



## Visualizations as Data Input?

Samuel Huron, Wesley Willett

### ► To cite this version:

| Samuel Huron, Wesley Willett. Visualizations as Data Input?. 2021. hal-03356575

**HAL Id: hal-03356575**

**<https://imt.hal.science/hal-03356575>**

Preprint submitted on 28 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Visualizations as Data Input?

Samuel Huron \*  
Telecom Paris - Institut  
Polytechnique de Paris  
CNRS I3

Wesley Willett †  
University of Calgary  
Calgary, AB, Canada

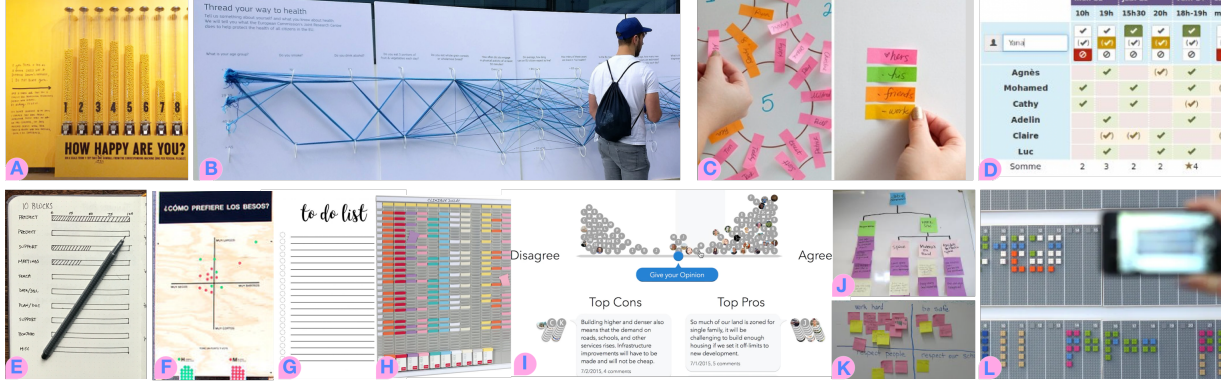


Figure 1: A small collection of input visualizations: a) reverse bar-chart by *Stefan Sagmeister*, b) parallel coordinates data strings by *Domestic Data Streamers*, c) wedding table mock-up, d) matrix poll for planning meeting from *Framadate*, e) bar chart in a bullet journal, f) Todo list, g) physical scatterplot by *Jose Duarte*, h) planning board (commercial product), i) spectrum from *consider.it*, j) tree of Post-it notes, k) scatterplot of Post-it notes, l) schedule with Lego blocks by *Vitamins*.

## ABSTRACT

We examine “input visualizations”, visual representations that are designed to collect (and represent) new data rather than encode pre-existing datasets. Information visualization is commonly used to reveal insights and stories within existing data. As a result, most contemporary visualization approaches assume existing data sets or data structures as the starting point for design, through which that data will be mapped to visual encodings to produce final visualizations. Meanwhile, the implications of visualizations as inputs and as data sources have received extremely little attention—despite the existence of visual and physical examples stretching back centuries—and the benefits, trades-offs, design patterns, and even the language necessary to describe them remain unexplored. In this paper we argue for the deeper examination of input visualizations, highlighting a set of recent examples and introducing vocabulary for describing them. Finally, we present a series of provocations which examine some of the challenges posed by input visualizations and suggest opportunities for better understanding this type of visual representations and their potential.

## 1 INTRODUCTION

Information visualization is typically thought of as a set of methods and approaches for giving visual structure to existing datasets, leveraging visual perception to enhance the analysis and interpretation of data. In most information visualization models, pipelines, and tools, data serves as the starting point for the design or analysis process, after which designers, developers, and analysts select visual mappings to make that data more legible and actionable. Over the past 50 years, a large body of research has successfully focused on optimizing visual mappings and interactions, creating a diversity

of different visualization genres tailored to unique data, tasks, audiences, and contexts. Most of these approaches implicitly assume that 1) the data (or its characteristics) are known in advance, 2) the principle goal of the visualization is to reveal trends and features in the underlying data, and 3) data interactions thought the visual structure (including filtering, computing new values, etc.) do not alter the underlying data.

