Insights into Electric Vehicle Adoption: A Data Analysis Perspective

Joaquin Ramirez, Seth Harris, Collin Real

2024-06-07

The structure below is one possible setup for a data analysis project (including the course project). For a manuscript, adjust as needed.

You don’t need to have exactly these sections, but the content covering those sections should be addressed.

This uses MS Word as output format. [See here](https://quarto.org/docs/output-formats/ms-word.html) for more information. You can switch to other formats, like html or pdf. See [the Quarto documentation](https://quarto.org/) for other formats.

* https://www.google.com/publicdata/directory
* https://www.data.gov/
* https://github.com/awesomedata/awesome-public-datasets
* https://registry.opendata.aws/
* https://www.kaggle.com/datasets
* https://www.kaggle.com/datasets/jainaru/electric-vehicle-population

**- (a) Briefly describe what the data is, how it was collected, where you will get (or got) it from. How many observations do you have, what was measured? Anything else important to report about the data?**

The dataset obtained from Kaggle provides detailed information about Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) registered through the Washington State Department of Licensing (DOL). It comprises 181,458 observations and 17 variables, encompassing key details such as Vehicle Identification Number (VIN), county, city, model year, electric range, and more.

**- (b) At this stage you are not required to already have and show the data, but if you do, even better. Then add a few lines of code which load the data and using some of the commands you learned about, provide summary descriptions of the data.**

setwd('C:/Users/Joaquin/School/DA - 6833 (Summer 2024)/Project')  
data <- read.csv("Electric\_Vehicle\_Population\_Data.csv")  
str(data)

'data.frame': 181458 obs. of 17 variables:  
 $ VIN..1.10. : chr "WAUTPBFF4H" "WAUUPBFF2J" "5YJSA1E22H" "1C4JJXP62M" ...  
 $ 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 98043 98119 98501 ...  
 $ Model.Year : int 2017 2018 2017 2021 2020 2023 2017 2020 2022 2017 ...  
 $ Make : chr "AUDI" "AUDI" "TESLA" "JEEP" ...  
 $ Model : chr "A3" "A3" "MODEL S" "WRANGLER" ...  
 $ Electric.Vehicle.Type : chr "Plug-in Hybrid Electric Vehicle (PHEV)" "Plug-in Hybrid Electric Vehicle (PHEV)" "Battery Electric Vehicle (BEV)" "Plug-in Hybrid Electric Vehicle (PHEV)" ...  
 $ Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility: chr "Not eligible due to low battery range" "Not eligible due to low battery range" "Clean Alternative Fuel Vehicle Eligible" "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 ...  
 $ Legislative.District : int 34 22 22 20 14 22 23 1 36 22 ...  
 $ DOL.Vehicle.ID : int 235085336 237896795 154498865 154525493 225996361 220675367 162720022 6293899 207620633 237392459 ...  
 $ Vehicle.Location : chr "POINT (-122.374105 47.54468)" "POINT (-122.943445 47.059252)" "POINT (-122.78083 47.083975)" "POINT (-122.85403 46.856085)" ...  
 $ 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 5.31e+10 5.31e+10 5.31e+10 5.31e+10 ...

head(data)

VIN..1.10. County City State Postal.Code Model.Year Make Model  
1 WAUTPBFF4H King Seattle WA 98126 2017 AUDI A3  
2 WAUUPBFF2J Thurston Olympia WA 98502 2018 AUDI A3  
3 5YJSA1E22H Thurston Lacey WA 98516 2017 TESLA MODEL S  
4 1C4JJXP62M Thurston Tenino WA 98589 2021 JEEP WRANGLER  
5 5YJ3E1EC9L Yakima Yakima WA 98902 2020 TESLA MODEL 3  
6 1C4JJXP66P 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  
 Legislative.District DOL.Vehicle.ID Vehicle.Location  
1 34 235085336 POINT (-122.374105 47.54468)  
2 22 237896795 POINT (-122.943445 47.059252)  
3 22 154498865 POINT (-122.78083 47.083975)  
4 20 154525493 POINT (-122.85403 46.856085)  
5 14 225996361 POINT (-120.524012 46.5973939)  
6 22 220675367 POINT (-122.89692 47.043535)  
 Electric.Utility X2020.Census.Tract  
1 CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA) 53033011500  
2 PUGET SOUND ENERGY INC 53067011100  
3 PUGET SOUND ENERGY INC 53067012226  
4 PUGET SOUND ENERGY INC 53067012620  
5 PACIFICORP 53077000800  
6 PUGET SOUND ENERGY INC 53067010802

summary(data)

