DATA VISUALIZATION

MILESTONE 1

Group: NAA

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

8 April 2022

DATASET

The main goal of our project is to prove or reject some common beliefs about the United-States, and to highlight interesting casual relationships between the life of the country and the public opinion. We believe this last point is directly expressed by the results of the US presidential elections, which occur every four years.

Hence, we chose our project to be about providing a visualization tool for the United-States US democratic elections.

Two datasets are used in this project (please click on the links to access the database):

Presidential elections results (1976-2020)

Official data precleaned by MIT, providing the number of votes for each candidate at a state level

Candidates personal information

Wikidata is used to retrieve personal information for all candidates, such as the birth date, gender, place of birth, etc.

We are merging the names of the candidates with the Wikidata information. Unfortunately, Wikidata turned out to be missing many popular candidates. A manual extraction had to be done for many of them, that is why we decided **to keep only the 10 most popular candidates for each election.**

On the other hand, the MIT dataset is pretty clean, and has some null values only for the least popular candidates that we decided to drop anyway.

The preprocess has been done in Google BigQuery.

PROBLEMATIC

The aim of this project is to create an interactive website highlighting the dynamics in the United-States democratic process. A focus will be made on the electoral behaviors, and an educational presentation will be given.

What are the dynamics in the US democratic process?

The tool should provide quick answers to questions such as the followings:

- *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?*

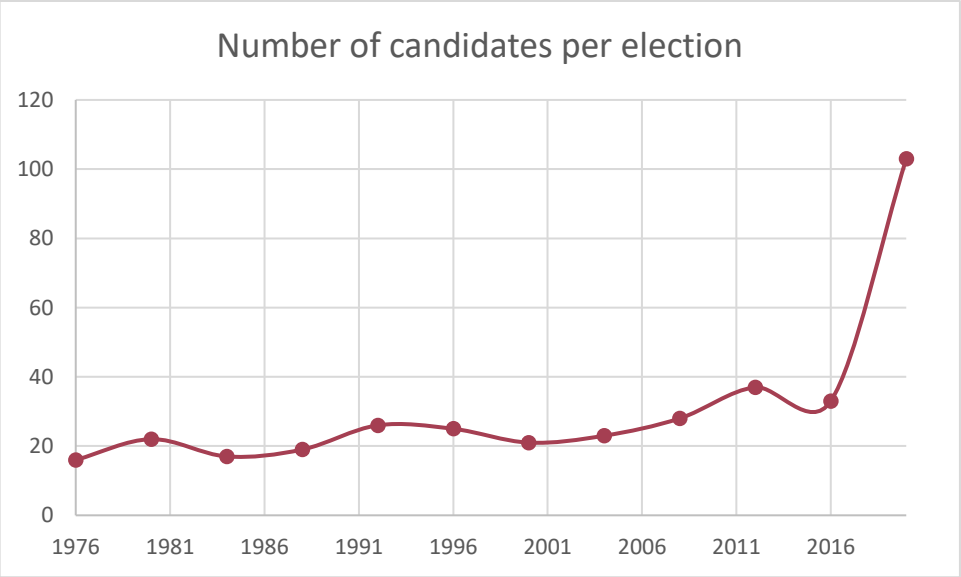
These points imply that the data should be explored with refined granularities, that is with yearly evolution, and at a state level.

Presidential elections

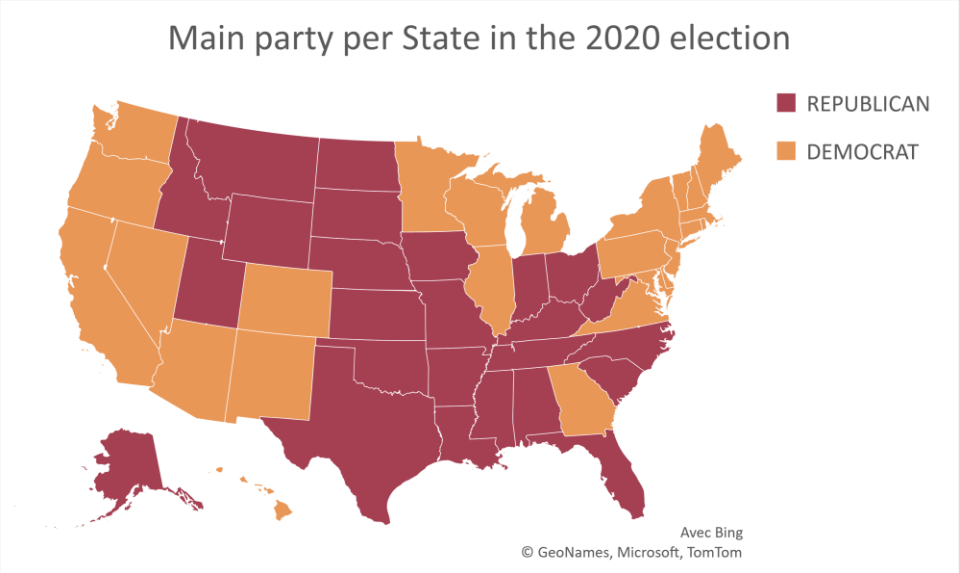
A US president is elected through the Electoral College every four years

EXPLORATORY DATA ANALYSIS

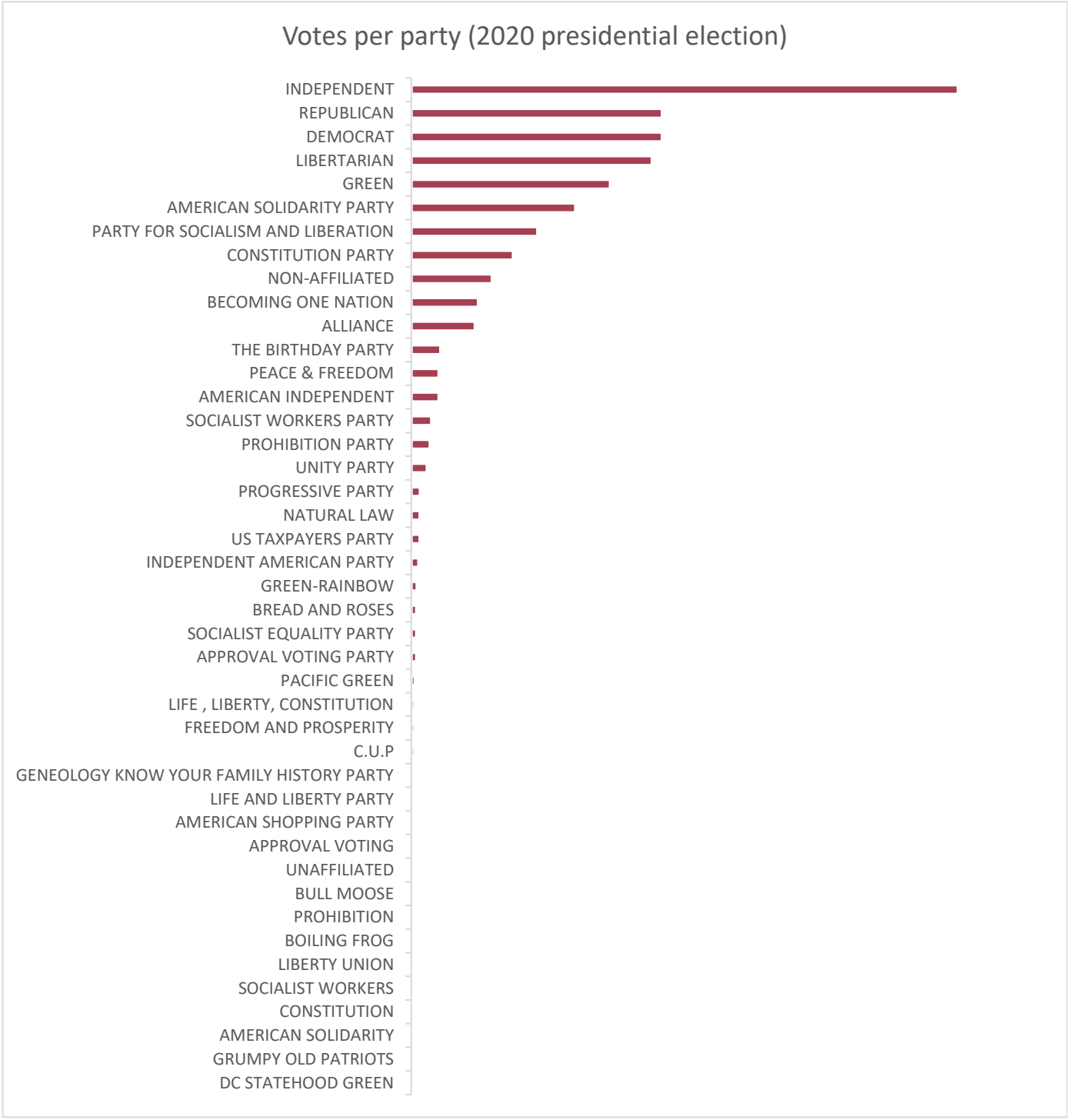
All the plots have been done in Excel.
We can see that the number of candidates has drastically increased in 2020. This strange behavior might be explained by the chaotic end of the Trump presidency.



Now let's focus on the 2020 election. Below is a map of the most popular parties per State.



Looking at the State level, we can see that, despite what we might believe, there are many more than two political parties in the US presidency.



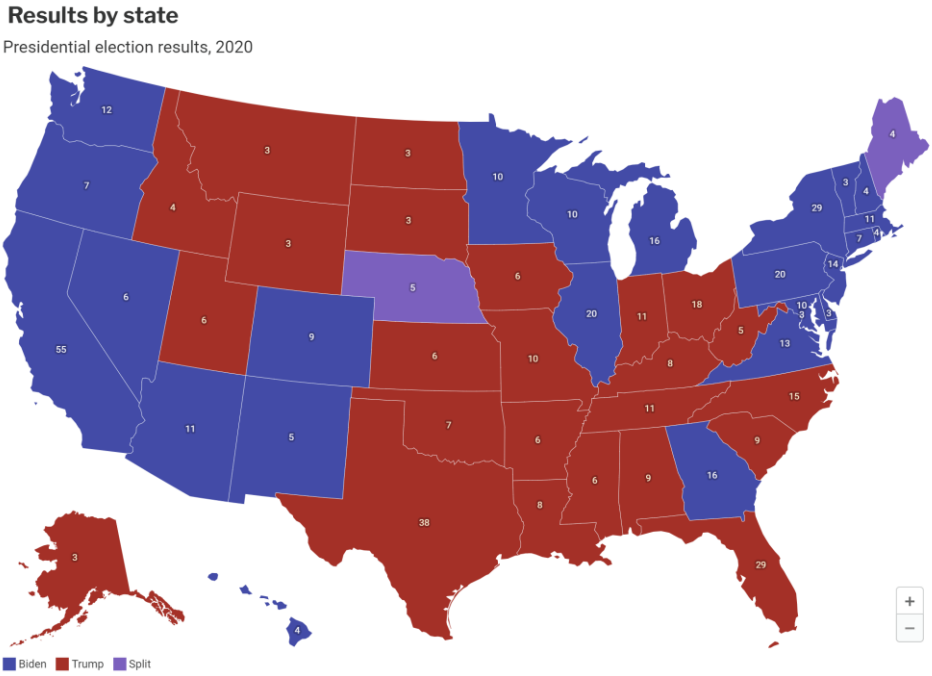
RELATED WORK

There exist many educational websites covering the US presidencies. An example is the official website https://www.whitehouse.gov/about-the-white-house/presidents/ which presents a chronological list of the US presidents, with some biographical information for each.

An example of related website tackling the same subject, from the whitehouse official page

What these websites do not provide is a more detailed view for the elections. We want our website to answer questions like "*Who were the other candidates and how close were they from the winner?*" or "*What states were for which candidate".*

This is covered in some other interactive tools, such as the website *ballotpedia.org.* However, these websites do not provide a visualization for the time evolutions of the dynamics. Instead, they focus on a per-election visibility.



An example of visualization proposed by ballotpedia.org

Our ambition is to propose both a time evolution understanding as well as a geographical visualization.