Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) "Review and Evaluation of Systems Supporting Data Journalism", in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

Review and Evaluation of Systems Supporting Data Journalism

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris

INTRODUCTION

In recent years, a new form of journalism has developed a strong dynamic: Journalism based on data. An important factor in its success is the availability of data from many public bodies and international organizations as part of the movement of open data worldwide. This form of journalism is based on the creation of stories that are based on data (storytelling), which are presented with attractive visualizations. Journalism of this type is practiced by both professional journalists and ordinary citizens (citizen journalism). The success of the move is so great that technologies and applications that support it have begun to emerge. Applications supporting data journaling have been developed in a wide range of domains that serve specific purposes in the data processing process and satisfy a wide range of users. Data processing involves many activities, but the key steps for completing the process are: data collection and cleaning, analysis and visualization, and ultimately the publication of the story. Although technologies are in a period of mature productivity, their evaluation is an area lagging in development. It is a fact that lack of evaluation based on a reliable methodology is evident in the literature. Finding a suitable methodology for this purpose is particularly important in order to (a) evaluate systems supporting data journalism resulting in a specific ranking of potential, and (b) make it easier for the user to choose the application that best suits his / her requirements and cognitive level. This study will attempt to evaluate specific applications with the assistance of appropriate methodology. A comparative evaluation of 9 applications related to the visual imaging component was attempted based on a methodological approach with usability and functionality criteria. The aim of this study is to provide a quick evaluation method, open to proposed improvements and further refinements, in order to establish a framework for qualitative assessment of data journalism applications.

BACKGROUND

A system's evaluation is the process of testing and validating whether the system achieves its predefined goals and in what degree. Although in literature there are several qualitative and quantitative techniques proposed for different kinds of systems, there is not yet enough research for data journalism systems evaluation, so a goal for the area is for an evaluation framework to be established. Relative background can be found in related areas (e.g. information visualization, HCI), while the most suitable criteria for the proposal of such a framework are: functional and usability features.

Yi et. al. (2007) have collected the techniques from various research papers by creating a classification for InfoVis (Information Visualization) related to the interaction techniques. The following three classifications focus on interaction techniques, while the fourth refers to the tasks of the system's user, i.e. it focuses on his/her goals without paying attention to the interaction. The four proposed classified techniques with their related research work are:

Low-level Techniques. Buja et al. (1996) classified the interaction techniques into three classes: *Focusing*, *Linking*, and *Arranging* views. Chuah & Roth (1996) summarized a set of basic visualization interaction (BVI) operations: *Graphical* operations, *Set* operations, and *Data* operations. Dix & Ellis (1998) condensed the different ways in which interactivity adds value and resolves conflicts in representation: Highlighting and focus, accessing extra information – drill down and hyperlinks, Overview and context - zooming and fish-eye views, Same representation / changing parameters, Same data / changing representation, Linking representation – temporal fusion. Keim (2002) proposed that in addition to the visualization technique, it is necessary to use some interaction and distortion techniques for an effective data exploration: Dynamic projections, Interactive filtering, Interactive zooming, Interactive distortion, Interactive linking and Brushing. A categorization of interactions can also be found in Wilkinson (2005). Interaction techniques are classified into seven categories: Navigating, Manipulating, Brushing and linking, Filtering, Rotating, Transforming, Animating.

Interaction Techniques. Tweedie (1997) considered interaction techniques based on the following taxonomical dimensions: Interaction types (manual, mechanized, instructable, steerable, and automatic) and Directness (direct and indirect manipulation). Spence (2007) concerned with taxonomy interaction techniques as follows: continuous, stepped, passive, and composite interaction.
Interaction Functions. Ward &Yang (2004) proposed a framework for interaction techniques, identifying distinct classes and shared concepts. The classification of interaction operations follows:

- interaction operators
- interaction operands & spaces, data value-spaces, data structure-space, attribute-space, object-space, visualization structure-space
- interaction parameters

User Tasks. Concerning the taxonomies of user tasks, Zhou & Feiner (1998) proposed that visual tasks serve as an interface between presentation intents and low-level visual techniques, they can be characterized along two dimensions: visual accomplishments and visual implications. The set of features which describe various aspects of visual organization is: Visual grouping, Visual Attention, Visual Sequence, Visual Composition, Visual Signaling. Finally, Amar et al. (2005) present a set of ten low-level analysis tasks such as: Retrieve value, Filter, Compute Derived Value, Find Extremum, Sort, Determine Range, Characterize Distribution, Find Anomalies, Cluster, Correlate.
Yi et. al. (2007) bridge these approaches and propose an evaluation scheme based on three axes for evaluating interaction models:

1. Descriptive power: The ability to describe a significant range of existing interfaces.
2. Valuable power: The ability to evaluate alternative design modes.
3. Generic power: The ability to offer help to design new interfaces.

Few & Edge (2012) list the features that visualization tools should support in certain categories: Visualizations, Interaction, Multi-Chart Displays, Statistical Calculations, Speed of Response, Data Access and Integration, Output and Content Management, Platform Options, Ease of Learning and Use, Programmability, Advanced Features.
Another approach of evaluating visualization software by Atwood & Reznik-Zellen (2018) is based on the use of Rubrics, that is, tables that contain characteristic criteria that are graded on a 3-level scale. The Rubric of writers is based on 13 criteria in order to evaluate technologies in a holistic high-level way. These criteria are divided into two categories with functional and usability features (as shown in detail in Table 1). The score ranges from 0 to 3 with 0 meaning poor, 1 modest, 2 good and 3 excellent. For the purpose of the evaluation in this article, the approach of Atwood & Reznik-Zellen (2018) was chosen due to greater applicability to data visualization technologies. It can be broadly applied to technologies that primarily target interactive graphs, maps (mainstream), networks or combinations of them, on which we will focus on this study.
We have added one usability criterion in the original approach (No. 14 - Table 1), which refers to the ability to store on a cloud-based platform (cloud service) with some temporal/spatial restrictions and the possibility of visualizations privacy settings. This means we can discern the difference between the storage of an image file format or other type and that of cloud-based storage that is individually rated with spatial/time constraints. In this category, if the manufacturer of the application did not refer to time or storage limitation, it was considered that there is no such limitation.

