17
Graphical Approaches to the Digital Humanities

Johanna Drucker

The digital humanities have adopted conventions of information visualization and user interface that come from disciplines whose epistemological premises are fundamentally at odds with humanistic methods. The implications of this permeate every aspect of digital work. The challenge of addressing graphical approaches in/to/from/for the humanities requires that we engage in a critical description of visualization and interface from a humanistic-critical perspective, that we analyze the epistemological assumptions built into their development, that we think through the issues in adapting these for the humanities, and that we envision alternatives.

Visualization and Interface

Visualization tools have been integrated into digital humanities projects with great rapidity, but this process has not been accompanied by an equal increase in attention to the intellectual implications of the use of graphical arguments built on tools borrowed from other disciplines. Meanwhile, the dependence of digital humanities on the basic operations of the graphical (and now tactile) user interface, rendered almost invisible by its familiarity, passes without substantive critical comment. Thirty years after WISYWYG, and more than twenty years after the browser-enabled display of networked materials, it's about time to reflect on the relations between graphical approaches to the humanities and the humanistic aspects of interpretation and knowledge production for which they are being used. Understanding the rhetorical force of graphical formats is a critical task to which humanities scholars are aptly suited by their training in close reading, though the language of visual modes of meaning production is still a foreign tongue for many. From analysis springs invention, and the push for innovation of graphical expressions suited to the needs and methods of humanists should get a boost from exposing the operations and limitations of current conventions.

Both major areas of graphical expression – information visualization and interface display – are premised on assumptions about data, knowledge design, content models, and file formats that need explicit attention if they are going to be understood from humanistic perspectives and reworked for humanities projects.

Start with a basic typology of visualization types and their appropriate uses: bar charts (comparison of value), pie charts and tree maps (percentage of values), scatter plots (discrete values), continuous graphs (change over time), network diagrams (relations and connections), directed graphs (influence or force), tree diagrams (relations of hierarchy), bubble diagrams (relative scale and value, though they tend towards distortion more than other formats). Galleries of good and bad, best and worst, hideous and prize-winning graphics abound online,

17
Graphical Approaches to the Digital Humanities

Johanna Drucker

The digital humanities have adopted conventions of information visualization and user interface that come from disciplines whose epistemological premises are fundamentally at odds with humanistic methods. The implications of this permeate every aspect of digital work. The challenge of addressing graphical approaches in/to/from/for the humanities requires that we engage in a critical description of visualization and interface from a humanistic-critical perspective, that we analyze the epistemological assumptions built into their development, that we think through the issues in adapting these for the humanities, and that we envision alternatives.

Visualization and Interface

Visualization tools have been integrated into digital humanities projects with great rapidity, but this process has not been accompanied by an equal increase in attention to the intellectual implications of the use of graphical arguments built on tools borrowed from other disciplines. Meanwhile, the dependence of digital humanities on the basic operations of the graphical (and now tactile) user interface, rendered almost invisible by its familiarity, passes without substantive critical comment. Thirty years after WISYWYG, and more than twenty years after the browser-enabled display of networked materials, it's about time to reflect on the relations between graphical approaches to the humanities and the humanistic aspects of interpretation and knowledge production for which they are being used. Understanding the rhetorical force of graphical formats is a critical task to which humanities scholars are aptly suited by their training in close reading, though the language of visual modes of meaning production is still a foreign tongue for many. From analysis springs invention, and the push for innovation of graphical expressions suited to the needs and methods of humanists should get a boost from exposing the operations and limitations of current conventions.

Both major areas of graphical expression – information visualization and interface display – are premised on assumptions about data, knowledge design, content models, and file formats that need explicit attention if they are going to be understood from humanistic perspectives and reworked for humanities projects.

