



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

Susan VanderPlas

Statistics Department

University of Nebraska Lincoln

United States

Adalbert F.X. Wilhelm

School for Business

Social and Decision Sciences

Jacobs University Bremen

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 caused by different mutations of the SARS-CoV-2 virus have swept across the globe, claiming lives, stressing healthcare systems, and changing many different facets of our day-to-day experience. According to the WHO, almost 600 million confirmed cases and over 6.45 million deaths

have been reported as of late August 2022¹. 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 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. 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, the general principles remain the same; ensuring clear understanding by organising graphics in such a way that the story of the data is told most effectively.

1.1. Data Journalism and the Media

In recent decades, many media organisations have established data teams that received unprecedented attention and wide-ranging opportunities during the pandemic to showcase their skills and abilities, not only in visualising data, but also in explaining their data collection and data analysis strategies and methods. COVID-19 allowed these data journalists to prove their utility in newsrooms, presenting data in digestible form to the public in news articles. Hence, data journalism will certainly be one of the beneficiaries of the COVID-19 pandemic and has already become an important part of news publishing, with COVID-19 delivering many excellent applications, often presented in an interactive visual format on the web, such as dashboards (Koch 2021).

At the same time, we still see a lot of defective graphics disseminated and shared; some that violate fundamental statistical and visualisation design 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 (UK EQUATOR Centre 2022). The use of numbers, their graphic representation and the systematic creation of charts and tables from data in the mass media have been accompanied from the beginning by efforts to improve the communication of quantitative and statistical information. While efforts continue to be made in this area, historical documents describe many of the same problems we still struggle with today (American Statistical Association 1915; Haemer 1949; Koverman 1961; Fienberg 1979; Hoffrage et al. 2000; Tufte 2001; Rosling and Zhang 2011; Otava and Mylona 2020). In other words, while our *technology* for recording, visually summarising, and communicating information distilled from data has vastly improved over the past 100 years, we still face many of the same problems when *designing* charts which communicate data accurately and effectively.

In other areas, such as visualisation of uncertainty, the pandemic presents new challenges for scientific communication. How do we effectively show the uncertainty in the data amid flawed tests, new strains, changing diagnostic criteria, and incredible variability in global public health infrastructure? Moreover, how do we do this without further decreasing public trust in institutions responsible for shepharding us through the pandemic? While most designers chose to side-step these questions in the moment, as we assess the pandemic in graphical form, we should also consider what was omitted

¹https://www.who.int/docs/default-source/coronavirus/who_mou_30_august-2022.pdf

from our visual renderings.

1.2. Effective Data Communication

A best practice for effective communication (graphical or otherwise) is to present information in an understandable way (Gigerenzer et al. 2007; Gigerenzer 2011), for example by saying "one in ten" instead of 10%, or using absolute rather than relative numbers. Presenting information in a visually appealing graphic form and highlighting the most important points in "fact boxes" is one extremely effective way to engage readers while focusing on key takeaway points. This approach was developed to increase effective communication between stakeholders such as healthcare providers and patients (Harding-Zentrum für Risikokompetenz 2022). 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 on 14 March 2020 in the Washington Post with the title "Why outbreaks like coronavirus spread exponentially, and how to flatten the curve" (Stevens 2020). Nevertheless, both the complexity of phenomena and the "bipolarity" of statistical thinking remain a challenge. While human thinking tends to simplify patterns, 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 important not only 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 confounding variables. 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 to support one side in a polarised debate (McConway and Spiegelhalter 2021).

Finally, it is critical that data communication take into account the intended audience. Graphics created to communicate with public health officials should be different from those created for the purpose of informing the public. Designers must carefully consider the visual capabilities, attention constraints, and mathematical sophistication of the audience when creating a chart, but in this era of instantaneous global communication, charts designed for specific audiences may still reach a large population of viewers with different capabilities. In this paper, we focus primarily on graphics which are intended for mass consumption, if only to reduce the scope to something which can be (non-exhaustively) covered in a paper.

1.3. Data visualisation Design and Testing

Having briefly discussed the media landscape surrounding the COVID pandemic and goals for communicating about data effectively, we should also briefly discuss the choices that go into creating visual representations of data intended to convey quantitative information. While the overall goal of a data visualisation is to tell the story of the data clearly and effectively, how this story is conveyed involves a series of design choices: Which variables are shown, What limits are imposed, What visual forms are used to represent the data? How will the chart (and any surrounding story, in journalism) be

presented to the viewer? As the designer makes these choices, they have in mind an audience, and the audience's assumed characteristics affect the design choices as well: language, education level, likely medium used to access the visualisation (computer, physical newspaper, tablet, cell phone) – all of these considerations also impact the design of the chart and what ideas the designer considers during the creative process.

Once the basic form of the chart is determined, there are additional considerations that may be specified by the stylistic conventions of different news organisations, but may also be deliberately chosen for visual effect; colour palettes, fonts, spatial layouts, and annotations may be used to highlight information, convey additional points of a narrative, or provide helpful information about how to read an unfamiliar chart. While each of these design choices could be (and have been) the subject of additional paper(s), it is important to acknowledge that the graphics we analyse in this paper did not simply pop into being – they are the product of a series of decisions, some consciously made and others imposed by convention or technological capabilities. We will highlight particular choices which were effective (or not) throughout this paper, but we will also leave some undiscussed for the sake of brevity.

Of course, the creative process does not only involve the creator. Most graphics that are created for a purpose are tested, either formally or informally, before they are published, even if the test results are never themselves published. Because so many of the charts we highlight in this paper were discussed and disseminated via social media, in some cases we can provide contemporary discussions of the design process and reactions to the charts we highlight in this paper. Where such examples are not provided, that should in no way be taken as evidence that the design and testing process was less formal; it may only have been less public.

1.4. Scope

Our focus in this paper is to highlight, analyse, and discuss a selection of charts and graphics which caught our attention during the initial, real-time experience of the pandemic. As we primarily consume media from western Europe and the United States, our assessment is primarily limited to news outlets catering to these audiences; a retrospective global analysis of COVID graphics would certainly be interesting, but is beyond the scope of this work. While the charts we have selected were of personal interest during the fear and uncertainty of the initial stages of the COVID pandemic, in this paper we assess these same charts retrospectively, from a world where COVID is an acknowledged fact of life. While it still has significant impacts on our day-to-day lives, the existence of treatments, vaccines, and other preventative measures, as well as the human tendency to adapt to changing circumstances mean that the novel coronavirus which emerged in 2019 no longer dominates our consciousness in the same way. Because of this perspective difference, we have enough distance from the initial cataclysm to be able to assess the emotional impact of the graphics at the time and to assess their utility retrospectively.

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 during this initial period were commonly visualised by choropleth maps, see for example Figure 1 showing the total number of infections reported in each country as of January 14, 2022. A central element of the narrative during March 2020 was the ubiquity of the disease and the accompanying global impact.

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. While choropleth maps based on raw numbers of cases might look convincing and are an obvious choice under the circumstances, the use of these charts ignores 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. Viewers tend to integrate similarly coloured areas of the map 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 as well as 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. colour-scheme selection: 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.

The issues with choropleth maps were highlighted in trade publications during the initial stages of the pandemic ([Field 2020](#)), but these publications did not penetrate the bureaucracy of public health officials suddenly tasked with data visualisation in real time at scale.

Proportional symbol maps, see Figure 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. Proportional symbol maps sacrifice familiar shapes and topological connectedness for equal visual weight ([Gao et al. 2019](#)); while this representation avoids the area bias, proportional symbol maps are subject to under-estimation biases ([Shim and Son 2008](#)) that may be of particular concern when visualising epidemic disease spread. In addition, the visual impression of any map-based illustration depends on the projection chosen for the underlying map; the examples in Figure 1 and Figure 2 have a western-centric bias.

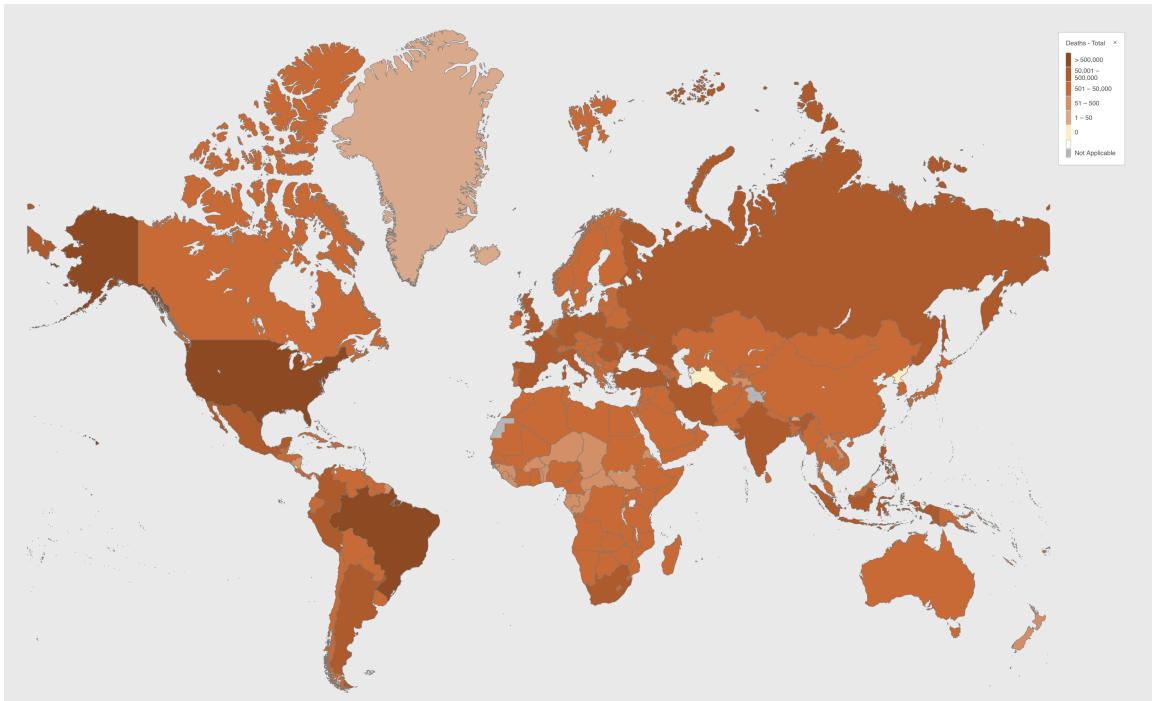


Figure 1: Choropleth map of the COVID-19 cases world-wide by country ([World Health Organization 2022](#)).

