Project Process Book: U.S. Traffic Accident Trends

Shrayas Raju, Landon Holcomb, Tong Liu, John Paul Lineberger

Overview and Motivation

Project Goals

This project aims to develop an interactive dashboard that visualizes traffic accident patterns across the United States at national, state, and county levels. By integrating a heat map and bar charts, the dashboard will highlight accident density and state-specific trends. Additionally, time distribution diagrams at various spatial scales will illustrate the temporal patterns of accidents, enabling the identification of high-risk travel periods for specific areas.

Motivation

Traffic accidents pose a significant public safety concern in the U.S. Understanding their geographical and temporal patterns, along with weather impacts, can lead to better decision-making for preventive measures. This visualization project serves to highlight key trends and provide an engaging way for users to interact with the data.

Related Work

This project drew inspiration from existing visualizations of spatial and temporal data, such as:

- Publicly available choropleth maps showing demographic or accident data.
- Visualization techniques discussed in class.
- Dashboards integrating maps and trend charts for real-time data analysis.

Questions

Initial Questions

- Which states experience the highest number of traffic accidents?
- How do accident counts vary across different months of the year?

Evolved Questions

- Which states experience the highest number of traffic accidents per capita?
- Are there significant seasonal patterns in traffic accidents?
- How do weather conditions correlate with accident frequencies and distributions?
- How do accident trends differ at state versus national levels?
- What insights can be gained from localized analyses at the county level?

Data

Source

The dataset was sourced from *US Accidents: A Countrywide Traffic Accident Dataset* (Moosavi, 2023)., containing detailed records of traffic accidents in the U.S. It includes variables such as state, month, weather conditions, and accident counts.

Preprocessing

Data preprocessing involved cleaning missing and erroneous values, aggregating accident counts by state and month, using python libraries such as pandas, geopandas and NumPy. For the prototype we had multiple levels of preprocessing due to lack of communication. For the final build of the project we will standardize our process.

Challenges

Handling large datasets required optimization techniques to ensure smooth visualization rendering in the browser. Additionally, aligning disparate data formats, such as merging weather data with accident statistics, posed integration challenges.

Exploratory Data Analysis

Initial Sketches

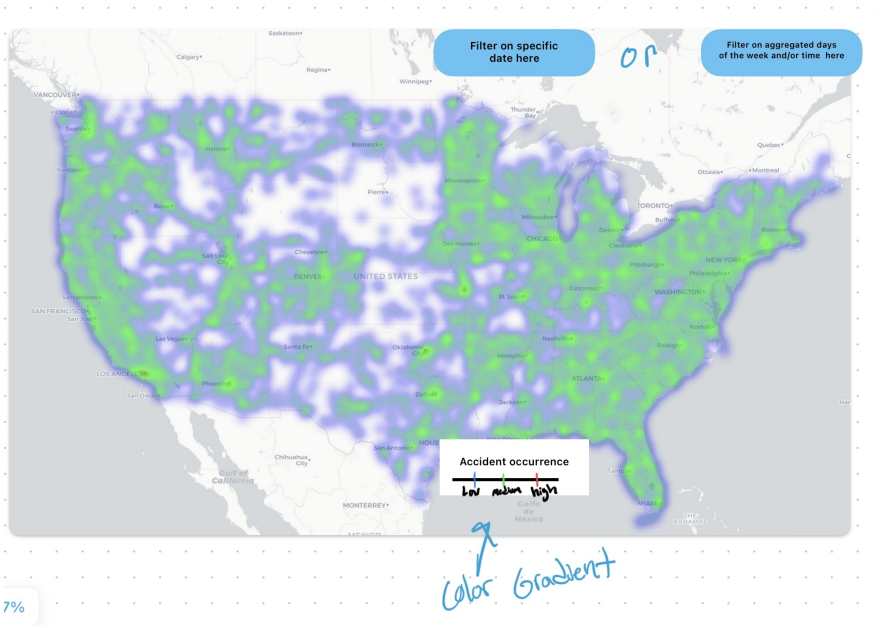
The initial design incorporated:

- A heat-map to show areas of high accident rates. We decided against this in favor of a choropleth map so we could more easily see the effects in individual counties and states.
- A sketched example of zoomed in heat-map, showing number of accidents within a certain radius for various severities. This would be useful for local data but its difficulty is outside the scope of this project.

