EconEX

Data Analysis Project

Mentor: Omer Arain

Han Jiang

COVID-19 Early Outbreak in NYC: Factors in High Community Infection Rate

I. Introduction

Since the global outbreak of COVID-19, the United States has made up approximately one third of the world's confirmed cases.[1] Specifically, by 12th June 2020, New York constituted about 20% of the confirmed cases nation-wide, while the cases of New York City alone take up about 10%. The gigantic numbers of infection make New York City one of the epicenters of the U.S.

Noticeably, race-related data of the pandemic manifests that the hospitalization rate of Black/African and Hispanic Americans is three times the rate of white Americans. Against the population ratio of the white majority to other ethnic minorities, the hospitalization rate along with other covid-19 data manifest a disproportionate damage to ethnic minorities. Moreover, when looking into the neighborhoods with distinct infection indexes, it would be surprising to notice that neighborhoods of high case/death counts correspond to neighborhoods of ethnic

[1] Johns Hopkins Coronavirus Resource Center.(2020). Retrieved 12 June, 2020 from https://coronavirus.jhu.edu/us-map

minority[2][3]. However, it is an oversimplification that the cause of COVID-19 infection is merely related to race or neighborhoods of enthic minorities, thereby dangerously misleading to a racial bias and a neglect of existing social-economic inequalities within the city.

Race is no more than a facade of the cause of suffering under COVID-19. When looking into the epicenters of this pandemic in NYC, they are not only areas of ethnic minorities but also areas with poor living conditions. For example, there is a neighborhood called Corona[4]. Despite the unfortunate coincidence of name with the Coronavirus, Corona has the 7th highest infection rate per capita among other areas in New York City, as well as a relatively high death rate.[5] The demographics of Corona is what makes it worth paying attention to. That is, within New York City, it has the second largest scale in family size, a household median income below $50,000, a young-oriented labor force with 40% of them in the services sector.

In this paper, one will argue that the contributors of this pandemic are not race but poverty, reliance on public transportation, lack of health insurance for employed labor, and community-wise inequality.

The argument is structured into two parts: the data analysis and the policy implication. In the data analysis part, I will develop my hypothesis based on features of Corona, and explain the sources of data, methodology, as well as the analysis. The methods include basic data

[2] Mollenkopf, J., Pereira, J., and Romalewski, S. (2013). Communities of interest and city council districting in New York 2012-2013. City University of New York. Retrieved 27 June, 2020 from https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Centers/Center%20for%20Urban%20Research/Resources/COI-Report_CURversionforwebsite.pdf

[3] City of New York. (2020). COVID19: Data summary. Retrieved 16 June, 2020 from https://www1.nyc.gov/site/doh/covid/covid-19-data.page

[4] ZIP code 11368. This area consists of two neighborhoods—Corona and North Corona.

[5] same as 3

visualization and OLS regression. In the policy implication part, based on the findings of data analysis, I will provide discussion of policy implication and a policy alternative matrix.

II. Data and methodology

In this report I gather data from two sources: the U.S. Census Bureau's American Community Survey (ACS) and the New York City Health Department Summary. In ACS there is geographic information called Public Use Microdata Areas (PUMAs) that helps to present data about demographics, economic and social characteristics based on specific regions—which are called PUMAs. According to ACS, the 2010 PUMA Guidelines divided New York City into 55 PUMAs, as subordinary regions of the five boroughs—Bronx, Brooklyn, Manhattan, Queens, and Staten Island[6]. In the NYC Department of Health COVID-19 summary, the data is provided with indexes such as cases & death counts, case & death rate per 100,000 people, total tests, and other related data. Specifically, the data about case counts, case rates and total testing are labeled with neighbor ZIP code. With this geographic information, the COVID-19 data can be merged with the demographic data according to ACS ZIP-PUMA conversion[7].

As for methodology, I use data visualization (maps, scatter plot, line plot and bar charts) and OLS regression on selected variables. In terms of COVID-19 data, I choose the case & death count, case & death rate per 100,000 people, and total COVID tests. In terms of demographic

[6] U.S. Census, 2010 Public Use Microdata Area Criteria and Guideline. Retrieved from https://www.census.gov/programs-surveys/geography/guidance/geo-areas/pumas.html

[7] The reason for using PUMAs is because the Census Bureau only provides 2018 ACS data based on PUMA while ZIP code is no longer available. Hence, the solution is to convert covid-19 data by ZIP code to Census PUMAs so that the data can be merged.

and economic data, I choose age, race, percent distribution—including household income, occupation, and means of transportation to work—to describe the neighbor of Corona. Meanwhile, I use mean household income, mean cash public assistance[8], population of employed people in the labor force without, percentage of race, reliance on public transportation, and occupation by type.

The analysis follows three steps: firstly, an overview of the NYC pandemic based on ZIP codes; secondly, my hypothesis of the contributors of COVID-19 within the epicenters of NYC based on a descriptive analysis of the place Corona; thirdly, a verification of my hypothesis by using an OLS regression analysis on COVID-19 data and NYC demographic & economic data based on PUMAs.

[8] Public assistance refers to assistance programs which offer either cash payment or similar benefits to individuals and households from any governmental entity.

