Common Sense Community: Scaffolding Mobile Sensing and Analysis for Novice Users

Wesley Willett[1], Paul Aoki[2], Neil Kumar[1],
Sushmita Subramanian[2], and Allison Woodruff[2]

[1] Computer Science Division, University of California, Berkeley, Berkeley, CA 94708 USA
[2] Intel Labs Berkeley, 2150 Shattuck Ave, Ste. 1300, Berkeley, CA 94704 USA
`willettw@eecs.berkeley.edu, aoki@acm.org,`
`neilkumar@berkeley.edu, sushmita.subramanian@intel.com,`
`woodruff@acm.org`

Abstract. As sensing technologies become increasingly distributed and democratized, citizens and novice users are becoming responsible for the kinds of data collection and analysis that have traditionally been the purview of professional scientists and analysts. Leveraging this citizen engagement effectively, however, requires not only tools for sensing and data collection but also mechanisms for understanding and utilizing input from both novice and expert stakeholders. When successful, this process can result in actionable findings that leverage and engage community members and build on their experiences and observations. We explored this process of knowledge production through several dozen interviews with novice community members, scientists, and regulators as part of the design of a mobile air quality monitoring system. From these interviews, we derived design principles and a framework for describing data collection and knowledge generation in citizen science settings, culminating in the user-centered design of a system for community analysis of air quality data. Unlike prior systems, ours breaks analysis tasks into discrete mini-applications designed to facilitate and scaffold novice contributions. An evaluation we conducted with community members in an area with air quality concerns indicates that these mini-applications help participants identify relevant phenomena and generate local knowledge contributions.

Keywords: Air quality monitoring, citizen science, environmental science, mobile sensing, participatory sensing, qualitative studies.

1   Introduction

Due to the increased availability of sensing technologies, citizens and novice users have new opportunities to pursue the kinds of data collection and analysis that were once handled almost exclusively by professional scientists and analysts [5]. Leveraging this citizen engagement effectively, however, requires not only tools for data collection but also mechanisms for understanding and utilizing citizens' "local knowledge" – the experiential and cultural context, insights, and expertise unearthed through collaboration between locals and experts [4]. For example, while sensing

Fig. 1. A personal air quality sensor (left). Community members with sensors (right).

systems may be able to detect the presence of a pollution source, local insight may be required to actually identify the source or reveal sensitive populations affected by it.

Currently, most tools for viewing and analyzing sensed data do not explicitly support collaboration and are not designed to elicit or compile these kinds of local questions and insights. Moreover, analysis tools are generally not accessible to novice users, since they tend to assume a high level of technical and scientific literacy. We seek to understand how interactive systems for supporting citizen science can facilitate input from novice users and provide scaffolding that allows them to make greater local knowledge contributions.

This research is one component of the Common Sense project [1][8], a mobile sensing program that aims to deploy distributed air quality sensors in the service of practical action. Whereas traditional air quality monitoring organizations utilize coarse, representative measurements from a relatively small network of fixed sensors, we advocate a complementary mobile participatory sensing [3] approach in which large numbers of personal, mobile sensors are deployed within communities. This approach allows the community members impacted by poor air quality to engage in the process of locating pollution sources and exploring local variations in air quality. It leverages citizens' desire to understand personal exposure and knowledge of their communities to help effect change. We have developed a research testbed to explore this approach, examining issues such as the relative accuracy and resolution of community-sensed data versus data collected in professional fixed installations. The project also focuses on developing models for facilitating engagement and cooperation between community members, citizen scientists, activists, and other stakeholders.

In this paper, we survey related work in citizen sensing, collaborative visual analysis, and air quality presentation. We then discuss our own research, focusing on four key contributions: First, we present principles for designing for novice users in a citizen science setting, based on the results of extensive interviews with community members and other stakeholders in the air quality ecosystem. Second, we propose a framework for describing the process of local knowledge creation in citizen science. Third, we demonstrate the Common Sense Community site, a set of collaborative web-based visual analysis tools designed to facilitate collaborative analysis of sensed data and the co-production of local knowledge. Unlike prior systems, ours breaks analysis tasks into discrete mini-applications designed to facilitate and scaffold novice contributions. Finally, we present an evaluation of an early prototype of the site that

indicates these mini-applications help participants identify relevant phenomena and harness local insights.

2 Related Work

Citizen science and community environmental monitoring efforts have a deep and varied history that has been well documented in the environmental justice literature, illustrated by numerous examples of "backpack studies" and volunteer monitoring programs [4]. These examples have demonstrated the effectiveness of community participation in the collection of environmental data. O'Rourke and Macey discuss the use of "bucket brigade" sampling in which a mix of participants in different roles coordinate to carry out observation, sampling, and analysis of refinery emissions [25]. Other work has documented the use of community air quality sensing to identify polluters and enforce standards for diesel bus emissions [21][19]. This citizen-centric ethos has also begun to surface in government monitoring programs for water quality and waste [11].

Interactive tools for collaborative visual analysis may help community members and experts analyze community-sensed data, but the design of these tools presents numerous challenges [13]. Web-based tools like sense.us [14] and Many Eyes [30] have sought to facilitate collaboration using free-text comments attached to visualizations. However, this work has typically addressed short-term exploratory analysis of small datasets, rather than the long-term, iterative analysis associated with environmental monitoring. Meanwhile, commercial products for collaborative visual analytics [27][28] are targeted at expert analysts and are not accessible to novice users. Luther et al.'s Pathfinder [20] is perhaps the closest to our work. It seeks to utilize collaboration and visualization tools to support citizen science, but focuses on small datasets and wiki-based collaboration.

