What Makes Data Visualizations Difficult to Understand? Not Their Dimensionality, Evidence Shows.

Gianna-Carina Grün*
TU Dortmund University

Christina Elmer†
TU Dortmund University

Marian Dörk‡
Potsdam University of Applied Sciences

Abstract

“This is too complicated” is a common criticism of journalistic data visualizations during editorial production, if the visual is not a simple chart or map. Data journalists and information designers face a conundrum: On the one hand, the data and story may require a chart that can display multiple data dimensions, such as a scatter plot or a Sankey diagram; on the other hand, practitioners want their audience to understand the chart. However, there is limited research investigating how people perceive and understand elaborate visualization types. Data from a representative online survey (n=1,041) indicate that the notion of ‘complicatedness’ is not related to the number of data dimensions shown. Visualizations showing two, three, or four dimensions of data perform equally well in users’ evaluations.

Index Terms: Data visualization, Understanding of charts, Data journalism.

1 Introduction

Vast amounts of data are generated across all sectors of daily life, and the COVID pandemic highlighted how important it is for people to make sense of this information. Data journalism makes such data accessible to a broader audience with the help of data visualizations, often using familiar formats like bar or line charts. However, the world’s increasing complexity is also reflected in its data, which in turn necessitates chart types that can represent more data dimensions than a simple line or bar chart. Interpreting radar or Marimekko charts, for example, can require a greater cognitive effort from viewers, but ideally they also offer deeper insights. A key question is how well elaborate chart types are understood and when they might prompt the reaction that it is “too complicated”.

This question touches on aspects covered by different research areas: visualization literacy, communication theory, design standards, and complexity research. All these domains shed light on when (i.e. under which conditions) and how people gain insight from common chart types [4, 1] and how to design charts well for users to understand them [5]. However, there appears to be little research on how people make sense of unfamiliar chart types [6] or why they sometimes fail to do so. With our work, we aim to evaluate if the number of represented data dimensions influences the affective and cognitive impact of data visualizations.

2 State of the Art

Previous studies on perception and sensemaking of data visualization show the importance of topic interest and emotional involvement [4] as well as familiarity with existing chart interpretation schemas [2, 6]. Some research suggests that introducing visual difficulties to data visualizations – while requiring more cognitive effort – may aid the understanding process, for example by prompting more accurate or less superficial interpretations [3]. However, most research on the understanding process of data visualizations either uses chart types with only a few data dimensions or does not specifically investigate the influence of the number of data dimensions represented. Instead, it often focuses on factors such as unfamiliarity [6], without differentiating between different types of visualizations and their levels of dimensionality.

3 Methodology

To investigate how users perceive charts representing multiple data dimensions, examples from journalistic online media were collected between January and May 2023. These examples were sourced from the entry lists of the Sigma Data Journalism Awards, the Global Investigative Journalism Network’s Top 10 DDJ of the week collection, and via the Datawrapper Dataviz Dispatch newsletter. All examples were stored in a database, and a subset was pre-selected for the survey, reviewed among all co-authors, and refined accordingly. The final survey included 18 examples in total, covering three levels of data dimensionality. The survey was tested in a pilot study with students of visualization courses (n=46) and adjusted, primarily to group related questions and to randomize the sequence of display for 2D, 3D, and 4D+ visualization examples.

The online panel was conducted by the market research company Bilendi over seven days in July 2024. Quotas were set for representativeness regarding age, gender, and education level for German media consumers. In total, 1,041 complete responses were obtained. On average, participants took 17 minutes to complete the survey and received a compensation of 2.20 euros.

4 Results

While the survey data offers a variety of insights, this poster focuses on how chart types representing multiple data dimensions are perceived, the relationship between perceived and actual insights, the influence of chart familiarity on both, and the reported obstacles.

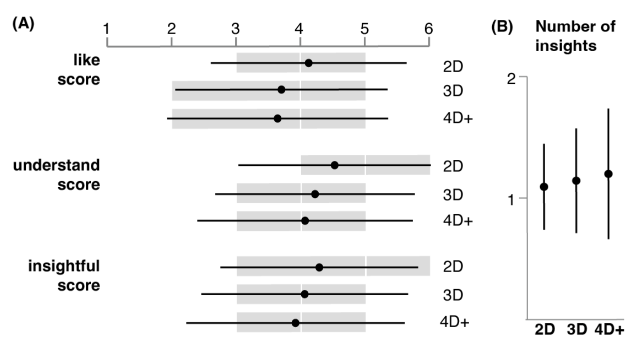

Figure 1: The circles represents the arithmetic mean, the black lines show the standard deviation. (A) Aggregated users’ perception of data visualizations (B) Aggregated number of insights users had.

