

# US ACCIDENTS (2016 - 2021) EDA

Havan Patel  
hawangpatel@lewisu.edu  
DATA-53000, Summer 2022  
Data Visualization  
Lewis University

## I. INTRODUCTION

For this project we will be analyzing US accidents from (2016 – 2021) data. Road accidents have become more common and dangerous now a days. Approximately 1.35 million people die in road crashes each year; on average 3,700 people lose their lives every day on the roads. [1] Not only that but 20-50 million people suffer non-fatal injuries which resulting in long-term disabilities.[1] This report utilizes multiple data visualizations for exploratory analysis; the overall goal of the report is to visualize the data attributes and seek potential relationships of the data attributes with prediction on what causes these accidents and the severity of them. Are we doing anything to stop them and save lives? Informative visualizations will be displayed below to educate the readers on the data and simply display the different data attributes.

## II. DATA DESCRIPTION

As described in Section I above, the dataset we are using is an Exploratory Data Analysis (EDA) to get insights from US accidents which consists of 1,5 million accident records with 47 columns, which covers 49 states of the USA [2]. Because the data has a high number of instances/records, data visualization will assist in the ability to comprehend the large amount of data. The dataset has been used for a variety of different studies for US accidents and prediction for where it is likely for an accident to happen. This will only help the US and state department of transportation to implement better solution to reduce the number of accidents. From this dataset we will get insight as to which feature have contributed to more accidents by performing data visualization in Tableau . It is always the best to practice and understand the data first and try to collect as many insights as possible from it. In the table 1 below I have listed all the important attributes that can help us in analyzing and extracting interesting facts from the dataset.

TABLE I. DATA ATTRIBUTES

Attribute	Type	Example Value	Description
Start Time	Date	2/8/2016 12:37:08 AM	Start time of accident
End Time	Date	2/8/2016 12:37:08 AM	End time of accident
Severity	Numeric	2	Severity of accident
State	String	NJ	Name of state in USA
Country	String	USA	country
City	String	Los Angeles	Name of city in USA
US Accidents	Numeric	12399	Total number of accidents
Weather Condition	String	Cloudy	Condition of weather when accident occurred
Year	Date	2020	Year accident happened
Temprature(F)	Numeric (Decimal)	42.1000	Temperature during accident
Wind Speed	Numeric (Decimal)	10.4000	Speed of wind in mph
Bump	Boolean	True	POI if bumps was present nearby location
Crossing	Boolean	False	POI if crossing was nearby location
Railway	Boolean	True	Presence of railway nearby
Precipitation (in)	Number (Decimal)	0.0200000	Precipitation amount in inches

## III. METHODOLOGY AND RESULTS

The data was input into Tableau, and multiple visualizations were created to conduct exploratory visualizations on the data. In Fig. 1 let's look at the most accident-prone cities. In the x-axis we have cities listed and on y-axis we total accidents. Each bar is

colored differently for each of the top accident prone cites. We can see that Los Angeles has the most accidents followed by Orlando and Dallas.

Top 10 cities in US with the most accidents.

