A web-based, interactive visualization tool for social environmental survey data

Amber Spackman Jones [a, *], Jeffery S. Horsburgh [b], Douglas Jackson-Smith [c], Maurier Ramírez [a], Courtney G. Flint [c], Juan Caraballo [a]

[a] Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, UT 84322-8200, USA
[b] Department of Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, UT 84322-8200, USA
[c] Department of Sociology, Social Work, and Anthropology, Utah State University, 0730 Old Main Hill, Logan, UT 84322-0730, USA

ARTICLE INFO

ABSTRACT

Understanding human motivations and actions related to environmental problems is central to modeling complex, human-natural systems. However, social science survey data on environmental issues are often presented in relatively static reports and figures and are not easily accessible for participatory deliberation. Federal data sharing mandates motivate innovative data visualization and sharing mechanisms. We developed an open-source, web-based Survey Data Viewer as a visual interface to explore quantitative social science survey data. We used the Python Django web framework and the D3.js visualization library to create and deploy the tool. The Viewer was implemented using a water-related survey administered to a large, random sample of Utah adults in public venues. The Viewer allows users to visualize question responses based on demographic variables with percentages and mean response levels. We developed a standardized template for encoding survey data and metadata that permits the generalization of the tool to other similar surveys.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Software availability

Name of software: Survey Data Viewer
Developers: Maurier Ramírez, Juan Caraballo, Amber Spackman Jones, Jeffery S. Horsburgh
Contact: jeff.horsburgh@usu.edu
Year first available: 2015
Hardware required: Web server capable of hosting a Python/Django web application
Software required: Python, Django Web Server
Software availability: All source code and documentation for the Survey Data Viewer can be accessed at https://github.com/UCHIC/SurveyDataViewer
Cost: Free. Software and source code are released under the New Berkeley Software Distribution (BSD) License, which allows for liberal reuse of the software and code.

1. Introduction

Recognition is growing for the importance of incorporating social science data in studies of complex coupled human-natural environmental systems (Hiwasaki and Arico, 2007; Pickett et al., 2007; Braden et al., 2009; Wagener et al., 2010; Sivapalan et al., 2014; Hale et al., 2015). Information on public environmental perspectives and reported use of natural resources is essential to modeling future environmental and resource conditions and informing environmental management and decision-making (Fath and Beck, 2005; Morehouse et al., 2010). Social science surveys bring the ordinary knowledge and behaviors of citizens and environmental actors into scientific understanding and participatory decision-making arenas (Coenen et al., 2012). While a growing number of initiatives have been undertaken to collect social science survey data as a component of integrated environmental studies and large-scale environmental observatories (Redman et al., 2004; Curtis et al., 2005; Braden et al., 2009), access to and interpretation of social science datasets have historically been limited primarily to the researchers who originally collected the data (Ryssevik and Musgrave, 2001; Fry et al., 2012). This paper describes an open-

* Corresponding author.
E-mail address: amber.jones@usu.edu (A.S. Jones).

source platform for presenting quantitative social survey data in an interactive format so that stakeholders can participate in the exploration and understanding of the data.

The rate of data sharing among researchers or between researchers and non-academic audiences is relatively low in the social sciences (Freese, 2007; Tenopir et al., 2011; Healy and Moody, 2014). In particular, social scientists surveyed by Tenopir et al. (2011) report lower levels of electronic data sharing and less satisfaction with available tools for data sharing and publication than for other scientific disciplines. Furthermore, privacy concerns and requirements for protecting identities of human subjects often restrict the interaction that a wider audience can have with these data. Because of this, social science datasets are often communicated using summary plots and reports that are not interactive, present only a subset of the results, and are limited to the insights extracted by the original data collectors (Hamilton, 2006; Wexler, 2014). This is in contrast to baseline biophysical environmental data (e.g., environmental observations made by sampling or sensing the biophysical environment) for which numerous open repositories for sharing data have emerged and for which there are strong initiatives and requirements for making datasets widely available (e.g., Horsburgh et al., 2009, 2011; Zaslavsky et al., 2011; Lehnert et al., 2011; Jones et al., 2015). Federal mandates and scientific journal requirements are helping promote data sharing and publication, and recent technological advances make data sharing more approachable (Healy and Moody, 2014). Low levels of data sharing by scientists may be due, in part, to logistical barriers, including a perceived paucity of mechanisms or tools for publication and communication of the data (Tenopir et al., 2011). New methods are needed for sharing social science datasets in ways that protect identities of human subjects of research while allowing environmental managers and broader audiences to access and interact with these data.

There are multiple types of social science data relevant to environmental studies, including information collected using censuses and other secondary data, surveys, key informant interviews, focus groups, and field observations (Braden et al., 2009, 2014; Corti, 2012). As a first effort in this space, we focused on the challenges and opportunities associated with visualizing and sharing datasets derived from quantitative responses to survey questions, which is an important class of social science data.[1] When attributes of respondents are included in survey data, such as demographic information, locality, and other environmentally relevant characteristics, perspectives can be aggregated and disaggregated to understand the commonalities and heterogeneities within local societies. These insights are key to understanding human dimensions of environmental issues, such as identifying locations and groups dominating resource use or varying levels of public support for environmental policy and management options. Tools are needed that allow stakeholders open access to explore survey data and ask their own questions about how particular social groupings or localities relate to perspectives and behaviors.

In this paper, we describe a web-based software tool called Survey Data Viewer for presenting dynamic visualizations of quantitative survey results to a broad audience. Instead of using static plots and reports that present limited aggregations and permutations of survey results, we developed a web-based software tool that enables users to interactively explore multiple dimensions of the data.

The Survey Data Viewer allows users to select survey questions and visualize aggregate responses in terms of the percentage of respondents falling into each response category or as a mean response score for that question. Users can also visualize results disaggregated by characteristics of respondents (e.g., by age, gender, etc.). Furthermore, the Survey Data Viewer permits simultaneous view of multiple questions, the ability to view whether observed differences are statistically significant, and a map viewer that aggregates responses based on the zip code within which respondents live. As a demonstration use case, the Survey Data Viewer was implemented for a major random sample survey of adults that assessed perspectives about water resources concerns and issues as part of the iUTAH (innovative Urban Transitions and Aridregion Hydro-sustainability) interdisciplinary research project (http://iutahepscor.org). We present the Survey Data Viewer in the context of this example survey and describe its architecture and implementation, which was designed in a general way to support visualization of any quantitative social science survey data. While we designed the Survey Data Viewer for environmentally-related social survey data, it could be applied to surveys in other sectors or domains.

Section 2 provides background on current methods for the communication and visualization of social science survey data. In Section 3, we present the context of our case study. Section 4 describes the software implementation for the Survey Data Viewer, including options and features as well as how the tool can be reused. Section 5 discusses the effectiveness of the Survey Data Viewer as applied to the case study and opportunities for improvement. Finally, we summarize our work in Section 6.

2. Background

Many studies of coupled human-natural systems are intended to provide data and tools to inform and assist resource managers in their efforts to address environmental challenges. Since human behaviors are often at the root of environmental concerns, efforts to quantify the perceptions, attitudes and behaviors of resource users can be important touchstones for decision-makers. These data are often gathered using surveys.

In the social sciences, sample surveys are a major data collection tool that can facilitate inferences about the characteristics of a larger population and enable analyses of relationships among respondent attributes, perceptions, attitudes, and behaviors (Fowler, 2013). Surveys can be administered using many modes, including mail, phone, Internet, and public intercept methods (Dillman et al., 2014). Survey data may be quantitative or qualitative. Quantitative survey questions capture responses in units that can be expressed using nominal, ordinal, or interval/ratio measures. Qualitative survey questions allow participants to provide openended responses that are recorded as free text, which are then processed and categorized by the researchers. The Survey Data Viewer described here was designed specifically for communicating quantitative survey data. Qualitative data could be converted (e.g., by categorizing/coding more detailed responses and anonymizing respondents if necessary) for incorporation into the Survey Data Viewer.

2.1. Communication of survey data

Widely available tools exist for developing, disseminating, and collecting responses for surveys. However, these tools generally do not provide access and functionality for broad audiences to explore patterns in resulting survey data (Wexler, 2014). Analysis of survey responses and reporting of results are typically controlled by survey researchers who utilize specialized software packages (e.g.,

---

[1] While most efforts to visualize social science data have, to our knowledge, emphasized quantitatively measured variables, there have been innovative efforts to create visualizations of qualitative data using grounded theory methods (Knigge and Cope, 2006).

Statistical Analysis System (SAS) or Statistical Package for the Social Sciences (SPSS)) to describe univariate results and analyze statistical relationships among responses to survey questions. Visualizations of social science data are less widely used than in the physical sciences (Healy and Moody, 2014). Graphical presentations of survey results are often produced using Microsoft Excel or specialized functions as part of the statistical packages listed above. Static plots and selected results are then typically communicated to stakeholders and the public via documents and reports (Hamilton, 2006). While the process can be effective for drawing conclusions from a completed survey, it is not easily reproduced, does not facilitate hypothesis testing, does not lend itself to the easy addition of new survey results, and precludes exploration by audiences wider than the principal investigators (Ryssevik and Musgrave, 2001; Schwabish, 2014; Wexler, 2014). In our development of the Survey Data Viewer, we sought to address these shortcomings and expand the possibilities for communicating quantitative survey data.

2.2. Visualization of survey data

Visualizations of data communicate complex information to enable quick and easy understanding and can emphasize static forms and explanatory functionality or facilitate more interactive exploration through rapidly emerging tools and technologies (Healy and Moody, 2014; Schwabish, 2014). For quantitative survey datasets, visualizations are most often used to describe response profiles for single survey items using measures of central tendency or the distribution of responses across categories. Perhaps the most frequently used visualizations for categorical survey data are simple bar charts or pie charts that provide a graphical representation of the proportion of responses in discrete categories. More complicated visualizations may use more elaborate bar charts to illustrate cross-tabulations or correlations of more than one survey question (Ryssevik and Musgrave, 2001; Wexler, 2014). Quantitative survey data can also be represented by graphics that highlight measures of central tendency (e.g., mean or median) of the responses as well as indicators of statistical variance or dispersion (e.g., error bars or box plots).

