WHAT MAKES TENANTS "TENANTS"?

Jiwon Park

FINAL REPORT
FALL 2020
SES5215

TABLE OF CONTENTS

I. INTRODUCTION

1. Research Question

"What are the relationships between the tenure and socio-economic factors in the regional level?"

For a long time, increasing homeownership and protecting tenant rights have been considered as two major axes of urban housing policy.

However, as planners, we need to go beyond merely supporting certain types of tenure or mediating some side effects of tenure. We have to go deeper into exploring what socio-economic factors make people end up living in certain types of tenure to come up with more systemic planning solutions. That being said, the geographical focus and the level of analysis in this research are the census tracts in Massachusetts. The reason for setting the level of analysis to the census tracts, not the person-level data, is to verify the geographical relevance in the relationships between tenure and socio-economic factors.

Through this research, I will deepen understanding of the systemic factors that contribute to determining the form of tenure (In this research, I limit the focus to the "tenant") and translate these findings to come up with policy implications and the need for further research.

I. INTRODUCTION

2. Hypothesis

My hypothesis is that median income, race/ethnicity, mode of transportation, and nativity have statistically significant relationships with the percentage of the tenant population. More specifically, I expect each variable to have the following relationships with the percentage of the tenant population.

- **Median Income**: In the tracts where the median income is higher, the percentage of tenant population is lower.

- **Race/Ethnicity:** In the tracts where the ratio of people of color outnumbers the ratio of the white population, the percentage of tenant population will be higher.

- **Mode of Transportation:** In the tracts where the rate of using alternative transport (public transportation, bicycle, or walking) is higher than cars, the percentage of tenant population will be higher.

- **Nativity:** In the tracts where the percentage of the foreign-born population is higher, the percentage of tenant population is higher.
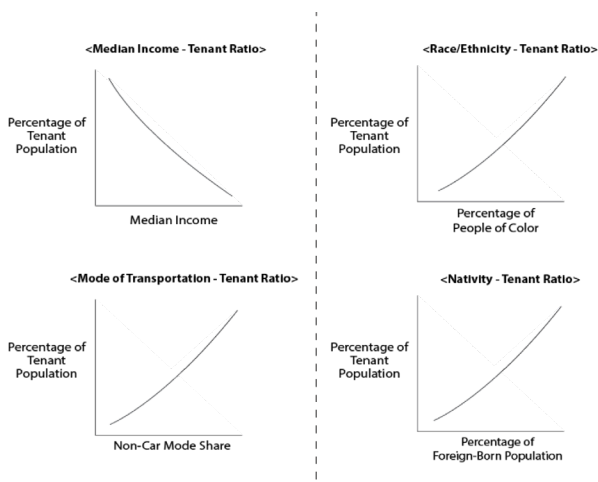
The graph on the left visualizes my initial thoughts on the relationships between the percentage of tenant population and the socio-economic variables.

II. DATA

1. Dataset

The sample for this study comprises all the census tracts in the state of Massachusetts. Therefore the unit of analysis in this research is the "tract", and total of 2,346 census tracts are included in the dataset.

The data was compiled from the American Community Survey 5-Year Data (2014-2018). The reason behind this choice is because the estimates for small geographic areas with populations less than 20,000 people are only available for the ACS5.

2. Continuous Variables (3)

- pct_ten : Percentage of Tenant Population

- med_income : Median Income

- pct_foreign : Percentage of Foreign-Born Population

3. Categorical Variables (2)

- maj_race : Race/Ethnicity with the Highest Ratio
  (Category: White, Black/African American, Hispanic-Latino, Asian, Other, No Majority)

- maj_tra : Mode of Transportation with the Highest Mode Share
  (Category: Car, Public Transportation, Walking, Bicycle, Other, No Majority)

II. DATA

How Does the Distribution Look Like?

To get the sense of the basic characteristic of the variables, I undertook the summary statistics of each variables and created a histogram for each variable.

<Table1. Distribution of Continuous Variables>

| Statistc	Percentage of Tenant Population	Median Income	Percentage of Foreign-Born Population
Sample Mean	34.3 %	39221.19	16.8 %
Median	27.3 %	37908	13.3 %
Standard Deviation	24.7 %	15517.17	12.5 %
Interquartile Range	13.4 % to 52.7 %	28982.25 to 47765.25	6.62 % to 24 %
Population Mean (95% Confidence)	33 % to 35.5 %	38425.13 to 40017.25	16.1 % to 17.4 %

