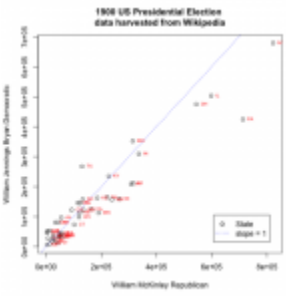
Visualizing US Presidential Elections

2013-08-08 BY GENE



I am a curious cat by nature, and a visual one at that, so back in 2012 I made a little scatter plot of the popular votes in the recent presidential election…

{ “tl;dr” : “https://github.com/ology/Election-Viz/ & 1900 to 2012 elections” }

It was mildly interesting. It showed which states favored which candidate and by how much.

A couple weekends ago, I noticed the project sitting incomplete on my shelf (A.K.A. `~/devel/R/`) and decided to resurrect it. I wanted to see the same type of charts for multiple years and with reference lines. Besides the *popular vote*, there are other interesting columns too: *percentage votes* and *electoral votes*.

Also, I desired to harvest and integrate other state facts for those years, like *population*, *voter turnout*, etc.

So I had my marching orders:

1. Harvest the election data
2. Process (sanitize) the election data
3. Plot the election data

Simple!

For **#1** I initially had just copy-pasted the popular vote data for 1012, by hand, to a file. But being a programmer, this type of manual drudgery is abhorrent. I'm far too lazy for that.

So first I looked for available, friendly comma-separated-value files to just download and skip to step #2… No luck. Then I decided that I trusted the wikipedia sources and their relatively consistent tabular format.

Because it seemed like an interesting challenge, I tried to parse the raw wikipedia markup but nearly went (more) insane doing it. Fortunately, I have used the handy Perl module HTML::TableExtractor to great success. So a bit of LWP plus some caching logic and I was harvesting!

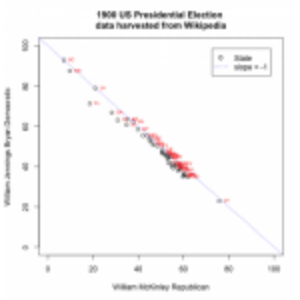
#2: Because wikipedia entries have varying ways to list the candidates, the table headers and the actual data itself, I had to (possibly) sanitize the row cells. Perl to the rescue again! Its regular expression engine makes text manipulation a breeze. (And yes, I know all about string operations, and even interweb fetching in R, but 1. I was already harvesting the table with Perl and 2. I wanted to keep the division of labor as independent as possible.)

For your inner geek, check out my wikipedia-scrape code at github.

#3: For plotting and stats in general, I use R. **I love R.** It is insanely powerful.

As such, a big part of this project was to learn more about R! And indeed I did: Fetching `list.files()`, using a custom `axis()` and `pretty()`, using `ifelse()` to toggle argument values, using `identify()` to allow mouse interaction… Also I made friends with `paste()` the cousin of Perls `join()` and learned about how to manhandle `legend()`.

Anyway, when I added a 45° slope=1 line, it made the division between states obvious. This is a reference line, not the result of linear regression, btw. When I graphed the vote percentage columns, the state division was even more obvious. Then reference line serves to show how many other candidates received votes *besides* the two being plotted.



For your inner geek, check out my presidential-candidates.R code at github.

Here are all the charts for the 1900 to 2012 elections.

FILED UNDER: DATA, INTERNET, SOFTWARE
TAGGED WITH: ELECTIONS, HTML::TABLEEXTRACTOR, PERL, R

Epistemologist-at-large

^ Top