Visualising the Balance Between Electric Vehicle Adoption and Charging Infrastructure in Washington State

1. Abstract

This project looks at how Electric Vehicle (EV) adoption and vehicle features are connected to the availability of charging stations in cities across Washington State. It compares city charging infrastructure with the number of EVs and their performance over time. The analysis brings together three datasets: Washington State EV registrations [7], a national list of alternative fuel stations [4], and technical details about EVs [2]. Combined, these datasets give a complete picture of vehicles, charging options, and performance.

Preparing the data took several steps, such as removing ID fields, handling missing information, making data types consistent, and combining data from vehicles, stations, and models. The final dataset includes more than 260,000 EV registrations, along with city-level charging information and technical details about each vehicle.

An animated bubble scatterplot was created using Plotly [5] to show how the average EV range compares to the number of charging ports per 100 EVs for different model years. The results suggest that cities with better charging infrastructure usually have more EVs and vehicles with longer ranges. However, some cities are seeing EV adoption grow faster than charging options, which creates challenges for city planning and policy.

2. Dataset(s)

This project uses three datasets that cover EV adoption, charging infrastructure, and vehicle technical features.

- The Washington State Electric Vehicle Population dataset, from the state's Department of Licensing [7], lists each registered EV. After cleaning, it has 264,628 records and 13 variables, such as make, model, year, EV type, range, price, and location details. This dataset is the main source for measuring EV adoption.

- The Alternative Fuel Stations dataset, published by the U.S. Department of Energy (NREL) [4], contains 27,922 station records across the United States. A subset of relevant fields was retained: fuel type, location, geographic coordinates, and counts of Level 1, Level 2, and DC fast chargers. This dataset enables the derivation of city-level charging infrastructure metrics for Washington State.

- The EV Specifications dataset [2] covers 478 vehicle models and 22 technical features, such as battery size, range, acceleration, top speed, efficiency, drivetrain, and body type. It adds detailed performance data not found in the registration records.

Big Data Characteristics:

- Volume: Over 260,000 vehicle records.
- Variety: Integration of administrative, infrastructure, and technical datasets at different granularities.

- Velocity (temporal): Use of model year and station opening year to analyse change over time.

Besides size and content, these datasets also vary in structure and reliability. The Electric Vehicle registrations dataset is highly consistent and complete but lacks technical details. The Alternative Fuel Stations dataset is semi-structured and has many missing values, which reflects differences in reporting across states and operators. The EV specifications dataset is detailed but smaller and needed careful matching because of differences in model names.

The integration of these datasets introduces genuine data variety, as it combines record-level vehicle ownership data, station-level infrastructure data, and model-level engineering data. This multigranularity integration allows adoption behaviour, infrastructural capacity, and technical performance to be analysed together within a single analytical framework.

3. Data Exploration, Processing, Cleaning & Integration

Data preparation followed a clear process: exploring, cleaning, filling in missing values, and combining sources to make sure the data was consistent and ready for visual analysis.
When exploring the Electric Vehicle registration dataset, we found both categorical and numerical variables, with very few missing values. We removed several identifier variables that did not add value for analysis. These included:

- Vehicle ID
- Truncated VIN
- Legislative district

Numerical attributes such as electric range and base price were imputed using median values to reduce the influence of extreme values associated with premium vehicle models. City and county fields were completed using modal imputation to preserve realistic geographic groupings, while postal codes were standardised as integers. Missing vehicle location values were labelled as “Unknown” to avoid unnecessary record deletion.

The alternative fuel stations dataset required substantially more cleaning due to very high missingness in administrative variables. All columns with more than 80-90% missing values were removed to reduce noise and dimensionality. Key processing steps included:

- Filling Level 1, Level 2 and DC fast charging port counts with zeros
- Completing missing categorical fields (city, address, network, connector type) with “Unknown”
- Converting station open dates to year-only values to preserve temporal context

This ensured that the dataset accurately represented charging infrastructure intensity without introducing artificial distortion.

