DATA VISUALIZATION

MILESTONE 3

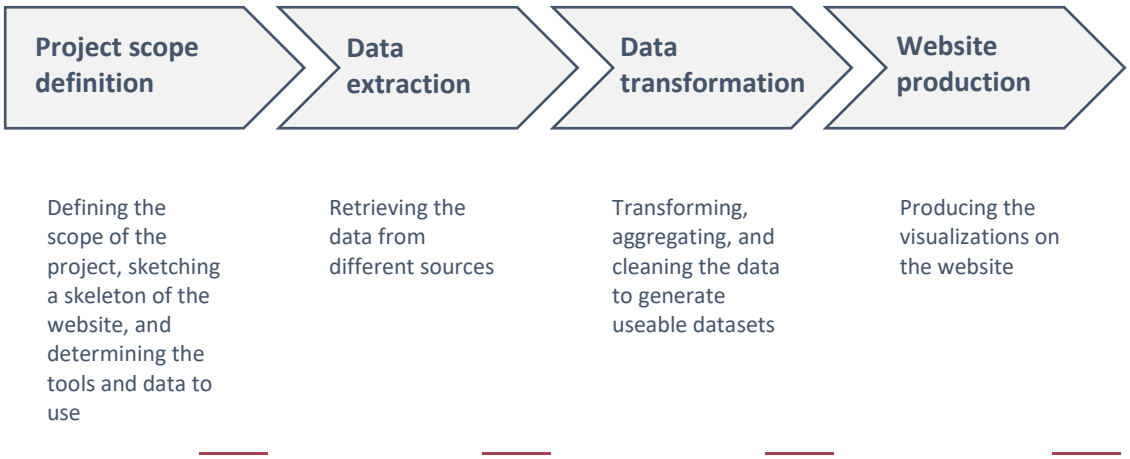
Group: NAA

Arsenii Gavrilenko (337943)
Anastasia Filippova (337602)
Nils Hasselmark (302267)

3 June 2022

ROAD TO FINAL WEBSITE

The project road consists of four main phases, each of which will be further detailed in this section.



The challenges we faced and the design choices we made will be described in each of these steps.

1. DEFINING THE PROJECT SCOPE

Our original idea was to come up with a tool aiming to draw an analytic picture of US politics. However, as the subject is relatively wide, we had to restrain the scope of the project to a smaller topic. We therefore decided to focus on the most emblematic process in a democracy: the presidential election.

We wanted our website to be an educative tool, designed to better understand the way US democracy works, as well as an analytical tool, allowing to visualize the dynamics behind the different elections.

Here are some of the questions we wanted or website to answer:

- *Can we highlight geographic states preferences toward a given party?*
- *How do people vote in urban areas compared to rural areas?*
- *How do electoral trends evolve with time?*
- *How do parties popularities evolve with time?*
- *Could we observe diversity among US leaders?*
- *Do we have any specific bias in education or religion of US leaders?*

From these points, we tried to design a first sketch of the website on Miro, using only simple charts copied on the internet. The goal was for us to quickly understand the tools and datasets we would need in the following phases, and to have already at the project's kick-off a clear view of the road ahead of us.

2. EXTRACTING THE DATA

Now that the project goals were clear for us, we had an exact understanding of the datasets we would need for our website. We figured out that we would need three distinctive and complementary types of information:

Presidential elections results at a State level

Candidates personal information

Past presidents personal information

We naturally needed data about the results of each presidential elections, at a State level so that we could analyze geographical disparities among the U.S. territory. This would allow us to analyze how US citizens are voting. But it was not enough to build the tool we thought about. We needed to retrieve data about the personal information of each candidate, such as their age, and their gender. And finally, we needed some data about the past presidents of the United-States: when they got in the office, a short biography, and a picture.

We found a cleaned dataset for the first pillar in the MIT website. We could not find data older than 1976 However, there was no satisfactory dataset for the second and third pillar.

We had to create the datasets by our own. This was made using the power of the WikiData API, for which we had to learn the SPARQL querying syntax. Unfortunately, the data extraction only worked partially: a lot of candidates of past presidential elections were missing, even some famous politicians, which had not been correctly labelled in Wikipedia. We had to complete manually the dataset, by looking individually at each personality. Considering the time it took, we decided to retain only the ten most popular candidates for each election.

The WikiData extraction of past presidents' personal information worked much better, but we still had to write manually the biography of each of the past presidents in order not to plagiarize Wikipedia.

3. TRANSFORMING THE DATA

Now that the three datasets were obtained, transformations were needed in order to create relevant and useable JSON files in the production phasis. This was done in parallel to the development of the website, as it was difficult to predict in advance which data format and schema would best fit the tools used to create the website.

Multiple JSON files were created, using different data engineering tools such as Google BigQuery and Python with pandas.

