How to survive on road？

Boya Zhou (bzhou@wpi.edu, github: boyazhou1993)

Caidan Liu (cliu7@wpi.edu, github: CaidanLiu2)

Yuting Liang (yliang3@wpi.edu, github: DarienLiang)

Final Project Process Book

CS573 (Fall 2016)

1. Overview and Motivation

Driving is an essential part of day-to-day life for most people. However, with so many people on the road with varied physical and psychological conditions, driving an automobile sometimes will be a risky and dangerous activity. In recent years, many safety actions have helped decrease the number of traffic deaths. However, newly released government data paint a sobering picture of safety on the nation's roads and highways. Despite after decades of safety improvements, traffic fatalities up sharply. The number of people who died in car accidents reached 35,092 last year, according to newly released figures from the National Highway Traffic Safety Administration. There a 7.2% increase over 2015. The last time there has a large single-year increase was back in 1966, when fatalities rose 8.1 percent from the previous year.

Based on the above information, impaired driving should be gained more attention following increasing amount of crash, but in a less “preaching” way(though we make it in a preaching way in the end). So our visualization design absorbs some basic user information in our visualization, such as which state users live in, and what commute mode they often use. Meantime, we focus on the relationship between fatalities and some key factors (weather and bad behavior) to help people gain some insights if they want to avoid fatal crashes.

2. Objectives

Visualization design is guided by following questions:

(1) How to absorb the basic information of users in visualization?

Failed at some level, we will talk about that latter in Data Flow part.

(2) What is the traffic death in each state? And how the commute mode, climate, and bad behavior change influence the traffic fatal tolls in each state?

(3) How the climate change influence the average fatal death in each crash?

(4) How the commute mode change impact the traffic fatalities?

(5) How the drivers’ bad behaviors influence the average fatal tolls in each traffic accident?

3. Dataset

3.1 Data Source

The dataset containing all detailed information of 2015 traffic fatalities were released by the the National Highway Traffic Safety Administration. (https://www.whitehouse.gov/blog/2016/08/29/2015-traffic-fatalities-data-has-just-been-released-call-action-download-and-analyze)

The dataset of recent 30 years traffic fatalities are collected from Wikipedia. (https://en.wikipedia.org/wiki/List_of_motor_vehicle_deaths_in_U.S._by_year)

3.2 Data Processing

Python is used for data processing tool. In the final visualization implementation, two datasets are used. byYear.csv contains traffic death tolls over years. final0.csv is the main dataset contains detailed information of 2015 traffic crashes.

3.2.1 byYear.csv

- **Year**: From 1900 to 2015.
- **Deaths**: The number of fatal death.

3.2.2 final0.csv

Data used in the visualization implementation is derived from accident.csv, distracted.csv, person.csv and vehicle.csv joined on ST_CASE.

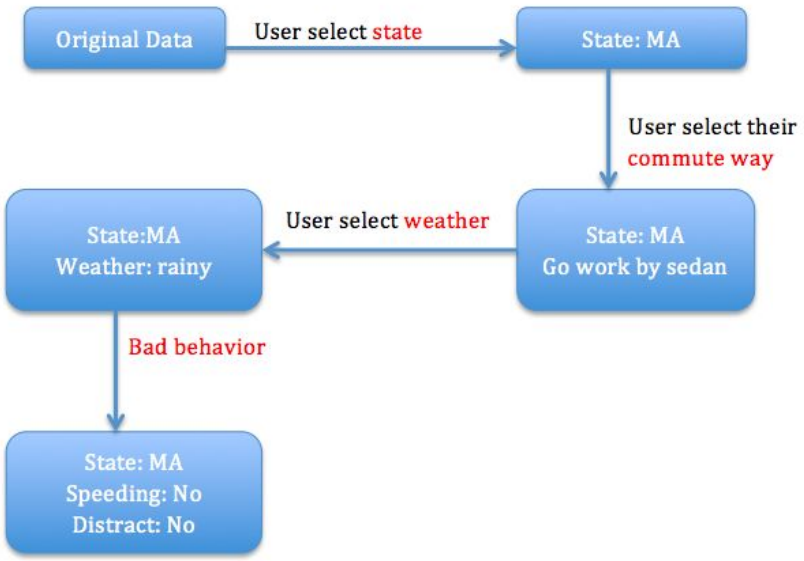
- **ST_CASE**: This is the unique case number assigned to each accident. It appears on each data file and is used to merge information from the data files together.
- **DAY_WEEK**: This data element records the day of the week on which the crash occurred derived from accident.csv. We remove accidents with unknown DAY_WEEK.
- **HOUR**: This data element records the hour at which the crash occurred derived from accident.csv. We remove accidents with unknown HOUR.
- **BODY_TYP:** This data element identifies a classification of this vehicle based on its general body configuration, size, shape, doors, etc. It is derived from vehicle.csv.
- **BODY_TYP2**: This data element record what the type of the car derived from BODY_TYP. As the original BODY_TYP attribute has almost a hundred categories, it

is meaningless to display them all. So we manually group them into six categories. 01 to 09 are labeled "Sedan/Hardtop/2-Door Coupe"; 10 to 19 are labeled "Utility"; 20 to 29 are labeled "Van"; 30 to 49 are labeled "Light Vehicle"; 50 to 59 and 80 to 99 are labeled "Others"; 60 to 79 are labeled "Truck". Particularly, 50 to 59 are labeled "Bus" initially, but we concatenate “Bus” to “Others” as “Bus” is only a small proportion, which would probably cause data deficiency for visualizing.

