21-01-2019

RELATION BETWEEN COMPANIES AND GROSS DOMESTIC PRODUCT IN SPAIN FROM 2008 TO 2017

Thematic Visualization

Carlos Bayarri Cebrecos

Content

1. Introduction

The purpose of this project is to collect, develop and analyze statistical data, and construct several suitable representations about this data, explaining a history, what is happened.

But let's tell a little bit history before. Some years ago began the Spanish financial crisis (2008 – 2014). *Spain* became a part of the European sovereign debt crisis because this country was unable to rescue its financial sector. The housing bubble exploded, and the high Gross Domestic Product (GDP) growth rate.

As the BBC explained in their news, "After *Spain* joined the euro, the country experienced a long boom, underpinned by a housing bubble, financed by cheap loans to builders and homebuyers […] So, although the Spanish government still had relatively low debts, it has had to borrow heavily to deal with the effects of the property collapse, the recession and the worst unemployment rate in the Eurozone."(Fund, Olympics, & Union, 2013).

Spain left behind the recession in 2014 with a growth of 1.4% of GDP, which would continue with a figure of 3.2% in 2015. In 2014, 433,000 jobs were created and unemployment fell to 23.78% in 2015. In 2016, Spain became the second country in the Eurozone that grew the most, registering a growth of 0.7% in the second quarter of that year.

As of June 2017, *Spain* eventually recovers all the GDP lost during the crisis, registering a growth of 0.9% in the second quarter of the same year, becoming the fastest growing developed country.

This is why we want to explain a little part of the business world in *Spain* along the crisis. All the people in *Spain* have been affected by this crisis: salaries were very low, there were less government aid, everything seemed very expensive. In chars to analyze a part of the crisis evolution, we are going to relate the companies that they have been active along a decade in each region with the GDP per capita produced.

2. Background

Talking about this fact, a big crisis is not easy and less talk about how to explain it graphically. For this reason, we are going to explain some basis.

In this project, we use an approach called “Thematic map”. A thematic map is used to display the spatial pattern of a theme or attribute (Terry A. Slocum, Robert B. McMaster, Frits C. Kessler, & Hugh H. Howard, 2009). We can use this kind of map to provide specific information about our world, information of patterns and compare patterns in different ways. We have two techniques for thematic mapping: a static map (like a traditional map) and a dynamic map presentation (like web mapping). After, we will see our theme is reflected with both techniques, and showing information of patterns and comparisons.

Steps for communicating map information

To develop these techniques, we have to follow the basic steps for communicating map information (Terry A. Slocum et al., 2009):

1. Consider what the real-world distribution of the phenomenon might look like. We can consider that there are many companies affected by the crisis, people has experimented a decline but an improvement also. Of course, regions have experienced a big change in their GDP.

2. Determine the purpose of the map and its intended audience. In our case, we want to analyze the relation between active companies and GDP in the crisis decade, check if the crisis is reflected in these variables, and if the companies’ number is decreased by this crisis.

3. Collect data appropriate for the map’s purpose. The main place to search this kind of statistical data is the most accepted institute called Statistics National Institute (SNI). Here we have many groups of information like agriculture and environment, science and technology, demography and population, but the selected group for our purpose is economy. Inside, we have some themes, and the selected one is “Statistical Use of Central Business Register “(CBR). The objective is to group all Spanish companies and their local units located in the national territory in a single information system, studying companies and local units by location, employee stratum, legal condition and branch of activity. Later, in the Data chapter, we will see how to analyze and develop this data.

Moreover, we want to explain the time component. We have to develop the data mentioned in a decade. This data is changing over each year; it is collected by years. Here we have different aspects for both techniques: in the static representation, we have symbols dividing their state for each year; in the dynamic representation, we will see a temporal change over the same positions.

4. Design and construct the map. There are some cartographic representation methods for representing data, like area, scattered, qualitative or quantitative data, but is this last one we are going to use mainly. Also, we will use the choropleth approach to represent a specific variable, but we will see later. So, we are going to map quantitative data by proportional map symbols: positional and areal.

