**Does Clean Electricity Drive Electric Vehicle Adoption?**

As the global push for decarbonization accelerates, electric vehicles (EVs) are often promoted as a key solution for reducing transport-related emissions. However, the environmental benefits of EVs depend heavily on the cleanliness of the electricity used to charge them. This analysis will explore the relationship between EV adoption (measured by electric cars sold per year) and the carbon intensity of national energy systems (CO₂ emissions per unit of energy consumed).

**Hypothesis:**

Countries with cleaner electricity, meaning lower CO₂ emissions per unit of energy, will have higher rates of EV adoption.

EV use grows faster in countries where electricity comes from clean sources like wind or solar. But if electricity is made from coal or oil, electric cars don’t help as much, so fewer people may buy them.

**Investigating this relationship is important because:**

* It may show if clean electricity grids are needed before many people start using EVs.
* It can help decide if countries should focus on cleaning their power grids first or promote EVs and clean energy together.
* It adds detail to the conversation about electric transport by looking not just at EV sales, but also at how clean the electricity is.

**Search strategy:**

I searched for good and reliable data about energy consumption and electric vehicles (EVs) from trusted websites. I mainly used Kaggle and government open data sites like data.gov. I also checked well-known sources like IEA and Our World in Data. I looked for datasets with country-level or regional information on EV use, energy consumption, and CO₂ emissions. I made sure the datasets cover several years and include the same countries or regions. Finally, I checked that the data can be used for research by looking at their licenses.

**Evaluated dataset pairs:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Pair** | **Dataset 1** | **Source** | **Dataset 2** | **Source** |  |
| 1 | Global EV Data Explorer (EV sales & stock) | IEA | World energy data 1990 - 2020 | Kaggle |  |
| 2 | Global EV Sales: 2010–2024 | Kaggle | Primary Energy Consumption per Capita | Our World in Data |  |
| 3 | Number of New Electric Cars Sold | Our World in Data (IEA-derived) | Energy Consumption by Source & Country | Our World in Data |  |
| 4 | Global EV sales 2010-2024 | Kaggle | Global Fossil Fuel Consumption | Our World in Data |  |
| 5 | Electric Vehicle Population Dataset | Kaggle | Global Energy Consumption | Kaggle |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pair** | **Dataset Pair** | **Completeness** | **Temporal Alignment** | **Geographic Alignment** |
| 1 | Global EV Data Explorer IEA / World energy data 1990 - 2020 | High – Minimal missing values | High – Multi-year, annual | High – Global coverage |
| 2 | Global EV Sales 2010–2024 Kaggle / Primary Energy Consumption per Capita Our World in Data | High – Mostly complete | Medium–High – Overlap exists | High – Global coverage |
| 3 | Number of New Electric Cars Sold — OWID / Energy Consumption by Source & Country OWID | High – Both clean datasets | High – Perfect overlap | High – Same country codes |
| 4 | Electric Vehicle Population Dataset Kaggle / Global Energy Consumption Kaggle | Medium – Missing entries | Medium – Limited overlap | Medium – EV dataset less global |
| 5 | Global EV Sales 2010–2024 Kaggle / Global Fossil Fuel Consumption — OWID | High – Minimal missing values | High – Overlapping years possible | High – Broad country coverage |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pair** | **Data Granularity** | **Source Reliability** | **Overall Suitability** |  |
| 1 | Annual, aggregated by country | Very High (IEA gold standard, Kaggle credible) | High | Perfect for EV adoption vs energy/emissions; clean and well-aligned |
| 2 | Annual per country | High (Kaggle + OWID credible) | High | Great for long-term trend comparisons; minor range differences |
| 3 | Annual per country | Very High (OWID) | High | Easy merge, lines up well, but sources are same |
| 4 | Annual but inconsistent for EV dataset | Medium (user-sourced Kaggle datasets) | Medium | Requires cleaning; limited alignment and completeness |
| 5 | Annual per country | High | High | Strong for EV adoption vs fossil fuel decline comparisons |

**Chosen Datasets:**

1. Global EV Data Explorer – IEA (Electric Vehicle sales & stock)

# World energy data 1990 - 2020 – Kaggle

1. Common Dimensions

* Temporal dimension: Both datasets include year (first dataset: *Year*, second dataset: *Year*), enabling alignment across time periods.
* Geographic dimension: Both datasets have country identifiers (first dataset: *country*, second dataset: *Entity/Code*) for matching countries.
* Numerical dimension:  
  + Dataset 1 contains energy and emissions metrics (*CO2 emissions from fuel combustion, energy production/consumption, renewables/electricity shares, oil/gas/coal production and consumption, energy intensity, etc.*).
  + Dataset 2 contains EV adoption metrics (*Electric cars sold*).

