Parking Ticket Analysis

Monday, May 22, 2023

1 INTRODUCTION 1

2 METHOD 2

3 RESULTS 5

4 CONCLUSION 10

5 REFERENCES 10

1 Introduction

Dear Ms. Weber,

Your article in The Morning Call, *Parking Problems*, highlighted and dissected a pertinent, important, and hot issue in Allentown: the steady and dramatic increase in parking tickets from January 2022 to February 2023.

We at LVJI are dedicated to researching, understanding, developing, and promoting a reimagined criminal justice system that is equitable and fair for all communities. While we focus on criminal justice, we strongly believe that all justice issues influence and impact criminal justice. Particularly influential to criminal justice is economic justice. The claims from Allentown residents that the Parking Authority is "attacking" people who are experiencing poverty invoke our attention.

LVJI is a data-driven organization, and we firmly believe that the best way to handle social justice issues is to use data to find what specific problem is occurring, why that problem is occurring, and what solutions address the root cause. The analysis in *Parking Problems* inspired us to examine these complaints. How are parking tickets and city demographics related? What are the strengths of those relationships? Are poorer areas of Allentown experiencing more enforcement and tickets than wealthier areas?

This document contains an analysis of the Allentown parking ticket data which you have graciously shared with us. I investigate which Allentown residents are most affected by parking enforcement through statistical correlations and geospatial mapping between parking tickets and area demographics like racial and ethnic composition as well as economic measures of poverty and income from the U.S. Census Bureau. The results are presented in the Results section. The Method section describes the details of my analysis from start to finish. The Method section is there for transparency and if you need clarification on setup or execution.

Thank you for your work in advocating for the residents of Allentown and for your generosity in allowing us to contribute to your conversation. I am more than happy to answer any questions you have on any material in this document at my contact information below.

With warm regards,

Victoria Wrigley
Data Scientist, Lehigh Valley Justice Institute
vwrigley@lvji.org
(484) 668 - 0434

2 METHOD

2.1 SETUP FOR GEOCODING

The dataset had tickets of the form #_ _ BLK STREET NAME, such as 900 BLK HAMILTON ST, which encompassed all 900 addresses on Hamilton St. I refer to these as BLKs. I wanted to match these BLKs with U.S. Census data to examine correlations with residential demographics.

The U.S. Census tabulates data into differently sized regions to be suitable for analysis on a scale as large as the state to as small as a city block. Figure 1 illustrates how the Census designates geographic entities. I chose to match BLKs to Census Block Groups, the smallest geography besides a Block. I chose this because Blocks are too small to accurately match to BLKs and Tracts are too big to see more granular differences in ticket issuances.

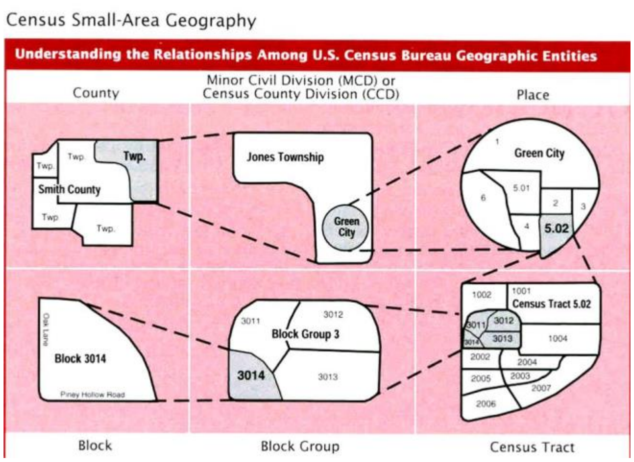


Figure 1: Diagram of U.S. Census Bureau geographies. Retrieved from https://pitt.libguides.com/uscensus/understandinggeography

Geocoding allows us to match BLKs with Block Groups. Geocoding is the process of matching a description of a location, such as an address, to a specific location on the Earth. The U.S. Census has a geocode service that takes as input a street address like 645 W Hamilton St, Allentown, PA 18101, and returns a multitude of descriptors for that location including its Block, Block Group, and Tract.

However, the parking ticket data provided ranges of addresses rather than specific addresses, which are what the Geocoder requires. Therefore, we obtained a database of addresses in Lehigh County from OpenAddresses.io, a free, open-source, and global address collection. The idea was to geocode every address in Allentown to obtain their Tract and Block Group and combine them across the BLK they belong to. Figure 2 below maps this process.

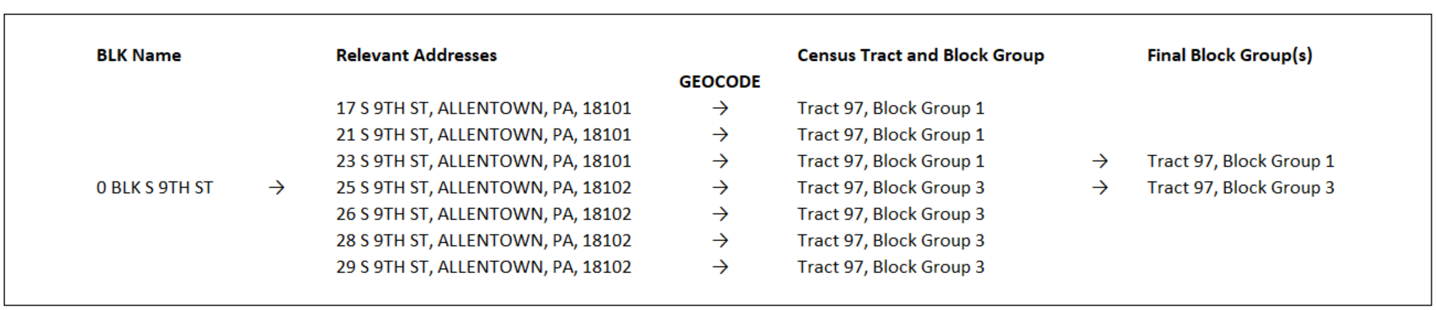


Figure 2: Flowchart of the geocoding process.

