Alysha Alloway
GIS 5577 Spatial Database Design and Administration
Spring 2019

Minneapolis Police Stops and Household Demographics

Introduction

The movement to make local government data open and accessible to constituents allows us as data scientists to explore new types of data and ask more complicated questions than we have ever been able to before. Police stops are some of the most controversial datasets released to the public in recent years, as more and more civilians question racial bias in policing and interrogate the failures of our police system to treat everyone equitably. I have been personally working with the Minneapolis Police Stops data since it was released in 2017, and through this project explored demographic characteristics in Minneapolis neighborhoods, asking whether police stops were reflective of the neighborhoods they occurred in, as well as looking at general questions related to racial bias in police stops, including total counts of stops by race and spatial patterns of stops by race.

Database Description

This project database contains four tables: Minneapolis police stop data symbolized by point, Minneapolis neighborhood boundaries, ACS 2017 household demographic and income tabular data, and ACS 2017 census tract geographies. I used both shared attributes and spatial relationships to query these tables and answer my main questions. Please refer to the 'Finalized Entity Relationship Diagram' PDF in this project's Github repository for further details.

Data

I acquired these data from three sources: the City of Minneapolis Open Data Portal for neighborhood boundaries and the police stop data, American Factfinder for the ACS 2017 household race and income data, and TIGER for the census tract geographies. All data were in shapefile format except the ACS household data, which I downloaded as a CSV. I did not generate my own data for this project.

Methods: Loading Data

I used shp2pgsql on the command line to generate SQL create table statements for my shapefiles, and loaded those into my database on the command line by connecting to my database and copying in the SQl create table statement. I made my own create table statement in my database to load in the ACS household data, and then used the /COPY command on command line to copy the csv into the newly created table.

Data Analysis

For my first query, I decided to look at the highest stop counts grouped by 'prerace' (race as identified by the police officer), 'race' (race as identified by the person stopped), and gender.

prerace character...	race character var...	gender character...	race_count bigint
Unknown	Black	Male	16828
Unknown	Unknown	Unknown	13520
Black	Black	Male	10566
Unknown	White	Male	10330
White	White	Male	6755
Unknown	White	Female	5801
Unknown	Black	Female	5712
Unknown	East African	Male	3145

As we all know, black men are by far the most over-policed people in America. After looking at this query, I began to question the role of the value 'Unknown' in this data. Considering this data is input by police officers who are face-to-face with people, I think most would find it difficult to believe that police officers are 'color blind' when it comes to race, and cannot identify that many people's race. I would find it more likely the overuse of 'Unknown' prerace in this data is an effect of this data being made public. It is easy to assume that the police as an institution as well as individual police officers want to avoid being considered racist, and so avoid entering their own data about people's race, knowing the data is public and will be scrutinized by people like me! However, I think this table speaks for itself when it comes to questioning whether racial bias is apparent in Minneapolis police stop data.

As a companion to this question, I also looked at simple counts of stops by race and the percentage of stops each race category represent of the total data.

race character varyi...	race_count bigint	stop_pct double precision
Black	35728	29.85
White	26147	21.84
Unknown	22370	18.69
East African	5160	4.311
Latino	4216	3.522
Native American	3043	2.542
Other	2889	2.414
Asian	1471	1.22

Minneapolis Racial Demographics, 2010 Census

- **White: 64%**
- **Black or African American: 19%**
- **American Indian: 2%**
- **Asian: 6%**
- **Native Hawaiian/Pacific Islander: 0.1%**
- **Other race: 6%**
- **Two or more races: 4%**
- **Hispanic or Latino (any race): 11%**