2. Possible Connections

* Temporal connection: Match each country’s energy and emissions indicators with its EV sales in the same year.
* Geographic connection: Link energy/emissions data for each country with the corresponding EV sales data.
* Numerical connection: Compare EV adoption trends with factors such as fossil fuel consumption, CO₂ emissions, renewable energy shares, and electricity production/consumption.

3. Documented Join Keys

* country ↔ Entity
* Year ↔ Year

**4. Data Linkage Strategy**

* Make sure country names/codes match in both datasets (for example, *“United States” vs “USA”*).
* Check that the *Year* format is the same in both datasets and only use years that exist in both.
* Create a carbon intensity value for each country and year:  
   CO₂ emissions from fuel combustion (MtCO₂)÷Total energy consumption (Mtoe)\text{CO₂ emissions from fuel combustion (MtCO₂)} \div \text{Total energy consumption (Mtoe)}CO₂ emissions from fuel combustion (MtCO₂)÷Total energy consumption (Mtoe)
* Join the datasets using (*country/Entity, Year*) so that EV sales line up with the energy and emissions data.
* Include energy breakdowns like coal, oil, gas, renewables, and wind/solar shares to compare clean vs dirty energy.
* After merging, look for missing values or mismatches and fix them (by filling in or removing).
* In the final table, each row will represent one **country-year** and will have:  
  + Electric cars sold
  + Carbon intensity
  + Energy source fractions (coal, oil, gas, renewables, wind/solar)
  + Other energy/emission info (like electricity use, energy intensity)

### 5: Exploratory Data Analysis

### The goal of this task was to perform an initial exploration of the merged EV and CO₂ emissions datasets to uncover patterns, outliers, and potential relationships. This step helps guide the final visualization and confirms whether our hypothesis about cleaner electricity and EV adoption holds across countries and years.

* **Line Chart:** Showed the trend of electric cars sold from 2015 onwards for the top 10 countries. This allows us to see which countries are increasing EV adoption fastest. Hover interactions highlight individual country trends.
* **Scatter Plot:** Plotted EV sales against CO₂ emissions from fuel combustion for the same countries. Outliers (countries with unusually high EV sales or emissions) are highlighted in red. Brushing allows selection of a subset of points to focus on specific ranges or countries.
* **Map Visualization:** Used Leaflet to show geographic distribution of total EVs for the top 10 countries. Marker size reflects total EV sales, and popups display country-specific totals. Color coding matches the line/scatter charts for consistency.

**Outlier Analysis:**

* Identified the top 5% of EV sales and CO₂ emission values as potential outliers.
* These points were emphasized in the scatter plot to check for unusual patterns or exceptions in the relationship between EV adoption and national emissions.

**Initial Insights:**

* Countries with cleaner electricity grids do not always have the highest EV sales, suggesting additional factors influence adoption.
* Some high-EV countries also have high CO₂ emissions, indicating that EV adoption may be driven by policy or incentives rather than electricity cleanliness alone.
* Line charts confirm that growth trends differ: some countries show steady adoption, while others spike in recent years.

**Potential Correlations:**

* A preliminary Pearson correlation coefficient was calculated between EV sales and CO₂ emissions, giving an initial quantitative measure of the relationship.
* While the correlation is weakly negative in some years (supporting the hypothesis), it is not uniform across all countries, highlighting the complexity of adoption patterns.

**Interactivity and User Exploration:**

* Hover effects on line charts and scatter points provide immediate insights into individual countries.
* Brushing in the scatter plot allows filtering to explore subsets of data interactively.
* Map popups link geographic locations with numeric summaries for easy cross-checking.
* 6 — **Justification & Design Choices**
* **Spatial vs. non-spatial:** Scatter plot reveals non-spatial relationships (energy consumption vs. CO₂). The two map-based views highlight spatial context and regional patterns.
* **Data volume & complexity:** With only 10 countries, all three are readable. D3 scatter scales to more variables (size/colour encodings). Maps remain clear with low point density, no clustering needed.
* **Audience needs:** Executives: quick map overview. Analysts: quantitative comparison (scatter, axes, tooltips). Educators: hybrid view demonstrates layering geospatial + multivariate marks.