Hackathon-IE6600-Sec05-Group1

Darshan Patel, Naveen Nanda Mohan, Navneet Kaur Gujjar

10/12/2021

Introduction and Problem statement

Traffic is a living, breathing organism which is constantly in flux. To be able to understand and interpret traffic patterns is to control a very important aspect of day-to-day life. One important aspect of traffic, which is especially relevant to crime statistics is road accidents along with cause and effect. From a data analytics point of view, traffic analysis has a lot of potential due to

- The size of the data
- Periodicity of the data
- Constant flux

For the purpose of simplicity and understanding, we have taken the city of Austin as a case study and using APD's fatality data for the years 2017,2018 and 2019, we try to understand the different cause and effect of road accidents over the given time span.

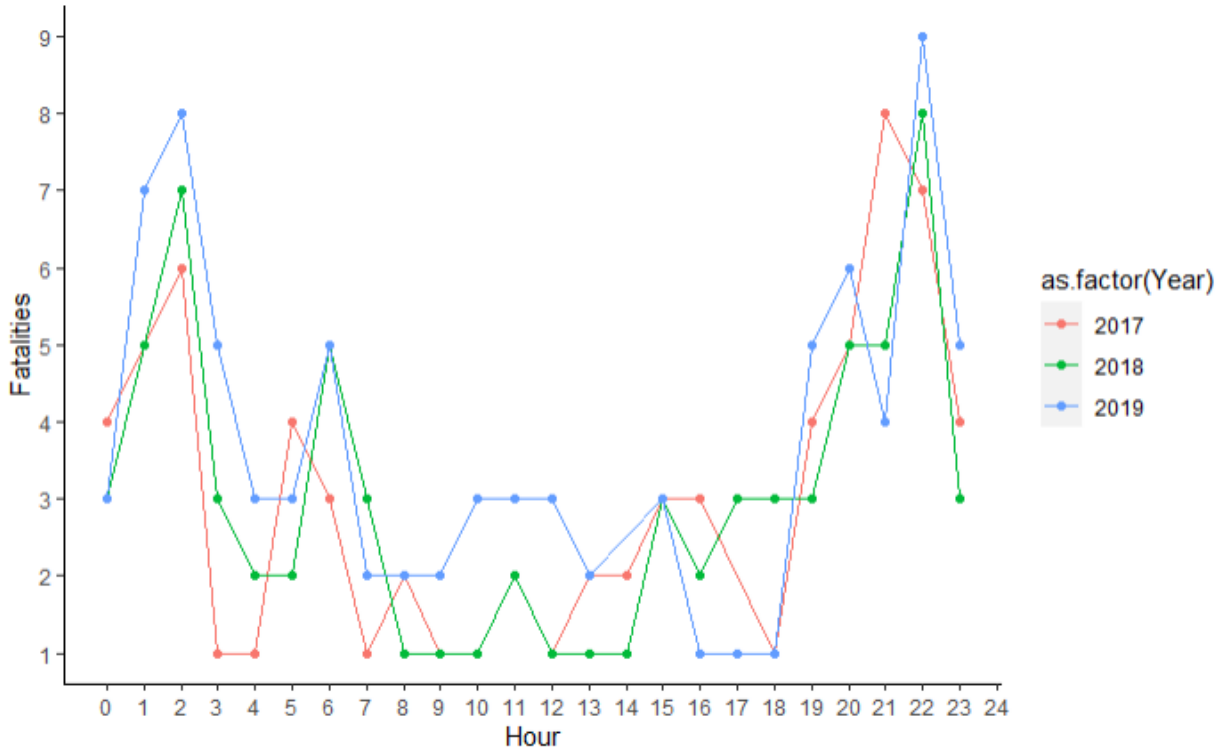
Using the given dataset, we were able to understand the correlation between different hours and months versus the number of fatalities, spread across different years. We plan to compare the fatality rate at multiple time levels to try and understand the overall trend. The objective was to figure out if timeline played an important role in fatalities, such as accidents during the day vs night, over the weekend versus weekday and during the earlier/later periods of the year. Although the data is limited to a specific city, there are a lot of possibilities for exploratory analysis.

We also plan to use the full spectrum of visualizations offered by R. One such example is a Google map plot which can help isolate the cluster of fatalities based on latitude and longitude coordinates. Plotting this graph requires a Google maps API key.

The final section deals with using an alluvial chart to map out an important parameter called as FSRA – Failure to Stop and Render Aid. This is a report filed by the APD to highlight the importance of victims receiving immediate help, including help from passer-by in the form of a 911 call or first aid.

