The following is an author's accepted manuscript of an article that appeared in Cartography and Geographic Information Science in 2017. The final publication is available through Taylor & Francis at http://www.tandfonline.com/doi/abs/10.1080/15230406.2016.1276479

A geovisual analytics exploration of the OpenStreetMap crowd

Sterling D. Quinn

Department of Geography, Central Washington University, Ellensburg, Washington, USA

Alan M. MacEachren

Department of Geography, The Pennsylvania State University, University Park, Pennsylvania, USA

sterling.quinn@cwu.edu, macheachren@psu.edu

A geovisual analytics exploration of the OpenStreetMap crowd

It is sometimes easy to forget that massive crowdsourced data products such as Wikipedia and OpenStreetMap (OSM) are the sum of individual human efforts stemming from a variety of personal and institutional interests. We present a geovisual analytics tool called Crowd Lens for OpenStreetMap designed to help professional users of OSM make sense of the characteristics of the "crowd" that constructed OSM in specific places. The tool uses small multiple maps to visualize each contributor's piece of the crowdsourced whole, and links OSM features with the free-form commit messages supplied by their contributors. Crowd Lens allows sorting and filtering contributors by characteristics such as number of contributions, most common language used, and OSM attribute tags applied. We describe the development and evaluation of Crowd Lens, showing how a multiple-stage user-centered design process (including testing by geospatial technology professionals) helped shape the tool's interface and capabilities. We also present a case study using Crowd Lens to examine cities in six continents. Our findings should assist institutions deliberating OSM's fitness for use for different applications. Crowd Lens is also potentially informative for researchers studying Internet participation divides and ways that crowdsourced products can be better comprehended with visual analytics methods.

Keywords: OpenStreetMap; volunteered geographic information; crowdsourcing; geovisual analytics; scenario-based design

Introduction

In the past two decades the increasing availability and interactivity of the Internet have facilitated the creation of enormous stores of crowdsourced information. Projects such as Wikipedia, Yahoo! Answers, and OpenStreetMap (OSM) rely on the collective experiences of contributors to improve the information product. Although it is easy to think of the crowd as amorphous and "out there" somewhere, each piece of information can be linked to an individual contributor in a physical location, and by the same logic, any large body of information can be traced to a finite contributor set.

These crowdsourced projects generate large, heterogeneous, and often messy data; thus, data for which visual analytics methods can facilitate sensemaking about the nature of the crowd as well as about the data's fitness for use. For example, with OSM, answers to the following might help users understand how much credibility can be placed in the map:

- How many contributors generated the map, and how does this number vary over time or space?
- What backgrounds and motivations do contributors bring to the work?
- What is the balance of hobbyists vs. contributors who map for a corporation?
- How do local and remote contributions compare?
- Who is likely to edit the data in the future?
- To what degree are edit wars, vandalism, and automated edits influencing the map?

Understanding each person's contribution can be cognitively and computationally challenging due to the amount of data and the number of individuals involved. Since many data items have multiple versions resulting from change over time, the complexity can even apply to a single entity such as a highway in OSM or an article in Wikipedia. Furthermore, obtaining the raw historical data archives can require technical skill and computing resources, as much of these data are in semi-structured text-based formats.

This paper describes *Crowd Lens for OpenStreetMap*, a visual analytics tool we designed, implemented, and evaluated for learning about construction of OSM in a place. Like Wikipedia, OSM is built online by volunteers; however, OSM is focused on geographic information. Any contributor, whether local or remote, can add or modify geographic features and adjust feature attributes using a set of community-defined "tags". OSM boasts over 2 million registered contributors (http://wiki.openstreetmap.org/wiki/Stats), and OSM coverage has come to rival commercial and government-produced alternatives for reference mapping in some areas of Europe and the United States (Haklay, 2010; Neis, Zielstra, & Zipf, 2011; Graser, Straub, & Dragaschnig, 2014), with more varied coverage in other parts of the world (Neis, Zielstra, & Zipf, 2013).

The dimensions of geographic and attribute space differentiate OSM from other purely text-based crowdsourced projects such as Wikipedia, and make OSM particularly interesting for visual analytics. The goal of understanding who is mapping what, from where, combined with a visual interface to computationally processed messy data, puts Crowd Lens in the category of "geovisual analytics" tools proposed by Andrienko et al. (2007).

Crowd Lens is place-based; it is a proof-of-concept application that demonstrates the potential of geovisual analytics to support decisions about OSM data use (thus it is not an enterprise system engineered to scale to all of OSM at once). We demonstrate its capabilities using six pre-processed towns of roughly the same size from six different continents, as well as two urban neighborhoods popular with tourists. The tool provides an overview of the contributor set behind each place, while enabling detailed qualitative inquiry into individual contributors. All data displayed in Crowd Lens is derived from publicly available OSM history files. Both Crowd Lens (http://tinyurl.com/crowdlens) and a demo video (http://tinyurl.com/crowdlensvideo) are available to view online.

Because detailed spatial data on a global scale takes an extraordinary investment to collect, OSM has become attractive to businesses and governments who want to supplement or replace their existing spatial data at minimal cost. The main goal of Crowd Lens is to help geospatial technology professionals make informed decisions about whether to adopt (or continue using) OSM data. Existing use cases of OSM adoption include a transit authority offering OSM base maps for its route planning application (McHugh, 2014), a cloud-based web mapping company offering a variety of custom-styled OSM base maps (Barth, 2015), and a location-based services company offering a mobile route-finding application using OSM street data (Van Exel, 2014).

Despite the cost and flexibility advantages of free and open data, each use case above is vulnerable due to the loose organization of OSM contributors. Incoming OSM data is not pre-checked in real time by any kind of gatekeeper for geometric or semantic accuracy. Although there are automated scans of OSM by bots and third-party quality assurance (QA) teams looking for logical flaws and other anomalies (Bhangar, 2016), one of the most effective bug-prevention mechanisms might be a large and active contributor set that will quickly spot and fix problems in the data (Raymond, 1999; Haklay, Basiouka, Antoniou, & Ather, 2010). The intent of Crowd Lens is to reveal the size, activity, and interests of the OSM contributor set in a given place, while offering a view into anomalies in data contribution including imports, bot activity, and vandalism. The tool gets beyond a simple "node counting" approach to understanding OSM activity by using qualitative metadata such as contributor comments and biographical text. Thus Crowd Lens accommodates the visual analytics mantra of Keim, Mansmann, Schneidewind, Thomas, and Ziegler (2008), facilitating an initial overview of the data, followed up with more detailed filtering and analysis of specific items of interest (i.e., the activities of individual OSM contributors).

The remainder of this paper develops as follows: First, we review cases where visual analytics has been applied to the study of crowdsourced projects like Wikipedia and OSM. We then describe the design and development of Crowd Lens, including two user studies and a scenario-based design exercise that were undertaken to guide development and assess the tool's effectiveness. We also offer insights about OSM contributors that we learned using Crowd Lens, and we conclude with future directions of development for Crowd Lens and similar tools.

Relevant literature

OSM falls into a category of spatial data often called volunteered geographic information (VGI), which allows individuals who may have little traditional training in cartography or geographic information systems (GIS) methods to create spatial data (Goodchild, 2007). When broadly considered, the term VGI can encompass phenomena as diverse as crowdsourced traffic speed databases, geotagged social media posts, and citizen engagement "Report a problem" apps (see for example http://www.fema.gov/mobile-app). Haklay (2014) proposed that OSM has become large and unique enough to merit its own strain of inquiry within VGI research. Areas of investigation within his proposed "OpenStreetMap studies" include data completeness, data trustworthiness, societal impacts of and social practices in OSM contribution, and OSM as an exemplar of big data for computing research.

Questions of data quality, trust, and credibility addressed in the broader VGI literature are informative for judging the utility of OSM. Flanagin and Metzger (2008) note that the credibility of VGI is based largely on the believability of the source, yet in digital media environments source metadata can be sparse, missing, or hard to interpret. Studying government adoption of VGI, Johnson and Sieber (2013) observed that although VGI holds great potential to help governments correct errors and fill in gaps in their own spatial data, officials hesitate to adopt VGI due to the notion that it is produced by error-prone hobbyists and is not a serious source of data. Credibility could potentially be increased by knowing more about contributors. Along these lines, the National Research Council suggests a key research question for VGI: “What are the characteristics of the producers of VGI and how should we evaluate the content and quality of what they have produced?” (NRC, 2010, p. 108). Drawing from literature examining the free and open source software (FOSS) and Wikipedia communities, Coleman, Georgiadou, and Labonte (2009) propose that content contributors could be characterized by their humanity (in other words, whether they are real or a bot), edit frequency, edit types, veracity of the edits, and reputation gained through longevity and activity in the project.

