Electric Vehicle Data Analysis

Anoop S Hari

2023-12-04

Link to Dataset

This comprehensive dataset provides detailed information on Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) currently registered through the Washington State Department of Licensing (DOL). It offers a thorough examination of electric vehicle ownership patterns and trends, including vehicle registration, make, model, electric vehicle type, clean alternative fuel vehicle (CAFV) eligibility, electric range, base MSRP, legislative district, DOL vehicle ID, vehicle location, electric utility, and 2020 Census tract.

- **Battery Electric Vehicles (BEVs):** BEVs run entirely on electricity stored in a battery pack. They don't have a traditional internal combustion engine; instead, they use an electric motor to power the vehicle.
- **Plug-in Hybrid Electric Vehicles (PHEVs):** PHEVs have both an electric motor and an internal combustion engine. They use a larger battery pack than conventional hybrids, allowing them to travel a certain distance on electric power alone.
- **Clean alternative fuel vehicle (CAFV):** Clean Alternative Fuel Vehicle (CAFV) eligibility typically refers to the qualifications or criteria that a vehicle must meet in order to be considered a clean alternative fuel vehicle. Eligibility for CAFVs can vary depending on the context, such as government incentives, tax credits, regulatory

definitions, or programs aimed at promoting environmentally friendly transportation.

- **Base MSRP:** Manufacturer's Suggested Retail Price, which is the initial price set by the vehicle manufacturer for a standard or base model of a vehicle. This price typically includes the cost of the vehicle itself with standard equipment and does not include any optional features, taxes, destination charges, registration fees, or other additional costs.

Objectives:

1. CAFV Eligibility and Trends:

- What percentage of the registered vehicles are considered Clean Alternative Fuel Vehicle (CAFV) eligible?
- Are there any noticeable trends or changes in CAFV eligibility among different vehicle types or over various model years?

2. Electric Range Distribution:

- What is the distribution of electric ranges among BEVs and PHEVs in the dataset?
- Are there specific clusters or ranges where most vehicles fall, and how do these ranges impact eligibility and adoption?

3. Base MSRP Analysis:

- What is the range and distribution of Base Manufacturer's Suggested Retail Price (MSRP) across different vehicle types and models?
- Are there correlations between MSRP and CAFV eligibility or electric range?

4. Geographical Insights:

- How are these electric vehicles distributed geographically across counties and cities within Washington State?

5. Utility Provider Preferences:

- Is there a preference for a particular electric utility among the registered electric vehicles?

```
# installing the required packages
install.packages(c("tidyverse","plotly","ggthemes", "bslib"), repos =
"https:\\cran.rstudio.com")
library(tidyverse)
library(plotly)
library(ggthemes)
library(bslib)
library(knitr)
```

```r
# Loading the data
ev <- read.csv("F:\\R practice\\R Projects\\Project 5\\EV
Project\\Electric_Vehicle_Population_Data.csv")

# checking head and str
str(ev)
```

```
## 'data.frame':    159467 obs. of  17 variables:
##  $ VIN..1.10.                                       : chr  "2C4RC1N71H"
"2C4RC1N7XL" "KNDC3DLCXN" "5YJ3E1EA0J" ...
##  $ County                                           : chr  "Kitsap"
"Stevens" "Yakima" "Kitsap" ...
##  $ City                                             : chr  "Bremerton"
"Colville" "Yakima" "Bainbridge Island" ...
##  $ State                                            : chr  "WA" "WA" "WA"
"WA" ...
##  $ Postal.Code                                      : int  98311 99114
98908 98110 98501 98367 98902 98901 98359 98370 ...
##  $ Model.Year                                       : int  2017 2020 2022
2018 2018 2019 2019 2022 2012 2021 ...
##  $ Make                                             : chr  "CHRYSLER"
"CHRYSLER" "KIA" "TESLA" ...
##  $ Model                                            : chr  "PACIFICA"
"PACIFICA" "EV6" "MODEL 3" ...
##  $ Electric.Vehicle.Type                            : chr  "Plug-in Hybrid
Electric Vehicle (PHEV)" "Plug-in Hybrid Electric Vehicle (PHEV)" "Battery
Electric Vehicle (BEV)" "Battery Electric Vehicle (BEV)" ...
##  $ Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility: chr  "Clean
Alternative Fuel Vehicle Eligible" "Clean Alternative Fuel Vehicle Eligible"
"Eligibility unknown as battery range has not been researched" "Clean
Alternative Fuel Vehicle Eligible" ...
##  $ Electric.Range                                   : int  33 32 0 215 151
239 12 0 6 0 ...
##  $ Base.MSRP                                        : int  0 0 0 0 0 0
36900 0 0 0 ...
##  $ Legislative.District                             : int  23 7 14 23 35
26 14 15 26 23 ...
##  $ DOL.Vehicle.ID                                   : int  349437882
154690532 219969144 476786887 201185253 478017067 146830148 207786505
284893416 211699309 ...
##  $ Vehicle.Location                                 : chr  "POINT (-
122.6466274 47.6341188)" "POINT (-117.90431 48.547075)" "POINT (-120.6027202
46.5965625)" "POINT (-122.5235781 47.6293323)" ...
##  $ Electric.Utility                                 : chr  "PUGET SOUND
ENERGY INC" "AVISTA CORP" "PACIFICORP" "PUGET SOUND ENERGY INC" ...
##  $ X2020.Census.Tract                               : num  5.30e+10
5.31e+10 5.31e+10 5.30e+10 5.31e+10 ...
```

```r
head(ev)
```

```
##   VIN..1.10.   County              City State Postal.Code Model.Year
Make
## 1 2C4RC1N71H   Kitsap         Bremerton    WA       98311       2017
CHRYSLER
## 2 2C4RC1N7XL  Stevens          Colville    WA       99114       2020
CHRYSLER
## 3 KNDC3DLCXN   Yakima            Yakima    WA       98908       2022
KIA
## 4 5YJ3E1EA0J   Kitsap Bainbridge Island    WA       98110       2018
TESLA
## 5 1N4AZ1CP7J Thurston          Tumwater    WA       98501       2018
NISSAN
## 6 KNDCC3LG6K   Kitsap      Port Orchard    WA       98367       2019
KIA
##      Model                   Electric.Vehicle.Type
## 1 PACIFICA Plug-in Hybrid Electric Vehicle (PHEV)
## 2 PACIFICA Plug-in Hybrid Electric Vehicle (PHEV)
## 3      EV6         Battery Electric Vehicle (BEV)
## 4  MODEL 3         Battery Electric Vehicle (BEV)
## 5     LEAF         Battery Electric Vehicle (BEV)
## 6     NIRO         Battery Electric Vehicle (BEV)
##              Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility
Electric.Range
## 1                     Clean Alternative Fuel Vehicle Eligible
33
## 2                     Clean Alternative Fuel Vehicle Eligible
32
## 3 Eligibility unknown as battery range has not been researched
0
## 4                     Clean Alternative Fuel Vehicle Eligible
215
## 5                     Clean Alternative Fuel Vehicle Eligible
151
## 6                     Clean Alternative Fuel Vehicle Eligible
239
##   Base.MSRP Legislative.District DOL.Vehicle.ID
Vehicle.Location
## 1         0                  23      349437882 POINT (-122.6466274
47.6341188)
## 2         0                   7      154690532    POINT (-117.90431
48.547075)
## 3         0                  14      219969144 POINT (-120.6027202
46.5965625)
## 4         0                  23      476786887 POINT (-122.5235781
47.6293323)
## 5         0                  35      201185253    POINT (-122.89692
47.043535)
## 6         0                  26      478017067   POINT (-122.6847073
47.50524)
##         Electric.Utility X2020.Census.Tract
```

```
## 1 PUGET SOUND ENERGY INC        53035091800
## 2            AVISTA CORP        53065950500
## 3             PACIFICORP        53077000904
## 4 PUGET SOUND ENERGY INC        53035091001
## 5 PUGET SOUND ENERGY INC        53067011720
## 6 PUGET SOUND ENERGY INC        53035092902
```

```
# Checking for null values
any(is.na(ev))
```

```
## [1] TRUE
```

```
# Count of NA's with column names
null <- colSums(is.na(ev))

null[null > 0]
```

```
##          Postal.Code Legislative.District   X2020.Census.Tract
##                    4                  361                     4
```

```
# imputing NA's with 0
ev$Postal.Code[is.na(ev$Postal.Code)] <- 0
ev$Legislative.District[is.na(ev$Legislative.District)] <- 0
ev$X2020.Census.Tract[is.na(ev$X2020.Census.Tract)] <- 0
```

