Geographic Distribution of Libraries in New York City

BY RACHEL DOLS

Introduction

- Public libraries are a valuable community resource that can help lessen the divide between richer and poorer communities by offering free access to resources like computers, Internet, and job searching aids, which can improve opportunities for people in less-privileged communities.

- Therefore, it is important that less-privileged communities should have as much access to public libraries as more-privileged communities.

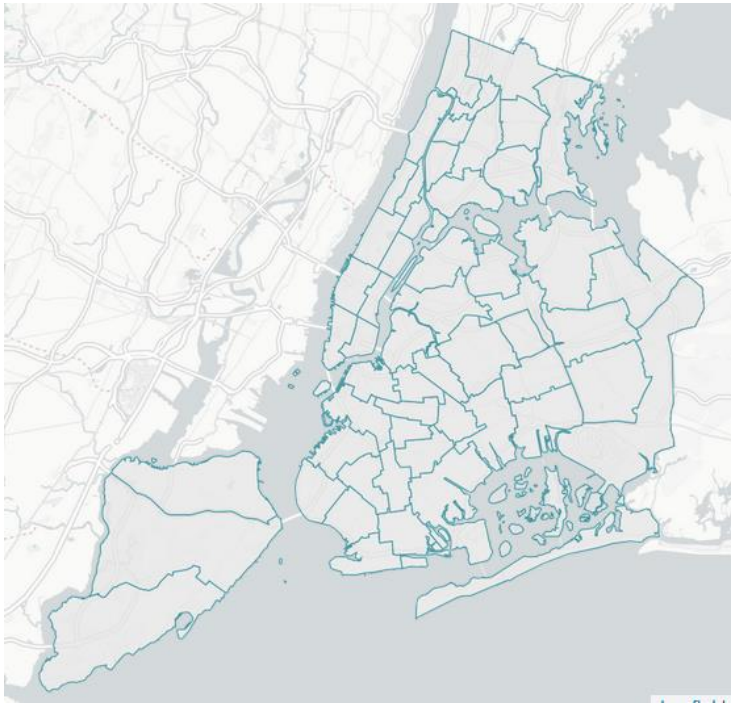
The Question:

Looking at New York City as one sample collection of communities:

- Are libraries in New York City evenly distributed between more-privileged and less-privileged communities?

Defining “Communities”

The United States Census has subdivided New York City into 55 zones called Public Use Micro Areas (PUMAs), for data aggregation purposes



