

Theme 4: Visualization and Dissemination

Sheelagh Carpendale,
Renee Miller,
Fanny Chevalier,
Christopher Collins

Break-out Session Notes

Participants:

- Sheelagh Carpendale, Chris Collins, Fanny Chevalier, Dimitri Litvin, Travis Windling, Renee Miller, Glenn Paulley, Leon Punambolam, Wayne Oldford, Nadine Kerrigan, Kelly Lyons, Pubudu Premewardena.

Questions to discuss

- What are the fundamental R&D problems in each of the areas in the core?
- What are the fundamental R&D problems that need to be tackled at the intersection of each area with every *other* area?

Theme 4

Dissemination and Visualization

- An essential component of data science is ensuring that both data and results are readily accessible to all interested or affected parties.

Methods

(Visualization & Dissemination)

- How to know that I am using the right analytical methods?
 - Can the system provide recommendation on methods to be more appropriate for my data/questions?
- How to know that I am using the right (best available) data?
 - Can we recommend better data?

Methods (cont'd)

- How to combine traditional visualization methods with other computational techniques (e.g., ML, search, AI, ...)

Methods (cont'd)

- To what extent are we, as a community thinking about what an ideal (programming) language for visualization would be? There are a zillion programming languages; for example in data management we're stuck with SQL.
 - Can we extend SQL to data discovery?

Methodology

- Research methods as a research question: when and how do you synthesize the process from many case studies
- How do we enable research outcomes to be applied in practice?
- How to deal with data that has more than 3 dimensions?

Visualization Research

- How to leverage new technologies (VR / AR) - immersive analytics
(see <https://www.immersiveanalytics.com/>)
- Overcoming challenges of resolution, interaction with different form factors (large screens, 3D interaction, VR / AR)
- Collaborative, engage multiple people, how to discuss about data together (co-located and distributed)

Visualization Research (cont'd)

- How to cope with different levels of data and visualization literacy
- How to use visualization as a mediator/facilitator across different expertise and disciplines when discussing data

Visualization Research (cont'd)

Apply Visualization throughout the process

- Visualization can be applied at any steps of the analytical / data processing cycle to help people understand raw data, how it's been transformed, what models have been applied, etc.... What is the right way to present these data and processes?

Dissemination Research

- How to share data with sufficient/appropriate metadata for reuse.
- How to cope with data shared at different levels of aggregation
- How to manage data versions (through visualization or other techniques)
- How to ensure authoring/provenance (watermarking, etc)

Design and Authoring

- How much do you show to your audience: sometimes you want to be efficient and show the final result (details can be distracting/confusing) vs. sometimes you want to have access to all details within the black box.
- More flexibility in tools to produce data visualizations. Expanding the number of available templates to increase capacity building of visualization (in business setups)

Design and Authoring (cont'd)

- Identify when a (drag-and-drop) template is not sufficient anymore? How to enable more flexibility? What does it look like?
- Understand when and how to open the black boxes: need to understand people's needs first.

Example R&D problems

Example 1: apply visualization to search

- Data search (google for data): How to search datasets (facilitate creating queries, seeing what is available, etc.) Then, how to present the results of a search of *datasets* in a *data lake* using visualization?

Example R&D problems (cont'd)

Example 2: user created data stories

- How to empower actual consumers of the visualization to pull on their own stories of the data, annotations, and insights that matter to them? Give the ability to the user to author their own sequence of views, together with annotations, to make a story.

Example R&D problems (cont'd)

Example 3: Corporate challenge:

- Providing tools for “self service” within companies - not all people involved in data analysis have strong technical aptitude, and in particular higher up exec teams need to see a “data story” which they can relate to. Make it clear how much work was involved to come to conclusions (provenance story).

Example R&D problems (cont'd)

Example 4: Convey the sophistication of the analysis

- When presenting a visualization, how to convey it's simple vs. complex, how to convey this is important vs. a cherry-picked minor result. Reveal the level of importance of any conclusions. Visualizations which are too simple can give the impression that the work was easy, or shouldn't be trusted.

Connections to other themes

Theme 1 - Trust and Usability

- Trust in data, in what ways can visualization help build appropriate trust in the data as well as appropriate trust in the visualization itself and trust in the models (sensitivity analysis)
- Helping customers trust that the data they provide will actually be for their benefit (trust) by informing customers of changes they can make to improve their situation (e.g. improve their insurance risk profile).
- Selecting trusted data for dissemination

Connections to other themes

Theme 2 - Management of Big Data

- Show the flow of data, some coming from legacy systems. Who is going to be affected by changes in the data flow pipeline.
- Know about downstream implications when you change the way data is captured.
- Pushing common statistical/viz functions into DBMS for improved performance and scalability.
- Most systems are limited in terms of built-in statistical analysis methods. How to know what analytical methods are available to pick from? How to know what's missing in systems that allows to perform data analytics? [not a usability question, but performance question]

Connections to other themes

Theme 3 - Modelling and Analysis

- Model diagnostics; pre-compute features to detect what might be interesting and present those.
- Where in the pipeline do you put the computation?
- Confidence into the presented results: reveal quality of data / uncertainty, reveal confidence in a model / ML process result / speculation or prediction

Connections to other themes

Theme 5 - Security and Privacy

- Differential privacy and data protection within a visualization
- Some of the PII data
- Anonymization without destroying the utility; progressive disclosure; compare your results with everyone else's without actually being able to see the data of others
- Privacy aware dissemination

Connections to other themes

Theme 6 – Ethics, Policy and Social Impact

- Visualizations used to persuade in problematic ways
- Visualization used to tell stories which may mislead; data propaganda
- Missing or uncertain data which is used in a visualization to make a critical decision
- Generate ethical do's and do not's about visualization and dissemination
- <https://www.linkedin.com/pulse/rise-horseshit-leaders-emil-kresl/>

■ ■ ■