

The Data-Wink 🙄 Ratio: EMOJI ENCODER for Generating Semantically-Resonant Unit Charts

Matthew Brehmer*
University of Waterloo

Vidya Setlur†
Tableau Research

Zoe‡
McGraw Hill

Michael Correll§
Northeastern University

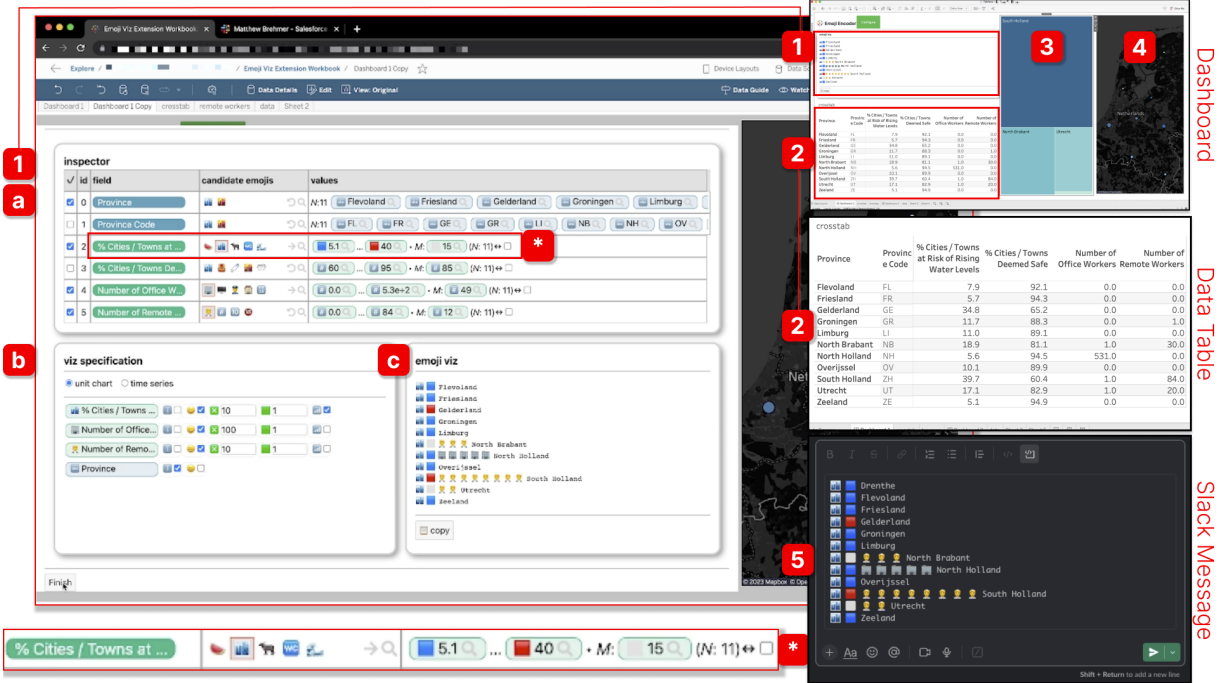


Figure 1: Beginning with the Dashboard in the top right (containing components 1 - 4), the EMOJI ENCODER is an interactive chart authoring interface (1) for Tableau that generates emoji representations based on field names and the values of categorical fields. In this example, a Data Table (2) contains flood risk values across the Netherlands along with the number and type of employees in each province. In the expanded inspector panel (1a), the author evaluates and selects from emoji recommendations for fields and values. For instance (*), 🏙️ and 🌊 are among the recommendations for the % cities / towns at risk of rising water levels field, for which the author has selected a diverging emoji scale for its values: 🟦🟤🟥. Once satisfied with an emoji palette, the author specifies (b) the type and parameters of a unit chart (c) or time series chart (Fig. 2). The completed chart can appear alongside other charts in the dashboard (3, 4) or be copied and pasted for text-based communication, such as in a Slack Message (5). See the supplemental video to see each chart construction step.

