2016 ELECTION

Tony Reyes. Jessica Gallardo, Lidzy Jimenez

DATA OVERVIEW

Key Data Points:

- **2016 Data**
- **Primary Results:** Candidate, Party, Vote Count & Pct.
- **County Demographics:** 3,144 counties (55 columns) - Pop., Gender, Race, Edu, Income, etc.
- **35.6% Missing Values**

MISSING VALUES

- Alaska
- Colorado
- Connecticut
- Illinois
- Kansas
- Maine

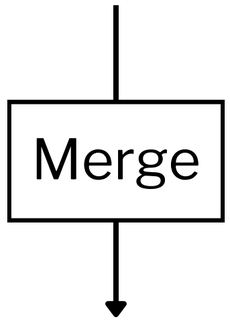
Democratic Votes	**3%**
Republican Votes	**15%**
Democratic Votes (Fraction)	**3%**
Republican Votes (Fraction)	**15%**
Fips	**33.5%**
...	**33.5%**

- Massachusetts
- North Dakota
- Rhode Island
- Vermont
- Wyoming

Primary results:

Include counties that aren't real:

- State House District 1, State House District 10, State House District 11, etc.

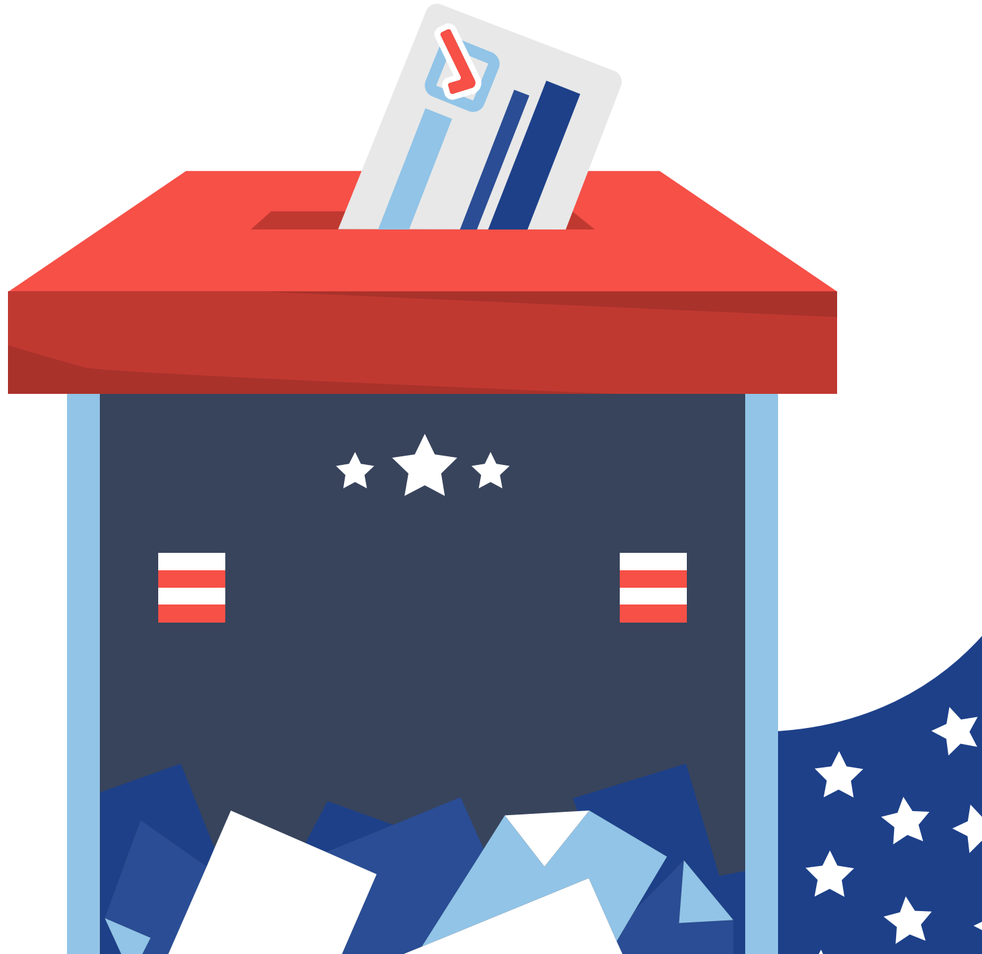


Merged df:

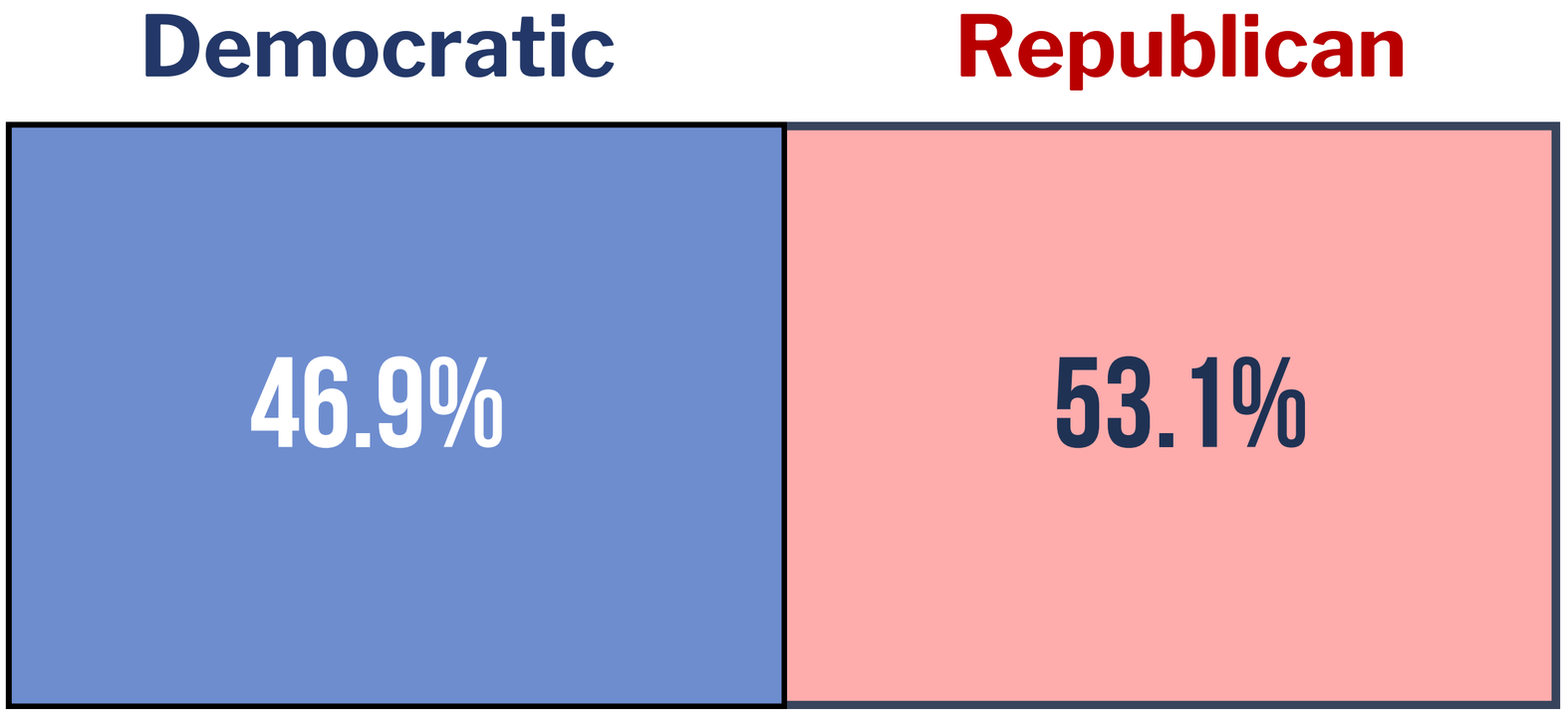
Counties with **no** demographic information

Q: CAN WE USE DEMOGRAPHIC FACTORS TO PREDICT VOTER TENDENCIES IN U.S. COUNTIES?