REVIEW OF SYSTEMS SUPPORTING DATA JOURNALISM

Our goal is to evaluate tools that help a journalist support their own story with the means provided by technology. Because technologies are enough and cover different sectors in the field, often overlapping with each other, we have used some inclusion/exclusion criteria for the systems to be evaluated. The following software systems were selected on the basis that a data journalist should have no specialized programming skills, but he can use a tool that can learn easily, in a short time in order to bring the story to his public in the best possible way. The inclusion / exclusion criteria are as follows:

- Tools that need code input to create visualizations were excluded (e.g. google charts).
- We included tools that could be used for a reasonable amount of time without restrictions on their basic functionality either as a desktop version or as web-based applications.
- Applications to be evaluated could be made available without any financial consideration for use by interested parties for some time.

The evaluation will be done separately for each application with a description of its purpose, a brief explanation of its functions, some observations we made when testing each one in individual attributes, and the score in the attributes selected with the individual comments in each category. To control entry capacity, we used 3 xlsx files of about 5 MB (120000 lines), 9 MB (240000 lines), 30 MB (865000 lines). The evaluation of data journalism systems has been conducted by the first author of the article be testing the tools on certain tasks (specifically the creation of interactive charts, maps and network graphs) for producing journalistic stories, using the aforementioned datasets.
Next, we present a short review for the applications being evaluated in the present article: Tableau, Gephi, Cytoscape and web-based applications Datawrapper, Plot.ly, Datavisual, Infogram, Datacopia & Filtergraph.

1. Cytoscape (desktop version 3.6.1)

Purpose of use: An open source software platform to visualize complex networks for each type of data.
Description: Although originally developed to assist research in biology, it is also used in other areas of network analysis (Shannon et. al., 2013). It is designed to analyze and visualize large-scale networks. The input accepts a variety of file formats (sif, nnf, gml, BioPAX etc.). The network can be created either sequentially by the user with the possibility of inserting pictures, shapes or text frames, or automatically by the system by entering data from a table where the user guided by the interface defines the details of the network. Tools for handling visualization network include functions such as scale, align, rotate, distribute and a variety of algorithms for graph layout. A feature of the menu is the access to available-to-install apps that perform specialized functions. Filtering can be applied to select a subset of nodes and edges that meet certain criteria Network layout algorithms support various representations with a variety of features. The publication of visualization is supported by several options.

2. Datacopia (web-based app)

Purpose of use: Web application or making interactive charts.
Description: Data input is done via URL in csv or tsv format, uploading the user file in csv, tsv, txt, xls, xlsx and finally by typing or pasting data to a form denoted by commas. Visualization supports many kinds of charts (including maps). The interface automatically calculates a number of charts of various types such as scatter plots, function plots, histograms, spider charts, etc. (in our case, 189 charts) that are displayed to the user. The user can choose a chart of his / her choice that can interfere with very basic formatting parameters. Saving is performed in low resolution png. The distribution of the chart can be done through Reddit, Google+, Pinterest, Twitter, Facebook with rudimentary cloud storage support, but without interactivity and the ability to create a personal account.

3. Datavisual (web-based app, beta version)

Purpose of use: Web platform for the design of selectable charts through a collection of interface templates.
Description: The user selects the type of chart he/she wants and through the appropriate filters he/she formats the chart. The design is aided by various tools such as sliders, color pickers, on/off function buttons and text boxes. The data input is done with an excel, csv or tsv file and through a special menu he can interfere with the degree of interaction of the chart through tools such as display of information when the pointer is at some point (hover display, hover label size, etc.) and animation techniques for chart presentation (Bar duration, Bar delay, Label Duration, etc.). Storing is done in low-resolution jpeg format while other file types and web embedding are supported after upgrading the account. Only the English language is supported while the output can be saved in low resolution jpeg. No support material is available in any form, and no information is provided about the hosting conditions of the charts on the server site.

4. Datawrapper (web-based app version 1.20.0)

Purpose of use: An online application of interactive charts and maps.

Description: The application has been created with the participation of a group of journalists who have worked in major media such as the New York Times, NPR, Deutsche Welle, Bloomberg, etc. It is available in a variety of versions from the trial one (free of charge and it is the one we tested) up to a business version. The free choice has the limitation of 10000 tests; the full functionality of the application (output formats and styling, etc.) is not supported, while canceling the registration can be done without limitations though the fate of visualizations after the cancellation is not specified. The user does not need a cognitive background and the result comes out through a few guided by the application steps. Account creation is required to store and share charts. The result is finalized through 4 directed steps that are: upload, check &describe, visualize and publish & embed.
Statistical analysis is not supported except a graphical trend line with 6 defaults in the Refine tab. Output formats are not supported except storage on a cloud-based platform. The version is available in 15 languages.