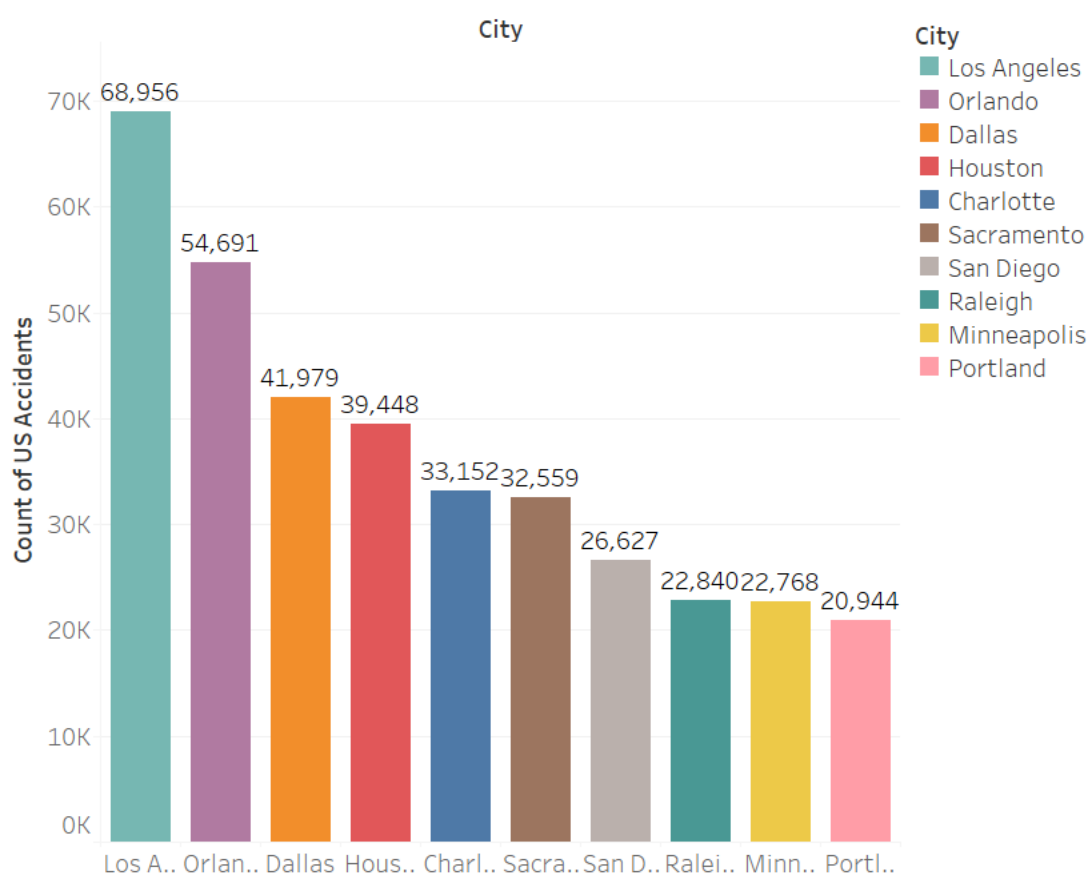


Fig. 1 Top 10 cities in US with the most accidents.

In Fig. 2, visualization we have a choropleth map which shows the most accident-prone states in USA. The darker the color the more accident occurred in that state and the lighter the color the less accidents. We can clearly see the correlation between Fig. 1 and Fig 2, where we know from Fig. 1 that Los Angeles is the city where most accidents happens and that is in California State. Therefore, California state is highlighted more darker followed by Florida and Texans. This also represents that a lot of accidents happen in these 3 states (California, Florida, and Texans).

