Mapping and Analyzing the Impact of Urban Agriculture in New York City

New York University, Center for Urban Science and Progress, Fall 2021

Nico Ampuero, Jeremy Rucker, Xiaolin Li

Abstract

Mapping Agricultural Production in New York City (M.A.P. NYC) is a project led by Stern's Center for Sustainable Business that aims to help the city and its partners reach equity and sustainability goals by cataloging all food growers within the city and mapping their attributes on a web-based platform. We use geospatial techniques and data gathered from official sources to implement that goal and explore the relationships between urban agriculture locations, socioeconomic factors, and health outcomes. Our platform provides administrative infrastructure for the city, crowdsources new data for future academic research, and creates a functional, user-driven interface between urban farmers, city agencies, and community members. Additionally, our statistical analysis offers an innovative perspective of urban agriculture as social infrastructure. Specifically, we focus on spatial relationships between non-commercial agriculture and an approximate measure of subjective well-being across census tracts in New York City. This research provides insight into the broad-reaching impact of urban agriculture in the city, particularly on communities that have been disproportionately affected by economic and health crises.

I. Introduction

A. Motivation

The United States Environmental Protection Agency defines urban agriculture as "part of a local food system where food is produced within an urban area and marketed to consumers within that area", which includes "animal husbandry, beekeeping, aquaculture, aquaponics, and non-food products such as producing seeds, cultivating seedlings, and growing flowers" (EPA, 2021). In the past several decades, this broad set of practices has emerged as a popular means to strengthen vulnerable, yet critical, urban systems (Reynolds, 2019). Columbia's Urban Design Lab, a leading research center in the space, also supports this approach: "[urban agriculture] lies at the nexus of a variety of issues which are seen as critical to the ongoing sustainability and livability of our urban environments: public health, healthy food access, green space, air and water quality, economic development, and community engagement." (Ackerman, 2012).

While academic, political, and community interest in urban agriculture has proliferated, information regarding *what* is being grown and *where* it is distributed has remained frustratingly scarce at scale. New York City is home to the largest number of active farms and gardens of any U.S. city, yet it lacks the metrics to quantify the impact of these assets, limiting its ability to achieve its outlined equity and sustainability goals for food production, distribution, and disposal networks (Food Forward, 2021). Despite persistent efforts from a diverse community of participants and researchers, the tangible impact of urban agricultural efforts remains largely unknown.

In 2016, researchers at the CUNY Urban Food Policy Institute pointed to this measurement gap as one of the major challenges to expanding agricultural activity, noting “there is no comprehensive, regularly updated urban agriculture database” in New York City (Cohen, 2016). Over five years later, there is still no such store of information, and the city still lacks the resources to build and maintain the requisite digital infrastructure. Meanwhile, a global health crisis has rendered urgent the need for food system resiliency and equitable access to healthy produce.

B. M.A.P. NYC Platform

M.A.P. NYC is a vital first step in building a centralized public network, engaging city farmers, and better informing researchers and policymakers focused on supporting and expanding the city’s self-sufficiency. At its core, it is a directory of all known New York City farms and gardens, with relevant details gathered from publicly available sources and an institutionally approved survey. That information is stored in a database that is editable directly from our permission based interface, and is presented visually with a design that facilitates exploration and information sharing. The platform is built with a modern cloud-based technology stack (Appendix Diagram 1) and makes extensive use of powerful visualization libraries for user-friendly interactivity (Appendix Image 1).

M.A.P. NYC is designed to be an evolving repository, crowdsourcing the most up-to-date information directly from farmers, gardeners, and city administrators, after initial database population by our team (Appendix Image 2). This ensures that data is current and accurate down to the weight of specific crops across thousands of agricultural lots in the city. The interactive map shows farms and gardens alongside food pantries, and farmer's market locations as these are direct-to-consumer points of interest for food access. Location counts, crop production, lot size are all aggregated to the community district level on the fly so as to account for any new, user-submitted data. The map also allows users to explore spatial relationships between agricultural production and select socioeconomic variables evaluated for significance in our regression analysis below (Appendix Image 3).

C. Urban Agriculture, Social Infrastructure, and Mental Health

While much research has focused on relationships between community agriculture and food access, the lack of detailed production and distribution data limits opportunity for statistical analysis. Moreover, when excluding the larger or more experimental commercial operations from consideration, "the literature strongly shows that a primary motivation of gardeners, managers and others was to produce fresh foods in a context of social interaction, community building and welfare." (Guitart et. al, 2012)

This means that non-commercial agriculture plays an important role not only as a budding stem of New York City's larger food system, but also as a pillar of the social infrastructure that supports the wellbeing of its residents. Flagship research of this relatively new concept comes from the work of sociologists Susan Lee Star and, more recently, Eric Klinenberg, who writes:

> *Public institutions, such as libraries, schools, playgrounds, parks, athletic fields, and swimming pools, are vital parts of the social infrastructure. So too are sidewalks, courtyards, community gardens, and other spaces that invite people into the public realm. Community organizations, including churches and civic associations, act as social infrastructures when they have an established physical space where people can assemble, as do regularly scheduled markets for food, furniture, clothing, art, and other consumer goods.*
>
> (Klinenberg, 2018)

As urban agriculture continues to permeate into discourse around urban planning and social equity, the role of non-commercial agriculture within the broader context of social infrastructure has earned theoretical precedent. For example, the MIT Media Lab's City Science

Group created a Well-Being Index using five social indicators—Community Connectedness, Safety & Security, Physical Health, Mental Health, and Diversity—as a way to address the impact of the built environment on overall psychological health at the city level (Orii, 2020), explicitly mentioning the positive impact of community gardens and green spaces on residents' subjective well-being (SWB). A case study in Melbourne, Australia corroborated this relationship, confirming the positive impact of social infrastructure on SWB through a distance-based GIS analysis of various educational, cultural, recreational, and health services across sub-geographies (Davern et al., 2018).

Our analysis builds on this research by regressing similar components of New York City social infrastructure (including non-commercial agriculture), along with various demographic and socioeconomic variables, against perceived mental health status, a proxy for SWB. We also incorporate a spatial component in the regression, expanding the theoretical basis of prior research to account for the spatial collinearity that impacts nearly all urban phenomena: "Most global statistical models require observations to be independent, but spatial phenomena, including health outcomes, are usually spatially correlated" (Ha, 2018). Research by Houlden et al. (2019) on the mental health effects of green spaces in London provides additional precedent for the spatial modeling approach.

II. Data

A comprehensive list of farms and gardens in the city was compiled by our team, with data from web searches, NYC Open Data, as well as shapefiles of NYCHA garden locations provided by the Department of Parks and school gardens from GrowNYC, a non-profit dedicated to food-related environmental programs. Tax lot sizes, an approximate measure of agricultural area, were pulled from the Department of City Planning's PLUTO database. Locations that share borough, block, and lot attributes were considered duplicates and only those originating from sources with the most spatially granular or reliable data were kept.

Measures of physical and mental health and socioeconomic data—some mentioned in the SWB literature above—were gathered from the US Census Bureau, the Centers for Disease Control and Prevention, and city agencies. Locations of educational, religious, cultural, and recreational buildings, derived from PLUTO building classification codes, were used to quantify the robustness of social infrastructure. For a full list of data sources, see Appendix Table 1. Data was compiled at the Census tract level (around 2,000 data points in the city), allowing for statistically meaningful regressions and providing a realistic understanding of distinct neighborhoods across the city.

III. Data Ethics and Impact Considerations

There is a checkered history between urban farmers and city agencies, so naturally privacy and ethics concerns arise when considering information-sharing initiatives. New York City has previously reneged on agreements with community gardens and land security remains a major concern for urban farmers, especially of minority communities (Guitart, et, al., 2012). Concerted efforts must be made to ensure that the value generated from this project is distributed equitably, especially since much of the requisite data has been collected from underserved areas throughout the city. To this end, our platform is designed not only for researchers and policymakers, but also as a community-building tool for urban growers to foster relationships with city officials, city residents, and among themselves.

Furthermore, it is important to consider methodological biases present in survey information collected from these populations. The social impact of urban agriculture is a primary concern for our analysis and the data used to develop our models may be disproportionately or inaccurately representing marginalized communities. For example, our dataset contains NYPD violent crime records, which may reflect crimes reported rather than crimes committed. Moreover, self-reported physical and mental health information is extremely subjective, especially when conducted on a broad scale and distilled into a single metric, as is the case with the CDC data we obtained. Finally, demographic information may be distorted by both the wording of subjective self-identifying categories as well as reticence by marginalized groups who lack trust in governmental institutions. Any conclusions or assessments derived from these metrics must take into account the nuanced context in which they were obtained.

IV. Methodology

A. Spatial Preprocessing

Each observation in our dataset represents, depending on the variable, either a census tract or service area buffer of half-mile radius around each tract population centroid in New York

City. We used centroids of 2010 vintage, as 2020 data has yet to be released by the Census Bureau. This circular area defines a widely used and accepted threshold of walkable distance to reach public spaces (Davern et al., 2018). Partially overlapping buffer areas ensure that social infrastructure is reachable by a larger swath of the population than simple tract boundaries. Other variables were obtained at, or aggregated to, Census tract level.

After excluding tracts in Staten Island, whose urban environment we deemed incomparable to the other boroughs, and those with low population (<1,000) and low walkability scores (<10), we were ultimately left with 1,818 out of 2,100 tracts with complete data across 26 variables, although just 19 were used for modeling. For a detailed explanation of our filtering methodology, see Appendix Figure 1.

B. Dependent Variable - Mental Health Status

In lieu of a more direct measure of SWB—obtained, for example, through a target survey encapsulating perceived "standard of living; health; achievements; personal relationships; community connectedness; safety and future security" (Davern et al., 2018)— we employed perceived mental health status from the ongoing CDC PLACES project as a proxy. This variable is also used by The Trust for Public Land to promote the expansion of green space in underserved areas (Chapman et al, 2021), and is one of the few mental health measures available at the census tract granularity. It represents the percentage of adults who responded as having poor mental health for at least 14 days over the last 30 days. Therefore, a *higher* value represents *worse* mental health and explanatory variables with *positive correlations* have an implied *negative impact* on mental health.

C. Independent Variable Preprocessing

After finalizing the list of independent variables deemed relevant to our analysis based on research cited above, we implemented a number of data processing steps to prepare for modeling. We then normalized all social infrastructure (SI) variables, i.e. counts of PLUTO classified buildings/open space and non-commercial agriculture, by population.

D. Modeling

With this dataset we built three types of regression models using Python modules: Lasso, Random Forest, and geographic weighted regression (GWR). For description of models, see Appendix Table 2. All three enjoy high interpretability - the first representing a hyperplane, and the second utilizing an ensemble of variable splits, the third providing localized interdependence measures. High interpretability is crucial in serving our primary objective of evaluating relative feature importance. We split the data 80/20 into training and test datasets, and used 5-fold cross validations to find optimal parameters for both the Lasso and Random Forest models (Appendix Table 2).

In preparing the GWR model, we performed a principal component decomposition (PCA) for the variables with high explanatory power and severe multicollinearity since a regularization penalty is not available for this spatial regression package. As a result, variables *cdc_physical*, *acs_income_log*, and *acs_bachelors* were replaced with a single PCA feature. Using the "sel_BW" feature of the mgwr package in Python, the GWR was fitted across the entirety of the dataset using a spherical coordinate parameter, applying the optimal bandwidth value of 215.

