6 Information visualization

6a Basics of visualization

Information visualizations are a part of everyday communications and scholarship. These graphics have powerful rhetorical force. The visualizations are often more easily consumed than the complex research data on which they depend. Understanding the process by which visualizations are made helps bring into focus what they show and what they conceal.

All information visualizations are metrics expressed as graphics. The implications of this simple statement are far ranging. Data can be very difficult to interpret in tabular form. Very few individuals are skilled at reading spread sheets, let alone relational databases, to make sense of information. A query might produce thousands of data points. Information visualizations are used to make this quantitative data legible. They are particularly useful for seeing patterns in large amounts of information, making these apparent in a condensed form.

Anything that can be quantified (given a numerical value) can be turned into a graph, chart, diagram, or other visualization.

Points, lines, and areas can be plotted using analog tools—paper and colored pencils—and many of the formats used in digitally produced visualizations are centuries old. The process of making graphs by hand is slow and deliberate. Each point has to be marked, each line created by connecting dots or using mathematical formulae, and each area calculated. At each step of hand-drawing a graph or chart, we reflect on how it is made.

But the ease of production afforded by computational means makes it possible to create polished and sophisticated graphics without critical reflection. We can easily overlook the fact that all parts of the process—from creating quantified information to producing visualizations—are acts of interpretation. In addition, the ability to *read* a visualization requires understanding the *semantics* of graphic formats. Visual forms create meaning, they don't just display it. A bar chart makes a different statement than a pie chart, for instance, and such insights are crucial to the critical engagement with information visualization (Lengler and Eppler 2007).

Benefits and liabilities

To begin, consider the two components of a visualization separately—the metrics and the graphics. Here are two versions of the same information, a table and a bar chart:

	A	B	C
1	1969	4	
2	1970	11	
3	1970	24	
4	1971	8	
5	1971	54	
6	1971	25	
7	1972	20	
8	1972	20	
9	1972	20	
10	1972	54	
11	1972	52	
12	1972	48	
13	1972	1	
14	1973	30	
15	1973	5	
16	1973	15	
17	1973	16	
18	1973	6	
19	1973	6	
20	1973	21	
21	1973	37	
22	1973	1	
23	1973	30	

Figure 6.1a Segment of a table and *6.1b* Bar chart generated from the same information (JD)

The table is not very complicated, it puts dates in one column and number of pages output by an author into a second one. All of the information in it makes good sense but trying to read columns of numbers to see a pattern in them is difficult. The chart makes clear that a steady output of pages occurred in 1972, matched by one spike in 1971, and followed by

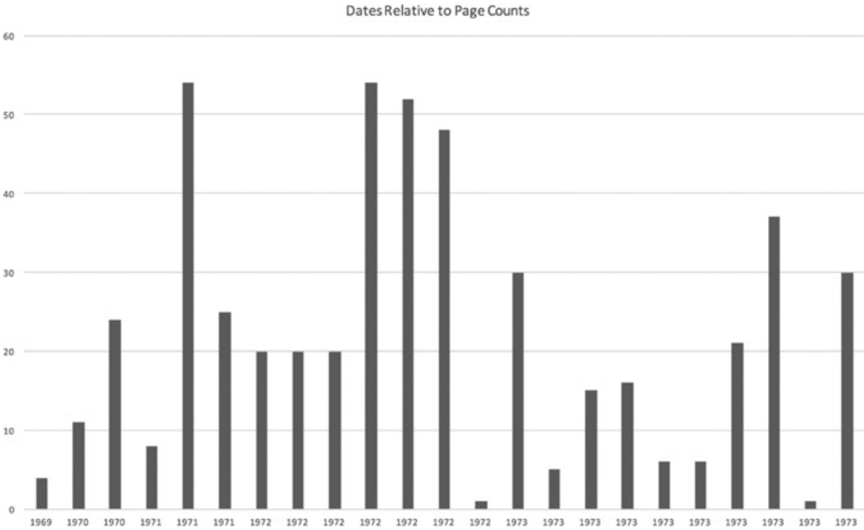


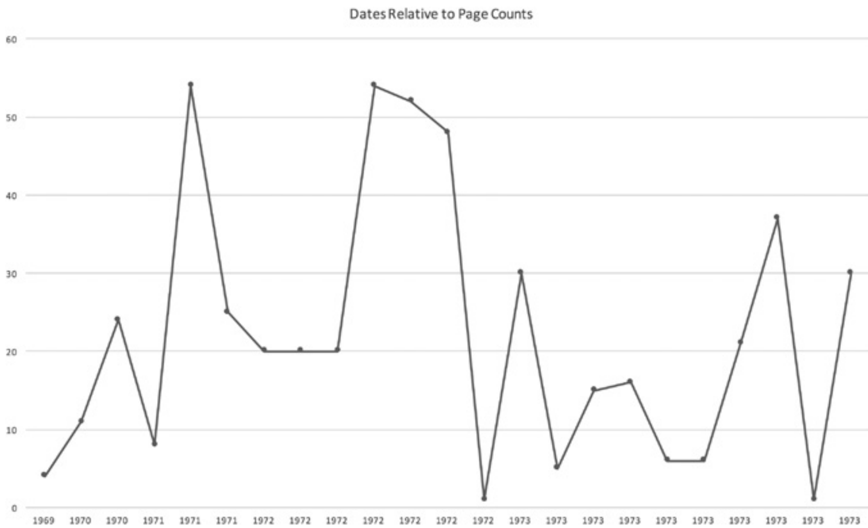
Figure 6.1a Continued

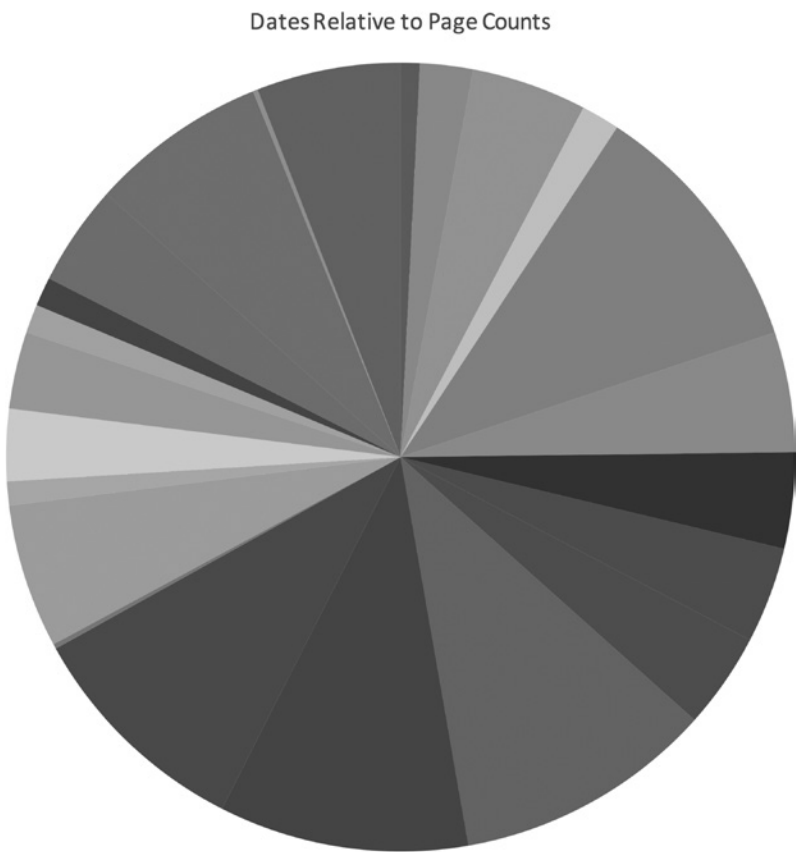
low output in 1973. The comparison of values is easily done in the visual format, and if we imagine extending the table to include hundreds or thousands of data points, this fact would be even more dramatically clear.