Yet, a variety of visualization and visualization-like approaches exist which eschew the “data-first” orthodoxy of the academic information visualization community and instead use visualization idioms as mechanisms for data input. We define *input visualizations*, visual representations that are designed to collect (and represent) new data rather than encode pre-existing datasets. These kinds of input visualizations invert traditional data encoding and design models, using visual structures to support the collection of data, the definition of new visual schemas, and the exploration of possible visual mappings. As a result, these approaches pose problems for classical information visualization reference models [4,5], interaction taxonomies [13,36], and design guidelines [7].

In this paper we present examples of several input visualizations which illustrate the potential and challenges posed by this approach. We then reflect on the similarities and differences between these approaches and classical output-driven information visualization. Informed by these examples, we also pose a set of provocative questions to drive discussion about the role of input visualizations, their relationship to traditional visualization designs and norms, and their utility. Finally, we call on the visualization community to help us identify additional examples of these kinds of visual structures and consider their implications.

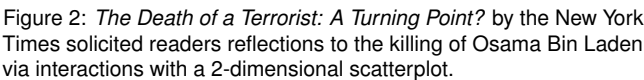
## 2 RELATED WORK

Using information visualization as a means of supporting data input and participation has been explored in multiple domains of human computer interaction, including work on civic participation [20,23,34], community engagement [12], online debate [16,30], personal reflection [32], planning [35], and polling [11]. Some of these examples, including tools like BitPlanner [35] and Thudt et

\*e-mail: samuel.huron@cybunk.com

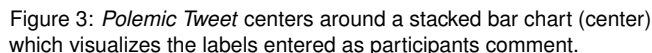
†e-mail: wesley.willett@ucalgary.ca

President Obama's announcement Sunday night about Osama bin Laden's death produced an outpouring of reaction. We asked readers the following questions: Was his death significant in our war against terror? And do you have a negative or positive view of this event? Readers — 13,864 of them — answered by plotting a response on the graph and adding a comment to explain the choice. Each light blue dot represents one comment. Darker shades represent multiple comments made on a single point. (Comments are no longer being accepted.) [Related Article »](#)



Information visualization research has also examined how visual marks can serve as interactive controls [2], including approaches like DimpVis [21] and À Table [27] in which viewers can manipulate marks to navigate between visualization views and change timespans. Similarly, “You Draw It” visualizations in which viewers articulate predictions by drawing on visualizations [19] have used input as a way of drawing attention to data values and encouraging recall. Other recent work has examined how direct manipulation interactions like changing or repositioning marks within a visualization [28] might support view transitions and visualization editing. Yet these approaches have mostly treated sketching and manipulation of visual marks ways of interacting with datasets, rather than as mechanisms for data collection.

To our knowledge, the only information visualization model to describe data input via visualizations is Jansen and Dragicevic’s *interaction model for visualizations beyond the desktop* [17]. Their model differentiates *concrete* rendering pipelines (in which existing



### 3 CASE STUDIES

### 3.1 The Death of a Terrorist: A Turning Point?

### 3.2 Polemic Tweet

<sup>1</sup><https://archive.nytimes.com/www.nytimes.com/interactive/2011/05/03/us/20110503-osama-response.html>

<sup>2</sup><https://polemictweet.com/rsln/polemicaltimeline.php>





Figure 4: Duarte's *Preguntas Sobre el Barrio* uses a variety of physical polling visualizations to solicit community input.

“??” for questions). These tweets were then displayed in a vertical list below the input box and in a stacked bar chart, both colored according to the tags they include. The vertical axis of the visualization corresponded to a time window, and bar heights showed the number of tweets emitted at that particular time slot. Unlike *The Death of a Terrorist*, *Polemic Tweet*'s input mechanisms is not spatially overlaid with the visualization. However, the two are tightly integrated and all data visualized in the *Polemic Tweet* interface was generated in that immediate context.

### 3.3 Preguntas Sobre El Barrio

In “Preguntas Sobre El Barrio” (“Questions about the Neighborhood”) (figure 4), Colombian designer Jose Duarte created multiple physical displays to engage the residents of a district and solicit opinions on issues such as transportation, security, and urban planning. These include a variety of visualization types including histograms, scatterplots, and density plots through which participants vote using physical tokens such as stickers or balls. Similar kinds of collectively-constructed visualizations and physical polls have been widely used in participatory settings (figure 1-a,b,f) and are perhaps the most straightforward examples of the input visualization paradigm, with individuals literally adding or changing visual marks to add new information.

