OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020

Theory

- Our theory:
  - A Single demographic category is an effective predictor of which political party wins
    - Education
    - Median Income
    - Race
    - Median Home Value
    - Employment
    - Age

Describe data sets used

- Data sets used
- 2016 Indiana election results by county Harvard Dataverse as maintained by the MIT Election Data and Science Lab
  - All 92 counties
  - 2016 Presidential Election between Hillary Clinton and Donald Trump
- 2016 annual American Community Survey (ACS) conducted by the U.S. Census
  - Used API
  - There were approximately 20,000 variables available to select
  - Data was available by country, state, county, and other geographic categories
  - We selected six
  - Methodology

Data Acquisition and Cleanup

```
# Census API Key
from config import census_api_key
c = Census(census_api_key, year=2016)
import pandas as pd
```

In [3]:
```
census_data = c.acs5.get(("NAME", "B01003_001E", "B02001_002E","B02001_003E","B02001_005E","B03001_003E",
                          "B25077_001E","B15003_002E","B15003_017E","B15003_018E",
                         "B15003_021E","B15003_022E","B15003_023E","B15003_024E","B15003_025E"), {'for': 'county:*'})
```

In [4]:
```
census_complete=pd.DataFrame(census_data)
census_complete=census_complete.rename(columns={"B01003_001E":"Total Population",
                                                "B02001_002E":"Population (White)",
                                                "B02001_003E":"Population (African-American)",
                                               "B02001_005E":"Population (Asian)",
                                               "B03001_003E":"Population (Hispanic)",
                                               "B25077_001E":"Median Home Value",
                                               "B15003_002E":"Education (None)",
                                               "B15003_017E":"Education (High School)",
                                               "B15003_018E":"Education (GED)",
                                               "B15003_021E":"Education (Associates)",
                                               "B15003_022E":"Education (Bachelors)",
                                               "B15003_023E":"Education (Masters)",
                                               "B15003_024E":"Education (Professional)",
                                               "B15003_025E":"Education (Doctorate)"})
census_complete.head()
#census_county = census_complete[1].str.split(' ').apply(Series, 1)
```

Out[4]:

	NAME	Total Population	Population (White)	Population (African-American)	Population (Asian)	Population (Hispanic)	Median Home Value	Education (None)	Education (High School)	Education (GED)	Education (Associates)	Education (Bachelors)	Education (Masters)	E (Profe
0	Carroll County, Arkansas	27690.0	25856.0	318.0	245.0	4021.0	118500.0	128.0	5458.0	1346.0	1162.0	2157.0	951.0	
1	Chicot County, Arkansas	11189.0	4778.0	6070.0	46.0	578.0	59600.0	96.0	2621.0	627.0	312.0	718.0	220.0	
	Clark													

Data Acquisition and Cleanup

Launcher | homevalue_plot3.ipynb | merged_file2.ipynb

Code

```
[252]: # import in and read indiana_census.csv and 2016_US_County_Level_Presidential_Results.csv files
import pandas as pd
import requests
import json
import csv

election_results_file = "in_clean_result2.csv"
indiana_census_file = "indiana_census.csv"
indiana_census2_file = "indiana_census2.csv"
indiana_education_file = "indiana_education2.csv"

election_results_df = pd.read_csv(election_results_file)
indiana_census_df = pd.read_csv(indiana_census_file)
indiana_census2_df = pd.read_csv(indiana_census2_file)
indiana_education_df = pd.read_csv(indiana_education_file)
```

```
[253]: election_results_df.head()
```

[253]:

	votes_dem	votes_gop	total_votes	per_dem	per_gop	per_dem2	per_gop2	state_abbr	county_name
0	2802	9642	13039	0.214894	0.739474	21.49	73.95	IN	Adams County
1	55222	83801	145787	0.378785	0.574818	37.88	57.48	IN	Allen County
2	9841	20637	32389	0.303838	0.637161	30.38	63.72	IN	Bartholomew County
3	860	2579	3654	0.235359	0.705802	23.54	70.58	IN	Benton County
4	1243	3349	4822	0.257777	0.694525	25.78	69.45	IN	Blackford County

```
[254]: # Clean up county_name category. Replace 'county_name' with 'County'
election_results_df.rename(columns={"county_name":"County"}, inplace=True)
election_results_df.columns
```

```
[254]: Index(['votes_dem', 'votes_gop', 'total_votes', 'per_dem', 'per_gop',
       'per_dem2', 'per_gop2', 'state_abbr', 'County'],
      dtype='object')
```

```
[255]: election_results_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 92 entries, 0 to 91
Data columns (total 9 columns):
votes_dem      92 non-null int64
votes_gop      92 non-null int64
total_votes    92 non-null int64
per_dem        92 non-null float64
per_gop        92 non-null float64
```

Launcher | homevalue_plot3.ipynb | merged_file2.ipynb

Code — Python 3

```
[302]: # Reorganizing the columns
merged_file = merged_file4[['State', 'State Abbr','County', 'DEM Votes', 'GOP Votes', 'Total Votes', 'DEM %', 'GOP %',
       'Total Population', 'Population (White)', 'Population (African-American)', 'Population (Asian)',
       'Population (Hispanic)', 'Median Age', 'Median Age (Male)', 'Median Age (Female)',
       'Education (None)', 'Education (High School)', 'Education (GED)', 'Education (Associates)',
       'Education (Bachelors)', 'Education (Masters)', 'Education (Professional)', 'Education (Doctorate)',
       'Median Income', 'Income Per Capita', 'Median Home Value',
       'Pop in Labor Force', 'Pop Not in Labor Force', 'Unemployed',
       'White Male (Bachelors or higher) %', 'White Female (Bachelors or higher) %', 'Af-Am Male (25 and over) %', 'Af-Am Female (Bachelors or higher) %']]
merged_file.head(2)
```

[302]:

	State	State Abbr	County	DEM Votes	GOP Votes	Total Votes	DEM %	GOP %	Total Population	Population (White)	...	Median Income	Income Per Capita	Median Home Value	Pop in Labor Force	Pop Not in Labor Force	Unemployed	White Male (Bachelors or higher) %	White Female (Bachelors or higher) %	Af-Am Male (25 and over) %	Af-Am Female (Bachelors or higher) %
0	Indiana	IN	Adams County	2802	9642	13039	21.49	73.95	34813.0	33743.0	...	47572.0	21173.0	114600.0	15975.0	8948.0	872.0	16.54	14.83	0.00	3.33
1	Indiana	IN	Allen County	55222	83801	145787	37.88	57.48	365565.0	290987.0	...	49574.0	26058.0	116400.0	186706.0	93216.0	13152.0	29.69	28.96	13.63	12.97

2 rows × 34 columns

```
[297]: merged_file.shape
```

```
[297]: (92, 34)
```

```
[ ]: merged_file.count()
```

```
[ ]: merged_file.columns
```

