Confidence as First-class Attribute in Digital Humanities Data

Max Franke*, Ralph Barczok†, Steffen Koch*, Dorothea Weltecke†

Abstract—Digital humanities scholars deal with many types of uncertainty in data. Especially historians work with sources that can be incomplete, hard to translate, or biased. This introduces uncertainties, which are difficult to assess or quantify, and typically require critical interpretation through experts. We, therefore, use the term *confidence*, which seems to be semantically more adequate as a measure of the validity of historical data and findings, thus reflecting both biases of the creators of historical artifacts, and human interpretation involved in determining sources' uncertainty. We emphasize the benefits of treating confidence as a first-class attribute of the data in the processing steps as well as during visual analysis. This allows us to interact with confidence directly, enabling analysis possibilities which can identify the need for additional data acquisition, or reveal open questions in the interpretation of historical sources. In views tailored to the depiction of specific attributes, data with very low confidence, or incomplete data, results in the omission of data items. To overcome this issue, we propose using auxiliary views reflecting data with high uncertainty to avoid misinterpretation. We demonstrate these improvements in the context of a multiple-coordinated-views prototype for visual analysis of coexistence of institutions of diverse religious communities in cities of the medieval Middle East.

Index Terms—Digital humanities visualization, confidence, uncertainty, historical data

1 Introduction

Uncertainty and trust have become a research focus in visual analysis, where these aspects require consideration on multiple levels and in different processing steps, as Sacha et al. [31] discuss. The stages where uncertainties can be introduced include the recording of the data, data (pre)processing, the application of automated analysis methods, visualization, interpretation, and manipulation. However, making uncertainty visually perceivable and helping analysts taking into account uncertainties in data visualization has the potential to increase trust in analysis approaches and the trustworthiness of analysis results. Accordingly, uncertainty aspects and their respective origin play an important role when visual approaches are developed to support scholars in answering research questions in the humanities.

After developing visualization support for a number of digital humanities projects, we argue that the importance of uncertainty in these endeavors not only justifies a visual acknowledgment or reflection of uncertainty, but deserves a more prominent role as a first-class data attribute in such visual analysis approaches. In addition, we discuss uncertainty characteristics that are typical for data sets and analysis tasks in the humanities.

1.1 Visual Analysis Using Confidence

We propose an explicit, multi-faceted model that can be used to interactively filter visualized data according to specific aspects of uncertainty, or their combination. This allows to incorporate uncertainty in the analysis process, which offers a number of benefits for data collection and analysis tasks: (i) Data quality enhancement is supported by focusing on specific uncertainties in data creation and refinement. Uncertainties can reveal systematic weaknesses in the data collection and suggest directed search for new or additional data sources to improve certainty considering combinations of data characteristics. Awareness of uncertain data can guide collaborative efforts in assessing information from uncertain sources. (ii) Consideration of uncertainty during analyses creates awareness of the certainty of analytic results, and sources of conflicting data can be assessed critically. The two basic tasks in complement can increase both data quality and analytic trustworthiness, which supports iterative foraging and sense-making loops in visual analytics or related models [28, 32].

* *Institute for Visualization and Interactive Systems, University of Stuttgart, Germany. E-mails: {firstname}.{lastname}@vis.uni-stuttgart.de*

† *Historical Seminar, Goethe-University in Frankfurt/Main, Germany. E-mails: {lastname}@em.uni-frankfurt.de*

Another challenge is low-confidence or even unknown data attributes. Geographical location is a good example to illustrate the problem: Interactive maps are a common, suitable visual representation for places. If a historical place is unknown—or its location is very uncertain—it is difficult to show it on a map adequately. The situation is similar for other attribute-specific views. We, therefore, propose auxiliary views that show representations of data items that cannot be expressed in particular views because of low-confidence attributes.

