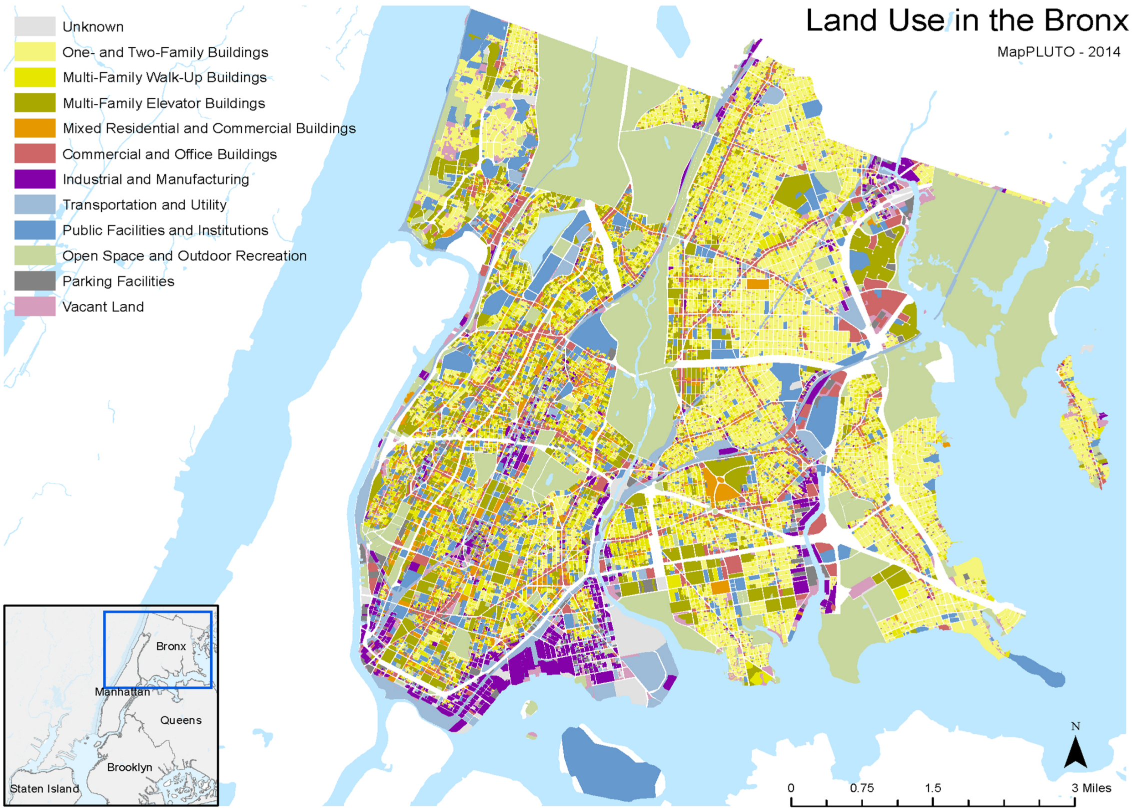
01

Measuring Segregation in Bronx by Integrating Social Media Data and Urban Spatial Network

by Xuan Liu



SENSEABLE CITY LAB, MIT

PROJECT DESCRIPTION

Abstract

Segregation has a significant impact on urban development, especially in terms of social equity. Various factors related to people's social activity and urban spatial structure could contribute to the segregation index in an urban area. This research deploys social segregation analysis and urban spatial network assessment by integrating spatial-social data, to quantify segregation in Bronx. Both home location and visit pattern based segregation are calculated by leveraging Twitter API. And two sample set with higher and lower entropy is selected for further analysis. Four indicators of spatial network are used for the experiment, including density, diversity, proximity, and land use type. An interactive interface was later applied to show the impact of potential physical interventions on social segregation. The scenario planning could be approached by manipulating live-work symmetry input to eliminate segregation. It turns out that higher entropy mobility mode with intensified, mixed-use development would achieve better social-spatial equity.