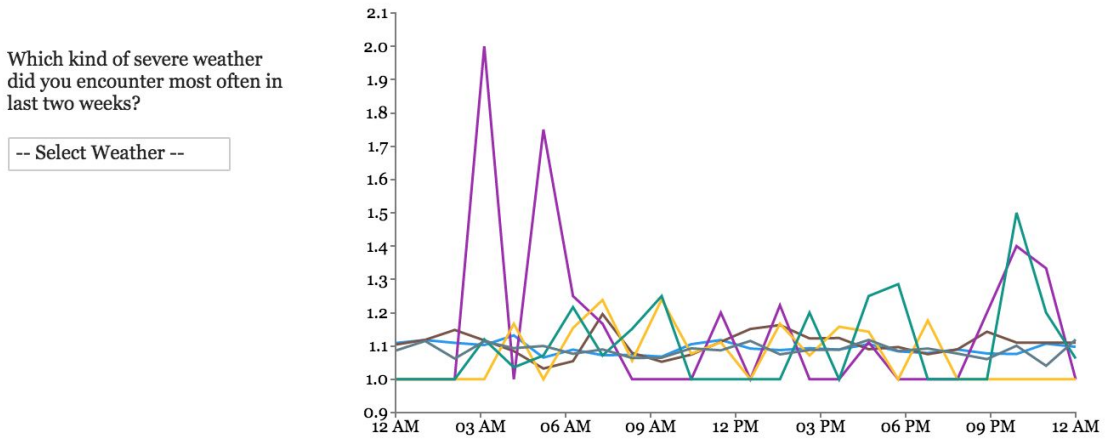
- **FATAL_COUNT**: In each crash, the total number of fatal death, derived from person.csv. The value is obtained by counting death people in INJ_SEV in each accident ignoring injured people.
- **FATAL_LEVEL**: In each crash, there has three levels of fatal death (“1” for one death, “2” for two deaths, “>2” for more than two deaths), which is derived from FATAL_COUNT.
- **WEATHER**: This data element records the prevailing atmospheric conditions that existed at the time of the crash, which is derived from accident.csv. Although there are four attributes for describing climate in accident.csv as there can be more than one weather in the same day, we use the first WEATHER attribute to define the climate for simplification consideration.
- **STATE**: This data element identifies the state in which the crash occurred. The codes are
  from the General Services Administration’s (GSA) publication of worldwide Geographic Location Codes (GLC).
- **SPEEDREL**: This data element records whether the driver’s speed was related to the crash as indicated by law enforcement, which is derived from vehicle.csv.
- **MDRDSTRD**: This data element records whether drive has distract at the time of the crash, which is derived from distract.csv.
- **SPEEDREL2**: This element records whether the driver has speeding or not, which is derived from SPEEDREL. As SPEEDREL describes different speeding level which is not required in our visualization, we treat 0 as “not speeding”, all other non-zero case as “speeding”.
- **MDRDSTRD2**: This element records whether the driver was distract or not, which is derived from MDRDSTRD. As MDRDSTRD describes different types of distraction, which is not required in our visualization, we treat 0 and 99 as “not distracted”, all other values as “distracted”.

3.3 Data Flow

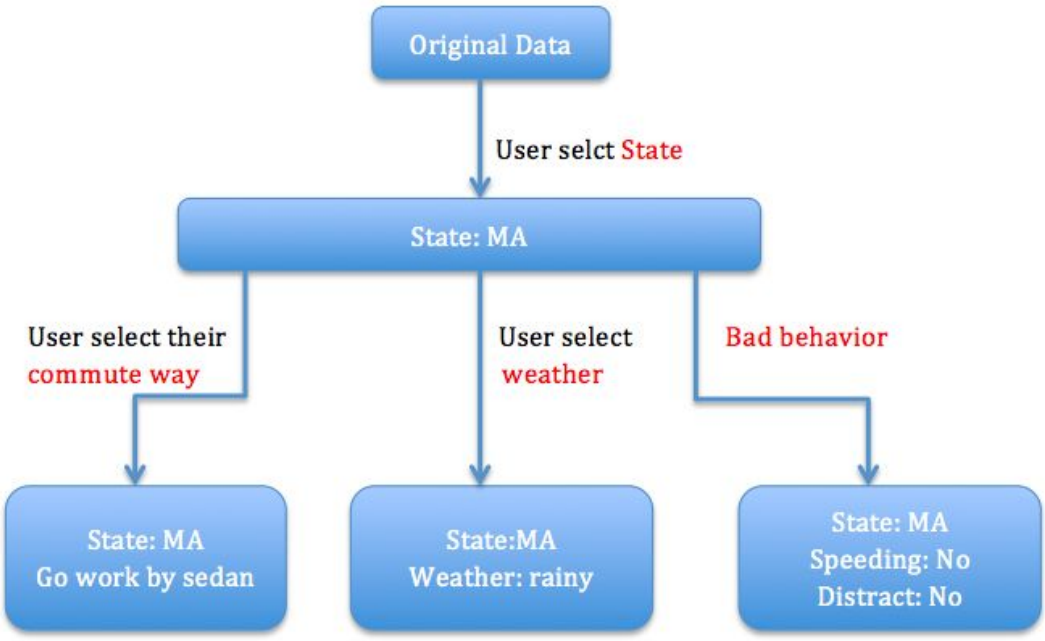
In order to strengthen the sense of participation of the user throughout the exploration process, the data flow is initially designed in a sequence structure with filtering conditions in each joint. For example, dataset used for commute type visualization is filtered by state which is identified by former visualization. The sequence-structured data flow can be described as the following.



However, with more filter conditions applied, less data can be displayed. To be more specific, for example, the multiple line graph of weather will appear gaps in most time of day after data is filtered by state and type of vehicle(many zeros, no pattern to be discovered at all). It could be worse and intolerable in the last visualization.



Therefore, we adjusted the data flow in final visualization design to a two-level hierarchical structure. The last three visualization performed based on data filtered by state, that is, instead of hierarchy relationship, the last three visualizations are parallel relationship to ensure data sufficiency.



In fact, if you do not select any specific state at the beginning, the last three visualizations are performed based on national dataset as the default option.

