Exploratory Data Analysis and Visualization for Surrey's Electric Vehicle Strategy

Data Science for Social Good (DSSG)
UBC Institute of Data Science

A Collaboration with the City of Surrey

Authors

Laura Greenstreet
Eugenie Lai
Alexi Rodríguez-Arelis

January 7, 2020

Contents

1 Introduction

The City of Surrey is in the early stages of developing an electric vehicle strategy as part of its approach to reducing greenhouse gas (GHG) emissions. Due to limited public transit and low-density development, Surrey is highly auto-dependent, with 60% of its GHG emissions coming from transportation (City of Surrey, 2018). Additionally, Surrey is growing rapidly, increasing in population by 10.6% in last census period compared to 6.5% for the Metro Vancouver area and 5% for Canada as a whole (Statistics Canada, 2016). Growth is expected to continue, with Surrey becoming the largest city in the Metro Vancouver region with over 800,000 residents by 2040 (City of Surrey, 2017; Western Investor, 2017). Surrey is currently planning how to support this projected rapid growth to design a sustainable and livable city.

A variety of factors make electric vehicles (EVs) attractive to consumers in the Metro Vancouver area. In other regions, the electricity used to charge vehicles comes from fossil-fuel reliant power plants, which result in electric vehicles having less than half of the lifetime emissions of a standard gasoline powered vehicle (Union of Concerned Scientists, 2015). However, 93% of electricity generated in British Columbia comes from renewable sources, resulting in nearly emission-free charging (City of Vancouver, 2019). Over 90% of BC's power comes from hydro-electricity whose abundance leads to competitive electricity rates, which in combination with the region's high gas prices make electric vehicles attractive to consumers (Government of Canada, 2019; DeMuro, 2019). Electric vehicles offer additional benefits including lower lifetime maintenance costs, better safety ratings, dedicated parking spots, and unrestricted HOV lane use (City of Vancouver, 2019; DeMuro, 2019). While electric vehicles offer many benefits, they currently make up less than 1% of the total vehicle stock in Surrey and face a variety of challenges to reaching widespread adoption.

While using electricity as a power source results in cheaper fuel prices for electric vehicles it also results in the need for new infrastructure. There is an inherent chicken-and-egg problem with EVs and charging infrastructure where a lack of charging infrastructure makes consumers reluctant to buy EVs and the small number of EVs makes private companies reluctant to build charging stations. Governments can play a key role in supporting EV technology by building initial infrastructure to encourage early consumers before letting the private sector take over. However, there is no easy answer to the question of how much support the government should provide to the fledgling EV industry.

In addition to lack of charging infrastructure, EV adoption faces challenges with the price and range of available models. In our data, we found an entry price point for an electric 4-door sedan is around $ 30,000 USD, which is around $ 10,000 USD more than the most popular models purchased in Surrey between 2016 and 2018. Additionally, most electric vehicles on the market today are small cars. Only around a quarter of Surrey's vehicle stock falls consists of small cars and there are currently no electric models SUV models on the market, which is both the largest and fastest growing category of Surrey's vehicle stock. Further, many potential customers are still unsure if an EV will fit their needs due to the time needed for charging and the limited driving range of the vehicles between charges.

The Province of British Columbia has passed legislation requiring all vehicles sold by 2040 to be zero emission and Surrey has set the goal of converting its entire passenger vehicle stock to zero-emission vehicles by 2050 (Province of British Columbia, 2018). To achieve these goals, the City of Surrey plans to support EV adoption through direct approaches such as building charging infrastructure, creating city fleet requirements, and simplifying the permitting and installation process for chargers as well as educational efforts to make consumers aware of the benefits of electric vehicles as well as rebates and programs to reduce the upfront cost. To develop a successful Electric Vehicle Strategy, Surrey needs to understand both the state of Surrey's vehicle stock and charging infrastructure and the social and economic landscape of the city.

1.1 Project Strategy

Our role is to support Surrey's Electric Vehicle Strategy by integrating vehicle stock, land-use, and demographic datasets and provide an easy way for policy makers to visually explore their interaction. As visualization struggles to compare more than two or three variables, we will also conduct deeper statistical analyses that allow us to consider many variables in a single model. We focused our modelling on the problems of where to place additional charging infrastructure and what demographic factors are associated with early EV consumers. We then compared the results from our models to existing literature on the Metro Vancouver area.

The electric vehicle stock in Surrey has doubled each of the last two years, increasing from a single vehicle in 2011 to over 1200 purely electric vehicles at the end of 2018. Due to the rapid change in Surrey's electric vehicle stock, it will be important to incorporate new data into the visualization tool in the future, as we don't expect future years to be similar to existing data. To streamline integration of future datasets, we developed a database and a web application which pulls data directly from the database. The database schema and app interfaces were designed to allow the comparison of datasets which were previously not comparable due to incompatible spatial scales. With this pipeline, the city will be able to easily add and visualize new datasets and incorporate vehicle stock, land-use, and demographics into their decision making.

1.2 EV Strategy Explorer: A Tool for Interactive Visualization

Our goal for the visualization tool was to allow users to interactively visualize the data spatially and over time. We chose to use an SQL database to capture the relationships between the datasets and to create a consistent standard for data which varied over time. The database was initially developed in Postgres but was migrated to SQL Server to integrate with the City's existing infrastructure. We chose to develop the front end of the app in R using the Shiny platform, as Shiny has many tools developed for attractive data visualization. This allowed our app to display a range of visuals including interactive maps and plots without our team needing to focus on front-end development.

The app consists of five dashboards, separating raw data from our models and findings to allow users to both view our results and analyze the data themselves. The main dashboard provides an introduction to the app and summarizes our findings on Surrey's vehicle stock, charging sites, and demographics. The modelling dashboard allows users to interact with our traffic-based charging site placement model. The vehicle stock, traffic, and census dashboards allow users to interactively visualize the datasets. We chose to present multiple dashboards for interactive data visualization to streamline the options available on each dashboard, so it was clear what each plot or map was visualizing.

2 Data Sources and Processing

We integrated six datasets covering vehicle stock, EV charging infrastructure, land-use, traffic flow, and demographics. The datasets used a variety of spatial measurements which were often incompatible, requiring mappings to be created between systems. We worked with City staff to divide Surrey into 47 areas and aggregated each dataset to the area level. This allowed our datasets to be compared at a level of granularity that is useful for decision making. Our primary datasets were ICBC vehicle registration records for the City of Surrey and location and charging information for charging sites in the Surrey area. Our secondary datasets provided socio-demographic information to help put the vehicle information in context.

2.1 Primary Data Sources

2.1.1 ICBC Registrations

The ICBC datasets consisted of individual vehicle registrations for the City of Surrey for the years 2006, 2011, 2016, 2017, and 2018. Postal codes were available for each registration, but the data was anonymized, so it was not possible to track vehicles across years. Each registration included a variety of information including the make and model of the vehicle, electric and hybrid flags, a general vehicle type, and two fields indicating whether the vehicle was a passenger or commercial vehicle and the primary use of the vehicle. Pickup trucks were always placed in the commercial category. Since we believe that many people own pickup trucks for personal use, we only considered pickup trucks to be commercial if both use fields indicated commercial use. We then restricted our analysis to passenger vehicles and the selected pickup trucks. This classification scheme was also used by the 2018 Data Science for Social Good transportation team for their GHG emission analysis on passenger vehicles (Anwar et al., 2018).

A significant challenge in working with the ICBC data was that consistent names were not used for makes, models, or categories of vehicles. This made it difficult to get total counts for vehicle models or categories. For example, the vehicle class provided for a Honda CR-V SUV varied by the registration across 7 categories including four door sedan, four door stationwagon, and hatchback. As an example with makes and models, the Toyota Highlander hybrid is listed with 9 different model names including Highlander, Highlander Hybrid, Highlander Hybrid 4DR 2WD, Highlander Hybrid Limited 4DR 2WD, and HLNDR. Additionally, plug-in hybrids are not clearly indicated in the ICBC data, with the plug-in status only indicated as part of the model name. So, it is unclear if only 8 of the 1737 Priuses recorded in the 2018 registration data are actually plug-in hybrids, or if the number is greater and the status was not recorded for many registrations.

Because there were issues with ICBC classifying the same vehicle in multiple, often inaccurate classes, we devised an alternate system for vehicle classification. Based on the work done by the 2018 DSSG transportation team, we merged the ICBC data with the EPA Fuel Economy dataset (EPA, 2018) which added MPG, GHG emission, and EPA vehicle class information. We used the scripts from the 2018 team to integrate the datasets, matching 93% of passenger vehicles registrations to an EPA model (Anwar et al., 2018). To make visualizations clearer, we simplified the 12 categories classification scheme used by the 2018 team to 6 categories: small cars, large cars, SUVs, trucks, vans, and special purpose vehicles.

In addition to the fields provided, we created a luxury flag in the ICBC dataset based on the make of the vehicle. Registrations associated with 23 makes of vehicles were considered luxury. However, the distribution of makes was highly skewed with the top five brands accounting for 79% of the vehicles in the 2018 dataset and BMWs alone accounting for over a quarter of all luxury vehicles.

2.1.2 Vehicle Price and Rating Data

We manually collected additional information on car prices and ratings of popular models in the Surrey vehicle stock from popular American car websites, primarily USNews, Kelley Blue Book, and Autotrader. These datasets include the MSRP and used prices of popular electric and gasoline vehicles by year as well as critic ratings. We chose to collect data from American car websites as we could get more reliable data for past prices, particularly historical MSRP prices. As data was collected manually, we could only collect information on electric vehicle models and the most frequently purchased vehicles in Surrey between 2016 and 2018. Prior to manual collection, we looked for an open pricing dataset and even attempted to match our data to a set of 11,000 vehicle prices available on Kaggle. However, we were only able to match around 25% of vehicles to a price and were often unable to make unique matches, so we abandoned the approach in favor of targeted collection on key models.

2.1.3 EV Charging Data

The charging site dataset included information on 24 charging sites representing 35 chargers in the City of Surrey run by the companies ChargePoint, Flo, and Greenlots. The data included charging session information for 13 of the 24 sites. The information provided varied by company, but all included the start time and location for each charging session. The ChargePoint datasets were most detailed including information such as end time of each session, the postal code of the driver, and the amount of time spent spent actively charging. To account for sites that are not partnered with the city, we manually collected data on charging site locations from ChargeHub, a website that helps EV owners find publicly available charging sites. The ChargeHub data provided an additional 28 sites in Surrey representing 85 chargers as well as 22 sites within 5km of Surrey representing an additional 52 chargers.

Existing charging data is key to determining charging site capacity and utilization and would be a useful input for charging site placement models. However, there were significant data quality issues which limited our analysis of the data. First, we only had data for 13 of the 70 public charging sites we are aware of in Surrey. Additionally, all of the sites we had data on were government buildings such as libraries, recreation centers, and museums, biasing our data away from commercial locations. While the full set of charging sites does include many government buildings, it also includes charging sites at hotels, parks, and businesses including Tim Hortons, Shoppers Drug Mart, IKEA, and Real Canadian Superstore. For the 13 sites we had data on, the data had consistency issues including charging sessions with lengths over 1 year or of negative duration and gaps of a year or more without a charging session being logged. Based on the quality of the data, we performed a basic analysis to capture trends in charging session frequency and duration and used the location of existing sites in our models for placing additional sites, but were unable to fully analyze capacity or utilization.

2.2 Secondary Data Sources

Our secondary datasets included business and property licenses for the City of Surrey, traffic flow based on the Metro Vancouver EMME transit model, and census data.

2.2.1 Business License Data

The business license dataset consisted of 30,000 business license applications received by the city of Surrey in 2018, including approved, pending, and rejected applications. Entries included an address, an employee total, an application type of Commercial/Industrial, Home Occupation, or Non-Residence, as well as a business category code. There were several hundred category codes of varying levels of specificity from as specific as taxi service, bank, and esthetician to as general as consultant and contractor. Multiple application types were used for the same category code, including odd combinations such as classifying welding as a home occupation.

