THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF POLITICAL SCIENCE

Accessible Public Transportation and Costs of Voting:

Does Nearby Public Transportation Increase Voter Turnout?

CALEB J. KLEMICK
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees in Political Science and Advertising/Public Relations
with honors in Political Science

Reviewed and approved* by the following:

Eric Plutzer
Professor of Political Science
Thesis Supervisor

Michael Berkman
Professor of Political Science and Director, McCourtney Institute for Democracy
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

What factors impact a potential voter's decision to cast a ballot in a given election? This question is the subject of much examination in the field of political science. This analysis seeks to contribute to that body of work through an investigation into a factor that could increase voter turnout—proximity to public transit. Using a dataset containing voter records for a random sample of 1,500 voters from 16 counties in Pennsylvania and public transportation stop data, this study creates a number of ordinary least squares regression models to examine the impact of proximity to transit on voter turnout in the 2016 and 2018 general elections. An analysis of the regression models reveals that proximity to transit does have a modest but reliable and statistically significant impact on voter turnout in low-salience elections and a smaller but still statistically significant impact on voter turnout in high-salience elections.

TABLE OF CONTENTS

LIST OF TABLES ……..iii

ACKNOWLEDGEMENTS …………………………………………………………….iv

Chapter 1 Introduction 1

Chapter 2 Literature Review 3

Weighing Costs of Voting 3
Benefits of Accessible Polling Places...4
Benefits of Accessible Public Transportation...6
Methods of Quantifying Public Transportation Accessibility 8

Chapter 3 Theory and Hypotheses 11

Chapter 4 Data and Methodology 13

Cases and Time Period...13
Voter Data...13
Experimental Variables...14
Individual-Level Control Variables..16
ZIP Code Tabulation Area-Level Control Variables..19
County-Level Control Variables...20
Regression Model 22

Chapter 5 Results and Analysis 23

Hypothesis 1 - Distance to Transit and Voter Turnout..23
Hypothesis 2 – Distance to Transit, Voter Turnout, and Socio-economic Status............26

Chapter 6 Conclusion...30

Appendix A Data 33

Experimental Dataset...33
10 Bin Regression Table..34

ACADEMIC VITA...39

LIST OF TABLES

Table 1. Individual-Level Descriptive Statistics 17

Table 2. ZIP Code Tabulation Area-Level Descriptive Statistics (for 569 ZCTAs)............... 19

Table 3. County-Level Descriptive Statistics (for 16 PA Counties) 21

Table 4. Regression Model Results Table.. 23

Table 5. Regression Model Results Table, Grouped by Median Household Income (for 2016) 27

Table 6. Regression Model Results Table, Grouped by Median Household Income (for 2018) 28

ACKNOWLEDGEMENTS

I would like to begin by thanking my thesis supervisor, Dr. Eric Plutzer, for his sage guidance, and constant support and understanding throughout this process. I would also like to thank my honors advisor and workshop professor, Dr. Michael Berkman, who provided insight and encouragement that helped me complete my thesis.

I'm also extremely grateful for my family and all my friends who tolerated long conversations with me about public transportation while I was working on this. Finally, I'd like to thank Alyssa, without whom I would not have finished this project.

Chapter 1

Introduction

In the 2016 United States presidential election, approximately 55.7% of America's voting age population cast a ballot. This 55.7% turnout percentage put the United States in 30th place out of 35 in a comparison of Organization for Economic Co-operation and Development countries conducted by the Pew Research Center (DeSilver, 2020). This percentage places the United States behind Japan, Australia, Germany, France, Mexico, Canada, and the United Kingdom, among many others, in terms of national voter turnout (DeSilver, 2020).

Low voter turnout can undermine government legitimacy (Goodman & Stokes, 2020), limit the ability of citizens to influence their government (Lijphart, 1997) and lead to a non-representative result that is biased toward people of a high socio-economic class (Lijphart, 1998). Given that responsive elected officials and representative election results are important features of a healthy democracy, and the United States' lagging voter turnout numbers when compared to its peers, it's logical that examinations into voting reforms and voter turnout are incredibly common among American political scientists (Aldrich, 1993; Grofman, 2016).

Attempting to explain macro-level voter turnout trends has proven to be a particularly thorny issue—it is clear is there are a variety of factors that go into an individual voter's decisions, however, one common hypothesis in this type of study is that the monetary/opportunity costs associated with voting play a role in an individual's decision to cast a ballot or not (Aldrich, 1993; Edlin et al., 2007). That is, in theory, as it gets easier to vote, more people will turn out, and as it gets harder to vote, less people will turn out.

Reforms hoping to exploit this phenomenon and increase turnout primarily aim to make voting more convenient through changes to electoral institutions like allowing early voting, expanding mail-in ballots, and making voter registration easier (Brennan Center for Justice, 2021). There are, however, other types of reforms that could achieve an increase in voter turnout through decreasing costs of voting that are worth considering. Take, for example, increasing the availability of public transportation. There is a variety of literature that examines the ways in which public transportation (PT) can have an outsized impact on the health, mobility, and economy of the cities they serve (Kenyon 2002, 207; Saif et al. 2019, 37; Bok and Kwon 2016, 1) by providing a reliable, affordable way to get around. Since there is an abundance of literature on the ways in which public transport can improve a variety of societal outcomes in daily life, it stands to reason that public transportation could also play a role in helping to improve political outcomes such as increased voter turnout and decreased costs of voting. In this thesis, I seek to test that argument by examining the impact the distance between an individual's home and the nearest public transportation station has on their voter turnout in the 2016 and 2018 general elections.

If there is a relationship between accessible public transportation and increased voter turnout, this research would give the field a new factor to consider when studying turnout, as well as identify a simple, original, and effective way to get more people to the polls while also making the city healthier, wealthier, and more sustainable.

Chapter 2

Literature Review

Weighing Costs of Voting

In 1957, American economist Anthony Downs published a book entitled *An Economic Theory of Democracy*, which contained a theory regarding how voters make the decision about whether or not to vote in a given election. Downs posited that potential voters weigh the perceived benefits of voting against the anticipated costs to determine whether they will cast a ballot (Sigelman & Berry, 1982). In his theory, Downs surmised that the potential benefits of voting were related to the civic value of voting (ensuring the continued health of the electoral system through participation), as well as seeing one's desired outcome in the election and reaping the benefits that come from that preferred outcome (policy preferences, ideological representation, etc.) (Downs, 1957b). The principal costs he discussed were opportunity costs as voting and preparing to vote, by informing oneself about the candidates and registering to vote, both take time (Downs, 1957b). According to Downs, voters consider the total value of all the benefits of voting and the likelihood that their vote will affect the outcome of the election against all the costs, and, "if the returns outweigh the costs, he votes; if not, he abstains" (Downs, 1957b, p. 260).

Downs' general theory of voting, that a voter decides whether to vote by weighing the perceived costs against the perceived benefits, has been widely accepted by the field as a tool for explaining voter behavior and has been expanded on by a number of scholars (Brady & McNulty, 2011; Feddersen, 2004; Sigelman & Berry, 1982).

Benefits of Accessible Polling Places

Downs' theory can be easily applied to much of the prior work considered for this thesis because one of the primary concepts that ties many of these studies together is the idea that more accessible polling places will result in increased voter turnout by decreasing costs of voting (Schur et al., 2017) and difficulties finding or getting to a polling place lower turnout because of the added opportunity costs (Fauvelle-Aymar & François, 2015; Hershey, 2009).