Some examples of transformations needed are:

- Join datasets, sometimes of different format (csv and JSON)
- Modify datasets schema
- Aggregate the data to a higher granularity (for example State to country level, or election to lifetime)
- Create new variables and rank the candidates in each State
- Clean missing or incorrect values

This phasis was more time consuming than we expected, as the datasets extracted from WikiData were not as cleaned as we first believed, and as the schemas were incompatible with the libraries we used.

4. PRODUCING THE WEBSITE

We decided to use the Javascript framework React for our project. We believed this project was a great opportunity for us to learn this popular frontend framework, and it appeared as well very flexible and robust. One of our team members already had a humble previous experience with React, but it still was a challenge for all of us to learn the basics of the framework, especially considering the fact that the two other team members had no frontend experience whatsoever.

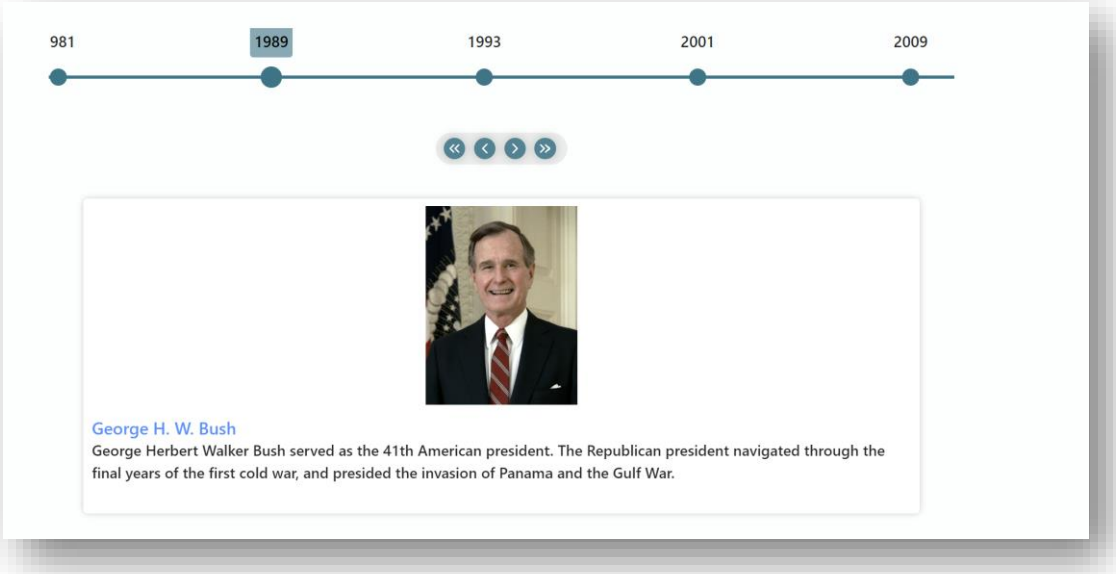
The website is divided into five main sections, each with its own visualizations.

❖ The American election process explained

This section is purely educative and shows no analytics at all. The goal is to make the user understand the way U.S. presidential elections work before showing the first visualizations.

❖ A look at the last American presidents

In this section, the objective is to display an interactive tool showing the recent successive American presidents. We thought showing famous portraits here before going into details with the analytics might be more adapted, to catch the user eyes, and to recontextualize each election.



The visualization was made using a horizontal timeline. We firstly built it using vertical timeline, but we soon realized that it was taking too much space compared to the other.

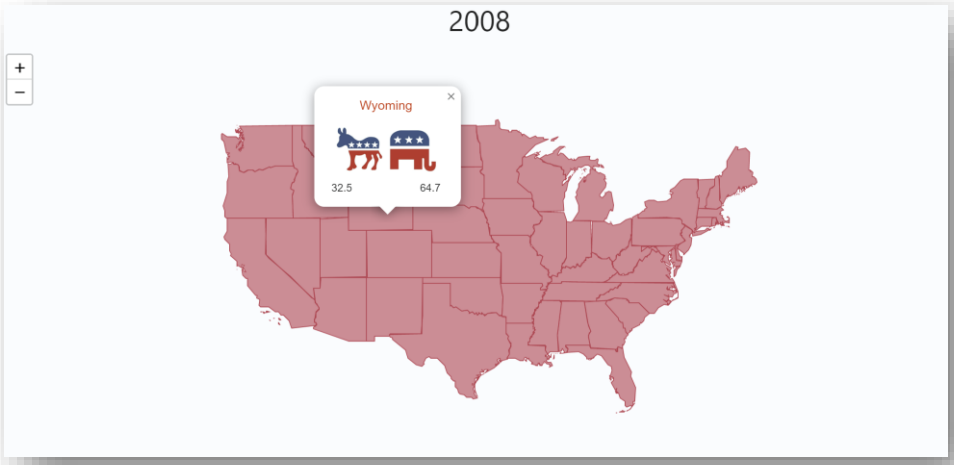
❖ Two historic parties competing for the presidency of the country

Before showing the first analytics charts, we thought it essential to explain who the two main actors of the elections were. We briefly described the two historic political parties of the United-States, and then added the description of a third more recent party. This was a crucial step as the next sections focus on the dynamics between these three parties.

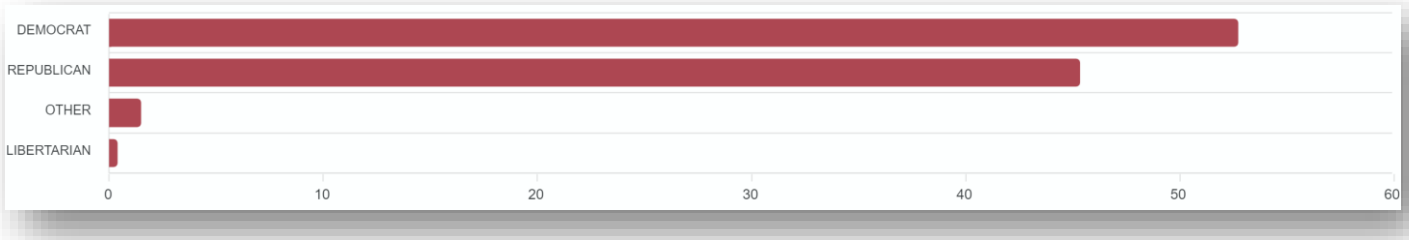
❖ Geographical trends showing a country divided in two parts

The study of the dynamics of the US presidential elections really starts in this section. Here, two visualizations are given:

- **A map chart**, intended to highlight the disparities of the votes in the U.S. territory. It gives insightful knowledge of the differences of votes between the different States of the country



- **Comparison bars** showing the results of the election by aggregating the votes at the country level. This is extremely interesting as it shows how candidates with the highest number of votes are not always the winners of the election

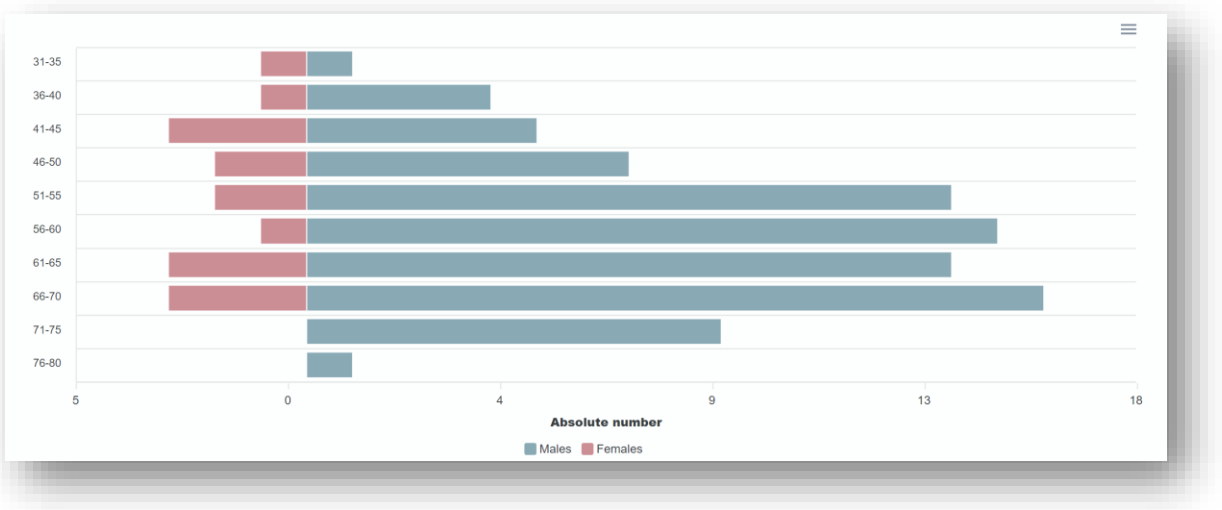


The website allows the user to select the year of the election he wishes to analyze. Our first idea was to create a scrollbar with play and pause options, but we realized it would be complicated to implement so we decided to build a Carosel component to select the different years instead. Unfortunately, it was not possible to change the color of the buttons, so they are not very visible.

The making of the map was very challenging, especially when it came to understanding how to make mouse on and pop-up events. It was done using the geojson and leaflet React libraries. The region projection was also tricky, and needed some specific format of data combining the geographical coordinates of each state.

