D.C. Car Fatality Analysis

Team 48 Project Proposal
CSE 6242 - Data & Visual Analytics

Adam Peir
Justin Schulberg
Maynard Miranda
Ryan Doogan

Goal (Q1/4/5)

Our goal in this project is to **create an interactive and robust visualization dashboard to communicate the true impact of traffic-related crashes in D.C.** to stakeholders who would be most interested in D.C.'s Vision Zero plan, such as policymakers, police, and residents (Q4) of Washington, D.C. (Q1).

Success of this project will contribute to the success of D.C.'s goal of zero crash-related fatalities by 2024 (Q5).

Plan of Activities (Q8)

Task	Status	Assigned To	Start Date – End Date
Project Ideation	Completed	All Team Members	10/02 – 10/08
Set-up/Training of GitHub	Completed	Justin Schulberg	10/05 – 10/09
Project Proposal Write Up	In progress	All Team Members	10/09 – 10/15
Project Proposal Slides	Not started	All Team Members	10/09 – 10/15
Project Video	Not started	Ryan Doogan, Maynard Miranda, Adam Peir	10/09 – 10/15
Identification of Datasets	In progress	All Team Members	10/16 – 10/29
Data Pre-processing	Not started	All Team Members	10/23 – 11/06
Progress Report	Not started	All Team Members	10/23 – 11/04
Data Integration	Not started	All Team Members	10/30 – 11/13
Data Analysis/Modeling	Not started	All Team Members	11/06 – 11/29
Data Visualization/Dashboard	Not started	All Team Members	11/06 – 12/03
Finalizing Final Report & Project Poster (Q9)	Not started	All Team Members	11/27 – 12/10

Checks for Success (Q9)

- Midterm
  - All data sources collected, all models run and validated, initial model analysis begun
- Final
  - All deliverables completed, including poster, final paper, video, slides, and readme

Areas for Improvement (Q2)

Although D.C.'s Vision Zero website includes numerous data visualizations and analytical products, they lack the following features, that we aim to improve upon (Q2):

- **Disjointed Visualizations** | Even though all of the visualizations are built in Tableau and portray insightful information, they are not connected in one seamless dashboard, allowing users to see relationships across multiple features at once.
- **Disparate Tools Used** | Currently, the Vision Zero Team creates its visualization products in both Tableau and ArcGIS. This allows for more powerful mapping capabilities in ArcGIS; however, the two products are kept separately, making it difficult to understand the relevance of the geospatial data in the context of other visualizations.
- **Lack of Diversity in Datasets Used** | Currently, only data on direct crashes are being used.
- **Incorporation of Geographic Analysis** | While the current data is mapped, there is substantial room for additional spatial analysis of existing crash data.

Expected Innovations (Q3/7)

To build upon the work already done, we will:

- Not only integrate multiple data sources, but we will **centralize all data and corresponding visualizations into one visualization software, Tableau**.
  - Because we will only be using Tableau, and it offers free licenses to those with University emails, the cost of this project is $0 (Q7).
- Additionally, we will use time-series modeling to project traffic-related crashes into the future and incorporate additional geospatial features into our analysis.
- Lastly, the team plans to integrate not only crash/fatality data, but also data on where innovative road infrastructure has been built and other land use metrics.

Risks & Payoffs (Q5/6)

- Potential risks include:
  - Dealing with different levels of detail can be very methodologically challenging and can lead to unexpected delays and increases in complexity.
- Potential payoffs
  - Improved understanding of DC traffic patterns that can be used to design interventions that reduce local crash volume.

Literature Review (Q2/3) - Time Series/Analytical Methods

On top of D.C.'s Vision Zero plan, many traffic safety researchers have approached the issue of reducing vehicle-related crashes using analytical approaches.

"Predicted crashes in Washington D.C. using two separate models: ARIMA and Heston."

- ARIMA assumes that any volatility in the data is constant, while Heston assumes that the volatility is arbitrary.
- Heston > ARIMA in terms of accuracy

"Evaluated crashes based on different collision types, such as rear-end crash, sideswipe crash, and angle crash"

- Bayesian analytics was used to perform the evaluation. The result shows that each collision type has different rates and risk factors.
- For example, # of lanes on the mainline and ramp length significantly impact rear-end crash and sideswipe crash but no impact on angle crash. Deceleration lane length positively impacts rear-end crash but has no relationship with sideswipe crash and angle crash.

"Analyzed what factors affect pedestrian crashes in Texas' county-level areas using OLS Regression"

- Suggests that homelessness, median household income, and poverty positively correlate with pedestrian crashes.

Areas for Improvement:
The studies above do not have an interactive dashboard that can better communicate their results to non-data stakeholders.

Literature Review (Q2/3) - Bicycle Focus

One component of our crash-related analysis is on bicycle safety.

“A data-driven approach for assessing biking safety in cities”

- Some of the most common forms of vehicle-related crashes are between cars and bicycles.

“Why cities with high bicycling rates are safer for all road users”

- Cites presence of cycling lanes as a causal link to reducing crashes for cyclists;
- However, some of these studies only look at the absolute presence of a cycling lane (does one exist or not) as a factor for reducing crashes. In actuality, different types of cycling lanes exist and have different effects on road safety.

“Bicycle Tracks and Lanes: a Before-After Study”

- Looks at different types of road-calming and cycling infrastructure measures, bit only uses observational methods to analyze the impact on vehicle-bicycle crashes between different cycling infrastructures.

Areas for Improvement:
Integrate advanced analysis on large datasets related to large swathe of bicycle infrastructure measures and road-calming measures.

Literature Review (Q2/3) - Geographic Analysis

Existing methodology makes heavy use of spatial analysis of crash data. Incorporating geographic data allows researchers to isolate problems with specific locations. Incorporating geographic data allows researchers to incorporate analysis of demographic, socioeconomic, and land use factors

“Pedestrian crash estimation models for signalized intersections”

- Researchers in North Carolina analyzed the factors predicting pedestrian crashes at signalized intersections, finding that while increased traffic and population increases crash frequency, iland use variables indicative of increased pedestrian traffic decrease crash frequency

“Macro-level safety analysis of pedestrian crashes in Shanghai, China”

- Researchers in China considered factors of Traffic Analysis Zones in Shanghai that predicted crash frequency, finding that pedestrian crashes were higher in TAZ’s with medium land use intensity, higher populations and longer major and minor roadway arteries.

“Intersection crash prediction modeling with macro-level data from various geographic units”

- Researchers in Florida analyzed crash prediction model performance at different levels of spatial aggregation, finding that models with zip-code tabulation data perform better than models incorporating data tabulated at other levels of detail.

Areas for Improvement:
Existing Vision Zero research lacks a major spatial analysis component

Literature Review (Q2/3) - Visualization

To make our analysis easily accessible and understandable, we plan on creating a visualization.

“A Data-Driven Safety Dashboard Assessing Maryland Statewide Density Exposure of Pedestrians, Bicycles, and E-Scooters”

- Researchers superimposed photo enforcement citation data on crash data to illustrate the effectiveness of automated enforcement.

“Using automated enforcement data to achieve Vision Zero Goals: A case study.”

- Researchers used transportation, mobile, and demographic data to determine safety risks.

Areas for Improvement:

Limited in scope. Visualizations can catch more types and layers of data.