Presidential Election Results

R and SAS Final Project

### Authors:

Andrew Dunne

Mahesh Gaya

Jonathan Rudnick

Presidential Election Results

Introduction

For our project we collected data from Github.1 The data shows a wide variety of information about the 2012 and 2016 presidential elections. We were interested in this information because it gives us the opportunity to better understand the factors that resulted in Donald Trump’s win over Hillary Clinton. We plan to learn about the voting habits by county, state, and region in 2012 and 2016 in order to help understand the dynamic for the election in 2020.

Data Description

The data was scraped from Townhall.com’s county-by-county map by Rony McGovern. Below is a detailed description of the data:

**Total number of rows:** 3142  
**Total number of columns:** 21

**Row description:**

Each row represents a county

**Column description:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Data Type** | **Description** |
|  | Integer | Numbering of each row |
| **Combined\_fips** | Integer | Combined county and state FIPS |
| **Votes\_dem\_2016** | Integer | Votes received by the Democratic candidate in 2016 |
| **Votes\_gop\_2016** | Integer | Votes received by the Republican candidate in 2016 |
| **Total\_votes\_2016** | Integer | Total votes in 2016 |
| **Per\_dem\_2016** | Double | Vote percentage received by the Democratic candidate in 2016 |
| **Per\_gop\_2016** | Double | Vote percentage received by the Republican candidate in 2016 |
| **Diff\_2016** | Integer | Vote margin in 2016 |
| **Per\_point\_diff\_2016** | Double | Percentage margin in 2016 |
| **State\_abbr** | String | Two-letter state abbreviation |
| **County\_name** | String | County name |
| **FIPS** | Integer | Federal Information Processing Standard (FIPS) |
| **Total\_votes\_2012** | Integer | Total votes in 2012 |
| **Votes\_dem\_2012** | Integer | Votes received by the Democratic candidate in 2012 |
| **Votes\_gop\_2012: Votes received by the Republican candidate in 2012** | Integer | Votes received by the Republican candidate in 2012 |
| **County\_fips** | Integer | County FIPS |
| **State\_fips** | Integer | State FIPS |
| **Per\_dem\_2012** | Double | Vote percentage received by the Democratic candidate in 2012 |
| **Per\_gop\_2012** | Double | Vote percentage received by the Republican candidate in 2012 |
| **Diff\_2012** | Integer | Vote margin in 2012 |
| **Per\_point\_diff\_2012** | Double | Percentage margin in 2012 |

Data Cleaning

We removed the columns combined\_fips, FIPS, county\_fips, state\_fips because they are unnecessary to do any analysis for our project.

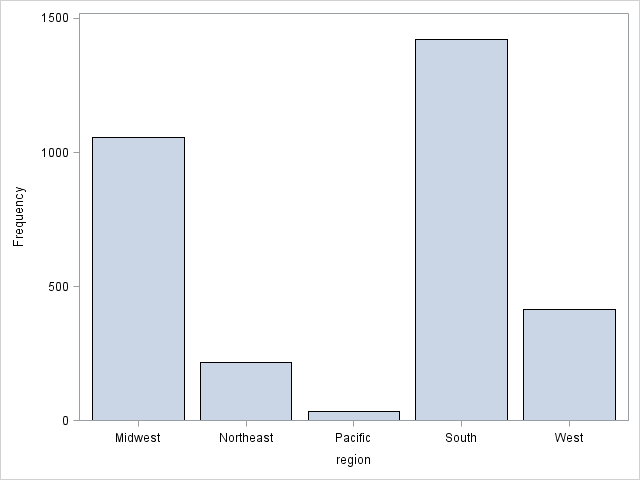
Binning

We binned the data into the following categories:

1. Region Categories 2
   1. Northeast
   2. South
   3. Midwest
   4. Pacific
   5. West
2. DEM County Winner 2012
   1. Yes
   2. No
3. GOP County Winner 2012
   1. Yes
   2. No
4. DEM County Winner 2016
   1. Yes
   2. No
5. GOP County Winner 2016
   1. Yes
   2. No
6. State Winner 2016 3
   1. Trump
   2. Clinton

