Impacte de l'activitat del Banc dels Aliments a la província de Barcelona

Dashboard - Presentation and Visualization
Data Science Master Degree @ Universitat de Barcelona

Alejandro Castrelo
Gerard Marrugat
Eduard Ribas
Pilar Santolaria

Table of Contents

Dataset

The proposed dashboard for this task aims to give insight information about the impact and scope of the activity of the *Banc dels Aliments* (Food Bank), a non-profit NGO working for a fair distribution of excess food between needy people, in the Barcelona region.

To do so, several sources of information have been used:

One of the datasets provided for the task (*foodpandas.csv*), which contains the data related with the activity of Banc dels Aliments Foundation in the province of Barcelona. Each row of the dataset contains information about the quantity of each food category provided by year at municipality level, except for Barcelona city, whose information is splitted at district level. Also:

- According to the types of data that Altair recognizes, *foodpandas.csv* contains all of them: quantitative (*Quantitat*), ordinal (*IDComarca, IDMunicipi* and *IDMacrofamilia*), nominal (*NomComarca, NomMunicipi* and *NomMacrofamilia*) and temporal (*Any*).
- The range for the quantitative values is widely variable depending on the region and its population. This aspect, the issues derived from it and how we have dealt with them will be explained in the dashboard section.
- Dataset values are collected at a year detail. The temporal span starts at 2007 and ends at 2017.
- There is no information for some specific regions and time (like Bages, 2008), in which cases we set up the field *Quantitat* to zero.

Another dataset (*beneficiarisComarca.pkl*), also provided for the task, that contains the number of beneficiaries by region (within Barcelona province) and year:

- The data types included in *beneficiarisComarca.pkl* are quantitative (B*eneficiaris*), nominal (*NomComarca*) and temporal (*Any*).
- As happened with *foodpandas.csv,* quantitative value ranges in this dataset also depend on the region. We will discuss this aspect later.
- Dataset values are collected at a year detail. The temporal span starts at 2008 and ends at 2017.

An open dataset from the Institut d'Estadística de Catalunya, with details of the population by region and year: www.idescat.cat/pub/?id=aec&n=246&t=2018. This data is modified in order to obtain a csv file where the population is given at region-year level. According to the types of data that altair recognizes the population dataset contains quantitative (*Població*), nominal (*NomComarca*) and temporal (*Any*).

Dashboard design

There are different types of dashboards according to their main purpose, such as analytical, strategic or operational, among many others. We consider this dashboard to be an analytical dashboard, whose aim is to provide the user with the most relevant information at a glance. Thus, the goal of this dashboard is to provide general key metrics and trends, so that the user can analyse further afterwards and take data driven decisions.

In order to create the dashboard, we started by individually studying the data obtained through the datasets mentioned in the previous section. Once we knew the different types of data, the relevant information and how we could aggregate and combine the datasets, we started to plan the dashboard.

