2015

Data Visualization as a Communication Tool

Susan [Gardner] Archambault, *Loyola Marymount University*
Joanne Helouvry
Bonnie Strohl, *University of Scranton*
Ginger Williams

Data Visualization as a Communication Tool

Introduction

The systematic monitoring and evaluation of data enables libraries to respond to the changing needs and demands of their users, and to demonstrate their support of departmental and institutional accreditation standards. There are advantages to presenting data visually rather than as a set of flat statistics. Proper data visualization facilitates the recognition of patterns and relationships in order to communicate a message in a more compelling and interesting way. It allows the complexity of the data to be understood more easily. The purpose of this article is to provide a framework for thinking about meaningful data visualization in ways that can be applied to routine data collected by libraries. It will also offer practical “best practice” tips and a list of free, easy to use online tools to help visualize data in an effective way in order to communicate your message.

Visual Processing

Our brains process visual information and images in various ways. Some processing is performed immediately and almost automatically, rather than consciously. These processing tasks, called “pre-attentive attributes,” stand out and are immediately detected. They include differences in color hue and color intensity, and certain form attributes like shape, size, line length, line width, enclosure, and orientation (Few, 2012, p.67-71) [place Figure 1 here]. Our brain immediately detects these unique visual properties and they “pop out” of a display. Data visualization takes advantage of this and allows people to process data faster using these innate visual processing capabilities.

In order for your audience to process data visualizations efficiently and effectively, it is best to include only the information and graphical elements that are absolutely necessary. The majority of the visualization should be essential to displaying data. Extraneous markings, such as grid lines and excessive labeling, should be avoided as much as possible while still expressing data accurately. Tufte (2001, p.107) calls such non-essential elements “chartjunk.” Taking time to edit visualizations for simplicity will lead to a graphical representation that is accurate and comprehensible. Also, color should be used strategically. The three best color schemes to use are 1. Analogous (colors are near each other on the color wheel to create harmonious designs); 2. Complimentary (colors are opposite each other on the color wheel and the high contrast stands out); and 3. Monochromatic (when colors are all across a single hue to achieve visual cohesion) (Clayton, 2014). Angela Zoss’ (2014) “Introduction to Data Visualization” Libguide offers even more tips for maximizing the effectiveness of data visualizations.

Quantitative Relationships

Good data visualizations help us understand the kind of quantitative relationships presented by data. These relationships are an association between quantitative values or multiple sets of values and categories. Different methods of data display are better at expressing particular quantitative relationships. The major types of relationships among quantities include:

- Nominal comparison (differences in particular values)
- Time Series (how values change over time)
- Ranking (the order of values)

- Part to whole (%)
- Deviation (differences from some standard value)
- Distribution (how a set of values are distributed over a range)
- Correlation (whether two different values change together) (Lown, 2010)

There are two types of charts used to create visualization data. The first, a **"table**," is used to display or compare individual values. A table should be used when precision is important, a huge amount of data has to be included, or when you need to show multiple units of measure. A table takes up only a small amount of space, but it has no visual impact (Wong, 2010, p.82; Lown, 2012). Use thin rules or lines after every three to five entries to help the reader follow the numbers across a table, and use shading to highlight a column of data or an entry. For a table with multiple data series, do not present the comparative data horizontally since it is easier to analyze data vertically (Wong, 2010, p.83-84).

The second type of chart is a "**graph."** In a graph, the meaning of the data is revealed by the shape of the values. Graphs can show relationships among many values, and they are good when there is a large amount of data (Lown, 2012). Here are some of the most common types of graphs.

Common Types of Graphs

Scatter Plot: A graph in which the values of two variables are plotted along two axes, the pattern of the resulting points revealing a correlation, relationship, or distribution. Each marker represents an observation, and the positioning represents the value. Variables are correlated if they have a dependency on or are influenced by each other. A strong relationship will be close to a straight line, rather than bouncing around. A

positive relationship moves upward from left to right, and a negative relationship moves downward from left to right (Few, 2012, p.31-34). [place Figure 2 here]. A *bubble plot* is a variation of the scatter plot in which markers are replaced with bubbles.

Line Chart: A chart that shows the position of data points relative to each other as they fluctuate over time to reveal trend, acceleration or deceleration, and volatility (Dando, 2014, p.80). Line charts do not have to start at "0." [place Figure 3 here] According to Wong (2010, p.51-57), there are many things you can do to make line charts easy to read, including using four lines or less, using increments that are natural and easy to count, not labeling the lines with more than one short sentence, and trying to label directly next to or on the line itself. Other things she suggests are using bright or dark colors such as red or black to emphasize the important line (Wong, 2010, p.42). For more than four data series, try using small multiples, a series of small charts that use the same graphic to display different slices of a data set for better comparison.

