IT'S MOVING! – A VISUALIZATION OF THE MOVEBANK DATASET ON STUDIES ABOUT WORLDWIDE ANIMAL MIGRATION

RUBEN BURDIN – 338367
LOVRO NUIC – 338203
FABRIZIO FORTE – 328932

PROF. LAURENT GILLES MARIE VUILLON

TABLE OF CONTENT

Online Website

Video

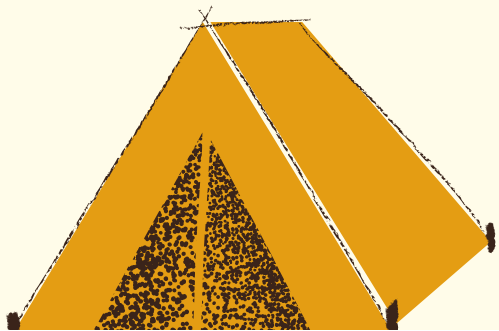
FROM IDEA TO PRODUCT

This section shows the progression of our team from ideation to product implementation. The challenges and decisions taken are documented in the relevant section below. The goal of this section is also for our team to reflect on the development of the project and identify the key conclusion to optimize future works that we will take on during our future studies and career.

IDEA

The initial idea was to create an engaging story about animals and wildlife, due to our personal interest in the topic. We found the exciting **movebank dataset** about animal migration, this demonstrated to be a perfect dataset for our purpose.
Our project aims to **focus on geographical visualization** (i.e. globes and maps) and describe geographical features and movement in **several various tree and chart data visualizations**. The initial idea was to create a globe visualization and gain experience on that, showing dynamic elements on that same visualization. The globe visualization would be then inspired from globe.gl, and other visualizations would be inspired from D3.js.

TARGET AURDIENCE



The goal of the project is twofold: to show fellow students the way some familiar (or less familiar) animation migrate across the globe in an **interesting story**, and to teach us how to **design and implement a data visualization** project with an industry and production-ready stack.
The **target audiences** are our fellow students, friends, and families that would like to learn about the topic.

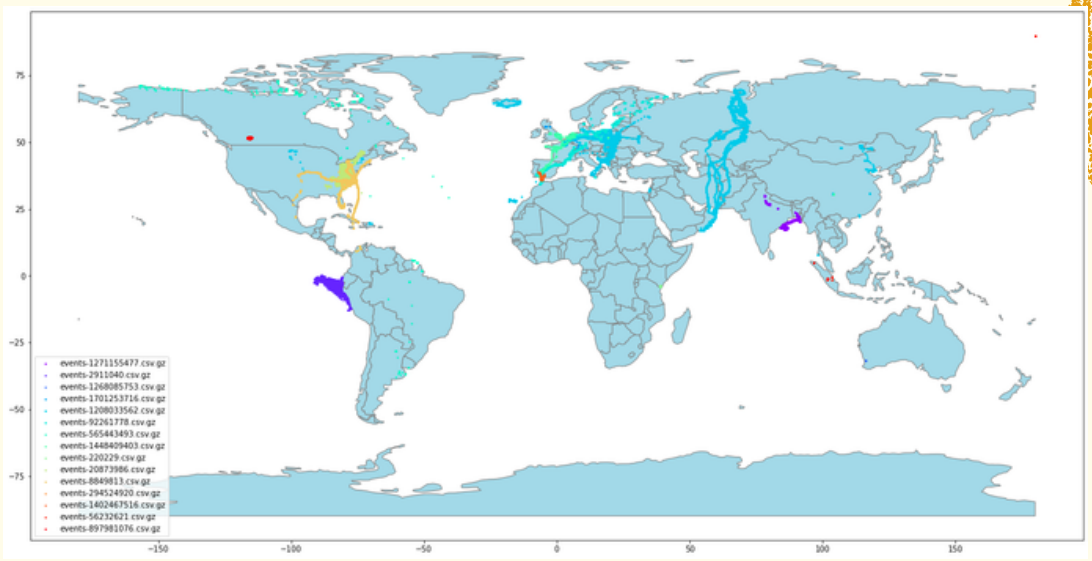
DATA EXPLORATION

The data exploration/data cleaning part was painful!
We downloaded our data from the Movebank.org dataset, which collects several studies about all kinds of animals, which are georeferenced (that includes Homo Sapiens, cats, cats, and ... plastic bottles!). In addition, information about the studies was separated by the actual records. As a first step of the data exploration, we aggregated all the data together, then we filtered the studies, retaining only those of interest. For instance, we excluded those studies where more than one type of animal was observed. Then we preprocessed the data, removing outliers, and interpolating the trajectories when necessary.