ABSTRACT

Communicating data insights in an accessible and engaging manner to a broader audience remains a significant challenge. To address this problem, we introduce the EMOJI ENCODER, a tool that generates a set of emoji recommendations for the field and category names appearing in a tabular dataset. The selected set of emoji encodings can be used to generate configurable unit charts that combine plain text and emojis as word-scale graphics. These charts can serve to contrast values across multiple quantitative fields for each row in the data or to communicate trends over time. Any resulting chart is simply a block of text characters, meaning that it can be directly copied into a text message or posted on a communica-

tion platform such as Slack or Teams. This work represents a step toward our larger goal of developing novel, fun, and succinct data storytelling experiences that engage those who do not identify as data analysts. Emoji-based unit charts can offer contextual cues related to the data at the center of a conversation on platforms where emoji-rich communication is typical.

Index Terms: Unit charts, pictograms, pictographic representation, emojis, word-scale graphics.

1 INTRODUCTION

Data storytelling can assume many forms [19], and some of these forms are more appropriate for certain media than others. For instance, magazine-style articles and data comics [3] may be better suited for a print medium, while the use of scrollytelling [14] and data videos [2] may be better suited for online journalism and broadcast media, respectively. In this paper, we address the medium of *text-based communication*, encompassing text messages, social media posts, blog posts, journalistic articles, and conversations on collaboration platforms such as Slack [23] or Teams [21]. Distinguishing aspects of this medium include the expressive use of emojis, memes, and animated GIFs to convey a certain tone, effect, or

*e-mail: mbrehmer@uwaterloo.ca

†e-mail: vsetlur@tableau.com

‡e-mail: zoezoezoe.cc@gmail.com

§e-mail: m.correll@northeastern.edu

nance that text alone cannot accomplish succinctly. Our research specifically investigates the potential of emojis for communicative data visualization, particularly due to their universality (there is currently a finite set of emojis in the Unicode standard), their broad familiarity [8], and their expressivity [12]. We further argue that data storytelling with emojis is appropriate for audiences who do not identify as data analysts and especially those who may be more receptive to fun or playful pictorial representations of data.

Constructing a bespoke emoji-based chart that reflects the underlying semantics of the data is currently possible in any emoji-compatible text editor, as exemplified by the use of emojis in social media posts to communicate proportions [24] or timelines [4]. However, this construction process can be tedious. To this end, we apply methods from the artificial intelligence community to generate a palette of semantically-resonant emojis based on the field and category names present in the tabular dataset provided by a chart author. Surfacing these recommended palettes in an interactive authoring interface as part of a Tableau dashboard extension (Fig. 1) allows for deeper integration with a chart author’s workflow. This EMOJI ENCODER interface is the primary contribution of this work. As charts generated using this interface are simply text blocks comprised of alphanumeric and emoji characters, they can easily be copied and pasted for use in any text-based communication, from social media posts and conversation platform threads to inline word-scale graphics [10] in documents or articles. This functionality allows for succinct and semantically suggestive data storytelling that is appropriate for any text-based medium.

2 BACKGROUND & RELATED WORK

Our work builds upon prior work in word-scale information graphics [10], a popular example of which being *sparklines* [22] for representing time-series data. We also take inspiration from pictorial statistics [5] and its rich history, from the use of semantically-suggestive symbols to communicate economic data in Central America prior to European colonization [6], or the 20th-century development of Isotype [11] to draw attention to economic inequality and promote group consciousness amongst the working class [7].

Apart from the manually-constructed emoji charts mentioned above [4, 24], we also draw inspiration from Gamio’s emoji-like symbol map of Chernoff Faces [9] and the Dark Sky app’s use of emojis in weather reporting [16]. Both of these instances involve predetermined mappings between quantitative values and visual representations, whereas in our approach, recommended mappings will vary given different fields names and categorical values.

Prior research in communicative data visualization has employed various NLP techniques to impart some semantic resonance with the underlying data. Techniques include the use of heuristics to populate data marks in charts with thematically-appropriate icons retrieved from the web [20] or the use of generative techniques to produce thematically-suggestive data glyphs [27] and mark textures [25, 26]. We continue this line of research by applying a neural network in the construction of emoji-based unit charts.

