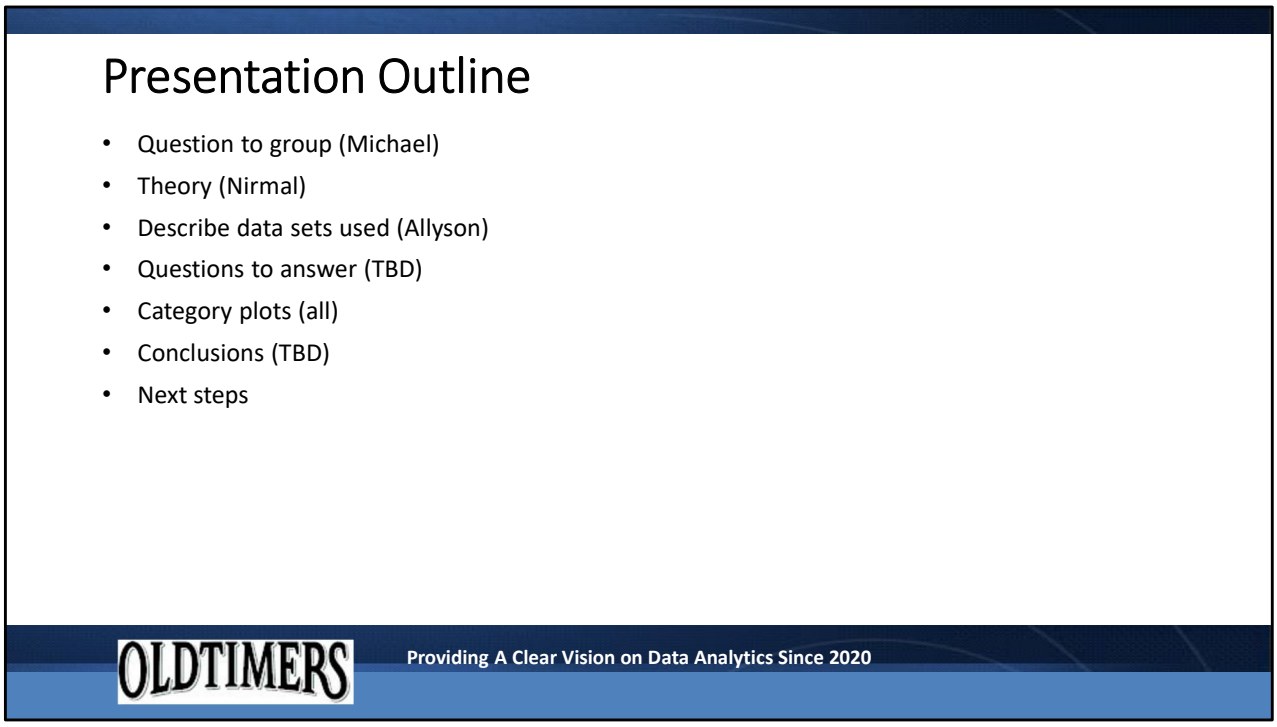
OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020

OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020

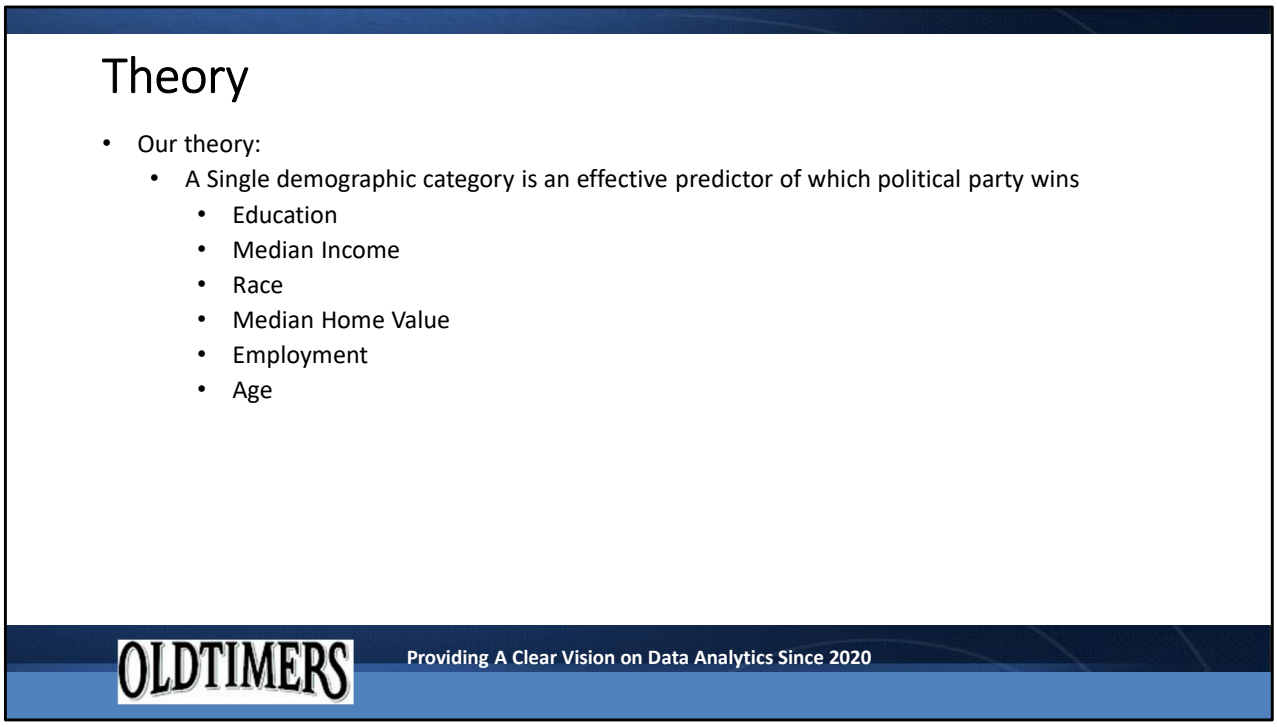


Presentation Outline

A. Question to group (Michael)
B. Theory (Nirmal)
C. Describe data sets used (Allyson)
D. Questions to answer (TBD)
E. Category plots (all)
F. Conclusions (TBD)
G. Next steps

