500-words Monday position statement
Maria-Cruz Villa-Uriol

In my research, I fully position myself in the exploratory data analysis end of the spectrum. I look at *relatively* large numbers of sequences of events that are data-rich. Sequences that are multivariate in the sense that have many attributes that might change over time. I am interested in deriving and representing meaningful insights in an easy manner without misleading analysts. To do this I combine visualisation and machine learning methods. I am always in conflict, trying to balance providing valuable insights while avoiding misleading analysts. Making it easy for analysts to derive valuable findings while ensuring that those are actually relevant. I keep going back to statisticians to determine if there are ways where statistics can provide reassurance about how solid my visualisations are.

Personally, in the problems I have worked on, exploratory data analysis always comes first. It represents the starting point of my work, and I see it as the necessary stepping stone before moving into the confirmatory space. Before moving on to deriving findings, I need first to understand the problem, what kind of questions are relevant, and the nature of the data.

The provocations and questions motivating this challenge are exciting, and I find myself constantly thinking about some of them:

- How to keep the context behind an analytic decision without making everything too complicated?
  I find that using multiple levels of detail using coordinated views is a must. Where I struggle is in keeping visualisation easy. Depending on the question you might want to visualise something as an absolute value, other times as a percentage, and on other occasions with respect to another value.
- In my work, there is always a human-in-the-loop involved, a data analyst. This analyst drives the analysis. The analyst starts with questions, but as the data exploration moves forward, the analyst keeps exploring the space not necessarily in an organised and exhaustive manner. How do we make sure that the findings are not just anecdotal?
  To solve this problem, I like to use machine learning techniques able to analyse the data in multiple manners and exhaustively. With those results, I provide the analyst with multiple scenarios for further analysis.
- How to embed in visualisations the uncertainty present in data? There are multiple sources of uncertainty. How to represent them accurately to avoid misleading analysts is a constant threat to the visualisation world. Visualizations are considered as ‘limited’ evidence because our senses can be easily manipulated.
- Statistics has been hand in hand with visualisations, how can we bring these two worlds together to increase confidence in visualisations?