3 EMOJI ENCODER

EMOJI ENCODER leverages Word2Vec [15] to generate a set of emoji recommendations for the field and category names appearing in a tabular dataset. We trained the Word2Vec model on emoji descriptions from Emojipedia [8], which has curated paragraph-length descriptions for many of the 3,782 emojis currently included in the Unicode standard (v 15.1).

A chart author can provide tabular data to the EMOJI ENCODER in two ways, either by ingesting data from a CSV table, such as the annual average global temperature differential values shown in Fig. 2, or by referencing tabular data in a Tableau dashboard, such as the table shown in Fig. 1.2: flood risk values across the Netherlands, along with the number and type of employees in

each province. Once ingested, the EMOJI ENCODER authoring interface surfaces emoji recommendations in the inspector interface (Fig. 1.1.a), which includes a paginated ranked list of candidate emojis for each field, as well as the top recommended emoji for categorical field values. If the model does not provide a recommendation, the field or value is assigned a categorical placeholder emoji, and at any point in the authoring process, a recommended or placeholder emoji assignment can be manually overridden by the author via an emoji search interface. In Fig. 1.1.a, 🏠 and 🏡 are among the recommendations for the % cities / towns at risk of rising water levels field, while 🧑 is among the top recommendations for the number of remote workers field. For numerical field values, we provide a menu of ordinal emoji palettes, such as our ‘emoji-10’ scale (😞😞😞😞😞😞😞😞😞😞) used in Fig. 2 or the diverging 🟡🟠🔴 scale used in Fig. 1.1.a.i.

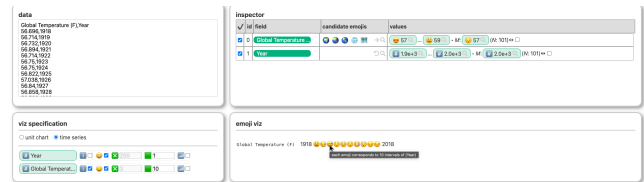


Figure 2: In contrast to the unit chart specification (c.f. Fig. 1.1.b), the EMOJI ENCODER can also be used to represent univariate time series, such as in this instance depicting global average temperature differences over ten-year intervals between 1918 and 2018 using our emoji-10 scale: 😞😞😞😞😞😞😞😞😞😞.

The selected set of emoji assignments for field names and values are automatically propagated to a chart specification panel (Fig. 1.1.b). Choices that an author makes using this panel are in turn reflected in a preview of the emoji chart in the lower right panel. The chart specification panel includes template configurations for stacked bar unit charts (Fig. 1.1.c) and univariate time series charts (Fig. 2). The former configuration includes affordances for specifying aggregation and grouping factors for numerical fields, while the latter allows for specifying a window size for the provided temporal field. Lastly, the author re-arrange the ordering of fields via drag and drop and opt to include field and value names as plain text. Once satisfied, the authoring interface can be dismissed, and the resulting chart can appear alongside other dashboard elements (Fig. 1.3 and Fig. 1.4) or be copied to the author’s clipboard as text and pasted into any text-based communication, such as a Slack message (Fig. 1.5), social media post, or document.

4 DISCUSSION & FUTURE WORK

Many chart construction interfaces assume that those using them already have a sense of how they want to represent their data [18]; in other words, they are authoring tools rather than design tools. The EMOJI ENCODER, like other recent interfaces that incorporate AI techniques to generate semantically-resonant visual encodings (e.g., [25, 26, 27]), represents a step toward closing the gap between visualization design and authoring.

Immediate next steps include conducting human-centered evaluations. First, prior reflections on the evaluation of chart authoring interfaces [1, 17] suggest that a chart reproduction or replication study could shed light on the interface’s learnability and usability. Second, the results of an open-ended chart construction activity with participants’ own data could provide us with a sense of the potential utility of our approach, particularly if complemented by retrospective interviews with participants as well as with the audiences for their charts. Finally, we can look to methodological precedents for evaluating the readability of pictographic representations [5, 11] to conduct similar experiments that specifically focus on the emoji-based charts constructed using the EMOJI ENCODER.