OVERVIEW

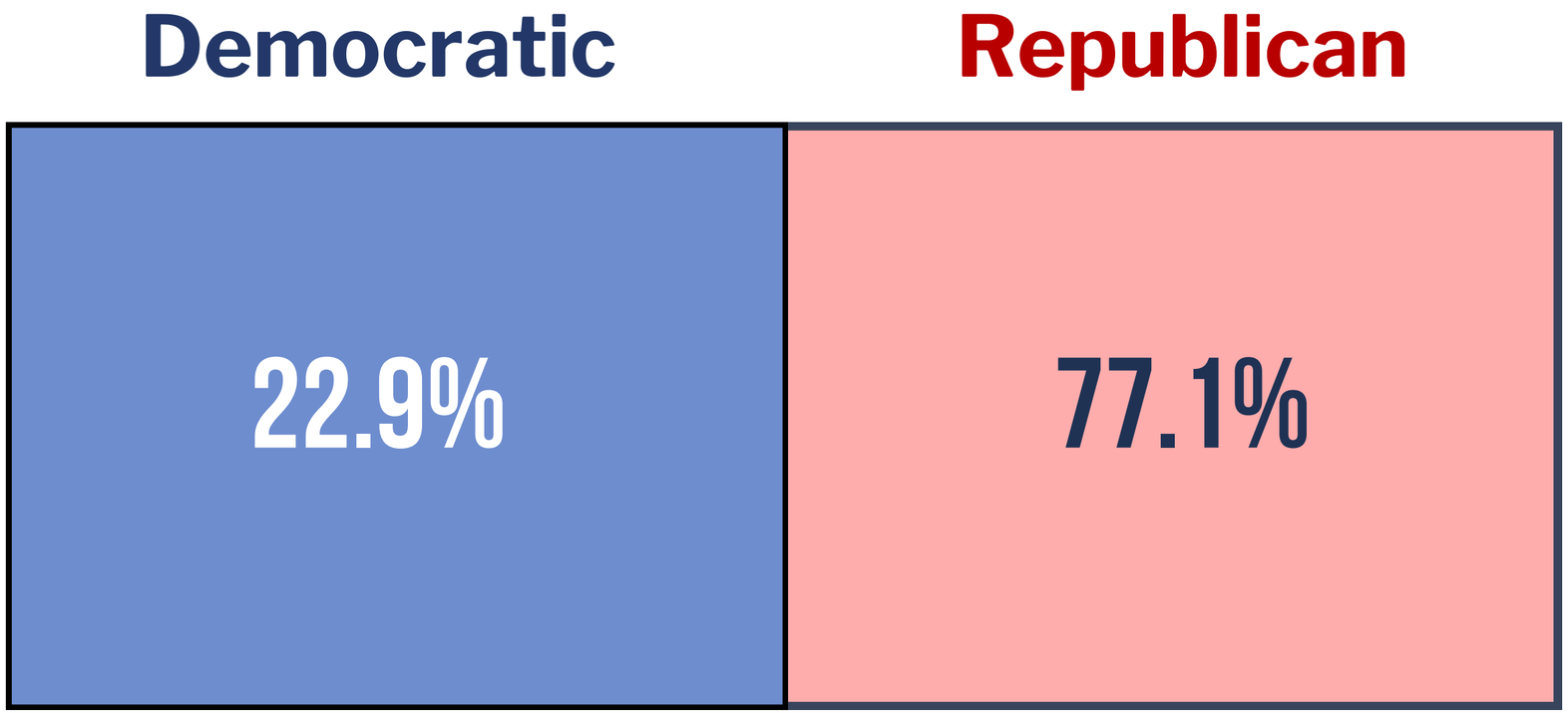


WINNING PARTY - OVERALL VOTES



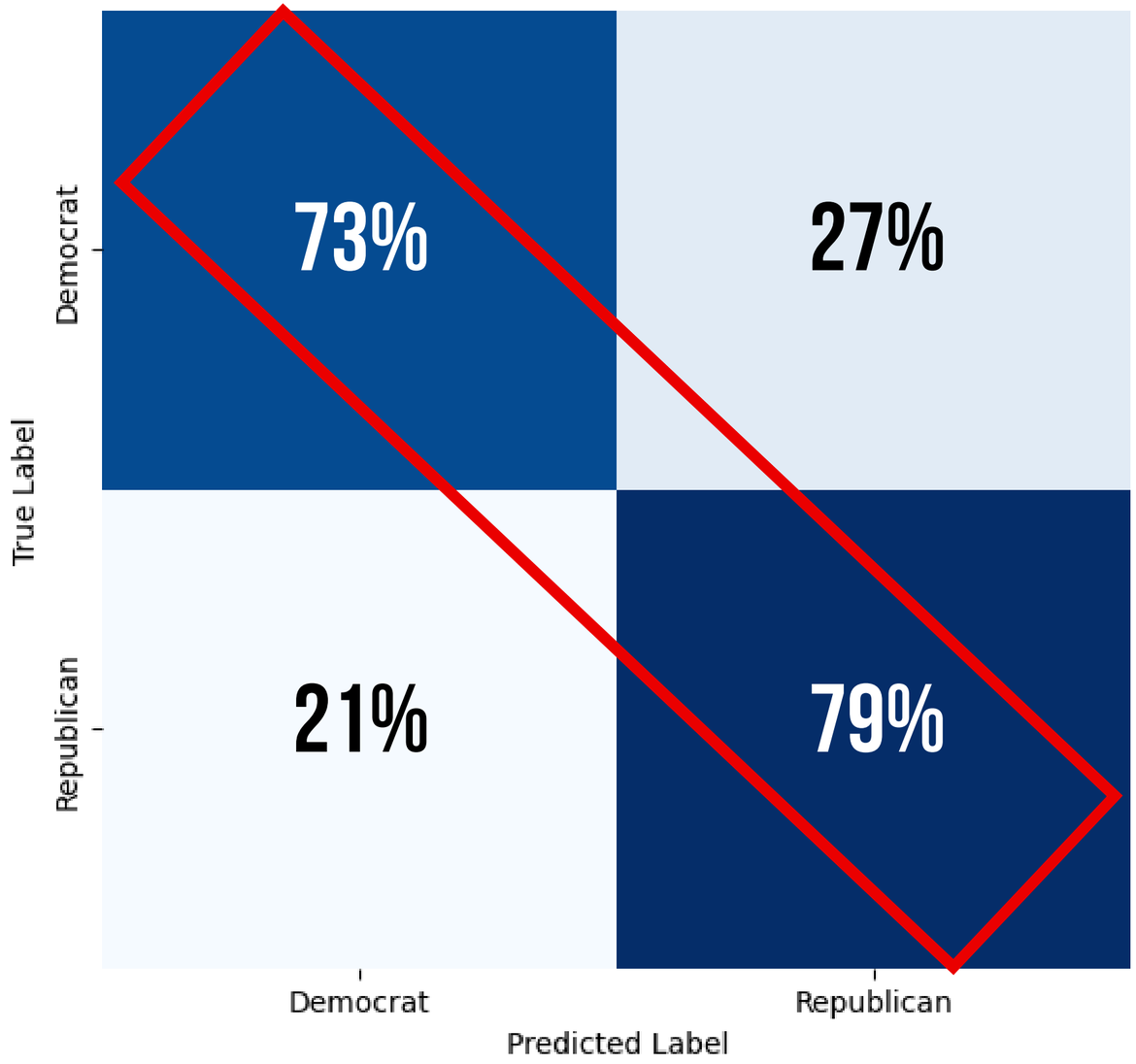
Total number of votes

WINNING PARTY SHARE - COUNTY LEVEL.