What is the relationship of the data to the visualization? In this situation, a line of dates is charted on the x-axis and a set of values is indicated by the y-axis. The conventions of charts make this easy to read and even intuitive in layout. But is there an inherent visual form in the data? One interesting exercise is to put the same data into other graphical formats to see what happens. Here are two examples of the same data but in a line chart and a pie chart.

We are immediately confronted with the question of what features of the graphical display are meaningful. For instance, the *continuous* line on the left graphing the dates suggests that the *rate of change* in the data about pages is a significant factor. But the "number of pages" data is actually a *discrete* value. While the bar chart *compares the values* of each segment to each other, the line chart makes these part of a continuous process, though this is not the case. By contrast, the pie chart suggests that each entry is *part of a whole*—that the sum total of pages is significant, not the difference in their value. The values are hard to compare, the dates are lost entirely, and the concept of the "whole" of the author's output has no meaning. Neither of these charts makes the correlation of date and page output as clear as the initial bar chart. These are both "bad" graphics (and possibly bad data as well).

The point is that nothing in the data dictates the form of the visualization. These and a host of other charts can be generated from the same data.





Figures 6.2 and 6.3 Other visualizations of the same data in Figure 6.1 (JD)

Any data set can be put into a pie chart, a continuous graph, a scatter plot, a tree map, and so on. The challenge is to understand how the information visualization creates an argument and then make use of the graphical format whose features serve your purpose. Any sense that data have an *inherent* visual form is an illusion. [See Exercise #1: A range of graphs.]

Data creation, as we noted in earlier (see Sections 2a and 2b), depends on parameterization. As stated before, this means that anything that can be measured, counted, or given a metric or numerical value can be turned into data. The concept of parameterization is crucial to visualization because the ways in which we assign value to the data will have a direct impact on the ways it can be displayed. Visualizations are convincing by virtue of their graphic qualities and can easily distort the data. While all visualizations are interpretations, some are more suited to the structure of a given data set than others.

Visualization basics

In many cases, the graphic image is an artifact of the way the decisions about the design were made, not about the data. Understanding some basics of the relation between graphics and metrics is essential.

Here are some fundamental guidelines for thinking about which chart to use:

- The distinction between discrete and continuous data is one of the most significant decisions in choosing a design. Example: in visualizing the height of students in a class, making a continuous graph that connects the dots makes no sense at all. There is no continuity between the height of one student and another. Individual height is a discrete value.
- If you are showing change over time or any other variable, then a continuous graph is the right choice. Example: Change in height for individual students over a five-year period.
- If a graph shows quantities with area, use it for percentages of a whole, like a pie chart, not comparative value. If you increase the area of a circle by length of the radius, or a square from the length of the side, you are introducing distortion into the relation of the elements. This is a common error. Example: The population in the town doubled from ten thousand to twenty thousand in five years. The data is visualized with two squares on a map, with the second having its sides twice the length of the first (10,000 to 20,000). But the area of the second square four times that of the first, not double.
- The way in which you label and order the elements in a chart will make some arguments more immediately evident. If you want to compare quantities, be sure they are displayed in proximity. Example: when comparing the population size of states should you put the states in

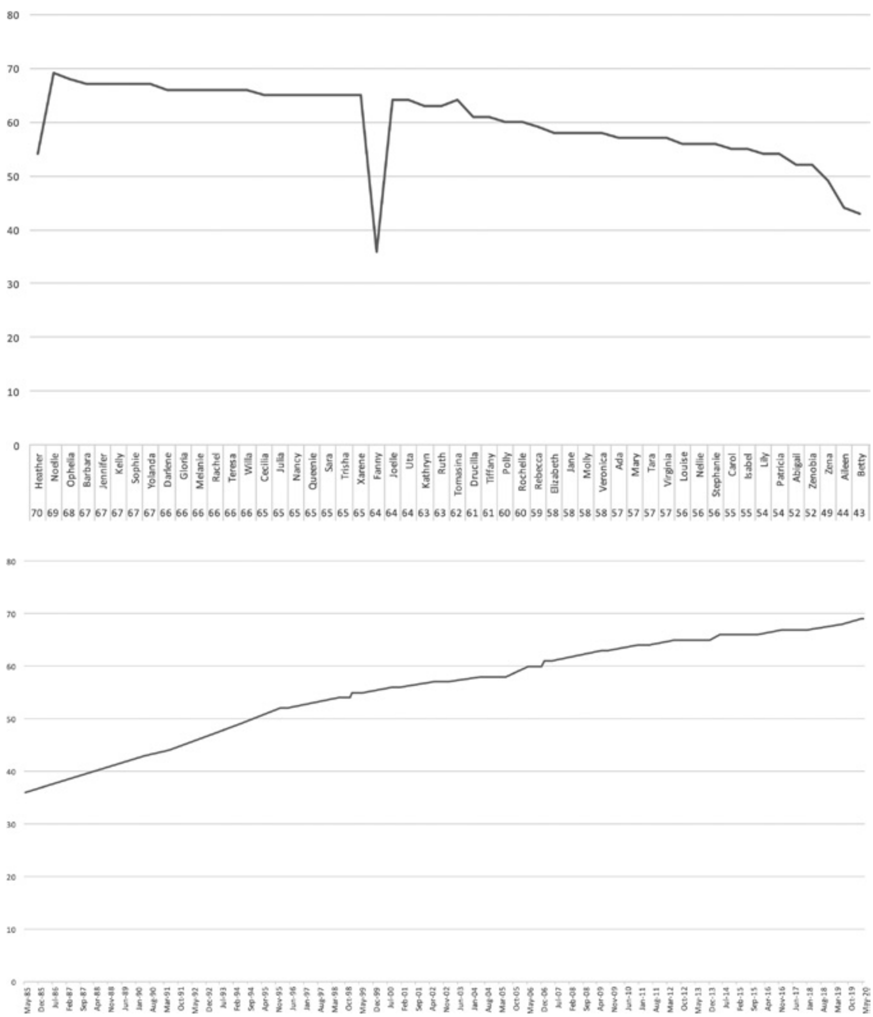


Figure 6.4a Meaningless graph of height among a group of girls graphed continuously and *6.4b* Graph of the change of one girl's height changing over time (JD)

alphabetical order or put the data in size order? Which is going to make the information more legible?