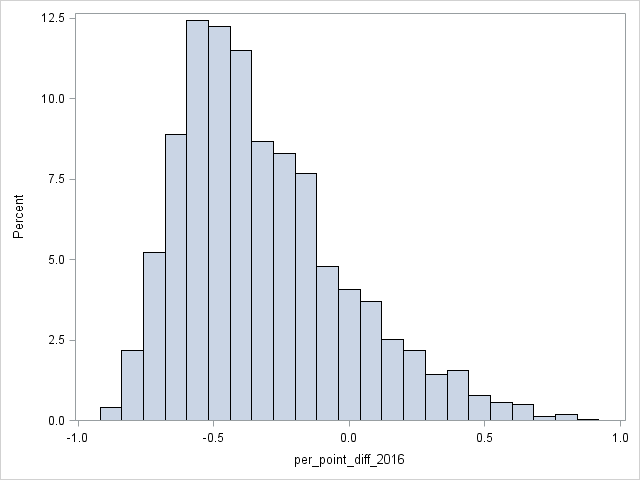
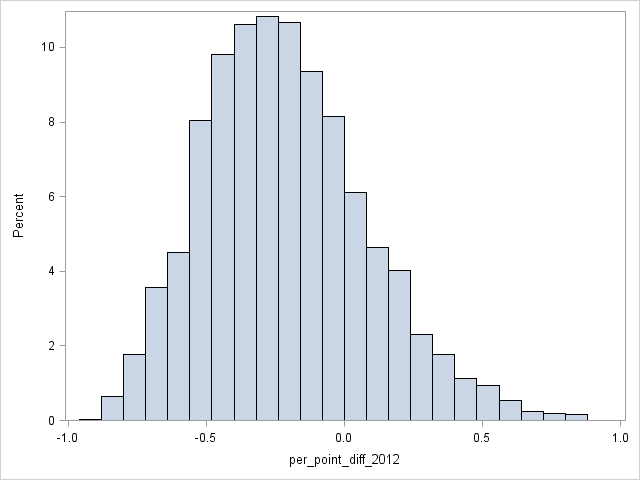
Visuals

Number of Counties in Each Region Bar Chart



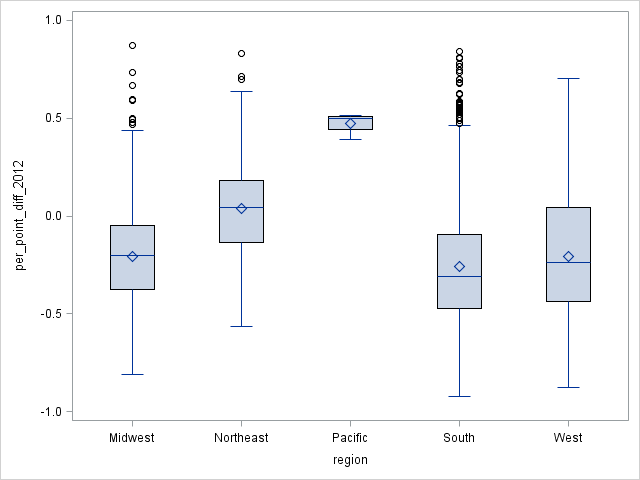
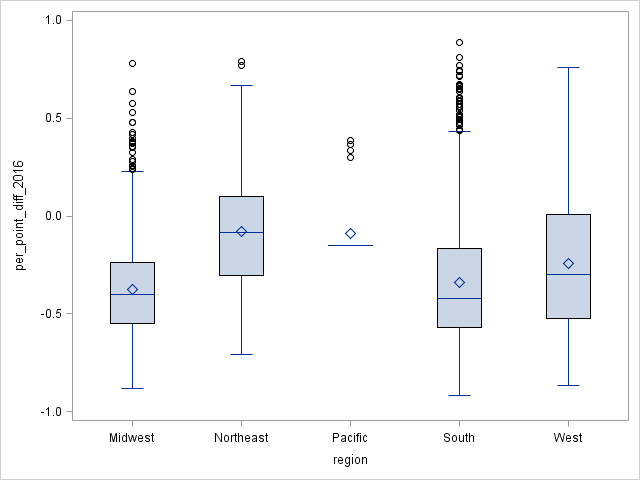
This chart shows the importance of each region and the impact of individual counties. For instance, we can see that the Pacific region is smaller and the south region is much larger. This impacts the amount of effort that should be devoted to each region.