(1) Percentage of Tenant Population
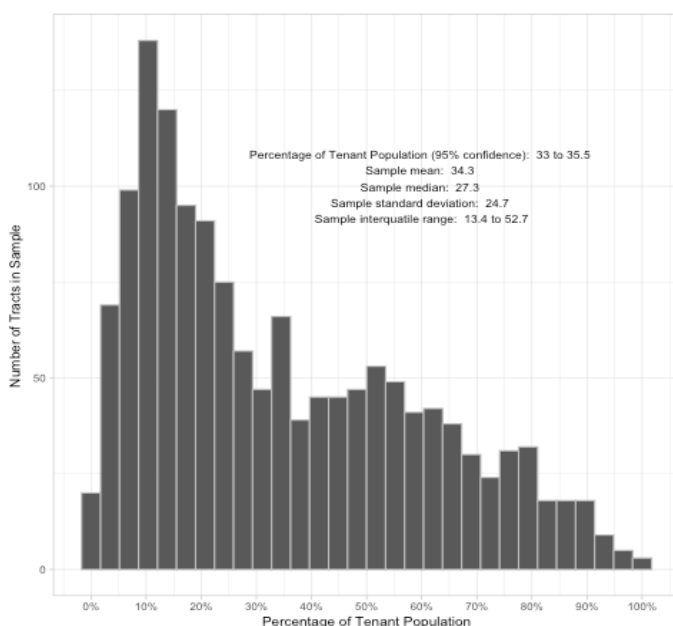
In 2018, the percentage of tenant population included in the sample was around 34.3%. The percentage of tenant population of half of those tracts were between 13.4% and 52.7%. The overall average was likely to be in the range from 33% to 35.5%. The diagram is right-skewed, indicating that lots of the sample tracts have low percentage of tenants.

II. DATA

(2) Median Income
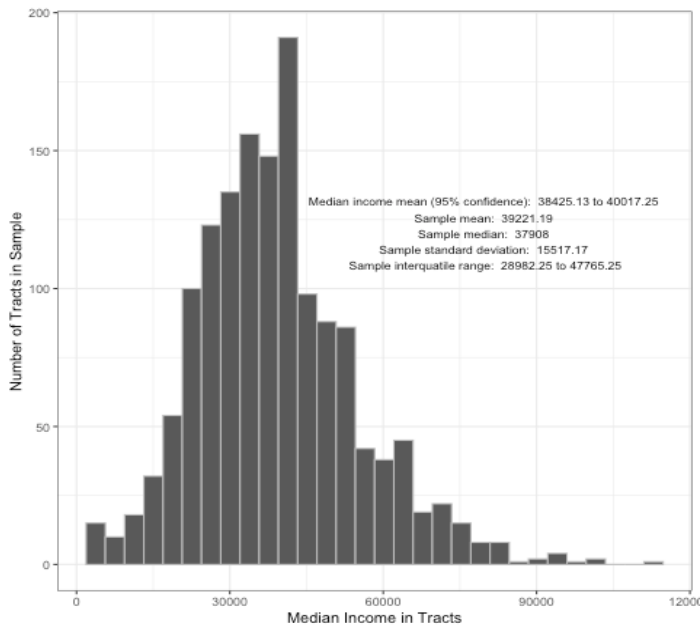
In 2018, the median income of tracts in the sample was around $39,221. The median income of half of those tracts were between $28,982.25 to $47,765.25. The overall average of median income was likely to be in the range from $38,425.13 and $40,017.25. The diagram is slightly right-skewed, indicating that there were a few tracts that have much higher median income than others.

(3) Percentage of Foreign-Born Population
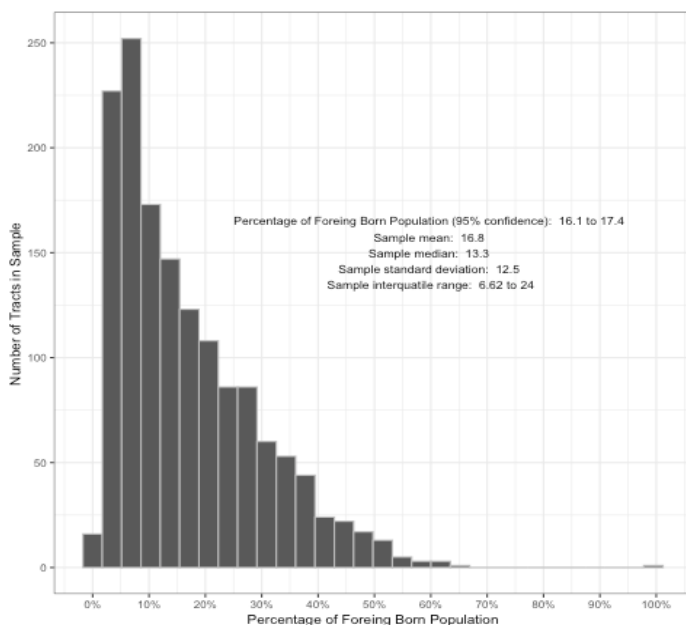
In 2018, the percentage of foreign-born population in the sample was around 16.8%. The percentage of foreign-born population of half of those tracts were between 6.62% and 24%. The overall average of the percentage of foreign-born population was likely to be in the range from 16.1% to 17.4%. The diagram is right-skewed, indicating that there a lot of tracts that have low percentage of foreign-born population.

