Project A: Visualization of Floods Data

INFO-633 Information Visualization
Professor Chaomei Chen
Group 9: Aviv Farag, Kelly Wetherbee, and Elizabeth Worsley
10/09/2022

Abstract – This paper examines worldwide floodings from January 1985 to October 2021. Tableau Public was utilized in order to visualize both the Main Cause and Severity Level of flood events, and Main Cause and Number of Deaths. By doing so, the aim of this study is to find insights that will help countries and communities prepare for future flood events.

Keywords – Floods, Geographic Information, Tableau Public, Data Visualization

Introduction

A flood is a natural hazard in which water covers land that is normally dry. Severe flooding can damage structures, such as houses, both while forming and receding. Floods can form in several ways. Heavy rain can increase water level in a river causing a flood in an area nearby. Another way flooding occurs is snow that melts too quickly, and thus overwhelms the local infrastructures. These two causes take time to develop, therefore residents can make preparations such as retention walls or evacuate in a timely manner. However, floods can also develop quickly, for example when a dam is damaged from causes such as tsunami, earthquake; or from storms, such as cyclones and hurricanes, that are accompanied with heavy rain. In those cases, there are little to no warnings, which can be very deadly. In order to better prepare for these disastrous events, governments and communities should know if (Q1) the main cause of floods is correlated to their severity and (Q2) which main cause is the most deadly (Boudreau, 2022).

Data Exhibition

The visualizations in this paper rely on a dataset that was created and published by [Dartmouth Flood Observatory (DFO)]() capturing worldwide flooding from January 1985 to October 2021 (Appendices 1 - 3). Each row represents a flood event and contains the following information:

- **Identifying Data -** Location, timeframe, and effect. There are 11 features.

  1. Glide number (Globally unique disaster ID)
  2. Country
  3. Other Country (Regions)
  4. Longitude
  5. Latitude
  6. Area (affected)
  7. Began
  8. Ended
  9. Validation (source)
  10. Dead
  11. Displaced

- **Main Cause -** Type of the event. There are 17 types:
  1. Avalanche Breach
  2. Dam Break
  3. Heavy Rain
  4. High Tides
  5. Hurricane
  6. Ice Jam
  7. Landslide
  8. Levee Failure
  9. Rain and Snowmelt
  10. Monsoonal Rain
  11. Storm Surge
  12. Tidal Surge
  13. Torrential Rain
  14. Tropical Cyclone
  15. Tropical Storm
  16. Tsunami
  17. Typhoon

- **Severity -** A discrete class of flood events with values:
  1. "Class 1.0: large flood events: significant damage to structures or agriculture; fatalities; and/or 1-2 decades-long reported interval since the last similar event.
  2. Class 1.5: very large events: greater than 20 yr but less than 100 year recurrence interval, and/or a local recurrence interval of at 10-20 yr.
  3. Class 2.0: Extreme events: with an estimated recurrence interval greater than 100 years (Archive Notes (DFO))."

Audience

The dataset report, these data visualizations, and this study over all may be of use for a wide audience. Government agencies might be interested to read this report to prepare for evacuations and emergency services. City planners may find this report useful in aligning on infrastructure needs. News outlets may find this useful in the dissemination of pertinent weather related data and death tolls, or possibly to decipher expected mortality rates for an impending flood. Historians, statisticians, forecasters, etc. may also find this data interpretation useful.

Tools

Three tools were utilized in this study: Microsoft Excel, Python and Tableau Public. There were several formats available for the dataset, and the Excel format .XLS was chosen for simpler integration with Tableau. The dataset spreadsheet contained noise, and Python was used to automate the cleaning process. Finally, Tableau Public was not only used to create the visualizations in this study, but to accurately decipher the data meaning.

Data Cleaning

A few inconsistency issues were found when reviewing the data, such as names assigned to storms. For example, a storm name such as Hurricane Ian that recently hit the United States' lower half of the Eastern Seaboard. This issue presents redundancy in the data due to the unique names assigned to each storm.