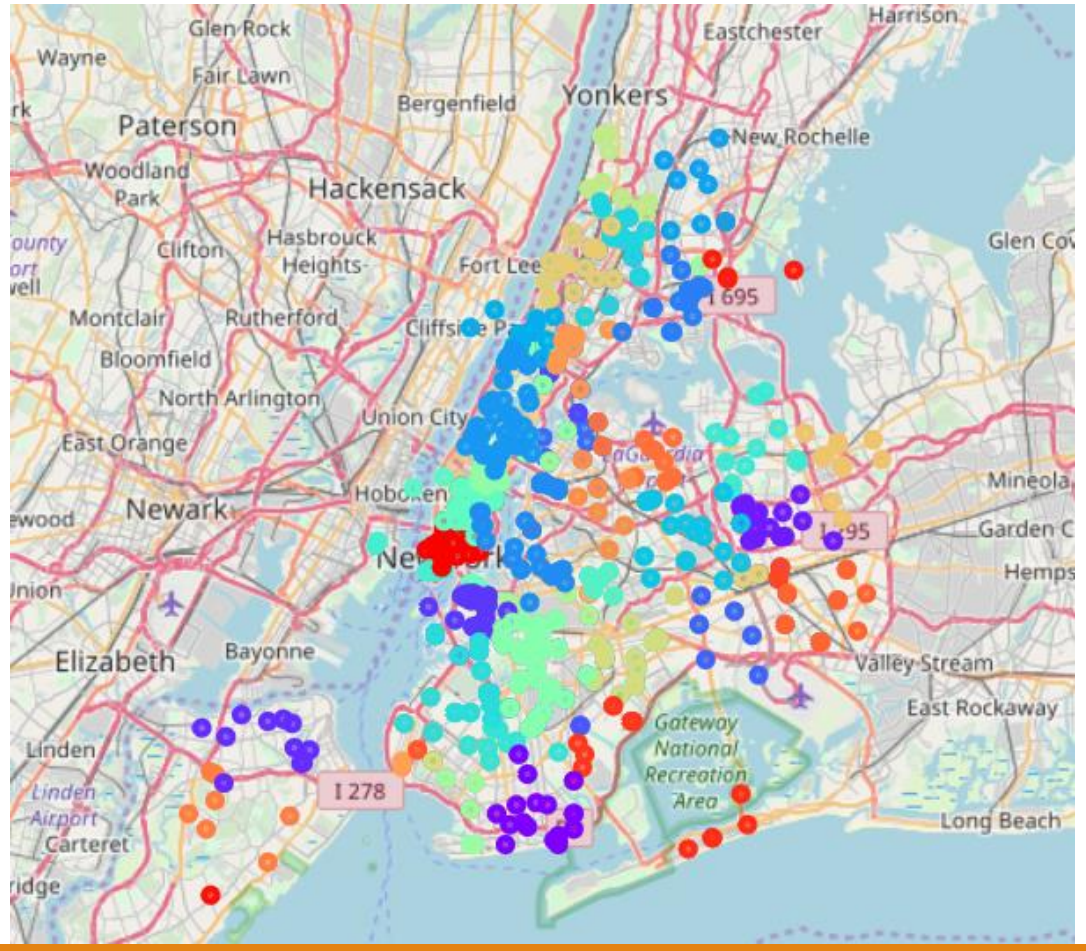
- Each PUMA consists of one or two neighborhoods, making a PUMA a reasonable size for a community in my research.
- Also, U.S. Census data is available in tables grouped by PUMA for ease of comparing these communities by their socioeconomic statistics.

Measuring “Privilege”

I chose 6 socioeconomic measures from the data in the United States Census:

- Median household income
- Percent of the population with a disability
- Percent of the population who is white
- Percent of the population below the poverty line
- Percent of the population born outside the United States
- Percent of the population over the age of 65

Library Distribution: Fair or Not?



- This map shows all the libraries in New York City, color-coded by PUMA.*
- We see that some are densely populated, and others have very few libraries per PUMA.

*The map was generated by k-means clustering of Foursquare search results. See Jupyter Notebook for full details.

Library Distribution: Fair or Not

- As we saw in the previous slide, some areas have more libraries than others, which leads to a reiteration of the research question:

- **The Question:** Is there a correlation between the number of libraries in each community and the socioeconomic measures for that community?

  - For example: does lower median income mean fewer libraries? More libraries? No correlation at all?

Methodology

Counting libraries

Used Foursquare API to count the number of libraries in each PUMA

- Calculated the center latitude and longitude for each PUMA
- Did a Foursquare “explore venues” search for the query “Library”, centered around each central lat/long pair and with a radius of 2500 meters
  - The 2500-meter radius was chosen because the typical PUMA size is about 5000 meters in length/width, so the radius was selected to be half that.
- For each search, the length of the Foursquare result list is the number of libraries in that PUMA.
- Appended the library count to an existing dataframe with the socioeconomic data for each PUMA

Analysis method

- I used linear regression to fit the number of libraries (the target variable) against each of the socioeconomic factors as independent variables.

- I then used the Pearson Coefficient to measure the strength of each variable's correlation.

