Using Voter File Data to Study Electoral Reform

Group Members: Carolyn Kolaczyk, Doma Ghale, Jin Ishizuka and Rodrigo Kreis de Paula
Mentor: Prof. Julia Payson

Problem

Many US cities have begun adopting an alternative voting system called ranked-choice voting (RCV), in which voters rank candidates rather than simply voting for one. Proponents of RCV argue that this is a more fair system than the typical "winner takes all" approach in US politics. Critics claim that it is confusing and might discourage turnout, especially among minority groups.

Goals

- Create city voter profiles to study turnout at different demographic levels
- Visualize voter turnout across demographic profiles and RCV vs. Non-RCV elections
- Examine the relative importance of various demographic features in predicting voter turnout

Data

- L2 voter file contains two types of files: Demographic and VoteHistory
  - Demographic file consists of 691 detailed demographic variables for each voter
  - VoteHistory file contains information on which elections each voter voted in between 1994 and 2022
- Census data consists of demographic data at the city level

Methods

Selecting states, cities, and elections

- Chosen states with RCV cities: *California, Colorado, Maine, Maryland, Minnesota, New Mexico, Utah, Vermont*
- Cosine similarity to select 5 most similar Non-RCV cities for each RCV city
- Select up to 4 most recent elections per election type for each city, occurring in 2008 or later

Creating voter turnout profiles

- Merging demographic and voter turnout data, we create voter turnout profiles for each election with information about turnout for different demographics such as age, income, education, and race/ethnicity

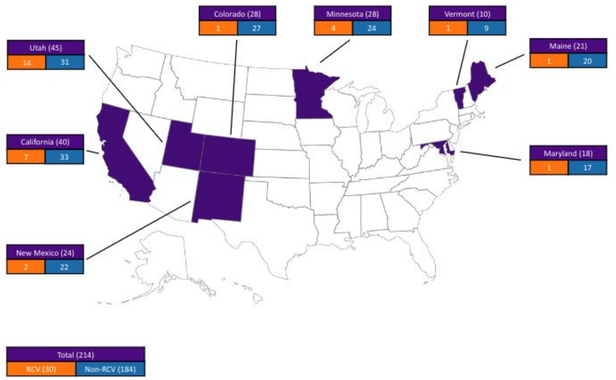
Visualizations

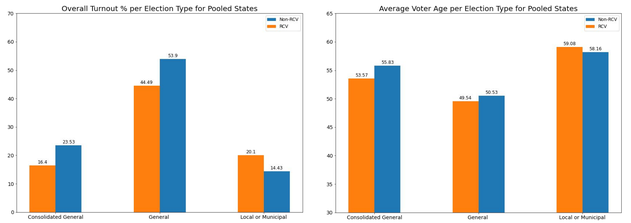
- We create bar graphs and other visualizations to display the difference in turnout for RCV and Non-RCV elections

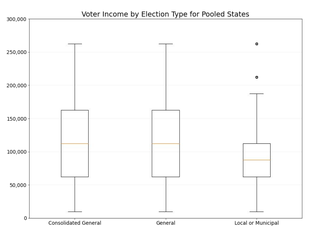
Model

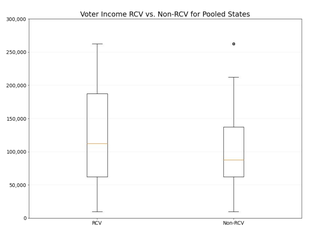
- We implement several random forest models using the targets and features listed below to examine the factors that predict different election outcomes
  - **Targets**: (1) overall turnout, (2) turnout among white voters, (3) average age of voters, turnout for voters with incomes (4) < 50K and (5) > 100k
  - **Features**: state, election type, RCV, city population, city % white, % with college degree or higher, median age, median income

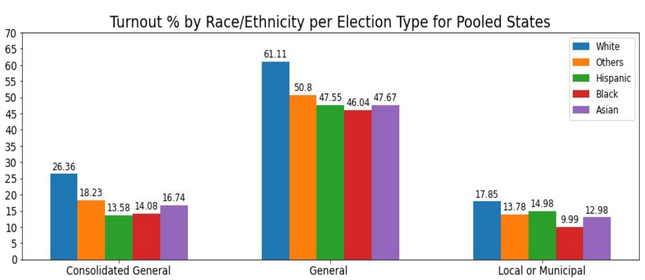
Results & Conclusions











RCV vs. Non-RCV findings

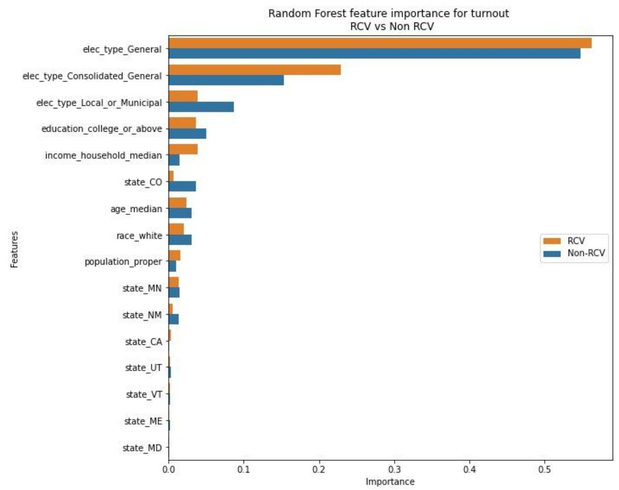
- RCV cities have higher turnout in Local or Municipal Elections, showing that it is potentially more effective at attracting voters in elections with more candidates (vs. polarized General Elections with fewer candidates)
- RCV cities did not have a substantial difference in the voter age when compared to non-RCV cities
- RCV cities have a wider and slightly higher range of voters incomes than non-RCV cities

Election type findings

- Local or Municipal Elections tend to attract lower-income voters compared to Consolidated General and General Elections, which have a higher and more distributed range of voter incomes
- Local or Municipal Elections tend to attract a more racially diverse set of Voters compared to Consolidated and General Elections, in which the gap between White voters and other races and ethnicities is more pronounced

Predictive modeling findings

- Here we focus on our primary modeling objective, examining the demographic factors that predict overall voter turnout. In order to see if ranked-choice voting affects the features that predict turnout, we train two random forest models, one using data from RCV cities and one using data from non-RCV cities. We examine the feature importance scores from these two models below:



- We see that election type is the dominant feature in predicting voter turnout across both RCV and Non-RCV cities. The influence of other demographic variables, such as education level, income, age, and race are comparatively minimal.
- Additionally, we implement a random forest model using combined RCV and Non-RCV data and find that whether a city utilized ranked-choice voting or not is not a significant predictor of overall turnout.

Future Work

Expanding to more states

- Our current analysis focused on 8 US states and 214 cities. In the future we would like to expand to include more states and cities.

Prediction of voter turnout at individual voter level

- The L2 voter file data contains information at the individual voter level. We would be interested in creating models to predict whether a specific individual will turn out to vote at a particular election, using information about both the individual and the election. Some examples of feature engineering we would perform are the number of elections each individual has voted in the past and whether the individual voted in the last election.

References

- DeSilver, D., Blazina, C., Chavda, J., & Leppert, R. (2021, June 29). *More U.S. locations experimenting with alternative voting systems*. Pew Research Center. https://www.pewresearch.org/fact-tank/2021/06/29/more-u-s-locations-experimenting-with-alternative-voting-systems/
- DeSilver, D. (2018, February 15). *Q&A: The growing use of 'voter files' in studying the U.S. electorate.* Pew Research Center. https://www.pewresearch.org/fact-tank/2018/02/15/voter-files-study-qa/
- Hajnal, Z., Kogan, V., & Markarian, G. (2022). Who Votes: City Election Timing and Voter Composition. *American Political Science Review, 116*(1), 374-383. doi:10.1017/S0003055421000915.
- Nickerson, D. W. & Rogers, T (2014, Spring). *Political Campaigns and Big Data.* Journal of Economic Perspectives, Volume 28, Number 2, Pages 51–74.
- Rentsch, A., Schaffner, B. F., & Gross, J. H. (2019, Winter). *The Elusive Likely Voter. Improving Electoral Predictions with more informed vote-propensity models.* Public Opinion Quarterly, Vol. 83, No. 4, pp. 782–804.

Acknowledgement

We would like to thank our brilliant mentor Prof. Julia Payson for her guidance throughout the project. Her great leadership, organization, insights and knowledge about the subject were crucial to all obtained results. We would also like to thank Prof. Julia Kempe for her advice and supervision.

github.com/jnshzk/capstone