**What are the patterns in population health and correlations with COVID-19 outcomes in NYC? (EDA, Feature Engineering)**

Merged/Combined dataset

1. NYC ZipCode COVID Cases (reported by April 1 at zip code level and borough)
2. Data2goNYC – by borough (**GEO\_LABEL),** Census Data[[1]](#footnote-1) (**GEO\_ID),** and PMUAs.

Data2GoNYC[[2]](#footnote-2) (<https://ssrc.formstack.com/forms/index.php>) [we use the 4th edition].

The data are available in spreadsheets (.xlsx) by sector. Each column is a separate unit of geography. Different units of geography sometimes have different sources or years, which is why we do not combine all geographies for a given indicator into a single column.

1. Data at the following levels: ACS Public Use Microdata Sample (PUMS) data; Boro; Community District; and Census Tract
2. Concentration of COVID-19 in NYC by ZIP code
3. Racial composition and COVID-19
4. Share of Adults with a college degree (ACS)
5. [NYC Data](https://www1.nyc.gov/site/doh/data/tools.page)/ [Community Health Profiles Home](https://a816-health.nyc.gov/hdi/profiles/) (<https://a816-health.nyc.gov/hdi/profiles/>). Notes can be found here - <https://a816-health.nyc.gov/hdi/profiles/>

**New York City Community Health Survey (CHS)** -  is a telephone survey conducted annually by the DOHMH, Division of Epidemiology, Bureau of Epidemiology Services. CHS provides robust data on the health of New Yorkers, including neighborhood, borough, and citywide estimates on a broad range of chronic diseases and behavioral risk factors. **Limitations** [The survey sampling methodology does not capture the following groups: adults in households without any telephone service, and prior to 2009 adults who could only be reached by cell phone. The CHS also excludes adults living in group quarters, such as college dormitories or nursing facilities].

**Citation**: New York City Department of Health and Mental Hygiene. [**Community Health Survey [YEAR(S)]**](https://www1.nyc.gov/site/doh/data/data-sets/community-health-survey.page); public use dataset accessed on [DATE].

Housing Quality: poor maintained housing is associated with poor health outcomes, including worsened asthma and other respiratory illness. “Housing cockroaches infestation” , which are a potential asthma trigger.

1. Data from the American Community Survey

[<https://www.census.gov/topics/preparedness/events/pandemics/covid-19.html>]

**NYC health data**

1. New York City Locations Providing Seasonal Flu Vaccinations [Updated February 7, 2020]
2. [Primary Care Access and Planning - Health Insurance Enrollment (Map)](https://data.cityofnewyork.us/Health/Primary-Care-Access-and-Planning-Health-Insurance-/j7nh-487d) [Updated September 11, 2018]

* Health insurance enrollment and assistance with SNAP benefits (Food Stamps)
* Data collected to promote health insurance enrollment among uninsured New Yorkers. Data collected manually. Each record represents a health center location where health insurance enrollment and assistance with SNAP benefits (Food Stamps) are offered. Data can be used by general public seeking assistance with signing up for health insurance or SNAP. Data may change as program needs are changed, e.g., opening of new site or change in hours of operation at a particular site.

1. [View of NYC Health + Hospitals patient care locations – 2011 (Map)](https://data.cityofnewyork.us/Health/View-of-NYC-Health-Hospitals-patient-care-location/4bew-e6kg) [Updated July 3, 2019]

* NYC Health + Hospitals is the largest municipal health care system in the country serving more than one million New Yorkers every year. The City’s public health care delivery system provides trauma, emergency, medical, mental health and substance abuse services across the five boroughs. This is a list of the public hospitals, skilled nursing facilities, and some of the community-based health centers that are part of the NYC Health + Hospitals system as of 2011.

1. NYC Quarterly Economic Update (<https://comptroller.nyc.gov/reports/new-york-city-quarterly-economic-update/>)

# New York City Census Data (<https://www.kaggle.com/muonneutrino/new-york-city-census-data>)

### Context: There are a number of Kaggle datasets that provide spatial data around New York City. For many of these, it may be quite interesting to relate the data to the demographic and economic characteristics of nearby neighborhoods. I hope this data set will allow for making these comparisons without too much difficulty. Exploring the data and making maps could be quite interesting as well.