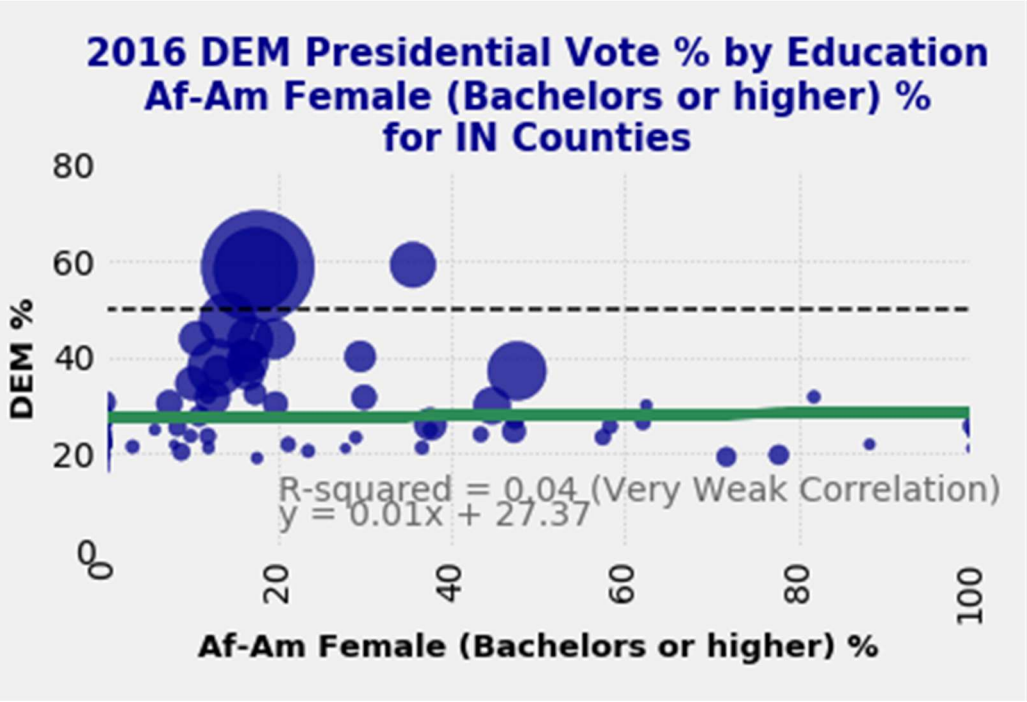
```
[303]: merged_file4.describe()
```

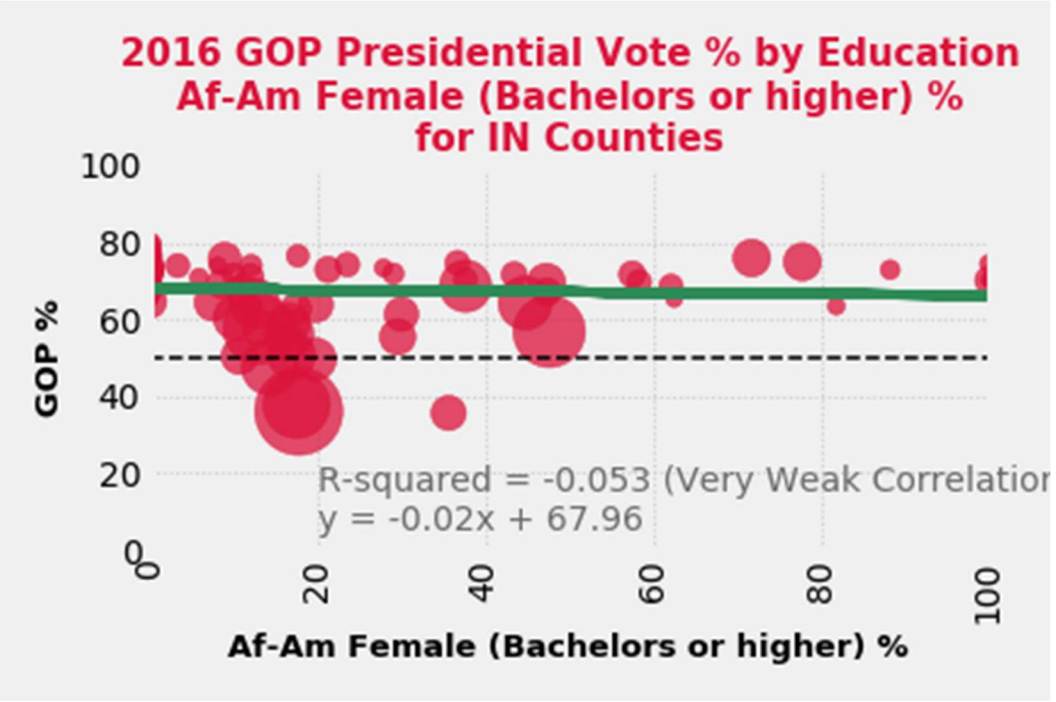
[303]:

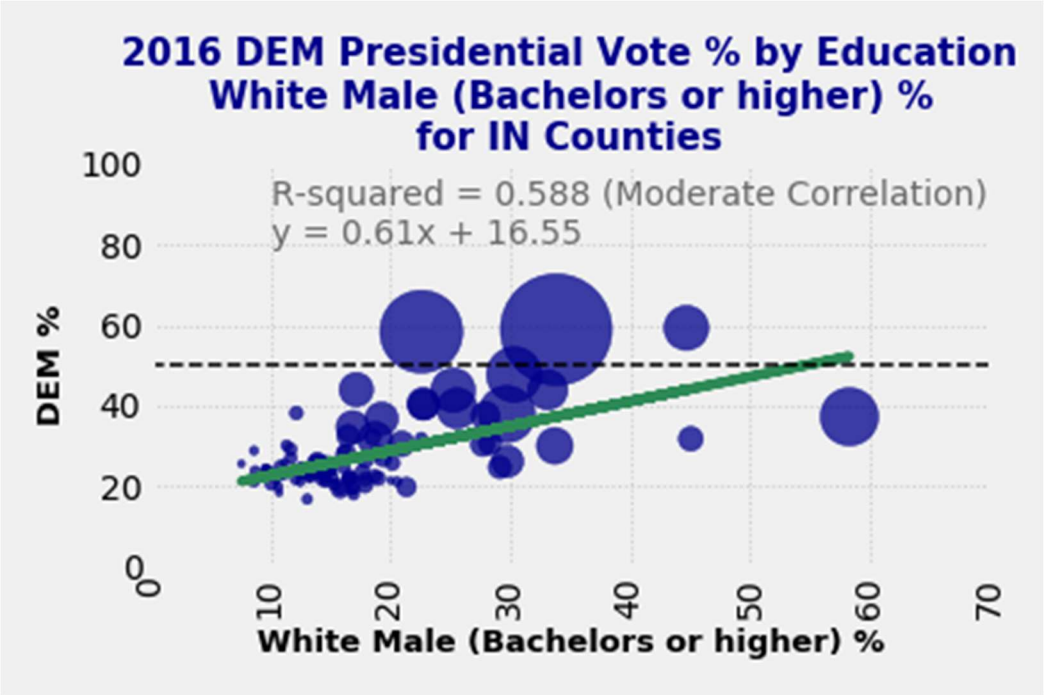
	DEM Votes	GOP Votes	Total Votes	per_dem	per_gop	DEM %	GOP %	Total Population	Population (White)	Population (African-American)	...	White Female (25 and over)	White Female (Bachelors or higher)	White Female (Bachelors or higher) %	Af-Am (25 and over)	Af-Am Male (25 and over)	Af-Am Male (Bachelors or higher)
count	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000	...	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000
mean	11216.880435	16915.434783	29587.271739	0.276277	0.676119	27.628261	67.611413	71625.847826	60160.423913	6611.152174	...	21090.847826	5348.989130	19.245326	3931.826087	1808.304348	267.565217
std	26630.651685	20249.244901	48057.304706	0.088005	0.089893	8.800259	8.988692	120038.509471	81181.583363	29600.642447	...	29107.845885	10118.831714	7.552899	17696.059450	7609.226506	1123.042123
min	686.000000	2118.000000	2917.000000	0.167660	0.356229	16.770000	35.620000	6003.000000	5811.000000	12.000000	...	2167.000000	274.000000	8.780000	8.000000	0.000000	0.000000

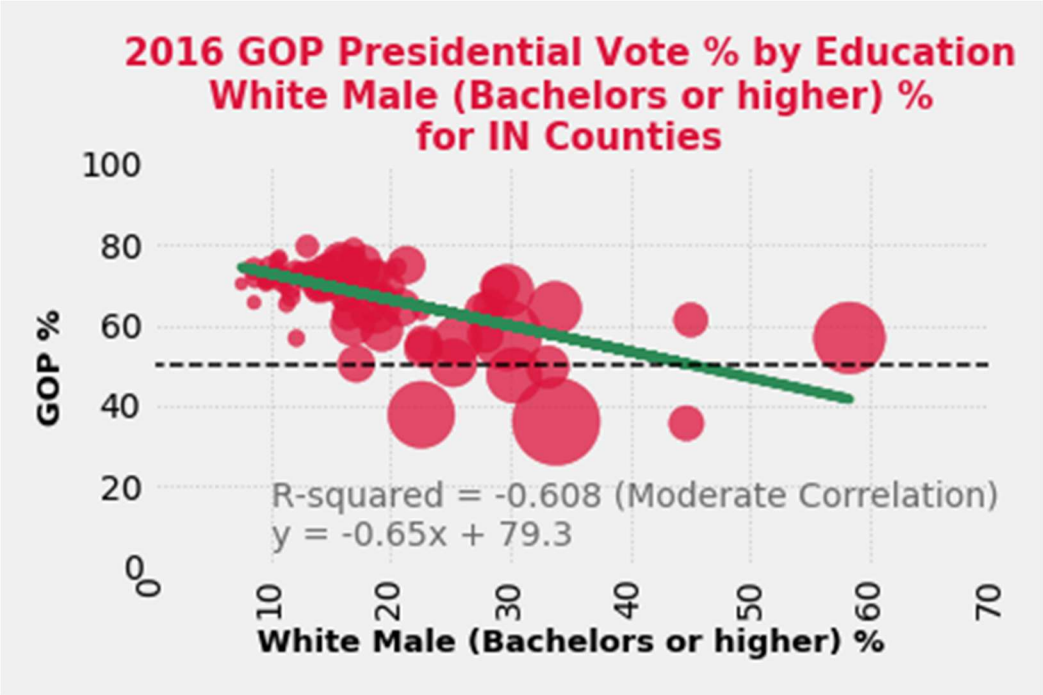
Questions to answer

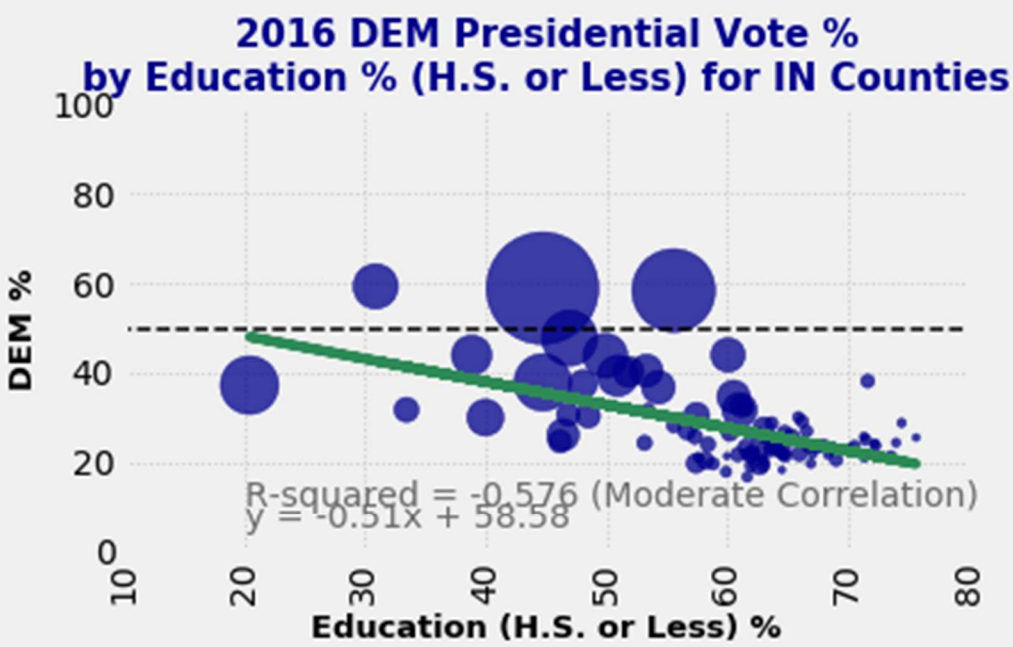
- Does median age/unemployment predict the DEM/GOP % vote in a county
- Does median home value/education predict the DEM/GOP % in a county
- Does race/median income predict the DEM/GOP % in a county

