
Forecasting the 2014 Toronto Mayoral Election

PsephoAnalytics



psephoanalytics.blogspot.ca/
[@PsephoAnalytics](https://twitter.com/PsephoAnalytics)
PsephoAnalytics@gmail.com

September 12, 2014



Overview

Campaigns have limited resources—both time and financial—that should be spent on attracting voters that are more likely to support their candidates. Identifying these voters can be critical to the success of a candidate.

Given the privacy of voting and the lack of useful surveys at the municipal level, there are few options for identifying individual voter preferences:

- City-wide polling, which is large-scale, but does not identify individual voters
- Voter databases, which identify individual voters, but are typically very small scale
- In-depth analytical modeling, which is both large-scale and helps to 'identify' voters at a neighbourhood level on average

In this paper, we describe the results of some modeling—at a neighbourhood level—of which candidates voters are likely to support in the 2014 Toronto mayoral race. All of our data is based upon publicly available sources (see the [appendix](#)). So, the resources required are *far* less than those of polling or creating voter databases. Through these attempts, we explore the limits of such data and how they can best be used for informing candidates' strategies.

Voters aren't stuck in their ways

The most effective strategy for a candidate depends a great deal on the extent to which individual voters change their minds during election campaigns. If the favoured positions of voters change, then candidates can focus on persuading voters that the candidate's position on an issue is the best one. Otherwise, the main objective is to "get out the vote" and ensure that those voters who already agree with you actually cast their votes. And, ideally, your opponent's voters just stay home.

Voters have revealed that their preferences aren't stuck in stone. As shown in Figure 1, the positions of voters writ large changed dramatically in 2010 from predominantly left positions to predominantly right.

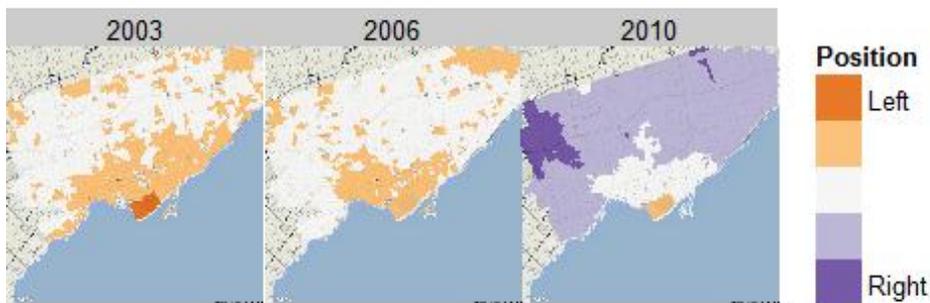


Figure 1 - The position of voters writ large changed dramatically from 2006 to 2010.

Is it the case that actual *individual* voters are changing positions or voting strategically (and are those the same?), or is it simply that different voters are showing up? We can see in Figure 2 that turnout increased in 2010, a particularly polarizing election. Perhaps more right-leaning voters simply turned out to vote in 2010, rather than individual voters changing their positions.

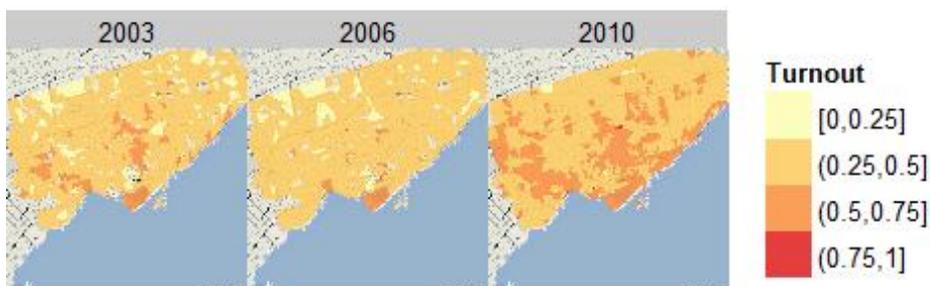


Figure 2 - Voter turnout increased substantially in 2010.