This is a difficult task; we have to make representations according to the data, according to what we want to show and according to what the final user has to see. Both representations are going to use the same data, but in different ways of representation and visualization. Later, we will see these different approaches.

5. Determine whether users find the map useful and informative. At the end of the document we will see the conclusion of both techniques.

Map projection

To project the World onto a flat surface is through a map projection. The map suffer from distortion, of course, when the Earth's two-dimensional curved surface (a glove of Earth that preserver areal, angular, distance and directional relationships) is projected to a map (Terry A. Slocum et al., 2009). For this reason, we have to search the right projection over our study zone, *Spain*. We have chosen Lambert azimuthal equal-area projection. This projection preserves the area of individual polygons while simultaneously maintaining a true sense of direction from the center. The general pattern of distortion is radial ("Lambert Azimuthal Equal Area—Help | ArcGIS for Desktop," n.d.). Even though shape is minimally distorted, less than 2 per cent, angular distortion is more significant; small shapes are compressed radially from the center and elongated perpendicularly. But we preserve areas and we have true direction radiating from the central point.

3. Data

As we have said, our main data is the number of active companies and the Gross Domestic Product per capita in *Spain*, over a range of years (2008 to 2017). This data is divided for each region (17 regions):

1. *Andalucía*
2. *Aragón*
3. *Principado de Asturias*
4. *Illes Balears*
5. *Canarias*
6. *Cantabria*
7. *Castilla y León*
8. *Castilla – La Mancha*
9. *Cataluña*
10. *Comunitat Valenciana*
11. *Extremadura*
12. *Galicia*
13. *Comunidad de Madrid*
14. *Región de Murcia*
15. *Comunidad Foral de Navarra*
16. *País Vasco*
17. *La Rioja*

The data is provided by the Statistics National Institute of *Spain*. We have collected it in excel format, to make calculus and plots for the analysis. After, we have converted the final data to JSON, to work with the different techniques we have used. There is an excel file in this project with all the data used:

- Tabs for each region contain all the active companies in each sector.
- Tab about *NaturalBreaksGDP* contains the process to classify the Gross Domestic Product by means of natural breaks.
- Tab about *ValuesStatic* contains the values for the static map.
- Tab about *RelativeValuesCompanies* contains the standardized values of companies.
- Tab about *AbsoluteValuesCompanies* contains the absolute values of companies.

Companies by Region, main activity (CNAE groups 2009) and employee stratum.

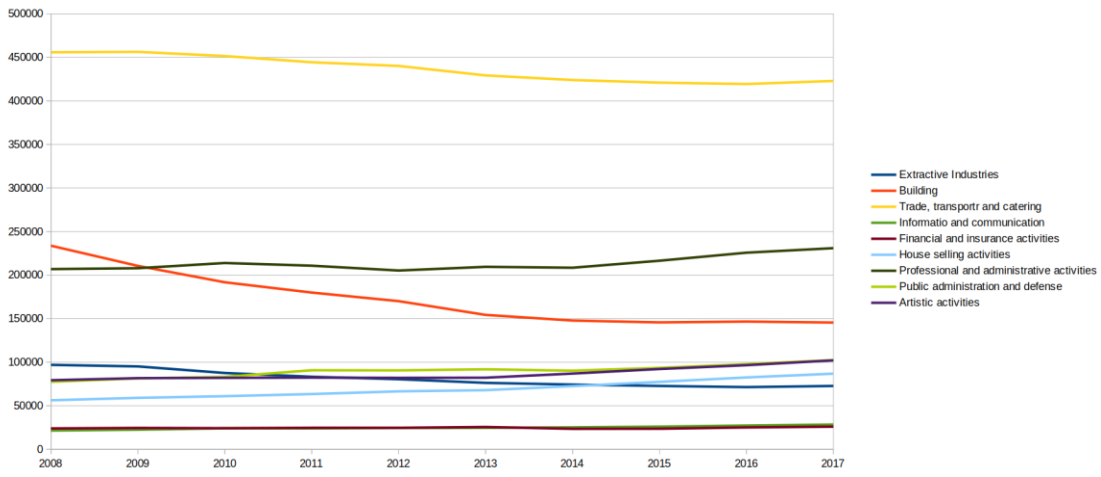
Companies are grouped in main activities, according to Statistical Use of Central Business Register (CBR):

