**Title**
A Geostatistical Analysis of Electric Charging Infrastructure in the United States



**Author**
Chen, Joy

**Publication Date**
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# A Geostatistical Analysis of Electric Charging Infrastructure in the United States

A thesis submitted in partial satisfaction
of the requirements for the degree Master of
Applied Statistics and Data Science

by

Joy Qiaoyi Chen

2023

# ABSTRACT OF THE THESIS


A Geostatistical Analysis of Electric Charging

Infrastructure in the United States

by

Joy Qiaoyi Chen

Master of Applied Statistics and Data Science

University of California, Los Angeles, 2023

Professor Frederic R. Paik Schoenberg, Chair

One of the greatest societal challenges of the twenty-first century is tackling the ever-growing issue of global climate change. Transportation alone accounts for nearly 15% of all global emissions[1], and implementing new, cleaner alternatives to traditional fuel-combustion engines is critical in the fight against a warming planet. Electric passenger vehicles play a significant role in a path to decarbonization, and, subsequently, there must be robust electric charging infrastructure to support the increasing adoption of electric vehicles. This analysis utilizes electric charging station data from the U.S. Department of Energy's Alternative Fuels Data Center and employs geostatistical methods to explore clustering of electric charging infrastructure in the United States. Additionally, the analysis aims to identify covariates in areas that exhibit clustering patterns on national and regional levels and utilizes point processes to model station intensity as functions of such covariates.


[1] United States Environmental Protection Agency, *Global Greenhouse Gas Emissions Data*, www.epa.gov.

The thesis of Joy Qiaoyi Chen is approved.

Xiaowu Dai

Vivian Lew

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES

# CHAPTER 1
# Introduction

The global climate crisis has become one of the greatest existential threats to humanity. Ninety seven percent or more of actively publishing climate scientists and numerous scientific organizations agree that human activity has largely been responsible for climate-warming trends in the past century [1]. Excessive warming of the planet above 1.5 degrees Celsius (ºC) can trigger severe consequences including extreme weather, rising sea levels, coral bleaching, and extensive loss of natural ecosystems [2] – all of which can threaten societal and economic stability across the globe.

To address this ever-growing threat, one hundred ninety-five nations gathered in Paris in 2015 and ratified the Paris Agreement which pledged each ratifying member will determine a path towards limiting carbon emissions to curb global warming to 1.5 ºC [3]. This global call to limit carbon emissions has precipitated the demand to shift energy infrastructure away from burning fossil fuels, particularly in the transportation sector. According to the United States Environmental Protection Agency, nearly 15% of global carbon emissions are due to transportation-related activities [4]. Electric vehicles (EVs) and associated electric charging infrastructure are key technologies that, with rapid, widespread adoption, can assist in the fight against climate change.

Progress for electric vehicle infrastructure has been positive since the turn of the century as seen with the meteoric rise of Tesla as well as an industry-wide shift to large investments in electric vehicle research and development. Some of the largest global car manufactures such as Volkswagen, Daimler, BMW, General Motors, Ford, and Toyota have all announced multi-billion-dollar investment efforts to develop electric vehicle infrastructure.

To power this future generation of electric vehicles, specifically in the United States, there must be an extensive electric charging network across the country. While most vehicles are used for short-range distances, a significant hurdle in electric vehicle adoption is "range anxiety" – the perception that EVs are not capable of travelling longer distances due to lack of charging resources. A decade ago, "range anxiety" was a legitimate concern given there were a mere 2,100 charging stations nationwide in 2011 [6]; however, the infrastructure has developed at a rapid speed to boast over 53,000 public charging outlets in 2022 [6]. This growth in charging resources has built a platform enabling even greater EV adoption.

Various private companies – including Tesla, Chargepoint, EVgo, and Volta – have entered the market to provide different networks that now spans across the country. While the stations are available nationally, they are not distributed evenly. Thus, there are areas with greater concentrations of stations than others. Utilizing data from the U.S. Department of Energy's Alternative Fuels database ("the data"), this analysis will apply geostatistical methods to assess clustering of nation-wide electric charging infrastructure, and, based on any prevalence of clustering, examine the geographic and sociopolitical characteristics of such areas to reveal any patterns and covariates of electric station clustering and utilizes point processes to model station intensity.

# CHAPTER 2

# U.S. Department of Energy Alternative Fuels Data

The data for electric charging stations in the United States is provided by the U.S. Department of Energy's Alternative Fuels Database [6]. The data holds nearly 56,000 observations for alternative fuels stations across the U.S. including information about biodiesel, ethanol, hydrogen propane and natural gas stations in addition to electric charging stations.

The original dataset includes sixty-four variables for each station including: station name, address, network information (i.e. Tesla), DC charging availability (fast charging), station opening date, latitude, longitude, operational status and various others. The data was filtered to reflect only operational, public electric charging stations in the continental United States (i.e. excluding Hawaii and Alaska), resulting in a trimmed dataset of nearly 51,400 stations. We focus on public chargers to better understand the network characteristics of chargers that all drivers can access compared to private chargers (i.e. company-specific, government-owned) that serve a standard, repeating user base. Additional variable selection, creation and cleaning narrowed the original sixty-four variables to twenty-seven.

# CHAPTER 3

# Previous Research

Given increasing public and governmental focus on shifting to renewable energy, there has been significant recent research in the battery electric vehicle and charging space. Specifically, there is much discussion surrounding optimal placement of charging stations within cities to address grid power availability or driver demand.

Shahraki et al. (2015) have proposed an optimization model for charging locations to increase vehicle-miles traveled for electric taxis [7], while Xi et al. (2013) have also proposed an optimization model but applied towards maximizing private electric vehicle use [8]. More recently, He et al. (2018) have developed an algorithm based on vehicle range to identify optimal station locations that maximize path flows to stations [9].

While this previous literature provides critical insight into the development of electric charging networks, they do not address the geo-sociopolitical and business patterns of network expansion. Additionally, while there may be mathematically optimal locations based on maximizing vehicle range or increasing utilization, network providers account for different factors when planning locations.

For example, providers must also consider their customers' geographic distribution and areas that are feasible for charger installation. If the majority of electric vehicle owners live in a certain area, it would be more logical to focus on grouping chargers more closely to such customers – even if doing so is not "optimal" for increasing coverage or maximizing vehicle range. Additionally, even if a certain location is identified as algorithmically "optimal", it provides little use if the same area is strategically infeasible to build a charger (i.e. private property, heavily industrial area, no customers in region, etc.).

This analysis differs in that it aims to employ geostatistical methods to identify sociopolitical and geographic characteristics that help inform the distribution of chargers in the United States – rather than suggesting where stations should be built. Based on station distributions, we will identify potential covariates (i.e. income, population, etc.) in areas that display heavy levels of geographic clustering and model station intensity as a function of such covariates. We will test the results of such models help reveal the if certain covariates could in fact be key to charging network providers' building strategies on national and regional levels. For example, are there greater concentrations of stations in areas with higher incomes or other economic factors? Understanding of these sociopolitical and geographic covariates could provide more societally contextualized insight into the future development and expansion of electric charging networks.

# CHAPTER 4

# Geostatistical Methods Overview

## 4.1 Overview of Point Processes

Point processes belong to a subsection of statistical theory that have many crucial and significant applications to real-world occurrences and events. They are commonly leveraged in various applications including epidemiology, land management, finance and various other industries. According to Schoenberg, point processes are random collections of points falling in some space and / or time [11]. They can be commonly classified into *spatial* and *temporal* point process which capture the locations in space (two-dimensional) and times (one-dimensional), respectively, where and during which certain points occur. For this analysis, we are primarily interested in utilizing *spatial* point processes.

One of the most prominent forms of point processes is the **Poisson Process**. According to Keeler (2018), a Poisson process is a point process where each point is stochastically independent from all others governed by an intensity function $\lambda(x)$ which describes the average rate of "events" that occur [11]. The intensity function defines the density or concentration of events within a designated space [11]. A *spatial* Poisson process is a point process defined on the plane $R^2$ [11].

Poisson Processes can be classified as *homogenous* if the intensity function $\lambda$ is constant over $R^2$ (events occur uniformly) or *inhomogenous* if the intensity function varies according to underlying spatial trends or patterns within $R^2$. According to Hartmann et al. (2018), the properties of a homogenous point process are often referenced as *complete spatial randomness* (CSR) [12]. According to Zhang et al. (2019), in most applications, the property of homogeneity is not practical and thus inhomogeneous point processes are the most widespread point process models [13]. Inhomogeneity indicates the underlying point pattern is subject to external factors, thus understanding the influence of covariates is key to modeling the intensity function [13].

## 4.2 Estimation Techniques

### 4.2.1 F, G, K and J Functions

<u>F Functions</u>

According to Hartmann et al. (2018), "the F function measures the distribution of all distances from an arbitrary reference location *u* in a plane to the nearest observed event *j*." Thus, the F function is commonly referred to as the *empty space function*. The empirical distribution function on a grid $u_j$, j = 1, …, *m* is:

$$\hat{F}(r) = \frac{1}{m}\sum_{j} 1\left\{d(u_j, x) \leq r\right\}$$

For a homogenous Poisson process with intensity λ, the value of the F function becomes (Hartmann):

$$F_{pois}(r) = 1 - \exp\,(\lambda \pi r^2)$$

To determine if a spatial pattern exhibits clustering, $\hat{F}(r)$ is compared with $F_{pois}(r)$ when plotted. Cases where $\hat{F}(r) < F_{pois}(r)$ suggest clustering is present in the pattern [12]

G Functions

The G Function is a measure of the distribution of distances from an arbitrary event to its nearest neighbors [12]. The empirical distribution function is [12]:

$$\hat{G}(r) = \frac{1}{n(x)} \sum_{i} 1\{t_i \leq r\}$$

For a homogenous Poisson point process [12]:

$$G_{pois}(r) = 1 - \exp\left(-\lambda \pi r^2\right)$$

Thus, the G function has the reverse interpretation as the F function: if $\hat{G}(r) > G_{pois}(r)$, there is evidence of clustering within the pattern [12].

K Functions

The K function calculates the expected number of other points within a certain distance *r* of a point within the point process [12]. The empirical distribution function is [12]:

$$\hat{K}(r) = \frac{1}{\hat{\lambda}\, area(W)} \sum_{i} \sum_{i \neq j} 1\{\|x_i - x_j\|\} e(x_i \cdot x_j)$$

For a homogenous Poisson point process [12]:

$$K_{pois}(r) = \lambda \pi r^2$$

When comparing $\hat{K}(r)$ and $K_{pois}(r)$, cases where $\hat{K}(r) > K_{pois}(r)$ suggests clustering within the point pattern.

J Functions

The J function is calculated using the F and G functions [12, 14]:

$$J(r) = \frac{1 - G(r)}{1 - F(r)}$$

For an homogenous Poisson process, the F function = G function and thus [14]:

$$J_{pois}(r) = 1$$

To determine if a pattern exhibits clustering, $J(r) < 1$ [14].

### 4.2.2 Kernel Smoothing (Kernel Density Estimation)

Kernel Smoothing is a nonparametric density technique to estimating an underlying probability density function for a point process [15]. It uses the principle of weighting distributions according to proximity to a set location utilizing a *kernel function* and a *smoothing bandwidth* [16]. The primary thesis of the method is to estimate a density function at a specific point using neighboring observations [17]. A critical part of kernel smoothing includes selecting the right bandwidth for the estimator given it is responsible for the amount of smoothing that occurs [17]. Bandwidth values that are too small can lead to undersmoothing while the inverse occurs with values too large [18]. With application of the right bandwidth, kernel smoothing can provide insight into the presence of clustering within a point process.

# CHAPTER 5

# Data Analysis and Results

## 5.1 National-Level Data

### 5.1.1 Graphical Analysis

We begin by examining the spatial characteristics of charging stations on a nationwide basis to understand broader distribution patterns across the continental United States. Figure 5.1 below is a graphical representation of stations across the U.S., with each blue dot representing one station:

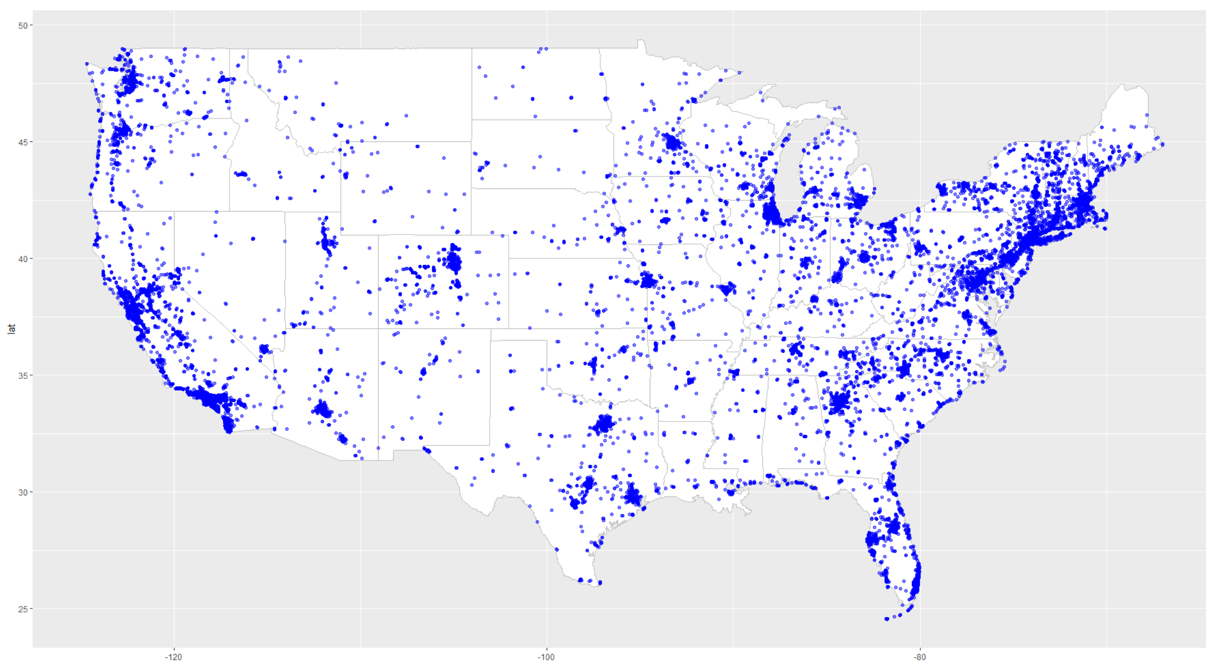



***Figure 5.1:*** *National Distribution of Electric Charging Stations*

Graphically, it appears there are heavy distributions of stations on the coasts – particularly California and the New York / Tri-State region – while there are noticeable gaps in the mid-Northern states such as Idaho, Montana, Wyoming, North and South Dakota and Kansas. Additionally, there are multiple large groupings across the country that correspond to locations for some of the largest cities in the nation. For example, the large congregation of points in Southern California correspond to Los Angeles and San Diego areas while the large grouping in Colorado to Denver. Figure 5.2 below juxtaposes a few names of large cities corresponding to some of the prominent station concentrations:

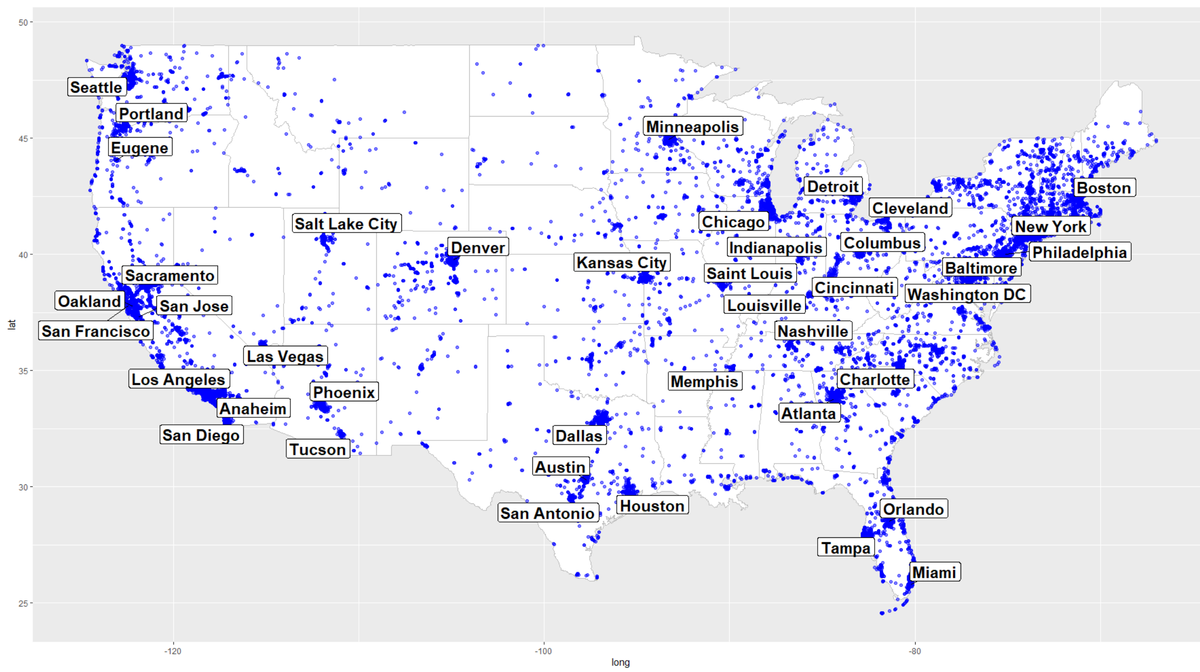



***Figure 5.2***: *National Distribution of Charging Stations with Superimposed City Labels*

Based on initial observations, it appears the national distribution of electric charging stations is inhomogeneous with exhibited clustering around large cities. It may appear qualitatively obvious that cities would foster higher concentrations of electric stations given the majority of people, and subsequently vehicles, reside in such metropolitan areas. However, we must present tangible evidence to evaluate this statement. We will utilize geostatistical methods to demonstrate, firstly, that stations exhibit clustering on a national level, and, secondly, that the clustering correlates with large metropolitan areas. Given cities typically have larger populations, county population will be used as a proxy for metropolitan areas.

### 5.1.2 Clustering Analysis

F, G, K and J functions are employed to identify the presence of clustering in the nationwide station data. Below are renderings for F, G, K, and J function curves for national data:

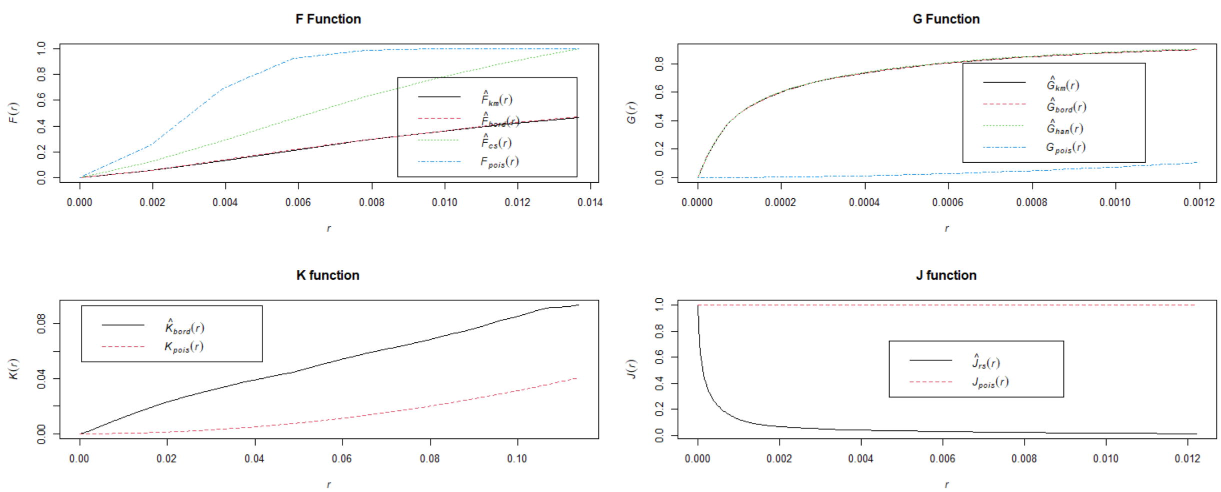



***Figure 5.3:*** *F, G, K, and J Functions for National Station Distribution*

The estimated function curves appear to support the presence of clustering in the national data – the J function < 1, the K and G functions are above the theoretical Poisson curve while, inversely, the F function is below. To better understand the presence of such clustering as the F, G, J and K functions entail, kernel smoothing was employed after normalizing latitude and longitude data to a [0,1] x [0,1] grid. The results are as follows:

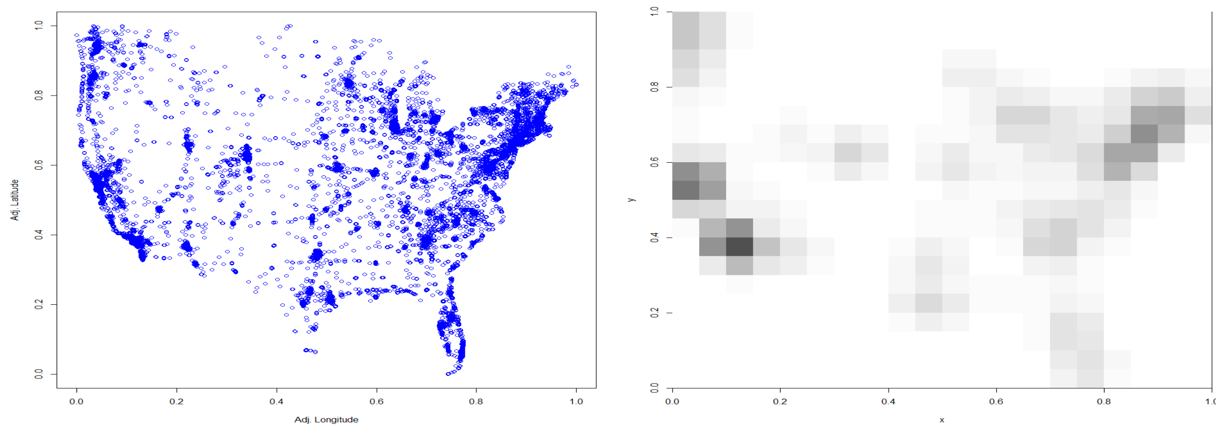



***Figure 5.4***: *National Distribution of Electric Charging Stations (left) Juxtaposed against Kernel Smoothing (right)*

The panel on the left displays the original, scaled data points and on the right, the kernel smoothing which displays estimated intensities on an increasing grey-scale (i.e. the darker the kernel, the higher the estimated intensity). The kernel smoothing appears to confirm the initial graphical analysis with heavy clustering on the coasts in California and the New York / Tri-State area while picking up on the larger cities in the mid-Western corridor (east of Texas).

### 5.1.3 Clustering Correlation with Population on a National Level

The kernel smoothing appears to support the initial hypothesis of heavy clustering in larger cities. Kernels are highlighting large groupings that correlate with locations of large metropolitan areas. Specifically, the kernel smoothing attributes the highest estimated intensities to Los Angeles, the San Francisco Bay Area, and the New York / Tri-State region – corroborating the visual analysis from Figure 5.2.

To further analyze the potential relationship between large metropolitan areas and station clustering, we utilize data from the 2014-2018 5-Year American Community Survey (ACS) available through the '*tidycensus*' package in R. Based on ACS data for population by county, we highlight and juxtapose the top 150 most populous counties against the original distribution of station points. The median population of these counties is approximately 796,000 people, while Los Angeles County – the largest in the United States – supports over 10 million.

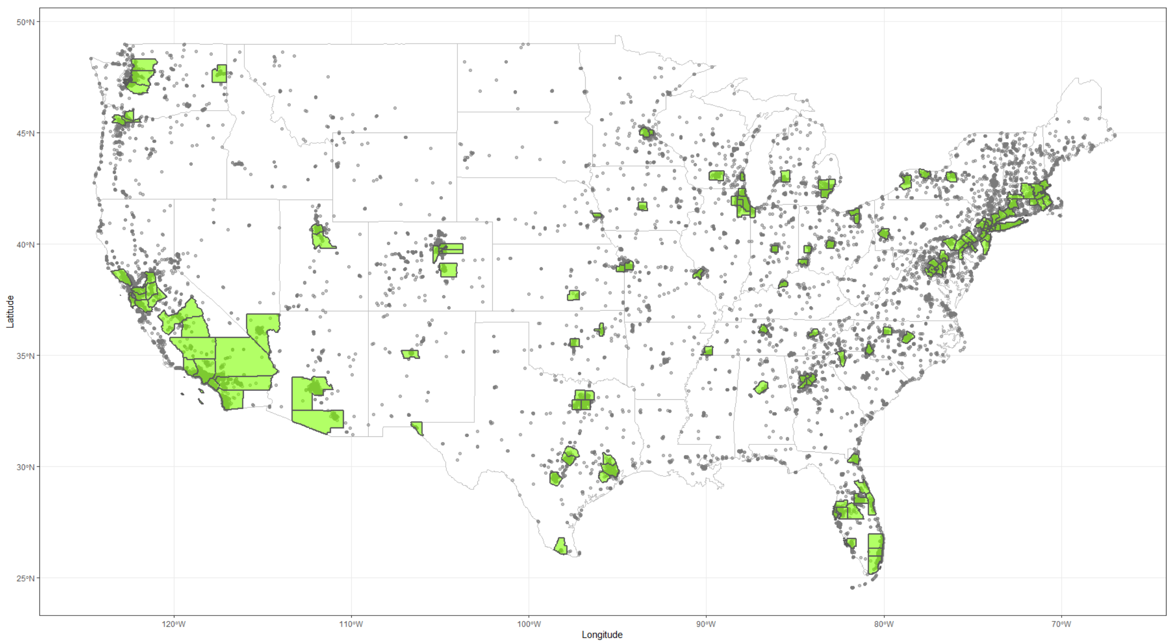



***__Figure 5.5__:*** *National Distribution of Electric Charging Stations with 150 most Populous Counties Superimposed*

Figure 5.5 depicts the top 150 populated counties shaded in green, and the areas they cover visibly appear to correspond with the previously identified point clusters. Notably, areas the previous kernel smoothing identified with high levels of clustering – Los Angeles, the San Francisco Bay Area, and Washington DC to Boston corridor – are all highlighted as areas with high populations. In addition, other large cities with station clusters are also highlighted including, but not limited to, Phoenix

(Arizona), Houston (Texas) and Atlanta (Georgia). This suggests, on a national level, electric charging stations exhibits clustering behavior more commonly in large metropolitan areas.

A national view, however, is expansive and difficult to identify how the stations are creating the clustering effects within highly populated metropolitan areas. While the analysis helps to confirm some underlying relationship between population and clustering of electric charging stations, that information may not be the most useful given its obvious nature. Generally, we can expect that electric charging stations are centralized around areas which large numbers of people are residing and traveling within.

Thus, we are interested in the more detailed factors within these large metropolitan areas that correlate to intra-city clustering. Using Los Angeles as an example, is the clustering in the city created by various collections of neighborhood-level clusters? Or are the stations more evenly spaced throughout Los Angeles, but at a higher density than other areas of California, which then appear clustered from a nationwide perspective? We are also interested in uncovering other potential covariate variables (in addition to high population) these clustered areas may share that could give greater insight into the geographic strategies that organizations employ when constructing stations.

We will further explore specific regions identified as high-cluster, high population areas to analyze how stations are distributed. Based on the national analysis results, we will focus on the Los Angeles and New York regions – two of the largest cities in the nation according to the US Census Bureau [19].

## 5.2 Analysis of Los Angeles, California

We will perform a closer analysis of Los Angeles, California, given its particularly high density of stations, to uncover more detailed covariates associated with station clustering in addition to population as identified in the national-level data.

### 5.2.1 Graphical Analysis

To examine the distribution of stations for Los Angeles, we filtered the national data by longitude and latitude coordinates within Los Angeles County south of the 34.35º latitude parallel. We choose to exclude areas north of the 34.35º latitude due to the more rural nature of the area and significant presence of forests. Below are two maps: on the left, a Google Map layout of Los Angeles and, on the right, the same map with charging station points superimposed:

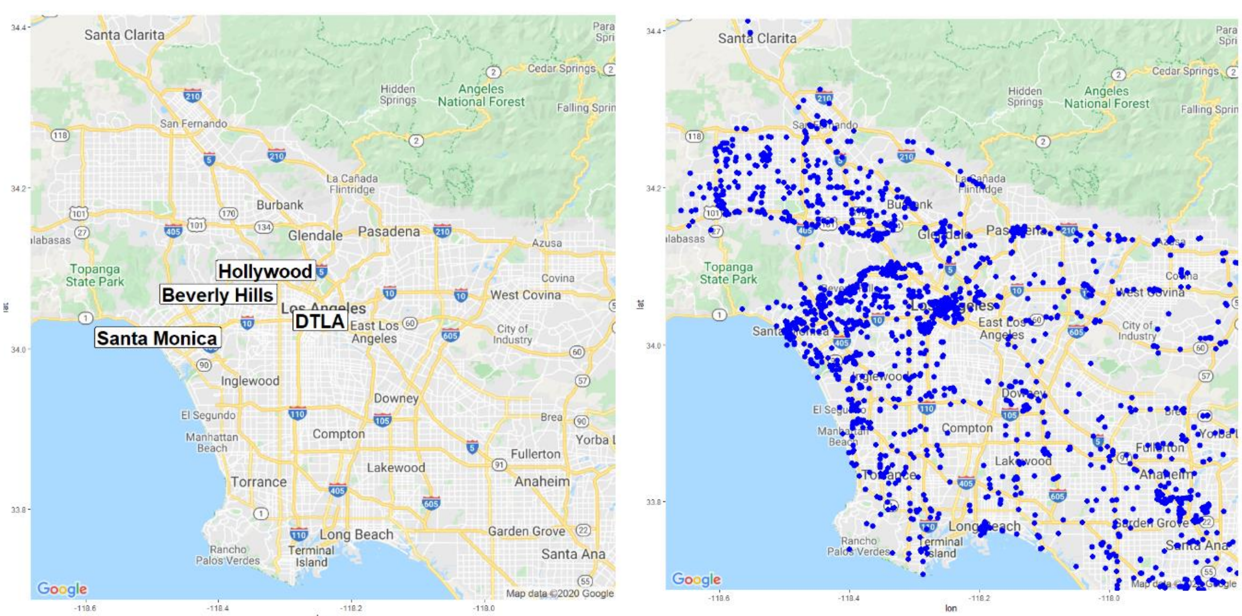



***Figure 5.6:*** *Map of Los Angeles (left) and Overlay of Station Locations (right)*

Visually, it appears there are groupings predominately along the north-western edge of the city (south of the Hollywood Hills) including in Santa Monica and Beverly Hills, as well as in Hollywood and the DTLA (downtown LA) region. In addition, it seems the density of stations generally decreases moving inland with a particularly noticeable "desert" south of DTLA through Compton bounded by the Interstate 405. From these visual observations, it appears charging stations are not evenly distributed across Los Angeles. We will conduct clustering analysis to statistically demonstrate the presence of clustering and present observations of the clustering areas.

### 5.2.2 Clustering Analysis

Visually, it appears there are station groupings predominately along the north-western edge of the city (south of the Hollywood Hills) including in Santa Monica and Beverly Hills, as well as in Hollywood and the DTLA (downtown LA) region. F, G, K and J functions are employed to identify the presence of clustering within the Los Angeles station data. Results for F/G/J/K functions and kernel smoothing are available below to statistically analyze if indeed such clustering is present:

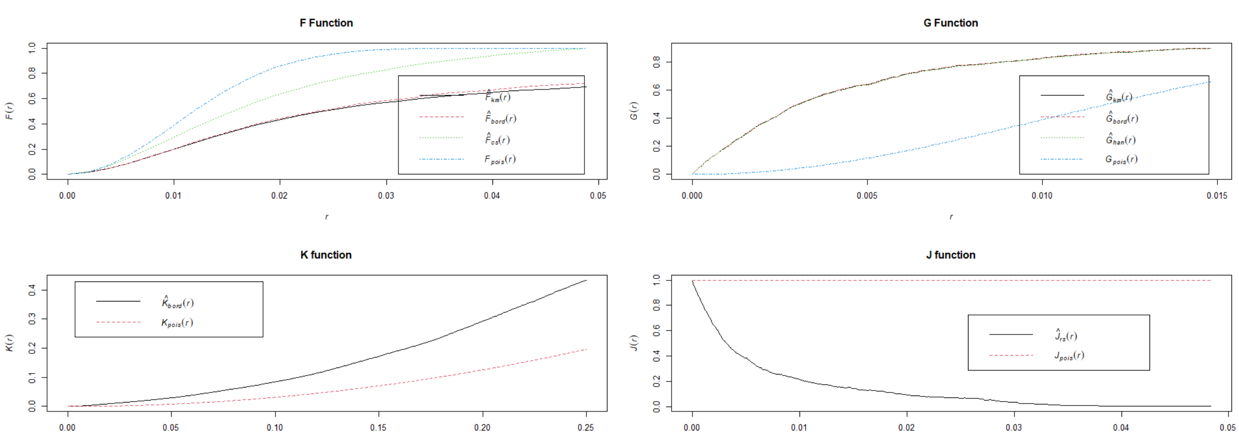
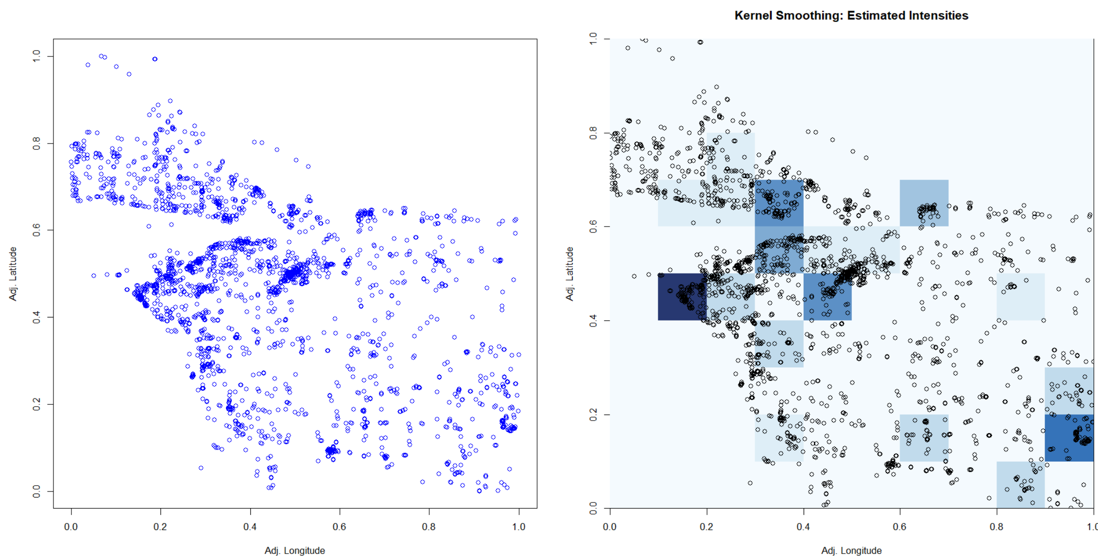



***Figure 5.7:*** *F, G, K, and J Functions for Los Angeles Station Distribution*




***<u>Figure 5.8</u>:*** *Los Angeles Distribution of Electric Charging Stations (left) Juxtaposed against Kernel Smoothing (right)*

The kernel smoothing and pseudo-likelihood models both attribute high levels of clustering in what corresponds to the Santa Monica and DTLA regions. We "zoom-in" to these regions to better understand where the stations are located that contributes to the higher intensities within the city. For Santa Monica, it appears many stations are congregated about what is the Third-Street Promenade, and, in DTLA, within the area bounded between the Interstate 110 and S Grand Ave. – the same area with Staples Center, Grand Central Market and various high-end restaurants.

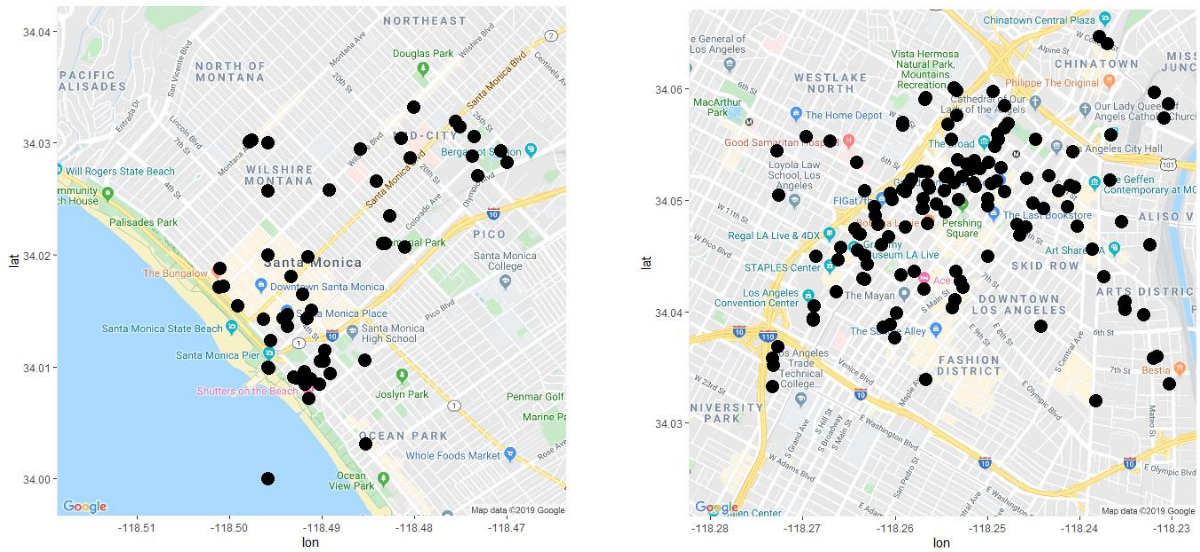



***Figure 5.9***: *Station Distribution in Santa Monica (left) and Station Distribution in DTLA (right)*

### 5.2.3 Observations

Similar to the national analysis, there appears to be evidence of clustering within Los Angeles. The estimated function curves appear to support the presence of clustering in the Los Angeles data – the F function is below the theoretical Poisson curve, the K and G functions are above the theoretical Poisson curve, and the J function is < 1. The kernel smoothing attributes high estinated densities to to DTLA and Santa monica., and there is a generally skewed distribution of stations towards the northern edges of the city (excluding North Hollywood). These regions all appear to commonly demonstrate higher levels of tourism and commerce, particularly for Hollywood, Santa Monica and DTLA.

## 5.3 Analysis of New York, New York

### 5.3.1 Graphical Analysis

Employing a similar analytical process to the Los Angeles data, we graphically examine the distribution of charging stations within the New York City area. A Google Map layout of New York City and, on the right, the same map with charging station points superimposed:

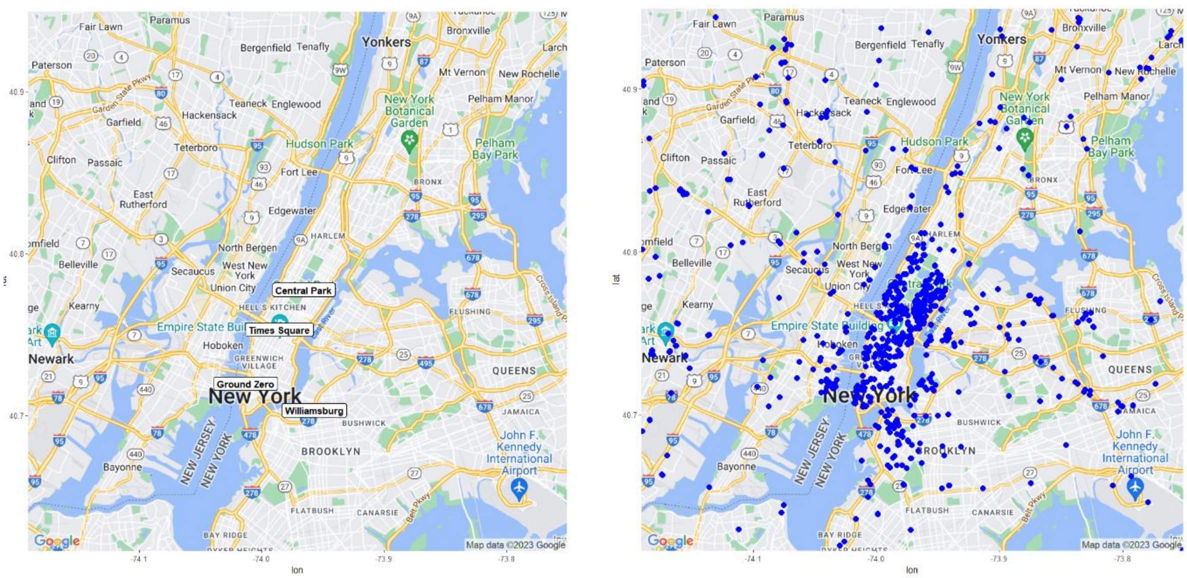



***Figure 5.10:*** *Map of New York City (left) and Map with Overlay of Station Locations (right)*

Visually, the concentration of stations within the Manhattan area it is immediately striking compared to the rather dispersed nature of stations everywhere else. It appears there is also some congregations within Williamsburg, Brooklyn (southeast of the tip of Manhattan). Given Manhattan has various distinct areas, we zoom in further specifically into Manhattan to better visualize the distribution. The below figure displays, on a closer scale, charging stations within Manhattan:

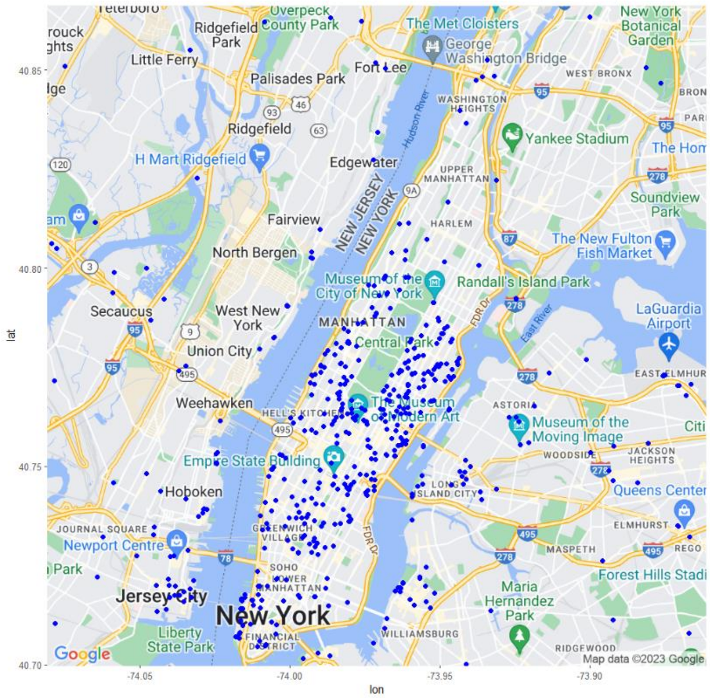



***Figure 5.11:*** *Map of Manhattan with Overlay of Station Locations*

From the close-up in Manhattan, it appears there are a significant number of stations around the Upper West and East Sides, the southern end of Central Park, Greenwich Village, and the Financial District (Ground Zero). These areas We will perform geostatistical analysis within Manhattan to demonstrate the presence of clustering and present observations of the clustered areas.

### 5.3.2 Clustering Analysis

F, G, K and J functions are again employed to identify the presence of clustering within the Los Angeles station data. Results forF/G/J/K functions and kernel smoothing are available below to analyze if indeed such clustering is present:

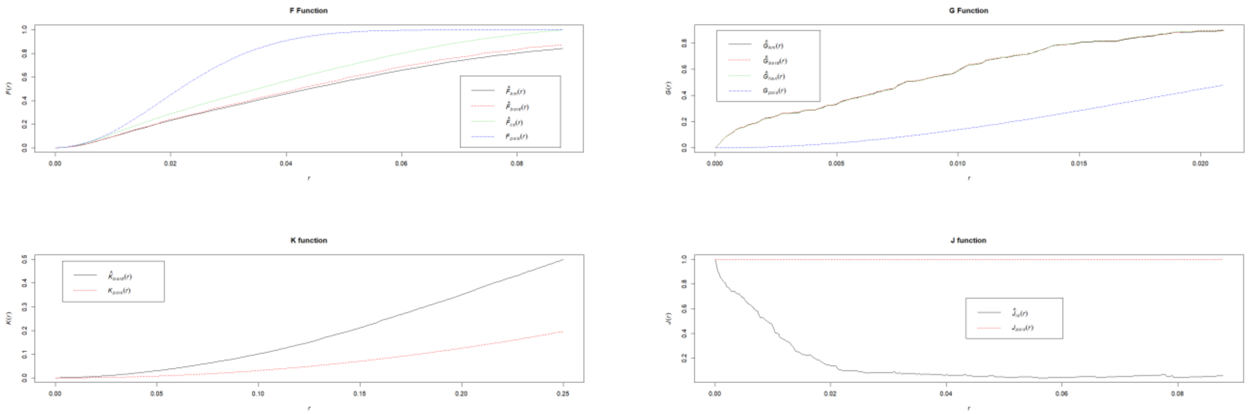



***Figure 5.12:*** *F, G, K, and J Functions for New York Station Distribution*

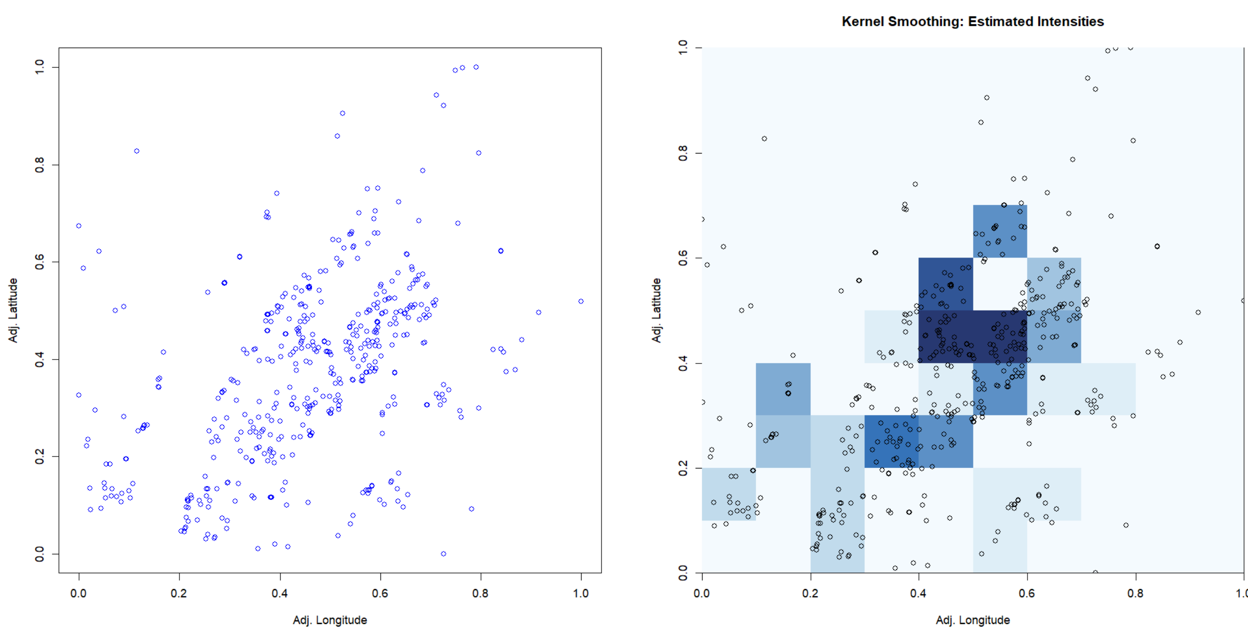



***Figure 5.13:*** *New York Distribution of Electric Charging Stations (left) Juxtaposed against Kernel Smoothing (right)*

### 5.3.3 Observations

Based on the analysis above, there appears to be evidence of clustering within New York City and Manhattan specifically. The estimated function curves appear to support the presence of clustering in the New York data – the F function is below the theoretical Poisson curve, the K and G functions are above the theoretical Poisson curve, and the J function is < 1. The pseudo-likelihood model and kernel smoothing both appear to demonstrate there is a generally skewed distribution of stations towards the southern borders of Central Park as well as the areas in Greenwich Village bordering Washington Square Park.

In relation to the broader New York City area, Manhattan is undoubtedly considered the "hub" of attractions and tourism. On that scale, as many stations are concentrated on Manhattan, it seems stations in New York City follow a similar pattern of being centralized in busy, tourist- and commerce-heavy areas. When "zooming-in" to Manhattan, clustering appears to be in areas that have higher values of real estate. For example, the area immediately south of Central Park, where the clustering analysis shows the highest estimated intensities, is known as "Billionaire's Row" [ref] while the south-western edge of the Park corresponds to Columbus Circle where some of the largest hotels such as the Mandarin Oriental and Trump Tower are located.

## 5.4 Observed Potential Covariates

For both Los Angeles and New York, there appears to be a pattern where stations are clustered around busy, downtown areas and touristic attractions. For example, New York appears to feature heavy concentrations of stations within Manhattan in comparison to the broader metropolitan area. This is particularly interesting given only 22% of households in Manhattan own a car compared to other regions with much higher car ownership (Queens 62%, Brooklyn 44%) [20] as well as the existence of the MTA subway system. Why would there be such a large clustering effect of electric charging stations in an area where there is low car ownership and high usage of public transportation? Similarly in Los Angeles, stations demonstrate clustering around Santa Monica, Downtown LA, Hollywood and Beverly Hills. This would appear counter-intuitive as these touristic areas typically correlate with difficult driving conditions, large crowds and limited or expensive parking options. Assuming most people would not be interested in driving through such areas, why are there such high concentrations of charging stations?

One hypothesis for why charging stations are centralized around tourist and sight-seeing districts could be due to the current time-requirements for charging vehicles. While technology has been improving steadily over the past decade around battery capacity and charging efficiency, charging an electric vehicle is still a time-consuming process that, depending upon the battery size, could take anywhere between 30 minutes to an hour. Thus, charging companies, to improve the experience for customers, strategically designate stations in areas that have attractions nearby where people can easily visit while waiting for their vehicles to charge. These "touristy" areas also correlate with high levels of restaurants, bars and other hospitality-related services.

When comparing more granularly between high-clustering areas in Los Angeles (Santa Monica, DTLA) and New York (Manhattan), it appears stations in Santa Monica / DTLA may have tourism / hospitality as a covariate as we see the high concentrations of stations in areas such as Third Street Promenade and the DTLA Financial District. On the other hand, Manhattan appears to have real estate value as a covariate as we see high density of stations around the upper West and East sides as well as Billionaire's Row.

To test the potential of these covariates on a city-level, we will fit Poisson point process models for both cities and analyze the results. It would be noted that this model fitting differs from the previously utilized kernel smoothing and pseudo-likelihood models in that we will be using original geographical data, not normalized coordinates on a [1, 1] grid. We will employ data from the US Census' 5-Year American Community Survey (ACS) in 2020.

# CHAPTER 6

# Modeling Station Intensity as a Function of Covariates

## 6.1 Overview

As previously mentioned, we are more interested in understanding the granular city-level implications and relationships of stations distributions given the association (on a national-scale) between population and station clustering appears qualitatively obvious. We will utilize ACS Data to fit Poisson point process models to Los Angeles using *Total Service occupations: Arts, entertainment, and recreation and accommodation and food services (variable id: C24050_040)* and New York data using real estate taxes paid by census tract (as a proxy for real estate value: *Estimate of Median Real Estate Taxes Paid – Total (US Dollars) (variable id: B25103_001)* [21].

## 6.2 Model Results

### 6.2.1 Los Angeles

We utilize *spatstat* in R to fit a Poisson point process model to the Los Angeles data using total median real estate taxes paid as a covariate. Below are the results of the model summary as well as a plot of model predictions juxtaposed with the observed station pattern:

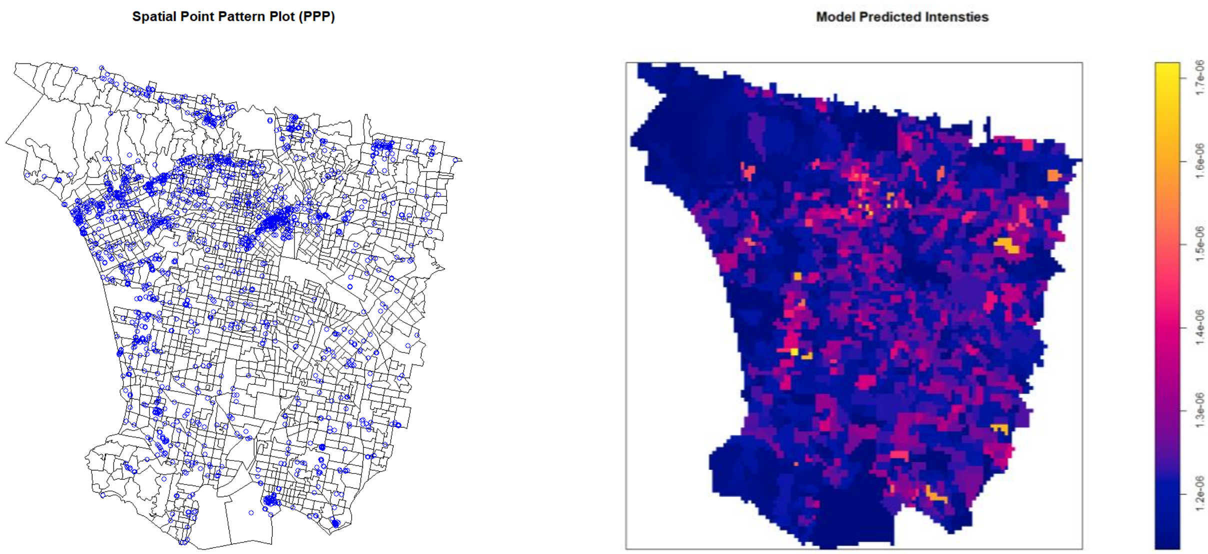



***Figure 6.1:*** *Los Angeles PPP Spatial Point Pattern (left) and Predicted Model Intensities (right)*

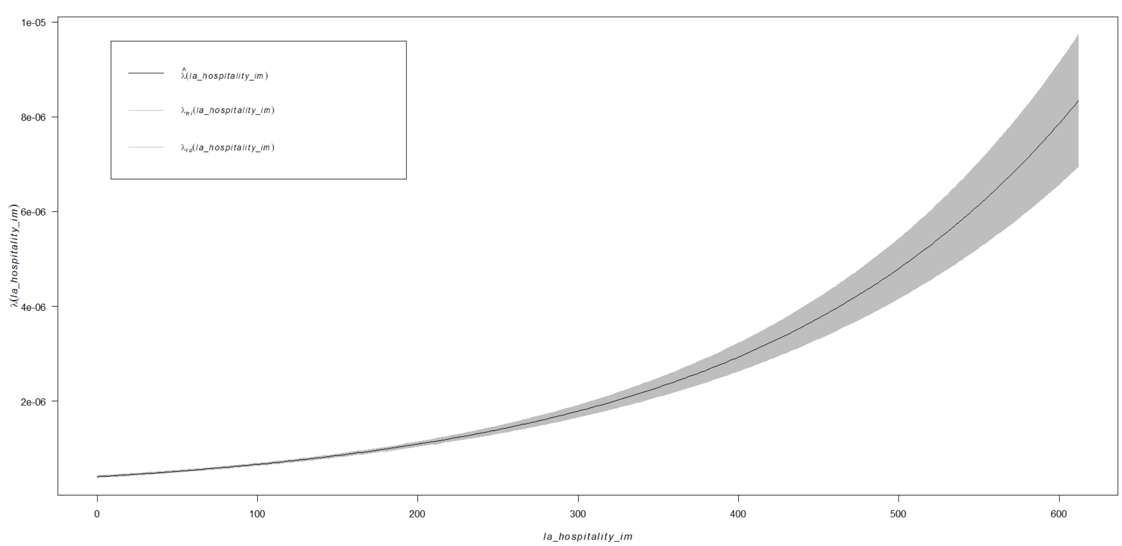



***Figure 6.2:*** *Effectfun() Plot of Los Angeles Model*

```
Log intensity: ~la_hospitality_im
Model depends on external covariate 'la_hospitality_im'
Covariates provided:
        la_hospitality_im: im

Fitted trend coefficients:
      (Intercept) la_hospitality_im
    -14.711236438       0.004929033

                       Estimate         S.E.       CI95.lo       CI95.hi Ztest        Zval
(Intercept)       -14.711236438 0.0254734526 -14.761163487 -14.661309388   *** -577.51247
la_hospitality_im   0.004929033 0.0001623249   0.004610882   0.005247184   ***   30.36523
```

***__Figure 6.3__:*** *Model Summary Statistics of Los Angeles Model*

Graphically analyzing the model results compared to the spatial point pattern, it appears that, the model captures some of the same high-density areas such as the Santa Monica and Hollywood regions. However, the model does not capture the high density to the same degree within DTLA while assigning higher densities to the more eastern edges of the county. The model outputs demonstrate there is a general positive relationship between station intensity and hospitality jobs as a covariate. This is further illustrated in the *effectfun()* plot in Figure 6.3.

### 6.2.2 New York

Below are the results of the model summary as well as a plot of model predictions juxtaposed with the observed station pattern:

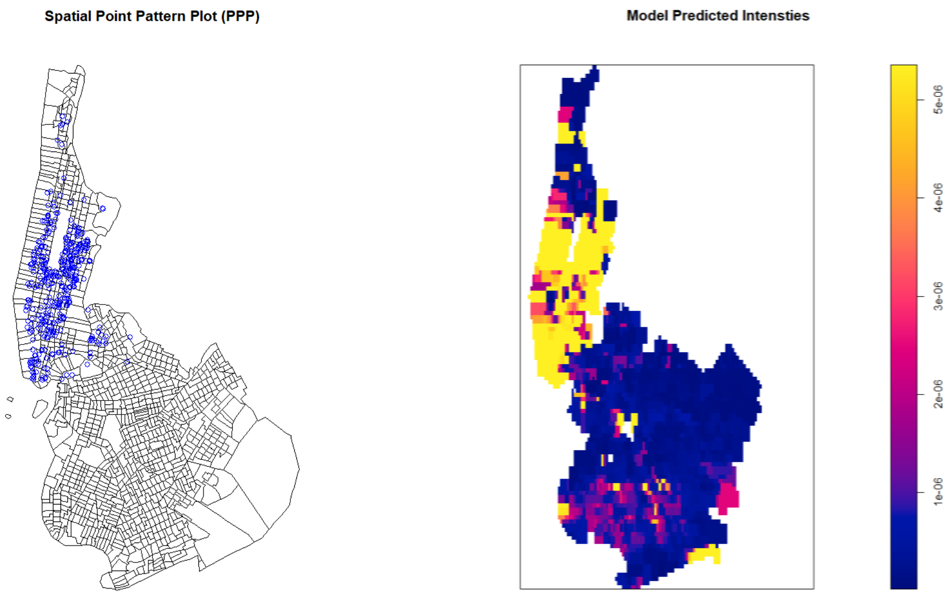



***Figure 6.4***: *New York PPP Spatial Point Pattern (left) and Predicted Model Intensities (right)*

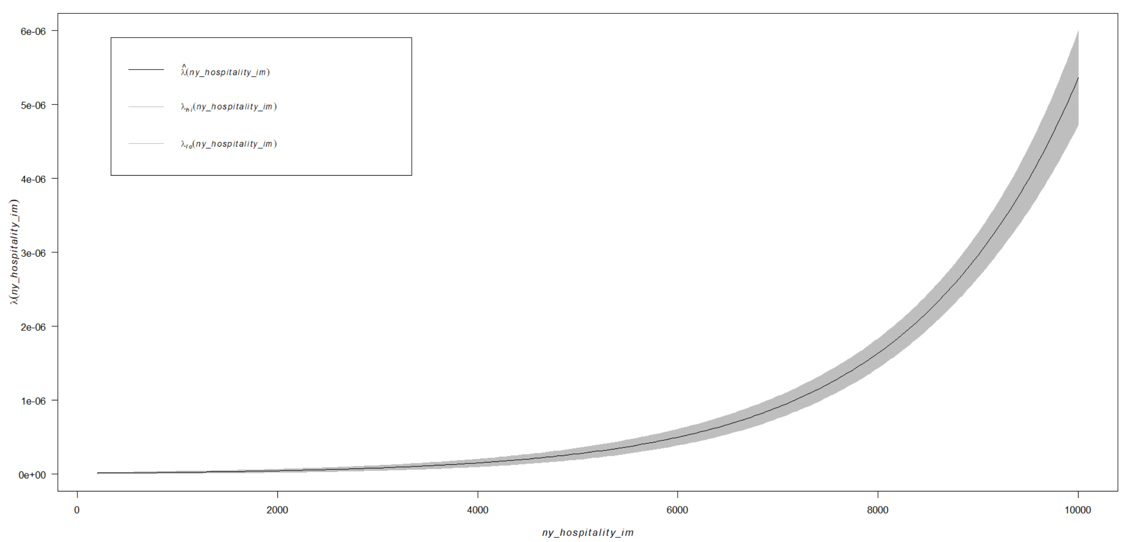



***Figure 6.5***: *Effectfun() Plot of New York Model*

```
Log intensity: ~ny_realestate_im
Model depends on external covariate 'ny_realestate_im'
Covariates provided:
        ny_realestate_im: im

Fitted trend coefficients:
     (Intercept) ny_realestate_im
   -1.805964e+01     5.922299e-04

                     Estimate         S.E.       CI95.lo       CI95.hi Ztest       Zval
(Intercept)      -1.805964e+01 2.893967e-01 -1.862685e+01 -1.749243e+01   *** -62.40445
ny_realestate_im  5.922299e-04 3.153031e-05  5.304317e-04  6.540282e-04   ***  18.78288
```

***Figure 6.6:*** *Model Summary Statistics of New York Model*

Graphically analyzing the model results against the spatial point pattern, it appears the model also captures corresponding areas in Manhattan – particularly around Central Park. The model also appears to highlight high densities in areas of Williamsberg (East of the southern tip of Manhattan). The model summary and effectfun() plot demonstrate a positive relationship between the real estate value (taxes) covariate and intensity.

# CHAPTER 7

# Conclusion and Further Research

Our analysis demonstrates that, for Los Angeles, the presence of hospitality jobs – and related tourism areas – are positively related to charging station intensity. When comparing between the spatial point pattern and the plotted predictions, the intensities of model predictions appear to capture areas in Santa Monica, Hollywood and northern Los Angeles. The model does not, however, capture the same levels of intensity in DTLA and assigns high densities to areas on the eastern edge of the city. From the results, it appears that there are other covariate interactions that may be at play that require further investigation. For New York, the use of real estate taxes paid as a proxy for real estate value appears to better capture intensities as reflected in the spatial point pattern as we can see the key areas of Billionaire's Row (immediately south of Central Park), the Upper and East Sides, as well as regions in the Financial District (southern tip of Manhattan). The model appears to higher densities to areas in the Brooklyn region compared to the observed spatial point pattern. Thus, there may be additional covariate variables to consider for that region.

This analysis was primarily concentrated on the two largest cities in the US [19] by population due to the heavy clustering in both regions within our national analysis. Both areas, while on opposite sides of the country, share many commonalities regarding economic size, wealth, political demographics and metropolitan constitution. There are thousands of additional cities and towns within the US with charging stations that have different characteristics which could lead to different covariate identification. Further research is needed among a broader diversity of areas geographically, economically, and socio-politically to gain greater insights into how charging stations are distributed.

The difference in covariates between the two cities could demonstrate that charging stations are built tailored to each specific city they are in and there may not be a "blanket" strategy across the nation. However, our sample size between Los Angeles and New York is too small and additional research is needed to fully investigate this hypothesis. Further guidance would also include a thorough understanding of the nation's electric grid infrastructure as charging stations are reliant upon availability and feasibility of electricity supply.

REFERENCES


[1] “Scientific Consensus: Earth’s Climate Is Warming.” NASA, 28 Feb. 2023, climate.nasa.gov/scientific-consensus/.

[2} Allen, Myles R., et al. “Summary for Policymakers - Special Report: Global Warming of 1.5C .” Global Warming of 1.5 oC, www.ipcc.ch/sr15/chapter/spm/.

[3] “UN, United Nations, UN Treaties, Treaties.” United Nations, treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27.

[4] United States Environmental Protection Agency, *Global Greenhouse Gas Emissions Data*, www.epa.gov.

[5] Bloomberg New Energy Finance, and The Business Council for Sustainable Energy. 2018 Sustainable Energy in America, data.bloomberglp.com/bnef/sites/14/2018/02/Sustainable-Energy-in-America-2018-Factbook.pdf.

[6] US Department of Energy. “Electric Vehicle Charging Infrastructure Trends.” Alternative Fuels Data Center: Electric Vehicle Charging Infrastructure Trends, afdc.energy.gov/fuels/electricity_infrastructure_trends.html.

[7] Narges Shahraki, Hua Cai, Metin Turkay, Ming Xu. Optimal locations of electric public charging stations using real world vehicle travel patterns, Transportation Research Part D: Transport and Environment, Volume 41, 2015, Pages 165-176, ISSN 1361-9209, https://doi.org/10.1016/j.trd.2015.09.011.

[8] Xiaomin Xi, Ramteen Sioshansi, Vincenzo Marano. Simulation–optimization model for location of a public electric vehicle charging infrastructure, Transportation Research Part D: Transport and Environment, Volume 22, 2013, Pages 60-69, ISSN 1361-9209, https://doi.org/10.1016/j.trd.2013.02.014.

[9] Jia He, Hai Yang, Tie-Qiao Tang, Hai-Jun Huang. An optimal charging station location model with the consideration of electric vehicle’s driving range, Transportation Research Part C: Emerging Technologies, Volume 86, 2018, Pages 641-654, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2017.11.026.

[10] Schoenberg, Frederic Paik. Introduction to Point Processes, www.stat.ucla.edu/~frederic/papers/encycpiece.pdf.

[11] Keeler, Paul. Notes on the Poisson Point Process, 20 Mar. 2018, hpaulkeeler.com/wp-content/uploads/2018/08/PoissonPointProcess.pdf.

[12] Hartmann, K., Krois, J., Waske, B. (2018): E-Learning Project SOGA: Statistics and Geospatial Data Analysis. Department of Earth Sciences, Freie Universitaet Berlin.

[13] Shen Zhang, Xin Liu, Jinjun Tang, Shaowu Cheng, Yinhai Wang, Urban spatial structure and travel patterns: Analysis of workday and holiday travel using inhomogeneous Poisson point process models, Computers, Environment and Urban Systems, Volume 73, 2019, Pages 68-84, ISSN 0198-9715, https://doi.org/10.1016/j.compenvurbsys.2018.08.005.

[14] van Lieshout, M.N.M. and Baddeley, A.J. (1996), A nonparametric measure of spatial interaction in point patterns. Statistica Neerlandica, 50: 344-361. https://doi.org/10.1111/j.1467-9574.1996.tb01501.x

[15] Chen, Yen-Chi. "A Tutorial on Kernel Density Estimation and Recent Advances." arXiv.Org, 13 Sept. 2017, arxiv.org/abs/1704.03924.

[16] Kile, Hakon. Bandwidth Selection in Kernel Density Estimation - NTNU, May 2010, ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/258553/348940_FULLTEXT01.pdf?sequence=1.

[17] Zambom, Adriano, and Ronaldo Dias. A Review of Kernel Density Estimation with Applications to Econometrics, 2012, dergipark.org.tr/en/download/article-file/259389.

[18] Chen, Yen-Chi. Density Estimation, 2017, faculty.washington.edu/yenchic/17Sp_403/Lec7-density.pdf.

[19] Bureau, US Census. "City and Town Population Totals: 2020-2022." Census.Gov, 15 May 2023, www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html#v2022.

[20] NYEDC. "New Yorkers and Their Cars." New Yorkers and Their Cars | NYCEDC, 5 Apr. 2018, edc.nyc/article/new-yorkers-and-their-cars#:~:text=Ownership%20is%20lowest%20in%20Manhattan,83%20percent%20of%20all%20households.

[21] US Census. "Census Data API: /Data/2021/ACS." Developers, api.census.gov/data/2021/acs.html.