|  |
| --- |
| Urban Science Intensive – Spring 2016 |
| A Seasonal Flu Social Vulnerability Index for NYC |
| March 14, 2016 |

|  |
| --- |
| Maria Filippelli, Priya Khokher, Nikhil Kishore, Ramda Yanurzha |

**INTRODUCTION**

Seasonal influenza (flu) is one of the most common disease outbreaks in New York City (NYC)[[1]](#footnote-1). Each year approximately 2,000 residents perish from flu and pneumonia, which can develop as a flu complication[[2]](#footnote-2). For 2014, this was more than homicides (328)[[3]](#footnote-3) and HIV/AIDS related deaths (1,473)[[4]](#footnote-4) combined. Nationally, the Centers for Disease Control (CDC) estimates that 5-20% of people contract seasonal flu annually, this is upwards of as many as 1.7 million NYC residents. There are 209 different types of human influenza A virus and the number is constantly increasing[[5]](#footnote-5). The flu virus has the ability to evolve with humans has made it difficult to develop a permanent solution to influenza.

In addition, the density, complex infrastructure, and international stature of the city increases the risk of spreading illness. Because of this, prevention measures such as educational outreach and vaccines are of utmost importance. Given that any agency needs to optimize their budgets, a targeted prevention outreach strategy would be helpful in decreasing the number of seasonal flu deaths. To assist in that effort this project creates a NYC Seasonal Flu Social Vulnerability Index (Flu SVI), using demographic and social variables to identify areas where the flu is likely to affect large portions of the population.

**BACKGROUND**

Currently efforts to track and quantify flu cases occur on the federal, state, and local levels. Nationally, the CDC track viral strains, pediatric mortality rates, influenza-associated hospitalizations, and other statistics on the flu. The New York State Department of Health tracks similar statistics and produces a weekly report on the county level. At the city level, the Department of Health and Mental Hygiene (DOHMH) monitors emergency department (ED) discharges to track trends over time.

While the seasonal flu outbreaks are tracked, there is no evidence of planning for their occurrence. There is, however, planning efforts for the pandemic (global epidemic) flu as it is estimated that it could have a fatality rate up to 40,000 residents in a single wave (or cycle)[[6]](#footnote-6). Both the CDC and DOHMH have developed ways to measure risk.

In 2011, the CDC published *A Social Vulnerability Index for Disaster Management [[7]](#footnote-7)*(Index) which uses 15 census variables, put into four categories:

* socioeconomic status,
* household composition and disability,
* minority status and language, and
* housing and transportation,

to better assist disaster management planning. The Index was a shift from traditional disaster management planning, which tended to focus on infrastructure only, and it incorporated social factors such as mobility and primary language. Every census tract in the United States with a non-zero population (65,081) was given an index number based on the percentile rank of the individual and grouped variables.

The DOHMH released *Vulnerable Populations: A Function-Based Vulnerability Measure for the New York City Region[[8]](#footnote-8)* in 2013 which took into account individual level data to create a score for each American Community Survey (ACS) Public Use Microdata Sample (PUMS), a demographic designation that has individual level data while still providing confidentiality. While both models are significant undertakings, the focus is on collecting information after people are diagnosed or on large-scale, global (pandemic) outbreaks.

With respect to outreach, the DOHMH offers information about flu vaccination locations and preventative measures for schools and employers on their website[[9]](#footnote-9). There is no evidence to suggest a more proactive outreach approach. Efforts have also been made to require mandatory flu shots for children to varying effectiveness. As such, flu vaccinations remain a voluntary activity.

This project borrows and builds upon both models, most specifically using census tract level data and percentile ranks like the CDC Index, to create the NYC Seasonal Flu Social Vulnerability Index (Flu SVI). The Flu SVI differs from previous models because it takes into account NYC specific variables like the high percentage of residents that commute to work via public transportation and high residential density.

**DATA**

The majority of variables used in the Flu SVI were collected from the 2010-2014 American Community Survey and utilized Social Explorer[[10]](#footnote-10) to extract a table of specific variables. Variables used are:

* *Age* – the most vulnerable age groups[[11]](#footnote-11), normalized by total population of NYC and added together
  + <18 years of age
  + >65 years of age
* *Insurance coverage* – percentage of residents without insurance coverage, normalized by total pop of NYC
* *Population density* – number of residents per square mile
* *Commute* – percentage of residents that commute to work using public transportation, normalized by total population; MTA and taxi are the modes covered; it should also be noted that this is commutes for residents only and does not include employees commuting into NYC for work
* *Household income* – median household income for the census tract
* *Poverty rate* – percentage of residents that live below the poverty rate; for this variable the threshold of $19,999 was used (the federal poverty rate for a three person household is $20,160 and the average household size in NYC is between two and three people) since ACS data is available in ranges only so the closest range max was selected (household income < $20k, normalized by total HH in all of NYC); normalized by…

Once the variables were normalized, they were each given a percentile rank. The percentile ranks were then summed for each census tract which created the index. Percentile ranking allows for even distribution among the census tracts. It can, however, overestimate skewed distributions.

Employment density was calculated using Longitudinal Employment-Household Dynamics Origin-Destination Employment Statistics (LODES). LODES data includes total jobs and is calculated at the census block level. Total jobs at the block level were summed to create the total jobs per census tract for 2014[[12]](#footnote-12).

Geographical boundary information is from the XX.

**METHODOLOGY**

In total, there are 2,064 census tracts in NYC. Census tracts with non-zero populations were removed from the index because they are largely parks and cemeteries and do not have population or employment density (or statistically significant values) available.

A percentile ranking and summation was used to create the census tract scores for the flu SVI. The value of each variable was ranked on a scale of 1 to 100. The rankings were then summed together and that summation was ranked to determine the percentile for each tract as shown:

Percentile Rank =

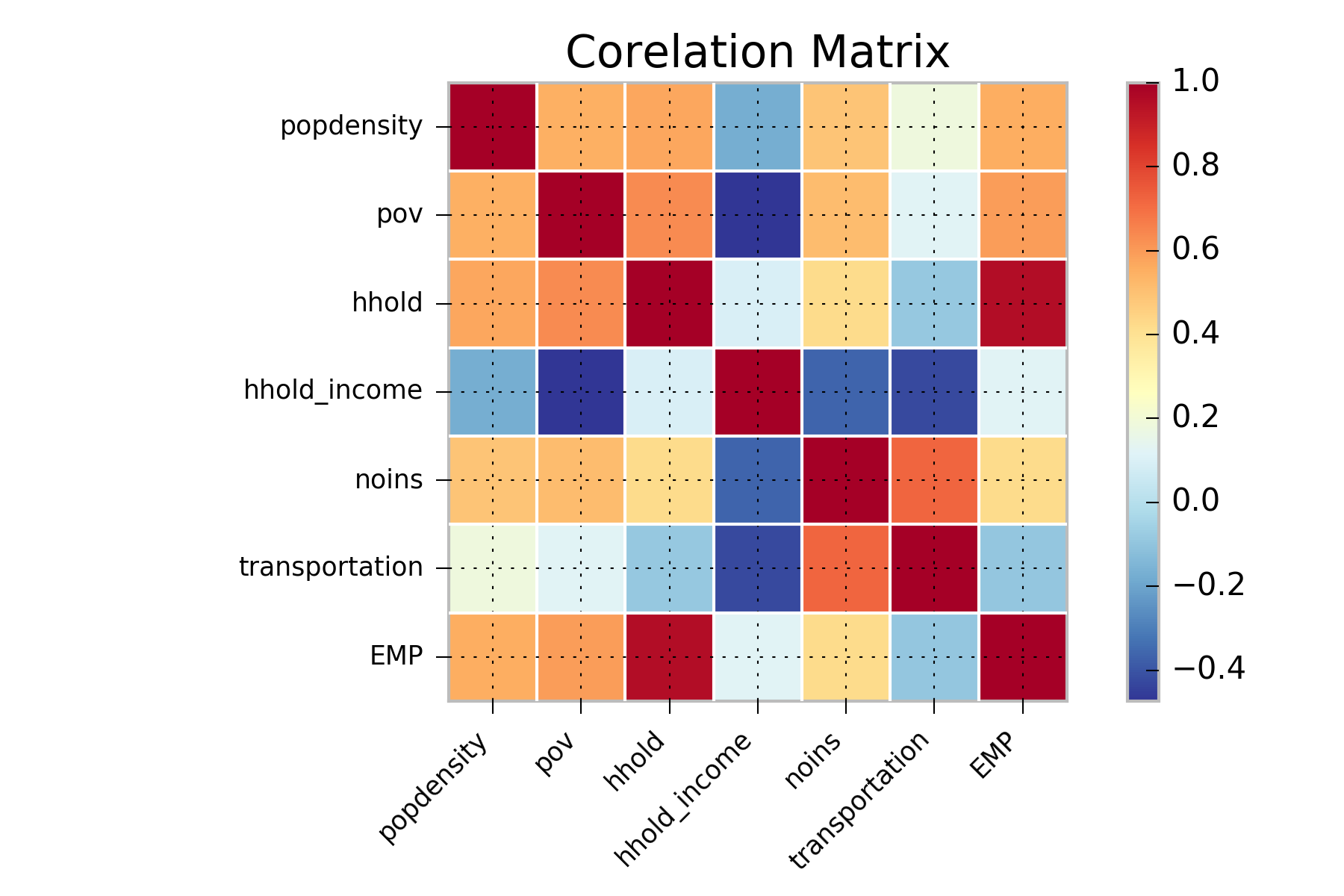
**RESULTS**

PERCENTILE RANKING

The percentile ranking creates the best Flu SVI. It takes into account all variables and highlights…

CORRELATION TESTING

Producing a correlation matrix allowed for an investigation into the dependence between multiple variables at the same time. Values near 1 and -1 show variables that are highly correlated. Table 1 shows the variables in the Flu SVI are mostly uncorrelated, with the exception of household size and employment density which appear to have perfect correlation. Additional investigation would have to occur to determine why these variables appear so closely linked.



*Table 1: Correlation Matrix for all 7 Variables*

SPATIAL AUTOCORRELATION

The percentile ranking use for the Flu SVI helps in detecting the outliers in the sense it helps to gauge the census tracts which are at a greater risk. Obtaining additional information about the surrounding census tracts (the ones perceived as most vulnerable), will help in knowing if there is a match in attributes in the surrounding areas. Basically, we assume that two census tracts to be independent, but measurements made at different locations may not be independent, or rather measurements at a closer distance might have similar values than the ones that are geographically located and this may or may not result in feature based causation.

Hence a spatial correlation mapping of the Flu SVI values was implemented to see if the census tracts near high vulnerability ones have similar rankings. Though this is not as basic as k-means clustering or advanced in attribute grouping, it gives a sense of correlated pattern measurement superimposed over the vulnerability map, helping to generalize better over a larger area the relative risk-prone areas when used in conjunction with flag map.

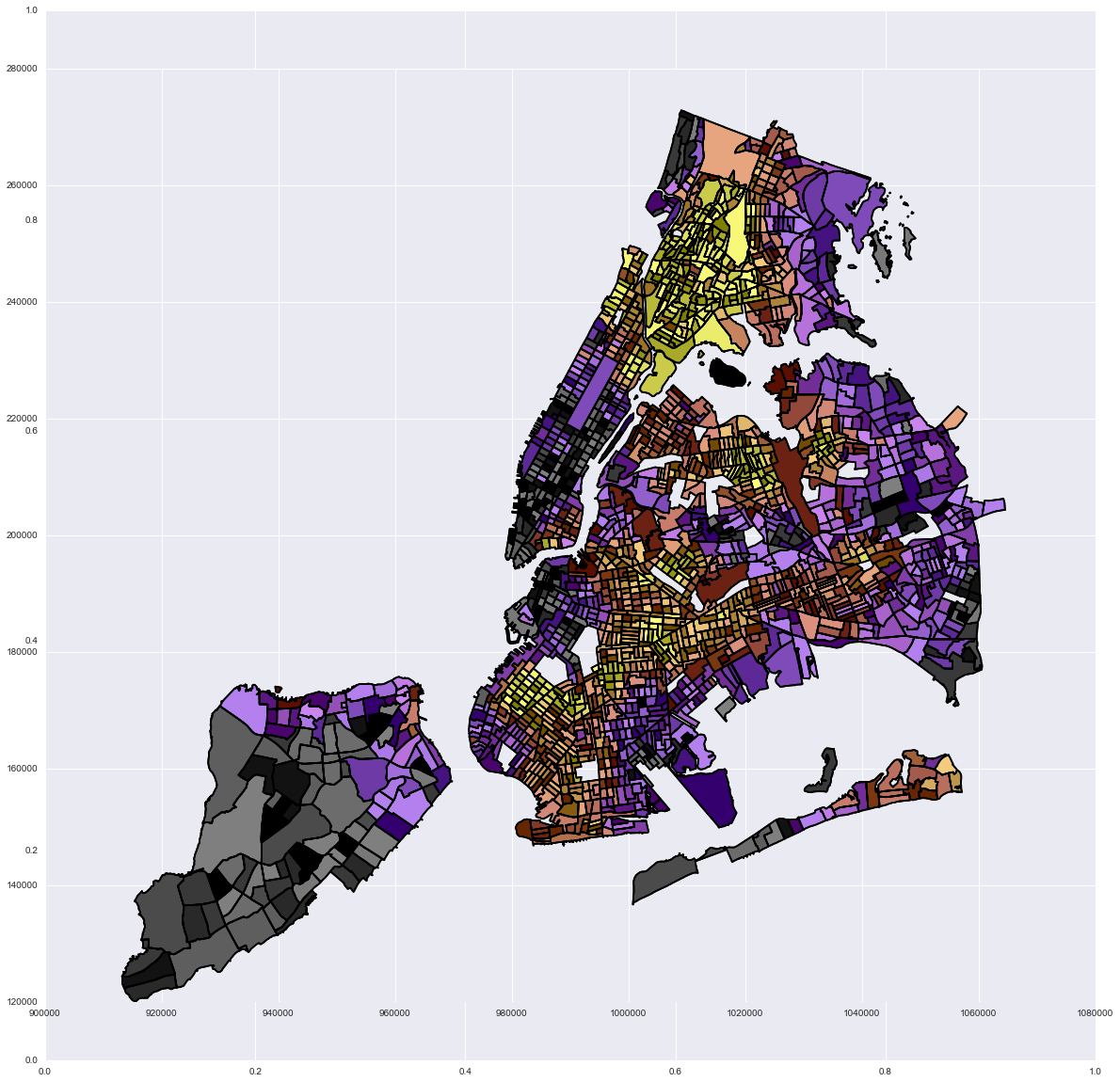
The spatial correlation that we have used here is based on Moran’s coefficient for detecting global correlations if any. This highlights census tracts of correlated rankings to be of shades of a particular color depending on their range of correlation as shown in figure XX. Moran's I (Moran 1950) tests for global spatial autocorrelation for continuous data. It is based on cross-products of the deviations from the mean and is calculated for observations on a variable at locations, as:

,



where is the mean of the variable, are the elements of the weight matrix, and is the sum of the elements of the weight matrix: .



[Longitude]

[Latitude]

*Figure XX: Spatially Correlated Map [90th Percentile]*

ACCESS GAP

One of the main ways to prevent the spread of flu is through vaccinations. The vaccination locations for NYC were overlaid on the Flu SVI, with a one half-mile buffer, giving a further indication of vulnerability. As shown in figure XX…

ILLNESS AND MORTALITY RATES

Efforts were taken to determine if the areas indicated on the Flu SVI as being highly vulnerable were also the areas that had high rates of flu illness and death. Due to privacy constraints, the lowest spatial level of data that could be obtained was at the county (borough) and could not be utilized.

**OUTREACH PLAN**

Based upon the Flu SVI, targeted outreach should be conducted in the following communities…with specific strategies to include…

Vaccination location could be better distributed throughout the city by locating sites…

**CONCLUSIONS**

The Flu SVI, in its current form, highlights areas of high vulnerability for seasonal flu outbreaks. [some additional recap here]

Next steps could include utilizing infection rates at a low level of granularity, such as census tract, to further refine outreach and healthcare needs. This SVI can also be used for other easily communicable airborne diseases. Outreach at a Community District level may be the best approach as it will cover the census tracts of high vulnerability and their surrounding area.

1. https://www1.nyc.gov/html/doh/downloads/pdf/ip/ip-death-all-rank.pdf [↑](#footnote-ref-1)
2. http://www.nyc.gov/html/doh/flu/html/public/general.shtml [↑](#footnote-ref-2)
3. http://www.nyc.gov/html/doh/downloads/pdf/dires/2014-hiv-surveillance-annual-report.pdf [↑](#footnote-ref-3)
4. http://www.nytimes.com/2015/01/01/nyregion/new-york-city-murders-fall-but-the-police-arent-celebrating.html?\_r=0 [↑](#footnote-ref-4)
5. http://www.fludb.org/brc/home.spg?decorator=influenza [↑](#footnote-ref-5)
6. http://www1.nyc.gov/assets/em/downloads/pdf/hazard\_mitigation/nycs\_risk\_landscape\_chapter\_4.9\_pandemic\_flu.pdf [↑](#footnote-ref-6)
7. http://gis.cdc.gov/grasp/svi/A%20Social%20Vulnerability%20Index%20for%20Disaster%20Management.pdf [↑](#footnote-ref-7)
8. http://www1.nyc.gov/assets/doh/downloads/pdf/em/regional\_hazards\_vulnerability\_measures.pdf [↑](#footnote-ref-8)
9. http://www1.nyc.gov/site/doh/health/health-topics/flu-seasonal-brochures-and-posters.page [↑](#footnote-ref-9)
10. http://www.socialexplorer.com/ [↑](#footnote-ref-10)
11. http://www.cdc.gov/flu/about/disease/high\_risk.htm [↑](#footnote-ref-11)
12. http://lehd.ces.census.gov/data/#lodes [↑](#footnote-ref-12)