Project Report CAP5738:

Lightning geographic and statistics visualization

Introduction (Lauren Dacre)

This project is intended to provide a demonstration of pedagogical tools that may be used to educate people about the prevalence and danger of lightning strikes. Knowing how common lightning strikes are in our area helps us to gauge our risk level when there is a potential for lightning strikes. In regions where lightning is more common, people must be even more vigilant and proactive during lightning storms. As technology improved our ability to track and notify people of dangerous weather events, we observed a downward trend in the number of lightning strike-related fatalities in the United States. We believe this is a testament to the importance of education and awareness when it comes to avoiding lightning strike casualties (and more generally, severe weather casualties).

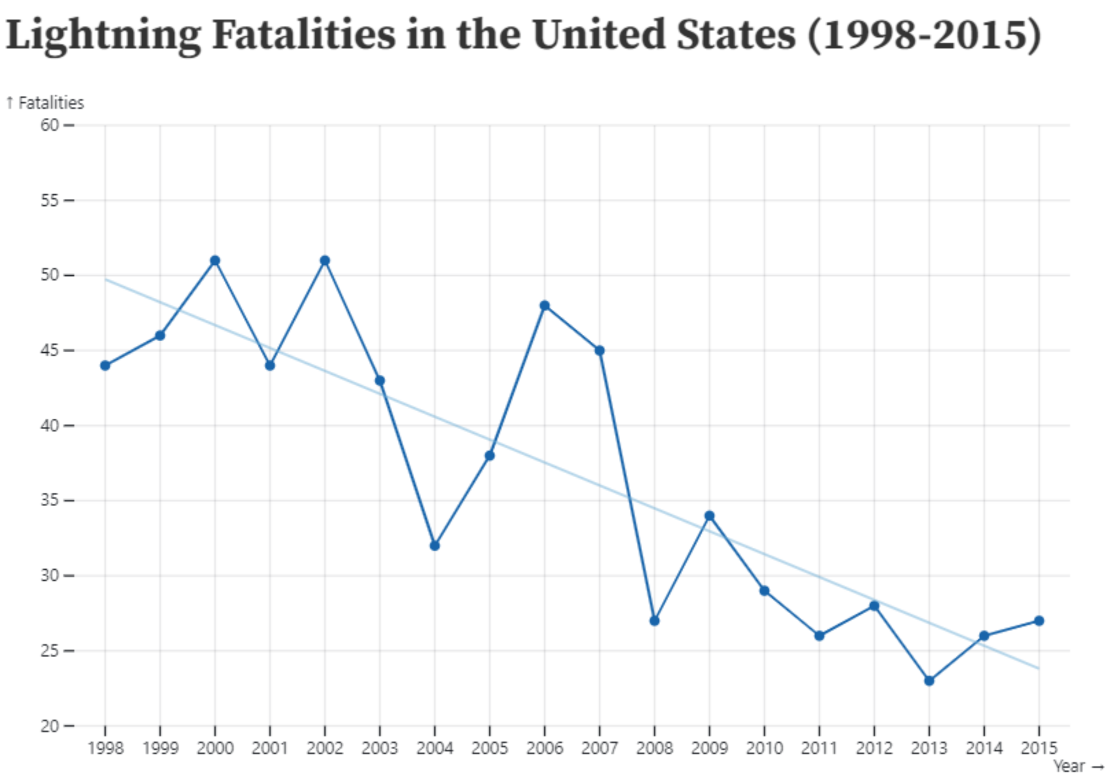

Figure 1: Line chart illustrating the downward trend of lightning-related fatalities in the U.S.

Data description

Gathering and concatenation [Sheikh Nabil]

The dataset is collected from NASA's Earth observing TRMM (Tropical Rainfall Measuring Mission) satellite. The dataset includes data availability from January 1998 to April 2015 for regions all over the world. The TRMM satellite project was terminated after that. Since then, the

lightning data has been collected by the International Space Station (ISS) using LIS sensors. Lightning data from March 2017 to October 2021 is available using the ISS Space Time Domain Search tool. For this project, we used the TRMM satellite dataset which can be downloaded from the LIS Space Time Domain Search Tool:
https://lightning.nsstc.nasa.gov/nlisib/lissearch.html

We downloaded all the dataset for the United states and some parts of Mexico and the Caribbean islands. We downloaded the data in three parts for each of the years from 1998 to April 2015 and then merged those three parts manually for each text file. For the area selection, we selected the center as (Lat 30, Long -115), (Lat 30, Long -96) and (Lat 30, Long -86). The search area is set to 20. The latitude is set to 30 because 40 is the highest latitude for LIS values.

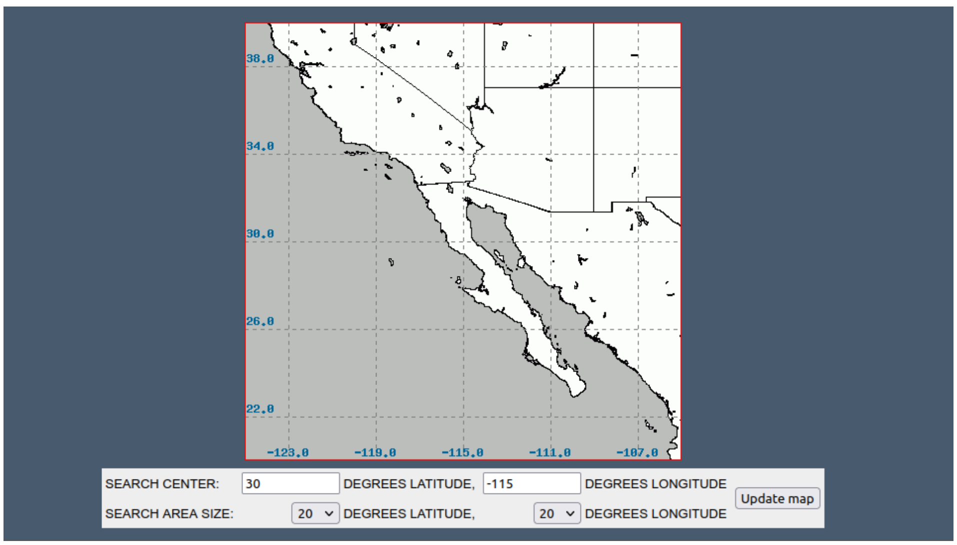

Figure 2: downloading dataset from LIS space dime domain search tool

The dataset is available in this shared google drive link:
https://drive.google.com/drive/folders/1oQKu8yPkHVhPoWw5zpaRkTvgx8BtFBk8?usp=sharing

Observable only allows us to upload files with a maximum size of 50MB, so we manually combined two or three years of data in a text file and uploaded them to Observable. After that we combined the dataset into several versions of different sizes. The final dataset is named data1998_2015; it contains all the lightning strike data for the United States (1998-2015). The state names are combined and unidentified state names are filtered out.

