



FAMOUS WORLD

PROCESS BOOK

COM-480 | Data Visualization| EPFL – 2021
Ravinithesh Annapureddy - Michał Bien - Anmol Prasad



INTRODUCTION

Communicating through graphics is one of the effective mediums to convey information. Visualizations help in unearthing new relations and gives a chance to the user *to* explore and interpret them. The visualizations range from textual/ statistical summaries that are obtained from raw data to displaying images and transformed data. In the era where every day petabytes of data are generated; visualization becomes key to interacting and understanding it. On the other hand, there are already huge data available in the storages that are not necessarily generated by computers. These existing data dumps include historical information about people, nations, wars, various scientific advancements, sports, trade, travel etc. Many times, these historical topics are related to each other, however, not many systems were available until recently (Google Knowledge Graph) to connect the data points. At the same time, even when such information is located on a central system (for example Wikipedia), access to a specific category of information is not available. Through this project, we aim to create a visualization to access bio about famous people around the world.

Through this process book, we present the journey through the development of our product i.e., the website detailing the steps and solutions to the challenges we encountered.

MOTIVATION

People are born worldwide, and only those who have done exemplary activities during their lifetime become famous. However, not many know great details about them or there is single-window access to see who these famous people are, where they came from, and what they did. Pantheon attempts to solve this problem up to some extent by manually curating a set of prominent people across nations. Using this data, we decided to create a webpage that shows the famous people placed on a world map and give foundational information about the person. Our motivation is to enable users to access the information of these memorable people in Spatio-temporal dimensions at a single stop.

The primary dimension of the visualization is a location through a map. The second dimension of interaction is the time where the map is modified according to the selected time period. The last dimension of interaction is via occupation. As a consequence of this multi-dimensional accessibility, the users can overlay various dimensions at the same time.



DATA

The dataset used for visualization is the Pantheon dataset¹, which was created to study the variations in the historic biographic information. The data connects the historic individuals to time and place. Additionally, the curators have proposed a Historical Popularity Index (HPI) to measure the individual's impact. This dataset has been used by others in various socio-economic and collective memory analysis.

DESCRIPTION

The data is available as comma-separated-values on the Pantheon website. It contains data on more than 88,000 famous/memorable people. Among various attributes, we have the name, age, occupation, HPI, Wikipedia id, date and place of birth and death (name and coordinates), for each entry. These entries would Help in creating the visualization in various Dimensions as conceived above.

PREPROCESSING

The data is already arranged into a table and does not require any text cleaning or pre-processing. However, we remove the entries that correspond to organizations rather than people. Additionally, we augment that dataset with the image and a summary of the person as available in Wikipedia.

EXPLORATORY DATA ANALYSIS

To understand and select the various attributes of a person from the data in designing the website we perform the Explorations in three different points of view: the place and time of birth and death and occupation.

Spatial analysis

The birth/death locations are specified by the name and the corresponding latitude and longitude in the spherical coordinate system. As many of the persons specified in the dataset are alive, we will scope the visualization through the birth locations. Figure 1 - 2 shows the birth and death locations of all persons in the dataset identified by a red mark on the map. It can be observed that many people are from Europe.

Temporal analysis

The time of birth/death is specified by the full date and only year. We find that the distribution of the years of birth and death range from 3500 BC to 2021 AD spanning more than 5000 years of data. Furthermore, 44.30% are the century years (divisible by 100) in the BC era, whereas it's the case only for 1.62% of AD samples suggesting that the many entries in the former period are approximations.

Categorical analysis - Occupations

The occupation forms an important part of the persona. People are deemed to be famous/prominent by displaying success in the work they do. In this data, a unique occupation is specified for each person by combining similar professions. There are 101 distinct occupations and no missing values. The most frequent one is "POLITICIAN" (15585 occurrences), and the least frequent - "BULLFIGHTER" (1 occurrence).

