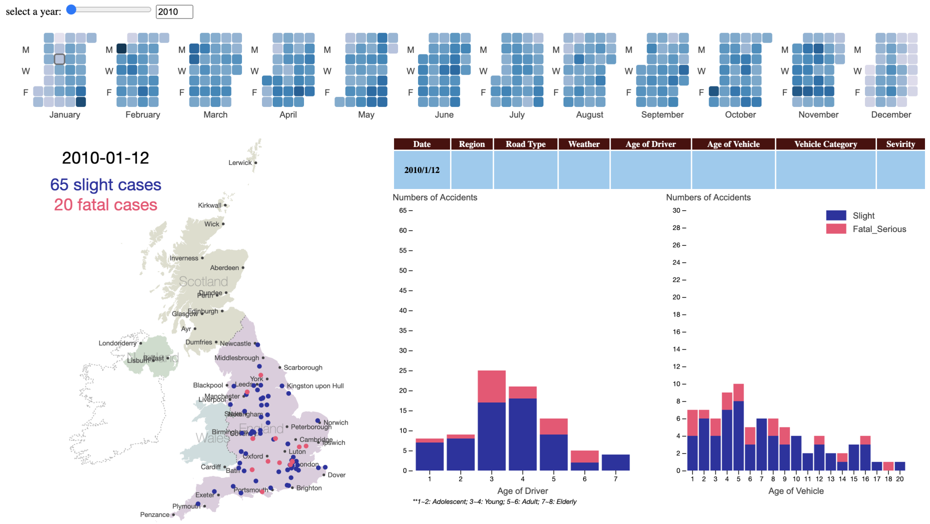
CSCI-SHU 235 Project Report - Group 8

Yuchen Wang (yw3642@nyu.edu), You Xu (yx1711@nyu.edu)

December 11, 2020

1 Overview

1.1 Screenshot of our visualization



select a year: 2010

2010-01-12
65 slight cases
20 fatal cases

Date	Region	Road Type	Weather	Age of Driver	Age of Vehicle	Vehicle Category	Sevirity
2010/1/12							

**1~2: Adolescent; 3~4: Young; 5~6: Adult; 7~8: Elderly

Figure 1: A screenshot of our visualization

1.2 Summary of our project

In this project, we designed a visualization of car accidents in the UK during the year from 2010 to 2014. With a large number of deaths in car accidents every year, measurements to improve road safety are essential. Our website-based visualization is targeted at the officers of the Department of Transport in the UK. The website contained four different views – calendar selection, map with interaction, multiple stacked bar chart, and information table – that fulfills the demands of our users. With the visualization, we expect the policymakers can plan for more effective measurements to prevent as many car accidents as possible.

2 Data

2.1 Description of our data

Our dataset is in tabular format. Technically, it is a *comma-separated values (.csv)* file. Each row corresponds to a particular case of traffic accident and each column corresponds to an attribute. It keeps a comprehensive record of the details of road accidents in the UK. This project is based on an excerpt of the data from 2010 to 2014. There are 251832 accidents recorded and each accident has 33 attributes. Some attributes are redundant and not very important. Specifically, we are interested in the attributes listed below:

Attribute	Scale/Cardinality	Description
Latitude	[49.915001, 58.893008]	the latitude where the accident happened
Longitude	[-6.530040, 1.758661]	the longitude where the accident happened
Year	[2010, 2014]	the year when the accident happened
Month_of_Year	[1, 12]	the month when the accident happened
Day_of_Month	[1, 31]	the day when the accident happened
Day_of_Week	[1, 7]	day of the week when the accident happened
Age_of_Vehicle	[1, 30]	age of the vehicle involved
Age_of_Driver*	[1, 8]	age of the driver involved
Region	11	the region where the accident happened
Road_Type	5	the type of road where the accident happened
Weather	6	the weather when the accident happened
Vehicle_Category	6	the type of vehicle involved in the accident
Accident_Severity	2	whether the accident was fatal or not

*The age of the driver is normalized by the data publisher. 1 2: Adolescent, 3 4: Young, 5 6: Adult, 7 8: Elderly.

2.2 Source of our data

Our data is downloaded from a public dataset on Kaggle (`https://www.kaggle.com/stefanoleone992/adm-project-road-accidents-in-uk`). It a is cleaned version derived from the original data provided by the UK government (`https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data`)[1]. The map of the UK is downloaded from `https://martinjc.github.io/UK-GeoJSON/`.