Bar chart: A chart in which the numerical values of variables are represented by the height or length of lines or rectangles of equal width to compare numerical data by size or importance (Dando, 2014, p.78). Bar charts can also be used to illustrate ranking**,** part to whole (percentages), deviation, or distribution between categories or groups (Lown, 2010). A bar chart needs to start at "0" on the Y axis so the data doesn't get distorted (Wong, 2010, p.41). There are several variations on the standard *vertical bar chart.* A *horizontal bar chart* is good for longer labels, but should be avoided if most of the values are negative (Wong, 2010, p.72). [place Figure 4 here] Another variation is the *stacked bar chart*, which consists of columns (bars) with subgroups, or stacks, comparing the parts to the whole. The bars in a stacked bar graph are divided into

categories, where components show relative contribution to the entire bar. A final variation is the *multiple bar chart*, good for four or fewer categories when the overall totals are not critical (Wong, 2010, p.67).

According to Wong (2010, p.41-67), *vertical bar charts* can be made easier to read by making the width of the bars twice the width of the space between the bars. Other tips include using different colors to represent different categories (while avoiding red for positive numbers since this represents a "loss" in the business world), shading the bars from lightest to darkest for easy comparison, and sorting data so it is not plotted in a random order.

Histogram: A *histogram* is a variation on the bar chart in which the data is "continuous" and adjacent. A histogram shows how frequently or infrequently a given value occurs, and it shows the data set's distribution. It is used to show how often values fall within certain ranges. [place Figure 5 here]

Pie Chart- A chart divided into sectors totaling 100% to illustrate the relationship of parts to a whole (Dando, 2014, p.80). [place Figure 6 here] Pie charts should be used sparingly because it is difficult to distinguish small differences in area that are similar in size but not located next to each other (Lown, 2012). Using high-contrast colors can help differentiate between segments (Dando, 2014, p.81), and using a darker shade or different color can highlight an important segment (Wong, 2010, p.42). Other disadvantages of pie charts are that they take up a lot of space, and they can be difficult to label. Always put the percentage in labels on or near the chart. According to Wong (2010, p.74-75), you should put the largest slice in the 12 o'clock position, and then order by size, with "other" coming last. Pie charts should have no more than five slices;

smaller slices can be combined into an “other” category. Alternatives for pie charts with too many slices include a *bar chart* or a *bubble chart.*

Infographic: A combination of textual and quantitative data used to highlight key images in order to tell a story or narrative. Infographics should reveal information that is otherwise lost in a crowd of data, and the message should be significant or surprising (Dando, 2014, p.84). [place Figure 7 here] Pictograms can be used as symbols for your data, but keep them simple, symmetrical, clear and crisp, and small enough to fit in a square. Avoid using partial icons (Wong, 2010, p.87-88).

Tips: Do’s and Don’ts for Data Visualization

DO’S:

- ✓ Pass the squint test: Is the information conveyed without reading any text? Does the right color pop out? Are the labels clear? Try printing out your visualization in black and white or make a copy in gray scale to check for visibility (Clayton, 2014).
- ✓ Show restraint and limit the data to your main point so you don’t overwhelm the audience (Dando, 2014, p.74).
- ✓ Organize: group, prioritize, and sequence data to help viewers understand.
- ✓ Provide a key to your visualization if necessary for the viewer to understand your data (Dando, 2014, p.81).
- ✓ Have a colleague preview your data visualization for clarity (Dando, 2014, p.80-81).
- ✓ Round to the nearest significant digit for clarity in labels, but use decimal places for accuracy in calculating and plotting graphs (Wong, 2010, p.22).

- ✓ Frame your data in a context that your audience can relate to and offer relevant reference points (Dando, 2014, p.77).
- ✓ Use colors sparingly and to help convey meaning rather than for decoration (Dando, 2014, p.81). Similarly, use as few font styles as possible and use those not highly stylized (Wong, 2010, p.143). Select a color palette that includes 3-5 shades of each hue (Wong, 2010, p.38).

DON'TS

- × Don't manipulate data to tell a story it doesn't actually tell. Tufte (2001, p.55-77) calls this "graphical integrity."
- × Don't use 3D or a "blow apart" effect- this reduces comprehension and makes it hard to compare elements (Few, 2012, p.197).
- × Visualize all of the important relationships and make large data sets coherent (Tufte, 2001, p.13).
- × Don't put a box around your graph. This is unnecessary ink that will visually distract the viewer (Tufte, 2001, p.127).
- × Don't use red/green or blue/yellow combinations because the lack of contrast in lightness makes it unreadable for the color-blind (Wong, 2010, p.44).

