4. Findings

# 5. Critical reflection

# References

Comfort, N. A. (1996). *Brewer's Politics: A Phrase and Fable Dictionary.* Weidenfeld Nicolson Illustrated.

1. https://es.slideshare.net/joseantoniobarriga/partidos-polticos-chilenos-51444635 [↑](#footnote-ref-1)