



Journal of Data Science, Statistics, and Visualisation

MMMMMM YYYY, Volume VV, Issue II.

doi: XX.XXXXX/jdssv.v000.i00

Visual narratives of the Covid-19 pandemic

Abstract

Covid-19 has sparked a worldwide interest in understanding the dynamic evolution of a pandemic and tracking the effectiveness of preventive measures and rules. For this reason, numerous media and research groups have produced comprehensive data visualisations to illustrate the relevant trends and figures. In this paper, we will look at a selection of Covid 19 data visualisations to evaluate and discuss the currently established visualisation tools in terms of their ability to provide a communication channel both within the data science team and between data analysts, domain experts and a general interested audience. Although there is no set catalogue of evaluation criteria for data visualisations, we will try to give an overview of the different core aspects of visualisation evaluation and their competing principles.

Keywords: exploratory data visualisation, logarithmic scales, visual comparisons, R.

1. Introduction

Over the past two years, several waves of Covid-19 infections with different mutants of the SARS-CoV-2 virus have swept across the globe, claiming many lives, causing numerous health damages and affecting our personal lives in many ways. According to the WHO, over 304 million confirmed cases and over 5.4 million deaths have been reported as of early January 2022 ¹. The pandemic has generated enormous interest in epidemiological data, its analysis and visualisation. From the beginning of the pandemic, data on the number of infections and covid-related deaths has been published

¹<https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---11-january>

daily and made available to the public. Media, politicians and individuals use this data to build their narratives about the pandemic, discuss its evolution, justify the measures taken and discuss different prevention strategies against the spread of the virus. So there are several goals to be achieved by visualising Covid-19 data. These goals and their priorities were adapted as dynamically as the virus mutated and the pandemic changed pace. But as with any other data visualisation two general principles remain the same: ensuring clear vision by optimising the data-to-ink ratio (Tufte 2001) (or signal-to noise ratio) and ensuring clear understanding by organising the graphics in such a way that the story of the data is told most effectively.

In recent decades, many media organisations have established data teams that received unprecedented attention and wide-ranging opportunities ~~during the pandemic as they have been showcasing their skills and abilities, not only in visualising data, but also in explaining their data collection and data analysis strategies and methods~~¹. Data journalism will certainly be one of the beneficiaries of the Covid-19 pandemic and it has become an innovative part of news publishing, with COVID-19 delivering many excellent applications, often presented in an interactive visual format on the web, such as dashboards.²

At the same time, we still see a lot of ~~defective~~ graphics disseminated and shared: some that violate fundamental visualisation and statistical reporting principles such as accuracy, relevance, timeliness, clarity, coherence, and reproducibility. These principles have been laid out in numerous standards for statistical reporting ~~in the application areas~~, such as the ESS standard for quality reports (Commission 2020), the CONSORT, PRISMA, CHEERs guidelines, and others (see <https://equator-network.org>). Numerous publications, initiatives, and ideas to improve the communication of quantitative and statistical information have been prepared, see for example Hoffrage et al. (2000); Tufte (2001); Rosling and Zhang (2011); Otava and Mylona (2020).

The Covid 19 pandemic has amply demonstrated that ~~policies are only effectively implemented and followed by the population if they are accepted by a large majority of the population. To achieve this goal, it is essential that scientists, governments, the media, and citizens all effectively communicate in order to justify the appropriateness, usefulness and relevance of the measures. This is as true for guidelines about the use and creation of charts as it is for policies like masks and vaccines.~~

A best practice for effective communication is to present information in ~~an understandable way~~¹ (Gigerenzer et al. 2007; Gigerenzer 2011), for example by saying "one in ten" instead of 10%. Using absolute rather than relative numbers, presenting information in an appealing graphic form and summarising the most important facts in "fact boxes" are some of the methods ~~that have been developed and advocated to increase transparency in communication between stakeholders such as healthcare providers and patients (see for example: <https://www.hardingcenter.de>)~~. Such "fact boxes" combined with the dynamic layer of the internet allowed for illustrative simulations of the spread of the ~~epidemic~~, for example as presented already on 14 March 2020 in the Washington Post² ~~with the title "Why outbreaks like coronavirus spread exponentially, and how to flatten the curve". Nevertheless, both the complexity of phenomena and the "bipolarity" of statistical thinking remain a challenge. While human thinking tends to simplify patterns~~

²<https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>

and political communication also prefers a simple cause-effect relationship, real phenomena are often multivariate. Thus, in studying COVID-19 and predicting its spread, it is not ~~very~~ important to consider symptomatology, disease incidence and geographic distribution, population behaviour patterns, government policies, and impacts on the economy, schools, people in nursing homes, and society at large, but also to incorporate these into data analyses and communication of results. Associations observed in the data can often be caused by third party variables (confounders). In addition, much of the data comes from observational studies, which usually makes robust causal attribution problematic. However, statisticians who point out these limitations are at risk of having their statements pulled out one-sidedly in a polarised debate (McConway and Spiegelhalter 2021).

2. The global narrative

On 11 March 2020, WHO declared the outbreak of the novel coronavirus disease (COVID-19) a pandemic, and since that date at the latest, the global distribution of the disease has been in the public eye. The spatial spread of the virus and the resulting cases and deaths are commonly visualized by choropleth maps, see for example Figure 1 showing the total number of infections reported in each country as of January 14, 2022. For the pandemic perspective, a central element of the narrative is the ubiquity of the disease and the accompanying global impact. Choropleth maps based on raw numbers of cases might look convincing and fit to the purpose, but neglect a number of well-known caveats for statistical reporting and visualisation:

1. Absolute value unsuitability: As explained in (Monmonier 2005; Slocum et al. 2008; Speckmann and Verbeek 2010) among others, choropleth maps are fundamentally unsuitable for the representation of absolute numbers. Especially in the case of similarly coloured areas of the regions, viewers tend to integrate them unconsciously and perceive choropleths as representations of density. They also do not help to convey the desired message as the absolute numbers of Covid-19 cases are strongly influenced by the population size of the country, but also by the number of tests performed and the accuracy of the recording and reporting system.
2. The area-bias: The visual impression is determined more by the colour and the geographical area of the individual countries than by the number of Covid cases. Since the countries of the world differ extremely in area, the visual assessment is distorted, especially in the case of neighbouring countries with similar numbers but different areas.
3. Color-scheme obstructions: Much research in visualisation is concerned with the appropriate choice of colour schemes, see (Brewer et al. 1997; Müller et al. 2021). The choice of a continuous scale or a categorical scale, the choice of a scale that promotes the recognition of patterns or a scale that supports the filtering out of specific map details, influences the quality of choropleth maps.

Choropleth maps are quite commonly used by media companies and governmental organisations and it is easy to find good and bad examples of their usage. Proportional

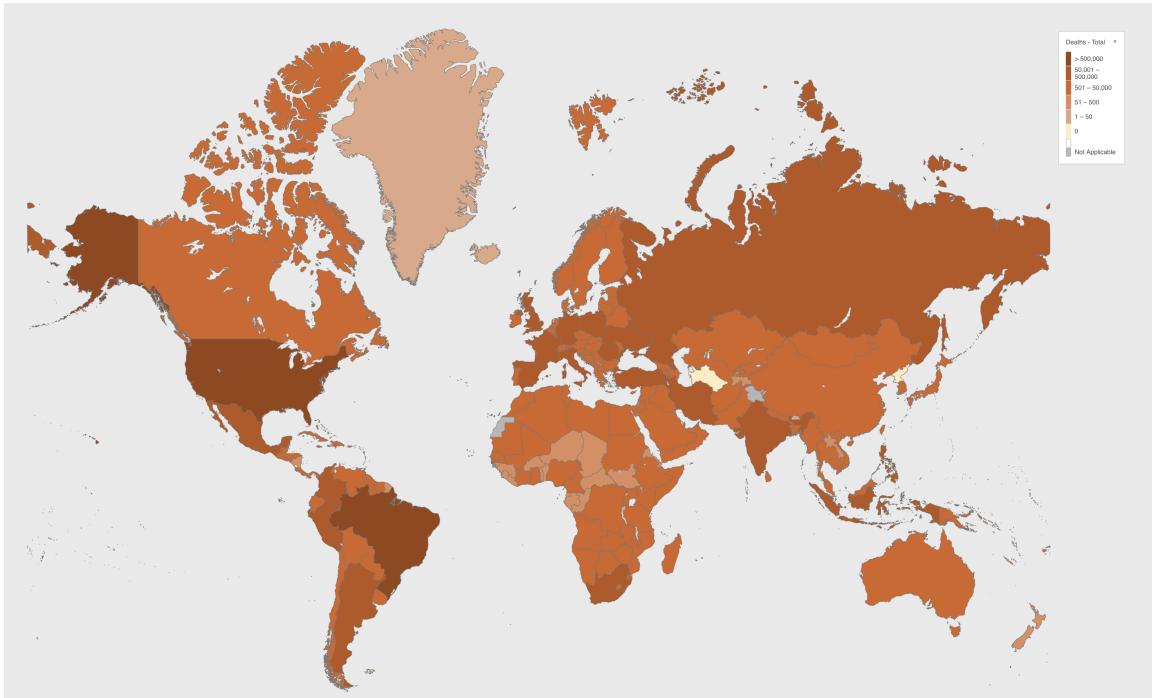


Figure 1: Choropleth map of the Covid-19 cases world-wide by country. Source: WHO
<https://covid19.who.int>

symbol maps, see Fig. 2 or graduated symbol maps place scaled symbols or diagrams directly on the input map, often on the centroid of the regions. The symbol, most commonly a disk or a square, is scaled such that its area corresponds to the data value of the region. Clearly, the visual impression of any map based illustration depends on the projection chosen for the underlying map as the examples quite often demonstrate a western-centric viewpoint.

