Visualizing the Storm Data

Nivedita Vattipalli, Ravali Basna, Rohitha Reddy Muppidi

Abstract— The aim of our project is to analyze the high dimensional Storm information published by National Weather Service(NWS) and visualize the results. The Storm Events Database published by NWS provides data about different event types occurred in different states along with the type of damages, injuries and deaths occurred, the cause of these events. We have taken the storm events of all the States in US and analyzed the damage occurred and precautions to be taken based on the intensity and frequency of storms. We are showing the obtained results in the form of various graphs and techniques like donut pie chart, heat map, bar graph, pie chart and interactive bar chart. Data visualization mainly helps people to understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software.

Index Terms— High Dimensional Data, Data Visualization, Charts, D3.js

1 Introduction

Visualizations can be greatly enhanced by interaction and animation. One way to classify a data visualization is by counting how many different data dimensions it represents. By this we mean the number of discrete types of information that are visually encoded in a diagram. The count of the number of data dimensions can be described as the level of complexity of the visualization. As visualizations become more complex, they are even more challenging to design as well, and can be more difficult to learn from. For that reason, visualizations with no more than three or four dimensions of data are the most common—though visualizations with more dimensions can be found. A primary goal of data visualization is to communicate information clearly and efficiently via statistical graphics plots and information graphics. Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message. Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. Data visualization is both an art and a science. Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects (e.g., points, lines or bars) contained in graphics. The goal is to communicate information clearly and efficiently to users. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Data visualization is closely related to information graphics, information visualization, scientific visualization, exploratory data analysis and statistical graphics.

Everyone deals with storms at some point or another. So, some expanded knowledge about them is always a benefit. Although storms cannot be prevented, being storm ready can help communities minimize the loses.

Looking at a high dimensional data, it is difficult for a person to analyze and make assumptions out of it. If such data is visualized, it gets easy for a person to get a better understanding of the data. When it comes to analyzing a real-world data, it is very difficult to imagine and analyze it manually. So, our project focuses on visualizing the Storm events dataset, analyzes the states where there are highest number of storms in United states, what are the reasons behind these storms, number of deaths and damages incurred so that respective precautionary measures can be taken. Our project visualizes the storm data from 2011 to 2017.

2 Related work

The data from the storm event database is plotted on the United States map to analyze different storm events in different states. Depending upon the location and the weather conditions, different states have different impact and occurrence of storms. We used heat map technique to plot data on the US map such that the state with the more number of storms can be shown with a color having highest intensity whereas the state with lowest number of storms is shown with color having least intensity. This is the first step we made to visualize the storm data by performing data visualization and represented the data visually in the form of different graphs which depicts some useful information. We have also performed data cleaning, data preprocessing, data analysis and attribute identification before plotting useful graphs.

3 Methodology and implementation

Data Cleansing and Data Pre-processing:

The dataset that we have considered for visualization contains some inconsistent and incomplete data. So, we have performed data cleansing and data pre-processing activity to change it to understandable format based on the trends observed in the data.
Below are the data cleansing activities that we have

performed. This task is performed by all three of us.

1. The damages to property and crops occurred in dollars is mentioned in K and M which means thousand and million. But when we write queries to visualize this data, the comparison could not be proper because different data is mentioned in different formats. So, we have added 0's by removing K and M and made it uniform.
2. For TOR_F_SCALE which represents intensities of tornadoes, the values should be in the range of EF0, EF1, EF2, EF3, EF4, EF5 and EFU. Where EFU is not a valid intensity. So, depending upon the behavioral pattern of the data, we have modified the TOR_F_SCALE value accordingly for tornado data.