Data processing (Alexis Benamira)

The data came as raw .txt files. I did the parsing function to gather the data into an array.
The information available per strike is: the date and time, the geographic coordinates (latitude & longitude), the intensity in radiant and the time the strike lasted in ms.
The parsing function is called `processdata`.

In order to perform per state statistics and visualizations, each lightning strike needed to be associated with a state. But the only information at our disposal were each state's geographic coordinates and each lightning strike's geographic coordinates. This problem is known as the "Point in polygon" problem [1]. Is a point inside or outside a polygon geometry in 2D. In order to solve this problem, I wrote a function called `inside`.
The approach I took is raycasting. You cast a ray towards the right and count the number of intersections with the polygon boundary. If it is even, then it is outside, otherwise it is inside.

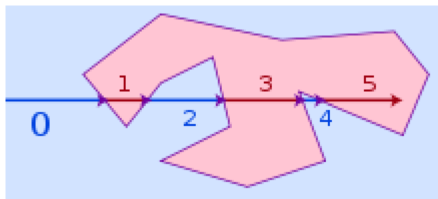

Figure 3: Illustration of the ray casting method to solve the point in polygon problem

With that I associated each lightning strike to a state which was very useful for all the visualization.

Geographic Visualization (Alexis Benamira)

Cartogram

The first map visualization is a cartogram where each state is distorted based on the number of lightning strikes that happened in the state. So the Area of the state is proportional to the number of lightning strikes events. The final code is simple, but it took me a long time to make it work. Indeed the initial Topological data of the US I was using was not working for reasons still not fully understood, I think it is either because there are 56 states in the state section of the US data he used originally (because territories like Guam are included which are not states but an unincorporated US territory). Or because the topological representation of the data was very detailed (so very large) and the topogram function I used did not like it. Once I changed the data to a simplified version of the topology of each state and only the 51 actual states, then it was working perfectly as seen on the capture underneath. For his geographic visualization, I used a subset of the full data, the lightning strikes of only the year 2015. This enables the visualization to appear faster and the interaction to run more smoothly.

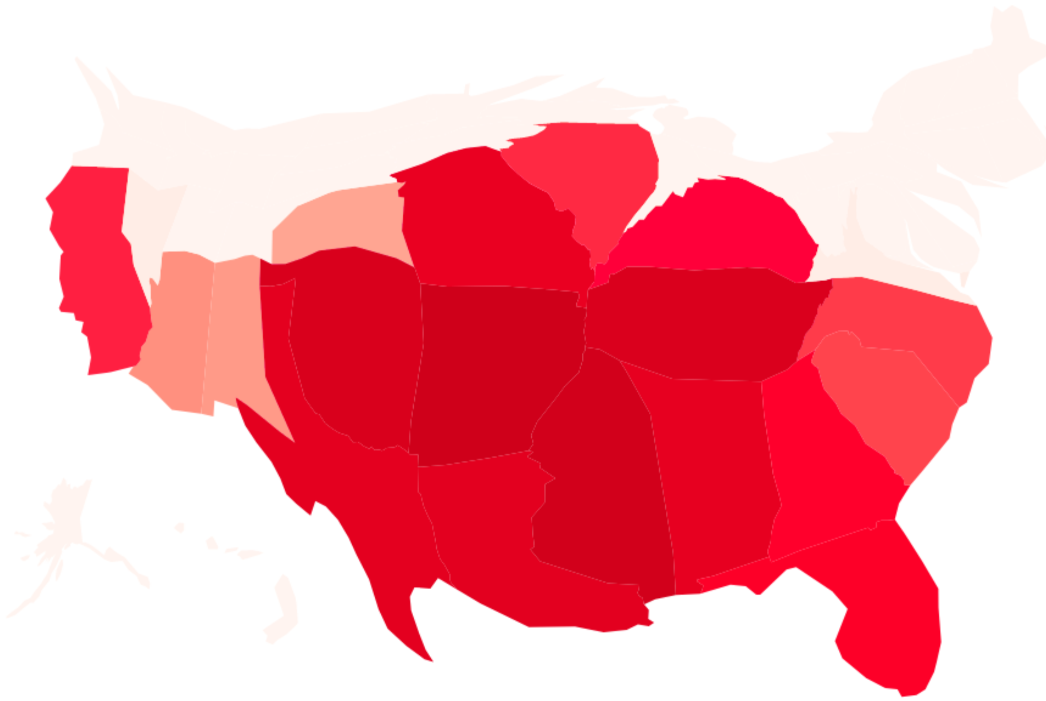

Figure 4: Cartogram of the US where distortion is proportional to the number of lightning strikes

Interactive map

The interactive map was fully made using D3.

The points represent the lightning strikes locations. The color scale for the dots represents the intensity. I have made a function to classify the strikes in 10 categories depending on their intensity in radiant. From category 1 (weakest intensity/paler color) to 10 (strongest intensity/darker colors).

