Data Visualization (COM-480)

Project Datart

Milestone 3 - Process Book

Abdeljalil MOUSSA (Sc. 250755)
Jeremy BENSOUSSAN (Sc. 269502)

Introduction

Overview

In 2020, accessing objective and understandable political information almost seems impossible. With social medias, fake news and biased facts spread even more easily and voters seem to be more and more influenced when making a political choice during elections. Moreover, access to correct and objective information can be dependant on the social background when it should be available to everyone.

In France, election results can drastically change from one region to another and also seem to be changing over time. Indeed, election outcomes are often influenced by the socio-economic situation of the voters as well as the important events which recently occurred. While political mechanisms can be tricky to understand, data visualisation helps to bring out the essential information and can easily show the results of complex data analysis.

Using data visualisation to put forward complex political analysis helps to make the election results more understandable for everyone and therefore help voters reflect on their own choices and the underlying reasons behind them.

Motivation

"Opinion is something intermediary between knowledge and ignorance" - Republic, Plato

This quote is at the basis of our work. Indeed, the main motivation behind our project is to make objective political knowledge more accessible and erase different preconceived pieces of information. As objective and understandable political information is becoming more and more difficult to find, this project aims at presenting different visualisations giving key insights on the voters' profiles. The data used are official data from *INSEE* (French National Institute of Statistics and Economics Studies) and *data.gouv.fr* and therefore shall not be discussed.

The goal of the project is to conduct an in-depth analysis of the impact of different elements on the results of elections in France. We studied the French Presidential, Legislative and Cantonal elections between 2002 and 2017. They are the three main elections in France and namely elect the President, the National Assembly representatives and the departments (sub-regions) representatives.

The visualisation tools chosen aim at showing the links between some elements and the election outcomes. To this matter, we used as elements the major events which occurred in the country in the studied time period but also several socio-economic indicators.

Even though the visualisations will be showing some correlations, the goal is not to influence the viewer in any way but for him or her to build his or her own opinion and knowledge based on the objective and easy to understand information we bring.

Target Audience

We wanted to make the visualisations easy to understand for all our target audience. There is no need for political knowledge to understand our work and the visualisations were made playful and practicals.

We believe voters might find interest in our work as it will help them better understand the mechanisms and reasons behind the votes and therefore help them be more aware of their own choices. The goal here is to reach out to any voters regardless of their education or social background. Indeed, we believe that understandable and objective political data should be available to everyone.

As there is much more to explore, some people interested in data visualisations might find interest in our work and want to pursue research on other elements which could have an impact on votes and find new correlations.

Discovering our path

Data Exploration

At the beginning of our project, we wanted to do our analysis on two countries : France and the United-Kingdom (in order to benchmark them). However, after getting some feedback on the first milestone, we understood that it was better to focus on only one country (France) and to have more metrics describing the population's features [1] as well as several elections (presidential, legislative and cantonale) [2].
In this regard, we discussed our data preprocessing in the previous milestones, but the final datasets used for this projects are as follows :

- Presidential Elections results between 1965 and 2012 per department
- Legislative Elections between 1958 and 2012 per department
- Cantonal Elections between 1988 and 2011 per department
- Mean Revenue/Taxes per region between 1994 and 2015

Before, we also processed metrics such as gender and age [3], however we did not find much correlation between these factors, and some voting groups.

Inspiration and related work

The mechanisms behind election results is an issue that has inspired many other researchers. Before we started to work on our project, we investigated some data visualisations work which had been done on the subject. You will find below some of our main inspirations :

- In 2017, **The New York Times** created a set of data visualisations illustrating the results of the 2017 French Presidential election. The study, called *"How France voted"* [4]. The visualisations show the election outcome [5] and turnout percentage by department and contextualize the results with an analysis of unemployment rate.
- The Political Visualisations done by **"le Monde"** [6] allowed us to truly understand their journalistic vision and how one can express political views and social influence in the most neutral way possible.
- Finally, the political visualisations done by **"FiveThirtyEight"** [7] allowed us to fully understand the importance of fun and interactive visualisations in order to make the viewer reflect on the message intended.

Visualizations Design

In this part, we would like to show you the evolution of our plots and visualisations, in order for you to better understand the magnitude of the changes we have made. We wanted to make these plots as interactive and informative as possible of course.

Horizontal Timeline plot

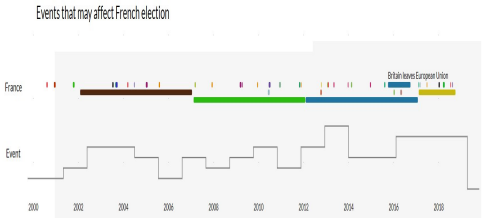

Figure 1: Timeline (before)

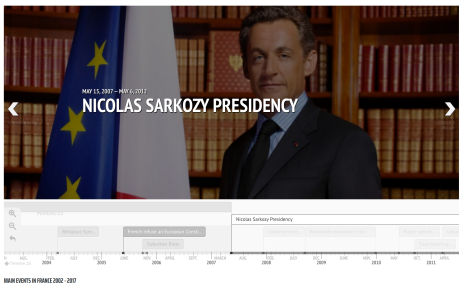

Figure 2: Timeline (after)

The basis of a population's ecosystem is the different events which affect the people at the national level. Therefore, we wanted to show the different major events which have occurred in France since 2002.
At first, as you can see above on the left, we wanted to create an horizontal timeline with colliding event, where you can click on them to be redirected to the event's Wikipedia page. However, we thought that this process was counter-intuitive because the user did not have access to the information directly so we chose a much more interactive plot, which has images. To this matter, we used a tool called *"Timeline JS"* which allowed us to build this plot. It is a lot more visual as you see above on the right. Moreover, the events are directly described in details and there is no need to visit another website.

French Map

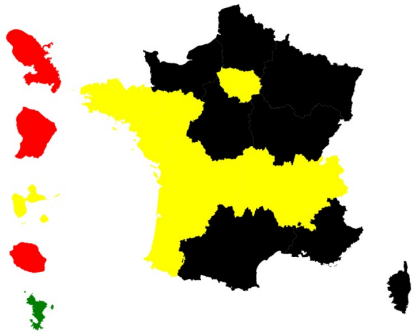

Figure 3: French Map (before)

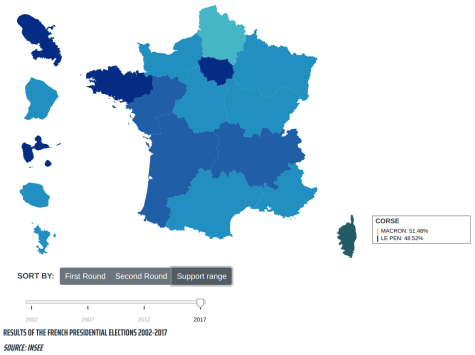

Figure 4: French Map (after)

Now that we know a little bit more about what happened in France since 2002, it remains necessary to inform about the French elections results. For this part, we decided to create a map of France with the results of the presidential elections (the most important election in France).

At first, the map was not interactive and the color mapping was not ideal. Hence, we came up with ideas in order to allow the viewer to directly interact with the map : we added a scrolling bar to select the year of the election and buttons to choose the election's round. We also used a new calculation method (we added votes percentage to distinguish certain regions), made the regions responsive to provide more information to the user, and we also chose a more appropriate color palette which better matches with our website's color code.

Statistical Plot

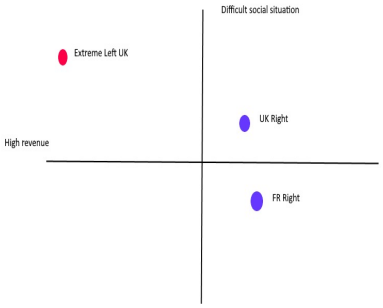

Figure 5: Statistical Plot (before)

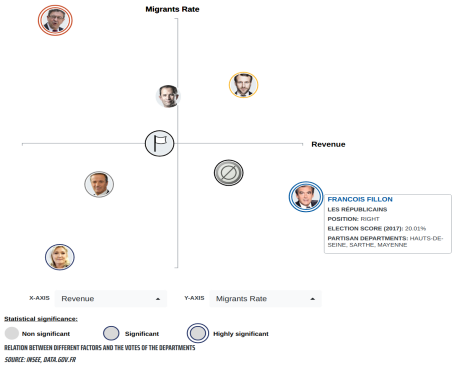

Figure 6: Statistical Plot (after)

Now that we have learned how the different regions of France vote, it is time to explore the population's ecosystem. We wanted to plot the correlation between the voters' profiles and the 2017 Presidential election results. To do so, we extracted the profiles of the voters registered in the different departments in 2016. We did some statistical analysis to determine for example whether a department supporting in majority the right wing had a population with a higher mean salary.

At first, this statistical plot was hard to understand because it contained only static points with little information. However, we enhanced this visualization by adding some interesting elements : pictures of the parties' candidates, responsive points with more information about the candidate and finally drop-down menus to select the region's population features. Statistical significance was added to be as objective as possible and avoid stereotypes.