- The use of labels is crucial and their design can either aid or hinder legibility. Where are the labels? How much work are you adding to your reader's experience?
- Another consideration and challenge is the choice of a scale. When values are relatively close, the scale of the chart can be kept consistent.

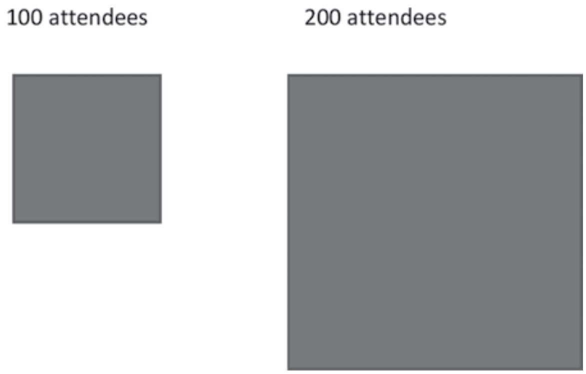
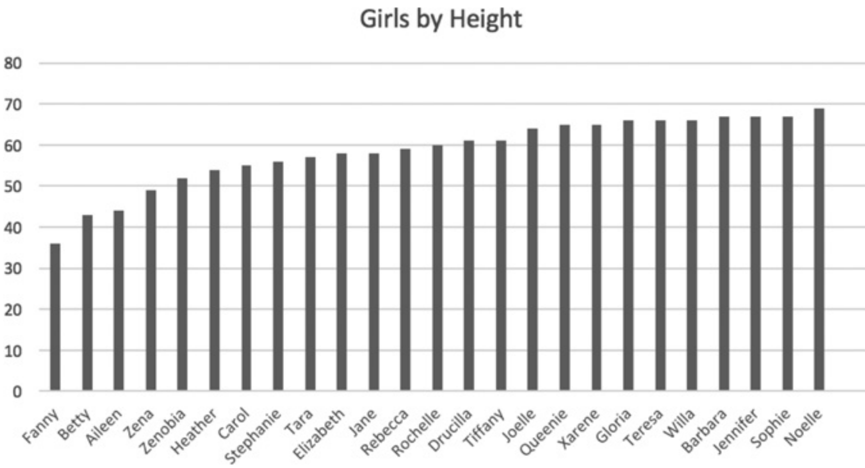


Figure 6.5 Classic error in which a value increases numerically but the area increases geometrically. The quantity on the right is twice that on the left, but the area is four times as large (JD)

But imagine the charts of date and page outputs in the example above if in one year the author produced 2000 pages. To show this value, the scale would need to extend to forty times its current height. The result would be that the difference between 20 pages and 50 pages would barely register. The legibility of the graph and patterns would be altered. To deal with such anomalies, charts are drawn with "broken" or modified scales, leaving a gap between lower and upper values. These gaps need to be noted and taken into account in some kind of legend, labeling, or documentation. [See Exercise #2: Reverse engineering a visualization.]

The rhetoric of graphics

Every visualization has a history to its format (Friendly 2007). The earliest forms of visual records seem to have been observations of the planets and other natural cycles. Early accounting systems for tracking inventory and also for taking census information used tabular forms. These allow easy correlation across values. The notion of continuous graphs, line charts, and other visual representations of information from natural or social phenomena did not appear until modern times. These emphasize continuous change. William Playfair, the 18th century statistician, is credited with the invention of many forms of bar chart and continuous graph still in use today. Playfair was working with what he called "Political Arithmetik," or the tracking of information relevant for guiding politics and policy in economic arenas (Norman 2004–2020). Playfair's visual solutions were very elegant as well as highly legible. Keep in mind that the science of statistics is also relatively modern, originating chiefly in the 17th century with techniques developed



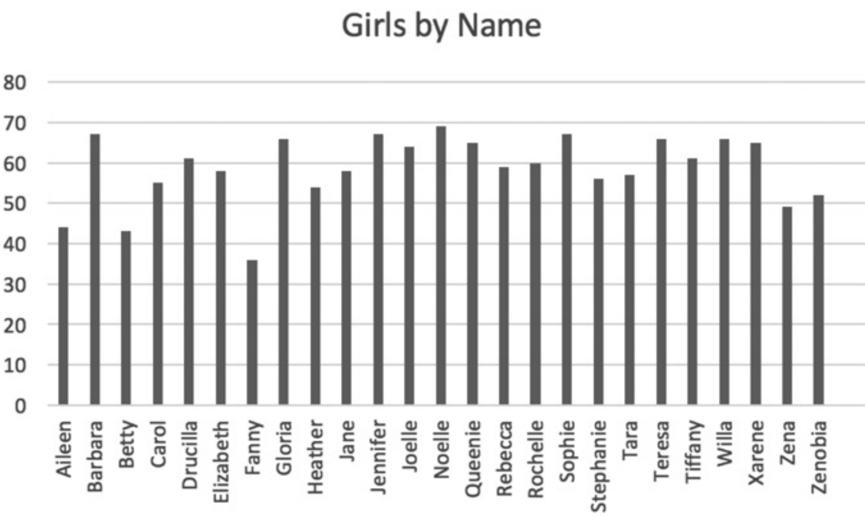


Figure 6.6a and 6.6b Charts showing ordering and labels: The first chart makes it easy to find individuals by name, the second makes it easy to compare heights and correlate with names (JD)

by the French mathematicians, Blaise Pascal and Pierre de Fermat, to gauge the risks of gambling (Apostol 1969).

The power of visualizations has been understood for a long time. In the 19th century, the nurse and activist Florence Nightingale created a specific format—known as the cockscomb because of its resemblance to the rooster's crown—to make her point about the fact that more deaths occurred among the wounded in field hospitals than on the battlefield. She deliberately chose

