The Stellar Chronicles of History

Data Visualization Project

Maria Tager
Killian Hinard
Elliot Jacquet-Francillon

May 2024

1 Introduction

As we orbit through the cosmos of data, our team has harnessed the power of visualization to craft an interactive atlas of global events that light up the dark void of history with the brilliance of insight. This journey, anchored in the boundless potential of D3.js and other cutting-edge technologies, allows us to travel through time and space, unveiling the dynamic tapestry of humanity's footprint on Earth. From the sparks of political upheavals to the echoes of social movements, our platform invites you to explore how these events ripple through the annals of time, influencing populations and reshaping our world. With every click, rotate, and zoom, discover a universe where data points transform into stories, and numbers into narratives.

Our mission was ambitious: to translate complex historical data into a galaxy of interactive experiences. The path wasn't linear—much like history itself. Through challenges faced and milestones reached, our process book aims to be a transparent chronicle of our voyage from conception to realization. Prepare for a thrilling excursion through our digital cosmos, where history is not just learned but experienced. Join us as we unfold the layers of the past, presenting a universe of events through a lens that is as educational as it is enthralling.

Dive into our process book to witness the evolution of an idea into an immersive exploration tool, detailed with the craftsmanship of our team's collaborative spirit. This is not just a tale of data visualization; it's a story of passion, perseverance, and the power of visual storytelling. **Let the journey begin!**

2 Data processing

In the early stages of our data processing, we encountered significant challenges due to the nature of our dataset, which was generated by a GPT-based AI model. This introduced a layer of complexity with its non-standardized data format, complicating the extraction of pertinent information. In this part, we present the solutions we found to overcome the different hurdles we found in our dataset throughout the process.

2.1 Country, place name and coordinates

A major challenge was that we did not have the coordinates for the locations of the events. The only geographic information we have are Place name and Country.

Additionally, the names of the places and countries in our dataset were not standardized and included some unusable information.

We used Python to clean up these data points, standardizing the names of countries and places as much as possible manually. After cleaning the data in these columns, we used Geopy for geocoding to obtain latitude and longitude information. We chose the Nominatim API for geocoding because it is free and integrates directly with Geopy.

2.2 GeoJson

For one of our visualizations, we required the country boundaries from around the world in a standardized GeoJson format, along with data on the number of events per country. To obtain this, we utilized geographic data from Natural Earth, which is both free and open-source.

To compile a GeoJson that includes event counts, we used two main tools: Geopandas and Pandas. With Geopandas, we manipulated the GeoJson objects to accurately reflect each country's borders. Meanwhile, Pandas allowed us to process our dataset.

This approach enabled us to merge our event data with the geographic outlines provided by the GeoJson files.

2.3 Classifying of the events by type

The data had a column called "TypeOfEvents" with more than 400 different kinds of entries. Our goal was to make sense of these entries by grouping them into broader categories, for the bubble graph and the mosaic bubble graph visualizations 4 . We decided to create a new column called "Broad Category" that would include 12 general types: Political Events, Military and Conflict, Economic and Infrastructure Development, Social and Cultural Events, Technological and Scientific Advancements, Environmental and Health, Legal and Judicial Changes, International Relations and Diplomacy, Crisis and Emergency Response, Historical and Monumental, and Other.

We did this mapping by hand using Python and the Pandas library. We also used ChatGPT-4 to help speed up the process by reading through the existing event types and grouping them into our new categories. This step helped us organize the data better and made it easier to work with for our visualizations.

2.4 Multiple type characterization of events

We further refined our classification methodology by utilizing advanced natural language processing techniques. Specifically, we represented each event using a BERT sentence embedding that captured the event's name, type, and impact. To accurately represent each broad category, we computed the average embedding of all events within that category. We then created a vector of cosine similarities between the embeddings of individual events and the embeddings of each category. For each event, we selected the smallest set of categories that collectively accounted for at least 35% of the total similarity score. This approach allowed for a more nuanced and accurate characterization of events. For example, the event "Fall of Berlin Wall," which had an impact statement of "Symbolized the end of the Cold War and led to German reunification; significant US involvement," was initially categorized by our first mapping method 2.3 as a "Technological and Scientific Advancement." However, our refined processing correctly reclassified it under the more appropriate categories of "Social and Cultural Events" and "Military and Conflict." This enhanced classification method also enabled us to eliminate the "Other" category, ensuring that each event was assigned to at least one relevant category.