Data Visualization Quiz: Test Yourself!

Match the situation with the best data visualization option. Check your answers in the Appendix.

1. ___You surveyed undergraduate students, graduate students, faculty, and staff at your institution and you want to show the percentages of each category that participated in your survey.
2. ___You have circulation statistics from the four branches in your library system, and you would like to show a comparison between them over a ten-year period.
3. ___You interviewed faculty for a liaison program from seven different areas at your school: liberal arts, business administration, communication, law, education, film and television, and science. You want to visualize a comparison of how many faculty were interviewed in each of these areas, from large to small.
4. ___You have all of the information literacy quiz scores for each student. You want to show the distribution of how many students received each grade (e.g. A, B, C, D, or F).
5. ___You need to display statistics showing the exact cost spent on all twelve academic departments for interlibrary loan requests.
6. ___You need to plot the number of book checkouts for students against their GPA to see if there is a correlation.
7. ___You need to tell the story of school librarians' involvement as leaders in their learning community within one school district drawing from a variety of statistics including professional development, planning, work time, and networking.

Answer Choices:

A. Scatter plot

B. Line Chart

C. Bar Chart

D. Pie Chart

E. Infographic

F. Histogram

G. Table

Free Online Data Visualization Tools

You can use Excel to create all of the graphs mentioned in this article, but there are also free online visualization tools available. Here are some of our top picks. [place Table 1 here]

Conclusion

Evidence-based libraries collect and use data in order to affect change and support departmental and institutional accreditation standards. Proper data visualization allows libraries to communicate their message in a more compelling and interesting way, while assisting in the understanding of complex data. The general rules and strategies for visualizing quantitative relationships presented in this article, when followed, will help data to be better understood by its audience. The free online data visualization tools can help librarians experiment with enhancing their data displays so the underlying message is more transparent, ultimately leading to proactive changes and improvements in the library.

References

Association of Research Libraries. (2014), "Statistical trends," available at: http://www.arl.org/focus-areas/statistics-assessment/statistical-trends#.U5H5UfldV8F (accessed 27 October 2014).

Clayton, T. (2014), "Dress your data for success: data visualization strategies for library assessment [ACRL webcast]," available at: http://www.ala.org/acrl/datavisualization (accessed 29 April 2014).

Cox, B. L. and Jantti, M. H. (2012). “Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement”, *Library & Information Science Research,* Vol. 34 No.4, pp. 308-316.

Dando, P. (2014), *Say It With Data: A Concise Guide to Making Your Case and Getting Results*, ALA Editions, Chicago, IL.

Few, S. (2012), *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (2nd ed.), Analytics Press, Burlingame, CA.

Maidenberg, K. and Kawula, B. (2012), “Ask a librarian - a year in review [Infographic]”, available at: http://vr.scholarsportal.info/ask/statistics (accessed 27 October 2014).

Lown, C. (2010), “Practical considerations for displaying quantitative data [presentation slides]”, available at: http://www.slideshare.net/corylown/data-visualization-7522083 (accessed 27 October 2014).

Lown, C. (2012), “Let the data talk”, presented at the American Library Association Annual Conference (ALA), 24 June, Anaheim, CA, available at: http://www.slideshare.net/corylown/let-the-data-talk-ala-2012 (accessed 27 October 2014).

Tufte, E. R. (2001), *The Visual Display of Quantitative Information* (2nd ed.), Graphics Press, Cheshire, CT.

Wong, D. M. (2010), *The Wall Street Journal Guide to Information Graphics: The Dos and Don’t’s of Presenting Data, Facts, and Figures* (1st ed), W. W. Norton & Co.,New York.

Zoss, A. (2014), “Introduction to data visualization”, available at: http://guides.library.duke.edu/datavis (accessed 27 October 2014).

