
Structured and Uncertainty-Aware Data Storytelling

Yuchen Zhao

Yuchen.Zhao@soton.ac.uk
University of Southampton
Southampton, UK

Laura Koesten

laura.koesten@theodi.org
The Open Data Institute
University of Southampton
London + Southampton, UK

Tom Blount

T.Blount@soton.ac.uk
University of Southampton
Southampton, UK

Elena Simperl

E.Simperl@soton.ac.uk
University of Southampton
Southampton, UK

ABSTRACT

Data stories are an increasingly popular medium for delivering information in the field of journalism. More and more journalists are telling stories directly from data, by creating and combining data visualisations. A good data story, as with any other types of news story, should convey the author's message accurately, and without distorting the underlying data. In this paper, we introduce our ongoing project, which aims to help data journalists create accurate stories by providing a logical story structure and informing authors of data-uncertainty through the whole life cycle of story creation. We identify key challenges in this space, and present our design considerations for addressing them.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '19 Workshop on HCI for Accurate, Impartial and Transparent Journalism: Challenges and Solutions, May 05, 2019, Glasgow, United Kingdom

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

CCS CONCEPTS

- **Human-centered computing** → **Interactive systems and tools.**

KEYWORDS

human-data interaction, data journalism, data visualisation

INTRODUCTION

Nowadays, more and more data has become openly available for people to extract facts, identify insights, and tell stories from it. Data stories are made up of more than simply numbers from a spreadsheet; they can reference huge webs of interconnected data about individuals and organisations, and contain interactive visualisations that drive the story towards the author's conclusion [2].

Creating data stories requires a diversity of skills such as domain knowledge about topics, story telling skills, data analysis as well as visualisation skills. However, domain experts may or may not be good story tellers. Journalists who are primarily professional story tellers may find it difficult to produce high quality data visualisations. Therefore, it is necessary to provide tools that have the potential to assist journalists in creating data stories.

The Data Stories¹ project aims to explore how people engage with data, and provide solutions to some of the challenges in data journalism. One of our goals is to develop a system to support data journalists in creating accurate and engaging stories. Our system will enable authors to create data stories in a logical structure, which aims to bridge the gaps among data, arguments, and facts. In addition, it will inform authors of underlying uncertainty in their data through recommendations and visualisations, in different stages of the story creation process. We believe that our proposed system would be of interest to authors who lack data analysis and visualisation skills.

DATA STORYTELLING SYSTEMS

Many systems have been proposed to help journalists create data stories more easily. The Contextifier system proposed by Hullman et al. [5] automatically generates annotations for line charts from stock news. The annotated visualisations created by the system are perceived to have better relevance and salience than those made by baseline methods such as randomly generated annotations. Similarly, Gao et al. [1] have proposed NewsViews, a system that generates annotated maps for news. The generated maps are also perceived to be interesting and relevant to the base story in a user study of online workers and university graduate students. These systems can assist data journalists to make specific types of annotated visualisations (i.e., line charts and maps). Some other systems that enable journalists to create data stories in a more general format have also been proposed. For instance, Satyanarayan and Heer [7] designed and prototyped the Ellipsis system, which helps journalists create

¹<http://datastories.co.uk>

data stories that have multiple sections. Hu et al. [4] have proposed the DIVE system that supports the whole process in data storytelling, including data exploration, visualisation generation, and story creation. Such storytelling tools are perceived helpful and valuable by professional journalists [7].

Although these existing systems can help storytelling in many aspects, they lack two useful supportive functions. One is to provide a story template that helps journalists to organise the content with a focus on the underlying logical structure of their arguments. Another is to inform journalists about the uncertainty inherent to data and data visualisations. We argue that functionality that supports a logical structure template, and awareness and understanding of uncertainty attached to the arguments based on data would help journalists report news more accurately. Therefore our project aims to implement a system that supports both structured and uncertainty-aware data storytelling.

CHALLENGES AND CONSIDERATIONS

We now discuss the challenges in structured and uncertainty-aware data storytelling, and how our system aims to address these.

Structures of Data Stories

To support structured storytelling, the first issue is to understand the aim of the author and provide a suitable corresponding structure. Although the aim may vary for different genres [8], the goal of many data stories is to convey an argument, and provide evidence with data to support it. Therefore, we aim to design our system to support this common goal.

Kosara [6] has analysed a number of data stories that convey arguments and has summarised a *Claim-Fact-Conclusion* (CFO) structure from these stories. The first part of this sequential structure is a *Claim* section, which contains a story's argument. To support the argument, a number of *Fact* sections will follow. Each of the Fact sections contains some evidence (e.g., visualisations) generated from data and partially contributes to the proof of the argument. The structure ends with a *Conclusion* section that summarises how the facts support the argument.

Our system will use this logical structure to guide authors in the process of storytelling. Based on the authors' prior decisions, and additional context inferred from the data, our system will help the authors decide if they should present a claim, seek evidence from data, or draw a conclusion, and provide appropriate supporting visualisations. We believe that the support of such a logical structure will make the created stories convincing and perceived as reasonable.

Uncertainty in Data Storytelling

In the process of data storytelling, different types of uncertainty exist in different stages. For instance, some uncertainty exists in the data itself, due to sample size or distribution amongst other factors. There may also be uncertainty in how this is expressed in visualisations, due to the analytic results

behind them (e.g., significance level, standard error). To improve the accuracy of a data story, the uncertainty in the whole life cycle of creating the story needs to be considered and the author of the story should be aware of such uncertainty.

Greis et al. [3] find that presenting uncertainty through visualisations can help people better estimate the value of discrepant information; therefore it is important to communicate uncertainty in an effective and intuitive way. Authors of data stories have to deal with similar tasks, such as deciding whether a data field is good enough, or if the result shown in a visualisation is convincing. Thus visualising uncertainty information such as the distribution of values and confidence intervals would be beneficial for authors.

In the design of the system, we consider recommending visualisations based on the uncertainty of the data used, and the analytical results drawn from them. We also apply uncertainty visualisations to our system in order to support authors' decision making in different stages (e.g., data exploration, visualisation selection). Such considerations could help authors choose data and visualisations with higher quality, which may lead to more accurate and significant conclusions.

CONCLUSIONS

As an increasingly popular medium of news, data stories can be entertaining, engaging, and informative. However, they need to present the information hidden in data accurately and, to some extent, convey the assumptions attached to any kind of data analysis. In this paper, we identified two challenges for existing tools to help authors create accurate data stories: providing a consistent and logical structure, and communicating uncertainty. We propose a system to address these issues, through adopting a well established story structure, uncertainty-aware recommendations, and uncertainty visualisations.

As an on-going part of this project, we plan to get feedback from professional journalists about the design and conduct user studies to evaluate the system to understand how helpful it is. We will iterate the design and functions in a co-design process based on the feedback and the results of evaluation.

ACKNOWLEDGMENTS

This work is funded by EPSRC Data Stories project (EP/P025676/1) and EU H2020 TheyBuyForYou project (780247).

REFERENCES

- [1] Tong Gao, Jessica Hullman, Eytan Adar, Brent Hecht, and Nicholas Diakopoulos. 2014. NewsViews: An Automated Pipeline for Creating Custom Geovisualizations for News. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3005–3014. <https://doi.org/10.1145/2556288.2557228>
- [2] Jonathan Gray, Lucy Chambers, and Liliana Bounegru. 2012. *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. " O'Reilly Media, Inc."

- [3] Miriam Greis, Aditi Joshi, Ken Singer, Albrecht Schmidt, and Tonja Machulla. 2018. Uncertainty Visualization Influences How Humans Aggregate Discrepant Information. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 505:1–505:12. <https://doi.org/10.1145/3173574.3174079>
- [4] Kevin Hu, Diana Orghian, and César Hidalgo. 2018. DIVE: A Mixed-Initiative System Supporting Integrated Data Exploration Workflows. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics (HILDA '18)*. ACM, New York, NY, USA, 5:1–5:7. <https://doi.org/10.1145/3209900.3209910>
- [5] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. 2013. Contextifier: Automatic Generation of Annotated Stock Visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2707–2716. <https://doi.org/10.1145/2470654.2481374>
- [6] Robert Kosara. 2017. An Argument Structure for Data Stories. In *Short Paper Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization (EuroVis '17)*. <https://research.tableau.com/sites/default/files/Kosara-EuroVis-2017.pdf>
- [7] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring Narrative Visualizations with Ellipsis. *Computer Graphics Forum* 33, 3 (Jun 2014), 361–370. <https://doi.org/10.1111/cgf.12392>
- [8] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (Nov 2010), 1139–1148. <https://doi.org/10.1109/TVCG.2010.179>