### 3.4 Cairn

Finally, Gourlet and Dassè's Cairn (figure 5) is a tangible tabletop that enables data collection, visualization, and analysis of activity in a shared makerspace [12]. However, in comparison to most other participatory examples, Cairn leverages a much more complex visual encoding schema. Using a variety of composable physical tokens and a more complex layout, this physicalization allows makers to record detailed information about their work in the shared space—including the type, duration, and form of their projects, as well as qualitative feedback about the skills and techniques they learned. This more complex encoding captures a considerable amount of information in each visual mark, while simultaneously giving contributors opportunities for creative expression as they construct “cairns” out of multiple tokens.

### 3.5 Case study analysis

These four case studies, as well as the examples highlighted in figure 1, showcase a diversity of physical and digital input visualization approaches—including both straightforward examples and ones that challenge visualization norms, as well as the boundaries of our own definition of input visualizations.



Figure 5: The *Cairn* tabletop records information about projects created in a shared makerspace using composable physical tokens.

**Visualization as Data Schema.** In all four case studies, the visual structure is defined by the visualization designers, and serves not only to reveal patterns but also to define the data schema. These schema are established using multiple visual variables (space, color, shape, etc.) and capture varying degrees of data complexity, ranging from simple binary choices (*Preguntas Sobre El Barrio*), to continuous and connected inputs (*The Death of a Terrorist*), category labels (*Polemic Tweet*), and complex multivariate data (*Cairn*).

**Input Mechanisms.** Our case studies also illustrate a variety of different input approaches. These include digital examples that rely on both direct manipulation on the visualization [29] (*The Death of a Terrorist*) and indirect input via interface elements closely associated with the visualization (*Polemic Tweet*). In the physical case studies the inputs are all direct, involving physical construction and/or placement of unique visual marks. However, indirect input mechanism for physical input visualizations are also possible, particularly if the visualizations are mechanically actuated.

**Why input visualizations?** All four case studies use input visualization approaches to solicit opinions, sentiments, and other data from viewers. This approach, in which input visualizations are used as survey tools, is common (but not universal) across the examples we are aware of. Other common use cases include planning (figure 1-d,g,k), personal reflection (figure 1-e), and supporting reasoning tasks (figure 1-c,j).

## 4 PROVOCATIONS

In this section we present several provocations intended to initiate discussion about controversial aspects of input visualizations, including their novelty, relationship to other approaches, and utility. We hope such provocations will lead to fruitful discussions and prompt future research ideas as well as new input visualization approaches.

### 4.1 Input visualizations aren't new!

Humans have designed artifacts that encode information into visual structures for millennia, including ones that share similarities with common visualization idioms (scatter plot, bar chart, parallel coordinates, matrices, etc.). In fact, *almost all* early examples of external visual representation of information (including tally marks, tables, astronomical diagrams, etc.) are arguably input visualizations according to our definition. Everyday tools like physical and digital calendars also fit this definition—specifying a data schema by encoding periods of time using daily, weekly, and yearly grids, then allowing individuals to define new events using that structure.

Yet despite this, the vast majority of the examples of contemporary input visualizations come from outside of the information

visualization community and little research has investigated the approach. Moreover, the practical utility of input visualization approaches for polling, planning, and thinking tasks suggests that they merit further study.

## 4.2 Are these even visualizations?

One could argue that while these examples may look like visualizations, they are not. After all, most of these input visualizations don't align neatly with garden variety descriptions of visualization, which usually explain visualization approaches as "visually encoding data to make it easier to understand". (This might also have something to do with why the visualization research community hasn't talked much about them.)

Some of the visualization research community's most beloved definitions of information visualization incorporate enough generality to include input visualization approaches, but others leave less room for them. For example, Card et al.'s definition of visualization as "the use of interactive visual representations of data to amplify cognition." [4], says nothing about where the data comes from and places the most weight on the more abstract goal of amplifying cognition. On the other hand, Keim et al.'s description of information visualization as "the communication of abstract data relevant in terms of action through the use of interactive visual interfaces" [18] places an emphasis on *communication* that might seem to exclude input visualizations. Data collection also doesn't fit neatly into any of the three major goals of visualization (presentation, confirmatory analysis, and exploratory analysis) that Keim et al. identify.