Perception Each survey participant viewed three data visualization examples, one from each level of data dimensionality (2D,

*e-mail: gianna.gruen@tu-dortmund.de
†e-mail: christina.elmer@tu-dortmund.de
‡e-mail: marian.doerk@fh-potsdam.de

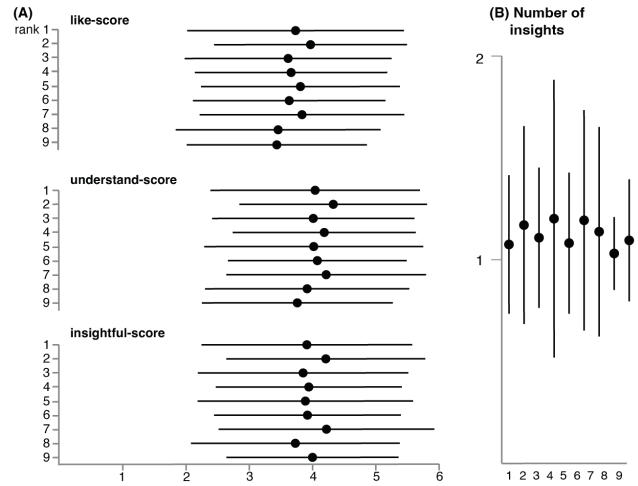

Figure 2: Influence of chart familiarity on (A) liking, understanding, and insightfulness, and (B) the number of insights produced.

3D, 4D+). They were asked to rate each example on a six-point Likert scale – one scale for whether they **like** the example (1 = not at all, 6 = very much), a second one for whether they **understand** an example, and a third for whether they find it **insightful**. As shown in Figure 1 A, visualizations with two (2D), three (3D) or four (4D+) data dimensions received similar ratings for liking, understanding, and perceived insightfulness.

Perceived vs. actual insights After rating each example, participants were asked to describe any insights they gained based on the visualization in free-text form. Responses were subsequently coded according to the framework established by Stokes et al. [7]. Overwhelmingly and across all dimensionalities, most participants reported only one insight, as shown in Figure 1 B. Visualizations with more data dimensions showed greater variability in the number of insights reported compared to those with fewer dimensions.

Chart familiarity Before evaluating the examples, participants ranked at least three out of nine different chart types according to their familiarity, placing the most familiar in the first position. Figure 2 relates these rankings to participants' ratings for liking, understanding, and perceived insightfulness, and the number of insights. The results suggest that chart familiarity does not notably influence any of these measures: charts across all familiarity ranks perform similar towards on all indicators.

Reported obstacles After viewing each visualization, assigning the liking, understanding, and insightfulness scores and noting insights, participants were asked to describe in free text any obstacles they encountered while understanding the chart or having insights. Responses were manually tagged and clustered.

Table 1: Normalized frequency of mentions of obstacle clusters per dimensionality (median number of mentions per 100 participants who saw an example of the respective data dimension)

obstacle cluster	2D	3D	4D+
no obstacles	56	55	55
design	18	23	17
understanding	8	7	11
information density	6	8	9

About half of all participants reported no obstacle whatsoever, regardless of data dimensionality, as shown in Table 1. When obstacles are reported, design-related issues are reported most frequently. Again, no significant differences emerged between visualizations with different numbers of data dimensions.

5 Discussion

Previous research has investigated the role of visual difficulties in visualizations, but in that context did not particularly look at the number of data dimensions shown [3]. This study contributes a new perspective, suggesting that the number of data dimensions – one potential source of visual difficulty – does not appear to have a significant impact.

The data shown here does not indicate that a users' familiarity with a chart type influences their perception of it or the number of insights gained. This contrasts with previous studies that highlight the role of existing graph schemas for understanding a visualization [2]. One possible explanation is that participants constructed schemas for interpreting less familiar chart types when encountering them in the survey [6], allowing them to gain insights comparable to those from familiar chart types.

Further research will be necessary to explore this hypothesis. A comparison of the representative online panel data presented here with the data analysis results of the pilot study (not shown) suggests that the outcomes may depend on the context in which participants are asked to evaluate visualizations [4].

6 Conclusion

Based on the data analyzed so far, there is no evidence that visualizations representing a higher number of data dimensions are perceived as more difficult or complex than visualizations with a lower number of data dimensions. Participants report obstacles to understanding or gaining insights with at similar rates across all levels of data dimensionality. Furthermore, participants' familiarity with the evaluated chart types did not appear to influence either their perception of the charts or the number of insights they reported.

Supplemental Materials

An overview of the selected examples can be found here.

References

[1] K. Börner, A. Bueckle, and M. Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *PNAS*, 116(6):1857–1864, 2019. 1

[2] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zackz, and J. Hullman. The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3):110–161, 2021. 1, 2

[3] J. Hullman, E. Adar, and P. Shah. Benefitting infovis with visual difficulties. *IEEE TVCG*, 17(12):2213–2222, 2011. 1, 2

[4] H. Kennedy, R. Hill, W. Allen, and A. Kirk. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday*, 21(11), 2016. 1, 2

[5] A. Kirk. *Data Visualisation: A Handbook for Data Driven Design*. SAGE Publications Ltd, Thousand Oaks, 2nd ed., 2016. 1

[6] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-A. Kang, and J. S. Yi. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE TVCG*, 22(1):499–508, 2016. 1, 2

[7] C. Stokes, V. Setlur, B. Cogley, A. Satyanarayan, and M. A. Hearst. Striking a balance: Reader takeaways and preferences when integrating text and charts. *IEEE TVCG*, 29(1):1233–1243, 2023. 2