2.5 The Continent problem

The dataset lacked a "Continent" column which was needed for the bubble graph and bubble mosaic visualizations 4 so we added it by mapping the "Country" column to it. This was also done using Python and the Pandas library, manually and with the help of ChatGPT-4 to speed up the process.

3 Exploring the globe

When we first looked at our dataset, we noticed it included locations for each event. This inspired us to create a visualization using a 3D globe. This method makes it easy and exciting for users to look through the data and understand the different events in our dataset.

To enhance user engagement, we implemented two distinct types of 3D globe visualizations. The first features interactive points on the globe, allowing users to click on each point to uncover more details about specific events. The second is an interactive heatmap, designed to illustrate the overall distribution of events across the globe in our dataset.

3.1 Historical events around the globe through the centuries

As previously mentioned, our initial visualization features a 3D globe where users can interact with event points, each colored according to their outcomes. Users have the ability to click on any point to get detailed information about that event. They can also filter the events by the outcomes they are interested in, and choose which century they want to display. This method of selecting events by century helps avoid cluttering the visualization with too many events at once, which would make it hard to understand.

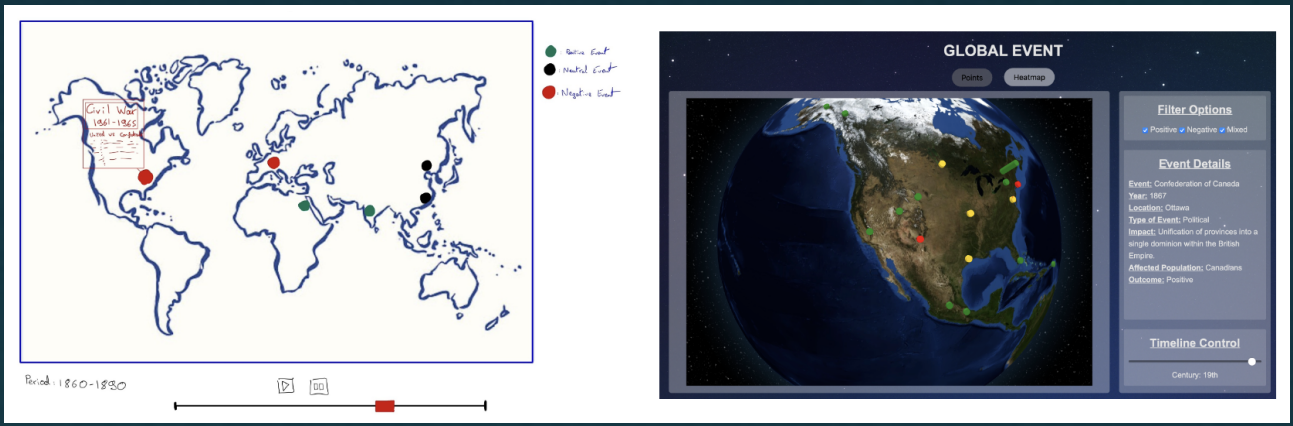

Figure 1: Milestone 2 idea (left) vs Milestone 3 implementation (right)

The purpose of this visualization is to give users a quick and clear understanding of the data, allowing them to immediately see specific events.

For this project, we used the open-source JavaScript library Globe.GL because it's easy to use and it was also suitable for creating the heatmap visualization.

The design has evolved from our initial concept outlined in milestone 2, as illustrated below. We introduced a side panel that displays event details and plotting options. This adjustment was made after realizing that providing this information in a side panel offers clearer visibility for the user, compared to overlaying the information directly on the globe, which was our original plan.

