PJIM PARSONS JOURNAL FOR INFORMATION MAPPING

THE PARSONS INSTITUTE FOR INFORMATION MAPPING
68 Fifth Avenue
New York, NY 10011
212 229 6825
piim.newschool.edu

POLIS: Designing a Visualization Tool for the Research of Complex Sociopolitical Landscapes

MAYA KRISHNAN
JOSIAH OBER, PHD
MARK PYZYK, MA

KEYWORDS *Data visualization, data exploration, cartographic representation, information mapping, sociopolitical visualization*

DATE *Work on the application began in 2012 and was completed in February 2014. Work on the data set began in 2005 and continued through late 2013. The preparation of the data set was the product of substantial collaboration between many Classics scholars.*

URL http://polis.stanford.edu

ABSTRACT While sophisticated information visualization tools promise exciting transformations in scholarship, translating those capabilities into practical solutions can prove challenging. The "POLIS" project, a the product of the Stanford Classics Department, is a multi-year effort to create a visualization application capable of enabling research into the sociopolitical landscape of the Ancient Greek world. Researchers working on the POLIS project have digitized Oxford Classical Dictionary entries regarding over 2300 people and have united this data with information from the Inventory of Archaic and Classical Greek Poleis on over one thousand Greek cities. Through an assessment of the POLIS application, we demonstrate the way data visualization can facilitate original scholarly research employing large and highly rich data sets.

An interactive cartographic map that provides its users with an overview of the Ancient Greek world constitutes the core of the application. This map allows the user to combine over forty filtering criteria to generate customized representations of the Greek world. Basic statistical tools enable the calculation of counts and correlations. The user can then see different maps side-by-side in a "comparison view" that enables pattern-detection and the discovery of trends. Although the particular data sets visualized in this application are of special interest to classicists, the application design itself can generalize to other domains.

One of the broader questions we address is how to use the tools of computer science to address research needs in the humanities. This project is a case study in using sophisticated solutions such as D3 to generate data-visualization tools capable of solving problems for humanities scholars.

INTRODUCTION

The availability of digitized data about the ancient world outstrips Classics scholars' abilities to organize it in a manner that facilitates the recognition of patterns and trends. Moreover, many classical scholars recognize the value of testing data for significant correlations and trends, yet they are unfamiliar with even the most user-friendly of statistical tools (e.g., Excel). Making data useful for humanists depends upon the development

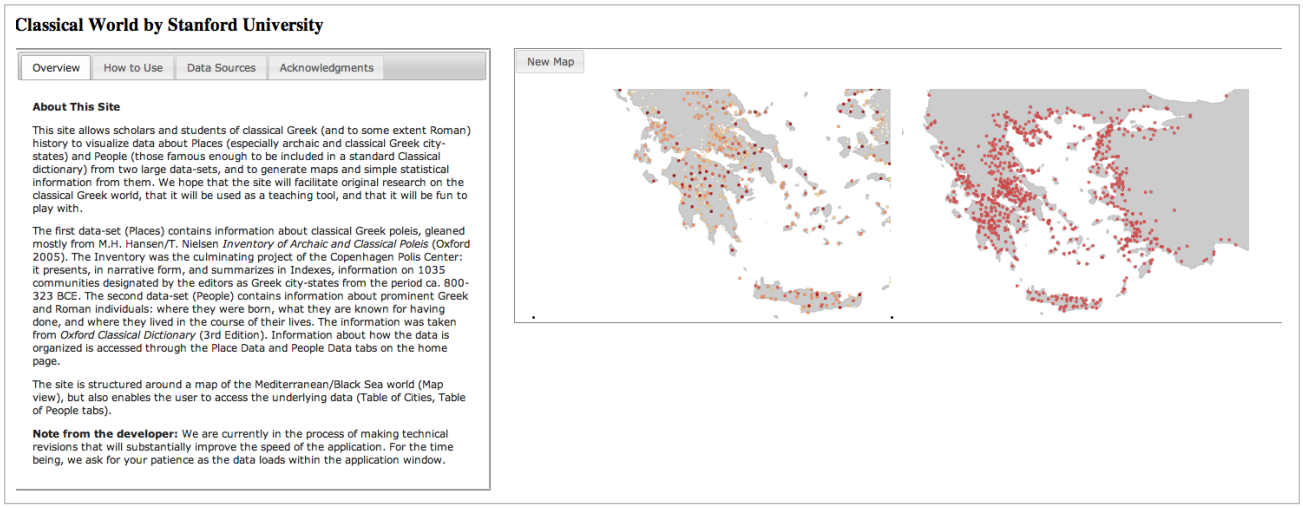

FIGURE 1: *Start screen for POLIS application. On left, introductory text and instructions. On right, map view panel. The user any existing map visualization to edit, or click "New Map" to create a new data visualization.*

of analytical tools capable of presenting complex information in an accessible and practical manner. However, designing suitable applications for humanists is challenging because of the diversity of research questions that scholars with different research agendas may pose. In order to make existing data more useful for scholars, applications must present it in a way that is at once unified and open-ended.