2.2 CLEANING THE ADDRESSES BEFORE GEOCODING

All real-world datasets require some form of data "cleaning" to ensure that analysis is straightforward, understandable, and consistent. The following describes the data cleaning process for the address database, so the Census Geocoder outputted accurate information. The goal is to input every address into the Geocoder and to have it return a Census Tract and Block Number.

We narrowed down the addresses in the database to the street names seen in the parking dataset and had ALLENTOWN listed as the city. The street names in the database were not always consistent with the streets in the parking dataset. For example, LINDEN ST in the parking dataset could mean E LINDEN ST or W LINDEN ST in the database. Further, if the street name had a direction identifier (ex. "W" for West) in the database, it was tacked onto the first letter of the street name (ex. W TURNER ST was coded WTURNER ST). To fix this for geocoding, I separated each street name by hand. Once this was complete, I put the addresses into the Geocoder.

2.3 MATCHING AFTER GEOCODING

There were three issues that arose that left a sizeable portion of BLKs unmatched. I implemented some more cleaning methods to remedy these.

First, not all parking BLK streets were represented in the address database. For example, there were tickets issued on E BIRCH ST, but the address database did not have any E BIRCH ST addresses because the street has no residential addresses there. Second, the Geocoder did not find a match for every address.

Any BLKs or parking street names that were not in the address database were searched for by hand via Google Maps and geocoded with the Census Geocoder. For all initially unmatched addresses and any unmatched addresses from the second round of geocoding, I matched their Census Tract and Block Group visually with this map of Lehigh County's 2020 Census Tracts and Blocks.

Third, some parking street names differed from the geocoded output. There were several reasons for this. First, directional identifiers were not always consistent. For example, TILGHMAN ST was the parking street name, the street name inputted into the Geocoder was TILGHMAN ST, and the geocoded output was W TILGHMAN ST. For many streets with a directional W, this corresponded to the base street name (ex. TURNER ST in parking input = W TURNER ST in geocoded output). To fix this, I created a "Street Decoder" which matches each parking street name to each geocoded street name. Second, In the parking dataset, some street names had ODD or EVEN attached to them representing the side of the street where the ticket was received. To approach this, I coded each geocoded street as either "EVE" or "ODD" depending on the street number of the address. Finally, some mismatches occurred because of typos in the parking street names. These included forgetting the "BLK" in the address name, having extra spaces, or misspellings. These were remedied by hand.

If there were no addresses in the database, no addresses on Google Maps, and I could not reliably visually assign a Block, I left the BLK unmatched. Out of 3,239 BLKs, 177 were unmatched (5%), accounting for 1,469 tickets (0.5% of all tickets). These tickets were not included in the analysis.

Tickets issued in a parking lot or deck were not included in the analysis. I made this choice because these spaces are more commercial than residential, and I do not believe residential Census data is accurate for these areas. The 4,528 tickets issued in 25 lots and 8 decks were not included in the analysis.

178 tickets had no street address listed. These were removed from the analysis.

The final analysis contained 277,364 tickets.

About 12% of all tickets (34,421) regarded overtime parking at parking meters, occurring mostly in downtown Allentown. These tickets likely do not reflect residential demographics. Therefore, for the demographic analysis, these tickets were removed.

2.4 CENSUS DATA

Census Block Group boundaries were obtained for mapping purposes via the 2020 Census.

Data on racial makeup, Hispanic ethnicity makeup, poverty rate, and median household income were obtained from the 2021 American Community Survey 5-year estimates. These tables were downloaded from the Census's data tabulation website. The following describes how I determined the values for the variables used in this analysis, along with definitions from the American Community Survey technical documentation (U.S. Census Bureau, n.d.).

NonWhite Proportion. The U.S. Census surveys race with a categorical question. "The racial categories included in the census questionnaire generally reflect a social definition of race recognized in this country and not an attempt to define race biologically, anthropologically, or genetically. In addition, it is recognized that the categories of the race item include racial and national origin or sociocultural groups" (page 115). The categories include White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and Some Other Race. "People may choose to report more than one race to indicate their racial mixture, such as "American Indian" and "White"" (page 115). I added all responses that were not "White" and divided them by the total number of residents to obtain the proportion of NonWhite residents in each Block Group.

Hispanic Proportion. "The 2021 ACS Hispanic origin question included three detailed checkboxes ("Mexican, Mexican Am., or Chicano," "Puerto Rican," "Cuban"), along with a "Yes, another Hispanic, Latino, or Spanish origin" checkbox, example groups, and a write-in area to collect additional detailed Hispanic responses" (page 77). This is a separate question from the Race inquiry, as "[p]eople who identify their origin as Hispanic, Latino, or Spanish may be of any race" (page 115). This variable represents the proportion of residents in each Block Group that checked at least one of those checkboxes.

Poverty Rate. The U.S. Census determines if a household is in poverty by comparing "the total income in the past 12 months of all family members with the poverty threshold appropriate for that family size and composition…[i]f the total family income is less than the threshold, then the householder together with every member of his or her family are considered as having income below the poverty level." I divided the number of households below the poverty threshold by the total number of households in each Group to obtain the poverty rate for that Group.