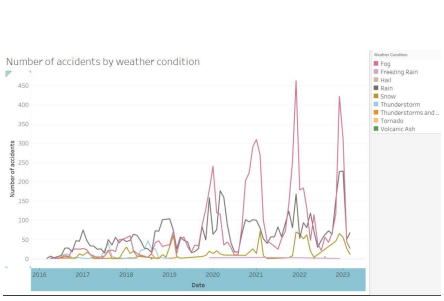
- A bar chart to highlight the top accident-prone states. We once again decided against this visualization in favor of a slightly more useful one, showing the top accident rates per capita of the states. Accident rate per capita is more useful than just accident rate per state because it accounts for population differences across states, providing a more fair comparison of traffic safety.
- A line chart showing weather vs accident rates.
- A radar chart showing the impact of weather on accident severity.

Initial Insights

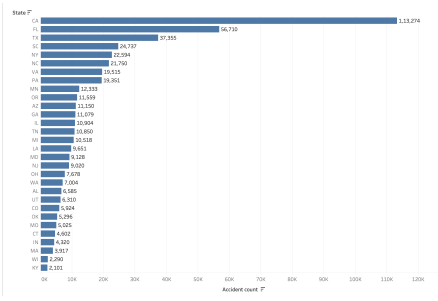
- California and Florida consistently rank as the states with the highest number of accidents, unsurprisingly.
- There is a notable increase in accidents during winter months.
- There is a significant increase in accidents for winter weather, which is to be expected given the spike in accidents found for winter months.



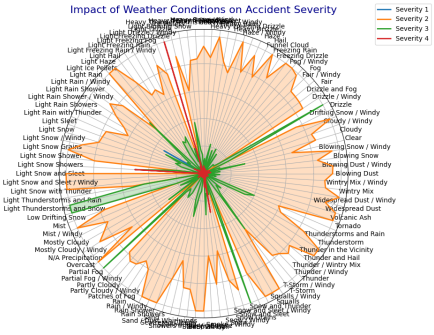
(a) Heat-map showing accident hotspots through the entire country



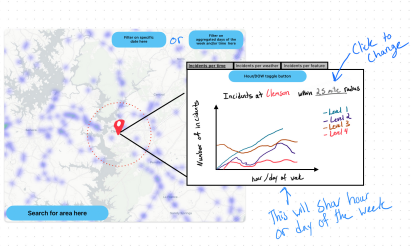
(b) Accidents as a function weather



(c) Bar chart showing numher of accident per state.



(d) Radar plot showing weather rate and severity.



(e) Zoomed heat-map showing local data.

Figure 1: Visualizations used in our initial analysis of the dataset

Feedback and Iterations

- Feedback indicated a need for clearer designs. For example, it was noted that image 3 masks the interesting data by forcing us to focus on California and Texas.
- Designs were revised to highlight more relevant data, such as accidents per capita.

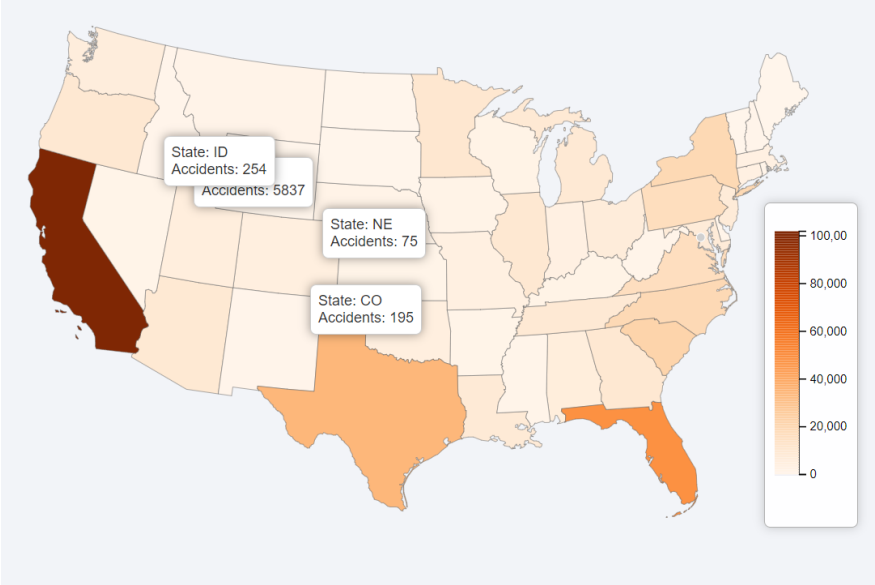
Dashboard Evolution: From Prototype to Version 3