The studies have different time spans, so we realized that visualizing studies with different time spans would not have been visually optimal. Single studies also have different sizes. This is something that we had to keep in mind in the implementation!
We also realized how different the statistics of different animals are: some species migrate over short distances, while others cross many countries. All these insights were crucial to decide how to best visualize our data and what studies to focus on.

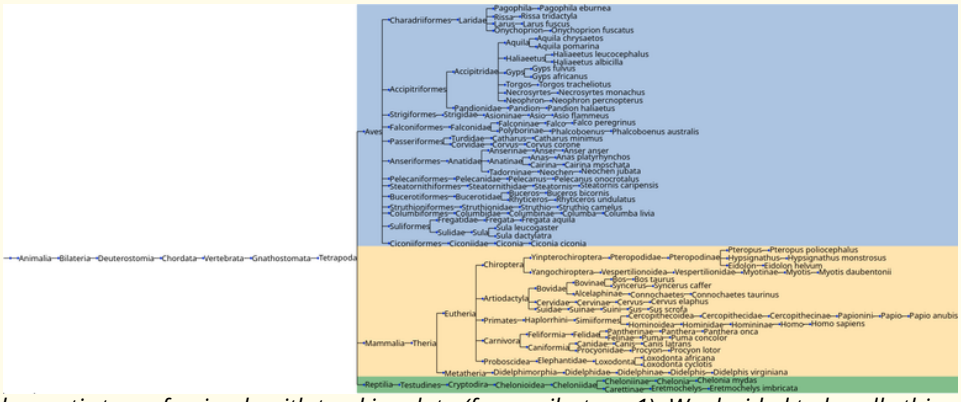
MOCKUPS

After some extensive discussion about the storyline, we found that our main visualization would be a dynamic globe showing the movement of animals over time. Some additional visualizations include a tree (selector), line charts, and text visualizations.
The initial mockup of the main globe visualization was as follows.



Several studies were plotted in the same world map. Every study is in a different color (from milestone 1)

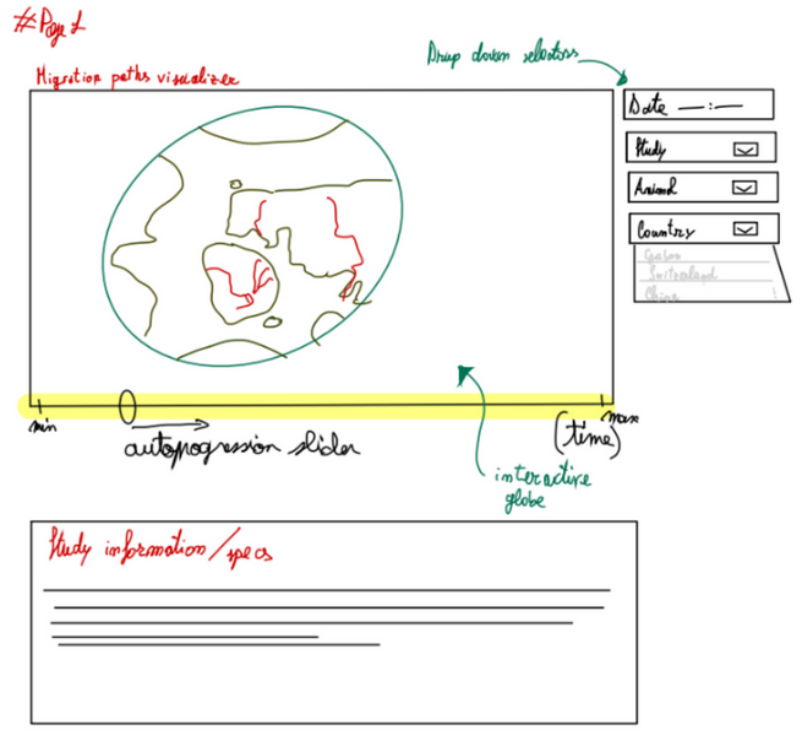
The geo-visualization will be bundled with interactive filters to focus on certain studies/species/single animals/time periods. Users will be able to zoom in and out of certain regions and simulate time travel.
We wanted to show the visualization of animals and their relationship between species in a graphical tree. We found it very interesting to actually turn this visualization into a selector and make it possible to visually understand and select (i.e. filter) the animals to display with a dynamic tree. So we could bundle the useful and the visually compelling and intuitive.