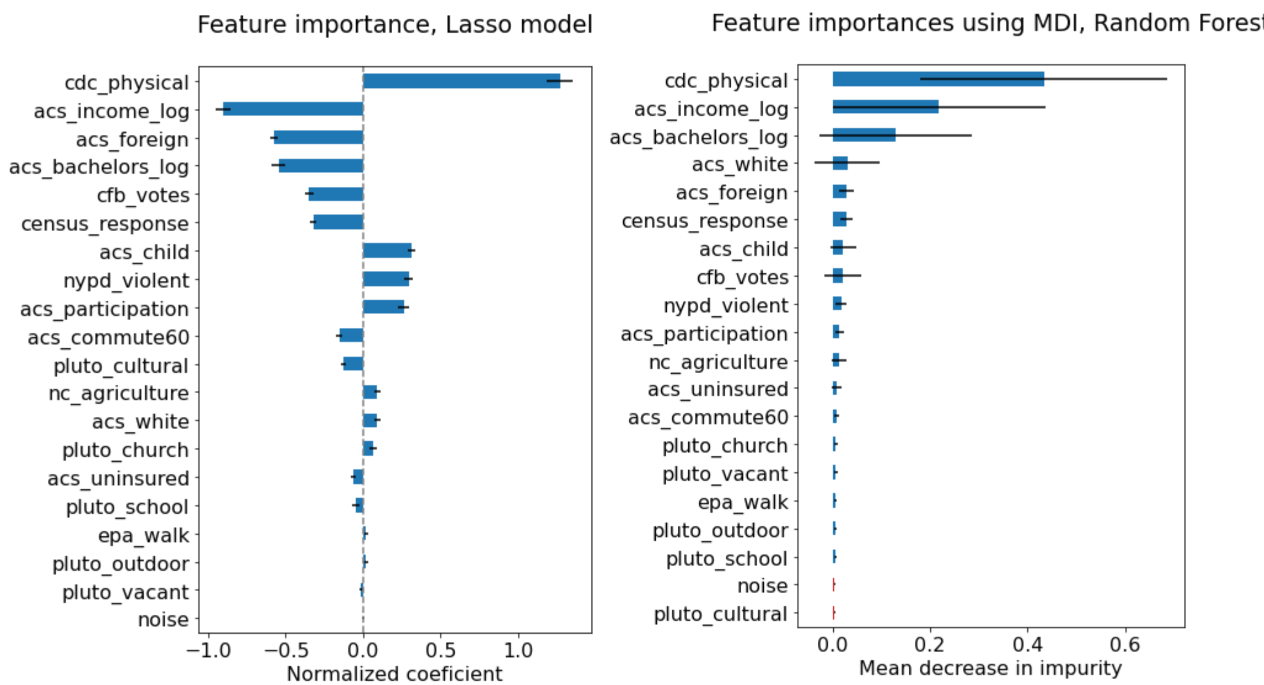
V. Results

A. Non-spatial Regression

Both the non-spatial OLS and tree-based models yielded high accuracy results: less than 0.7% out-of-sample error and 0.91 out-of-sample R2 score (Appendix Figure 2). This implies that reportedly poor mental health, at the census tract level, is quite predictable given certain information. Physical health, income, foreign born, and education variables have an outsized impact in both models and provide useful validation of existing literature relating socioeconomic factors to health outcomes.

The presence of certain social infrastructure, or lack thereof in the case of vacant lots, also explained more of the variance in mental health than random noise. Notably, non-commercial agriculture (*nc_agriculture*) had a significant role in the models (inline Figure 1, below), but in the opposite direction than expected. Along with churches and other outdoor spaces, they were associated with negative impacts (positive correlation with poor mental health), while cultural centers, schools, and vacant lots had a positive impact (negative correlation). Importantly, the random forest model has a slightly higher accuracy, but does not associate features with a sign relative to the outcome, because the relationship is non-linear. This may imply that the linear model does not adequately account for non-linear structure in social infrastructure.

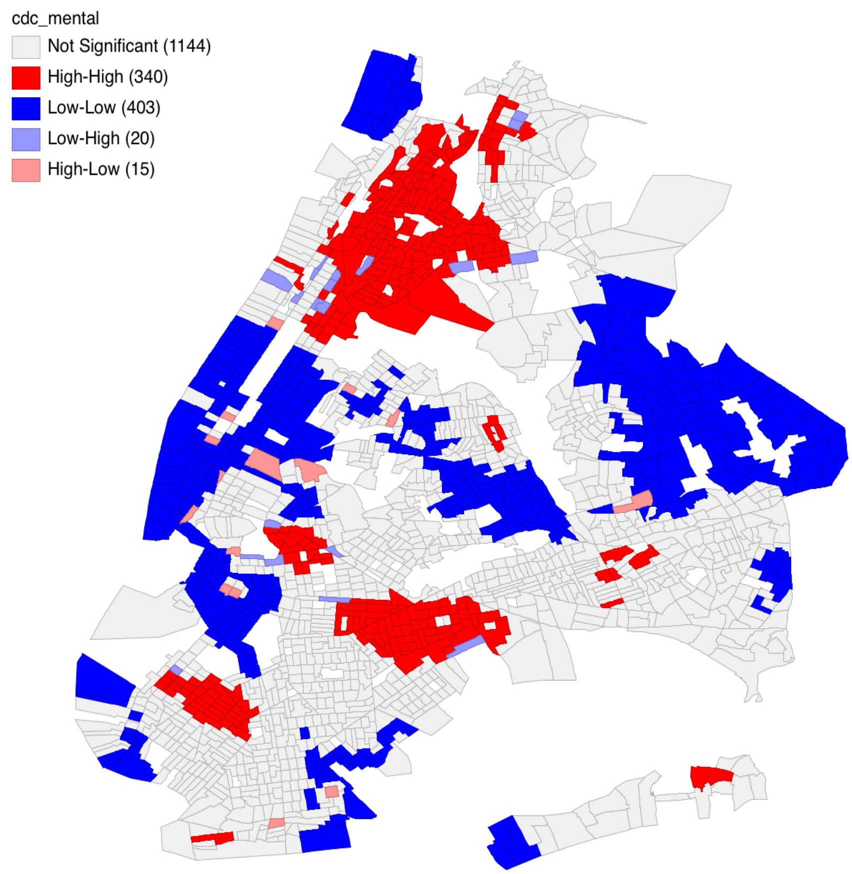
Figure 1: Regression Feature Importances



B. Spatial Analysis

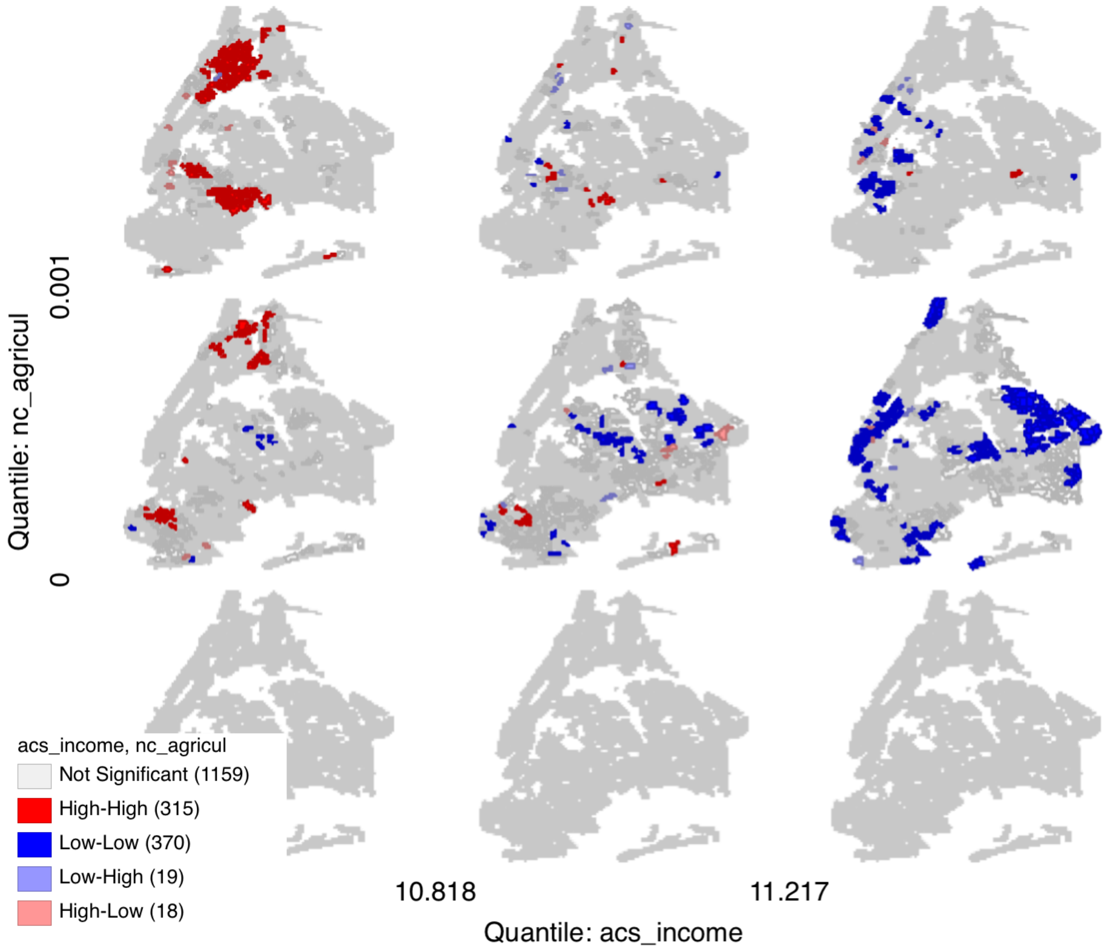
Moreover, the Global Moran's I of the target variable, calculated in ArcGIS, was 0.544 with a near-zero p-value, signaling a very high likelihood of positive spatial autocorrelation (Appendix Figure 3). Calculation of Local Moran's I values, performed using spatial analysis software GeoDa, also identified significant clustering (inline Figure 2, below). This spatial dependence may explain some of the counterintuitive results produced by the standard regression models, which treat census tracts as individual observations untethered to geographic location and relative distance.

Figure 2: Mental Health Local Moran's I Plot



Conditional Local Moran's I cluster maps reveal different clustering patterns when controlling for high and low levels of certain variables. Among the lowest income tracts in our analysis, a high number of farms and gardens correlates with clusters of worse mental health outcomes, despite correlating with clusters of better mental health among high income tracts (top left and right small multiples in inline Figure 3, below). This same pattern appears for education and physical health and for all PLUTO counts (Appendix Figures 4a-f), implying that non-commercial agriculture may have a compounding effect on other explanatory variables of mental health.

