



Virginia Voter Voices

An Analysis of Voter Data in Virginia

Team NULL:

André Zazzera (alz9cb), Annie Williams (maw3as), Hannah Frederick (hbf3k),
Sean Grace (smg2mx)



★ VIRGINIA ★
DEPARTMENT *of* ELECTIONS

Introduction

- **Changing Ideologies**
 - “Turning Virginia Blue” narrative
 - Demographic shifts
 - 2019 Elections
- **Motor Voter Act**
- **Sociopolitical Context**
 - Elections In Danger?
 - Registration & Turnout

Virginia Registration Data

- Scraped from the Virginia Department of Elections Registration Statistics for November General Elections (1976-2019)
- Data has all the acceptable properties: comes from a good source, has a large scale, is fairly clean, and has many types of data
- Variables in the Virginia Registration Dataset:
 - Year: the year of the election
 - Total Registered: the number of registered voters in that election
 - Percentage Change From Previous Year: the percent change in registered voters
 - Total Voting: the number of total voters that voted in that election
 - Voting Absentee (Included in Total Voting): the number of absentee voters
 - Turnout (% Voting of Total Registered): the percent of registered voters that voted

Virginia Elections Data

- Scraped from the Historical Elections Database for Presidential and Gubernatorial General Elections (1976-2017)
- Data has all the acceptable properties: comes from a good source, has a large scale, is fairly clean, and has many types of data
- Variables in the Virginia Elections Dataset:
 - Year: the year of the election
 - Office: the office that the candidate was running for
 - Winner (VA): the candidate's name
 - Party: the candidate's party
 - Winner's Vote Count: the number of all votes cast for them
 - Winner's Percent of Total Votes: the percent of all votes cast for the candidate

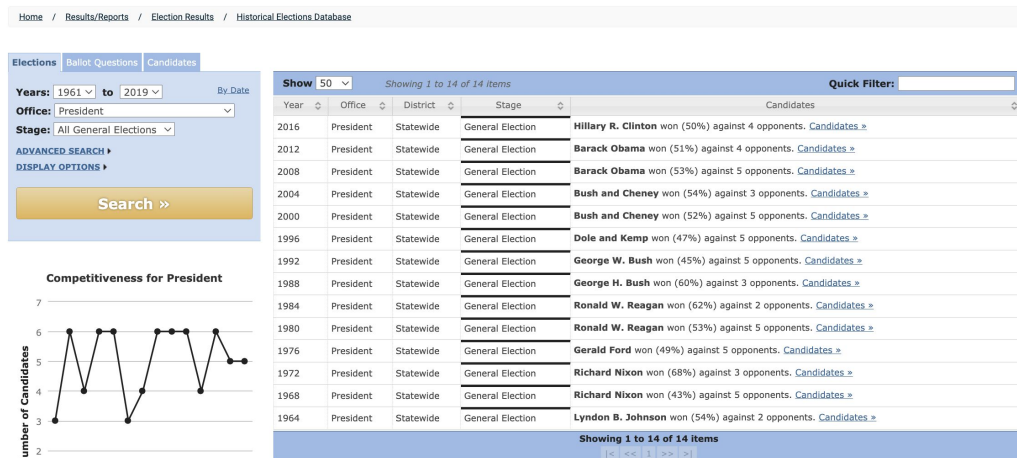
Experimental Design

1. Find the data
2. Scrape/clean the data
3. Merge/clean the data
4. Test throughout
5. Query the data
6. Graph the data
7. Interpret results



Web Scraping Process

- Used BeautifulSoup and python web request to scrape three separate pages on the Virginia Department of Elections Website



Results/Reports

Home / Results/Reports / Registration/Turnout Reports

Registration Statistics and Polling Places

Registration/Turnout Reports

Election Night Reporting

Election Results

Post-Election Reporting

Voter Registration List Maintenance

Election Security

Registration/Turnout Reports

Search Elections Go

SUMMARY OF VIRGINIA REGISTRATION & TURNOUT STATISTICS

The Department of Elections only publishes turnout statistics for November Elections and turnout figures are only published after the election is certified.

