INFO 4310 HW #1 Static Visualization Design: 50 Years of Deadly Natural Disasters

Artina Maloki (am2495)

This visualization was designed using JavaScript and the D3 library.

This visualization shows the intensity of deadly natural disasters over a 50 year period in different country income groups. It reveals patterns in the death toll caused by for each natural disasters over the years, how it differs across income groups, and how it compares to other disasters.

From this visualization we can immediately see the large effect natural disaster have on lower income countries from the use of color and size. I think it's also interesting to see which disaster types are more prominent in the recent years compared to 50 years ago (for example epidemics and extreme temperature).

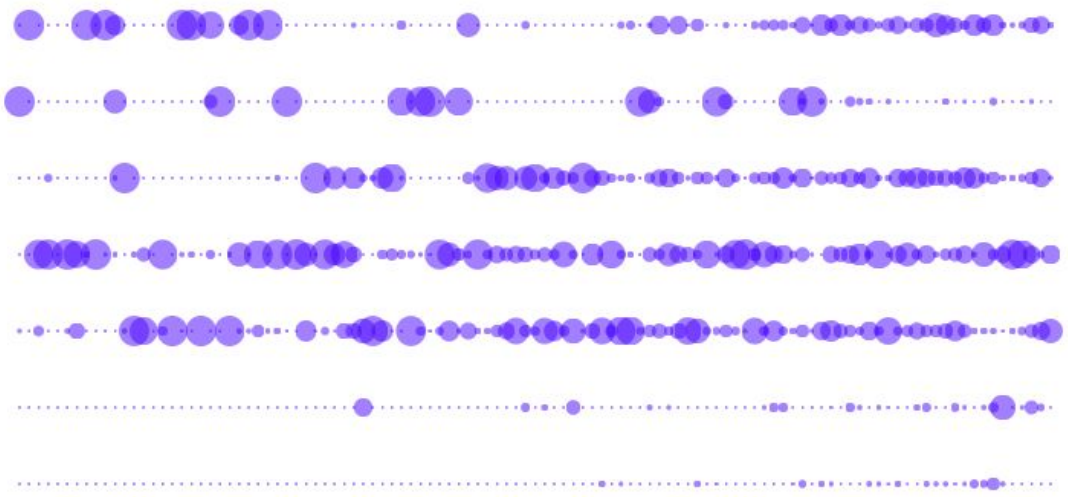
The goal of the visualization was to show some sort of pattern existing in the dataset over an interval of time. To show this I used several different elements to portray a pattern.
For each data point we have:

- Color represent income classification based on gross net income (White for High, Orange for Middle, and Pink for Low)
- Size for death toll (Smallest is 100, Largest is 5000+)
- X-Position corresponds to year of disaster
- Y-Position corresponds to type of disaster

Some preprocessing was done before designing this visualization. I categorized the countries into three different income groups by their 2014 Gross Net Income as classified by the UN. I included a new column in the dataset called "Income" that takes on three ordinal values, "Higher". "Middle", and "Lower". After performing some data exploration on the three categories, it was clear that there were some very large outliers that dramatically skewed visualizations. To account for this I capped the number of people killed in the "Killed" column to 5,000 and created a new column called "roundKilled" that took on quantitative values between 100 and 5,000. I think it was necessary to cap the large values and outliers *(instead of removing them)* because they are essential to the patterns observed. Removing them would be removing significant events. The goal of the visualization is to represent intensity, not exact numbers. Since I used scale / area to represent intensity you can immediately pick out the outliers as major events.

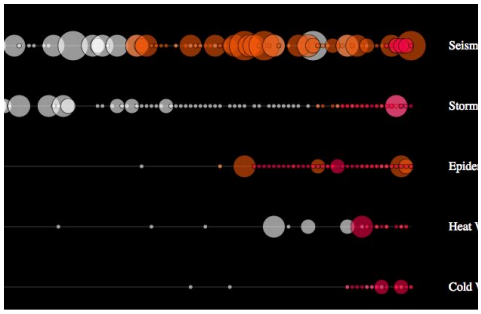
It intuitively made sense to let the x-axis of the visualization correspond to "Year" since people can easily interpret and read the progression of time from left to right. I chose to only include data from 1950 and aggregated the data by year into each income class. Using all of the data points made the visualization very cluttered and harder to interpret. Additionally the country income classifications in the early 1900's were drastically different than the one's today. I think this would confuse people. More importantly, the final condensed dataset still painted a distinct pattern.

The next question became, how do I show this trend and what elements should I use to represent my data points? I wanted to show intensity and not exact values to portray a pattern (if any). I thought a great way to show changing rates over time is by using circles and have the area correspond to the intensity of the disaster (number of people killed).



After inspection there was clearly some underlying patterns in the dataset! This layout not only lets you compare how a disaster's rates change over the years but also how disasters compare to each other.

I wanted to dig a little bit deeper and add new dimension, income class. I wanted to keep a similar style as the first mockup and also clearly make a distinction between different income classes. I incorporated color to distinguish between the classes and shifted the y-axis for each type so the circles would not overlap and hide data. I recognize the tradeoff here. By representing the data points as circles and changing their areas accordingly, it is hard to determine their exact value. However it is an effective way to show change and patterns which is the goal of the visualization.



Each disaster type has an associated horizontal section. I used negative space and padding to separate these sections so they are distinct, but not too far apart, so the reader's eyes can travel horizontally along the x-axis but also make comparisons vertically across different types. I used a horizontal thin line at the top and bottom for the x-axis and to aid the reader.

You'll also notice that all other elements (except for the main title) are in grey. This is because they are secondary elements to the visualizations and live in the backdrop. This allows the data points to really stand out. A lot of time was spent on determining the style of the data points. Because there was no way to get around overlapping it was essential to lower the opacity levels so no data was hidden. The range for the point's radius was determined as [smallest we can possibly see, largest that can fit but not completely overlap another class]. If all of the circles were too small, it was impossible to see a trend. Same if they were too big. It was essential to have a large enough range that captured the differences and intensities present in the dataset.

Finally I added a corresponding color coded map that shows income classification for the countries represented in the dataset. The map on the left hand side not only serves as a “legend” for the visualization but it also provides the viewer more context and helps familiarize the viewer with the different types.