Prototype Overview

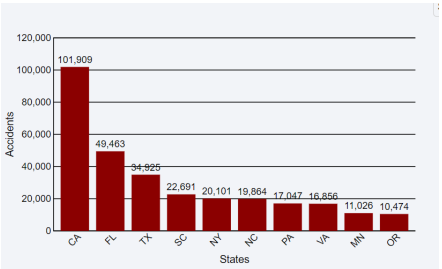
The initial prototype of the dashboard featured:

- **Choropleth Map:** Displays accident counts by state with hover-over tooltips for detailed statistics. We use this visualization as it is extremely general, showing the results per state for all variables in the dataset.
- **State Choropleth Map:** The zoomed-in map provides a detailed analysis of traffic accidents at the state and county levels, offering a granular view of accident data that complements the broader national overview. This functionality allows users to drill down into specific regions to identify localized patterns, trends and severity level.
- **Country level Chorpleth Map:** The country level map gives even more detailed statistics on the accidents, including locations, time and duration, weather conditions and severity.
- **Bar Chart:** Shows the top 10 states with the highest accident rates. In the next iteration of our project we will use accident rates per capita which will provide more statistically relavent information.
- **Line Chart:** Visualizes accident trends over the months of a selected year. This shows a steady increase over winter months. Future improvements will allow us to filter using weather conditions.

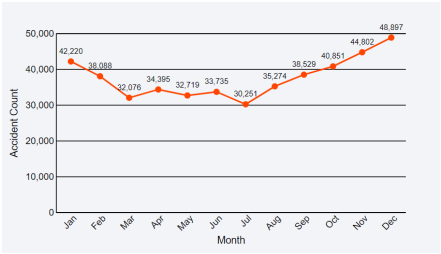
We chose to go with these because it shows the most relevant information, starting off broad in the chorpleth map before filtering information with the bar chart and line chart, allowing us to specify dates, and weather in the near future.



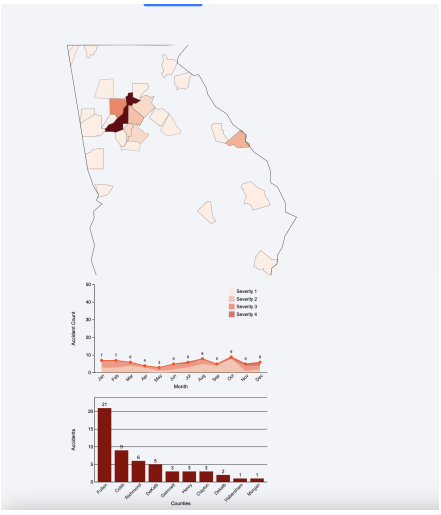
(a) Choropleth Map: Displays accident counts by state with hover-over tooltips for detailed statistics.



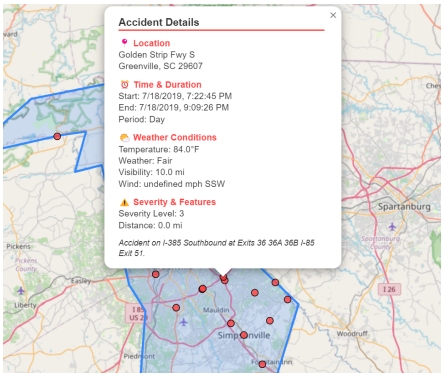
(b) Bar Chart: Shows the top 10 states with the highest accident counts.



(c) Line Chart: Visualizes accident trends over the months of a selected year.

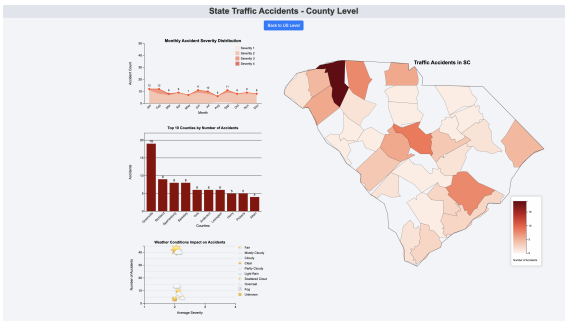


(d) Choropleth Map: State level map shows accident rates and severity level for individual counties in a state.