Around 2750 businesses (8.9% of total data) did not have address information and another 100 provided addresses that could not be geocoded through Geocoder.ca's open dataset or the Google Maps API. Eleven hundred applications (3.5% of total data) used one of 400 postal codes which were changed to a different postal code when the address was geocoded using the Google Maps API. As the postal codes were also missing from the Geocoder.ca dataset, we chose to use the postal codes found by the Google Maps API as they had reliable latitude and longitude information, which was needed for data visualization.

We have two possible explanations for how postal codes were provided that are not recognized by our geocoding sources. Some of the postal codes were likely entered incorrectly as they did not match the A1A 1A1 pattern of Canadian postal codes. Second, as Surrey is rapidly developing, new postal codes are frequently created. So, in some cases the given postal code could be correct and Google Maps may not be aware of the new postal code. However, in this case we still think it is better to use the outdated Google Maps postal code as we have no geolocation data for newly created postal codes.

2.2.2 Property Dataset

The property dataset consisted of 135,000 buildings on record with the city at the end of 2018. Each entry included an address, estimated population and employment, floorspace, building type, year of construction, intended usage, and inclusion status in various neighborhood and development plans. This dataset does not capture the secondary housing market, such as rented basement suites. The City of Surrey has a dataset on the secondary housing market that was requested but not received during the project that would be a good future addition to the app.

2.2.3 EMME Traffic Flow Data

The EMME model is a statistical model designed to predict traffic flow for the Metro Vancouver area. The EMME model is part of the larger Greater Vancouver Regional Travel Model and can be run on a variety of parameters at different spatial scales (Translink, 2016). We were provided origin-destination matrices of traffic flows for single occupant vehicles (SOV), high occupancy vehicles (HOV), commercial trucking, public transit, bike, walking, and rail trips for the peak periods of AM (7:30-8:30AM), midday (12:00-1:00PM), and PM (4:30-5:30PM) for a typical fall workday in 2016, 2035, and 2050 (Translink, 2016). The model is built based on data from the 2011 Metro Vancouver trip diary, 2011 Screenline Survey, and other Metro Vancouver models based on socio-demographic data. Of these datasets, the screenline study provided the most important data, measuring actual traffic volumes at a variety of measurement stations in Surrey. The origins and destinations in the EMME model are traffic analysis zones (TAZs) which range in size from several square city blocks totalling 0.05km^2 in dense areas to over 20km^2 in rural areas. There are a total of 374 TAZs in Surrey and 1700 in the Greater Vancouver area.

While the EMME model is widely used by transit planners there are some oddities in the data. First, trips between TAZs are not integers, as you would expect for predications of how many vehicles travel from one zone to another. A fractional value could be obtained by dividing an integer number of total trips by an integer number of days. For example, the model might predict 1.20 daily trips from one zone to another if 6 trips were recorded in a work week. However, many of these values were less than 0.05, translating to a prediction of one trip every 20 days or more. This seems like a very strong prediction given that estimates project at minimum 5 years from the original data.

As many of the fractional values were quite small, we tried rounding values to the nearest integer. However, this drastically decreased the total traffic. For example for SOV vehicle traffic in the AM peaktime of 2016, the model predicts a total of 234,766 trips. Rounding to the nearest integer, this total decreases to 63,618 trips. Entries representing fewer than one trip account for 2/3 of travel and trips with a value less than 0.05 account for 10% of travel.

A total of 235,000 SOV trips during peak commute time (7:30-8:30am) seemed low for the entire Metro Vancouver area. So, we compared this value to the publicly available results of the Metro Vancouver 2011 Regional Trip Diary Survey (Translink, 2013). The trip diary measured traffic flows by the hour, recording 400,000 total trips from 7:00-7:59am and 700,000 total trips from 8:00-8:59am. The trip diary found 54% of traffic during the peak AM commute was SOV traffic. Assuming the traffic from 7:30-8:30am was between the 7:00-7:59am and 8:00-8:59am totals, we would conservatively expect 216,000 - 378,000 SOV trips in the peak AM commute time. Thus, the total we get from the EMME model is on the low end, even with the assumption that the model chose a peak AM commute time to be 7:30-8:30am when more trips were taken from 8:00-8:59am. As the trip diary data is from 2011 and the model data is a prediction for 2016, we would expect the totals to be even higher as the region grew 6.5% in the interim (Statistics Canada, 2016).

Given our reservations about the trip totals of the EMME data we received, we focused on visualizing and modelling relative volumes. For example, in the app we use heat maps to show the popularity of origins and destinations for each peaktime and we used chord diagrams to show where traffic is flowing to and from relative to a single area.

2.2.4 Canadian Census

The Canadian Census was our primary source of demographic data. We used data from the three most recent censuses in 2016, 2011, and 2006. While the census collects data on hundreds of measures, we considered only 12 key measures to make it easier to visualize and interact with the demographic data. We considered eight socioeconomic measures: population, household size, age, income, education, dwelling type, property ownership status, and workplace. We also considered four measures related to commuting: commuting mode, total commute time, commute destination, and commute start time.

While in most cases data were available for all three editions of the census, the commuting values were only measured during the 2016 census. For measures which occurred in multiple editions, the bins often changed from census to census. For example, in 2006 the highest value for income was $100,000+ which was increased in 2011 to $150,000+ and was further raised in 2016 to $200,000+. To keep the data comparable over time, we aggregated bins in later editions of the census to match the lower resolution of the 2006 census.

Even simplifying to the 2006 bins, many of the measures were hard to visualize as they contained over ten bins. For example, income was measured in $10,000 intervals, so the initial plot for income included 11 bins. To improve visualization and prepare data for the consumer profiling process, we further aggregated some categories to match the those used in Axsen et al. (2016) as closely as possible. For categories not included in the Axsen study, such as commute time, we aggregated bins until only 4-5 categories remained.

The census uses a separate spatial scale than our other datasets, with its most detailed public data being available at the dissemination area (DA) level. DAs are defined as “small, relatively stable geographic unit[s]... with a population of 400 to 700 persons” (Statistics Canada, 2015). Polygons for DAs are freely available from the census, and though DAs can be redefined in each census period, we found the DAs for Surrey for all three editions of the census to be identical.

(a) TAZs (b) Areas (c) Communities

Figure 1: Traffic Analysis Zones (TAZs) (a) were aggregated by city staff to form 47 areas (b). Areas can be further aggregated to form the communities familiar to Surrey residents (c).

2.3 Spatial Rebasing

A significant challenge in integrating the data was that the six datasets used three spatial systems. The EMME model used 374 TAZs, ranging in area from a fraction of a square kilometre to over 20km^2. The finest resolution we could obtain for the census was at the DA level, where areas ranged in size from 0.02 km^2 to 20 km^2. The remaining four datasets provided data at the postal code level. We were unable to find a comprehensive list of postal codes, but found over 12,000 postal codes in Surrey in our data. While postal codes provided the finest spatial measurement, there were no freely available boundaries for the postal codes zones.

We decided to use the TAZs as the base unit for our spatial system and worked with planning staff to aggregate TAZs into 47 Areas ranging from 1 km^2 to 50 km^2. We aggregated to a coarser spatial level so each Area could be interpreted as a neighborhood, making it reasonable for the city to develop plans for each Area and reducing spuriously significant results arising from running statistical tests on hundreds or thousands or regions. While Areas were our primary spatial level for analysis, we allow users to view data at the TAZ, DA, and Area levels in the app. Surrey identifies seven communities within the City and the Areas were chosen so they can be aggregated to form the communities, with most communities containing 5-7 Areas (see fig. 1).

As we did not have the boundaries of the postal codes, we assigned each postal code to a single TAZ based on the TAZ the centroid of the postal code intersected. Centroid coordinates for postal codes were obtained using the Google Maps API and the Geocoder.ca dataset. Polygons with spatial coordinates for each TAZ were provided by Metro Vancouver with the EMME model. Intersections between postal code centroids and TAZs were found using QGIS. In some cases the true boundary of the postal code may have overlapped multiple TAZs, but as we did not have boundaries for the postal codes we had no way to determine this. In figure 2, you can see that the majority of postal codes are likely contained in a single TAZ, suggesting our centroid intersection method is reasonable.

The census DAs were also aggregated to the Area level. The census provides shapefiles that define the boundaries of each DA for each census year. However, using QGIS we found these boundaries were identical for 2006, 2011, and 2016. Using QGIS to perform the spatial analysis, we found there were 719 regions defined by the intersections of DAs and Areas. Of the 594 DAs, 103 (17.3%) overlapped more

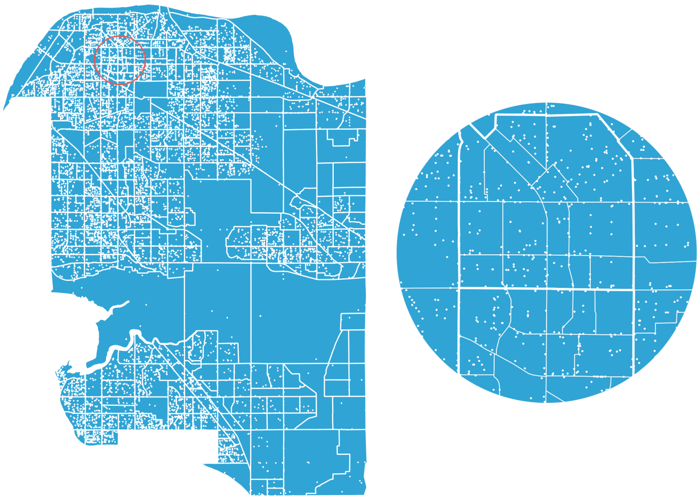


Figure 2: Map of Surrey by TAZ with overlay of postal code centroids. Most centroids are far from TAZ borders and the area associated with the postal code is most likely contained in a single TAZ. The magnified region corresponds to the circled portion on the map.

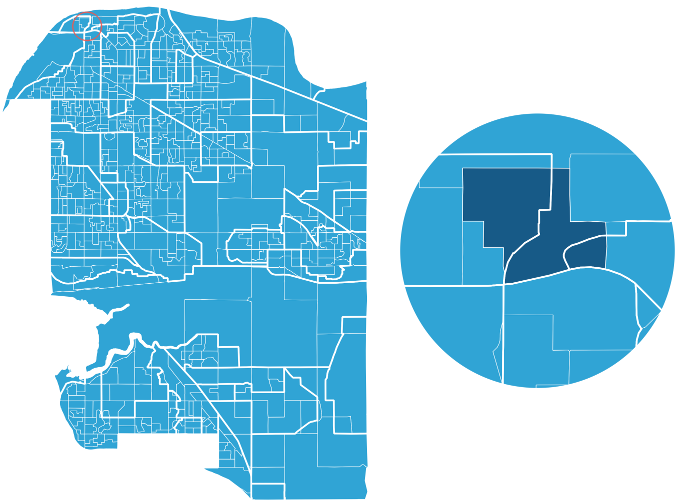


Figure 3: Map of Surrey by DA with Areas shown with bold white lines. The magnified region shows an example of a DA which intersects multiple regions. In these cases, we assigned to each Area the count for every DA intersecting the Area times the portion of the DA within the Area.

