Cheat Sheets for Data Visualization Techniques

Zezhong Wang[1]**, Lovisa Sundin**[2]**, Dave Murray-Rust**[1]**, Benjamin Bach**[1]

[1] University of Edinburgh, UK, [2] University of Glasgow, UK
{zezhong.wang, d.murray-rust, bbach}@ed.ac.uk, 2092502s@student.gla.ac.uk,
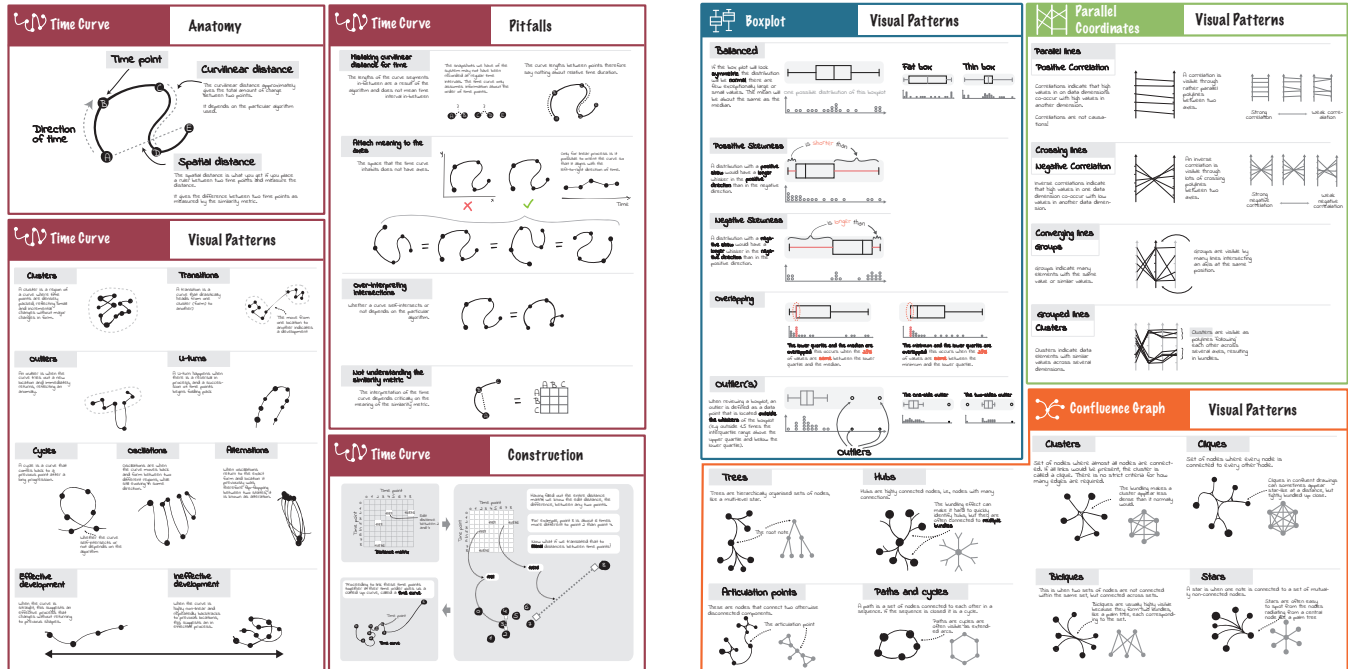
Figure 1. Examples of customized cheat sheets combining several aspects of visualizations. Left: example of look-up summary sheets for the time curve visualization showing *anatomy*, *visual patterns*, *construction*, and *pitfalls*. Right: Cheat sheet showing *visual patterns* for several techniques: boxplots, parallel coordinates, and confluent graphs. More cheat sheets and information can be found at: https://visualizationcheatsheets.github.io

ABSTRACT

This paper introduces the concept of 'cheat sheets' for data visualization techniques, a set of concise graphical explanations and textual annotations inspired by infographics, data comics, and cheat sheets in other domains. Cheat sheets aim to address the increasing need for accessible material that supports a wide audience in understanding data visualization techniques, their use, their fallacies and so forth. We have carried out an iterative design process with practitioners, teachers and students of data science and visualization, resulting six types of cheat sheet (*anatomy, construction, visual patterns, pitfalls, false-friends and well-known relatives*) for six types of visualization, and formats for presentation. We assess these with a qualitative user study using 11 participants that demonstrates the readability and usefulness of our cheat sheets.

CCS Concepts

•**Human-centered computing** → **Visualization theory, concepts and paradigms; Empirical studies in visualization;**

INTRODUCTION

Visualization techniques, also called visual representations, turn patterns in data into visual patterns that can be read and understood by people that are central to the practice of data visualization. While visualization techniques are intended to make information comprehensible, many techniques require some degree of learning and understanding of the visual mappings, graphical conventions, symbols, marks, layouts and patterns in order to properly decode the representation [23, 49, 52]. With a rapidly increasing array of techniques and applications across different fields, we see a natural demand for instructional resources to support a growing and heterogeneous audience in creating and decoding data visualisations.

There is no shortage of resources to learn about data visualization, including books (e.g., [46, 56, 27]), survey papers, workshops (e.g., [60, 69]), classes, and online catalogues of techniques [59]. However, much of this material talks about general topics in visualization such as deception, design, presentation, tasks, storytelling or layouts. We find a lack of

material that supports the understanding of specific visualization techniques: *How to decode visualizations correctly? What patterns can we see and how should we interpret them? What are common misinterpretations to avoid for a given technique?*. There is a specific need for resources that are tightly structured and standardized, which align with established visual design principles and which are easy to understand.

In order to support the development of material for learning, teaching, and the regular use of visualization techniques, this paper presents *cheat sheets for visualization techniques* (Fig. 1). In particular, we are interested in how to design cheat sheets that are understandable and usable. Our cheat sheets are inspired by infographics, data comics [73, 15, 74], assembly instructions, and the wide range of examples of cheat sheets for programming languages. In the context of visualization techniques, we **define a cheat sheet as** *"a set of concise, annotated graphical explanations of aspects of a specific visualization technique"*. They aim to provide carefully designed visual and textual explanations in a concise way, while focusing on the most important knowledge necessary for given tasks. We imagine cheat sheets to support two scenarios: *i)* first-time *learning* aided through slides, posters, books, or activities, and *ii)* as *look-up* material during an actual data exploration process. Our cheat sheets, in their current form, do not aim to overview and classify visualization techniques, nor to support finding the "right" technique for a task.

In this paper, we present cheat sheets for a selection of non-trivial visualization techniques representative to different data types (temporal, relational, hierarchical, statistical/quantitative, multidimensional) and taught by our collaborators: *parallel coordinates plots* (PCP), *adjacency matrices*, *Whiskers Plots* (box-plots), *tree maps*, *confluence graphs* [16] and *time curves* [17]. Using an iterative design process that involved 8 experts and regular feedback, we designed six different *types* of cheat sheets for each technique, each of which explains a specific aspect of a technique and supports a different usage scenarios: (a) *anatomy* explaining visual elements, (b) *construction* explaining the general idea of a technique, (c) its *visual patterns*, (d) common *pitfalls*, (e) *well-known relatives* relating a visualization to a more commonly known representation, and (f) *false-friends* which show visualizations that look similar but have to be interpreted differently. *Introduction* is to give a high-level overview over the technique. Results from the controlled user study and co-design sessions show consistently high ratings for readability and usefulness, encouraging extension to other techniques and the development of class-room activities around cheat sheets. We refined our design guidelines and respective visual design-materials in our co-design workshops, eventually inviting everyone to create cheat sheets for teaching, personal learning, and for any new visualization technique that gets designed.

RELATED WORK

Visualization Literacy
Many forms of literacy are considered essential in a data society, such as data literacy, numeracy, and visual literacy [26, 62]. *Visualisation literacy* relates to these, but is specifically situated around making use of visualizations: Boy et al. [25] provide a formal description for visualization literacy as *"the ability to confidently use a given data visualization to translate questions specified in the data domain, as well as interpreting visual patterns in the visual domain as properties in the data domain"*. This definition emphasizes the ability to translate information from the specific application domain to the visual domain and back. The translation requires correctly interpreting the graphical conventions of a visualization in order to create a consistent mapping between data and graphics, so that higher-level domain tasks can be mapped onto lower-level visualization tasks. Consequently, Börner et al. [23] define visualization literacy at a task level including *"interpreting patterns, trends, and correlations in visual representations of data"*. Bertin [22] describes three sequential stages of interpreting a visualization: an *external identification stage* where a reader recognizes the frame of visual encoding (e.g. axes, labels and dimensions); an *internal identification stage* where a reader identifies visual characteristics (visual patterns); and a final *perception of correspondence stage* where a reader analyzes the content and extracts messages from the visualization. Similarly, the NOVIS Model [49], describes readers encountering unfamiliar visualizations as wandering through the following five activities: *encountering*, *constructing a frame*, *exploring the visualization*, *questioning the frame* and *floundering on the visualization*.

