Data Visualization About Traffic Accidents

Qianqian Tang, Yixiang Cao, and Haonan Xu

DSCI 554 Project Report

1 Introduction

Traffic accidents have become a serious social problem worldwide. The consequences of heavy traffic accidents have forced people to attach great importance to traffic safety and apply the continuous advancement of science and technology to the research work of traffic safety. So it is necessary to analyze the data of various dimensions of existing traffic accidents, and visualization is a very effective analysis method.

Our visualization is provided for anyone who cares about the cause of a traffic accident, such as drivers, tourists, government workers, etc. From our visualization, they can judge traffic safety risks based on their own situation.

We mainly used D3 library [1] to draw graphs and used Vue.js to construct our system. Unlike some existing visualizations, we have established a linkage between the map and other charts, which will make our system more intuitive and user friendly.

in section 2 we present related visualization work, and compared it with our visualization. In section 3 we discuss the data sources we used, in section 4 we explain our design process and considerations, in section 5 we show the main aspects of the system we have built and finally conclude with 6).

2 Related Work

In our research, we expected to find out several important features other than the number of accidents that are relevant to car accidents. Prior to working on our project, we got some inspiration from visualization examples of others' works in order to best utilize these data and stand out from our own project. For example, other visualizations have used maps to show the number of accidents in specific regions, and created a timeline to reflect the trend on the start time of accidents in 24 hours, and drew a proportional area to show how weather conditions are related to accidents.

While others‘ works tend to treat this dataset more as time-series data, we managed to do more on the causes and other features of each accident itself. The original part of our project is that we integrated maps and line charts so that we can look at the number of accidents among states and get more details when choosing a specific state. To show the relevance of accidents and weather conditions, instead of using the total number of accidents versus every weather condition, we have an interactive bar chart to look at how each weather condition affects the number of accidents in each state.

3 Data

Our dataset is from the Kaggle database. It is a countrywide car accident dataset, which covers 49 states of the USA. The accident data are collected from February 2016 to June 2020, using two APIs that provide streaming traffic incident (or event) data. These APIs broadcast traffic data captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. There are about 3.5 million accident records in this dataset. [2] [3][4]

The dataset contains 49 columns with features of accidents recorded. The features consist of location information (such as state, city, and coordinates), time information, weather conditions (such as temperature, humidity, and strength of winds), and other information like the severity of each accident.

To create a better visualization of our data, we first did some data cleaning and data retrieval jobs. For instance, when encountering missing values, we decided either to ignore them or use a mean value to replace them. Because of the nature of data visualization, we need to retrieve the subset data from the entire dataset constantly. Thus, using a python Jupyter notebook, we did all the above preprocessing jobs and other works like selection and aggregation as needed.

4 Approach

Many factors affect traffic safety. After discussion, we decided to analyze these factors from four dimensions: Drivers, Time&Date, Location and weather.

For Time&Date, what we have is the exact time of each accident, such as "2018-01-01:03:00:00". We designed to use a multiple-line chart to represent the number of car accidents at various times of the day and use a calendar chart to represent the number of accidents on each day so that users could know the traffic safety risks' time feature and date feature.

For Infrastructure, to visualize location-related data, maps are indispensable. We designed to use Mapbox to draw a map of the United States and mark the average severity of accidents in each county. What's more, for the data which shows the feature of location such as "Crossing", "Traffic_Signal", etc. we designed to use Pie chart to represent the ratios of different features.

For Weather, We have the temperature and visibility of each accident. In order to represent the relationship between them and the probability of accidents, we designed to draw a multi-dimensional bar chart, which also supports sorting by each dimension.

For Population, the problem we encountered was that there was not enough information about the driver in our data source. So we first want to represent the relationship between population and accidents. So we designed to draw a choropleth map to represent the population and number of accidents in each state and use linked scatter plots to show their relationship. And for the age of the driver in the accident, we used the bar chart to show it.

5 System

We used Python to do the data wrangling first. We used HTML, CSS, and javascript, Node.js to develop our website on Vue App. To draw the chart and d3 map, we mainly used d3.js. To draw the Mapbox map, we used mapbox-gl.js, mapbox-gl.css, and vue-mapbox.js.

We divided the website into different pages based on the topic we analyzed. We analyzed the relationship between traffic accidents and time, the relationship between traffic accidents and weather, and the relationship between accident severities and infrastructures. Most of our charts are interactive. The d3 choropleth map we made is able to interact with a bubble chart and a line chart simultaneously. The content of the map is also changeable. The Traffic Severity Mapbox choropleth map interacts with an animated pie chart. We made highlight and tooltips for the charts. Additionally, information on the Mapbox map in different layers can be shown by zooming in and out. We also have a bar chart that is animated, interactive, and responsive.

5.1 Time&Date&Population

Below is the screenshot of this section. There is a d3 map on the top , which is a choropleth map representing the number of accidents in different states. And from the type selector, user could select to see the choropleth map representing the population in each state.

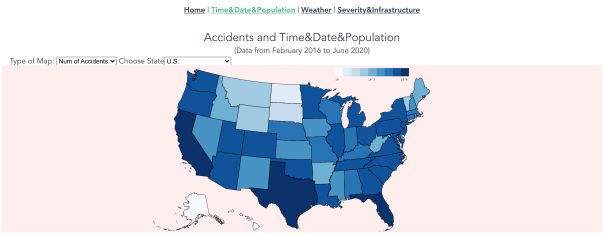

Fig. 1. Figure of D3 map

And here is a scatter chart, whose x-axis is population and y-axis is accident. This scatter graph is linked to the map and the state selector.And in the right t is a timeline chart. whose x-axis is time zone in one day, and y-aixs is the number of accidents. Also this time line chart is interactive with map and state selector.

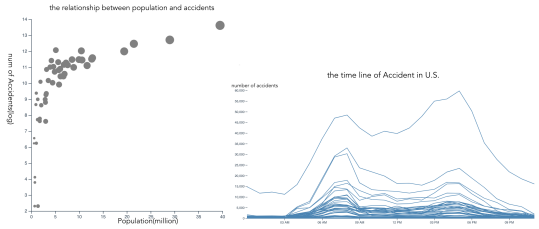

Fig. 2. Figure of scatter and time line chart

And we used calendar to represents the number of car accidents that occurred during that day.

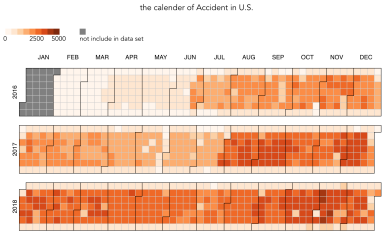

Fig. 3. Figure of Calendar chart

5.2 Weather

When analyzing car accident data, the weather condition is necessary to be taken into consideration. This is a grouped bar chart that shows the relationship between the number of accidents, temperature, and visibility. These data collected from 2016 to 2020 are presented in three different sorted order. Users can press each button to change the sorted order. To look into the details of weather conditions in each state, one can hover over a bar to see the value of each of the data.

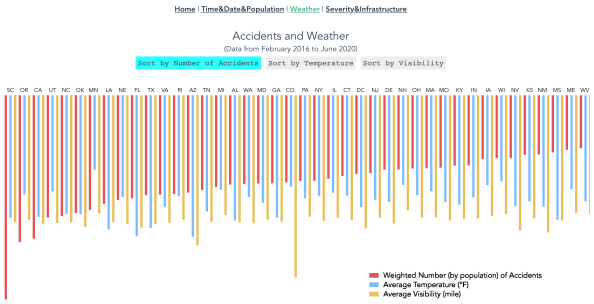

Fig. 4. Figure of Weather section

5.3 Severity&Infrastructure

We draw a Mapbox map that shows the severity of traffic accidents of the states and the counties in the US. And the pie chart shows the percentage of the appearance of the 10 different infrastructures when accidents happened in the US in the 5 years.

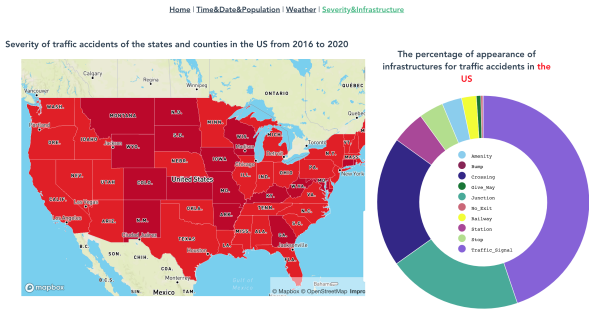

Fig. 5. Figure of Mapbox map and Pie chart

The bar chart shows the count of traffic accidents that happened near these 10 infrastructures in the 5 years with 4 different severities.

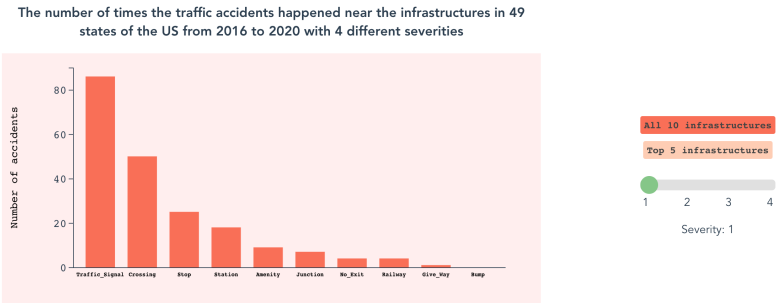

Fig. 6. Figure bar chart of infrastructures

6 Conclusion

As for contribution, Qianqian Tang was responsible for analyzing the relationship between traffic accident severity and infrastructures. Qianqian Tang made the Mapbox choropleth map, the animated pie chart, and the interactive animated responsive bar chart. Yixiang Cao was responsible for analyzing the relationship between traffic accidents and time and population. Yixiang Cao made the interactive d3 choropleth map, timeline chart, bubble chart, and d3 calendar chart. Haonan Xu was responsive for analyzing the relationship between traffic accidents and weather. Haonan Xu made the interactive multi-bar chart.

To comprehensively analyze the relationship between traffic accidents, time, environmental stimuli, and road conditions, we would combine all the features to make a confusion matrix. Besides, we would randomly select data to draw dots representing the occurrence of traffic accidents to show the hotspot locations on the map. To predict accidents, we would use feature selection and then do the prediction using the data and show them in charts. What's more, we would use recent data to study the impact of COVID-19 on traffic behavior and accidents.

References

1. d3js. http://d3js.org, last accessed 10 Aug 2019
2. Kaggle:us-accidents. https://www.kaggle.com/sobhanmoosavi/us-accidents, last update 6 mounths ago
3. Moosavi, Sobhan, M.H.S.S.P., Ramnath., R.: "a countrywide traffic accident dataset" (2019)
4. Moosavi, Sobhan, M.H.S.S.P.R.T., Ramnath, R.: "accident risk prediction based on heterogeneous sparse data: New dataset and insights". In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,ACM (2019)