From phone booths to Wi-Fi kiosks: the spatial inequality of public connectivity in New York City

Ryan Y. Wang

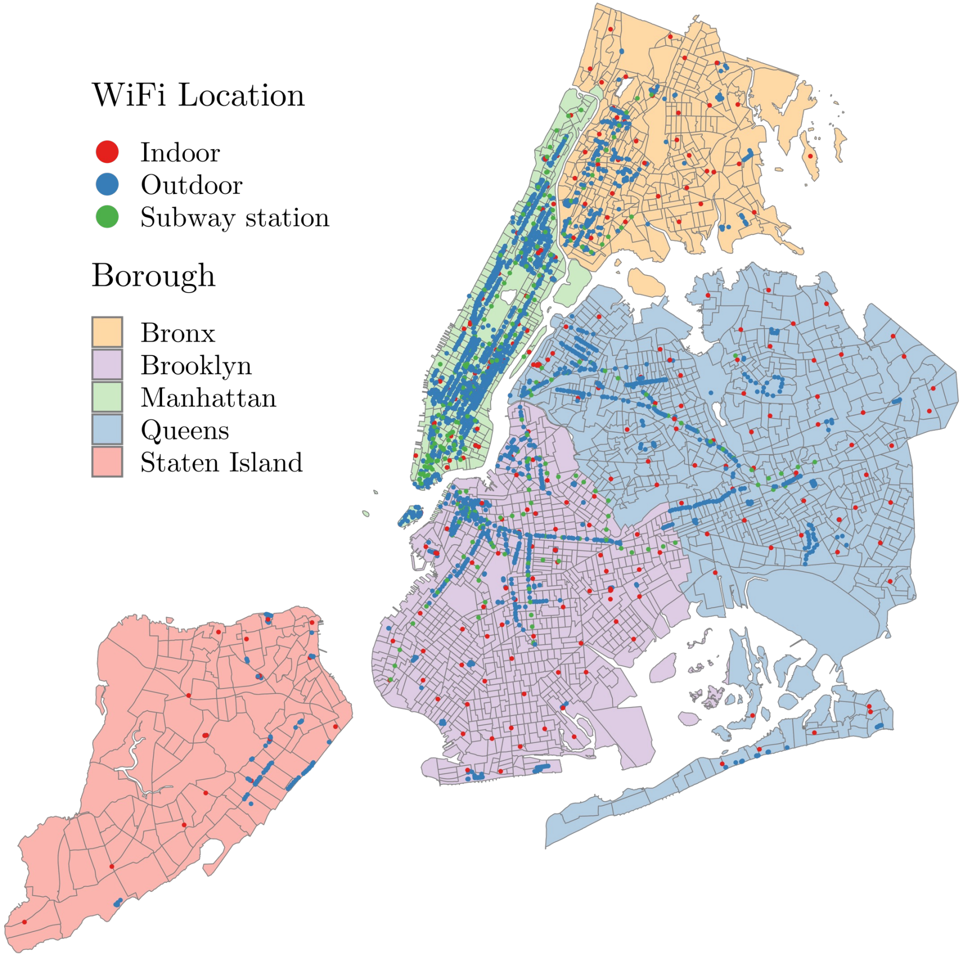
PTC'23, Honolulu, HI

PennState
Donald P. Bellisario
College of Communications

INTRODUCTION

Public Internet access in NYC

- Public Internet access as essential community infrastructure (Simpsons, 2004, McClain et al., 2021)
- Digital access inequities in NYC
    - 29% no home broadband subscription
    - 46% in poverty lack both home and mobile broadband
- The *Internet Master Plan* by de Blasio administration
    - 42% have at least one “hotspot” (free Wi-Fi access point) in a commercial area (out of 196 neighborhoods)
    - The free public Wi-Fi has been deployed at 3,319 locations by around 17 providers on streets, open space, transit and public facilities across the five boroughs
- Does it mitigate vs. reflect existing digital inequalities?



Source: The New York City Internet Master Plan

Does it work?

- RQ1: What is the spatial pattern of the public Wi-Fi hotspots in New York City?
- RQ2: What is the relationship between (a) the Asian American population; (b) the African American population; (c) Hispanic population and the Wi-Fi hotspot deployment on census tract level in NYC?
- RQ3: What is the relationship between education attainment and the Wi-Fi hotspot deployment on census tract level in NYC?
- RQ4: What is the relationship between (a) median household income; and (b) GINI index (income inequality) and the Wi-Fi hotspot deployment on census tract level in NYC?
- H1: The census tract in NYC with more commercial business with have denser Wi-Fi deployment.
- RQ5: What is the relationship between digital infrastructure and the Wi-Fi hotspot deployment on census tract level in NYC?
- RQ6: What is the relationship between political orientation and the Wi-Fi hotspot deployment on census tract level in NYC?
- RQ7: Are any relationships above sensitive to spatial structure on census tract level in NYC?

