Graphic Detail

Geographic Information Systems (GIS) organize and clarify the patterns of human activities on the Earth's surface and their interaction with each other. GIS data, in the form of maps, can quickly and powerfully convey relationships to policymakers and the public. This department of Cityscape includes maps that convey important housing or community development policy issues or solutions. If you have made such a map and are willing to share it in a future issue of Cityscape, please contact alexander.m.din@hud.gov.

Using HUD Crosswalk Files to Improve COVID-19 Analysis at the ZIP Code and Local Level

Alexander Din
U.S. Department of Housing and Urban Development

Ron Wilson
University of Maryland, Baltimore County

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research, the U.S. Department of Housing and Urban Development, or the U.S. Government.

Abstract

As the novel coronavirus disease (COVID-19) continues to infect, harm, and kill thousands of Americans, many jurisdictions and institutions are publishing data at the ZIP Code-level, including counts of tests performed, people infected, hospitalizations, and deaths. These data are leading to quickly produced publications with strong conjectures about the forming of geographic patterns. We present an alternative to ZIP Codes when working with local COVID-19 data.

The large, ambiguous shapes and skewed underlying data of ZIP Codes adversely affect statistical analyses, which can lead to incorrect conclusions, particularly in the health sciences (Beyer, Schultz, and Rushton, 2007; Cudnick et al., 2012; Grubesic and Matisziw 2006; Krieger et. al., 2002; Oregon Health Authority, 2020; Sadler, 2019; Wilson, 2015). In particular, a recent study by Harris (2020) from the National Bureau of Economic Research (NBER) simply overlaid New York City (NYC) subway ridership patterns with ZIP Code data to suggest the subway is responsible for COVID-19 concentration patterns across the city.

Using COVID-19 data from the NYC Department of Public Health[1] and subway exit turnstile data from the NYC Metropolitan Transit Authority (MTA),[2] we examine if any spatial relationships exists between the two, with a more statistically robust analysis than other authors. The COVID-19 data are cumulative to April 30, 2020. The turnstile data are between November 1, 2019, and March 15, 2020; Staten Island data were unavailable for this analysis.[3]

Using the HUD 2020 quarter 1 ZIP-to-census-tract crosswalk file,[4] we created positive COVID-19 case density estimates for a more local-level analysis at the census tract-level, thus escaping the adverse effects of ZIP Codes. Our disaggregated estimates from these larger geographic units are robust because of numerous empirical results that exemplify Gibrat's law, which states that a growth rate is proportional to the size of the distribution with which it is in contact (Santarelli, Klomp, and Thurik, 2006; Yigit, 2020). With respect to the growth of COVID-19 in specific neighborhoods, it is expected that the virus growth is clustered in census tracts with a higher population.

The bivariate cluster map in exhibit 1 shows the statistical relationship between census tract distance to the subway station and positive COVID-19 estimate densities (per square kilometer). The Short Distance-High Density (dark orange) are the clusters of interest, which represent census tracts near subway stops that are surrounded by census tracts with high positive COVID-19 estimate densities.

The Bronx and upper Manhattan are the only two boroughs that show a systemic relationship between proximity to the subway stations and high COVID-19 density estimates. However, the subway exit averages, shown as red circles, in these positive density hot spots vary from low to high, with many non-hot-spot subway stops having consistently high exit averages. This is inconsistent with the idea that COVID-19 hot spots would be near high-use subway stops and lines.