Another issue encountered was related to case sensitivity as well as spelling. For example, "Heavy Rain", "Heavy RAin", "Heavy rain" and "Heavy ran". Those are all the same Main Cause, but since they are written differently, they also increase redundancy. Spelling issues were also presented in the Country feature, so it was necessary to clean it before uploading to Tableau Public.

Q1: Is there a correlation between the flood's main cause and severity level?

Analysis

Based on the data used, there is no relation between the Main Cause and Severity Level of . The six most occurring Main Causes are Heavy Rain, Torrential Rain, Tropical Storm, Monsoonal Rain, Rain and Snowmelt, and Dam Break, with drastically more occurrences of Heavy Rain than the others. The rest are negligible in occurrences. Most of the Main Causes have all Severity Levels in dwindling order from least severe (1.0) to most severe (2.0). Therefore, there are no Main Causes that are specific to a certain Severity Level, outside of all Main Causes have most occurrences in Severity Level 1.0.

Results

Below we broke the question down into two visualizations. First, utilizing the Law of Common Fate, the Main Causes were grouped by Severity Level (Chen, 2022). In other words, for each value of severity there are separate bars each of which is representing one main cause. Since they are all bounded by a specific value of severity, they share a region.

The treemap (Fig. 1.1) displays the number of occurrences for each Main Cause which is separated by color that represents Severity Level. Heavy Rain has roughly 3.5 times as many occurrences as the next closest Main Cause, and that more than half of all storms have a Severity Level of 1.0. Additionally, Heavy Rain events with a Severity Level 1.0 account for about 40% of all floods in this dataset for all Main Causes and Severity Levels because of how large the first leaf is compared to the overall treemap. Each Main Cause has the most occurrences of Severity Level 1.0.
(see Fig. 1.1 on next page)

Visualization 1.1: Treemap of Main Causes sorted by the number of occurrences (size) and Severity (color)

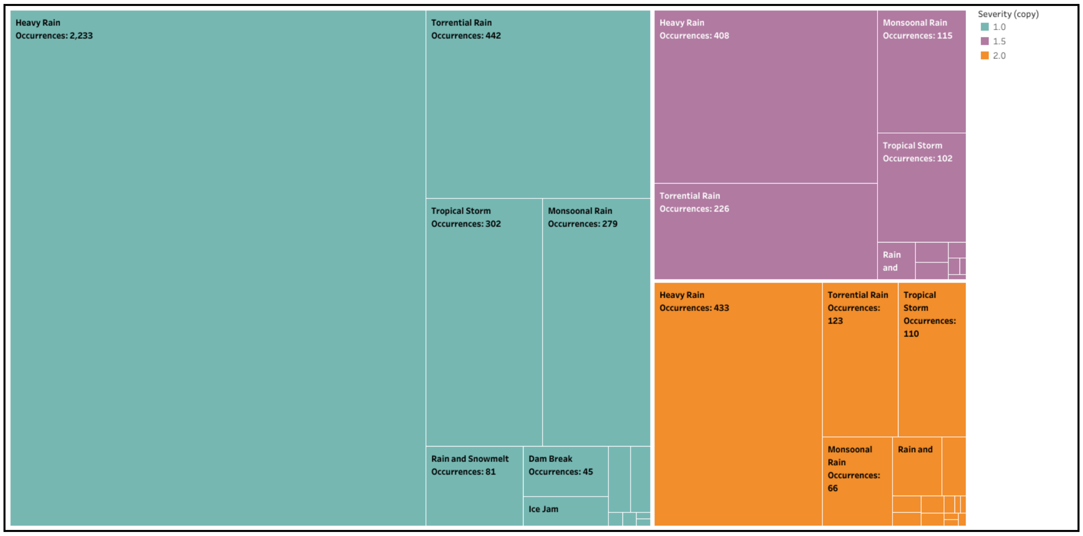

Figure 1.1 Treemap of occurrences separated by severity and main cause, displaying that most main causes have had occurrences of all Severity Levels.