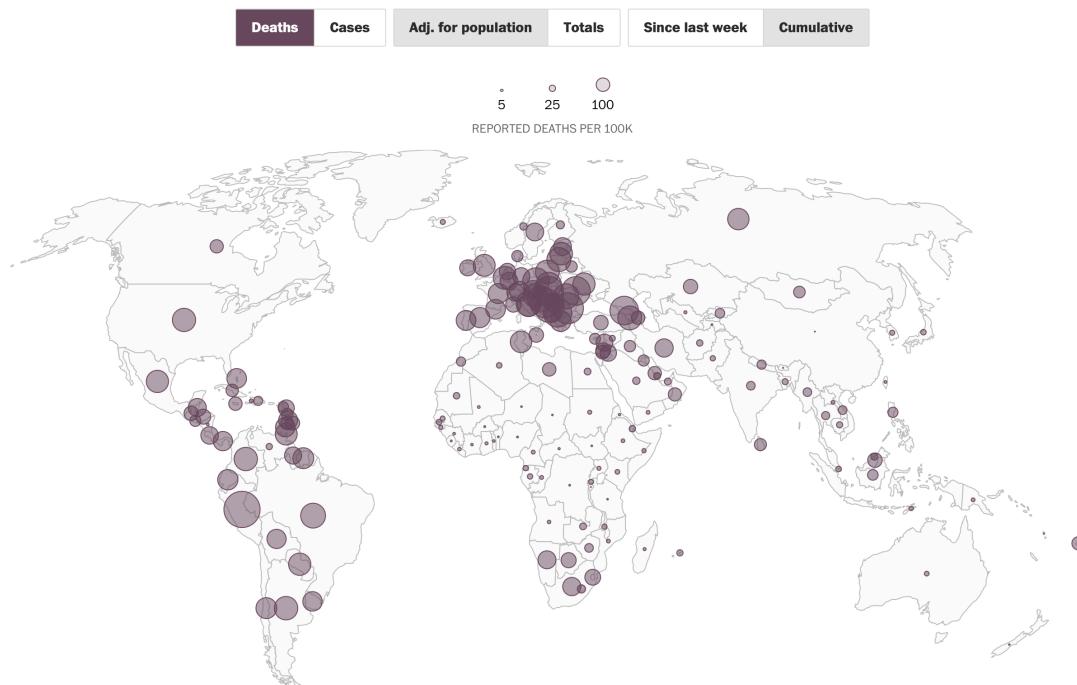


Figure 2: Proportional symbol map showing the cumulated reported COVID-19 related deaths adjusted for population world-wide by country ([Alcantara et al. 2022](#)).

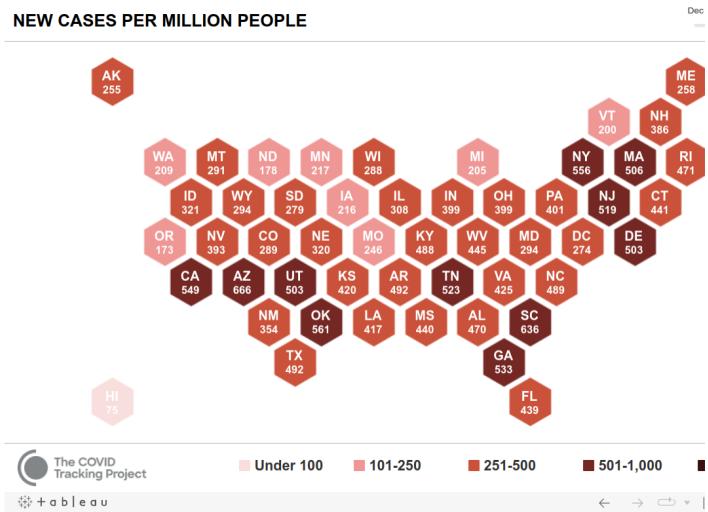
Another modification of the standard choropleth to address the area-bias problem is to relax the spatial borders of a region in order to show magnitude using area (as well as colour). Various algorithms can be used to modify geographic boundaries to maintain spatial continuity, or shapes such as hexagons or squares can be used to represent geographic entities (as in Figure 3a); with either choice, geometry, topology, or both must be sacrificed. While these methods are more popular in the academic visualisation literature, including analysis of the spread of COVID-19 (Yalcin 2022), they were not overly common in the media for global representations of COVID information. This may be because cartograms require more visual effort on the part of the viewer; in addition, news organisations may be reluctant to assume their readers have the geographic sophistication necessary to translate between the distorted map of the globe as drawn on the cartogram, and the undistorted mental map of the globe used for mental comparison purposes. A spatial dot density map might have allowed for analysis of data across geographic boundaries, but this approach was not commonly used to show global case counts, though it was highlighted as a useful option at the national level (Field 2020), as shown in Figure 3b.

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 to assess the global pandemic 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 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. It is not that often that day-to-day design changes (or lack thereof) cause fundamental problems with the interpretation of data visualisation, but during the COVID 19 pandemic, this situation was both extremely common and had real consequences; people would glance at the current COVID risk map and assume that "red" meant the same thing that it did yesterday, despite the fact that the underlying values had changed (Abiad 2020; Rubel 2020; Matthews 2020). One example of this issue over two months of NY Times graphics is shown in Figure 8; the overall colour scheme used changes once, but the variable displayed changes (at least) 4 times over the course of a two month period.

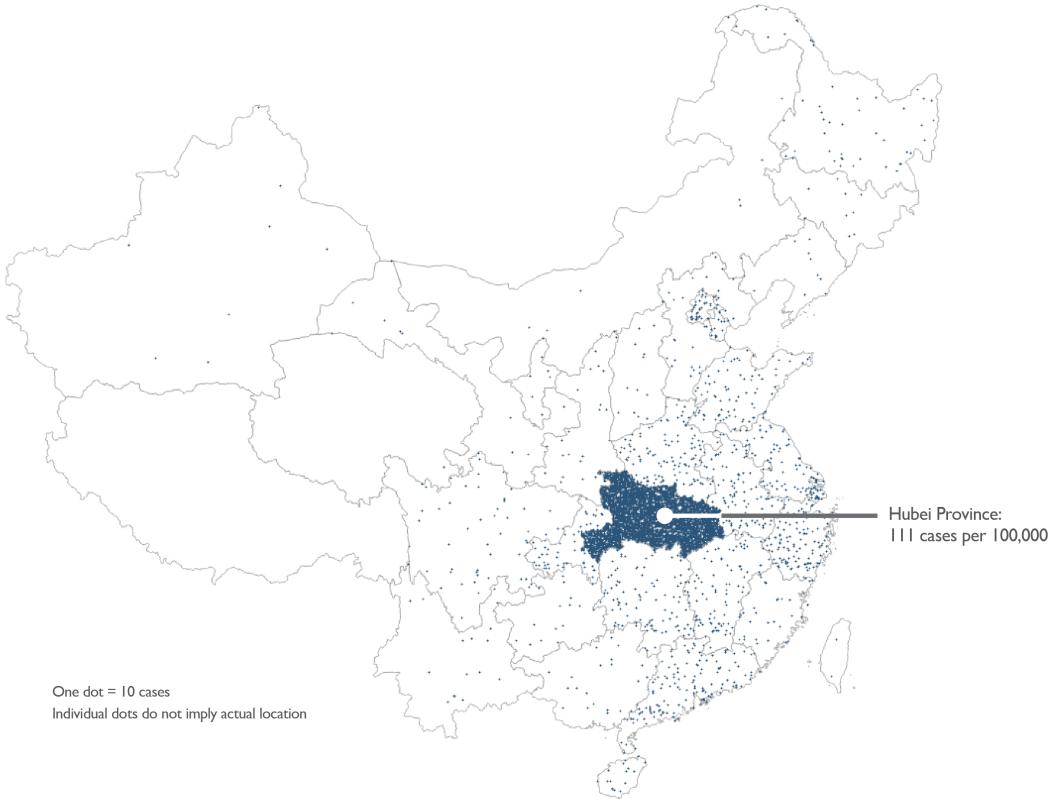
This problem is more noticeable in choropleths and other charts which show case counts based on colour scale alone than in time series charts, where we are more likely to expect and notice a scale change, but it is still a fundamental issue across chart types and of particular concern when there is no natural upper bound on the variable displayed. Some designers leaned into this by arranging their graphics to highlight the fact that cases were "off the charts", as in Figure 4, implicitly cuing readers that the situation was exceptional and worthy of their attention.

