Become One With the Data: Technological Support of Shared Exploration of Data in Informal Settings

J. Roberts, L. Lyons, J. Radinsky.

Abstract: The "stuff" of modern social science is data, but museums are often ill-equipped to allow visitors to appreciate and explore data as an "artifact" of science. Interpreting data visualizations can include both the ability to detect patterns and the ability to hypothesize possible causal mechanisms that would produce such patterns. We are utilizing embodied interaction as a means of augmenting visitors' personal connection with a data subset in an effort to promote shared curiosity and hypothesis generation as multiple visitors explore the data collaboratively. Here we share initial findings of the impacts of embodiment on visitors' sense-making around the display.

Introduction

Geographic Information Systems (GIS) technology is a powerful tool with great potential for teaching and learning in formal contexts (e.g. Radinsky et al., submitted). However, many of the GIS-based visualizations we encounter are *outside* of the school context, without the aid of a teacher to interpret them for us. Therefore it is important to pay attention to the how people are interacting with and interpreting these visualizations in informal contexts, such as museums.

The CoCensus project seeks to engage visitors in exploring data from the United States census and American Community Survey (ACS) in a museum environment by using *embodied interaction*. Unlike a typical digital display where users control input using a mouse or keyboard, CoCensus uses remote sensing technologies – an Xbox Kinect camera and passive radio frequency identification (RFID) antennae – to track visitors' movements within an interaction space and change the display according to individuals' locations within the room. This provides multiple users access to the controls of a display, rather than limiting control to just one user and relegating other participants to the periphery (Heath & vom Lehn, 2008). A thorough description of the modes of interactivity in various phases of the project can be found below.

This design aims to help visitors engage collaboratively with a complex data set by first narrowing the scope to just one small subset (e.g. first ancestry data) of the larger set and by fostering *connections* to the data through two primary means: *personal connections* and *physical connections*.

Personal connections

U.S. census data has only recently been made readily available to the general public through websites such as Social Explorer and NHGIS. Previously only researchers and experts were able to explore these data in raw form. Now novices have access to both the tables and maps of the spatially distributed data. Exploration of these data maps is not easy, however, and users need support. This support varies based on the context – museum learners' needs are different from those of classroom teachers and students. Because the museum is a "free choice" learning environment (Falk, 2001) intended to spark interest and conversation among visitors rather than explore deep content, rather than providing the entire corpus of census data to visitors, we have stripped it down to one particular category and one subset of that category – in our case first ancestry. Visitors are asked to identify an ancestry or ethnic heritage that represents them, and when they enter the interaction area, they see scaled centroids ("bubbles") appear on the map illustrating the distribution of the population of ACS respondents who also listed that ancestry first on their questionnaire. By letting users self-identify their ancestry and displaying just that subset of the one census category, we are trying to give museum visitors a way in to this complex data set by letting them see the distribution of people "like them" in that area.

Physical connections

In order for an embodied interaction experience to be successful, users need to be able to "instrumentally operate" within the system (Williams and Dourish, 2005), first by recognizing that they are in control and ultimately how to navigate as they wish. We have tested multiple means of engagement in a lab setting (see Cafaro et al., 2013) and implemented two controls *in situ*. The first is related to the users' distance from the display, along the vertical axis (perpendicular to the display, see Figure 1). When a user first enters the interaction area from the back, her data bubbles appear but are semi-transparent. As she moves closer to the display they become increasingly opaque. This connection serves two purposes: 1) It is a visual cue to highlight the users' control of the data set, showing the user that as she moves the display changes. 2) It allows multiple users in the space to see their own unique data sets concurrently. Without some level of transparency, larger data sets would obscure smaller populations. To further mitigate the occlusion problem, the z-order, or stacking order, of the bubble layers is also mapped to this axis – whoever is closest to the display sees his bubbles "on top." This interactivity allows users to negotiate the space in order to complete their desired exploration. For example, if two visitors want to look at a pattern of one data set, one person might go forward while the other retreats and fades out their bubbles. Alternatively, if they want to compare the two distributions they might both stand near the middle to make their bubbles semi-transparent.

The second mode of physical connectivity tested *in situ* utilizes the horizontal access (parallel to the display, see figure 1) as a timeline, allowing visitors to control the census decade displayed on the map by stepping side-to-side into marked spaces on the floor. This interactivity allowed visitors to explore changes in the data over time.