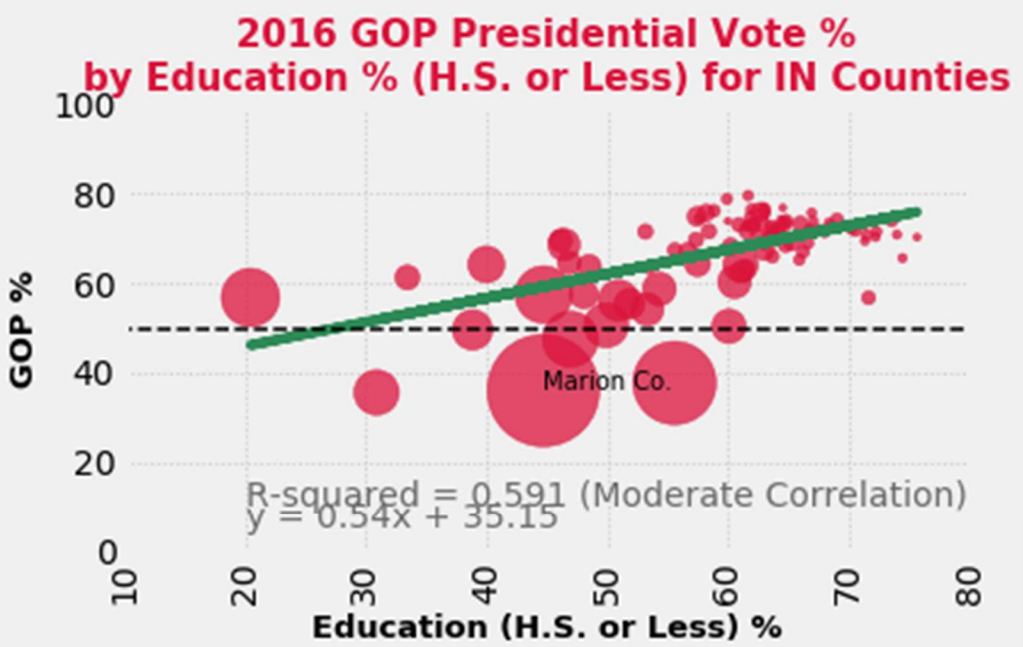

2016 DEM Presidential Vote % by Education
Af-Am Female (Bachelors or higher) %
for IN Counties
DEM %
0
20
40
60
80
20
40
60
80
100
R-squared = 0.04 (Very Weak Correlation)
y = 0.01x + 27.37
Af-Am Female (Bachelors or higher) %

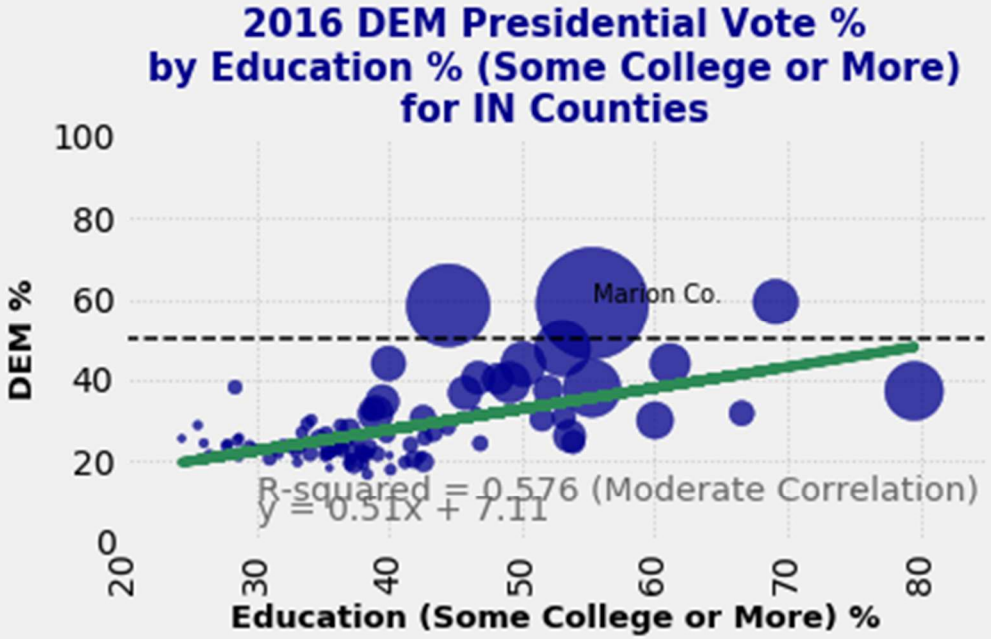

2016 GOP Presidential Vote % by Education
Af-Am Female (Bachelors or higher) %
for IN Counties
GOP %
0
20
40
60
80
100
20
40
60
80
100
R-squared = -0.053 (Very Weak Correlation
y = -0.02x + 67.96
Af-Am Female (Bachelors or higher) %

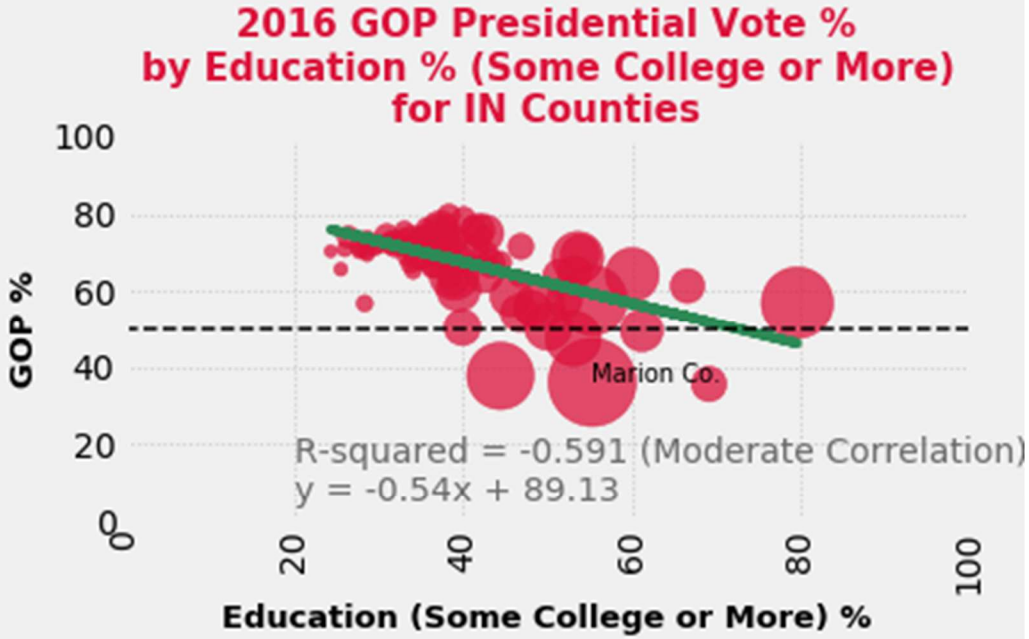

2016 DEM Presidential Vote % by Education
White Male (Bachelors or higher) %
for IN Counties
R-squared = 0.588 (Moderate Correlation)
y = 0.61x + 16.55
DEM %
White Male (Bachelors or higher) %