 The narrative of the global perspective seems to be highly limited to the aspect of a pandemic affecting the entire globe. None of the above visualisations intends to provide a deeper insight into the spatial distribution of the phenomenon, neither within the administratively motivated spatial borders nor across them. The use of maps in this context more often focuses on the comparative aspects: how are we doing as opposed to our neighbors, which strategy to fight Covid-19 is better? We will examine these topics more closely in Section 4.

 Throughout the pandemic, both in choropleth maps and time-series graphs, creators have had to grapple with scale. When scales automatically adjust, mental comparisons across time can become misleading, because e.g. the colour scale from yesterday does not mean the same thing as it did today, as the range has expanded and the colours have stayed the same. This problem is more noticeable in choropleths than in time-series charts, where we are more likely to expect and notice a scale change, but it is still a fundamental issue in many different types of charts. Some publications leaned into this by arranging their graphics to highlight the fact that cases were “off the charts”, as in Figure 3, implicitly cuing readers that the situation was exceptional and worthy of their attention.

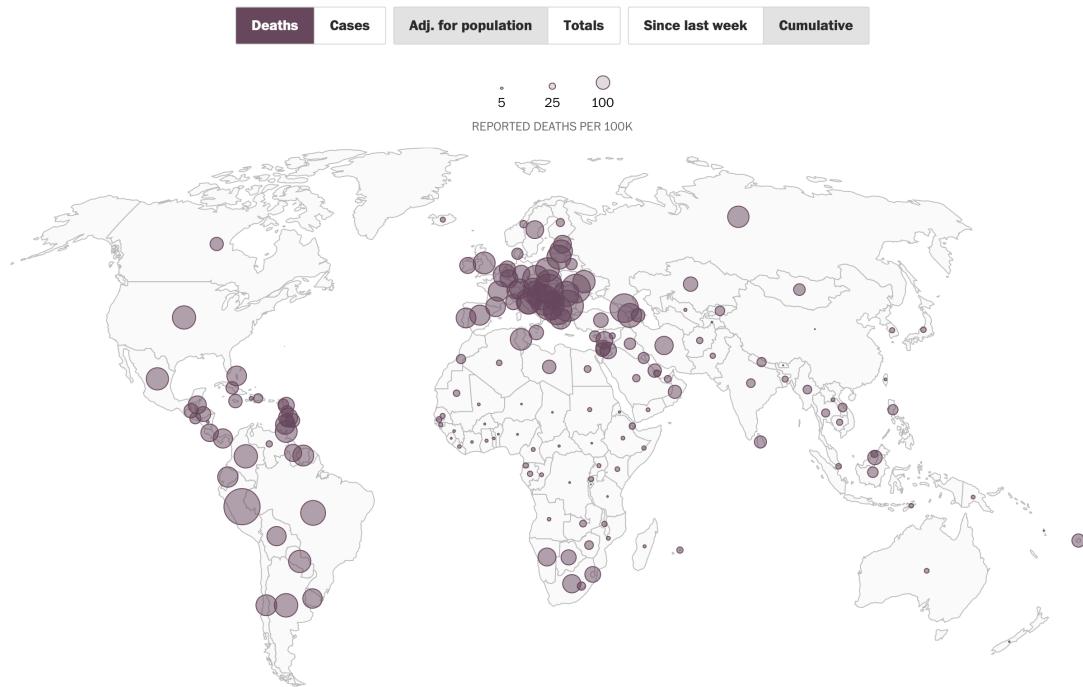


Figure 2: Proportional symbol map showing the cumulated reported Covid-19 related deaths adjusted for population world-wide by country. Source: Washington Post <https://www.washingtonpost.com/graphics/2020/world/mapping-spread-new-coronavirus/>

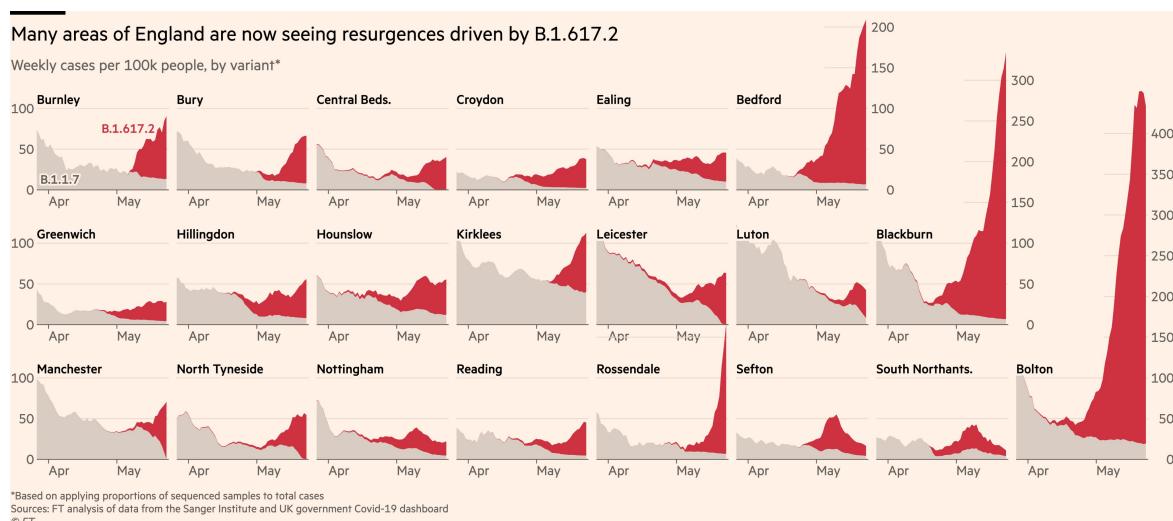


Figure 3: Chart showing the increase in cases as a result of the B.1.617.2 variant in regions of the UK. Notably, this chart arranges areas to accommodate certain regions whose cases are “off the chart” in order to emphasize the exceptional situation. Source: Tweet by John Burn-Murdoch (@jburnmurdoch), designer for the Financial Times. <https://twitter.com/jburnmurdoch/status/1397995388267810818>.

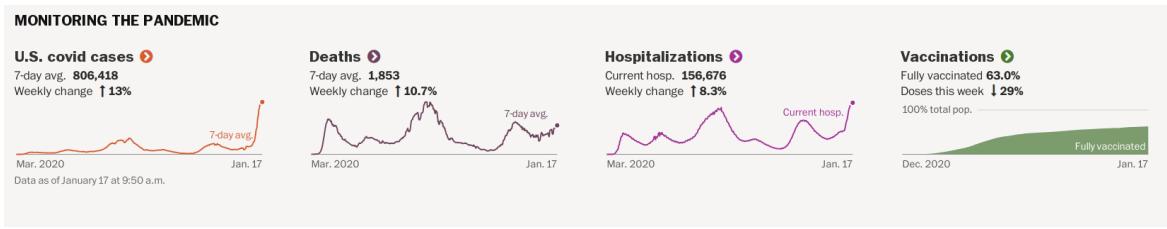


Figure 4: COVID monitoring summary charts on the Washington Post’s Coronavirus page. Source: <https://www.washingtonpost.com/coronavirus/>, 2022-01-17.

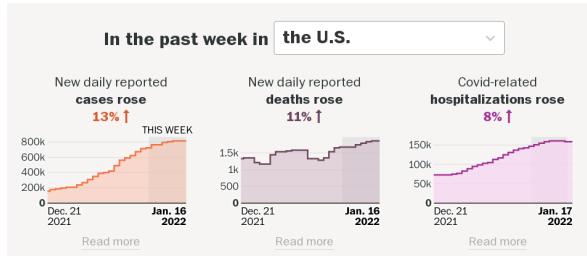


Figure 5: Some outlets provide contextual information to assist readers with interpreting time-series plots. Source: Washington Post Coronavirus Tracking, <https://www.washingtonpost.com/graphics/2020/national/coronavirus-us-cases-deaths/>.

3. The temporal narrative

Visualisations are typically created for a specific purpose, and during a pandemic it is particularly critical to show the evolution of the situation over time, as there are new developments on a daily basis. Typically, this means that cases, hospitalizations, and/or deaths are plotted on the y-axis, with date shown on the x-axis. This not only allows individuals to compare the present to past case counts, but also allows for forecasting and prediction of the future based on the current state. Time series plots are the most natural choice for these goals, and over the course of the pandemic, many different attempts have been made to show time series information using different graphical forms.

