Communication is Design is Analysis is Communication

Jo Wood *giCentre, City, University of London*

Provocation Theme 3: "from analysis to communication"

It is true that visualizations (noun) are limited as evidence. They are indeed the result of many subjective design choices and vulnerable to subjective perceptual and cognitive interpretation.

I want to argue that this is not unique to visualization, nor is it necessarily undesirable. I will suggest in fact it is a virtue of the data analysis process of the kinds considered this week that this subjectivity exists.

How do we resolve the tension between wanting to standardise our analysis and communication processes and providing the contextual flexibility necessary to address complex and nuanced data-driven science? I will argue that key to that resolution is to reduce the friction involved in documenting and communicating our design and analysis choices.

All Data Science is Subjective

When we proposed *literate visualization* (Wood *et al* 2019) as an approach to visualization design, we confronted the challenge of making the often informal and hidden parts of the visualization design process explicit. We were motivated by the desire to improve how visualization design can build upon prior work and experience. We did this in part by providing a mechanism to build *schemas* to structure *design exposition*. This recognised the value in providing rich description of a design process that can be subjective, opinionated and yet still informed by theory and empirical best practice. It provides one means of injecting rigor into the design process and its communication (Meyer and Dykes, 2020). And importantly it recognised that 'communication' and 'design' are part of the same overall process of codifying understanding.

While literate visualization focussed on the visual design process, we found the same approach was useful in the data assembly and shaping process (*literate data*) and in the analytics performed numerically and visually (*literate analysis*). It recognises that data shaping and analysis are also the result of choices in a wide space of possibilities, choices that are to a large extent subjective. The 'literate' approach (after Knuth, 1984) is characterised by deep integration of the construction process (assembling data, implementing a visualization, performing numeric analysis) with exposition justifying and explaining the process. Thus, the acts of *doing*, of *reasoning* and of *communicating* become unified.

Reducing Expository Friction

Of course, reasoning about one's design and analysis process is not new. It sits at the heart of much academic work in visualization and allied fields. But the cost of communicating and justifying a piece of data science through the academic publishing model is high and its rewards limited to those within the academic system. A key challenge in widening uptake in more structured data science exposition is to reduce the friction involved in explicitly reasoning and justifying. New structured schemas would help, as does infrastructure that supports integrated production and exposition (computational notebooks being an obvious example). But a key challenge remains: how do we develop the 'literate' process to ease the structuring and communication of our analytical design endeavours?

References

Knuth, D. (1984) Literate programming. *The Computer Journal*, 27(2), 97–111 doi:10.1093/comjnl/27.2.97

Meyer, M and Dykes, J. (2020) Criteria for rigor in visualization design study. *IEEE transactions on visualization and computer graphics* 26, no. 1 (2019), 87-97. https://ieeexplore.ieee.org/document/8809711

Wood, J., Kachkaev, A., and Dykes, J. (2019). Design Exposition with Literate Visualization. *IEEE Transactions on Visualization and Computer Graphics,* 25(1), 759–768. https://openaccess.city.ac.uk/20081