The second visualization (Fig. 1.2) tells a similar story. Utilizing the Law of Similarity, the divided Severity Levels easily displays that each Main Cause had occured in each Severity Level (Chen, 2022). Again, the most frequent floods are Heavy Rain, more specifically, Severity Level 1.0 Heavy Rain events. Thus, there is no correlation between Main Cause and Severity Level because all Main Causes had occurrences of all Severity Levels, 1.0, 1.5, and 2.0. Since there was no correlation between Main Cause and Severity Level, the study goes to the next question that dives into whether there is a correlation in Main Cause and Number of Deaths.

Visualization 1.2: Main Causes separated by their severity, and represented in percentage

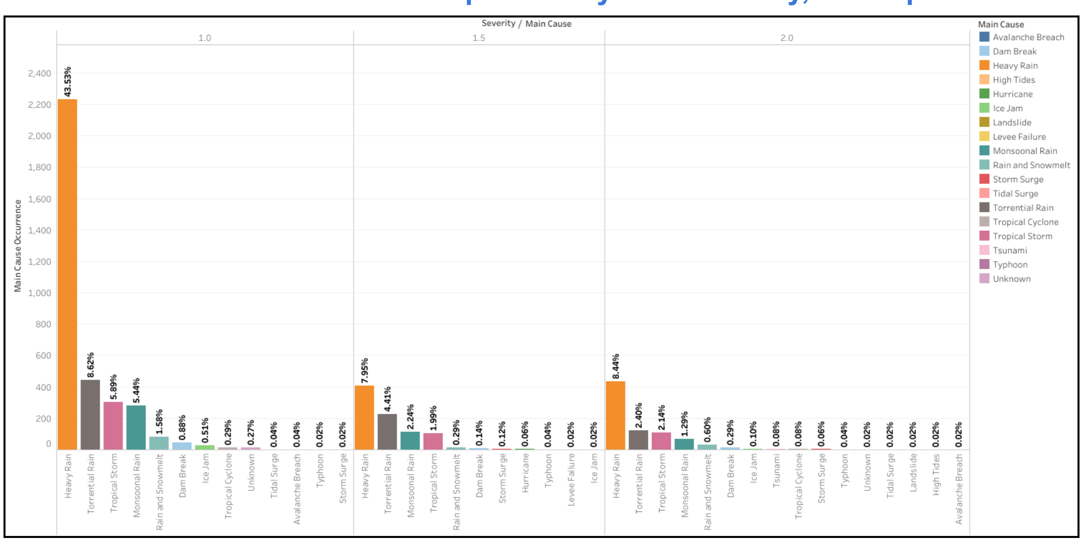

Figure 1.2 Bar chart of occurrences separated by severity and main cause, and their representative percentages to all main causes captured.

Q2: Is there a correlation between the flood's main cause and the number of dead?

Analysis

Unlike the first question, the dataset used provides clarity that there is a correlation between Main Cause and Number of Dead. The seven most deadly Main Causes are Tropical Storm, Tidal Surge, Heavy Rain, Monsoonal Rain, Torrential Rain, Tsunami, and Dam Break. The rest are negligible in occurrences. The most Dead per Main Cause Occurrence is from Tropical Storms at 53,367 deaths per occurrence. The next highest at 2,818 Dead per Occurrence is Tsunamis, making the gap between Tropical Storms and Tsunamis 50,549 Deaths. In other words, Tropical Storms are at least 18.9 times higher in Deaths than others. The other six most deadly Main Causes, though not as large in their death rates, are still significant in their Dead per Occurrence. Therefore, these seven Main Causes have the most deaths per occurrence.

Results

The visualizations for this question follow the seven tasks that were suggested in the paper, "The Eyes Have It." First, Main Causes were reviewed for anomalies. One such anomaly was there are several Main Causes with little to no occurrences, and therefore are negligible. Hence, the zoom-in process was used on values that have more occurrences, filtering out the negligible values, and creating the bubble chart displayed in Figure 2.1 (Shneiderman, 1996).