The most obvious time-series plot is also the most common: plotting cases against date, often with a 7 or 14-day smooth to handle periodicity in case reporting due to the work week, as shown in Figure 4. This allows for individuals to assess the current situation and easily compare the current number of cases to past times for which the individual has a direct reference for what to expect in impact to daily life. Often, as in Figure 4, these time series show multiple variables: cases, deaths, and hospitalizations (and sometimes vaccinations) to provide a more comprehensive view of the situation; this is particularly useful as there has been some decoupling of case numbers and hospitalizations/deaths with the emergence of effective vaccinations and strains which are more contagious but potentially also less severe. Many outlets also provide additional charts designed to provide some immediate context to readers, as shown in Figure 5.

However, there are some limitations with this presentation: if we are interested in showing the change in counts over time, viewers must assess the slope of the line,

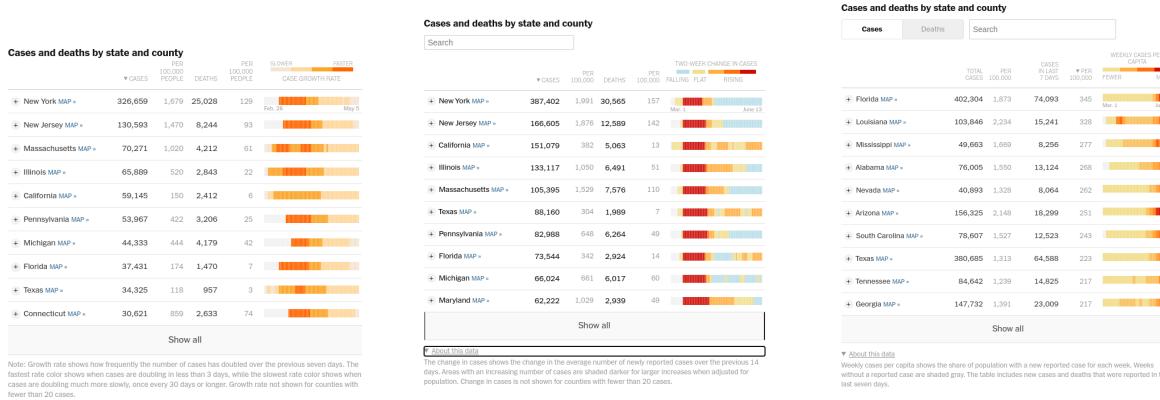


Figure 6: Sparklines style heatmaps embedded in a table of coronavirus data by state. Over the course of Summer 2020, the specific measures shown changed, even though the form of the data remained highly similar. Source: <https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html> (May 6, 2020; June 15, 2020; and July 24, 2020, respectively)

rather than its position. We know that slope judgments are much less accurate than position (Cleveland and McGill 1987), even in relatively simple situations; when we add in our ability to assess exponential growth, our perceptual accuracy is even more suspect.

Other time-series representations were more abstract, sacrificing exact data representations for a higher-level summary. It allowed viewers to make comparisons between regions and points in time without the burden of numerical calculations. Several versions from the New York Times are shown in Figure 6. The advantage of these displays is that they are very simple and allow for viewers to gain an intuitive understanding of the data, however, they do not present precise numerical information and even the color scale is fairly opaque – it would be extremely difficult to translate the visualization into any sort of accurate numerical estimate. In addition, by hiding these details from the user, it is very difficult to identify when the representation changes over time, as it did several times during the summer of 2020. The use of a similar color scheme as well as similar representations made it extremely difficult for users to identify that the value represented had changed. Clearly, the purpose of these charts is not to provide numerical precision, but rather a comparative assessment of the status of one region relative to others as well as relative to previous time points.

One unique time-series plot created a representation of the case counts in the frequency domain: instead of plotting cases as they occurred, the chart instead shows line segments from 0 to N in y , where N is a predefined number of cases; then the chart resets to show the next angle. This produces a sense of intensity of the waves of the pandemic over time as shown in Figure 7. Interestingly, this frequency representation can also be easily transferred into an audio domain, providing access for visually impaired users as well as a multimodal representation of the data for sighted users.

There are many different ways that designers can use time-series data to provide additional contextual information and facilitate comparisons. Figure 8 shows covid cases, positivity rate, and hospital admissions from South Africa's Gauteng province, which

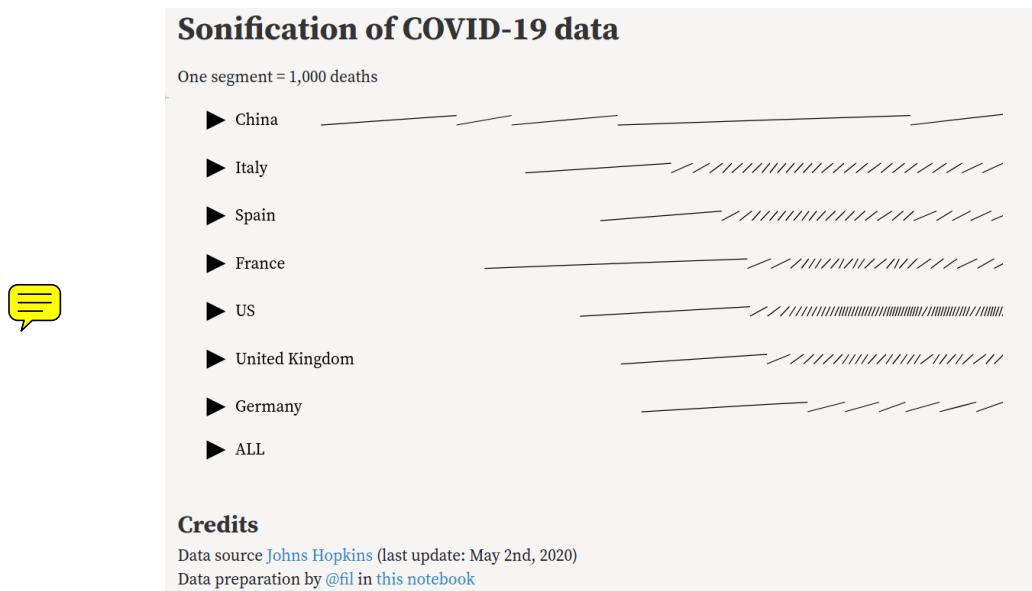


Figure 7: Frequency-domain case count graphics. Here, each 1000 deaths are represented by a single line segment whose slope represents the time taken to reach the next 1000 deaths. This provides viewers with a sense of the pace of the epidemic, rather than the raw case counts in more standard time-series representations. This representation can also be shown in the auditory domain, providing access to those who are typically excluded from visualizations due to vision loss. Source: Romain Vuillemot, <https://observablehq.com/@romsson/sonification-of-covid-19-data>.

was one of the first areas to experience a large wave of the Omicron variant. This chart provides context for the Omicron wave, showing that while infections are occurring at a much faster rate than in previous waves, hospital admissions seem to be approximately following previous trends.

While traditionally, time-series information has been limited to charts with time on the x-axis, there have been many attempts to make use of the capabilities of web graphics and animation in order to convey temporal information in less traditional time-series forms. One controversial chart appeared in the New York Times; the x-axis was wrapped around an Archimedean spiral to provide a sense of periodicity in the year-by-year evolution of case counts, as shown in Figure 9. The original form of this chart is prone to the line-width illusion (VanderPlas and Hofmann 2015), in addition to the well-known problems with polar charts (Hofmann et al. 2012; Waldner et al. 2020). The re-envisioned version¹, which aligns counts on the outside of the spiral, and displays them as radial lines, mitigates some of the issues with the line-width illusion by directly showing each line as its own entity, facilitating direct comparisons.² In addition, the re-envisioned chart includes a reference line at 100K cases/day, which allows the reader to compare the severity of different waves directly. Still, it is more work to interpret and compare peaks on this chart than on a similar linear time series chart with the same data, as the reader must assess line width and then do a mental rotation and shift operation in order to compare with any other time period. In addition, this type of chart devotes less area to previous years of the pandemic, which may decrease the amount of focus given to that data; whether this is an advantage or disadvantage

