Spatial Trends in Parking Traffic in North America

Sahil Singh, Dohyun Lee

Yale University – ENV 781 / S&DS 674

Introduction

The purpose of this study is to investigate spatial patterns and trends of areas within a city where drivers have trouble finding parking in North America at the continental, country, and city levels. Our dataset (Shin) contains aggregated parking statistics dating from April 2020 to October 2020 with information on the specific parking location represented as point data with latitudinal and longitudinal coordinates and variables such as average time taken to search for parking, total number of drivers searching for parking, and percentages of different types of vehicles with parking issues. Additionally, only cities with a population of more than 100,000 people are included.

In our study, we utilize R and a Google Maps API to create data visualizations, such as barplots, boxplots, and histograms, and implement spatial autocorrelation, clustering, and kriging to geospatially visualize and elicit insights on parking difficulty in North America during the initial onset of COVID-19. Given that we are interested in the time it takes a driver to find parking in a given area, the Null Hypothesis driving our study is that high parking times are randomly distributed, which is tested using Global Moran's I and Geary's C.

Key Questions

- What will mapping parking times at the continental, country, and city levels reveal? Is it harder to find parking in one country versus another or one state versus another?
- What cities are the most prevalent in the dataset? How is the data for these points distributed and what are their means and variances?
- Are high or low parking times random chance occurrences and can spatial predictions be made given our data?

Methods and Results

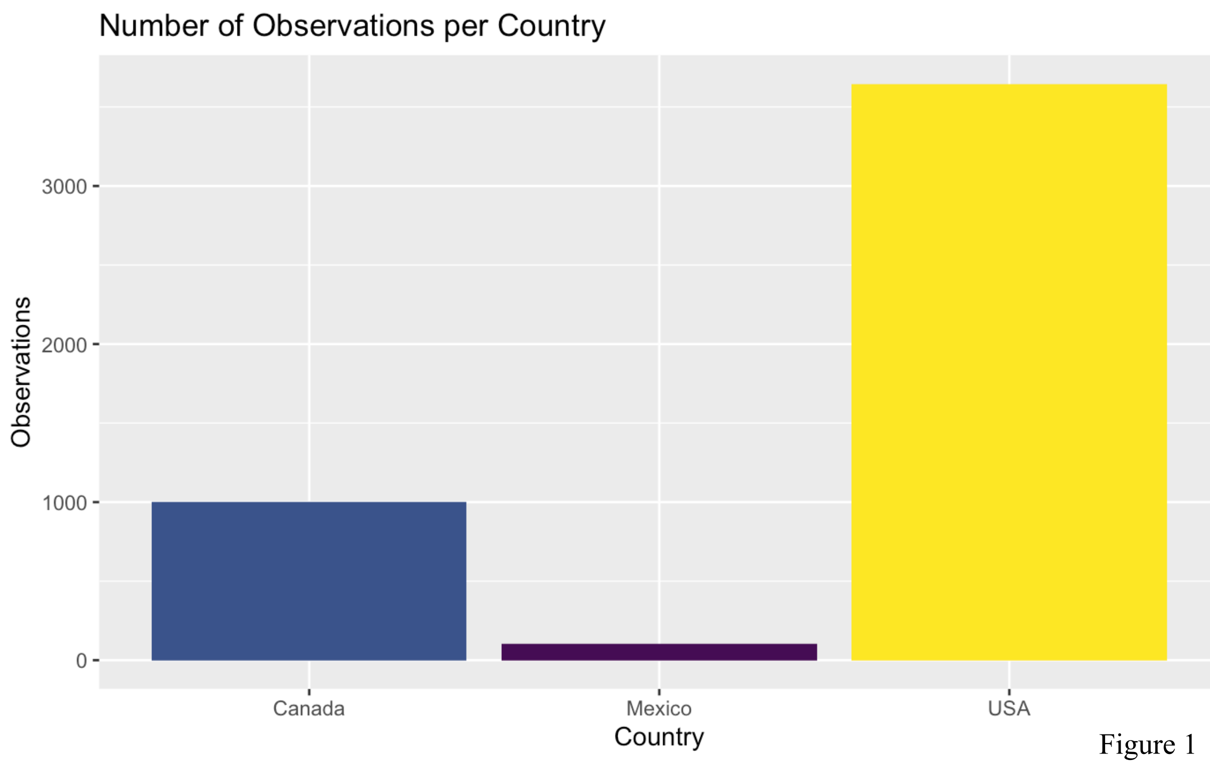

Figure 1

After plotting a barplot of the number of occurrences by country (Figure 1), we can see that Mexico does not have many recordings (only 101), so it might not be interesting to compare those observations with locations in the US and Canada that have 3647 and 1002 observations, respectively. Even though Canada has 1002 observations, the number of US observations far outweigh the Canada ones, so it still might not be very insightful to compare parking difficulty between those two countries. Figure 2 below shows an interactive Leaflet plot that maps every location in the dataset and with this we spatially visualize the information from Figure 1. Since we can’t include a dynamic graphic in a static PDF file, Figure 2 is a screenshot of what you would see if you hover your cursor over any of the points, in which a pop-up will display the county, city, state, country, coordinates, the total number of people searching, and the average time to park.

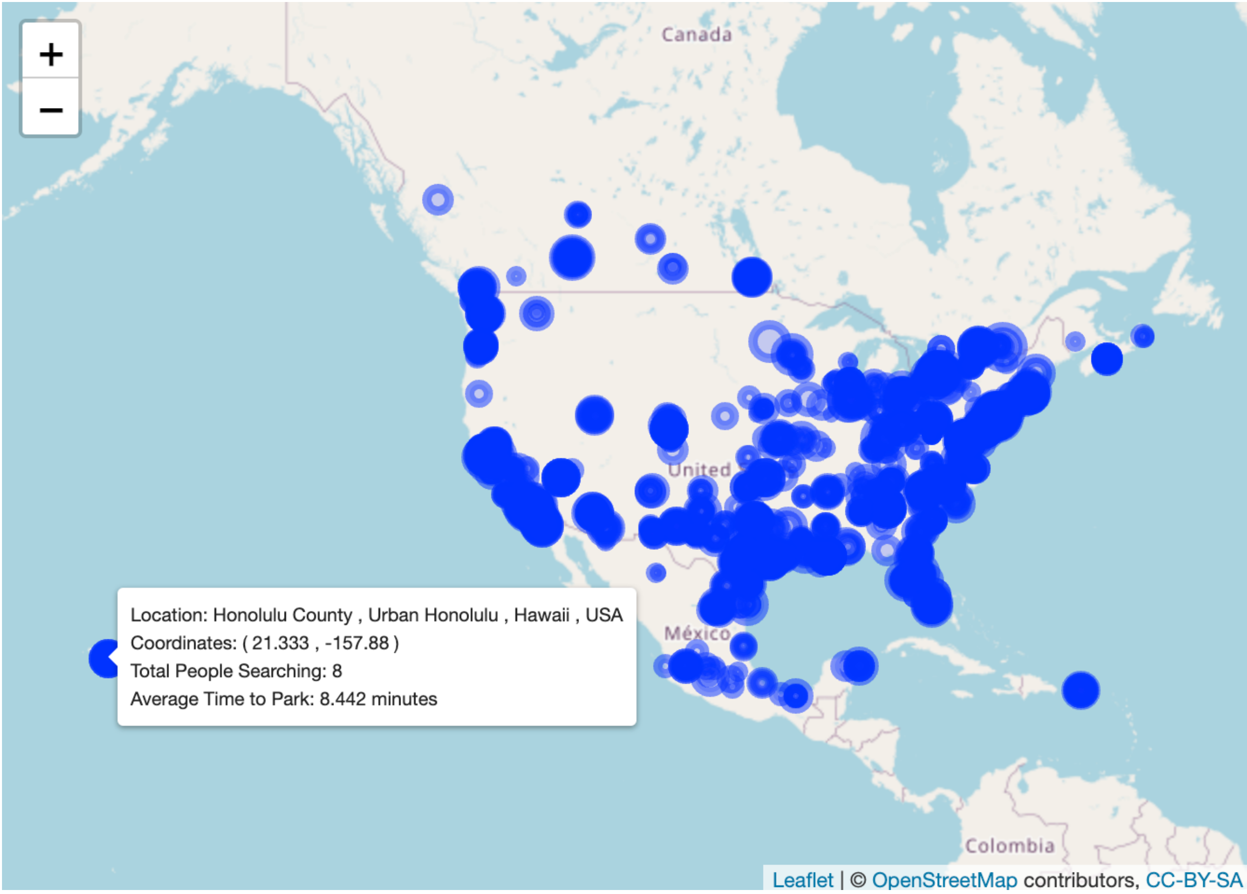
+
−
Canada
United
México
Colombia
Location: Honolulu County , Urban Honolulu , Hawaii , USA
Coordinates: (21.333 , -157.88)
Total People Searching: 8
Average Time to Park: 8.442 minutes
Leaflet | © OpenStreetMap contributors, CC-BY-SA

Figure 2

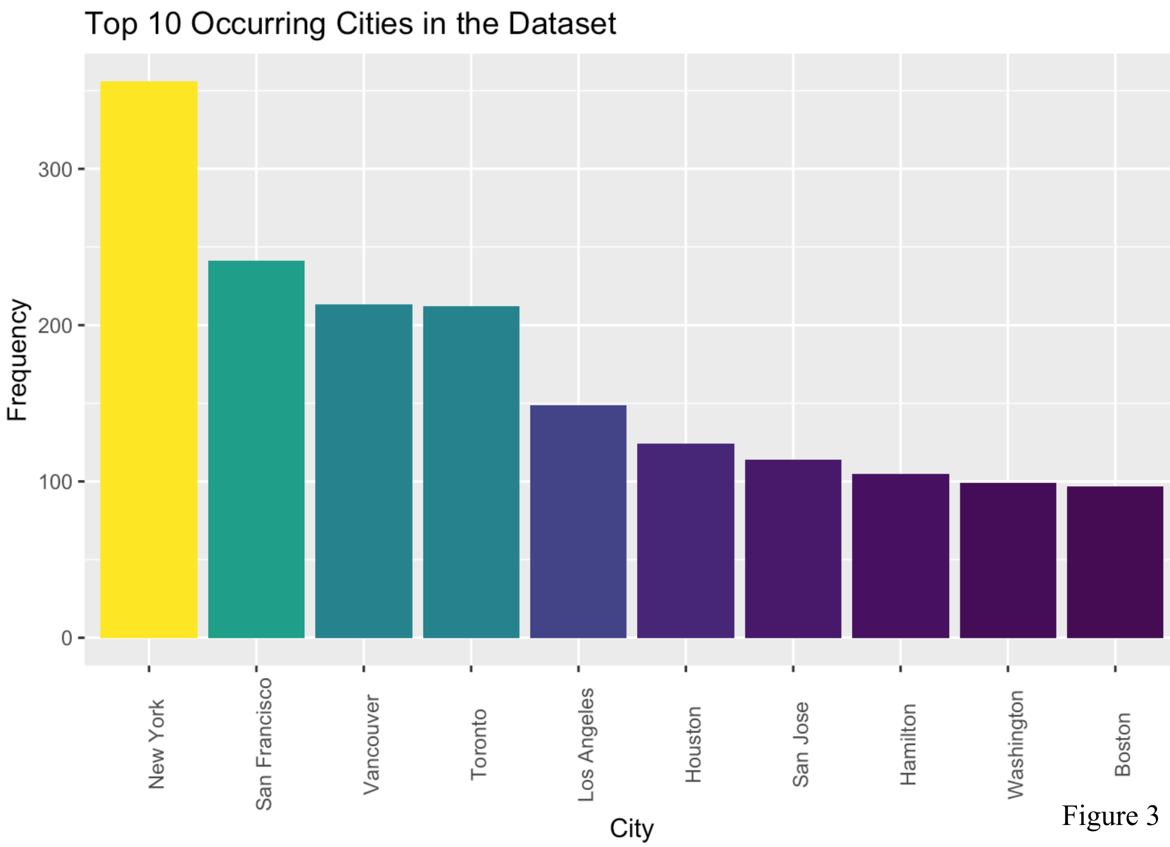

Figure 3

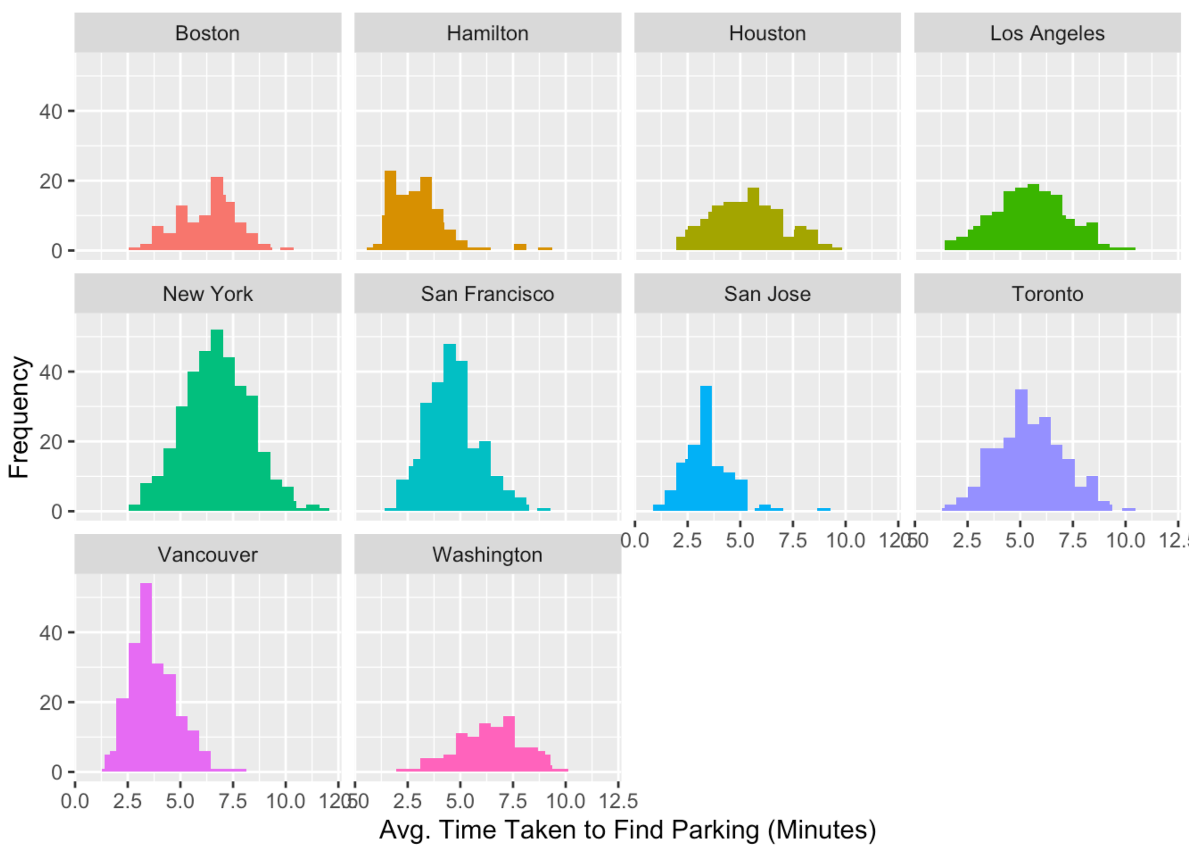

Figure 4

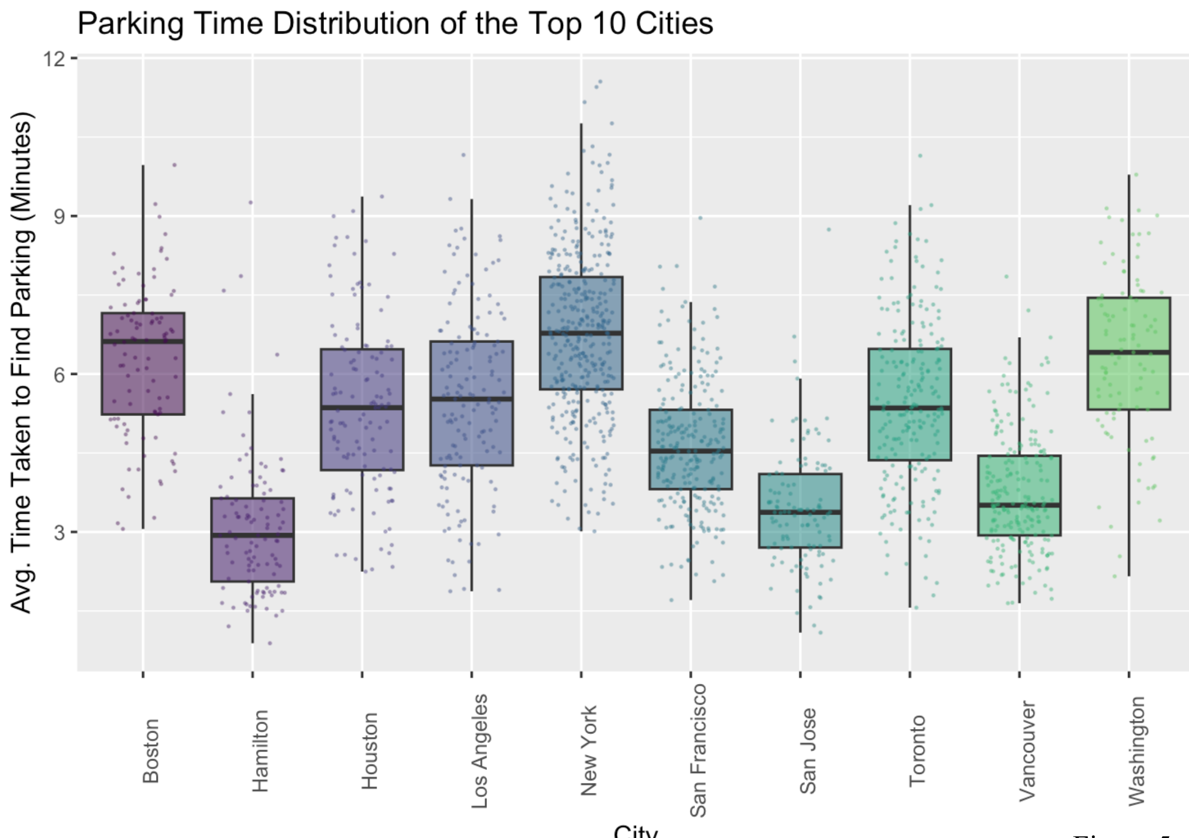

Figure 5

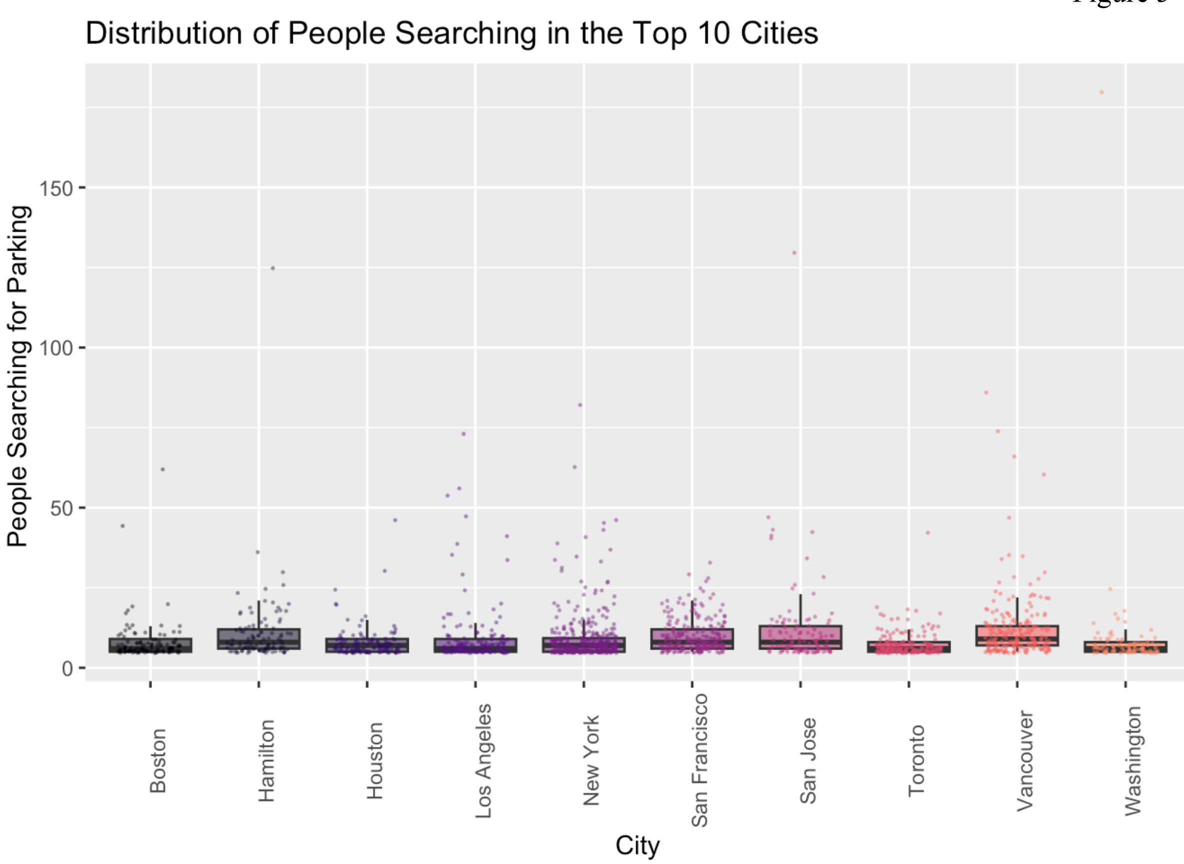

Figure 6

Figure 7

To narrow down our points of interest, we decided to look at the top ten occurring cities in the dataset since it might give us some interesting insights. To visualize this, we created a barplot (Figure 3), which shows us that New York City had the most data (356 points) and among the top 10 there are three Canadian cities: Vancouver, Toronto, and Hamilton. We made boxplots for each city plotted with their points against the average time taken to find parking (Figure 5). The mean time for New York, Boston, and Washington appears to be about 7 to 7.5 minutes and cities like Hamilton, San Jose, and Vancouver have mean times of about 3 to 4 minutes (Figure 4, 5). Our grid of histograms (Figure 4) also reveals that the parking times for New York are approximately Normally distributed along with Los Angeles and Boston, while the others are more skewed toward lower times. Cities with lower parking times could be easier to navigate in or could have more accessible parking. Looking at the boxplot for people searching for parking the numbers are mostly low, but you can see that there are some outliers in which a few locations in some cities had more than 100 drivers looking for parking during the first 2020 lockdown (Figure 6). Because these outliers stretch the boxplots out, it is difficult to see whether the mean number of people searching for parking is different across each of the ten cities. So, we applied a log transformation on the number of total people searching to make it easier to see the differences. From this, we can see that there are more people searching for parking in Vancouver, San Francisco, and San Jose, but the differences among the cities are still low (Figure 7). A reason for these outliers could be that perhaps these locations could be near hospitals or commercial districts with establishments that sell essential items like food, personal protective equipment, or personal hygiene products. On that note, this data was accumulated during the initial onset of COVID-19 when people all around the world were stuck inside their houses, so the relative low times and number of drivers make sense and would not be reflective of traffic today.

In the next step of our analyses, we focus on New York City exclusively to find spatial traffic patterns. Since New York City had the most observations, we thought it was appropriate to use it as our city-level example. Data exploration and online research also revealed that it has the worst parking time among the major US and Canadian cities and continues to be an immense problem. An article from Bloomberg published in August 2022 states that New Yorkers are "shunning the subway in the wake of the Covid outbreak" and that car ownership surged 224%

in 2021 (Sheinerman). Furthermore, a USA Today article from 2017 revealed that the average driver spends about 107 hours annually searching for parking in New York City (McCoy).

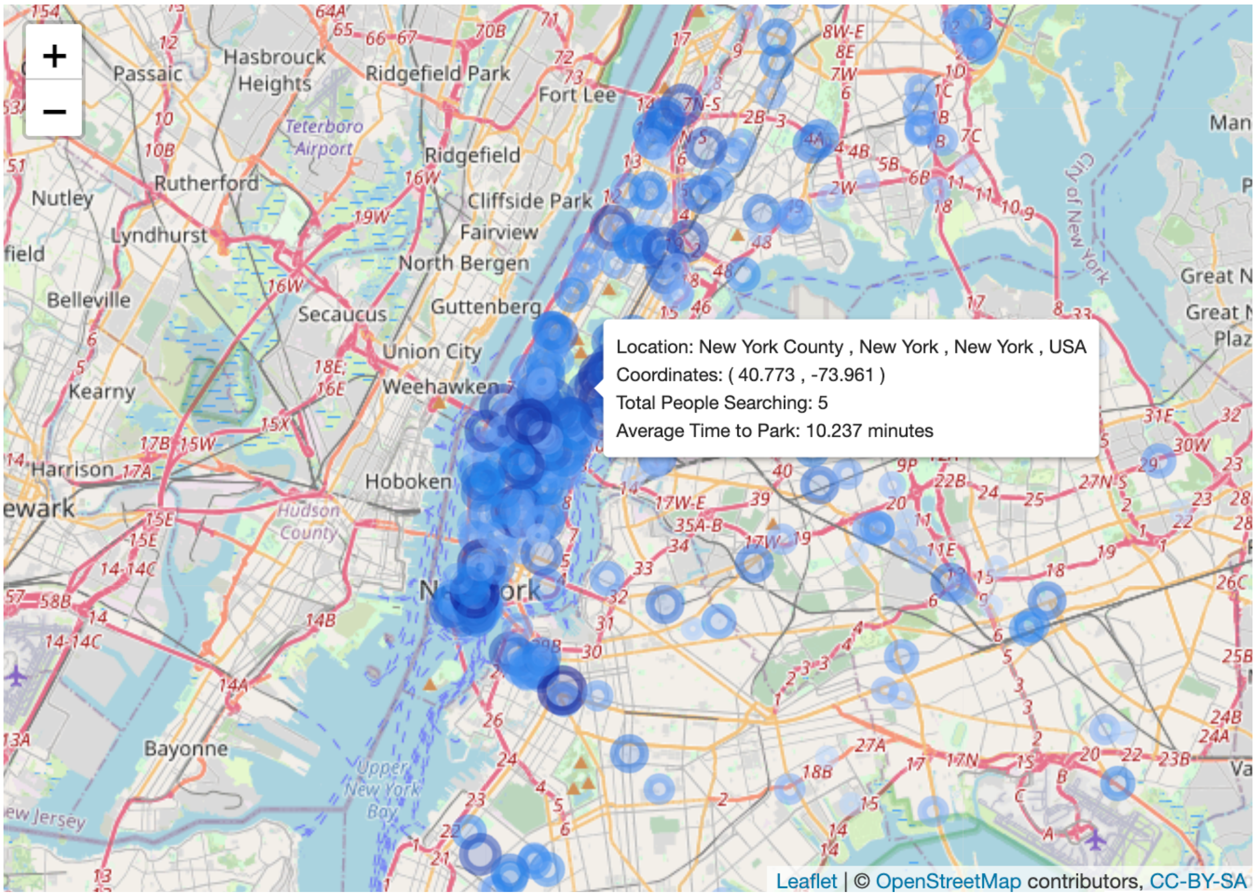

Figure 8

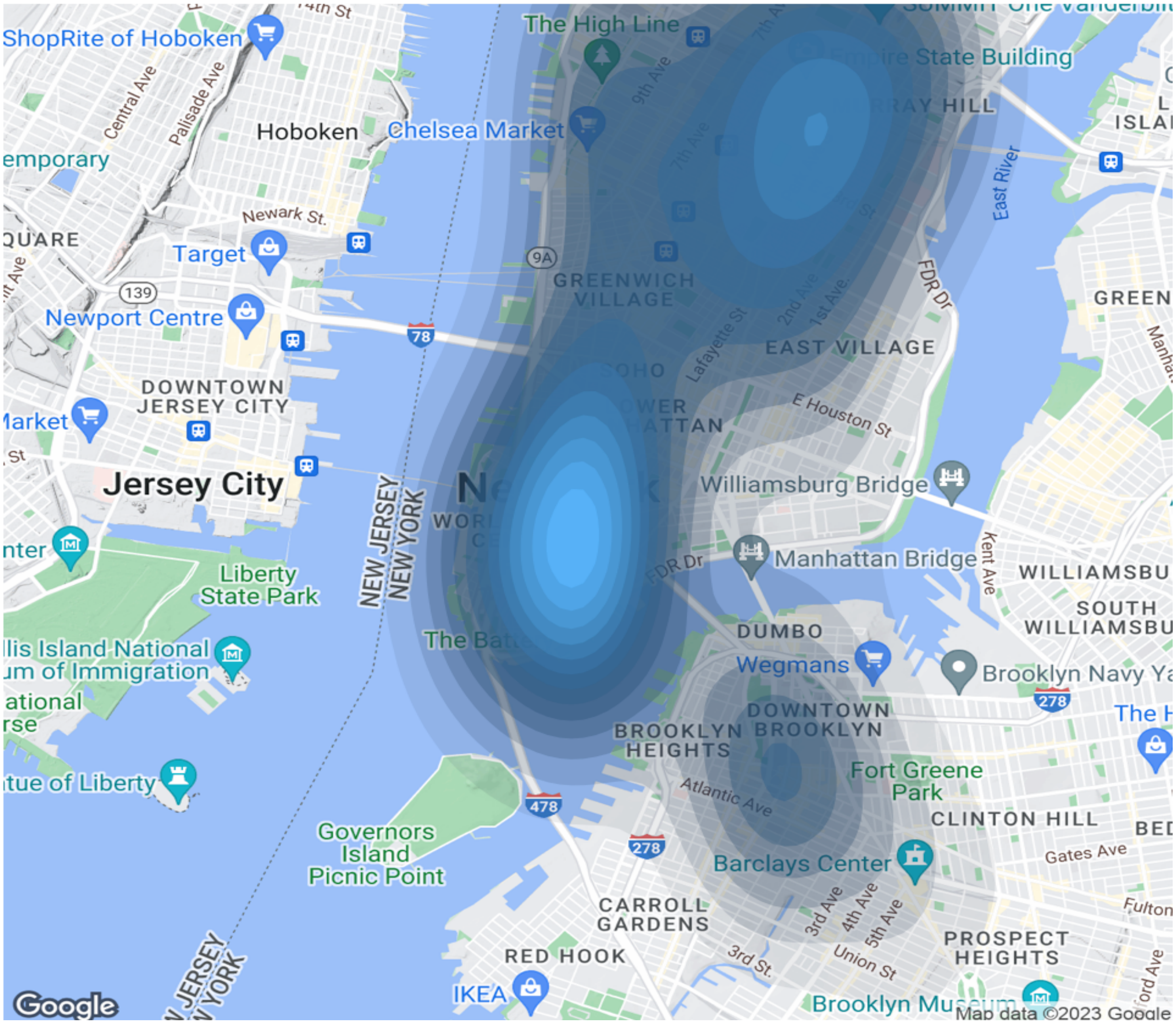

Figure 9

Like we did in Figure 3, we include a screenshot of an interactive Leaflet Plot in Figure 8 except this time we filtered for observations only in New York City. The shade of blue represents the average time it took to park at that specific location – the darker the shade of blue the longer the time. In Figure 9 we use a kernel density mapping based on the average time taken to find parking in New York City. From a city-level view, we see signs of spatial clustering, especially in Manhattan which we would expect. Though this, along with our point plot in Figure 8, are good ways to identify and visualize spatial clustering trends in our data, it is a little difficult to pinpoint what certain areas observe the most difficulty for drivers to find parking due the massive amount of area the points and kernel density cover on the map. Despite this the algorithm still follows the insight we expect – a higher concentration of traffic in and around Manhattan. To dive deeper into definitively finding spatial clustering, we want to analyze spatial autocorrelation using the Global and Local Moran's I values as well as the Geary's C statistic.

```
##  Moran I test under randomisation
##
## data:  park$avgtimetopark
## weights: soiNBW
##
## Moran I statistic standard deviate = 3.060861791603, p-value =
## 0.0011035047177
## alternative hypothesis: greater
## sample estimates:
##    Moran I statistic          Expectation             Variance
##  0.14308391130162901 -0.00281690140845070  0.00227210282558898
```

```
##  Geary C test under randomisation
##
## data:  park$avgtimetopark
## weights: soiNBW
##
## Geary C statistic standard deviate = 2.803144806492, p-value =
## 0.0025303470819
## alternative hypothesis: Expectation greater than statistic
## sample estimates:
##   Geary C statistic         Expectation            Variance
## 0.86155248351716485 1.00000000000000000 0.00243937895364295
```

Figure 10

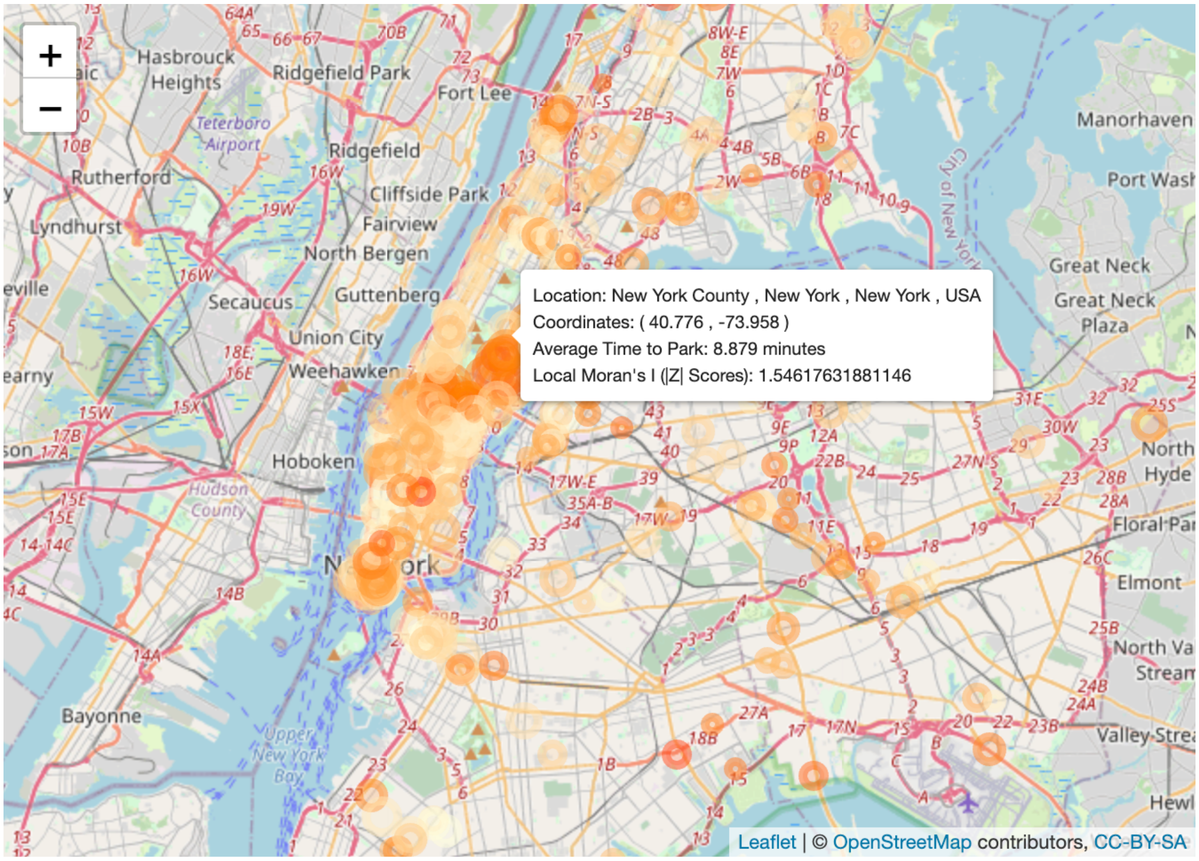

Figure 11

A Global Moran's I value greater than 0 indicates positive spatial autocorrelation and that neighboring regions tend to have similar values and can be clustered whereas a negative Moran's I value indicates negative spatial autocorrelation and that neighboring regions tend to have different values. In the output summaries above (Figure 10), our Global Moran's I value of 0.143 is greater than 0 which indicates positive spatial autocorrelation and that neighboring regions tend to have similar values and can be clustered. We can also see that the positive autocorrelation is statistically significant as the p-value is $0.0011 < 0.05$. A Global Geary's C value of 0 indicates perfect positive spatial autocorrelation and a C value of 2 indicates perfect negative spatial autocorrelation. From the Geary's C test, we observe that the Geary's C value is 0.86 and the p-value is $0.0025 < 0.05$, indicating statistically significant positive spatial autocorrelation.

So, we reject the null hypothesis and conclude that parking times in New York are not randomly distributed and that they are not a random chance spatial process. In other words, locations that observe longer times for parking are more likely to be clustered together and similarly locations that have shorter parking times are more likely to be clustered together. Local Moran's I is computed locally by evaluating the spatial autocorrelation between a location and its neighbors and looking at the spatial autocorrelation among the average times it took to find parking in a certain location. In Figure 11 we can see that locations with higher Local Moran's I values (darker orange) have longer parking times thus are clustered together. Similarly, locations with lower Local Moran's I values (lighter orange) have lower parking times and surround each other more.

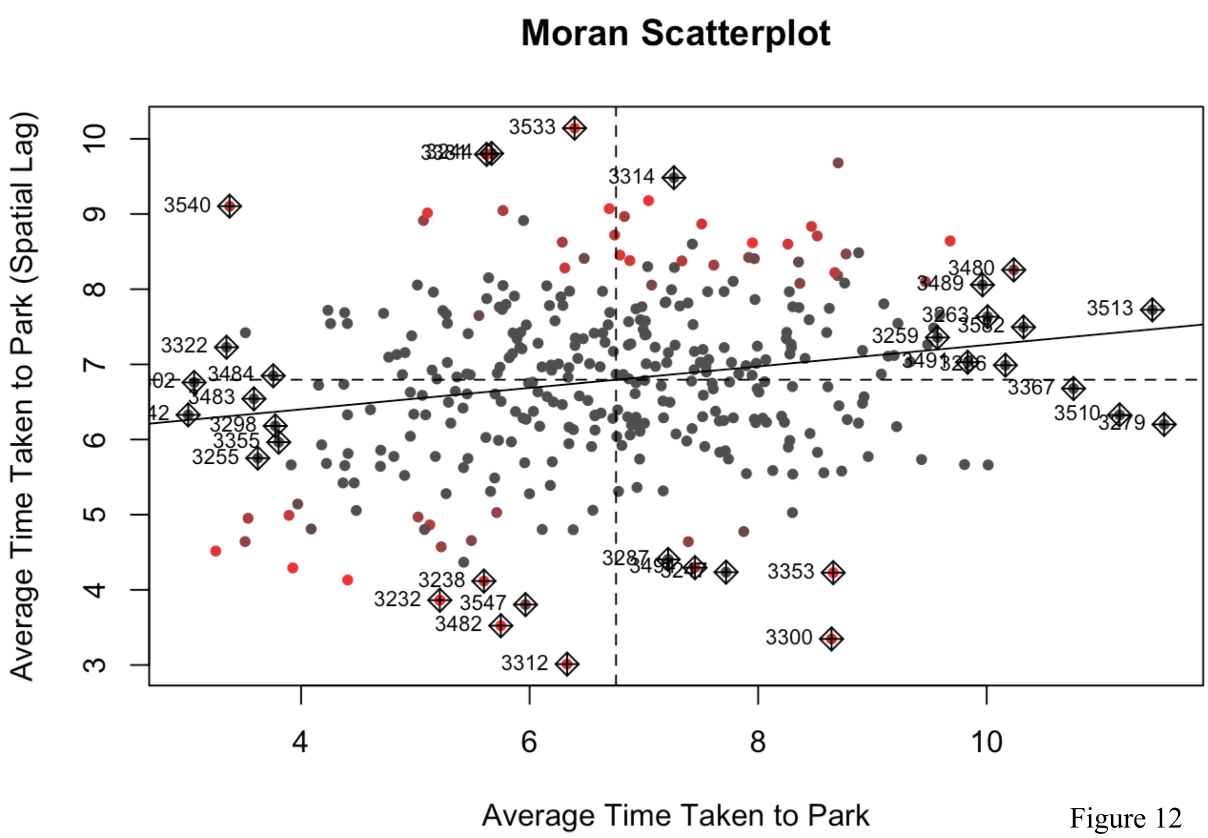

Figure 12

In the Moran scatterplot above (Figure 12) the points in the upper right and lower left quadrants are those in which there is a positive association between the location and the spatially lagged counterparts. Given this information and the positions of the points we can see that there is positive spatial autocorrelation which does agree with the global metrics. However, it is important to note that the Moran scatterplot shows that a positive value of the Global I does not necessarily correspond with values exclusively in the upper right and lower left quadrants – there were a considerable amount of points in the upper left and lower right quadrants that had a negative association between the location and the spatially lagged counterparts, which would explain the best-fit line that shows a weak, but positive relationship.

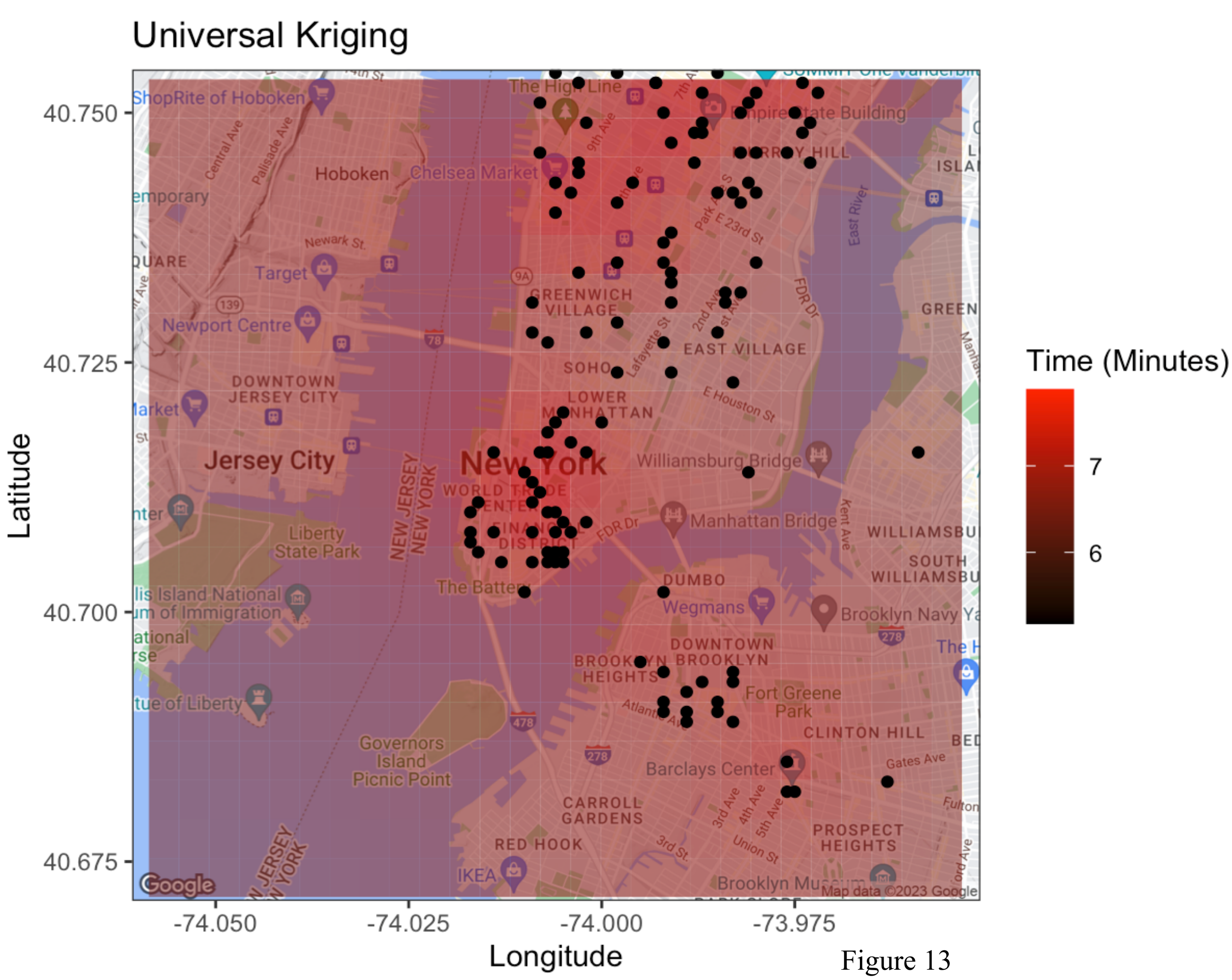

Figure 13

Lastly, we use kriging to predict the parking times of the surrounding areas given what we know about the observed points. The algorithm uses universal kriging as opposed to ordinary kriging because the input data is already marked by an overriding trend of higher times in points where there is more clustering whereas ordinary kriging assumes that there is no trend in the data. As we would expect, we can see that, based off the color scale, the areas surrounding the main clusters of points are predicted to have higher parking times (brighter red) and areas farther away are predicted to have lower times thus less difficulty parking (darker red). From our map, the predicted time to find parking in the clustered Manhattan points appears to be around 7 to 7.5 minutes, which is consistent with the mean time we observed from our plots earlier.

Conclusion

By narrowing down our case study to focus on New York City traffic, we discovered that the amount of time it takes to find parking is not randomly distributed and that there is positive spatial autocorrelation using a kernel density mapping and Local Moran's I. We also found that spatial clustering does occur in areas with higher observed parking times. As a result, we were able to use universal kriging to make predictions of parking times in the surrounding areas, which ended up being consistent with what we derived in the initial data exploration process. Going forward, we would like to gather parking data from the present day and even before the pandemic and conduct the same analyses to see how traffic and search times compare. We would also love to acquire public transportation data and see how mapping parking destinations, bus stops, and subway stations altogether might contribute to gaining more insight on parking difficulty – we believe that our efforts can offer meaningful information on a spatiotemporal level, help predict future traffic outcomes, and contribute to urban planning to accommodate the needs of the inhabiting communities.

References

- Shin, Terence. "Parking Statistics in North America." *Kaggle*, 17 Oct. 2020, https://www.kaggle.com/datasets/terenceshin/searching-for-parking-statistics-in-north-america.
- Sheinerman, Marie-Rose. "NYC's Parking 'Nightmare' Deepens with 224% Increase in New Cars." *Bloomberg.com*, Bloomberg, 16 Aug. 2022, https://www.bloomberg.com/news/articles/2022-08-16/where-to-park-in-nyc-rise-in-new-cars-outdoor-dining-deepen-nightmare.
- McCoy, Kevin. "Drivers Spend an Average of 17 Hours a Year Searching for Parking Spots." *USA Today*, Gannett Satellite Information Network, 13 July 2017, https://www.usatoday.com/story/money/2017/07/12/parking-pain-causes-financial-and-personal-strain/467637001/.