One might also take the position that many of these examples *are* visualizations, but that they are trivial ones—and that considering calendars, token voting systems, or bullet journals as visualization tools is reductionist or simply not useful. However, doing so runs the risk of drawing a boundary around visualization that excludes most of the pre-industrial history of visualization, as well as a considerable slice of contemporary work on physicalization [9], infographics [3], and other visualization-adjacent topics.

## 4.3 Isn't changing the data lying?

Despite recent efforts to dislodge it, visualization research and practice is still often accompanied by an ideology of data objectivity in which raw data is seen as inherently truthful [22]. This fixation on data and visualizations of it as arbiters of truth, as well as persistent concern over the potential for "lying with data" (epitomized by the Tufte's oft-repeated *graphical integrity* principles [33] or the VIS community's own long-running VisLies series [1]) often brings with it the implication that any tool which allows end-users to manipulate the data creates opportunities for deception. This has led to outcry over tools (including versions of Excel) that allow users to directly manipulate points in graphs to alter the underlying data tables.

Yet this disdain ignores many common and well-justified reasons for modifying data values, including tuning model parameters and projections, correcting invalid values, or inputting new ones. From this perspective, input visualizations could have considerable potential to improve data entry and cleaning. After all, data wrangling *almost always* involves changing the data, and visual feedback in data entry interfaces has long been viewed as an important mechanism for improving input data quality [14].

## 4.4 Could input visualizations reduce data quality?

In some obvious situations, input visualization seem likely to make data entry easier. After all, defining two dimension in one click is certainly faster than using two sliders. However, what do we know about the potential impact of these approaches on the data gathering process, or the quality of the data they produce? In fact, one could argue that the data they produce could be biased by the presence of existing data, or that particular input visualization designs could even encourage duplicative or overly-localized data entry.

Is a visualization like "*The Death of a Terrorist?*" a poll? Maybe not. Most credible polls don't show prior results before soliciting a response. Showing prior data in advance clearly has the potential to influence how individuals express their opinions. Additionally, showing the data in advance might encourage attempts to actively pollute or manipulate data in situations where input visualizations permit input from large or anonymous audiences.

One could argue that not all input visualizations need to display their results before allowing the user to input data. However, for input physicalizations like *Preguntas Sobre El Barrio* or *Cairn*, doing so would be quite difficult. Moreover, hiding the data removes opportunities for viewers to calibrate their responses based on existing data (which might allow them to catch data errors at entry time) or use their inputs to respond rhetorically to existing data points (as in discourse-oriented visualizations like *The Death of a Terrorist*). In any case, better understanding how input visualizations might bias data entry or influence the opinions and behaviors of the people entering it may be a promising area for future research.

## 4.5 Is there even data here?

When do we consider data to *be* data? Does it need be recorded and encoded in a digital file or tabulated in a structured format? While some examples of input visualizations do indeed produce structured and easily-interpretable data, many others—including systems that rely on unstructured input, physical materials, or ambiguous encodings—may not. In these cases, the lack of any underlying data structure can mean that the data are manifest *only* in the visual artifact, which may or may not be easily measurable or reproducible. A lack of formalized data schemas or visual encodings can also mean that critical aspects of the data may exist only in the relationships between visual marks or in other intangible aspects of the visual representation and thus resist objective quantification.

For example, the two axes (negative↔positive and not-significant↔significant) used in *The Death of a Terrorist* lack absolute values or landmarks, and the significance of individual points is largely implied by their relationship to those around them. Physical installations like *Preguntas Sobre El Barrio* and *Cairn* introduce further ambiguity. For example, how should we interpret a mark that intersects both "yes" and "no" (figure 4)? What information, if any, does an elaborate and intentional token stack (figure 5) communicate if ordering isn't formalized in the instructions? Projects like Thudt et al.'s exploration of personal physicalization [32] raise similar questions, noting that "[transforming] an experience directly into a visual and physical manifestation makes it more difficult to create alternate representations later on".

Moreover, input visualizations, as mechanisms for collecting and displaying new information, collide with deeper epistemological discussions about the nature of "data" itself. Already, humanities researchers such as Latour [24] and Drucker [10] have criticized the implications of the term *data*, whose very etymology—from the Latin *datum* "(thing) given"—implies that information is somehow objective in nature and obscures the myriad biases, errors, and sources of uncertainty intrinsic to any attempt to observe or record external phenomena. Drucker advocates instead for the notion that all data is in fact "*capta*", which is actively "taken" from the world and reflects the unique tools, approaches, and biases implicit in each mode of knowledge production or inquiry. The notion of *capta* and considerations of the constructivist nature of data collection and visualization production are already important veins of discussion within the visualization community. However they become even more salient in the context of input visualizations, which explicitly surface the mechanisms of data collection.