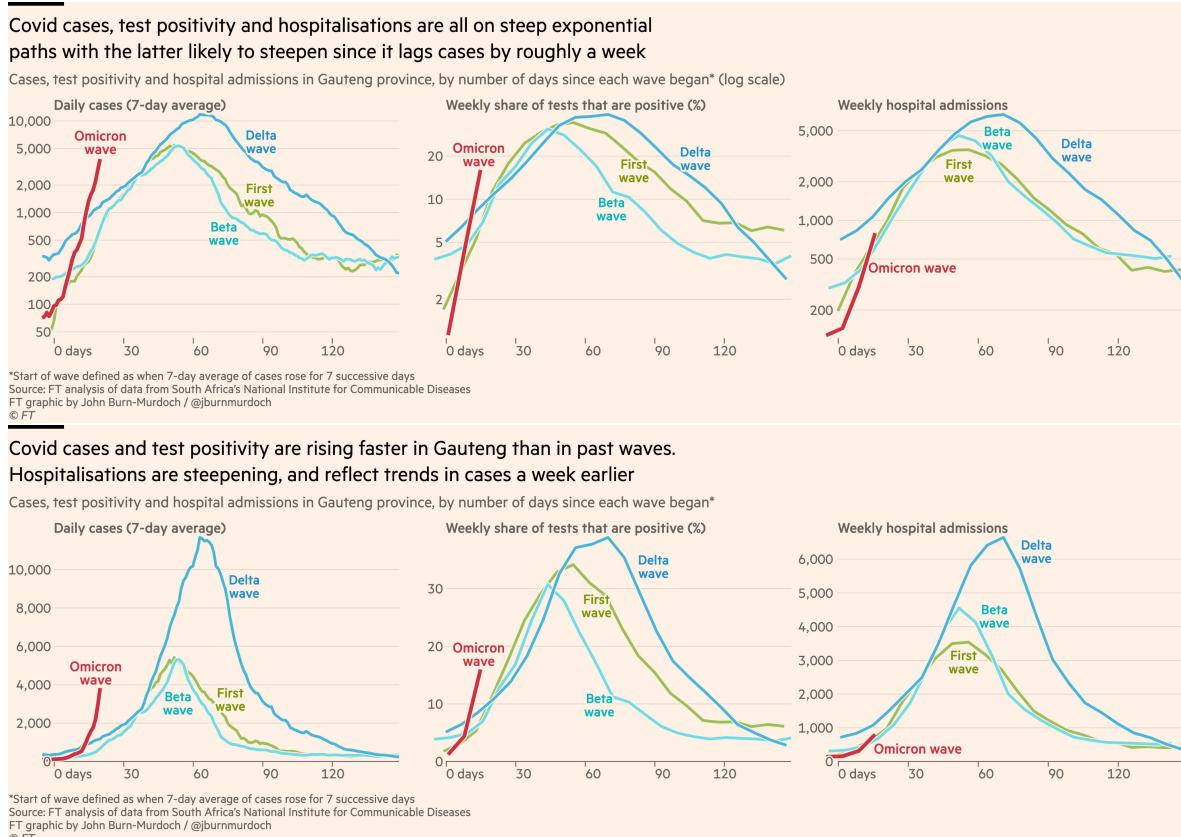


Figure 8: Covid Cases, positivity rate, and hospital admissions from South Africa's Gauteng province, on a log scale (top) and linear scale (bottom). Successive waves are overlaid, showing that the increase in cases due to the Omicron variant was sharper than any previously measured increase, while hospitalizations were much more comparable to previous waves. Source: John Burn-Murdoch, <https://twitter.com/jburnmurdoch/status/1466480113487392769>

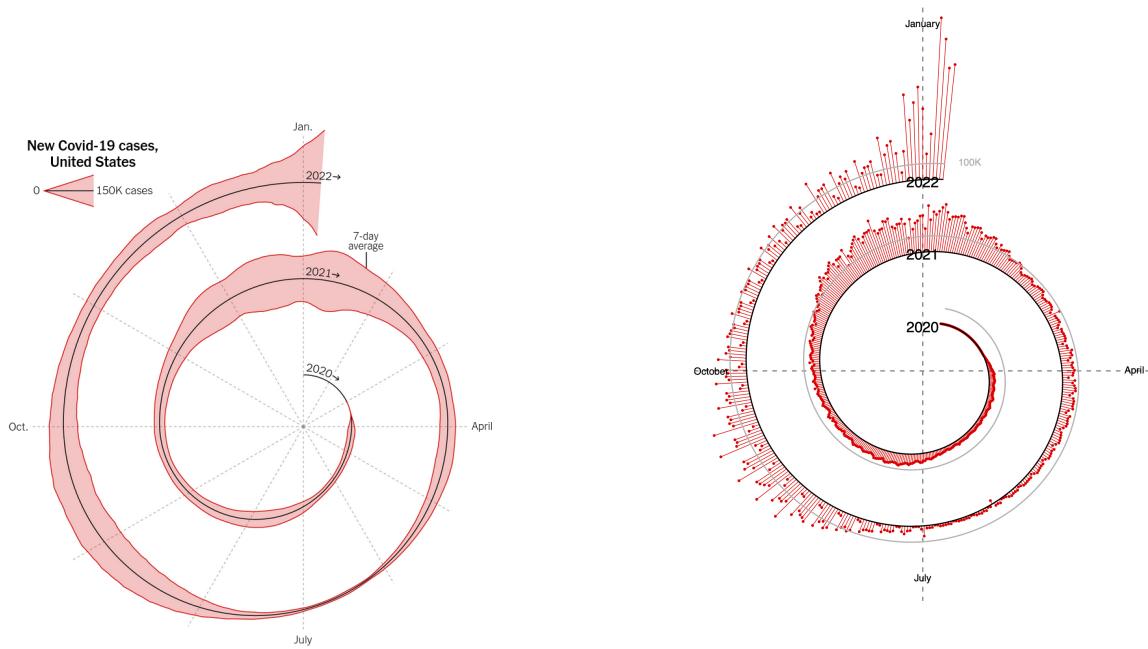


Figure 9: Covid cases in the United States, 2020-2022. The original graph from the New York Times is on the left, and a reenvisioned version which is more perceptually friendly is on the right. Source: (Left) New York Times <https://www.nytimes.com/2022/01/06/opinion/omicron-covid-us.html>, (Right) Covid Spiral, by Sourya Shrestha, https://github.com/souryashrestha/covid_spiral.

~~depends on your perspective and goals when using the chart.~~

Another attempt to creatively show time-series data that has been extremely common online throughout the pandemic is the bar chart race: an animated series of bar charts over time that show how case counts in each country are increasing. A typical example can be found on YouTube at <https://www.youtube.com/watch?v=8WH-bTH6GqI>; a screenshot is shown in Figure 10. This type of chart is effective in showing instantaneous changes in counts over days for countries with an extremely high number of cases, but may make it difficult for the viewer to process how cases are changing in countries that do not dominate case counts. By focusing viewer attention on changes in relative totals of case counts, the graphic tends to hide steady, proportional increases in cases across the globe, making it somewhat difficult to perceive large global trends relative to small changes in comparative case counts.

As people grappled with the scale of the pandemic and the lives lost as a result, a different set of data displays attempted to provide context to this loss over time. In the New York times, this took the form of a story showing the accumulation of the first 100,000 deaths due to covid, with occasional short quotes from individuals' obituaries to humanize the names(Barry et al. 2020). While this is not a typical time-series chart, as the user scrolled through the page the pace of the names increases, showing the scale of the pandemic in a visceral way.

A slightly different approach by the BBC's Visual and Data Journalism team, shown

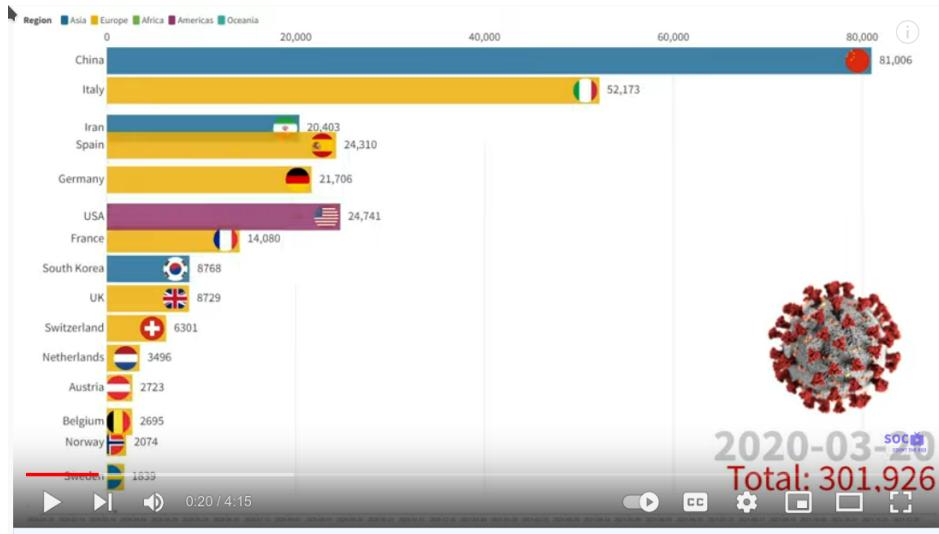


Figure 10: A screenshot from Stats on Clock’s bar chart showing counts of covid cases over time throughout the pandemic. Source: <https://www.youtube.com/watch?v=8WH-bTH6GqI>

in Figure 11, displayed the Covid cases and deaths as a growing flower, with the stem proportional to the number of cases and the flower petals showing the deaths over time(The BBC Visual and Data Journalism team 2020). This visualization is intended to evoke emotion, with the flower representation as an explicitly recognized symbol of grief; it is clearly not about showing specific case counts and deaths numerically. The visualization comes with sound as well, an attempt to make the pandemic an auditory experience as well as a visual one.

4. The ranking narrative

The plague is always the others: This is the short formula for dealing with infectious diseases from a historical perspective (Thiessen 2021). Sociologists have coined the term “othering” for such attributions of the others (Mountz 2009). This refers to the observation that above all new, unknown threats are projected onto “strangers” and “the others”. Closing the borders and restricting access to the country became a popular means of controlling the spread of the disease at several different points during the pandemic. Unsurprisingly, the prevailing visual narrative focused on comparisons often fueled by political rivalry, historical dependencies, recent withdrawal from supranational institutions or regional competitions. An ongoing debate about the true extent of the dangers of covid-19 and how best to combat it, combined with daily availability and public access to data across administrative levels, fostered ongoing competition and the use of leaderboards to show how bad case counts were somewhere else.

The bar charts used for this purpose (see top image in Fig. 12) are straightforward and provide a good way to show the “leading” countries or regions. Emphasis on specific countries is either preconfigured by the creator of the chart or can be modified interactively in the online version by check boxes and drop-down menus. These ~~bar charts~~ also allow easy adjustment of the count data to population size or other meaningful

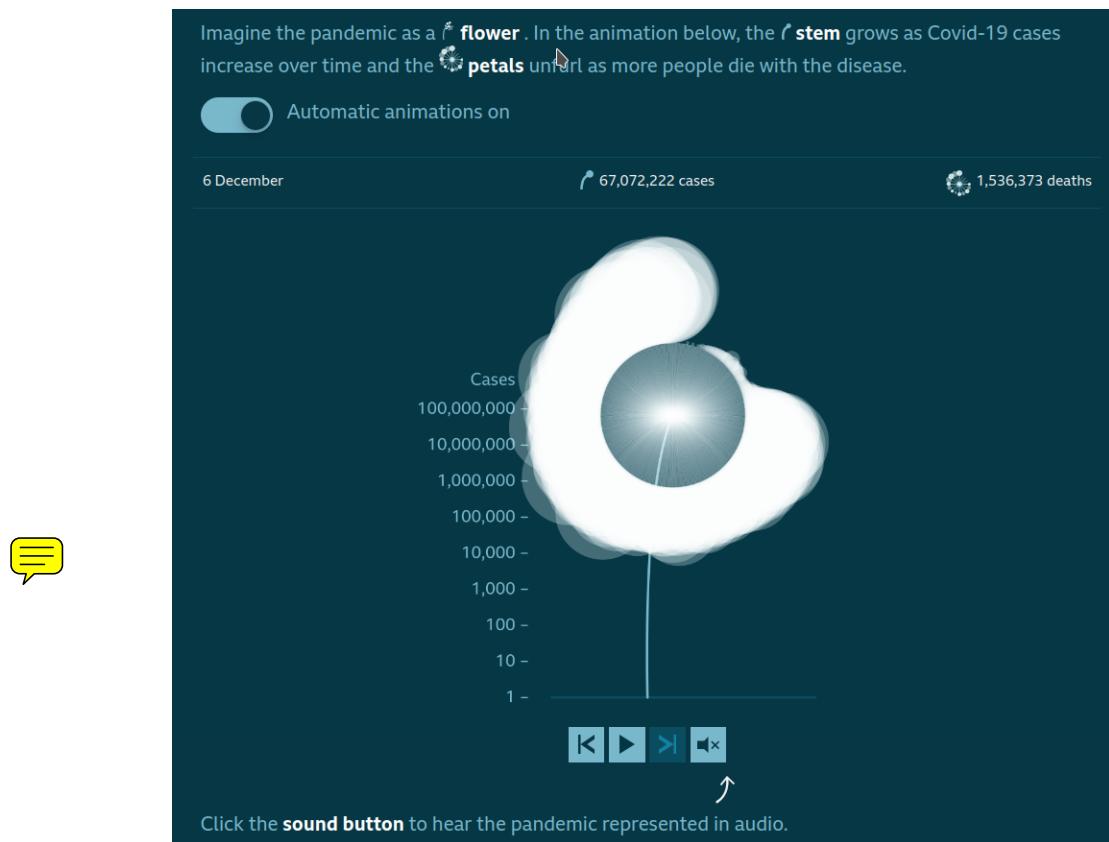


Figure 11: The BBC Visual and Data Journalism Team's Covid flower, showing the cases and deaths over time. Source: <https://www.bbc.co.uk/news/resources/idt-7464500a-6368-4029-aa41-ab94e0ee09fb>

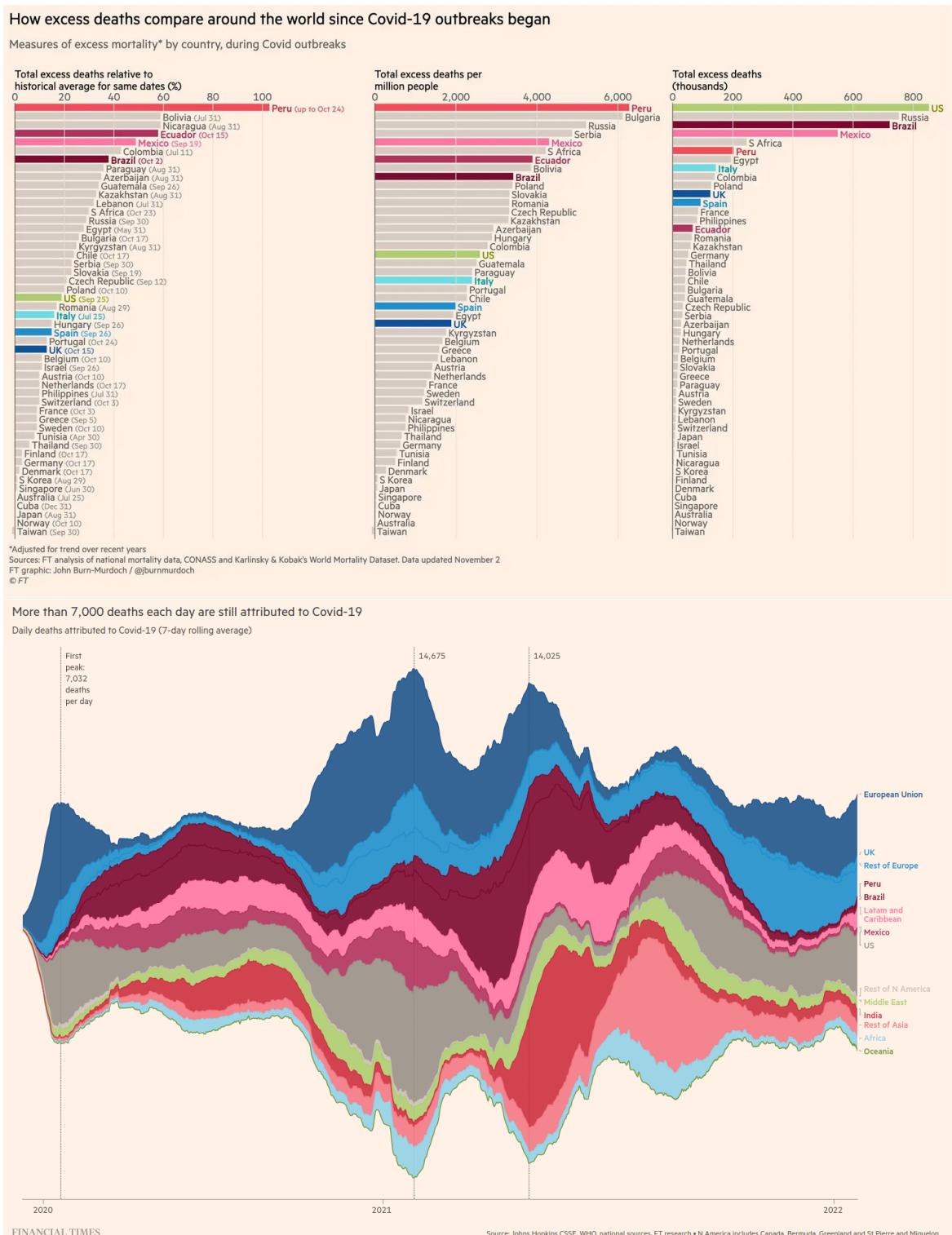


Figure 12: Ordered death rates in countries and the temporal evolution of the death toll share in major world regions. Source: Financial Times, <https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938>

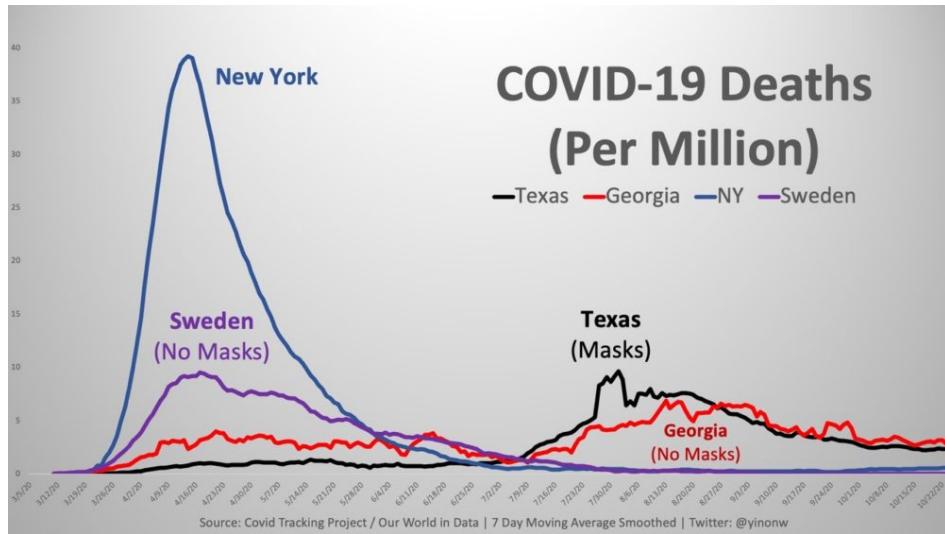


Figure 13: Deaths per million people in New York, Sweden, Texas, and Georgia, as shown in the Federalist (Weiss 2020). The goal of this comparison chart is to lead the reader to conclude that masks are ineffective, as there is not a clear difference between Texas and Georgia. New York is also shown (though not labeled as requiring or not requiring masks), and Sweden is shown as well, despite not being a US state at all, leading the critical reader to conclude that the time series shown in this chart have been cherry-picked to make a rhetorical point.

standard. Streamgraphs (see lower image in Fig. 12) are a special type of stacked area graphs and became quite popular around 2008 when they were used by the New York Times to visualise box office results of movies (Di Bartolomeo and Hu 2016). Their usefulness depends heavily on the clarity of the pattern. Various design choices such as the ordering of the different groups have a strong impact on the visual impression and message conveyed by the graph. Small variations in the proportions of the groups are almost impossible to detect, but streamgraphs give a good indication of which group is predominant at any given time, even though they are subject to the line-width or sine illusion (VanderPlas and Hofmann 2015).

Often, these comparative charts were combined with policy discussions, with individuals challenged to spot certain policy interventions on the case-count graphs of different localities with different policies (Weiss 2020). Figure 13 shows a chart from an article which appeared in The Federalist, a conservative outlet in the United States, suggesting that mask mandates are ineffective because it is difficult to spot the impact of the mask mandate on the overall trend of cases. Of course, mask mandates are but one component of a much broader pandemic management strategy, but the goal of the chart is clear: the reader is supposed to conclude that cases and masks are not associated. These charts were so common on social media that media outlets ran stories identifying them as misinformation (Reuters Staff 2020); clearly, comparative charts were effectively deployed to misinform as well as to inform.

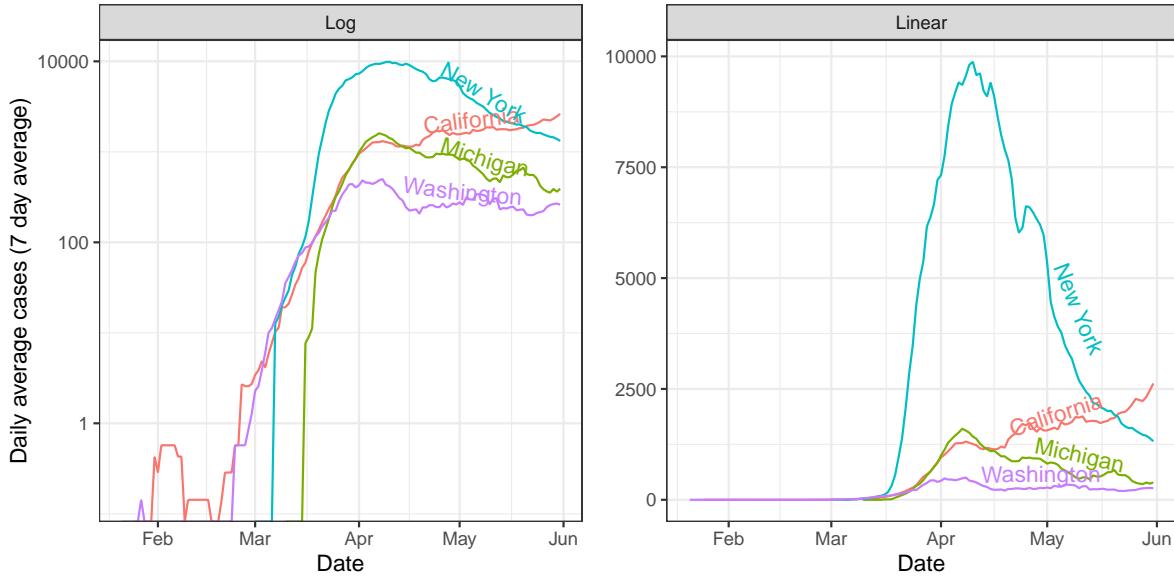


Figure 14: In the early stages of the pandemic, log scales allowed the comparison of raw case counts in locations with vastly different population and case counts.

5. To log or not to log

As COVID cases grow quasi-exponentially while there are susceptible members of the population (subject to the effectiveness of mitigation measures and testing availability), it seems natural to use log scales to allow for more effective comparisons of slight changes in case counts over time. In addition, log scales make it possible to compare regions with different populations or infection rates in the same chart. As noted previously, however, interpreting log scales requires levels of numerical sophistication that may not be appropriate for the general public. Even researchers do not always read and interpret log scales correctly(Menge et al. 2018); expecting the general public to do so is difficult under normal circumstances(Heckler et al. 2013) is difficult. When panic, fear, uncertainty, and doubt about the situation are added to the mix, it is easy to imagine that we become even worse when interpreting graphics.

One issue with assessing the use of log scales is that their effectiveness changes with ~~the stage of the pandemic and the amount (and varieties)~~ of data shown. Initially, log scales were incredibly useful at showing case counts, because minimal mitigation measures were in place and the growth of case counts (or presumptive positive cases, in absence of available testing) was fairly close to exponential. In addition, the use of log scales allowed for the comparison of nominal cases across entities with large population differences: in the US, we could compare cases in New York and California with cases in Michigan and Washington, even though the population of Michigan and Washington are much lower than the population of either New York or California.

While log scales are not necessarily intuitive, many outlets tried to make the graphs more intuitive by adding reference lines, as shown in Figure 15.

However, after the first wave of COVID, the issues with log scales became more apparent: it was difficult to detect slight increases in case counts that indicated the beginning of a new wave amid a background level of spread, as demonstrated in Figure 16. Di-

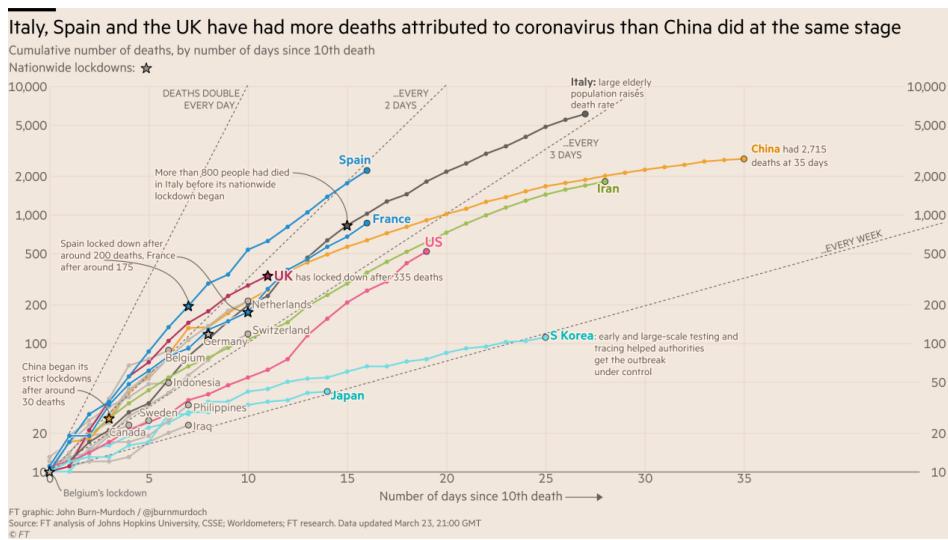


Figure 15: Reference lines to compare exponential growth rates of deaths in different countries. This provides some additional context that may help individuals use log scale data more successfully. This approach was first featured in the Financial Times, but was quickly adopted by the New York Times, 91-DIVOC, and other outlets. Graph from the Financial Times (March 23, 2020), image from [Kosara \(2020\)](#).

agonal reference lines from the origin were also less helpful, as the growth of cases or deaths was no longer approximately exponential and varied over time; for these reference lines to be effective there would need to be a clear idea of when the case counts started to increase exponentially, which is difficult to determine whilst in the thick of a potential COVID wave. Occasionally graphs with manually drawn reference lines were made available, but usually only in a retrospective manner, as in Figure 17.

While log scales have their problems, linear scales are not immune from issues either. It can be very difficult to adequately compare past situations when looking at the full time series of case counts. For example, in Figure 18, it is difficult to tell whether the first wave of COVID cases in March 2020 had an increase as fast as that in January of 2021; it is even more difficult to compare the order-of-magnitude of change in case rate growth of January 2021 relative to January 2022 when the more contagious omicron variant became prevalent.

It is not clear that the use of log or linear scales during the COVID-19 pandemic had a large effect on public opinion. Several studies were conducted in the early stages of the pandemic ([Romano et al. 2020](#); [Sevi et al. 2020](#); [Ryan and Evers 2020](#)) and results seem to suggest that while individuals have difficulty understanding log scale graphs, these issues do not tend to affect their support for intervention measures, perhaps in part because COVID-related news saturated the news and opinions were set outside of information provided in the experiments. This saturation makes it difficult to study the graphical influences while removing effects of popular opinion, political leaning, and emotional sentiment relating to the pandemic.

