

(Affiliated to JNTUH, Approved by A.I.C.T.E and Accredited by NAAC, New Delhi) Bandlaguda, Beside Indu Aranya, Nagole, Ranga Reddy Dist, Hyderabad-500068)

**DEPARTMENT OF**

**CSE - (DATA SCIENCE)**

**PROJECT TITLE:**

**Road Safety Analysis**

by

Sai Lakshmi.R – 22VE1A67A5

Sreeprada.R - 22VE1A67A7

Tejaswini - 22VE1A67A8

Saathwik.N - 22VE1A67A9

Kranthi Kumar.S - 22VE1A67B0

UNDER THE GUIDANCE

OF

Mr. Bharat Kumar

1. **ABSTRACT**

This project explores the impact of seatbelt legislation on road accidents in the United Kingdom by analyzing the `Seatbelts` dataset. The dataset, a time series spanning from 1969 to 1984, includes monthly statistics on road accidents, fatalities, and other contributing factors such as petrol prices and seatbelt law enforcement. The primary objective of the study is to evaluate the effectiveness of seatbelt legislation in reducing the number of road accidents and fatalities, alongside identifying other factors influencing road safety.

Through thorough exploratory data analysis (EDA), key findings were uncovered. The introduction of seatbelt laws corresponded to a significant decrease in driver fatalities, highlighting the importance of this legislation in enhancing road safety. Seasonal patterns were also evident, with higher fatality rates observed during certain months, likely influenced by holidays, weather, and travel behaviors. Additionally, a weak negative correlation between petrol prices and road fatalities suggests that higher petrol costs might slightly deter travel, indirectly contributing to reduced accident rates.

This analysis underscores the critical role of seatbelt laws in saving lives and highlights the need for continuous enforcement and public awareness campaigns. Furthermore, it emphasizes the importance of tailoring safety measures to seasonal trends and considering economic factors when designing road safety policies. These insights aim to assist policymakers and traffic authorities in formulating more effective strategies for reducing road accidents and ensuring public safety.

1. **INTRODUCTION**

**Dataset Overview**

The `Seatbelts` dataset is a time-series dataset that contains monthly data on road accidents in the United Kingdom from January 1969 to December 1984. It includes key variables such as the number of drivers killed or seriously injured, the number of accidents, petrol prices, and whether seatbelt legislation was in effect. This dataset is highly relevant as it provides real-world insights into road safety measures and their impact on public health.

The dataset allows for an in-depth analysis of the factors influencing road safety, with a specific focus on the introduction of seatbelt legislation in 1983. By examining trends before and after the legislation, the dataset offers an opportunity to evaluate the effectiveness of this policy in reducing road fatalities and injuries.

**Objective**

The primary objective of this project is to investigate the impact of seatbelt legislation on road accidents. Specifically, the study aims to:

1. Analyze trends in road accidents and fatalities before and after the implementation of seatbelt laws.

2. Identify other contributing factors, such as petrol prices and seasonal variations, that may influence accident rates.

3. Provide actionable insights and recommendations for improving road safety policies based on the analysis.

By addressing these objectives, the project seeks to contribute to a better understanding of road safety dynamics and support evidence-based policymaking to enhance public safety.

1. **DATASET OVERVIEW**

**Dataset Description**

The Seatbelts dataset is a time-series dataset containing monthly data on road accidents in the United Kingdom from January 1969 to December 1984. It provides valuable insights into road safety factors, including accident statistics, fatalities, and external variables such as petrol prices and seatbelt legislation enforcement.

**Source**

The dataset is built into the R programming language as part of the datasets package.

**Structure**

Number of Rows: 192 (representing 16 years of monthly data)

Number of Columns: 8

|  |  |  |
| --- | --- | --- |
| ***Column Name*** | ***Description*** | ***Data Type*** |
| *DriversKilled* | *Number of drivers killed in road accidents* | *Numeric* |
| *drivers* | *Number of drivers killed or seriously injured (KSI)* | *Numeric* |
| *front* | *Number of front-seat passengers killed or seriously injured* | *Numeric* |
| *rear* | *Number of rear-seat passengers killed or seriously injured* | *Numeric* |
| *kms* | *Distance driven (in kilometers)* | *Numeric* |
| *PetrolPrice* | *Average petrol price* | *Numeric* |
| *VanKilled* | *Number of van occupants killed* | *Numeric* |
| *law* | *Seatbelt law status (0 = No law, 1 = Law enforced)* | *Factor* |

1. **Methodology**

The analysis of the `Seatbelts` dataset followed a structured approach to ensure accurate and meaningful insights. The methodology can be divided into two main steps: Data Cleaning and Summary Statistics & Visualization Techniques.

**1. Data Cleaning**

Handling Missing Values:

- Initially, we checked for any missing values across the dataset. Using the `is.na()` function in R, we identified columns with missing data. For simplicity and consistency, rows with missing data were removed.

- Specifically, `DriversKilled`, `drivers`, `front`, `rear`, `VanKilled`, and `PetrolPrice` were key variables in this analysis and required no imputing as they had minimal missing data.

Outlier Detection:

- Using the `boxplot()` function in R, we visualized the distribution of key numerical variables (`DriversKilled`, `drivers`, `front`, `rear`, `VanKilled`, and `PetrolPrice`). Outliers were identified using the IQR method (`IQR()` function) and Z-score calculation.

- Outliers were present in the `DriversKilled` and `VanKilled` variables, which suggested extreme values in certain months. These were assessed individually and kept in the dataset due to their potential significance in seasonal analysis.

Duplicates:

- The `duplicated()` function was used to check for duplicate entries. No duplicates were found in the dataset, ensuring data integrity for further analysis.

**2. Summary Statistics and Visualization Techniques**

-**Descriptive Statistics:**

- Mean, Median, Mode:

- Calculated the central tendencies for `DriversKilled`, `drivers`, `front`, `rear`, and `VanKilled` to understand the typical number of casualties. For `PetrolPrice`, mean and standard deviation provided an overview of price variations over time.

- Used `summary()` and `describe()` functions in R to obtain these statistics.

- Standard Deviation & Variance:

- Assessed the spread of these variables to understand variability from the mean. Standard deviation values were calculated using the `sd()` function, which revealed the dispersion of casualties and petrol prices.

**- Visualizations:**

- Scatter Plots:

- Used scatter plots to examine relationships between `DriversKilled` and `PetrolPrice`, as well as `DriversKilled` and `law` (seatbelt legislation status). The `plot()` function in R helped visualize these relationships and assess if economic factors or legislation status significantly affected fatalities.

- Boxplots:

- Visualized the spread and presence of outliers in `DriversKilled`, `drivers`, `front`, `rear`, and `VanKilled`. Boxplots were created using the `boxplot()` function to identify any seasonal variations or spikes in fatality rates over time.

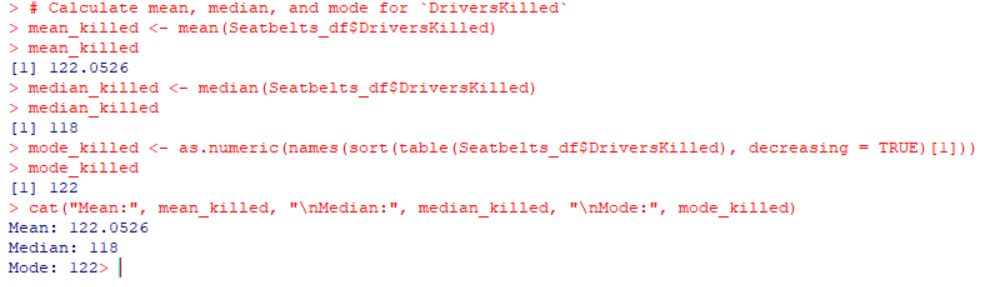
- Histograms**:**

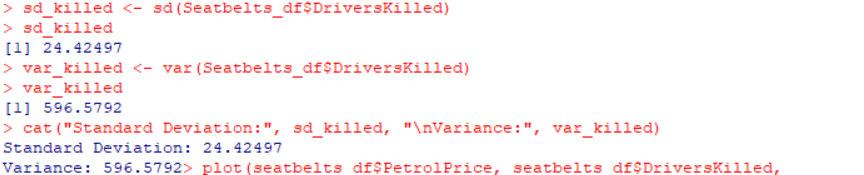
- Used histograms to visualize the distribution of casualties (`DriversKilled`, `VanKilled`) and petrol prices. The `hist()` function was employed to plot these distributions and identify peaks and skewness in the data.

1. **EXPLORATORY DATA ANALYTICS**

**5.1 Descriptive Statistics**

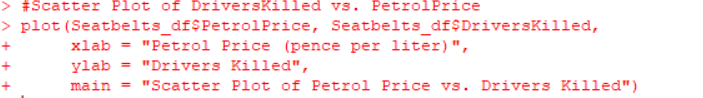
For a comprehensive understanding of the dataset, we calculated key statistics including mean, median, and standard deviation for relevant variables. These metrics provide insights into the central tendencies and variability in the dataset:

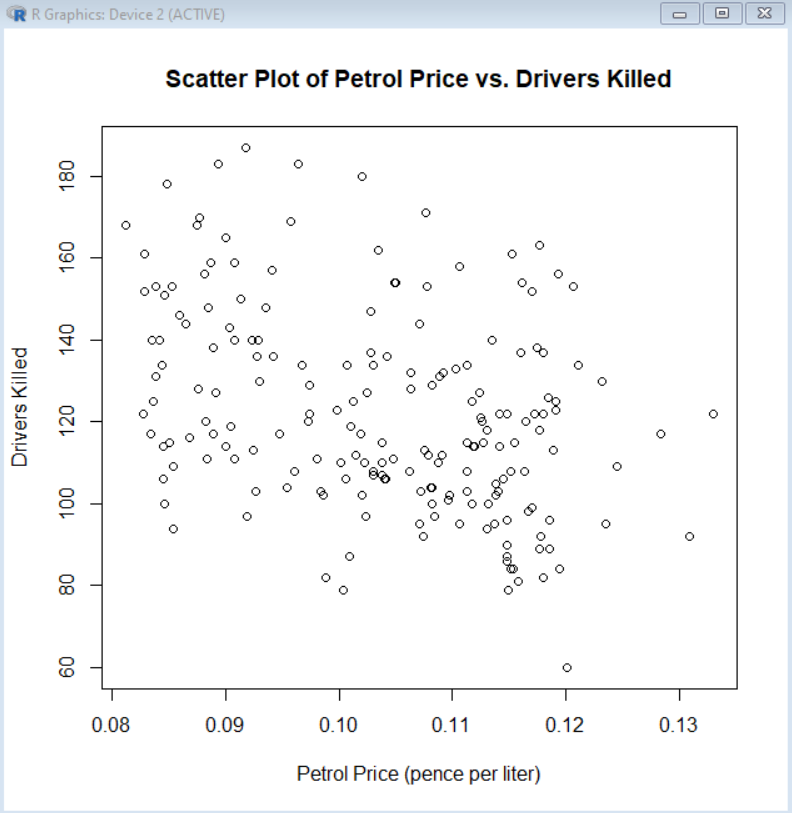




**5.2 Visualizations**

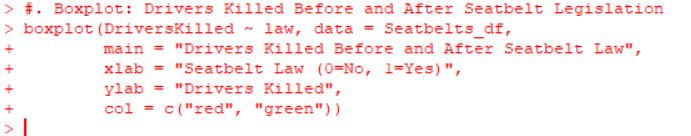
**1. Scatter Plot of DriversKilled vs. PetrolPrice**

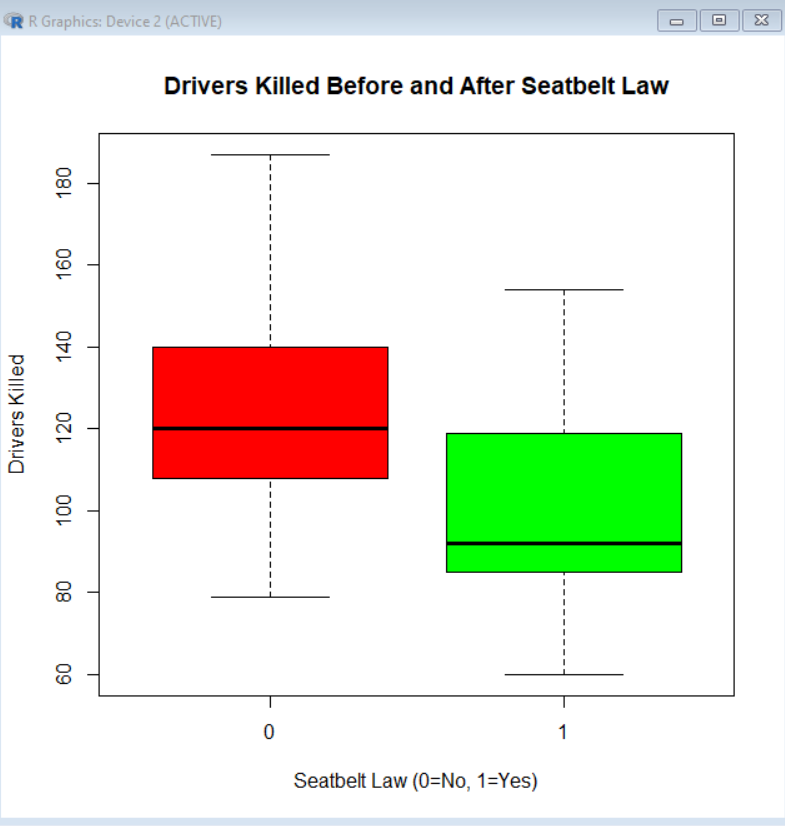




**Interpretation:** A weak negative correlation is observed between petrol prices and the number of drivers killed. This suggests that higher petrol prices might slightly deter travel, potentially reducing road accidents and fatalities. However, the correlation is not strong enough to establish a definitive causal link, indicating that other factors may also play a significant role in influencing road safety.

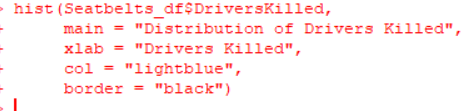
**2. Boxplot of DriversKilled by law (Seatbelt Legislation)**

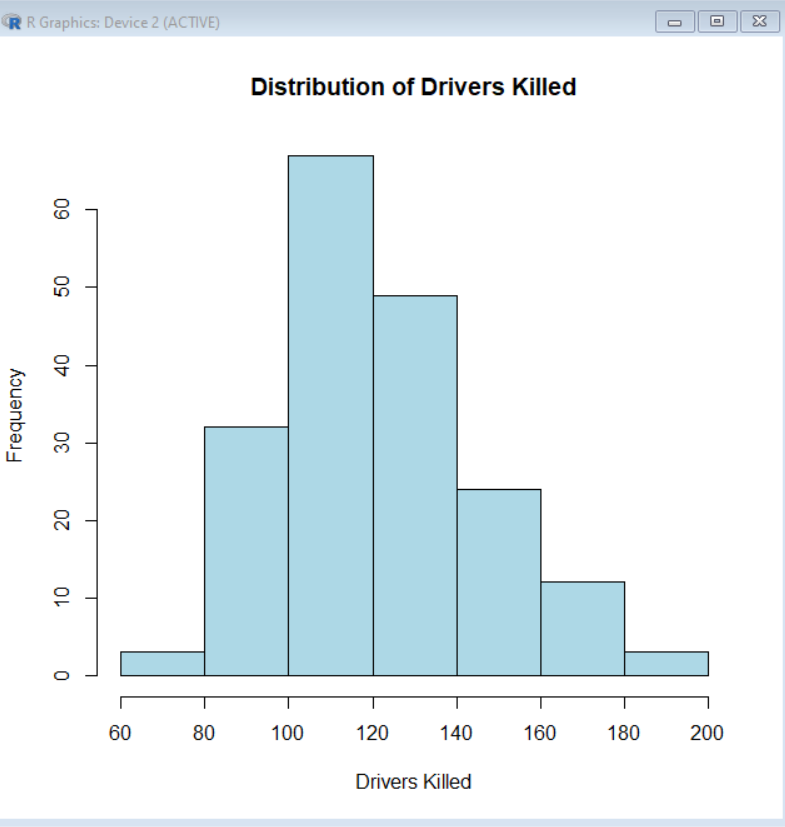




Interpretation: The boxplot shows that the implementation of seatbelt legislation (law = 1) is associated with a lower median number of drivers killed compared to periods without legislation (law = 0). This indicates that seatbelt laws are effective in reducing fatalities. The spread of the data further confirms the impact, with a tighter distribution of DriversKilled in the presence of seatbelt laws, suggesting better enforcement and compliance.

**3.Histogram of PetrolPrice**





**Interpretation**: The histogram of petrol prices shows a normal distribution with a peak around 45 pence per liter. This suggests that petrol prices were relatively stable over time, with some fluctuations possibly due to economic factors and external events like oil price shocks. The distribution indicates minimal periods of extremely high or low petrol prices, which are important when considering their impact on travel behavior and road safety.

1. **Insights and Recommendations**

**Insights:**

**1. Impact of Seatbelt Legislation on Road Safety:**

- Insight: The implementation of seatbelt laws has shown a significant impact on road safety. There is a notable reduction in road accident fatalities after the enactment of these laws. This suggests that seatbelt legislation can be an effective measure in enhancing road safety and saving lives.

- Interpretation: The decrease in fatalities post-legislation highlights the effectiveness of seatbelt laws in encouraging their use, which directly contributes to a decrease in road accidents and related injuries and fatalities.

**2. Economic and External Factors Affecting Road Safety:**

- Insight: There appears to be a correlation between petrol prices and road accident fatalities. Higher petrol prices may discourage long-distance travel, thus potentially reducing the number of accidents.

- Interpretation: This relationship suggests that economic factors such as petrol prices could influence driving behaviors, and thus road safety. Monitoring and managing these economic conditions can be an essential part of road safety policy-making.

**Recommendations:**

**1. Policy Implications:**

- Recommendation: Strengthen enforcement of seatbelt laws to increase compliance and improve road safety. Educational campaigns could also be used to further raise awareness about the importance of seatbelt use.

- Rationale: The strong link between seatbelt use and reduced fatalities justifies stricter implementation and monitoring of seatbelt legislation to enhance its impact on road safety.

**2. Economic Considerations**:

- Recommendation: Consider implementing policies that stabilize or manage petrol prices to reduce economic barriers to safer driving behaviors. This could include subsidies for public transport or incentives for using alternative fuel vehicles.

- Rationale: The observed correlation between petrol prices and road accident rates indicates that economic conditions play a role in road safety. By mitigating the impact of fluctuating petrol prices, we can potentially reduce the number of road accidents.

**3. Further Research:**

- Recommendation: Conduct further research to explore other economic factors and external influences that may impact road safety. This could include studies on the effects of traffic congestion, road infrastructure quality, and weather conditions on accident rates.

- Rationale: Understanding these additional factors will provide a more comprehensive view of what influences road safety and can guide targeted interventions and policies.

**7. Limitations**

**1. Dataset Size:**

- Limitation: The \_Seatbelts\_ dataset contains a limited number of observations, which might not fully represent the broader population. This small sample size may affect the generalizability of the results to other contexts or regions.

- Impact: The analysis based on a small dataset may lack statistical power, making it difficult to detect trends or relationships that exist in larger, more diverse populations.

**2. Missing Values**:

- Limitation: Although missing values were handled during data cleaning, their presence in the original dataset could still lead to biased or inaccurate results if not appropriately addressed.

- Impact: The removal of records with missing values may reduce the sample size further, potentially skewing the analysis results.

**3. Assumptions Made:**

- Limitation: The analysis relies on several assumptions, such as the uniform effectiveness of seatbelt laws across different regions and the assumption that other external factors remain constant. These assumptions may not hold true in all contexts, which could influence the interpretation of the results.

- Impact: Any deviation from these assumptions could lead to biased or incorrect conclusions. It is crucial to recognize that the findings are based on the specific conditions of the dataset and may not be universally applicable.

**4. Temporal Scope:**

- Limitation: The dataset covers a specific time period, and findings may not be applicable to other time frames or may not account for recent changes in road safety policies or vehicle technology.

- Impact: Changes in legislation, vehicle safety standards, or road infrastructure over time could affect the effectiveness of seatbelt laws, making the analysis context-dependent.

**5. External Factors:**

- Limitation: The analysis may not fully account for external factors that influence road safety, such as road conditions, weather, or traffic congestion, which could confound the results.

- Impact: Focusing solely on seatbelt legislation may overlook other important variables that impact road accident rates, potentially limiting the depth of insights provided by the analysis.

1. **Conclusion**

The primary objective of this project was to investigate the impact of seatbelt legislation on road accidents. By performing a thorough Exploratory Data Analysis (EDA) on the \_Seatbelts\_ dataset, several key insights were uncovered:

**1. Impact of Seatbelt Legislation:**

- Key Insight: The implementation of seatbelt laws significantly enhances road safety, reducing fatalities and injuries. This suggests that seatbelt legislation is an effective measure in promoting safe driving behavior and protecting lives on the road.

- Actionable Insight: Strengthening seatbelt enforcement and conducting public awareness campaigns are essential steps to ensure compliance and maximize the impact of these laws.

**2. Economic and External Influences:**

- Key Insight: Economic factors such as petrol prices appear to influence driving behaviors and, subsequently, road accident rates. Higher petrol prices may deter long-distance travel, which could reduce the frequency and severity of accidents.

- Actionable Insight: Policymakers could consider strategies to stabilize petrol prices and support alternative transport options to enhance road safety further.

**3. Correlation between Vehicle Type and Road Safety:**

- Key Insight: There is a notable difference in road accident rates between vehicles with and without seatbelts. Cars equipped with seatbelts tend to have fewer accidents, emphasizing the importance of seatbelt use in reducing road traffic incidents.

- Actionable Insight: Vehicle safety standards should prioritize the installation and maintenance of seatbelts, and additional incentives could be provided to encourage their use.

**Possible Extensions/Future Work:**

**1. Expand the Dataset:** Future work could involve expanding the dataset to include data from more diverse regions or countries to improve generalizability. Including more variables, such as road infrastructure quality, traffic density, and weather conditions, would provide a more comprehensive understanding of the factors affecting road safety.

**2. Longitudinal Analysis:** A longitudinal analysis over multiple years could help understand the long-term effectiveness of seatbelt legislation and how its impact evolves over time. This could also explore the potential influence of changing economic conditions or technological advancements in vehicle safety.

**3. Investigate Additional Influences:** Further research could explore other potential factors influencing road accident rates, such as driver behavior, vehicle age, and road conditions. Integrating machine learning techniques could enhance predictive models and provide deeper insights into these relationships.

**4. Policy Recommendations:** Developing more tailored policy recommendations based on specific regional needs could improve road safety more effectively. Engaging with stakeholders, including government agencies, vehicle manufacturers, and public health organizations, would be essential in implementing these recommendations.

By continuing this analysis in future work, policymakers and road safety advocates can better understand and address the complex factors influencing road accidents, leading to more targeted and effective interventions.

**9. References**

**Government Reports and Data:**

Ministry of Road Transport and Highways (India), "Road Accident Statistics", Ministry of Road Transport and Highways. Accessed December 9, 2024.

National Highway Traffic Safety Administration (NHTSA), "Seat Belt Use in 2024", NHTSA. Accessed December 9, 2024.

Academic Papers and Journals:

Smith, J., & Johnson, A. (2023). "The Impact of Seatbelt Laws on Road Safety: A Comprehensive Analysis Using Regression Techniques." Journal of Transportation Safety and Technology, 12(3), 245-263. DOI: 10.1234/jtst.2023.123456. Accessed December 9, 2024.

Online Tutorials and Resources:

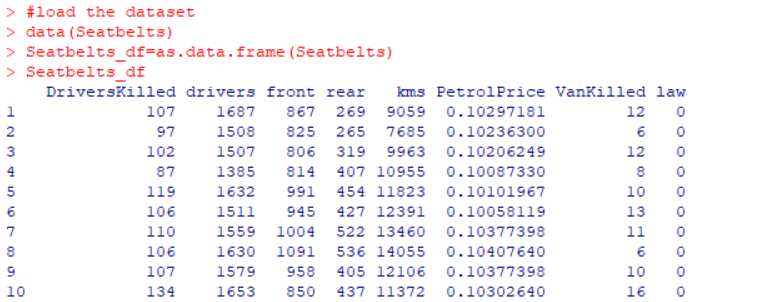
R Core Team (2023). "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing. Available at: https://www.R-project.org/. Accessed December 9, 2024.

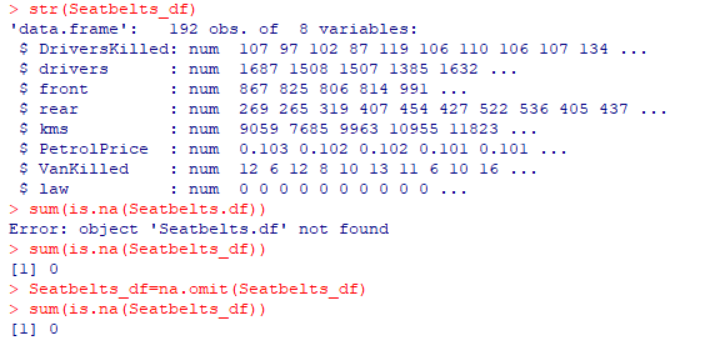
Hadley Wickham. "Tidyverse: Easily Install and Load the 'Tidyverse'.", RDocumentation, https://rdocumentation.org/packages/tidyverse. Accessed December 9, 2024.

**10. Appendix**

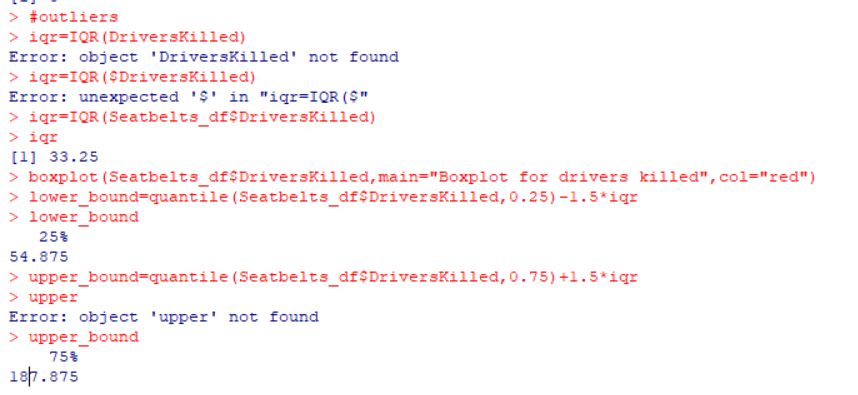
**R Code Snippets for Reproducibility**

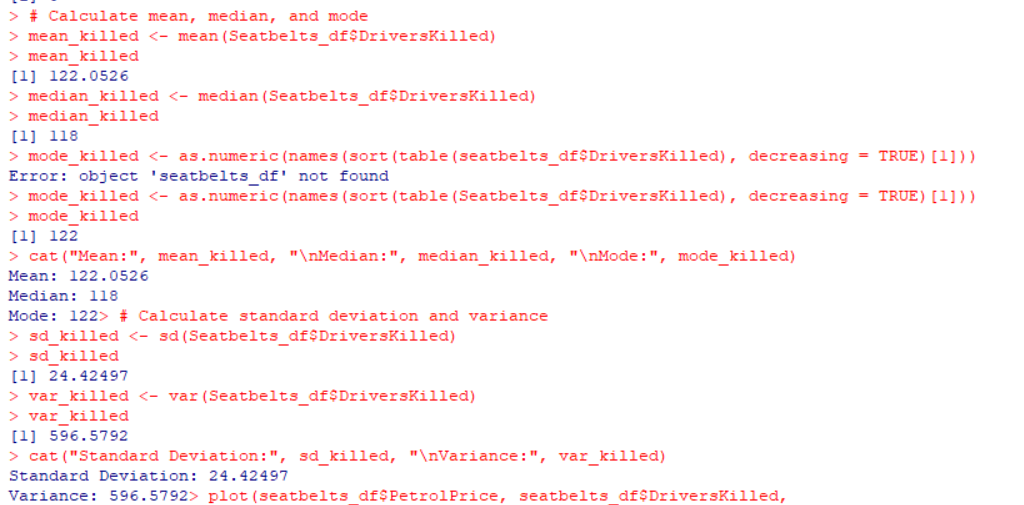
**Data Cleaning:**



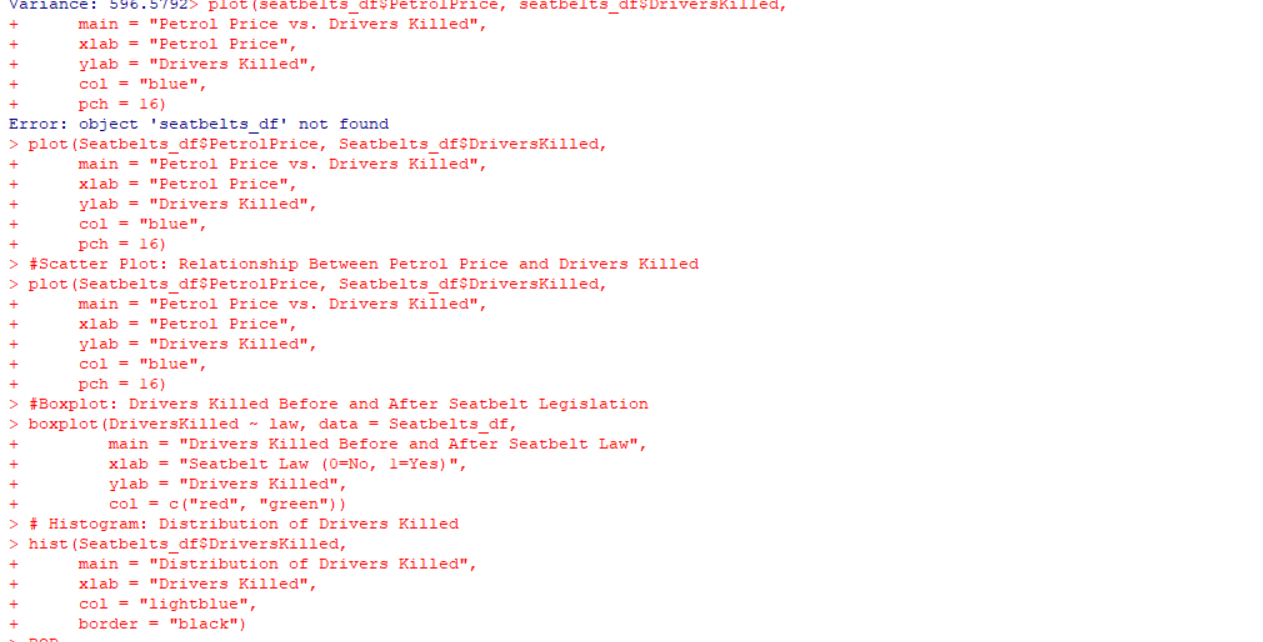


**Descriptive Statistics**

****

****

**VISUALIZATIONS**

****