# The Influence of Individual Characterisitcs on Public Transportation Planning*

Iris Zhong

**Abstract**

xx

## 1   Introduction

Air pollution has become a global concern. Approximately 91% of the population around the world live in places where air quality levels exceed WHO limits in 2016 (Ambient (Outdoor) Air Pollution, n.d.). Severe air pollution increases the risk for cardiovascular and respiratory diseases, cancer and adverse birth outcomes, and is generally associated with higher death rates (WHO | Air Pollution, n.d.). Transportation is a major source of outdoor air pollution. For example, road transport contributes to up to thirty percent of particulate emissions in European cities. Developing public transit is an effective way to reduce air pollution. When people switch their travel mode from private vehicles to public transport, the concentration of air pollutants such as carbon monoxide decreases significantly (Chen & Whalley, 2012; Zheng et al., 2019).

Constructing a public transit often requires majority approval from the local voters. Voters either choose yes or no in the ballot, and if more than half of them vote for yes, the initiative is approved. In public transit referendums, supporters benefit from public transport by saving travel time or expenditure, or believe it can relieve traffic congestion and reduce pollution (Manville & Cummins, 2015). Meanwhile, opponents concern with the increase of tax. While some argue that ballot result does not provide enough information for empirical analyses because it is a binary response void of the reasoning behind the voter's choice (Manville & Levine, 2018), various research studying voting outcomes has come to insightful conclusions (Burkhardt & Chan, 2017; Kinsey et al., 2010, etc.). When enough data is collected and shows patterns of individuals with

---

*xx

similar characteristics making the same decision, researchers can infer how these factors shape voters' minds. Such an investigation helps policymakers recognize how different populations understand public transit and improve the proposition accordingly to benefit a wider variety of people.

This paper examines a public transit referendum in Gwinnett, Georgia, on March 23rd, 2019. The referendum is about whether to join Metropolitan Atlanta Rapid Transit Authority (MARTA) and expand public transit service. My research asks how sociodemographic characteristics such as race and partisanship influence voters' decisions. After collecting data from the government website and census bureau, I convert a few variables to census tract level, so that they are measured at the same unit. I create two model specifications, one using the original dataset, the other using a modified dataset with skewed variables transformed. After age, commute time, percentage of people using public transit, and voter turnout are held as controls, partisanship significantly predicts the percentage of supporters for the referendum – the tracts with a higher proportion of Trump supporters tend to vote against the public transit plan. Race is also a significant predictor, such that the tracts with a higher proportion of the white population favor the referendum. Finally, counterintuitively, the tracts newly included in the public transit system by the proposition tend to reject the plan, even though they gain more benefit from it.

A possible explanation for my last finding can be related to racial discrimination, as addressed in popular press articles (Binkovitz, 2017). Gwinnett had already rejected MARTA twice before, mainly because the white residents in the suburban areas refused to take the same transport with urban Black people. It is likely that the same reason applies to the 2019 referendum as well.

The remaining paper is structured as follows. Section 2 reviews previous literature. In Section 3, I include a background of Gwinnett's current public transit system, the content of the proposition, and the referendum. In Section 4, I describe the data sources, methods, and models. Section 5 presents the results and discussions. Finally, Section 6 concludes this paper.

## 2 Literature Review (Unfinished)

Referendum data, characterized as dichotomously coded from the survey, has been widely used in economic research. The estimation outcome with the referendum approach is compared to that of the individual stochastic valuation approach in Bollino's paper, which estimates the Italians' WTP of using renewable energy to produce electricity (2010). The estimated WTP from the referendum approach depends on how the original Likert-scale data is coded into a binary variable, and most of the outcomes are similar to the results from the individual stochastic valuation approach. Even though he ignores the sociodemographic

factors in the referendum approach, the study justifies that it can provide similar conclusions with other models.

Similarly, Burkhardt and Chan (2017) also utilize referendum data to estimate WTP for public goods, but at an aggregate zip code level with a log-likelihood function. They look into thirteen referenda in California and find out the price of the public good is negatively correlated with support for the proposition. For the referenda that failed, they calculate the amount of cost that needs to be cut off for the proposition to pass by finding the difference between WTP and price. Even though my project does not involve assessing voters' WTP, this procedure should be considered in future studies so that an estimation of the necessary reduction in cost to reach enough support can be computed.

Allen et al. (2016) study the reasoning of the failure of a referendum on a congestion charging scheme in Edinburgh. Instead of using direct voting data, they conduct a survey after the referendum, which allows them to ask more specific questions. Researchers can gain detailed data by surveying, because the unit of measurement is each individual; however, a possible disadvantage of surveying is that respondents who turn in the questionnaire tend to have stronger attitudes towards the proposal, generating sampling bias. They conclude that people who use cars as the primary transportation mean, demonstrate a misconception of the pricing plan, or question the effectiveness of the scheme at reducing congestion are more likely to oppose it. Their findings can give insights to the similar failure in the Gwinnett referendum. Voters against the proposal could be those who rarely use public transportation and those who are not convinced by the effectiveness of expanding public transit in alleviating the traffic.