### **Acknowledgements:** The data here was taken from the American Community Survey 2015 5-year estimates (<https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml)>.

The census block coordinate data was taken from the FCC Census Block Conversions API (<https://www.fcc.gov/general/census-block-conversions-api>)

As public data from the US government, this is not subject to copyright within the US and should be considered public domain.

1. Data2GoNYC[[3]](#footnote-3) (<https://ssrc.formstack.com/forms/index.php>) [we use the 4th edition]

DATA2GO.NYC data are available in spreadsheets (.xlsx) by sector. Each column is a separate unit of geography. Different units of geography sometimes have different sources or years, which is why we do not combine all geographies for a given indicator into a single column.

* Data at the following levels: ACS Public Use Microdata Sample (PUMS) data; Boro; Community District; and Census Tract.

Data and links:

1. Data2GoNYC (<https://ssrc.formstack.com/forms/index.php>) [we use the 4th edition]
2. NYC Quarterly Economic Update (<https://comptroller.nyc.gov/reports/new-york-city-quarterly-economic-update/>)
3. Data from the American Community Survey [<https://www.census.gov/topics/preparedness/events/pandemics/covid-19.html>]
4. NYC Shapefiles for the maps – census tract (<https://data.cityofnewyork.us/City-Government/2010-Census-Tracts/fxpq-c8ku>)
5. NYC Quarterly Economic Update (<https://comptroller.nyc.gov/reports/new-york-city-quarterly-economic-update/>)
6. [NYC Data](https://www1.nyc.gov/site/doh/data/tools.page)/ [Community Health Profiles Home](https://a816-health.nyc.gov/hdi/profiles/) (<https://a816-health.nyc.gov/hdi/profiles/>).

1. Public Use Microdata Areas (PUMAs) are statistical geographic areas defined for the dissemination of Public Use Microdata Sample (PUMS) data. They are also used for disseminating American Community Survey (ACS) and Puerto Rico Community Survey period estimates. [↑](#footnote-ref-1)
2. [https://data2go.nyc/map/?id=107\*36047015900\*ahdi\_puma!undefined!ns\*!other\_pop\_cd\_506~ahdi\_puma\_1~sch\_enrol\_cd\_112~age\_pyramid\_male\_85\_plus\_cd\_20~median\_household\_income\_puma\_397~median\_personal\_earnings\_puma\_400~dis\_y\_perc\_puma\_102~poverty\_ceo\_cd\_417~unemployment\_cd\_408~pre\_k\_cd\_107!\*air\_qual\_cd~ahdi\_puma\*family\_homeless\_cd\_245#10/40.8278/-73.9586](https://data2go.nyc/map/?id=107*36047015900*ahdi_puma!undefined!ns*!other_pop_cd_506~ahdi_puma_1~sch_enrol_cd_112~age_pyramid_male_85_plus_cd_20~median_household_income_puma_397~median_personal_earnings_puma_400~dis_y_perc_puma_102~poverty_ceo_cd_417~unemployment_cd_408~pre_k_cd_107!*air_qual_cd~ahdi_puma*family_homeless_cd_245#10/40.8278/-73.9586) [↑](#footnote-ref-2)
3. [https://data2go.nyc/map/?id=107\*36047015900\*ahdi\_puma!undefined!ns\*!other\_pop\_cd\_506~ahdi\_puma\_1~sch\_enrol\_cd\_112~age\_pyramid\_male\_85\_plus\_cd\_20~median\_household\_income\_puma\_397~median\_personal\_earnings\_puma\_400~dis\_y\_perc\_puma\_102~poverty\_ceo\_cd\_417~unemployment\_cd\_408~pre\_k\_cd\_107!\*air\_qual\_cd~ahdi\_puma\*family\_homeless\_cd\_245#10/40.8278/-73.9586](https://data2go.nyc/map/?id=107*36047015900*ahdi_puma!undefined!ns*!other_pop_cd_506~ahdi_puma_1~sch_enrol_cd_112~age_pyramid_male_85_plus_cd_20~median_household_income_puma_397~median_personal_earnings_puma_400~dis_y_perc_puma_102~poverty_ceo_cd_417~unemployment_cd_408~pre_k_cd_107!*air_qual_cd~ahdi_puma*family_homeless_cd_245#10/40.8278/-73.9586) [↑](#footnote-ref-3)