VIN..1.10. County City State   
 Length:181458 Length:181458 Length:181458 Length:181458   
 Class :character Class :character Class :character Class :character   
 Mode :character Mode :character Mode :character Mode :character   
   
   
   
   
 Postal.Code Model.Year Make Model   
 Min. : 1545 Min. :1997 Length:181458 Length:181458   
 1st Qu.:98052 1st Qu.:2019 Class :character Class :character   
 Median :98122 Median :2022 Mode :character Mode :character   
 Mean :98174 Mean :2021   
 3rd Qu.:98370 3rd Qu.:2023   
 Max. :99577 Max. :2024   
 NA's :3   
 Electric.Vehicle.Type Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility  
 Length:181458 Length:181458   
 Class :character Class :character   
 Mode :character Mode :character   
   
   
   
   
 Electric.Range Base.MSRP Legislative.District DOL.Vehicle.ID   
 Min. : 0.00 Min. : 0 Min. : 1.00 Min. : 4385   
 1st Qu.: 0.00 1st Qu.: 0 1st Qu.:18.00 1st Qu.:183068667   
 Median : 0.00 Median : 0 Median :33.00 Median :228915522   
 Mean : 57.83 Mean : 1040 Mean :29.11 Mean :221412778   
 3rd Qu.: 75.00 3rd Qu.: 0 3rd Qu.:42.00 3rd Qu.:256131982   
 Max. :337.00 Max. :845000 Max. :49.00 Max. :479254772   
 NA's :398   
 Vehicle.Location Electric.Utility X2020.Census.Tract   
 Length:181458 Length:181458 Min. :1.001e+09   
 Class :character Class :character 1st Qu.:5.303e+10   
 Mode :character Mode :character Median :5.303e+10   
 Mean :5.298e+10   
 3rd Qu.:5.305e+10   
 Max. :5.603e+10   
 NA's :3

colSums(is.na(data))

VIN..1.10.   
 0   
 County   
 0   
 City   
 0   
 State   
 0   
 Postal.Code   
 3   
 Model.Year   
 0   
 Make   
 0   
 Model   
 0   
 Electric.Vehicle.Type   
 0   
Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility   
 0   
 Electric.Range   
 0   
 Base.MSRP   
 0   
 Legislative.District   
 398   
 DOL.Vehicle.ID   
 0   
 Vehicle.Location   
 0   
 Electric.Utility   
 0   
 X2020.Census.Tract   
 3

**- (c) Explain the question you want to answer using the data. What will be your outcome(s) of interest (if any)? What (if any) specific predictors will you focus on? What relations/patterns are you looking for in the data?**

Question: What insights can be gleaned from the dataset regarding the electric vehicle landscape ?

## 1 Objectives:

* Understand Geographic Distribution: Determine the distribution of electric vehicles across different counties and cities in Washington State.
* Identify Popular Models: Identify the most commonly registered electric vehicle makes and models.
* Analyze Electric Range: Investigate the distribution of electric range among registered electric vehicles.-
* Explore Clean Alternative Fuel Vehicle Eligibility: Examine the eligibility for clean alternative fuel vehicle incentives among different vehicle types.

## 2 Outcomes of Interest

* Geographic Distribution: The distribution of electric vehicles across counties and cities in Washington State.
* Popular Models: The most frequently registered electric vehicle makes and models.
* Electric Range Analysis: The range distribution and average electric range of registered electric vehicles.
* Clean Alternative Fuel Vehicle Eligibility: The proportion of vehicles eligible for clean alternative fuel vehicle incentives.

## 3 Predictors

* Geographic Variables: County, City, Postal Code.
* Vehicle Attributes: Model Year, Make, Model, Electric Vehicle Type, Electric Range.

