Quantifying the Lumps of Livability in New York and New Jersey

Gaurav Bhardwaj and Charlie Moffett
November 27, 2017
Paul M. Torrens, NYU
Urban Spatial Analytics

Study Design

The primary goal of our project was to examine the role of geography in determining patterns of livability in New York and New Jersey. Defining livability is a subjective task, however, so an important first step in our study design was selecting meaningful proxy variables for quality of life. We chose five variables from the U.S. Census Bureau's American Community Survey, a continuous investigation of key statistics related to housing, economics, education and more (US Census Bureau, 2017). For this project, we analyzed information from the 2011-2015 ACS 5-year estimates. These self-reported details, derived from 60 month's worth of collected data, provide a portrait of social and demographic community conditions.

In designing our study, we selected a handful of variables from a set of 20,000 features that reflect in our opinion the quality of life for individuals in a given community. Economic prosperity is a key component of livability, and for a large fraction of Americans, net worth is still by and large connected to homeownership. According to the National Association of Home Builders (Neal, 2013), 62% of the median American homeowner's assets are tied up in their primary residence. To represent this angle on prosperity, we collected median house value data for each Census Tract. We also collected Gross Rent as a Percentage of Household Income (GRAPI) data for NY and NJ residents that pay 35% or more of their paycheck for their place to live - a measure of excessive rent burden.

Median travel time to work is a measure we obtained to estimate the impact of commuting, which can take a toll on one's life for myriad reasons. Unemployment rate was another factor of community stability we took into consideration. The last livability measure in our study, percent bachelor's degree or higher, was an attempt to understand educational opportunity at the neighborhood level.

Data

The American FactFinder, a portal provided by the U.S. Census Bureau for streamlined access to community data, breaks down large volumes of survey response information into more manageable tables that are separated by topic (i.e. people, housing, etc.), geography (state, county, etc.), and race or ethnicity. Learning how to navigate the plethora of resources produced by an agency like the Census Bureau was a major component of our overall experience. We used FactFinder to distill our variables of interest at the Census Tract level for the states of New

York and New Jersey. This level of granularity, comparable to neighborhood size and stability, ensured that the geographies we compared were meaningful (MSU Libraries, 2017).

Our data span financial and occupancy characteristics, including information specific to employment status, commuting, educational attainment, home value, and renter housing costs. In collecting data that could serve as indicators of livability, we were careful to select metrics that also had meaningful coverage. In the assessment of variables for the project, for example, a negligible number of rows in our variable calculations for each Census Tract contained null values. Since respondents in both states encounter the same survey questions, we were able to ensure a continuous range for each variable over the entirety of our study area.

Unemployment numbers for civilians aged 16 years and older are presented as rates, which estimate the portion of a Census Tract that was able to work but could not find employment during the week of the survey. Information on education and rent were also collected as percentages, which proved useful as a means of accounting for population size in each tract. To assess educational attainment, we looked at the percent estimate of individuals with a bachelor's degree or higher. For renter housing costs, we captured the percentage of occupied units paying 35% or more of their household income toward rent. Our commuting metric was collected as the mean travel time to work in minutes, and home value as the median value in dollars.

Methodology

Our main intention, as discussed above, was to find correlation between the geographies with respect to certain variables. For appropriate granularity, we considered Census Tracts as they determine the geographies with similar number of people in a tract and would therefore be useful for comparison. We tabulated the data from American FactFinder in different spreadsheets where we kept only the attributes and the GEOID of each tract and exported it to ArcGIS as a geo-database file. At the start, we set the basemap coordinate system for ArcGIS to WGS 1984 so that all the other datasets that we import would automatically get converted to this coordinate system. We loaded the cartographic shapefile that we obtained from Census website and transformed it WGS 1984 coordinate system. On top of this, we loaded our geo-database file which was to be joined with the shapefile. Initially we tried extracting the shapefile for New Jersey from their open data portal, but that resulted in further difficulties, as the GEOIDs differed between our dataset and that shapefile.

Once we had merged our features data with the shapefile, we were set to run Local Moran's I on our data. We considered conceptualization of spatial relationships as contiguity edges only and ran the model to get another shapefile with our Local Moran's I attributing the high-high, low-low, low-high and high-low areas, as well as the areas without any significant correlation whatsoever. From the map, we can tell what are the clusters of high livability and what clusters have the mixed setup of low and high features. We ran Local Moran's I on all the features we selected for our project and saved the shapefiles for all of them. It was interesting to see how some of the features were concentrated around the metropolitan areas while some of the

features were sparsely distributed across the state. What we were more interested in was what happens behind the software and how we get the hotspots of association and how significant these hotspots are.

Results and Discussion

We had some encouraging system-wide results by way of our Global Moran's I testing. Spatial autocorrelation determines relation using calculated simple multivariate statistics and tells us about the strength and type of relationship between points (or Census Tracts) in space. In the case of bachelor's degree holders, commute time, and median house value, we found strongly positive spatial autocorrelation across the bi-state New York and New Jersey area with high confidence in our statistics. GRAPI and unemployment rate were also positively correlated, but not nearly as significantly as the others. The weak tendency to cluster for these two variables could mean that rent burden and lack of work aren't paired only with metropolitan areas but perhaps manifest independently of urban influence. Though in sociodemographics it is very unlikely to find negative spatial autocorrelation, our strongly positive results for the three variables led us to consider how significantly our measures of livability are dependent of values in the surrounding tracts.

After some meaningful analysis of our variables we also came up with local measures of correlation between them, such that residents with bachelor's degree or higher had lower rent burden as compared to other residents, and also (as one might expect) a person with a bachelor's degree or higher living in a decent enough place near to their workplace, his or her commute time to work was also low. Once we were done with correlation between variables, we started our analysis across the states to see how variables correlate in two different states. For New Jersey as well, the high-high and low-low areas were concentrated near the metropolitan areas and for the other parts there was either no correlation or variable high-low and low-high correlation.

Team Contributions

Our team worked together to transfer a lot of what we learned in lab work into meaningful analysis for the data we collected on our own. After exploring our options presented by the the 900 tabulations from the most recent ACS 5-year estimates, we worked together to apply our statistical approaches to detecting and measuring clusters, and to make sense of our urban science question around livability.

Charlie's Contribution: We inspected a variety of data formats to get our variable data, and Charlie helped with data collection and cleaning. He worked with the team in ArcGIS to load in our ACS tables and perform measures of global spatial autocorrelation. Charlie also contributed to the project report write-ups and interpretation of results.

Gaurav's Contribution: Testing for autocorrelation and measuring the strength of ACS variable clusters was paramount to our analysis, and Gaurav built maps of local clusters, showing high-

high relationships and low-low relationships. He also wrote about his experience with the methods set out by the project design and communicated results of the analysis.

References/Resources:

Neal, M. (2013, September 4). Housing Remains a Key Component of Household Wealth. National Association of Home Builders. Retrieved from https://www.eyeonhousing.org

MSU Libraries. Finding Census Tract Data: About Census Tracts. Retrieved on November 21, 2017, from http://libguides.lib.msu.edu/tracts

US Census Bureau. What is the American Community Survey? Retrieved on November 17, 2017, from https://www.census.gov/programs-surveys/acs/about.html