To try to separate out these effects, we specifically looked at areas where turnout was above 60% in all three elections to see if those areas changed positions. Turns out that they do (by 0.1 on the position scale, on average)—suggesting that it's not simply voters waiting for the 'right' candidate to come along. Many voters are, in fact, supporting different positions from one election to the next.

These results suggest good opportunities for candidates to, at least partially, "run on the issues", rather than relying on any traditional voting patterns. Unfortunately for this work, these results also introduce significant variation—and therefore, uncertainty—into the modeling.

Having looked at the past, we can now move to analyzing the upcoming election, based on these patterns.



Proximity voting and calibration

In order to analyze the upcoming election, we use previous election results and apply both the theory of 'proximity' voting (see the [appendix](#), but roughly a sense of taking similar positions on issues) and a 'voteability' variable to calibrate to the most recent polls.

'Voteability' covers numerous broad factors, including:

- How well known the candidate and their positions is/are
- How 'likable' and 'trustworthy' the candidate is
- Perceived viability of the candidate winning
- How strategically voters are voting (e.g., choosing the 'lesser of evils' if their preferred choice is perceived to be nonviable)

This does not yet say *why* voters vote the way they do, but does identify *where* those voters may be.

Candidate positions *alone* don't predict outcomes

If strictly based on currently¹ stated positions, John Tory would win a three-way Toronto mayoral election in 2014 with 50% of the vote, followed by Olivia Chow at 32.8%, and Rob Ford would place a distant third at 17.2%.

As we can see in Figure 3, support for Tory is widespread, while Chow's support is concentrated around the Bloor-Danforth subway line, roughly. Ford has moderate support around the downtown core.

¹As of 4 September 2014

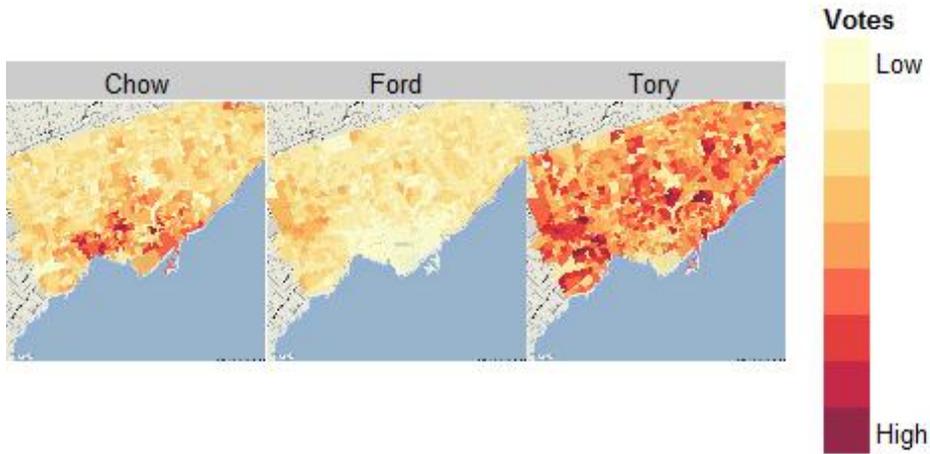


Figure 3 - Votes cast for each candidate, based only on their stated positions.

This assumes that voters only consider the positions taken by candidates on all of the issues in the campaign. But, of course, this is not true. As described above, the 'voteability' of the candidate is also important to voters. Calibrating to recent polls² suggests that John Tory and Olivia Chow have roughly equivalent 'voteability', but both are about 46% of Ford's score.

Taking this into account, the election—if held today—might see the distribution of votes cast for each of candidates shown in Figure 4.

