Particulates Matter: Assessing Needs for Air Quality Visualization

Aspen Hopkins * University of Utah

Pascal Goffin † University of Utah

Miriah Meyer ‡ University of Utah

Abstract

We present the results of a qualitative study aimed at improving current understanding of the needs and desires of Salt Lake City (SLC) residents in regards to air quality. Currently, government websites provide only coarse data. With the opportunity to get more fine grained data it is important to provide a system that supports the interests of the general public. In addition, we also provide a design mockup of how such a visualization system could look like based on the results of our study.

Index Terms: air quality, sensor data, thematic analysis

1 Introduction

The World Health Organization estimated that, in 2012, approximately 3.7 million people died as a result of ambient air pollution, and by 2014 only 1 in 10 people world-wide breath clean air [9]. Among common air pollutants, fine particulate matter ($PM_{2.5}$) – particles with diameters smaller than 2.5 microns – has the greatest adverse health effects. In response to increasing health and economic concerns related to elevated $PM_{2.5}$ levels , citizens and grassroots organizations are taking advantage of new fixed-location, low-cost commodity sensors to better understand their local air quality [1].

A new project at the University of Utah is making use of low-cost sensors to densely instrument SLC, which periodically experiences the worst air quality in the nation [6]. The goal of the project is to provide residents, organizations, and government officials with real-time, high-resolution estimates of $PM_{2.5}$ levels. The estimates are produced from a statistical model that combines sensor measurements with geospatial land-use information to produce a real-time, 2D continuous field of $PM_{2.5}$ levels. These $PM_{2.5}$ estimates will be publicly accessible through a web-based visualization interface.

2 Related Work

Currently, air quality data is represented as either discrete measurements at specific geospatial locations, or as a continuous field interpolated from discrete measurements. For the former, visualizations often color-code $PM_{2.5}$ measurements using landmarks on a map with details-on-demand linked views – an example of this is the PurpleAir visualization shown in Fig. 1(a). By clicking on a marker either a sidebar [5] or a tooltip [8] shows temporal details of the sensor measurements along with weather data such as temperature, atmospheric pressure and wind. Similarly, Willett et al. [7] show discrete sensor measurements along a person's trajectory during the day – in the project, people carried air quality sensors with them around a neighborhood. Furthermore, Hsu et al. [3] use a map to show the locations of six air quality sensors around a coke plant, with an additional view to show the evolution of $PM_{2.5}$ counts.

The second type of air quality data – a continuous field – overlays a geographical map with a colormap of air quality levels. One such example is the EPA's AirNow visualization [4], shown in Fig. 1(b).

*e-mail: akh0901@westminstercollege.edu

†e-mail: pascal.goffin@sci.utah.edu

‡e-mail: miriah@cs.utah.edu

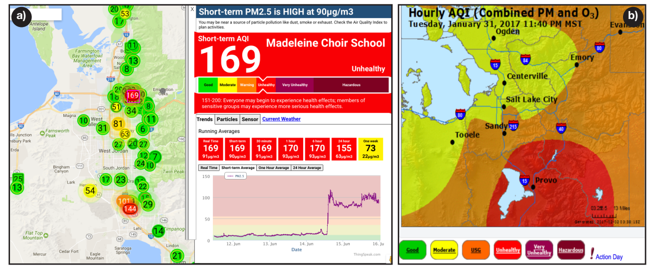

Figure 1: Purple Air [5] using discrete measurements at specific locations (a), AirNow [4] using a continuous map with color scale (b).

These government-based visualizations, however, tend to rely on measurements taken from just a few gold-standard sensors placed over a large geographic region. For example, the interpolated air quality data shown in Fig. 1(b) has 10 sensors for an approximately $\approx$9500 square mile region. In contrast, the continuous fields in this project will be interpolated from roughly 500 commodity sensors placed in a $\approx$115 square mile region. As such, the $PM_{2.5}$ estimates will capture air quality microclimates never before seen by SLC residents, such as potential hotspots around interstates during rush hour and in the vicinity of polluting industries. This high-resolution data offers an opportunity for new visualization approaches to engage residents about air quality. The formative study described in this poster is a first step to assess visualization possibilities.