Median Household Income. Total income is measured by "the sum of the amounts reported separately for wage or salary income; net self-employment income; interest, dividends, or net rental or royalty income or income from estates and trusts; Social Security or Railroad Retirement income; Supplemental Security Income (SSI); public assistance or welfare payments; retirement, survivor, or disability pensions; and all other income" (page 85). The median is the household income value that is "in the middle" of all the household income values in the Block Group; 50% of all household incomes are below the median, and 50% are above the median.

Direct links to the datasets are in the References.

2.5 Matching with Census Block Group Data

Some BLKs existed in one Census Block Group, while others existed in up to four. To distribute BLK tickets between shared Groups, I gave each Group a proportion of the tickets equivalent to the proportion of households in that Group out of the collection of Groups. For example, there were 36 tickets issued on 1000 BLK N QUEBEC ST, and 1000 BLK N QUEBEC ST passes through Block Groups 1 and 2 of Tract 1.01. Groups 1 and 2 have 481 and 891 households respectively So, Group 1 received $36 * \frac{481}{481+891} = 12.3$ tickets and Group 2 received $36 * \frac{891}{481+891} = 23.3 = 23$ tickets.

2.6 Mapping the Data

I used ARCGIS Online, a data analysis and mapping program, to plot the parking ticket data and Block Group demographics over a street map of Allentown.

2.7 Correlation

I use correlation to measure how strongly Block Group ticket quantities are related to their demographic compositions. In statistics, correlation measures the strength of the relationship between two variables.

Correlation is a number between -1 and 1. Correlations close to 1 indicate strong positive relationships: as one variable increases, the other increases as well. Correlations close to -1 indicate strong negative relationships: as one variable increases, the other decreases. Correlations close to 0 indicate weak or no relationships: as one variable increases or decreases, there is no discernable change in the other. Table 1 below shows how correlations are often interpreted.

Correlation Value	Strength of Relationship
-1 < correlation < -0.75	Strong negative relationship
-0.75 < correlation < -0.50	Moderate negative relationship
-0.50 < correlation < -0.25	Weak negative relationship
-0.25 < correlation < 0.25	No relationship
0.25 < correlation < 0.50	Weak positive relationship
0.50 < correlation < 0.75	Moderate positive relationship
0.75 < correlation < 1	Strong positive relationship

Table 1: Correlations and their interpretations

In this analysis, I find the correlations between the number of tickets issued in Block Groups and four demographic variables: poverty rate, NonWhite proportion, Hispanic proportion, and median household income. I restrict the Groups to ones that have had tickets issued; if there were no tickets issued in a Block Group, it was not included in the correlation calculation.

3 RESULTS

3.1 TICKET TYPES AND QUANTITIES

I begin with an introductory analysis of ticket types, quantities, and locations. This section complements your initial analysis should you be interested in displaying your results graphically.

3.1.1 Most tickets concerned inspection, street cleaning, prohibited parking, and parking meter violations.

Five tickets made up 67% of all tickets. Ten tickets make up 85% of all tickets. The most common ticket issued was for out-of-date inspections (17% of all tickets). Combining the street cleaning and street cleaning repeat offender tickets, street cleaning violations become the most common ticket, with 55,951 tickets issued (20% of all tickets).[1] Table 2 tabulates the offenses; Figure 3 below graphs them.

Ticket	Count	Percentage
3JB - No Current Inspection	48,981	17%
6B - NP Street Cleaning	46,332	16%
2E - NP Anytime	35,716	13%
7A - OT Parking at a Meter	34,421	12%
3JA - No Current Reg/Inoperable	25,604	9%
4D - NP On Private Property	13,650	5%
2G - Too Close to Corner/Intersection	12,147	4%
8B - St Cleaning Repeat Offender	9,382	3%
2V - No Parking on A Sidewalk	7,928	3%
2C - NP This Street	6,841	2%

Table 2: Top 10 most frequent tickets issued and their counts.

[1] Please note that our numbers may be slightly different based on how we cleaned and prepared the data.