If we evaluate the use of log and linear scales under the more general context of exponential (or near-exponential) growth rates, however, we can gain some clarity as to their use in this particular context. We are abysmal at forecasting exponential growth, vastly

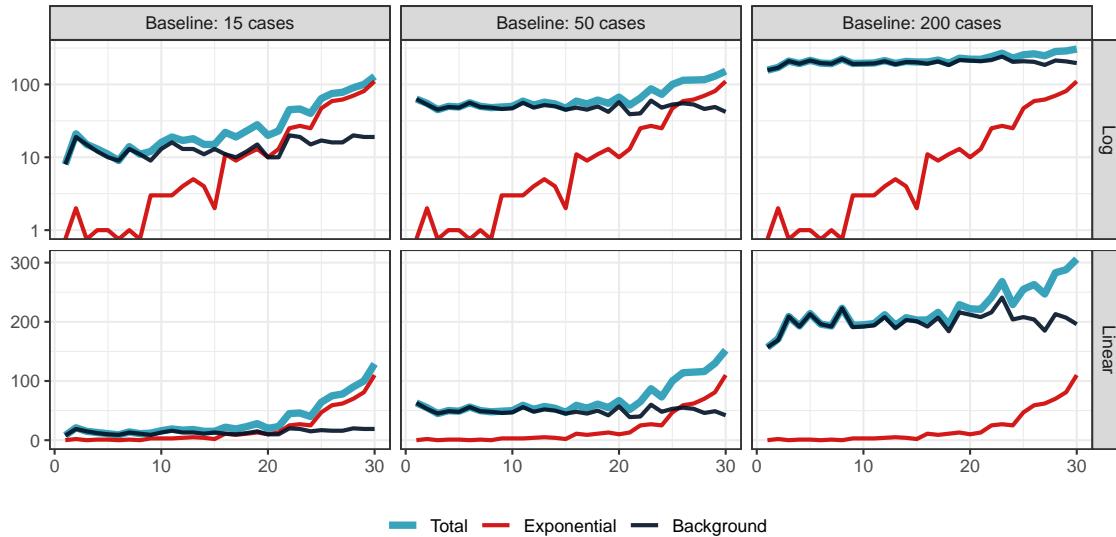


Figure 16: One problem with log scales is that if there is a background level of spread, it can be hard to notice the introduction of an additional source of exponential spread. Linear scales do not have this problem - the exponential source is noticeable very quickly in the total line, but on the log scale it is much harder to discern when the exponential source causes the total line to diverge from the background. In the top-right corner, it is difficult to identify that there is an exponential increase in cases amid the baseline, even though the exponential source makes up approximately 50% of the cases at the end of the time period shown.

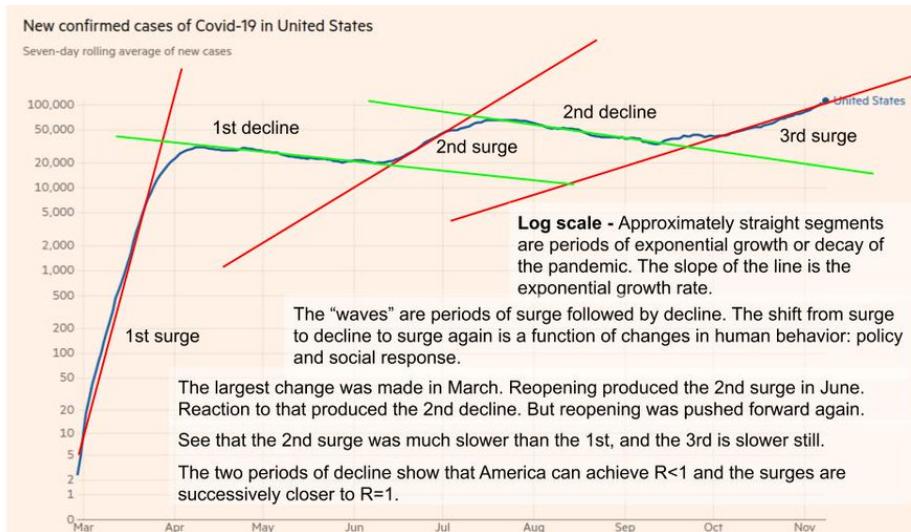


Figure 17: An annotated time series from the Financial Times which appeared on Twitter on November 12, 2020. This chart has annotations which show the decreasing R_{eff} in successive peaks of the pandemic, with periods of $R_{eff} < 1$ between surges in cases. These annotations assist the reader with drawing complex conclusions from the chart, but are difficult to automate and thus tend to be manually curated. Source: Mark Gubrud (@mgubrud), <https://twitter.com/mgubrud/status/1326784082399944704>.

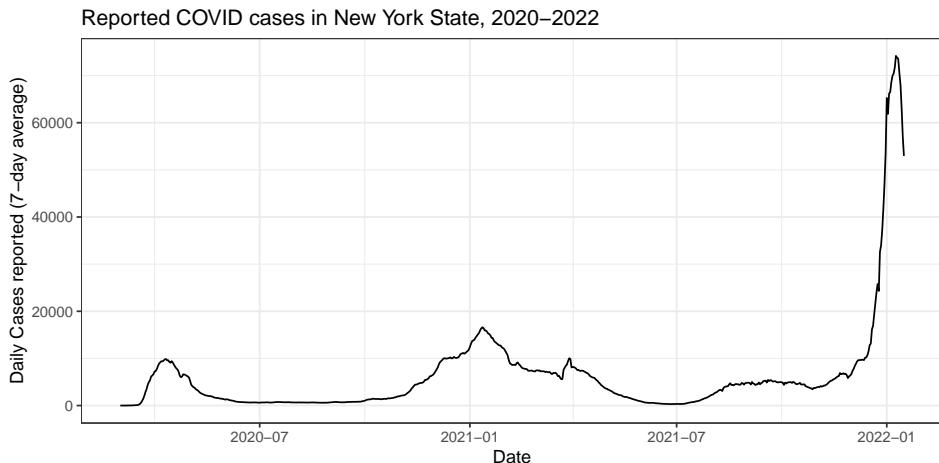


Figure 18: Reported COVID cases in New York State, 2020-2022. The linear scale makes it difficult to compare the trajectory of different waves to determine how severe the current status is relative to the past, because the primary contrast is the height of the relative peaks, rather than the growth *rate*. A similar graph on the log scale would have the peaks at much more similar heights (though there would still be a difference), allowing the reader to focus on other information, such as the slope of the relative lines.

under-estimating future growth by using linear or quadratic approximations ([Wagenaar and Timmers 1978](#); [Lawrence and O'Connor 1992](#); [Timmers and Wagenaar 1977](#)). If the goal of a chart of COVID case counts is to allow individuals to forecast the trend and make decisions accordingly, then using a log scale (with all of its pitfalls in understanding) may be the best option, as it at least replaces the need to forecast along an exponential curve with the need to forecast along a straight line. However, it is not clear that individuals can transfer that prediction back to a linear scale. If the goal of presenting a chart of case counts is to tell the story of what has happened (rather than supporting future decision-making), then it is undoubtedly better to present the data on a more familiar linear scale that requires less cognitive load on the part of the viewer.

Unfortunately, there are not a lot of good options for supporting forecasting decisions in mathematically unsophisticated viewers. Guide lines, like those used in the Financial Times, were helpful in the initial phases of the pandemic, but no outlets that we are aware of shifted the lines to show the growth rate of the current peak (as opposed to the initial peak), and due to the issues shown in Figure 16, these guide lines may not have been all that successful in any case. Those hoping to create successful time-series charts that allow forecasting are thus left with two options: deal with our inability to forecast exponential growth, or deal with most individuals' inability to translate log scales into practical reality.

6. Summary and discussion

Throughout the pandemic (thus far), visual representations of data have been an integral part of scientific communication. While not always optimal, these graphics have attempted to provide meaningful context, to encourage individual and collective ac-

tion, and to help individuals grapple with the scale of the pandemic in cases, deaths, vaccinations, and interventions. As in any developing situation, choices made at some points in the pandemic did not always persist - many news outlets refined their charts and approach to the design of graphics over the course of the pandemic. This evolution of graphical forms has provided us with an opportunity to evaluate what worked and did not – and why – in a context of broad general interest.

The visual narratives of the pandemic which have persisted over time are due in part to the rise of data journalism and interactive graphics platforms which ensure that every outlet has the ability to host engaging and visually appealing graphics; however, while these graphics are nice to look at, not all are equally functional or useful for communication purposes. As in any situation, good tools can be used and mis-used freely; we have seen charts and graphics used to perpetuate misinformation and misleading claims, even though it is far more common for charts and graphs to be used for good - to educate and inform the public.

The pandemic and graphics used to show the ebb and flow of COVID cases highlight the challenges of data visualization in the modern era: even with numerous guidelines and standards for graphics, it is not straightforward to communicate about complex data, and creating charts which fully represent the intended message or messages is a challenge. It is essential that we continue to study how graphics are perceived and interpreted, supporting the guidelines and standards which exist with user studies that ground these rules in science.

Computational Details

Graphics included in this paper not sourced from media outlets were created using data published by the New York Times at <https://github.com/nytimes/covid-19-data/>. Graphics were created using ggplot2(Wickham 2016) and R 4.1.2. R itself and all packages used are available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/>.

Acknowledgments

First ideas have been presented at DSSV 2020. We thank our colleagues and the editors for valuable feedback and encouragement to complete this manuscript.