(4) Race/Ethnicity with the Highest Ratio
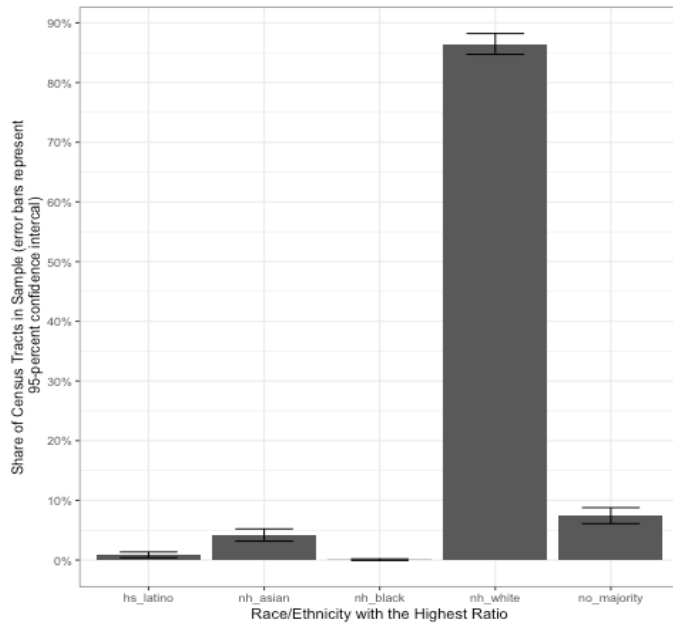
In 2018, in nearly 80% of the census tracts in Massachusetts, the majority race/ethnicity was white. Only about 4% of the census tracts in Massachusetts, hispanic-latino was the majority of population. There were very few tracts where the majority race is Black or Asian (2% and 0.4% for each). The ratio of tracts where there is no majority race/ethnicity was about 11%.

(5) Mode of Transportation with the Highest Mode Share
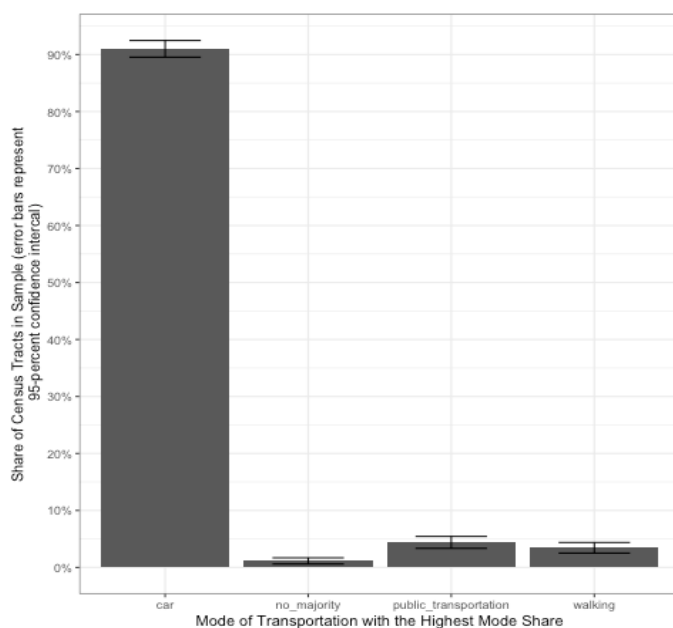
In 2018, in nearly 91% of the census tracts in Massachusetts, transportation with the highest mode share was driving. Only about 4% and 3% of the census tracts in Massachusetts, public transportation and walking accounted for the highest mode share. The ratio of tracts where there is no majority transportation was about 1%.

III. METHODS

1. Relationship Test

I tested the relationships between variables through Pearson's Correlation test and Tukey HSD test. The questions addressed through this process are 1) Is it statistically significant? (p-value) 2) How are the direction and strength of the relationship? (95-percent confidence interval)

(1) Pearson's Correlation
- Percentage of Tenant Population and Median Income
- Percentage of Tenant Population and Percentage of Foriegn-born Population

(2) ANOVA/Tukey HSD
- Percentage of Tenant Population and Race/Ethnicity with the Highest Ratio
- Percentage of Tenant Population and Transportation with the Highest Mode Share

2. Linear Regression

Following the test on the relationships, I examined more complex relationships between these variables controlling other variables to identify how much these socio-economic factors have predictive power on the percentage of tenant population.