Multiple interactions function were created:

- Hovering over
- Click to zoom
- Hovering over when zoomed in
- Double click when zoomed in (zoom out)

When hovering over the state,its color changes and the name of the state as well as the number of lightning strikes for this state is printed. When exiting the state, the color reverts back to normal. (cf image 3, hover over Florida)

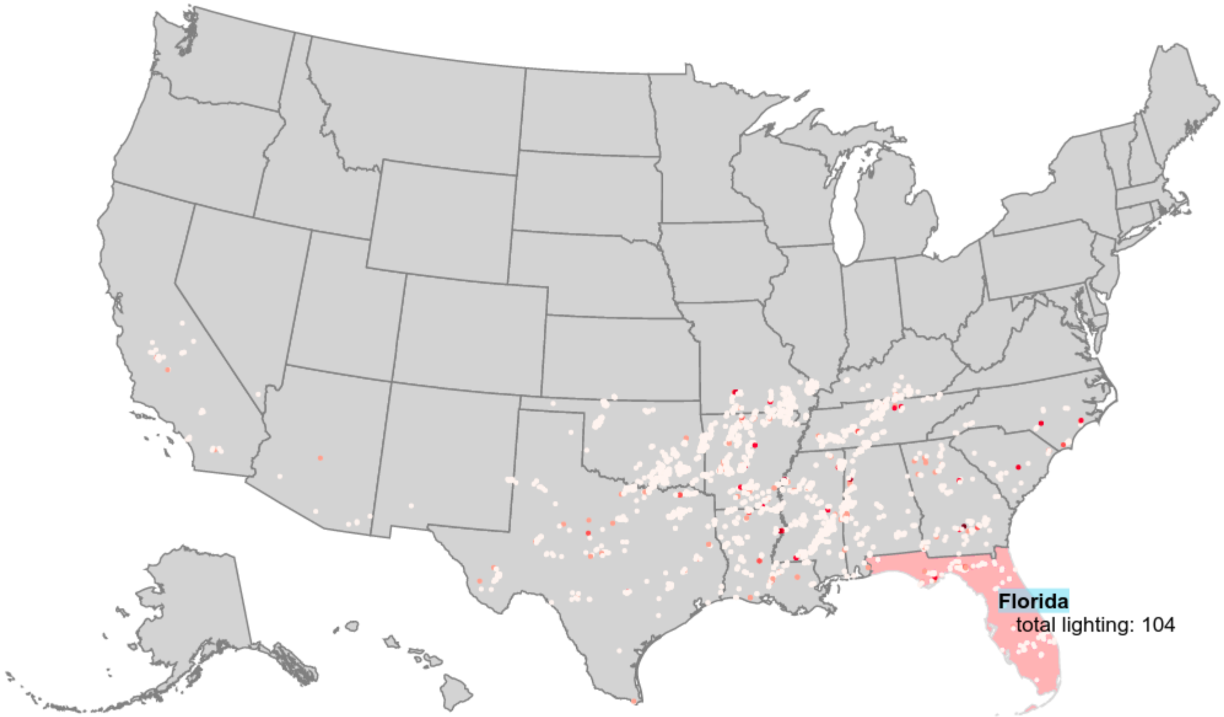

Figure 5: Interactive US map, illustration of the hovering over function

When clicking on a state the state is rescale to appear bigger and the rest of the map disappears. The proper translation and scaling factor based on the state location and area was implemented. This clicking to zoom function does not work for Florida and Alaska. (cf image 4)

Once zoomed in, if you hover over a point representing a lightning strike, then detailed information about this lightning strike is displayed : the date, the intensity, the category (1 to 10) and the time it lasted. (cf image5).

If you double click while zoomed in, the initial full US map is displayed again.

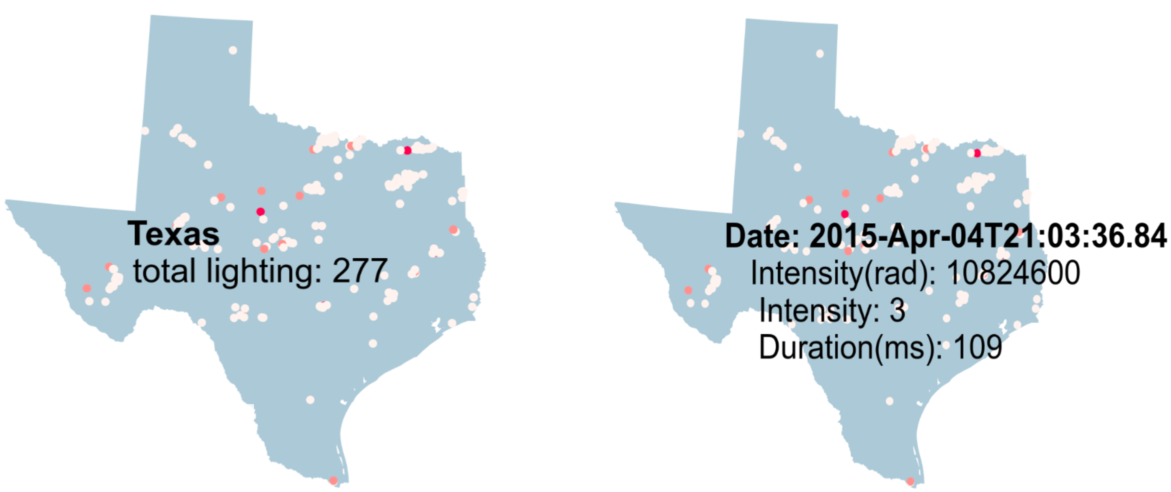

Figure 6: Interactive US map, illustration of the zoom in and zoomed hovering over function

With these interactive visualizations I was able to explore D3 functionality and experiment with it which is what I wanted to do in this project.

Statewise Analysis and Over the year trend analysis (Sheikh Nabil)

After merging and concatenating the dataset, I parsed the dataset using Alexis's parsing functions. The initial version of data1998_2015 had 3111500 rows. After filtering the rows for which the state is unknown or invalid, there are 2010354 rows in total.