The spatiality of ICT access

- ICTs could be “situated, localized and specialized” in space (Rodriguez-Amat & Brantner, 2016), as the physical space or the place-based attributes that might affect people’s ICTs practices as well (Fast et al., 2019; Liu, 2019).
- The geography of ICT (i.e., the spatial location of Wi-Fi), provide a tangible representation of information infrastructure in the urban space (Kim, 2018).

The missing dimension

- What have been missing in the digital inequality literature is the presence of spatial effects, as the geographical unit with high or low values of ICT access or utilization could be influenced by their neighboring units with similar value of ICT practices.
- Spatially Aware Technology Utilization Model (Pick & Sarkar, 2016) incorporates the dimension of spatial inequality into the digital inequality
    - ICTs practices as the outcome variables
    - Some demographic factors might influence the utilization and availability of ICTs within the geographical unit and adjacent units as well
        - Such as average age of the population, employees in professional, scientific and technical sectors, income, GDP per capita, education, gender, race and ethnicity (Azari & Pick, 2005; Billon et al., 2017; Pick et al., 2015; Song et al., 2020)

(Main)Research questions

What is the spatial pattern of the public Wi-Fi hotspots in New York City?

This study examines the hot (and cold) spots in public Wi-Fi distribution density with the Getis-Ord Gi* statistics across the census tracts of New York City.

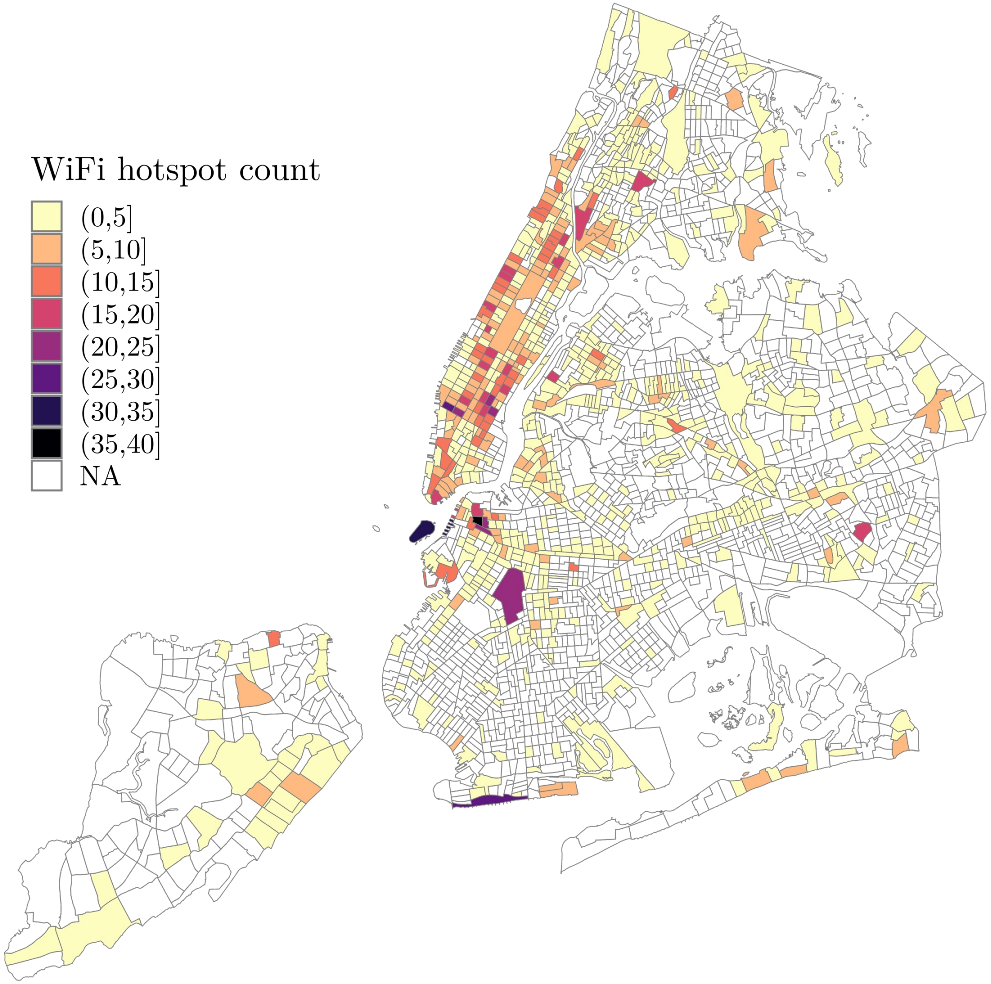
Would demographic and socio-economic factors, digital infrastructure, political affiliation drive the public Wi-Fi hotspots' spatial pattern?