The questions I'm going to address through this process are 1) Which model has the highest R-squared value?, 2) What is the 95-percent confidence interval of each variable?, 3) How are the results of linear regression and Pearson's Correlation/Tukey HSD test meaningfully different?
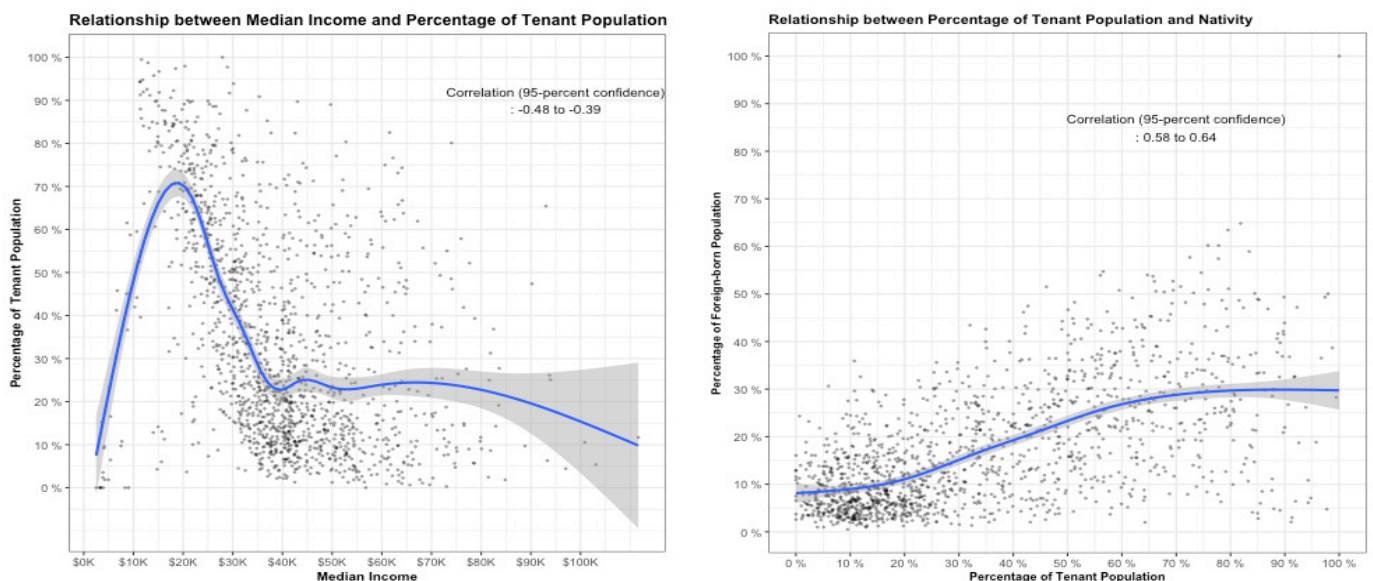
IV. RESULTS

1. Pearson's Correlation Test

According to the result of Pearson's correlation test (Table2, Figure1), the median income had a negative relationship and the percentage of foreign-born population had a positive relationship with the percentage of tenant population. The relationships were relatively moderate considering 95-percent confidene interval for correlation were around (+/-) 0.5. Both relationships were statistically significant with sufficiently low p-values.

<Table2. The Result of Pearson's Correlation Test>

Variable	95-percent confidence interval for correlation with the percentage of tenant population
Median Income	-0.48 to -0.4
Percentage of Foreign-born Population	0.58 to 0.64

Bold text indicates significance at 95-percent confidence level

<Figure1. Scatterplot of the result of Pearson's Correlation Test>

IV. RESULTS

2. ANOVA/Tukey HSD

According to the result of the ANOVA/Tukey HSD test (Table3, Figure2), the tracts with a majority Hispanic-Latino population or black/African American population had a higher percentage of tenant population than white-majority tracts with statistical significance. For transportation mode, the tracts where the public transportation shares the highest mode share tended to have a higher percentage of tenant population than the tracts where driving shares the highest mode share with statistical significance. However, other relationships between the categories under both variables were not statistically significant.

<Table3. The Result of ANOVA/Tukey HSD Test>

Variable	95-percent confidence interval for correlation with the percentage of tenant population
Majority Race	
White- Latino	**-0.49 to -0.37**
White - Asian	-0.47 to -0.14
White - Black	**-0.36 to -0.21**
White- Other	-1.03 to 0.13
Transportation with the highest Mode Share	
No Majority-Car	-0.53 to 0.17
Public Transportation-Car	**0.3 to 0.45**
Walking-Car	0.06 to 0.23