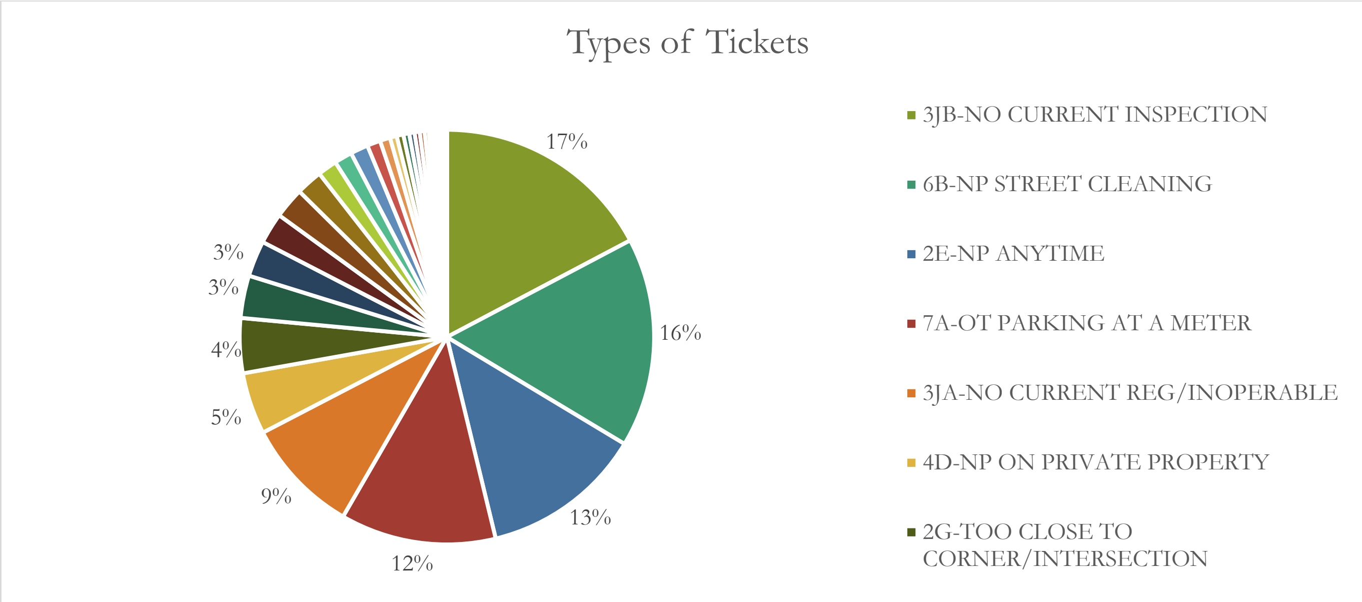
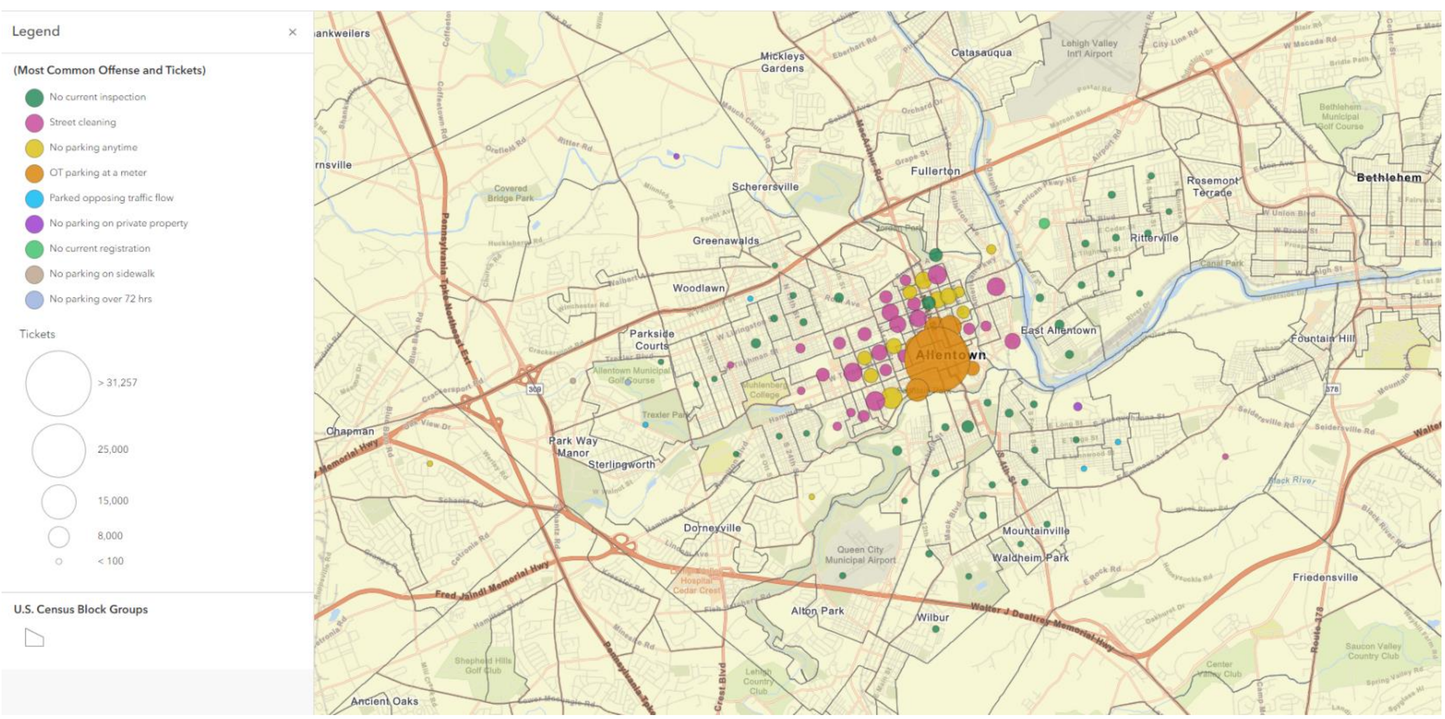


Figure 3: Pie chart of types of tickets issued.

3.1.2 Common ticket types varied by city area.

Map 1 shows the most common type of ticket issued in each Block Group and the total number of tickets issued in that Group. The largest share of tickets was issued in downtown Allentown, Tract 97 Block Group 1 (31,257 tickets; 11% of all tickets). Notably, this area contains the 900 block of Hamilton St, the first block of N 6th St, and the 600 block of Hamilton St, as you mention in your article.

If you were to get a ticket in Center City, it would most likely concern metered parking (orange). Farther out, you're most likely to get a ticket regarding street cleaning or prohibited parking (pink and yellow). On the outskirts of the city, you're most likely to get a ticket regarding an out-of-date inspection (green).



