

# Final Project

DS Pros

2024-03-21

**Research Topic:** Electric Vehicle Statistics in Washington  
**Project Idea:** Analyze electric vehicle consumers

**Team Members:** Vedant Pungliya, Michaela Beck, Ekaterina Kurkalova

## Introduction:

The goal of this project is to explore consumers' preferences when looking to purchase an electric vehicle. For this project, we focused on Washington State and analyzed electric cars registered through the Washington State Department of Licensing. In recent years, electric vehicles have become more popular and accessible. We aim to determine what factors are most important to consumers and what impacts their buying decision when shopping for a battery or plug-in hybrid electric vehicle. Our analysis will address consumer preferences regarding popular electric vehicle brands, trends over time, fuel range significance, average range and age of EVs, changes in electric range among top brands, the influence of EV type on consumer choices, and the impact of Washington state location on EV adoption rates.

## Questions to be addressed:

- 1) What vehicle brand is most popular for consumers? (top 20) Histogram / bar chart showing the car brand and quantity purchased. Who dominates the EVs market?
- 2) What brands are the trends for consumers over the years?
- 3) What brand has the highest fuel range? Is fuel range a selling point for consumers?
- 4) Average age of EV? How has the electric range changed based on model year?
- 5) Does the electric vehicle type matter to consumers BEV, PHEV? Which vehicle type has better electric range?
- 6) Compare zip codes with quantities of EVs within different areas of the city. Will demonstrate where in the city EVs are most popular?

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4    v readr     2.1.5
## vforcats   1.0.0    v stringr   1.5.1
## v lubridate 1.9.3    v tibble    3.2.1
## v purrr    1.0.2    v tidyverse  1.3.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Here we have the variable description-

- 1) VIN..1.10.: Vehicle Identification Number (VIN), a unique identifier for each vehicle.
- 2) County: The county where the vehicle is registered.
- 3) City: The city where the vehicle is registered.
- 4) State: The state where the vehicle is registered.
- 5) Postal.Code: The postal code of the vehicle's registration address.
- 6) Model.Year: The year of the vehicle's model.
- 7) Make: The manufacturer or brand of the vehicle.
- 8) Model: The specific model of the vehicle.
- 9) Electric.Vehicle.Type: Type of electric vehicle (e.g., Plug-in Hybrid Electric Vehicle, Battery Electric Vehicle).
- 10) Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility: Eligibility status for Clean Alternative Fuel Vehicle (CAFV) incentives.
- 11) Electric.Range: The electric range of the vehicle in miles.
- 12) Base.MSRP: The Manufacturer's Suggested Retail Price (MSRP) of the vehicle.
- 13) Legislative.District: The legislative district associated with the vehicle's registration.
- 14) DOL.Vehicle.ID: A unique identifier assigned by the Department of Licensing (DOL) for the vehicle.
- 15) Vehicle.Location: Geographic coordinates (latitude and longitude) of the vehicle's location.
- 16) Electric.Utility: The electric utility company associated with the vehicle.
- 17) X2020.Census.Tract: The 2020 Census Tract code associated with the vehicle's location.

## Cleaning the data:

```

##      County    City State Postal.Code Model.Year   Make     Model
## 1      King Seattle    WA      98126      2017 AUDI      A3
## 2 Thurston Olympia    WA      98502      2018 AUDI      A3
## 3 Thurston   Lacey    WA      98516      2017 TESLA MODEL S
## 4 Thurston   Tenino    WA      98589      2021 JEEP WRANGLER
## 5 Yakima    Yakima    WA      98902      2020 TESLA MODEL 3
## 6 Thurston Olympia    WA      98501      2023 JEEP WRANGLER
##
##           Electric.Vehicle.Type
## 1 Plug-in Hybrid Electric Vehicle (PHEV)
## 2 Plug-in Hybrid Electric Vehicle (PHEV)
## 3 Battery Electric Vehicle (BEV)
## 4 Plug-in Hybrid Electric Vehicle (PHEV)
## 5 Battery Electric Vehicle (BEV)
## 6 Plug-in Hybrid Electric Vehicle (PHEV)
##
## Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility Electric.Range Base.MSRP
## 1          Not eligible due to low battery range            16      0
## 2          Not eligible due to low battery range            16      0
## 3 Clean Alternative Fuel Vehicle Eligible            210      0
## 4          Not eligible due to low battery range            25      0
## 5 Clean Alternative Fuel Vehicle Eligible            308      0
## 6          Not eligible due to low battery range            21      0
##
##           Vehicle.Location
## 1 POINT (-122.374105 47.54468)
## 2 POINT (-122.943445 47.059252)
## 3 POINT (-122.78083 47.083975)
## 4 POINT (-122.85403 46.856085)
## 5 POINT (-120.524012 46.5973939)
## 6 POINT (-122.89692 47.043535)

## 'data.frame': 181060 obs. of 12 variables:
## $ County : chr "King" "Thurston" "Thurston" "Thurston" ...
## $ City   : chr "Seattle" "Olympia" "Lacey" "Tenino" ...
## $ State  : chr "WA" "WA" "WA" "WA" ...
## $ Postal.Code : int 98126 98502 98516 98589 98902 98501 98345 ...
## $ Model.Year : int 2017 2018 2017 2021 2020 2023 2017 2020 2023 ...
## $ Make   : chr "AUDI" "AUDI" "TESLA" "JEEP" ...
## $ Model  : chr "A3" "A3" "MODEL S" "WRANGLER" ...
## $ Electric.Vehicle.Type : chr "Plug-in Hybrid Electric Vehicle (PHEV)" ...
## $ Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility: chr "Not eligible due to low battery range" ...
## $ Electric.Range : int 16 16 210 25 308 21 53 322 23 53 ...
## $ Base.MSRP  : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Vehicle.Location : chr "POINT (-122.374105 47.54468)" "POINT (-122.943445 47.059252)" "POINT (-122.78083 47.083975)" "POINT (-122.85403 46.856085)" "POINT (-120.524012 46.5973939)" "POINT (-122.89692 47.043535)"
```

