Utah Accident Visualization Process Book

Alan Bird, Eric Sims

Fall 2022

Overview and Motivation

It seems to be a common conversation topic of how terrible “Utah Drivers” are. How true is that? We believe that Utah roads are plagued with people going too fast, running red lights, disobeying traffic laws or being just plain ignorant of their surroundings on the road. We think it is worth considering looking at the accidents that have happened in Utah in a given year to determine what are the most common types of crashes and when and where they tend to happen.

We have heard that it is best to avoid the roads at night, but is that true? When do most DUI’s happen? Are teenage-involved accidents more common in urban areas than rural areas? Do elderly drivers become more involved in accidents during times of inclement weather?

Related Work

Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

Questions

With this data visualization we are trying to answer many questions about diving and accidents in Utah in general, but specifically we want to show the correlation between different types of accidents, when they are likely to happen and what the causes of them are. With this data law enforcement will be able to better predict optimal times to be watching the streets and citizens using the Utah roadways will be more prepared and aware of potentially dangerous times to be on the road and locations with higher than average accident rates.

Data

Data Source

The data we are using is from this website: https://dev.socrata.com/foundry/opendata.utah.gov/herb-zqda. Socrata is an open source data site that has hundreds of sets of collected data. We are focusing on the Utah data crash info that contains information of crashes from thousands of car accidents across a few years spanning from 2017 to to 2020.

Data Cleanup

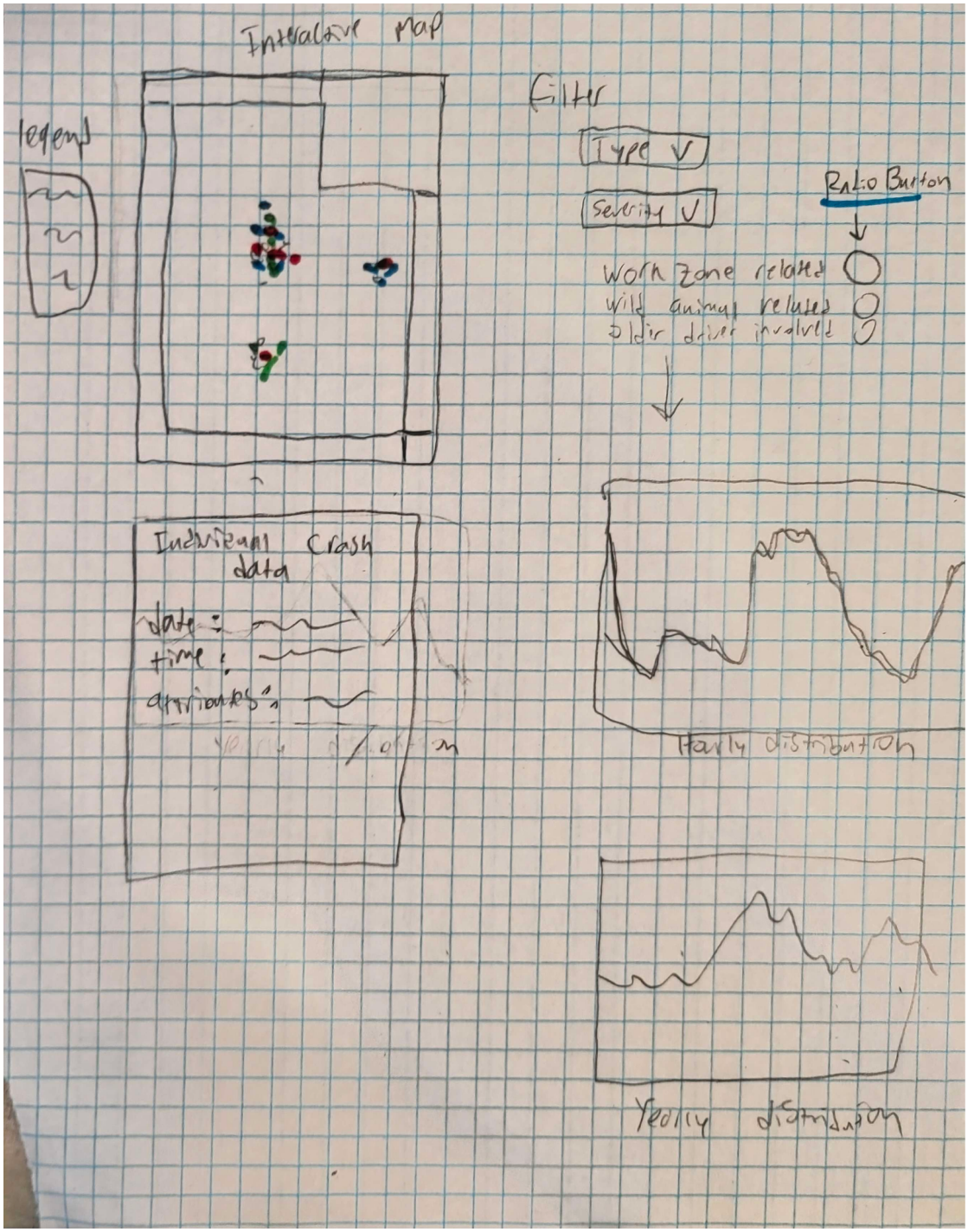
The data arrived in an extremely usable format, the CSV file provided easily converts to a large array of JSON objects that is quite straightforward to use. The one data issue we found is that the location data is stored in UTM and our implementation through the google maps api required latitude and longitude. We created a python script to take the initial csv and convert the UTM coordinates to new “lat” and “lng” columns in a new CSV file.