Keywords

Social segregation; social media data; urban network analysis; scenario planning; mix-use development; social-spatial equity

Introduction

In an urban environment, the neighborhood vibrancy derives from reasonable mobility crossing different places. This cross-reference enables people with daily activities, such as going to work, school, public transport, shopping, living, etc. During this mobility, people with different social status occur and encounter, which provides better social communication. To quantify and deeply understand the isolation and segregation effect between neighborhoods, this paper conduct multifaceted data analysis, by integrated social media data and urban spatial network. Based on this quantified evaluation system, future scenario planning to promote social equity is prospective.

1.1 Socio-economic Status on Segregation

Geotagged Twitter data could be served as a proxy for co-occurrence of people within the city. A metric for assessing connection across neighborhoods was proposed using this method [1]. In this study, the main result is that socio-economic similarity is crucial for connectivity. The "linkage strength" is defined by analyzing residents' median income, educational level, and immigration history, to show mobility homophily.

1.2 Connectivity Analysis Using Social Media Data

Social media check-in data indicates people's activity and preference in a city, thus could be utilized to uncover latent connectivity within neighborhoods. For instance, by identifying hidden places, such as restaurants and utilities, it is beneficial to show the local culture and make improvements to a more livable neighborhood [2]. Also, social media check-in data could help portray urban structure to enhance prediction of network accessibility [3], and apprise urban planning in high density cities to increase vibrancy [4].

1.3 Social-spatial Integrated Segregation

Segregation could also be analyzed by integrating urban physical-social space. In this research, a computational framework was implemented, coupling mobility, social-network and socio-economic status [5]. The framework could be used to depict segregation dynamics across multiple levels, from individual scale, to aggregated scale of places and cities. In the experiment in Singapore, a highly correlated result was achieved at large number of individual aggregation. Two indicators were classified. One is CSI (communication segregation index), employed to quantify the segregation level of individuals in social space, and the other one is PSI (physical segregation index), accepted to measure individual segregation in urban space.