Percent of counties that voted Democrat or Republican

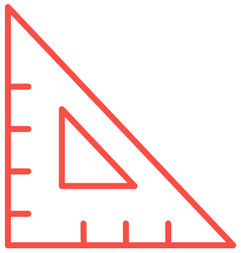
CONFUSION MATRIX NORMALIZED



METRICS

ACCURACY

77.5%



HIGH PRECISION

90%

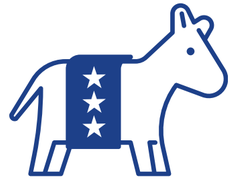
GOOD RECALL

79%

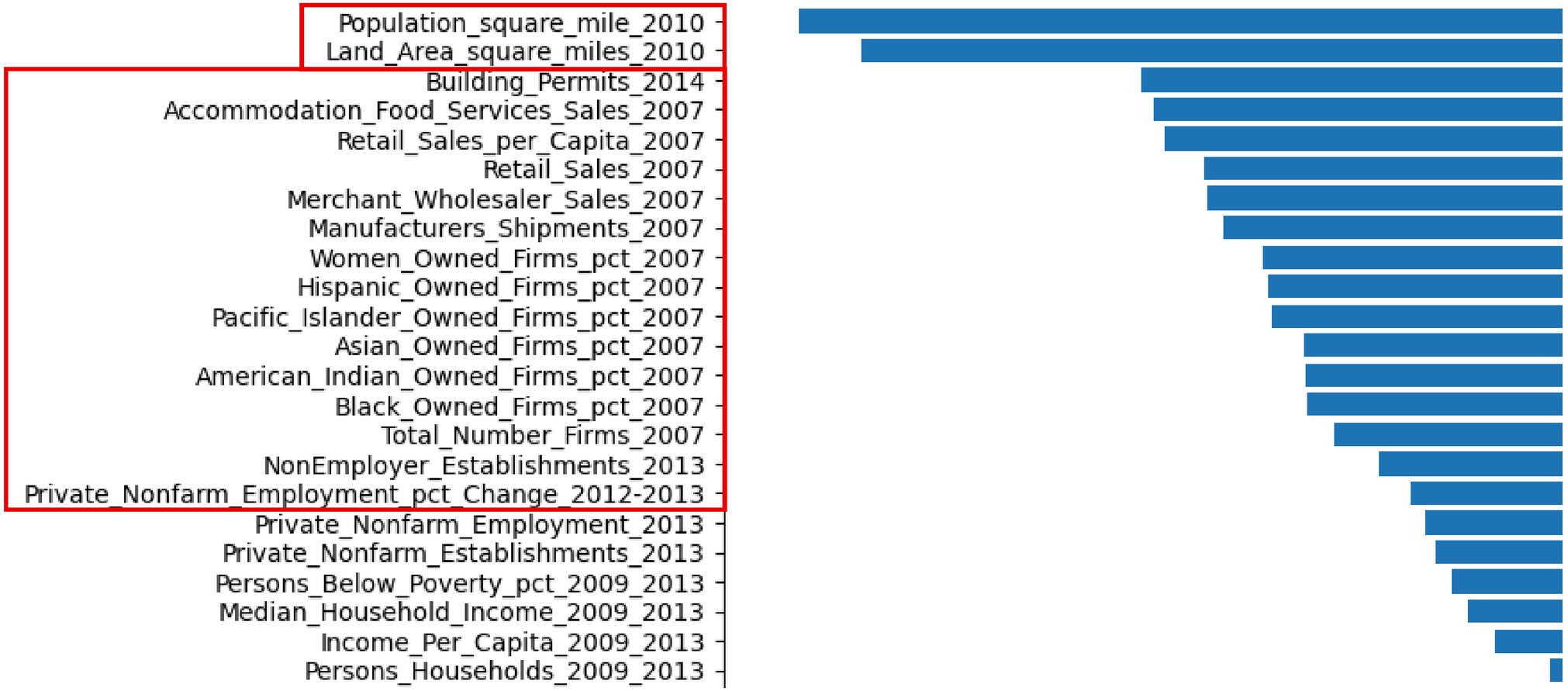
AUROC

82%

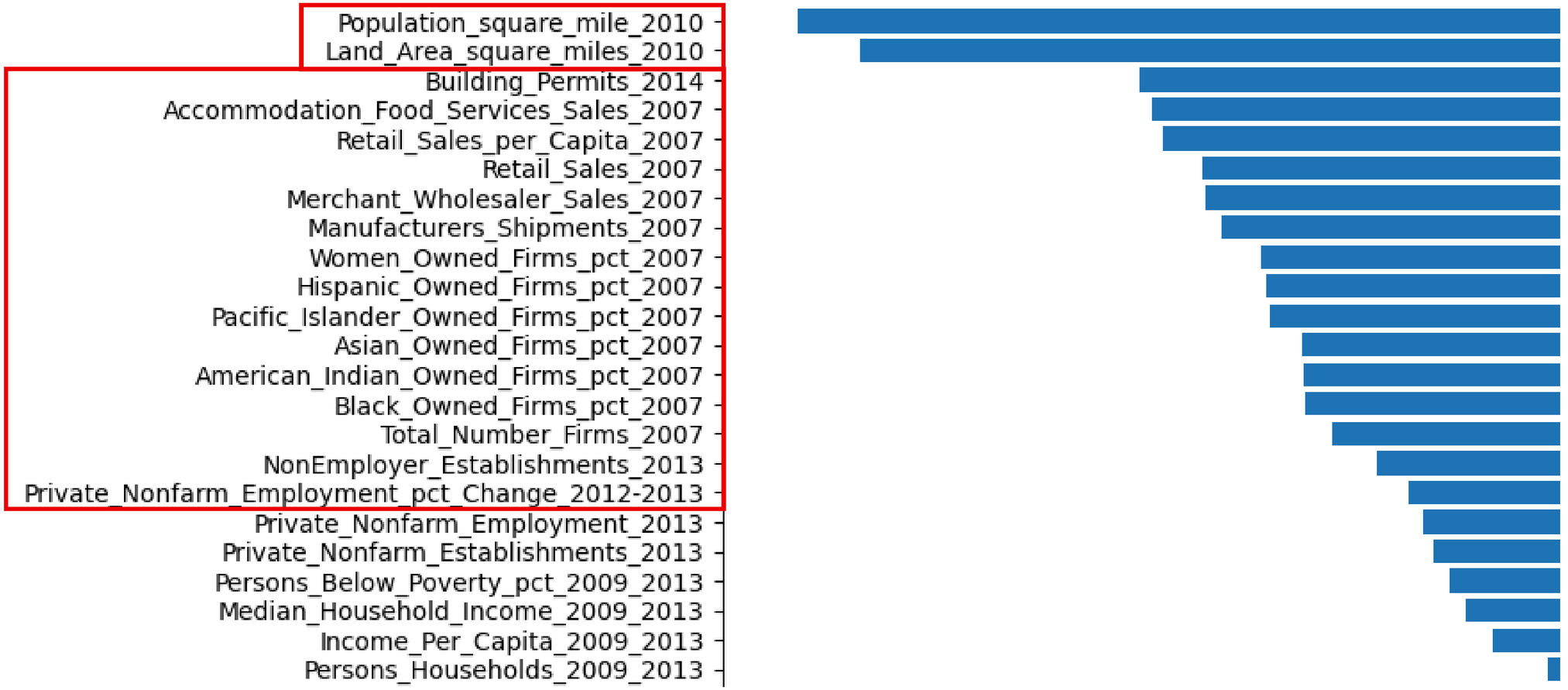
DEMOCRATIC



COEFFICIENTS - DEMOCRATIC PARTY

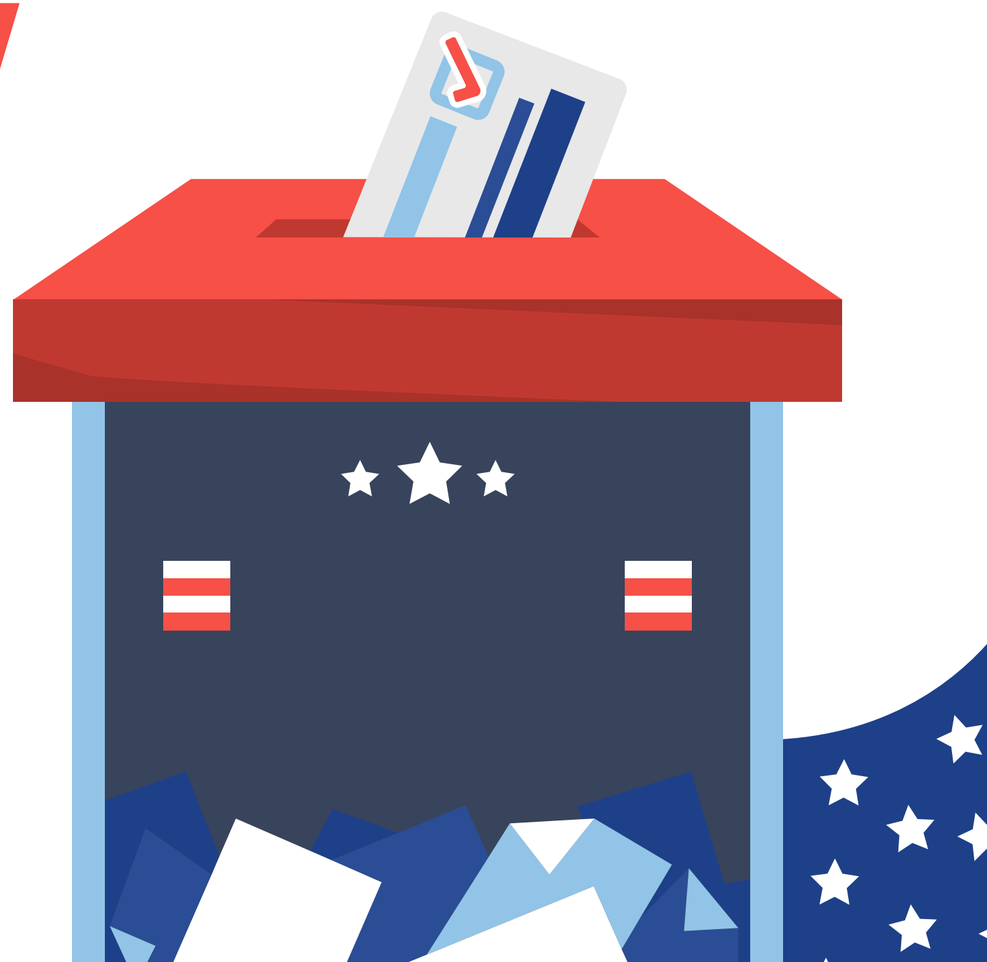


COEFFICIENTS - DEMOCRATIC PARTY

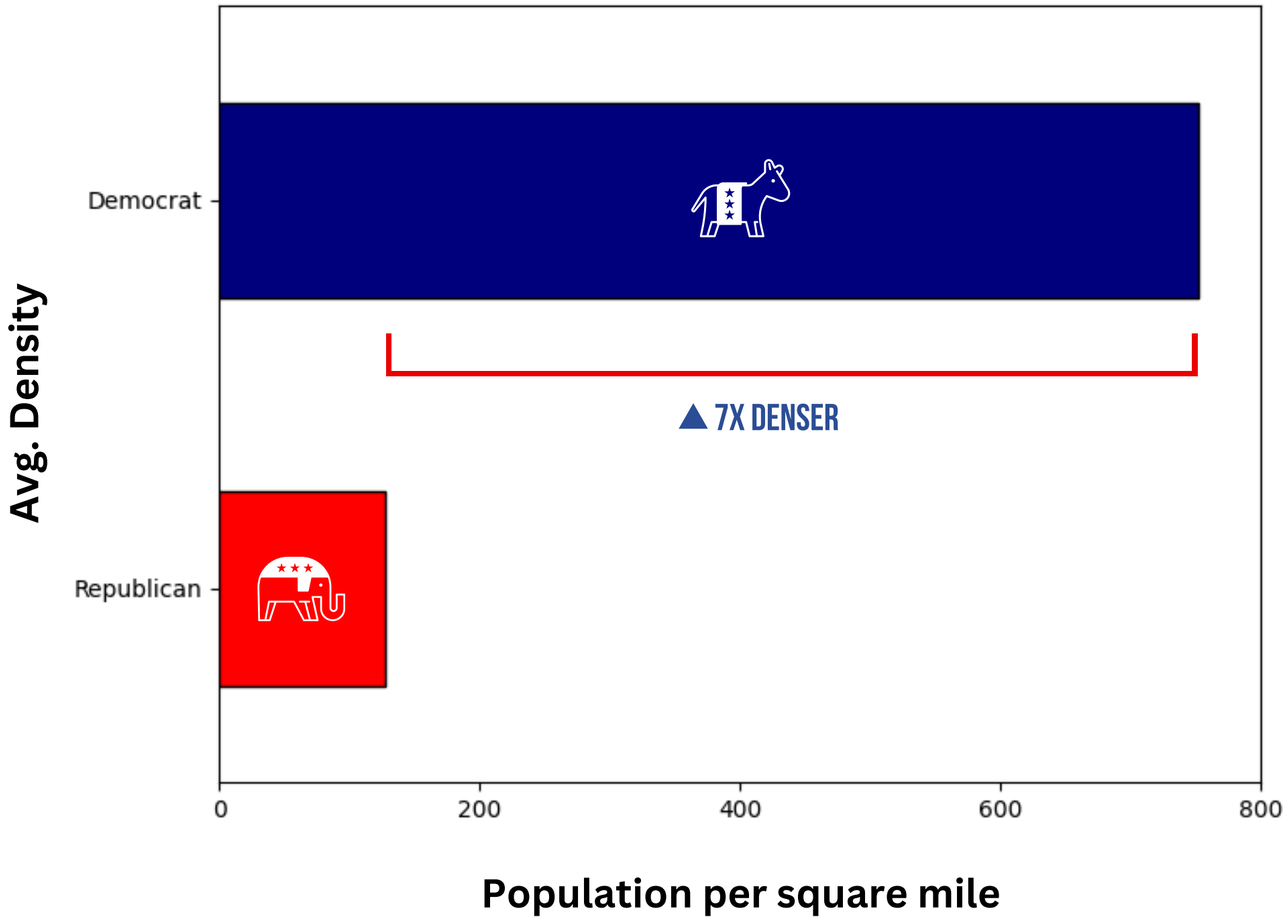


