OXFORD BROOKES UNIVERSITY

Assessment cover

STUDENTS, PLEASE COPY THIS PAGE AND USE AS THE COVER PAGE FOR YOUR SUBMISSION

Module No:	COMP6037	Module title:	Foundations of Data Analytics

Assessment title :	Coursework 2 – Data Visualisation

Due date and time**:**	1:00PM,16 Dec 2024

Estimated total time to be spent on assignment:	30 hours per student

LEARNING OUTCOMES

On successful completion of this assignment, students will be able to achieve the module following learning outcomes (LOs): *LO numbers and text copied and pasted from the module descriptor*
1. Critically analyse data visualisation approaches with respect to human sensory modalities
2. Create appropriate visualisations for temporal, dynamic, and high dimensionality data
3. Devise methodologies for data interaction to facilitate exploratory data analysis

Engineering Council AHEP4 LOs assessed (from S2 2022-23) *LOs copied and pasted from the AHEP4 matrix (add rows as required)*		
LO number	**LO text**	**Met? (Y/N)**

STUDENT NAMES (ONLY IF GROUP ASSIGNMENT, OTHERWISE ANONYMOUS)

Student No:	**Student Name:**	**Group Name and Number:**
1.		
2.		
3.		
4.		
5.		

Statement of Compliance *(please tick to sign)*

☑ I declare that the work submitted is my own and that the work I submit is fully in accordance with the University regulations regarding assessments (*www.brookes.ac.uk/uniregulations/current*)

RUBRIC OR EQUIVALENT (BELOW)

FORMATIVE FEEDBACK OPPORTUNITIES

SUMMATIVE FEEDBACK DELIVERABLES

Deliverable content and standard description and criteria	**Weighting out of 100%**
Individual reflection section: • Propose further work that would offer improvements and enhancements. • Evaluate personal learning and development in terms of technology/hardware/software/group work.	**10%**

Marking grid and peer marking form are attached at the end of this assignment.

ASSIGNMENT IN DETAIL

Table of Contents

Introduction

This coursework analyses electric vehicle (EV) adoption in Washington State using interactive visualizations. Trends by county and variations over time for vehicle types are investigated using R, ggplot2, and plotly, ensuring clarity and accessibility to a wide audience.

Part 1: Context and EDA

Source and context of the data

The data was sourced from the United States Government Website. The dataset is publicly available and can be accessed here: *https://catalog.data.gov/dataset/electric-vehicle-population-size-history-by-county*. It is titled "*Electric Vehicle Population Size History by County,"* and the data includes monthly historical registration counts of vehicles across various counties from January 2017 through October 2024 registered by the Washington State Department of Licensing (DOL). DOL generates this information by combining data from the National Highway Traffic Safety Administration (NHTSA) with Environmental Protection Agency (EPA) fuel efficiency ratings, DOL ownership and registration information (data.gov, 2024).

Although the original dataset covers multiple states and years across the United States, this analysis focuses exclusively on Washington State and spans the years 2021 to 2023. Washington State was chosen because it continuously ranks among the states with the highest adoption rates of electric vehicles. Also, the timeframe allows for exploration of recent trends coinciding with technological advancements and increased public awareness.

Figure 1: County Map of Washington State (Whereig.com, 2024)

Summary statistics

Data size

The selected dataset contains 1,873 rows and 10 columns, detailing electric and non-electric vehicle registrations in Washington State from 2022 to 2023. The exported CSV file size is approximately 112 KB.

Description of variable and datatype (class)

Variable Name	Type	Description	Notes
Date	Date	This day (the end of this month) is used to count the number of registered vehicles.	Converted from character to date.
County	Factor	The name of the county where the vehicle was registered.	39 unique counties.
State	Character	The state code (in this case, all are "WA").	Single value across all rows.
Vehicle Primary Use	Factor	Primary intended use of the vehicle. (Passenger or Truck).	Two levels: "Passenger" and "Truck".
Battery Electric Vehicle (BEVs)	Numeric	Number of Battery Electric Vehicles registered.	Continuous numeric data
Plug-in Hybrid Electric Vehicle (PHEVs)	Numeric	Number of Plug-In Hybrid Electric Vehicles registered.	Continuous numeric data
Electric Vehicle (EV) Total	Numeric	Total number of Electric Vehicles (BEVs + PHEVs).	Continuous numeric data
Non-Electric Vehicle Total	Numeric	Total number of Non-Electric Vehicles registered.	Continuous numeric data
Total Vehicles	Numeric	Total number of all vehicles (Electric + non-electric).	Continuous numeric data
Percent Electric Vehicles	Numeric	Percentage of Electric Vehicles out of Total Vehicles.	Continuous numeric data

