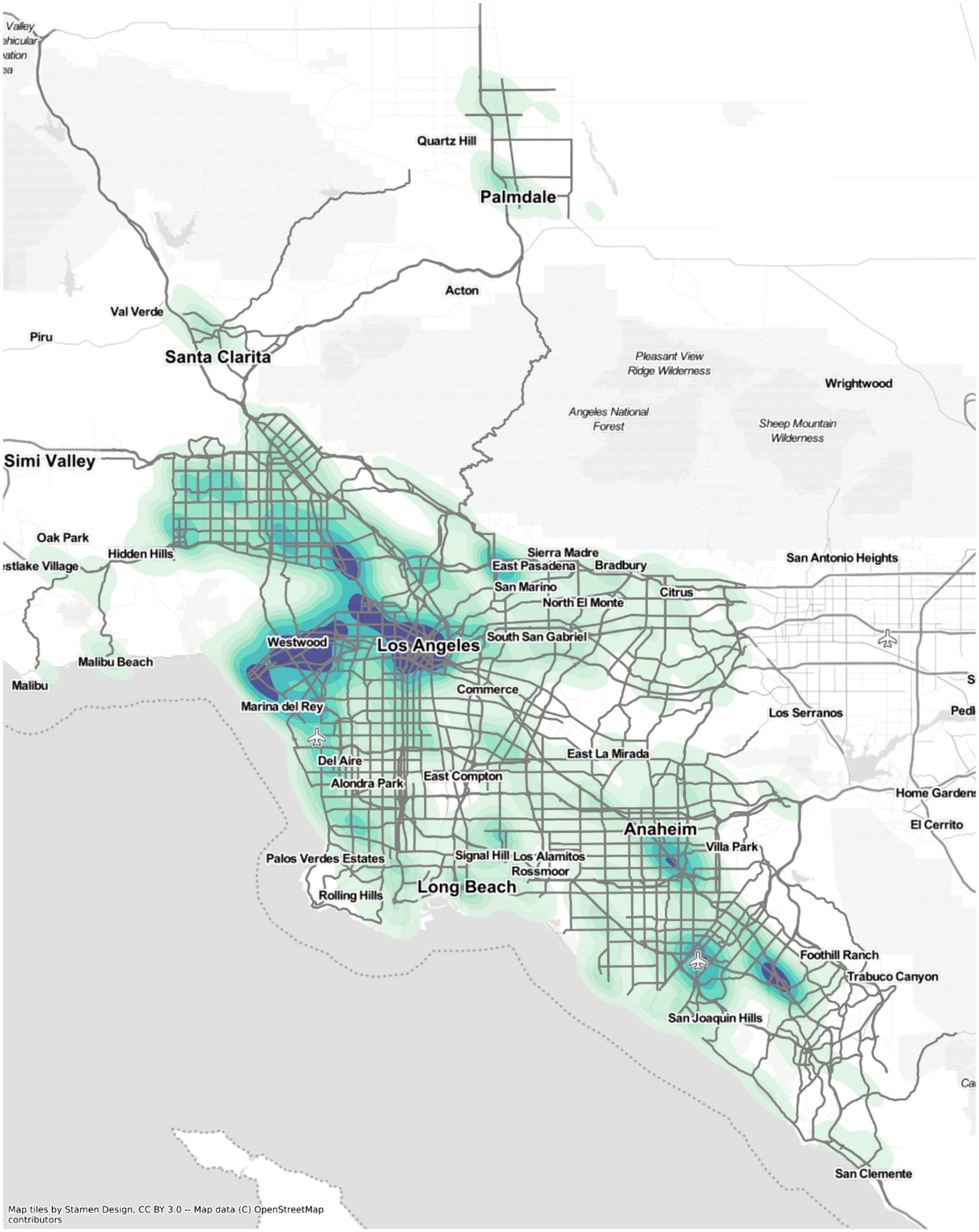
Southern California Electric Vehicle Charging Station Network Analysis

Authors: Jarvis Yuan, Justin Wong, Danny Ha, Jing Xu



I. Abstract

In this project, our group examines the **EVCS (Electric Vehicle Charging Stations)** with major roadway street networks in the urban Southern California region to extract its complex network properties and to find the correlation between various data attributes using data science methods. Some data inputs include the 2019 American Community Survey (ACS) census data for demographic correlation, electric vehicle (EV) registration as well as Longitudinal Employer-Household Dynamics (LEHD) commute flow data for estimating local charging demand. Through this data aggregation, we find many electric vehicles to be located in West Los Angeles and in Orange County where higher income individuals also reside. In addition, we find an imbalance in the number of job destinations with the number of chargers available in some Zip Codes. Using open-source python packages, we clustered the EVCS into six groups within our study region to study their correlated characteristics both demographically and geospatially.

II. Introduction

Motivation and Background

In September of 2020, California's Governor Newsom announced an executive order requiring sales of all new passenger vehicles to be zero-emission by 2035. Progressive laws are pushing for a faster transition to electric vehicles (EVs) in light of increasing CO2 concerns in the transportation sector. However, before phasing out internal combustion engine vehicles (ICEVs), many improvements must be made in the existing transportation and utility sectors, such as the ability to balance and anticipate demand in the electric grid, battery technology relating to the "range-anxiety" consumer mindsets, and the current high upfront cost of affordable EVs. Furthermore, the efficient management of the electricity grid will be critical. Meanwhile, there does not seem to be an equal distribution of EVCS throughout California.

In order for us to transition away from ICEVs, we need a reliable network of EVCS in terms of equity, accessibility, and meeting local charging demand while preventing local grid overload. The electrification of transport is in progress and will transform public commute and our basic need of simply getting to places. Unlike bus/subway stations, charging points in the urban settings are relatively mobile and flexible depending on the demand and the technology, which make them a very interesting network analysis problem that connects the need of people to the optimization of the operation.

However, EVCS alone do not provide much information on neither its accessibility nor the commute and the public charging demand they are addressing. Since most public EVCS are driving-accessible and located at street-side parking lots, the centrality measures on the nearby street nodes could be approximated as the centrality measures of the EVCS on the network, which would entail more information on the nodes' connectedness with network science. This motivated us to also take a closer look at more granular demographics on each charger and their characteristics that might correlate with the street network performance. By picking and incorporating a set number of 2019 ACS census data as our variables, we attempt to cluster the EVCS within the study region into groups and closely examine their characteristics both demographically and geographically.

Network Description

The geographic focus of this study is in the Southern California bounding box; the South West corner is (33.311935, -118.793825), while the North East corner is (34.507458, -117.592195). In this network study, we define two types of nodes, A and B. Type A nodes are street network nodes, such as intersections, on-ramps, junctions, etc. Meanwhile, Type B nodes are EV/PHEV charging stations queried in the bounding box region.

We built an initial network demand analysis of the EV charging stations in Los Angeles and Orange County. We used the ”primary” and ”motorway” OSM streets within the boundary box to build a directed spatial network to approximate the actual street network, adding physical distance as edge weights and intersection, on-ramps, and exits as the network nodes. After this initial network creation, there was a lower bound of 12,000 nodes within the street network.

We then overlay the EV/PHEV registration share (either the absolute share in count or percent share among all vehicle types) by Zipcode from California Department of Motor Vehicles and observe the correlation between the EV/PHEV registration distribution and the street network centrality measures. Additional analysis incorporated the socio-economic census tract data to provide more insights.

One limitation of our approach is that it does not entail the full picture of the charging infrastructure demand, as most of the charging infrastructure are built for long-distance trips and workplace charging in mind. The ideal approach would be to simulate the trip distribution on the given network. However, due to our constraints on computing power and calibration abilities, we simplify the trip distribution by using regression methods to estimate the demand of EV/PHEV trips on analysis geo-units (census tract/zip code) using the share of EV/PHEV vehicles based on the DMV registration data and Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination (LODES) data.

Relevant Literature and Previous Work

Multiple studies have been conducted in attempting to determine optimal EV locations in an urban city and analyze EV charging station networks.

Sebastian Wagner, Markus Götzinger, and Dirk Neumann use a point of interest approach to determine optimal locations of EV charging stations. They analyzed charging stations in Amsterdam by gathering data on utilization rates and surrounding points of interests to determine weights for maximum coverage optimization to determine optimal locations.

Yunfei Mu, Jianzhong Wu, Nick Jenkins, Hongjie Jia, and Chengshan Wang compare mapping and routing algorithms that route to the nearest charging station. They investigate how different technologies and algorithms behind navigation APIs differ from each other, focusing on routes that lead to electric vehicle charging stations. They found that algorithms used a variety of ways that consider real time availability information of charging stations, prioritizing highways, calculating temperature and altitude impact on the battery, and even certain charger types, such as Tesla’s superchargers.

Yunfei Mu, Jianzhong Wu, Nick Jenkins, Hongjie Jia, Chengshan Wang use a spatial temporal model (STM) that runs based on systematic integration of power system analysis and transportation analysis. The STM provides average and probabilistic values to determine critical network components to upgrade to offset the negative grid impacts from plug-in electric vehicles. This study primarily evaluates the impact of EV charging on power systems using the STM method.

Yanyan Xu, Serdar Çolak, Emre C. Kara, Scott J. Moura, and Marta C. González present a method to estimate the individual mobility of plugin EV(PEV) drivers at fine temporal and spatial resolution by combining various datasets, such as mobile phone, census, and PEV survey data. They provide statistical and visual representations of the EV demand forecast while coupling the charging profile with urban mobility. In this paper, the trip distribution simulation was performed and calibrated with the CVRP data, but the focus was more on the complex topic of charging behaviors to address a gap in the existing PEV management literature.

Focus of This Study

In this study, we look into Southern California's street network with the EV/PHEV charging infrastructure, focusing on the EV charging demand (work & home) and existing charging stations to investigate charging station accessibility and its relationship with demographic information.

III. Data and Methods

Figure 1. Study Area in Bounding Box, Southern California Street Network

Driving Street Network

Since we are hoping to evaluate the network performance of the EVCS, we first started with constructing the driveable street network in the study area. Using OSMnx, we queried the street network within a custom bounding box that largely joins Los Angeles County, CA and Orange County, CA for the analysis (Figure 1); the South West corner was (33.311935, -118.793825), while the North East corner was (34.507458, -117.592195). This spatial extent includes various landscapes, dense metropolitan and residential areas, and different demographics, providing us with a rich background for our analysis. Due to computation power and memory constraints, we limited our street network to only include primary and motorway type streets, according to OSM standards, but the entire network still largely resembles the urban commute network in the Southern California region.

Our nodes in the network represent intersections and joints between roads and the network like onramps, and the edges represent OSM streets, which are links between street nodes represented by LineString objects. For the analysis, we also appended the EVCS onto our street network as nodes to approximate their centrality performances according to that of their nearest nodes, which method will be detailed in later sections.

The queried MultiDigraph network had 14,198 nodes and 24,089 edges, which includes a significant amount of residential street nodes and four-way intersections. Following Geoff Boeing's guide on [simplifying graphs and consolidating nodes](), we are able to reduce the complexity of the network down to 4,025 nodes with 8,871 edges. Although this significantly improved the computational efficiency on running centrality analysis, it came at the cost of losing some degree of accuracy in terms of topology of the network. Since we are using a higher-level network for performance analysis, such effect is outweighed by the benefit of a more efficient analysis.

Overall, our street network covers nearly 16,000 square kilometers in area, spans 2,152 intersections, each node has an average degree of 4.4, and stretches near 100 miles north-south and 65 miles east-west.

EVCS and Street Network Centrality Approximation

EVCS data is queried from [Open Charge Map](), an open source project that provides an API to consume and contribute charging station data in the entire world. We then obtained the EVCS needed for the study by filtering all charging stations within the bounding box in the Southern California region. There were in total 3,317 unique charging stations within our study region.

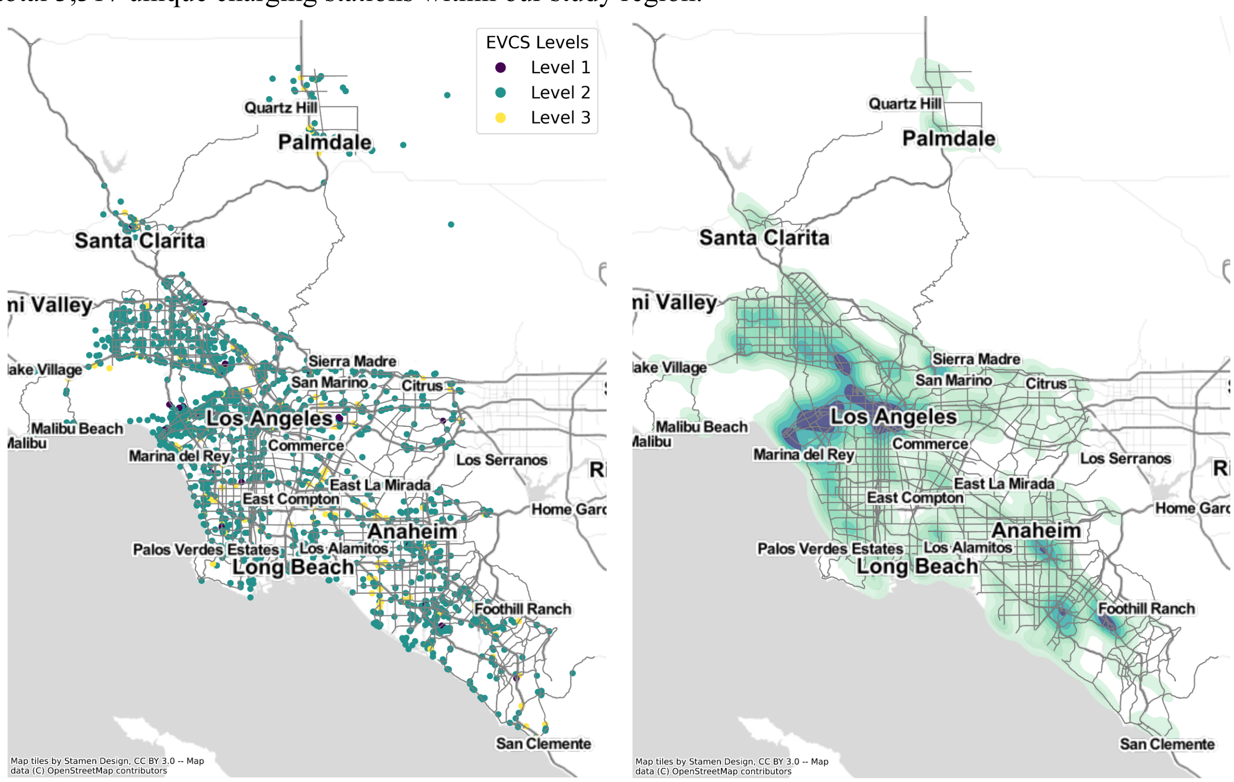


Figure 2. Southern California Charging Stations and Kernel Density Clustering with City Basemap. On the left, the individual dots correspond to individual stations categorized in color on their charging levels. On the right, a kernel density heatmap is displayed to better understand the geographic distribution of EVCS.

From the kernel density plot on the right, we could clearly see that the EVCS in the study region are mostly clustered into two large pockets: One in the downtown Los Angeles area and Westwood, and the other near Irvine in Orange County, CA with two nearby small pocket at downtown Orange and John Wayne Airport next to I-405 in Santa Ana. With this distribution in mind, we expect to see the density of the chargers to correlate with other data sources like population demographics as well as network centrality measures, which we will be exploring in the results section of the report.

EVCS are generally divided into three types of charging levels (Level 1, 2, and 3), with different associated charging rates, typical use cases, and supporting infrastructures. The first EV charging level is the basic Level 1 charger. A Level 1 charger is simply charging from a standard 120V household outlet, which only provides on average 4 to 5 miles of range per hour. Level 2 chargers take in a voltage supply over 200V, and will charge a typical EV at a rate between 12 to 60 miles of range per hour. These chargers are suitable for office/workplace charging and also overnight charging for commuters living in the suburbs. Usually, a small transformer would need to be installed at households/parting lots to achieve a high voltage to enable the Level 2 charging rates. Most EVs accept Level 2 charging. Lastly, Level 3 chargers, also known as "DC fast chargers" use direct current (DC), which is different from the alternating current (AC) that is available in households and commercial buildings. DC charging is available at a much higher voltage and can charge some plug-in electric vehicles with as high as 800V. This allows for very rapid charging but comes at a cost of higher expenses for the infrastructure and higher stress for the grid. Currently, Tesla Level 3 Superchargers dominate the 6% share of all Level 3 chargers in the SoCal study area, while most other public EVCS are rated at Level 2 at 93% (JuiceBlog).

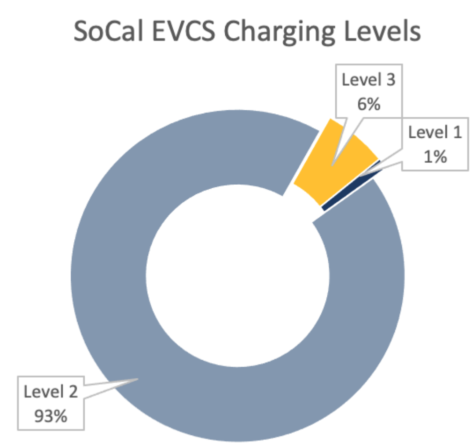


Figure 3. SoCal EVCS Charging Levels

In order to approximate the network centrality measures (betweenness, closeness, and degree) of the EVCS, we first computed all node centrality of the consolidated street network and added to the nodes as attributes. This process is quite computation intensive given the large scale of our street network, but we are able to benefit from consolidating the intersection and reduce this analysis from nearly 15 minutes to around 2 minutes. Then, we unified the network projection and the coordinates of the EVCS to make sure we could correctly locate the nearest neighbor on the same map projection. Using the built-in function from OSMnx, we are able to locate the street network nodes that are nearest in haversine distance after the projection for each and every EVCS point using an iterative approach. Finally, we extracted the nearest neighbor's centrality measures and appended to the EVCS as a reasonable approximation. The network centrality provided us with an extra layer of information that is inherent to the street networks yet not apparent from the EVCS data (on charging level, power, and counts) nor the census data.

EVCS K-Means Clustering Analysis

We investigated how the different datasets correlated with each other. We followed the following process for our data analysis pipeline:

1. Select relevant census variables for clustering exploration, in addition to the selected attributes of the EVCS and network centrality approximations;
2. Extract the ACS 2019 census data on the census tract level for the two major counties in the study region, LA County and Orange County, CA;
3. Merging their census data frame together geo-spatially, and clip within the custom bounding box to keep the analysis' consistency;
4. Spatially interpolate the EVCS data that falls into the shape boundary of its census tract. Each charging station node thus contains the census demographics of the tracts that they represent;
5. Using the Elbow Method to determine the optimal value of k, the number of clusters, for the K-Means clustering algorithm;
6. Run a K-means algorithm to cluster the EVCS based on their collected charging attributes and census attributes;
7. Combined with traffic/commute flow (LEHD) and the clustering of the census tracts themselves providing a background, we hope to extract more insights on the accessibility and demographic backgrounds of the EVCS within the SoCal study area.

After cleaning and processing the clustering variables, we plotted their spatial distribution as a reference for the clustering analysis, which you can refer to in the Appendix section, Figure A1. Further analysis is discussed in detail in the Results section of this paper.

ACS 2019 Census Tract Data and EVCS Clustering

Using the [Cenpy Python library](), we obtained ACS 2019 census tract data in Orange County and Los Angeles County and combined them for our clustering and geospatial analysis. There are a total of 2,851 unique census tracts and 6 million in population.

Some of the initial variables obtained include total households, total populations of different ethnicities (White and Hispanic), total population with bachelor degrees, median age, total workers over 16 that drive alone to work by some vehicle, median household income, and allocation of travel time to work. To ensure the quality of our input data, we begin our data processing pipeline by replacing invalid and empty values by the column means. To better represent the data and reduce the skew due to the unnormalized absolute values, we convert most of our variables into percent share of the total tract population to compare on the relative scale.

The following sections describe the breakdown of the final clustering variables extracted from 2019 ACS data and SoCal street network:

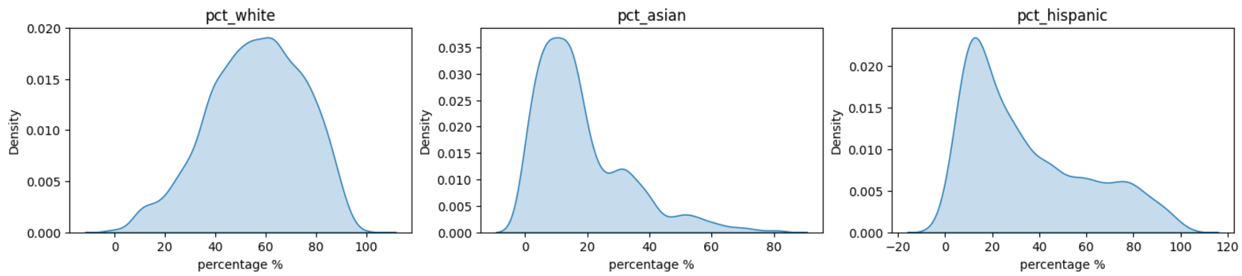


Figure 4-6. Distribution of Percent Population of White, Asian, and Hispanic by Tract onto EVCS

Percent Population of White, Asian, and Hispanic

Racial distribution would likely be a consideration for the planning process of EVCS as it usually correlates with charging demands and income.

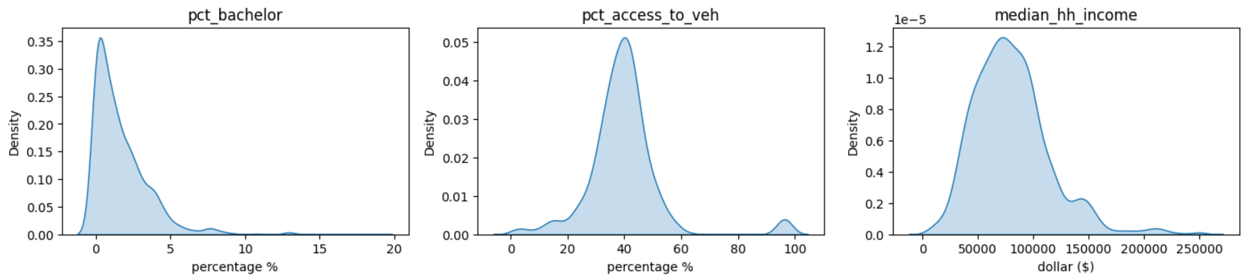


Figure 7-9. Distribution of Percent Population with Bachelor Degrees, Access to Vehicle, and Average Household Income by Tract onto EVCS

Percent Population with Bachelor Degrees

Education level is also another strong indicator of wealth and income, and also correlated with higher EV adoption rates among others (Benjamin K. Sovacool et al). From Figure 8, we could see that we have a median of < 5% across all tracts in the study area with a long tail to the right and capped at around 20%.

Percent Population with Access to Vehicle

Higher percentage of share of the population with access to vehicles would also suggest a higher demand for vehicle usage for commute and personal trips, which brings EV charging demands for both commuters and residents. Within our study area, the distribution from Figure 9 shows a median of 40% population with access to vehicles.

Median Household Income

According to Quartz, the average cost of a new car in June 2019 in the U.S. was $36,600, while the average cost of an electric vehicle was $55,600 (Hearst Autos Research). Therefore, a higher household income would likely bring more purchasing options and power of EVs for families considering the next new vehicle, which is likely to correlate with a higher EV and public charging demands. From Figure 10, we can see that the median household income ranges from around > $50,000 to < $250,000, which is likely to be the main distinguishing factor among the clusters.

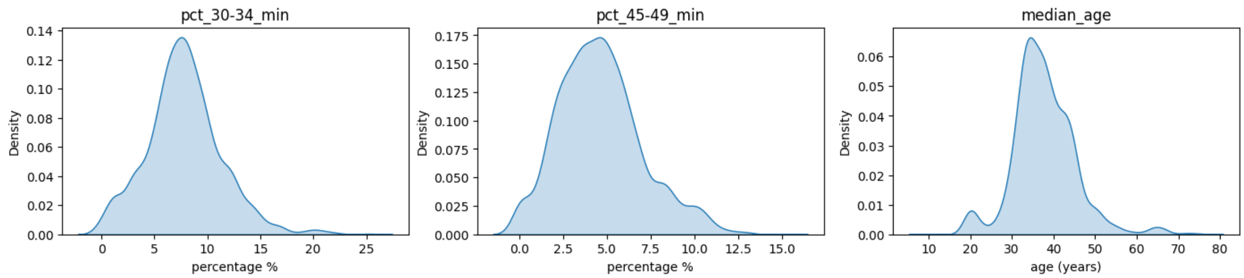


Figure 10-12. Distribution of Percent Population with Commute Time of 30-34 minutes, 45-59 minutes, and Population Median Age

Percent Commute Time = 30-34 minutes and 45-39 minutes

The average one-way commute time among the commuting population in Los Angeles is 32 minutes (Leonard). These two variables would present the opportunity to cluster other variables in relation to a span of commute times. To reflect the commute density, we would ideally be querying data for the destination census tract, as most workers usually converge to a number of office plaza "hubs" from the suburbs. Due to the time constraints and data availability, we decided to work with origin commute data provided by the ACS 2019 for this analysis

Population Median Age

Since our study area spans multiple urban settings, we are expecting to see a spatial variation of population median age, which is likely to be correlated with the demand of EV charging.

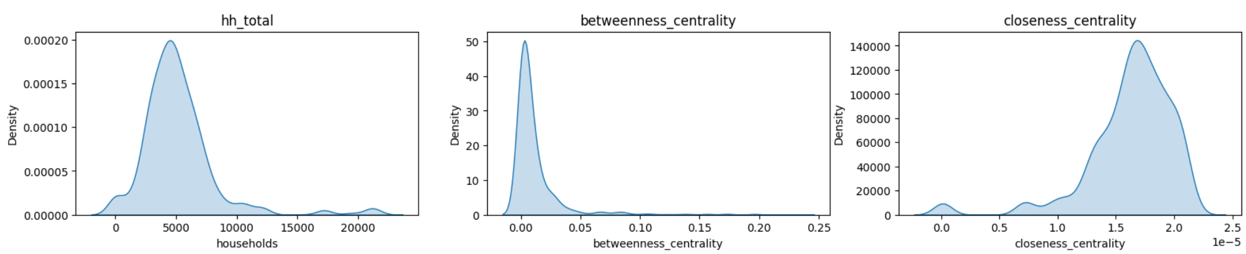


Figure 13-15. Distribution of Total Households, Approximated Betweenness and Closeness Centrality.

Total Household

Total household count is included as an indicator of population density and to help separate suburban settings from metropolitan settings.

Network: EVCS Betweenness and Closeness Centrality Approximation

As detailed in the above section on EVCS and Street Network Centrality Approximation, we approximated the EVCS node betweenness and closeness centrality with the nearest neighbor method since chargers are largely attached to the street network in reality. Betweenness measures node importance when considering shortest paths while closeness centrality quantifies the ease of access to all other nodes in the network. Street nodes with high betweenness usually reside on major roadways like freeways and street nodes with high closeness often correlate with downtown areas and hubs at large freeway junctions.

Since our ACS 2019 census data were queried on the census tract level, we need to interpolate them onto the EVCS to prepare for the clustering analysis. Using the geopandas Python library, we merge the EVCS attributes with the census tract variables whose boundary contains the coordinates of the EVCS. We expect there will be multiple EVCS nodes falling into the same tracts thus having repetition of the census datas, but it is not a concern for the analysis given the large number of both census tracts & EVCS within the study area. After the tract-to-EVCS interpolation, we are able to run a K-means algorithm to cluster the EVCS based on their charging attributes *and* census attributes, which would further provide insights on who and how these EVCS serve its population.

Combined with traffic/commute flow and the clustering of the census tracts themselves providing a background, we hope to extract more insights on the different clusters of the EVCS within the SoCal study area.

DMV EV Registration Data

Vehicle population data was downloaded from the DMV. The figure to the right shows the fuel type of the different vehicles in the SoCal region. 83% of vehicles were gasoline, which is a large problem when relating back to the decarbonization of the transportation sector. However, it is worth noting that this may be skewed by the possibility that many vehicles registered are not running anymore.

For the purposes of our analysis, vehicles were filtered to contain only light duty vehicles as the Census data we planned to analyze would be based on residents and not businesses which typically have heavy duty vehicles on the road.

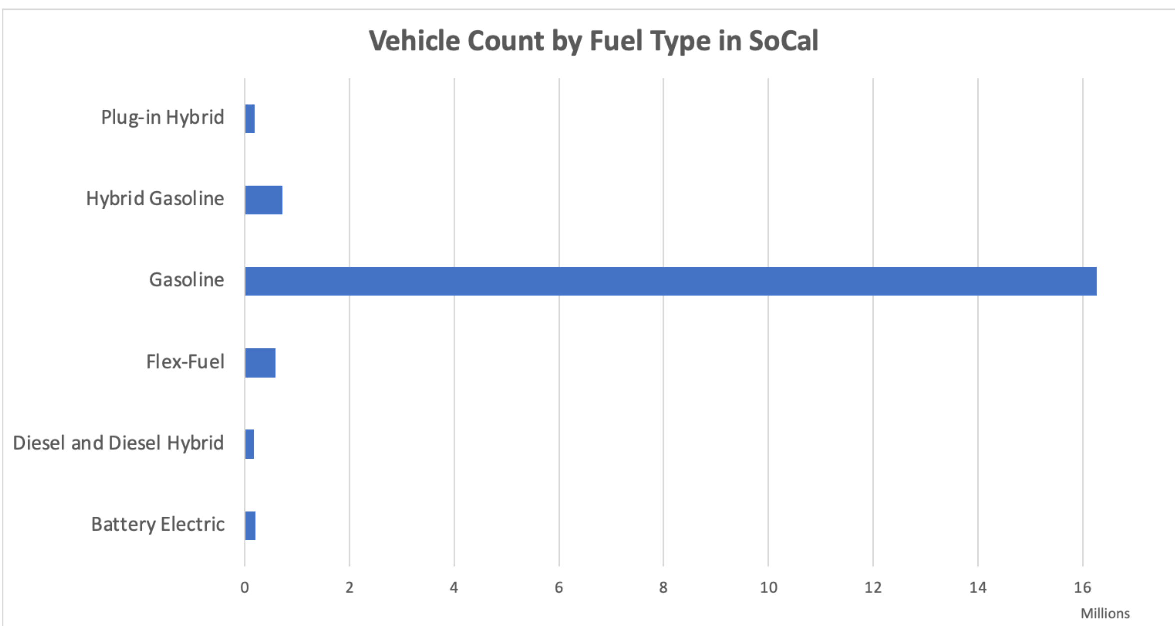


Figure 16. Fuel type by share of vehicles in Southern California, as reported by California DMV as of January 1, 2020.

In addition, we defined 'Battery Electric' and 'Plug-in Hybrid' to encompass EVs in our analysis as they are the two fuel types for vehicles requiring electric chargers.

For some introductory analysis, the vehicle registrations were aggregated by Zip Code obtaining the total number of EVs, total number of vehicles, and the percentage of EVs in each Zip Code. The map below visualizes the aggregated information.

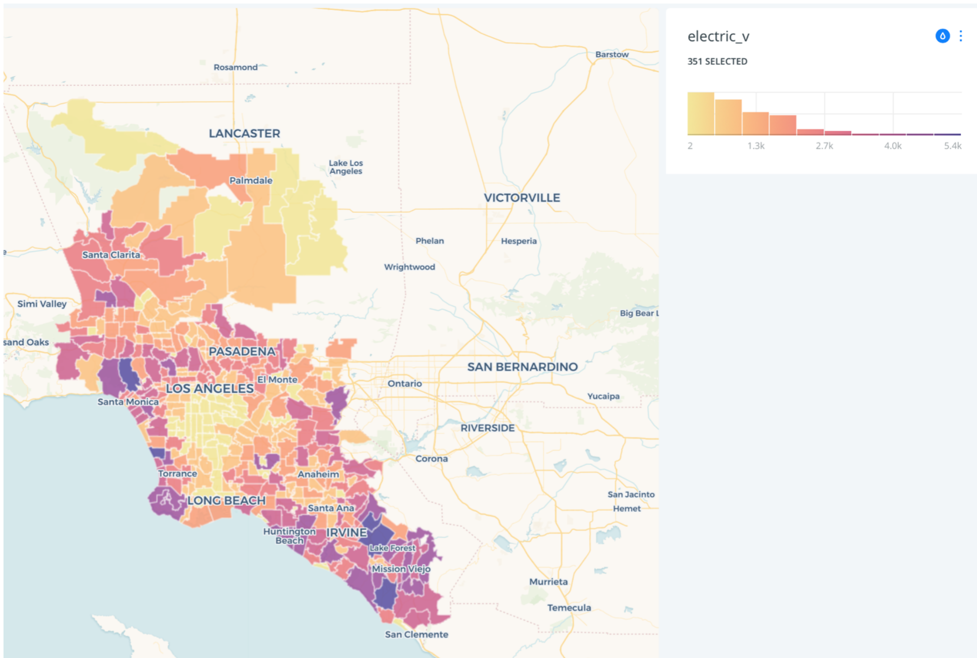


Figure 17. EV vehicle registrations by Zip Code.

There are large concentrations of EVs in West Los Angeles and Orange County areas like Irvine. There are weak concentrations of EVs in Central and Downtown Los Angeles. It is important to recognize that Zip Codes do not have the same number of residents and also car owners and thus the number of EV registrations can largely be due to greater populations. Nonetheless, the map can be used to see general concentrations of EVs in the SoCal Region.

Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Data

The DMV data enables us to look at where the EVs are stationed but does not capture the movement of them. Movement of EV vehicles is important because a good EV charging network accommodates travel and commute patterns.

LEHD Origin-Destination Employment Statistics (LODES) uses Census data to get the number of those that reside in a said Residence Census Block and work in a said Workplace Census Block. Although we cannot assume that all these workers drive to their workplaces, we can use LODES data to get a gauge of general work commute movement, some of those of course being EVs.

To stay consistent in granularity, the LODES data was interpolated to Zip Codes by mapping the centroids of tracts to Zip Codes. The following map was created aggregating the Workplace Zip Codes which shows the Zip Codes where people are commuting to. We can see that there is lots of commuting to Orange County Zip Codes as well as Northwest, West, and Downtown Los Angeles.

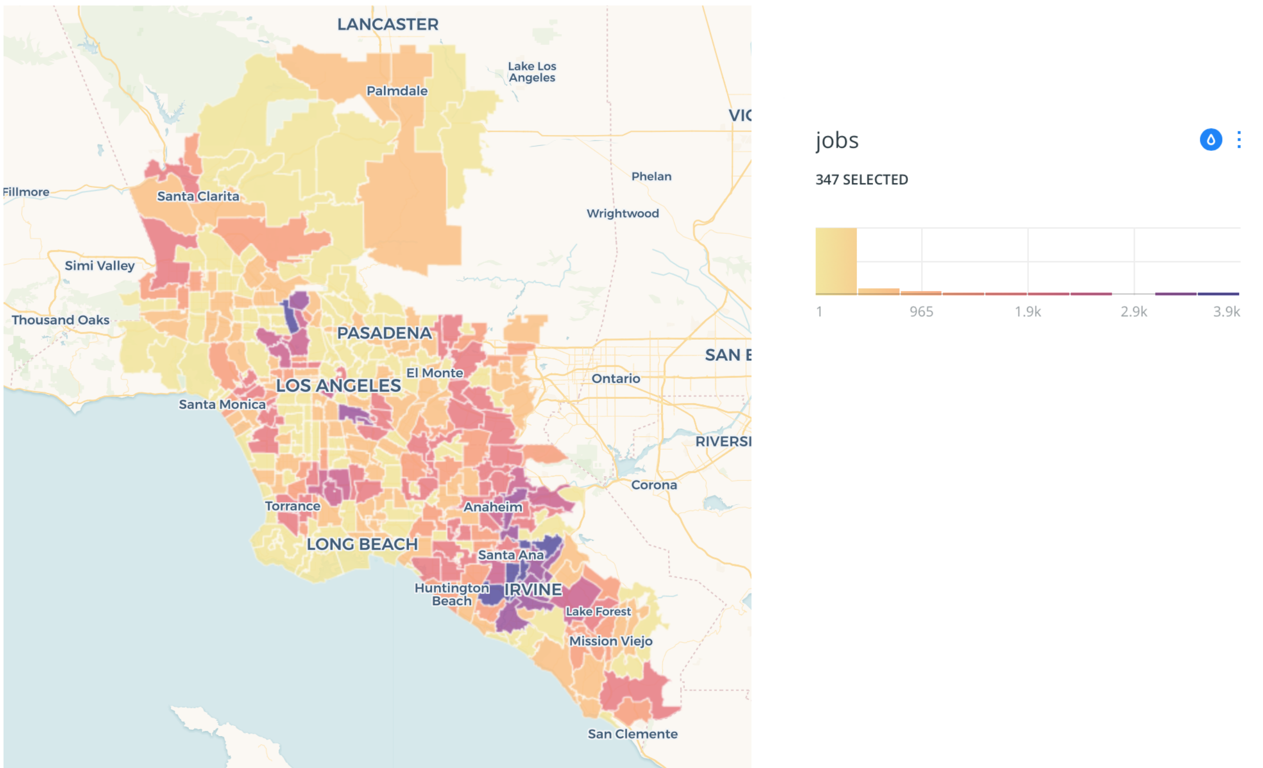


Figure 18. Map showing the distribution of jobs in Zip Codes.

IV. Results

Network Analysis

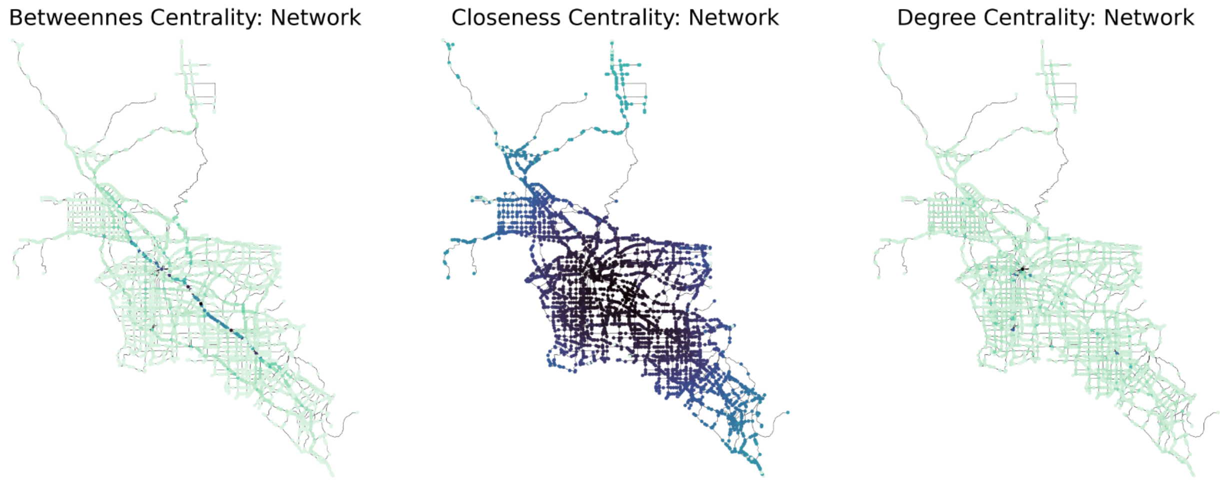


Figure 19. Southern California Street Network, Betweenness, Closeness, and Degree Centralities

After appending the centrality measures onto the network edges, we can clearly see that the I-5 freeway routing through the central Orange region has the highest betweenness centrality, which indicates a high utility and usage for trips taken in the area. It also indicates that the I-5 has a considerable influence on the network since more paths go through the freeway, which is also in line with the reality where congestion during the peak periods tends to take place on the I-5 freeway. However, the connection with traffic congestion seems to end there, as other highly congested major freeways during peak periods like the I-405 and I-10 do not show clear indications of high betweenness on the street network. The urban and downtown areas show a high degree of closeness centrality as more street nodes are clustered around these settings. However, only a few street nodes at major junctions and Los Angeles downtown areas indicated a high value of degree centrality measure, which is not very much helpful in determining the importance and connectedness of this street network.

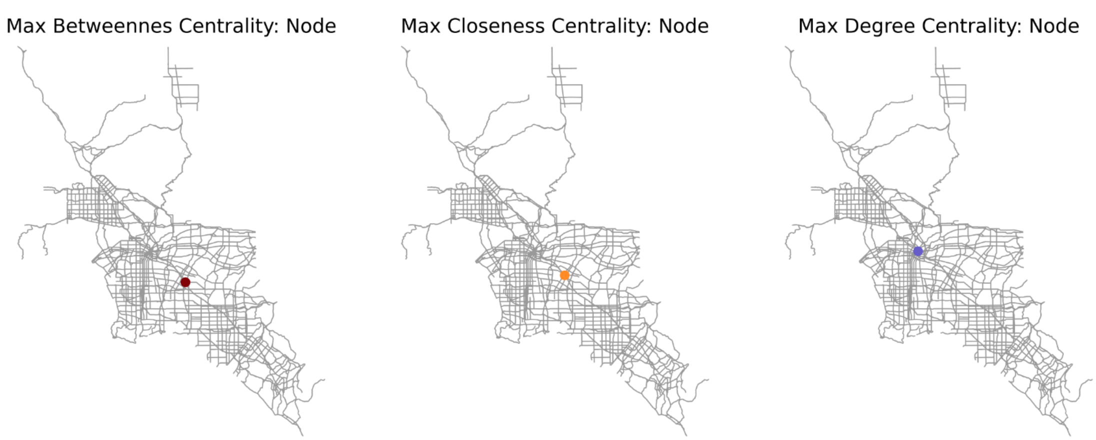


Figure 20. Street Nodes with Highest Centrality Measures

LEHD Origin-Destination Analysis

Combining findings from the preliminary LEHD Origin-Destination Analysis with data of the EVCS, a ratio of number of chargers to number of jobs in a destination Zip Code was conducted. We have decided to use the number of jobs in a destination Zip Code to signify some level of travel demand to that Zip Code. We recognize that work commute is not the only traveling going on and that there are lots of leisure driving.

However with that said, we have found an interesting cluster of high charger-to-job ratio in West Los Angeles and in the Long Beach area. There are lower ratios in the yellow areas in the majority of South Los Angeles and Orange County. Alluding back to Figure 17, Orange County held the highest EV ownership which leads to a hypothesis that this area is underserved in EVCS and could use more. West Los Angeles seems to be meeting demands currently and we assume in the future as ownership rises as it has a high charger-to-job ratio and given it already has higher EV ownership seen in Figure 21.

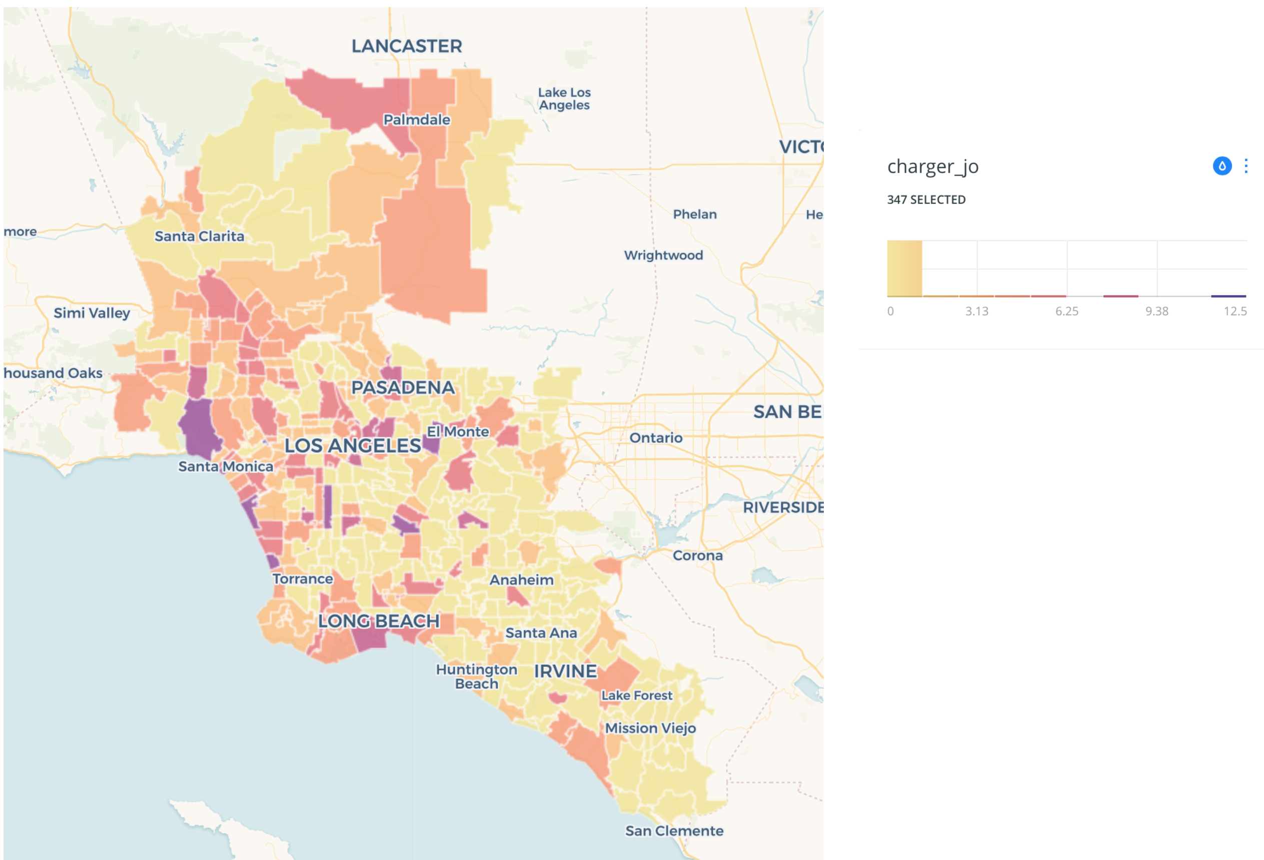


Figure 21. Map showing charger-to-job ratio per zipcode from LEHD Origin-Destination data

Exploratory Data Analysis

In the electric vehicle charging stations (EVCS) data, we plotted a kernel density plot overlaid with city names and the street network. Each dot is a charging station, while the heatmap shows the density of charging stations throughout the region and the color of the dots represents the charging level, which ranges from 1-3. There is a very clear cluster of charging stations in Los Angeles and Westwood, which we see clearly overlaps many streets.

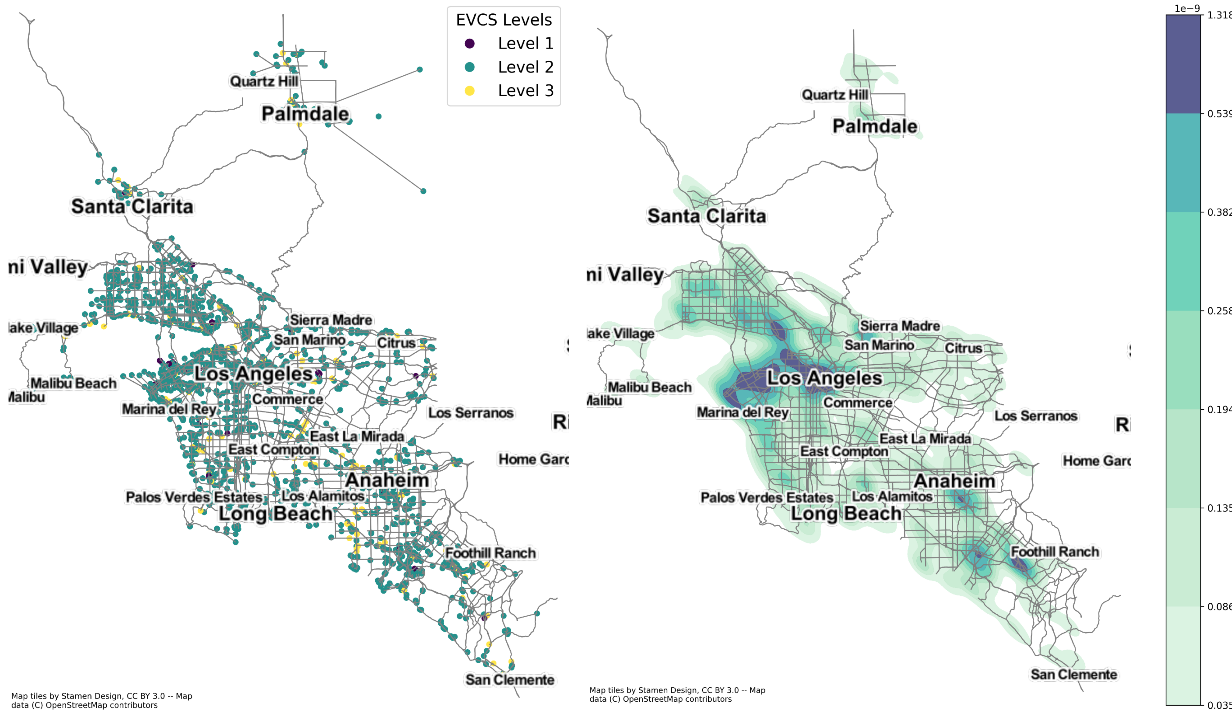


Figure 22. Appended EVCS Street Network (left) and EVCS Kernel Density Map

As we look at the shortest path lengths, we see a long right tail curve. In this right skewed graph, it is worth noting that the mean shortest path distance is right of the mode shortests path distance. This seems to indicate that most paths are fairly easy to get to, which is a good attribute when trying to make EV charging stations easily accessible.

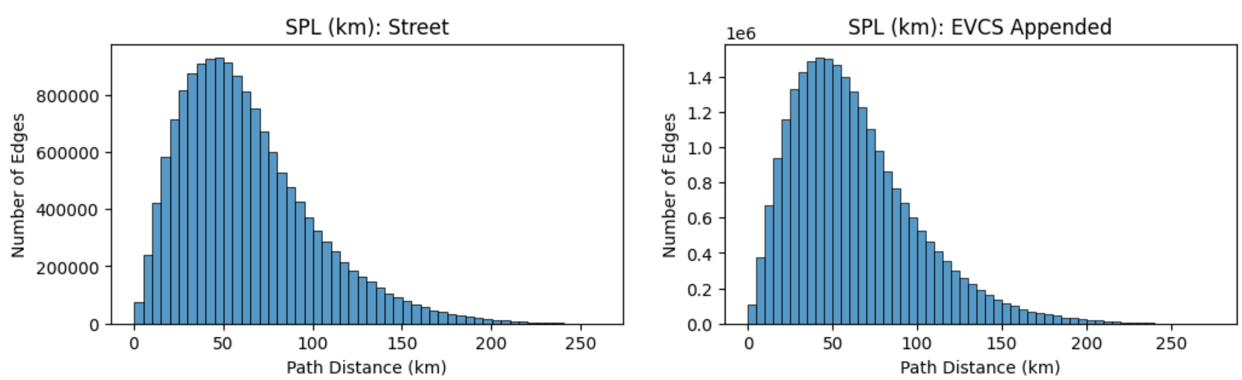


Figure 23. Physical Shortest Path Distribution: Street Network Only (Left), EVCS Appended Network (Right)

Plotting the physical shortest path distributions comparing the original street network (Figure 23. Left) and the EVCS appended street network (Figure 23. Right), we noticed that the overall shape and spread of the distribution are very similar. However, the appended street network has almost doubled the density for each bin. This is expected since each EVCS is added to the nearest street node with haversine distance in between as the edge distance. Since we have 3,317 unique charging stations and 2,851 unique census tracts in the study area, the doubling of density that we've observed is reasonable given the a particular shortest path from A street node to B street node is relative the same that from A1 EV node to B street node, given A is the nearest neighbor node to A1.

We were also interested in seeing if the network qualified as a "small world". We created a Watts Strogatz graph with the same network property as the SoCal Street Network, which yielded the following table. The average lengths significantly differ and the clustering coefficient is very low, which indicates the network does not resemble a small world network. In other words, most nodes are not neighbors.

	#nodes	**#links**	**<C>**	**<K>**	**<L>**
network					
SoCal Street Network	16753	26032	0.026921	3	51.618255
Small World Network Model	16753	16753	0.000000	2.0	192.015645

Figure 24. "Small World" Network Attributes. This network does NOT resemble a "Small World"

EVCS K-Means Clustering Analysis

On the left, the light blue dots are where charging stations exist. Meanwhile, the right figure shows the percent of the census tract with access to vehicles. By laying these figures adjacently, we can observe that areas with less access to vehicles have fewer charging stations.

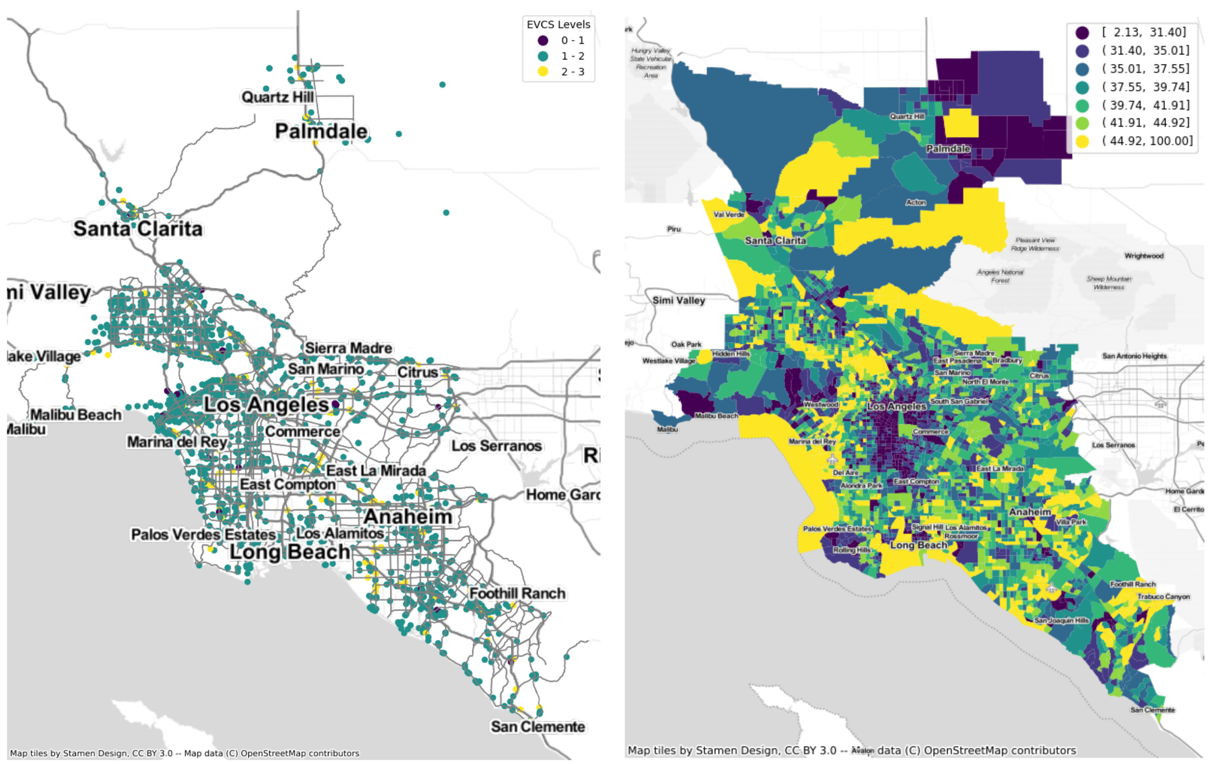


Figure 25. Adjacent comparison of EV Charging Stations(left) and Access to Vehicles, defined as "percentage of workers over 16 that drove alone to work by car, van, or truck" (right).

Since we have a total 3000+ EVCS nodes in the study area, we are able to run a K-means algorithm to cluster the EVCS based on their charging attributes and census attributes, which would further provide insights on who and how these EVCS serve its population. The next page shows the results of such clustering. Comparisons can be made across any of the variables. Some notable overlays include:

- Population and Connections quantity: Areas with higher populations tend to hand have more charging stations. This makes sense because areas with more people would correspond to higher EV charging station demand and thus require more charging stations.
- Income and Connections quantity: Areas with higher income have more clusters of charging stations. Such finding shows that higher income communities tend to have higher EV adoption rates
- Ethnicity and Connections quantity: White and asian populations seem to have the biggest overlap with charging station quantity. Hispanics have some clusters that overlap with charging station quantity. Black population appears to have the biggest difference. This has significant implications for environmental justice and involving diverse populations in the fight against climate change.

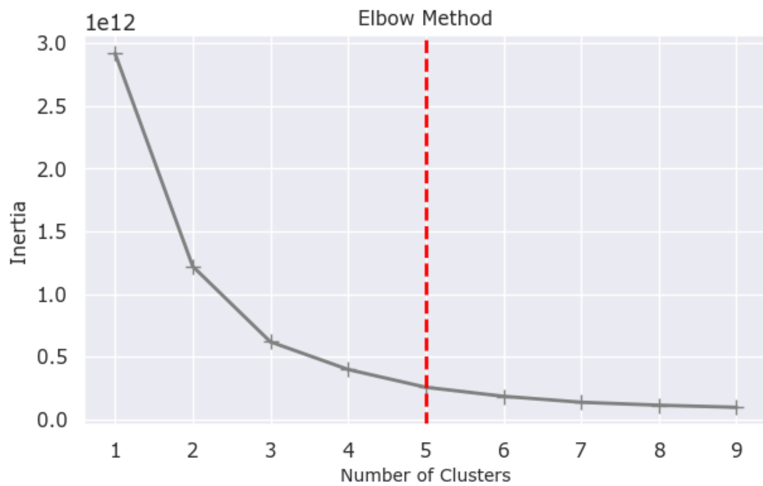


Figure 26. Elbow method for Determining the Optimal Number of Clusters

Using the Elbow method, a heuristic used in determining the number of clusters in a data set, we are able to set on K = 5 for our clustering analysis where it cuts "elbow" at a relatively optimal cut-off value. The Elbow method shows the captured variance by setting a range of cluster numbers. There is usually a trade-off between the extent of captured variance and interpretability of the clustered information. Given the large number of cluster variables inputted into the algorithm, K = 5 is a relative optimal value.

Finally, we fitted KMeans Clustering to account for demographics, centrality, and traffic flow data. By fitting five distinct clusters, we are able to see how different attributes are similar to each other and correlate to charging station accessibility. From Figure 27, notice cluster 3 has a significantly lower count of EVCS categorized compared to others, which suggests that it have some unique properties behind the cluster variables that defines this group, which we will be discussing in the next section.

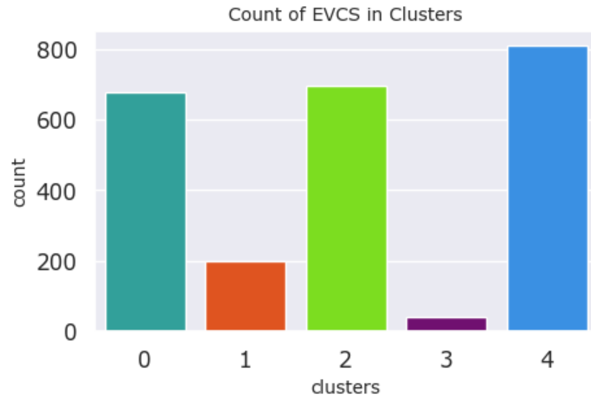


Figure 27. Count of EVCS in Clusters

The below figures show the clusters displayed geographically throughout the SoCal bounding box we defined earlier.

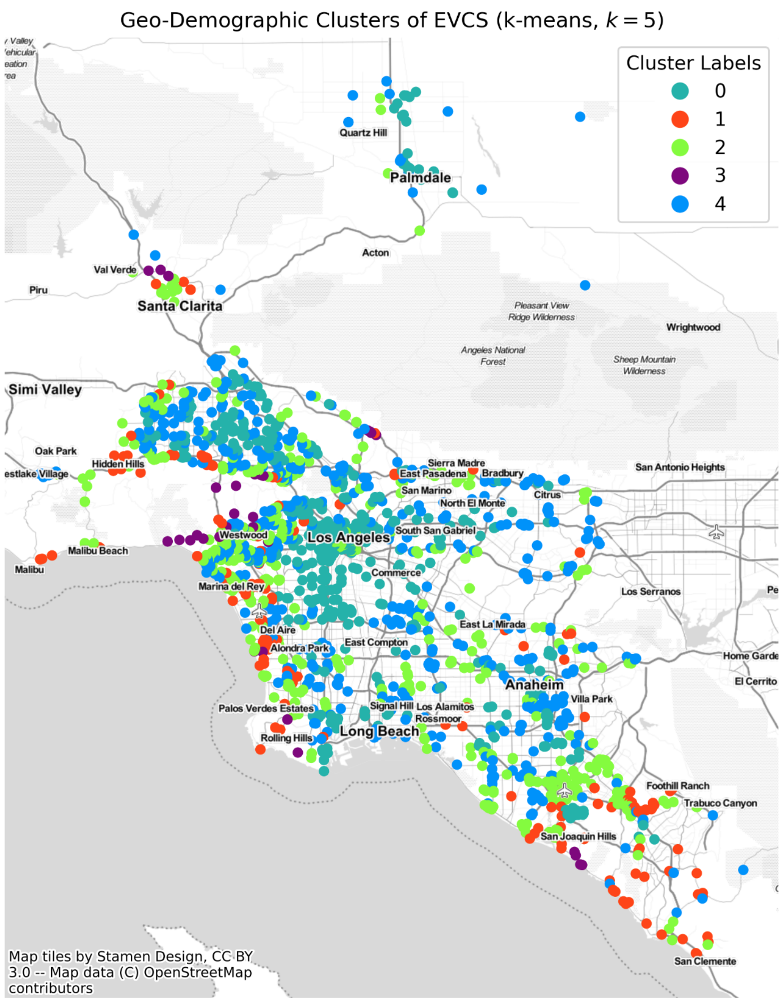


Figure 28. Distribution of 5 KMeans Clusters in Southern California

Additionally, there is clear separation across certain properties, as displayed in multiple histograms visualizing unique properties about certain clusters. By looking at the five clusters above, distinct characteristics emerge.

Figure 29 shown on the next page uncovers some extent of spatial divides among different working/commute populations groups, combining spatial distribution, network centrality, and census clustering. The unique clusters show very distinct characteristics:

Cluster 0

Cluster 0 has high closeness centrality. This means that city centers match spatial distribution, and correspond to a low median age of roughly 30. Additionally, this group had the lowest household income, which indicates that this cluster represented poor communities. Demographically, the cluster consisted of a relatively low percent share of white population and a higher hispanic population percentage, and relatively low percent access to vehicles compared to Cluster 4. However, there was no significant difference on commute time and slightly higher education level, which served well when looking at the count of EV charging stations in this cluster.

Cluster 1

Cluster 1 had low closeness centrality and betweenness centrality much less than .01. Additionally, many of the data points in this cluster were mostly near freeways and junctions, and had a fairly high median income with median age of about 45. Additionally, some areas had a large number of households with a high white percentage of the population. An interesting finding was that this cluster was suburban, but is typically underserved, in terms of EV charging stations.

Cluster 2

Cluster 2 appeared to be geospatially on the outer edge of city centers and downtown with high density around airports. Demographically, data points in this cluster had a relatively high asian population, median household income, and access to vehicles similar to Cluster 4.

Cluster 3

Cluster 3 had a very small number of EV charging stations by count. Additionally, data points appear to be scattered on the outskirts of SoCal urban regions, which tend to have very low betweenness centrality. Demographically, this group had the highest median age and household income and tended to be white dominant with a very low hispanic population and average Asian percentage of the population. The median access to vehicle was roughly 30% and had a fairly low commute time. These findings indicate that seniors could either be underserved or have low charging demand.

Cluster 4

Cluster 4 was the largest cluster size and appears to mostly be in urban-suburban transition zones. This geographic trait seems to explain the relatively high closeness centrality. Demographically, the median age is roughly 35 and is likely to have a segregated hispanic dominant population. Additionally, this cluster had higher vehicle access with some tracts having 100% average vehicle accessibility.

Finally, Figure 30 and 31 shows the relationship between EV adoption with median income and population density respectively. Income can be understood from a socioeconomic perspective because EVs are new and the cost of a new EV is higher than ICEVs, making it a vehicle for those who can afford the vehicles, presumably those with higher incomes. Meanwhile, population density was another interesting finding because perhaps having less population density means sufficient space or EV charging stations which encourage EV adoption.

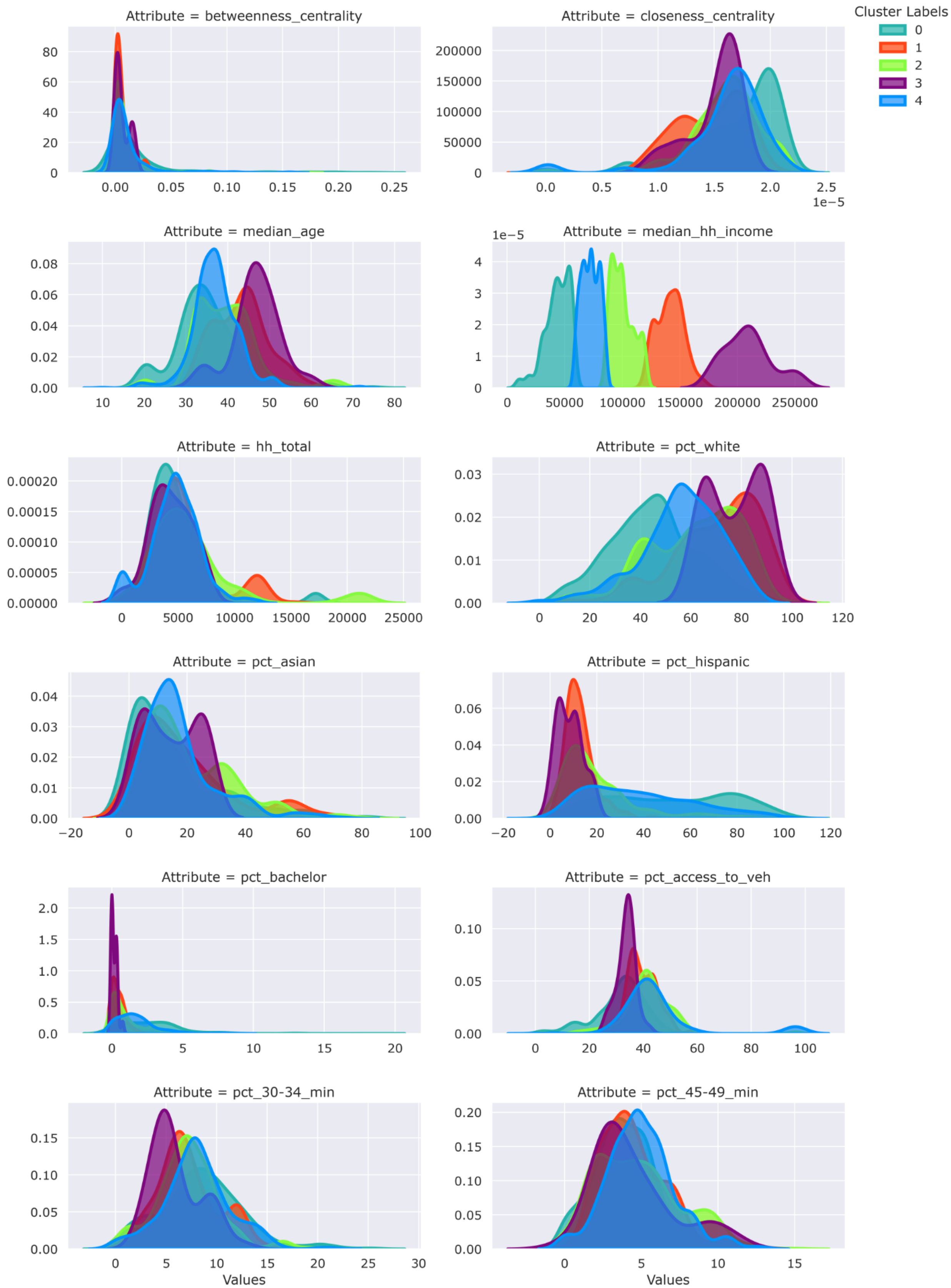


Figure 29. Cluster Characteristic Comparison Distribution

Clusters and EV Ownership

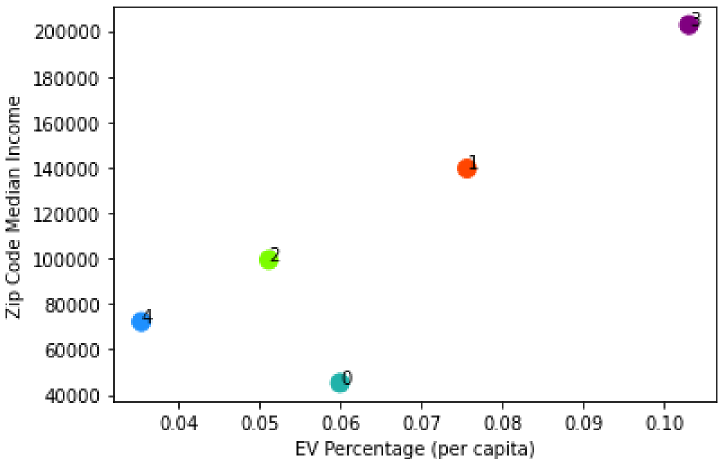


Figure 30. EV Ownership Percentage and Median Income Correlation (by Cluster)

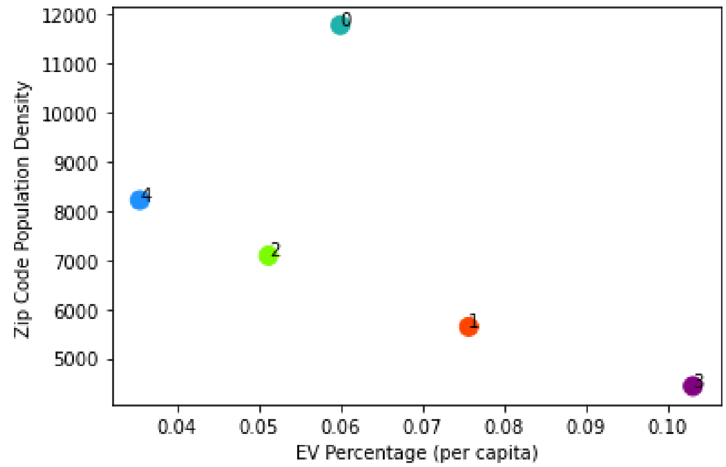


Figure 31. EV Ownership Percentage and Population Density Correlation (by Cluster)

V. Conclusions and Future Work

Findings

This study finds that there is a strong relationship between the existing charging stations and various characteristics in population, ethnicity, and traffic flows. The relationship between ethnicity and charging stations raises environmental justice issues. Additionally, the existing charging station fleet reflects demand, as determined by population and vehicle population. This study also provides additional evidence that EV adoption is limited by EV infrastructure. In order to address the decarbonization of the transportation industry and meet a net-zero economy, incentives to grow the existing EV infrastructure and charging network must be a priority.

Future Work

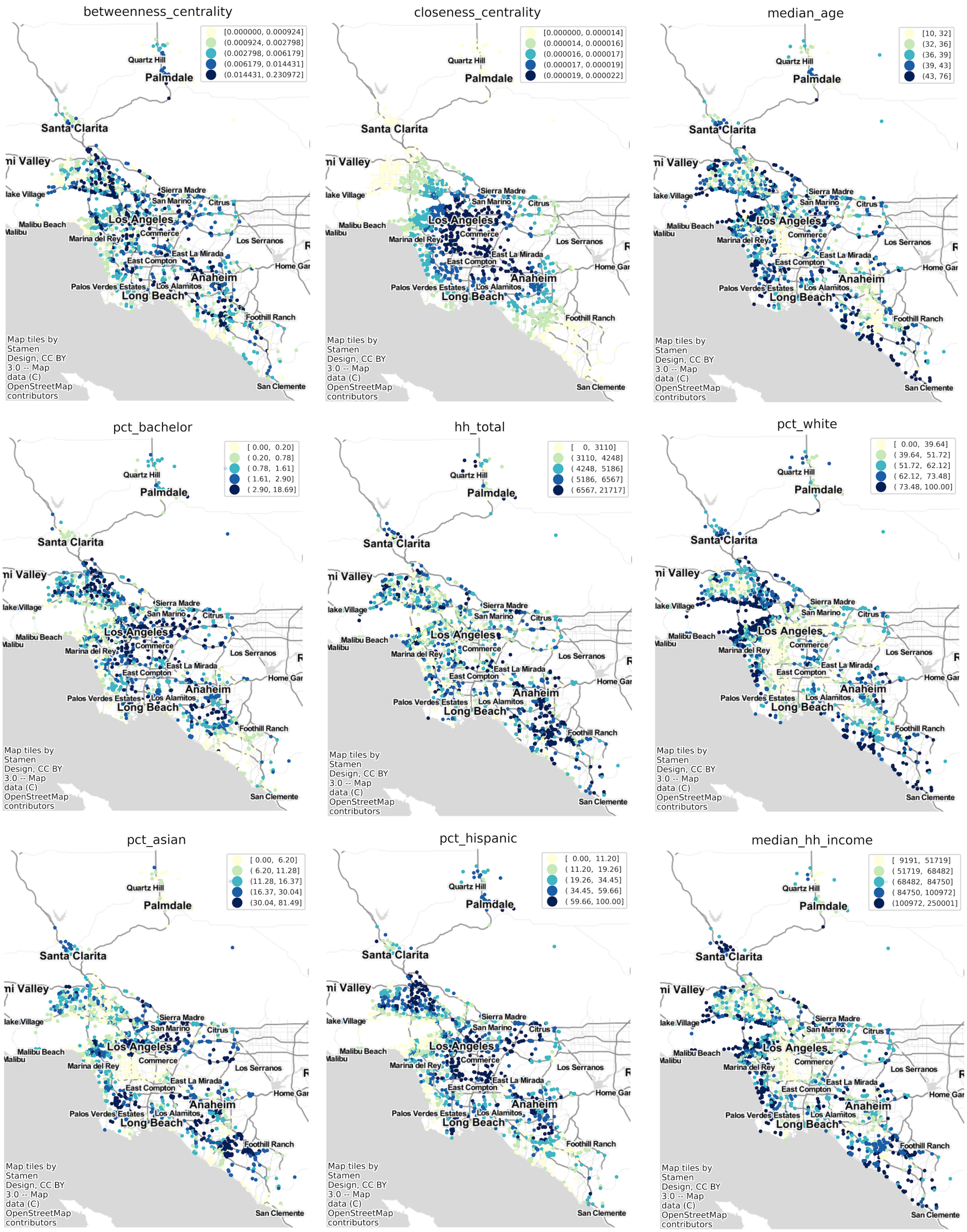
Much of our analysis was driven by the relationship between different demographics and the existing charging network in Southern California.

Further work can extrapolate and upscale this study to the entire commuting area centered at Southern California urban area, extending the street network into Riverside County and eastward as well as across different temporalities. Other regions could include the other parts of California, other states, or even other countries. Different spatial scales could be as granular as cities or as broad as states. However, the broader the scale, the more computationally expensive the study becomes, as the network grows in size exponentially rather than linearly. Investigating further into how these characteris change over time can help project the direction in which EV infrastructure is trending and if it's going in a fair, efficient, and optimized direction.

In fact, a nation-wide correlation study between the EVCS and gas station network across different temporalities would be significantly helpful in determining clusters regions where EVCS infrastructure are most needed and testing out the resilience of the EVCS network compared to that of the gas station network. By partnering up with private EV infrastructure or automotive system companies and the public transportation and infrastructure sectors, the methods used in this study would aid the electrification of transportation systems in terms of planning and evaluation.

Additionally, scaling up the study variables, not limited to ACS, LEHD, and network centrality measures, could also bring in interesting findings as EVCS are essentially serving as small "hubs" that are ingrained in the street network that commuters traverse everyday. Further characteristics to incorporate to the network analysis are air pollutant levels in each of the census tracts to see if and how air pollution is affected by the properties of charging networks.

VI. Appendix


betweenness_centrality
[0.000000, 0.000924]
(0.000924, 0.002798]
(0.002798, 0.006179]
(0.006179, 0.014431]
(0.014431, 0.230972]
closeness_centrality
[0.000000, 0.000014]
(0.000014, 0.000016]
(0.000016, 0.000017]
(0.000017, 0.000019]
(0.000019, 0.000022]
median_age
[10, 32]
(32, 36]
(36, 39]
(39, 43]
(43, 76]
pct_bachelor
[0.00, 0.20]
(0.20, 0.78]
(0.78, 1.61]
(1.61, 2.90]
(2.90, 18.69]
hh_total
[0, 3110]
(3110, 4248]
(4248, 5186]
(5186, 6567]
(6567, 21717]
pct_white
[0.00, 39.64]
(39.64, 51.72]
(51.72, 62.12]
(62.12, 73.48]
(73.48, 100.00]
pct_asian
[0.00, 6.20]
(6.20, 11.28]
(11.28, 16.37]
(16.37, 30.04]
(30.04, 81.49]
pct_hispanic
[0.00, 11.20]
(11.20, 19.26]
(19.26, 34.45]
(34.45, 59.66]
(59.66, 100.00]
median_hh_income
[9191, 51719]
(51719, 68482]
(68482, 84750]
(84750, 100972]
(100972, 250001]
Quartz Hill
Palmdale
Santa Clarita
Sierra Madre
San Marino
Citrus
Los Angeles
Malibu Beach
Malibu
Marina del Rey
Commerce
Los Serranos
East La Mirada
East Compton
Home Gar
Anaheim
Palos Verdes Estates
Los Alamitos
Long Beach
Foothill Ranch
San Clemente
Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

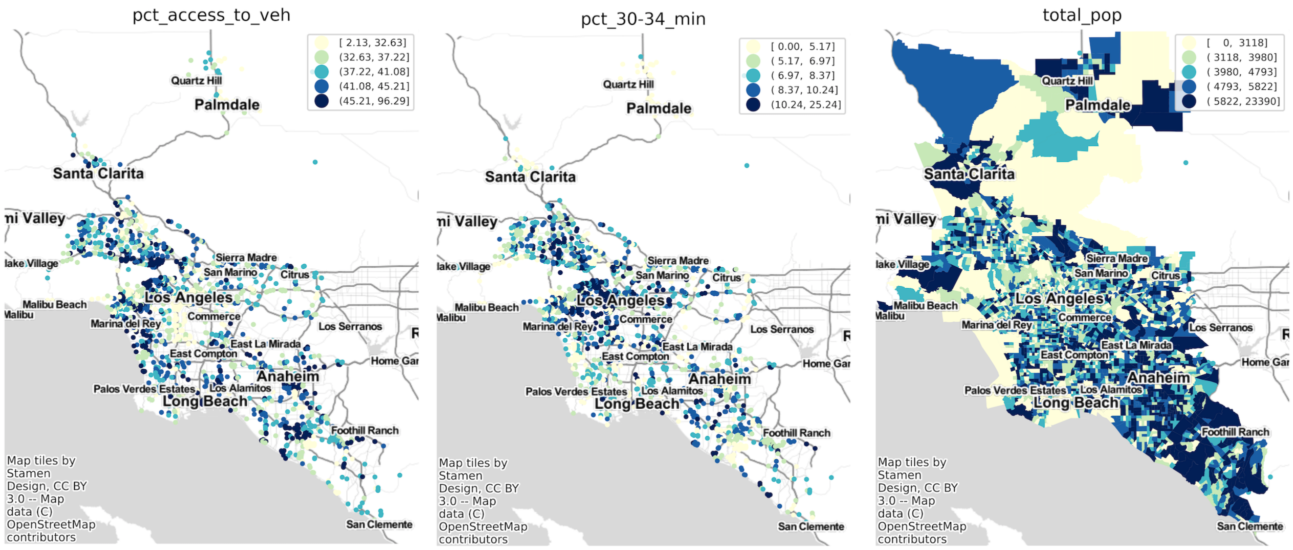


Figure A1. Cluster Variable Spatial Distributions

VII. References

Amanda Axelsson, Erik Andreasson (2020). Comparing technologies and algorithms behind mapping and routing APIs for Electric Vehicles. Jönköping School of Engineering. https://hj.diva-portal.org/smash/get/diva2:1452378/FULLTEXT01.pdf

Benjamin K. Sovacool, Johannes Kester, Lance Noel, Gerardo Zarazua de Rubens, The demographics of decarbonizing transport: The influence of gender, education, occupation, age, and household size on electric mobility preferences in the Nordic region, Global Environmental Change, Volume 52, 2018, Pages 86-100, ISSN 0959-3780, https://doi.org/10.1016/j.gloenvcha.2018.06.008.

Hearst Autos Research. “How Much Is an Electric Car?” Car and Driver, 2020, p. 1, www.caranddriver.com/research/a31544842/how-much-is-an-electric-car/#:~:text=According%20to%20Quartz%2C%20the%20average,decrease%20from%20the%20year%20before. (https://www.sciencedirect.com/science/article/pii/S095937801830030X)

Leonard, Christian. “Irvine Workers Spend Nearly $10,000 a Year Commuting, Study Says.” NBC Los Angeles, 2019, pp. 1–2, www.nbclosangeles.com/news/irvine-workers-spend-10000-commuting/2222810/#:~:text=Several%20ranks%20down%2C%20Los%20Angeles,14.5%25%20of%20their%20annual%20income.

JuiceBlog. “Understanding the Different EV Charging Levels.” Enelx, 2019, p. 1, evcharging.enelx.com/resources/blog/550-different-ev-charging-levels.

Sebastian Wagner, Markus Götzinger, Dirk Neumann (2013). Optimal location of charging stations in smart cities: A point of interest based approach. *Thirty Fourth International Conference on Information Systems.* https://www.researchgate.net/publication/287014278_Optimal_location_of_charging_stations_in_smart_cities_A_point_of_interest_based_approach/citations.

Yanyan Xu, Serdar Çolak, Emre C. Kara, Scott J. Moura, and Marta C. González (2018). Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nature Energy*. https://doi.org/10.1038/s41560-018-0136-x.

Yunfei Mu, Jianzhong Wu, Nick Jenkins, Hongjie Jia, Chengshan Wang (2018). A Spatial–Temporal model for grid impact analysis of plug-in electric vehicles. *Applied Energy*, 456-465. https://doi.org/10.1016/j.apenergy.2013.10.006.