The EV technical specifications dataset was comparatively clean. Small amounts of missing data in numerical fields were filled using median imputation, while categorical gaps were handled using either modal values or “Unknown” labels, ensuring a fully complete technical dataset.

Dataset integration was carried out in a staged manner. Vehicle registrations were first linked to technical specifications using cleaned and standardised make-model keys to minimise mismatch caused by formatting differences. Charging infrastructure data were then:

- Filtered to electric-only stations
- Restricted to Washington State

- Aggregated to the city level to generate infrastructure intensity metrics

These city-level infrastructure measures were merged into the combined vehicle specifications dataset using a left join on city and state. Missing infrastructure values were set to zero to represent cities without recorded charging stations. A left-join strategy was deliberately chosen to avoid accidental loss of valid vehicle registration records. Final validation confirmed that the row count remained unchanged at 264,628 records with 35 attributes, ensuring that integration introduced no sampling bias.

4. Visualisation

4.1 Visual Question and Chart Choice

Research Question:
How does the intensity of city-level charging infrastructure relate to EV adoption and vehicle capability over time in Washington State?

To answer this question, I created an animated bubble scatterplot using Plotly [5]. This type of chart works well because all the main variables are quantitative:

- X-axis: Charging ports per 100 EVs (infrastructure intensity)
- Y-axis: Average EV range (vehicle capability)
- Bubble size: Number of registered EVs (adoption scale)
- Colour: Infrastructure intensity
- Animation: Model year (temporal evolution)

This encoding allows four dimensions of information to be shown simultaneously in a perceptually efficient way.

4.2 Design Process & Sketch

Early exploration involved bar charts of EV counts, static scatters of EVs versus chargers, and distribution plots of EV ranges. While informative, these lacked explanatory power. Additional experiments included parallel coordinates [1] and dumbbell comparisons [3], which revealed imbalances but failed to communicate temporal evolution effectively.

Next, a hand-drawn sketch was created for a time-based bubble plot, which included the following features:

- Infrastructure on the horizontal axis
- Vehicle capability on the vertical axis
- Adoption scale via bubble size
- Colour for infrastructure
- An animation control for time

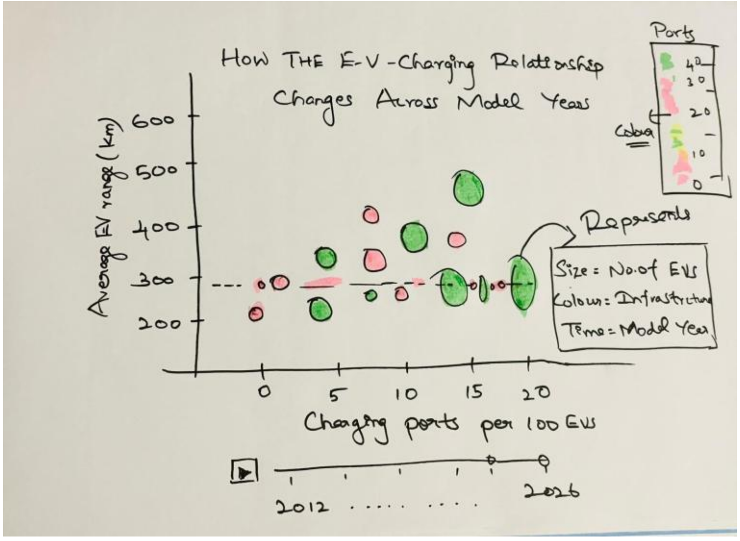

Figure 1: Design sketch of the animated bubble scatterplot for EV range, charging intensity, and adoption over time.

This design ensured that the most important relationships are encoded using high-accuracy perceptual channels such as position and motion.