3.2 Interactive Heatmap of event distribution

Another visualization we made is a heatmap to visualize the distribution of events again using a 3D globe from Globe.GL. This feature allows users to see a color-coded representation of data density across different regions. By interacting with the heatmap, users can hover on any country to reveal its name and the number of event we have in our dataset for this country.

The primary goal of this heatmap is to provide a comprehensive and immediate visual summary of the data distribution, enabling users to quickly grasp patterns and anomalies at a glance.

For this aspect of our visualization, we continued to utilize the open-source JavaScript library Globe.GL due to its straightforward implementation and versatility. Not only is it effective for point-based data visualizations, but it also excels in rendering complex heatmaps on a global scale. This also allowed us to use the same globe for the two visualization preventing two many loading of this 3D object.

The design has undergone significant refinements since our early planning stages described in milestone 2. Due to the sparse nature of our data and the necessity for clearer interaction, we revised our original approach. Instead of plotting the heatmap based on the precise locations of events, we opted to aggregate the data by country. Additionally, we implemented a hover feature that allows users to instantly see the number of events and the name of a country simply by moving their cursor over the area. This enhancement makes the visualization more user-friendly and informative.

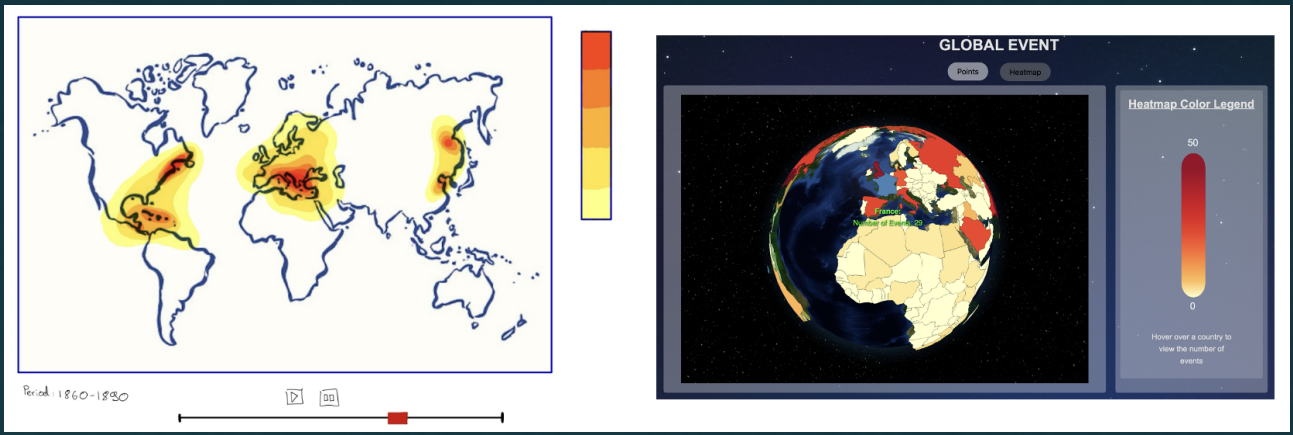

Figure 2: Milestone 2 idea (left) vs Milestone 3 implementation (right)

4 Type of events analysis

In our project, we undertook a detailed analysis of the types of events using two main visualizations: the Bubble Graph and the Mosaic Bubble. These tools are designed to provide an academic and comprehensive understanding of the distribution and characteristics of global events over time and across different regions.

4.1 Continent Distribution Bubble Graph

The Bubble Graph serves as an interactive tool for exploring the distribution of historical events by continent. Users can select a continent and the graph dynamically displays bubbles representing different categories of events such as military conflicts or cultural festivals. Each bubble's size and color vary according to the frequency and type of event, providing an intuitive visual representation of the data. This visualization was initially conceptualized in our Milestone 2 sketch and has been successfully implemented as planned, capturing the intended functionality and interactive elements, as shown below.