Analysis:

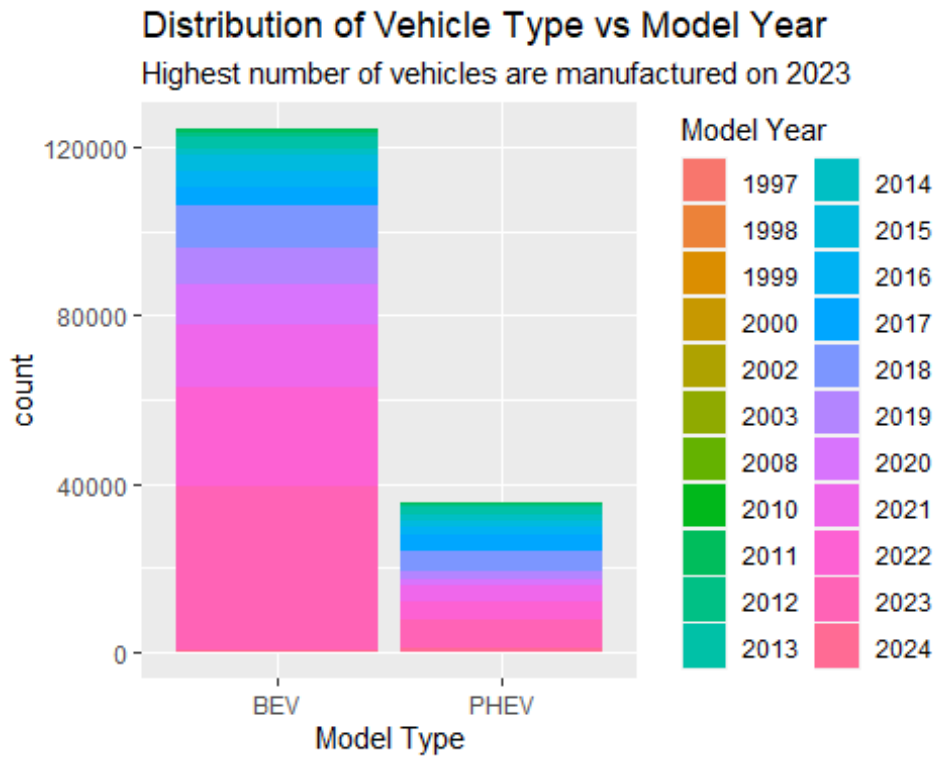
1. CAFV Eligibility and Trends:
   - What percentage of the registered vehicles are considered Clean Alternative Fuel Vehicle (CAFV) eligible?
   - Are there any noticeable trends or changes in CAFV eligibility among different vehicle types or over various model years?

```
# For better readability the values of 'Electric.Vehicle.Type' column is
changed to
# 'BEV' for Battery Electric Vehicle
# & PHEV for Plug-in Hybrid Electric Vehicles
ev_clean <- ev %>%
  mutate(Electric.Vehicle.Type = ifelse(Electric.Vehicle.Type == "Battery
Electric Vehicle (BEV)", "BEV", "PHEV"))

# Changing column name Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility to
CAFV_Eligibility
ev_clean <- ev_clean %>%
  mutate(CAFV_Eligibility =
Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility)

ev_clean <- ev_clean %>%
  select(- Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility)

# Clean Alternative Fuel Vehicle percentage
CAFV_perc <- ev_clean %>%
```

```
  group_by(CAFV_Eligibility) %>%
  summarise(count = n()) %>%
  mutate(Perc = (count / sum(count))* 100)
```

