

Comprehensive Analysis of Fatal Traffic Accidents with Data-Driven Visualizations

TEAM 5

Nidhi Vinodbhai Patel

Gokulraj Muthukumar

Kirubaharan Joseph Abraham

Moumita Baidya



★ PROJECT OVERVIEW

→ Fatal traffic accidents remain a major public safety issue in the U.S., with thousands of lives lost each year due to factors like weather, driver impairment, and distraction. Our project addresses this by building a web-based tool to visualize key patterns and support smarter road safety decisions.

★ GOAL

→ Our goal is to transform complex fatal traffic accident data into accessible, insightful visualizations. By analyzing the 2022 FARS dataset, we aim to uncover key patterns such as when, where, and under what conditions fatal crashes occur. The tool we're developing will support both analysts and safety advocates in making informed, data-driven decisions to improve road safety.

★ DATASET OVERVIEW

- For our project, we used the 2022 FARS dataset from NHTSA, which provides a nationwide view of fatal traffic accidents.
- ACCIDENT.csv – crash time, location, and conditions
- VEHICLE.csv – vehicle type, movement, and damage
- PERSON.csv – info on people involved (age, gender, role)
- RACE.csv – race and ethnicity data
- WEATHER.csv, DAMAGE.csv, SAFETYEQ.csv – environmental and safety context
- VSOE.csv, NMCRASH.csv – special vehicle use and non-motorist cases

★ Methodology

- Data Processing
- Data Cleaning
- Filtering
- Column Selection
- Data Merging
- Aggregation
- Visualization
- Web Integration

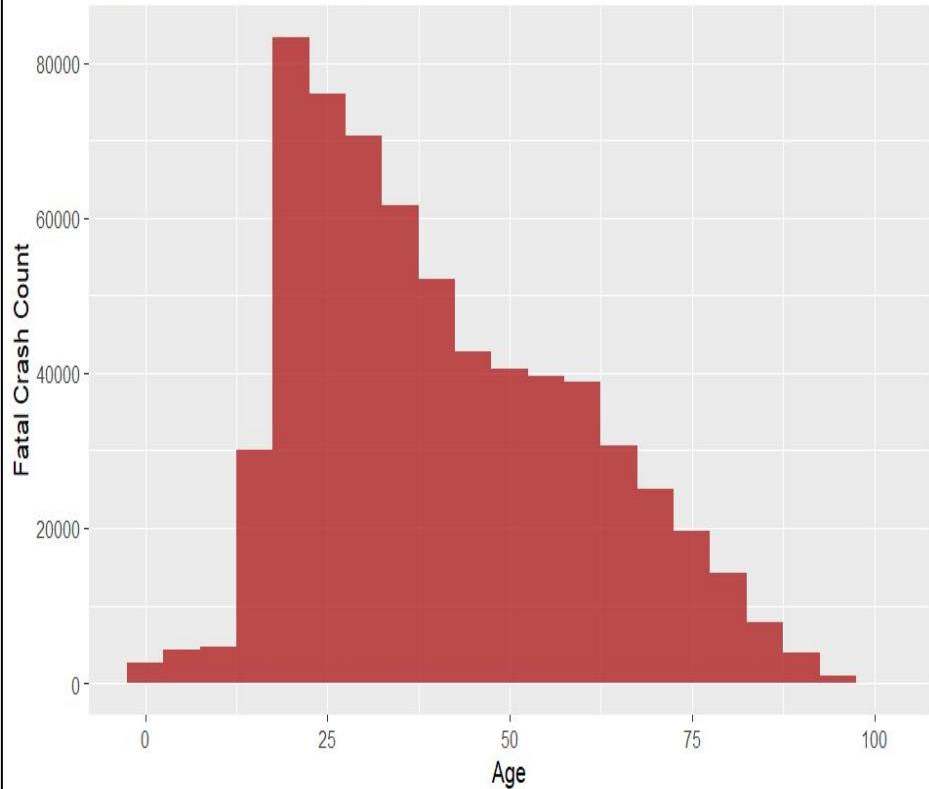
❖ TOOLS AND TECHNOLOGY

- Python/Flask – creating backend API
- HTML/CSS – For designing a simple, user-friendly front end
- R – For generating clear and insightful visualizations

