City Research Online

City, University of London Institutional Repository

Citation: Hutchinson, M., Jianu, R., Slingsby, A. & Madhyastha, P. (2024). From LitVis to Reasoning in Data Visualisation: A Research Plan. Poster presented at the Computer Graphics & Visual Computing (CGVC) 2024, 12-13 Sep 2024, London, UK.

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/33687/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

From LitVis to Reasoning in Data Visualisation: A Research Plan

Maeve Hutchinson, Radu Jianu, Aidan Slingsby, Pranava Madhyastha
Department of Computer Science, City, University of London

Motivation

This project aims to use litvis notebooks to create a dataset that encapsulates the decision-making processes in visualisation design, asking the following **research questions**:

- **How can litvis notebooks be used to capture and analyse the reasoning behind visualisation design choices?**
- **Can litvis notebooks be used to develop a more realistic automated reasoning dataset?**

Background

Natural Language (NL) is an essential part of data visualisation, not just as titles, annotations, and labels, but also as a means of conveying the reasoning behind design choices. Current datasets combining Natural Language Processing (NLP) and data visualisation often focus on the final visualisation design, neglecting the underlying reasoning processes that lead to them.

Literate Visualisation (LV) [1] is a visualisation design approach that focuses on documenting the reasoning behind visualisation choices, integrating visual outputs with NL narratives to capture the design process. The platform *litvis* supports LV by combining code, visuals, and text in a notebook format, often following a narrative schema — a structured framework that guides users in documenting their design process.

Data Collection Methodology

We are collecting data from about **800 courseworks from students** learning data visualisation. Students analyse a variety of datasets and produce a comprehensive visualisation design and narrative that reflect their choices and findings.

Students followed a narrative schema designed for this project guiding them to:

- ask research **questions**;
- design **visualisations**;
- describe the resulting **insights**;
- and articulate the **design justification** behind their choices.

The dataset comprises visualisations and their specifications, textual descriptions, and the relationships between each of these elements.
Metadata will also be included, encompassing the mark and textual feedback.
We will use this feedback as a measure of the quality of each submission.

Example Student *litvis* Notebook

Exploring the 2022 Six Nations Championship

What are the research questions that your data visualization will help you to answer?

Research Questions

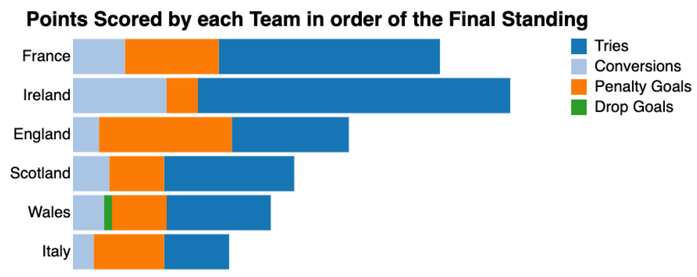
- How does the experience and structure of each team vary?

...

1. The Visualization

Insert your visualization here.

1.1 Context: Final Standing and Points Scored



...

2. Insights

What has your visualization allowed you to discover about your data that help you answer your research questions? Identify a maximum of the 3 most important insights that result directly from your visualization.

2.1 Insight One - Team Experience and Structure

From an initial look at visualisation 1.2, ...

3. Design Justification

Why have you designed your visualizations this way? Itemize up to 3 design choices providing a justification for each supported by literature or evidence of good practice.

3.1 Choice One - Colour

Throughout my visualisation, I have used colour to distinguish nominal variables...

Intended Applications and Future Research

- **Analyse Visualisation Design:** to gain a deeper understanding of visualisation design processes in-practice.
- **AI Model Evaluation:** as a benchmark to evaluate analytical reasoning in models.
- **AI Model Building:** as a knowledge base about data visualisation for AI models.
- **Visual Data Science Discourse:** studying how people communicate about data through multiple modalities (visually, verbally, and through code).
- **Methodology Transfer:** generalise methodology to gather data from more diverse, real-life domains.

References

[1] J. Wood, A. Kachkaev, and J. Dykes, 'Design Exposition with Literate Visualization', *IEEE Trans. Visual. Comput. Graphics*, vol. 25, no. 1, pp. 759–768, Jan. 2019, doi: 10.1109/TVCG.2018.2864836.

giCentre

CITY UNIVERSITY OF LONDON EST 1894