UNIVERSITAT DE BARCELONA

FUNDAMENTALS OF DATA SCIENCE MASTER'S THESIS

Analysis and visualization of data from the Banc dels Aliments Foundation

---

Author:
Carla DIVÍ CUESTA

Supervisor:
Mireia RIBERA
Eloi PUERTAS

A thesis submitted in partial fulfillment of the requirements for the degree of MSc in Fundamentals of Data Science

in the

Facultat de Matemàtiques i Informàtica

September 2, 2018

UNIVERSITAT DE BARCELONA

Abstract

Facultat de Matemàtiques i Informàtica

MSc

Analysis and visualization of data from the Banc dels Aliments Foundation

by Carla DIVÍ CUESTA

Nowadays information overflows our brains, especially when you can receive any kind of information at any time. With this in mind, The Banc dels Aliments Foundation wants to find a way to make people understand the work that they do.

The aim of this project consists on finding the best way to design and develop a visualization for the Banc dels Aliments showing the maximum information about the food distributed during the last ten years in the province of Barcelona. Proving that with a good visualization the user can understand everything in a short time.

Acknowledgements

First of all, I would like to thank my supervisors Mireia Ribera and Eloi Puertas for considering me for this project, for their help and for all the hours they put in this project.

Second, I would like to thank my family, friends and boyfriend for their help and patience that they have always dedicated to me; without them, this would not have been possible.

Finally, I would like to thank my colleagues for their motivation and for all the good moments shared during this MSc, making it possible to enjoy and love the things that we do.

Contents

Chapter 1

Introduction

Nowadays information overflows our brains. You can receive information from many sources, such as television, radio or smartphones. The main problem of all the information that we received during the day is that not all of it will remain in our heads. This happens because there is a huge amount of information and sometimes is not well explained or we are just not interested in it.

For this reason, a good visualization of data can help to understand and process the information with a new view of the data. This can be achieved by taking advantage of the way the human perception system work by converting the information into a graphical representation.

A lot of companies, organizations and people that want to share some information decide to it do by graphic visualization, minimizing the amount of texts and tables. The Banc dels Aliments Foundation (Food Bank Foundation) is the organization that proposed to us creating a new visualization to present their data.

The Banc dels Aliments de Barcelona [1] is managed by an independent charitable foundation, apolitical, non-denominational and non-profit, the main objective of which is to recover food surpluses and to distribute them among local entities so that they can reach those in need.

This foundation needs to recruit more people to contribute to gather more food. For this reason, they decided to do a graphic visualization with its data history to show the evolution of the foundation and what has happened over the last 10 years in the province of Barcelona.

1.1 Motivation

The main problem that motivates this project is that there is no visualization for the Banc dels Aliments that clusters the information per zones and food.

Until now, the information was presented by graphics and text, but they had no proper visualizations. For that reason, this project will endow the Banc dels Aliments with a visualization of their data that will give the users more information about the places where more food is distributed, and the number of people who they are helping.

1.2 Objectives

The main objective of this project is to implement a visualization of the data of the Banc dels Aliments de Barcelona, which is compatible with their web system.

From this main objective we can extract some specific goals:

- To design a good visualization per location and food for the Banc dels Aliments that sensitize the people.
- To analyze the data so that they can improve on the organization.

The technical goals that this project should achieve are:

- Analysis of the final users to determine the requirements for the user interface (UI).
- Analysis of the data to study which type of visualization can and should be done.
- Implement the functionalities to show the data from location and type of food.

1.3 The Banc dels Aliments de Barcelona

The Banc dels Aliments de Barcelona has a management section on its website to show the transparency of the organization with data about their budget management, volunteers, associated entities and food distribution. They have a good and transparent visualization of this type of data, showing the evolution of the foundation over the years, because they generate a detailed report every year that describes all the inner workings, the amount of food that they have distributed and even some statistics and graphics about the food and volunteers.

The foundation has the opportunity to show to the people all their data and try to convince more volunteers to participate. If they manage that, they can help a larger group of people who do not have money to buy food for themselves. For that reason, they want to reach the broadest possible audience so they can raise more money and collect more resources to help as many people as possible.

The current information of Banc dels Aliments is like a newsletter, explaining the results of the year with some graphic elements, but it is not a visualization, it is strictly a report of the results. Some examples of the graphics can be seen below.

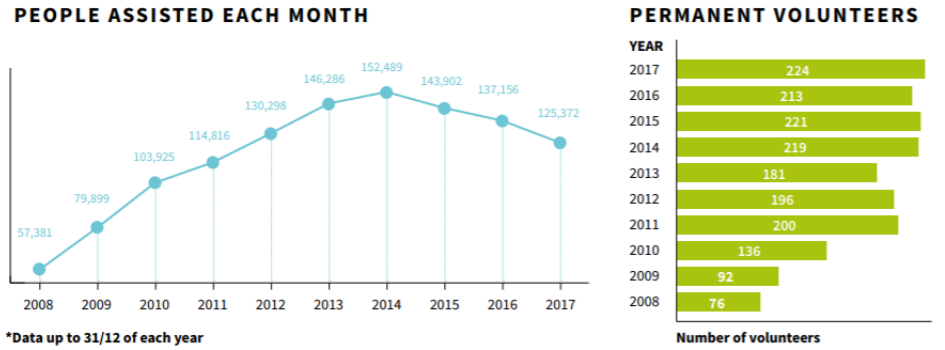

FIGURE 1.1: Example of current graphics of Banc dels Aliments: People assisted and volunteers

Figure 1.1 shows the evolution over the years of the people assisted by the foundation and of the volunteers working for it. This kind of graphics is good to show an evolution and to summarize the results in a static way.

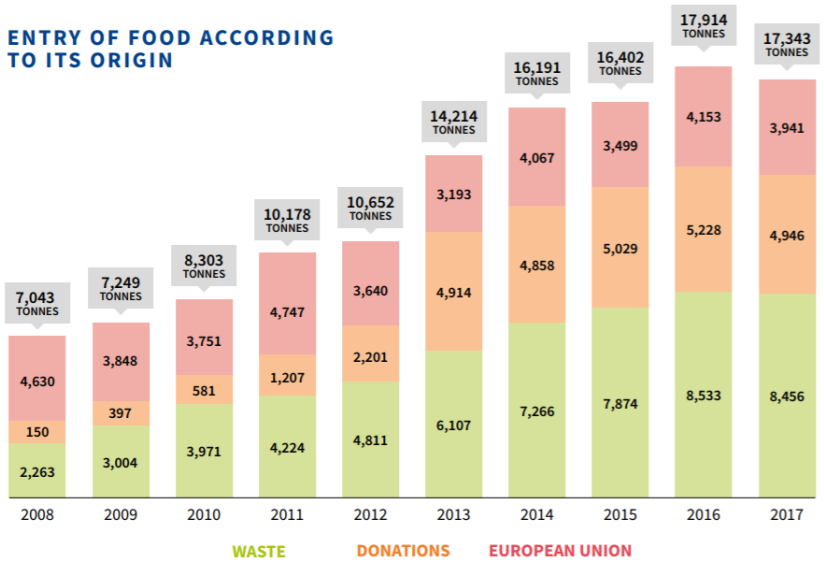

FIGURE 1.2: Example of a current graphic of Banc dels Aliments: Entry of food by origin

Another graphic that they had is the Figure 1.2, showing the origin of the food they work with.

These graphics are good to show some information about the foundation, but if we want to reach as many people as possible, we need to make our visualizations more appealing, while explaining to them how they could participate and what they could contribute with.

1.4 Visualization: A tool of communication

Visualization can be defined as the way *"to show information in a spatial or graphical representation, in order to facilitate comparison, pattern recognition, change detection, and other cognitive aptitudes of our visual system"*, quoting Marti Hearst, professor in the School of Information at the University of California, Berkeley.

A visualization is a great tool for communication. It has a huge potential to show information in an easy way because it takes advantage of the human perception system by converting information to a graphical representation. But is it always necessary?

A visualization might not always be necessary, depending on the context, the intended audience and the data. Sometimes, the data is easy to understand and is not necessary to present it using a visualization or maybe the intended audience are people that work in the field and thus visualization is not needed. However, when faced with new data, intended audiences or context, a visualization is a perfect tool to show this information and, if it is well done, it makes it self-explanatory. To present information when you want to sell, to explain or to understand is the best way to do because it used the visual potential of humans.

In our case, we will use visualization to transmit to the general public all the information about Banc dels Aliments Foundation.

1.4.1 State of the art

Before starting with our visualization, research was done to find other visualizations to take as an example as inspiration for the final one.

Hunger Heat Map

The first interesting visualization we would like to comment on is the Hunger Heat Map from Food Bank of Washington [2]. This visualization displays the hunger levels of the city to make people aware of it, because hunger is often invisible.