For the whole website, we concentrated on **fatal crash rather than normal accident.**

Another thing worth mention is that, in the last three views (our main views), we basically use the idea of average death toll per accident.

How we understand average death toll per accident? Here are some possibilities.

- A accident more likely involved more than one car.
- A fully concentrated driver were involved in accident due to others fault.

- The driver are all fully concentrated, however, due to the bad circumstance, rain, fog, high gravity car which are more likely to roll over.

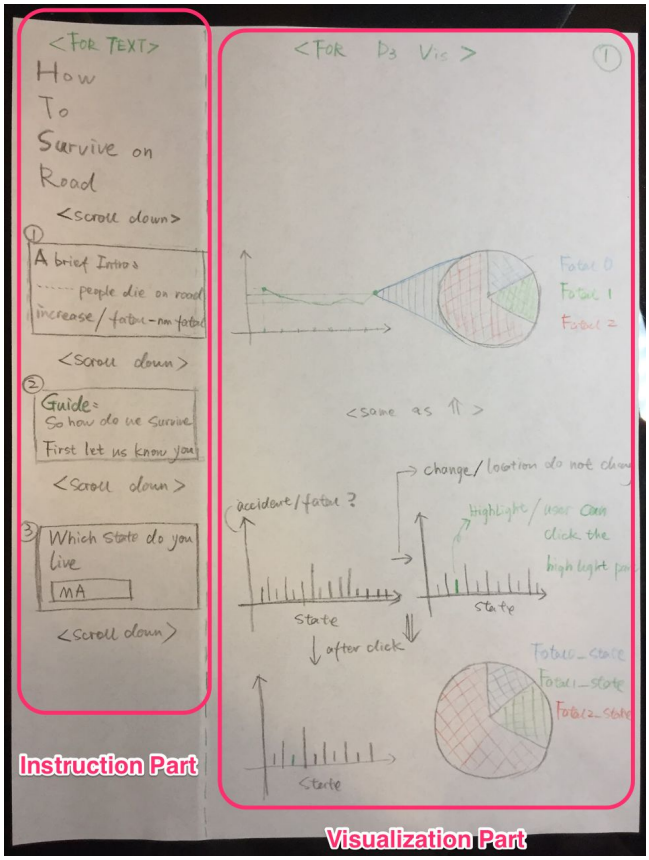
4. Design Evolution and Implementation

4.1 Proposal Design

The initial website design mentioned in the proposal which is shown below includes two parts: instruction (left area) and visualization (right area).

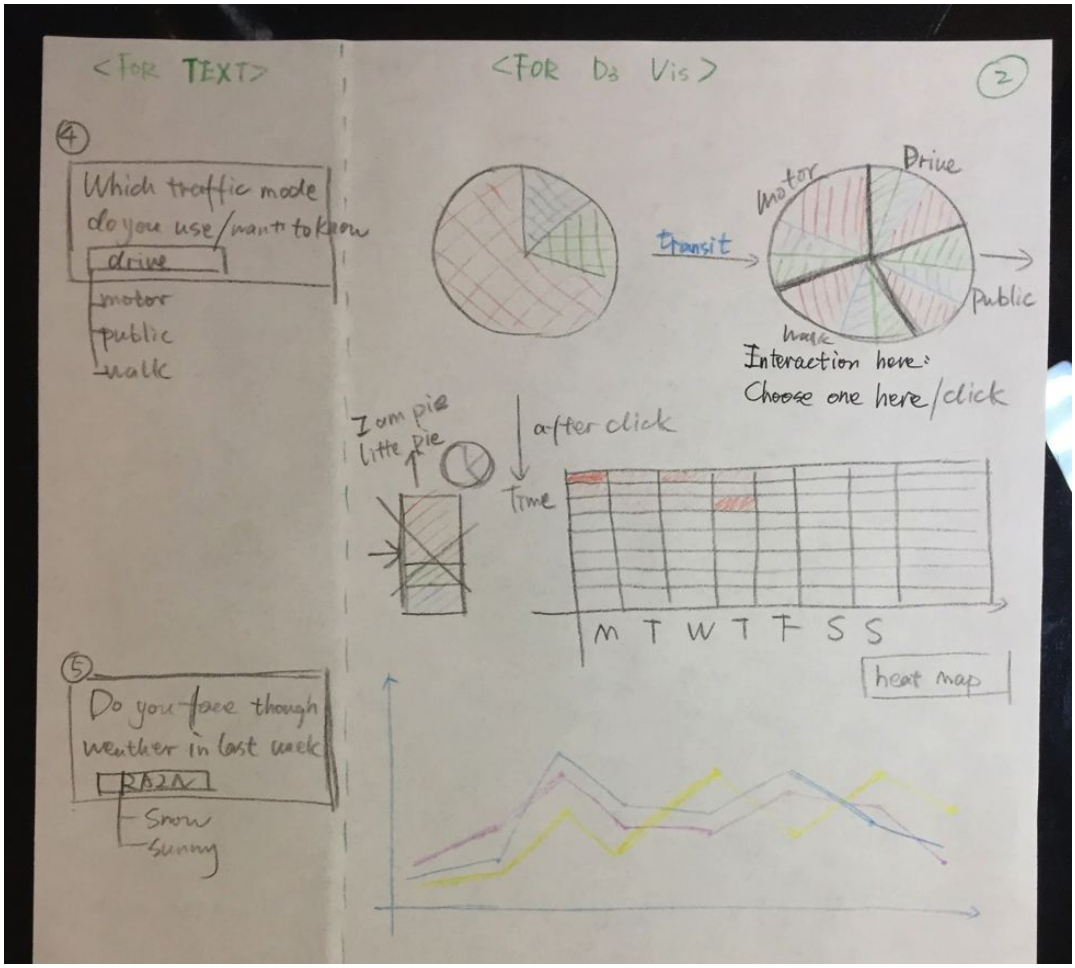
The first visualization shows the trend of number of traffic fatalities over around a hundred years. A pie chart displaying the proportion of different severe levels of crashes occurred in 2015 will appear after the user click the twinkling dot for 2015. This visualization is aimed to provide a macroscopic view of traffic fatality of 2015.

