

École Polytechnique Fédérale de Lausanne

Barcelona, a city of Data

A Data Visualization Project

Léopaul Boesinger, Marc Egli, Louis Perrier

01/06/2021

Process Book



CONTENTS

1	Birth of the project	2
1.1	The choice of Barcelona.....	2
1.2	A map as a tool of visualization.....	2
2	Three pages for three functionalities	3
2.1	A basic interactive map of Barcelona.....	3
2.2	Bar charts and choropleth	4
2.3	District comparison.....	5
3	Merging into one single page	6
4	Technical challenges	7
5	Peer assessment	8

1 BIRTH OF THE PROJECT

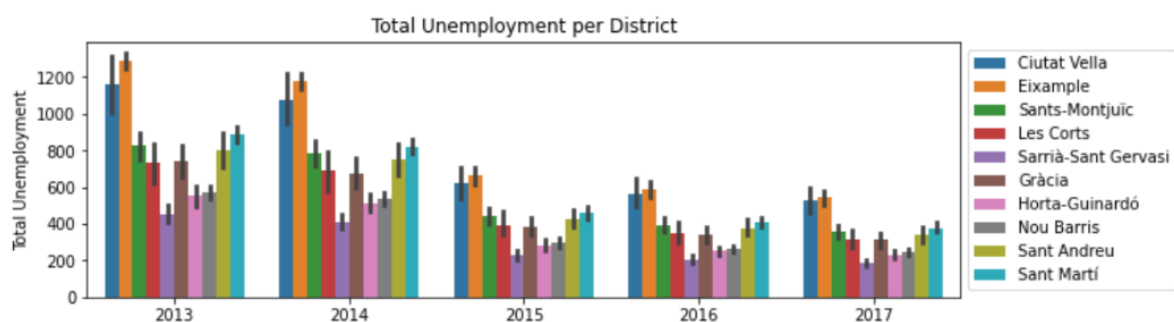
1.1 THE CHOICE OF BARCELONA

Large metropolis are always evolving. Their attractiveness makes people come or leave each year, requiring the city to adapt to all these changes. New residencies are built, the traffic management and the transport network must be updated... With these changes come new challenges, accidents may rise, air pollution can increase, requiring the administration of the city to adapt their measures to these constraints.

Of course all these evolutions involve data. Each change can be interpreted as a number increasing or decreasing, and comparing those numbers can be useful to understand the dynamism of a city. It can be especially helpful for the administration of the city, allowing them to analyze these values, predict their evolution and react to it. But for these numbers to be meaningful, they must be represented in a way that allows to highlight evolution, correlation or causality in a user-friendly manner. This is the point of data visualization, and the goal of this project.

Barcelona is one of the most dynamic cities in Europe. It's the second most populated city in Spain, is an economic center and the eighth most visited city on the continent. Its population is always evolving, and each of the ten districts of the city has its own dynamic. Therefore it was the ideal topic for our project, especially considering we found a large and clean dataset containing all the information we needed to achieve our goal.

After pre-processing some of the data we managed to build charts, like the one below about unemployment across districts and over the years.



We now had access to all the necessary data, but we still had to find the best way to display it.

1.2 A MAP AS A TOOL OF VISUALIZATION

Through our researches, we found several data analysis displaying statistics about Barcelona using different kinds of charts. However those were just presenting a list of graphs without a real

structure and were in the end quite difficult to read. We wanted to construct our website around a single, interactive, visualization to make it the most user-friendly possible.

Since we wanted a visualization that would be helpful to the administration of Barcelona, we thought a good way to represent the data was to separate it by districts. There are ten of those in Barcelona, and our dataset tells us for each data at which district it corresponds, so it was an information we could easily retrieve.



With this in mind, we decided that the main feature of our website would be a map of Barcelona where each district is clearly distinct. Such a representation would be beneficial both for readability and interactivity:

- Having a map would allow us to use visualization techniques such as choropleth, which is used to visualize how a variable varies across a geographic area in a more readable way than with a bar chart for example.
- It also makes the website more user-friendly in term of interactivity. Indeed, clicking on a district on the map is easier than accessing it through a list of hyperlinks.

2 THREE PAGES FOR THREE FUNCTIONALITIES

2.1 A BASIC INTERACTIVE MAP OF BARCELONA

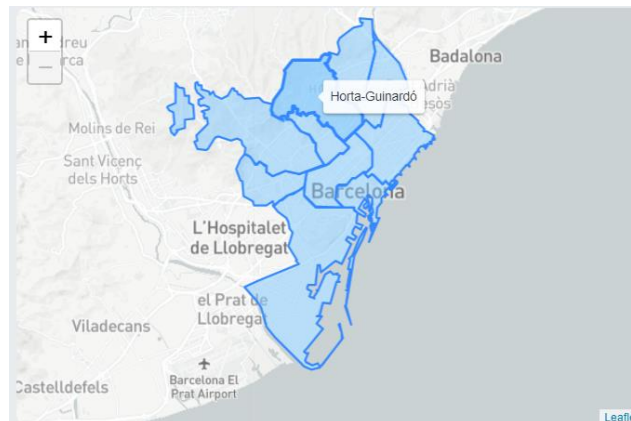
We started building our website with its central component: the map. To display the map and be able to draw polygons on it, we used the Leaflet package. We had to make sure that the map was

centered on Barcelona and adequately zoomed in. Furthermore we wanted the user to be able to zoom and drag the map, without leaving Barcelona or zooming out too much.

Once the map was properly displayed, we drew on it polygons corresponding to the ten districts of Barcelona, using the coordinates given in our dataset.

Of course we wanted our map to be interactive. The first step towards this goal was to be able to hover over districts. Doing so would make the polygon stand out and display the name of the district.

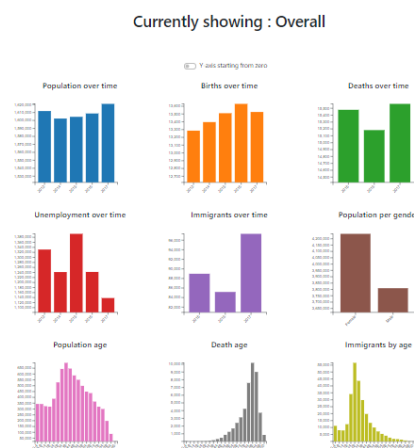
You can see the first draft of our visualization below.



2.2 BAR CHARTS AND CHOROPLETH

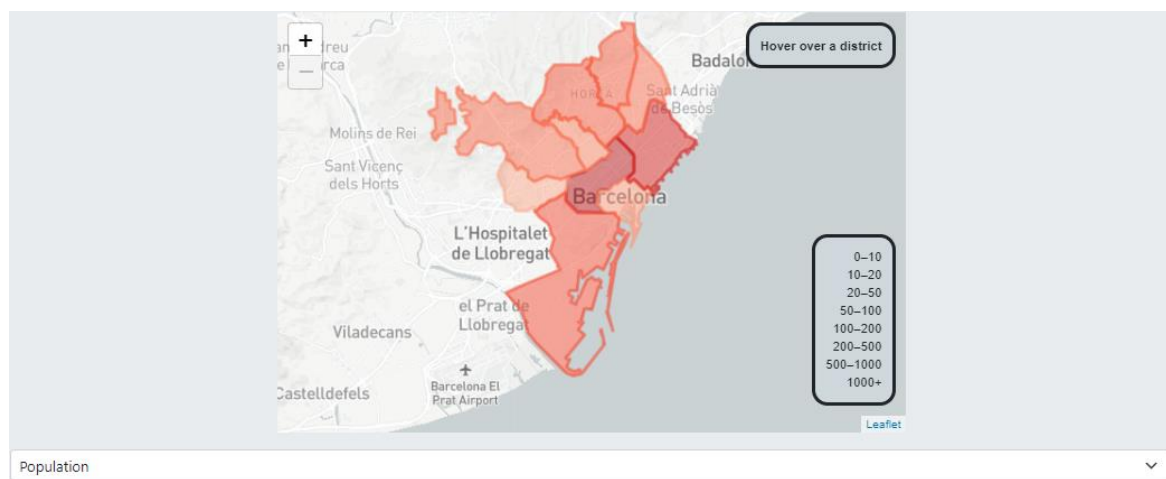
Now that we had a working map we finally had to add data to our web page. For data aggregation and manipulation, we used the Lodash library.

The first functionality we added to our site was a variety of bar charts showing the evolution of several statistics in Barcelona between the years 2013 and 2017: one for population, one for births, one for deaths, one for unemployment and one for immigrants. We also have one bar chart showing the distribution of population per gender and age, as well as one for the distribution of the ages of death and the age of immigrants.



On click on a certain district on the map, the graphs would be updated (in a smooth way) to display the same statistics but only within the selected district.