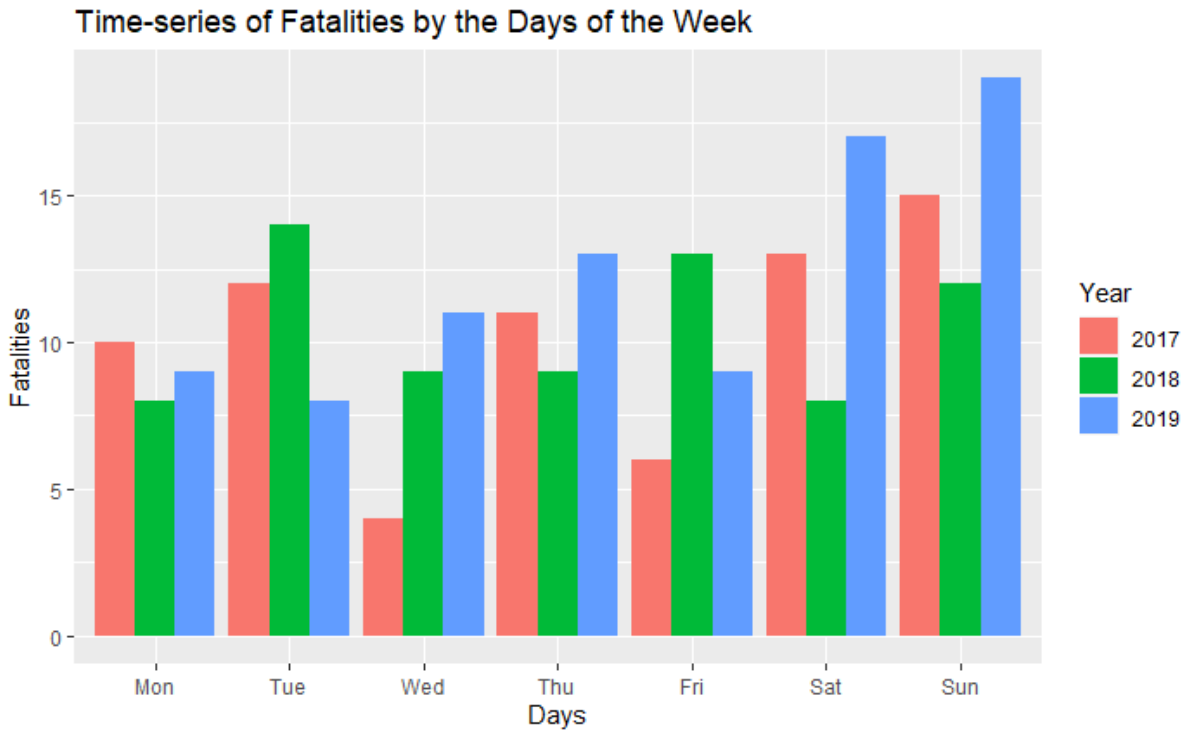
Time Series Analysis of Fatalities

Q1. What is the Peak hour of fatalities across all months each year?



Conclusion: Based on the line graph, we understand that peak fatality hours occur during the night-time, probably due to low visibility. One possible next step is to take floodlight and street light data to see if the roads and areas in this dataset coincide with any places with low visibility due to insufficient lighting or other hazards that require proper vision to avoid.

Q2. What is the split up of fatalities for weekend vs weekday?



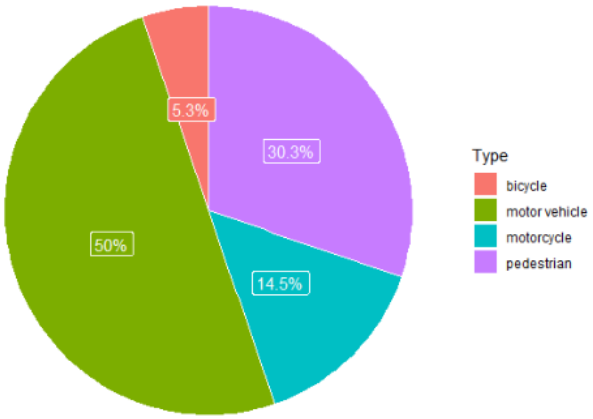
Conclusion:

In 2017 and 2019, fatalities occur more frequently during the weekends rather than the weekdays, whereas in 2018, fatalities occur more during the weekdays. This is interesting as there is a jump in time line, hence we require data related to police weekend patrols to understand if certain actions where taken during the weekdays which resulted in the above result.

