



European Public Sector Information Platform

Topic Report No. 2013 / 07

DATA PROCESSING AND VISUALISATION TOOLS

Author: datos.gob.es

Published: August 2013

Table of Contents

- Keywords: 4
- Abstract/ Executive Summary: 4
- 1 Introduction 5
- 2 Tool features 6
 - 2.1 Processing tools 6
 - A. Refinement tools..... 6
 - 2.1.1 DataWrangler 6
 - 2.1.2 Google Refine..... 7
 - B. Conversion tools..... 8
 - 2.1.3 Mr. Data Converter..... 8
 - 2.2 Statistical analysis tools 9
 - 2.2.1 The R Project for Statistical Computing 9
 - 2.3 Display services 10
 - A. Generic visualisation applications..... 10
 - 2.3.1 Google Fusion Tables 10
 - 2.3.2 Tableau Public 11
 - 2.3.3 Many Eyes 12
 - 2.3.4 CartoDB..... 14
 - 2.3.5 GeoCommons 15
 - B. Wizards, libraries, API..... 16
 - 2.3.6 Google Chart Tools..... 16
 - 2.3.7 JavaScript InfoVis Toolkit..... 17
 - 2.3.8 D3.js 19
 - 2.3.9 Protovis 20
 - 2.3.10 Recline.js 21
 - C. Geospatial visualisation tools..... 22
 - 2.3.11 OpenHeatMap 22
 - 2.3.12 OpenLayers 23
 - 2.3.13 OpenStreetMap 24
 - D. Temporal data visualisation tools..... 25
 - 2.3.14 TimeFlow..... 25
 - 2.4 Tools for network analysis..... 26

2.4.1	Gephi.....	26
2.4.2	NodeXL.....	27
3	Comparison.....	28
4	Conclusions and recommendations.....	30
	About the Author	33
	Copyright information	33

Keywords:

visualisation, charts, library, application, graphics, API, display, processing, tools, toolkit, refinement, cleansing, data

Abstract/ Executive Summary:

Raw data can be hard for the average internet user to understand, even for those with advanced technical skills. In order to make this data easily understandable and user-friendly, it must be processed and prepared. Data processing and visualisation are essential in facilitating the interpretation of data and the story behind the information.

This document contains a compilation of free of charge tools for data processing, analysis and visualisation. These tools will be assessed by category and results and conclusions will be shown at the end of this report.

1 Introduction

In recent years, there has been a huge proliferation of raw data that must be processed and prepared for the end user in a format that is easy to understand. This information is often difficult to understand; therefore several visualisation tools have been developed in order to facilitate interpretation and understanding.

However, these tools do not solve the problems generated by the low quality of the source data. This implies that there is need to work correctly with data, before it receives graphical treatment. Prior to the execution of any analysis, classic data processing states that we should pay attention to the data acquisition techniques, and carry out a study of the data obtained to ensure a correct representation of the universe of information. After assuring the quality of information, we can proceed to assess the set of data (exploratory, qualitative, etc.), and get the results and visualisation that best fits the results, and information to be transmitted.

This document contains a compilation of the best free of charge tools for data processing, analysis and visualisation, currently available on the market.

2 Tool features

In order to carry out a good data analysis and visualisations, it is essential to know and understand the tools available as well as their correct application in the corresponding fields. There are several tools to turn data into graphics, but some of them may be costly.

Below is a selection of the best free tools for data processing and display. They are grouped by target use and application.

2.1 Processing tools

The three tools shown below have been designed to assist in the debugging and the transformation of data. They are useful to clean and refine messy data, and convert it into appropriate formats. Often, large data sets represented in tabular formats contain typos, inaccuracies –e.g., dates expressed in different formats, cells with abbreviated/expanded names, encoding errors, blank cells, etc.–, whose manual correction is unfeasible.

These tools accelerate the process that enhances the quality of the information, and makes the data complete and easy to re-use.

A. Refinement tools

2.1.1 DataWrangler

TYPE. Web application

TECHNOLOGY. HTML

LICENSE. Free to use

AUTHOR. The Stanford Visualization Group (United States)

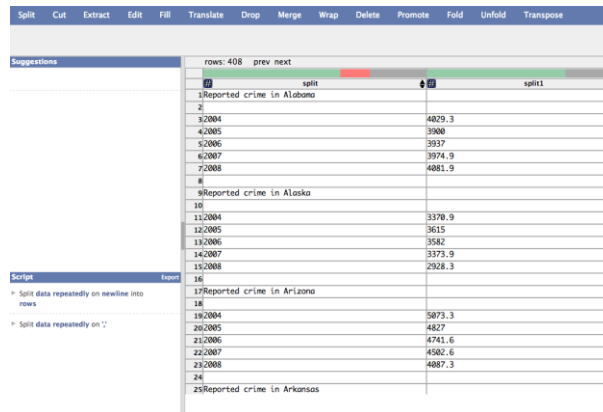
LINKS.

- Website. <http://vis.stanford.edu/wrangler/>
- Research work. <http://vis.stanford.edu/papers/wrangler>

An interactive web application for data cleaning and transformation, Wrangler combines direct manipulation of visualised data with automatic inference of relevant data transformation. It enables analysts to repeatedly scan the space of applicable operations and anticipate its effects.

It leverages semantic data types (geographical locations, dates, classification codes) to aid

validation and type conversion.



split	split1
1Reported crime in Alabama	
2	
3/2004	4829.3
4/2005	3900
5/2006	3937
6/2007	3974.9
7/2008	4881.9
8	
9Reported crime in Alaska	
10	
11/2004	3378.9
12/2005	3615
13/2006	3582
14/2007	3373.9
15/2008	2328.3
16	
17Reported crime in Arizona	
18	
19/2004	5873.3
20/2005	4827
21/2006	4741.6
22/2007	4582.6
23/2008	4887.3
24	
25Reported crime in Arkansas	

2.1.2 Google Refine

TYPE. Desktop application

TECHNOLOGY. Java

LICENSE. BSD

AUTHOR. Google Inc. (United States)

LINKS.

- Website. <http://code.google.com/p/google-refine/>
- Documentation for users. <http://code.google.com/p/google-refine/wiki/DocumentationForUsers>
- Documentation for developers. <http://code.google.com/p/google-refine/wiki/DocumentationForDevelopers>

A Free tool designed with the objective to assist in understanding the structure and quality of the data, allowing the correction of certain common errors in data.

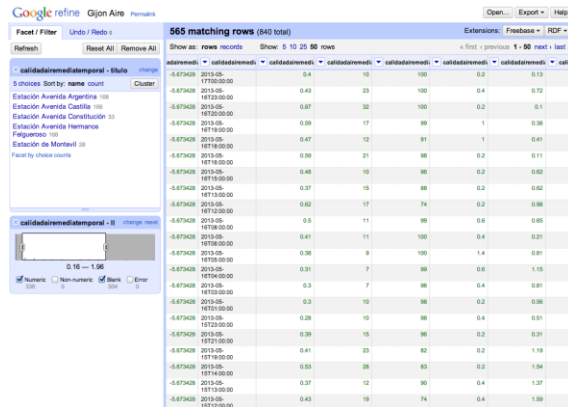
It supports a wide range of formats: TSV, CSV, *SV, Excel (. xls and xlsx), JSON, XML, RDF-XML, and Google Data documents. The data source can be provided in 4 ways: upload a local file, from a URL (importing data from tables in web pages, in XML documents, etc.), paste data from the clipboard, and link a Google Docs document. After treatment of the information, data can be exported in TSV (Tab Separated Values), CSV (comma separated values), and Excel formats, and in HTML table.

Google Refine has three key features:

- Data Cleansing. It enables changing cell content and field unification. This action may be

performed manually or assisted by the program (the system can suggest optimisations). It offers predefined operations such as collapsing consecutive whitespaces in texts, scape/un-scape HTML entities, changing letter case, converting text to dates, blanking out cells, among others.

- Data transformation. Transformations through GREL (Google Refine Expression Language) instructions. It enables the splitting of columns, creating new columns based on values of other columns and combining cells to create new columns among other features.
- Creation of new data fields. New data fields may be created by external services to obtain new data from existing data, or using Freebase (free collaborative database) to complement the data.



calidadatimediatemporal - titulo	calidadatimediatemporal - id	calidadatimediatemporal - id	calidadatimediatemporal - id	calidadatimediatemporal - id	calidadatimediatemporal - id
Estación Avenida Argentina 100	100	0.43	23	100	0.4
Estación Avenida Castilla 100	100	0.07	32	100	0.2
Estación Avenida Constitución 33	33	0.08	17	99	1
Estación Avenida Hermandad	100	0.07	12	91	1
Estación de Montevideo 20	20	0.08	21	98	0.2
		0.48	10	98	0.2
		0.27	15	88	0.2
		0.02	17	74	0.2
		0.5	11	99	0.5
		0.41	11	100	0.4
		0.08	9	100	1.4
		0.31	7	98	0.8
		0.3	7	98	0.4
		0.3	10	98	0.2
		0.28	10	98	0.4
		0.38	15	98	0.2
		0.41	23	82	0.2
		0.53	28	83	0.2
		0.37	12	90	0.4
		0.43	19	74	0.4

B. Conversion tools

2.1.3 Mr. Data Converter

TYPE. Library

TECHNOLOGY. JavaScript

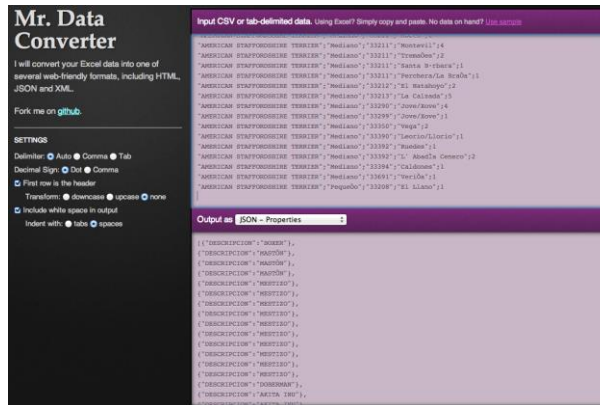
LICENSE. MIT

AUTHOR. Shan Carter (United States)

LINKS.

- Website. http://shancarter.com/data_converter/
- GitHub Repository. <https://github.com/shancarter/Mr-Data-Converter>

A Web application that can convert Microsoft Excel data into various web-friendly formats, this includes HTML, JSON and XML.



2.2 Statistical analysis tools

Tools for combining graphical representations of data along with a strong numerical analysis.

2.2.1 The R Project for Statistical Computing

KIND. Programming Language

TECHNOLOGY. R

LICENSE. GPL

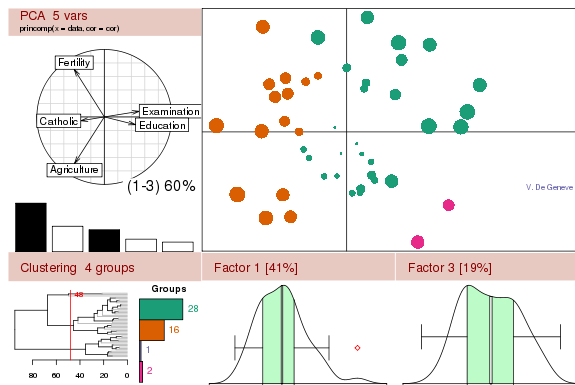
AUTHOR. R Foundation (Austria)

LINKS.

- Website. <http://www.r-project.org/>

R is a free, open source programming language and environment for statistical computing and graphics.

This is a command-based language, which allows the creation of tailored graphics. It is not based just on standardized chart types, but includes new kind of graphics for different problem addressed.



2.3 Display services

The following describes some free visualisation tools, classified according to their technological features.

A. Generic visualisation applications

There is a number of tools available that offer visualisation options. Although some of them use conventional tables and charts, many others offer new options such as tree diagrams, and word clouds.

2.3.1 Google Fusion Tables

TYPE. Web application and API

TECHNOLOGY. JavaScript, Flash

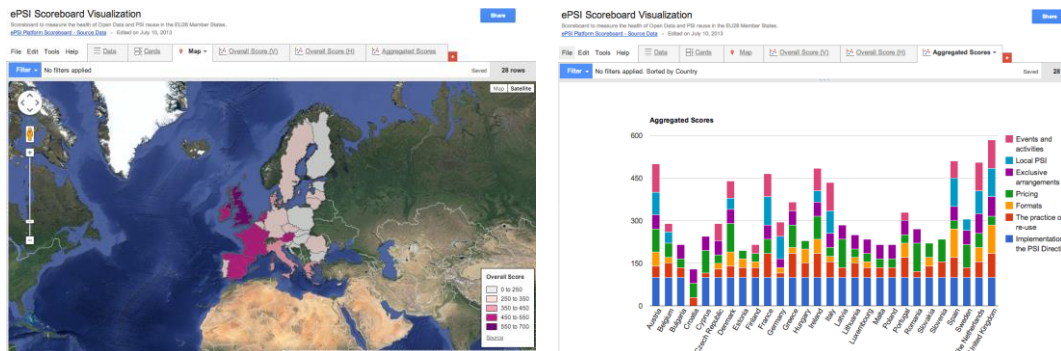
LICENSE. Free to use

AUTHOR. Google Inc. (United States)

LINKS.

- Website. <http://www.google.com/fusiontables/>
- Gallery. <https://sites.google.com/site/fusiontablestalks/stories/>
- API documentation. <https://developers.google.com/fusiontables/>

A Web application for organising, managing, visualising, curating and publishing data on the web in a simple way. It manages large collections of data to be standardised and stored in Excel, .ods, .csv or .kml files. This application displays data using pie charts, bar charts, scatter plots and timelines as well as represented on Google Maps.



2.3.2 Tableau Public

TYPE. Desktop application

TECHNOLOGY. Windows, JavaScript

LICENSE. Free to use

AUTHOR. Tableau Software (United States)

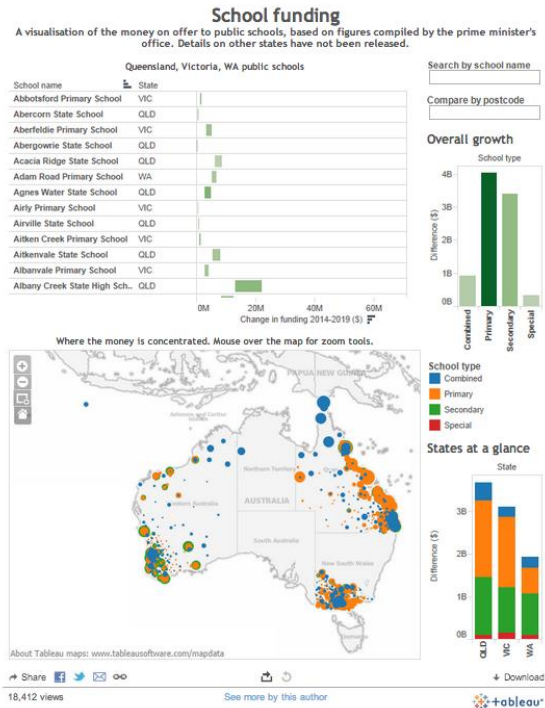
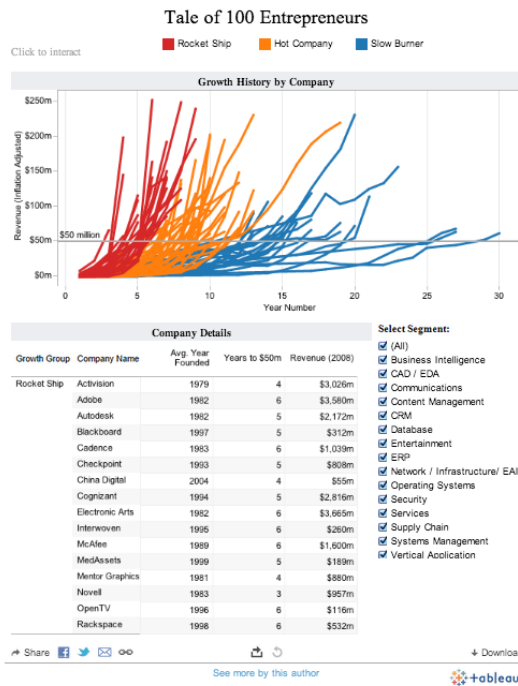
LINKS.

- Website. <http://www.tableausoftware.com/public/>
- Gallery. <http://www.tableausoftware.com/public/gallery>

A Free tool for data visualisation through graphics that combines an appealing, fast, and efficient graphical interface with traditional elements of business intelligence tools, such as the organisational model of variables using dimensions and measures, or connection with other information management systems –i.e., databases, and spreadsheets.

Some of the most relevant features of this tool are:

- Quick and easy data acquisition. It allows working with databases and spreadsheets of any size. It accepts Microsoft Excel, Access, and plain text formats.
- Work with a variety of graphics: fever, bars, stacked bars, pie, maps with polygons, lines or points, etc.
- Publication of interactive graphics.
- Combination of different data sources in a single view.
- Data are public.
- Raw data can be downloaded from the visualisation.



2.3.3 Many Eyes

TYPE. Web application

TECHNOLOGY. Java, Flash

LICENSE. Free to use

AUTHOR. IBM (United States)

LINKS.

- Website. <http://www-958.ibm.com/software/data/cognos/manyeyes>

A Web application that enables the user to create, share and discuss the graphical representation of data downloaded by users. This data visualisation tool is made by IBM.

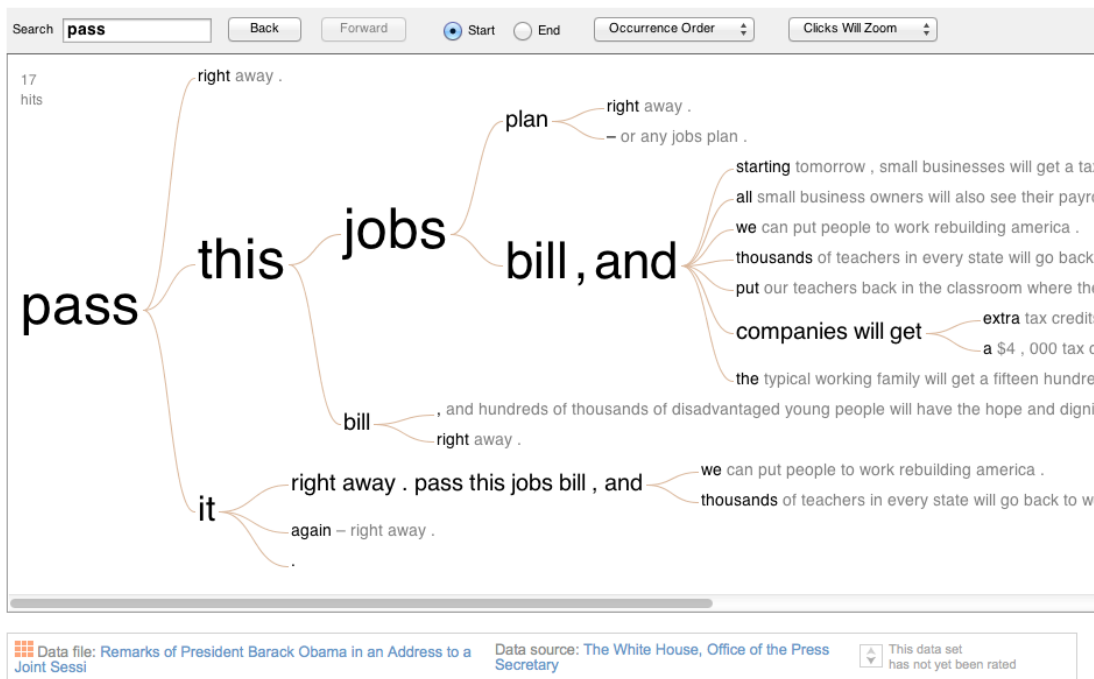
With Many Eyes users can share their visualisations, encouraging discussions through different approaches from the same data. It is a tool for public use –i.e. all data and visualisations will be made available to all users–, and it cannot be used privately.

It allows many kinds of views:

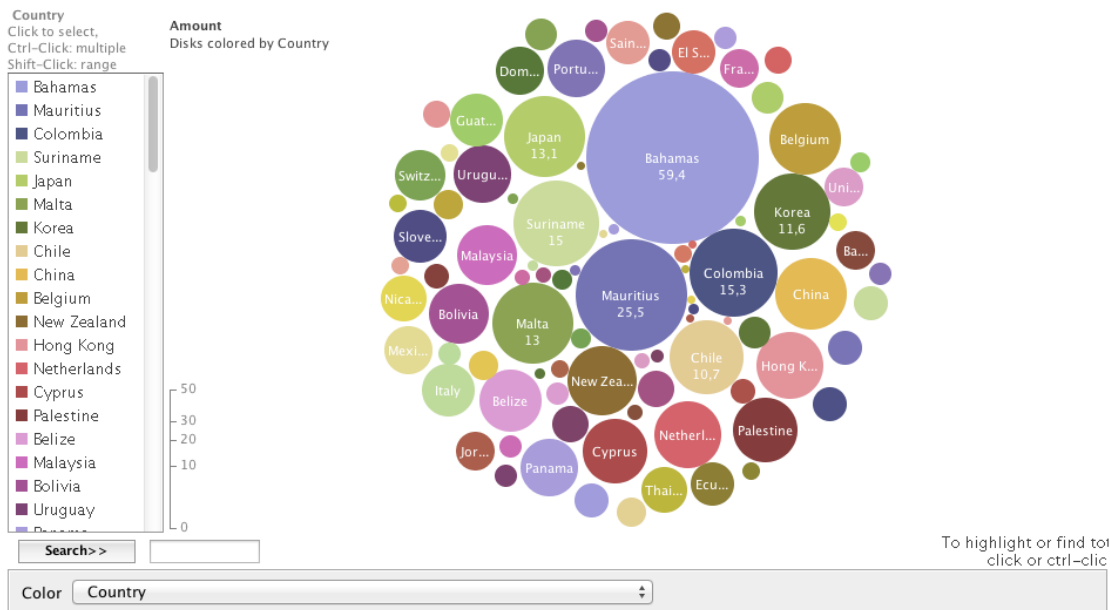
- Relations between points (scatter plot, matrix charts and network diagrams).
- Comparing values (bar, histograms and bubble charts).
- Trend changes over time (line, bar and category bar graphs).
- Parts of a whole (pie chart, treemap, and treemap for comparisons).

- Text analyser (word tree, tag cloud, phrase net, word cloud).
- Geographical graphics (charts on maps).

One of the most famous examples showing the potential of this tool is Obama's speech about work in the form of a tree and word cloud.



<http://www-958.ibm.com/software/data/cognos/manyeves/visualizations/word-tree-for-president-obamas-job>



<http://www-958.ibm.com/software/data/cognos/manyeves/visualizations/kg-bubble-chart>

2.3.4 CartoDB

TYPE. Web application

TECHNOLOGY. JavaScript, PostgreSQL and its PostGIS geospatial extension.

LICENSE. Commercial

AUTHOR. Vizzuality (United States)

LINKS.

- Website. <http://www.cartodb.com>
- Tutorials. <http://vimeo.com/channels/cartodb>
- Blog CartoDB. <http://blog.cartodb.com/>
- Blog Vizzuality. <http://blog.vizzuality.com/>
- GitHub Repository. <https://github.com/Vizzuality>

CartoDB is a geospatial database in the cloud, running on Amazon Web Services, allowing scalability and flexibility of their services. It is an open source project, also offered as a service on demand.

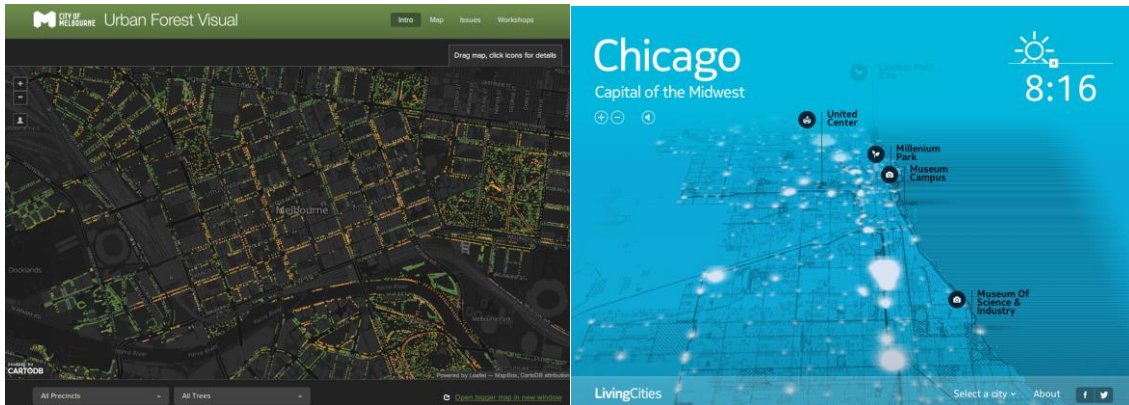
CartoDB aims to facilitate the development of geolocated applications and maps. It allows the design and development of real-time maps that work on universal web and mobile platforms.

Among its features, we can highlight:

- Map design for data layers. We can use CartoCSS in order to easily edit formats and the look and feel of the maps.
- Integration with other mapping services. CartoDB produces the data layers that are represented on Google Maps and MapBox (since version 2.0) mapping layers. These maps include the basic functions –zoom, scroll, etc.
- Integration with other libraries. CartoDB has several libraries that can extend its use or integrate with other services.
- Geocoding. Geographical information can be obtained from elements different from coordinates.
- Easy data import. CartoDB enables direct input of data into tables from the dashboard, adding data via SQL or reading from URLs. Other data collection may be imported from various formats.
- SQL queries based on spatial components. By using PostGIS, CartoDB can query and combine data sets using geospatial data.

- Public and private tables. CartoDB allows users to define the privacy of the tables, selecting between public and private use.

It is aimed at developers without experience in geospatial information systems, with a friendly interface. Various prestigious institutions such as UN, Google, NASA, the Oxford University, Yale University, among others use CartoDB.



2.3.5 GeoCommons

TYPE. Web application and API

TECHNOLOGY. JavaScript, Ruby

LICENSE. Various (http://geocommons.com/help/Open_Source)

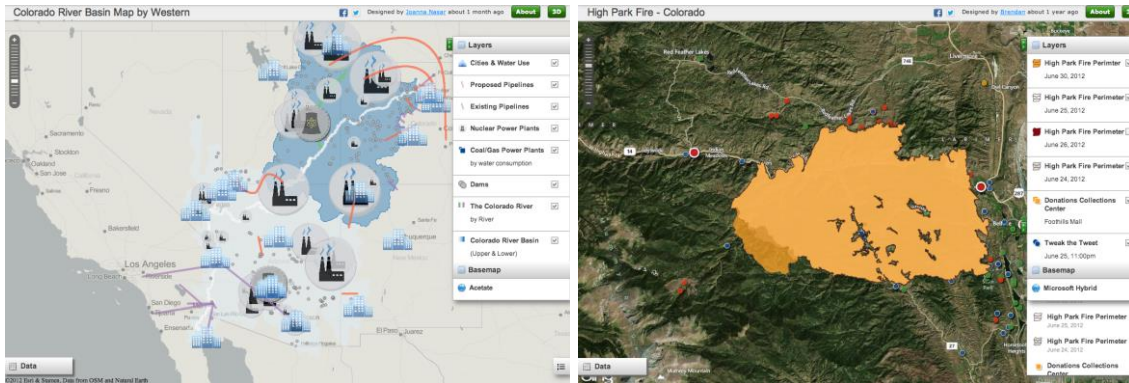
AUTHOR. Esri (United States)

LINKS.

- Website. <http://geocommons.com/>
- API documentation. <http://geocommons.com/api/>

A platform for geospatial data management, visualisation, mapping and spatial analysis. It supports the loading data from different types of data sources: spreadsheets, KML files, shapefiles, database servers with spatial support, OGC services such as WMS and TMS, and from its own public repository.

This tool supports techniques for the cartographic representation of choropleth maps, enabling customisation of the symbols –including the size, colour and transparency, the shape and style of the icons and lines, and the colour sequences– on the maps. GeoCommons also includes timing animation capabilities. Maps can be exported to KML format, its data to KML, spreadsheet or shape files, among others. Maps can be also embedded in a web page.



B. Wizards, libraries, API

A wide range of libraries and APIs available to help developers to create their own visualisations.

2.3.6 Google Chart Tools

TYPE. Library

TECHNOLOGY. JavaScript

LICENSE. Free to use

AUTHOR. Google Inc. (United States)

LINKS.

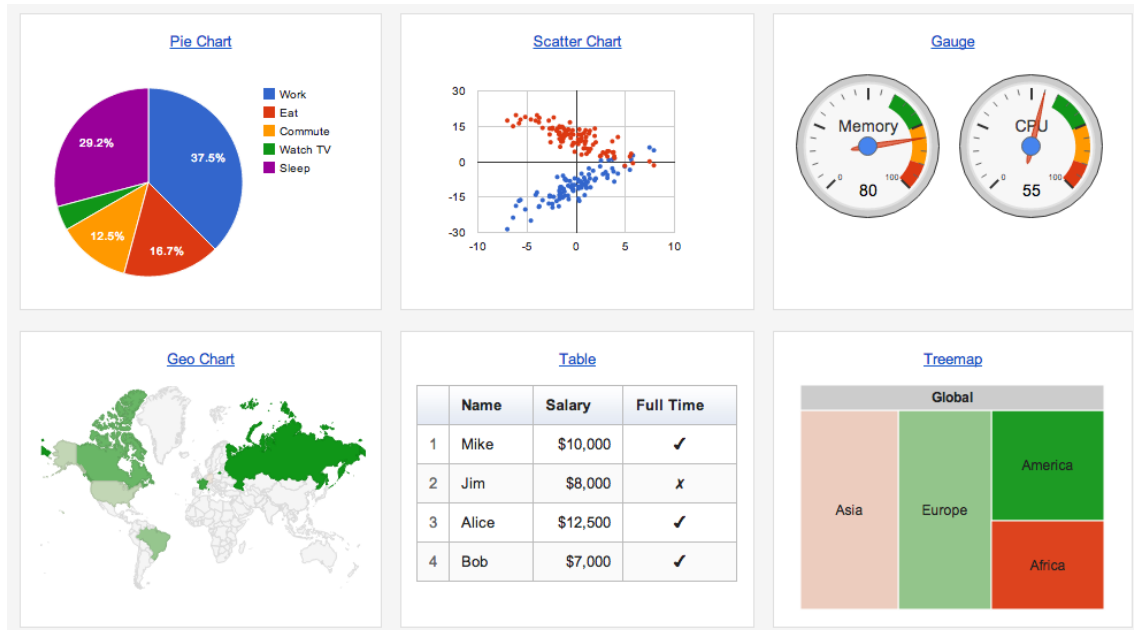
- Website. <https://developers.google.com/chart/>
- Screenshot gallery. <https://google-developers.appspot.com/chart/interactive/docs/gallery/>
- Code. <https://code.google.com/apis/ajax/playground/?type=visualisation/>
- API documentation. <https://google-developers.appspot.com/chart/interactive/docs/reference>

This Google Developers tool enables the creation of graphic images as PNG. Its operation is based on HTTP requests to a specific URL (<http://chart.apis.google.com>).

It is free to use but with some limitations. Initially, its use was limited to 50,000 requests per URL and day, but now this limit stands at 250,000. In order to avoid this limitation, generated images may be stored on an external server running as a cache of images.

There is a variety of graph types, offered as JavaScript classes. One advantage with this graphic generation system is that users do not need to install any component in environment or server,

so that each plot can be generated on the fly.



2.3.7 JavaScript InfoVis Toolkit

TYPE. Toolkit

TECHNOLOGY. JavaScript, Python

LICENSE. MIT

AUTHOR. Nicolas Garcia Belmonte (United States)

LINKS.

- Website. <http://thejit.org/>
- GitHub Repository. <https://github.com/philogb/jit>
- Google Group. <https://groups.google.com/forum/?fromgroups#!forum/javascript-information-visualization-toolkit>

A JavaScript library that provides tools to create interactive data visualisations within web applications (strategic maps, hierarchical trees, relational maps, etc.). Because of its extensive variety of representations, this tool fits any developer's need.

Some of the most relevant features of this library are:

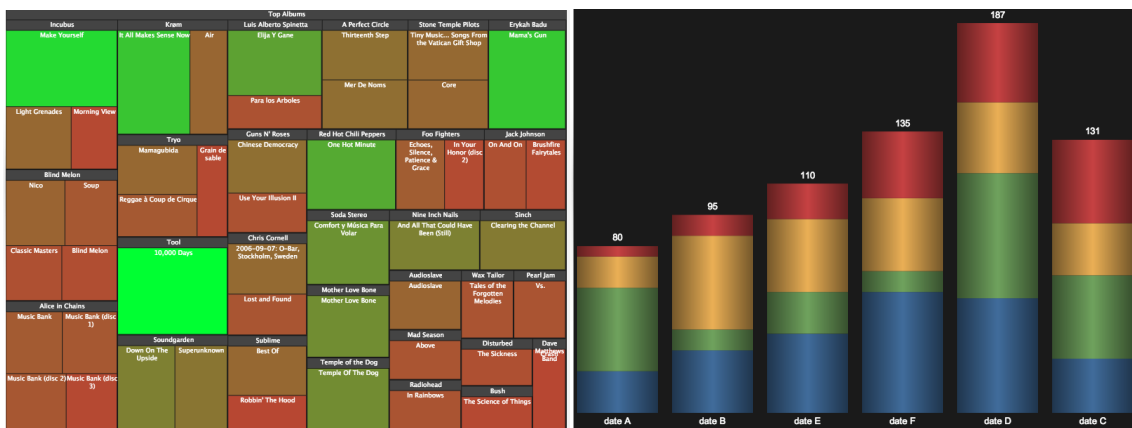
- Different types of data representations.
- Interaction with data in real time.

- Compatible with most browsers.
- Open Source resource easily integrated into web development.
- Extensible.
- Combines the visualisations to create new forms of representation.
- High processing speed for complex structures.

From the technical point of view, the representation of the data to be shown is noted in JSON (JavaScript Object Notation) format. This lightweight data exchange format is based on two structures: a collection of name/value pairs (object, record, structure, dictionary, hashmap, etc.), and an ordered list of values (arrays, lists or sequences). These universal JSON structures allow all programming languages to be easily adapted.

This toolkit has many possibilities and use cases:

- Development in BI (Business Intelligence) environments.
- Organisational charts.
- Strategic maps for dashboards (Balanced Scorecard).
- Statistical data maps.
- Relational Maps.



2.3.8 D3.js

TYPE. Library

TECHNOLOGY. JavaScript

LICENSE. BSD (allows use of the source code in non-free software)

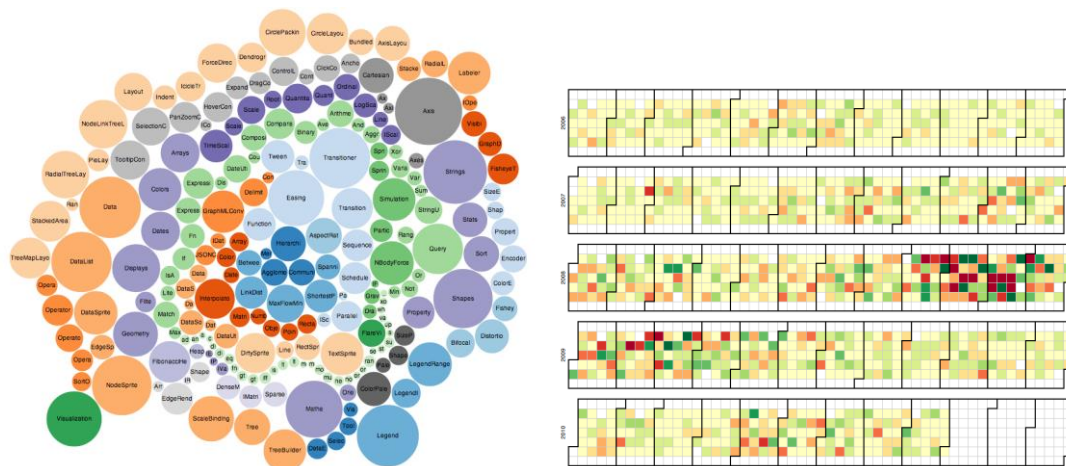
AUTHOR. Mike Bostock (United States)

LINKS.

- Website. <http://d3js.org/>
- GitHub Repository. <https://github.com/mbostock/d3>
- Gallery. <https://github.com/mbostock/d3/wiki/Gallery>
- Tutorials. <https://github.com/mbostock/d3/wiki/Tutorials>

A JavaScript library for creating complex visualisations and interactive graphics. Basically, the library allows users to manipulate data based documents using open web standards. Browsers may render complex visualisations without relying on proprietary software. Developments are open and can be used and adapted by other users. The possibilities are as vast as the geometry itself (bubbles, Chord diagrams, node links, etc.)

D3 allows binding data to the DOM (Model Objects for Document Representation) and apply transformations. For example, generating an HTML table from a set of numbers, and using the same data to create an interactive SVG graphic with transitions and interactions.



More examples.

- "Paths to the White House" (<http://elections.nytimes.com/2012/results/president/scenarios>)
- "Size of China's manufacturing industry" (<http://www.nytimes.com/interactive/2013/04/08/business/global/asia-map.html>)
- "Increased surveillance forces on the border between the U.S. and Mexico"

(<http://www.nytimes.com/interactive/2013/03/01/world/americas/border-graphic.html>)

- “Among the Oscar Contenders” (<http://www.nytimes.com/interactive/2013/02/20/movies/among-the-oscar-contenders-a-host-of-connections.html>)

2.3.9 Protovis

TYPE. Library

TECHNOLOGY. JavaScript

LICENSE. BSD

AUTHOR. Stanford Visualization Group (United States)

LINKS.

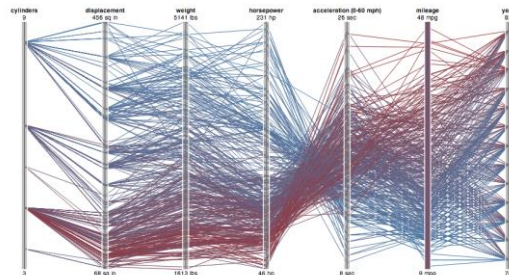
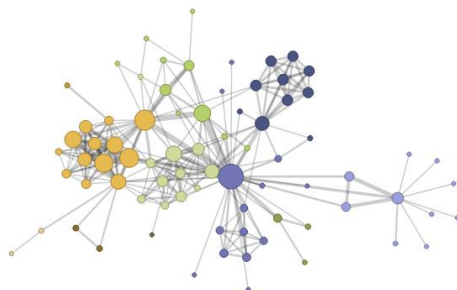
- Website. <http://mbostock.github.com/protovis/>
- GitHub Repository. <https://github.com/mbostock/protovis>

A JavaScript-oriented graphics library performing visualisations. It provides developers a large set of components and tools, enabling customisation of the displays with direct control.

Some of the most relevant features of this library are:

- Unlimited flexibility. It is based on a declarative grammar and data-driven framework.
- Simple graphics settings, based on chaining method.
- Focused on statistical graphics, its development method also enables structured, data-driven visualisations.
- It incorporates some statistical functions for data preparation.

The main Protovis' disadvantage is that it is a heavy library (weighing more than 700 Kb), designed for either Intranets or fast connections.



2.3.10 Recline.js

TYPE. Library

TECHNOLOGY. JavaScript

LICENSE. MIT

AUTHOR. Open Knowledge Foundation Labs (United Kingdom)

LINKS.

- Website. <http://reclinejs.com/>
- GitHub Repository. <https://github.com/okfn/recline/>

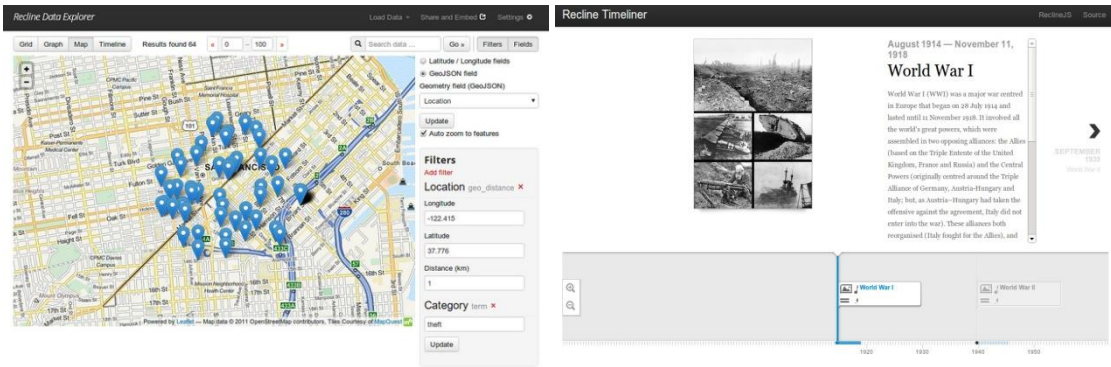
A Library for developing applications based on HTML and JavaScript. It is designed for integration, making it easy to integrate into other websites and applications. Aimed at developers with minimal much knowledge of programming, who use simple interfaces to view (and edit) data. The displays are available in graphical mode, map and timelines.

Recline runs on *Backbone*, this structure provides excellent support for building applications that handle relevant data loads, using models for the management of information and views to display them. Moreover, it is easily extensible through new *back-ends* for connecting a database or storage layer.

This library has many features for database manipulation, including insertion, search and update. It supports data loads from CSV, Excel, Google Docs, ElasticSearch, CouchDB and DataHub, among others. It features data cleaning and updating mechanisms, using a simple script.

The Recline library is composed of three modules:

- Model. Definition of the data structure (e.g., definition of the dataset to be used according to its source and data type).
- Backend. Connection of data by Recline.js API directly with the data source –i.e., a database, a CSV file, etc.
- Views. Sample and management of the information obtained and managed in the two previous instances.



C. Geospatial visualisation tools

The following tools can be used for representing geographic data.

2.3.11 OpenHeatMap

TYPE. Web application and API

TECHNOLOGY. JavaScript

LICENSE. GPL 3

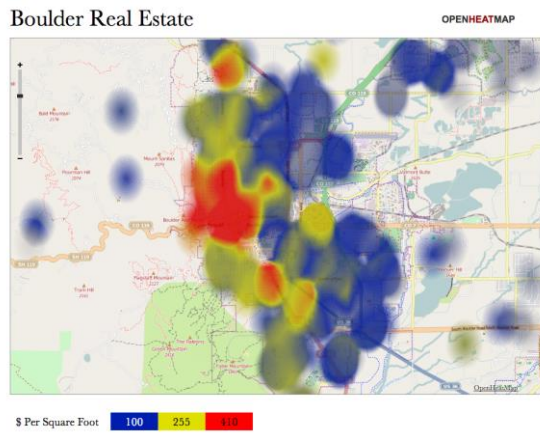
AUTHOR. Pete Warden (United States)

LINKS.

- Website. <http://www.openheatmap.com/>
- GitHub Repository. <https://github.com/petewarden/openheatmap/wiki>

A Web application used to convert statistical data in the form of spreadsheets into thermal maps. Its operation is simple and supports various formats as source files: Excel, CVS and linked Google Docs documents.

In order to locate data, the files must contain a specific column with the address or geographical location related to the data. OpenHeatMap enables sharing via email, and social networks, and even embedding maps in web pages.



2.3.12 OpenLayers

KIND. APIs

TECHNOLOGY. JavaScript

LICENSE. BSD

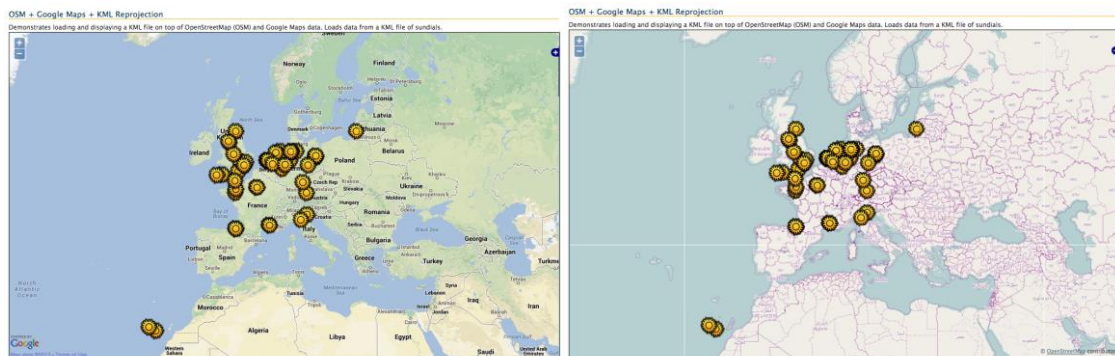
AUTHOR. Open Source Geospatial Foundation (United States)

LINKS.

- Website. <http://www.openlayers.org/>
- Documentation. <http://trac.openlayers.org/wiki/Documentation>

Open Source JavaScript library that allows adding maps in any web page with geographical references. It is a map viewer in JavaScript, therefore as a client-side library browsers can download directly all the resources via Ajax. No traffic is generated on the server; the maps are downloaded directly from the server of maps.

OpenLayers allows overlapping layers on a base, adding indicators or points on the map with legends and polygons. It also provides its own API to draw maps in a simple way. It includes a set of basic controls and a toolbar with advanced controls, fully customisable using the API.



2.3.13 OpenStreetMap

TYPE. Web application and API

TECHNOLOGY. Ruby, PostgreSQL,

LICENSE. CC BY-SA

AUTHOR. OpenStreetMap Foundation (United Kingdom)

LINKS.

- Website. <http://www.openstreetmap.org/>

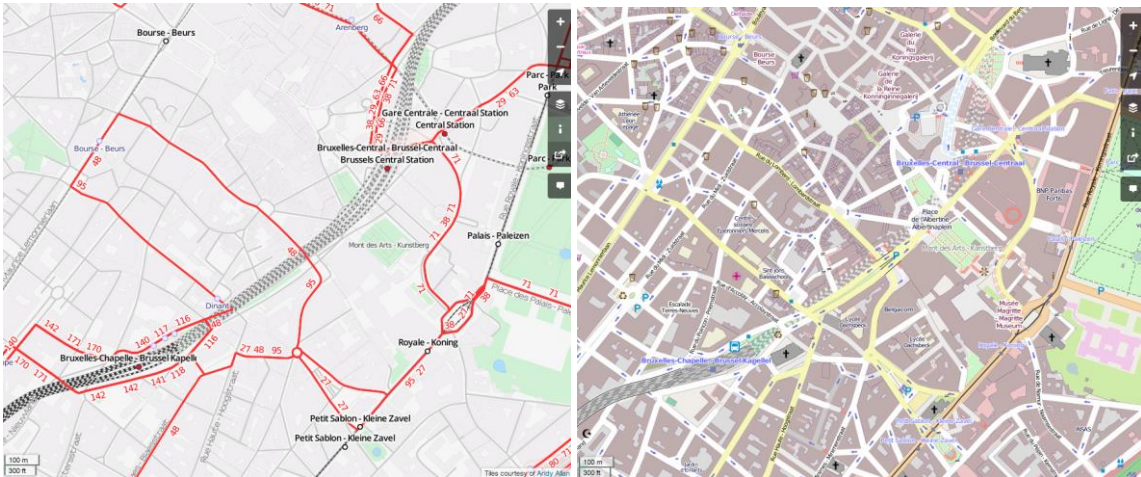
A collaborative project that contains free and editable maps. Maps are created using geographic information captured with mobile GPS devices, orthophotos and other free sources. This cartography, both the images created as vector data stored in the database, is distributed under the Open Database License (ODbL).

Registered users can upload GPS tracks, create and edit vector data using tools created by the OpenStreetMap community.

OpenStreetMap uses a topological data structure. Data are stored in WGS84 datum lat/lon (EPSG: 4326) Mercator projection format. The basic elements of OSM maps are:

- Nodes. Points collecting geographical positions.
- Ways. Ordered list of nodes representing either poly-lines or polygons (when a poly-line starts and ends at the same point).
- Relations. Groups of nodes, paths and other relationships that can include specific common properties. For example, all those roads that are part of the Camino de Santiago.
- Tags. Key/value pairs which can be assigned to nodes, ways or relations. For example: highway=trunk

Data attributes follow a more complex than the social folksonomies. The ontology to describe map features (mainly the meaning of the labels) is maintained from a wiki.



D. Temporal data visualisation tools

A set of tools for data analysis when time is an important component.

2.3.14 TimeFlow

TYPE. Desktop application

TECHNOLOGY. JavaScript

LICENSE. Free

AUTHORS.

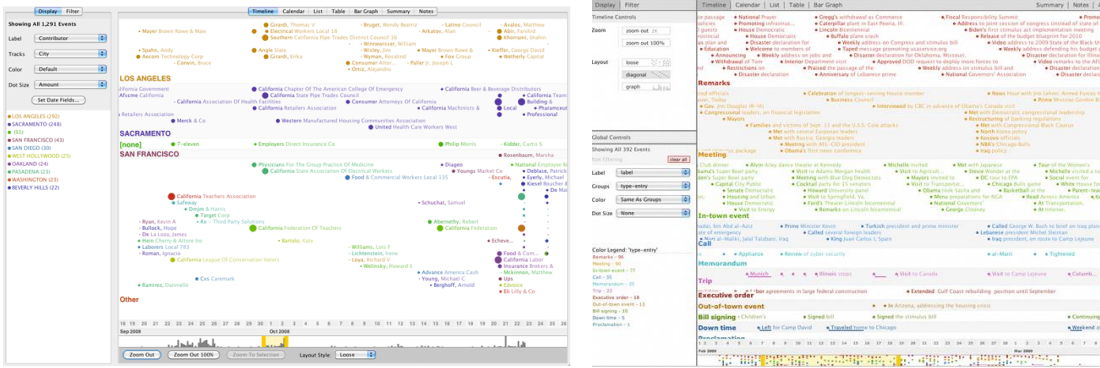
- Fernanda Viegas, Martin Wattenberg (Flowing Media, United States),
- Sarah Cohen (Duke University, United States)

LINKS.

- Website, GitHub Repository. <https://github.com/FlowingMedia/TimeFlow/wiki>

A Visualisation tool for temporal data. The current release is in "alpha" version, so it may contain errors. This tool helps to analyse temporal data through five different views:

- *Timeline view.*
- *Calendar view.*
- *Bar chart view.*
- *Table view.*
- *List view.*



2.4 Tools for network analysis

This kind of tools is interesting for social networks analysis, where people and connections among them can be represented from different data sets.

In order to use this software category it is needed to understand the statistical theory for network nodes analysis.

2.4.1 Gephi

TYPE. Desktop application

TECHNOLOGY. Windows, Linux, MacOS X, Java

LICENSE. CDDL, GPL3

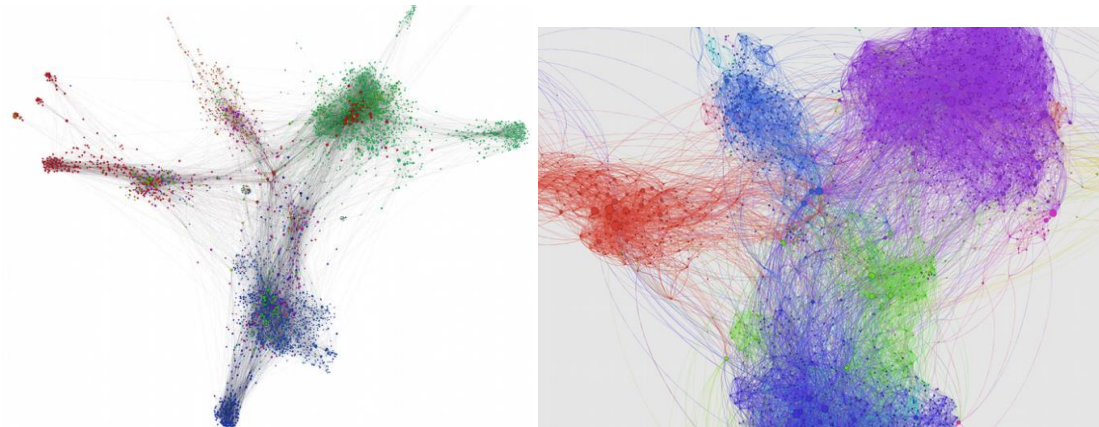
AUTHOR. Gephi Consortium (France)

LINKS.

- Website. <http://gephi.org/>
- Documentation. http://wiki.gephi.org/index.php/Main_Page/

A platform for the interactive visualisation and exploration of networks and complex, dynamic and hierarchical graphs. It displays the relationship between data and its evolution, grouping sets, representing hierarchies, exporting and importing tables, among other functions.

It can handle large graphs, and networks with up to 50,000 nodes and one million edges.



2.4.2 NodeXL

TYPE. Desktop application

TECHNOLOGY. Microsoft

LICENSE. Microsoft Public License (Ms-PL)

AUTHOR. Social Media Research Foundation (United States)

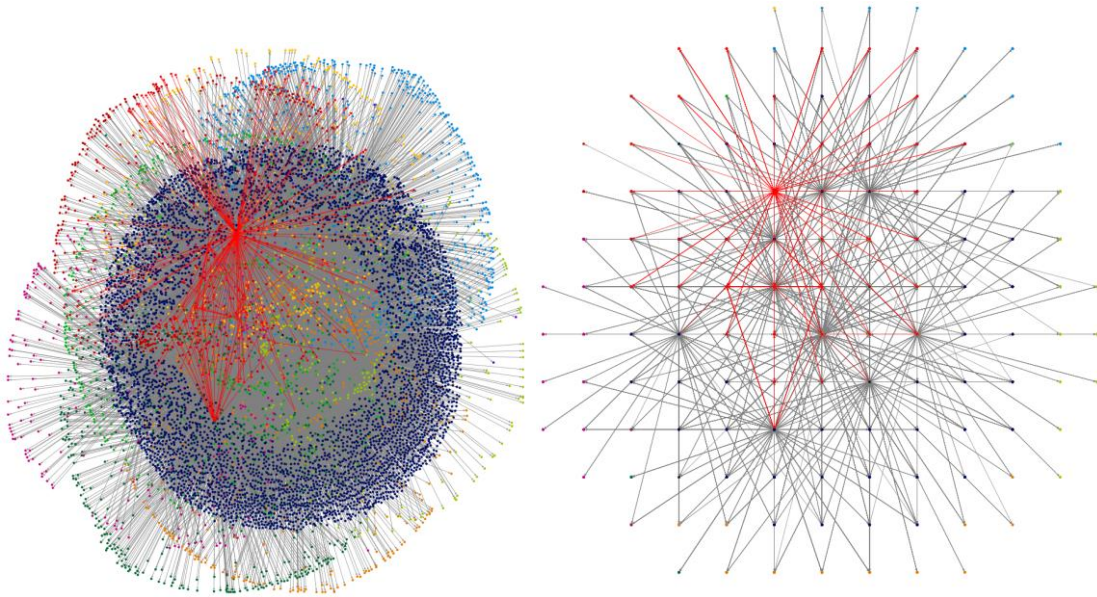
LINKS.

- Website. <http://nodexl.codeplex.com/>

A Powerful analysis and representation tool that works with Excel. It renders network graphics from a given list of connections, helping in the analysis and discovery of patterns and relationships in data.

Some of the most relevant features of this tool are:

- Flexible import and export. Importing data from multiple sources.
- Direct connection with social networks. Optimised for analysing online social media, including connections to query built in APIs from Twitter, Flickr and YouTube.
- Flexible design.
- Duplicate links combination.
- Metrics calculation and network analysis.
- Image insertion of network sub graphs.
- Automating tasks.



3 Comparison

Below is a summary table presenting a comparison of all the visualisation tools stated and assessed in this document.

Tool	Category	Multi-purpose	Type	Technology	License	Platform	Data Storage	Web Publication?
DataWrangler	Data cleansing	No	Web application	HTML	Free	Browser	External server	No
Google Refine	Data cleansing	No	Desktop application	Java	BSD	Browser	Local	No
Mr. Data Converter	Data converter	No	Library	JavaScript	MIT	Browser	Local	No
The R Project for Statistical Computing	Statistical analysis	Yes	Programming language	R	GPL	Linux, Mac OS X, Unix, Windows XP	Local	No
Google Fusion Tables	Visualisation application/service	Yes	Web application, API	JavaScript, Flash	Free	Browser	External server	Yes
Tableau Public	Visualisation application/service	Yes	Desktop application	Windows, JavaScript	Free	Windows	External public server	Yes
Many Eyes	Visualisation application/service	Yes	Web application	Java, Flash	Free	Browser	External public server	Yes
CartoDB	Visualisation application/service	Yes	Web application	JavaScript, PostgreSQL	Commercial	Browser	External server	Yes
GeoCommons	Visualisation application/service	Yes	Web application, API	JavaScript, Ruby	Several	Browser	Local or external server	Yes
Google Chart Tools	Visualisation library, service	Yes	Library	JavaScript	Free	Code editor and browser	Local or external server	Yes
JavaScript InfoVis Toolkit	Library	Yes	Toolkit	JavaScript, Python	MIT	Code editor and browser	Local or external server	Yes
D3.js	Library	Yes	Library	JavaScript	BSD	Code editor and browser	Local or external server	Yes
Protovis	Library	Yes	Library	JavaScript	BSD	Code Editor	Local	No
Recline.js	Library	Yes	Library	JavaScript	MIT	Code Editor	Local or external server	No
OpenHeatMap	GIS	No	Web application, API	JavaScript	GPL 3	Browser	External server	Yes
OpenLayers	GIS	No	API	JavaScript	BSD	Code editor and browser	External server	Yes
OpenStreetMap	GIS	No	Web application, API	Ruby, PostgreSQL	CC BY-SA	Browser or desktop running Java	Local or external server	Yes
TimeFlow	Analysis of temporal data	No	Desktop application	JavaScript	Free	Desktops running Java	Local	No
Gephi	Network analysis	No	Desktop application	Windows, Linux, MacOS X, Java	CDDL, GPL3	Desktops running Java	Local	As picture
NodeXL	Network analysis	No	Desktop application	Microsoft	Microsoft Public License (MS-PL)	Excel 2007 and 2010 on Windows	Local	As picture

4 Conclusions and recommendations

As a general conclusion, a large number and diversity of free visualisation tools is available on the market. Thus, it can be stated that this is a period of great proliferation of raw data and there is a growing interest in finding the most appropriate way to present this information in an attractive, clear, concise and understandable way for the end user.

Although there are many information and data visualisation tools, below are listed the most recommended, based on the capabilities provided and the level of experience required to use them.

In the category of Web applications, we have selected:

- **Google Fusion Tables.**

An excellent tool for beginners or for those with no programming skills. In the case of the more technical users, API is available that can produce graphs or maps from information.

One advantage of this application is the variety of data representations provided to the user. In addition, it can create graphics and maps without being very time consuming, It offers GIS functions to analyse data by geography. This service automatically provides geocoding addresses, which is useful when locating many points on a map.

Google allows users to use data: private, unlisted or public, even though the data remains stored at Google's servers. The external storage of data becomes a drawback, considering the problem of privacy.

- **CartoDB.**

An Open Source service addressed to a variety of users, regardless of technical level, and with a user- friendly interface.

It is important to highlight that there is an active group of developers that provide extensive documentation and a large number of examples. The openness of the API fosters the continuous development of new integrations and the enhancement of the capabilities, with new additional libraries.

Among its customers are highly respected institutions such as the UN, Google, NASA,

the University of Oxford, and Yale.

The use of libraries and APIs allows the developer to create tailored views, according to the project's needs.

- **Google Chart Tools.**

This API has two performance modes, providing static graphics as pictures (simpler) and interactive graphics (more powerful).

Generation of static graphics in images is based on requests to Google's servers. It is easy to use, and it offers a variety of chart types and options. *Live Chart Playground* is a tool which enables the generation of URLs of graphs to embed in HTML code, as well as preview the changes of parameters made in real-time. Also, there is no need to install any component or server in local/external environment because graphs are generated dynamically on the fly.

The mode of interactive graphics generation, through a JavaScript library, is more complete and can generate functional graphics. The drawback of this tool is that, as with other JavaScript libraries, it requires additional scripting code.

It is free to use but with some limitations of requests per URL and day. The current limit stands at 250,000 requests.

- **Recline.js.**

This library is easy to use for users who do not have extensive programming skills. It is considered a versatile library due to its modularity. This means that only the needed modules are used to build the application.

Another advantage is that views can be embedded in other applications, just as done for CKAN and DataHub.

Among the tools available for representing geographic data, we can highlight:

- **OpenLayers.**

A powerful library that requires advanced knowledge in the GIS field. The big advantage is that it does not require the use of licensing like in the case of Google maps.

It is an interesting option for those who are used to programming in JavaScript and they prefer not using commercial platforms such as Google or Bing. It is fully compatible with proprietary technologies. In addition, it is WMS and WFS standard compliant.

It allows a huge range of possibilities: importing features in KML, like polygons with islands; positioning features from RSS feeds; integration with jQuery Mobile; adding animations to polygons to represent, for example, trajectories.

About the Author

datos.gob.es is the Spanish Open Data Portal, launched in 2011 and promoted by the Government of Spain through the [Ministry of Industry, Energy and Tourism](#), and the [Ministry of Finance and Public Administrations](#). This portal is directly managed by the State Secretariat for Telecommunications and the Information Society (SETSI).

Copyright information

© 2013 European PSI Platform – This document and all material therein has been compiled with great care. However, the author, editor and/or publisher and/or any party within the European PSI Platform or its predecessor projects the ePSIplus Network project or ePSINet consortium cannot be held liable in any way for the consequences of using the content of this document and/or any material referenced therein. This report has been published under the auspices of the European Public Sector information Platform.



The report may be reproduced providing acknowledgement is made to the European Public Sector Information (PSI) Platform.