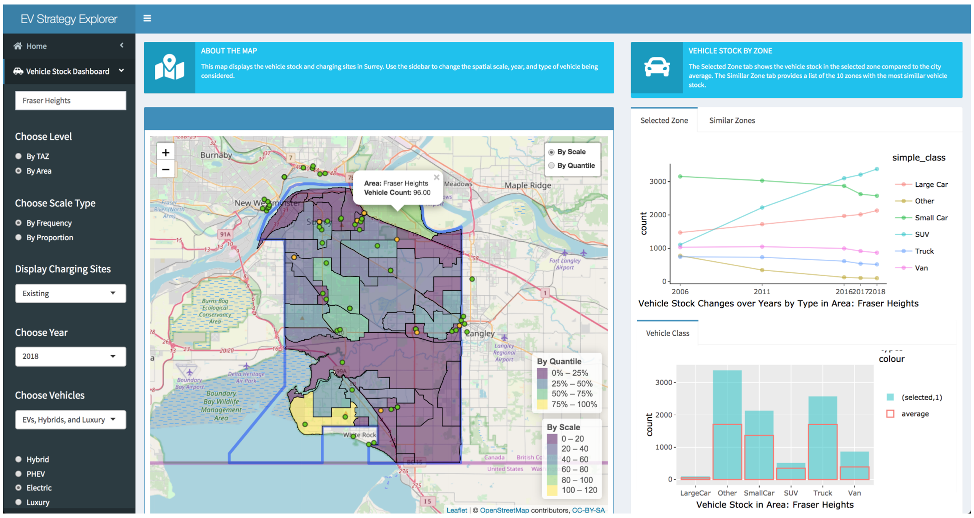


Figure 4: A screenshot of EV Strategy Explore. Tthe map on the left is exploring the number of EVs per Area in 2018, while the plots on the right show the statistics specific to the selected Area, Fraser Heights. The line graph on the top right shows the yearly changes of vehicle counts by class, while the bar chart on the bottom right shows the vehicle composition compared to the city average.

than one Area. For each of the 719 regions, we assigned demographic values equal to the count for the DA times the proportion of the DA in the region (see fig. 3).

Proportionally weighting the DAs by land area assumes both the population density and demographics are uniform. While these assumptions are likely not fully satisfied, the census strives to have DAs represent geographically stable units (Statistics Canada, 2015). Using our other datasets, we could have computed a finer measure of population density and weighted the DA-Area intersections by these values. However, due to the time constraints of the project, and to keep our work simple and reproducible, we chose to weight by land area.

2.4 Visualization Methods

Since the primary purpose of the app is to understand the spatial distribution of the data, our visualization heavily relies on maps. However, maps are prone to subjectivity. For example, choosing different colour scales for a heat map can minimize or exaggerate the differences in EV density among areas. To mitigate subjectivity, we enabled the tool to visualize the data on several scales to give the audience several perspectives. For example, on the vehicle stock tab, users are able to see where EVs are located in each area by either count or proportion. By count, Cloverdale Industrial does not even show up in the top quantile, while by proportion, it becomes an outlier with 4% electric vehicles compared to 2% in the area with the next largest percentage. The app also allows the user to visualize by the TAZ and Area levels to provide different levels of spatial resolution.

While maps are effective for spatial visualization, it is inconvenient for users to switch between maps to see changes over time. Hence we complemented the maps with visuals such as box plots and line graphs to provide information on temporal change. For example, the line chart on the top right shows that the total number of passenger vehicles has been growing rapidly in Surrey. Other plots are also helpful for visualizing comparisons in addition to changes over time. For example, a solid bar plot with an outline for a Surrey-wide average can easily communicate if an area has few or many of each class of vehicle (see fig. 4).

3 Single Dataset Analysis

While our goal was to integrate the datasets, our primary datasets had not been individually explored in depth. We believe that it is valuable to understand the trends in vehicle stock and charging site usage before trying to correlate these trends with demographics and land-use patterns.

We were interested in studying the passenger vehicle stock as a whole and with respect to low-emission vehicles. For the whole of Surrey, we were interested in how the number of vehicles is growing with respect to the population and how the market is distributed between vehicle classes. With respect to low-emission vehicles, we were interested in the number, location, and change in EVs over time, the distribution of other low-emission vehicles such as hybrids, and how the prices and rating of electric vehicles compare to popular models in the Surrey vehicle stock.

For charging sites, we were interested in how the number, duration, and amount of energy consumed during sessions changed over time. We were also interested in how charger use varied by season and day of the week. The findings below are a summary of some of our most compelling results.

3.1 Vehicle Stock Trends

3.1.1 Stock Growth

The passenger vehicle stock of Surrey grew 50% between 2006 and 2018, increasing from 190,538 to 286,987 vehicles. In the same period, the population only increased by 44%, resulting in a rise in vehicles per capita from 0.48 to 0.51.

3.1.2 Class Growth

Stock growth was not distributed evenly across the six classes in the vehicle stock. In terms of count, there was modest growth in the number of large cars and vans, with the stock increasing by 19% and 15% respectively from 2006 and 2018. Small cars stayed nearly constant with less than 1% growth over the same period. The passenger truck stock saw a modest decrease of 9%. Other vehicles, which consist mostly of special use vehicles such as golf carts and ATVs, saw a drastic decrease of 85%, falling from 16,862 vehicles in 2006 to only 2465 vehicles in 2018. Finally, SUVs saw a drastic increase of 345%,

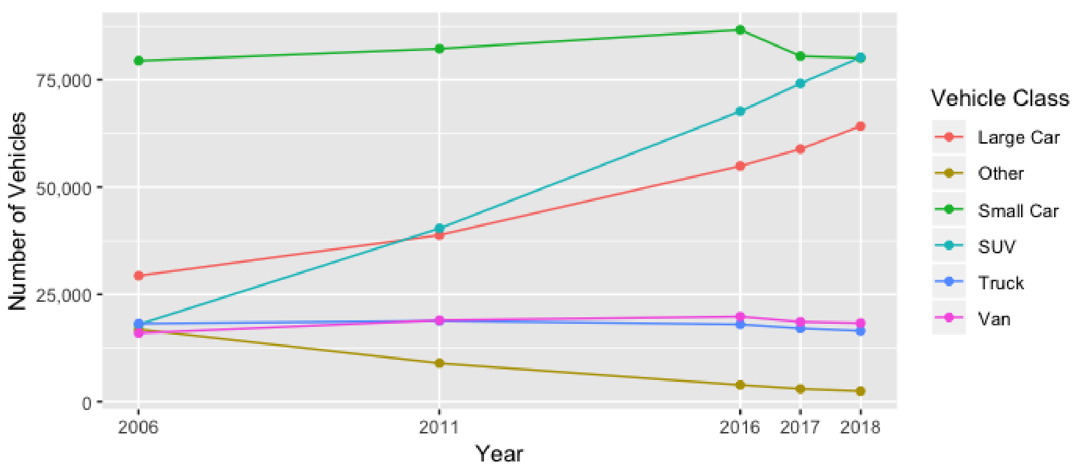


Figure 5: Changes in Surrey's vehicles stock by class and year. A large increase is SUVs and no growth in small cars resulted in SUVs surpassing small cars as the largest class in 2018.

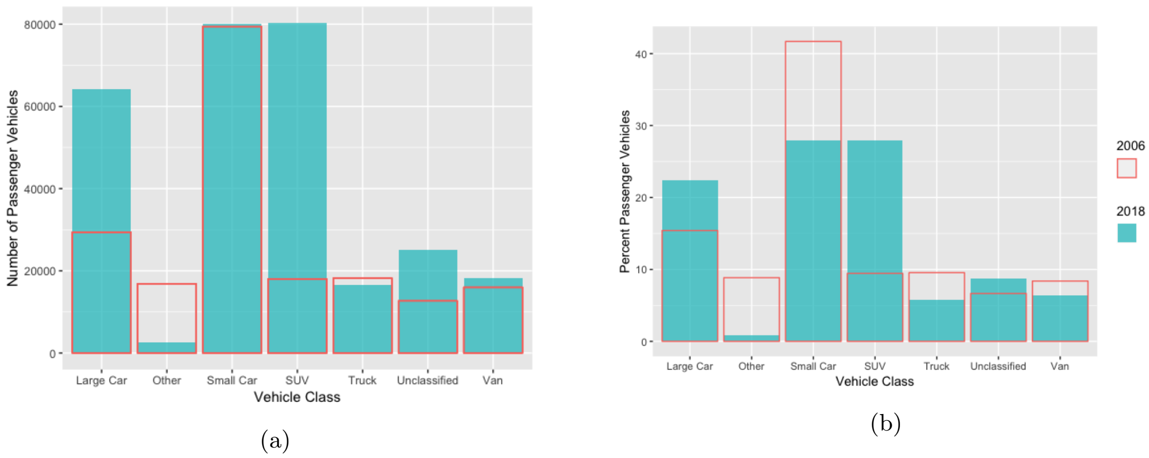


Figure 6: Change in the vehicle stock between 2006 and 2018 by count (a) and percent market share (b). In both cases, we see growth in the SUV and large car classes. By count, we see nearly no change in small cars and small changes in trucks and vans. However, from the perspective of market share these classes suffered significant loses. The unclassified category represents registrations that could not be matched to a vehicle model in the EPA dataset.

increasing from 18,000 SUVs in 2006 to 80,094 SUVs in 2018. These changes are visualized in figure 6.

As the population increased by nearly half during this period, we would expect that the counts of all of the vehicle classes to grow if buyer preferences had stayed the same. As the number of vehicles per capita slightly rose during this period, the classes with negative growth or slight positive growth do not represent a trend away from vehicle ownership, but instead are part of a shift in market share between vehicle classes.

When we look at the classes from the perspective of market share, we see that small cars, trucks, vans, and other vehicles have all lost market share to large cars and SUVs. Small cars, other vehicles, trucks, and vans lost 14%, 8%, 4%, and 2% of market share respectively (see fig. 6). The loss of small car market share is a challenge for electric vehicle adoption, as almost all electric vehicle models currently available are small cars. The market share lost in these categories was spread between SUVs, large cars, and vehicles that we were unable to classify with our model. SUVs gained a huge 19% of market share while large cars and unclassified vehicles increased by 7% and 2% respectively. There has been recent growth in the number of EVs available in the large car class, suggesting consumers will soon have more choice in this area. Currently the only electric SUV on the market in the Tesla Model X, but other manufacturers are developing competing models.

3.1.3 Luxury Growth

The luxury vehicle market in Surrey also grew rapidly between 2006 and 2018. In 2006, luxury vehicles represented 4% of the vehicle stock but grew to 11% by the end of 2018. Increased market share combined with overall growth means that this increase represents 285% growth and 23,727 additional luxury vehicles on the road. Luxury vehicles are not evenly distributed through Surrey, with many areas in the south of Surrey exceeding 20% luxury vehicles. The Central Semiahmoo Peninsula is a notable outlier with over 30% luxury vehicles.

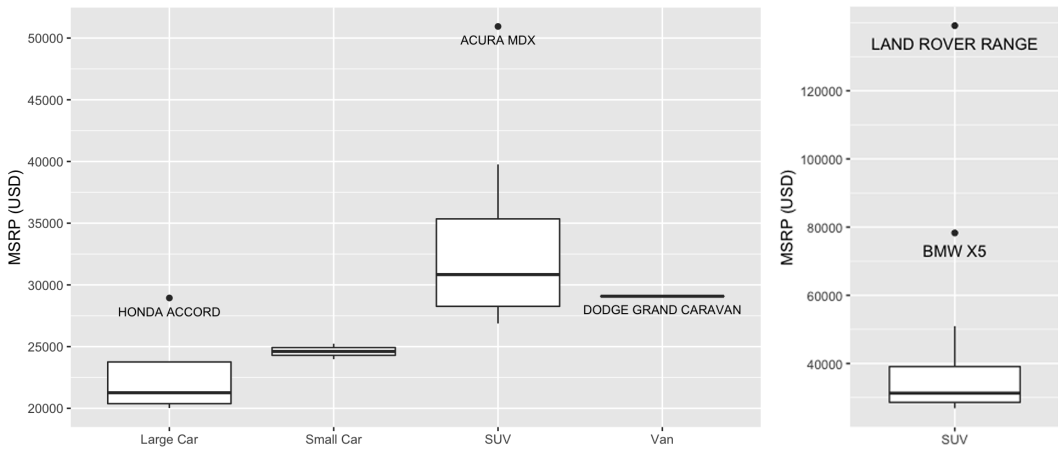


Figure 7: Mean MSRP prices of popular new vehicles in Surrey. Consumers are willing to pay more for SUVs and vans than large and small cars. Three luxury models show up as outliers in the data, all in the SUV class.

