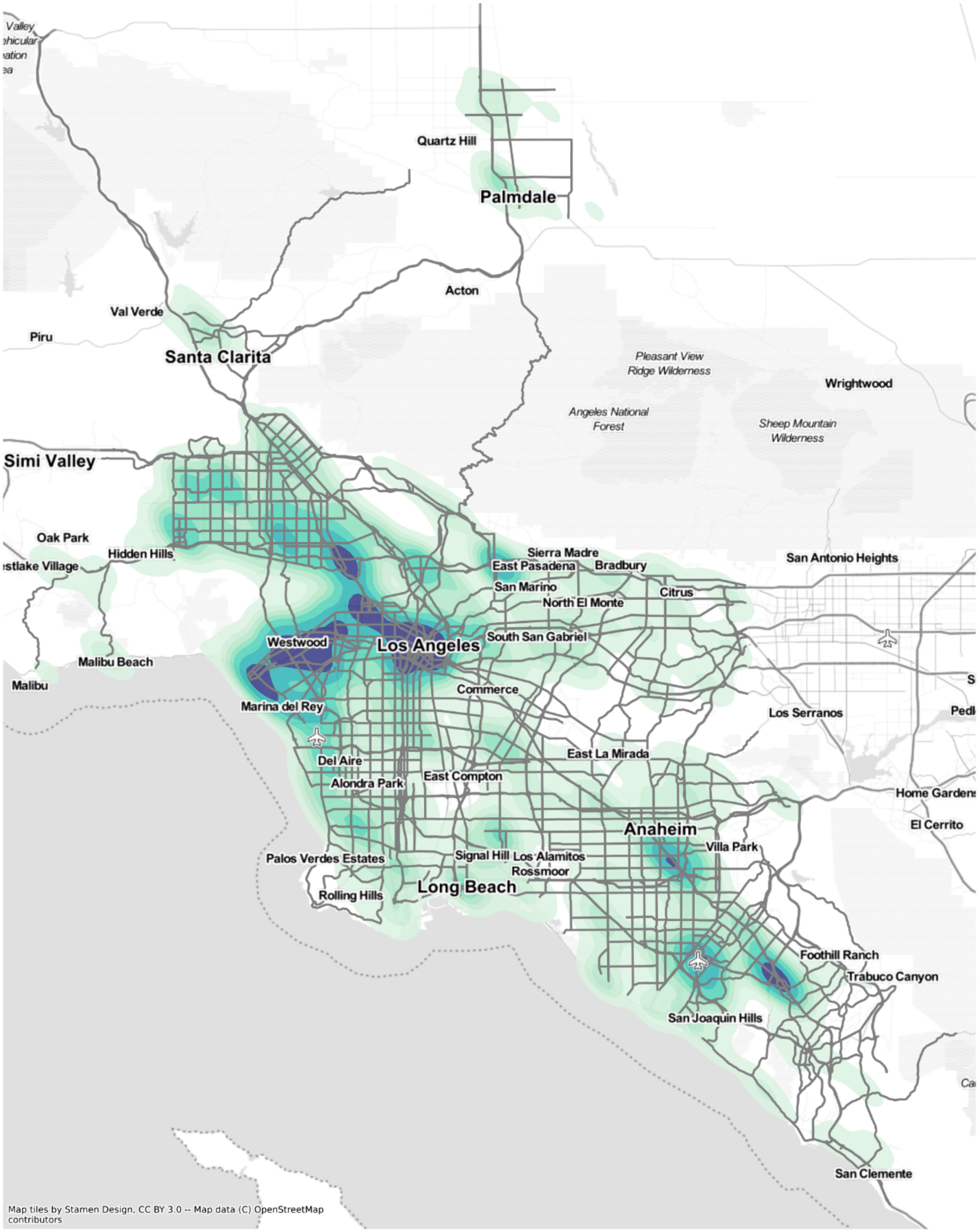
Southern California Electric Vehicle Charging Station Network Analysis

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



I. Abstract

In this project, our group examine the EVCS (Electric Vehicle Charging Station) with major roadway street network in the Southern California region to evaluate the network performance of the existing charging stations and to find the correlation between various data attributes, including the 2019 ACS (American Community Survey) census, electric vehicle registration for local charging demand, as well as LEHD traffic flow data for commute charging demand. Using open-source datasets and python packages, we are able to cluster the EVCS into six groups. …

...as supply to learn more about the accessibility and the correlation between them.

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. 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 are relatively mobile and flexible depending on the demand and the technology, which makes it a very interesting network analysis problem that connects the need of people to the optimization of the operation.

