

US Accidents:

Names of team members:

- Ratan Deep Singh, Alan Zhang, Haoyu Zhang

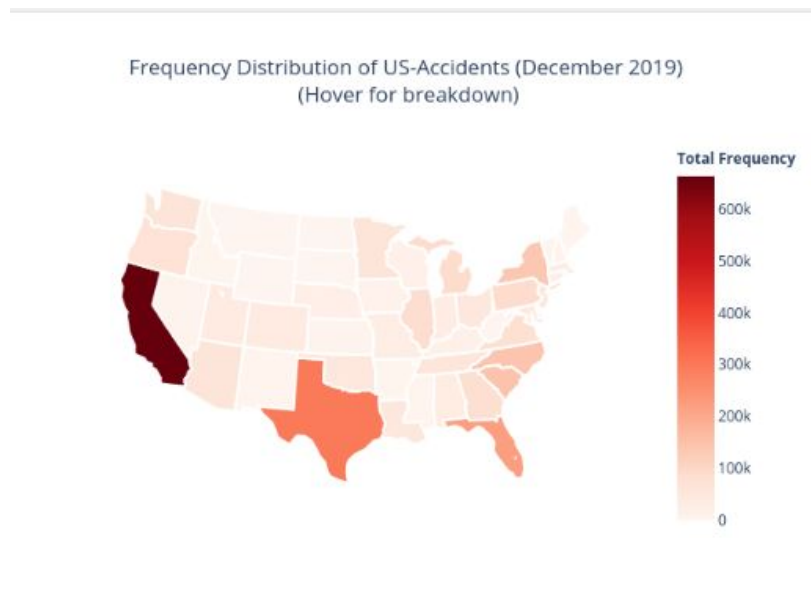
Name of your team's GitHub repository:

- Project2_Singh_Zhang_Zhang

A primary dataset you intend to analyze:

- US Accidents (2016-2020)
 - <https://www.kaggle.com/sobhanmoosavi/us-accidents>

Initial plots, figures, or tables: Work in Progress



Some of the variables (column names) you intend to explore and what kind of insights you expect to glean:

- For the US Accidents dataset, we have a detailed table for the column names. See the link below.
 - https://smoosavi.org/datasets/us_accidents
 - Example below

#	Attribute	Description	Nullable
1	ID	This is a unique identifier of the accident record.	No
2	Source	Indicates source of the accident report (i.e. the API which reported the accident.).	No
3	TMC	A traffic accident may have a Traffic Message Channel (TMC) code which provides more detailed description of the event.	Yes
4	Severity	Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).	No
5	Start_Time	Shows start time of the accident in local time zone.	No
6	End_Time	Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed.	No
7	Start_Lat	Shows latitude in GPS coordinate of the start point.	No

Supplemental datasets, if any, to complement your primary dataset - this means links, columns that you'll join on, etc.,

What you plan to cover in the final report and how you plan to organize it:

- For US Accidents, we plan to analyze:
 - The distribution of the reported accidents to understand the bias
 - Average response time by looking at the start/end time
 - Location of the accidents
 - The relationship between temperature and # of accidents
 - Types of road where accidents happen more often
 - When are accidents mostly likely to happen; time of day