¹ Yu, A. Z., Ronen, S., Hu, K., Lu, T., & Hidalgo, C. A. (2016). Pantheon 1.0, a manually verified dataset of globally famous biographies. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2015.75>

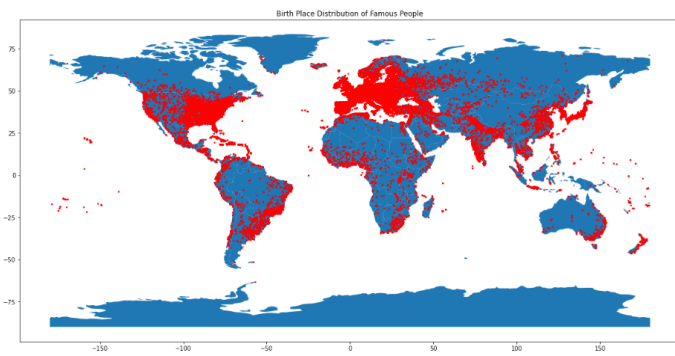


Figure 1: Birth location distribution

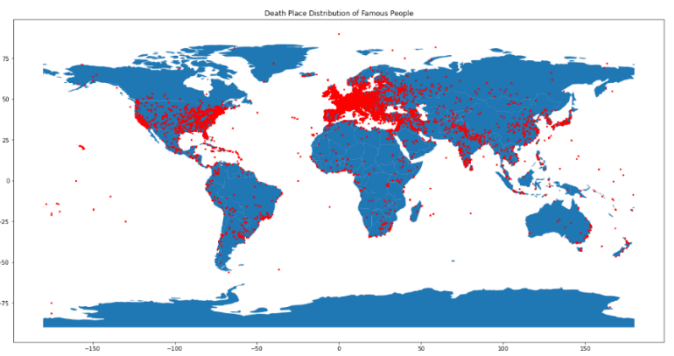


Figure 2: Death location distribution

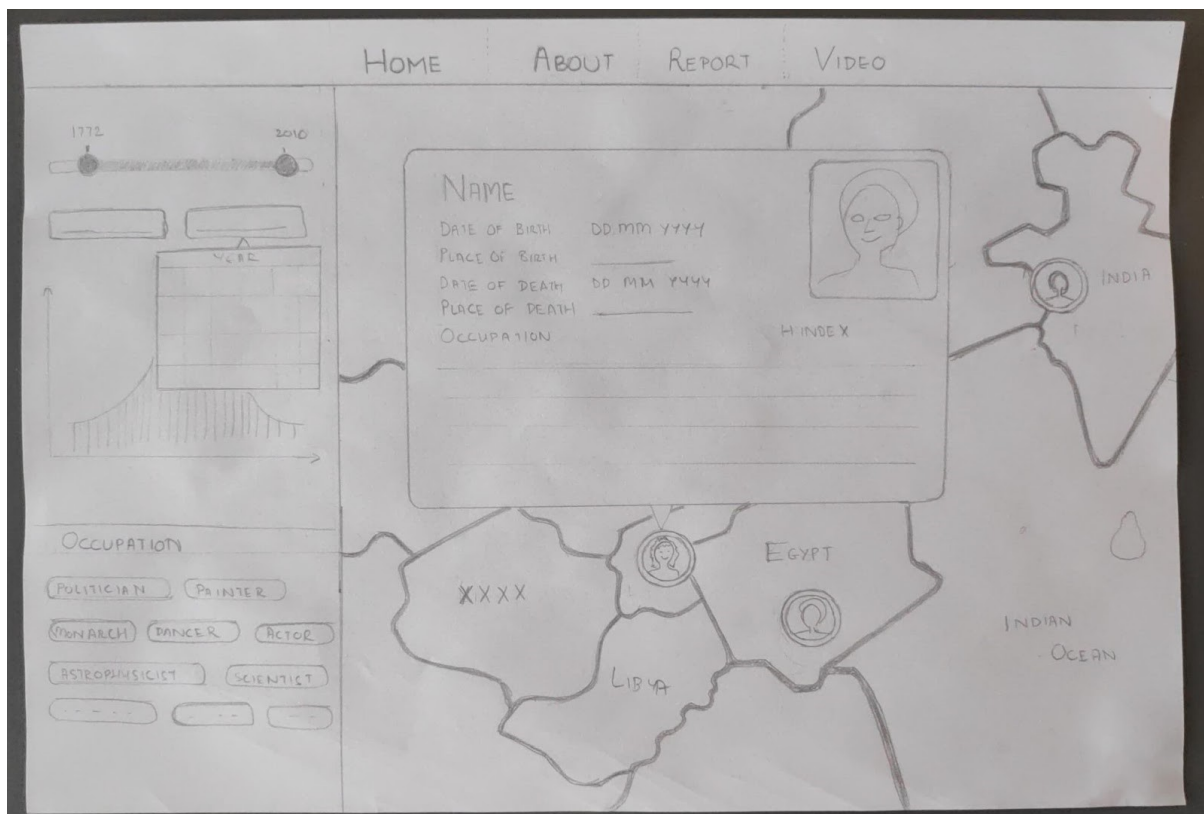


Figure 3: Proposed website design (Hand Drawn sketch)