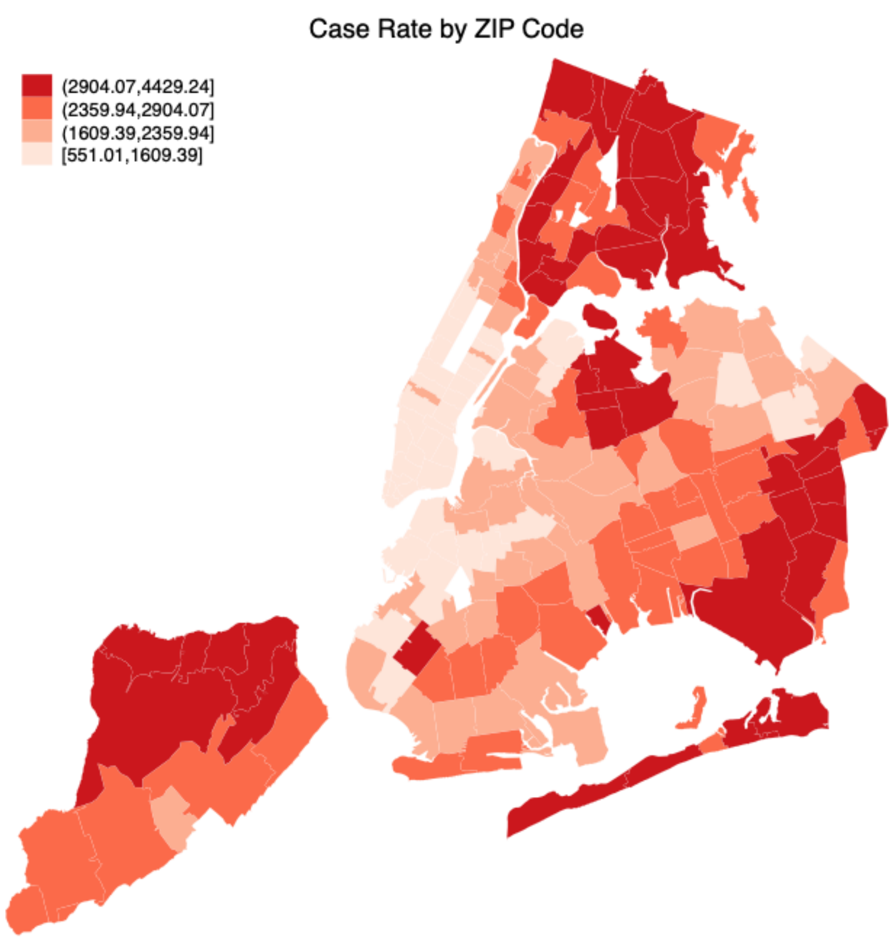
III. Results and Analysis

A. Overview of the pandemic in New York City

Similar to what NYC health department's summary, I plot the case rate per 100,000 people based on ZIP code.

Figure 1: Case Rate by ZIP code

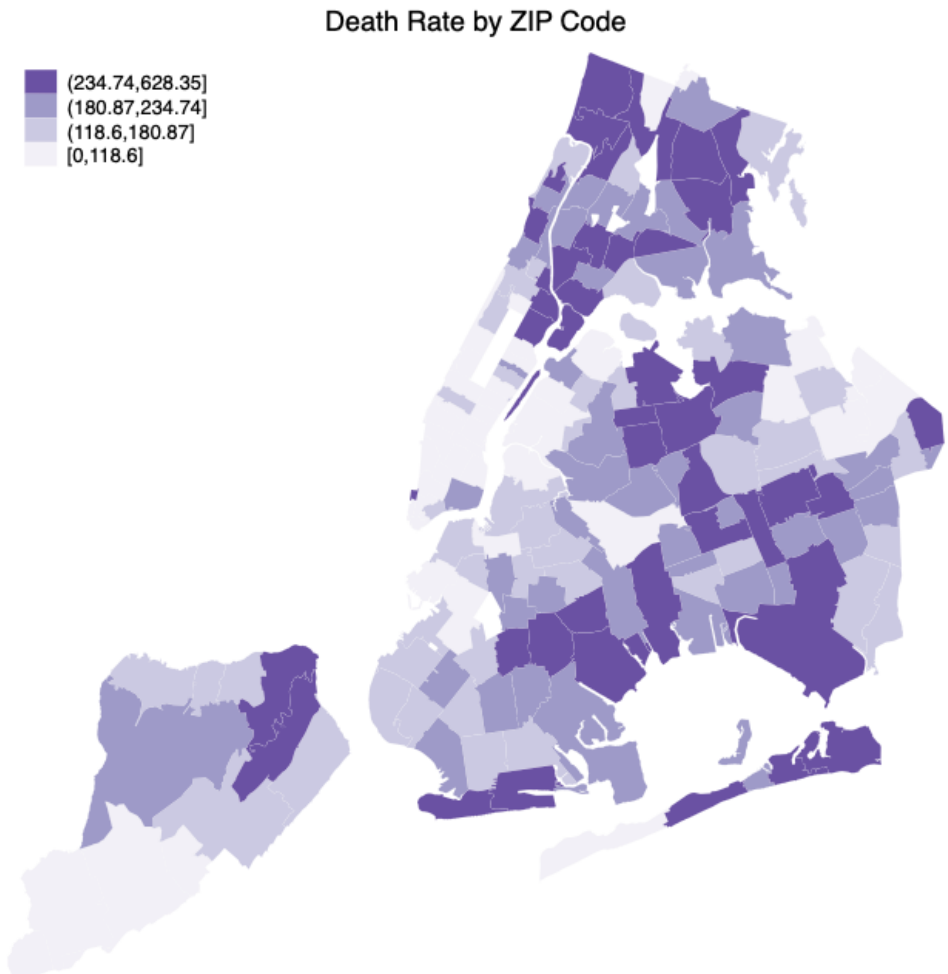
Data: NYC Health Department, COVID-19 Summary at 14th June, 2020



Corona is the dark red area at the center area (marked with a blue star) with a death rate of 4143.59 per 100,000 people, which is one of the highest rates among the city. As for death rate, the plot is shown in figure 2.

