VISUALISE THE AVIATION ACCIDENTS

SPRING 2017 IN CSCI 55200
DATA VISUALIZATION

Submitted By,
Mani Manjusha Kottala
Kartik Sooji
INDIANA UNIVERSITY PURDUE UNIVERSITY INDIANA | IUPUI,USA

TABLE OF CONTENTS

ABSTRACT

Data is Market, Big Data – Now is the era of big data and data mining. As the technology develops, the data behind the earlier versions of the solution to a problem needs to be stored. However, for any general user the same will not be able to understand the details rather such users wouldn't have relevant software experience to store and view the data locally. Hence, visualization of such data in a simplest way that can make any user to understand the minor details without having the technical insights. We initiated a visualization project to make tourists plan their travel based on the aviation accidents that occur due to various conditions. In this project we considered few of the visualization techniques to represent the data

INTRODUCTION

Motivation behind this project idea was to make tourists aware of the accidents prone areas across the globe. This application provides various visualization of the accidents prone zones considering many factors from the data set.

There are many airplanes accidents that happen at a given point of time at some part of the world for which the respective governments are taking precautions for the same by analyzing the previous data records. Few of them are adverse weather, pilots and other technical issues. Hence the severity of the accident matters. This application shows us the various kinds of accidents based on the type of the vehicle, and the time stamp details as to where and when the accident occurred. Hence this application helps in better understanding of the causes and types of accidents.

DATA SET[1]

The dataset source is an online open database named **National Transportation Safety Board(NTSB)**, **USA**. A USA based firm that has the accident history since 1962. The data set has most data with USA civil accidents with respect to states, data was not only about the accidents but also the detailed investigation reports of the accidents. The dataset is almost up to date as the reports will be generated in a span of a week which helps for the further improvements in the methodology of the bird's design.

The dataset consists of nearly **79K records of 32 columns** out of which here is the sample of the header columns. There is not much data collected during initial phases of the firm as, the aviation technology was not that prominent than how it is been engineered now. Few of the fields that are considered for the project are highlighted below.

> *Event Id;Investigation Type;Accident Number;* ***EventDate****;* ***Location****;* ***Country****;* ***Latitude****;* ***Longitude****;Airport Code;Airport Name;****Injury Severity****;Aircraft Damage; Aircraft Category; Registration Number;Make;Model;Amateur Built;Number of Engines;Engine Type;FAR Description; Schedule;Purpose of Flight;Air Carrier;****Total Fatal Injuries****;****Total Serious Injuries****;****Total Minor Injuries****;****Total Uninjured****;****Weather Condition****;****Broad Phase of Flight****;Report Status;Publication Date;geo_point*

The data set is part of the csv. NTSB has the records yearly, monthly and is always under continuous update.

DEVELOPMENT TOOLS

There are few visualization tools and techniques that can be used to visualize that we have learn during the course namely VTK and D3.JS.

We chose to work with d3 as we gained basic knowledge on the same.

- *Front End Development* – D3.js (Version 3 and Version 4) and HTML5 [2]
- *Page Styling* – CSS
- *Data Set formats* – CSV, JSON
- *Data segregation* – MS Excel, SQL, MS Access

On a general note, D3.js has few components that are utilized while generating any visualization layout are:

- Selection
- Dynamic Properties
- Enter and Exit
- Transformation
- Transitions

VISUALIZATION TECHNIQUES

Figure 1: Index Page

Index page has visualization techniques implemented. Clicking on each of them would redirect to the respective graphs.

Geo-spatial Visualization:

We chose this type of visualization as it is one of the best way to represent the huge data that involves a global data. Geo visualization provides lot of information just on one screen that makes the observer understand much more with a lesser knowledge.

World map[3] was generated using **d3.geo.mercator** projection. Later, the geo-location of the dataset was mapped with the geo-location in the TopoJSON[5] file using the projection of longitude and latitude. USA map is created using the d3.geo.albersUsa projection.

Data Input: JSON data file**,** aggregated whole data set of **75K records**
Data Attributes Described:

Event Year, Geo-location of accident, Damage level of Aircraft and Aircraft Category

Components Implemented:

Point – Geo-location plot: All data points are mapped onto the SVG area based on the longitude and latitude information from the dataset

Drop Down Menu Selector: To visualize data according to the year, select any year in the dropdown to have the data points updated on the map

Legend: Shows the categories of the type of aircraft accidents

Zoom: click on the SVG would lead to a better visibility of the plots by zooming in and zooming out on countries and states.

Hover: hover over any data point on the screen to have more insights of the accident occurred which are Year, Damage and Airplane Category