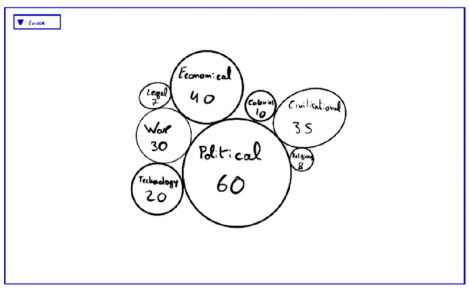

Figure 5: Bubble graph for event types

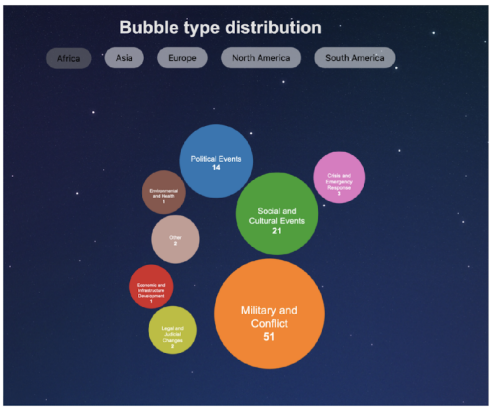

Figure 3: Initial sketch (left) and resulting visualization (right) of the bubble continent graph

Following the initial concept, we employed a series of libraries and technologies to bring the Bubble Graph to life. We used D3.js to manage data loading, filtering, and visualization. Data is loaded from the data CSV file and filtered based on the user's continent selection. We then use D3 to aggregate the data by event type, which influences both the size and the color of the bubbles on the graph.

In terms of interactivity, the graph utilizes D3.js's simulation functions to animate and organize the bubbles. This includes forces that prevent bubble overlap and ensure that each bubble is positioned clearly within the visualization space. Additionally, the colors are mapped to specific event types using a predefined color palette, enhancing the visual distinction between categories. Textual overlays within each bubble provide further details, making the visualization not only informative but also engaging. This approach has allowed us to translate our initial sketches into a fully functional and interactive visualization that effectively communicates complex data through an accessible and visually appealing format.

4.2 Mosaic Bubble Graph

The Mosaic Bubble Graph is implemented to enhance our understanding of the impact of global events. This visualization is structured hierarchically and unfolds across three layers to deepen the user's interaction and analysis of the data.

At the first layer, the graph displays bubbles categorized into three outcomes: Positive, Negative, and Mixed, representing the overall impact of events on global populations. Users can interact with this layer by clicking on any of the outcome bubbles, which then transitions them to the second layer of the visualization. This second layer breaks down the events into continental distributions, allowing users to see how different regions contribute to the specified outcomes of global events.

Once a user selects a continent from the second layer, the visualization dives deeper into the third layer, which details the distribution of event categories within the chosen continent, tailored to the outcome selected in the first layer. This allows for a nuanced exploration of how different types of events are distributed across continents, further specifying the nature of their outcomes.

This visualization was added during the development process and was not part of the original Milestone 2 sketches. It employs advanced D3.js techniques such as hierarchical data structuring, dynamic transitions, and interactive zoom functionalities to provide a multi-layered exploration of data. The technical implementation involves creating a packed bubble layout using D3's packing algorithms, adjusted for interactive exploration. This includes setting appropriate color scales, padding, and transitions for zooming into different layers of the hierarchy. The SVG elements are dynamically generated and manipulated based on the user's interactions, providing an interactive and informative way to visualize complex datasets.

