Environmental Modelling & Software 84 (2016) 412–426

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

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

Article history:
Received 20 November 2015
Received in revised form
6 June 2016
Accepted 14 July 2016

Keywords:
Social science
Survey data
Visualization
D3.js
Python
Django

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.

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).

http://dx.doi.org/10.1016/j.envsoft.2016.07.013

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.

[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).

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 open-ended 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.,