## State with the most number of accident in US

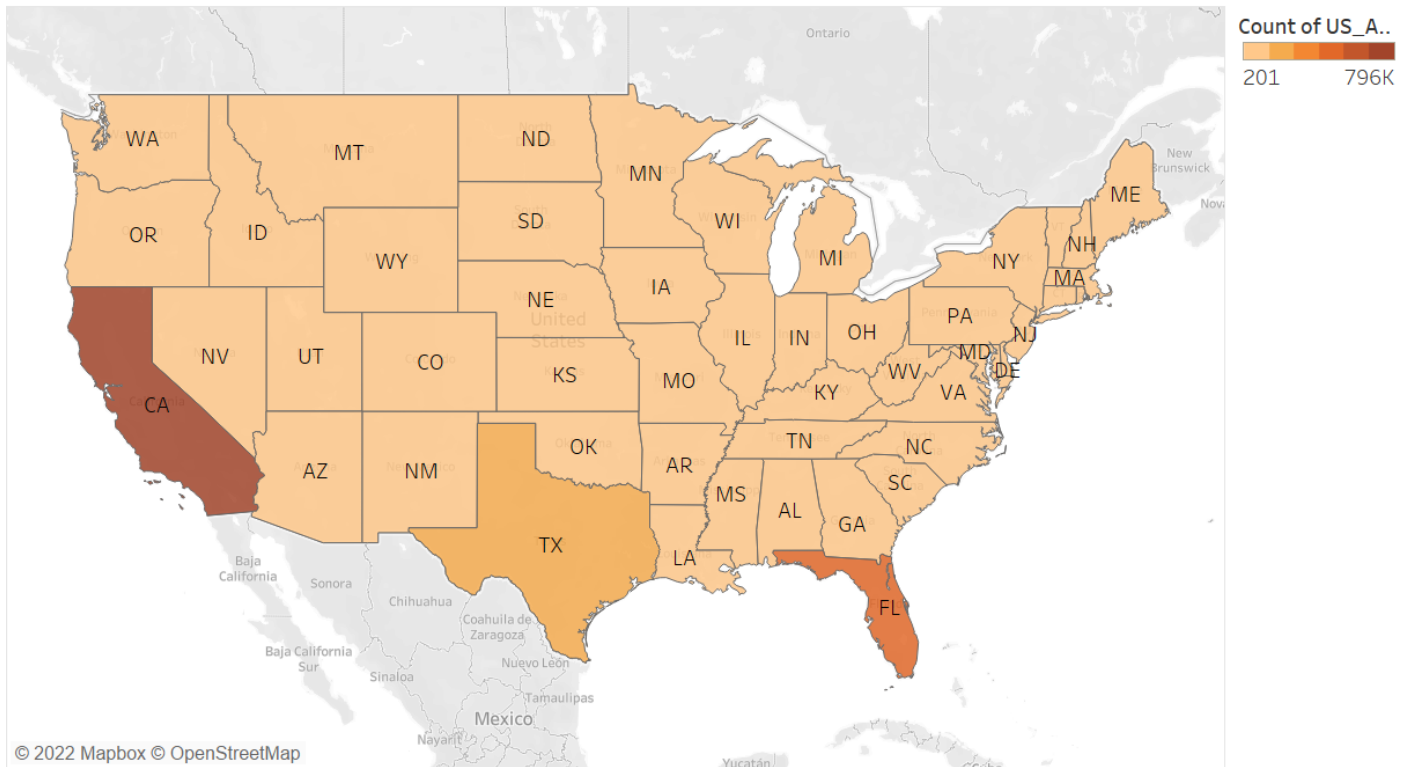


Fig. 2 State with the greatest number of accident in US

In Fig. 3, we take a different approach and look at the severity of accidents in the cities for each state. With this density map in Fig. 3 We see something interesting where we would expect more sever accidents in California, Florida, and Texans but the average severity of these 3 states is from 1to 2. The severity scale is from 1 to 4 where 1 represent not too sever and 4 being sever accident. Once again, we see the color scheme of dark to light color. Dark color is severity level 4 and light color is severity level 1. It is interesting that a lot of sever accident happen in northeast part of US.