2.3 Data preprocessing pipeline

The original data is *Accidents_categorical.csv* (80.6MB). It is a little bit large for *d3.js* to process. We use the popular data processing package *Pandas* in *Python* to preprocess the data. We first remove redundant attributes and derive a much smaller data file *data.csv* (25.0MB), which is more efficient for *javascript* to process. Then we group the data by day and sum up the number of accidents. The result is saved in *dates.csv* (29.0KB). This process is beneficial to the calendar filter, which will map the number of cases each day to color hue. By doing so, the browser will not need to do the summation before rendering the view and the runtime will be improved.

3 Goals and tasks

Assume Mark is an officer of the Department of Transport in the UK. Instead of generally paying attention to the cases that happened throughout the whole year, he is more interested in a specific day of the year so that he can do a comparative analysis of these cases. Therefore, we design a calendar view for the users to select years by toolbar and select a specific day in a year by many calendar blocks organized by calendar orders. Each rectangle block represents a day and all the days in the same month are organized together. With notation, users can also tell which day it is easily from the calendar view directly.

In addition, Mark also cares about the seasonal trend of accident numbers throughout the year so that he can pay more attention to the days that cases happen more often. To make this information intuitive, we add color saturation to the calendar blocks. Darker indicates more cases. The specific number of the day will also display on the screen when he clicks to select the day on the calendar.

Also, Mark wants to know the geographical distribution and density of car accidents that happen in the UK so that he can arrange more resources to the areas that cases happen more often. So, we plot each of the accidents that happen in a day onto the national map of the UK. Each dot represents a case. In this way, users can easily tell the areas that the cases cluster.

Besides, Mark also wants to do statistical cause and effect analysis on the factors that may result in car accidents. With this information, Mark can apply more regulations and pay more attention to those that significantly impact the possibility of a road accident. We build multiple stacked bar charts about two factors – age of the driver and age of the vehicle – as a tool and reference to users to achieve this goal. With ordered causal factors on the x-axis and the numbers of accidents on the y-axis, users can intuitively tell the ones that correspond to more cases. Case severity is split in the y-axis.

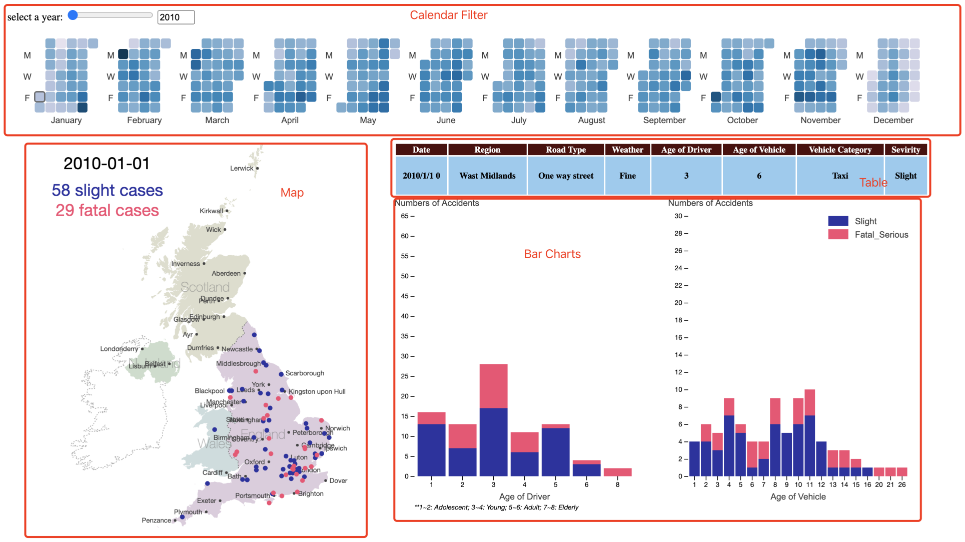
It is also very essential for Mark to know the severity of an accident since a more serious case is more likely to cause death. We tell the slight case and fatal serious case apart in both map view and stacked-bar chart view by different color hues. Blue represents slight cases while red represents fatal serious cases.