The research domain of visual analytics can aid understanding of the OSM contributor crowd. Visual analytics emerged in the same time frame as the popularization of the term "big data", and was fueled by the need to interpret incoming streams of text, video, images, and other data sources, often in real time (Thomas & Cook, 2005). Visual analytics has been applied toward understanding large corpora of news stories and social media posts (MacEachren et al., 2011; Dou, Wang, Skau, Ribarsky, & Zhou, 2012), cycles of geopolitical events (Robinson, Peuquet, Pezanowski, Hardisty, & Swedberg, 2016), histories of judicial decisions (Collins, Viegas, & Wattenberg, 2009), and more. Andrienko et al. (2007) promoted the incorporation of mapping and location awareness components into visual analytics to help understand data with a spatial component. "Geovisual analytics", therefore, recognizes that much big data is associated with location coordinates or geometries, and that geographers are well suited to addressing big data analysis tasks (Burns & Thatcher, 2014).

Visual analytics, as well as "traditional" information visualization, have been applied toward both understanding crowdsourced data generation processes and making sense of the resulting data. Wikipedia has perhaps received the most attention, as articles may be constructed by dozens of contributors, each with their own motives. In an early example, Viegas, Wattenberg, and Dave (2004) created history flow diagrams showing how individual efforts meld to form an article over time. Other studies of Wikipedia have visualized “edit wars” and interactions between contributors (Suh, Chi, Pendleton, & Kittur, 2007; Brandes & Lerner, 2008; Borra et al., 2015). Many of these tools are designed for analysts with extensive domain knowledge of Wikipedia contribution patterns; however, Boukhelifa, Chevalier, and Fekete (2010) describe a “skin” for Wikipedia that allows a more casual user to assess the amount of attention and controversy an article has generated and to make inferences about content quality.

Beyond Wikipedia, a smaller body of research has employed visual analytics to study OSM data contributions. Some inquiries have focused on developing cartographic renderings (whether static or interactive) for OSM metadata. These include "version contour lines" (Van Exel, 2011a), a "temperature" map of OSM community attention (Van Exel, 2011b), and maps of the time since the most recent edit of any feature (Barron, Neis, & Zipf, 2013).

In the realm of interactive cartography, Roick, Hagenauer, and Zipf (2011) and

Roick, Loos, and Zipf (2012) used hexagonal cells to aggregate OSM metadata, including the date of most recent feature edit, the number of points of interest, and so forth. These maps are bundled in an interactive tool called OSMatrix (http://koenigstuhl.geog.uni-heidelberg.de/osmatrix). The tool reveals stark variations in mapping activity along some international boundaries, and illuminates hotbeds of activity. In related research, Trame and Keßler (2011) use an interactive heat map to display which OSM entities have been edited the most. “Hot” areas of the map could identify controversial regions, or they could indicate a generally high interest in the area from tourism, etc.

Exploratory visualizations of OSM data and metadata can also help identify data completeness, as well as anomalies and errors. For example, interest in humanitarian applications of OSM led to OpenStreetMap Analytics (http://osm-analytics.org) that visualizes how coverage levels of roads and buildings have changed over time throughout the world. The project, whose sponsors include the American Red Cross and the Knight Foundation, indicates where mapping should be encouraged to achieve a more complete OSM for crisis response and prevention. In the area of quality assurance, Geofabrik’s OSM Inspector (http://tools.geofabrik.de/osmi/) presents a navigable map with an overlay of detected problems such as "self-intersecting ways" and "duplicate node in a way". These might lead the analyst toward a decision about whether and how to fix the issue.

Visual analytics tools for OSM have been mostly concerned with geographic data characteristics and rate of data production, whereas many of the Wikipedia tools include analysis of the contributors in an attempt to understand more about article credibility and author motives. Exploratory visualizations of OSM contributor work and habits have been more limited. Some promising avenues for identifying the most active contributors with local knowledge were proposed by Napolitano and Mooney (2012) but not developed into a tool. A series of interactive websites by Pascal Neis (http://resultmaps.neis-one.org/) including *How Did You Contribute to OSM?* and *Your OSM Heatmap* offer exhaustive statistics and some generalized maps summarizing the work of individual contributors. We believe that it is the connection to the individual that makes Neis's maps popular in the OSM user community, and that additional research on visualizing contributor activity is needed to complement emerging OSM query frameworks such as epic-osm (J. Anderson, Soden, K. Anderson, Kogan, & Palen, 2016). The Crowd Lens tool described here advances OSM visual analytics in the human realm, allowing an exploration of the composition and activity patterns of the OSM contributor crowd and its impact on fitness for use. While a few of the summary statistics reported by Crowd Lens are similar to those in Neis's tools (such as the contributor's total number of OSM edits), our focus is on places and the contributors' relationships with those places. These relationships are expressed through comments, tags, the proportion of the contributor's effort dedicated towards that place, user profile pages, and various other metadata items not exposed as completely by other tools.

Design and development of Crowd Lens

Crowd Lens is an interactive tool that runs in a web browser. In a single display containing multiple linked views, it provides a filterable overview of the OSM contributor crowd behind any given place, as well as a range of drill-down options to explore detailed aspects of the data and the actions of any one selected contributor (Figure 1).

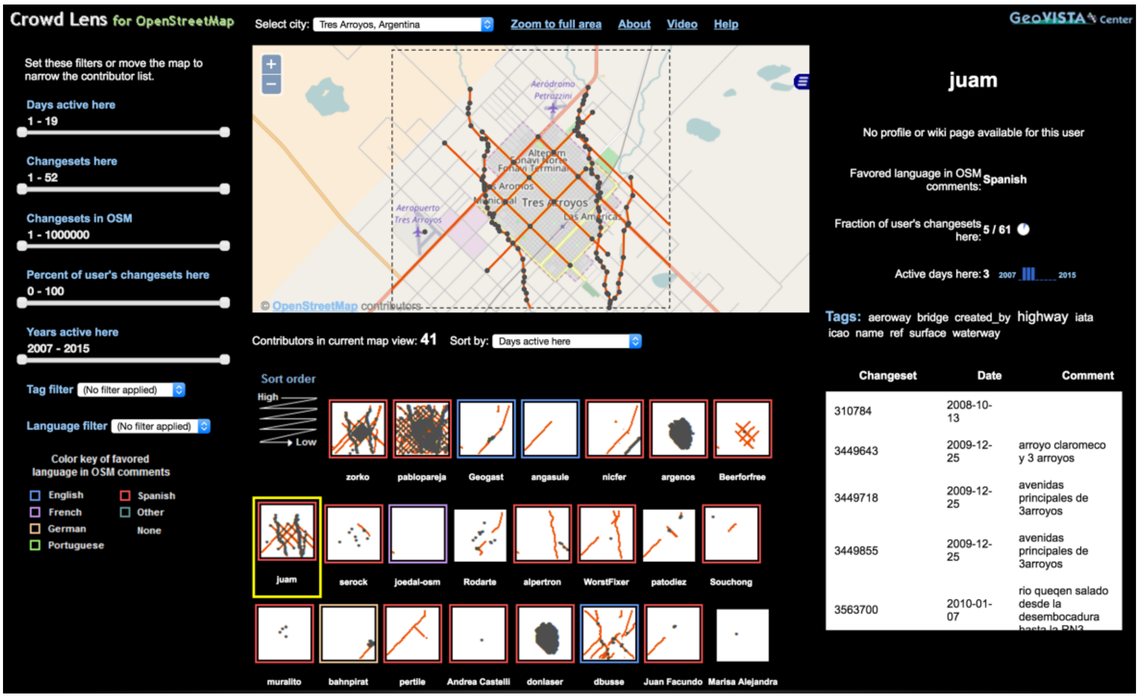

Figure 1. Crowd Lens user interface.