We sought to contribute to the more interactive forms of data visualization to promote participatory exploration of survey data in line with developments highlighted by Schwabish (2014). The Survey Data Viewer falls into the "exploratory" and "interactive" sections of the visualization taxonomy presented by Schwabish (2014) and articulated by Kirk (2013), which is in contrast with typical visualizations of survey data that use "explanatory" and "static" plots to reinforce the interpretation of the creator. In this way, stakeholders accessing the data can ask their own exploratory questions and visualize the corresponding summary information rather than understanding the data only through the filter of a scientific researcher's perspective. As the onus of interpretation is on the consumer rather than the creator, users may discover unique insights based on their context (Kirk, 2013). This is in line with the recommendations of Healy and Moody (2014) that visualizations themselves need not provide the answers to our questions, but should be incorporated into the exploration and confirmation process of the social science workflow. Furthermore, the exploratory nature of the Survey Data Viewer makes it suitable for use in conjunction with other types of social science data in the iterative process of grounded theory development suggested by Knigge and Cope (2006).

A common error in visualizing quantitative data is to utilize counts instead of percentages to disaggregate patterns of responses in one survey question by characteristics of the respondent (Shah and Hoeffner, 2002). When respondents are not equally distributed across characteristics, the visual images presented by count data can be cognitively misleading (i.e., because the total count of one group may be larger than a comparison group, though the percentage of the group is not necessarily larger). Another important consideration is that graphical cues utilizing color and symbol size to illustrate the direction, intensity or relative importance of statistical relationships can assist in communicating accurate messages from quantitative analysis (Tufte, 1983; Healy and Moody, 2014; Schwabish, 2014). For the development of visualization options in the Survey Data Viewer, we worked to use simple, graphical representations of survey responses that rely on percentages and that provide intuitive and visually appealing methods for exploring the quantitative results. For example, we implemented views and features that allow users to visualize patterns of responses for individual survey questions and graphics that disaggregate response patterns by characteristics of survey respondents, including spatial context. We also included functionality to assess statistical significance in relationships between variables.

More general purpose, proprietary visualization software (e.g., Tableau: http://www.tableau.com/) and academically developed software programs (e.g., NESSTAR: http://www.nesstar.com/, SDA: http://sda.berkeley.edu) exist that provide visualization tools that researchers can use to illustrate survey results. The typical workflow for this type of software involves using a desktop software application to connect to data stored in internal spreadsheets. Researchers can use tools included with these programs to format data and develop customized visualizations, and graphics can be updated as new data are added. Because these software programs are general purpose and do not prescribe a particular file format or data model for the underlying data, complex steps may be required to format the data and develop customized visualizations. Some general purpose visualization software programs do enable deployment of visualizations to a web server from which users can then interact with the views defined by the creator. However, while permitting a great deal of flexibility and power in visualization, the financial cost, complexity, and learning curve associated with the software may be prohibitive for many research groups. Furthermore, some proprietary software programs only permit access by those holding software licenses, restricting broad or public dissemination of results. Additionally, few programs allow users to manipulate the visualizations, and those that do require a higher level of expertise in the underlying software interface and statistical methods than is common for a lay user. Some of these programs are designed to be repositories for complex, large-scale surveys that include numerous options for analysis and a complex interface, limiting the use and interpretation to social science experts. Our design for the Survey Data Viewer was driven to fill a niche for a freely available and open-source option for hosting quantitative survey data and providing visualizations able to be manipulated and explored by untrained users.

3. Utah Water Survey

The Survey Data Viewer was conceptualized to satisfy needs for storing, managing, visualizing, and disseminating to non-specialists the data from environmentally focused social science surveys conducted as part of the iUTAH project. iUTAH is an interdisciplinary research program focused on water resources within Utah, particularly across landscapes that transition along mountain-to-urban gradients with population centers undergoing varying rates and types of urbanization. This region faces growing challenges to meet the water needs of a rapidly growing population in the face of predicted land use and climate changes. Utah residents currently use 167 gallons of water per capita per day (GPCD),

double the national average and second highest in the nation (Maupin et al., 2014). Perceptions of water supplies and concerns about water shortages are factors that influence water conservation behaviors, and also shape attitudes and behaviors related to other water resource issues (like water quality and flooding). As part of the overall effort, iUTAH researchers have conducted scientific surveys of the general public to measure current attitudes and behaviors about water resources, and to compare levels of awareness and concern across diverse households, landscapes, and settings.

One of the iUTAH project's objectives is the storage, sharing, and dissemination of biophysical and social datasets. Project efforts have generated baseline biophysical datasets and a sophisticated set of associated web-based tools for data sharing and visualization (e.g., Jones et al., 2015). We desired to develop a parallel interface for baseline social science datasets, including public surveys designed to assess the drivers of water use behavior and decisions. We needed a tool to provide access to survey datasets for a broad audience having varying levels of technical expertise. Target users included researchers from various scientific domains (e.g., social scientists, hydrologists, ecologists), municipality and state government agency partners, educators, the survey respondents themselves, and the general public.

As a particular case of survey data, we used the iUTAH "Utah Water Survey," which was implemented by participating researchers from several Utah institutions of higher education. The objectives of the survey were to document how a representative cross-section of Utah's adult population thinks about water issues. The survey included three core blocks of questions: perceptions of the adequacy of local water supplies, perceptions of the quality of local water resources, and concern about a range of water and non-water issues. A number of additional questions captured information about respondents' familiarity with water cost, lawn-watering behaviors, participation in water based recreation, and demographic attributes. Supplementary material to this paper includes a document with a description of the dataset as a whole, a document containing the complete survey instrument, and two data files containing the results and an associated codebook (see Section 4.3).

The survey was administered by trained teams of students from six Utah universities, who randomly selected adults at grocery store entrances across a wide range of urban communities in Utah. Survey teams invited respondents to complete the survey on electronic tablet computers using Qualtrics Offline Survey Application Software (http://www.qualtrics.com/). At the time of writing, the survey had been implemented at 30 stores across all major urban counties in the state. Almost 18,000 adults had been approached, providing 6881 useable responses (a 40.7% response rate after eliminating responses from people under age 18 and non-Utah residents). The survey is ongoing, and the number of respondents is growing; however, analysis of age and gender characteristics of respondents at the time of this writing closely match proportions in the 2010 Census of Population for the state. This survey was designed to contain no personally identifiable information, so publicly sharing the results would not reveal the identity of individual respondents.

4. Software implementation

Our goal in developing the Survey Data Viewer was to present simple, interactive, and easily accessible visualizations of quantitative social science survey results. The following requirements motivated the implementation:

1. An open-source, reusable, graphical user interface (GUI) for use on any computing platform

2. Ability to implement viewing for multiple, unique surveys as well as append new results to each survey as new data become available without modifying the software code
3. Visualization of responses to survey questions by percentage of respondents in each response category and mean response score
4. Disaggregation/faceting of responses based on collected demographic variables (e.g., by age, gender, education level, etc.)
5 Geospatial visualization of data based on respondents' residential zip codes
6 Calculation of chi-square values and statistical significance of cross-tabulated data

The Survey Data Viewer was implemented as a web application with a web browser-based GUI. The architecture of the web application consists of a user interface layer, a web framework layer, and a data storage layer (Fig. 1). In the following sections, we describe each of these layers, their key components, and functionality. Instructions for deployment of the Survey Data Viewer are documented in the Github repository.

4.1. User interface

The user interface layer was implemented primarily using HTML5 and JavaScript, which function in any modern web browser. This makes the Survey Data Viewer cross-platform compatible for users on any device with a web browser. For the user interface design and styling, we used the Bootstrap framework (http://getbootstrap.com), which provides an effective and responsive user experience along with consistency between all browsers. Symbology for visualizing survey results was implemented using the D3.js JavaScript library (http://d3js.org), a flexible and powerful open-source tool for creating interactive data visualizations using hypertext markup language (HTML), cascading style sheets (CSS), and scalable vector graphics (SVG) technology.

The user interface consists of two web pages for selection and exploration of surveys. The first displays the set of surveys for which the Survey Data Viewer has been deployed (Fig. 2a). Each survey includes a link for the user to select a single survey to explore and display results. This page also permits viewing a copy of the original survey instrument and an "About" page that can be customized to provide information about the background and purpose of the survey. The second web page is the results viewer interface for exploring the results of the selected survey. As shown in Fig. 2b, the Results page consists of a selection facet panel on the left side of the screen listing the survey questions, and the main visualization panel that presents the results. Buttons in the Results panel allow users to select visualization options. When a survey is selected, its codebook (i.e., a coded representation of the survey instrument) is automatically parsed and used to populate the facet panel with the survey questions. When a question is selected from the facet panel, the data file is parsed to obtain the results for that particular question, which are then displayed in the main visualization panel. The content and format of these files are described in Section 4.3. Specific visualization features and functionality are described in Section 4.4.

