OLDTIMERS

Providing A Clear Vision on Data Analytics Since 2020



1

## Presentation Outline

- Question to group (Michael)
- Theory (Nirmal)
- Describe data sets used (Allyson)
- Questions to answer (TBD)
- Category plots (all)
- Conclusions (TBD)
- Next steps



2

?



3

## 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



4

## 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



5

## 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_024E":"Education (Masters)",
                                                        "B15003_025E":"Education (Doctorate)"})
        census_complete.head()
        #census_county = census_complete[1].str.split(' ').apply(lambda, 1)
```

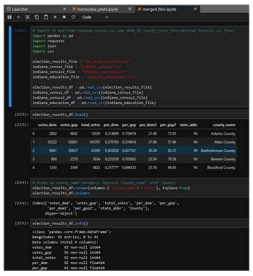
Out[4]:

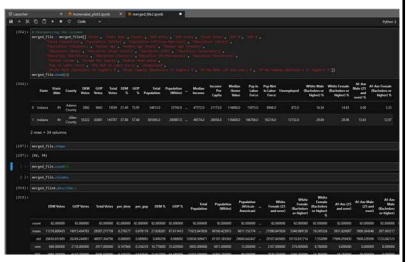
| | NAME | Total Population | Population (White) | Population (African-American) | Population (Asian) | Population (Hispanic) | Median Home Value | Education (None) | Education (High School) | Education (GED) | Education (Associates) | Education (Bachelors) | Education (Masters) | Ed (Prof |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Carroll County, Arkansas | 27890.0 | 25850.0 | 316.0 | 245.0 | 4021.0 | 116500.0 | 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 | 229.0 | |



6

## Data Acquisition and Cleanup





OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

7

## 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

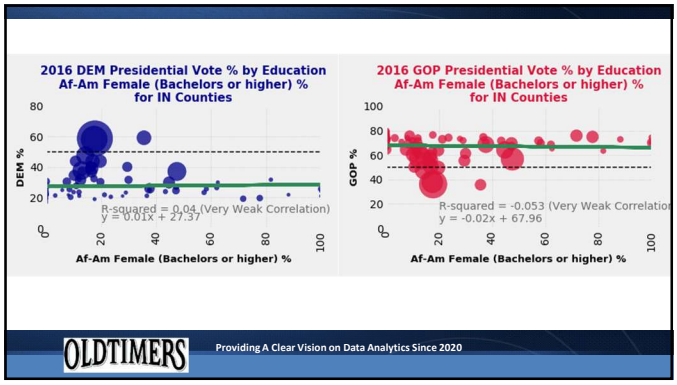
8

## Category plots

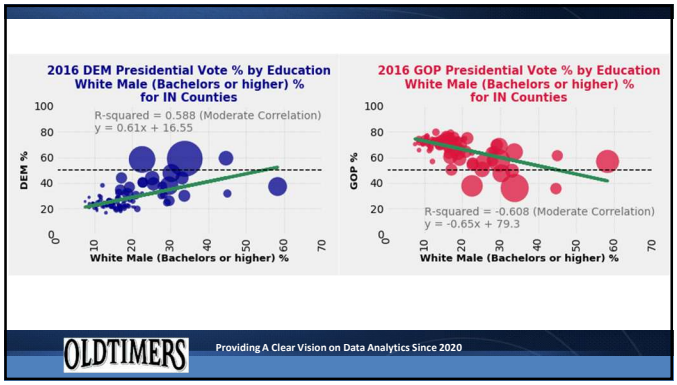
- Education & Home Value (Michael)
- Income & Race (Nirmal)
- Age & Employment (Allyson)

OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

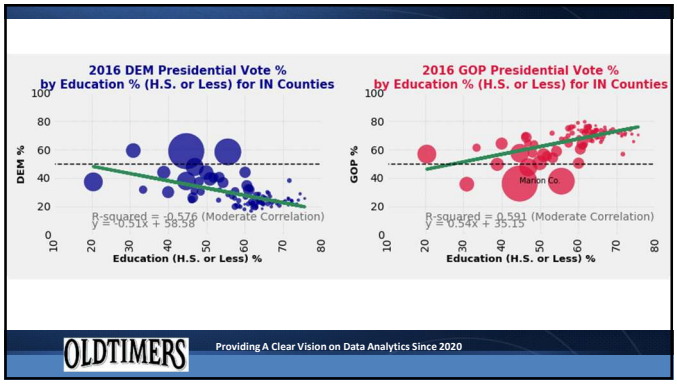
9


2016 DEM Presidential Vote % by Education
Af-Am Female (Bachelors or higher) %
for IN Counties
DEM %
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 %
R-squared = -0.053 (Very Weak Correlation)
y = -0.02x + 67.96
Af-Am Female (Bachelors or higher) %
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

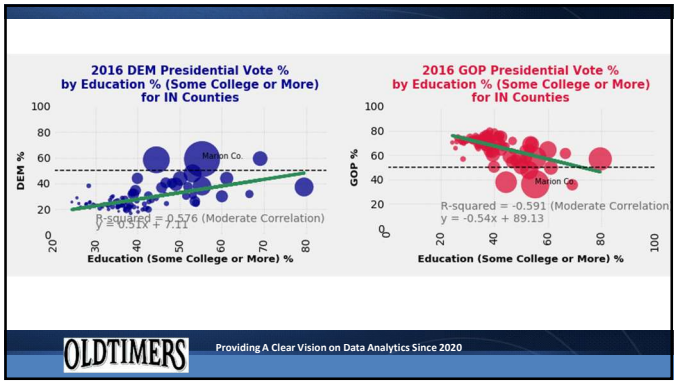
10


2016 DEM Presidential Vote % by Education
White Male (Bachelors or higher) %
for IN Counties
DEM %
R-squared = 0.588 (Moderate Correlation)
y = 0.61x + 16.55
White Male (Bachelors or higher) %
2016 GOP Presidential Vote % by Education
White Male (Bachelors or higher) %
for IN Counties
GOP %
R-squared = -0.608 (Moderate Correlation)
y = -0.65x + 79.3
White Male (Bachelors or higher) %
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

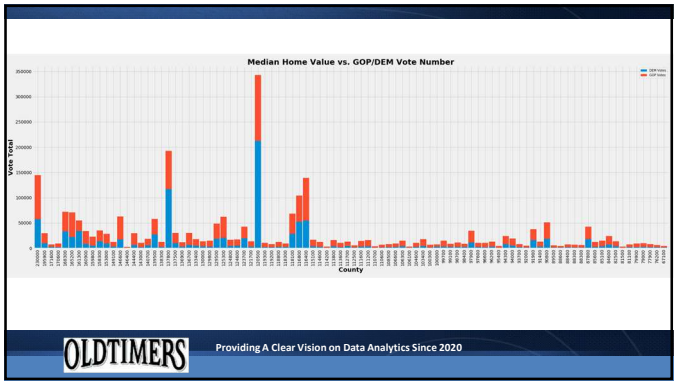
11


2016 DEM Presidential Vote %
by Education % (H.S. or Less) for IN Counties
DEM %
R-squared = -0.576 (Moderate Correlation)
y = -0.51x + 58.58
Education (H.S. or Less) %
2016 GOP Presidential Vote %
by Education % (H.S. or Less) for IN Counties
GOP %
Marion Co.
R-squared = 0.591 (Moderate Correlation)
y = 0.54x + 35.15
Education (H.S. or Less) %
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

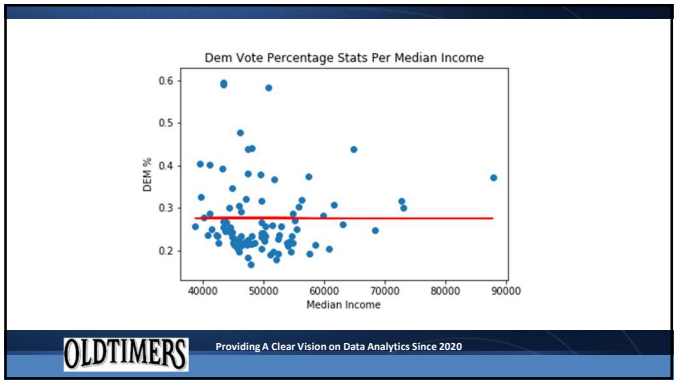