In this paper we describe POLIS (FIGURE 1, previous page), the result of a multi-year effort on the part of the Stanford Classics Department to find an appropriate means of presenting two rich data sets about people and places in the ancient Greek (and to some extent Roman) world. The location data is based on a digitized version of the 2005 Hansen/Nielsen Inventory of Archaic and Classical Greek City-States, which includes detailed data about 1035 Greek cities and is widely considered by classicists to be the most comprehensive overview of Greek city-states available to date. The Inventory was the product of a very large-scale data-collection and organization effort, and represents a uniquely valuable source of information for the size, location, and various aspects of the social, economic, and political history of all Greek city-states known to have been in existence by 323 BCE. Although the Inventory was conceived by its compilers as book-form "hard-copy" resource, quantifiable information on a number of variables were readily entered into a spreadsheet. This work was done by hand under the general direction of one of the co-authors of this article.

The data about individuals is based on the Oxford Classical Dictionary. As in the case of the Inventory the compilers conceived of the Dictionary as book-form, but much of the information was quantifiable and was entered by hand into a spreadsheet. The result of this process is a data set of over 2300 people associated with the Classical world. Classicists at Stanford categorized these people into six broader categories of occupation or "endeavors": philosophy, culture, economics, military, religion, and politics. These categorizations are intended to enable analysis of the distribution of human capital over the Ancient Greek world.

Because both the Hansen/Nielsen Inventory and the Oxford Classical Dictionary are widely known as reference works within the community of classicists, the data sets included in the application are of substantial intrinsic interest to existing scholarly communities and individual users can readily validate the information in the data sets. By placing these data sets in a form that facilitates comparative visual analysis, while simultaneously enabling direct access to the underlying data, the POLIS application can serve a diverse array of scholarly and educational agendas. Using data sets that have already gained a certain level of trust or recognition within a scholarly community can help increase the chance that the resulting visualization tool will find acceptance for employment in substantive research projects.

This paper describes a tool for the analysis of data associated with POLIS that suggests potential avenues for addressing the broader challenges in facilitating humanistic inquiry by use of large data sets. By allowing a user to generate interactive queries to create custom data maps of the Ancient Greek world, the tool opens itself to addressing a broad variety of research questions. The tool enables the user to compare different customized data maps within one overall view in order to facilitate the visual identification of patterns and trends. This feature also allows the user to generate and to test hypotheses about change over space and time. By focusing on enabling a wide range of comparisons within a single data set, this visual tool addresses the question of how to generate the maximum scholarly utility given a particular set of data.

The decision to craft a visual tool for these particular data sets is responsive to the idiosyncratic nature of humanistic data, and its demands for an extensive degree of design customization. By designing a tool that focuses on extracting the greatest utility from one particular data set our project can address the visualization challenges associated with achieving the best possible analytic tool for a particular time span and geographical region. We anticipate that the design choices we have made in the process of optimizing the tool for the Ancient Greek world will prove relevant to other problem domains as well.

THE BASICS: GENERATING VISUAL QUERIES

Maximizing access to the underlying data is the first priority of POLIS. The visual focus of the application is therefore an interactive map of city-states and their inhabitants, which is itself linked to a dynamically updating data table. This map allows the user to combine over forty filtering criteria to generate customized representations of the Greek world. For example, a user can choose to visualize only those cities in the Delian League whose populations exceeded a certain size, or only those cities that experienced over three civil wars. Similar filtering operations are available for the people, enabling users to craft different views of the social and the spatial aspects of the Greek world (FIGURE 2 and FIGURE 3, next page).