References

- Barry, D., Buchanan, L., Cargill, C., Daniel, A., Delaquérière, A., Gamio, L., Gianordoli, G., Harris, R., Harvey, B., Haskins, J., Huang, J., Landon, S., Love, J., Maalouf, G., Matthews, A., Mohamed, F., Moity, S., Royal, D.-C., Ruby, M., and Weingart, E. (2020). Remembering the 100,000 Lives Lost to Coronavirus in America. *The New York Times*, ISSN: 0362-4331, <https://www.nytimes.com/interactive/2020/05/24/us/us-coronavirus-deaths-100000.html>.
- Brewer, C. A., MacEachren, A. M., Pickle, L. W., and Herrmann, D. (1997). Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers*, 87(3):411–438, ISSN: 00045608, 14678306, <http://www.jstor.org/stable/2564061>.
- Cleveland, W. S. and McGill, R. (1987). Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data. *Journal of the Royal Statistical Society. Series A (General)*, 150(3):192, ISSN: 00359238, DOI: 10.2307/2981473, <https://www.jstor.org/stable/2981473?origin=crossref>.
- Commission, E. (2020). ESS Standard for Quality Reference Line. https://unstats.un.org/unsd/dnss/docs-nqaf/eurostat-esqr_final.pdf.
- Di Bartolomeo, M. and Hu, Y. (2016). There is more to streamgraphs than movies: Better aesthetics via ordering and lassoing. *Comput. Graph. Forum*, 35(3):341–350, ISSN: 0167-7055.
- Gigerenzer, G. (2011). What are natural frequencies? *BMJ*, 343(oct17 2):d6386–d6386, ISSN: 0959-8138, 1468-5833, DOI: 10.1136/bmj.d6386, <https://www.bmjjournals.org/lookup/doi/10.1136/bmj.d6386>.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., and Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest*, 8(2):53–96, ISSN: 1529-1006, DOI: 10.1111/j.1539-6053.2008.00033.x.

- Heckler, A. F., Mikula, B., and Rosenblatt, R. (2013). Student accuracy in reading logarithmic plots: The problem and how to fix it. In *2013 IEEE Frontiers in Education Conference (FIE)*, pages 1066–1071. DOI: [10.1109/FIE.2013.6684990](https://doi.org/10.1109/FIE.2013.6684990).
- Hoffrage, U., Lindsey, S., Hertwig, R., and Gigerenzer, G. (2000). Communicating statistical information. *Science*, 290(5500):2261–2262, ISSN: 0036-8075, DOI: [10.1126/science.290.5500.2261](https://doi.org/10.1126/science.290.5500.2261), <https://science.sciencemag.org/content/290/5500/2261>.
- Hofmann, H., Follett, L., Majumder, M., and Cook, D. (2012). Graphical tests for power comparison of competing designs. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2441–2448, DOI: <https://doi.org/10.1109/TVCG.2012.230>.
- Kosara, R. (2020). In Praise of the Diagonal Reference Line. <https://eagereyes.org/blog/2020/in-praise-of-the-diagonal-reference-line>.
- Lawrence, M. and O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, 8(1):15–26, ISSN: 0169-2070, DOI: [10.1016/0169-2070\(92\)90004-S](https://doi.org/10.1016/0169-2070(92)90004-S), <http://www.sciencedirect.com/science/article/pii/016920709290004S>.
- McConway, K. and Spiegelhalter, D. (2021). Sound human, steer clear of jargon, and be prepared. *Significance*, 18(2):32–34, DOI: [10.1111/1740-9713.01508](https://doi.org/10.1111/1740-9713.01508).
- Menge, D. N. L., MacPherson, A. C., Bytnarowicz, T. A., Quebbeman, A. W., Schwartz, N. B., Taylor, B. N., and Wolf, A. A. (2018). Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature Ecology & Evolution*, 2(9):1393–1402, ISSN: 2397-334X, DOI: [10.1038/s41559-018-0610-7](https://doi.org/10.1038/s41559-018-0610-7), <http://www.nature.com/articles/s41559-018-0610-7>.
- Monmonier, M. (2005). Lying with maps. *Statistical Science*, 20(3):215–222, ISSN: 08834237, <http://www.jstor.org/stable/20061176>.
- Mountz, A. (2009). The other. In *Key concepts in political geography*, page 328–338. SAGE.
- Müller, A., Lausser, L., Wilhelm, A., Ropinski, T., Platzer, M., Neumann, H., and Kestler, H. A. (2021). A perceptually optimised bivariate visualisation scheme for high-dimensional fold-change data. *Advances in Data Analysis and Classification*, 15(2):463–480, ISBN: 1862-5355, DOI: [10.1007/s11634-020-00416-5](https://doi.org/10.1007/s11634-020-00416-5), <https://doi.org/10.1007/s11634-020-00416-5>.
- Otava, M. and Mylona, K. (2020). Communicating statistical conclusions of experiments to scientists. *Quality and Reliability Engineering International*, 36(8):2688–2698, DOI: <https://doi.org/10.1002/qre.2697>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.2697>.
- Reuters Staff (2020). Fact check: Mask mandates are not causing a spike in COVID-19 infections. *Reuters*, <https://www.reuters.com/article/uk-factcheck-mask-mandates-not-causing-c-idUSKBN2802W4>.

- Romano, A., Sotis, C., Dominion, G., and Guidi, S. (2020). The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. *Health Economics*, 29(11):1482–1494, ISSN: 1099-1050, DOI: [10.1002/hec.4143](https://doi.org/10.1002/hec.4143), <https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.4143>.
- Rosling, H. and Zhang, Z. (2011). Health advocacy with gapminder animated statistics. *Journal of Epidemiology and Global Health*, 1:11–14, ISSN: 2210-6014, DOI: <https://doi.org/10.1016/j.jegh.2011.07.001>, <https://doi.org/10.1016/j.jegh.2011.07.001>.
- Ryan, W. H. and Evers, E. R. K. (2020). Graphs with logarithmic axes distort lay judgments. *Behavioral Science & Policy*, 6(2):13–23, ISSN: 2379-4615, DOI: [10.1353/bsp.2020.0011](https://doi.org/10.1353/bsp.2020.0011), <https://muse.jhu.edu/article/799814>.
- Sevi, S., Aviña, M. M., Péloquin-Skulski, G., Heisbourg, E., Vegas, P., Coulombe, M., Arel-Bundock, V., Loewen, P. J., and Blais, A. (2020). Logarithmic versus Linear Visualizations of COVID-19 Cases Do Not Affect Citizens' Support for Confinement. *Canadian Journal of Political Science*, 53(2):385–390, ISSN: 0008-4239, 1744-9324, DOI: [10.1017/S000842392000030X](https://doi.org/10.1017/S000842392000030X).
- Slocum, T., McMaster, R., Kessler, F., Howard, H., and Mc Master, R. (2008). *Thematic Cartography and Geographic Visualization*. Prentice Hall, 3rd edition.
- Speckmann, B. and Verbeek, K. (2010). Necklace maps. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):881–889, DOI: [10.1109/TVCG.2010.180](https://doi.org/10.1109/TVCG.2010.180).
- The BBC Visual and Data Journalism team (2020). Coronavirus: How can we imagine the scale of Covid's death toll? <https://www.bbc.co.uk/news/resources/idt-7464500a-6368-4029-aa41-ab94e0ee09fb>. Section: uk.
- Thiessen, M. (2021). *Auf Abstand: Eine Gesellschaftsgeschichte der Coronapandemie*. Campus Verlag.
- Timmers, H. and Wagenaar, W. A. (1977). Inverse statistics and misperception of exponential growth. *Perception & Psychophysics*, 21(6):558–562, ISSN: 1532-5962, DOI: [10.3758/BF03198737](https://doi.org/10.3758/BF03198737), <https://doi.org/10.3758/BF03198737>.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 2 edition.
- VanderPlas, S. and Hofmann, H. (2015). Signs of the Sine Illusion—Why We Need to Care. *Journal of Computational and Graphical Statistics*, 24(4):1170–1190, ISSN: 1061-8600, DOI: [10.1080/10618600.2014.951547](https://doi.org/10.1080/10618600.2014.951547), <https://doi.org/10.1080/10618600.2014.951547>.
- Wagenaar, W. A. and Timmers, H. (1978). Extrapolation of exponential time series is not enhanced by having more data points. *Perception & Psychophysics*, 24(2):182–184, ISSN: 0031-5117, 1532-5962, DOI: [10.3758/BF03199548](https://doi.org/10.3758/BF03199548), <http://link.springer.com/10.3758/BF03199548>.

- Waldner, M., Diehl, A., Gračanin, D., Splechtna, R., Delrieux, C., and Matković, K. (2020). A Comparison of Radial and Linear Charts for Visualizing Daily Patterns. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1033–1042, ISSN: 2160-9306, DOI: [10.1109/TVCG.2019.2934784](https://doi.org/10.1109/TVCG.2019.2934784).
- Weiss, Y. (2020). These 12 Graphs Show Mask Mandates Do Nothing To Stop COVID. <https://thefederalist.com/2020/10/29/these-12-graphs-show-mask-mandates-do-nothing-to-stop-covid/>.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, ISBN: [978-3-319-24277-4](#), <https://ggplot2.tidyverse.org>.

Affiliation: