Advancing Sustainable Mobility : A Data Visualization Analysis of the U.S. EPA Automotive Trends

Charles Fredric G. Inventado
University of Santo Tomas
Malolos City, Bulacan
9625037088, (+63)
charlesinventado@gmail.com

Rodelas, John Vincent B.
University of Santo Tomas
Deparo, Caloocan
9563959525 (+63)
jvincent4215@gmail.com

Valles, James Vincent V.
University of Santo Tomas
Mandaluyong, Manila
96547776746, (+63)
jamesvincr@gmail.com

ABSTRACT
The transport sector remains a primary contributor to global greenhouse gas emissions, presenting a critical challenge to sustainable development. The study leverages historical automotive data from the U.S. Environmental Protection Agency (EPA) to analyze trends in fuel economy, CO_2 emissions, and technological adoption from 1975 to the present. By aligning with United Nations Sustainable Development Goals (SDG) 7 (Affordable and Clean Energy) and 13 (Climate Action), the paper employs advanced data visualization techniques—including bump charts, treemaps, and dual-axis plots—to decouple technological efficiency gains from market shifts towards heavier vehicles. The findings aim to reveal the "masking effects" where engineering advancements are offset by increased vehicle mass and power, providing data-driven insights for future policy and consumer awareness.

CCS Concepts
- Human-centered computing -> Visualization -> Visualization application domains -> Information visualization (500)
- Applied computing -> Physical sciences and engineering -> Earth and atmospheric sciences -> Environmental sciences (500)
- Mathematics of computing -> Probability and statistics -> Statistical paradigms -> Exploratory data analysis (300)
- Information systems -> Data management systems -> Data cleaning (300)

Keywords
Data Preprocessing, Data Cleaning, Data Visualization, Sustainable Development Goals (SDGs), Automotive Trends, CO2 Emissions, Fuel Economy, Sustainable Mobility, Exploratory Data Analysis, Cleaner Technologies

1. PROJECT BACKGROUND

1.1 Global Issue
The study tackles global issues prevalent in the transport sector as one of the largest contributors of greenhouse gas emissions affecting the air quality from urban areas and catalyzes the impacts of climate change. Considering that mobility today is on demand and heavily reliant on traditional fossil fuels, we will face long-term consequences regarding sustainability as resources may go scarce and be limited in the future.

1.2 SDG Goals
The project aims to contribute on SDG 7 (Affordable and Clean Energy) by examining the fuel economy to have a better insight into how vehicles advance toward clean and reliable modern energy, and SDG 13 (Climate Action) by analyzing the Carbon Dioxide (CO2) emissions data and the adoption of other fuel-efficient and cleaner vehicle technologies to promote sustainable consumption and to take urgent action on CO2 emission reduction to and take combat climate change.

2. Statement of the Problem
The sector of transportation is a primary cause of global carbon dioxide (CO2) emissions, contributing for a substantial portion of consumption of fossil fuels, climate change, air pollution, and insecurity of energy. In the U.S., many fleets of vehicles still continue to rely on internal combustion engines despite the regulatory advancements since 1975. These results in persistent challenges to achieving much higher fuel economy standards and widespread adoption of cleaner technologies and also slows down the progress toward United Nations Sustainable Development Goals, specifically the SDGs 7 (Affordable and Clean Energy) and 13 (Climate Action), and other commitments involving international climate agreements such as the Paris Accord. The transition to a sustainable and low-emission mode of transport remains suboptimal without having any deeper insights into historical trends, patterns, and barriers, increasing environmental degradation and delaying global policy objectives for emission reductions

2.1 Objectives
The project will utilize data visualization techniques to discover important trends and occurrences and discover long-term patterns in vehicle fuel and economy efficiency and emissions. Addressing the key challenges in the transport sector will help us uncover relevant insights and strategies that could be possibly done aligned with global sustainability efforts, thus, we could successfully contribute to the target SDGs by committing to these objectives:

- Examine relevant variables and be able to visualize historical trends in fuel economy, CO2 emissions, and the adoption of fuel-efficient and cleaner vehicle technologies using the data available spanning from 1975 to preliminary 2024 data.
- Identify key patterns, progressions, and significant factors in the U.S. vehicle sector aligned with SDG 7 to sustain clean energy, and SDG 13 to prevent the severity of impacts of climate change.

- Brainstorm to form and be able to address SDG-aligned questions by having data-driven analysis, exploring the economic, environmental, and technological implications of emission reductions.
- Provide recommendations that can be improved by policymakers, industry stakeholders, and researchers to promote sustainable consumption, accelerate technology adoption, and support global development goals and commitments in the transportation sector.

2.2 Scope

For the scope of the study, first, the vehicles fleets and regulations were based only from the country the dataset originated from, the United States (U.S.), thus excluding any other countries than the stated one. Second, the data covers only from 1975 until the preliminary of 2024, which may not include the latest EV technologies and innovations made recently of late 2024 and the rest of 2025. Lastly, regards to contributing pollution features are focused only on the CO2 emissions and not other contributing factors such as Nitrogen Oxide and Sulfur Oxide.

2.3 Limitations

Regarding the limitation of the study, first, the dataset is containing missing values on columns requiring imputation of data on some records such as when a vehicle will not have complete value on all of powertrain columns and the number of gears, there are also missing years on manufacturers such as Tesla's data were available since only 2011. Second, percentages or fraction values are to describe technological adoption such as hybrid systems and transmission type present on the dataset, which needs the team to carefully interpret to use the appropriate visualization or model. Third limitation is that vehicle type values were aggregated averages, that not a specific car model is represented thus limiting the possibility of a more detailed analysis.