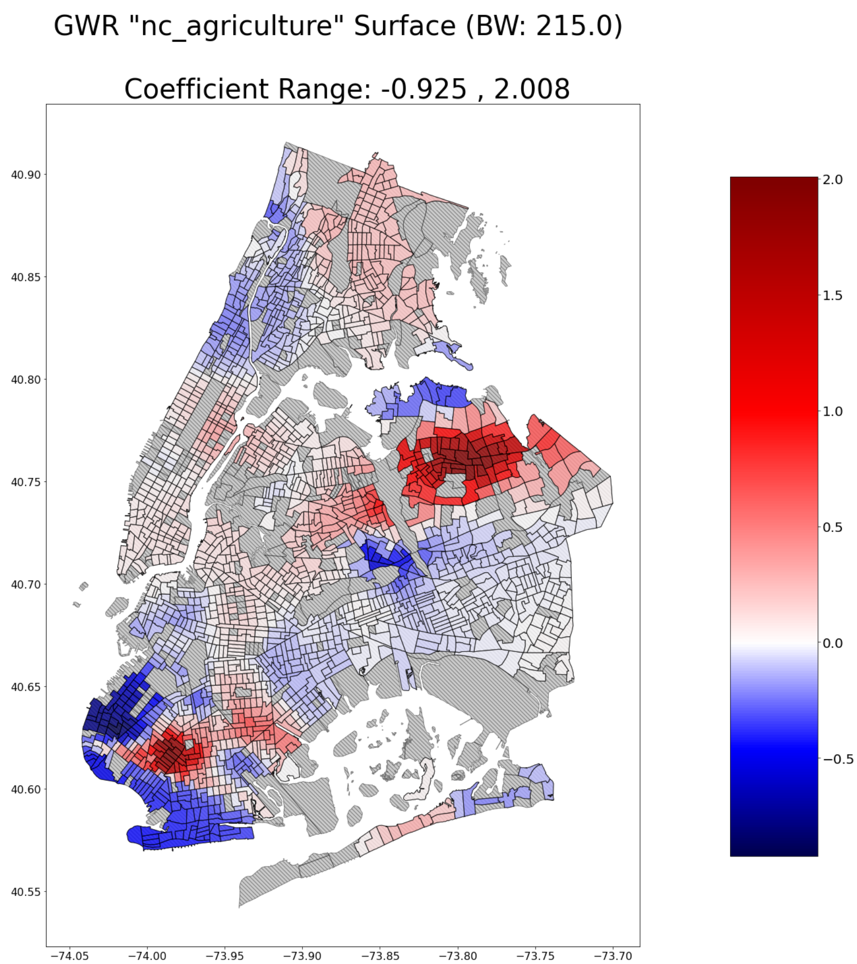
Figure 3 - Mental Health Local Moran's I Plot
Conditional on income and non-commercial agriculture



C. Geographic Weighted Regression

Results from the geographic weighted regression confirmed the spatial non-linearity of our variables (Appendix Figure 5). While a standard OLS regression yielded an R2 value of 0.90, the GWR model increased the R2 value to 0.96 and reduced the residual sum of squares (RSS) by more than half. Outside of the three most significant features, nearly all variables, including community gardens (inline Figure 4, below) and each of the PLUTO counts (Appendix Figure 6) had both positive and negative relationships with mental health depending on geographic location.

Figure 4: Spatial Distribution of GWR Coefficients for Non-commercial agriculture

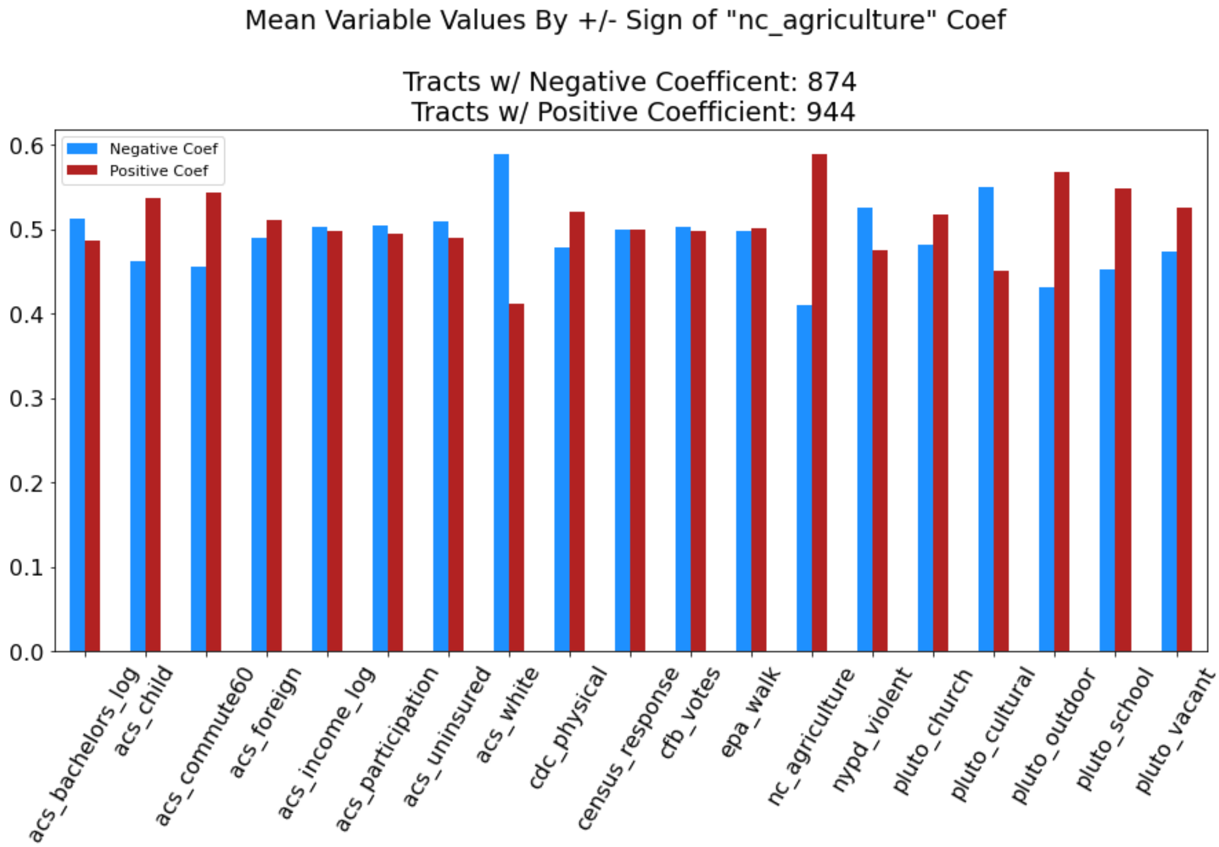


VI. Discussion

The impact of social infrastructure variables (including non-commercial agriculture) on mental health outcomes are significant across nearly all regression models. Moreover, these features are strongly spatially clustered, exhibiting positive correlations with mental health in some areas while demonstrating negative correlations in others. Despite the strong spatial autocorrelation, prevailing assumptions do not necessarily explain the spatial behavior exhibited by the variables.

To glean further insight, we separated census tracts by the sign of their coefficient values for each SI feature, visualizing the comparative mean values for each independent variable in the dataset (inline Figure 5, below). Tracts where farms and gardens strongly correlated with better mental health (negative coefficient) had higher income, higher education attainment, better physical health, fewer vacant lots, and a much higher percentage of white residents than those with a positive coefficient. While far from conclusive, this seems to further suggest that farms and gardens have compounding effects, improving mental health in advantaged areas while deteriorating it in more disadvantaged ones.

Figure 5: Mean Variable Values by Non-commercial Agriculture Coefficient Sign



Another possible explanation for our results is that gardens and certain types of social infrastructure tend to cluster around populations that are more vulnerable from a health perspective. In other words, the causal impact is flipped. In some instances, new community gardens are built as a social investment in response to adverse economic conditions or physical urban decay. In this case, there may be some time lag before the effects on individual and collective well-being are reversed. Unfortunately, at present, we do not have a large enough set of farm/garden founding dates to test this hypothesis.

While there is ample theoretical precedent linking community gardens to improvement in social cohesion and subjective well-being, these outcomes are difficult to measure and require complex methodologies that differ across studies. Some prior academic literature has claimed a positive impact on mental health specifically both from proximate greenspaces (Houlden, 2019) and from green transformations of vacant lots (South, 2018), but these fall short of providing quantitative correlation between mental health wellness and community gardens specifically.

Similarly, while quantitative studies on social infrastructure have found positive impacts on SWB, their definition of wellbeing enompasses more than reported mental health status. Our methodology attempted to account for some of this nuance by incorporating socioeconomic status and built environmental features into our regressions as independent features, but the models proved difficult to interpret in the context of a real urban system. Future research should attempt to develop a metric that better reflects the perceived social wellbeing of New Yorkers while limiting the number of explanatory variables used in any model built to determine its influencing factors.

Conclusion

Our capstone project provides the city, NYU Stern's *Invest NYC SDG Initiative*, and other partnered organizations an easily updatable and interactive web tool to assist in their efforts to promote urban agriculture. Through collaborative research and statistical analysis, we have created the most comprehensive catalog of urban farms and gardens in New York City and produced novel research on the mental health impact of social infrastructure on communities at the census tract level. Results from spatial analysis and geographic weighted regression reveal strong spatial clustering among mental health outcomes, corroborating previous spatial analyses of health metrics. Our results suggest that the quantifiable impact of community farms and gardens varies greatly by geographic location and community characteristics of a given census tract. While future research is needed to better isolate the impact community farms and gardens have on subjective well being, this research provides insight into the broad-reaching impact of social agriculture and will hopefully inform future efforts to effectively and equitably harness the power of urban agriculture for a better New York City.

Bibliography

Ackerman, K. (2012). The Potential for Urban Agriculture in New York City: Growing Capacity, Food Security, & Green Infrastructure. *Urban Design Lab at the Earth Institute* [PDF]. http://urbandesignlab.columbia.edu/files/2015/04/4_urban_agriculture_nyc.pdf.

Champman, R., Foderaro, L., Hwang, L., Lee, B., Muqueeth, S., Sargent, J., Shane, B. (2021) Parks and an equitable recovery. TPL. https://www.tpl.org/parks-and-an-equitable-recovery-parkscore-report

Cohen, N. (2016). Policy Brief: New Directions for Urban Agriculture in New York City. *CUNY Urban Food Policy Institute* [PDF]. https://www.cunyurbanfoodpolicy.org/s/Policy-Brief-New-Directions-for-UA-in-NYC.pdf.

Davern, Melanie, Gunn, L., Whitzman, C., Higgs, C., Giles-Corti, B., Simons K., Villanueva, K., Mavoa, S., Roberts, R., & Badlands, H. (2017). Using Spatial Measures to Test a Conceptual Model of Social Infrastructure That Supports Health and well-being. *Cities & Health*, 1(2). DOI: 10.1080/23748834.2018.1443620

Environmental Protection Agency (2021). *Urban Agriculture.* https://www.epa.gov/agriculture/agricultural-crops#UrbanAgriculture

Food Forward NYC: A 10-Year Food Policy Plan (2021, February). *New York City Mayor's Office of Food Policy* [PDF]. https://www1.nyc.gov/assets/foodpolicy/downloads/pdf/Food-Forward-NYC.pdf.

Guitart, D., Pickering, C., & Byrne, J. (2012). Past Results and Future Directions in Urban Community Gardens Research. *Urban Forestry & Urban Greening,* 11(4). DOI:10.1016/j.ufug.2012.06.007

Ha, H. (2018). Using Geographically Weighted Regression for Social Inequality Analysis: Association Between Mentally Unhealthy Days (MUDs) and Socioeconomic Status (SES) in U.S. Counties. *International Journal of Environmental Health Research,* 29(2). https://doi.org/10.1080/09603123.2018.1521915

Houlden, V., Porto de Albuquerque, J., Weich, S., & Jarvis, S. (2019). A spatial analysis of proximate greenspace and mental wellbeing in London. *Applied Geography*, 109, 102036. https://doi.org/10.1016/j.apgeog.2019.102036

Klinenberg, E. (2018) Palaces for the People: How Social Infrastructure Can Help Fight Inequality, Polarization, and the Decline of Civic Life. Crown Books

Latham, A., & Layton, J. (2019). Social Infrastructure and the public life of cities: Studying urban sociality and Public Spaces. *Geography Compass*, 13(7). https://doi.org/10.1111/gec3.12444

NYC Well-Being Index and Changes Over Time (2019). *Center for Innovation Through Data Intelligence (CIDI)* [PDF]. https://www1.nyc.gov/assets/cidi/downloads/pdfs/2019-NYC-Well-Being-Index.pdf

Orii, Lisa, Alonso, L,. Larson, K. (2020). Methodology for Establishing Well-Being Urban

Indicators at the District Level to be Used on the CityScope Platform. *Sustainability,* 12(2).
DOI:10.3390/su12229458

Reynolds, K. (2019, June 24). Realizing a comprehensive urban agriculture plan in New York City?. *CUNY Urban Food Policy Institute*. https://www.cunyurbanfoodpolicy.org/news/2019/6/20/realizing-a-comprehensive-urban-agriculture-plan-in-new-york-city.

