Explaining Vote Patterns in the Chilean Presidential Elections of 2017

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.

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. Because of this, our initial approach was to specialize ourselves in one of the two 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.

Research Question

We would like to answer the following questions:

- What are the most relevant demographic variables explaining the results of the first and second rounds of the 2017 presidential election?

- Which variables explain better the “swing to right” behaviour?
- Is this relationship the same across geographic areas in Santiago de Chile?

Tasks and approach

Transform and merge datasets

Select relevant variables by visualization

Analytical steps

Findings

Critical reflection