We loaded and installed the libraries and packages necessary to perform our analysis. The libraries and packages included, ggplot2, tidyverse, dplyr, forcats, readr, and leaflet. Our data set was found on Data.Gov. It describes battery and plug- in hybrid electric vehicles registered through the Washington State Department of Licensing. We cleaned the data by filtering to only display vehicles still in Washington State. We removed the ‘X2020.Census.Ttract’, ‘DOL.Vehicle.ID’, ‘Legislative.District’, ‘VIN..1.10.’, and ‘Electric.Utility’ columns that were unnecessary for our analysis.

```
ev_data_2024 <- ev_data %>%
  filter(Model.Year == 2024)

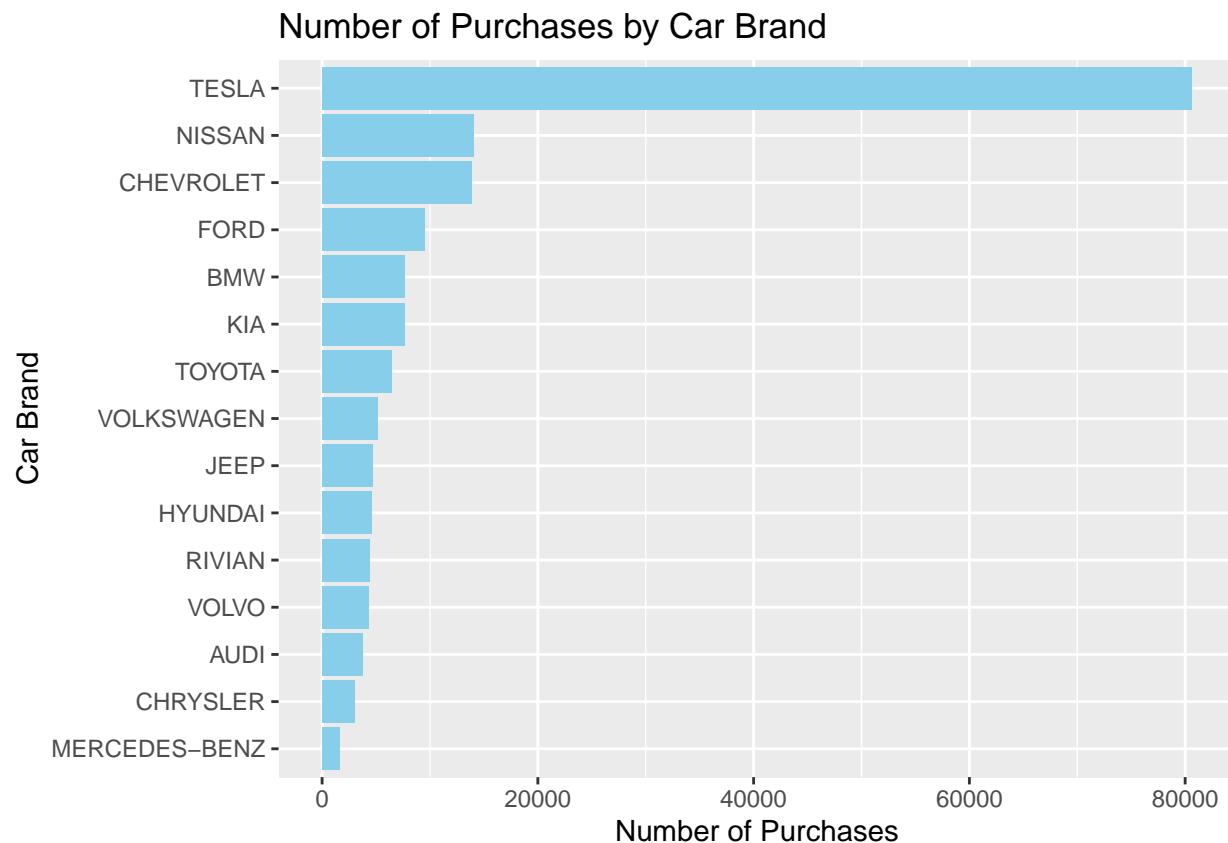
total_sales_2024 <- nrow(ev_data_2024)

print(total_sales_2024)

## [1] 9787
```

Anomalies: The year 2024 cannot be compared to other years since the dataset was uploaded in early 2024 and hence we do not have enough data. With the EV trend growing, the sales for 2024 should definitely be near or more than 2023.