Venturing even further down the epistemological rabbit hole, some pre-digital definitions of data, including from the Diderot and d'Alembert's 18th-century *Encyclopédie* [6], further differentiate data which are given (*data*) from "those which are unknown, and

which one seeks” (*quaesita*)<sup>3</sup>. Given this perspective, one could argue that the information captured in an input visualization are only data (or *capta*) *after* they have been input. Up until that point, including during the design of the visual representation, these future pieces of information remain *quaesita*—*sought* but not yet obtained. With these semantics in mind, perhaps we shouldn’t be discussing input visualizations at all, but rather *capta visualizations* or even *quaesita visualizations*?

## 5 CONCLUSION

As it stands, input visualizations remain a niche and underconsidered corner of the visualization universe, but one that we suspect is full of research opportunities, including:

- Developing new **conceptual models** for visualization that better accommodate input visualizations and examine the relationship between them and other existing visualization, interaction, and analysis approaches.
- More systematically examining the **design space** of input visualizations to better understand the range of possible forms and applications.
- Assessing the extent to which input visualizations impact **data quality** as well as the **experience** of people using them.
- Developing **new tools** for authoring and deploying input visualizations, particularly within existing interfaces and workflows.
- Exploring **new applications** of input visualizations, including platforms for data entry, analysis, and structured reasoning.

With this in mind, we encourage the visualization community to further examine this space—responding to these provocations, sharing new and existing input visualizations, and building an understanding of the potential of this approach, one input at a time.

## ACKNOWLEDGMENTS

Thanks to Pauline Gourlet, Jeremy Boy, and Lora Oehlberg for their helpful discussions and input.