Bold text indicates significance at 95-percent confidence level

IV. RESULTS

<Figure2. Boxplots of the results of ANOVA/Tukey HSD Test>
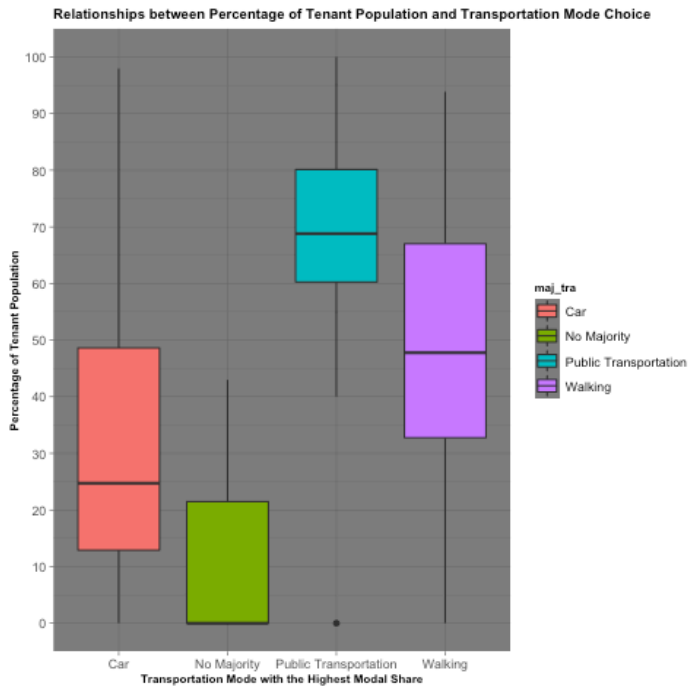
The tracts where the public transportation share the highest mode share tend to have a higher percentage of tenant population than the tracts where driving shares the highest mode share with statistical significance.
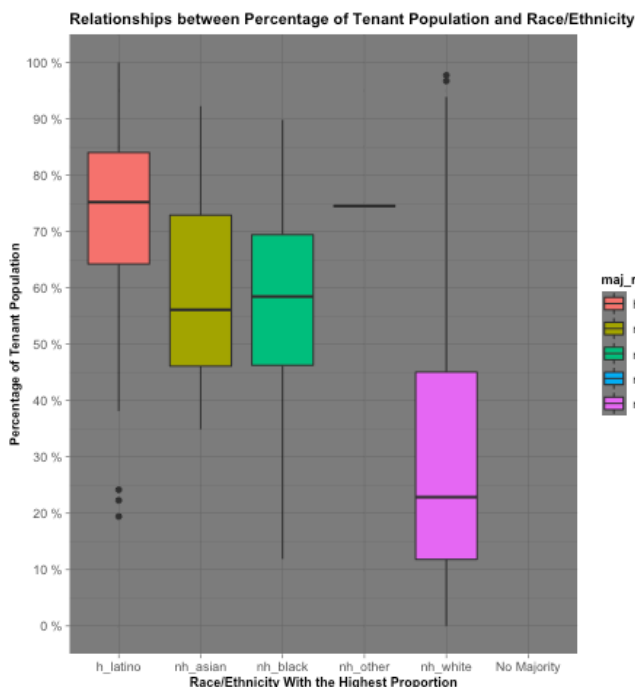
The tracts with majority Hispanic-latino population or black/African American population have higher percentage of tenant population than white-majority tracts with statistical significance.

3. Linear Regression

Based on the result of the relationship test, I have built four models. Model1, the initial model, sets the percentage of the tenant population as a dependent variable and median income, race/ethnicity, nativity and mode of transportation as independent variables.

Percentage of Tenant Population
= a*Median Income (10K) + b*Majority Race + c*Transport with the highest
Mode Share + d*Percentage of Foreign-born Population + e

For model2, I log-transformed the median income. For model3, I additionally mutated the variable, "percentage of foreign-born population" to the categorical variable by whether foreign-born population is the majority (>50%) or not. Lastly, for model4, I added the interaction term between the median income and transportation mode to the initial model (Table4&Figure4, p.14).

Shown in the graph comparing the adjusted R-squared values of four models (Figure3, p.13), as the R-squared value is the highest in the model4, I selected model4 as the best-fit model. With the R-squared value of 0.61, these independent variables explain 61% of the variation in the percentage of tenants in this dataset.

Median income turns out to have a statistically significant and negative relationship with percentage of tenant population. As the median income increases by $10,000, the percentage of tenant population decreases by 0.06%p, when controlling other variables.

In terms of transportation mode, percentage of tenant population is an aver-

IV. RESULTS

age of 0.38%p lower in the tracts where walking accounts for the highest mode share than car-dominant tracts. Also, the percentage of tenant population is an average of 0.55%p lower in the tracts where no single transportation mode takes the highest mode share than car-dominant tracts. These differences are all significant at a 95% confidence level. The other relationships between different categories and percentage of tenant population were not statistically significant.

This is a notable difference when compared to the result of Tukey HSD test which showed the statistically significant and positive relationship between public transportation and the percentage of tenant population, while other relationships were not even statistically significant. This implies there's more complex relationship between the tenure and mode of transportation, presumably influenced by other factors including median income or race/ethnicity.