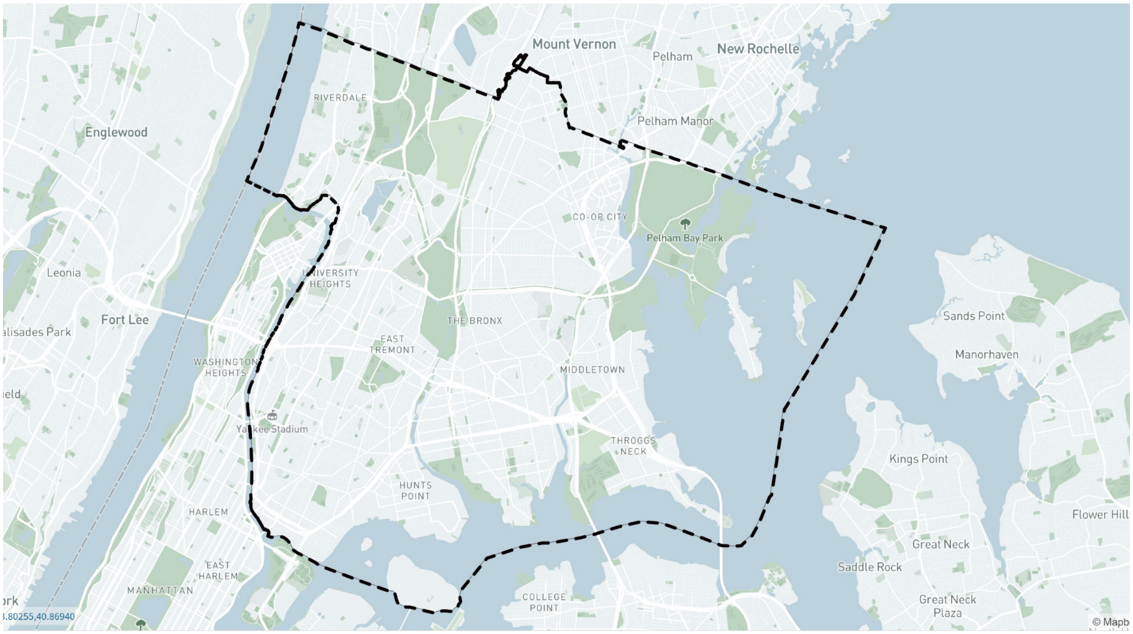


Figure 1. Study Area in Bronx

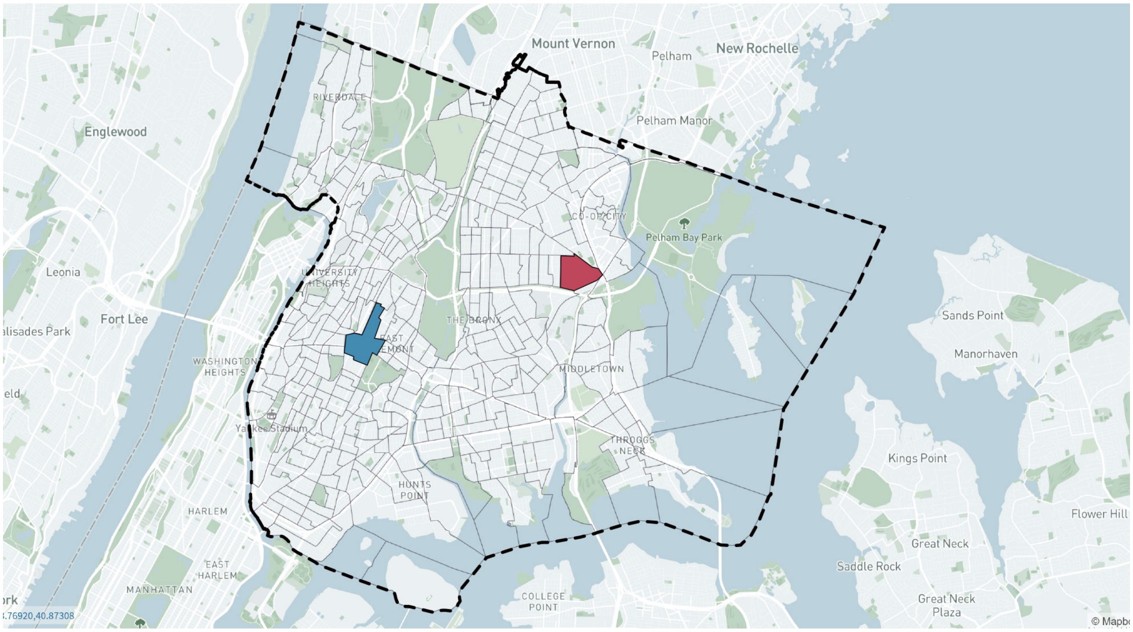


Figure 2. Two Selected Area for Comparison Study

a -1. Area A Basemap

a-2. Area A Sa

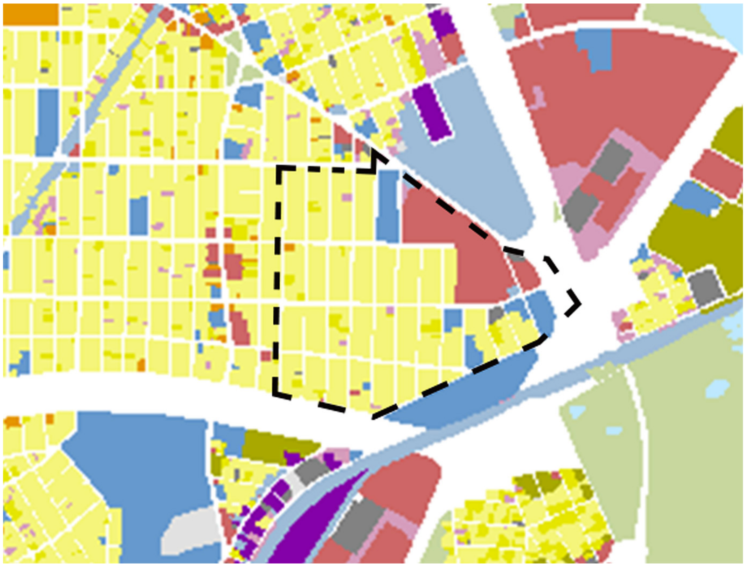
b -1. Area B Basema

b-2. Area B Sa

atellite Image

a-3. Area A Land Use Type

atellite Image



b-3. Area B Land Use Type

Figure 3. Geographic Information
of Two Selected Area

DATA AND METHODOLOGY

2.1 Study Area

The study area is Bronx county, north borough of NYC. The area of Bronx is mainly with residential function, while the riverside area condenses industrial and manufacturing functions. For the purpose of data aggregation, division of Bronx area was further selected by census tract and census block group with higher granularity.