By adopting the Spatially Aware Technology Utilization Model (SATUM, Pick & Sarkar, 2016), it analyzes how demographic and socio-economic factors, digital infrastructure and political affiliation might impact public Wi-Fi locations, while incorporating spatial strcture.

METHOD

Data collection

- The number of public Wi-Fi location data: NYC Open Data ($n = 3319$)
- Shapefile: Census (Unit of analysis: Census tract, $N = 2324$).
- *Demographic and socio-economic factors*
  - American Community Survey (ACS) 2016 – 2020 5-Year Estimate: Population, African American population, Asian American population, Hispanic population, population with less than high school education, population with disability, GINI index, income level
- *Commercial density*
  - NYC's open data portal: the aggregated number of the commercial business in the particular census tract
- *Digital infrastructure*
  - American Community Survey (ACS) 2016 – 2020 5-Year Estimate: Household without any devices (i.e., computer/smartphone/tablet) at home; household without any broadband subscription; household without any Internet connection at home
  - FCC 477 data: residential with fixed high-speed connections over 200 kbps in at least one direction per 1000 households
- *Political orientation*
  - ESRI's ArcGIS Business Analyst: population affiliated with Democrats and Republicans


WiFi hotspot count
(0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
NA

Analytical strategy

- Hotspot analysis: Geti-Ord Gi^* statistics
- Bayesian spatial modeling
    - Negative binomial regression for count data
    - Integrated nested Laplace approximation approach
        - Computationally much less expensive compared to MCMC
        - Focus on controlling the potential bias led by the spatial structure
    - Baseline (negative binominol) model, IID model, ICAR model and BYM model
    - Model selection: DIC, WAIC, CPO
    - Standardization on the same scale
    - Multicollinearity issues: The variables of population, and household without any devices (i.e., computer/smartphone/tablet) at home are removed

Model specifications

$$\overline{y_i} = \exp\left(\beta_0 + \sum_{j=1}^{n} \beta_j x_{ij} + \delta_i + \varepsilon_i\right),$$

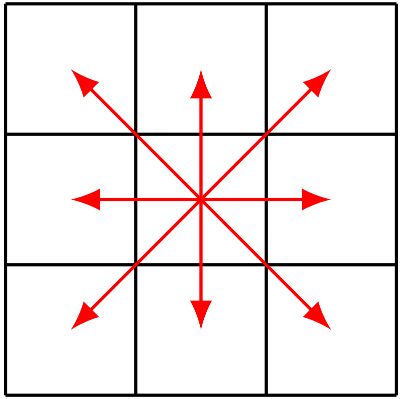
$$where\ \delta_i \sim \mathbb{N}(0, \tau_\mu)\ and$$

$$\varepsilon_i | \varepsilon_{-i} \sim \mathbb{N}(\frac{\sum_{j\sim i} \varepsilon_i}{n_i}, \frac{1}{n_i \tau_\varepsilon})$$

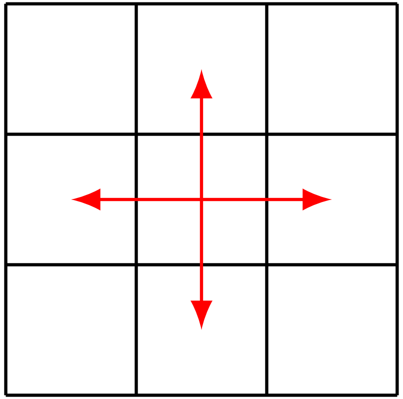
Where: δ_i = a random error for the corresponding census tract i, which is independent and identically distributed (IID) that follows a normal distribution with a mean of 0 and a variance parameter; and τ_μ = a precision parameter for the IID component.

