Multiple Testing, Modern Inference, and Replicability
Final Project: Gerrymandering
(Progress Report)

Peter Gao, Casper Neo, Nishesh Sharma

3 March 2017

1. Motivations of a partisan re-district-er

   (a) A partisan incumbent will to try to win as many districts while still maintaining a margin of safety. To accomplish this goal the re-district-er will draw districts such that the other party's voters are concentrated in a few districts (voter packing) while their party's voters are spread out. The effect would be to win many districts with a narrow margin while losing few with large margins.

   (b) Bipartisan re-district-er may want to increase stability and protect the incumbents. To accomplish this goal they will draw districts such that all incumbents win by margins large enough that their seat will not be contested.

   (c) Something about the notion of stability

2. Methods Implemented

   (a) Efficiency Gap (Stephanopoulos and McGhee 2015)

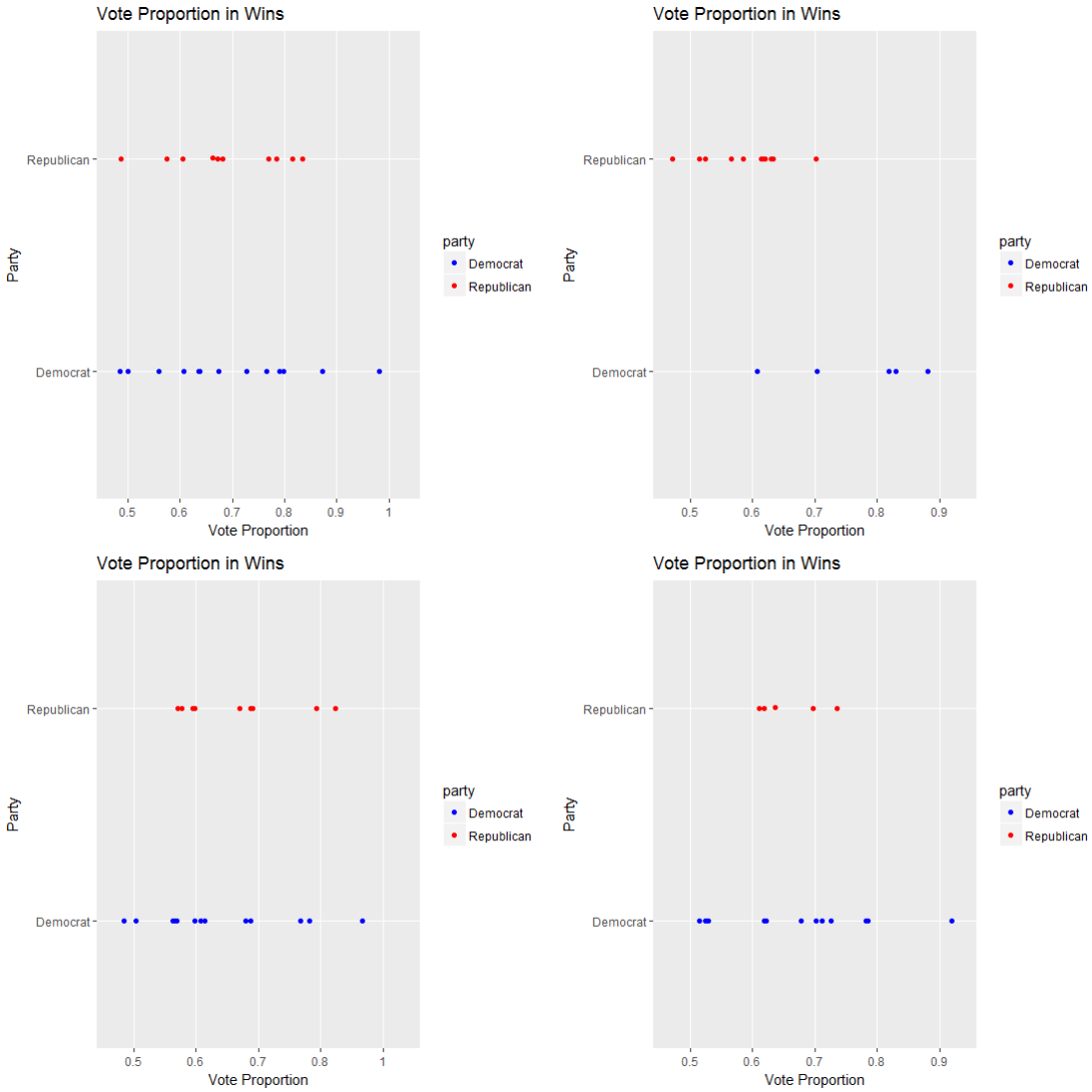
   Any votes cast towards a losing candidate is considered wasted. And all the votes cast towards a winning candidate in excess of the number of votes required to win is considered wasted. Let $wasted(A)$ be the number of wasted votes cast by party A. Then define the efficiency gap as

$$\text{Efficiency Gap}(A, B) \stackrel{\text{def}}{=} \frac{wasted(A) - wasted(B)}{\#\text{votes}}$$

   (b) Partisan Bias

   (c) Lopsided Victories (Wang)

   If there is partisan gerrymandering then as evidence of voter packing, one party will win by large margins while the other party's will win by small margins. We use a two sided T test to compare the means of the winning margin of the two parties. In order to deal with elections where one party ran unopposed, we added an imputed value parameter that represents the assumed percentage of votes that the winning party would have received had there been an opponent. In most cases, tweaking this parameter's value does not significantly alter the p-value for a given state and year. However, it is possible that across many years and many states, the choice of imputed value is significant. We plotted the vote proportions in wins for both Democrats and Republicans for many different states and years. Some sample graphs and p- values are displayed on the following page:



These graphs correspond to PA 1984 (p-value of 0.679), PA 2012 (p-value of 0.03), IL 1984 (p-value of 0.843), and IL 2012 (p-value of 0.766) respectively from left to right.

3. Preliminary Results and challenges

   (a) One of the initial challenges was finding data. "http://projects.iq.harvard.edu/eda/" provides shape file and election data. However we found that the data formats were not uniform over different states (i.e. there were columns missing or named differently). "https://www.ncsbe.gov/Election-Results" provided data over just North Carolina.

4. Simulations (To be completed)

   (a) Sampling from Simulated Delegations - Excess seats (Wang)

   (b) Perturbation of Existing Map

   For a given election outcome consider for N simulations, exchanging k voting blocks along the boundary of voting districts. If the original voting districts were gerrymandered we would expect our metrics to decrease in a majority of the simulations assuming the gerrymander-er would have, in gerrymandering,

chosen a local maximum of our metrics by choice of map.