Visualization 2.1: Bubble chart of Main Cause, the Number of Occurrences, and Number of Dead

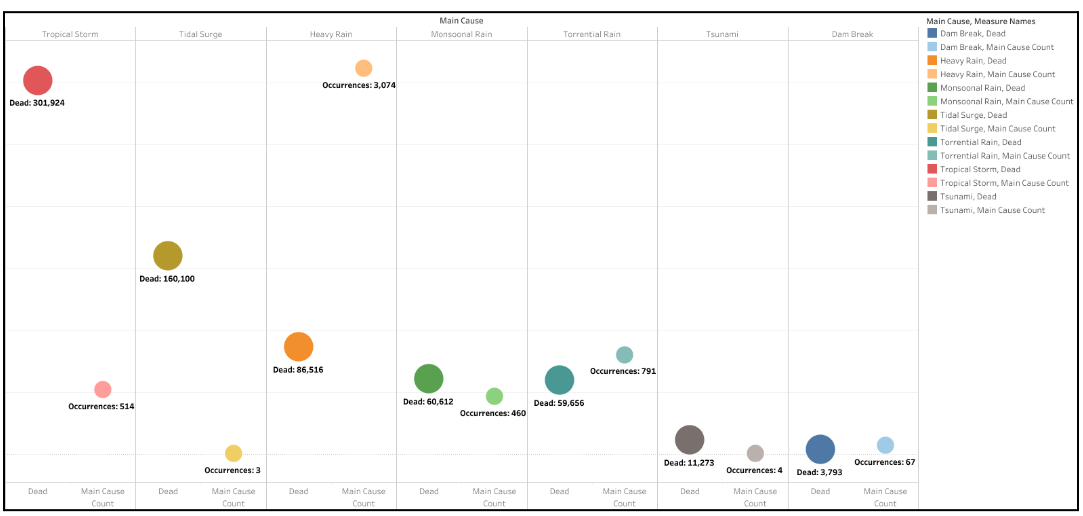

Figure 2.1 Main Cause, their number of occurrences (smaller circle) and their number of dead (bigger circle)

Again, the visualization uses the Law of Common Region to bound the circles together in a rectangle that represent a single Main Cause, as well as the Law of Similarity to use circle size for representation of the amount of Dead as larger circles, and the smaller circles as representation of the number of Main Cause Occurrences (Chen, 2022).

Then, we continue to get "details-on-demand" by simply calculating the number of dead per occurrence for each Main Cause. By doing so, we created the Pie Chart (Fig. 2.2A) that shows which Main Causes are more deadly than others, thus we find relationships among items and order them. This also allows us to extract information such as grouping the three deadliest Main Causes (Shneiderman, 1996).

While creating the Pie Chart there were scaling issues, i.e., Tidal Surge with 53,367 is much greater than other values, and therefore covered most of the Pie Chart's area (Fig. 2.2A). To overcome this issue, rescaling the numbers by calculating their logarithm was used which is a common practice in engineering, and used that calculation for drawing the angle of each Main Cause (Fig. 2.2B).

For example:

$$Tidal\ Surge\ \ -->\ \ log_{10}(53,367)\ =\ 4.727$$

$$Monsoonal\ Rain\ \ -->\ \ log_{10}(132)\ =\ 2.12$$

Eq. 1 Rescaling values using logarithm

This new scale allows all Main Causes to be apparent in the logarithmically scaled Pie Chart. As can be seen in Eq. 1 above, the calculated numbers are within a small range, and therefore can visualize the data without losing information.