Data Analysis and Attribute Identification:

This task is performed by Nivedita Vattipalli.
There are 51 attributes in total and depending upon the use cases we planned to implement, I have narrowed it down to 9 attributes by analyzing the use of each attribute. Below is the sample data sheet snapshot:

STATE	MISSISSIPPI	INDIANA	NORTH CAROLINA
YEAR	2011	2015	2016
EVENT_TYPE	Flood	Flash Flood	Flood
DEATHS_DIRECT	0	0	2
DEATHS_INDIRECT	1	2	2
DAMAGE_PROPERTY	1000000	20000	62100000
DAMAGE_CROPS	100000	1000	20000000
MAGNITUDE			
TOR_F_SCALE			
FLOOD_CAUSE	Heavy Rain / Snow Melt	Heavy Rain	Heavy Rain / Tropical System

The Headers in the column 1 are the attributes identified and below is the significance of each attribute:
STATE: This represents the state name where the event occurred.
YEAR: This is the four-digit year for the event in the storm database.
EVENT TYPE: It describes the type of event occurred like Tornado, Thunder storm wind, Lightning etc., There are 52 such event types in storm database that are most frequently occurred.
DEATHS_DIRECT: It describes the number of deaths directly related to the weather event.
DEATHS_INDIRECT: It describes the number of deaths indirectly related to the weather event.
DAMAGE_PROPERTY: The estimated amount of damage to property incurred by the weather event in dollars.
DAMAGE_CROPS: The estimated amount of damage to crops incurred by the weather event in dollars.
MAGNITUDE: It is the measured extent of the magnitude type ~ only used for wind speeds and hail size.
FLOOD_CAUSE: It is the reported or estimated cause of the flood. E.g.: Ice Jam, Heavy Rain etc.,

As the data contains lots of information, we planned to visualize the given data in the form of various graphs. Heat map, bar chart, pie chart and interactive bar chart are generated using d3.js and donut pie chart is generated through high charts using java script implementation. We have used html to integrate all the visualization charts generated. Charts with multiple dimensions are generated through Tableau.
D3.js: D3 means Data Driven Documents. D3.js is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards.
High charts: High charts is a charting library solution built using pure JavaScript, providing users with a quick and simple platform of integrating charts to their web sites or web-based applications. High charts offer a vast selection of supported chart types like pie chart, donut char etc.,
Tableau: Tableau is an industry leading BI tool that focuses on data visualization, dashboarding and data discovery. Tableau can help anyone see and understand their data, connect to almost any database to create visualizations.

In this project, we focused on analyzing the storm data to
1. find the states where there are highest number of storms.
2. Top 5 states with more number of tornadoes for each intensity level – EF0, EF1, EF2, EF3, EF5 and EF5.
3. find the total number of deaths caused from 2011 to 2017 for different levels of tornado severity and the total number of deaths happened in each year for different tornado intensities.
4. the damages to the property incurred in dollars by the weather event in each state for different event types from 2011 to 2017.

Individual contribution of Nivedita Vattipalli:

1. Created a web application as below:

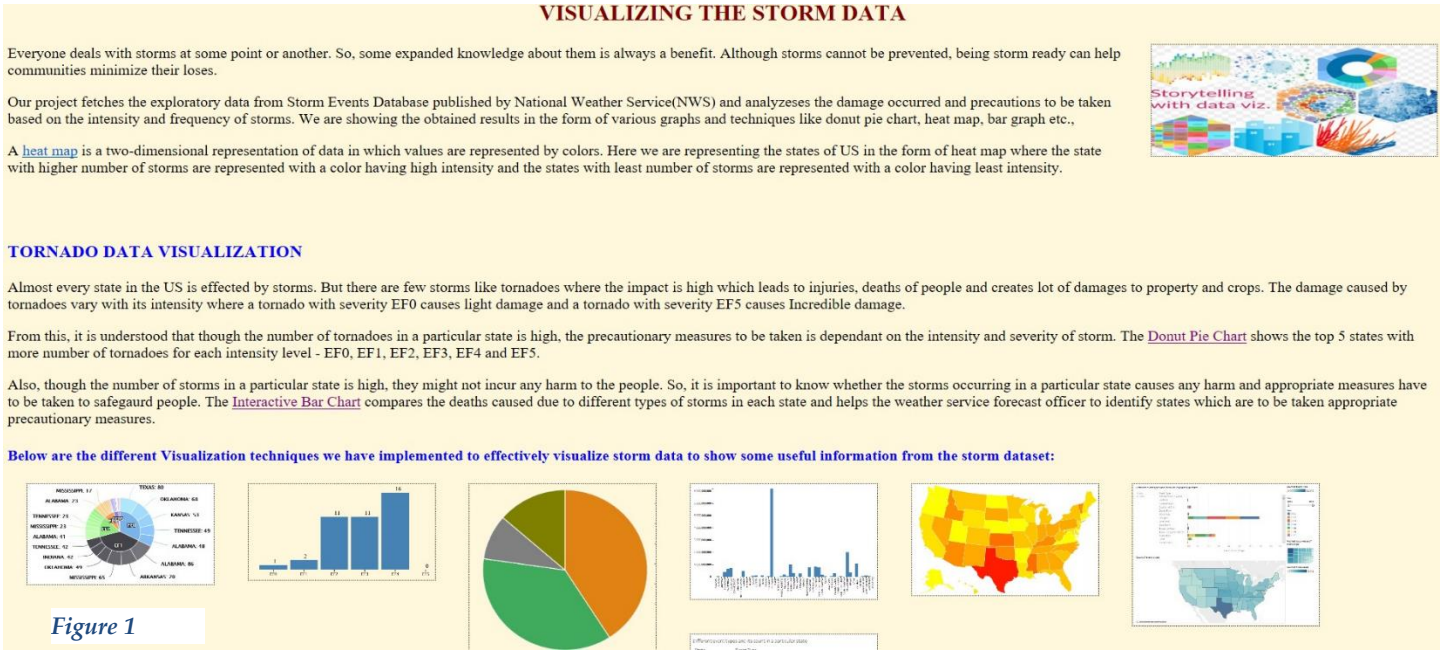


Figure 1

Above snapshot (*Figure 1*) represents the Home page of our application which depicts all the visualization techniques we have implemented in this project.

2. Donut Pie Chart:

Just like a pie chart, a doughnut pie chart shows the relationship of parts to a whole.

Donut pie chart generation is done for the F-Scale damage level of top 5 states using High Charts which use java script code implementation. Our graph (*Figure 2*) shows the top 5 states for each F-Scale intensity for the storm data collected from 2011 to 2017.

Use case: *As a National Weather Service forecast officer, one would like to know the areas which are severely affected by the EF4 and EF5 tornadoes (these are the tornadoes with more intensity) such that more precautionary measures can be taken in those areas to avoid damages to people and property.*

Here we have visualized the states which are mostly affected by tornadoes based on the F-Scale Intensity.

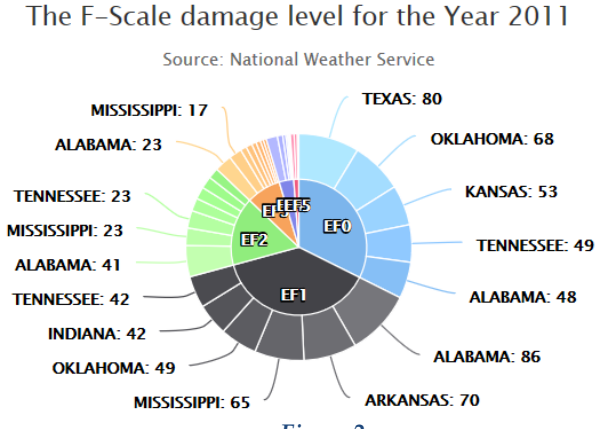


Figure 2

Similar donut pie chart is implemented for all the tornado information that we have from years 2011 to 2017.

3. Interactive Bar and Pie Charts:

Interactive Bar Chart and Pie Chart implementation is done using D3.js which makes use of the widely implemented SVG, HTML5, and CSS.