2.2 Data Source

For social related variable, Census tract geographic information is from nyc planning, census population and average income data is from Social Explorer. Tweet username, user id, time stamp, and geo-location is scraped by Twitter API. For physical related variable, block group shapefile data is from census.gov, LODES data is from LEHD (residence area and workplace area characteristics). OSM and MapPLUTO are also used for road network and land use data.

2.3 Social Segregation Analysis

Before segregation analysis, socio-economic data, including population and average income, is merged by each census tract. And Twitter API is leveraged by a large set of IDs in Bronx.

For home location based segregation, first the home location and income for each user is estimated. DBScan algorithm is adopted for identifying clusters of night-time activity, and then this method finds the census tract that contains this home cluster, and attach the user to its average income. Second, the diversity of users who visit each census tract is calculated. Here, entropy is defined by measuring randomness of distribution, where higher entropy indicates higher randomness across income diversity. For visit pattern based segregation, neighborhood-level, mobility based connection is measured.

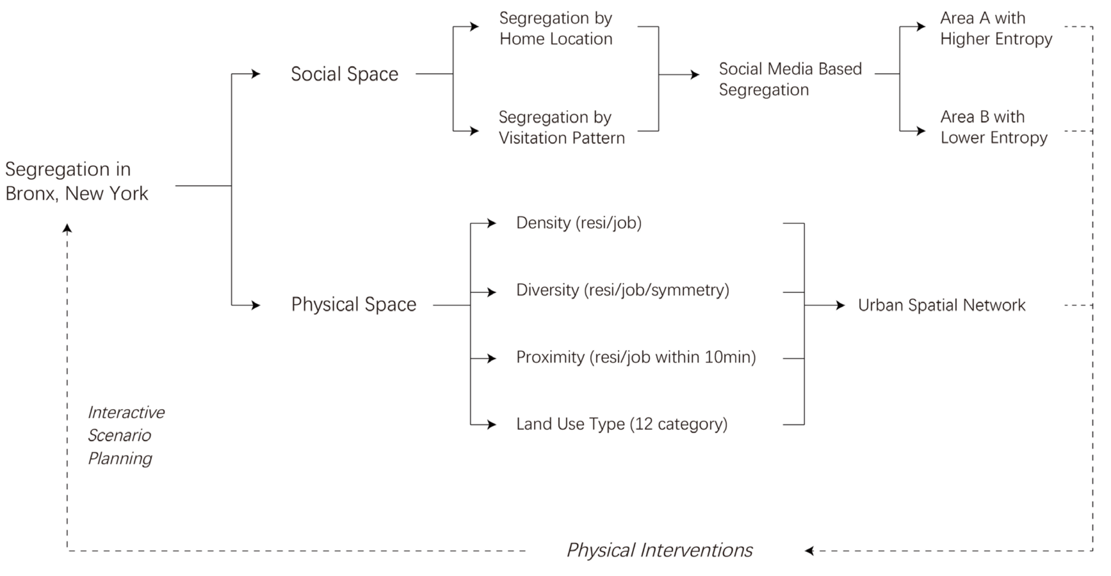


Figure 4. Research Methodology Framework

2.4 Urban Network Analysis

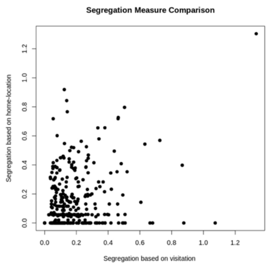
There are four physical-related variables in terms of urban network analysis: 1) Density, which calculates the number of employees and residents in each area; 2) Diversity, which constructs Shannon equitability score for measuring variation of jobs based on NAICS code, residential income, as well as live-work symmetry; 3) Proximity, which conducts accessibility analysis for jobs/residence within 1000m (10min walk); 4) Land Use Type, which represents urban development functions in each area.

