Visualizing Uncertainty in Human Geography Data

2017 IQT Mission Challenge

Report prepared by Andrea Brennen
IQT Mission Challenge Team: Andrea Brennen, Nick Miller, Ravi Pappu, Kelly Pernia, Lisa Porter
Please direct questions or comments about this material to the author at abrennen@iqt.org

IQT
IN·Q·TEL

ABSTRACT

Satellite imagery data will never fully capture the complexity of today's digital landscapes. Open source digital data sources can be a valuable supplement to traditional image collections, but these data sources vary considerably in quality and reliability. Meaningful integration of disparate data sources requires a robust understanding of the uncertainties that exist within them. However, analysts are limited by today's data visualization tools, which do not provide substantial capabilities for conveying uncertainty, particularly in the realm of geospatial data.

In this report we offer a systematic approach to developing capabilities for visualizing uncertainties in human geography data. This approach involves creating a taxonomy of common uncertainties and then designing a visual vocabulary -- a set of standardized, reusable techniques -- that can be used to display these uncertainties. The report also describes how we have applied this approach to create an interactive visualization prototype that helps users explore the Armed Conflict Location and Event Data Project (ACLED), a publicly-available human geography dataset.

ACKNOWLEDGMENTS

Many people contributed to our understanding of the topics discussed in this report. In particular, we would like to thank Todd Bacastow, Peter Beshai, Tony Frazier, Jessica Hullman, Kimberly Kagan, Robert Kosara, Jessica Lewis McFate, Todd Mostak, Ed O'Donnell, Chris Parmer, Jack Parmer, Desiree Petrie, Irene Ros, Diane Staheli, Jim Vallandingham, Katherine Zimmerman, and our US Government partners for their insight and feedback.

Visualizing Uncertainty in Human Geography Data

2017 IQT Mission Challenge

Satellite imagery data is foundational to the Intelligence Community's understanding of world events. However, despite increasing coverage, resolution and availability, this imagery will never fully capture the complexity of today's digital landscapes. Open source digital data sources are a well-recognized supplement to traditional image collections and a vast array of new tools and techniques provide means to collect and process this data. However, it is not obvious how to integrate these open data sources with traditional imagery collections. Some of the challenges are technical — for example, mapping tools like GIS were not designed to handle high-volume, streaming data — but there are methodological challenges as well, as the quality and reliability of open source data differs substantially from foundational, trusted data sources.

Today, a significant obstacle to meaningful integration of disparate data is a lack of capabilities for tracking, measuring, and conveying uncertainty. Libraries of algorithms and analytics offer ways to extract statistical insights from large volumes of data and to compute probabilities that can be compared and — if not controlled — at least controlled for. But data-driven work flows are not immune from uncertainty; each step in the process of collecting, cleaning, processing, analyzing and interpreting data introduces the potential for different types of error and bias. And when different types of data from different sources are integrated, so too are the uncertainties they contain. While it is not possible to remove this uncertainty from open source data collection and analysis, it is critical to represent it, to the extent that it is understood.

Visualizations are an important means of communicating information to decision-makers. Maps, charts, illustrations and infographics are used to represent large quantities of data, to emphasize or reveal relationships in that data, and to help decision-makers digest this information quickly. Increasingly, visualizations serve as a form of evidence, conveying the results of quantitative and qualitative analyses in succinct and legible ways to a variety of stakeholders. However, today's data visualization tools offer few capabilities and no representational standards for conveying uncertainty. These capability gaps are particularly acute in the realm of geospatial data; for maps, there is no standard "error bar" equivalent.

Too often, even when robust analytical methods are used, the nuances of confidence, data quality, and error potential are not visualized simply because today's tools do not provide capabilities to convey this information. Uncertainties may be explained in supplemental text or annotations, but the visualizations themselves often inadvertently imply more certainty than is (or can be) known about underlying data. The IC needs data visualization capabilities that are as robust as their analytical methodologies. And for this, uncertainty must become a "first class citizen" in data visualization.

Looking ahead, the need to visualize errors, biases and uncertainties will only become more important. As aspects of data exploration and analysis are outsourced to automatic or semi-automatic processes, analysts and decision-makers must fully understand the output of those processes. Without a robust representation of uncertainty, there will be no way to assess the integrity or the relevance of results.

In 2016, IQT defined an unclassified "Mission Challenge" focused on visualizing uncertainty in open source human geography[1] data.

1] "The study of the interrelationships between people, place, and environment....concentrat[ing] on the spatial organization and processes shaping the lives and activities of people, and their interactions with places and nature. "Castree, N., Kitchin, R., & Rogers, A. (2013). "Human geography." In A Dictionary of Human Geography: Oxford University Press. Retrieved 14 March 2017.