I further delved into the potential interaction between median income and mode of transportation with the interaction term. The positive and statistically significant coefficient for the interaction between median income and transportation mode (except the "No Majority" with statistically insignificant p-value) tells us that median income has more of an effect on the percentage of tenant population for the tracts where public transportation and walking is the major transportation than it does for ones with cars.

More specifically, the interaction plot (Figure5) shows that given the condition that sample tracts' median income is zero, the tracts where the majority of people drive tend to have higher percentage of tenant population than the tracts with public transportation or walking. However, as the median income of tracts increases, the trend begins to reverse at some point that tracts where alternative
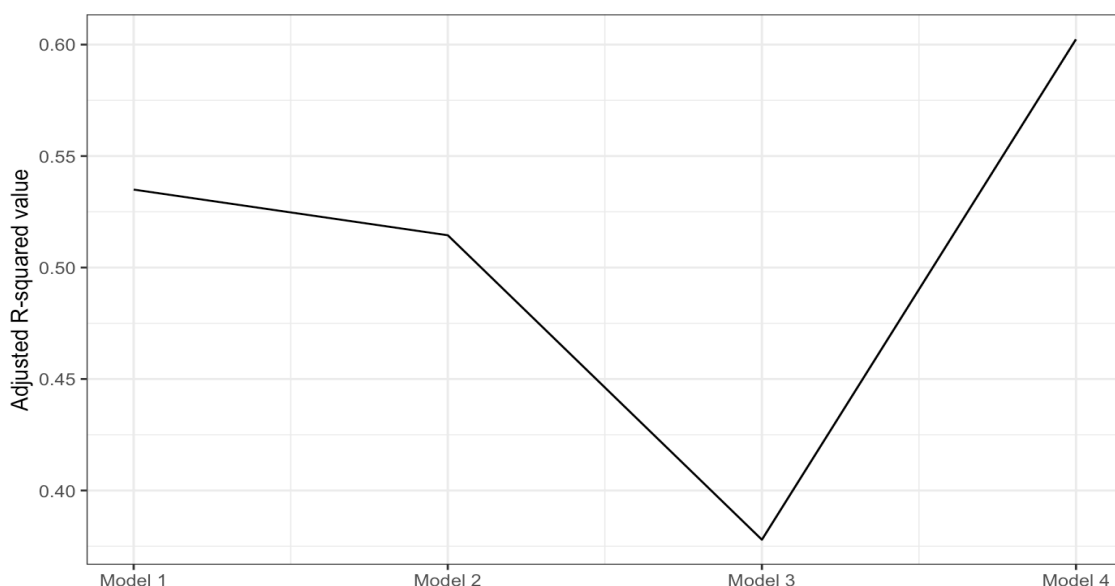
IV. RESULTS

mode is mainly used have higher percentage of tenant population than majority-driving tracts.

For race/ethnicity, with "white" population as the base case, the percentage of tenant population is an average of 0.17%p higher in the tracts where Hispanic-Latino population accounts for the highest racial proportion. However, the relationships between white and Black/Asian majority tracts were not statistically significant, when controlling other variables.

Lastly, the percentage of foreign-born population had a positive relationship with the percentage of tenant population. For every 1%p increase in the census tract's percentage of foreign-born population, the value of percentage of tenant population increases by 0.86%p. It is statistically significant at a 95% confidence level.

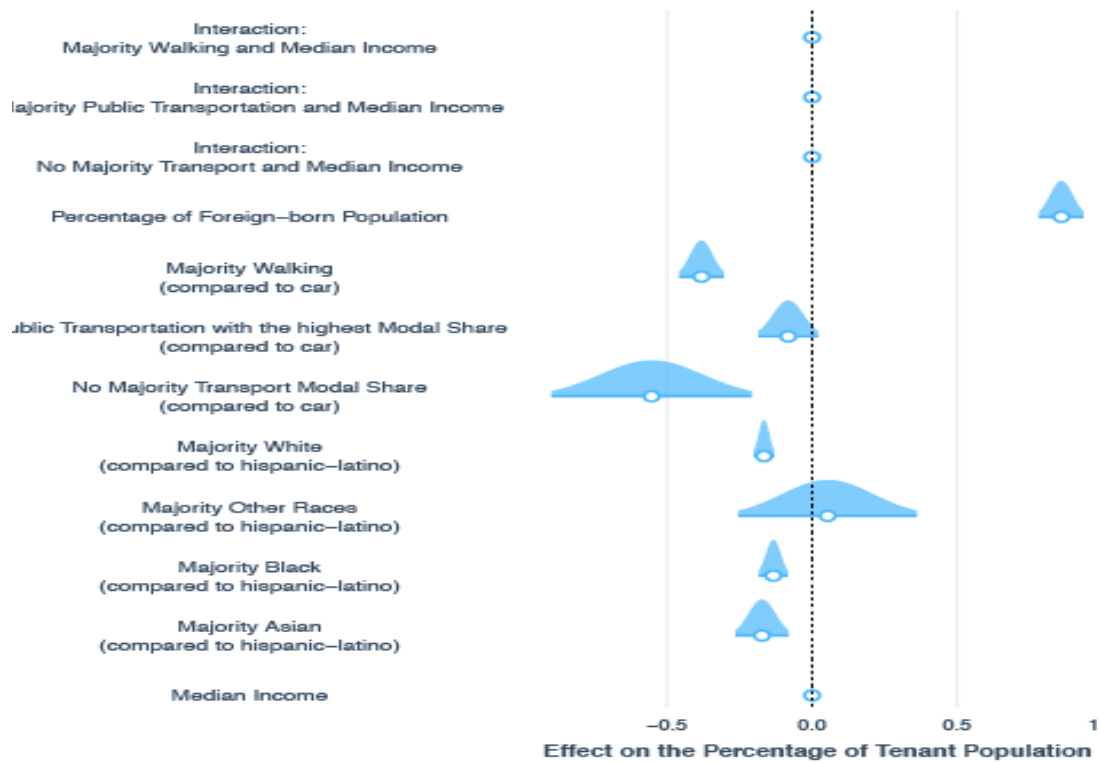<Figure3. Comparison of Adjusted R-squared Values>

IV. RESULTS

<Table4. Result of Linear Regression>

	Model1	Model2	Model3	Model4
Median Income(/10K)	-0.04 *** (p = 0.00)			-0.06 *** (p = 0.00)
Log(Median Income(10K)		-0.12 *** (p = 0.00)	-0.13 *** (p = 0.00)	
Majority Hispanic-Latino (compared to White)	0.19 *** (p = 0.00)	0.19 *** (p = 0.00)	0.32 *** (p = 0.00)	0.17 *** (p = 0.00)
Majority Asian (compared to White)	-0.01 (p = 0.81)	-0.01 (p = 0.91)	0.24 *** (p = 0.00)	-0.01 (p = 0.88)
Majority Black (compared to White)	0.05 * (p = 0.03)	0.07 ** (p = 0.01)	0.21 *** (p = 0.00)	0.03 (p = 0.15)
Majority Other (compared to White)	0.27 (p = 0.11)	0.27 (p = 0.12)	0.35 (p = 0.07)	0.22 (p = 0.16)
No Majority Transport Modal Share (compared to car)	-0.31 * (p = 0.01)	-0.39 ** (p = 0.00)	-0.37 ** (p = 0.01)	-0.55 ** (p = 0.00)
Public Transportation with the highest Modal Share (compared to car)	0.20 *** (p = 0.00)	0.19 *** (p = 0.00)	0.30 *** (p = 0.00)	-0.08 (p = 0.12)
Majority Walking (compared to car)	0.09 *** (p = 0.00)	0.04 (p = 0.16)	0.10 *** (p = 0.00)	-0.38 *** (p = 0.00)
Percentage of Foreign-born Population	0.85 *** (p = 0.00)	0.86 *** (p = 0.00)		0.86 *** (p = 0.00)
Majority US-Born			-0.03 (p = 0.52)	
Interaction: No Majority Transport and Median Income				0.15 (p = 0.13)
Interaction: Majority Public Transportation and Median Income				0.07 *** (p = 0.00)
Interaction: Majority Walking and Median Income				0.13 *** (p = 0.00)
N	1462	1462	1462	1462
R2	0.54	0.52	0.38	0.61