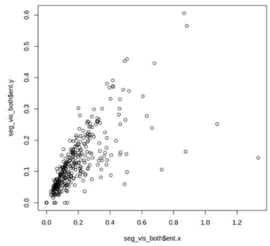
2.5 Interactive Scenario Planning

For lower entropy area, an interactive scenario planning platform is built to alleviate segregation. First, unfolded.ai is used for mapping the selected area. Second, residence area and workplace area characteristics are joined within each census block group. Third, density and live-work ratio is calculated for baseline model. Fourth, proximity analysis is visualized by creating new live-work POI, and manipulating its parameters for optimization such as GFA, area per person, live-work ratio, etc.

Variable Code	Variable Description	Data Source
Social-related Variable		
Home Location Based Segregation	Cross-reference home locations with census information, and identify socioeconomic diversity of users that gather in each neighborhood.	Census tract geographic information from nyc planning, census population and average income data from Social Explorer. Tweet username, user id, time stamp, and geo-location from Twitter API.
Visit Pattern Based Segregation	Measurement of neighborhood-level, mobility based connection.	
Physical-related Variable		
Density	The number of employees and residents in each area.	Block group shapefile data from census. gov LODES data from LEHD (residence area and workplace area characteristics)
Diversity	Construct Shannon equitability score, for diversity of jobs based on NAICS code, diversity of residential income, and diversity of live-work symmetry.	Industry classification by NAICS code
Proximity	Accessibility analysis for jobs/residence within 1000m (10 min walk).	Road network data from Open Street Map
Land Use Type	Indicators for urban development functions of each area.	New York City MapPLUTO

Table 1. Variable Description and Data Source





Indicator Type	Numerical Summary	Value
Density	Simply aggregate all the zones together and compute density as normal.	0.00247
Diversity	Take the average of all diversity scores (each area is given equal importance).	0.36886
Proximity	Simple Average over nodes.	5564.17
	Weight the importance of each node by the population living there.	6914.72

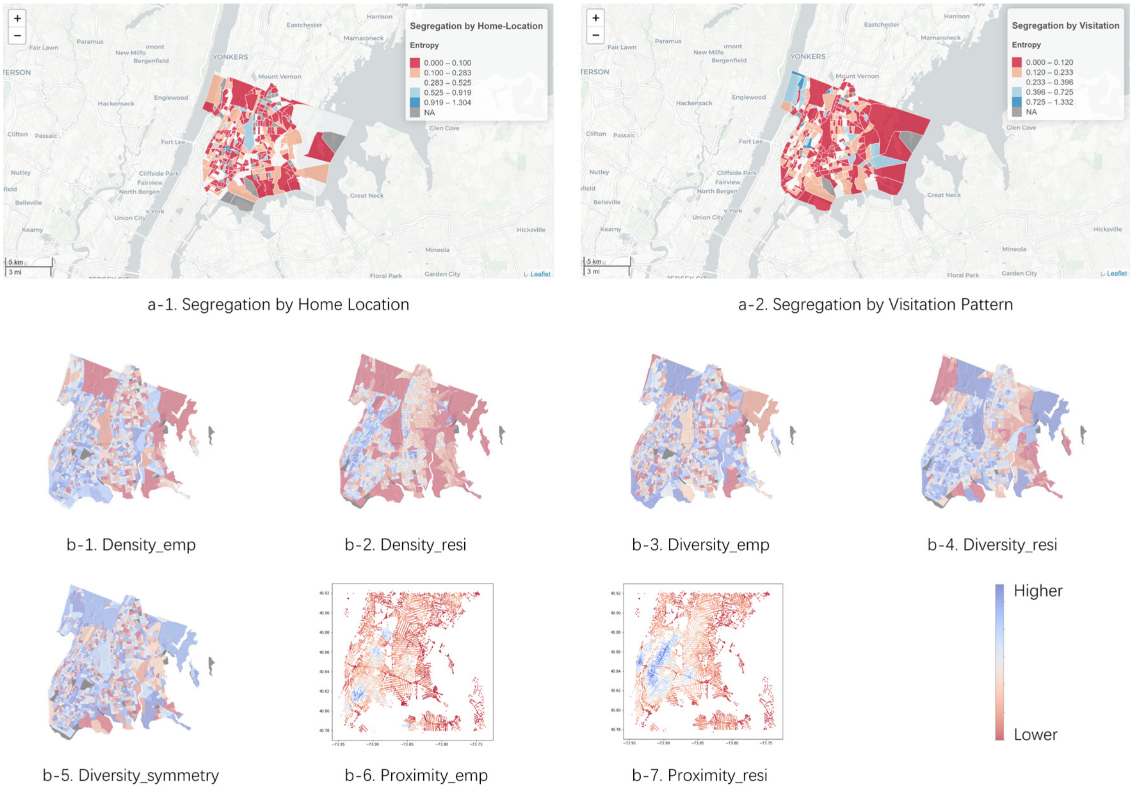
Landuse Type	Description
1	Unknown
2	One and Two Family Buildings
3	Multi-family Walk-up Buildings
4	Multi-family Elevator Buildings
5	Mixed Residential and Commercial Buildings
6	Commercial and Office Buildings
7	Industrial and Manufacturing
8	Transportation and Utility
9	Public Facilities and Institutions
10	Open Space and Outdoor Recreation
11	Parking facilities
12	Vacant Land