The goals of this effort were to investigate state-of-the-art capabilities for visualizing uncertainty; to determine how these visualization capabilities might be applied to human geography datasets; and to work towards a standardized, generalizable approach to visualizing uncertainty that would facilitate integration of open source data with other data sources, such as satellite imagery.

We, the project team, then narrowed the scope to focus on a specific dataset that would serve as a representative proxy for other open source human geography data. We selected the Armed Conflict Location and Event Data Project[2] (or "ACLED"), a publicly available dataset that is generally considered a reputable source of information. We then set out to build an interactive visualization prototype (in this case, a web application) to represent the ACLED dataset in a way that would expose known uncertainties in the data. In doing so, we hoped to demonstrate "the art of the possible," but also to call attention to the prevalence of uncertainties that exist, even in trusted datasets. While the visualization prototype was never intended to be a production-ready tool, it was envisioned as a way to document the output of the Mission Challenge while reinforcing the need for robust, systematic techniques for visualizing uncertainties in open source human geography data.

APPROACH

Too many visualization projects attempt to solve the challenges of data collection and storage, of building infrastructure, of buying analysis platforms and of hiring data scientists, before even considering how to convey and communicate results. Coming at the end of laborious and complex data work flows, too often, visualization is an afterthought. But without sustained attention, data visualizations often miss the mark. They fail to deliver meaningful value to decision-makers, or worse, they convey information in a way that is confusing, misleading or inaccurate. In contrast, we began this project by foregrounding the desired visualization product — an interactive web application that would expose uncertainties in the ACLED dataset. Instead of limiting our solution space to visualizations that were easy to generate with current tools, our intention was to work like designers, first envisioning a desired outcome and then defining the capabilities needed to implement it.

Foregrounding visualization design — the appearance, legibility, style, and means of interacting with a visual display of data — can drastically improve the user experience of visualizations. However, it can also lead to specific, customized solutions that are not generalizable across datasets or applications. As the goal of this Mission Challenge was to develop generalizable techniques for conveying uncertainties, we developed and followed a systematic design approach: (1) We extracted specific uncertainties from the ACLED dataset. (2) We abstracted those uncertainties to create a generalizable taxonomy of uncertainty.[3] (3) We mined academic publications and identified a set of known visualization methods for

2] http://www.acleddata.com/

3] The aim of the taxonomy was not to exhaustively describe all possible uncertainties, but rather, to help distinguish between different types of uncertainties that might coexist in a dataset. While a comprehensive depiction of uncertainty is likely a fool's errand, this does not absolve us of the responsibility to visualize the many types of uncertainty that are understood; there is considerable value in distinguishing known unknowns from unknown unknowns. Additionally, open source datasets contain many different types of uncertainty, which vary widely in their quantifiability. In the first version of the taxonomy, we focused primarily on uncertainties resulting from measurable errors; these uncertainties are much more straightforward to characterize than uncertainties resulting from biases and judgments made during the analysis process.

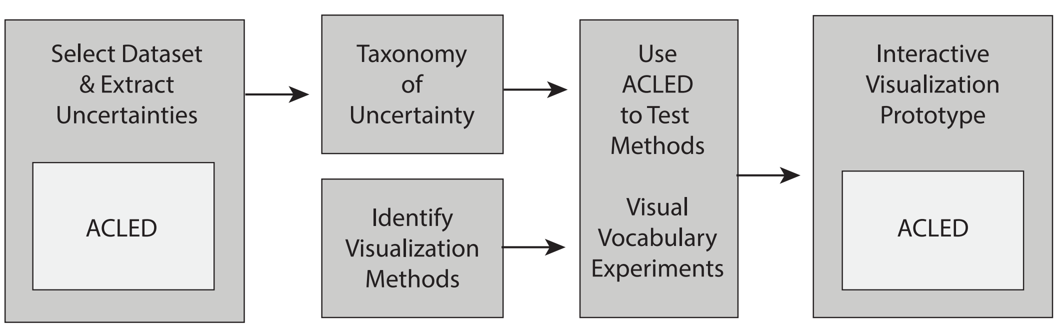


Figure 1: Project Approach

conveying uncertainties. (4) We tested[4] these methods by applying them to the ACLED dataset over a series of quick experiments. And (5) we built an interactive visualization prototype to showcase several visualization methods concurrently.

Following this systematic approach, we aimed to develop a reusable visual vocabulary of uncertainty, a re-combinable set of visualization techniques that can be assembled to convey the different types of uncertainty commonly found in open source human geography datasets. A composite visualization, then, might leverage multiple elements from this vocabulary in order to help decision-makers understand and distinguish among different types of uncertainty. Ultimately, the goal was to work towards a publicly-available, open source visual vocabulary, that could benefit from the contributions of others and might inform (or potentially be incorporated into) existing data visualization tools.

SOUTH SUDAN AND THE ACLED DATASET

ACLED is a collection of conflict events in Africa and Asia, curated from a variety of local, regional, national and international news sources and NGO reports. The dataset is maintained by a team of researchers at the University of Sussex[5] and updated weekly. Each event in the dataset has a location and time of occurrence, a brief description of primary actors, and when applicable, a corresponding fatality estimate. Each event is assigned an event type from a predefined list of categories (Battle, Riot/Protest, Violence against civilians, etc.) and information about the source of the data is provided. When aggregated, the frequency and severity of events in a region provide a timely overview of crisis, conflict and political violence.

For this project, we focused on South Sudan, where conflict event data might be used to anticipate human migration trends within the country. For example, one might use this data to estimate the locations of Internally Displaced Persons (or "IDPs"), people who have been forced to leave their homes but who stay within the borders of South Sudan. The Internal Displacement Monitoring Center (IDMC) estimates that as of December 2015, at least 1.69 million people were internally displaced in South Sudan alone, mostly due to conflict.[6] Better awareness of the locations of internally displaced people could have tremendous humanitarian value as it could greatly improve the ability of aid organizations to deliver food and medical supplies. The ACLED dataset can assist with this goal, but only if the accuracy and reliability of that data is well understood.

The *ACLED Codebook*[7] contains a thorough explanation of the methodologies used to collect, curate and ensure the quality of ACLED data. Two types of uncertainty are explicitly quantified: spatial precision,[8] a scale of 1-3 that indicates how precisely the location of an event is known and time precision,[9] a scale of 1-3 that indicates how precisely the date of an event is known. Additionally, the *Codebook* describes cases where information is often missing or known to be biased. For example, Section 7 (on fatalities) explains that "very often, no fatality information is available for events from sources and such reported fatality totals are often erroneous, as the numbers tend to be biased upward."[10]

TAXONOMY OF UNCERTAINTY

Generalizing from the three types of uncertainty explicitly referenced in the ACLED Codebook — spa-

4] Further work is needed to develop and apply evaluation methods to more rigorously determine the effectiveness of various techniques.
5] ACLED is overseen by Prof. Clionadh Raleigh and operated by Andrea Carboni for Africa and Hillary Tanoff for Asia.

6] Globally, the two most common causes of internal displacement are conflict and disasters. Source: IDMC (Internal Displacement Monitoring Center) http://www.internal-displacement.org/sub-saharan-africa/south-sudan/figures-analysis
7] http://www.acleddata.com/wp-content/uploads/2017/01/ACLED_Codebook_2017.pdf
8] From Section 3.2 of the ACLED Codebook: "If the source notes a particular town…the highest precision level '1' is recorded….if the source material notes that activity took place in a small part of a region…a town…to represent that area is chosen and the geoprecision code will note '2' …if a larger region is mentioned, a provincial capital is chosen…and noted with precision level '3'."
9] From Section 4.1: "If sources include an actual date, '1' is chosen as the precision level. If sources note a week, '2' is noted…and the first date of that week is used as the reference date. If sources note only that an activity took place within a particular month…the month mid-point is chosen…and '3' [is listed] as the precision level unless the beginning or end of month is noted (in which case, the first and last date are used, respectively)."
10] It may be that news sources are prone to sensationalize events and/or that aid organizations are financially incentivized to overestimate fatality counts. If biases such as these are present in news reports, the fatality counts in ACLED would reflect this bias. From the Codebook: "ACLED only codes estimated casualties when reported by source materials. It cannot verify the numbers reported from sources and does not use fatalities as the basis for event inclusion. Very often, no fatality information is available for events."

tial precision, time precision and estimated fatality counts (which have a known bias) — we worked to categorize and define additional types of uncertainty that are generalizable across human geography datasets. The initial taxonomy of uncertainty is shown in Figure 2 and working definitions of each term are included below.

Spatial Uncertainty refers to the physical location of an event, object or person. When an exact location is not known, the precision (or imprecision) is often described in one of three ways: (1) The uncertainty around the true location is described by a circular error ***probability***. This enables us to make statements like “the probability that the estimated location is within a given radius of the true location is 50%, or 90%, or 95%.” This approach is used to set an upper bound on how far the reported or estimated location is from the true location. (2) The true location of the event may be within a region that is described by an abstract geometric construct such as a ***square grid area***[11] projected onto some geographic region. In such cases, visualization systems might adopt a convention of simply placing the location at either one corner or at the center of a square within the grid. (3) The true location may be within a region that is described by a (non-geometric) boundary such as a geographic feature, political boundary, or ***administrative region***. The ACLED dataset’s spatial precision codes are an example of this third category.

Temporal Uncertainty refers to the date and time when an event occurred. A common way of indicating temporal uncertainty is to describe the known precision in date or time ranges (as with ACLED time precision codes). For example, an event may have happened the week of March 1, or sometime between Wednesday and Friday. In the taxonomy, this is referred to as ***Calendar Uncertainty***. An alternative way of indicating temporal imprecision, ***Chronological Uncertainty***, involves the order of events; for example, it may be known that a particular event

11] Two examples of grid systems are the Military Grid Reference System (MGRS) and the Universal Transverse Mercator (UTM). More information can be found at: http://earth-info.nga.mil/GandG/coordsys/mmr201.pdf

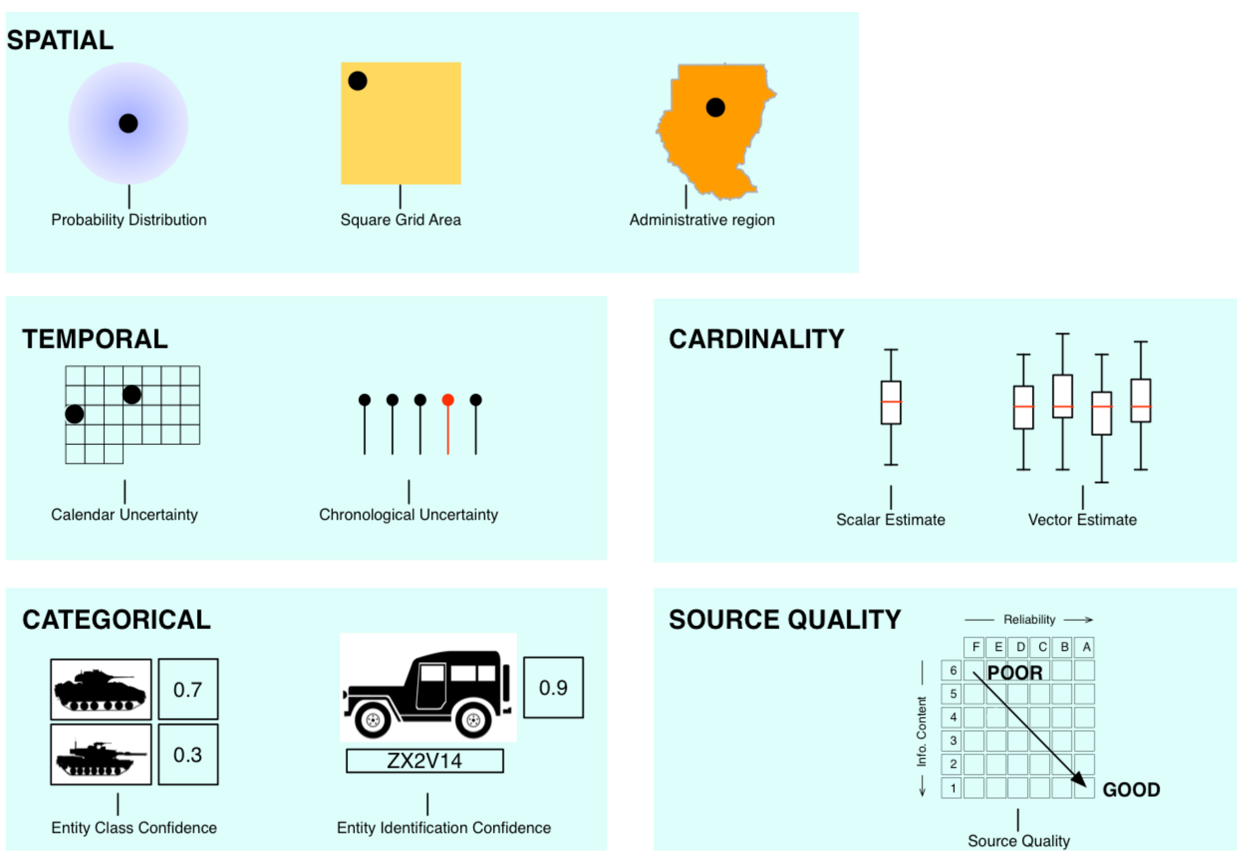


Figure 2: Taxonomy of Uncertainty

happened before (or after) another event. Even if precise dates or times are not known, the order of events may be.

Cardinality refers to uncertainty about counts or amounts of people or things. A ***scalar estimate*** — such as the number of fatalities associated with an event — expresses variability within a single dimension or quantity. A ***vector estimate*** incorporates variability across multiple dimensions. The fatality counts in ACLED are an example of a scalar estimate.

Categorical Uncertainty arises when things are sorted into predefined categories and an attempt is made to quantify the probability of misclassifying something as the wrong type of thing. For example, if a vehicle is known to be either a Type A vehicle or a Type B vehicle, the vehicle might be labeled "Type A," with an accompanying ***Entity Class Confidence*** of 0.7. This confidence level — 0.7 — implies that whoever assigned the vehicle to the "A" category is 70% certain that the vehicle is, in fact, a Type A vehicle. ***Entity Identification Confidence*** refers to the probability that a specific vehicle is identified correctly, for example, as having a particular VIN number. Events in ACLED are categorized according to predefined types, but no indication of entity class confidence is provided.

Source Quality refers to the degree of trust placed in the source of information. For this, we chose to use an existing two dimensional scale known as the Source and Information Reliability Matrix (SIRM)[12]. In this framework, information is rated on two independent scales: the reliability of the source and the credibility of the information (often, whether it has been confirmed by other sources). ***Reliability*** is rated on a scale of A to F, where A is "completely reliable," E is "unreliable," and F is unknown. ***Credibility*** is rated on a scale from 1-6, where 1 is "confirmed by other sources," 5 is "improbable," and 6 is unknown or cannot be judged. ACLED event data is curated from a variety of media sources and those sources, presumably, vary in their reliability. However, an explicit ranking of sources by reliability is not provided. Instead of making judgments about various sources (which would have introduced additional uncertainties), we decided to simulate source quality for the purposes of this project. Events were arbitrarily assigned reliability and credibility scores following the SIRM for demonstration purposes only.

VISUALIZATION EXPERIMENTS

We reviewed existing research papers, tools and examples to collect a set of known visualization methods for conveying uncertainties. Then, we worked with Bocoup[13], an open source software development consultancy, to conduct a series of experiments to test these methods by applying them to uncertainties in the ACLED dataset.

The experiments were small and quick — a few lines of code were written to visualize a single aspect of uncertainty using a particular visualization technique. And the outcome of each experiment was visual — the team created either a static mock-up, an animation, or an interactive visualization, depending on the method tested. The purpose was to create a series of visualization options that we could compare before deciding which to include in a composite visualization prototype.

The table in Figure 3 summarizes the 13 experiments that were conducted, showing which visualization methods were tested against which types of uncertainty. Red boxes show the experiments that were ultimately included in the prototype. All of the experiments were conducted over a two week period and archived on an internally hosted code repository at IQT. Given the limited time, several promising options were not tested, but we plan to expand upon these experiments in future work.

Despite considerable variation in graphics and appearance, existing strategies for visualizing uncertainty fall into a few categories. Most often, uncertainty is treated as an additional dimension or attribute of data[14] and visualized through commonly used graph-

12] A full definition of the SIRM may be obtained from Appendix B of the US Army Field Manual FM 2-22.3 Human Intelligence Collector Operations, available at https://fas.org/irp/doddir/army/fm2-22-3.pdf

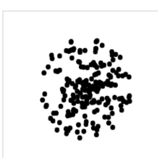
13] www.bocoup.com

14] Several publications address this strategy, including: (1) Alan M. MacEachren, Anthony Robinson, Susan Hopper, Steven Gardner, Robert Murray, Mark Gahegan, and Elisabeth Hetzler. "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know." Cartography and Geographic Information Science. Vol. 32, No. 3, 2005. (2) K. Potter, J. Kniss, R. Riesenfeld, and C.R. Johnson. "Visualizing Summary Statistics and Uncertainty." Eurographics/ IEEE-VGTC Symposium on Visualization. Vol. 29, No. 3, 2010. And (3) Alan M. MacEachren, Member, IEEE, Robert E. Roth, James O'Brien, Bonan Li, Derek Swingley, Mark Gahegan. "Visual Semiotics & Uncertainty Visualization: An Empirical Study." IEEE Transactions on Visualization & Computer Graphics.

ical techniques such as color, transparency/opacity, size, line type, or special "glyphs," etc. The visualizations in Figure 4, which were produced during Experiment #1, provide an example. In the image on the right, circle size is used to show spatial uncertainty. A larger circle indicates a less certain event location. Circle size, however, could just as easily be used to encode a different data attribute, such as the population of a city or the number of conflict events in a given region. During Experiment #1, we also realized that this technique is not ideal for showing spatial uncertainty; it inadvertently over-emphasizes less-certain events by highlighting large areas around those events. This example also illustrates the difficulty of developing a visual vocabulary of uncertainty, as choosing effective visual representations is not straightforward.

Type of Uncertainty *ACLED example*	**Treat as a data attribute** *assign uncertainty a graphical dimension -- color, size, etc.*	**Use distinct chart type** *such as box plot or violin plot*	**Use blurry or fuzzy graphics** *make uncertain data appear blurry or fuzzy*	**Animate** *interpolate between possible outcomes*	**Simulate outcomes** *visualize multiple possibilities*
Spatial *location of event*	1 2 3		4	5 6	
Temporal *date of event*	7 8		9		
Cardinality *number of fatalities*				10	11
Categorical *type of event*				12	
Source Quality *reliability & credibility*	13				

5

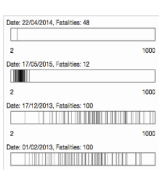


"Wandering dots"
animate the position of dots to "wander" around the area where an event is known to have occurred.

10

HOPs
or "Hypothetical Outcome Plots" show a finite set of outcomes; animated marks move between possible values.

11



Rug plots
show distributions of a variable along an axis with tick marks, reminiscent of the tassels on a rug.

13

Check boxes
allow users to select different categories of uncertain-to-certain data to display.

Figure 3: Visualization Experiments and (in Red) Techniques Selected for Prototype

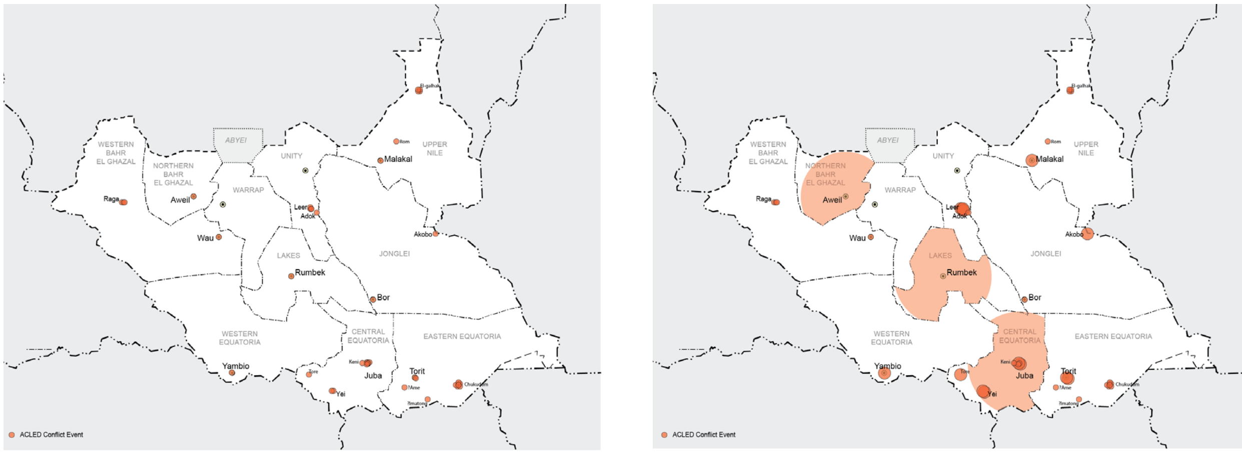


Figure 4: Two Images Produced During Experiment #1.
The image on the left shows the locations of 41 conflict events that occurred in South Sudan between October 1 and October 15, 2016. The image on the right shows the same 41 events, but the orange circles are scaled by spatial precision code so that the size of each circle represents the area in which an event is known to have occurred.

An alternative strategy is to treat uncertainty as distinct from other data types. This may lead to the use of distinct chart types like box plots[15] or violin plots[16]. In other cases, visualizations might incorporate a unique graphical treatment — like blurriness or fuzziness — that is metaphorically associated with the idea of uncertainty.[17]

A third strategy is to treat uncertainty as a probability distribution and visualize possible outcomes. In static visualizations, multiple possible outcomes may be simulated and then super-imposed on a single chart. This strategy is commonly seen in weather maps that show projected hurricane trajectories. It is also used in "Rug" plots,[18] which use different colored tick marks to show distributions of a variable along an axis. Animation[19] may also be used to show different possibilities, as in Jessica Hullman's Hypothetical Outcome Plots (HOPs).[20] In these plots, animated bars cycle between different possible data values. While conducting the experiments, we expanded upon this idea to develop a new animated technique to visualize spatial uncertainty that we called "Wandering dots." With this technique, the location of dots (representing conflict events) are animated to "wander" around an area where the event is known to have occurred.

VISUALIZATION PROTOTYPE

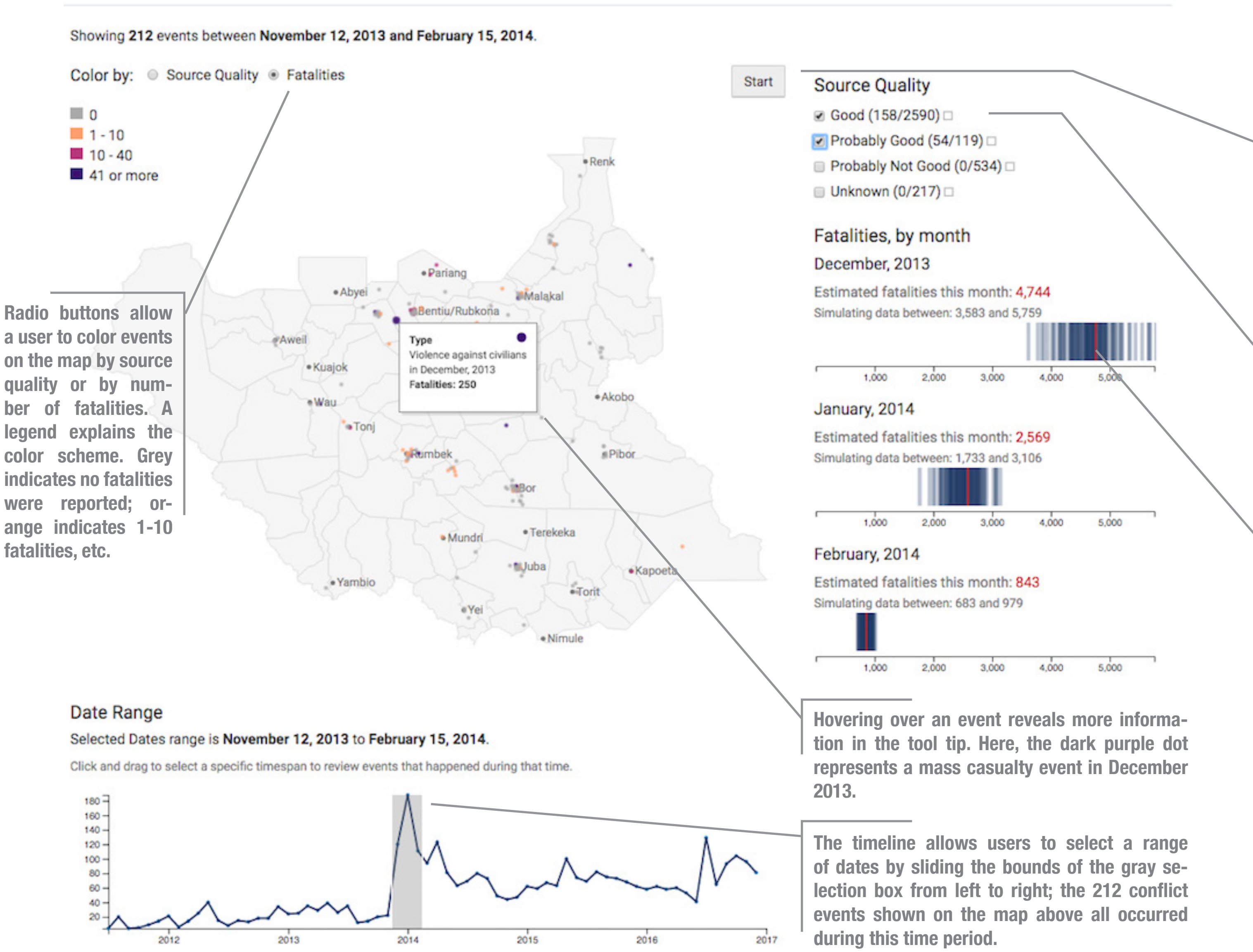
Working with Bocoup, we then integrated several visualization techniques into an interactive prototype. The output was a data-driven web application that serves as a visual interface or "front-end" to the ACLED dataset. Several types of data are shown in the prototype (see Figure 5) — event locations, event types, dates of occurrence, fatalities, the total number of events over time, and aggregated fatality counts. And three types of uncertainty are visualized -- spatial uncertainty surrounding the precise location of events, fatality estimates, and source quality. The prototype includes three principle views: a map showing the locations of conflict events, where each event is represented with a small dot; a timeline showing the number of events over time; and a

15] http://www.physics.csbsju.edu/stats/box2.html
16] http://www.datavizcatalogue.com/methods/violin_plot.html
17] The use of blurry graphics is addressed in: Alan M. MacEachren. "Visualizing Uncertain Information." Cartographic Perspectives. No. 13, Fall 1992. http://www.cartographicperspectives.org/index.php/journal/article/viewFile/cp13-maceachren/1039. Graphics that appear fuzzy and/or "sketchy" are also addressed in: (1) http://openaccess.city.ac.uk/1274/ (2) https://hal.inria.fr/hal-00717441 and (3) https://eagereyes.org/blog/2012/visweek-2012-digest-part-2
18] https://www.mathworks.com/matlabcentral/filee-change/27582-rug-plots?requestedDomain=www.mathworks.com
19] This "Jittery Gauge" blog post is an example of one possible use of animation to convey uncertainty in a possible set of values : http://vis4.net/blog/posts/jittery-gauges-election-forecast/

20] Jessica Hullman, Paul Resnick and Eytan Adar compare Hypothetical Outcome Plots (HOPs) to error bars in: Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering (https://doi.org/10.1371/journal.pone.0142444) series of horizontal plots showing aggregated fatality counts. These three views are linked and provide several options for filtering the subset of data that is displayed. Within these views, wandering dots show spatial uncertainty on the map, check boxes allow users to filter information by (simulated) source quality, and fatality estimate distributions are shown through a combination of HOPs and Rug plots. When an animation feature is activated, fatality estimates are shown in a HOP; when the animation is stopped, all of the possible estimates are super-imposed in a static Rug plot.

By integrating multiple visual techniques into a composite visualization, this prototype illustrates how visual vocabulary elements can be combined to show different types of uncertainty that coexist within a dataset. Displaying these uncertainties in different ways can help decision-makers distinguish among them. This is important because the relevance of a particular uncertainty depends on how the data is used. For example, the screenshot in Figure 5 highlights a mass casualty event described as "violence against civilians". The location of the event is indicated by a dark purple dot near the town of Bentiu and according to the tool tip,

Figure 5: Interactive Visualization Prototype



Radio buttons allow a user to color events on the map by source quality or by number of fatalities. A legend explains the color scheme. Grey indicates no fatalities were reported; orange indicates 1-10 fatalities, etc.

Hovering over an event reveals more information in the tool tip. Here, the dark purple dot represents a mass casualty event in December 2013.

The timeline allows users to select a range of dates by sliding the bounds of the gray selection box from left to right; the 212 conflict events shown on the map above all occurred during this time period.

there were an estimated 250 fatalities. After seeing this information, an analyst might want to take some action — such as tasking a satellite or other platform to collect additional imagery data about the place where this event occurred. For this decision, uncertainty about the precise number of fatalities might not be relevant; it may be that knowing the order of magnitude is sufficient. However, it would be important for the analyst to understand any uncertainties about the precise location of the event.

This ACLED event has a spatial precision code of 3, meaning that only the administrative district where the event occurred is known. The dataset includes latitude and longitude coordinates for the event, but these point to Bentiu, the (former) capital of Unity state. (As the *ACLED Codebook* explains, for events coded 3, "a provincial capital is chosen" for the location). Most visualization tools only provide the ability to display a location, not the uncertainty surrounding the location. But here, if only the given lat/long coordinates were visualized, an analyst might draw the (erroneous) conclusion that this mass casualty event happened in Bentiu.

In contrast, this prototype visualization provides an animation feature to show the spatial uncertainty associated with this event; when this feature is activated (by pressing the "start" button), the location of the dark purple dot "wanders" around the region where the event is known to have occurred. In this case, that region is the entire (former) Unity state[21], an area of 14,995 square miles. Visualizing this spatial uncertainty — the margin of error associated with the location provided in the dataset — underscores the need for additional information before any collection platforms are tasked.

The "Start" button activates an animation that conveys the spatial uncertainty associated with the location of events. When the animation is running, event dots "wander" around the zone in which they are known to have occurred. The size of this zone is determined by the spacial precision code listed in the ACLED dataset. If an event has a spatial precision code of 1, the dot will not move; if it is labeled "2" it will move within a small zone; if it is labeled "3" it will wander around an entire administrative district to reflect the wide range of possible locations.

Check boxes allow a user to filter the events by source quality. For ease-of-use, only four simplified categories are shown: Good, Probably Good, Probably Not Good, and Unknown. Currently, only events backed by "Good" or "Probably Good" data are shown.

Rug plots show aggregated fatality estimates for the selected time period. The plots are either aggregated by month, by quarter, or by year depending on the range selected. The red band indicates the sum of fatalities for all selected events. According to "Good" and "Probably Good" quality data, there were 4,744 reported fatalities during December 2013. However, these counts are known to be estimates. The blue bands show the distribution of possible fatalities, where darker blue indicates a higher probability of occurrence. Here, the number of fatalities might have been as low as 3,583 or as high as 5,759, but was probably between 4,500 and 5,000. When the animation feature is activated, each Rug plot is replaced by a HOP and instead of showing all possible outcomes (all blue bands) simultaneously, only one possibility is shown at a time. The position of the visible blue band is animated to reflect different possible outcomes.

CONCLUSION & FUTURE WORK

There may be good reason to limit data visualization capabilities for methodological, privacy or security reasons, but technological capabilities should support the implementation of intentional policies, not artificially restrict which aspects of available data analysts can see. In order to work towards robust capabilities for visualizing uncertainty, we plan to extend our work to create a publicly-available, open source visual vocabulary of uncertainty that is thoroughly documented and illustrated with examples. We also anticipate expanding the taxonomy of uncertainty to incorporate data types from other domains and testing a variety of additional visualization techniques. In parallel, we must work towards a robust approach to evaluating visualization techniques, so that we may develop best practices around the effective usage of vocabulary elements. Ideally, these visualization efforts will be supported by the development of standardized metadata formats for capturing uncertainty and easy-to-use computational techniques for extracting uncertainties from a variety of datasets.

Visualizing uncertainty does add complexity to the way open source human geography data is displayed, but this complexity exists in the underlying data, regardless of whether or not it is displayed. Displaying these uncertainties is the only way to ensure that our tools do not inadvertently fool us into thinking we know more than we do, by showing us more certainty than can be known.

21] Bentiu was the capital of Unity state in December 2013, when this event occurred. However, in 2015 the states of South Sudan were reorganized by Presidential decree and Unity was divided into Ruweng, Northern Liech, and Southern Liech.