A number of projects have mapped air quality data using mobile sensors, typically with an emphasis on improving environmental awareness [9][17][26]. Some have taken creative approaches to presenting and collecting this data through artful visual presentation [6], provocative platforms [7], and gameplay [2]. While we also provide web-based tools to visualize data from mobile sensors, our tools focus instead on facilitating more direct engagement in the process of data analysis.

3 Motivating Fieldwork

Before deploying our mobile sensing platform with community members, we wanted to understand how those members factor into discussions about air quality and what roles they could play in data collection, analysis, and outreach. To gauge this, we conducted a concentrated investigation of the communities we hoped to engage with.

3.1 Method

Over the course of several months, we interviewed novice community members as well as scientists, remediation consultants, government representatives and other stakeholders in order to understand their perspectives on air quality and assess the role

that technological interventions could play in their environmental decision-making processes [1]. This included 14 formal, in-person interviews and approximately 30 informal interviews conducted either in person, by phone, or at community meetings. In these interviews, we discussed existing practice and used prototype sensors and interface mockups to explore people's reactions to potential mobile sensing tools. We recorded the formal interviews and took detailed field notes describing all of our interactions. Using these, we performed affinity clustering to identify a general set of emergent themes and design principles. We also performed more targeted clustering to identify common user needs, tasks, and motivations for community participation and engagement with environmental data.

3.2   Personas

Based on this fieldwork, we developed a set of personas to characterize the relevant stakeholders and identified a set of common tasks and questions associated with each. Because the system presented here is targeted primarily at community members and novice users, we will limit our discussion to the three most relevant personas: an *activist* or *community organizer* responsible for orchestrating actions and publicizing environmental issues, a *browser* who has an interest in environmental quality but is not directly involved with sensing, and a novice community member who might act as a *data collector* (Table 1). While we focus here on tools for these community members and novice users, it is also clearly valuable to provide tools for (and promote dialog with) other expert stakeholders with different needs, such as *scientists* and *government regulators*.

Table 1. Some of the key personas derived from our inital fieldwork

	Activist/Organizer	Browser	Data Collector
Motivation	Specific concerns about the community with an emphasis on political change.	Likely to be interested in environmental and/or societal issues. Possibly concerned with political change.	Likely to have personal health issues.
Goals	Prove there is a problem. Determine neighborhood exposure. Pursue political change.	Understand broader environmental and societal impacts. See trends.	See personal, immediate data. Modify personal behavior. Pursue political change.
Desired Tools	Tool for community understanding and presentation.	Summaries, Interactive tools for exploring data.	Glanceable s ummaries, Alarms, Forecasting.

3.3   Design Principles

Based on our fieldwork, we also extracted a set of design principles for developing tools to support visual analysis of sensed data. Some of the key issues are:

Support specific, goal-directed tasks.   Participants were highly goal-oriented and motivated by specific issues such as "What is my personal exposure throughout the

day?" or "What are hotspots in this area?". "General" exploration did not tend to engage them. As one interviewee put it, "You don't want to look at the interface and say, 'What is this supposed to tell me?'"

Show local and personally relevant data. Participants were most interested in data close to their homes and other locations they frequented, rather than the aggregate regional data typically provided by current air quality monitoring solutions. The interviews further suggest that many users may not engage unless they are driven by health concerns or some other issue that personally connects them to the data. As one participant said, "Make the data as local as possible. People want to see their house, their block, not a general neighborhood, not a general area."

Elicit latent explanations and expectations. Community members have local knowledge and expertise, such as beliefs about sources of pollution in their neighborhood. However, our interviews suggest that it is often difficult for them to translate this knowledge into specific queries. While community members were good at generating high-level or vague questions (e.g. "How does the freeway impact air quality?"), they had fewer immediate instincts about how to break these questions down. Therefore, it is important to provide tools that help community members draw on their personal knowledge, for example by making suggestions about possible formulations of queries or by guiding them in their exploration of the data.

Prompt realizations. As mentioned above, community members have significant local knowledge that could be helpful in interpreting local environmental data. Accordingly, it is valuable to present views of the data that are perceptually suggestive of various possible patterns, and therefore prompt spontaneous realizations that draw on the users' local knowledge. For example, a view that aligns readings from multiple days may prompt a user to realize that repeated spikes at a site are the result of a recurring event – for example, a delivery truck unloading.

Beware of "language" barriers. Current tools to which community members have access, such as the EPA EnviroMapper [11], are technically complex and require a moderate level of scientific knowledge (for example an understanding of pollutant concentrations in parts per million). Novice users may benefit from scaffolding to introduce scientific language, and tools that target novice users should not require an understanding of such language.

"You don't want to be inundated." Understandably, participants did not want to be overwhelmed with unnecessary information and complexity (particularly if the information was somewhat new to them or was beyond their level of expertise). Therefore, staged or gradual presentation of information is desirable.

3.4 Framework

Drawing on our personas and design principles, we derived a framework for describing data collection and local knowledge generation in a citizen science setting. This framework does not just describe the existing ecosystem or citizen science applications. Rather, it builds on the key findings and user needs we indentified in our

fieldwork and describes operations an ideal citizen science solution might address. As such, the framework serves as a potential blueprint for designing new citizen science tools and for assessing existing ones.