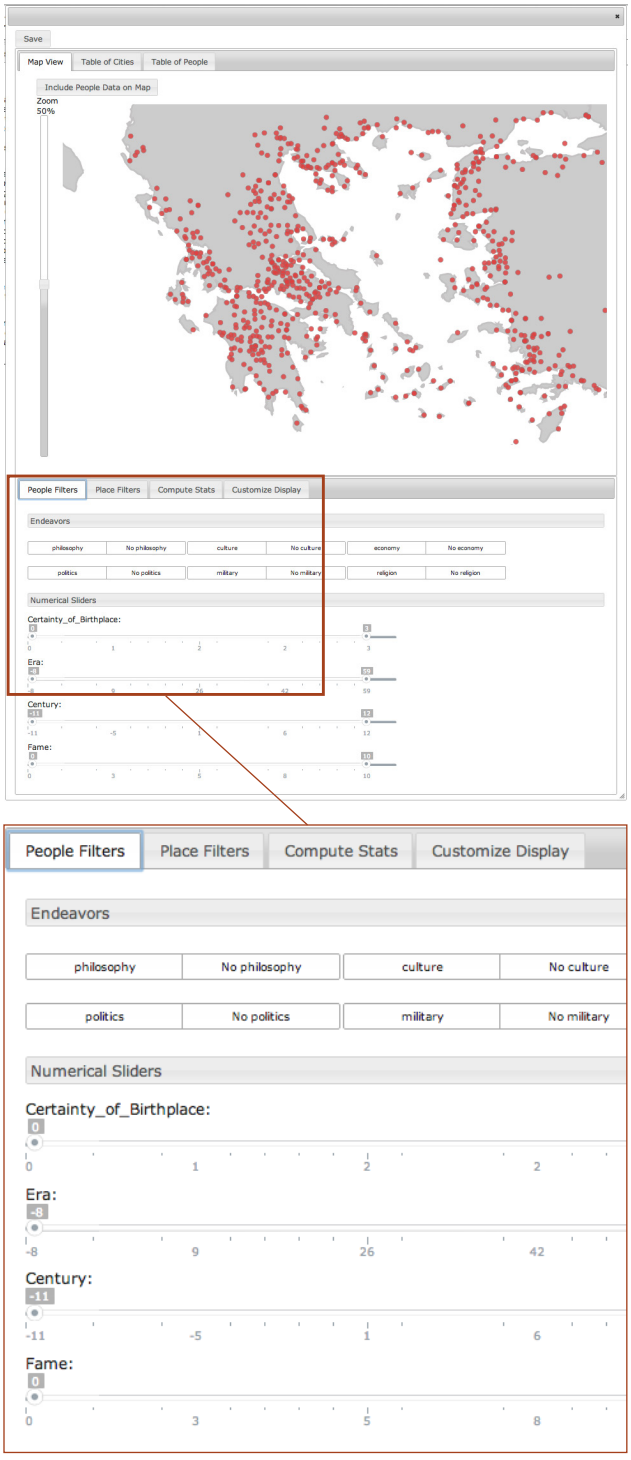

FIGURE 2: *Starting screen with map. User can toggle buttons and move sliders at bottom of the screen filter cities on map. Updates show up in real-time.*

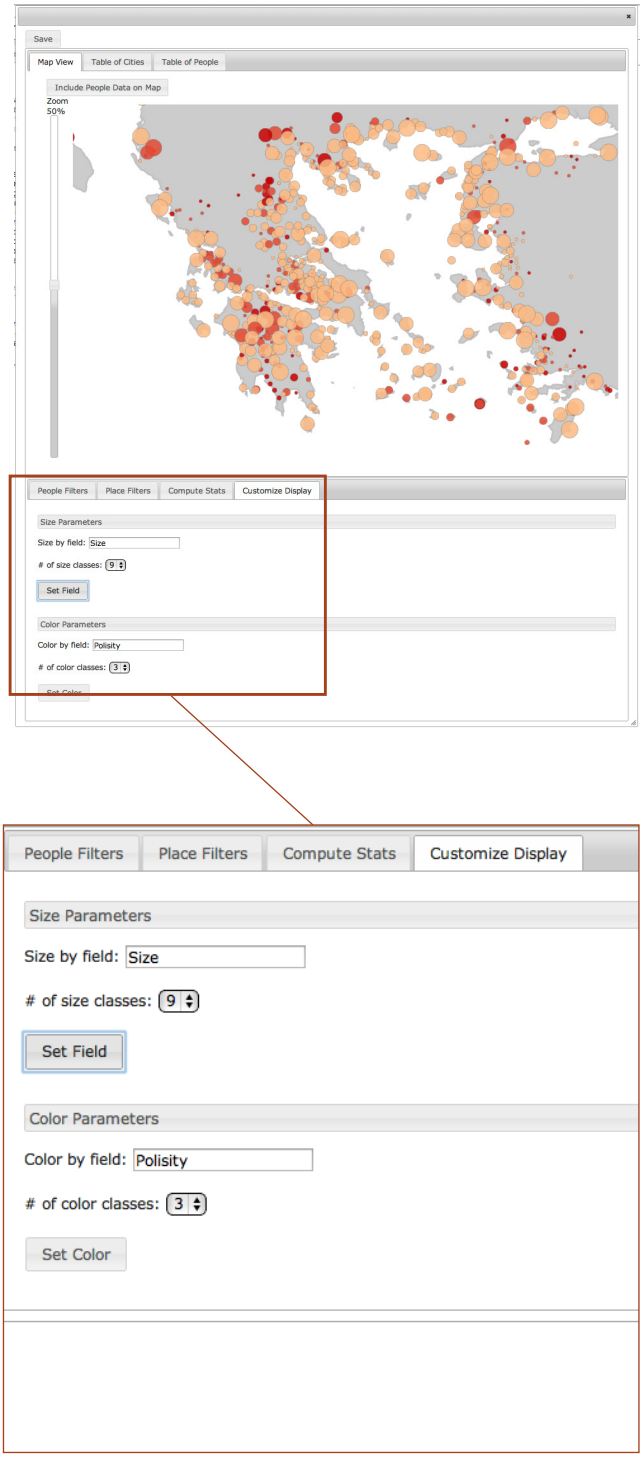

FIGURE 3: *Starting screen with data table. Filtering options performed on map simultaneously filter the data table. User can interact sort the data, search for particular data points, and calculate counts and correlations.*

The map initially plots all of the place data in geographical space on a cartographic map. The map updates in real-time as the user specifies query parameters. The user can choose to customize the data map by coloring or resizing the cities on the basis of any numerical field. This feature helps the user create representations that visually highlight patterns of interest to their particular project—for example, all cities could be resized based on the number of colonies they helped to found and then color-coded based on whether or not they were members of the Delian League, a military alliance of the Athenian Empire's tribute-paying subjects (FIGURE 4).

Place data and people data can be linked visually within the data map. The user can choose to resize places to render the radius proportional to the number of people who lived, worked, or were born there. The number of people associated with each place will change according to the filtering parameters the user selects for people. Thus, the user could choose to search for all of the people who are philosophers, and by including people data on the map, could resize each of the cities to reflect the number of philosophers who are residents. This feature allows users to visualize the distribution of human capital over space. (FIGURE 5)

The user generates queries to the database by interacting with checkboxes, radio buttons, and sliders. Generating

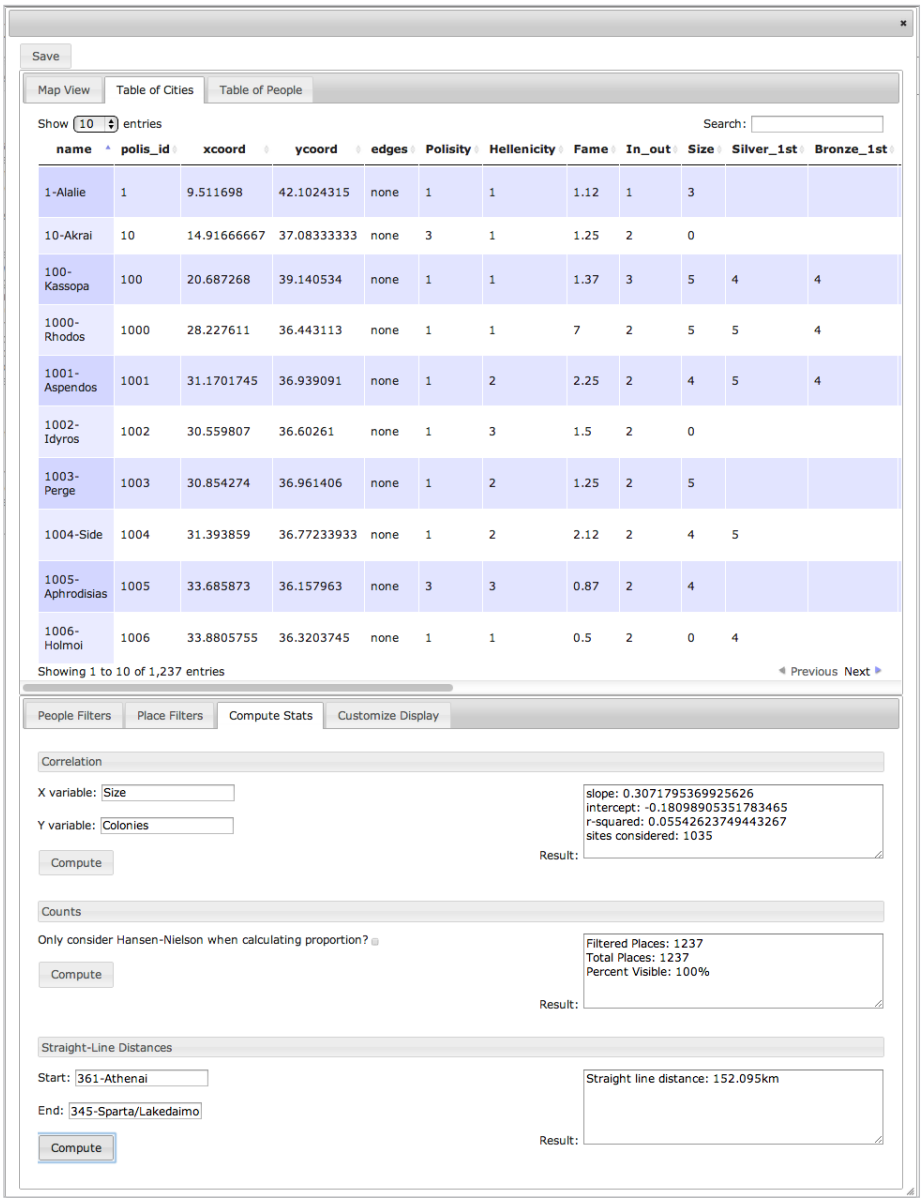

FIGURE 4:
Customize appearance of visualization in order to facilitate exploration of patterns in the underlying data. User can change the size and color of data points based on any numerical field.

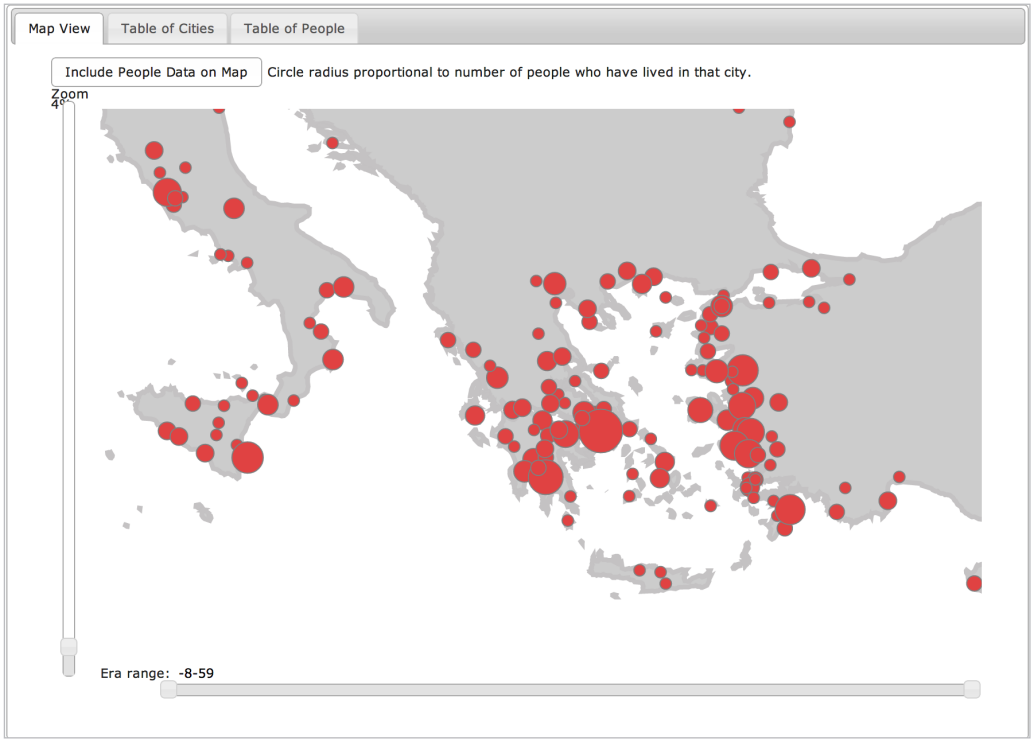

FIGURE 5:
Map with people data included. The radius of the city is proportional to the number of people associated with that city. Both people and places can be filtered by further queries.

representations is rapid and intuitive because the user can toggle the parameters of the query and see the impact of those changes in real-time on the map. Users generate queries on quantitative data by interacting with sliders. Quantitative data captured within this tool include factors such as size of territory, number of colonies founded, number of violent civic conflicts, and number of victors in the Panhellenic games. The visual tool also captures nominal data such as membership in the Delian League (yes/no) and presence of civic architecture (yes/no). The viewer can toggle these options using checkbox buttons. The real-time updates provide the user with immediate visual feedback regarding the effect that queries have on the data set as a whole. This basic functionality provides the foundation for the higher-level analysis POLIS supports through its multiple-view feature.

COMPARATIVE ANALYSIS OF VIEWS

POLIS is designed to facilitate comparative analysis within a single data set. Once users generate a custom representation based on their queries of the data, they have the option of saving that representation and creating further representations. These representations can be visually compared against one another within a central map "sandbox" that juxtaposes all of the user's representations for a given session with the POLIS tool.

This approach is inspired by Edward Tufte's idea of "small multiples,"[1] which is based on the notion that users can quickly spot differences between side-by-side representations that share a common background. This dashboard enables users to change filtering parameters for different maps and see the results of the operations next to one another. Allowing the user to create multiple customized versions of a single basic map graphic enables the detection of significant patterns and trends that can only emerge from comparative analysis (FIGURE 6).

Users may also choose to return to maps and edit them once they have saved them to the dashboard. The POLIS tool can thereby facilitate a cycle of exploration, hypothesis formation, and further exploration. This system was designed with the needs of humanities scholars in mind, whose open-ended style of inquiry demands research tools that are correspondingly supportive of exploration and revision.

This feature was also designed with many different kinds of research questions in mind. A user might choose to hold all variables constant except the date, thereby using the small multiples to show change over time. This use is particularly important to humanists, who are frequently interested in using data analysis to understand large-scale change from one historical era to another. Users might also choose to generate small multiples that compare develop-

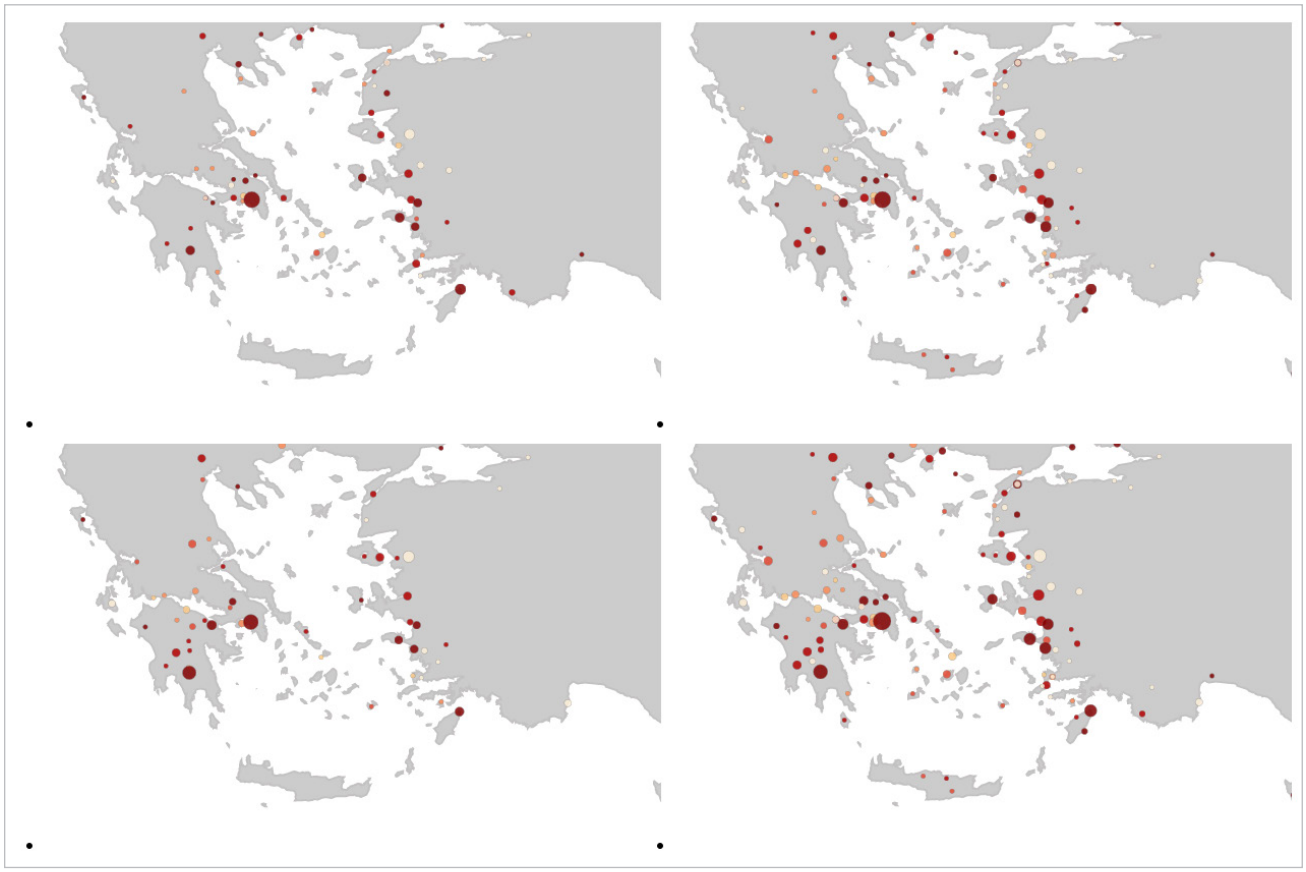

FIGURE 6: *Compare visualizations side-by-side. This comparative view shows the diffusion of human capital across the landscape of the Ancient Greek world. Top left shows cities with philosophers, top right shows cities with important cultural figures, bottom left shows cities with important political figures, bottom right shows all cities associated with people in the Oxford Classical Dictionary. Any of the maps can be opened in the full-screen view for closer examination.*

ment factors of cities that were, for example, members of the Delian League and cities that were not members, or to create small multiples that compare colonizing states to non-colonizing states.

By enabling the user to first customize the data map and then choose which customized data maps to juxtapose, POLIS enables visual analysis and exploration for a wide array of scholarly questions. Users may choose to return to any saved data map to further refine it through additional queries. This process enables one stage of exploration to give rise to a refinement of research questions and thereby provides material for further exploration.

BASIC COMPUTATION AND CORRELATION TOOLS

While we anticipate that most scholars using the site will focus on the visualization tools, others may be equally interested in the capacity to do basic computations—this is likely to be particularly useful for the Hansen/Nielsen Inventory data. By using the computation tools, scholars can easily discover, for example, exactly how many cities in the Delian League coined silver beginning in the sixth century BCE, compared to the number that began minting silver coins in the fifth or fourth century. Likewise, they may discover whether or not there is a statistically meaningful correlation between, for example, being an early minting state and being a relatively large state. These sorts of calculations are quite tedious to undertake on a spreadsheet, even for scholars who have the data and know how to operate simple spreadsheets. The application makes these sorts of calculations, which are important to many historical topics, very simple and intuitive. This simplicity gives scholars the means to do basic tests on a variety of more or less intuitive hypotheses at very low cost, and thus potentially allows for the discovery of surprising new relationships among the many variables captured by the application.

DISCUSSION

The process of designing this tool involved substantial collaboration between the technology specialist, who worked on the site and interface design, and the team of classicists who were familiar with the subject matter and underlying data. Although the work on the interface could proceed largely independently from work on the data, the POLIS site benefitted from close collaboration between the technical specialist and the subject matter experts. The result is an application that fuses the expertise of humanists with the tools of computer science, as opposed to a straightforward application of technical resources to a humanistic domain.

Validating the data sets with the input of classicists was an important step in the creation of the application. Teams of Classics graduate students hand-checked the accuracy of transcription for entries recorded from Oxford Classical Dictionary data set. Location information for the poleis was likewise hand-checked since the Hansen-Nielsen Inventory data, which offered only approximate locations for sites, was supplemented by more precise location information from sites such as Pleiades, Nomisma, and Google Maps.

The validation of information by subject-matter experts was crucial since the exciting increase in the availability of data on the Internet does not guarantee the reliability of that data. Because this application is ultimately intended to serve members of the Classics scholarly and educational community, it was important to ensure that the data used in the visualization process answered to the standards of members within those communities. Data corruption and errors, requiring a good deal of manual correction, arose with a surprisingly high frequency throughout the multi-year process of preparing the data set for the application. This situation underscores the need for technology specialists and humanists to remain in close contact as the application itself is prepared for release.

The final application presents detailed information about the data sources in a text sidebar. The ready accessibility of this information is very important for scholars who will be using the application. Since any information gathered about individuals and sociopolitical structures existing thousands of years ago is subject to uncertainty, it is important that scholars themselves know the sources of the data in the visual representations. This information will allow them to calibrate their own expectations for certainty accordingly.

The design process can itself yield insights that are useful to humanistic disciplines. The process of creating this visualization revealed key features about the underlying data set. Initially we had anticipated using network representations to display the relationships between people and places. However, upon providing visual links between individuals and their associated cities, we realized that individuals were clustering around three key population centers: Athens, Sparta, and Syracuse. This clustering reduced the usefulness of network representations, since attempts to provide a node-link structure to the data merely resulting in visual "clumping" around these sites.

This visual pattern could suggest a tendency on

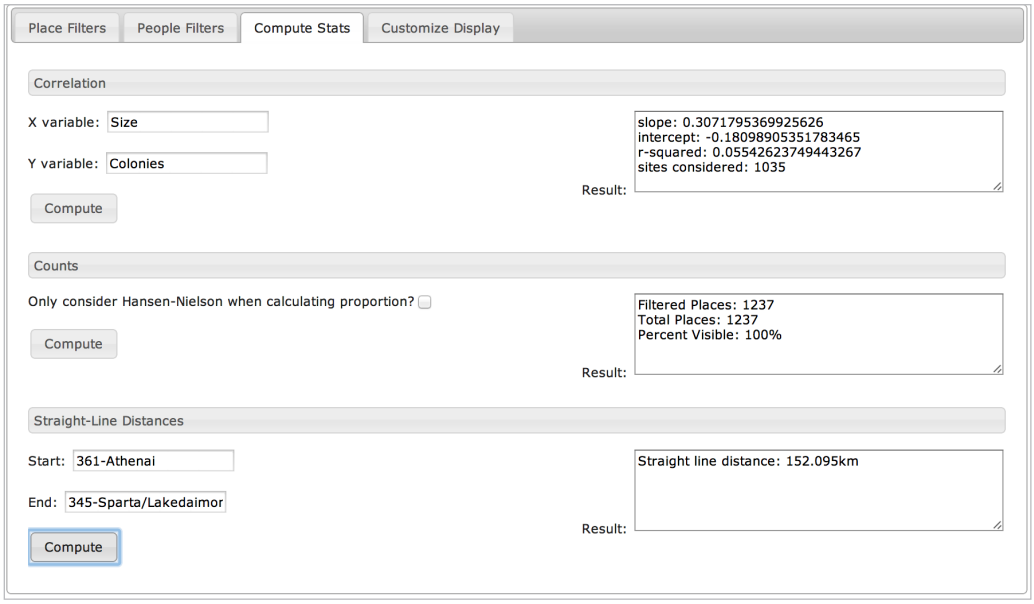

FIGURE 7: *Statistics panel. User can calculate correlations, counts, and straight-line distances directly in the application.*

the part of the Oxford Classical Dictionary to skew its focus toward large population centers, which would be consistent with the trends of mid-twentieth century classical historiography. Our visual identification of this potential bias is therefore of interest to classical scholars, and also suggests the need to expand our data set to include information on individuals beyond that provided by the Oxford Classical Dictionary. Unlike the Hansen/Nielsen Inventory, which is near comprehensive in its collection of data on individual city-states and unequalled in its scope, the Oxford Classical Dictionary lists entries for only a small fraction of the documented individuals who are known to have lived in the Greco-Roman era. Other, more comprehensive, compendia of individual ancient biographies exist, and could potentially be digitized along the lines of our current sample. If the Dictionary's tendency to focus on Athens, Sparta, and Syracuse at the expense of other city-states reflects mid-twentieth century historiographical biases rather than the actual state of the ancient Greek world, our visualization could help reveal the nature and extent of this bias.

FUTURE DIRECTIONS

There are multiple directions for further expansion and exploration on the basis of the existing POLIS tool. The first possible direction involves an expansion of the existing data set. Attempts at visualizing network linkages between places and people revealed that the Oxford Classical Dictionary does not provide sufficient information regarding the birthplaces, workplaces, and living places to support a network visualization. As noted above, the Oxford Classical Dictionary is far from comprehensive. Supplementing location data with information from other reference works could facilitate the creation of network visualizations to accompany the data map and data table. Classicists working on the project have indicated interest in expanding the data set to include the information from the significantly more extensive Neue Pauly encyclopedia in order to overcome the shortcomings identified as a result of this visualization.

name	Polisity	Hellenicity	Size	Victors	Proxenoi
	parallel	parallel	parallel	binary	binary
	1,3,1	1,5,1	0,5,1		
45-Sileraioi	3	3	0	0	0
87-Artichia	3	1	0	0	0
111-Zmaratha	3	1	0	0	0
152-Phylea	3	1	0	0	0
155-Therminea	3	1	0	0	0
246-Alion	2	1	0	0	0
252-Eupagion	2	1	0	0	0
277-Koila	3	1	0	0	0
284-Nestane	3	1	1	0	0
290-Phara	3	1	0	0	0

FIGURE 8:
Statistics panel. User can calculate correlations, counts, and straight-line distances directly in the application.

Another possibility is to apply the same visualization model to other historical eras and geographic regions. While the POLIS site was developed for the data set consisting of the Oxford Classical Dictionary, individuals, and the Hansen/Nielsen polis inventory, the POLIS team can create a duplicate visualization for any other data set simply by changing the spreadsheet of data loaded into the application. The code is non-proprietary and therefore available to any other research group, or researcher, interested in visualizing different spatiotemporal data with the POLIS site interface. This kind of technical adaptability is an important design consideration when producing academic visualizations, since it allows template solutions that are effective in one domain to quickly and easily assist researchers in another.

As scholars from the Stanford Classics Department and the broader community of classicists begin to interact with this tool, we anticipate incorporating their feedback into further iterations of POLIS.

TECHNICAL BRIEF

The project is a web application written primarily in Javascript. The D3 Javascript package was used for map design. Since the user can have many different maps open at once, multiple application states need to be simultaneously accessible to the application. The Backbone Javascript package was used to manage information about application states. Jquery was used for the design of the application interface. HTML and CSS were also used for interface design.

BIOGRAPHY

Maya Krishnan currently studies computer science and philosophy at Stanford University. She has served as the technology specialist for the Classical World project.

Josiah Ober holds the Constantine Mitsotakis Chair in the School of Humanities and Sciences. He writes on various topics conjoining Greek history, classical philosophy, and political theory and practice, and is currently developing projects on rational cooperation in the Greek world on the relationship between democracy and dignity.

Mark Pyzyk is a PhD candidate in Stanford University's Classics Department. His dissertation addresses the treatment of expertise in texts from the 5th and 4th century BCE Greek world. He is interested in the ways the codification of expert knowledge assists its transfer to new social and political environments.

NOTE

1 Edward M. Tufte, *Envisioning Information*. 1990.

BIBLIOGRAPHY

Pauly, August Friedrich von, Hubert Cancik, and Helmuth Schneider. 1996. *Der Neue Pauly : Enzyklopädie Der Antike*. Stuttgart: J.B. Metzler.

Hansen, Mogens Herman, and Thomas Heine Nielsen. 2004. *An Inventory of Archaic and Classical Poleis.* Oxford: Oxford University Press.

Hornblower, Simon, and Antony Spawforth. 1999. *The Oxford Classical Dictionary.* Oxford ; New York: Oxford University Press.