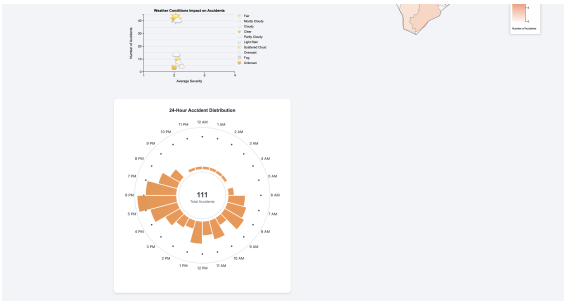


(e) Choropleth Map: Country level map showing

Figure 2: Three designs used in our project prototype.



(a) Overview of dashboard showing multiple figures.



(b) second overview of dashboard

Figure 3: Overview of state level dashboard

Although these visualizations provided a solid foundation, feedback indicated several areas for improvement.

- **Granularity of Data:** The lowest level of data, the state level, it too high. The dashboard should show county or city level data.
- **Bar chart:** The bar chart only shows 10 states with no clear reason.
- **Single dashboard view:** The dashboard should be on a single screen, removing the need to jump between separate visualizations or scroll to see the entire dashboard.
- **Dynamic Updates:** Visualizations should dynamically adjust based on user interaction rather than switching to new views.
- **Drop-down:** Drop-down menus should be removed in favor of interactive visualizations
- **Time chart:** The time chart was appreciated but should work at multiple levels of granularity.

Version 2 Overview

The second version of the dashboard addressed the limitations of the prototype and incorporated the following improvements:

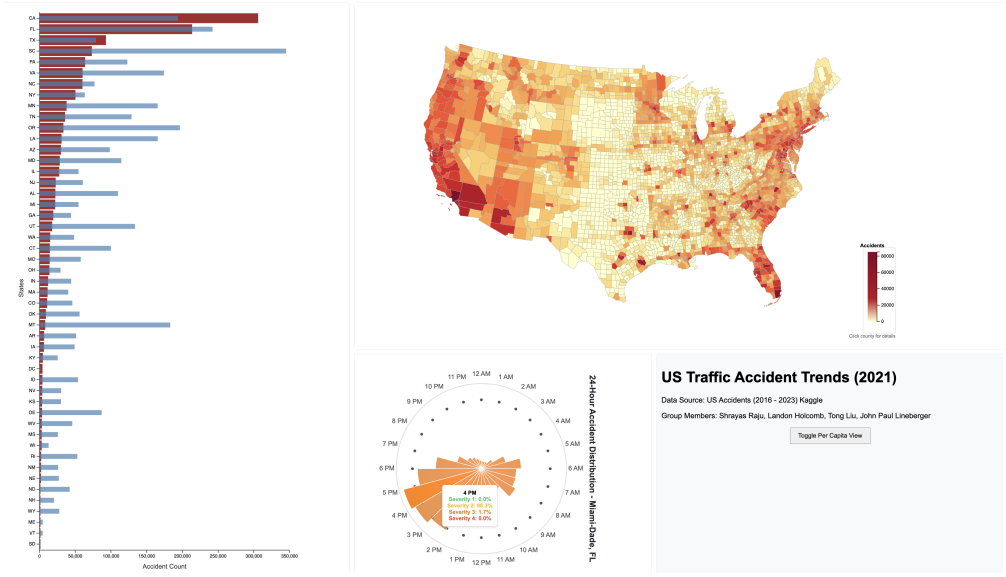


Figure 4: Version 2 of the dashboard

- **Granularity of Data:** The chlorpleth map now displays data at the county level rather than the state level, allowing for more detailed geographical analysis.
- **Improved Interaction:** A toggle button was introduced to switch between total accident counts and per capita views, allowing for fairer comparisons across states with varying populations.
- **Integrated Design:** The dashboard layout now consolidates all visualizations into a single cohesive view, removing the need for switching between pages and supporting seamless exploration of relationships between different dimensions.
- **Focus on Usability:** All visualizations now dynamically interact within the same view:
  - Clicking on a county updates the radar plot.
  - Users can explore regional and temporal dimensions seamlessly within the dashboard.
- **Data Clarity:** Clear labeling of data sources and team member contributions ensures transparency and user-friendliness.

Version 2 feedback

- **Dynamic updating of Bar chart:** The bar chart should reorder when toggling the per capita view to allow a comparison of the highest per capita accident rates.
- **Time chart levels:** Again it was suggested that we include multiple levels of granularity for the time chart, such as state, county and city level for easy comparison.

Version 3 Overview

We once again address the feedback given to us to desing our final version of the dashboard. Below we list the improvements over version 2.