In a brainstorming session, we discussed the main metrics that should be included in the dashboard and the best graphs to show them. Afterwards, we designed a first draft version of what the dashboard would look like, which can be seen below.
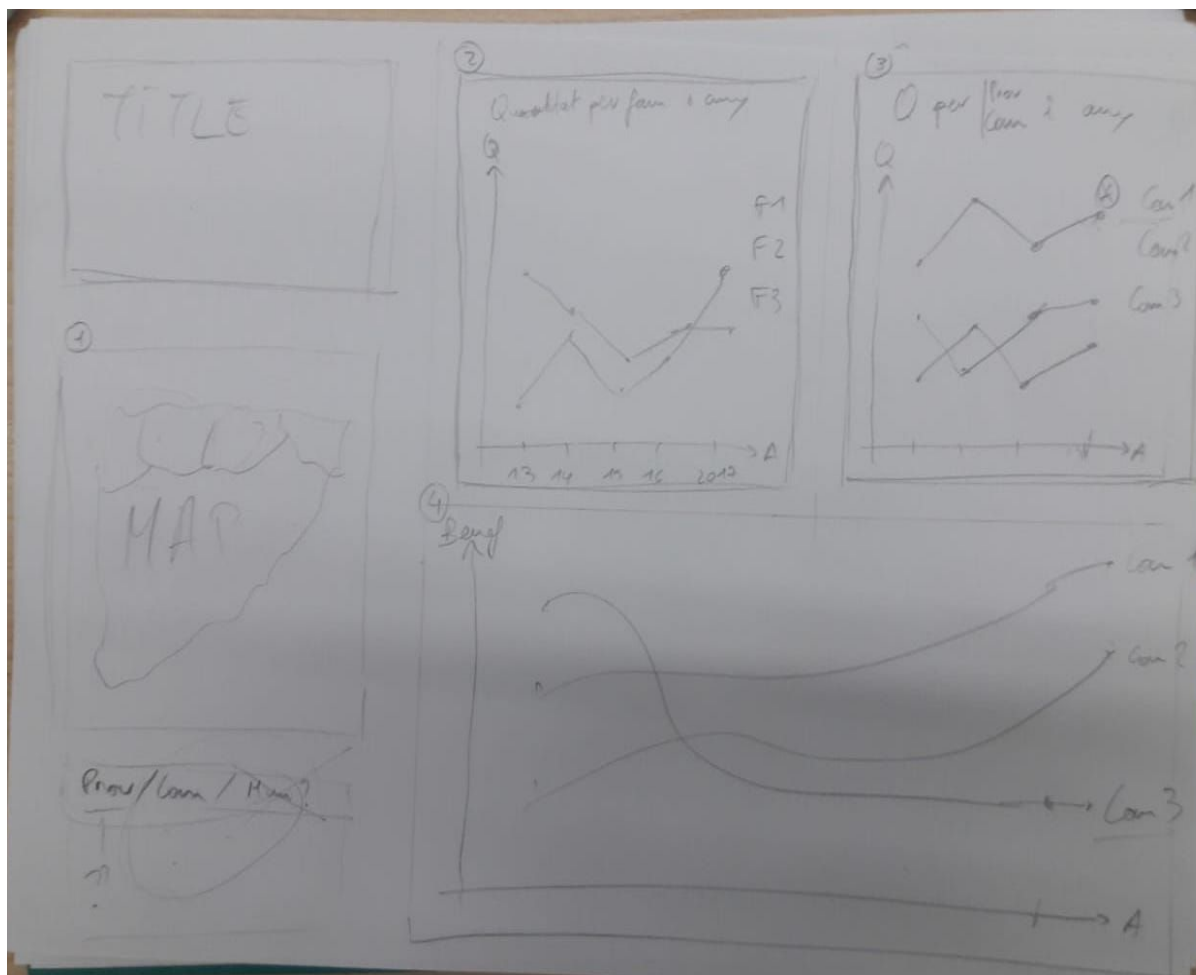

Fig. 1: First iteration of the dashboard

As it can be seen in the picture, on the left-side of the dashboard we decided to include a map of Catalunya with the different regions of the province of Barcelona in colors. The reason was to facilitate the contextualization of the rest of the metrics of the dashboard. We wanted that anyone who take a quick look at the dashboard could easily understand that the metrics they see are grouped by or related in some way to the regions seen on the map. At the end, geographical data displayed as a map is easier to understand and contextualize.

We also decided to put it on the left because, since many languages read and write from left to right, the left-side of the screen is considered to be the place where people would frequently look at and where they would normally expect to see general information. It is believed that users who are scanning for content in a first look at the screen, tend to gravitate toward the left side of the screen. This is why in many websites, this space is left for the menu or filters, but in our case, we wanted to prioritize the map that contextualizes the relevant metrics on which the user is going to focus afterwards.

Regarding the rest of the space, we decided the three main metrics that we thought a general user would need to see at-a-glance level: quantity of food by year aggregated by food category, quantity by year aggregated by region, and beneficiary by year aggregated by region. In the following paragraphs, these three metrics will be explain in greater detail. Our first initial approach included some applicable filter or clickable regions of the map (on the bottom left corner), so that the second level of the dashboard showed the metrics at a municipality level based on the specific interest of the user.