Another crucial factor is the accessibility of the proposed transit system. Kinsey et al. (2010) examine the relationship between the distance to the scheduled railway station and voter turnout by studying the Seattle monorail referendum. They introduce the concept of diffused and concentrated benefit/cost. People who live far from the monorail enjoy the diffused benefit of less traffic congestion, and bear the diffused cost of increased tax. People living close to the rail experience the same diffused benefit and cost, but they also gain the concentrated benefit of easily accessing the public good. Finally, those who live very close to the railway have the same benefits and costs, but they also face the concentrated cost such as inconvenience during construction. Since "people are more strongly motivated to avoid losses than to approach gains," they expect a higher turnout rate in farther places with votes for "no," which is verified from their analyses. Besides distance, they also find out precincts with a higher percentage of people of lower socioeconomic status or young people have a lower turnout rate. Interestingly, there is a significant interaction between partisanship and distance, which would be also tested in my study. In essence, the effect of distance on turnout is weakened by partisanship, and vanishes beyond a threshold of distance. Even though my dependent variable

is voters' responses rather than turnout, it can be inferred from Kinsey et al.'s findings that people farther away from the transit system would vote against the referendum more. However, the relationship might be non-linear and requires some form of transformation. Regarding the methods, they utilize the spatial lag model to correct for autocorrelation, which is proper to use in my project as well since both studies use precinct-level data.

# 3   Background

## 3.1   Gwinnett County and Transportation

Gwinnett County sits in the north-central part of Georgia in the United States (Fast Facts | Gwinnett County, n.d.). The estimated population was 936,250 in 2019 (U.S. Census Bureau QuickFacts, n.d.), with approximately 55% identified as white, 29.3% identified as Black or African American, 12.4% as Asian, and 0.9% identified as other races. For ethnicity, 21.5% identified themselves as Hispanic or Latino.

The primary transportation means for Gwinnett County residents is driving. According to Census Bureau statistics in 2018 (Census - Means of Transportation to Work by Vehicles Available, n.d.), of the 456,465 households who are workers over 16 years old, over 90% drive alone or carpool to work. On the other hand, only 0.8% take public transit to work. Since approximately 4.9% of Americans use public transportation to commute (Principal Means of Transportation to Work | Bureau of Transportation Statistics, n.d.), the usage of public transit in Gwinnett is lower than other parts of the country.

The major roads in Gwinnett include two interstate highways, three U.S. routes, and eighteen state routes ("Gwinnett County, Georgia," 2020). In the public transit sector, Gwinnett County Transit currently operates seven Local bus routes and five Express routes (Routes and Schedules | Gwinnett County, n.d.) (see Figure ???). The heavy rail transit (HRT) operated by the Metropolitan Atlanta Rapid Transit Authority (MARTA) does not go through Gwinnett. Yet, the nearest rail station is very close to the border, and several bus lines start from there (Connect Gwinnett Transit Plan Recommendations Report, 2018).

## 3.2   Connect Gwinnett: Transit Plan

The population in Gwinnett has nearly doubled from 2000 to 2016, but public transit service has not expanded (Connect Gwinnett Transit Plan Recommendations Report, 2018). As a result, Gwinnett County Transit drafted a public transit expansion plan named "Connect Gwinnett: Transit Plan" in 2018. All of

the following information comes from its recommendations report (2018).

The plan consists of four stages: short-range (2019 - 2024), mid-range (2025 – 2029), long-range phase 1 (2030 – 2049), and long-range phase 2 (2050 onward). The most exciting improvement is the first HRT service proposed to build during long-range phase 1. It is an extension of an existing line from MARTA.

Nevertheless, I decided to study the short-range plan in this paper in order to minimize the effect of discounting on future costs and benefits. Short-range recommendations target on the most immediate demands. During this period, Gwinnet transit service will increase to seventeen local bus routes, nine express commuter bus routes, two direct connect lines, and two flex routes. Direct connect is a new service introduced in the short-range that takes passengers to and from MARTA stations with limited stops to shorten travel time. Flex is an on-demand service that carries passengers upon request.

The entire Connect Gwinnett: Transit Plan is estimated to cost 527.7 million dollars. The plan is going to be funded mainly by sales tax revenue, assuming a 1.5% increase of tax in the first five years and a 1% increase afterward.

## 3.3 Referendum

A special referendum was held on March 19th, 2019 to decide whether the county will join MARTA and expand the public transit system following Connect Gwinnett: Transit Plan. A vote for yes "would be a vote in support of ratifying Gwinnett's pending transit service contract with MARTA — and one in favor of enacting a new 1 percent sales tax to pay for transit projects and operations in the county." (Estep, 2019)

The referendum eventually failed. Among the 92,243 votes, 42,156 voted yes, and 50,087 voted no (Results - Gwinnett - Election Night Reporting, 2019). The turnout rate was 16.98%.