5. Filtergraph (web-based app)

Purpose of use: Web platform drawing charts and maps.
Description: Filtergraph is an interactive web-based tool for displaying large data sets. Although originally designed to meet the needs of astronomers in a wide range of projects, it can be used to manage data in any situation where large-scale visualization of data, including academic research and data-based publishing, is required. The input accepts a variety of file formats. Initially, the user has the ability to specify names and restrictions of access to the content of the portal (filtergraph.com/NameOfPortal), and also it is possible to delete all of its saved content. He can invite readers or registered users via mail using a link with view or view and edit permissions. The interface contains column data filter and the possibility to highlight them, undo function and 10 graph options (map, scatter, line, error bars, histogram, stacked histogram, heatmap, surface map, bar chart, pie chart). Storing is done in 6 different formats and a statistical description function is provided. There is a difficulty in displaying categorical data (setting capability, displaying the desired number on the axes), while hover mode is not supported.

6. Gephi (desktop version 0.9.2)

Purpose of use: Open source software for exploring and manipulating networks.
Description: The application runs on Linux, Windows and Mac and requires installed version of Java 7 or 8. Data entry is done with a spreadsheet, other types of graph files or via online connection to MySQL database, SQLite, PostgreSQL etc (Bastian et. al. 2009).
Gephi uses 3D graphics to showcase large networks in real time, speeding up their exploration. Provides individual functions that facilitate the import, visualization, filtering, fixation, spatial representation and output of all types of networks. The user interface consists of workspaces where individual tasks can be performed. Nodes or edges can be obtained manually or by using the filter system. Filters can choose nodes or edges with boundaries, range, and other properties. The frames of the filters are designed in a chain way so that the next frame accepts as input the output of the previous one. The graphics elements accept as input a group of nodes and modify the visualization parameters. The architecture of the system supports graphs whose structure or content varies over time. A data source can send network data to the potential controller at any time and visualize the result in real time. Architecture is interoperable. Nine languages are supported including English.

7. Infogram (web-based app)

Purpose of use: Web platform for charting, maps and infographics.
Description: Use of the service requires a simple registration or use of a google / facebook account. Data can be imported using excel spreadsheets, cloud services like dropbox, OneDrive or json files. An open data finder is supported. The interface enables the creation of some type of infographic, chart or map. In the infographic option the user can select one of the available templates, add title and text, adjust the font, align, and other options that text editing software offer. In the field of charts, there is

available a variety of types of charts with color adjustment functions of their size, with the addition of the labeling of axes and the appropriate range of values to the axes. The construction of maps in free version is only supported in two types: Global and United States. The result can be publicized using a hyperlink in social media or embedded on a website or blog. Five languages are supported.

8. Plot.ly (web-based app)

Purpose of use: Web platform creating charts and control panels (dashboards).
Description: The platform is used after signing up with a Google account, twitter, facebook, GitHub, or simple email. Data entry is done in csv, xls, xlsx, via URL, or linked to a SQL platform. A diversity of chart types is supported, which are divided into 6 main categories. Basics (scatter, line, bar, Area, pie), stats (histogram, box plot, 2d histogram heatmap, 2d histogram contour plot, parallel coordinates), scientifically (error bars, heatmap, contour, ternary plot, carpet, Sankey), Maps (Chloropleth, Atlas maps, satellite maps), 3d (3d Scatter , 3d line, 3d surface, 3d mesh), economical (range slider, candlestick, OHLC, animations). It supports filtering in the dataset with immediate overview of the changes. A wide range of formatting functions is supported. Storage in the chart and the data has privacy settings.

9. Tableau public (Desktop version 10.5)

Purpose of use: A software for data analysis through interactive and shared visualizations depicting trends, variations and density of data in the form of a plurality of graphs, charts and maps.
Description:The basic steps for creating a data processing report include:

1. Linking to a data source (files, SQL databases or other types of cloud-based platforms).
2. The choice of dimensions and measures (dimensions are categorical data while the measures are numerical) for the analysis.
3. Applying visualization techniques using filters, grouping, color management, captions etc.
4. Creating multiple worksheets that create different views on the same or different data.
5. Create and organize a dashboard that contains multiple worksheets that are linked to each other.
6. Creating a story based on a sequence of worksheets or dashboard that work collaboratively to transfer information.

Also, a statistical analysis of the data as well as a wide range of operators, functions and calculations in numbers, strings and dates is supported.

EVALUATION OF SYSTEMS SUPPORTING DATA JOURNALISM

Table 1 shows the evaluation criteria (Rubric) and summary tables 2 and 3 show the comparative evaluation of the applications with the corresponding score that resulted after their evaluation by the first author of the article. Figure 1 shows the bar graphs of the overall comparative evaluation.

Total evaluation

Desktop applications due to more and more powerful features show higher scores. The web-based applications generally exhibit lower levels of usability and functionality features as described in detail.

Note that the joint evaluation in Tableau public, Plot.ly & Datawrapper was made independently of that of Atwood & Reznik-Zellen (2018) with some differences in individual scores in the 13 common evaluation criteria. The other tools were evaluated by the author without any other similar assessment in the literature. A comparative table of scores is shown in Table 5. In Table 5, yellow color depicts our evaluation, while the pink is that of Atwood &Reznik-Zellen (2018) and the shades of green show the differences in the score. The dark shade is a difference of 3 points, the middle shows a difference of 2 points and the low is a difference of 1 point.