Results

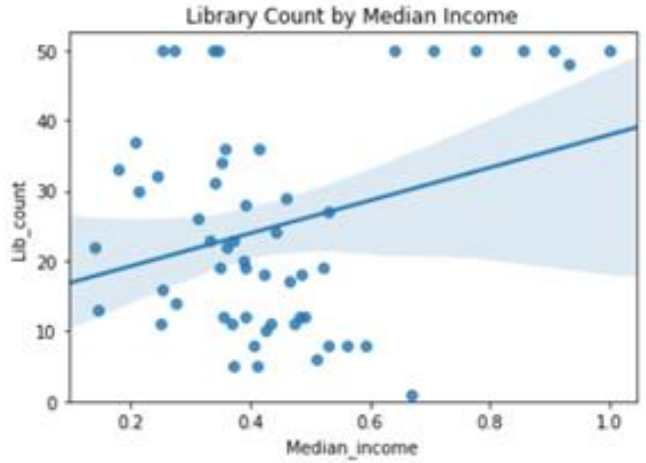
Pearson Coefficients and P-Values

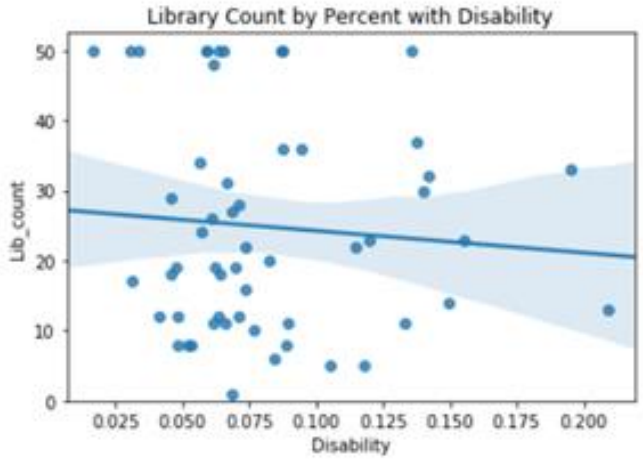
Variable	Pearson Coeff	P-Value	Correlation?	Certainty?
Median Income	0.2905	0.0314	Small-medium	Moderate
Disability	-0.0823	0.5503	Very small	Very Weak
White	0.2064	0.1305	Small	Weak
Poverty Rate	0.1397	0.3090	Small	Very Weak
Foreign-Born	-0.4148	0.00164	Medium	Strong
Over 65	-0.1692	0.2170	Small	Weak

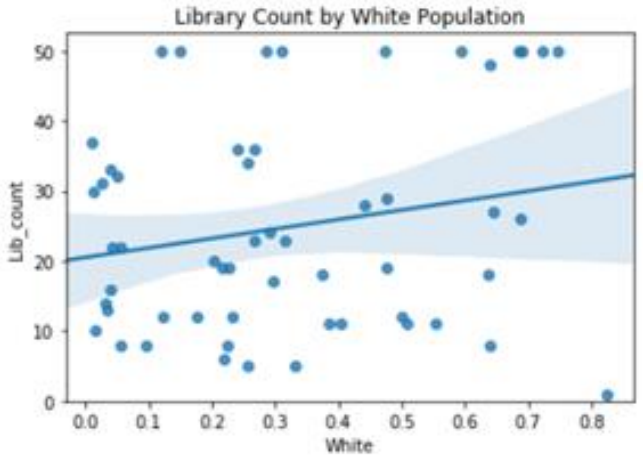
Note that none of these independent variables are strongly correlated with the Library Count variable (all Pearson Coefficients are less than 0.5.)

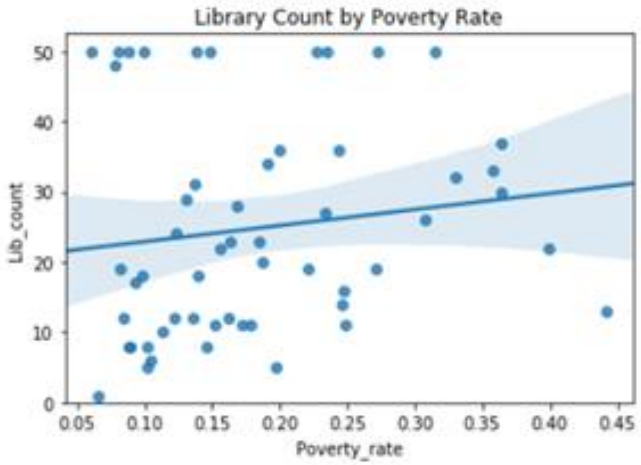
The two highest correlations are Median Income and Foreign-Born.