Table1: Variables and Datatypes

Continuous variable summary

Variable	Minimum	Mean	Median	Maximum	Standard Deviation
Battery Electric Vehicles (BEVs)	0.0	1185.9	26.0	69524.0	5953.3
Plug-In Hybrid Electric Vehicles (PHEVs)	0.0	368.8	1.0	16679.0	1550.7
Electric Vehicle (EV) Total	0.0	1554.8	34.5	86203.0	7495.4
Non-Electric Vehicle Total	965	77421	25210	1354565	174257.3
Percent Electric Vehicles	0.0	0.691	0.240	6.150	1.007168

Table 2: Continuous variables summary

The above table shows summary statistics for continuous variables. The data shows great variance and wide ranges in values, with BEVs ranging from 0 to 69,524 and a high standard deviation of 5,953.3. PHEVs exhibit a comparable spread, with a standard deviation of 1,550.7 and a maximum of 16,679 values. Non-Electric Vehicles range from 965 to more than 1.3 million, whereas EV totals indicate a wide range, from 0 to 86,203. These figures show the diverse distribution of vehicle counts, both low and high vehicle usage in the selected dataset.

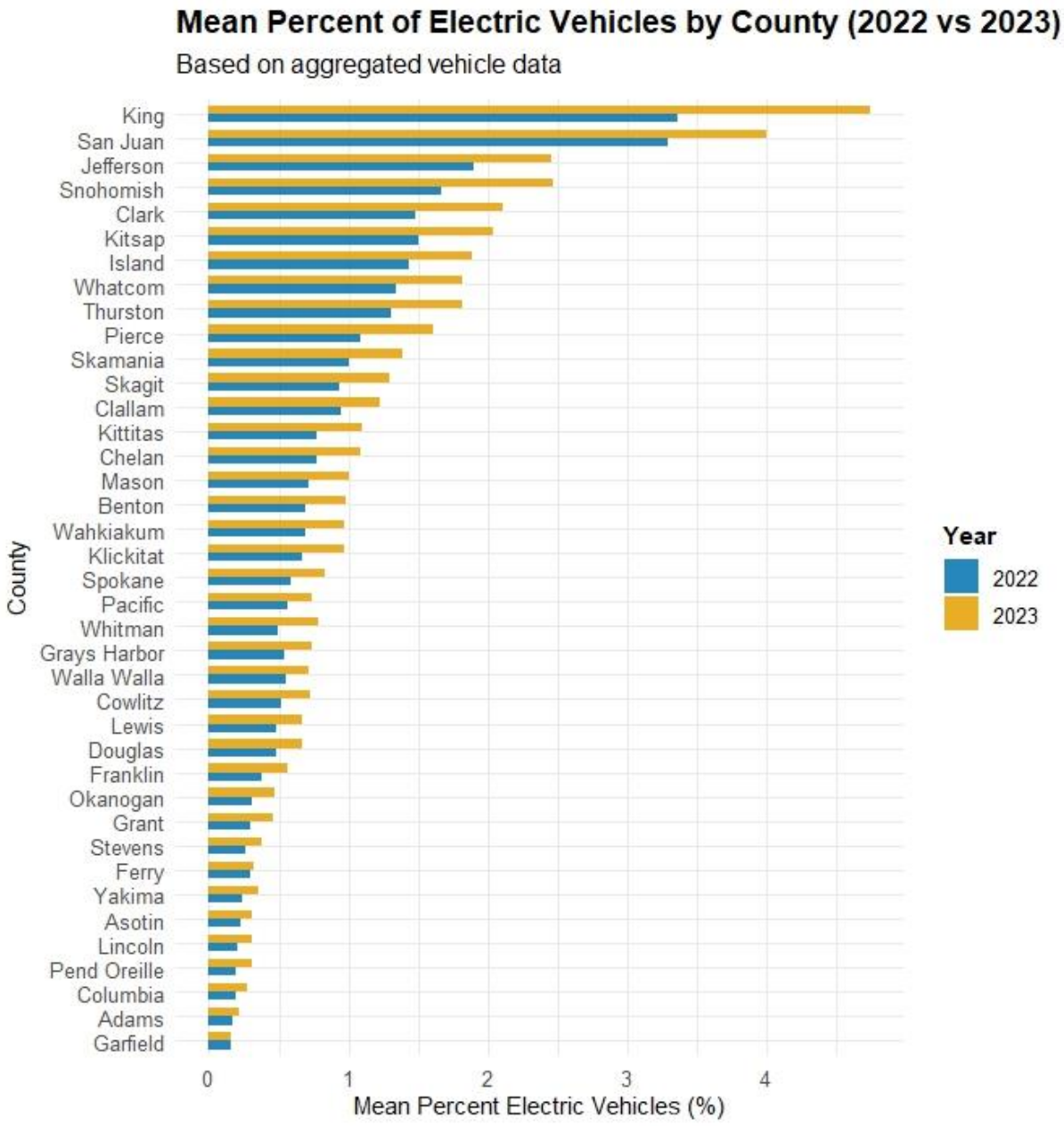

Figure 2: Summary Statistics

To summarize electric vehicle (EV) adoption trends across the 39 counties, which are categorical values, mean EVs percentage for each county across 2022 and 2023 is calculated. The counties' EV adoption rates were arranged from highest to lowest. Garfield County had the lowest mean percentage in both years, whereas King County continuously displayed the highest, indicating significant adoption rates. This demonstrates a notable variation in EV adoption between counties. Additionally, the data shows that all counties saw a rise in EV registrations in 2023 over 2022.

No Missing Data

The dataset does not contain any missing values. This is confirmed by the absence of NA values across all columns. A viz_miss plot, generated using the visdat package, confirms that no data is missing for any variables.

Part 2: Design

The data for these visualizations will be looked into electric vehicle adoption over time, focusing on different counties. By grouping the data by County or Date, comparing the number of EVs registered in various locations or at different times could help us better understand the changes in EV adoption.

Purpose of the Visualization

The purpose of this visualization is to investigate the trends and behaviours related to adoption of electric vehicle (EV) in Washington State. It will aim to highlight two key insights and visualize in two separate plots:

1. **Geographic Distribution**: The distribution of electric cars (EVs) in Washington State by county, showing regions with total registration numbers.
2. **Temporal Trends**: How EV registration numbers have changed month over month over the past two years, divided down into Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs).

Prospective Audience

The prospective audience can include policymakers, environmental organizations, general public and researchers. Based on geographical and temporal trends, policymakers can utilize the data to create infrastructure (such as charging stations) or targeted subsidies. Environmental advocate can comprehend how the use of EVs in Washington State affects the decrease of emissions. The general public, including locals and prospective EV purchasers, may view how their county compares to others. For additional investigation, such as examining transportation trends and their effects on the environment, researchers can utilize the image as a foundation.

Design Plan of the Visualizations

The visualization will consist of an **interactive map** created using the Plotly library, along with accompanying **interactive time-series bar-line chart**. Below is the description of each visualization:

1. Interactive County-Level Map:

Each county will be shaded with the logarithmic scale depending on its electric vehicle adoption. The map will have tooltips that provide comprehensive data on the number of EVs in each county. The goal is to highlight areas with high and low adoption rates and display spatial dispersion.