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) “Review and Evaluation of Systems Supporting Data Journalism”, in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) “Review and Evaluation of Systems Supporting Data Journalism”, in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

No	Category	Feature	Poor (0)	Medium (1)	Good (2)	Excellent (3)
1	Usability	Ability to handle	There is no control of the output design	Low handling capacity. (modification of one or two elements)	Some ability to handle. (modification of many items but with limitations)	Full control of all aspects of the output design.
2	Usability	Support materials	There is no documentation.	Insufficient documentation with little information about key features	Almost sufficient documentation. It leaves out some features.	Well documented. Available technical support (online chat, troubleshooting, videos, webinars, etc.
3	Usability	Ease of use	Graphics and hyperlinks are not labeled; fonts are difficult to read, poor navigation, need of a cognitive background, that is, not accessible to all users.	Some available tags, navigation is not clear.	Many tags available, navigation is simple	Graphics and hyperlinks are well-marked, fonts are legible and do not require extensive knowledge.
4	Usability	Learning time	It takes a long time (> 5 hours)	It requires a moderate time (1 hour - 5 hours)	It takes less than 1 hour. (from 10 minutes to 1 hour)	It requires a minimum of time. (totally intuitive or less than 10 minutes)
5	Usability	Making use of previous knowledge	Previous experience with other software, code, or previous versions is required	Previous experience with other software, code or previous versions is recommended. Quite intuitive.	It requires little experience with other visualization platforms	Totally intuitive
6	Functionality	Processing power	Only accepts files smaller than 5MB.	It can accept files of limited size.	It can upload/download file sizes of any size with a login	User can upload any file size without limitations
7	Functionality	Login requirements	A connection is required for all functions.	A connection is required for most functions.	A connection is required for some specialized functions but access is possible for routine functions.	No connection is required for full function.
8	Functionality	File type requirements (Input)	Only accepts one or two types of files in specific formats.	Accepts only a few types of files	Accepts multiple types of files of various types.	Accepts all file types.
9	Functionality	Cost for “full” version	High cost. (over $ 100 / year)	Medium cost. (between $ 100 - $ 20 / year)	Low cost. (below $ 20 / year)	Totally free.
10	Usability	Built-in analytics	No analysis supported.	Available basic mathematical functions.	Many functions available.	Complete range of available analysis methods.

11	Functionality	Output	Output is not supported.	Limited output supported (1 or 2 formats you can export), low resolution.	Some supported output- either with many file types, but with limited numbers or with many downloads available with very limited output formats.	Many formats and unlimited downloads, high resolution.
12	Functionality	Data collection	It gathers and maintains all information that has been uploaded, allows third parties access to information.	It collects personally identifiable information.	Collect cookies.	It keeps little or no information.
13	Usability	Aesthetics	Minimalist style, few options, limited color palette, limited fonts...	Minimalist style but some flexibility, more charting choices, or style options.	Various style choices and chart types.	Many style options available, such as color, font, chart type, etc.
14	Usability	Save on a cloud-based platform for public sharing with or without limitations	Paid or no storage capability.	Free for a period of time or limited to the number of visualizations without any privacy settings.	Free for some time or with a limitation of the number of visualizations with privacy settings.	No limitations and privacy settings.

Table 1: Modified evaluation criteria (Rubric).

Evaluation Feature	1. Cytoscape (3.6.1)	2. Datacopia	3. Datavisual	4. Datawrapper	5. Filtergraph
Ability to handle	Full control of all aspects of output design. **(3)**	Insufficient control in the output design of visualization with interference in some elements. **(1)**	The user has partial control over the final output of his / her work mainly in formatting the chart. **(2)**	The user has partial control over the final output of his work through guided steps. **(2)**	Some ability to handle. (Modification of many elements but with limitations). **(2)**
Support materials	A user manual is provided with detailed documentation, online presentations, FAQs, mailing list support **(3)**	Only one demonstration of the basic charts in a gallery without explanations is provided. **(0)**	Documentation is not supported in the basic version. **(0)**	A knowledge base with a search engine is supported with relevant articles on application functionality and a blog coordinated by the production team with news, updates and ideas from the community. **(2)**	A video and a help page are provided with detailed information about the functions. **(2)**

Ease of use	There are many available and comprehensible labels to facilitate navigation. **(2)**	Few labels without additional assistance to help navigate. **(1)**	Accessible to the beginner user with direct result through two steps: selecting & importing data. **(2)**	Very good UI, user friendly, with the ability to instantly track the changes made **(3)**	It takes a lot of time to familiarize the novice user with the interface menu. **(0)**
Learning time	It takes a long time to familiarize the average user with the capabilities of the program. **(0)**	Minimum time is required to visualize the data. But it takes time for selecting the appropriate graph that satisfies the user. **(3)**	Minimum time is required < 1 hr to fully enrich the novice user with the application interface. **(2)**	It requires time < 1 hr for a novice user to familiarize with its functionality through strictly defined steps. **(2)**	It requires moderate time. The user must experiment with the capabilities of the application in order to achieve the desired result. (1 hrs - 5 hrs). **(1)**
Making use of previous knowledge	Graphics of the interface are quite functional but for the navigation menu the previous experience with similar software can reduce the learning time. **(1)**	The construction of the diagrams is automated with few intervention options by the user so it can be described as intuitive. **(1)**	Intuitive to a great extent without requiring previous experience from similar applications. **(3)**	Intuitive to a great extent, since possible actions of the user are guided by the interface diligently. **(3)**	Previous experience with other software or earlier versions is suggested to allow the user to quickly use it. **(1)**
Processing power	Editing the graph binds memory to the user's PC, so the performance depends on the characteristics of RAM and the system graphics card. **(3)**	We attempted to upload the 5MB file into the application without success. **(0)**	A 5MB file slows the browser to a large extent, so it does not finish loading the file within a reasonable amount of time. **(0)**	A 5MB file is significantly slowing the browser (firefox quantum 63.0.1 64 bits), so the upload of the file is not completed within a reasonable amount of time **(0)**	We uploaded a 30MB file (approximately 860000 lines) successfully. **(2)**
Login requirements	No connection required for full operation. **(3)**	No connection required for full operation. **(3)**	Editing and formatting are done by connecting to a user's personal account. **(0)**	Editing and formatting are done by linking to a user's personal account. **(0)**	A connection is required for all functions. **(0)**
File Type Requirements (Input)	Although it accepts various types of files, ODF formats such as .ods spreadsheets are not included. **(2)**	Accepts files such as .csv, .tsv, .txt, .xls, .xlsx. **(1)**	Formats such as xls, csv, and tsv are supported **(1)**	Formats such as xls, csv, google spreadsheets, and a csv file hosted on another server are supported, with the ability to automatically refresh the file and make changes in the visualization from time to time for the month following publication. **(1)**	It accepts files of various types such as ASCII text files (space separated, comma separated, tab separated, and fixed length), Microsoft Excel files (*.xls, *.xlsx), SQLite files (*.sqlite), VOTable files (*.vot, *.xml), FITS files (*.fits), IPAC files (*.tbl), Numpy files (*.npy), HDF5 files (*.h5) **(2)**
Cost for “full” version	Free. **(3)**	Free. **(3)**	Free for basic version and for full (pro) 15 $ / month.**((0)**	Free for single version with a limit of 1000 views for 1 user while for full (custom) 279 €/month. **(0)**	Free. **(3)**

