Untitled

2025-04-05

## Fundamentos de Estadística para Analítica de datos

Integrantes

* Jimmy José Díaz Leyton
* Jessica Hasbleidy Pinilla Segovia
* Yenny Marisol Sánchez Sánchez

El uso de vehículos eléctricos ha experimentado un crecimiento significativo en los últimos años debido al interés en reducir la dependencia de los combustibles fósiles. Por ello, se ha tomado un conjunto de datos de vehículos eléctricos registrados en el estado de Washington con el propósito de examinar distintos aspectos relacionados con su comercialización y desempeño.

Se analizarán variables como el tipo de vehículo eléctrico (BEV - totalmente eléctrico y PHEV - híbrido enchufable), el precio base (MSRP) y el alcance eléctrico (Electric Range), con el fin de identificar patrones y diferencias entre estos modelos.

Además, se explorará la evolución de los precios a lo largo del tiempo, la relación entre la autonomía y el tipo de vehículo, así como la influencia de factores geográficos en la elegibilidad para incentivos de energía limpia. A través de este análisis, se espera obtener información valiosa sobre la adopción y el desarrollo de los vehículos eléctricos en Washington, contribuyendo a una mejor comprensión de su impacto en el mercado y en la infraestructura de transporte sostenible.

### **Variables cualitativas:**

1. **City** (Ciudad) - **Nominal**
2. **State** (Estado) - **Nominal**: “Washington”.
3. **Make** (Marca) - **Nominal**: Diferentes marcas de vehículos, como Tesla, Nissan, etc.
4. **Model** (Modelo) - **Nominal**: Diferentes modelos de vehículos.
5. **Electric Vehicle Type** (Tipo de vehículo eléctrico) - **Nominal**: Puede ser BEV (Battery Electric Vehicle) o PHEV (Plug-in Hybrid Electric Vehicle).
6. **Clean Alternative Fuel Vehicle (CAFV) Eligibility** (Preferencia para vehículo de combustible alternativo limpio) - **Nominal**
7. **Vehicle Location** (Ubicación del vehículo) - **Nominal**
8. **Electric Utility** (Utilidad eléctrica) - **Nominal**: Es el nombre de la empresa de servicios eléctricos que abastece el área.
9. **Legislative District** (Distrito legislativo) - **Nominal**: Representa los diferentes distritos legislativos en el estado de Washington.

### **Variables cuantitativas:**

1. **Postal Code** (Código Postal) - **Discreta**: Es un número que representa un código de área específico, pero es una variable discreta porque los códigos postales están en un conjunto limitado de valores.
2. **Model Year** (Año del modelo) - **Discreta**: Es un número entero que indica el año en que se fabricó el vehículo. Aunque es numérica, es discreta porque los valores son enteros y finitos.
3. **Electric Range** (Alcance eléctrico) - **Continua**: Es la distancia que puede recorrer un vehículo con una sola carga. Como puede tomar cualquier valor real dentro de un rango, es una variable continua.
4. **Base MSRP** (Precio base MSRP) - **Continua**: Es el precio base del vehículo, que puede tomar cualquier valor dentro de un rango, lo que lo convierte en una variable continua.
5. **2020 Census Tract** (Tracto censal de 2020) - **Nominal o discreta**: Aunque parece ser un identificador geográfico, en algunos casos puede ser tratado como una variable nominal (es un código de área), pero dependiendo del análisis, puede también considerarse discreta si se le usa para representar zonas geográficas.

### **Pregunta global:**

**¿Cuál es la relación entre el tipo de vehículo eléctrico (BEV vs PHEV), el precio base (MSRP) y el alcance eléctrico (Electric Range) de los vehículos registrados en el Estado de Washington?**

Esta pregunta abarca una perspectiva amplia que examina las diferencias entre los tipos de vehículos eléctricos (BEV y PHEV) en relación con dos variables cuantitativas clave: el precio y el alcance eléctrico. Esto te permitirá ver si existen patrones o diferencias significativas entre estos dos tipos de vehículos en cuanto a su precio y rendimiento.

### **Preguntas adicionales para apoyar la respuesta:**

1. **¿Cómo varía el precio base (MSRP) de los vehículos eléctricos en función del año del modelo?**

* Esta pregunta te permitirá analizar cómo ha cambiado el precio a lo largo de los años. Esto es útil para entender si los vehículos más nuevos tienden a ser más caros y si esto está relacionado con el avance de la tecnología.