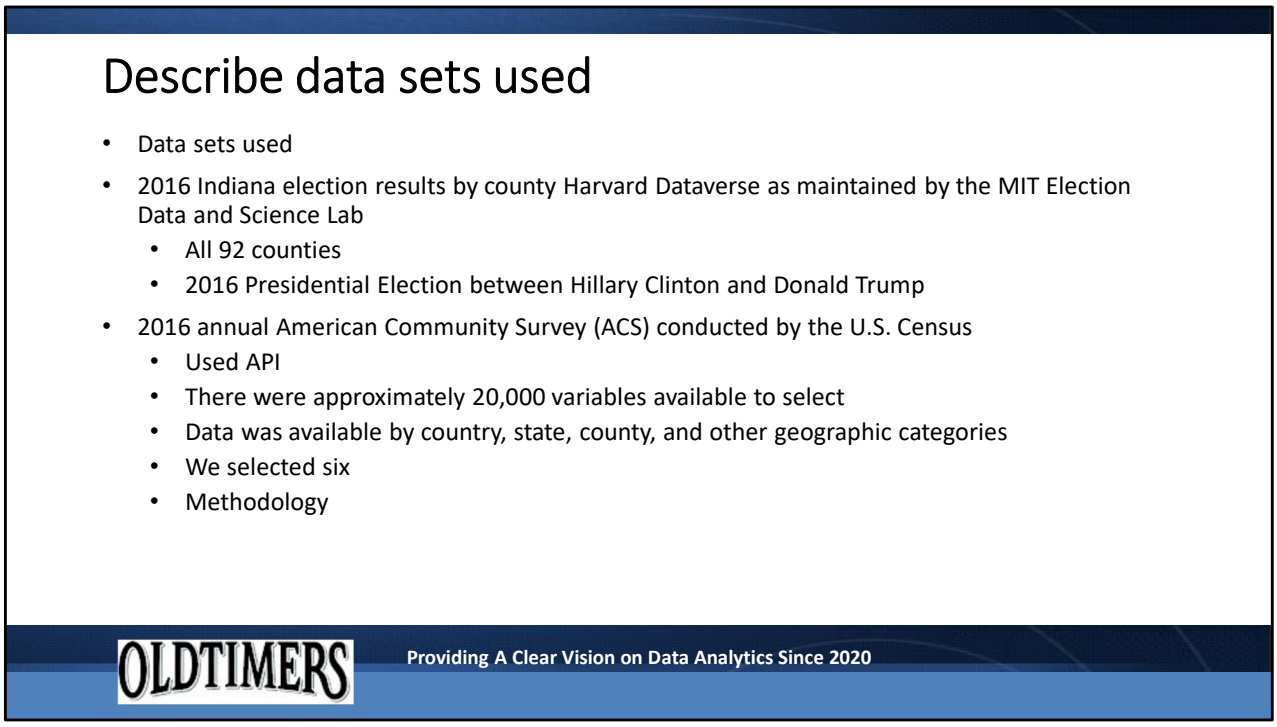
A. Question to group (Michael)

- Let's start off by asking a question
- Everyone raise your hand
- How many of you could predict who is going to win an election based on Census data?



B. Theory (Nirmal)

- Our theory:
  - Single demographic categories are effective predictors of which political party wins



C. Describe data sets used (Allyson)

- Data sets used
- 2016 Indiana election results by county from Harvard Dataverse as maintained by the MIT Election Data and Science Lab
  - How many counties
  - What race/candidates
- 5-year continuous American Community Survey (ACS) covering 2012-2016 conducted by the U.S. Census
  - Used API

- There were approximately 20,000 different variables available to select
- Data was available by country, state, county, and other geographic categories
- We selected six broad categories

- Methodology
  - Reach of 3.5M households per year
- Data cleanup & restructuring
  - Merging
  - Column headings
  - Finding correct Census codes to specific categories we wanted

Data Acquisition and Cleanup

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

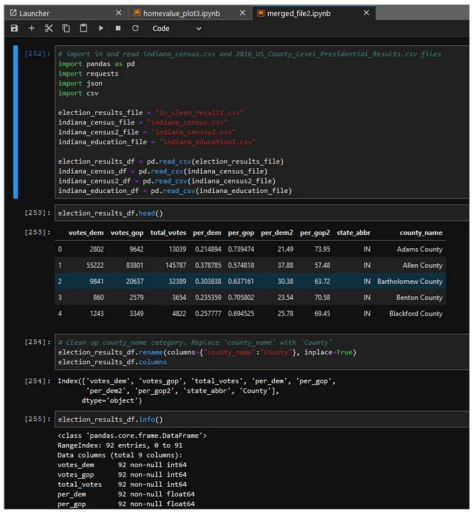
In [3]:
```
census_data = c.acs5.get(("NAME", "B01003_001E", "B02001_002E","B02001_003E","B02001_005E","B03001_003E",
                          "B25077_001E","B15003_002E","B15003_017E","B15003_018E",
                          "B15003_021E","B15003_022E","B15003_023E","B15003_024E","B15003_025E"), {'for':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