Built-in analytics	A complete set of metric statistical analysis is available such as: AverageShortestPathLength/BetweennessCentrality/ClosenessCentrality/ClusteringCoefficient/Degree/Eccentricity/IsSingleNode/NeighborhoodConnectivity./NumberOfDirectedEdges/NumberOfUndirectedEdges/PartnerOfMultiEdgedNodePairs/Radiality/Stress/TopologicalCoefficient/EdgeBetweenness. **(3)**	Not supported. **(0)**	Not supported. **(0)**	Only graphical display of trend lines with 6 presets (in the Refine panel with application in scatterplots type diagrams). **(1)**	Many available functions such as: Single variable statistics (percentiles: 5th, 10th, 25th, 75th, 90th, 95th, Minimum, Median, Maximum, Mean, Standard deviation, Variance, Sum), Two-variable statistics: (Linear regression, Quadratic regression, Cubic regression), Correlation matrix. **(2)**
Output	Many formats and unlimited downloads, high resolution. **(3)**	One type is only supported. (.png) **(1)**	Public sharing is supported through a hyperlink to the user's account in the application and low-resolution .jpeg. **(1)**	Only publicly shared hyperlinks to the user's account in the application or integration with code on a personal site are supported.	Many formats and unlimited downloads, high resolution. **(3)**
Data collection	When running Cytoscape, the usage statistics are recorded with the date, time, and IP address of the workstation running Cytoscape. This recording can be disabled when installing Cytoscape. **(3)**	It gathers little information for statistical purposes. **(0)**	Storage is based on personally identifiable information with mandatory publication on the server site **(0)**	Storage is based on personally identifiable information with mandatory publication on the server site **(0)**	It collects personally identifiable information. **(1)**
Aesthetics	Exceptional. **(3)**	Limited features and option. **(0)**	Various style choices and chart types. **(2)**	Fairly good interface with clear and comprehensible design and functionality features. **(2)**	Minimalist style, few options, limited color palette, limited fonts.**(0)**
Save on a cloud-based platform for public sharing with or without limitations	The NDEx platform (http://ndexbio.org/#/) provides an open source framework where scientists and organizations can share, store, handle and publish knowledge of biological networks. **(3)**	Functionality is not fully supported. **(0)**	Privacy settings are not supported. An account is required by the reader in order to see a user's graph. **(1)**	No privacy settings. Storage is done via the user's account with a limit of 1000 views for 1 user in the free version. **(1)**	No time or space constraints are specified. **(2)**

Table 2: Comparative evaluation of apps: Cytoscape (3.6.1)/Datacopia/Datavisual/Datawrapper/Filtergraph.

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) "Review and Evaluation of Systems Supporting Data Journalism", in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

Evaluation Feature	6. Gephi (0.9.2)	7. Infogram	8. Plot.ly	9. Tableau public (10.5)
Ability to handle	The user has complete control during the processing of the graph **(3)**	There are restrictions on options that are removed only if paid. (storage options, logo removal, more types of maps, etc.) **(2)**	Many graph formatting capabilities are supported through the following tabs: general, traces, axes, legend, colorbars, annotation, shapes, images, themes, built in. **(3)**	The user has full control over the final output format of his/her work with a satisfactory format menu of the imaging features through tabs easily managed by the average user. **(3)**
Support material	Knowledge base is provided (wiki), forums/facebook group (with registration) & online documentation. **(3)**	Chat, search engine, tutorial, videos, FAQ are supported. **(3)**	There are detailed instructions for using the application and community support.. **(2)**	A wide range of support material is supported (tutorials, webinars, chatting, demos, etc.). **(3)**
Ease of use	The menu is easy to read and functional without demanding extensive knowledge. **(3)**	User-friendly interface quite intuitive. **(3)**	Some menu features require more specific knowledge. (analysis, grouping and aggregation functions). **(2)**	Very good GUI, friendly to the average user, with the ability to instantly track the changes made. **(3)**
Learning time	It requires time > 5 hrs for a beginner to familiarize themselves with the features and capabilities of the software. **(0)**	It requires time < 1 hr for a novice user to become familiar with the application's functionality. **(2)**	It requires moderate time. (1 hr - 5 hrs). **(1)**	It requires time > 5 hrs for a novice user to become familiar with its functionality, which offers a lot of experimentation opportunities. **(0)**
Making use of previous knowledge	Previous experience with similar graph creation software reduces learning time. **(1)**	The novice user can be helped with the previous experience of similar platforms. **(2)**	Previous experience with other software or previous versions is recommended. Quite intuitive. **(1)**	A previous experience on similar platforms works positively as it potentially accelerates the learning process **(1)**
Processing power	The processing of the graph captures memory on the user's PC, so the performance depends on the RAM and the system's graphics card characteristics. **(3)**	At the 5 MB limit the platform does not respond. **(0)**	We successfully uploaded files up to 20 MB. **(1)**	We tried to upload a 900,000 line file (30MB) (with line limitation for excel 2010 at 1,048,576 lines) without success. **(2)**
Login requirements	No login is required at any stage of network processing and storage. **(3)**	An account is required to use the platform. **(0)**	Login is required for all chart studio functions in the community plan. **(0)**	Editing and formatting do not require login while storage and publishing require a connection to the server. **(2)**
File Type Requirements (Input)	It accepts files of many and different types except .ods spreadsheets. **(2)**	It accepts xls, xlsx, csv, json, upload from google drive & dropbox.. **(2)**	It accepts csv & excel files. **(0)**	Formats such as xls, xlsx, csv, tab, txt, json, kml, shp, sav, R files and pdf are supported. **(2)**