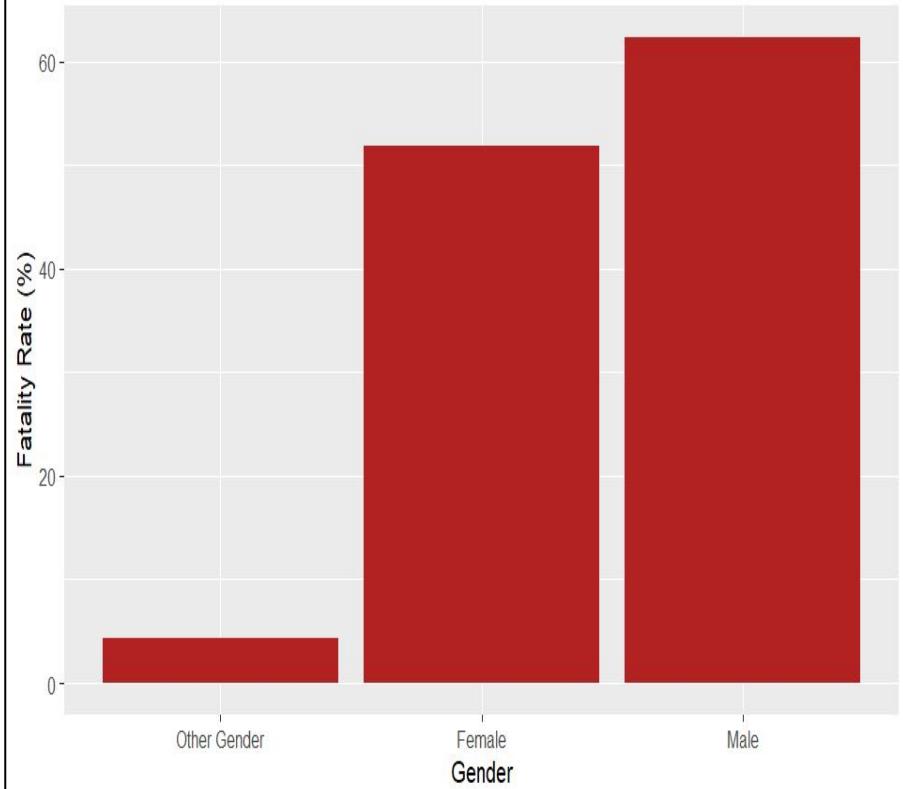


AGE AND GENDER

Age Distribution of Fatal Injuries



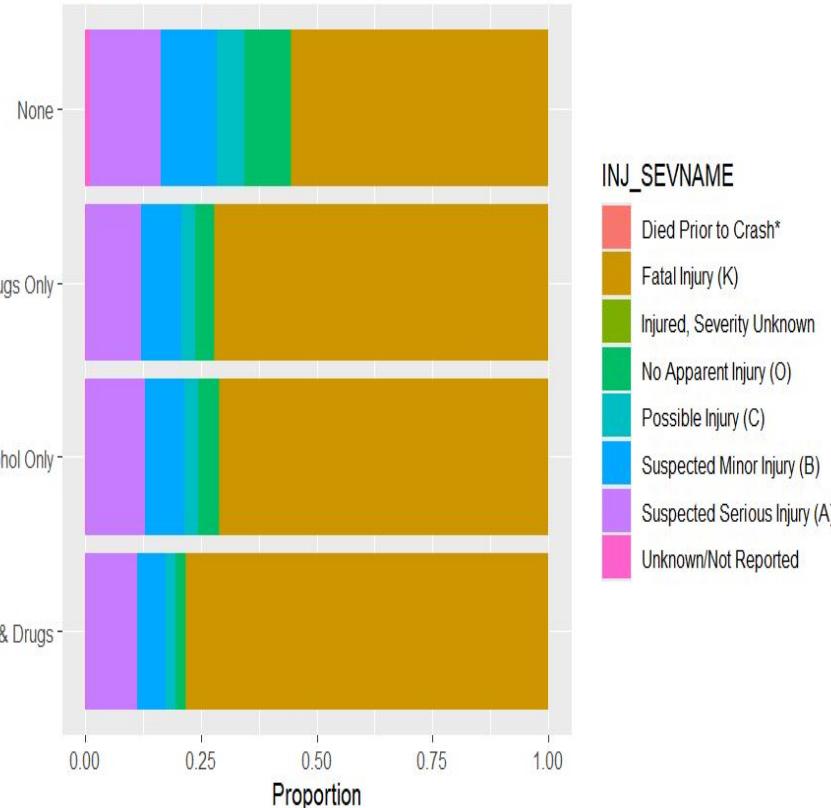
Fatality Rate by Gender