Percent Per Point Difference Histogram

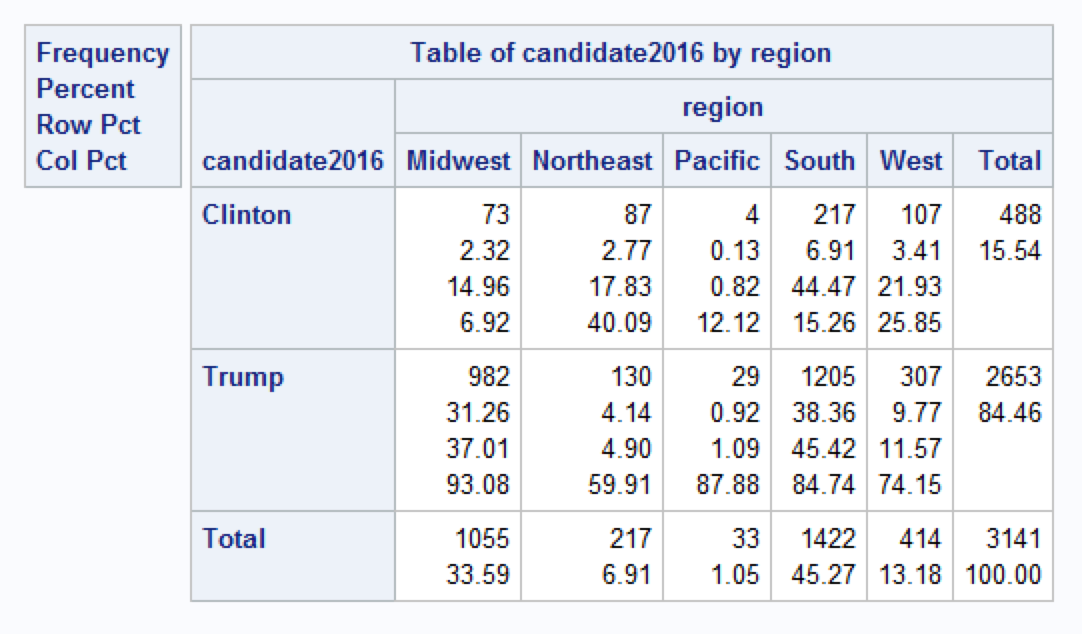
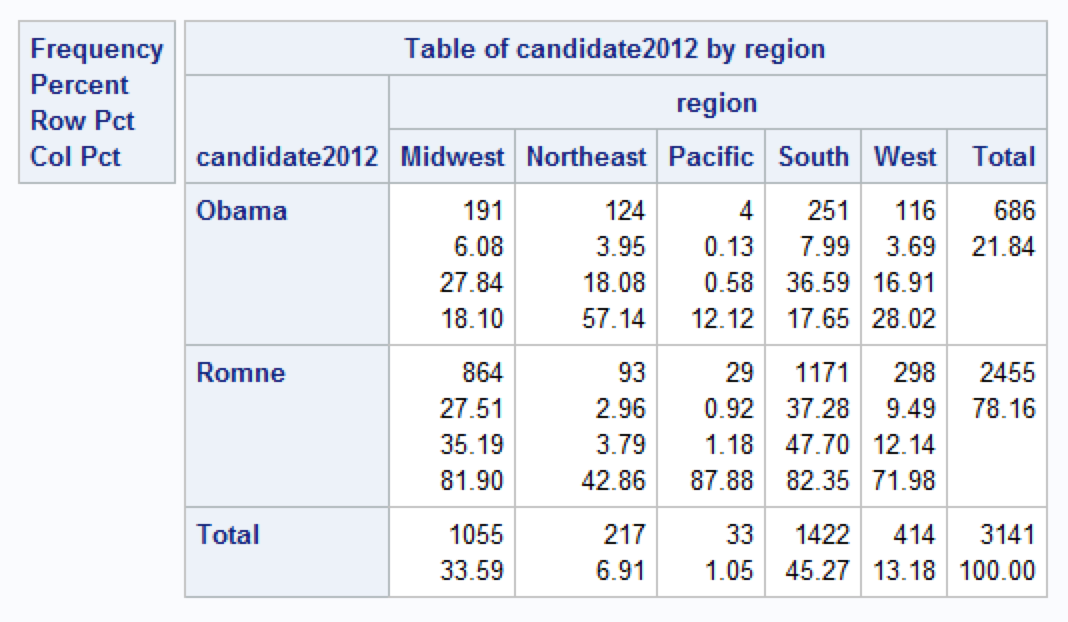


This is a histogram of every county showing number of points that winning candidate won by. Negative points are republican and positive are democrats. This shows that the counties are skewed towards the republican side despite the democrats winning total votes.