1. **¿Existen diferencias en el alcance eléctrico (Electric Range) entre los vehículos BEV y los PHEV?**  
     
   Este análisis podría ayudar a determinar si los vehículos completamente eléctricos (BEV) tienen un mayor alcance en comparación con los híbridos enchufables (PHEV) o si no hay diferencias significativas.
2. **¿Cómo varía la elegibilidad para los vehículos de combustible alternativo limpio (CAFV) en función del tipo de vehículo y su ubicación geográfica (estado o ciudad)?**

* Esta pregunta te permitiría explorar si hay diferencias en la elegibilidad para incentivos o beneficios dependiendo del tipo de vehículo (BEV o PHEV) y cómo esto se relaciona con la ciudad o región en la que están registrados los vehículos.

1. **¿Qué relación existe entre el tipo de vehículo (BEV o PHEV) y la utilidad eléctrica que abastece la región en la que está registrado el vehículo?**  
     
   Esta pregunta busca explorar si existe alguna correlación entre el tipo de vehículo eléctrico y el proveedor de electricidad que tiene la región, lo que podría influir en la accesibilidad y la infraestructura de carga disponible.

### Cargar librerías necesarias

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.4.3

## Warning: package 'lubridate' was built under R version 4.4.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(dplyr)  
# Instalar TinyTeX  
#install.packages("tinytex")  
#tinytex::install\_tinytex()

### Cargar datos (ajusta el nombre del archivo)

datos<- read.csv("C:/Users/OptiPlex/OneDrive/Documentos/Universidad\_Central/Opcion Estadistica/Proyecto/Electric\_Vehicle\_Population\_Data.csv")

### Ver las primeras filas

head(datos)

## VIN..1.10. County City State Postal.Code Model.Year Make Model  
## 1 5YJ3E1EBXK King Seattle WA 98178 2019 TESLA MODEL 3  
## 2 5YJYGDEE3L Kitsap Poulsbo WA 98370 2020 TESLA MODEL Y  
## 3 KM8KRDAF5P Kitsap Olalla WA 98359 2023 HYUNDAI IONIQ 5  
## 4 5UXTA6C0XM Kitsap Seabeck WA 98380 2021 BMW X5  
## 5 JTMAB3FV7P Thurston Rainier WA 98576 2023 TOYOTA RAV4 PRIME  
## 6 5YJSA1DN0C Thurston Olympia WA 98502 2012 TESLA MODEL S  
## Electric.Vehicle.Type  
## 1 Battery Electric Vehicle (BEV)  
## 2 Battery Electric Vehicle (BEV)  
## 3 Battery Electric Vehicle (BEV)  
## 4 Plug-in Hybrid Electric Vehicle (PHEV)  
## 5 Plug-in Hybrid Electric Vehicle (PHEV)  
## 6 Battery Electric Vehicle (BEV)  
## Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility Electric.Range  
## 1 Clean Alternative Fuel Vehicle Eligible 220  
## 2 Clean Alternative Fuel Vehicle Eligible 291  
## 3 Eligibility unknown as battery range has not been researched 0  
## 4 Clean Alternative Fuel Vehicle Eligible 30  
## 5 Clean Alternative Fuel Vehicle Eligible 42  
## 6 Clean Alternative Fuel Vehicle Eligible 265  
## Base.MSRP Legislative.District DOL.Vehicle.ID Vehicle.Location  
## 1 0 37 477309682 POINT (-122.23825 47.49461)  
## 2 0 23 109705683 POINT (-122.64681 47.73689)  
## 3 0 26 230390492 POINT (-122.54729 47.42602)  
## 4 0 35 267929112 POINT (-122.81585 47.64509)  
## 5 0 2 236505139 POINT (-122.68993 46.88897)  
## 6 59900 22 186637195 POINT (-122.92333 47.03779)  
## Electric.Utility X2020.Census.Tract  
## 1 CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA) 53033011902  
## 2 PUGET SOUND ENERGY INC 53035091100  
## 3 PUGET SOUND ENERGY INC 53035092802  
## 4 PUGET SOUND ENERGY INC 53035091301  
## 5 PUGET SOUND ENERGY INC 53067012530  
## 6 PUGET SOUND ENERGY INC 53067010600

### Ver estructura de los datos

str(datos)