Table 2. Social and Physical Variable Characteristics

RESULTS AND DISCUSSION

The segregation analysis based on home location shows that most of the neighborhoods are with low entropy, with a few set of relatively higher entropy area located in west Bronx. The other segregation analysis based on visit pattern also shows that the entropy is higher in west Bronx than that in east Bronx, while the entropy in east Bronx is more unevenly distributed. For urban network analysis, the employment density and diversity are similar in west and east Bronx, while both the density and diversity of residence are higher in west Bronx. Thus west Bronx achieves a better live-work symmetry. Similarly, the proximity is significantly higher in west Bronx. Besides, the land use data indicates that the east Bronx consist of a large proportion of residential buildings, while west Bronx contains diversified function areas including residential, commercial, office, and mixed-use. In the two selected area, the one with higher entropy locates in west Bronx, showing higher quality of urban network connectivity, and more multi-function zones. For the other area with lower entropy, the interactive planning process could decrease segregation by providing better proximity and live-work symmetry. This manipulation is promising to be implemented in other lower entropy area in Bronx in future research.

While this research finds a strong correlation between social media data and urban spatial network in terms of segregation, certain methodology could also be considered for more comprehensive analysis. For instance, natural language processing (NLP) could be utilized to identify user preference for specific areas and an emotional mapping could be produced to show variant segregation [6]. Also, google street view image could be used for analyzing human perception of the environment [7], existing neighborhood patterns could be leveraged using machine learning to better understand gentrification [8], housing selling/renting price could be correlated to analyze its impact on mobility choice [9]. Furthermore, except for live-work POI such as residential/office buildings, other attributes could also be deployed for scenario planning, including commercial, amenities, open space, etc.