## Accident severity of cities in US

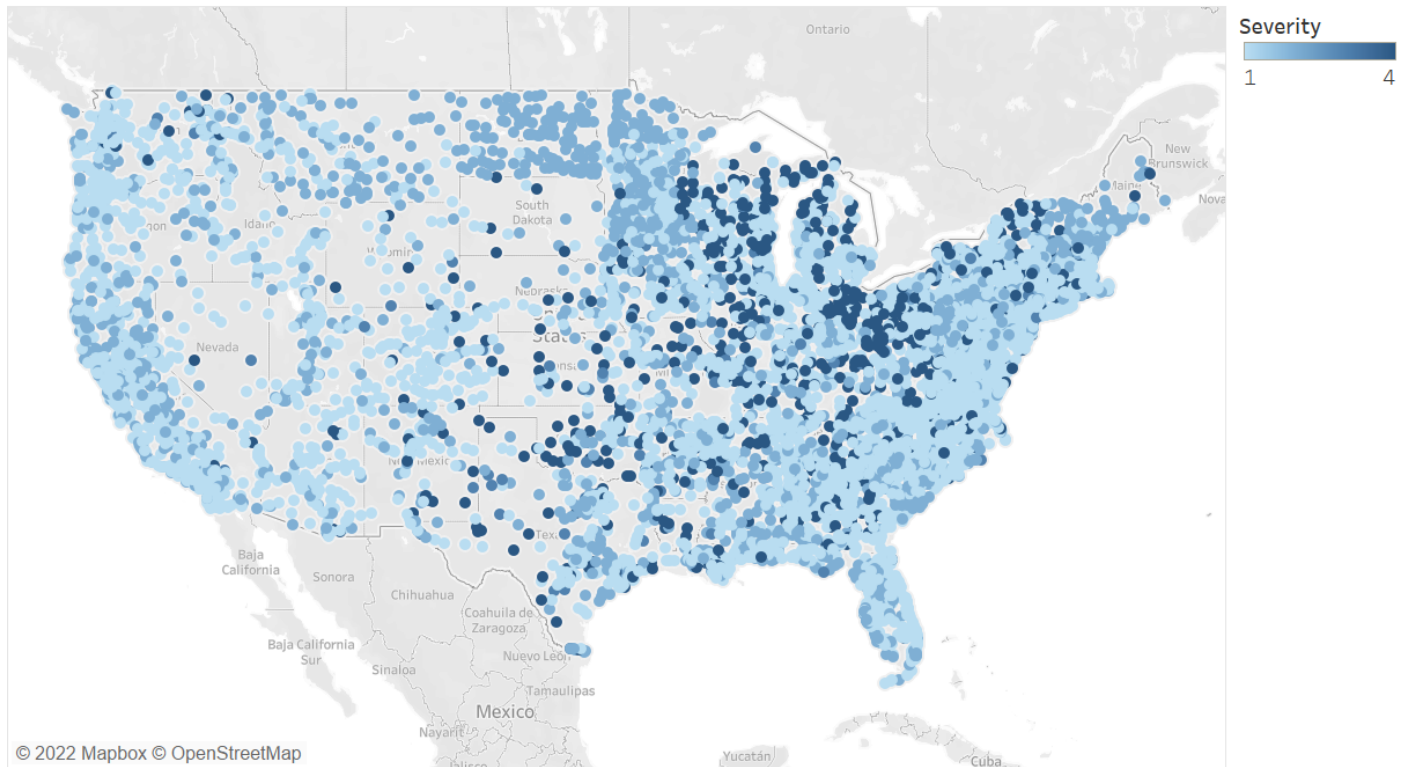


Fig. 3 Accident severity of cities in US.

In Fig. 4, we are again continuing with the theme of accidents but what are the cause for it. It looks like that weather does play in a little role for the happening of the accidents but some of the situation doesn't make sense where there is more accident on clear weather then let's say a light rain. Therefore, it does not seem like weather plays an important role based on the visualization here.