ε_i = a spatially structured error that follows a normal distribution conditional on the neighboring census tract ε_{-i}; and τ_ε = a precision parameter for the spatial component, while $j \sim i$ denotes the census tract j is a neighboring census tract of the census tract i; and n_i is the total number of the neighboring geographical unit of the census tract i.

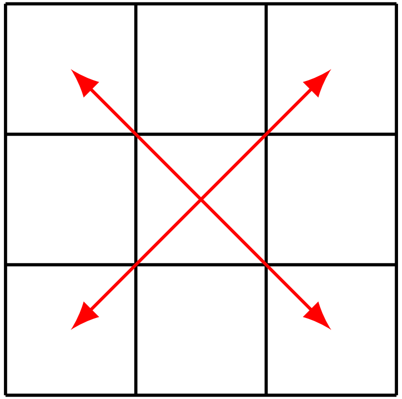
Define neighbors



Queen’s contiguity (left)



Rook’s contiguity(middle)

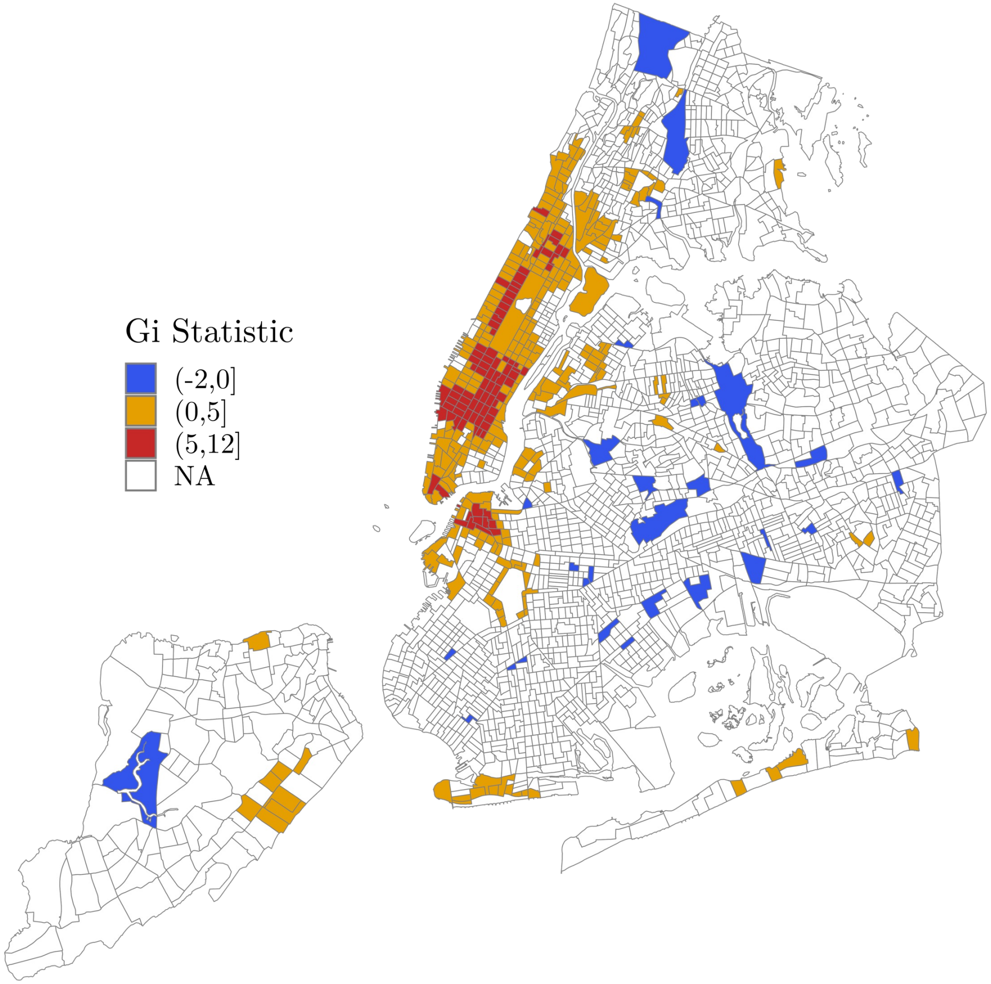


Bishop’s contiguity(right)