South, E. C., Hohl, B. C., Kondo, M. C., MacDonald, J. M., & Branas, C. C. (2018). Effect of greening vacant land on mental health of community-dwelling adults. *JAMA Network Open*, 1(3). https://doi.org/10.1001/jamanetworkopen.2018.0298

APPENDIX

Diagram 1 - Architecture of M.A.P. NYC web application

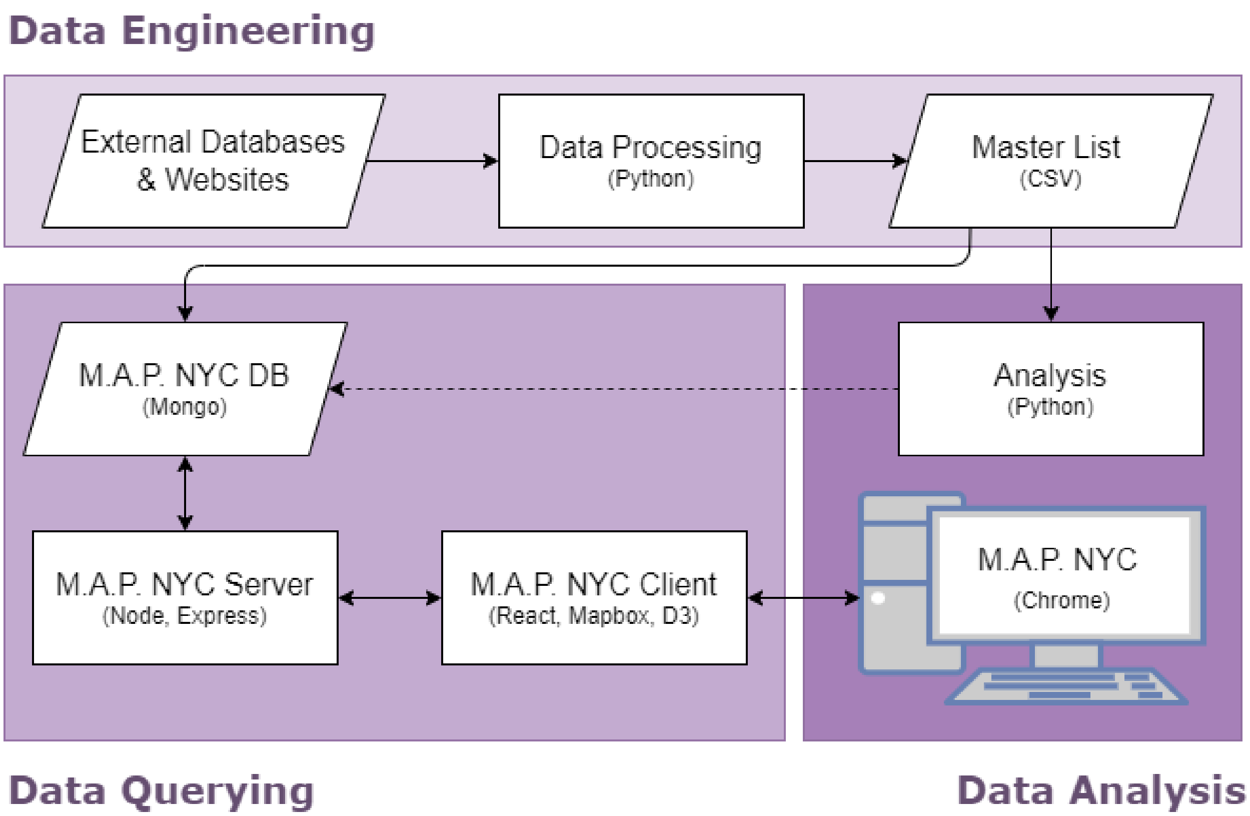


Image 1 - Screenshot of the M.A.P. NYC interface

Showing farm and garden locations colored by type and sized by tax lot area

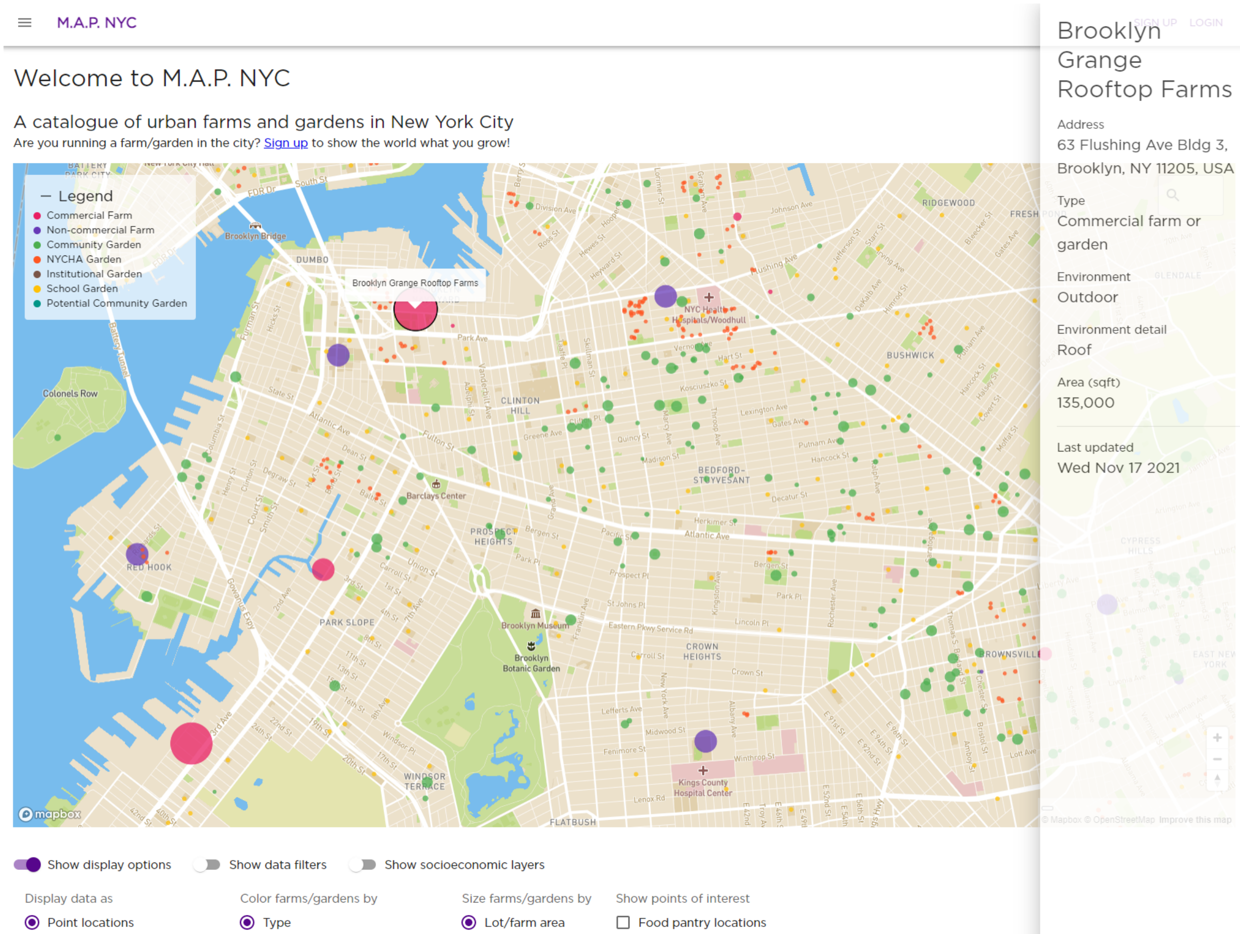


Image 2 - Screenshot of M.A.P. interface

Showing part of data collection form

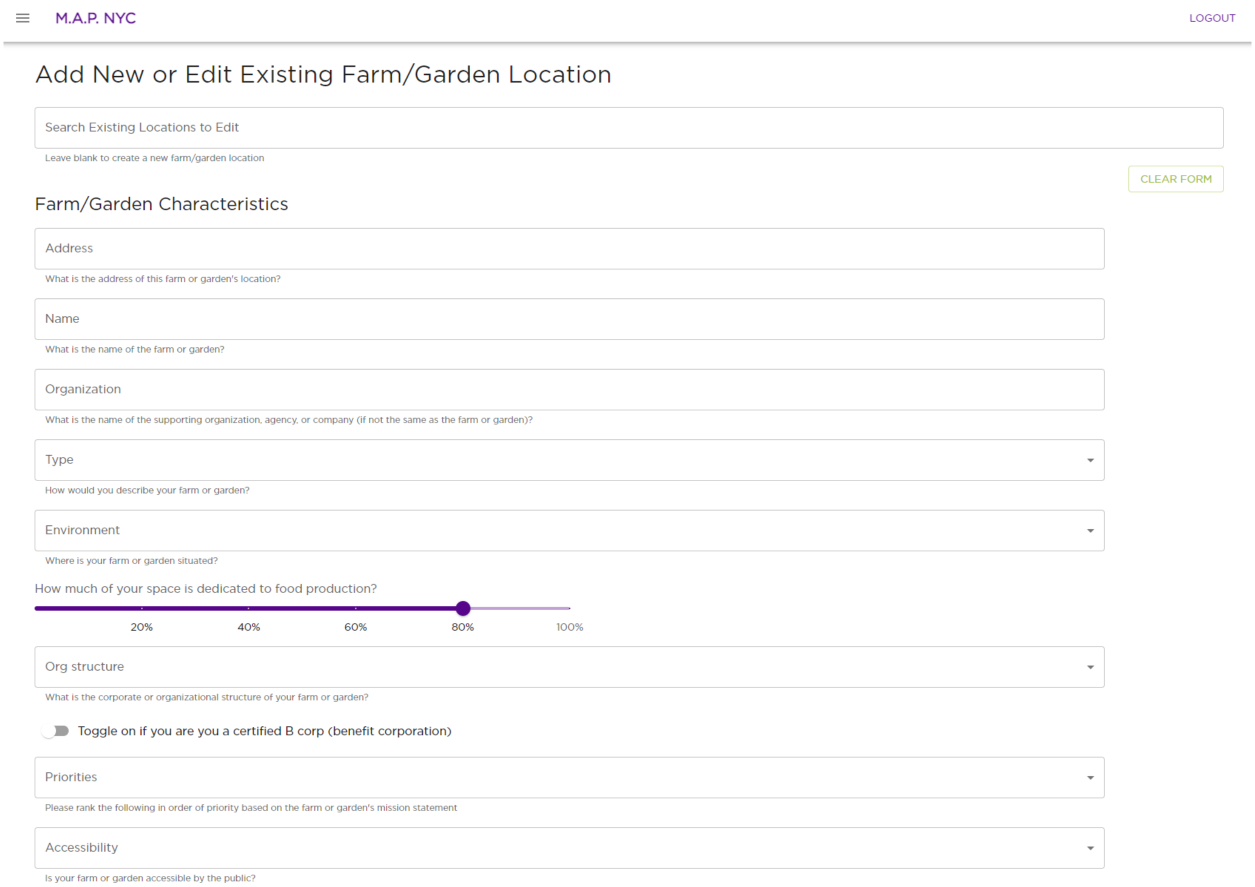


Image 3 - Screenshot of the M.A.P. NYC interface

Showing community district aggregates and socioeconomic layers

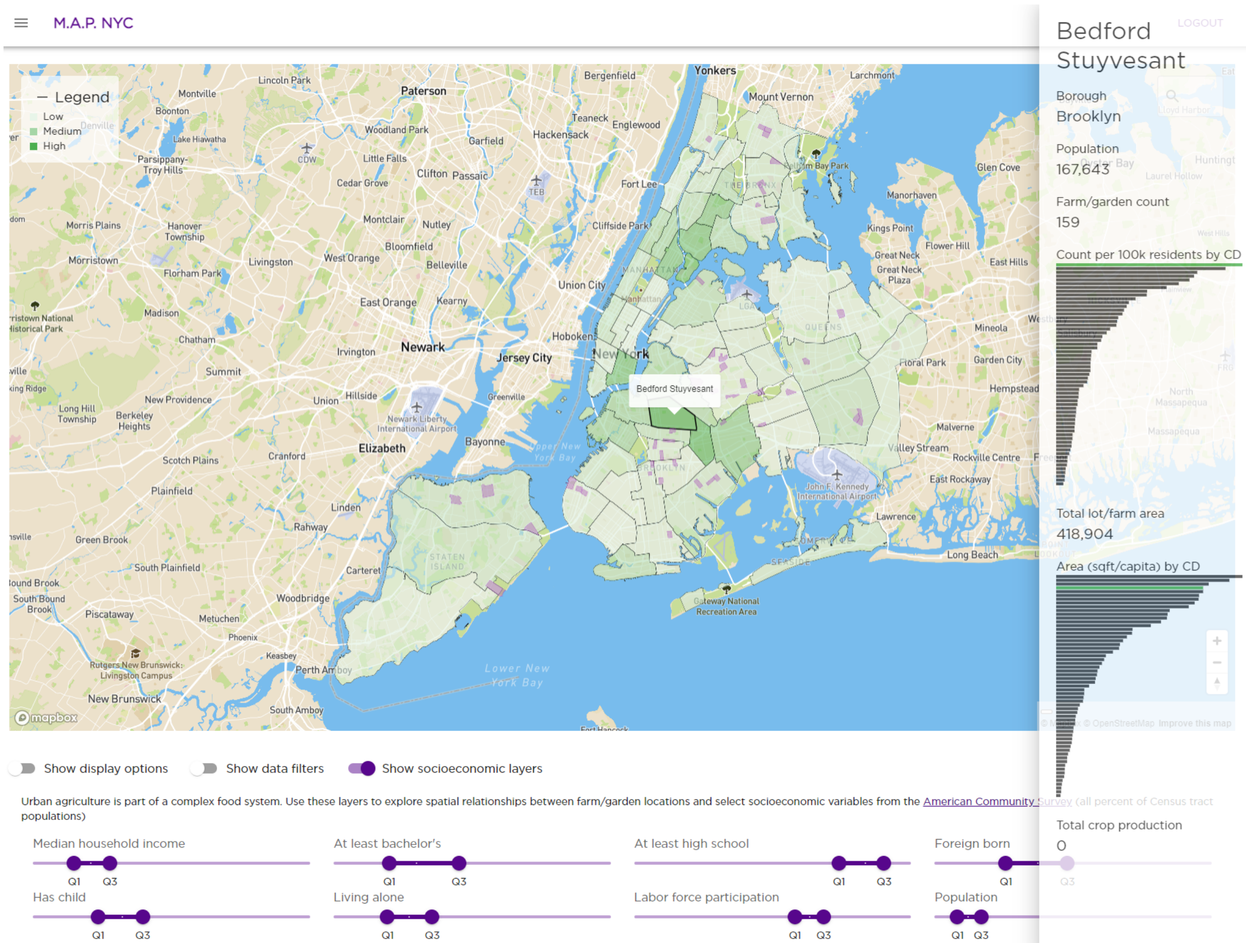


Table 1 - Data Sources

Data	Source
Urban Agriculture Data	
GreenThumb Gardens	NYC OpenData (Green Thumb)
NYCHA gardens	CSV Provided by NYC Department of Parks
School Gardens	CSV Provided by GrowNYC
Data from ~100 large farms and gardens	Survey responses with supplementary data from public websites and reports
Community garden list curated by CUSP team	GrowNYC, Brooklyn Queens Land Trust, Bronx Land Trust, Manhattan Land Trust
Potential garden locations	NYC Open Data (Local Law 46, 2018)
Modeling Data	
2018 NYC PLUTO tax lot sizes	NYC Department of City Planning
Health, demographic, and housing data	Data2Go (American Community Survey)
CDC PLACES Physical and Mental Health Data	CDC Open Data
Census Self-Response Rate	Census API
NYPD Violent Crime Data	NYC OpenData (NYPD), filtered by NYS Penal Code, aggregated to census tract
Voter Participation Score	NYC OpenData (CFB), aggregated to census tract
EPA Walkability Score	edg.epa.gov

Table 2 - Name, source and description of modeling variables

Variable Name	***Source***	***Description***
cdc_mental	CDC Places Data Portal (2020, data from 2017-18)	Percent of adults with mental health not good for ≥14 days in last 30
cdc_physical	CDC Places Data Portal (2020, data from 2017-18)	Percent of adults with physical health not good for ≥14 days in last 30
acs_income_log	American Community Survey 5-Year estimate, 2014-2018 (via Data2Go.NYC)	Natural logarithm of median income
acs_bachelors_log	American Community Survey (2014-2018)	Natural logarithm of percent of population with bachelor's degree
cfb_votes	NYC Campaign Finance Board - Voter Analysis 2008-2018 (via NYC Open Data Portal)	2018 voter participation score aggregated to census tract level
census_response	United States Census Bureau (2010)	Self-response rates mapped to census geographies