Visualization 2.2A: Pie chart of Deaths per Main Cause Occurrence in Ratio

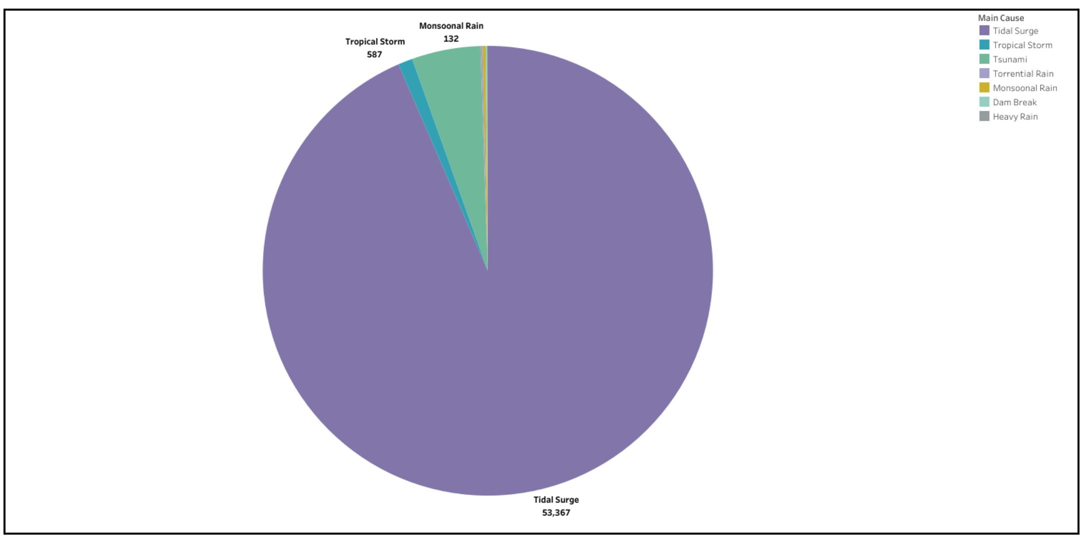

Figure 2.2A Pie chart of Deaths per Main Cause Occurrence (In ratio scale)

Visualization 2.2B: Pie chart of Deaths per Main Occurrence after rescaling

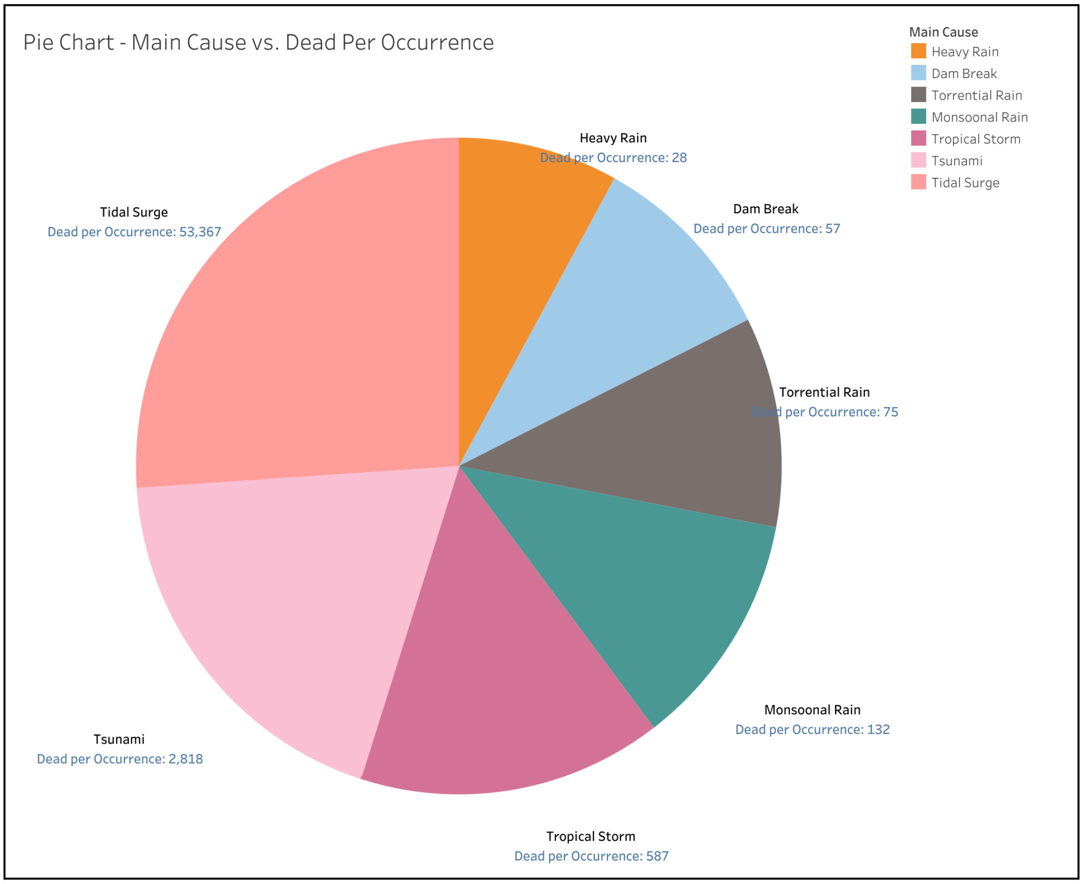

Figure 2.2B Pie chart of Deaths per Main Cause Occurrence (Logarithmic scale)

