RELATIONSHIP BETWEEN GEORGIA VOTER TURNOUT, POPULATION CHARACTERISTICS AND KEY VOTING ISSUES INTEREST ON TWITTER

Lan Dau '23, advised by Prof. Cassandra Pattanayak & Prof. Eni Mustafaraj

BACKGROUND

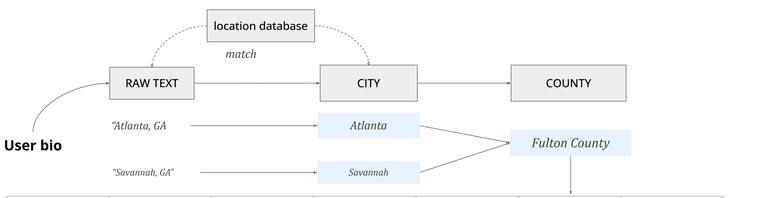
US elections have recently experienced a rise in voter turnout, with population change majorly influencing shifts in states political outcomes. Most notably, Georgia experienced a tight Senace race in the 2022 midterm election that called for a runoff. Thus, understanding voter interest is critical to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

What is the relationship between voter turnout, population characteristics, and key voting interests tweets?

DATA PROCESSING & GEOCODING

1. **Tweets:** Tweets and author profiles were collected through Twitter API between 2022-10-24 and 2022-11-08 for the following terms: *abortion, inflation, unemployment, wages*.
2. **Voter Turnout:** Voter turnout data on county level was collected from Georgia Secretary of State's election website.
3. **Population Characteristics:** This dataset used 2021 data from the US Census County Population by Characteristics dataset.

User location information is not universally formatted and ranges from city to country level, so further work was done to produce location on county level:

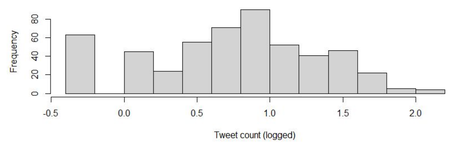


County	Population Characteristic 1	Population Characteristic 2	...	Population Characteristic 12	Tweet Count	Voter Turnout
Fulton County	X_1	X_2	...	X_{12}	3771	56%

County-level Population Characteristic Data — County-level Voter Turnout

The final dataset has sample size n = Due to this limitation, 85/159 Georgia counties are in the dataset.

Distribution of tweet counts across Georgia sampled counties



Assumption:

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

MODELS & ANALYSIS

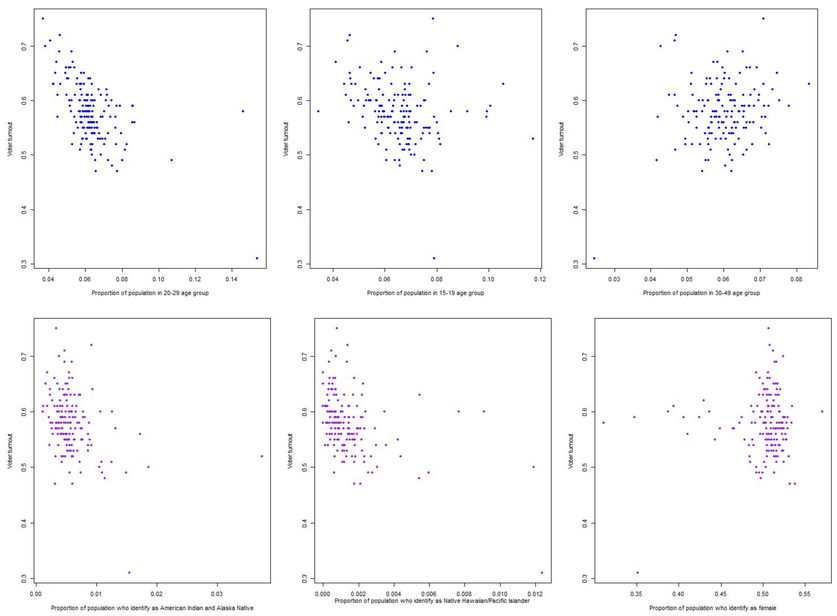
Population characteristic & tweet count as predictors

Upon checking for multicollinearity, the variables for proportion of African Americans, Asian Americans, Caucasian Americans were removed.
Subset of the variables were screened through automatic selection. Among these subsets, the models with the best R^2, adjusted R^2, Mallow's Cp value and AIC were selected. The models with the best adjusted R^2, Mallow's Cp value and AIC were found to be identical.

	Model	R^2	Adjusted R^2	AIC	P-value
1	Model with best R^2 value	0.694	0.6899	-2457	P-value < 0.0001
2	Model with best adjusted R^2 value	0.695	0.6891	-2453	P-value < 0.0001

Cross validation based on PRESS value was used to select the best first-order model, which include the following variables:

1. Proportion of population in the 20 - 29 age group
2. Proportion of population in the 15 - 19 age group
3. Proportion of population in the 30 - 49 age group
4. Proportion of population who identify as American Indian and Alaska Native
5. Proportion of population who identify as Native Hawaiian/Pacific Islander
6. Proportion of population who identify as female.
7. Population density
8. Number of tweets



Relationship between voter turnout and population characteristics

Model were checked and found to satisfy all linear regression assumptions. No significant interaction was found.

CONCLUSION

The best model was determined to be:

Mean (Voter Turnout | 20-29 age group, 15-19 age group, 30-49 age group, Female, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, density, Tweet.Count = e^k *)*

= 1.45
- **– 4.71** 20-29 age group
- **– 1.00** 15-19 age group
- **+ 0.97** 30-49 age group
- **- 0 .79** Female
- **- 9.67** American Indian/Alaska
- **+ 5.46** Native Native Hawaiian/Pacific Islander
- **+ 0.8 x 10^{-4}** Density + k

Based on the model, voter turnout was found to be negatively correlated with proportion of population who are in the youth age group, female, American Indian/Alaska Native, and positively correlated with population density and proportion of population who are Native Native Hawaiian/Pacific Islander

It is worth noting that Georgia was reported by the Election Law Journal to have the second highest in the Cost of Voting Index, indicating of how restrictive voting laws in this state is. This can be an explanation for the negative relationship observed between voter turnout and demographics with lower socioeconomic status,

The fact that the number of voting issues tweets wasn't found to be significant to the model, as twitter user demographics is not reflective of the entire US population. However, the inclusion of this variable yields the best adjusted R^2 value, Mallows' Cp value, and AIC value implies that between areas that have the same population characteristics, online vocalness can be an useful variable in predicting voter turnout.

LIMITATION

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.

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