After parsing the dataset, we get the following attributes:

- Date and time of the event
- Location (Latitude and Longitude)
- **Radiance** (the electromagnetic force of the lightning event)
- **Duration** (the flash duration in milliseconds)
- Year
- State

From the above list of attributes, we used the radiance and duration attributes for our analysis. To work further with the full dataset, we need to understand the distribution of values in the attributes. To effectively observe the spread of values inside those attributes, we first generated boxplots for the whole dataset for these two attributes:

Box Plots visualization for Radiance and Duration values

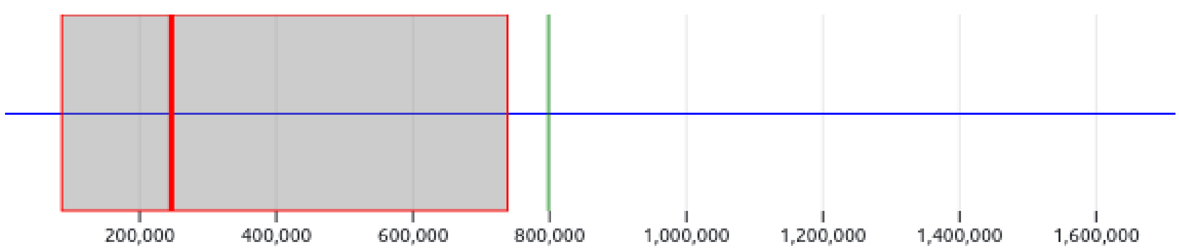

Figure 7: distribution of values for radiance attribute (from 1998 to 2015)

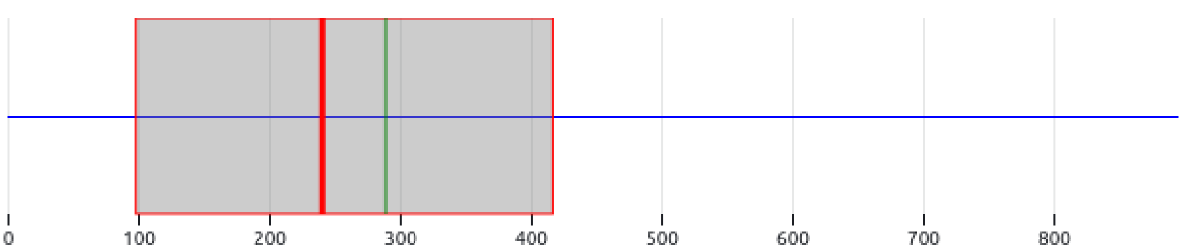

Figure 8: distribution of values for duration attribute (from 1998 to 2015) using box plots

The mean value is also marked in the figure (in green) to identify how spread the distribution of values is for the radiance attribute. For the duration attribute, the distribution is not sparse, the mean is closer to the median. However, for the radiance values, the quartile 3 values are more spread than the quartile 2. Even the mean value is located above the quartile 3, which indicates higher outlier values in the dataset. Researchers and the public may be interested in whether the lightning event has enough energy to cause fatalities or property damage. For this, we filtered the dataset and kept the upper quartile radiance values (above quartile 3). For the duration values, the distribution of data is evenly spread, so we only filtered the lower quartile data for our analysis. The final filtered dataset contains a total of 462364 rows.

Apart from boxplot analysis, we also created a group bar chart to visualize the spread of radiance values for the top ten states with the highest number of lightning occurrences. Figure 7 shows the mean value is located much higher than the median value. This indicates that the outliers in the original dataset are very high.

Group Bar Chart: statewise mean and median Radiance analysis

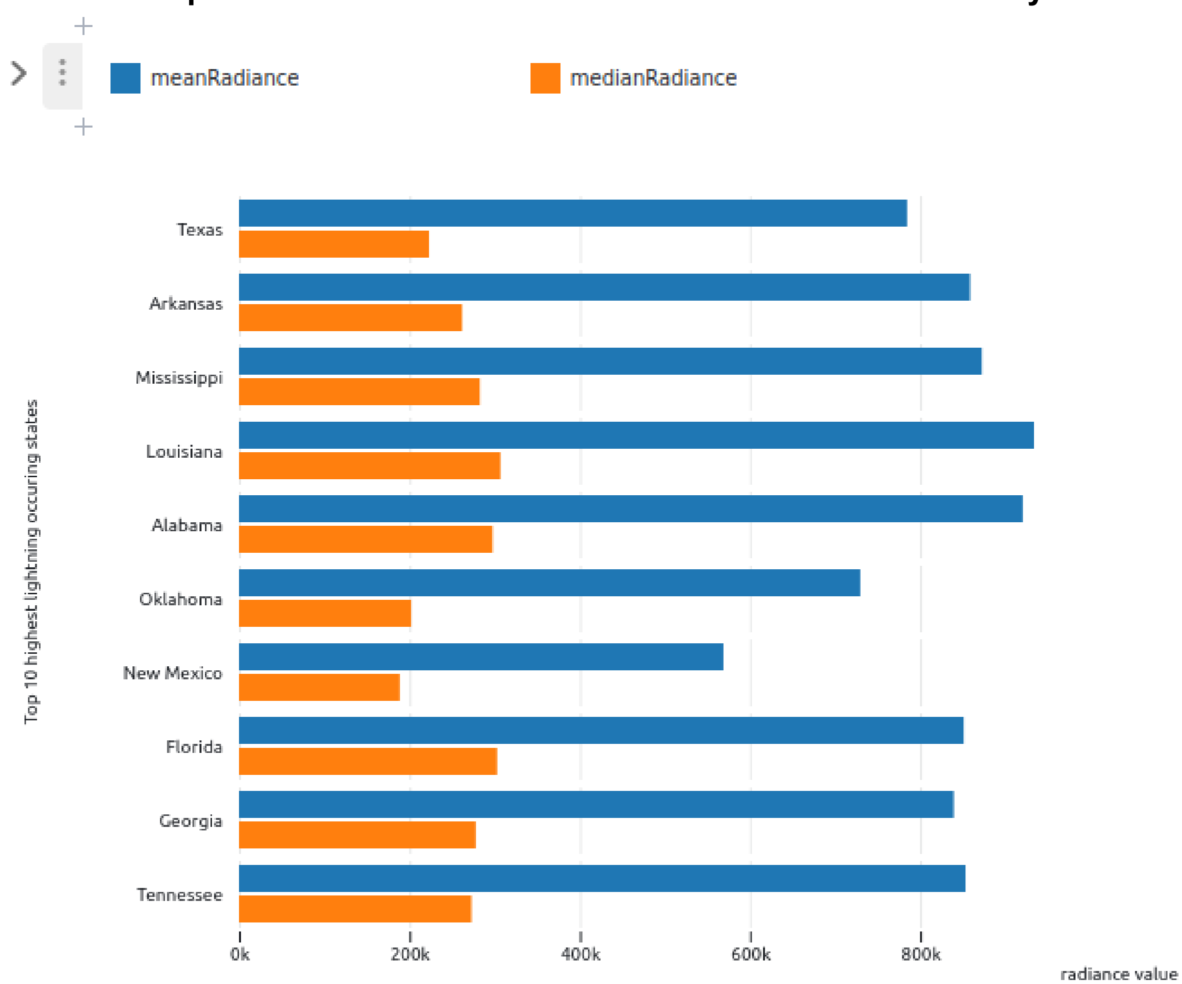

Figure 9: mean and median radiance values for the top ten highest lightning occurring states