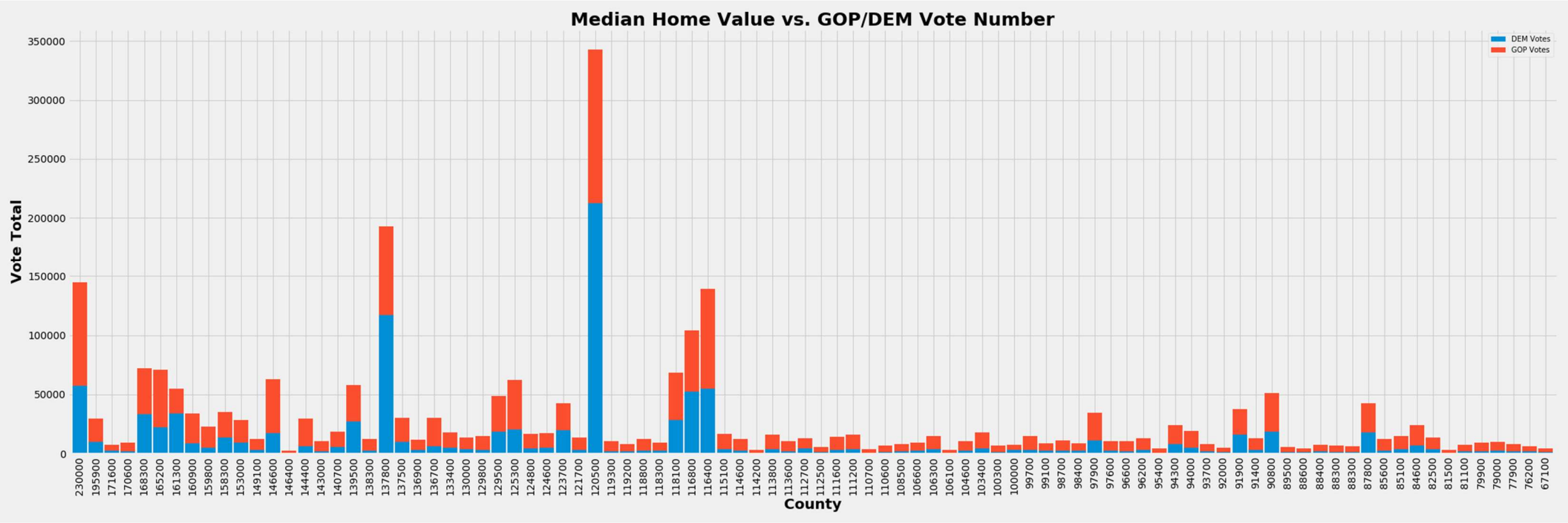

2016 GOP Presidential Vote % by Education
White Male (Bachelors or higher) %
for IN Counties
R-squared = -0.608 (Moderate Correlation)
y = -0.65x + 79.3
GOP %
White Male (Bachelors or higher) %

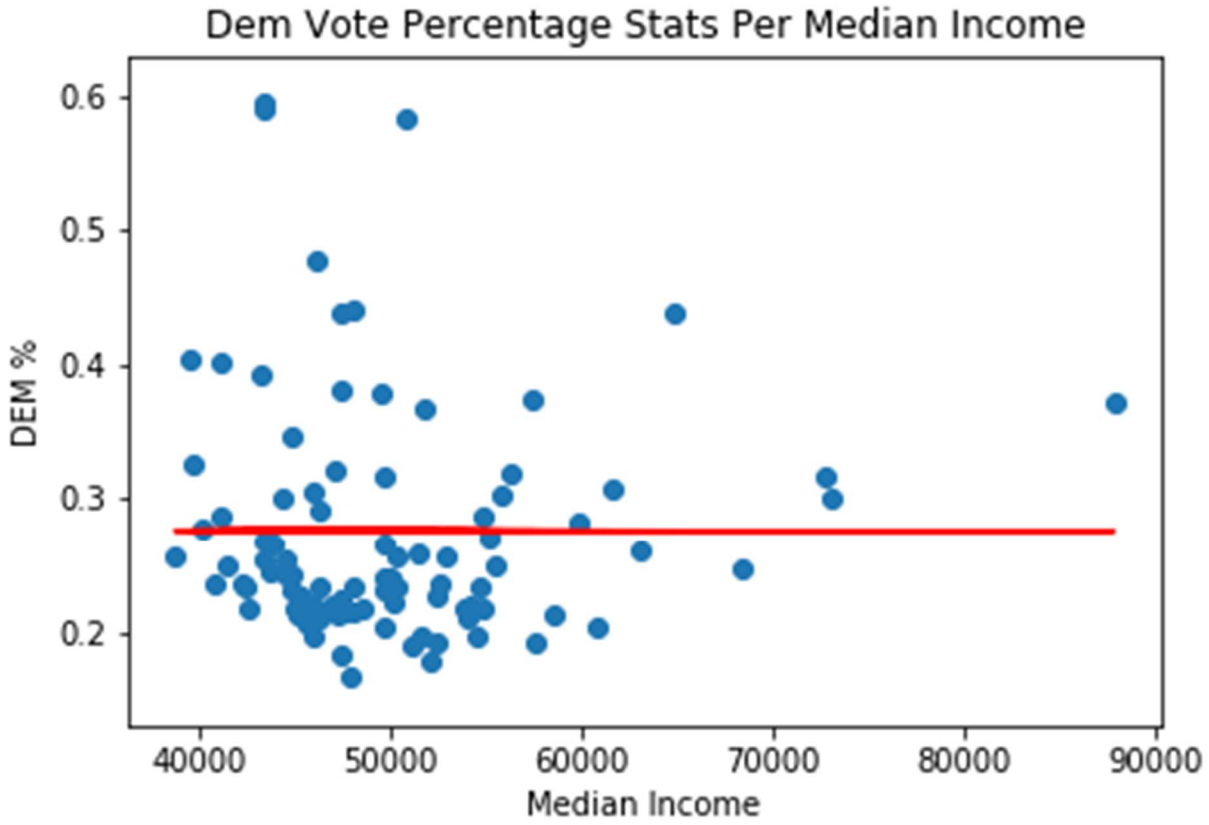

2016 DEM Presidential Vote %
by Education % (H.S. or Less) for IN Counties
DEM %
Education (H.S. or Less) %
R-squared = -0.576 (Moderate Correlation)
y = -0.51x + 58.58

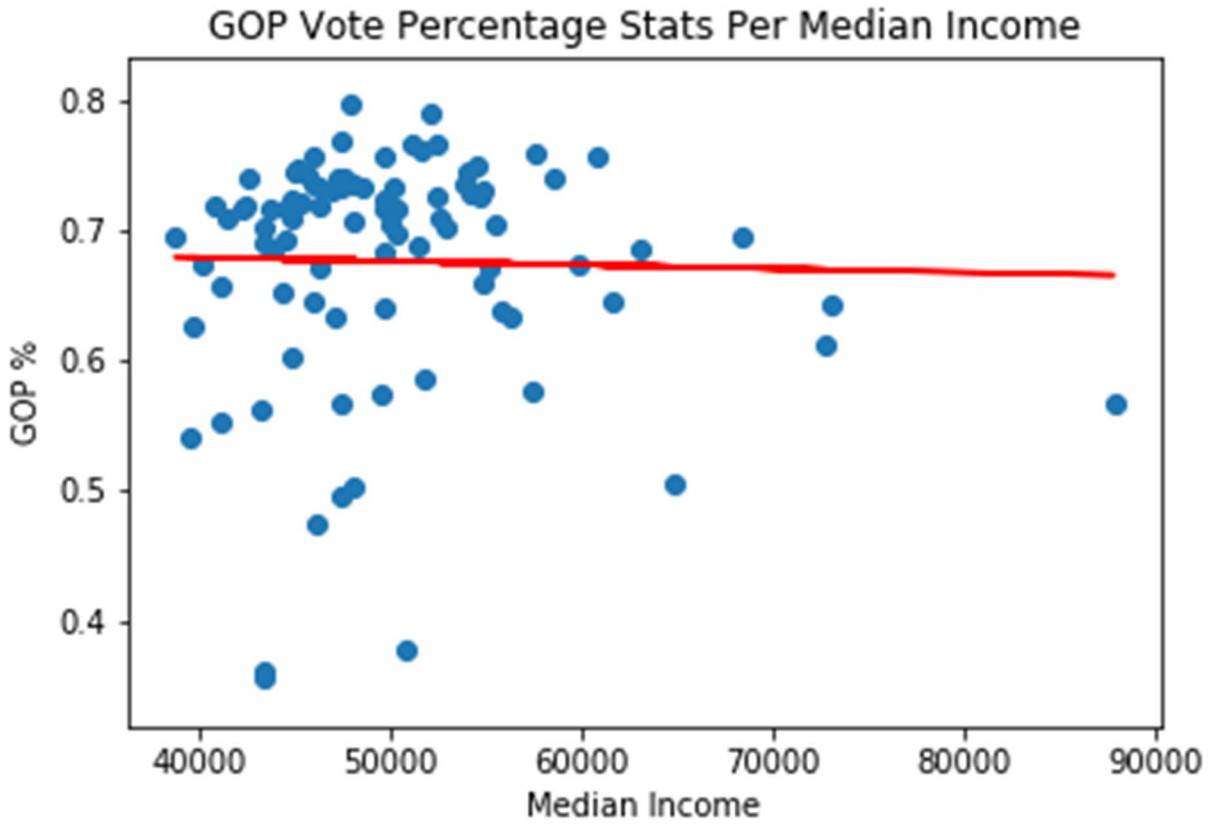

2016 GOP Presidential Vote %
by Education % (H.S. or Less) for IN Counties
GOP %
Education (H.S. or Less) %
Marion Co.
R-squared = 0.591 (Moderate Correlation)
y = 0.54x + 35.15