Early in the pandemic, as case numbers grew exponentially without appropriate public health measures to restrain them, the scale problem for choropleth maps was particularly complex; over time either the colour mapping changes or the map becomes useless



(a) Hexagonal Cartogram of COVID-19 cases in the United States by state ([The Atlantic 2021](#)).

Coronavirus in China: 24th February 2020



(b) A dot-density map of COVID cases in China ([Field 2020](#)).

Figure 3: Modifications of choropleth maps which attempt to address the area bias of choropleths in different ways. Both methods are applied at the national scale rather than at global scale because of limitations with the visual form or data quality.

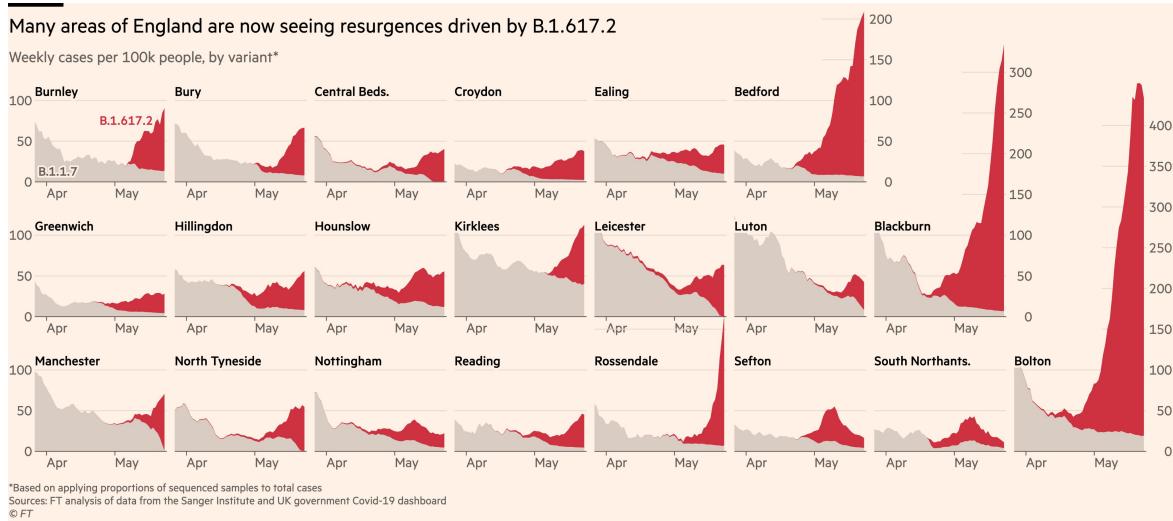


Figure 4: Chart showing the increase in cases as a result of the B.1.617.2 variant in regions of the UK, May 27, 2021. Notably, this chart arranges areas to accommodate certain regions whose cases are “off the chart” in order to emphasize the exceptional situation. ([Burn-Murdoch 2021a](#))

due to compression against the upper limits of the scale, as in [Panchadsaram \(2020\)](#). The magnitude of the problem also depends on whether the scale uses a set of bins (a discrete mapping to a continuous variable) or uses a continuous colour mapping; in some cases, outlets switched between the two options over time or changed the underlying unit entirely. Interestingly, media outlets and municipal dashboards were subject to criticism when the colour scales changed in response to increasing case counts ([Boeck 2020](#)) as well as when the colour scale did not change ([Sandalow 2020](#)), illustrating the difficulty of designing visualisations amid competing constraints and a constantly evolving pandemic.

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, hospitalisations, and/or deaths are plotted on the y-axis, with date shown on the x-axis, as in Figure 4 and Figure 5. 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 5. 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.

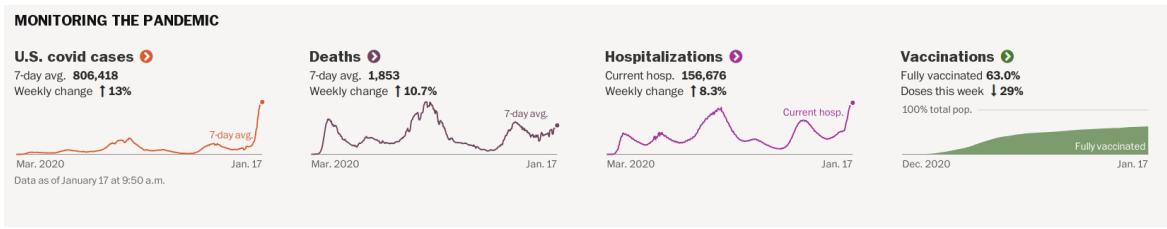


Figure 5: COVID monitoring summary charts on the Washington Post’s Coronavirus page ([Washington Post 2022a](#)).

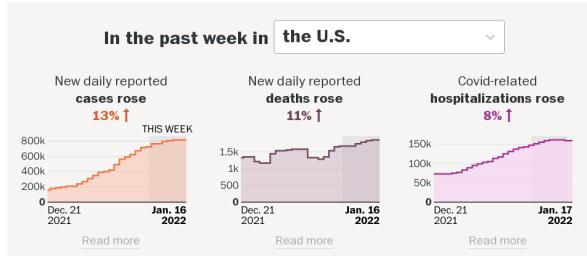


Figure 6: Some outlets provide contextual information to assist readers with interpreting time series plots ([Washington Post 2022b](#)).

Often, as in Figure 5, these time series show multiple variables: cases, deaths, and hospitalisations (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 hospitalisations/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 6.

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, 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. This issue will be discussed further in Section 5 when we discuss log scales, which are one solution to our misperception of exponential growth.

Yet another approach to showing the temporal component of the pandemic was to leverage web graphics to show the temporal component through animation; this freed up the x axis which in the default configuration would have been devoted to time and allowed another variable to be displayed. While this approach was common on social media, it was used more selectively, with carefully chosen variables by professional media outlets. Figure 7 shows an example from the Financial Times which uses the x -axis to show GDP per capita and the y -axis to show vaccination rate, bubble size indicates a country’s population. With so many variables and ratios of variables, these charts are complex to read and use, even though they are visually appealing.

Other time series representations were more abstract, sacrificing exact data representations for a higher-level summary comparing regions and points in time without the burden of numerical calculations. Several versions from the New York Times are shown

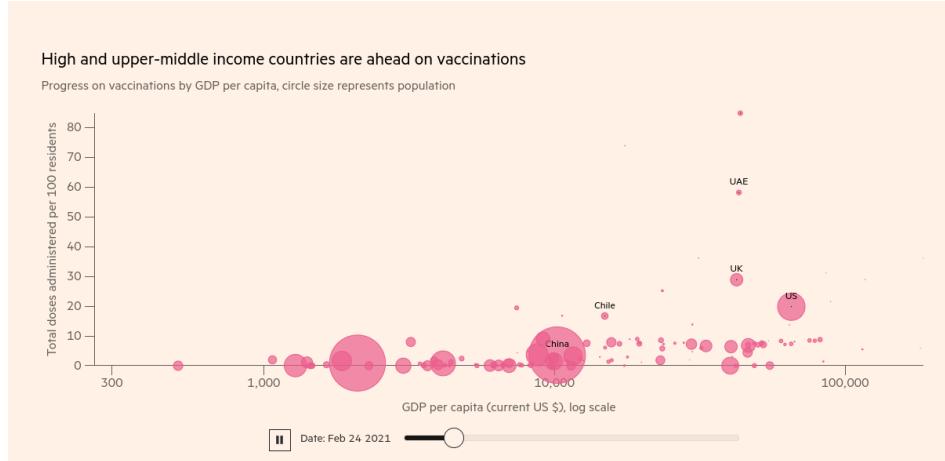


Figure 7: Animated bubble chart showing vaccination rate relative to per capita GDP for each country ([F.T. Visual and Data Journalism Team 2022](#)).

in Figure 8. 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 colour scale is (perhaps intentionally) opaque – it would be extremely difficult to translate the visualisation into any sort of accurate numerical estimate. In addition, by hiding these details from the user, it is very difficult to identify when the underlying mathematical representation changes over time, as it did several times during the summer of 2020. The use of a similar colour scheme as well as similar geometric 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 9. 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 10 shows COVID cases, positivity rate, and hospital admissions from South Africa's Gauteng province, which 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. In addition, it is clear that the chart is highlighting the Omicron variant from the colour selection; the red used to show the Omicron numbers is a sharp contrast from the green and blue shades used to show the first three waves of the pandemic. Notably, even though red and green in combination are usually to be avoided for the sake of colourblind viewers, the shades of blue/green variants which are chosen are light enough that the red stands out even to someone