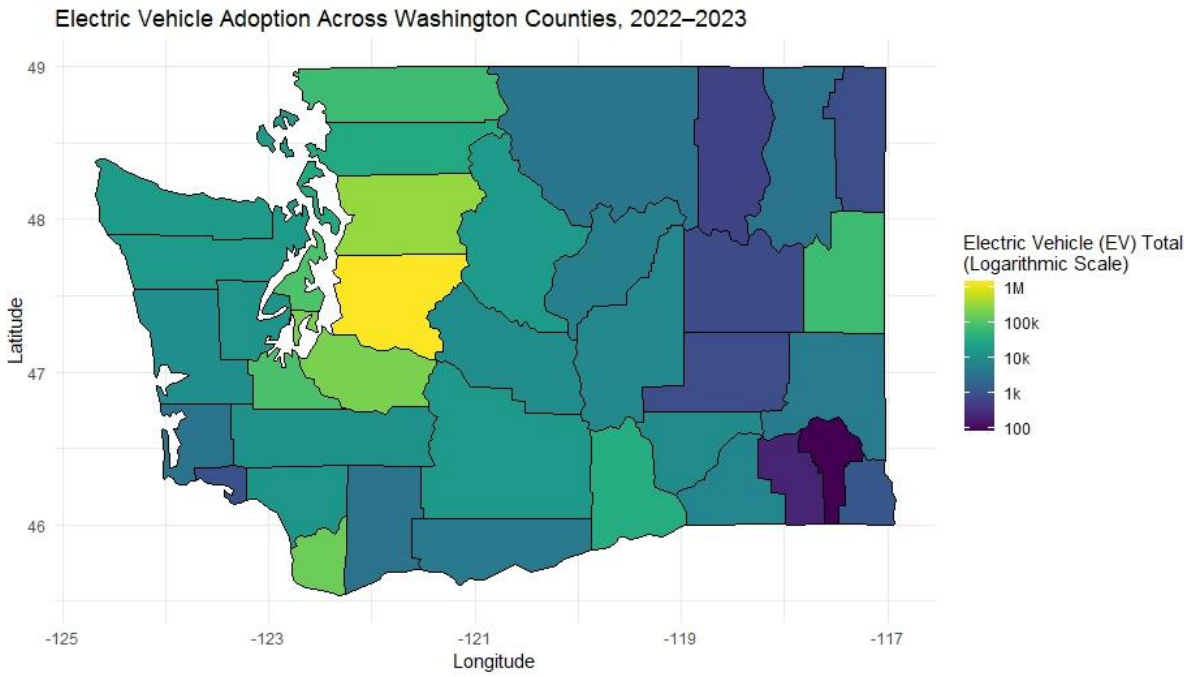

Figure 3: Static Map Representation

2. Time-series Bar-line Chart

It will be BEV and PHEV stacked bar charts with a line plot overlay for the entire EV population. Time (monthly intervals) will be shown by the x-axis, and the population of cars in the y-axis. Interactive features will be included with tooltips, vehicle-type filtering, and zooming.

The goal is to compare trends across BEVs and PHEVs and show how EV adoption has grown over time.

Time-Series Line Chart: shows the trend of EV registration growth over time.

Stacked Bar Chart: displays how different EV types—BEVs and PHEVs—contribute to the total number of registrations each month.

