



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

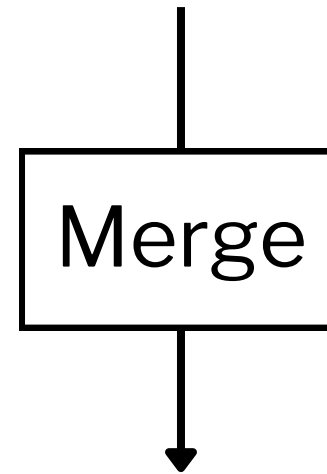
Democratic Votes	3%
Republican Votes	15%
Democratic Votes (Fraction)	3%
Republican Votes (Fraction)	15%
Fips	33.5%
...	33.5%

- 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.



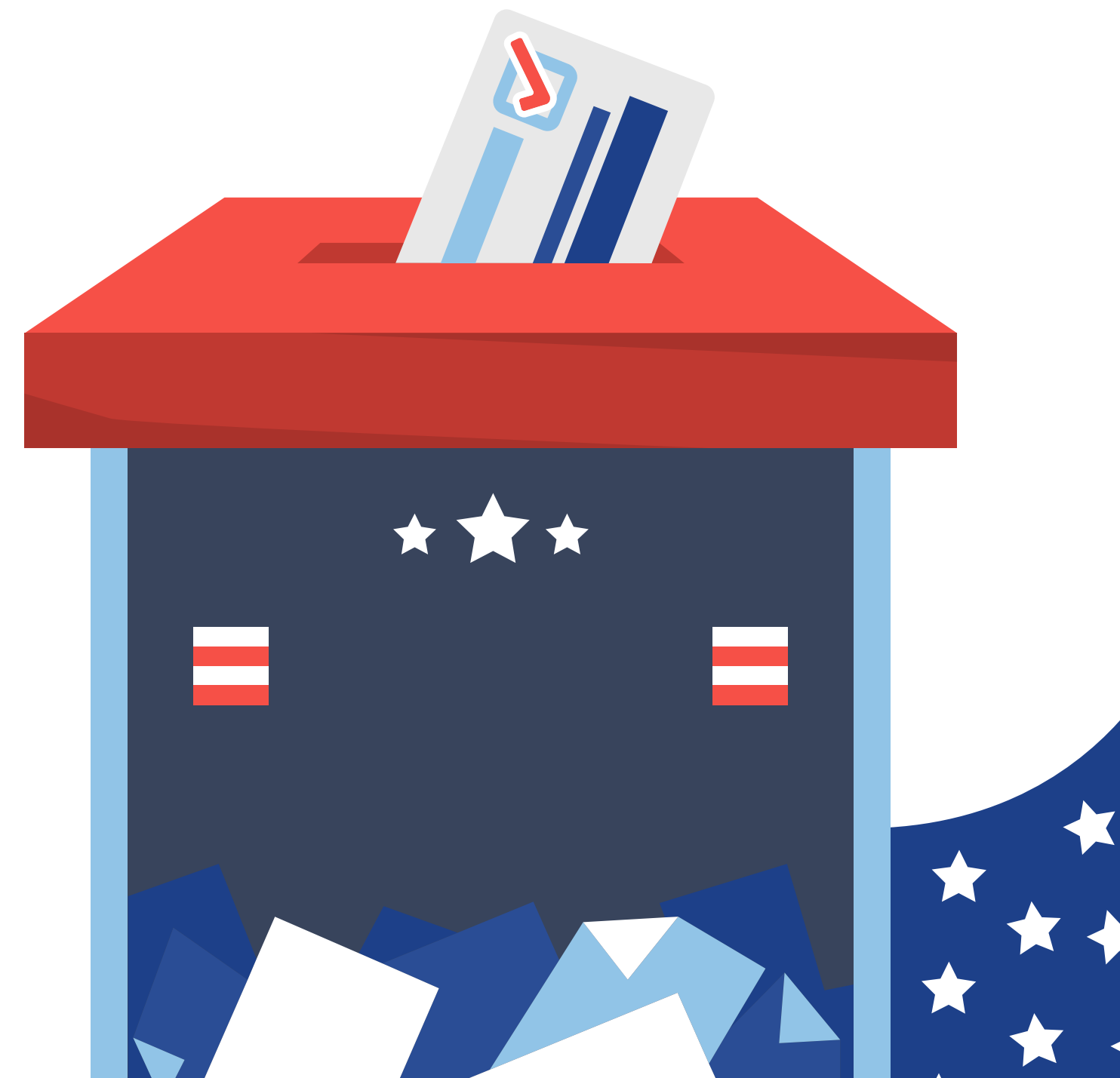
Merged df:

Counties with **no** demographic information



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

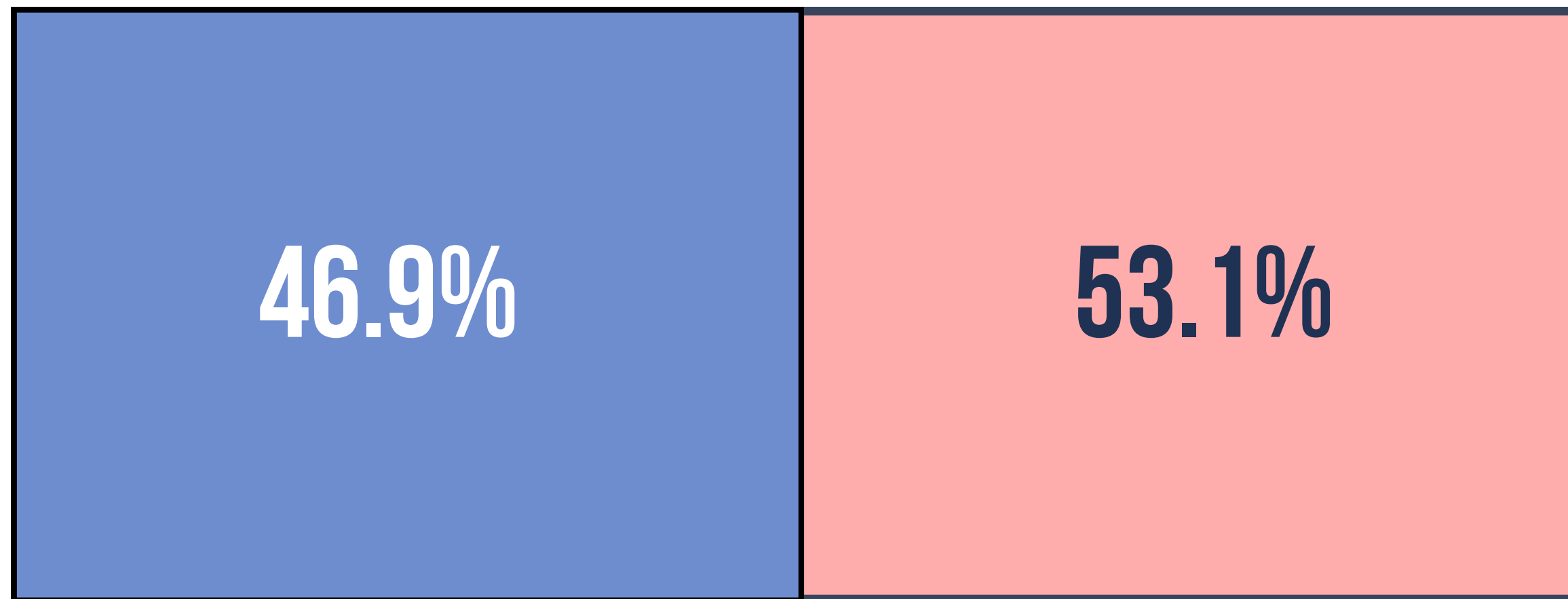
OVERVIEW



WINNING PARTY - OVERALL VOTES

Democratic

Republican

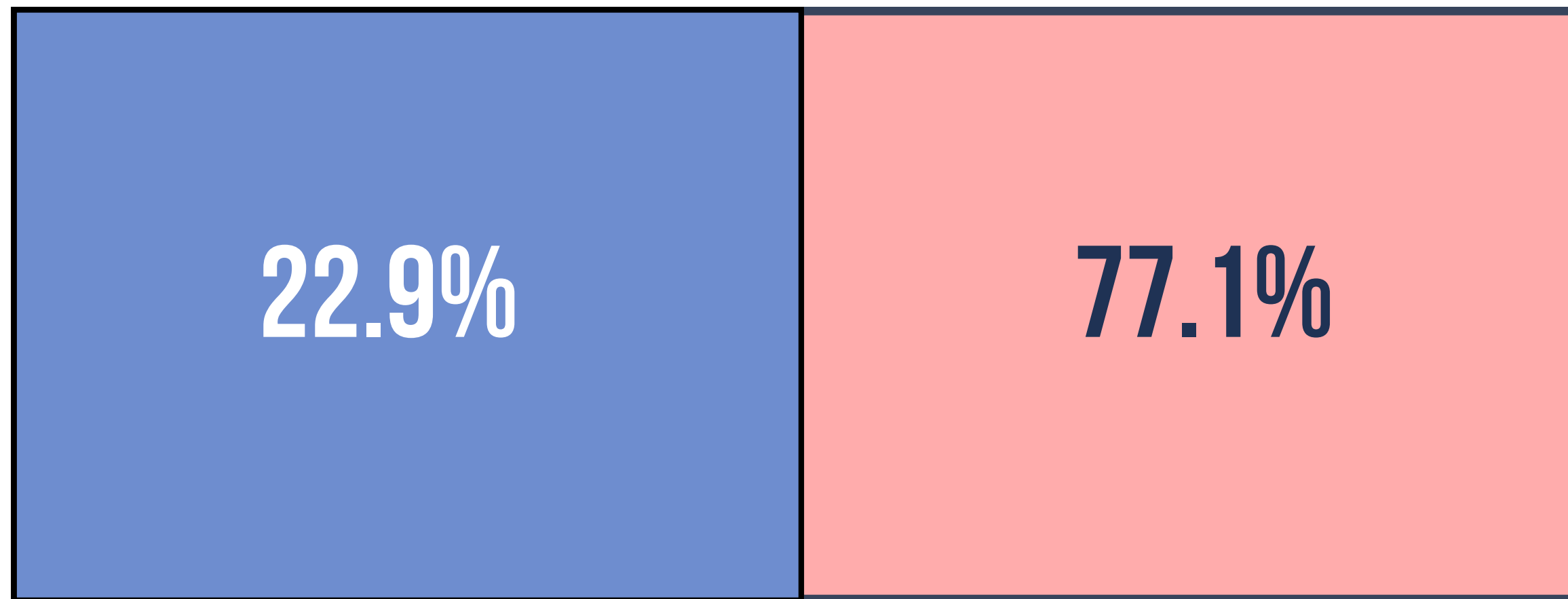


Total number of votes

WINNING PARTY SHARE - COUNTY LVL.