²As of 4 September 2014

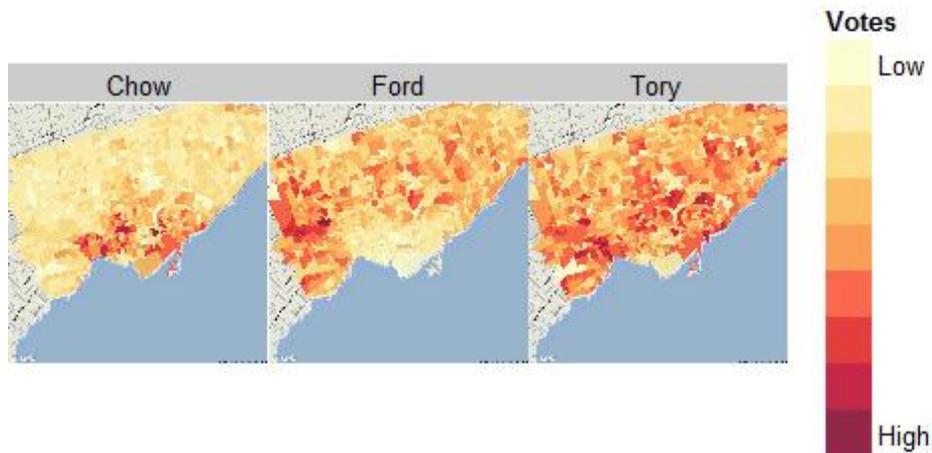


Figure 4 - Votes cast for each candidate, after taking voteability into account.

This distribution of votes gives Tory the lead with 42.3% of the votes. Chow is far behind with 26.9% and Ford is second with 30.8%.

A couple of issues have been important

So far, we've treated all issues the same and derived an overall right-left score for each voter and candidate. Now, we turn to considering each issue separately. We do this by estimating ward-level support for each issue through the votes that each candidate received in the ward. So, for example, if a particular candidate had a strong left position for airport expansion (i.e., fully against expansion) and received many votes in a particular ward, that ward has a strong left position for airport expansion.

Of course, since *candidates*—not issues—receive votes, we need to use variation in positions taken by candidates and the votes they receive to simultaneously estimate support for particular issues. This was done as a linear model of each major issue (roads/bike lanes called "transportation", transit, finance and budget, waste management, and airport expansion) on the number of votes for each ward.



We find that, over the last three elections, various issues have been influential in the way wards have voted (see Figure 5), but two key issues certainly stand out:

- airport expansion: top issue overall; top issue in 30 of 44 wards; close second in another 6. Voters have typically voted against expansion on average
- transportation (roads and bike lanes): Top issue in 10 wards; close second in another 6. Votes split depending on ward (indicating a fight)

Finance and budget, transit, waste management, and other issues have been much less important historically, on average.

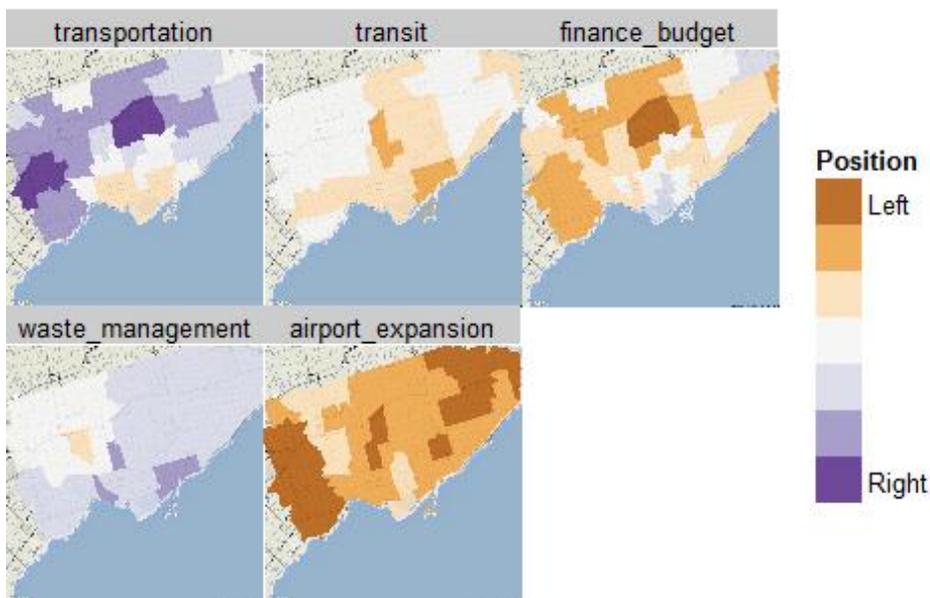


Figure 5 - The distribution of right-left scores for each issue.

Looking at the distribution of these top issues in Figure 6, we can see the dominance of the airport expansion. We can also see that transportation issues are most important in the north-west of the city.

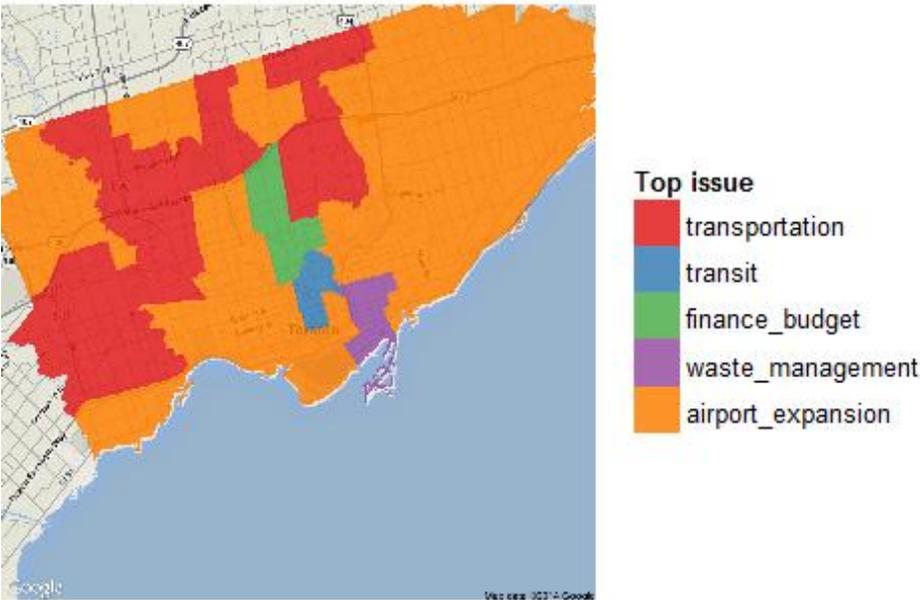


Figure 6 - The top issue for voters changes by ward, with airport expansion and transportation dominating.

If these issues were still important...

If voters placed the same importance on issues as they have over the last three elections (i.e., not weighted equally as above), then strictly based on currently stated positions³, Tory would still win with 49.3% of the votes, but Chow would narrow the gap significantly with 40.5% (see Figure 7). Ford would come in third at 10.2%. While this may not be entirely realistic, the distribution of results would be:

³As of 4 September 2014

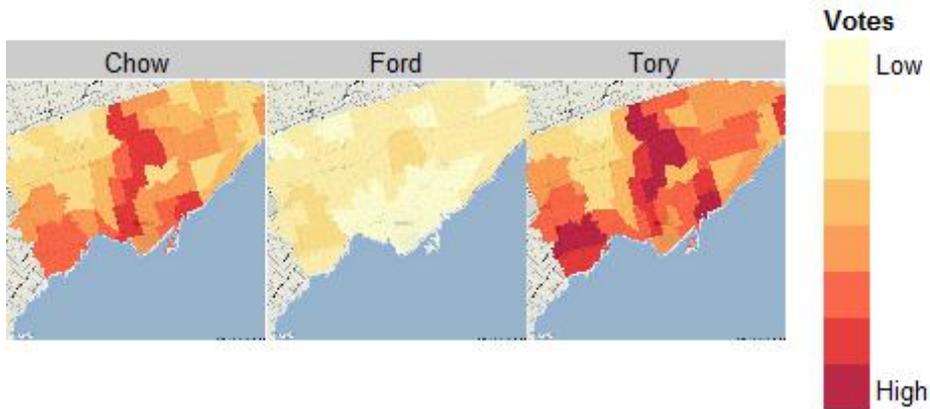


Figure 7 - The distribution of votes, based on specific issues.

This implies that Chow has more closely targeted her positions to specific issues. If voters only assess the overall platform of each candidate, Tory is the clear winner. But, if voters dissect platforms into specific issues and vote based on this dissection, Chow is much more competitive.

Neighbourhood characteristics are descriptive

So far, we've focused simply on historical voting patterns and stated positions by candidates. Now, we try to use census data from Statistics Canada to better understand the characteristics of neighbourhoods that could explain why people favour certain candidates.

To do this, we use a statistical method called principal-component analyses (see the [appendix](#)) along with the 2011 Census data. There are three general neighbourhood components that have driven voting over the last three elections (see Figure 8). Neighbourhoods with:

- Younger families, fewer children, many transit users, and below-average home values tend to vote left (e.g., Trinity-Spadina). Those with older families with more children, who tend to drive and have expensive homes tend to vote right (e.g., Don Valley West).
- Fewer children and above-average home values tend to vote left (e.g., Toronto-Danforth). Those with more children and inexpensive homes tend to vote right (e.g., York West).



- Parents, transit users, and expensive homes tend to vote left (e.g., Eglinton-Lawrence). Those with fewer children, many drivers, and less expensive home tend to vote right (e.g., Etobicoke North).

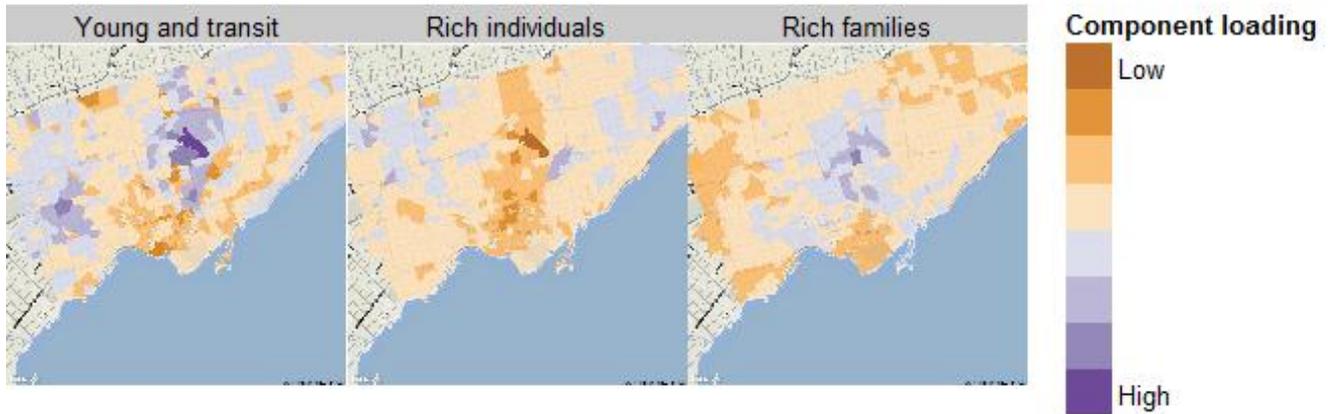


Figure 8 - The distribution of PCA scores, based on the 2011 Census. For each component, high loadings are correlated with left positions.

Although these three components are statistically significant parameters in a model of voting patterns, it is important to point out that, overall, the model only explains a small portion of variation in voting ($R^2 = 0.1$). So, the results might only be taken as "descriptive".

Limitations of these approaches

We think this modelling is interesting and has certainly helped clarify our thinking about both this election and the general problem of modeling elections. But, there are at least two major limitations:

1. Because these are linear models, they assume linear changes in support (i.e., tends towards the extremes). Non-linear (e.g., parabolic) models might be more appropriate, but there are many to choose from and we haven't yet looked at their strengths and weaknesses in the context of voting preference.



2. Voters writ large appear to readily change their positions from election to election, with the 2010 election seeming particularly different. The main issue that we don't know individual preferences, which adds significant noise to the model estimates. More data on voter preferences would obviously be really helpful, but we aren't aware of any public sources for Toronto voters.

These concerns don't invalidate the approaches. But they led us to moving on to something completely different, where we use an agent-based approach to simulate entire elections. We are actively working on this now and hope to share our progress soon.

Conclusions

This work demonstrates that significant insights on the upcoming Mayoral Election in Toronto can be obtained from an analysis of publicly available data. In particular, we find that:

- Voters will change their minds in response to issues. So, "getting out the vote" is not a sufficient strategy. Carefully chosen positions and persuasion are also important.
- Despite this, the 'voteability' of candidates is clearly important, which includes voter's assessments of a candidate's ability to lead and how well they know the candidate's positions.
- The airport expansion and transportation have been the dominant issues across the city in the last three elections, though they may not be in 2014.
- A combination of family size, mode of commuting, and home values (at the neighbourhood level) can partially predict voting patterns.

Appendix A - Data

There are two data sources for the issue model:

1. The **elections data** from the city tells us how many votes each mayoral candidate received in each of the past three elections
2. We've collated the positions taken by each of the major mayoral candidates on each of the major issues over those elections. Each position is ranked on a left-right score and taken from public sources (predominantly media). While certainly subjective, there is at least internal consistency.

Votes

- 2003 Election: 692,085 votes cast for 44 candidates and 38% turnout, split by 1,926 polling subdivisions
- 2006 Election: 584,484 votes cast for 38 candidates and 39% turnout, split by 1,946 polling subdivisions



- 2010 Election: 813,984 votes cast for 40 candidates and 50% turnout, split by 1,870 polling subdivisions

Candidate positions

- 2003 Election: positions of four candidates (Miller, Tory, Hall, Nunziata) ranked on 6 issues
- 2006 Election: positions of three candidates (Miller, Pitfield, and LeDrew) ranked on 5 issues
- 2010 Election: positions of four candidates (Ford, Smitherman, Pantalone, and Rossi) ranked on 17 issues
- 2014 Election: positions of five candidates (Ford, Chow, Tory, Stintz, Soknacki) and 'rest of field' ranked on 11 issues so far, where positions are public

Voters

- 2011 Census: commuting, family income, age, and sex by census tract (1,149)

Appendix B - Theory

Proximity voter theory

Spatial voting theory supposes that candidates and voters have positions in a policy space and that these positions determine the voter's preferences and behaviour.

Given that the Toronto mayoral race is unlikely to focus on any significantly polarizing issues, use of the proximity voting model is reasonable. Further, given that Toronto elections are non-partisan (and therefore focus on evaluations of incumbents or specific policy issues), incorporating specific issues the proximity approach seems appropriate.

We apply a normal distribution around the average position for each area to calculate the likely support for each candidate within that area. This is scaled up or down using the 'voteability' variable, and then all support is normalized to the average number of votes in previous elections - essentially estimating how many votes each candidate would garner.

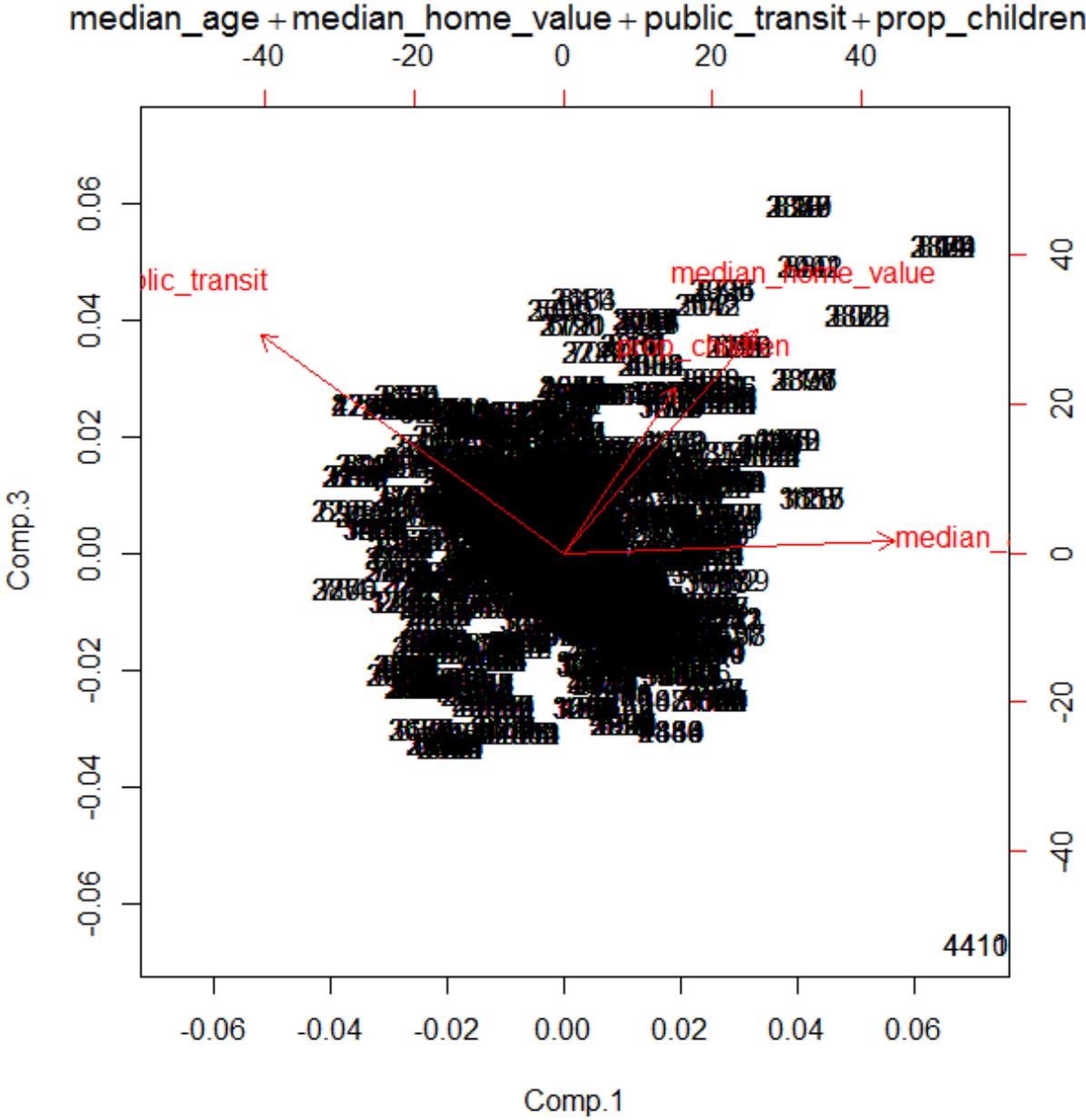
Principal-component analyses

Principal-component analysis (PCA) is a useful statistical technique commonly used for finding patterns in data with many variables. It is a fundamental part of modern data analysis, and is used in the likes of facial recognition, meteorology, and computer graphics.

PCA filters out noise and redundancy to better express 'messy' data by providing components that account for, in order, the most variation between variables. This is true even when the researcher does not know how many relationships there may be.



The main benefit is that PCA can take multi-dimensional data and reduce it to two or three 'components', allowing for graphical representation with a minimal loss of useful information. These components typically account for most of the variability in the data.



An example of the output from a principal-component analysis.