Appendix: Data Visualization Quiz Answers

1. D. Pie Chart
2. B. Line Chart
3. C. Bar Chart
4. F. Histogram
5. G. Table

6. A. Scatter Plot

7. E. Infographic

Figure 1: Pre-Attentive Attributes

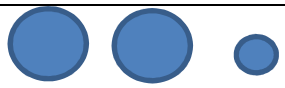
Attribute	Visual
Color	
Size	
Shape	
Line Length	
Line Width	
Enclosure	
Orientation	

Figure 2: Scatter Plot (Cox & Jantti, 2012, 312)

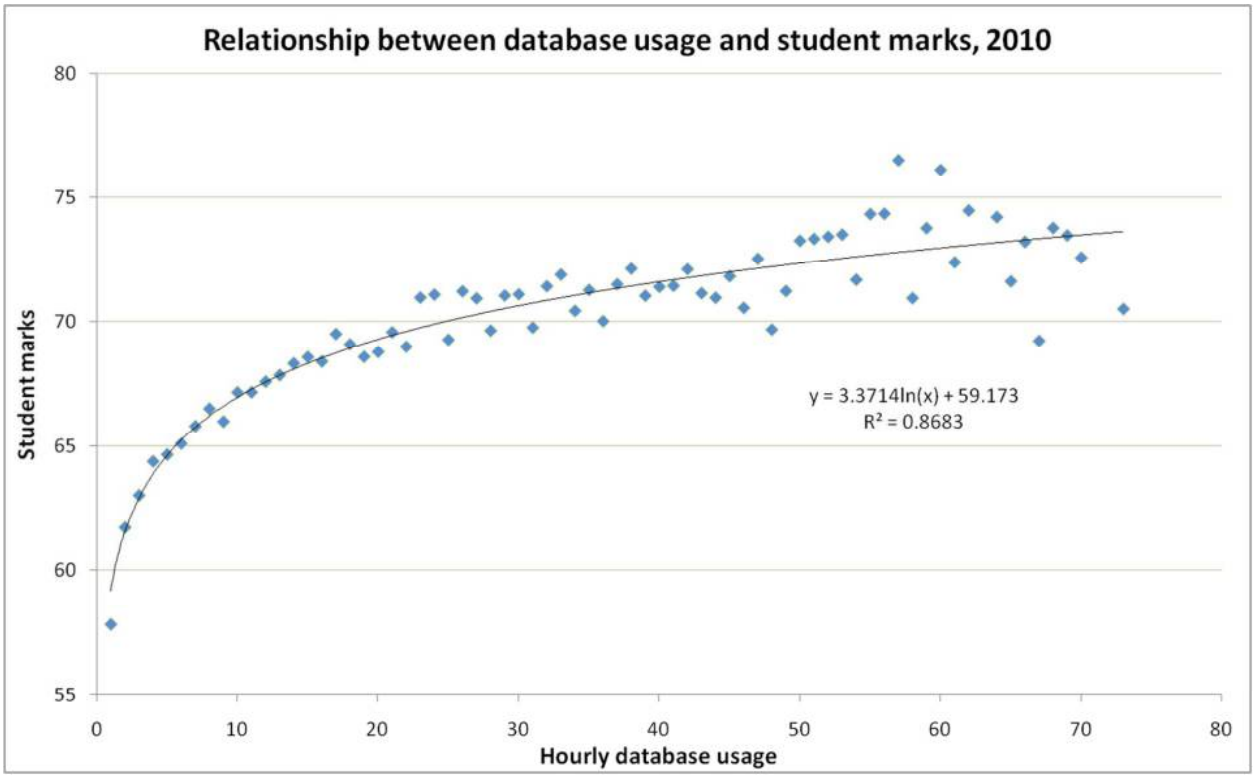


Figure 3: Line Chart (ACRL, 2014)

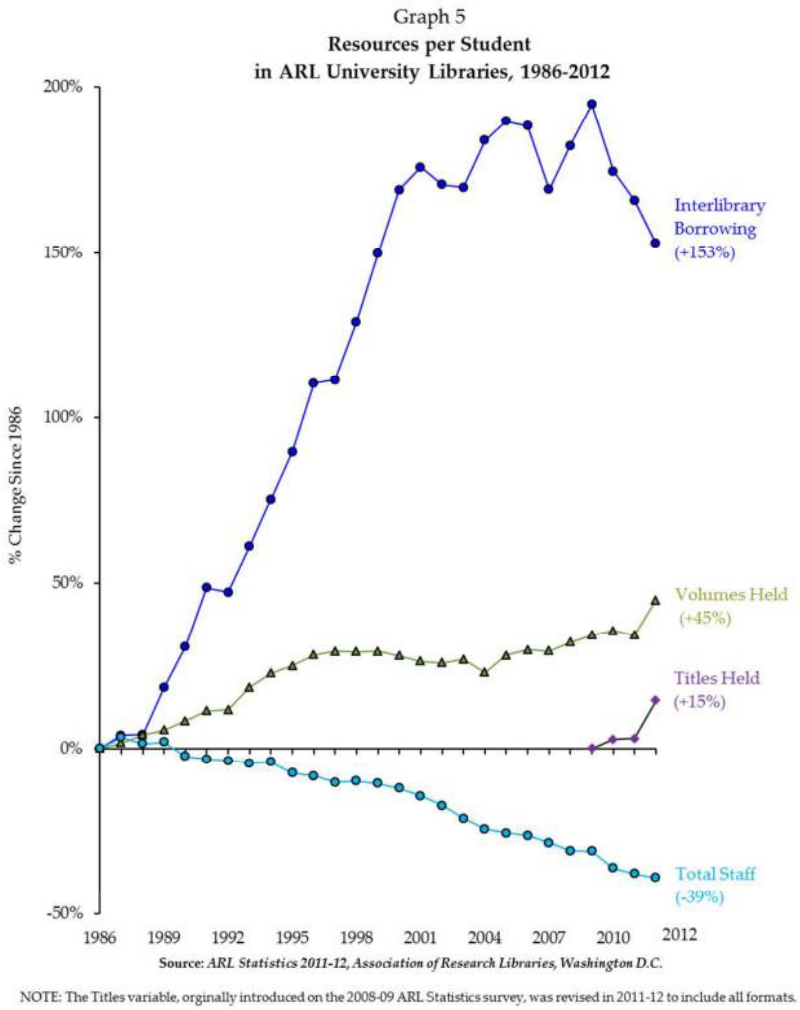


Figure 4: Bar Chart Example of a horizontal bar chart

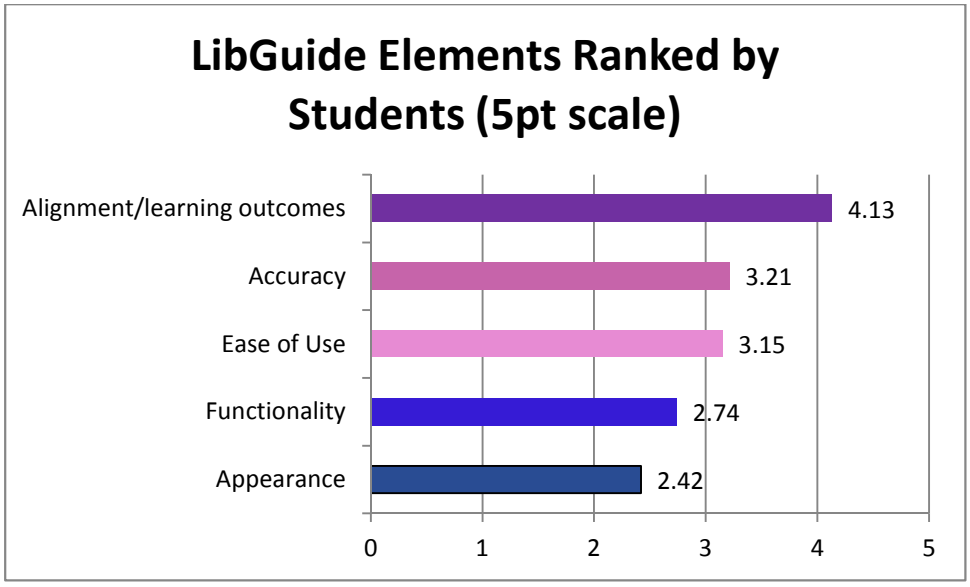


Figure 5: Histogram

Figure 6: Pie Chart

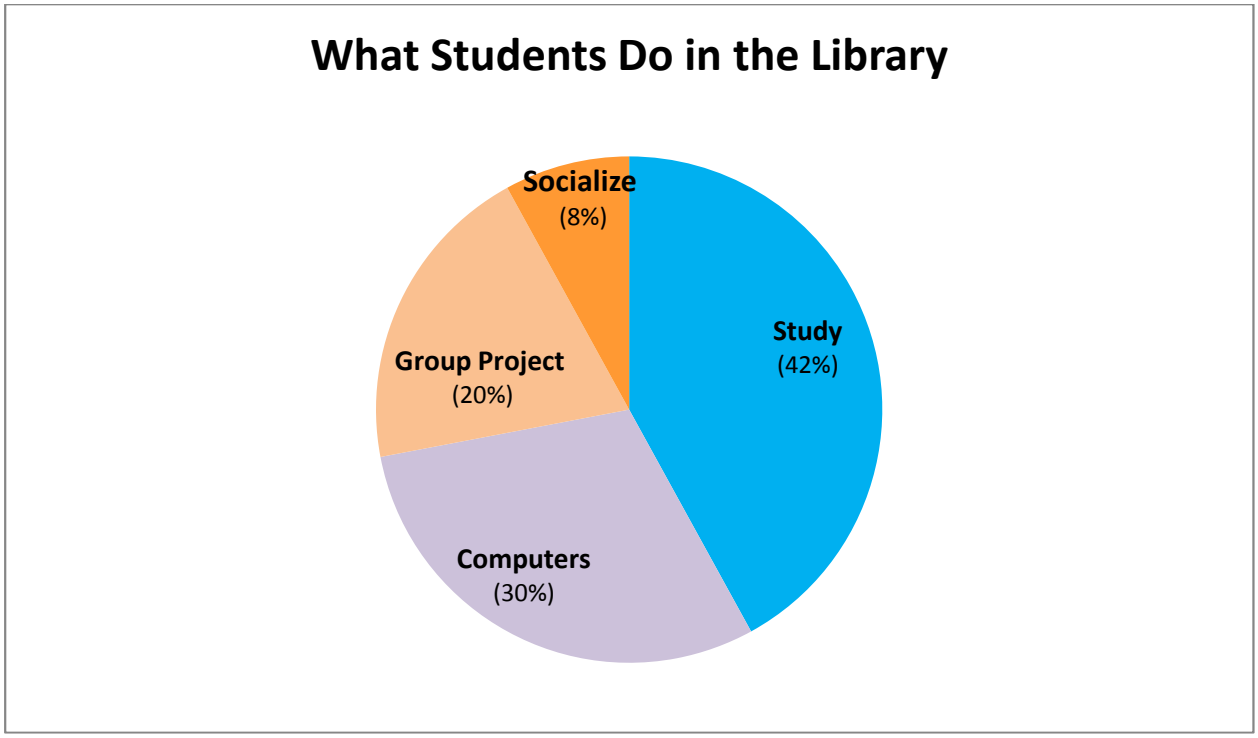


Figure 7: Infographic (Maidenberg & Kawula, 2012)

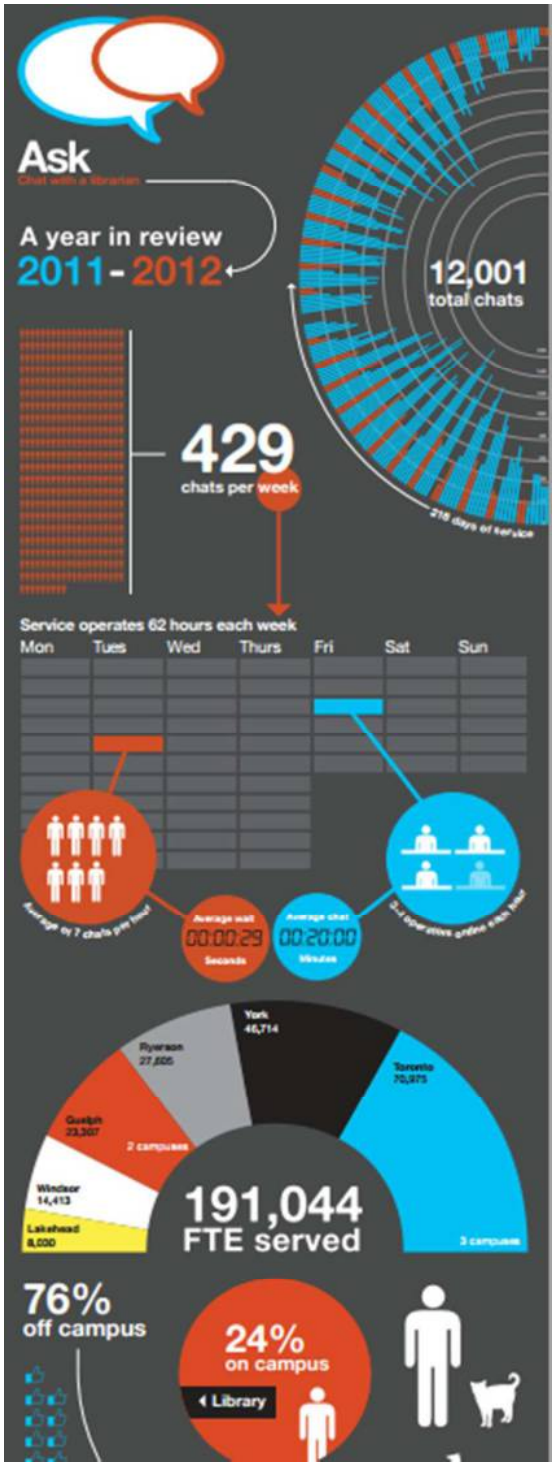


Table 1: Comparison of Free Online Data Visualization Tools

L	***DATA IMPORT OPTIONS***	***CHART SUPPORT*** ***S****CATTER****P****LOT*** ***L****INE* ***B****AR* ***H****ISTOGRAM* ***P****IE* ***I****NFOGRAPHIC*	***EXPORT/ SHARE OPTIONS***	***ABILITY TO CUSTOMIZE***	***SIZE LIMITS***	***TECH LIMITATION***	***PRIVATE OR PUBLIC***	***HELP/ SUPPORT***	***PRO OR PAID VERSION***
fogr.am p://infogr.a	Upload data in XLS, XLSX, or CSV formats; upload images	***S, L, B, P, I***	Publish to social media, view on the web, or embed into web page. Download in PNG format	6 infographic templates; can put in your own data	None	Does not support IE 8 or older	Public, except unpublished infographics are private	Chat support, email	Price varies; cheapest Pro = $18/mo. Includes private sharing, 4 new themes, save infographic as PDF, teams feature
ısel.ly p://www.eas ly/	Upload your own charts and graphs, but must be saved as an image (PNG, JPG, PDG, and SVG). Text can be added	***I*** S,L,B,H, P will be supported starting 9/19/14	Saved as "my visuals" to cloud server. Export in a PDF or high-res image, share in browser, on social media, or embed into web page	Choose or modify one of 30,000 public templates or create your own infographic from scratch (choose from 15 themes)	None	None	Private unless you specify that it is public	Blog, online form	Paid version coming soon. Expanded library of images

L	***DATA IMPORT OPTIONS***	***CHART SUPPORT*** ***S****CATTER****P****LOT*** ***L****INE* ***B****AR* ***H****ISTOGRAM* ***P****IE* ***I****NFOGRAPHIC*	***EXPORT/ SHARE OPTIONS***	***ABILITY TO CUSTOMIZE***	***SIZE LIMITS***	***TECH LIMITATION***	***PRIVATE OR PUBLIC***	***HELP/ SUPPORT***	***PRO OR PAID VERSION***
eau iblic nline or esktop) ww.tableaus ware.com/	Excel, Access, or Text files. Can also connect to an Open Data Protocol (OData) data source	***S, L,B,H,P or I*** (Also simple tables and variety of other charts)	Tableau dash-boards may be embedded in blogs or in web page; shared directly (using a unique URL) via social media or email	Color, size text and details can be changed; "Show Me" feature presents a choice of charts based on the data; change or add data easily with drag & drop	1 GB overall, and individual workbooks have a 1M record limitation	Installed on Mac OS 10.8.1 or later and Windows. Works on these browsers: Android 3.2+, Safari 3.x+, IE 7+, Firefox 3.x+, Chrome, iPad/Android iPad app	Public/ shared and is available to cmty of users	Online training videos, active cmty of users' online forum, knowledge base of articles, free weekly webinars	Premium annual subscription/ price based on views per month (25K/yr or 6K/yr) Disable download of workbooks, storage space is 10GB, and row limitation is 10M rows
any Eyes ww.ibm.com/ **anyeyes**	Data must be formatted as a table (use a spreadsheet program like Excel or a text file), then pasted into a text box	***S, L, B, H, P*** (Also other charts including matrix, network diagram, maps)	Export link to email, share on social media, embed either a static image or live version into web page, or save as PNG	Choose visualization type; customization options depend on chart type, highly interactive; can modify color, fonts, layouts	5 MB per data set	Users must have Java and Flash to view	Public. You can also reuse data sets others have uploaded	Email, blog, wiki, FAQs	N/A