Thomas and Cook’s (2005) visual analytics research agenda identifies three main tasks associated with visual analytics tools: Assess, Forecast, and Develop Options. Crowd Lens is designed to facilitate each. As an assessment tool, it helps the analyst conceptualize what has happened with OSM in any particular place. These findings could facilitate educated guesses about which OSM contribution trends may continue (or not) in the near future, thus supporting at least rudimentary forecasting. Crowd Lens could also support decisions about whether (and how) OSM could be integrated into an organization's existing suite of geospatial services and applications. Furthermore, findings from Crowd Lens might help develop strategies for dealing with conflict between contributors, bias in the data, or sparsely represented geographies.

Crowd Lens is a data foraging and sensemaking tool that, in addition to finding answers to questions about OSM construction, also helps formulate new questions, a goal articulated for visual analytics more generally by Bivand (2010). For example, a

Crowd Lens user might notice OSM contributors who appear in different cities on opposite corners of the globe, leading to further investigation about the origins and influence of "power users" on the worldwide map. A foray into these user profiles might reveal mappers who are paid to improve the data on behalf of a company that relies on OSM (see example in Barth, 2015), leading to additional inquiries into the political economies of free and open data and how they affect projects like OSM.

Crowd Lens evolved from a set of static maps and graphics derived by Quinn (2015) in a study of OSM contributor patterns in small cities. That research used small multiple maps to compare the amount of activity between OSM contributors in a given place. Abnormal contribution patterns such as automated imports and very active contributors jump out of these graphics and beg a deeper analysis through an interactive visual interface.

Crowd Lens is built with OpenLayers, an open source library for JavaScript development of geospatial web applications. Data for the visualization comes from two archive files: (1) the OSM "full history dump" (48 GB compressed), and (2) the OSM changeset history (1.3 GB compressed). Our demonstration uses the December 28, 2015 release of the data. Rectangular envelopes of data for each study area were extracted and processed into a series of GeoJSON data files for web display and query. The bulk of this work was automated through the Python scripting language. We used a mix of ArcGIS Python modules and Geospatial Data Abstraction Library (GDAL) command line tools to construct vector geometries and small maps of each contributor's work.

The OSM history dump contains the geometries and tags of nearly all current and past geographic features contributed to the OSM database. (Some very early items are missing, while others were redacted when some contributors refused to accept a license change in 2012.) The OSM changeset history is a complementary record of how these features connect to different contributors' efforts in the project. A changeset represents one session of a contributor's work. When saving a changeset, a contributor can type a comment describing and justifying the edits made. These comments provide qualitative information about contributor motives and habits. They also give clues about the preferred languages of each contributor, which we derived using the Python-based langid.py language identification module (Lui & Baldwin 2012; Quinn, 2016).

The top menu bar of Crowd Lens allows the user to select one of the predefined city or neighborhood areas for analysis. To support the user evaluations reported here, six small cities were selected from different continents. They are Hereford, UK; Hervey Bay, Australia; Johnstown (Pennsylvania), USA; Kadiri, India; Suhum, Ghana; and Tres Arroyos, Argentina. These are non-suburban cities of regional importance judged to be representative of OSM coverage levels in similar sized towns nearby. The rectangular study areas surrounding each city were positioned to contain about 50–100 thousand inhabitants. This choice of cities builds on Quinn's (2015) inquiry into small cities as a more telling barometer of OSM status in a region than might be observed in large metropolises.

Although some concepts from Crowd Lens could inform a broad geographical analysis, our interest here is more place-focused. The application of the tool toward the scales of block, neighborhood, and small city allows the user to conceptualize all the OSM activity occurring in a place, enabling the analysis of microgeographies where contributor attention to the map may be sparse or disproportionately influenced by a small number of active contributors.

The Crowd Lens interface displays the contributor activity in the city and is divided into four user interface components detailed below.

Contributor list

The contributor list (Figure 2) acts as the focal point for analysis. It conveys the size of the contributor crowd active in any one place using small multiple map images (one per contributor) that provide an overview of each contributor's relative amount of influence on OSM map contents. It also allows the selection of any one contributor to learn further details.

Each contributor's small map depicts *all* the work performed by that contributor in the OSM history file; the maps do not represent tiled fragments, deltas, or individual changesets. The contributor maps are sortable by activity criteria such as a contributor's number of active days in this city, number of changesets in this city, number of changesets in the OSM project, and percent of OSM changesets that occur in this place.

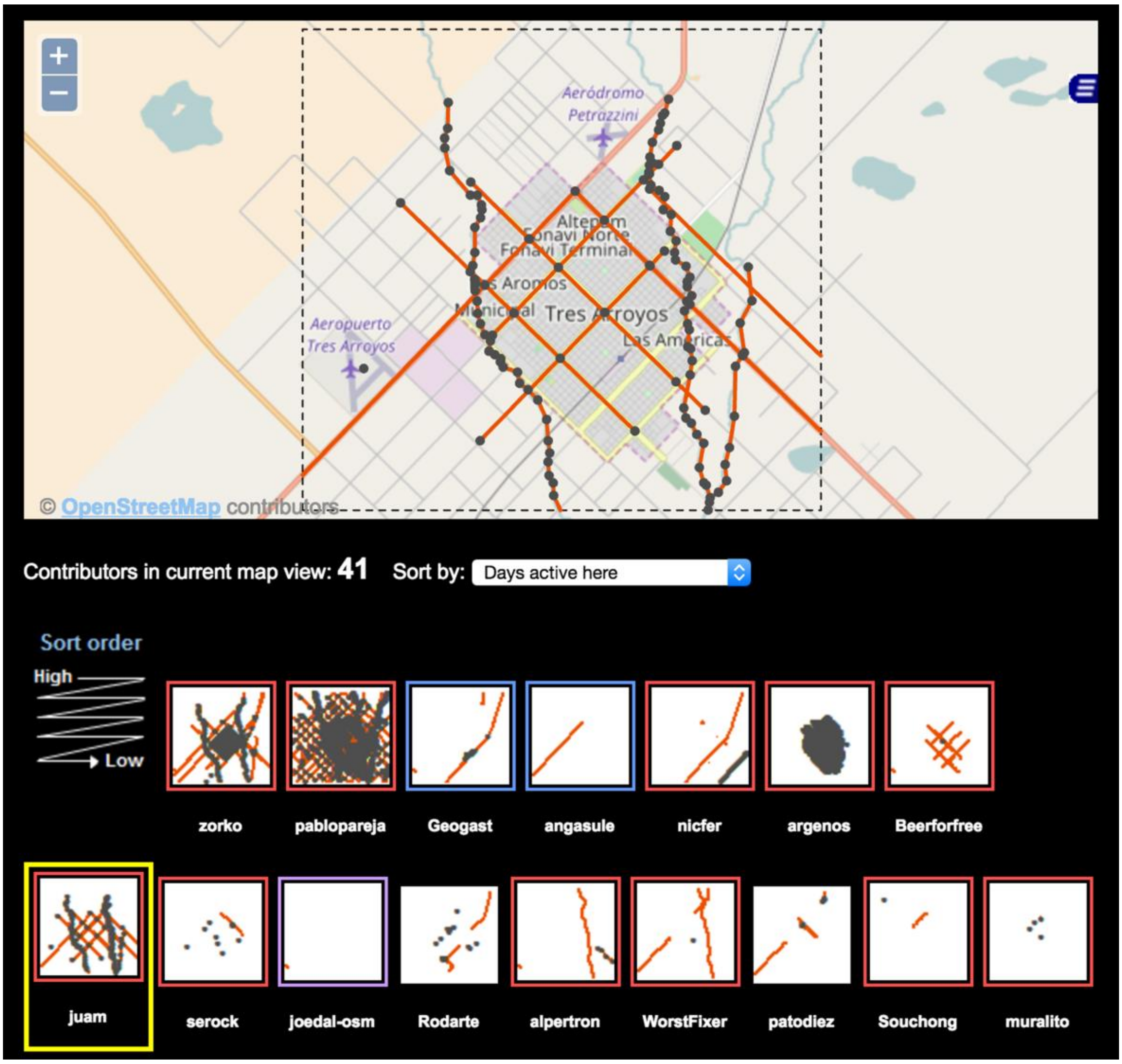

Figure 2. Contributor list of small map images, topped by the navigable main map.