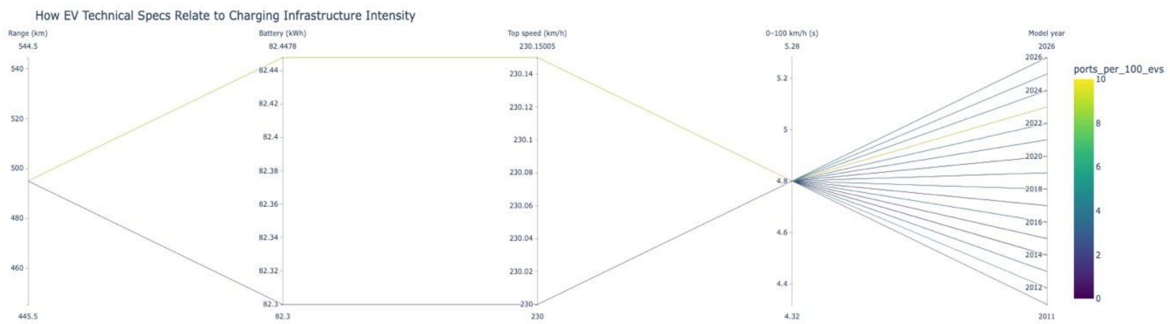

Figure 2: Exploratory parallel coordinates plot of EV technical specifications coloured by infrastructure intensity

This view highlights how multiple technical attributes (range, battery capacity, top speed and acceleration) vary across different infrastructure conditions. While useful for multivariate comparison, it does not effectively communicate temporal change or adoption scale, so it was rejected as the final explanatory visualisation.

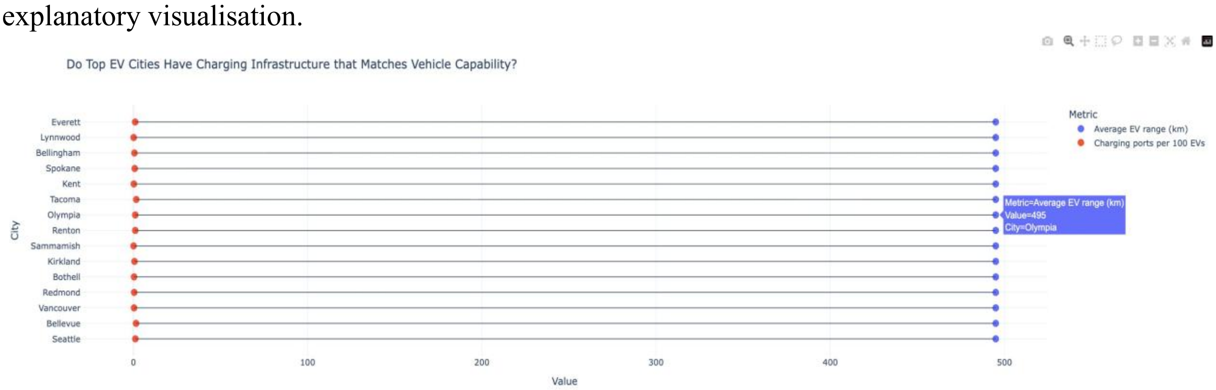

Figure 3: Exploratory dumbbell chart comparing average EV range and charging ports per 100 EVs for leading cities

This chart exposes mismatches between vehicle capability and infrastructure at a single time snapshot, but it cannot illustrate how this relationship evolves over time. Therefore, it was used for early insight only and not selected as the final explanatory chart.

4.3 Design Justification & Interactivity

The final interactive visualization was built using Plotly. Before plotting, data were grouped by city and model year. Cities with very few EVs were excluded to keep the results clear. The following design principles were used:

- Position shows the two key quantitative variables
- Area (bubble size) communicates adoption scale effectively
- Sequential colour reinforces infrastructure intensity
- Hover tooltips show exact values without making the plot look crowded
- Animation allows users to observe how city positions evolve across years

The animation turns the chart from a simple comparison into a story over time, showing how infrastructure and vehicle technology improve together in different cities [6].