RESULTS

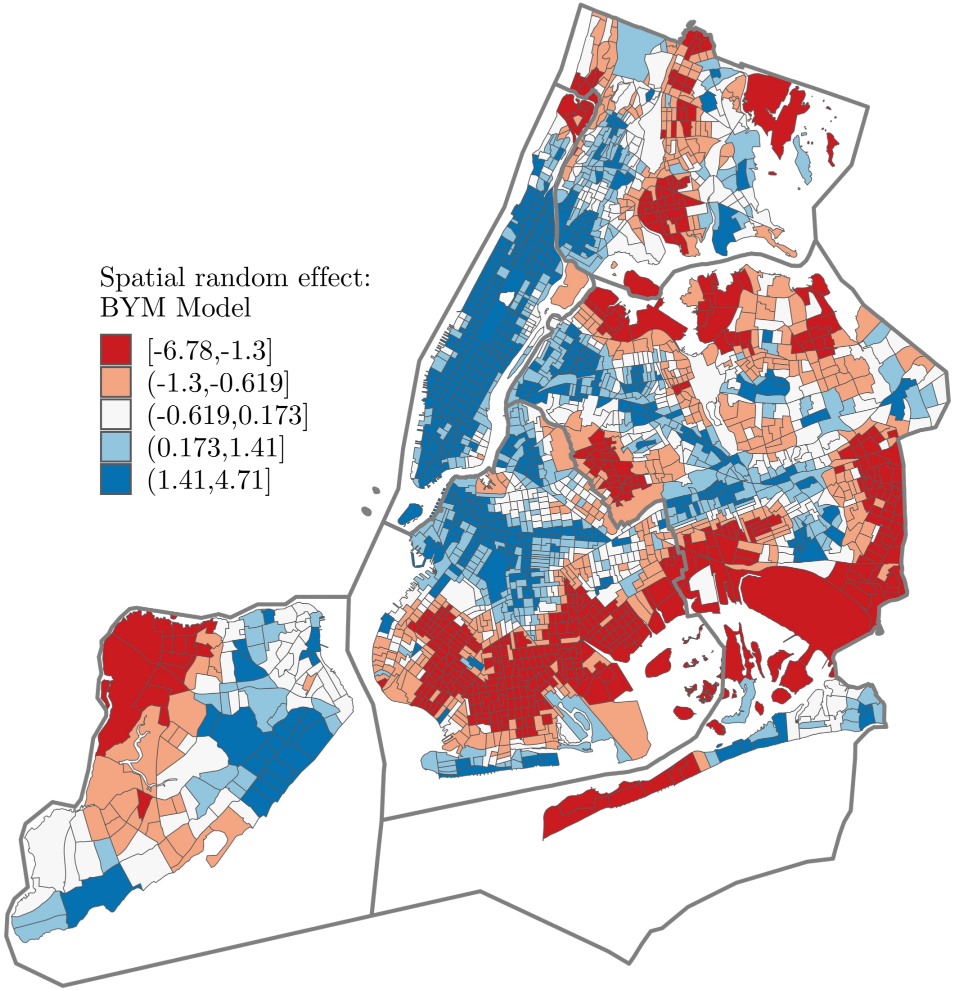
Hot spot analysis

- The Moran's *I* of the counts of public Wi-Fi hotspots is 0.463 and statistically significant under the Monte-Carlo simulation with a randomization of 999 permutations ($p < .001$), which indicates a positive spatial autocorrelation across New York City.
- The z-scores of Geti-Ord Gi^* statistics has identified the spatial cluster of high values of Wi-Fi hotspots count is mainly located in Manhattan and a smaller cluster in Brooklyn Height and the pier. Meanwhile, the "cold spot", or the regions with low count of Wi-Fi hotspots are pretty scattered in the rest of four boroughs other than Manhattan