3 Methods

We conducted six semi-structured interviews to discover assumptions and motivations regarding air quality among SLC residents. The participants were selected using a convenience sampling of the social network of one of the authors (4 participants) as well as a from a local coffee shop (2 participants). We focused on selecting participants that diversify relevant characteristics for our study: gender identity (3 female, 1 genderfluid, 2 male); income (ranging from \$20k per year to \$310k per year); race (3 white, 2 asian, 1 black); education (2 some college, 3 college degree, 1 graduate degree); and activities outside (4 exercise regularly, 3 walk dogs, 4 participate in mountain activities such as skiing, 2 cycle). Interviews took from 15 to 35 minutes and were completed during the months of May and June 2017. All interviews were voice-recorded.

We transcribed the interviews by using a combination of quotes and interpreted summarizations. To analyze these interpreted transcriptions, we used thematic analysis [2] to first code, then generate themes. Two of the authors separately coded 3 of the interviews, then compared resulting codes to ensure validity of results – the remainder of the interviews were coded by just 1 of the authors.

Following the coding process, all 3 authors generated a first round of low-level themes for 25% of the codes, with the remainder of the codes analyzed by just 1 author. The forty-one low-level themes were then iteratively grouped into 7 mid-level themes by all authors, with a final round of analysis producing 3 high-level themes.

We selected a prevalent low-level code to build upon in the creation of a low-fi visualization prototype. This code was chosen for the consistent interest expressed by the interviewees, the cross-cutting nature of the code to the identified themes, and the potential for novel visualization solutions.

4 Emergent Themes

Through the process of thematic analysis, seven mid-level themes emerged: *data for good*, *personal empowerment*, *quality of life*, *learning about air quality*, *impacts on air quality*, *data accessibility*, and *visualization ideas*. These themes are grouped into three high-level themes: *motivations*, *understanding*, and *design*.

The **motivation** themes focus on the reasons why residents care about having access to air quality information. **Data for good** emphasizes the needs of individuals who wish to use air quality data to improve their communities and society, such as the environmental advocacy, green technology, and policy. Similarly, **personal empowerment** captures the needs of individuals who want to make more informed decisions based on the air quality. One participant said *"The biggest factor for me is running and the help it might be in planning the day to day... If it showed maybe an area where there is better air quality, it could definitely affect where Id choose to live."* A third motivating theme is **quality of life**, particularly with respect to the impact of poor air quality on health and economics. The motivation themes may provide guidance on target user groups for engaging with the visualization and evaluating its efficacy.

Understand themes capture the assumptions and questions that residents have about the science, analysis, and communication of air quality levels. **Learning about air quality** was particularly prevalent as participants had a variety of questions regarding the science of air quality and its impact on health, suggesting a need for a trusted source of information. General assumptions stated within the interviews further supported the need for reliable data. One participant said: *"From a pure curiosity point of view, I would like to know what the distribution is. Its probably a bad assumption to think that it is homogeneous. And then if one thinks it is due to cars, itd be good to see if the pollution is worse say along State Street as compared to maybe the center of the university, where there is not a lot of automobiles. Or if there are things that are spewing out garbage, Id like to know that too, and for gods sake let's shut them down."* There was also misinformation about air quality. Related to learning, **impacts on air quality** were often discussed by participants, from the impact of development and weather to vegetation. The understand themes highlight the need for educational material within the visualization interface, as well as for illustrative examples that educate residents on how to interact and engage with the air quality data.

The final class of themes, those that could directly inform the **design** of the visualization, consisted of the wants and recommendations of residents. The **visualization ideas** theme includes a desire to understand microclimates, opinions regarding the EPAs Air Quality Index, and requests to visualize the relationship of elevation to air quality. Elevation in particular proved to be interesting, as existing air quality data sources are not fine-grained enough to capture the role that elevation plays in the levels of $PM_{2.5}$. As one participant mentioned, *"I'd like to know exactly the altitude of [where the air quality becomes good]. I want to know where it is. So if I needed to get out of it, I know where would I go."* Another design theme is **accessibility**, which focused on different user groups and their ability to effectively use a tool to gather information. Ideas ranged from at-a-glance information to access to the underlying raw data. The design themes are not all compatible in a single tool, but taken together describe a rich space of requirements for designing air quality visualizations to a broad set of users.