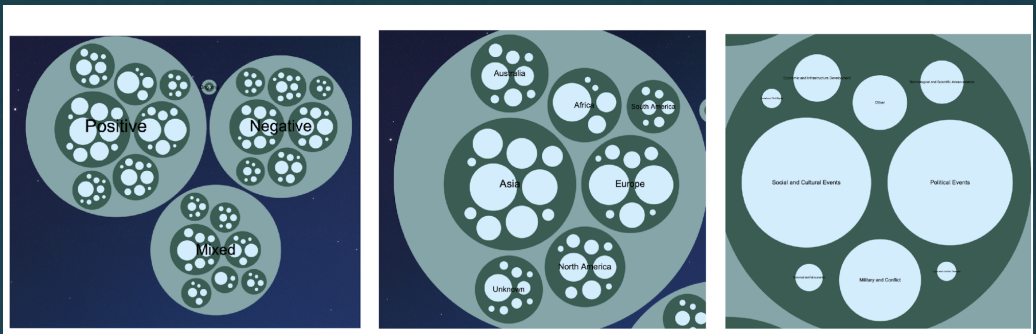

Figure 4: The three layers of hierarchy in our mosaic bubble graph

5 Dynamic Event Categorization

In our project, we developed an interactive visualization called "Dynamic Event Categorization" to analyze the relationships between various global events and their corresponding categories within selected continents. This tool aims to provide a clear and intuitive understanding of how different events are interconnected through shared categories, offering users a focused and engaging way to explore regional event data.

The visualization features a central circle containing small dots, each representing an individual event. Surrounding this circle are broader event categories, each depicted as a labeled circle. The placement of each event within the central circle is based on its proximity to the categories that best describe it. Users can hover over any event to see a detailed description and lines connecting it to the relevant categories, visually emphasizing the connections and enhancing the interactive exploration of the data.

We initially conceptualized this visualization in our Milestone 2 sketch, where the original idea was to represent global events in a similar circular layout. However, we decided to make the visualization available for one continent at a time only, as including all events globally would have resulted in an overcrowded and less informative display, as you can see below:

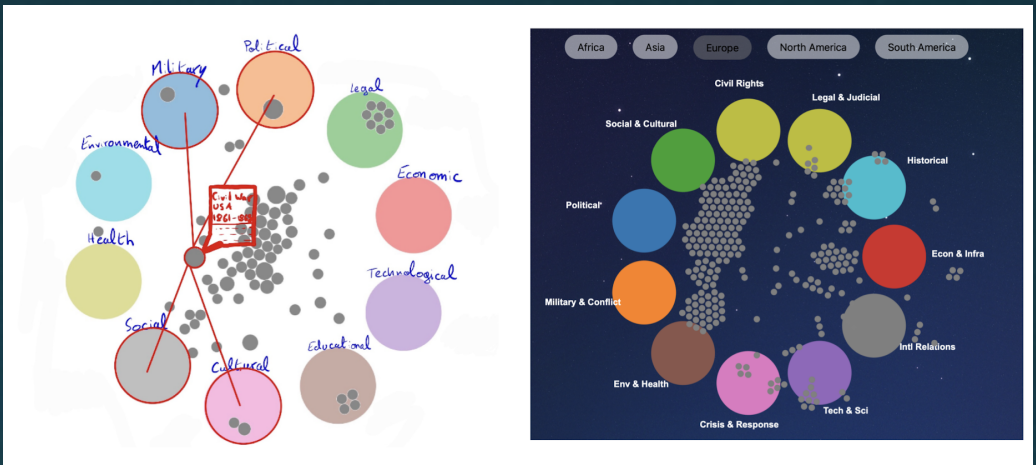

Figure 5: Initial Sketch (left) and resulting visualization (right) of Dynamic Event Categorization

One of the significant challenges we faced was accurately categorizing the events. The solution we found involved using natural language processing techniques, as described in the data processing section of this report. By employing BERT embeddings to capture the nuances of each event's name, type, and impact, we ensured a more precise categorization.

To bring this visualization to life, we utilized D3.js for data loading, filtering, and rendering. Data is sourced from a processed CSV file and filtered based on the user's continent selection. D3.js is then used to calculate the positions of events and categories, ensuring a clear and informative display. The visualization employs D3.js's force simulation functions to animate the positioning of events and prevent overlap, while also ensuring events are placed near their relevant categories. This dynamic and interactive approach translates complex data into an accessible and visually appealing format.

Users begin by selecting a continent, allowing the visualization to focus on regional data. This selection ensures that users can explore trends and patterns specific to different parts of the world. The interactive nature of the tool, including hover effects and dynamic linking, enhances the user's ability to delve into the data. The detailed information provided upon hovering over events, combined with the visual connections to categories, offers a comprehensive and engaging user experience.