Start with a basic typology of visualization types and their appropriate uses: bar charts (comparison of value), pie charts and tree maps (percentage of values), scatter plots (discrete values), continuous graphs (change over time), network diagrams (relations and connections), directed graphs (influence or force), tree diagrams (relations of hierarchy), bubble diagrams (relative scale and value, though they tend towards distortion more than other formats). Galleries of good and bad, best and worst, hideous and prize-winning graphics abound online,

and they are useful for teaching and research.[1] Many standard platforms or packages, from basic Excel charts to ManyEyes (www-958.ibm.com/software/analytics/manyeyes) or Tableau (www.tableau.com), make use of these same formats. They are the stock in trade of creating what Edward Tufte (2001), the high-profile designer, called the "visual display of quantitative information." The history of information graphics tracks the development of these formats from an impulse towards "political arithmetik" that arose in the eighteenth century as a part of bureaucratic administration of economies, demographics, and resources.[2] Some are older, of course, such as the spreadsheet grid whose content modeling capacities were already embraced by Babylonians for accounting purposes as well as the study of geometry (Drucker, 2014). And some are newer – such as the rectangular area "tree maps" whose surfaces would be difficult to calculate without the automated methods that are one of the strengths of digital processing.

The design of accurate and appropriate representations of quantitative information is an art, not a mere push-the-button mechanical task. Graphical organization, dimensions, scale, the labeling and ordering of lines, bars, symbols, textures, and so on all create artifacts whose visual form is what is present to a reader/viewer. The adage that an information visualization allows complex datasets to be perceived in an efficient manner, rendering patterns legible across sets of numbers whose relation is almost impossible to discern in spreadsheet form, should always be accompanied by a warning to be wary of reading the artifactual features of the graphic as if they are an unmediated presentation of underlying information. A simple change in the scale of the *x* or *y* axis exaggerates or minimizes the significance of any difference across values. Those interested in creating effective graphics can look to the classic works of practitioners (Calvin Schmid is an excellent example) for fundamental principles (Schmid, 1983). Get the basics right, and the rest will follow. But all visualizations are artifacts in their own right, and knowing how to read them *as graphical expressions* is crucial. The means by which a graphic produces meaning is an integral part of the meaning it produces. While excellent guides to production of graphics exist (albeit out of sight of the mainstream of digital platforms, whose automatic parameters and packaged solutions come without intellectual instructions), few, if any, guides articulate the meaning production mechanisms of graphical formats. Much work can and will be done in this area ahead. Here are a few suggestive thoughts towards that engagement.

The basic formats of information visualization retain the imprint of their disciplinary origins. Graphics come with arguments in their form, before they are linked to specific information sets. The columnar form of the spreadsheet, for instance, which, as noted above, goes back into Mesopotamian times, structures the assignation of values in columns and rows that can be read against each other. The power of this invention is mind-boggling, really, as the generative potential of the grid arrangement allows for combinatoric selection of values across its entire contents – you can read the areas of intersection for each column and row from a single graphic presentation. The format has been used for sorting and managing information for millennia. The discrete boxes and boundaries and the grouping of "data" in a field or column with a single designation allow for powerful epistemological moves. The articulation of the row/column format is complex, even if the presentation is simple, and the structure allows for generative

reading of values in varying combinations. The use of a grid in this structure emphasizes the spaces/fields for content, while a continuous graph makes use of a grid for the points at intersections of lines/values.

Grid forms do not express a hierarchy in their graphical system. Tree diagrams (not to be confused with tree-maps, which are area-based charts), by contrast, declare a parent–child relation linked by derivation and continuity. Nodes in a tree always have some connection to each other, and thus the articulation of the graphic form determines meaning as an aspect of these relations. Tree diagrams track their history to ancient images of the Tree of Life, but achieve widespread use in genealogical tables from the Tree of Jesse to the tracking of bloodlines across generations in humans, pedigree animals, and other systems where direct connections are being mapped (Watson, 1934; Cook, 1974; Klapisch-Zuber, 2007). In an interesting twist, the Sephirotic Tree of Jewish mysticism, though it maps a hierarchy with a divine being at its top, is often abstracted, combined with the traditional squares of opposition to create a logical structure whose meaning production relies on multiple crossings from node to node. Squares of opposition are formal graphic structures invented in classical antiquity to represent propositions in Aristotelian logic. While they enjoyed particular prominence in European medieval philosophy, squares of opposition were part of George Boole's formal studies in the nineteenth century as well as being adopted for structuralist and semiotic analyses (Boole, 1854; Parsons, 2012). The apparent simplicity of their form belies the complexity of argument and issues to which they can be put, and, like grids, they are generative in their provocation of multiple readings and interpretations.

