

This readme file was generated on 2023-12-12 by Bartosz Bonczak

## GENERAL INFORMATION

Title of Dataset: Zipcode level COVID-19 case rate data for selected cities in the US

### Author

Name: Boyeong Hong

ORCID: 0000-0003-0748-2893

Institution: Marron Institute of Urban Management, New York University

Email: boyeong.hong@nyu.edu

### Principal Investigator Information

Name: Constantine E. Kontokosta

ORCID: 0000-0003-4831-9996

Institution: Marron Institute of Urban Management, New York University

Email: [ckontokosta@nyu.edu](mailto:ckontokosta@nyu.edu)

### Author/Associate/Co-investigator/Collaborator Information

Name: Bartosz Bonczak

ORCID: 0000-0003-4107-4711

Institution: Marron Institute of Urban Management, New York University

Email: bartosz.bonczak@nyu.edu

Date of data collection: September 2020

Geographic location of data collection: Various municipalities across the United States

Funding sources: National Science Foundation awards no. 2028687 and no. 2040898.

\*\*\*

## SHARING/ACCESS INFORMATION

Licenses/restrictions placed on the data: MIT License

Links to publications that cite or use the data:

Kontokosta, C. E., Hong, B. & Bonczak, B. J., Socio-Spatial Inequality and the Effects of Density on COVID-19 Transmission in U.S. Cities. *Nature Cities*  
<https://doi.org/10.1038/s44284-023-00008-2> (In Press)

Links to other publicly accessible locations of the data: N/A

Links/relationships to supporting or related data sets:

<https://github.com/UrbanIntelligenceLab/socio-spatial-inequality-and-the-effects-of-density-on-covid19-transmission-in-us-cities>

Was data derived from another source? Yes

If yes, list source(s):

Full list of sources is listed below and complemented with the COVID-19 sources in table 2.

Additionally, all of the relevant information is included in the Supplementary Information document accompanying the published article and in the corresponding GitHub repository.

- Zip Code Geometries:  
U.S. Census Bureau, 2019, TIGER/Line Shapefile, 2010 Census 5-Digit ZIP Code Tabulation Area (ZCTA5), url:  
<https://catalog.data.gov/dataset/tiger-line-shapefile-2019-2010-nation-u-s-2010-census-5-digit-zip-code-tabulation-area-zcta5-na>, accessed on 08.17.2020
- Demographic Information:  
U.S. Census Bureau, 2019, 2019 5-year estimate American Community Survey, url:  
<https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2019/>, accessed on 08.17.2020

**Table 1.** Administrative boundaries data.

City	Name	Provider	Link	Year
New York City	Borough Boundaries	City of New York	<a href="https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm">https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm</a>	2020
Chicago	Boundaries	City of Chicago	<a href="https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-City/ewy2-6yfk">https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-City/ewy2-6yfk</a>	2017
Philadelphia	City Limits	City of Philadelphia	<a href="https://opendataphilly.org/datasets/city-limits/">https://opendataphilly.org/datasets/city-limits/</a>	2014
Jacksonville	Duval County Boundary	Duval County	<a href="https://koordinates.com/data/global/north-america/united-states/florida/duval/">https://koordinates.com/data/global/north-america/united-states/florida/duval/</a>	2020
Charlotte	Mecklenburg County Boundary	Mecklenburg County GIS	<a href="http://maps.co.mecklenburg.nc.us/opendata/MecklenburgCounty_Boundary.zip">http://maps.co.mecklenburg.nc.us/opendata/MecklenburgCounty_Boundary.zip</a>	2020
Indianapolis	Marion County Boundary	City of Indianapolis and Marion County	<a href="https://data.indy.gov/datasets/marion-county-boundary/explore?location=39.754804%2C-86.069834%2C10.92">https://data.indy.gov/datasets/marion-county-boundary/explore?location=39.754804%2C-86.069834%2C10.92</a>	2017
Seattle	Seattle City Limits	City of Seattle	<a href="https://data-seattlecitygis.opendata.arcgis.com/maps/seattle-city-limits-2">https://data-seattlecitygis.opendata.arcgis.com/maps/seattle-city-limits-2</a>	2020
Washington	Washington DC Boundary	City of Washington, DC	<a href="https://opendata.dc.gov/datasets/washington-dc-boundary/explore">https://opendata.dc.gov/datasets/washington-dc-boundary/explore</a>	2020
Boston	City of Boston Boundary	City of Boston	<a href="https://bostonopendata-boston.opendata.arcgis.com/datasets/142500a77e2a4db94a86f7e0b568bc_9/explore?location=42.350673%2C-71.093157%2C12.88">https://bostonopendata-boston.opendata.arcgis.com/datasets/142500a77e2a4db94a86f7e0b568bc_9/explore?location=42.350673%2C-71.093157%2C12.88</a>	2020
	Municipal Boundaries	Town of Brookline	<a href="https://maps.brooklinema.gov/datasets/f308d0841afa4b41b0d2448a01b925a2_0/explore">https://maps.brooklinema.gov/datasets/f308d0841afa4b41b0d2448a01b925a2_0/explore</a>	2020
	City Boundary	City of Cambridge	<a href="https://www.cambridgema.gov/GIS/gisdatadictionary/Bou">https://www.cambridgema.gov/GIS/gisdatadictionary/Bou</a>	2020

			ndary/BOUNDARY_CityBoundary	
	City Limits	City of Somerville	<a href="https://data.somervillema.gov/GIS-data/City-Limits/pz4k-wh6e">https://data.somervillema.gov/GIS-data/City-Limits/pz4k-wh6e</a>	2017
Miami	Miami-Dade Boundary	Miami-Dade County	<a href="https://gis-mdc.opendata.arcgis.com/datasets/miami-dade-boundary/explore">https://gis-mdc.opendata.arcgis.com/datasets/miami-dade-boundary/explore</a>	2019
Las Vegas	MCD	Clark County GIS Management Office	<a href="https://clarkcountygis-cgismo.hub.arcgis.com/datasets/cgismo::mcd/explore?location=36.339490%2C-114.867170%2C9.27">https://clarkcountygis-cgismo.hub.arcgis.com/datasets/cgismo::mcd/explore?location=36.339490%2C-114.867170%2C9.27</a>	2020

**Table 2.** COVID-19 data sources used

No.	Geographical Area	State	Data Accessed Date	Granularity	Data Source	URL
1	New York City	NY	08/17/2020	Zipcode	New York City Department of Health	<a href="https://github.com/nychealth/coronavirus-data">https://github.com/nychealth/coronavirus-data</a>
2	Chicago	IL	08/17/2020	Zipcode	Chicago data portal	<a href="https://www.chicago.gov/city/en/depts/cdph/providers/health_data_and_reports/svcs/covid-19-data.html">https://www.chicago.gov/city/en/depts/cdph/providers/health_data_and_reports/svcs/covid-19-data.html</a>
3	Philadelphia	PA	08/17/2020	Zipcode	Philadelphia open data	<a href="https://www.opendataphilly.org/dataset/covid-cases/resource/d4d1e48a-d401-405c-972b-c45292c3d4f5">https://www.opendataphilly.org/dataset/covid-cases/resource/d4d1e48a-d401-405c-972b-c45292c3d4f5</a>
4	Austin	TX	08/18/2020	Zipcode	Travis county data portal	<a href="https://experience.arcgis.com/experience/3793562ab87e4299b106e0c282bb1fc5">https://experience.arcgis.com/experience/3793562ab87e4299b106e0c282bb1fc5</a>
5	Jacksonville	FL	09/09/2020	Zipcode	Florida Department of Health Open Data	<a href="https://open-fdoh.hub.arcgis.com/datasets/florida-cases-zips-covid19/data?geometry=-104.147%2C24.356%2C-63.519%2C31.156">https://open-fdoh.hub.arcgis.com/datasets/florida-cases-zips-covid19/data?geometry=-104.147%2C24.356%2C-63.519%2C31.156</a>
6	Charlotte	NC	08/25/2020	Zipcode	NC Department of Health	<a href="https://covid19.ncdhhs.gov/dashboard/about-data">https://covid19.ncdhhs.gov/dashboard/about-data</a>
7	Indianapolis	IN	08/25/2020	Zipcode	Indiana Management Performance Hub	<a href="https://hub.mph.in.gov/dataset/covid-19-cases-by-zip/resource/3ea01356-42e4-42aa-8935-493709313ca3">https://hub.mph.in.gov/dataset/covid-19-cases-by-zip/resource/3ea01356-42e4-42aa-8935-493709313ca3</a>
8	Seattle	WA	08/19/2020	Zipcode	King county website	<a href="https://www.kingcounty.gov/depts/health/covid-19/data/daily-summary.aspx">https://www.kingcounty.gov/depts/health/covid-19/data/daily-summary.aspx</a>
9	Washington	DC	08/17/2020	Zipcode	Open data DC	<a href="https://opendata.dc.gov/datasets/dc-covid-19-total-positive-cases-by-neighborhood/data">https://opendata.dc.gov/datasets/dc-covid-19-total-positive-cases-by-neighborhood/data</a>
10	Boston	MA	08/18/2020	Zipcode	Boston Public Health Commission	<a href="https://www.bphc.org/onlinenewsroom/Blog/Lists/Posts/Post.aspx?ID=1282">https://www.bphc.org/onlinenewsroom/Blog/Lists/Posts/Post.aspx?ID=1282</a>
11	Miami	FL	08/17/2020	Zipcode	Florida Department of Health Open Data	<a href="https://open-fdoh.hub.arcgis.com/datasets/florida-cases-zips-covid19/data?geometry=-104.147%2C24.356%2C-63.519%2C31.156">https://open-fdoh.hub.arcgis.com/datasets/florida-cases-zips-covid19/data?geometry=-104.147%2C24.356%2C-63.519%2C31.156</a>
12	Las Vegas	NV	09/09/2020	Zipcode	Southern Nevada Health District	<a href="https://covid.southernnevadahealthdistrict.org/data/city-reports/">https://covid.southernnevadahealthdistrict.org/data/city-reports/</a>

Recommended citation for this dataset:

Hong, B., Bonczak, B. J. & Kontokosta, C. E. Zipcode level COVID-19 case rate data for selected cities in the US *New York University* <https://doi.org/10.58153/avp10-a8h86> (2023)

\*\*\*

## DATA & FILE OVERVIEW

File List: `multicity_covid_19_data.csv`

Relationship between files, if important: N/A

Additional related data collected that was not included in the current data package: N/A

Are there multiple versions of the dataset? No

If yes, name of file(s) that was updated: N/A

Why was the file updated? N/A

When was the file updated? N/A

\*\*\*

## METHODOLOGICAL INFORMATION

Description of methods used for collection/generation of data:

In order to examine neighborhood-level COVID-19 infection rates, we collect available data on confirmed COVID-19 cases for twelve (12) cities from multiple data sources at the zip code level. Specifically, the infection rate is defined as the number of confirmed cases per 100,000 residential population. COVID-19 case data sources are presented in table 2.

Methods for processing the data:

Selection of the zip codes was obtained by performing spatial join between the administrative boundaries and the US zip code geometries using Python with Geopandas library. COVID-19 data was collected manually, combined together and merged with the resulting zip code geometries based on the zip code 5-digit identifiers. Zip code level population information from American Community Survey was appended with the same method. Finally, COVID-19 rates were calculated by dividing the number of infections by the size of the local population and multiplied by 100,000.

Instrument- or software-specific information needed to interpret the data:

Python v. 3.8, Pandas 1.1.2, Geopandas 0.8.1, Numpy 1.19.2

Standards and calibration information: N/A

Environmental/experimental conditions: N/A

Describe any quality-assurance procedures performed on the data: N/A

People involved with sample collection, processing, analysis and/or submission: Boyeong Hong, Bartosz Bonczak, Constantine Kontokosta

\*\*\*

DATA-SPECIFIC INFORMATION FOR: multicounty\_covid\_19\_data.csv

Number of variables: 6

Number of cases/rows: 5,080

Variable List:

Zip Code (integer): FIPS Zip code number - 5 digits.

City (string): Name of the main city / municipality.

State (string): FIPS State code - 2 characters.

Cases (integer): Number of COVID-19 cases as of August 2020.

Population (float): Number of residents of the zip code as defined by the  
American Community Survey.

Case Rate (float): Zip code level COVID-19 case rate by 100,000 people.

Missing data codes: Blank / NaNs. For the "Population" column it indicates the unsuccessful match between the reported COVID cases data and the ACS population information. For the "Case Rate" column it can also indicate the undefined value due to the reported population of "0".

Specialized formats or abbreviations used: N/A