Cost for “full” version	Free. **(3)**	$228/year for the pro version. **(0)**	$840/year for the pro version. **(0)**	Free in public edition. **(3)**
Built-in analytics	A complete set of metric statistical analysis of the network is available: Average Clustering Coefficient/Average Path Length/Betweenness Centrality/Closeness Centrality/Connected Components/Degree/Degree Power Law/Diameter/Eigenvector Centrality/Graph Density/HITS/Modularity/PageRank. **(3)**	Not supported. **(0)**	Many functions available such as curve fitting with choices of basic & advanced & moving average. **(2)**	A satisfactory set of statistical analysis is contained for all basic functions. **(3)**
Output	Many unlimited formats of output. **(3)**	Not supported. **(0)**	Supported export in several formats such as: Image: png, jpeg / Data: csv, xls, json, ppt / Code: R, MATLAB, Julia, plot.js, node.js / HTML: html, Zip archive, Embed URL. **(3)**	It is possible to output the table in csv and / or the workbook with the data (twbx format). Also visualization can be downloaded by the user in pdf, png format. **(1)**
Data collection	It maintains little or no information.. **(3)**	It gathers and maintains all information that has been uploaded, allows third-party access to information. **(0)**	It collects personally identifiable information and allows third-party access to visualizations in the community plan we evaluated. **(0)**	Storage is based on personally identifiable information while third party access is allowed in visualizations. **(0)**
Aesthetics	Many features of graph display (development algorithms) and tag and font formatting **(3)**	Various style choices and chart types. **(2)**	Many style options available, such as color, font, chart type, etc. **(3)**	Many style options available, such as color, font, chart type, etc. **(3)**
Save on a cloud-based platform for public sharing with or without limitations	There is no such possibility. **(0)**	Storage limit for testing is 10 projects. Privacy settings are also not supported. **(1)**	The community plan we evaluated offers storage space up to 25 charts without the possibility of privacy settings. **(1)**	The storage is done through the user's Tableau account with privacy setting capabilities. (on the user's account, it is possible to switch the display to others when viewing the user's profile or search). **(3)**

Table 3: Comparative evaluation of apps: Gephi (0.9.2)/Infogram/Plot. Ly/Tableau Public (10.5).

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) “Review and Evaluation of Systems Supporting Data Journalism”, in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

We list the pivot Table 4 with the scores per application.

No	Feature	Tableau	Data wrapper	Datavisual	Infogram	Plot.ly	Filtergraph	Gephi	Cytoscape	Datacopia	Category
1	Ability to handle	√√√	√√	√√	√√	√√√	√√	√√√	√√√	√	Usability
2	Support material	√√√	√√		√√√	√√	√√	√√√	√√√		Usability
3	Ease of use	√√√	√√√	√√	√√√	√√		√√√	√√	√	Usability
4	Learning time		√√	√√	√√	√	√			√√√	Usability
5	Making use of previous knowledge	√	√√√	√√√	√√	√	√	√	√	√	Usability
6	Processing power	√√				√	√√	√√√	√√√		Functionality
7	Login requirements	√√						√√√	√√√	√√√	Functionality
8	File Type Requirements (Input))	√√	√	√	√√		√√	√√	√√	√	Functionality
9	Cost for “full” version	√√√					√√√	√√√	√√√	√√√	Functionality
10	Built-in analytics	√√√	√			√√	√√	√√√	√√√		Usability
11	Output	√		√		√√√	√√√	√√√	√√√	√	Functionality
12	Data collection						√	√√√	√√√		Functionality
13	Aesthetics	√√√	√√	√√	√√	√√√		√√√	√√√		Usability
14	Save on a cloud-based platform for public sharing	√√√	√	√	√	√	√√		√√√		Usability

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) “Review and Evaluation of Systems Supporting Data Journalism”, in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

Table 4: Overall comparative rating.

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) “Review and Evaluation of Systems Supporting Data Journalism”, in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

No	Feature	Tableau public (10.5)	Tableau public [Atwood et al. 18]	Datawrapper	Datawrapper [Atwood et al. 18]	Plot.ly	Plot.ly [Atwood et al. 18]
1	Ability to handle	√√√	√√√	√√	√√	√√√	√√
2	Support material	√√√	√√√	√√	√	√√	√√
3	Ease of use	√√√		√√√	√√√	√√	√√
4	Learning time		√	√√	√√√	√	√√
5	Making use of previous knowledge	√	√√	√√√	√√√	√	√√√
6	Processing power	√√	√√			√	√
7	Login requirements	√√	√√				√√
8	File Type Requirements (Input)	√√	√√	√			
9	Cost for “full” version	√√√	√√√				
10	Built-in analytics	√√√	√√√	√		√√	√√
11	Output	√	√		√	√√√	√
12	Data collection						
13	Aesthetics	√√√	√√√	√√	√√	√√√	√√

Table 5: Differences in both application evaluations.