Bar charts came relatively late into the family of graphics, invented for accounting and statistical purposes, and thus pressed into service in the eighteeenth century, with only rare exceptions beforehand. They depend on underlying statistical information that has been divided into discrete values before being mapped onto a bivariate graph. The techniques belong to the natural and social sciences, to the externalized standards of empirical metrics, to the presumption of repeatable results and observer-independent phenomena. Bar charts express discrete quantitative values as visual features (one value per bar), but the size, scale, and height of a bar becomes a significant aspect of its impact. By contrast, scatter plots display discrete values as points on an *x*,*y* grid, as single points in a set. Bar charts imply that a value is achieved by aggregation, a collection that has reached a certain level (like water in a rain column), while scatter plots merely mark a value produced against a metric. Though all metrics are constructs, of course, the scatter-plot point expresses an intrinsic value (like temperature, date, or weight) rather than an accumulated one (17 inches of snow). These distinctions are the essence of graphical literacy, and should be used in determining the fitness of any particular visualization to a particular task.

Bubble charts and area graphs are inventions of nineteenth- and twentieth-century information designers, though their precedents in using plane geometry, cartographic and architectural foundations of plans and maps of property, also have ancient origins. The difference between using geometry to represent and divide territory on a representation and the creation of graphics that use area as an effect of changing variables to generate a display requires an intellectual leap. The amount of calculation necessary to refigure the area of the circles in a

bubble diagram or the subdivisions of area in a tree-map is nontrivial in analog calculations but formulaic in digitally automated ones. Distortions abound in any representation of value that relies on circles, since the tendency to vary the radius results in a dramatic increase in area that does not correspond to the actual proportion of change. Florence Nightingale is credited with inventing the coxcomb diagram to communicate the dramatic level of fatalities from post-battle infection and disease. She made deliberate use of their distorting properties to be persuasive. Tree maps are more accurate, less distorting, and allow nesting and hierarchy to be displayed in the same system as proportion or percentage of value. The hard-edged rectangular format is highly legible, within reason, and can be remarkably efficient as a presentation of enormous amounts of complex information.

Network diagrams, directed graphs, and other depictions of relationships have become extremely popular forms of visualization in the humanities. They seem to work well to show communication systems, social relations, power and influence, or markets, as well as other kinds of relation-based phenomena. The edges and nodes that constitute the language of networks are reductive, and though they can be loaded with attributes that provide extra information and description to the depiction, they are basically static expressions of complex systems. The level of complexity necessary to model dynamic systems introduces another order of challenge into the analysis of information and its transformation into visualization. The family resemblance of one network diagram to another seems so much stronger than its individual character that it should give pause. Though usefully compact for communicating information, network diagrams often create their graphical display using algorithms that optimize legibility within the limits of screen real estate, organizing their nodes and neighborhoods for efficiency rather than for semantic or meaning-driven origins within the original materials. When we turn our attention, as we will in a moment, to the fundamental issue of data creation and extraction that lies under any visualization, these issues become even more clear.

The use of geographic information systems (GIS) and conventional maps for humanities projects opens a whole Pandora's box of issues – from historical anachronism to the assumptions that space is a static, *a priori* given, rather than a culturally contested and socially constructed phenomenon. The use of *Google Maps*, in particular, produces a set of images based on unquestioned assumptions about the ideology of cartographic rendering that no serious geographer would tolerate. Humanists tumble wholesale into the mapping exercise with rampant enthusiasm for sticking virtual pins in virtual maps, using terms like "geo-rectification" as if they were mere mechanical conveniences rather than acts of interpretative distortion (often enacting hegemonic values). Space and time, the two great philosophical concepts, are suddenly treated as givens in the digital humanities, as if the full weight of millennia of reflection on their philosophical complexity – the distinctions, for instance, between time as a container and temporality as a relational system, with its parallel in space/spatiality – were of no use whatsoever in the new enterprise. The challenges are central to digital humanities as a field, but will require more effort than is currently being put to their conceptual formulation if humanistic approaches are to find their way into visualizations of these profoundly significant and substantive issues. For now, this brief mention will have to

suffice, standing in for elaborate discussions ongoing elsewhere (Thrift, 2007; Drucker, 2009).