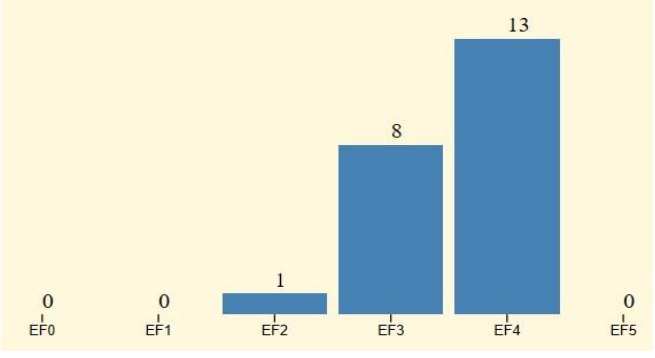


Figure 3

Use case: *The assessment of damage level based on the deaths and damage levels helps the Government and the public to take necessary precautions to protect and save themselves from the upcoming storms.*

BAR CHART in *Figure 3* above shows the total number of deaths caused from 2011 to 2017 for different levels of tornado severity.

PIE CHART in *Figure 4* below shows the total number of deaths happened in each year for different tornado intensities.

These charts show the total number of deaths caused from 2011 to 2017 for different levels of tornado severity and the total number of deaths happened in each year for different tornado intensities.