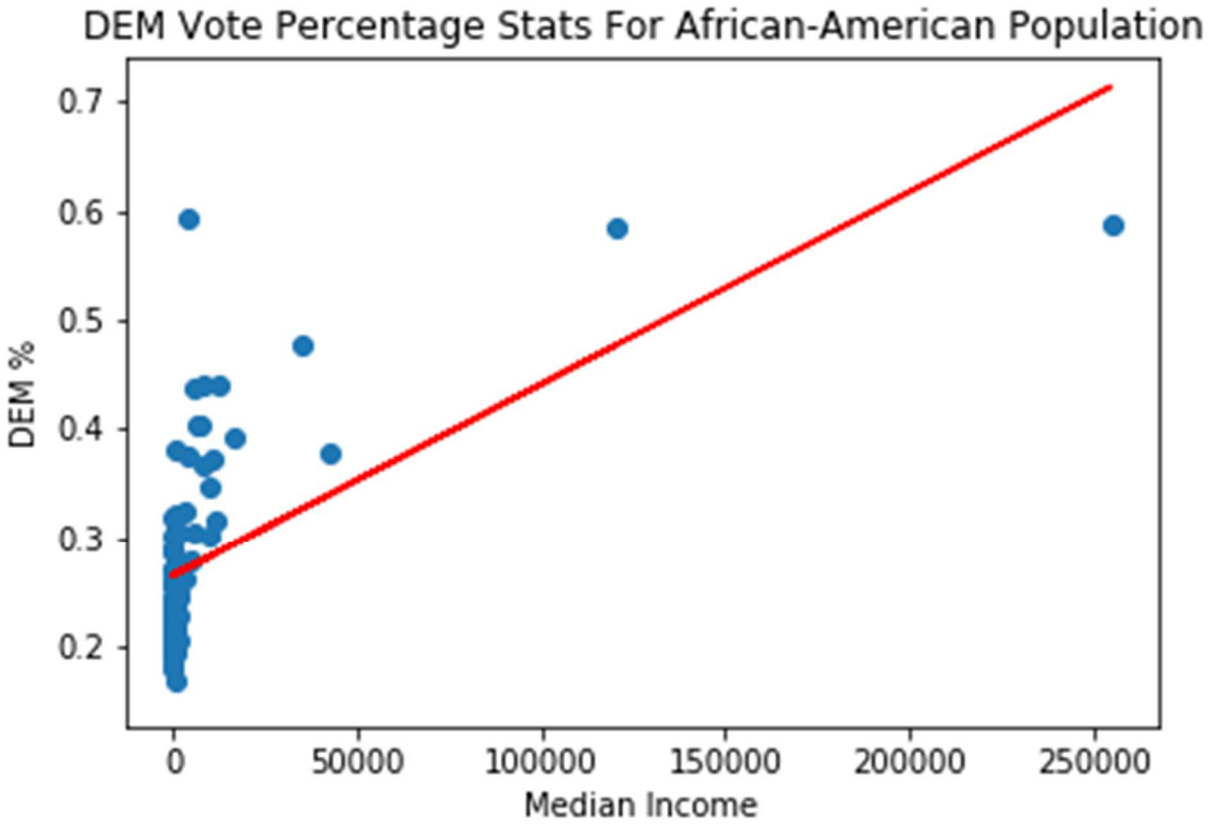

2016 DEM Presidential Vote %
by Education % (Some College or More)
for IN Counties
DEM %
100
80
60
40
20
0
Marion Co.
R-squared = 0.576 (Moderate Correlation)
y = 0.51x + 7.11
20
30
40
50
60
70
80
Education (Some College or More) %

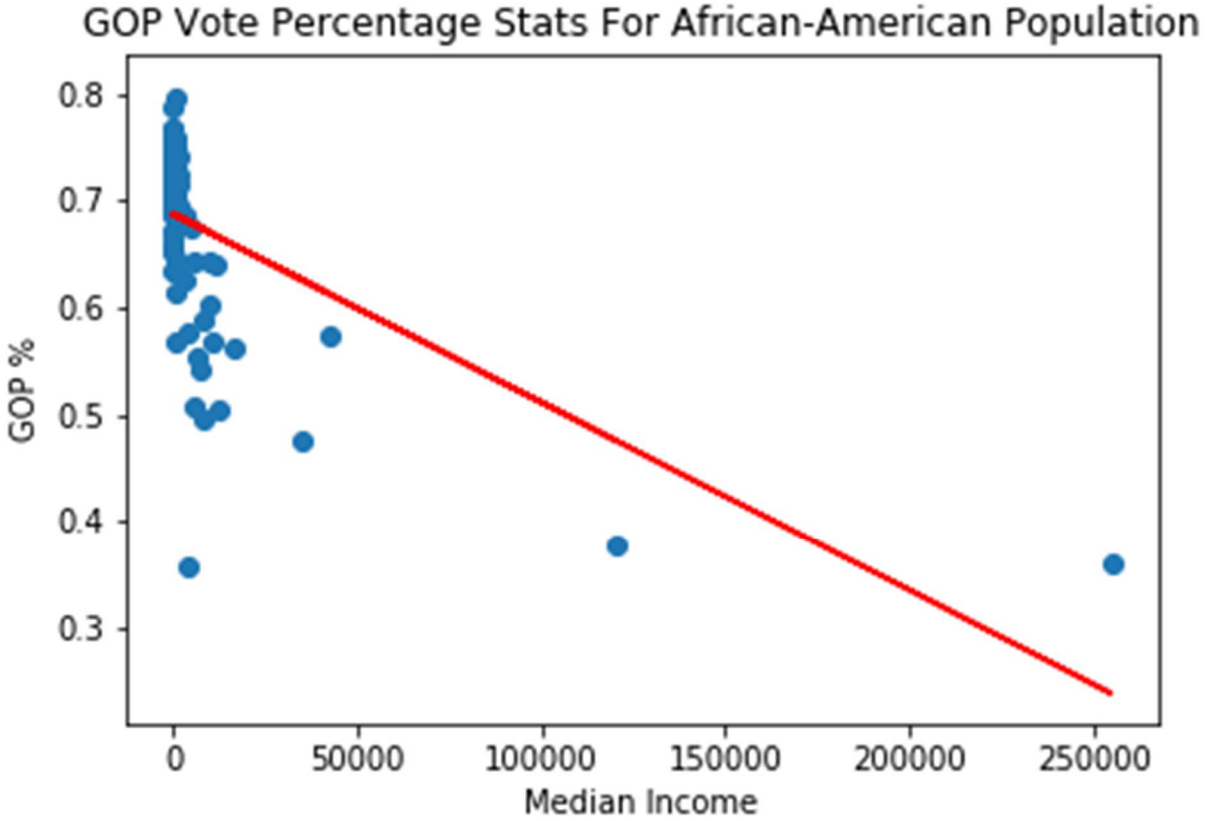

2016 GOP Presidential Vote %
by Education % (Some College or More)
for IN Counties
GOP %
100
80
60
40
20
0
Marion Co.
R-squared = -0.591 (Moderate Correlation)
y = -0.54x + 89.13
20
40
60
80
100
Education (Some College or More) %

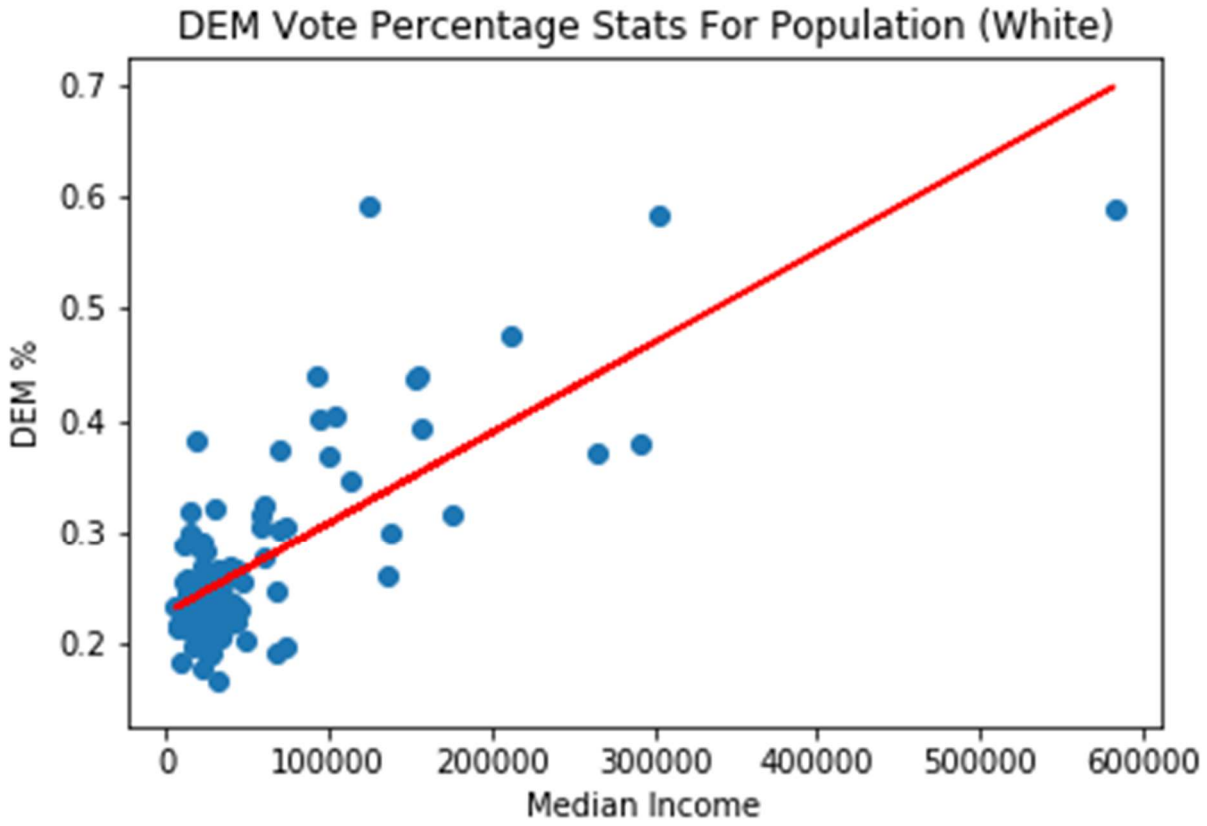

Median Home Value vs. GOP/DEM Vote Number
DEM Votes
GOP Votes
Vote Total
0
50000
100000
150000
200000
250000
300000
350000
County
230000 195900 171600 170600 168300 165200 161300 160900 159800 158300 153000 149100 146600 146400 144400 143000 140700 139500 138300 137800 137500 136900 136700 133400 130000 129800 129500 125300 124800 124600 123700 121700 120500 119300 119200 118800 118300 118100 116800 116400 115100 114600 114200 113800 113600 112700 112500 111600 111200 110700 110600 108500 106600 106300 106100 104600 103400 100300 100000 99700 99100 98700 98400 97900 97600 96600 96200 95400 94300 94000 93700 92000 91900 91400 90800 89500 88600 88400 88300 88300 87800 85600 85100 84600 82500 81500 81100 79900 79000 77900 76200 67100

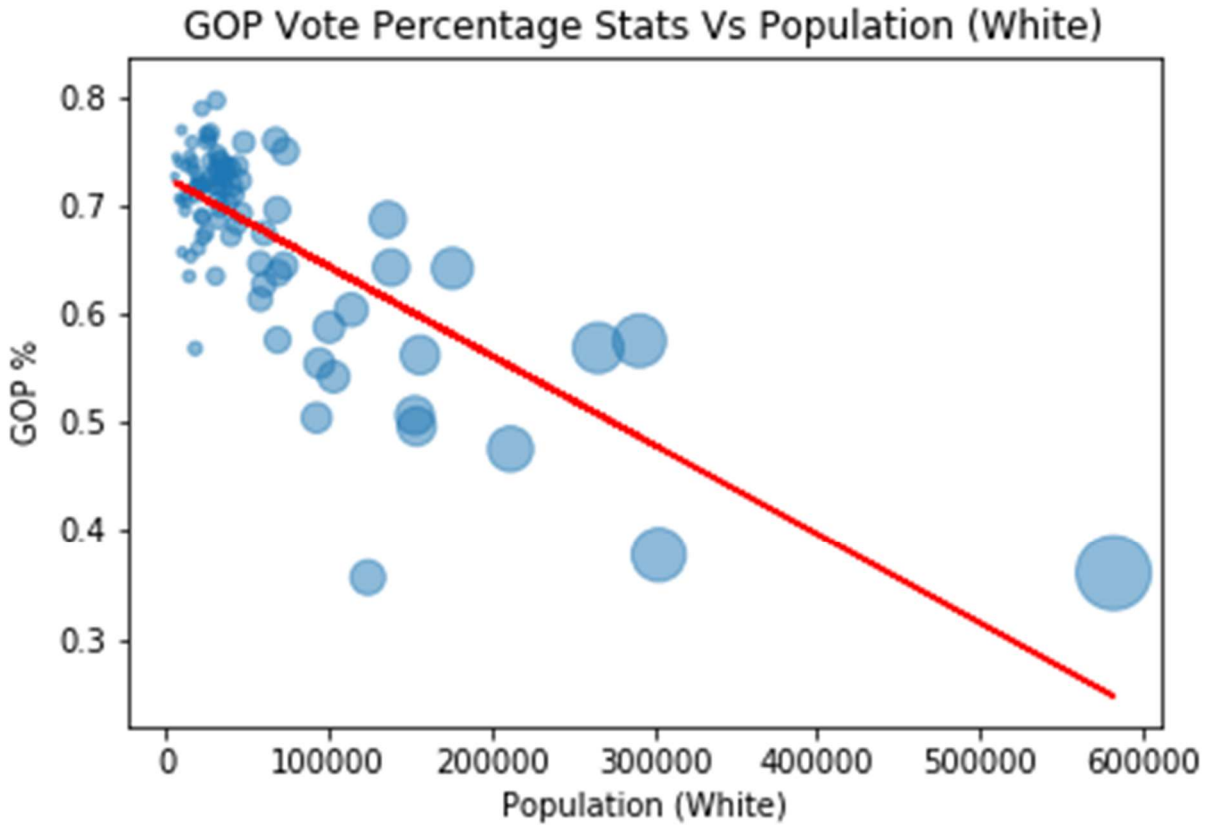

Dem Vote Percentage Stats Per Median Income
DEM %
0.6
0.5
0.4
0.3
0.2
40000
50000
60000
70000
80000
90000
Median Income


GOP Vote Percentage Stats Per Median Income
GOP %
0.8
0.7
0.6
0.5
0.4
40000
50000
60000
70000
80000
90000
Median Income

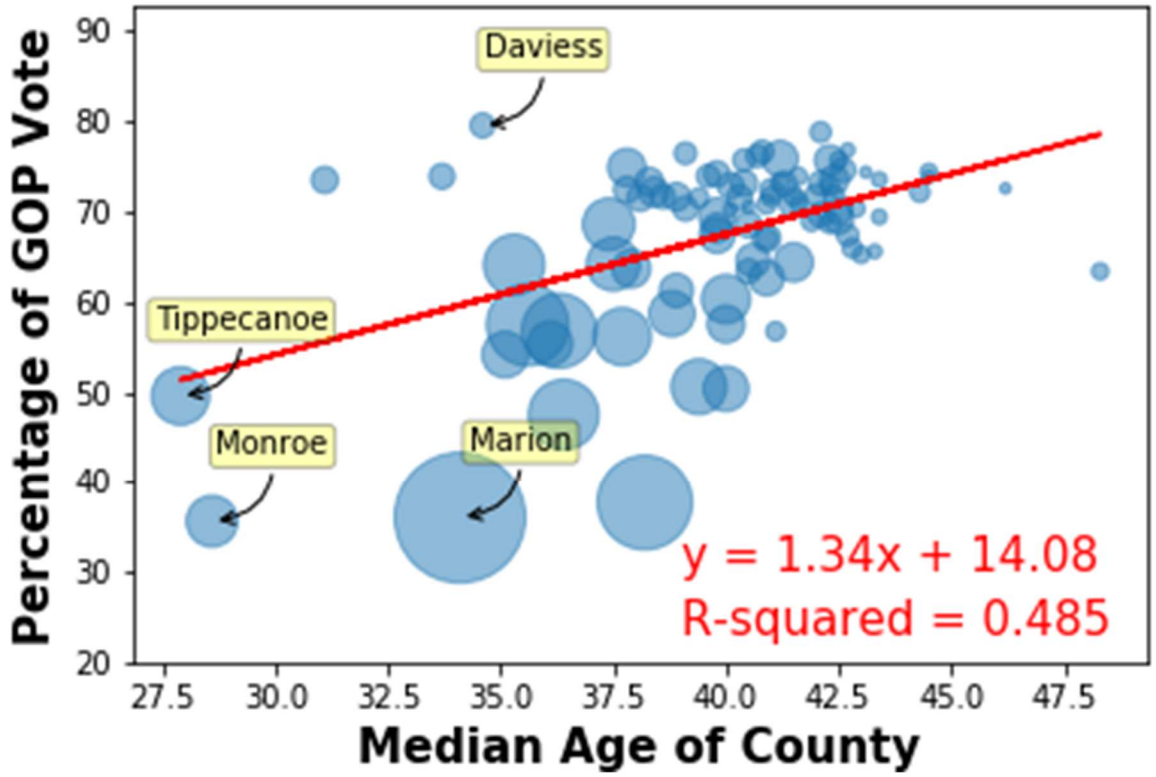

DEM Vote Percentage Stats For African-American Population
DEM %
0.7
0.6
0.5
0.4
0.3
0.2
0
50000
100000
150000
200000
250000
Median Income


GOP Vote Percentage Stats For African-American Population
GOP %
0.8
0.7
0.6
0.5
0.4
0.3
0
50000
100000
150000
200000
250000
Median Income


DEM Vote Percentage Stats For Population (White)
DEM %
0.7
0.6
0.5
0.4
0.3
0.2
0
100000
200000
300000
400000
500000
600000
Median Income


GOP Vote Percentage Stats Vs Population (White)
GOP %
0.8
0.7
0.6
0.5
0.4
0.3
0
100000
200000
300000
400000
500000
600000
Population (White)

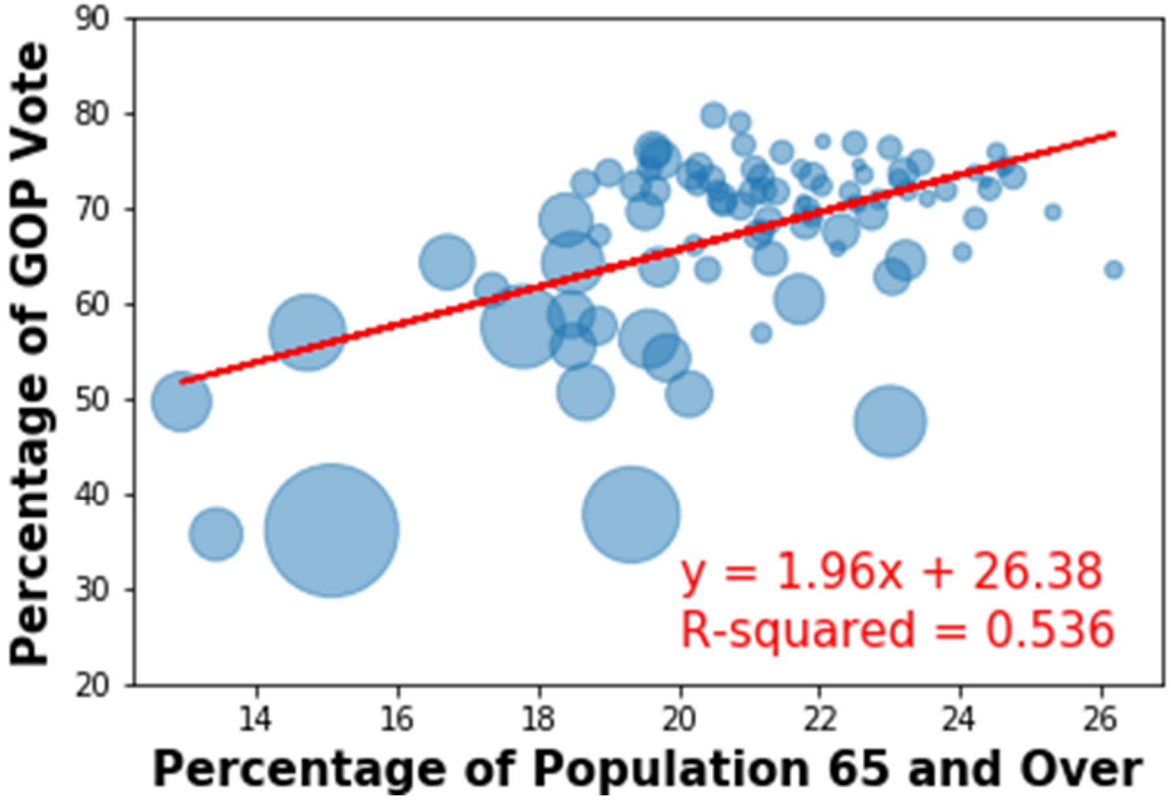
What effect does median age have on the GOP vote?

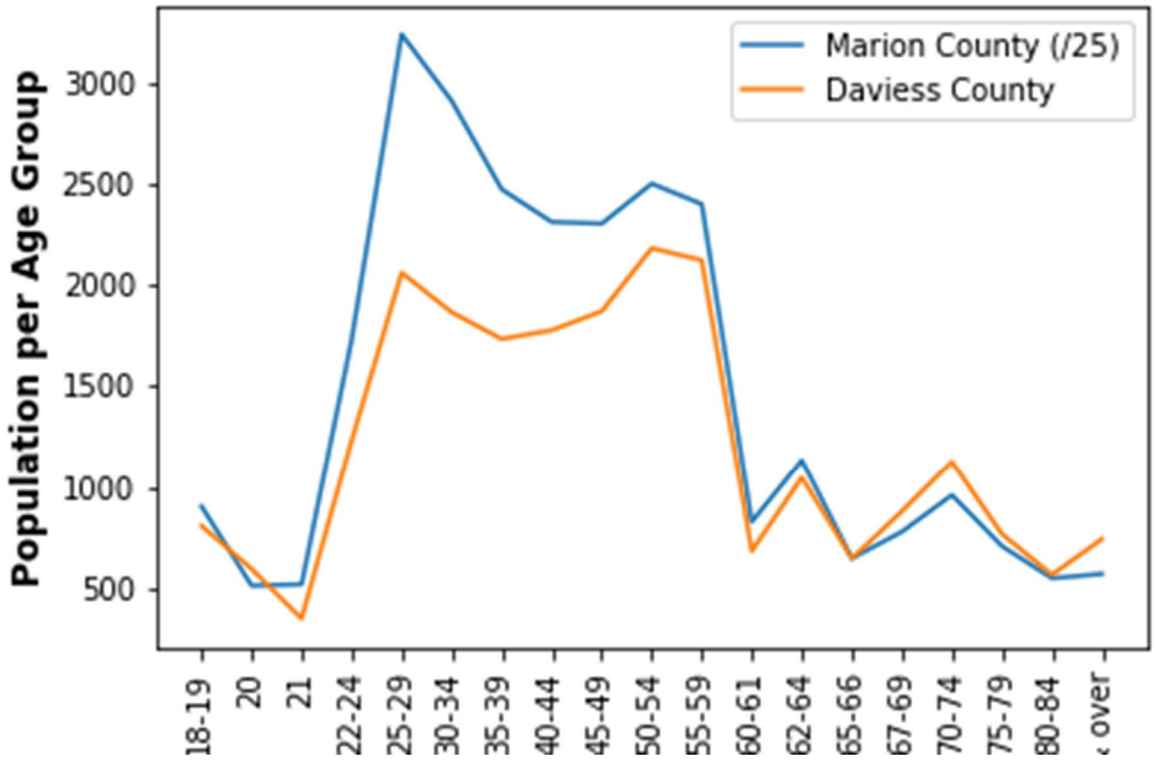


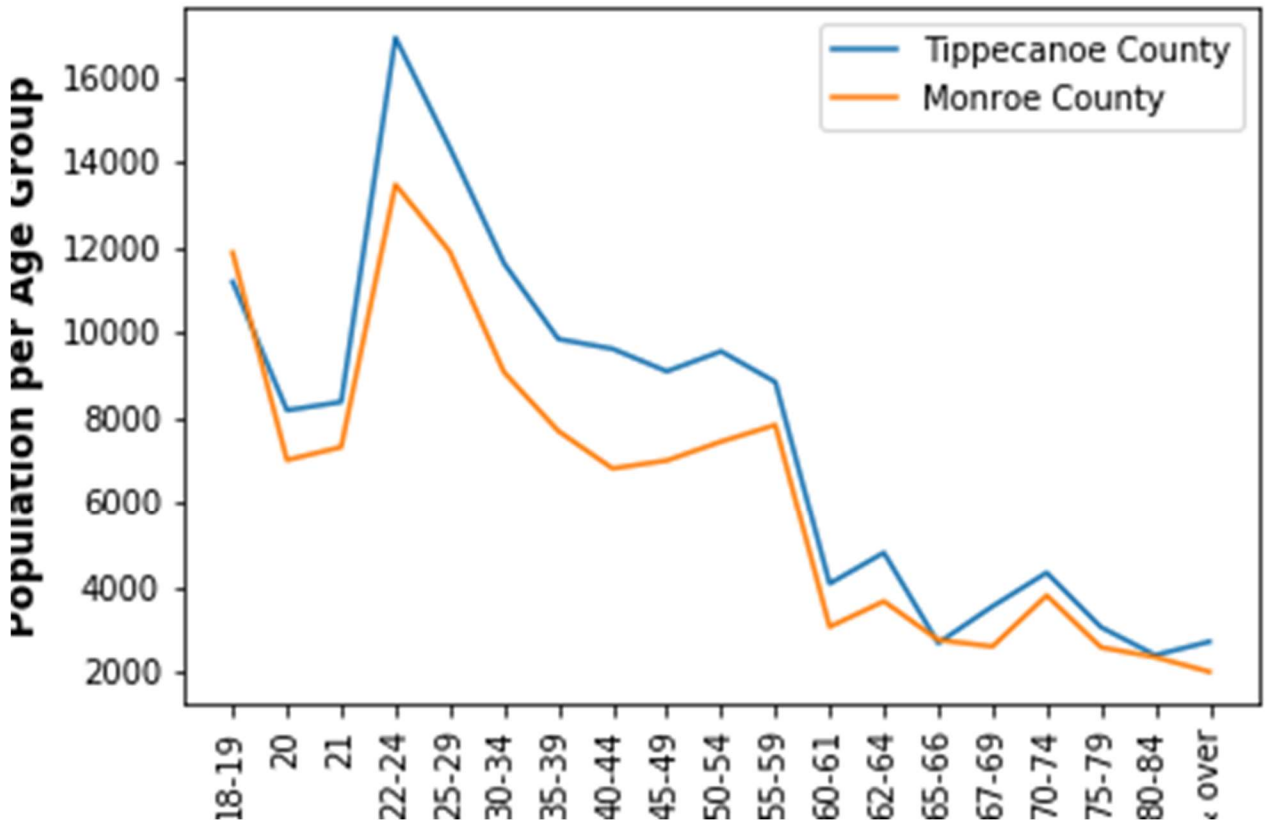
Voter Turnout per Age Group

- 18-29 year olds: 46.1%
- 30-44 year olds: 58.7%
- 45-64 year olds: 66.6%
- 65 years and oldes: 70.9%