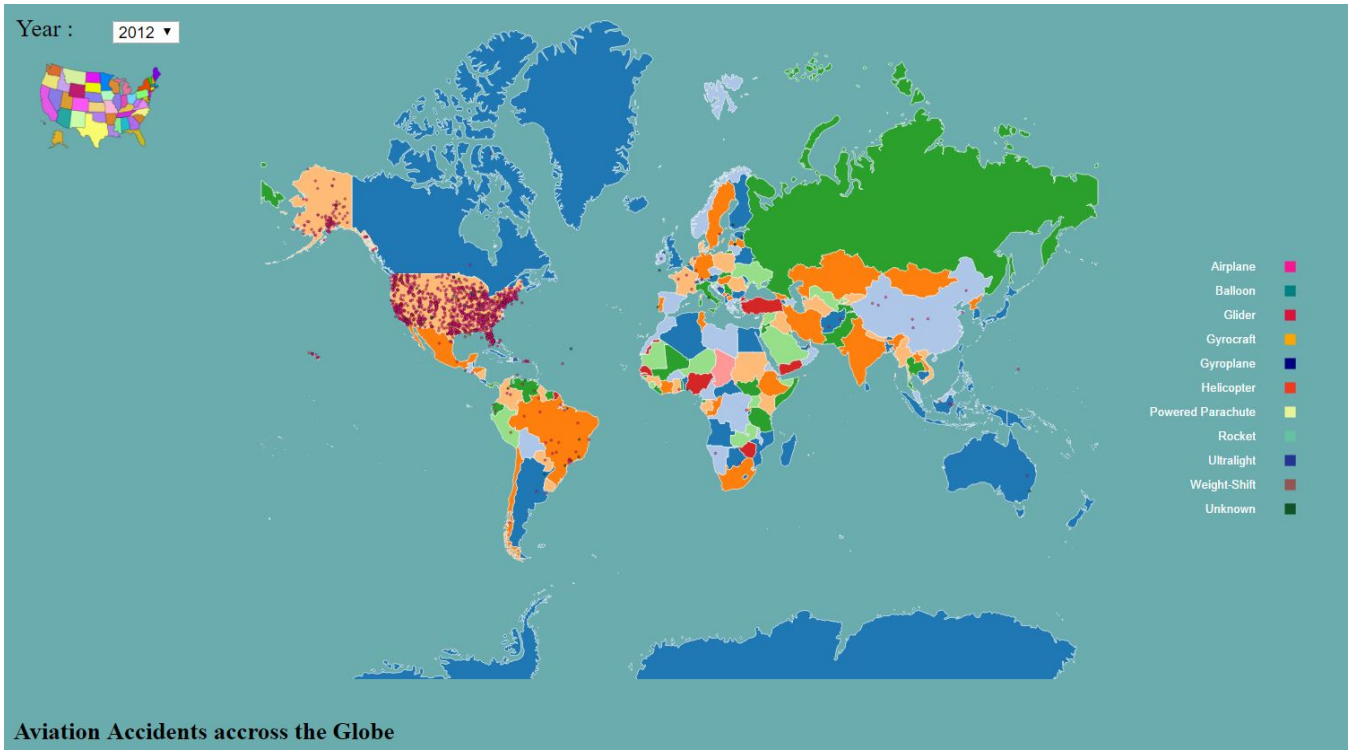

Figure 2: Accident – World Map

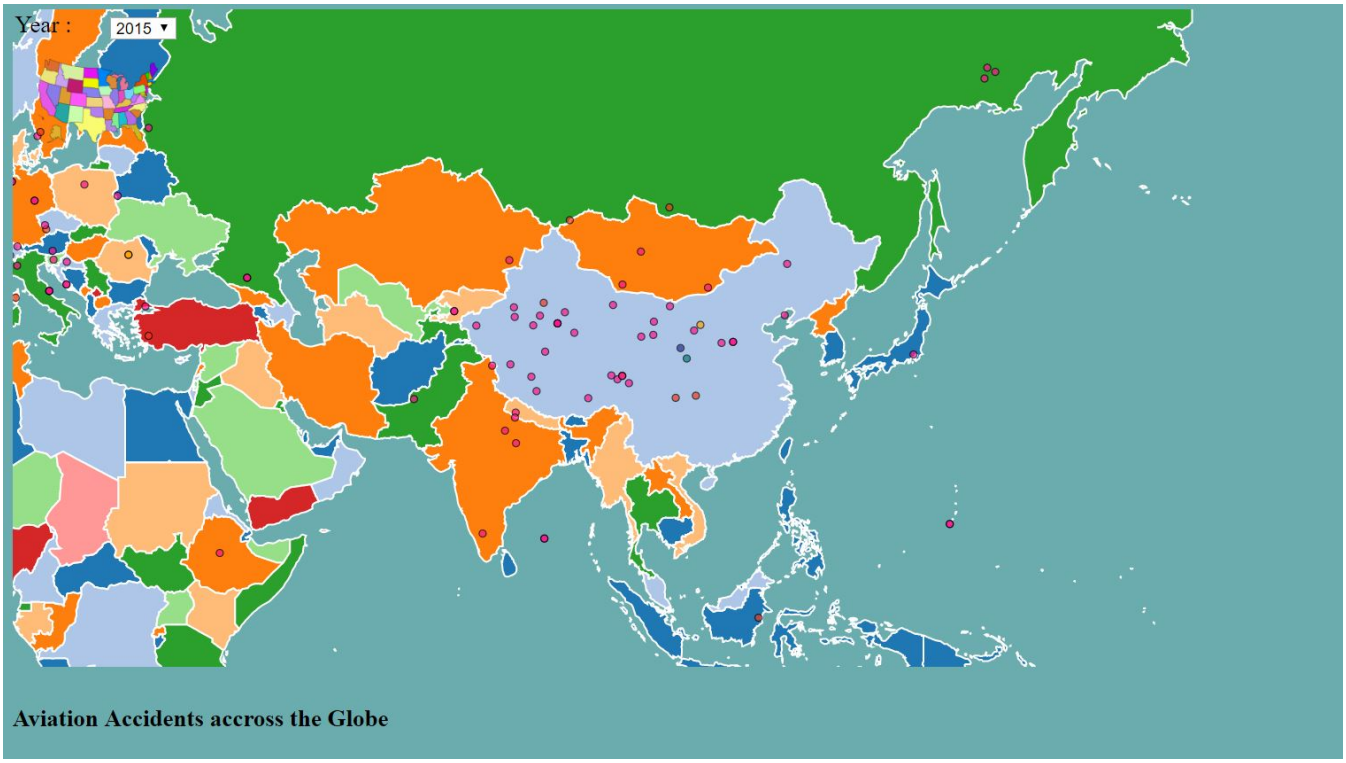

Figure 3: Zoom feature

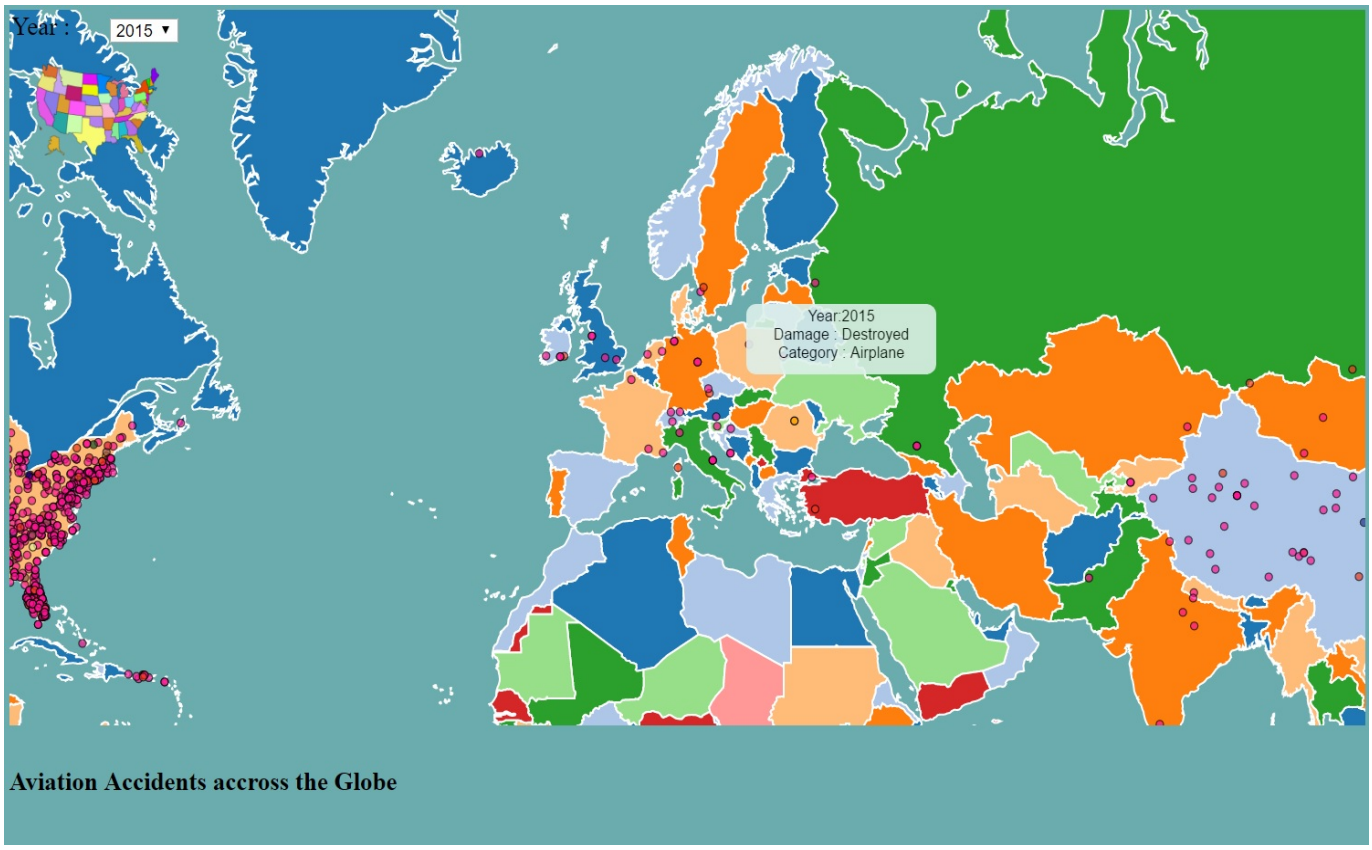

Figure 4: Hover feature

As we can see in figure 1, The USA map has lot of accidents recorded in the dataset. It was hard to visualize the whole USA data in that screen hence we provided a hyperlink which shows the same data just for the USA map. All the properties and components are inherited from the world map.