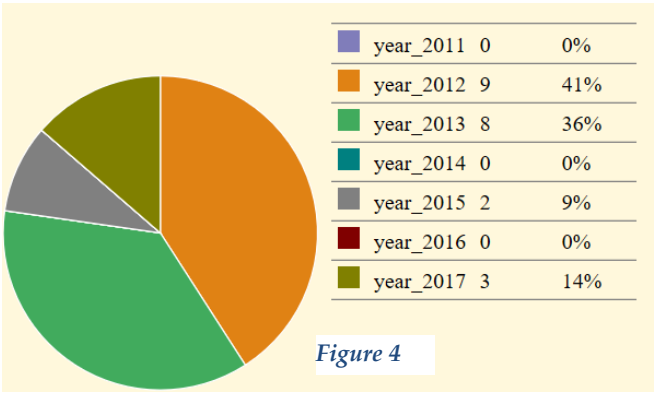


Figure 4

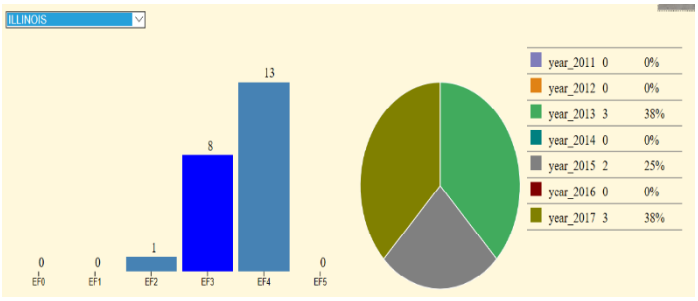


Figure 5

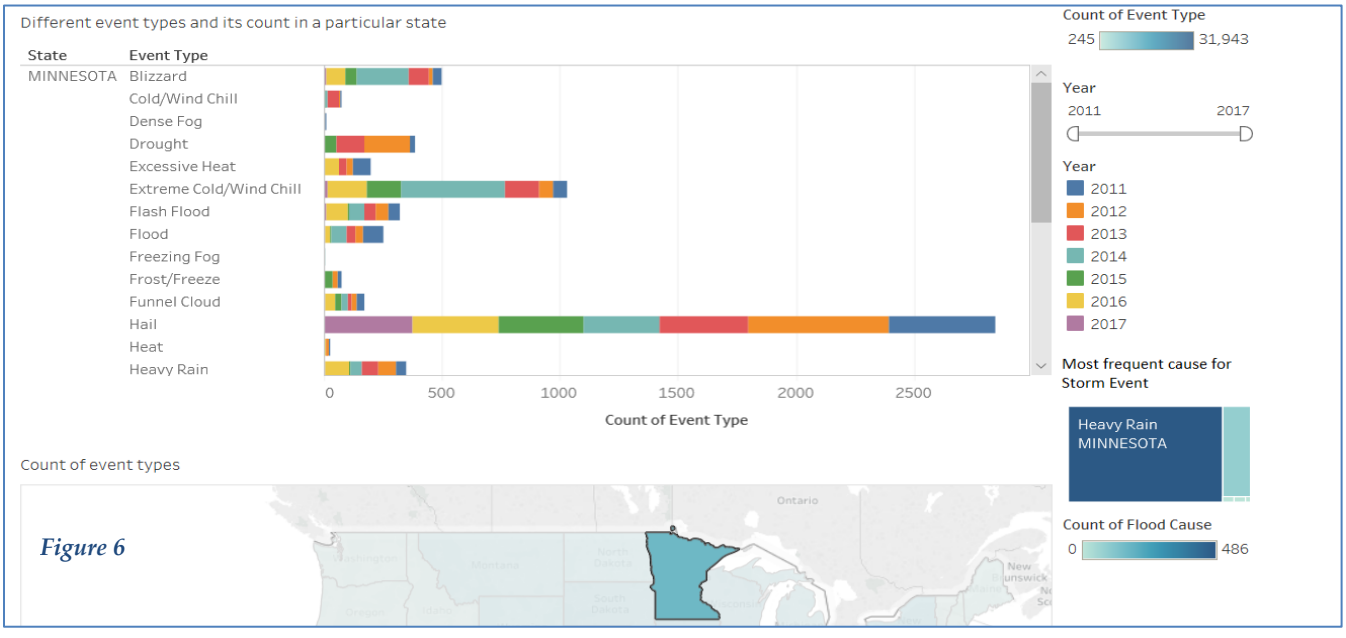


Figure 6

On hovering any of the bars in bar graph (*Figure 5*) shows the number of deaths occurred in different years for that tornado severity in pie chart. Here I have hovered on EF3 bar and it showed the number of deaths in different years in the pie chart.
I have added a dropdown which lists all the states in US and user will be able to view the statistics for each state.
Each Year is identified by a different color coding scheme in the generated Pie Chart.

4. Tree Map implementation using Tableau

Usecase: To help the weather service forecast officer in identifying states which are to be taken appropriate precautionary measures based on the reason for occurrence of storm.

This interactive chart help in
1. finding out the total number of storms in a particular state from 2011 to 2017.
2. different storm types that occurred in each state from 2011 to 2017
3. the major reason for the occurrence of storms in that state.
4. For how many storms, this cause was the reason for the storm occurrence.

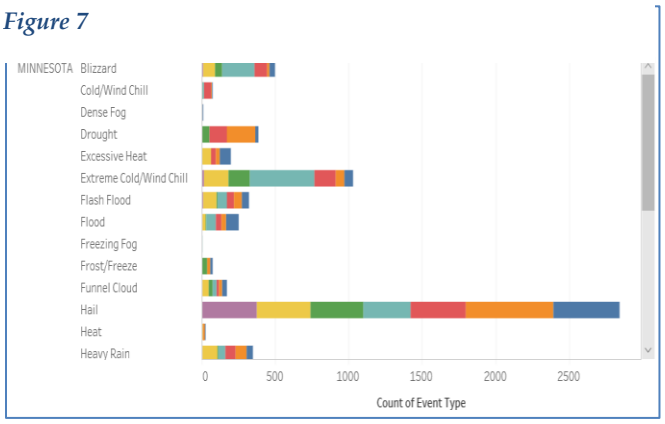
Above snapshot(*Figure 6*) shows the tree map to the right which shows the most frequent cause for storm event whenever a state in the US map is selected. Here Minnesota is selected in the US heat map, then the tree map shows the frequent cause and the bar below the tree map shows the count of floods which occurred because of this cause.

Individual Contribution of Ravali Basna

1. Heat map and Bar chart implementation in Tableau

Refer to *Figure 6* wherein heatmap is generated for the number of storms occurred in US from 2011 to 2017 and the state with highest number of storms is shown a color having highest intensity and a state with least number of storms is shown with a color having less intensity.
I have also generated a bar chart (*Figure 7*) which shows the different storm events occurred in each state from 2011 to 2017 where data corresponding to each year is highlighted with a different color. On hovering on the graphs shows the corresponding counts.

