Economic Map of New York City

Final Report

New York University

Center for Urban Science + Progress

Andrea Kanner

Awais Malik

Kara Leary

July 28, 2014

Contents

List of Figures

1 Introduction

The goal of our project was to develop an economic profile of New York City. This project was part of a larger effort at CUSP to provide New York City agencies with analysis that facilitates better understanding of the City's economic condition. The success of this project required integration of the Primary Land Use Tax Lot Output (PLUTO) data with other sources such as the U.S. Census and American Community Survey (ACS), as well as data from relevant New York City Agencies including the Departments of Finance, and Parks and Recreation. Our group created an interactive visualization tool, which could assist City agencies in achieving greater understanding of the key drivers of economic prosperity in the City.

The initial stages of the project involved building the base layers of a comprehensive platform for analysis and visualizations of economic data through data cleaning and integration. The analysis stage began with simple geo-spatial correlations, and progressed towards area classification and correspondence, with the primary objective being to identify "main street" metrics and spatial patterns of equity. Our group also engaged in research on the analysis aspects of the project: summary statistics, metric correspondences, and neighborhood feature vectors for classification. Throughout the process, we were also involved in visualizations including simple plotting, interpretation, and map making. Our final step was to create an interactive web-based visualization tool, which highlights some of the economic drivers the govern New York City.

Future steps include enabling the identification of "like" locations based on demographic and economic parameters of interest. Once the economic map is complete, an example of its application would be to evaluate the effect of proximity to parks on real estate market value and business turnover. Similarly, one could also analyze the economic effects of closeness to a Business Improvement District (BID) on a particular neighborhood, or the factors that make a BID more or less successful. We hope our product will assist City agencies in achieving a better understanding of the economic needs and assessments of New York City.

2 Background

The New York City Economic Development Corporation (NYC EDC) published an "Economic Snapshot" of the city in August 2011,[1] with particular attention paid to the impact of parks on property values within the city. While this document serves as a good starting point and provides valuable insight into the topic at hand, we are aiming to produce results that are more granular and broader in scope than the NYC EDC study.

The NYC EDC study focuses mainly on high-level numbers such as the 29,000 acres of land controlled by the Parks Department (just under 1/5 of all City land), and how New York City's parks rank against those in other large and similarly dense US cities (seven of the top 50 visited parks in the country are in New York City), but does little to provide the kind of 'data-driven' analysis we are working on. For example, the study broadly states that "land values of residential properties increase the closer they are to a park," and discusses the difference between land values "between a 5 and 10 minute walk" and those "within 5 minutes" of Central Park's East side, without providing an explanation of how walking times were determined. It also fails to consider the impact of smaller or medium-sized parks, paying attention primarily to the effects of Central Park, Prospect Park, and the High Line, which are unusually large compared to the rest of the parks of New York City as a whole, and thus do not provide the same insight which we are hoping to achieve (see Figure 1).

The NYC EDC study provides qualitative and higher-level information without delving too deeply into the quantitative details, which leaves much room for new analysis on our part. Other relevant studies have also taken place, such as one conducted in Portland, Oregon, which found that the addition of public parks has a "positive and statistically significant effect" on property values and varies with distance from the park.[2] Another study, focused in New York City, looks specifically at the impact of community gardens and incorporates garden quality as a key factor.[3] Overall, there is much to learn from existing work but also a good deal of unchartered territory that we aim to explore.

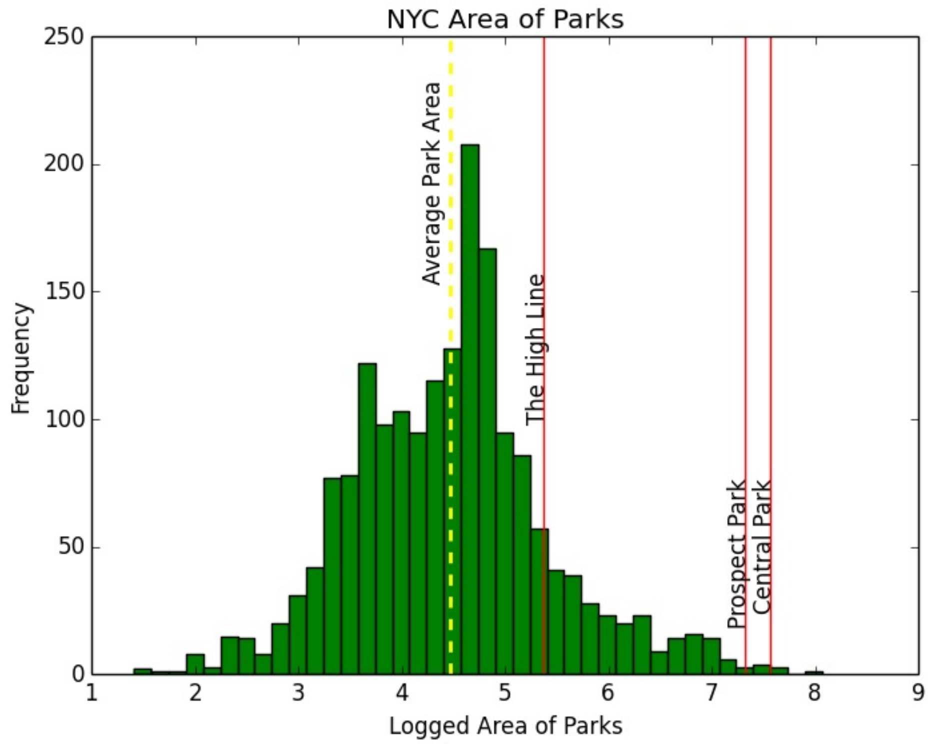

Figure 1: Histogram of New York City Parks by Size

3 Methodology

3.1 Data Sources

All of the data used in this project is open and readily available to the public. Datasets and their sources are listed below:

- PLUTO (Primary Land Use Tax Lot Output): NYC Department of City Planning
- Park Boundaries: NYC Department of Parks & Recreation
- Park Quality: NYC Department of Parks & Recreation
- 2010 Census: US Census Bureau
- 2008 - 2012 American Community Survey: US Census Bureau
- Zip Code Business Patterns (ZBP): US Census Bureau
- Tax Lot Market Value: NYC Department of Finance

3.2 Metrics

In looking at creating an economic map for the city, we looked at features found in or derived from the Zip Code Business Patterns (ZBP). At the zip code level, these features included total number of employees, total number of establishments, and annual payroll. Rough estimates of employees per establishment and average annual salary were calculated using a combination of the features provided. These were selected based on data availability and their ability to provide insight into the overall economic condition of a given zip code. It is important to note that these numbers are reflective of the people who 'work' in a zip code, as opposed to those features given for people who 'live' in a zip code in the US Census and ACS data.

In looking specifically at the impacts of parks, the metrics used included distances to the nearest park, park quality and size, market values for residential tax lots, median household income, and percentage of land in a given zip code covered by parks. ZBP data described above was also included in this analysis, with the addition of a ratio of "small" to "big" businesses in an area, where small businesses are defined as those having fewer than 250 employees. This metric was chosen with the idea that the presence of small businesses in a locality is a positive economic driver.

3.3 Analytical Methods

3.3.1 Building an Economic Map of the City

With the initial goal of creating an overall economic profile of the city in mind, we set out to determine what this might look like. Looking at the years 2000 - 2012, a time series was created of the data for each zip code with the goal of determining if there were any zip codes that exhibited similar time-like behavior or trends. Each time series has been mean-subtracted and variance-normalized because the data is going to be at least partially dependent on the size of the zip code within which it falls. With these normalized vectors, we can look at similar relative trends as opposed to being distracted by absolute numbers.

In order to find these similar areas, unsupervised learning and k-means clustering techniques were used. A fallacy of k-means clustering is that the number of clusters is arbitrary and left to the discretion of the analyst, and for this reason both two and five clusters were analyzed. Further discussion of the results is included in the "Analysis" section below.

The presence of a lot of parks in an area could be reasonably argued to have either a positive or negative impact on local business. On the one hand, park presence is likely to bring more pedestrians to an area and so may benefit local businesses. However, park presence also takes up space that could have otherwise been occupied by stores and restaurants. With this in mind, the Census ZBP data was analyzed further to determine if any relationship existed between the various metrics and the percentage of a zip code covered by park land. ArcGIS's geo-processing tools were used to determine the percent of a zip code that was covered by parks. Median household income was also looked at in this way. The scatter plots produced, although very basic, provide a good starting point for future work. They are included and discussed in detail in the "Analysis" section.

3.3.2 Impact of Parks on the Local Economy

In order to evaluate the distance from each residential tax lot to the nearest park, we used the PLUTO data and found the minimum distance from a lot belonging to the land use category of 1, 2, or 3 (residential one, two, multi family homes) to a lot in land use category 9 (Open Space and Outdoor Recreation). We were also able to merge the PLUTO data with the Market Value data for each BBL (Borough Block Lot): a unique ten digit identification number for each tax lot in New York City.

Once these two datasets were joined, we were able to analyze the variation in Market Values for homes as distance to the nearest park is varied. The initial Euclidean distance to the park was not an accurate calculation, since the method was using the centroids of the parks and tax lots. Also, not every lot that was in the land use 9 category was relevant to our analysis (e.g. cemeteries). From the NYC Parks and Recreation data file, we were able to extract the vertices, and calculate the distance from the center of the tax lot to closest park vertex. However, this method was also not accurate since it only used the corners of the park lots. Therefore, we calculated the distances from each residential lot to the edge of the nearest park and removed the irrelevant park types (Cemeteries, Waterfront Facilities, Buildings/Institutions, Historic House Parks, and Undeveloped). We made a data set for each borough, with all the distances from every PLUTO lot in the borough to the edge of every park in the city. Merging the distance data set with the parks and PLUTO data makes it computationally easy to look at certain buildings and/or certain parks and get their minimum distances.

This more accurate distance calculation method showed that the distribution peaked at a lower distance than we had seen with the first calculations (see Figure 2 for Brooklyn). For the denser boroughs (Manhattan, Brooklyn and the Bronx), the peak distance is less than 500 feet; while in the sparser boroughs (Queens and Staten Island), it is less than 1000 feet. We also looked at the minimum distances by weighting the bins by the number of units in each building. The distributions were similar, but more noisy. We also observed the distribution of Market Values/Square Foot in each Borough. Most Boroughs had a nice distribution curve with the peaks around $200 - 300 per square foot, except for Manhattan which was very noisy with no clear trend.

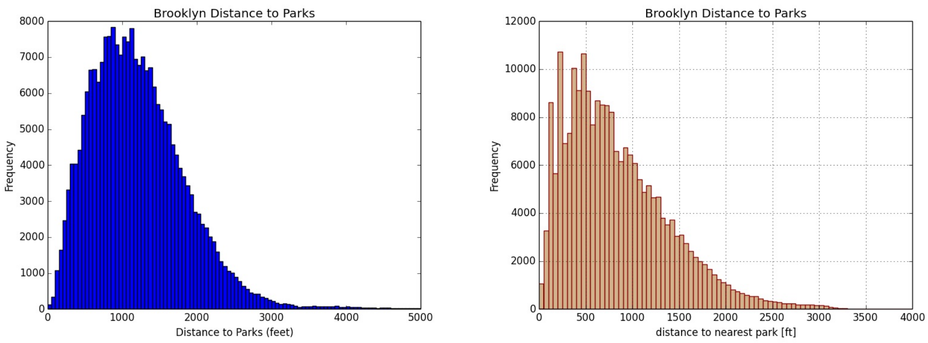

Figure 2: Brooklyn Distances to the nearest Park (Old Method (left) & New Method (right))

Now that the groundwork had been laid and all the data was in an accessible and cohesive form, any calculations that need to be made in the future can easily be done. For instance, one can get the distance from any types of park to any building types with certain characteristics by zip code or Borough. It is therefore very easy to see how the distribution of distances to parks in a given zip code is similar or dissimilar to its borough as a whole. For instance, Figure 3 (Zip Code 11219 - Borough Park) is very different from the rest of Brooklyn, and in general much farther from parks. Figure 4 shows that the characteristic distances to parks in the outer Boroughs is much greater than that of the central Boroughs.

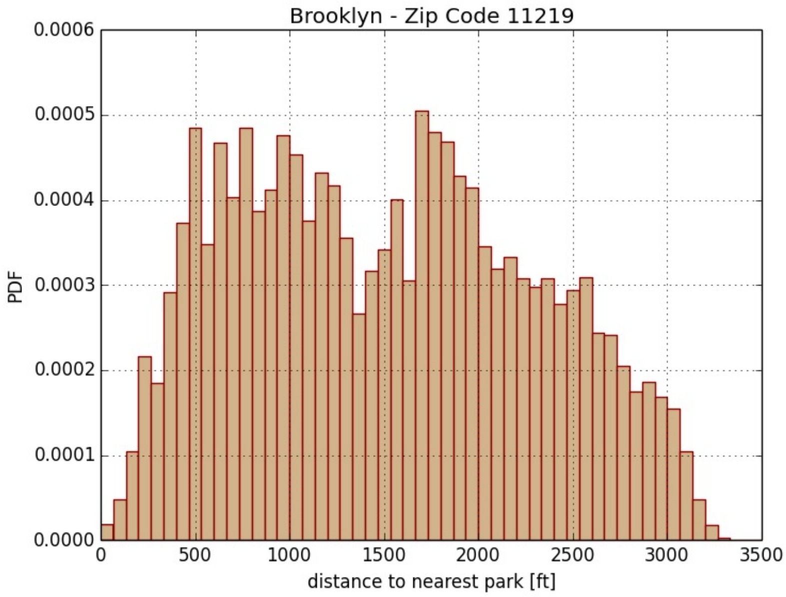

Figure 3: Brooklyn Zip Code 11219 - Borough Park

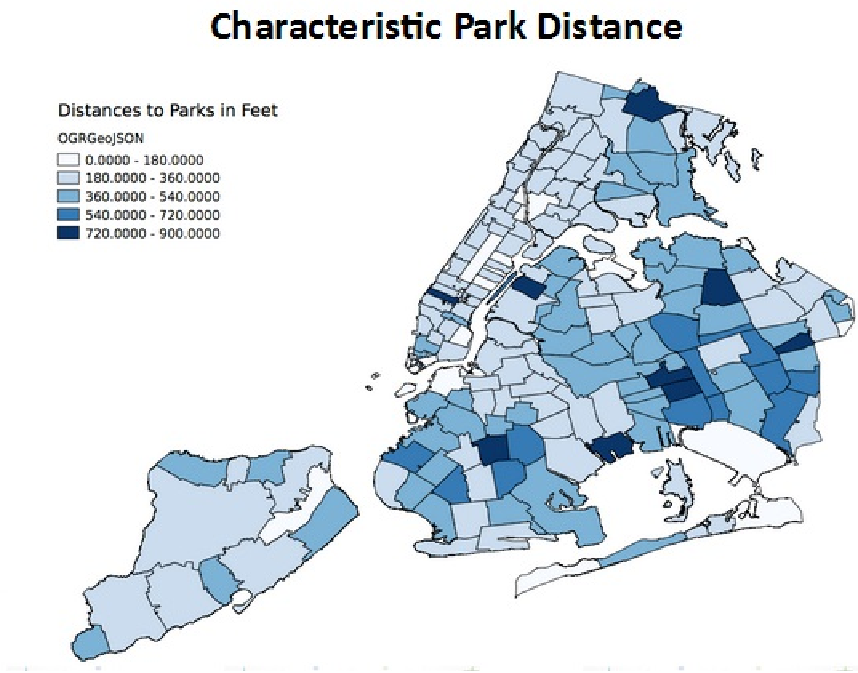

Figure 4: New York City Characteristic Distances to Parks

Park quality measures were also included in our analysis. The data available left much to be desired, and came in the form of a rating of either "acceptable" or "unacceptable" for both "Overall Cleanliness" and "Overall Quality" of a park. Unfortunately, the vast majority of ratings were "acceptable," and did not provide too much insight. The ratings were combined to create a trinary score leaving each park with a score of 0.0, 0.5, or 1.0 based on a combination of the two ratings. These trinary scores were combined with the area of each park to create an area-weighted average of the overall quality of the parks within any given zip code. The following equation explains this calculation:

$$Q_{ave} = \frac{\sum_i Q_i * A_i}{\sum_i A_i}$$

where Q is the park quality, and A is the area.

4 Analysis

4.1 Static Visualizations

Initial attempts at creating an economic profile of the city are demonstrated in Figures 5 - 11 below. In the plots, which show the output of the k-means clustering as described above, each gray line represents the normalized time series of the data for a specific zip code. Each colored line represents the average behavior or overall trend for the zip codes that behaved like that cluster, and corresponds to the coloring of the map included on the right. You will notice a good amount of covariance in some of the five-cluster results, particularly in that of total number of employees, which gives more motivation to exploring a two-cluster alternative.

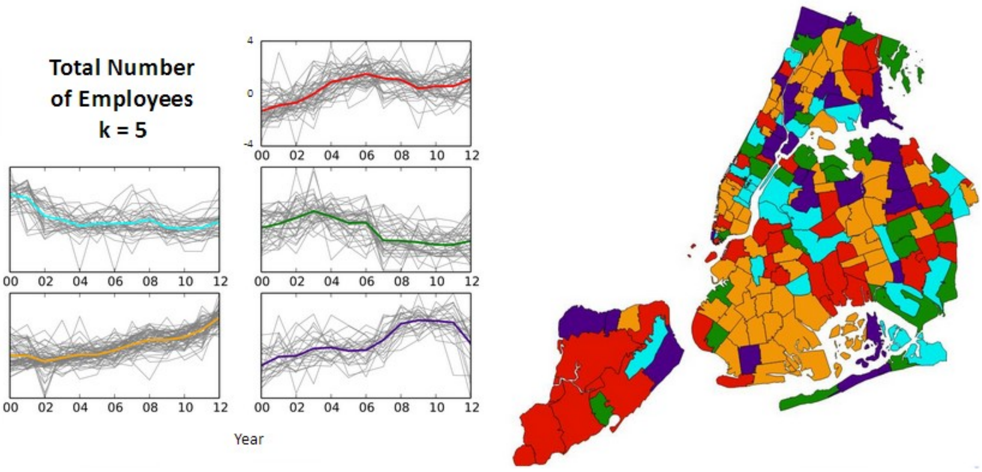

Figure 5: Total Number of Employees (k = 5)

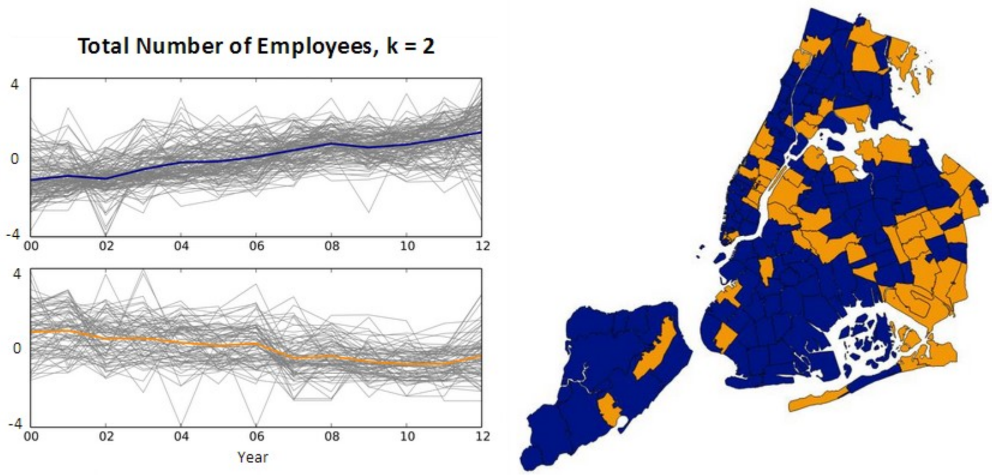

Figure 6: Total Number of Employees (k = 2)

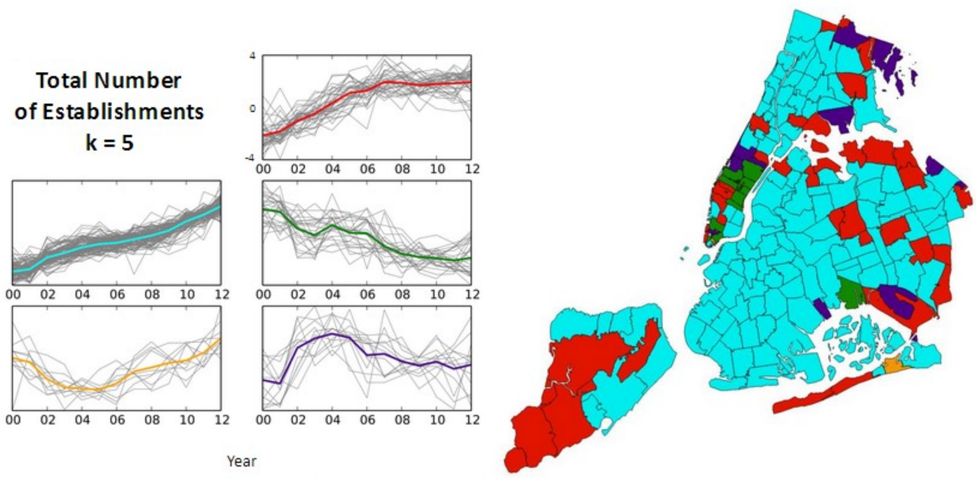

Figure 7: Total Number of Establishments (k = 5)

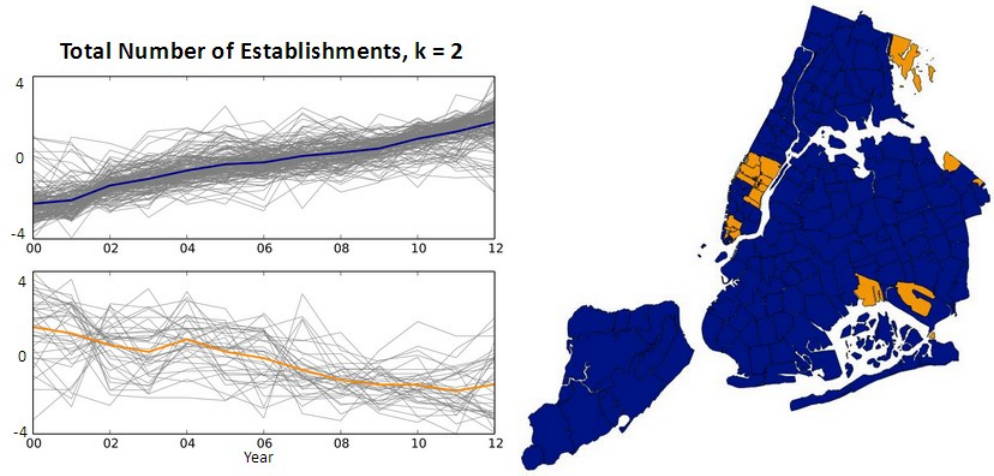

Figure 8: Total Number of Establishments (k = 2)

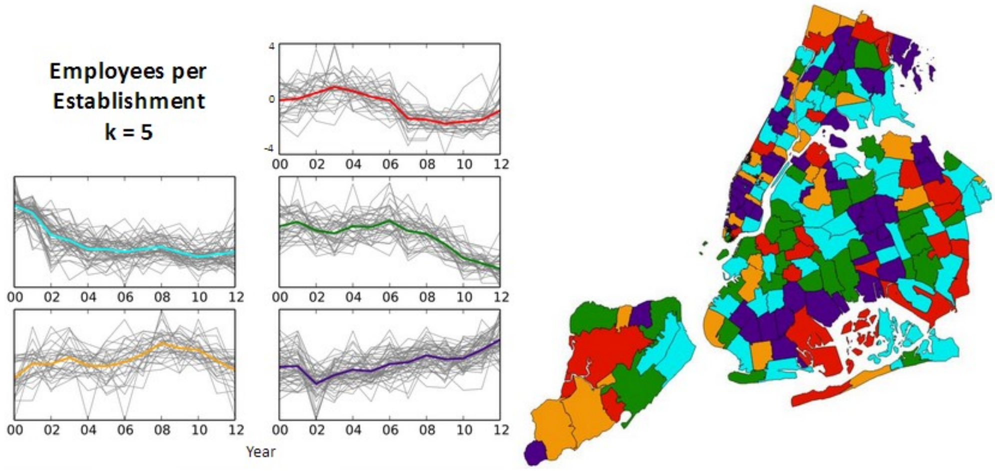

Figure 9: Employees per Establishment (k = 5)

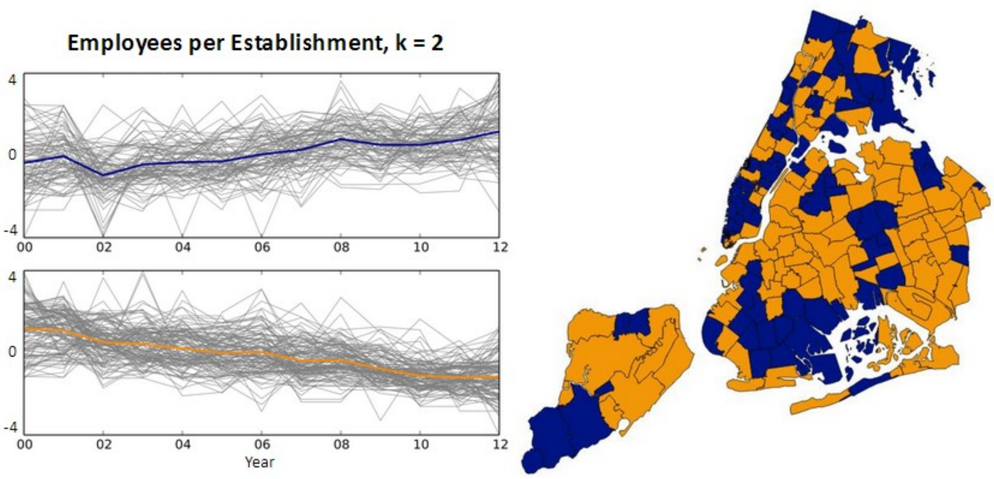

Figure 10: Employees per Establishment (k = 2)

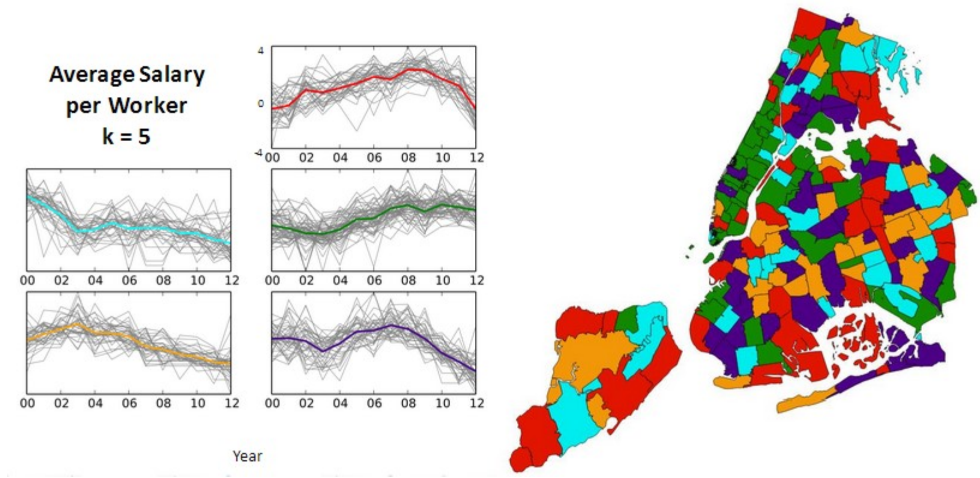

Figure 11: Average Salary per Employee (k = 5)

Looking at the results, it is clear that there are definite neighborhoods or areas of the city that behave in similar temporal ways. The plots of total number of employees and establishments show that midtown Manhattan appears to be either stagnant or slightly decreasing. This might come as a surprise to anyone familiar with midtown, as it is a major hub of business for the city as a whole. However, it may be that because midtown has been a well-established business district for years it was already saturated with both employees and establishments by 2000, leaving little room for further growth. The total number of establishments appears to be increasing for the overwhelming majority of the city, particularly in the outer boroughs, which is consistent with what we observe as New Yorkers every day. Looking at average salary in two clusters did not provide any meaningful insight, and thus has been omitted from this report.

Overall, k-means clustering on the ZBP data has provided a good first step towards understanding what a comprehensive economic map of the city might look like. Unfortunately, time and data limitations halted our progress at this point as we pivoted towards our first use case of the economic effects specifically made by the presence and quality of parks and open spaces throughout the city.

The plots of ZBP data versus percentage of a zip code covered by parks can be seen in Figures 12 to 14 below. Only two zip codes are almost 100% covered by parks, corresponding to the zip codes containing Central Park and Randall's Island. We can consider the motivating question of whether increasing the number of parks inversely affects the total numbers of employees and establishments. Those points circled in red seem to suggest that perhaps increased percentage of parks does negatively affect business, as they all fall towards the center of the x-axis and leave a noticeable void in the upper right corner of each plot. However, further investigation reveals that these points all correspond to zip codes in midtown Manhattan, and those zip codes lie near or adjacent to Central Park. Thus, these results cannot be taken at face value and are inconclusive.

The same goes for the plots of annual payroll and ratio of small to big businesses. The plot of percent of a zip code covered by parks against median household income shows a few interesting points, again circled in red, that correspond to Manhattan zip codes in the Upper East Side and Upper West Side, and Tribeca. A plot of the area-weighted average park quality against median household income as well as a map of the area-weighted average are shown in Figures 14 and 15. The results may suggest a slight relationship between the two metrics, but future analysis and perhaps better park quality data will be needed to investigate further.

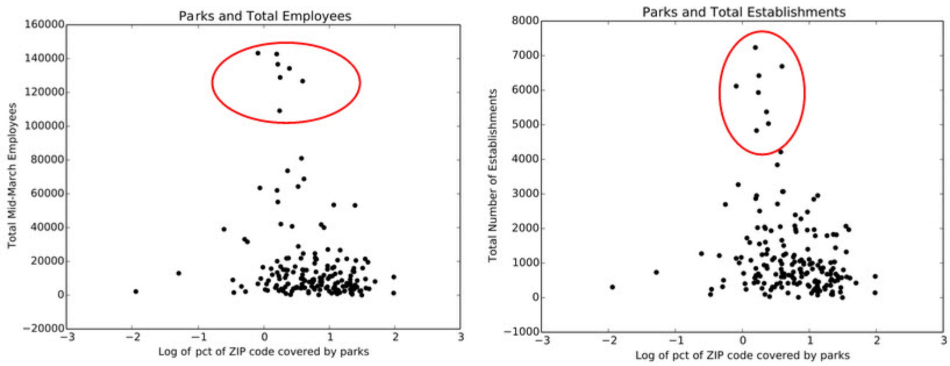

Figure 12: Parks vs Total Employees (left) and Parks vs Total Establishments (right)

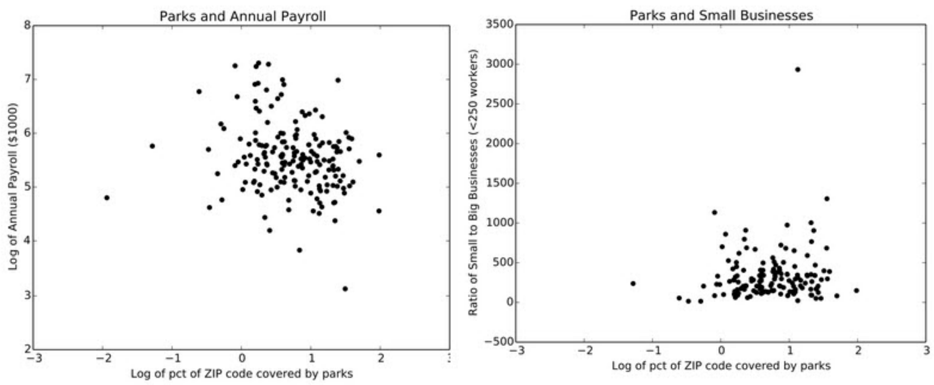

Figure 13: Parks vs Annual Payroll (left) and Parks vs Small Businesses (right)

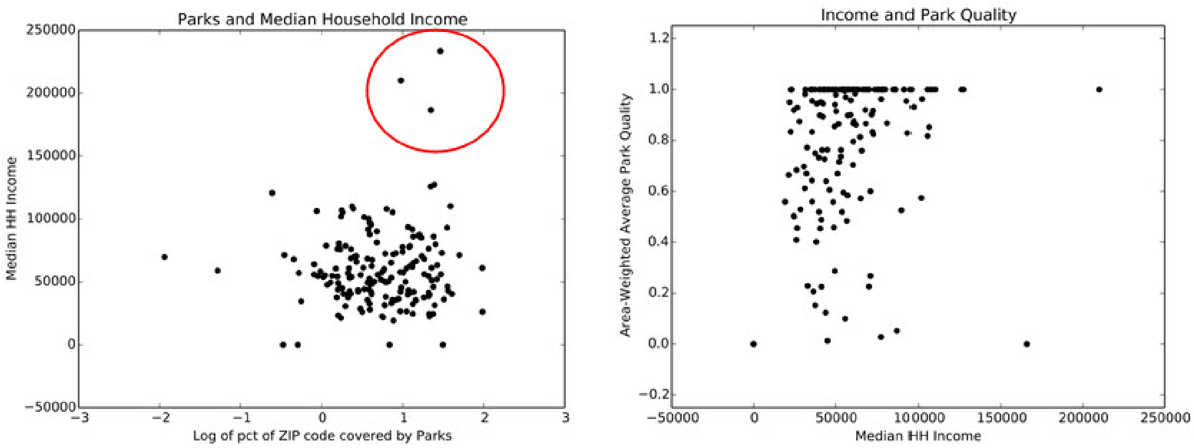

Figure 14: Parks vs Median Household Income (left) and Median Household Income vs Park Quality (right)

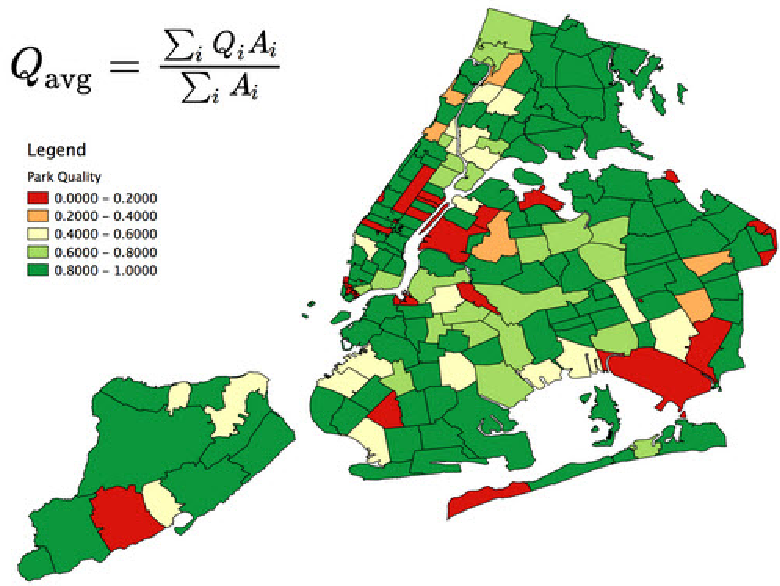

Figure 15: Map of Average Park Quality

4.2 Interactive Visualization

We used JavaScript, specifically D3.js, to create an interactive web-based visualization which would highlight some of key parameters that affect or represent economic activity in the city. Figure 16 is a screen shot of the interactive map with the mouse (invisible here) placed over zip code 10004 in Lower Manhattan. The key attributes of that zip code are highlighted. The heat map of the zip codes is color coded according to the percentage of zip code covered by parks (lighter green means higher percentage of zip code is covered by parks). Future improvements to this tool would include allowing the user to click on the zip code, and additional statistics related to that zip code will appear in the form of graphs (scatter plots, histograms etc). Another important feature which needs to be incorporated in this tool is the ability to lasso a region and have similar regions (based on our feature vectors) be highlighted for comparison purposes.

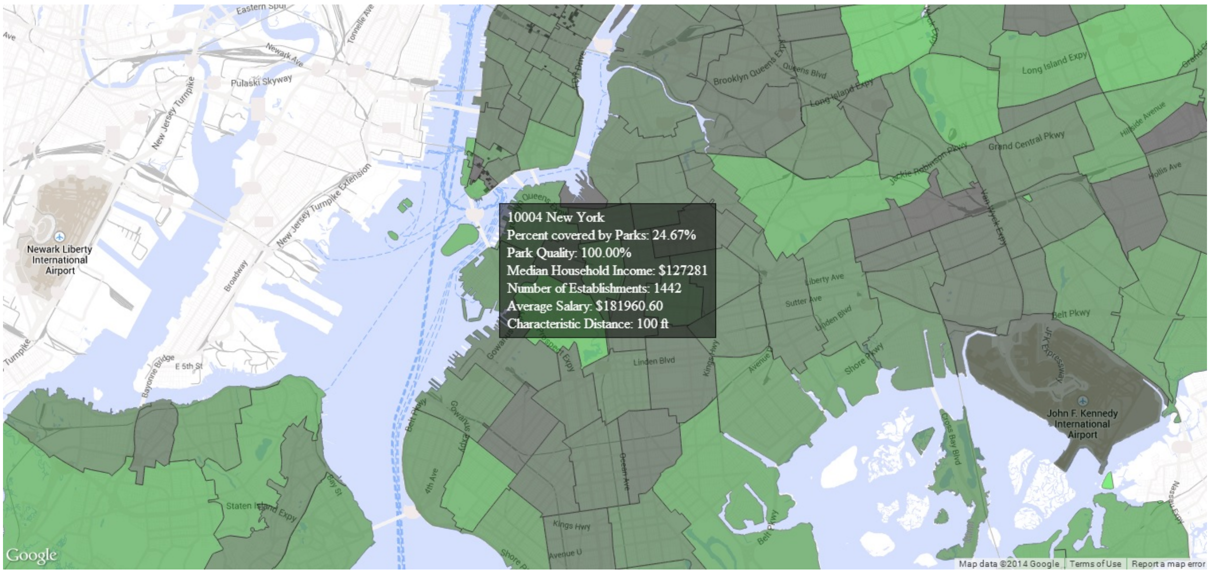

Figure 16: Screen shot of Interactive Web-based Economic Map of the City

4.3 Errors & Uncertainties

Now that we were able to get accurate park distances, we can easily see the distance from every lot in a borough or zip code to the closest park. But that does not account for how many people live in that lot. We did look at a histogram of the distances with the density being the number of units per lot, but that does not necessarily capture the dynamic population, because it is in essence considering a fixed number of people per unit.

Improvements can be made to the k-means clusters of ZBP data through employing algorithms to determine the optimal number of clusters. The calculation of average salary is a very rough one, and better data would likely contribute to a more comprehensive analysis. A more complete dataset of park quality would also greatly improve our analysis and likely allow us to gain much more insight into the relationships between the various metrics and park quality.

5 Conclusion

This project has the potential to have a significant impact in New York City, particularly for the Department of Parks and Recreation. Although future analysis will be necessary, the Parks Department may be able to use this work to make the case for increased funding to improve park access and quality. This may particularly be the case in the outer boroughs, which have been shown to have disproportionate access to parks and open spaces compared to Manhattan. Our work, in particular the web-based interactive visualization, has the potential to be used as a tool for real estate as it could provide meaningful insight to those looking to move to or relocate within New York City who may be concerned about access to open spaces. Down the road, a more complete economic profile of the city as a whole will hold great potential for impact to the city. This can be seen preliminarily with the ability to identify regions of the city that are growing or changing in similar ways. The potential to expand on this simple concept will create new ways to understand the city's dynamic growth.

References

[1] S. Huynh, U. Kaul, F. Peach, and E. Tumalad, "A summary of new york city's economy," *Economic Snapshot*, August 2011.

[2] B. Bolitzer and N. R. Netusil, "The impact of open spaces on property values in portland, oregon," *Journal of environmental management*, vol. 59, no. 3, pp. 185–193, 2000.

[3] I. Voicu and V. Been, "The effect of community gardens on neighboring property values," *Real Estate Economics*, vol. 36, no. 2, pp. 241–283, 2008.