## 'data.frame': 235692 obs. of 17 variables:  
## $ VIN..1.10. : chr "5YJ3E1EBXK" "5YJYGDEE3L" "KM8KRDAF5P" "5UXTA6C0XM" ...  
## $ County : chr "King" "Kitsap" "Kitsap" "Kitsap" ...  
## $ City : chr "Seattle" "Poulsbo" "Olalla" "Seabeck" ...  
## $ State : chr "WA" "WA" "WA" "WA" ...  
## $ Postal.Code : int 98178 98370 98359 98380 98576 98502 98004 98271 98034 98052 ...  
## $ Model.Year : int 2019 2020 2023 2021 2023 2012 2017 2022 2018 2018 ...  
## $ Make : chr "TESLA" "TESLA" "HYUNDAI" "BMW" ...  
## $ Model : chr "MODEL 3" "MODEL Y" "IONIQ 5" "X5" ...  
## $ Electric.Vehicle.Type : chr "Battery Electric Vehicle (BEV)" "Battery Electric Vehicle (BEV)" "Battery Electric Vehicle (BEV)" "Plug-in Hybrid Electric Vehicle (PHEV)" ...  
## $ 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 220 291 0 30 42 265 81 22 215 215 ...  
## $ Base.MSRP : int 0 0 0 0 0 59900 0 0 0 0 ...  
## $ Legislative.District : int 37 23 26 35 2 22 48 39 45 45 ...  
## $ DOL.Vehicle.ID : int 477309682 109705683 230390492 267929112 236505139 186637195 196789610 204822761 2039222 474817283 ...  
## $ Vehicle.Location : chr "POINT (-122.23825 47.49461)" "POINT (-122.64681 47.73689)" "POINT (-122.54729 47.42602)" "POINT (-122.81585 47.64509)" ...  
## $ Electric.Utility : chr "CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)" "PUGET SOUND ENERGY INC" "PUGET SOUND ENERGY INC" "PUGET SOUND ENERGY INC" ...  
## $ X2020.Census.Tract : num 5.30e+10 5.30e+10 5.30e+10 5.30e+10 5.31e+10 ...

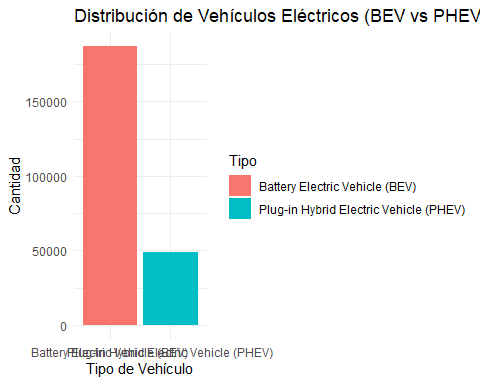
### Pregunta 1:

Filtrar datos para BEV y PHEV

vehiculos\_filtrados <- datos %>%  
 filter(Electric.Vehicle.Type %in% c("Battery Electric Vehicle (BEV)", "Plug-in Hybrid Electric Vehicle (PHEV)"))

### Crear el gráfico

#windows()  
ggplot(vehiculos\_filtrados, aes(x = Electric.Vehicle.Type, fill = Electric.Vehicle.Type)) +  
 geom\_bar() +  
 labs(title = "Distribución de Vehículos Eléctricos (BEV vs PHEV)",  
 x = "Tipo de Vehículo",y = "Cantidad",fill = "Tipo") + theme\_minimal()

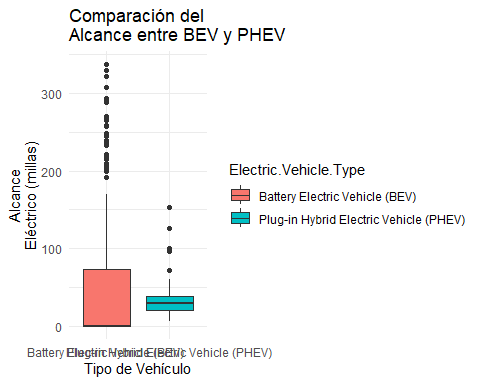


###Pregunta 3: Diferencias en el alcance eléctrico entre BEV y PHEV

### Gráfico de boxplot comparando BEV y PHEV

ggplot(datos, aes(x = Electric.Vehicle.Type, y = Electric.Range, fill =  
Electric.Vehicle.Type)) + geom\_boxplot() + labs(title = "Comparación del  
Alcance entre BEV y PHEV", x = "Tipo de Vehículo", y = "Alcance  
Eléctrico (millas)") + theme\_minimal()