Looking for trends in Vehicle Type and Model Year

```
# Columns to look for:
        # CAFV eligibility
        # Electric.Vehicle.Type
        # Model.Year

ggplot(ev_clean, aes(x = Electric.Vehicle.Type)) +
  geom_bar(aes(fill = factor(Model.Year))) +
  labs(title = "Distribution of Vehicle Type vs Model Year" ,
       x = "Model Type",
       subtitle = "Highest number of vehicles are manufactured on 2023",
       fill = "Model Year")
```

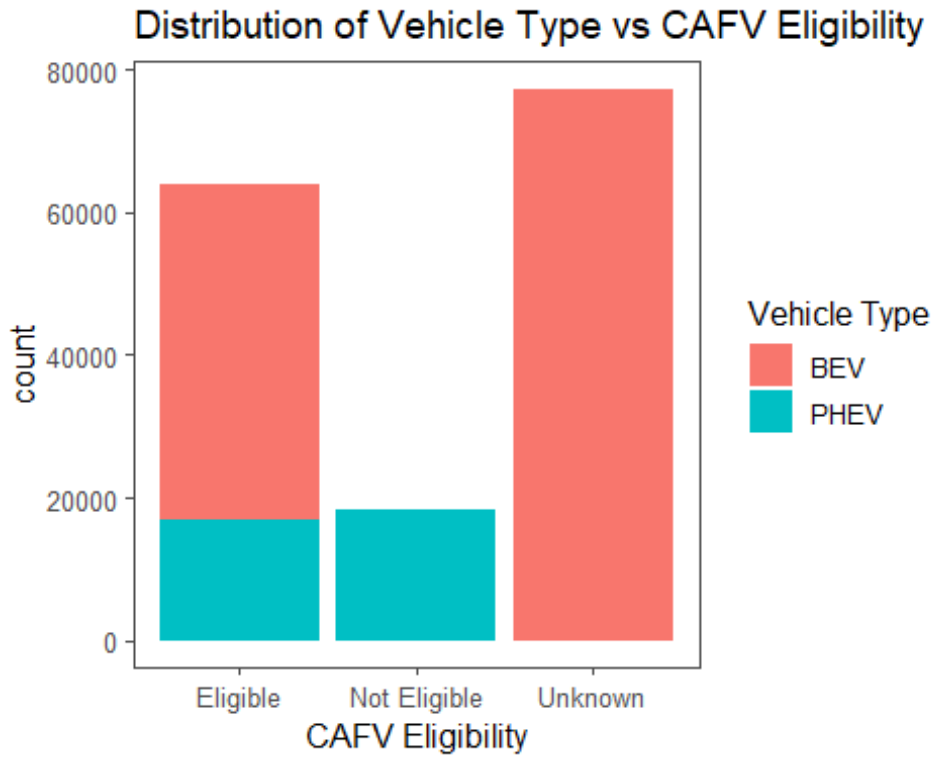


The analyzed data underscores a prevailing dominance of Battery Electric Vehicles (BEVs) throughout 2022 to 2024, totaling 66,032 units, with a consistent increase in BEV production. Intriguingly, 2024 presents a notable surge in planned manufacturing of Plug-in Hybrid Electric Vehicles (PHEVs) over BEVs, signaling a potential market shift. The peak production year for both BEVs and PHEVs aligns in 2023, reflecting the zenith in vehicle output for both categories. These insights illuminate an evolving landscape favoring all-electric models, accompanied by a marked rise in interest or planned production of plug-in hybrid variants in the most recent year.

```r
# Looking for distinct values in CAFV_Eligibility column
table(ev_clean$CAFV_Eligibility)
```

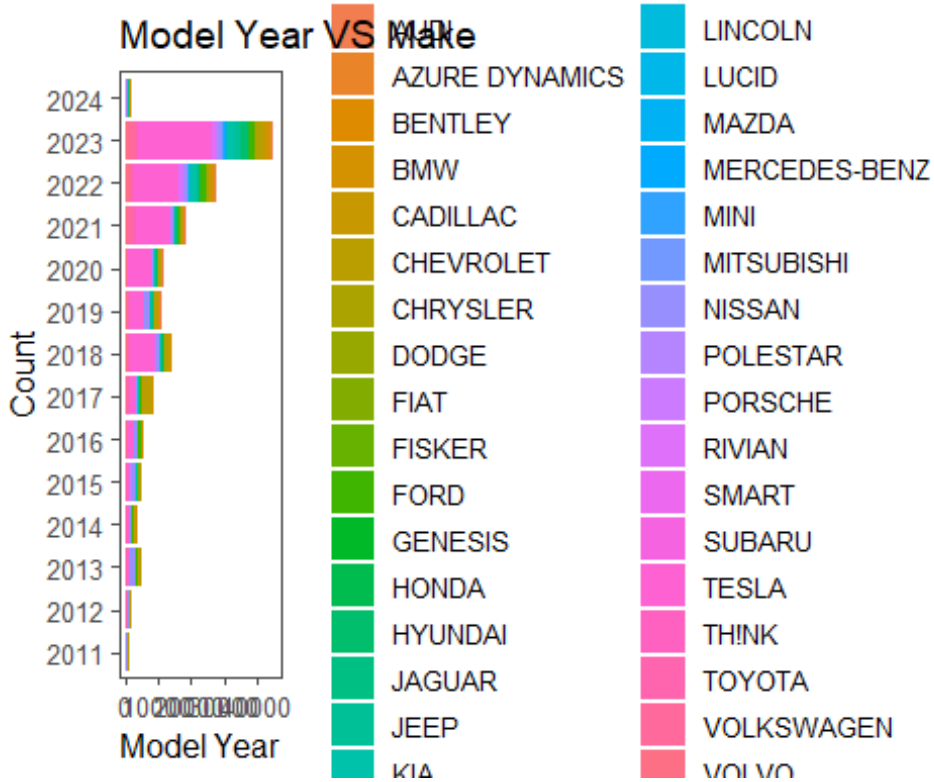
```
##
##                      Clean Alternative Fuel Vehicle Eligible
##                                                       63824
## Eligibility unknown as battery range has not been researched
##                                                       77195
##                        Not eligible due to low battery range
##                                                       18448
```

```r
# For better readability changing the values of CAFV_Eligibility to
# Eligible, Not Eligible and Unknown
ev_clean$CAFV_Eligibility[ev_clean$CAFV_Eligibility == "Clean Alternative
Fuel Vehicle Eligible"] <- "Eligible"
ev_clean$CAFV_Eligibility[ev_clean$CAFV_Eligibility == "Eligibility unknown
as battery range has not been researched"] <- "Unknown"
ev_clean$CAFV_Eligibility[ev_clean$CAFV_Eligibility == "Not eligible due to
low battery range"] <- "Not Eligible"

#setting a common theme for all ggplot visuals
theme_set(theme_few())

# Checking the Distribution of Vehicle Type with CAFV Eligibility
ggplot(ev_clean, aes(x = CAFV_Eligibility)) +
  geom_bar(aes(fill = Electric.Vehicle.Type)) +
  labs(title = "Distribution of Vehicle Type vs CAFV Eligibility",
       x = "CAFV Eligibility", fill = "Vehicle Type") +
  theme_few()
```