November General Elections: 1976 – Present

Year	Total Registered	Percentage Change from Previous Year	Total Voting	Turnout (% Voting of Total Registered)	Voting Absentee (Included in Total Voting)
2019	5,628,035	-1.0%	2,383,646	42.4%	144,360
2018	5,666,962	3.31%	3,374,382	59.5%	337,315
2017	5,489,530	-0.73%	2,612,309	47.6%	192,397
2016	5,529,742	6.41%	3,984,631	72.05%	566,948
2015	5,196,436	-1.60%	1,509,864	29.1%	62,605
2014	5,281,011	0.78%	2,194,346	41.6%	123,221
2013	5,240,286	-3.5%	2,253,418	43.0%	121,359
2012	5,428,833	6.1%	3,858,043	71.06%	447,907
2011	5,116,929	1.68%	1,463,761	28.61%	99,519
2010	5,032,144	1.54%	2,214,503	44.01%	106,888
2009	4,955,750	-1.57%	2,000,812	40.4%	88,182
2008	5,034,660	10.7%	3,723,260	74.0%	506,672

Data Cleaning Process

- Small amount of pre-processing in the web scrapers to remove extra symbols

Candidate	x	Vote Count	%
Ralph Shearer Northam ✓ Democratic		1,408,818	53.9%

- Pandas merge to condense data into a single csv

```
# specify outer join with how='outer'  
df_merge_col = pd.merge(voter_registration, combined_elections, on='Year', how='outer')  
df_merge_col.head()
```

Data Querying Process

- Calculated statistics for our complete dataset with all elections
- Calculated statistics for our dataset of elections with democratic winners vs elections with republican winners
- Calculated statistics for our dataset of presidential elections vs gubernatorial elections
- Calculated statistics for our dataset of elections pre-Motor Voter Act vs elections post-Motor Voter Act
- Calculated percent change in the number of total voters from one year to the next year for gubernatorial elections

Democratic
Statistics

	Year	Total Registered	Percentage Change from Previous Year	Total Voting	Turnout (% Voting of Total Registered)	Voting Absentee (Included in Total Voting)
count	10.00000	1.000000e+01	10.000000	1.000000e+01	10.000000	10.000000
mean	2002.70000	4.283457e+06	0.006480	2.497381e+06	0.583430	208131.300000
std	13.25854	1.306769e+06	0.053089	1.003899e+06	0.125009	213644.146323
min	1981.00000	2.214926e+06	-0.049000	1.377966e+06	0.430000	30584.000000
25%	1992.00000	3.080287e+06	-0.033500	1.842309e+06	0.467000	52802.250000
50%	2006.50000	4.743442e+06	-0.011150	2.126735e+06	0.589500	98670.500000
75%	2012.75000	5.381696e+06	0.048000	3.445522e+06	0.699200	384029.500000
max	2017.00000	5.529742e+06	0.107000	3.984631e+06	0.740000	566948.000000

Republican
Statistics

	Year	Total Registered	Percentage Change from Previous Year	Total Voting	Turnout (% Voting of Total Registered)	Voting Absentee (Included in Total Voting)
count	12.000000	1.200000e+01	11.000000	1.200000e+01	12.000000	11.000000
mean	1991.333333	3.206238e+06	0.060845	2.147801e+06	0.691233	105142.090909
std	10.611600	9.302905e+05	0.063335	5.233627e+05	0.136907	54331.709821
min	1976.000000	2.022619e+06	-0.048000	1.267208e+06	0.404000	29970.000000
25%	1983.000000	2.584026e+06	0.027150	1.804452e+06	0.623000	69357.500000
50%	1992.500000	3.015304e+06	0.073000	2.090664e+06	0.717450	108237.000000
75%	1997.750000	3.692684e+06	0.094000	2.452148e+06	0.809725	128864.500000
max	2009.000000	4.955750e+06	0.148000	3.198367e+06	0.837000	222059.000000