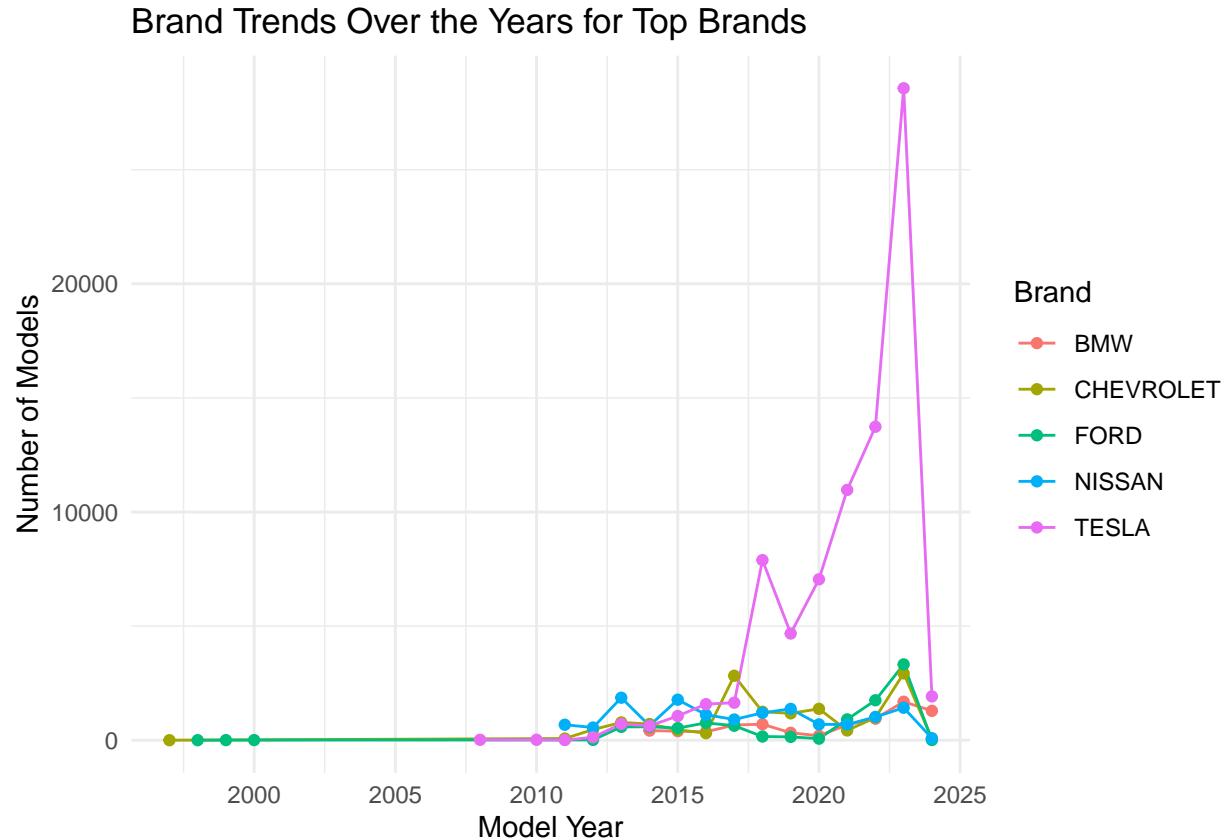
1. What vehicle brand is most popular for consumers? (top 20) Histogram/bar chart showing the car brand and quantity purchased. Who dominates the EV market?



You can see in the bar chart above that the Tesla car brand dominates the market. The number of EVs purchased from Tesla is nearly four times that of its leading competitors, with its number of car purchases reaching around 80,000. The next popular car brand from consumers next to Tesla is Nissan and Chevrolet. The number of cars on the market for Nissan and Chevrolet are very similar, with the number of purchases reported for the brands being around 15,000 each. The other car brands shown on the bar chart are all within a similar range of sales, between 1,000 and 10,000 for the leading car brands.