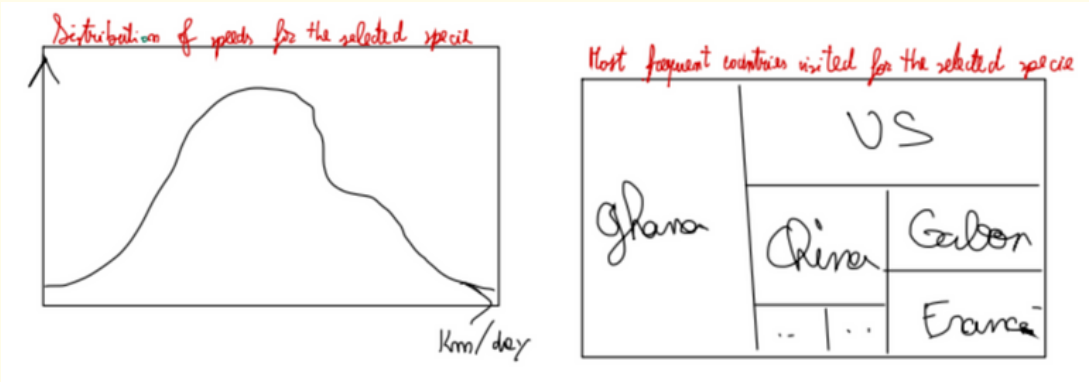
Phylogenetic tree of animals with tracking data (from milestone 1). We decided to bundle this tree into a visualization and a selector with a single graphical element (see foam-tree later on).

REFINEMENT 1

We engaged in a user journey and storytelling workshop so we could define exactly what would be our needed visualizations, taking into account the time and computing resources available to develop these. The visualizations below refer to some additional charts with UI and filtering options. Some additional visualization and animations were also designed (too many drawings to display in this report, only a small selection of them is shown below).



Main globe visualization sketch with filtering and dynamic date range options. The globe is animated (very complex animation). Sketches from milestone 2 report.



Sketches of line charts and text visualizations that will strengthen the storytelling and insights for the user. Sketches from milestone 2 report.

At this point of the project, most of the UI and interactions were defined. We were confident that the storytelling would be fully supported with our initially proposed visualizations, and even more with the additional ideas (extra-mile visualizations).