Exploratory Data Analysis

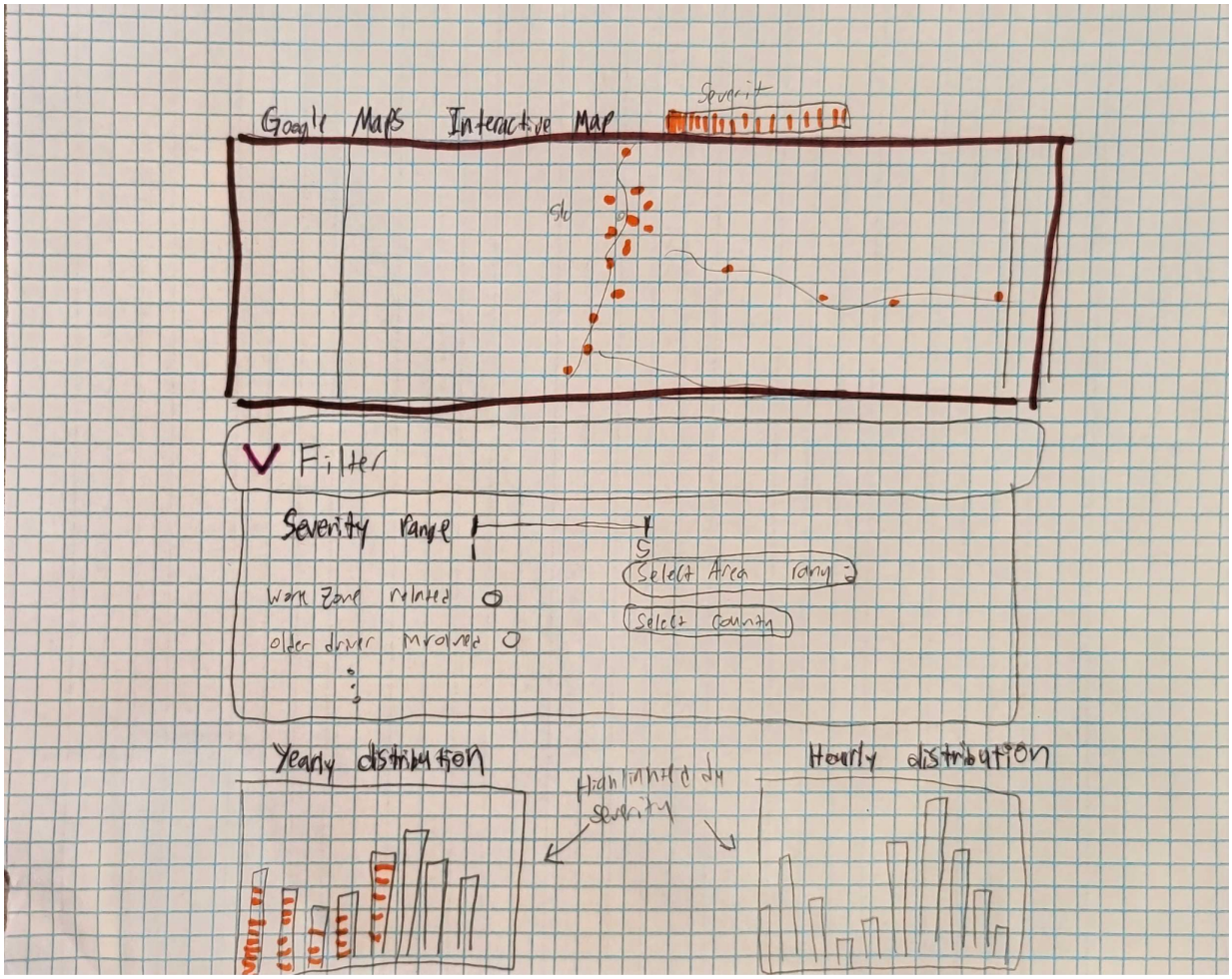
What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

Design Evolution

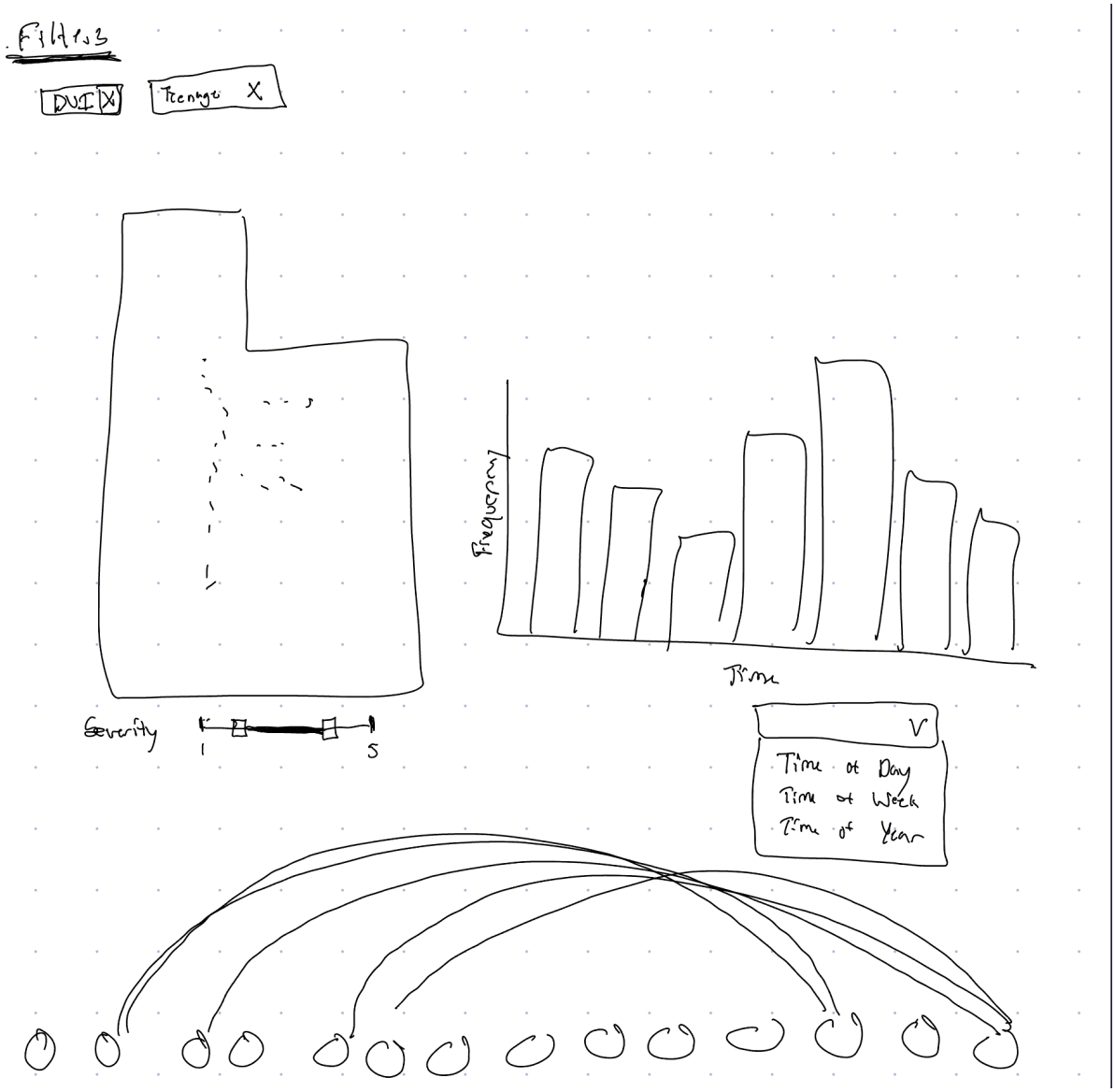
We knew that we wanted to have a few key features in our visualization. First, we wanted to have an interactive map that plots where the crashes have been. We also wanted a filtering area that users could focus on certain types of crashes. Furthermore, we wanted a detailed page for a selected crash and finally we wanted bar charts showing the most common times that crashes tend to occur.



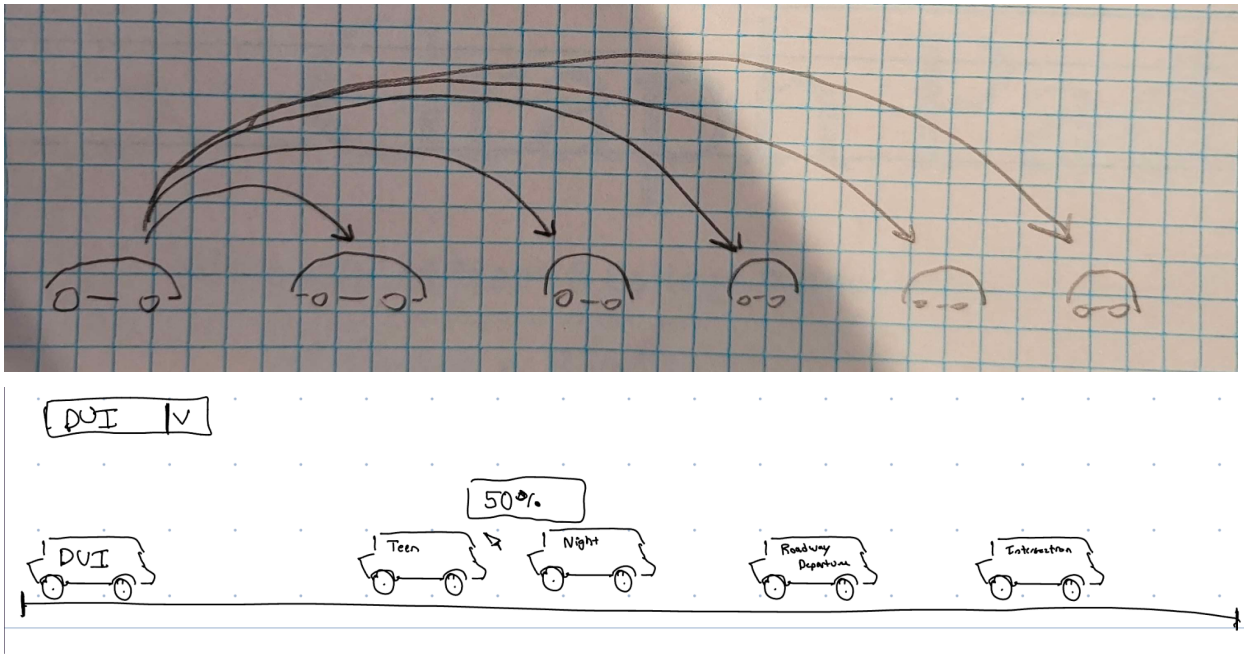
Prototype #1 - This was a pretty good home run of what our final design is. Although it did not consider a data derivation.



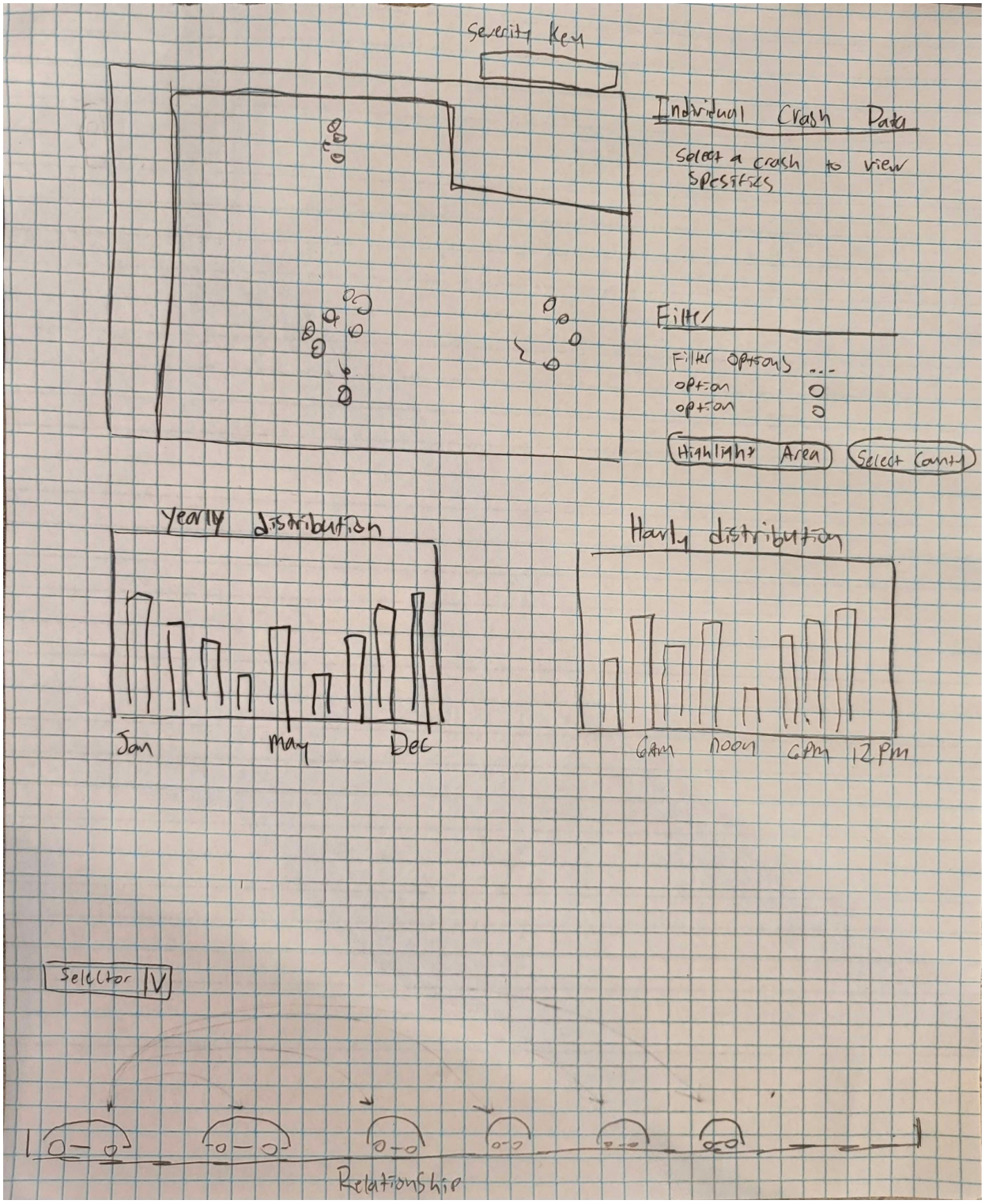
Prototype #2 - Experimenting with using a severity scale on the map to differentiate the accident plots



Prototype #3 - General layout design being finalized, brainstorm idea of data derivation



Required Feature 5 prototyping: Relationship Correlations, mouse hover to show percentage correlation between selected feature



Final Design- relationship correlation display at bottom, ⅔ screen interactive Google map with overlaid accident data and individual crash data and filter options next to it and above Yearly and hourly distribution plots.

Implementation

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

Project Peer Feedback with Kyle Woods and Kasidy Fernandes

General observations from Kyle and Kassidy

-The project looks good and interesting

-The split between Must-Have features and Optional features looks balanced

-The data looks simple and good to work with, but the ideas you have should be scalable with more data quite easily

Suggested Changes (and our responses)

-There should be more data, getting data from multiple years will help with your time distributions, not to mention 2020 will likely be an outlier for driving/accident data because of the pandemic.

-Our assessment: This is a good suggestion, the place we found our data in only has the data from 2020 but as a part of our goals for our first week we want to get the data for multiple years as well as 2020 so that our averages better reflect and can predict current driving patterns.

(Update: It turns out that our data contains crashes from 2016 to 2019, even though it is called “Utah 2020 Crash Data”. What is interesting is that it seems that only 2019 has a full dump of crash data with 2016-2018 being sparse.

-Is there an ability to dynamically add new accidents as they happen?

-Our assessment: While this would be an interesting feature it lies outside the scope of what we believe our assignment should cover. The data we have has been curated to specifically be in the format we need it and the original source of the data may not be even available for a certain period of time. We will be increasing our data set but not to this extent.

Feedback analysis

Kasidy and Kyle did a great job with their feedback, they were insightful with their suggestions and their general observations were positive and encouraging. They could have been a bit more critical of our design look or had more suggestions for us to work with but their contributions were sufficient.