Some argue that the failure could be due to the delay of the referendum from November 2018 to March 2019 (Yeomans, 2018). The change from voting in a general election to a special referendum would reduce voter turnout and affect the result.

# 4 Data & Methods

## 4.1 Conceptual model

According to previous research, sociodemographic elements can influence people's voting decisions in the referendum. For example, the effect of income is mixed: on the one hand, people with higher income will

pay a smaller portion of their earnings for the implementation of the plan; on the other hand, they will pay a larger amount of tax. Bollino (2008) finds a positive correlation between income and people's willingness to pay for renewable resources. Burkhardt and Chan (2017) separate the influence of income from tax, and discover their opposite effects on voting. Therefore, it is worth considering the relationship between income and percentage of supporters in this referendum. Voters' partisanship attachment is found to be a significant factor as well in Burkhardt and Chan's (2017) paper. Areas with higher proportions of Republicans are less supportive of fiscally costly propositions. In my project, it can be hypothesized that tracts that have a higher proportion of Trump supporters tend to have a lower percentage of agreement to the proposal.

In addition, some factors related to transportation can intuitively shape people's attitudes towards public transit. For example, the areas in which people do not use public transit at all might have a higher percentage of refusal of the proposal, because they have alternative transportation means. People who have to travel a long time to work are more likely to support the extension plan if it helps save time. These two factors serve as controls in the models.

Finally, people favor the proposition if it benefits them. Specifically, tracts that are not covered by public transport at present but will be covered in the expansion plan are predicted to support the proposal more.

## 4.2 Data

The ballot results of the Gwinnett County referendum on Mar 19th, 2019 is obtained from a website powered by Scytl, a trusted source of election outcomes. The cross-sectional dataset contains the voting information of all 157 precincts in the county. The dependent variable – the proportion of supporters of the referendum, and the voter turnout rate are calculated from this data source.

The result of the 2016 Presidential Election is chosen to reflect partisanship. A cross-sectional precinct-level election data is obtained from the MIT Election Data and Science Lab website. The number of votes for Trump at each precinct in Gwinnett county comes from this dataset.

Next, cross-sectional sociodemographic characteristics at the census tract level are found in Census Bureau via an R package called `tidycensus`. The data comes from the American Community Survey 5-year estimate published in 2018. The median age and median income are collected here. In addition, the proportion of white, the percentage of people who go to work by public transportation, and the percentage of people who travel more than an hour to work are calculated by dividing the relevant variables by the total population or the survey sample size.

**Table 1:** Variable definitions

| Variable name | Description |
|---|---|
| GEOID | The geographic identifier of the census tract |
| medage | The median age of the population in the tract |
| medincome | The median income of the population in the tract |
| white_pct | The percentage of white population in the tract |
| public_pct | The percentage of people who go to work by public transportation (excluding taxi or cab) |
| time_pct | The percentage of people who travel more than an hour to work |
| trump_pct | The estimated percentage of votes for Donald Trump in that tract |
| voter_turnout | The estimated percentage of voters who voted in this referendum in the tract |
| yes_pct | The estimated percentage of voters who voted yes in this referendum in the tract |
| plan_yes | Whether the tract is covered by the public transportation now and in the short-range, defined by whether any transportation is available within 500 meters. |
| | 1 stands for the tract doesn't have transit both now and in the short-range plan. |
| | 2 stands for the tract has transit now but not in the short-range plan. |
| | 3 stands for the tract that doesn't have transit now and will have in the future. |
| | 4 stands for the tract that has public transit both now and in the future |

The information on whether the tract enjoys the proximity of public transportation now and future can be acquired from spatial maps and analyses. First, a precinct-level shapefile of Gwinnett County made in 2018 is obtained from the Georgia General Assembly. Gwinnett County maps with current and proposed future public transit systems are available on the Gwinnett County government website. I select the short-range expansion plan (Y2020 – 2025) because the cost and benefit of the expansion in the far future are discounted more. After modifications in QGIS software, the maps are then transformed into spatial data readable in R. Five-hundred-meter buffer zones are created around bus and railway routes. The categorical variable *current_plan* has four levels: 1 represents no transportation near this tract at present and in the future; 2 represents the tracts that are accessible to public transit currently but not in the future; 3 stands for the tracts that do not have transit at present but will do in the expansion plan; 4 represents the tracts that have and will have public transit for now and for the future.

As noted above, both referendum and 2016 election data are collected at the precinct level. However, the other datasets are performed at the census tract level. Therefore, referendum and election data are redistributed by the areas shared by the precinct and the tract. See the data appendix for detailed steps of transformation.

The final dataset joins the datasets above by GEOID. It is cross-sectional, measured with the unit of the census tract. Since there are 113 census tracts in Gwinnett, the dataset has 113 observations, with no missing data. A description of the variables can be found in *Table 1*.