However, EVCS themselves don't provide much information on its accessibility nor the commute and 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 ACS census data as our variables, we could 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 grid, adding physical distance as edge weights and intersection, on-ramps, and exists as the network nodes. After this initial network creation, there will be a lower bound of 12k 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 income/demographic census tract data to provide more insights.

A 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, we simplify the trip distribution by using regression methods to estimate the demand of EV/PHEV trips on analysis geo-units (TAZ/census tract/zip code) using the share of EV/PHEV vehicles based on the DMV registration data. The base trip generation forecast data for regression can be sourced from local Metropolitan Transportation Commission (MTC) agencies, but some geo-spatial interpolation could be applied to unify a geo-unit for the final analysis.

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 compares 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.

III. Data and Methods

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; 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.

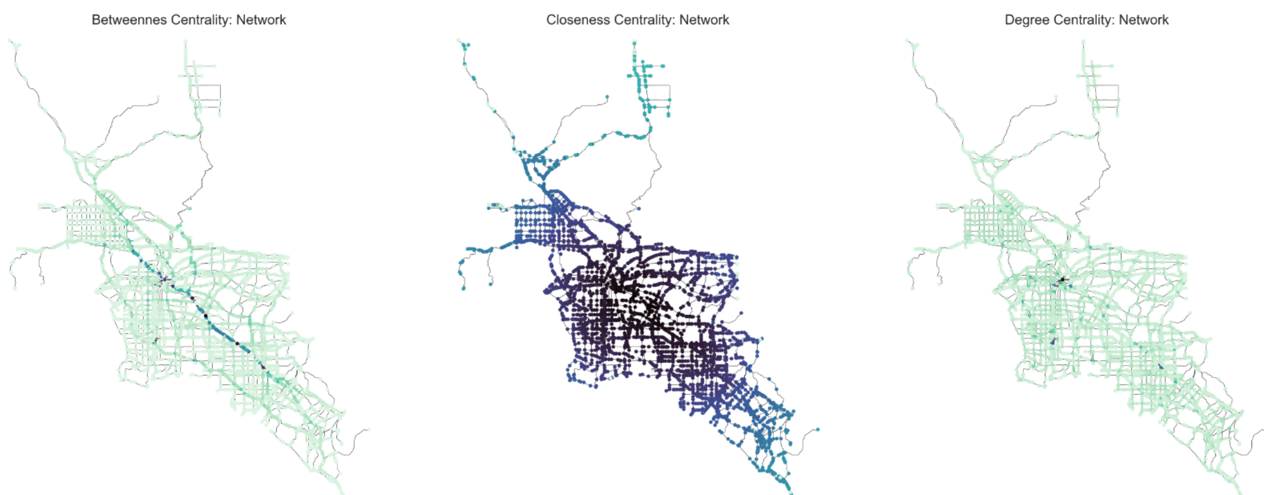


Figure 1. Street Network, Betweenness, Closeness, and Degree Centrality

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 billion squared meters, spanning 2,152 intersections and each node has an average degree of 4.4. After appending the centrality measures onto the network edges, we can clearly see that the I-5 freeway routing through the 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 have 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 tend 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. One possible explanation could have a connection with the freeway's geographic location [...]

Figure 2. Street Nodes with Highest Centrality Measures

[Will be adding in more basic analysis of the network and the data processing pipeline]

EVCS and Street Network Centrality Approximation

Charging stations data came from Open Charge Map, an open source project that provides an API to consume and contribute charging station data in the entire world. Charging stations were accessed using the [Open Charge Map API] 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.

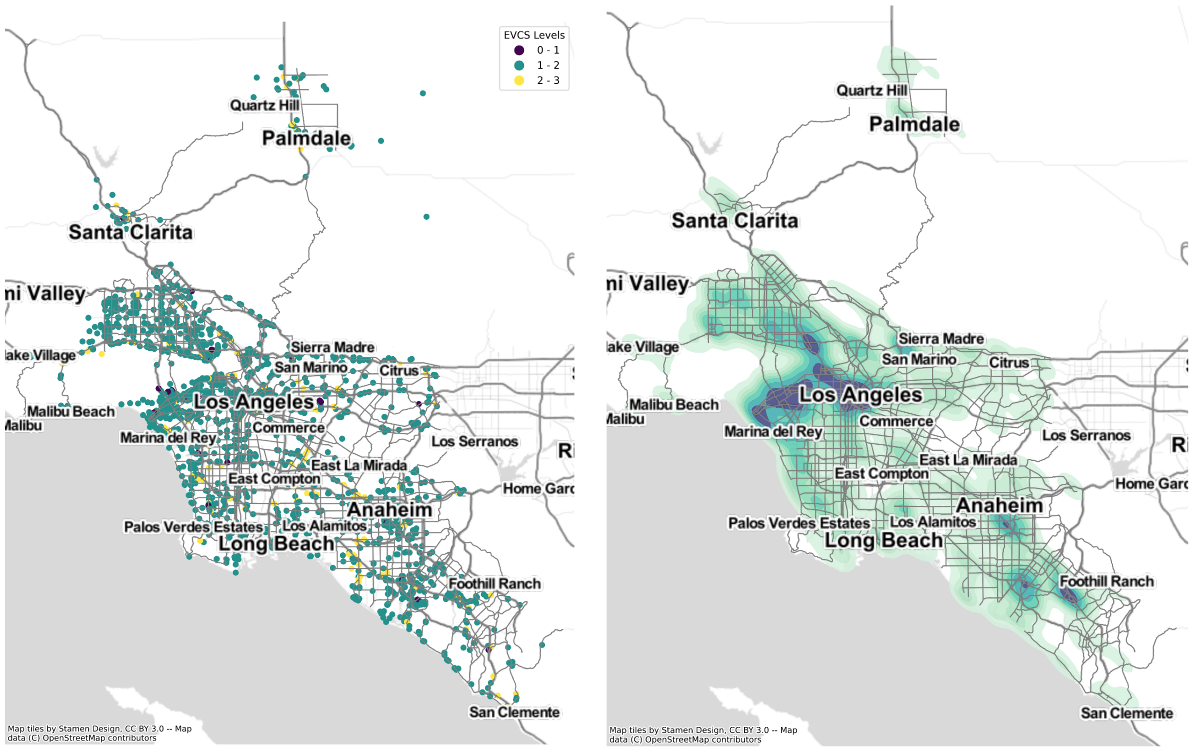


Figure 3. Southern California Charging Station Nodes and Kernel Density Clustering with City Basemap. On the left, the individual dots correspond to individual stations. On the right, a kernel density heatmap is displayed to better understand the geographic distribution of EVCS.

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.

In order to approximate the network centrality measures 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.

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. Some of the initial variables obtained included total households, total populations of different ethnicities (white and black), total households with bachelor degrees, median age, median household income, workers over 16 that drive alone to work by some vehicle, allocation of sec, allocation of travel time to work. We will be building our clustering analysis on top of this layer of census data as well.
[Adding in more details on the pipeline and data cleaning process for the final report]

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 huge problem when relating back to the decarbonization of the transportation sector.

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.

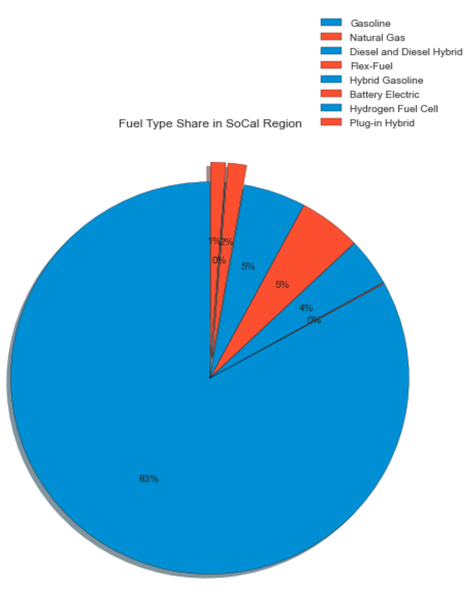


Figure 4. Fuel type by share of vehicles in Southern California, as reported by California DMV as of January 1, 2020.

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.

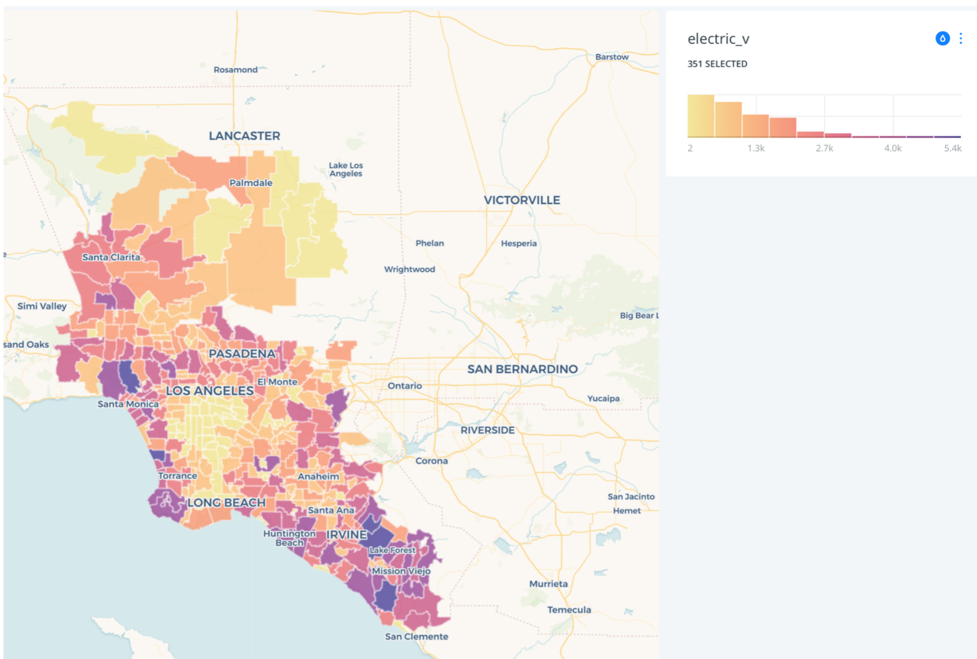


Figure 5. EV vehicle registrations by Zip Code.

There are large concentrations of EVs in the Orange County area like Irvine. There are weak concentrations of EVs in central and downtown Los Angeles. It is important to recognize that Zipcodes do not have the same number of residents and also car owners. Nonetheless, the map can be used to see general concentrations of EVs in the SoCal Region.

Longitudinal Employer-Household Dynamics (LEHD) Traffic Flow 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 being EVs.

The result of preliminary analysis of the Origin-Destination pairs shows that travel from Census Tract 6053014105 to Census Tract 6053014104 is the greatest with 374 workers.

It is interesting to note that although the same Origin-Destination tracts, Census Tract 6059062614 has 1355 workers.

Residence Census Tract	Workplace Census Tract	Total Jobs
6059062614	6059062614	1355
6053014105	6053014104	374
6071007302	6071007303	369
6081610202	6081611700	355
6083002930	6083002922	343
6073008363	6073008305	317
6051000200	6051000200	271
6075061500	6075061500	263
6085505008	6085505008	263
6059052518	6059052518	245
6085511609	6085511705	241
6037267901	6037310400	240
6071007306	6071007303	231
6073010015	6073010015	205
6067008508	6067008504	204
6037265510	6037310400	202
6087100400	6087100400	201
6087100500	6087100400	194
6059062614	6059076102	192
6113010606	6095253300	191

Figure 6. Example of LEHD LODES data which shows the number of jobs where people from a census tract commutes to another.

IV. Results

Various Analyses

We developed various analyses that range from simple exploratory data analysis, cluster analysis, centrality analysis, and census tract clustering.

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 0-3. There is a very clear cluster of charging stations in Los Angeles and Westwood, which we see clearly overlaps many streets.

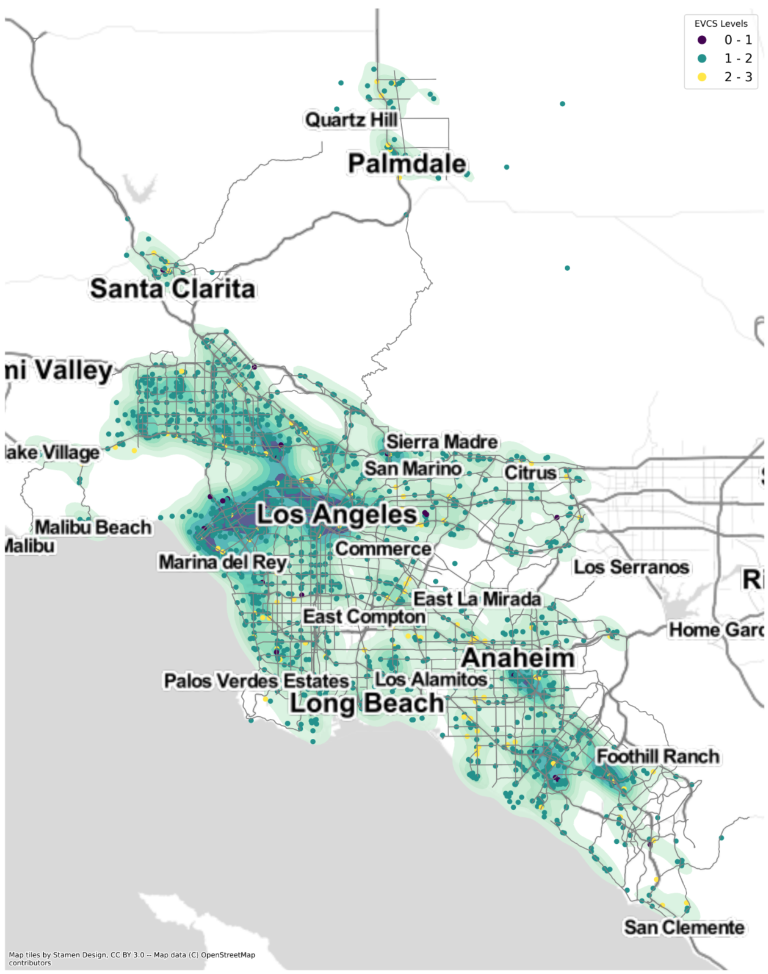


Figure 7. Overlay of EVCs nodes, KDE plot, and Street Network

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.

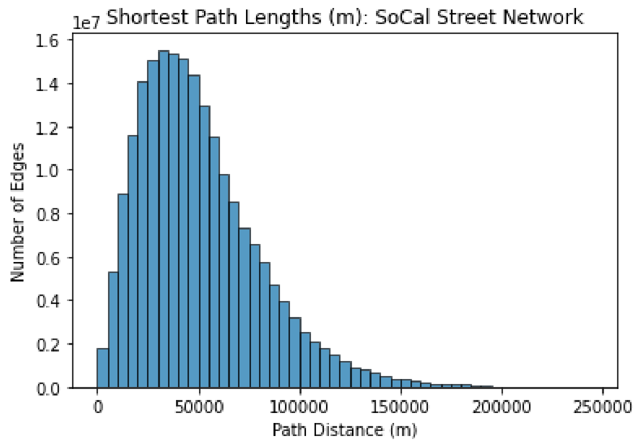


Figure 8. Physical Shortest Path Distribution.

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, but the clustering coefficient seems to closely resemble a small world network.

	#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 9. "Small World" Network Attributes.

Cluster Analysis

We investigated how the different datasets correlated with each other. We followed the following process:

1. Select relevant census variables for clustering exploration, in addition to the selected attributes of the EVCS
2. Extract census data on the census tract level for the two major counties in the study region, LA County and Orange County, CA
3. Combine their census data frame together geo-spatially, and clip within the custom bbox to keep the analysis consistency
4. Spatially interpolate the EVCS data with the census data where it lands. Each charging station represents the census demographics of the tracts that they belong to.
    a. We expect there will be 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 amount of both census tracts & EVCS within the SoCal region.
5. Run a K-means algorithm to cluster the EVCS based on their charging attributes and census attributes.

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

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.

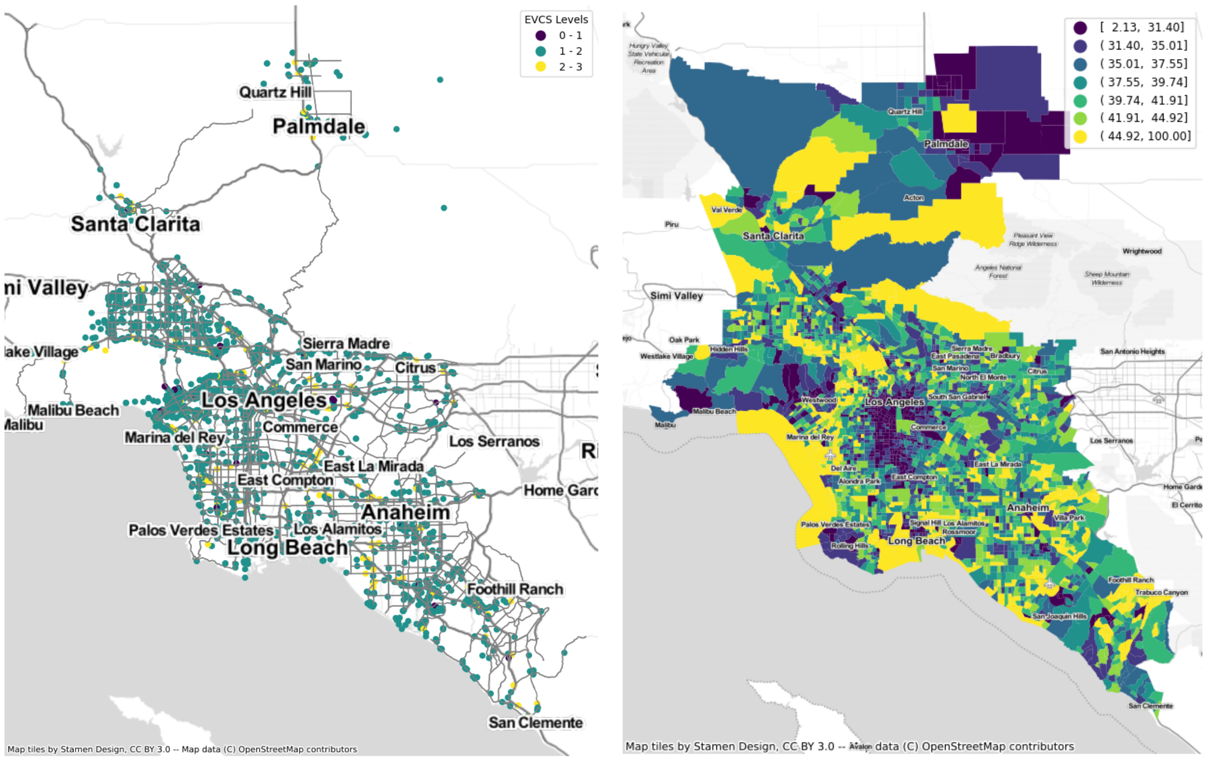


Figure 10. Adjacent comparison of EV Charging Stations(left) and Access to Vehicles, defined as "percentage of workers over 16 that drove alone to work by car, van, or truck" (right).

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.

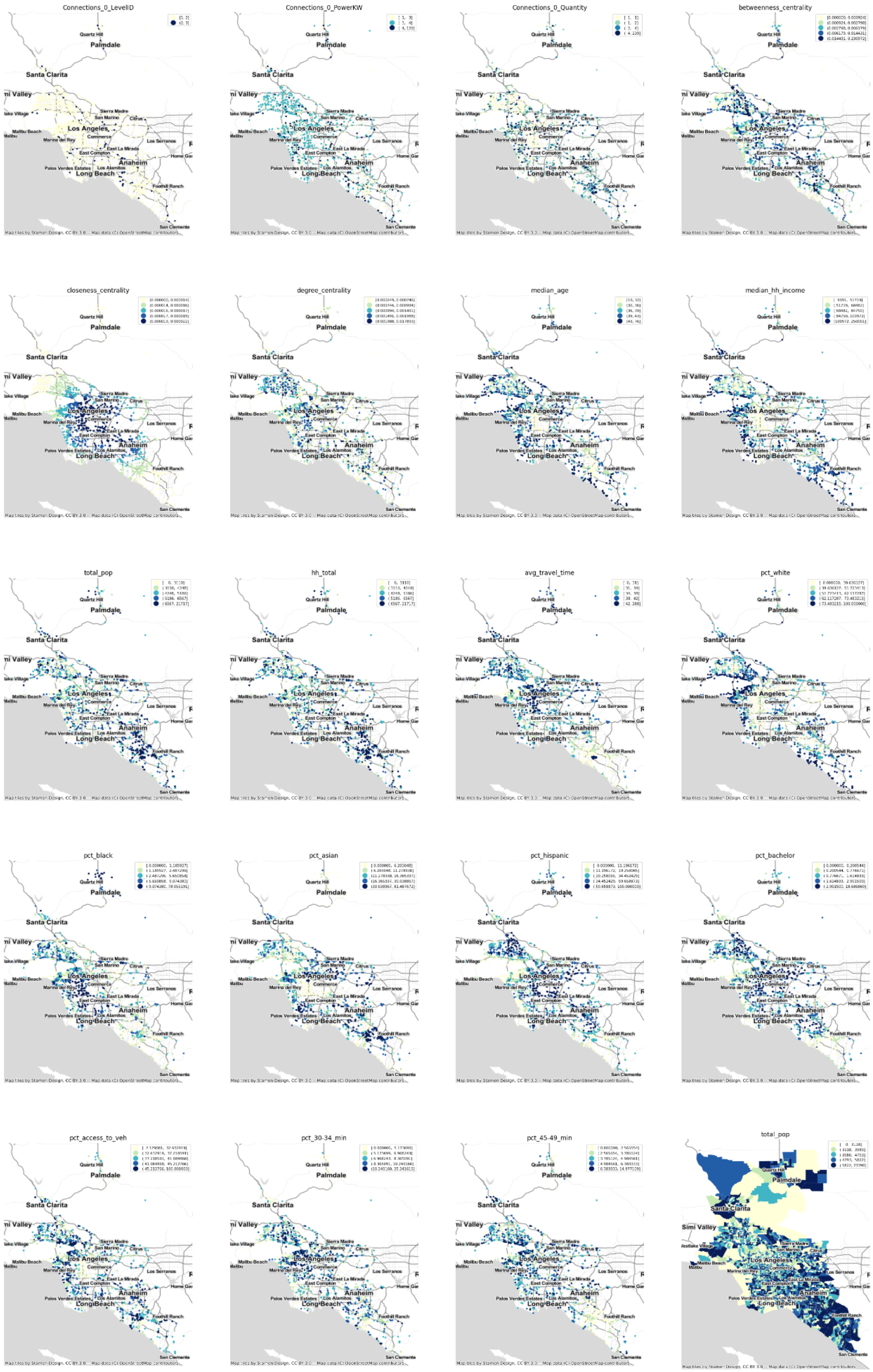


Figure 11. Comparison of different census tract characteristics and charging stations per census tract in Southern California.

The following figure shows kernel density plots of various properties and attributes in the network and data. As shown, some properties, such as household total, median household income, and commute times follow a near perfect normal distribution. Meanwhile, others, such as the different centralities, minority ethnicity, and education have apparent tails in the distributions.

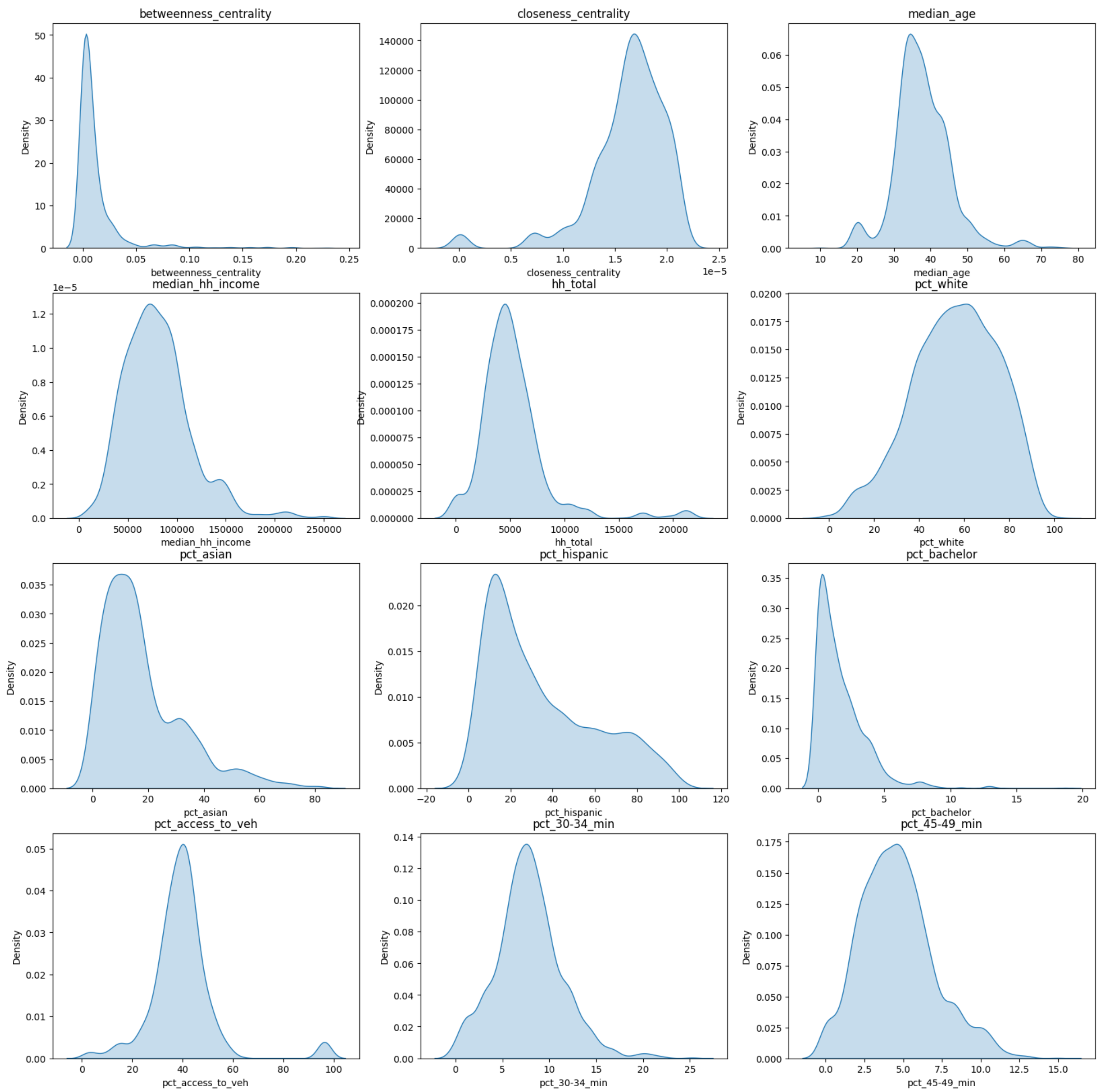


Figure 12. Kernel Density plots of different characteristics from charging stations, census tract, and traffic flows.

Finally, we fitted KMeans Clustering to account for economic, centrality, and traffic flow data. By fitting four distinct clusters, we are able to see how different attributes are similar to each other and correlate to charging station accessibility. The below figures show the clusters displayed geographically throughout the SoCal bounding box we defined earlier.

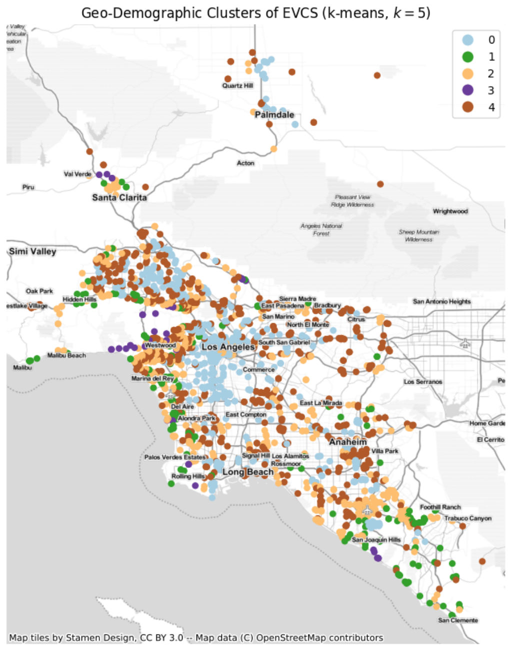


Figure 13. Distribution of 5 different clusters found by KMeans in Southern California.

Additionally, there is clear separation across certain properties, as displayed in multiple histograms visualizing unique properties about certain clusters.

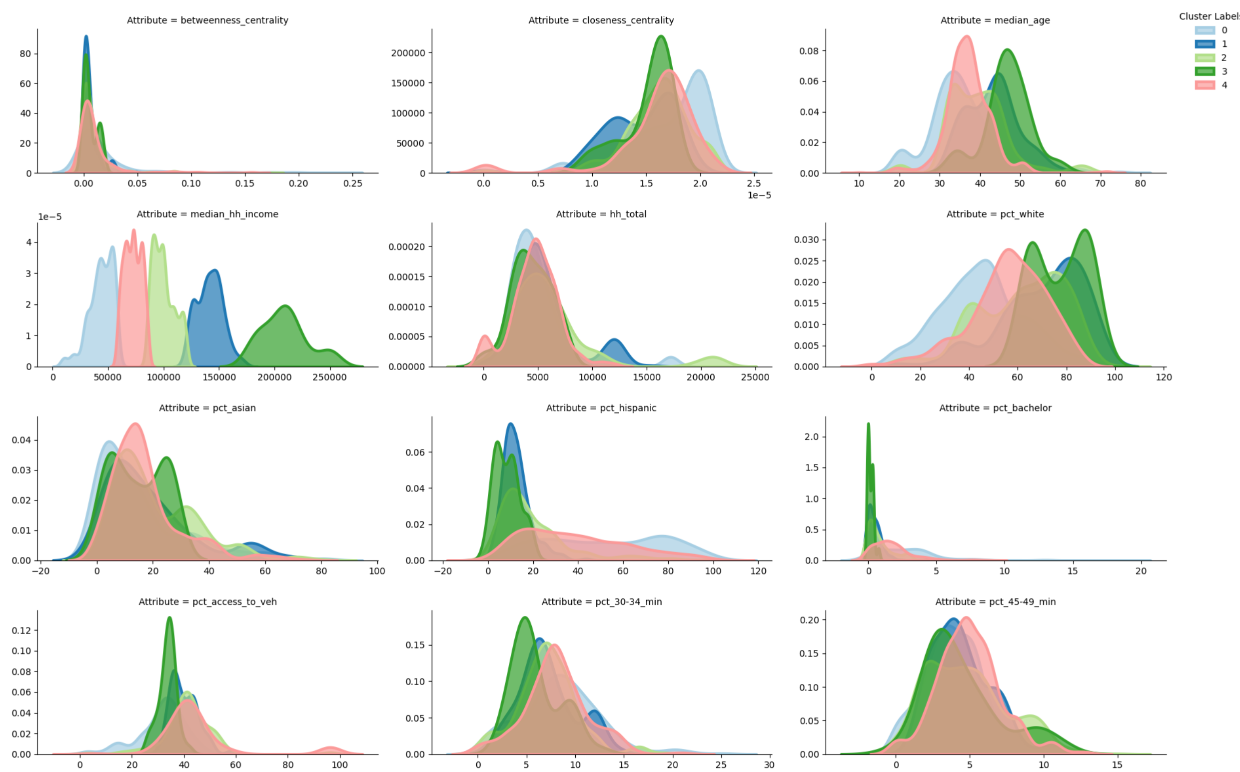


Figure 14. Cluster characteristic comparisons.

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 thee 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 this study to other regions at various spatial scales 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. Additionally, 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.

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. References:

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

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

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

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