- **Dynamic Sorting in Bar Chart:** The bar chart now dynamically reorders states when toggling between absolute accident counts and per capita views, making comparisons more intuitive.
- **Three Levels of Time Charts:** To enhance temporal analysis, time charts were added for national, state, and county levels. This allows users to explore trends at varying geographical granularity.
- **Visual Optimization:** Adjustments were made in the organization of the dashboard to allow three time charts to be displayed in parallel.

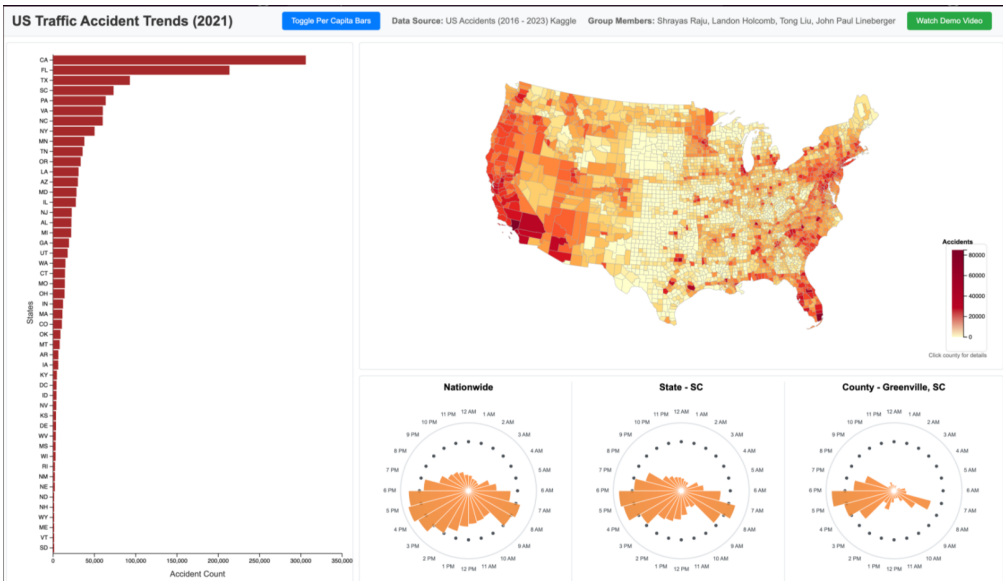


Figure 5: Version 3 of the dashboard

Tools Used

The project was implemented using:

- **D3.js:** For dynamic and interactive visualizations.
- **JavaScript, HTML, and CSS:** To build the dashboard interface.
- **Python:** For data cleaning and preprocessing.

Insights

- **Rush Hour Risk:**

  Accidents are primarily concentrated over the morning and evening peak hours, with a higher likelihood of occurring during 4-6 PM at the national level and in most states, such as Florida and South Carolina. Interestingly, South Carolina has an above-average high number of accidents between 7-8 AM, which is comparable to evening peak levels. This insight is significant and should allow for further research into how how to mitigate accidents at these hours.

- **State-Level Trends:** Exploring the accident rates per capita shows states with abnormally high accident rates. These are potential outliers that need further research to show what is causing the higher accidents rates, such as road conditions, local weather or driver behavior.

- **Impact of Granularity:** Related to the above, we can examine which counties have the highest accident rates in these outliers, drilling down the focus to specific hot spots and high risk areas.

Future Improvements

While the visualizations look good, they are slow to load. There are several aspects that need to be improved upon for future iterations of our dashboard. The following list outlines improvements we will make for the final version.

- **Performance Optimization:**
  - Implement data aggregation techniques to improve the speed and performance of the dashboard when handling large datasets.
  - Explore caching strategies to store pre-aggregated accident counts at different levels (state, county, city) to avoid recalculating data during user interaction.
- **Weather Condition Filters:**
  - Add filters that allow users to explore how weather conditions affect accident rates. For example, users should be able to toggle between weather types (e.g., rain, snow, fog) and see the impact on accident counts and severity.
  - Introduce interactive elements (like sliders) to filter accident data by weather conditions and visualize how different conditions (snow, rain, fog, etc.) impact accident rates.
- **Accessibility and User Experience:**
  - We could improve accessibility by insuring we follow WCAG (Web Content Accessibility Guidelines)
  - Add export options so users can download the dashboard's current state as an image or PDF.

References

- Moosavi, S. (2023). *US Accidents: A Countrywide Traffic Accident Dataset*. Retrieved November 17, 2024, from `https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents/data`.