❖ A closer look at the candidates of the presidential elections

This final section focuses on the characteristics of the candidates of the elections. We thought that this part was very interesting so that we could draw a typical portrait of a US politician. It turns out that politicians in the United-States are mainly old males, and women still seam behind of the politics.



As stated above, we had to use data taking into account only the top 10 most popular candidates for each election.

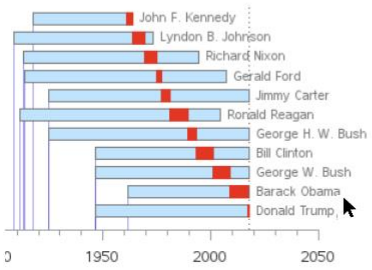
CHANGES FROM PREVIOUS MILESTONES SKETCH

The first skeleton is given again below.

Understanding US Presidential elections

This website was developped for the EPFL course Data Visualization given in 2022.

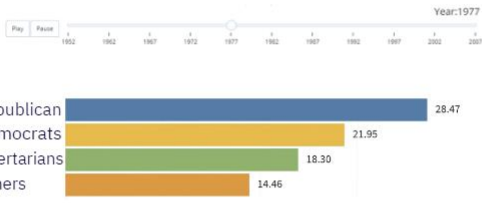
US Presidents history



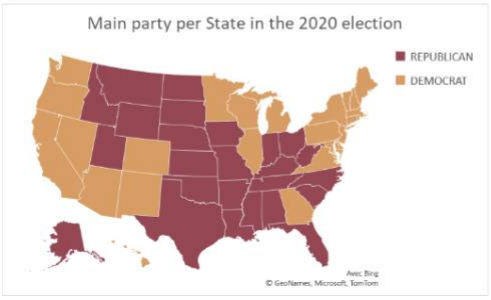
1

Barack Obama
(1961 -)

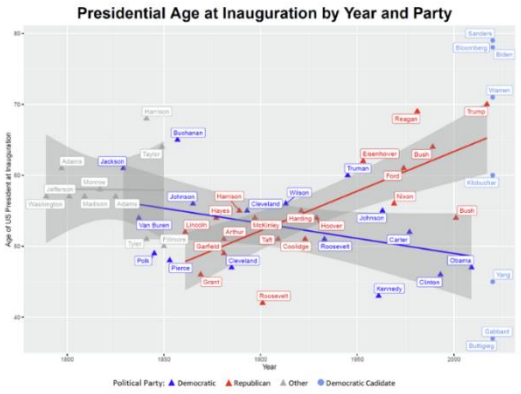
Party popularity



2

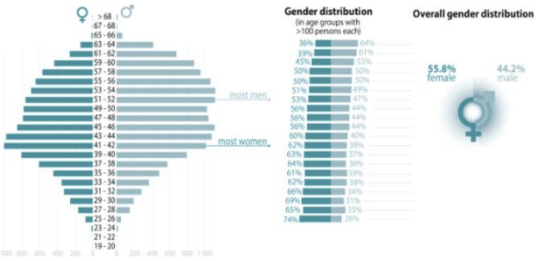


Candidates characteristics



3

US Candidates gender and age repartition



You can observe some changes in the final website:

1. We decided to remove the graph showing the presidents mandate durations, and replace it with a timelapse showing the successive presidents. It was not a question of technical difficulties, we simply realized that the design would look nicer that way.

2. Unfortunately, we judged it too hard to implement a time cursor bar which could be played and stopped as a video. Instead, we added buttons to select manually the year of the election.

3. This part was put under “Extra Ideas”, and the design was not very clear at that time. We finally decided to implement a graph showing the age and gender distribution in the same plot. We did not have time to do the other plots.

PEER ASSESSMENT

After having defined together the subject and the datasets we were going to use, we tried as much as possible to split the workload between the group members. We regularly organized meetings to share our advancements and to help each others. Thanks to her past experience, Anastasia was the leader of the project and guided us a lot in the project. She was the pillar of the project.

- Nils mainly worked on:
  - Storyline and structures of the three milestones reports
  - Data transformation and cleaning for the reduced MIT dataset
  - Texts for the website
  - Graphical drafts of all visualizations
  - **JS**: US presidents horizontal timeline

- Arsenii mainly worked on:
  - Making the screencast video
  - Data scrapping from WikiData
  - Data processing & Merging data from MIT Lab and Wikidata
  - **JS**: Bar chart for Parties
  - **JS**: Gender charts

- Anastasia was in charge of:
  - Managing the GitHub repository
  - Creating the skeleton of the website
  - **JS**: Geographical analysis (Map, scroll bar)
  - **JS**: Enhancing overall website design
  - **JS**: Code review for Nils and Arsenii visualizations
  - **JS**: Assist with deploying code of Nils and Arsenii