## 4 Relations/Patterns

* Geographic Distribution: Analyzing clusters of electric vehicles in specific regions.
* Popular Models: Identifying dominant electric vehicle makes and models.
* Electric Range Analysis: Understanding the distribution of electric ranges and potential implications for consumer preferences.
* Clean Alternative Fuel Vehicle Eligibility: Examining eligibility patterns and potential factors influencing eligibility.
* Correlations: Investigating any associations between electric vehicle characteristics and legislative districts or census tracts, which may offer insights into policy implications and consumer behavior.

**(d) As much as you know, suggest how you will analyze it. At this stage in the course, we haven’t covered analysis approaches yet, so you can keep things vague and non-technical here.**

* Descriptive Analysis: Summarize the dataset to understand its contents, including summary statistics and frequency tables.

-Visualization: Use plots and charts to visualize aspects like geographic distribution, electric range, and vehicle types.

* Comparative Analysis: Compare different groups within the dataset, such as BEVs vs PHEVs or eligibility for incentives by vehicle type.
* Correlation Analysis: Explore potential relationships between variables, such as electric range and base MSRP.
* Geospatial Analysis (optional): Use R packages to analyze spatial patterns in the data.
* Exploratory Data Analysis (EDA): Engage in exploratory analysis to identify interesting patterns or trends.

**(e) You are allowed, but not yet required, to provide background information for the question you plan to answer. For instance you can describe why it’s an interesting question, who else has done similar analyses, how your analysis will be new/different, etc. Similar to what you read in an introduction to a research paper. For the final report, you’ll need these parts. For part 1, they are not required, but you are welcome to already write down some of that.**

## 5 Primary Research Question: What insights can be gleaned from the dataset regarding the electric vehicle landscape in Washington State?

Sub-questions:

* How are electric vehicles distributed across different counties, cities, and States?
* Which EV makes and models are most popular?
* What is the range of electric vehicles in terms of electric range and base MSRP?
* How does eligibility for clean alternative fuel vehicle incentives vary among different vehicle types?
* Are there any significant correlations between EV characteristics and geographic or legislative factors?

**(d) Eventually, for your final report, what you write for this part will go into different sections of the full report. Some will go into the introduction, some in the methods section. You can already place these items there, or for now just write them as a single section.**

**Introduction** Electric vehicles (EVs), including Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs), are crucial in the global effort to reduce greenhouse gas emissions and dependence on fossil fuels. This dataset, sourced from the Washington State Department of Licensing (DOL) and obtained from Kaggle, provides a comprehensive overview of registered electric vehicles in Washington State. Analyzing this data offers valuable insights into the adoption patterns and characteristics of EVs, informing policies and infrastructure development to further promote sustainable transportation.

**Objectives:**

**Geographic Distribution:** Determine the distribution of electric vehicles across different counties and cities. **Popular Models:** Identify the most commonly registered EV makes and models. **Electric Range Analysis:** Investigate the distribution of electric range among registered EVs. **CAFV Eligibility:** Examine the eligibility for clean alternative fuel vehicle incentives among different vehicle types. **Correlation Analysis:** Explore potential correlations between EV characteristics and geographic or legislative factors.

**Methods**

**Data Cleaning and Preparation:**

* Address missing values and ensure correct data types.
* Transform variables as necessary for analysis.

**Descriptive Statistics:**

* Compute summary statistics (mean, median, mode, standard deviation) for numerical variables.
* Generate frequency tables for categorical variables.

**Data Visualization:**

* Create bar charts, histograms, and pie charts to visualize distributions.
* Use scatter plots and box plots to explore relationships between variables.
* Generate geographic maps to visualize EV distribution across the state.

**Comparative Analysis:**

* Compare BEVs and PHEVs in terms of electric range and base MSRP.
* Analyze differences in CAFV eligibility across makes and models.

**Correlation Analysis:**

* Calculate correlation coefficients for numerical variables.
* Use cross-tabulations to explore relationships between categorical and numerical variables.

**Geospatial Analysis (if applicable):**

* Use geographic mapping tools to identify clusters or patterns in vehicle registrations.

**Exploratory Data Analysis (EDA):**

-Perform in-depth analysis to identify trends, patterns, or outliers. -Use advanced visualization techniques for deeper insights.

**Conclusion:**

By analyzing the electric vehicle registration data for Washington State, this project aims to uncover significant patterns and trends that can inform policy decisions and infrastructure planning. The insights gained from this analysis will contribute to the broader understanding of EV adoption and support the transition towards a more sustainable transportation future.