The border color of the small map images signifies the contributor's preferred language in OSM changeset comments. Cities with a higher percentage of contributors speaking the native language are expected to contain more map features that would be of value to the daily routines of the city's residents (Quinn, 2016). In some cases the preferred language detected by the langid.py module does not match that of the contributor's changeset comments displayed in Crowd Lens. Although miscodings are possible, we have found that these mismatches are most common when a multilingual contributor tailors his or her comment language to the place being mapped. Thus, comments in a language other than the personal preference might signal a non-local contributor (e.g., a German speaker using English when contributing outside Germany).

Crowd filters

To better identify subsets of the crowd for further analysis, Crowd Lens offers multiple filtering tools. Dropdown filters allow users to narrow the contributor list by preferred language and by tag applied (e.g., highway). Additionally, the contributor list can be narrowed down by adjusting the double-ended dynamic query filters (Shneiderman, 1994) (Figure 3).

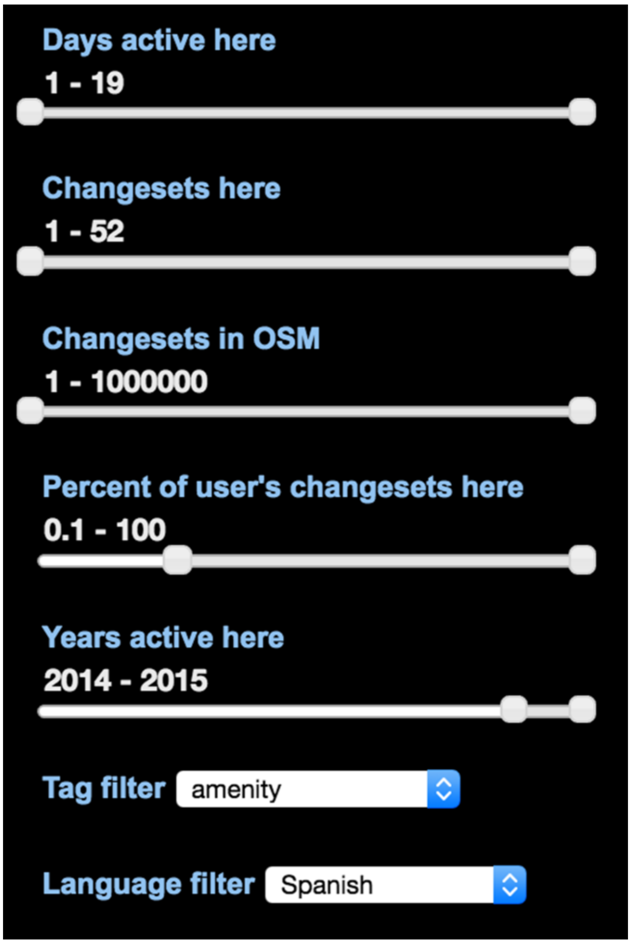

Figure 3. Crowd filters on the left hand side of the interface allow the narrowing down of the contributor list.

Map

The large map in Crowd Lens details a selected contributor's work, overlaid on the current OSM map tiles for geographic context. Clicking one of the geometries highlights all associated features in the changeset. The map also acts as a filter for the contributor list; therefore, panning and zooming the map offers a way to see the crowd size in any particular neighborhood or other sub-geography of the city.

Individual contributor panel

Users can analyze any contributor in more detail by selecting a small map image in the contributor list. Figure 4 shows how the individual contributor panel displays the selected contributor’s preferred language, number of changesets (in this city and in all of OSM), and years active in this city. A “tag cloud” shows the relative amounts of attention the contributor devoted toward mapping different types of entities.

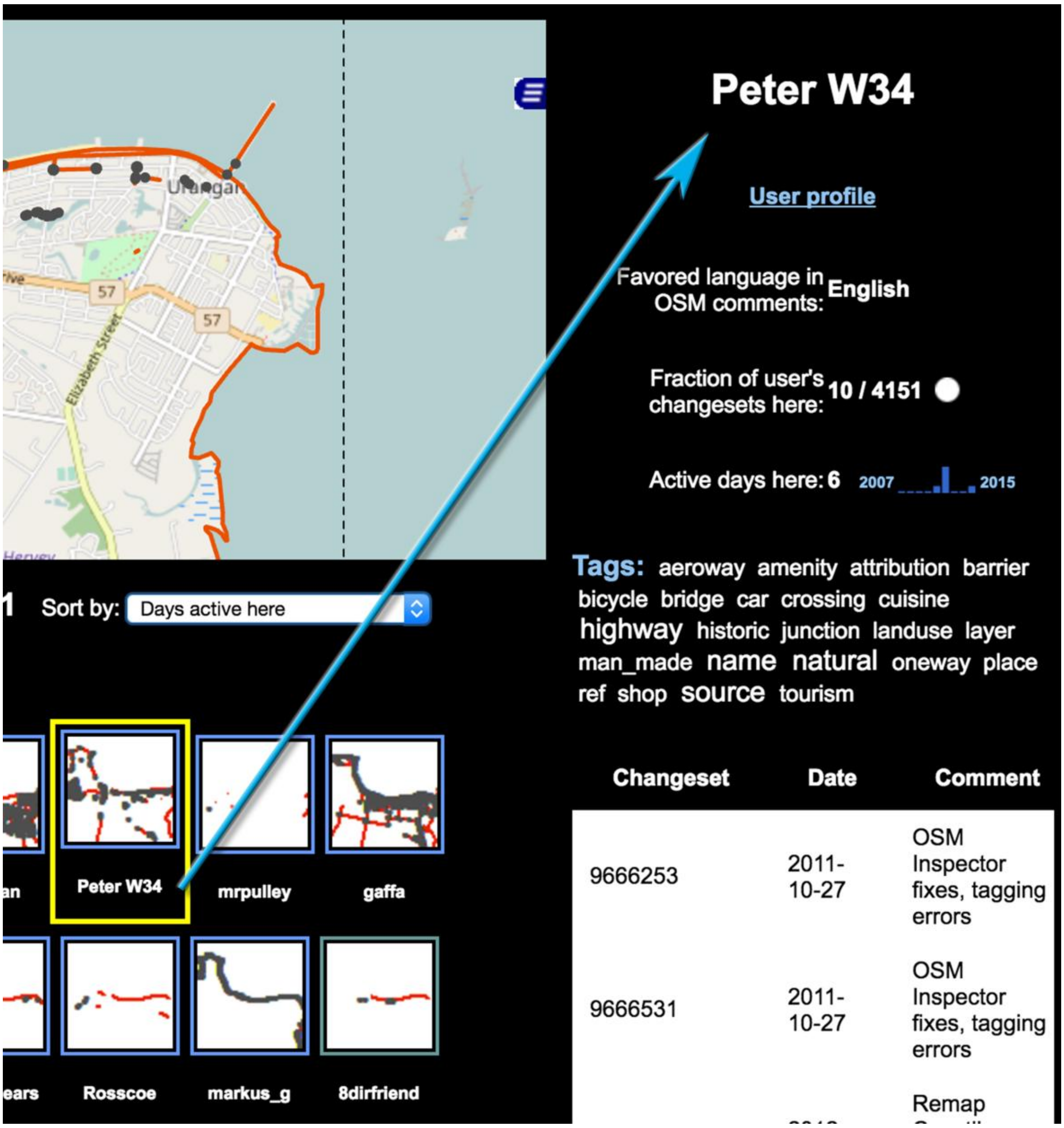

Figure 4. Individual contributor details displayed when a map image from the contributor list is selected. The arrow is shown for emphasis only.

Additionally, this panel offers a data table showing all changesets produced by the selected contributor. The table reports each changeset's date, identification number, and comment text, sometimes revealing information about a contributor's relationship with the mapped entities. This table is interactive, such that clicking a changeset record highlights its associated map edits, and clicking an item on the map highlights the corresponding changeset record.

Sometimes further information about a contributor (e.g., profession, motives, and hobbies related to OSM) is available in publicly-available OSM profile or wiki pages that he or she created. When these pages exist, hyperlinks to them are provided in the individual contributor panel.