DESIGN

PROPOSAL

As mentioned in the motivation, the central messaging of the visualization is to not only provide an interface to view the famous people but also to provide interactive tools in multiple dimensions through mouse events. The sketch of the website can be seen in Figure 3. Firstly, to be able to see where the famous people come from, we place markers for each individual on a world map allowing users to get the spatial distribution of the data. The markers would contain the picture of the person as available on Wikipedia. As we have long and detailed information about each person, it would hinder the user's message perception if everything is presented at once. Hence, using the philosophy of "Overview first and Details on Demand", the details about the person are provided as a popup on clicking any marker. The sketch of the pop up is also presented in Figure 4 and the idea was to make it look like a mini resume outlining the synopsis about the person. At the same time, for the user to be able to navigate on the map to see people from a region we add to the pan and geometric zoom options.

Secondly, to see the evolution of the places of famous people, we provide a timeline to the users to select two dates and then the map is populated with the people from that time period on the left side of the screen. The timeline is designed to be a slider with two markers. The position of the left marker would indicate the start time and the end time on the right marker. Then, to view the people from the profession of interest, we provide the list of occupations as tags. On selecting the tags, the people of those professions are displayed on the map. It should be noted that the occupation tags can be used in conjunction with the timeline i.e., the users can view people of a specific occupation in a selected time interval.

Lastly, we provide links to the about page of the project pages in the navigation bar placed on the top. The prototype of the website with the proposed sketch can be seen in Figure 4 and an instance of a pop up can be seen in Figure 5.



Figure 4: Prototype of the visualization: Home screen with map and markers

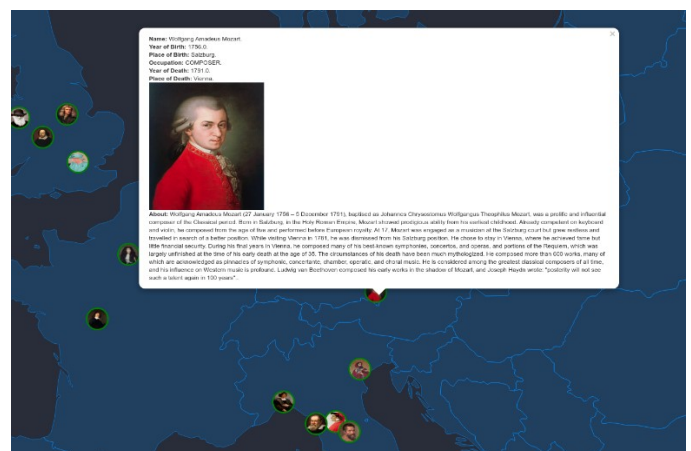


Figure 5: Example of a Person information pop-up

CHALLENGES

It is a difficult process to translate the ideas on paper to the screen and it is one of the most common problems in visual communication fields like movies, art/data visualization. We faced challenges in this project as well. Firstly, the exact date of birth of many people, especially those born in BC were only approximations. To mitigate this and include as many people as possible, we decided to use the birth year only as the temporal attribute in the timeline. The map is implemented using OpenStreetMap² and the markers and popup using Leaflet³ as it contains the options for plug and play the map rendering, markers, and popups on a map which are key elements of our visualization.

Secondly, there were more than 85,000 people and placing all of the markers at once would not only clutter the map but would also take significant time to load the webpage due to the dataset size. As seen in Figure 1, having so many markers will make the navigation difficult as there could be thousands of people in the same location and the user cannot select the desired marker. To overcome the problem of dataset size, we limited the visualization data to 5,000 people with the highest HPis. Since loading all the markers on the map would take considerable time and clutter the map, we decided to present only 50 markers, which are the most relevant to the currently visible region of the map, at the same time.

Finally, we discovered that there is a possibility that multiple markers can appear at the same location on the map, especially if the only known birth location is a big city (e.g., Paris). To decrease the chance that these markers would stack into each other, we add a small random value to both latitude and longitude of every marker on the map. We discovered that this action fixed the problem and made the city-related markers disperse nicely in the whole area of the city.