The Hunger Heat Map is a tool that visually represents data related to poverty and food insecurity, and the food distribution efforts that Capital Area Food Bank (CAFB) is leading across its service area. The four metrics we want to focus on in this map are: Food insecurity Rate, Food distribution, Pounds Per Food Insecure Person and Unmet Pounds.

The first map displays the food insecurity rate for each census tract in the Capital Area Food Bank's service area. Food Insecurity is measured annually by the US Census Bureau.

The second map shows the points in the city of Washington where one can find community partners. This map is for the people who want to participate in the programs of the Food Bank. These points are the food distribution locations.

The Pounds Per Food Insecure Person map displays the results of the food bank's successes in food distribution, which is a combination of the layer of need and the

layer of impact. This distribution is calculated by spreading out our pounds of food through each of our 444 partner locations.

The final map displays the metric Unmet Pounds. To calculate unmet need, we simply subtract our total food distributed from the estimated total food needed. This map displays areas of high unmet need, which the Capital Area Food Bank uses to plan and make programmatic decisions.

This type of visualization is good to empathize a bit more with the people in need since it makes easy to see the problems in each zone and to better understand them. The interactivity is always attractive for the users. In addition, people always want to see what type of zone they live in, and it is good to obtain more volunteers.

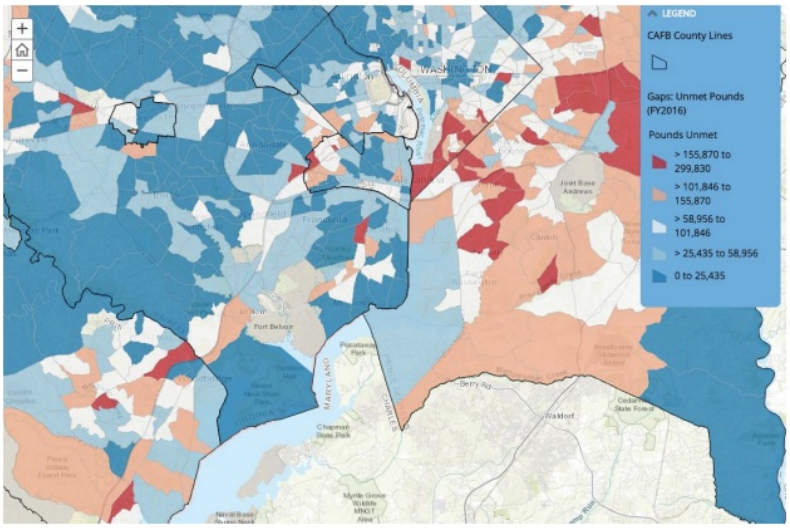

FIGURE 1.3: Hunger Heat Map

U.K. Energy Consumption Guide

Another interesting visualization for this project is the United Kingdom Energy Consumption Guide [3]. This visualization is not about the same subject as this project, but it is interesting for how they communicate the information.

The U.K. Energy Consumption Guide is a visualization that visually displays the data related of how energy consumption changes every ten years. The information is divided by percentages of fuel consumption (electricity, gas, petroleum and solid fuel) and by sector consumption (industry, domestic, transport and others).

This visualization explains a story with the graphic, helping the user to understand the evolution of the data itself. This is an example of how to design an attractive visualization and how to represent the data in a visual way. Also, the way that this graphic represents the timeline is a good practice to take into account.

Figure 1.4 shows an extract of the visualization.

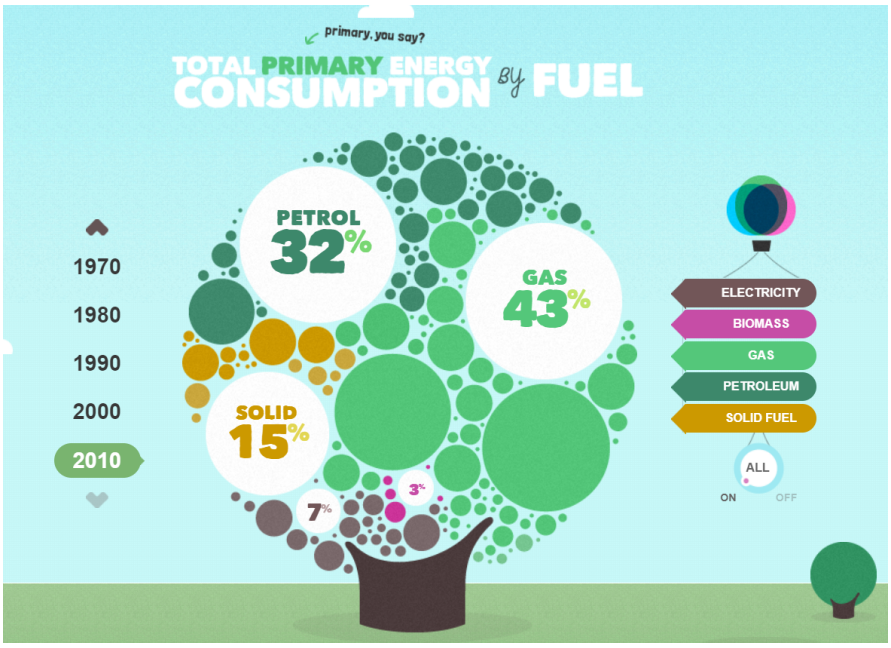

FIGURE 1.4: U.K. Energy Consumption Guide

Chapter 2

Analysis

In this chapter, the previous analysis of both the project and the data is explained in order to understand the situation and the reasons of the decisions that were made before the design process.

This project has two different type of stakeholders: Banc dels Aliments Foundation, an organization that benefits from the project, and final users, people that will watch and interact with the visualization. It is important to keep this in mind since the visualization should satisfy the requisites of Banc dels Aliments while maintaining a good ability of interaction for the final users.

2.1 Human-Computer Interaction

Human-computer interaction (HCI) is the research of how the user interacts with the machine interface. To do this research we should ask some questions about the different parts of the project: stakeholders (Banc dels Aliments and final users) and the data itself.

2.1.1 Banc dels Aliments Foundation

The foundation is the principal client of this project because it is the visualization of their own data, and they can exploit this visualization as much as they want. First of all, it was necessary to contact them and ask questions to understand which kind of visualization they need and determine what they want. The contact of Banc dels Aliments is Josep Maria Font, former professor of UB and volunteer of Banc dels Aliments. He is the direct contact and the person who met us during the development of this project.

The first meetup with Josep Maria Font was to ask all the necessary questions to start with the design of the project.

The first question was “Who are the final users?”, this question is the most important one because it is not the same to do a visualization for the employee of Banc dels Aliments or to other external public, that don’t knows anything about the foundation. He answered that they don’t need a visualization for the employee, they want a visualization to show on their website to the people that want to know about them. Another important question was “What do you want?”, this question sometimes is the most difficult one because normally they don’t know what they want or if they do, they don’t know how to explain it. So, we decided to make this question

different to catch more information, "Which are your objectives? Inform, sensitize, monitoring?" In this case, he answered that their objectives, with the implementation of the new visualization, were to inform and show the evolution of Banc dels Aliments over the years.

Other questions were "Which kind of data do they have?", "How will we do the deployment?" and "How does the data update?". These more technical questions will be answered during this document.

2.1.2 Final users

When the final users of the visualization are known, it is necessary to learn a little bit about them to understand which kind of users they are and which are they main interests in the visualization.

So, the questions that we want to answer are: Which is the age range? Race? Gender? Which are they interests? How much time will they spend in the visualization?

First of all, the only information that we know about the final users is that they will connect to the Banc dels Aliments website and then they can see directly this visualization. So, the range of age, race and gender are not relevant for this visualization because everybody with a computer, smartphone or other technology with internet can connect to the website and watch the visualization. They only need interest in the Foundation and want to know more about it's propose. Therefore, since Banc dels aliments works in Catalonia, we can suppose that the final user will be from there. In reference to how much time they will spend in the visualization, we couldn't know it but the visualization should attract them to maintain their interest and even increase their interest.

In summary, it supposes that final users will be:

- Age range: from 12 to 80. It is an approximation of the age of the people with Internet connection and interest.
- Race and gender: it is not relevant because any of them will alter the visualization.
- Demographic: the highest percentage will be from Catalonia.
- Interests: Foundation, help people and work with them.

Therefore, the visualization should take care of all the details that we know about the users. First of all, the most important thing is that our final users are not from an specific range of years, so the visualization should be easy for all the people that can see, and also the interaction should be well explained for the people that do not use the computer frequently. Another thing to take into account is the demographic part; a lot of people will be from Catalonia so it should close down to this demographic part. Finally, the interests of the people that will watch the visualization are helpful to understand what they want to know and how we catch their attention.

2.1.3 Data

This part solves the questions about data that it presents before. To refresh, the questions are: "Which kind of data do they have?", "Is the data yours or is from other community?" and "Can we exploit it directly?".