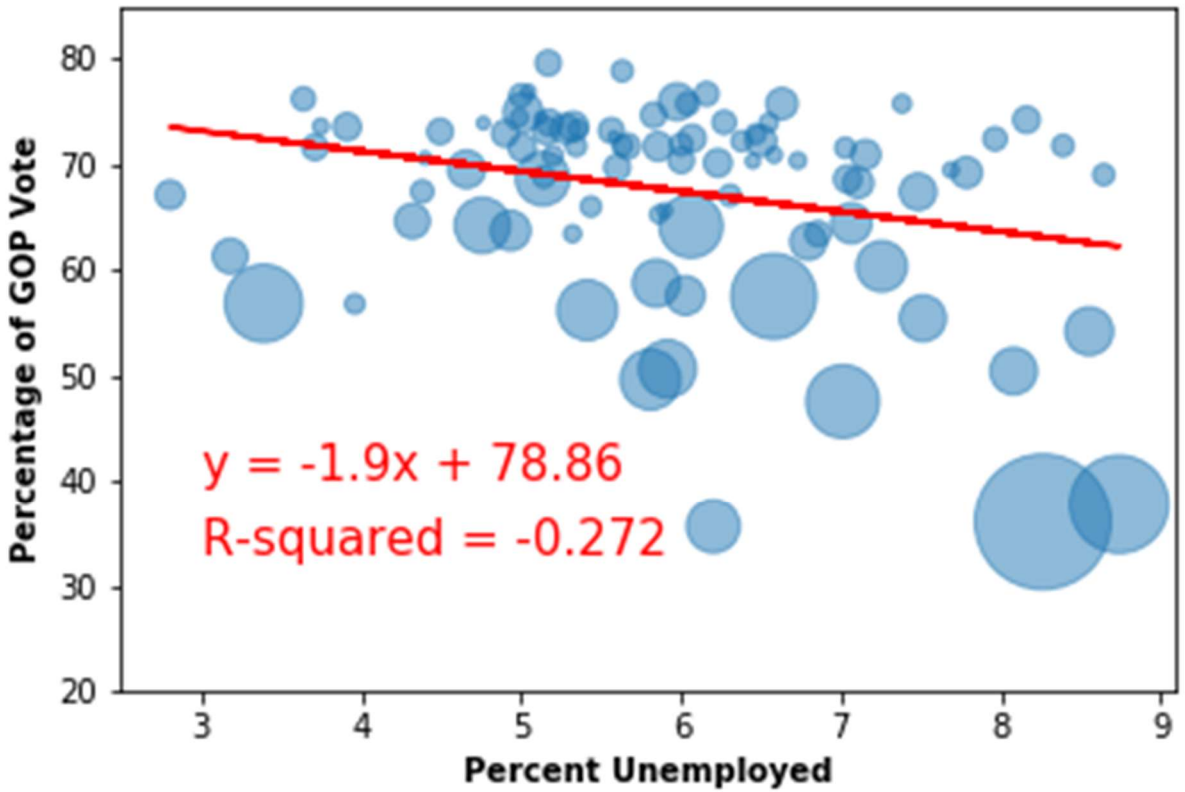
https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html


Percentage of GOP Vote
90
80
70
60
50
40
30
20
14
16
18
20
22
24
26
Percentage of Population 65 and Over
y = 1.96x + 26.38
R-squared = 0.536

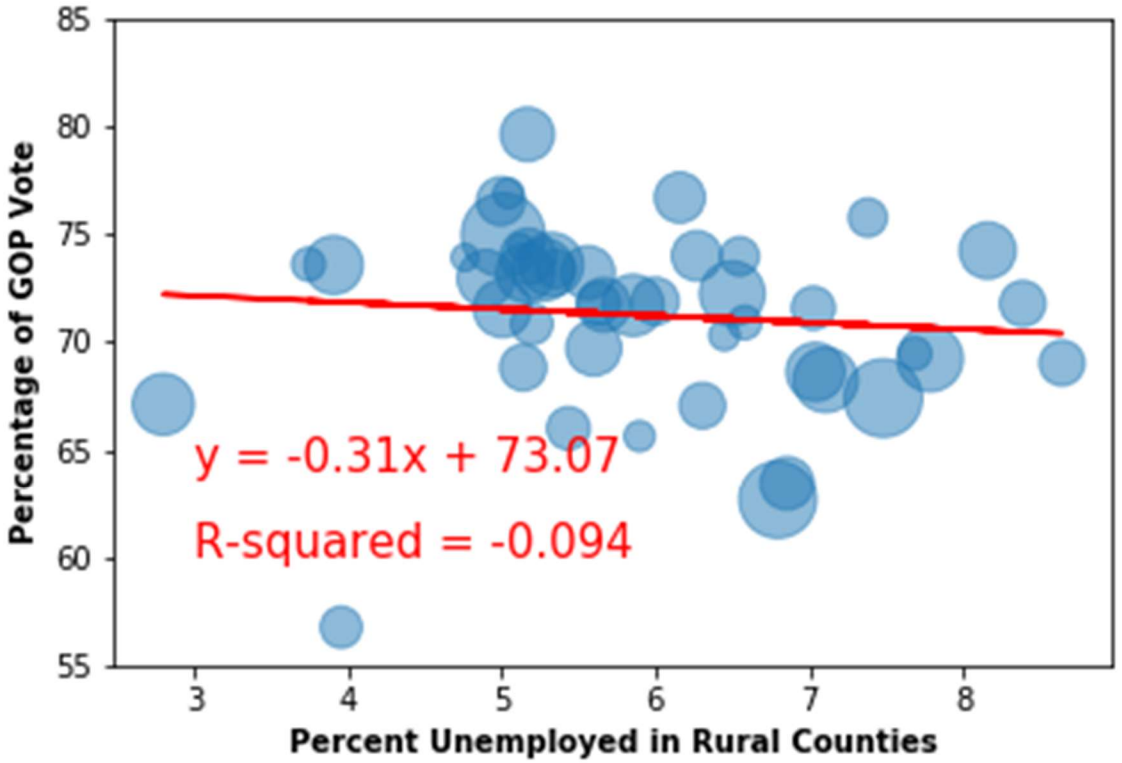

Marion County (/25)
Daviess County
Population per Age Group
3000
2500
2000
1500
1000
500
18-19
20
21
22-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-61
62-64
65-66
67-69
70-74
75-79
80-84


Tippecanoe County
Monroe County
Population per Age Group
16000
14000
12000
10000
8000
6000
4000
2000
18-19
20
21
22-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-61
62-64
65-66
67-69
70-74
75-79
80-84

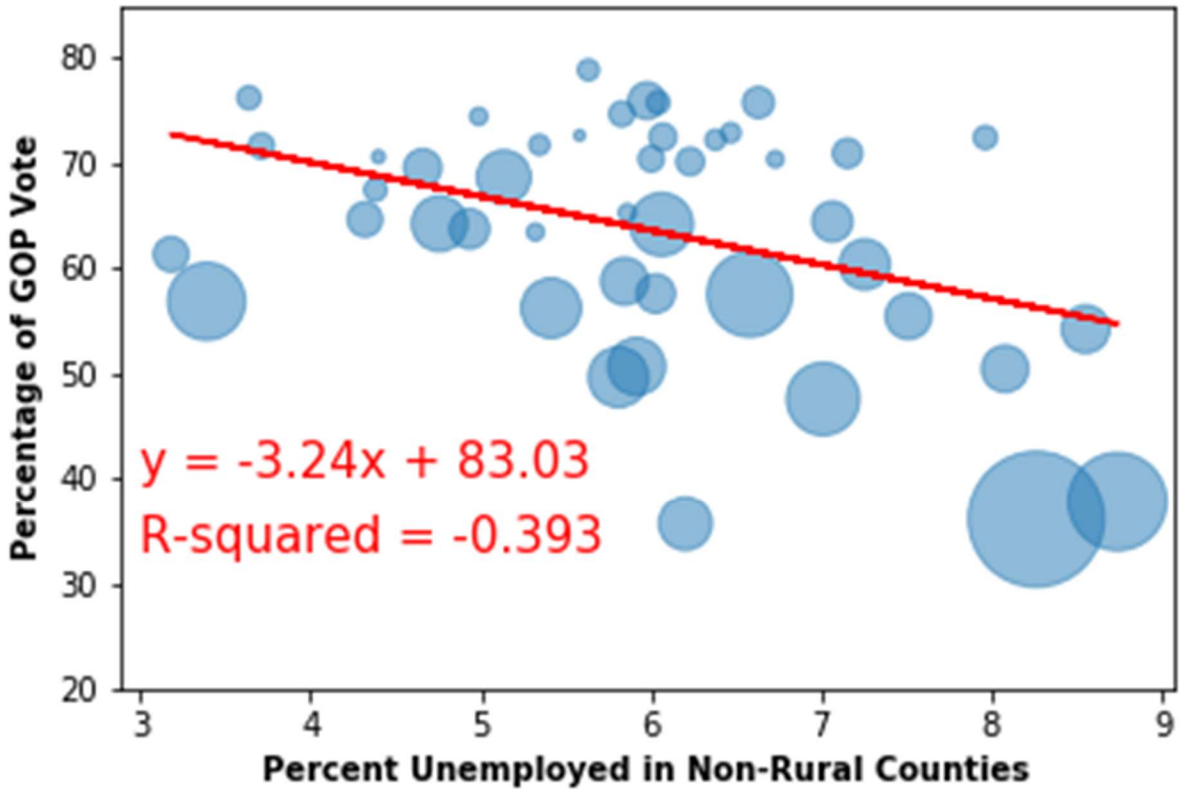
What effect does unemployment have on the GOP vote?



Unemployment in Rural Counties



https://www.ers.usda.gov/webdocs/DataFiles/53180/25569_IN.pdf?v=0


Percentage of GOP Vote
80
70
60
50
40
30
20
3
4
5
6
7
8
9
Percent Unemployed in Non-Rural Counties
y = -3.24x + 83.03
R-squared = -0.393

Conclusions

1) Any single demographic category is NOT a good predictor of which 2016 presidential candidate won any Indiana county.
    1) Pearson's correlation testing provided us with either weak or correlation
2) When multiple variables are introduced, we saw improved correlation between the census variable and the election results
    1) Pearson's correlation testing provided us with moderate to strong correlation on several of the plots we ran
        1) Education with race vs. DEM/GOP vote
        2) Race (white) vs. GOP vote
        3) Age (65+) vs GOP vote

G. Next steps

- With more time, we would develop a prediction model comparing actual vs. expected results and run ttests (Michael)
  - We would add:
    - More exit polling data
    - Election results from additional years and races
- We would incorporate and test more Census variables
- We would combine Census variables