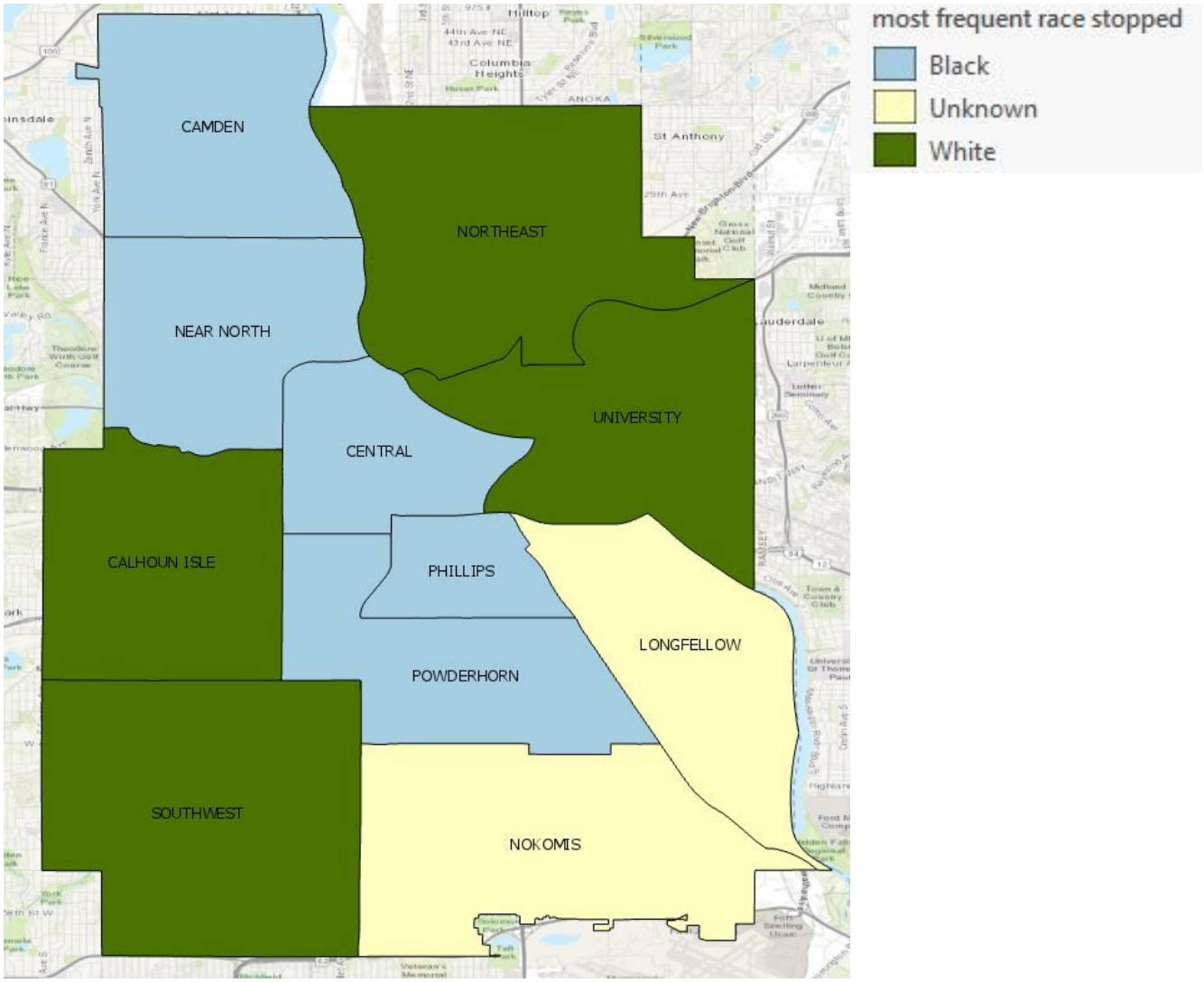
Comparing these percentages to the overall demographics of Minneapolis according to the 2010 Census, it is particularly notable that Black and East African people make up 34% of police stops, but only 19% of the population, while White people make up 22% of stops and 64% of the population.

For my second query, I looked at stop counts by race for each neighborhood.

stop_count bigint	race characte...	neighborhood character varying
10603	Black	Near North
6242	Black	Central
4563		Central
4389	White	Northeast
4035	Black	Powderhorn

Null and unknown values comprise a significant portion of the police stop data, but as I discussed earlier I chose not to exclude these values from my queries because I believe it says a fair amount about the people creating this data. Near North is by far the most policed neighborhood in the city.

I also looked into the most common race stopped in each neighborhood, and visualized it as a layer in Pro:

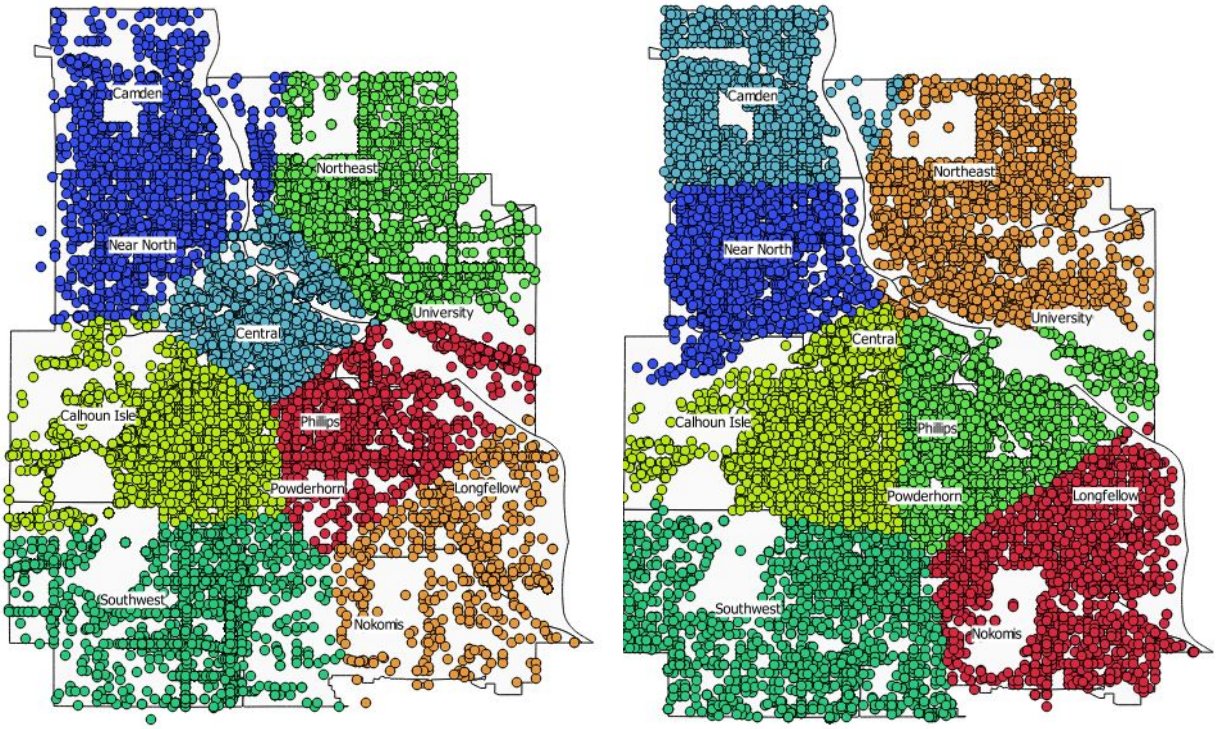


For my third query, I decided to look deeper into specific census tracts. I looked at the number of stops, by race, for each census tract, while also selecting income and household race variables for each tract.

stops_count bigint	race character v...	geography_label character varying	neighborhood character var...	median_income integer	nonwhite_households integer	total_households integer
2191	Black	Census Tract 1044 Hennepin County Minnesota	Central	59500	240	1140
1633	Black	Census Tract 22 Hennepin County Minnesota	Near North	43542	284	386
1560		Census Tract 1044 Hennepin County Minnesota	Central	59500	240	1140
1357	Black	Census Tract 1021 Hennepin County Minnesota	Near North	39777	525	707
880	Black	Census Tract 1028 Hennepin County Minnesota	Near North	31827	794	995
867	Black	Census Tract 1029 Hennepin County Minnesota	Near North	35833	274	440

From this query, I explored specific census tract ‘hotspots’, asking who is considered ‘suspicious’ where, and by whom.

For my final query, a spatial query, I used ST_ClusterKMeans partitioned by race to look at significant clusters of stops and overall stop density in Minneapolis.



K= 7, Race = White K = 7, Race = Unknown

Example Outputs From KMeans Query

6. Challenges and Limitations

I had some challenges loading all the police stop data into my database (in terms of the time it took) but now know that could have been alleviated using the shp2pgsql GUI instead of command line. Otherwise, Postgis documentation and StackOverflow posts were, as always, very helpful in helping me overcome my lack of fluency in SQL.

7. Solution that this database and queries provide.

The tools provided to citizens in the Minneapolis OpenData Portal are, in my opinion, insufficient. As a person familiar with working with datasets, the tools provided are difficult to use effectively, particularly when it comes to mapping visualizations. It is difficult to imagine a person who is less familiar with computing trying to use these tools in an easy and effective manner. The solution for mapping the data that they link to opens an ArcGIS Online web map, and I am certain that someone unfamiliar with GIS would become quickly overwhelmed, since it automatically maps the data my unique incident number and errors out trying to draw the complete layer.

While the data landing page does open on a density map of the stops (seen at right), the stops are not constrained to any familiar geography or labeled by neighborhood, and are instead gridded over Minneapolis. This is not a very effective way to communicate the density of stops to a lay person. The loading page map is not interactive beyond zooming in and out.

All this is to say that the solution this database provides is an easier way to explore, visualize, and query the data than what is currently provided. The queries allowed me to ask more specific questions than I could with the tools available on OpenData, looking deeper into how race corresponds with police stop density and counts, and also allowed me to have control over visualizing and representing the data.