In developing this theory, it's important to first recognize that accessibility challenges make up a small part of the variety of reported reasons eligible non-voters chose to stay home. According to a survey of over 18,000 eligible non-voters interviewed by the Census Bureau after the 2016 election, the prevailing reason for not voting was a dislike of the candidates or campaign issues (25% of respondents), followed by a lack of interest or feelings of individual insignificance (15% of respondents) (López & Flores, 2017). However, another important factor in the decision to vote or not is an assessment of the personal costs associated with registering to vote and casting a ballot (Downs, 1957b; Li et al., 2018).

A 2017 article by Schur et al. set out to examine the impact that polling place inaccessibility has on voter turnout, especially among people with disabilities. The authors rely on prior work to establish that people with disabilities are already less likely to vote in general and that problems finding or accessing a polling place reduce turnout as well (Schur et al., 2017). The authors theorize that people with limited mobility (not necessarily disabled) are less likely to turn out because of physical challenges (finding the polling place, standing in line for a long time, navigating flights of steps) and other potential challenges like trouble reading the ballot, and difficulty operating a digital voting machine (Schur et al., 2017). Researchers coded survey responses and used probit and ordinary least squares regression analyses to draw their conclusions. The authors found that the voter participation disability gap in the 2012 election was 5.7% and could not be entirely explained away by standard predictors of turnout they controlled for. Therefore, they concluded that polling place inaccessibility contributed to that gap in participation (Schur et al., 2017).

Similar to Schur et al., Christine Fauvelle-Aymar and Abel François set out to examine how polling place accessibility impacted voter turnout, but from a monetary/opportunity cost perspective. Specifically, the researchers were examining how concurrent vs separate elections impact voter turnout. They felt turnout would be higher in districts with concurrent elections because it is easier for people to organize mobilization efforts and reduces the costs of voting. The authors tested this theory using an idiosyncrasy in the French electoral process in which some years, about half of France's electoral districts hold their regional and departmental elections at the same time, while the other half of districts only participates in regional elections and have departmental elections three years later (Fauvelle-Aymar and François, 2015). After conducting their analysis, the authors found that turnout is higher for concurrent elections because of more coordinated mobilization efforts and lower costs, especially transportation costs (Fauvelle-Aymar and François, 2015).

Increased voting costs were also found to impact voter turnout in the 2009 literature by Marjorie Hershey, in which she sought to build on previous findings that increased costs of voting lower voter turnout by looking into what particular voting costs affect what types of individuals and under what circumstances. The author theorizes that most voting costs affect people of color and low socioeconomic status, "but the burden of added costs could also disproportionately affect some ethnic, language, and age groups as well as Americans with disabilities and the residentially mobile" (Hershey, 2009, p. 87). Hershey specifically looks at how needing to arrange private transportation to and from a voter registration office, polling place, or other government office associated with the voting process can discourage some, and totally disqualify others, from being able to vote (Hershey, 2009). After compiling this body of work, the author found support for the claim that increased costs of voting decrease turnout, but interestingly, Hershey found, "less evidence, however, that reducing the costs of voting necessarily increases the turnout of these groups, probably because advance registration and photo-ID rules are only two of multiple burdens on their likelihood of voting" (Hershey, 2009, p. 90).

When considered collectively, this second group of studies lends credibility to the idea that voter turnout will be lower if voting costs are higher and also lends slightly more nuanced support to the idea that voter turnout will be higher if costs of voting are lower and polling places are more accessible.

In addition to a consensus regarding the potential added costs imposed by the electoral process, there is also scholarly support that these costs are not evenly distributed amongst the electorate. Hershey found mixed support for the idea that voter ID laws disproportionately affect Black, Latinx, and Asian-American voters and fairly clear support for the idea that there are substantive racial differences in the administration of election laws that impose costs on voters (Hershey, 2009). The findings of a 15-state study published in 2008 studying the administration of state voter ID laws comported with Hershey's assertions. The Alverez et al. study found that Black and Hispanic voters were approximately 14 to 19 percent more likely to be asked to produce photo ID on Super Tuesday 2007 and 2008, when compared to white voters, regardless of the state's voter ID laws (Alvarez et al., 2009).

In the Census Bureau's non-voter survey, 14% of respondents reported not voting because they were too busy, 12% had an illness or disability that prevented them from voting, 7% had transportation or voter registration problems, and 2% said they were unable to vote because their polling place was in an inconvenient location. (López & Flores, 2017). The fact that these logistical and accessibility challenges affected 35% of the sample again lends credibility to the idea that there are a number of formal and informal costs associated with the electoral process that can negatively impact the likelihood of some voters casting a ballot.

Benefits of Accessible Public Transportation

The other concept that ties much of this literature together is the finding that that an efficient, accessible public transportation system can have noticeable benefits for individual riders and the community at large (Kenyon, 2002; Saif et al., 2019; Bok and Kwon, 2016).

A 2002 literature review by Kenyon et al. examined a potential correlation between lack of mobility and social exclusion. The authors theorize that people without access to adequate mobility are likely to lack access to "opportunities, social networks, goods and services" (Kenyon et al., 2002). After giving robust and detailed definitions for social exclusion, mobility, and mobility-related exclusion the authors began a detailed explanation, drawing on a wide body of previous literature, to explain why accessible transport (both public and private) can alleviate some social exclusion by facilitating increased mobility for everyone (Kenyon et al., 2002).

Similarly, the 2019 Saif et al. literature review examined available scholarly research related to the study of public transport accessibility, how it can be measured, and the impact it can have on societal outcomes. The study consisted of an in-depth examination and discussion of the most important findings from seven articles studying public transportation. The authors found that the body of literature they reviewed supported claims that more accessible public transport is more attractive to citizens so, therefore, the level of accessibility should be of paramount importance when considering changes in transport policy (Saif et al., 2019). They also found that more accessible public transport provides more economic, educational, and social opportunities to its users (Saif et al., 2019).

The 2016 study by Jinjoo Bok and Youngsan Kwon briefly discusses the economic, environmental, and sustainability benefits of accessible public transportation and comes to similar conclusions as Kenyon et al. and Saif et al. about the positive impact that public transportation can have on societal outcomes (Bok and Kwon, 2016).

All three of these three studies establish public transportation systems as a service that can improve important areas of individual and community life. They specifically show PT benefits economic, public health (Bok and Kwon, 2016), social (Kenyon, 2002), and environmental outcomes (Saif et al., 2019).

Methods of Quantifying Public Transportation Accessibility

Bok and Kwon also propose a new way to compare the accessibility of PT systems across urban areas using Google's General Transit Feed Specification format, which is a standard way of publicly sharing transportation data between a wide variety of transportation agencies. In order to accomplish their goal, they collected all relevant data for the urban areas they were interested in studying: Washington, D.C; Chicago, IL; Portland, OR; Vancouver, Canada; Toronto, Canada; and Daejeon, South Korea. The researchers were able to use the data to assemble a complete, detailed analysis of the public transportation systems in all six case cities both in terms of infrastructure (number of stops, vehicles, transportation modalities) and access (share of population near a PT stop, the share of geographical area near a PT stop). They found that, on average, American functional urban areas (FUAs) score much worse in both accessibility (length of wait) and access (share of population near a stop) than European cities because Americans are much more reliant on personal vehicles and the efficiency of US public transit services reflect this fact (Bok and Kwon, 2019).

The Bok and Kwon study also shares the same unit of analysis and general metric for PT accessibility as the 2020 study by Verbavatz and Barthelemy. Verbavatz and Barthelemy and Bok and Kwon both used a spatial threshold-based measure of public transportation accessibility.

