

## BACKGROUND

US elections have recently experienced a rise in voter turnout, with population change majorly influencing shifts in states political outcomes. Most notably, Georgia experienced a tight Senace race in the 2022 midterm election that called for a runoff. Thus, understanding voter interest is critical to a decisive voter turnout. As Twitter has previously been used as a means of gauging election results, this project seeks to determine if there is the relationship between voter turnout, population characteristics, and key voting interests tweets.

## DATA

1. **Tweets:** Tweets were collected through Twitter API between 2022-10-24 and 2022-11-08 for the following key terms: *abortion*, *inflation*, *unemployment*, *wages*. Tweets from users with Georgia based location in their profile were selected and cleaned.
2. **Voter Turnout:** Voter turnout data on county level were collected from Georgia Secretary of State's election statistics website.
3. **Population Characteristics:** This dataset used 2021 data from the US Census County Population by Characteristics: 2020-2021 dataset.

### Data cleaning:

Extensive data cleaning was conducted to merge information between tweets and users location, as Twitter deliver these data separately. User location information ranges from specific location to country level, so further work was done to produce location on county level. Tweet count was aggregated and merged with population and voter turnout datasets on county level. Due to limited amount of location provided, only 85 out of 159 Georgia counties appear in the dataset.

### Assumption:

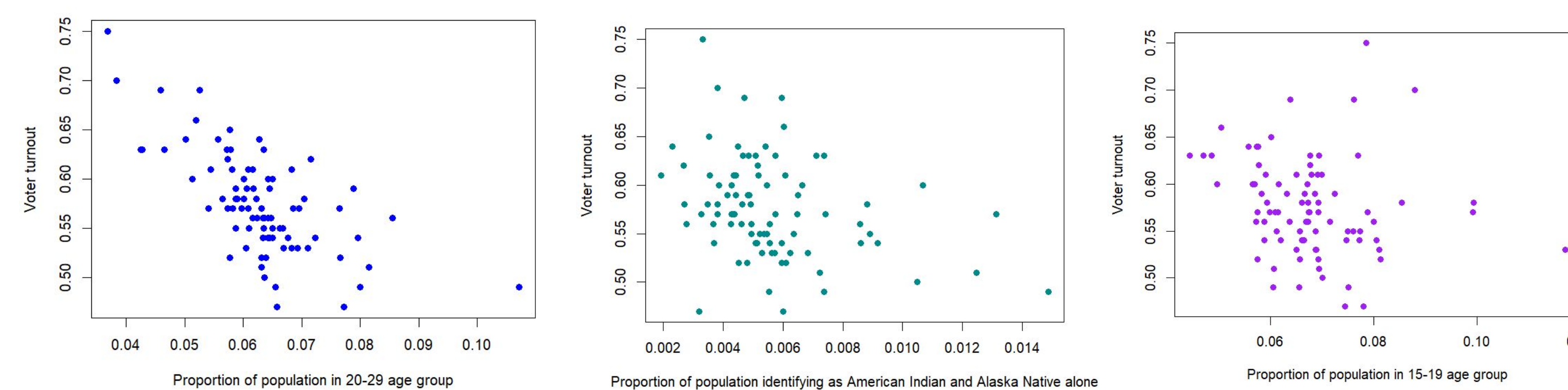
1. Counties selected are representative of Georgia counties.
2. Tweets are made independently of each other.

## MODELS & ANALYSIS

### Population characteristic as predictors

Subset of the variables were screened and automatically selected based on  $R^2$ , adjusted  $R^2$ , and BIC values:

1. Proportion of population in the 20 - 29 age group (*Age.20.29*)
2. Proportion of population who identified as only American Indian and Alaska Native (*Proportion.IA*)
3. Proportion of population in the 15 - 19 age group (*Age.15.19*)



Relationship between voter turnout and population characteristics

Model	Predictors	$R^2$	Adjusted $R^2$	BIC	P-value for new predictor from ANOVA *
1	<i>Age.20.29</i>	0.45	0.44	-300	$6.4e^{-14}$
2	<i>Age.20.29 + Proportion.IA</i>	0.54	0.53	-311	0.0001
3	<i>Age.20.29 + Proportion.IA + Age.15.19</i>	0.57	0.55	-312	0.019

\*:  $\alpha = 0.05$

### Tweets count as a potential significant predictor

Model	Predictors	$R^2$	Adjusted $R^2$	BIC	P-value for new predictor from ANOVA
4	<i>Age.20.29 + Proportion.IA + Age.15.19 + log (Tweet.Count )</i>	0.60	0.58	-313	0.027

All models were checked for multicollinearity, linearity, and equal variance. Based on  $R^2$ , adjusted  $R^2$ , BIC, and p-value, county local tweets count on key voting interests offers an improvement to using population characteristics as sole predictors.

Interaction models were found to offer no significant improvement.

## CONCLUSION

It can be concluded that there is a relationship between voter turnout and population characteristics and local tweets. The best model was determined to be:

$$\begin{aligned} \text{Voter Turnout} = & \\ & 0.904 - 3.688 \text{ Age.20.29} \\ & - 6.802 \text{ Proportion.IA} \\ & - 1.035 \text{ Age.15.19} \\ & + 0.006 \log (\text{Tweet.Count}) \end{aligned}$$

## LIMITATION & FUTURE STUDIES

### Limitations:

Only location data formatted with Twitter's default location selection were used in this dataset. Since Twitter is a discourse based platform, tweets are not guaranteed to be made independently from each other. Twitter prohibits sharing of user and tweet information, so these data are also anonymized and not accessible to public use. The results have not yet been checked for cross validation.

### Future Studies:

1. From Google Trends data, *healthcare* and *social security* were among the top searched queries in Georgia during the election. Adding these queries and extending collection duration can help yield more accurate results.
2. County level voter turnout statistics are also available for other states. Studying models for these states can allow more insights in comparison with Georgia.
3. Using NLP models to clean freeform location data and derive tweets sentiment can allow more meaningful data to be incorporated into the dataset.

### Data sources:

1. Twitter API v2: <https://developer.twitter.com/en/docs/twitter-api>
2. County Population by Characteristics (2020-2021) from US Census: <https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-detail.html>
3. Georgia Secretary of State's Election statistics: <https://results.enr.clarityelections.com/GA/115465/web.307039/#/summary>