The personal exploration process starts from the second visualization. Before any interaction, a bar chart of traffic death tolls over states is shown. After the user specifies a state from the dropdown options, the corresponding bar will be highlighted and a corresponding pie chart displaying the proportion of different severe levels of crashes occurred in that state will appear.

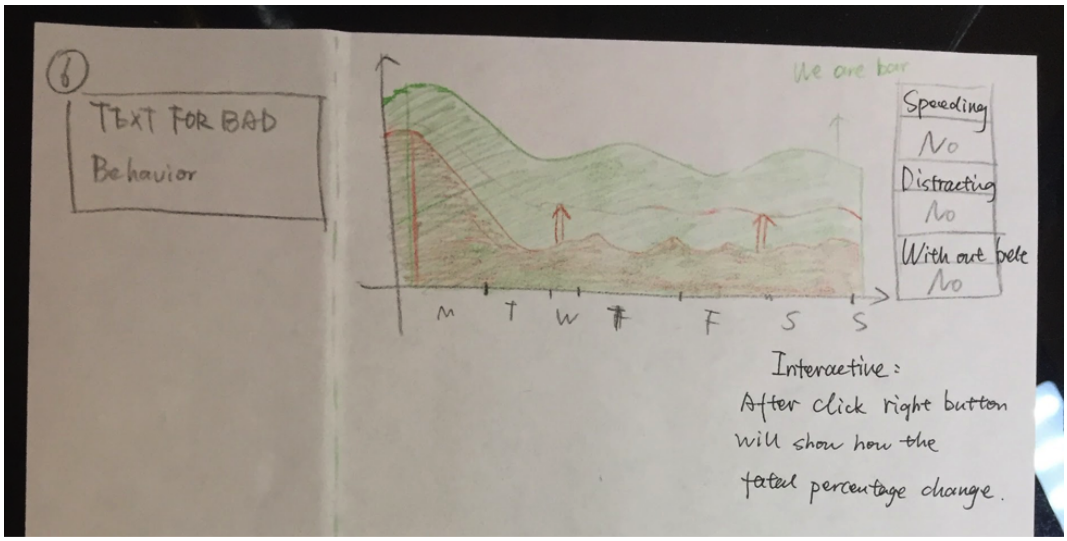


Then goes to the third visualization. Before any interaction, a pie chart displaying the proportion of different severe levels of crashes occurred in the specified state is shown. By clicking a transition button, a sub-pie chart displaying the corresponding proportion of different types of vehicle for each severe level. After the user specifies a type of commute type from the dropdown options, the corresponding heatmap of traffic death tolls will appear.

Next is the fourth visualization. Before any interaction, a multiple line chart of traffic death tolls for different weathers. After the user specifies the weather from the dropdown options, the corresponding line will be highlighted and other lines will be blurred.



The last visualization is a bar chart of average death per accident without any bad behaviors over times of day in weekdays by default and also performs as a benchmark. The three buttons on the right side are parallel and can be overlaid. After making any new combination of bad behaviors by pressing the buttons, a new group of bars will appear.

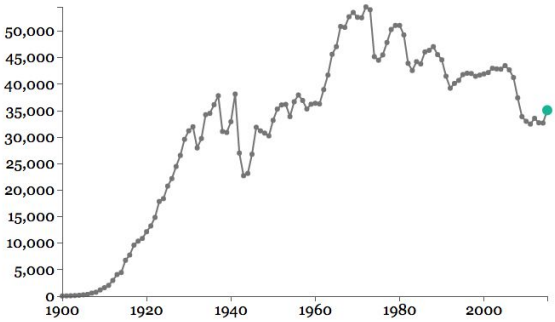


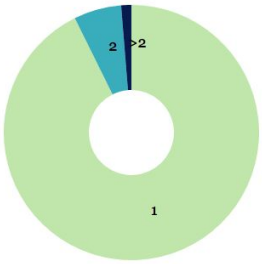
4.2 Prototype Design

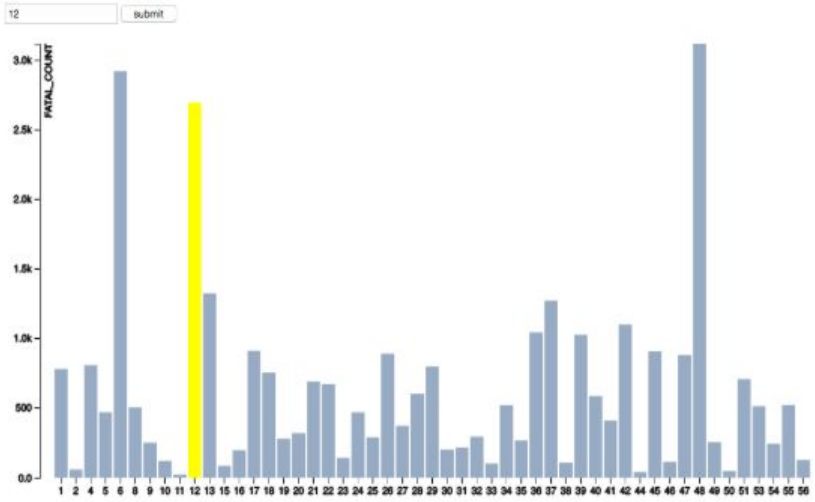
In the prototype, the first two visualizations are almost the same with the initial design.