In conclusion, the Dynamic Event Categorization visualization provides a powerful tool for exploring the intricate connections between global events and their categories within specific continents. By focusing on regional data and offering interactive features, it allows users to gain deeper insights into historical trends and event categorization, making it an invaluable resource for academic analysis and data exploration.

6 Impact on population analysis

The Impact Analysis visualization provides an interactive way to explore the distribution of historical events characterized by their outcomes—positive, negative, or neutral—over specified periods and on specified populations. Users can select any country of interest via a dropdown menu, and the 100 years range of interest via the widget under the graph. If there is no country selected, then the graph shows the distribution of the global events' outcomes. This implementation is quite similar to what we intended to do through our sketch in milestone 2, as shown in the figure below **??**. That was not specified in the initial sketch, but we chose the dropdown list to be the most practical and user friendly tool to choose a country.

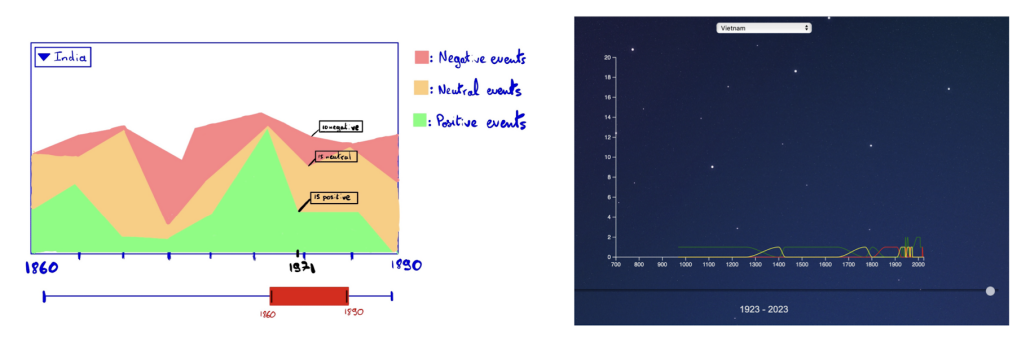

Figure 6: Initial Sketch and resulting visualization of the impact analysis

The implementation is powered by D3 and required a multi-layered SVG setup where each type of event outcome—positive, negative, and mixed—is represented by a distinct color-coded line (green for positive, red for negative, and yellow for mixed). The user interface includes a slider component that allows the adjustment of the time range, showing data from month-to-year spans, which is crucial for observing changes over time. This slider updates the graph dynamically, reflecting shorter or broader historical spans as chosen by the user.

7 Peer Assessment

In this section, we review each group member's contributions to the project. We outline the roles and responsibilities undertaken by each individual and assess how their efforts impacted the project's success. This evaluation recognizes the dedication of team members and serves as constructive feedback for future collaborations.

Maria:

- For Milestone 1, Maria was responsible for compiling the Dataset and defining the Problematic sections of the report.
- In Milestone 2, she crafted the Introduction, drafted the Visualization Sketches, and outlined the Task Breakdown for the report.
- During Milestone 3, Maria focused on the Bubble type distribution, the Bubble Mosaic, and the Impact Analysis. She managed both the implementation on the website and the detailed documentation in the report. Additionally, Maria was in charge of the introduction and the overall styling of the final report.

Elliot:

- In Milestone 1, Elliot contributed the Related Work section of the report, providing a thorough review of existing literature and similar projects.
- For Milestone 2, he was tasked with drawing the sketches and detailing the Visualization Sketches section in the report.
- For Milestone 3, Elliot handled the Event Category Visualization, which included the processing and implementation on the website, and documented these parts in the report. He also produced the project's screencast.

Killian:

- Killian conducted the Exploratory Data Analysis for Milestone 1, delving deep into the data to uncover initial insights and trends.
- In Milestone 2, he contributed to the Tools section of the report and developed the prototype of the website.
- For Milestone 3, Killian developed the Globe visualization, which included both the heatmap and point displays, implemented these features on the website, and documented the process in the report. He also oversaw the general structure and code cleanup of the website.