5 Visualization Design

While several codes pulled from the interviews could be connected to a visualization system, one in particular stood out in regards to being regularly mentioned, and in the unique perspectives it offered. The relationship of elevation to $PM_{2.5}$ levels is currently missing from the existing coarse-grained data, is generally known among SLC residents to be important, and offers unique visualization challenges.

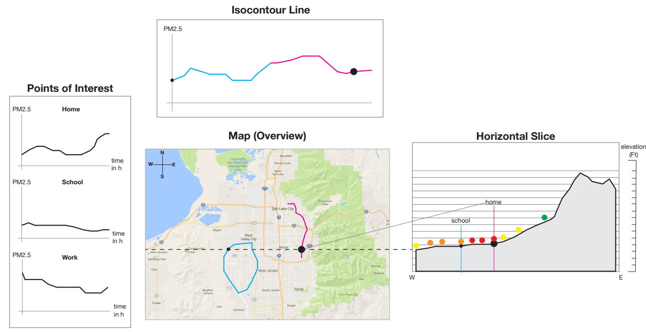

Figure 2: Visualization System Mockup.

Previous attempts at representing air quality data has consistently involved the use of a two-dimensional scalar map, however this view is limiting to users that wish to understand the relationships between elevation and air quality. As an alternative, we propose a system wherein the user has four linked views (see Fig. 2): the first is the standard, two-dimensional map with options to select any point of interest, the second is an associated, transverse profile of the Salt Lake Valley, the third is a small multiples grouping of an individuals points of interest, and the fourth is a line graph of $PM_{2.5}$ values along an isocontour. The map provides an overview of the data, and each additional linked view gives further insight into the data.

The transverse profile represents elevation through a horizontal slice of SLC. The user selects points along the profile to produce line graphs of both historic and real-time data on the PM levels. Users may choose multiple points, allowing them to compare air quality data across the valley. The small multiples allow users to essentially bookmark places that are important to them, for the sake of monitoring the values. The final view, a line graph for the isocontours, gives users a sense of the data, creating an environment conducive to self-directed exploration.

6 Conclusion and Future Work

We presented a first set of design considerations for interacting with air quality data based on the thematic analysis of multiple interviews with SLC natives. We focused on representing elevation, air quality, and microclimates for the diverse community needs of SLC. As future work, we plan to integrate the aforementioned views into an environment for public use and evaluate these approaches with end users.

References

[1] P. Aoki, A. Woodruff, B. Yellapragada, and W. Willett. Environmental protection and agency: Motivations, capacity, and goals in participatory sensing. In *Proc. of CHI*, pp. 3138–3150. ACM, 2017.

[2] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[3] Y.-C. Hsu, P. Dille, J. Cross, B. Dias, R. Sargent, and I. Nourbakhsh. Community-empowered air quality monitoring system. In *Proc. of CHI*, pp. 1607–1619. ACM, 2017.

[4] Office of Air Quality Planning and Standards (U.S. EPA). AirNow, Last read: June 2017.

[5] PurpleAir. PurpleAir Air Quality Map, Last read: June 2017.

[6] M. Shenefelt. Standard Examiner: Northern Utah cities have nation's worst air pollution for second straight day, feb 2016.

[7] W. Willett, P. Aoki, N. Kumar, S. Subramanian, and A. Woodruff. Common sense community: scaffolding mobile sensing and analysis for novice users. In *Proc. of Pervasive Computing*, pp. 301–318. Springer, 2010.

[8] World Air Quality Index. Worldwide air quality, Last read: June 2017.

[9] World Health Organization. Ambient air pollution: A global assessment of exposure and burden of disease, 2016.