Well … it depends — on the notion of context in data-intensive practices

“Well… it depends…” — this is a statement that we often utter or hear when working with data and models to reason about a phenomena. Data-driven practices involve numerous decisions around parameters, techniques and models, i.e., garden of forking paths (Gelman and Loken 2013), but also involve many other contextual factors such as the capabilities and the “defaults” of the analysis platform or the tacit knowledge, experience and the biases of the analyst/scientist. And then there is the contextual information about the phenomena being studied/modelled and how well the data is representing and framing the nuances of the world. Inferences, forecasts and decision made based on models depend heavily on such contextual factors and how well they are understood and considered. Any observation, evidence or claims informed by the data and the produced data artefacts are then “situated” (Tkacz 2021) in such multi-faceted context that is in and out of the data. Does this dependency then render all these outcomes invalid and spurious? I would like to say no and argue instead that embracing the situatedness and meaningfully acting on it is an opportunity for establishing a “better” data science practice.

One recent project where I have experienced how context matters in data-informed decisions is our work on user modelling for cyber security (Nguyen et al 2019). In that project, we worked with cyber security analysts on how they flag a series of actions done by a user on a digital system as being “suspicious”. The analysts had a general behaviour model (a hidden-markov-model) that can label activities as suspicious by assigning them a score. This is akin to identifying outliers in a dataset. But given the inherent complexities of what drives people’s behaviour on the system and without knowing the context for an individual, these scores were meaningless. The normal varied from people to people and from team to team, and sometimes day to day, without understanding the roles in the organisation, the different kinds of everyday tasks and routines, putting these data models to work was impossible. And this was what we worked on in that project by looking for ways to better unravel the context for a particular data result. Our approach was one that put *human in the loop*, giving them affordances to contextualise what they observe by providing multiple models that characterise behaviour at different levels and relating them through visualisation.

Trying to surface and understand the various factors that contribute to the context and how the patterns are situated within them can only make our analyses richer. Only through attempting to understand the “situations” and piecing together the contextual learnings in which things operate or not we can unravel the complex phenomena that we study through data. But are these also the factors that can make the practice less rigorous, less replicable? How to then navigate this trade-off or is there a trade-off? How to deliver what Tracey and Hinrichs call “rich rigour” (Tracy, S.J. and Hinrichs, M.M., 2017)? I see a few ways that I am keen to explore further:

- Building on Leo Breiman’s Rashomon’s Effect to advance our methods to reconcile multiple models - bringing together models that operate at different levels, scales, parameters, those that doesn’t necessarily share a common space but tell partial stories about the world and make this a part of the exploratory data analysis practice
- Learning from understandings of rigour within both the qualitative and quantitative methods literature to establish a refreshed data science practice that is critical, reflective and reflexive
- And rethinking visualisation’s role and contribution in facilitating this envisioned practice

But these all come with unique challenges and will they get us where we want? Well…it depends…

REFERENCES

Gelman, A. and Loken, E., 2013. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*, *348*(1-17), p.3.

Tkacz, N., Henrique da Mata Martins, M., Porto de Albuquerque, J., Horita, F. and Dolif Neto, G., 2021. Data diaries: A situated approach to the study of data. *Big Data & Society*, *8*(1), p.2053951721996036.

Nguyen, P.H., Henkin, R., Chen, S., Andrienko, N., Andrienko, G., Thonnard, O. and Turkay, C., 2019. Vasabi: Hierarchical user profiles for interactive visual user behaviour analytics. *IEEE transactions on visualization and computer graphics*, *26*(1), pp.77-86.

Breiman, L., 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, *16*(3), pp.199-231.

Tracy, S.J. and Hinrichs, M.M., 2017. Big tent criteria for qualitative quality. *The international encyclopedia of communication research methods*, pp.1-10.