In this framework, we divide the process of collecting, analyzing, and synthesizing environmental data and local insights into six phases: *collect, annotate, question/observe, predict/infer, validate,* and *synthesize.* While these phases can build on one another, they are not necessarily linear and individual participants do not necessarily participate in all of them. Rather, each involved stakeholder may engage in the process at a few phases and the various members of the community together carry out activities at all phases. The various phases each serve different functions and can build on one another but do not always do so. These phases may also be iterative - for example, answering questions and validating predictions may require additional data collection.

The phases detailed here dovetail with formulations of the scientific method, and some steps (*question, predict*, and *validate*) echo the question-hypothesize-test formulations seen in the science education literature. However, our framework describes a more general set of operations, many of which need not necessarily be formulated in the language of scientific discourse. Questions, predictions, and inferences generated by community members are often pre-scientific and can contribute valuable insights that inform a more formal and rigorous process of scientific analysis without necessarily being framed as such.

Finally, while we frame this process in terms of air quality monitoring for the sake of this discussion, the framework itself is applicable to a broad range of citizen science projects including other environmental and health monitoring efforts.

Collect

In this phase, data collectors engage in various collection activities. These may include using sensors to record raw data or observing phenomena and making manual observations (as in traditional citizen science activities like the Christmas Bird Count [24]). Most existing citizen science places a strong emphasis on this collect phase.

Annotate

After data has been recorded, data collectors provide additional insights that contextualize and supplement it. This can include additional information that helps explain the data; for example, if a peak in the data corresponds to an event they observed during collection. Collectors can also include information about the data gathering process (when, where, and under what conditions was the data collected) or comments about data quality.

Question/Observe

Using their own data and data collected by other participants, data collectors (as well as browsers and activists) can begin to ask basic questions and identify trends. These questions can be introspective ("What is my personal exposure to pollutants?", "Is air quality bad at my home?") or generally inquisitive ("Where is air quality good and bad?", "Are there block-by-block trends in air quality?"). Some of these questions, including those dealing with personal exposure, can often be answered directly using the collected data, while others are more abstract. These questions can be implicit or

explicit and may be driven by the data or by existing assumptions and expectations. Users may also observe and note apparent trends (for example, higher levels of a pollutant at different times of day) or other phenomena of interest (high levels at an unexpected intersection).

Infer/Predict

Building on these questions and observations, data collectors, browsers, and activists can begin to make predictions and inferences about the observed phenomena ("I think values will get worse towards this intersection.", "Higher readings here seem to indicate a source."). The observations and inferences made by community members may be less clearly articulated than in a formal analysis, but can contain local insights. While this phase often resembles the "hypothesize" stage seen in formulations of the scientific method, participants' predictions and insights may not necessarily be framed as clearly testable hypotheses. They may only suggest the existence of a trend or its repeatability rather than proposing a mechanism for it. In these predictions, regardless of their precise formulation, lie some of the most important pieces of local knowledge that community members can contribute.

Validate

At this phase, contributions from data collectors are more likely to overlap with those of activists and organizers. Here, data collectors, browsers, and organizers may look for additional data to corroborate their own findings and organizers may also make requests for additional data. Additionally, organizers may enlist the help of outside entities including domain experts and professional analysts to help verify insights and predictions generated by collectors and browsers.

Synthesize

At the highest level, activists and organizers must integrate the data and knowledge generated in prior phases to produce documentation, reports and other deliverables. Again, organizers may involve domain experts and professional analysts, along with administrators and regulators, in order to generate summary documentation that can be used to support activism, inform policy decisions, and enforce regulations.

This framework (and particularly the *annotate, question/observe,* and *infer/predict* phases) provides a blueprint for scaffolding novice users' progression from initial elicitation through more involved and integrated questions and contributions. In this paper, we focus on applications that engage novice users and guide them through these initial phases. We defer discussion of validation and synthesis, which tend to utilize more specialized sets of tools for more expert users.

4   The Common Sense Community Site

Building on the framework and our design principles, we designed and built the Common Sense Community site, a suite of task-oriented mini-applications that allow community members to participate in the collaborative analysis of local air quality data. While the site is targeted primarily at novice data collectors in a low-income urban area, it is also designed to be accessible to more specialized participants
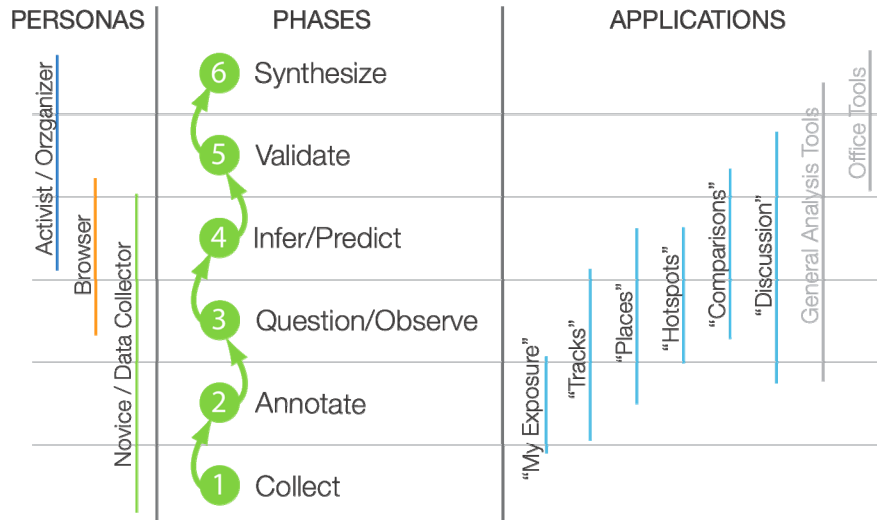
Fig. 2. Our framework for knowledge generation in citizen science (center). Personas (left) and tools (right) are shown in their intended phases.