We found that placing the navigation bar on the top and the filtering options on the left is hindering the space that the map can use. Thus, we shifted the filtering options to the navigation bar and moved the bar to the left side of the screen with an option to hide them to utilize the space to the maximum extent.

The two major technical problems were loading images and presenting the details of the person whose marker is selected. As the images are obtained through the URL to Wikimedia sequentially, the time to load the page was long. To avoid this, we have used lazy loading⁴. To decrease the size of thumbnail images used in markers, we used Wikimedia thumb API, which allowed us to request image resize to 64x64 server-side, which considerably decreased the weight of data. The full-size images are loaded only when the specific pop-up is clicked. While making a popup is easy using Leaflet, arranging the elements aesthetically is not possible only by using HTML code. To alleviate this, we decided to use an iframe⁵ that embeds an HTML page into the current one.

FINAL DESIGN

The final visualization of the website displays a world map that has markers with the photograph of the person at their birth location. On opening the website, a little description about the makers and the features of the website was presented on a left side toolbar with an option to hide it (Figure 6). On the same left toolbar, there are options to set a range of years and professions to

² OpenStreetMap contributors. (2017). Planet dump retrieved from <https://planet.osm.org>.
<https://www.openstreetmap.org>.

³ <https://leafletjs.com/>

⁴ https://developer.mozilla.org/en-US/docs/Web/Performance/Lazy_loading

⁵ <https://developer.mozilla.org/en-US/docs/Web/HTML/Element/iframe>

filter the data (Figure 7, 8). The profession filtering option was implemented using the Select2⁶ library as it's a common library to incorporate forms without many dependencies. The slider in the timeline filter was implemented using a Nouislider⁷ library as it's a lightweight UI library with no JavaScript framework dependencies. Next, on clicking the markers on the map, the user can see the mini-bio of the selected person including a summary from Wikipedia. To differentiate between the people born during different times, we have designed four popup types with different backgrounds and fonts (Figure 9). Lastly, to see the evolution of these people in time, we have added a time-lapse play option in the timeline filter page. In the time-lapse feature (Figure 8), the users can select a period using the from and to options. Then on clicking play, the markers of people born between the “from” and “to” years first appear on the map. Then the widow is advanced further until 2021 while maintaining its size. The users can moderate the speed of appearance (or disappearance) of the markers on the map using the fast forward or backward buttons. In addition, the feature to join people having same profession using a link and an animated line from a person's birth place to death place were removed in the final product as they were cluttering the viewport and it was not intuitive to the users to interpret.