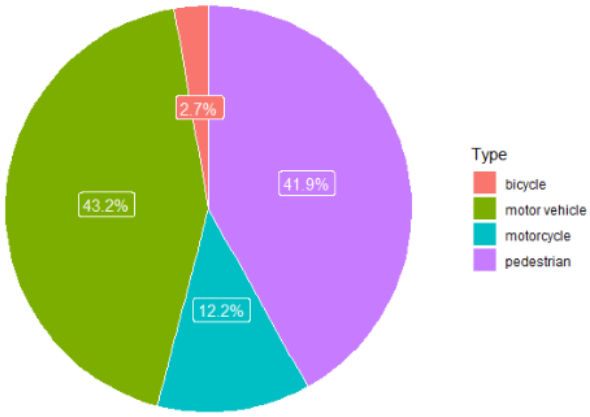
<u>Vehicles involved in Fatalities</u>

Q1. What are the Vehicle types involved for each year?

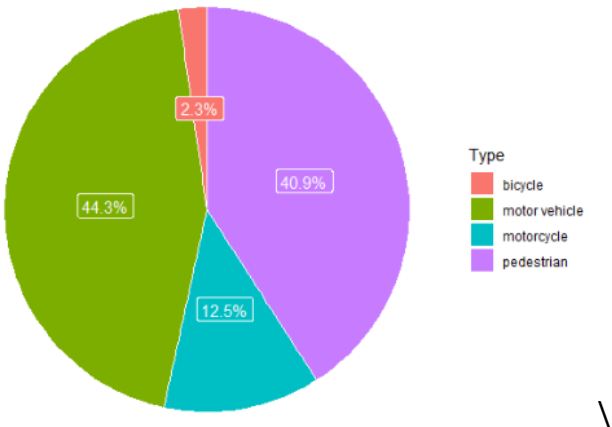
Total Number of Fatalities by the Types of Vehicle (2017)



Total Number of Fatalities by the Types of Vehicle (2018)



Total Number of Fatalities by the Types of Vehicle (2019)