POPULATION DENSITY

How does county population density influence party vote partiality?



AVERAGE COUNTY POP. DENSITY BY WINNING PARTY



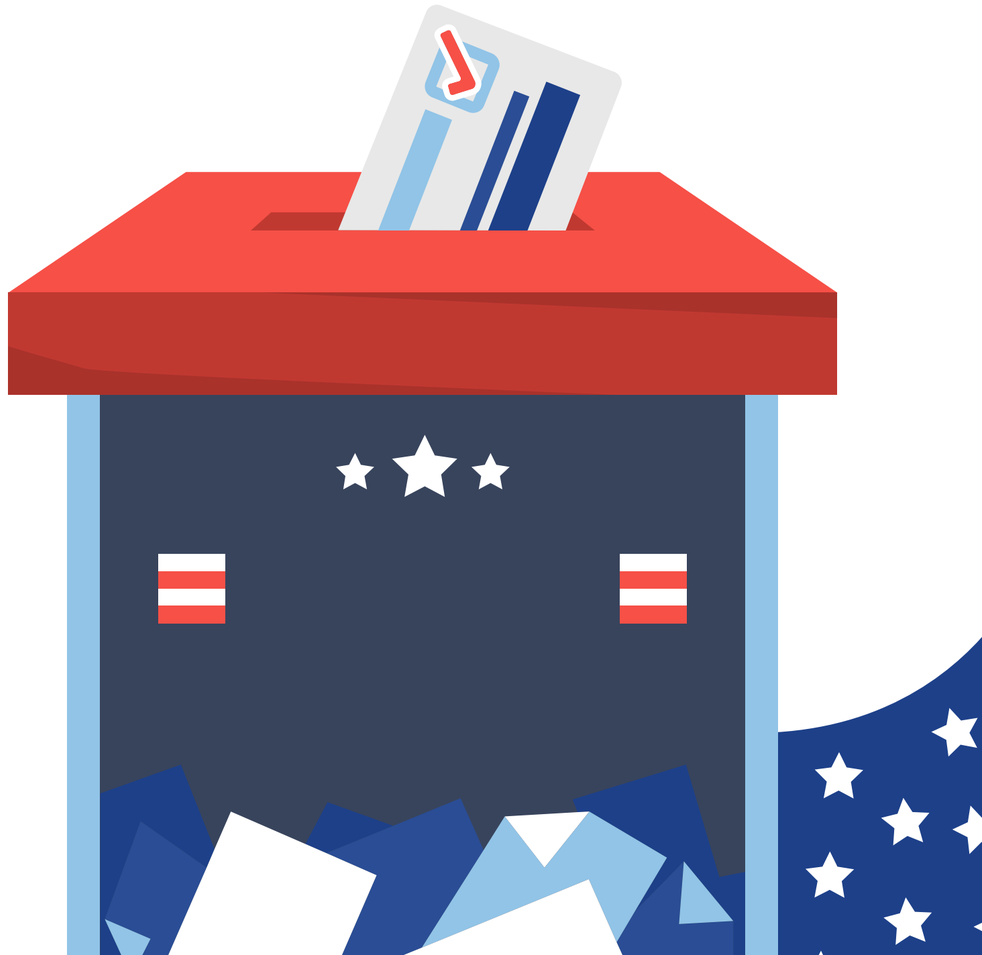
2016 Political Context:

- Historical Robust Rural Support
- 2X Rep Rural Support (‘10)
- Strong Urban Dem. Support (65%)
- Suburban = Divided Split

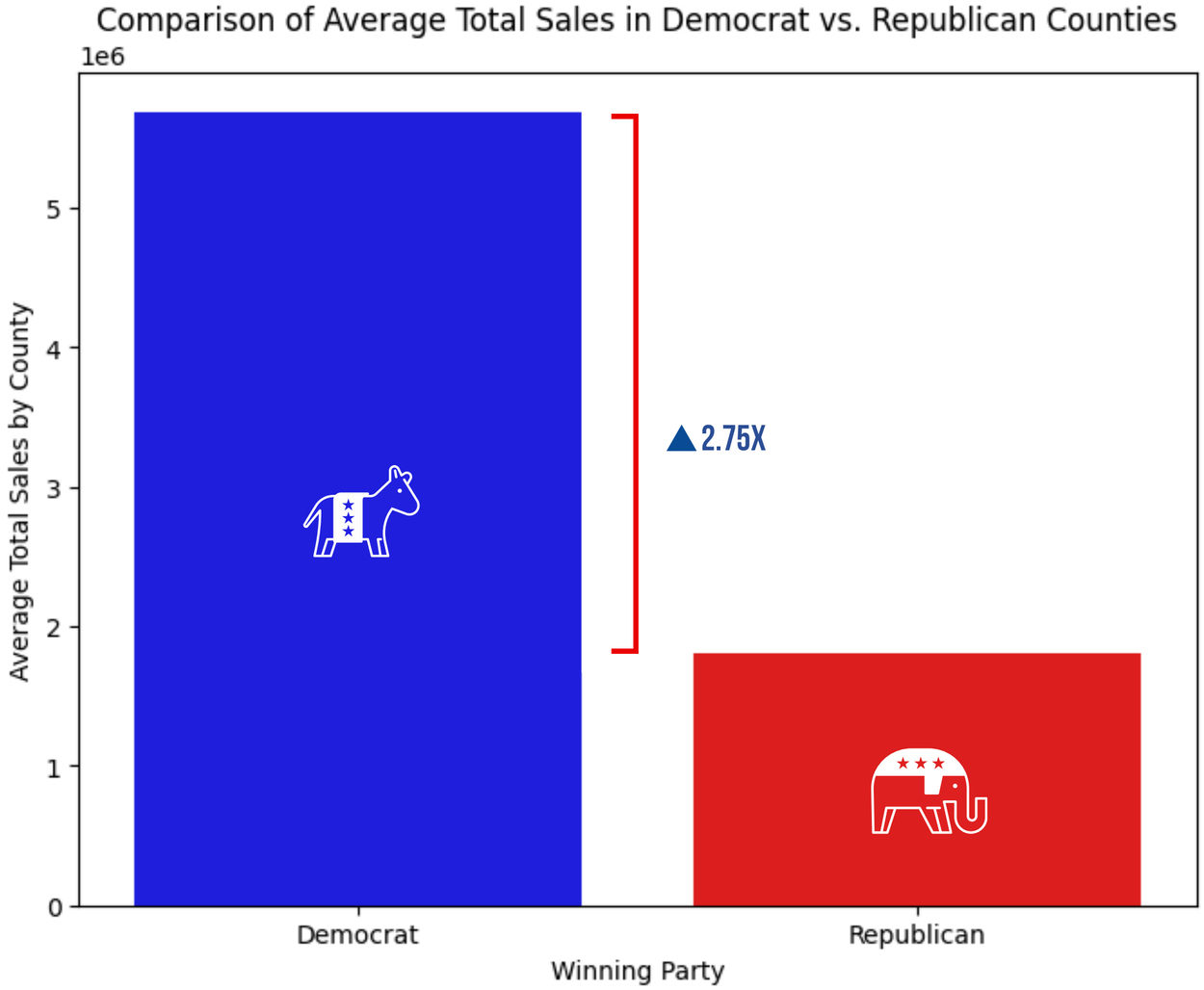
(Pew Research Center, 2024)

COUNTY ECONOMICS

How does a county's total business sales influence vote shares?



AVG. COUNTY-WIDE SALES



2016 Political Context:

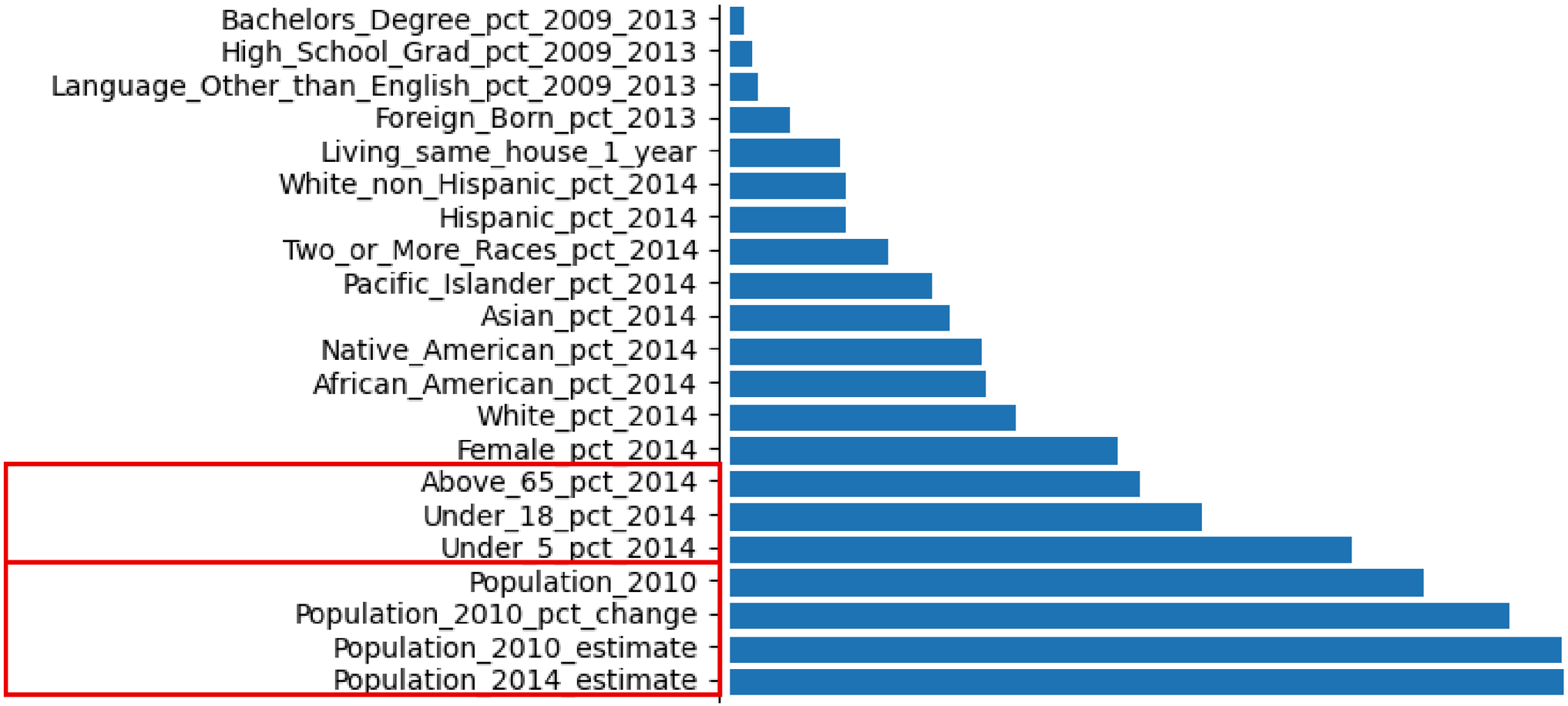
Clinton vs. Trump (extra info.)

- 2584 Counties Won = 36%
- 472 Counties Won = 64%

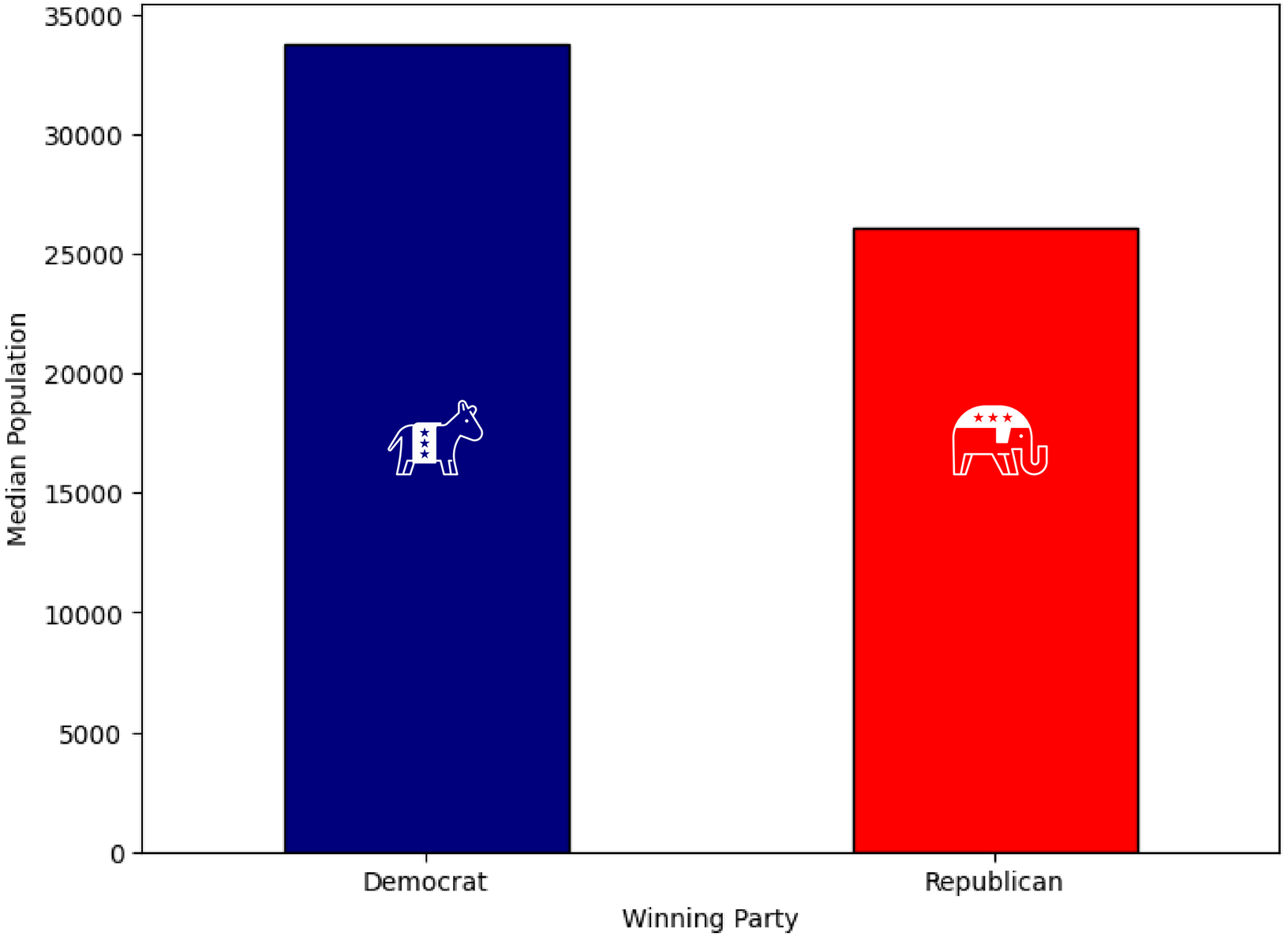
(Muro, Whiton, & Maxim, 2020)

REPUBLICAN

COEFFICIENTS - REPUBLICAN PARTY



MEDIAN POPULATION BY WINNING PARTY



WHY IS THIS THE CASE?