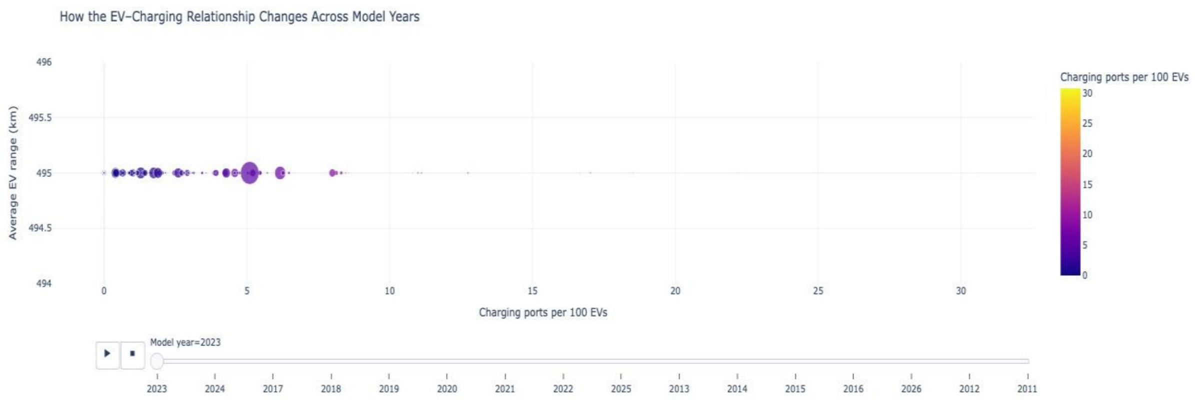

Figure 4: Final explanatory animated bubble scatterplot of average EV range versus charging ports per 100 EVs across model years

To improve readability and interpretability, the visualisation uses a clean white background, consistent typography, and carefully scaled axes to avoid visual distortion. City labels and interactive tooltips reduce cognitive load by presenting detailed information only on demand. These choices support efficient visual comparison while maintaining a clear and uncluttered layout.

The choice of a bubble scatterplot is further supported by perceptual research which identifies spatial position as the most accurate visual encoding channel for quantitative comparison. Encoding infrastructure intensity and vehicle capability on orthogonal axes allows users to rapidly assess correlation and clustering patterns. Bubble size introduces a third quantitative channel while maintaining readability due to careful filtering of low-frequency cities. Colour is used redundantly with horizontal position to reinforce infrastructure intensity without overloading visual cognition. These layered encodings enable multivariate analysis while retaining interpretability for non-technical audiences [6].

5. Conclusion

The integration of vehicle registration data, charging infrastructure data, and electric vehicle (EV) technical specifications produced a unified dataset representing the EV ecosystem in Washington State. Systematic data cleaning, transformation, and multi-source integration established a reliable foundation for explanatory visualisation. The resulting animated bubble scatterplot demonstrates the concurrent evolution of EV adoption, vehicle capabilities, and charging infrastructure across cities and over time.

The results show that cities with stronger charging infrastructure generally support larger EV fleets and higher-range vehicles, highlighting the importance of infrastructure in enabling sustainable transport. The visualisation also reveals locations where adoption is growing faster than infrastructure, signalling potential future pressure on charging networks. These findings are relevant for planners and policymakers, and future work could extend the study through geographic mapping, charger-type analysis, or finer-grained temporal modelling.

6. References

[1] An Introduction to Parallel Coordinates (2020) https://www.youtube.com/watch?v=FuMK_x4rmz4

[2] Electric Vehicle Specifications 2025 Dataset, Kaggle, accessed 2025 https://www.kaggle.com/datasets/urvishahir/electric-vehicle-specifications-dataset-2025

[3] Matplotlib Dumbbell charts (2021) https://www.youtube.com/watch?v=qsOcLIo5vYc

[4] National Renewable Energy Laboratory (NREL), Alternative Fuel Stations API, U.S. Department of Energy, accessed 2017-2025 https://developer.nrel.gov/docs/transportation/alt-fuel-stations-v1/

[5] Plotly Technologies Inc., Plotly Express Documentation, accessed 2025 https://plotly.com/python/plotly-express/

[6] Rosling, H. "Gapminder: Visualizing Global Development with Animated Bubble Charts." Gapminder Foundation (2006) https://www.gapminder.org/tools/

[7] Washington State Department of Licensing, Electric Vehicle Population Data, accessed 2025 https://catalog.data.gov/dataset/electric-vehicle-population-data