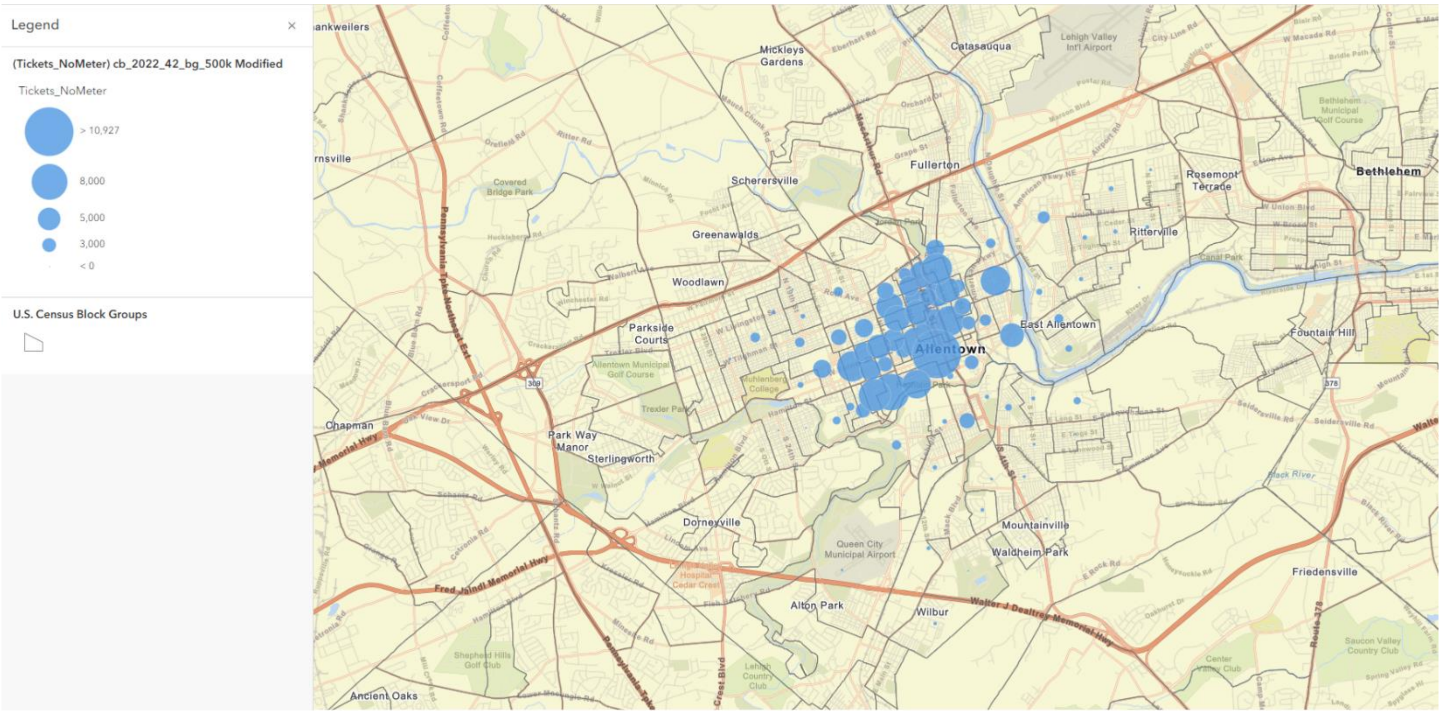
Map 1: Most common ticket types (color) and quantity of tickets (size).

3.2 Tickets and Demographics in Allentown

Metered parking spaces are generally not residential. If we want to examine city demographics and ticket issuances, we need to remove parking meter tickets so that the tickets are more representative of Allentown residents.

3.2.1 With metered parking tickets removed, ticket issuances were spread more evenly across Block Groups.

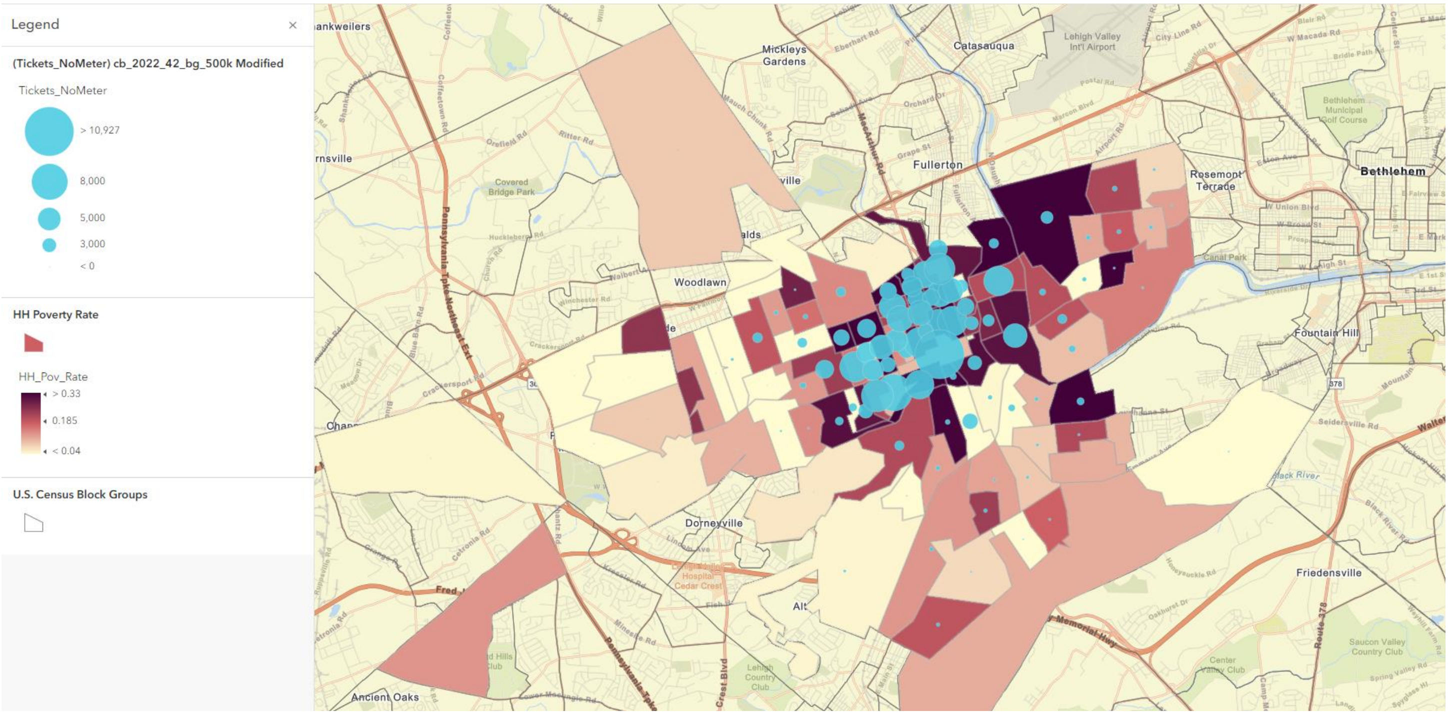
Map 2 shows that tickets were more spread out with metered parking tickets removed, though most still occurred in Tract 97 Block Group 1. There were 10,927 tickets issued here (5% of all tickets).