4.2. Web framework layer

The Survey Data Viewer uses the Python Django web framework (https://www.djangoproject.com) as a mediator between the front-end visualization layer and the underlying web server. Although the Viewer could have been implemented using any web framework, we chose to use Python Django because it is freely available, open-source, and supports rapid and straightforward development
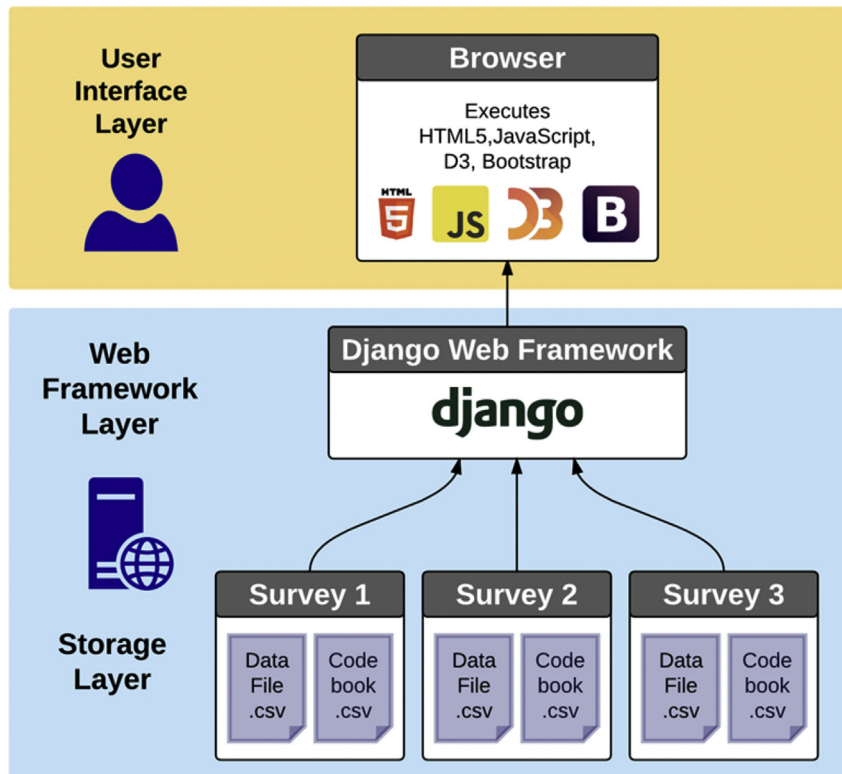
Fig. 1. Architectural diagram of the Survey Data Viewer web application.

with interchangeable and scalable components. Because it is Python based, it can run on multiple server platforms (e.g., Windows or Linux) as well as a variety of web servers (e.g., Apache, NGINX, and Internet Information Services (IIS)), providing multiple options for deployment. For this application, we deployed the Django web framework with a Microsoft IIS web server using the Helicon Zoo module (http://www.helicontech.com/zoo/).

4.3. Storage layer

By using Django's Object-Relational Mapping functionality, the data for each survey can be stored in any relational database management system (e.g., SQLite, MySQL) or as basic text files. We chose to store the survey data as comma separated (CSV) text files for simplicity of implementation and to facilitate appending new results as they became available. The Survey Data Viewer was designed to be generalizable for multiple surveys consisting of different questions. For each unique survey, the tool requires two CSV files that provide the underlying data: a data file and an associated codebook file. For data collection efforts that are ongoing, new results may be appended to the data file and new questions added to the data file and the codebook at any time. Because the codebook and data files are not hard coded into the web application, but rather parsed when the application is launched, the visualizations automatically update after the underlying data files have been updated. These files, the survey instrument document, and the general survey information can be easily added and updated to the web application using Django's default administrative functionality. A survey data file and codebook for the iUTAH Utah Water Survey are contained as electronic supplements to this paper.

The data file is a table that consists of a single column for each survey question (variables) and a row for each respondent (cases).

The values in each of the fields are the associated individual responses to each survey question. As the Survey Data Viewer was designed for quantitative survey data, these response values should be numeric. If free text responses were permitted in the survey, they must first be re-coded to correspond to a numeric entry. Otherwise the question should be omitted. Text strings may be used if the response corresponds to a disaggregation category (see Section 4.4.6). In this case, the number of occurrences of the same text string is calculated. This data file format is consistent with output options from most online or desktop survey data collection and analysis software (e.g., Qualtrics). For some visualization packages, it is necessary to reshape raw survey data, transposing it from a cross-tabulation format and splitting cases into separate rows for each variable (Wexler, 2014). This additional step is not needed for the Survey Data Viewer.

The codebook, or metadata file, provides the interpretation and substantive meanings of the numeric codes associated with each variable in the data file and defines the 'type' of each variable. Each row in the codebook file represents a question in the survey to be displayed in the Survey Data Viewer and must correspond to a column in the data file. The columns in the codebook file provide the necessary information for display of the questions in the Survey Data Viewer. The columns, their format, and a specification of coding for features are described in Table 1. A general template for the codebook file is shown in Table 2 and can be accessed on the Survey Data Viewer GitHub site. This template should be followed to develop metadata for any survey to be visualized with the Survey Data Viewer.

4.4. Visualization features and options

The Survey Data Viewer provides three main views for baseline visualization of survey results: a percentage view, a mean result
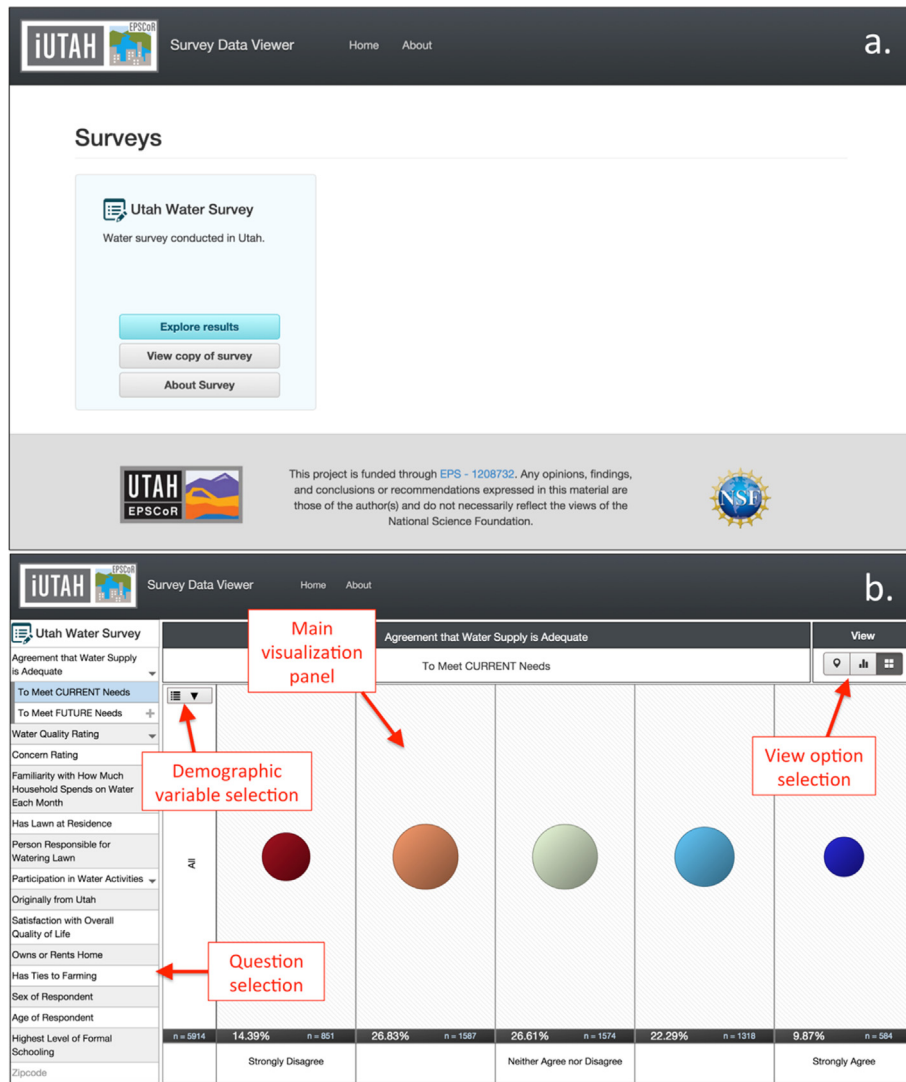
Fig. 2. Survey selection webpage (a) and results webpage with features highlighted (b).

view, and a heat map view. On the Results pane, users may toggle between these views for the questions to which they correspond. Additional features include the disaggregation of results within the views by demographic variables, display of a statistical significance indicator, color schemes based on question type, and orientation/instructions. Fig. 2b shows the locations of these view options and several features on the Results pane.

We selected these views and the corresponding visualization elements based on availability in existing software libraries, applicability to quantitative survey data and our study survey, and consensus within our research group following unstructured experimentation with users. In development of visualizations, we followed basic principles to keep plots simple and uncluttered (Tufte, 1983; Schwabish, 2014; Healy and Moody, 2014). We also sought to provide multiple mechanisms for viewing data as recommended by Knigge and Cope (2006) and to permit overview of data along with the ability to drill down to more detailed views as recommended by Schwabish (2014). As noted by these authors, effective visualization is a subjective process, and there is no prescriptive, generic formula for its successful implementation. The elements and attributes (position, size, shape, color) that we selected may not be the only way to visualize these data, and other

visualization types may be equally effective. In fact, of our visualization options, the map view (chloropleth) is the only one to appear in a survey of visualization types by Heer et al. (2010).

4.4.1. Percentage view

Percentage view shows the percent distribution of survey responses across response categories for each variable or survey question (Fig. 3a). The grid in the percentage view displays questions on the y-axis and response categories on the x-axis. When a question is selected, the results pane shows circles that are sized to visually convey the relative proportions of respondents in each response category. Text is also displayed to give the total number and numeric value of the percentage of respondents in each response category.

4.4.2. Mean result view

Another option for visualization is the mean result view (Fig. 3b). The axes are implemented similarly to percentage view. A slider on the scale of possible responses is used to illustrate the mean score for a question, which is determined by calculating the mean of the numeric responses associated with the question. The numeric values for the mean are not shown in the view, as the

Table 1
Survey Data Viewer codebook template description.