The data is from the Banc dels Aliments Foundation, but it has four different parts, one per each province. Each one of these province has their own management and organization. Our contact is from the province of Barcelona and he only has access to the data from this part because each province has their database and their own data. Therefore, as the data used is from the province of Barcelona, this visualization will be from the there though the public can be from the other parts of Catalonia.

About data exploits they decided to send us a copy of their data relating to food distribution and we can exploit as we wish.

About which kind of data will be explained in the next section.

After all the analysis from Banc dels Aliments Foundation, final users and data, we can have a first impression of which kind of visualization they want and what they need to visualize. The visualization needed to catch the attention of the public that visits their website and explains the evolution of their Foundation, more concretely in the province of Barcelona.

2.2 Analysis of the data

In this section, the analysis of the data is explained describing the data, range of the values and the format and transformation that was done.

As it tells in the previous section, the data came from the Banc dels Aliments Foundation of the province of Barcelona. They send us the data relating to a place, food and quantity.

2.2.1 Description of data

The data is divided into five tables: Entities, Orders, Lines, Foods and Beneficiaries.

Entities: has all information of entities, know entity as a place to distribute food. This table has the following columns:

- Id: primary key, represents the row. Integer variable.
- Receiving_entity: code represents the entity in the old database. Integer variable.
- Code_ent_banc: code represents the entity. Integer variable.
- Typology: type of entity. Categorical variable, the list of types: Entitats Receptores Ordinàries, Traspassos, Regulacions, Altres Banc dels Aliments, Altres Entitats Receptores.

- City: town where it is. Text variable.
- Address code: address code of the entity. Integer variable.
- Municipality: municipality that belongs to the town. Text variable.
- Region: region of the entity, territorially in Catalonia is a *comarca*. Text variable.
- Municipal district: if it is from Barcelona, it is necessary the neighborhood. Text variable.

An example of Entity: receiving entity of Ciutat Vella (Barcelona).

Receiving_entity	Code_ent_banc	Typology	City
1006032	21	Entitats Receptores Ordinàries	Barcelona

Address code	Municipality	Region	Municipal district
08002	Barcelona	Barcelonès	Ciutat Vella

Orders: information about each order of entity. It is like a ticket for a purchase. This table is general, the columns are:

- Id: primary key. Integer variable.
- Date: Date for when the entity receives the food. Date time variable. The range of this is from 2000 to 2017.
- Receiving_entity: Entity which receives the food. Id of Entity, foreign key.

An example of Order: Order of receiving entity of Ciutat Vella (Barcelona).

Id	Date	Receiving_entity
2195	2011-05-04	1006032

Lines: information about the food and quantity sent in an Order. If we compare Order like a ticket, this table contains each line of the ticket: food, quantity and price. The columns are:

- Id: primary key. Integer variable.
- Order: order belongs to the line. Id of order, foreign key.
- Food: which food. Id of food, foreign key.
- Quantity: quantity of kg of food send it. Double variable.
- Line: line in the Order. Integer variable.

An example of Line: Line from Order of receiving entity of Ciutat Vella (Barcelona).

Id	Order	Food	Quantity	Line
1896	2195	1075	20	1

Foods: information about food. The columns are:

- Id: Primary key. Integer variable.

- Code_food_banc: new code represents the food. Integer variable.
- Food: description of food. Text variable.
- Food_family: Family of food that belongs to. Categorical variable, the list of families: *Làctics i derivats; Farines i derivats; Arròs, pasta, sucre, llegums; Conserves i plats preparats; Olis i greixos; Carn,peix,embotits; Begudes i infusions; Fruites i verdures fresques*.

An example of Food:

Id	Code_food_banc	Food	Food_family
1128	98	LLET EN POLS	Làctics i derivats

Beneficiaries: information of the people attended in each entity per year. The columns are:

- Id: Primary key. Integer variable.
- Entity: entity where attend the people. Id of Entity, foreign key.
- Year: year of the people attended. Date variable, range of values from 2008 to 2017.
- People: number of people attended this entity and the year. Integer variable.

An example of Beneficiary:

Id	Entity	Year	People
14	100	2011	420

The data is well organized and easy to understand. The previous work done by our contact was helpful, and allowed us to have a simple database with only the needed data.

2.2.2 Format and transformation

The data from the Foundation has two different formats, depending on the year.

In their organization, they have an Access database from 2001 to 2016 and in 2016 they decided to change to a SQL database. Therefore, they sent the data in different formats.

Entities and Foods tables were sent in Excel format, they only sent us a copy because for all the years are more or less the same. But for the other tables the data was in different formats:

- Data from 2001 to 2016 (Access database): Orders and Lines tables in Access format.
- Data from 2017 (SQL database): Orders and Lines tables in a unique table in Excel format.

Beneficiaries table was not in their database as a table. They had the information about people attended in other site and they sent us as a document with Excel format.

Tables	Data from 2001 to 2016	Data from 2017
Entities	Excel format	
Orders	Acces database	Excel format
Lines	Acces database	Excel format
Foods	Excel format	
Beneficiaries	Excel format	

After seeing the different formats, it was necessary to decide a format to standardize all the data for later exploitation.

The decision was easy to take because seeing that the new database of the Foundation is a SQL database the best option was to do a SQL database saving all the data from 2001 to 2016. Then, it will be easy to standardize the new database and maybe take real time data.

To create a new database, we used a SQL database platform. For the import, the data with Excel was not a problem because the platform has an option for this type of imports and for Access database we had to use an application to transform from Access to SQL format.

Therefore, in the process to transform and standardize the data we encountered some problems. In the 2017 data new codes were found and it was necessary to create a new table to transform the codes. Also was needed some data cleaning.

2.2.3 Data cleaning

It is always necessary to be sure that the data is cleaned before starting to use it. The tables that were necessary to clean were:

- Lines table: exists some rows with a negative quantity that it had been deleted.
- Entities table: rows with entities from another province that will not be used for the visualization. They also have been deleted.

2.2.4 Transform data

To merge the data from the different databases it was necessary to do some transformations.

The data of Beneficiaries was transformed because each year was a new column and it was not scalable. Thus, we decided to transform each column of number of people attended per year in a unique column and add a column for the year. Moreover, reading the data of this table is easier with this transformation.

The table of Benefeiciaries data before applying any changes was:

Code_ent_banc	2008	2009	...	2017
21	532	620	...	368

After the changes is:

Code_ent_banc	Year	People
21	2008	532
21	2009	620
21	...	...
21	2017	368

The data from the first years was in two tables: Orders and Lines, and the data from 2017 was in a unique table: Orders_Lines that contains: date, entity, food and quantity. The main problem was in the relation of this tables with the Entities table, they had different codes and the foreign key is to a different column of entities. Thus, we decided to maintain both tables and to do a view with the union of these tables.

The tables before the changes were:

Data from 2001 to 2016

Orders		
Id	Date	Receiving_entity
2195	2011-05-04	1006032

Lines				
Id	Order	Food	Quantity	Line
1896	2195	1075	20	1

Data from 2017

Orders_Lines				
Id	Date	Code_ent_banc	Code_food_banc	Quantity
1896	2017-05-05	21	98	25

The codes of Entity and Food in the data from 2017 are different from the old data. Remember that in the tables of Entities and Foods we had two different codes. For that reason, we decided to do new views and merge these tables.

The new views will be the union of two different selects the data from 2001 to 2016 and the data from 2017. The terms selected are: Year, Zone, Food family of food and Quantity. There is a view per zone, to split the data and to make easy for the implementation, the reasons will be explained in Chapter 4, implementation.

An example of the view of the regions zone are:

Region_Order_view				
Id	Year	Region	Food_family	Quantity
1	2011	Barcelonès	Làctics i derivats	20
1111	2017	Barcelonès	Làctics i derivats	25

To a better understanding, Figure 2.1 shows the changes in Beneficiaries table and the zone views mentioned before Region_order_view and a view for each municipality.

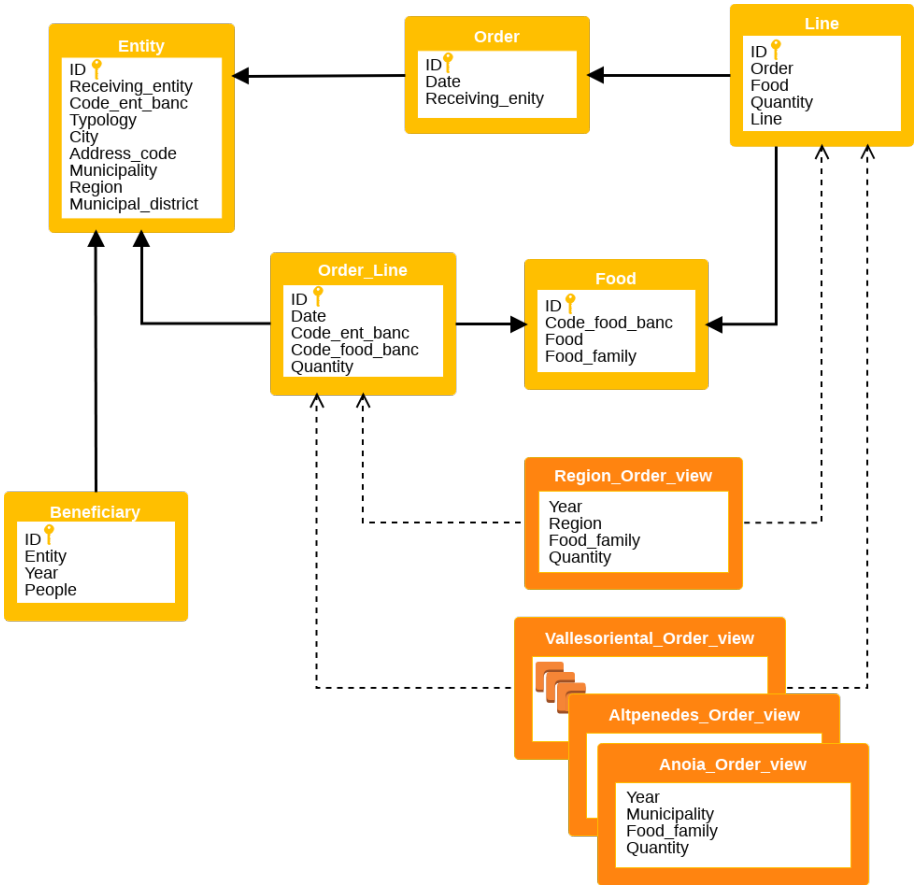

FIGURE 2.1: Database diagram

Chapter 3

Design

In this section is explained the design of the visualization. Starting from the macro level, followed by the micro level explanation. This chapter also illustrates the interactions of the design and the mock-ups done.

3.1 Macro-level

The macro-level design was referred to as the type of visualization chosen. After knowing the data, the final users and the visualization proposal, it is necessary to remember the different types of existing visualizations before choosing one. There are so many types of visualization, but in this case, only three of them can be a possibility: Storytelling, Dashboard and Infographics. [4]

3.1.1 Storytelling

Storytelling is a type of visualization that explains a story on the screen. There are many ways to do a Storytelling, but two of them are the most known:

- "**Static**": using only graphics and images, and the user should pass the slides.
- **Video**: explain a story with audio and images or graphics and the final user does not interact, just looks.

This type of visualization is used to present a site or explain a story. An example: [5]

3.1.2 Dashboard

A dashboard is a visual display of the most important information needed to achieve one or more objectives that has been consolidated on a single screen so it can be monitored at a glance. As a rule, a dashboard is a screen with all the graphics that wants to show. This type of visualization normally is used for monitoring and compare some results. Usually, the graphics are connected between them and one single screen transmits all the information. Each graphic answers something.

Dashboards must be designed to support the process of performance monitoring by:

- Update high level situation awareness.

- Identify and focus on particular items that need attention.
- If action is required, access additional information that is needed, if any, to determine an appropriate response.
- Respond.

The Perceptual Edge Dashboard (Figure 3.1) is an example of dashboard, extracted from a competition [6].

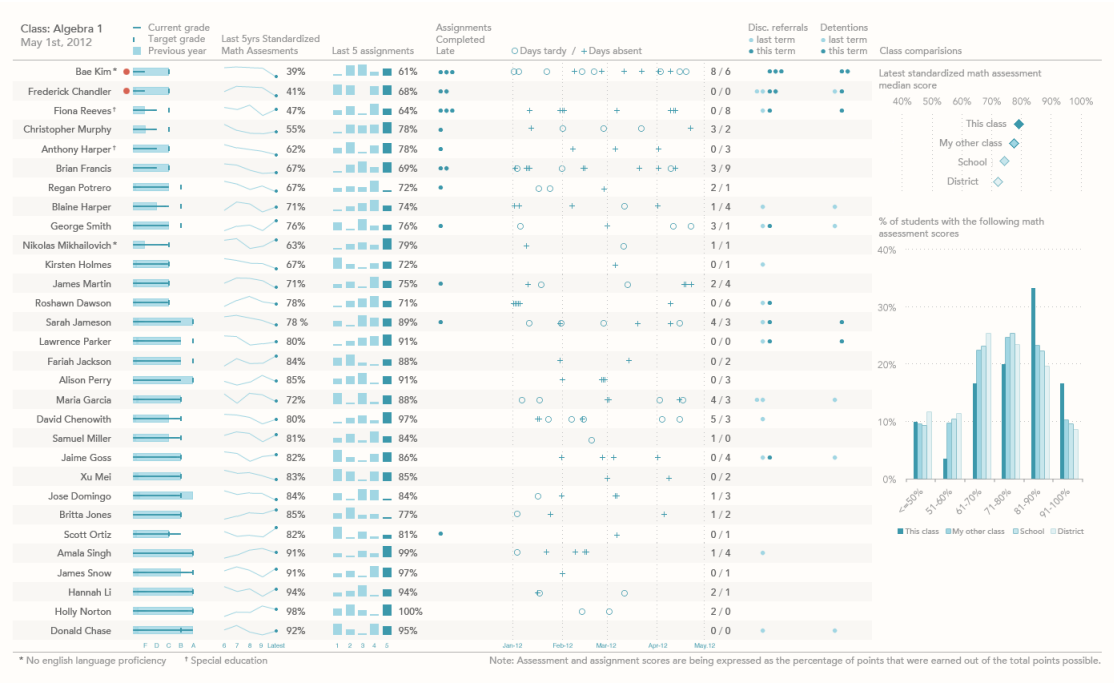

FIGURE 3.1: Perceptual Edge Dashboard

3.1.3 Infographic

An Infographic is a standalone summary of information. It is characterized by illustration, large typography, and long, vertical orientation displaying an assortment of facts. Opposite to dashboards, infographics are for external audiences, we should not expect any expertise for the general public. This type of visualization is useful for posters in the street to present something in a completely static way and for any audience.

The next example (Figure 3.2) is extract from Factographs, collection of facts in graphics [7].

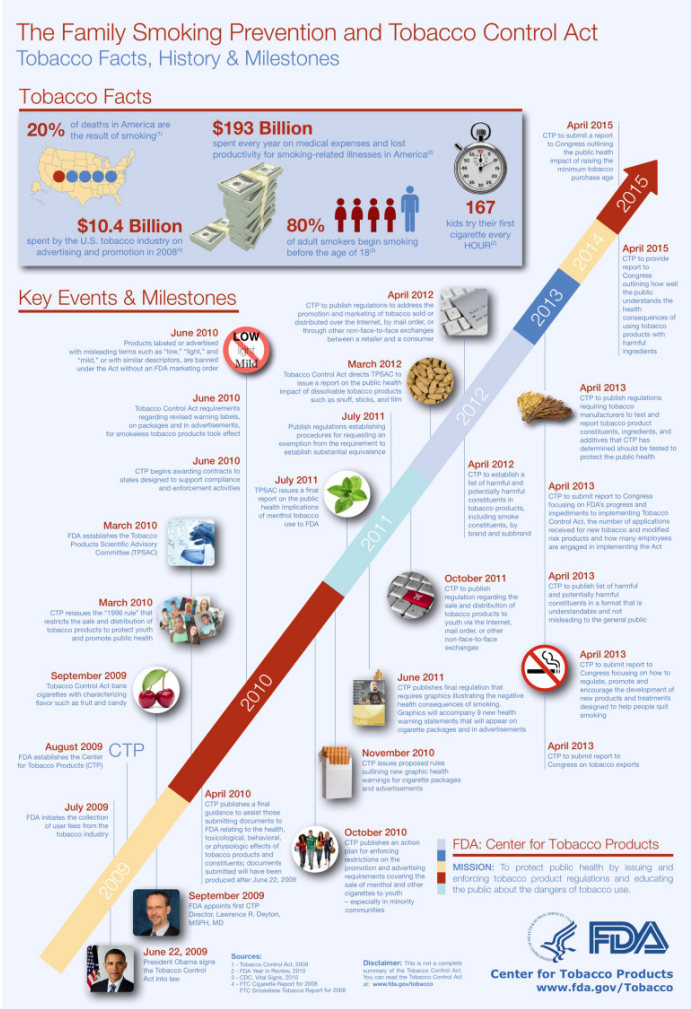

FIGURE 3.2: Infographic

In this project, we chose the dashboard because infographic and storytelling were options that in one year will be obsolete, they are static and Banc dels Aliments wanted a dynamic visualization. In addition, the interaction was an option that they wanted to include.

Furthermore, the dashboard is a good decision to show everything in a single view. It transmits everything at once and, in a few seconds, you can have a general idea of the information that its being transmitted.

3.2 Micro-level

Micro-level design refers to the design of each graphic and the decisions that where taken. Once we know which type of visualization is necessary then the design phase is started. The dashboard was chosen. Then, the next step is to decide which graphics will appear on the screen and which data.

Knowing the data is necessary to do some aggregation and decide how these aggregations will be shown.

3.2.1 Aggregations

The aggregations that were made and the decisions:

- **Geographically**: when the data has places, is an option to do an aggregation per zone. In this case, Entities has an address code, municipality and region. The data is only from Barcelona province, but it can be useful to show the aggregation per region and per municipality. It is discarded to aggregate per address code because there are not so many entities in the same city, but in Barcelona as has so many Entities we decided to do an aggregation per districts.
- **Time**: when it has time data an aggregation per time is a good option, to see the evolution from the other years. In this case, the Orders table has the date in a daily format and it will be useful to make an aggregation per year when the Banc dels Aliments Foundation said that it will be nice to see the improvement of the Foundation.
- **Family of food**: the quantity distributed is per food, and it has food families to make a classification. This decision was taken because there exists hundreds of foods and it be impossible to understand the classification, and per food family, the classification was reduced to eleven options and it will be more clear. For example the food family "*Làctics i derivats*" has 36 terms like: *Llet en pols, iogurts, xocolata defeta,* etc.

3.2.2 Types of graphics

After deciding which aggregations of data were needed, is necessary to decide which type of graphics are appropriate per each aggregation.

A map is a clear and clean way to present the geographical aggregation. The map is easy to understand and in a single view, you can see immediately everything. There are many ways to present the information in a map (choropleth, proportional symbol map, flow maps..). To show the relation between amount and zone the best map is the choropleth map, a thematic map that charts phenomena that are evenly distributed within a specific area. It has a visual impact because color and position are preattentive properties, easily perceived.

Aggregation of time is a good choice to represent the data in a line chart, bar chart or some other graphic with an axis for the timeline. Both of them are visual to

see the years and the evolution of the data itself. For time series, is more appropriate the use of a bar graph or a line chart. The decision of which chart is suitable for the data will depend on what aspect of the information you want to focus on: trends and patterns or individual values.

The food family classification is a classification per types, for this type of data can be represented by different charts: waffle chart, pie chart, donut chart, bar chart... All of them can present the proportions and it is easy to see the information. We need a chart to show the proportions so that we can identify which types of food are the most distributed and in which amount. Pie chart or donut chart are not the best decision, because our perception of dimensions and angles is not very precise and they are used when one attribute or gross comparison among very few values (2 to 4), not this case. The best one to see in a simple view the proportions is a waffle chart.

People served was another data that should be represented, but Banc dels Aliments only wanted the total of served people. This information is going to be shown with an icon and a label indicating how many people has been served.

3.2.3 Codification

All types of graphics should explain how they are encoded and what decisions are made from them.

Geographic

As it was explained before, the graphic will be a choropleth map, but what should it show and how? was the principal question. The data per zone has a date, type of food and quantity. In a map, the data that can be easily represented is the quantity of food. This data will be presented in distributed kilograms per zone, but with this option, there is a problem; the zones with more population always will have a huge quantity of distributed food. In a Barcelona province is obvious that the regions and municipalities with more population always will be the first ones. Then, the decision with Banc dels Aliments was to show two different maps, one per kilograms distributed (because they wanted to show the total kilograms distributed) and another map with kilograms distributed per habitants, because this map has a different interpretation: which zone is more generous. Both of these maps are choropleth maps with a gradient palette to represent the highest values with dark zones. The maps have different palettes of colours because each map represents different units.

Time

To present the time it was clear to do it with an X axis, a timeline. This timeline can present the type of food and quantity. Notice that the zone is already represented in the map. Therefore, we want to represent the amount of food distributed per year, this graph can be represented using a bar chart or a line chart. The line chart is better to see the evolution of the data and the bar chart for timely comparisons.

In this case, it is necessary to represent the maximum information in each graphic, so a decision taken here was to show the total quantity or quantity per food families. In the second case, remember that it has eleven food families, with a bar chart it be

a mess and visually is not an option to represent eleven classifications per year. For all these reasons, the line chart is the best option to represent the time and quantity of kilograms distributed and families of food, showing the evolution. Also in this chart will be represented the improvements in the passage of time.

Food families

In the last graphic this aggregation was used, but in the line chart is not always easy to see the difference between some items. So it is necessary another graphic to only represent the dimensions of the food families classifications. There are some graphics to present this information, but the best one, in this case, is a waffle chart because visually is easy to watch the dimensions and proportions of the different families. The waffle chart is a graphic to represent in different colors the classification and, in a view, you can say which one has a huge proportion. In this graphic will be representing the quantity of food family.

People attended

To represent the attended people is only necessary to show the total number of people, but to be more visual will be supported by a kind of waffle chart with a people icon.

In summary, two maps will appear in the dashboard: one that represents kilograms per zone and an other one with kilograms/population per zone. A line chart will represent the kilograms per year, a waffle chart will represent the quantity per food families and a waffle chart of people attended. Moreover, a menu to select the food families and the years.

3.3 Interactions

In this section we are going to explain the interactions between all the graphics.

The final user can interact with the dashboard filtering by zone, time and food family. The ways to filter are:

Menu

The dashboard will contain a menu from where the user can select the food families and years that he wants to see. The food families will be in a check-box list and the years will be within an slider. Both of them are filters for the visualization that select a time period or a type of food.

When the user changes the configuration of the shown food type, using the mentioned check-box list, then the other graphics reacts by:

- Maps: changing the number of kilograms per zone because it is the sum of the number of selected food families.
- Line chart years: adding the food families selection with a new line per each family.

- Waffle chart food: showing only the families selected.

In the case of the years, a slider is used to select the desired time. This interaction effects the following graphics by:

- Maps: only showing the kilograms distributed in the selected time range.
- Waffle chart food: showing the proportion of quantity per food family in the selected period.
- Waffle chart people: showing the people attended during the selected period.

Maps

The maps will appear divided by regions and the user can click on one of them to zoom in and change the view of the province of Barcelona to the selected region. This interaction is used to do a zoom and change the view. The graphics affected are:

- Maps: both maps will zoom-in into a new region map, showing the municipalities from there.
- Line chart years: change the number of kilograms distributed in all Barcelona province to the next level of selected region.
- Waffle chart food: showing the proportion of quantity per food family in the selected region.
- Waffle chart people: showing the people attended in this zone.

Line chart

Exists an interaction in the line chart: to select the years by brushing, sectioning a range of years. The affected graphics by this interaction are the same as the slider case, because both of them select a range of years and induce the same interaction.

Breadcrumbs

Another interaction is with the breadcrumbs, a line with the path that indicates where we are inside the web application. When in the map the interaction is zoom-in per region, the user may want to come back to the Barcelona province map. So in the breadcrumb, the interaction will be zoom-out and the graphics affected are:

- Maps: both maps will zoom-out to a Barcelona province map, showing the regions from there.
- Line chart years: change the number of kilograms distributed in the region to the before level of Barcelona province.
- Waffle chart food: showing the proportion of quantity per food family in the Barcelona province.

- Waffle chart people: showing the people attended in this zone.

All of these interactions use all the graphics from the dashboard, making it possible to display all the information on a single screen.

3.4 Mock-ups

The design finishes with the mock-ups, where some images simulate the final result of the visualization. Including all the requisites and decisions taken before that.

The mock-up showed to Banc dels Aliments Foundation was the next image:

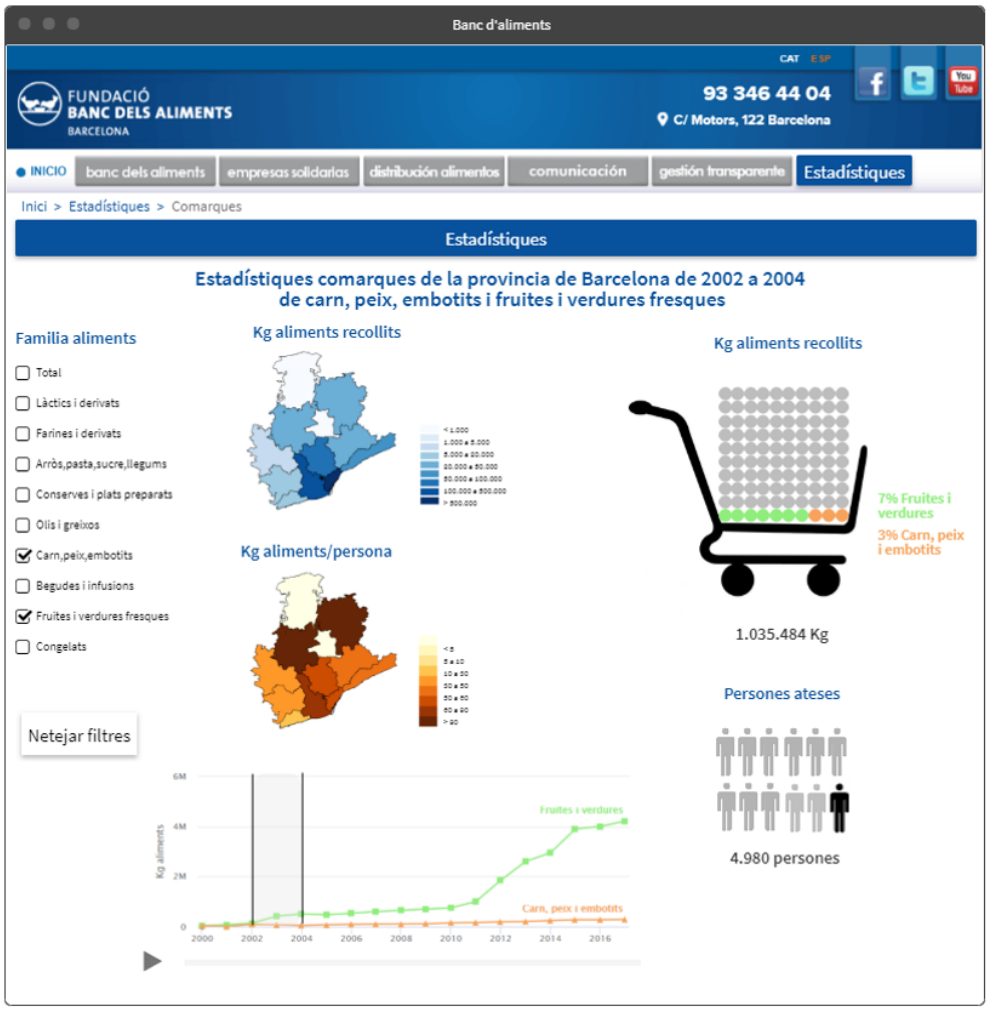

FIGURE 3.3: Mock-up design

As the above Figure shows, the dashboard has all the elements explained in the lasts sections:

- The left menu with the selectors of food families and the slider for the years.

- The two maps of the province of Barcelona divided by regions, one of them expresses the quantity of food distributed and the other one expresses the same quantity but divided by the number of population.
- The line chart with the timeline. The x-axis contains the years and the y-axis the quantity of distributed food. If some food families are selected, the timeline will contain one line per each family.
- The waffle chart of food, showing the proportion of each selected food family and the total quantity in kilograms.
- The waffle chart of people with the total of people attended during the selected period.

Other details to comment: The colors of the maps are different because each map represents a different unit.

Every food family has a color defined and in the line chart and waffle chart the same color is always shown, to maintain consistency.

This mock-up is the final mock-up done after a meeting with our contact from Banc dels Aliments (on 7th of May of 2018). In the meeting, he commented some things to improve the design and help us to understand some of the business rules. The comments done were:

- The title needs to show where we are, showing the zone and years selected.
- Delete a graphic that contains a diet of a person, he said that was not relevant for this visualization and it will be static.
- Do an explanation of each graphic.
- Delete food families of products that were not food. Example: "Complements dietetics" or "Productes no alimentaris".
- Add breadcrumbs to come back on a previous page.
- Add a "total" option in the food families check-box.

Chapter 4

Implementation

In this chapter, it is explained the process and implementation of this visualizations.

4.1 Physical architecture and technologies

This section is centered in the physical architecture that the visualization needs, and the technologies used to build it.

The solution proposed is a visualization that will be in a Banc dels Aliments Foundation website. Therefore, our visualization should follow the same architecture of the Banc dels Aliments website.

The website has an Apache server to save all the pages of the website and MySQL database. This server is in the Banc dels Aliments environment, due to that, in order to implement this project, it was necessary to simulate this environment.

The technology used to simulate it is XAMPP[8]: a free and open source cross-platform web server solution with an Apache HTTP server, MySQL database and the necessary tools to program with PHP. This technology allows to simulate the final environment and shows that the proposed solution will work on the website.

The technology used to manipulate the database is MySQL Workbench, a unified visual tool for database architects, developers, and DBAs (Database Analyst).

Finally, to implement all the logical code the tool that has been used is PHPStorm, a PHP Jetbrains IDE.

4.2 Logic architecture

In this section, it is explained the logic architecture used for the implementation.

The pattern used is Model – View – Controller (MVC). This pattern consists in dividing the data layer, the business logic and the user interface. There are three main components:

- Model: representation of information, defining the model of data and is the only part that has a connection with the database.
- View: presents the information on the screen in a way that the user can interact with it.

- Controller: part of the code in charge of asking the data from the model, process it do some logic and then send it to the view.

In this project, the pattern used is applied through the different parts of the code. There is a unique view because the visualization is a dashboard, a unique screen. Although, to be more neat a Controller and Model has been created per each graphic. To better understand in the Figure 4.1 shows the pattern with the type of document codes.

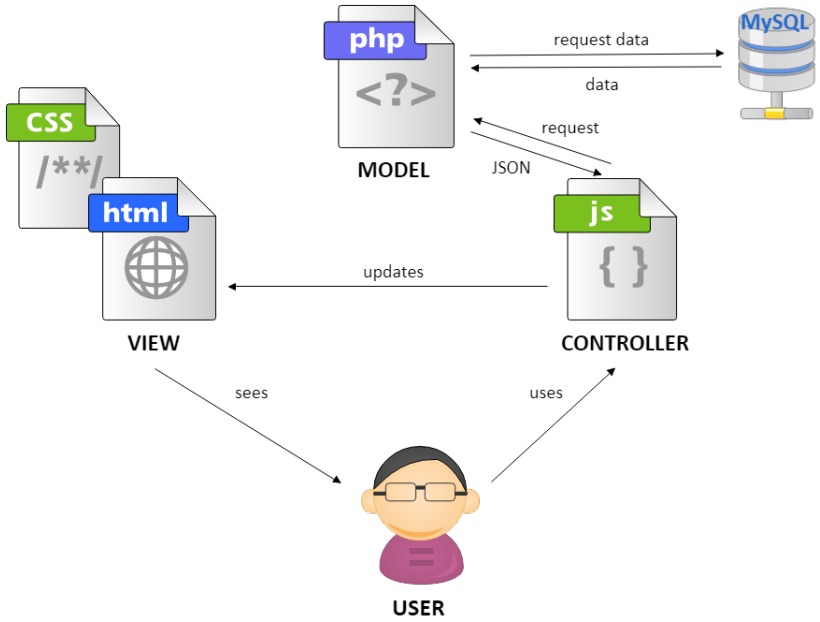

FIGURE 4.1: Pattern Model - View - Controller

In addition, an example for each component will be explained. The code shown is only the necessary to understand the communication between the components. The complete code of this project is in Github [9].

Model

The model component is a PHP code, this code is the responsible to get the data from the database, filter it, parse it in a JSON format and send it to the controller. As for each graphic, different data is needed, each graphic has it is own model. The listing 4.1 is the waffle chart model. It depends on years and regions, and the data is grouped by food family.

```php
<?php
  include 'db_connection.php'
  [...]
  if( isset($_GET['param']) ) $comarca = $_GET['param'];

  $sql = "SELECT MACROFAMILIA, sum(QUANTITAT) AS QUANTITAT FROM $
      comarca WHERE YEAR > $Y1 AND YEAR < $Y2 AND MACROFAMILIA IN ($
      families) GROUP BY MACROFAMILIA";

  $result = mysqli_query($conn, $sql);
  [...]
  echo json_encode($data);
?>
```

LISTING 4.1: Model code

View

The view component is an HTML code, this code is the responsible to show the information on the screen, where it specifies the elements and the style that user will see. This is the unique view, that shows the dashboard with all the graphics. The listing 4.2 shows how to reference the graphic that the view will display.

```html
<!DOCTYPE html>
[...]
  <body class="keen-dashboard" style="padding-top: 80px;">
    div class="col-md-5">
        <div class="row">
           <div class="chart-wrapper">
          <div class="chart-title">
            Kg aliments repartits
          </div>
          <div class="chart-stage">
            <script src='js/d3-waffle.js'></script>
            <script src='js/wafflevf.js'></script>
            <div class="container" id="chartwf-1"></div>
            <div style="margin-left: 150px;"><p>
              <input type="text" id="numkilos" readonly>
            </p></div>
          </div>
[...]
</html>
```

LISTING 4.2: View code

Controller

For each Model a controller component has been implemented. Each one of this controllers is in charge of managing and transforming the data the correspond graphic. The controller is the responsible to execute the actions that the user wants to do through the interaction and is the element that requests the data to the Model and returns the updated data to the view. This class is implemented in JavaScript. The listing 4.3 is the part of the controller from the waffle chart that request for the data to the Model component using Ajax, asynchronous HTTP request from jQuery library [10], a JavaScript library.

```
1   var waffleAjax = function(name,year1,year2,families) {
2     $(function () {
3         $.ajax({
4           url: name+'.php',
5           data: {'param' : year1, 'param2': year2, 'param3': families},
6           type: "GET",
7           dataType: 'json',
8           success: function(data) {
9             dataw1 = data;
10            }
11        });
12    });
13    return dataw1;
14  };
```

LISTING 4.3: Controller code: request data

The listing 4.4 is the part of the controller from the waffle chart that manage the data and updates the view.

```
1  var waffleDraw = function() {
2
3      var domain = dataw1.map(function(el) {
4        return {
5            name: el.MACROFAMILIA,
6            value: el.QUANTITAT,
7        };
8      });
9      var chart4 = d3waffle()
10                     .rows(10)
11                     .height(170);
12     d3.select("#chartwf-1")
13       .datum(domain)
14       .call(chart4);
15 };
```

LISTING 4.4: Controller code: updates view

4.3 Specific parts

In this section, it is explained the back-end and front-end parts of the code. The problems, decisions and they are implemented.

4.3.1 Back-end

The back-end is the part of the code that refers to the database and manages the data.

In the implementation of this project, the main problem of the back-end part was the huge amount of data. The database saves all data without any problem. The issue was during the modeling step of the data, the part of the code that was explained as Model. This code contains SQL queries and the server can not hold much time waiting a response. For that reason, it was necessary to make views in the database. A view is the result set of a stored query on the data.

The views made were:

- **Region_Order_view**, a view that contains: Region, Year, Food family and Quantity. This view is used in all the graphics of the province of Barcelona screen.
- **[Concrete region]_Order_view**, for each region there is a view that contains: Year, Municipality, Food Family and Quantity. This view is used by all the graphics when the user chooses a region.

When the visualization is about province of Barcelona, an aggregation by region is needed. Year and food family for all the graphics can react with the interactions and change the data showing. The view Region_Order_view is the data used in the province of Barcelona visualization.

In the other case, when the visualization is about a certain region, uses the [Concrete region]_Orders_view of the concrete region for all the graphics.

In addition, this views are the way to merge all the data from the first years, 2001 to 2016 that are in the two tables, with the data from 2017, that is in a unique table. As it explains the Chapter 2 Format and Transformation data.

4.3.2 Front-end

The front-end is the part of the code that refers to the presentation layer, the part that shows the information to the user and he interacts with it.

This part is in the View and Controller codes of this project. Both of this parts have some elements to take into account.

The view has dashboard structure, which has the code to build the layout and is the responsible to maintain the organization of the elements from the dashboard. To make this it is necessary a layout that divides the screen in three columns and three rows, using Bootstrap library [11] to get a nice design web. In addition, the layout has the title and breadcrumbs. These elements change when the user interacts with the visualization, but the responsible of updating these elements is the controller.

The controller is the part of the code where the interactions are made. This code is in JavaScript, more concretely with the library D3. D3 is a JavaScript library for producing dynamic, interactive data visualizations in web browsers.

D3 has a lot of facilities to implement the graphics in the visualization and their interactions. Below is shown how was implemented each graphic.

Map

To build the maps in D3 is needed a TopoJSON document, a JSON format to draw maps, divisions of territory and the information about them. This document was built with QGIS and mapshaper tools, both permits to visualize the map and shows the metadata for each region. QGIS was used to change the map projection and to add data of the population for each region. The data was from Idescat [12].

When the document is ready, it must process the data from the database, send it to the model and paint the map according to the data. In addition, the remote map, the province of Barcelona divided by regions is necessary to add an interaction. This interaction happens when clicking on a region, then loads the new map and repeats the process explained before (loads the data and paints it the new map). Also, it has to add the legend.

Line chart

The line chart has a line per Food Family and shows the quantity of distributed food per year. This graphic seems simple but the difficulty was the implementation of the interaction. The user can interact with this chart by brushing, which means selecting a part of the chart using a gesture such as clicking or dragging the mouse. The **brushing** in this graphic is used to select a period of years. To implement this part was necessary to add a D3 interaction in the x-axis, a brush function and define the function specifying that the brush should stop in an existing x-axis element. An example: The region selected with the brush must start and end in a specific year, it is not possible to stop holding the brush in the middle of two years. This had to be implemented because the data was grouped by years.Additionally, the brush function has to update the data and refresh the other graphics.

Waffle chart

In this graphic before drawing we need to define the maximum rows for then calculate the proportions for each food family. Also, in the second waffle chart of the people attended, the font awesome it is been used to draw the person icon instead of the normal square form. The font awesome [13] is a library that provides icons and styling tools. This library has been integrated with D3 in order to paint the person icon.

Title and breadcrumbs

The title and breadcrumbs are updated in each user interaction. A function was implemented that updates the names of where we are in the web application and which type of family food is selected at the moment.

Chapter 5

Conclusions

This chapter explains the conclusions of this project through the achievement of objectives, the contributions to Banc dels Aliments Foundations and the personal learning with this project.

5.1 Achievement of objectives

Once finished the project, it must be checked whether all of the objectives described at the beginning have been achieved.

The principal objective of implementing a visualization of the data of Banc dels Aliments de Barcelona, **it has been achieved**.

The secondary objectives extracted from the principal one were:

- To do a good visualization per location and food for the Banc dels Aliments that sensitize the people.
- To analyze the data so that they can do a better organization of their resources.

The first one is achieved because the visualization complies with all the requisites that Banc dels Aliments told us, but we must wait for the response of the final users to know if the objective has been really achieved. The other one it is in hands of Banc dels Aliments, the visualization shows an analysis of their data geographically and by time, so they can see how is the food distributed. With this new information Banc dels Aliments can take more and better decisions regarding the management of their resources. Thus, we are **waiting for an answer**.

And the technical objectives of this project were:

- Analysis of the final users. **It has been achieved**, the analysis of the users it's done and the results have affected most of the decisions taken during the project.
- Analysis of the data to study which type of visualization can be done. **It has been achieved**, the analysis of the data is done and the reasons are explained.
- Implement the functionalities to show the data from location and type of food. **It has been achieved**, all the functionalities are implemented.

5.2 Contributions to Banc dels aliments

This project has been finished with a visualization that uses the data from Banc dels Aliments Foundation. The contribution that it has achieved is a visualization that shows the evolution of this Foundation, also their resources and volunteers.

Although not yet published, the organization is going to publish the visualization on their website because the visualization satisfies all the Banc dels Aliments requisites and the users that have already seen it give a good feedback.

5.3 Personal learning

Personally, this project has been very satisfying for the type of client and to understand and participate in the Banc dels Aliments Foundation. Seeing how they work, the attitude they put on and the whole functioning of the foundation.

In addition, the work I have done really like me and although I sometimes had some complications with the implementation or how I should do something, I am very happy with the result and the time I work on it.

It has been a good experience to work with two professors of UB and collaborate with Banc dels Aliments Foundation to help them to do a project that it can contribute in their visualization and sensitize the final users.

Furthermore, I have learned more technical knowledge and I put into action some knowledge learn it in this master. How to make decisions, analyses some data or learn how to manage a huge amount of data.

Chapter 6

Future work

In this last chapter is explained the future work that could be done.

6.1 Deployment

Be integrated with the Banc dels Aliments system is a future work to do. The visualization complies all the software requisites to integrated with the Foundation system, but the deployment will depend on Banc dels Aliments Foundation, they should approve the visualization and then it will be necessary to upload all the code and the intermediate database in their server.

6.2 Real time data

Other future work after talk with Banc dels Aliments Foundation it can be to use the data of the Foundation in real time. Now the data used is old data, "frozen" data, but it will be interesting to use real-time data to see the evolution of the current year and how it is growing.

Appendix A

Screen-shots of the visualization

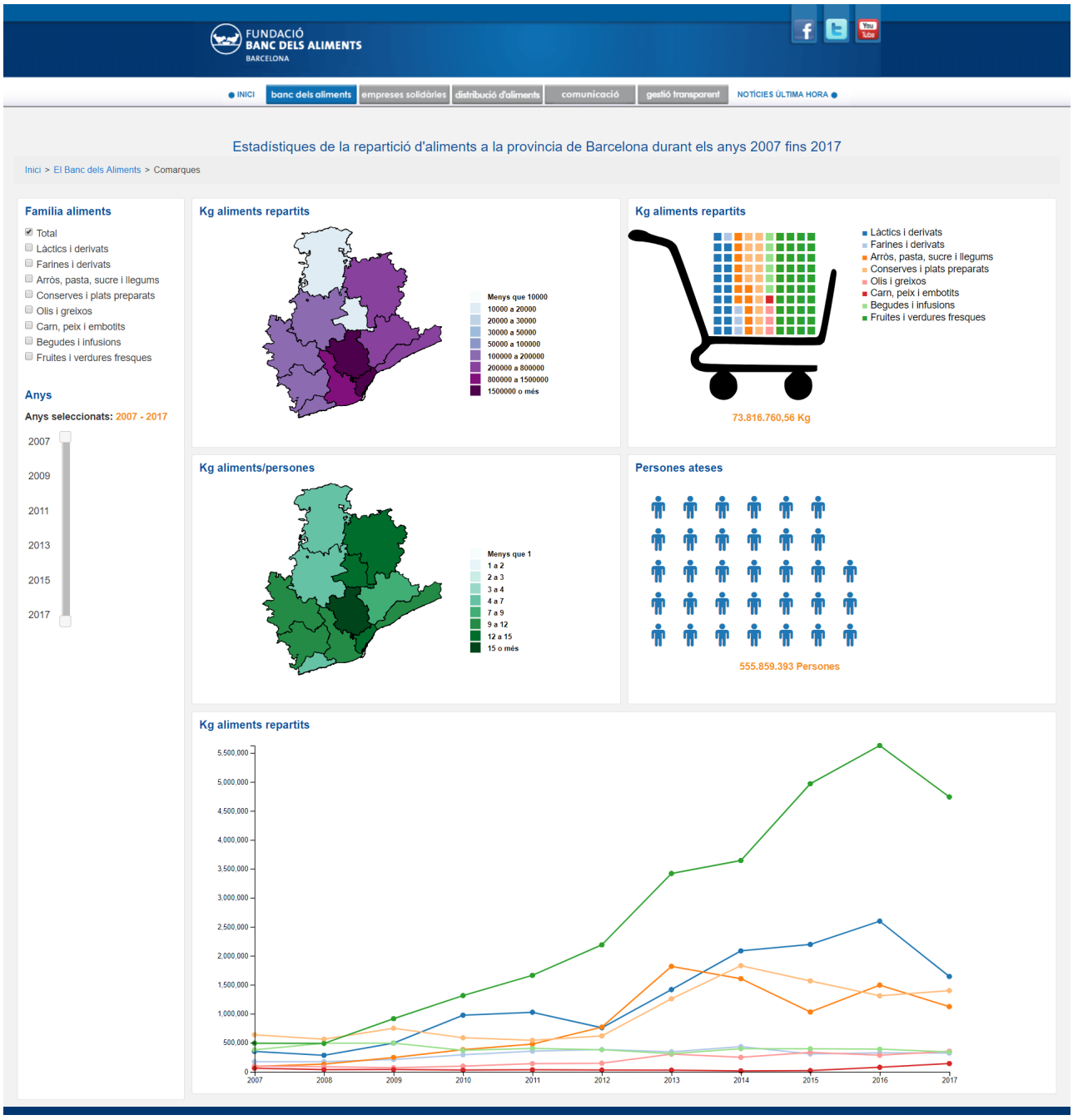

FIGURE A.1: Visualization of province of Barcelona

Visualization filtering by zone, years and food family.

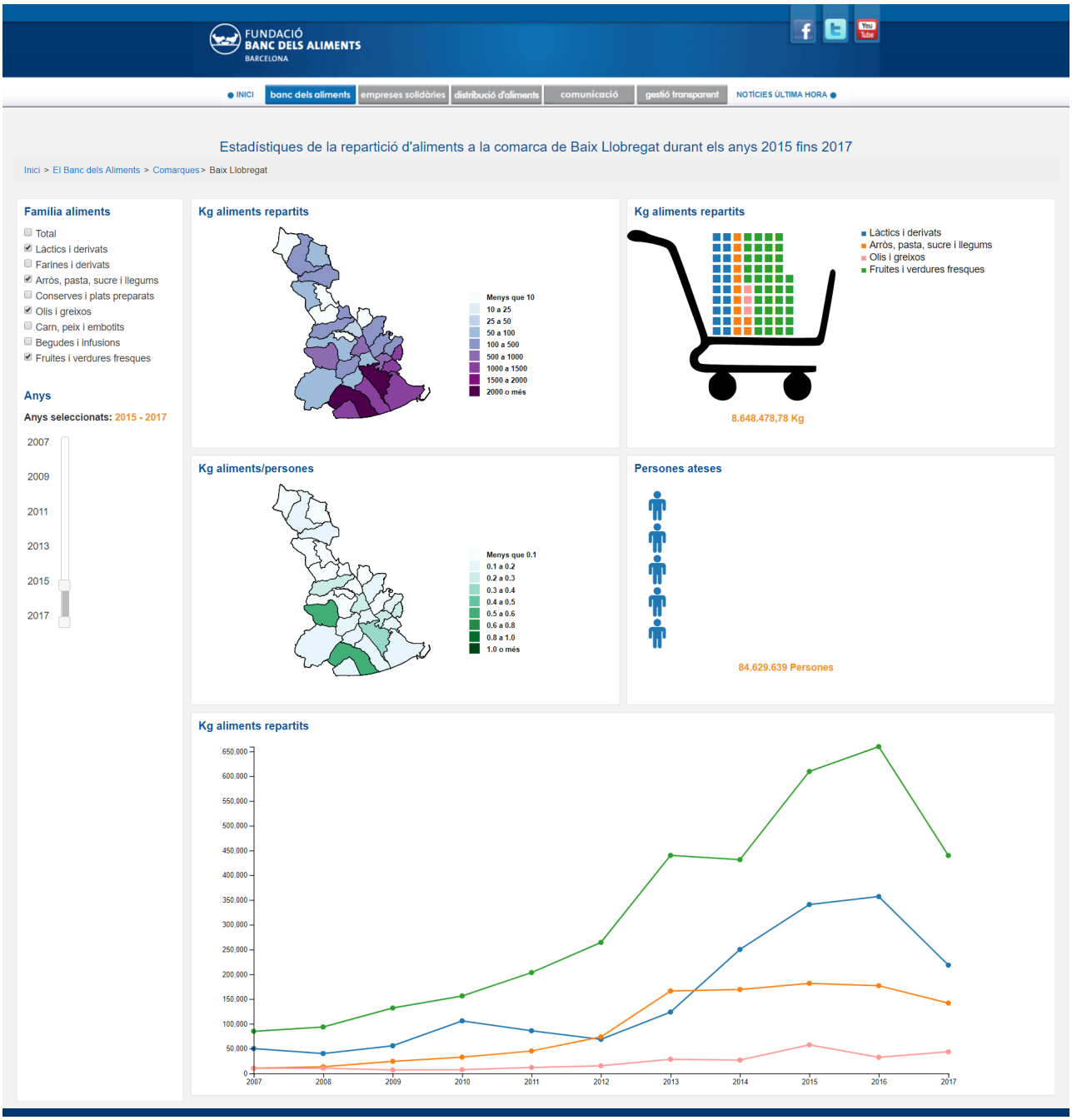

FIGURE A.2: Visualization of Baix Llobregat

Appendix B

GitHub repository

In the following link you can find the code of the project:

https://github.com/carladivicuesta/tfm [9]

Bibliography

[1] Banc dels aliments de Barcelona. *Fundació Banc dels aliments Barcelona*. https://www.bancdelsaliments.org/. [Accessed: March of 2018]. 2011.

[2] Capital Area Food Bank. *CAFB Hunger Heat Map*. http://cafb.maps.arcgis.com/apps/MapJournal/index.html?appid=b4906ac11bf74cd781c5567124be9364. [Accessed: April of 2018]. 2015-2017.

[3] UK National Statistics. *U.K. Energy Consumption Guide*. https://www.evoenergy.co.uk/uk-energy-guide/. [Accessed: May of 2018]. 2018.

[4] Mireia Ribera. *Presentation and visualization material course (2018)*. . [Accessed: March of 2018]. 2018.

[5] Atlassian. *Time wasting at work*. https://www.atlassian.com/time-wasting-at-work-infographic. [Accessed: May of 2018]. 2017.

[6] Perceptual Edge. *Perceptual Edge Dashboard Design Competition*. http://www.perceptualedge.com/blog/wp-content/uploads/2012/10/dashboard-competition-winner.png. [Accessed: March of 2018]. 2012.

[7] Suresh Suthar. *Family smoking prevention*. https://factographs.wordpress.com/2012/04/26/family-smoking-prevention-and-tobacco-control-act/. [Accessed: April of 2018]. 2012.

[8] Apache friends. *XAMPP tool*. https://www.apachefriends.org/es/index.html. [Accessed: April of 2018]. 2018.

[9] Carla Diví Cuesta. *Code of the project*. https://github.com/carladivicuesta/tfm. 2018.

[10] jQuery. *jQuery*. https://jquery.com/. [Accessed: July of 2018]. 2018.

[11] Bootstrap. *Bootstrap*. https://getbootstrap.com/. [Accessed: June of 2018]. 2017.

[12] Idescat. *Institut d'estadística de Catalunya, Generalitat de Catalunya*. https://www.idescat.cat/. [Accessed: June of 2018]. 2008.

[13] Font awesome. *Font awesome*. https://fontawesome.com/. [Accessed: July of 2018]. 2018.