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) “Review and Evaluation of Systems Supporting Data Journalism”, in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

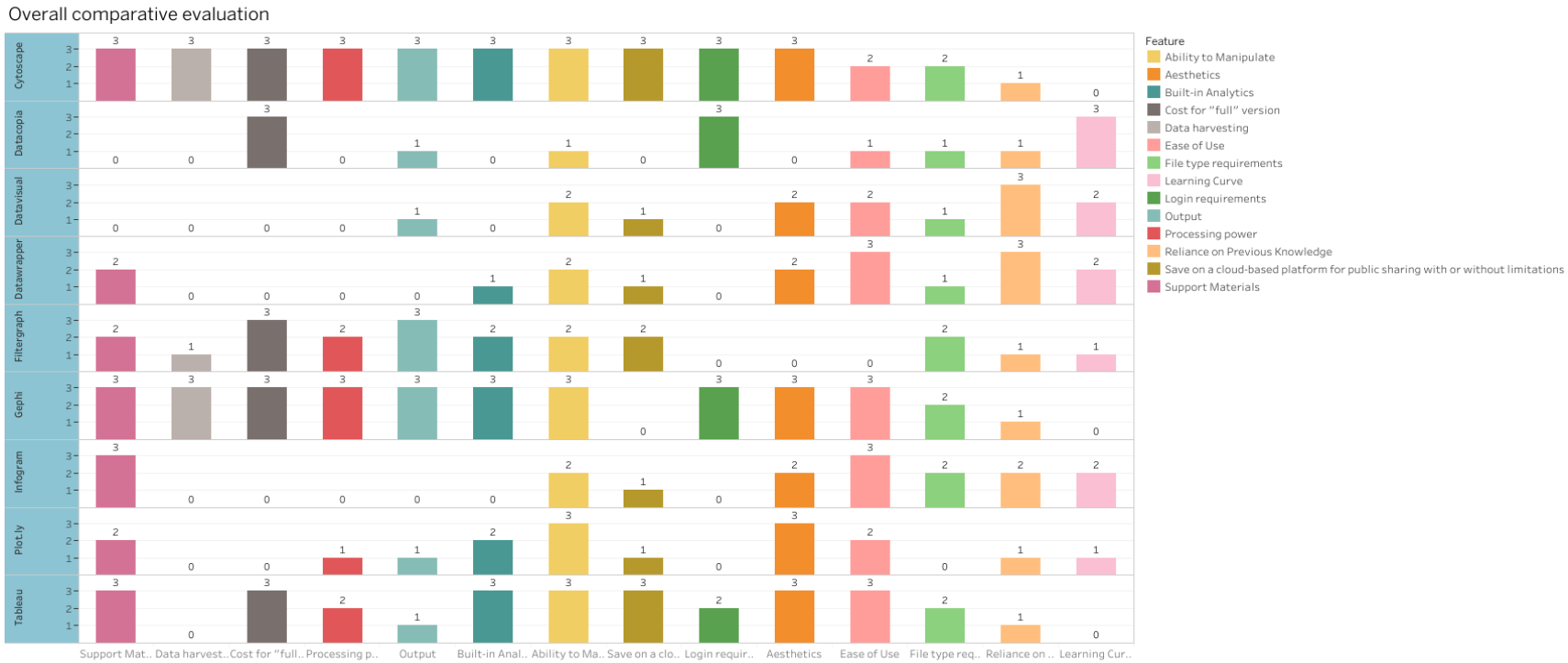

Figure 1: Overall comparative evaluation.

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) “Review and Evaluation of Systems Supporting Data Journalism”, in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

SOLUTIONS AND RECCOMENDATIONS

The presented method for evaluating data journalism systems can be used as a quick evaluation method to be used improvements should be proposed in order to achieve the ultimate goal, which is a well-established objective method for such software. A basic limitation of the method is that in its current form it is applied by a single author, which makes it a subjective opinion in some cases.
Comparing to the work of Atwood & Reznik-Zellen (2018), our study showed (in terms of common criteria) that although there are many subjective characteristics, the overall scores were relatively close together. The method could gain more from some improvement. Some points we can suggest are:

1. Stricter refinement of the criteria to minimize the strongly subjective nature of some such as:
    i. Learning time that is related to the user's cognitive and empirical level. It can be more accurately assessed if a target is set where the user-evaluator is required to achieve a visualization of specific standards based on specific data that will be given to them. Evaluation scale times may be revised because they are related to the target.
    ii. The use of prior knowledge: this criterion should also be referred to in relation to the novice user (more in relation to their empirical level of similar applications). The user-evaluator should evaluate the criterion based on whether experience in similar applications would speed up the learning process.
    iii. Ease of use with a novice user again and provided the necessary minimum time to familiarize with the application should not exceed 1-2 hrs.
2. The criterion "cost for full version " must determine whether the prices are for individual, student or professional use since the prices in many of them have a big difference.
3. Additional independent comparative evaluations in order to evaluate and improve existing criteria and to possibly supplement them with new ones.
4. It should be examined whether the above methodology can be applied with credible results in other technological fields of data journalism such as applications aimed at collecting and filtering data (e.g. OpenRefine, Nomenclature), scraping technologies (e.g. Artoo.js, Elex, Zup) publishing (e.g. Listify, Tarbell) etc.