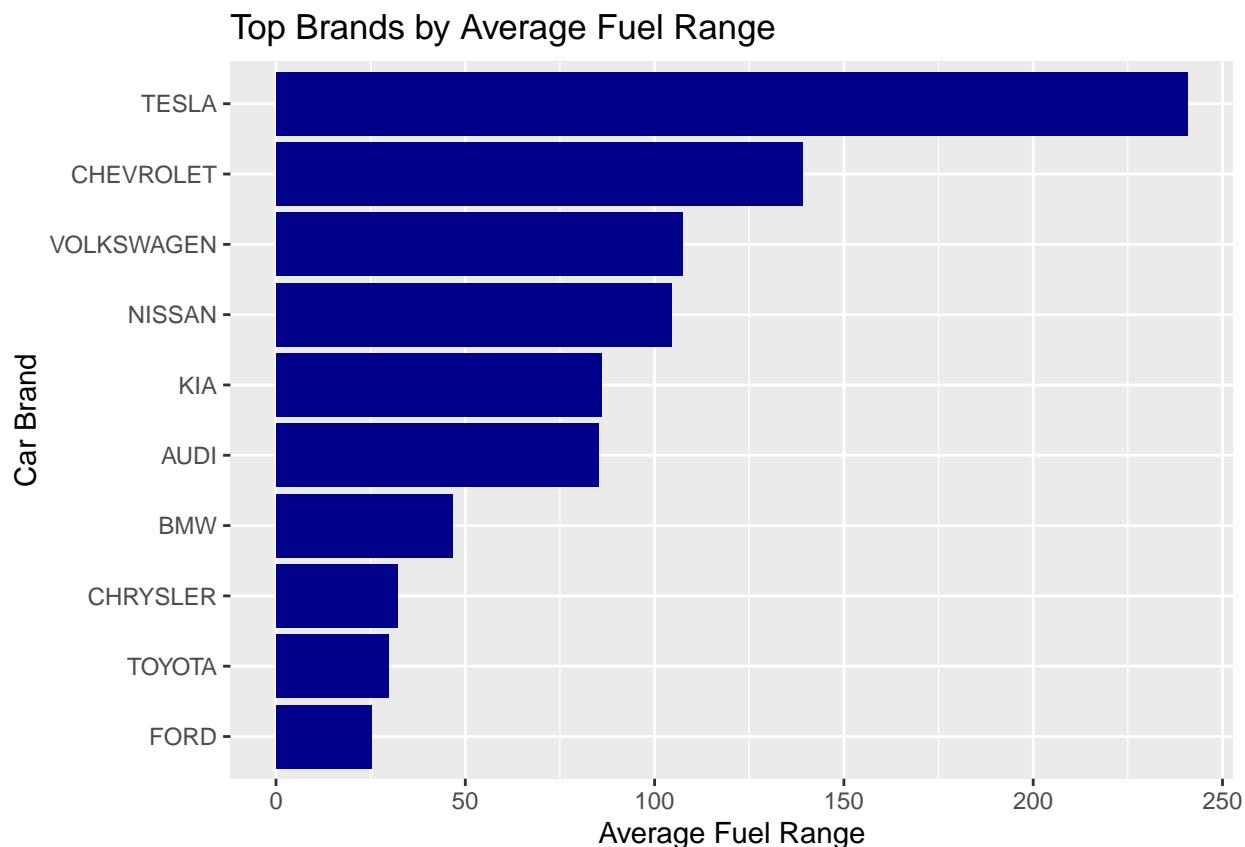
## 2. What brands are the trends for consumers over the years?

```
## `summarise()` has grouped output by 'Model.Year'. You can override using the
## `.groups` argument.
```



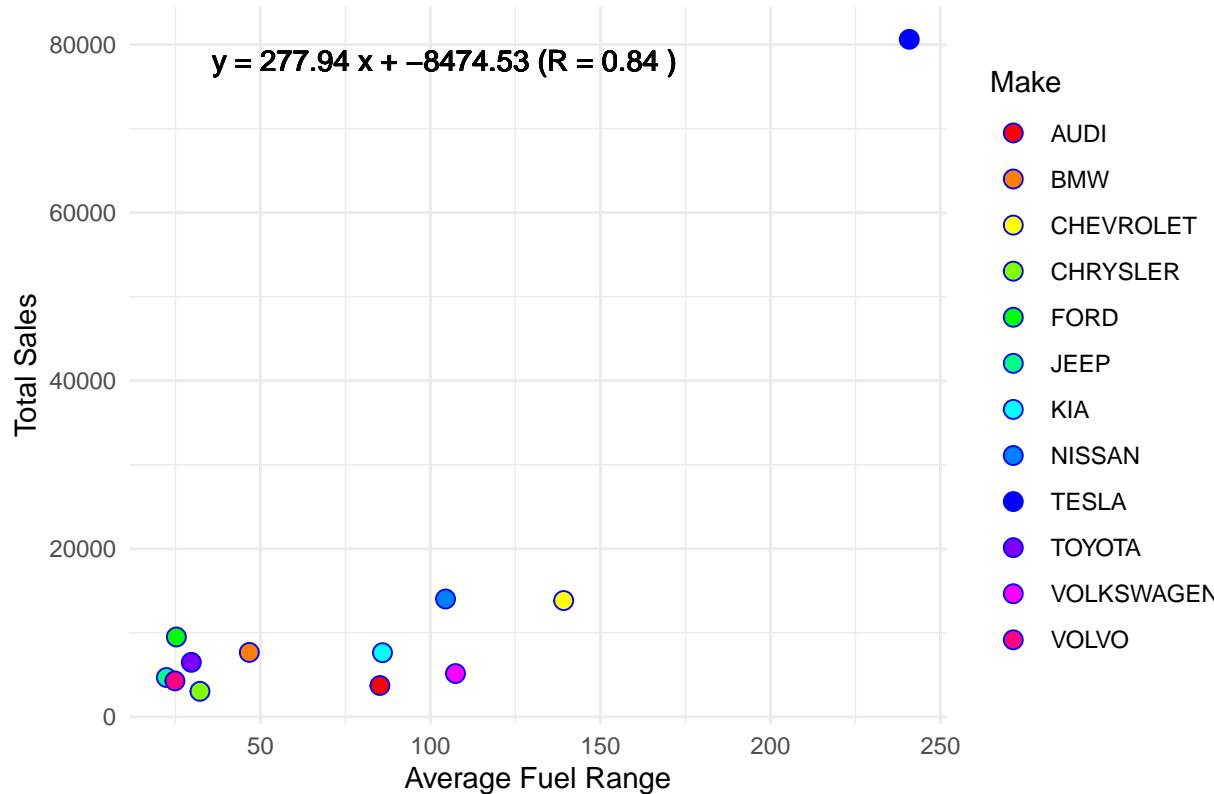
Over the years, many car brands have stayed consistent in their sales. Tesla has quickly risen in popularity. Around 2019, the number of sales took flight and increased significantly faster than its competitors. Based on the quick incline in Tesla's sales, their electric vehicles appear to be what consumers are looking for and choosing over their competitors. It is worth noting Chevrolet had a spike in sales around 2019 but dropped back down to its regular levels a few years later.