acs_foreign	American Community Survey (2014-2018)	Percent of pop foreign born
acs_child	American Community Survey (2014-2018)	Percent of households with children under 18
acs_white	American Community Survey (2014-2018)	Percent of pop that is white
acs_uninsured	American Community Survey (2014-2018)	Percent of pop with no health insurance
acs_participation	American Community Survey (2014-2018)	Labor force participation rate
acs_commute60	American Community Survey (2014-2018)	Percent of pop with commute > 60 mins
nypd_violent	NYPD Complaints 2016-2018 (via NYC OpenData), NYS Penal Law Offenses	Avg yearly complaints of felony violent crimes
epa_walk	Environmental Protection Agency - National Walkability Index	Walk score
pluto_vacant	NYC Department of City Planning PLUTO Tax Lot Data (2018)	Building classifications V0-6
nc_agriculture	Multiple sources	population-normalized community farm and garden count
pluto_church	PLUTO 2018	population-normalized PLUTO church count

pluto_school	PLUTO 2018	population-normalized PLUTO school count
pluto_outdoor	PLUTO 2018	population-normalized PLUTO outdoor count
pluto_cultural	PLUTO 2018	population-normalized PLUTO cultural count

Figure 1 - Distribution of tract level populations and EPA walkability scores

*(Lower whisker = lower quartile - 1.5 * interquartile range)*

The distribution of tract populations was skewed right so we applied a cube root transform to better approximate a normal distribution and removed tracts below the lower quartile minus the interquartile range. This is slightly different from the more common 1.5 * IQR outlier threshold, but given the subjective nature of these considerations, we believe it is reasonable. We applied the same logic for walkability but without the transformation, because the distribution is symmetric. Outliers on the high end were not removed because they do not limit the availability or accessibility of social infrastructure.

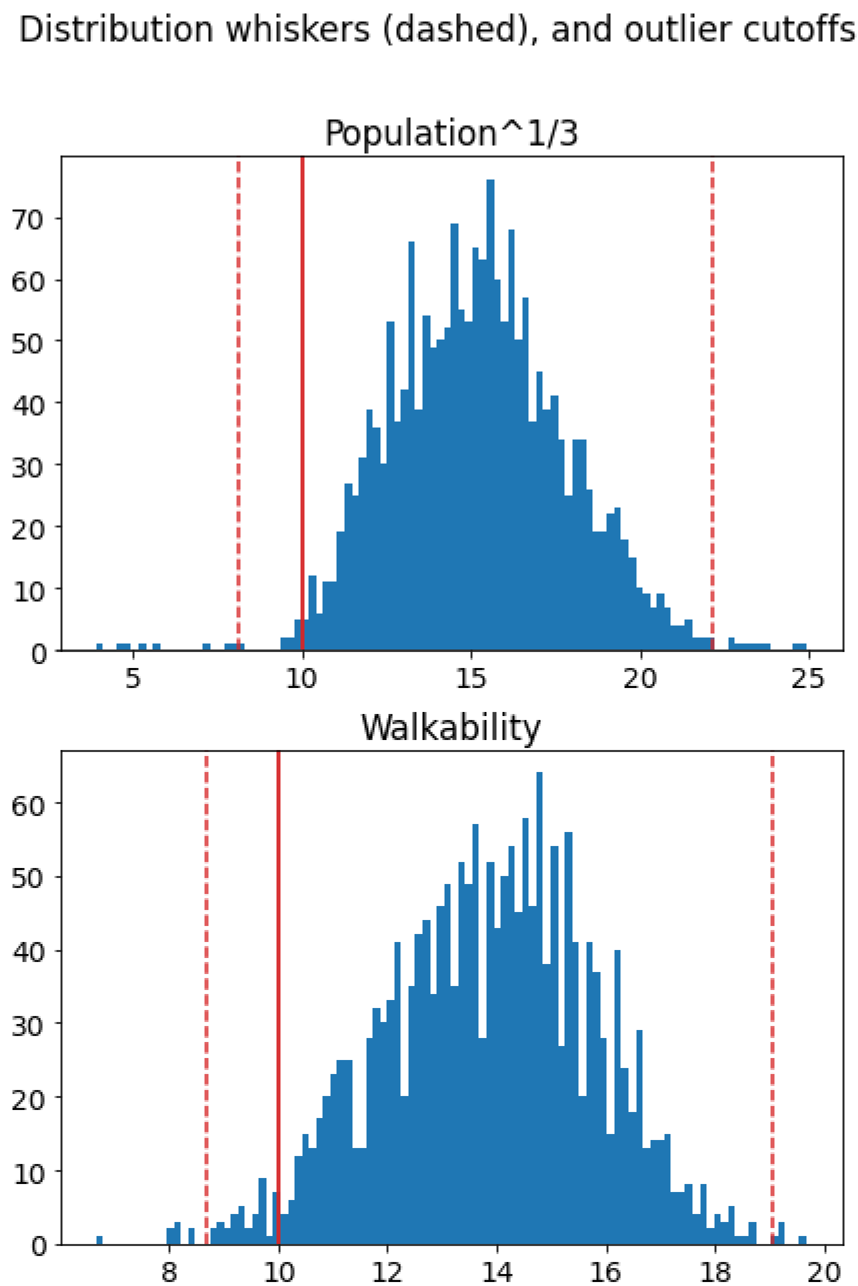


Table 2 - Description and Optimal Parameter Selection of Regression Models

Lasso regression is a regularized linear model that mitigates multicollinearity—common in socioeconomic data—by shrinking variable coefficients, thus preventing overfitting. Random Forest is a non-linear model that also performs well with high dimensionality. GWR is a weight-based regression model that uses geospatial relationships to estimate coefficients based on the spatial proximity of each observation, allowing for accurate assessment of geographic datasets.