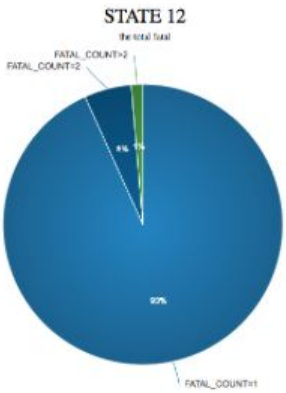
How
To
Survive On
Road

In recent years, many safety actions helped decrease the number of traffic deaths. However, newly released government data point that the number of people who died in car accidents increase 7.2% over 2015.

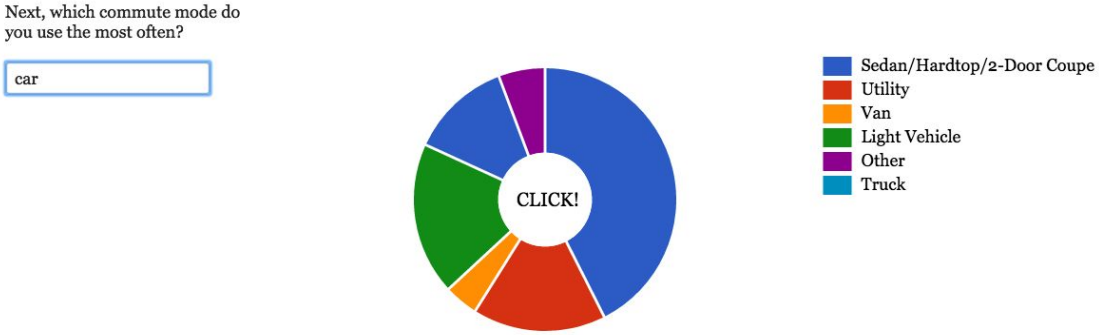


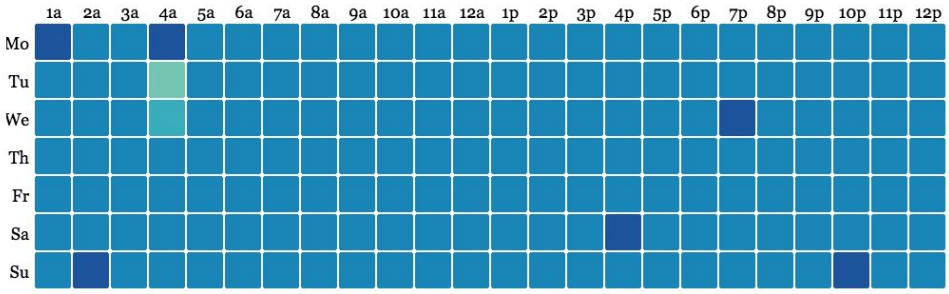




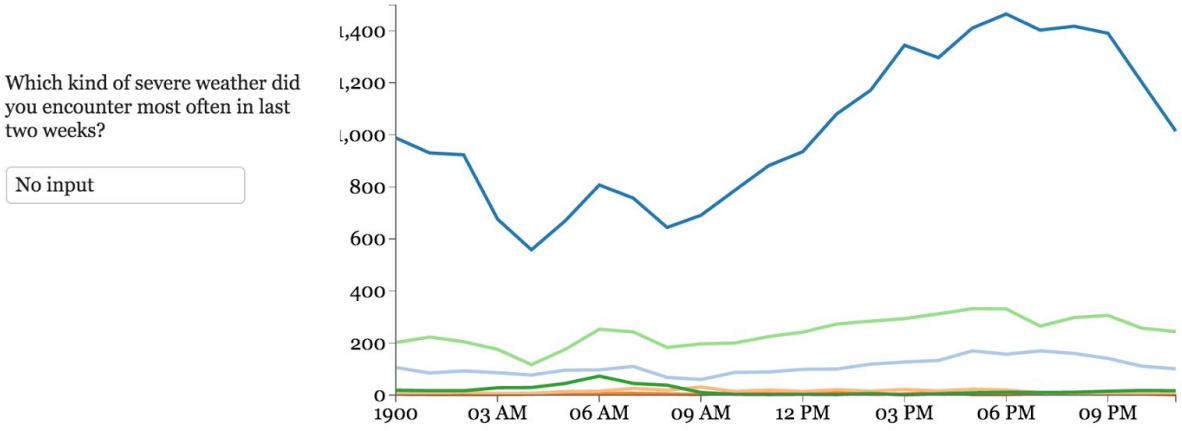


For the second visualization, first, instead of displaying the proportion of accidents of different severe levels of traffic fatalities, the pie chart displays the proportion of accidents of different types of vehicles. Second, instead of displaying the sub-pie chart inside the original pie chart, a sub-donut chart will appear outside the original pie chart after user clicking “CLICK!”. Other features are the same with the initial design.

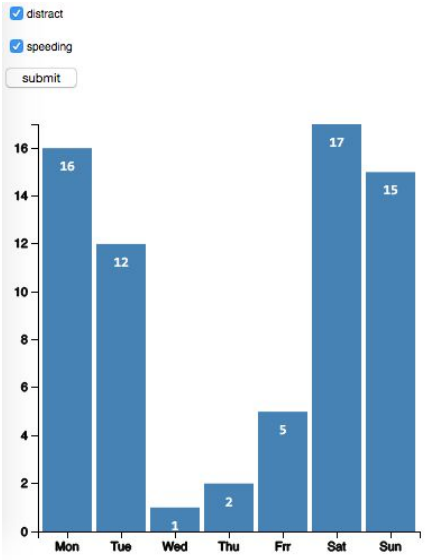




The fourth visualization is also basically identical with the initial design.