The basic rules of information graphic design suggest adopting a graphic format appropriate to the argument and the "data" (more on the scare quotes in a moment) using only as many graphic variables/elements as there are dimensions in the information (color, size, shape, tonal value, texture, orientation, position), and making the graphic as legible as possible, with as little distortion as can be managed. These are admirable principles, but they are grounded in an assumption of transparency in which the visualization is meant to "show" the form of the pre-existing dataset. One problem is that datasets, while they have structure and form, are not graphical, so the graphic expression is always a translation and remediation. In many cases, the resulting graphic has so much extra information that it gets read for its artifactual qualities rather than its informational ones. From a functionalist point of view, the directive to digital humanists is to learn the basic language of graphics and use it in accord with the professional guidelines developed by statisticians. From a critical point of view, however, the message is more skeptical and suggests a radical rethinking of the epistemological assumptions that the statisticians have bequeathed us. The fault is not with the source, since it is the borrowing for humanistic projects that is problematic, not the statistical graphics themselves. They work just fine for statistical matters (Börner, 2010).

This tiny glimpse into the graphical rhetoric structured in visualization conventions is meant to make the point that the use of any of these formats engages processes of meaning production that are built into its structures and derived from disciplinary origins whose epistemological viewpoint is embodied in the form. The same can be said at the micro-level, where the distinction between formats has implications for the ways information is encoded. Graphic image formats are either pixel-based or scalable vector graphics. This distinction divides the digital world into "pictures" – that is, images that can describe anything through patterns of color values in a tapestry on screen – and "shapes" – images with properties that constrain them. Many techniques are used for rendering images, but the rendering effects conform to data structures that are either descriptive or scalable. The challenge for the humanities is to consider ways the attributes in a scalable graphic can be given values that enact humanistic properties. One concrete way to think about this is to consider the "edges" or lines of connection (relations) in the network diagrams discussed above. "Nodes" (points or entities) can carry attributes that vary in a complex way, and modeling a relationship so that factors affect each other, as in an adaptive system, makes more sense than merely giving them weight or a value or quality that remains static. To create visualization programs that accommodate these considerations would be a move towards asserting the methodological foundations of humanities and their role in a computational environment.

In addition to the ongoing use of information visualizations, humanists are involved, as are all users of screen-based technologies, in the mediated experience of the user interface. The graphicality of interface has become almost invisible in the digital environment. Even though its conventions are recent, they are so familiar we take for granted the navigation bars, search boxes, drop-down and sidebar menus. Even when our movement through them is performed by pinching or swiping, the information structures are graphic. The basic tensions in interface between offering cues for behavior and presenting an information structure are part of digital

humanities projects just as much as they are present in commercial, news, or entertainment sites. The difference of domains is not marked by functional features, but by the absence of ads, a consistency throughout the site, subtle but distinctive branding, and other elements of graphic design. Similarly, the difference between a scholarly book and a storybook does not reside in its structure – both are bound pages with an open fore-edge and book block within covers – but in the details of type, font, layout, paratext, scholarly apparatus, and content. The conventions for ordering and organizing information in digital humanities projects are less established than the codes of interface. In other words, the use of navigation features, breadcrumbs, and so forth may be more or less standard, but the content model of a project on a major historical event will vary radically from one on the letters of an important political figure or the study of a collection of publications. Communicating the contents of a digital project is a knowledge design problem – since the multifaceted aspect of database structures offers different views into the materials. The decisions about presentation of a project's contents constitute an argument de facto, though they may or may not be considered in using the more or less default features of an interface with its "search," "home," and "contents" tabs. An interface usually has to balance between showing a model of the intellectual content or creating cues for user behaviors (Garrett, 2002). The question of making the organization of knowledge accessible and legible is a difficult one to address through design without overwhelming a user or constraining their engagement through a too-limited menu of access points. These are problems in any interface design, and in the humanities, where the shape of content is itself content, the question carries a certain extra weight.

Our understanding of the design issues in a digital environment can benefit from retrospective consideration of the ways analog artifacts provide a semantically inflected armature for meaning production. Or, put another way, we ought to pay more attention to the ways form is meaning, or presentation encodes the instructions that already determine how meaning is produced through such basic features as juxtaposition, hierarchy, sequence, proximity, continuity, rupture, position, scale, orientation, and so on. These basic graphical principles construct meaning. We read the top header on a page differently from a line within a text block, a title differently from a footnote, and so on. Each is positioned within the semantic field according to hierarchies that structure meaning as part of the performance of the image/text files/content that are presented in that armature.