## Accident by weather

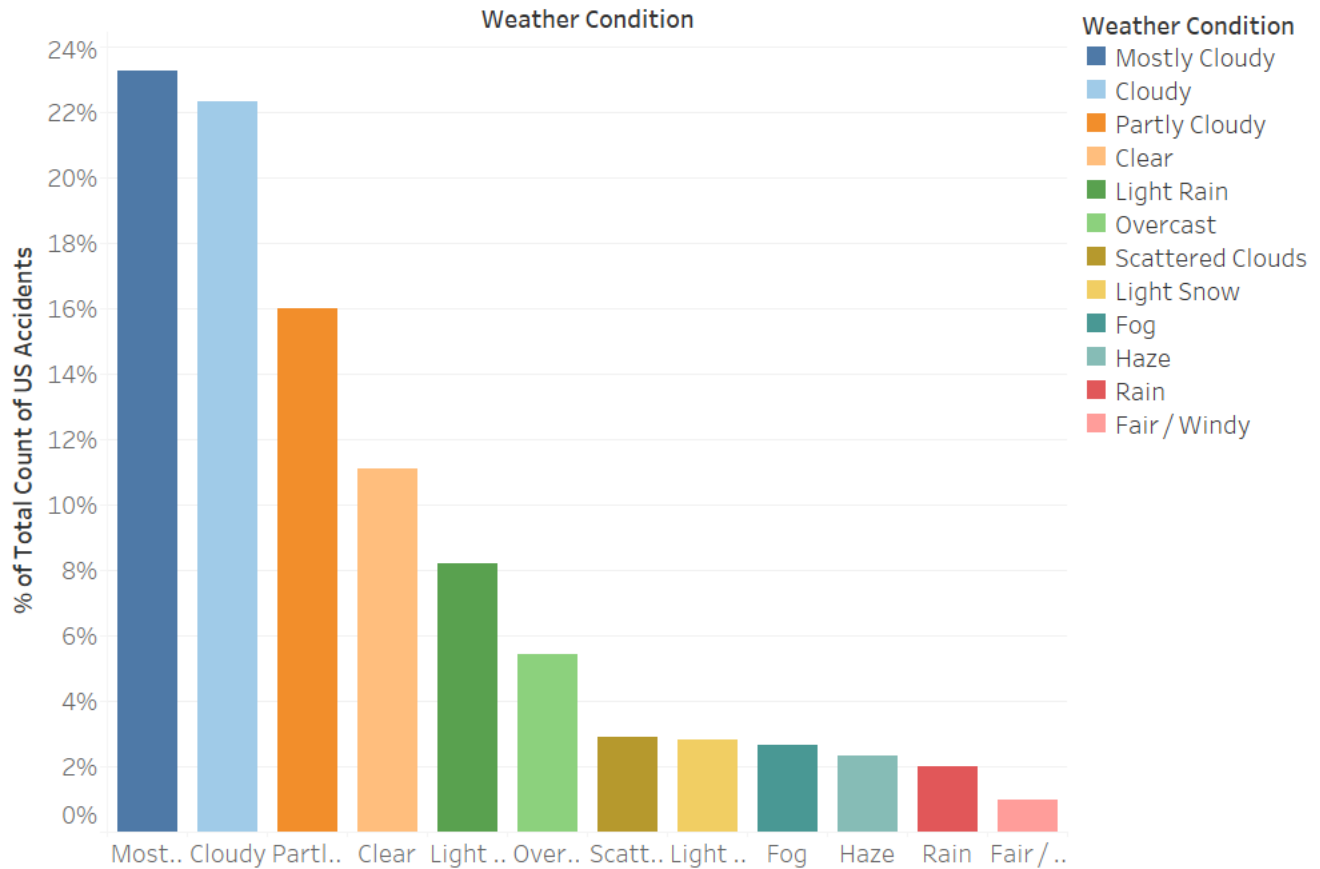


Fig. 4 Accident by weather.

In Fig. 5, we are looking at how over the time the accidents are increasing, and we see a massive leap of increase from 2019 to 2020. Due to covid-19 we see less increase from 2018 to 2019 because of the lockdown but I would assume that it would have the same impact and increase in accident rate as it had in 2020. In this visualization we see that the bigger the bubble the more percentage of accidents for given year.