## Warning: Removed 36 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).



###Pregunta 4: Relación entre elegibilidad CAFV y ubicación

### Contar registros por ciudad y elegibilidad

ciudad\_elegibilidad <- datos %>%  
 group\_by(City, Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility) %>%  
 summarise(Conteo = n(), .groups = "drop") %>%  
 arrange(desc(Conteo))

### Obtener la moda (ciudades con mayor frecuencia por tipo de elegibilidad)

moda\_ciudades <- ciudad\_elegibilidad %>%  
 group\_by(Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility) %>%  
 slice\_max(Conteo, n = 5) # Tomar las 5 ciudades con más registros  
# Tomar las 5 ciudades con más registros

### Obtener la mediana del conteo de registros por ciudad

mediana\_conteo <- median(ciudad\_elegibilidad$Conteo)

### Obtener la media del conteo de registros por ciudad

media\_conteo <- mean(ciudad\_elegibilidad$Conteo)

### Mostrar resultados

print("Moda (Top ciudades por elegibilidad):")

## [1] "Moda (Top ciudades por elegibilidad):"

print(moda\_ciudades)

## # A tibble: 15 × 3  
## # Groups: Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility [3]  
## City Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility Conteo  
## <chr> <chr> <int>  
## 1 Seattle Clean Alternative Fuel Vehicle Eligible 12269  
## 2 Bellevue Clean Alternative Fuel Vehicle Eligible 3136  
## 3 Vancouver Clean Alternative Fuel Vehicle Eligible 2840  
## 4 Redmond Clean Alternative Fuel Vehicle Eligible 2180  
## 5 Renton Clean Alternative Fuel Vehicle Eligible 2078  
## 6 Seattle Eligibility unknown as battery range has not been researched 21596  
## 7 Bellevue Eligibility unknown as battery range has not been researched 7690  
## 8 Redmond Eligibility unknown as battery range has not been researched 5565  
## 9 Bothell Eligibility unknown as battery range has not been researched 5490  
## 10 Vancouver Eligibility unknown as battery range has not been researched 4666  
## 11 Seattle Not eligible due to low battery range 3545  
## 12 Vancouver Not eligible due to low battery range 1004  
## 13 Tukwila Not eligible due to low battery range 826  
## 14 Renton Not eligible due to low battery range 709  
## 15 Bellevue Not eligible due to low battery range 683

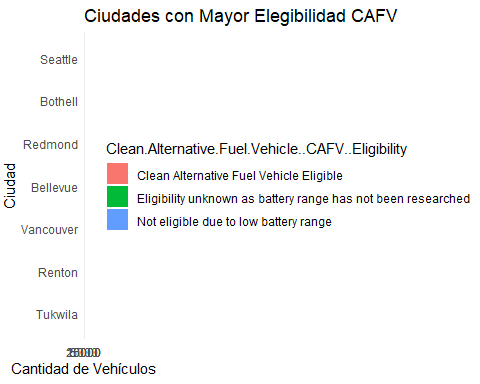
print(paste("Mediana de registros por ciudad:", mediana\_conteo))

## [1] "Mediana de registros por ciudad: 6"

print(paste("Media de registros por ciudad:", round(media\_conteo, 2)))

## [1] "Media de registros por ciudad: 145.13"

#windows()  
ggplot(moda\_ciudades, aes(x = reorder(City, Conteo), y =  
Conteo, fill = Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility)) +  
geom\_bar(stat = "identity", position = "dodge") + coord\_flip() +  
labs(title = "Ciudades con Mayor Elegibilidad CAFV", x = "Ciudad", y =  
"Cantidad de Vehículos") + theme\_minimal()



### Pregunta 5: Relación entre tipo de vehículo y proveedor eléctrico

### Contar registros por Tipo de Vehículo y Proveedor Eléctrico

vehiculo\_proveedor <- datos %>%  
 group\_by(Electric.Vehicle.Type, Electric.Utility) %>%  
 summarise(Conteo = n(), .groups = "drop") %>%  
 arrange(desc(Conteo))

### Obtener la moda (las combinaciones más frecuentes)

moda\_vehiculo\_proveedor <- vehiculo\_proveedor %>%  
 group\_by(Electric.Vehicle.Type) %>%  
 slice\_max(Conteo, n = 5) # Tomar las 5 combinaciones más comunes

### Calcular mediana y media del conteo de registros

mediana\_conteo <- median(vehiculo\_proveedor$Conteo)  
media\_conteo <- mean(vehiculo\_proveedor$Conteo)

### Mostrar resultados en consola

print("Moda (Top combinaciones de Tipo de Vehículo y Proveedor Eléctrico):")