'Unknown' is the predominant eligibility type for most vehicles, with Battery Electric Vehicles (BEV) leading in manufacturing.

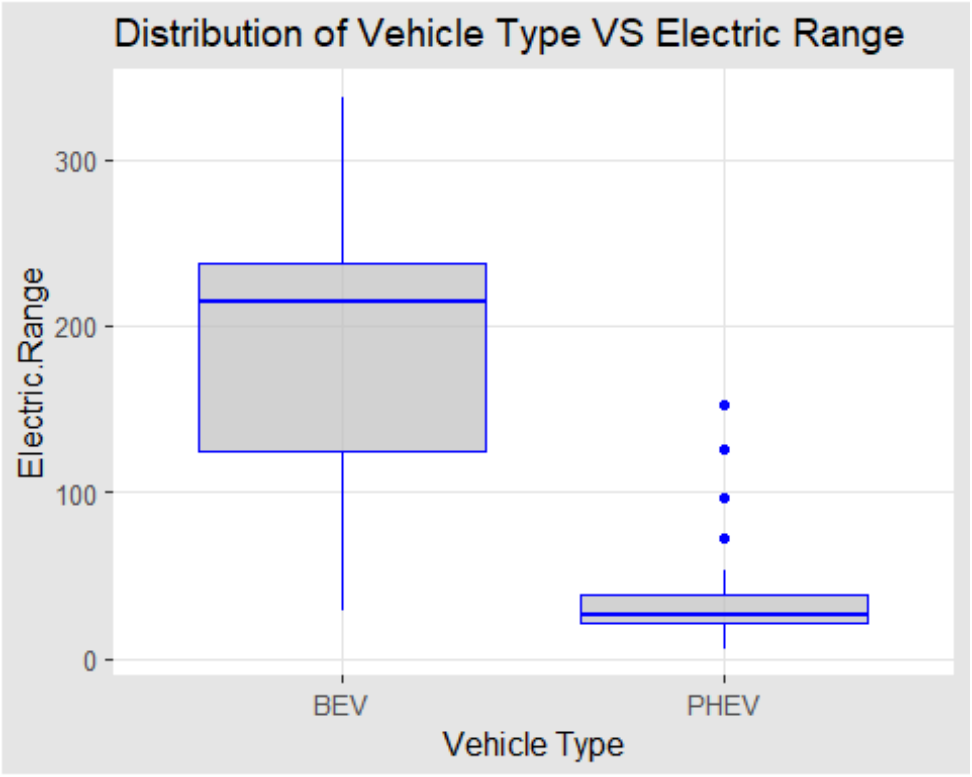
```
# Subsetting 'Model Year' is necessary due to lower data counts before 2011
year <- c(2011:2024)

ev_year_adjusted<- ev_clean %>%
  subset(Model.Year %in% year)

# Finding out distribution of 'Make' with 'Model Year'
ggplot(ev_year_adjusted, aes(x = factor(Model.Year))) +
  geom_bar(aes(fill = Make)) +
  theme_few() +
  coord_flip() +
  labs(title = "Model Year VS Make",
       x = "Count",
       y = "Model Year")
```




```
'Tesla' stands as the most frequent 'Make' among registered vehicles.
```