We then added a second page to our website which consisted of a choropleth map. On this page the user can select a statistic (population, birth, death or unemployment) as well as a year to visualize how this statistic differed between districts during this year. Each of the ten polygons corresponding to the districts is filled in a different color depending on the value of the data. We chose to use shades of the same color (red in the first draft) to make it more readable to the user, as it is not always instinctive to understand which color corresponds to the biggest value.

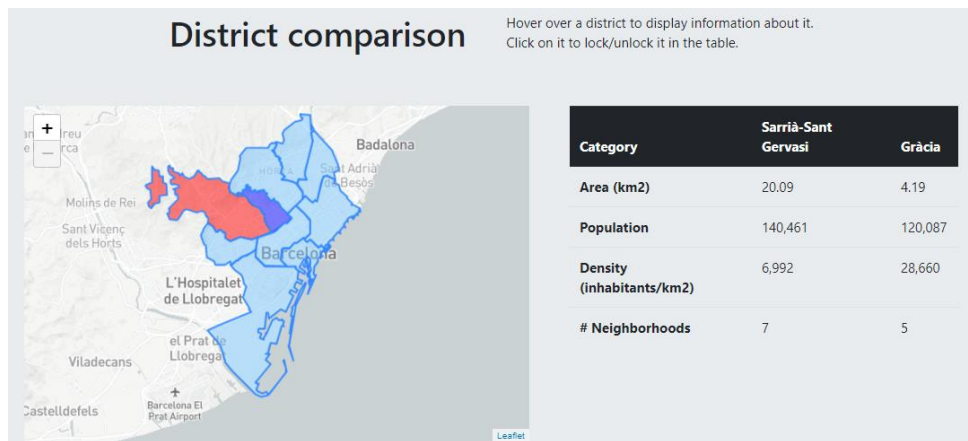


Such a visualization allows the user to easily compare data between districts, and be able to see if a particular district stands out compared to the others for a particular statistic. With only the bar charts this would be more difficult to see, as the user would have to switch between districts and check the values for each one in order to compare between them, making it way less convenient.

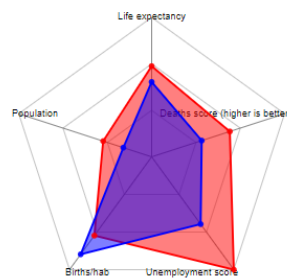
2.3 DISTRICT COMPARISON

The choropleth map is a good way to see how districts behaves regarding a certain statistic, however we wanted to be able to compare two districts with more insight on their differences.

We then added a third page to our website called District comparison. This page also consisted of an interactive map of Barcelona where it's possible to click on a district to select it, and afterwards click on another one that we want to compare with the first district. The page shows a table containing various information about the district area, population and density as well as the number of neighborhoods (subdivisions of a district).



Another feature of this page was a radar chart, or spider web chart. On this graphic, each district is given a score for several statistics. When two districts are selected, the score of both appear on the chart making it easy which one behaves the best in what statistic.

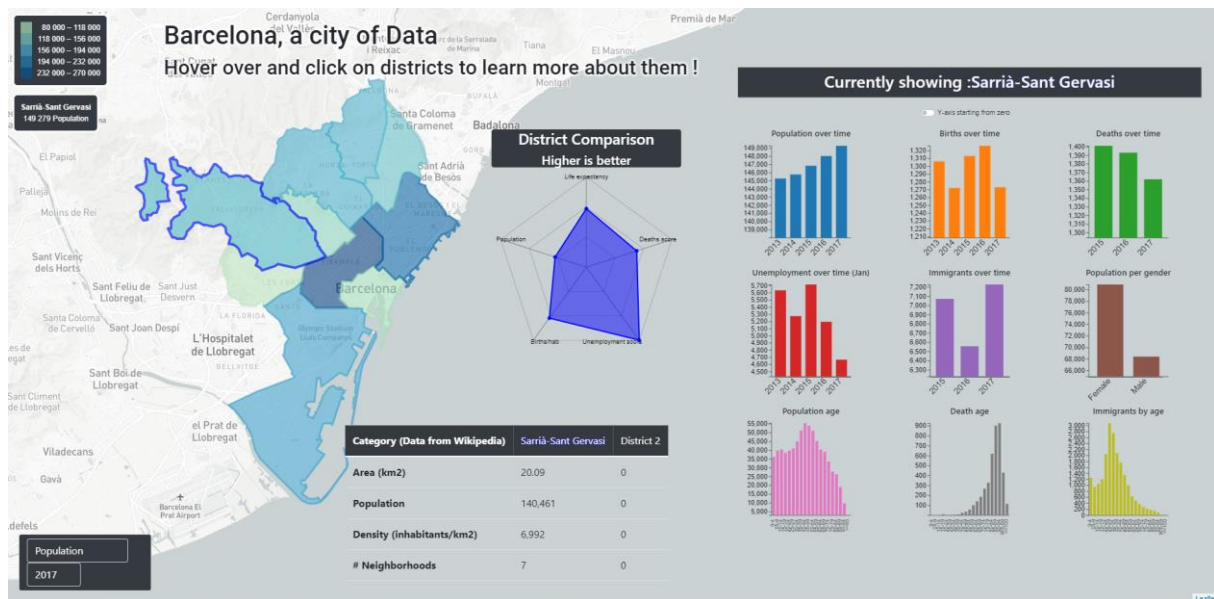


3 MERGING INTO ONE SINGLE PAGE

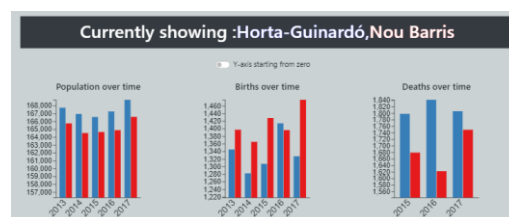
While we were satisfied about the three functionalities we created, we thought there was a better way to structure the website. Indeed our three pages all displayed the same map of Barcelona with its districts, so it would be even better to have one single page with a map and all the functionalities available to the user. It would make the website more practical to use, allowing the user to switch easily between two functionalities if they want more insight about a particular statistic.

We also wanted our page to be entirely visible on the screen of the user without them having to scroll down, which would mitigate the advantages of having one single page.

Managing to fit every functionality on the screen of the user while making our website readable was a challenge, but in the end we managed to make it in a satisfying way. Here is what the final version of the website looks like.



Having a single page could also allow us to adapt our bar charts when comparing two districts. When printing the statistics of two districts on the same bar charts, it's possible to check if a statistic behaves differently in time, or if the age or gender distribution isn't the same for those two districts.



With this done, we have achieved a satisfying visualization of the data of Barcelona. All what's left to do were minor changes to make it better, such as updating our choropleth map to show a statistic with respect to the population of the district, because showing the brute value was not giving much insight (for example it is obvious that the district that has the most deaths in a year is the most populated one, dividing the value by the number of inhabitants in the district allows to better understand this statistic).

4 TECHNICAL CHALLENGES

While working on the project, we faced several technical difficulties and had to find our way around those. Here is a non-exhaustive list of those challenges.

- TopoJson / Geojson: Firstly fetching the Geojson data was not as easy as we thought. We had to make use of deprecated APIs to obtain the one we currently use. Then using the data turned out to be hard as well: while for Leaflet we can simply give the path to a Geojson file and be done with it, we had to abandon some ideas like the anamorphic chart because the d3 library used was deprecated and could not handle the form of Topojson we had.

- Bar chart: Having a bar chart that transition from data of a single district to a comparison of two districts was painful to make. The d3 library needs big boilerplate code which wasn't easy to work with in this case.

- Responsiveness: This is another issue that we didn't completely fix: While we tried to make some elements change size with the screen size, it was simply not possible to make everything fit together with no overlap, but we figured that this is the case for many websites of the sort, which are meant to be viewed in full screen.

- Inconsistencies and coarse grained categories in dataset: While for some csv files we could simply aggregate by years and be done with it, multiple levels of granularities are given across files: for the unemployment files we are given monthly data, so we shouldn't aggregate over the whole year, sometimes we need to filter out district values because of "No data" elements, and sometimes totals are directly included in the dataset. We thus had to be very careful when integrating multiple files in a single visualization.

We also wanted to add more interactivity by allowing the user to get a boxplot with whiskers when comparing two districts (this would show the variation of population across the times) but since we only had yearly data for few years, creating a box plot with only 5 data points would not make for a very honest visualization.

5 PEER ASSESSMENT

Here are the main tasks performed during this project and how they were divided between team members:

- Research of the dataset : Léopaul, Marc and Louis
- Map : Léopaul and Louis
- Bar charts : Léopaul
- Choropleth : Marc and Louis
- District comparison : Léopaul
- Merging three pages into a single one : Léopaul and Marc
- Various adjustments and code cleaning : Marc
- Process book : Louis
- Screencast : Marc and Louis