Huma Tejaswi Akkem

Dinesh Dammalapati

Project 2 CS 424

Analysis of traffic crashes in City of Chicago

Background Information

Open chicago data information is taken from chicago data portal service which consists of more than 500 information of accidents occured in chicago that includes minor and major crashes. Records include numerous attributes such as road condition, Weather condition, Crashed Date, Posted speed limit, device condition, road number, damage information and more excluding the personal information.

In this project, we study the dataset that contains the accidents occured in the time period of January 1st of 2019 to January 1st of 2020.

Goal: To Study the vehicle collision of two different complete situations,one being as a normal year and other being as pandemic year

Cleaning the dataset and exploratory data analysis

According to Data Portal, the dataset has 42 attributes and we wanted to use only a few attributes in the dataset. We dropped the half of the attributes that contain string unknown, null values and empty values using df.drop. To simplify the dataframe we dropped the column and rows that we are not using. Exploratory data analysis consists of studying the main roles of a data set usually by means of visualization that contains specific interests such as time, space, space + time, distributions, and comparisons.

Domain-Question:

How did COVID-19 pandemic affect the road crashes and crash outcomes in Chicago?

Data-Questions

- Based on the dataset in which year has a higher amount of crashes occurred?

- How did the pandemic change the frequency of crashes based on the area.

- How did the number of crashes change over the time period by month?

- How did the number of monthly injuries change from the year 2019 to the year 2020?

Answering the questions

1. Based on the dataset in which year has a higher amount of crashes occurred?

    - To study the distribution of crashes per year, we used the d3.arc to create pie charts and d3.pie to send the data and we previously created an array which has number crashes for year one and year two which made it easier for us to create the graph.

    - First we grouped by year and then calculated the highest occurrence in each month.

    - **Conclusion** : As expected we have a higher number of crashes in 2019 then 2020. But we do not see a major difference and the reason can be, reduced crowd can lead to increase in speed.