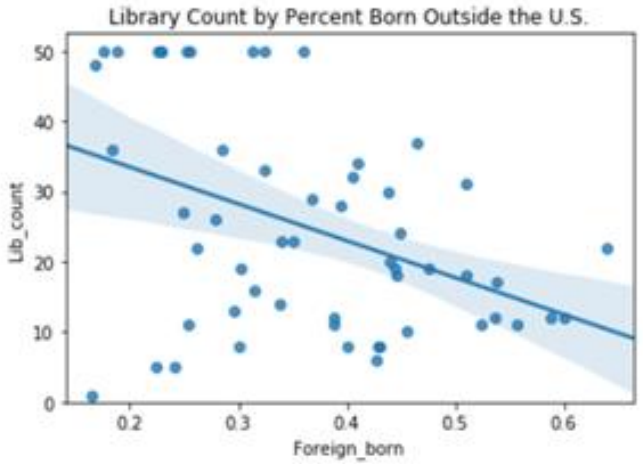
Regression Plots

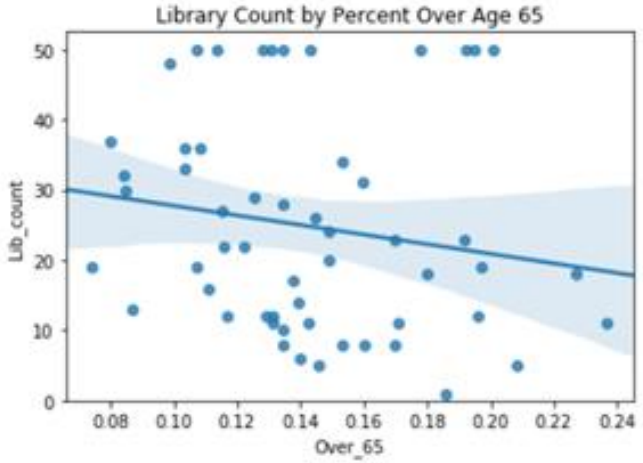












Discussion

- The regression plots show a great deal of scattering: the points do not fall very close to the fit lines.
  - This means that the number of libraries is fairly random and tends not to be very easily predictable by any of the independent variables.
- The variables with the steepest fit lines are “Median Income” and “Foreign-Born” (i.e., immigrant population)
- This agrees with the Pearson Coefficients, which had the highest magnitude for those two variables.

Discussion

- **Conclusion:** Libraries in New York City are distributed mostly fairly across neighborhoods, given that there are no *strong* correlations between the number of libraries and the socioeconomic factors of the neighborhoods.

- **However:** There are still *moderate* correlations between library count and two of the variables:
  - Neighborhoods with higher median income have more libraries than those with lower median income
  - Neighborhoods with more immigrants have fewer libraries than those with more American-born citizens

Conclusion

This is an analysis of only one city (New York City). It would be worth repeating the same analysis for other major cities to see if the results are similar (i.e., no strong correlations between library count and socioeconomic factors), or if other cities might have more of a gap in library access between more-privileged and less-privileged areas.

References

American Library Association. "Core Values of Librarianship." Retrieved from http://www.ala.org/advocacy/intfreedom/corevalues

NYC OpenData. "Public Use Microdata Areas (PUMAs)." Retrieved from https://data.cityofnewyork.us/Housing-Development/Public-Use-Microdata-Areas-PUMA-/cwiz-gcty

NYC.gov. "New York City PUMAs and Community Districts." https://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-population/census2010/puma_cd_map.pdf

NYU Furman Center. "CoreData.nyc." Retrieved from http://app.coredata.nyc/?mlb=true&ntii=hh_inc_med_adj&ntr=Sub-Borough%20Area&mz=14&vtl=https%3A%2F%2Fthefurmancenter.carto.com%2Fu%2Fnyufc%2Fapi%2Fv2%2Fviz%2F691a2b7c-94d7-46ac-ac4d-9a589cb2c6ed%2Fviz.json&mln=true&mlp=true&mlat=40.718&ptsb=&nty=2017&mb=roadmap&pf=%7B%22subsidies%22%3Atrue%7D&md=table&mlv=false&mlng=-73.996&btl=Sub-Borough%20Area&atp=neighborhoods

NYU Spatial Data Repository. "2016 New York City Public Use Micro Areas (PUMAs)." Retrieved from https://geo.nyu.edu/catalog/nyu-2451-34562