Koen Venema

Data-visualization and Storytelling with Relevance to Microbiota Research - the Application of AI

Introduction

The presentation is inspired by the "1 dataset, 100 visualizations" project, where a dataset on World Heritage Sites of only 3 columns (3 countries) and 2 rows (2 years) are plotted in (at least) 100 different, (arguably) meaningful ways. If data visualization (or storytelling) in many different ways is already possible with such a small dataset, imagine the possibilities of a large dataset on a probiotic or prebiotic trial, in which many different parameters are measured. In addition, the use of machine learning (a.k.a. artificial intelligence) to predict the outcome of a clinical intervention, and to deduce which parameters are explaining the outcome of such a study, are becoming increasingly important. Therefore, data visualization (storytelling) is a crucial addendum to a researcher's bag of expertise.

Methods

The presentation will use data visualization and machine learning tools to provide examples on how to use these to make sense of multidimensional data. It will both be a high abstract overview, but at the same time providing some state-of-the-art tools and scripts that can be used by others to get insight in their data. It will address 'dashboard creation' platforms (open source and commercial), ways of model deployment (*e.g.*, in a Docker-container, or cloud-based), and interactive graphs. Open source tools have been implemented as much as possible in R. Commercial platforms include Power BI (Microsoft) and Tableau (which also has a free version, but with limited functionality).

Results

An excerpt of the "1 dataset, 100 visualizations" project will be provided to exemplify the power of data visualization.
Next, two *in vitro* examples of existing datasets will be presented. One dataset used the TNO *in vitro* model of the colon (TIM-2), regarding different response of the microbiota when fed different (doses of) (potential) prebiotics, and one dataset from a Colon-on-a-plate® project from Prodigest, regarding the effect of a mix of probiotics on L-Dopa production by the microbiota of healthy individuals versus production by a dysbiotic microbiota. Furthermore, the results of a placebo-controlled *in vivo* intervention study on stunted children and normal age- and gender-matched controls, performed in collaboration with Binus University in Jakarta, Indonesia will be presented. The interventions were a probiotic, and the same heat-killed probiotic (postbiotic). The power of machine learning (artificial intelligence) and data visualization in combination with interactive graphs, and due to that storytelling, will become clear.

Discussion

Particularly since journals nowadays are almost exclusively published online, (as an editor-in-chief myself) I predict that the possibility of inserting interactive graphs in publications will take flight soon. It will allow researchers to focus/zoom in on those aspects of the data that ***they*** are particularly interested in, which doesn't have to be the same as the story that ***the authors*** want to tell.
In conclusion, data science, data engineering and data visualization are crucial to make sense of high dimensional (microbiota) data.

References

- https://100.datavizproject.com/