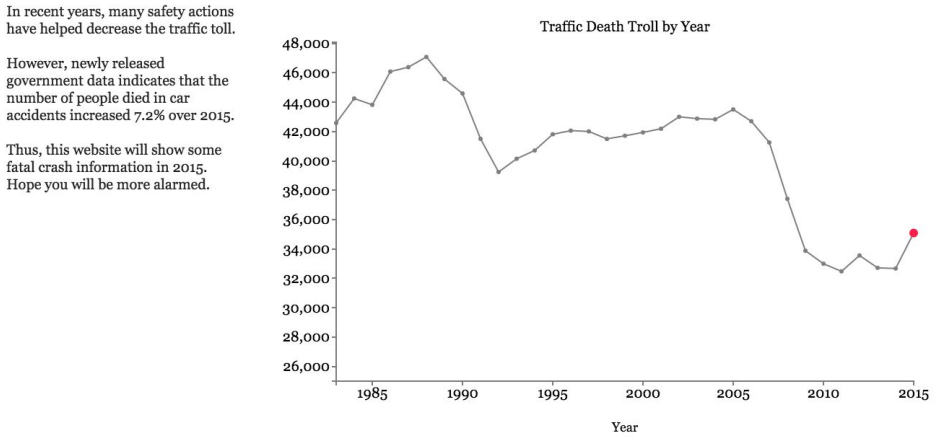
The final visualization is under developing in the prototype design.



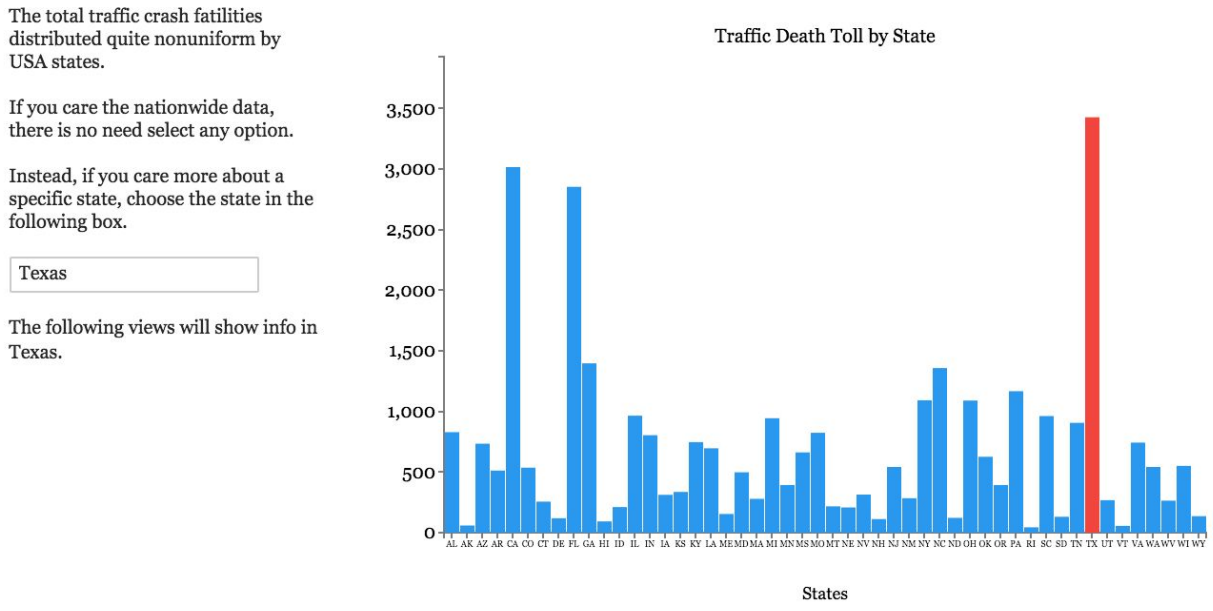
4.3 Final Design

First of all, we provide a general perspective of U.S. traffic fatalities over years. In our final design, we want users focus on the trend in recent 30 years, so data in even earlier time is

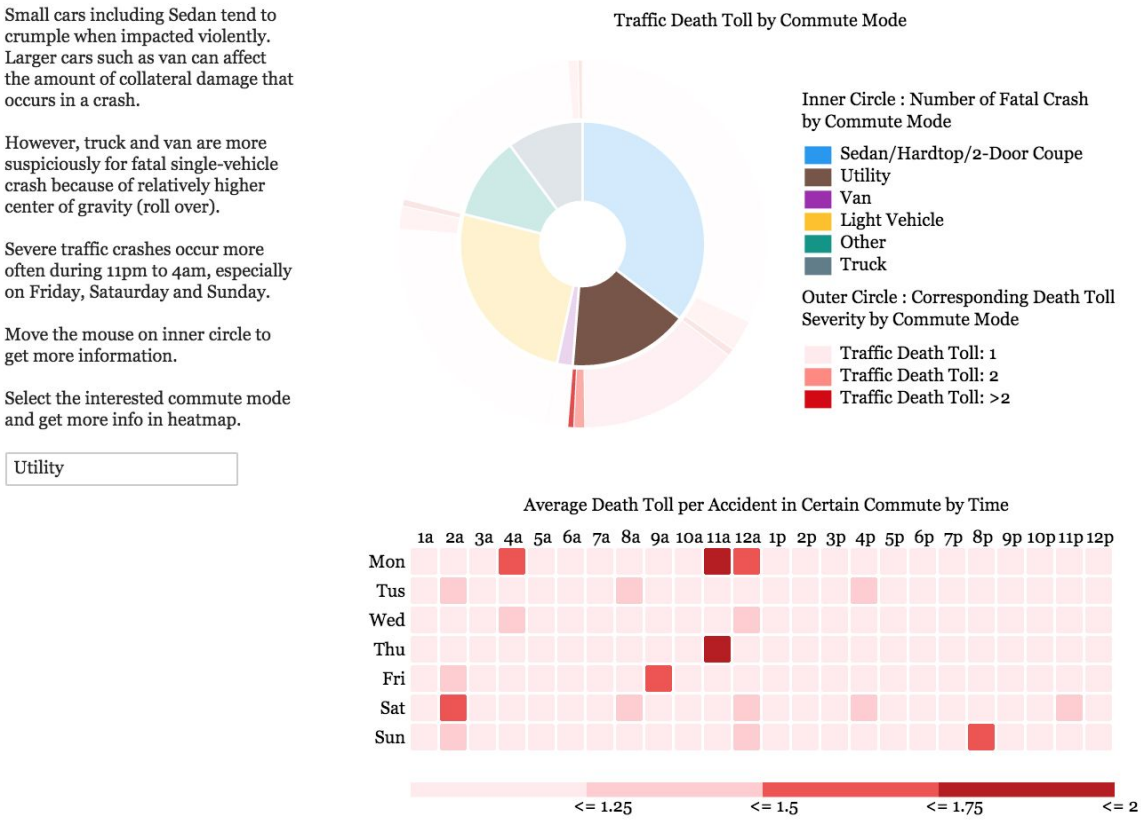
removed. Besides, the original pie chart is meaningless after further consideration, we remove it in final design.



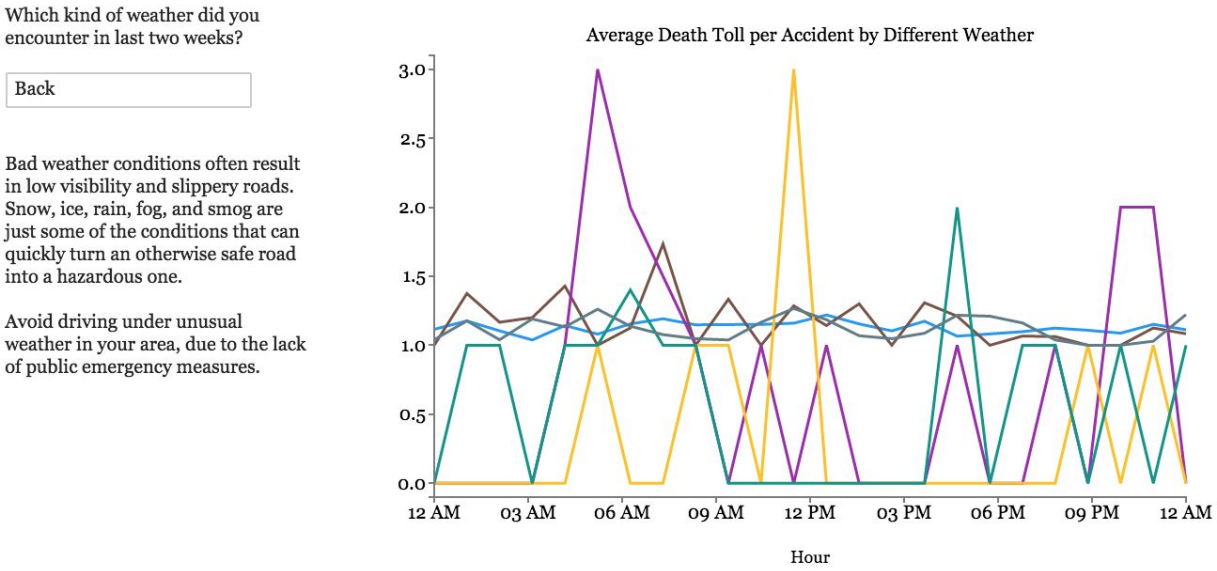
The second view is generally identical with prototype design. After a state specified, the corresponding bar will be highlighted. Then, a line will appear under the select down box, remind the user of which state he/she will explore. If you select other state, this highlight and text will also change corresponding.



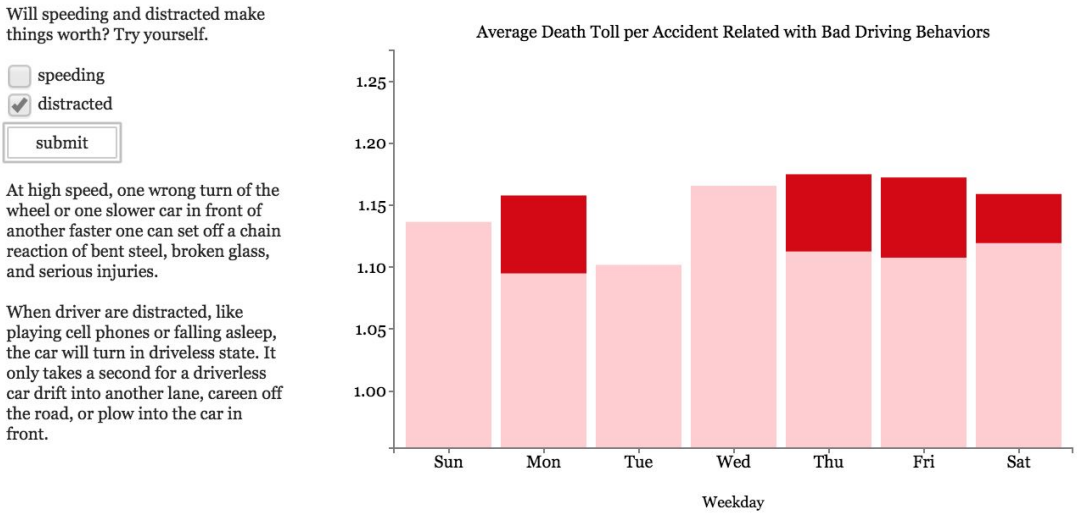
As for the third view, the click event was removed for simplification consideration. The outside donut is shown with the inner donut. Besides, instead of traffic fatalities, the heatmap displays the average fatality per accident in order to deliver that at what time people would likely to encounter a destructive crash.



As for the fourth view, we will asked users choose one weather category. At first, the multiple line chart shows the trend of traffic fatalities over the time of day by weather. After user clicking a weather category they are interesting, others will be blurred. In our prototype design, the y-axis represent the total fatalities tolls. However in final design, we decided to shows the average fatalities in every accident to let users focus on the trend of traffic fatalities over the time of day by weather. Furthermore, we modify the axis range dynamically, to help show more detail of pattern. As for some missing data after filter, group and reduce, we put zero in this missing place. This modification will still help user to find certain pattern, especially for pattern under unusual weather conditions.



Finally, users will select whether they have speeding or distracted. The default option is driver has speeding. This design is to remind the user this view has interactive function and encourage user to use the function. The pink bar represent the average death toll per accident without driving bad behaviour? Compare with our prototype visualization, despite the modified detail such as colors and legend, the most modification is the dynamic change of axis range. This modification highlight the change after select certain bad behaviour.



5. 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?

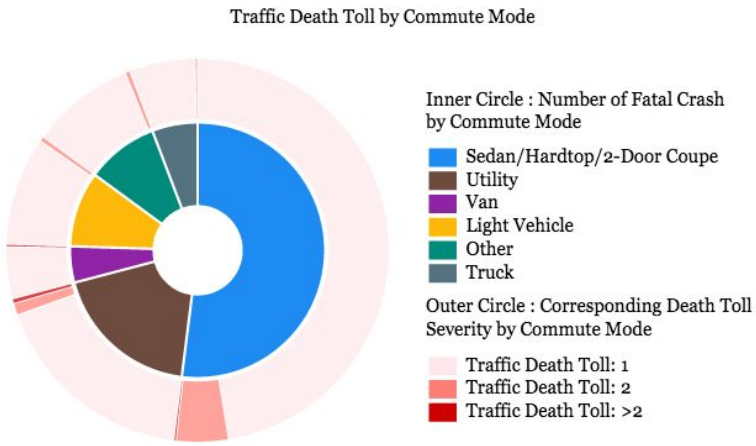
- What did you learn about the data by using your visualizations?
    - Small cars including Sedan tend to crumple when impacted violently. Larger cars such as van can affect the amount of collateral damage that occurs in a crash.
    - However, truck and van are more suspiciously for fatal single-vehicle crash because of relatively higher center of gravity (roll over).
    - Severe traffic crashes occur more often during 11pm to 4am, especially on Friday, Saturday and Sunday.
    - Bad weather conditions often result in low visibility and slippery roads. Snow, ice, rain, fog, and smog are just some of the conditions that can quickly turn an otherwise safe road into a hazardous one.
    - Avoid driving under unusual weather in your area, due to the lack of public emergency measures.
    - At high speed, one wrong turn of the wheel or one slower car in front of another faster one can set off a chain reaction of bent steel, broken glass, and serious injuries.
    - When driver are distracted, like playing cell phones or falling asleep, the car will turn in driverless state. It only takes a second for a driverless car drift into another lane, careen off the road, or plow into the car in front.

- How did you answer your questions?
    - First, how we understand average death toll per accident? Here are some possibilities. (Repeat part with some part of the data flow, for easy reminder)
        - A accident more likely involved more than one car.
        - A fully concentrated driver were involved in accident due to others fault.

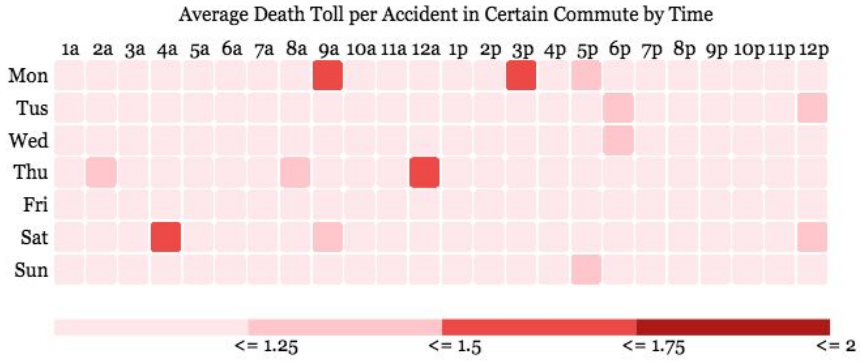
- The driver are all fully concentrated, however, due to the bad circumstance, rain, fog, high gravity car which are more likely to roll over.

- For weather and bad behaviour view, it's more obvious how to use view s to answer questions. So let's talk about Death toll and commute mode. Use the data from New Jersey as example.

  - The darker pink part of the outer donut of sedan is larger than the others, it can show 'Sedan tend to crumple when impacted violently'. The same idea for truck and van.



  - The heatmap show when did the more destructive accident will occur, more dark, more severity.(We want to change this to median, since there will be super severe crash and affect many. Though forgot this...)



- How well does your visualization work?
    - As for the code, we are confident at the robust of the code, test many times.(While not 100% confident, since each time add a new function, other view need change because they are all linked)
    - As for the detail and style of the whole website, we are confident as well.
    - As for the data flow, we want to make it more simple and straightforward. We think a scroll down order of the website will not make the user feel too complex of the website. Well when talks about many 'average' level data at the following states, we discuss quite a lot about how to make it more simple, more intuitive, but at the end, there still not a good solution.
    - Another thing worth mention is that we add many 'instruction' for users, like legend or instruction about when you should select and mouse on. We try to make those words on a obvious place while not harm the appearance of the views.
- Further improvement
    - Still the data problem, the average death toll per accident is not that straightforward. Though Traffic Death Toll is the main topic of our website.
    - We want to show the data flow to the user, not have a good idea right now.
    - Actually, at first, we want to make a traffic lore database, and show random correspond ones when user select certain option(like something will show at the video game loading screen). Like when user select Texas and choose weather snow, it will appear something like 'unusual weather should combine with high cautious '. That's another problem.
    - The loading time is a bit slow, but we have to load all data in. Do not know how to solve this problem.

6. Reference

http://www.injuryclaimcoach.com/multi-vehicle-accident.html
http://www.ohiotiger.com/suvs-vs-cars-crash-survival/

http://journalistsresource.org/studies/environment/transportation/comparing-fatality-risks-united-states-transportation-across-modes-time