President
Statistics

	Year	Total Registered	Percentage Change from Previous Year	Total Voting	Turnout (% Voting of Total Registered)	Voting Absentee (Included in Total Voting)
count	11.000000	1.100000e+01	10.000000	1.100000e+01	11.000000	10.000000
mean	1996.000000	3.722627e+06	0.091810	2.771752e+06	0.757173	245846.300000
std	13.266499	1.249958e+06	0.028422	8.036520e+05	0.055125	186083.341671
min	1976.000000	2.123849e+06	0.061000	1.716182e+06	0.672000	84811.000000
25%	1986.000000	2.776392e+06	0.070250	2.206196e+06	0.715550	114416.000000
50%	1996.000000	3.322740e+06	0.088500	2.558665e+06	0.740000	145768.500000
75%	2006.000000	4.776320e+06	0.103750	3.460814e+06	0.811450	391445.000000
max	2016.000000	5.529742e+06	0.148000	3.984631e+06	0.837000	566948.000000

Governor
Statistics

	Year	Total Registered	Percentage Change from Previous Year	Total Voting	Turnout (% Voting of Total Registered)	Voting Absentee (Included in Total Voting)
count	11.000000	1.100000e+01	11.000000	1.100000e+01	11.000000	11.000000
mean	1997.000000	3.669139e+06	-0.016727	1.841650e+06	0.527291	70855.727273
std	13.266499	1.248136e+06	0.034619	3.915778e+05	0.094434	48816.567991
min	1977.000000	2.022619e+06	-0.049000	1.267208e+06	0.404000	29970.000000
25%	1987.000000	2.667622e+06	-0.038000	1.600929e+06	0.456600	40257.000000
50%	1997.000000	3.565697e+06	-0.026000	1.821242e+06	0.495000	53904.000000
75%	2007.000000	4.703988e+06	-0.011150	2.000432e+06	0.619000	82082.000000
max	2017.000000	5.489530e+06	0.073000	2.612309e+06	0.665000	192397.000000

Pre-Motor
Voter Act

	Year	Total Registered	Percentage Change from Previous Year	Total Voting	Turnout (% Voting of Total Registered)	Voting Absentee (Included in Total Voting)
count	20.00000	2.000000e+01	19.000000	2.000000e+01	20.000000	16.000000
mean	1985.50000	2.553199e+06	0.020474	1.610476e+06	0.630295	57691.500000
std	5.91608	3.588868e+05	0.056057	4.153958e+05	0.128214	36921.335496
min	1976.00000	2.022619e+06	-0.049000	1.059158e+06	0.427000	22589.000000
25%	1980.75000	2.229240e+06	-0.013500	1.263649e+06	0.520625	30430.500000
50%	1985.50000	2.633555e+06	0.009000	1.512869e+06	0.622250	42145.000000
75%	1990.25000	2.813096e+06	0.032000	1.836344e+06	0.713750	82429.750000
max	1995.00000	3.055486e+06	0.148000	2.558665e+06	0.837000	141123.000000

Post-Motor
Voter Act

	Year	Total Registered	Percentage Change from Previous Year	Total Voting	Turnout (% Voting of Total Registered)	Voting Absentee (Included in Total Voting)
count	24.000000	2.400000e+01	24.000000	2.400000e+01	24.000000	24.000000
mean	2007.500000	4.696537e+06	0.026633	2.275088e+06	0.483000	153832.750000
std	7.071068	6.993378e+05	0.038469	8.426050e+05	0.156103	155233.342278
min	1996.000000	3.322740e+06	-0.035000	1.229139e+06	0.286100	10686.000000
25%	2001.750000	4.190202e+06	-0.002575	1.498338e+06	0.353250	51150.750000
50%	2007.500000	4.755216e+06	0.019900	2.204424e+06	0.444650	111747.000000
75%	2013.250000	5.250467e+06	0.061775	2.644094e+06	0.614250	160909.750000
max	2019.000000	5.666962e+06	0.107000	3.984631e+06	0.740000	566948.000000

