

This README file explains the data sets and codes from the exercise of scraping the website [‘The Political Graveyard’](#).

The data on the website does not have tags. So, I have tried to use the pattern in which they have been rendered to impose a structure. This inevitably leads to some errors which need to be suitably cleaned before any analysis.

- `scrape_updated.ipynb`: Scrapes the following excel spreadsheet
- `data_born_pol.xlsx`:
 - `state`: state of birth
 - `county`: county of birth
 - `pol_name`: name of politician

Does not contain data on politicians from Wyoming

- `scrape_office.ipynb`: Scrapes the following excel spreadsheet
- `data_office.xlsx`:
 - `state`: state in which the position was held
 - `name`: name of politician
 - `office`: office held
 - `years`: years in which office was held by that politician
 - `city`: for Mayors and Postmasters

Some observations do not have end date for ‘years’

Some observations have the text ‘*as of*’ in ‘years’ – without any start/end date for the office held by the politician

- `scrape_new_clergy.ipynb`: Scrapes the following excel spreadsheet
- `data_state_county_religion_final.xlsx`:
 - `lived_state`: state of residence
 - `born_state`: state of birth
 - `born_county`: county of birth
 - `religion`: religious denomination for the politicians
 - `pol_name`: name of politician

For politicians who were born outside the US, ‘born_state’ is empty and their country of birth is entered into ‘born_county’

- name_match: matches the politician names from data_born_pol.xlsx and data_state_county_religion_final.xlsx and generates the following excel spreadsheets:
- data_name_match.xlsx:
 - residence state: state of residence
 - birth state: state of birth
 - birth_county: county of birth
 - pol_name_religion: name of politician in the file data_state_county_religion_final.xlsx
 - pol_name_county: name of politician in the file data_born_pol.xlsx
 - religion: religion of the politician
- data_name_no_match.xlsx:
 - residence state: state of residence
 - birth state: state of birth
 - birth_county: county of birth
 - pol_name_religion: name of politician in the file data_state_county_religion_final.xlsx
 - religion: religion of the politician

Author: Nikhil Kumar (nikhilklath@gmail.com)

Created on: 1/11/2021