Figure 6: Final Design of the visualization: Home screen with map and markers

⁶ <https://select2.org/dropdown>

⁷ <https://refreshless.com/nouislider/>

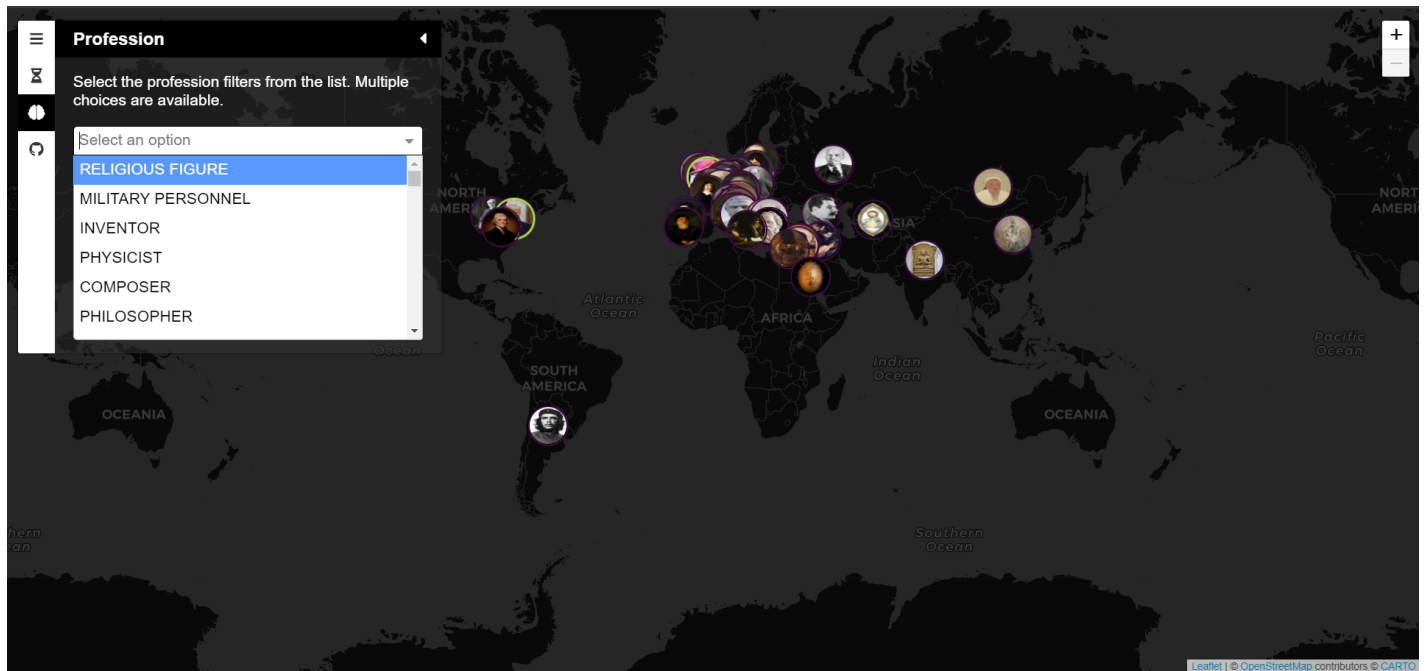


Figure 7: Final Design of the visualization: Occupation filter

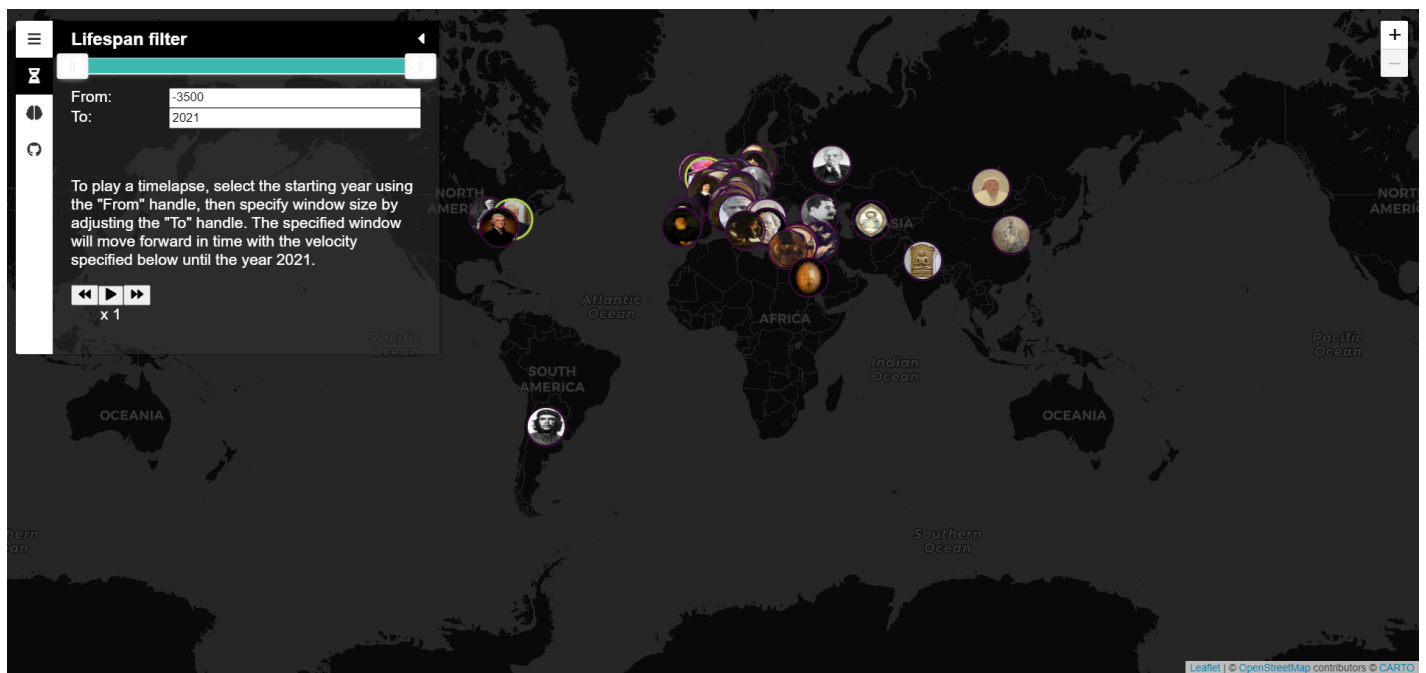


Figure 8: Final Design of the visualization: Timeline filter

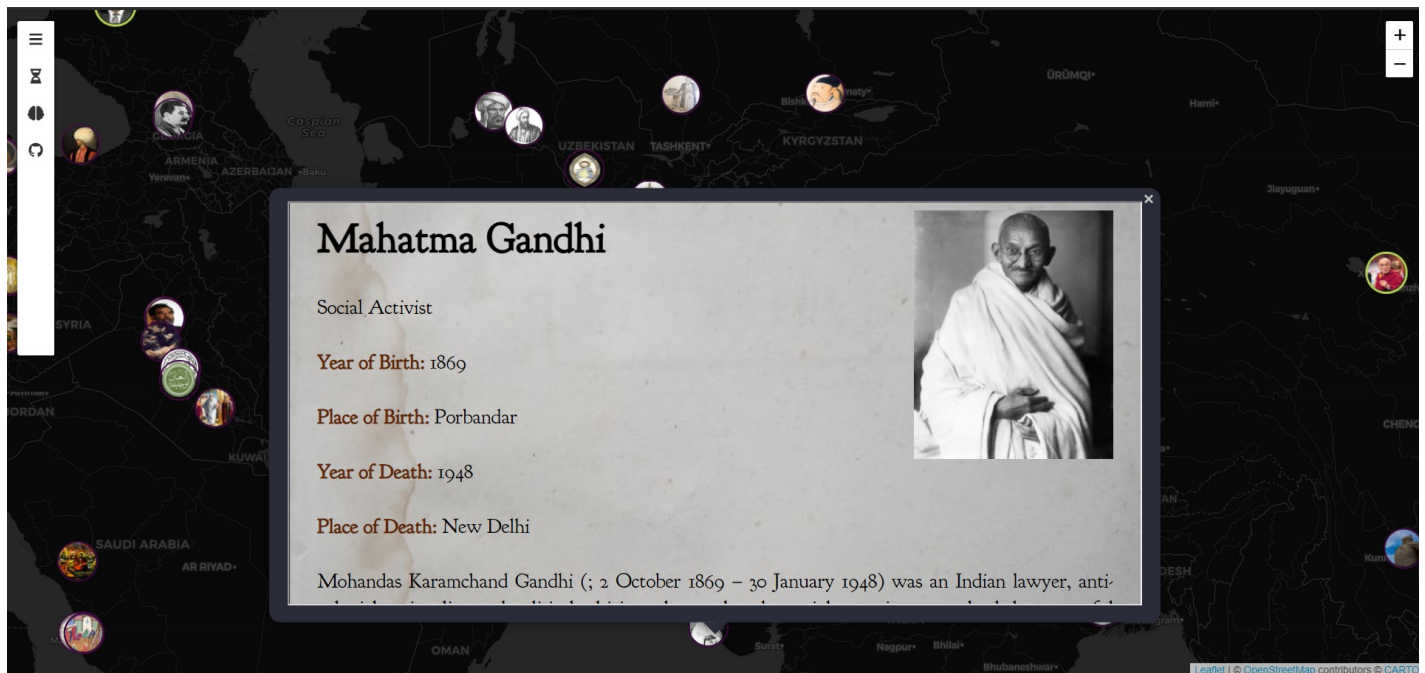


Figure 9: Final Design of the visualization: Person Bio



PEER ASSESSMENT

On a larger scale, all team members have been part of important design decisions and jobs were divided after consent with an aim to achieve the end goal of the project. The main contributions of each team member are as following:

Anmol: Sketched website design and built the basic structure of the website with map and markers containing images. Improved popup formatting. Edited process book.

Michal: Data pre-processing pipeline, designed and implemented the sidebar along with the implementation of filters and time lapse feature on the map. Edited script for screencast.

Ravinithesh: Initial project idea, prototyped the website with popup, modified styles of the popup and implemented dynamic resizing of the popup, created the process book and recorded the screen cast.