a-1. Segregation by Home Location

a-2. Segregation by Visitation Pattern

b-1. Density_emp

b-2. Density_resi

b-3. Diversity_emp

b-4. Diversity_resi

b-5. Diversity_symmetry

b-6. Proximity_emp

b-7. Proximity_resi

Figure 5. Social Segregation and Urban Network Analysis Result

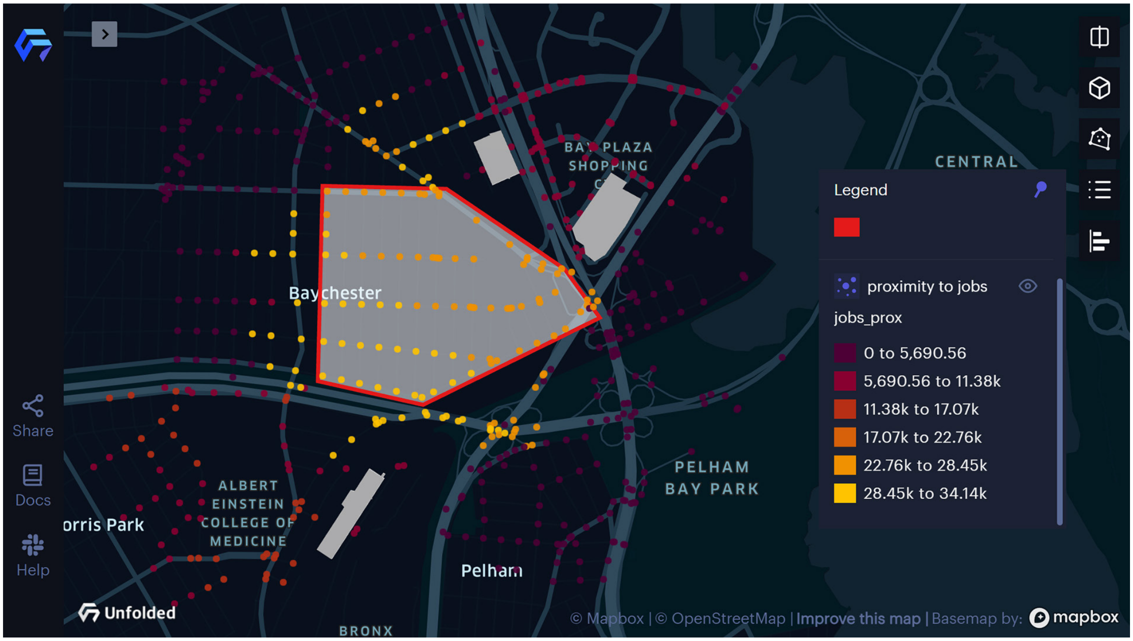


Figure 6. Interactive Scenario Planning
Live-Work POI 2D View

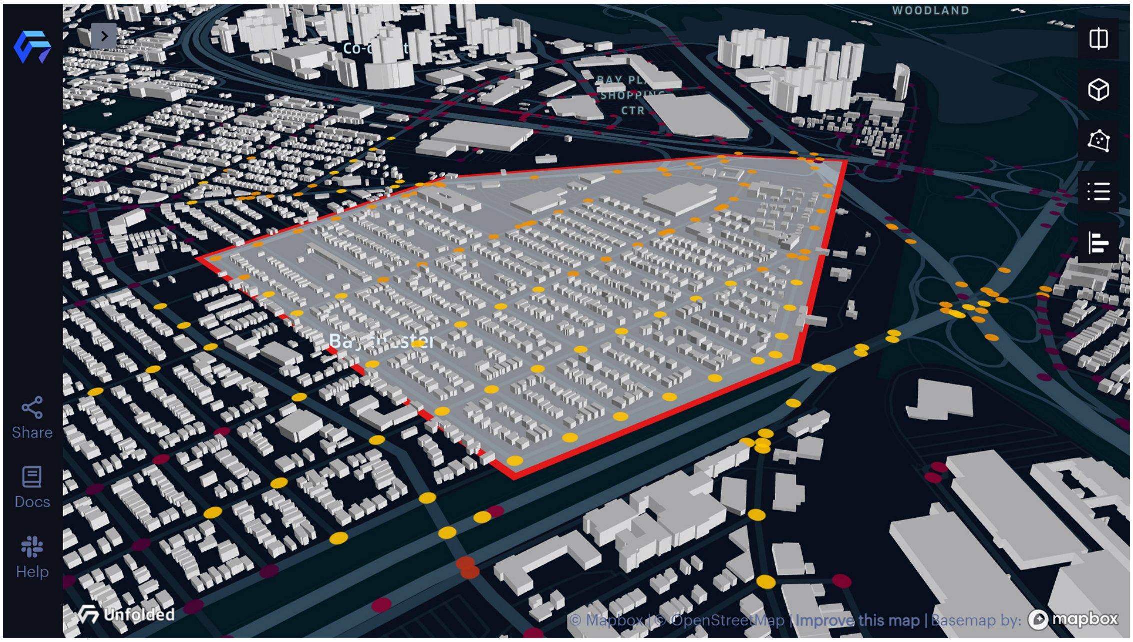


Figure 7. Interactive Scenario Planning
Live-Work POI 3D View

CONCLUSION

In this paper, a series of experiment is performed to analyze segregation in Bronx by integrating social media data and urban spatial network. Social segregation is calculated by entropy based on home location and visitation pattern, while physical segregation is measured by density, diversity, proximity and land use. The social and physical indicators show strong correlation with each other. Two typical area, with higher and lower entropy respectively, is selected for further comparison. In this study, the area in west Bronx indicates higher entropy across socio-economic groups, with better quality of urban network connectivity and diversified land use. For the areas with lower entropy of segregation, interactive scenario planning by adding new live-work POI is conducted and contributive to alleviate neighborhood segregation. Future research could adopt certain methodology to analyze segregation more comprehensively. From the social perspective, NLP could be used for identifying user preference to a specific area. From the physical perspective, GSV image for human perception, ML for urban gentrification, housing price for mobility choice could be implemented. Moreover, scenario planning model could be modified by integrating other functional attributes.

This project represents an important advancement of depicting segregation in Bronx from a social-spatial view, thus further exploration is promising and encouraging for data-driven decision-making process for environment equity in the near future.

Reference