12


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
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

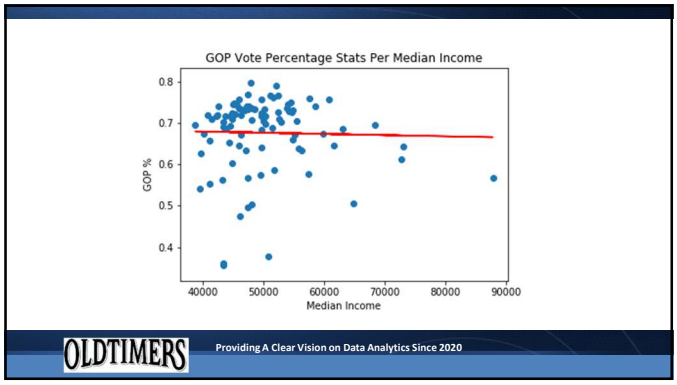
13


Median Home Value vs. GOP/DEM Vote Number
Vote Total
County
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

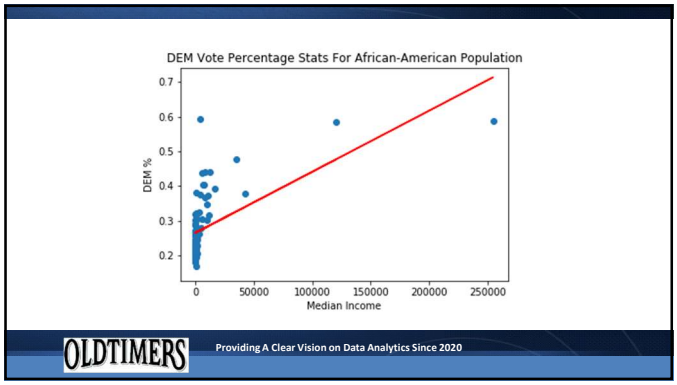
14


Dem Vote Percentage Stats Per Median Income
DEM %
Median Income
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

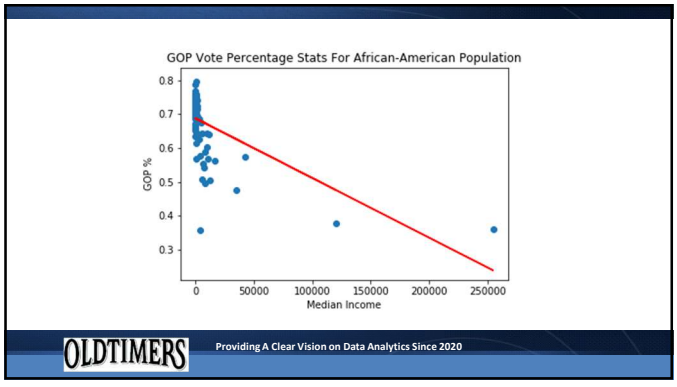
15


GOP Vote Percentage Stats Per Median Income
GOP %
Median Income
0.8
0.7
0.6
0.5
0.4
40000
50000
60000
70000
80000
90000
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

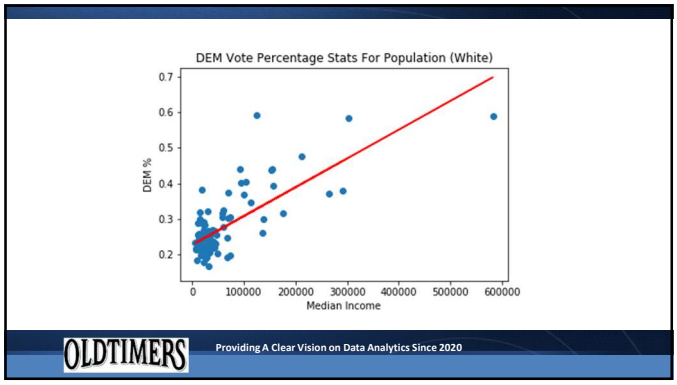
16


DEM Vote Percentage Stats For African-American Population
DEM %
Median Income
0.7
0.6
0.5
0.4
0.3
0.2
0
50000
100000
150000
200000
250000
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

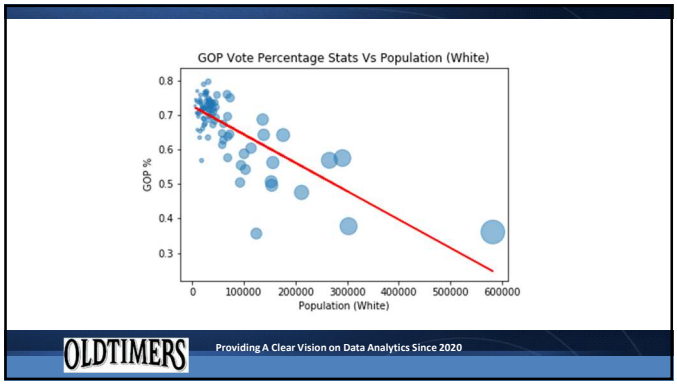
17


GOP Vote Percentage Stats For African-American Population
GOP %
Median Income
0.8
0.7
0.6
0.5
0.4
0.3
0
50000
100000
150000
200000
250000
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

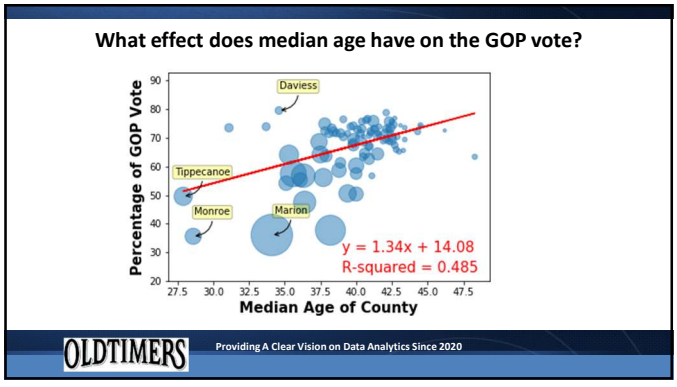
18


DEM Vote Percentage Stats For Population (White)
DEM %
Median Income
0.7
0.6
0.5
0.4
0.3
0.2
0
100000
200000
300000
400000
500000
600000
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

19


GOP Vote Percentage Stats Vs Population (White)
GOP %
Population (White)
0.8
0.7
0.6
0.5
0.4
0.3
0
100000