L	***DATA IMPORT OPTIONS***	***CHART SUPPORT*** ***S****CATTER****P****LOT*** ***L****INE* ***B****AR* ***H****ISTOGRAM* ***P****IE* ***I****NFOGRAPHIC*	***EXPORT/ SHARE OPTIONS***	***ABILITY TO CUSTOMIZE***	***SIZE LIMITS***	***TECH LIMITATION***	***PRIVATE OR PUBLIC***	***HELP/ SUPPORT***	***PRO OR PAID VERSION***
gle **arts** ps://develop s.google.co 'chart/	Any Google spreadsheet or CSV file	***S, L, B, H, P***	Embed into website or Google spreadsheet; generate static image	Options change depending on chart type. Modify colors, sorting, text, data. Start from a template and customize	No limit	Users must be able to understand HTML and paste their data into the chart templates provided, or use Google spreadsheets to create charts automatically	Public	FAQs, Community forum	N/A
oogle **ision** **ables** p://tables.go lelabs.com/	CSV, KML & TSV formats, spreadsheets (XLS, .XSLX, ODS, and Google). Developers can use Fusion Tables API to insert/delete data. Upload into Google Drive	***S, L, B, P*** (Also supports timelines, geomaps & network graphs)	CSV, KML, embed in a web page, share link, share link on social media, allow viewer to click "File > Make a copy"	Combine data from public sets w/ your own data; cite source of data/ attribution/ add a reuse license; control download option; change colors & fonts, filter/ aggregate data	250 MB per user; 100 MB per data set	Can't copy/paste data in	Private or public in Google Drive. Drive allows sharing with specific people for editing or viewing	Help center (FAQs), feedback form, email, add a feature form	N/A

L	***DATA IMPORT OPTIONS***	***CHART SUPPORT*** ***S****CATTER****P****LOT*** ***L****INE* ***B****AR* ***H****ISTOGRAM* ***P****IE* ***I****NFOGRAPHIC*	***EXPORT/ SHARE OPTIONS***	***ABILITY TO CUSTOMIZE***	***SIZE LIMITS***	***TECH LIMITATION***	***PRIVATE OR PUBLIC***	***HELP/ SUPPORT***	***PRO OR PAID VERSION***
wrapper ps://datawra er.de/	Cut and paste, upload CSV file. Data needs to be cleaned.	***L, B, P*** S coming in October (Option to create SVG-maps)	Embed HTML code via iframe, link/URL, download the data, export to static image	Allows first rows to be headers with a checkbox. Add a "pre-append" to the data (e.g. $ or #), sort data, customize color, highlight data, list source, select European or US-based number	None-Can download free soft-ware to server	Data has to be clean, no commas, dots, currency symbols, not a lot of formatting. Optimized for HTML5- won't display on older browsers	Public, (but charts can be published behind firewall)	Quick Start guide and tutorial, fee based training	Several paid options coming soon. "Single" 30 days/ 12 Euros Includes hosting of charts, no watermark, extended chart and map options Datawrapper "Team" (10 users) 1.200 Euros/year
harts ww.icharts.n	Upload data directly from Excel, Google spreadsheet, API, or manually	***L, B, P*** Create your own I	Publish to social media, share URL, embed code to web page or blog, down-load as PNG image	14 templates; can add data or comments into chart; choose color and thickness; sort order	No limit on storage	Works on all major browsers; flash required for interactivity on chart	Public only	FAQs, get started video, contact form, email, blog	Platinum $75 per user per month to upload images, private charts, chart books Enterprise-starts at 15K per year

L	***DATA IMPORT OPTIONS***	***CHART SUPPORT*** ***S****CATTER****P****LOT*** ***L****INE* ***B****AR* ***H****ISTOGRAM* ***P****IE* ***I****NFOGRAPHIC*	***EXPORT/ SHARE OPTIONS***	***ABILITY TO CUSTOMIZE***	***SIZE LIMITS***	***TECH LIMITATION***	***PRIVATE OR PUBLIC***	***HELP/ SUPPORT***	***PRO OR PAID VERSION***
ochart ps://magic.p ochart.com/	XLS, XLSX, or CSV files or use Google spreadsheets. Also add data manually	***I*** S, L, B, P can be inserted into the Infographic	Link, embed, email, or share on social media. Download as JPEG or PNG file.	4 infographic formats and 5 themes; change colors; upload 20 images. Watermark branding. Can make chart interactive, choose resolution	No limit other than on image size:2mb or less per file	Works best on Chrome or Firefox	Public, but can "unpublish"	Online form, email, FAQs, Skype, live chat, phone	Pro - $29/month w/ no contract Or $290/year 125+ themes, no watermark, upload 200 personal images, download high- res, privacy control
nngage ps://vennga .com	No import option- data entry is manual or copy/paste from spreadsheet	***S, L, B, P, I***	Link with URL, embedded in an iFrame, or shared on social media	A few templates available. Can add images, backgrounds, shapes, pictogram, font style, and color.	Maximum of 5 infographics	Chrome and Firefox supported. IE1+, Opera, and Safari should work. Internet Explorer 10 and under not supported	Private by default, must be changed to public to share.	Blog, email	$19/Month or $49/3months and $190/12month Export to PDF/PNG, and access premium content which includes more Templates, Icons, and Charts.