[1] C. Heine, C. Marquez, P. Santi, M. Sundberg, M. Nordfors, and C. Ratti, "Analysis of mobility homophily in Stockholm based on social network data," PLoS ONE, vol. 16, no. 3, p. e0247996, Mar. 2021, doi: 10.1371/journal.pone.0247996.

[2] F. Zhang et al., "Uncovering inconspicuous places using social media check-ins and street view images," Computers, Environment and Urban Systems, vol. 81, p. 101478, May 2020, doi: 10.1016/j.compenvurbsys.2020.101478.

[3] Y. Shen and K. Karimi, "Urban function connectivity: Characterisation of functional urban streets with social media check-in data," Cities, vol. 55, pp. 9–21, Jun. 2016, doi: 10.1016/j.cities.2016.03.013.

[4] T. Chen, E. C. M. Hui, J. Wu, W. Lang, and X. Li, "Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data," Habitat International, vol. 89, p. 102005, Jul. 2019, doi: 10.1016/j.habitatint.2019.102005.

[5] Y. Xu, A. Belyi, P. Santi, and C. Ratti, "Quantifying segregation in an integrated urban physical-social space," J. R. Soc. Interface., vol. 16, no. 160, p. 20190536, Nov. 2019, doi: 10.1098/rsif.2019.0536.

[6] Y. Song, J. Fernandez, and T. Wang, "Understanding Perceived Site Qualities and Experiences of Urban Public Spaces: A Case Study of Social Media Reviews in Bryant Park, New York City," Sustainability, vol. 12, no. 19, p. 8036, Sep. 2020, doi: 10.3390/su12198036.

[7] Y. Yao et al., "A human-machine adversarial scoring framework for urban perception assessment using street-view images," International Journal of Geographical Information Science, vol. 33, no. 12, pp. 2363–2384, Dec. 2019, doi: 10.1080/13658816.2019.1643024.

[8] J. Reades, J. De Souza, and P. Hubbard, "Understanding urban gentrification through machine learning," Urban Studies, vol. 56, no. 5, pp. 922–942, Apr. 2019, doi: 10.1177/0042098018789054.

[9] M. Yurrita et al., "Dynamic Urban Planning: an Agent-Based Model Coupling Mobility Mode and Housing Choice. Use case Kendall Square," arXiv:2106.14572 [cs], Jun. 2021, Accessed: May 01, 2022. [Online]. Available: http://arxiv.org/abs/2106.14572