Gi Statistic
(-2,0]
(0,5]
(5,12]
NA

	Baseline NB model 1		NB with IID model 2		NB with ICAR model 3		NB with BYM model 4	
	Mean	95% CR	Mean	95% CR	Mean	95% CR	Mean	95% CR
Intercept	-0.097	(-0.181, -0.013)	-0.483	(-0.627, -0.328)	-1.146	(-1.383, -0.940)	-1.258	(-1.374, -1.125)
Demographic factors								
African American population	-0.070	(-0.206, 0.067)	-0.076	(-0.219, 0.066)	-0.102	(-0.273, 0.069)	-0.088	(-0.261, 0.084)
Asian American population	0.096	(-0.016, 0.209)	**0.140**	**(0.024, 0.255)**	0.123	(-0.014, 0.261)	0.119	(-0.019, 0.257)
Hispanic population	**0.393**	**(0.234, 0.553)**	**0.477**	**(0.313, 0.643)**	**0.414**	**(0.224, 0.605)**	**0.413**	**(0.221, 0.605)**
Median age	-0.020	(-0.110, 0.070)	**-0.138**	**(-0.250, -0.027)**	0.020	(-0.091, 0.132)	0.003	(-0.112, 0.118)
Population with less than high school education	**-0.383**	**(-0.599, -0.168)**	**-0.509**	**(-0.741, -0.278)**	**-0.298**	**(-0.530, -0.066)**	**-0.304**	**(-0.541, -0.068)**
Socio-economic factors								
GINI index	**0.142**	**(0.040, 0.245)**	**0.150**	**(0.046, 0.254)**	0.054	(-0.045, 0.154)	0.054	(-0.048, 0.157)
Income	**0.508**	**(0.365, 0.651)**	**0.553**	**(0.423, 0.683)**	0.087	(-0.045, 0.220)	0.117	(-0.019, 0.253)
Commercial density	**0.272**	**(0.156, 0.388)**	**0.239**	**(0.144, 0.335)**	**0.120**	**(0.053, 0.187)**	**0.125**	**(0.057, 0.193)**
Digital infrastructure								
Population without broadband subscription	-0.035	(-0.008, 0.233)	-0.040	(-0.125, 0.046)	-0.016	(-0.100, 0.068)	-0.026	(-0.112, 0.061)
Population without Internet	0.112	(-0.117, 0.047)	**0.148**	**(0.022, 0.273)**	-0.034	(-0.153, 0.086)	-0.020	(-0.143, 0.103)
Residential with high-speed Internet connections	**-0.317**	**(-0.404, -0.229)**	**-0.258**	**(-0.353 , -0.163)**	**-0.299**	**(-0.382, -0.215)**	**-0.271**	**(-0.360, -0.182)**
Political orientation								
Population affiliated with Democrats	**0.563**	**(0.394, 0.733)**	**0.666**	**(0.503, 0.830)**	**0.357**	**(0.207, 0.507)**	**0.370**	**(0.217, 0.523)**
Population affiliated with Republicans	-0.215	(-0.350, -0.080)	-0.230	(-0.375 , -0.086)	-0.062	(-0.205, 0.080)	-0.060	(-0.206, 0.085)
Model hyperparameters								
Overdispersion hyperparameter	0.353	(0.315, 0.394)	0.590	(0.462, 0.728)	7.388	(2.269, 20.93)	13.858	(4.124, 29.825)
Precision for IID component			1.530	(1.086, 2.184)			3.996	(1.482, 10.973)
Precision for spatial component					0.169	(0.141, 0.200)	0.213	(0.163, 0.271)
DIC	6150.91		6023.32		5260.43		5262.84	
WAIC	6158.85		6012.39		5257.2		5452.08	
CPO	1.33		1.65		10.58		9.17	
Marginal log-Likelihood	-3159.61		-3146.11		-4836.40		-2704.29	

Note: NB: negative binomial regression; CR: credible region; IID: independent and identically distributed; ICAR: intrinsic conditional autoregressive regression; BYM: Besag, York, and Mollié.


Spatial random effect:
BYM Model
[-6.78,-1.3]
(-1.3,-0.619]
(-0.619,0.173]
(0.173,1.41]
(1.41,4.71]

Main findings

- The deployment of public Wi-Fi hotspots under the Internet Master's Plan is not spatially equal across the five boroughs of New York City.
- It is associated with areas with denser commercial business
- It favors neighborhoods with more population identified themselves as Democrats
- It helps with mitigating the digital inequalities in:
  - Hispanic neighborhoods
  - Neighborhoods with more residentials who do not have high-speed Internet connections
- It reinforces the existing digital inequalities in
  - Communities with more population with lower educational attainment
- It does not improve broadband access disparities between digitally marginalized areas and the rest of the city
  - Communities with more population without Internet or broadband subscription
- Without considering the spatial dimension, the relationships as above would not be accurately estimated (i.e., median household income, GINI)

Summary

- The distribution of the public Wi-Fi hotspots deployment in New York City, is spatially uneven.
- One of the main purposes for the project is to serve the commercial area
- It helps with mitigating digital inequalities to some extent, but also actually reinforces the existing gaps in the digitally marginalized areas and the rest of the city.
- These findings confirm the role of space in shaping the inequality of ICT access in the form of public Wi-Fi deployment while supporting the utility of spatially informed modeling strategies in future research studying ICT access as well.

THANK YOU!

@ryn_wang
ryanwang@psu.edu