Development process and evaluation of Crowd Lens

The development of Crowd Lens followed a user-centered design process, wherein an iterative series of user studies were applied to assess and refine the tool's usability and utility (Robinson, Chen, Lengerich, Meyer, & MacEachren, 2005; Goodwin et al., 2013). Roth, Ross, and MacEachren (2015) describe how a series of smaller studies each using 5–10 participants during the formative stages of the tool can lead to more positive changes than a single large summative study executed after the tool's deployment.

The Crowd Lens evaluations included two components with human participants: (1) a usability test early in the development cycle to assess the intuitiveness of the user interface, and (2) a later evaluation by geospatial technology professionals to assess the utility of the Crowd Lens and understand how the tool affected their confidence in OSM data. Between these two rounds, we conducted a scenario and claims analysis exercise (Rosson & Carroll, 2002) to help determine which remaining development tasks would be most relevant to end-user workflows.

Early usability testing

An early development version of Crowd Lens was tested for usability by seven social sciences students at a major US university. Participants were required to have some previous exposure to GIS and OSM because the ultimate target audience for Crowd Lens is professionals already familiar with OSM concepts. The testers were asked to use Crowd Lens to answer a series of objective questions pertaining to facts about the OSM data (see supplementary materials). Finding the correct answers required exercising the sorting and selection functions in Crowd Lens across the user interface components. Examples of questions include, “Which contributor to ‘City X’ has over 10,000 changesets in OpenStreetMap?” and, “Which of the cities had the most users active over 10 days in the project?” As testers attempted to answer these questions with the tool, their interactions were captured by screen recording software.

We terminated each tester’s Crowd Lens session after a maximum of 15 minutes. This was followed by a semi-structured interview wherein we asked testers to identify the most difficult and easy tasks and comment on the user interface components that facilitated or impeded their work. We also invited testers to suggest new features.

Results of early usability testing

Testers were generally able to figure out the answers to the questions without any assistance or previous exposure to the user interface. Five of the seven testers got all seven questions correct within the allotted time period. One tester missed a single question and one tester missed two questions. Testers commented that they liked the ability to sort the small map images, and that the colored border of the map images was helpful for quickly getting a view of the contributors’ preferred languages.

The most common items of feedback were that not enough small map images were visible on the screen at a time, making it difficult to navigate through the images or get an overview of which users were predominant within a city. In response to this, the percentage of screen space devoted to the user list was increased in future iterations of the tool, while the size of the map images was decreased.

Other users reported lags in responsiveness of the tool in some of the cities with high OSM activity, impeding interpretation of the display. This was eventually mitigated by developing a more scalable data structure to store each contributor's work. An animated "loading" graphic was also added to give users a visual cue about when they should begin to interpret the map results.

Other tester suggestions eventually applied to Crowd Lens included adding an explanatory graphic to the user list explaining the map sort order (see Figure 2), and making the (originally static) tag cloud into a clickable filtering mechanism to show all users who applied a particular tag.

Scenarios of use and claims analysis

The usability testers submitted more suggestions than could be reasonably implemented in the time available. To prioritize features for development, we undertook a thought exercise called a "scenario of use". Rosson and Carroll (2002) describe scenarios of use as a way to form a design by telling a story about how an end user would interact with a tool's user interface. Critical user interface features and their interaction points in the scenario can then be evaluated through a "claims analysis" identifying the pros and cons of the design. See MacEachren, Crawford, Akella, and Lengerich (2008) for an example applied to an exploratory geovisualization web application. Although scenarios of use are often implemented in the early stages of design, they became important at the mid-point of Crowd Lens development to keep the project focused on its core objectives.

We developed two scenarios of use, both available in the supplementary materials. They describe hypothetical interactions with Crowd Lens by geospatial technology professionals with specific use goals. The first involves a GIS analyst trying to decide whether to incorporate OSM as a base map in municipal government applications. The second describes a GIS manager in a non-profit organization that conducts operations in rural areas. Both use Crowd Lens to evaluate the amount of recent local influence going into OSM in their areas of interest. Although the scenarios are fictional, the contexts of OSM use described therein were influenced by a range of formal and informal input collected by the first author at industry conferences such as State of the Map and FOSS4G.

Scenario claims analysis

Following the development of the above scenarios, a claims analysis was performed to evaluate advantages and drawbacks of critical aspects of the user interface. These pros and cons are enumerated using + and – symbols, respectively, in Table 1.

The ability to narrow down the contributor list through a series of filters...
+ Allows the quick calculation and further exploration of the subset of contributors who participated in a particular way
+ Gives the end user a way to explore which parameters affect the crowd size the most
- May get the end user into a state where they forget they are only viewing a subset of contributors
- May tempt the user into getting overly interested in tweaking settings and controls and/or exploring the work by specific individual contributors, rather than thinking about the broader context of and patterns of contributions
Showing the crowd filters and the individual user details on the same screen...
+ Lets the end user immediately know that he or she can explore the general characteristics of the crowd as well as the individual nuances of each contributor
+ Cuts down on the clicks and "window management" that the end user needs to perform to open and shut interface components
- Makes the view of the screen more visually cluttered
- May be annoying for the analyst who solely wants to focus on crowd characteristics, or who is only interested in individual characteristics
Showing just a single user's contributions as a vector overlay on the map (rather than showing all the data)...
+ Cuts down on visual clutter in the map screen
+ Allows a direct link between clicks on map features and the contributor changeset table
+ Keeps map loading speed and responsiveness reasonable
- May mislead people into thinking that a single person's contribution actually represents all the available data
Representing each contributor with a small map image...
+ Gives the analyst a quick view of each person's relative contribution and how it compares with others' edits
+ Gives a visual clue that the analyst might be able to change the main map by clicking the thumbnail
- Takes up additional space and makes it difficult to see all the contributors listed in the view at the same time
- May be initially confusing to those who do not understand that the image is showing a (very abstracted) map

Table 1. Claims analysis derived from the scenarios of use.

At the time these scenarios were created, the dynamic crowd filter controls had not been developed. From the scenarios of use and claims analysis, it was decided that the filters would introduce many new possibilities for understanding subsets of the crowd, and that it would be wise to invest most of the remaining development time in implementing the filters.

Testing by geospatial technology professionals

A final round of testing evaluated how effectively Crowd Lens addressed tasks faced by technical OSM users. We also assessed how insights from Crowd Lens might affect professionals' confidence levels in OSM data. This round of testing occurred after we deployed a beta version of the tool incorporating improvements prompted by the earlier evaluation and scenario exercises.

We invited 10 geospatial technology professionals working with digital maps or GIS on a regular basis to try Crowd Lens. To be considered for participation, testers were required to have used or evaluated OSM in some way to support their work. We identified participants through a snowball process that began by contacting acquaintances from our professional networks. Testers represented a broad range of domains including cartographic design, software development, environmental consulting, utilities management, federal and municipal government, and others.

Testers completed the evaluation entirely online. (The full evaluation procedures are available as supplementary material.) In a pre-assessment, testers rated their level of confidence in OSM for their work purposes. Subsequently, they were instructed to use the Crowd Lens tool for as long as they liked to form an understanding of how OSM had taken shape in different towns, and consider how this related to the use of OSM in their own projects. Testers were then asked to share any insights gained through the tool and suggest improvements. Finally they evaluated how the tool had affected their confidence in the quality of OSM data for their work purposes.

Questions in the survey were largely open-ended because we knew that the testers used OSM differently in their professions. The purpose of this stage in Crowd Lens development was to assess the utility of the tool and we did not want to create closed-ended questions reflecting any preconceived notions from us about OSM use. We did not ask testers to complete any specific task during the evaluation because we wanted them to tailor the use of Crowd Lens to their own interests, and we felt the intuitiveness of the interface had already been adequately vetted through the earlier usability test.

We compiled the results of the survey by carefully reading the responses to each question and tallying recurring themes. We also noted tester insights that were particularly detailed, innovative, or outside the realm of what we had expected. The depth of the responses indicated to us that the testers were generally able to navigate the new crowd filters and the associated contributor statistics in the ways postulated in the claims analysis exercise.

Results of testing by geospatial technology professionals

All testers reported having used OSM at some point to support their work, although the amount of experience varied. For instance, some occasionally used OSM as a basemap, others used it to supplement datasets from governments and commercial sources, and several worked directly with OSM source data they had extracted for use in GIS analysis projects. No testers indicated using OSM as their primary data source at work, although this was not asked explicitly.