For some very special case, other information that is not included in the graphs are also important for Mark. Therefore, we put this information into a simple and straight-forward table. When the user hovers over a point on the map, the corresponding additional information about this case will automatically be displayed in the table. This will further help them with policy-making decisions.

4 Visualization

4.1 General view

There are four sections in our design (figure 2): the calendar filter, the map, the bar charts and the table. All these sections are linked together through interactions. When the user select a day on the calendar, the corresponding traffic accident cases are displayed on the map below and the statistics are shown in the bar charts. Moreover, when the user hover over a point on the map, the detailed information of that case will appear on the table.

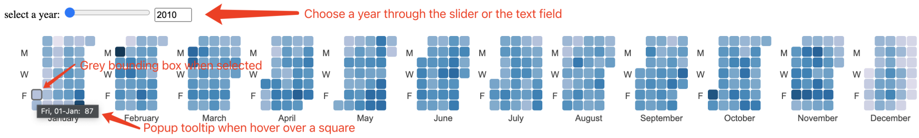


Date	Region	Road Type	Weather	Age of Driver	Age of Vehicle	Vehicle Category	Sevirity
2010/1/1 0	Wast Midlands	One way street	Fine	3	6	Taxi	Slight

Figure 2: Four sections of our design

4.2 Calendar Filter

A slider on the top allows the user to select a specific year. The calendar the channel color hue to encode the number of cases on each day throughout the year selected. A darker square stands for more cases on that day. When the user hover over a square, he will see a popup tooltip showing the date and number of cases. When he clicks, the map and the stacked bar chart will change correspondingly. This design is similar to the contribution calendar on *Github* (figure 4), by which the user can get an intuitive sense of the year's data. Calendar-like filter is easy to use, easy to get started.


select a year: 2010

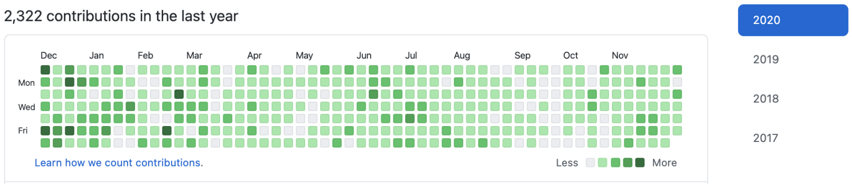
Choose a year through the slider or the text field

Grey bounding box when selected

Fri, 01-Jan: 87

Popup tooltip when hover over a square

Figure 3: The calendar filter in our design


2,322 contributions in the last year

Learn how we count contributions.

Figure 4: The contribution calendar of Glenn Jocher, the executive developer of YOLO-v5, a SOTA object detection algorithm

4.3 Map

On the map (figure 5), different regions are encoded with different colors. Slight and fatal cases are represented by blue and red point respectively. On the top left corner, there is a general statistic illustrating the date and number of cases. Through the map the user can easily know geographical distribution of all cases on the day he selected.

4.4 Stacked bar charts

The stacked bar charts (figure 6) shows the distribution of number of cases over the age of the driver and the vehicle. The horizontal position of the rectangles encodes the age. And the length of the rectangle encodes number of traffic accidents. The color saturation encodes accident severity, and it is consistent with the map: red for fatal and blue for slight. When the user hover over a component of the chart, a tooltip will popup showing the detailed information. The rationale of this design is that bar charts can precisely and intuitively illustrate quantitative information.