The 2020 study by Verbavatz and Barthelemy, "Access to Mass Rapid Transit in OECD Urban Areas", set out to provide a standard measure of public transport accessibility that can be used in a variety of studies. To accomplish this goal, the authors modified an existing measure called the People Near Transit (PNT) metric, originally developed for a different study. PNT uses the FUA unit of analysis and quantifies accessibility by reporting the proportion of a population that lives within 500m, 1000m, or 1500m from a public transportation station. In order to form their PNT, the researchers first acquired data on global population distribution disaggregated from national censuses and mapped those data onto a collection of 250 square meter "population tiles," which represent the Earth's land area (Verbavatz and Barthelemy, 2020). The researchers then mapped those population tiles onto a map of global FUAs, an

OECD standard international definition for a city, which is the urban core plus its commuting zone (Verbavatz and Barthelemy, 2020). Once the FUAs of interest were identified, the researchers overlaid a map with the public transit routes and stops onto the gridded population tiles and used that combined map to do the PNT calculations in Python (Verbavatz and Barthelemy, 2020). There is one notable caveat to the usefulness of the dataset provided by these researchers, though. The PNT method as modified by Verbavatz and Barthelemy only includes tram, streetcar, or light rail services; Subway, Metro, or any underground service; and suburban rail services in its definition of public transport. This excludes busses, a modality that accounts for approximately half of all public transportation ridership in the United States (American Public Transport Association, 2020). Given the popularity of busses in the United States public transportation market, it will be important for this research to take them into account when calculating accessibility.

Bok and Kwon used a spatial-temporal threshold of no more than 5 minutes walking time to a road-based PT medium and 10 minutes walking time to a rail-based PT medium as their threshold for accessible public transport, while Verbavatz and Barthalemey were focusing on travel distance. Though constructed slightly differently, the variables are attempting to measure the same things. It's clear a spatial-threshold model is not the universal metric of choice, though.

In 2007, Damalesh Abreha set out to examine the efficiency of public transport systems and identify opportunities and challenges associated with the high cost of building, maintaining, and expanding PT networks in the developing world. The researcher investigates the case of Addis Ababa, Ethiopia, a city of 3.3 million people in which urban bus service is provided by a company called Anbessa (Abreha, 2007). The author then took all the data they collected to build a model using descriptive statistics and GIS-analysis to quantify the overall performance of Addis Abba's PT network based on the following criteria: infrastructure availability, physical proximity, quality of service, degree of equity, and customer satisfaction (Abreha, 2007). The author found that many of the variables they used to quantify performance are interrelated; for example, lack of functional infrastructure can mean more traffic on roads in good repair, which can cause congestion, which decreases quality of service, and in turn, decreases customer satisfaction. Abreha concluded that the best way to assess PT accessibility is with a combined metric of multiple variables, not just one measure of ridership, physical proximity, etc. (Abreha, 2007).

Abreha's position that merely measuring physical proximity to a PT stop is insufficient to effectively determine PT accessibility seems to be in the minority. Many authors, like the ones discussed above, have found success conducting PT accessibility studies using a quasi-PNT metric. However, Abreha's argument is a needed reminder that there are a lot of inter-related factors that go into these issues, so it will be important to control for as many confounding variables in the model as possible in order to lend more credibility to the results.

Chapter 3

Theory and Hypotheses

The previous section illustrates that there is scholarly support for the idea that voters decide whether to cast a ballot or abstain from a given election by weighing the perceived benefits of voting against the anticipated costs, and that, in general, as costs of voting increase, voter turnout will decrease. Some of the costs of voting discussed were related to informing oneself about the candidates, completing the voter registration process, and finding and accessing one's polling place.

The review of prior literature also showed that, in general, transportation systems provide the basis on which many societies develop and interact socially (Abreha, 2007) and influence how people move throughout a community (Murray et al., 1998).

In addition to the collective societal benefits efficient transportation systems provide, there are substantial benefits for individual community members if they have access to an efficient means of public transportation as well. There is scholarly consensus that access to a public transportation network provides access to educational, employment, and recreational opportunities (Murray et al., 1998); improves access to other public resources (Abreha, 2007); and provides a means for people to carry out activities in their daily life (Abreha, 2007).

So generally, efficient public transportation is designed to provide access to other places, people, and opportunities for its users. That fact is important in the context of this research because more accessible polling places could potentially increase turnout and difficulties finding or getting to a polling place lowers turnout (Schur et al., 2017).

Given all the significant ways a robust public transportation can improve the environmental, economic, and public health conditions of a city, and given that citizens rely on the system to get to and from, work, school, and other commitments in daily life, it stands to reason that a strong PT system could improve voter turnout by facilitating easy, affordable access to a polling place on Election Day.

However, the people that would be most reliant on transit to facilitate their movement to and from a polling place obviously have limited mobility without access to public transportation. Given this, the distance from an individual's residence to the nearest public transportation stop impacts the costs of voting they incur and how much access to public transportation mitigates those costs. Therefore, the distance between a voter's residence and the nearest PT stop will be an important factor in determining the magnitude of this effect, which leads to the second hypothesis for this study.

H_1: The closer an individual's home is to the nearest public transportation stop, the greater the likelihood they turn out to vote.

Finally, while costs of voting are primarily time and opportunity costs, there are also direct monetary costs sometimes associated with obtaining voter identification and transporting oneself to the voting precinct and any other official location associated with the registration and voting processes that some voters cannot afford. In places without strong public transport, the effect of needing to arrange private transportation can deprive some of the ability to vote (Hershey 2009, 90). These costs are most commonly insurmountable for low-income voters, who have less disposable income to cover monetary costs (voter identification fees, transportation costs, etc.), and a decreased ability to withstand the opportunity costs (taking off work to travel to their polling place, etc.). Given this, the second hypothesis for this research is below.

H_2: The positive impact of accessible public transportation on voter turnout will be most pronounced in low-income communities.

Chapter 4

Data and Methodology

Cases and Time Period

The empirical analysis for this research consists of 24,000-n ordinary least squares regression models with public transportation accessibility as the primary independent variable and voter turnout in the 2016 and 2018 general elections as the dependent variable. The models also include a number of control variables measured at the individual, zip code tabulation area, and county levels. Both experimental variables are measured at the individual level and each case is a registered voter from one of 16 counties in Pennsylvania for which public transportation stop location data were readily available. Those counties are Adams, Allegheny, Beaver, Bucks, Butler, Centre, Chester, Dauphin, Delaware, Erie, Lehigh, Monroe, Montgomery, Northampton, Philadelphia, and York.

The control variables were collected at three different levels of analysis because the degree of granularity available for each variable varied. The level of analysis for each control variable was as precise as possible at the time of collection. The only variables available at the individual-level were contained in the Full Voter Export dataset, the ZIP Code Tabulation Area data were all collected from the US Census Bureau, and the county-level data was collected from the Pennsylvania Department of Motor Vehicles and the Center for Public Integrity, a Washington DC-based journalism non-profit (Center for Public Integrity, 2021).

Voter Data

All voter data for this study were sourced from the October 11th, 2021 version of the Pennsylvania Department of State's Full Voter Export. The Full Voter Export is a dataset containing the name, voter ID, gender, date of birth, date registered to vote, registration status, date of last registration status change, political party, residential address, mailing address, polling place, date last voted, and voter history for a selection of local, state, and federal primaries, general elections, and special elections of every registered voter in each of Pennsylvania's 67 counties. The data within the Full Voter Export is collected by individual county boards of election and submitted to the PA Department of State for compilation into the whole-state dataset, with the voter records for all of Pennsylvania's approximately 8.7 million registered voters, grouped by county, contained within.

For the purposes of this study, the only variables taken from the Full Voter Export were voter ID, gender, date of birth, political party, residential address, date last voted, 2012 general election turnout, 2014 general election turnout, 2016 general election turnout, and 2018 general election turnout. For each of the 16 counties this study is examining, 1,500 voters were randomly selected by importing each county voter export dataset into RStudio and using the 'sample_n' function from the 'quanteda' package, set to seed '2468'.

Experimental Variables

The dependent variable for this study is individual-level voter turnout for the 2016 and 2018 general elections. As The Full Voter Export reports whether or not a voter turned out in 40 previous elections, it was important to start by indexing out all the irrelevant voter history. To begin this, I used the election map files for each county to determine which of the 40 turnout entries corresponded to the 2016 and 2018 elections, then removed the 36 other elections (excluding 2012 and 2014, which are being used as control variables) from the experimental dataset using indexing in RStudio.

In the FVE, turnout is reported by using a two-character string to denote if and how a person voted in a given election. If the entry for a given election is blank, that means the voter did not vote. If the cell has a two-character string in it, that means they did vote and the letters denote how the ballot was cast. (For example, "AP" means appeared in-person and "MB" means mail-in ballot.) In order to make empirical analysis possible and more straightforward, the FVE turnout variables were converted to a binary variable with a 0 meaning that a voter did not cast a ballot and a 1 meaning they did. The specific medium by which each voter cast their ballot was not recorded for this study, so a 1 for turnout could mean a voter cast their ballot in-person or through the mail. The choice was made to study turnout for both the 2016 presidential election and the 2018 mid-term congressional elections in order to study if/how election salience alters the impact accessible public transportation has on voter turnout.

The dependent variable was also binned into quintiles using the R function ntile() to more clearly see the nature of the relationship and see at what distances, if any, public transportation could positively impact voter turnout. The fifth bin, containing the voters who live furthest away (over 3.32 miles) from the nearest public transport stop, has not been included in order to make comparison between bins possible.

The primary independent variable for this study is public transportation accessibility, which is measured as the straight-line distance between a registered voter's residential address and the closest bus, subway, or light rail station. In order to measure the accessibility of public transportation within these FUAs, it will be important to gather reliable data about public transportation service within that city. Since 2008, the most common format for sharing public transportation data has been Google's General Transit Feed Specification, or GTFS (Verbavatz and Barthelemy 2020, 3). GTFS data usually consists of a .zip archive containing several .csv files that report general information about the transit agencies, schedule data, and the coordinates for every stop on each route the transit agency operates.

The GTFS data underwriting this accessibility metric were sourced from OpenMobilityData. OpenMobilityData is a free repository of public transportation data accessible at transitfeeds.com and maintained by Canadian-based non-profit MobilityData IO. OpenMobilityData currently hosts GTFS data for 1297 public transportation providers in 675 locations in more than 50 countries (MobilityData IO n.d., 1). For many agencies, OpenMobilityData has all this information across multiple years. This repository was the obvious choice for sourcing public transportation data because it is the largest source of GTFS files currently available. It is also mentioned in numerous empirical studies examining public transportation (Verbavatz and Barthelemy 2020, 4). Additionally, OpenMobilityData is recommended for GTFS developers by Google, the entity primarily responsible for maintaining the GTFS format (Google Developers 2020).

Once all the available Pennsylvania public transport stop location data were downloaded from OpenMobilityData, the coordinates of each stop were imported into an ArcGIS Pro map alongside the residential address for all 24,000 registered voters who were randomly selected for inclusion in this study. For each voter, the straight-line distance from their public transport stop was calculated using the “Find Nearest” tool in ArcGIS Pro. Once ArcGIS had calculated the nearest stop to each voter’s residence, the straight-line distance for each voter was exported from ArcGIS Pro, imported back into RStudio, matched to the appropriate voter by voter ID, and bound to the primary experimental dataset.

Individual-Level Control Variables

The individual-level control variables included in this study are age, political party, gender, 2012 general election turnout, and 2014 general election turnout. All of these control variables rely on data from the Pennsylvania Full Voter Export as there was no practical way to collect new information about all 24,000 study participants. Measures of central tendency and dispersion for all five of these variables are shown below.

Table 1. Individual-Level Descriptive Statistics

Variable	Valid N	Mean	Median	St. Dev.	Min	25th %tile	75th %tile	Max
2018 election turnout	24000	0.527	1	--	0	0	1	1
2016 election turnout	24000	0.607	1	--	0	0	1	1
2014 election turnout	24000	0.328	0	--	0	0	1	1
2012 election turnout	24000	0.491	0	--	0	0	1	1
age	23997	50.827	51.021	18.604	18.425	34.718	65.222	103.144
political party	24000							
...dem	10976	45.7%						
...ind	3231	13.5%						
...oth	605	2.5%						
...rep	9188	38.3%						
gender	24000							
...NA	2991	12.5%						
...fem	9607	40%						
...male	8676	36.1%						
...oth	2726	11.4%						

Age

As is shown in Table 1, all individual-level control variables have observations for the entire sample except age, which is missing three observations. The age of those three voters were excluded because they were outside the range of possible voting ages (one voter was reported as one year old, and two voters were reported as over 150 years old). This was seemingly the result of a clerical error during the data compilation process because all three voters had voted in an election since 2000.

As the FVE dataset only gave each voter's date of birth, age was calculated using RStudio's difftime() function, which calculated the time difference in weeks between their date of birth and March 12th, 2022 (the date the analysis was completed) and then divided each difftime() output by 52.25 to get each voter's age in years.

Political Party

The data for the political party variable were collected from the Pennsylvania Full Voter export. In the FVE dataset, there were 30 unique responses to the political party variable which ranged from the major political parties to incredibly small, fringe parties. In order to make the analysis more manageable, all the minor parties were aggregated into a single *other* category, while *democrat* and *republican* stayed as their own individual response. The *independent* category is an aggregation of the "independent" and "no party affiliation" responses. *Independent* is the comparison category for the *democrat, republican,* and *other* responses.

Gender

The data for the gender variable were again sourced from the Pennsylvania Full Voter export. There are four possible response categories—*male*, *female*, *unspecified*, and *no response*. The fourth group, *no response*, is present because responding to the gender question is not required to successfully complete a voter registration application. The decision was made to include the *no response* category in this analysis because of the relatively substantial portion of the sample that did not respond to the gender question. Voters who did not respond to the gender question are the comparison category for *male, female, and unspecified.*

2012/2014 Election Turnout

Like the experimental turnout variable, the 2012 and 2014 election turnout data were sourced from the Pennsylvania Department of State and were measured at the individual-voter level. Also similar to the experimental variable, turnout was recorded using two-character strings, all of which were recoded as a 1, while NAs (representing an individual who did not turn out) were recoded as a 0. These data were included in the model in order to capture a voter's political interest, which is another import factor in the a potential voter's decision to actually cast a ballot or not.

ZIP Code Tabulation Area-Level Control Variables

The ZIP Code Tabulation Area (ZCTA)-level control variables included in this study are median household income, proportion of residents in poverty, and proportion of white residents. ZIP Codes do not have defined areal boundaries but are instead a group of mail delivery routes; therefore, they cannot reliably be used to compare different zip code areas to each other. ZIP Code Tabulation Areas were created by the US Census Bureau to serve as approximate representations of each ZIP Code service area to make this type of comparison possible (US Census Bureau, 2021). Summary statistics for all three variables are below.

Table 2. ZIP Code Tabulation Area-Level Descriptive Statistics (for 569 ZCTAs)

Variable	Valid N	Mean	Median	St. Dev.	Min	25th %tile	75th %tile	Max
2016 median household income (in 10,000s of dollars)	23817	6.463	5.946	2.168	1.063	5.128	7.557	21.144
2016 proportion of residents in poverty	23944	0.117	0.090	0.089	0.000	0.055	0.142	0.670
2016 proportion of white residents	23958	0.820	0.882	0.186	0.022	0.785	0.936	1.000
2018 median household income (in 10,000s of dollars)	23802	6.947	6.473	2.351	1.105	5.460	8.105	25.000
2018 proportion of residents in poverty	23944	0.112	0.085	0.087	0.000	0.051	0.135	0.622
2018 proportion of white residents	23958	0.814	0.865	0.185	0.028	0.779	0.930	1.000

Table 2 shows that none of the three ZCTA-level variables have observations for all 24,000 cases. This is due to the difference between ZIP Codes and ZCTAs and the way in which the Census Bureau creates ZCTAs. In order to approximate the geographic area of ZIP codes, the Census Bureau examined all the addresses in each census block in order to find the most common ZIP code in each block and assign that ZIP code to the entire block. The Census Bureau then aggregated the blocks with common assigned ZIP Codes together to create larger areas (US Census Bureau, 2021). This method means that it is possible for a given area to have a ZCTA that is different from its ZIP codes, and it is also possible for some places to have no ZCTA at all. This lack of a ZCTA is most common for areas of land or water without ZIP codes but was also common for sparsely populated ZIP codes (US Census Bureau, 2021). The latter explains the vast majority of the missing cases for these variables.

Variable Construction

All three variables rely on data from the US Census Bureau's American Community Survey. For each variable, the one-year American Community Survey estimate for every ZCTA in Pennsylvania for the years 2016 and 2018 were downloaded and imported into Microsoft Excel. Using Excel's vlookup() function, the ZIP code for each resident in the experimental dataset was matched with the corresponding ZCTA value. Once matched, the ZCTA values for median household income were divided by 10,000 and the values for poverty and white residents, originally reported as percentages, were divided by 100. This was done in order to make the regression coefficients for these variables easier to interpret. After transformation, the values were important into RStudio and bound to the experimental dataset.

County-Level Control Variables

The two variables measured at the county-level that are included in this research are the number of registered passenger vehicles per capita and the number of polling places. Descriptive statistics for both variables are included below.

Table 3. County-Level Descriptive Statistics (for 16 PA Counties)

Variable	Valid N	Mean	Median	St. Dev.	Min	25th %tile	75th %tile	Max
number of registered passenger vehicles per capita	24000	0.614	0.647	0.081	0.415	0.565	0.673	0.696
number of polling places	24000	353.06	160	458.47	52	120.5	336.75	1703

Passenger Vehicles Per Capita

The data for the passenger vehicles per capita variable were sourced from the Pennsylvania Department of Motor Vehicles' Annual Report of Registrations, which details the number of active Pennsylvania vehicle registrations, by county, at the end of each year. The total number of passenger vehicles for each of the 16 counties included in this study was recorded, inputted into RStudio, divided by the total population of each county according to the US Census Bureau, and bound to the experimental dataset. Transforming this into a per capita variable was done in order to make the regression model easier to interpret.

Number of Polling Places

The polling place data were sourced from the Center for Public Integrity, a Washington DC-based investigative journalism nonprofit that focuses on inequality. The dataset contains the name, address, and precinct ID, for every polling place in every federal general election in Pennsylvania since 2012. The Center for Public Integrity acquired the data from the 2012 and 2014 general election from the Voting Information Project, a partnership between a number of state governments and Democracy Works, a nonpartisan election information nonprofit (Rebala, 2021; Voting Information Project, 2021). The data from the 2016, 2018, and 2020 elections were sourced from the Pennsylvania Secretary of State's office through a "Right to Know" public information request (Rebala, 2021).

Regression Model

After all the data were collected, homogenized, and compiled, multiple ordinary least squares regression models were constructed using RStudio to test the asserted hypotheses. There are a total of 24,000 cases in each model. The dependent variable in both models was voter turnout in the 2016 and 2018 elections, and the primary independent variable was public transportation accessibility, which has been binned into quintiles. The first model contains all the experimental and control variables as described above in order to test whether distance to transit has an impact on voter turnout. The second model has median household income removed and splits cases into thirds according to median household income. The low, medium, and high household income subsamples will then each be run through a model to test if the effect is more pronounced for low-income voters.

Chapter 5

Results and Analysis

Hypothesis 1 - Distance to Transit and Voter Turnout

The results from the first set of regression models are below in Table 4. The first model has 2016 general election voter turnout as the dependent variable and the second model has 2018 general election voter turnout as the dependent variable. Both models have distance to transit as the primary independent variable. In order to appropriately interpret the results, it's important to remember that the distance variable was binned into quintiles and the fifth bin is not included in the model, so the coefficients seen for distances bins in Table 4 are a comparison between the given bin in the model and the reference bin. Distance bin 5 contained the voters who lived furthest from transit (between 3.32 and 26 miles).

Table 4. Regression Model Results Table

2016 and 2018 Election Regression Results (with Prior Turnout) 5 Bin

	Dependent variable:	
	2016 Voter Turnout	2018 Voter Turnout
	(1)	(2)
distance bin 1 (0.000-0.079 mi)	-0.016^{*}	0.012
	(0.010)	(0.010)
distance bin 2 (0.079-0.234 mi)	-0.002	0.023^{**}
	(0.008)	(0.009)
distance bin 3 (0.234-0.816 mi)	0.013^{*}	0.038^{***}
	(0.008)	(0.009)
distance bin 4 (0.816-3.316 mi)	0.016^{**}	0.030^{***}
	(0.008)	(0.009)
age	0.002^{***}	0.003^{***}
	(0.0002)	(0.0002)
gender (female)	0.077^{***}	0.057^{***}

	(0.008)	(0.009)
gender (male)	0.060***	0.068***
	(0.008)	(0.009)
gender (unspecified)	0.081***	0.057***
	(0.011)	(0.012)
political party (democrat)	0.056***	0.097***
	(0.008)	(0.009)
political party (republican)	0.060***	0.063***
	(0.008)	(0.009)
political party (other)	0.014	0.006
	(0.017)	(0.019)
2016 median household income (in 10,000s of dollars)	0.002	
	(0.002)	
2016 proportion of residents in poverty	-0.046	
	(0.050)	
2016 proportion of white residents	0.024	
	(0.020)	
2018 median household income (in 10,000s of dollars)		0.010***
		(0.002)
2018 proportion of residents in poverty		-0.080
		(0.055)
2018 proportion of white residents		0.033
		(0.021)
number of registered passenger vehicles per capita	-0.116***	-0.156***
	(0.044)	(0.048)
number of polling places	0.00000	-0.00000
	(0.00001)	(0.00001)
2012 election turnout	0.419***	0.244***
	(0.007)	(0.007)
2014 election turnout	0.181***	0.298***
	(0.007)	(0.007)
Constant	0.176***	0.044

	(0.044)	(0.047)
Observations	23,814	23,799
R^2	0.389	0.312
Adjusted R^2	0.389	0.312
Residual Std. Error	0.382 (df = 23795)	0.414 (df = 23780)
F Statistic	842.271*** (df = 18; 23795)	599.398*** (df = 18; 23780)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4 shows that distance to transit does have a statistically significant relationship to voter turnout in certain circumstances. For the 2016 model, distance bins 3 and 4 both have a positive, statistically significant impact on voter turnout. Specifically, voters in distances bins 3 and 4 are 1.3% and 1.6% more likely to cast a ballot in an election when compared to a voter from bin 5. In the 2018 model, bins 2, 3, and 4 have the same impact; voters in these groups are 2.3%, 3.8%, and 3.0% more likely to cast a ballot than a voter from bin 5. Additionally, some demographic controls including age, gender, political party, and prior election turnout were statistically significant in both models.

The smaller impact, both in magnitude and scope, of distance from transit on voter turnout for the 2016 general election makes sense and seems to indicate that election salience is another important factor in this relationship and a potential voter's calculus about whether or not they will actually cast a ballot. That is, the more prominent the election is or important it seems to a voter (like a presidential election), the more likely they are to put in effort to overcome barriers to voting like transportation. For lower salience, less prominent elections (like congressional midterms), this willingness is lessened as the benefits can seem less substantial to the voter, so they are much more sensitive to any time/opportunity costs associated with casting a ballot.

While it is clear distance to transit does have an impact on voter turnout, the direction and magnitude of this relationship are not consistent across the entire range of distances in the model. For example, Table 4 shows a slightly negative coefficient for the first distance bin in the 2016 model. I was unable to determine the specific reasons for that through examination of the dataset and regression model. The most likely explanation seems to be an anomaly in the distribution of voter addresses somewhere inside that bin. However, when looking at a map of the addresses in bin 1, I was not able to identify any major anomalies.

Even with the anomaly with distance bin 1 and the more limited impact of the relationship on high-salience elections, the positive, statistically significant impact of distance bin 2 in the 2018 model and distance bins 3 and 4 in both models is significant. Approximately 136.8 million Americans voted in the 2016 presidential election, a 1.5% percent turnout increase would have meant approximately 2.1 million more people cast a ballot (Clerk of the House of Representatives, n.d.). In 2018, a 3.8% increased would have brought 4.4 million voters to their polling place (Clerk of the House of Representatives, n.d.). Given this, the positive regression coefficients, and the high degree of statistical significance for the effects of bins 3 and 4, I am able to find mixed support for my first hypothesis.

Hypothesis 2 – Distance to Transit, Voter Turnout, and Socio-economic Status

In order to test my second hypothesis, which posited that proximity to transit will have a greater positive impact on turnout for voters of a low socio-economic status, the experimental dataset was subset by election year and then into thirds by median household income. A total of six models were created, one for each combination of year (2016 or 2018) and median household income (bottom, middle, or top 1/3). Once subset, all the regression models were fit and two summary table were created, one for 2016 and one for 2018, both of which are below in Tables 5 and 6.

Table 5. Regression Model Results Table, Grouped by Median Household Income (for 2016)

2016 Election Regression Results (Grouped into Thirds by Median HHI)

	Dependent variable:		
	2016 General Election Voter Turnout		
	(bottom 1/3 median HHI)	(middle 1/3 median HHI)	(top 1/3 median HHI)
	(1)	(2)	(3)
distance bin 1 (0.000-0.079 mi)	-0.047**	-0.026	0.002
	(0.020)	(0.019)	(0.021)
distance bin 2 (0.079-0.234 mi)	-0.019	-0.030**	0.007
	(0.019)	(0.015)	(0.016)
distance bin 3 (0.234-0.816 mi)	-0.019	-0.001	0.030**
	(0.020)	(0.014)	(0.014)
distance bin 4 (0.816-3.316 mi)	-0.009	0.010	0.021*
	(0.022)	(0.013)	(0.013)
age	0.002***	0.002***	0.002***
	(0.0003)	(0.0003)	(0.0003)
gender (female)	0.086***	0.094***	0.043***
	(0.015)	(0.015)	(0.015)
gender (male)	0.067***	0.076***	0.025*
	(0.015)	(0.015)	(0.015)
gender (unspecified)	0.107***	0.091***	0.052***
	(0.020)	(0.020)	(0.019)
political party (democrat)	0.093***	0.034**	0.041***
	(0.015)	(0.015)	(0.013)
political party (republican)	0.099***	0.045***	0.038***
	(0.016)	(0.015)	(0.013)
political party (other)	0.076**	-0.010	-0.036
	(0.031)	(0.030)	(0.030)
2016 proportion of residents in poverty	-0.093	-0.255*	0.114
	(0.060)	(0.150)	(0.183)
2016 proportion of white residents	-0.001	0.029	0.135*
	(0.027)	(0.046)	(0.073)
number of registered passenger vehicles per capita	-0.119	-0.036	-0.130
	(0.073)	(0.080)	(0.103)
number of polling places	-0.00001	0.00004**	-0.00003
	(0.00001)	(0.00002)	(0.00002)
2012 election turnout	0.402***	0.416***	0.436***
	(0.012)	(0.012)	(0.011)
2014 election turnout	0.185***	0.188***	0.169***
	(0.013)	(0.013)	(0.012)
Constant	0.233***	0.143*	0.136
	(0.066)	(0.079)	(0.108)
Observations	7,310	7,047	7,665
R^2	0.366	0.394	0.405
Adjusted R^2	0.364	0.392	0.404
Residual Std. Error	0.393 (df = 7292)	0.382 (df = 7029)	0.370 (df = 7647)
F Statistic	247.161*** (df = 17; 7292)	268.363*** (df = 17; 7029)	306.317*** (df = 17; 7647)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6. Regression Model Results Table, Grouped by Median Household Income (for 2018)

2018 Election Regression Results (Grouped into Thirds by Median HHI)

	Dependent variable:		
	2018 General Election Voter Turnout		
	(bottom 1/3 median HHI)	(middle 1/3 median HHI)	(top 1/3 median HHI)
	(1)	(2)	(3)
distance bin 1 (0.000-0.079 mi)	-0.005	-0.005	0.028
	(0.021)	(0.021)	(0.024)
distance bin 2 (0.079-0.234 mi)	0.015	-0.006	0.032*
	(0.021)	(0.016)	(0.018)
distance bin 3 (0.234-0.816 mi)	0.023	0.016	0.052***
	(0.021)	(0.015)	(0.015)
distance bin 4 (0.816-3.316 mi)	0.010	0.007	0.052***
	(0.023)	(0.014)	(0.014)
age	0.002***	0.002***	0.003***
	(0.0003)	(0.0003)	(0.0003)
gender (female)	0.052***	0.055***	0.065***
	(0.016)	(0.017)	(0.016)
gender (male)	0.080***	0.057***	0.068***
	(0.016)	(0.017)	(0.017)
gender (unspecified)	0.081***	0.060***	0.031
	(0.022)	(0.022)	(0.021)
political party (democrat)	0.105***	0.108***	0.076***
	(0.016)	(0.016)	(0.015)
political party (republican)	0.073***	0.073***	0.040***
	(0.017)	(0.016)	(0.015)
political party (other)	0.077**	0.019	-0.054
	(0.033)	(0.033)	(0.033)
2018 proportion of residents in poverty	-0.159**	-0.255	-0.308
	(0.066)	(0.187)	(0.215)
2018 proportion of white residents	0.046*	-0.027	0.105
	(0.028)	(0.049)	(0.081)
number of registered passenger vehicles per capita	-0.176**	-0.049	-0.376***
	(0.078)	(0.087)	(0.114)
number of polling places	-0.00001	0.00003	-0.00004*
	(0.00001)	(0.00002)	(0.00002)
2012 election turnout	0.224***	0.243***	0.257***
	(0.013)	(0.014)	(0.013)
2014 election turnout	0.321***	0.314***	0.267***
	(0.014)	(0.014)	(0.013)
Constant	0.136*	0.117	0.243**
	(0.070)	(0.083)	(0.116)
Observations	7,310	7,047	7,665
R^2	0.301	0.317	0.300
Adjusted R^2	0.300	0.315	0.298
Residual Std. Error	0.418 (df = 7292)	0.413 (df = 7029)	0.413 (df = 7647)
F Statistic	184.854*** (df = 17; 7292)	191.630*** (df = 17; 7029)	192.783*** (df = 17; 7647)
Note:			*p<0.1; **p<0.05; ***p<0.01

Tables 5 and 6 clearly show that the effect of proximity to transit on voter turnout does not grow in magnitude for voters of low socioeconomic status. In fact, it seems that the exact opposite is true. The regression coefficients for the distance bins slightly grow in positive magnitude and statistical significance as median household income increases. Therefore, I can confidently conclude that this analysis does not support hypothesis 2.

Chapter 6

Conclusion

This research set out to better understand the relationship between proximity to public transportation and voter turnout, drawing on a variety of literature about costs of voting, voter turnout, public transportation, and political psychology to posit that more accessible (closer) public transportation might lessen the costs of voting for some potential voters, which would in turn make it more convenient/possible for certain people to vote. Specifically, this paper put forward and tested the following hypotheses.

H_1: The closer an individual's home is to the nearest public transportation stop, the greater the likelihood they turn out to vote.

H_2: The positive impact of accessible public transportation on voter turnout will be most pronounced in low-income communities.

The results of this study suggest that greater proximity to public transportation produces a modest, but reliable and statistically significant positive impact on voter turnout in low-salience elections and a smaller and narrower, but still noticeable and statistically significant positive impact on voter turnout in high-salience elections, which lends credence to my first hypothesis and allows me to determine there is mixed support for H_1 in this study.

These findings comport with the prior literature reviewed for this study. Namely, the group of studies by Schur et al, Fauvelle-Aymar & François, and Hershey that leant credibility to the idea that voter turnout will be lower if voting costs are higher and also lends slightly more nuanced support to the idea that voter turnout will be higher if costs of voting are lower and polling places are more accessible. This study strengthens this finding and indicates that accessible public transport could play a small, but significant, factor in lowering costs of voting and increasing turnout.

My expectation in H_2 was that the impact of proximity to transit on voter turnout would be most pronounced for voters living in the poorest areas, and that was not at all supported by the data collected and analyzed for this study. In fact, the exact opposite seemed to be true—the higher a ZIP Code Tabulation Area's median household income, the greater the impact of proximity to transit. Therefore, I must reject hypothesis 2.

Given that the wealthiest ZCTAs in the model are most likely to contain voters who can afford their own private transit, these results seem counterintuitive. One would expect that the greater the median household income of an area, the less reliant that area would be on public transport, but that is not the case. The data underwriting this research does not itself provide any clear explanations of what could be causing this phenomenon. There are, however, two possible explanations that I was able to uncover. First, there is substantial evidence that public transportation increases the value of nearby real estate (National Association of Relators, 2014). This effect could put some property near transit out of the price range for many low-income voters, which would limit the magnitude of this effect. More granular demographic information for the voters in this, or a similar future, study could easily support or refute this claim. The second possibility is that there are low-income voters in ZCTAs that have a high aggregate median household income who benefit greatly from public transportation, but I just cannot identify them because of a lack of granularity in the data.

Future research into this topic would be incredibly valuable as there has not been much empirical work on this topic and the results in this study make clear that there is a statistical phenomenon to examine here. Additionally, this research had a number of limitations that could and should be corrected for future work. First, while I originally set out to run this analysis on every county in Pennsylvania, I had a great deal of difficulty finding public transport stop location data for other counties/public transport authorities. It is possible to collect a more complete dataset, it might just require pulling the stops off Google Maps or a similar platform one at a time, which was not practical for this paper. Second, I think the model and analysis could be improved through a greater number of control variables measured at the individual level, especially some level of income and car ownership. As I was relying on the Full Voter Export for all my individual-level data, I did not have an incredibly robust selection of data to choose from. While I was able to find workarounds, it would be good to avoid the differing levels of granularity for the control variables as much as possible.

Appendix A
Data

Experimental Dataset

Publication of data from the Pennsylvania Full Voter Export is prohibited by 4 Pa. Code sections 183.13 (g) & 183.14 (k), so I will unfortunately be unable to publish the experimental dataset I used for this research. If you'd like to purchase a copy of the FVE for yourself to attempt to replicate this study, you can do so at this link.

10 Bin Regression Table

2016 and 2018 Election Regression Results (10 Bins)

	Dependent variable:	
	2016 Voter Turnout	2018 Voter Turnout
	(1)	(2)
distance bin 1 (0.000-0.033 mi)	-0.010	0.025*
	(0.013)	(0.014)
distance bin 2 (0.033-0.079 mi)	-0.031**	0.001
	(0.012)	(0.013)
distance bin 3 (0.079-0.137 mi)	-0.014	0.021*
	(0.012)	(0.013)
distance bin 4 (0.137-0.234 mi)	-0.001	0.029**
	(0.011)	(0.012)
distance bin 5 (0.234-0.418 mi)	0.0002	0.024**
	(0.011)	(0.012)
distance bin 6 (0.418-0.816 mi)	0.019*	0.063***
	(0.011)	(0.012)
distance bin 7 (0.816-1.685 mi)	0.011	0.038***
	(0.011)	(0.012)
distance bin 8 (1.686-3.316 mi)	0.014	0.038***
	(0.011)	(0.012)
distance bin 9 (3.317-6.070 mi)	-0.008	0.010
	(0.011)	(0.012)
age	0.002***	0.003***
	(0.0002)	(0.0002)
gender (female)	0.078***	0.059***
	(0.008)	(0.009)
gender (male)	0.060***	0.069***
	(0.008)	(0.009)
gender (unspecified)	0.083***	0.060***
	(0.011)	(0.012)
political party (democrat)	0.058***	0.097***
	(0.008)	(0.009)
political party (republican)	0.063***	0.061***
	(0.008)	(0.009)
political party (other)	0.022	0.013
	(0.017)	(0.018)
2016 proportion of residents in poverty	-0.082**	
	(0.041)	
2016 proportion of white residents	0.014	
	(0.021)	
2018 proportion of residents in poverty		-0.281***
		(0.044)
2018 proportion of white residents		0.019
		(0.022)
number of registered passenger vehicles per capita	-0.127***	-0.180***
	(0.044)	(0.048)
number of polling places	-0.00000	-0.00001
	(0.00001)	(0.00001)
2012 election turnout	0.416***	0.241***
	(0.007)	(0.007)
2014 election turnout	0.183***	0.299***
	(0.007)	(0.007)
Constant	0.212***	0.164***
	(0.043)	(0.045)
Observations	23,941	23,941
R^2	0.388	0.310
Adjusted R^2	0.387	0.310
Residual Std. Error (df = 23918)	0.382	0.415
F Statistic (df = 22; 23918)	688.134***	488.765***
Note:		*p<0.1; **p<0.05; ***p<0.01

BIBLIOGRAPHY

Abreha, D. A. (2007). *Analysing Public Transport Performance Using Efficiency Measures and Spatial Analysis: The case of Addis Ababa , Ethiopia* [International Institute for Geo-information Science and Earth Observation]. https://webapps.itc.utwente.nl/librarywww/papers_2007/msc/upla/abreha.pdf

Aldrich, J. H. (1993). Rational Choice and Turnout. *American Journal of Political Science*, *37*(1), 246–278. JSTOR. https://doi.org/10.2307/2111531

Bok, J., & Kwon, Y. (2016). Comparable Measures of Accessibility to Public Transport Using the General Transit Feed Specification. *Sustainability*, *8*(3), 224. https://doi.org/10.3390/su8030224

Brady, H. E., & McNulty, J. E. (2011). Turning Out to Vote: The Costs of Finding and Getting to the Polling Place. *American Political Science Review*, *105*(1), 115–134. Cambridge Core. https://doi.org/10.1017/S0003055410000596

Brennan Center for Justice. (2021, October 4). *Voting Laws Roundup: October 2021 |*. Brennan Center for Justice at NYU Law. https://www.brennancenter.org/our-work/research-reports/voting-laws-roundup-october-2021

Center for Public Integrity. (2021). *About Us*. Center for Public Integrity. http://publicintegrity.org/about/

Clerk of the House of Representatives. (n.d.). *Election Statistics: 1920 to Present | US House of Representatives: History, Art & Archives*. Clerk of the United States House of Representatives. Retrieved March 21, 2022, from https://history.house.gov/Institution/Election-Statistics/

DeSilver, D. (2020, November 3). *In past elections, U.S. trailed most developed countries in voter turnout | Pew Research Center*. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/11/03/in-past-elections-u-s-trailed-most-developed-countries-in-voter-turnout/

Downs, A. (1957a). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, *65*(2), 135–150. JSTOR.

Downs, Anthony. (1957b). *An economic theory of democracy*. Harper; /z-wcorg/.

Edlin, A., Gelman, A., & Kaplan, N. (2007). Voting as a Rational Choice: Why and How People Vote To Improve the Well-Being of Others. *Rationality and Society*, *19*(3), 293–314. https://doi.org/10.1177/1043463107077384

Fauvelle-Aymar, C., & François, A. (2015). Mobilization, cost of voting and turnout: A natural randomized experiment with double elections. *Public Choice*, *162*(1), 183–199. https://doi.org/10.1007/s11127-014-0212-0

Feddersen, T. J. (2004). Rational Choice Theory and the Paradox of Not Voting. *Journal of Economic Perspectives*, *18*(1), 99–112. https://doi.org/10.1257/089533004773563458

Goodman, N., & Stokes, L. C. (2020). Reducing the Cost of Voting: An Evaluation of Internet Voting's Effect on Turnout. *British Journal of Political Science*, *50*(3), 1155–1167. Cambridge Core. https://doi.org/10.1017/S0007123417000849

Google Developers. (2020, November 13). *Testing GTFS Feeds | Static Transit*. Google Developers. https://developers.google.com/transit/gtfs/guides/tools

Grofman, B. (2016). Perspectives on the Comparative Study of Electoral Systems. *Annual Review of Political Science*, *19*(1), 523–540. https://doi.org/10.1146/annurev-polisci-020614-092344

Hershey, M. R. (2009). What We Know about Voter-ID Laws, Registration, and Turnout. *PS: Political Science and Politics*, *42*(1), 87–91. JSTOR.

Kenyon, S., Lyons, G., & Rafferty, J. (2002). Transport and social exclusion: Investigating the possibility of promoting inclusion through virtual mobility. *Journal of Transport Geography*, *10*(3), 207–219. https://doi.org/10.1016/S0966-6923(02)00012-1

Lijphart, A. (1997). Unequal Participation: Democracy's Unresolved Dilemma. *The American Political Science Review*, *91*(1), 1–14. JSTOR. https://doi.org/10.2307/2952255

Lijphart, A. (1998). *The Problem of Low and Unequal Voter Turnout—And What We Can Do About It. IHS Political Science Series No. 54, February 1998* (Working Paper No. 54; IHS Political Science Series). The Institute for Humane Studies at George Mason University. http://www.ihs.ac.at/vienna/IHS-

Departments-2/Political-Science-1/Publications-18/Political-Science-Series-2/Publications-19/publication-page:8.htm

López, G., & Flores, A. (2017, June 1). Why registered voters say they didn't vote in 2016 [Fact Tank]. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2017/06/01/dislike-of-candidates-or-campaign-issues-was-most-common-reason-for-not-voting-in-2016/

MobilityData IO. (n.d.). *OpenMobilityData*. OMD - Public Transit Feeds from around the World. Retrieved November 26, 2020, from http://transitfeeds.com/

MobilityData IO. (2020). *About*. MobilityData. https://mobilitydata.org/about/

MURRAY, A., Davis, R., Stimson, R., & Ferreira, L. (1998). PUBLIC TRANSPORTATION ACCESS. *World Transit Research*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.426.2996&rep=rep1&type=pdf

National Association of Relators. (2014, June 16). *Public Transportation Boosts Property Values*. www.nar.realtor. https://www.nar.realtor/articles/public-transportation-boosts-property-values

Rebala, P. (2021, December). *Polling Place Location Data*. GitHub. https://github.com/PublicI/us-polling-places

Saif, M. A., Zefreh, M. M., & Torok, A. (2019). Public Transport Accessibility: A Literature Review. *Periodica Polytechnica Transportation Engineering*, *47*(1), 36–43. https://doi.org/10.3311/PPtr.12072

Schur, L., Ameri, M., & Adya, M. (2017). Disability, Voter Turnout, and Polling Place Accessibility. *Social Science Quarterly*, *98*(5), 1374–1390. https://doi.org/10.1111/ssqu.12373

Sigelman, L., & Berry, W. D. (1982). Cost and the Calculus of Voting. *Political Behavior*, *4*(4), 419–428. JSTOR.

US Census Bureau. (2021, October). *ZIP Code Tabulation Areas (ZCTAs)*. United States Census Bureau. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html

US Polling Places (2012-2018). (2020). Center for Public Integrity. https://github.com/PublicI/us-polling-places (Original work published 2020)

Verbavatz, V., & Barthelemy, M. (2020). Access to mass rapid transit in OECD urban areas. *Scientific Data*, *7*(1), 301. https://doi.org/10.1038/s41597-020-00639-3

Voting Information Project. (2021). *About Us*. The Voting Information Project. https://www.votinginfoproject.org/about

ACADEMIC VITA

Caleb Klemick

cjk5872@psu.edu

Skills: strong oral and written communication skills; empirical and informal research skills; interpersonal and problem-solving skills; Muck Rack Media Relations certificate; LinkedIn Learning Crisis Communications Certificate; demonstrated interest in Democratic politics

Penn State Division of Development and Alumni Relations / Development/Alumni Relations Intern
MAY 2021 - JULY 2021, UNIVERSITY PARK, PA

Responsible for maintaining strong, mutually beneficial relationships with donors to the College of Arts and Architecture at Penn State University Park. Conducted more than 40 donor interviews and wrote in-depth biographies for each donor. Was also responsible for writing donor proposals, providing administrative support for front-line fundraisers, and digitizing archived records.

Penn State HUB-Robeson Center / Student Production Supervisor
DECEMBER 2018 - PRESENT, UNIVERSITY PARK, PA

Responsible for the safe and efficient daily operations of a ~900 person and ~150-person theater--Schwab Auditorium and the HUB Flex Theater, respectively. As a member of a team that operates theater sound and lighting equipment, manages all guests and talent, and ensures the satisfaction and safety of everyone who enters our event spaces.

The Penn State IFC/Panhellenic Dance Marathon / Entertainment Director
APRIL 2021 - PRESENT, UNIVERSITY PARK, PA

Responsible for leading 26 Captains and 50 Committee Members to provide the highest quality music and live entertainment at THON events throughout the year, including THON Weekend, a 46-hour live event. Personally responsible for contracting more than 50 live performers and managing a budget of nearly $200,000.

The Pennsylvania State University / Bachelor of Arts in Public Relations and Bachelor of Science in Political Science
AUGUST 2018 - MAY 2022, UNIVERSITY PARK, PA

Admitted to the Schreyer Honors College as a rising junior. Dean’s List honoree since Spring 2019. Provost Award academic merit scholarship recipient. Anticipated graduation in Spring 2022.