On the five-point scale of confidence in OSM data quality for work purposes (with 5 being the highest), the mean result was 3.8, with seven out of the ten testers reporting a score of 4 and only one tester with a score on the lower end of the scale (Table 2). Most testers reported that they had found OSM suitable for their work purposes and three mentioned that OSM was typically better than other data sources they encountered; however, many of these same people maintained concerns about errors persisting in the OSM data. These misgivings were characterized by a tester who remarked, "Open Street Map [sic] data looks fairly clean for my area, but I have found areas with older information which makes me question the data as a whole." Other issues mentioned included OSM's lack of certain types of data (such as elevation), difficulty convincing colleagues of OSM data quality, and legal pitfalls of using crowdsourced data.

One objective of testing was to understand what (if any) insights about OSM testers derived from their interaction with Crowd Lens. The most common reported insight was an appreciation of the magnitude of edits supplied by the most active OSM contributors. Several testers looking closely at the habits of these contributors noticed there are those who tend to stay local in their edits and those who edit a wider variety of places across the globe. One tester also observed a middle group of contributors that tended to focus on one type of task, such as adding addresses or working on water features.

	MAJOR CONCERNS				HIGHLY CONFIDENT
Score	**1**	**2**	**3**	**4**	**5**
Number of respondents	0	1	1	7	1

Table 2. Responses to the pre-assessment question "How confident are you in the quality of OpenStreetMap data for your work purposes?"

Four of the 10 testers mentioned the potential for Crowd Lens to be used as a tool for assessing OSM data suitability or quality. One tester noted the pitfalls of relying on a raw count of editors, musing that "the number of editors may be a good indicator of the accuracy of the data, but it also may not..." Another commented, "one has to keep in mind that there may be 'hit and runs' where a one-time or inexperienced user may make a change." This remark demonstrates that a large set of contributors displayed in Crowd Lens may not entirely dispel a fear of errors permeating the dataset.

Two testers suggested that Crowd Lens could be used for identifying experts in particular topics such as hiking, parks, or pedestrian infrastructure, who could then be recruited to help with targeted mapping projects or other initiatives. We did not anticipate this use of Crowd Lens; it was enabled by the tag cloud. Future development work in this vein could focus on grouping tags to create theme-based filters, or highlighting the map features and changeset table records associated with a selected tag.

Despite all the testers sharing at least one new insight from Crowd Lens, several reported that the tool would not affect the way they worked with OSM because they already tended to use OSM data as a "last resort" when other sources were not available. Thus, quality seems to be less of a concern when the alternative is to have no data at all. Overall, however, testers' confidence levels in the quality of OSM for their work purposes went up after using Crowd Lens, as indicated by a mean response of 3.7 to a question asking whether they were more or less confident after using the tool (Table 3). We were surprised to see that no one submitted a score lower than 3 (which would indicate a decrease in confidence) because we thought some testers might be alarmed by the low number of contributors in some cities and the fact that most contributors performed relatively small amounts of work on the map. Asked to explain their answers, testers focused on the places with the most contributors and the amount of attention rendered by the most active contributors. One tester remarked, "seeing the large number of different contributors makes me more confident that errors will be fixed", while another was "very impressed at the scale of edits and dedication by volunteers." Testers whose level of confidence remained unmoved (ie., submitted a score of 3) cited lingering concerns with OSM data accuracy.

Testers offered numerous suggestions for improving Crowd Lens to provide additional insights. A popular request was to select any geographic bounding box for analysis. Achieving this capability will require an engineering effort beyond the scope of the current research. The challenge is not only large data, but also messy data that currently requires much data wrangling (Kandel et al., 2011) to clean and preprocess. Meeting this challenge will require specialized geospatial and multi-lingual text data wrangling tools. Once data are usable, search and indexing optimization techniques will also be needed that can scan the entire OSM history files and construct necessary data structures on the fly.

Even given the ability to select any bounding box for analysis, a limit on the size of requested study area would be needed, not only to ensure an acceptable level of performance, but also to maintain the utility of the tool towards understanding the OSM crowd in a place. For instance, even if an entire state or country could be selected, an interactive exploration and side-by-side cartographic comparison of each contributor's work might prove infeasible due to the sheer number of persons involved in the project.

	LESS CONFIDENT				MORE CONFIDENT
Score	**1**	**2**	**3**	**4**	**5**
Number of respondents	0	0	4	5	1

Table 3. Responses to the post-assessment question "To what degree has this tool affected your perception of the quality of OpenStreetMap data for your work purposes?"

Two testers mentioned a desire for faster performance of the tool. Following this feedback, we performed a full code review that resulted in efficiency gains; however, alternative technical approaches could potentially increase responsiveness. For example, the most computationally intensive portion of Crowd Lens is the back-end spatial processing that determines on the fly which contributors influenced the current map bounding box. Spatial indexing methods might make this step more efficient. Another barrier to performance occurs with drawing the contributor's mapped features as vector graphics in the web browser. This is typically a boon to web visualization performance and interactivity, but with the contributors who have added thousands of features, the number of graphics to be loaded and drawn can slow down the browser. This might be alleviated by producing the map layer from a web map service (WMS) or a pregenerated image cache. Such an approach would sacrifice the interactive re-styling that is possible with vector graphics, and would introduce a new geospatial server tier into the application architecture. It was avoided for this version but could be implemented for larger datasets. Approaches using WebGL and/or vector tiles (http://wiki.openstreetmap.org/wiki/Vector_tiles) also hold promise for rendering many graphics in the browser, and would retain the interactive benefits of client-drawn vectors.

Finally, a tester commented that "the tool had too many adjustable elements that made the app confusing." This presumably refers to the many filters available through slider widgets and dropdown menus. Although the claims analysis exercise had anticipated that some users might not like the visual clutter of the filters, we had perhaps underestimated how overwhelming the multitudinous available settings would seem to some users. A potential way to address this confusion was suggested by another tester: define some standard filter settings that would guide the end user toward learning certain things. For example, there may be thresholds of certain measures of participation (such as number of days active, percentage of OSM changesets in the target city, etc.) that would characterize the most active contributors likely to have local knowledge of the place being mapped. Crowd Lens could offer some pre-generated analysis patterns (Fowler, 1996; Filho & Iochpe, 1999) with default choices to help end users find these interesting subsets of the crowd. These patterns could be constructed through the domain knowledge of OSM experts, or they might be generated through data mining approaches trying to automatically detect and suggest the most "interesting" trends in the metadata (Shneiderman, 2002; Beale, 2007).

Overall, the view of each OSM contributor's work offered by Crowd Lens had the effect of increasing some testers' confidence levels in the quality of OSM data. Others felt the tool was interesting and could possibly help them assess data quality in individual situations, but it did not alleviate their worries about errors infiltrating the crowdsourced data.

Case study analysis

We complement the user studies here with a case study analysis of the six original sample cities that we loaded into Crowd Lens, along with two urban neighborhoods popular with tourists that we added after the user tests (Seattle Center, Seattle, USA; and Copacabana, Rio de Janeiro, Brazil). A glance at any of these areas in Crowd Lens confirms that most of the work in OSM is done by a relatively small percentage of individuals, a statistic noted throughout the project's history (Neis & Zipf, 2012; Wood, 2014). Each city has a group of individuals who made many contributions, and a large group who only created one changeset; yet many users in the latter group have thousands of other changesets in OSM and came to these cities to make a single fix or addition. In other cases they mapped in these places during the process of applying automated fixes and imports, as the text “bot” appears in several of the user names with large numbers of nodes and ways (e.g., street polylines) modified. Visiting the profile and wiki pages of these bots reveal that they are often designed to add, remove, or adjust attribute tags to conform to a certain semantic standard.