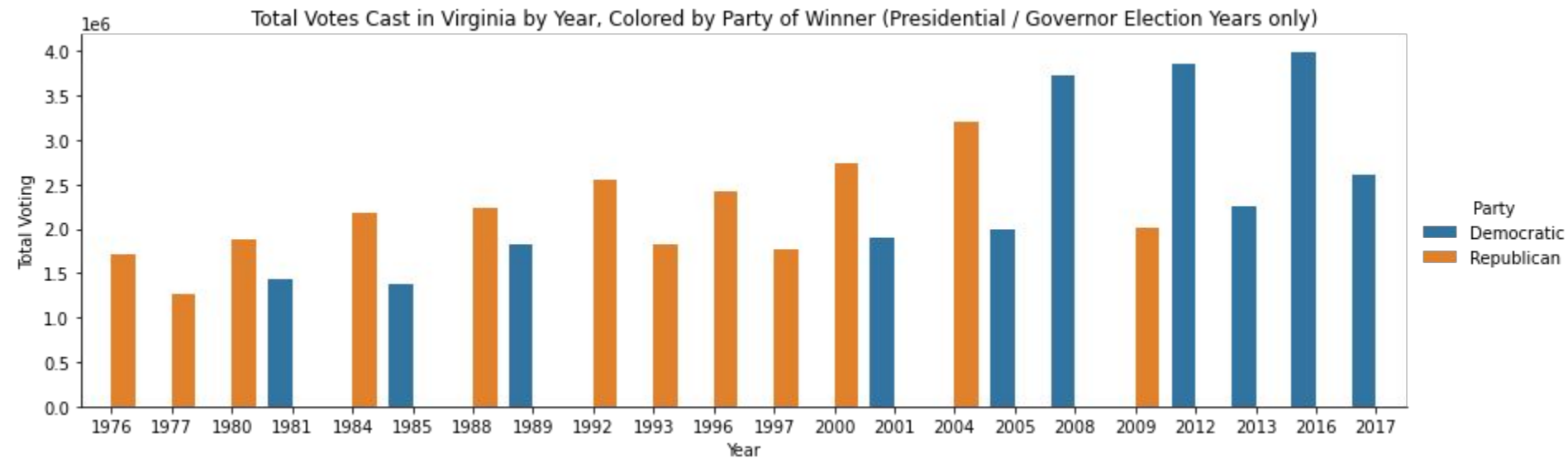
Other Queries

- For number of voters in only gubernatorial elections:
 - smallest percent change was in 2009
 - greatest percent decrease was in 1985
 - greatest percent change was in 1989
- 1989 election might have seen such a dramatic increase in the number of voters compared to the previous election because of Douglas Wilder
- Douglas Wilder was not only the first black governor in Virginia, but the first black person elected governor in the history of the U.S.