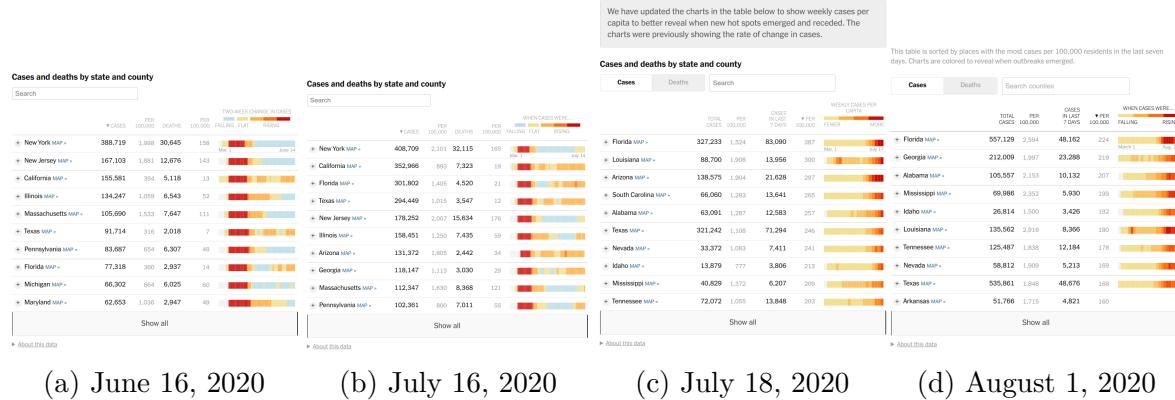


Figure 8: A series of screenshots of the NY Times state-by-state display showing evolving colour schemes and even metrics; initially, a discrete scale of two-week change in cases is shown; then, a similar colour scheme is used with a different metric. Around July 17, the colour scheme and variable shown changed to weekly cases per capita (from "fewer" to "more"), and by August 14, the variable shown changed back again ("falling" to "rising") while the colour scheme stayed the same.

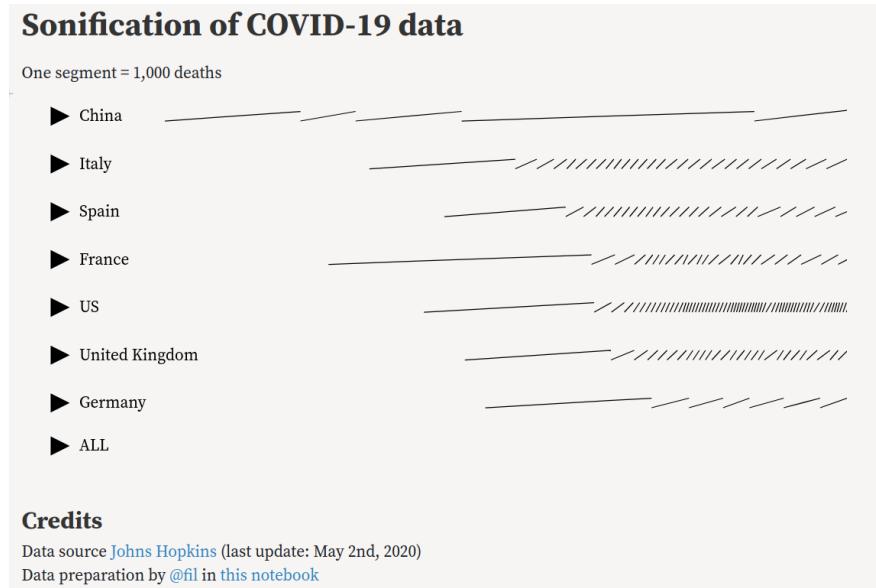


Figure 9: Frequency-domain case count graphics. Here, each 1000 deaths are represented by a single line segment whose slope represents the time taken to reach the next 1000 deaths. This provides viewers with a sense of the pace of the epidemic, rather than the raw case counts in more standard time series representations. This representation can also be shown in the auditory domain, providing access to those who are typically excluded from visualisations due to vision loss (Vuillemot 2020).

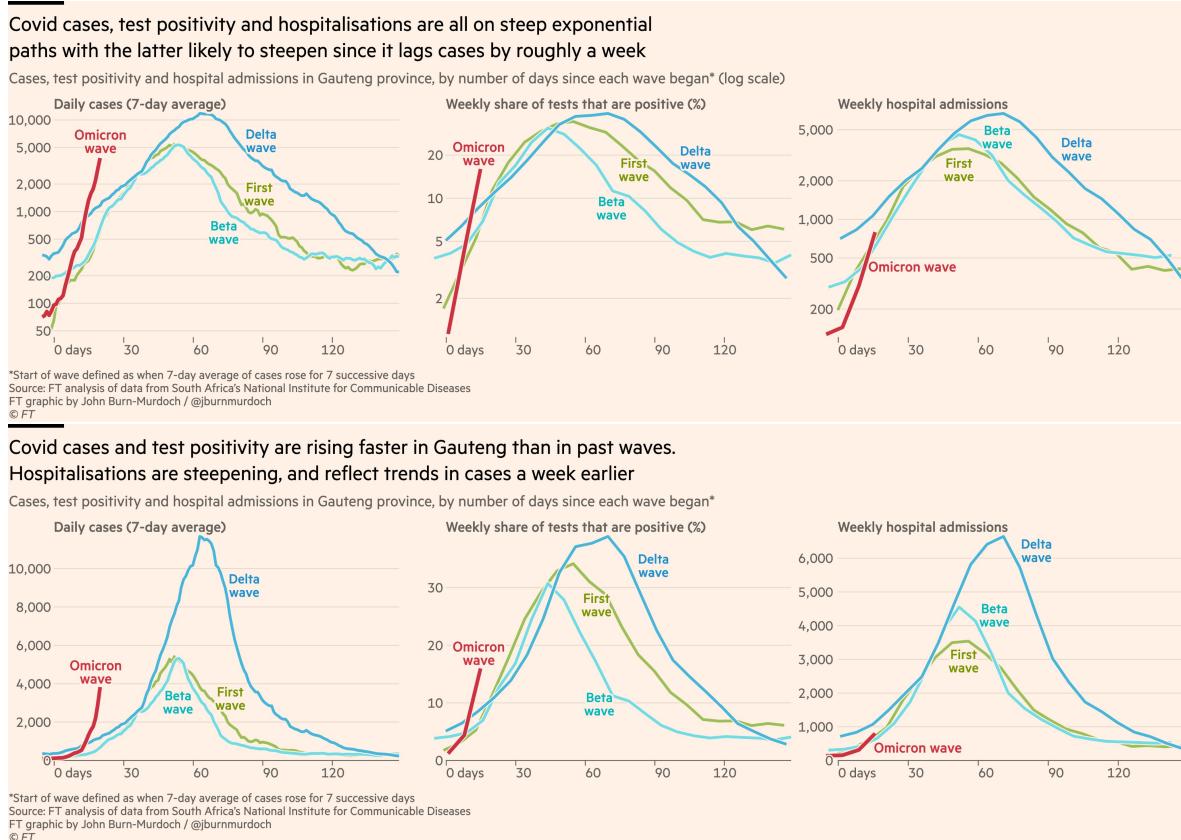


Figure 10: COVID Cases, positivity rate, and hospital admissions from South Africa's Gauteng province, on a log scale (top) and linear scale (bottom). Successive waves are overlaid, showing that the increase in cases due to the Omicron variant was sharper than any previously measured increase, while hospitalisations were much more comparable to previous waves (Burn-Murdoch 2021b).

with colour vision deficiency.

While traditionally, time series information has been limited to charts with time on a linear x-axis, many charts manipulated time to highlight different aspects of the pandemic. In some cases, these modifications provided large amounts of utility for a relatively small increase in complexity; for example, John Burn-Murdoch's decision to index Financial Times COVID graphs to the days since the Nth case or death. This decision is a small modification of the basic time series plot, but has important consequences in the chart's readability as well as the visual emphasis for users: the graphs in the Financial Times were designed to emphasize "... that there's an inevitability about how coronavirus spreads... even if there are only a few cases in your country today, based on all the data we have, you will end up going along that path, the same path that the likes of Italy and Spain have been on so tragically" (Hannen 2020). By placing all countries on an even footing divorced from the temporal randomness of when the first case appeared, the viewer can effectively compare the effect of government policies and other interventions on the trajectory over time (Burn-Murdoch 2020a,b).