Regression	***Optimal Parameter Value***
Lasso (L1 Regularization)	Alpha = 0.00583
Random Forest	Max_depth = 20, Max_features = 0.5 N_estimators = 500
Geographic Weighted Regression	Bandwidth (w/ Adaptive Bisquare Spatial Kernel) = 215

Figure 2 - Lasso Regression Out-of Sample Error

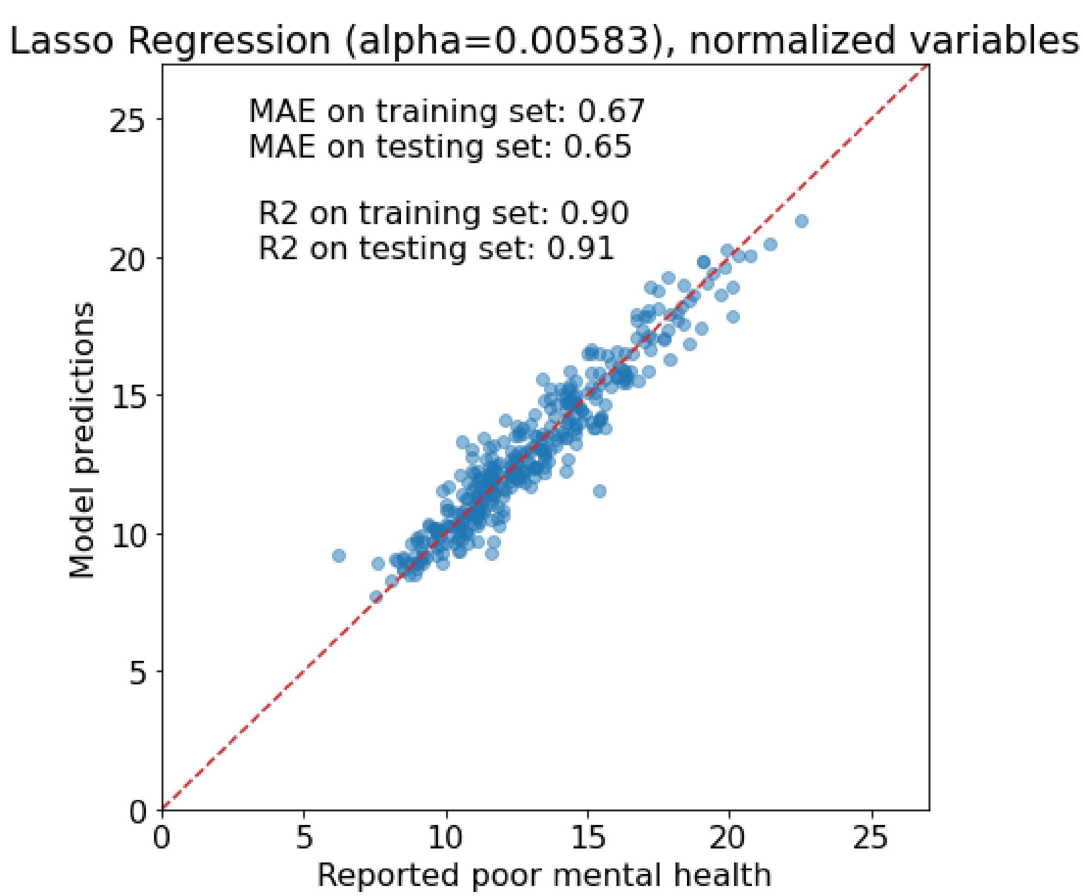
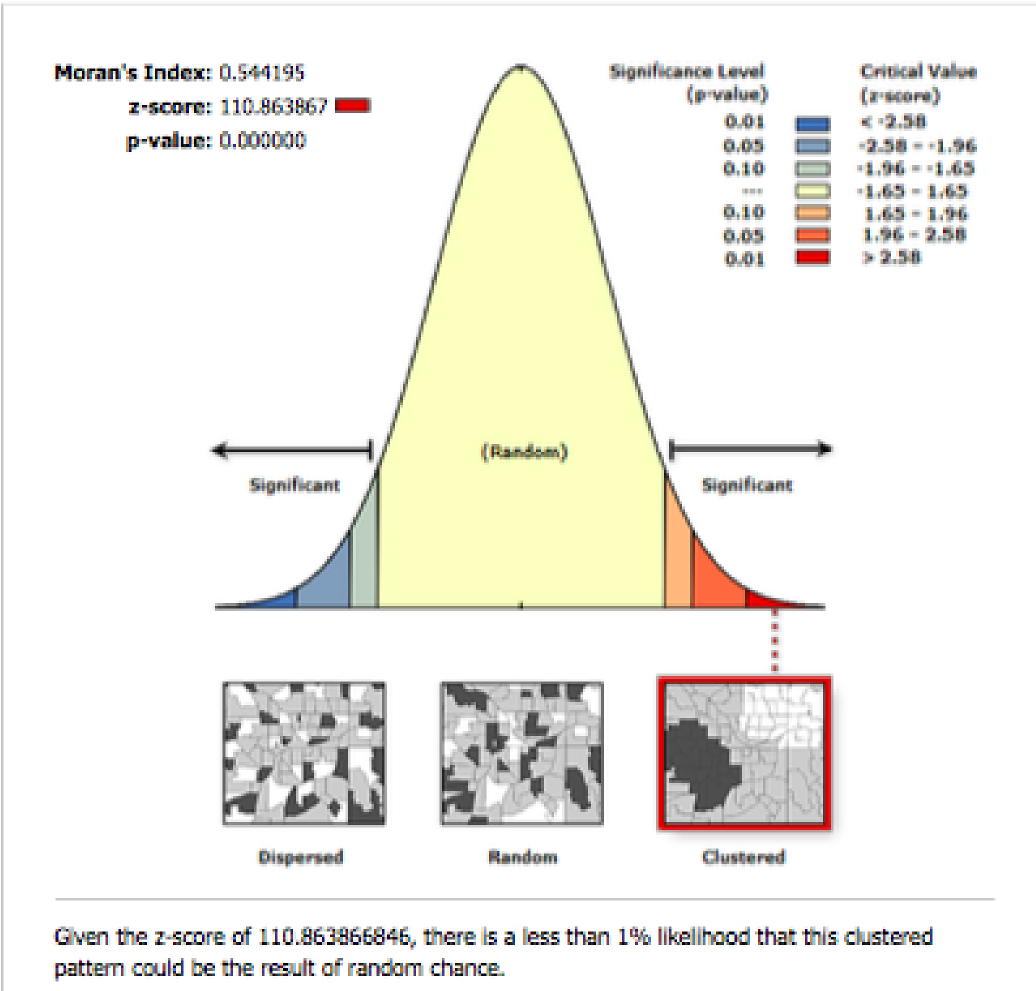


Figure 3 - ArcGIS Global Moran's I Report

Spatial Autocorrelation Report



Global Moran's I Summary

Moran's Index:	0.544195
Expected Index:	-0.000520
Variance:	0.000024
z-score:	110.863867
p-value:	0.000000

Table 3 - Subjective Well-Being indicators from NYC Well-Being Index (p. 15)

DOMAINS AND INDICATORS

The seven indicators included in this report were selected to cover a wide array of distinct factors related to well-being. To ensure this distinctness, the report selected indicators with minimal topical overlap by carrying out a correlation analysis using a large number of possible indicators. In cases where multiple variables provided essentially the same information, only one variable was chosen. Based on this research and analysis, the final domain and indicator list includes below.

DOMAINS	INDICATORS
1 ECONOMIC SECURITY	1. Household Income 2. Household Poverty 3. Unemployment Rate
2 HEALTH AND WELL-BEING	1. Current Asthma 2. Did Not Get Needed Medical Care 3. Health Insurance Coverage 4. Late or No Prenatal Care 5. Poor Health (Composite) 6. Poor Mental Health (Composite) 7. Preterm Births 8. Self-Reported Health Status
3 EDUCATION	1. Bachelor's Degree and Above 2. Chronic Absenteeism 3. On-Time High School Graduation Rate 4. Preschool Enrollment 5. State Test Proficiency: ELA 6. State Test Proficiency: Math
4 HOUSING	1. Owner Cost Burden 2. Renter Cost Burden 3. Noise Complaints 4. Overcrowded Housing
5 PERSONAL AND COMMUNITY SAFETY	1. Index Crime Rate 2. Pedestrian Injuries 3. Perception of Neighborhood Safe
6 CORE INFRASTRUCTURE AND SERVICES	1. Commute Time 2. Internet Subscription 3. Pothole Complaints
7 COMMUNITY VITALITY	1. Disconnected Youth 2. General Election Voter Turnout Rate 3. Helpful Neighbor 4. Jail Incarceration