(browsers, organizers, scientists, administrators, and regulators) who may engage in the analytic process at different phases.

The set of visualizations is designed specifically to facilitate the incremental progression of novice community members through multiple phases of analysis. A person may begin by collecting data or asking questions about data collected by other community members and progress through structured phases, triggering new kinds of insights. Over time, this can allow novices to become more adept contributors.

Providing a suite of simple task-oriented applications rather than a more general analysis tool has several benefits. First, it lowers barriers to entry. Participants do not need to learn a complicated tool in order to contribute. This encourages legitimate peripheral participation [18] and allows novice users and participants with little computing experience to engage in the process. Whereas more general analysis tools such as Excel, Tableau [27], or Matlab require greater familiarity with formal analysis processes, these individual applications allow users to answer specific questions and can guide them towards particular kinds of insights. Figure 2 shows approximate mapping between our mini-applications and the framework discussed previously.

4.1 Collecting Data

Users collect air quality data using mobile sensors designed as part of the broader Common Sense project [8]. These sensors (Figure 1) are designed to be self-contained and unobtrusive monitoring devices that can be clipped to a bag or carried as an accessory. The units feature a custom board design and embedded software that can be deployed with commercial carbon monoxide, nitrogen oxides, and ozone gas sensors. As users carry these sensors with them throughout the day, the units transmit

live sensor reading and GPS data to a database server over a GSM data network connection. Users can also upload data from offline air quality sensors.

4.2 Applications

To display this data, we built mini-visual analysis applications that target common, representative tasks and questions that we identified through our fieldwork. These included: monitoring personal exposure, inspecting recorded tracks, identifying locations with poor air quality, and eliciting possible sources. These targeted applications exemplify our approach to designing for citizen science – modular, accessible applications that serve specific needs and which together scaffold the process of local knowledge production. Users begin by selecting an application that serves a particular need (e.g. "see my personal exposure") from a portal site. They then move between applications via a tabbed interface. We also provide gateways designed to allow participants to build familiarity with simpler, more targeted tools and then transition in a natural way to more complex tools designed to elicit different types of insights. This facilitates the transitions between *annotation* and *questioning* or *questioning* and *inference* we described in our framework.

In each of these applications, users can record their questions and insights by leaving comments attached to individual views of data. Each application features a commenting panel (Figure 4c) that participants can use to annotate and discuss their findings. This panel also provides intelligent prompts designed to elicit questions and observations, along with educational prompts designed to help scaffold novice users' understanding of the domain.

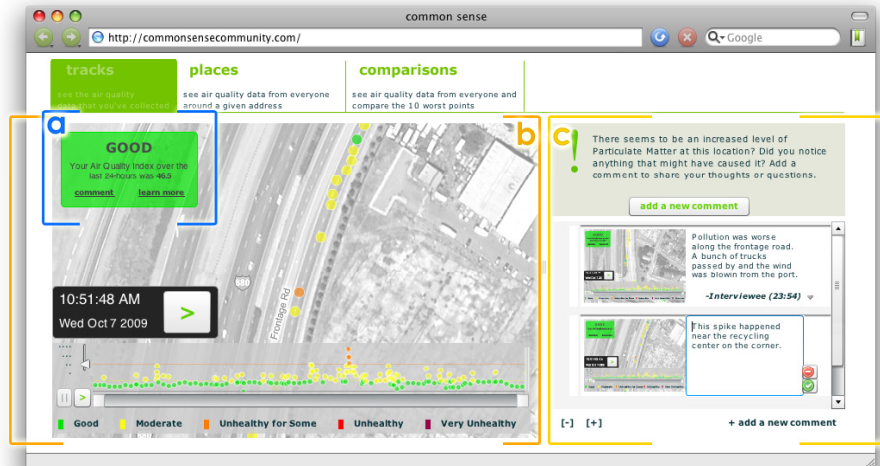We describe several applications in detail below.

Fig. 3. Two views of the *My Exposure* application

My Exposure

The first application provides a widget that helps users answer one of the most common questions we observed in our fieldwork: "What is *my exposure* to a pollutant?" Many of the community members we interviewed suffered from allergies or respiratory disease exacerbated by the poor air quality in their neighborhood, and expressed a desire for tools that would help them gauge and mitigate their exposure. To meet this need, we developed the *My Exposure* widget (Figure 3, Figure 4a). *My Exposure* shows a single aggregated measure of the pollutants measured by a participant's sensor, normalized over time to the EPA's Air Quality Index (AQI) [22]. (Because many people are not familiar with raw pollutant concentrations, all of the visualizations on the site also use the AQI color encodings and category descriptors – "Good", "Moderate", "Unhealthy for Sensitive Groups", "Unhealthy", "Very Unhealthy", and Hazardous" – in addition to providing actual values).

For community members carrying our air quality sensors, this application acts as an entry point to the site and serves an ongoing need that is likely to garner repeat

Fig. 4. The Common Sense Community Site showing data collected by a single user. The *My Exposure* widget (a) and *Tracks* visualization (b) are visible along with the commenting panel (c).

visits. To encourage participants who are initially only curious about their exposure to further explore their data, we placed the *My Exposure* view adjacent to the *Tracks* application (discussed momentarily).

Tracks

The *Tracks* application (Figure 4b) provides a simple way for novice users to observe and ask questions about pollution data from their own sensor. In this visualization, pollution measurements are plotted on a map and also appear in a timeline below the map view. The application behaves like a media player and provides a play/pause button, a playback speed control, and a draggable thumb on the timeline that can be used to scrub back and forth in the dataset.

As mentioned above, in each of our applications, participants use the commenting panel (Figure 4c) to annotate and discuss their findings. This panel is collapsed by default to avoid overwhelming the user, but expands to display intelligent prompts designed to elicit questions and observations. For example, when a participant plays back data from their own sensor in the *Tracks* application, the interface pauses briefly whenever a dramatic spike occurs in the data and actively prompts the user to document the change. The user can choose to either enter a comment or continue playback. If no action is taken, playback resumes after a brief interval. Users can also pause playback at any point to enter comments or questions.

Places

Our fieldwork indicated that users' initial inquiries about air quality are often location-centric ("What is air quality like in *my* neighborhood?", "Are we protecting our 'treasures', our schools, hospitals, libraries, parks, etc.?"). To help facilitate questions and observations of this type, we provide a location-centric *Places* visualization

(not pictured). When a user starts the visualization, they are prompted to enter an address and a time range. The application then produces an interactive map showing all data collected by any sensor near the specified address during those times. Whereas the *Tracks* application is designed to mimic the functionality of a media player, *Places* is designed to feel similar to online mapping tools like Google Maps [12]. The map can be panned and zoomed and the data points plotted on it can be played back chronologically.

We include gateways that allow users to enter the *Places* view from within other applications. When using another application, a user can click a "see more for this location" button to transition to the *Places* view, centered on the location visible in their current application.

Hotspots

The *Hotspots* visualization (not shown) helps users identify regions with the best and worst air quality over a period of time. The application is intended to help users answer questions about where and when levels are high and low. It draws on the notion, frequently seen in our initial interviews, that "worse things are exciting" and uses this to provoke insights regarding new locations and unexpected sources.

Using a range slider, users select whether to show regions with high or low pollution levels. Readings that match the specified thresholds are then plotted on a map similar to the one used in the *Places* view. Users can also transition to this visualization by clicking the "see other places with readings this high/low" gateway from within the *Tracks* or *Places* applications.
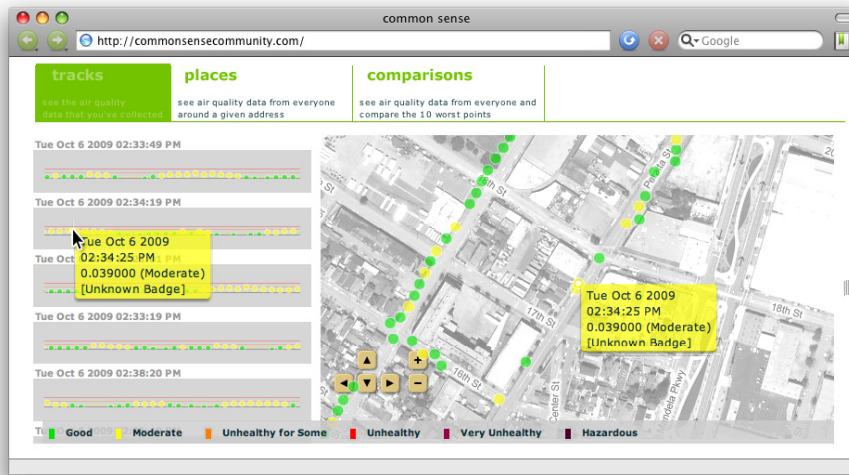
Comparisons

The *Comparisons* visualization is designed to support inference and help users identify repeated sources and relationships between them. The *Comparisons* visualization presents users with a set of discrete 'episodes', short windows of time in which some notable event occurred in the recorded air quality data. These can be the largest spikes seen in an area over the course of a period of time, or the periods of time with the highest variance.

The notion of focusing on spikes was driven by two observations from our fieldwork. First, we noted that people often wanted to "examine an event, not a timeline," seeing detailed data at the scale where the event was apparent, rather than at the level of the entire dataset. Second, we hoped that by grouping together sets of episodes that would otherwise appear separately, this view would prompt noticings and inferences that might not emerge otherwise. In the *Comparisons* view, these episodes are displayed as a set of small multiples [29] alongside a map that also plots that same data (Figure 5). The small multiples are linked to the map so that brushing a plot focuses that event in both views. This allows users to compare the events spatially as well as temporally.

Discussions

In addition to the collapsible commenting pane that accompanies each one of the visualizations, the site features a *Discussions* view – a separate application that serves as a central location for viewing all comments and provides a forum-like interface for

Fig. 5. The *Comparisons* view. Small multiples of the timeline (left) showing the five highest episodes recorded during the past day. The commenting panel is hidden at right.

further discussion. All comments and annotations left by users in the other applications are visible here as separate threads and users can compare and build on observations and insights from multiple applications.

Other Candidate Applications

The visualizations described here cover a large subset of the kinds of questions and observations specified by our framework. However, other visualizations are clearly possible (e.g., tools for understanding variations in air quality over time), and we expect to build examples of them in the future to support other relevant questions.

We also anticipate tools that will assist participants in the *validate* and *synthesize* phases of a citizen science task. For example, applications might include automated pattern matching to help locate sets of similar sources or identify characteristic pollution signatures. Similarly, tools to support "crowdsourcing" could allow organizers to request new samples or ask community members to identify sensitive locations like schools and day care centers.

4.3   Implementation Details

The Common Sense Community site and all of the visualizations within it were constructed using Adobe Flash with the Modest Maps toolkit [23].

5   Evaluation

We deployed an early version of the site with community members in a low-income urban neighborhood with poor air quality. There, we carried out interviews and think-aloud assessments to help characterize participants' use of the tools. We wanted to

understand which visualizations were perceived to be useful and approachable and assess whether this set of tools facilitated activities at the various phases identified in our framework, such as emergent prediction and observation.

5.1 Method

During our assessment we carried out seven interviews with nine community members. We recruited participants through a local non-profit organization that focuses on environmental monitoring and awareness. Five of the participants were affiliated with the non-profit and had participated in air quality monitoring activities through the organization. Most of the participants we surveyed were members of a small and relatively tightly knit community and the majority knew one another in some capacity. Participant ranged in age from the mid-teens to late 40's and had a variety of education levels, including some middle- and high-school students and some participants without high school degrees.

We conducted all of the interviews at the office of the non-profit. We started each session with a brief interview designed to assess participants' knowledge of air quality issues and the impact of air quality on their community. In our discussions, we emphasized the impacts of particulate matter and described its sources. We then gave the participants a particulate matter sensor and asked them to take samples in a several block radius around the office. We asked participants to choose a route that they thought would maximize the amount of particulate matter detected. During the sampling process, the interviewer walked with the participants and asked them to describe their route choice and identify potential sources in the area. We used a commercial particulate matter sensor rather than our custom hardware since particle pollution is of particular interest in the target neighborhood.

Once they returned to the non-profit, participants used an early version of the Common Sense Community site to examine their data as well as data gathered by other participants. We conducted a one-hour think-aloud evaluation with each participant in which they were instructed to interact with the site and verbally relate their thought processes and any questions or insights that occurred to them. Participants used a version of the site that included the *Tracks*, *Places*, and *Comparisons* visualizations detailed above. In the *Places* and *Comparisons* views, each participant had access to his or her own measurements as well as measurements taken by all of the previous participants. Because users only had access to data collected by a small group of participants in short windows over the course of a few days, we were unable to test the *Hotspots* visualization, which was designed to leverage larger datasets.

We recorded each of these interviews and coded participants' interactions with the site to assess whether or not they fit within our framework. We also performed clustering to extract key findings that emerged. These are discussed in Section 6.

5.2 Scaffolding and Navigation Strategies

Most participants were able to explore the visualizations and inspect the data that they had collected without much confusion. The majority began by identifying their current location on the map and followed the track they had recorded, looking for peaks either on the map or in the timeline. Most voiced questions and observations about the

data and a few made additional inferences or predictions. We report key observations that correspond to each of the phases in our framework.

Collect. Almost all of the users identified a nearby freeway and trucking lots as the most likely sources of pollution and most chose routes that took them along a nearby frontage road. The students we interviewed all minded their sensors attentively as they walked, looking for spikes and actively seeking out areas with higher readings. All other participants used the sensor more passively and traversed areas that they predicted would be more polluted without actively noting the levels there.

Annotate. Using the Tracks view, several participants observed distinct peaks and ascribed them to events that occurred or features that they passed while they were walking ("All the trucks [get on the highway there].", "That's the new construction there."). Participants also tended to note readings taken adjacent to locations that interested them ("At least we don't have any red marks near the park…"). In two cases, participants had observed increased particulate matter levels on the sensor as they walked and directly attributed a peak to a particular source.

Question/Observe. Most participants asked questions and made remarks about locations ("Where was that again?"), data ("Was [that spike] at an intersection?"), and other participants ("Where did she go?", "Which person did that come from?"). Participants also asked broader questions about day-to-day and month-to-month trends. For example, one wondered whether pollution levels would change during the rainy season and another asked "Would it be different if there was wind?" A few participants also noted locations on the map without data and contributed additional anecdotes and pieces of information about them.

Infer/Predict. Based on the data and their initial questions and observations, several participants made inferences about the behavior of phenomena they observed. For example, one participant compared her readings with those from a participant earlier in the day and noticed that her own were higher. She inferred that the level of particulate matter might be impacted by the change in temperature.

   Another participant investigated the data he had collected and extrapolated from it to predict air quality readings further along the frontage road saying, "I wouldn't doubt that it gets worse around the bend." Talking about a several-block radius, he also made a prediction about the health impacts of pollutants in the area. He noted, "Just in this radius I can honestly say [...] at least half the kids have asthma. At *least* half." He supplemented this prediction with a quick calculation, "Fifteen residences per area so … that's probably about a good 500 kids."

Validate and Synthesize. This set of interviews involved only novice community members and incorporated only data collected during their sessions. As such, we did not emphasize the validate and synthesize phases in this study.

5.3 Usability

Based on our fieldwork, we were mindful in our design process of the computer literacy of the target population. As one participant in our initial interviews noted,

"There's still that big digital divide in [our city] and all poor neighborhoods." Therefore, we were pleased that the system was generally usable by all participants. The study did reveal a few straightforward usability issues, which we are addressing, such as the need to make the playback controls more visible. These issues did not appear to impact the results discussed below.

6   Discussion

Here we discuss trends and activities we observed across all of our interviews.

6.1   Health and Personal Safety

As expected, displays tailored to personal use proved to be an effective tool for engaging users in the process of citizen science. The most interested and receptive participants each had a personal or family health concern (asthma, allergies, or some other reaction) that they attributed to air quality. One asthmatic participant who bicycles and does not own a car expressed a desire to use the data to vet safe cycling routes, stating, "This has brought to mind – you're gonna get exercise, but what are you breathing in?" Participants with small children also expressed a strong desire to use the tool on a regular basis to help minimize exposure.

6.2   Socializing

Although we conducted interviews separately and the sequential nature of the interviews did not facilitate conversations or dialogues using the commenting tools, we did see social interactions between participants when they viewed one another's data. Several participants asked questions like, "Which person did these come from?" and "Whose was whose?" and were eager to compare their tracks against those recorded by previous participants. In particular, those from the same social circle were interested in knowing which of their friends had collected data, where they'd walked, and how "well" they had done. For example, one participant located a friend's track and followed it for the entire length, noting each location she'd visited and commenting, "She was pretty good, [she found a few orange ones]." Comparing tracks in a competitive way was also common, particularly among the students we interviewed. One group of younger students, for example, was excited to discover that their readings were higher than those of other participants. This suggests a competitive impulse that we might also leverage to encourage participation.

During the interviews, several participants attributed their continued awareness and investment in air quality to a particular community organizer. One participant observed, "You could say *she's* our resource when things are happening. If she feels we need to know, then it's up to us to get involved." This suggests that, at least within this community, maintaining long-term interest and investment depends, in part, on leveraging these kinds of key community members.

While we observed users' reactions to one another's data, the linear nature of our interviews did not allow us to observe exchanges or evolving social use of the system. A longitudinal study with more users is needed to understand these social aspects of

the system and to gauge the impact of larger amounts of data and discussion on the analyses that participants undertake.

6.3   Exposing Preconceived Notions

A number of our participants approached the data not from an inquisitive standpoint, but rather expecting to find validation of their expectations about air quality. We noted comments from a number of participants that suggested implicit assumptions about areas ("On Fourth Street, that makes sense.") and expectations about how bad pollution levels would be ("[If you sampled this area] you'd see lots of red"). One participant, in particular, was surprised that the level of particulate matter she recorded was low, stating, "I feel like it should be a little stronger with picking up certain particulates and fumes. I *know* there should be a lot more out there because there are a lot of businesses and industrial stuff." To test this, the participant requested to take the sensor out again and collected additional data.

In some cases these kinds of assumptions may function as implied hypotheses and predictions that participants can immediately begin to validate and build on. However, as in the case of the latter participant, preconceptions can sometimes generate mistrust in sensors and tools that do not reinforce these existing notions.

6.4   Visualizations as a Catalyst for Discussion

We also observed several participants who used the map extensively as a catalyst for discussion. These users would point and navigate to areas with strong personal relevance including their homes, schools, and public areas, even when no air quality data for that particular region was present.

One interviewee, in particular, used the map to discuss pollution sources outside the zone in which he had collected data and to make predictions about sources and impacts there. He first predicted that there might be "really high values" in main intersections adjacent to a nearby port and shipping terminal, stating, "I can only imagine [it gets worse toward the intersections.]" He then contributed a number of anecdotes about locations in and around the port including spots where diesel trucks idle, areas where water quality has been impacted by dredging, and an isolated residential building in the industrial zone. These anecdotes were often very specific and drew on his experience as a port worker and volunteer air monitor – for example:

> "Here - definitely this intersection - we did some of the survey in this area last year. Here, right here - this is a fuel station. It's a truck fuel station. This is where all the trucks get on the freeway. All the trucks are always right here - along [Street 1] and [Street 2] and um, [Street 3] and [Street 2]. I know for sure, these monitors are not going to catch moderate here. Lucky enough, nobody lives on these blocks. All business, all industry."

These kinds of observations are key examples of the types of local insights community members may bring to the table and which we hope to elicit.

7 Conclusions and Future Work

In this paper, we have presented design principles for targeting novice users in a citizen science setting and supplied a framework that describes data collection and knowledge generation in these conditions. We have described the genesis of this model through interviews with community members and activists, as well as its application in the user-centered design of a system for mobile air quality monitoring. Unlike prior systems, ours breaks analysis tasks into discrete mini-applications designed to facilitate and scaffold novice contributions. Based on our initial evaluations, this strategy helps novice users identify relevant phenomena and generate local knowledge contributions.

Although the applications discussed here focus on air quality, we believe that the approach we advocate can be applied to other domains with a citizen science component. Monitoring of other environmental indicators including water and soil quality as well as epidemiological monitoring should be equally applicable, particularly when the object of study is of strong significance to the participants. As we move forward to deploy mobile sensors more broadly and develop mobile interfaces for accessing and interacting with the data, we expect to employ similar techniques and build on these frameworks and tools.

Acknowledgements

References

1. Aoki, P.M., Honicky, R.J., Mainwaring, A., Myers, C., Paulos, E., Subramanian, S., Woodruff, A.: A Vehicle for Research: Using Street Sweepers to Explore the Landscape of Environmental Community Action. In: Proc. CHI 2009, pp. 375–384. ACM, New York (2009)
2. Black Cloud, `http://www.blackcloud.org`
3. Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.B.: Participatory Sensing. In: SenSys 2006 WSW Wksp. ACM, New York (2006)
4. Corburn, J.: Street Science: Community Knowledge and Environmental Health Justice. MIT Press, Cambridge (2005)
5. Cuff, D., Hansen, M., Kang, J.: Urban Sensing: Out of the Woods. Comm. ACM 51(3), 24–33 (2008)
6. Da Costa, B., Schulte, J., Singer, B.: AIR – Area's Immediate Reading, `http://www.pm-air.net`
7. Da Costa, B., Hazegh, C., Ponto, K.: PigeonBlog, `http://pigeonblog.mapyourcity.net`
8. Dutta, P., Aoki, P.M., Kumar, N., Mainwaring, A., Myers, C., Willett, W., Woodruff, A.: Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors (demonstration). In: Proc. SenSys 2009, pp. 49–50. ACM, New York (2009)

9. Eisenman, S.B., Miluzzo, E., Lane, N.D., Peterson, R.A., Ahn, G.S., Campbell, A.T.: The BikeNet Mobile Sensing System for Cyclist Experience Mapping. In: Proc. SenSys 2007, pp. 87–101. ACM, New York (2007)

10. Environmental Protection Agency: Volunteer Monitoring | Monitoring and Assessing Water Quality, `http://www.epa.gov/volunteer/`

11. EPA EnviroMapper, `http://www.epa.gov/enviro/html/em/index.html`

12. Google Maps, `http://maps.google.com`

13. Heer, J., Agrawala, M.: Design Considerations for Collaborative Visual Analytics. Information Visualization 7(1), 49–62 (2008)

14. Heer, J., Viégas, F., Wattenberg, M.: Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. In: Proc. CHI 2007, pp. 1029–1038. ACM, New York (2007)

15. Honicky, R.J., Brewer, E., Paulos, E., White, R.: N-SMARTS: Networked Suite of Mobile Atmospheric Real-Time Sensors. In: SIGCOMM 2008 NSDR Wksp. ACM, New York (2008)

16. Irwin, A.: Citizen Science: A Study of People. In: Expertise and Sustainable Development. Routledge, London (1995)

17. Kanjo, E., Benford, S., Paxton, M., Chamberlain, A., Stanton Fraser, D., Woodgate, D., Crellin, D., Woolard, A.: MobGeoSen: Facilitating Personal Geosensor Data Collection and Visualization Using Mobile Phones. Pers. Ubiq. Computing 12(8), 599–607 (2008)

18. Lave, J., Wenger, E.: Situated Learning: Legitimate Peripheral Participation. UP, Cambridge (1991)

19. Levy, J.I., Houseman, E.A., Spengler, J.D., Loh, P., Ryan, L.: Fine Particulate Matter and Polycyclic Aromatic Hydrocarbon Concentration Patterns in Roxbury, Massachusetts: A Community-Based GIS Analysis. Environmental Health Perspectives 109(4), 341–347 (2001)

20. Luther, K., Counts, S., Stecher, K., Hoff, A., Johns, P.: Pathfinder: An Online Collaboration Environment for Citizen Scientists. In: Proc. CHI 2009, pp. 239–248. ACM, New York (2009)

21. Minkler, M., Wallerstein, N. (eds.): Community Based Participatory Research for Health. Jossey-Bass, San Francisco (2003)

22. Mintz, D. (comp.): Technical Assistance Document for the Reporting of Daily Air Quality - the Air Quality Index (AQI). Tech. Research Triangle Park, U.S. Environmental Protection Agency (2009)

23. Modest Maps, `http://modestmaps.com`

24. National Audubon Society. History & Objectives. Christmas Bird Count, `http://www.audubon.org/bird/cbc/history.html`

25. O'Rourke, D., Macey, G.: Community Environmental Policing: Assessing New Strategies of Public Participation in Environmental Regulation. Journal of Policy Analysis and Management 22(3), 383–414 (2003)

26. Paulos, E., Honicky, R.J., Hooker, B.: Citizen Science: Enabling Participatory Urbanism. In: Foth, M. (ed.) Handbook of Research on Urban Informatics, pp. 414–436. IGI Global, Hershey (2008)

27. Tableau Server, `http://www.tableausoftware.com`

28. TIBCO, Spotfire Decision Site, `http://spotfire.tibco.com`

29. Tufte, E.: Envisioning Information. Graphics Press, Cheshire (1990)

30. Viégas, F.B., Wattenberg, M., van Ham, F., Kriss, J., McKeon, M.: ManyEyes: a site for Visualization at Internet Scale. IEEE Transactions on Visualization and Computer Graphics 13(6), 1121–1128 (2007)