Codebook column	Encoding description
Variable	Must correspond to the labels of columns in the Data File. For a question to be displayed in the Survey Data Viewer, it should be prefaced by a 'Q' (e.g., Q1, Q7, Q10, etc.). Other 'demographic' variables can be included to disaggregate response characteristics of respondents but are not displayed as questions. For grouped questions, each sub-question must have its own row. Questions of this type retain the same number and are differentiated with letters (e.g., Q2a, Q2b, etc.).
VariableLabel	Contains the text displayed as the title for each question in the Survey Data Viewer. For nested questions, the same parent question/variable label must be repeated for the row corresponding to each sub-question.
SubVariableLabel	Contains the text displayed as labels for each sub-question. Should be left blank for questions that are not nested.
ValueLabels	Contains text labels related to numeric response values in the data file. An " = " associates a numeric code with its text label. A "; " separates each possible response. All responses may be listed, or only the minimum and maximum numeric values - e.g., a question presented for response on a scale from 1 to 5, where "1 = Not at all Concerned; 5 = Very Concerned". The Survey Data Viewer interprets that integer responses between 1 and 5 are valid, but that labels should not be included for response categories 2, 3, and 4.
Features	Assigns Survey Data Viewer features that determine the visualization and color scheme for the question. If a question does not include a feature flag, a categorical color scheme is assigned and the variable is only displayed in percentage view. At the time of writing, the features options are "Demographic", "Bidirectional", "Unidirectional", "Spatial", and "Spatial Threshold".

numbers corresponding to each response category may not have been part of the survey presented to participants, though mean values could be added for surveys that used different question-answer formats.

4.4.3. Heat map view

Surveys may be used to explore how the geographic characteristics of the participants influence answers to survey questions (Fry et al., 2012). The Survey Data Viewer includes a map view option

Table 2
Survey Data Viewer codebook template.

| Variable	VariableLabel	SubVariableLabel	ValueLabels	Features
Venue	Venue where Data Collected			isDemographic
City	City where Data Collected			isDemographic
Investigator	Investigator Leading Data Collection			
Q1	Question 1 Title		1 = yes; 2 = no	
Q2a	Grouped Question 2 Title	SubQuestion 2a Label	1 = Strongly Disagree; 3 = Neither Agree nor Disagree; 5 = Strongly Agree	bidirectional
Q2b	Grouped Question 2 Title	SubQuestion 2b Label	1 = Strongly Disagree; 3 = Neither Agree nor Disagree; 5 = Strongly Agree	bidirectional
Q3a	Grouped Question 3 Title	SubQuestion 3a Label	1 = very bad; 3 = neither good nor bad; 5 = very good; 6 = not sure	bidirectional
Q3b	Grouped Question 3 Title	SubQuestion 3b Label	1 = very bad; 3 = neither good nor bad; 5 = very good; 6 = not sure	bidirectional
Q3c	Grouped Question 3 Title	SubQuestion 3c Label	1 = very bad; 3 = neither good nor bad; 5 = very good; 6 = not sure	bidirectional
Q3d	Grouped Question 3 Title	SubQuestion 3d Label	1 = very bad; 3 = neither good nor bad; 5 = very good; 6 = not sure	bidirectional
Q4a	Grouped Question 4 Title	SubQuestion 4a Label	1 = Not at all concerned; 5 = Very Concerned	unidirectional
Q4b	Grouped Question 4 Title	SubQuestion 4b Label	1 = Not at all concerned; 5 = Very Concerned	unidirectional
Q4c	Grouped Question 4 Title	SubQuestion 4c Label	1 = Not at all concerned; 5 = Very Concerned	unidirectional
Q4d	Grouped Question 4 Title	SubQuestion 4d Label	1 = Not at all concerned; 5 = Very Concerned	unidirectional
Q5	Question 5 Title		1 = Not at all familiar; 5 = Very familiar	unidirectional
Q6	Question 6 Title		1 = Yes; 2 = No	
Q7	Question 7 Title		1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often; 5 = Unsure	
Q8a	Grouped Question 8 Title	SubQuestion 8a Label	1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often	bidirectional
Q8b	Grouped Question 8 Title	SubQuestion 8b Label	1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often	bidirectional
Q9	Question 9 Title		1 = Yes; 2 = No	isDemographic
Q10	Question 10 Title		1 = Very Dissatisfied; 5 = Very satisfied	bidirectional
Q11	Question 11 Title (Own or Rent Home)		1 = Own; 2 = Rent	isDemographic
Q12	Question 12 Title		1 = Yes; 2 = No	isDemographic
Q13	Question 13 Title (Sex)		1 = Female; 2 = Male	isDemographic
Q14	Question 14 Title (Age)		1 = 18 to 29; 2 = 30 to 39; 3 = 40 to 49; 4 = 50 to 59; 5 = 60 and over	unidimensional; isDemographic
Q15	Question 15 Title (Education)		1 = Some High School or High School Diploma/GED; 2 = Some College and/or Vocational School; 3 = 4 Year College Degree; 4 = Graduate Degree	unidimensional; isDemographic
Q16	Question 16 Title (Spatial)		Threshold = 15	spatial
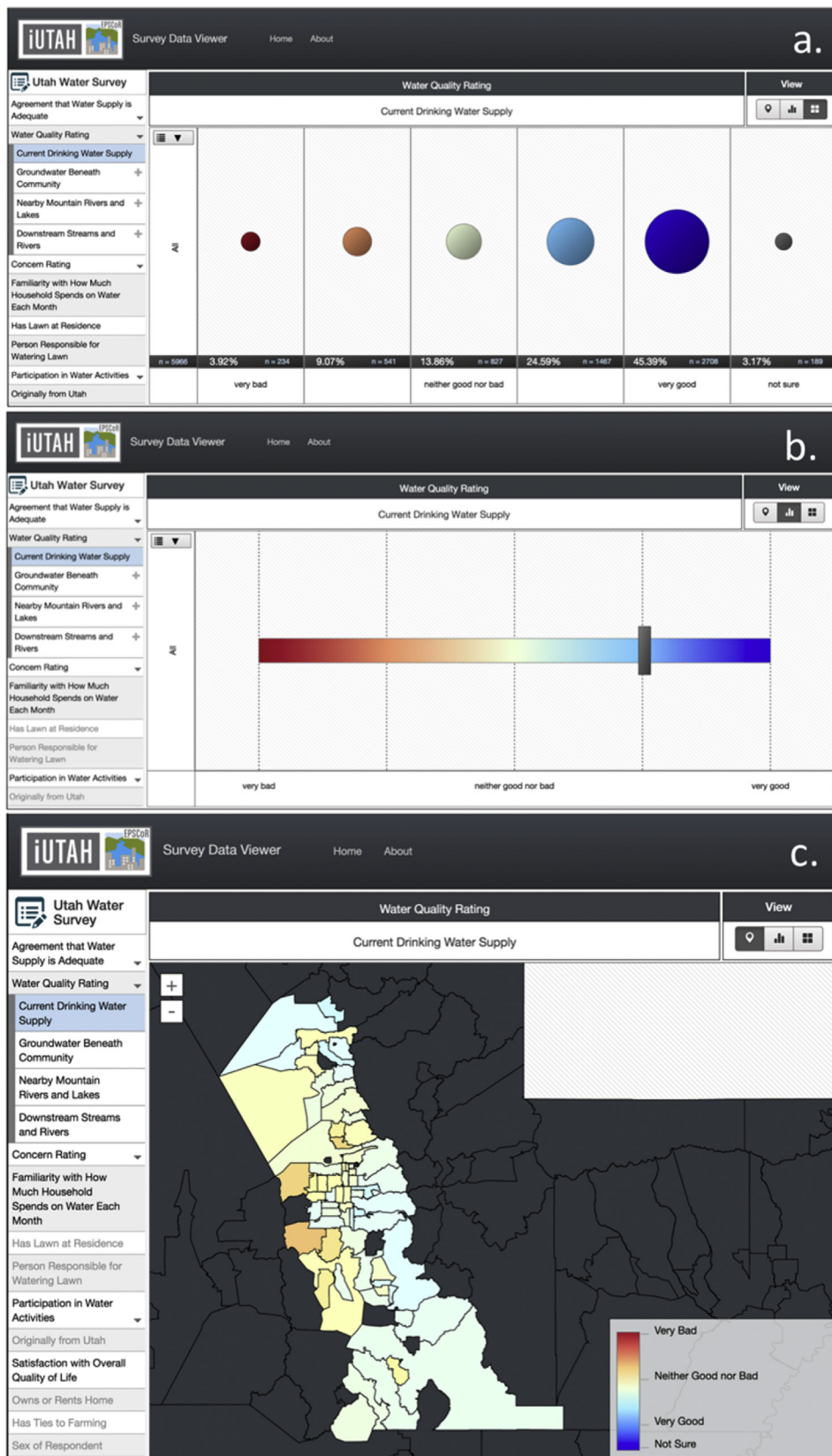
Fig. 3. Percentage view (a), mean result view (b), and heat map view (c) for the same question selected. Results indicate participants' level of concern with drinking water supply.

that uses participants' zip codes to illustrate spatial patterns of responses to individual survey questions (Fig. 3c). The heat map view displays the mean scores to a selected question for participants aggregated at the zip code level. To avoid problems associated with making inferences from small sample sizes, the heat map view only provides zip code level results where the number of responses within a zip code is greater than a threshold specified in the codebook. The total number of survey participants in each zip code is also visible via the heat map when no question is selected.

To implement the heat map view for a particular question, the "Spatial" flag is used in the codebook to indicate the variable associated with spatial data (i.e., zip code). Currently, the heat map view is implemented for zip codes in Utah. However, Utah zip codes were loaded to the map viewer as a GeoJSON (Butler et al., 2008) file that could be replaced with zip code boundaries for other states or geographic areas (or other polygon features on which survey responses are to be aggregated). This enables developers to adapt the map viewer to use other spatial extents and other spatial divisions (e.g., census tract, city boundaries, counties, etc.).