Another direction for future work is to consider additional recent advances in AI for generating a broader variety of emoji-rich representations: (1) Beyond unit charts for comparing quantities and sparkline-inspired emoji sequences for time series, symbol maps and set diagrams are two particularly alluring possibilities. (2) We could evaluate the utility of large language models to generate explanatory narratives around patterns observed in the data: text-based explanations that embed recommended or selected emojis throughout as word-scale graphics [10]. (3) We might use these models to generate additional ordinal emoji palettes. (4) We could replace or complement the current template-based chart specification interface (Fig. 1.1.b) with a prompt-based interface for describing and adjusting the layout of an emoji-based chart, or a multi-modal interface that replaces abstract visual elements in an existing chart image with emojis. (5) Lastly, if we relax the constraints of using existing emojis and generating entirely text-based output, we could generate palettes of novel or hybrid emoji-based icons resulting from traversals of the latent space between emojis [13].

Irrespective of whether recent generative models or prior approaches are used to construct emoji-based representations of data, the use of emojis has two additional benefits that merit discussion. First, relative to chart construction techniques that leverage text-to-image models, emoji-based charts are more likely to be stylistically consistent without significant refinement. Second, charts constructed using text-to-image models could (intentionally or unintentionally) incorporate styles, motifs, or visual elements associated with particular artists who did not consent to their works being used in this way; in contrast, all Unicode emojis are in the public domain, alleviating some potential ethical concerns when publishing or sharing an emoji-based chart.

SUPPLEMENTAL MATERIALS

We include a 2.5-minute video demonstration of the EMOJI ENCODER extension for Tableau, beginning with the table of data featured in Fig. 1.2 and concluding with the resulting pictographic unit chart pasted into a Slack message Fig. 1.5.

ACKNOWLEDGMENTS

Matthew Brehmer, Zoe, and Michael Correll contributed to this work while they were affiliated with Tableau. We thank Jason Bekkedam for his assistance with the Tableau Dashboard Extensions API, as well as Ana Crisan and Kate Mann for their feedback.

REFERENCES

- [1] F. Amini, M. Brehmer, G. Bolduan, C. Elmer, and B. Wiederkehr. Evaluating data-driven stories & storytelling tools. In N. Henry Riche, C. Hurter, N. Diakopoulos, and S. Carpendale, eds., *Data-Driven Storytelling*. A K Peters / CRC Press, 2018. tinyurl.com/amini2018.
- [2] F. Amini, N. Henry Riche, B. Lee, C. Hurter, and P. Irani. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *Proc. ACM Conf. on Human Factors in Computing Systems (CHI)*, 2015. doi: 10.1145/2702123.2702431
- [3] B. Bach, Z. Wang, M. Farinella, D. Murray-Rust, and N. Henry Riche. Design patterns for data comics. In *Proc. ACM Conf. on Human Factors in Computing Systems (CHI)*, 2018. doi: 10.1109/MCG.2017.33
- [4] J. Brogan. Kyle MacLachlan Just Brilliantly Re-created the Plot of *Dune* in Emojis, 2016. *Slate.com*. Last accessed on Aug. 14, 2024. tinyurl.com/twitterdune.
- [5] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. Designing with pictographs: Envision topics without sacrificing understanding. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 28(12), 2021. doi: 10.1109/TVCG.2021.3092680
- [6] M. del Mar Navarro. Pictorial statistics in Mexico — The Isotype Institute in 1930s México and the Sixteenth Century Huejotzinco Codex as an example of pre-Isotype pictorial statistics. , 2021. Information+Conference. vimeo.com/592184578.

- [7] H. Doudova and S. Jacobs, eds. *Image Factories: Infographics 1920-1945: Fritz Kahn, Otto Neurath et al.* Spector Books, 2018.
- [8] Emojipedia, 2024. Last accessed on Jul. 9, 2024; emojipedia.org.
- [9] L. Gamio. The Emoji States of America, 2017. Last accessed on Aug. 14, 2024. <https://www.axios.com/2017/12/15/the-emoji-states-of-america-1513302318>.
- [10] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Exploring the placement and design of word-scale visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis)*, 20(12), 2014. doi: 10.1109/TVCG.2014.2346435
- [11] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)*, 2015. doi: 10.1145/2702123.2702275
- [12] S. Khandekar, J. Higg, Y. Bian, C. Won Ryu, J. O. Talton Iii, and R. Kumar. Opico: A study of emoji-first communication in a mobile social app. In *Companion Proceedings of the World Wide Web conference (WWW)*, 2019. doi: 10.1145/3308560.3316547
- [13] Y. Liu, E. Jun, Q. Li, and J. Heer. Latent space cartography: Visual analysis of vector space embeddings. In *Computer Graphics Forum (Proc. EuroVis)*, vol. 38, 2019. doi: 0.1111/cgf.13672
- [14] S. McKenna, N. Henry Riche, B. Lee, J. Boy, and M. Meyer. Visual narrative flow: Exploring factors shaping data visualization story reading experiences. *Computer Graphics Forum (Proc. EuroVis)*, 36(3), 2017. doi: 10.1111/cgf.13195
- [15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems (NeurIPS)*, 26, 2013.
- [16] A. Pope. Interactive map tells weather forecast entirely in emojis, 2017. canadiangeographic.ca. Last accessed on Aug. 14, 2024. tinyurl.com/darkskyemoji.
- [17] D. Ren, B. Lee, M. Brehmer, and N. Henry Riche. Reflecting on the evaluation of visualization authoring systems. In *Workshop Proc. Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, 2018. doi: 10.1109/BELIV.2018.8634297
- [18] A. Satyanarayan, B. Lee, D. Ren, J. Heer, J. Stasko, J. Thompson, M. Brehmer, and Z. Liu. Critical reflections on visualization authoring systems. *IEEE TVCG*, 26(1), 2020. doi: 10.1109/TVCG.2019.2934281
- [19] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE TVCG*, 16(6), 2010. doi: 10.1109/TVCG.2010.179
- [20] V. Setlur and J. D. Mackinlay. Automatic generation of semantic icon encodings for visualizations. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)*, 2014. doi: 10.1145/2556288.2557408
- [21] Statista. Number of daily active users (DAU) of Microsoft Teams worldwide as of April 2021, 2022. Last accessed May 24, 2022. statista.com/statistics/1033742/worldwide-microsoft-teams-daily-and-monthly-users/.
- [22] E. R. Tuft. *Beautiful Evidence*. Graphics Press Cheshire, CT, 2006.
- [23] L. S. Vailshery. Slack’s number of users from February 2014 to October 2019, by paid status, 2020. Last accessed on Apr. 30, 2024. statista.com/statistics/652779/worldwide-slack-users-total-vs-paid/.
- [24] J. von Bergmann. Could not resist, Vancouver land use in 100 emojis, 2022. Last accessed on Aug. 14, 2024. x.com/vb_jens/status/1497754055271092227.
- [25] J. Wu, J. J. Y. Chung, and E. Adar. viz2viz: Prompt-driven stylized visualization generation using a diffusion model. *arXiv preprint arXiv:2304.01919*, 2023.
- [26] S. Xiao, S. Huang, Y. Lin, Y. Ye, and W. Zeng. Let the chart spark: Embedding semantic context into chart with text-to-image generative model. *IEEE Transactions on Visualization and Computer Graphics (Proc. VIS)*, 30(1), 2024. doi: 10.1109/TVCG.2023.3326913
- [27] L. Ying, X. Shu, D. Deng, Y. Yang, T. Tang, L. Yu, and Y. Wu. Metaglyph: Automatic generation of metaphoric glyph-based visualization. *IEEE Transactions on Visualization and Computer Graphics (Proc. VIS)*, 29(1), 2023. doi: 10.1109/TVCG.2022.3209447