However, after having a first version of the dashboard, we evaluated the general overview in order to see if the at-a-glance overview gave all the general information that we needed to communicate. We realised that there were things that could be improved, and we designed a final version of the dashboard (draft attached below).
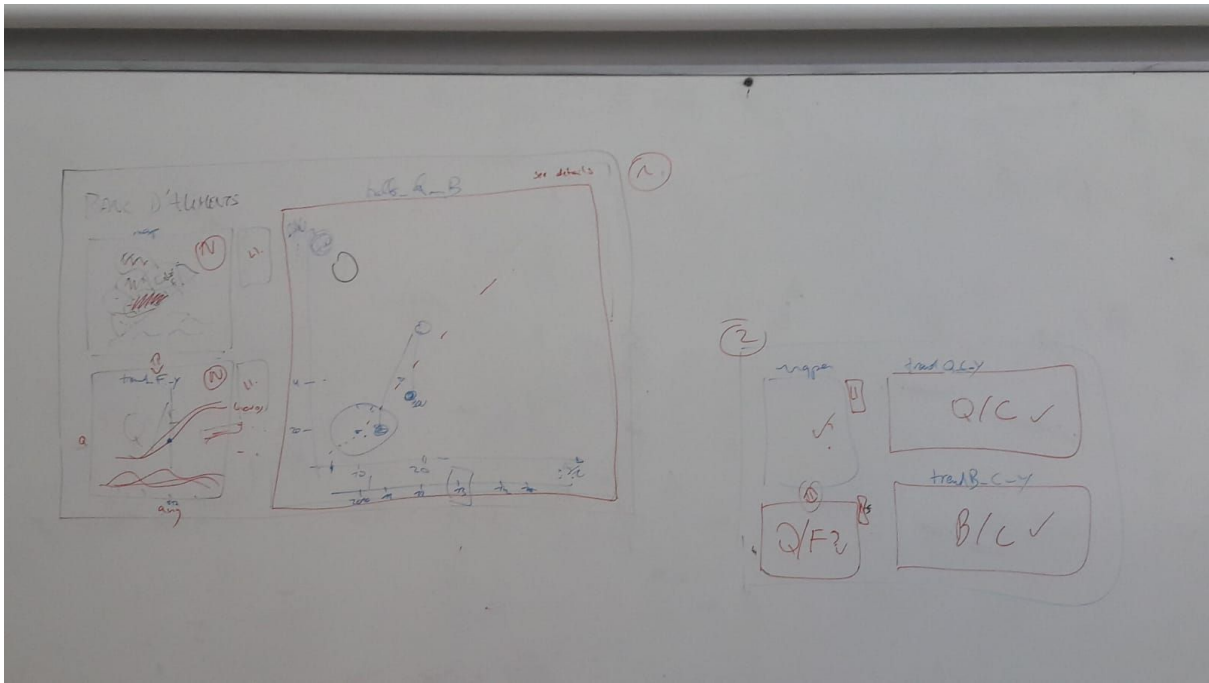
Fig. 2: Final iteration of the dashboard

As the pictures show, we decided to visually separate the graph that shows the quantity by year aggregated by food category (down left corner). We separated it from the other two metrics, since they were aggregated by region and aimed to show different data trends (beneficiary and quantity trends per region). The reason to do so was that it also helps to contextualize the other metrics and it gives the user a clear idea of what categories of food are the datasets taking into account and the general trend of quantity per food category over the years.

In addition to this, we decided to add a two-way interaction between the quantity by year aggregated by food category line chart and the map. This also supports the idea of having actionable parameters gathered on the left-side of the screen. Moreover, the interaction allows the user to see much more information personalised based on their personal interests (as they select or deselect fields) without changing the screen or using more space.