4.4.4. Question types and multiple question selection

Surveys may incorporate several question types. We have implemented options for symbology and features so that response data can be viewed in a way that is most appropriate for the type of question. Surveys often include Likert scale questions with possible responses spanning a scale so that respondents can indicate level of agreement, concern, or satisfaction; frequency; or degree. These questions should either be tagged as "bi-directional" or "uni-directional" in the survey codebook file. Bi-directional variables have responses in two directions, with positive and negative values centered on a neutral midpoint (e.g., from very bad to very good or from strongly disagree to strongly agree). Uni-directional variables consist of responses that represent intensity of a single facet (e.g., level of concern ranging from not at all concerned to very concerned). Because the mean value of these questions has significance as a central tendency of the survey participants, questions designated as bi-directional or uni-directional are displayed on the mean result and heat map views, as well as the percentage view. Nominal, categorical variables that do not have the properties of ordinality (or results that span a scale of values) are only displayed via the percentage view.

Nested variables, or questions that contain sub-questions with similar answer structure (e.g., short matrices; Dillman et al., 2014), are also supported by the Survey Data Viewer. These variables are organized with the parent question as a header in the Question panel of the Viewer with sub-questions nested underneath. When the parent question is clicked, the sub-questions are expanded for selection. Furthermore, we implemented functionality to view related sub-questions simultaneously for both percentage view and mean view. After a single sub-question is selected, additional related sub-questions can be added to the view by clicking a "+" sign on each sub-question of interest. When multiple sub-questions are selected, the results are cross-tabulated, with each question shown as a row on the y-axis with response categories on the x-axis. Sub-questions are removed from the view by using the corresponding "-" sign. Multiple sub-question selection disables the demographics options. Similarly, multiple selected sub-questions cannot be displayed simultaneously in the heat map view. Fig. 4 shows the comparison of multiple sub-questions selected for visualization.
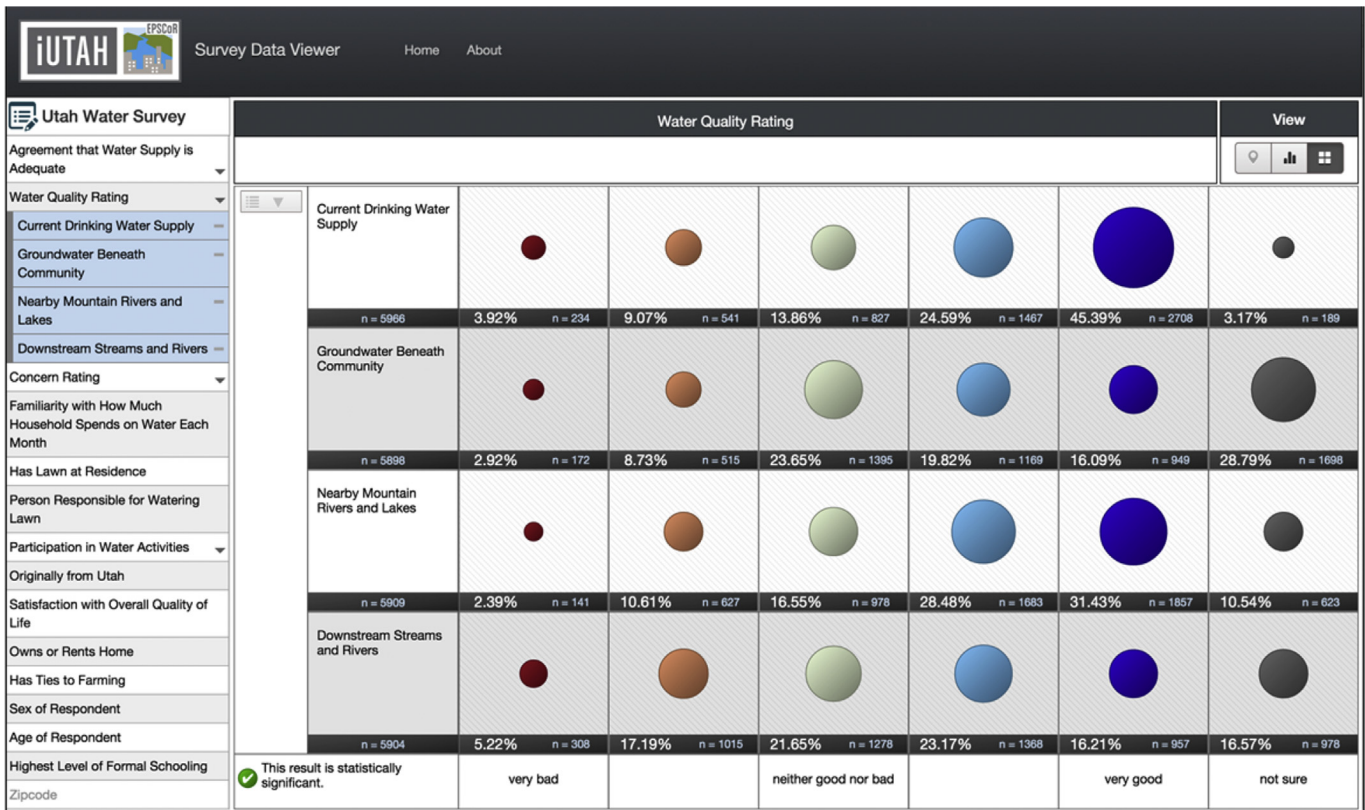

Fig. 4. Survey Data Viewer visualization with multiple, nested sub-questions selected, which are associated with a bi-directional color scheme. The selected question illustrates participants' rating of water quality of various water sources. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.4.5. Color schemes

Several color schemes were implemented to provide context to the responses and to illustrate the types of questions. These schemes are consistent for the percentage, mean result, and heat map views. Color schemes were selected to both indicate a gradient from one end of a response scale to another (e.g., bad to good) (Borland and Taylor, 2007) and to make them distinguishable to viewers regardless of common types of color-blindness (Light and Bartlein, 2004). For bi-directional questions, a diverging blue-to-dark red scheme is used. In this scheme, dark red corresponds to low values in the coded survey responses, which for our case study, are associated with negative connotation to indicate disagreement, dissatisfaction, and low quality. On the other hand, blue corresponds to higher values in the coded survey responses to indicate agreement, satisfaction, and high quality. Neutral responses are highlighted at the middle of this color spectrum (e.g., neither agree nor disagree, neither good nor bad). A dark gray is used to represent "Not Sure" where applicable. These colors are shown in Fig. 4.

For uni-directional questions, a sequential light blue-to-dark blue color scheme was implemented. Light blue is used for lower values in the coded survey data to represent low degree or intensity. Dark blue is used for higher values in the coded survey data to represent high degree or intensity. This color scheme is illustrated in Fig. 5. For variables not designated as bi- or uni-directional, a simple categorical color scheme is used. As these questions are only visible in the percentage view, distinguishing them side-by-side is less important than for bi- or uni-directional questions (i.e., for color blindness).

4.4.6. Variables used to disaggregate results

It is often important to disaggregate survey results based on demographic characteristics of participants (Wexler, 2014). The Survey Data Viewer allows users to select from a set of respondent demographic attributes that were included in the survey. These variables are distinguished by a flag in the codebook file, which makes them selectable on the Survey Data Viewer demographic dropdown menu. When a "demographic" variable is selected from the dropdown menu, the results for the variable of interest are disaggregated into demographic groupings on the y-axis. Demographic disaggregation is only available for percentage and mean views and can only be visualized for a single selected question at a time. Fig. 6 shows a visualization of the results of the same question with two different demographic variables selected.

4.4.7. Statistical significance flag

The overall statistical significance for patterns in cross-tabulated categorical data is typically determined by comparing observed frequencies to a chi-square distribution (Frankfort-Nachmias and Leon-Guerrero, 2014). We implemented a flag to provide an indication of the statistical significance of data displayed categorically in percentage view (i.e., whether or not the results are different from expected random behavior). In percentage view, when a set of variables is selected for cross-tabulation (i.e., multiple questions or demographic variables), the statistical significance is calculated on the fly based on metrics associated with the data selected for display. A message is shown on the view indicating whether or not the differences between categories in the results meet the thresholds for statistical significance (Figs. 4–6). Clicking on the flag in the Survey Data Viewer provides users with details on the determination and interpretation of statistical significance.

Statistical significance is evaluated using the following steps:

1. A table of observed frequencies is constructed (equivalent to the percentage view display).
2. A table of expected frequencies is calculated using Equation (1):

$$E_{ij} = \frac{T_i T_j}{N} \tag{1}$$

where E_{ij} is the expected frequency for each category, T_i is the total of the frequencies for the ith row, T_j is the total of the frequencies for the jth column, and N is the overall total frequency.

3. A calculated chi-square value is determined using Equation (2):

$$\chi^2 = \sum_0^{i,j} \frac{(E_{ij} - O_{ij})^2}{E_{ij}} \tag{2}$$

where χ^2 is the calculated chi-square statistic and O_{ij} is the observed frequency for each category.

4. This calculated statistic is compared with values on a chi-square distribution with the significance level of interest and the associated degrees of freedom. The Survey Data Viewer uses $\chi^2 0.05$ to represent the 95% confidence interval. If the calculated statistic is greater than that of the chi-square distribution, then the difference is statistically different than what would result from chance, and the results are considered significant.