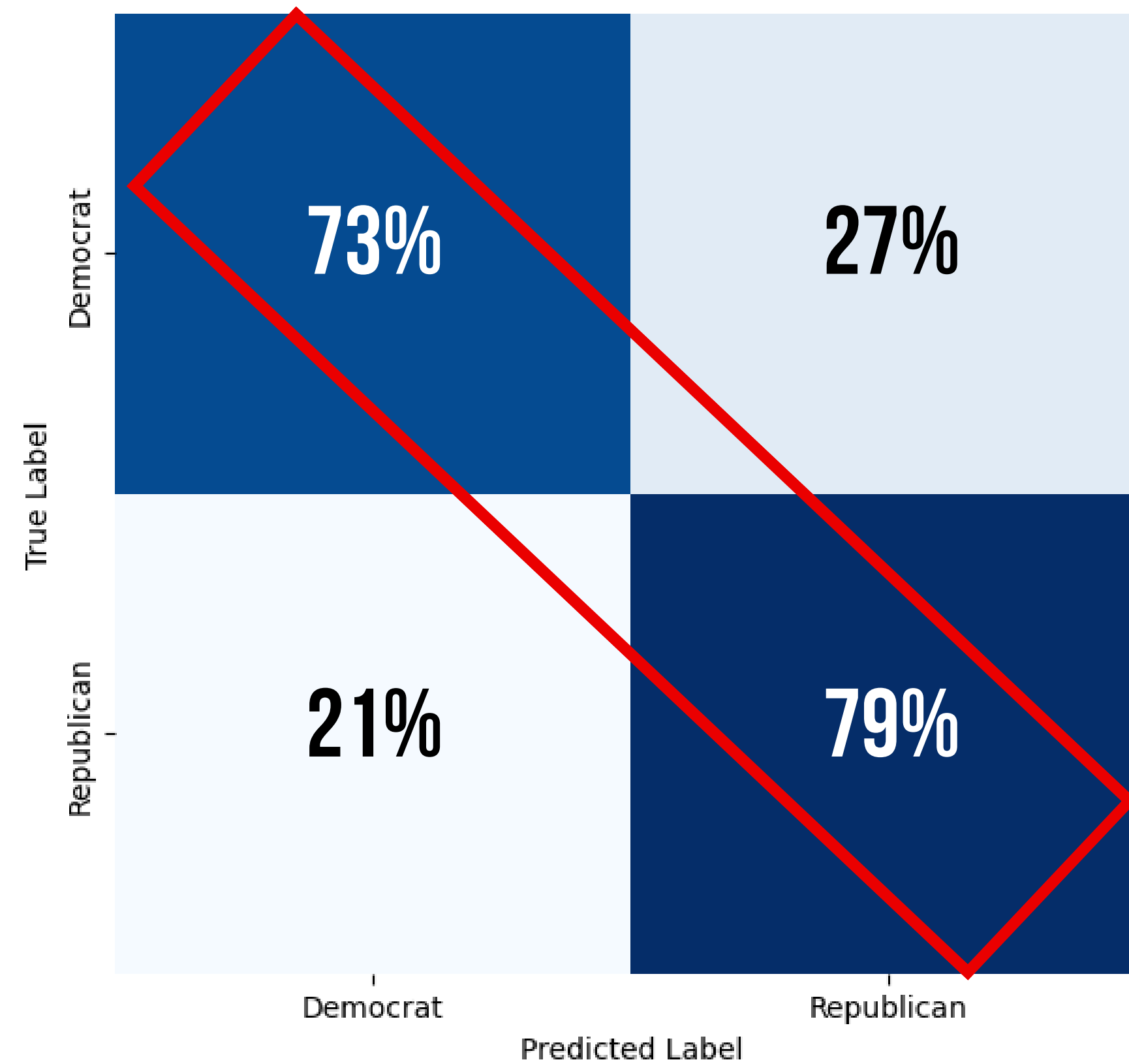
Democratic

Republican



Percent of counties that voted Democrat or Republican

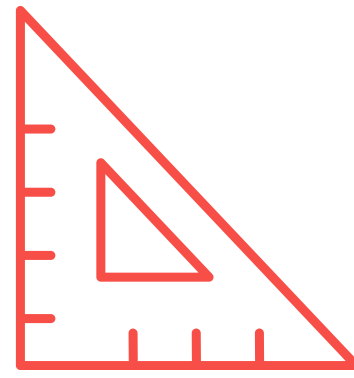
CONFUSION MATRIX NORMALIZED

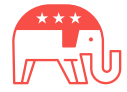


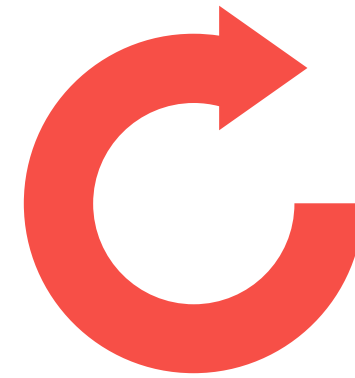
METRICS

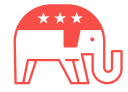


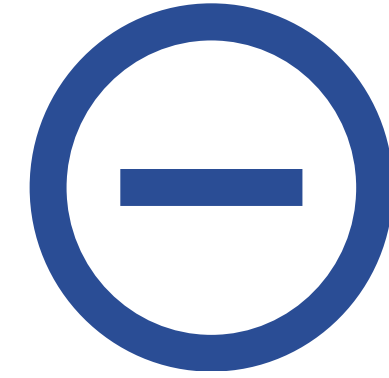
ACCURACY
77.5%



HIGH PRECISION
90% 

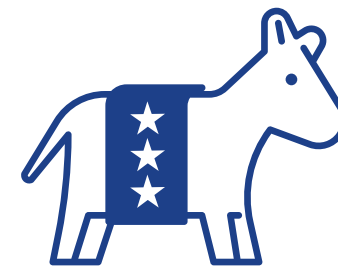


GOOD RECALL
79% 



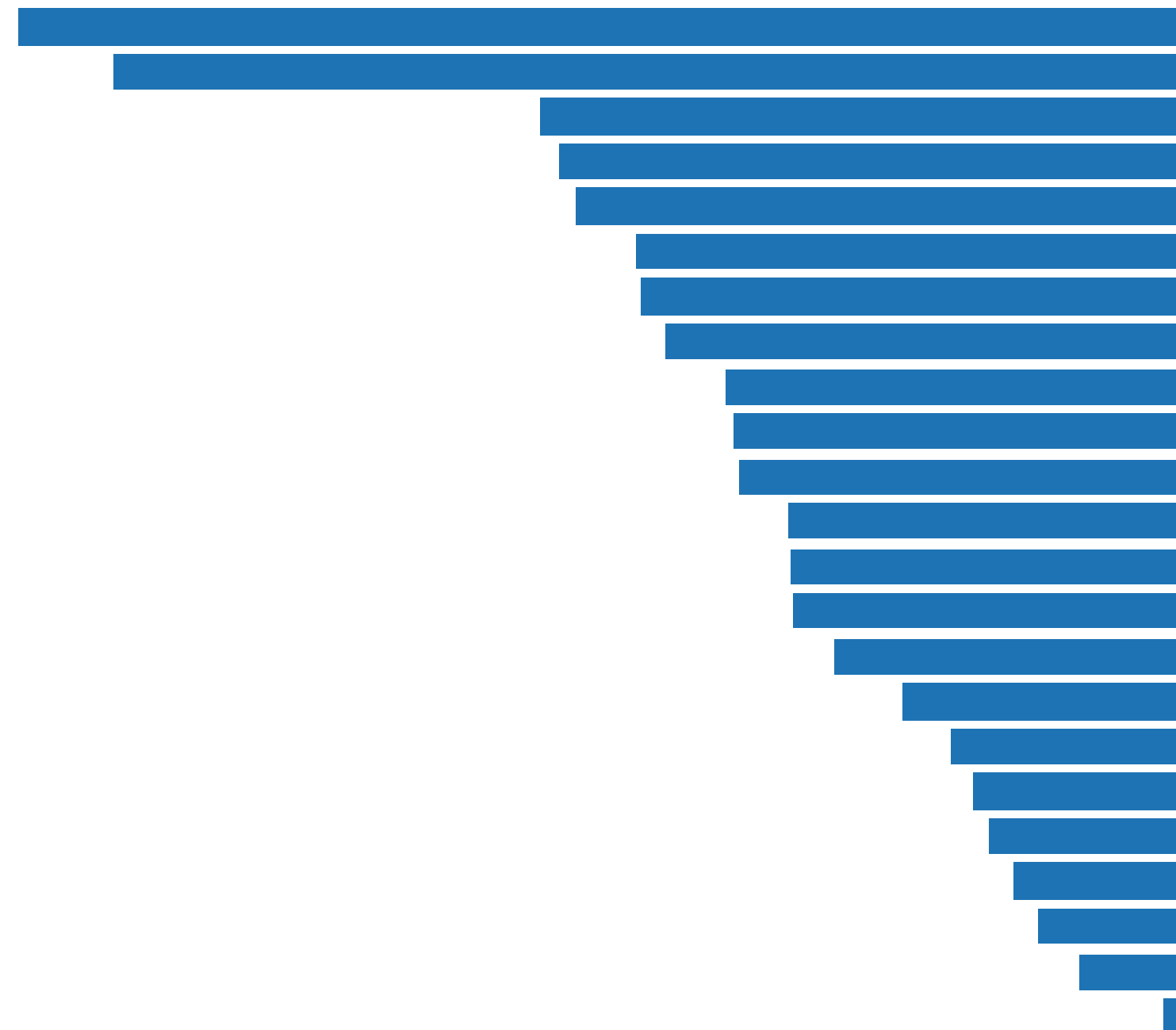
AUROC
82%

DEMOCRATIC



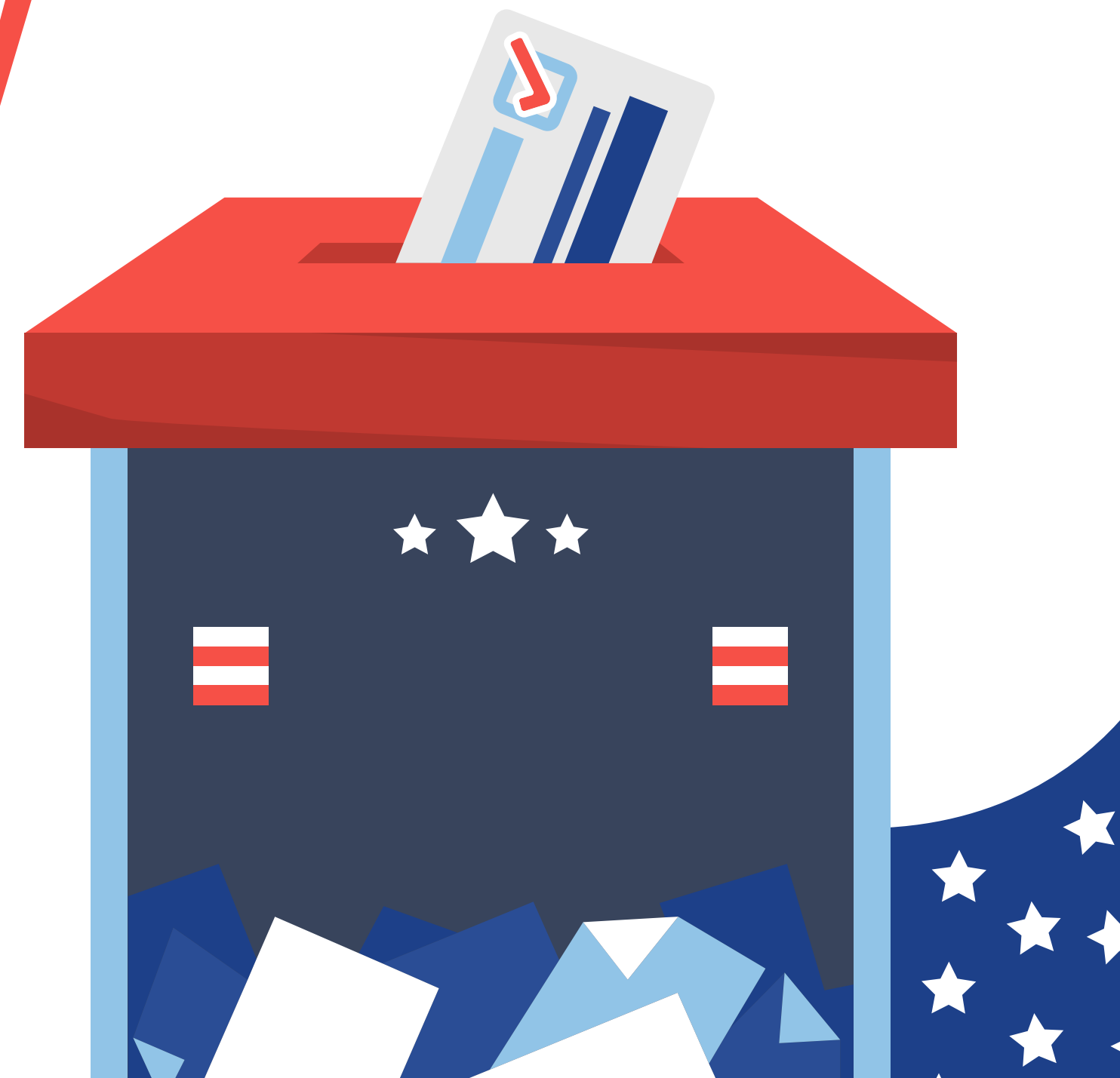
COEFFICIENTS - DEMOCRATIC PARTY

Population_square_mile_2010	-
Land_Area_square_miles_2010	-
Building_Permits_2014	-
Accommodation_Food_Services_Sales_2007	-
Retail_Sales_per_Capita_2007	-
Retail_Sales_2007	-
Merchant_Wholesaler_Sales_2007	-
Manufacturers_Shipments_2007	-
Women_Owned_Firms_pct_2007	-
Hispanic_Owned_Firms_pct_2007	-
Pacific_Islander_Owned_Firms_pct_2007	-
Asian_Owned_Firms_pct_2007	-
American_Indian_Owned_Firms_pct_2007	-
Black_Owned_Firms_pct_2007	-
Total_Number_Firms_2007	-
NonEmployer_Establishments_2013	-
Private_Nonfarm_Employment_pct_Change_2012-2013	-
Private_Nonfarm_Employment_2013	-
Private_Nonfarm_Establishments_2013	-
Persons_Below_Poverty_pct_2009_2013	-
Median_Household_Income_2009_2013	-
Income_Per_Capita_2009_2013	-
Persons_Households_2009_2013	-