```
[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()
```

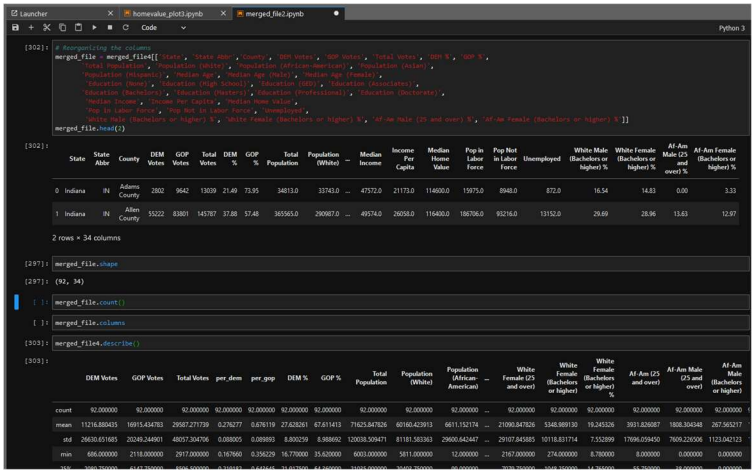
	votes_dem	votes_gop	total_votes	per_dem	per_gop	per_dem2	per_gop2	state_abbr	county_name
0	2802	9642	13039	0.214894	0.739474	21.49	73.95	IN	Adams County
1	55222	83801	145787	0.378785	0.574818	37.88	57.48	IN	Allen County
2	9841	20637	32389	0.303838	0.637161	30.38	63.72	IN	Bartholomew County
3	860	2579	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
```



```
[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()
```

```
[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	0.080005	0.089893	8.000259	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

OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

D. Questions to answer (TBD)

1) ?
2) ?
3) ?

Category plots

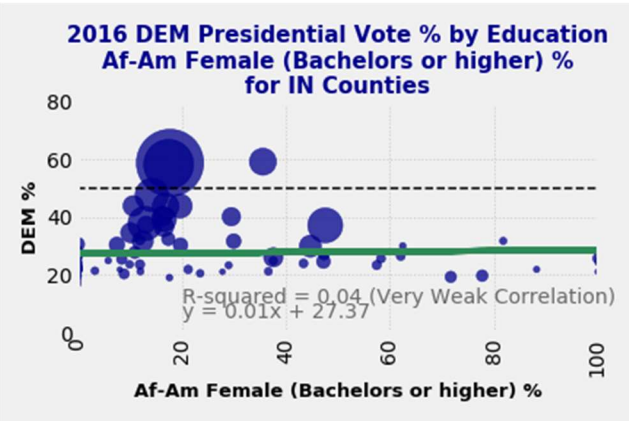
- Education & Home Value (Michael)
- Income & Race (Nirmal)
- Age & Employment (Allyson)

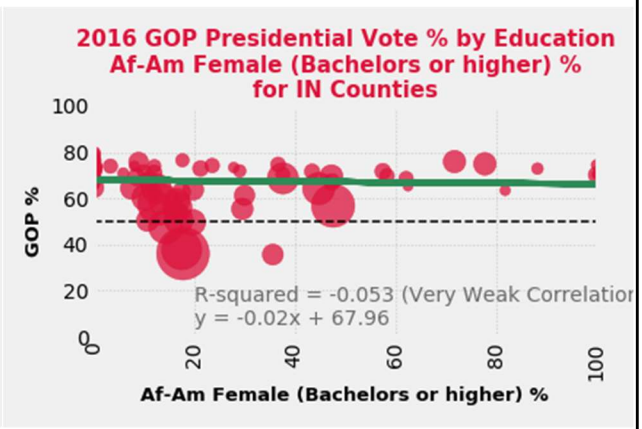
OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020

E. Category plots (all)

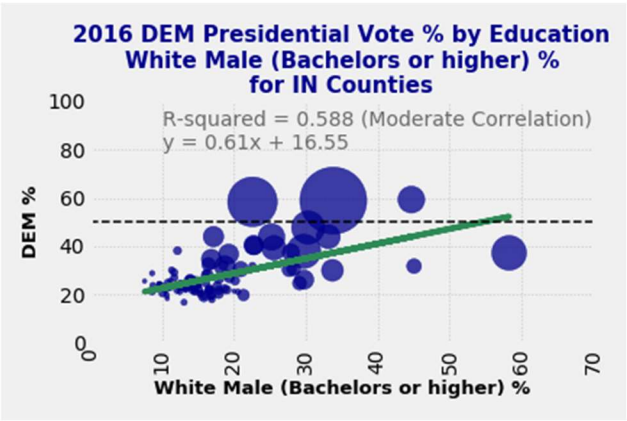
- Education & Home Value (Michael)
- Income & Race (Nirmal)
- Age & Employment (Allyson)

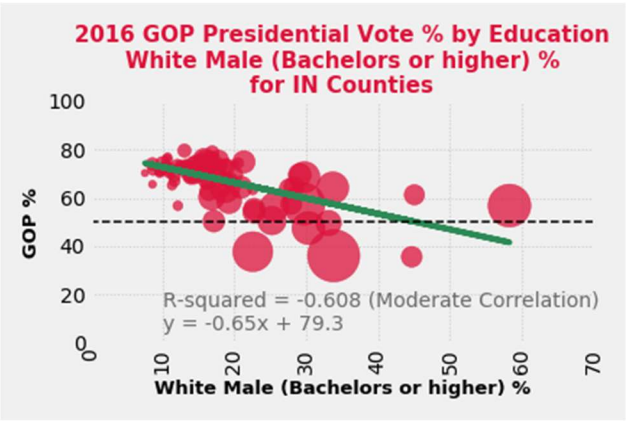




OLDTIMERS

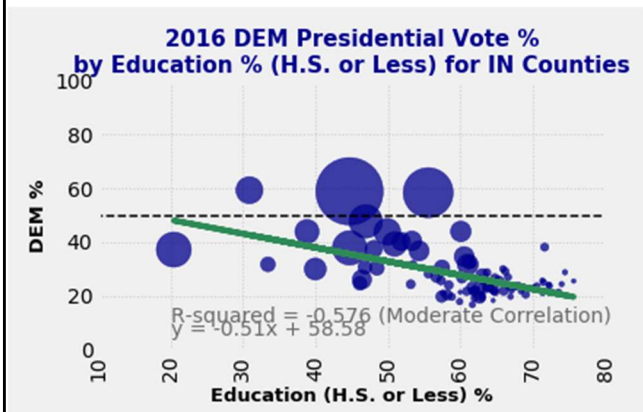
Providing A Clear Vision on Data Analytics Since 2020