4.4.8. Orientation guidance

As first time users may not be familiar with the features and options available in the Survey Data Viewer, we implemented a brief orientation to provide guidance for initial visits to the site. The orientation consists of tips that point to features and provide text instructions and images to describe important functionality. The orientation begins with a welcome message and includes options to bypass the steps and prevent them from displaying in the future. The first tip highlights the selection of survey questions, the second tip demonstrates toggling between the three available views, the third tip describes the selection of demographic variables, and the fourth tip points out the statistical significance flag. In testing our implementation, we found that providing this level of guidance was needed by most users, regardless of the level of their technical skill.

5. Discussion

In this section, we describe the application of the Survey Data Viewer for the iUTAH Utah Water Survey (Section 3) and highlight insights gained by the ability to easily visualize the data for this example with the Survey Data Viewer. Although not all results are particularly interesting, the exploratory nature of the tool facilitates investigation and testing of hypotheses, important functionality described by Wexler (2014), Knigge and Cope (2006), and Healy and Moody (2014). We also explain the how the Survey Data Viewer has been tested, how it is being extended, and further development.

To implement the Survey Data Viewer for the iUTAH Utah Water Survey, results were uploaded from tablet computers to the Qualtrics server by teams in the field. A single investigator then downloaded and compiled responses from multiple iterations of the survey into a consolidated data file, which was lightly edited to remove unneeded columns and formatted to be consistent with the data table format specified in Section 4.3. Investigators then used the Survey Data Viewer codebook file template to translate the questions in the original survey instrument into the questions and demographic variables for display. These files for the iUTAH Utah Water Survey are published in supplemental electronic files to this paper.

After creating and formatting the input files for the Survey Data Viewer, investigators loaded them to the website along with the
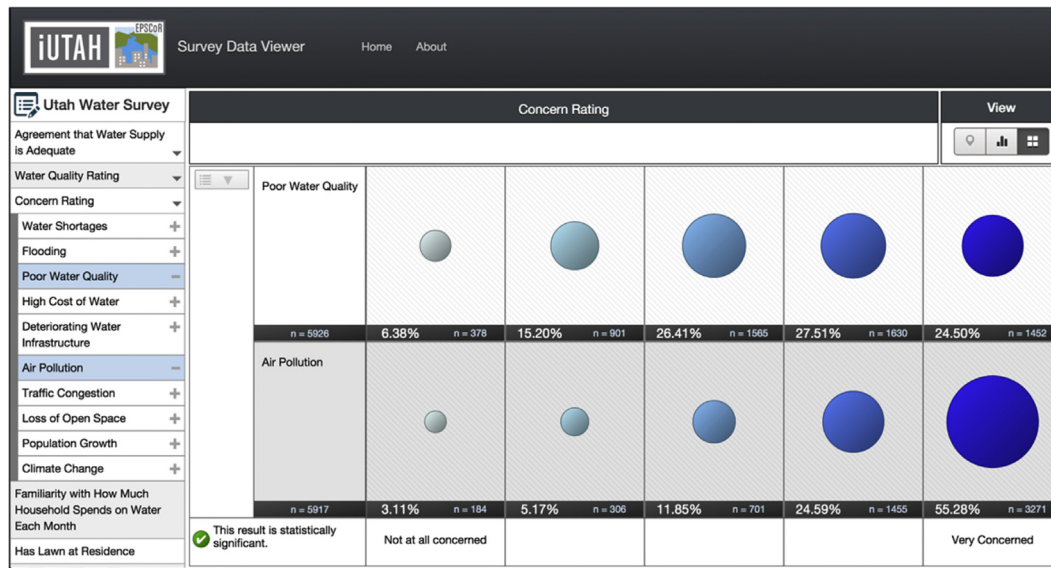
Fig. 5. Survey Data Viewer visualization of a question associated with an uni-directional color scheme. This question illustrates participants' reported concern about water quality and air pollution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

survey instrument and description of the survey for the "About" page. As the survey was ongoing at the time the Survey Data Viewer was originally deployed, new cases were appended to the data file and reloaded to the website as they were received. Re-formatting of the data file and the codebook was only required if additional or different questions were presented in newer iterations of the survey; otherwise, it was simple to append new results as rows to the data file. In this manner, preliminary data could be viewed, with updated visualizations of survey results immediately available as soon as an updated data file was uploaded to the server. The implementation of the Survey Data Viewer for the iUTAH Utah Water Survey results is accessible at http://data.iutahepscor.org/surveys.

For the iUTAH Utah Water Survey, many of the questions in the facet panel represent sets of sub-questions with similar answer structure (e.g., different prompts to which respondents indicated levels of agreement/disagreement). There were cases where it was desired to compare the responses for one question with that of another, related question. The Survey Data Viewer has facilitated exploration of these nested questions. For example, it is straight-forward to view the level of concern regarding poor water quality in conjunction with the level of concern toward other issues, such as air pollution. The Survey Data Viewer quickly reveals that in general, participants were more concerned about air pollution than about water quality (Fig. 5). Another interesting result gleaned from this visualization is that a majority of participants rated the water quality of their drinking water supply and water in nearby mountain rivers and lakes as good, but were less sure about the water quality of local groundwater and of downstream streams and rivers (Fig. 4).

Not only were traditional demographic variables (e.g., age, education level, sex, place of origin) collected as part of the iUTAH Utah Water Survey and implemented in the Survey Data Viewer, but several other measured variables were also used as disaggregation categories for questions of interest. This allows users to explore how responses to core survey questions vary in relation to the extent to which a respondent has family ties to farming, participates in water related activities, owns or rents their home, and whether they have a lawn or responsibility for lawn-watering decisions. When these questions are defined as 'demographic'

variables, the Survey Data Viewer allows users to disaggregate results against these characteristics of respondents. In our Water Survey example, we find several interesting patterns: participants with farm ties (i.e., whether or not a family member in the current or most recent generation was a farmer) were less concerned about climate change and air pollution than those without farm ties (Fig. 7). Participants originally from Utah were also less concerned about air pollution than those originally from outside of the state. These results are quickly accessible with both the percentage view as well as the mean result view through the click of a button. The ability to visually explore these data and quickly interact with the results has helped launch further investigation into some unexpected results of the iUTAH survey (Baji and Jackson-Smith, 2016; Barnett and Jackson-Smith, 2016).

It was anticipated that perceptions and concerns about water issues among respondents to the iUTAH Utah Water Survey would relate to social and geographic attributes of the places where respondents live. To maintain anonymity, the survey only asked respondents to provide the zip code of their residence, and the zip codes provide a basis to visualize spatial patterns in responses to key survey questions. The current instance of the Viewer uses a map of Utah as the extents of a dynamic heat map view, with a zip code layer for the map category aggregation. The map view highlights some interesting spatial patterns (e.g., more people living near mountains reported participating in hiking and snowsports) as well as helps identify spatial 'hotspots' where certain types of water concerns were higher than average. For example, a few zip codes stood out as having residents reporting relatively poor quality for current drinking water supply, while neighboring zip codes generally reported good quality (Fig. 3c).

The statistical significance flag for cross-tabulations helped the investigators and users of the website identify situations where apparent differences between groups are (or are not) statistically reliable. The addition of a warning flag provides a mechanism to alert users to situations where the patterns in the table they are viewing should not be treated as a statistically meaningful result. For example, in the Utah Water Survey, the comparison of concerns about climate change by store type (i.e., the locations at which the surveys were administered) reveals some apparently interesting differences. However, these are not statistically valid (e.g., not
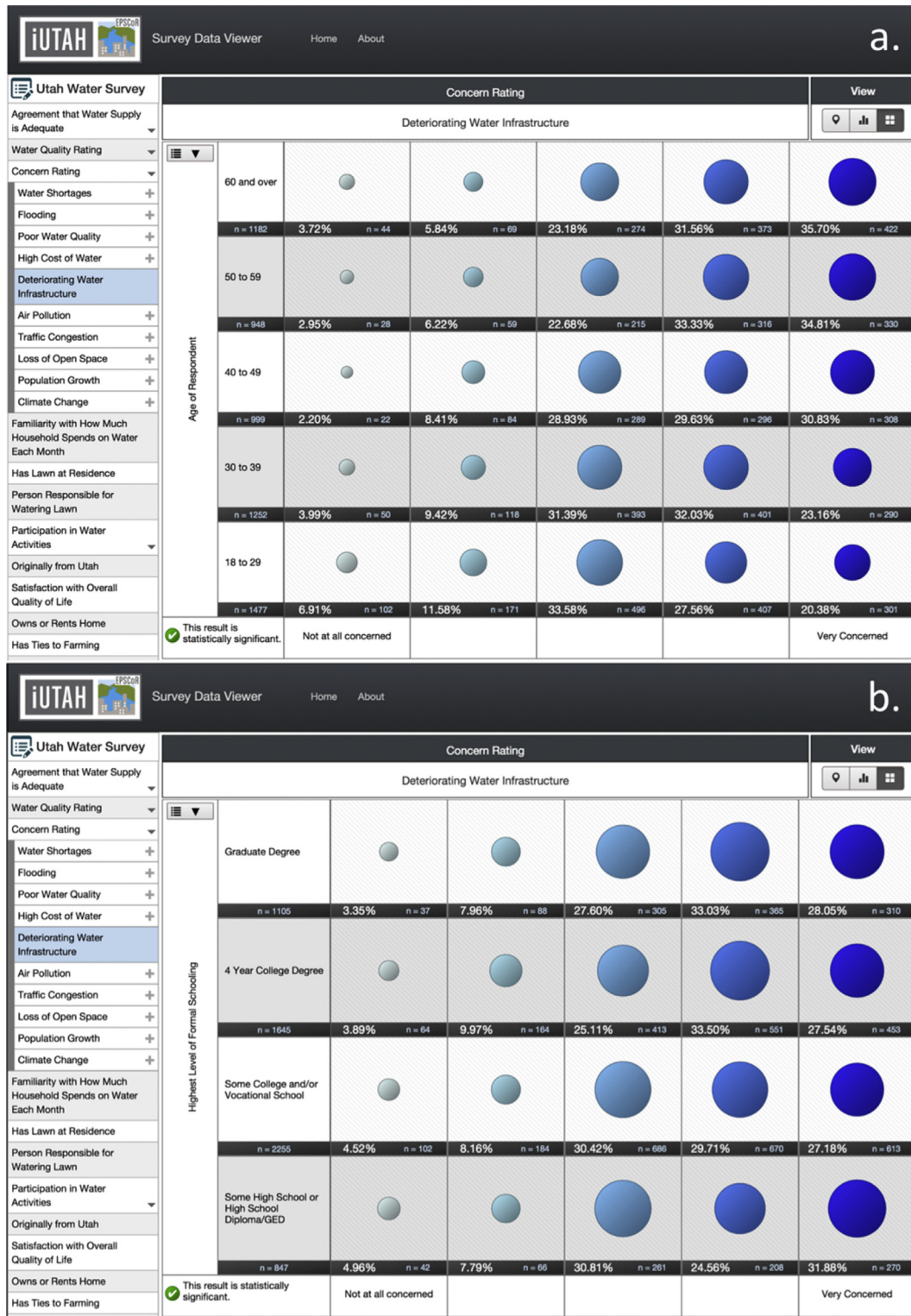
Fig. 6. Survey Data Viewer showing data for a single question but disaggregated by two different demographic variables: age group (a) and education level (b).