The current conventions of graphical interface design are not the natural expression of any order in the social or natural world. They are conveniences that make optimal use of screen space. Techniques for introducing humanistic perspectives would begin with parallax views or any point-of-view system that inscribes a non-singularity within the information field and display. Difference, the recognition of the incommensurable distinction between one reading and another, one interpretative act and every other, can be rendered in subtle graphical ways, marking the trace of a subject's presence within the field of display. Point-of-view systems are central to the history of visual representation. They exist in all images, whether marked or not, whether they are conspicuously present, as in one-point perspective renderings, or rendered neutral and omniscient, as in axonometric and isomorphic projections. A point-of-view system places any and all representations into a subject position, away from value-neutral and

observer-independent claims to simply be what *is*. This shift, from presentation to representation, from an image that presents itself as *what is* to one that marks its *constructed-ness*, is the crucial shift from a non-humanist to a humanist perspective. This is the move that registers the presence of an individual or collective subject within the system of knowledge production.

In the early days of hypertexts and before the conventions of the Web, interface conventions were not fixed. The archives of the Electronic Literature Organization (http://eliterature.org) are an excellent resource for looking at the imaginative engagement with design of experience and navigation. What was being communicated and how in the structure of those early designs, and how might some of their experiments rejuvenate our imagination? The very fact that the designs force an awareness of the act of reading, of making one's way, and of being in the digital work seems useful as a way to de-naturalize the experience. Transparency has its place, of course, and not every work should be a meta-work, one that trips us at every turn, forces extremes of self-consciousness about the user experience. Such novelties grow tedious. But for interface design to come of age within a humanistic framework, to embody the subjective, interpretative, and user-dependent, historically situated conditions of knowledge will require some design work that has not yet been done. Likewise, for information visualizations to carry the inflected, affective, and multidimensional variety and subtlety of humanistic documents into a display for purposes of analysis, re-representation, mediation, or study requires an extension of the current techniques. Affective metrics are crucial, as are co-dependent variables, so that the scene that emerges does not stand in static relation to an always-fictive *a priori* given, but embodies the made-ness of interpretation as a practice. We'll return to this below.

Analysis of Assumptions

As stated at the outset, information visualization techniques borrowed from statistical social sciences, natural sciences, and the business of bureaucratic management work poorly for the humanities. The bad fit begins at the very moment of parameterization (using a metric to generate quantitative information). This is when the demonstration that data is *capta*, constructed and not given, is most evident and most critical, since the initial decisions about *what* will be counted and *how* shape every subsequent feature of the visualization process.

Content modeling (giving shape to the data through a database or other structure), builds layers of interpretative framing on top of the original decisions. Consider, for instance, the example of census data and the counting of members of different ethnic or racial groups, then consider the use of such quantified "data" in contrast to other parameters, such as education, longevity, income levels. The first set of decisions about how to determine ethnicity is highly problematic, but once it is used to generate information, the information separates itself from that initial process. The "data" appear to be self-evident, rather than being, as they are, the effect of a construct. In the humanities, the process of data creation often depends on an act of remediating knowledge and/or experience from analog form to digital form (transcription of texts, re-recording of film or vinyl sound to digital file formats). The continuous spectrum of analog phenomena is chunked into the discrete form of digital units. Once these exist, they are

highly tractable to all manner of analytical processing, but the information is at a remove from the original.

Born-digital humanities materials present themselves with an analog face. Texts, images, music, video, documents of all kinds may be made with a digital camera, keyboard, or program, but they are experienced in the same way as analog humanities texts. All digital formats are fungible, and the form of input or source does not have to determine the output. Sound signals can be output as light, text files as music, and so on, and any file can be subject to some kind of quantitative analysis that allows some variable to be charted against another (e.g., length of track and range of notes) in a graphic form.

Visualizations are all based on this sequence: parameterization (assigning a metric), quantification (counting or measuring what has been parameterized), and translating this captured, constructed information into a graphic. Visualizations are interpretative translations, but they pass themselves off as *images of data*. It is not too strong a statement, therefore, to say that *almost all information visualizations are reifications of mis-information*, and this is particularly true in the humanities, where the initial parameterization is often a radical intervention into and reductive extraction from an original artifact, corpus of documents, or other phenomena. Stated another way, visualizations are all *representations* (substitutes and surrogates) that pass themselves off as *presentations* (the information itself), as if the "form follows data" dictum of Tufte (2001) were accurate, and as if the artifact on the screen were an actual image of the data. The misinformation which is at the heart of data creation becomes amplified, and each act of display creates an artifact. This often presents viewers with a situation where they are reading the features of the artifact and taking incidental elements as expressions of a dataset. The dataset is already an extraction from a corpus, text, or aesthetic work, and a remediation. The image is another level of translation, further removed from the original act of creating *capta*.

Humanities documents and aesthetic artifacts are not "data" and they don't contain "data." They have to be remediated to become "data" – quantified and discrete information units, and in the process several issues come into play. For one thing, the plane of discourse and plane of reference are conflated. Data mining can be performed on any digital file, but only on the discourse – the literal information encoded in the ASCII coding or the stream of bits or other features of the file. The file does not have to be in a database or higher order of organization to be subjected to processing. Once the data-mining operation is run, it produces a "derivative" that then passes as "information" to be visualized. Not only is this information removed from and different from the source, but it is fraught with other problems that build on each other. Take a simple-seeming example of performing data mining on a text. A search algorithm can find every instance of a word, even create a keyword-in-context analysis, but the search is being performed on what we call, in semiotic terms, the plane of discourse. On the plane of discourse, the information of the digital file is simply code. But what can we capture of the plane of reference? The distinction of discourse/reference is the same as the difference between the telling and the told. We cannot capture the told because it is a performance, made anew in every instance, but also, it has no material instantiation in the substrate, only in the reader/viewer's experience. And yet, the told, not the telling, is where meaning is produced by

a reading or viewing of any text, image, artifact, sound or music file. The content of scientific texts, legal texts, or business documents is similarly constructed, though the aim of most of these is to create as close a connection between discourse and reference as possible to avoid ambiguities (as in contracts or treaties).

My point is that every instance of the word “amoeba” might be more closely related to every other across a body of scientific texts than every instance of the word “feeling” across even a single aesthetic text. And much of the data mining that leads to visualization ignores this fact, so that the process of word counting, or string searching, is based on a flawed method that conflates literal discourse and symbolic/interpreted reference. In an art-historical context, this would be the equivalent of counting instances of the color red across a collection of images without discriminating between symbolic and representational functions. The reds are not the same, and can not be counted the same way, put into the same category, or re-represented as data for visualization in a graph or chart, without monstrous distortion.

The next set of critical issues in using visualizations from outside the humanities involves the distinctions between quanta and qualia. Data are discrete, not continuous; they are explicit, not ambiguous; they are modular and bounded, not vaguely defined; they are sorted into categories that do not support contradiction; they are put into relations according to hierarchies, structures, or other ordering principles that have a very limited and highly defining set of qualities. In other words, “data” are antithetical to humanistic artifacts, they are fundamentally different in nature from the artifacts from which they are derived. Creating a humanities dataset, or culling quantitative or statistical information from humanities documents or corpora is problematic on many levels. Humanistic data are rarely discrete. A word is not reducible to its letters, for instance, and 20 instances of “the same” word are likely 20 different linguistic formulations in which the word is being given value and meaning through use. Some words have an enormous resonance, and inflect an entire paragraph, while others might serve a helper function, or a relatively passive role. None of this can be recorded by a string search, and again, counting words for data mining produces qualified results that need to be treated according. Not all instances of the same word mean the same thing (homonyms and puns are the most striking instances). Neither are words always explicit, and much of the impact of language use is in its subtlety, its implications, suggestions, nuances, and these are created by proximity to other words – a delicate situation is not delicate in the same way as a delicate piece of lace. Likewise, humanities data are rarely bounded – When does an event begin? End? What are its contributing factors and forces? What do we do to measure effects that exist without actual causes?

Phenomena in the world of humanistic experience and also in the varied and complex discourse fields of aesthetic documents do not lend themselves to representation within bounded, carefully delimited parameters. The metrics used to weight or characterize humanities phenomena are more complex than single value systems can represent, so a network diagram that shows “relations” among various nodes in a cultural system, among documents, authors, concepts, and so on, that is grounded in a single metric value for the edge-node relations, is painfully reductive. Relationships, whether among human beings or humanistic concepts, are dynamic, fluid, flexible, and changeable. They are always in flux, not static or

fixed. Humanistic phenomena are co-dependent with their conditions of production: Is a news event an effect or a cause, a representation or a driver, force or a reflection of the system of social conditions in which it participates?

Obviously these are questions that cannot be answered; they are posed to expose the limits of representational systems built on *a priori* or outset conditions of decision making on which subsequent analyses are made. By the time we are looking at a network diagram, a bar chart showing frequency of word use, or a scatter plot mapping dates of historical events, we are in deep complicity with the process whereby the artifact of visualization is mistaken for the phenomena it has (mis)represented.

As we have seen, the process of information visualization involves a series of distinct and dependent phases: extraction of data as information through parameterization and quantification, their remediation in forms and format that express the statistical, quantitative features of a dataset according to conventions. These have been adopted (almost exclusively) from fields that not only have nothing to do with the humanities, but that are often theoretically and methodologically antithetical to its core values and beliefs. The range of information visualizations that allows data to be created, displayed, and analyzed is comprised of charts, graphs, and diagrams whose pedigrees link them to statistical, managerial, and bureaucratic domains. The simple act of swapping contents from the humanities for those of the business or government offices is not sufficient to change the epistemological imprint. And therein lies the crux of the problem with using these techniques in the humanities. The difficulty is not just that the suit of clothing is ill-fitted, but that the body of evidence and argument on which it needs to hang is constructed in fundamentally different ways in the humanities than in fields whose relation to statistical processing is less problematic.

The basic graphical environment for visualizations seems counter to the principles by which interpretation works. The graphic conventions of information visualizations are almost all Euclidean. They are all structured on regular divisions of space and standard units of measure. When the value of a word, image, note, or other unit of meaning production varies by context, by inflection, by conditions of reception, its value needs to be represented in a graphical system that reflects these nuances and complexities. The content models that create structured data, in database or other formats, require that knowledge representations get sorted into named, identified boxes or fields. The imprint of that nomenclature over-determines the value of the information entered into the field, and thus the data structure becomes as powerful a part of the argument to be represented as the information it contains. A similar observation extends to the ways graphical interface scaffoldings create semantic value through their structuring principles.

The structuring principles for designing experience on the screen through the creation of a human–computer interface is driven largely by engineering sensibilities that prioritize efficiency, user satisfaction, and short-cycle results and rewards that are grounded in a model of the user as consumer, not producer or scholar. A “user” is not the same as a “subject” or “interpreter.” The premises of a consumerist model of the user experience are different from those that would inform an interpretative one. And yet, the graphical formats of screen display

that have become familiar conventions for reading, processing, understanding, and meaning production within networked environments come as much from commerce and entertainment as from literary or scholarly precedents. We now navigate, search, orient, and understand the materials we encounter in a networked environment according to a set of codes whose graphic conventions are only occasionally an object of critical study.

Towards Alternatives

Rethinking graphical display in humanistic terms would involve designing point-of-view systems, partial knowledge representation, scale shifts, ambiguity, uncertainty, and observer dependence into our visualizations and interface. These could be custom-built boutique projects, but it would be better to develop conventions designed to engage and expose principles of cultural conditions, hegemonies, and power structures.

To address this we have to address first principles: how to create methods for generating capta that have some of the characteristics of humanities documents and expressions. These have to embody ambiguity, complexity, fluidity, dynamic change, co-dependence, and other features of humanistic phenomena. If we take seriously even the most basic premises about meaning production from twentieth-century philosophy of language, that meaning is dependent on use, then modeling conditions of use is a prerequisite for mapping word frequency and usage over a corpus.

Beyond the basic modeling of phenomena, and the creation of mathematical and conceptual designs that are more appropriate to their specific character and quality, we face the challenge of creating conventions of visualization that are legible, rhetorically useful, and effective in communicating arguments about influence, development, slippage, and other interpretative dimensions.

Most conventions of interface act as concealments: they are devices for hiding what has been structured in the back-end of the site. No matter how simple (a mere HTML outline and hierarchy) – or complex (a vast and elaborate content management system customized to the last degree of granularity) – the back-end structures what the front-end displays. The display covers the design process, decision making, and all the many aspects of the content model in order to provide a means of access through the user experience. This is not pernicious, but it does have implications, and the ideology of document design, the rhetoric of the database, and the hegemonic force of information structures all play their part in the ongoing instrumentalization of knowledge regimes. We know this, and recognize it in architecture, text production, spectacle, and performance – and the critical study of new-media artifacts intersects here with digital humanities in useful ways. The point is to figure out how the workings of concealment act, what are their techniques, modes, and capacities, and what techniques for allowing interpretation to engage with the design of knowledge in database and digital formats might look like. That would be the beginning of a humanistic interface.

As for visualizations, they are the reification of misinformation, representations passing themselves off as presentations. All data is *capta*, made, constructed, and produced, never

given. What counts is what can be counted, what can be parameterized. So the first act of creating data, especially out of humanistic documents, in which ambiguity, complexity, and contradiction abound, is an act of interpretative reduction, even violence. Then, remediating these "data" into a graphical form imposes a second round of interpretative activity, another translation. An original question, how many of "x" are there in this text, becomes a statement in the bar chart graphic showing exactly how many "x's" were in these texts. But what are these presumed equivalences really based on? Are the various "x's" really the same? Data translations are fictions, distortions, misrepresentations, and then they become reified as visualizations, statements that pass as self-evident. The graphic shows just how many "x's" were in a body of texts and everyone forgets how they got there. Reading the image for its rhetorical force requires yet another set of critical exercises, for the scale of the graph, its own interior metrics, its regularized and static metrics. The use of visualizations from outside the humanities, from fields whose foundations are based on empirical observation, suggests that the presentation of interpretative analysis can be performed with the same tools as those of business management or census taking. What an impoverished point of view. One might as well write poetry with a table of weights and measures, or perform a sonata using a mechanical watch. The tools are too crude for the task. The challenges to the humanities are clear: construct systems of graphic designs to show humanistic values and methods within the visualizations and interfaces of our work.

References and Further Reading

Boole, G. 1854. *An Investigation of The Laws of Thought*. London: Macmillan.

Börner, K. 2010. *The Atlas of Science: Visualizing What We Know*. Cambridge, MA: MIT Press.

Cook, R. 1974. *The Tree of Life*. New York: Avon Books.

Drucker, J. 2009. *SpecLab*. Chicago: University of Chicago Press.

Drucker, J. 2014. *Graphesis*. Cambridge, MA: Harvard University Press.

Garrett, J.J. 2002. *The Elements of the User Experience*. Upper Saddle River, NJ: Peachpit Press.

Klapisch-Zuber, C. 2007. The tree. In *Finding Europe: Discourses on Margins, Communities, Images ca. 13th – ca. 18th Centuries*, ed. A. Molho, D.R. Curto, and N. Koniordos. New York: Berghahn Books, 293–314.

Parsons, T. 2012. The traditional square of opposition. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/square (accessed June 20, 2015).

Schmid, C. 1983. *Statistical Graphics: Design Principles and Practices*. Hoboken, NJ: Wiley.

Serna, S.P., Scopigno, R., Doerr, M., et al. 2011. 3D-centered media linking and semantic enrichment through integrated searching, browsing, viewing and annotating. *VAST'11: Proceedings of the 12th International Conference on Virtual Reality, Archaeology and Cultural Heritage*. Aire-la-Ville: Eurographics Association, 89–96.

Thrift, N. 2007. *Non-Representational Theory: Space | Politics | Affect*. New York: Routledge.

Tufte, E. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Watson, A. 1934. *The Early Iconography of the Tree of Jesse*. Oxford: Oxford University Press.

Notes

1 For best and worst visualizations, see: Visualising data, http://www.visualisingdata.com/index.php/2013/02/best-of-the-visualisation-web-january-2013/; EagerEyes, Visualization and Visual Communication, Robert Kosara, http://eagereyes.org/blog/2008/ny-times-the-best-and-worst-of-data-visualization; and many others.

2 The term is generally traced to the seventeenth-century economist, William Petty, whose book *Political Arithmetic* was posthumously published in 1690: http://en.wikipedia.org/wiki/William_Petty. See also the work of William Playfair, *An Inquiry into the Decline and Fall of Powerful and Wealthy Nations* (London: Greenland and Norris, 1807).