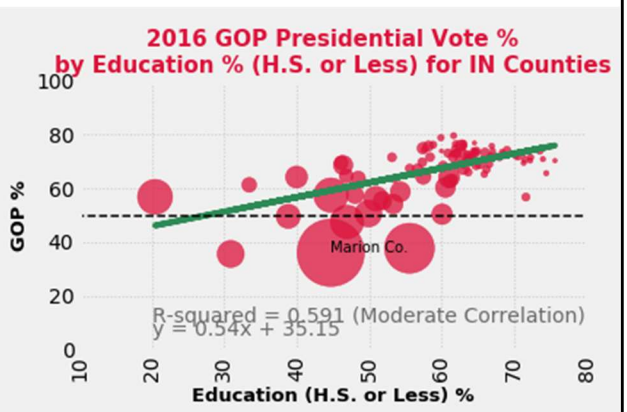




OLDTIMERS

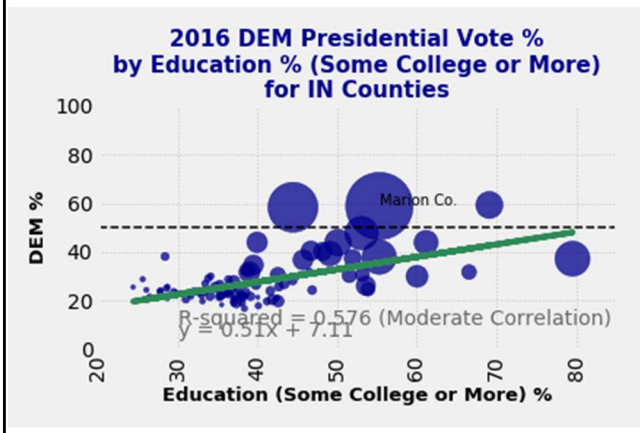
Providing A Clear Vision on Data Analytics Since 2020

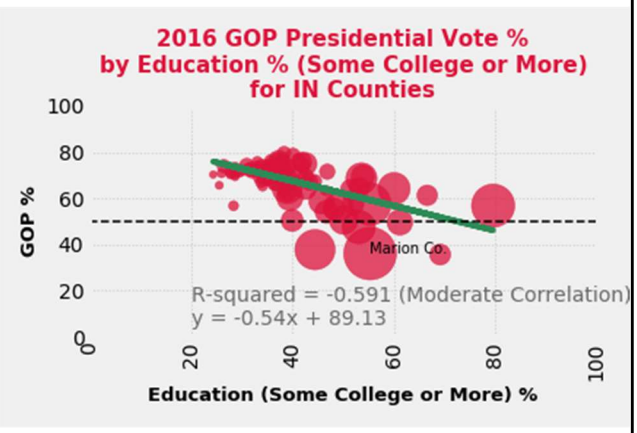


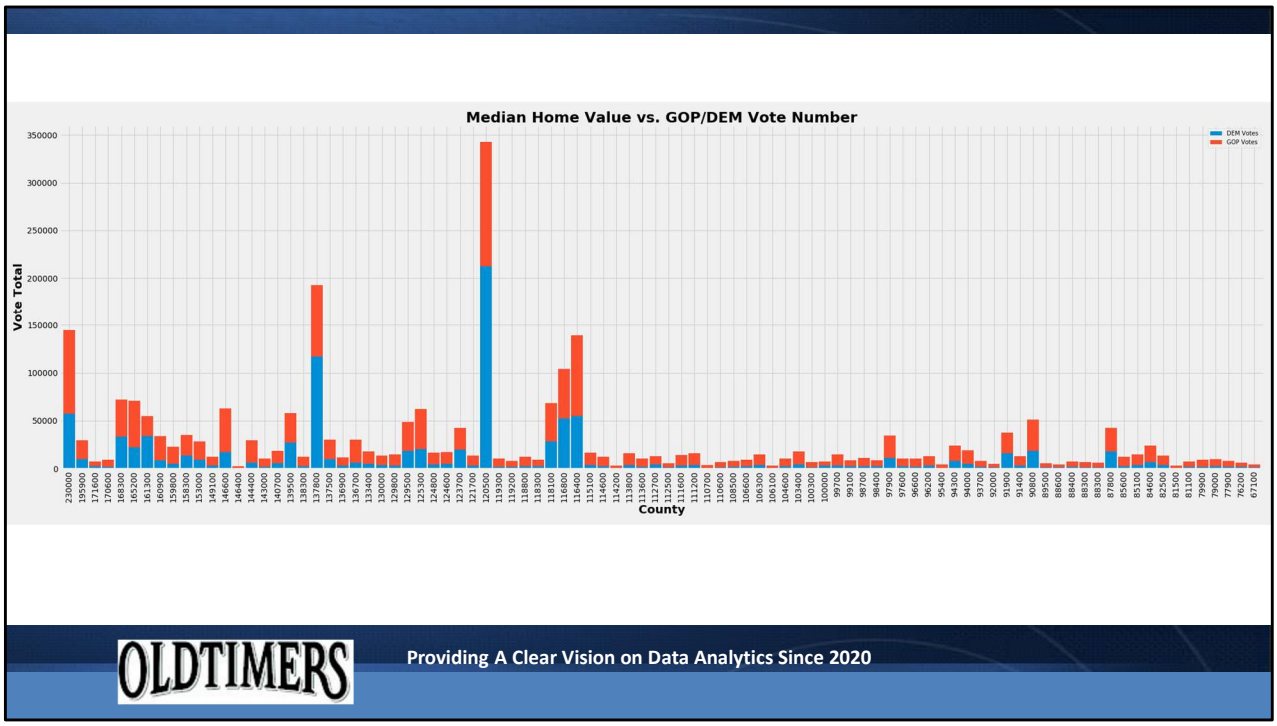


OLDTIMERS

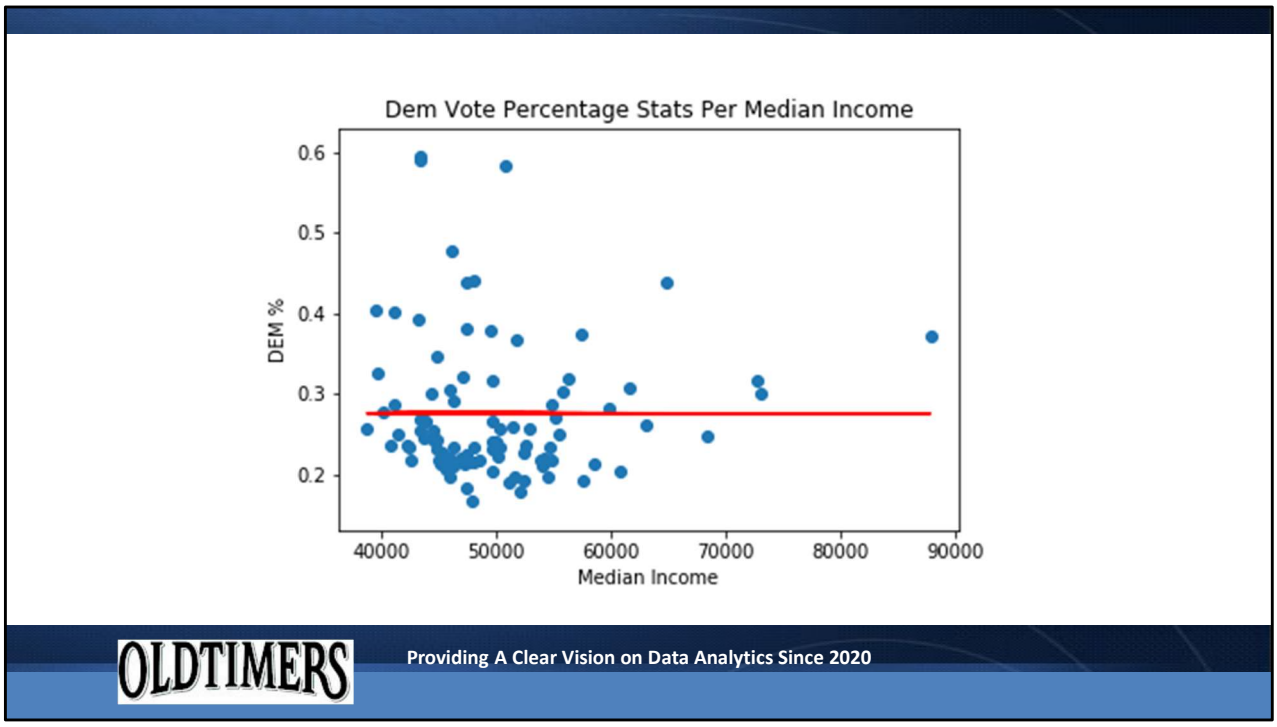
Providing A Clear Vision on Data Analytics Since 2020


2016 DEM Presidential Vote %
by Education % (Some College or More)
for IN Counties
DEM %
Education (Some College or More) %
Marion Co.
R-squared = 0.576 (Moderate Correlation)
y = 0.51x + 7.11

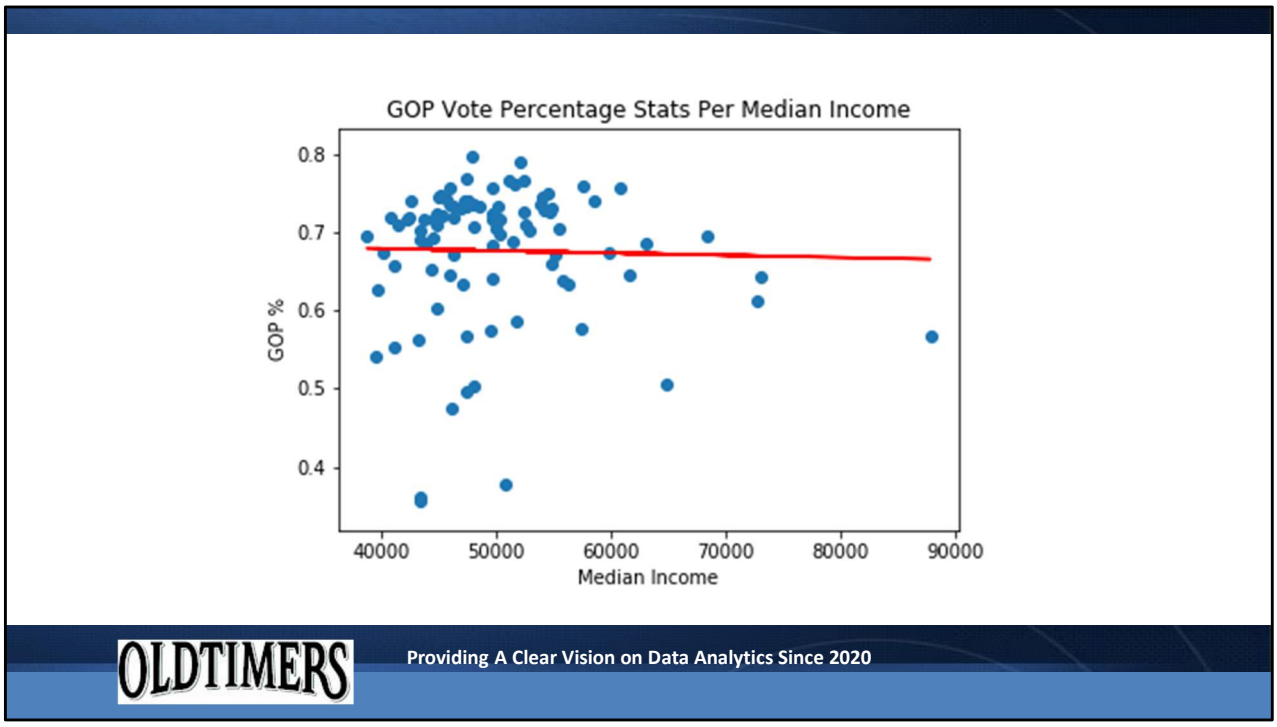

2016 GOP Presidential Vote %
by Education % (Some College or More)
for IN Counties
GOP %
Education (Some College or More) %
Marion Co.
R-squared = -0.591 (Moderate Correlation)
y = -0.54x + 89.13



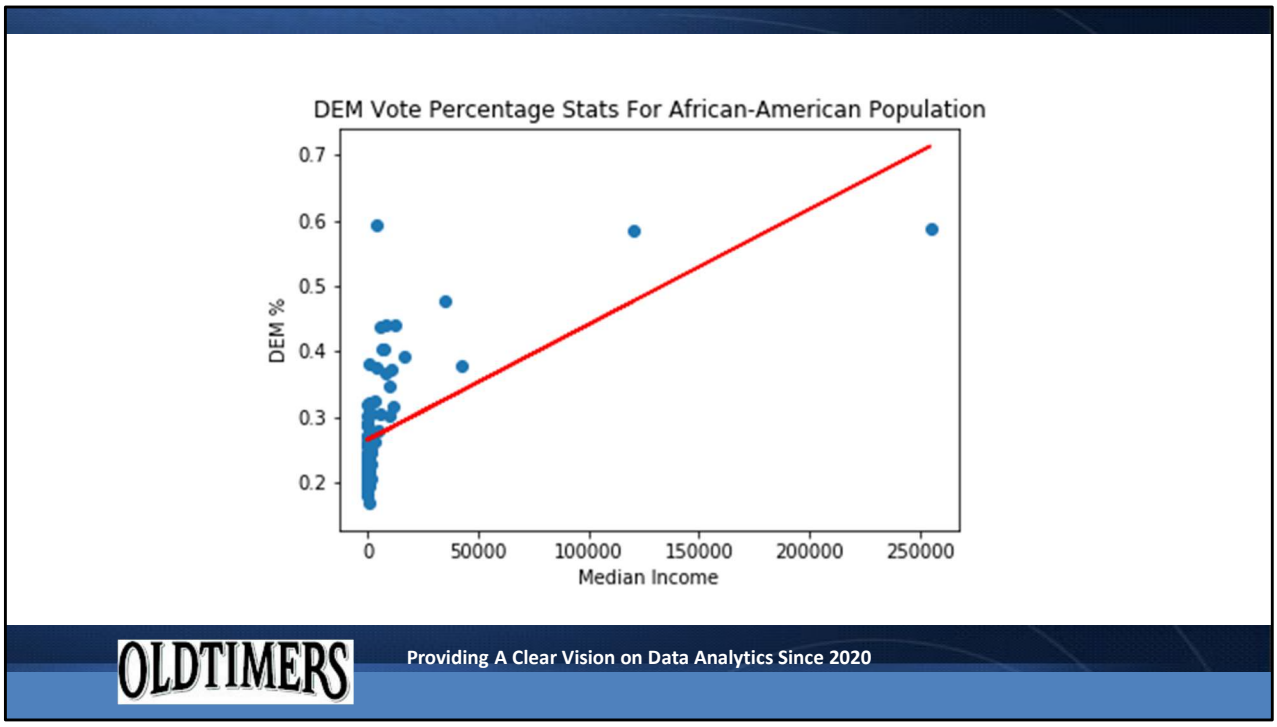
- Income & Race (Nirmal)



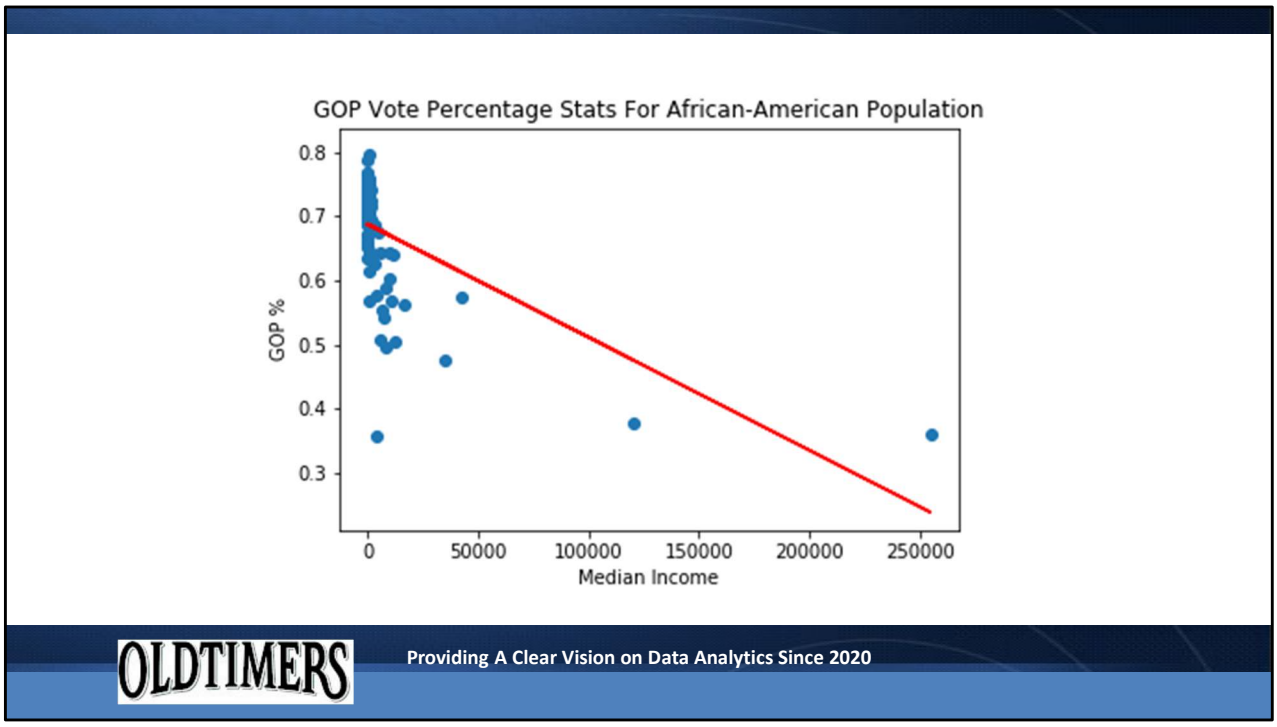
- Income & Race (Nirmal)



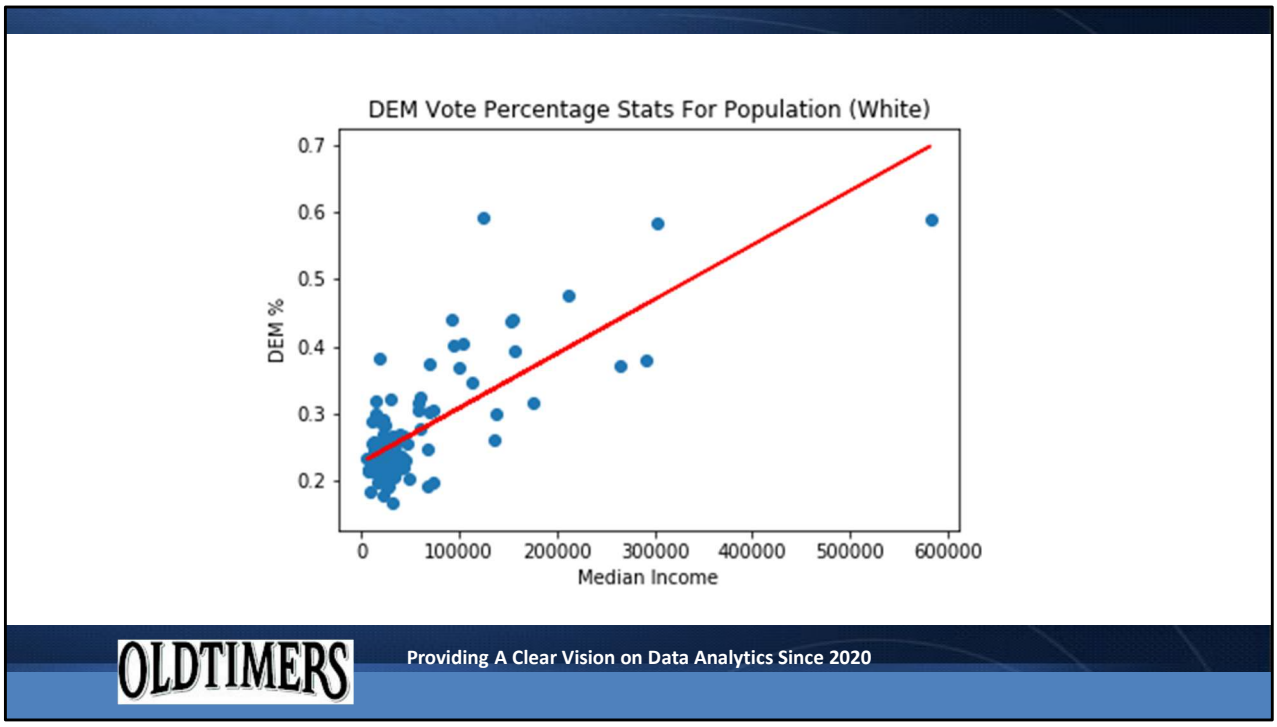
- Income & Race (Nirmal)



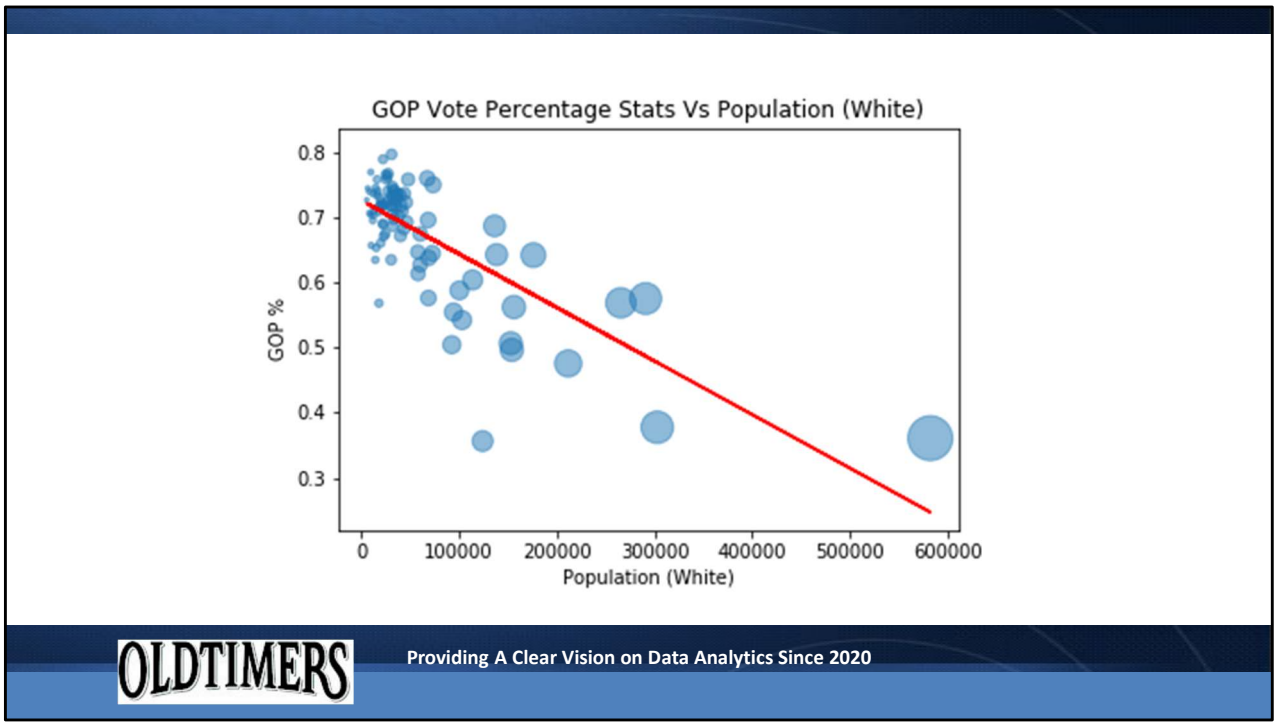
- Income & Race (Nirmal)



- Income & Race (Nirmal)

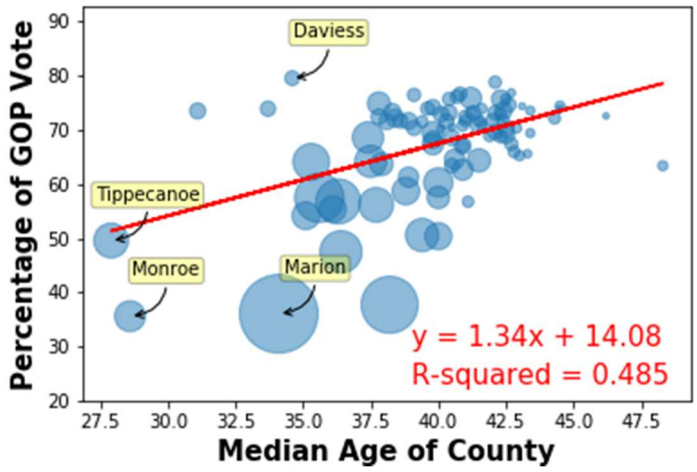


- Income & Race (Nirmal)



- Income & Race (Nirmal)

What effect does median age have on the GOP vote?



OLDTIMERS

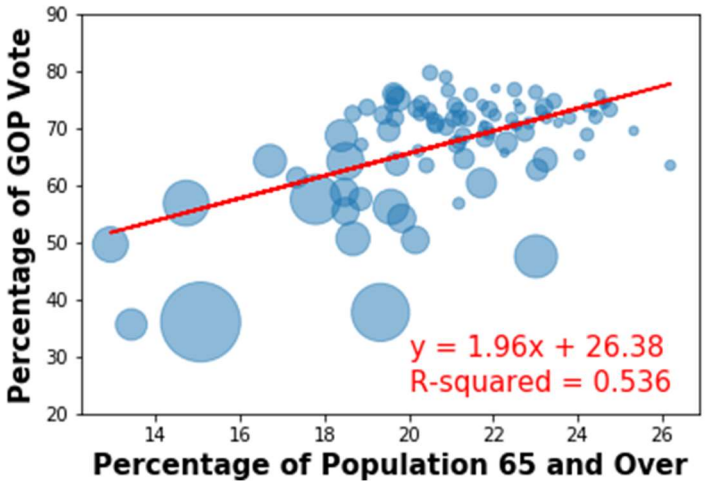
Providing A Clear Vision on Data Analytics Since 2020

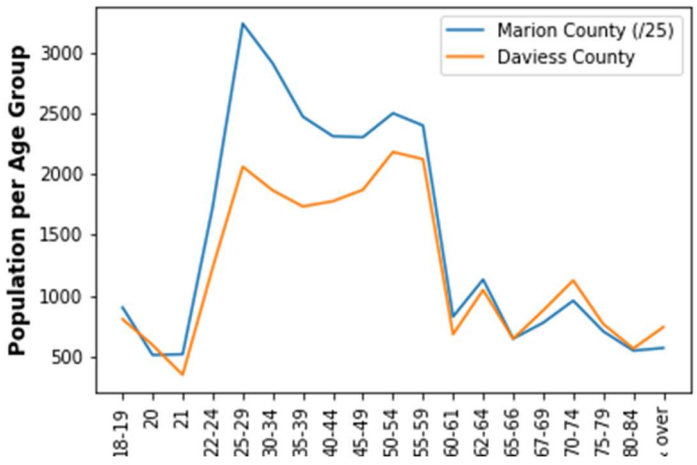
- Age & Employment (Allyson)

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%

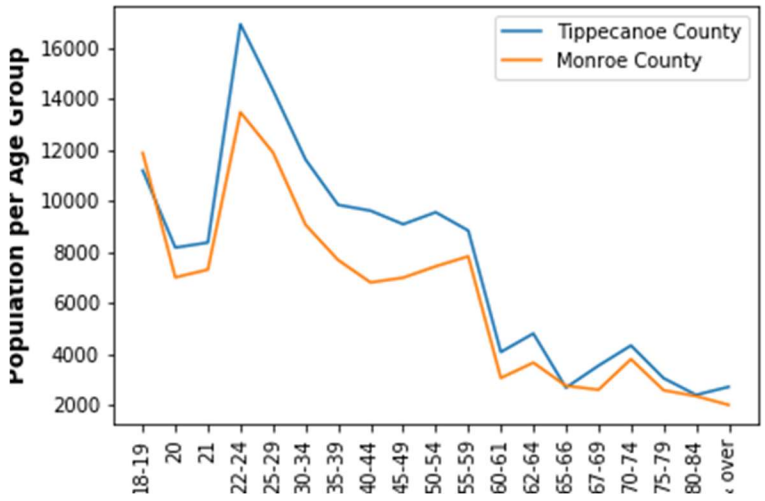
https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html


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



OLDTIMERS

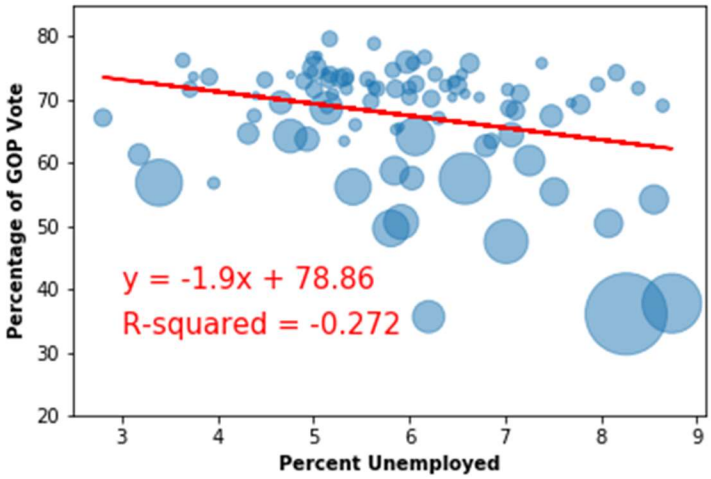
Providing A Clear Vision on Data Analytics Since 2020



OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020

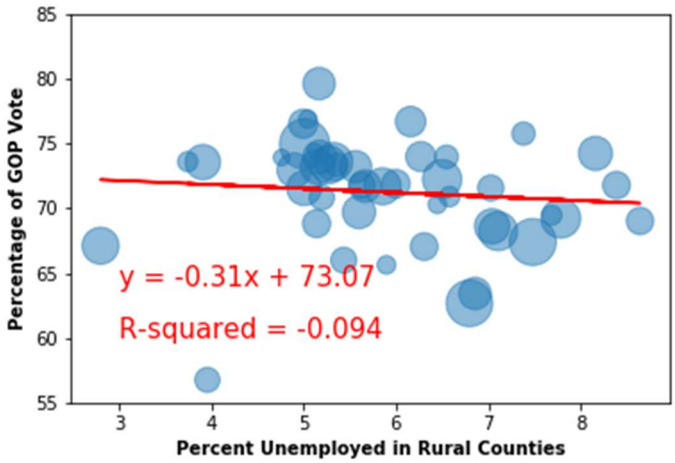
What effect does unemployment have on the GOP vote?



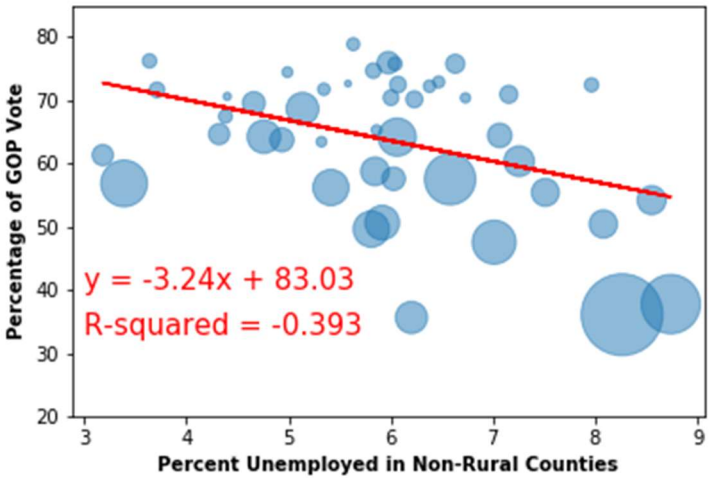
OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020

Unemployment in Rural Counties



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



OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020

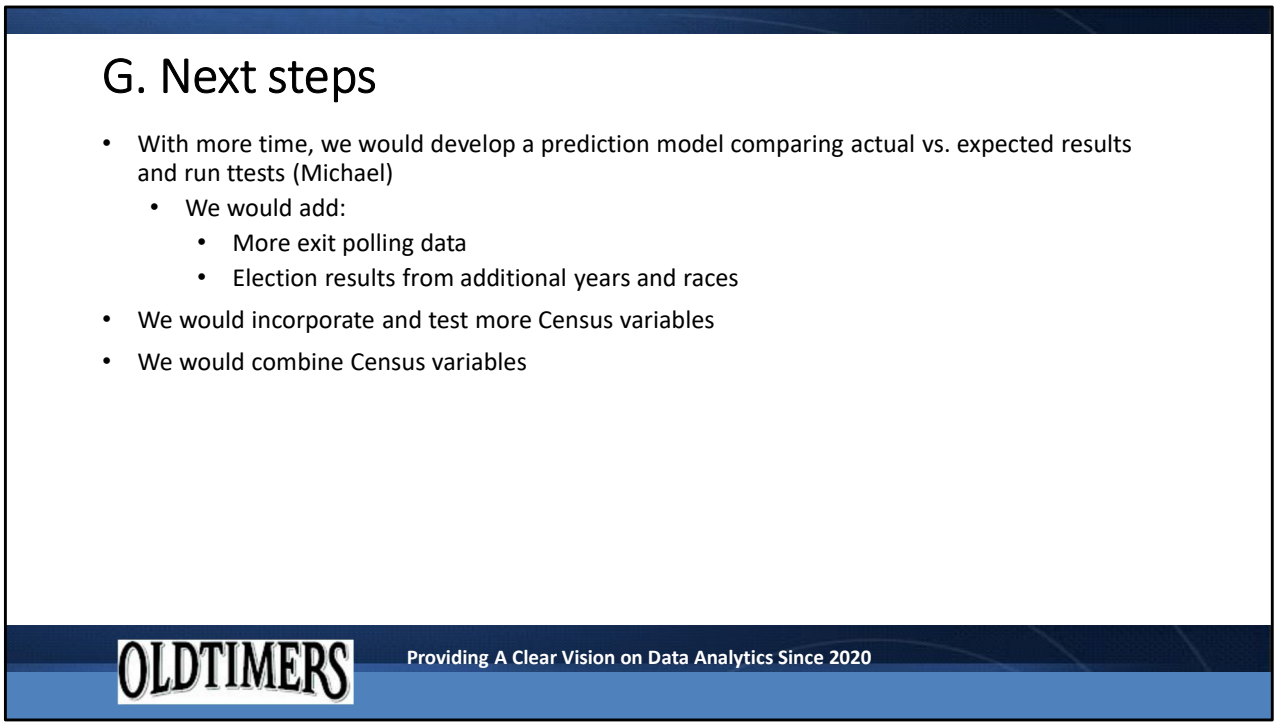
Conclusions

1) Any single demographic category is NOT a good predictor of which 2016 presidential candidate won any Indiana county.
    1) None of the plots we ran provided us with a moderate or strong 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

OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

F. Conclusions (TBD)

1) Any single demographic category is NOT a good predictor of which 2016 presidential candidate won any Indiana county.
2) ?
3) ?



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