significantly different from what we would expect by sheer random chance). In such cases, it would be unwise to draw inferences from those patterns. Conversely, as we added cases to the dataset, an increasing number of cross-group comparisons emerged as statistically significant, even when the size of the different response patterns did not change substantively. In such cases, the larger sample size enables us to say with greater confidence that the patterns we see are not simply a product of random chance but appear to be real differences.

5.1. Testing and feedback

The Survey Data Viewer was developed over multiple iterations of input from a broad group of researchers consisting of social
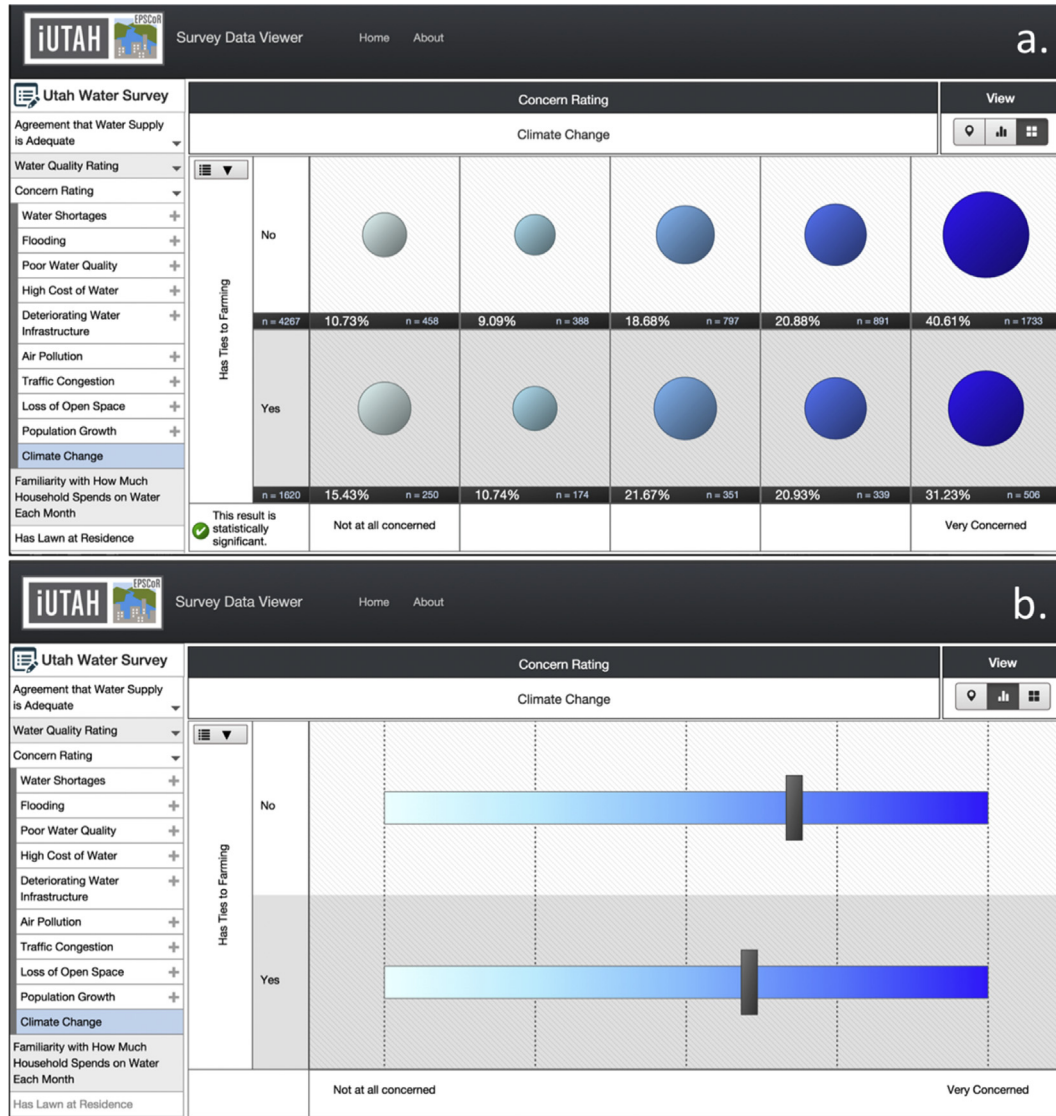
Fig. 7. Participants' reported level of concern about climate change disaggregated by whether participants have family farm ties in percentage view (a) and mean result view (b).

scientists, environmental engineers, and computer programmers from the undergraduate to faculty level. At the time of writing, the Survey Data Viewer was being initially presented to project stakeholders, participants, and the public throughout the state of Utah. Informal focus groups of students from several scientific disciplines, including some who administered the survey and some who did not, tested and used the tools and provided feedback, which provided the developers with suggestions for improvement and usability. Input from these groups particularly motivated the implementation of the orientation instructions to highlight functionality for first time users. The tool has not been thoroughly vetted by all user groups, which is outside of the scope of this paper; however, we are encouraged that the data are accessible to broad audiences without requiring them to recreate visualizations or run complex software.

The Survey Data Viewer was conceptualized and initially implemented for the iUTAH Utah Water Survey, but designed to be generalizable to any quantitative survey dataset. Since the initial development of the Utah Water Survey and the Survey Data Viewer, we have implemented the Survey Data Viewer for a completely separate survey. This was straightforward and carried out by a

student unfamiliar with the development of the Survey Data Viewer. Also, the Survey Data Viewer could be implemented for other types of survey data (e.g., containing personally identifying information or qualitative results) after appropriate anonymization or aggregation procedures have been completed for results with identifying information or after coding procedures have been completed for qualitative responses.

5.2. Potential improvements

Developers and users have identified several visualization options that would be useful for implementation in future versions of the Survey Data Viewer. One important next step is the functionality for users to define their own base maps and aggregation polygons for the heat map view. Another desirable feature is the ability to compare results of the demographic breakdown to overall totals while on the same view as well as to view by multiple demographic variables (e.g., sex and age at the same time). The ability to collapse categories for demographic variables in the y-axis and response categories in the x-axis would also be a useful addition. For example, users may want to view two age categories rather

than the five used in the original survey instrument, or users may want to combine all positive responses and all negative responses to a question rather than view the full gradient of responses. Another feature that may be desirable are error bars around the mean marker on the mean result view to give an indication of the spread of the data. This could also be used to report statistical significance of comparisons of means across groups. Finally, some users may desire alternate methods for viewing the distribution of responses (e.g., add a view that shows results in different styles of bar charts). Any of these features as well as other visualization types (e.g., Heer et al., 2010) could be implemented via contributions to the Survey Data Viewer code repository. We welcome further development on this tool while recognizing that separate tools/interfaces may be preferred for interaction and visualization of different types of data.

6. Summary and conclusions

The Survey Data Viewer provides a mechanism for making quantitative, social science survey data accessible to a variety of users with a broad range of technical and social science expertise. The web-based interface means that no specialized software is required for users to visualize the data, and the interactive aspects of the Survey Data Viewer make social science survey data available in new and dynamic ways. The Survey Data Viewer includes functionality to display responses to various survey question types by percentage of respondents, mean response result, and spatial distribution. Additional features include disaggregation by demographic variables, symbology to aid in interpretation of responses, the ability to view multiple questions simultaneously, and the display of the statistical significance of cross-tabulated results.

We designed the Survey Data Viewer with the flexibility to include results from multiple surveys and to dynamically accept updated results as they become available. The Viewer can be easily adapted for reuse by other research groups conducting their own social science surveys by using the data and metadata templates developed and deploying on a local web server according to the instructions we developed. The source code for the Survey Data Viewer is open-source, and the use of an open web development platform including Python and Django provides flexibility for deployment on multiple server platforms.

The Survey Data Viewer contributes a potentially valuable tool to facilitate the growing web-based interactions between citizens and governments (Conroy and Evans-Cowley, 2006). In the case of the Utah Water Survey, the Survey Data Viewer allows stakeholders from any vantage point to explore survey data and make their own assessments of the important patterns in water related perspectives and behaviors reported by different demographic groups or in particular localities. Other potential applications include use by environmental scientists seeking to explain human drivers of observed variation in geo-referenced biophysical data, or as an educational tool to demonstrate to students the complexity of human factors related to environmental problems. As social science data are increasingly being collected in concert with biophysical data as part of interdisciplinary environmental science projects, we anticipate that the viewer will be a useful tool to provide opportunities for data sharing and facilitating public data access and visualization in mechanisms analogous to those used for biophysical data. The interactivity provided by this tool promotes the multi-directional information sharing critical to making participatory decisions for environmental sustainability. Though our focus was on environmental projects, the functionality of the Survey Data Viewer could be successfully applied to survey data of other domains.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envsoft.2016.07.013.