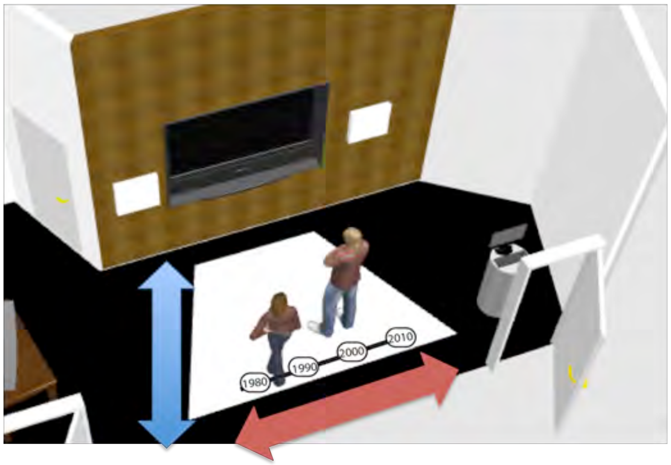

Figure 1. The physical setup of CoCensus. Movement along the axis perpendicular to the screen (blue arrow) affects transparency and z-order (stacking order) of the data bubbles. Movement along the timeline on the axis parallel to the screen (red arrow) allows visitors to look at change over time. Unlike the transparency that is individually controlled, one user at a time controls the timeline for all data on the display.

Methods

During this ongoing iterative design process, we have implemented enactments of this exhibit in two museums: the Jane Addams Hull House (a small urban history museum), and the New York Hall of Science (NySci, a large interactive science center). Here I will focus on two specific interviews conducted in two separate enactments, one at each site, in order to illustrate the types of hypotheses and reasoning we have seen around the exhibit. One interview, conducted at the Hull House Museum, emphasized personal connectivity by allowing visitors to explore personally selected ancestry data. The alpha transparency and z-order positioning were in effect during this interview. The second interview I discuss here took place at NySci and emphasized physical connectivity by incorporating the timeline feature to allow visitors to explore an impersonal data set – median household income. Both interviews consisted of multiple visitors (two colleagues visiting the Hull House and a family of four at NySci) who participated in semi-structured interviews conducted by members of the research team. Audio and video recordings were taken and transcribed. See figure 2 for images of the physical spaces.

Figure 2. Visitors interacting with CoCensus at the Jane Addams Hull House Museum (left) and the New York Hall of Science (right).

Analysis

Narrative hypotheses: Belle and Peg

Belle and Peg (all names are pseudonyms) more than any other participants in this enactment embodied the data and took on a personal identity, putting themselves in the map by using first person pronouns (Roberts et al 2013). The ancestries chosen by these two participants had very different distributions (see Figure 3). Belle and Peg hypothesized about the differences in distributions by taking a narrative approach – they assigned identities to their data bubbles, as demonstrated in the following excerpt:

> BELLE: It looks like I'm along the lake.
> PEG: And I'm not. ((laughter))
> PEG: Polish are inland. We're farming folk!
> BELLE: We're sailors!
> PEG: You're sailors, yeah. ((laughing))

The two adopted the narrative of Belle, embodying the British population located closer to Lake Michigan, as a sailor and of Peg, of Polish ancestry distributed further inland, as "farming folk" quickly and returned to it several times throughout the interview. A minute after the first mention, they furthered the hypothesis by attributing a financial correlation to the ancestry data:

…

> PEG: You're right, you're sailors!... So right, we have no idea... that's... if you're in the, if you're in the upscale? ((laughs))
> BELLE: I would think so…
> PEG: I think so too. You're by the water. It's more expensive.

And again six minutes later they attributed dispositions to their data sets:

> PEG: So you're partying and I'm there in the fields.
> BELLE: Right, I'm having (all the fun).
> PEG: Shucks. ((laughs))
> BELLE: I'm out on the lake, I'm much cooler then you.

Belle and Peg, both librarians in town for a conference, used the data distributions on the map as springboards for a narrative about the people the data represented. They first noticed the differences in the distributions, with Belle "along the lake" and Peg "inland" and hypothesized about what those differences might mean in terms of occupation, income, and lifestyle. Their hypotheses were not based in facts from the data but nonetheless provided a meaningful interaction with the display.