Map 2: Ticket quantities (with metered parking tickets removed) by Census Block Group.

3.2.2 Ticket quantities are weakly positively correlated with the Groups' poverty rates.

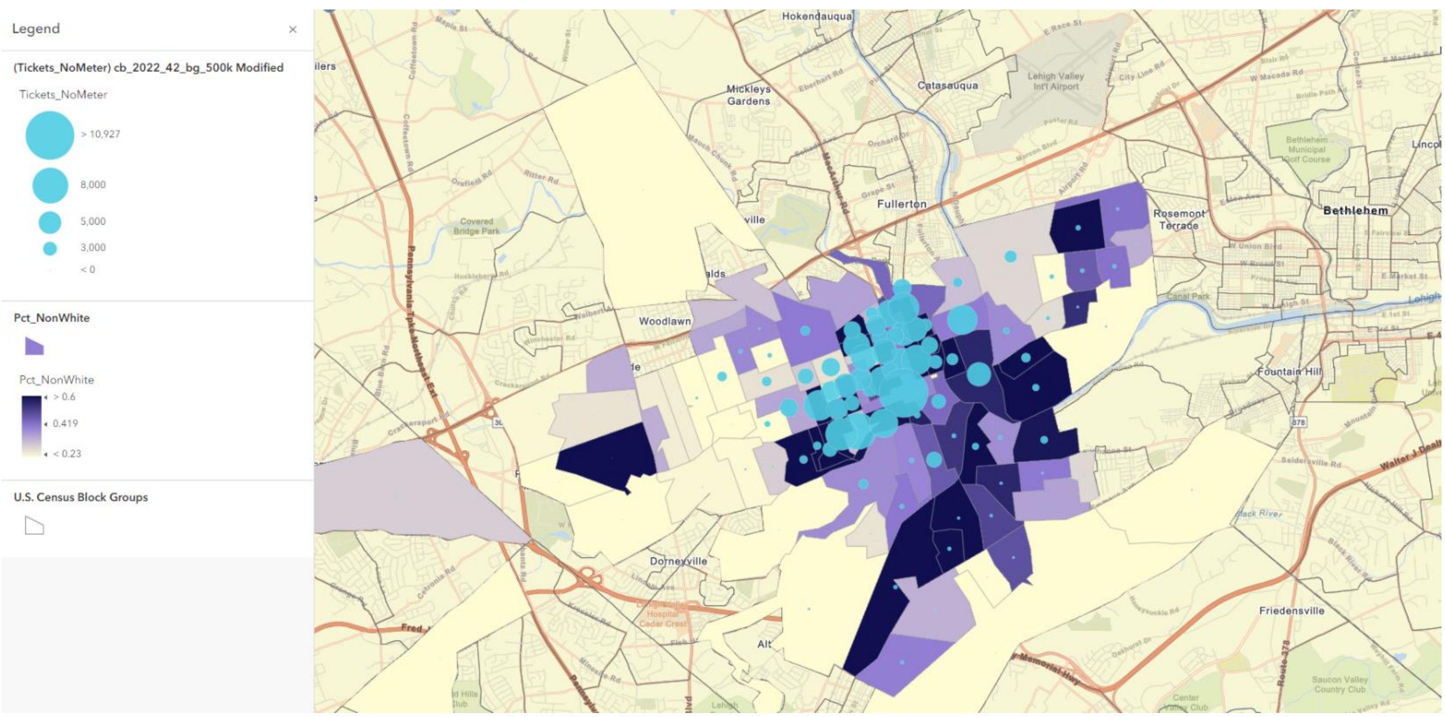
Map 3 examines the poverty rate of Block Groups (the proportion of households in the Block Group that are below their poverty threshold) with the number of tickets issued in that Block Group. As we've seen before, the blue circles demonstrate that tickets are clustered around Center City. The darker regions on the map correspond to Block Groups with higher poverty rates. The correlation is 0.41, indicating a weak positive relationship between the two variables. This means that Groups with higher proportions of households experiencing poverty tend to have more tickets issued.



Map 3: Ticket quantities (blue circles) and household poverty rate (shaded regions) by Census Block Group.

3.2.3 Ticket quantities are weakly positively correlated with the Groups' proportion of NonWhite residents.

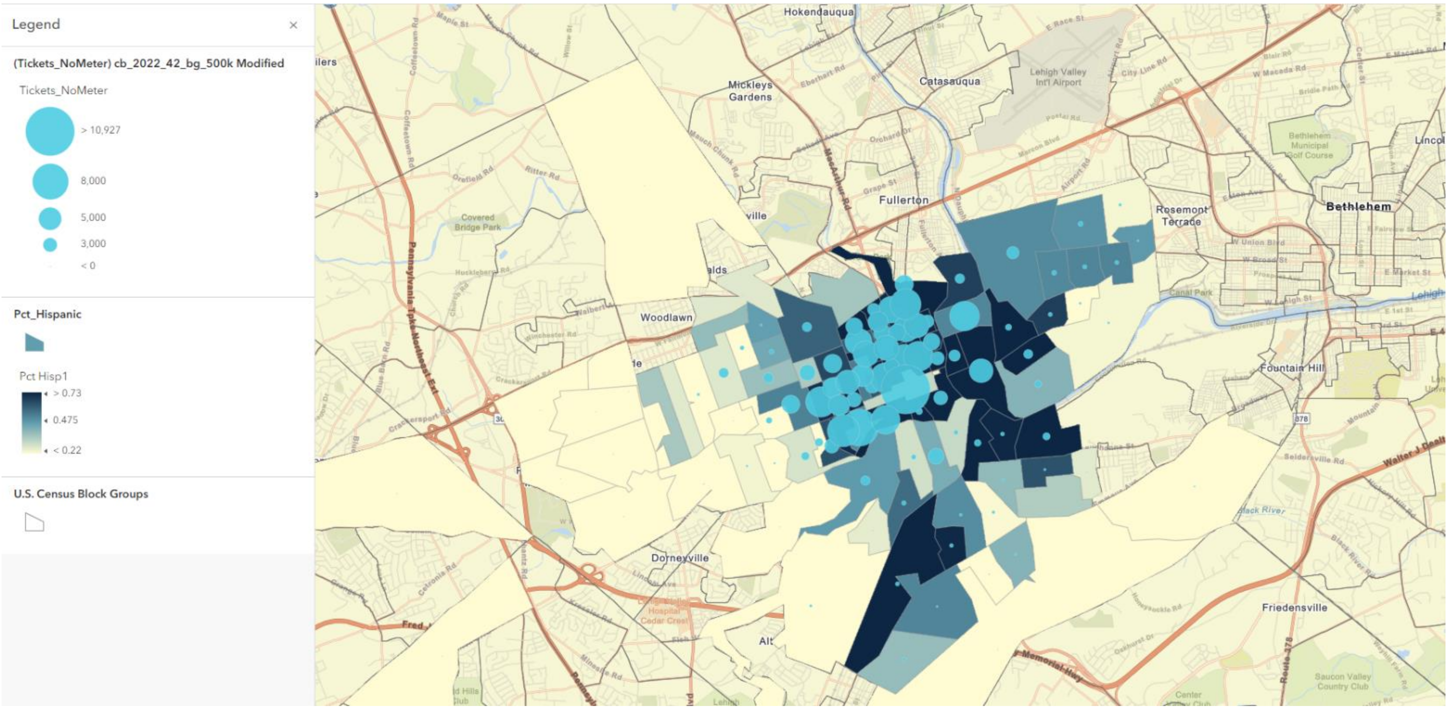
Map 4 shows that Block Groups with higher NonWhite populations tend to cluster near Center City, along with the tickets. The correlation is slightly higher, at 0.44.



Map 4: Ticket quantities (blue circles) and proportion of NonWhite residents (shaded regions) by Census Block Group.

3.2.4 Ticket quantities are moderately positively correlated with the Groups' proportion of Hispanic residents.

Map 5 depicts Hispanic populations and ticket quantities. These two variables have the strongest correlation at a moderate 0.57.



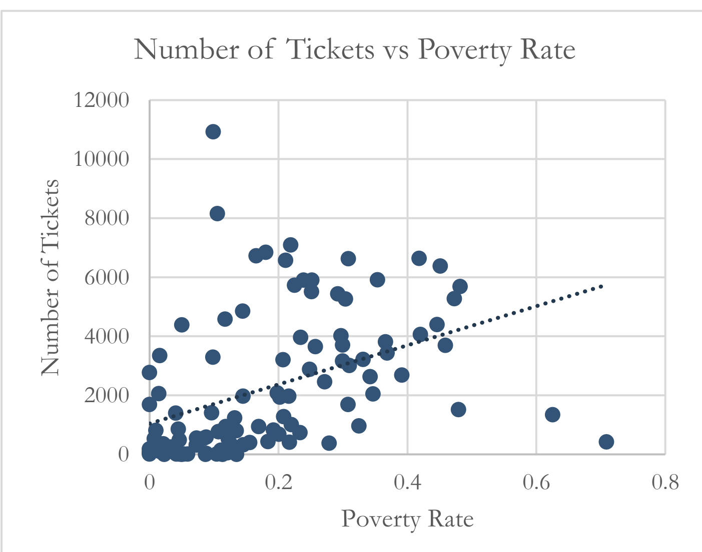
Map 5: Ticket quantities (blue circles) and proportion of Hispanic residents (shaded regions) by Census Block Group.

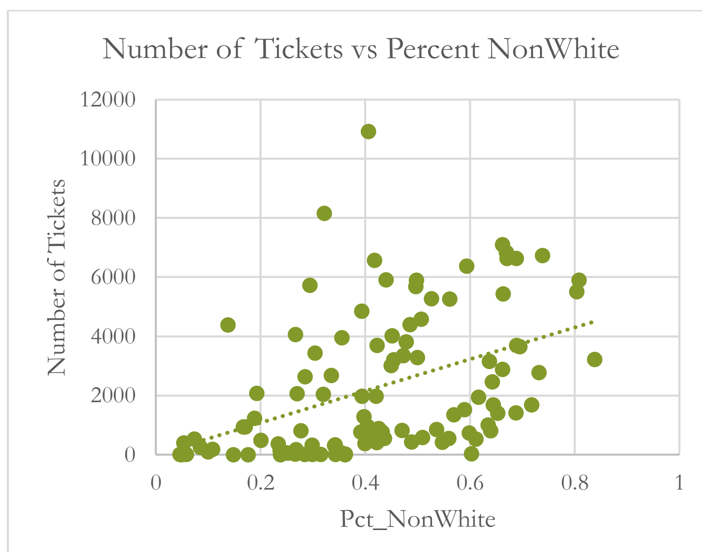
3.2.5 Ticket quantities are moderately negatively correlated with median household income.

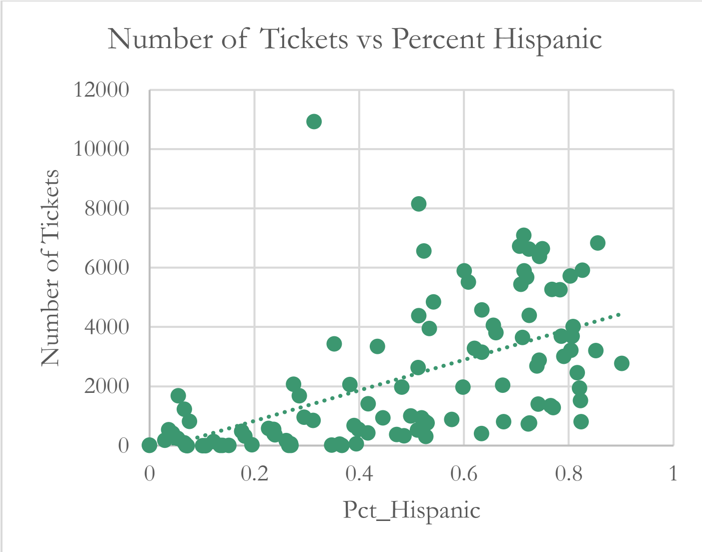
The correlation between ticket quantity and median household income is -0.52. This means that more tickets were issued to lower-income areas.

3.3 Scatterplots

To visualize the relationships between ticket quantities and Block Group demographics, Figure 4 below presents four scatterplots that depict how the number of tickets relates to the demographic variables.







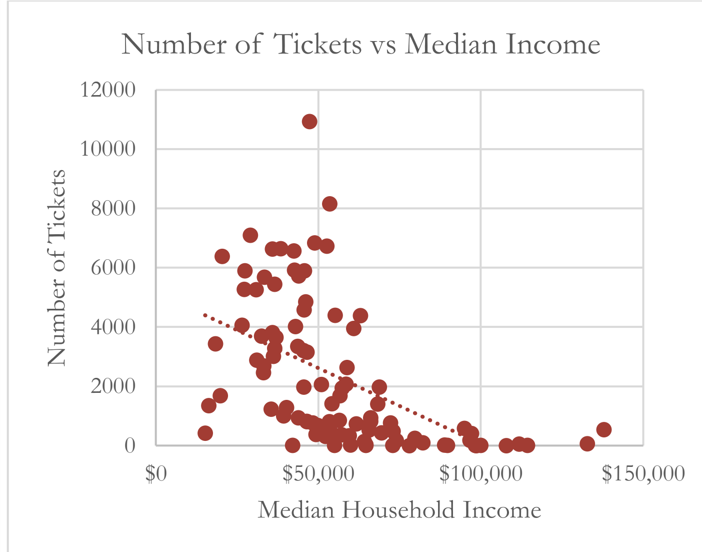


Figure 4: Scatterplots of (A) poverty rate, (B) NonWhite proportion, (C) Hispanic proportion, and (D) median household income by number of tickets issued in each Block Group.

4 Conclusion

These results demonstrate that Allentown's parking enforcement tends to have greater impact on the poorer, more diverse regions of the city. While reading these results, it is important to keep in mind that correlation does not mean causation. These results do not imply that the Allentown Parking Authority purposefully targets lower-income areas and areas with higher NonWhite or Hispanic populations. The correlations show that these areas receive greater numbers of tickets than their counterparts, particularly the regions of the city with greater Hispanic populations. Whether more violations naturally happen in these areas due to congestion or structure or because the Authority increases patrol and enforcement in these areas is still unknown. It is likely some combination of both.

5 References

American Community Survey and Puerto Rico Community Survey: 2021 Subject Definitions. (n.d.). U.S. Census Bureau. https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2021_ACSSubjectDefinitions.pdf

Openaddresses.io. (2022, November 11). *us/pa/lehigh - addresses – county* [GeoJSON Data file]. Retrieved from https://batch.openaddresses.io/job/314792

U.S. Census Bureau. (2021). *2017-2021 American Community Survey 5-year Estimates: Hispanic or Latino Origin* [CSV Data file]. Retrieved from https://data.census.gov/table?t=Poverty&g=050XX00US42077$1500000&tid=ACSDT5Y2021.B17017

U.S. Census Bureau. (2021). *2017-2021 American Community Survey 5-year Estimates: Median Household Income in the Past 12 Months (In 2021 Inflation-Adjusted Dollars)* [CSV Data file]. Retrieved from https://data.census.gov/table?t=Income+(Households,+Families,+Individuals):Income+and+Earnings&g=050XX00US42077$1500000&y=2021&tid=ACSDT5Y2021.B19013

U.S. Census Bureau. (2021). *2017-2021 American Community Survey 5-year Estimates: Poverty Status in the Past 12 Months by Household Type by Age of Householder* [CSV Data file]. Retrieved from https://data.census.gov/table?t=Poverty&g=050XX00US42077$1500000&tid=ACSDT5Y2021.B17017

U.S. Census Bureau. (2021). *2017-2021 American Community Survey 5-year Estimates: Race and Ethnicity* [CSV Data file]. Retrieved from https://data.census.gov/table?t=Race+and+Ethnicity&g=050XX00US42077$1500000&y=2021&tid=ACSDT5Y2021.B02001