*** p < 0.001; ** p < 0.01; * p < 0.05.

IV. RESULTS

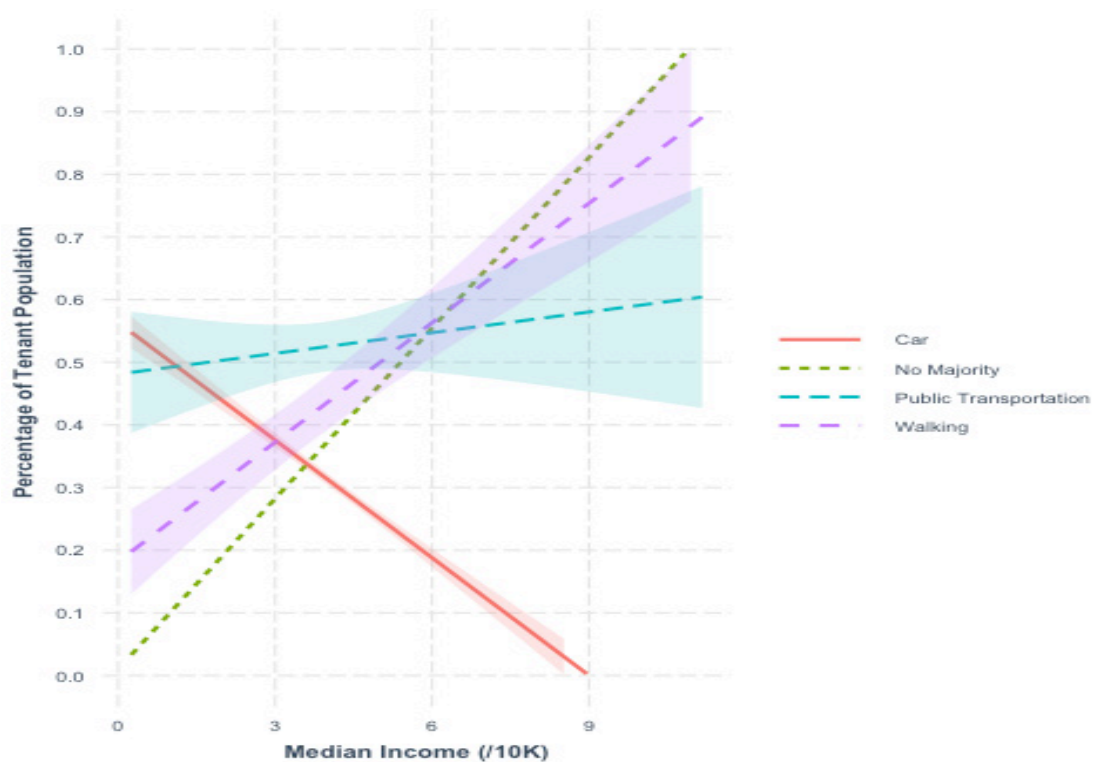<Figure4. 95-percent Confidence Intervals for Model 4 Coefficients>

<Figure5. Interaction Plot between Median Income and Mode of Transportation>

V. DISCUSSION

1. Hypothesis Testing

My hypothesis presuming the relationship between each socio-economic factor and the percentage of the tenant population turned out to be generally correct according to the result of Pearson's correlation and the Tukey HSD test.

However, the result of regression analysis that controlled other variables showed the relationships are more complex than they seem to be. More specifically, what rejected my hypothesis was that the percentage of tenant population was rather lower in the tracts where walking or no single majority transport account for the highest mode share than car-dominant tracts when controlling other variables. Public transportation had no meaningful value in predicting the ratio of the tenant population. This contradicted my initial hypothesis that I supposed the tracts using more alternative transport modes will have a higher percentage of tenant population, and it was also different from the result of the Tukey HSD test.

The relationship between the tenure and mode of transportation turns out to be not so simplistic when the interaction between the mode of transportation and median income is taken into consideration. In lower median income tracts, car-dominant tracts tend to have a higher percentage of the tenant population. However, as median income goes up, the tracts where public transportation or walking are the majority modes of transportation tend to have a higher percentage of the tenant population. This result implies that we cannot simplify the relationship between tenure and the mode of transportation as "positive" or "negative", and we have to consider more variables in predicting the relationship between the both.

V. DISCUSSION

2. Limitation and Weakness

If this research is to be more accurate in predicting the percentage of tenant population, the model should include more variables and interaction terms. In this study, I limited the number of independent variables to four and only added one interaction term. I will explore more variables that might have relationships with the percentage of the tenant population such as the rate of people aged from 0-17, the number of public schools, or the ratio of people living with their family. Also, additional interaction terms could be added between the mode of transportation and different independent variables to explore the reason behind the difference between the regression result and the relationship test.

Lastly, the fact that my research limited the sample data to the tracts of Massachusetts limits the generalizability of the findings of this research to different urban context of the United States.

VI. CONCLUSION

What can planners, policymakers and researchers do with these results? A primary contribution of this work is to identify whether socio-economic factors are associated with variation in the percentage of the tenant population. Considering that most of them (except a few categories under variables) had statistically significant relationships, this research confirms the common sense that tenure is highly shaped by the socio-economic status.

The interesting result that the relationship between the mode of transportation and tenure is highly affected by the median income of the given area can provide some meaningful insights for urban policy. In the urban contexts, say Manhattan, where the median income is relatively high, the mode share of public transportation and walking can predict a higher tenancy rate. However, in a more impoverished urbanized area, a higher mode share of public transportation and walking rather may be associated with a lower percentage of the tenant population. This could potentially imply - from the policy perspective - that we have to be more cognizant about the median income of certain areas when we try to make an impact on the composition of tenure through transportation interventions.

Considering the limitation of this research, the result calls for further research on what different factors might affect the relationship between the mode of transportation, income, and tenure. Beyond that, more variables must be included in the model to increase the model predictability of the percentage of the tenant population. To generalize this research implication to the different urban context of the US, the sample tracts should also be expanded.