DATA IMBALANCE

Multiple small **Republican counties** may **skew model** to associate pop. as **(+) correlated** with Republican outcomes.

OTHER VARIABLES

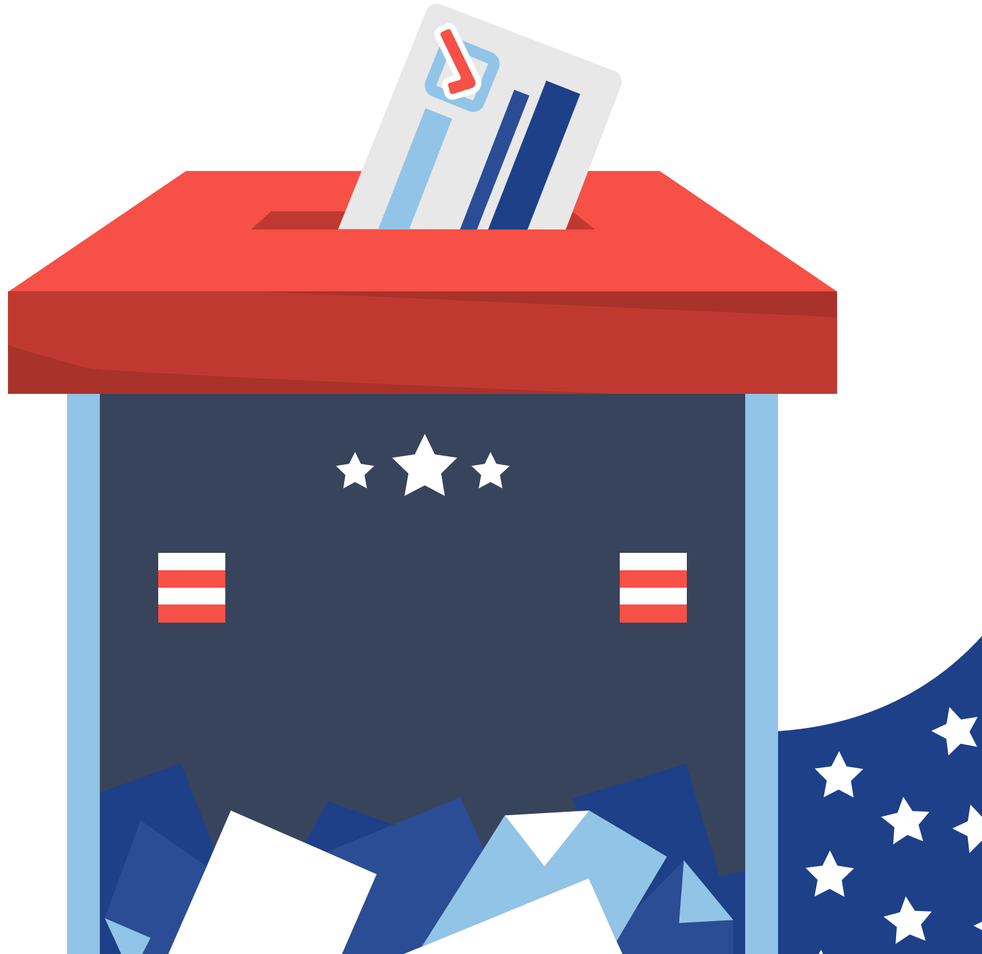
When controlling for race, poverty, age groups, etc., the **partial relationship with pop. can flip sign.**

POP. VS. DENSITY

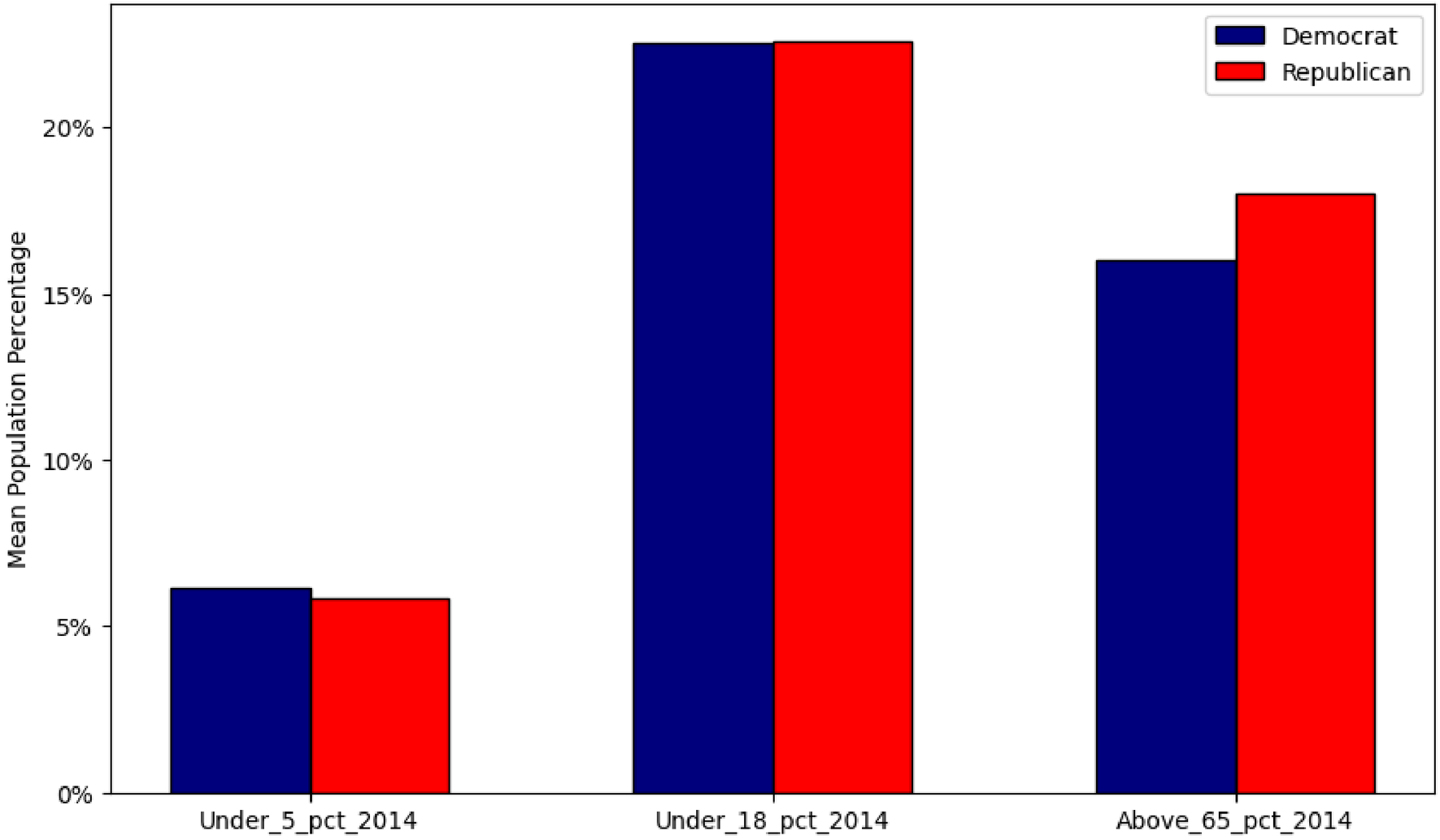
Large counties **aren't** necessarily **dense** (urban).