[1] *Icons from the Noun Project.*

Figure 4a - Mental health local Moran's I plot

Conditional on education and non-commercial agriculture

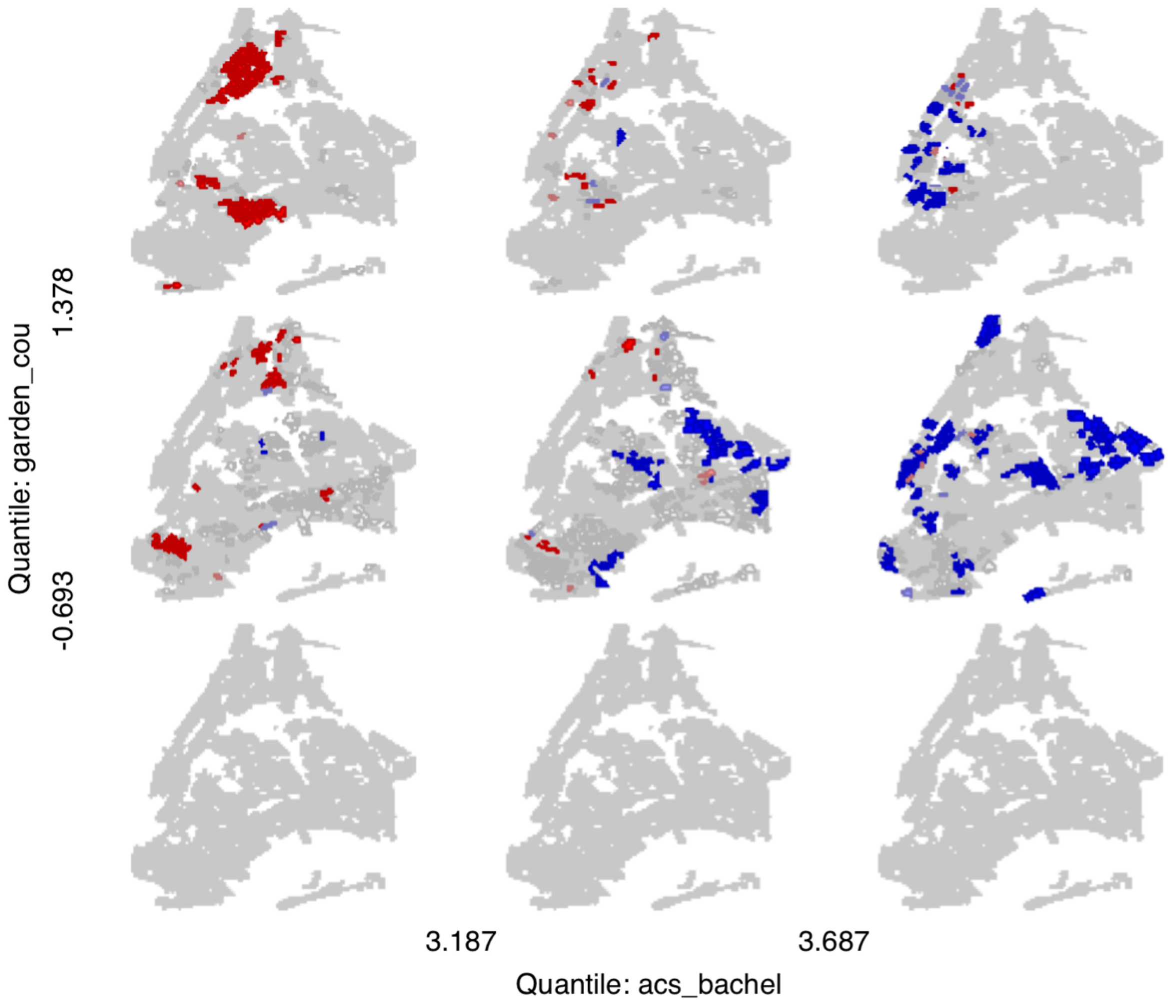


Figure 4b - Mental health local Moran's I plot

Conditional on physical health and non-commercial agriculture

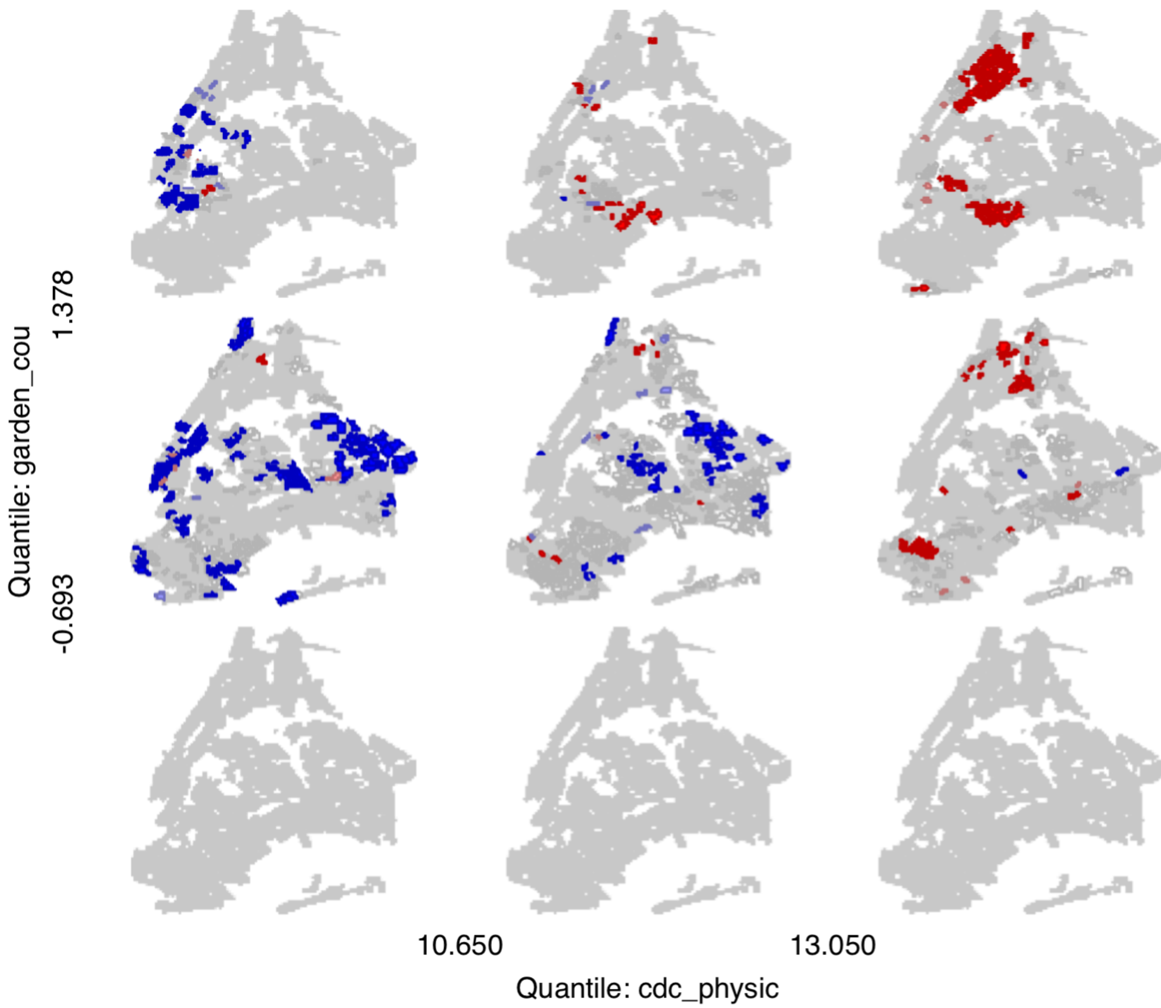


Figure 4c - Mental health local Moran's I plot

Conditional on PLUTO church classification and non-commercial agriculture

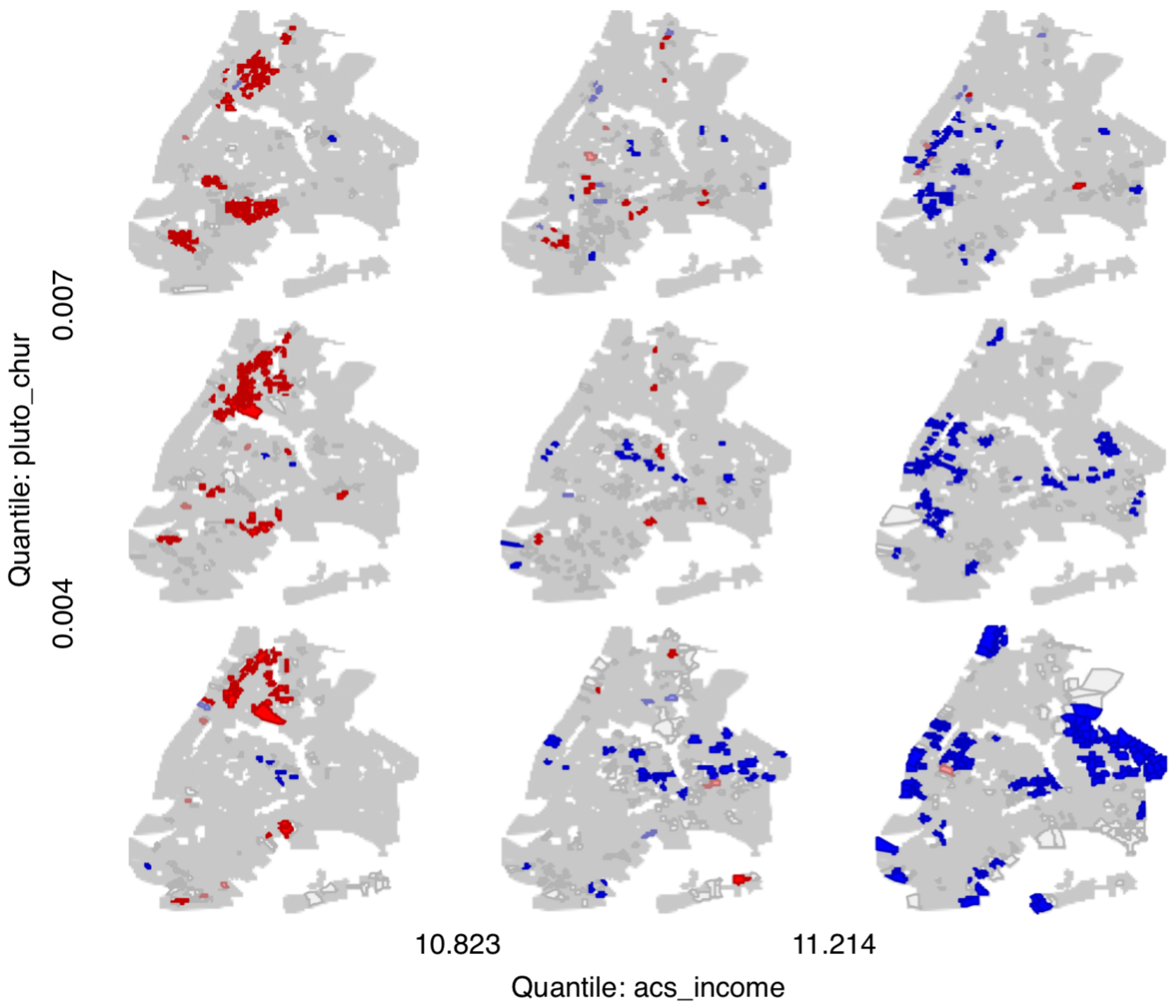


Figure 4d - Mental health local Moran's I plot

Conditional on PLUTO public school classification and non-commercial agriculture

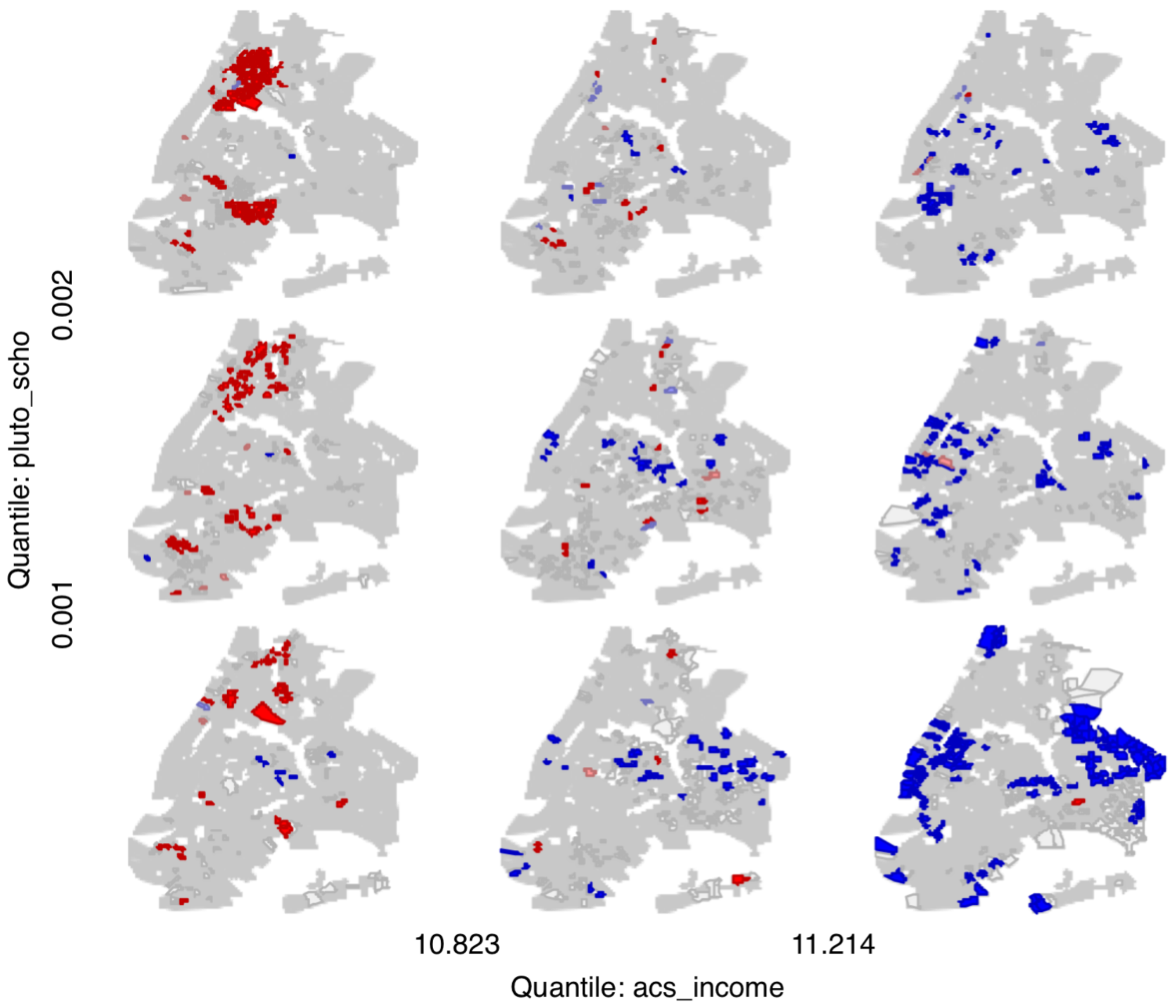


Figure 4e - Mental health local Moran's I plot

Conditional on PLUTO cultural center classification and non-commercial agriculture

