

Historical Context and Current Trends in Predicting Voter Turnout in Douglas County, Nebraska

Yuxuan Chen* Ramesh Danuwar† James M Geiger, MA‡ A.J. Guevara§
Pratik Kandel¶

16 December, 2020

Abstract

Historically, the trend in voter participation for elections in the United States is based on the expansion of voting rights. Voter turnout becomes gradually larger over the years as access to voting is made easier and more equitable. In this research, the analysis of voter turnout will be focused on Douglas County, Nebraska. Using public data from the Douglas County Election Commission and the United States Census Bureau, we identify a positive trend in voter registration in the past 20 years, especially among voters who identify as other- or non-partisan. We furthermore identify an upward trend in population growth among White residents while the population of other demographic groups remains flat. Ultimately, we present a preliminary model that local election commissions could employ to predict voter turnout by predicting voter turnout for the 2020 general election. The results of the preliminary model are promising—the model was able to predict turnout to a difference within 3,000 voters.

*University of Nebraska Omaha, Architectural Engineering

†University of Nebraska Omaha, Management and Information Systems

‡University of Nebraska Omaha, Data Science

§University of Nebraska Omaha, Data Science

¶University of Nebraska Omaha, Management and Information Systems

1 Data

1.1 Sources

Detailed voter registration and election result data are made openly available by the Douglas County, Nebraska Election Commission (election commission) on their website (Douglas County Election Commission, 2020b). Because the Election Commission is responsible for the fair and accurate conduction of elections in the county as a whole, they provide data on both federal and municipal elections, including special issues elections that don't fall within a specific election year. Since this project will focus on federal elections, we've limited our inclusion of data to presidential and non-presidential election years. Douglas County provides voter turnout data from the 1996 election onward, however tabular data have only been available since 1998. To avoid text mining, only federal election years since and including 1998 have been included in this analysis.

Analyses of voter turnout easily lend themselves to exploration of results by geographic area. Douglas County divides its election areas into wards and precincts. In 2020, there were 451 precincts within eight wards. The election commission provides, upon request, ESRI shapefiles (ArcGIS, n.d.) of the local ward/precinct map (Douglas County Election Commission, 2020a).

The United States Census Bureau is most widely known for its decennial national census. While congressional redistricting relies solely on the 10-year census, the bureau also surveys and provides data in-between census years, known as the American Community Survey (ACS) (United States Census Bureau, n.d.). The ACS data are generally provided in aggregate, however another Census Bureau product—the Public Use Microdata Sample (PUMS)—provides the same data gathered by the ACS in de-identified hyper-local form, down to the individual or housing unit (United States Census Bureau, 2020). The PUMS can be paired with the Public Use Microdata Areas (PUMA) (United States Census Bureau, 2019) to aggregate the ACS into small areas at the county level. While the decennial census is able to provide more detailed geographical information, the ACS combined with the PUMS provides a level of detail sufficient for exploratory analysis.

1.2 Cleaning and Preparation

Data for this analysis were relatively clean by default—all data from the Douglas County Election Commission was available in Microsoft Excel format and was free from character encoding errors. Data from the US Census Bureau was likewise available in comma-separated format and was easy to incorporate without further processing. However, the election turnout data were in separate files across the 23 total elections included in this analysis. Furthermore, column labels were not consistent across years and were often named differently or in a different order making it difficult to easily bind these files together. Another unique difficulty in cleaning the data presented itself in the changing political landscape over the 20-year analysis period. A number of “third parties” have grown and lost popularity since 1998 and the individual data files account for the most prevalent third parties that voted in that election.

To process the data from the election commission we first selected only the variables of interest, dropping any columns and rows that the commission had computed. To handle the “third party” issue mentioned previously, each of these individual parties were collapsed into a single column. Although many analysis presented in this paper will further collapse third party voters into the larger “nonpartisan” category, we elected to retain more specificity during the data cleaning process to provide the most flexibility in exploring the data. Prior to finally combining the data, additional variables were added to specify for each row the election year, whether it was a presidential or gubernatorial year, and whether it was a primary or general election.