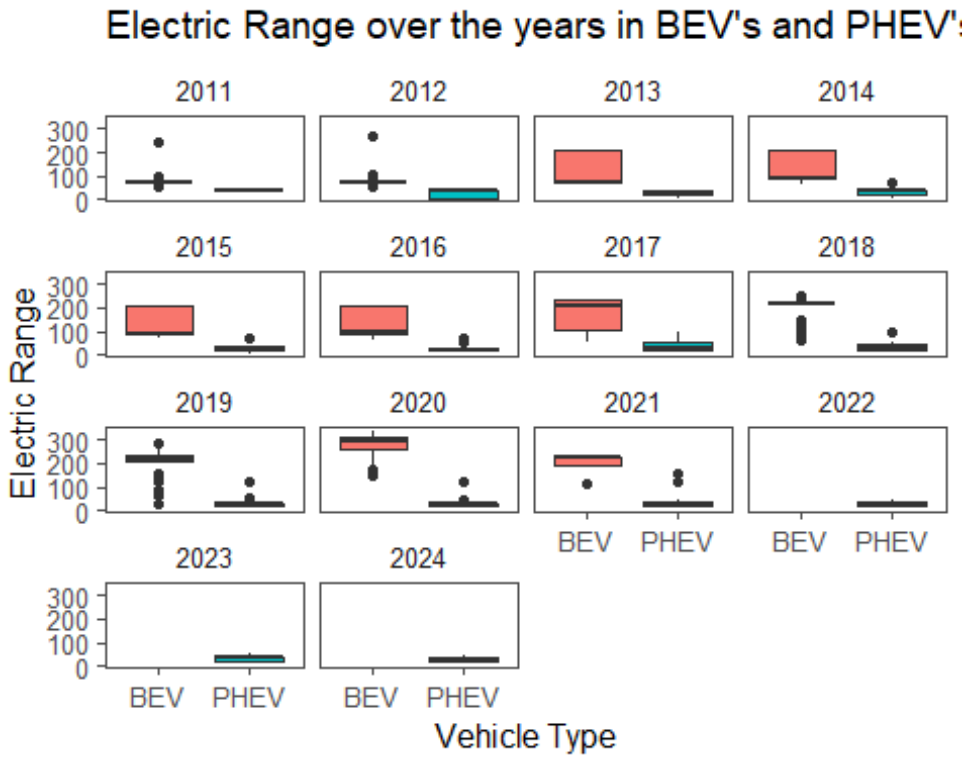
2. Electric Range Distribution:
   - What is the distribution of electric ranges among BEVs and PHEVs in the dataset?
   - Are there specific clusters or ranges where most vehicles fall, and how do these ranges impact eligibility and adoption?

```
# Filtering out PHEVs with an electric range of 0, as they operate on both
electric charge and gasoline.
ev_range <- ev_clean %>%
  filter(Electric.Range > 0) %>%
  subset(Model.Year %in% year)

# boxplot to understand the distribution
ggplot(ev_range, aes(x = Electric.Vehicle.Type, y = Electric.Range)) +
  geom_boxplot(color = "blue", fill = "grey",alpha = 0.7) +
  labs(title = "Distribution of Vehicle Type VS Electric Range",
       x = "Vehicle Type") + theme_igray()
```



For BEVs, half of the electric range falls between 125 to 238 miles, whereas for PHEVs, 50% of the range lies
between 21 to 38 miles.

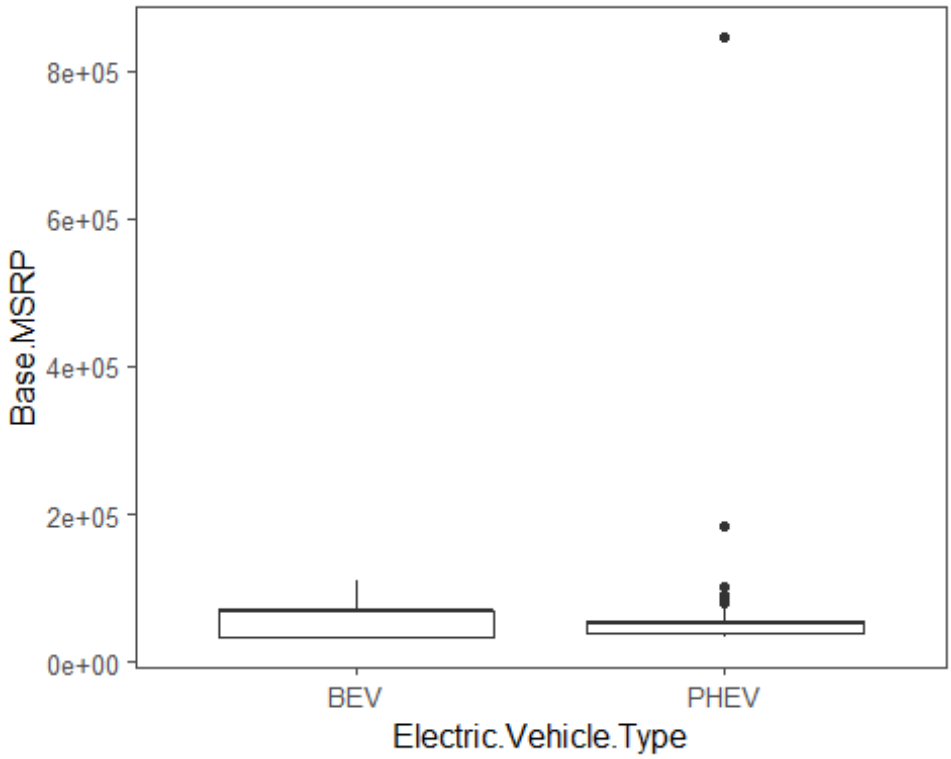
```
# Electric range variation across different model years
ggplot(ev_range, aes(x = Electric.Vehicle.Type, y = Electric.Range)) +
  geom_boxplot(aes(fill = Electric.Vehicle.Type)) +
  facet_wrap(~Model.Year) +
  guides(fill = FALSE) +
  labs(title = "Electric Range over the years in BEV's and PHEV's",
       x = "Vehicle Type",
       y = "Electric Range")
```




```
Electric range for BEVs shows a consistent rise over the years, but this
upward trend diminishes after 2020.
Conversely, electric range for PHEVs doesn't exhibit significant growth
across the years.
```

3. Base MSRP Analysis:
   - What is the range and distribution of Base Manufacturer's Suggested Retail Price (MSRP) across different vehicle types and models?
   - Are there correlations between MSRP and CAFV eligibility or electric range?

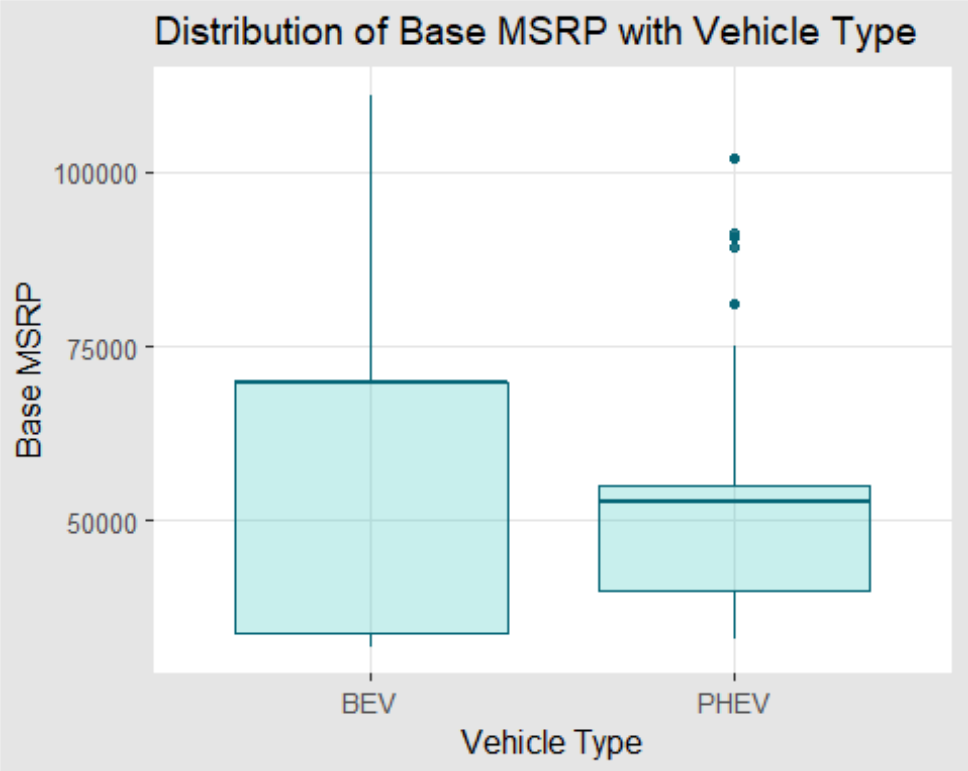
```r
# I opted to filter out the numerous zero values to obtain a clearer
depiction.
msrp <- ev_clean %>%
  filter(Base.MSRP > 0)

ggplot(msrp, aes(x = Electric.Vehicle.Type, y = Base.MSRP )) +
 geom_boxplot() # noticed an outlier
```




```
# finding outlier
sorted_msrp <- msrp %>%
  arrange(-Base.MSRP)
```

Typically, the Base MSRP of PHEV vehicles is lower than that of BEVs. As the maximum Base MSRP of a BEV in this
dataset stands at 110950, I chose to filter out PHEV values exceeding this threshold for a refined analysis.

```
msrp_at <- msrp %>%
  filter(!(Base.MSRP > 110950))

# Distribution of Base MSRP with Vehicle Type
ggplot(msrp_at, aes(x = Electric.Vehicle.Type,
                    y = Base.MSRP )) +
  geom_boxplot(color = "#046576",
               fill = "#45C8C4",
               alpha = 0.3) +
              theme_igray() +
              labs(x = "Vehicle Type",
              y = "Base MSRP",
              title = "Distribution of Base MSRP with Vehicle Type")
```



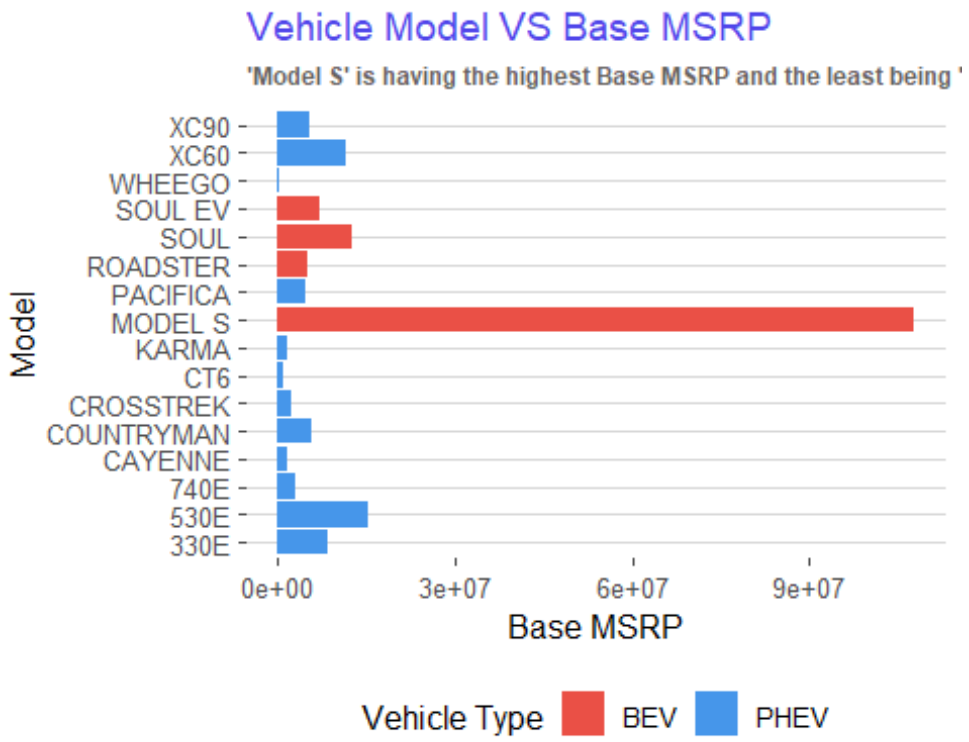
Half of the BEVs exhibit a base MSRP falling between 33,950 and 69,900, while 50% of PHEVs showcase a base MSRP ranging from 39,995 to 54,950.

From this information, it can be inferred that, on average:

- BEVs tend to have a broader range in terms of base MSRP, spanning from 33,950 to 69,900, indicating a wider spectrum of pricing among these vehicles.
- Conversely, PHEVs demonstrate a narrower pricing range, with 50% of the vehicles falling within the range of 39,995 to 54,950. This suggests a relatively more clustered pricing distribution among PHEVs compared to BEVs.

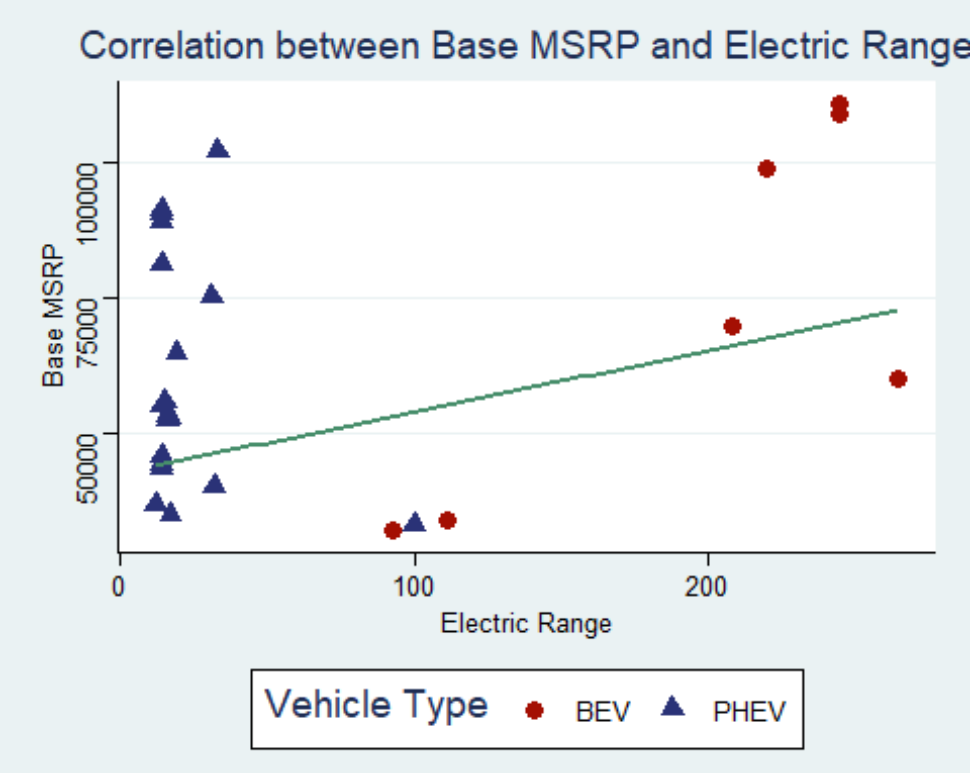
Overall, this analysis provides insight into the diverse pricing structures between BEVs and PHEVs, showcasing different concentration patterns within their respective price ranges.

```
# Vehicle Model VS Base MSRP
ggplot(msrp_at, aes(x = Base.MSRP, y = Model)) +
  geom_col(aes(fill = Electric.Vehicle.Type)) +
  labs(title = "Vehicle Model VS Base MSRP",
       x = "Base MSRP",
       fill = "Vehicle Type",
       subtitle =  expression(bold("'Model S' is having the highest Base MSRP
and the least being 'Wheego'")))+
  scale_fill_manual(values = c("#EA5046", "#4696EA")) +
  theme_hc() +
  theme(plot.subtitle = element_text(color = "#666260",size = 9),
        plot.title = element_text(color = "#5046EA"))
```



Base MSRP of 'Model S' is the highest and 'Weego' the lowest.

```r
# finding correlation between MSRP and electric range
correlation1 <- cor(msrp_at$Base.MSRP, msrp_at$Electric.Range)

# Plotting Correlation between Base MSRP and Electric Range
ggplot(msrp_at, aes(x = Electric.Range,
                    y = Base.MSRP))+
  geom_point( aes(color = Electric.Vehicle.Type,
                  shape = Electric.Vehicle.Type),
                  size = 3) +
  geom_smooth(method = lm, se = F, color = "#478E6B") +
  scale_color_manual(values = c("#A40F03", "#2A3277"),
                     name = "Vehicle Type")+
  labs(title = "Correlation between Base MSRP and Electric Range",
       x = "Electric Range", y = "Base MSRP",
       shape = "Vehicle Type") +
  theme_stata()
```




```
There is a moderate correlation (0.5982857) observed between Base MSRP and
Electric Range, signifying a notable
association between these two attributes.
```

4. Geographical Insights:
   - How are these electric vehicles distributed geographically across counties and cities within Washington State?

```
To simplify chart labels, I opted to filter out state and county names by
focusing on the top 20 entries.
Given that the majority of sales occur in WA state, I won't be examining
state-specific trends.
```

```
county_filt <- ev_clean %>%
  group_by(County) %>%
  summarise(Vehicle_Count = n())

print(county_filt)
```

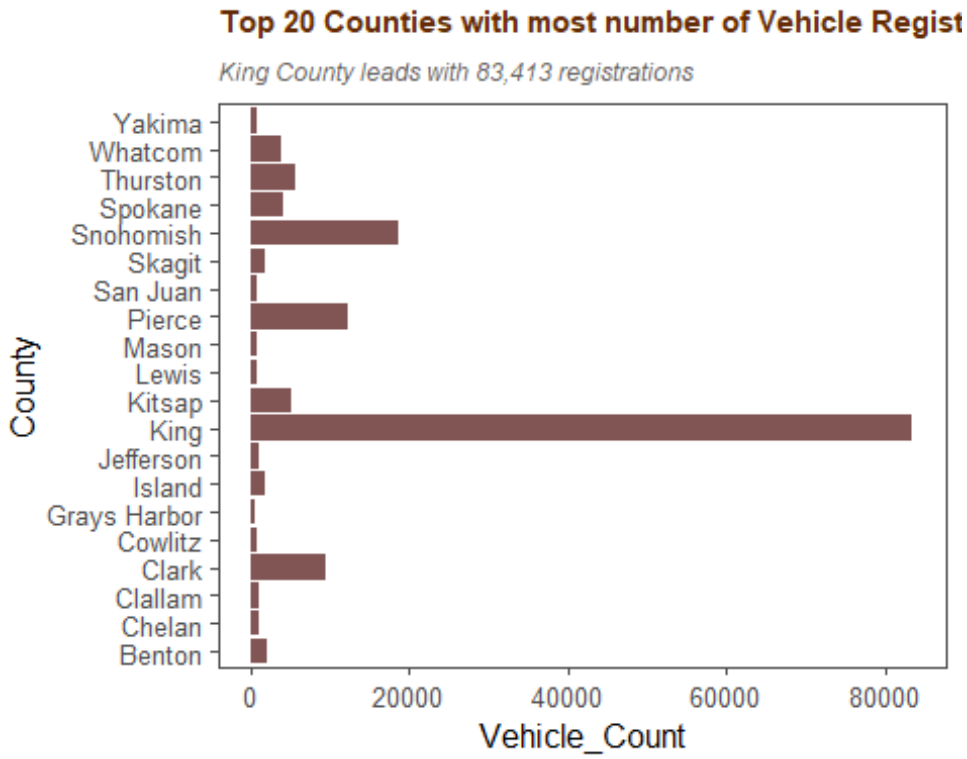
```
## # A tibble: 185 × 2
##    County         Vehicle_Count
##    <chr>                  <int>
##  1 ""                         4
##  2 "Adams"                   46
##  3 "Alameda"                  4
##  4 "Albemarle"                1
##  5 "Alexandria"               3
##  6 "Allen"                    1
```

```
##  7 "Anchorage"                1
##  8 "Anne Arundel"             9
##  9 "Arapahoe"                 2
## 10 "Asotin"                  66
## # ℹ 175 more rows

state_filt <- ev_clean %>%
  group_by(State) %>%
  summarise(Vehicle_Count = n())

print(state_filt)

## # A tibble: 45 × 2
##    State Vehicle_Count
##    <chr>         <int>
##  1 AE                1
##  2 AK                1
##  3 AL                3
##  4 AP                1
##  5 AR                2
##  6 AZ                9
##  7 BC                2
##  8 CA               95
##  9 CO               13
## 10 CT                7
## # ℹ 35 more rows

# Finding the top 20 Counties
county_filt_desc <- county_filt %>%
  arrange(desc(Vehicle_Count))

top_20_county <- head(county_filt_desc,20)


kable(top_20_county)
```

County	Vehicle_Count
King	83413
Snohomish	18544
Pierce	12315
Clark	9370
Thurston	5711
Kitsap	5216
Spokane	4016
Whatcom	3865
Benton	1942

County	Vehicle_Count
Skagit	1759
Island	1721
Clallam	965
Chelan	926
Jefferson	907
Yakima	882
San Juan	875
Cowlitz	791
Mason	742
Lewis	652
Grays Harbor	564

```
# Top 20 Counties with most number of Vehicle Registrations
ggplot(top_20_county, aes(x = County,
                          y = Vehicle_Count)) +
  geom_bar(stat = "identity",
           fill = "#825555") +
  coord_flip()+
  labs(title = expression(bold("Top 20 Counties with most number of Vehicle
Registrations")),
       subtitle = expression(italic("King County leads with 83,413
registrations"))) +
  theme(plot.subtitle = element_text(color = "#666260",size = 9),
        plot.title = element_text(color = "#632B02", size = 11))
```




```
King County leads with the highest registrations, totaling 83,413, followed
by Snohomish with 18,544
registrations.
```

5. Utility Provider Preferences:
   - Is there a preference for a particular electric utility among the registered electric vehicles?

```
eu_count <- ev_clean %>%
  group_by(Electric.Utility) %>%
  summarise(Count = n())

eu <- eu_count %>%
  arrange(-Count)

# Top 3 Electric Utilities
top_3_eu <- head(eu,3)

kable(top_3_eu)
```

Electric.Utility	Count
PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)	58884
PUGET SOUND ENERGY INC	31869
CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)	28634

Conclusion

The analysis encompassed various aspects of the electric vehicle dataset, providing several key insights:

- **Manufacturing Trends:** Over the observed years (2022-2024), Battery Electric Vehicles (BEVs) significantly outnumber Plug-in Hybrid Electric Vehicles (PHEVs), with a consistent increase in BEV production. Conversely, PHEV production doesn't exhibit a steady rise, remaining relatively stable.
- **Vehicle Eligibility:** The majority of vehicles have an 'Unknown' Clean Alternative Fuel Vehicle (CAFV) eligibility type, while BEVs dominate as the most manufactured vehicle type.
- **Popular Make:** Tesla emerges as the most registered vehicle make, indicative of its widespread presence in the dataset.
- **Electric Range:** BEVs tend to have a wider range with an average of (215-302 miles) compared to PHEVs (21-38 miles), with a steady increase in BEV range over the years until a decline post-2020. PHEVs, however, show limited range variation over time.
- **Base MSRP:** Half of the BEVs fall within a price range of $33,950 to $69,900, while 50% of PHEVs range between $39,995 and $54,950. 'Model S' boasts the highest Base MSRP, while 'Weego' records the lowest.
- **Correlation:** A moderate correlation (0.5982857) exists between Base MSRP and Electric Range, suggesting a noticeable association between these variables.
- **Registration Distribution:** King County holds the highest number of registrations (83,413), followed by Snohomish with 18,544 registrations.

Considering these findings collectively, the dataset reveals a dominance of BEVs in both production and electric range expansion, while PHEVs maintain a more stable presence. Tesla stands out as a popular make among registered vehicles. The moderate correlation between Base MSRP and Electric Range highlights a notable relationship between pricing and vehicle range. Moreover, geographical distribution is concentrated in King County, emphasizing localized registration patterns. These findings collectively illustrate the dynamic landscape of electric vehicles, portraying shifting trends in production, pricing, and range among different vehicle types over the observed period.