REFINEMENT 2

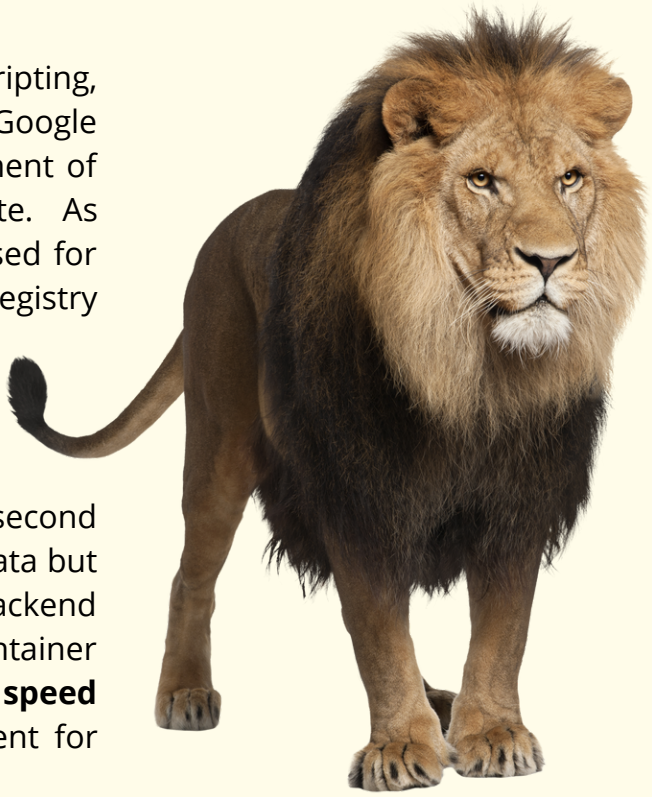
To make the selector tree even more compelling and original, we decided to visualize it as a foamtree. A screenshot of our MVP can be found below.

MVP of the foam tree selector, after the second refinement phase.

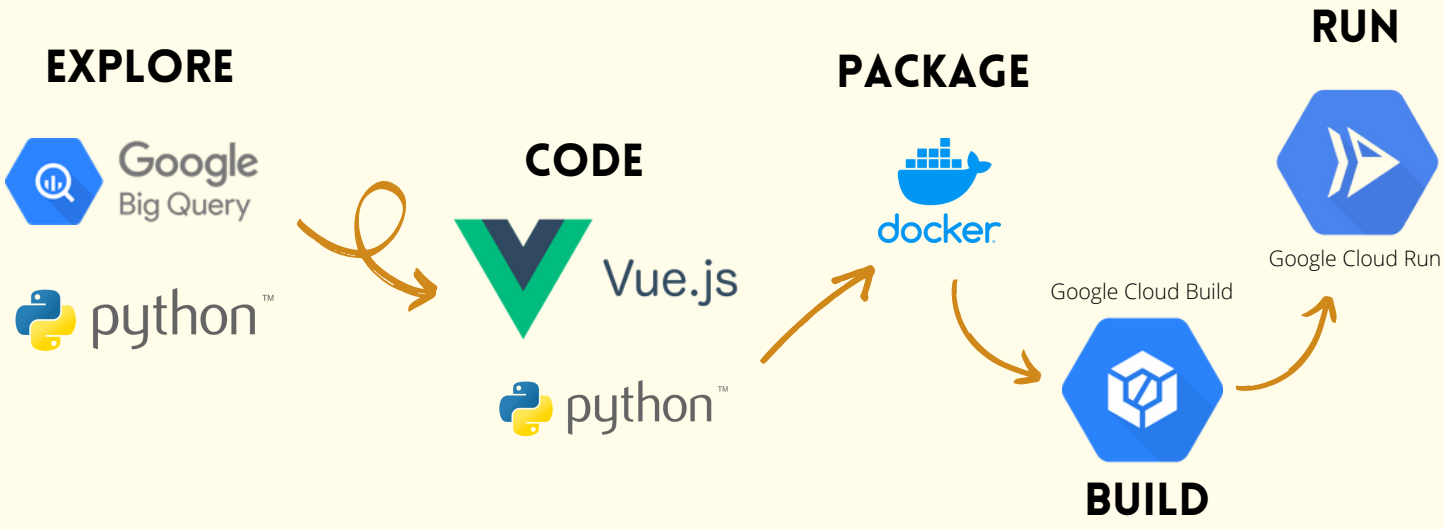
THIS WAS THE LAST STEP OF THE UI PLANNING AND MVP. THE FINAL IMPLEMENTATION CAN START!

TECHNICAL IMPLEMENTATION

The website has been implemented in Vue.js, an increasingly popular framework. We chose this framework since it is easy to get started with, versatile and was something we wanted to learn for our careers. The coding has been done in Javascript, HTML and CSS.
Styling has been firstly designed and mocked-up in Bootstrap studio, a visual editor for design elements. We then transcribed the page into Vue.js. For the visualizations we have used different libraries such as **globe.gl** for the main globe visualization, **carrotsearch** for the foamtree, **leaflet** for the flatmap and **chartjs** for the rest. We chose to explore several libraries in order to learn how to use them and develop our skills at maximum. We have tried **plotly** too but after having some challenges in the compilation of the docker image in Google Cloud Run, we decided to use chart.js instead.