    - Assumed reasons are, 2019 was the start of pandemic and had strict rules implemented and people had a fear of covid-19 death rate.
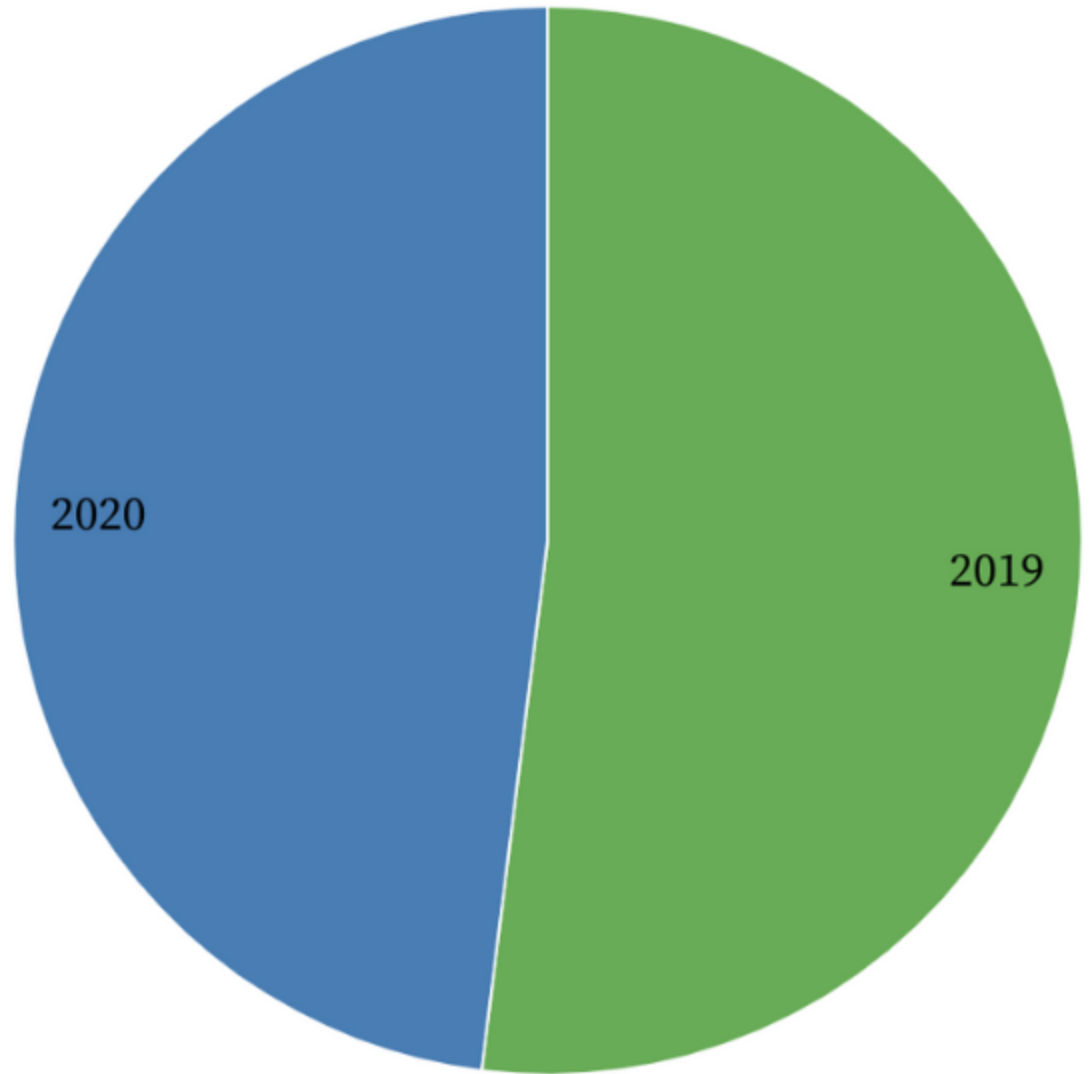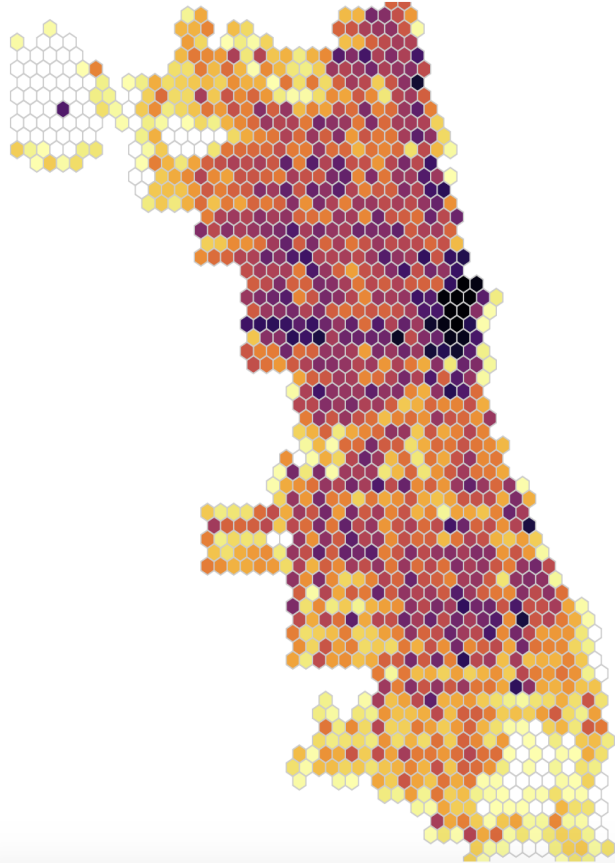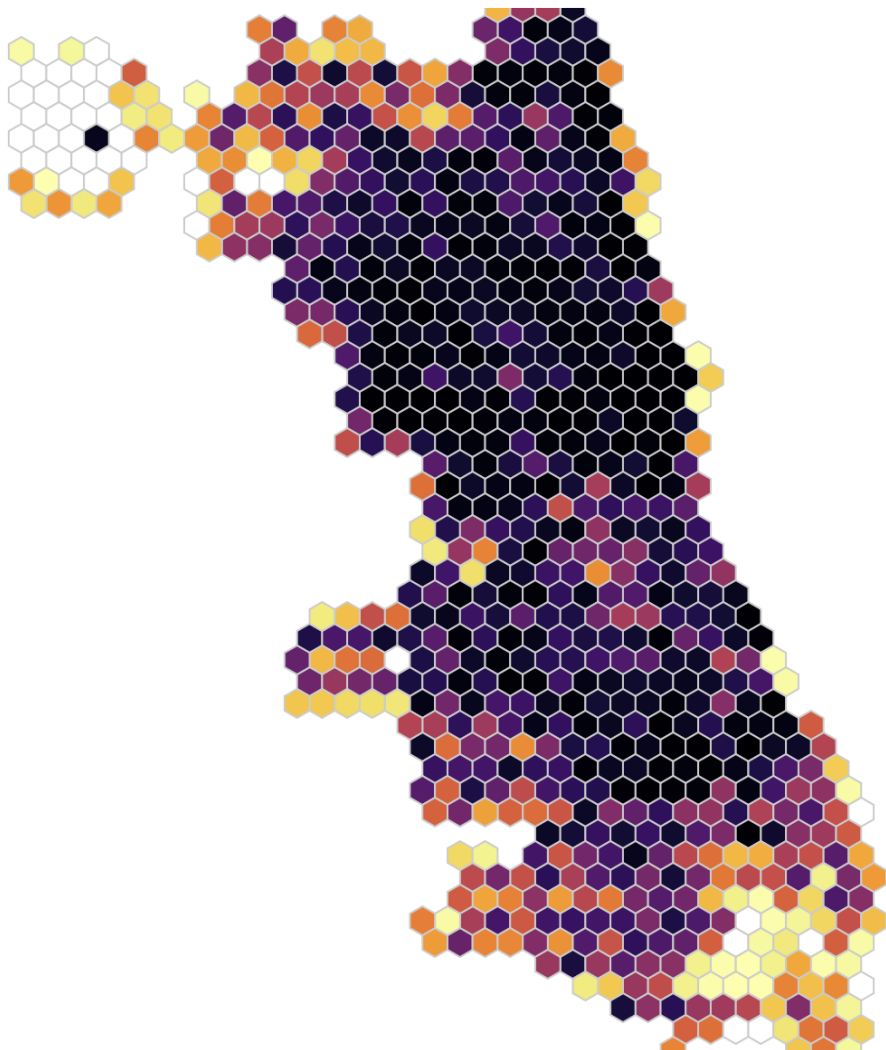
Interactive Graph 1:

2 . How did the pandemic change the frequency of crashes based on the area.

- We are presenting on project 3, We imported the first two graphs from project 2 as
  required. First map shows the number of crashes that occurred in 2019 and 2021.
- D3.js handles the geographical information more powerful/ efficiently than other
  visualization tools, manipulating and cleaning the geographical data was a hard
  part in our project.

- We stored the data in JSON in a compatible data format, then defined the size of the map, projection, SVG element, and appended it to DOM.

- We made three Illinois maps: one as an entire dataset, second as data from 2019 then finally data from 2020.

- We made it interactive because graphical representation is always better to be in control of user to have a clear data-understanding

- We made it interactive using the scale which can change the map's hex grid and colourScale according to the user's preference. In the first picture, the hex grid is 6 and the color scale is 5. Likely, in the second picture the hex grid scale is at maximum level and colourScale is at 16.

colourScale = ƒ(t)

6

5
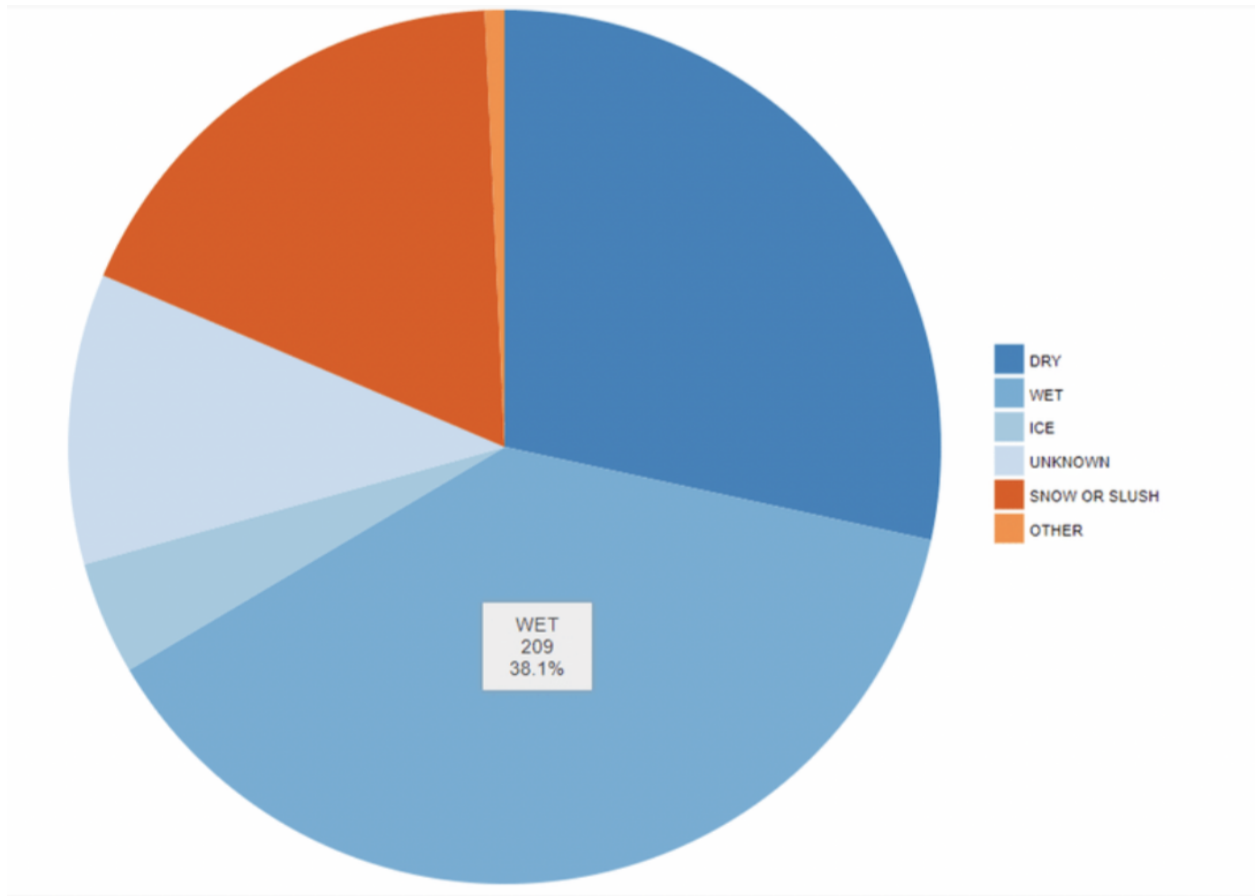
-
- In addition, we also did surface condition based on the geographical map.
- Majority of the crashes that occured during 2019 and 2020 have occured when the surface_condition was wet