3.1.4 Popular Models

We decided to look at vehicle models which had more than 1,000 registrations from model years 2016 or later to get an idea of what vehicles consumers have been buying and their prices. We chose models from 2016 onwards to reflect the new vehicle market, as potential EV buyers will mostly need to buy new EVs due to the small potential pool of used vehicles. Twenty-two models met our requirements, consisting of 2 small cars, 4 large cars, 1 van, and 15 SUVs. By far the most popular vehicle was the Honda Civic, where Civics with a model year from 2016-2018 account for 3% or the Surrey vehicle stock alone.

The price of popular vehicles varies by vehicle class, with consumers willing to spend more for SUVs and vans than small and large cars (see fig. 7). Large and small cars generally fall in price between $20,000 and $25,000, with the Honda Accord being an outlier at $28,000. SUVs generally ranged in price between $26,000 and $36,000 with a mean price of $32,000. However, there were three distinct outliers with the Acura MDX around $50,000, the BMW X5 around $ 60,000, and the Land Rover Range around $100,000.

3.1.5 Trend Toward Higher MPG

While there has been a trend away from small cars towards large cars and SUVs, the loss in market share of trucks and vans combined with increases in fuel efficiency across the board have resulted in an increase in average fuel efficiency (see fig. 8). The average mpg of a vehicle registered with ICBC increased from 21mpg in 2006 to 24mpg in 2018. Increases in fuel efficiency were seen in every vehicle class, through the increase for large cars was particularly pronounced, with the average mpg of a large car registration increasing from 21mpg to nearly 27mpg. This increase for large cars closed the gap in fuel efficiency between large and small cars from 4mpg to under 1mpg.

3.2 Low-Emission Vehicle Stock Trends

3.2.1 Hybrids

Hybrids are defined as any vehicle with multiple fuel sources. The vast majority of hybrids in the dataset are traditional hybrids, such as the Toyota Prius, that use regenerative breaking to capture and reuse some of the energy the car would otherwise dissipate. Plug-in hybrid electric vehicles (PHEVs) are a small

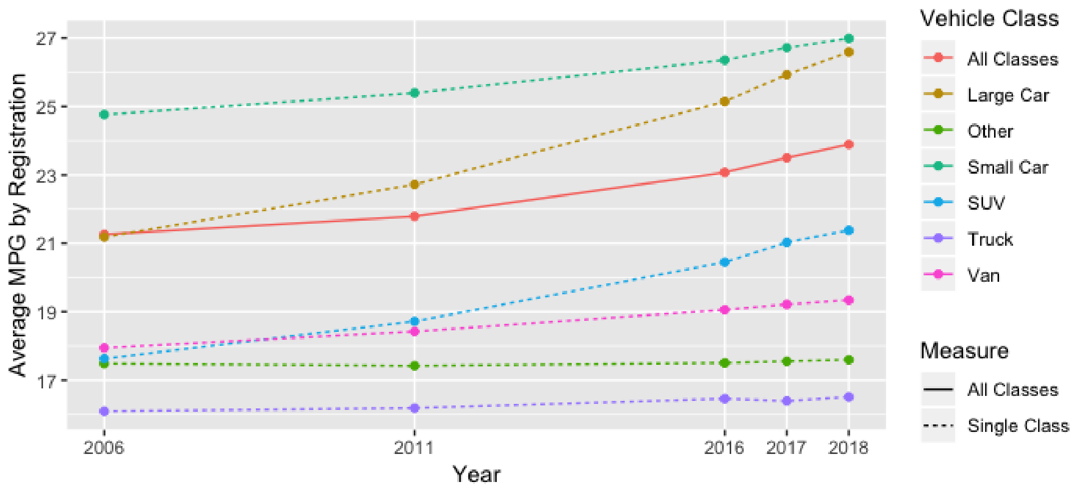


Figure 8: The mpg of vehicles increased for all vehicle classes between 2006 and 2018. This resulted in a 4% increase in mpg for the average ICBC registration.

subset of the hybrid dataset that consists of vehicles that default to using electricity from an on-board battery but can also use gasoline to extend their range.

In 2006 there were only 206 hybrids in Surrey representing 0.1% of the vehicle stock but by the end of 2018, hybrids had made up over 2% of the vehicle stock. PHEVs made up just over 0.1% of the vehicle stock at the end of 2018. However, as ICBC does not flag PHEVs, the true numbers may be greater than the number we were able to identify in the data. See section 2.1.1 for more information on the challenges of identifying PHEVs. Hybrids make up 2-4% of the vehicle stock in most Areas, with a few notable outliers such as 11% in East Whalley.

3.2.2 Electric Vehicle Totals

Electric vehicles make up less than half a percent of the current vehicle stock in Surrey, however the market is growing rapidly. The first electric vehicle appears in 2011 as a single Smart car. By 2016, there were 313 electric passenger vehicles in Surrey and the total doubled each of the next two years, reaching a total of 1226 EVs by the end of 2018. Electric vehicles are not evenly distributed in Surrey, with five Areas where over 1% of the vehicle stock is electric. These Areas are mostly in the south and include Panorama Ridge, Crescent Beach, Grandview Heights, and North Grandview Heights. The Cloverdale Industrial area bordering on Langley is a strong outlier with 3% of the vehicle stock being electric.

Areas differ greatly in population, so the Areas with the largest number of electric vehicles do not correspond to the Areas with the largest percentage of electric vehicles. South Semiahmoo Peninsula has the largest number of electric vehicles with 104 at the end of 2018. Fraser Heights and East, West, and South Newton all had over 50 electric vehicles at the end of 2018. These 6 Areas account for less than a third of the EV stock, indicating EV adoption has not been isolated to a few Areas.

3.2.3 Popular Electric Vehicle Models

In addition to total counts, we were interested in how many models of EV existed in Surrey and which models were most popular. Only 21 models of EV were found in the ICBC data. Of these vehicles, the top five models account for two thirds of the vehicles registered in 2018. At the end of 2018, there were 266 Nissan Leafs and 195 Chevrolet Volts, together making up a third of the EV stock. The next three

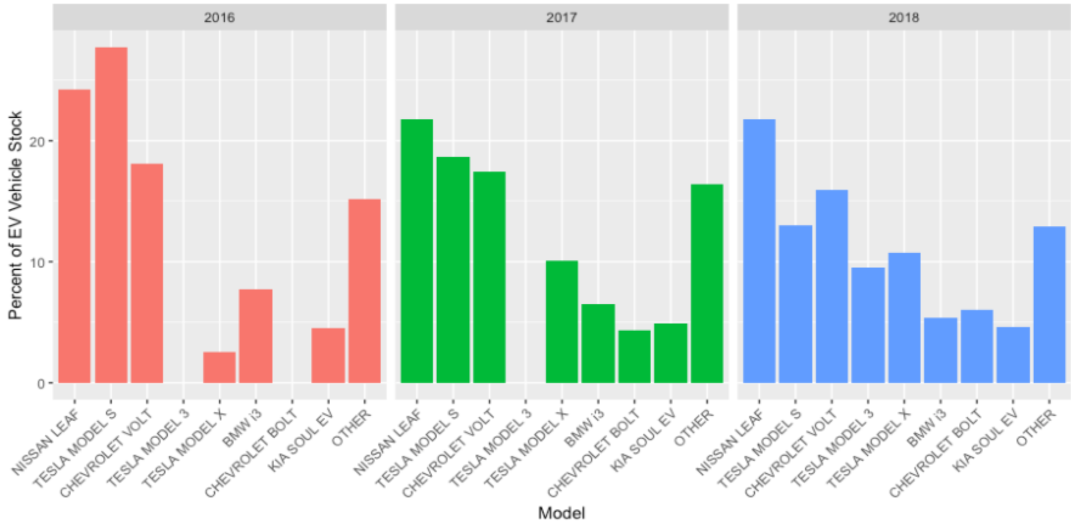


Figure 9: EV Models by percentage of EV market by year. The electric vehicle market has been diversifying through more models and more even market share.

most common models were the Tesla S, X, and 3 making up another third of the EV stock.

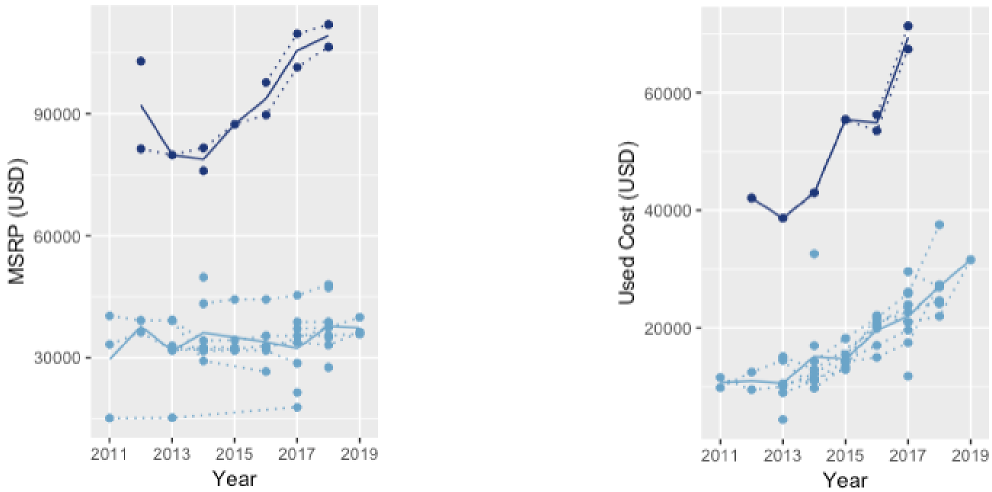
While EV market is currently dominated by a few models, it has shown a trend towards more diversity over time (see fig. 9). In 2016 there were only 15 EV models and the top three models made up 70% of the EV stock. In 2017 the number of models increased to 20. The top three models remained the same but fell to only 58% of the EV stock. In 2018, the number of models increased to 21 and the same top three models fell to 50% of the EV stock.

3.2.4 Electric Vehicle Pricing Trends

We wanted to explore the prices of the EV stock in Surrey to find the entry price for an EV, see what price range of EVs are most popular, and test if EV prices have been decreasing over time. Using US MSRP data from 2011 to 2019 on models found in the Surrey vehicle stock, we found that electric vehicles fell into distinct low and high price clusters (see fig. 10). The low price cluster ranged from a two-seater Smart car for around $15,000 to the BMW i3 for around $45,000. The mean price of an EV in the low price group was $34,500. The most numerous models in the low group were the Nissan Leaf around $33,000 and the Chevrolet Volt around $35,000. The high price cluster consisted of only the Tesla Model S, the Tesla Model X, the Cadillac ELR, and the Fisker Karma. These models ranged in price from $ 76,000 to $112,000 with a mean price of $94,000.

The two groups are also present in the used prices, with low priced used EVS ranging from $4,500 to $37,500 USD with a mean of $18,000 and high priced EVs ranging from $38,500 to $71,000 with a mean of $ 53,000. While these prices are much more affordable, the market would quickly saturate if there was more demand as there are only 1,200 EVs in all of Surrey.

If we look at the change in price overtime, we don't see distinct price decreases for MSRP in either high or low priced models. The MSRP price of the high priced models is generally increasing overtime, though due to the low number of models the 2011 mean MSRP value was skewed high by a single Fisker Karma. The MSRP mean for the low priced category was bounded between $30,000 and $ 37,500, with individual vehicles showing slight increases or decreases in price. The used dataset unsurprisingly shows that the resale value of a model rapidly decreases as it ages.