Region Per Point Difference Box and Whisker

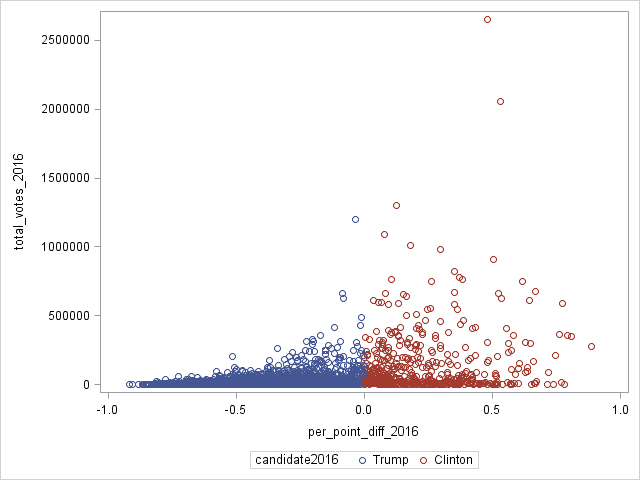
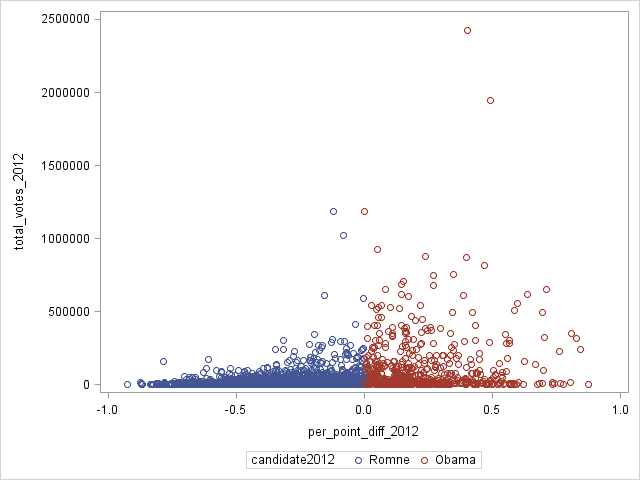


This chart shows negative for republican and positive for democrats. This shows that Trump won the majority of counties. It also shows that Obama won the majority of counties in 2012. Hillary had more votes, but she focussed too heavily on large urban counties. She lacked support in rural areas.This shows the importance of winning overall counties.



These charts show how many counties each candidate won in each region like the previous Box and Whisker plot. It shows data as a count to better understand the implications.

Total Votes Per Point Difference Scatter Plot



These charts show the total number of votes in each county against the percentage the candidate won by. Blue is republican and red is democrat. This shows the larger the county was, the larger the percentage the democratic candidate won by. This further shows that the democratic party needs to focus on smaller counties.

Summary

Based on our data, we can make recommendations for the 2020 presidential campaign. The above data and charts show the importance of counties. Clinton lost the 2016 election because she did not focus on the overall number of counties even though she had a majority vote. This is because Clinton pulled in a large number of votes in urban counties, but lost rural counties. We recommend that the 2020 candidates focus on both rural and urban counties. The goal is to even out the points won by for each county.

Works Cited

1 Tonmcg. "County\_Level\_Election\_Results\_12-16." GitHub. N.p., 20 Nov. 2016. Web. 8 Dec. 2016. <https://github.com/tonmcg/County\_Level\_Election\_Results\_12-16/blob/master/US\_County\_Level\_Presidential\_Results\_08-16.csv>.

2 U.S. Census Bureau. "Census Regions and Divisions of the United States." United States Census Bureau. N.p., n.d. Web. 8 Dec. 2016. <https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us\_regdiv.pdf>.

3 CNN. "Presidential Results." CNN.com. N.p., n.d. Web. 11 Dec. 2016. <http://www.cnn.com/election/results/president>.