After many discussions, we agreed on using the term *confidence* instead of uncertainty in our collaborative project. The main reason is that this term describes the nature of the sources of uncertainty we are dealing with better and fits the terminology used by historians. Already the initial steps in humanities projects, i.e., manual data collection, often involve interpretation or judgment by an expert. A historian who has worked extensively with the respective sources has gained an understanding of the sources' trustworthiness that will influence the historian's judgment of the information collected from them. Accordingly, this context knowledge—which is not expressed explicitly in the data or the sources—is worth being represented in the form of explicit confidence for a number of reasons including subsequent visual analysis tasks. Not only human statements need to be assessed with respect to confidence, often other data attributes—such as time or location—require human interpretation if they are not stated explicitly in historical documents. We are aware that the term confidence is already used in the context of uncertainty visualization. This is the case in particular for statistical methods and automatic procedures that are capable of producing confidence values, but also for human interpretation of the visualization and the analysis. We use the term "*confidence*" rather than "*uncertainty*" to emphasize the holistic perspective on historical data, but we do include its usage in other uncertainty contexts [15].

1.2 Project Background

In our project, visualization researchers collaborate with historians who study the diversity of institutionalized religions in medieval Muslim cities, specifically from the 7th to the 14th century. During this time, specific non-Muslim groups were tolerated, accepting, in turn, a lower legal status. This legal construction and political pragmatism led to a great religious diversity in these cities. The research interests of the historians include analyzing the coexistence of institutionalized religious communities in cities of the medieval Middle East, understanding the evolution of multi-religious variety in the Middle Ages, and comparing different perspectives of scholars of that time period, amongst other goals. The historians collect pieces of historical evidence hinting at the existence of institutionalized religious communities from contemporary documents (primary sources) and literature (secondary sources). These sources can be written in different languages, and are available in different analog or digital qualities. Suitable information on such a religious community includes information on the religious affiliation, e.g., a reference indicating institutionalization such as sacred buildings or religious dignitaries, as well as time, and the location or city. Such historical sources implicitly contain uncertain data, where the uncertainty has multiple sources and requires a human assessment to determine confidence.

Our collaboration is part of an ongoing project, where data collection—as well as the development of the visualization—are still in progress. As part of this development, explicit confidence attributes were just recently added. Previously, the historians entered information about confidence in a general-purpose, free-text comment field available in the data model, for example as "*Location unknown, approximately here based on 'literature X'*". As a result, the fields for confidence are only sparsely filled currently, but this is changing quickly. We discuss and develop the visual analysis of confidence aspects in our collaboration iteratively, using dummy confidence data.

This development—and the feedback on it—supports our suggestion to make confidence a first-class attribute of data for interactive visual analysis approaches. We describe an interactive visual prototype that incorporates confidence analysis for understanding coexistence of institutionalized religious communities in the medieval Middle East. While we consider several aspects relevant for assessing confidence, we also make some pragmatic choices that reduce the theoretical model to a feasible practical one. The two main objectives of the approach are (1) to make scholars aware of data confidence during visual analyses, and (2) to facilitate the incremental improvement of data quality.

2 Related Work

The communication of uncertainty to the viewer of a visualization, as well as the sources and propagation of uncertainty and the trust-building process between visualization and viewer, have been much-discussed topics in recent years. MacEachren et al. [22] survey several approaches to visualizing and dealing with uncertainties in geovisualization, and list seven research challenges in uncertainty visualization. Kinkeldey et al. [18] compare several approaches to visualization of uncertainty in geospatial data. Further work [7, 11, 24, 26, 33, 34] also emphasizes the importance of the communication of uncertainty and better interaction techniques to facilitate objective analysis of the data, or evaluate [17, 23] the effectiveness of different visual variables for communicating uncertainty. Hullman et al. [15] survey evaluations of uncertainty visualization. They list *confidence* as an expected effect thereof, describing it as the "*degree of belief in the validity or truth of a judgment, data set, visualization, etc.*". Considering the human factors involved in the creation, transmission and collection of historical sources, this matches our use of the term closely.

Muir [25] emphasizes the importance of trust between humans and machines. Sacha et al. [31] expand the knowledge generation model for visual analysis [32] to model how uncertainties in the data are propagated, further uncertainty can be introduced, and how trust is built on the human side of the analysis. Further work discusses missing data [10, 16, 35, 37] or visual encodings [4] as sources of uncertainty and mistrust.

Thomson et al. [39] categorize uncertainty in analytics into nine types: accuracy/error, precision, completeness, consistency, lineage, currency/timing, credibility, subjectivity, and interrelatedness. Further work also classifies types and sources of uncertainty on different levels [12, 13, 35, 38], and Griethe and Schumann [13] discuss the challenges of visualizing and dealing with multi-dimensional uncertainty.

Treating confidence as a first-class data attribute in visual analysis allows for filtering by confidence. Paradis and Beard [27] propose a filter by data quality for geographical information systems which allows users to modify the shown data based on quality attributes. De Lange et al. [9] work with very similar historical data, and their web interface for the data also visualizes aspects of uncertainty. While their approach allows to filter by different data aspects, filtering by uncertainty is not supported. In digital humanities data, where uncertainty is often less precise, several works [1, 2, 9, 29] model uncertainty to a discrete set of labels—or colors [8]—as well as use the labels to model alternative truths. Wrisley and Jänicke [44] emphasize the uncertainties originating from historical documents, specifically difficulties with placing historical toponyms on a map. Windhager et al. [42] discuss difficulties in modeling spatio-temporal uncertainties in cultural heritage collections, as well as visualization techniques.

Visualization of uncertainty alongside other data has been explored in depth by many works. These map uncertainty to color hue [5, 6, 8, 23], color value [5, 23, 36], color saturation [5, 8, 23, 40], transparency [6, 19, 29], texture [3, 5, 21], blurriness [6, 20, 23, 33], squiggliness [14], or sketchiness [43]. These techniques require that the visual variable in question is not yet used, and does not interfere with other visual variables already. Contrary to those works, our approach shows confidence alternatively when the *confidence coloring mode* is enabled, replacing the meaning of color hue.

We aim to integrate confidence directly into the visual interface, treating it as a regular, first-class data attribute. Consequently, typical interactions in visual analysis, such as orthogonal filtering by different attributes or brushing and linking, can be seamlessly extended to confidence. Unknown attribute values are represented in auxiliary views in our solution. To our best knowledge, confidence or uncertainty have not been integrated into visual analytics approaches in this form so far.

3 Confidence in the Data Processing Pipeline

Sacha et al. [31] identified sources of uncertainty in visual analytics, and how uncertainty propagates from the data to the visualization and further to the user. Their approach assumes the data is already available in digital form, and considers human factors and uncertainty in the visual analytics model. McGurdy et al. [24] discuss sources of *implicit error* in human-collected data, and how domain experts can judge the truth content of a piece of information better with context knowledge. In the domain of historical research, human factors must be considered from the first recording of a piece of information. We use the term *confidence* when talking about the reliability of recorded historical facts to emphasize the human factors and the nature of the data.

We define six aspects of confidence, which are described in detail in Sect.s 3.1 and 3.2: *time*, *religion*, *location*, *place attribution*, *source*, and *interpretation confidence*. We add additional levels of abstraction to the confidence aspects, which allows for coarser-grained searches if required: *Total place confidence* abstracts *location* and *place attribution confidence*, and *evidence confidence* abstracts *source* and *interpretation confidence*. *Total confidence* abstracts all aspects of confidence. Fig. 1 shows our model of the pipeline for historical data from the event to the visual analysis stage, and how confidence in the data is affected in the different stages. In the following, we will discuss the different stages, and how confidence in the data is affected in each stage.

3.1 Data Generation and Propagation

First issues with confidence in the data arise with the creation of historical documents (Fig. 1 1 and b). For one, the author of the historical source might have a perspective that does not reflect objectivity as we

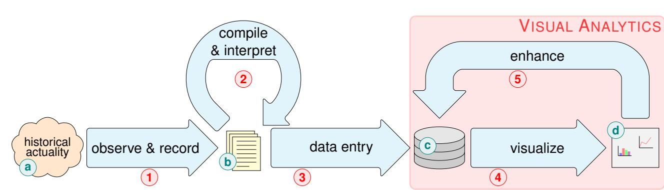

Fig. 1. The historical data processing pipeline. A a historical actuality is 1 recorded as a b primary source. b Secondary sources may 2 compile and interpret primary and other secondary sources. Historians 3 enter the evidence collected from both primary and secondary sources into the c database. The data from the database is 4 visualized, and historians 5 find and correct errors and inconsistencies they find via the d visualization. Colleagues review the already present data, and 5 verify and corroborate their veracity, improving the source interpretation confidence. The red box is an abstraction of the Visual Analytics knowledge generation model described by Sacha et al. [31].

understand it today. Thomson et al. [39] categorize this as *deception*, but also as *subjectivity* and *credibility*, the latter of which is also used by Therón et al. [38]. Such deliberately biased representations arise from an author's wish to represent the facts in a specific light. This ranges from minor acts of framing, over deliberate omissions of facts, to plainly made-up and untrue descriptions. Such modifications can be difficult to detect in historical documents, since the author would not want to advertise ideological or religious biases whether they are aware of a bias or not. Historians use ancillary sources as well as context knowledge to infer the confidence level which can be attributed to the source. Confidence is entered as a qualitative measure from a small set of labels, to avoid a false sense of precision. We use the term *source confidence* for the confidence in the authenticity, objectivity and veracity of the source a piece of historical evidence is based upon.

Lack of interest by the author of a source can also introduce bias. For example, a traveling Christian monk might report on a Christian community he encounters, but omit the Jewish or Islamic communities in that city. Analysis of that source as a free-standing piece of evidence can leave the impression that the city in question is entirely Christian. This bias is ameliorated by other sources, which complete the picture, and by context knowledge of the historian. As additional sources do not exist in every case, the source's bias still needs to be recorded. In the categorization of Thomson et al. [39], this bias can be categorized as *subjectivity* and *completeness*, and Therón et al. [38] use the terms *incompleteness* and *ignorance*. The incompleteness does not affect confidence in the existing data, but rather implies further, missing data.

The confidence level of a piece of historical evidence is further affected by whether the information is acquired from a primary or a secondary source (Fig. 1 ⓑ). A secondary source is a piece of work by a scholar who has compiled and analyzed (Fig. 1 ②) other, primary or secondary, sources. Thorough research by the scholar can compensate for bias in the sources, but they might add their own bias, as described in Sect.s 3.1 and 3.2. Thomson et al. [39] categorize this as *lineage*.

3.2 Data Entry

From the historical sources, the historians extract single pieces of historical evidence (Fig. 1 ③). One piece of evidence is, loosely defined, a tuple of a time or time range, a place, and a religious affiliation. Each of those attributes is annotated with an aspect of confidence, as well as the *source confidence* discussed in Sect. 3.1. We use five distinct labels for confidence: *unknown*, *uncertain*, *probable*, *confident* and *true*. We choose a small set of labels to avoid different interpretation of the labels by different scholars as much as possible. The data entry is a manual process, which is facilitated by a web interface where existing religious denominations, as well as locations with their alternative names, are suggested on entry.

The manual process of extracting historical evidence from hard copies is susceptible to data entry errors. Data entry is an unavoidable challenge for confidence in data, and must, therefore, be considered as an implicit source of error by the historians. Thomson et al. [39] categorize this as *collection accuracy*. Through explorative use of the visualization, scholars can detect disparities between their understanding of the historical facts and their interpretation of the visualization (Fig. 1 ⓓ). That way, data entry errors can be found and corrected (Fig. 1 ⑤). For instance, we found double entries of cities—"*Cusae*" and "*Qusqam*" for Qusiya in Egypt, and "*Kallinikos*" and "*Qalliniqos*" for Ar-Raqqah in Syria—and merged the instances.

Confidence in a piece of historical evidence also depends on the historian's expertise. This aspect of confidence, which the historians can self-report on for each piece of evidence, is affected by multiple factors: The historian's grasp on the language of the source may be intermittent. Further factors are the ambiguity of entities mentioned in the source, incompleteness due to damage, or aspects of a piece of evidence that are only implicit. Historians can complete the picture through context knowledge and conjecture, which however affects the confidence. We use the term *interpretation confidence* for the historian's confidence in their interpretation. We further aspect the place attribute of a piece of evidence with the *place attribution confidence*, which encodes ambiguity in the assignment of a place to a toponym. Thomson et al. [39] use the categories of *completeness* and *credibility* for this kind of uncertainty. However, they only consider the source's author's expertise, whereas we must consider multiple subsequent situations where interpretation can affect confidence.

The historian's confidence regarding the interpretation of a piece of historical evidence changes over time, as the historian spends more time working with the data. Treating confidence as a first-class attribute of the data—and integrating it accordingly in the visualization—allows historians to review data specifically in their geospatial, temporal or religious context, and to reflect on data with low interpretation confidence. Improved domain knowledge then allows correction of that data (Fig. 1 ⑤) and a boost of the interpretation confidence level.

4 Prototype

Our multiple-coordinated-views (MCV) [30, 41] prototype, which is being developed as part of the project, visualizes different aspects of the data and is shown in Fig. 2. The views are fully interactive and connected via brushing and linking, such that hovering over any part of the visualization will highlight the represented data in other views. All views can be constrained interactively, resulting in a restriction of data items in other views. Filters from multiple views apply orthogonally.

The distribution of evidence tuples on levels of different aspects of confidence is visualized in the *confidence view* (Fig. 2 ⓐ). The religious affiliations of evidence tuples are visualized in the *religious affiliations view* (Fig. 2 ⓑ). Religious denominations and entire subtrees can be disabled, thus constraining the visualized data. The *map view* (Fig. 2 ⓒ) shows the geographical aspect of the data by placing glyphs on a map. The glyphs represent sets of locations, which are grouped to avoid overlaps, and consist of up to four parts representing the four main subtrees of religions shown in Fig. 2 ⓑ. A *timeline view* (Fig. 2 ⓓ) shows the number of evidence tuples for each year, and can be zoomed, panned and constrained. In a *list of locations* (Fig. 2 ⓔ), all locations are listed by their main name. Locations without geographical location—which are missing from the map—are listed separately. *Tooltips* (Fig. 2 ④) show more information about those locations. Evidence without time data is shown in a separate *untimed data view* (Fig. 2 ⓕ).

4.1 Filtering Data by Confidence

We store the six specific confidence attributes—which are discussed in Sect. 3—alongside the other data attributes in the relational database. These six attributes are hierarchically abstracted over two levels. Both the leaves and the parent nodes of the confidence hierarchy can be used to filter by selecting the range of acceptable values (see Fig. 2 ⓐ). Selecting a range on a parent node activates that node and deactivates its children. Subsequently, an evidence tuple is considered a match for the filter if *any* of the leaf confidence levels under that parent aspect matches the parent node's selected range.

The setup allows historians to filter the data by very specific confidence criteria, and on different levels of abstractions, depending on the task at hand. An example with the developed prototype could be the following: A user selects data items of interest by constraining data attributes in the coordinated views interactively to a region, time period, and religious affiliations of interest. Afterwards, only high *total confidence* values are picked to understand which of the results have enough certainty to base a hypothesis on them. As an alternative, only evidence tuples with low *evidence confidence* can be searched for. This helps to plan new data acquisition steps, or the consultation of another expert for the respective data. With the described mechanism, confidence can be seamlessly integrated into various analyses.

4.2 Visualizing Confidence

The confidence hierarchy is visualized as a horizontal tree view (see Fig. 2 ⓐ), where the nodes visualize the distributions of the confidence aspects as normalized, stacked bar charts. For the size of the bars, the number of evidence tuples matching the respective level of confidence for that aspect are counted. For parent nodes, tuples are counted analogously to the filtering described in Sect. 4.1.

In other views, confidence is visualized in *confidence coloring mode*, which can be switched on (see Fig. 2 ⓖ). In *confidence coloring*

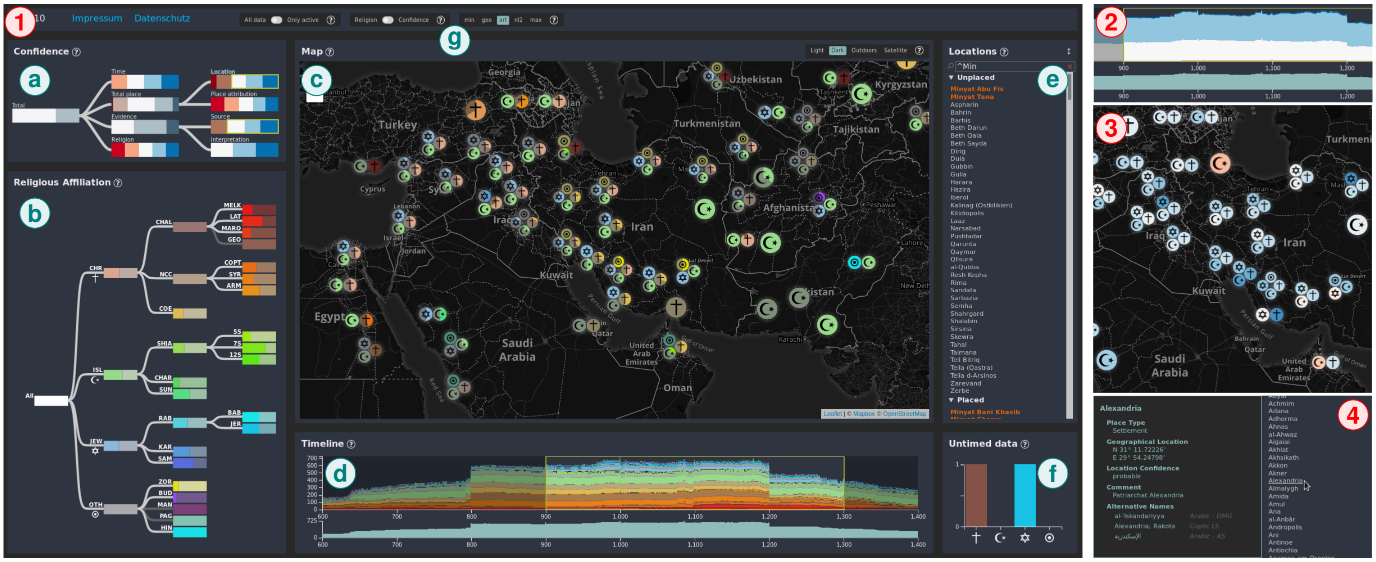

Fig. 2. The ① visualization prototype. Here, some aspects of ⓐ confidence and the ⓓ timeline have been constrained. Aspects of the data are further shown in a ⓑ hierarchy of religious affiliations and a ⓒ map. A ⓔ list of unplaced locations and a ⓕ histogram of untimed evidence tuples make *known unknowns* [35] in the data explicit. A ⓖ set of controls allow switching between religion and confidence coloring mode, as well as between showing or hiding filtered-out data. In confidence coloring mode, visual primitives are colored based on the average confidence of their data, as shown in ② and ③. Hovering over locations from the ② location list shows a ④ tooltip with additional information.

mode, a diverging color scheme encodes the average confidence of the represented data, and color hue no longer encodes religious affiliation, as shown in Fig. 2 ② and ③. The aspect of confidence used for coloring can be selected from the *confidence view* (Fig. 2 ⓐ).

4.3 The Effect of Missing or Hidden Data

Data that is either missing or hidden from viewers of a visualization can affect the overall confidence in the shown data [10, 35]. This is especially true for geospatial data, where viewers wonder whether the empty regions on the map signify sparse population or missing data. Skeels et al. [35] describe three levels of uncertainty awareness: "*unidentified unknowns*", "*unknown knowns*" and "*known knowns*". Without clear communication of missing data, viewers can quickly lose confidence in the visualization itself, which is difficult to reestablish [25].

Incompleteness is a pervasive issue with historical data. As the project particularly investigates the coexistence of religious institutions, the collected data is generally limited to cities with this characteristic. To avoid the impression of unpopulated areas, our prototype has a suitable disclaimer for the *map view*. Future work may include recording and visualizing cities where no evidence has been entered. That would reduce ambiguity between *unidentified unknowns* and *unknown knowns*, and make areas without cities (*known knowns*) explicit.

In an interactive context, temporarily hiding filtered-out data can help reduce visual clutter and allow for showing the still-visible data at a higher level of detail. In our approach, we provide the historians with the choice of whether they want to see the filtered data in a standalone setting (Fig. 2 ② and ③), or within the context of the full data set (Fig. 2 ①), by offering a switch (Fig. 2 ⓖ). In the latter case, filtered-out data is still visualized with reduced saturation and brightness. Consequently, locations on the map whose data are fully filtered out are still indicated, and their symbols do not disappear.

We make low confidence because of missing data in already-entered evidence tuples explicit by showing places with missing geographical locations (see Fig. 2 ②), as well as evidence tuples with missing temporal data (Fig. 2 ③), in separate views. Those views allow for a directed search of entered data in need of further revision.

5 Future Research Directions

One main focus for the continued efforts of the project is data entry with added confidence aspects on the *humanities* side of the cooperation. On the *digital* side, we will improve visual interactive support for data acquisition. This includes integrating the data entry into the visualization, to shorten the corrective feedback interval indicated in Fig. 1 ⑤. The availability of some historical sources in digital format also makes the use of semi-automated text annotation possible. To increase visualization confidence, we would like to add plausibility checks for probably missing data, e.g., by recording locations of larger cities where we do not have evidence for religious communities.

We will also collect more metadata during the data entry phase, to allow for provenance on the data. Such metadata may include revision history of the data. Tracking revision history could allow automatic improvement of *interpretation confidence* for evidence that has been reviewed by multiple historians. Until now, the historians annotated metadata in a free-text comment field, which impedes, in particular, collaborative analysis and reassessment of confidence.

6 Conclusion

While we presented our idea of dealing with explicit confidence information in the context of the described visualization prototype, we believe that making confidence a first-class data attribute, which can be used for filtering, is beneficial for other digital humanities endeavors too, including those employing hermeneutic approaches. We already incorporate possibilities to change uncertainty aspects through the data collection interfaces, but more research effort needs to be spent on how confidence models or values can be revised over time through iterative, collaborative, human visual analysis. The need for revision can arise by considering new sources that either strengthen or reduce confidence, e.g., by contradicting current findings. Confidence in the interpretation of historical sources can further be strengthened by research associates, who review extracted pieces of evidence, and corroborate the interpretation. If reassessment of confidence is part of an analysis cycle, this has consequences for related topics such as recording analysis provenance for traceability and reproducibility.

We believe that making confidence (or uncertainty) an integral part of visual analyses—by seeing it as a first-class data attribute that can be used actively in analysis—can advance not only visual analysis support in digital humanities projects, but many visualization approaches that rely on human assessments and the confidence in these assessments.

Acknowledgments

This work has been developed as part of the "Dhimmis & Muslims" project funded by the VolkswagenStiftung in the "Mixed Methods" program. In addition, it builds upon many discussions and experiences made in the Stuttgart Center for Reflective Text Analytics (CRETA) funded by the German Ministry for Education and Research (BMBF), and the Oceanic Exchanges project, which is funded by the German Science Foundation (DFG) as part of the Digging into Data Challenge.

References

[1] F. Binder, B. Entrup, I. Schiller, and H. Lobin. Uncertain about uncertainty: Different ways of processing fuzziness in digital humanities data. In *Proc. DH*, pages 95–98. EPFL/UNIL, jul 2014.

[2] A. M. Bisantz, S. S. Marsiglio, and J. Munch. Displaying uncertainty: Investigating the effects of display format and specificity. *Human Factors*, 47(4):777–796, dec 2005.

[3] R. P. Botchen, D. Weiskopf, and T. Ertl. Interactive visualization of uncertainty in flow fields using texture-based techniques. In *Proc. ISFV*, pages 4051–4056, 2006.

[4] K. Brodlie, R. A. Osorio, and A. Lopes. A review of uncertainty in data visualization. In *Expanding the frontiers of visual analytics and visualization*, pages 81–109. Springer, 2012.

[5] B. P. Buttenfield. Visualizing cartographic metadata. In B. P. Buttenfield and S. B. Clapham, editors, *Visualization of the Quality of Spatial Data: Report on the Specialists Meeting For NCGIA Research Initiative 7*, pages c17–c26, Santa Barbara, 1991. National Center for Geographic Information and Analysis.

[6] C. Collins, S. Carpendale, and G. Penn. Visualization of uncertainty in lattices to support decision-making. In *Proc. EuroVis*, pages 51–58. Eurographics Association, 2007.

[7] C. D. Correa, Y.-H. Chan, and K.-L. Ma. A framework for uncertainty-aware visual analytics. In *Proc. VAST*, pages 51–58. IEEE, 2009.

[8] M. Correll, D. Moritz, and J. Heer. Value-suppressing uncertainty palettes. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, page 642. ACM, 2018.

[9] N. de Lange, A. Panayotov, and G. Rees. Mapping the Jewish communities of the Byzantine Empire. `http://www.byzantinejewry.net/`, 2013.

[10] C. Eaton, C. Plaisant, and T. Drizd. The challenge of missing and uncertain data. In *Proc. VIS*, pages 100–101. IEEE, 2003.

[11] B. J. Evans. Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers & Geosciences*, 23(4):409–422, 1997.

[12] P. F. Fisher. Models of uncertainty in spatial data. *Geographical Information Systems*, 1:191–205, 1999.

[13] H. Griethe and H. Schumann. The visualization of uncertain data: Methods and problems. In *Proc. SimVis*, pages 143–156. SCS Publishing House e.V., jan 2006.

[14] J. Görtler, C. Schulz, D. Weiskopf, and O. Deussen. Bubble treemaps for uncertainty visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):719–728, 2017.

[15] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, 2018.

[16] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. van Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, 2011.

[17] C. Kinkeldey, A. M. MacEachren, M. Riveiro, and J. Schiewe. Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, 44(1):1–21, 2017.

[18] C. Kinkeldey, A. M. MacEachren, and J. Schiewe. How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51:372–386, 2014.

[19] J. Liem, E. Goudarouli, S. Hirschorn, J. Wood, and C. Perin. Conveying uncertainty in archived war diaries with GeoBlobs. In *Proc. VIS*. IEEE, oct 2018. Poster.

[20] A. M. MacEachren. Visualizing uncertain information. *Cartographic Perspectives*, 13:10–19, 1992.

[21] A. M. MacEachren, C. A. Brewer, and L. W. Pickle. Visualizing georeferenced data: representing reliability of health statistics. *Environment and Planning A*, 30(9):1547–1561, 1998.

[22] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32:139–160, 2005.

[23] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, 2012.

[24] N. McCurdy, J. Gerdes, and M. Meyer. A framework for externalizing implicit error using visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):925–935, 2018.

[25] B. M. Muir. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5):527–539, 1987.

[26] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, nov 1997.

[27] J. Paradis and K. Beard. Visualization of spatial data quality for the decision-maker: A data-quality filter. *Journal of the Urban and Regional Information Systems Association*, 6(2):25–34, 1994.

[28] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. International Conference on Intelligence Analysis*, volume 5, pages 2–7, McLean, VA, USA, may 2005.

[29] G. Rees, N. de Lange, and A. Panayotov. Mapping the Jewish communities of the Byzantine Empire using GIS. *Migration and Migrant Identities in the Near East from Antiquity to the Middle Ages*, pages 104–121, 2018.

[30] J. C. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Proc. CMV*, pages 61–71. IEEE, 2007.

[31] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, 2015.

[32] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014.

[33] R. Scholz and Y. Lu. Uncertainty in geographic data on bivariate maps: An examination of visualization preference and decision making. *ISPRS International Journal of Geo-Information*, 3(4):1180–1197, 2014.

[34] K. Seipp, X. Ochoa, F. Gutiérrez, and K. Verbert. A research agenda for managing uncertainty in visual analytics. *Mensch und Computer 2016—Workshopband*, 2016.

[35] M. Skeels, B. Lee, G. Smith, and G. G. Robertson. Revealing uncertainty for information visualization. *Information Visualization*, 9(1):70–81, 2010.

[36] A. Slingsby, J. Dykes, and J. Wood. Exploring uncertainty in geodemographics with interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 17:2545–2554, 2011.

[37] H. Song and D. A. Szafir. Where's my data? Evaluating visualizations with missing data. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):914–924, 2018.

[38] R. Therón, A. G. Losada, A. Benito, and R. Santamaría. Toward supporting decision-making under uncertainty in digital humanities with progressive visualization. In *Proc. TEEM*, pages 826–832. ACM, 2018.

[39] J. Thomson, E. Hetzler, A. M. MacEachren, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. In *Visualization and Data Analysis*, volume 5669, pages 146–158. International Society for Optics and Photonics, mar 2005.

[40] R. Twiddy, J. Cavallo, and S. M. Shiri. Restorer: A visualization technique for handling missing data. In *Proc. VIS*, pages 212–216. IEEE, oct 1994.

[41] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proc. Working Conference on Advanced Visual Interfaces*, pages 110–119. ACM, 2000.

[42] F. Windhager, V. A. Filipov, S. Salisu, and E. Mayr. Visualizing uncertainty in cultural heritage collections. In *Proc. EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization*, pages 7–11. Eurographics Association, 2018.

[43] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18:2749–2758, 2012.

[44] D. J. Wrisley and S. Jänicke. Visualizing uncertainty: How to use the fuzzy data of 550 medieval texts? In *Proc. DH*. Center for Digital Research in the Humanities, jul 2013.