Figure 7



2. Interactive Bar Chart implementation in D3.js

Use case: To help the weather service forecast officer in identifying states where there are more damages so that respective precautionary measures can be taken to protect the properties from the damages caused due to different storms.

Here we have visualized the damages to the property incurred in dollars (*Figure 8*) by the weather event in each state for different event types from 2011 to 2017.

The Values in X-axis represents different states in US and the Values in Y axis represents the damages incurred in dollars for different event types.

I have also added a dropdown which shows the different types of storm events such that the damage incurred by each storm can be compared across all the states in US.

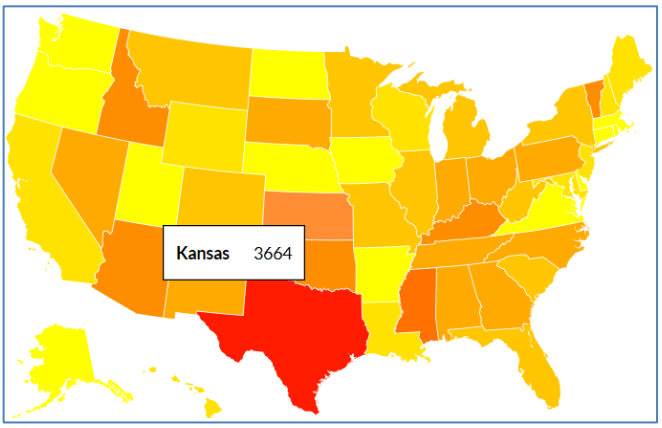


Figure 9

Use case: To find out the state with highest number of storms to predict the climatic conditions of a particular state.

Here I have visualized the number of storms versus each state to understand which state is hit by more number of storms.

I have used heat map color scheme (*Figure 9*) such that the state which is hit by more number of storm is represented by a colour having higher intensity.

The same is represented in the bar graph as well where the states with more number of storms is shown with a bar with greater length.