(a) New

(b) Used

Figure 10: Average price of low-cost (light blue) and high-cost (navy) electric vehicle models for both new and used vehicles. The dashed lines indicate the prices of individual models, while the solid lines indicate the average over all models in the cluster.

If we look at the clusters with respect to number of vehicles and market share, we see that both categories have been growing in number of vehicles between 2016 and 2018. However, low priced vehicles have increased in market share from two thirds to over three quarters during the same period.

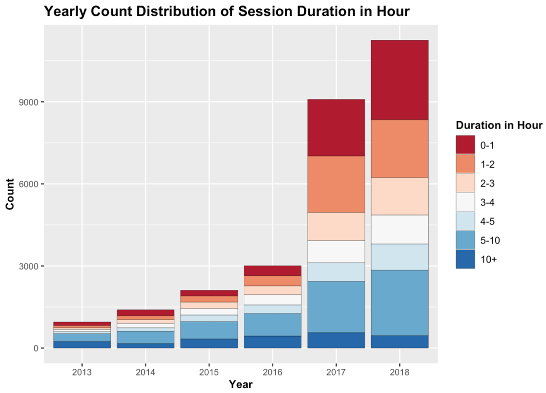
3.2.5 Comparison of EV Prices to Popular Models

Popular electric vehicles from the low-cost cluster cost in the $30,000-40,000 USD range, exceeding the cost of popular traditional small and large cars which genrally fall in the range of $20,000-25,000 USD. Electric vehicles from the low-cost cluster are in the same price range as popular SUVs. However, it is unclear if consumers are willing to pay the same price for a small electric car as a traditional SUV. In terms of luxury vehicles, only luxury SUVs showed up in our popular vehicle data and ranged from $80,000-130,000 USD. The high-cost electric vehicles were around the low-end of the cost range for the luxury SUVs, costing $90,000-100,000 USD. Like the low-cost cluster, it is unclear if consumers will pay the same amount for luxury vehicles of different classes.

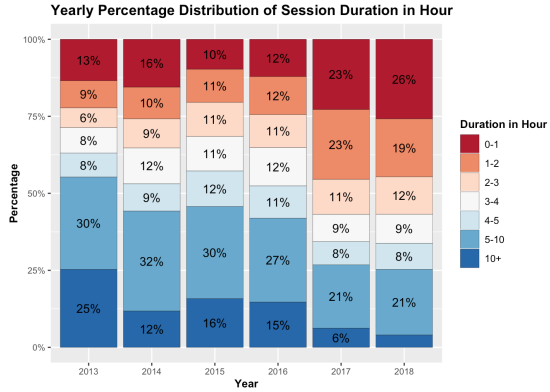
3.3 Charging Session Trends

The analysis in this section is based on incomplete data. One of the issues was that we only had data from 13 of 24 sites run by the companies ChargePoint, Flo, and Greenlots. Further, for the sites where we had data, the data often had peculiarities such as sessions with negative charging times or gaps in data of a year or longer. See figure 2.1.3 for more information on the limitations of the charging site data.

To deal with the data quality issues, we filtered sessions based on several criteria. First, any charging session with a kWh ranging from 0kWh to 500kWh was classified as an ordinary charging session. A charging session with over 500kWh would be considered an outlier and was dropped from the dataset. Second, any charging session with a charging duration ranging from 0 minutes to 48 hours was classified as an ordinary charging session. A charging session that lasted over 72 hours would be considered an outlier and dropped from the dataset. Lastly, any session with a charging duration ranging from 0 minutes to



(a)



(b)

Figure 11: The number of charging sessions and the distribution of charging session length by year from 2013 to 2018.

2 hours was categorized as a short charging session and any sessions lasting more than 5 hours were categorized as long charging sessions.

3.3.1 Increasing Number of Sessions and kWh

We are interested in whether the existing charging sites are being utilized in terms of the amount of (1) vehicle charging time and (2) the amount of energy being consumed. The number of charging sessions tripled in the past three years with a sharp increase from 2016 to 2017. The changes in the total amount of kWh by year follow the same trend as the number of charging sessions. While our data only represents a small portion of the charging sites in Surrey, these trends suggest that infrastructure is being utilized.

3.3.2 Decreasing Charging Session Duration

In addition to the amount of time a charger is in use, we are interested to see if chargers are serving many consumers, in other words, whether individual vehicles are using the charger for most of the day. We used the average length of the charging sessions as a proxy for whether this was occurring and found that there is an increasing number of shorter charging sessions as well as a decreasing number of longer sessions. Combining with the increasing number of charging sessions over the years, there is a clear trend that both consumer demand for EV charging sites in Surrey and the utilization of existing charging sites are increasing.

3.3.3 Weekly Charging Patterns

The three peak times in a day for people to start charging are 7-8 am, 12-1 pm, and 4-6 pm. While there are more charging sessions in the day time, the average charging session duration is drastically shorter during the day than the night.

On a weekly scale, the share of different charging durations approximately stays the same, while a peak is observed in the middle of the week for the number of charging sessions.

4 Electric Vehicle Adopter Modelling

The purpose of the EV adopter model is to answer the following questions:

1. Where are the current EVs located in Surrey?
2. Is EV adoption correlated with hybrid and luxury vehicle ownership in Surrey?
3. What are the demographic features of the current EV owners in Surrey?

We attempted to answer these questions with statistical modelling and machine learning approaches such as generalized linear models, hierarchical clustering, and decision trees.

4.1 Modelling Workflow

Axsen et al. (2016) and Axsen et al. (2017) conducted a study to understand the profile of the EV adopters in BC and the Metro Vancouver area. Income, education, age, family size, housing ownership, and housing type are identified to have a positive correlation with EV adoption. These six demographic features are used as a starting point in our modelling.

We used three methods are to understand the correlation between the vehicle stock and demographics and the EV adoption in an Area. We first used generalized linear regression models to check the p-values of the demographic features shown to be significant in literature. Then, we used hierarchical clustering on EV count and proportion. We decided to cluster the Areas into four groups based on the EV count and proportion histograms using hierarchical clustering. We then visualized the distribution of demographic features for each cluster using box plots. If a demographic feature is strongly correlated with EV adoption, there should be little to no overlap between clusters. Furthermore, we took the labels from hierarchical clustering and used decision trees, a common feature selection technique, to verify our findings. Finally, we took the features identified by the decision trees, creating new linear regression models and performing additional statistical tests.

4.2 Results and Analysis

4.2.1 Linear Regression Models

We chose continuous, Poisson, quasi-Poisson, and negative binomial models to model the number of EVs in each Area. Continuous model is a simple linear regression model with proportion variables, whereas Poisson regression model is a type of linear regression model commonly used to model counts assuming a Poisson distribution. Quasi-Poisson and negative binomial models are variations of Poisson model with more general assumptions.

We first fitted the linear models with the luxury and hybrid vehicle data by Area. Based on the results from all four models, the p-values of luxury vehicles are smaller than hybrids, which indicates that luxury vehicles have a stronger correlation with EVs than with hybrids. Hybrid vehicles were also found to be negatively correlated with EV adoption. Furthermore, out of all the demographic variables mentioned in the literature, income was the only statistically significant factor in the proportion model. While most of the demographic features are significant in the count models, our count variables are highly correlated with the population. It seems trivial that for Areas with high EV counts, there tend to be more people with the corresponding demographics.

As mentioned in our workflow above, our feature selection step later also feeds back to linear modelling. We ran the same regressions on the features selected by our decision trees. Although workplace and morning commute is not studied in the literature, they show to be significant in both proportion

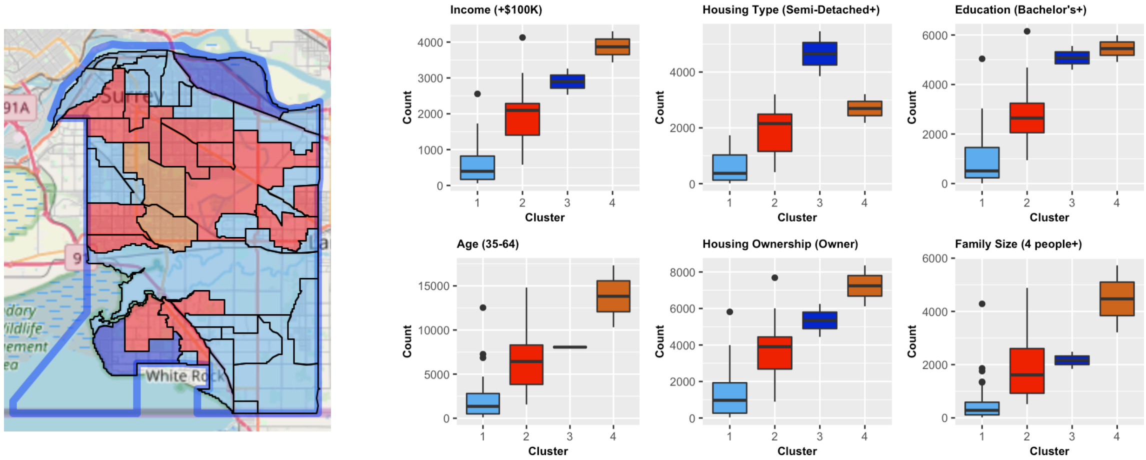


Figure 12: Demographic distributions for the four clusters obtained using hierarchical clustering on EV count. There is a clear separation among the clusters in the distribution of all the demographic features shown to be significant in the study by Axsen et al. (2016) and Axsen et al. (2017).

models and count models. This finding indicates that in terms of demographic features related to EV adoption, Surrey may be different from the Metro Vancouver studies. Moreover, there is a negative correlation between morning commute and EV adoption, meaning that people who live in areas with higher EV adoption are less likely to commute from 5 to 7 am and more likely to work from home.

4.2.2 Hierarchical Clustering on Count

We used hierarchical clustering to group areas with similar EV adopter composition. The hierarchical clustering algorithm would start by putting each data point in its own cluster. Next, it would identify the closest two clusters and combine them into one cluster. This step would be repeated until all the data points are in a single cluster. The way our model determines how close two clusters are is to find all possible pairwise euclidean distances for points belonging to two different clusters and then calculate the average. The method is also known as mean linkage clustering.

To explore how the vehicle stock and demographic features reflect the EV adoption in Surrey, we first clustered areas based on EV count per Area. The clustering algorithm groups Areas with similar EV counts. Out of the four clusters, the dark blue Areas have the highest mean EV count of 100. Although the brown and dark blue cluster has equally high luxury vehicle counts, their ranking in EV counts and hybrid counts are the opposite of one another. Additionally, despite the lack of overlaps when clustered on EV count, the separation of clusters is much less distinguishable when clustered on EV proportion.

The findings can be further interpreted with the following box plots on the demographic features. Based on the literature, the EV adoption of an Area should have a strong positive correlation with the demographic features in the box plots. However, the brown cluster ranks higher than the dark blue cluster in all features other than the housing type. Combining with the brown cluster's high ranking in hybrid vehicle count, these two factors might strongly correlate with the brown cluster's lower EV count.

Further, we created box plots for features not mentioned in the literature. From the plots, the separation is also clear for morning commute and workplace, which suggests a strong correlation. The box plots above indicate that people who live in Areas with a high EV count (i.e. dark blue Areas) are more likely to work from home and are less likely to commute from 5 to 7 am in the morning.