Backend has been using Python for the scripting, BigQuery (BQ) for the data warehousing and Google Cloud Build/Run/CI-CD for the automatic deployment of a production-ready and hyper-scalable website. As mentioned in the report 2, google cloud run is used for deployment, google cloud build and container registry for the build of the software into a docker image and BQ for the database as a service. Since our data was large, we opted for a hybrid of BQ for the data warehouse and loading of local files.
For fast data serving, BQ can serve as fast as 1 second response time for query, which is good for large data but not for lighter datasets. For instance, having the backend to load the data from local files in the docker container instead of querying BQ helped us **gain 1 second speed for each data query**, a considerable improvement for the end-user experience.
Our project distinguishes since our stack is fully production-ready, as reflects industry best practice for serverless and scalable deployment.



CHALLENGES AND DECISIONS

We faced numerous challenges during the project and we had a large learning curve due to these. They will be very important for our upcoming career since these are likely challenges and design choices faced in almost every industry project.
We decided to split the issues we face into 2 categories: (1) the backend issues that were relevant for technical stack choices and performance of the website mainly, and (2) the frontend issues including the technical implementation of frontend elements, serving and design & choice of visualization + storytelling.

BACKEND CHALLENGES

Framework to use for the Frontend
One discussion in the group was about the framework to use for the frontend. The easy solution would have been to use React since most of the community is on this framework but we decided to use Vue.js instead since we wanted to learn something new. Vue is actually becoming more and more popular in the industry. This framework is flexible, easy to learn with previous experience in javascript and is very versatile.

Data Serving Speed
First, data processing has not been a hurdle since we started right with the right (almost clean) data and the right setup for data exploration. For the data processing we used a mix of Python and Google Big Query (called BQ, a NoSQL cloud data warehouse) that would enable us to quickly clean, process, and deeply explore the dataset.
The first version of the backend was implemented with Python and a data storage in BQ. The problem was that since that data was big, it would take around 10 to 30 seconds to actually receive the data in the frontend. The user would have to wait for all that time before seing the first visualization appearing. We understood that using BQ was great for data processing and exploration but not for serving user in real-time since the BQ solution is designed to be an analytics data warehouse and not a serving data solution. The challenge was to reduce the waiting time before the data would start to display to the user. To improve the user journey, we kept the python backend which was already very efficient but loaded the cleaned data directly as a file in the docker container so that we would not have to wait for the bigquery processing and load over the network to start processing the backend data request.
This resulted in a reduction of waiting time, namely less than 2 seconds waiting time from the website.
We learnt that using the right stack for fast data processing is crucial to great performance on the website.

Backend vs. frontend processing?

For the visualizations related to the direction of the animal, we started to have a separate table with pre-computed values in the backend, that would serve only the visualization about the direction of animals. After doing some work on the frontend, we realized that the visualization would be more compelling with extra information and would require modifying both the backend and data processing to get this extra information. We realized that all the information could actually be computed in the frontend without the need for a special preprocessing in the backend at all. We decided to move the processing of the data to the frontend with the data that was already fetched for other visualizations instead of doing it in the backend and required additional data to move over the network, potentially creating some waiting time for the end-user.

This showed us the need for extensive planning before starting to code any of the backend or frontend elements. We also learned that optimizing for network traffic is very important to provide a seamless experience for the user and less waiting time.

FRONTEND CHALLENGES

The development of the frontend was a challenge in many aspects. In fact, not only the design and the large database were real challenges for us, we also had to learn javascript from a novice level.