2 Methods

Analysis of the data were conducted using the R statistical computing language (R Core Team, 2020) inside the RStudio integrated development environment (RStudio Team, 2020). An array of R packages was used to

ease data transformation and statistical analysis including: `tidyverse` (Wickham, H., et al., 2019), `readxl` (Wickham, H. and Bryan, J., 2019), and `reshape2` (Wickham, H., 2007) for file processing and data cleaning; `ggplot2` (Wickham, H., 2016) and `scales` (Wickham, H. and Seidel, D., 2020) for plot graphics and labels; `sf` for geographic data (Pebesma, E., 2018). This paper has been compiled with reproducibility in mind. The codes and data necessary to inspect our findings or conduct one’s own analysis has been placed into a GitHub repository.

We conducted an exploratory analysis of the data to inform the selection of variables to include in a predictive model. Our exploration focused on voter registrations and turnout—first independently and then with respect to their geography within the county. Results of our analyses are in four parts: 1) an exploration of the trend in voter registration over time and party, 2) registration within demographic and geographic contexts, 3) voter turnout with regard to election cycle and type, and 4) turnout by geography. Finally, the findings of our analysis come together to produce a predictive model for voter turnout.

3 Results

3.1 Voter Registration

In order to be able to participate in an election, a potential voter must be registered to vote by a certain deadline that is set prior to the election. We started, then, by exploring changes in voter registration. Figure 1 displays a plot of voter registrations at the deadline for each general election since 1998. Overall, registrations have grown from 280,693 in 1998 to 357,159 registered voters in 2018, the most recent year that general election data were available. Over the same period, the group of voters realizing the most growth were nonpartisan/third party voters. While nonpartisan and third party voters accounted for only 15.81% of the electorate in 1998, the same group twenty years later represented 25.56% of the electorate. Partisan registrations have remained relatively uniform.

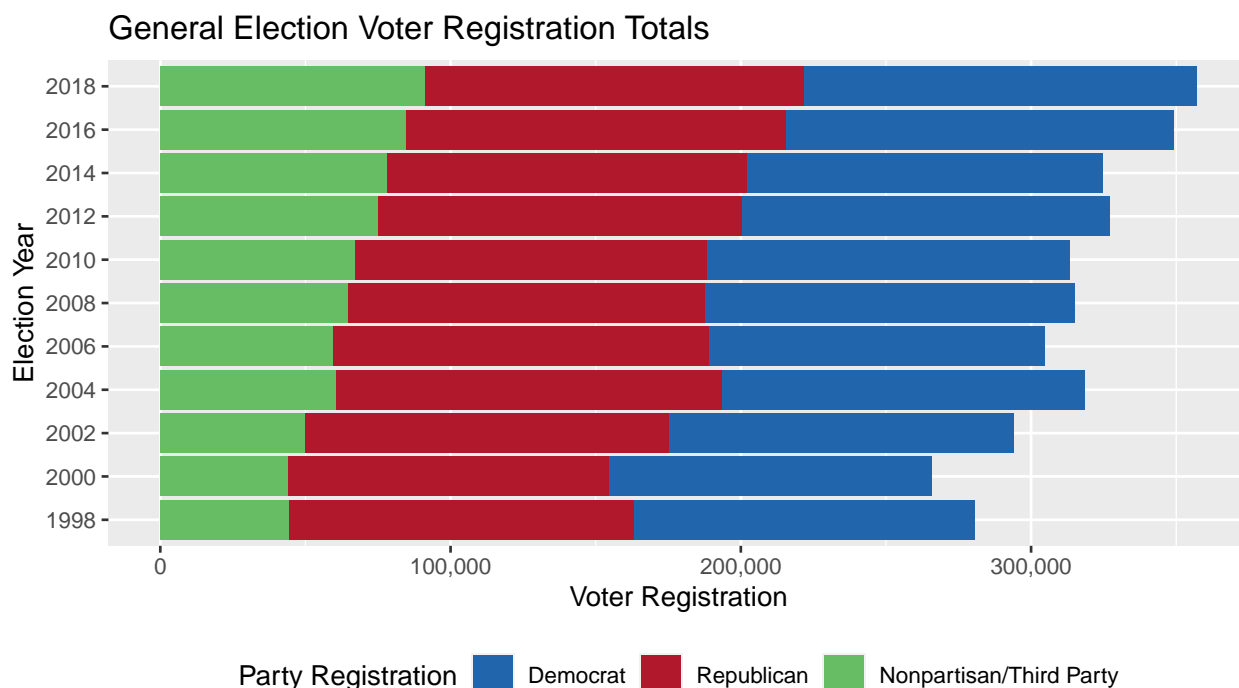


Figure 1: Plot of voter registration by party affiliation. Since 2000, voter registration has increased. These increases are mostly seen in nonpartisan/third party registrations.

Discussion of voter registration growth must be understood within the broader context of population growth within the county overall. The earliest data available from the American Community Survey (ACS) that

included detailed geography began in 2006 and extended to 2018. Figure 2 displays these data over time with a linear model overlaid to display trend. Over this period, the population in Douglas County grew 16.86% from 360,904 in 2006 to 421,737 in 2018. Over this same period, voter registrations increased 16.39%, which is in-line with population growth overall.

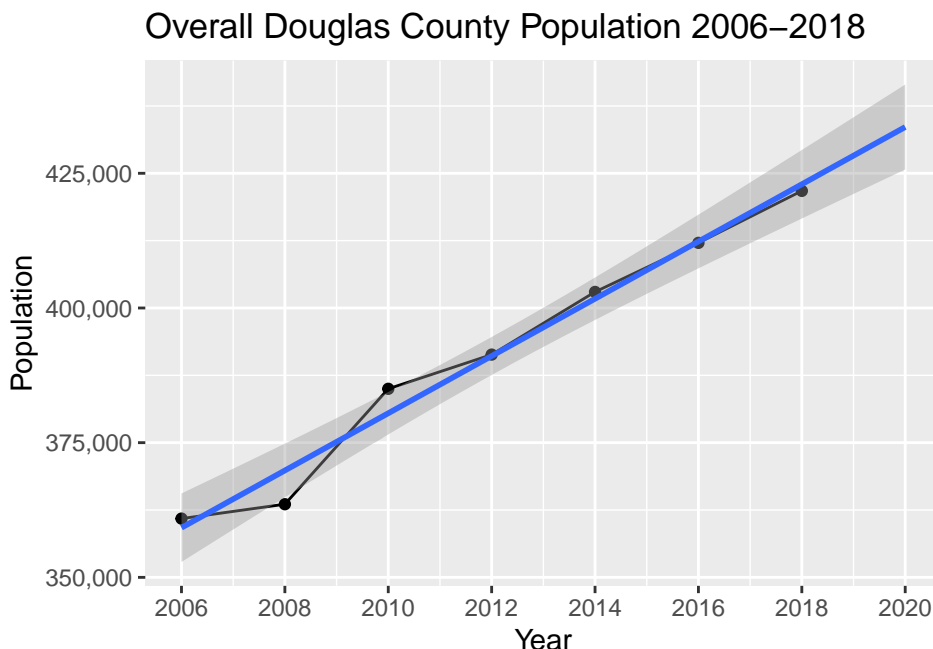


Figure 2: Plot of Douglas County population growth between 2006 to 2018. Growth is projected using a linear model to 2020. The regression line from the linear model is shown in blue, with 95-percent confidence intervals displayed in the grey shaded band.

With this in mind, we turned our attention to any shifts in demographics of the county populous that could inform or impact voter behavior overall. Because the aggregate voter registration and turnout data available through the election commission do not include any demographic information such as age, sex, race, we make certain inferences based on demographic changes observed in the ACS data. We identified race as one such salient factor upon which to draw inference and explored further. (The ACS makes no distinction between race and ethnicity in their data gathering. For the purposes of this analysis, we will use the terms interchangeably, recognizing that doing so is a limitation of the data overall.) Figure 3 displays the change in population by race over the same period. The plot makes clear that persons who identify primarily as White/Caucasian far outnumber all other identified ethnic groups. However, growth in this group also appears to be contributing most to population growth overall.

Although this observed disparity in population growth cannot directly be correlated with changes in voter registration or turnout, a geographic analysis of voter registration and population distribution may provide additional context. We first mapped the proportion of Democrat vs. Republican registrants by voting precincts, shown in Figure 4. Taking only partisan voter registration into account, deeper shades of blue indicate precincts with higher proportion of Democrats, where deeper shades of red indicate a higher proportion of registered Republicans. The lightest shades on the map mark an even balance between the two parties.

The geographic distribution of Democrats tends to be clustered in the north- and southeastern portions of the county. Moving from east to west, registered voters begin to lean more Republican, with the highest concentration in the westernmost parts of the county. The area in-between these two groups, roughly correlating with central Omaha, appears to be evenly distributed between the two parties. While this plot is interesting on its own, it becomes more revealing when paired with a similar plot exploring the distribution of racial groups in the county. Figure 5 maps the proportion of persons identified as primarily White/Caucasian vs. all other racial groups overlaid over the outline of voting precincts in the county. Areas which tend to be

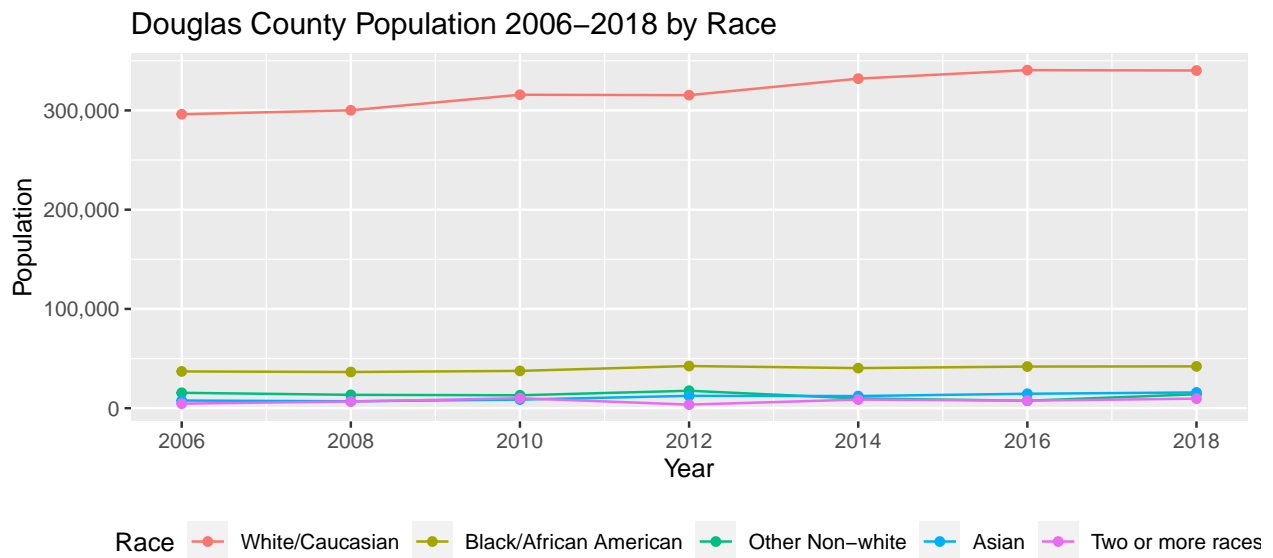


Figure 3: Plot of Douglas County population growth from 2006-2018, grouped by race. White/Caucasian persons far outnumber any other racial group individually and in aggregate.

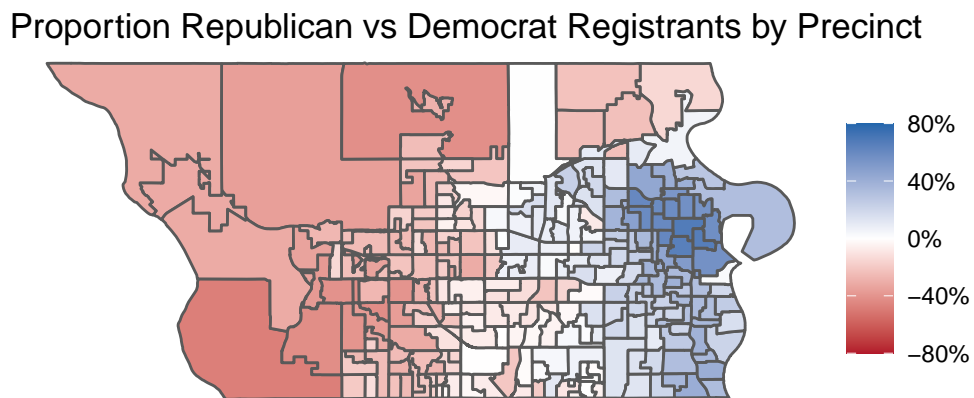


Figure 4: Plot of voter registration weight by party. Precincts with higher Democrat registrants are displayed in darker shades of blue, precincts with higher Republican registrants are displayed in darker shades of red. Third party/Nonpartisan registrants are not included.

less White also tend to have more registered Democrats.

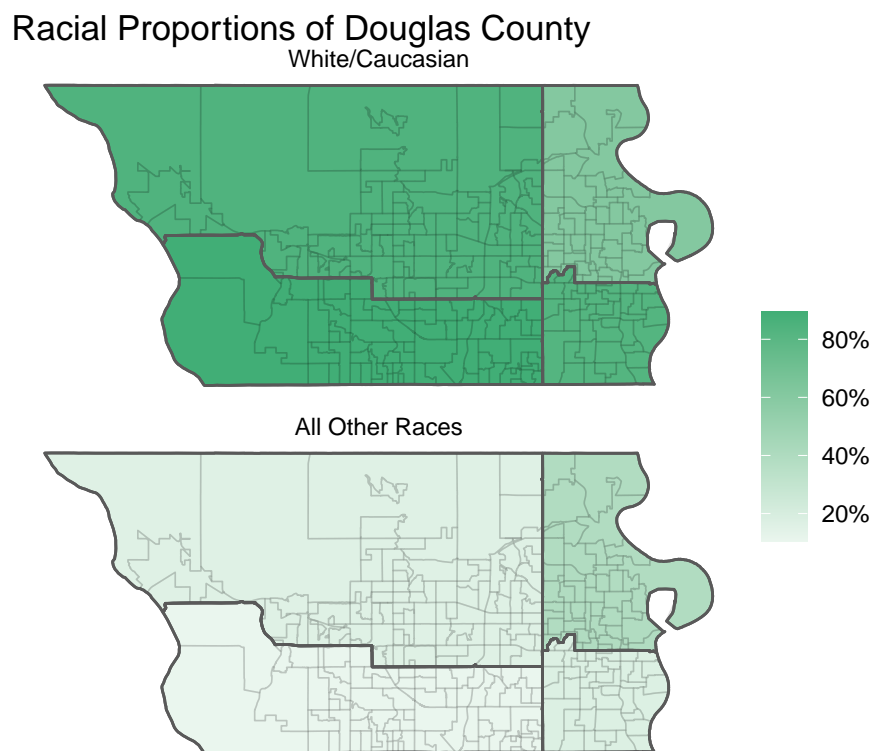


Figure 5: Geographic distribution of the Douglas County population as of 2018, faceted by race. Deeper shades of green indicate a higher proportion of persons of that racial group residing in the area.

3.2 Voter Turnout

Next, we turned to explore voter turnout which is the proportion of people who were registered to vote in any given year and who subsequently voted in an election. We conducted two analyses, first examining turnout in presidential compared to non-presidential election years, then evaluating differences in turnout between primary and general elections.

Figure 6 displays election turnout grouped by year of election type—whether presidential or non-presidential—and faceted by a voter’s political affiliation. For this exploration, turnout was considered for each primary and general election in sum. Voter turnout in Douglas County has been consistently higher for presidential elections compared to non-presidential elections over all the elections included in this analysis. Broken out by political party, Republicans tended to turn out in higher numbers overall, averaging 45.56% turnout compared to Democrats who averaged 38.74. Although certain parties broke trend in certain years—for instance, Republicans turned out noticeably higher in 2006—trends remained fairly flat.

To better understand how the primary and the general election turnouts have changed over the years, a plot comparing the primary and the general election turnouts faceted by different parties is displayed in 7. As is seen in the plot, the general election turnout rate has been higher than the primary election turnout rate for every election in the past 20 years, averaging 56.31% turnout in general elections compared to an average turnout in primary elections of 21.34%. This surge in turnout during the general election could be due to greater passion and interest in voting in general elections. However, the changing pattern of turnout rate is shown to be different for each type of election. The seasonality in turnout is even more evident in Figure 7—presidential election years align with the spikes in each line. The group who tended to turn out the least overall was nonpartisan/third party voters, averaging just 43.73% turnout in general and NA% turnout in primary elections.

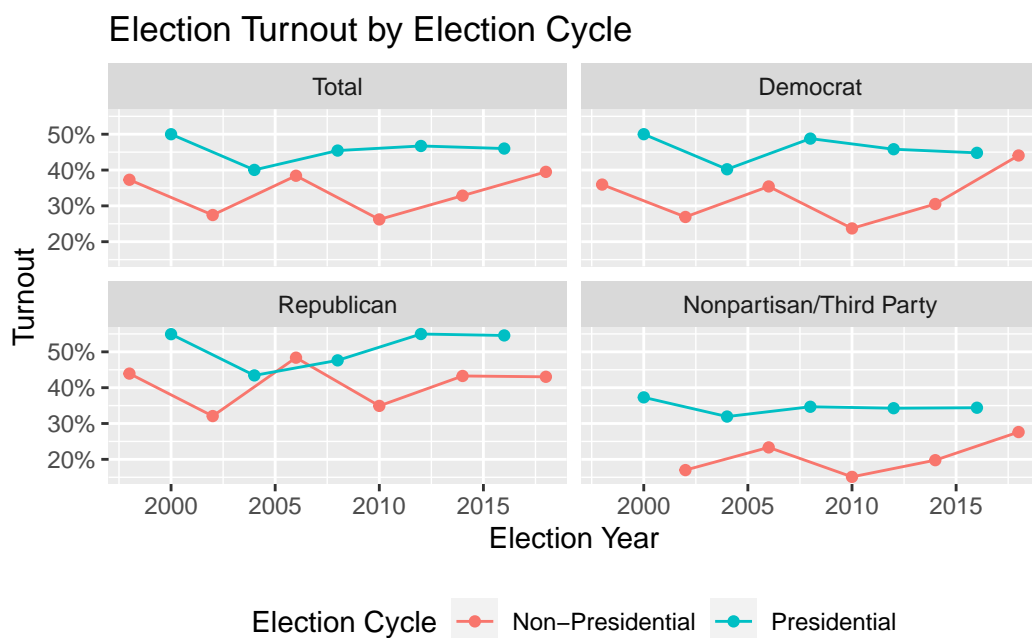


Figure 6: Plot of election turnout by election cycle faceted by party registration. Election turnout is consistently higher during presidential election cycles both overall and across all party affiliations.

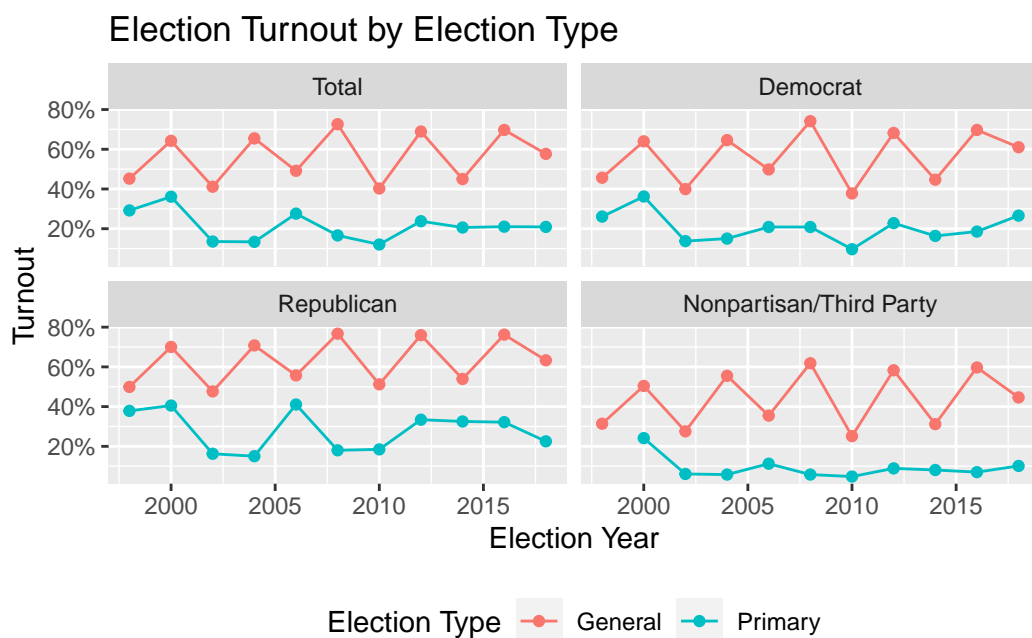


Figure 7: Plot of election turnout by election type faceted by party registration. Election turnout is consistently higher during general elections both overall and across all party affiliations. Overall, the highest turnout observed to date occurred during the 2008 general election.

4 Creating a Predictive Model

Given our findings on voter registration and turnout in certain areas of the county and across different types of elections, we sought to answer whether a predictive model of voter turnout could be generated that would have good predictive power. While our model is specific to Douglas County, a model that uses similar variables is likely to provide similar results. Although a number of predictive models could have been fit to this data set, we selected a linear regression model for its simplicity in training and ease of explanation.

The findings from our exploratory analysis indicated that the variables likely to be most important to prediction were the number of people registered to vote along with the election type and cycle. Certainly a wide variety of other exogenous variables could provide additional context to the model, including precinct-level trends and demographics. For instance, if we identify that Democrats tend to vote more than Republicans, precincts with more registered Democrats would tend to have higher voter participation. However, we take the position that a simple model with good accuracy can be applied easier and more consistently than a complex model that might only marginally increase predictive accuracy.

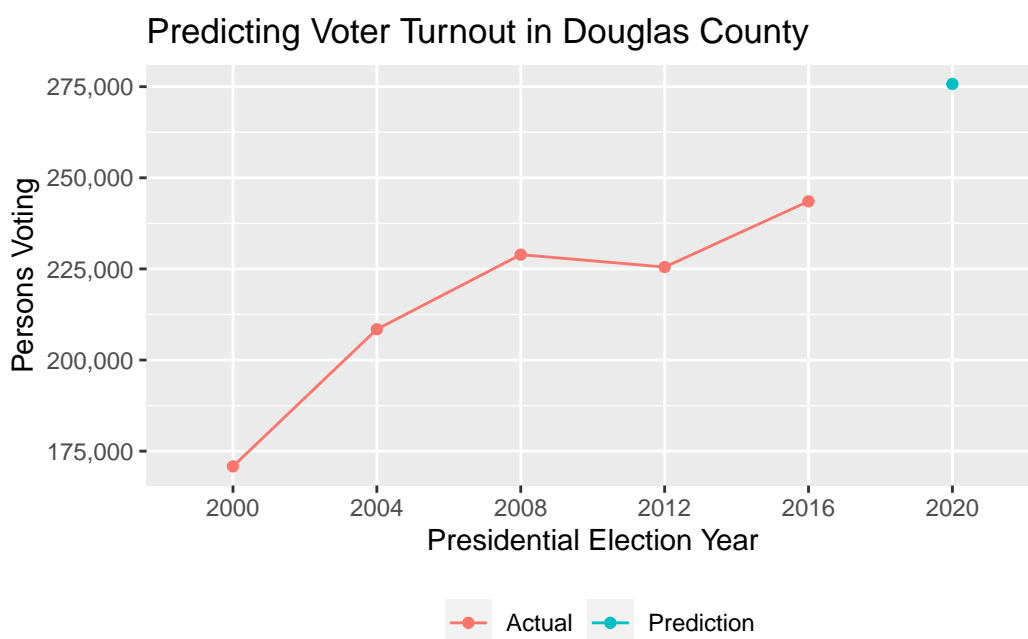


Figure 8: Plot predicting voter turnout in Douglas County for the 2020 presidential general election. A linear model was fit including the number of voter registrations, turnout in the primary election, and the change in voter registrations between the primary and general election to predict general turnout. The model predicted a sharp increase in turnout compared to the previous election.

Voter participation in presidential general elections is presented with the predicted value of our model in Figure 8. The total number of voters participating in the 2020 election as predicted by our model was 275,741. The model was trained and predicted prior to the election commission releasing their official turnout results. The actual number of people who voted in the most recent election was 278,509—a difference of only 2,768 voters compared to our prediction. Thus, this model could be employed as a starting-point for the Douglas County election commission—and perhaps other election commissions across the country—to predict voter turnout in presidential elections.

5 Discussion

With few exceptions, elections in Douglas County display relatively stable trends. Voter registrations have grown in line with population growth since 2006; however, similar growth in voter turnout has not been

realized. Instead, increases in the number of persons voting is likely to do more with an increase in population overall. Voters tend to turn out in higher numbers for presidential general elections, and turn out the least for non-presidential primary elections. Republicans tend to vote more consistently than Democrats and other nonpartisan/third party voters. A model trained on these factors was able to predict overall voter turnout in the 2020 presidential general election to within 3,000 voters. Election commissions that observe similar patterns in their electorate to those discussed in this paper could employ a similar model to predict turnout in their own elections.

5.1 Limitations & Future Directions

A number of minor technical limitations did exist for this analysis—chiefly among them the lack of detail in recent Census Bureau data and the lack of historical Douglas County ward/precinct maps. A more thorough exploration of geographic features of these data was therefore limited. Furthermore, an analysis of how turnout may have an effect on the *outcome* of an election was not possible with this data set. Although Douglas County is the largest portion of Nebraska’s second congressional district, certain portions of Sarpy County to its immediate south are also included. Future reports could work to aggregate the data from the election commissions of both Douglas and Sarpy counties to make such an assessment.

With regard to our predictive model, the 2020 primary and general elections have been unlike any other in recent history, and certainly unlike any election included in this analysis in terms of the continuously fraught, polarized, and divisive nature of electoral politics leading up to election day¹. These factors are not easily included in any predictive model. It is certainly possible that the model presented in this paper over-predicts turnout, but, due to higher-than-average interest in the electoral process, more people participated in the election thus masking any over-prediction. This limitation could be explored in future work by including more historical elections allowing for a wider range of data to be split for model training and testing. Attempts to do so with the current data set would begin to run into issues with the $n - k$ assumption in linear modeling.

Finally, future work should take into greater account the demographic nature of voter registration and electoral participation; these factors are inexorably linked in US politics. Many election commissions keep more detailed information about voters and their individual participation in elections over time. These individual “voter files” could allow for more thorough analysis, including broader or more hyper-specific geographical trends that don’t fit well into a ward/precinct structure.

¹The authors are unable to provide a specific citation. To compile a sufficient list of citations supporting this point would require more time than the analysis itself.

Bibliography

- Ansolabehere, S. and Hersh, E. (2011, August). Gender, Race, Age, and Voting: A Research Note. *Proceedings of the American Political Science Association Annual Meeting*. Retrieved from <https://ssrn.com/abstract=1901654>
- ArcGIS. (n.d.). Shapefiles. Retrieved October 2020, from <https://doc.arcgis.com/en/arcgis-online/reference/shapefiles.htm>
- Douglas County Election Commission. (2020a). *Douglas County Election Voter Precinct Maps* [Data set]. Retrieved from <https://www.votedouglascounty.com/maps.aspx>
- Douglas County Election Commission. (2020b). *Douglas County Voter Turnout* [Data set]. Retrieved from https://www.votedouglascounty.com/voter_turnout.aspx
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10. Retrieved from <https://doi.org/10.32614/RJ-2018-009>
- R Core Team. (2020). R: A Language and Environment for Statistical Computing. Retrieved from <https://www.R-project.org/>
- RStudio Team. (2020). RStudio: Integrated Development for R. Retrieved from <https://www.rstudio.com/>
- United States Census Bureau. (2019). *2010 Census Public Use Microdata Area (PUMA) Reference Maps* [Data set]. Retrieved from <https://www.census.gov/geographies/reference-maps/2010/geo/2010-pumas.html>
- United States Census Bureau. (2020). *Public Use Microdata Sample (PUMS)* [Data set]. Retrieved from <https://www.census.gov/programs-surveys/acs/microdata.html>
- United States Census Bureau. (n.d.). *American Community Survey (ACS)*. Retrieved from <https://www.census.gov/programs-surveys/acs>
- Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21, 1–20. Retrieved from <http://www.jstatsoft.org/v21/i12/>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4, 1686.
- Wickham, H. and Bryan, J. (2019). readxl: Read Excel Files. Retrieved from <https://CRAN.R-project.org/package=readxl>
- Wickham, H. and Seidel, D. (2020). scales: Scale Functions for Visualization. Retrieved from <https://CRAN.R-project.org/package=scales>