Responsive Scatter Plot

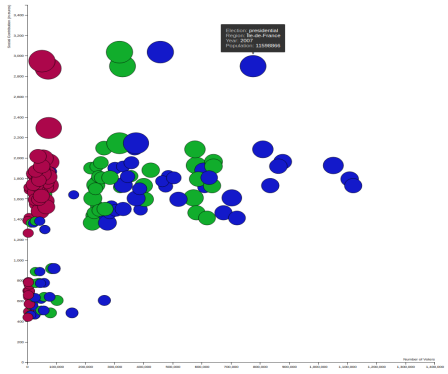

Figure 7: Responsive Scatter-Plot (before)

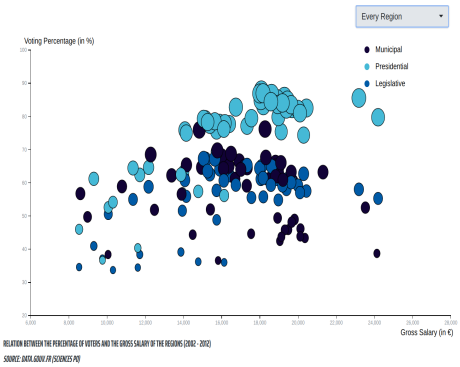

Figure 8: Responsive Scatter-Plot (after)

Finally, in here, we implemented one of our extra ideas from the milestone 2, and that we thought was a nice addition to our data story. The core component of this plot is the voting percentage (y-axis) depending on the gross salary of the population (x-axis).

At first, this plot was not easily understandable because it contained too much information (gross salary, population count, year, region, election, social contribution) and the color palette did not make sense because we did not have any legend.
However, after having receiving some feedback on this plot, we made it more intuitive by adding several components : drop-down menu for selecting a region, focus on the gross salary and voting percentage (number of voters over the number of people registered) with radius-sensitive points, more appropriate color palette which matches the website's color code as well as a legend.

In addition to that, we wanted to create more fun visualisations around the topic of gross salary such as a bar race (with a brush function) between regions, or even a bubble chart with each bubble for a region and a color for the type of election. However, we gave up these ideas presented in the last milestone because we considered that they were too basic, and also the informative bonus was not sufficient to make the viewer see our main objective : the correlation between election results and socio-economical status.

Project Implementation

In this part, we will discuss the technical implementation of our project (code implementation), by going through the challenges that we faced when we built our website and visualisations.

Code Architecture

First of all, you will find below our code architecture, this will help you have a better understanding on how we broke down the structure of our website. Indeed, our code is divided into main topics, and consequently we divided our website (and project in general) into these topics :

```
datart
|   README.md
|
└──dataset
|   └──French_Election/
|   └──all_elections_departments/
|   └──population-france/
|
└──Visualisations
|   └──Dimensionality/
|   └──French_Election/
|   └──french-population
|       └──bar-race-salaries/
|       └──bubble-salaries/
|       └──responsive-scatter
|               └──gender/
|               └──salary/
|
└──Milestone-reports
|   └──milestone1.md
|       milestone2.md
|       milestone3.md
|       Report_Milestone_2_Datart.pdf
|       Process_Book_Datart.pdf
|
```

Figure 9: Project Architecture - Datart

For all our visualisations, we mainly focused on using **D3.js** in its last version (v5), and this was clearly a challenge because we had to learn the latest practices with this newest version of D3. However it was a nice experience because we could also use the newest features such as promises and asynchronous data loading. Moreover, we had the opportunity to use several visualisation tools we discovered lately such as :

- bootstrap and Jquery
- bootstrap-select
- "Iwanthue" color palette picker
- Timeline JS [8]

Website and interactive elements

For the website implementation, we started by taking a basic template from *github.io* and then we tried to make it personal by starting with drawing some sketches of what we wanted. Then we came up with a skeleton with our main topics :

- INTRODUCTION & CONCLUSION
- HISTORICAL TIMELINE
- VOTER PROFILES
- VOTES & SALARIES

Following that, we had to come up with features which would distinguish our website from other data stories. Hence, we first created a common color code, which will set the tone for all our visualisations. Also, we added a side-menu (on the right) in order to have an easier navigation for the viewer.
As a final touch, we introduced a website logo, with an ending gif making the website much appealing in general.

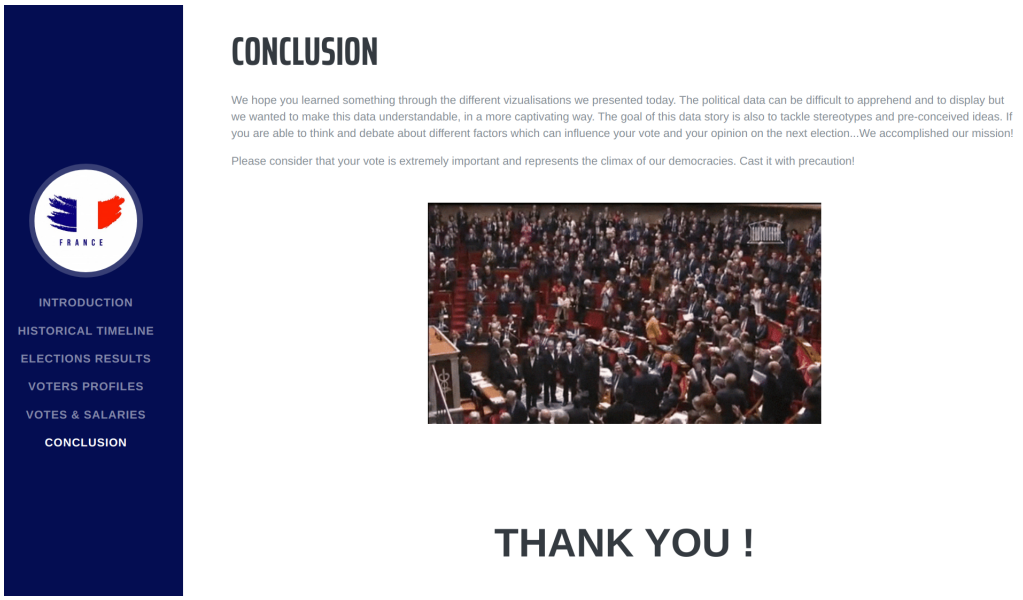

Figure 10: Website Implementation Example - Datart

Conclusion

Finally, this project was our opportunity to display our project's goal : tackle stereotypes and pre-conceived ideas on the French Elections. Throughout our visualizations, the goal was always to show the correlation between environment and political decisions and votes.

Regarding the technical implementation, we could add further improvements such as diving more deeply into the interaction between us and the viewer. For example, if the viewer comes from our target audience (French citizen or voter), we can create his/her anonymous profile into an artificial database, and give to this person a voting prediction, based on his/her personal features previously studied in this project.

Peer Assessment

In general, we had a pretty good distribution of the tasks and the workload. During the whole process, we always helped each other when needed by correcting each other's code or even sharing some experience learned from the lectures and exercise sessions.
Also, we were meeting every week, and we would set up goals for the following weeks every time we met. Consequently, the implementation of the visualisations have been split evenly.
Regarding the last milestone, Jeremy was in charge of the last changes on the website, and the creation of the screencast. Abdeljalil was in charge of the process book, and the refactoring/-cleaning of the project's repository.

References

[1] INSEE, "Revenu disponible brut des ménages par région de 1994 à 2015." https://www.insee.fr/fr/statistiques/2043717#consulter. (accessed: 28.05.2020).

[2] data-gouv-fr & Ministry of the Interior, "Les données des élections." https://www.data.gouv.fr/fr/posts/les-donnees-des-elections/. (accessed: 28.05.2020).

[3] INSEE, "Estimation de la population au 1[U+1D49]r janvier 2020." https://www.insee.fr/fr/statistiques/1893198#consulter. (accessed: 28.05.2020).

[4] grego AISCH & matthew BLOCH & rebecca LAI & benoit MORENN, "How france voted." https://www.nytimes.com/interactive/2017/05/07/world/europe/france-election-results-maps.html?_r=0. (accessed: 28.05.2020).

[5] karen YOURISH & adam PEARCE, "How large is the divide between red and blue america?." https://driven-by-data.net/2016/11/04/red-blue.html. (accessed: 28.05.2020).

[6] pierre BRETEAU, "Elections européennes 2019 : 6 cartes pour voir la progression ou la chute des principaux partis en france." https://www.lemonde.fr/les-decodeurs/article/2019/05/27/l-effondrement-de-lr-la-progression-d-eelv-et-la-stabilite-de-la-gauche-radic 5468011_4355770.html. (accessed: 28.05.2020).

[7] "Fivethiryeight politics." https://fivethirtyeight.com/politics/. (accessed: 28.05.2020).

[8] N. U. KNIGHT LAB, "Timeline js." https://timeline.knightlab.com/. (accessed: 28.05.2020).