## REFERENCES

- [1] VisLies. <https://www.vislies.org/>, 2021. Last accessed 2021-07-29.
- [2] T. Baudel. From information visualization to direct manipulation: extending a generic visualization framework for the interactive editing of large datasets. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, pp. 67–76, 2006.
- [3] L. Byrne, D. Angus, and J. Wiles. Acquired codes of meaning in data visualization and infographics: Beyond perceptual primitives. *IEEE TVCG*, 22(1):509–518, 2015.
- [4] M. Card. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.
- [5] E. H.-h. Chi and J. T. Riedl. An operator interaction framework for visualization systems. In *Proc. IEEE Symposium on Information Visualization*, pp. 63–70. IEEE, 1998.
- [6] D. Diderot and J. L. R. d’Alembert. *Encyclopédie, ou, Dictionnaire raisonné des sciences, des arts et des métiers*. University of Chicago: ARTFL Encyclopédie Project (Spring 2021 Edition).
- [7] A. Diehl, A. Abdul-Rahman, M. El-Assady, B. Bach, D. A. Keim, and M. Chen. VisGuides: A forum for discussing visualization guidelines. In *EuroVis (Short Papers)*, pp. 61–65, 2018.
- [8] P. Dragicevic and Y. Jansen. List of physical visualizations. [data-phys.org/list/tags/participatory](https://data-phys.org/list/tags/participatory), 2012. Last accessed 2021-07-18.
- [9] P. Dragicevic, Y. Jansen, and A. Vande Moere. Data Physicalization. In J. Vanderdonckt, ed., *Springer Handbook of Human Computer Interaction*, Springer Reference. Springer, 2021.
- [10] Framacloud. Framadate. [framadate](https://framadate.org/), 2012. Last accessed 2021-07-18.
- [11] <sup>3</sup><https://artflsrv03.uchicago.edu/philologic4/encyclopedie1117/navigate/5/221/>
- [12] J. Drucker. Humanities approaches to graphical display. 5.
- [13] P. Gourlet and T. Dassé. Cairn: A tangible apparatus for situated data collection, visualization and analysis. In *Proc. ACM DIS*, pp. 247–258, 2017.
- [14] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4):45–54, 2012.
- [15] J. M. Hellerstein. Quantitative data cleaning for large databases. *United Nations Economic Commission for Europe (UNECE)*, 25, 2008.
- [16] S. Huron, S. Carpendale, A. Thudt, A. Tang, and M. Mauere. Constructive visualization. In *Proc. ACM DIS*, pp. 433–442, 2014.
- [17] S. Huron, P. Isenberg, and J. D. Fekete. PolemicTweet: Video annotation and analysis through tagged tweets. In *IFIP Conference on Human-Computer Interaction*, pp. 135–152. Springer, 2013.
- [18] Y. Jansen and P. Dragicevic. An interaction model for visualizations beyond the desktop. *IEEE TVCG*, 19(12):2396–2405, 2013.
- [19] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler. Challenges in visual data analysis. In *Tenth International Conference on Information Visualisation (IV’06)*, pp. 9–16. IEEE, 2006.
- [20] Y.-S. Kim, K. Reinecke, and J. Hullman. Explaining the gap: Visualizing one’s predictions improves recall and comprehension of data. In *Proc. ACM CHI*, pp. 1375–1386, 2017.
- [21] L. Koeman, V. Kalnikaitė, and Y. Rogers. “Everyone Is Talking about It!” a distributed approach to urban voting technology and visualisations. In *Proc. ACM CHI*, pp. 3127–3136, 2015.
- [22] B. Kondo and C. Collins. Dimpvis: Exploring time-varying information visualizations by direct manipulation. *IEEE TVCG*, 20(12):2003–2012, 2014.
- [23] D. Kosminsky, J. Walny, J. Vermeulen, S. Knudsen, W. Willett, and S. Carpendale. Belief at first sight: Data visualization and the rationalization of seeing. *Information Design Journal*, 25(1):43–55, 2019.
- [24] T. Kriplean, J. Morgan, D. Freelon, A. Borning, and L. Bennett. Supporting reflective public thought with considerit. In *Proc. ACM CSCW*, pp. 265–274, 2012.
- [25] B. Latour. *Pensée retenue, pensée distribuée*. 2007.
- [26] H. Michael. On LEGO powered time-tracking; my daily column. [web.archive.org/web/20170421133649/http://jexp.de/blog/2008/08/on-lego-powered-time-tracking-my-daily-column/](http://web.archive.org/web/20170421133649/http://jexp.de/blog/2008/08/on-lego-powered-time-tracking-my-daily-column/), 2008. Last accessed 2017-04-21.
- [27] D. Offenhuber. Data by proxy—material traces as autographic visualizations. *IEEE transactions on visualization and computer graphics*, 26(1):98–108, 2019.
- [28] C. Perin, R. Vuilleumot, and J.-D. Fekete. À table! improving temporal navigation in soccer ranking tables. In *Proc. ACM CHI*, pp. 887–896, 2014.
- [29] B. Saket, S. Huron, C. Perin, and A. Endert. Investigating direct manipulation of graphical encodings as a method for user interaction. *IEEE TVCG*, 26(1):482–491, 2019.
- [30] B. Shneiderman and P. Maes. Direct manipulation vs. interface agents. *interactions*, 4(6):42–61, 1997.
- [31] H. Stiegler and M. D. de Jong. Facilitating personal deliberation online: Immediate effects of two considerit variations. *Computers in Human Behavior*, 51:461–469, 2015. doi: 10.1016/j.chb.2015.05.018
- [32] S. R. Thielke and A. Højmoose. Daily stack. [ciid.dk/education/portfolio/09/courses/tangible-user-interface/projects/daily-stack/](https://ciid.dk/education/portfolio/09/courses/tangible-user-interface/projects/daily-stack/), 2008. Last accessed 2017-04-21.
- [33] A. Thudt, U. Hinrichs, S. Huron, and S. Carpendale. Self-reflection and personal physicalization construction. In *Proc. ACM CHI*, pp. 1–13, 2018.
- [34] E. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 2001.
- [35] N. Valkanova, R. Walter, A. Vande Moere, and J. Müller. Myposition: sparking civic discourse by a public interactive poll visualization. In *Proc. ACM CSCW*, pp. 1323–1332, 2014.
- [36] Vitamins. Bit planner. [www.bit-planner.com](http://www.bit-planner.com), 2012. Last accessed 2021-07-18.
- [37] J. S. Yi, Y. ah Kang, J. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE TVCG*, 13(6):1224–1231, 2007.