200000
300000
400000
500000
600000
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

20


What effect does median age have on the GOP vote?
Percentage of GOP Vote
Median Age of County
Daviess
Tippecanoe
Monroe
Marion
y = 1.34x + 14.08
R-squared = 0.485
90
80
70
60
50
40
30
20
27.5
30.0
32.5
35.0
37.5
40.0
42.5
45.0
47.5
OLDTIMERS
Providing A Clear Vision on Data Analytics Since 2020

21

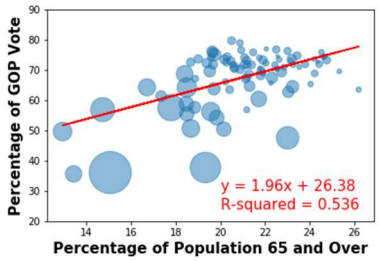
**Voter Turnout per Age Group**

- 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

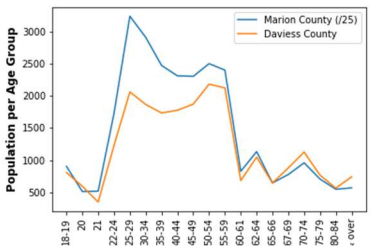
OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

22



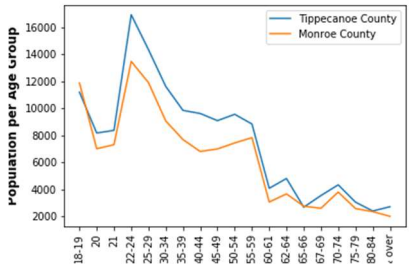

OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

23




OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

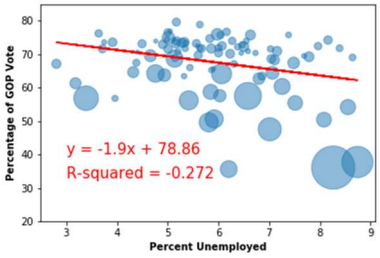
24




OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

25

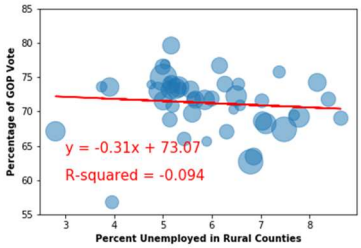
## What effect does unemployment have on the GOP vote?




OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

26

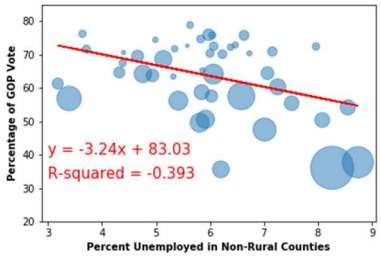
## Unemployment in Rural Counties




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

OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

27




OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

28

## 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

29

## 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

OLDTIMERS Providing A Clear Vision on Data Analytics Since 2020

30