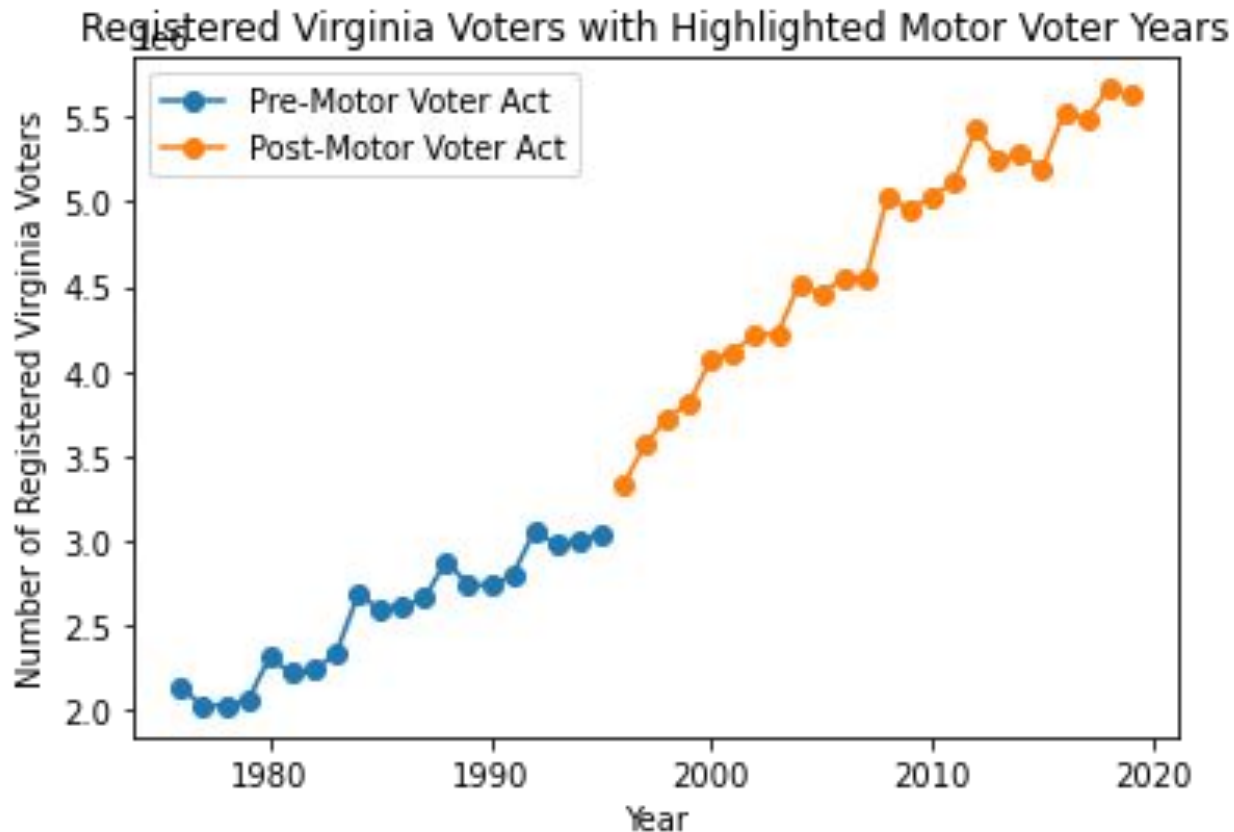
Data Visualization Process

- Visualize how party winners changed over time
- Observe the changes in voter registration pre- and post- Motor Voter Act
- Visualize how voter turnout changed over time
- Compare different election types
- How has absentee voting changed over time?

