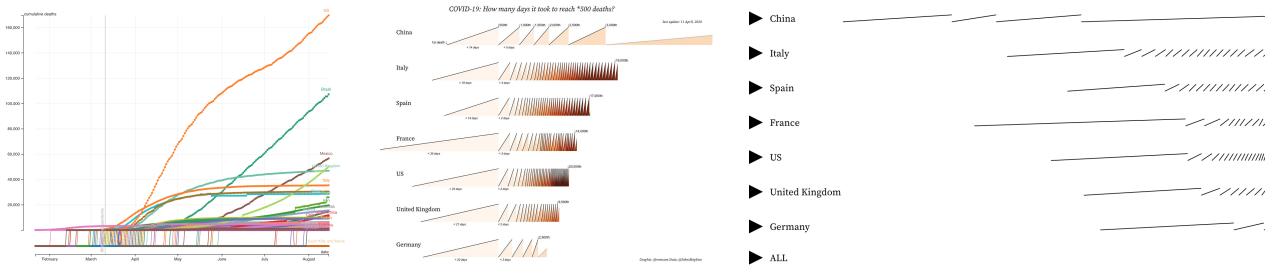
The Sawtooh Chart for Compact Cumulative Data Visualization

Romain Vuillemot

The Sawtooh Chart for Compact Cumulative Data Visualization

Romain Vuillemot*
École Centrale de Lyon, LIRIS, France



(a) (b) (c)

Figure 1: A standard cumulative chart **(a)** displays the sum of quantities over time as a line chart (here COVID-19 deaths by country between January-April 2020). We introduce a design variation for such cumulative data, the *sawtooth chart* **(b)** which splits the line chart into segments once a given value has been reached (e. g., 500 deaths) over an horizontal layout, similar to an horizon chart. As a result, a country with 6 segments reflects that 3,000 deaths have been recorded and the slopes of those segments the time needed. The color is used to re-inforce the message by also encoding the cumulative value over the time period. In this paper we also report on design variations such as black and white version **(c)**, resulting into a slope chart-like design.

Abstract

We introduce the sawtooth chart, a compact representation of cumulative data that splits a continuous line chart into segments once a given value has been reached. All segments are then aligned over an horizontal layout and their slope indicates the time to reach the given value. This design originated as an alternative to the standard COVID-19 cumulative deaths charts to convey how much time it took for countries to reach 500 deaths. The main benefit of the sawtooth chart is that it provides a compact graphics and the slope encoding as segments provide an easier way to compare trends. We report on the technical details and feedback we collected during the design process of this technique which is applicable beyond COVID-19 data and available as an open-source project.

1 Introduction

The COVID-19 death records have been the subject of scrutiny over the past months. One the most widely used representation of this dataset has been the *cumulative chart* of those deaths (Figure 1, a) which is a line chart number of deaths are added over the previous ones, over time. Cumulative line charts enabled to immediately grasp the total count of deaths, and compare countries growth rate over days, weeks or months. Many versions of cumulative charts have been released during the pandemic, but the most popular being Financial Times[1] due to is constant data and design updates. Others examples were created[2] using the Johns Hopkins' CSSE[3] dataset. Cumulative charts make it easy to grasp the high impact of the virus since the vertical slope shows the grotwh rate. However, due to the variable growth values, it is challenging to compare countries growth rates. Logarithmic scales have been used to limit the vertical spread of the chart but many concerns were raised on its understandability by the general public [6]. Many other variations of cumulative charts have been proposed to improve their reading but did not address the scale issue: e. g., with X-axis shift so that countries are aligned by 10th deaths; normalization by capita; rolling average; and using % of growth over the past 7 days rather than absolute.

We introduce the *sawtooth* chart (Figure 1, b) for the compact representation of cumulative values. Its basic idea is to split a cumulative line chart every time it reaches a certain value (e. g., 500 deaths by COVID-19), and then to draw a segment in which X-axis is the duration to reach this value and the Y-axis is from 0 to 500. Overall it results in small segments within a bounded visual interval similar to as a Horizon Chart [5]. The name stems from its visual appearance to the saw tool, but it is also a characterization of a signal in the signal process domain. Color is optional and used as a way to encode the original cumulative value. It differs however from the slope chart [7] as here the dates are not fixed, only the Y-axis points. An interactive prototype is available online[4] as an Observable notebook as an open-source project.

We argue this is an efficient technique as the encoding of the growth slope using a segment is an efficient encoding [3] for quantitative values. Also, the result of multiple vertical segments sides by sides generates a visual density encoding (similar to color) which is efficient for temporal values comparison tasks [2] where the value, in this case, is the slope which is a quantitative value. Finally, the break down by values enables *unitize* large values technique [1] so the general public pays more attention as it is easier for her to relate to smaller values.

2 Design Process

The graphics originate from our experiments by our research group with cumulative line charts especially in domains like sports where landmark values are important (e. g., visualize the time it took for a player to score 100 goals). As we found journalists are eager to report on those, we built internal exploratory tools to support this

*e-mail: romain.vuillemot@ec-lyon.fr

[1] http://ft.com/coronavirus-latest
[2] https://observablehq.com/@fil/covid-19-derived-chart
[3] https://systems.jhu.edu/

[4] https://observablehq.com/@romsson/sonification-of-covid-19-data

way of unitizing quantitative values and simplify their representation (as sometimes the message or the data quality do not require to represent all data points). For instance just visualizing the outcome of championships and not the rankings[5] as a tail scatterplot[6]. With the beginning of the pandemics, we applied those experiments to COVID-19 data.

The whole design process was progressively published as a Twitter thread[7] as a way to rapidly publish the designs and improve them based on community feedback. The sawtooth chart was created using Observable notebook using the D3 library for the data binding and redendering, and the d3-gridding toolkit for the layout [8]. It applies to any cumulative chart, however, it is restricted to monotonic curves so that it does not support a decrease of values.

The main technical challenge was to implement *cumulative binning*, to distribute the death counts into regular bins. This data processing step was required as only *aggregated values* of deaths were reported by days. And not each death once they occurred. Figure 2 illustrates cumulative data groups (A, B, C) which are the equivalent of days where data have been reported. The cumulative binning process consisted in assuming deaths were uniformly spread during days and thus progressively filling bins. Once one was fulled, its overflow was filling the following bin, and so on. We released it as an open-source module[8] independently from the chart.

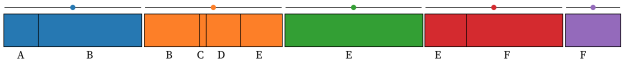


Figure 2: The cumulative binning process splits multiple categories (A, B, C) which are cumulative values (e. g., by days) into regular bins (represented as top horizontal lines). As a result, bins contain the same number of values (except probably the last one which is not filled, yet), but that may be from multiple categories/days (e. g., blue A, B) or the same day (e. g., green E).

3 Feedback and Perspectives

Our approach to constantly publish our work-in-progress design of the chart as a Twitter thread[9], which generated many discussions and enabled to significantly improve the visual aspect of the chart. A total of 8 design variations have been released as Tweets and received around 10 thousand impressions on average, with a peak of 50 thousand for the first one, along with many comments and discussions.

Overall the feedback was consistent: the visualization was considered as a novel and well suited for the application domain[10]. It was even considered to be added in a visualization toolkit[11]. We have received many requests for re-implementation in other languages (R, Python).

The main critics of the technique were related to the data. First, some countries were criticized for not reporting accurate data. Also, the *first cases* of deaths that occurred for many countries in January were mostly isolated and due to travel, so it was requested to discard them. In particular, because it created the first segment with a flat slope which was not reflecting the sudden increase of the pandemics for certain countries. Many other cumulative visualizations faced this problem and aligned countries once a death threshold was reported (e. g., 10 deaths). Surprisingly there were little concerns on the arbitrary choice of 500 deaths which was made because it was a relatively small number people could relate to and it also was not generating too many/few segments. Finally, the cumulative binning process was not criticized despite it assumed a uniform spread of death during the day. This is why the main perspective of this work is to support user interaction to dynamically change the split value similar to Interactive Horizon Graphs [4] so that the unit being used can be adjusted manually based on the user task or the selected countries. This is needed as the COVID-19 dataset is (unfortunately) constantly updating and need adjustment once values have been updated. Another perspective is using *familiar values* [1] to split the cumulative values, instead of rounded values, e. g., values such as total city or a neigboordhood population nearby[12] so the general public may be even more attentive to the visualization message.

[5] https://twitter.com/GabinRolland/status/1220721044199223297

[6] https://xeno.graphics/tail-scatter-plot/

[7] https://twitter.com/romsson/status/1243300312758259714

[8] https://observablehq.com/@fil/cumulative-binning

[9] https://twitter.com/romsson/status/1243300312758259714

[10] https://www.tableau.com/about/blog/2020/4/if-data-could-talk-visualizing-covid-19-data

[11] https://twitter.com/Elijah_Meeks/status/1248824809531666432

[12] http://www.thefunctionalart.com/2020/07/new-project-what-if-all-covid-19.html

Acknowledgments

We thank Philippe Rivière for providing feedback and implementing the cumulative binning algorithm. We also thank Twitter users for their constructive feedback. This project was partially supported by the M2I project on Urban Mobility funded by the French Agency for Durable Development (ADEME).

References

[1] F. Chevalier, R. Vuillemot, and G. Gali. Using Concrete Scales: A Practical Framework for Effective Visual Depiction of Complex Measures. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2426–2435, Dec. 2013. doi: 10.1109/TVCG.2013.210

[2] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of Alternative Glyph Designs for Time Series Data in a Small Multiple Setting. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pp. 3237–3246. ACM, New York, United States, Apr. 2013. doi: 10.1145/2470654.2466443

[3] J. Mackinlay. Automating the Design of Graphical Presentations of Relational Information. *ACM Transactions on Graphics*, 5(2):110–141, 1986. doi: 10.1145/22949.22950

[4] C. Perin, F. Vernier, and J.-D. Fekete. Interactive horizon graphs: improving the compact visualization of multiple time series. p. 3217. ACM Press, 2013. doi: 10.1145/2470654.2466441

[5] H. Reijner and P. Software. *The Development of the Horizon Graph*. 2008.

[6] A. Romano, C. Sotis, G. Dominioni, and S. Guidi. The Scale of COVID-19 Graphs Affects Understanding, Attitudes, and Policy Preferences. SSRN Scholarly Paper ID 3588511, Social Science Research Network, Rochester, NY, Apr. 2020. doi: 10.2139/ssrn.3588511

[7] E. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 2nd ed., 1983.

[8] R. Vuillemot and J. Boy. Structuring Visualization Mock-Ups at the Graphical Level by Dividing the Display Space. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):424–434, Oct. 2017. doi: 10.1109/TVCG.2017.2743998