HOUSEHOLD AGE

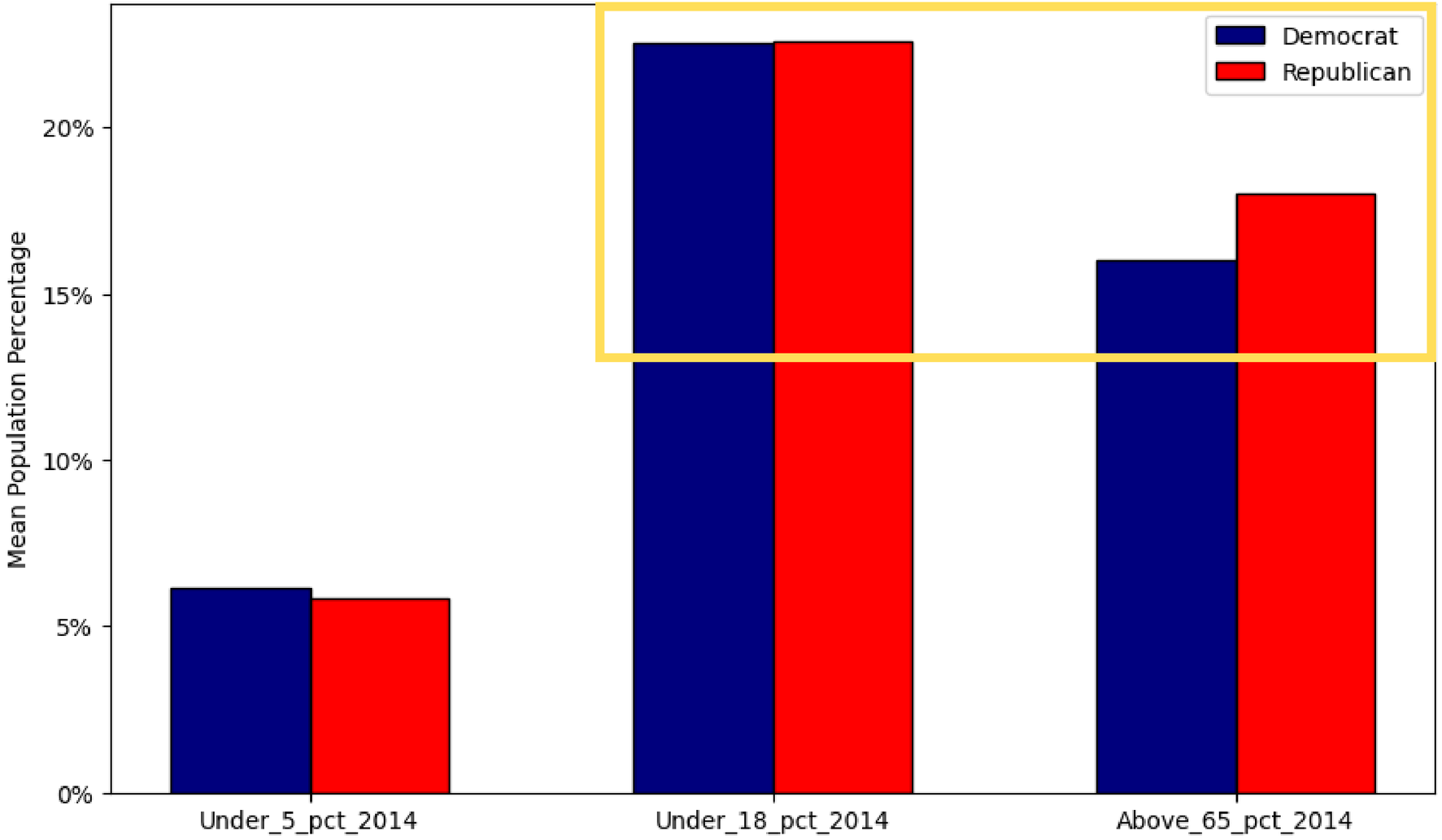
How do age ranges within a household affect voter tendencies?



HOW DO HOUSEHOLD AGE GROUPS CHANGE VOTING?

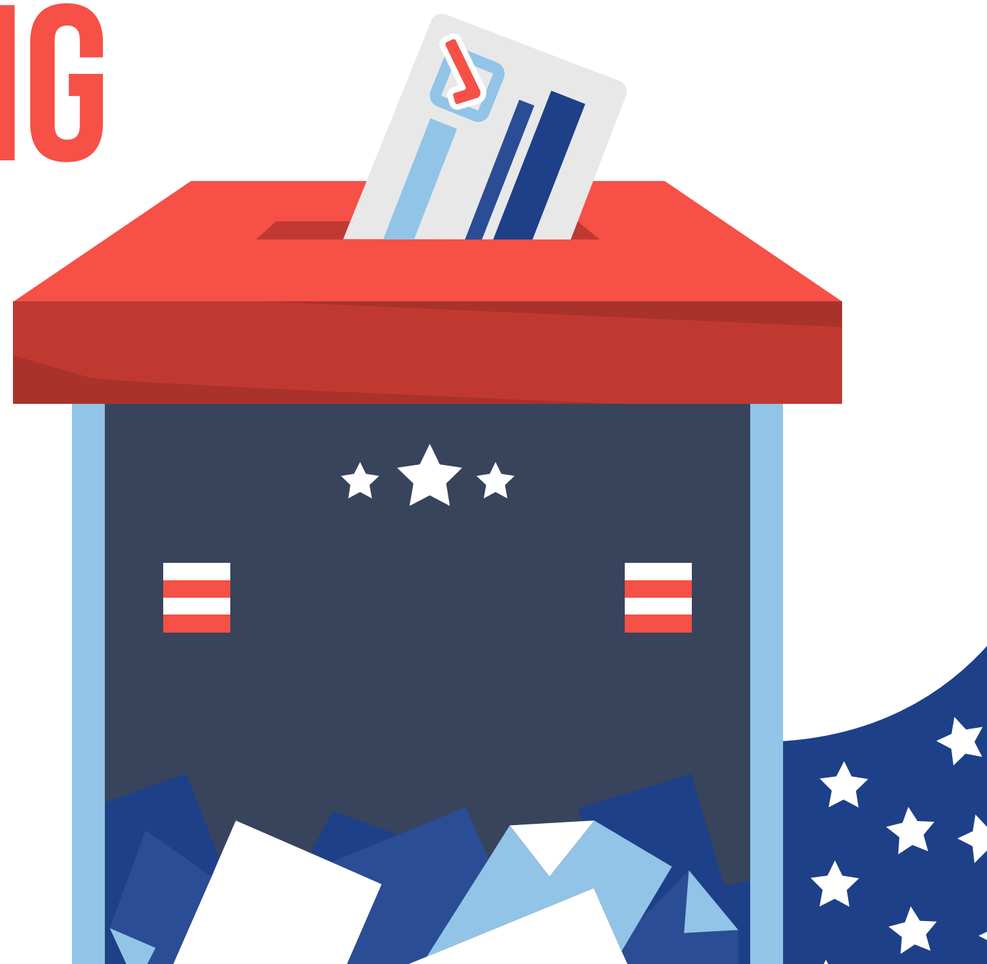


HOW DO HOUSEHOLD AGE GROUPS CHANGE VOTING?



FEATURE ENGINEERING

Can we fine-tune the model for better performance?



VARIANCE INFLATION FACTOR (VIF)

Purpose: Evaluates level of correlation between predictor (feature) in a regression model

A **high ▲VIF** means a feature is **redundant** because it **shares too much info**. with other **features**

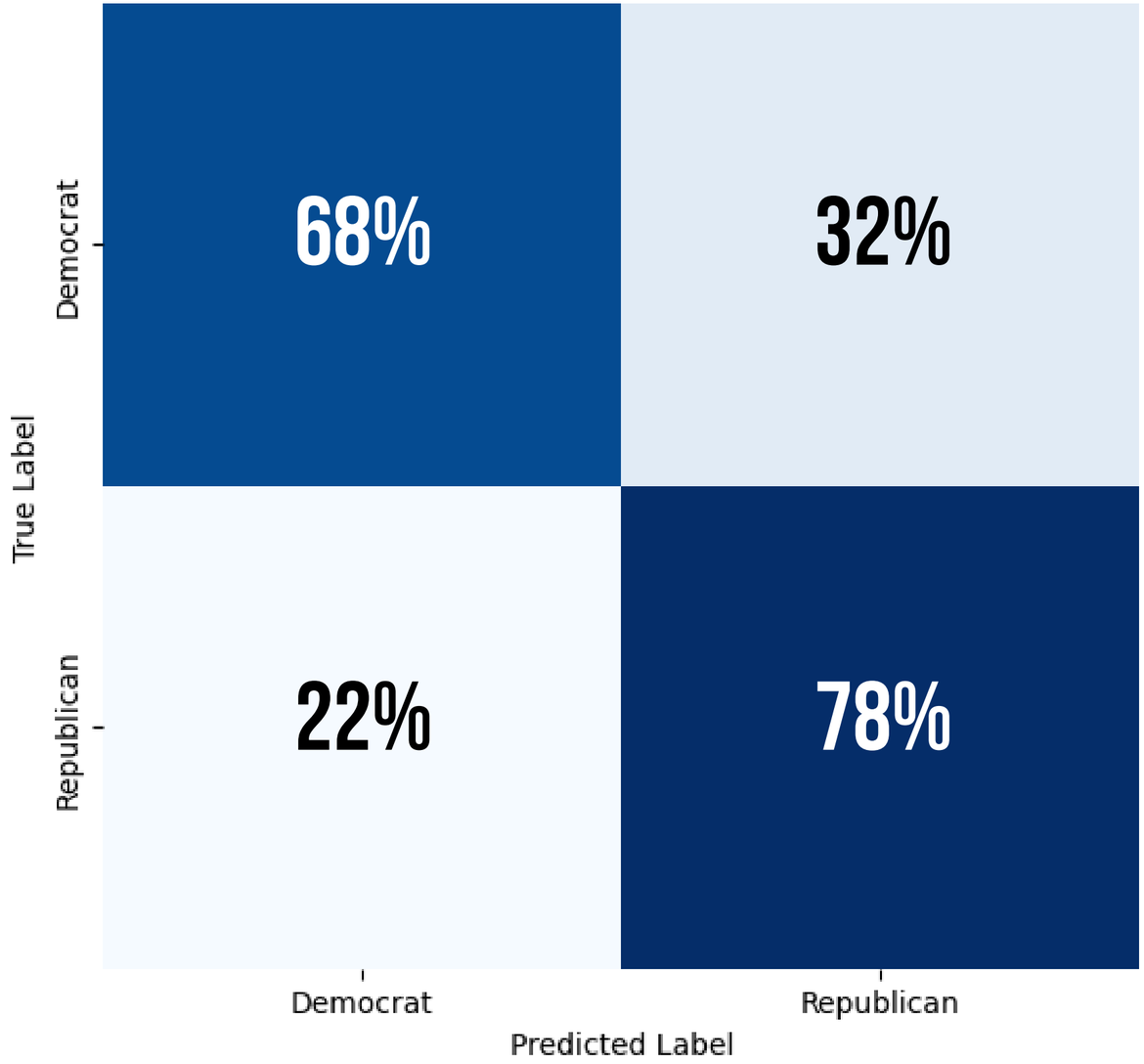
Can make the model unstable.