- Extractive industries
- Building
- Trade, transport and catering
- Information and communication
- Financial and insurance activities
- Professional and administrative activities
- House selling activities
- Public administration and defense
- Artistic activities

This classification is organized according to National Classification of Economic Activities (in Spanish CNAE). Of course, these nine groups are the most important; meanwhile each one has many subgroups. Here we have many groups, so we have to group some of this, because some of them have big values and symbols result complex. The suitable way to accomplish with our symbols is to take into account that these groups do not have similar values at all. For example, the right table show us the number of companies divided into the nine activity groups and for 2008, in *Cataluña*.

CATALUÑA	2008
Extractive industries	96926
Building	233730
Trade, transport and catering	455808
Informatio and communication	21554
Financial and insurance activities	24002
House selling activities	56274
Professional and administrative activities	206896
Public administration and defense	77518
Artistic activities	79332

In the next chart, we can see differentiate groups:



As we can see, trade, transport and catering has 455808 companies (a differentiate group), building has 233730 companies and professional and administrative activities has 206896 companies. Meanwhile, the other groups have less than 10000 companies. If we group these latest groups in only one, we have four groups with similar quantities. These phenomena occur in all the years, and in each region (in their proper measure). The result, following with the previous example, is like:

CATALUÑA	2008
Trade, transport and catering	455808
Building	233730
Professional and administrative activities	206896
Other	355606

Also, we have done standardization over these values. This is why we have different symbols to represent the number of companies in the different techniques:

- In the static representation, we have the start or spider graph. This is a multivariate point symbol map. The multivariate ray-glyph or star is constructed by extending rays from an interior circle, with the lengths of the rays made proportional to values associated to each variable (Slocum, 1999). These symbols are an example of nonrelated variables and are not part of a larger whole. That means we want to know the number of companies without a relation between the others.

  Moreover, we have a problem with the presented representation: in the static map, if we want to see three years overlapping with this graph, we have visualization lacks, because the change is not too big. For this reason, we have standardized these variables by means of the population of each region. The procedure is dividing the number of companies between the population of its region in the right year, and multiplying this by a thousand. So we have the number of companies per 1000 inhabitants. Now, instead of high values, we have values more similar, for instance:

CATALUÑA	2008
Trade, transport and cattering	62
Building	32
Professional and administrative activities	28
Other	48

  As an example of operation, we have 455808 companies about Trade, transport and catering in *Cataluña* in 2008. This region has 7364078 inhabitants in 2008. The operation is 455808 / 7364078 * 1000, resulting 62 (rounded).

- In the dynamic representation, we have the pie chart. This is another multivariate point symbol map, but with related variables. Such variables can be depicted using the pie chart, in which a circle is divided into sectors representing the proportion of each variable (Slocum, 1999), as a part of a larger whole. In this way, we are dividing a circle in parts proportional to the variables according to the sum of all of them. In the dynamic representation chapter, we will see this kind of symbol with non-standardized and standardized values. This is why is interest the result in the dynamic one, how we can see the change of the quantities in the number of companies or the whole group.

  Moreover, the reason why we have used three years in the static one and all the years in the dynamic one is simple: the static representation is limited, and we cannot see all the years overlapping. We have chosen the beginning of the crisis (2008), the worst transition (2012) and the last year like as a boom in the economy (2017); in the dynamic one, we can see all the transition over the decade with animations.

Gross Domestic Product[1]

This variable by region is obtained by the Regional Accounting in Spain. The measure unit is Euros. We have worked with this variable because this is an economic indicator that measures the relationship between the level of income of a country and its population. For this, the Gross Domestic Product (GDP) of said territory is divided by the number of inhabitants (Bankinter. Blog de referencia financiera, 2017).

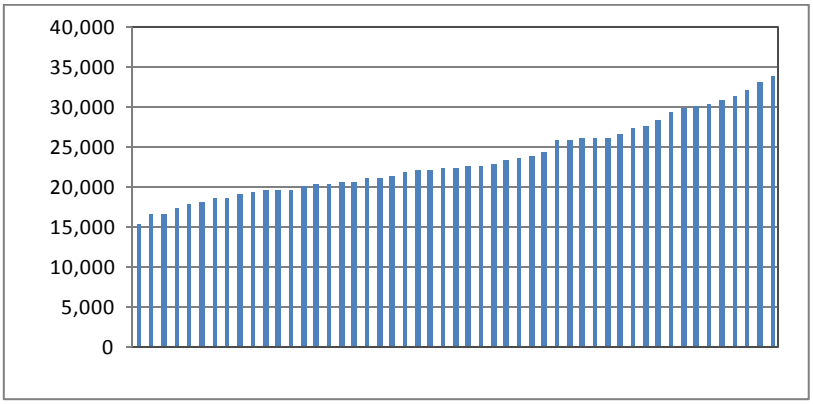
This variable we have obtained is represented by choropleth in the dynamic one and by a Doughnut graph in the static one.

In addition, we have to classify the data. According to Terry A. Slocum, "Data classification involves combining raw data into classes or groups, with each class represented by a unique symbol. For instance, data for a choropleth map might be grouped into five classes, with each class depicted by a different shade of gray"(Terry A. Slocum et al., 2009). In this way, we have done a classification by natural breaks. These natural breaks allow us to see the groups or cluster considering natural groupings of the data, minimizing differences between data values in the same class and maximizing the differences between classes. As an example, we can choose the 2008, 2012 and 2017 years, like the right image.

	2008	2012	2017
ANDALUCÍA	18,625	16,666	18,470
ARAGÓN	26,650	24,267	27,403
ASTURIAS, PRINCIPADO DE	22,336	20,005	22,046
BALEARS, ILLES	25,717	23,224	25,772
CANARIAS	21,186	19,017	20,425
CANTABRIA	22,850	20,559	22,513
CASTILLA Y LEÓN	22,421	21,150	23,555
CASTILLA - LA MANCHA	19,697	17,864	19,681
CATALUÑA	28,332	26,041	29,936
COMUNITAT VALENCIANA	21,878	19,288	22,055
EXTREMADURA	16,633	15,313	17,262
GALICIA	21,226	19,533	22,497
MADRID, COMUNIDAD DE	32,155	30,455	33,809
MURCIA, REGIÓN DE	20,354	18,168	20,585
NAVARRA, COMUNIDAD FORAL DE	30,128	27,485	30,914
PAÍS VASCO	31,243	29,275	33,088
RIOJA, LA	25,986	23,917	26,044

1 The data obtained is located in the data.xls file, added with the project.

If we group all this values, ordered from least to greatest, we can see the differences between numbers. In the next graph we have all the values:



In the following table, we can see in red these differences we are searching:

VALUE	DIFFERENCE	VALUE	DIFFERENCE
15313		22497	76
16633	1320	22513	16
16666	33	22850	337
17262	596	23224	374
17864	602	23555	331
18168	304	23917	362
18470	302	24267	350
18625	155	25717	1450
19017	392	25772	55
19288	271	25986	214
19533	245	26041	55
19681	148	26044	3
19697	16	26650	606
20005	308	27403	753
20354	349	27485	82
20425	71	28332	847
20559	134	29275	943
20585	26	29936	661
21150	565	30128	192
21186	36	30455	327
21226	40	30914	459
21878	652	31243	329
22046	168	32155	912
22055	9	33088	933
22336	281	33809	721
22421	85		

Analyzing this information, we can affirm that with five groups we can work. These groups are:

- Minimum to 17000
- 17000 to 22000
- 22000 to 25000
- 25000 to 28000
- 28000 to maximum

4. Static representation

Our static representation is based on a traditional map making, representing temporal data.

Cartographic representation methods

As we have said in the last chapter, for the static representation we have a polar or start graph for the number of companies and a pie or Doughnut graph for de GDP, circle divided in same parts. We have three years in this map: 2008, 2012 and 2017. And we use standardized values for the number of companies.

We have to say the radius of the Doughnut symbol or the size of the star graph is not relevant. For instance, the radius of the Doughnut is adjusted to the star graph, if it is possible. The size of the star graph is built depending on the number of companies in each group.

Colors

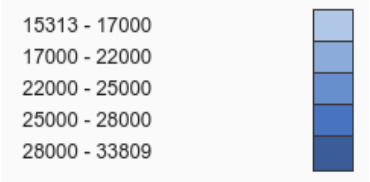
On the one hand, we have ordered the companies classes by Intensity, with the same Hue and Saturation. The oldest has the darkest, and from lighter to darker, because the Intensity has the perceptive property ordered (Irigoyen, 2012). Colors we have chosen are in HSL directive, and we have changed the Intensity depending on the year:

- 2008 has hsl (340, 100%, 70%).

- 2012 has hsl (340, 100%, 45%).

- 2017 has hsl (340, 100%, 30%).

On the other hand, we have ordered GDP with different blues by Intensity also, but in classified ranges. Following the HSI directive, we have 210º in hue, 50% of saturation and a difference between 80% and 40% in intensity, 80% for the first class (minimum to 1700) and 40 for the latest class (28000 to maximum). The legend of these values would like the right image.



Base layer

The base layer we have used is a GeoJSON with Spain, *France*, *Portugal*, *Andorra* and a part of *Africa*. This vectorial layer[2] comes from Natural Earth, a community effort to make visually pleasing, well-crafted maps with cartography or GIS software at small scale. Of course, we have discarded all the remaining countries. Also, we have generalized the polygons: we need less detail.

2 Data obtained by http://datahub.io/docs/core-data

Map elements

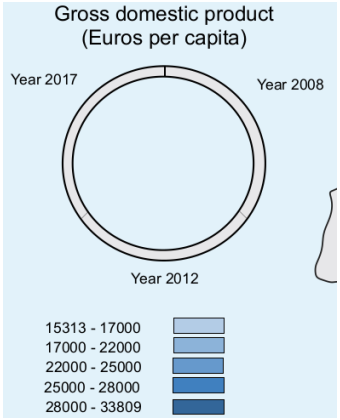
The basic elements the mapmaker has to work with are the subject area, the title, the legend, the scale indicator, the graticule or north arrow, supplementary text, frame/border and insets (Judith A. Tyner, 2010). In this representation we have some important elements from these:

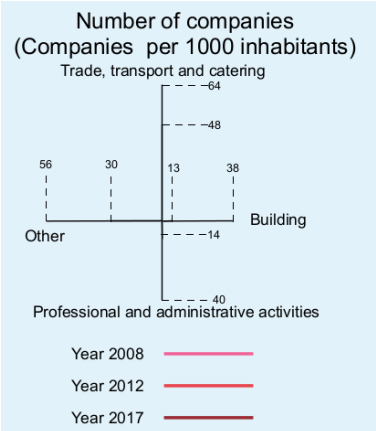
- Title. The title is "*Relation between companies and Gross Domestic Product in Spain in 2008, 2012 and 2017*".
- Numerical and graphical scale. The scale for this view is 1:5000000
- Author and data source as text.
- Projection. The projection used is Lambert azimuthal equal-area projection.
- Legend of both kinds of data.
- Inset map showing *Canarias.*
- Map of *Spain*.

Approach

This representation has been built with Quantum GIS V3.4 Madeira[3]. With the "*Layout Manager*" we can edit a representation in an A4 page format, add the map done with the layers and we can add scale, titles, etc. Moreover, all the symbols have been done with JavaScript as SVG (Scalable Vector Graphics). These symbols are introduced in the different styles for each region. Therefore, the legend has been created from these symbols:

- For Gross Domestic Product we have a circle without a specific radius, changing color by means of classes. In the legend we show a circle divided in three equal parts (2008, 2012 and 2017), and the five classes with their colors and ranges, as we can see in the right image. In the right legend we can see the minimum and maximum values, and the different natural breaks from the data classification.





- For Companies number we have star graphs showing the quantities and depending on the year. In the left legend we show a cross by axes representing the star for the groups. As we can see in the left image, there is the minimum and the maximum value for each group, and the colors for each year.

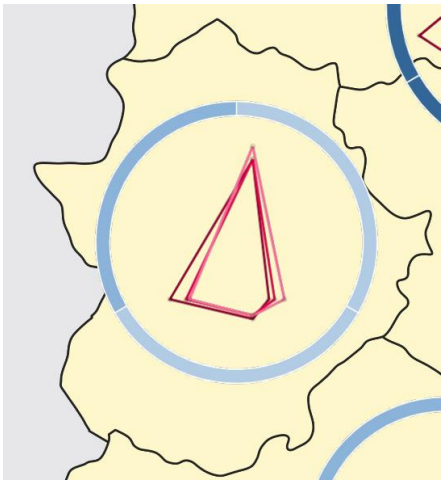
3 QGIS page: https://qgis.org/en/site/

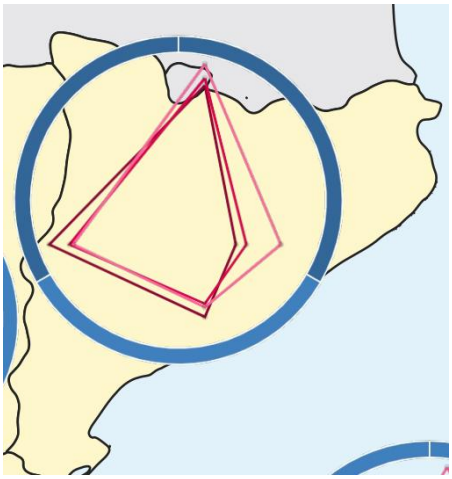
Result

The static map representation can be seen in the Annex. With this result we have noticed Spanish crisis is seen a lot in the GDP variable: in many regions it is clearly seen a low value in 2008 and a high value in 2017. As we said at the beginning, *Spain* experienced a decrease in GDP since 2008 and then began to recover it on 2014.

In addition, we have to say in some regions the star graph is longer and smaller than the others, with a low GDP, such as *Extremadura* and *Andalucía*. However, in other regions this symbol is fatter and bigger than the previous one (and less elongated), with a high GDP.

Also, we can see two activity groups that decrease through time, like Trade, Transport and catering and Building, and the other two increases through time. As an example, we have in the left image *Extremadura*, with low values on GDP and companies, and we have in the right image *Cataluña*, with high values on GDP and companies.





5. Dynamic representation

Our dynamic representation is based on a web map showing symbols and animations to explain our data and claim. An animated map has a continuous change while the map is viewed. We use some visual variables for animation (Terry A. Slocum et al., 2009):

- Duration: length of time that a frame of animation is displayed. We have ten years.
- Order. Sequence in which frames are presented. Our order is chronological.
- Display date. The time some display change is initiated. We start showing information in 2008 when the user open the page.

Moreover, we intend to emphasize in the change, so we have time series, or an animation that emphasizes changes through time. We have developed a tool that continuously changes the year displayed every second; when in finished in 2017, returns to 2008 and continuous.

Also, we have developed another animated tool to emphasize the location through time. This tool shows us the highest number of companies for each activity group in all the country. The highlighted group change over the years as the higher number changes with the time slider (temporal brushing).

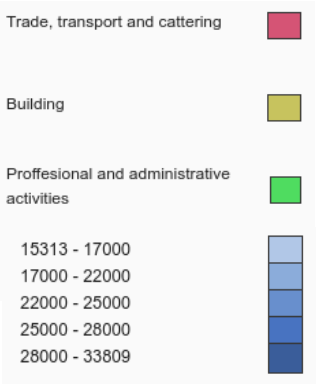
Cartographic representation methods

As we have said in the data chapter, for the dynamic representation we have a pie graph for the number of companies and a choropleth map for the GDP. We have ten years, since 2008 to 2017. Moreover, we have used proportional symbols to represent the whole set of companies; proportional areas as the sum of the companies groups in a region

Colors

On the one hand, we have companies represented by a pie chart. The colors we have chosen are in HSL directive, and for the four groups we have maintained the Saturation and Lightness, changing the Hue, depending on the activity group of companies:

- Trade, transport and catering has hsl (350, 50%, 60%).
- Building has hsl (50, 50%, 60%).
- Professional and administrative activities have hsl (110, 50%, 60%).
- Others have hsl (180, 50%, 60%).



On the other hand, the GDP is showed with the same colors that the static one, but represented with by choropleth map. The legend showed in the right we can see in the dynamic map.

Base layer

The base layer is the same as the static one, a generalized GeoJSON with *France*, *Andorra*, *Portugal*, *North Africa* and *Spain*.

Utilities

Here we have some utilities the user can interact with the map and understand better. In addition to the default controls, such as zoom and full screen, we have made a button to reset the center and zoom of the map. It is true that we have limited pan and zoom for the user, because it is not interesting what is not Spain. Also, we have the scale bar in the bottom right, and the coordinates in meters in the Lambert azimuthal equal-area projection in the bottom left. Also, we have different tooltips to show information to the user, like the definition of Gross Domestic Product per capita when the mouse passes (hover) over that title, or the value of the companies number when passes through the pie chart.

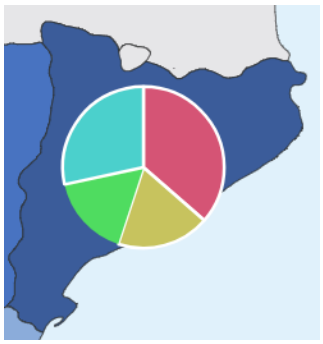
Nevertheless, we also have some tools:

- Change between standardized values and absolute values. As we have said before, we have standardized and non-standardized values for companies' number. So we have a button to change automatically between these particularities.

  On the one hand, we can see with standardized values circles with almost a similar radius, without a big spatial difference. However, as the population plays an important role, there is a little more differences between regions. For instance, the highest value of an activity group is changing over the years.

  On the other hand, with non-standardized values we can appreciate a little bit more the spatial patter and each region with different quantities. For instance, the whole companies set of *Aragón* is much smaller than the whole companies set of *Cataluña*; we have more differences in the absolute numbers of companies

- Show highest values of groups in Spain. This function allows us to see the highest value of an activity group in all Spain. For instance, in 2008 *Cataluña* has three groups with the highest values, as we can see in the right image, with the edges marked in white. Through the years, this highlight is changing and the user can see that region with high companies' number in different sectors. Of course, we can see four highlight per year, like the right image.



- Time slider between 2008 and 2017, as temporal brushing, with buttons: backward, play/pause and forward. This slider works like an animation when the user presses the play button. Automatically, the data is changing according to the year and the active layers. If the time slider is in pause and the user presses the backward and forward buttons to change the year, it interacts more.

Approach

This dynamic representation has been constructed with JavaScript, together with the use of a framework called *AngularJs* (V1.6.9)[4]. We also use the *Material (V1.1.12)*[5] library for styles. Moreover, we have used *Openlayers* (V4.0.1)[6] as web mapping library, and *ChartJs (V2.7.2)*[7] as chart library.

On the one hand, we have made the choropleth style with *Openlayers*, setting the colors for each in a function and depending on the values the function collect from the JSON data.

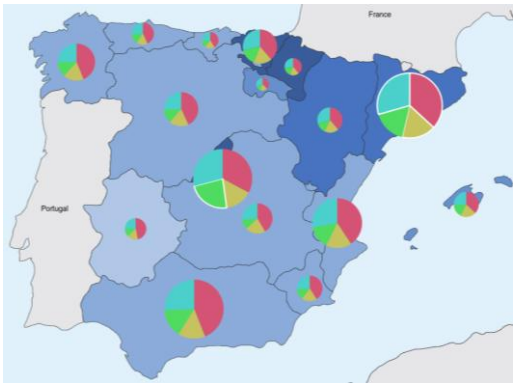
On the other hand, we have done pie charts according to the directives of ChartJs, putting up these in overlays located in the map points of the regions. The radius of each symbol is calculated on the fly (quivering, change in size dynamically): it is calculated the sum of the whole companies (Trade, transport and catering, Building, Professional and administrative activities and Others), and this value is divided by a parameter, resulting the area. The radius is square root of the area divided by **π** (Pi). This radius is introduces in the ChartJs option for each pie symbol.

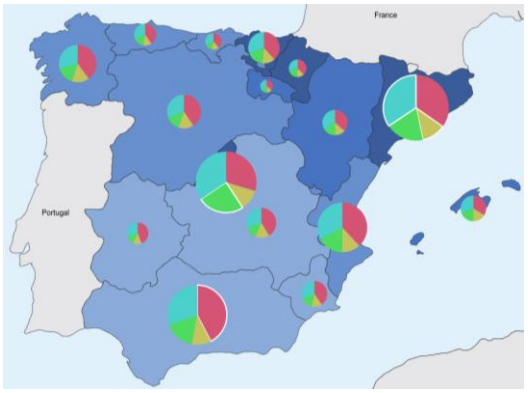
Result

The dynamic map representation can be seen in the following link:

https://carlosbayarri.github.io/TMCompanies/

In this representation, the animations allow us to do a comprehensible temporal analysis, where we can focus on a change in size and properties of the symbols. It is true that with the time slider we can easily follow the change in GDP values: we can see how the regions have lower values in 2008 than in 2017. For instance, here we have the map for 2009 (left) and 2017 (right):





Also, it is useful how change the highest value in the activity groups for the different regions in different years.

4 AngularJs page: https://angularjs.org/

5 Material page: https://material.angularjs.org/latest/

6 Openlayers page: https://openlayers.org/

7 ChartJs page: https://www.chartjs.org/

6. Comparison of both depictions

Both representations are complex and understandable, but with different results.

On the one hand, the static one might be easier to see: we can get an idea about the companies change with the size and form of the start graph. Also, we can see much better the decrease and increase of the number of companies in each group. It could also be thought that when a region has a high GDP, also has many companies in most cases, and vice versa.

On the other hand, the dynamic one show us entertaining animations where variables change as a function of time. Thus, GDP values look better in this representation being a choropleth map, we can see how it evolves. Nevertheless, standardized values do not contribute much in companies because the circles are almost equal and we cannot see the difference between sectors. Therefore, without the standardizing values the change of the set is better appreciated. Also, in the dynamic one we cannot see the relation between companies and GDP.

7. Bibliography

Bankinter. Blog de referencia financiera. (2017). PIB per cápita de España por Comunidades Autónomas. Retrieved January 19, 2019, from https://blog.bankinter.com/economia/-/noticia/2017/9/21/pib-per-capita-espana-comunidades-autonomas-grafico

Fund, I. M., Olympics, A., & Union, E. (2013). Eurozone crisis explained. *BBC News*, (November 2012), 11–13. Retrieved from https://www.bbc.com/news/business-17549970

Irigoyen, J. (2012). *Diseño y Producción Cartográficos*. Universitat Politècnica de València.

Judith A. Tyner. (2010). *Principles of Map Design*.

Lambert Azimuthal Equal Area—Help | ArcGIS for Desktop. (n.d.). Retrieved January 19, 2019, from http://desktop.arcgis.com/en/arcmap/10.3/guide-books/map-projections/lambert-azimuthal-equal-area.htm

Slocum, T. A. (1999). *Thematic Cartography and Visualization* (First edit).

Terry A. Slocum, Robert B. McMaster, Frits C. Kessler, & Hugh H. Howard. (2009). *Thematic Cartography and Geovisualization* (Third edit).

8. Annex

Static map representation

(Map page)