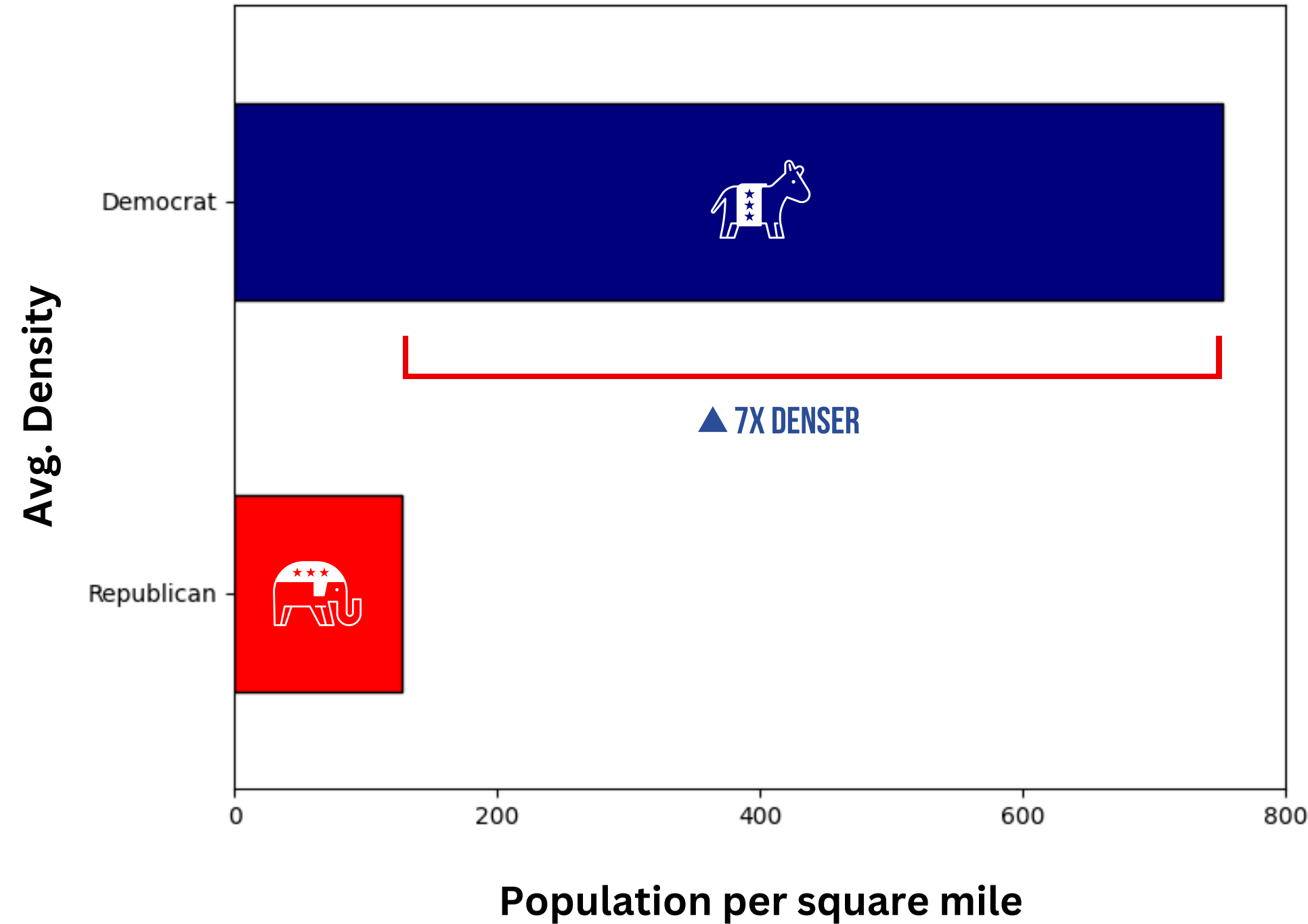


POPULATION DENSITY

How does **county population density** influence party vote partiality?



AVERAGE COUNTY POP. DENSITY BY WINNING PARTY

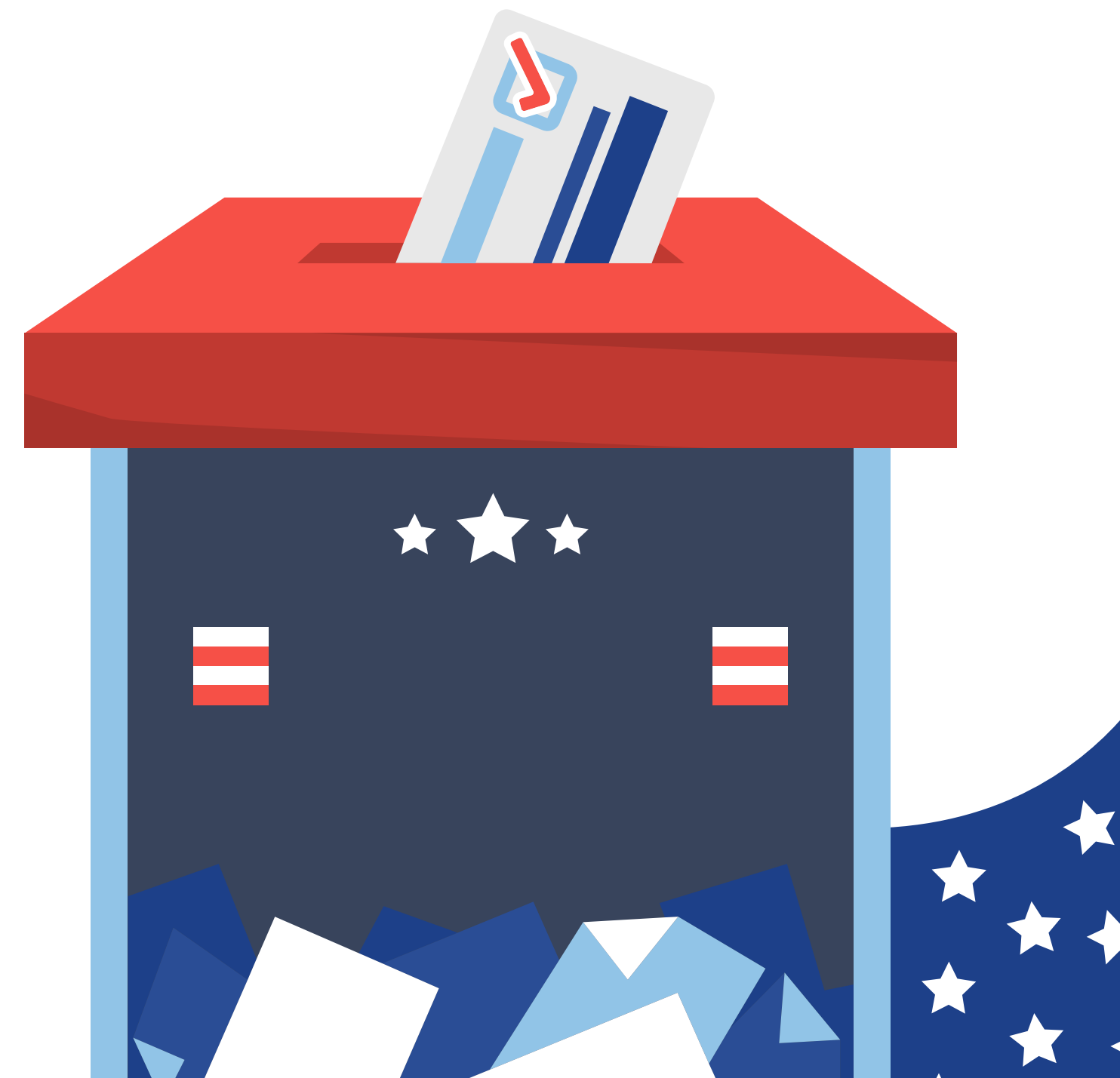


2016 Political Context:

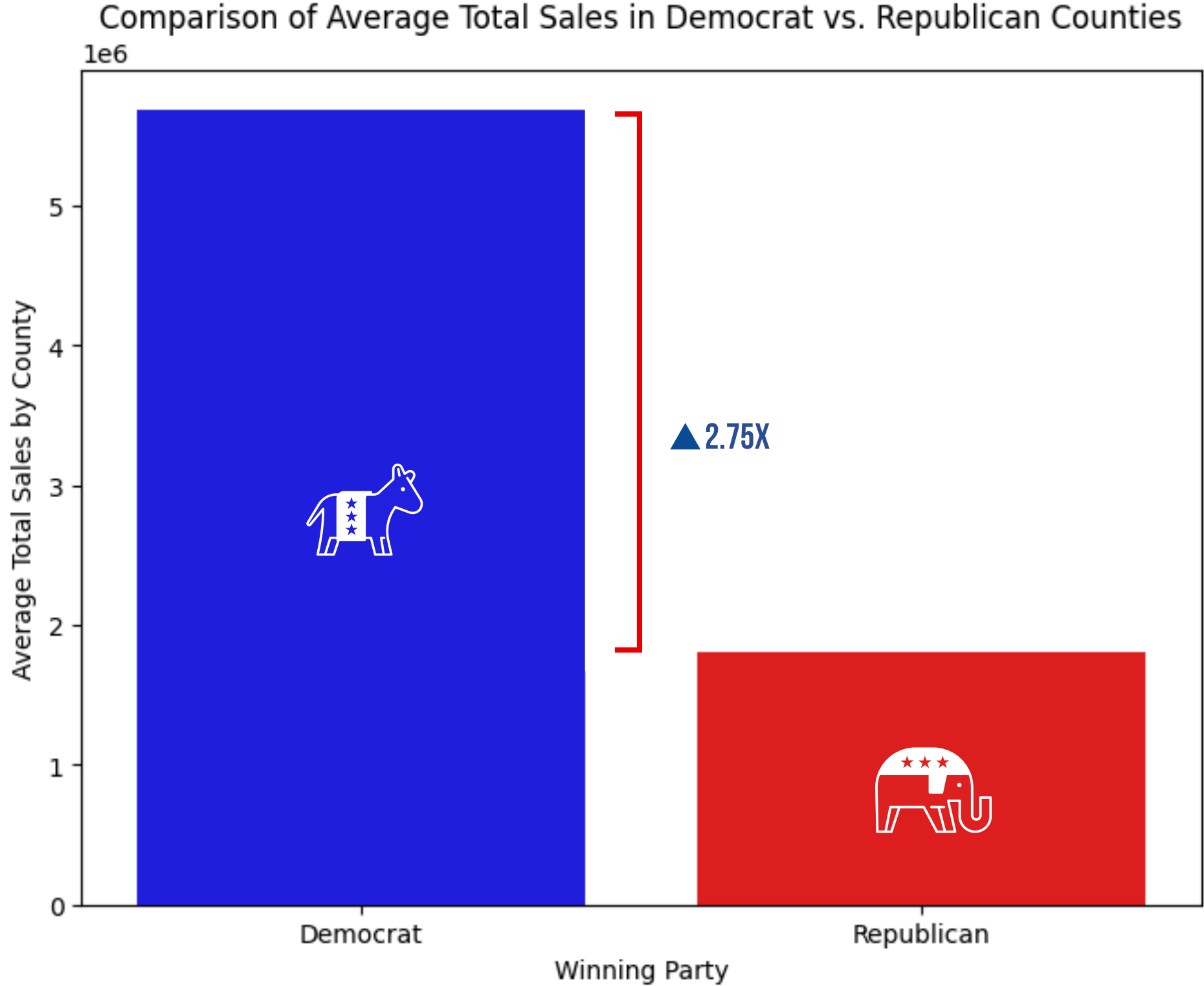
- Historical Robust Rural Support
- 2X Rep Rural Support ('10)
- Strong Urban Dem. Support (65%)
- Suburban = Divided Split

COUNTY ECONOMICS

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



AVG. COUNTY-WIDE SALES

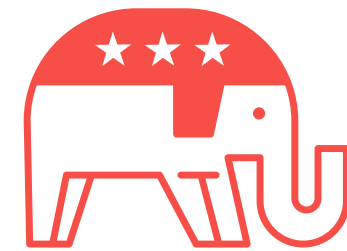


2016 Political Context:

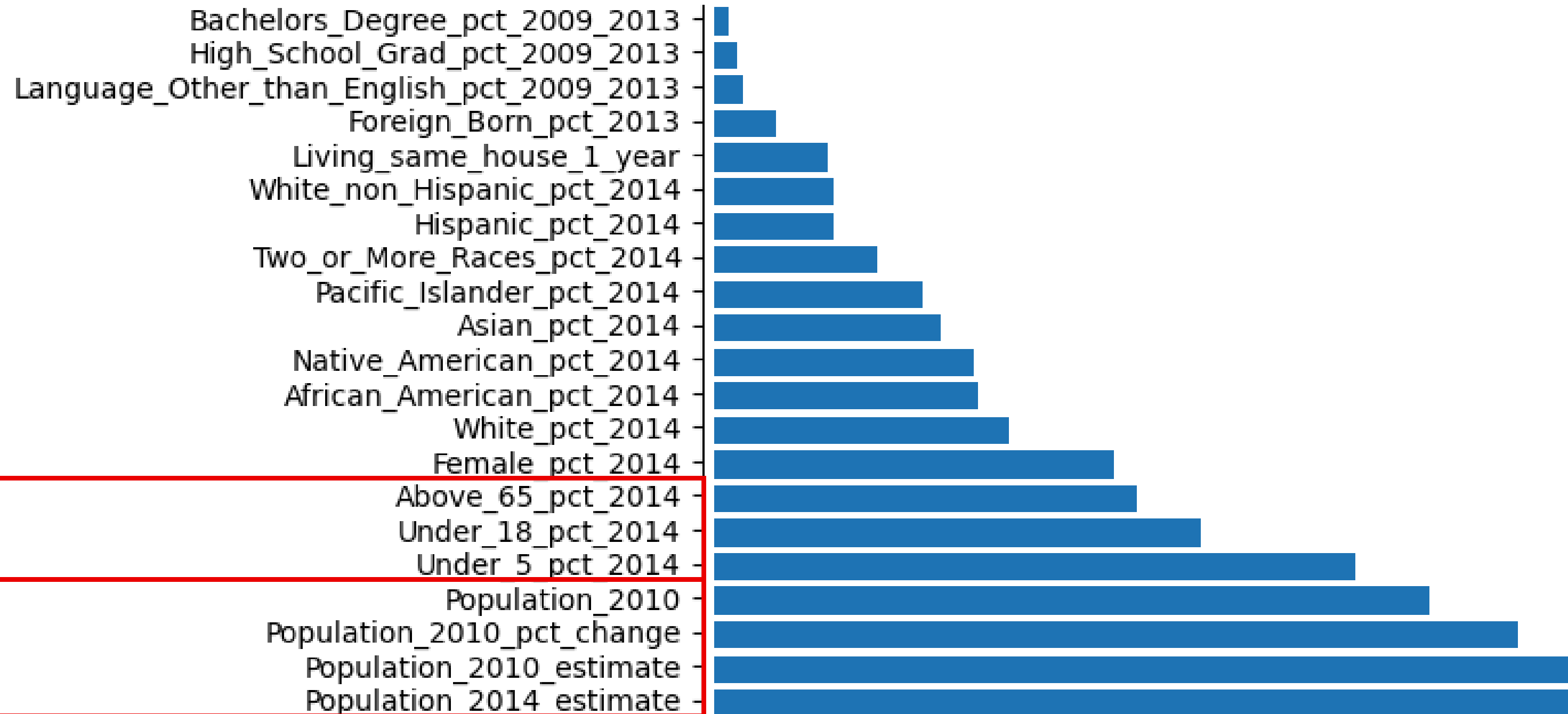
Clinton vs. Trump (extra info.)