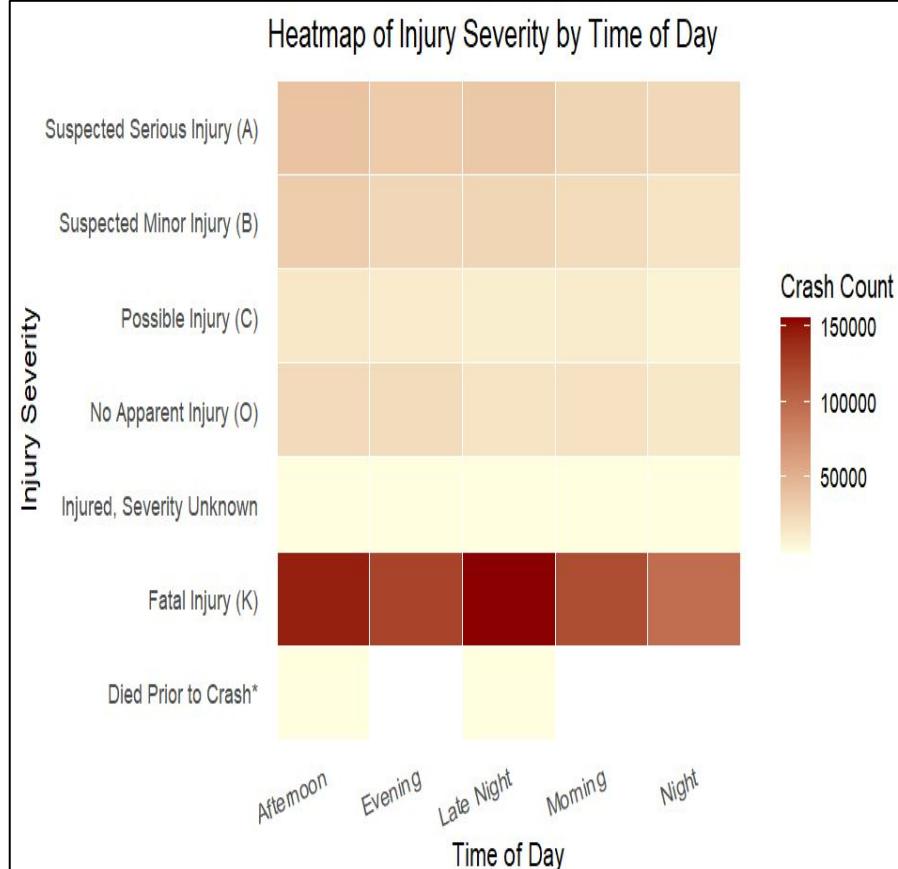


SUBSTANCE AND DAYTIME

Injury Severity by Substance Involvement



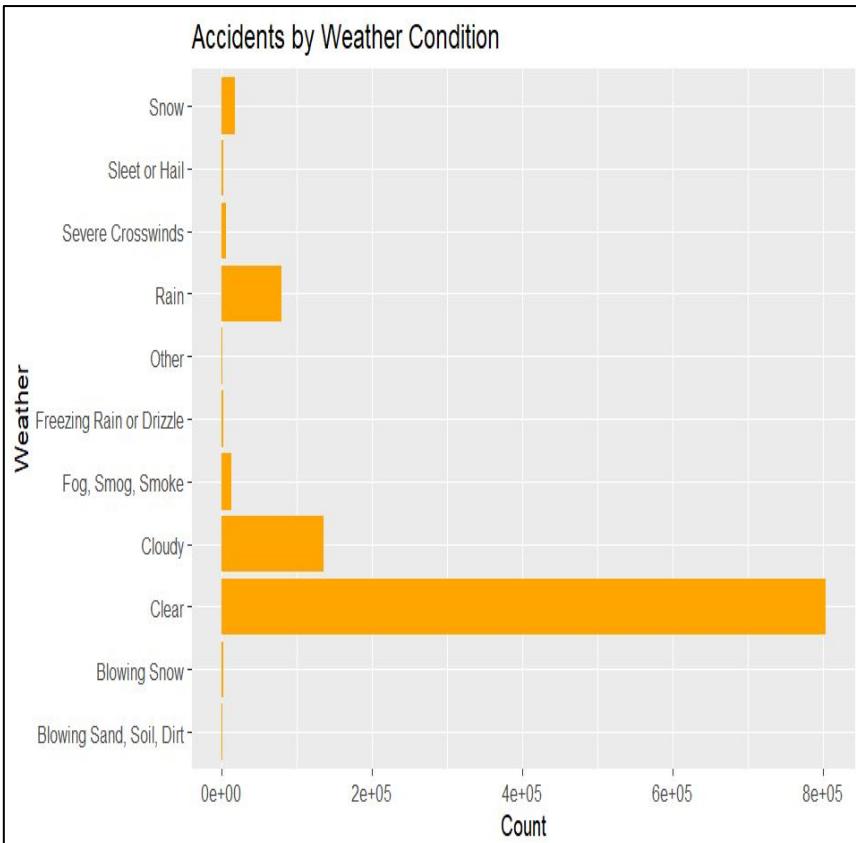
Heatmap of Injury Severity by Time of Day



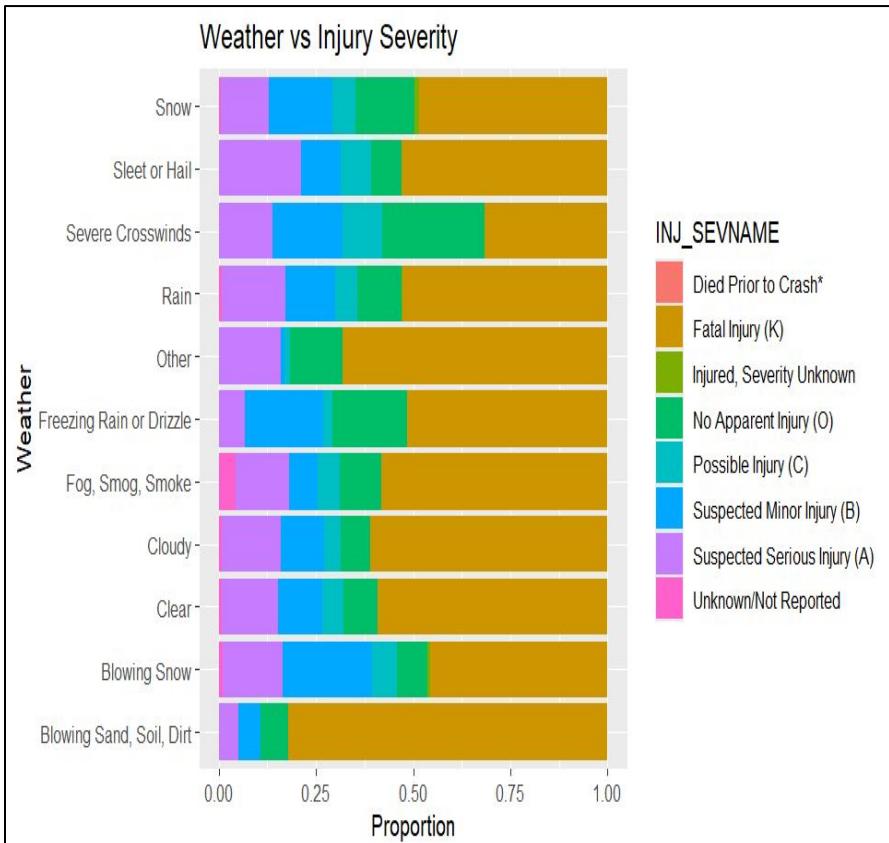


WEATHER

Accidents by Weather Condition



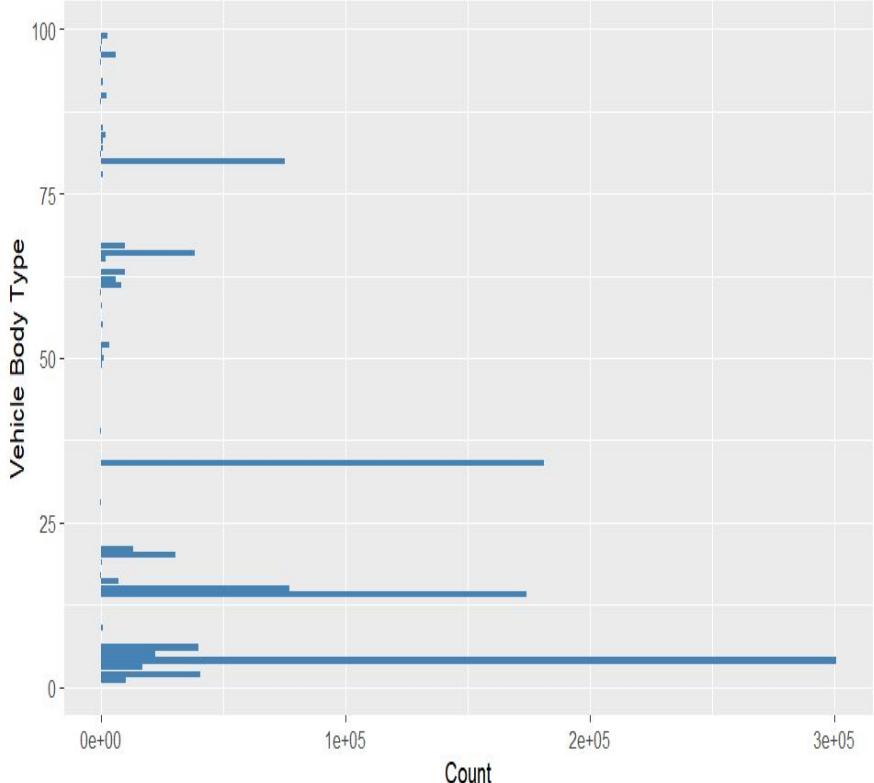
Weather vs Injury Severity



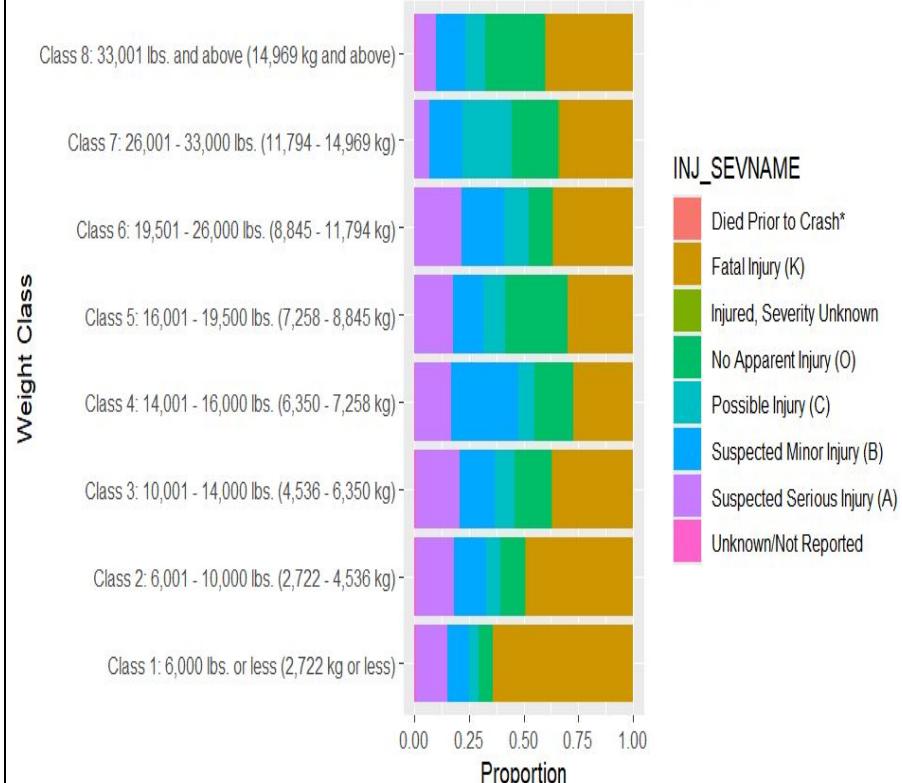


VEHICLE TYPE

Crashes by Vehicle Body Type



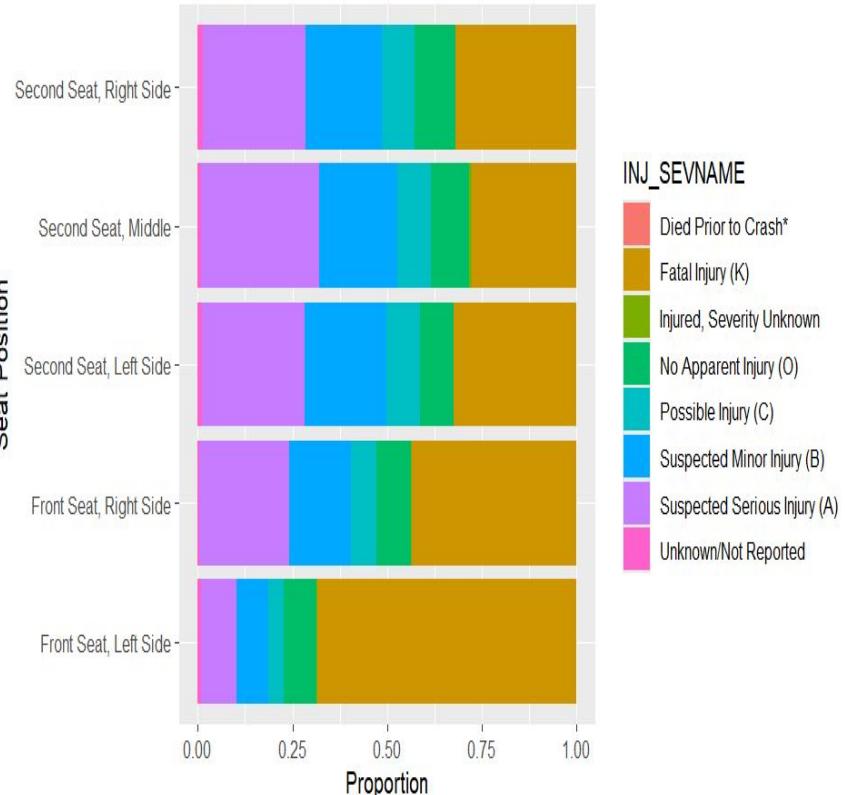
Vehicle Weight Class vs Injury Severity



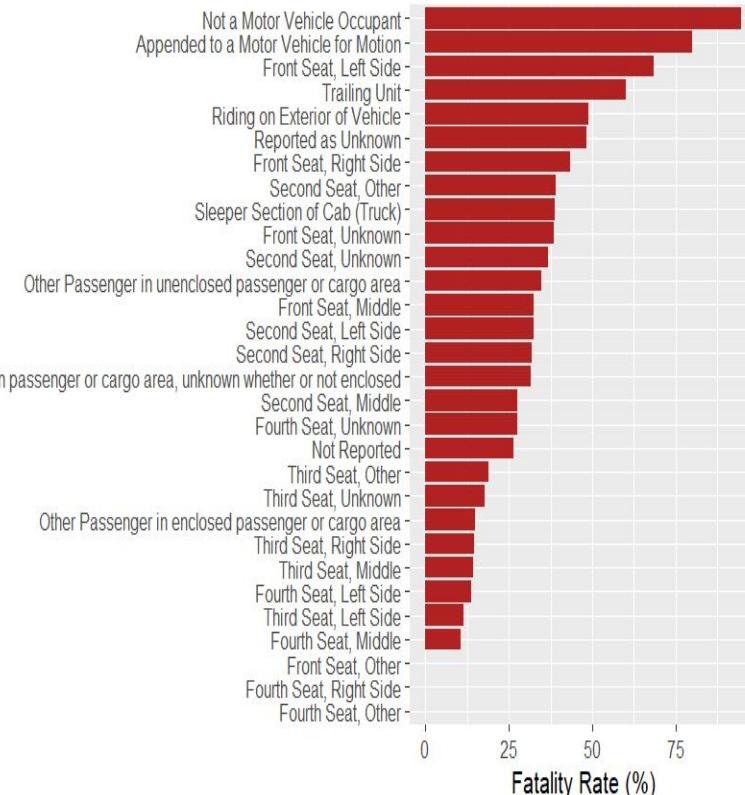


SEAT POSITIONS

Injury Severity by Major Seat Positions



Fatality Rate by Seat Position

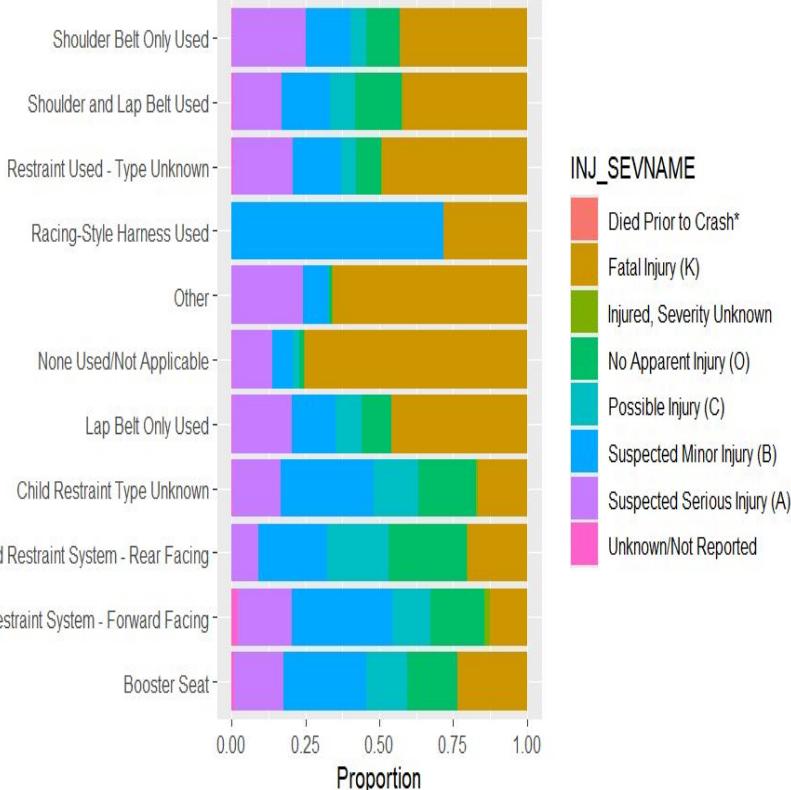




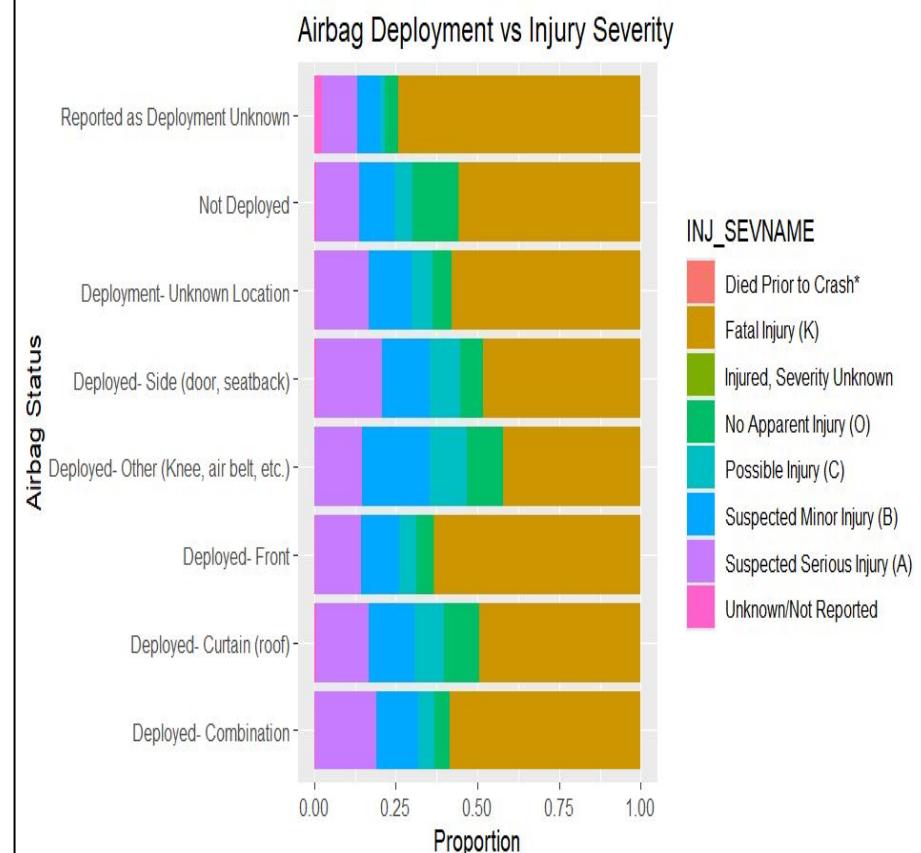
SAFETY EQUIPMENT

Seatbelt Usage

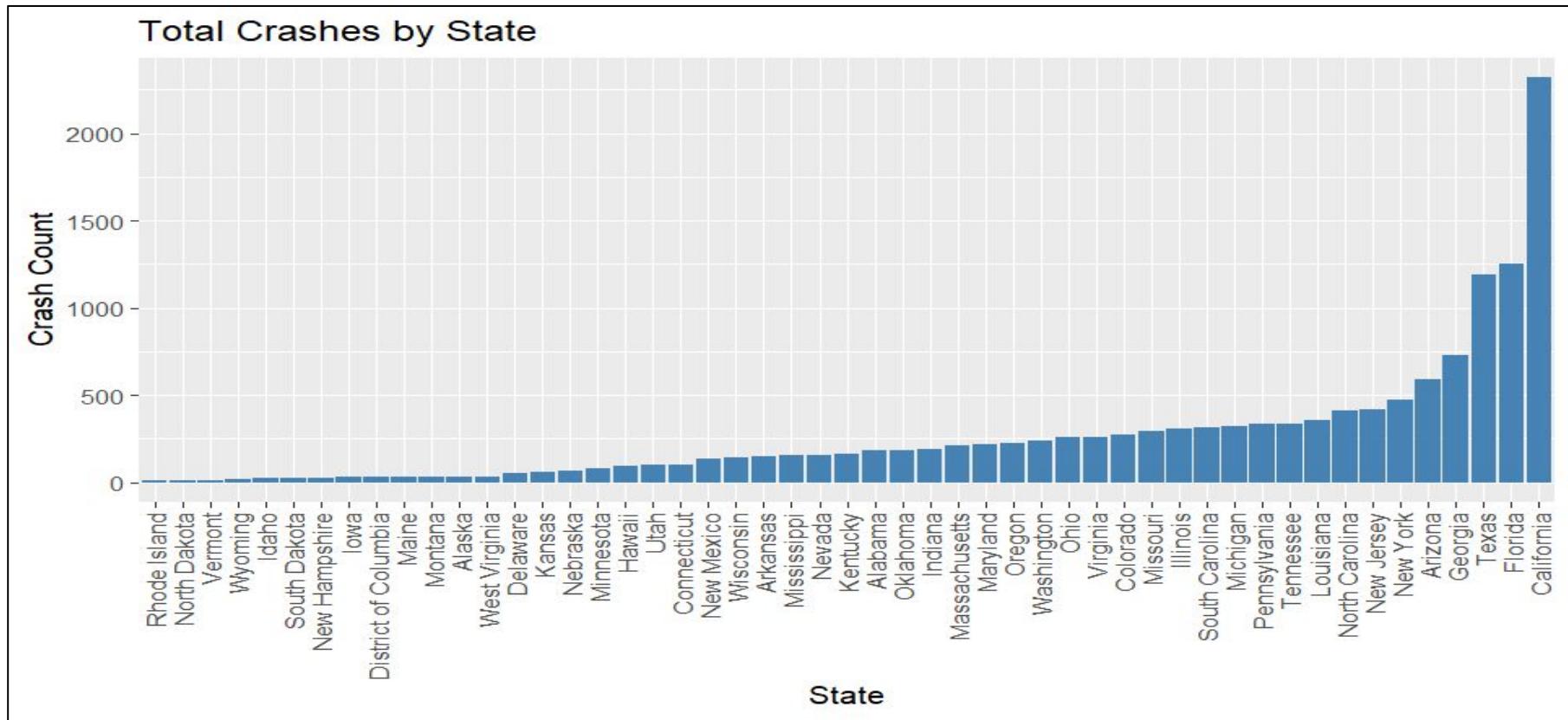
Seatbelt Use vs Injury Severity



Airbag Deployment vs Injury Severity

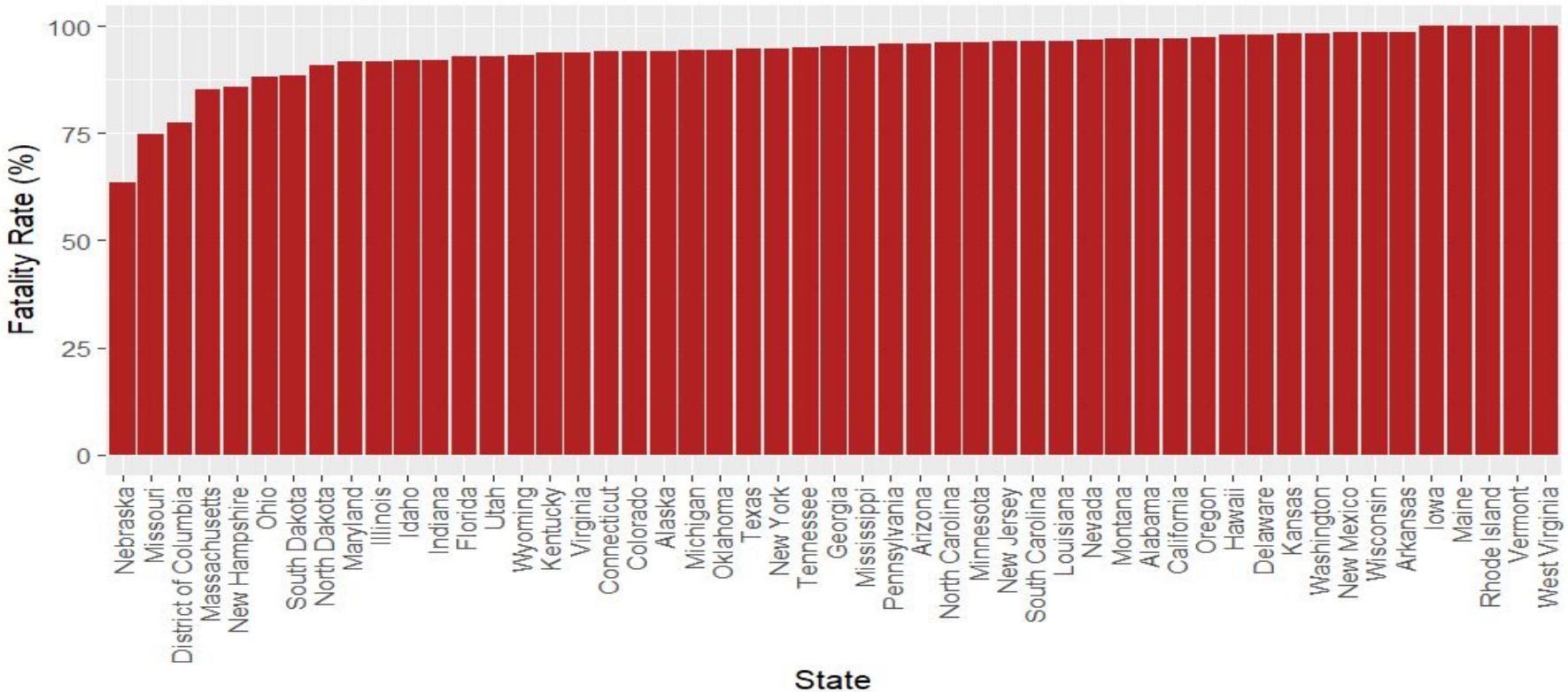


★ STATE WISE REPORT



★ STATE WISE REPORT

Fatality Rate by State





Web interface to Generate Traffic Accident Data Analysis with Preprocessing

Traffic Accident Data Analysis with Preprocessing

Start Date (2022):
12/01/2022

End Date (2022):
12/31/2022

State (optional):
-- All States --

Report Type:
Time of Day vs Severity

Generate Report

Heatmap of Injury Severity by Time of Day

Injury Severity

Crash Count

Suspected Serious Injury (A)

Suspected Minor Injury (B)

Possible Injury (C)

No Apparent Injury (D)

Injured, Severity Unknown

Fatal Injury (E)

Crash Prior to Crash?

Morning Evening Noon Night

Time of Day

Traffic Accident Data Analysis with Preprocessing

Start Date (2022):
12/01/2022

End Date (2022):
12/31/2022

State (optional):
-- All States --

Report Type:
Age vs Injury Severity

Generate Report

Age vs Injury Severity

INJ_SEVNAME

- Dead Prior to Crash?
- Fatal Injury (E)
- Injured, Severity Unknown
- No Apparent Injury (D)
- Possible Injury (C)
- Suspected Minor Injury (B)
- Suspected Serious Injury (A)
- Unknown/DK Required

Count

Age



Challenges

- Merging Complex Data: Combining multiple FARS tables using keys like `ST_CASE`, `VEH_NO`, and `PER_NO` was tricky—missing values and mismatches often caused merge issues.
- Missing & Unknown Values: Fields like injury severity, drinking status, and safety equipment often had vague or missing entries like “Unknown” or “Not Reported,” which we had to carefully filter out.
- Column Confusion After Merge: Merged tables produced columns with `.x` and `.y` suffixes, requiring manual checks to avoid plotting the wrong data.
- Imbalanced Categories: Some outcomes (like fatal crashes) were rare, so we used proportional charts to show meaningful comparisons instead of just counts.
- Coded & Wordy Labels: Many variables had long, detailed labels (e.g., “Deployment - Unknown Location”), which we needed to simplify for cleaner, easier-to-read visuals.
- Blank Visuals from Filters: Some charts failed when filters removed all rows—fixing this meant debugging filters and checking values with `unique()`.



Future Expansions

- Predictive Modeling: Use machine learning to estimate injury severity or fatality risk.
- Map Visualization: Plot crashes on maps to identify high-risk zones.
- Time Trend Analysis: Compare crash patterns over multiple years for long-term insights.
- Crash Risk Index: Create a visual “risk score” for each crash based on key factors.



REFERENCES

1. <https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/FARS/2022/National/>
2. National Highway Traffic Safety Administration. (2022). Fatality Analysis Reporting System (FARS) 2022 Data. Available at: NHTSA FARS.
3. McCartt, A. T., et al. (2010). “Trends in Fatal Crashes Involving Young Drivers.” *Journal of Safety Research*, 41(3), 123-130.
4. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.