VARIANCE INFLATION FACTOR (VIF)

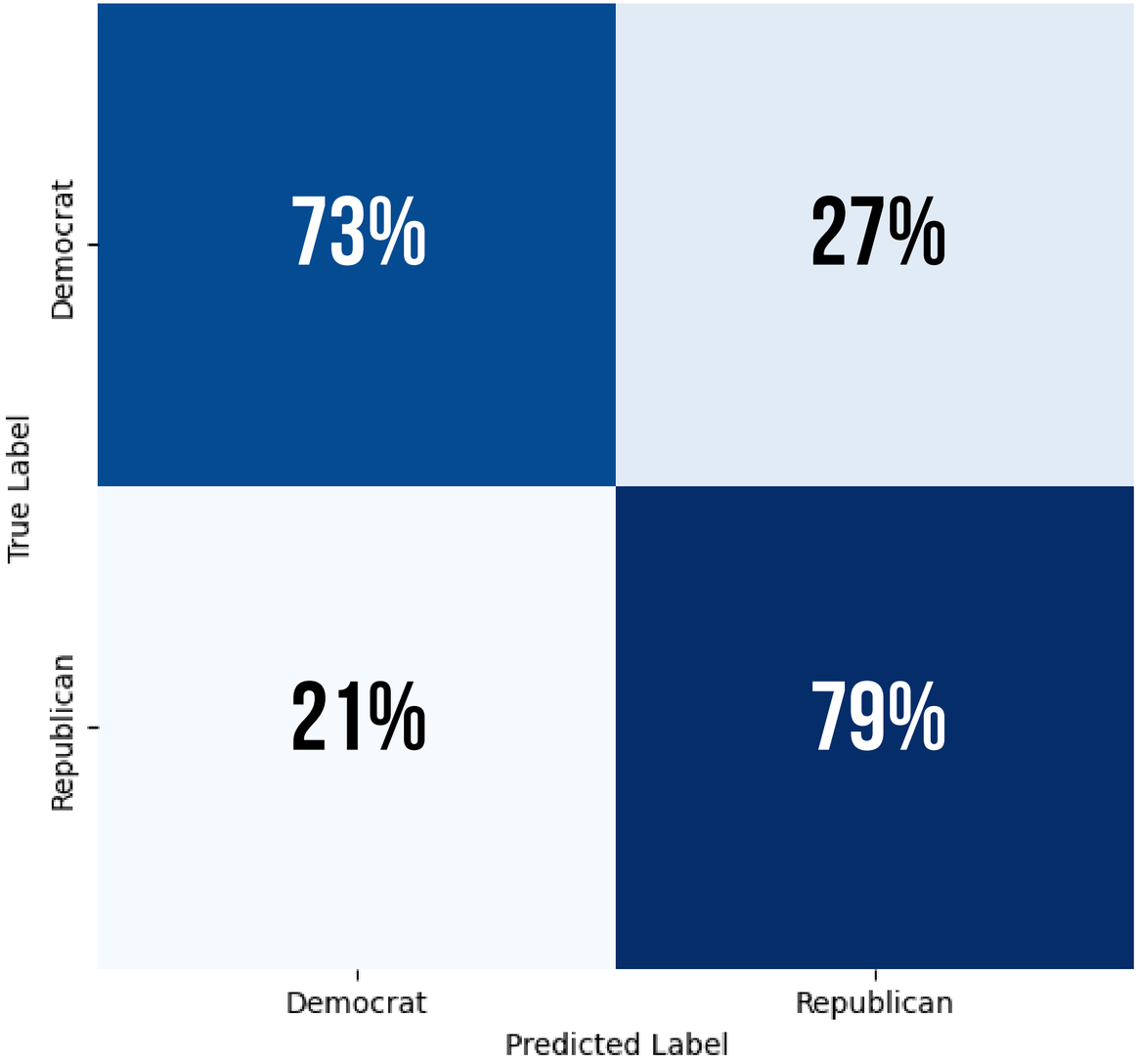
County_Size_Index	Population 2010	Households 2009_2013	Population 2014 estimate	Housing_Units 2014	...
Youth_pct_2014	Under_5_pct_2014	Under_18_pct_2014			
Income_Index	Median_Household_Income 2009_2013	Income_Per_Capita_2009_2013			

Correlation threshold = 0.8

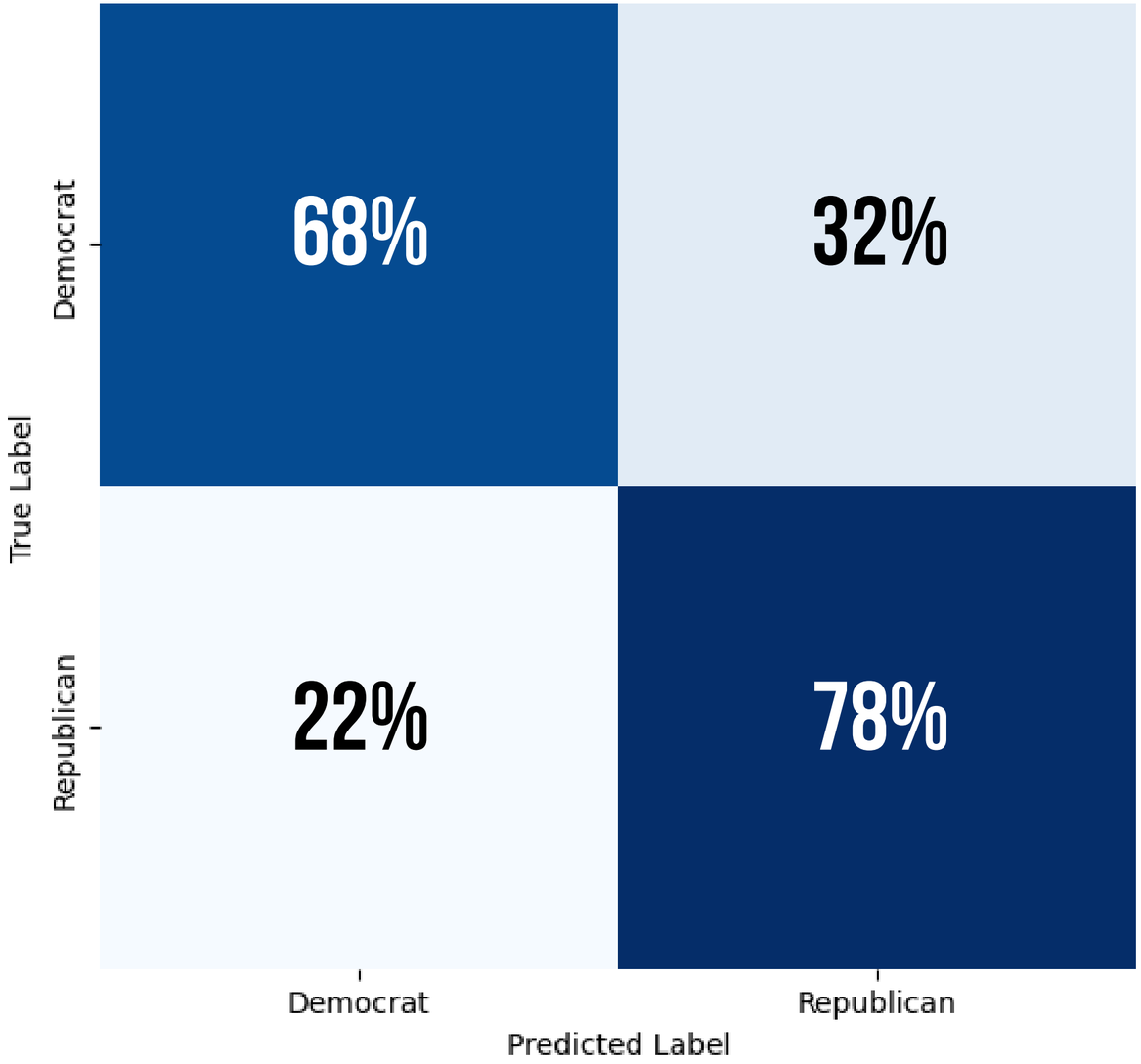
CONFUSION MATRIX



OLD


73%
27%
21%
79%
Democrat
Republican
True Label
Democrat
Republican
Predicted Label

NEW


68%
32%
22%
78%
Democrat
Republican
True Label
Democrat
Republican
Predicted Label

SUMMARY

- **Data Limitations:**
  - 35 % missing values
  - Data Imbalance: Republican-leaning
- Dense counties & Total Sales = More Democratic Tendencies
- Household Age Groups Influence Voting Dynamics
- Feature Engineering = We Tried - Accuracy Levels Dropped

THANK YOU

Questions?