3. BACKGROUND ON THE DATASET

Name : A. Detailed Real-World Fuel Economy, CO2 Emissions, and Vehicle Attribute and Technology Data

Source : EPA United States Environmental Protection Agency

https://www.epa.gov/automotive-trends/explore-automotive-trends-data#SummaryData

3.1 Data Coverage

Regarding coverage, notice from the Model Year column, the time span of the records are usually from 1975 to 2024. The records were also further organized into the 15 (including "all" on the count) manufacturers of the automobile, the corresponding regulatory class (All, Car, and Truck), and the vehicle classification. (All, All Car, All Truck, Car SUV, Minivan/Van, Pickup, Sedan/Wagon, Truck SUV)

3.2 Indicators

Aside from the stated four column identifiers stated above, from the original dataset gathered, there are over 56 indicators that could be themed by production (ex: Production (000), Production Share) , fuel efficiency performance (ex : 2-Cycle MPG, Real-World MPG, etc…), Carbon Emissions (ex : Real-World CO2 (g/mi), etc…), Vehicle specifications (ex : Weight (lbs), Footprint (sq. ft.), Engine Displacement, Horsepower (HP), etc…), Transmission / Drivetrain Features (ex : Cylinders in Gasoline ICE Vehicles, Drivetrain - Front, Transmission - Manual, 4 or Fewer Gears, etc…) , and Fuel Delivery Type and Powertrain Types (ex : Fuel Delivery - Diesel, Powertrain - Diesel, etc…)

3.3 Data Availability

The data can be accessed through the hyperlink given on the source through CSV download. There were no secondary datasets used, the primary dataset focused thoroughly on the vehicle's characteristics, vehicle performance, fuel efficiency , and carbon emissions.

4. Literature Review

The **Energy Policy, 146, 111783 (Greene et al., 2020)** study offers a historical evaluation of the U.S. Corporate Average Fuel Economy (CAFE) and Greenhouse Gas (GHG) standards since 1975, utilizing EPA data to quantify impacts, attributing a significant portion of fuel economy gains to regulations. Its longitudinal trend analysis directly inspires the project's use of the EPA data to visualize historical patterns in fuel economy and CO2 emissions, aligning with Sustainable Development Goal (SDG) 7 (clean energy transition) and SDG 13 (emission avoidance). The study's limitations include a lack of interactivity, multi-faceted overlays (e.g., combining CO2 with technology adoption timelines), and exclusion of explicit SDG framing or global policy comparisons. The project will address these by incorporating SDG-aligned questions and advanced, dynamic visualizations up to preliminary 2024 data.

The **Proceedings of the National Academy of Sciences, 120(14). (DiStefano and Zeitler, 2023)** case study assesses the influence of National Academies reports (2002–2021) on EPA/NHTSA regulations, using EPA Automotive Trends data for fleet-average fuel economy timelines and showing implementation rates via static bar charts and tables. The study's recommendation-tracking framework inspires the project's development of SDG-aligned policy questions, using EPA data visualizations to evaluate progress toward cleaner technologies (SDG 7) and urgent climate action (SDG 13). The primary gaps are static visualizations, no integration of recent 2024 data, and the absence of SDG metrics. The project will mitigate these by extending the analysis to preliminary 2024 figures and creating dynamic, SDG-explicit visualizations.

The **Harvard Law School (Greenstone et al., 2020)** paper critiques the inefficiencies of the 2012 National Program, proposing a cap-and-trade system and using extensive EPA compliance data to depict line graphs of projected versus actual Miles Per Gallon (MPG)/CO2, scatter plots of lifetime consumption, and market share trends. The study's detailed EPA-derived visualizations of projections and loopholes inspire data exploration techniques for plotting CO_2 reductions against technology adoption, supporting SDG 13 targets. The limitations are descriptive and non-interactive figures, a focus on policy critique without SDG integration, and a lack of global context for trading visualizations. The project will utilize the full 1975-preliminary 2024 dataset for interactive visualizations and incorporate SDG-specific brainstorming.

The **Intergovernmental Panel on Climate Change (IPCC, 2022)** reviewed links to SDG 7 and SDG 13, using EPA-aligned U.S. data within broader analyses for visualizations like bar charts comparing modes and scenario projections. The chapter's use of SDG-metric tables and emission comparisons inspires the project's objective to use EPA data to quantify U.S. progress on SDG 7 and SDG 13 urgency within a global context. The gaps include global aggregation of EPA data with a lack of granular historical trends or interactive tools, and insufficient detail on regulatory specifics. The project will address these by focusing on the EPA's 50-year dataset to deliver detailed, U.S.-centric visualizations with SDG-explicit overlays.

5. METHODOLOGY

For our visualization to be accurate, the researchers employed a straightforward methodology to process the dataset into more reliable data to visualize. The researchers first identified what columns to be used; then prepared the data by formatting, imputation, duplicate checking, and more; next is conducting exploratory data analysis to identify characteristics and patterns that may help in the creation of a visualization; finally is the visualization of the data where the researchers used what they learned in the previous steps of the process and create an accurate and informative visualization.

5.1 Data Set

World Development Indicators (WDI) to Use:

- Average CO_2 Emissions per Mile (g/mi)
    - [Real-World CO_2 (g/mi)]
- Average Vehicle Fuel Economy (MPG)
    - [Real-World MPG]
- Average Vehicle Weight (lbs)
    - [Weight (lbs)]
- Average Engine Horsepower (HP)
    - [Horsepower (HP)]
- Annual Change in Fleet Emissions (%)
    - [Real-World CO_2 (g/mi) across diff. Model Year]
- Urban vs. Highway Emissions Disparity
    - [Real-World CO2_City (g/mi) and Real-World CO2_Hwy (g/mi)]

Other important columns for visualization:

Manufacturer, Model Year, Regulatory Class and Vehicle Type, Production (000), Powertrain Types, Fuel Delivery Types, Footprint Class

5.1.1 Relevance of Chosen Indicators

- Average CO_2 Emissions per Mile (g/mi)
    - The indicator measures the mass of greenhouse gases emitted for every mile driven. It serves as the primary metric for SDG 13 as it tracks the direct impact of vehicle operation on global warming. Lowering the value is essential for meeting international climate targets.
- Average Vehicle Fuel Economy (MPG)
    - MPG calculates the distance a vehicle travels per unit of fuel consumed. It aligns with SDG 7 by measuring energy efficiency and the reduction of fossil fuel waste. Gains in MPG indicate progress toward more affordable and clean energy usage in transport.
- Average Vehicle Weight (lbs)
    - Tracking vehicle mass is necessary to identify if engineering gains are being lost to the production of larger vehicles. Increasing weight typically requires more energy to move which can negate improvements in engine efficiency, revealing the trade-offs between vehicle size and environmental goals.
- Average Engine Horsepower (HP)
    - Horsepower monitors the trend toward higher performance and acceleration in the consumer market. Excessive power often leads to higher fuel consumption and prevents maximum CO_2 reductions. It helps explain why technological advances do not always result in better fleet-wide MPG.
- Annual Change in Fleet Emissions (%)
    - The percentage tracks the rate of improvement or decline in total vehicle pollution from one year to the next. It shows whether current regulations are moving fast enough to reach climate milestones. It provides a clear view of the momentum behind decarbonization efforts.
- Urban vs. Highway Emissions Disparity
    - Engines operate differently in stop-and-go city traffic compared to steady high-speed driving. Comparing these two values shows the efficiency of hybrid systems and traditional engines under different conditions. The data informs urban planning and infrastructure needs for cleaner air.
- Manufacturer, Model Year, Regulatory Class and Vehicle Type, Production (000), Powertrain Types, Fuel Delivery Types, Footprint Class
    - The columns for Manufacturer, Model Year, and Regulatory Class allow us to sort the data. We need these to see if specific companies or types of cars like trucks are holding back progress. These columns help the researchers identify who needs to improve to help the country meet the sustainability targets.

5.2 Data Preparation

1. **Fix data formatting** - On Model Year, Replace the term “Prelim. 2024” to “2024” on column “Model Year”

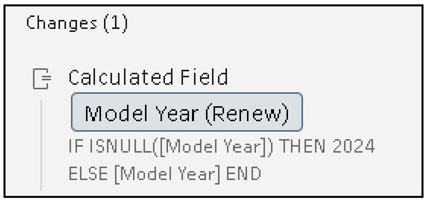

Figure 5.2.1 : Tableau Prep Builder Snippet - 2024 Model Year Label Adjustment

2. **Replace missing or null values** - Replace values of cells that are currently “-”. Replace these values with a simple imputation using 0. An additional attempt was also made to impute the respective mean value of the numerical data column.

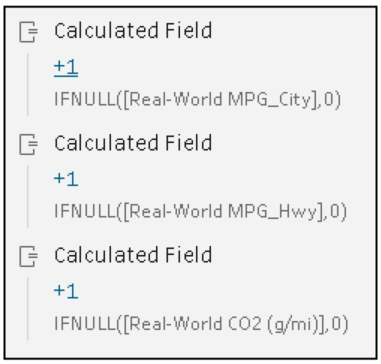

Figure 5.2.2 : Tableau Prep Builder Snippet - Impute 0 Value

3. **Check for Duplicates** - using tableau prep builder remove records that are detected as duplicates. Currently there were no recorded duplicates.

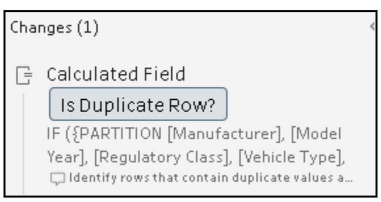

Figure 5.2.3 : Tableau Prep Builder Snippet - Check for Duplicate Records

4. **Removal of columns not significant on the study** - only include the columns mentioned beforehand on World Development Indicators (WDI) to be used and other important columns for analysis and visualization.

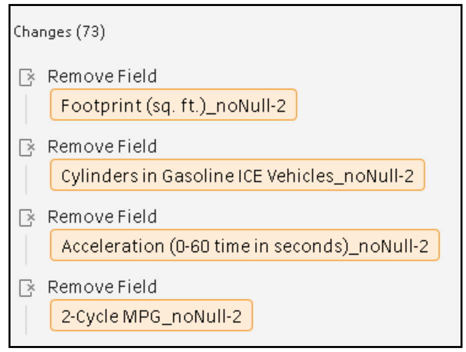

Figure 5.2.4 : Tableau Prep Builder Snippet - Removal of Insignificant Columns

5. **Monitoring outliers** - using Box and Whisker Plot, check the outliers on columns; although we will not pursue replacing nor removing the value on these upper (90% quartile) or lower (10% quartile) extremities.

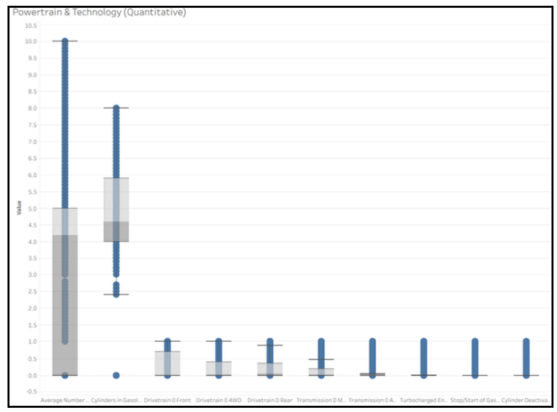

Figure 5.2.5 : Sample of Box and Whisker Plot to Monitor Outliers

6. **Scale the values for standardization** - some columns undergo normalization, scaling values using StandardScaler() where the indicator columns are standardized to have zero mean and a standard deviation of 1.

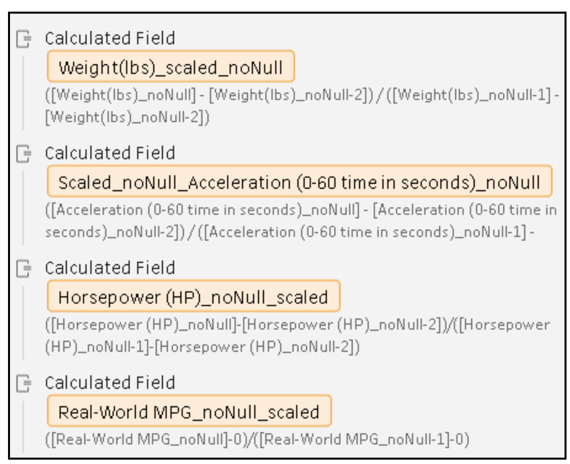

Figure 5.2.6 : Tableau Prep Builder Snippet - Normalization of Values using 0 Imputed Data

7. **Proper column labels** - if any, rename columns that may be difficult to understand, especially those with technical terms.

   Put indicators such as :

   - "noNull" → null to 0
   - "AveImp" → 0 to Average of the numerical column
   - "Scaled" -> values were normalized
   -

```
Real-World CO2 (g/mi)_noNull_scaled
```

Figure 5.2.7 : The Shown CO2 Emissions Column undergoes 0 Imputation and Normalizations

```
Acceleration (0-60 time in seconds)_noNull_AveImp
```

Figure 5.2.8 : The Acceleration Column undergoes 0 Imputation and then Average Imputation

5.3 Exploratory Data Analysis (EDA) Procedures

After cleaning the data, the researchers undergo Exploratory Data Analysis (EDA) which includes procedures to analyze the data attributes, identify outliers, and assess existing relationships between variables.

1. **Descriptive Statistics** - computing for the columns statistical values such as mean, median, mode, mind, max, and standard deviation for numerical data and make frequency counts for categorical data.

2. **Visualize outliers of Columns** - using box and whisker plots (under data preparation too), we can visualize onto what extent values at lower and upper extremities affect the affected statistical values we have.

3. **Correlation of Columns** - use scatter plot to show the relationship between two numerical columns and determine whether the pair has a strong correlation or not.

4. **Distribution Analysis** - on numerical data, use histograms to check for distribution such as skewness, and bar charts for categorical columns to visualize class frequency.

5.4 Data Visualization

Based on the data composition analyzed in the previous step, appropriate visualizations were selected to represent the findings. The charts chosen encompass categorical, time-series, and distribution plots.

1. **Bump Chart**: Shows fuel economy rank of vehicle types over time. Reveals efficiency leaders and laggards. Assesses SDG 7 progress in market segments.

2. **Treemap Chart**: Displays manufacturer and class contribution to the fleet (size) and their CO2 intensity (color). Aggregates high-volume producers and their CO2 cleanliness.

3. **Stacked Bar Chart**: Tracks the rate and mix of cleaner technology adoption over time. Visualizes technological transition. Evaluates SDG 7 and SDG 13 commitment.

4. **Dual Axis Line Chart**: tracks CO2 reduction (SDG 13) against vehicle mass/size trends. Identifies technology masking effects, where heavier vehicles necessitate aggressive CO2 cuts.

6. DATA ANALYSIS

6.1 Analysis and Visualizations in Exploratory Data Analysis (EDA)

	Real-World MPG_noNull	Real-World CO2 (g/mi)_noNull	Weight(lbs)_noNull	Horsepower (HP)_noNull
count	5390.000000	5390.000000	5390.000000	5390.000000
mean	17.719864	327.862479	2969.543213	143.584466
std	13.519341	201.022471	1714.050065	98.723607
min	0.000000	0.000000	0.000000	0.000000
25%	14.072798	277.095230	2529.992000	74.561650
50%	19.866220	382.202660	3534.459000	153.367300
75%	23.602197	457.660565	4128.883750	207.294975
max	129.831470	988.023890	6668.898000	566.456600

Figure 6.1.1 : Descriptive Statistics Using 0 Imputed Data

From Figure 6.1.1, to characterize the distribution of the dataset's continuous attributes, we computed descriptive statistics for columns such as Real-World MPG, CO2 emissions, vehicle weight, and horsepower. The analysis includes measures of central tendency (mean and median) and dispersion (standard deviation), alongside the minimum, quartiles, and maximum to facilitate the identification of spread and potential outliers.

```
Top 5 counts for Regulatory Class:
Regulatory Class
Truck    2499
Car      2156
All       735
Name: count, dtype: int64

Top 5 counts for Vehicle Type:
Vehicle Type
Truck SUV      735
Sedan/Wagon    735
All Truck      735
All            735
All Car        735
Name: count, dtype: int64

Top 5 counts for Manufacturer:
Manufacturer
Ford          392
GM            392
Toyota        392
Stellantis    392
Mazda         392
Name: count, dtype: int64
```

Figure 6.1.2 : Frequency Counts of Categorical Columns

From Figure 6.1.2, frequency counts analysis of the categorical attributes of the Regulatory Class variable shows a slight imbalance, consisting of 2499 Truck and 2156 Car entries. In contrast, Vehicle Type and Manufacturer display a uniform distribution among the top categories, with counts of 735 and 392, respectively. Notably, the appearance of 'All' and 'All Truck' labels indicates that the dataset contains both granular observations and pre-aggregated summaries.

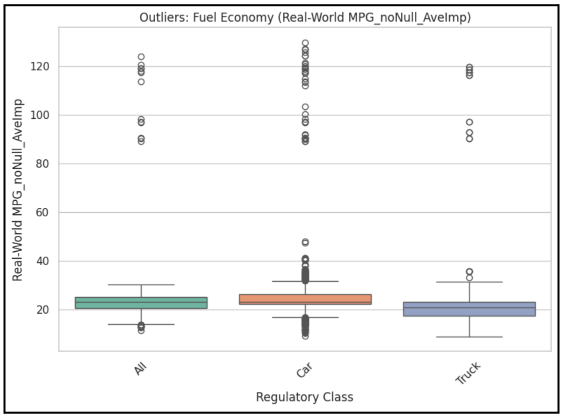

Figure 6.1.3 : Box and Whisker Plot forMPG Using 0 Imputed Data

The most obvious finding in Figure 6.1.3 is the dots at the very bottom of the charts at the zero line. These are the null values we identified earlier and they squash the rest of the box plot making it hard to read. Apart from the errors we can see that the "Car" category has a tighter box meaning their efficiency is more consistent.

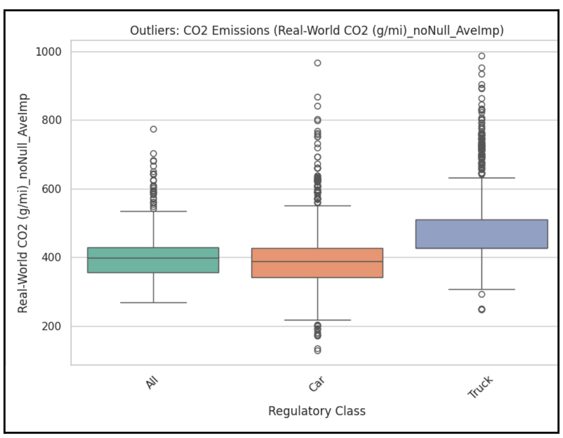

Figure 6.1.4 : Box and Whisker Plot CO2 Emissions Using Average Imputed Data

The "Truck" category has a taller box for CO2 emissions which means there is a lot of variation in how much they pollute. Some trucks are okay but some are extreme polluters. The visualization suggests that regulations for SDG 13 might need to target the worst performing trucks specifically rather than just the average since the spread is so wide.

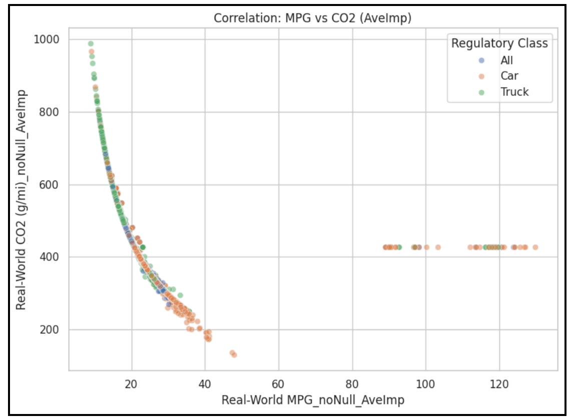

Figure 6.1.5 : Correlation between MPG and CO2 Emissions Using Average Imputed Data

Figure 6.1.5 is a smooth curve with no distractions. The visualization confirms a very strong mathematical relationship where improving MPG is the most direct way to reduce CO2 emissions for SDG 13. There are no outliers drifting far from the curve which means the physics is consistent across all manufacturers.

Figure 6.1.6 : Correlation between Weight(lbs) and Horsepower Using Average Imputed Data

Figure 6.1.6 shows a dense cloud of green dots for trucks in the upper right corner. The area represents heavy and high power vehicles. The fact that so many vehicles are clustered in the high resource zone confirms that the industry trend is moving away from the lightweight and low power ideal that would be best for sustainability.

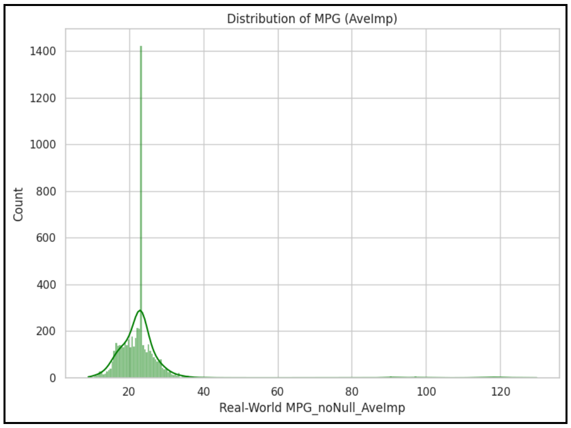

Figure 6.1.7 : Distribution Analysis of MPG Using Average Imputed Data

Figure 6.1.7 curve is skewed to the right which means most cars have low to medium fuel economy and only a few have very high MPG. The skewness is a barrier to SDG 7 because it shows that high efficiency vehicles are still the exception rather than the norm.

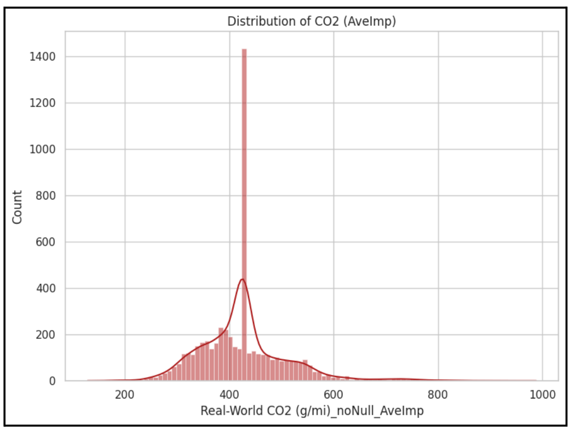

Figure 6.1.8 : Distribution Analysis of CO2 Emissions Using Average Imputed Data

Figure 6.1.8 curve is more centered but it is quite wide. The smooth red line helps us see that the average emission is around 400 grams per mile. The baseline number is what we need to lower to meet the climate action goals and now that the data is clean we can trust the average.

6.2 Planned Visualizations - Analysis and using 0 Imputed Data

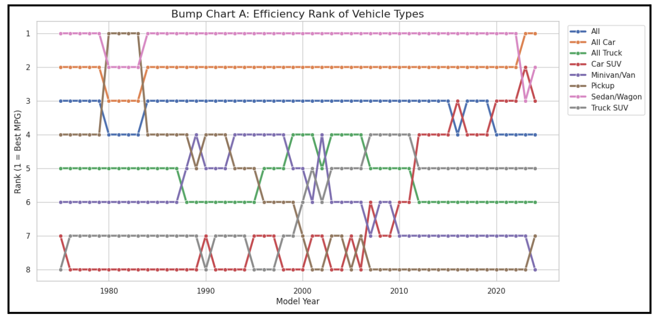

Figure 6.2.1 : Bump Chart on Efficiency Ranks of Vehicle Types from 1975 to 2024

Figure 6.2.1 tracks the ranking of different vehicle types based on their fuel economy performance over the years. We can see that the Sedan/Wagon category usually stays at the top rank which means it is the most efficient type of vehicle. The Truck SUV and Pickup categories are consistently at the bottom ranks like 7 or 8. The visualization answers the research question about patterns because it proves that heavier vehicles like trucks are a persistent barrier to SDG 7. Even if technology gets better the trucks represent a group that slows down the whole fleet efficiency because they are naturally less efficient than cars.

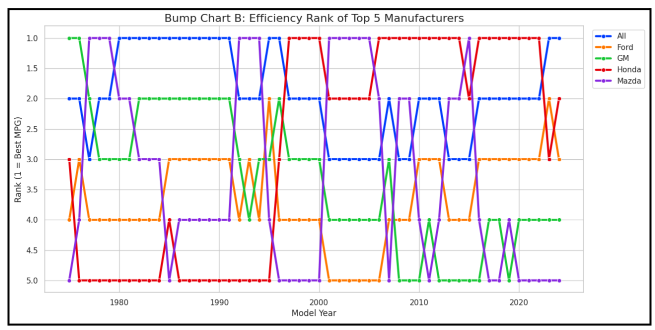

Figure 6.2.2 : Bump Chart of Top 5 Manufacturers in Efficiency Ranks from 1975 to 2024

Figure 6.2.2 compares the top 5 manufacturers to see who is the most efficient in the market. Honda and Mazda often fight for the top spot while companies like Ford and GM are usually in the lower ranks around 4 or 5. The visualization relates to the project objective of identifying which companies need to improve to meet sustainability goals. It seems foreign manufacturers prioritize fuel economy more which aligns better with SDG 13 for reducing emissions compared to American brands that focus on power and size.

Figure 6.2.3 : Treemap Chart of Market Share and CO2 Emissions by Manufacturer in 2023

Figure 6.2.3 shows the size of the box is the number of cars produced and the color is the CO2 emissions. We see that GM and

Stellantis have very large boxes that are dark red which means they produce many cars and those cars have high pollution intensity. Toyota has a big box too but it is lighter meaning they are a bit cleaner. The finding highlights the problem that the biggest producers are also the biggest polluters which makes it hard to achieve the climate action goals of SDG 13.

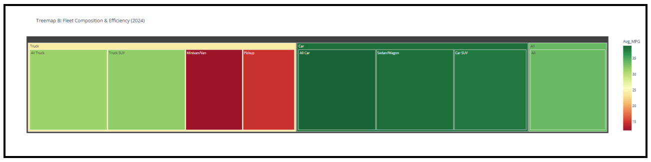

Figure 6.2.4 : Treemap of Fleet Composition and Efficiency in 2023

Figure 6.2.4 visualization groups the fleet by Truck vs Car to show us the efficiency difference. The Car box is dark green which means high MPG but the Truck box is light green or yellow which means lower MPG. The visualization confirms that the shift to trucks mentioned in the problem statement is bad for efficiency. The trend implies that as long as people keep buying trucks the average fuel economy will struggle to go up, a clear difference between vehicle classes that affects the national energy targets.

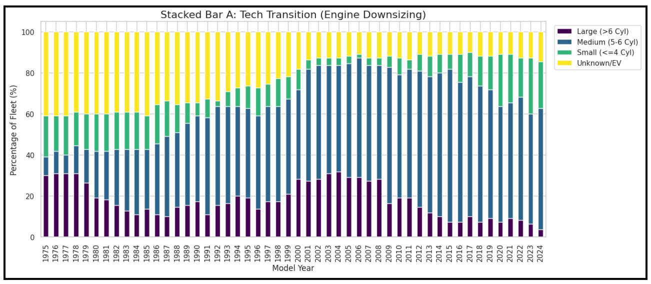

Figure 6.2.5 : Stacked Bar Chart of Tech Transition or Engine Downsizing from 1975 to 2024

Figure 6.2.5 graph tracks the engine size over the years. We can see a clear trend where the large engines with more than 6 cylinders which are the purple bars are disappearing. They are being replaced by small engines with 4 or fewer cylinders shown in green. The technological progression that supports SDG 7 because smaller engines generally use less fuel. However we have to be careful because manufacturers might just be adding turbos to make them powerful anyway which uses more gas.

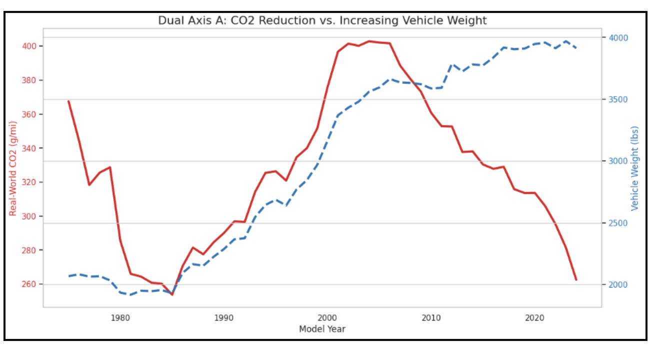

Figure 6.2.6 : Dual Axis Line Chart of CO2 Emissions and Vehicle Weight From 1975 to 2024

Figure 6.2.6 red line shows the CO2 emissions going down and the blue dashed line shows the vehicle weight going up. Around the year 2000 the weight started shooting up while the CO2 reduction slowed down. The visualization suggests that the engineering gains are being cancelled out by the weight increase, proves the masking effect mentioned in the proposal. If cars were not getting so heavy we would probably meet the SDG 13 targets much faster because lighter cars need less energy to move.

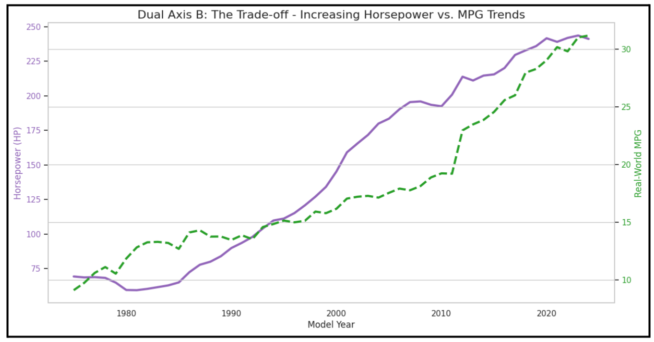

Figure 6.2.7 : Dual Axis Line Chart of Horsepower and MPG from 1975 to 2024

In Figure 6.2.7, we see Horsepower in purple and MPG in green plotted together. The horsepower has increased massively since 1980 almost doubling in value. The MPG has increased too but much slower than the power. The visualization shows a trade off where manufacturers use technology to make cars faster rather than just making them efficient. The technology prioritizes consumer performance desires over the sustainable energy goals of SDG 7 which is a big challenge for policymakers.

6.3 Planned Visualizations - Analysis using Average Imputed Data

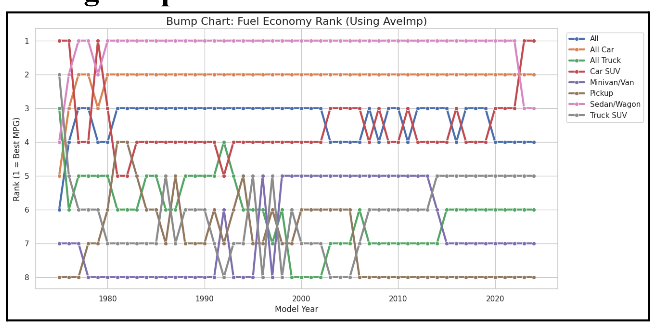

Figure 6.3.1 Bump Chart of Fuel Economy Ranks Using Average Imputed Data

The rankings in Figure 6.3.2 are consistent with the 0 Imputed Data. The All Car category remains higher ranked than All Truck. An unexpected finding here is how volatile the middle ranks are especially for Minivans. The Minivan class jumps up and down a lot which suggests that their efficiency depends heavily on specific model releases that year. For policymakers it implies that they need consistent regulations for all types of vehicles, not just the popular ones to ensure stability in achieving SDG 7.

Figure 6.3.2 : Treemap Chart of Market Share and CO2 Emissions by Manufacturer in 2024 Using Average Imputed Data

Figure 6.3.2 uses the imputed data to fix missing values but it looks like there is a weird error with Tesla. The box for Tesla is red which means high emissions but electric cars should be zero so the imputation might be wrong there. Ignoring that outlier we still see that the Truck sections for Ford and GM are very large and orange. The visualization reinforces that the US market relies heavily on high emission vehicles. The visualization is good for identifying the biggest contributors to pollution for SDG 13 but the data cleaning limitation on EVs needs to be fixed.

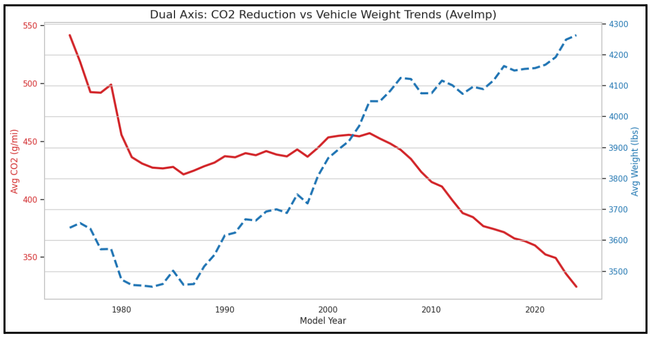

Figure 6.3.3 Dual Axis Line Chart of CO2 Emissions and Vehicle Weight from 1975 to 2024 Using Average Imputed Data

Figure 6.3.3 uses the imputed data to smooth out the lines so we can see the trend better. The trend is even clearer here than in the initial one. The blue line for weight is at the highest point in history right now in 2024. The red line for CO2 is going down but it is flattening out at the end which is a worrying sign for global development because it means our progress is stalling. We cannot achieve clean energy targets if vehicles keep getting heavier and heavier.

The comprehensive analysis of the generated visualizations reveals a conflicting trend where technological advancements are being overshadowed by market shifts towards larger and more powerful vehicles. The bump charts consistently show that sedans and wagons remain the most efficient vehicle types while truck SUVs and pickups are stuck at the bottom rankings which proves that the structural design of a vehicle is a major limit to its fuel economy performance . The ranking aligns with the treemaps which visualize that the manufacturers producing the largest volume of vehicles like GM and Stellantis are also responsible for the highest intensity of CO2 emissions because their fleets are dominated by the inefficient truck category . The stacked bar chart regarding fleet mix clarifies the historical progression by showing that the market share of efficient sedans has shrunk massively while the share of truck SUVs has expanded which directly opposes the affordable and clean energy targets of SDG 7 . The negative trend is explained by the dual axis charts where we see that the potential CO2 reductions from better engines are being cancelled out because vehicles are hitting their highest historical weight in 2024 and horsepower is rising much faster than MPG which creates a masking effect that slows down the urgent climate action needed for SDG 13

7. CONCLUSION

7.1 Key Findings

The research utilized historical EPA data to analyze the trajectory of the U.S. automotive sector and its alignment with global sustainability goals. The visualizations revealed a clear conflict between engineering progress and consumer market trends. While internal combustion engines have become more efficient technologically, the dual-axis analysis demonstrated that these gains are largely negated by a consistent increase in vehicle weight and horsepower since the year 1975. The bump charts and treemaps confirmed that the market has shifted significantly from efficient sedans to heavier truck SUVs which results in a stagnating fleet-wide fuel economy . The trend reveals a critical barrier to achieving Sustainable Development Goal 7 regarding energy efficiency because the widespread adoption of heavier vehicles requires more energy to operate regardless of engine innovation. Furthermore, the persistence of high emission intensities in major manufacturers hinders the urgent decarbonization required for Sustainable Development Goal 13, as the reduction in tailpipe emissions is too slow to meet climate action targets.

7.2 Limitations of the Study

The validity of these findings is subject to specific data and methodological limitations encountered during the study. The primary dataset contained significant missing values for recent model years and specific technologies which required the use of average imputation methods that may smooth out extreme variations in real-world performance. Regarding the geographical scope, the analysis was restricted to the United States market. However, the study serves as a critical case study for global application, as the lessons learned from U.S. regulatory challenges and market behaviors can be adapted by other nations to design more resilient policies that avoid the pitfalls of vehicle upsizing. Additionally, the study focused exclusively on Carbon Dioxide emissions and did not account for other harmful pollutants like Nitrogen Oxide or Sulfur Oxide which are also relevant to the broader environmental impact of the transport sector. The aggregation of vehicle types into broad categories such as "Car SUV" or "Truck" also masked the performance of specific car models which limits the granularity of the insights.

7.3 Policy Recommendations

Based on the identified trends, policymakers should consider implementing regulations that specifically target vehicle weight and size rather than focusing solely on tailpipe emissions. The strong correlation between mass and fuel consumption suggests that disincentivizing the production of excessively heavy passenger vehicles is necessary to accelerate progress toward international climate commitments.

7.4 Research Recommendations

Future research should expand on the work by incorporating machine learning algorithms to better impute missing data and predict future emission trends with higher accuracy. Researchers could also broaden the scope to include global datasets to see if the American preference for high horsepower is a unique anomaly or a worldwide pattern. Finally, including economic indicators such as fuel prices or inflation rates in future visualizations could help explain the consumer behavior shifts that drive the demand for inefficient truck SUVs.

8. REFERENCES

[1] J. G. DiStefano and E. L. Zeitler. 2023. Implementation of National Academies' study recommendations: A fuel economy case study. Proceedings of the National Academy of Sciences 120, 14 (2023). https://doi.org/10.1073/pnas.2303242120

[2] D. L. Greene, J. M. Greenwald, and R. E. Ciez. 2020. U.S. fuel economy and greenhouse gas standards: What have they achieved and what have we learned? Energy Policy 146 (2020), 111783. https://doi.org/10.1016/j.enpol.2020.111783

[3] M. Greenstone, C. R. Sunstein, and S. Ori. 2020. Fuel economy 2.0. Harvard Public Law Working Paper No. 17-27. Harvard Law School. Retrieved from https://journals.law.harvard.edu/elr/wp-content/uploads/sites/79/2020/04/44.1-Greenstone-Sunstein-Ori.pdf

[4] IPCC. 2022. Chapter 10: Transport. Retrieved from https://www.ipcc.ch/report/ar6/wg3/chapter/chapter-10/