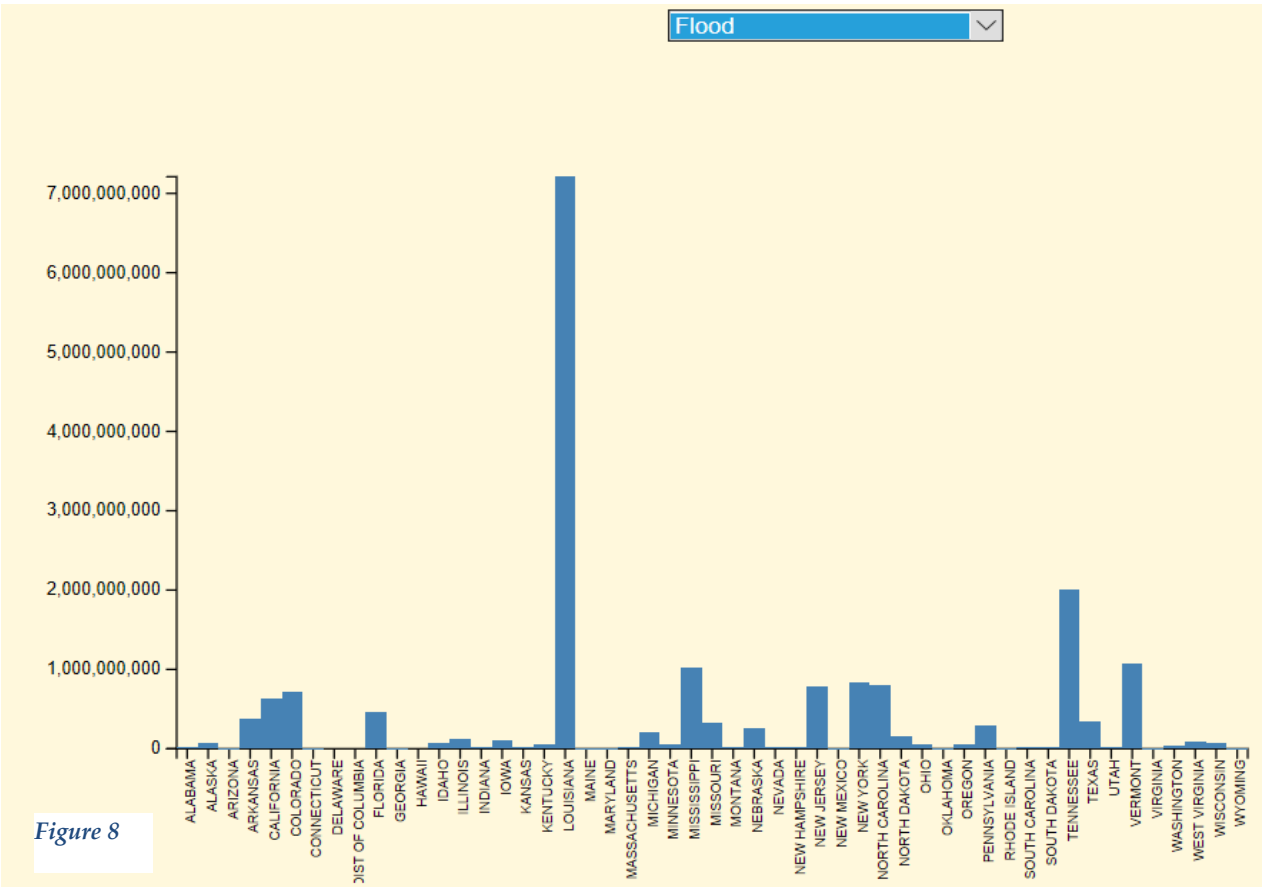


Figure 8

3. Heat map implementation using D3.js

Low values are represented by using one hue and high values by using a different hue.

Individual Contribution of Rohitha Muppidi

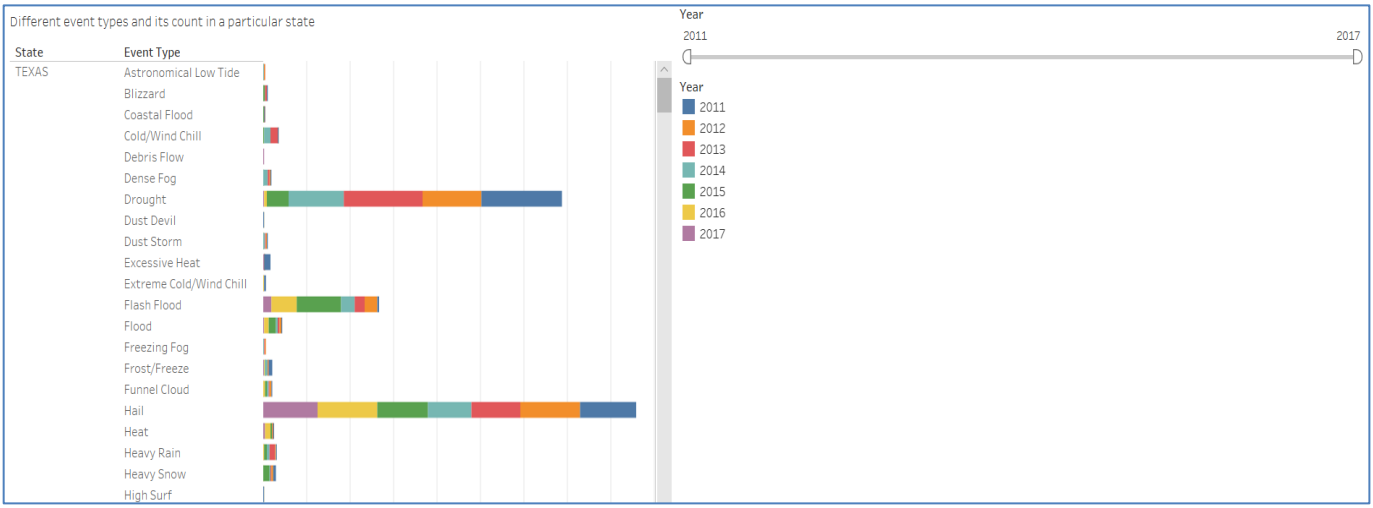


Figure 10

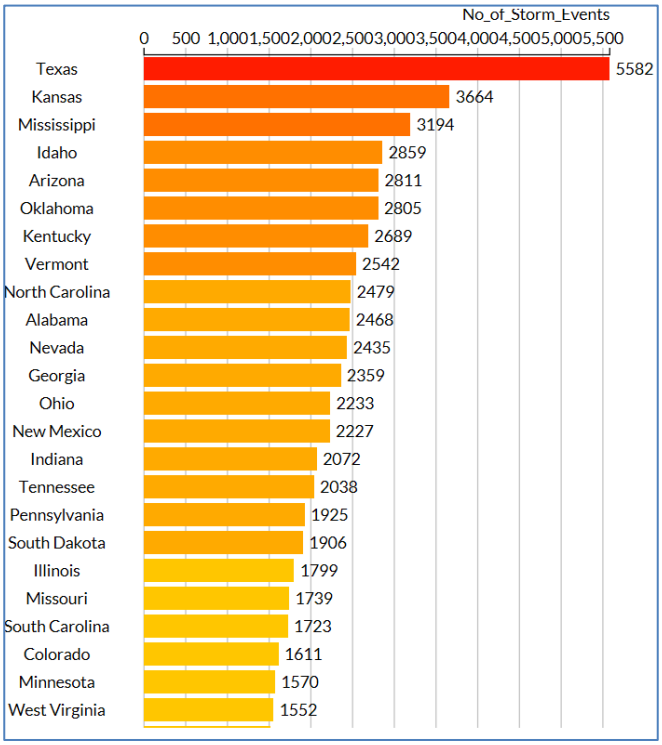


Figure 11

1. Implementation of Bar chart showing the total number of storms in D3.js

The heat map is represented in the bar graph as well where the states with more number of storms is shown with a bar with greater length (*Figure 11*)

2. Implementation of Bar Chart showing the different event types for each state using Tableau

Created a bar chart with States, Event Type and Year data which shows the different storms that occurred in each state of US from year 2011 to 2017.

This graph lets the user select the data corresponding to each year to visualize the data in specific and know the places which are affected by a particular type of storm more often (*Figure 10*)

Results and Discussions

The results of each visualization technique are mentioned along with the implementation. Below is the dataset preparation for each visualization technique implemented.

1. Heat Map: Prepared an excel which includes the count of storm events for each state
2. Donut Pie chart: Prepared a dataset which includes details of the top 5 states for each intensity level of tornado ranging from EF0 to EF5 from 2011 to 2017.
3. Pie Chart and Bar Chart: Dataset is prepared which contains information of deaths occurred for each intensity of tornado in each state.
4. Interactive Bar Chart: Created a file which includes the damages to property and floods caused by each storm in every state of US.

Conclusion and Future Work

After observing all the results, we have drawn below conclusions:

1. Though the number of storms in a particular state is more, it does not mean that the damages or deaths in that place is more. Damages and deaths mainly depend upon the intensity of storms and the magnitude of hails.
2. Identifying the frequent flood cause from the huge data set is difficult. Visualization techniques implemented helped in finding out such information easy.
3. Comparison of damages across each state for a particular storm event helps the weather forecast officials to know which state must be taken more precautionary measures.

As part of future work, we are planning to include one more dimension to the graphs that we have visualized to visualize and compare more information. Also, we are planning to compare the results with the area of each state in case of damages and injuries/deaths with population of the state

REFERENCES

National Centers for Environmental Information [US] – https://www.ncdc.noaa.gov/stormevents/

Introduction to D3.js
https://d3js.org/

High charts
http://www.highcharts.com/

Tableau
https://public.tableau.com/s/