The New York Times intentionally violated typical design guidelines in a controversial chart; the x-axis was wrapped around an Archimedean spiral to provide a sense of

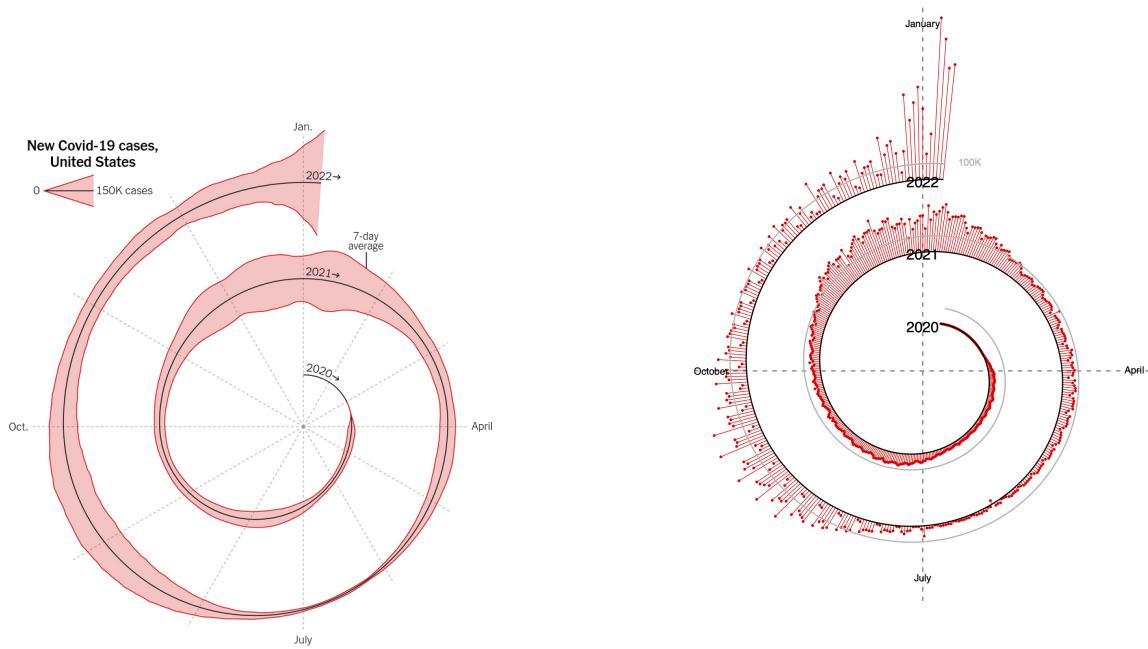


Figure 11: COVID cases in the United States, 2020-2022. The original graph from the New York Times is on the left ([Shaman 2022](#)), and a reenvisioned version which is more perceptually friendly is on the right ([Shrestha 2022](#)).

periodicity in the year-by-year evolution of case counts, as shown in Figure 11. 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](#)). A re-envisioned version created by Sourya Shrestha, aligns counts on the outside of the spiral, and displays them as radial lines, mitigating some of the issues with the line-width illusion by directly showing each line as its own entity, facilitating direct comparisons.

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. By design, this type of chart devotes less area to previous years of the pandemic, which may decrease the amount of focus given to that data.

Other designers used the capabilities of web graphics to show the time component in the data through animation. In some cases, this took the form of animated bubble plots, while others dispensed with the traditional form of a line graph altogether, showing rising case counts using a "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 ([Stats On Clock 2022](#)); a screenshot is shown in Figure 12. 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

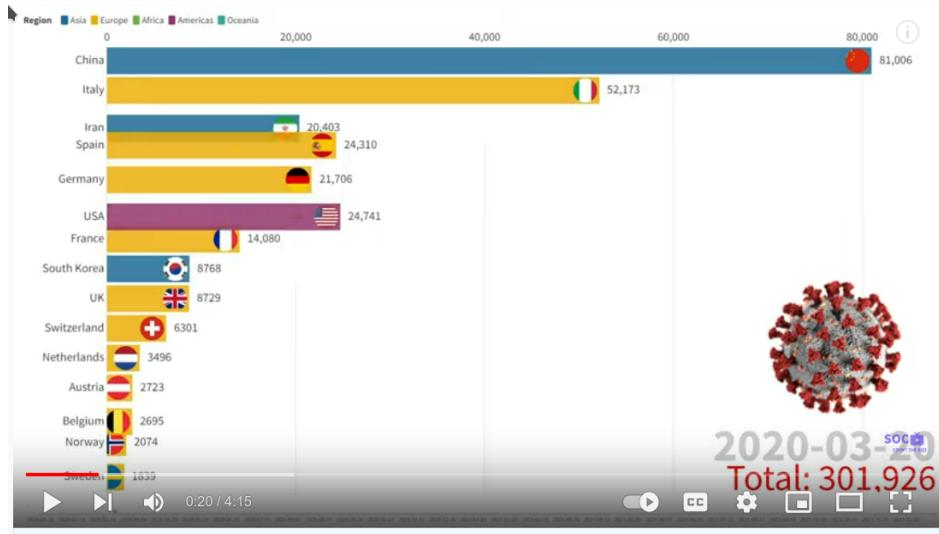


Figure 12: A screenshot from Stats on Clock's bar chart race showing counts of COVID cases over time throughout the pandemic ([Stats On Clock 2022](#)).

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 in the US, with occasional short quotes from individuals' obituaries to humanise 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 less sentimental dot-density timeline by the New York Times about 9 months later chronicled the nation reaching 450,000 deaths; the difference between the two visualisations brings to mind the quote oft attributed to Stalin: "One death is a tragedy; a million deaths is a statistic" ([Quote Investigator 2016](#)).

A slightly different approach by the BBC's Visual and Data Journalism team, shown in Figure 13, 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](#)). The visualisation comes with sound as well, an attempt to make the pandemic an auditory experience as well as a visual one. The form, animation, and auditory experience are all indicative of an emotional appeal ([Kostelnick 2016](#); [D'Ignazio and Klein 2020](#)); with the flower representation as an explicitly recognized symbol of grief; it is clearly not about showing specific case counts and deaths numerically.

4. The ranking narrative

"The plague is always the others" ([Thiessen 2021](#)). This is the short formula for dealing

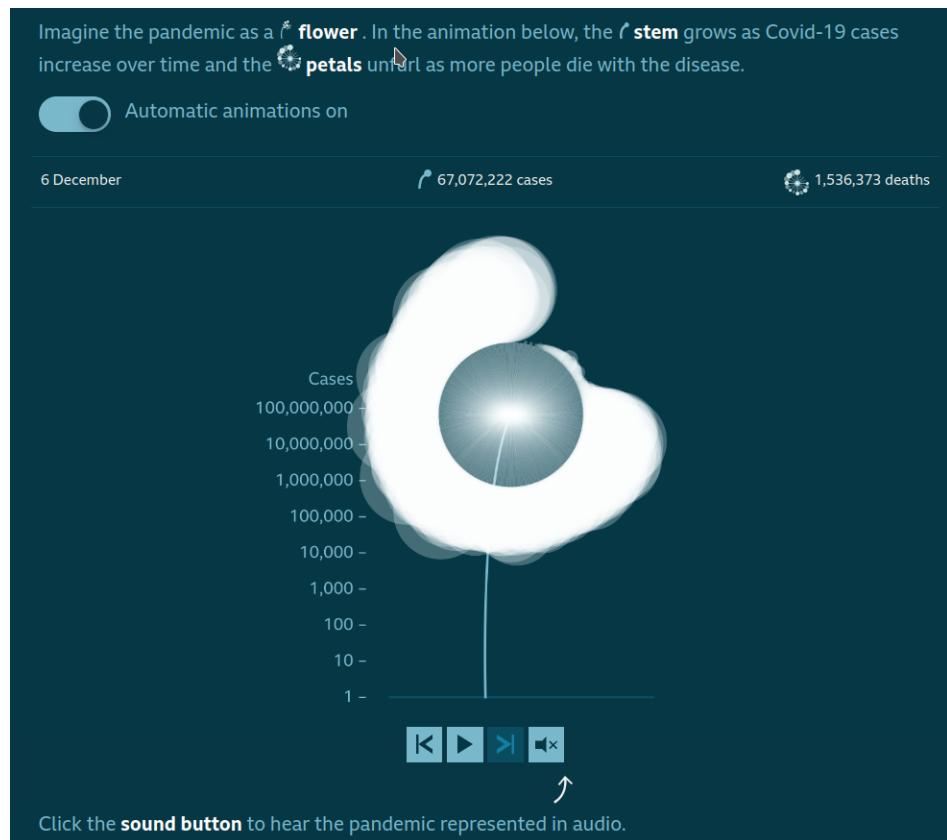


Figure 13: The BBC Visual and Data Journalism Team's COVID flower, showing the cases and deaths over time (The BBC Visual and Data Journalism team 2020).

with infectious diseases from a historical perspective. 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 Figure 14) 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 standard. Streamgraphs (see lower image in Figure 14) 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, as well as design choices such as the ordering of the different groups. 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](#)). When paired with the bar charts showing excess mortality, the streamgraph provides temporal context to the instantaneous information in the bar graph.

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 15 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.

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

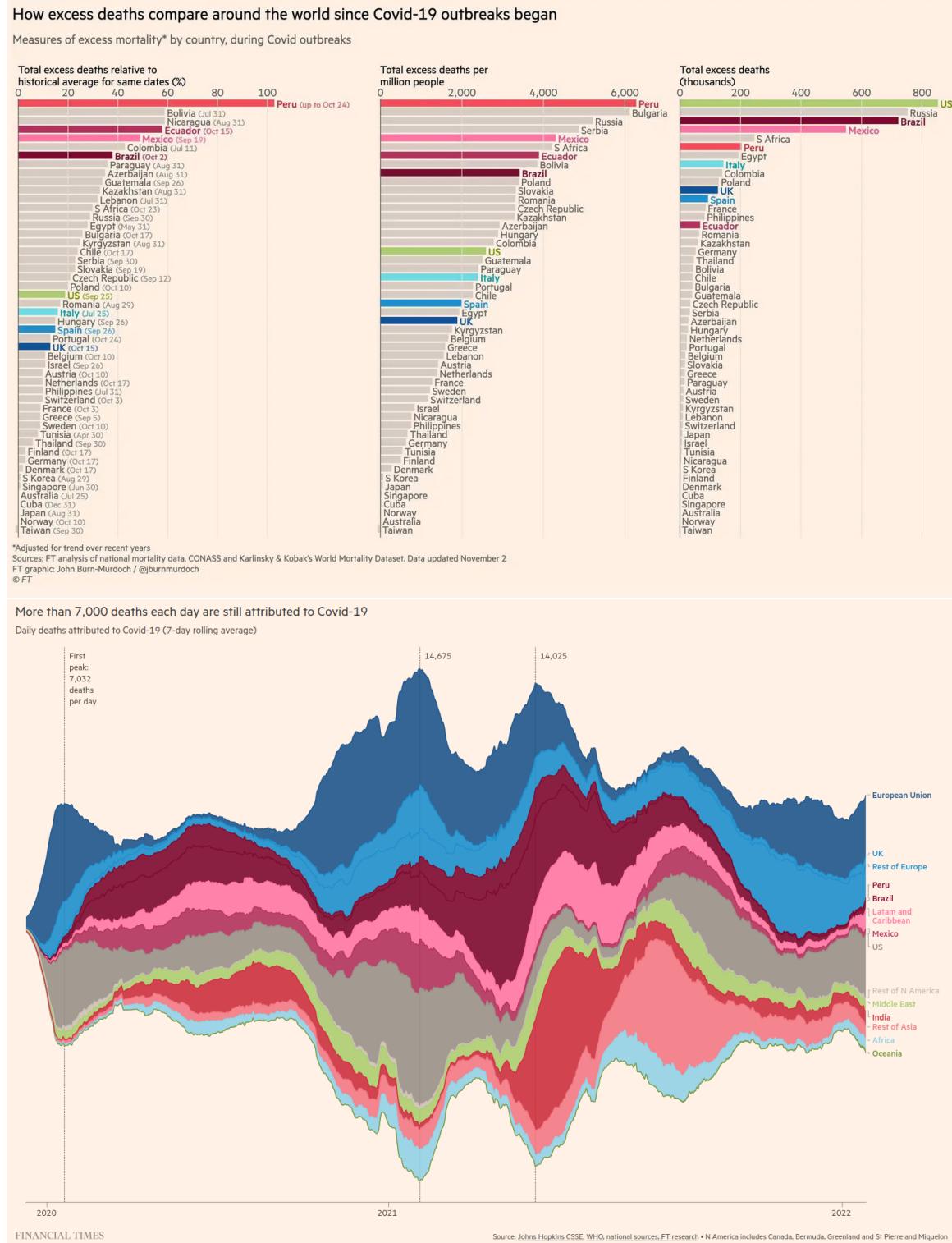


Figure 14: Ordered death rates in countries and the temporal evolution of the death toll share in major world regions (FT Visual & Data Journalism team 2021).

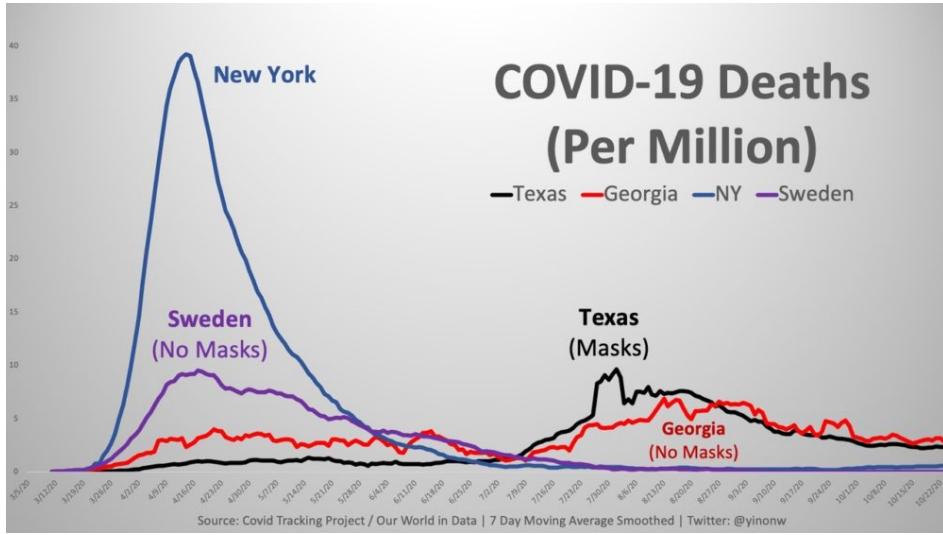


Figure 15: Deaths per million people in New York, Sweden, Texas, and Georgia, as shown in the Federalist (Weiss 2020). The goal of this comparison chart is to lead the reader to conclude that masks are ineffective, as there is not a clear difference between Texas and Georgia. New York is also shown (though 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.

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 variety 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 total 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 populations (and corresponding case counts) of Michigan and Washington are much lower than the populations (and case counts) 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 17.

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 18. Diagonal 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

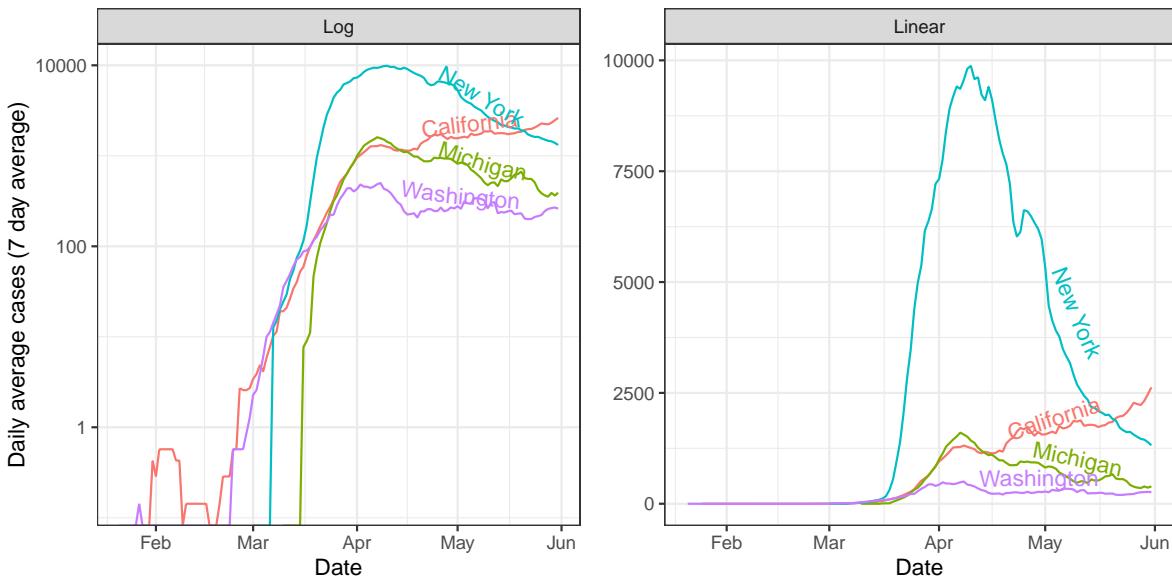


Figure 16: In the early stages of the pandemic, log scales allowed the comparison of raw case counts in locations with vastly different population and case counts.

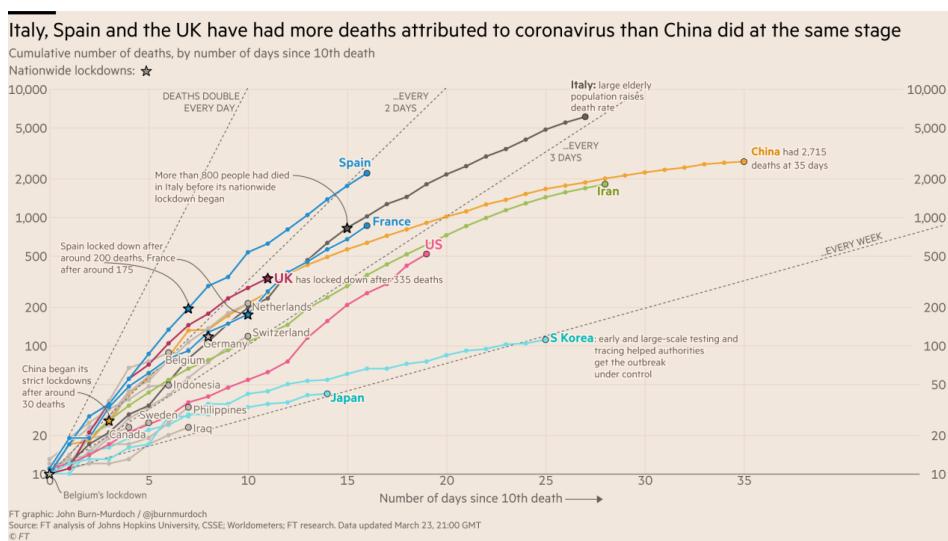


Figure 17: Reference lines to compare exponential growth rates of deaths in different countries. This provides some additional context that may help individuals use log scale data more successfully. This approach was first featured in the Financial Times, but was quickly adopted by the New York Times, 91-DIVOC, and other outlets. Graph from the Financial Times (March 23, 2020), image from [Kosara \(2020\)](#).

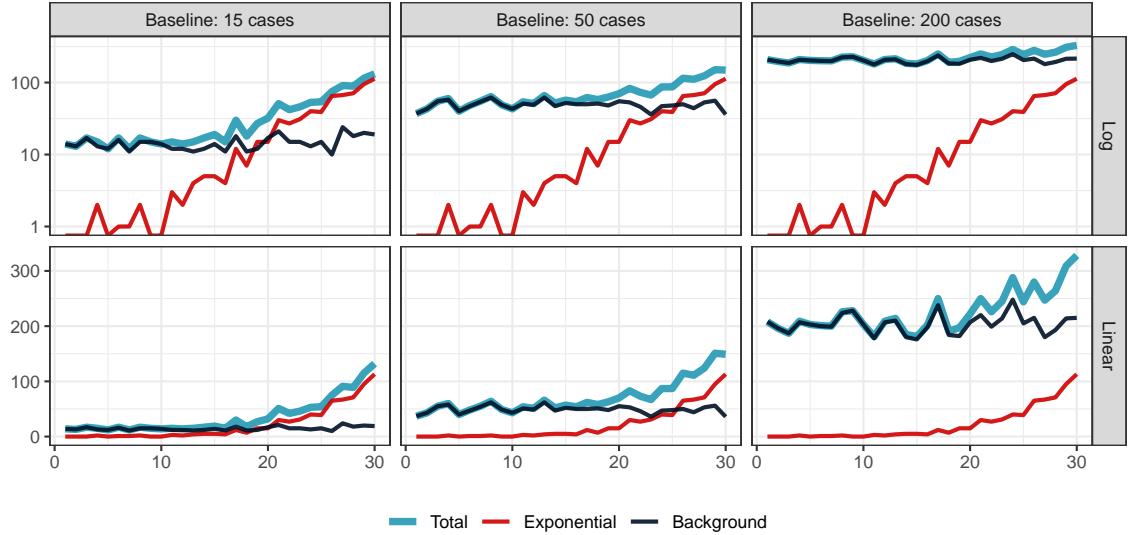


Figure 18: One problem with log scales is that if there is a background level of spread, it can be hard to notice the introduction of an additional source of exponential spread. Linear scales do not have this problem - the exponential source is noticeable very quickly in the total line, but on the log 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 of a uniform level of spread over time, even though the exponential source makes up approximately 50% of the cases at the end of the time period shown.

made available, but usually only in a retrospective manner, as in Figure 19.

While log scales have their problems, linear scales are not immune from issues either. It can be very difficult to adequately compare to past situations when looking at the full time series of case counts. For example, in Figure 20, 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, we can gain some clarity as to their use in this particular context. We are abysmal at forecasting exponential growth, vastly under-estimating future growth by using linear or quadratic approximations (Wagenaar and Timmers 1978; Lawrence and O'Connor 1992; Timmers and Wagenaar 1977). If

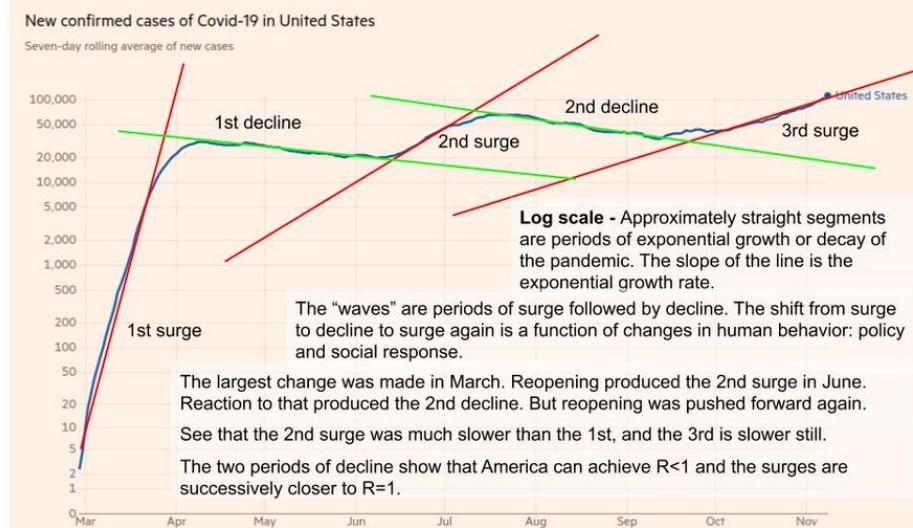


Figure 19: An annotated time series from the Financial Times which appeared on Twitter on November 12, 2020. This chart has annotations which show the decreasing R_{eff} , the effective rate of exponential spread, in successive peaks of the pandemic, with periods of $R_{eff} < 1$ between surges in cases. These annotations assist the reader with drawing complex conclusions from the chart, but are difficult to automate and thus tend to be manually curated (Gubrud 2020).

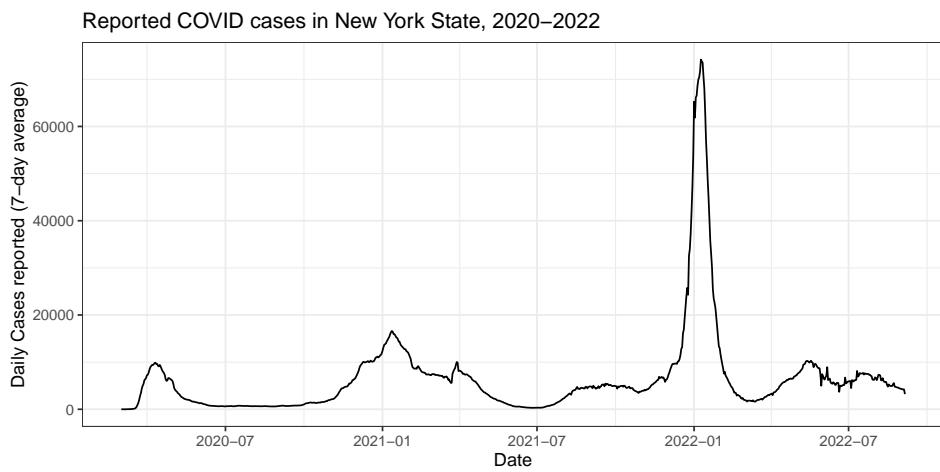


Figure 20: Reported COVID cases in New York State, 2020-2022. The linear scale makes it difficult to compare the trajectory of different waves to determine how severe the current status is relative to the past, because the primary contrast is the height of the relative peaks, rather than the growth rate. 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.

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 18, 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 action, 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. Thus, analysing the rich collection of visualisations produced as part of the pandemic provides a unique opportunity to examine how these graphics were (and were not) effective, as well as the design decisions made during their creation. As in any situation, good tools can be used and misused 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 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.

The pandemic and graphics used to show the ebb and flow of COVID cases highlight the challenges of data visualisation in the modern era. Even with numerous guidelines and standards for graphics, it is not straightforward to communicate complex data; creating charts which fully represent the intended message or messages remains 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/>.

Archived data and code to create the figures in this paper can be found at <https://github.com/srvanderplas/covid-visualization-dssv>.

Archived Information

This paper references many news articles which contain interactive graphics. In order to preserve these graphics (which depend on a great many external files), the authors have preserved many of the interactive graphics as self-contained html files (Vanderplas 2022). The authors have also made extensive use of the Internet Wayback Machine, which aims to preserve the state of important web pages as they change over time.

Acknowledgments

These ideas were first presented at DSSV 2020. We thank our colleagues, the reviewers, and the editors for valuable feedback and encouragement to complete and revise this manuscript.

References

- Abiad, F. (2020). A color diagram by @WHOLebanon for the distribution of Covid cases by district (September 21 Vs. 25). The rise and spread of cases was such that the scale had to be changed. The graph on the right actually shows worsening numbers though the change in the color can be misleading. <https://t.co/IZKVi9ePjH>. <https://twitter.com/firassabiad/status/1310202386821132293>.
- Alcantara, C., Shin, Y., Shapiro, L., Taylor, A., and Emamdjomeh, A. (2022). Mapping the worldwide spread of the coronavirus. <https://www.washingtonpost.com/graphics/2020/world/mapping-spread-new-coronavirus/>.
- American Statistical Association (1915). Joint committee on standards for graphic presentation. *Publications of the American Statistical Association*, 14(112):790–797, ISSN: 15225437, DOI: [10.2307/2965153](https://doi.org/10.2307/2965153), <http://www.jstor.org/stable/2965153>. tex.ids= JointCommitteeStandards1915 publisher: [American Statistical Association, Taylor & Francis, Ltd.].
- 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>.
- Boeck, T. (2020). #DataViz decisionmaking in action: was confused why the NYT COVID 19 map was showing many states lighter than I remembered (lighter = fewer cases). Upon closer inspection, they've been forced to change their color scale due to exploding case numbers. <https://t.co/p9Le8eTIQx>. <https://twitter.com/toriboeck/status/1329327634619228162>.
- 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, <http://www.jstor.org/stable/2564061>.
- Burn-Murdoch, J. (2020a). All countries' death toll trajectories: Increasingly looks like India tracking S Korea, Australia's curve tracking Japan, 38 countries now shown here, more every day. <https://twitter.com/jburnmurdoch/status/1244721960946409472/photo/1>.
- Burn-Murdoch, J. (2020b). Latest case trajectories for major countries: US cases continue to soar. unquestionably the worst outbreak in the world. reporting from @yuanfenyang & co on whether you can trust the Chinese data: <https://ft.com/content/4aa35288-3979-44f7-b204-b881f473fca0>. <https://twitter.com/jburnmurdoch/status/1244722153041387520>.
- Burn-Murdoch, J. (2021a). NEW: B.1617.2 is fuelling a third wave in the UK, with not only cases but also hospital admissions rising. Vaccines will make this wave different to those that have come before, but it remains a concern, and one that other countries will soon face. Thread on everything we know: <https://t.co/4825qOqgrl>. <https://twitter.com/jburnmurdoch/status/1397995388267810818>.
- Burn-Murdoch, J. (2021b). NEW: today's update from Gauteng, now on a log scale to better show current trajectories. Steepness of lines shows how much faster the growth in cases and positivity is now vs past waves, and hospital admissions are now steepening too as the acceleration in cases feeds through. <https://t.co/xBNGzxmwhC>. <https://twitter.com/jburnmurdoch/status/1466480113487392769>.
- 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.

- D'Ignazio, C. and Klein, L. F. (2020). *Data Feminism*. Ideas Series. The MIT Press, Cambridge, Massachusetts, ISBN: 978-0-262-35852-1.
- Field, K. (2020). Mapping coronavirus, responsibly. <https://www.esri.com/arcgis-blog/products/product/mapping/mapping-coronavirus-responsibly/>.
- Fienberg, S. E. (1979). Graphical Methods in Statistics. *The American Statistician*, 33(4):165–178, ISSN: 0003-1305, DOI: 10/chmrvh, <https://www.jstor.org/stable/2683729>.
- FT Visual & Data Journalism team (2021). Coronavirus tracker: the latest figures as countries fight the Covid-19 resurgence. *Financial Times*, <https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938>.
- F.T. Visual and Data Journalism Team (2022). Covid-19 vaccine tracker: the global race to vaccinate. <https://ig.ft.com/coronavirus-vaccine-tracker/>.
- Gao, P., Li, Z., and Qin, Z. (2019). Usability of value-by-alpha maps compared to area cartograms and proportional symbol maps. *Journal of Spatial Science*, 64(2):239–255, ISSN: 1449-8596, DOI: 10.1080/14498596.2018.1440649, <https://doi.org/10.1080/14498596.2018.1440649>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/14498596.2018.1440649>.
- 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.
- Gubrud, M. (2020). @DrEricDing We've basically cycled a small fraction of the reduction we had achieved, a smaller fraction in the 3rd surge than the second. Again, this is evidence that more use of better masks, ventilation, testing, etc. can make the difference & drive the pandemic into steep decline. <https://t.co/x9NoOgRDA7>. <https://twitter.com/mgubrud/status/1326784082399944704>.
- Haemer, K. W. (1949). Presentation Problems: Area Bias in Map Presentation. *The American Statistician*, 3(2):19–19, ISSN: 0003-1305, 1537-2731, DOI: 10.1080/00031305.1949.10483438, <http://www.tandfonline.com/doi/abs/10.1080/00031305.1949.10483438>.
- Hannen, T. (2020). Coronavirus trajectory tracker explained. <https://www.ft.com/video/9a72a9d4-8db1-4615-8333-4b73ae3ddff8>.
- Harding-Zentrum fur Risikokompetenz (2022). Faktenboxen zur mRNA-Schutzimpfung gegen COVID-19. <https://www.hardingcenter.de/de>.

- 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.scienmag.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>.
- Koch, T. (2021). Welcome to the revolution: COVID-19 and the democratization of spatial-temporal data. *Patterns*, 2(7), ISSN: 2666-3899, DOI: [10.1016/j.patter.2021.100272](https://doi.org/10.1016/j.patter.2021.100272), [https://www.cell.com/patterns/abstract/S2666-3899\(21\)00098-2](https://www.cell.com/patterns/abstract/S2666-3899(21)00098-2).
- Kosara, R. (2020). In Praise of the Diagonal Reference Line. <https://eagereyes.org/blog/2020/in-praise-of-the-diagonal-reference-line>.
- Kostelnick, C. (2016). The Re-Emergence of Emotional Appeals in Interactive Data Visualization. *Technical Communication*, 63(2):116–135.
- Koverman, H. C. (1961). PLANNING THE VISUAL PRESENTATION. *The Professional Geographer*, 13(6):24–26, ISSN: 0033-0124, 1467-9272, DOI: [10.1111/j.0033-0124.1961.136_24.x](https://doi.org/10.1111/j.0033-0124.1961.136_24.x), http://www.tandfonline.com/doi/abs/10.1111/j.0033-0124.1961.136_24.x.
- 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>.
- Matthews, C. (2020). A lesson in futzing with data to make a country seem less covid infected: just change the # scale of what your colors mean. As if by magic, things look better! Until someone like me comes along and redraws the map in photoshop using last week's color scale. cc @markrutte @rivm <https://t.co/Oi1IOGJP6>. <https://twitter.com/matthewscd/status/1315604109819551749>.
- 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>.

- 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>.
- 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>.
- Panchadsaram, R. (2020). If we are at war with this virus, we need to push away partisanship and attack it with the interventions that work. **Things are getting worse.** We just updated <http://covidexitstrategy.org> and added a new color, "Bruised Red" to capture the new extremes that we are seeing. <https://twitter.com/rypan/status/1282020934044401664>.
- Quote Investigator (2016). A Single Death Is a Tragedy; A Million Deaths Is a Statistic. <https://quoteinvestigator.com/2010/05/21/death-statistic/>.
- 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://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>.
- Rubel, L. (2020). hey @UpshotNYT – i'm curious about your COVID choropleth maps? as the data changes, do you shift the scale so to maintain five 5 color categories? In other words, is today's red tomorrow's orange? i don't know how to see old maps to figure it out myself #maps #dataviz <https://t.co/nCYagMS6ip>. https://twitter.com/Laurie_Rubel/status/1327982127749722112.
- 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://muse.jhu.edu/article/799814>.
- Sandalow, S. (2020). ChiTribGraphics please update the color scale on your COVID by ZIP map. Practically every zip in the city has 500 cases. <https://twitter.com/lastyear/status/1290835141825265664>.
- 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).
- Shaman, J. (2022). Opinion | Here's When We Expect Omicron to Peak. *The New York Times*, <https://www.nytimes.com/2022/01/06/opinion/omicron-covid-us.html>.
- Shim, J.-B. and Son, I. (2008). An Analysis of the Middle School Students' Abilities to Recognize the Proportional Symbol Maps. *Journal of the Korean Geographical Society*, 43(4):638–654, ISSN: 1225-6633, <https://koreascience.kr/article/JAK0200801440605855.page>. Publisher: The Korean Geographical Society.
- Shrestha, S. (2022). Covid spiral. https://github.com/souryashrestha/covid_spiral.
- 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).
- Stats On Clock (2022). Coronavirus Cases Bar Chart Race (Till 2022, JAN) | COVID-19 Spread Around the World | SOC. <https://www.youtube.com/watch?v=8WH-bTH6GqI>.
- Stevens, H. (2020). These simulations show how to flatten the coronavirus growth curve. <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>.
- The Atlantic (2021). State Cases per 1M People. <https://covidtracking.com/data/charts/us-daily-positive-cases>.
- 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.
- UK EQUATOR Centre (2022). The EQUATOR Network Enhancing the QUAlity and Transparency Of Health Research. <https://www.equator-network.org/>.
- Vanderplas, S. (2022). COVID-19 data graphics archive. DOI: [10.6084/m9.figshare.21152971.v1](https://doi.org/10.6084/m9.figshare.21152971.v1).

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>.

Vuillemot, R. (2020). Sonification of COVID-19 data / Romain Vuillemot / Observable. <https://observablehq.com/@romsson/sonification-of-covid-19-data>.

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).

Washington Post (2022a). Coronavirus. <https://www.washingtonpost.com/coronavirus/>.

Washington Post (2022b). U.S. coronavirus cases: Tracking deaths, confirmed cases by state. <https://www.washingtonpost.com/nation/2020/03/27/us-coronavirus-cases-tracking-deaths-confirmed-cases-by-state/>.

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>.

World Health Organization (2022). WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int>.

Yalcin, M. (2022). Mapping the global spatio-temporal dynamics of COVID-19 outbreak using cartograms during the first 150 days of the pandemic. *Geocarto International*, 37(13):3791–3800, ISSN: 1010-6049, DOI: [10.1080/10106049.2020.1844310](https://doi.org/10.1080/10106049.2020.1844310), <https://doi.org/10.1080/10106049.2020.1844310>.

Affiliation:

Susan Vanderplas
Department of Statistics
University of Nebraska–Lincoln
349A Hardin Hall
Lincoln, NE 68583-0963, USA
E-mail: susan.vanderplas@unl.edu
URL: <https://statistics.unl.edu/susan-vanderplas>

Adalbert F.X. Wilhelm
School of Business, Social and Decision Sciences
Jacobs University Bremen gGmbH
Campus Ring 1
28759 Bremen, Germany
E-mail: a.wilhelm@jacobs-university.de
URL: <http://www.jacobs-university.de/directory/wilhelm>

<https://jdssv.org/>

published by the International Association for Statistical Computing

<http://iasc-isi.org/>

MMMMMM YYYY, Volume VV, Issue II
[doi:XX.XXXXXX/jdssv.v000.i00](https://doi.org/XX.XXXXXX/jdssv.v000.i00)

Submitted: yyyy-mm-dd
Accepted: yyyy-mm-dd
