

DATA VISUALIZATION

Definition, technologies and tools



GOBIERNO
DE ESPAÑA

MINISTERIO
DE INDUSTRIA, ENERGÍA,
Y TURISMO

MINISTERIO
DE HACIENDA
Y ADMINISTRACIONES PÚBLICAS



INDEX

1. Data visualization

1.1. Visualization: data exploitation

1.1.1. Types of data visualization

1.2. Data visualization and Open Data

2. Data visualization software

2.1. Web Standards

2.1.1. HTML5 Standard and Canvas

2.1.2. SVG Standard

2.1.3. WebGL Standard

2.2. Libraries and framework for visualization

2.2.1. Libraries for 2D graphs

2.2.2. Libraries for 3D graphs

2.2.3. Libraries for maps

2.2.4. Other types of visualizations

3. State of art in tools and platforms for data visualizations

3.1. General description

3.2. Examples of tools

3.2.1. Tableau Software

3.2.2. Tabulae

3.2.3. QlikView and QlikSense

3.2.4. CartoDB

3.2.5. Socrata

3.2.6. CKAN (Visualization units)

4. Examples of data visualization

- 4.1. Datos.gob.es
- 4.2. World Bank
- 4.3. Transparency portal of Principality of Asturias
- 4.4. openREDBAG Platform
- 4.5. Google Public Data
- 4.6. Eurostat
- 4.7. Fundación Civio
- 4.8 Other examples of visualizations
 - 4.8.1. Evolution of the Web
 - 4.8.2. Global warming
 - 4.8.3. Retirement of Kobe Bryant
 - 4.8.4. History of music

5. Conclusions

- 5.1. Future trends

1. DATA VISUALIZATIONS

We live in a world surrounded by data that our brain processes continuously to construct reality, understand the environment around us and make decisions about our future. At the present time our information consumption has increased exponentially due to two factors: **more and more information is produced** (social networks, devices, etc.) and we have increasingly **more capacity to access such information**, especially via the Internet and the web. The ability to exploit and understand raw information is closely linked to our ability to exploit and transform it into something more than pure data: the **data acquires meaning**.

However, the data, understood as single records, do not provide a specific meaning. Only when we approach it and apply an interpretation does it make sense and become knowledge. In the field of technology, data mining has evolved in recent decades to design interpretation mechanisms that are increasingly robust and affordable. And among these exploitation mechanisms, the most important is **data visualization**.

1.1. Visualization: data exploitation

Data visualization is the graphical presentation of information for two purposes. On the one hand, interpretation and construction of meaning from the data (i.e. analysis); and on the other hand, communication.

Visualization is a powerful tool to discover and understand the logic behind a set of data and to share this interpretation with others from an objective point of view. As they say, "a picture is worth a thousand words", especially when the meaning you want to communicate is represented better graphically than verbally, and this graphical representation is designed according to formal principles for data visualization.

We can stare for hours at a table of numbers and not see what is immediately obvious when this information is presented graphically. For example, imagine the following sets of historical sales data of mortgages in Spain (INE Source: 2003-2011 series), in which the sale volume of mortgages are broken down by year and financial institution:

| Period | Entity | Mortgages |
|--------|----------------|-----------|
| 2010 | Other entities | 103.055 |
| 2008 | Other entities | 122.332 |
| 2009 | Bank | 367.298 |
| 2011 | Bank | 258.667 |
| 2008 | Bank | 408.712 |
| 2010 | Bank | 337.825 |
| 2003 | Savings bank | 685.464 |
| 2011 | Savings bank | 279.461 |
| 2006 | Bank | 576.777 |
| 2006 | Savings bank | 1.016.728 |
| 2004 | Bank | 583.213 |
| 2009 | Other entities | 111.066 |
| 2005 | Bank | 594.477 |
| 2003 | Bank | 486.579 |
| 2004 | Savings bank | 791.349 |
| 2003 | Other entities | 119.617 |
| 2005 | Savings bank | 924.882 |
| 2007 | Bank | 518.638 |
| 2008 | Savings bank | 683.546 |
| 2007 | Savings bank | 1.010,548 |
| 2011 | Other entities | 86.420 |
| 2007 | Other entities | 154.809 |
| 2009 | Savings bank | 549.558 |
| 2010 | Savings bank | 475.536 |
| 2005 | Other entities | 184.773 |
| 2004 | Other entities | 147.843 |
| 2006 | Other entities | 202.889 |

Although we understand the data shown in the table, for the end user it will be much easier to interpret such information if a graph is created with the evolution of the property market in Spain and the bubble effect, as shown in the following figure.

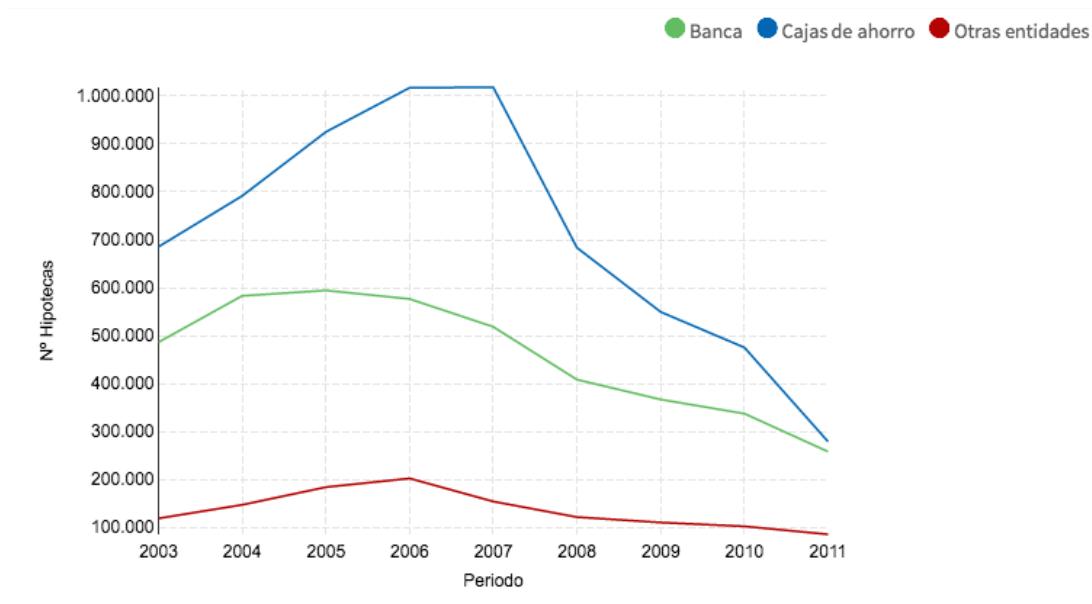


Fig. 1. Historic range of mortage sales in Spain 2003-2011 (Source: INE)

What numbers cannot communicate when they are presented in a table becomes visible and intelligible when they are communicated visually. **This is the "power" of data visualization.**

It is important to note that while data visualization is used to generally represent quantitative variables and relationships between them, it can also be used to represent relationships between entities of a qualitative nature. For example, relations between people of a certain social network, which may be also "typecast" according to the nature of this relationship: friendship, family, work, etc. These visualizations representing entities and relational properties are based on the typology of the structure to be represented and use graphs based on nodes and arcs.

Historically visualization has existed consubstantially with data. Especially in the field of cartography. However, it is the **late eighteenth century and early nineteenth century when the first studies and applications of data visualization appear** in order to construct narratives and understand real phenomena: from economic indicators to historical events. In this regard, we must highlight the pioneering work of Scottish economist William Playfair and his book: *The Commercial and Political Atlas and Statistical Breviary*.

Playfair is considered to be the founder of the application of graphical techniques for statistical analysis, inventing graphics such as lines, areas (as in Figure 2), bars and pie charts.

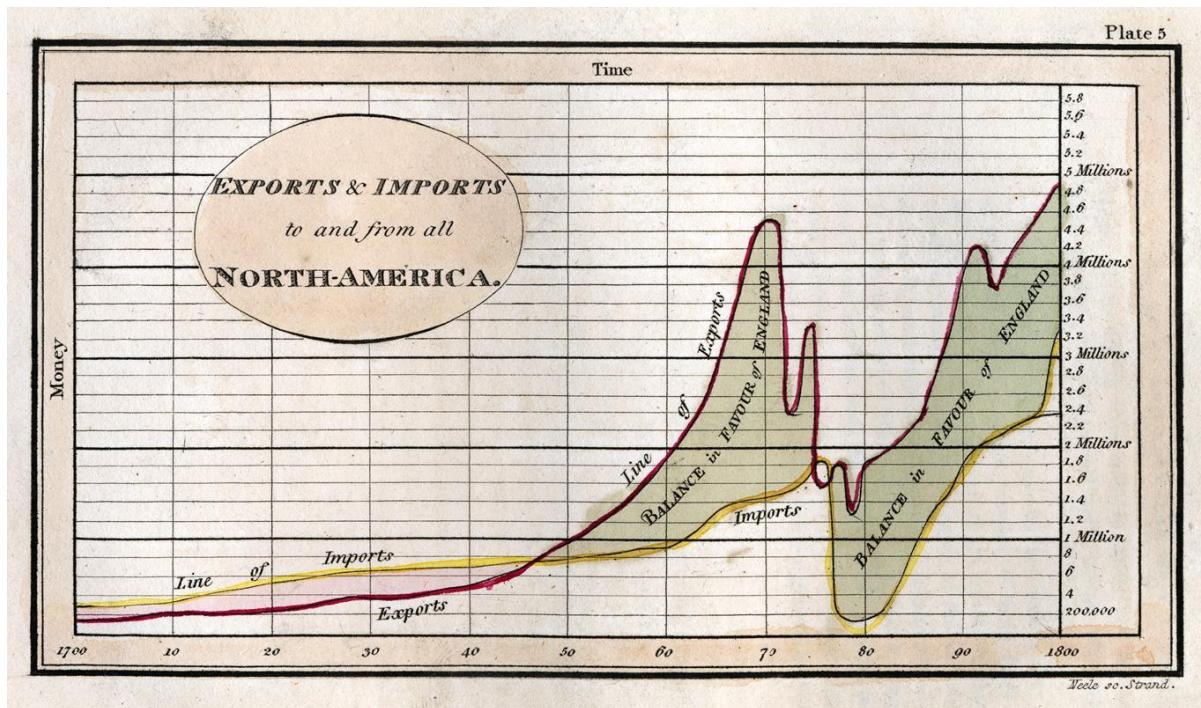


Fig. 2. Historic range (1700 - 1800): trade balance bewteen England and North America

Playfair's innovative work was accompanied by the work of other engineers and economists of the time, as is the case of Charles Minard and his famous graphic narrative about Napoleon's march on Russia in 1812, M.H.P.R. Sankey diagrams and flow charts, Jon Snow and Charles Dupin and the use of thematic or Choropleth maps to represent quantitative variables in cartography (see Figure 3), and so much more.

Already in the twentieth century, we can highlight two authors of reference: Tukey¹ and Tufte² and their respective works, which established the formal principles of visualization. Tukey established what is known as EDA (Exploratory Data Analysis) and explained how to use visual methods to understand the data and formulate hypotheses without using statistical models.

¹ Tukey, John W (1977). *Exploratory Data Analysis*. Addison-Wesley.

² Tufte, Edward R (2001) [1983], *The Visual Display of Quantitative Information* (2nd ed.), Cheshire, CT: Graphics Press

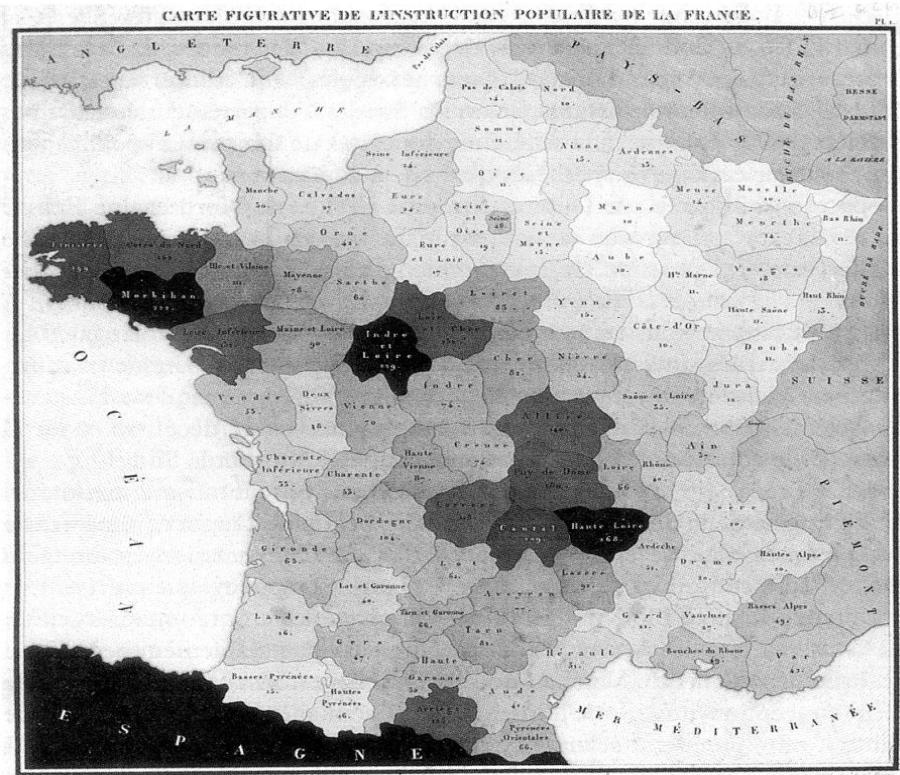


Fig. 3. Dupin representation of population distribution in France in 1826

Nowadays, as we shall see in the following sections of this report, there are **many tools to generate graphs and maps** in an assisted way that allow value to be built on the data. Although visualization can still be a traditional process based on design and manual processing, as some examples of infographics are, in most applications specific software is used for the construction of these graphical representations.

1.1.1. Types of data visualization

There are any number of techniques and approaches for visualization depending on the nature of the data information. From the point of view of the data, especially structured data (or semi-structured) and its visual exploitation, we can establish roughly the following classification³ of **types of visualization according to complexity and information processing.**

³ This classification is not base don academic criteria. It is intended for illustrative purposes only.

1. Basic elements of data representation

This is the simplest case. A basic element of visualization can be a chart, a map, a KPI, data tables, a graph, etc. Here there are some basic types of visualizations:

- **Graphs:** bars, lines, bullet points, “tree maps”, pie charts, semi-pie charts, etc.
- **Maps:** bubbles, choropleths (or thematic map), heat map, aggregation (or drill-down analysis).
- **Tables:** with nesting, dynamic, drill-down, transitions, etc.

2. Dashboards

A dashboard is a complex composition of single visualizations that have a coherence and a thematic relationship among them. They are widely used in organizations to analyze groups of variables and decision making.

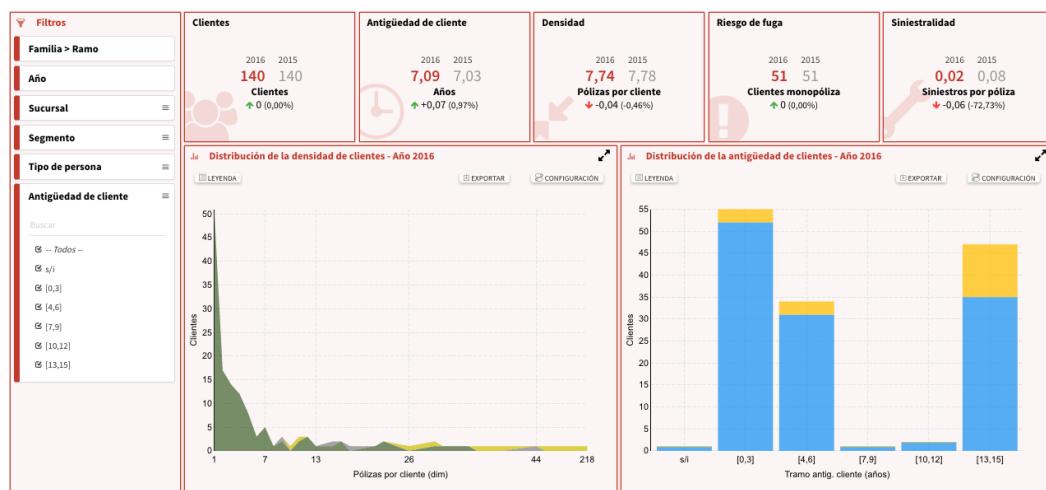


Fig. 4 –Example of business dashboard

3. Infographics

An infographic is also a composition made of visualizations that builds up, from different elements, a complex meaning for the user. However, infographics are not intended for the

analysis of variables but **for the construction of narratives from the data**; i.e. infographics are used to tell "stories".

This narrative is not built through text, but by providing the information in a way that the visualizations are combined with other elements such as symbols, captions, drawings, synthetic images, etc.

An infographic has a very high component of manual development and design in its construction.

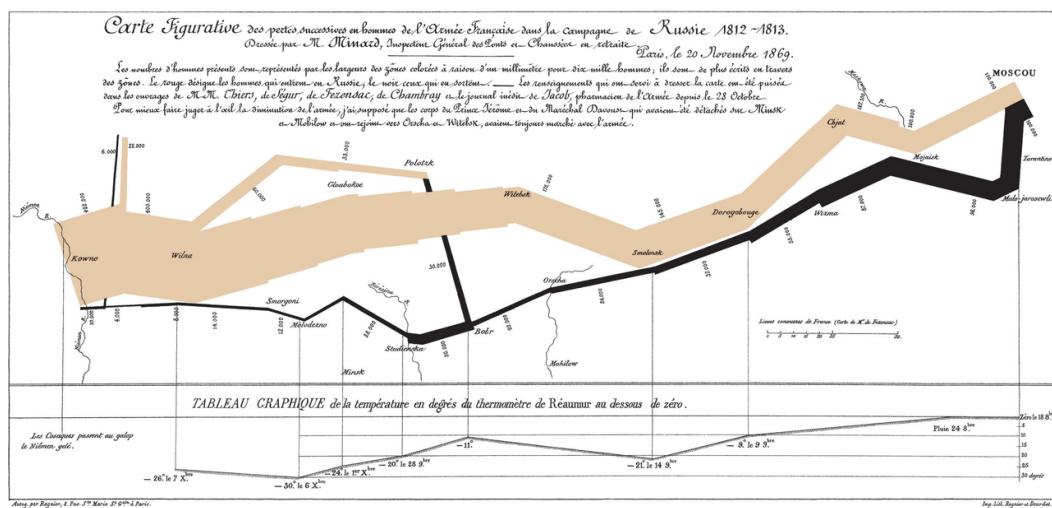


Fig. 5 – March of French army in Napoleon's campaign to invade Russia (C. J. Minard). This visualization can be considered the first infographic in history.

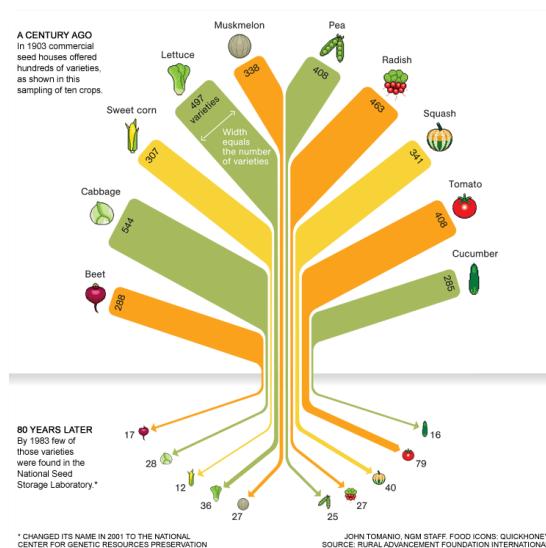
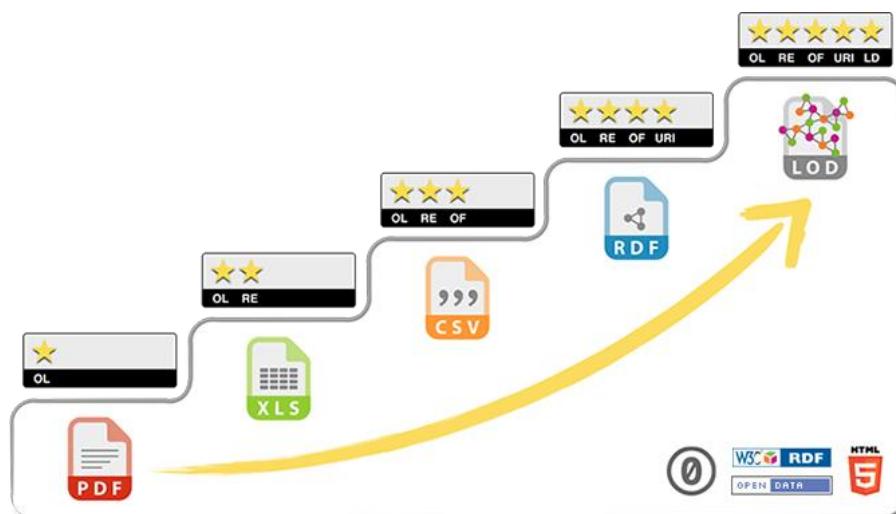


Fig.6 - Changes in alimentary diet in the last century.

1.2. Data visualization and open data

The movement of open data within administrations has focused in recent years on the issues of their publication. On the one hand, policies and regulations have been defined to make authorities open their data to society, both for transparency commitments and seeking a return on the re-use of information through other organizations and the development of a productive system around it.



Graph 4. Open data publication categories (image of 5stardata.info)

On the other hand, [formats and best practices](#) have been defined for administrations to know not only what to publish, but how and when, so that search, exchange and access to information processes obey the principles of standard design to facilitate the **automated re-use of open data**.

In this regard, we can highlight the cataloging of the different levels of data publication and the identification of good practices in re-use (for example, the 5 star distinction of Tim Berners-Lee⁴).

Organizations such as the W3C, the Open Knowledge Foundation (OKFN) or the World Wide Web Foundation have been decisive for evangelization on the need to open up public data and promote technical work on their publication. In Spain in particular and Europe in general, this has resulted in numerous initiatives and projects as well as the emergence of specific legislation. The European Union very clearly is actively promoting the harmonization of these dynamics through

⁴ Tim Berners-Lee (2006): Linked Data - Design Issues. <https://www.w3.org/DesignIssues/LinkedData.html>

concrete [technology proposals such as DCAT-AP](#) and promoting business impact through funding programmes such as H2020.

Nevertheless, one of the most common criticisms⁵ about open data is that the current effort is focused on publishing data and not on its **usability**; i.e. how this data will be consumed by the end user. Many open data projects have been overly focused on these technical issues: formats, endpoints, etc., unfamiliar to the potential user, and they have not paid attention to how this data can be used and value can be extracted from it. This has caused many portals to become mere repositories of data, with marginal traffic and with a very low social impact.

Beyond a critical valuation in this regard, which is completely beyond the scope and intent of this document, it is crucial to understand that it is natural for the first steps in the open data movement to be focused precisely on the definition of what open data is and what the publication procedure for it is. Right now, however, both the available technology and the maturity of the movement as well as the demand from citizens need to take a further step and evolve the open data concept to a wider dimension in which not only issues related to publication and re-use from the technical point of view are taken into account.

The next milestone is to **provide tools for users to be able to consume and exploit data independently**, making real the initial objective that information from administrations should have a real positive impact on our society. In this respect, visualization, as mentioned before, is the most powerful tool to bring data to any user.

Currently visualization is part of any standard management process and data analysis in the business world, especially when referring to issues related to statistical analysis and "Business Intelligence" (hereinafter BI), although not exclusively. In any case, we can define these processes as a **set of techniques and tools for the acquisition, processing and transformation of raw data** into useful information and knowledge for a particular purpose of analysis.

In the field of business, the purpose is usually linked to processes and business aspects, but the application of techniques and tools is generic with respect to the intent of the analysis. In these

⁵ Helbig, N., Cresswell, A.M., Burke, G.B. and Luna-Reyes, L. (2012) *The Dynamics of Opening Government Data: A White Paper*. Centre for Technology in Government, State University of New York, Albany. <http://www.ctg.albany.edu/publications/reports/opendata/opendata.pdf>

cases, visualization is part of the life cycle of the data, in which, after preparation and processing, the information is exploited by the end user through graphical representations that allow the meaning of the data to be interpreted.

The open data movement can reuse both the technology available on the current market, and the lessons learned in this area in the last fifteen years. It is not a question of starting from scratch, but of applying the knowledge gained in these years to make a qualitative leap and ensure that data acquires the usability expected by the citizenship.

2. DATA VISUALIZATION SOFTWARE

This section describes the web standards that have been developed in recent years for the development of web applications, essential for creating web-based visualizations based on data. In turn, different JavaScript libraries that use these standards for the creation of the fundamental parts of the display are also described.

2.1. Web Standards

In recent years, the development of web content and applications has undergone a revolution thanks to the new web standards. Since in 2014 the W3C published the final version of the HTML5⁶ standard, new standards have been added that allow developers to create not just static displays but powerful visualization applications which include a high degree of dynamism and aesthetic customization. Standards like HTML5 or the new version of standard Cascading Style Sheets (CSS3⁷) together with the evolution of modern browsers allow developers to create multi-device web applications without having to worry, as often happened, what browser or device is used to see the visualization.

| Standard | Last version | Funtion |
|----------|--------------|--|
| HTML5 | v5 | Canvas: HTML element to draw 2D graphs |
| CSS3 | v3 | Allows differentiation of the content of websites from the presentation of that content. |
| SCV | v2 | Used to make 2D graphs |
| WebGL | v1 | 3D graphs using Canvas |

⁶ <https://www.w3.org/TR/html5/>

⁷ <https://www.w3.org/Style/CSS/>

What follows explains in detail each of the above in detail:

2.1.1. HTML5 Standard and Canvas

In 2014 the W3C published version 5 of the most important web standard: **HTML5**, ending six years of development in collaboration with leading technology companies. **HTML is the basic language on which all contents are built.** This new version was a qualitative leap in this standard including new APIs for interaction with the browser and device, and most importantly, forcing all web browsers to be updated to support this new version.

This resulted in companies responsible for each of the modern browsers being forced to improve the compatibility of their products with this new version. Thus **homogenization of web content** was provided and, at least partially, it put an end to the coding "hacks" necessary to correctly display the contents in each of the browsers.

Within this new version HTML5 an important element for the visualization of content was defined, Canvas. This HTML element is used to draw graphics, typically 2D, but it can also be used along with WebGL 3D to visualize graphics, using scripts, usually written in JavaScript. Canvas can be used to **draw graphs, make photo compositions and animations**. There are many JavaScript libraries that facilitate the creation of graphics for this item.

2.1.2 CSS3 Standard

Cascading Style Sheets or CSS (Cascading Style Sheets) is a language that defines the presentation of a structured document such as HTML or CSS. This language is a standard of the W3C consortium to differentiate the contents of websites from the presentation of the content. After many years of development, in 2011 version 3 of this language was published, a breakthrough in terms of power and functionality.

Due to its modular definition, not all elements of the CSS3 language have the same level of technological maturity. Regarding data visualization, the most important modules are the following:

- **Backgrounds and colors:** Makes it possible to add multiple wallpapers, as well as new elements that facilitate better control of their sizes and positions.

- **Animations:** Evolution of animations that can be written for different language elements.
- **Positioning:** This module is an evolution of the classic positioning that allows the elements to be distributed easily and more flexibly.

2.1.3. SVG⁸ Standard

Resizable Vector Graphics or SVG are a standard specification for describing two-dimensional vector graphics, both static and animated (the latter using SMIL) in XML format. In 2001 the W3C made them into a recommendation and they became **one of the most used standards to create 2D graphics**. This standard allows us to define basic concepts such as point, line, polygon, then to combine them and form the complex vector graphic. It is especially useful for creating statistical charts components that allow the visualization and analysis of data sets.

Currently version 2 of SVG is in development and promises to add a comprehensive set of features that facilitate and expand the use of the new features of modern browsers and web standards.

2.1.4. WebGL⁹ Standard

WebGL is a standard that came initially from Mozilla but now includes many large technology companies. This standard allows 3D graphics and animations to be defined using the Canvas HTML5 component. WebGL is based on OpenGL ES 2.0 and provides an API for 3D graphics. It is now widely supported by browsers and is becoming the most used language for web 3D visualizations.

As WebGL is a technology designed to work directly with the graphics processor of the device it is quite difficult to code compared to other more accessible web standards. Multiple JavaScript libraries that facilitate the use of this specification have appeared in order to resolve this.

⁸ <https://www.w3.org/Graphics/SVG/>

⁹ <https://www.khronos.org/registry/webgl/specs/1.0/>

2.2. Libraries and framework for the visualization

2.2.1. Libraries for 2D graphics

There are many libraries that facilitate the creation of 2D graphics for data visualizations. Within this group we can distinguish those based on Canvas, and those based on SVG.

| | Standard | Graphic palette | Final finishing | Ease of use | Extensible | Modality of use |
|----------------------------|----------|----------------------------|-----------------|-------------|------------|-----------------|
| D3.js | SVG | Highly extensible | Good | Difficult | Yes | Free |
| Google Charts | SVG | Very wide | Good | Simple | No | Free |
| Chart.JS | Canvas | Lines, bars, radar and pie | Very good | Simple | No | Free |
| HighchartsJS | Canvas | Muy amplia | Very good | Simple | No | Paid/ Free |
| JavaScript InfoVis Toolkit | WebGL | Quite reduced | Regular | Difficult | No | Free |

2.2.1.1. D3.js¹⁰

D3.js is a JavaScript library for manipulating data based documents. D3 helps give life to data using HTML, SVG and CSS. The key point of D3 is the use of these web standards to enable the publisher / developer to exploit all the new capabilities of modern web browsers, making DOM manipulation easier to create powerful data-driven visualizations.

D3 gives us an abstraction layer above SVG offering a number of utilities for data manipulating and later visualization. Thus, D3 offers the **creation of interactive graphics that can be customized using standards with CSS3.**

¹⁰ <https://d3js.org/>

For these reasons, D3 has become one of the **most commonly used libraries** not only to create visualizations but as a basis of a large set of more specific libraries based on this library.

Some examples are n3-charts, NVD3¹¹, etc. This type of packaging minimizes the main problem we find in D3, genericity. As it is a generic library for 2D graphics, it is complicated to use. Therefore libraries such as NVD3 offer Javascript classes for the main types of graphs, such as: bar graph, line, area, bubble, etc., facilitating the creation of such visualizations.

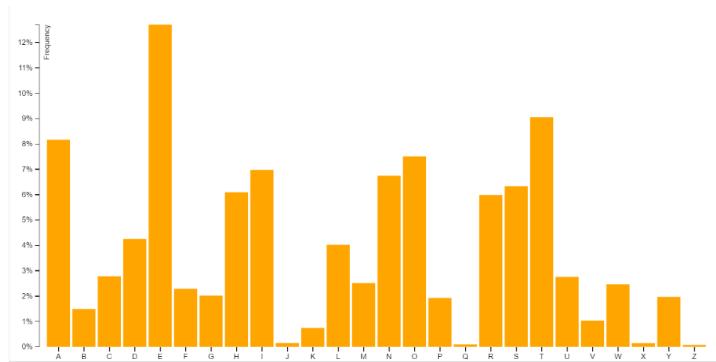


Figura 5 - Ejemplo de gráfica de barras de D3

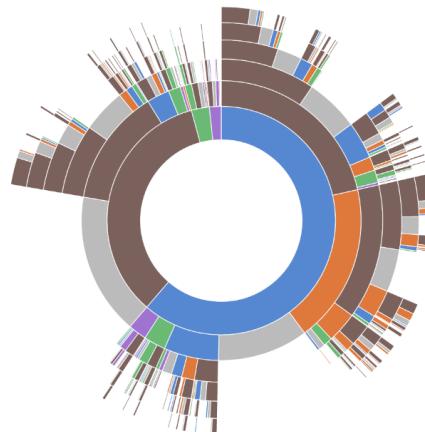


Fig. 6 – Example of more complex 3D visualization

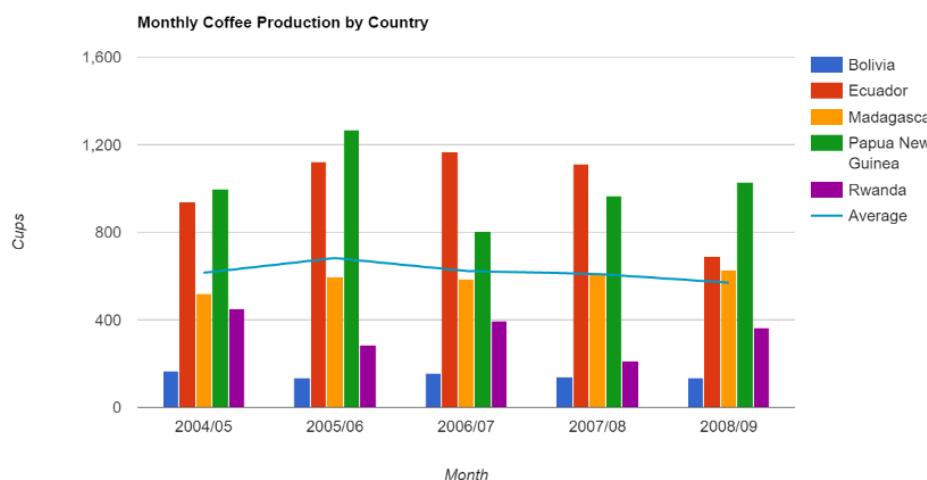
2.2.1.2. Google Charts¹²

Google has invested heavily in data visualization and has created its own library for creating visualizations. This library is also based on SVG and has a large number of types of charts and maps. It also has what they call dynamic data that allow us to make SQL queries on data stored

¹¹ <http://nvd3.org/>

¹² <https://developers.google.com/chart/>

in Google Spreadsheets, Google Fusion Tables or SalesForce and visualizations. Google also provides, albeit limited, some components for creating interactive dashboards.



Graph 7- Example of Google Charts graphic

2.2.1.3. Chart.JS¹³

Chart.JS is the first of the included examples using Canvas instead of SVG. This type of libraries offers JavaScript classes to draw the graph in the new HTML5 Canvas element. The use of Canvas instead of SVG hinders the interoperability of graphics created with standards like CSS3 but has the advantage of requiring fewer computational resources for the web browser to represent them:

¹³ <http://www.chartjs.org/>

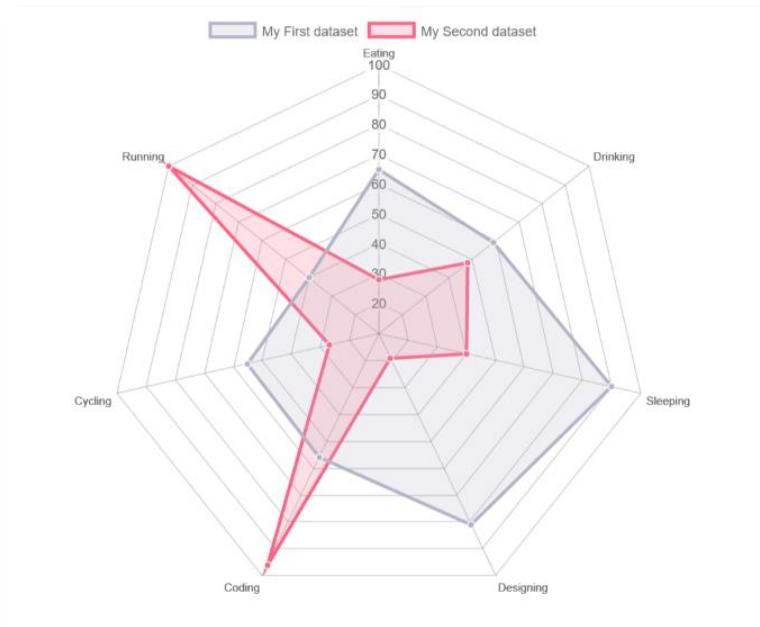


Fig. 8 - Example of graphic created with Chart.JS

2.2.1.4. HighchartsJS¹⁴

HighchartsJS is a free commercial library, for noncommercial use, which has a large number of followers. This library also uses the SVG standard for graphics. As a main point in its favor, **it has a large number of components and a very thorough documentation**. The main disadvantage is the license which has to be acquired for commercial purposes.

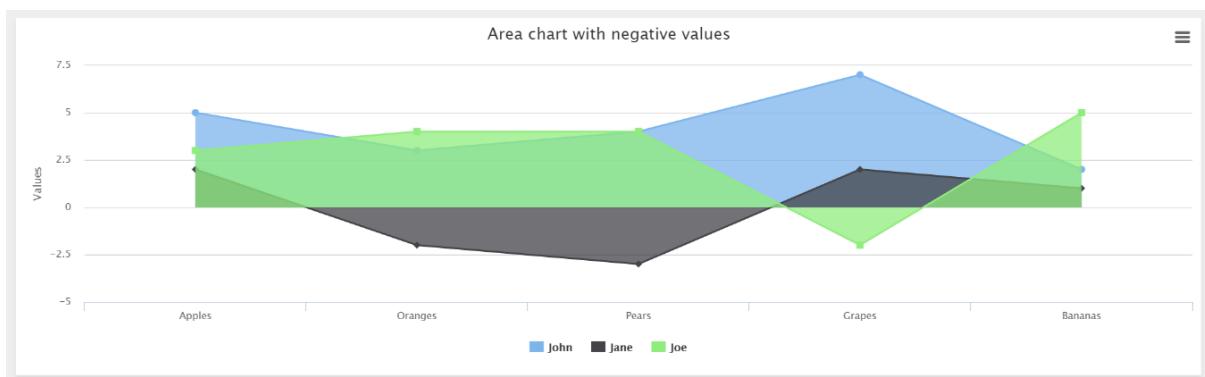


Fig. 9 - Example of graphic created with HightchartsJS

¹⁴ <http://www.highcharts.com/>

2.2.1.5. JavaScript InfoVis Toolkit¹⁵

Example of graphic library that uses the WebGL standard. This library is developed by Nicolas Garcia Belmonte copyrighted by Sencha Company and it incorporates some 3D aspects to the types of most commonly used graphics. This type of feature allows interactive maps to be created offering spectacular 3D animations.

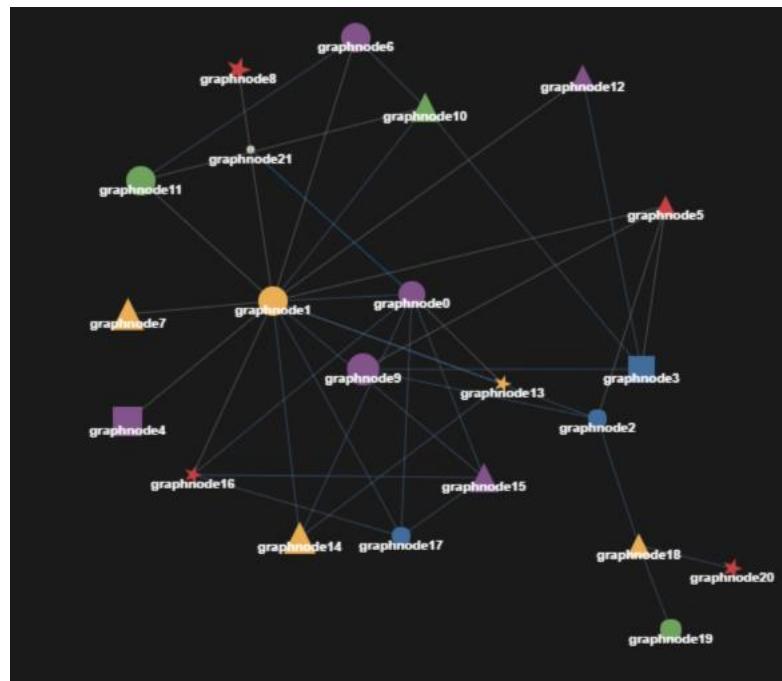


Fig. 10 - Example of graphic created with JavaScript InfoVis Toolkit

¹⁵ <http://philogb.github.io/jit/>

2.2.2. Libraries for 3D graphics

There are multiple libraries used to create WebGL 3D graphics, all of which are aimed at creating HTML5 animations and games but are very useful to create visualizations with high impact.

| | Purpose | Standard | Graphibs and 3D animations | Dependences |
|------------------|----------------------|-----------------|----------------------------|-------------|
| Three.js | General | WebGL | Yes | No |
| BabylonJS | Animations, games | HTML5, WebGL | Yes | Hand.js |

2.2.2.1. Three.js¹⁶

Three.js is a library to build 3D scenes with WebGL. We can see it with a WebGL 3D, ie, a library that eliminates the complexity of writing a visualization in WebGL. It offers all the necessary classes for common aspects of 3D scenes such as lights, cameras, materials and objects, etc.

Fig. 11 - Example of animation created by Google using Three.js



2.2.2.2. BabylonJS¹⁷

Framework to create 3D games with HTML5 and WebGL which can be also used to create spectacular visualizations. As with Three.js it provides the necessary kinds of camera, light and figure, so it frees us from the complexity of WebGL.

¹⁶ <http://threejs.org/>

¹⁷ <http://www.babylonjs.com/>

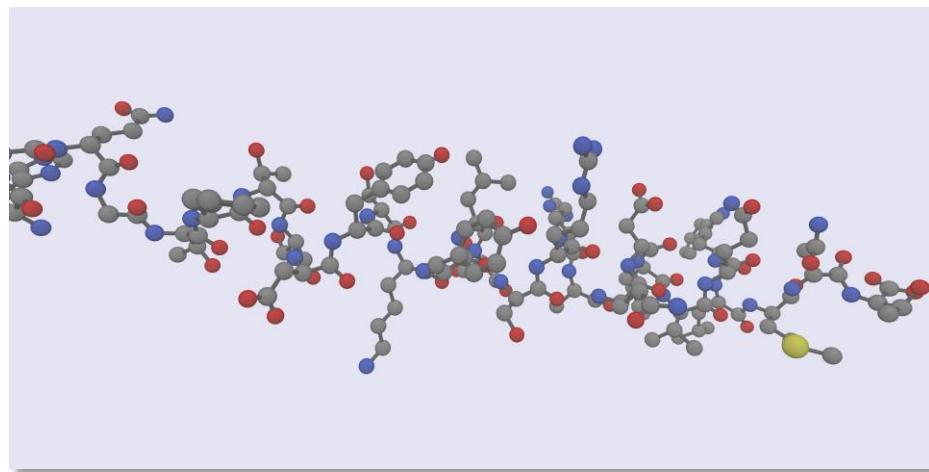


Fig. 12 - Example of 3D visualization with BabylonJS

2.2.3. Libraries for maps

With the rise of mapping services like Google Maps, OpenStreetMap and GIS (Geographic Information Systems), multiple libraries for visualization and analysis of geo-referenced data have appeared. Many of these libraries and clients are based on standards defined by the Open Geospatial Consortium (OGC)¹⁸ and WMS services (Web Map Service) and WFS (Web Feature Service). It should be noted that GeoJSON¹⁹ vocabulary is used as the de facto standard for describing geo-referenced layers and objects in the web environment. This vocabulary defines the basic concepts of point, path and polygon allowing the combination for describing the visualization layers. There are multiple libraries for the visualization of geo-referenced data, including: polymaps.js and leaflet as two of the most commonly used.

| | Vector elements | Layers GeoJson | Connection to WMS layers | Connection to WFS layers | Extensible (plugins) |
|--------------------|-----------------|-------------------|--------------------------|--------------------------|----------------------|
| Polymaps.js | Yes | Yes | No | No | No |
| Leaflet | Yes | Yes | Yes | Yes | Yes |

¹⁸ <http://www.opengeospatial.org/>

¹⁹ <http://geojson.org/>

2.2.3.1. Polymaps.js²⁰

Polymaps is a JavaScript library for creating dynamic and interactive web maps. It provides mechanisms to visualize vector data layers used on the most commonly used maps such as OpenStreetMap, CloudMade, Bing, etc.

Polymaps uses the SVG standard to create these layers from, for example, data defined with GeoJSON, also allowing further customization through the definition of visualization rules written in CSS.



Fig. 13 – Example of bubble map created with Polymaps

²⁰ <http://polymaps.org/>

2.2.3.2. Leaflet²¹

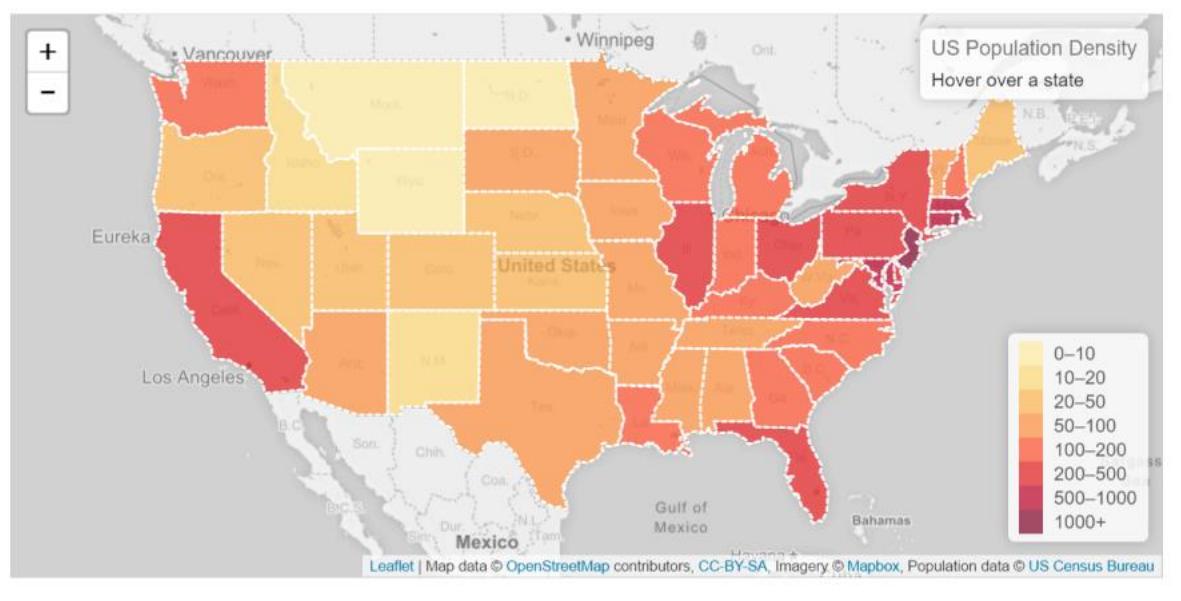


Fig. 14- Example of a Leaflet map

Leaflet is one of the libraries for visualization and creation of interactive maps most commonly used nowadays. It enables the creation of animated layers from multiple data formats, including GeoJSON or WMS layers. It also offers a large set of JavaScript classes to create and modify these layers.

²¹ <http://leafletjs.com/>

2.2.4. Other types of visualizations

Although many of the visualizations are based on graphs and maps, there are other types of visualizations that, depending on the problem, can be very useful to interact with the data. Examples of these visualizations are graphs, time lines or ontology viewers.

2.2.4.1. Sigma.js²²

Sigma.js is a JavaScript library for the visualization of graphs for their use in web environments. This library uses the HTML5 Canvas element for the visualization and provides a set of utilities for the visualization and analysis of graph elements.

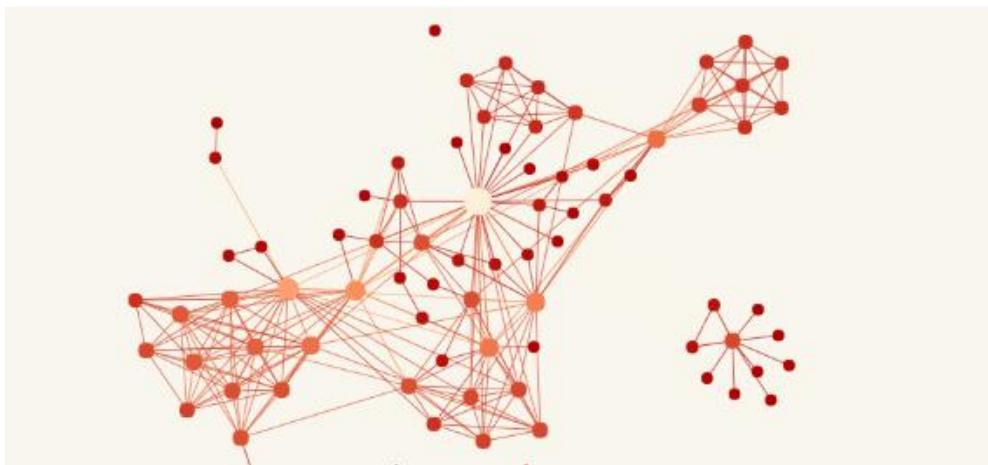


Fig. 15- Example of graph created in Sigma

²² <http://sigmajs.org/>

2.2.4.2. TimelineJS²³

One of the recurring components in visualizations are timelines, that is, the location of a set of events within a timeline. TimelineJS offers a simple way to create a timeline component that allows us to navigate among the events represented.

Nelson Mandela's Extraordinary Life: An Interactive Timeline

By TIME Staff | Dec. 05, 2013 | Add a Comment

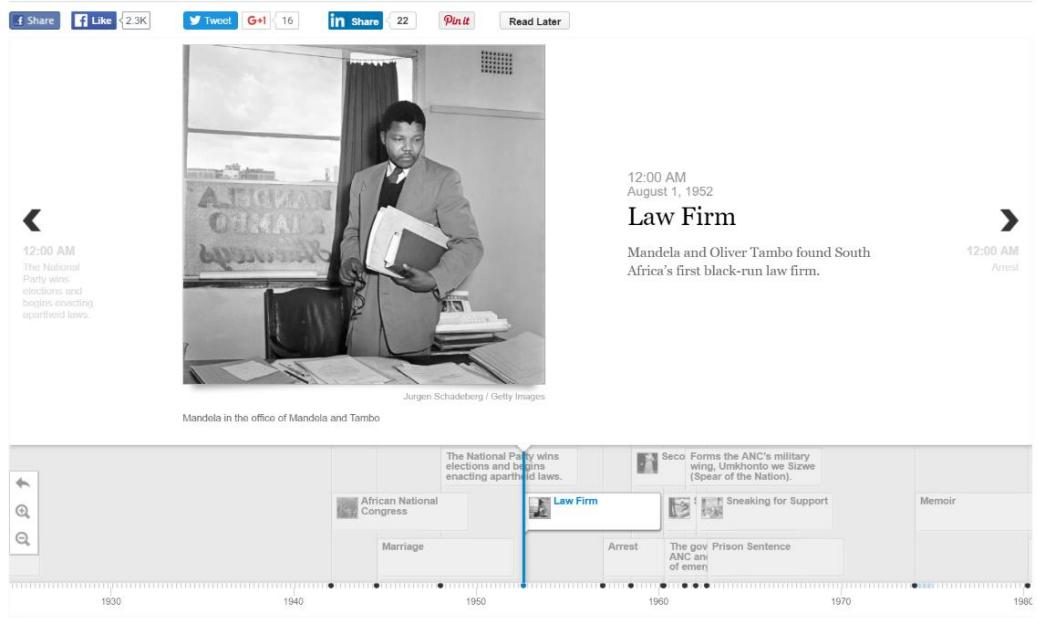


Fig. 16 – Example of visualization on a timeline

²³ <https://timeline.knightlab.com/>

3. STATE OF THE ART IN TOOLS AND PLATFORMS FOR DATA VISUALIZATION

Currently the number of tools and platforms for data visualization that can be accessed is very extensive. Depending on user requirements you can choose from a number of cutting-edge options. Given the magnitude and existing variety, it is difficult to make a rigorous classification. However we can distinguish several distinct groups of well differentiated tools according to the **data management** and the **type of end result** desired:

- **Office tools**- Allows you to move / copy content among different applications.
- **Web** - Orientation to create web content, not files.
- **Analytics** - (Advanced) calculations with data can be performed.
- **Coding** - The user has full control via programming language.
- **Open data** - Guidance to use open data (available on the web).
- **Maps** - Native mapping (data on a base layer).

| | Office tools | Web | Analytics | Coding | Opendata | Maps |
|----------------------|---------------------|------------|------------------|---------------|-----------------|-------------|
| Excel | ✓ | ✓ | | ~ | | |
| Openoffice | ✓ | | | ~ | | |
| Google Sheets | ✓ | ✓ | | | ✓ | |
| SAP BI | | ~ | ✓ | | | ~ |
| SAS | | ~ | ✓ | ~ | | ~ |
| MS Power BI | | ~ | ✓ | | | ~ |
| IBM Gognos | | ~ | ✓ | | | ~ |
| Pentaho | | ~ | ✓ | ~ | | ~ |
| Qlik | | ✓ | ✓ | | | ✓ |
| Tableau Sof. | | ✓ | ✓ | | | ✓ |
| R | | | ✓ | ✓ | | ~ |
| Weka | | | ✓ | ✓ | | ~ |
| Matlab | | | ✓ | ✓ | | ~ |

| | | | | | | |
|--------------------|--|---|---|---|---|---|
| Mathematica | | | ✓ | ✓ | | ~ |
| Maple | | | ✓ | ✓ | | ~ |
| Tabulae | | ✓ | ~ | | | ✓ |
| CartoDB | | ✓ | | | | ✓ |
| Socrata | | ✓ | | | ✓ | ✓ |
| CKAN | | ✓ | | | ✓ | ✓ |

✓=100% ~=50%

3.1. General description

As the beginning of the description or state of the art of data visualization tools, first of all the historically most popular visualization tool must be presented: the spreadsheet -**MS Excel**, **OpenOffice Calc** or **Google sheets**- among others, used daily by millions of people for all kinds of visual representation based on data. Its ease of use and the fact that it often forms part of office suites makes it the ideal choice for easy, quick use. In addition, there are web versions with cloud support which equate them with more modern tools, so they are resistant to leaving their place of preference.

When spreadsheets "fall short" (large volume of data, advanced operations, or aesthetic design options, etc.) more advanced tools and platforms appear. In general, in this second group tools associated with business intelligence are included, designed to be the "professional" version of the spreadsheet.

These allow complete data management not as individual files but as information systems (possibly) with some complexity, and the advanced exploitation of data at analytical level, which ends with the visualization options. SAP, SAS, Microsoft, IBM, Pentaho, Tableau Qlik among others, are present in thousands of organizations that make extensive use of them.

Another option, with a large number of enthusiastic followers, is the use of programming languages focused on data representation. R, Weka, Matlab, Mathematica or Maple among others, represent the "technical" version of spreadsheets and BI tools, in the sense that they **require some essential technical knowledge**, while the latter are accessible to users without special knowledge. By contrast, the possibilities for creating visualizations are virtually endless.

Thus far the review of the "traditional" tools, because with the improved web technology in recent years, **data visualization is moving to the web**. Indeed the three groups above have web presence, something that today is a prerequisite. However there is a truly differentiated group consisting of native web applications. These tools understand data visualization as a natural part of a web page (either as an embedded content or directly as an actual page), adapt to the current needs of dynamic and responsive content, and also simplify the process of creating display both at a purely data management level and as the aesthetic or visual part. An example in this line is the [Tabulae platform](#).

Finally, and highlighting from the point of view of open data management, there are several web portals that have been created to cover the need for online data management. Socrata or CKAN are representatives of a less company focused tool than BI tools, and understand data management natively on the web especially data from public sources, and manipulation and later viewing thereof.

3.2. Examples of tools

Some of these tools are closer to the realm of analysis and pure data visualization (on the web). Others, though created with the initial purpose of supporting the publication of data, then add to their functionality capabilities of graphical representation of data for their consumption. What are the main tools?

3.2.1. Tableau Software²⁴

Tableau Software is a company entirely dedicated to the data analysis and visualization. Through its product Tableau Desktop and Tableau Server it offers users a powerful tool for data analysis and enrichment. In addition, its powerful graphics engine can generate stunning visualizations on



²⁴ <http://www.tableau.com/es-es>

large volumes of data. Tableau Software is widely used in the communication sector (New York Times, O'Reilly Media), pharmaceuticals (Bayer CropScience) and education (mainly American universities).

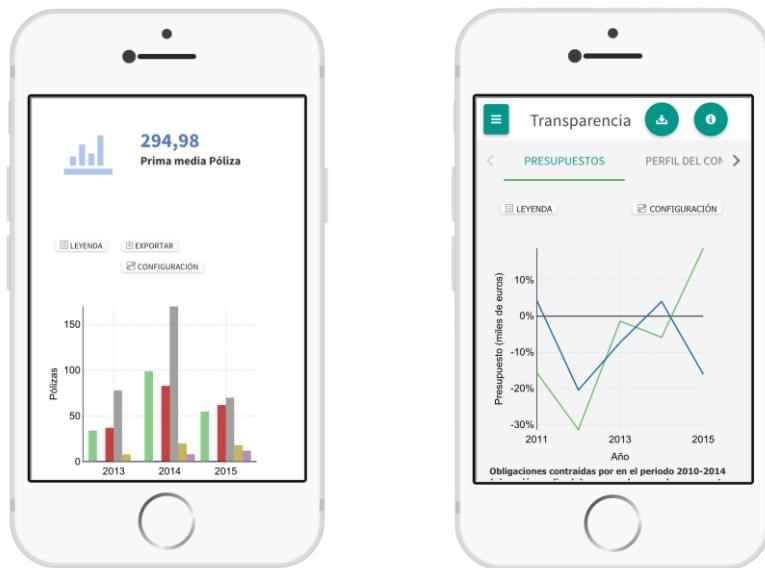
3.2.2. Tabulae²⁵

Tabulae is a web platform based on W3C standards, for data exploitation and visual analysis. It is a flexible tool that stands out for its ease of use and ability to fully customize the presentation of information through dashboards and interactive reports.

Tabulae provides dynamic mechanisms of exploitation, enriching the users' experience and their ability to interpret the data. Without programming knowledge, you can transform any data set into an interactive web application accessible from any device.



²⁵ <http://tabulaeapp.com/>



3.2.3. QlikView y QlikSense²⁶

QlikTech offers two main products QlikView and QlikSense. Its flagship product, QlikView is a business intelligence software that stands out for its ease of use and visuality allowing business users to make decisions based on data. QlikView enables the consolidation of data from multiple sources in a single application, which facilitates the exploration of associations between data. On the other hand QlikSense is a lighter and simpler version more oriented to visualization than analytics. Both solutions allow the construction of dashboards and reports easily. The website has a varied catalog of demos, which allows us to quickly discover functionality and visual appearance

²⁶ <http://wwwqlik.com/>

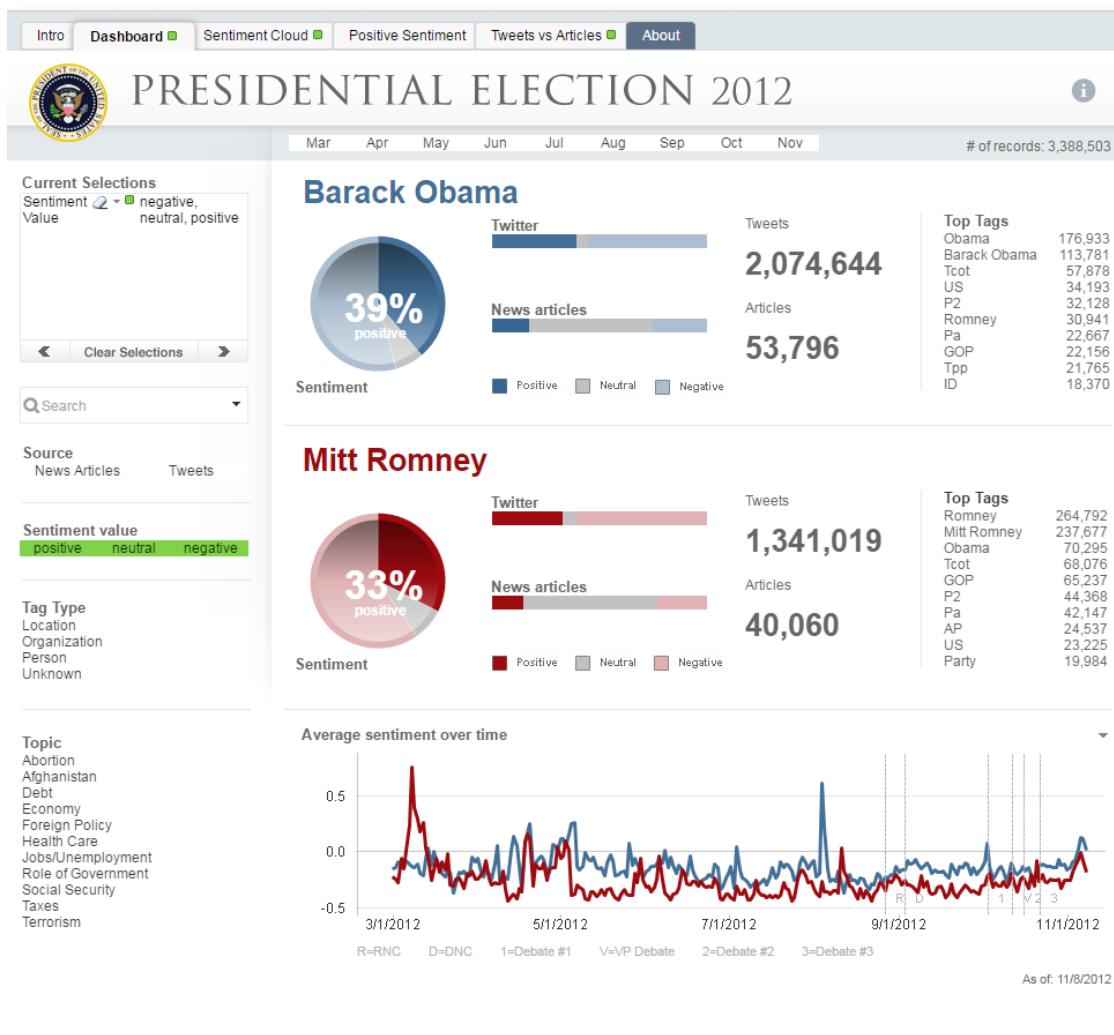
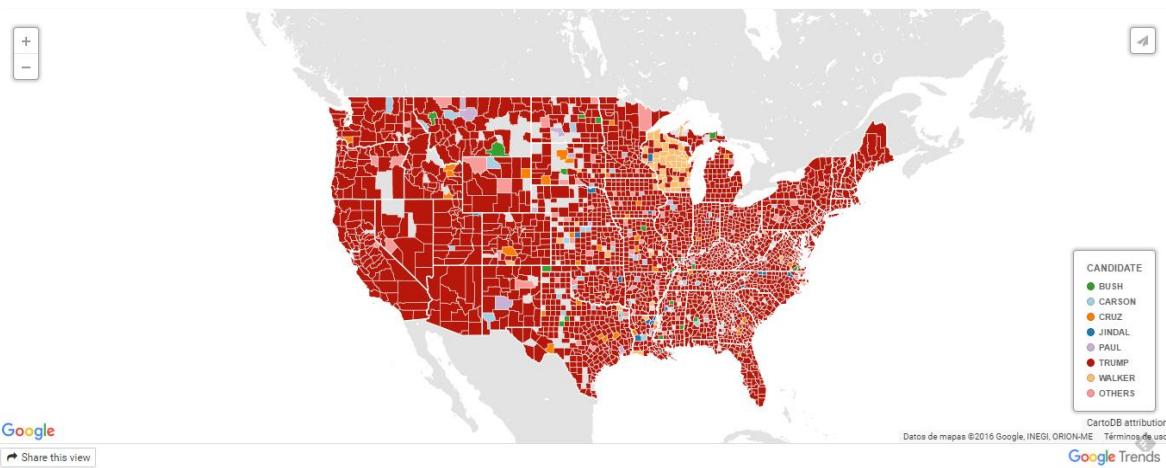


Fig. 17 – Example of dashboard on presidential election 2012 in USA with QlikView

3.2.4. Carto²⁷

Carto is a cloud service that provides GIS capabilities and tools to build maps accessible through a web browser. Carto is a service widely used for creating **thematic maps** because of its ease of use and visually attractive results, so we can find maps created with this tool in many news digital newspapers, blogs, etc.

Fig. 18 – Example of a thematic map developed with Carto of google searches on republican



candidates to USA Presidency.

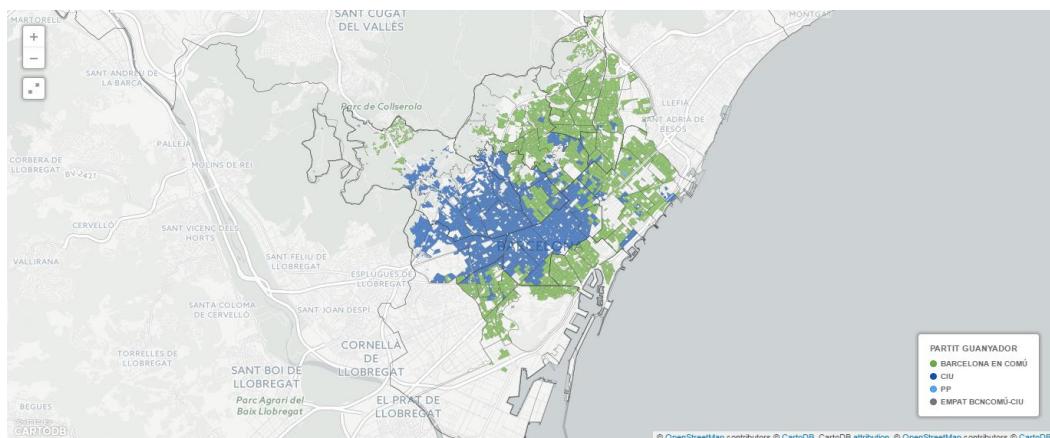


Fig. 19 – Example of visualization of the results of local elections in Barcelona, in May 2015

²⁷ <https://cartodb.com/>

3.2.5. Socrata

Socrata is a company that offers to non-technical users who want to share information the ability to easily create visualizations of their data. Founded in 2007 it is specialized in supporting the public sector offering a comprehensive service ranging from support for **data collection** to its **conversion into applications**. It is currently present in a number of open data portals such as the data transparency portal of Gijón (Fig. 5).

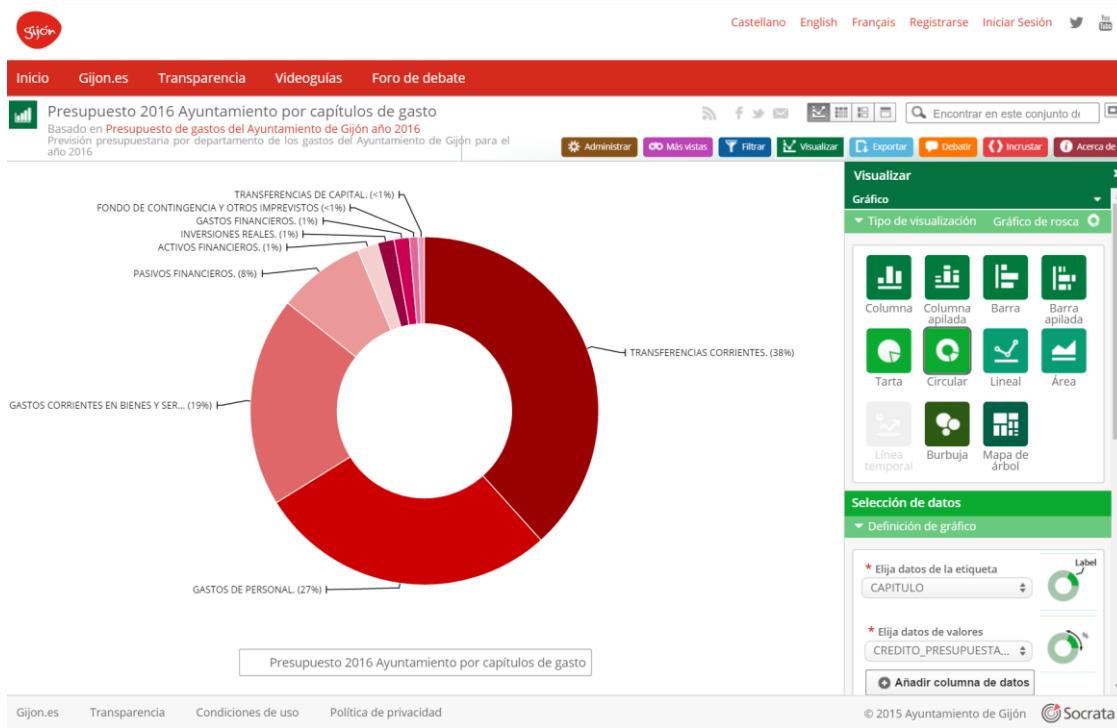


Fig. 20 - Visualization of budget, Gijón Town Hall: <https://transparencia.gijon.es/>).

3.2.6. CKAN (Visualization Modules)

The widely known platform for publishing data catalogs, CKAN, offers a set of visualization modules that can generate different graphical representations.

The **Table view** allows us to have a tabular version, in table format, of the data of a particular dataset, offering **listing tools and faceted view**. It also has a module for data visualization in graph form, enabling the comparison of variables through a series within the same graph.

Similarly, if the data contains geo-referenced data the module maps can be used to create interactive visualizations.

These visualization modules within the platform are CKAN Data Explorer, DataStore Grid, DataStore Graph, DataStore Map, Text view, Image view and Web page view. They are installed together with the platform. Moreover, there are also numerous extensions to complete this basic functionality, such as *ckanext-basiccharts*, *ckanext-dashboard*, *ckanext-map*, *ckanext-MapSearch*, *ckanext-mapviews*, etc.

4. EXAMPLES OF DATA VISUALIZATION

In this section reference examples of visualizations are described, both national and international, which are pioneers in the field of open data. The aim is therefore to illustrate through specific applications the power of graphic exploitation to provide value to the data from the point of view of the user (either citizens or employees of the administrations themselves). These visualizations have been built using libraries and tools presented in the previous sections of the document.

4.1. Datos.gob.es

The national initiative that organizes and manages the **National Catalogue of open data**, is the single point of access to data sets that the government made available for their re-use in Spain. This catalog offers a set of data produced or held by public bodies available in electronic and RF formats. The National Catalogue offers a multitude of datasets from a broad range of themes: transport, industry, health, education, finance, environment, etc. The publication of this data offers not only the possibility of citizens' access to a large set of relevant information, but also enables the creation of products and services based on such data.

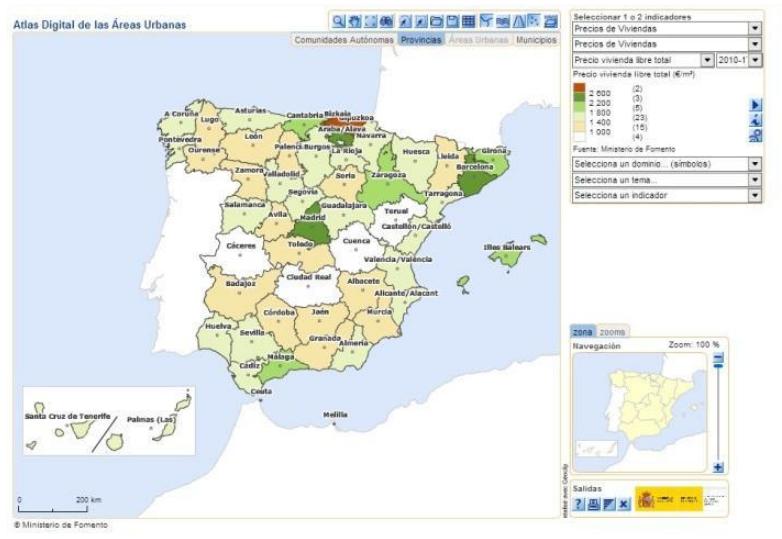


Fig. 21 - Example of visualization of urbanistic information in each Spanish region from the data of the Ministry of Public Works & Transport

4.2. World Bank²⁸

The World Bank is one of the main references regarding the publication of open data. They offer a wide range of data and the ability to view the data through graphs and maps. One of their objectives is to fight poverty in the world so they offer a great deal of data on the development of all countries in the world. Thus we have a mine of data on agriculture, economics, education, health, etc.

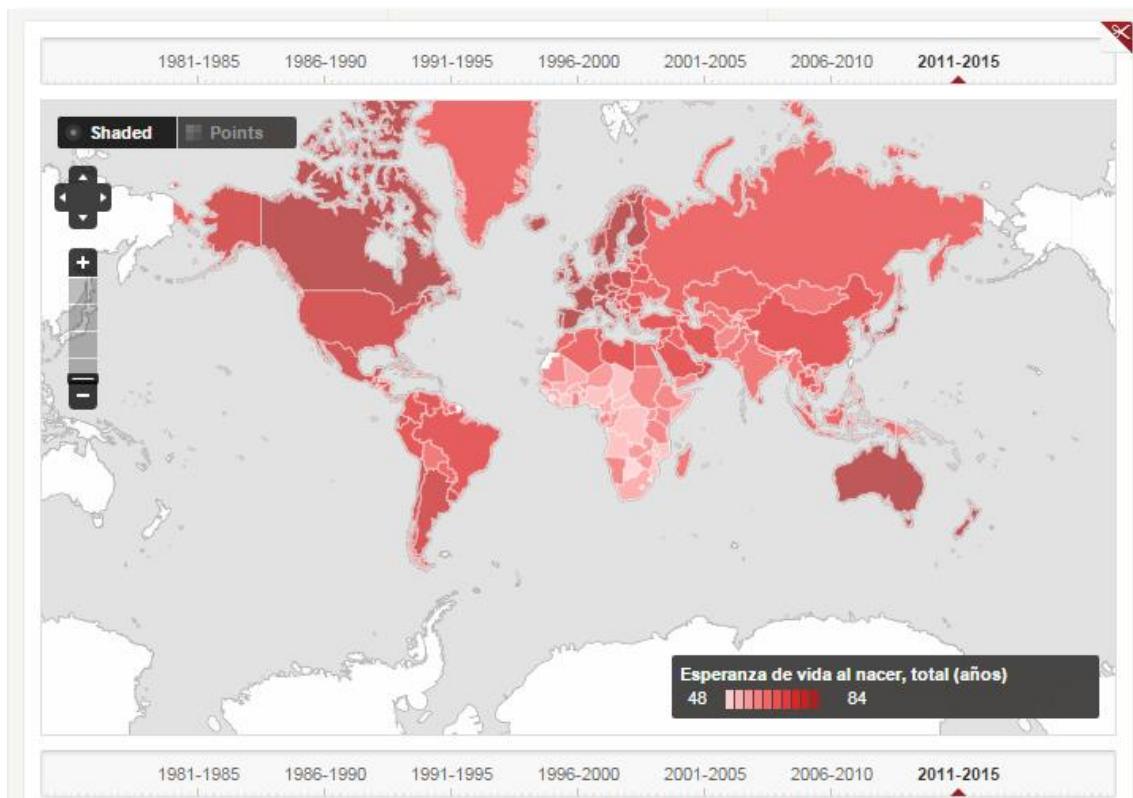


Fig. 22 – Example of visualization by World Bank. Life expectancy.

4.3. Transparency portal of the Principality of Asturias²⁹

The Government of the Principality of Asturias offers its citizens a transparency portal with a variety of data and dynamic visualizations. These visualizations, developed with the Tabulae tool

²⁸ <http://www.bancomundial.org/>

²⁹ <http://asturias.transparenciaendatos.es/>

allow the user to explore in a simple way information on social affairs, economy and finance, employment, environment and health.

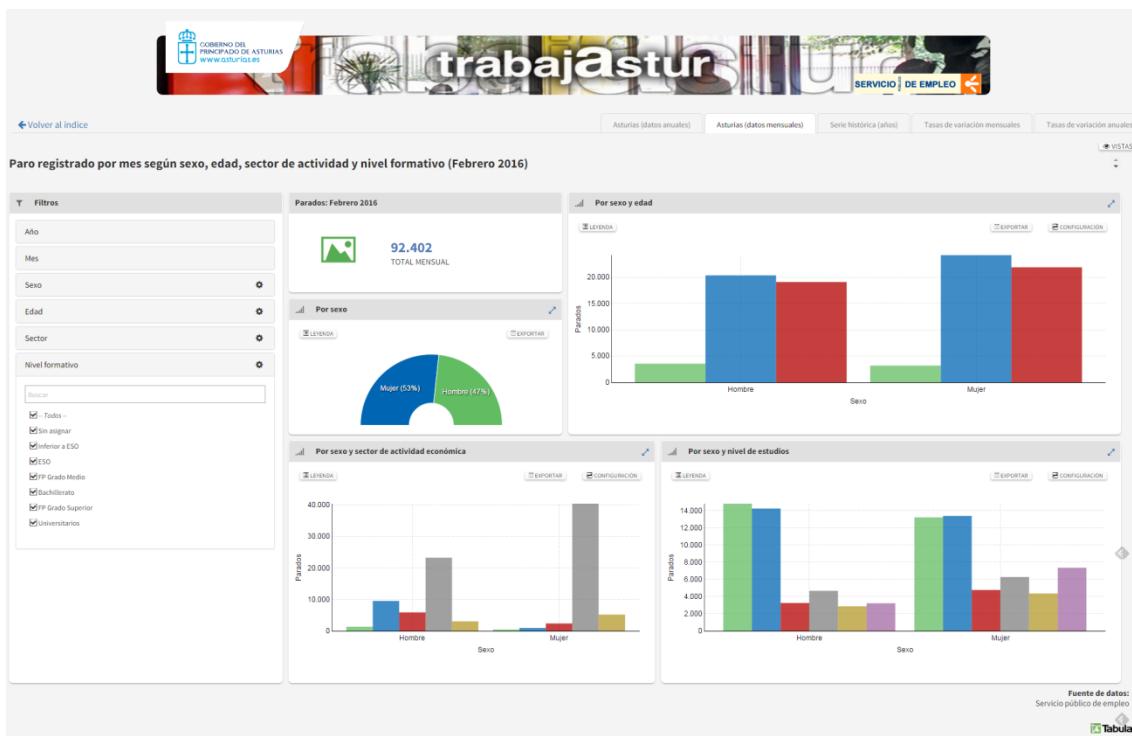


Fig. 23 – Example of interactive dashboard on unemployed population in the Principality of Asturias (Tabulae)

4.4. openREDBAG³⁰ Platform

openREDBAG is a project of the Ibero-Macaronesian Association of Botanic Gardens (AIMJB), MAGRAMA and the Biodiversity Foundation to build a **platform for access to data on wild plant genetic resources** held in genebanks. This platform offers users an application which through data from AIMJB and GBIF (Global Biodiversity Information Facility, the open database of global biodiversity), enriches them and allows them to visualize through maps and tables providing information on the species conservation, catalog protection, endangered species, etc. the national territory.

³⁰ <http://www.redbag.es/openredbag/index.php>

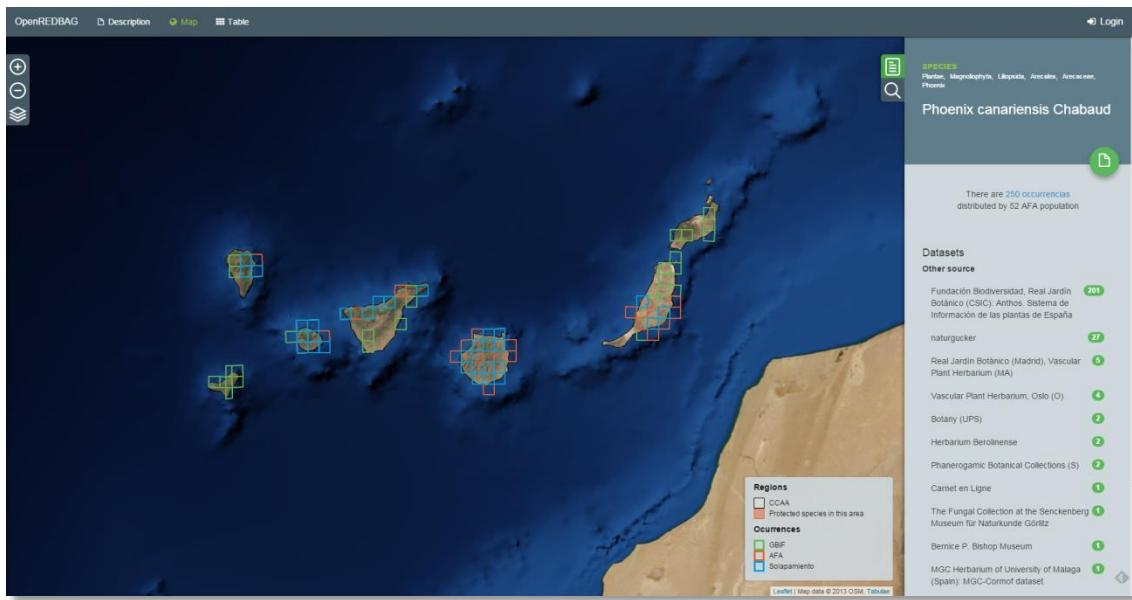


Fig. 24 – Example of visualization of species retained in germoplasm banks of REDBAG

4.5. Google Public Data³¹

Google Public Data is a Google service that allows the visual exploration of a variety of data sets from multiple sources. It offers the ability to view both graphs and maps and even to animate the visualization from the time dimension.

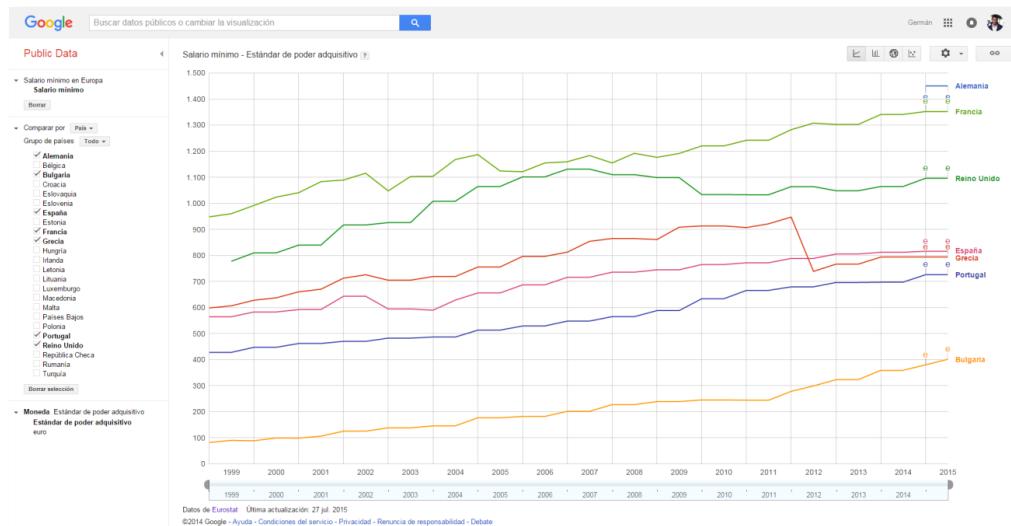


Fig. 25 – Example of visualization of minimum salary by country based on Eurostat data.

³¹ <https://www.google.com/publicdata/directory>

4.6. Eurostat

Eurostat (European Statistical Office) is the statistical office of the European Commission. Its main function is to provide statistical information on the institutions of the European Union and the harmonization of statistical methods of member states. On the Eurostat web we can find many datasets and visualizations on economic, social data, industry, etc. both globally and regionally.

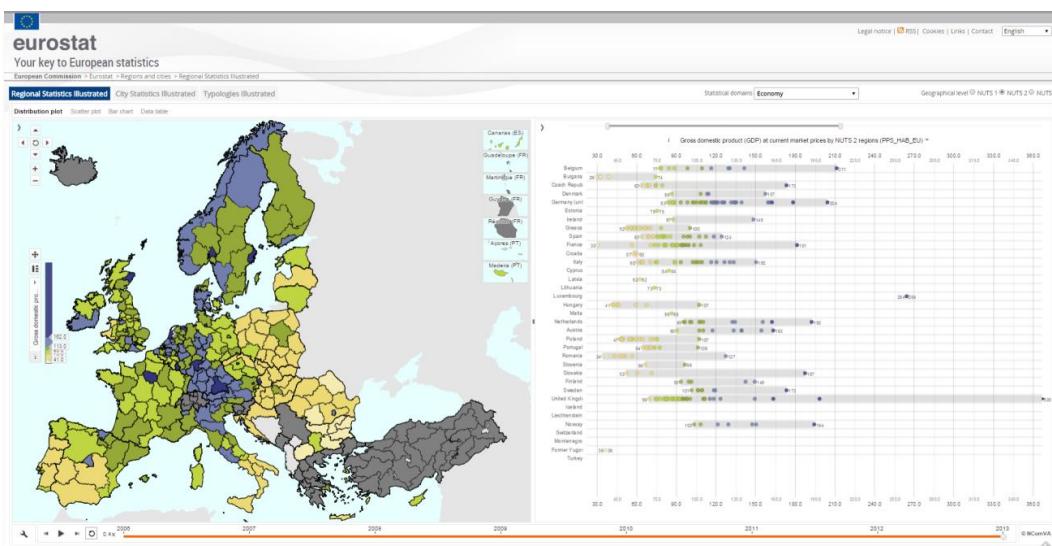


Fig. 26 – Example of visualization from Eurostat of gross domestic product.

4.7. Civio³² Foundation

The CIVIO foundation is a nonprofit organization that aims to achieve real transparency and free access to public data for both citizens and organizations. To this end, it develops tools and generates reports that help extract value from public data and promote transparency. CIVIO uses **available public data sources** such as INE or the BOE and researches to generate relevant information on public management. The ultimate goal is accountability on the part of institutions in order to improve democracy. Some of the most prominent CIVIO Foundation projects are:

- **¿Dónde van mis impuestos?** (Where do my taxes go?) It allows citizens to know the distribution of regional budgets for the years 2006 to 2015. Citizens can see the distribution of budgets by regional governments compared with the total budget or its

³² <http://www.civio.es/>

distribution per capita, and visualize that distribution for each of the functional areas (health, education, culture, etc.)

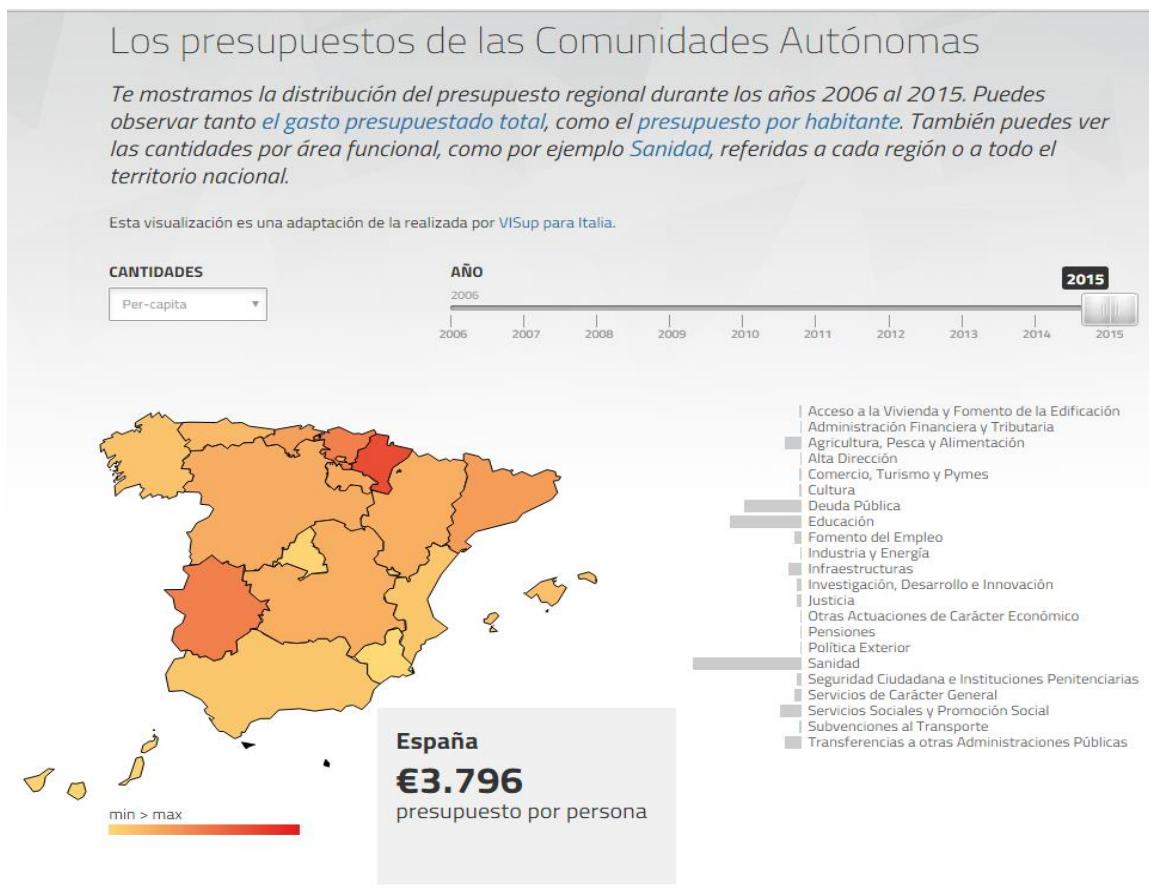


Fig. 27 – Where do my taxes go? Example of visualization of budget distribution based on data from Autonomous Region Budgets of the Ministry of Public Administration

España en llamas (Spain in Flames): This tool displays the data of all forest fires in Spain between 2001 and 2013. It offers many fire-related data, such as the cause, the number of hectares burned, whether they are intentional or not, whether they have caused deaths, etc.

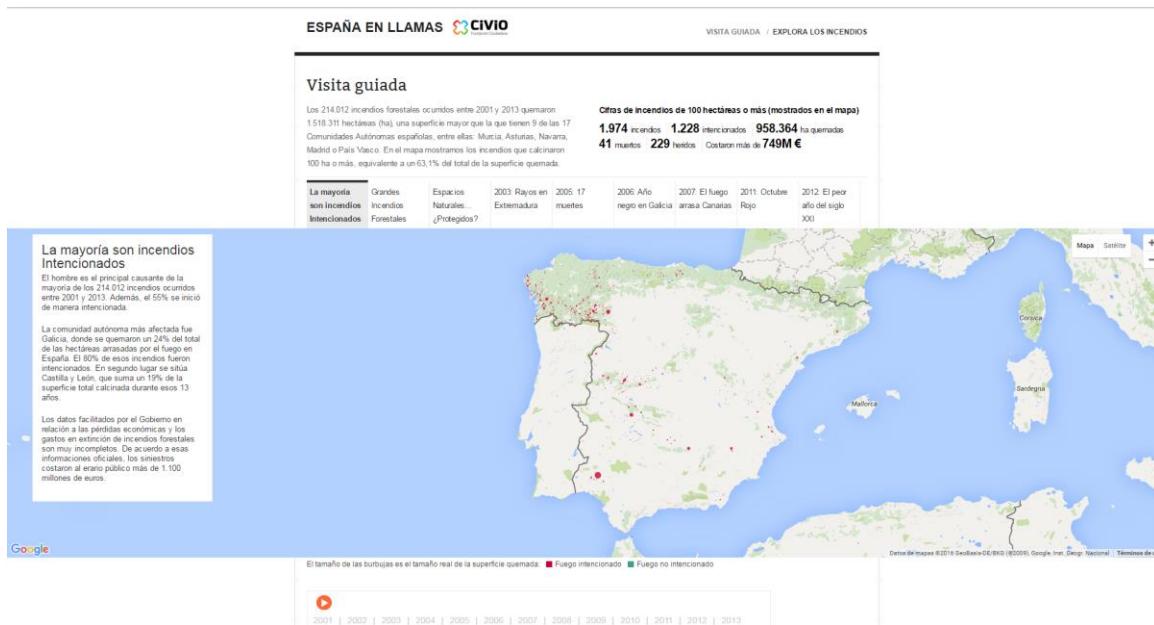


Fig. 28 – Spain in flames. Example of visualization of forest fires in Spain from 2001 to 2013.

- **El indultómetro (the pardon meter):** Collects and classifies all the information published in the BOE on the pardons granted in Spain since 1996.



Fig. 29 – Example of visualization on a timeline of main pardons.

4.8 Other examples of visualization

In recent years interactive visualizations have revolutionized the way we communicate on the Internet. In addition to institutions and initiatives there are many media, businesses and even people who use visualizations, infographics or dashboards to communicate. In many cases, these visualizations are not developed with libraries of conventional graphic components but they are

implemented on a more ad-hoc basis and with particular emphasis on **finding visual impact**. Here are some examples:

4.8.1. Evolution of the Web³³

Some members of the development team of Google Chrome browser developed an attractive visualization of the evolution of the Web.

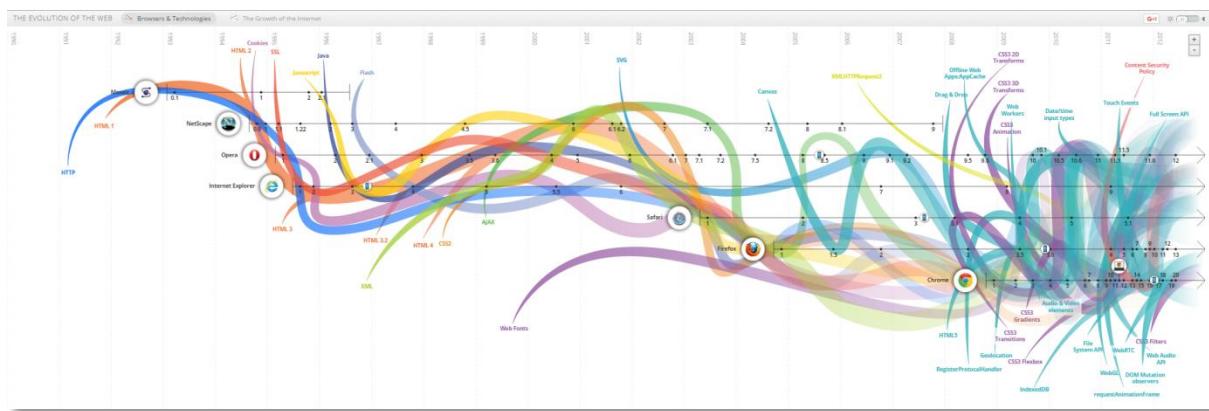


Fig. 30 - Visualization of a timeline of the evolution of the Web

This visualization shows the evolution of the web according to the different standards, browsers and technologies that have emerged over the years. The user can interact with the visualization to highlight each of the visualized items and access images and screenshots of the different versions of each.

4.8.2. Global warming³⁴

Ed Hawkins, a scientist and professor of meteorology at the University of Reading, recently published a shocking visualization of global temperature increase from 1850 to the present.

³³ <http://evolutionofweb.appspot.com/>

³⁴ http://www.climate-lab-book.ac.uk/files/2016/05/spiral_optimized.gif

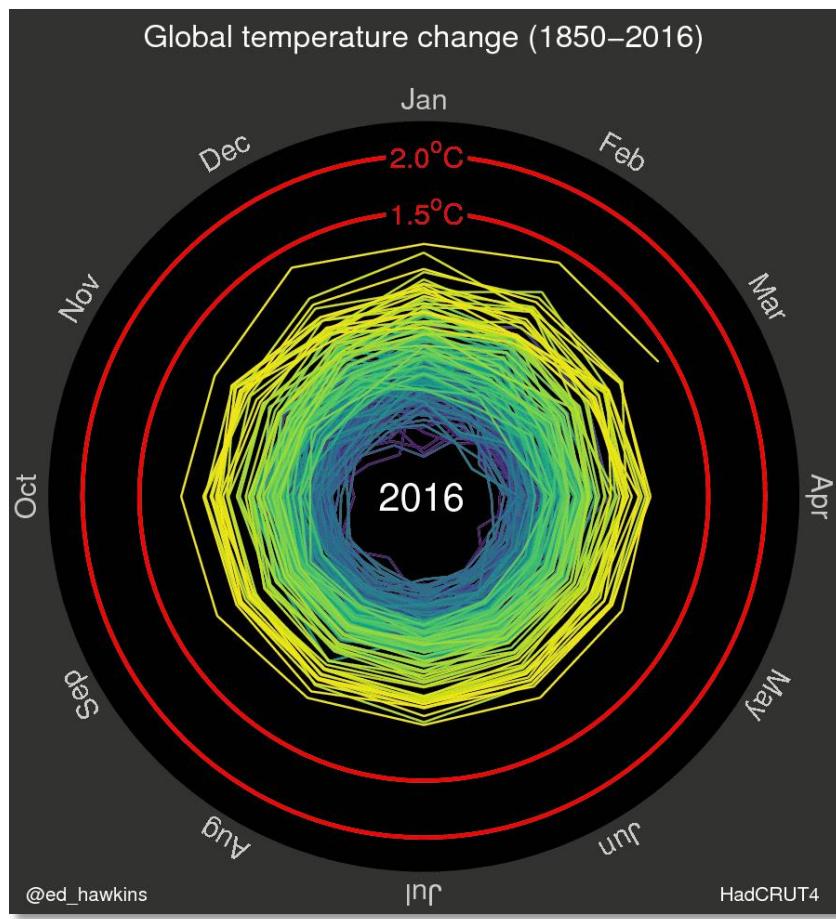


Fig. 31 - Visualizaton of global temperature increase since 1850 to date.

This visualization shows by means of a spiral the variation in the global temperature of the planet. This is an animated visualization in which the passage of years is simulated and which shows how in some periods, especially in the early years, the spiral shows closely spaced lines and indicating small variations and even contracts to indicate cooling. However with the passage of time and especially in recent decades a greater separation between the lines is shown and how they are moving further away from the center, clearly showing that global warming is accelerating.

4.8.3. Retirement of Kobe Bryant³⁵

On the occasion of the retirement of Kobe Bryant, one of the best players in NBA history, the LA Times created a fantastic interactive display that shows over 30,000 shots that he made during his sports career.

³⁵ <http://graphics.latimes.com/kobe-every-shot-ever/>

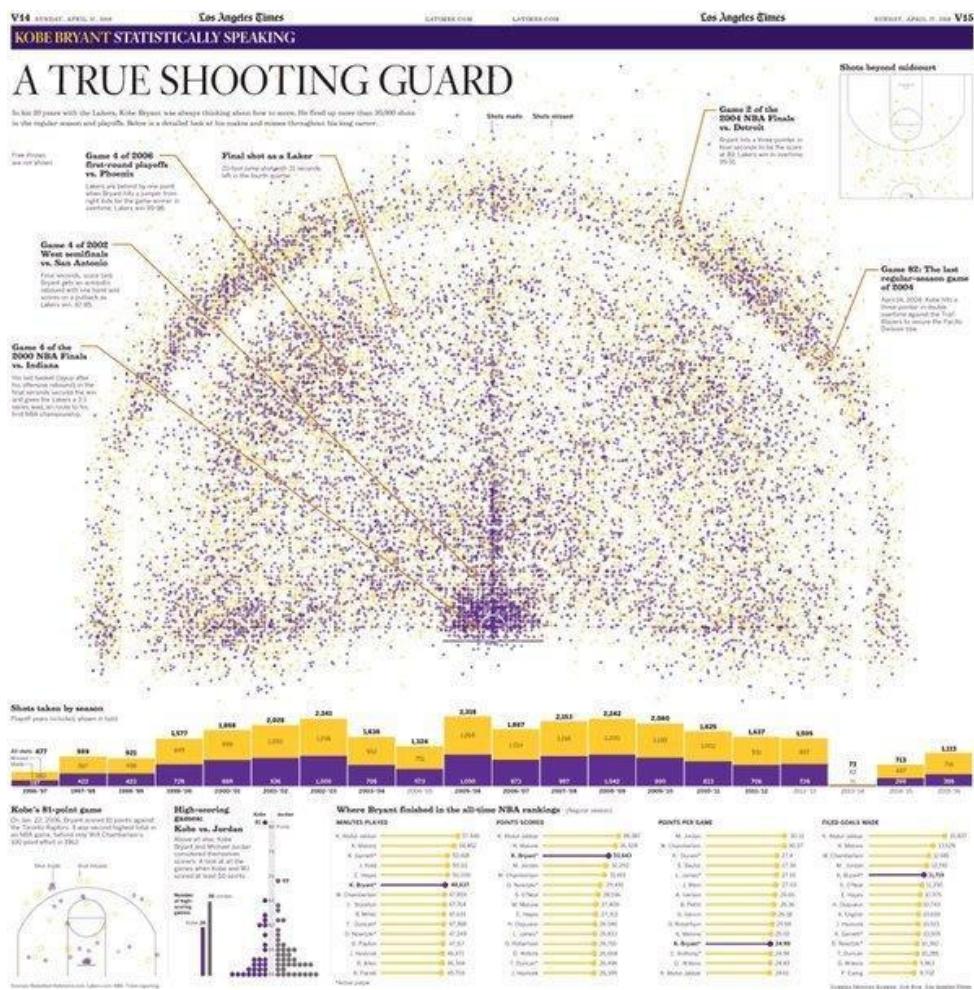


Fig. 32 - Visualization of Kobe Bryant's shots over his sporting career

This visualization shows on a basketball court the position of each shot by Kobe Bryant and allows readers to place their cursor on each of these points to see the details (type of shot, distance to the hoop and rival). The color of the dot indicates whether the shot was successful or unsuccessful.

4.8.4. History of music³⁶

Google has developed an impressive visualization showing the popularity of each genre through each decade since 1950.

³⁶ <http://research.google.com/bigpicture/music/>

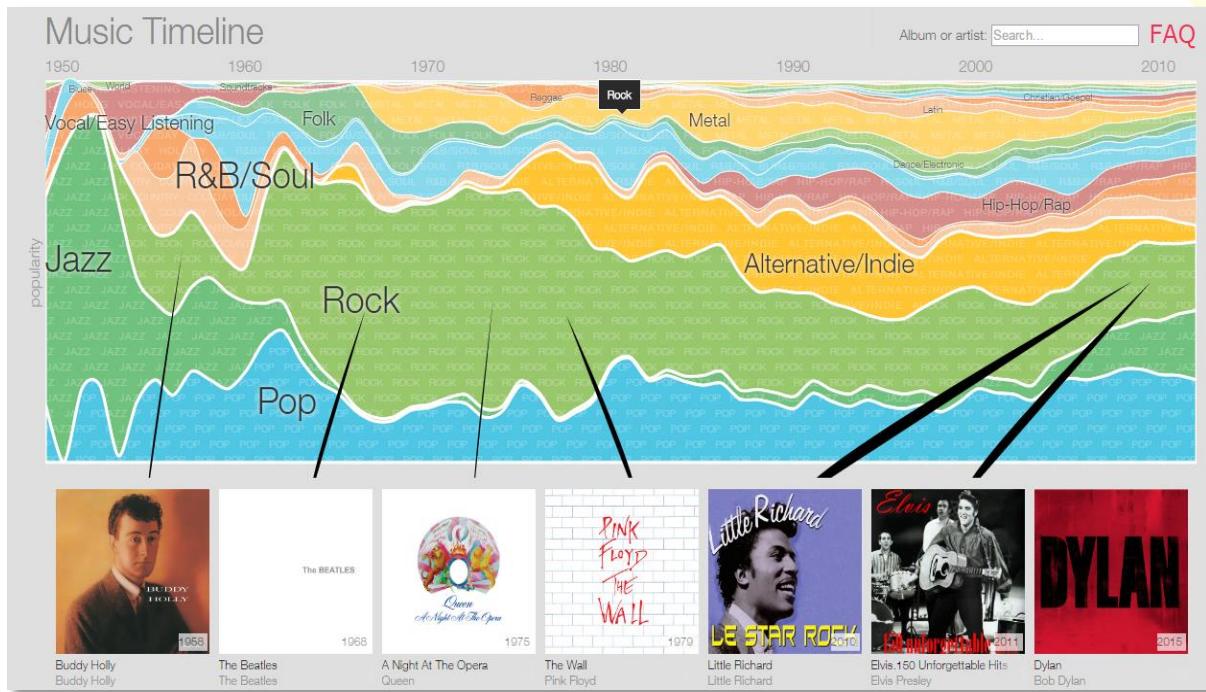


Fig. 33 - Visualization of music genders and their popularity over time

This example also allows users to interact to break down every musical genre in subgenres. It also shows artists and albums for each of the categories and lets them be played by the Google music service, acting as an original and attractive search engine.

5. CONCLUSIONS

Once you have advanced significantly in the publication of open data - legislation, formats, best practices, defining standard vocabularies, etc. - the next step is to facilitate usability and data access by any user. **Data visualization is one of the most powerful mechanisms to exploit and analyze autonomously the implied meaning in the data**, regardless of the degree of technological knowledge of the user. Visualization allows us to construct meaning from the data and create narratives based on the graphical representation.

Data visualization, from a formal and scientific point of view, is a mature discipline. In addition, the degree of implementation in industry is very wide, from the industrial and processing field to aspects more closely linked to financial reporting and business management. Its application is transversal to any process in which there are data, as in the case of the open data movement.

This is a great advantage from the point of view of the introduction of data mining tools based on visualization. The methodological and usability principles are well known. In addition there is now a wide variety of software available for the development of visualization projects within the open data portals.

In this report we have focused our analysis on two distinct blocks:

1. **Visualization technology.** Frameworks and programming libraries that allow the construction of applications and data based services. The report focuses exclusively on cutting-edge web technology and especially that which is built on standards, such as HTML5 (Canvas), SVG and WebGL.
2. **Visualization platforms.** Applications that allow the construction of dashboards and comprehensive interactive visualizations (such as the case of Carto maps). Those platforms that are more web-oriented, which can be divided into two main sections, are presented in detail. On the one hand, those coming from the world of **BI and data analysis**, and that can be applied to the exploitation of open data: Tableau Soft, Qlik and Tabulae. On the other hand, open data publishing tools that incorporate certain features of visualization: CKAN and Socrata.

5.1. Future trends

Related technologies and data visualization techniques continue to evolve significantly. This report is a state of the art that reflects what most relevant alternatives are on the market for work in this discipline. From the point of view of innovation and future lines, there are several paths being followed both by the software industry and the academic sector, including:

1. Visualization and large volumes of data

One problem related to visualization is working with large volumes of data. This problem does not exist only at the level of back-end, which is more linked to the analysis and processing of large volumes of structured and unstructured information (ie., Big Data), but also with the browser's ability to handle large data sets (eg, hundreds of thousands of records). Computing restrictions of clients in which visualization occurs can be a barrier to the construction of certain visualizations. This is a research field which is currently being worked on.

2. Visualizations that can be built by the expert data user, and even by the end user (in the case of open data, by the citizens themselves).

Many tools are only accessible for technical profiles, whether at computer or data processing and analysis level. This technology barrier is one of the current major challenges: allowing the user that understands the information to have enough autonomy to build through intuitive wizards (WYIWYG) their own visualizations and to exploit information dynamically.

3. 3D visualizations

In recent years, as identified in the report, a technology of 3D graphics acceleration for the web environment has appeared. The application of this technology in data visualization is still at a very early stage. Most of the tools are still working with 2D perspective. It is expected that once the technology reaches a more solid state of maturity, new visualization techniques integrated with the user tools will appear.

4. Data visualization and augmented reality

One field which has still been little explored until today, but which surely would be a breakthrough in the way data are consumed, is augmented reality technology, which uses devices to combine real and physical information in an interactive environment for the user. Currently the state of the art of visualization is concentrated in a classic and multi-device consumption of data: whether

web applications or mobile applications for a specific purpose. Augmented reality technology with its possibilities of 3D rendering and in real-time, is an uncharted territory at this time as regards the potential applications and impact on the discipline of visualization. However, a vast space opens up here for imagination and design of new solutions in an environment until recently "futuristic".