FUTURE RESEARCH DIRECTIONS

Data journalism is an umbrella term that encompasses an ever-growing set of tools, techniques and approaches to storytelling. It can include everything from traditional computer-assisted reporting to the most cutting-edge data visualization and news applications. As such, data journalism tools and technologies should be tested and evaluated more formally and by more experts in several fields, especially their final users: journalists. Moreover, future research should be directed towards three axes:

- evaluation of other components / functions of data journalism tools (apart from the visual imaging component)
- inclusion of more systems in future evaluations (without applying exclusion criteria as those described in this chapter)
- refinement of current tools and technologies, driven by formal evaluations and journalists’ needs.

The goal of future research should be the establishment of a framework for qualitative assessment of data journalism applications, as well as widely accepted standards for evaluating data journalism tools.

CONCLUSION

The technologies that support this powerful and highly interesting stream of data journalism are in a phase of intense productivity. Technology users have many options to best support a story without using specialized knowledge. On the other hand, there is a lack of methodological evaluation criteria

from the user's point of view in order to help those interested in finding the application that best meets their needs.
We performed a comparative evaluation of nine applications covering the thematic fields of graphs and interactive charts based on a methodological approach on the basis of 14 usability and functionality criteria. The evaluation conducted showed that desktop applications show higher scores due to more and more powerful features. The other web-based applications generally exhibit lower levels of usability and functionality features as it has been described in detail.
As a general conclusion, it is a fact that literature shows a lack of studies and methodological tools for evaluating the technologies used by data journalism, which include several different types of applications. Research should therefore continue in the direction of standardizing a commonly accepted framework for evaluating the various aspects of systems supporting data journalism (data, processing stages, methods and applications).

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) "Review and Evaluation of Systems Supporting Data Journalism", in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

REFERENCES

Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization*, 2005. INFOVIS 2005, pp. 111-117. IEEE.

Atwood, T. P., & Reznik-Zellen, R. (2018). Using the Visualization Software Evaluation Rubric to explore six freely available visualization applications. *Journal of eScience Librarianship*, 7(1): e1122.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Icwsm*, 8(2009), pp. 361-362.

Buja, A., Cook, D., & Swayne, D. F. (1996). Interactive high-dimensional data visualization. *Journal of computational and graphical statistics*, 5(1), pp. 78-99.

Chuah, M. C., & Roth, S. F. (1996). On the semantics of interactive visualizations. In *Proceedings IEEE Symposium on Information Visualization'96*, pp. 29-36. IEEE.

Dix, A., & Ellis, G. (1998). Starting simple: adding value to static visualisation through simple interaction. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 124-134. ACM.

Few, S., & Edge, P. (2012). Criteria for Evaluating Visual EDA Tools. *Visual Business Intelligence Newsletter. Perceptual Edge.*

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics*, 8(1), pp. 1-8.

Shannon, P. T., Grimes, M., Kutlu, B., Bot, J. J., & Galas, D. J. (2013). RCytoscape: tools for exploratory network analysis. *BMC bioinformatics*, 14, 217, doi:10.1186/1471-2105-14-217.

Spence, R. (2007). *Information Visualization: Design for Interaction, 2nd ed.* Prentice Hall.

Tweedie, L. (1997). Characterizing interactive externalizations. *In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pp. 375-382. ACM.

Ward, M. O., & Yang, J. (2004). Interaction spaces in data and information visualization. In *VisSym*, pp. 137-145.

Wilkinson, L. (2005). *The Grammar of Graphics, 2nd ed.* New York, USA: Springer.

Yi, J. S., ah Kang, Y., Stasko, J., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics*, 13(6), pp.1224-1231.

Zhou, M. X., & Feiner, S. K. (1998). Visual task characterization for automated visual discourse synthesis. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 392-399. ACM Press/Addison-Wesley Publishing Co..

Thisis a **pre-print version** of the following article:

Georgios Vassis, Dimitrios G. Margounakis and Efthimios Tambouris (2021) "Review and Evaluation of Systems Supporting Data Journalism", in Encyclopedia of Organizational Knowledge, Administration, and Technology, (In Mehdi Khosrow-Pour D.B.A., Ed.), IGI Global, pp. 679-694.

KEY TERMS AND DEFINITIONS

Data journalism: Journalism based on structured data in order to support a story that makes sense to the public.

Functionality: A system evaluation criterion aimed at identifying certain functional characteristics.

Interactive Chart: A chart on which a user can achieve a certain degree of interaction in order to reshape it, e.g. hover a point, zoom, play with axis, add a button etc. It allows the reader to go deeper in its understanding of the data, since he can customize it and try to answer his own question.

Rubric: A set of evaluation criteria that are valued based on a rating scale.

Storytelling: The social and cultural activity of sharing stories, with great application to journalism.

Usability: A criterion for the evaluation of the system aimed at identifying usability features from a user's point of view.

Visualization: The visual representation of journalistic story that aims to convey the information to the reader in the best possible way.

ADDITIONAL READINGS

Aitamurto, T. (2011). The impact of crowdfunding on journalism: Case study of Spot. Us, a platform for community-funded reporting. Journalism practice, 5(4), 429-445.

Bradshaw, P. (2014). Data journalism. Ethics for digital journalists: Emerging best practices, 202-219.

Chen, C. (2010). Information visualization. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 387-403.

Gray, J., Bounegru, L., & Chambers, L. (2012). Data Journalism Handbook. European Journalism Centre and the Open Knowledge Foundation.

Knight, M. (2015). Data journalism in the UK: a preliminary analysis of form and content. Journal of Media Practice, 16(1), 55-72.

Lorenz, M. (2010). Data driven journalism: What is there to learn. In IJ-7 Innovation Journalism Conference (pp. 7-9).

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. IEEE transactions on visualization and computer graphics, 16(6), 1139-1148.

Uskali, T. I., & Kuutti, H. (2015). Models and streams of data journalism. The Journal of Media Innovations, 2(1), 77-88.