Date Processing: First, two maps are generated for mean and median values for each of the states; then, these values are added as a column in the dataset. After that, wide-to-long technique is used to transform these mean and median columns for a single column named groupColums for each of the states. Also, the state list is sorted based on the amount of lightning occurrences and the top 10 states are displayed. We used the original unfiltered dataset for 1998 to 2015 to generate this plot.

After filtering the dataset, we created several visualizations based on the lighting event's radiance energy and total occurred events.

Choropleth (Lauren Dacre)

People located in regions where lightning is more prevalent need to be more proactive during severe weather events. I believe that the choropleth most clearly visualizes the distribution of lightning strikes across the United States. A choropleth map uses color and geographic area to convey information about a region, in this case, the United States. The choropleth legend, which is a range of colors each with a corresponding range of numeric values, gives a high-level indication of the number of strikes in a given state - the exact number is not conveyed. With the choropleth we can also see that data for northern US states is missing. From the available data, southern states, particularly Texas, Louisiana, Alabama, Mississippi, and Arkansas have the highest number of strikes. States that are slightly further north, including Nevada, Utah, Colorado and Kansas, tend to experience less lightning events.

Bubble Chart vs Choropleth (Lauren Dacre)

The bubble chart communicates similar information to the choropleth: the distribution of lightning strikes across the United States; however, the bubble plot does not give a numeric indicator of the number of strikes, instead the number of strikes is mapped to the area of each circle (bubble) and each state for which data is available has a bubble. I believe the choropleth conveys the information more clearly, but the bubble plot was included for comparison and some viewers may find it easier to see the data points represented by areas rather than color.

Choropleth: Strikes by State (1998-2015)

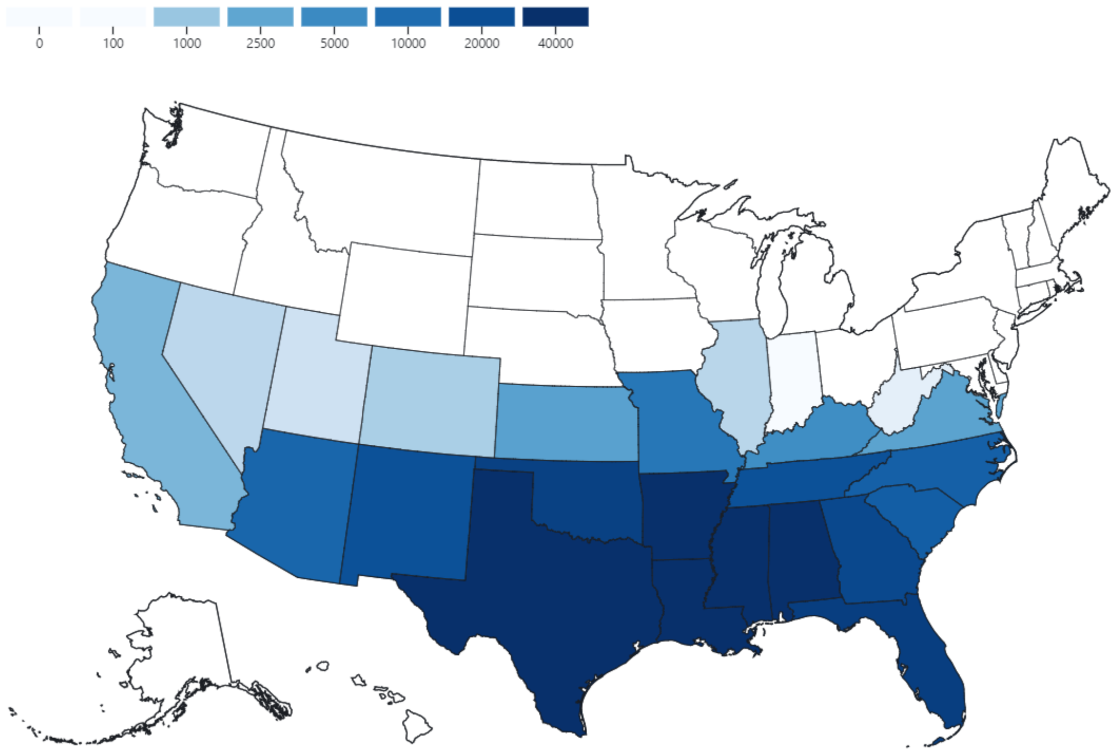

Figure 10: Choropleth illustrating lightning strike data for the United States

Bubble Chart: Strikes by State (1998-2015)