3) What brand has the highest fuel range? Is fuel range a selling point for consumers?



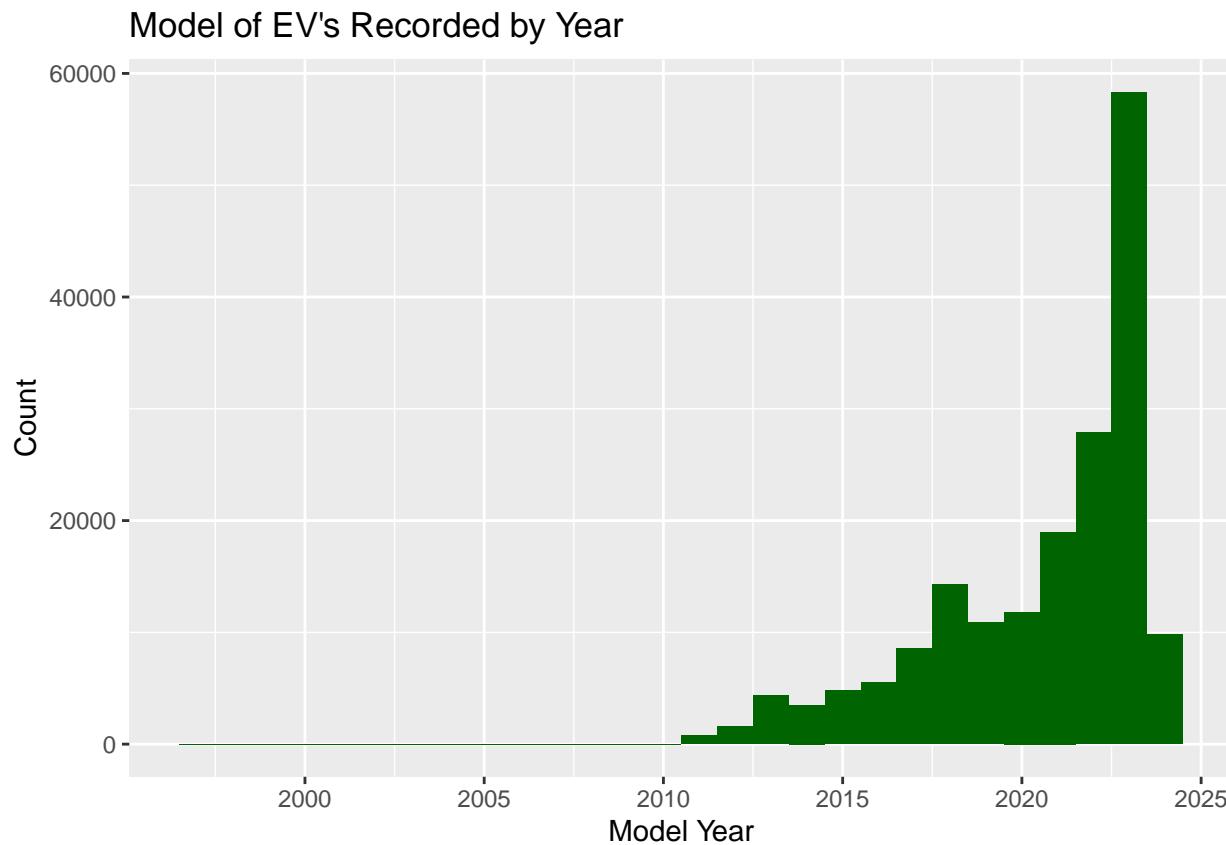
Out of the top brands for EVs, Tesla has the top average fuel range. We can see in the bar chart above that Tesla's fuel range reaches nearly 250 miles while its next competitor, Chevrolet, has a fuel range of around 140. Based on the data, Tesla has a significantly more extensive average fuel range. Leading competitors like Volkswagen, Nissan, Kia, and Audi are similar, with around 90-120 miles of fuel range. Tesla's fuel range is nearly double the fuel range of its competitors. With Tesla having significantly higher sales than any other EV car brand, Tesla's higher fuel range is likely a selling point for consumers.

## Average Fuel Range vs Total Sales

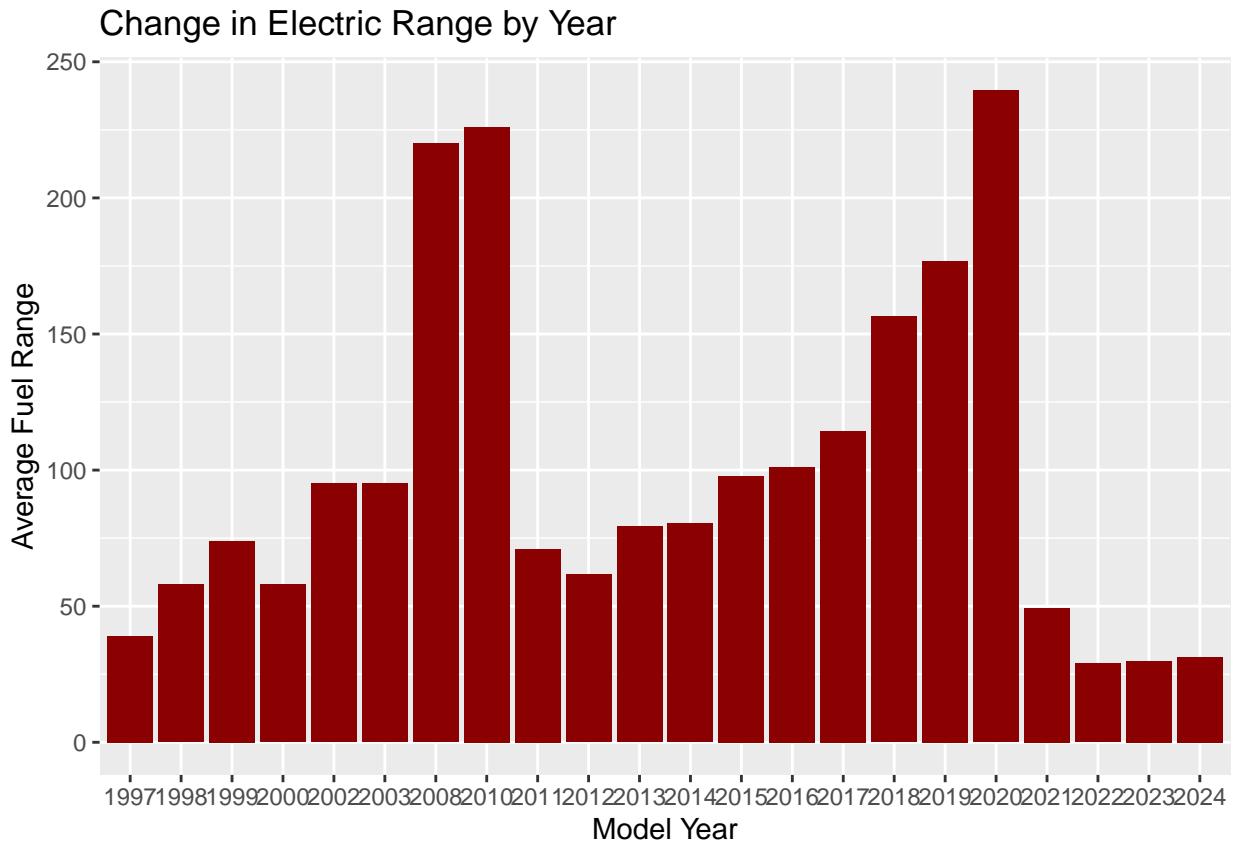


We can see in the scatterplot above a positive relationship between Total Sales and Average Fuel Range. As the fuel range increases, the total number of sales increases as well. We can see the Tesla(Blue dot) at the top, with a significant distance between it and the leading car brands. The leading competitors, such as Nissan(Light Blue) and Chevrolet(white), are also moving in a positive direction upward, with the fuel range positively influencing the total sales.

4) What is the average age of EV? How has the ‘Electric.Range’ changed based on model year?



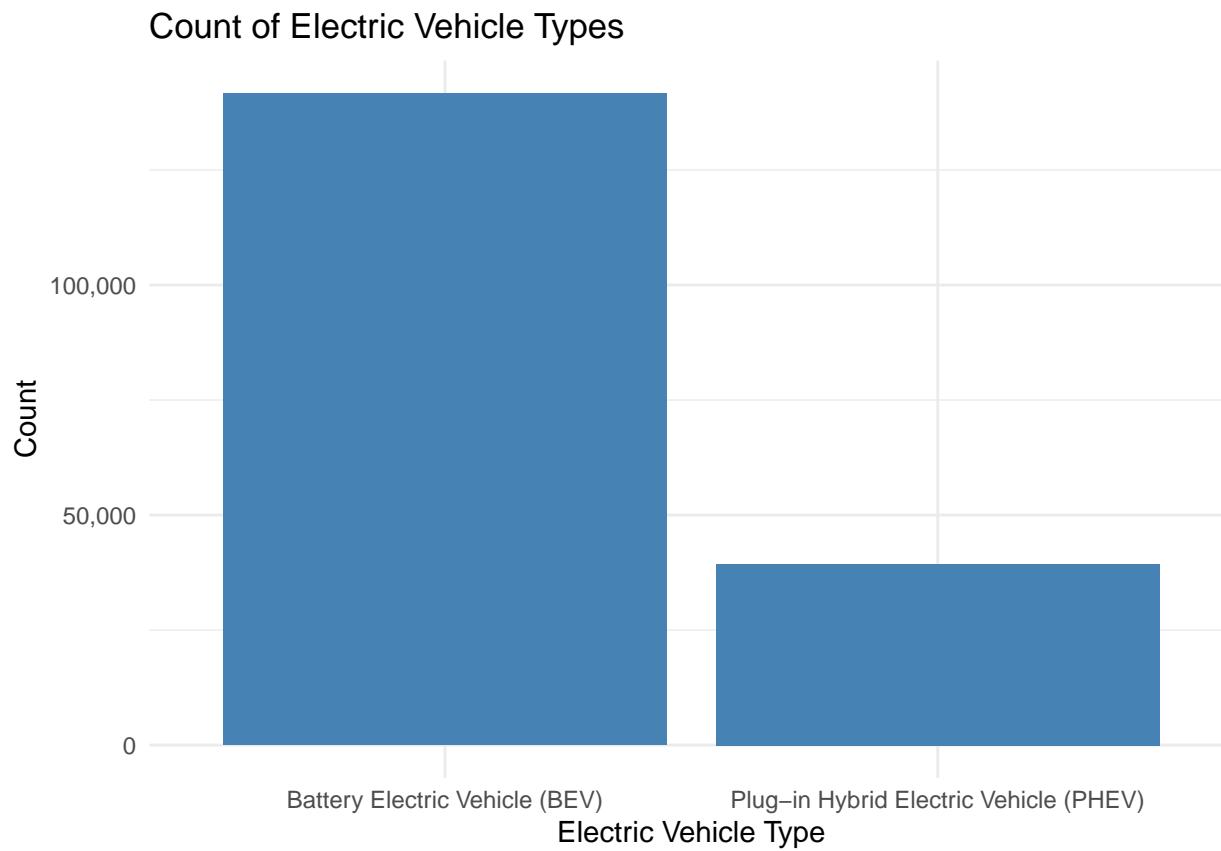
The spread of EV model year is skewed left with a mode of cars being a 2023 model by around double the previous year. There was a small spike in sales of 2018 EV models shortly before the rapid increase began around 2021. The boxplot shows a clearer spread, with any model before 2013 being an outlier and the earliest model being before 2000. This gives the impression that EVs became popular around 2018 and also that there is potentially insufficient data for 2024 vehicles.



Here we see the average electric range of electric vehicles for each model year. No consumers reported their electric range for 2011, 2004-2007, and 2009 models. The highest average electric range, in miles, is for the 2020 model which makes sense because the 2020 Tesla Model S has the highest reported range at 337 miles and has 76 instances, which brings the average up.

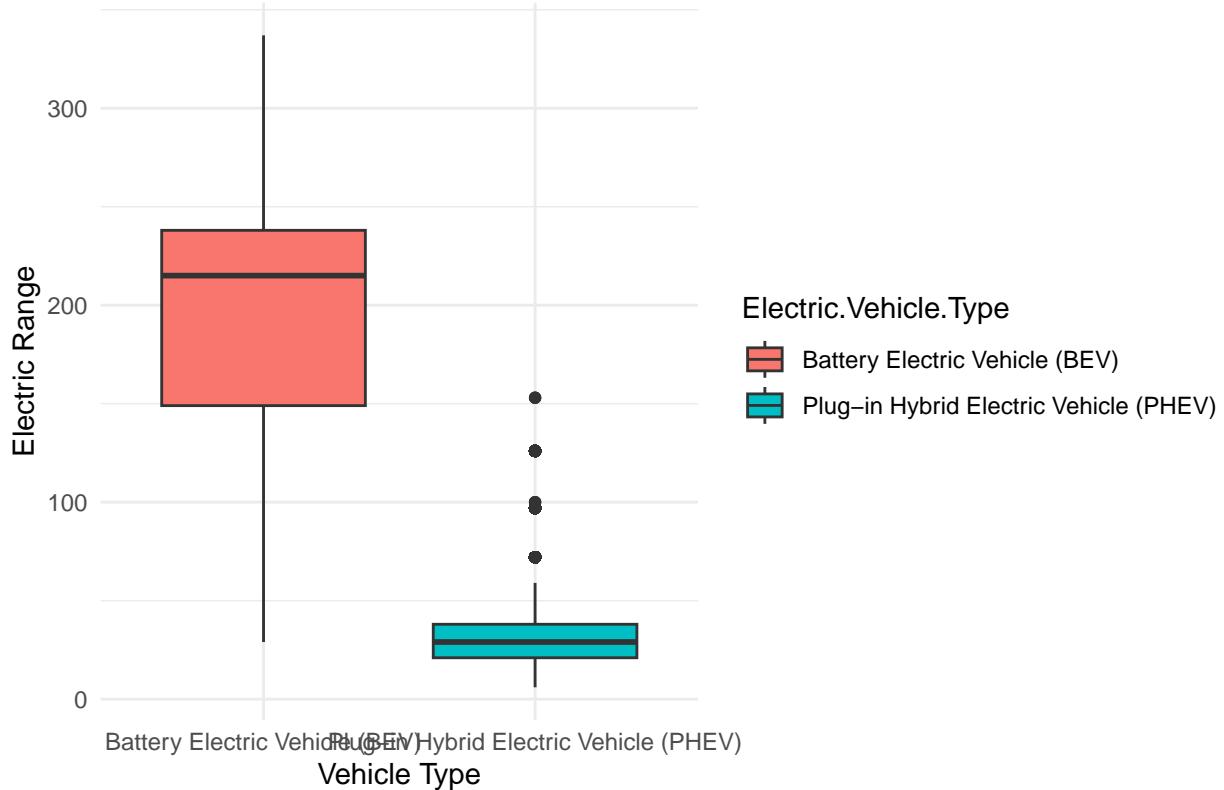
5. Does the electric vehicle type matter to consumers? Which vehicle type has better electric range?

```
## [1] "Battery Electric Vehicle (BEV)"  
## [2] "Plug-in Hybrid Electric Vehicle (PHEV)"
```



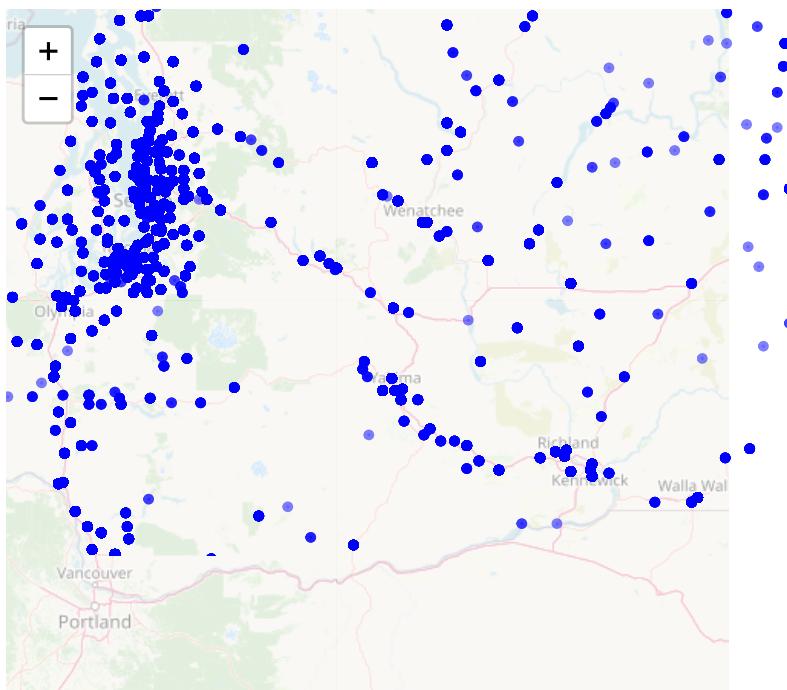
Based on the bar chart most electric vehicles were Battery electric vehicles instead of Plug-in Hybrid electric vehicles. The amount of Battery Electric vehicles in the data was approximately three times that of Plug-in Hybrid electric vehicles. Because of this there is potential for further analysis in whether the vehicle type is a deciding factor for consumers.

## Comparison of Electric Range by Vehicle Type



The box plot shows that Battery electric vehicles (BEV) have a better range than Plug-in Hybrid Electric vehicles (PHEV). However, the range of electric mileage is more spread out for BEV's than PHEVs which means that hybrids have less variation in this category. The average range for BEVs is over 200 miles which is significantly more than the average PHEV range of approximately 25 miles. The box plot presents no outliers for BEVs while there are five outliers presented for PHEVs

**6. How does location in Washington influence the amount of EVs?**



Leaflet | © OpenStreetMap, ODbL

In Washington, there is a wide distribution of EVs across the state. Based on the map above,

**there is the greatest amount of EVs on the state's western side in the Seattle and Tacoma area. The concentration of EVs follows along I-90 from the state capitol, Olympia, up to Everett. The areas with a high density of EVs are urban areas close to the northwestern coast of Washington. On the state's eastern side, there is also a high concentration of EVs around Spokane. The map shows a high concentration of EVs in urban areas or larger cities with more people. In Washington state, the western side has the most significant EVs compared to the rest.**

### **Conclusion:**

After modeling the electrical vehicle dataset of EVs registered by the Washington State Department of Licensing, we have better understood the trends in cars that people buy. Initially, we saw that brand is essential, with Tesla leading the market in recent years. That led us to consider which qualities made Tesla so appealing and found that, on average, Tesla cars had much better electrical fuel range. We also wanted to analyze the model year of EVs that people had purchased and found that models starting from the year 2021 are becoming a more and more popular option. This led us to wonder if EVs had improved over the years and compare the electric ranges of each model year. Surprisingly, the average fuel range of the vehicles in the dataset plummeted after the 2020 models, so we investigated if it had to do with the differences between hybrid (PHEV) and full electric (BEV) vehicles in recent years. Our suspicions were confirmed when we saw the extreme difference between the average electric range of BEVs against the range of PHEVs, which would significantly bring down the average range since averages are not robust against outliers. This conclusion left potential for further exploration of vehicle type preferences over the years since there were more fully electric cars in the dataset. Lastly, we also looked at the geographic spread of where people owned EVs, and it revealed sections concentrated in metropolitan centers, with fewer being in more rural parts of the state. Ultimately, we have better-understood consumer values and trends in EV qualities as the industry grows.