**Table 2:** Summary statistics

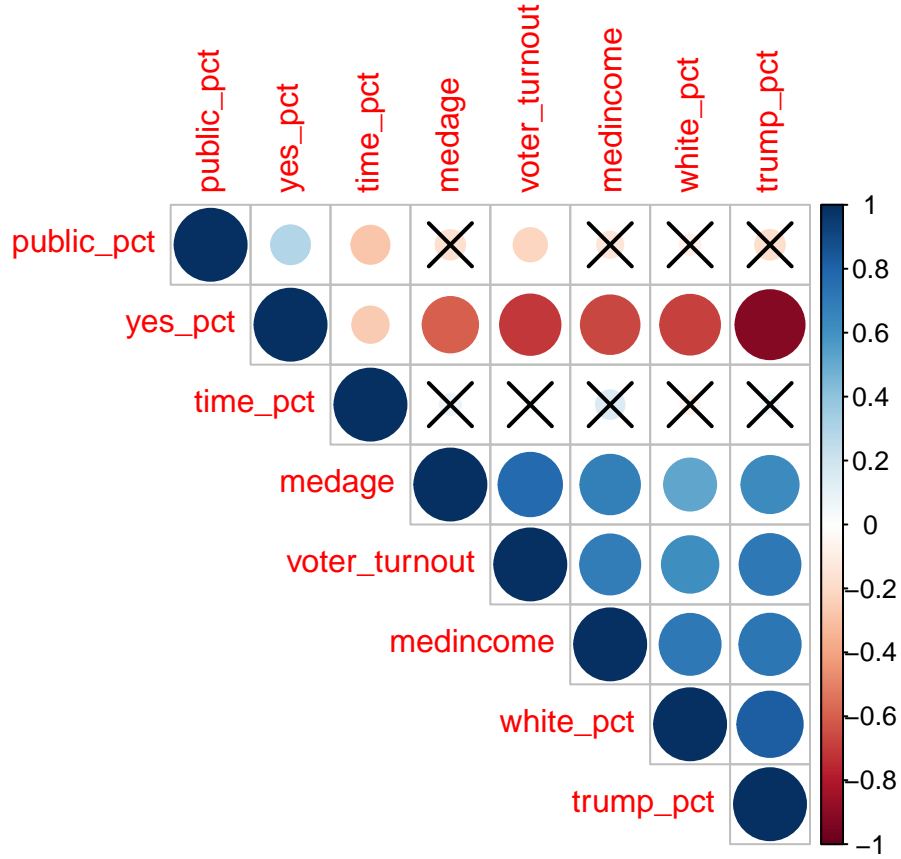| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| medage | 113 | 35.56 | 4.58 | 26 | 32.8 | 38.8 | 52 |
| medincome | 113 | 69,439.24 | 24,358.44 | 33,020 | 51,429 | 82,845 | 156,136 |
| white_pct | 113 | 0.48 | 0.15 | 0.17 | 0.38 | 0.61 | 0.89 |
| public_pct | 113 | 0.01 | 0.01 | 0 | 0.002 | 0.02 | 0.06 |
| time_pct | 113 | 0.16 | 0.05 | 0.04 | 0.12 | 0.20 | 0.31 |
| trump_pct | 113 | 0.40 | 0.15 | 0.11 | 0.27 | 0.52 | 0.69 |
| voter_turnout | 113 | 0.16 | 0.06 | 0.05 | 0.13 | 0.18 | 0.37 |
| yes_pct | 113 | 0.53 | 0.14 | 0.27 | 0.42 | 0.61 | 0.84 |



*Table 2* lists the summary statistics. Median income is skewed to the right, with a few tracts demonstrating high income levels. Such a pattern of inequality is universally observed.

The usage of public transit is low. On average, only one percent of the population relies on public transportation to go to work. As proved by later analyses, the distribution is highly skewed, and will be transformed in some models. **public_pct max needs to be fixed**

The variable *current_plan* is not in *Table 2* because it is a categorical variable. In sum, 31 tracts do not

have access to public transport within 500 meters, both now and in the short-range. There is one tract that is categorized as 2, indicating that it has public transit at present, but not in the short-range proposition. It could be due to the plan of reducing circuitous routing. Twenty-two tracts do not enjoy the proximity of public transport now, but will in the short-range. Lastly, 59 tracts have public transit available both at present and in the short-range plan. Since a level of this variable contains only one value ("2"), problems with degrees of freedom will potentially arise.

A correlation matrix is created from `corrplot` package to investigate the correlation between each pair of factors (see *Figure ???*). Positive correlations are in blue colors, and negative correlations are in red colors. The magnitude is reflected by the color intensity and the size of the dot. Non-significant ($p > 0.05$) correlations are marked with a cross. *Current_plan* is omitted in the matrix because it is categorical. The dependent variable *yes_pct* is significantly correlated with every independent variable. *Medage*, *voter_turnout*, *medincome*, *white_pct*, and *trump_pct* are all strongly positively correlated with each other, and all of them are negatively associated with *yes_pct*. Given the number of strong, significant correlations among the factors, it is essential to check collinearity in the regression model.

The major limitation of the data is the unit conversion from precinct to census tract. Such a method assumes that residents in one precinct have the same characteristics, and population density is identical. Clearly, the assumptions cannot be satisfied in a real dataset.

## 4.3 Model specification

Two models are tested in the analysis.

Model 1 adopts linear regression model with all the original variables:

$yes\_pct_t = \beta_0 + \beta_1 medincome_t + \beta_2 white\_pct_t + \beta_3 trump\_pct_t + \beta_4 current\_plan_t + \beta_5 medage_t + \beta_6 public\_pct_t + \beta_7 time\_pct_t + \beta_8 voter\_turnout_t + \epsilon_t$

where $t$ indexes census tract. Since all variables are measured in the same scope, no fixed effects are tested. The dependent variable is *yes_pct*, the proportion of supporters of the referendum in a tract. Independent variables include *medincome*, *white_pct*, *trump_pct*, and *current_plan*. For more information about these variables, refer to *Table ???*. I hypothesize that the effect of *medincome* is ambiguous. As explained in the Conceptual Model section, people with higher income pay a larger amount of tax, but the tax takes up a smaller portion of their earnings compared to those with lower income. *White_pct* is hypothesized to have a negative effect on *yes_pct*, based on the negative correlation between the two variables. A higher percentage of Trump supporters is expected to predict a lower percentage of votes for "yes" in the referendum, as

evidenced by common sense and previous literature (???). Finally, since tracts that do not have public transit now and will have it in the short-range benefit the most from the proposition, I predict that the tracts with this feature will have a higher level of *yes_pct* than the others.

*Medage*, *public_pct*, *time_pct*, and *voter_turnout* add to the model as controls. For example, tracts with higher voter turnout are hypothesized to have a lower *yes_pct*, according to the rule of loss aversion – people who believe the referendum incur losses to them are inclined to participate in the referendum actively and vote against it, but people who like the proposition are less motivated to vote in nature.

*Medincome*, *public_pct* and *voter_turnout* are found to be positively skewed. Thus, Model 2 uses multiple linear regression after log or square root transformation on these three variables. Among them, since many of the values in *public_pct* are 0, a constant has to add to the variable before taking log transformation.

$$yes\_pct_t = \beta_0 + \beta_1 log(medincome_t) + \beta_2 white\_pct_t + \beta_3 trump\_pct_t + \beta_4 current\_plan_t + \beta_5 medage_t + \beta_6 log(public\_pct_t + 0.01) + \beta_7 time\_pct_t + \beta_8 \sqrt{voter\_turnout_t} + \epsilon_t$$

# 5   Results & Discussion

## 5.1   Regression results

*Table ???* provides the results of Model 1. The equation is statistically significant ($F(10, 102) = 113.22$, $p < 0.05$). Adjusted $R^2$ is 0.909, indicating that the independent variables in this specification explain a large portion of the variance in the dependent variable. Consistent with the hypothesis, *medincome* is not a significant predictor of *yes_pct*. *Trump_pct* significantly predicts *yes_pct* as expected, and results in a large coefficient: a one percent increase in the proportion of Trump supporters decreases the percentage of voters in favor of the referendum by 0.886%, holding other variables constant.

However, unlike the outcome from the correlation matrix (see *Figure ???*), *white_pct* significantly predicts *yes_pct* in a positive direction. A higher percentage of the white population in the tract is associated with higher support of the proposition when holding other explanatory variables constant.

Another unanticipated result is *current_plan*. In contrast to baseline tracts that do not access public transit service both now and in the short-range (*current_plan 1*), only *current_plan 3* tracts differ from them significantly, in the opposite direction to the hypothesis. That is to say, tracts planned to be newly added to the public transit service have a significantly higher rejection of the proposition than the other tracts. The control variables are all significant factors. *Medage* and *public_pct* positively predict the level of *yes_pct*,

**Table 3:** Primary regression

| | *Dependent variable:* |
|---|:---:|
| | yes_pct |
| medincome | 0.00000 |
| | (0.00000) |
| | |
| white_pct | 0.114** |
| | (0.053) |
| | |
| trump_pct | −0.886*** |
| | (0.057) |
| | |
| current_plan2 | 0.019 |
| | (0.044) |
| | |
| current_plan3 | −0.032*** |
| | (0.012) |
| | |
| current_plan4 | 0.016 |
| | (0.012) |
| | |
| medage | 0.003* |
| | (0.001) |
| | |
| public_pct | 0.817** |
| | (0.338) |
| | |
| time_pct | −0.306*** |
| | (0.090) |
| | |
| voter_turnout | −0.326** |
| | (0.126) |
| | |
| Constant | 0.804*** |
| | (0.047) |
| | |
| Observations | 113 |
| $R^2$ | 0.917 |
| Adjusted $R^2$ | 0.909 |
| Residual Std. Error | 0.042 (df = 102) |
| F Statistic | 113.224*** (df = 10; 102) |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

while *time_pct* and *voter_turnout* negatively predict *yes_pct*.

**Table 4:** Transformed regression

|  | *Dependent variable:* |
| --- | --- |
|  | yes_pct |
| log_medincome | −0.008 |
|  | (0.022) |
| white_pct | 0.117** |
|  | (0.051) |
| trump_pct | −0.857*** |
|  | (0.059) |
| current_plan2 | 0.021 |
|  | (0.043) |
| current_plan3 | −0.033*** |
|  | (0.012) |
| current_plan4 | 0.014 |
|  | (0.011) |
| medage | 0.004** |
|  | (0.002) |
| log_public_pct | 0.022*** |
|  | (0.008) |
| time_pct | −0.281*** |
|  | (0.090) |
| sqrt_voter_turnout | −0.319*** |
|  | (0.103) |
| Constant | 1.032*** |
|  | (0.219) |
| Observations | 113 |
| $R^2$ | 0.920 |
| Adjusted $R^2$ | 0.912 |
| Residual Std. Error | 0.041 (df = 102) |
| F Statistic | 117.561*** (df = 10; 102) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Model 2 uses multiple linear regression as well, except that the skewed variables *medincome*, *public_pct*, and *voter_turnout* are transformed. The overall outcome is identical to Model 1 (see *Table ???*). After log transformation, the effect of median income is still insignificant.

A series of assumptions are examined. First, linearity and homoscedasticity assumptions are met, as illustrated by the residual plots. Next, because several factors correlate with each other (see *Figure ???*), VIFs are calculated to detect multicollinearity. Since all of the independent variables exhibit a VIF below 5 in both models, no collinearity issue is detected. Lastly, there is a potential fault of using linear regression in this dataset. The value of the dependent variable *yes_pct* is restricted to the range of 0 to 1, because it represents a percentage. On the other hand, the predicted outcome from linear regression is unbounded. As a result, I calculate the predicted values from the two models with the actual datasets. The predicted outcome is close to the actual value of *yes_pct*, all between 0 and 1. Then, I continue testing by finding the minimum and maximum values of each independent variable in the dataset (to see the actual minimum and maximum values, see *Table ???*), and plug them into the models accordingly to gain the extreme predicted *yes_pct*. The extreme values for Model 1 are 0.09 and 0.96, and 0.11 and 0.95 in Model 2, all within the interval [0,1]. Therefore, the range of the dependent variable in these two regression models meets the assumption.

## 5.2    Discussion

Overall, most of the explanatory and control variables predict the proportion of supporters of the expansion proposition as hypothesized.

In both models, the median income does not significantly predict *yes_pct*. The finding is consistent with the hypothesis that the effect of income level is complicated by taxes. If more information on the financing plan of the expansion and the details of taxation were available, it would be possible to separate the influence of taxation from income.

Partisanship attachment is a significant factor, even after controlling for other variables such as income and race. The tracts that have a higher proportion of votes for Trump during the 2016 Election tend to reject the expansion of public transport more. Republicans are usually reluctant to agree on public spending due to the increase in tax incurred.

As for the control variables, the median age positively associates with *yes_pct*, but the coefficient is relatively small. The percentage of people taking public transit to work predicts *yes_pct* positively as well. People who do not regularly utilize public transit benefit less from the expansion because they are likely to have other transportation available, and switching to public transportation generates costs such as registering, and spending time to get familiar with the locations and operation. *Time_pct* negatively predicts *yes_pct*, indicating that the areas in which more people spend over an hour to work tend to disapprove with the proposal. Voter turnout rate is negatively related to the dependent variable, conforming with previous

research conclusion that those who disagree with the proposal are more likely to express their opinions.

### 5.2.1 Complications with the *current_plan* variable
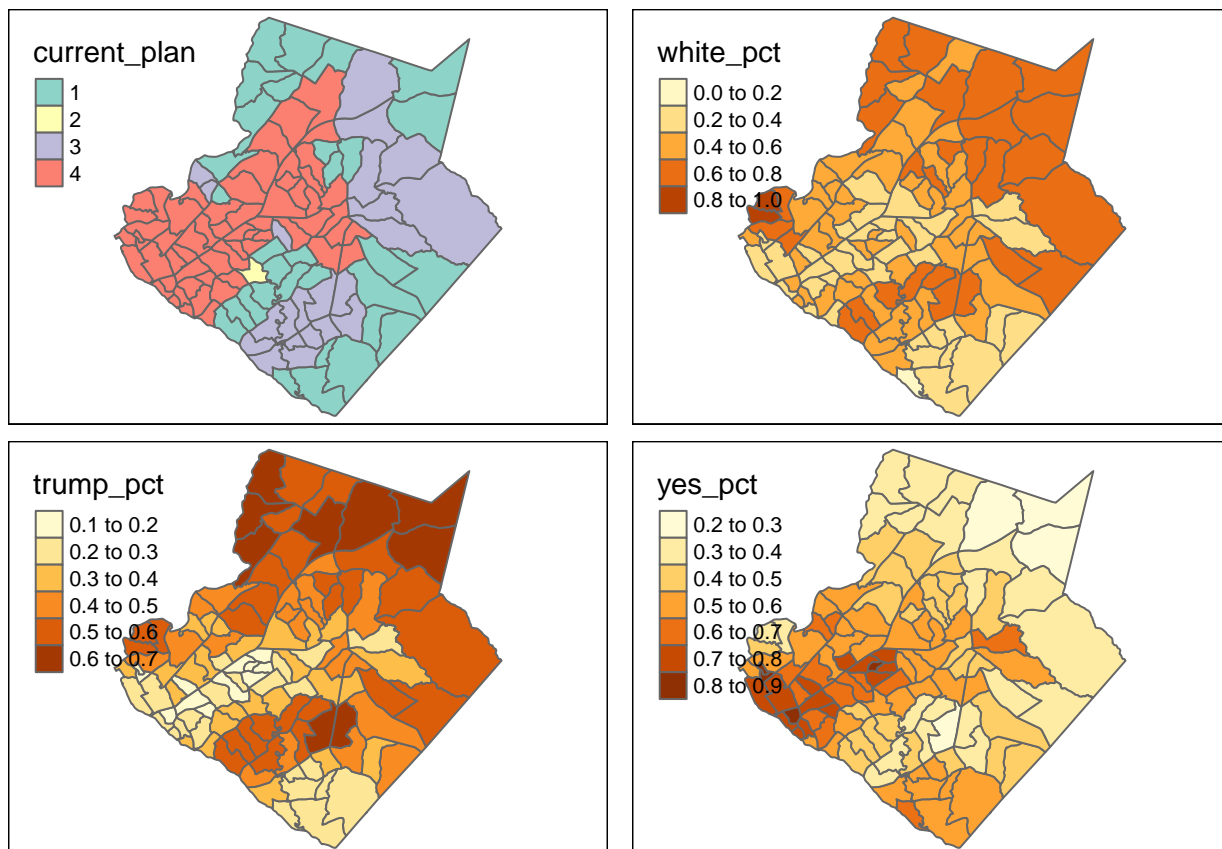


One of the most unforeseen findings is that the tracts that are newly included in the public transit system in the short-range plan have a significantly higher rate of refusal to the proposition. Intuitively, they should support the plan the most, because they are the direct beneficiaries.

A possible explanation is they are worried about the short-term disturbances due to the construction such as traffic congestion and noises. However, since short-range expansion proposal does not include services that require heavy construction, this speculation is less persuasive.

Another supposition is that residents in these tracts do not like the idea of "Connecting Gwinnett." As can be seen from *Figure ???*, the tracts marked with purple in the first map are mostly high in *trump_pct*, *white_pct*, and low in *yes_pct*. It is possible that some sort of segregation still exists, that people in these areas refuse to get connected with other communities that are less "white." In future research, I recommend adding another factor such as the number of hate crimes against racial or ethnic minorities, as a measure of "the level of racism."

### 5.2.2 Partisanship as a mediating factor (Unfinished)
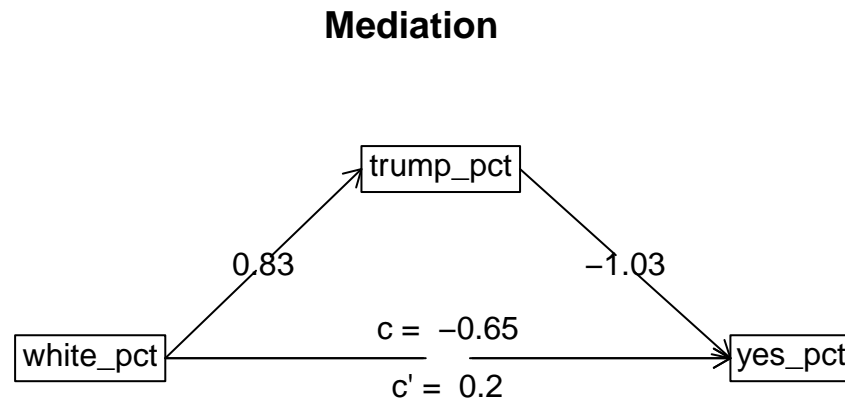
## Mediation



- *yes_pct* ~ *white_pct* positive: transportation history, favor the white - *trump_pct* is a mediating factor that turns the relationship between *white_pct* and *yes_pct* to negative

Call: lm(formula = yes_pct ~ log_medincome + white_pct + cp_3 + trump_pct + medage + log_public_pct + time_pct + sqrt_voter_turnout, data = data_tf_2)

Residuals: Min 1Q Median 3Q Max -0.100052 -0.035836 0.002392 0.033678 0.099693

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.092370 0.212447 5.142 1.28e-06  *log_medincome -0.011993 0.022030 -0.544 0.587326 white_pct 0.122586 0.049971 2.453 0.015823*

cp_31 -0.039802 0.010454 -3.807 0.000238  *trump_pct -0.871285 0.056783 -15.344 < 2e-16*  medage 0.003628 0.001507 2.408 0.017801 *

log_public_pct 0.022960 0.008133 2.823 0.005702  time_pct -0.320043 0.083614 -3.828 0.000221 *

sqrt_voter_turnout -0.321077 0.103100 -3.114 0.002383 — Signif. codes: 0 '*' *0.001* '' *0.01* ' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04115 on 104 degrees of freedom Multiple R-squared: 0.919, Adjusted R-squared:

0.9128 F-statistic: 147.5 on 8 and 104 DF, p-value: < 2.2e-16

Call: lm(formula = yes_pct ~ log_medincome + white_pct + cp_3 * trump_pct + medage + log_public_pct + time_pct + sqrt_voter_turnout, data = data_tf_2)

Residuals: Min 1Q Median 3Q Max -0.099292 -0.034462 0.001796 0.033772 0.098852

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.082925 0.214382 5.051 1.91e-06  *log_medincome -0.011154 0.022200 -0.502 0.616438*

*white_pct 0.122825 0.050170 2.448 0.016048*

**cp_31 -0.054186 0.034688 -1.562 0.121330**

**trump_pct -0.879980 0.060408 -14.567 < 2e-16**  *medage 0.003714 0.001525 2.435 0.016621*

log_public_pct 0.023097 0.008171 2.827 0.005652 ** time_pct -0.314020 0.085076 -3.691 0.000359 *

**sqrt_voter_turnout -0.321009 0.103505 -3.101 0.002485**   cp_31:trump_pct 0.032815 0.075431 0.435 0.664444

— Signif. codes: 0 '*' *0.001* '' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04131 on 103 degrees of freedom Multiple R-squared: 0.9192, Adjusted R-squared: 0.9121 F-statistic: 130.1 on 9 and 103 DF, p-value: < 2.2e-16

Analysis of Variance Table

Model 1: yes_pct ~ log_medincome + white_pct + cp_3 + trump_pct + medage + log_public_pct + time_pct + sqrt_voter_turnout Model 2: yes_pct ~ log_medincome + white_pct + cp_3 * trump_pct + medage + log_public_pct + time_pct + sqrt_voter_turnout Res.Df RSS Df Sum of Sq F Pr(>F) 1 104 0.17610

2 103 0.17577 1 0.00032298 0.1893 0.6644

Call: lm(formula = yes_pct ~ log_medincome + white_pct + trump_pct + medage + log_public_pct + time_pct + sqrt_voter_turnout, data = data_tf_3)

Residuals: Min 1Q Median 3Q Max -0.077770 -0.030289 0.004854 0.025092 0.073703

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.390563 0.830343 1.675 0.11618

log_medincome -0.038839 0.077578 -0.501 0.62440

white_pct 0.118582 0.192143 0.617 0.54704

trump_pct -0.854305 0.229601 -3.721 0.00228 ** medage 0.001382 0.003327 0.416 0.68405

log_public_pct 0.057258 0.028091 2.038 0.06087 . time_pct -0.137833 0.186997 -0.737 0.47324

sqrt_voter_turnout 0.025029 0.452743 0.055 0.95669

— Signif. codes: 0 '*' **0.001** '*' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04297 on 14 degrees of freedom Multiple R-squared: 0.9083, Adjusted R-squared: 0.8625 F-statistic: 19.81 on 7 and 14 DF, p-value: 2.992e-06

Call: lm(formula = yes_pct ~ log_medincome + white_pct + trump_pct + medage + log_public_pct + time_pct + sqrt_voter_turnout, data = data_tf_4)

Residuals: Min 1Q Median 3Q Max -0.098698 -0.032922 -0.003093 0.034279 0.102060

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.002039 0.240801 4.161 7.68e-05 *log_medincome -0.004292 0.025786 -0.166 0.868215*

*white_pct 0.124772 0.054313 2.297 0.024120*

**trump_pct -0.890271 0.062985 -14.135 < 2e-16** *medage 0.004028 0.002053 1.962 0.053086 .*

*log_public_pct 0.019245 0.008889 2.165 0.033254*

time_pct -0.365807 0.101423 -3.607 0.000528 * **sqrt_voter_turnout -0.347868 0.110835 -3.139 0.002351** — Signif. codes: 0 '*' **0.001** '*' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04169 on 83 degrees of freedom Multiple R-squared: 0.9164, Adjusted R-squared: 0.9093 F-statistic: 130 on 7 and 83 DF, p-value: < 2.2e-16

# 6   References

xxx