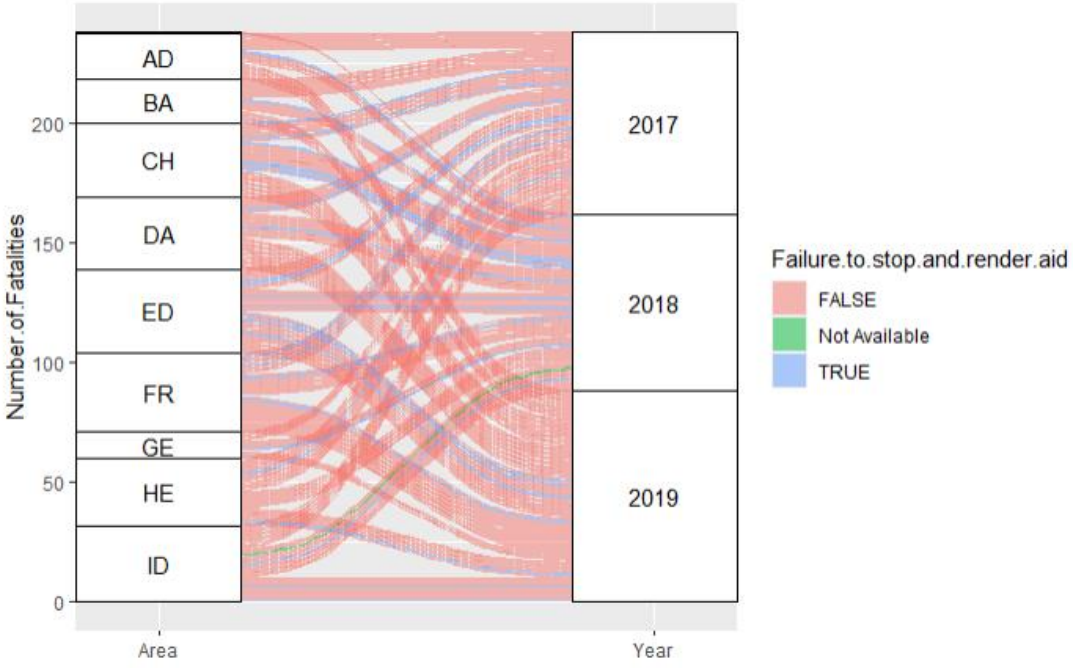
\

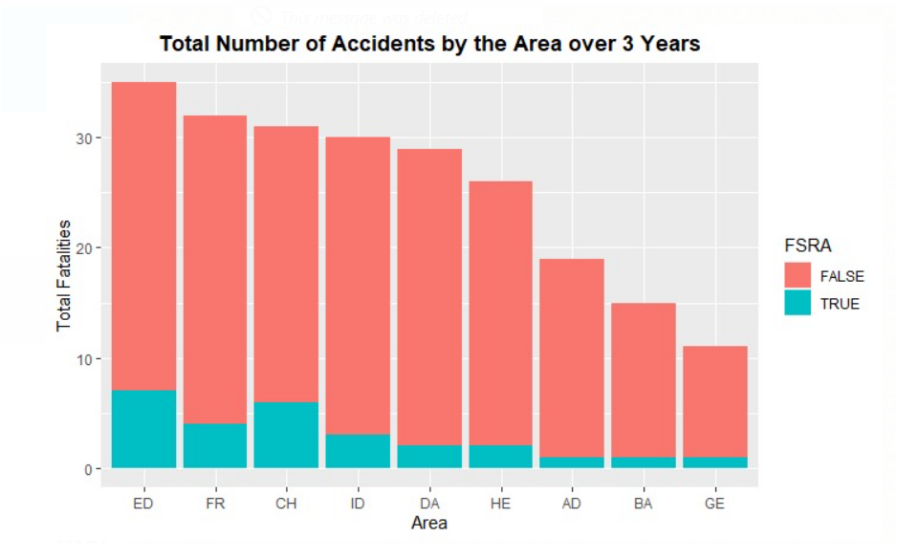
Conclusion:

Based on the analysis performed, we can see that motor vehicles irrespective of year, is the highest cause of accidents, covering nearly half of the total accidents caused in each year. Pedestrians ar affected next as a result of accidents and vewhicular mishaps affecting innocent bystanders.

Failure to Stop and Render Aid

Q1. What is the FSRA report over time considering each area under Austin?



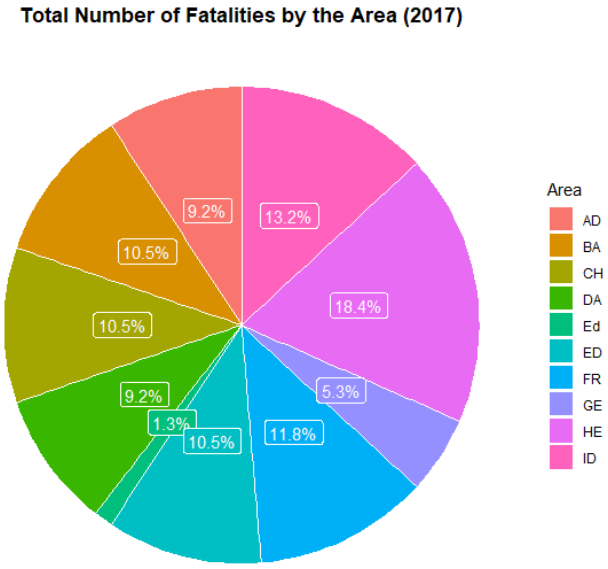


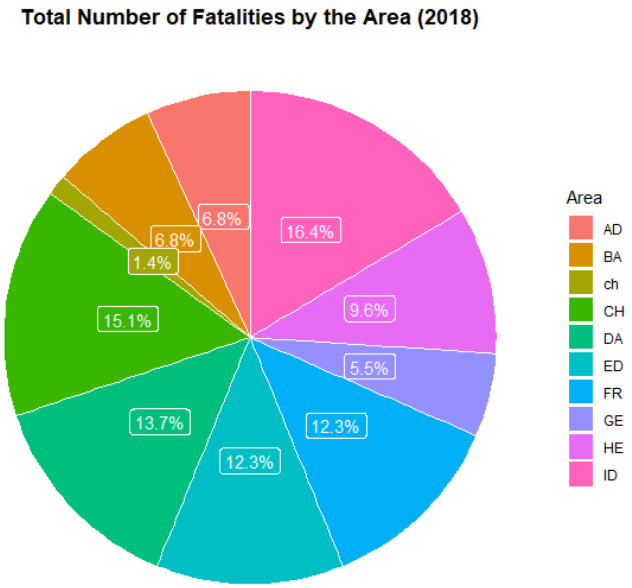
Conclusion:

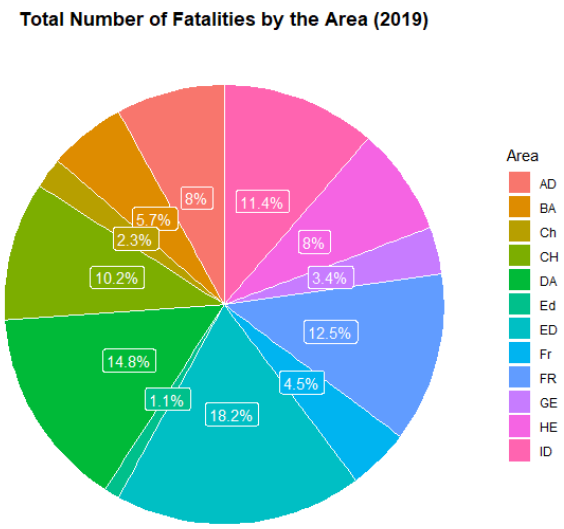
Based on the given graph, we understand that there is a significant amount of FSRA report filed positively, due to which the number of fatalities is impacted. Overall attitude towards immediate help is considered as intrusion or immediate help is not timely, resulting in potential needless loss of lives. The reverse can also equally hold true, according to numerous surveys.

Area Wide Distribution of Fatalities:

Q1. What is the breakage of fatalities across given areas for each year in the dataset?



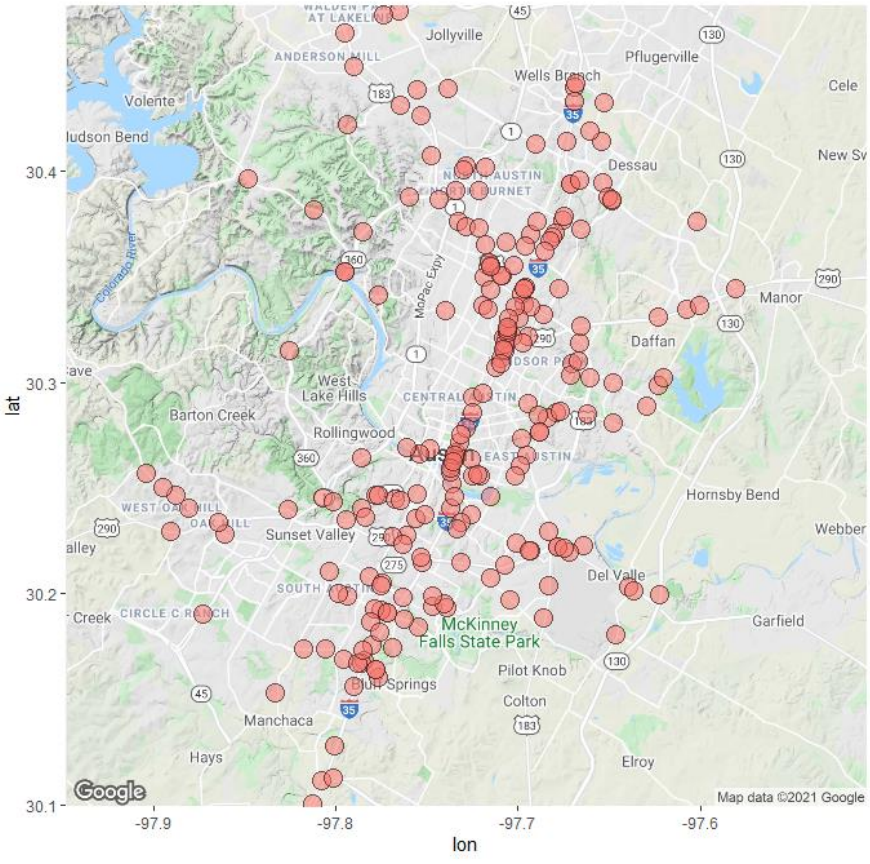




Conclusion:

Area ID is the most unsafe area in Austin during 2018 and 2019, whereas HE was the most unsafe back in 2017. But, HE has slowly started declining in crime rate at a steady pace over time, with ED catching up in crime rate pretty quickly.

Q2. What is the overall clustering of fatalities across Austin for all years combined?



Conclusion:

Using Google maps, we can visualize a clustering of data points across the different areas of Austin Texas, which also gives us a glimpse of the possible visualization mechanisms we can use to plot data and understand the total spread.

Possible Future Work:

- Performing the same kind of analysis with similar parameters across all major cities in US
- Including datasets such as road conditions, existence of floodlights and such (PWD data)
- Post-mortem reports indicating presence of alcohol or other substances will help understand the data better
- Maintaining a singular repository of historical traffic fatality data to understand patterns
- Expanding collection to include different dimensions of traffic data

References and Dataset Link:

- https://catalog.data.gov/dataset/2017-apd-traffic-fatalities
- https://data.austintexas.gov/Public-Safety/2018-APD-Traffic-Fatality-Data-021219/9jd4-zjmx
- https://catalog.data.gov/sv/dataset?tags=fatality&publisher=data.austintexas.gov