Graph Analysis

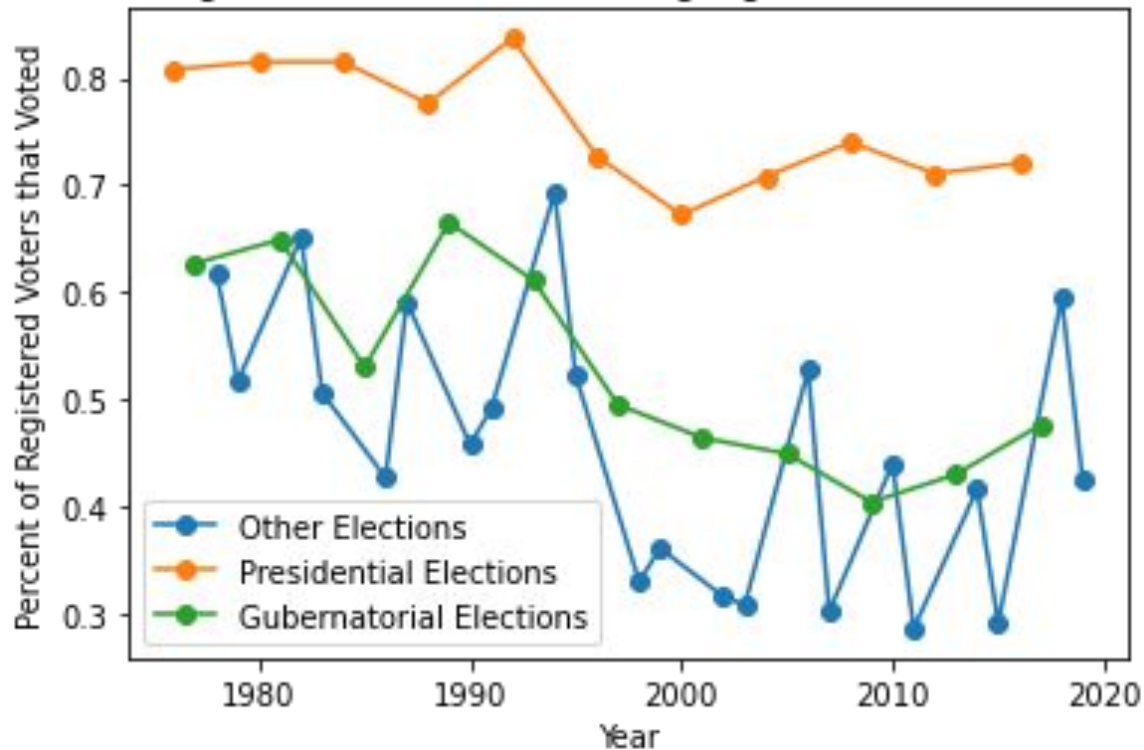


Graph Analysis

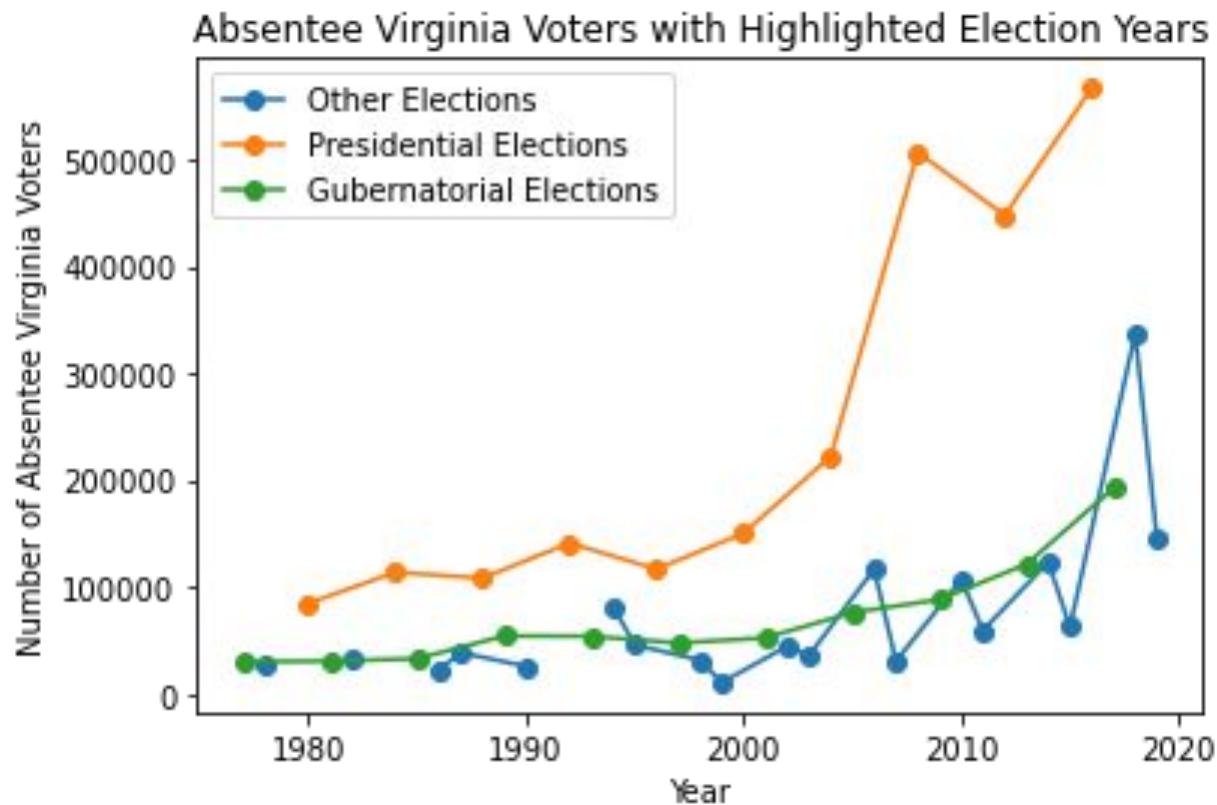


Graph Analysis

Virginia Voter Turnout with Highlighted Election Years



Graph Analysis



Testing Process

- Ran 9 unit tests on our 3 web scrapers in TestWebScraper.py:
 - The voter_registration.py web scraper converted the numbers from strings to floats as they were read in, so we tested and passed that criteria

