

Problem

In this project, we deep dived into the fatal car crash records in the U.S. in year 2016, collected by National Highway Traffic Safety Administration (NHSTA). By wielding this dataset, research, and interactivity, we developed a visual essay trying to thoroughly explore the top risk factors that are highly correlated to fatal motor vehicle crashes.

Motivation

Each year fatal motor vehicle crashes in the U.S. lead to an estimated societal burden of more than \$230 billion from medical and other costs. Motor vehicle crashes are also the leading cause of death for persons every age from 5 to 32 years old. Identifying the top risk factors that contribute to fatal motor vehicle crashes is important in developing interventions that can reduce the risks associated with those factors and promoting safe alternatives.

Analyzing and visualizing the NHSTA traffic record dataset is a challenging task: the dataset consists of 20 data tables which collectively describe over 400 known attributes of 34,619 fatal motor vehicle crashes. Furthermore, as a motor vehicle crash usually results from a combination of factors including vehicles, road users, environment, and the way they interact, it possess challenges to underpin their individual influences in road traffic fatalities. It is our goal to create visual narratives and exploratory tools that help people digest and learn from this dataset more easily.

Dataset

Google BigQuery public dataset:
https://bigquery.cloud.google.com/dataset/bigquery-public-data:nhtsa_traffic_fatalities

Lesson Learned From 2016 U.S. Fatal Crash Data

Chun Wang, Kexin Yu

Approach

1. We first explored, cleaned, transformed the data with BigQuery and Jupyter Notebook and collected the most interesting insights uncovered from the data.

2. Then we formulated an interactive document in the New York Times style consisting mainly of visual explainers produced in Altair (a Python library built on top of the Vega-Lite visualization grammar), textual narratives (which incorporate NHSTA research notes and other data sources to explain patterns and outliers observed in visualizations) as well as actionable takeaways such as safety tips for drivers, passengers and pedestrians.
3. We employed design techniques such as sticky navigation menu, annotations, tabs and tooltips to facilitate reading and reduce visual clutters.

4. We also constructed and provided readers with **1)** an interactive 3D map where users can learn about case details of every fatal car crash occurred in 2016 and filter records by user-selected attributes; **2)** an interface to play with a predictive model that we developed based on our analysis in order to further understand how much changing a factor will affect overall fatality risk quantitatively.

Results

Future Work

- Evolve the visual essay to polish the existing stories and incorporate more interesting factors, particularly those that highlight common misconceptions. May use reader quizzes as a means of education.
- Incorporate the predictive model into the interactive document.
- How might we better capture the way various factors interact with each other, e.g. identify factors that work together to contribute to fatal crashes?
- How would readers like to interact with the open-ended exploration map and what can we do better to facilitate their explorations?