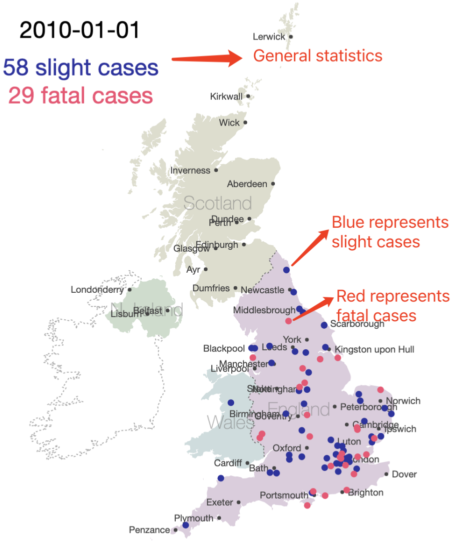


Figure 5: The map in our design

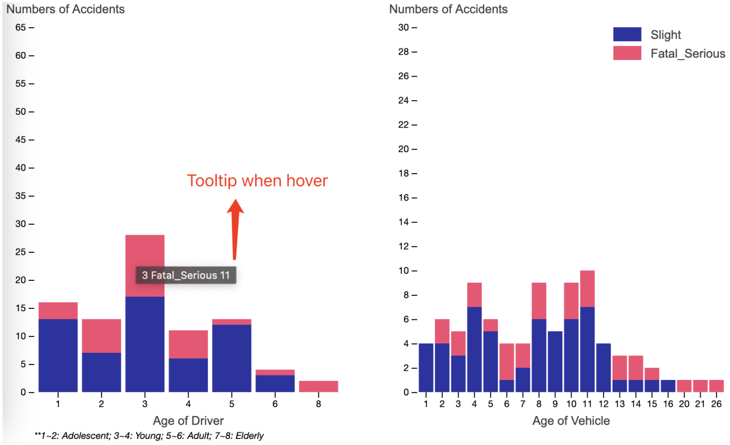
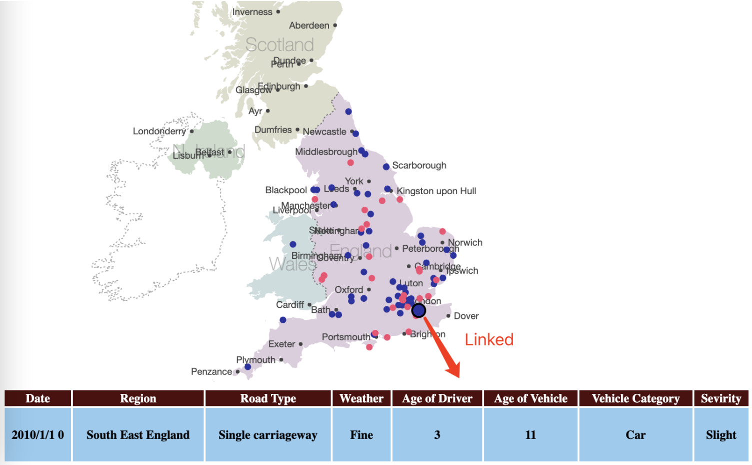


Figure 6: The stacked bar charts in our design

4.5 Table

The table (figure 7) is linked with the map. When the user hover over a point on the map, the point will be enlarged and the detailed information corresponding the point is displayed on the table. The design is beneficial if the user what to dig deeper into the dataset and get some insight.



Date	Region	Road Type	Weather	Age of Driver	Age of Vehicle	Vehicle Category	Sevirity
2010/1/1 0	South East England	Single carriageway	Fine	3	11	Car	Slight

Figure 7: The table in our design

5 Reflection

Different from the work breakdown table we proposed, we built the calendar filter first. Later, we add interactions on the calendar to display different materials on the other views. We then split our work into map view and chart view and each of us takes responsibility for either one. After three of the views are done, we add the information table view last which has an interaction with points on the map. At last, we spent much time constantly modifying the structure of our program to make sure the initialization and interactions between different views work well.

Due to the limited time that we work on the project, we dropped a few functions that are proposed in the proposal. There are zoom-in on the map, hover region on the map, multiple selections on the calendar. Besides the time limitation, there are also other reasons that we fail to achieve the hover and zoom-in feature. Not like China, the division of geographical areas in the UK are not quite unified. It is hard to find a reasonable way to split the areas of a country for us as non-UK residents. Moreover, the dataset only presents a general description of “location” so that it is even harder to categorize a location into a region only by its longitude and latitude.

Besides the functions that we proposed but not achieved, we think our visualization could have worked better with some additional features. Some possible future improvements are as follows. First, it would be better to allow users to choose which attributes they want to compare on the stacked-bar chart view so that they may know more about what they personally are more interested in. Second, it could look better if we add more animation to transitions. Third, by allowing users to switch between the stacked-bar chart and normalized stacked-bar chart. This would make more sense when comparing the percentage rate between having a slight case and a fatal serious case.

In all, we mostly follow the proposal and make some slight changes during the programming process. However, there is still space for potential future improvements to this visualization project.

References

[1] UK Department for Transport. Road safety statistics releases, 2020.