■ Ex:

4,549,864

- The voter_registration.py web scraper converted the percents from strings to floats as they were read in, so we tested and passed that criteria

■ Ex:

- 0.1%

-4.9%**

Testing Process

- Each web scraper makes a request to the Virginia Department of Elections website, so we tested that each request succeeded with a 200 status code

```
$ python3 TestWebScraper.py
Is 1,000,000*** converted to and returned as 1000000.0?
.Is 1,000,000 converted to and returned as 1000000.0?
.Is -10.0%*** converted to and returned as -0.1?
.Is -10.0% converted to and returned as -0.1?
.Is 10.0%*** converted to and returned as 0.1?
.Is 10.0% converted to and returned as 0.1?
.Gubernatorial election site status code: 200
.Presidential election site status code: 200
.Voter registration site status code: 200
.
-----
Ran 9 tests in 3.631s

OK
```

Beyond Original Specifications

- **Significance between presidential and nonpresidential election years' turnout**

```
pres_years_data = voting_data.loc[voting_data['Year']%4==0]
nonpres_years = voting_data.loc[voting_data['Year']%4!=0]
# 2 tailed t-test to see if this is a significant difference in means
print(stats.ttest_ind(pres_years_data['Turnout (% Voting of Total Registered)'],
                      nonpres_years['Turnout (% Voting of Total Registered)'], equal_var=False))
```

```
Ttest_indResult(statistic=10.429998517323842, pvalue=1.4126047812803902e-12)
```

Beyond Original Specifications

- **Significance between pre- and post-Motor Voter Act total registration**

```
premotor = voting_data.loc[voting_data['Year'] < 1996]
postmotor = voting_data.loc[voting_data['Year'] >= 1996]
# 2 tailed t-test to see if this is a significant difference in means
print(stats.ttest_ind(premotor['Total Registered'],
                      postmotor['Total Registered'], equal_var=False))
```

```
Ttest_indResult(statistic=-13.088118538685828, pvalue=3.715525965573648e-15)
```

Beyond Original Specifications

```
# Did VA elect the winning candidate in any given year?
## establish dict of presidential election years and winners
pres_winners = {'1976': 'Jimmy Carter', '1980': 'Ronald W. Reagan', '1984': 'Ronald W. Reagan',
                '1988': 'George H. Bush', '1992': 'Bill Clinton', '1996': 'Bill Clinton',
                '2000': 'Bush and Cheney', '2004': 'Bush and Cheney', '2008': 'Barack Obama',
                '2012': 'Barack Obama', '2016': 'Donald Trump'}

## define function to:
## take an election year
## check data against pres_winners dict
## return a boolean
def va_elect(year):
    return pres_winners[year] == pres_years_data.loc[pres_years_data.Year == int(year), 'Winner (VA)'].values[0]

# user inputs a presidential election year to see in virginia results aligned with the nation's results
check_year = input('Please enter a presidential election year: ')
try:
    print('Did Virginia results align with the nation in this election? ' + str(va_elect(check_year)))
except KeyError:
    print('You did not enter a presidential election year.')
```

Please enter a presidential election year: 1996

Did Virginia results align with the nation in this election? False

Conclusions

- **Significant effect of National Voter Registration Act**
- **Turnout trends**
- **Effect of turnout on winning party**
- **Absentee voting**

Extensions

- **Expand surveyed region (South, USA)**
- **Focus on county-level data**
- **Predictive modeling**
- **Additional user interactivity**

Data Sources

- <https://www.elections.virginia.gov/resultsreports/>
- <https://historical.elections.virginia.gov/>
- https://github.com/mawilliams098/team_null