A Spatial Analysis of the Commercialization of New York & it's Effect on the Demand For Children's Day-Care Facilities

Aneri Patel[1]

[1]Center for Urban Science and Progress, New York University

May 17, 2022

Contents

1 Problem Statement

With increasing commercialization in Manhattan, most adults - including parents with young children are engaged in paid work that involves commuting away from their children [1]. There are increasing number of cases where both parents are in the work force [1]. The effect is worse in single parent households [1]. Thus, there is an increasing need for child care facilities in Manhattan.
In this project, I first assess the change in commercialization of Manhattan by calculating fractal dimension of commercial and non-commercial areas over a decade. Later, I assess the current state of affairs - relationship between distribution of childcare facilities, median household income and number of children below 5 years of age. This analysis is done at a census tract level. Finally, the project focuses on using the Huff Model to visualize and understand the alleviating impact of a new childcare facility on pre-existing ones.

2 Data

- Child Care Facilities in New York
- MapPluto (2000, 2012, 2021) - Land use map of Manhattan
- ACS Census Tract Container & Data (2012)
    1. B01001e3 : Number of females below 5
    2. B01001e27 : Number of males below 5
    3. B19049e1 : Median Household Income

3 Methods Considered

In order to assess the change in commercialization of Manhattan over time, I planned on using the pafrac function in R to calculate fractal dimension.

The second part of the project focuses on assessing the current state of affairs - are children's day care facilities distributed based on prosperity or number of children in an area(census tract)? For this part, I considered performing a Geographically Weighted Regression (dependent variable: No. of children's day care facilities, independent variables: median household income, number of children below 5 years old). However, I faced issues in manipulating the shapefile containing child care facility information to derive the number of child care facilities per census tract, thus I tried performing the same analysis indirectly.

Finally, in this project, to analyze how a new child care facility would reduces the stress of pre-existing childcare facilities is analyzed using a Huff Model.

4 Methodology

4.1 Quantifying Commercialization in New York over Time using Fractal Dimensions

MapPluto data was used to analyse the Land Use in New York over time. MapPluto data classifies Land Use into 11 classes. The following steps were followed for each year (2000, 2012, 2021)

1. The polygon data is filtered to select data corresponding to Manhattan `"BOROUGH" = 'MN'`.
2. Then, the data is converted from polygon to raster using the `polygon to raster` tool.
3. Then, I reclassify the data into 2 classes using the `reclassify` tool:
   - Commercial Area (1): (`04`: Commercial Buildings, `05`: Commercial and Residential Buildings)
   - Non-Commercial Area (0): (`01, 02, 03, 06, 07, 08, 09, 10, 11`)
4. The resulting raster data is then exported as a `.TIF` file.
5. The path to this file is entered into the `landscapemetrics.R` file provided in lab 2 to produce fractal dimensions for both classes.

4.2 GWR - Is access to Child Care Facilities dependent on Income of a Neighborhood?

In this section, we follow an indirect analysis to assess the relationship between presence of childcare centers, median income of a census tract, and the number of children below 5 years in the census tract. The following steps were followed:

1. ACS tract container for 2012 is merged with the `X01` and `X19` data. All property fields are turned off except `GEOID, GEOID_Data, B01001e3, B01001e27, B19049e1`.
2. A new variable is created `below5 = B01001e3 + B01001e27`
3. The childcare facility shapefile is overlayed with the ACS tract data. Both layers are selected to represent only Manhattan and not all of NYC. The child care facility features are selected using the select within source data option.
4. The distribution of childcare centers is spatially overlayed with a heatmap of the median household income and number of children below 5 years respectively.
5. I then assess the global spatial autocorrelation of below5 and Median Household Income.

6. Then, I perform a LISA analysis to analyze presence of spatial clusters in both variables.
7. Then, I perform an Ordinary Least Squares (OLS) Regression and plot the residuals to find spatial autocorrelation in them.
8. Lastly, I perform a Geographically Weighted Regression (GWR) and plot the residuals to look for spatial autocorrelation in them.

4.3 Huff Model - How does a new Child Care Facility Impact Stress on Already Existing Facilities?

To analyze the impact of a new child care facility on pre-existing facilities, the following steps were followed:

1. The resulting shapefile of step 3 in the previous section is imported into the environment as a layer.
2. 2 childcare facilities (CCF) (`Mary Walton CC, Hudson Guild CC`) are selected as samples for this analysis and exported to another layer. Two distant childcares are chosen to visually enhance the effect of a new CCF.
3. Using the edit tool, a new CCF is added to the layer containing the selected day care facilities.
4. The tract layer and selected day care facilities layer data are then projected into the UTM projection `NAD 1983 UTM ZONE 18N`.
5. In `View Dataframe Properties`, the distance unit is set to `kilometres`. A near table is then generated to calculate the proximity of each census tract to the selected CCF's.
6. The generated near table is then `split by attributes` into 3 files corresponding to the selected CCF's. These split files are imported into the environment and joined to the file imported in step 1 of this section. All irrelevant features are turned off and the distance attributes are renamed to `DIST_0, DIST_1, DIST_2`
7. The attraction values of each of the selected CCF's is calculated ($\texttt{attract}_i = \frac{(\texttt{below5} \times \texttt{CC Capacity})}{\texttt{DIST}_i^2}$) and entered as new attributes into the tract layer that contains all the joined data. Total old attraction is calculated and entered as a new variable as $\texttt{old\_attraction} = \texttt{attract\_0} + \texttt{attract\_1}$. Total new attraction is calculated and entered as a new variable as $\texttt{new\_attraction} = \texttt{attract\_0} + \texttt{attract\_1} + \texttt{attract\_2}$. Old and new probabilities of someone visiting each of the CCF's is calculated. $\texttt{old\_prob\_0} = \frac{\texttt{attract\_0}}{\texttt{old\_attraction}}$ and $\texttt{old\_prob\_1} = \frac{\texttt{attract\_1}}{\texttt{old\_attraction}}$, $\texttt{new\_prob\_0} = \frac{\texttt{attract\_0}}{\texttt{new\_attraction}}$ $\texttt{new\_prob\_1} = \frac{\texttt{attract\_1}}{\texttt{new\_attraction}}$ $\texttt{new\_prob\_2} = \frac{\texttt{attract\_2}}{\texttt{new\_attraction}}$. The changes in probabilities for CCF's 0 and 1 is now calculated as $\texttt{change\_0} = \texttt{new\_prob\_0} - \texttt{old\_prob\_0}$, $\texttt{change\_1} = \texttt{new\_prob\_1} - \texttt{old\_prob\_1}$.
8. All of these attributes are set to be viewed as percentages represented by a calculated fraction.

5 Conclusions

Reported as per each section of the project.

1. Fractal Dimensions -
   - The fractal dimensions are found to remain mostly the same over the decade. However, the fractal dimension for non-commercial area decreases in 2012 by 0.01.
   - Through visual analysis, we can observe commercialization in Hudson Yards, North Harlem and South Manhattan. Roosevelt Island gets converted from almost totally commercial to almost no commercial buildings. There were asylums on the island that have now been shut down. This change could be because these were technically considered as commercial.
2. Spatial Autocorrelation, LISA, OLS & GWR -

- When the relationship between median household income and number of children below 5 years old in a census tract is assessed using:
  - Global Autocorrelation: Moran's I value for median household income is higher than that for children below 5 years old.
  - LISA: Most of the Low-High, High-High, High-Low, and High-High clusters remain the same for both the variables.
  - OLS: On performing OLS and plotting residuals, an R^2 score of 0.003 is achieved. The residuals of the regression, on visual analysis, show spatial autocorrelation.
  - Thus, Geographically Weighted Regression was performed to account for the spatial autocorrelation in the variables. The residuals of this regression on visualization don't show spatial autocorrelation. However, the R^2 value improves only to 0.3. The rest of the information of number of children in a neighborhood would be explainable by other variables.
- Spatial Interaction Model -
  - Incorporating the number of children below 5 per census tract in the calculation of attraction to a selected CCF doesn't have much impact as this is scaled by a larger variable `distance`^{2} in the denominator.

6 Figures

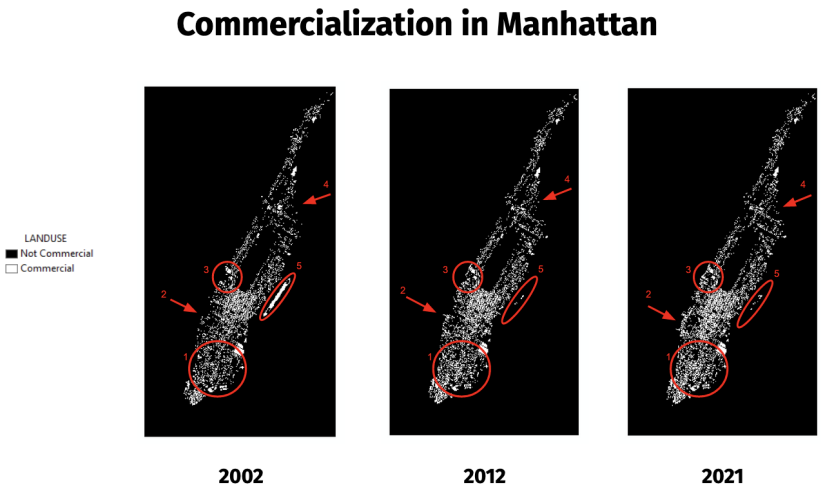


Figure 1: Commercialization

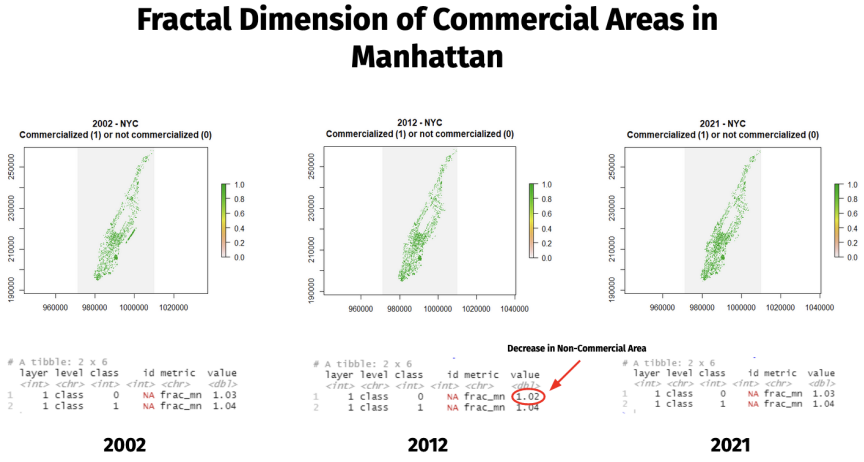


Figure 2: Fractal Dimensions

Figure 3: Heatmaps of Variables and CCF distribution

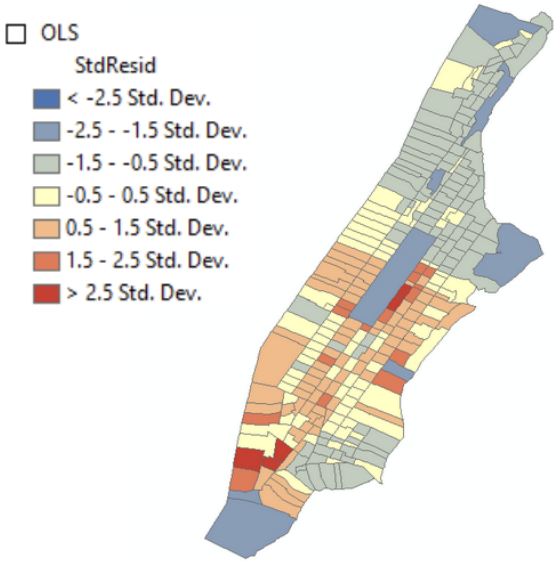


Figure 4: OLS

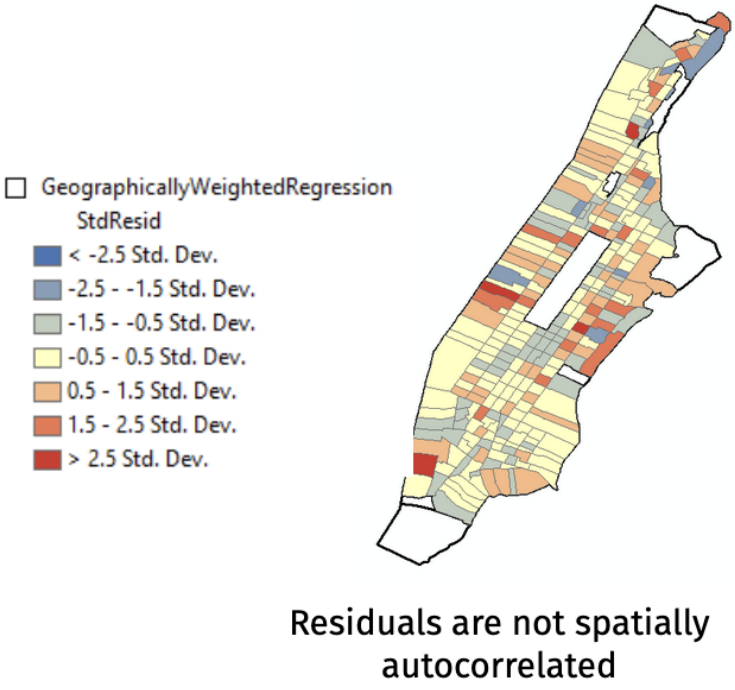
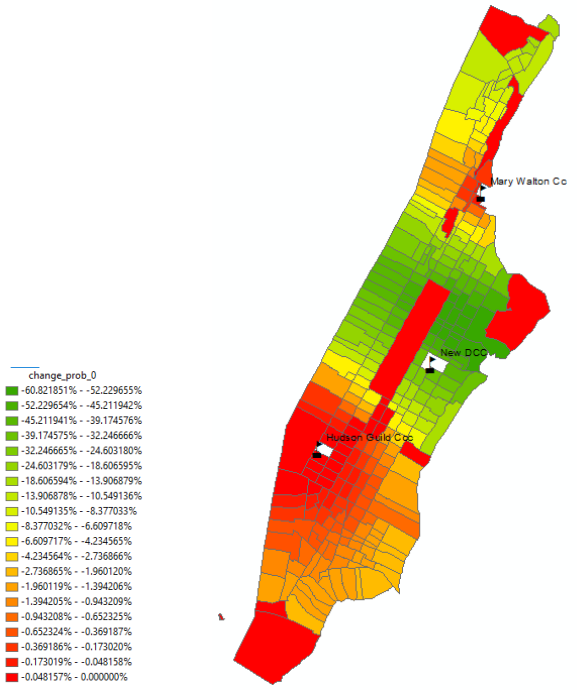
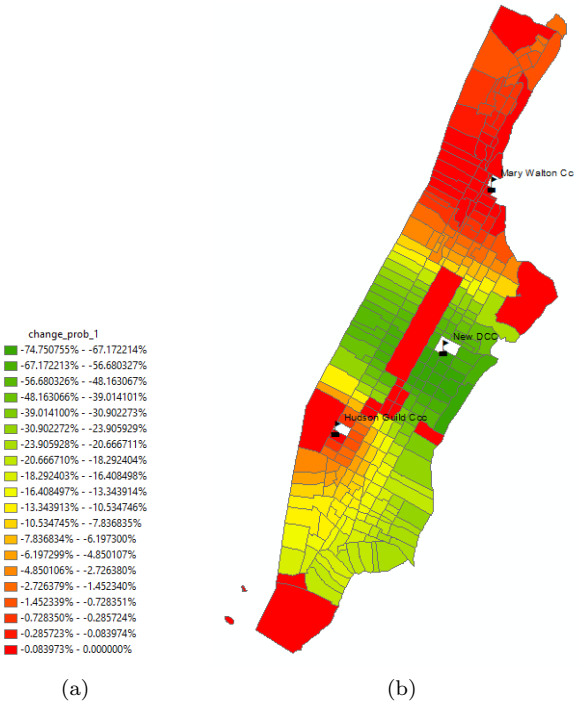


Figure 5: GWR



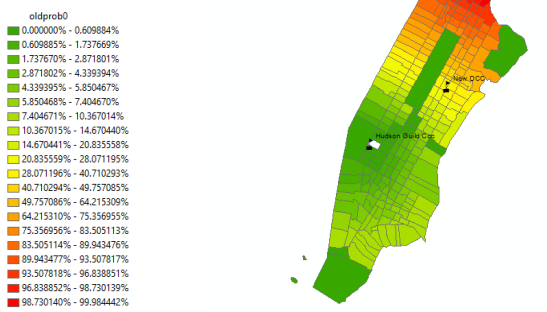
(a) (b)

Figure 6



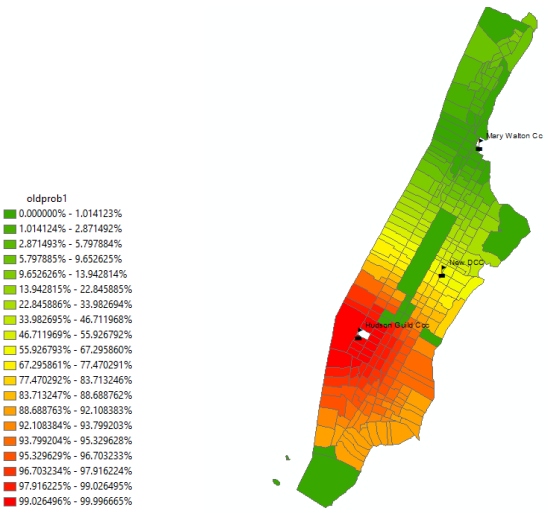
(a) (b)

Figure 7



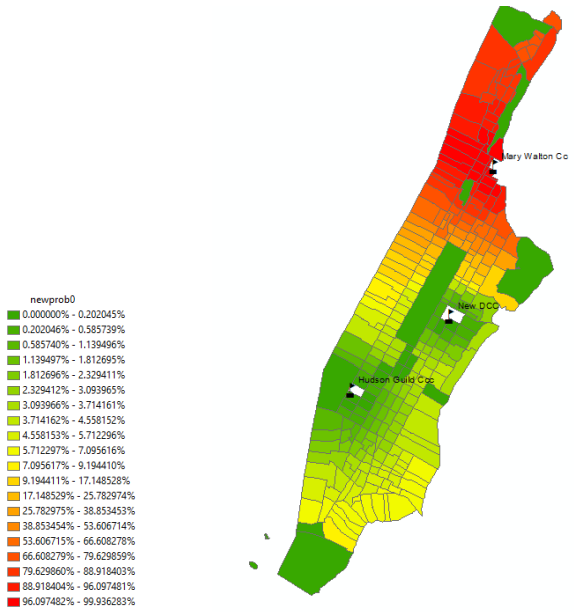
(a) (b)

Figure 8



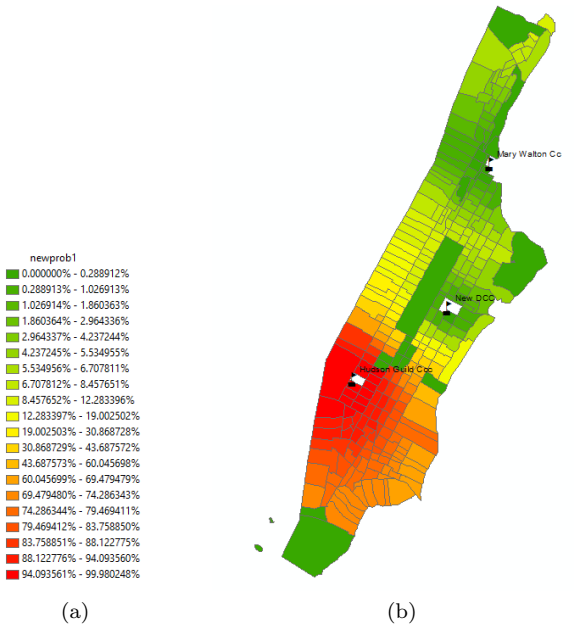
(a) (b)

Figure 9



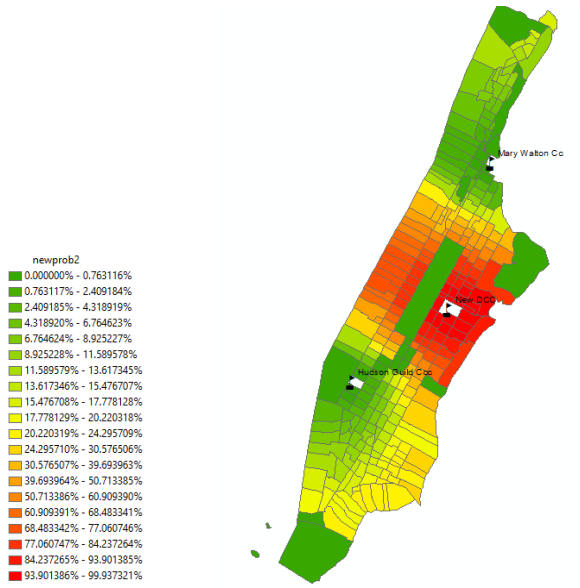
(a) (b)

Figure 10



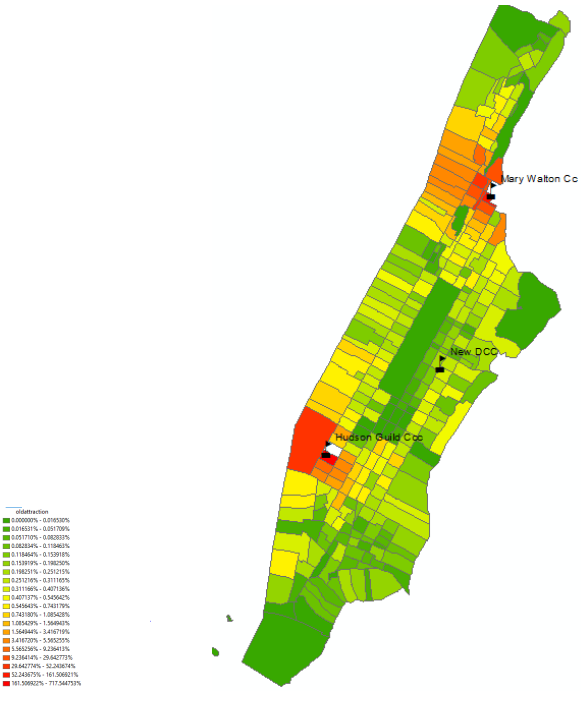
(a) (b)

Figure 11



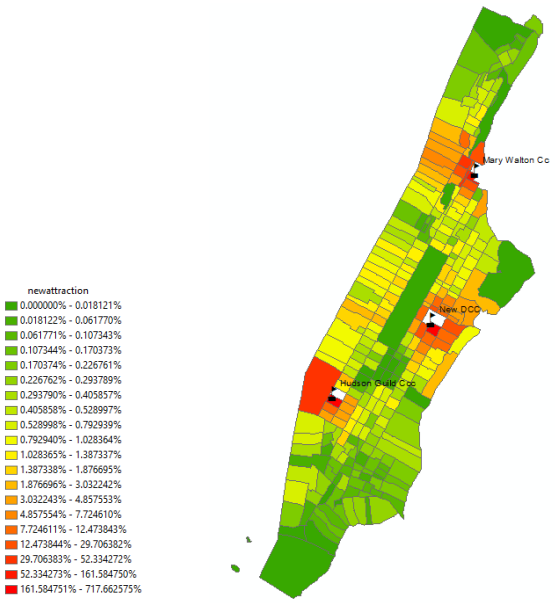
(a) (b)

Figure 12



(a) (b)

Figure 13


newattraction
0.000000% - 0.018121%
0.018122% - 0.061770%
0.061771% - 0.107343%
0.107344% - 0.170373%
0.170374% - 0.226761%
0.226762% - 0.293789%
0.293790% - 0.405857%
0.405858% - 0.528997%
0.528998% - 0.792939%
0.792940% - 1.028364%
1.028365% - 1.387337%
1.387338% - 1.876695%
1.876696% - 3.032242%
3.032243% - 4.857553%
4.857554% - 7.724610%
7.724611% - 12.473843%
12.473844% - 29.706382%
29.706383% - 52.334272%
52.334273% - 161.584750%
161.584751% - 717.662575%
Mary Walton Cc
New DCC
Hudson Guild Cgc

(a) (b)

Figure 14