---

[1] https://github.com/nychealth/coronavirus-data

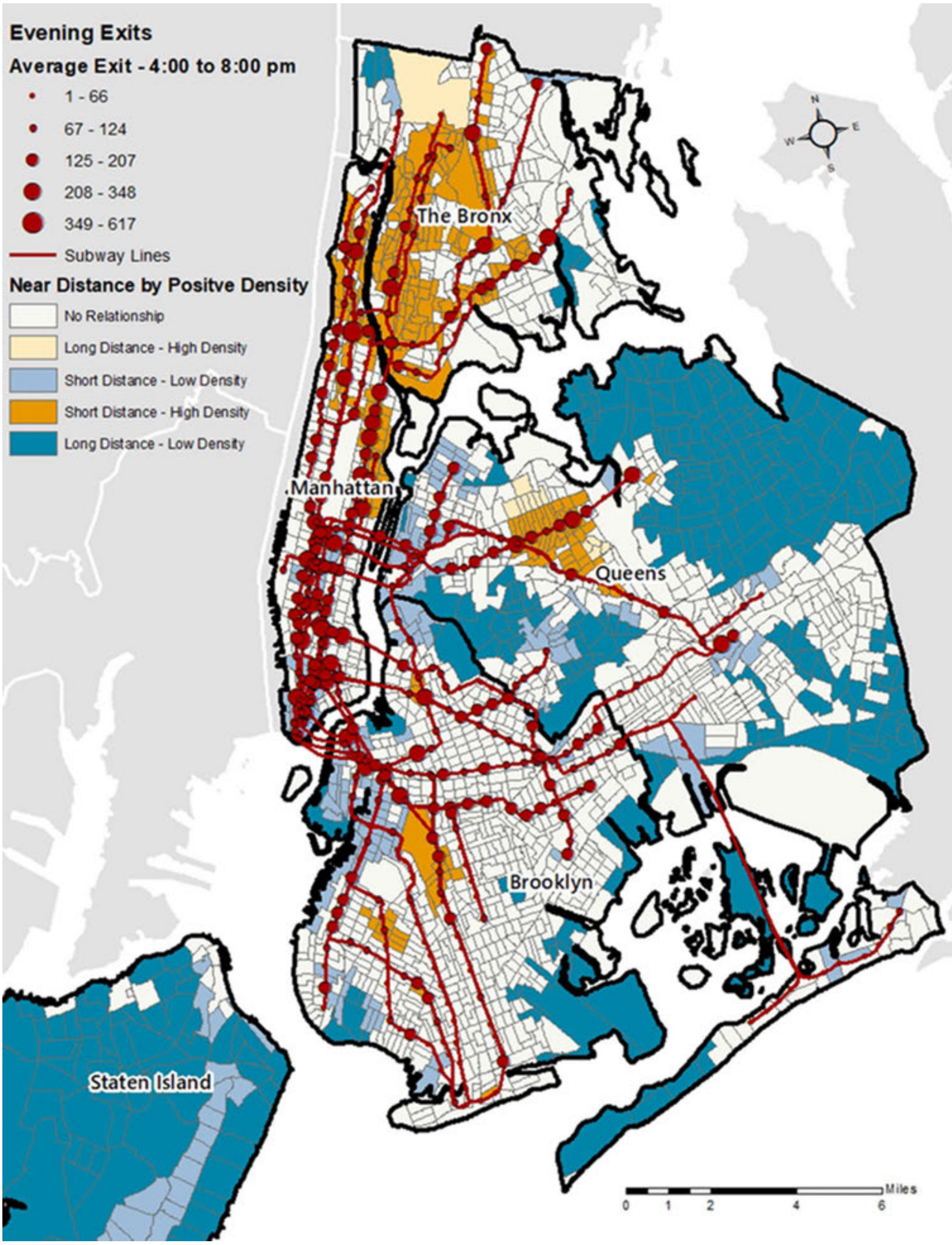
[2] http://web.mta.info/developers/developer-data-terms.html

[3] The New York City Subway does not connect to Staten Island. While Staten Island does have passenger rail service via the Staten Island Railway, it does not connect to any of the other four boroughs.

[4] https://www.huduser.gov/portal/datasets/usps_crosswalk.html

Exhibit 1

COVID-19 Cluster Relationships between Subway Exits Levels and Positive Density Estimates



Source: Metropolitan Transit Authority Turnstile Data November 1, 2019 – March 15, 2020 *http://web.mta.info/developers/turnstile.html*

Lower Manhattan does not show hot spots near subways, while Brooklyn and Queens only show three hot spots. Further, Brooklyn and Queens either show no relationship between distance to the subway and positive densities (light grey) or have Long-Distance Low-Density (dark blue) cold spots that are census tracts far from the subway, surrounded by tracts with low positive densities. Queens has two large cold spots, one in-between a series of subway stops, indicating no in-fill of positive density clusters emanating from being surrounded by the subway.

With some of the above studies suggesting a relationship between high positive densities and population density (population per square kilometer), we examine this relationship in conjunction with distance to the subway (exhibit 2). The first two correlations are the distance to the subway with positive densities and population densities, which primarily indicate that positive and population densities are moderately associated with closer proximity to the subway at about the same levels. However, the third correlation between positive and population densities shows a strong relationship, suggesting that COVID-19 densities are more associated with population density than proximity to the subway.

Exhibit 2

Correlation of Densities of COVID-19 Cases with Distance to the Subway and Population Densities

Geography	Distance & Positive Densities			Distance & Population Densities			Positive & Population Densities		
	r	*t*	*p*	*r*	*t*	*p*	*r*	*t*	*p*
The Bronx	-0.41	-8.13	< 0.001	-0.42	-8.33	< 0.001	0.88	33.52	< 0.001
Brooklyn	-0.32	-6.47	< 0.001	-0.37	-10.92	< 0.001	0.72	27.99	< 0.001
Manhattan	-0.15	-2.63	0.009	-0.12	-1.93	0.055	0.77	19.97	< 0.001
Queens	-0.37	-10.31	< 0.001	-0.49	-14.47	< 0.001	0.82	37.85	< 0.001
Staten Island	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
New York City	-0.32	-15.64	< 0.001	-0.43	-21.35	< 0.001	0.75	51.48	< 0.001

A simple regression of positive COVID-19 density estimates (y) in relation to the distance to the subway (x_1) and the population density (x_2) shows no statistical relationship between the positive estimates when controlling for the population density (exhibit 3). This baseline model indicates that increased COVID-19 estimates are related more to increased population density than to proximity to subway stations. The adjusted R-squared of 0.57 shows that this initial baseline model indicates that it has a strong explanatory power of the positive estimates being unrelated to tracts near subway stations.

Exhibit 3

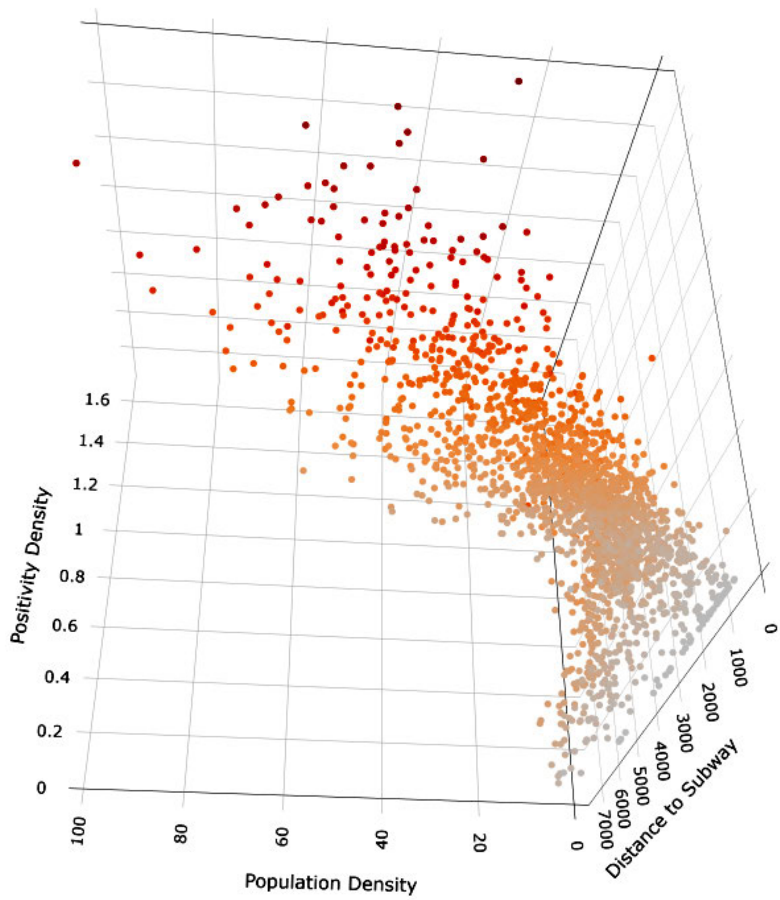
Regression Coefficients Between Distance to Subway, Positive Case Density, and Population Density

Variable	Estimate	Std. Error	t-Value	Pr(>|t|)
intercept	0.0524	0.0081	6.429	0.0000 ***
Distance to Subway	-1.2E-06	3.2E-06	-0.371	0.7100
Population Density	0.0124	0.0003	46.295	0.0000 ***

Exhibit 4, a 3D scatter plot, exemplifies the regression relationship between near distance to subway (x), population density (y), and positive estimates (z). The pattern in the data cloud reveals a higher correlation between the positive COVID-19 estimates and population density than with distance to the subway. The relationship trend in exhibit 4 shows that the positive estimate density rises steeply at very close distances to subway stations but is pulled away and spread widely on the population density axis.

Exhibit 4

Correlations between Distance to Subway, Positive Case Density, and Population Density



This pattern indicates that positive COVID-19 densities increase with population density, and population density increases as distance to the subway stations decreases. With the color gradient showing changes in positive estimate density, the pattern shows that the highest positive density estimates correspond with the highest population densities near the subway stops.

Correspondingly, the pattern revealed in the data cloud in exhibit 4 corresponds with the regression results in exhibit 3. That is, the positive COVID-19 estimates are more associated with higher population density than being close to subway stations.

Our analysis aligns with other research that suggests COVID-19 clusters may be related to something other than public transportation such as places where people spend a more significant amount of time (Bromage, 2020; Kay, 2020). With COVID-19 primarily requiring longer periods of exposure than typical subway rides, it is not yet proven that public transportation is the culprit for spreading the virus. More so, the virus seems to be associated with higher population densities, which is in line with the nature of a communicable outbreak (Yigit, 2020).

Acknowledgements

The authors thank Mark Shroder for his helpful comments.

Authors

Alexander Din is a social science analyst in the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development.

Ron Wilson is an adjunct faculty member of the Geographic Information Systems Program at the University of Maryland, Baltimore County.

References

Beyer, Kirsten, Alan Schultz, and Gerard Rushton. 2007. "Using ZIP Codes as Geocodes in Cancer Research." In *Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice*, edited by G. Rushton, M. P. Armstrong, J. Gittler, B. R. Greene, C. E. Pavlik, M. M. West, and D. L. Zimmerman. Boca Raton, FL: CRC PRESS: 37–64.

Bromage, Erin S. 2020. The Risks - Know Them - Avoid Them. Working paper. https://www.erinbromage.com/post/the-risks-know-them-avoid-them?fbclid=IwAR31ts4XTZnIJh4mJlZXX8_rZLAYgMDhCw6chpUT6F12avhv4n17LeBAXkQ.

Cudnik, Michael T., Jing Yao. Dana Zive, Craig Newgard, and Alan T. Murray. 2012. "Surrogate Markers of Transport Distance for Out-of-Hospital Cardiac Arrest Patients," *Prehospital Emergency Care* 16 (2): 266–272.

Grubesic, Tony H. and Timothy C. Matisziw. 2006. "On the Use of ZIP Codes and ZIP Code Tabulation Areas (ZCTAs) for the Spatial Analysis of Epidemiological Data," *International Journal of Health Geographics* 5 (1): 58.

Harris, Jeffrey E. "The Subways Seeded the Massive Coronavirus Epidemic in New York City." April 2020. working paper. *National Bureau of Economic Research*. https://www.nber.org/papers/w27021.pdf.

Kay, Jonathan. 2020. COVID-19 Superspreader Events in 28 Countries: Critical Patterns and Lessons. Working paper. https://quillette.com/2020/04/23/covid-19-superspreader-events-in-28-countries-critical-patterns-and-lessons/.

Krieger, Nancy., Pamela Waterman, Jarvis T. Chen, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson. 2002. "ZIP Code Caveat: Bias Due to Spatiotemporal Mismatches Between ZIP Codes and US Census–Defined Geographic Areas—The Public Health Disparities Geocoding Project," *American Journal of Public Health* 92 (7): 1100–1102.

Oregon Health Authority. (2020). COVID-19 Weekly Report, May 5, 2020. https://www.oregon.gov/oha/PH/DISEASESCONDITIONS/DISEASESAZ/Emerging%20Respitory%20Infections/COVID-19-Weekly-Report-2020-05-05-FINAL.pdf.

Sadler, Richard. 2019. "Misalignment Between ZIP Codes and Municipal Boundaries: A Problem for Public Health." *Cityscape, A Journal of Policy Development and Research* 21 (3): 335–340.

Santarelli, Enrico, Luuk Klomp, and Roy Thurik. 2006. "Gibrat's Law: An Overview of the Empirical Literature." In *Entrepreneurship, Growth and Innovation, The Dynamics of Firms and Industries*, edited by Enrico Santarelli: 41–73.

Wilson, Ronald E. 2015. "The Neighborhood Context of Foreclosures and Crime," *Cartography and Geographic Information Science* 4 (2): 162–177.

Yigit, Aydogan. 2020. "Are Covid-19 Cases Independent of the City Sizes?" *MPRA Paper No. 99697*: 1–9. https://mpra.ub.uni-muenchen.de/99697/.