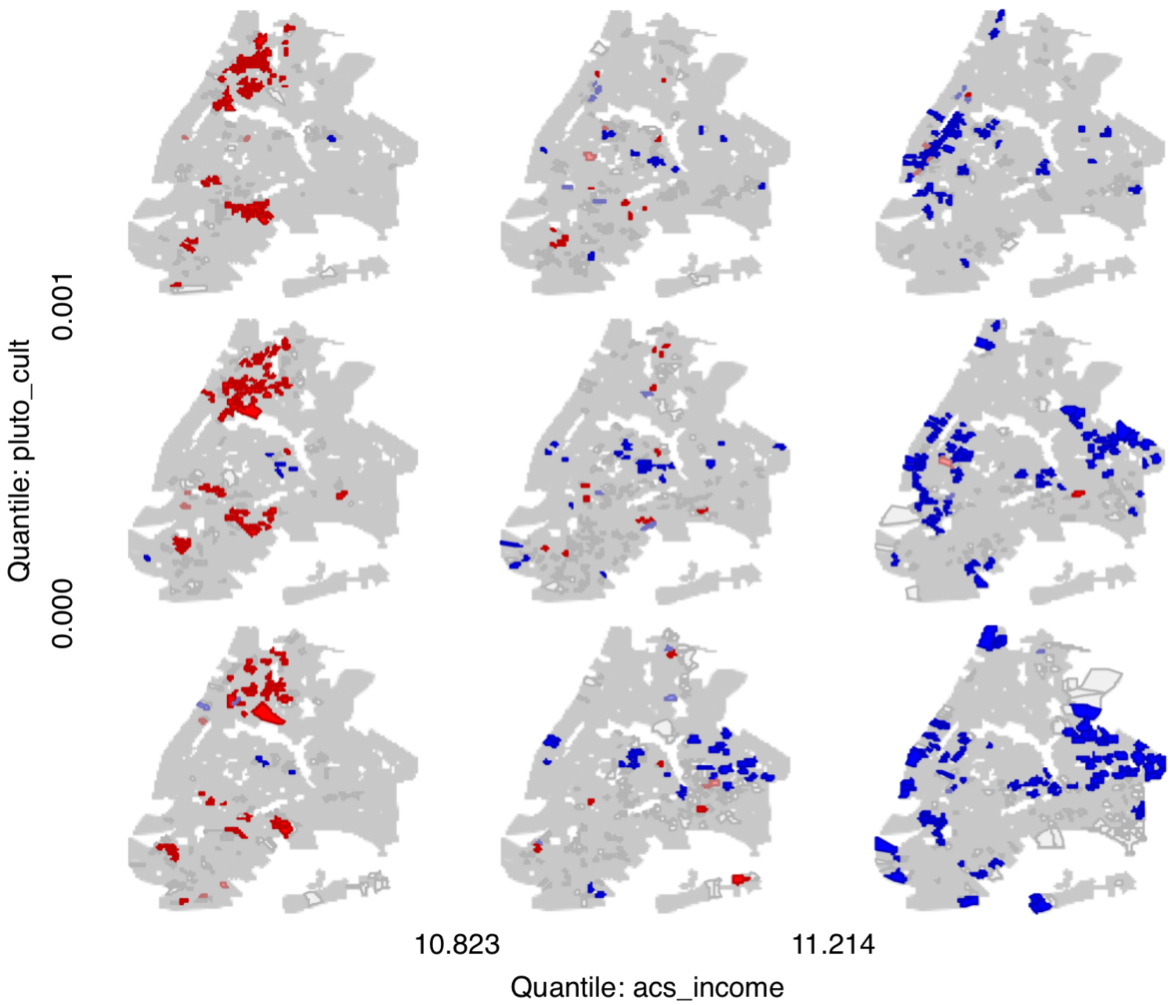


Figure 4f - Mental health local Moran's I plot

Conditional on PLUTO outdoor space classification and non-commercial agriculture

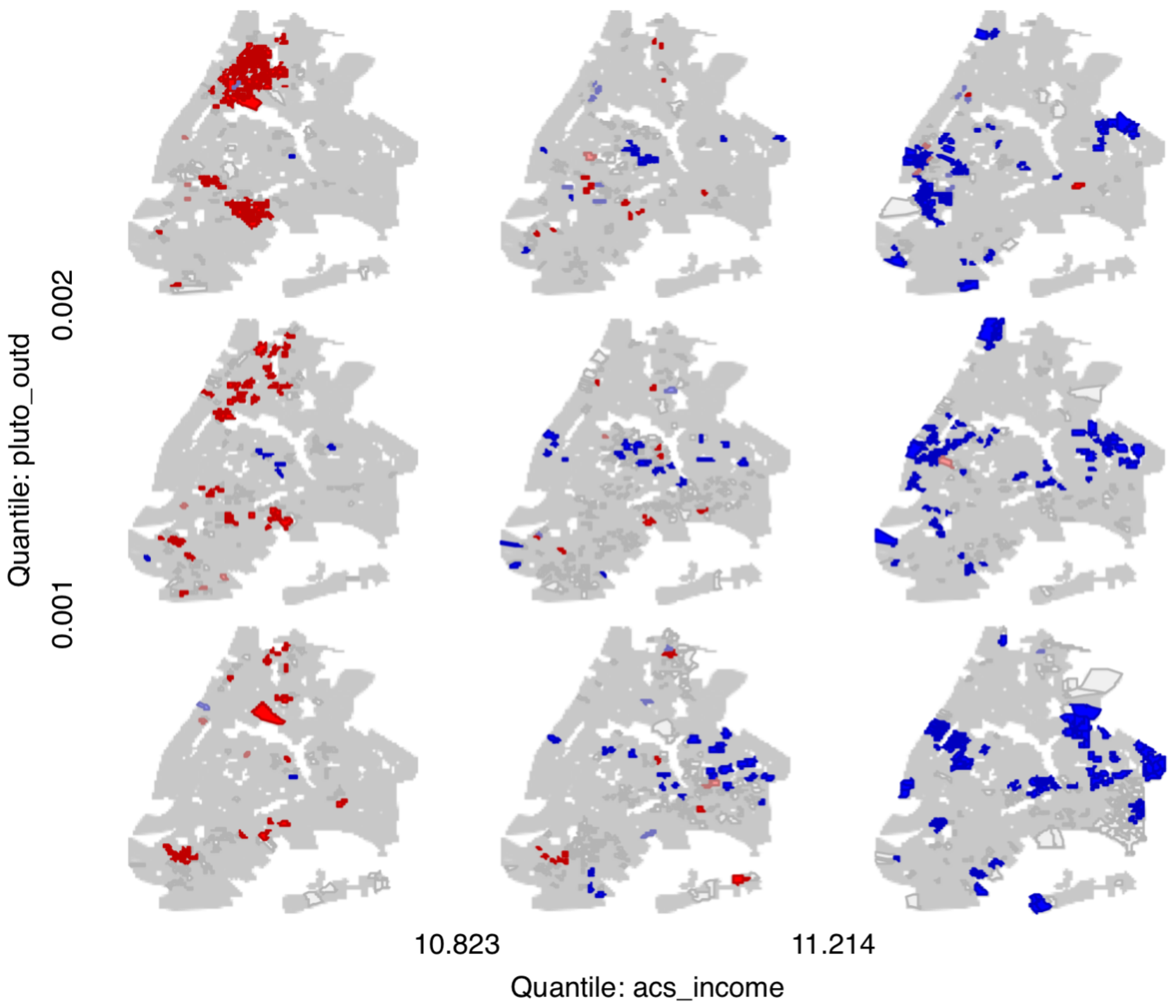


Figure 5 - Feature Importance of GWR Coefficients for Social Infrastructure

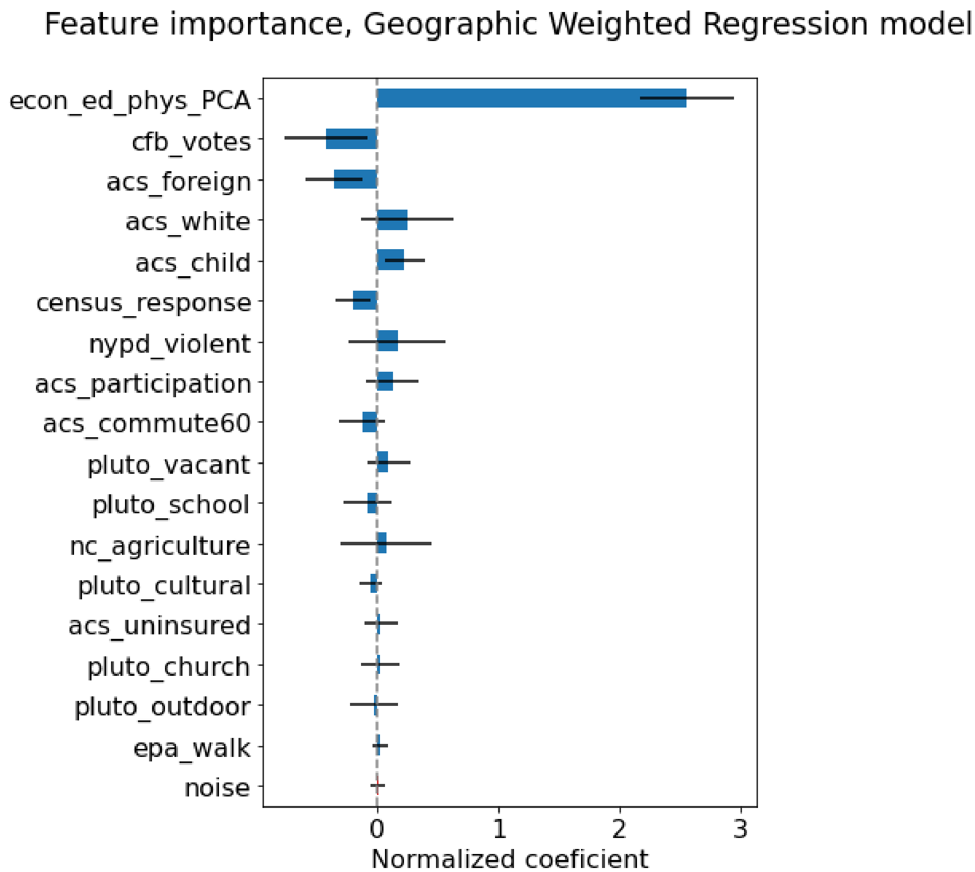


Figure 6 - Spatial Distribution of GWR Coefficients for Social Infrastructure