Many of the study areas have contributors who have made one or just a few edits in the entire OSM project. Crowd Lens may be especially helpful in revealing the activities of contributors who come to OSM for one specific mission, or decide to leave the project after a single mapping session. Investigations of one-time contributors are sparse in OSM literature, although Anthony, Smith, and Williamson (2005) showed that in Wikipedia such isolated “Good Samaritan” edits can have more staying power than those produced by more frequent contributors. In Crowd Lens, the city of Tres Arroyos shows six out of 41 contributors (14.6%) who produced 100% of their changesets in the city. All made fewer than 10 changesets. Hervey Bay, Australia saw six contributors out of 85 (7.1%) who made 100% of their changesets in the city. One added an automobile electric shop, while another added a gem/mineral club and a museum. These types of establishments might be valuable to the livelihoods of contributors from the perspectives of business or leisure, and could provide similar benefits to local residents viewing the map. The same applies to the over 30 works of public art added to OSM by two relatively novice contributors in the Seattle Center area.

In other situations, a contributor appeared to create an OSM account solely to ensure that a local business was placed on the map. Sometimes this was executed with much enthusiasm, as in the case of the contributor named “Spacious Copacabana Beachfront Penthouse” whose only addition was a single rental property, complete with website, phone, and wheelchair accessibility tags. This blatant self-promotion, which would be taboo in Wikipedia, may be tacitly accepted in OSM if the feature does indeed exist on the landscape. As yet the OSM community has not articulated elaborate rules about these practices.

The Crowd Lens language filters reveal that English is widely used for OSM changeset commenting in all locations studied. Only Tres Arroyos had a non-English majority of changeset comments, with most contributors there using Spanish. German-speaking contributors appear in all study locations, confirming the popularity of the OSM project among Germans identified by Neis and Zipf (2012). The cities of Suhum, Ghana and Kadiri, India saw no contributors favoring regional languages (e.g.,Twi, Ewe, Telugu, or Hindi), a finding that could reflect both a desire from the more local mappers to reach a broader audience via English, as well as a heavy influence from mappers living elsewhere, particularly the Global North. The Copacabana neighborhood in Rio de Janeiro has seen extensive editing from both Portuguese and English-speaking contributors. Future research might study if English use declines in nearby neighborhoods less popular with tourists.

When evaluating the six small cities of roughly uniform size, there are stark differences in participation between cities in the Global North (here including Australia), and those in the Global South (Table 4). Profile information and contributor comments are sparse in Suhum and Kadiri, and it appears that bot activity has only affected the top three cities in Table 4. Since bots and imports are controversial and have stirred debates about their effects on OSM communities (McConchie, 2015), it will be interesting to see if OSM develops in different ways in the Global South, perhaps relying on smaller contributions that contain more local knowledge. Tres Arroyos may be an example of this phenomenon, having seen systematic OSM editing activity by local-language contributors, who have created a detailed street map without being driven by bots or mass imports.

City	Number of OSM contributors identified by Crowd Lens from 2007–2015
Hereford, UK	189
Johnstown, USA	107
Hervey Bay, Australia	85
Tres Arroyos, Argentina	41
Kadiri, India	19
Suhum, Ghana	14

Table 4. Number of unique contributors in OSM history files for years 2007–2015 by study area.

Another potential scenario is that undermapped areas are filled in by professional mappers working for companies whose reputation depends on OSM data quality. For example, our analysis with Crowd Lens revealed that 15 out of 34 (44%) contributors who modified Johnstown, USA in 2015 identified themselves in their profile pages as mapping on behalf of Mapbox, a USA-based company selling web map services based largely on OSM data. These were easily detected because Mapbox data team employees openly state their affiliation in their profile pages. It is unknown how many mappers were working on behalf of other institutions and did not state an affiliation (although these cases might be detected through OSM analytics algorithms in the future). The growing presence of corporate interest in OSM is evident from the speaker, exhibitor, and sponsor lists at the annual OSM State of the Map conferences. The associated proliferation of paid mappers is likely a boon for data coverage and quality, but could cause shifting community dynamics that eventually dilute the influence of hobbyist mappers.

Conclusion

It can be easy to think of the "crowd" behind a crowdsourced information product as something amorphous without known or knowable bounds. This leads to a risk that the crowd is overly feared or revered, with end users of the crowdsourced product not knowing how much faith they can put in the resulting information quality and “fitness for use” (Chrisman, 1990). With the introduction of the Crowd Lens geovisual analytics tool, we have opened new views of the OSM contributor set that we believe can help users better comprehend the people who underlie the data. Qualitative data such as contributor comments are visually linked to the corresponding bits of modified geography, allowing the user to understand connections between a database edit and the contributor's view of the edit's purpose. Tools such as Crowd Lens might complement or guide deeper sociological studies on OSM contributors.

Crowd Lens was shaped by a multi-step user-centered design process. Early usability testers offered many suggestions about how the tool could be made easier to operate and interpret. We developed scenarios of use and an accompanying claims analysis to understand which of these suggestions to prioritize in order to facilitate the most salient workflows. Upon showing a refined version of the tool to geospatial technology professionals, we found that it generally raised their confidence in OSM data (Fig. 6), but did not eliminate concerns about errors creeping into OSM. Through an in-depth case study, we demonstrate the potential of Crowd Lens to identify bots, one-time contributors, corporate mappers, and different language communities, thus building a comprehensive understanding of OSM activity in a place.

In addition to the evaluations described in this paper, we suggest several more approaches for testing and refining this tool: (1) On the usability side, a task analysis could be performed wherein the items in the claims analysis would be directly tested. For example, do the crowd filters really cause end users to forget that they are viewing just a subset of the contributors? (2) From the utility side, the tool could be shared with researchers and professionals who work with other crowdsourced data products such as Wikipedia on a regular basis. They could be asked how their insights from Crowd Lens might apply to other crowdsourced data, and how a similar visual analytics tool might be designed and tailored for their own crowdsourced datasets of interest. Elements of our design and evaluation process could be applied toward analyses of other large information repositories, such as Wikimapia, the millions of geotagged articles on Wikipedia (see http://www.geonames.org/wikipedia/), or even "citizen science" projects such as those on Zooniverse.org. As OSM and other crowdsourced datasets increase in volume and popularity compared to more expensive alternatives, we anticipate that interactive visual analysis of contributor crowds will only grow in demand.

Acknowledgements

We are grateful to Greg Milbourne for assistance with data processing and early application prototyping. We also thank the editor and anonymous reviewers for their suggestions on the text and graphics.

References

Anderson, J., Soden, R., Anderson, K. M., Kogan, M., & Palen, L. (2016). EPIC-OSM: A Software Framework for OpenStreetMap Data Analytics. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 5468–5477). https://doi.org/10.1109/HICSS.2016.675

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M.-J., MacEachren, A., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839–857.

Anthony, D., Smith, S. W., & Williamson, T. (2005). Explaining quality in internet collective goods: Zealots and good samaritans in the case of wikipedia. *Hanover: Dartmouth College*. Retrieved from http://web.mit.edu/iandeseminar/Papers/Fall2005/anthony.pdf

Barron, C., Neis, P., & Zipf, A. (2013). iOSMAnalyzer – ein umfassendes Werkzeug für intrinsische OSM Qualitätsuntersuchungen. In *AGIT 2013*. Salzburg, Austria. Retrieved from http://koenigstuhl.geog.uni-heidelberg.de/publications/2013/Barron/Barron_et_al_iOSMAnalyzer@agit_2013.pdf

Barth, A. (2015). The paid mappers are coming. In *State of the Map US 2015*. Retrieved from http://stateofthemap.us/the-paid-mappers-are-coming/

Beale, R. (2007). Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. *International Journal of Human-Computer Studies*, *65*(5), 421–433. https://doi.org/10.1016/j.ijhcs.2006.11.012

Bhangar, S. (2016). Validating the Map. In *State of the Map US 2016*. Seattle, Washington, USA. Retrieved from http://stateofthemap.us/2016/validating-the-map/

Bivand, R. S. (2010). Exploratory Spatial Data Analysis. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis* (pp. 219–254). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-03647-7_13

Borra, E., Weltevrede, E., Ciuccarelli, P., Kaltenbrunner, A., Laniado, D., Magni, G., … Venturini, T. (2015). Societal controversies in Wikipedia articles. In *ACM CHI Conference on Human Factors in Computing Systems*. Seoul, South Korea.

Boukhelifa, N., Chevalier, F., & Fekete, J. (2010). Real-time aggregation of wikipedia data for visual analytics. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on* (pp. 147–154). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5652896

Brandes, U., & Lerner, J. (2008). Visual Analysis of Controversy in User-Generated Encyclopedias∗. *Information Visualization*, *7*(1), 34–48.

Burns, R., & Thatcher, J. (2014). Guest Editorial: What’s so big about Big Data? Finding the spaces and perils of Big Data. *GeoJournal*. https://doi.org/10.1007/s10708-014-9600-8

Chrisman, N. (1990). The error component in spatial data. In D. Maguire, M. Goodchild, & D. Rhind (Eds.), *Geographic Information Systems: Principles and Applications* (pp. 165–174). London: Wiley.

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*, *4*(1), 332–358.

Collins, C., Viegas, F. B., & Wattenberg, M. (2009). Parallel tag clouds to explore and analyze faceted text corpora. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* (pp. 91–98). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5333443

Dou, W., Wang, X., Skau, D., Ribarsky, W., & Zhou, M. X. (2012). Leadline: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (pp. 93–102). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6400485

Filho, J. L., & Iochpe, C. (1999). Specifying Analysis Patterns for Geographic Databases on the Basis of a Conceptual Framework. In *Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems* (pp. 7–13). New York, NY, USA: ACM. https://doi.org/10.1145/320134.320139

Fisher, D. (2007). Hotmap: Looking at geographic attention. *Visualization and Computer Graphics, IEEE Transactions on*, *13*(6), 1184–1191.

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, *72*(3–4), 137–148.

Fowler, M. (1996). *Analysis Patterns: Reusable Object Models* (1 edition). Menlo Park, Calif: Addison-Wesley Professional.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211–221.

Goodwin, S., Dykes, J., Jones, S., Dillingham, I., Dove, G., Duffy, A., … Wood, J. (2013). Creative User-Centered Visualization Design for Energy Analysts and Modelers. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2516–2525. https://doi.org/10.1109/TVCG.2013.145

Graser, A., Straub, M., & Dragaschnig, M. (2014). Towards an Open Source Analysis Toolbox for Street Network Comparison: Indicators, Tools and Results of a Comparison of OSM and the Official Austrian Reference Graph. *Transactions in GIS*, *18*(4), 510–526.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning. B, Planning & Design*, *37*(4), 682.

Haklay, M. (2014, August 14). OpenStreetMap studies (and why VGI not equal OSM). Retrieved from https://povesham.wordpress.com/2014/08/14/openstreetmap-studies-and-why-vgi-not-equal-osm/

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, *47*(4), 315–322.

Johnson, P. A., & Sieber, R. E. (2013). Situating the Adoption of VGI by Government. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 65–81). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/978-94-007-4587-2_5

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., Ham, F. van, Riche, N. H., … Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, *10*(4), 271–288. https://doi.org/10.1177/1473871611415994

Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual Analytics: Scope and Challenges. In S. J. Simoff, M. H. Böhlen, & A. Mazeika (Eds.), *Visual Data Mining* (pp. 76–90). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-71080-6_6

Lui, M., & Baldwin, T. (2012). langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 25–30). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2390475

MacEachren, A. M., Crawford, S., Akella, M., & Lengerich, G. (2008). Design and Implementation of a Model, Web-based, GIS-Enabled Cancer Atlas. *The Cartographic Journal*, *45*(4), 246–260. https://doi.org/10.1179/174327708X347755

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., … Blanford, J. (2011). Senseplace2: Geotwitter analytics support for situational awareness. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (pp. 181–190). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6102456

McConchie, A. (2015). Tracing patterns of growth and maintenance in OpenStreetMap. In *State of the Map US 2015*. New York, NY, USA. Retrieved from http://stateofthemap.us/tracing-patterns-of-growth-and-maintenance-in-openstreetmap

McHugh, B. (2014). Government as a contributing member of the OpenStreetMap (OSM) community. In *FOSS4G 2014*. Portland, Oregon. Retrieved from https://vimeo.com/album/3606079/video/106226528

Napolitano, M., & Mooney, P. (2012). MVP OSM: a tool to identify areas of high quality contributor activity in OpenStreetMap. *The Bulletin of the Society of Cartographers*, *45*(1), 10–18.

National Research Council. (2010). What Are the Societal Implications of Citizen Mapping and Mapping Citizens? In *Understanding the Changing Planet: Strategic directions for the geographical sciences* (pp. 105–112). National Academies Press.

Neis, P., Zielstra, D., & Zipf, A. (2011). The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, *4*(1), 1–21.

Neis, P., & Zipf, A. (2012). Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *1*(2), 146–165.

Quinn, S. (2015). Using small cities to understand the crowd behind OpenStreetMap. *GeoJournal*, 1–19. https://doi.org/10.1007/s10708-015-9695-6

Quinn, S. (2016). A Geolinguistic Approach for Comprehending Local Influence in OpenStreetMap. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *51*(2), 67–83. https://doi.org/10.3138/cart.51.2.3301

Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, *12*(3), 23–49.

Robinson, A. C., Chen, J., Lengerich, E. J., Meyer, H. G., & MacEachren, A. M. (2005). Combining Usability Techniques to Design Geovisualization Tools for Epidemiology. *Cartography and Geographic Information Science*, *32*(4), 243–255. https://doi.org/10.1559/152304005775194700

Robinson, A. C., Peuquet, D. J., Pezanowski, S., Hardisty, F. A., & Swedberg, B. (2016). Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data. *Cartography and Geographic Information Science*, *0*(0), 1–13. https://doi.org/10.1080/15230406.2016.1139467

Roick, O., Hagenauer, J., & Zipf, A. (2011). OSMatrix–grid-based analysis and visualization of OpenStreetMap. *Proceedings of the 1st European State of the Map*. Retrieved from http://koenigstuhl.geog.uni-heidelberg.de/publications/2011/Roick/Roick_2011_SotM.pdf

Roick, O., Loos, L., & Zipf, A. (2012). A technical framework for visualizing spatio-temporal quality metrics of volunteered geographic information. Presented at the Geoinformatik. Retrieved from http://koenigstuhl.geog.uni-heidelberg.de/publications/2012/Roick/Roick_OSMatrix_Geoinformatik2012.pdf

Rosson, M. B., & Carroll, J. (2002). Scenario-based design. In J. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (pp. 1032–1050). Lawrence Erlbaum Associates.

Roth, R. E., Ross, K. S., & MacEachren, A. M. (2015). User-Centered Design for Interactive Maps: A Case Study in Crime Analysis. *ISPRS International Journal of Geo-Information*, *4*(1), 262–301. https://doi.org/10.3390/ijgi4010262

Shneiderman, B. (1994). Dynamic queries for visual information seeking. *IEEE Software*, *11*(6), 70–77.

Shneiderman, B. (2002). Inventing Discovery Tools: Combining Information Visualization with Data Mining1. *Information Visualization*, *1*(1), 5–12. https://doi.org/10.1057/palgrave.ivs.9500006

Suh, B., Chi, E. H., Pendleton, B. A., & Kittur, A. (2007). Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* (pp. 163–170). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4389010

Thomas, J. J., & Cook, K. A. (2005). *Illuminating the path: The research and development agenda for visual analytics*. Pacific Northwest National Laboratory (PNNL), Richland, WA (US). Retrieved from http://www.osti.gov/scitech/biblio/912515

Trame, J., & Keßler, C. (2011). Exploring the lineage of volunteered geographic information with heat maps. *GeoViz, Hamburg, Germany*. Retrieved from http://carsten.io/trame-kessler-geoviz2011.pdf

Van Exel, M. (2011a, June 20). A New OpenStreetMap Visualization: Version Contour Lines. Retrieved from https://oegeo.wordpress.com/2011/06/20/a-new-openstreetmap-visualization-version-contour-lines/

Van Exel, M. (2011b, September 19). Taking the Temperature of local OpenStreetMap Communities. Retrieved from https://oegeo.wordpress.com/2011/09/19/taking-the-temperature-of-local-openstreetmap-communities/

Van Exel, M. (2014). OpenStreetMap and Telenav; Past, Present and Future. In *State of the Map 2014*. Buenos Aires, Argentina. Retrieved from https://vimeo.com/album/3134207/video/112305387

Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 575–582). Retrieved from http://dl.acm.org/citation.cfm?id=985765

Wood, H. (2014). The Long Tail of OpenStreetMap. In *State of the Map 2014*. Buenos Aires, Argentina. Retrieved from http://vimeo.com/album/3134207/video/112438218