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

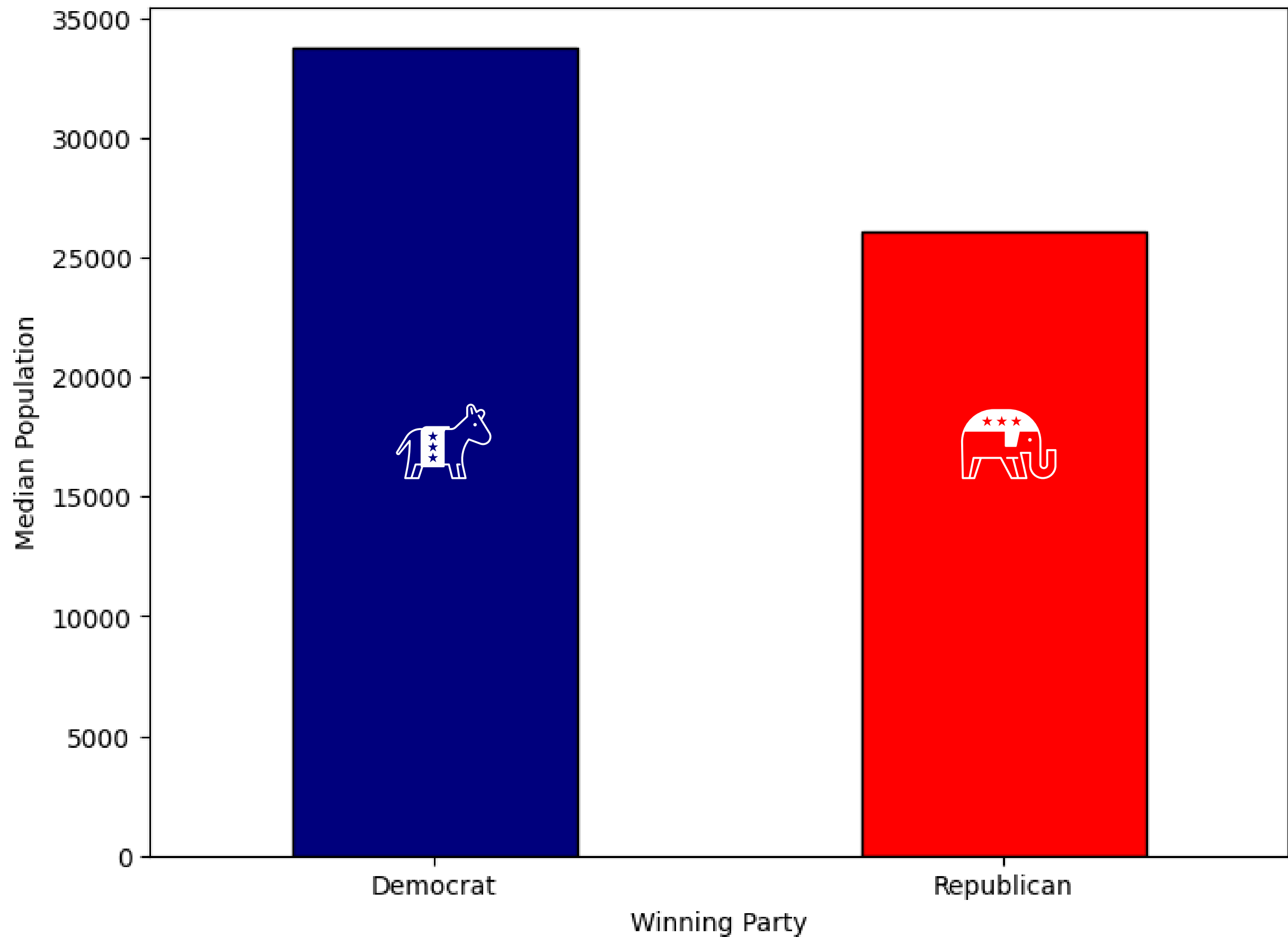
REPUBLICAN



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

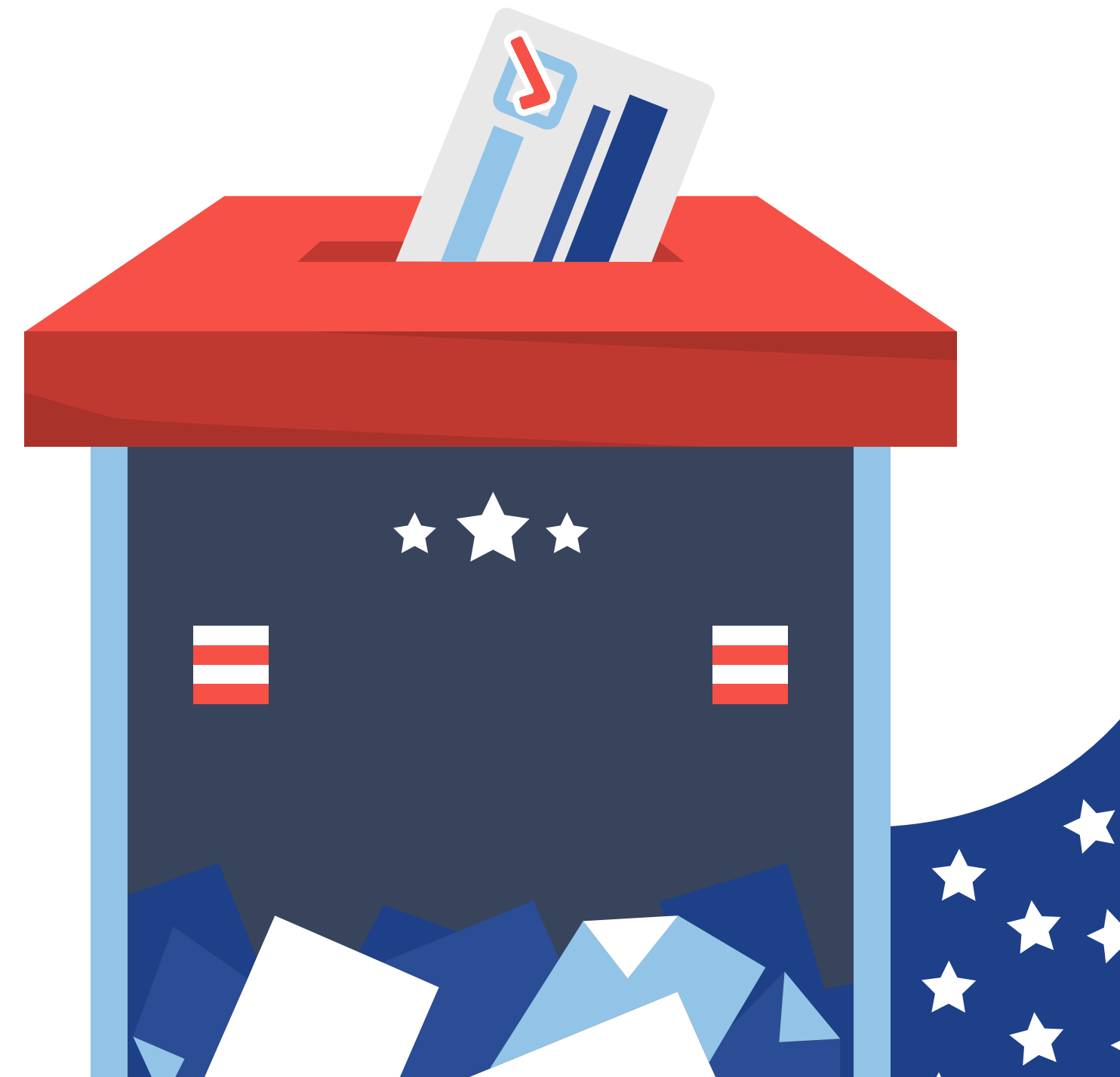
When controlling for race, poverty, age groups, etc., the **partial relationship with pop. can flip sign.**

POP. VS. DENSITY

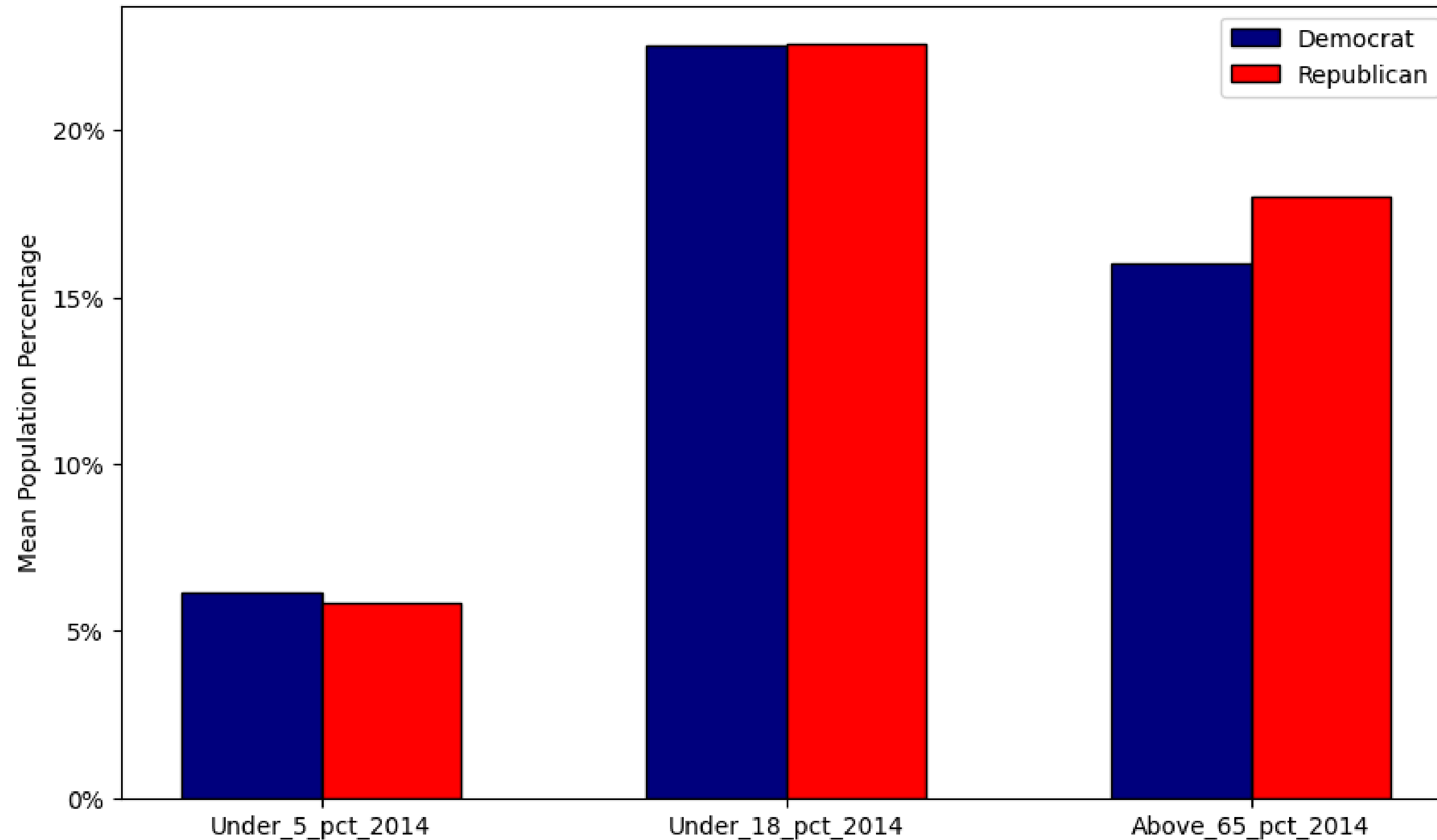
Large counties **aren't** necessarily **dense** (urban).

HOUSEHOLD AGE

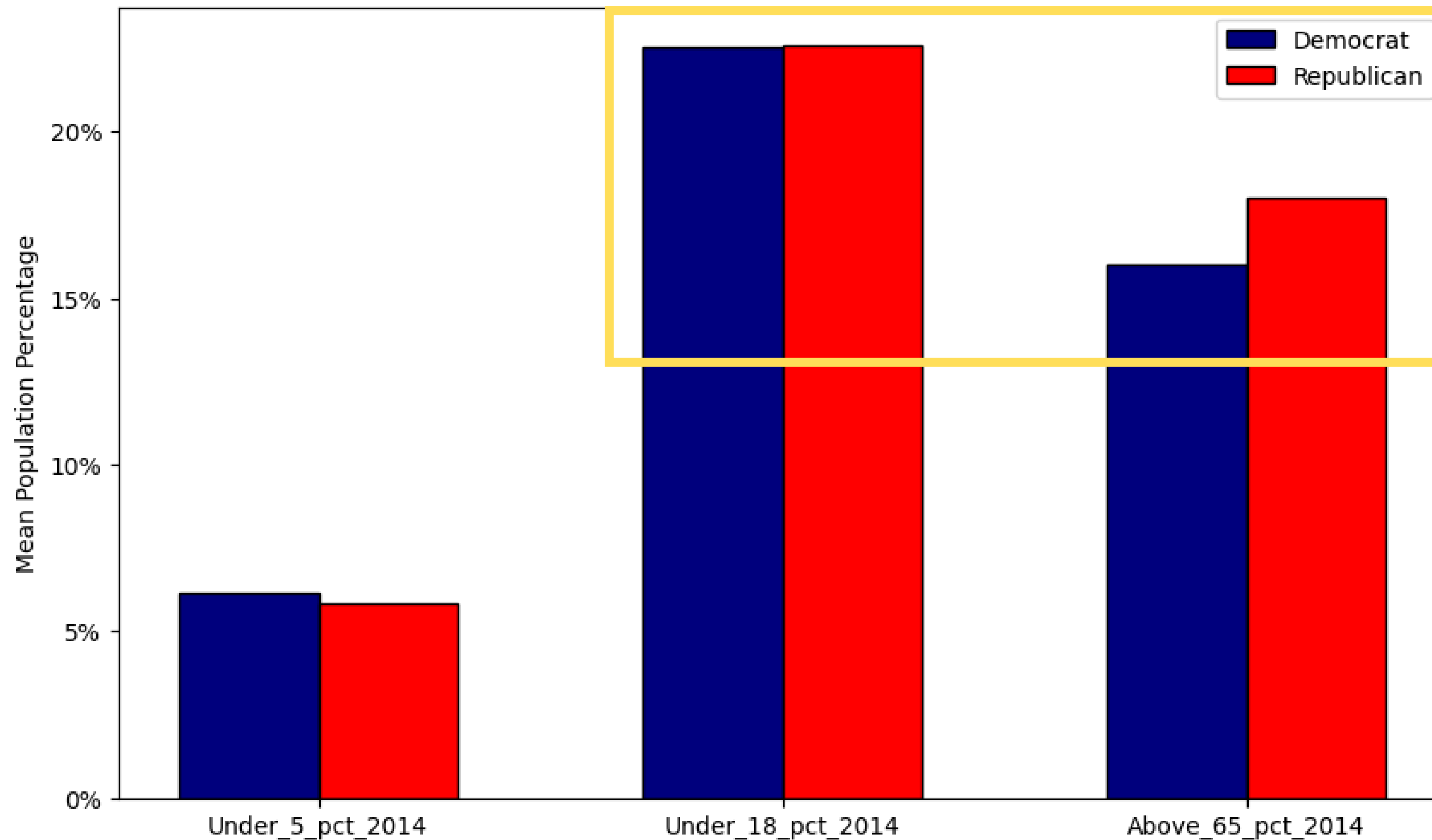
How do **age ranges** within a household affect voter tendencies?



HOW DO HOUSEHOLD AGE GROUPS CHANGE VOTING?

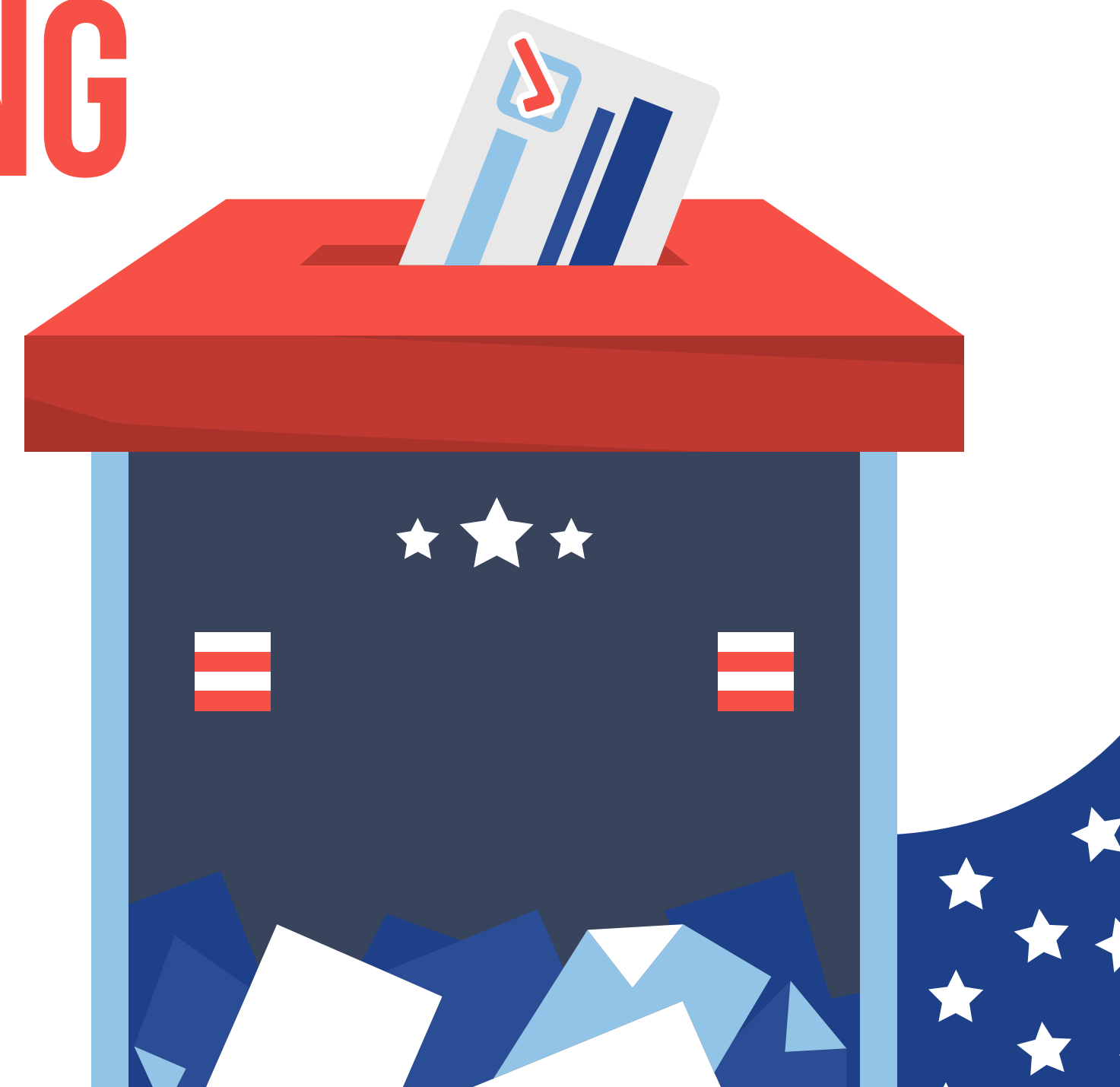


HOW DO HOUSEHOLD AGE GROUPS CHANGE VOTING?



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

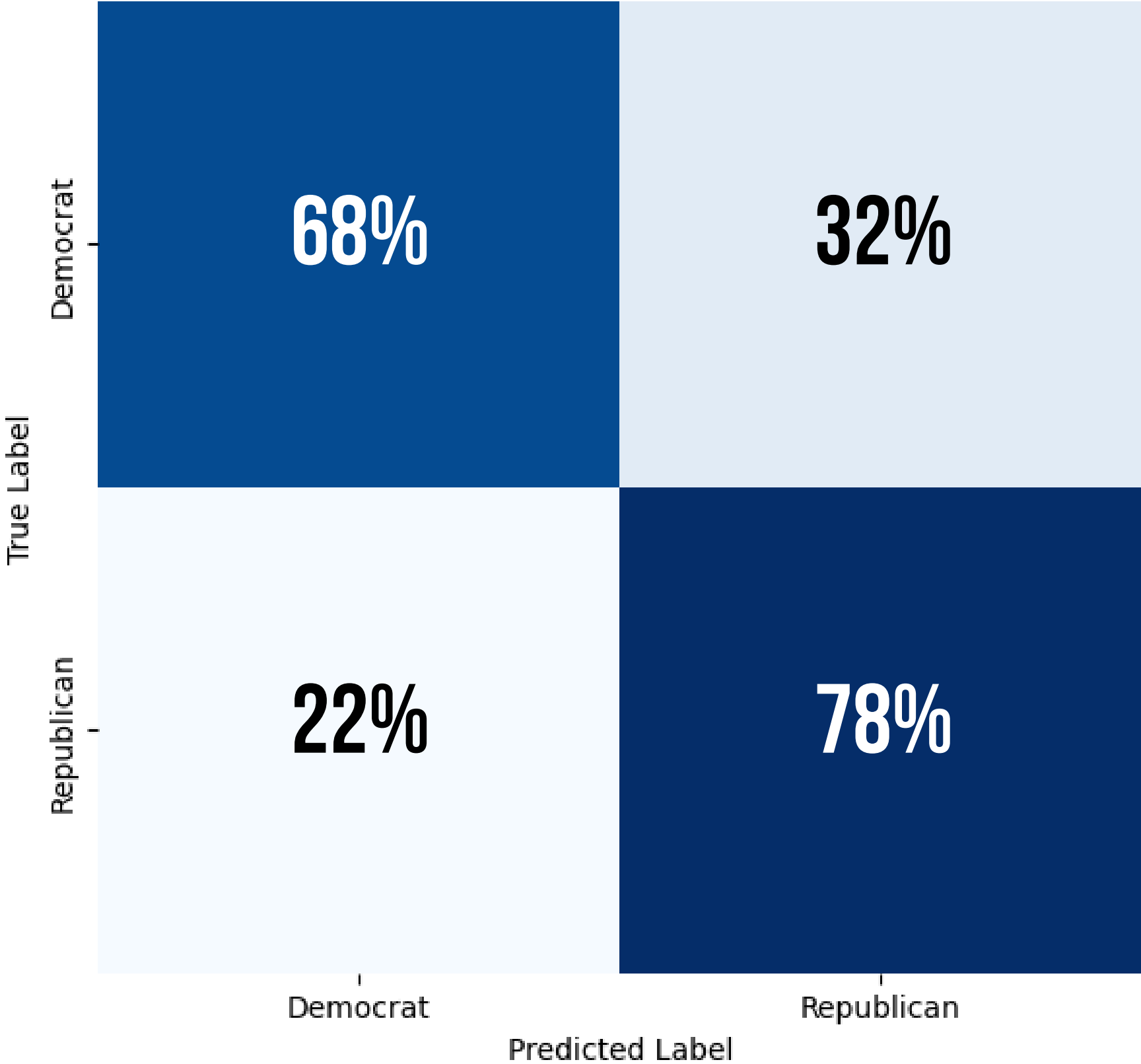
Under_5_pct_2014	Under_18_pct_2014
------------------	-------------------

Income_Index

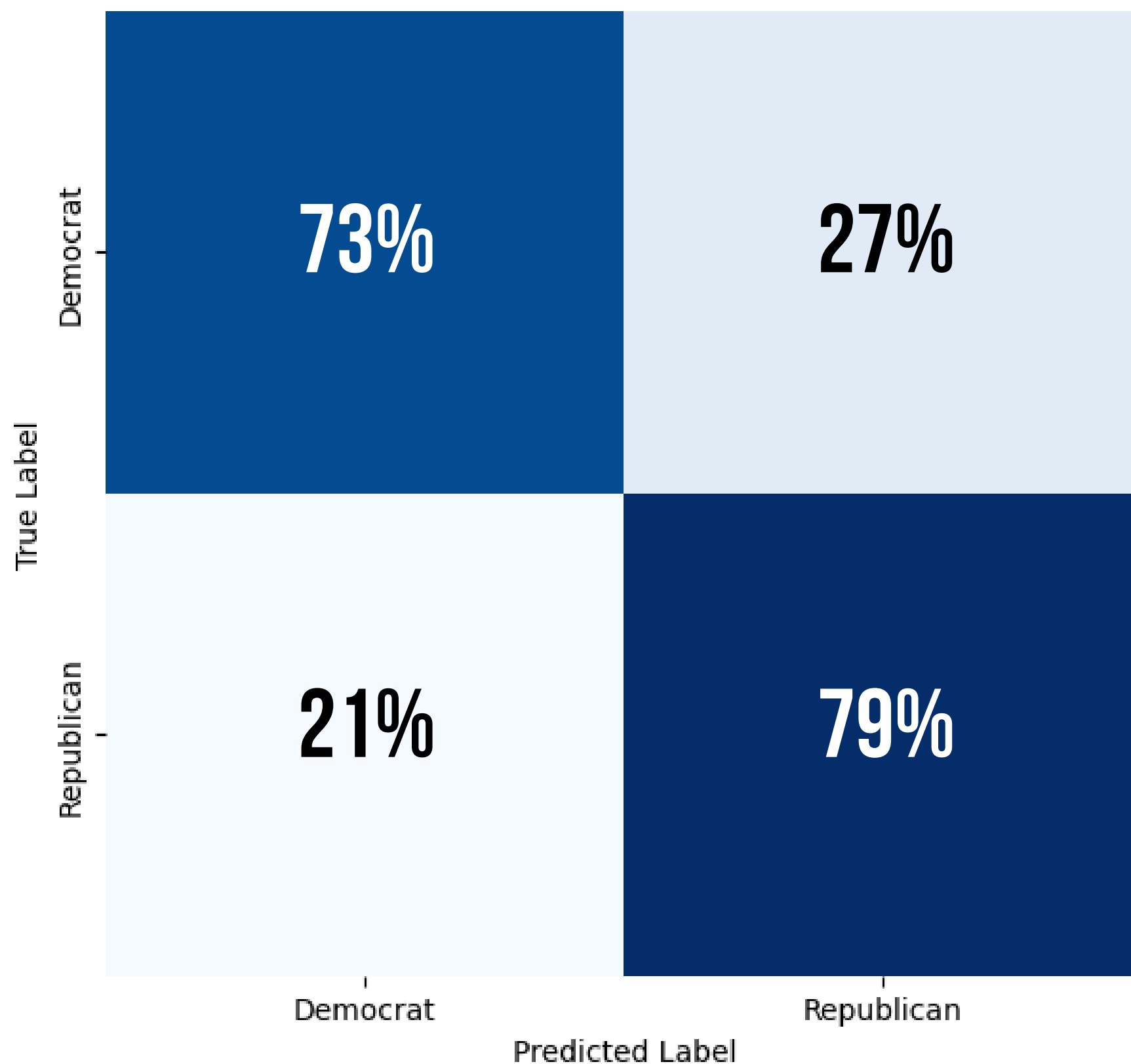
Median_Household_Income 2009_2013	Income_Per_Capita_2009_2013
--------------------------------------	-----------------------------

Correlation threshold = 0.8

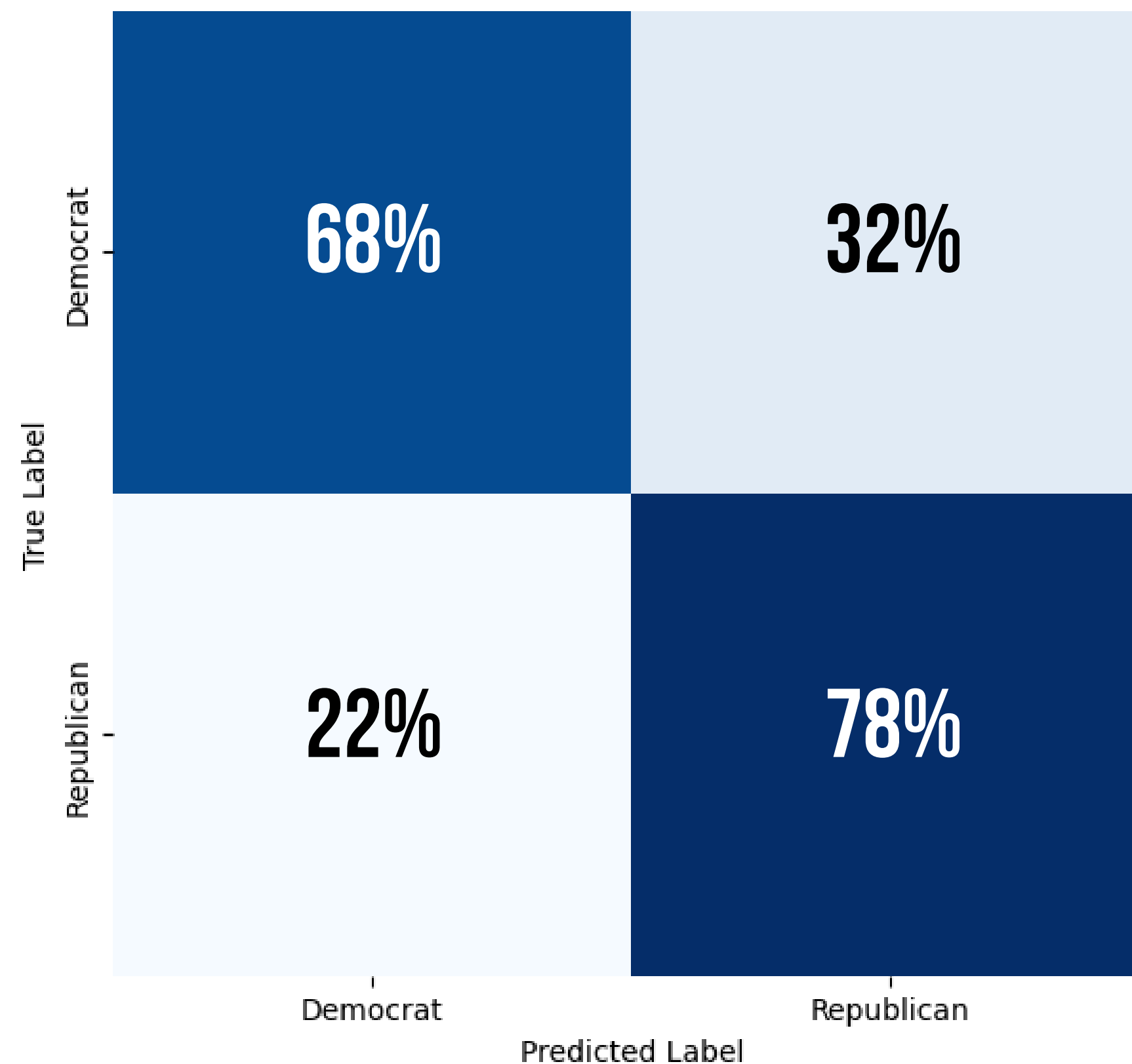
CONFUSION MATRIX



OLD



NEW



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?