Huge database

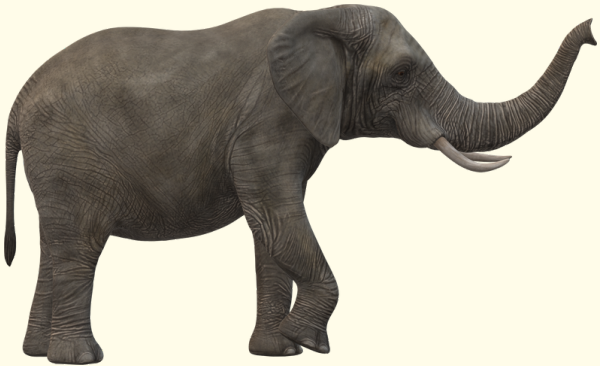
The dataset used was rather clean on the download and needed only a few days to be cleaned, but since we were working with geographical and timestamped data, the volume was very large (2 GB in total). The issue was transferring this amount of data over the network, thus slowing down the entire website and user experience. We thus had to do further data processing to reduce the network traffic but also some additional data manipulation in javascript, which is not a very easy language to manipulate data. In the end, we managed to do a real production-ready setup to serve requests quickly and efficiently.

Display GPS trajectories on the Globe and the Maps

Geographical data is known to be difficult. We wanted to do the extra mile for this project by exploring several types of maps and globes. We had a lot of issues with the data formatting and do working versions of the charts on our website. This took a lot of our time, this is a remarkable effort that distinguishes our project.

Design and responsiveness

Designing a good-looking and responsive website was a long journey. We did a lot of trials in bootstrap studio to preview designs as we wanted to do something visually appealing for this project. This took also a lot of time to design and implement with all the stylings in HTML and CSS.



Learning a new language

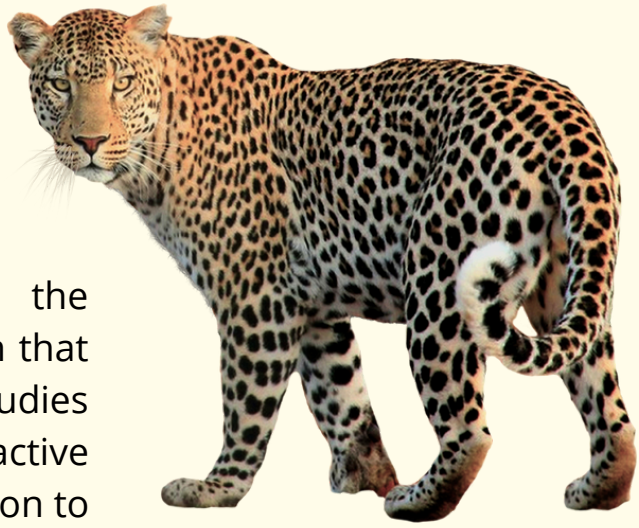
Our team was a novice with Javascript and we put all of our best effort to learn it from scratch. Vue.js and Javascript were completely new for 2 of us. This was also a great experience to help each other and learn together in a hands-on project!

RELATED WORKS

Other studies on the same data

- ***Movebank*** *dataset summarize*r (link): Summarizes data sets showing bounding boxed of animal movement in studies and general information about study.
- ***DynamoVis*** (link): An interactive tool for the exploratory visualization of animal movement data. Made for biologists in the form of desktop applications.

OUR CONTRIBUTION TO THE COMMUNITY



Why is our approach original?

While there is scientific research on the individual studies, there is no visualization that would combine the data of different studies and show animal migrations in an interactive way. Our goal is to present data on migration to non-experts in an accessible way and show the dynamism of the animal world that is greater than we usually think.

We are planning to publish and leave this website serving for a longer period of time for the community to use it for some better analysis and data exploration.

PEER ASSESSMENT

Lovro
Main Globe
Foam tree
Design implementation
Website

Fabrizio
Data Exploration
Storyline & Viz
Website

Ruben
Deploy & Production
Flat Maps,
Spider Chart,
Proc. Book

UTILITIES

Github: https://github.com/com-480-data-visualization/datavis-project-2022-vizsquad

Website: https://data-viz-course-frontend-e4sglwkopa-uc.a.run.app/

Video: https://drive.google.com/drive/folders/1wq44vAApNfggwb3BWWg4OhasPRtrd84G?usp=sharing

THANK YOU!