## Road Accident Percentage over past 5 years

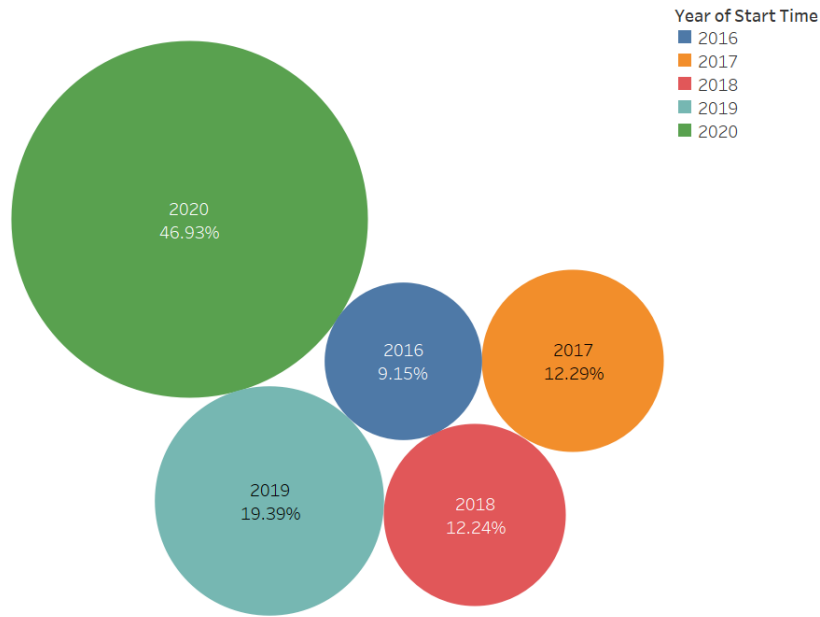


Fig. 5 Road Accident Percentage over past 5 years

In Fig. 6, let's look at the severity and corresponding accident percent for past 5 years. We see some consistency in the severity levels over the years. On the x-axis we have the percent of total count of US accident, and on y-axis represent years.

## Severity and corresponding accident percent for past 5 years

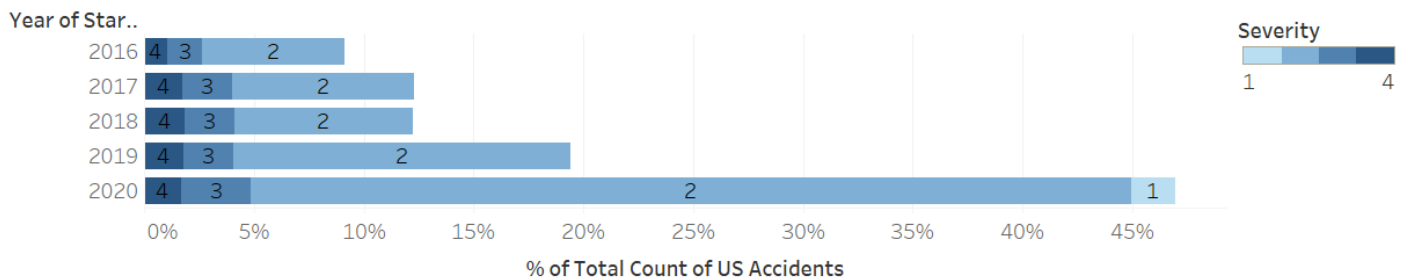


Fig. 6 Severity and corresponding accident percent for past 5 years.

## IV. DISCUSSION

In week 4 lecture we learned about different visualization maps and how each different map is used for different purpose when trying to convey information to public. In this specific report I have used 2 different type of visualization maps. One of them is called choropleth map and another one is density map. The difference is that a choropleth map shows a change across a geographic landscape within enumeration units such as countries, states, or watersheds. A heat map shows a change across a geographic landscape as a rasterized dataset—conforming to an arbitrary, but usually small, grid size. In Fig. 2 I utilized choropleth depicting states with most accidents. The reason I was able to utilize choropleth map was because I used the data that was spatial enumeration units.

## V. CONCLUSIONS

In this report, we used the dataset that contains data about car accident in US and created visualizations to conduct exploratory analysis to better understand the dataset. With our data visualizations, we were able to visualize how certain attributes can affect the

accidents in US. We came across some exciting information for example, we came to know which city or state saw the greatest number of accidents in the USA, we even plotted the results on a map and considered the severity of an accident.

#### REFERENCES

- [1] "Associate for safe international road travel" [Online]. Available: <https://www.asirt.org/safe-travel/road-safety-facts/>
- [2]"SOBHAN MOOSAVI" "US Accidents (2016 - 2021)" [Online]. Available: <https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents>