

You guessed it! Reflecting on preconceptions and exploring data without statistics

Pablo León-Villagrà
University of Edinburgh
pablo.leon@ed.ac.uk

Sarwar Islam
University of Leicester
si113@le.ac.uk

Megan Lucero
The Bureau of
Investigative Journalism
meganlucero@tbij.com

Brooks Paige
Alan Turing Institute
bpaige@turing.ac.uk

Tomas Petricek
Alan Turing Institute
tomas@tomasp.net

Abstract: We live in times in which information is abundant, but trust in expert analysis is low. How can we make complex issues accessible for readers and overcome their preconceptions? We propose a novel way of presenting readers with data and raising awareness for individual bias and preconceptions. In our application, readers freely choose potentially relevant factors in societally relevant issues and reflect on their choices in an engaging, non-threatening way. We suggest that such an approach allows for more engaged and open-minded readers and as a result can facilitate democratic, data-centric debate.

Keywords: data-driven storytelling, open data, interactivity, transparency and trust, bias

Introduction

In times of partisan debates and filter bubbles shaping our news, trust in traditional journalism seems to be declining. At the same time, the amount of information and publicly available data is larger than ever before, in principle allowing anybody to gain insight and participate in an informed, democratic debate. Here we explore how journalism can build trust and make data more approachable by implementing data-driven environments for exploration and reflection.

Data-driven documents

Articles often try to simplify complex issues and in an attempt to make the argument approachable present only aggregate statistics to support their argument. As a result, these articles do not allow the readers to check the sources or assumptions made during analysis, which can lead to distrust. Acknowledging these shortcomings, there has been a surge of data-driven reporting tools (see for example Petricek, 2018). These tools can make articles more transparent and open, and encourage readers to scrutinize the analysis and explore further. The reader then can transform from a passive recipient into an active and informed participant in a debate. In this work, we follow the example of data-driven, transparent storytelling but focus more crucially on allowing the reader to engage and reflect on her own beliefs and bias.

Environments for engagement and reflection

One element of a good explanation is its predictive quality. For example, say we are interested in regional differences in the quality of healthcare and potential sources and factors for better care. A thorough analysis of the matter will require us to examine measures of good care, for instance the time it takes to access care. In general, we will be interested in determining relevant factors that, say, successful regions share, as these are crucial for the evaluation of policy decisions. To find factors we might use domain knowledge to search data sources, potentially transforming or discarding inconclusive variables to derive our piece. In many situations the reader will not be aware of the steps preceding the analysis, nor the reasons for the transformations and as a result could dismiss or distrust our work. Additionally, readers rarely encounter an argument without personal views about what *the right* predictors should be or what the author intends with her piece.

Research has highlighted that correcting bias and engaging readers requires more than just a clear presentation of facts. In fact, corrections can sometimes increase misconceptions (Nyhan & Reifler, 2010). In general, corrections have a greater chance of being effective if they are communicated in non-challenging and appealing way (for an overview see Lewandowsky, Ecker & Cook, 2017). These results can not simply be explained on the basis of a lack of information literacy, as even numerically-savvy readers might choose to maintain their erroneous beliefs to protect their political self identity (Kahan et al., 2017). Given that corrections are difficult and mere presentation of facts can be unconvincing - how do we create an environment that encourages reflection and critical engagement? We propose that to encourage engagement and reflection it is crucial to allow the reader to freely voice and explore their preconceptions. Our application attempts to be a non-

challenging, open, exploration of the issue and the readers assumptions. In what follows we present a first prototype of such an environment.

Application

We chose NHS waiting times (for cancer patients) as our example topic and collected publicly available datasets². Our focus was acquiring a breadth of variables that might be relevant for differences in waiting times instead of finding few, statistically relevant factors. The selection of variables has to allow readers to make their own choices without creating the impression that the document presupposes a particular interpretation. The application consists of three stages: selecting factors, repeatedly predicting a target variable and an overview of one’s accuracy.

Selecting factors: After framing the issue, readers are presented with the dataset in form of (potentially predictive) variables. Readers are prompted to select variables that they deem important (see Figure 1). Since absolute numbers will not be informative for non-expert readers, all numerical quantities are expressed verbally (in terms of quantiles, e.g “slightly lower than average”).

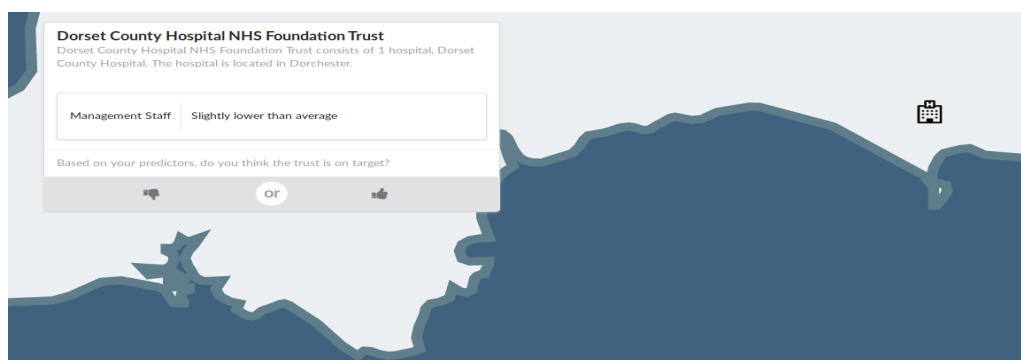


Figure 1. Readers try to predict if the NHS trusts fulfil the waiting time targets. They are given the trust location and general information, as well as the variables they selected. Here the reader chose Managerial Staff as a relevant predictor.

Predicting the target: Readers are shown the variables selected in the previous section for several instances of the target outcome and are asked to predict the target (in our example they had to judge if a particular NHS trust achieved the target for cancer waiting times). For an example, see Figure 2.

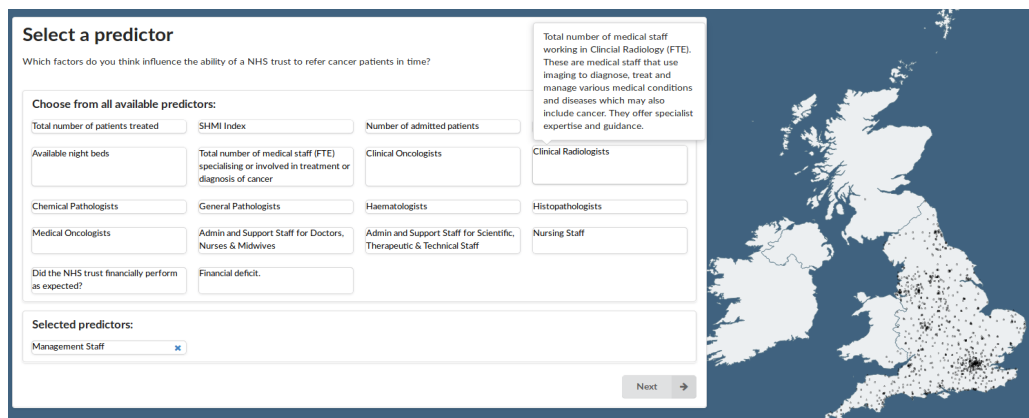


Figure 2. This article explores the ability of NHS trusts to fulfil waiting time targets for cancer patients. The variables the reader could select contained characteristics of NHS trusts (number of patients treated, staff numbers for different staff, day and night beds, financial indicators etc.).

Evaluation: After performing a number of predictions, readers are confronted with their predictive accuracy. This evaluation informs the reader about the predictive quality of their preconceptions. In

² Waiting times data: <https://www.england.nhs.uk/statistics/statistical-work-areas/cancer-waiting-times/>. Additional factors used: <https://data.gov.uk/dataset/nhs-workforce-medical-staff> and <https://digital.nhs.uk/catalogue/PUB30092>.

the future we plan to offer additional feedback at this stage - for example hints recommending features with stronger predictive accuracy or a comparison to other readers' performance. This evaluation and feedback can also be used as introduction to a traditional journalistic piece.

Conclusions

In this work we developed a prototype that allows readers express their preconceptions and reflect on their predictive accuracy. Our application is designed to be neutral - it does not attempt to instil an expert opinion, but instead tries to raise awareness for preconceptions and complex relationships. We suggest that reflection on preconceptions and awareness of complexity are fundamental for more engaged and open-minded readers and can facilitate more democratic debate. In the future we would like to test the efficiency of our application and explore further ways in which preconceptions and reader engagement can be incorporated into reporting.

References

- Lewandowsky, S., Ecker, U.K. and Cook, J., 2017. Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition*, 6(4), pp.353-369.
- Kahan, D.M., Peters, E., Dawson, E.C. and Slovic, P., 2017. Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), pp.54-86.
- Nyhan, B. and Reifler, J., 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), pp.303-330.
- Petricek, Tomas. "Tools for Open, Transparent and Engaging Storytelling." *Proceedings of the International Conference on the Art, Science, and Engineering of Programming - Programming '17*, 2017