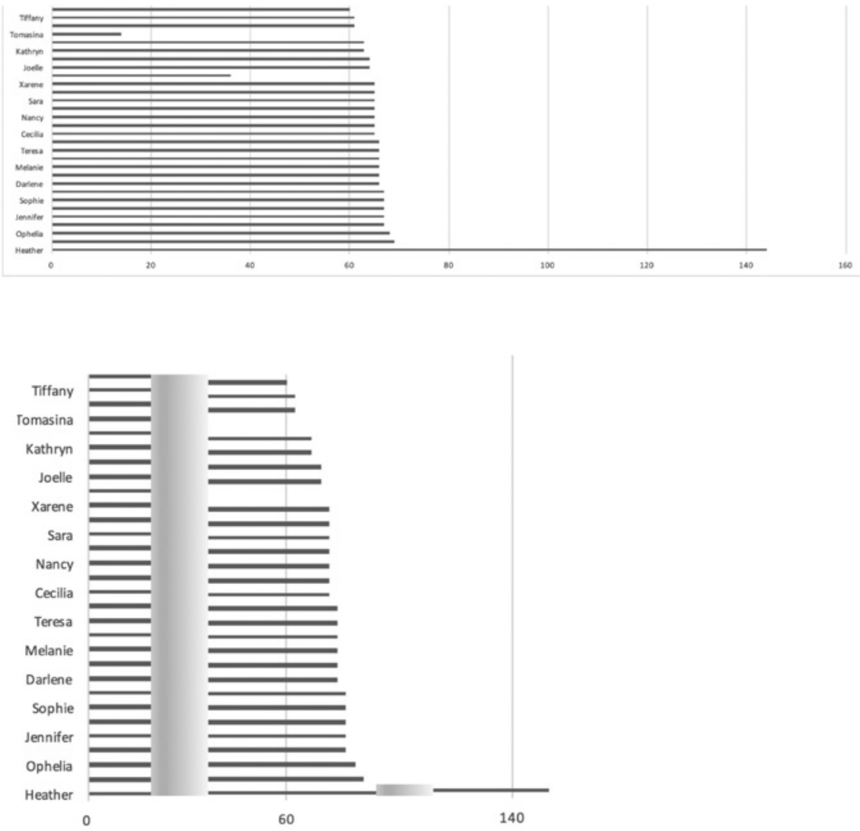


Figure 6.7 The scale has to stretch to include the height of the outlier and makes it difficult to compare the differences among the close values in the middle range. Making a "break" in the scale could allow focus on the area in which the meaningful information is present (JD)

a format that exaggerated this information. She used the difference in her data values to set the length of a radius in a circular form, also known as a polar area diagram, thus distorting the area. (This is similar to the example of the square, above, but here the area is calculated by the standard formula $A = \pi r^2$ (area = pi x square of the radius r). The contrast was dramatic, and she won her argument.

This kind of exaggeration can be very misleading in any chart that uses area as a feature of its graphical form. As already noted, when using graphics that are based on area, such exaggerations are built in. This distortion is a regular feature of information display on maps, as will be seen ahead.

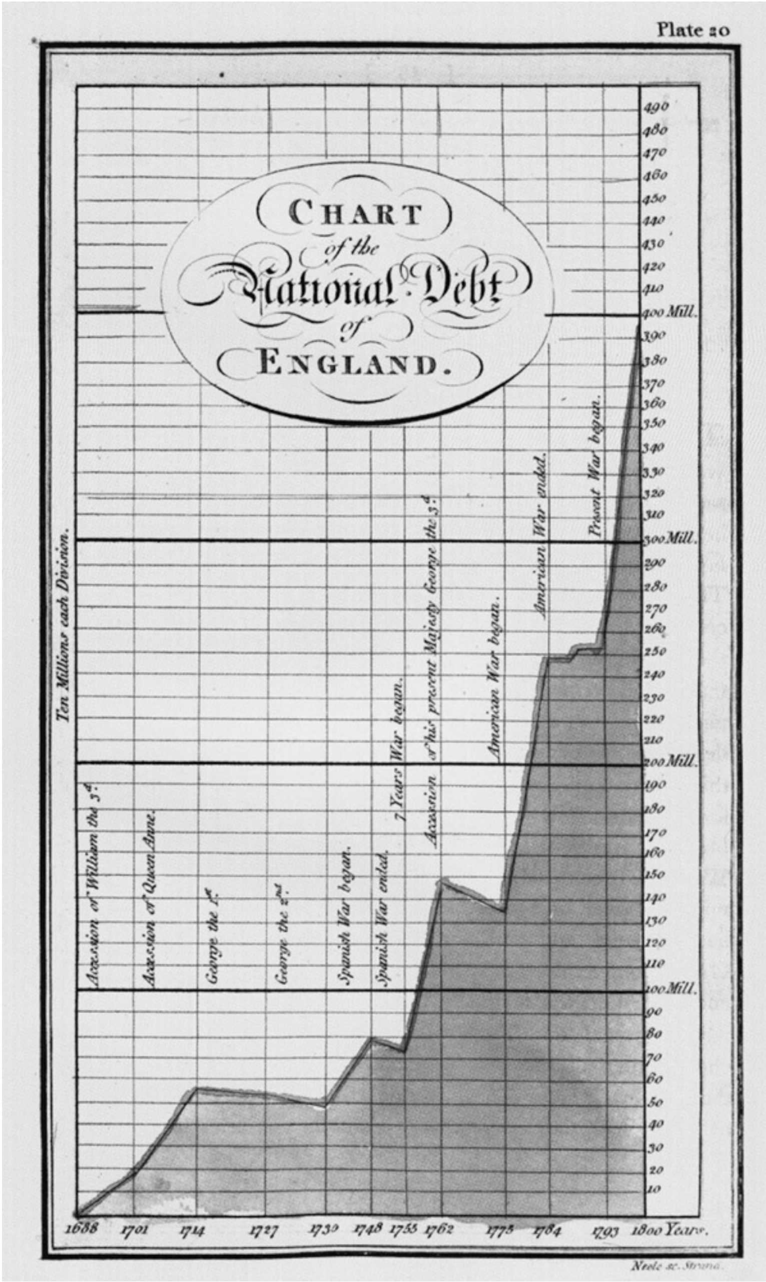


Figure 6.8 William Playfair Chart of the National Debt, *The Commercial and Political Atlas*, 1786 (Public domain)

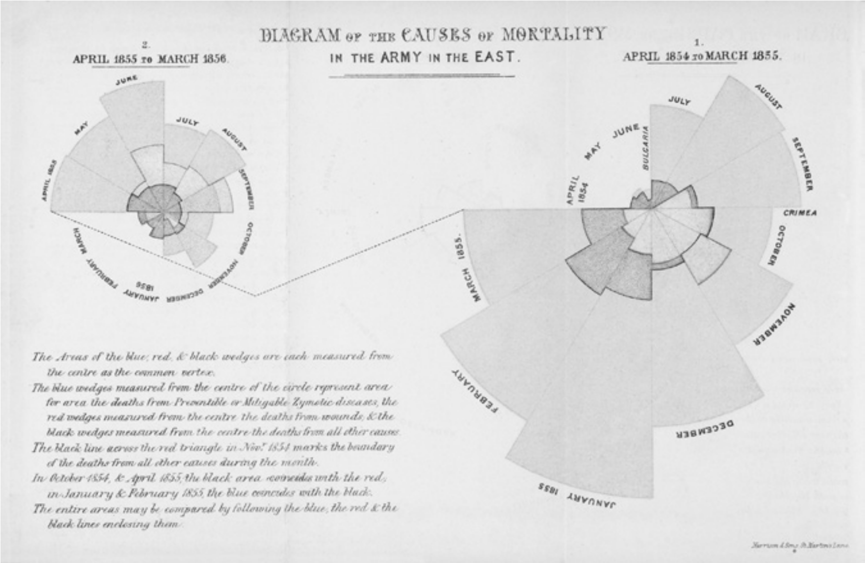


Figure 6.9 Florence Nightingale, cockscomb diagrams, 1854–55 (Public domain)

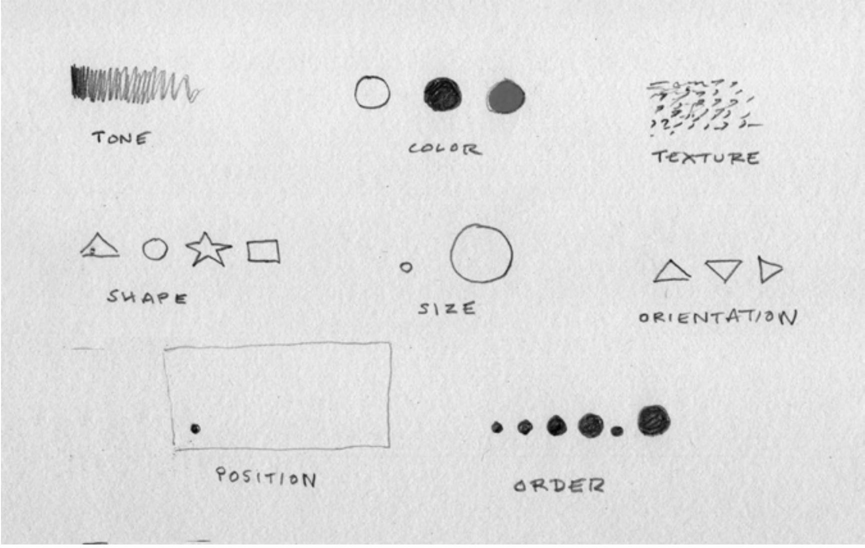


Figure 6.10 Graphic variables (JD)

Components of visualizations

When considering visualizations, a few fundamentals besides the type of chart and the rhetoric of its impact can be useful in guiding design decisions. The components of visualizations include *axes, elements, scales, order/ sequence, values of coordinates*, and the *graphic variables*.

Axes establish the basis for mapping values. Typically, the x-axis (left to right) is used to graph a value that changes over time (dates) while the y-axis is used to chart a specific value (cost of living, sea levels, etc.) against it. These are sometimes, but rarely, augmented with a z-axis that gives depth to the chart. However, mapping a third variable is trickier than it sounds, and it is often easier to simply have this information displayed as a second set of points or lines (earnings might change, for instance, as page outputs increased in our reference example). The basic coordinate x-y system was invented by René Descartes and is sometimes referred to as a Cartesian grid. (The apocryphal version of the story is that he was trying to figure out how to pinpoint the location of a fly on his ceiling (Wild Maths n.d.).) The grid uses standard metrics that remain the same across the full extent of the chart. One question to ask is whether you can imagine situations in which a metric might need to change? What conditions might require an alteration in one area of a neutral grid. Is every square the same—even if a spider is lurking in one?

Elements are the bars, lines, points, symbols, or other features that express value. They are always read against the axes. Even in a pie chart, the percentage is read in relation to an axis—this is the circumference, which forms the 100% boundary of the whole.

Scales set the specific metric to be used—inches, feet, number of units, dates, and so on. Scales have a start and endpoint. If I am measuring the difference in height among a group of giant statues, all of which are over twenty feet high but less than twenty-one feet high. Should I measure them in light years? Millimeters? Some scales are too large or too small to be useful. If I am measuring the occupancy of an airport, a scale of years might be too large, but a scale of seconds will be too small.

Order and sequence are generally determined by data and given as logical an expression as possible. Putting the work of an author into size order might be trivial and putting all of the paintings in the world into a single date sequence might be meaningless. The order and sequence should be meaningful to the research—and for communicating information in a visualization.

Values of coordinates are generated by the axes. But a major difference exists between the value of crossing points—where one axis intersects another—and continuous values within the lines, grid, or tick marks. Are discrete or continuous values being gauged and presented?

Finally, graphic variables are the features of visual language: color, tonal value, size, shape, orientation, position, and texture. These will be revisited in the discussion of mapping but designating a variable for a specific purpose makes good sense. Shape is very legible, so distinguishing different data types with stars, circles, squares, triangles, and other icons in a chart makes for legibility—provided there are not too many types of data. Tonal value is useful for showing changes of intensity, as in heat maps. Size generally indicates quantity, but can signal importance, particularly with typography. Color, like shape, is very legible and makes distinctions highly visible, as does texture. Use color to distinguish themes or topics (the height of freshmen relative to the height of seniors). Dotted lines are easily distinguished from solid ones. These allow information to be carried by the visual elements, not just their labels. Orientation should be used when a feature of the data correlates to it—like wind direction. Position is generally determined by coordinates but can also be part of the overall design—what is near what and why when proximity is significant.

Using graphic variables systematically increases the communicative legibility of your visualization. [See Exercise #3: Analyze the data-graphic connection.]

Checklist for visualizations

- Assess your data: Is it composed of discrete or continuous information?
- Choose the appropriate scale: too small a scale may make the important differences in value hard to spot and too large may exaggerate it. If outliers stretch the scale for a few data points, consider a gap in the scale and an explanation.
- Is the labeling efficient for use? What order should the information take to be meaningful and usable (alphabetical order of country names makes them easy to find but might separate values and make them hard to compare visually)?
- Use graphic variables carefully: shapes carry information readily, tonal values should be used for data that has a gradient, texture has little "meaning" in itself, and color can carry symbolic value or simply be used for differentiation.
- Proximity of labels to values is optimal for reducing cognitive load; make it easy for the viewer to correlate information.
- Never use changes in area to show a simple arithmetic increase in value.
- Review the graphic to see if it contains elements that are "incidental" artifacts of production rather than semantically meaningful ones.
- While illustrations, images, or exaggerated forms may be considered "junk," they can also help set a theme or tone when used effectively.

A few last thoughts

Visualizations do not usually show the lifecycle of the data. Decisions about parameterization, even the way samples were taken and what elements of the data were "cleaned" up and removed are all missing from the final visualization. Similarly, the history of the data within its institutional or research context may not be documented. Finding the source for the information can be difficult once the visualization exists. Thus, the question of whose authority—whose voice and point of view—is represented in the visualization can be very difficult to answer. A process of reduction, simplification, and what is known as *reification*—making a concept appear to be a *thing* (solid, tractable, and understandable)—takes place in the production of visualizations. The statement, "Information visualizations are reifications of misinformation," suggests that the apparent straightforward communication in a visualization should be treated with skepticism, rather than simply accepted, in spite of the value of these images for data presentation (Fenton 2015). In data journalism, these concepts are referred to as "the lie factor," and ethical practitioners work conscientiously to avoid misleading graphics. [See Exercise #4: Misleading graphics.]

Recent scholarship draws attention to critical concerns in this area of digital research. The work of feminist scholars questions some of the assumptions about who controls the technology of production and whose values are embodied in the information design process (D'Ignazio and Klein 2016). A cache of hand-drawn works by the African-American activist, W.E.B. Du Bois, sheds light on this formerly little-known aspect of his work and the way he made use of data visualization for advancing critical discussions of race (Mansky 2018). Their hand-drawn quality inflects their presentation, raising questions of equitable access to resources. A very different approach to hand-drawn visualizations appeared in a "Dear Data" project of letters exchanges between Georgia Lupi and Stephanie Posavec, both sophisticated information designers who used the experiment as a way to explore the possibilities of analog presentation (Lupi 2017). Many artists have been intrigued by data flows and visualizations as opportunities for aesthetic investigation, some of which will be touched on ahead in the discussion of complexity.

Takeaway

Information visualizations are metrics expressed as graphics. Information visualizations allow large amounts of (often complex) data to be depicted visually in ways that reveal patterns, anomalies, and other features of the data. No data has an inherent visual form. Any data set can be expressed in any number of standard formats, but only some of these are appropriate for the features of the data. Certain common errors include misuse of area, continuity, and other graphical qualities. The rhetorical force of visualization is often misleading. All visualizations are interpretations, not presentations of fact. Some graphic features of visualizations are artifacts of the display, not of the data, and can contribute to the reification of misinformation. Understanding the language of graphics is an art that combines conceptual insight with design acuity. Still, even a novice can produce useful graphics with current platforms and tools. The challenge is to produce graphics that are appropriate to the research task and communication of arguments.

Exercises

Exercise #1: arange of graphs

Try various visualizations for suitability. Take one of these data sets through a series of Microsoft Excel visualizations. Which make the data more legible? Less?

- United States AKC Registrations
  http://images.akc.org/pdf/archives/AKCregstats_1885-1945.pdf
- Sugar Content in Popular Halloween Treats
  www.popsugar.com/fitness/Calories-Halloween-Candy-Fun-Size-Treats-5452936

Exercise #2: reverse engineering a visualization

Look at Google's Public Data directory and the visualizations generated from the files. Can you locate the basic components (axes etc.) and evaluate them for common errors? Consider where the data comes from and what may be missing from its visualization.

www.google.com/publicdata/directory

Exercise #3: analyze the data-graphic connection

Imagine you are collecting data from the classroom on 1) classroom use, 2) attention span of students, 3) snack preferences, 4) age, height, and weight comparisons in a group? For what kind of data gathered in the classroom would you use a column chart? Browse this D3 gallery of visualizations for other formats: https://observablehq.com/collection/@observablehq/visualization

Exercise #4: misleading graphics

What is the concept of the "lie factor" and how is it visible at the following link?

www.datavis.ca/gallery/lie-factor.php

In each case consider legibility, accuracy, or the argument made by the form. What is meant by a graphic argument?

Recommended readings

D'Ignazio and Lauren Klein. 2016. "Feminist Data Visualization." *IEEE*. www.academia.edu/28173807/Feminist_Data_Visualization.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly*. www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Lupi, Giorgia. 2017. "Data Humanism: The Revolutionary Future of Data Visualization." *PRINT*. www.printmag.com, www.printmag.com/post/data-humanism-future-of-data-visualization.

6b Networks and complex systems

The concept of a network has become ubiquitous in current culture (Zer-Aviv 2016). Almost any connection to anything else can be called a network, but properly speaking, a network has to be a system of elements or entities that are connected by explicit relations. The term network is frequently used to describe the infrastructure that connects computers to each other and to peripherals, devices, or systems in a linked environment. While that is an accurate description, the networks we are concerned with in digital humanities are created by relationships in an information system. This might be the connection of books to authors, paintings to collections, people in communication with each other, or objects and ideas in circulation.

Unlike other data structures we have looked at—databases, markup systems, classification systems, and so on—networks are defined by the specific *relations* among elements in the system rather than simply by the content types or components. The elements of networks are *nodes* (points or entities) and *edges* (links or relations that connect the nodes).

Good examples of networks are social networks, traffic networks, communication networks, and networks of markets and/or influence. Many of the same diagrams are used to show or map these networks, and yet, the content of the relations and of the entities might be very different in each case. Standardization of graphic methods can create a problem when the same techniques are used across disciplines and/or knowledge domains, so a critical approach to network diagrams is useful.

Technically, networks are graphs, not visualizations. The distinction is important because graphs can include the feature of *directed* or *undirected* connections. These indicate a one-way (or two-way) movement in the connection. For example, money may flow from a parent to a child, but more rarely flows back in the other direction. Influence may move from a predecessor in a field to a new development, but might not flow both ways, particularly if the author of the earlier work is deceased.

The computational process by which graphs are produced requires that data be structured in a specific way: source > relationship > target. The vocabulary of nodes and edges is used to differentiate entities (source and target) from their relationships (edges). Particular features of networks are used to process the data in relation to notions of *centrality, closeness*, and *between-ness*. Centrality is the measure of how important any particular node is, measured by the number of connections and type (to or from other nodes) (Bhasin 2019). In graph theory, which governs the description and production of networks, other factors are gauged to assess the factors of between-ness and closeness based on the pathways established among and through nodes. The important principle here is that while some features of the display can be read literally (numbers of connections and directions), the literal distance of nodes from each other in a visualization can only be read logically. This is because the display algorithms try to preserve the statistical features of the data but are often optimizing legibility at the same time. As with all visualizations, it is important to be careful about reading the visual display literally. A node pulled out to a great distance might simply be far from the center so that its label can be seen.

Sketching network concepts

You can sketch a network on paper quite easily. Imagine yourself as a node and then draw lines to everyone you know in your immediate circles (family, friends, clubs, and groups) around you. Draw their links to each other. Think about degrees of proximity and also connections among the individuals in different parts of your network. How many of them are linked to each other as well as to you? If you can code the lines that connect persons to indicate something about the relationship, how does that change the drawing? What attributes of a relationship are readily indicated? Which are not? Think about the difference between how *often* you exchange communications with someone and how *central* they are to the exchanges among others. A parent might be someone to whom everyone is connected, but your own communications might be more frequent with your siblings. When a network algorithm processes data, it tries to calculate these properties.

Social networks are familiar and the use of social media has intensified our awareness of the ways social structures emerge from interconnections among individuals. A network may or may not have emergent properties, may or may not be dynamic, and may have varying levels of complexity. Simple networks, like the connection of your computer to various peripheral devices through a wireless router in your home environment, may exhibit very little change over time, at least little observable change. But a network of traffic flow is more like a living organism than it is like a set of static connections. Though nodes may stay in place, as in airline hubs and transfer points, the properties of the network have capacity to vary considerably. This is certainly true with social networks, most of which are highly dynamic, even volatile.

Properties of networks

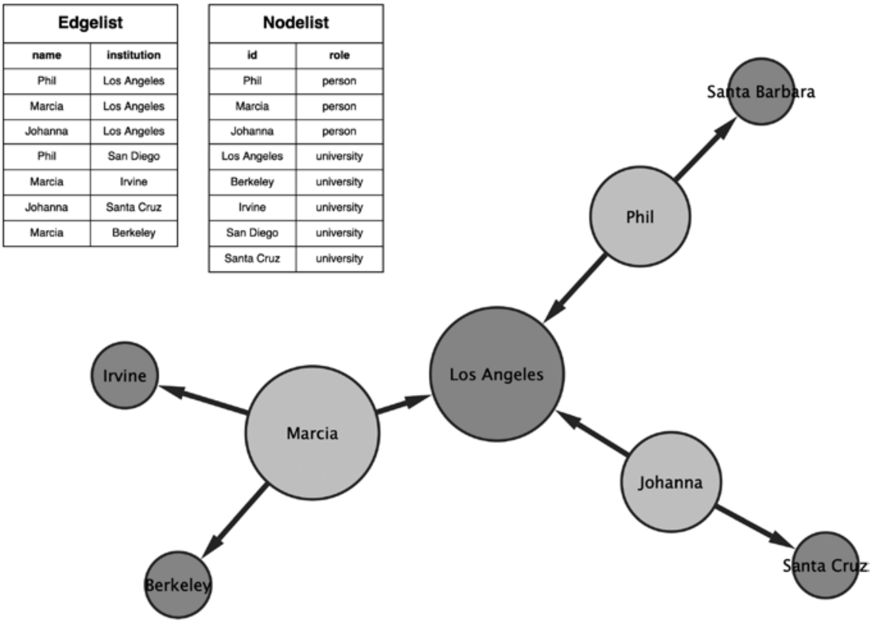
Networks exhibit varying degrees of closed-ness and open-ness. Researchers interested in complex or emergent systems are attentive to the ways boundary conditions are maintained under different circumstances, helping to define the limits of a system. Social networks are almost never closed, and like kinship relations or communications, they can quickly escalate to a very high volume of connections. Epidemiologists trying to track the spread of a disease are aware of how rapidly the connections among individuals grow exponentially in a very short period of time. Network analysis is an essential feature of textual studies, particularly of citations and influences. Network analysis plays a large role in policy and resource allocation as well as in other kinds of research work.

To reiterate what was already stated, the basic elements of any network are nodes and edges. The degree of agency or activity assigned to any node and the different attributes that can be assigned to any relation or edge will be structured into the data model. The data for linked "nodes" are understood as "source" and "target" (even though these can be reciprocal, and also, unrelated). Edges are the connections specified between the nodes.

For an example of this in action, look at the project, Kindred Britain, which studies connections of about 30,000 British individuals. The project is meant to show the many ways in which connections form through social networks, family ties, business, and political circumstances.

Another interesting example looks at the genre of "exchange poems" that were part of medieval Chinese culture. These had traditionally been characterized by schools and styles. But new research positioned them in social networks. To paraphrase the work of the project director, Tom Mazanec, it turns out that the Buddhist monks in the 7th to 10th centuries of the Tang dynasty were central "nodes" in the network of literary production (Mazanec 2017). Graphing these has changed the way this form of Chinese poetry is understood and its place in cultural and social life. Relations between literary forms and social activity that were not noted before were revealed through the analysis.

Art historians Pamela Fletcher and Anne Helmreich used network analysis to study the London art market, and found surprising insights from sales records and auction catalogs (Fletcher and Helmreich 2012). Artists and styles that have not necessarily been seen as important by later art historians turned out to play a significant role in the markets of the time, even if they have largely vanished from the canon. [See Exercise #1: Kindred Britain, a social network project.]



Edgelist	
name	**institution**
Phil	Los Angeles
Marcia	Los Angeles
Johanna	Los Angeles
Phil	San Diego
Marcia	Irvine
Johanna	Santa Cruz
Marcia	Berkeley

Nodelist	
id	**role**
Phil	person
Marcia	person
Johanna	person
Los Angeles	university
Berkeley	university
Irvine	university
San Diego	university
Santa Cruz	university

Figure 6.11 Network graph, edgelist, and nodelist (Image courtesy of Nick Schwieterman) (NS)

Tools and tutorials

As is the case with most digital methods, the fundamental principles of networks are well-established but the tools and platforms change over time. The principles that need to be understood are those of the data structure just mentioned (source > relation > target). The relations among entities, known as edges, can carry weight and also annotations. Network graphs are generated using statistical processes. In a network, the graph is generated by statistical assessment of frequency and weight of relations. With a massive data set—think of a day of Twitter feeds—these calculations become highly complex. Though a simple network can be sketched by hand (you know who is in your immediate social circles), generating a graph of a complex data set would be almost impossible without computation since it involves calculation of relative values across a number of variables (frequency, weight, directionality, etc.).

One tool frequently used in digital scholarship for generating network diagrams is Cytoscape, an open-source platform that can be downloaded directly from the Web and installed on a laptop or desktop. Gephi is another. Learning to use such a tool takes some time but has the advantage that it is a professional level program designed to handle data at every scale from small to large. Understanding the data model before you begin—what is connected to what and how are you characterizing the links or relations—is essential if the digital tools are to be used effectively. Conceptualizing the network before it is visualized automatically helps keep a critical view of what the program produces. [See Exercise #2: Comparing network diagrams.]

Cytoscape and Gephi make use of data in a number of structured formats, among them, some specifically designed for graphs (Rana 2018). These include GraphML, or Graph Markup Language, standards for networks. They readily store information about labels and attributes of the nodes and edges. CSV, Comma Separated Values, a common format in all spreadsheets, can be used to define nodes and edges in weighted to/from pairs (or in linked pathways). GEXF, or Graph Exchange XML Format, was designed by the developers of Gephi, another free, open-source tool for network production and analysis. Networked data formats have specific requirements and a strict syntactic structure. Creating a small data set to work through the tutorials with your own examples is the best way to see what is happening at each step. [See Exercise #3: Cytoscape tutorial.]

Though the data structure is critical in network diagrams, learning to read the graphical output generated is also important. In an initial visualization, especially of large datasets, networks tend to look like "hairballs." They are tangles of lines connecting points, often very densely packed into a small image. Working to open up the nodes and stretch the edges allows insight into the ways the network is branching and where the main areas of connection lie. As has been mentioned above, keep in mind that a network diagram display conforms to protocols that optimize screen space for legibility. While the relationships in a network display are generally accurate, the literal distances on the screen are not. Attaching semantic value—meaning—to the spatial placement of the nodes, can be misleading.

A final challenge is visualizing dynamic systems in static form. Very few social networks are static, though the analysis of historical materials, connections, and activities may be. Information in the data set needs to be carefully scrutinized to be sure that events from different time periods or unrelated events are not conflated into a single graph.

Complex systems

Systems that follow non-linear processes of development are called complex. This does not mean complicated. A complex system can be as simple as a relationship between two people, a person and an environment, or an environment and changing conditions (Clemens 2019). What makes it complex is that the development of the system cannot be predicted—because the processes are non-linear and/or non-deterministic from a statistical standpoint.

The conditions in which they emerge continue to change and elements in the system interact in unpredictable ways. Weather systems are a paradigmatic example of complex systems, but so are stock markets, political processes, social relations of all kinds, and cultural activities. Who could have predicted that a conceptual artist named Marcel Duchamp would confound the conventions of the Western art world in 1917 by displaying a urinal upside down in an exhibit? Or that Mao Tse Tung would come to power in the Chinese Revolution? Or that the presence of the Missions in Australia would create an opportunity for art practices that were 20,000 years old to become codified in the medium of paint on oil and board? (Artlandish n.d.). These are examples of complexity at work. Many—even most—cultural processes are complex but modeling these requires more than creation of a data set. This work involves modeling behaviors of agents and conditions in a system.

Information designers—and artists—have been intrigued with visualizing complexity. Art exhibitions featuring data aesthetics have become common (Remondino, Stabellini, and Tamborrini 2018). The result has been a rich vocabulary of vivid and dynamic information visualizations—as well as some "eye candy" that may be more seductive than meaningful (Lima 2013). The process of constructing data and formulae for visualizing complexity is more complicated than it is for other visualizations (Yau 2007–2020). [See Exercise #4: Complexity.]

Advanced network theory pays attention to emergent properties of systems. The capacity of networks to "self-organize" using very simple procedures that produce increasingly complex results makes them useful models for looking at many kinds of behaviors in human and non-human systems. Networks do not have to be dynamic, but complex systems almost always are. The study of systems theory and of networks is relatively recent and only emerged as a distinct field of research in the last few decades. We might argue, however, that novelists and playwrights have been observing social networks for much longer, as have observers of animal behavior, weather and climate, and the movements of heavenly bodies held in relation to each other by magnetism, gravity, and other forces. Most dynamic phenomena are complex systems governed by non-linear processes.

Takeaway

Networks consist of nodes (entities) and edges (relations). The data model for a network is a simple three-part formula of entity-relation-entity. This can be structured in a spreadsheet and exported to create a network visualization. Networks emphasize relations and connections of exchange and influence. Refining the relations among nodes beyond the concept of a single relation is important and so is the change of relations over time. Social networks change constantly, as do communication networks, and the relations among the technology that supports a network and the psychological, social, or affective bonds can alter independently.

Exercises

Exercise #1: Kindred Britain, a social network project

Explore the site and then discuss the selection of individuals, the character and quality of relations, explicit assumptions and implicit ones, and the diagrams and their rhetorical power.

http://kindred.stanford.edu/#

Exercise #2: comparing network diagrams

Go to: https://linkedjazz.org/network/ Determine what information you can reasonably extract from this graph. Now toggle between modes. Does this change your understanding? Or go to: www.databasic.io/en/connectthedots/ Network visualization with interactive sample data sets created by Rahul Bhargava and Catherine D'Ignazio.

Exercise #3: Cytoscape tutorial

This manual can be accessed without downloading and goes step by step through the basics of network graph construction. It is provided free of charge by the people who designed and maintain the standard platform for this work. Read through the table of contents and introduction to get oriented. http://manual.cytoscape.org/en/stable/Introduction.html

Exercise #4: complexity

Look at half a dozen examples on Nathan Yau's site: https://flowingdata.com/about/

What are the dimensions added in complex systems that are different from those of static visualizations? What is the correlation between graphic expression and information?

What role does aesthetics play in these projects?

Recommended readings

Grandjean, Martin, and Aaron Mauro. 2015. "A Social Network Analysis of Twitter: Mapping the Digital Humanities Community." *Cogent: Arts and Humanities* 3 (1). www.tandfonline.com/doi/full/10.1080/23311983.2016.1171458.

Weingart, Scott. 2011. "Demystifying Networks, Parts I & II Journal of Digital Humanities." *Journal of Digital Humanities* 1 (1). http://journalofdigitalhumanities.

References cited

Apostol, Tom. 1969. "A Short History of Probability." In *Calculus*, Vol. II. John Wiley & Sons. http://homepages.wmich.edu/~mackey/Teaching/145/probHist.html.

Artlandish. n.d. "Australian Aboriginal Art." www.aboriginal-art-australia.com/aboriginal-art-library/the-story-of-aboriginal-art/.

Bhasin, Jasin. 2019. "Graph Analytics—Introduction and Concepts of Centrality." *Towards Data Science*. https://towardsdatascience.com/graph-analytics-introduction-and-concepts-of-centrality-8f5543b55de3.

Clemens, Marshall. 2019. "Visualizing Complex Systems." *New England Complex Systems Institute*. https://necsi.edu/visualizing-complex-systems-science.

Fenton, William. 2015. "Humanizing Maps: An Interview with Johanna Drucker." *PC*. www.pcmag.com/news/humanizing-maps-an-interview-with-johanna-drucker.

Fletcher, Pamela, and Anne Helmreich. 2012. "Local/Global: Mapping Nineteenth-Century London's Art Market." *Nineteenth Century Art Worldwide* 11 (3). www.19thc-artworldwide.org/autumn12/fletcher-helmreich-mapping-the-london-art-market.

Friendly, Michael. 2007. "DataVis." www.datavis.ca/index.php.

Lengler, Ralph, and Martin J. Eppler. 2007. www.visual-literacy.org/periodic_table/periodic_table.html.

Lima, Manuel. 2013. *Visual Complexity: Mapping Patterns of Information*. New York, NY: Princeton Architectural Press. https://medium.com/@mslima/visualcomplexity-com-ad9a12fa2c1a.

Lupi, Georgia. 2017. "Dear Data, the Project." http://giorgialupi.com/dear-data.

Mansky, Jackie. 2018. "W.E.B. Du Bois's Visionary Infographics Come Together for the First Time in Color." *Smithsonian Magazine*. www.smithsonianmag.com/history/first-time-together-and-color-book-displays-web-du-bois-visionary-infographics-180970826/.

Mazanec, Tom. 2016–17. "Chinese Exchange Poems." https://cdh.princeton.edu/projects/chinese-exchange-poems/.

Norman, Jeremy. 2004–2020. "The History of Information." www.historyofinformation.com/detail.php?entryid=2929.

Rana, Ashish. 2018. "Getting Started with Network Data Sets." *Towards Data Science*. https://towardsdatascience.com/getting-started-with-network-datasets-92ec54958c07.

Remondino, Chiara L., Barbara Stabellini, and Paolo Tamborrini. 2018. "Exhibition: Visualizing Complex Systems." https://systemic-design.net/wp-content/uploads/2019/05/RSD7Exhibition_VisualizingComplexSystems.pdf.

Wild Maths. n.d. https://wild.maths.org/rené-descartes-and-fly-ceiling.

Yau, Nathan. 2007–2020. "Flowing Data Site." https://flowingdata.com.

Zer-Aviv, Mushon. 2016. "If Everything Is a Network, Nothing Is a Network." Visualising Information for Advocacy, visualisingadvocacy.org. https://visualisingadvocacy.org/node/739.html.

Resources

Cytoscape https://cytoscape.org/.
Gephi ttps://gephi.org/.
Kindred Britain http://kindred.stanford.edu/#.
Network Graphs (Flourish Studio) https://app.flourish.studio/@flourish/network-graph.
Social Network Graphs https://gwu-libraries.github.io/sfm-ui/posts/2017-09-08-sna.