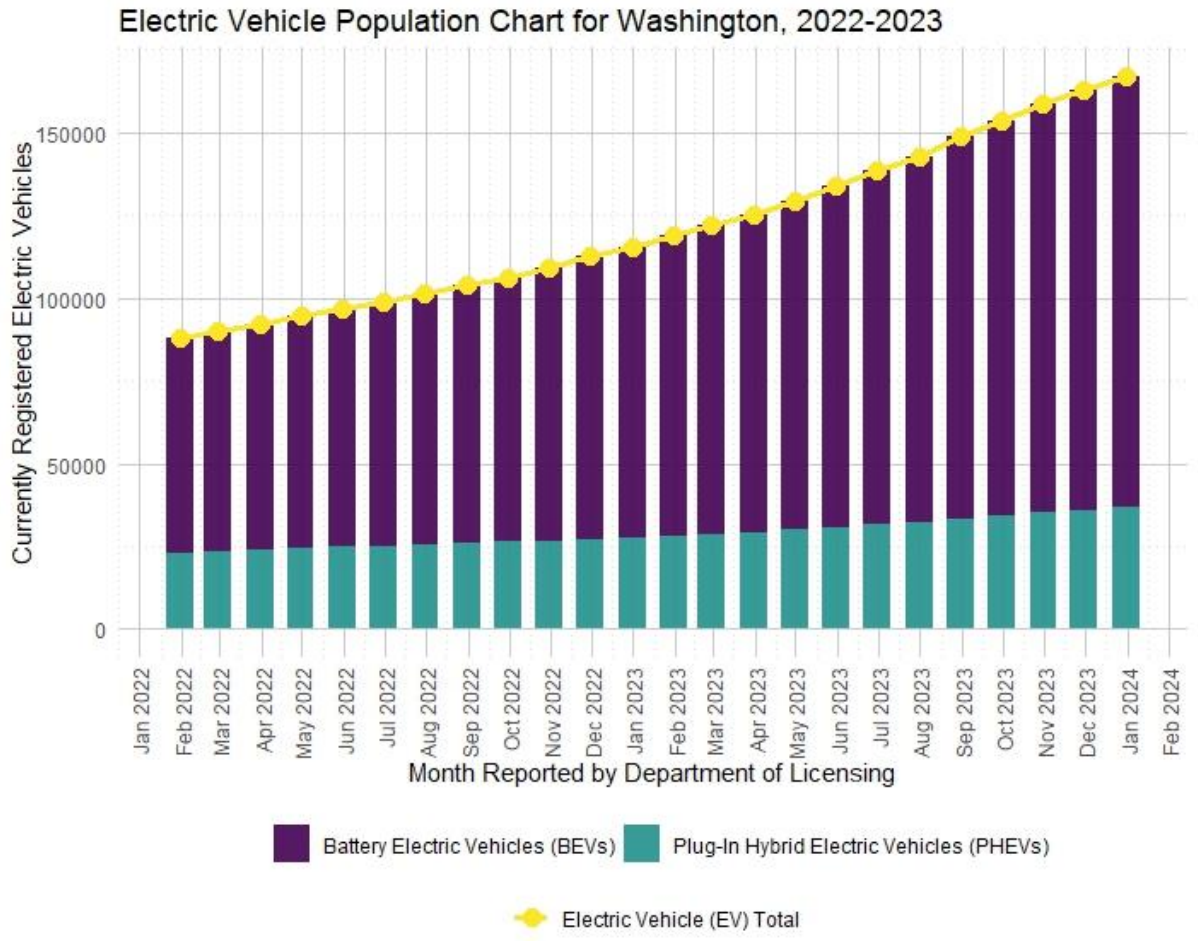

Figure 4: Static Time-series Bar-line Chart Representation

Part 3: Final Visualisations

Final Visualisation Commentary

The two interactive charts in the final visualizations, which were made with Plotly, are saved as an HTML file called visualization_2.html.

The legends for both graphs are positioned on the right. Software constraints prevented them from being placed adjacent to each graph, however the legends remain easily accessible.

The legends in Plot 2 have extra characters like commas, brackets, and "1". This occurred as a result of ggplot2's automatic handling of legends. Despite efforts to fix, this issue was unable to resolve. It may need further adjustments in future work.

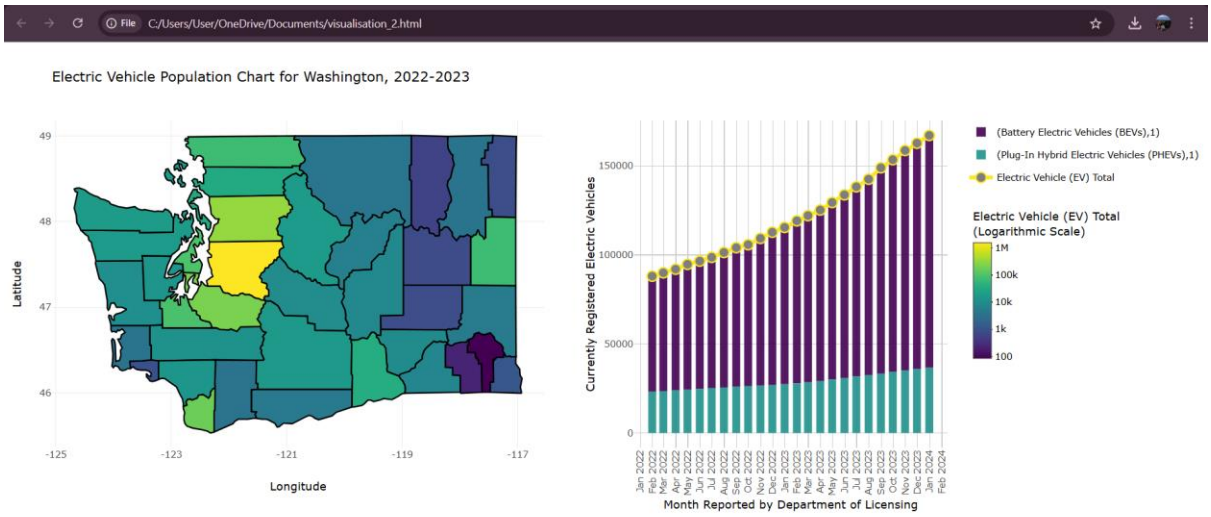

Figure 5: Final Visualization in HTML file

Accessibility

Accessibility for those with colour vision impairments is guaranteed by a colourblind-friendly palette. For ease of reading and uniformity in perception, the geographic map employs a viridis colour scheme. Usability is further improved by interactive tooltips and clear legends.

File Optimization

Data was combined by year and county to minimize file size, removing unnecessary information while keeping important insights. This method guarantees easier file exchange and quicker processing.

Audience Comprehension

Clear layouts and logical colour palettes highlight attention to the trends in EV adoption. A logarithmic scale improves the comprehension of data over a wide range of values.

Constraints

In accordance with project specifications, a reproducible script was created, guaranteeing platform compatibility and functionality.

Conclusion

This project produced interactive visualizations to highlight Washington's adoption trends for electric vehicles. Utilizing R with Plotly and ggplot2, the visualizations enable future policies to promote EV adoption by presenting key insights in an understandable manner.

References

Whereig.com, 2024. *Washington County Map, List of Counties in Washington with Seats.* [Online]
Available at: https://www.whereig.com/usa/states/washington/counties/
[Accessed 28 11 2024].

data.gov, 2024. *Electric Vehicle Population Size History By County.* [Online]
Available at: https://catalog.data.gov/dataset/electric-vehicle-population-size-history-by-county
[Accessed 23 11 2024].