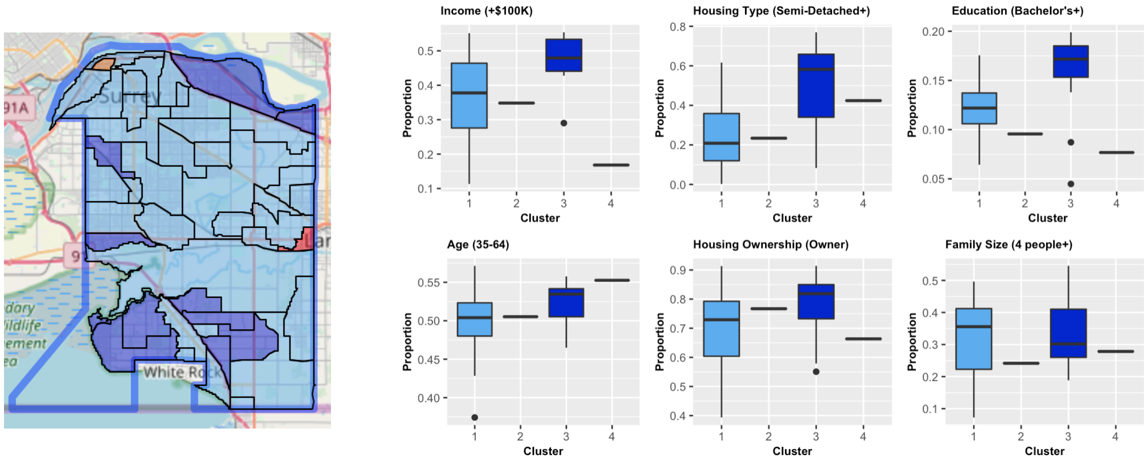


Figure 13: Demographic distributions obtained from hierarchical clustering on EV proportion. There is a clear separation among the clusters in the distribution of selected demographic features shown to be significant in the study by Axsen et al. (2016) and Axsen et al. (2017).

Since our count variables are highly correlated with population, and Surrey has an uneven distribution in the population, the results from the count model could be biased and are difficult to interpret. For example, although some evidence suggests the brown areas satisfy the features that are correlated with high EV adoption, due to the nature of the count model, the reason could be the brown areas simply have a larger population. Furthermore, fitting a model with proportion variables is also a good way to test the findings in the literature, since demographic features should work regardless of the population size. In the next section, we present a different way to understand the correlation between the EV adoption and the vehicle stock and demographics in an area with a proportion model.

4.2.3 Hierarchical Clustering on Proportion

To mitigate the potential bias introduced by differences in population size, we clustered Areas based on EV proportion. The clustering algorithm groups Areas with similar EV proportions.

There are two outliers, Cloverdale Industrial and Bridgeview. Although the total number of EVs in Cloverdale Industrial is just above the average, Cloverdale Industrial has an EV proportion of over 2% and has been a leading area in EV adoption in the past three years (i.e. 2016-2018), whereas Bridgeview is the only area in Surrey with an EV proportion of close to 0% in 2018. On the map, the dark blue Areas have a mean EV proportion of roughly 0.01 (1%), and the light blue areas have a mean EV proportion of roughly 0.004 (0.4%).

The clusters have distinct proportion ranges on the box plots for hybrid and luxury proportion, which suggests a stronger correlation between these factors and the EV adoption in an Area. This matches the results from the linear regression, where both luxury and hybrid vehicles show a statistically significant correlation to EVs. However, out of the features shown to be significant in the literature, although there is a clear separation in features such as income, education, and housing type, there are a lot of overlaps in age, housing ownership, and family size, which suggests a weaker correlation between these demographic features and the EV proportion in an Area.

Lastly, for the demographic features not mentioned in the literature, the separation is much more clear for morning commute and workplace, which suggests a stronger correlation. Further, there is an inversion in the ranking of the morning commute data, meaning people live in areas with a higher EV proportion are less likely to commute early in the morning (i.e. 5-7 am) and are more likely to work from home, which aligns with our previous finding in the count clustering model.

4.2.4 Decision Trees

We used decision trees to discover features that could potentially be correlated to the EV adoption. Decision trees are a machine learning method commonly used in feature selection. The algorithm would decide what factors are important by seeing how successful a factor is at predicting the EV adoption in an area. EV adoption is measured by either EV count of EV proportion. To control overfitting, our decision tree would return a pair of the most successful factors. We consider a pair of factors to be successful when the Areas could be classified into the correct quantile 90% of the time.

We took this approach for two reasons. First, feature selection with decision trees allows us to explore demographic features that are not included in the literature yet could potentially be correlated with EV adoption. Second, using a machine learning method is also a good way to explore a large number of variables and verify the previous findings obtained from the statistical approach.

We separated the count and proportion variables and fit two decision trees. Out of all the count variables, the most successful pair of factors is the number of luxury vehicles and the number of people who work from home. Out of all the proportion variables, the most successful pair of factors is the proportion of luxury vehicles and the proportion of people who commute from 5 to 7 am. We then did statistical tests on the result factors. Both pairs are shown to be statistically significant by all four regression models.

4.3 Conclusion

In the study done by Axsen et al. (2016) and Axsen et al. (2017), EV adopters are grouped into three categories, pioneer, early mainstream adopter, and late mainstream adopter, taking up 1%, 24%, and 75% of the market respectively. However, since every Area has a combination of all three types of adopters, the terminology does not apply to our model.

Based on our analysis of the EV distribution and vehicle stock in Surrey, the highest EV percentage per Area is less than 2.5%. In order to distinguish the difference in EV adoption progress between areas, we introduce the following terms to describe the adopter composition of an area.

- **Current Growth:** EV takes up more than 2% of total passenger vehicles
- **High Potential:** EV takes up between 1 and 2% of total passenger vehicles
- **Mid Potential:** EV takes up between 0.5 and 1% of total passenger vehicles
- **Low Potential:** EV takes up less than 0.5% of total passenger vehicles

Based on the result from our hierarchical clustering model on EV proportion, the dark blue areas with an EV percentage ranging from 0.85 to 1.3% is a mix of high to mid potential Areas, whereas the light blue Areas with an EV percentage range of 0.15 to 0.75% is a mix of mid to low potential Areas.

Although the count models show that the vehicle stock and demographic features have a stronger correlation with the EV adoption in Surrey, the results could be biased due to the uneven distribution in the population of the city. However, both of our count models and proportion models agree on the fact that there is a statistically significant correlation between EV adoption and luxury vehicle ownership, income, education, housing type, morning commute, and workplace. Moreover, EV adoption is shown to have a stronger correlation with luxury vehicles than with hybrid vehicles. A negative correlation was found between EV adoption and both hybrid vehicles and early morning commute times.

Given the count and proportion of the clustering models, the results can be used to support the development and progression of Surrey's EV transformation strategy in different ways. The count model shows the areas with a higher number of EVs and can be used to increase the total number of EVs in Surrey. The proportion model suggests areas with a higher EV proportion and can be used to explore areas with a higher chance to adopt EVs in Surrey. More generally, the modelling results can also be used to understand the current state of Surrey's EV adoption and target potential EV adopters through public awareness campaigns.

5 Charging Site Placement Model

Electric vehicles face a chicken-and-egg problem where consumers are reluctant to purchase vehicles if there is not developed charging infrastructure, while private businesses do not want to invest in charging infrastructure if there are not enough electric vehicle owners to make their investment worthwhile. The City can help address this problem by developing early infrastructure to encourage initial consumers and then let private business continue to develop the charging network. Additionally, as the City's goal is not to maximize profits, it can develop infrastructure in under-served areas that are less likely to see commercial investment.

5.1 Motivation and Scope

During our project, the City was preparing a funding application for twenty curb-side chargers and was interested in finding potential locations. While the majority of charging sessions occurs at home, public chargers are critical for consumers without access to home charging, help reduce range anxiety, and increase public awareness of electric vehicles (Fleetcarma, 2019). We chose to break the problem into two scenarios: looking at placing chargers for employees to use while at work and looking at placing chargers for people to use during other activities including shopping, dining, and recreation.

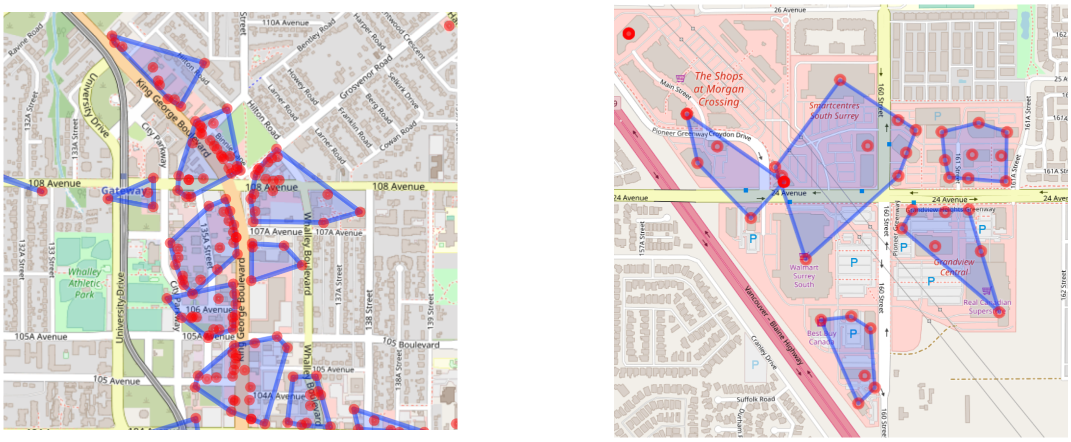
Outside of placing curbside chargers, we had the freedom to choose the criteria for the model, including our optimization criteria and how large of an area we provided for our recommendation. While this allowed us to explore a variety of models, it also created a challenge as different criteria would require different data and would likely lead to different conclusions. Ideally, we could use data from existing sites to understand which sites are used most often and how and when the sites are used. Unfortunately, due to bias and quality issues, we decided that we could not build a reliable model based on existing charging data (see section 2.1.3). As Surrey hopes to move towards a fully zero-emission vehicle stock, we decided to use current vehicles as a proxy for the potential number of electric vehicles in each area.

We searched for existing work on electric vehicle infrastructure in BC and found that UBC's Urban Predictive Analytics lab (UPAL) is studying electric vehicle infrastructure in BC, though their focus has been on infrastructure in residential buildings (Lopez-Behar et al., 019a,b). We met with Dr. Jerome Mayaud, a postdoctoral researcher with UPAL, who suggested we use a gridded model and map the travel time to the nearest charging site for each grid cell in Surrey, similar to recent work done by the lab on access to schools and hospitals in Surrey (Mayaud et al., 2018).

As our data was already aggregated at the TAZ level, we initially used these as our grid cells, calculating travel times between the centroid of each TAZ. While this model placed some chargers in sensible locations, it often placed chargers in rural TAZs because the centroid of the TAZ was in an area without roads, resulting in long travel times. This could be partially addressed by weighting each TAZ by its population or workforce. Even with weighting, the model assumes that chargers have no capacities. So it only placed in single charger in busy urban areas. To address these challenges, we decided to build a second model which identified smaller potential areas for charger placement and then scored these areas using known traffic patterns instead of calculating travel times.

5.2 Site Identification

We decided to increase the resolution of our model by identifying groups of buildings that could utilize a charging site, allowing the City to more easily interpret who would likely use chargers placed in each location. Using the Surrey building inventory and a standard clustering algorithm, we identified clusters of buildings that we no more than 500m apart. We then calculated employment and traffic totals for each cluster using data from the Surrey building inventory and Metro Vancouver traffic model. The



(a) King George Boulevard

(b) Morgan Hills Shopping Centre

Figure 14: Examples of clusters found along King George Boulevard and in Morgan Hills Shopping Centre. While our method only had the constraint that no points were further than 500m apart, the clusters tend to show other desirable properties such as rarely crossing large highways and avoiding intersections or thin parallel polygons.

maximum distance between buildings of 500m ensures that wherever a charger is placed in the cluster, an EV owner would need to walk a maximum of 500m to reach their destination. A minimum variance clustering algorithm as used to favor spherical clusters, which would allow walking distances of much less than 500m if a charger was placed near the centroid of the cluster. For the work scenario, we only clustered based on buildings classified for governmental, office, retail, and industrial use, identifying a total of 636 clusters. For the recreation scenario, we only clustered on retail buildings, resulting in 312 clusters.

Visualizing the clusters, we found that they generally seemed reasonable, forming non-overlapping polygons that often respected natural boundaries, such as not clustering stores on opposite sides of a large highway into the same cluster (see fig. 14). From the visualization, we also found that parking areas were often more than 500m from the locations they served, suggesting that we could increase our clustering radius. However, as smaller clusters would allow us to place more chargers in busy areas without modelling charger capacity, we chose to retain the 500m limit.

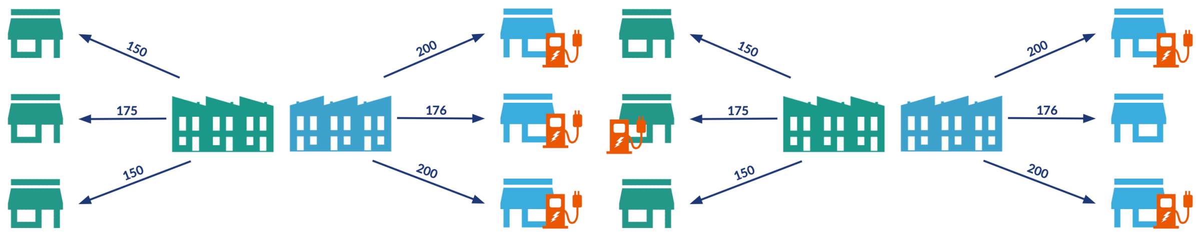
5.3 Scoring Methods

To rank the clusters, we started with a simple objective which used single vehicle traffic as a proxy for potential charger use. As we only had traffic data for TAZs as opposed to clusters or individual businesses, we used the proportion of businesses from the TAZ in the cluster to weight the amount of traffic from the TAZ assigned to the cluster. More precisely, let C_i be a destination cluster, T be the set of all TAZs, T_i be the set of TAZs C_i intersects, and $F(TAZ_a, TAZ_b)$ be the amount of traffic from TAZ_a to TAZ_b.

$$\text{Score for } C_i = \sum_{j \in T_i} \left[\frac{\#\text{ properties in } TAZ_j \text{ and } C_i}{\#\text{ properties in } TAZ_j} \cdot \sum_{k \in T} F(TAZ_k, TAZ_j) \right]$$

To account for larger stores likely receiving more traffic, we revised the score to use the portion of employees instead of the portion of businesses:

$$\text{Score for } C_i = \sum_{j \in T_i} \left[\frac{\#\text{ employees in } TAZ_j \text{ and } C_i}{\#\text{ employees in } TAZ_j} \cdot \sum_{k \in T} F(TAZ_k, TAZ_j) \right]$$



(a) Maximum Traffic Scenario

(b) Improved Access Scenario

Figure 15: Consider placing three charging sites between the six retail locations shown, where traffic is shown by the arrows from the residential areas. The maximum traffic model would place all three chargers at the stores used by residents of the blue area, providing no access for residents of the green area. As EV owners do not need to charge on every trip they make, it would be sensible to place one charging site at a business used by the green area. In this case, there was only one trip difference between the highest traffic site used by the green area and the lowest traffic site used by the blue area. However, the choice of whether to move a charger becomes less clear as the difference in traffic between the sites increases.

For the work scenario, we used morning traffic as a proxy for work commutes. For the recreation scenario, we used midday traffic as a proxy for shopping trips. We refer to the score using the proportion of employees as the maximum traffic objective.

Finally, we created a third objective to incentivize improved access across the city. Because our score does not capture who uses the chargers, it could place many chargers in an area with a population slightly greater than another area and no chargers in the second area. See fig. 15 for an illustration. To incentivize more even access, we created a third objective that weights the traffic score from the first objective by the proportion of traffic without charger access:

$$\text{Score for } C_i = \sum_{j \in T_i} \text{Traffic from } TAZ_j \text{ to } C_i \cdot \text{Proportion of traffic from } TAZ_j \text{ without charger access.}$$

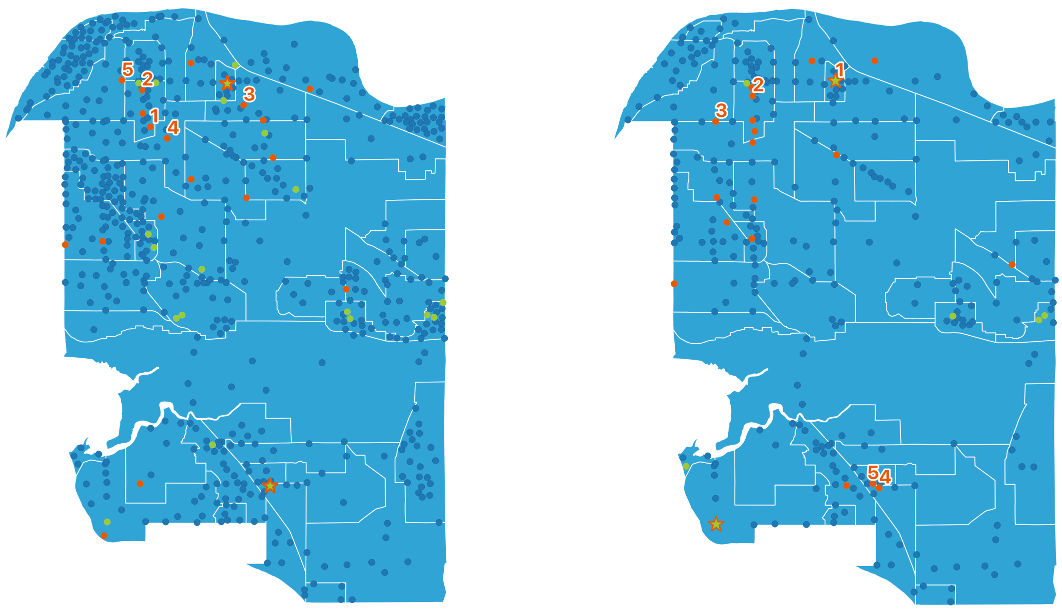
Since the score for each area changes as new areas are added to the set of selected sites, a simple forward selection model may not choose the optimal set of sites. Instead, we implemented a selection model which revisits selected sites after each new site is chosen and removes the lowest scoring selected site if there is an unselected site with a greater score than it.

5.4 Results

5.4.1 Work Scenario

The same top twenty sites were selected by both the maximum traffic and improved access objectives for the work scenario, however the order varied slightly by objective (see fig. 16). Sites were mostly placed in City Centre, Guildford, and Newton, though several sites were placed on the Semiahmoo Peninsula and one site was placed in Cloverdale. The top five sites sites represented the areas near Surrey Memorial Hospital, Central City Arena, William F Davidson Elementary School, Surrey School District and the North Surrey Learning Centre, and the intersection of 104th and 132nd St.

There are no existing chargers at any of the top five sites and only two existing chargers across the top 20 sites. Both sites with existing chargers were primarily retail locations. We were surprised that there was no charger at Surrey Memorial Hospital, as we recalled one in the dataset. However, we found



(a) Work Scenario

(b) Recreation Scenario

Figure 16: Placement of top 20 charging sites for the work and recreation scenarios under the imporved access model (orange) compared to existing chargers (green) and potential locations (blue). Stars indicate sites that were chosen by our model and have an existing charging site. The maximum traffic and equal access models placed the same 20 sites in the work scenario and placed 19 of the same sites in the retail scenario.

that the charger for Surrey Memorial Hospital was not within the 500m polygon of the hospital, instead being located in a parking garage further away.

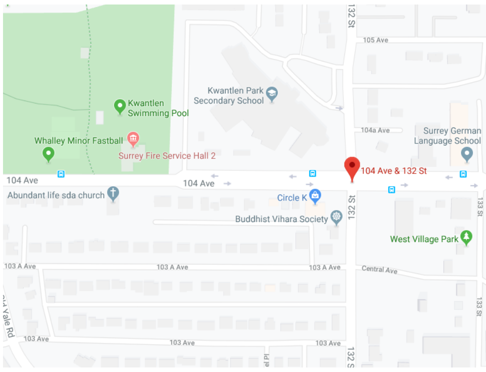
We believe that the third site, William F Davidson Elementary, was selected due to issues with the data, as only 37 employees work at the school. The score for the site was inflated due to a large traffic volume in the Metro Vancouver traffic model, which we cannot readily explain. Only the final destination is counted in the Metro Vancouver traffic model, missing parent drop-offs of students, and we do not see similar numbers for other schools. Outside of the school, the area consisted of a park and single-detached houses.

Finally, the fifth site at the intersection of 104th and 132nd is an interesting selection as the lack of a flagship business would make the site easy to overlook. However, the site receives large amounts of traffic spread between multiple locations including a pool, a fire hall, a church, and a street of small businesses (see fig. 17).

5.4.2 Recreation Scenario

Similar to the work scenario, 19 of the top 20 chargers were placed in the same location using by both objectives in the recreation scenario (see fig. 16). The placement was also similar to the workplace model, with the majority of sites being placed in City Centre and Newton, with a handful also placed in Guildford and on the Semiahmoo Peninsula. The top five sites were placed in Guildford Town Centre, Central City Shopping Centre, Cedar Hills Shopping Centre, and two in Morgan Hills Shopping Centre.

Three of the locations in our top twenty already have charging sites, all of which occur in our top



(a) 104th and 132nd

(b) 72nd and King George Boulevard

Figure 17: The intersections of 104th and 132nd and 72nd and King George Highway were in the top six choices for the work and retail models respectively, despite lacking an obvious flagship location. It is particularly interesting that our model identifies these locations as they could be easily missed by planners.

five choices. However, as only twelve of the 312 retail clusters currently include a charging site, there is certainly room for growth.

While the top five sites were all in large shopping centres, from the sixth site onwards, the recreation model began to choose areas that included groups of businesses that received large amounts of traffic but weren't grouped as a shopping centre. For example, the sixth site was at the intersection of King George Boulevard and 72nd Ave. encompassed over 15 businesses including several restaurants, banks, and a grocery store (see fig. 17).

5.5 Conclusion

We were encouraged that our model first placed sites in areas with flagship establishments, such as shopping centres, hospitals, and sports arenas, and then found areas with large volumes of traffic, which by lacking a single large business could have been easy for planners to overlook. While we were initially surprised that the maximum traffic and equal access models selected the same sites, we believe that due to the size of Surrey, we would likely need to place many more than the 20 sites required for the application to see the same groups utilizing multiple sites.

While the number of charging sites we placed was decided by the funding application we were supporting, this raises a larger question of how many sites would be needed to serve Surrey. While we did not explore the question in depth, our utility scores provide some guidance. The utility scores from the improved access model for both the work and retail scenarios form heavily left skewed distributions, with a relatively small number of sites having much greater utility than the others. While the utility of each site changes as additional charge sites are added in the improved access model, outliers persist even as hundreds of charging sites are placed (see fig. 18). The utility score for a charging site can be used as a heuristic for the benefit of additional charger. For example, in the work scenario when no chargers have been placed the utility of an additional charger is 860. When 50 chargers have been placed the utility of an additional charger is only 263, and by the time 200 chargers have been placed the utility of an additional charger falls to 53.

While our exploration of charging site placement was limited due to the length of the project, we

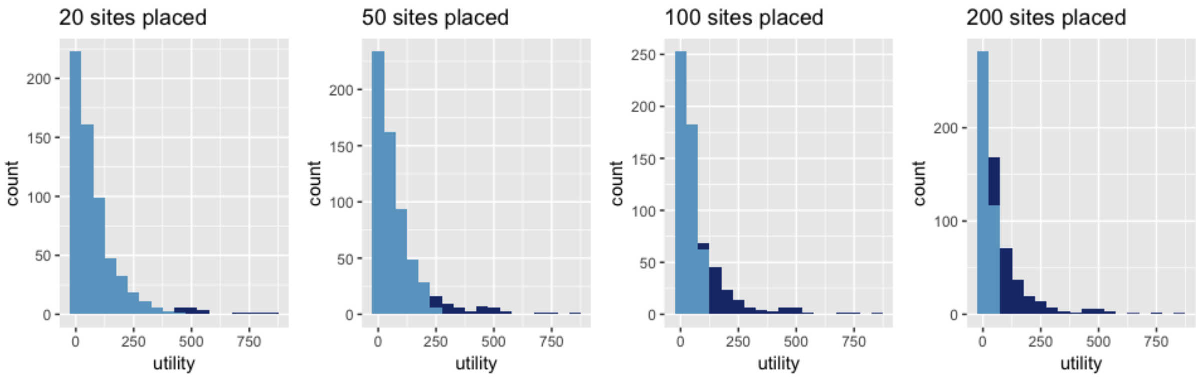


Figure 18: Histograms of utility scores for the 636 work clusters after the given number of charging sites have been placed. The dark blue portions of the histogram indicate sites with chargers. The utility of each site decreases as more chargers are placed, as more people have access to at least one charger. However, outliers with much greater utility scores persist even as hundreds of chargers are placed.

are excited by the initial results and hope the problem is explored further. With additional time, we would have explored how placement changed as we varied the cluster size and where sites would be placed based on current EV ownership instead of overall traffic. Additionally, we hope that with more detailed data collection, future work can be done to explore how current chargers are used, so capacity can be incorporated into placement models.

6 Conclusions and Future Work

With the population of Surrey expected to increase by over 40% by 2040, decisions made by the City now will have a large impact on the city's future landscape (City of Surrey, 2017). The Province of British Columbia has passed legislation requiring 100% of vehicles sold by 2040 to be zero-emission and the City of Surrey has set the goal of transitioning its entire passenger vehicle stock to zero-emission vehicles by 2050 (Province of British Columbia, 2018). Surrey's Electric Vehicle Strategy will help support these goals by increasing consumer awareness, streamlining the process for permitting and installation of chargers, and developing charging infrastructure. However to develop an effective Electric Vehicle Strategy the City needs to understand the vehicle stock, land-use, and demographic landscapes of Surrey and their interactions.

6.1 EV Strategy Explorer

The EV Strategy Explorer app helps support the development of Surrey's Electric Vehicle Strategy by capturing the relationships between existing datasets and providing a simple interface to visualize and explore the data. Through a relational database, the EV Strategy Explorer captures the spatial and temporal relationships between vehicle stock, land use, and demographic data as well as provides a structure to add new datasets as they become available. Further, the database and app allow multiple users to access the data maintained as a single source, limiting confusion arising from multiple versions of the data and streamlining the cleaning process. Creating a common spatial scale and rebasing the data was key to being able to visualize and analyze the interactions between the datasets. The app interface allows users to easily visualize the data as well as the interaction between datasets. By providing several scales, such as visualizing variables by count or proportion, the user can see the data from multiple perspectives.

As both the population and number of electric vehicles have been rapidly increasing in Surrey, it is important that new data is added to the app as it becomes available. In addition to updating the database with new versions of existing datasets, such as ICBC registrations or census data, it would be exciting to add additional datasets such as charger usage or information on secondary housing units. Additionally, as the Electric Vehicle Strategy takes shape, it would be helpful to update the dashboards of the app to make it easy to track metrics measured by the strategy.

6.2 Vehicle Stock Analysis

Our analysis of ICBC vehicle registrations found several trends which could be important for electric vehicle adoption. The market share of vehicle classes significantly shifted in Surrey between 2006 and 2018, with small cars losing 14% or market share and SUVs and large cars gaining 19% and 7% respectively. As most current models of electric vehicle are small cars, this change could pose a challenge for electric vehicle adoption. However, as new models, such as the Tesla Model X SUV, are introduced to the market, consumers may have more options.

The luxury vehicle market is also growing rapidly, increasing from 4% of the market in 2006 to 11% in 2018. In some areas, the percentage of luxury vehicles exceeds 20%, with some areas on the Semiahmoo Peninsula exceeding 30%. This trend is potentially promising for EV adoption, as popular entry-level electric vehicles cost from $30,000-40,000 USD. Popular traditional small and large vehicle models in Surrey generally cost from $20,000-25,000 USD, while popular luxury sedans often exceeded $ 50,000 USD. While many luxury vehicles are sedans or large cars, SUVs are still a significant part of the luxury market with over 1000 each of Land Rovers, BMW X5s, and Acura MDXs entering the vehicle stock since 2016.

The electric vehicle market in Surrey has grown rapidly since the first electric vehicle appeared in the vehicle stock in 2011. The number of electric vehicles has doubled each of the last two years with over 1200 EVs registered in Surrey at the end of 2018. However, electric vehicles still make up less than 1% of the overall vehicle stock. These numbers only capture purely electric vehicles, as plug-in hybrids are not distinguished from other hybrids in the ICBC data.

Currently only 21 electric vehicle models are present in Surrey; however the market has been diversifying both in terms of number of models and market share. In 2016, there were only 15 electric vehicle models present and three models made up over 70% of the stock. By 2018, an additional six models appeared in the vehicle stock and the top three fell to 50% of market share.

Due to the short duration of the project, we were only able to analyze the passenger vehicle stock. Commercial vehicles make up 12% of Surrey's electric vehicle stock and are growing slightly faster than the passenger vehicle stock. Commercial vehicles likely account for much more than 12% of vehicle emissions, as the vehicles are often heavier and used for longer periods of time, leading to lower fuel efficiency and more hours on the road. Studying the commercial vehicle stock and challenges to its electrification is an important direction for future work.

6.3 Charging Session Analysis

Our analysis of charging site usage in Surrey was limited by the quantity and quality of charging data. Unfortunately, We only had access to data for 13 of the 85 charging sites in Surrey, with all data collected at city owned locations such as libraries and recreation centres. Additionally, we encountered data quality issues such as charging sessions taking over a year or of negative duration. However, filtering the data to remove anomalous sessions, we were able to observe several trends. The number of charging sessions at the 13 sites greatly increased from around 1000 in 2013 to over 10,000 in 2018. Additionally, the duration of charging sessions was found to be decreasing, leading to more vehicles using each charger. The peak start times for charging sessions were found to be 7-8am, 12-1pm, and 4-6pm. Charging sessions were found to occur all days of the week with a slight peak in the middle of the work week.

6.4 Electric Vehicle Adopter Modelling

The goal of the electric vehicle adopter modelling was to understand where current electric vehicle owners are located in Surrey, what demographic factors are associated with the areas with high electric vehicle ownership, and to compare these results to previous studies of electric vehicle consumers such as Axsen et al. (2016). Using continuous, Poisson, quasi-Poisson, and negative binomial models we found many of the factors identified by Axsen et al. (2016) to be significant in distinguishing areas with high EV ownership from areas with low EV ownership. Significant factors included income, housing type, education, and housing ownership. While the distinctions were clear when modelling using counts, the results were less clear using proportions. In addition to the factors previously identified in the literature, we found several factors present in our data to be highly correlated with EV ownership, including the number of luxury vehicles in an area, the typical starting commute time, and the number of people that work from home. Interestingly, the the number of hybrids was found to be negatively correlated with the number of electric vehicles.

6.5 Charging Site Placement

The goal of the charging site placement model was to help select sites for curb-side chargers for a funding application. While we hoped to use data from existing chargers to understand usage and demand, due to data quality issues and Surrey's goal to work towards a fully zero-emission vehicle stock, we used traffic as a proxy for potential charger use. We identified sites that would benefit from a charger, clustering buildings with a maximum of 500m between any two buildings in a cluster. We developed a simple scoring method which assigned charging sites to areas with high traffic volumes. As this model did not account for the origin of the traffic, we developed a second scoring method which accounted for the origin's access to a charger, resulting in improved charger access.

Both scoring methods had nearly identical top 20 sites for both work and recreation scenarios. In both cases, the models initially placed sites near flagship locations that would draw large amounts of traffic, such as hospitals, shopping centres, and sports arenas. After identifying these sites, the model began to chose locations that would be more difficult for planners to identify, such as areas where a single charger could serve a pool, church, and several shops.

Due to the short duration of the project, we were only able to explore a handful of scenarios. With additional time, we would have liked to explore placing chargers based on current EV ownership instead of total traffic and study the literature to get a better value for how large of an area a single charger can serve. Additionally, we were unable to incorporate the capacity of charging each charging site into the model and believe this is an important direction for future work.

We are excited to support the City of Surrey's Electric Vehicle Strategy through the EV Strategy Explorer app, our analyses of the ICBC and charging session datasets, and our profiles of electric vehicle adopters and models for charging site placement. We would like to thank our partners at the City of Surrey for their help throughout the project, connecting us with data, helping us shape the project, and helping us test and deploy the app. Additionally, we would like to thank our mentors in the Data Science Institute for their guidance in scoping our project and designing our models. Finally, we would like to that our mentors from Boeing Vancouver and our peers for their ideas and feedback throughout the project.

References

Anwar, N., A. Hong, M. Kramer, and K. Wong (2018). Data Science for Social Good 2018: Transportation Vehicle Modelling for Policy Analysis - Final Report. Technical report.

Axsen, J., S. Goldberg, and J. Bailey (2016). How might potential future plug-in electric vehicle buyers differ from current Pioneer owners? *Transportation Research Part D: Transport and Environment 47*, 357–370.

Axsen, J., S. Goldberg, Z. Long, and C. Kormos (2017). Canadian Zero-Emissions Vehicle Survey Metro Vancouver Analysis. Technical report, Burnaby, B.C.

City of Surrey (2017). Population Estimates & Projections.

City of Surrey (2018). Community Energy & Emissions Plan. Technical report.

City of Vancouver (2019). Electric Vehicles (EVs).

DeMuro, D. (2019). Electric Cars: Benefits and Disadvantages.

EPA (2018). Data on Cars used for Testing Fuel Economy.

Fleetcarma (2019). Charge the North: Results from the world's largest electric vehicle charging study.

Government of Canada (2019). Provincial and Territorial Energy Profiles – British Columbia.

Lopez-Behar, D., M. Tran, T. Froese, J. R. Mayaud, O. E. Herrera, and W. Merida (2019b). Charging infrastructure for electric vehicles in Multi-Unit Residential Buildings: Mapping feedbacks and policy recommendations. *Energy Policy 126*, 444–451.

Lopez-Behar, D., M. Tran, J. R. Mayaud, T. Froese, O. E. Herrera, and W. Merida (2019a). Putting electric vehicles on the map: A policy agenda for residential charging infrastructure in Canada. *Energy Research & Social Science 50*, 29–37.

Mayaud, J. R., M. Tran, R. H. Pereira, and R. Nuttall (2018). Future access to essential services in a growing smart city: The case of Surrey, British Columbia. *Computers, Environment and Urban Systems 126*.

Province of British Columbia (2018). Provincial government puts B.C. on path to 100% zero-emission vehicle sales by 2040.

Statistics Canada (2015). Dissemination Area (DA).

Statistics Canada (2016). Population and Dwelling Count Highlight Tables, 2016 Census.

Translink (2013). 2011 Metro Vancouver Regional Trip Diary Survey. Technical report.

Translink (2016). *Greater Vancouver Regional Travel Model - Development and Validation Report* (Draft ed.).

Union of Concerned Scientists (2015). Cleaner Cars from Cradle to Grave. Technical report.

Western Investor (2017). Surrey: Fastest-growing B.C. city to become future core of the Lower Mainland.