Figure 2: Death rate by ZIP code

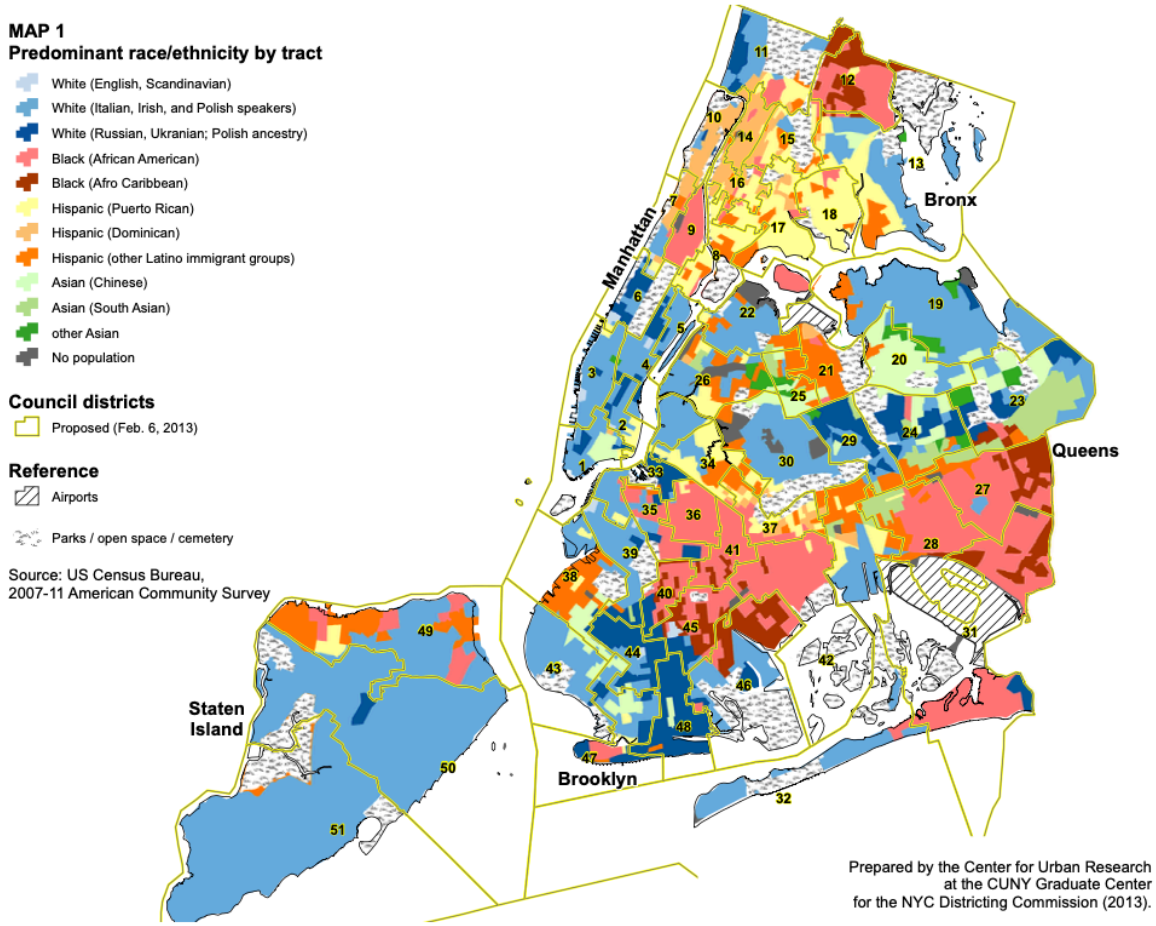
Data: NYC Health Department, COVID-19 Summary at 14th June, 2020



The map by death rate indicates several areas (dark purple) that face the highest lethal risk. Comparing the overlap of the places that face both high case rate and death rate, it is sufficient to say that the location of neighborhoods under severe damage from COVID-19 are highly relevant to the locations of neighborhoods of ethnic minorities, shown in figure 3.[9]

[9] Mollenkopf, J., Pereira, J., and Romalewski, S. (2013). Communities of interest and city council districting in New York 2012-2013. City University of New York. Retrieved 27 June, 2020 from https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Centers/Center%20for%20Urban%20Research/Resources/COI-Report_CURversionforwebsite.pdf

Figure 3: Predominant race by tract



To answer the question why geographic locations of communities of ethnic minorities are under a high danger of infection, we need to look into their demographics and economic status.

Hypothesis of COVID-19 contributors based the case of Corona

Firstly, Corona is a young neighborhood since over 70% of its inhabitants are aged between 16-64 years. Secondly, the majority of Corona's residences is Hispanic (over 60%) based on figure 4.

Figure 4: Distribution of Race (2018)

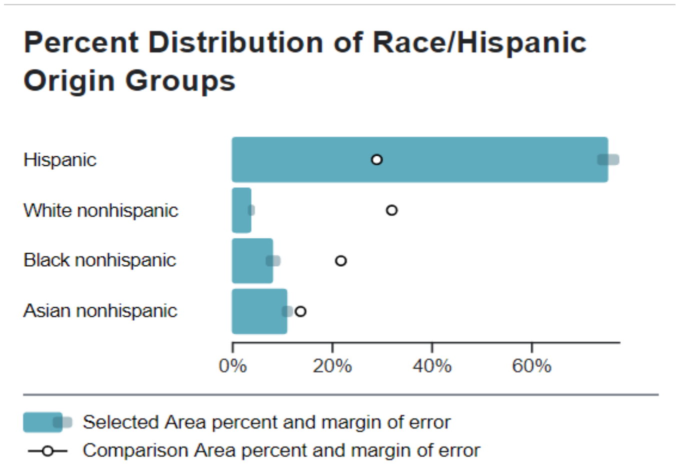
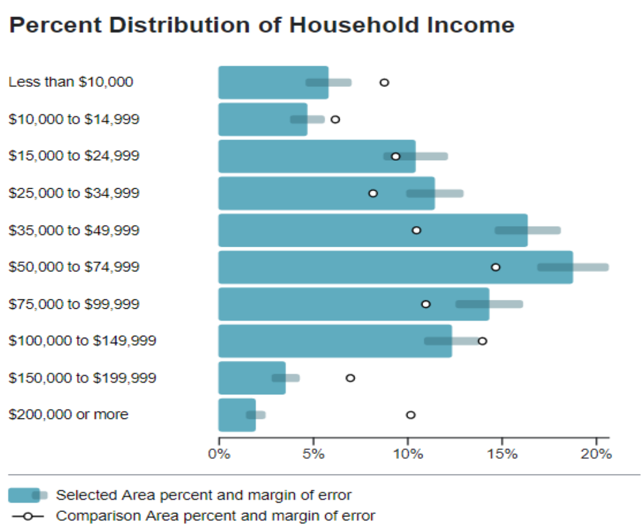


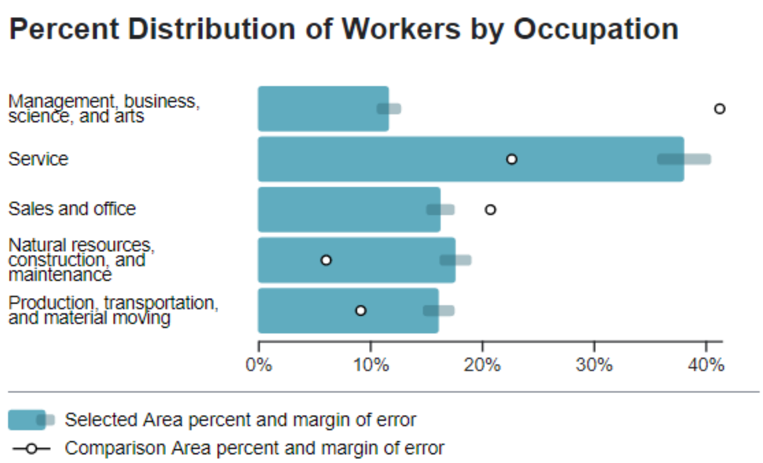
Figure 5: Distribution of Household Income



Thirdly, figure 5 manifests that about 17.9% of the households in Corona has an annual income that is lower than the poverty level ($32,402)[10 11], while a rate of 15.6% for the entire city.

Fourthly, as shown in figure 6, about 38% of the workers in Corona work in the service sector, which means their working condition is relatively riskier due to more frequency of human contact. Meanwhile, about 27% of NYC's total population work in the services sector.

Figure 6: Distribution of Workers by Occupation

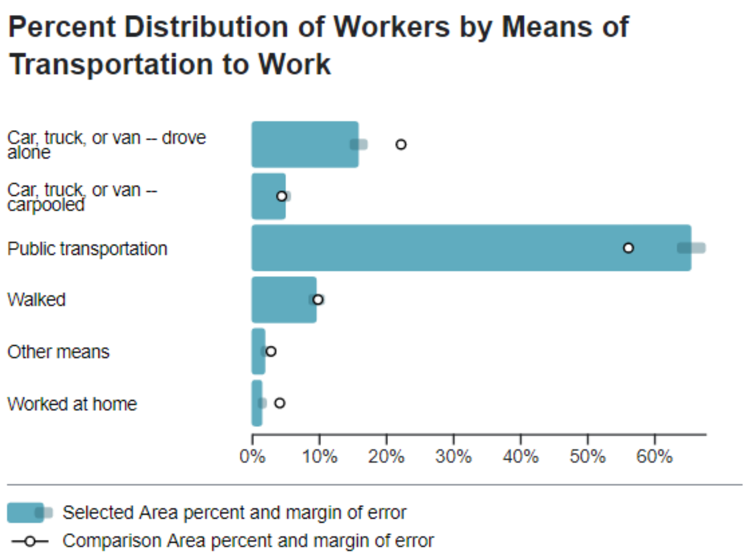


Fifthly, the bulk of the workers heavily rely on public transportations to commute to work, which is known as a way of transportation with a high risk of infection, according to figure 7. In addition, NYC has about 56% of its population relying on public transport to work while about 22% of NYC's total population rely on cars, trucks or vans for commuting.

[10] New York City, 2018 NYC Opportunity 2018 Poverty Report. Retrieved 16 June, 2020 from: https://www1.nyc.gov/assets/opportunity/pdf/NYCPov-Brochure-2018-Digital.pdf, Assuming the family size is two adults with two children, the household poverty line is estimated to be $32,402.

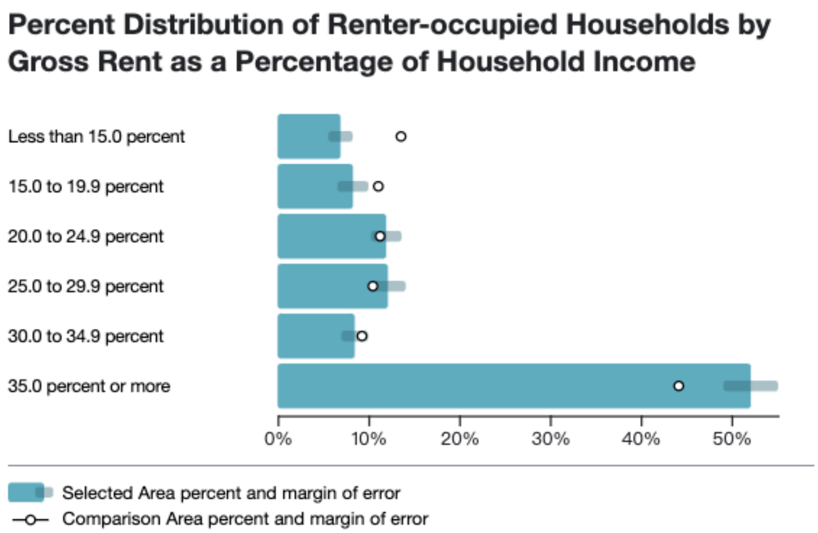
[11] NYC Population FactFinder. Based on American Community Survey (ACS) 2018. Retrieved 16 June, 2020 from https://popfactfinder.planning.nyc.gov/profile/4023/economic

Figure 7: Distribution of Workers by Means of Transportation to work



Sixthly, considering the social-distancing order, it is necessary to look at the housing data in order to know whether people are capable of having housing accommodation to follow the social distancing order. Although nearly 92% of the housing units in Corona are occupied, it is not sufficient to conclude that the social distancing order can be effectively conducted because many of the tenants are bearing a high rent.

Figure 8: Distribution of Renter-occupied households by rent-income ratio



Seventhly, from figure 8, it says that at least 50% of the renter households in Corona are having a house rent burden that is over 35%. Considering the 25% household poverty rate, one believes a considerable amount of people have to continue working in order to afford the housing rent.

Lastly, with regards to housing, only 76.2% of the population in Corona have health insurance coverage. Among the labor force aged between 18 to 64 years, 61.6% of the employed and 64.7% of the unemployed have health insurance coverage. As for those who are out of the labor force, 26% of them don't have health insurance. Therefore, approximately one fourth of the residences live under a lack of accessible medical resources. Furthermore, only about a half of the labor force (employed and unemployed) have medical insurance coverage. Given the fact that a large group of people in Corona still go out to work, it is a worrying situation for these people to go out while their lives are in jeopardy.

Given all the information above, instead of race, there are a lot more factors specifying the facts that Corona is facing numerous sources of danger from infection. In terms of social distancing, residences in Corona can hardly comply because it is not economic-sustainable given the high house rent debt ratio to income and low-income level to acquire living necessities. In terms of safety to work, over 30% of residences in Corona work in the services industry and about 60% of them rely on public transport. In addition, approximately 40% of the employees do not have health insurance.[12] Consequently, no matter on the way to work, venues of work, or insurance welfare of work, the target group is under-protected and vulnerable by exposing themselves to high risk of infection.

Therefore, **my hypothesis** is that the contributors of this pandemic are poverty, reliance on public transportation, and the lack of health insurance for employed labor. To verify my assumptions, a quantitative analysis, the OLS regression, is applied.

B. OLS Regression

1. Data Description

The data is from Census Bureau's American Community Survey (ACS) 2018 and the NYC health department on 14 June 2020. As stated in the data and methodology part, I choose several demographic and economic factors as variables, including mean household income, mean public assistance per household, population without health insurance coverage (total population and only the labor force). As for COVID-19 data, I use the counts of death and case, rates of death and cases. However, due to the complexity of the medical causes of death, the case count is

[12] Other factors such as education and government intervention might also be relevant. However, due to difficulty of acquiring data and quantifying, this paper will merely focus on a limited set of factors.

preferred in my regression analysis because it has a more straightforward connection to the risk of infection and other estimators.

Due to the difficulty of data gathering, the NYC health department was not able to provide demographic information like race by borough or PUMA. Hence, it is not feasible to analyze covid-19 data based on the patients' races and locations of residency. The alternative is to discuss the relationship between high risky neighborhoods and their ratios of race distribution.

Table 1: Data Description

```
  obs:            55
 vars:            36                          27 Jun 2020 02:26
```

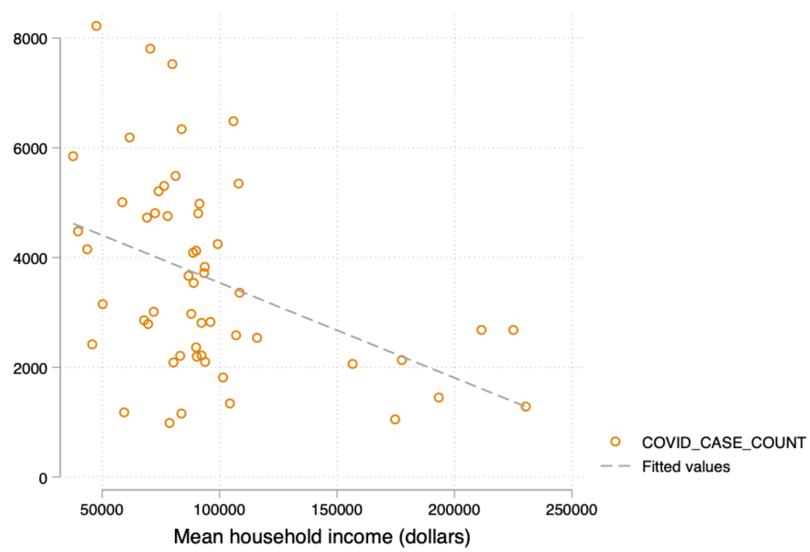
variable name	storage type	display format	value label	variable label
puma	str99	%99s		PUMA
id	float	%8.0g		
mh_income	long	%12.0g		Mean household income (dollars)
pubtrans	long	%12.0g		Public transportation (excluding taxicab)
pubtrans_perc~t	float	%9.0g		Public transportation in percent (excluding taxicab)
job_mgmt_bus_~t	float	%9.0g		Occupation Management, business, science, and arts occupations
job_services	float	%9.0g		Occupation Service occupations
job_sales_off~e	float	%9.0g		Occupation Sales and office occupations
job_nature_co~n	float	%9.0g		Occupation Natural resources, construction, and maintenance occupations
job_productio~g	float	%9.0g		Occupation Production, transportation, and material moving occupations
borough_group	str13	%13s		BOROUGH_GROUP
mhd_inc	long	%12.0g		mean household income per dollar
mhdpub_inc	int	%8.0g		Mean cash public assistance income per household (dollars)
pcap_inc	long	%12.0g		Per capita income (dollars)
lbr_NotInsured	int	%8.0g		19-64 years Labor Force Employed No health insurance coverage
covid_case_co~t	int	%8.0g		COVID_CASE_COUNT
covid_case_rate	float	%9.0g		COVID_CASE_RATE
covid_death_c~t	int	%8.0g		COVID_DEATH_COUNT
covid_death_r~e	float	%9.0g		COVID_DEATH_RATE
percent_posit~e	float	%9.0g		PERCENT_POSITIVE
total_covid_t~s	int	%8.0g		TOTAL_COVID_TESTS
pop_denominator	float	%9.0g		POP_DENOMINATOR
p_Hispanic_La~o	float	%9.0g		Percent Hispanic Latino
p_white	float	%9.0g		Percent White
p_Black_African	float	%9.0g		Percent Black or African American
p_Native_Indi~a	float	%9.0g		Percent American Indian and Alaska Native
p_Asian	float	%9.0g		Percent Asian
p_Native_Hawa~c	float	%9.0g		Percent Native Hawaiian and Other Pacific Islander
p_other_race_~e	float	%9.0g		Percent Other race alone
p_over_one_race	float	%9.0g		Percent Two or more races

2. Plotting

As mentioned in the Data and Methodology session, I chose variables including mean household income, mean cash public assistance[13], total population of people with health insurance, population of employed people in the labor force with health insurance, population of underage inhabitants with health insurance.

To better understand the relationship, I use Stata to generate scatter plots between the case/death counts and selected variables.

Figure 9: Case Count and mean household income



---

[13] Public assistance refers to assistance programs which offer either cash payment or similar benefits to individuals and households from any governmental entity.

Figure 10: Case Count and mean cash public assistance

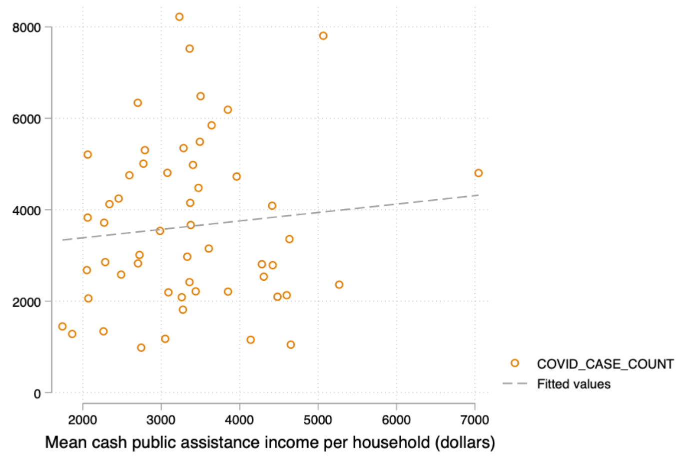
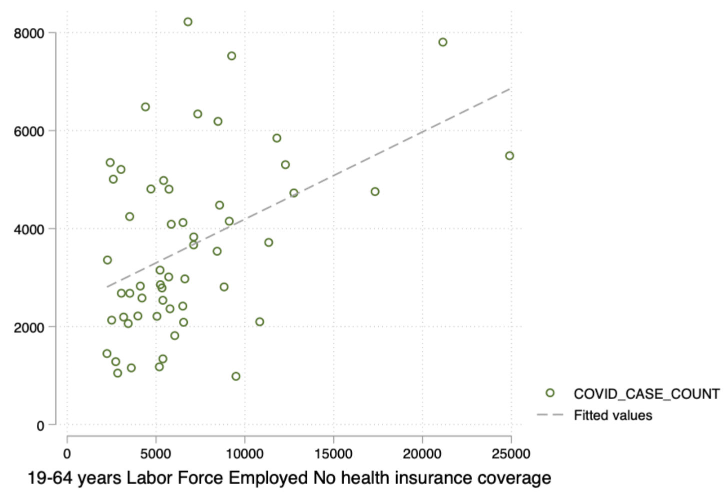


Figure 9 reveals that the mean household income is negatively correlated to the cases counts among all 55 PUMAs in NYC on 14 June, 2020. However, at the same time, figure 10 indicates a relatively weaker positive correlation between the case count and public assistance.

Figure 11: Case Count and the employed labor force without health insurance



As for figure 11, similarly, the case count is positively correlated with the population of employed labor force that is not medically insured.

Figure 12: Death count and employed labor force without health insurance

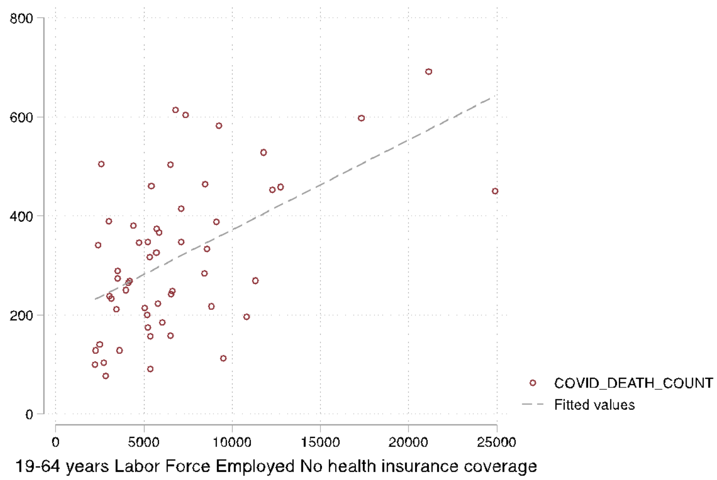


Figure 12 points out that there's a negative correlation between the death rate. Noticeably, the observations tend to cluster at the range 2,000 to 10,000, which is similar to the data characteristics in figure 9.

In summary of the scatter plot, the figures verify that the count of COVID cases and death are correlated with variables of mean household income, conditions of health insurance coverage, and questionably the mean cash public assistance per household. To check the statistical significance of the impact of variables, I use OLS regression in the next session.

3. Regression Analysis

Firstly, I want to know whether public assistance works as a way of government intervention.

Table 2: Regression on COVID-19 case rate against household income and public subsidy

Source	SS	df	MS
Model	28209551	2	14104775.5
Residual	145110255	51	2845299.11
Total	173319805	53	3270185.01

Number of obs	=	54
F(2, 51)	=	4.96
Prob > F	=	0.0108
R-squared	=	0.1628
Adj R-squared	=	0.1299
Root MSE	=	1686.8

covid_case~t	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
mhdpub_inc	.0740866	.2345836	0.32	0.753	-.3968596	.5450328
mh_income	-.0178101	.0058462	-3.05	0.004	-.0295468	-.0060735
_cons	5044.708	1046.884	4.82	0.000	2943.002	7146.414

The results indicate that the public assistance is irrelevant while the household income and conditions of medical insurance for employed labor force are significantly relevant.

Table 3: Regression based on hypothesis

Source	SS	df	MS
Model	57697206.8	4	14424301.7
Residual	116519014	50	2330380.28
Total	174216221	54	3226226.31

Number of obs	=	55
F(4, 50)	=	6.19
Prob > F	=	0.0004
R-squared	=	0.3312
Adj R-squared	=	0.2777
Root MSE	=	1526.6

covid_case_c~t	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
mh_income	.0027026	.0103768	0.26	0.796	-.0181399	.023545
lbr_NotInsured	.1777978	.0554825	3.20	0.002	.066358	.2892377
pubtrans	-.0456748	.0192345	-2.37	0.021	-.0843085	-.0070412
job_services	38.77022	45.58838	0.85	0.399	-52.79672	130.3372
_cons	3090.86	1826.024	1.69	0.097	-576.8173	6758.537

In table 3, based on my hypothesis[14] I ran the COVID-19 case count, mean household income, mean public assistance payment per household, and the population of employed labor force without health insurance, number of people who rely on public transport in this PUMA, as well as the number of people working in the services sector. Part of my hypothesis is verified because the result indicates that labor health insurance and reliance on public transport are significant factors of covid-19 case count. However, the p-value of mean household income and types of occupation implies a correlation between variables, which will be resolved later.

Table 4: Regression on COVID-19 case counts against distributions of PUMA inhabitants by race

Note: Hispanic-Latino as the constant variable

Source	SS	df	MS			
Model	158503.708	6	26417.2846	Number of obs	=	55
Residual	1088173.67	48	22670.2849	F(6, 48)	=	1.17
				Prob > F	=	0.3405
Total	1246677.38	54	23086.6182	R-squared	=	0.1271
				Adj R-squared	=	0.0180
				Root MSE	=	150.57

covid_death_~t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_white	-.118256	1.187202	-0.10	0.921	-2.505286	2.268774
p_over_one_r~e	3.750583	21.02739	0.18	0.859	-38.52781	46.02898
p_Native_Ind~a	-50.41553	97.86768	-0.52	0.609	-247.1917	146.3606
p_Native_Haw~c	-248.8079	186.556	-1.33	0.189	-623.904	126.2881
p_Black_Afri~n	-2.042239	1.421105	-1.44	0.157	-4.899562	.8150832
p_Asian	-3.26967	2.185113	-1.50	0.141	-7.663135	1.123794
_cons	415.8688	80.1541	5.19	0.000	254.7082	577.0294

[14] "Therefore, my hypothesis is that the contributors of this pandemic are poverty, reliance on public transportation, and the lack of health insurance for employed labor. To verify my assumptions, a quantitative analysis, the OLS regression, is applied."

By regressing the case counts and the percentage of each race living in the PUMAs, we can refute the null that the race itself is the cause why the case/death counts of African and Hispanic-Latino American are disproportionately higher than data of the white. However, it still fails to explain the phenomena that high-risky areas of COVID-19 correspond to the ethnic-minority-predominant neighborhoods.

To answer this question, as well as to avoid omitted variable bias in table 3, I generate a new dummy variable called under poverty (udpoverty). If udpoverty =1, it indicates that the observed PUMA is under the 25% percentile of household income among all 55 PUMAs[15].

Table 5: Regression on Case rate against PUMAs under poverty line; the PUMAs under 25% percentile of household income

Source	SS	df	MS			
				Number of obs	=	55
				F(1, 53)	=	8.06
Model	23002543.6	1	23002543.6	Prob > F	=	0.0064
Residual	151213677	53	2853088.25	R-squared	=	0.1320
				Adj R-squared	=	0.1157
Total	174216221	54	3226226.31	Root MSE	=	1689.1

covid_case~t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
udpoverty	1484.613	522.8571	2.84	0.006	435.8947	2533.332
_cons	3239.244	263.7945	12.28	0.000	2710.139	3768.349

[15] The 25% percentile household income among 55 PUMAs is $71974, calculated by Stata and shown in table A.1.

puma
NYC-Bronx Community District 1 & 2--Hunts Point, Longwood & Melrose PUMA; New York
NYC-Bronx Community District 11--Pelham Parkway, Morris Park & Laconia PUMA; New York
NYC-Bronx Community District 12--Wakefield, Williamsbridge & Woodlawn PUMA; New York
NYC-Bronx Community District 3 & 6--Belmont, Crotona Park East & East Tremont PUMA; New York
NYC-Bronx Community District 4--Concourse, Highbridge & Mount Eden PUMA; New York
NYC-Bronx Community District 5--Morris Heights, Fordham South & Mount Hope PUMA; New York
NYC-Bronx Community District 7--Bedford Park, Fordham North & Norwood PUMA; New York
NYC-Bronx Community District 9--Castle Hill, Clason Point & Parkchester PUMA; New York
NYC-Brooklyn Community District 13--Brighton Beach & Coney Island PUMA, New York
NYC-Brooklyn Community District 16--Brownsville & Ocean Hill PUMA, New York
NYC-Brooklyn Community District 5--East New York & Starrett City PUMA, New York
NYC-Manhattan Community District 11--East Harlem PUMA, New York
NYC-Manhattan Community District 12--Washington Heights, Inwood & Marble Hill PUMA; New York
NYC-Queens Community District 4--Elmhurst & South Corona PUMA, New York

Noticeably, these PUMAs under 25% percentile of household income are also the ones with high case/death counts. The correlation between the geographic locations and covid-19 data is positively correlated with poverty rather than ethnicity.[16] To interpret the table, it means if you live in the listed PUMA--areas under 25% percentile of household income--the case count is expected to be about 1484 cases more than other areas.

So far, my hypothesis is almost validated by the data, by refuting the influence of ethnicity on COVID-19 infection, as well as verifying the impact of income, type of occupation, and way of transportation to work. Hence, to improve the regression model, I use a method called partialling-out to avoid the case that mean household income is correlated with the type of occupation.

[16] A new comparison between the influence of ethnicity and poverty is given in table A.2 in the appendix.

Table 6: Regression on case count against variables in hypothesis by using partialling-out

```
. reg udpoverty job_services

      Source |       SS           df       MS      Number of obs   =        55
-------------+----------------------------------   F(1, 53)        =     45.84
       Model |  4.84023016         1  4.84023016   Prob > F        =    0.0000
    Residual |  5.59613348        53  .105587424   R-squared       =    0.4638
-------------+----------------------------------   Adj R-squared   =    0.4537
       Total |  10.4363636        54  .193265993   Root MSE        =    .32494

------------------------------------------------------------------------------
   udpoverty |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
job_services |    .031994   .0047254     6.77   0.000      .022516     .041472
       _cons |  -.5023749   .1200748    -4.18   0.000    -.7432144   -.2615355
------------------------------------------------------------------------------

. predict rudpoverty, residual

. reg covid_case_count  lbr_NotInsured  pubtrans rudpoverty

      Source |       SS           df       MS      Number of obs   =        55
-------------+----------------------------------   F(3, 51)        =      9.68
       Model |  63213454.1         3  21071151.4   Prob > F        =    0.0000
    Residual |   111002767        51  2176524.84   R-squared       =    0.3628
-------------+----------------------------------   Adj R-squared   =    0.3254
       Total |   174216221        54  3226226.31   Root MSE        =    1475.3

------------------------------------------------------------------------------
covid_case_c~t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
lbr_NotInsured |   .2211643   .0465648     4.75   0.000     .1276816    .3146469
      pubtrans |  -.0505937   .0158133    -3.20   0.002    -.0823403   -.0188472
    rudpoverty |   1256.964   632.3706     1.99   0.052    -12.57334    2526.501
         _cons |   4171.867   684.0007     6.10   0.000     2798.678    5545.056
--------------------------------------------------------------------------------
```

As shown in table 6, the rudpoverty is the residual of regressing the dummy variable udpoverty against the number of people in the services sector by PUMA. By doing so, we rule out the correlation between these two variables and get the result purely about the impact if the observation is a PUMA under 25% percentile of household income.

The result shows that all variables in my hypothesis are statistically significant. Considering the small sample size (n=55), the p-value of residual (rudpoverty) is expected to be lower if we obtain a larger sample because of the rule of normality, which can also be verified if using the robust standard errors.[17] [18]

IV. Recommendations

As for policy alternatives, we need to focus on three criteria: ability to flatten the curve, financial cost for the government, and political feasibility to implement.

The policy alternatives are as follows.

Firstly, regulating the employer companies to ensure employees are covered by health insurance, as employed labor is legally protected to have health insurance coverage.

Secondly, improving the level of sanitation of public transport, such as buses, coaches, city metro, commuter rail and others. Measures may include free hand-sanitizers, free masks, routine disinfection on seats, ticket machines, and so on.

Thirdly, as the regression analysis provided, it is the area with poverty issues that is highly risky of infection. Although social-economic conditions of these neighborhoods may also be associated with other factors like systemic racism (which can be hardly quantified), it is of vital importance to give special assistance to targeted areas. Not only does these places deal with clustered residences under severe risk of infection--due to type of occupation, way of

[17] Also, for the coefficient of public transport, an anti-intuitive result indicates the more people rely on public transport, the less case count it is expected to be. Two explanations might be given. Firstly, the data might be biased and the sample size is too small (n=55). Secondly, people are more cautious when they use public transport.
[18] However, seen in the table #x, if we control for only PUMAs under 25% percentile of household income, it is a positive correlation between public transport reliance and case count.

transportation, and lack of insurance--but also lack the financial feasibility to carry out community-based precautions against COVID-19.

Fourthly, as mentioned in my analysis of Corona, the large debt-rate ratio due to house rent and low income are one of the primary reasons that people violate the social-distancing order. This tenant rent burden is drastically rising for both employees and private owned businesses for a high rate of lay-off and shut-down order in June. In addition, people without available private vehicles face more danger of infection because they cannot avoid human contact when purchasing daily necessities, commuting or working.

Table 7: Policy Alternatives Matrix

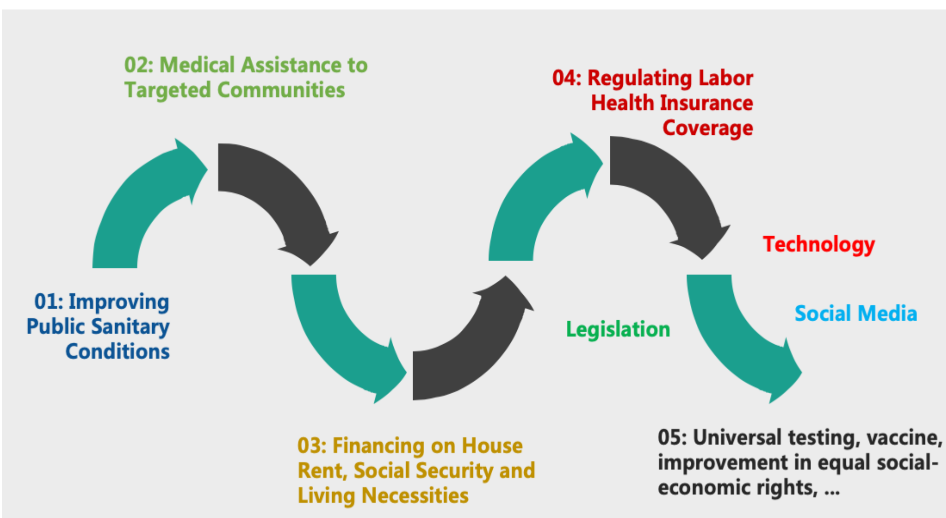
Note: scale 1 to 4; 1 indicates the highest or easiest

Policy Alternatives Matrix for Government Response to COVID-19				
	Policy			
	Regulating Labor Health Insurance Coverage	Medical Assistance to Targeted Communities	Improving Public Sanitary Conditions	Financing on House Rent, Social Security and Living Necessities
Flattening the Curve	3	1	2	4
Cost	4	3	2	1
Political Feasibility	4	1	2	3

In terms of ranking, it is believed that starting with targeted communities and targeted sectors of public services (public transportation in this case) are the easiest with regard to the financial and political resources of the NYC government. Arranging government resources is an executive-wise matter which has fewer obstacles compared to policies that require cooperation from the legislators, namely, regulation on workers' health insurance.

My recommendation is to implement these policies in a chronological manner (shown in figure 13). The NYC government can start by giving special assistance to the targeted neighborhood with financial disability to fund the fights against COVID-19. Later, the NYC government can invest more resources to the sanitary conditions of public transportation, such as more free hand sanitizers, frees masks, and routine disinfection on public facilities. Then, as the situation of COVID-19 in all epicenters are alleviated, the government can start to address regulations on employee's welfare related to legislative measures. For instance, a penalty fine to companies that refuse to pay for workers' health insurance. Lastly, speaking in the long term, we can expect diverse sources of support from technologies, legislation, business and other aspects. Not only should people unite to combat the pandemic, but also the social-economic inequalities.

Figure 13: The process of policy implementation



Suggestion for Future Studies and Conclusion

The R-square in the adjusted regression is 0.46, meaning that only 46% of the observations can be explained by this model. Factors such as education, age, and temperature are potential variables. However, many of the factors in social-economic fields are unlikely to be quantified. Not to mention our ignorance about the mechanism of the virus COVID-19. Another limitation of this study is that the sample size is too small (n=55) and the data is collected merely from New York City. Limited by the Census Bureau's primary data for non-scholar, it undermines the precision and available sample sizes of this study. In the future study, I look forward to testing the regression model if there are county-level or state-level data.

Based on the data analysis, the poverty issue is crucial to the neighborhoods with a high case/death rate. Behind the test data, it is a sophisticated problem associated with racism, social inequality, laws and so on. As supported by the data, Race is just a facade in this devastating pandemic. Like a mirror it is, the social-economic inequality inside is what we should reflect and act on.

Bibliography

City of New York. (2020). COVID19: Data summary. Retrieved 16 June, 2020 from https://www1.nyc.gov/site/doh/covid/covid-19-data.page

Johns Hopkins Coronavirus Resource Center.(2020). Retrieved 12 June, 2020 from https://coronavirus.jhu.edu/us-map

Mollenkopf, J., Pereira, J., and Romalewski, S. (2013). Communities of interest and city council districting in New York 2012-2013. City University of New York. Retrieved 27 June, 2020 from https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Centers/Center%20for%20Urban%20Research/Resources/COI-Report_CURversionforwebsite.pdf

NYC Population FactFinder. (2018). American Community Survey (ACS)-2018. Retrieved from https://popfactfinder.planning.nyc.gov/profile/4023/social

U.S. Census Bureau. (2010). Public use microdata area criteria and guideline. Retrieved from https://www.census.gov/programs-surveys/geography/guidance/geo-areas/pumas.html

U.S. Census Bureau. (2018). American Community Survey ACS 1-year estimate data profile DP-05. Retrieved from https://data.census.gov/cedsci/table?q=dp%20new%20york%20city%20demogrpahic%20&tid=ACSDP1Y2018.DP05&hidePreview=true&g=0400000US36.795000&moe=false&tp=false

Appendix Figures

Figure A.1 Summary of the variable called mean household income

Mean household income (dollars)

	Percentiles	Smallest		
1%	37647	37647		
5%	43638	39741		
10%	50237	43638	Obs	55
25%	71974	45796	Sum of Wgt.	55
50%	88711		Mean	95488
		Largest	Std. Dev.	43544.32
75%	101486	193326		
90%	174745	211515	Variance	1.90e+09
95%	211515	225079	Skewness	1.655956
99%	230362	230362	Kurtosis	5.439076

Figure A.2 Regression based on hypothesis with controlled group

. reg covid_case_count job_services pubtrans lbr_NotInsured if udpoverty==1

Source	SS	df	MS			
				Number of obs	=	14
				F(3, 10)	=	1.46
Model	12751424.1	3	4250474.69	Prob > F	=	0.2842
Residual	29153921.6	10	2915392.16	R-squared	=	0.3043
				Adj R-squared	=	0.0956
Total	41905345.7	13	3223488.13	Root MSE	=	1707.5

covid_case_c~t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
job_services	43.13423	81.53244	0.53	0.608	-138.5314	224.7998
pubtrans	.0067566	.049393	0.14	0.894	-.1032978	.1168111
lbr_NotInsured	.1805658	.1307162	1.38	0.197	-.1106881	.4718197
_cons	1479.053	3027.096	0.49	0.636	-5265.737	8223.844