Exploring Cheat Sheets for Data Visualization Techniques

Exploring Cheat Sheets for Data Visualization Techniques

Zezhong Wang* University of Edinburgh
Lovisa Sundin† University of Glasgow
Dave Murray-Rust‡ University of Edinburgh
Benjamin Bach§ University of Edinburgh

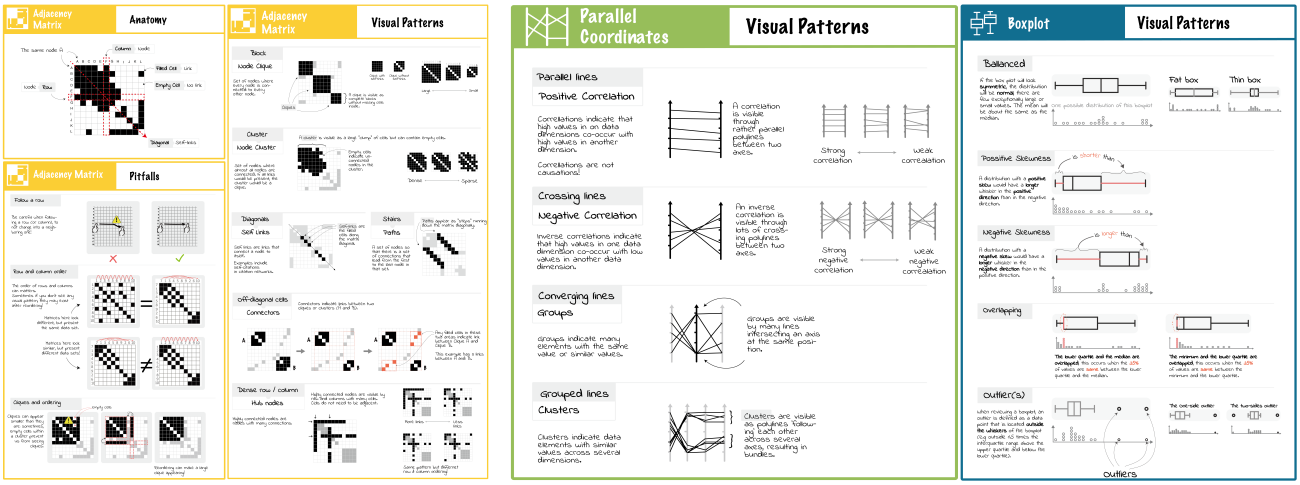

Figure 1: Examples of customized cheat sheets combining several aspects of visualizations. Left: example of sheets for the adjacency matrix showing *anatomy*, *visual patterns*,and *pitfalls*. Right: Cheat sheet showing *visual patterns* for parallel coordinates and boxplots.

Abstract

A cheat sheet is a set of concise graphical and textual explanations, inspired by infographics, data comics, and cheat sheets in other domains. Cheat sheets aim to support learning, teaching, and the regular use of both common and novel visualization techniques in a variety of contexts. To design cheat sheets for visualization techniques, we describe four components of a cheat sheet: *anatomy, visual patterns, pitfalls and introduction* and show examples for three visualization techniques. Our cheat sheets address the increasing need for accessible material that supports understanding data visualization techniques, their use, their fallacies and so forth by a larger audience.

1 Introduction

Visualization techniques are central to the practice of data visualization and its applications and huge number of techniques have been designed. These visualization techniques are generally defined by a specific visual encoding for marks, layout, and visual variables [6]. Although visualization techniques are intended to make information *more* comprehensible, most of them require some degree of learning and understanding of graphical conventions in order to properly decode the representation and 'make sense of the data' [7, 12]. With a rapidly increasing array of techniques and applications across different fields, we see a natural demand for instructional resources on how to learn, teach, and successfully use visualization techniques to an ever growing and heterogeneous audience.

In this paper, we explore *cheat sheets for visualization techniques* (Fig. 1). In particular, we are interested in how to design such cheat sheets that are understandable and if they can be usable. Cheat sheets are used both for summary and support, providing an overview, learning, and casual look up during related tasks. Generally, a cheat sheet consists of graphical explanations for a specific visualization technique. Our cheat sheets are partially inspired by infographics, data comics [3,4,19] and their strengths [17] as well as cheat sheets for programming languages. They aim to provide carefully designed visual and textual explanations, serving as resources which can be adapted for different contexts and purposes. We present cheat sheets for three non-trivial techniques, commonly taught in visualization classes: parallel coordinates plots (PCP), adjacency matrices (AM), and Whiskers Plots (boxplots). For each technique, we designed four different sheets, each of which explains a specific aspect of a technique and supports a different usage scenarios: (a) *introduction* for giving a high-level overview over the technique, (b) *anatomy* (explaining visual elements), (c) *visual patterns* and (d) *pitfalls*.

2 Related Work

The related works include data visualization literacy and data visualization learning supports. Data visualization literacy have been defined in conceptual level [8] as well as task level [6,7]. It describes an ability translating information mutually from an application domain to a visual domain, during which, an audience processing the visualization by understanding 1) the frame of visual encoding; 2) identifying visual patterns and 3) analyzing the content. To aid this process, text books (e.g. [11]) give general concepts, examples and design principles for visualizations, scientific papers (e.g. [5]) demonstrate visual patterns and online resources ([1,2,15]) provide categorizations, textual descriptions as well as aid to select alternatives. In data visualization application, cheat sheets have been used to help designers choose a chart type [1,10]. However, similar to the visualization chooser cards [9] and flashcards [16], these approaches either do not focus on techniques, or concentrate on cataloguing techniques, rather than supporting their understanding. Although the collection is broad, many explanations are lacking in detail, the

*e-mail: zezhong.wang@ed.ac.uk
†e-mail:2092502S@student.gla.ac.uk
‡e-mail: D.Murray-Rust@ed.ac.uk
§e-mail:bbach@inf.ed.ac.uk

patterns that can be seen in the graphs are rarely discussed and only superficially. To the best of our knowledge, there is no resource or format that addresses the details of our cheat sheets.

3 Cheat sheet types

Each cheat sheet type consists of a header that includes an icon and a unique color for a technique, the technique name, as well as the type of sheet (see the example in supplementary materials).

Anatomy *explains the visual elements of a visualization technique, their composition, their specific terminology, and how they relate to the data.* The visual components of a visualization can include individual visual marks (in the sense of Bertin [6]) as well as groups of marks, axes, locations in a visualization etc. Often, visualizations use similar visual marks to refer to different parts of a visualization, e.g., lines in a PCP are used for both dimensions and elements. Other visualizations use visual marks with very specific meaning such as the Whiskers in a boxplot. Understanding the individual visual elements components is key to understanding, correct interpretation as well as communication and collaboration. we decided to show the visual term on a gray background while the data term is shown on white background associated through a jigsaw shape. For simplicity we did not include alternative terminology (e.g., 'node', 'vertex'). Additional explanations are added in sparse textual annotations. Red lines are used whenever too much information is overlaid.

The ***visual patterns*** *sheet (or "patterns" for simplicity) provides a catalogue of meaningful patterns visible in a visualization.* Patterns are key in correctly interpreting visualizations and making discoveries. A visual pattern can be any sort of configuration in the visualization with a specific meaning. We found many patterns in the literature describing a specific technique (e.g., [5, 18]). A **visual explanation** that shows the 'idea' of a pattern and can include annotations. We use dark shading to highlight the pattern and gray shading to indicate its context if necessary. **Variations** of the patterns are added for each pattern, showing differences in pattern size, expression, location etc.

Pitfalls *show possible misinterpretations of a visualization.* While there is a long list of technique independent pitfalls [13, 14] (including missing labels, different but hard to perceive colors, not colorblind safe, or deceptive titles), we aim to list pitfalls specific for a given technique. Moreover, while some pitfalls are introduced by bad or intentional design of an individual visualization (e.g., truncated or inverted axes), other pitfalls are intrinsic to the technique and require knowledge, rather than better design.

Our design for the *pitfalls* sheets includes three choices: i) highlighting the part that could be overlooked with arrows, circles (e.g., in PCP a scales are not necessarily start from '0'), or an exclamation mark symbol ⚠, ii) using comparison to address the difference (e.g., equal versus not equal in a matrix and ascending scale versus descending scale in a PCP) and iii) using '✓' and '✗' for things to do and things to avoid (e.g., do not change to a neighbouring row while following a path in a matrix).

An ***introduction*** *introduces a visualization technique in a specific context, explains its purpose, and highlights important aspects described in the individual sheets.* An introduction is a more elaborate component that can include information from other sheets to provide an introduction into the technique and overview over information in the related sheets. For our introductions, we were mainly inspired by data comics [3, 19].

4 Discussion and Future Work

The sets of cheat sheet have been designed around simple, extensible templates with the intention that there is a flexibility in the way in which they are used. In particular, the modularity of our cheat sheets makes it possible to assemble their content in new ways, collecting and organizing the components to fit a particular audience and format. We believe our initial investigation is showing the potential of cheat sheets and the current gap in resources for visualization literacy. Hence, we invite designers, educators, and researchers, to follow our example and to create cheat sheets and adapt them for their needs. There are several directions for the expansion of cheat sheets that merit exploration. Firstly, we want to create cheat sheets for other rather unfamiliar visualizations, and explore alternative designs and styles to our current sheets. Secondly, we need to obtain feedback from in the wild studies to assess the value of cheat sheets for understanding and using data visualizations. This will help us improving our designs and decide how to best represent complex content in a concise way. Eventually, we need to run more ethnographic studies to assess which information is required on the cheat sheets and for which user group.

References

[1] Chartdoctor. online: `https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf`,, 2019.

[2] Datavizproject. online:https://datavizproject.com,, 2019.

[3] B. Bach, N. H. Riche, S. Carpendale, and H. Pfister. The emerging genre of data comics. *IEEE computer graphics and applications*, 37(3):6–13, 2017.

[4] B. Bach, Z. Wang, M. Farinella, D. Murray-Rust, and N. Henry Riche. Design patterns for data comics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 38. ACM, 2018.

[5] M. Behrisch, B. Bach, N. Henry Riche, T. Schreck, and J.-D. Fekete. Matrix reordering methods for table and network visualization. In *Computer Graphics Forum*, vol. 35, pp. 693–716. Wiley Online Library, 2016.

[6] J. Bertin, W. J. Berg, and H. Wainer. *Semiology of graphics: diagrams, networks, maps*, vol. 1. University of Wisconsin press Madison, 1983.

[7] K. Börner, A. Maltese, R. N. Balliet, and J. Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016.

[8] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics*, 20(12):1963–1972, 2014.

[9] S. D. H. Evergreen and G. McMahon. Data visualization chart chooser cards. http://chartchoosercards.com/.

[10] S. Franconeri. Which visualization? a quick reference. http://experception.net/Franconeri_ExperCeptionDotNet_ChartChooser.pdf, 2018.

[11] A. Kirk. *Data visualisation: a handbook for data driven design*. Sage, 2016.

[12] A. V. Maltese, J. A. Harsh, and D. Svetina. Data visualization literacy: Investigating data interpretation along the noviceexpert continuum. *Journal of College Science Teaching*, 45(1):84–90, 2015.

[13] M. Monmonier. *How to lie with maps*. University of Chicago Press, 2018.

[14] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1469–1478. ACM, 2015.

[15] S. Ribecca. The data visualization catalogue. online: `https://datavizcatalogue.com`.

[16] J. Schwabish. The graphic continuum flash cards. https://policyviz.com/product/the-graphic-continuum-flash-cards, 2017.

[17] Z. Wang, S. Wang, M. Farinella, D. Murray-Rust, N. H. Riche, and B. Bach. Comparing effectiveness and engagement of data comics and infographics. 2019.

[18] H. Wickham and L. Stryjewski. 40 years of boxplots. *Am. Statistician*, 2011.

[19] Z. Zhao, R. Marr, J. Shaffer, and N. Elmqvist. Understanding partitioning and sequence in data-driven storytelling. In *International Conference on Information*, pp. 327–338. Springer, 2019.