One way of interacting between both graphs is to click on the legend. With a Click (or SHIFT+Click for multiple selection) the user is able to filter the graph and map by the given food category, in order to see a less cramped graph.

On the other hand, by clicking on any year in the trend, the user changes the map to show the top food categories by region at this year.

On the right side of the main level there is a bubble chart whose aim is to provide a relation between the number of beneficiaries and the amount of food received by region

and year. These quantities are shown by each 1000 inhabitants: the values of beneficiaries and amount of food have been normalized according to the population of each region and multiplied by 1000. That way, the normalized amounts become comparable between regions. Each ball represents a region, and the size of the ball is proportional to the population of the region. Then, in this graph the user is able to see a lot of information in a quick view; how big is a population, how many of these inhabitants benefits from *banc dels aliments* and how much food they receive. Using the slider the year selected can be changed and an evolution of the relation between beneficiaries and food amount over the time can be observed. Before any interaction all the values for all the years are shown, that is why so many bubbles appear at the initialization, this is because *altair.selection_singles* does not allow an initial value for the slider.

Regarding the second level of the dashboard, we decided to keep the left-side of the dashboard as it was on the first level, and only substitute the bubble chart by two more detailed metrics. We decided that the bubble chart gave a general overview of the data and its overall trend over the years, but a user consumer of that data would probably need more detail on the charts as regards beneficiaries and quantity per region.

Thus, these two line charts were added on the second level in order to provide the user with a more specific information per region, since important conclusions and decisions can be taken after studying these two metrics, such as analyse the trends for each region, where the need is greater, in which regions people is more willing to help, in which regions should they invest their advertising campaigns in order to be more successful, etc.

Thus, it can be said that the final version consists of a two-level interactive dashboard that includes the following metrics distributed in two levels:

Level 1:
- Map of Catalunya
- Line chart: Quantity by year aggregated by food category
- Bubble chart: Beneficiary against Quantity by region filtered by year. The third dimension (bubble size) is the population of the region.

Level 2:
- Map of Catalunya
- Line chart: Quantity by year aggregated by food category
- Line chart: Quantity by year aggregated by region
- Line chart: Beneficiary by year aggregated by region

Most of the graphs are line trends, as we find them the simplest, most intuitive way to show and interpret data changing along periods of time. Also, as several categories might appear on each of them, we discarded considering bar charts.

Code and execution

The attached files contain all the code needed to replicate the computations and run the dashboard. The folders are:
- <u>data</u>: which contains all the datasets considered in the dashboard
- <u>library</u>: with a set of functions to load and process the data
- <u>notebooks</u>: jupyter notebooks to perform most of the data processing. These notebooks generate the final json files that are used to build the dashboard.
- <u>web</u>: with all the resources to host and launch the dashboard web application

To execute the dashboard, the following command needs to be executed from the command line (UNIX), within the "web" folder:

./run

Or similarly:

python -m http.server 8000

Then, the dashboard is accessible from a web browser on:

<u>http://localhost:8000/</u>

Distributions of task and workload

The dashboard creation has been executed in a extremely close collaboration between all the members in the team, however implementation tasks have been assigned individually in order to accomplish with the delivery date.

Regarding the decision about the design of the dashboard; how it should look like, what should be in each level, which information should be in each graph and how the user could interact with the graphs, was taken in group.

In relation to technical implementation, the tasks were divided as follows:

Alejandro Castrelo:

        Line chart: Beneficiary by year aggregated by region

        Web layout

Gerard Marrugat:

        Line chart: Quantity by year aggregated by region

        Bubble chart: Beneficiary against Quantity by region filtered by year

        Report documentation

Eduard Ribas:

        Map of Catalunya

        Two-way interaction between the map and the line chart

        Structure of the general functions to process datasets uniformly

        Report documentation

Pilar Santolaria:

        Line chart: Quantity by year aggregated by food category

        Collaboration with the interaction in the dashboard

        Report documentation

Finally, a final revision of all the work done, from the dashboard to the report, as well as the documentation and description of all components and aspects, was done all together by all members of the team.