Discussion

In this study, we examined historic data that is related to floodings, exploring any correlation between Main Cause, Dead and Severity Level in order to help the audience be better prepared for future occurrences. The two visualizations were created to find relationships between Main Cause and Severity Level, but found none. Therefore, one can't rely on the Main Cause to predict potential damage. In the second visualization (Fig. 1.2), the Law of Common Region was used to group them together under a single Severity Level, and thus show the same data in a different way that is easier to see that each Main Cause has occurrences in each Severity Level.

Further study could examine the relationship between the Main Cause and the Geographic Location (Longitude and Latitude), and possibly link between a country's wealth and the damage done by the storm. To put it simply, the aim is to show whether third-world countries are less prepared, and thus the Number of Dead is higher. If so, what type of support could help them in future events?

Assessment of the total damage is not precise, and thus limits this study. For instance, the features "Displaced" and "Area" aim to assess the damage, but they are just an estimation, and therefore cannot be relied upon. The former is an estimation of the number of people affected by the flood event, and the latter is the overall area affected by the flood event. Those features could potentially reveal great insights, if they were measured more precisely.

Another limitation faced was the scaling issue while drawing the Pie Chart. One value was much greater than others, and therefore took over most of the chart. In that case, information loss occurred because only one value was clearly visible, while others were hidden. This issue was resolved by calculating the logarithm of each value, thus rescaling the data to a small range of values, and displaying both the Pie Chart at Ratio and the Pie Chart Rescaled.

Conclusion

The worldwide floodings dataset from January 1985 to October 2021 display the devastating effects these events have, from displacement and damage to death. Tableau Public helped visualize the data in a specific way in order to help answer two questions: (1) Is there a correlation between the main cause of a flood and the severity of the flood? (2) Is there a correlation between the main cause and number of deaths per occurrence? While this particular dataset shows no correlation between main cause and severity, it does show a significant correlation between main cause and death. Three other learnings that stemmed from these two questions and their relevant visualizations are: (1) Most of the Main Causes have all severity levels in dwindling order from least severe (1.0) to most severe (2.0). (2) Only six Main Causes have most of the overall occurrences, while the other eleven have significantly less to no occurrences. (3) Only seven Main Causes have the most amount of dead per occurrence, while the other ten have significantly less to no dead. These findings can help governments plan for flood disasters, help people prepare their homes and/or evacuate when necessary, and help the news to accurately predict and inform their consumers about impending flood events.

References

Boudreau, D., McDaniel, M., Sprout, E., & Turgeon, A. (2022, May 20). *Flood*. National Geographic Society. Retrieved September 2022, from https://education.nationalgeographic.org/resource/flood/.

Chen, C. (2022, September 27). Week 2C: Natural Groupings and the Simplicity of the Whole [PowerPoint Slides]. Philadelphia, PA; Drexel University.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–343. https://doi.org/10.1109/vl.1996.545307.

Archive Notes. Dartmouth Flood Observatory (DFO), (n.d.). Retrieved September 2022, from https://floodobservatory.colorado.edu/Archives/ArchiveNotes.html.

Appendix

Appendix 1: Master List: FlooddataMasterListrev.xlsx

Appendix 2: Raw List: Fixed_FloodArchive - Original.xlsx

Appendix 3: Cleaned List: Fixed_FloodArchive - Updated.xlsx

Appendix 4: Dashboard: Tableau Public Flood Events Dashboard:

https://public.tableau.com/views/FloodArchiveFinal/FloodEventsDashboard?:language=en-US&:display_count=n&:origin=viz_share_link.