References

Baji, V., Jackson-Smith, D., 2016. Age and water concern. In: Presentation at the International Symposium on Society and Resource Management. June 2016, Houghton, MI.

Barnett, M., Jackson-Smith, D., 2016. Water-based recreation and water quality perception and concern among Utahns. In: Presentation at the International Symposium on Society and Resource Management. June 2016, Houghton, MI.

Borland, D., Taylor, R.M., 2007. . Rainbow color map (still) considered harmful. IEEE Comput. Graph. Appl. 27 (2), 14–17. http://dx.doi.org/10.1109/MCG.2007.323435.

Braden, J.B., Brown, D.G., Dozier, J., Gober, P., Hughes, S.M., Maidment, D.R., Schneider, S.L., Schultz, P.W., Shortle, J.S., Swallow, S.K., Werner, C.M., 2009. Social science in a water observing system. Water Resour. Res. 45, 1–11. http://dx.doi.org/10.1029/2009WR008216.

Braden, J.B., Jolejole-Foreman, M.C., Schneider, D.W., 2014. Humans and the water environment: the need for coordinated data collection. Water 6 (1), 1–16. http://dx.doi.org/10.3390/26010001.

Butler, H., Daly, M., Doyle, A., Gillies, S., Schaub, T., Schmidt, C., 2008. The GeoJSON Format Specification (last accessed 16.10.15.). http://geojson.org/geojson-spec.html.

Coenen, F., Huitema, D., O'Toole, J., 2012. Participation and the Quality of Environmental Decision Making. Springer Science & Business Media.

Corti, L., 2012. Recent developments in archiving social research. Int. J. Soc. Res. Methodol. 15 (4), 281–290. http://dx.doi.org/10.1080/13645579.2012.688310.

Conroy, M.M., Evans-Cowley, J., 2006. E-participation in planning : an analysis of cities adopting on-line citizen participation tools. Environ. Plan. C Gov. Policy 24, 371–385. http://dx.doi.org/10.1068/c1k.

Curtis, A., Byron, I., Mackay, J., 2005. Integrating socio-economic and biophysical data to underpin collaborative watershed management. J. Am. Water Resour. Assoc. 41, 549–563. http://dx.doi.org/10.1111/j.1752-1688.2005.tb03754.x.

Dillman, D., Smyth, J.D., Christian, L.M., 2014. Internet, Phone, Mail, and Mixed-mode Surveys: the Tailored Design Method, fourth ed. Wiley.

Fath, B.D., Beck, M.B., 2005. Elucidating public perceptions of environmental behavior: a case study of Lake Lanier. Environ. Model. Softw. 20, 485–498. http://dx.doi.org/10.1016/j.envsoft.2004.02.007.

Fowler, F.J., 2013. Survey Research Methods, fifth ed. Sage Publications, Inc.

Frankfort-Nachmias, C., Leon-Guerrero, A., 2014. Social Statistics for a Diverse Society, seventh ed. Sage Publications.

Freese, J., 2007. Reproducibility standards in quantitative social science: why not sociology? Sociol. Methods Res. 36 (2), 153–172. http://dx.doi.org/10.1177/0049124107306659.

Fry, R., Berry, R., Higgs, G., Orford, S., Jones, S., 2012. The WISERD geoportal: a tool for the discovery, analysis and visualization of socio-economic (meta-) data for wales. Trans. GIS 16, 105–124. http://dx.doi.org/10.1111/j.1467-9671.2012.01308.x.

Hale, R.L., Armstrong, A., Baker, M.A., Bedingfield, S., Betts, D., Buahin, C., Buchert, M., Crowl, T., Dupont, R.R., Ehleringer, J.R., Endter-Wada, J., Flint, C., Grant, J., Hinners, S., Horsburgh, J.S., Jackson-Smith, D., Jones, A.S., Licon, C., Null, S.E., Odame, A., Pataki, D.E., Rosenberg, D., Runburg, M., Stoker, P., Strong, C., 2015. iSAW: integrating Structure, Actors, and Water to study socio-hydro-ecological systems. Earth's Future 3, 110–132. http://dx.doi.org/10.1002/2014EF000295.

Hamilton, E.C., 2006. The impact of survey data: measuring success. J. Assoc. Inf. Sci. Technol. 58, 190–199. http://dx.doi.org/10.1002/asi.

Healy, K., Moody, J., 2014. Data visualization in sociology. Annu. Rev. Sociol. 40, 105–128. http://dx.doi.org/10.1146/annurev-soc-071312-145551.

Heer, J., Bostock, M., Ogievetsky, V., 2010. A Tour through the Visualization Zoo: a

survey of powerful visualization techniques, from the obvious to the obscure. ACM Queue 8 (2), 1—22. http://queue.acm.org/detail.cfm?id=1805128.

Hiwasaki, L., Arico, S., 2007. Integrating the social sciences into ecohydrology: facilitating an interdisciplinary approach to solve issues surrounding water, environment and people. Ecohydrol. Hydrobiol. 7, 3—9. http://dx.doi.org/10.1016/S1642-3593(07)70184-2.

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environ. Model. Softw. 24, 879—888. http://dx.doi.org/10.1016/j.envsoft.2009.01.002.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2011. Components of an environmental observatory information system. Comput. Geosciences 37, 207—218. http://dx.doi.org/10.1016/j.cageo.2010.07.003.

Jones, A.S., Horsburgh, J.S., Reeder, S.L., Ramírez, M., Caraballo, J., 2015. A data management and publication workflow for a large-scale, heterogeneous sensor network. Environ. Monit. Assess. 187, 348. http://dx.doi.org/10.1007/s10661-015-4594-3.

Kirk, A., 2013. Discussion: Storytelling and Success Stories. http://www.visualisingdata.com/index.php/2013/04/discussion-storytelling-and-success-stories/.

Knigge, L., Cope, M., 2006. Grounded visualization: integrating the analysis of qualitative and quantitative data through grounded theory and visualization. Environ. Plan. A 38, 2021—2038. http://dx.doi.org/10.1068/a37327.

Lehnert, K.A., Carbotte, S.M., Ryan, W.B.F., Ferrini, V., Block, K., Arko, R.A., Chan, C., 2011. IEDA: integrated earth data applications to support access, attribution, analysis, and preservation of observational data from the ocean, earth, and polar sciences. Geophys. Res. Abstr. 13, EGU2011—13113.

Light, A., Bartlein, P.J., 2004. The end of the rainbow? Color schemes for improved data graphics. EOS Trans. Am. Geophys. Union 85 (40), 385.

Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L., Linsey, K.S., 2014. Estimated Use of Water in the United States in 2010: U.S. Geological Survey Circular 1405, p. 56. http://dx.doi.org/10.3133/cir1405.

Morehouse, B.J., O'Brien, S., Christopherson, G., Johnson, P., 2010. Integrating values and risk perceptions into a decision support system. Int. J. Wildland Fire 19, 123—136. http://dx.doi.org/10.1071/WF08064.

Pickett, S.T.A., Belt, K.T., Galvin, M.F., Groffman, P.M., Grove, J.M., Outen, D.C., Pouyat, R.V., Stack, W.P., Cadenasso, M.L., 2007. Watersheds in baltimore, Maryland: understanding and application of integrated ecological and social processes. J. Contemp. Water Res. Educ. 136, 44—55. http://dx.doi.org/10.1111/j.1936-704X.2007.mp136001006.x.

Redman, C.L., Grove, J.M., Kuby, L.H., 2004. Integrating social science into the long-term ecological research (LTER) network: social dimensions of ecological change and ecological dimensions of social change. Ecosystems 7, 161—171. http://dx.doi.org/10.1007/s10021-003-0215-z.

Ryssevik, J., Musgrave, S., 2001. The social science dream machine. Soc. Sci. Comput. Rev. 19, 163—174.

Schwabish, J.A., 2014. An economist's guide to visualizing data. J. Econ. Perspect. 28, 209—234. http://dx.doi.org/10.1257/jep.28.1.209.

Shah, P., Hoeffner, J., 2002. Review of graph comprehension research: implications for instruction. Educ. Psychol. Rev. 14 (1), 47—69.

Sivapalan, M., Konar, M., Srinivasan, V., Chhatre, A., Wutich, A., Scott, C.A., Wescoat, J.L., Rodríguez-Iturbe, I., 2014. Socio-hydrology: use-inspired water sustainability science for the Anthropocene. Earth's Future 2, 225—230. http://dx.doi.org/10.1002/2013EF000164.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M., 2011. Data sharing by scientists: practices and perceptions. PLoS ONE 6 (6). http://dx.doi.org/10.1371/journal.pone.0021101 e21101.

Tufte, E.R., 1983. The Visual Display of Quantitative Information. Graphics Press, Cheshire, CT.

Wagener, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H.V., Kumar, P., Rao, P.S.C., Basu, N.B., Wilson, J.S., 2010. The future of hydrology: an evolving science for a changing world. Water Resour. Res. 46, 1—10. http://dx.doi.org/10.1029/2009WR008906.

Wexler, S., 2014. Visualizing Survey Data. Data Revelations (Last accessed 19.10.15.). http://www.datarevelations.com/visualizing-survey-data.

Zaslavsky, I., Whitenack, T., Williams, M., Tarboton, D.G., Schreuders, K., Aufdenkampe, A., 2011. The initial design of data sharing infrastructure for the Critical Zone Observatory. In: Proceedings of the Environmental Information Management Conference, Santa Barbara, CA, 28—29 September, EIM 2011. http://dx.doi.org/10.5060/D2NC5Z4X.