

What the presidential candidates' data can tell us about Trump and Clinton

October 18 2016, by Rick Hutley

It's election season, and the candidates' and campaigns' eyes are on you, the voter. Figuring out what you think about something a candidate said last night or tweeted this morning is very big business. All this gathering of data, from [statewide and national polls](#) and social media alike, can make it seem as if everything we do – or even think – is under scrutiny. In fact, it is.

As a result, elections seem very one-sided: Campaigns can get detailed data allowing them to read, see, hear and analyze almost everything we do. But what we, the people, get for analysis is mostly pundit commentary, not the kind of real analysis that uses data as its source. We are, therefore, left to decipher and discern among often-conflicting perspectives amid the cacophony of online reports, newspaper articles or TV broadcasts.

[Fact checking](#) the [candidates](#) is also [big business](#), but it tells us more about what the candidates say than about the candidates themselves. If only we could get access to data about the candidates! Then we could do our own analysis, just as they do.

To a large degree, it turns out that we can. Thanks to the vast scope of the internet, we can now turn the tables on the candidates and their campaigns and obtain a wide variety of data, such as [voter preferences](#), [campaign profiles](#); [corporate and foundation annual reports](#); and corporate tax information. As I'm teaching my Data Science students,

this broad range of factual data allows us to do our own analysis of the candidates, even as the campaigns analyze us.

Determining what to analyze

Some of the data you might like to collect for analysis about individual candidates simply are not going to be available – to you or anyone else – unless the candidates choose to make such information available. For example, health or tax records. But some data are available that are unequivocal: debate transcripts.

Debate transcripts are like court transcripts – they are an accurate, factual rendition of who said what. That makes them a very reliable source of information about candidates – devoid of bias or other influence that may be presented in third-party blogging or reporting about the debate.

Similarly, [social media](#) postings from the candidate directly or on official campaign accounts are excellent sources of data. When we subject them to computer analysis, we can learn many things about the candidates based on how they express themselves.

Initial analysis

The transcript can certainly tell us who spoke most, but that's not the whole picture. How much someone is talking isn't enough. What are they talking about, and how are they using language to discuss their topics? And how about emotion?

The field of [natural language processing](#) offers a wide range of techniques for summarizing large blocks of text, identifying names, identifying core topics and so on. Google has recently released two

programs that make this much easier for nontechnical users to explore: "[SyntaxNet](#)" and "Parsey McParseFace" ([its real name](#)).

A simple word count of the words spoken during the [16 primary debates](#) that took place up to February 2016 suggests that Hillary Clinton spoke about 20 percent more words than Donald Trump. By a simple count, she was the most prolific speaker of all of the candidates in these debates. But that's not the whole picture. Some candidates may have fielded more questions than others, or been given more leeway to speak at length. When we account for these and other factors – such as how many debates a candidate attended and how many other participants there were – a very different picture emerges: Trump is in fact the most verbose candidate, and exceeds Clinton by around 18 percent.

The quantity of talking isn't enough. We also need to look at the issues they are talking about, [their vocabulary](#) and the emotions they apply. Clinton uses a wider vocabulary: Using the combined data from these primary debates, she used around 2,300 distinct word bases or stems (counting related terms such as "vote," "voter" and "voting" as a single term). Trump used a much smaller vocabulary of only 1,750 stems.

Clinton uses lengthier, more sophisticated sentence constructions – scoring around 12 on the [Gunning Fog Index](#), which measures the complexity of language – while Trump uses tweet-like short phrases that score a 7. This suggests Clinton is seeking to communicate with a more educated and socially sophisticated audience, while Trump makes an effort to be readily understood at all socioeconomic levels.

We can also use sentiment analysis to get a sense of the language and emotion in the debate. We can determine whether a candidate is under stress or remaining calm by looking at the tone of the words used, or whether they are imparting a positive or negative message. [Analysis of the first presidential debate](#) shows the two candidates were close: Clinton

used 53 percent negative terms while Trump used 55 percent. She is also [more positive when tweeting](#).

Turning to social media

We could also delve deeper into the debate transcripts to look at things like the frequency with which specific topics are addressed, or how the candidates' debate styles, messages and sentiments change over time. But let's take a quick look at another valuable source of information: social media.

Twitter, Windows Messenger, Instagram and other sites provide a new and exciting window onto what is being said and thought by society at large. [These platforms](#) allow us to [download streams](#) of [data for analysis](#). With just a few lines of programming code you could, for example, get the latest tweets from either or both of the candidates – and often at no cost.

A sentiment [analysis](#) of their tweets could reveal how the candidates use social media, and what they're saying to their audiences on those services. As was found in an [analysis of which device Trump's account tweeted from](#), they can even reveal whether a candidate is [tweeting personally](#), or whether it's a campaign staffer standing in.

The internet and social media give us access to a wide variety of [data](#) that gives the public insight into facts and tendencies behind the public statements and claims. Even as the candidates and campaigns scrutinize our every click and post, we can keep our own eyes on them too.

This article was originally published on [The Conversation](#). Read the [original article](#).

Source: The Conversation

Citation: What the presidential candidates' data can tell us about Trump and Clinton (2016, October 18) retrieved 26 April 2024 from <https://phys.org/news/2016-10-presidential-candidates-trump-clinton.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.