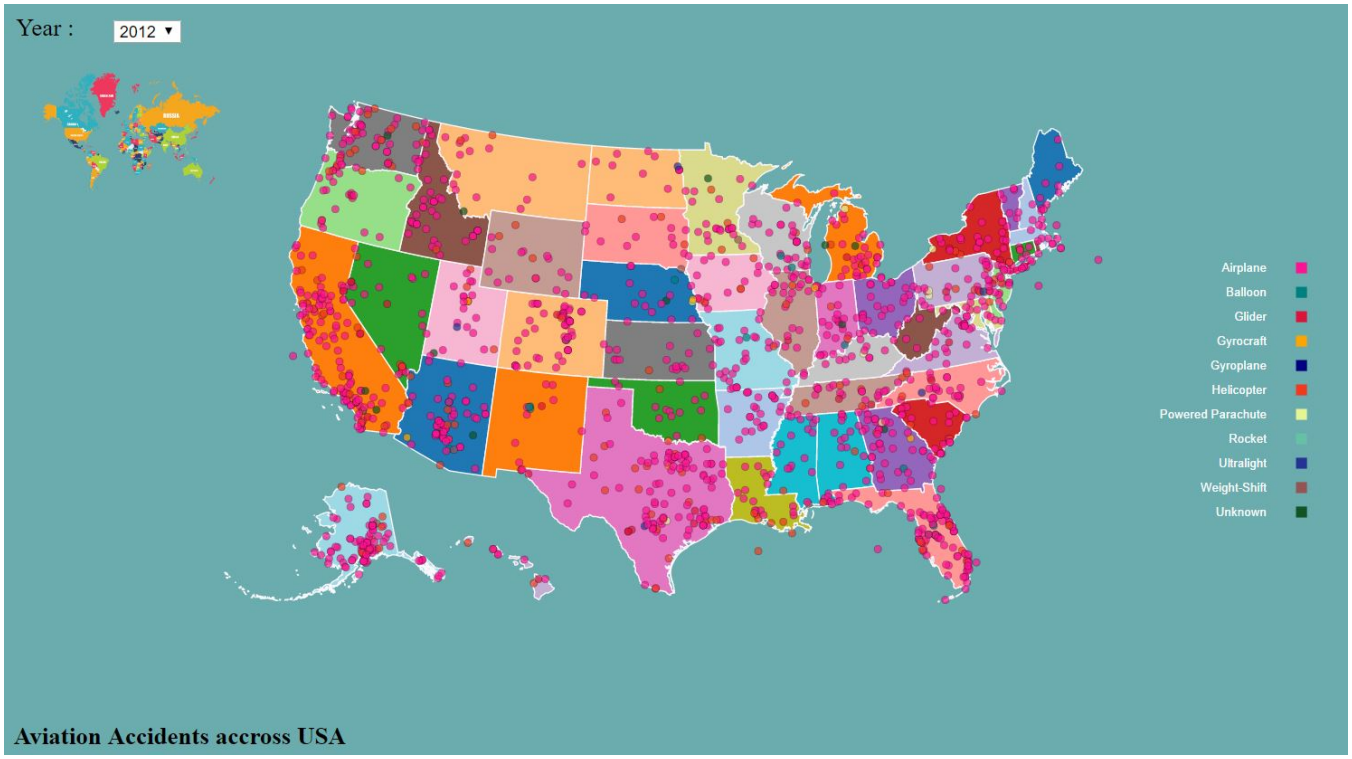

Figure 5: Accidents - USA map

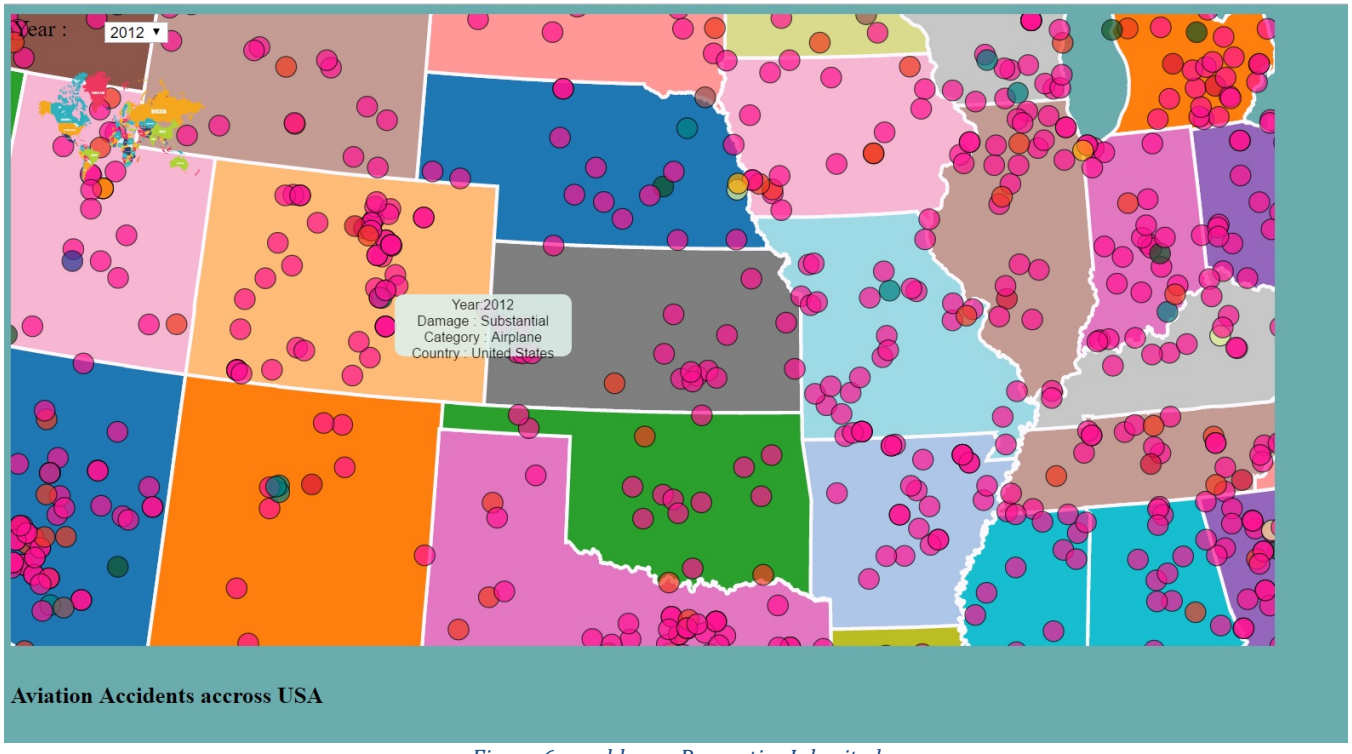

Figure 6: world map Properties Inherited

Bubble Chart[7]:

We chose this type of visualization as it has many features that can be used to represent the data. One of them is **Force Simulation**. This property has an interactive way to operate on the data points. Hence we described various phases in which the accidents occurred.

Data Input: CSV data file**,** aggregated whole data set to **8K records**
Data Attributes Described: Event Year, phase of accident, count
Components Implemented:

Circle plot: Each circle represents a count of the accidents by phase and year
Button Selector: To visualize data in a different way, click any button to have to visualization
Split by Year: Force simulate the bubbles to split according to the year
Merge: Revert to initial state
Split by phase: Force simulate the bubbles to split according the several phases of accidents
Hover: hover over any data point on the screen to have more insights of the accident occurred which are Year and Phase.

Figure 7: Bubble Chart

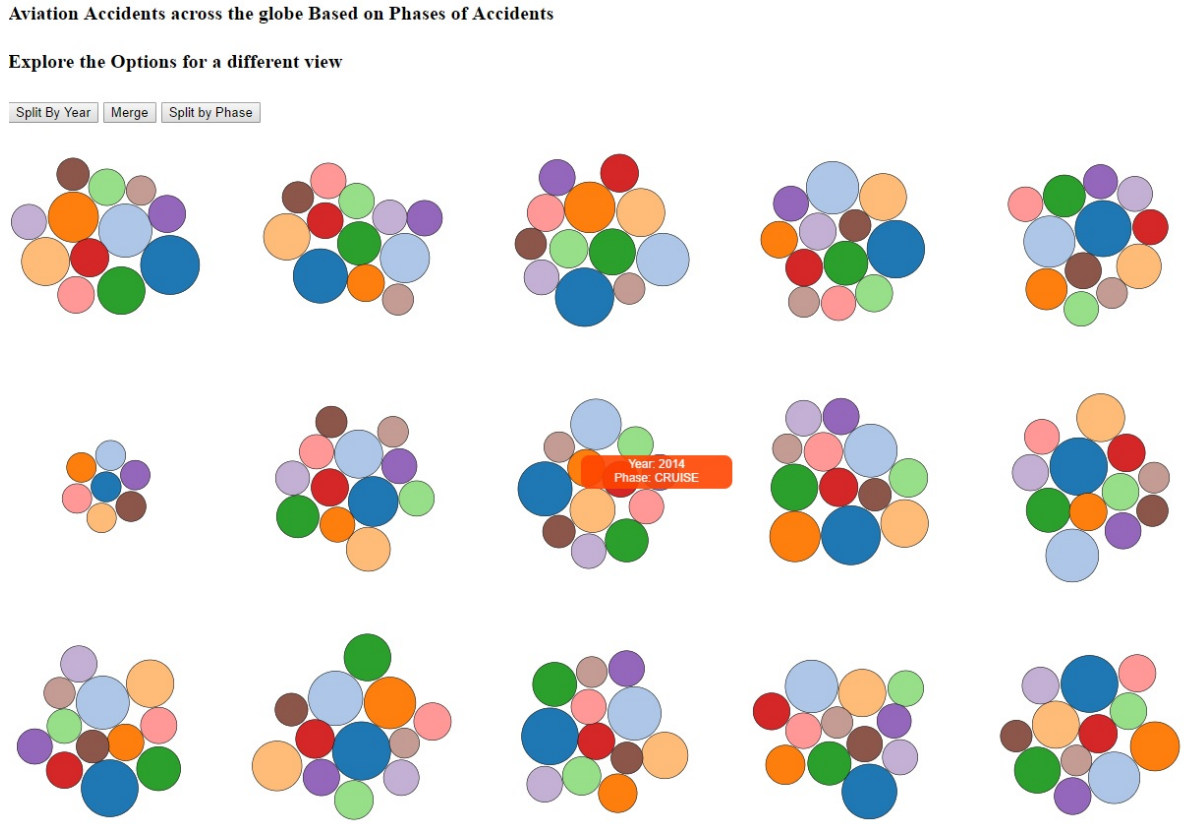

Figure 8: Split bubbles by Year

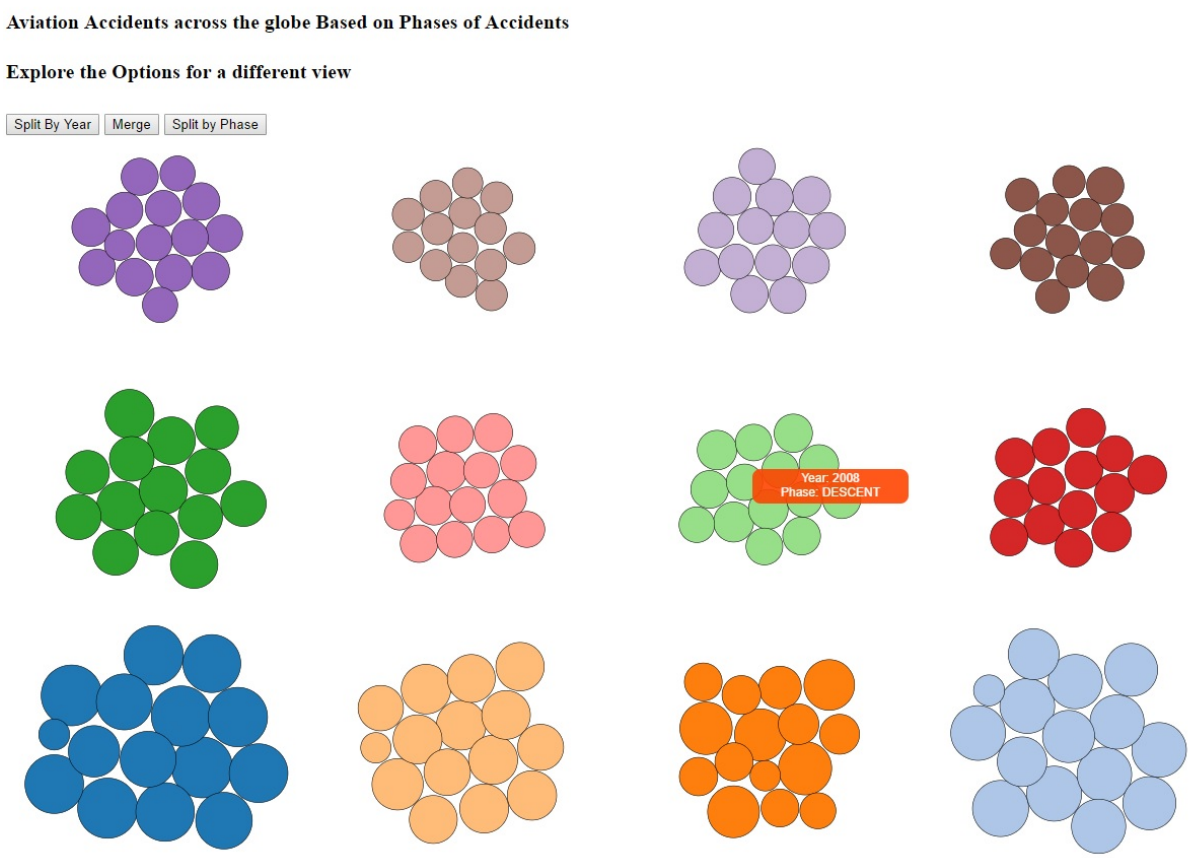

Figure 9: Split bubbles by Phase

HeatMap[4]:

We chose this technique to represent the severity of the accident. The severity of the accident can be distinguished base on the color on the rectangle. Higher the frequency of the color higher the severity of the accident.

Data Input: CSV data file, **,** aggregated whole data set to **1K records**
Data Attributes Described: Accident Year and month, Fatal injuries, Minor injuries, Serious injuries and total uninjured.
Components Implemented:

Rectangle Plot: Each rectangle in the map represents the count of accidents and intensity of injuries for every month of the year (2003 – 2017).

Button Selector: To visualize data in different views

Count of Accidents Distribution: Shows the distribution of accidents counts across all the months of the year.

Fatal Injuries Distribution: Shows the distribution of total fatal injuries across all the months of the year.

Serious Injuries Distribution: Shows the distribution of total serious injuries across all the months of the year.

Minor Injuries Distribution: Shows the distribution of total minor injuries across all the months of the year.

Uninjured Distribution: Shows the distribution of total uninjured across all the months of the year.

Legend: Shows the color intensity that specifies the distribution of the count and severity of accidents

Click: click on any rectangle on the heat map to have value of total accidents or severity of injuries.

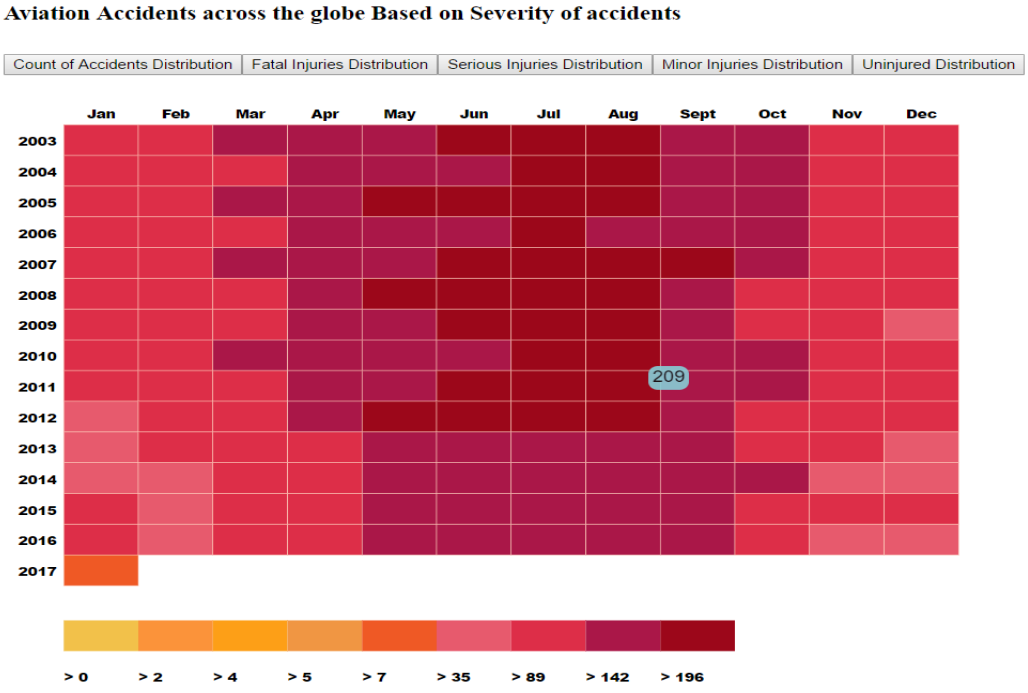

Figure 10 : Heat map - Severity of Accident

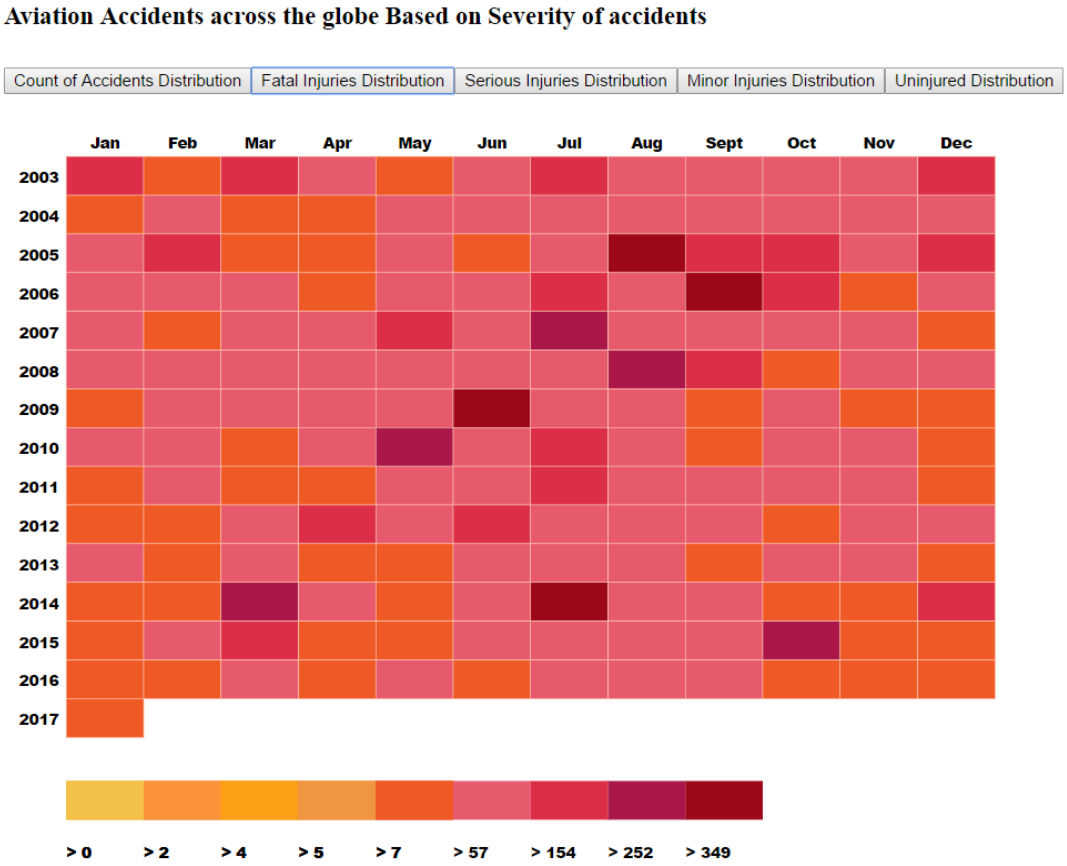

Figure 11: Fatal Injuries distribution

Stream Graph[6]:

We chose this technique to represent another attribute on the dataset which is number of accidents that occurred based on various weather conditions. Stream graph gives a collective visualization where the thickness of the graph shows the count of the accidents

Data Input: JSON data file**,** aggregated whole data set to **500 records**
Data Attributes Described: Type of weather condition, severity of accidents for all types of weather conditions and year of the accidents.
Components Implemented:

Stream Area: Area covered under each stream represents the severity of each category of injuries with respect to the weather condition across the range of years.

Button Selector: To select one category among seriously injuries, fatal injuries, minor injuries and uninjured.

TFI: Shows the Fatal injuries stream according to weather conditions.
TSI: Shows the Severe injuries stream according to weather conditions.
TMI: Shows the Minor injuries stream according to weather conditions.
TUI: Shows the uinjuried stream according to weather conditions.

Hover: hover on the stream path displays the type of weather condition in tooltip and highlights the stream area of that weather condition

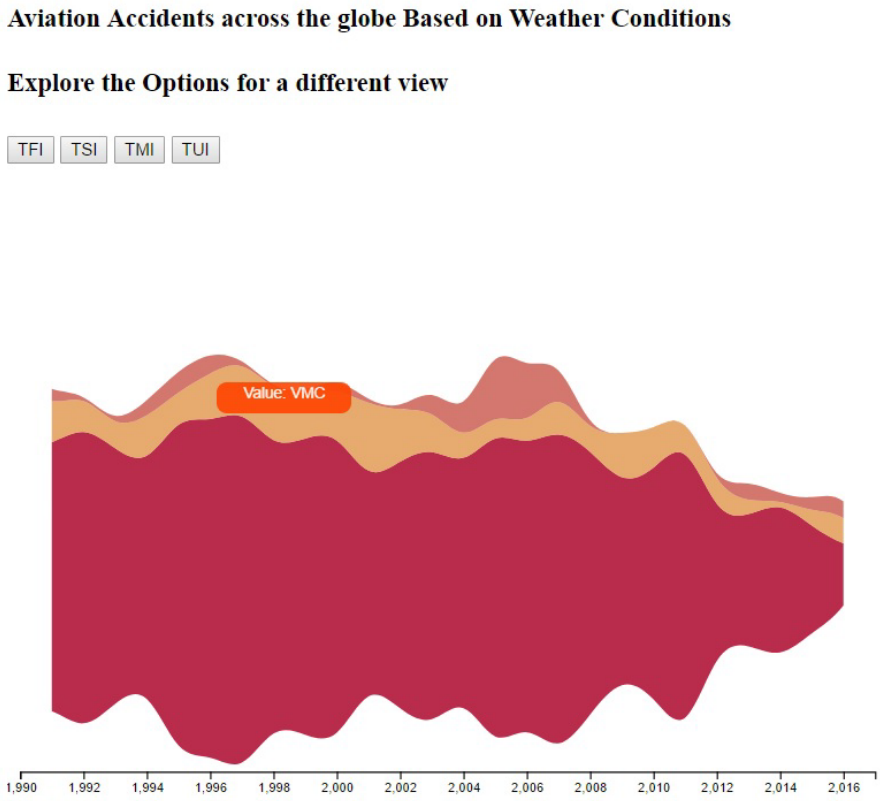

Figure 12: Stream Graph - Weather Conditions

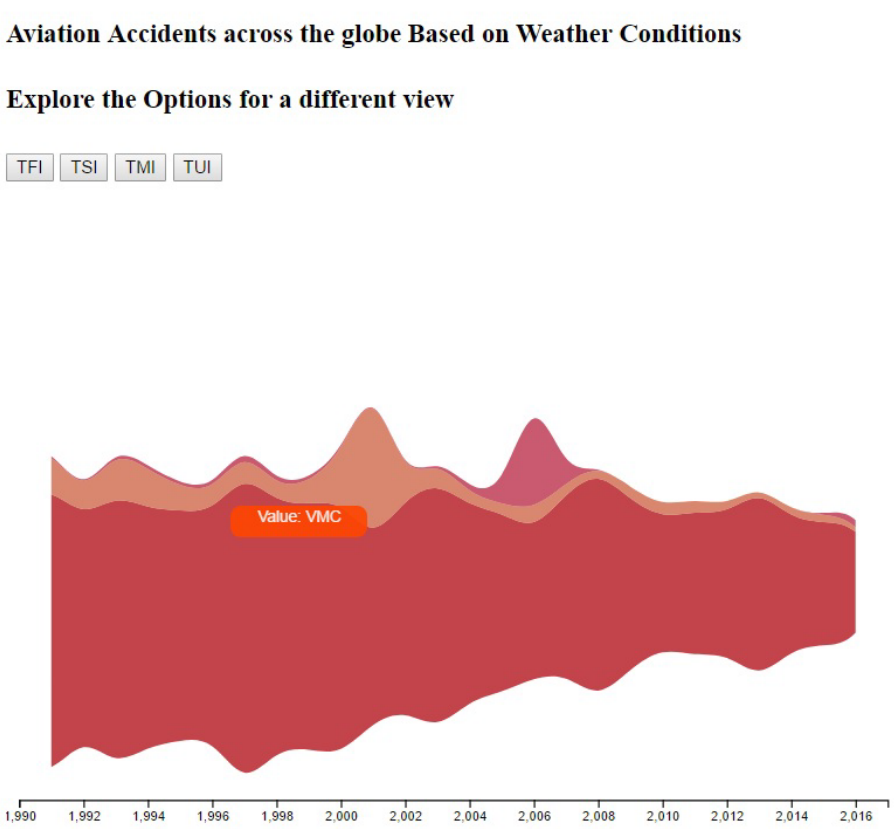

Figure 13: Button Click - TUI data

LEARNINGS AND ISSUES

While executing this project we got acquainted majorly with D3 which is a powerful visualization tool. When worked deeper for the projects it was fun to learn about the SVG, Selects and the type of a shape or structure that needs to be inserted as a basic unit that helps in aggregating the data to represent an attribute. Few of the shapes that are used Path, Circle, Rectangle each of which hold a data value that can be distinguished based on the Size, Color decided by the aggregated data values.

Analysis and data separation was a tedious task. Huge data was slow in processing. We gave trails on different types of data inputs JSON, CSV, TSV to get the data binding correctly. Geo visualization needed more time for mapping the geo location and projecting on the map layout. Bubble chart force simulation library issues. Data creation for stream graph was another as data was away for the requirement. Hence, we used queried the dataset in SQL to generate the required set.

REFERENCES

[1] https://www.ntsb.gov/Pages/default.aspx
[2] https://d3js.org/d3.v4.min.js
[3] https://vida.io/gists/TWNbJrHvRcR3DeAZq
[4] http://bl.ocks.org/tjdecke/5558084
[5] https://bl.ocks.org/mbostock/4090848
[6] https://bl.ocks.org/mbostock/4060954
[7] http://vallandingham.me/bubble_chart_v4/#