Visualization Cheat Sheets are our approach to support visualization literacy by providing a concise and consistent set of explanations for understanding specific visualization techniques, their visual marks and visual patterns. As learning is a process, our sheets aim to support a range of tasks and steps related to visualization literacy while focusing on supporting learning and lookup during use.

Learning Support for Visualization Techniques
Materials to support the learning of visualization techniques are presented in various formats with different scenarios in mind. Scientific papers introducing a novel technique can often be considered a primary source of learning about a technique. To this end, many papers include conceptual descriptions of the technique, diagrams on the creation of examples and demonstrate visual patterns (e.g. [40, 39, 14, 17, 16]), although some focus exclusively on the visual patterns (e.g. [37, 72, 21]). Text books (e.g. [46, 28, 58, 71]) tend to offer summaries of techniques alongside comprehensive knowledge about visualization design principles, goals, work flows, visualization categories, encoding and collections of examples and specific visualizations techniques. For example, Munzner [56] explains visualization techniques on a by-case bases, showing example implementations and discussing advantages and drawbacks. Researchers also propose the concept of learning unfamiliar visualizations by analogy [61], and tool for progressively constructing visualizations from visual channels and units [68].

More systematic description of visualizations are found in Wilkinson [71], who demonstrates a catalog with mathematical principles and fundamental components of quantitative statistics graphics. However, to quickly understand and use visualization techniques, more succinct and novice-friendly instructions are required. Kirk [46] systematically presents 39 visualisation techniques, each being presented on a single page, with primarily textual descriptions, but with consistently structured information covering *description*, *examples*, *how*

to read it, *presentation tips* and *variations & alternatives*. Similarly, some online resources gather information about very specific techniques, prepared by different educators and without general structure, e.g, on WikiPedia [11] and online videos [41], or collections of pitfalls, but without graphical explanations [5]. The DataVizProject [7] and Chart doctor [4] list types of visualization techniques with examples, categorization and descriptions, and backed up with abstract pictograms, but do not provide specific aid for understanding and learning.

The project closest to our visualization cheat sheets are some examples in the Data Visualization Catalogue [59]. The catalogue lists 60 visualization techniques, each of which coming with an example, a textual description, an anatomy, links to tools, and a lose tag list (e.g., 'data-over-time', 'patterns'). For some examples, it lists similar techniques to aid selection. With the same goal of identifying suitable chart types, *data-to-viz.com* [38] provides similar contents yet adds 'common mistakes' and 'build your own' with R and python galleries. While the collection covers many techniques, many explanations lack detail, depicted patterns are rarely and only superficially explained, and there is no deeper support in engaging with learning and using the technique. To the best of our knowledge, no resource or format exists with cheat sheets in the detail we propose.

Cheat Sheets and Visual Explanations
The traditional usage of 'cheat sheet' is *"a piece of paper bearing written notes intended to aid one's memory, typically one used surreptitiously in an examination."* [10, 31]. Using—but more so preparing—cheat sheets has been found to improve students' exam performance [67, 66], especially when students spend time with preparing cheat sheets [32, 42]. Suggestions for good cheat sheets include hand-made content and good (visual) organization [66]. Consequently, cheat sheets have been used to learn programming languages and support their daily use [1, 50, 19, 42], data analysis [6], and types of neuronal networks [1]. These cheat sheets provide the most important information at a glance, and are intended to be looked up during a specific task. In data visualization, types of cheat sheets have been used to help designers choose a chart type [36, 4, 35] and design effective visualizations [64, 43]. However, similar to the visualization chooser cards [33] and flashcards [63], these approaches either do not focus on techniques, or concentrate on cataloguing techniques rather than supporting use and understanding.

Our cheat sheets draw on this practice of providing important information at a glance in a structured form, and aim to support the learning of visualization techniques. This is distinct from the process of *choosing* a specific visualization type, instead dealing with the construction and metaphor, the variety of visual patterns, possible visual misinterpretations [57] specific to a particular technique. To provide appropriate visual illustrations, we draw inspiration from infographics and comics. Comics in particular, with their unique combination of complementary text and pictures, sequential narrative, and general familiarity, have been found beneficial for science education [34, 45], communicating complex data-related concepts [69] including confidence intervals, hypothesis tests, regressions, and simple statistical graphs [47, 65]. More recently, comics

have been used as general format for data-driven storytelling medium [73, 15], coming with specific design patterns [18].

DESIGN PRINCIPLES
Based on early exploratory prototypes and background literature, we identified the following initial design principles for creating cheat sheets:

- **D1—Modularity:** To introduce structure, we propose *types* of cheat sheets, each focusing on explaining a specific aspect such as visual patterns or visual parts. This is to reduce overcrowding, and to allow for different combinations of sheets to support different situations and different *formats*.
- **D2—Context independence:** Sheets should not rely on specific data examples or external contexts.[1] Our goal is to explore the *content* of cheat sheets and which concise graphical illustrations are required to communicate this content. By keeping our sheets context-free and abstract, we aim for reusability across examples and to learn about how abstract sheets *can* be while still allowing people to apply them to their specific examples at hand.
- **D3—Clear graphics:** As data visualizations are a form of visual languages and aim to facilitate memory and lookup, our goal with cheat sheets is to explain as much as we can through graphical content and use text where graphical content alone is not sufficient.
- **D4—Style neutrality:** Strong graphic styles can have implications and can render a graphics work specific to a narrower audience [51]. Following our principle of context-independence, with this paper, we aim to explore cheat sheets for a general audience and to allow for simple adaptation to more specific cases where necessary. For this work, we used stylistically consistent simple black-white graphics, partly inspired by the very minimalist but clear *xkcd* style [55], while at the same time avoiding overly minimalist designs [20]. Although color is an important visual variable, we focus on the basic structure and layout of visualization techniques; color is a general design decision whose application is often independent from a specific technique and therefore greatly varying with the specific visualization instance.

RESEARCH AND DESIGN METHODOLOGY
We designed a collection of cheat sheets using an iterative seven-step design process.

1) Idea generation: we selected several common yet non-trivial visualization techniques such as treemaps, adjacency matrices and horizon graphs, and used these to explore a range of cheat sheet types to explore various aspects: comparison, variation, advantages and drawbacks, concrete examples etc. **2) Expert workshop and co-design:** we conducted a two-hour co-design workshop with three senior university lecturers specialising in geographic visualization, medical informatics, and data visualization. Participants were asked to introduce a visualization technique (a PCP) to students unfamiliar with that techniques using sketches and as little text as possible (D3). All participants created material in a comic style with concrete examples and a story to motivate and introduce the specific

---
[1]Note that following the Expert Workshop in Section 4 we relaxed this principle

technique (e,g., peoples' biological characteristics, fruits and nutrition facts). Participants were then asked to feedback on the draft sheets prepared in step 1 (visual patterns, pitfalls and, construction, variations, and comparison) for PCPs, boxplots, and treemaps. The workshop highlighted the need for concrete examples (contrary to our principle of context-independence) and helped us narrow down our choice to the final set of six sheet types. **3) Expert interviews:** after redesigning our sheets based on feedback, we consulted individually with four university lecturers (medical informatics, graphic design, illustration and digital learning), on potential use cases. As well as small design tweaks, suggestions included: handing out the *Introduction* before a lecture and *visual patterns* during and after the lecture; handing out all sheets in a creative course on physicalization to inspire students; adapting the style and data examples from the *Introduction* to the specific context of their own lectures. **4) Readability feedback:** after another set of internal iterations, we conducted individual readability walk-throughs with four graduate students in design and computer science, all of whom had briefly encountered the respective techniques in their lectures. We obtained feedback on ambiguities that required further explanation and how much they have understood the concepts. **5) Design-by-example:** we asked a researcher and designer unfamiliar with our research to create cheat sheets for additional visualizations based on our examples (they are now a coauthor of this paper). The example's structure was found to be straightforward and salient enough for the new researcher to produce sheets for three additional visualizations with only minimal consultation. The main question involved the *Introduction* sheet (explained in Section 5), the structure of which was deliberately left open to encourage tailored story-lines for the specific visualization and the data examples necessary to explain a visualization. **6) Readability study with non-experts:** to obtain further feedback on readability, usability and to refine our design, we conducted an empirical user study following the rapid iterative test and evaluation method [53] with 11 participants novel to visualization (Section 7). **7) Guideline workshop:** we eventually, conducted two workshops with a total of five visualization experts and non-experts to apply, evaluate, and refine design guidelines for creating additional cheat sheets.

CHEAT SHEET TYPES

This section reports on six types of cheat sheets informed by the expert workshops, expert interviews and the readability study with non-experts. According to our design principle of modularity (D1), this section describes cheat sheets for individual aspects of visualization techniques. Figure 1 gives an idea how different sheet types can be combined into a *format* for a final presentation. For each type we give a brief definition, explain its motivation and its design.

ANATOMY *explains the visual elements of a visualization technique, their composition, their specific terminology, and how they relate to the data (Fig. 2).* The visual components of a visualization can include individual visual marks (in the sense of Bertin [22]) as well as groups of marks, axes, locations in a visualization etc. Our design for *Anatomy* was inspired by graphics-first explanations in patents using a bare minimum of ink while relying on black, white, and shades of gray only. Text labels are connected to their components through simple
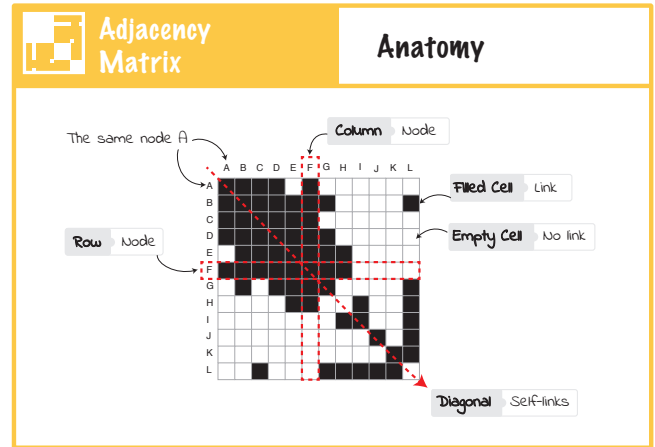

Figure 2. *Anatomy* **sheet for adjacency matrix; labels for visual elements are shown on a gray background for highlighting while data terms are shown on a white background linked to the visual terminology; arrows are curved to not interfere with the rather geometric visualization; red is used to show information overlaid on black elements.**
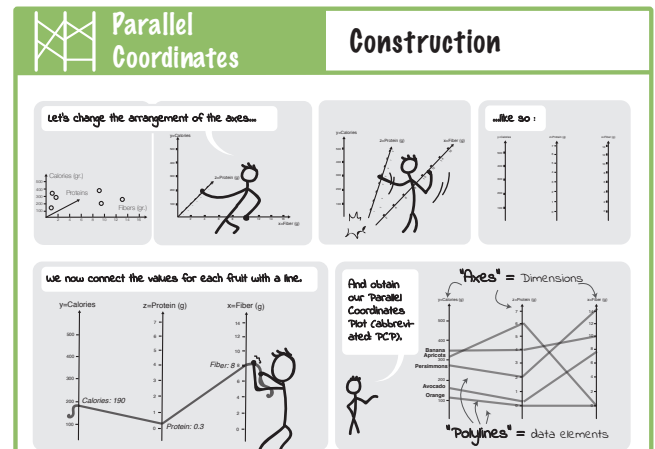

Figure 3. **Part of Construction for PCP, showing its "creation" from three-dimensional scatterplot in a comic-strip.**

leader lines. We found this design to increase readability, communicating on an abstract level without confounding the reader with specific but unimportant details (e.g., thickness of axes, number of lines, 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. Constructing precise terminology here is a challenge, as *i)* competing terminologies exist ('cell' vs. 'mark' in an adjacency matrix); *ii)* some elements do not have defined terminology (the box in a box plot); *iii)* visual representations have distinct names from the data concepts they represent ('axes' as visual representations in PCP versus 'dimensions' as the underlying concept). We 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.
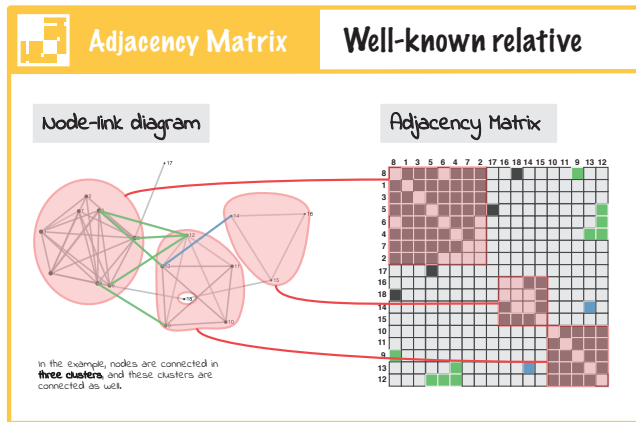
Figure 4. Well-known relative for adjacency matrix, adapted from [40]. Node-link diagram are familiar to most people, helping to understand the less familar marices: Blue and green represent the connectors between clusters in both charts.

CONSTRUCTION *explains the conceptual idea behind a visualization design and how it encodes data.* It reflects the relation between the data and its visual encoding by demonstrating the process of constructing or providing a familiar metaphor (e.g., Fig. 3) to help understand and memorize the encoding mechanism. While *Anatomy* explains the visual elements and their terminology, *Construction* provides a procedural explanation of a visualization design and delivers a blue-print for how to explain the visualization to a larger audience. For example, in a report, a presentation, or video, a presenter can show and explain a *construction* to the audience *before* talking about results in a visualization. Our constructions were largely inspired by data comics (e.g. the *build-up pattern* [18]) and videos that explain the visual encoding process in animated presentations (e.g., [2]) as well as explanations in the papers introducing visualization techniques, e.g., [40, 17, 16, 70])

WELL-KNOWN RELATIVES *highlights other approaches to visualizing the same data.* Relating new knowledge to existing knowledge is fundamental for learning, and there are often complementary techinques for visualising a given dataset to draw on [30], e.g., boxplots, barcharts and swarmplots for distributions, scatterplots and PCPs for multidimensional data. Figure 4 shows node-link diagrams as a relative of adjacency matrices, inspired by Riche et al. [40]. Node labels and groups are the same in both diagrams, while elements are highlighted and linked either implicitly (red clusters) or implicitly (element highlighted in red and green).

VISUAL PATTERNS *sheets (or "patterns" for simplicity) provides a catalogue of meaningful patterns that may occur in a visualization.* Spotting patterns can be key to correctly interpreting visualizations and making discoveries—for example, network cliques in adjacency matrices are visible as solid blocks (Fig 5, relies on a good ordering [21]). By providing a set of patterns and their explanations to users, cheat sheets can support users in developing understanding using a visualization.

Many patterns are described in the literature for specific techniques (we draw on [21, 17, 16, 70] here). Terminology is not always clear, e.g., naming either the visual representation or the concept being represented: 'block' versus 'clique' for
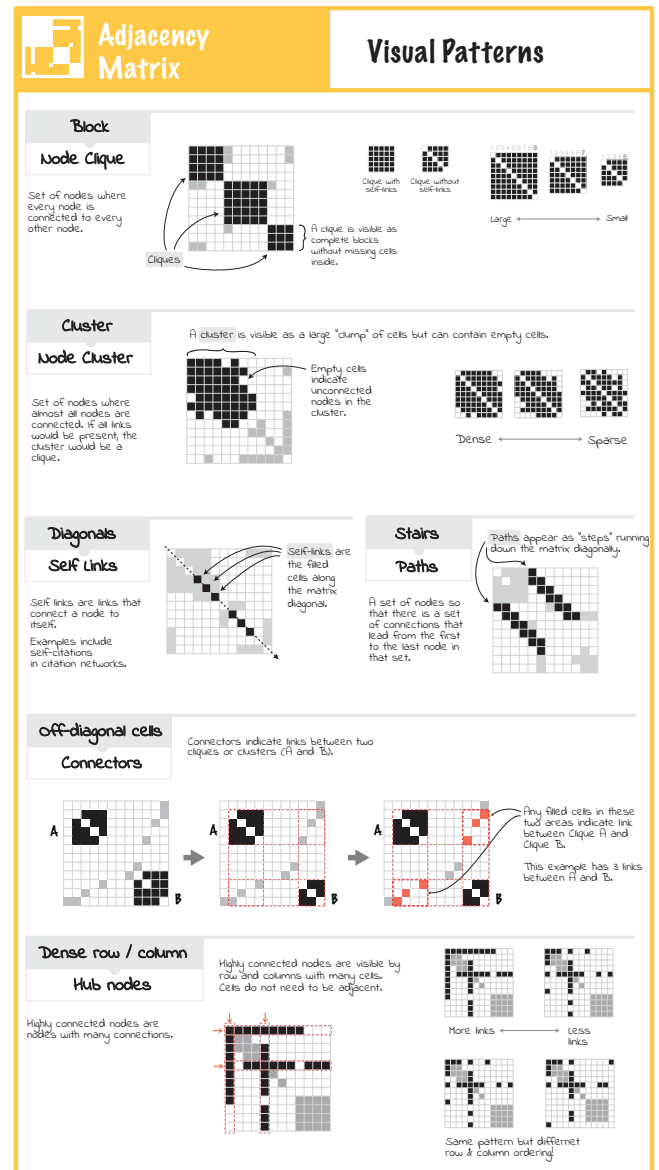

Figure 5. *Visual patterns* **for adjacency matrix. Patterns are separated from each other by gray dividers. Patterns are given two names, a visual name (e.g., 'Full Block') and a data name (e.g., 'Node clique'). Most pattern are shown by an example, complemented with text. Pattern variations (e.g., dense *vs.* sparse clusters) use scaled-down graphics.**

networks, 'crossed lines' versus 'inverse correlation' and so on. For our examples, we chose the 4-8 most common patterns for each visualization and describe them as follows:

- **Title** gives the visual appearance ('block') while **subtitle** gives the interpretation ('clique') with a **description**.
- A **visual explanation** that shows an abstract example of a pattern, using annotations and shading to explain concepts.
- **Variations** of the patterns are used to show differences in size, expression, location etc. From the user study we found *variations* to be essential for understanding patterns and to correctly interpret a pattern.

PITFALLS *show possible misinterpretations of a visualization.* Distinct from ways that visualizations can be poorly designed
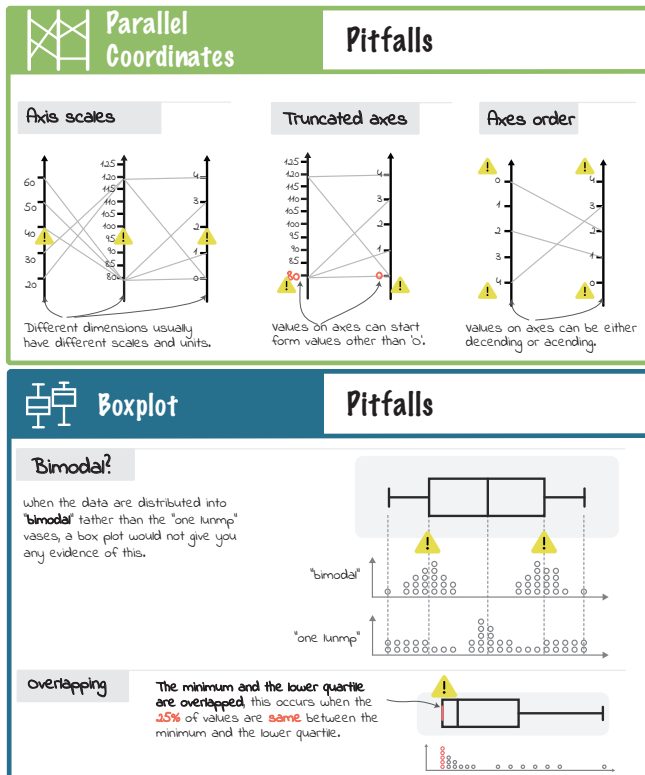
Figure 6. Examples of pitfalls for PCP (top) and boxplot (bottom). Yellow exclamation mark signs show areas where a pitfall can occur.

Figure 7. *False-Friends* for boxplots: the target visualization is shown on the left on a grey background to separate it from the rest. Next to it, separated by a '≠' sign, are shown visualizations that look similar but have different meanings and uses, annotated with main differences and brief explanations of their use.

(e.g., missing labels, different but hard to perceive colors, not colorblind safe, or deceptive titles [57, 29, 44, 54, 48]), pitfalls are specific to a given technique, and can occur even with well designed outputs. We can group pitfalls on our sheets into three major groups: human inattention (e.g., missing to check for min, max value on an axis), visual encoding (e.g., in a boxplot where the minimum and the lower quartile overlap), and data features obscured by particular visualization techniques (e.g., boxplots are not able to show whether a distribution is bi-modal or normal distributed). Our design for the *pitfalls* sheets uses three strategies (Figure 6): *i)* highlighting problematic parts of the visualisation (e.g., scales in PCPs do not always start from '0') with arrows, circles 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 when constructing visualisations (e.g., do not show axes for time curves [17]).

FALSE-FRIENDS *lists visualization techniques that are visually similar but functionally different.* For example, boxplots share visual similarities with candlestick charts and error bars in bar chars (Fig 7); parallel coordinates can look similar to line charts, time curves can look similar to connected scatterplots [37]. Our *false-friends* module aims to clarify these differences by placing the 'false-friend' beside the actual visualization, separated by a '≠' sign. For each 'false-friend', we add one sentence to introduce the applications and annotated the different components.
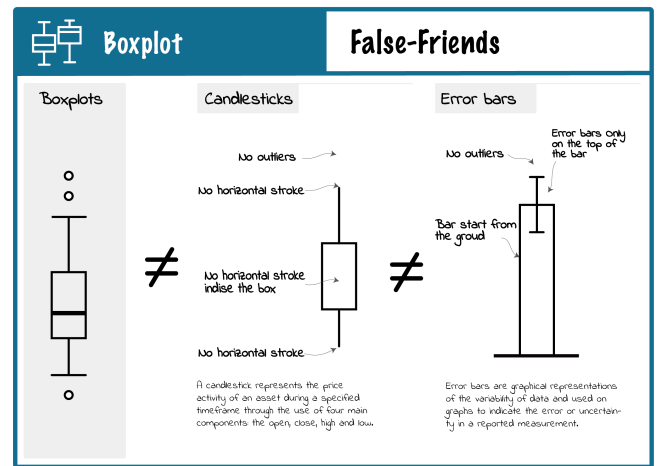
FORMATS

Figures 1 and 8 show *formats* of cheat sheets, i.e., cheat sheet types organized for a specific presentation purpose. While the examples in Figure 1 show two ways of grouping cheat sheets—by type (left) and by technique (right), Figure 8 shows a more elaborate example. An ***Introduction*** *is a format that* introduces a visualization technique within 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 (Fig. 8). For our *Introduction*, we were mainly inspired by data comics. After several iterations and feedback from our expert workshop, we settled on the following structure for an *Introduction*:

1. **Context:** start with a real-world example (e.g., fruits) that introduces the data type (e.g., multivariate data) and high-level tasks (e.g., finding fruits for with balanced nutrition values). This can include characters if necessary.
2. **Well-known relative:** then, introduce a visualization technique familiar to your audience (e.g., node-link diagrams) and explain its limitation to motivate why an alternative techniques is required.
3. **Construction & anatomy:** then, show a *construction* to introduce the novel visualization and explain its visual components with the *anatomy*.
4. **Visual patterns & pitfalls**: explain how to read the visualization by explaining some visual patterns in the context of your data example from step (1). If necessary, mention some pitfalls in the data example.
5. **Joke:** Inspired by stories to provoke engagement, a final joke has been highlighted by the participants in our study

EVALUATION

We conducted a readability study with non-experts, in which participants used our material to support answering questions about three different complex visualisations. This was not a comparative study for two reasons: firstly, as mentioned in Section 4, we were carrying out an iterative process involving
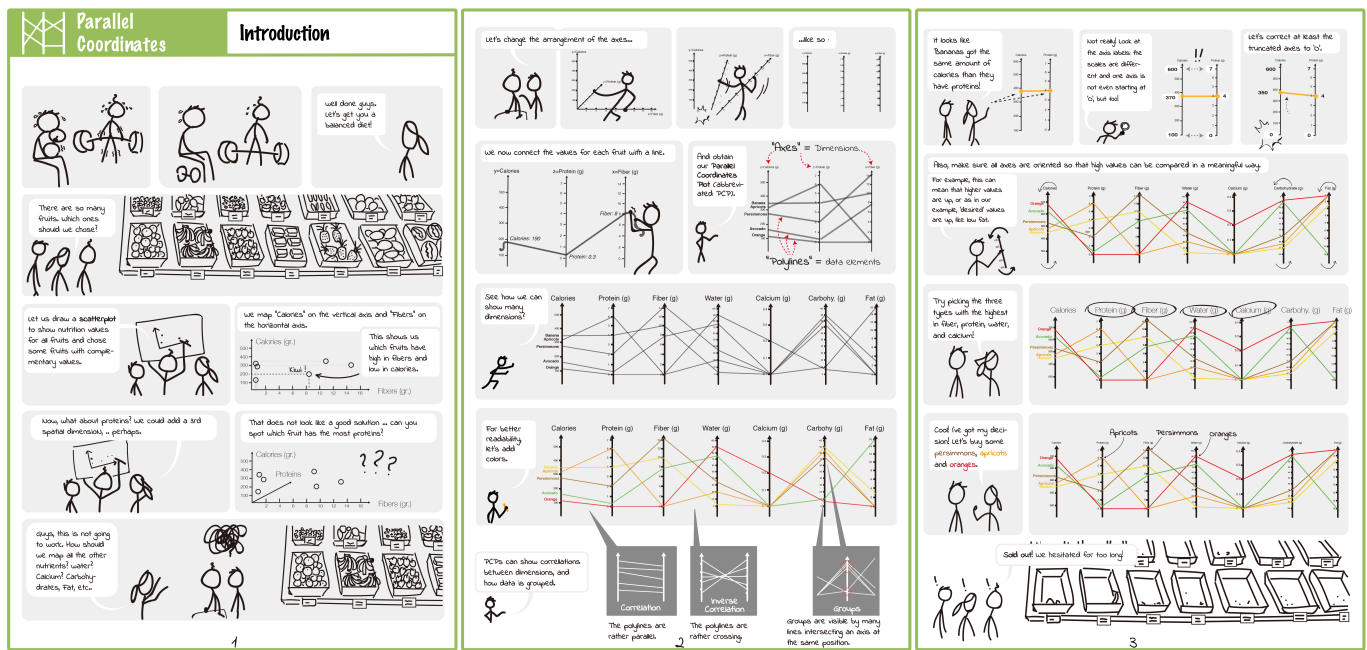
Figure 8. *Introduction* for parallel coordinates, featuring the following parts: introducing a data example and giving an example of a common visualization (scatterplot), constructing the new visualization (PCP), explain some patterns and pitfalls, then finish with a joke.

co-design, so the sessions served to support design iterations (similar to Bach et al. [13]) and identify missing patterns or pitfalls, or other issues with the cheat sheets. Secondly, it was not clear which baseline techniques would be appropriate: purely textual techniques would suffer from a lack of the necessary graphical illustrations, while the linear format of video explanation precludes rapid lookup during use.

The study tested for only three visualizations to keep the overall study length to the reasonable time of 1h per participant. We chose the three visualizations with real-world examples for the quiz. During the study, we gave participants cheat sheets for each of three visualization techniques: *boxplots*, parallel coordinates (*PCP*) and adjacency matrices (*matrix*). Participants were given our material and then asked to answer a quiz on each visualization type. The study included two distinct uses of the cheat sheets: *learning*, where the participants read the *Introduction* format, and *lookup* where the participants used the material while solving the quiz. During the study, participants were encouraged to annotate the sheets with any omissions, points of confusion or suggestions they had. After the study we gathered subjective feedback on aesthetics, readability, usefulness, and general feedback about preferences and misinterpretations. As highlighted in Section 4, we included the feedback from these sessions in our final designs wherever possible.

Participants

Participants were recruited from a local university. To select potential participants with little pre-knowledge about the visualization types that would be tested in the study, we asked respondents to rate their familiarity with the visualizations to be tested on a 5-scale Likert scale (*never seen* (0) to *very familiar* (4)). From the 91 replies we received, we selected the 11 with the lowest rankings. Eight participants rated their familiarity with all visualizations as 0 (*never seen*), while 3 scored 1 for box plots but 0 for the rest. From the 11 selected participants

(10 female), 9 were European, 1 Northern American and 1 South African. Four were undergraduates, 3 graduates and 3 doctoral students with background including cognitive science, psychology, philosophy, linguistics art, and university administration staff. Eight participants were native English speakers, 3 staying in a English speaking country for more than half year.

When arriving at the lab, participants were again asked for their familiarity with visualizations in general. Four participants rated themselves 1 out of of 8 on a scale of expertise with 1 being the lowest, 4 rated 2, 1 rated 3 and 1 rated 6. Most participants reported they had only seen graphs like bar and piecharts in news or academic papers. One participant had experience in reading fMRI visualizations but no experience with abstract information visualizations. 5 participants indicated reading visualizations on yearly basis, 1 on weekly basis, 3 on monthly basis and 1 read visualizations daily.

Procedure and Material

For each of the three techniques (*boxplots*, *PCP*, *matrix*) we produced a cheat sheet with a sheet for *Introduction*, *anatomy*, *visual patterns* and *pitfalls*. Each sheet was printed in color on its own A4 sheet. After a brief introduction, participants carried out an exercise with each visualization technique in increasing order of difficulty (based on formal ratings in pilot tests):*boxplot*,*PCP*,*matrix*. For each technique, participants first read the *introduction* while thinking aloud, and were encouraged to annotate anything that was unclear. Next, they were given an example visualization for the respective technique and asked to answer a small quiz to test their understanding. During the quiz, participants were given printouts of *anatomy*, *visual patterns*, and *pitfalls*. Again, participants were asked to think aloud and annotate the material where unclear. Finally, participants were given a questionnaire asking them to rate *understandability*, *aesthetics*, *usefulness*, *fun*, and *engagingeness* (5-point Likert scale). Each participant obtained £15 for the 1 hour study.

Data and Questions

Each quiz contained between 6 and 7 questions about that particular visualization example, chosen with respect to the cheat sheet components available to the participants. Questions covered low-level reading tasks such as finding single values, making comparisons and identifying patterns. Answers included multiple choice, true-false and short answers, and each question contained an 'I am not sure' option to minimise guessing. The full set of materials and questions can be found in the appendix. In detail:

- We showed 27 **boxplots** of varying shapes, covering the *1908 summer Olympic athlete ages* [9]. Example questions included true or false questions ('More than half of female archery athletes were older than than 40 years old.') and short answer questions ('How many groups were positively skewed?'). Some questions were constructed to be challenging, e.g. 'More than 25% of male boxing athletes were 30 or above 30 years old' required participants to understand overlapping quartile ranges and their meaning.
- for **PCP** we chose two data sets to cover our questions; Fisher's Irish data [8] (with 4 dimensions) and the car parameters data set [3] (8 dimensions). Example questions included 'What's the relation between E and F?' and 'Whats the biggest values recorded at D?'.
- For **matrix**, we chose Les Misèrables co-occurrence network [24], showing visual patterns such as cliques, clusters, or bridges. Example questions for the *matrix* included 'Are the nodes 14 and 77 connected?' or 'Are cliques A and B connected?'.

As the study went on, we made some modifications to the example datasets to improve readablility, as well as better fitting the examples to our tasks. For example, for the *matrix*, we applied an ordering algorithm to facilitate the identification of clusters [21]. For *boxplot* we removed some overlapping outliers to avoid confusion and allow precise counting where necessary. For *PCP*, we reversed one axis to introduce a typical pitfall. In all cases, we anonymized the data labels.

Results and Findings

Our data included quantitative results from the quiz, subjective feedback, and findings from thinking aloud.

Cheat sheets usability and readability—Subjective feedback on our cheat sheets was very positive. On a scale between -2 and 2, it shows consistently high and very high rankings for understandability, usefulness, and the other measures with rather small confidence intervals (98% Bonferroni-corrected for pairwise comparison). On average, **understandability has been rated 1.6** (Introduction=1.5, *Visual patterns*=1.5, *Anatomy*=1.9, *Pitfalls*=1.6); **aesthetics has been rated 1.8** (*Introduction*=1.8, *Visual patterns*=1.6, *Anatomy*=2.0, *Pitfalls*=1.7); **usefulness has been rated 1.4** (*Introduction*=1.8, *Visual patterns*=1.7, *Anatomy*=1.0, *Pitfalls*=1.1); **fun with 1.8** and **engaging with 1.7** had only been asked for *Introduction*. *Introduction* and *anatomy* received the highest ratings overall: *"Introduction are surprisingly clear. I felt like I understood enough after looking at them."*(P5). Participants particularly liked narration and humor in the *Introduction* (*"having a story line to follow made it much easier to understand what were the*
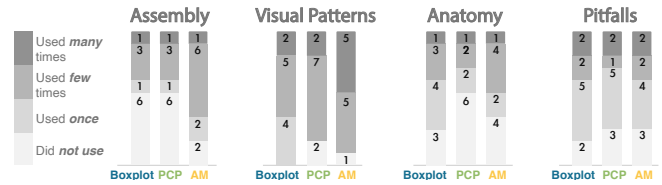

Figure 9. Frequencies of using different cheat sheets types for individual techniques during the quiz, as reported by participants.

purposes of plotting information in these ways"(P5), *"Good, clear illustrations. Humour is important as it is the application to real-life situation."*(P7). Other comments highlighted the use of *anatomy* and *visual patterns*: *"the anatomy sheets were very useful, the pattern sheets provided the most help when doing the task"*(P10).

For the individual techniques, those for *matrix* were ranked lower on average, especially the usefulness of *pitfalls* and *anatomy*. From the comments we learned that some concepts ('nodes', 'connections', 'filled-cells') were not explained well on the sheet and that matrices were generally seen as the most complicated visualization (*"The more difficult concept— adjacent matrix was least attractive but I think that is because I find the concept difficult to understand."*(P7)). Despite including specific explanations for it in the *Introduction*, participants had problems to *"determine similarity [between rows] in an adjacency matrix"*(P10).

Quiz results—Most participants scored well on the quizzes. Correctness-rates were 81% for *boxplot* (distribution of correct answers: **1** ▁▁▂▁▇ **7**), 77% for *PCP* (**1** ▁▁▂▇▃ **7**), and 75% for *matrix* (**1** ▁▃▅▃▂ **6**). For boxplots, the two questions about skewness caused frequent errors: *'More than 25% of male boxing athletes were 30 or above 30 years old.'* and *'How many groups were positively skewed?'* with 6 errors each. Although skewness was explained on the *visual patterns* sheet in both text and picture, from the think-aloud we learned that participants likely ignored the texts. As a consequence, we added larger bold-font keywords and added red color to highlight the message. This lead to an improved understanding with the last two participants. For *PCP*, the concept of correlation caused problems for some participants (*'What's the relation between dimensions B and F?', 'For values increasing at dimension A, do they increase for F as well?'.* Since we had only shown patterns for correlation and inverse correlation for neighboring axes, questions that asked for correlations between non-adjacent axes would have required a pattern and explanation we did not consider.

Cheat sheet use—From observation and think-aloud, we found that all participants used at least one of the modules during the quiz. Usage was highest for *visual patterns* which was used by 90% of the participants, followed by *pitfalls* (75%) and *anatomy* (60%) (Fig. 9). As expected, *Introduction* was least used (58%) during the quiz since participants already read through the *Introduction* prior to the quiz and had memorized most of the concepts.

Strategies of using the sheets differed somewhat. Before attempting the task, three participants had a glance but did not read the Visualization Cheat Sheets, five read part of the sheets, while three read almost all types of the sheets. When encoun-

tering a problem, a majority of participants first tried to solve the questions themselves before consulting the sheets. If they consulted the sheet, they first tried to answer their question from the graphics and only then consulting the textual annotations.

Design changes—We engaged in four major design iterations during the study, informed by participants' feedback. Changes included: highlighting data and terms in *Introduction*; using red to highlight visual elements in pictures with more than two visual hierarchies; adding keywords or annotation to clarify the message; adding terminology for both visual elements and data elements in *anatomy*. We also added variations for each pattern (*"does a cluster have to have self links?"*(P3)), which resolved issues after P3 (*"very easy and quick to find out important information that helped me answer questions"*(P5).) We found that the 'bridge' pattern on the matrix *visual patterns*contained too much information at once, so it was split across several panels, despite the additional visual complexity.

GUIDELINE WORKSHOP

To further evaluate how well our designs and cheatsheets support our goal, we decided to call for another co-design session. This time we prepared a set of design guidelines and visual instructions for creating cheat sheets (call them 'meta cheat-sheets'), to see if these guidelines help participants to create cheat sheets for their own purposes. We created general guidelines as well as guidelines specific to each sheet type with step-by-step instructions, examples, design suggestions, and practices to avoid.

We invited three visualization experts with teaching experience from academia and industry (female, all new to the study), and two participants with more casual experience (non-experts). During the 2-hour workshop, we *i)* introduced our cheat sheets (10min), then *ii)* asked participants to annotate existing cheat sheets for 3 visualisations (30min), then *iii)* discussed usefulness and possible application scenarios, and eventually *iv)* using our guidelines to construct a cheat sheet for a visualisation technique a participant was familiar with.

Similar to the study, participants' annotations and comments showed points of appreciation: *"really clear and explicit description"*(P1) as well as points where additional explanation was needed: *"mention 'Cluster' along an axis after introducing correlation"*(P2) which fed into our design process. Participants appreciated the *introduction* comics, as the stories got past perceptions that *"data is boring and visualization is difficult and I can not do it"*(P1); *"Low in complexity, but easy to understand in here [pointing to her head]"*(P5).

Using our guidelines, participants created *new* cheat sheets from scratch for *timelines*, *connected scatter plots*, *word clouds*, *bar charts* and *sankey diagrams* (see supplementary material). These sheets were coherent with our guidelines; participants organized the contents by dividers or space, and included clear graphics and meaningful annotations. Symbols (arrows, '✓', '✗', '+' and '-') are used in all of their drafts in proper cases. Some participants invented new types of cheat sheets which we will consider to explore in future work, e.g., design variations, colour encodings, and user interactions. Participants also raised open challenges related to using a limited space efficiently for

the most important information; preparing materials for other audiences than oneself; and creating *construction* sheets for abstract concepts such as word clouds.

Most importantly, participants reported that the process of creating cheat sheets helped them learn and think about a visualization technique, as the guidelines provided a structure to cover many aspects of a visualization, which evoked actively thinking (e.g., *"if I were to have a test in the next hour, on various types of data visualization, what are the likely test questions?"*(P4). Informed by discussions with the participants, we envisage the following scenarios for using cheat sheets:

- **Teaching**: visualization cheat sheets provide 'out-of-the-box' teaching materials for presentations in off-line and online courses or workshops, presented and explained on slide shows, handouts or videos (*Anatomy, Construction, Visual Pattern, Well-known relative, Pitfalls*).
- **Learning**: cheat sheets can support independent learning, e.g., the *introduction* introduces basic knowledge about a visualization with comics.
- **Analysis**: Cheat sheets can provide a rapid overview of advanced techniques for analysts to use during exploratory analysis: as print-outs at an analysts desk, or through an online website.

DISCUSSION AND FUTURE WORK

Usefulness of Cheat Sheets—Our study set out to show if cheat sheets are a possible way to support visualization literacy, and to refine our final designs and design principles. Our study indicates that we are able to create understandable cheat sheets that novice participants liked and found useful for developing understanding of complex visualisations and that creating cheat sheets helps understanding and memorizing novel techniques. By iterating with experts and novices, we have refined the protocol for creating cheat sheets to enhance clarity and relevance. Together with specific feedback, this has led to the revised guidelines summarised at the end of this section.

Limitations—Our study has two key limitations. Firstly, we cannot say *how much* cheat sheets improved real performance, or which parts were most important. There are many dimensions to any graphical work from aesthetics to conceptual structure, and resource-intensive experimental designs would be needed to truly establish the causal effect of each aspect. Our study did not have a factorial design, but instead explored multiple design decisions at once for the sake of rapid iteration and convergence, for which 11 participants were enough. Moreover, the use of cheat sheets in a lab setting is rather artificial. As a first step, we have investigated cheat sheets in isolation, but there are questions about the complementarities between cheat sheets and other formats (e.g., videos, tutorials) and how these can be brought together for particular learning pathways. However, cheat sheets are intended as concise and structured *support material* for learning, and as a lookup aid when deciphering visualizations, rather than replacements for traditional learning activities and explanations. Future studies should thus assess the use of cheat sheets as general medium for teaching in the classroom or at the office, and investigate which teaching activities they support best. Eventually, we have used simple, closed tasks, that do not require much context, so future studies

should assess how cheat sheets support higher level tasks that require exploration, comprehension and domain knowledge.

Cheat Sheets to Deal with Complexity—The visualization techniques that we are exploring here are relatively complex, both in their presentation and construction. For example, in adjacency matrix, *cliques* are conceptually simple, but the way in which graphics are constructed to make them apparent—re-ordering matrix based on similarity measures, mapping links to cells—is both conceptually and cognitively complex for novices. Participants highlighted these difficulties, but still successfully answered the quiz. Cheat sheets show promise here, as the basic concepts can be presented in a way that helps novices get to grips with the general principles, with more subtle details presented both in patterns and potential pitfalls.

There are many design decisions that have gone into our cheat sheets, and we cannot claim completeness or universality here. In particular, there are many *patterns* that could be highlighted, and their importance is often relative to a particular data context. Similarly, there are decisions to make about whether to classify an example as a *pitfall* or a *variation* (e.g. row ordering in matrices, axis ordering for PCPs). In the absence of categorically correct approaches, we hope that by highlighting both the patterns and the confusions around them, our cheat sheets can help to build up a considered approach to visual interpretation.

Cheat sheet formats—There is potentially a huge design space for cheat sheet formats to explore. Besides *Introduction* and examples in Figure 1, examples include posters, slideshows, videos, or embedded in interactive visualizations as pop-up reading-guides.

Using, Adapting, Improving, Extending—Our cheat sheets are ready to use for everyone and can be found online. They have been designed around simple, extensible templates, with the intention that there is a flexibility in the way in which they are used. The modularity of our cheat sheets makes it possible to assemble their content in new ways (Fig. 1), collecting and organizing the components to fit a particular audience and format, e.g., posters, mini cheat sheets with only vital information; explanatory notes in infographics or data comics [18]; inserted into slide shows or data videos [12].We therefor invite designers, educators, researchers, and practitioners to use our material, to create their own cheat sheets, and to suggest directions for the expansion of cheat sheets that merit exploration. *Firstly*, **more examples of patterns, pitfalls and illustrative examples** will emerge for each sheet. Through experience with deployment and learning, these can be noted and shared. With appropriate support, learners can be enlisted to explain the issues that confused them, adding a broad empirical base to the collection of issues. *Secondly*, **more visualization types** can be added, and we hope that the cheat sheet structure helps the creators of new visualizations (e.g., researchers, designers) to bring their work to a wider audience. *Thirdly*, **additional sheet types** can be added, such as explaining interactions or variants of a technique. While we considered some of these in our early designs, we found these types require more research while not being essential to explain a visualization technique and the concept of cheat sheets as in this paper. More (including non-graphical) information could be added around a visualization technique, such as examples, tools, and data types, explaining algorithms for layouts and re-ordering or covering specific usage strategies such as identifying clusters in matrix or tracing correlation in PCPs. Some of these may be specific to particular visualizations, but practise will help to draw out commonalities. *Finally*, we have worked with a general explanation of each visualization technique, but there is space for additional sheets, or particular configurations aimed at specific contexts such as medical patient data or nutrition. Similarly, high level tasks and specific analysis strategies should be considered.

Towards Guidelines for Creating Cheat Sheets—As already mentioned, there are far too many factors involved in our cheat sheets to assess the value of each individual decision. Importantly, there is a lot of value in *creating* cheat sheets by students as making cheat sheets requires a good understanding of technique, a critical eye, the structuring of knowledge, and a clear presentation. To facilitate the creation of cheat sheets, we summarize some core advice we have learned during our research and the workshops. The full set of instructional material for creating visualization cheat sheets ('meta cheat sheets') including the full list of guidelines and accompanying graphical explanations are found online.

- **Abstraction** is good to make examples general and make this generalizability understood.
- Concrete **examples** help understanding a visualization in the first place such as in our *Introduction*.
- **Visual hierarchy**, using fonts, sizes, gray backgrounds, grouping, and careful use of colors and separators helps structure content and make information retrievable.
- **Maximize graphic-text ratio** to support lookup and visual examplification; use text to give additional information.
- If necessary, **split information** across several comic panels to reduce visual complexity and aid explanation.
- **Show pattern variations** wherever possible to help understanding the main characteristic of a pattern.
- Use **complementary terminology** to help linking concepts from visual space, abstract data space, and concrete application space.

CONCLUSION

This paper introduces the concept of cheat sheets to support literacy around visualization techniques. Each cheat sheet has visual material covering the aspects that are vital for understanding a technique, such as *anatomy*, common *patterns* or *pitfalls*. We defined six types of sheets for six common, yet non-trivial visualization techniques and refined them through an iterative method involving expert interviews and workshops, readability feedback, design-by-example, a readability study and a guideline workshop. To aid creation of new cheat sheets, we present a refined set of design guidelines along with our freely available and modifiable library of cheat sheets. We hope that this material helps the community to build a common language and repository of material towards defining and develop a plan for visualization literacy for complex and non-trivial visualization techniques, and that this inspires an uptake and popularization of the many different visualization techniques.

REFERENCES

[1] 2017. Cheat Sheets for AI, Neural Networks, Machine Learning, Deep Learning & Big Data. https://www.mercuryminds.com/blog/cheat-sheets-ai-neural-networks-machine-learning-deep-learning-big-data/. (2017). Last access: 2019-03-11.

[2] 2019. Box and whisker plot. `https://www.brainingcamp.com/lessons/box-and-whisker-plots/lesson.php`. (2019). last access: 2019-03-07.

[3] 2019. Car parameters. https://stackoverflow.com/questions/19213961/parallel-coordinates-program-written-with-processing-cant-show-anything-in-mac. (2019). last access: 2019-03-12.

[4] 2019. Chartdoctor. `https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf`. (2019). last access: 2019-03-07.

[5] 2019. Data and Visualization Hazards. `http://visualizingrights.org/kit/hazards/index.html`. (2019).

[6] 2019a. DataCamp. `https://www.datacamp.com/community/data-science-cheatsheets`. (2019).

[7] 2019b. DataVizProject. `https://datavizproject.com`. (2019).

[8] 2019. Fisher's irish data. `https://en.wikipedia.org/wiki/File:ParCorFisherIris.png`. (2019). last access: 2019-03-12.

[9] 2019. Olympics athlete ages in summer 1908. `http://www.zingchart.com/gallery/chart/#!vertical-boxplot-html-tooltip`. (2019). last access: 2019-03-12.

[10] 2019. oxford dictionaries. `https://en.oxforddictionaries.com/definition`. (2019). Last access: 2019-03-12.

[11] 2019. Parallel coordinates. `https://en.wikipedia.org/wiki/Parallel_coordinates`. (2019).

[12] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. 2017. Authoring data-driven videos with dataclips. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 501–510.

[13] Benjamin Bach, Natalie Kerracher, Kyle Wm Hall, Sheelagh Carpendale, Jessie Kennedy, and Nathalie Henry Riche. 2016. Telling stories about dynamic networks with graph comics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3670–3682.

[14] Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. 2014. Visualizing dynamic networks with matrix cubes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 877–886.

[15] Benjamin Bach, Nathalie Henry Riche, Sheelagh Carpendale, and Hanspeter Pfister. 2017a. The emerging genre of data comics. *IEEE Computer Graphics and Applications* 37, 3 (2017), 6–13.

[16] Benjamin Bach, Nathalie Henry Riche, Christophe Hurter, Kim Marriott, and Tim Dwyer. 2017b. Towards unambiguous edge bundling: Investigating confluent drawings for network visualization. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 541–550.

[17] Benjamin Bach, Conglei Shi, Nicolas Heulot, Tara Madhyastha, Tom Grabowski, and Pierre Dragicevic. 2016. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 559–568.

[18] Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust, and Nathalie Henry Riche. 2018. Design patterns for data comics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 38.

[19] Ashok R Basawapatna, Kyu Han Koh, and Alexander Repenning. 2010. Using scalable game design to teach computer science from middle school to graduate school. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education*. ACM, 224–228.

[20] Scott Bateman, Regan L Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2573–2582.

[21] Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. 2016. Matrix reordering methods for table and network visualization. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 693–716.

[22] Jacques Bertin, William J Berg, and Howard Wainer. 1983. *Semiology of graphics: diagrams, networks, maps*. Vol. 1. University of Wisconsin press Madison.

[23] Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* 15, 3 (2016), 198–213.

[24] Mike Bostock. 2012. Les Misérables Co-occurrence. `https://bost.ocks/mike/miserables/`. (2012). last access: 2019-03-12.

[25] Jeremy Boy, Ronald A Rensink, Enrico Bertini, and Jean-Daniel Fekete. 2014. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1963–1972.

[26] Valerie J Bristor and Suzanne V Drake. 1994. Linking the language arts and content areas through visual technology. *THE Journal (Technological Horizons In Education)* 22, 2 (1994), 74.

[27] Alberto Cairo. 2012. *The Functional Art: An introduction to information graphics and visualization*. New Riders.

[28] Alberto Cairo. 2016. *The truthful art: Data, charts, and maps for communication*. New Riders.

[29] David Carr. 1999. Guidelines for designing information visualization applications. In *ECUE'99: 01/12/1999-03/12/1999*.

[30] Chunlei Chang, Benjamin Bach, Tim Dwyer, and Kim Marriott. 2017. Evaluating perceptually complementary views for network exploration tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1397–1407.

[31] Thomas N Dorsel and Gary W Cundiff. 1979. The cheat-sheet: Efficient coding device or indispensable crutch? *The Journal of Experimental Education* 48, 1 (1979), 39–42.

[32] Brigitte Erbe. 2007. Reducing test anxiety while increasing learning: The cheat sheet. *College teaching* 55, 3 (2007), 96–98.

[33] Stephanie D. H. Evergreen and Gavin McMahon. 2019. Data visualization chart chooser cards. `http://chartchoosercards.com/`. (2019).

[34] Matteo Farinella. 2018. The potential of comics in science communication. *Journal of science communication* 17, 1 (2018), Y01.

[35] Stephen Few. 2012. Designing Effective Tables and Graphs. (2012).

[36] Steve Franconeri. 2018. Which Visualization? A quick reference. http://experception.net. (2018). Last access: 2019-03-11.

[37] Steve Haroz, Robert Kosara, and Steven L Franconeri. 2016. The connected scatterplot for presenting paired time series. *IEEE Transactions on Visualization and Computer Graphics* 22, 9 (2016), 2174–2186.

[38] Conor Healy. 2019. data-to-viz.com. `https://www.data-to-viz.com/`. (2019).

[39] Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. 2009. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1303–1312.

[40] Nathalie Henry and Jean-Daniel Fekete. 2006. Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 677–684.

[41] Sarada Herke. 2019. Graph Theory FAQs: 03. Isomorphism Using Adjacency Matrix. `https://www.youtube.com/watch?v=UCle3Smvh1s&t=27s`. (2019).

[42] I-Han Hsiao and Claudia López. 2016. Lessons Learned from Students' Cheat Sheets: Generic Models for Designing Programming Study Guides. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 209–211.

[43] Noah Iliinsky. 2019. Visualization Design Thinking for Effective Data Visualization. `https://datascientistinsights.com/2012/12/05/design-thinking-for-effective-data-visualization/`. (2019).

[44] Gerald Everett Jones. 2011. *How to lie with charts*. LaPuerta Books and Media.

[45] Jiyoon Kim, Min Suk Chung, Hae Gwon Jang, and Beom Sun Chung. 2017. The use of educational comics in learning anatomy among multiple student groups. *Anatomical sciences education* 10, 1 (2017), 79–86.

[46] Andy Kirk. 2016. *Data visualisation: a handbook for data driven design*. Sage.

[47] Grady Klein and Alan Dabney. 2013. *The cartoon introduction to statistics*. Hill and Wang, a Division of Farrar, Straus and Giroux.

[48] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 438.

[49] Sukwon Lee, Sung-Hee Kim, Ya-Hsin Hung, Heidi Lam, Youn-ah Kang, and Ji Soo Yi. 2016. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 499–508.

[50] Makayla Lewis. 2018. Sketchnotes in UX Design cheat sheet. https://makaylalewis.files.wordpress.com/2018/09/ux-cambridge-2018-cheat-sheet.pdf. (2018). Last access: 2019-03-11.

[51] Matt Madden. 99. ways to tell a story: Exercises in style. *New York, NY: Chamberlain Bros* (99).

[52] Adam V Maltese, Joseph A Harsh, and Dubravka Svetina. 2015. Data visualization literacy: Investigating data interpretation along the novice—expert continuum. *Journal of College Science Teaching* 45, 1 (2015), 84–90.

[53] Michael C Medlock, Dennis Wixon, Mick McGee, and Dan Welsh. 2005. The rapid iterative test and evaluation method: Better products in less time. In *Cost-Justifying Usability*. Elsevier, 489–517.

[54] Mark Monmonier. 2018. *How to lie with maps*. University of Chicago Press.

[55] Randall Munroe. 2015. *Thing explainer: complicated stuff in simple words*. Hachette UK.

[56] Tamara Munzner. 2014. *Visualization analysis and design*. AK Peters/CRC Press.

[57] Anshul Vikram Pandey, Katharina Rall, Margaret L Satterthwaite, Oded Nov, and Enrico Bertini. 2015. 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*. ACM, 1469–1478.

[58] Sandra Rendgen, Julius Wiedemann, Paolo Ciuccarelli, Richard Saul Wurman, Simon Rogers, and Nigel Holmes. 2012. *Information graphics*. Taschen Cologne, Germany.

[59] Severino Ribecca. 2019. The data visualization catalogue. `https://datavizcatalogue.com`. (2019). last access: 2019-03-07.

[60] Jonathan C Roberts, Chris Headleand, and Panagiotis D Ritsos. 2016. Sketching designs using the five design-sheet methodology. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 419–428.

[61] Puripant Ruchikachorn and Klaus Mueller. 2015. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on Visualization and Computer Graphics* 21, 9 (2015), 1028–1044.

[62] Milo Schield. 2004. Information literacy, statistical literacy and data literacy. In *IASSIST QUARTERLY (IQ)*. Citeseer.

[63] Jon Schwabish. 2016. The Graphic Continuum Flash Cards. https://policyviz.com/2016/10/31/introducing-graphic-continuum-flash-cards/. (2016).

[64] Jon Schwabish. 2019. Core Principles of Data Visualization Cheatsheet. https://policyviz.com/product/core-principles-of-data-visualization-cheatsheet/. (2019).

[65] Woollcott Smith and Larry Gonick. 2005. *The cartoon guide to statistics*. Collins Reference.

[66] Yang Song and David Thuente. 2015. A quantitative case study in engineering of the efficacy of quality cheat-sheets. In *2015 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–7.

[67] Yoav Wachsman and others. 2002. Should cheat sheets be used as study aids in economics tests. *Economics Bulletin* 1, 1 (2002), 1–10.

[68] Qianwen Wang, Zhen Li, Siwei Fu, Weiwei Cui, and Huamin Qu. 2018. Narvis: Authoring Narrative Slideshows for Introducing Data Visualization Designs. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 779–788.

[69] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing Effectiveness and Engagement of Data Comics and Infographics. (2019).

[70] Hadley Wickham and Lisa Stryjewski. 2011. 40 years of boxplots. *Am. Statistician* (2011).

[71] Leland Wilkinson. 2012. The grammar of graphics. In *Handbook of Computational Statistics*. Springer, 375–414.

[72] Leland Wilkinson, Anushka Anand, and Robert Grossman. 2005. Graph-theoretic scagnostics. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. IEEE, 157–164.

[73] Zhenpeng Zhao, Rachael Marr, and Niklas Elmqvist. 2015. Data Comics: Sequential Art for Data-Driven Storytelling. *tech. report* (2015).

[74] Zhenpeng Zhao, Rachael Marr, Jason Shaffer, and Niklas Elmqvist. 2019. Understanding Partitioning and Sequence in Data-Driven Storytelling. In *International Conference on Information*. Springer, 327–338.