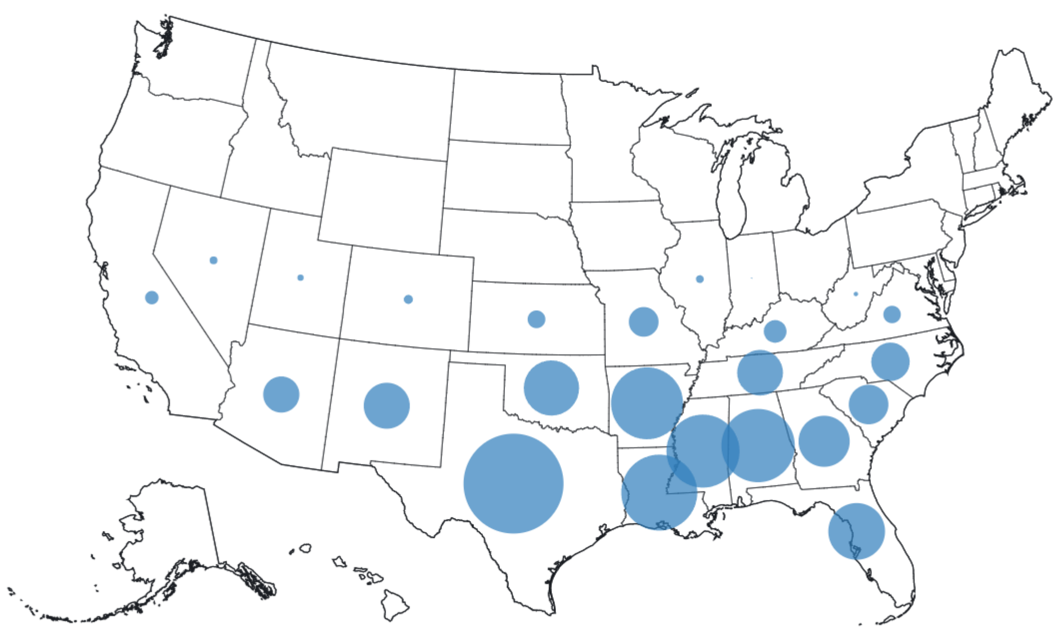

Figure 11: Bubble chart illustrating lighting strike data for the United States

Ranked Bar Chart (Lauren Dacre)

The ranked bar chart is a simple and concise visualization. It allows viewers to quickly scan the Y-axis and see which states experience the highest volume of lightning strikes as well as how other states compare. For readability purposes, this chart is limited to the 20 states with the most lightning strikes; it is important to be aware of the regions in which people are at the greatest risk of lightning threat.

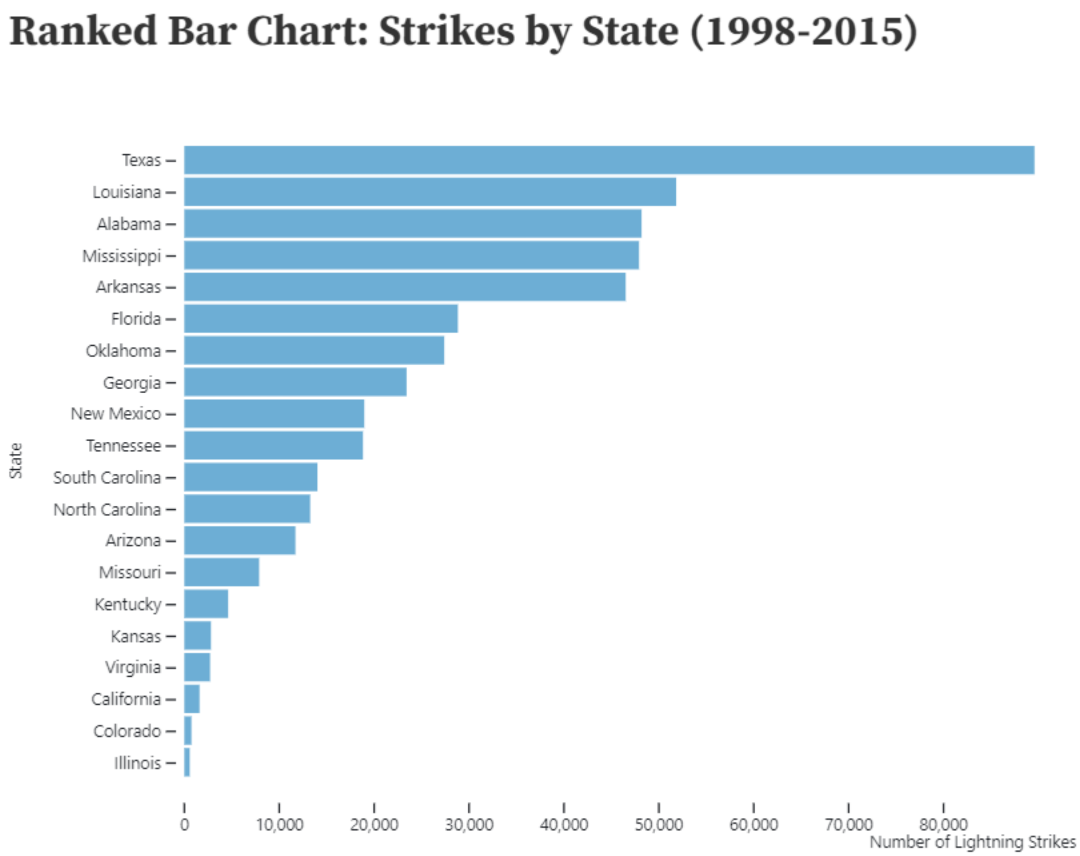

Figure 12: Bar chart which ranks states by the number of lightning strikes

Ranked Bar Chart: statewise mean and median Radiance value analysis

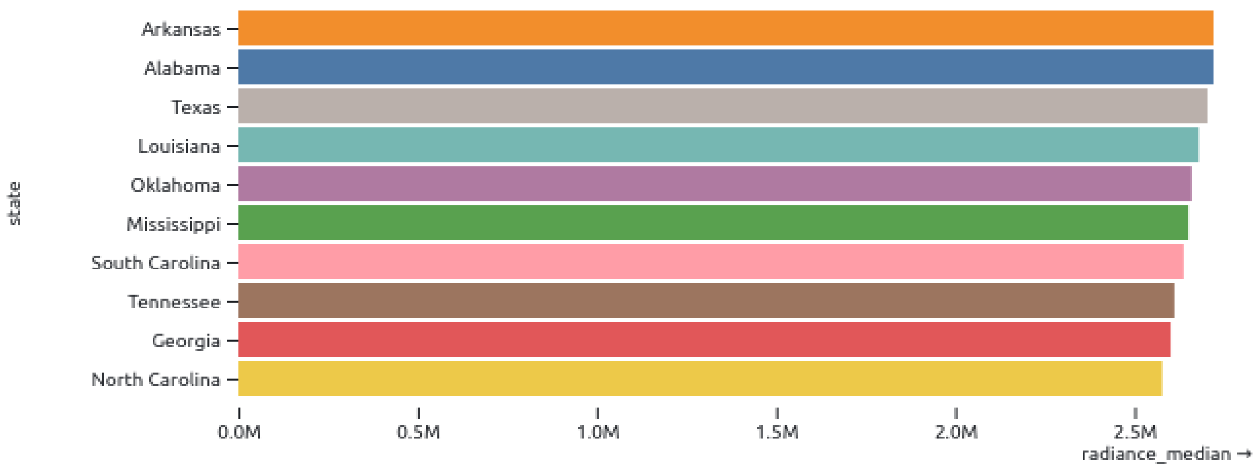

Figure 13: statewise mean and median Radiance value analysis using Ranked bar chart

Data processing: First we tried d3.group to summarize the data, but later found that d3.rollup and d3.rollsup provides better facility to perform this job. d3.rollsup returns an array instead of a map.

For creating the top state list, we took the mean radiance value from the upper quartile. Below, we create the same chart using bar chart (using svg). In this figure, the y axis is not started from zero, so it's highlighting the differences between the states. However, this does not fulfills Tufte's principle.

Bar Chart: statewise mean and median Radiance value analysis

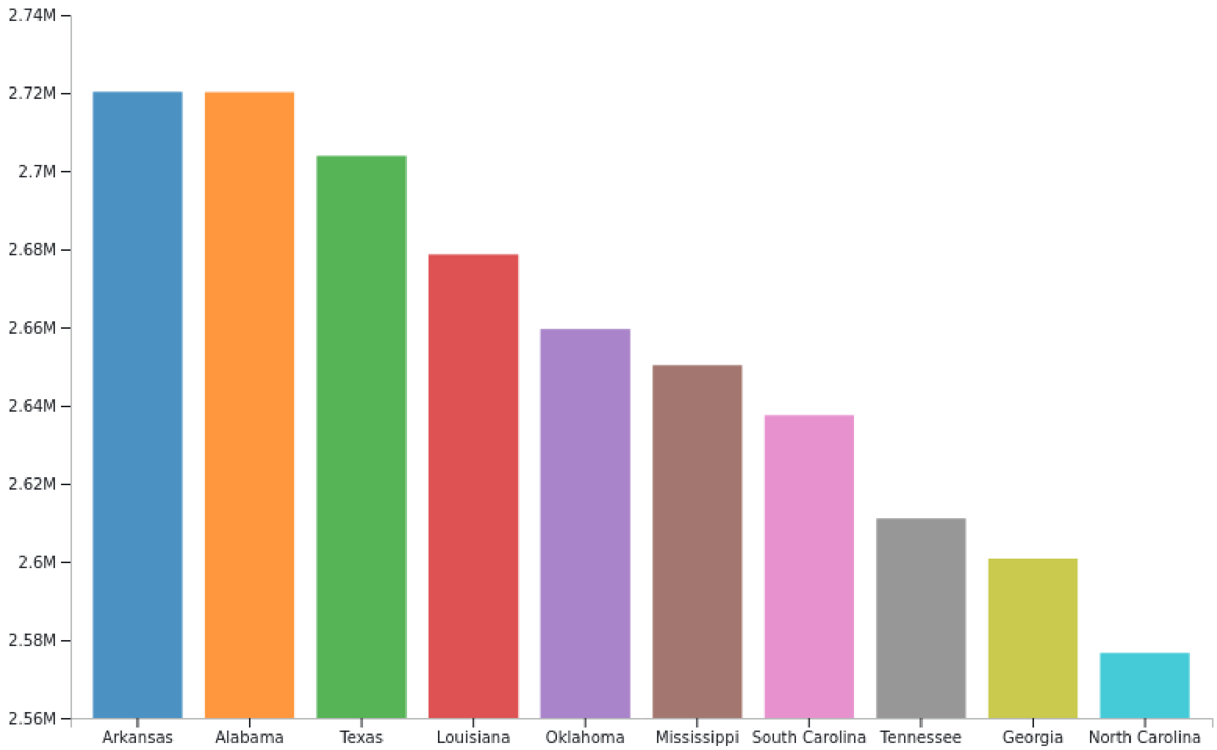

Figure 14: statewise mean and median Radiance value analysis using bar chart

From the above two figures, we can observe that Arkansas, Alabama and Texas are top three states, where the lightning occurs with the highest radiance energy. So, these places are more for lightning related incidents. Infact, if for the top ten states, the difference among each other is not significantly different.

However, if we consider the number of incidents occurred in 1998-2015, then Texas becomes the top state. The below pie chart shows the top ten states that has the highest amount of lightning events. Texas, Louisiana and Alabama is among the top 3. In terms of total lightning events, Florida places 6th in the list.

Pie Chart: statewise total number of Lightning events analysis

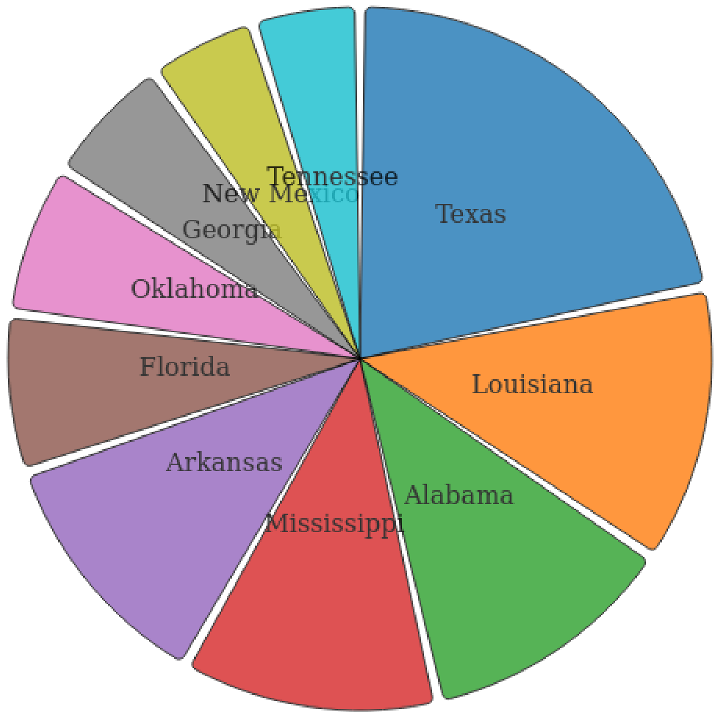

Figure 15: statewise total occurred radiances using Pie chart

***Data processing*: We used d3.rollups for the above chart. Also, using the d3.groupSort, we can also generate the same list of states (only state names)

Finally, we draw the multi line chart to visualize the over year trend for radiance values among different states. In this chart, we only showed the top five states for the year 1998 to 2015. We took the mean radiance value from the upper quartile to create this figure. From the figure, we can observe that there is change in 2000 to 2002 for the highest radiance values. Other than these two years, the range is similar for the rest of the years (with the exception of 2015).

Multi-line Chart: over the year trend analysis for radiance values for different states

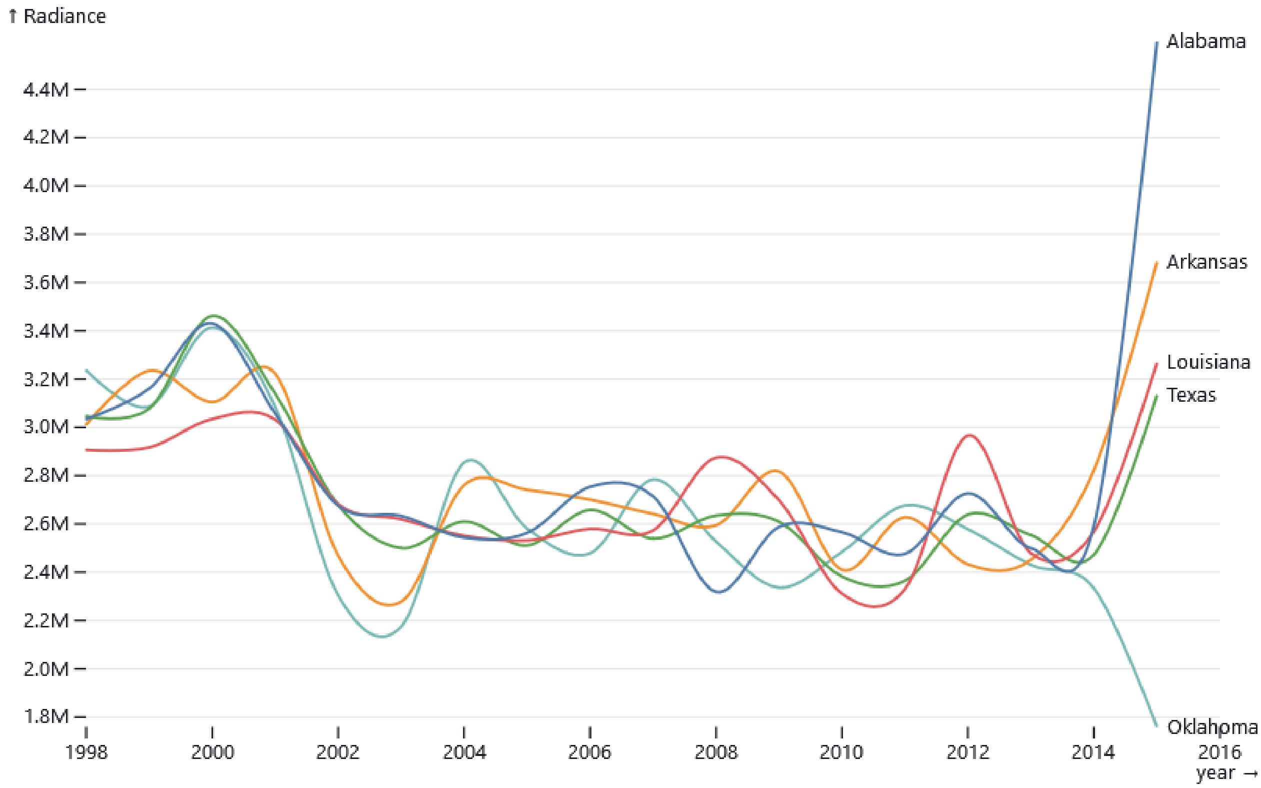

Figure 16: over the year trend analysis for radiance values

Data processing: Unlike the previous charts, the multiline chart requires three attributes. I tried to do that using the d3.rollups, but it cannot create more than one group. Using the d3.flatRollup, the data processing became much easier.

Limitations (Lauren Dacre)

We encountered a few limitations while working on this project. The two most notable limitations are data availability and Observable data storage and processing. The dataset from the TRMM (Tropical Rainfall Measuring Mission) satellite is missing data for some months and some states. The final year, 2015, only contains data for the first quarter. Related data points, such as fatalities, injuries, and property damage costs were difficult to obtain online. These data gaps prevent us from gaining a complete understanding of the prevalence, distribution, and effects of lightning strikes in the United States. Observable's data storage limitations also forced us to filter out data points (lightning strikes) below a certain radiance (kept values were in the upper quantile). Additionally, we initially planned to include data from other countries but this seems infeasible; Observable has a file upload size limit and the software also has difficulty processing high volumes of data. Despite these limitations, the project still does well in serving as a demonstration of how data visualization may be used to educate the public about the dangers of lightning strikes.

Conclusions (Lauren Dacre)

Many people may be unaware of the commonality of and risks associated with lightning strikes. Education and awareness can go a long way in helping people to understand lightning threats and take them seriously. We hope our visualizations help individuals in the United States understand the level of lightning risk in their area. A more comprehensive public lightning strike dataset and additional visualizations including all states may further this goal.

References

[1] Point in polygon https://en.wikipedia.org/wiki/Point_in_polygon

[2] TRMM space time domain search tool https://lightning.nsstc.nasa.gov/nlisib/lissearch.html

[3] Processed dataset https://drive.google.com/drive/folders/1oQKu8yPkHVhPoWw5zpaRkTvgx8BtFBk8

[4] Fatalities data https://www.weather.gov/media/hazstat/80years_2020.pdf