## [1] "Moda (Top combinaciones de Tipo de Vehículo y Proveedor Eléctrico):"

print(moda\_vehiculo\_proveedor)

## # A tibble: 10 × 3  
## # Groups: Electric.Vehicle.Type [2]  
## Electric.Vehicle.Type Electric.Utility Conteo  
## <chr> <chr> <int>  
## 1 Battery Electric Vehicle (BEV) PUGET SOUND ENERGY INC||CITY O… 69806  
## 2 Battery Electric Vehicle (BEV) PUGET SOUND ENERGY INC 39356  
## 3 Battery Electric Vehicle (BEV) CITY OF SEATTLE - (WA)|CITY OF… 32278  
## 4 Battery Electric Vehicle (BEV) BONNEVILLE POWER ADMINISTRATIO… 10361  
## 5 Battery Electric Vehicle (BEV) BONNEVILLE POWER ADMINISTRATIO… 8271  
## 6 Plug-in Hybrid Electric Vehicle (PHEV) PUGET SOUND ENERGY INC||CITY O… 15492  
## 7 Plug-in Hybrid Electric Vehicle (PHEV) PUGET SOUND ENERGY INC 9567  
## 8 Plug-in Hybrid Electric Vehicle (PHEV) CITY OF SEATTLE - (WA)|CITY OF… 7828  
## 9 Plug-in Hybrid Electric Vehicle (PHEV) BONNEVILLE POWER ADMINISTRATIO… 3400  
## 10 Plug-in Hybrid Electric Vehicle (PHEV) BONNEVILLE POWER ADMINISTRATIO… 2414

print(paste("Mediana de registros por combinación:", mediana\_conteo))

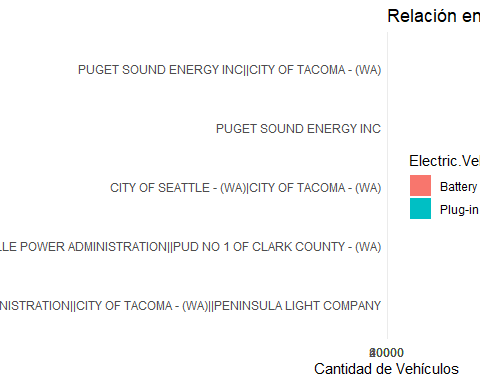
## [1] "Mediana de registros por combinación: 108"

print(paste("Media de registros por combinación:", round(media\_conteo, 2)))

## [1] "Media de registros por combinación: 1581.83"

### 🔹 Visualización de las combinaciones más comunes

#windows()  
ggplot(moda\_vehiculo\_proveedor, aes(x = reorder(Electric.Utility, Conteo), y = Conteo, fill = Electric.Vehicle.Type)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 coord\_flip() +  
 labs(title = "Relación entre Tipo de Vehículo y Proveedor Eléctrico",  
 x = "Proveedor Eléctrico",  
 y = "Cantidad de Vehículos") +  
 theme\_minimal()



#### **Conclusiones**

* Los lugares donde los proveedores de electricidad han desarrollado una buena red de carga y existen incentivos claros son los que tienen más vehículos eléctricos. Como nos demuestran las gráficas de la ciudad de Seattle.
* Aunque los híbridos enchufables (PHEV) se encuentren en el mercado, la mayoría de las personas están optando por los vehículos 100% eléctricos (BEV). Esto indica que la tecnología ha avanzado permitiendo mayor autonomía en estos, que los incentivos y lugares de carga cada vez están mejor distribuidos.
* Las empresas de electricidad tienen un rol clave en el futuro de la movilidad. No es solo una cuestión de que las personas quieran cambiar a autos eléctricos, sino que las compañías eléctricas deben estar preparadas para abastecer esa demanda. Aquellas que ya han trabajado en infraestructura están viendo cómo en sus regiones el número de autos eléctricos crece más rápido.
* Las compañías eléctricas deben estar preparadas para abastecer la demanda con su infraestructura ya que  están viendo cómo en sus regiones el número de autos eléctricos crece más rápido.
* Aunque hay avances importantes, aún hay proveedores con muy pocos autos eléctricos en sus redes, lo que indica que en algunas zonas falta infraestructura o incentivos atractivos.

#### **Referencias**

* Data.gov. (n.d.). Electric Vehicle Population Data. Retrieved from<https://catalog.data.gov/dataset/electric-vehicle-population-data>