Interactive Graph 2:

3. How many crashes took place each month in the year 2019.

- Interactive Zoomable graph that shows the number of crashes that took place in each month of the year 2019
- First we imported the data to JSON file and we used SVG to plot the graphs then we defined the dimensions for the axis

- We read the data and blended the data into the assigned axis. After that, we added the interactions and zoomable effects to the graph using zoom = f(svg)

- View DOM used to control the translation and zooming the graph.

- **Conclusion** : As expected we have a higher number of crashes in 2019 then 2020. Also, during the summer we had more crashes than other months of 2019

Additional graph : In addition, we added a dumbbell graph that shows the comparison of crashes by each month from 2019 to 2020

Comparison of crashes by each month from 2019 to 2020
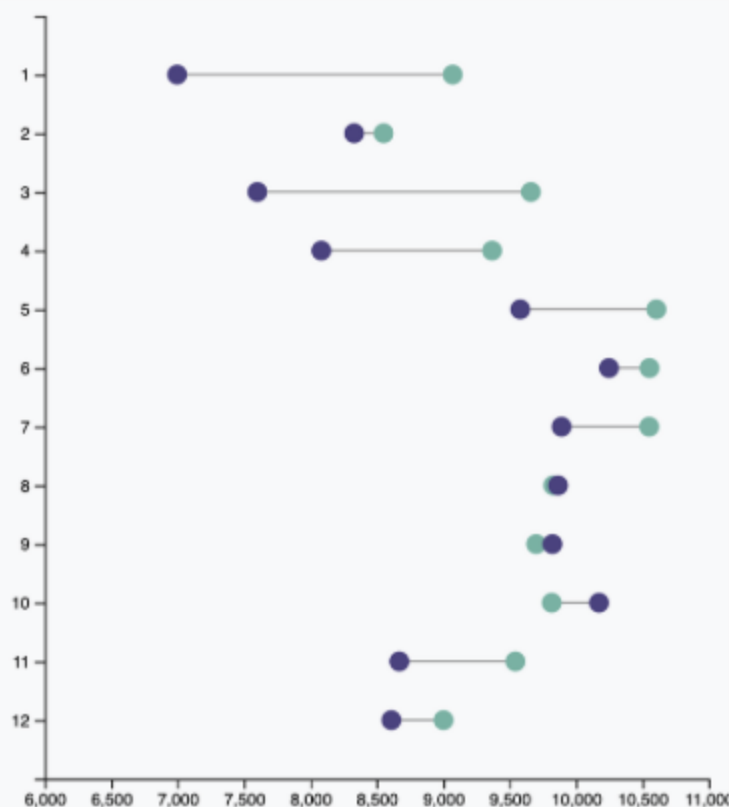
Dumbell Graph

4.How did the number of monthly injuries change from the year 2019 to the year 2020?

- We used the bar graph to represent the comparison between two years of injuries happened.

- Like before, we did the selected the graph and we cleaned the data before we used.

- We set the margins and set the scale, for our axis we used Linear Scale then we set the domains.

- We set the domains for both x-axis and y-axis then we used a range according to the dataset.

- After that we were able to display the graphs and then we formatted the axis such as calling the axisBotton as our x-axis and we read the particular attribute. Finally we created the dynamic bars.

- **Conclusion :** As expected, 2019 has more injuries than 2020 and the reasons we assumed because empty road speeds the vehicle chances.