**Visualize Motor Vehicle Collisions**

**Team Members**

Xu Wu 109607168

Chongxuan Yin 110445665

**Introduction**

Do you know that there were 210,110 motor vehicle collisions in New York City in 2016. This means that 576 vehicle collisions occur every day in New York City. If you are a New York City citizen, do you think this will affect you? Of course the answer is yes. Thus, we get the data from <https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95> which contains all motor vehicle collisions in NYC that have occurred in the past few years. We will select only the data in 2018 and visualize the data to show the view of it.

**Technique**

1. Python
2. Javascript
3. D3

**Mission**

1. There are 1.48 million rows and 29 columns in the data, so we first should do data cleaning and data reduction.
2. Layout
3. User Interaction

**Approach and Implementation**

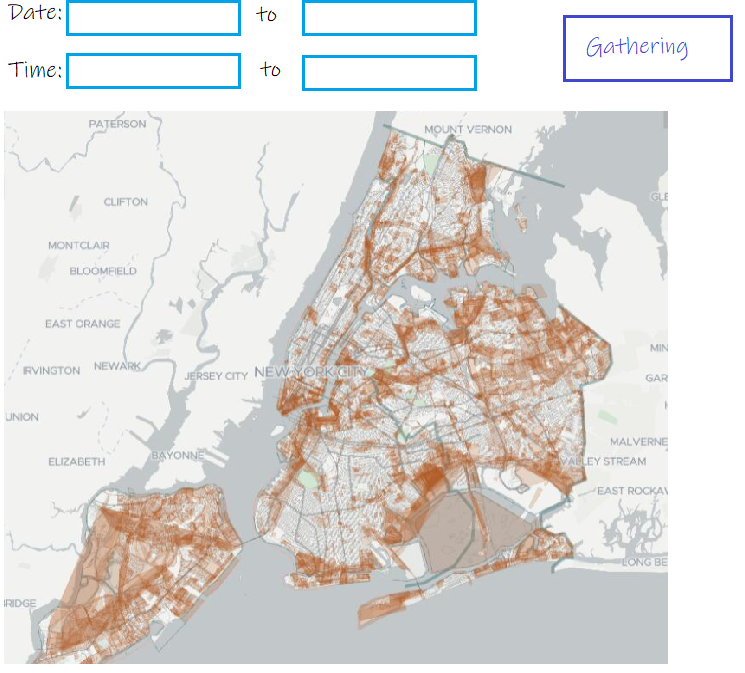
**Data Cleaning and Reduction**

1. Remove all the columns that are not in 2018.
2. Remove the attributes that we don’t care about or contain too many missing values by deleting these columns in EXCEL.
3. Since we have a large number of data, we remove the rows that do not have all the attribute values by filtering and deleting the rows in EXCEL.
4. Translate all the types of the attribute values into numbers. For example, if the value of attribute CONTRIBUTING FACTOR VEHICLE 1 is Following Too Closely, we can translate it as 2. We can use replacement in EXCEL to do this step.
5. Implement random sampling. Then, reduce the data to around 10000 rows.

**Layout**

Based on the NYC map, we can locate the collisions by latitude and longitude. Then, use d3 to put all the collisions as points on the map.

1. At the top left of the page we will set 4 top-down menus. Two of them are used to select date range, the other two are used to select time range. The date menus contain only January to December, and the time menus contain only 00 to 23.
2. At the top right of the page, we will set a gathering button.
3. After we click one gathering point, the page will pop up a interface that contains a line chart with date as x axis and number of collisions as y axis. There will be a swap button that swap the attribute of x axis to time.



**User interaction**

1. User can click the gathering button, then the point can be clustered to a gathering point. User can click the gathering button again to cancel the clustering. The color of the gathering point will states the number of collisions. For example, black point means there are more than 100 collisions in this area. It has mouse over action to show it clustered data such as collision number, number of people injured.
2. User can click the gathering point to show more detail in a line chart with date or time as the x axis.
3. User can click the line chart to display detail information in that month like using the days as axis to show the detail of the month.
4. User can manually select the points by dragging a rectangle. It will provide exactly the same function as the clustered point.

**Conclusion**

After you visualize the data, you can know which area of New York City has many collisions so that you can pay more attention when you are there.