**Voter Analysis Assignment #2**

This assignment requires these tasks:

**1. Carry out the analysis that you suggested in Voter Analysis Assignment #1, or modify**

**that analysis strategy based on class discussions.**

See attached table from submission named: **Group 8 Voter Assignment 2.xlsx**

**2. Did you find that you were missing other data that was not mentioned in your original**

**plan? What was the needed additional data and why didn’t you foresee the need for the**

**data? Find the data on the Internet and list the URL(s) and where you found the data and**

**describe the data sources and the relevant fields contained in each source.**

* All required data could be found in the files provided by Dr. Bradley
* However, the ‘**OBAMA** **V** **ROMNEY** **ONLY**’ spreadsheet was missing data for the states Colorado, Connecticut, Florida, Georgia, South Carolina, Utah and Wyoming, which was added from the ‘**FULL DATA**’ spreadsheet.
* Data was combined from the following spreadsheets from the data Dr. Bradley provided:
  + **OBAMA V ROMNEY ONLY**
  + **FULL DATA**
  + **PILF DATA**
  + **STATE ABBREVIATIONS**
* We used the following fields, which were sourced from the above files:
  + All Fields were used from **OBAMA V ROMNEY ONLY** (*State Postal, County Name, FIPS, Obama vote, %, Romney vote, %)*
  + We created the fields *State and County*, and *Total Voters* using VLOOKUP to pull data from the **FULL DATA** spreadsheet
  + We used INDEX and MATCH to create a *Registration Rate* field in the main table using the *State and County* fields as an index to pull the registration rates over from the **PILF DATA** table

**3. Did you run into other problems with your analysis? If yes, what were the problems?**

* Most of the problems we encountered involved issues with properly executing certain formulas in excel, namely VLOOKUP, INDEX and MATCH.
* When creating the *State and County* field, using the FIND function we found random spaces that caused problems with VLOOKUP, INDEX, and MATCH.
* One county, DeBaca, New Mexico, was spelled improperly on the **PILF DATA** spreadsheet (it included a space) and had to be adjusted to allow for proper analysis.
* These issues took extra time to troubleshoot and while copy and paste would have worked in a small dataset like this, VLOOKUP and similar techniques are crucial to master for larger datasets.
* We encountered a few formatting issues for select values once VLOOKUP formulas were properly functioning, and were able to fix them to retrieve the proper values.