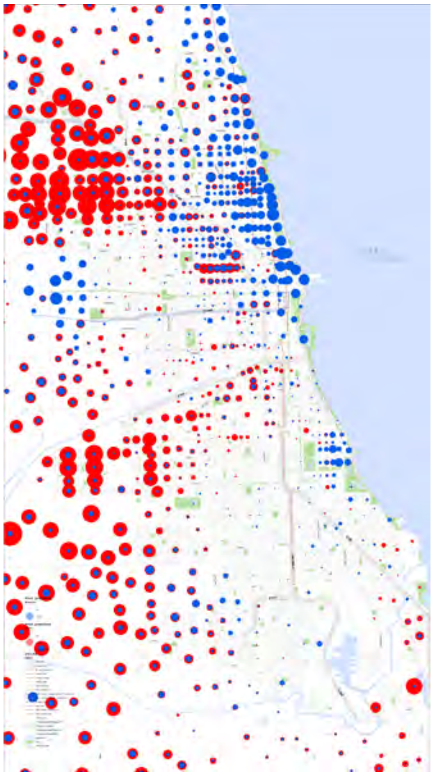

Figure 3. Re-creation of map viewed by Belle and Peg. The two noticed the differences between Belle's British bubbles in blue and Peg's Polish population in red.

Data-driven hypotheses: The Allen family

The next example demonstrates a very different type of hypothesis generation conducted by a family visiting NySci. This implementation featured not scaled centroids of personalized data as viewed by Belle and Peg, but a choropleth map showing median household income at the census tract level in New York City represented as 5 quantiles (see Figure 4). Because only one data set was being displayed at a time, the horizontal axis interactivity of transparency and z-order positioning was disabled, and only side-to-side movements along a timeline marked on the floor affected the data display. Therefore only one participant interacted with the map at a time; others were relegated to the periphery. However, the large screen size (approx. 6 ft. wide) allowed all visitors in the room to see the display.

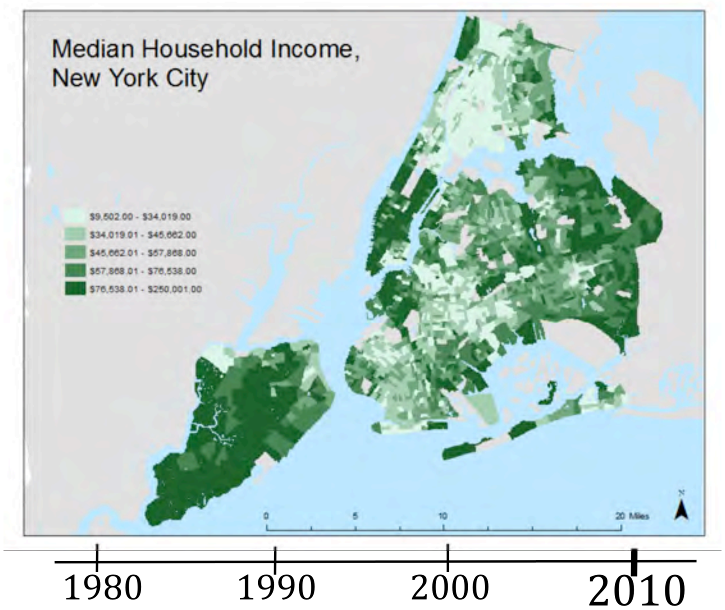

Figure 4. Re-creation of display viewed by the Allen family.

In this interview, the college-aged daughter, Kim, was the first to approach the exhibit. After about 3 minutes of interacting and participating in an interview her mother, Julie, joined her, followed a minute later by her father, Ken. Their teenaged son appeared briefly at the end of the interview but was not present during the portions described here. Kim immediately noticed and was interested in the class breaks used in each decade and how much they changed from year to year. Her questions before any other family members showed up were about the key and the quantile representation, and as soon as her mother entered, Kim pointed this detail out to her and explained the representation. Then, around 5 minutes into the interview, she noticed a phenomenon of interest:

> KIM: Oh that's interesting.
> KIM: I - I don't know what that bottom sector is, but it changes drastically. Right there.
> JULIE: White, from white to...
> KIM: Goes from bottom 25% to you know, middle or so.
> …
> KIM: It's kind of hard to distinguish between those last three colors, I can tell which one is darkest. Dad, do you know what area that is?
> KEN: Which?
> KIM: That area that's turning from white to middle green? Right there?
> KEN: Um, let's see. At the bottom… No, no wait. I'm sure that's Queens, but I'm not sure if that's near us or not.
> KIM: Really?
> KEN: I don't think it is.

Kim and Ken continued discussing this particular region of interest for some time. The census tract was a large tract consisting of JFK airport, but not being from the area, they were not immediately able to identify it. Instead, Kim toggled the timeline back and forth multiple times from 1990 to 2000 to 2010 – the time period over which she noticed the change – and they began offering observations and hypotheses:

> JULIE: And it went back to white!
> KIM: Interesting.
> JULIE: In twenty years
> KIM: Ten. Well, within the space of twenty years, it has uh, from white to green to white again.
> JULIE: That's what I mean.
> KIM: Um. ... and I mean, it completely skips that middle range, so, um.
> KEN: I wonder if that white area, what's white now, could be an area where is actually no data. Because the rest of the map outside of New York City is also white.
> KIM: If it's a different shade of white, you can just barely tell. Um, it's a bit more
> KEN: Yeah, yeah I see what you mean.
> KIM: But uh, I was just thinking about whether or not this has to do with data, definitely.

After about three minutes of this discussion, Ken pulled out his smart phone to investigate the area on Google Maps, eventually determining that it "looks like it's mainly an industrial area. That may be JFK airport." This led to further discussion about the size of the airport, the population of its census tract, and how the decision to use quantiles affected the interpretation of the numbers.

Discussion

This family self-described as "probably not your typical subjects" – Ken and his son both identified as programmers and Kim admitted to enjoying investigating data maps as a hobby. Throughout the interview they drew on their own knowledge, each others' ideas, and outside resources such as Google Maps to pose and investigate hypotheses about this one particular area of interest. This type of collaborative analysis and knowledge building is exactly the type of interaction we hope to support with this exhibit. We should not, however, look at it as a dichotomy with the narrative hypothesis example above, where the Allens' fact-based hypotheses are "good" compared to Belle and Peg's "bad" narrative hypotheses. To the contrary, both types of interactions should be viewed as valuable learning experiences for the participants. In both examples, visitors identified a phenomenon of interest in the spatial data, contributed observations about the phenomenon, and posed hypotheses about its origin and meaning. And importantly in this museum context, both groups reported enjoying the experience afterward.

Ongoing Work

Here I have described two markedly different interactions with a GIS data display in two markedly different situations. In one case, visitors were personally tied to an individualized data set designed to represent their first ancestry using scaled centroid "bubbles" in Chicago in one particular point in time. In the other, visitors were able to navigate an impersonal data set of median household income by physically moving along a timeline in front of the display. Ongoing work in this project combines physical and personal connections. Visitors will now be asked not just one census question, but four and will have the ability to utilize transparency, z-order layering, and a timeline to negotiate space and explore data collaboratively. We hypothesize that this type of interactivity and connectivity will further support the types of exploration shown here.

References

Cafaro, F., Panella, A., Lyons, L., Roberts, J., & Radinsky, J. (2013) I See You There! Developing Identity-Preserving Embodied Interaction for Museum Exhibits. Accepted for publication in the Proceedings of the 2013 ACM int'l conference on Human Factors in Computing Systems. ACM, New York, NY, USA.

Falk, J. H. (2001). *Free-Choice Science Education: How We Learn Science outside of School. Ways of Knowing in Science and Mathematics Series*. Teachers College Press, PO Box 20, Williston, VT 05495-0020.

Roberts, J., Cafaro, C., Kang, R., Vogt, K., Lyons, L., & Radinsky, J. (2013). That's Me and That's You: Museum visitors' perspective-taking around an embodied interaction data map display. In *Proceedings of the 10th International Conference on Computer-Supported Collaborative Learning*. Madison, WI, USA. 343-344.

Heath, C., & Vom Lehn, D. (2008). Configuring'Interactivity': Enhancing Engagement in Science Centres and Museums. *Social Studies of Science*,*38*(1), 63-91.

Williams, A., Kabisch, E., & Dourish, P. (2005). From interaction to participation: Configuring space through embodied interaction. Proceedings of the Ubicomp 2005, LNCS 3660 (pp. 287-304).