



Data Science in a Pandemic

REVIEW

DENNIS F. X. MATHAISEL 

*Author affiliations can be found in the back matter of this article

 ubiquity press

ABSTRACT

Data Science has the potential to provide humanity with critical insight into the massive data being collected during a pandemic. The COVID-19 pandemic presented that opportunity, and Data Science supported an international audience promptly, reliably, effectively, and frequently during that difficult time. The most significant contributions were data visualizations and data dashboards, however, other tools, such as predictive and prescriptive analytics, were equally critical to the effort. The urgency at the start of the pandemic was to quickly communicate information to citizens, governments, and institutions. The change in modality from traditional statistical metrics and tables to data visualizations was extremely significant and helpful to so many. This paper reviews these contributions by demonstrating how the COVID-19 story unfolded through author-generated data visualizations and dashboards, and by providing the community with open-source access to the scripts that generated these visualizations. The open-source access to the (R language) scripts reflects this article's novelty in the literature. Using publicly available datasets from multiple sources, and employing R toolkits, the author validates the role that Data Science can play in a pandemic, and that can be implemented by anyone with some basic knowledge of scripting languages, like R. The intent is to provide these valuable tools to the community and to demonstrate their effectiveness in the likely event when there is another crisis.

CORRESPONDING AUTHOR:

Dennis F. X. Mathaisel

Research Scientist (retired)
M.I.T. and Professor of
Management Science,
Babson College, Babson Park,
Massachusetts, US

Mathaisel@Babson.edu

KEYWORDS:

pandemics; data dashboards;
data visualizations; open-
source R scripts; data sources

TO CITE THIS ARTICLE:

Mathaisel, DFX. 2023. Data
Science in a Pandemic. *Data
Science Journal*, 22: 41,
pp. 1-12. DOI: [https://doi.
org/10.5334/dsj-2023-041](https://doi.org/10.5334/dsj-2023-041)

1 INTRODUCTION

Data Science provides individuals, institutions, and regulatory authorities with insight into the massive online data they are collecting. This potential is especially critical during a crisis, and it was the COVID-19 pandemic that unleashed that potential. The problem is big data does have the tendency to overwhelm the human capacity to understand the story being told within the data (Umesh & Kagan 2015). However, with advances in easy to use ‘drag-and-drop’ software tools, such as Tableau™ and the open-source languages of Python and R, data dashboards and data visualizations are relatively easy to generate and are now prevalent. Patterns and trends, which might otherwise go undetected in tabular data or statistical measures, can be exposed more easily through data visualization technology (Rouse 2018). Dynamic visualizations of data have become effective and sought-after tools for these informative trends.

Now that the COVID-19 pandemic appears to be in the rearview, this paper acknowledges the contributions of Data Science to that emergency. The goal is to demonstrate that Data Science is indeed playing an extremely significant role in informing humanity and aiding decision-makers whenever an emergency occurs. The paper also provides a roadmap of open-source scripts¹ for others desiring to use these critical toolsets in future disasters, thereby distinguishing itself from other articles on the topic, such as Berinato (2019), Saxena et al. (2020), and Leonelli (2021). Using publicly available datasets from multiple reliable sources and employing a number of R toolkits, the author developed a series of data dashboards and visualizations for the COVID-19 story.

2 LITERATURE REVIEW

The literature is replete with articles suggesting the power of data dashboards and visualizations (e.g., Berinato 2016, 2019; Dahan et al. 2010; Hildago & Almossawi 2014; Lane 2017; Slamka et al. 2012). Three authors, Matzler (2013), and, also suggest that, despite advances in predictive analytics, forecasting during an emergency is still challenging. Attempts to predict the path of the COVID-19 outbreak (Ayyoubzadeh et al. 2020) serve as an example. Rather, Post, Nielson, and Bonneau (2003) and Friedman (2008) recommend using data visualization techniques to quickly assimilate a critical situation and employ predictive models to statistically validate the visual trends. As a result, the focus has shifted toward ‘visualizations that really work’ (Berinato 2016; Lane 2017).

Even before the World Health Organization (WHO) declared the SARS-CoV-2 virus as a pandemic (Gumbrecht & Howard 2020), the world of Data Science was actively collecting COVID data (Kent 2020). Articles attempting to digest the data flooded the media (Chu 2020; Counts 2020; Danielson 2020; IHME 2020; Mayo 2020; Roser & Ritchie 2020). Many more data sets and articles have appeared since then, but what is noteworthy is that nearly every presentation of the data made use of data dashboards and visualizations for communication and diagnosis of the pandemic. Data Science exploded in notoriety, and never has this technology been so significant to so many at the same time: researchers, health care professionals, policy-makers, academics, decision-makers, and ordinary citizens. It is most likely the first case where visualizations were widely used in critical global decision-making. *The Washington Post*’s ‘corona simulator’ article in March 2020 (Stevens 2020) was the most-viewed article ever written by that newspaper. John Burn-Murdoch, a Senior Data-Visualization Journalist for the *Financial Times*, has seen his social media followers balloon (Gossett 2020), thanks to his visualizations (Burn-Murdoch 2020).

The ‘COVID-19 Dashboard’² provided by Johns Hopkins University & Medicine was unique and filled a void in international public health systems. It impacted the way the public accessed real-time health information (Gardner 2022). Similar data dashboards were widely replicated by governments, enterprises, and media outlets, and these dashboards are expected to change how society improves coordinated responses to future pandemics (Dong, Du, & Gardner 2020; Dong et al. 2022). As an example, Marivate and Combrink (2020) present a data dashboard to inform the public about the COVID outbreak in South Africa. The Centers for Disease Control

¹ All author-generated scripts for the visualizations are available open-source at: <https://faculty.babson.edu/mathaisel/>, then choose Data Science and Data Science Journal.

² <https://coronavirus.jhu.edu/map.html>.

and Prevention (CDC) in the United States provides a public dashboard³ to help communities assess the impact of COVID-19 and to help individual states take appropriate action. The WHO developed a dashboard⁴ to provide COVID information by country, and its Public Health and Social Measures (PHSM) program underscored the steps that need to be taken by countries, territories, and areas to enforce rules or guidelines to limit the spread of COVID-19.

3 METHODS

There are two contributions driving the methods used in this paper: the sources of publicly available data and the open-source toolkits for assimilating these data. For the former, the volume of data that agencies and institutions were tracking on COVID-19 initially overwhelmed our capacity to quickly digest the impacts. The data were not only the number of confirmed cases, but also related patient data, the virus' impacts on these patients, and the effects on nations and their economies. It takes time and the proper tools to completely understand how the virus was affecting humans and how to mitigate or remedy the ensuing problems. Big data in crises like this is so complex that traditional data-processing approaches and software tools are not as effective in assimilating the situation as Data Science toolkits. The challenges include capture, storage, analysis, search, transfer, query, updates, sourcing, and privacy. Marr (2014) simplified these issues into the Five V's: Volume, Velocity, Variety, Veracity, and Value. Adherence to the Five V's also implies that the data must be 'tidy', a reference to the cleanliness of the data.

A number of open-source COVID data repositories were created by reputable organizations. Table 1 provides a comprehensive list of those data sources.

Table 1 COVID-19 public data sources.

SOURCE	URL
Johns Hopkins University	2018 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE
World Health Organization	https://covid19.who.int/
European Centre for Disease Prevention and Control	https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
European Centre for Disease Prevention and Control	https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide
The Covid Tracking Project	https://covidtracking.com/
U. S. Centers for Disease Control and Prevention (CDC)	https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html
ESRI (Environmental Systems Research Institute) COVID-19 Resources	https://coronavirus-disasterresponse.hub.arcgis.com/
COVDATA: a data package for R that collects and bundles datasets related to the COVID-19 pandemic from a variety of sources.	https://kjhealy.github.io/covdata/index.html
BeOutBreakPrepared-nCoV2019	https://github.com/beoutbreakprepared/nCoV2019/tree/master/latest_data
Kaggle COVID-19 Open Research Dataset Challenge (CORD-19)	https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge
Kaggle Novel Corona Virus 2019 Dataset	https://www.kaggle.com/sudalairaikumar/novel-corona-virus-2019-dataset
Towards Data Science: Fighting the Covid-19: All the datasets and data efforts in one place	https://towardsdatascience.com/fighting-the-covid-19-all-the-datasets-and-data-efforts-in-one-place-4d6aeb0157ab
GitHub Our World in Data	https://github.com/owid
GitHub New York Times COVID-19 Data	https://github.com/nytimes/covid-19-data
GitHub CSSE (Center for Systems Science and Engineering at Johns Hopkins) COVID-19 Daily Reports	https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports
GitHub MIDAS Network Coordination Center	https://github.com/midas-network/COVID-19
University of Oxford Covid-19 Government Response Tracker	https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker

³ <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>.

⁴ <https://covid19.who.int/>.

The data contain thousands of observations with a number of attributes, some of which usually are not significant to the analysis. Thus, the first step is to investigate the data structure. To help accomplish this task, the author employed 'DataExplorer' in R, which provides a visualization of the structure (Figure 1).

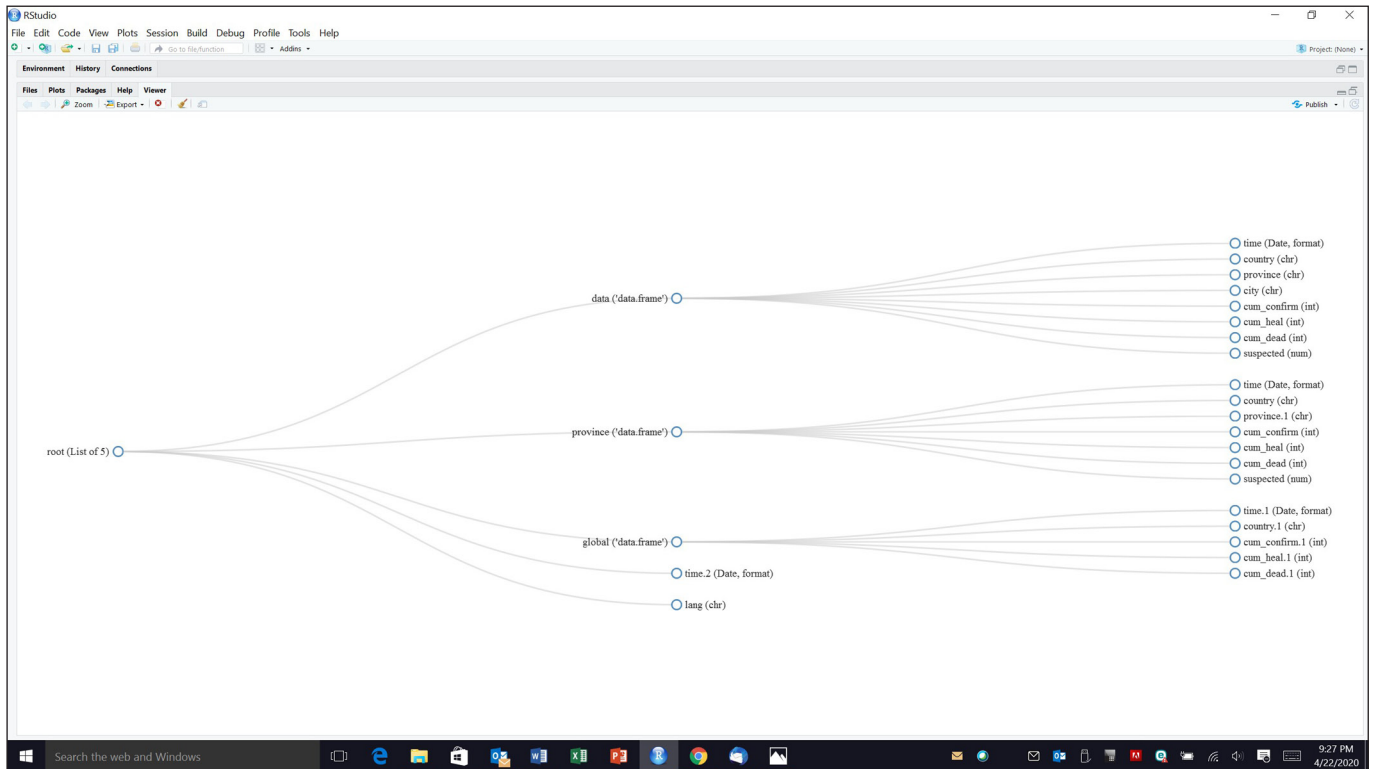


Figure 1 DataExplorer visualization of the underlying data structure (author generated).

In addition, raw data usually requires some data 'wrangling': that is, taking care of missing values, dealing with obvious outliers due to transcription errors, duplicate records, etc. All of these data tasks can be corrected using pre-defined functions in R or Python. For example, R has an excellent set of data cleaning packages in 'Tidyverse',⁵ which is worthy of investigating. In the case of all of the COVID-19 data repositories used in this paper, the organizations that provided these data already had published 'tidy' data.

Since the focus is on visualizations, of the many that are available, the question becomes: which visualizations are appropriate, and which visualizations tell the story that the data scientist wishes to analyse? Every visualization tells a different story, and every scientist interprets the visualization differently. Thus, there can be multiple ways to visualize data. As a guide, Lengler and Eppler (2007) developed a systematic overview of 100 visualizations. Their efforts culminated in a structure for the visualizations that replicated the logic, look, and use of the periodic table of elements in chemistry.

4 ANALYSIS OF A CRISIS

4.1 STATIC VISUALIZATIONS

Data were confirming the exponential growth of COVID that is to be expected with a pandemic. After the first trickle of confirmed cases, the trickle turned into a torrent (Stevens 2020). From the onset, the world fixated on its exponential growth, and the WHO began to underline the importance of 'flattening the curve' (Roberts 2020). The result was a study of what will undoubtedly become one of the most infamous data visualizations recorded in the public media for COVID. China became the first country to provide the curve and then attempt to flatten it (Figure 2).

Since the WHO declared the pandemic, every country followed China (Figure 3).

⁵ The tidyverse (<https://www.tidyverse.org/>) is a coherent system of packages for data manipulation, exploration and visualization that share a common design philosophy.

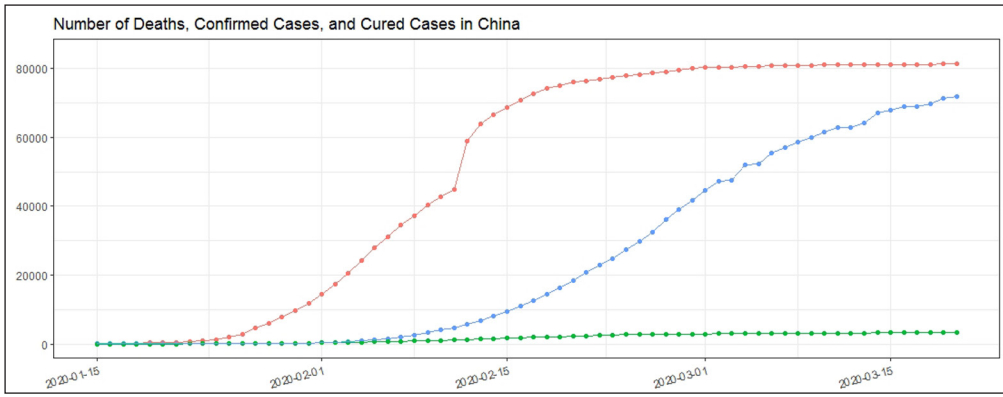


Figure 2 Number of deaths, confirmed cases, and cured cases in China (author generated).

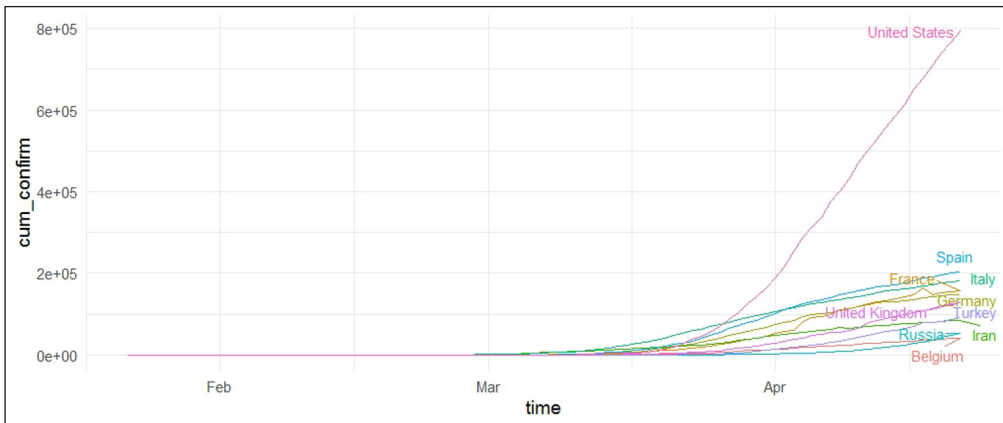


Figure 3 Total number of cases world-wide except China (author generated).

A different variation of this story can be told, not through traditional bar graphs, but through an adaptation of a 'wind rose' to visually depict the spread by country (Figure 4). In this wind rose adaptation, the United States, the country with the highest number of confirmed cases (highest speed), is represented as a hybrid of a bar chart and a pie chart.

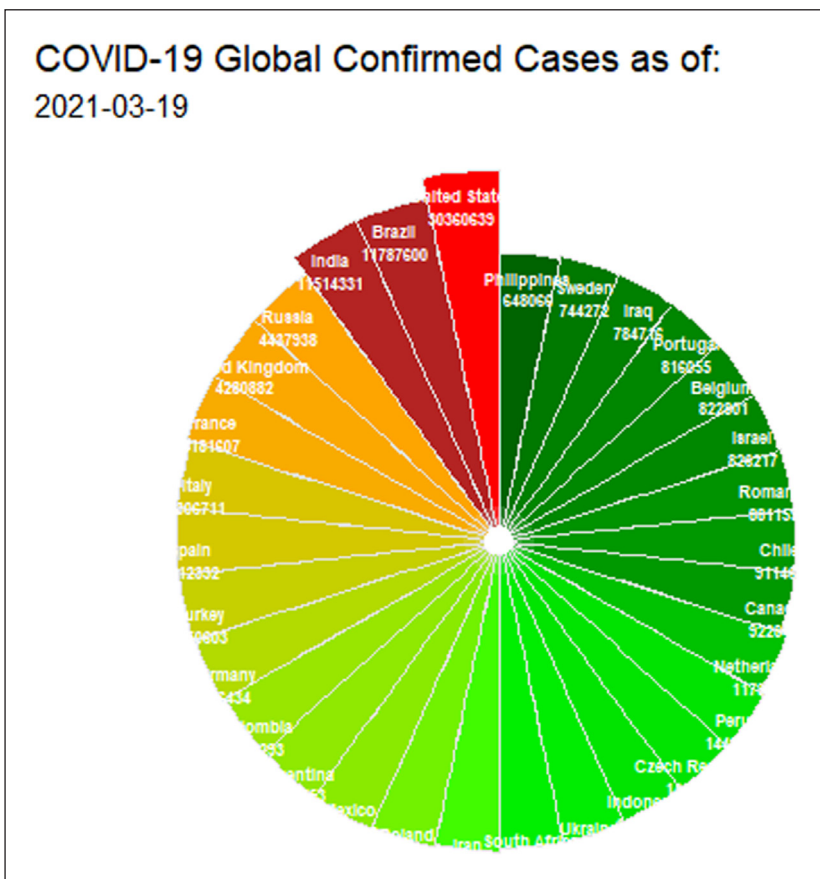


Figure 4 Global confirmed cases by country using a wind rose (author generated).

To depict location, the story can be told through another familiar and popular visualization: a geographic map (Figure 5).

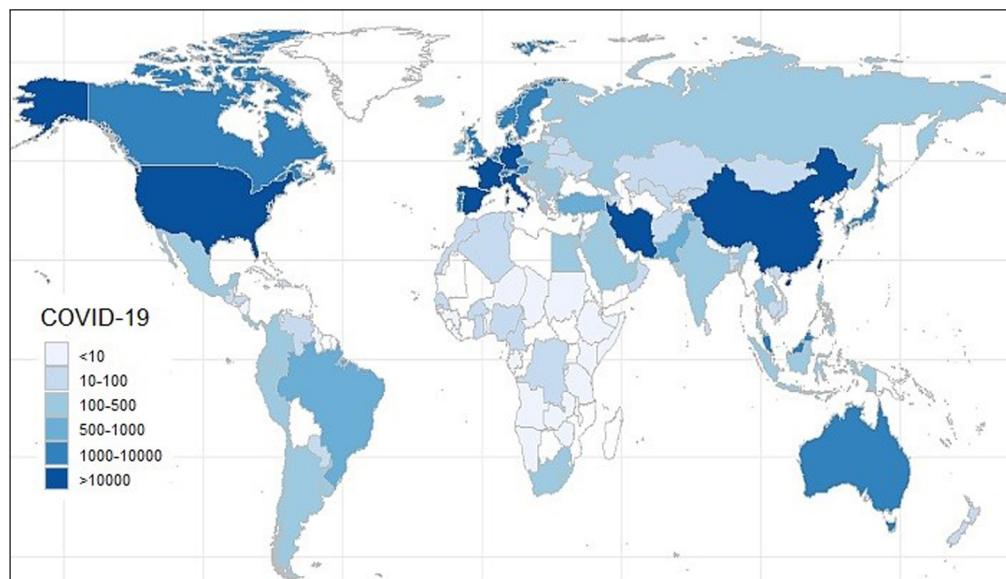


Figure 5 Global confirmed cases by country using a geographic map (author generated).

The velocity and the direction/location of the virus were only part of the story. Perhaps the most significant part was the medical and economic risk. Risk is the statistical story of variation. For that, a violin plot, which is a hybrid of a box plot and rotated kernel density plot, reveals both the density and spread of the virus. Figure 6 is a series of violin plots for the six leading countries in terms of confirmed cases. The United States leads the five other countries represented.

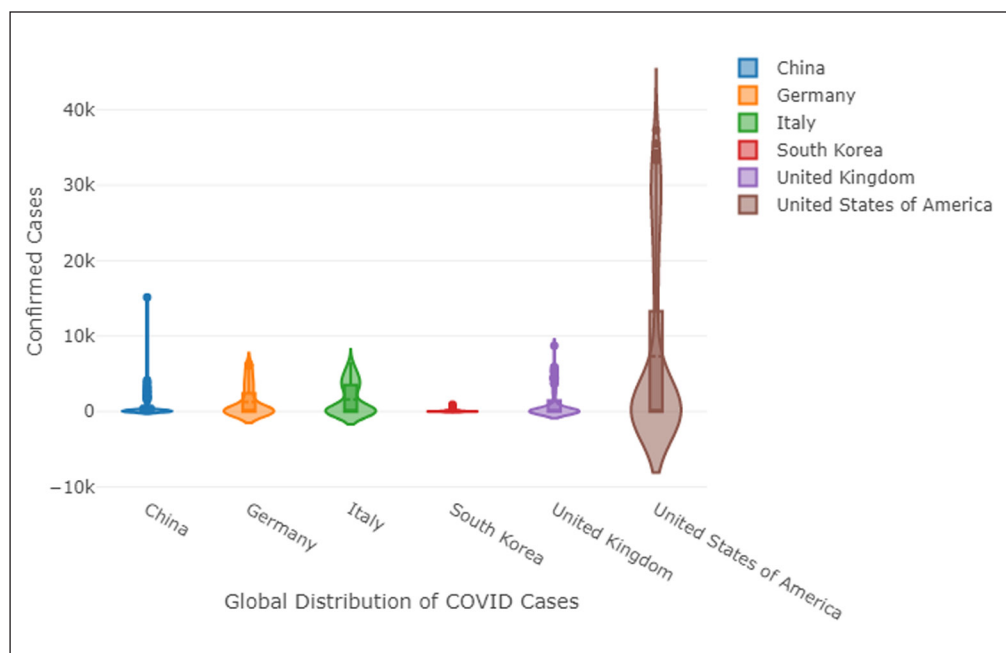


Figure 6 Global confirmed cases for selected countries using a violin plot (author generated).

The story imbedded in the violin plot of Figure 6 is not only the peak of the violin but also the width, or variation, in the number of cases for each country. Compare, for example, South Korea to the others, and compare the United States to the European countries.

Since the populations of these countries are very different, one begins to question the rate per capita. Figure 7 is the same violin plot as Figure 6, except that it shows the number of confirmed cases *per capita* for the top 10 countries, and Figure 8 is the number of deaths *per capita*. In the visualization of Figure 7, Spain and the United States reveal the highest peaks of the violin. The United States does fairly better in Figure 8, which showed deaths.

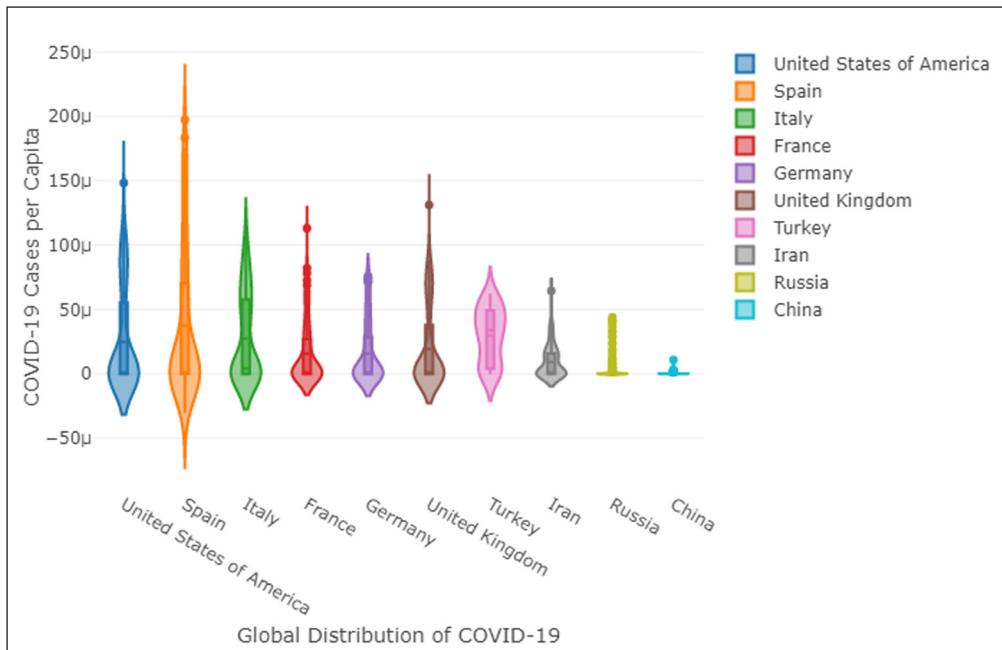


Figure 7 Confirmed cases per capita for selected countries using a violin plot (author generated).

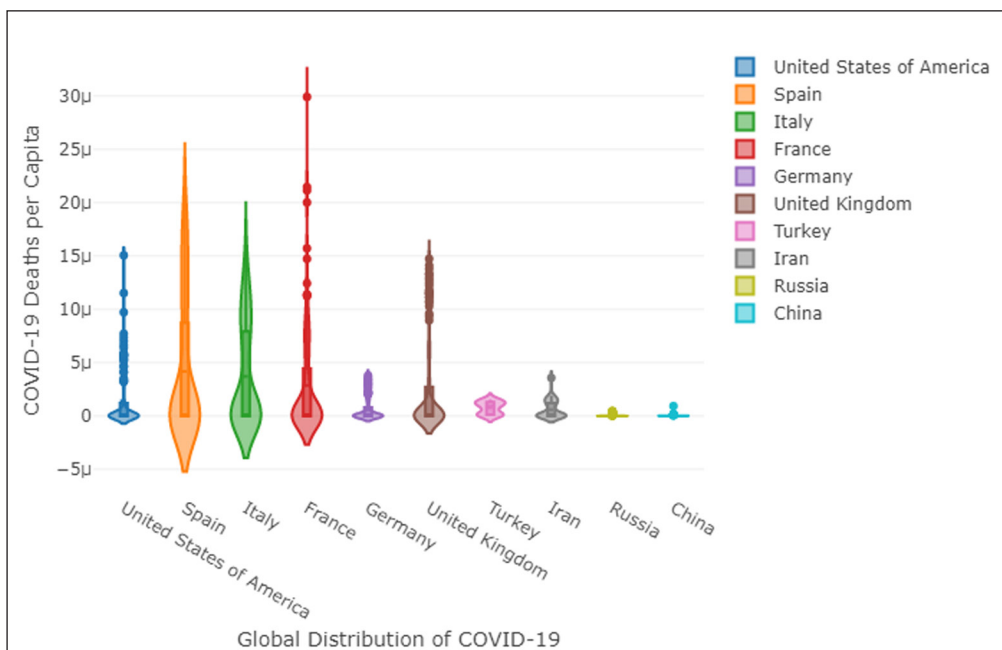


Figure 8 Confirmed deaths per capita for selected countries using a violin plot (author generated).

4.2 INTERACTIVE VISUALIZATIONS

While static charts can tell a compelling story at a point in time, researchers, medical professionals, decision-makers, and policy makers often seek to further investigate trends. Interactive visualizations allow investigators to directly interact with the ‘live’ data. As an example, Figure 9 shows two kernel density plots of the number of confirmed cases. The graph on the left has a high bandwidth⁶ for the density estimation, and it appears to show a rather normal distribution of the number of cases. The data include all countries. That is a clue because when the bandwidth is relatively high, it obscures much of the underlying data structure. When the bandwidth is adjusted, as shown on the right, the plot reveals that the distribution is not at all normal. Rather, it reveals more than one wave. Since multiple countries are represented, the first wave is actually China, followed by Italy, and then the other countries. When the chart was first created in March 2020, the United States was a small wave. Since then, the United States surpassed the other countries, but it demonstrates that being able to interact with the data dynamically reveals the underlying picture.

⁶ The bandwidth on a kernel density plot controls how closely the density matches the underlying distribution.

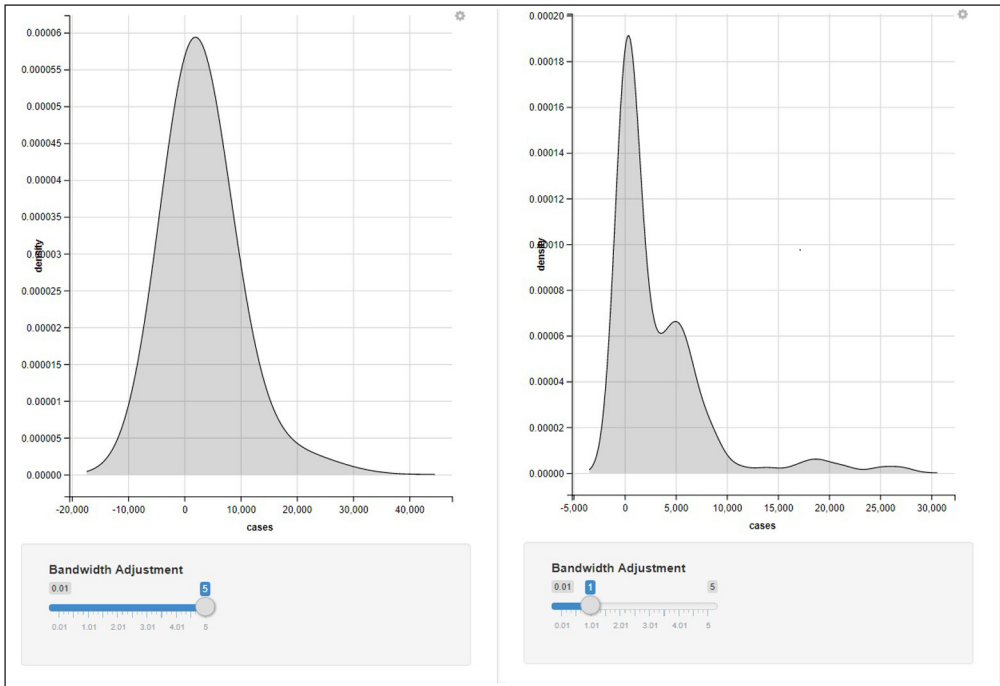


Figure 9 Global confirmed cases by country using an interactive density chart (author generated).

4.3 NETWORKED VISUALIZATIONS

As illustrated above, there are numerous and different ways of creating a visualization, like an artist, but, in this era of rapid dissemination of information, communicating the story to a wide audience often involves the web. Shiny is an R package that makes it relatively easy to build interactive web apps. For example, [Figure 10](#) is a horizontal bar chart arranged by the decreasing number of confirmed deaths in countries outside of China. The chart, produced in a Shiny dashboard, shows that the United States quickly surpassed Spain and Italy in the number of fatalities.

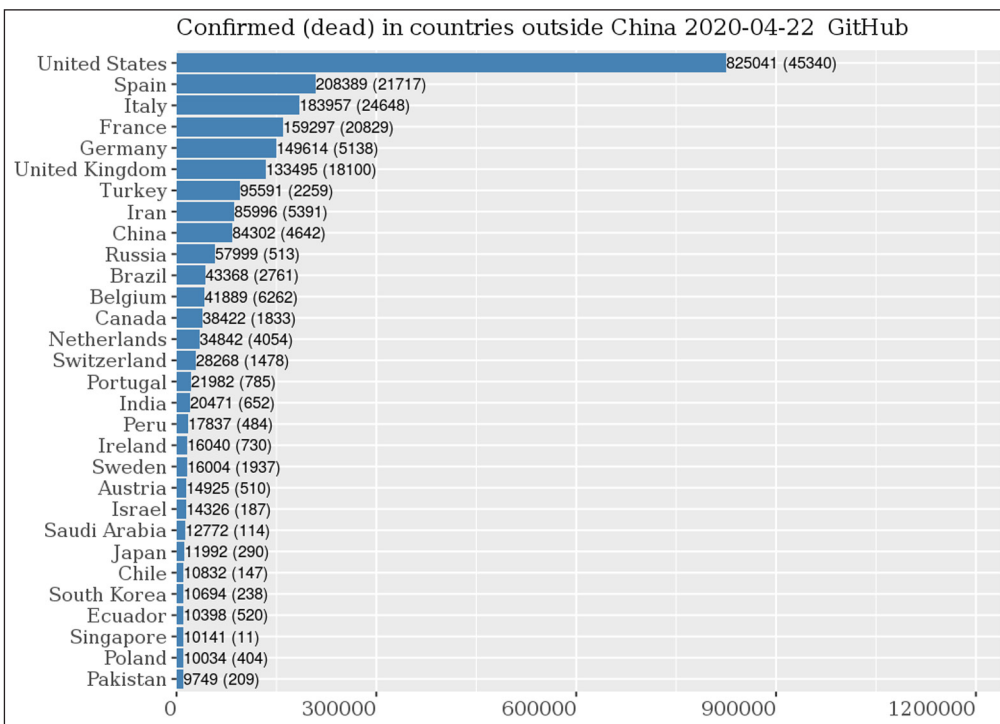


Figure 10 Global confirmed deaths by country using a Shiny app horizontal bar chart (author generated).

The interactive app of [Figure 11](#) shows the trajectories of growth in different countries, generated for the number of confirmed cases by the number of days since the 100th confirmed case. The data analyst can select an area to zoom in and/or ‘mouse over’ to see the name of the country, as illustrated in the figure. At the time the visualization was created, all of the countries were on a similar exponential growth trajectory. Unlike China, the United States seeded the virus broadly but is also showing signs of slowing down, that is, ‘flattening the curve’.

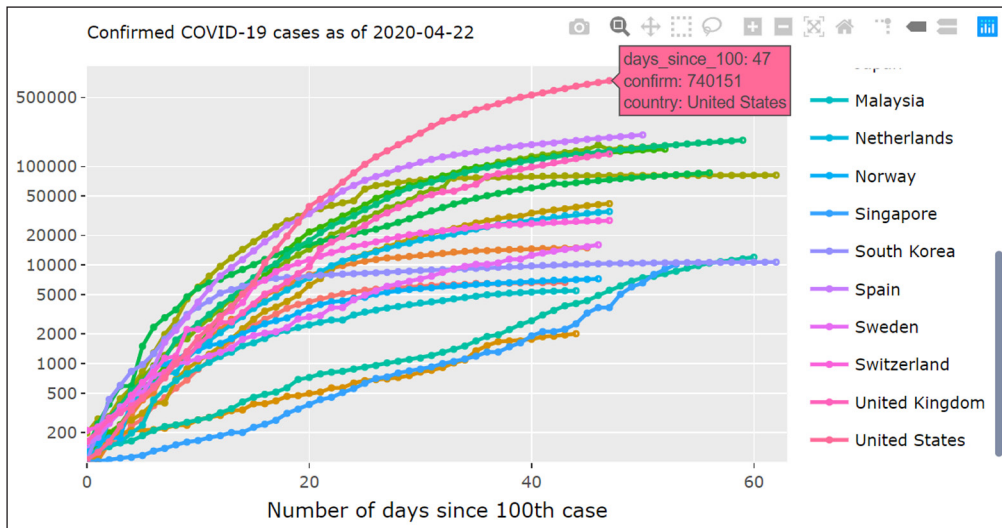


Figure 11 Confirmed COVID-19 cases by country using an interactive line chart web app (author generated).

Figures 10 and 11 were created using the Shiny ‘dashboard’ command in R.

This web-based interactive visualization tool is a great example of how Data Science and visualizations can help identify the volume, variety, and veracity of data (Bresciani & Eppler 2009; Eppler & Platts 2009; Hildago & Almosawi 2014; Romer 2015). The visualizations help to mitigate data challenges by summarizing the vast amounts of data (Volume), simplifying the different types of data (Variety), and identifying errors (Veracity). However, the biggest challenge is transforming the data into the fifth ‘V’, Value, especially when the investigation involves a pandemic. An interactive web visualization app can be a powerful tool for a media-hungry audience. In a recent *Harvard Business Review* article, Berinato (2019) maintains that the technical part of putting together charts has advanced rapidly by teams who develop them, but he advocates more focus needs to be on the presentation of the data to public audiences. That is the role of data visualizations. In other words, interactive visualizations are the tools that will allow an audience to better assimilate the complex data they are being given (Boost Labs 2015).

4.4 DATA DASHBOARDS

Dashboards were very helpful during the pandemic. Figure 12 is an author generated re-imagining of a typical COVID dashboard, the COVID-19 Global Tracker Dashboard (Schönenberger 2020). The purpose of dashboards is to track the pandemic in real-time in both a tabular and visual form.

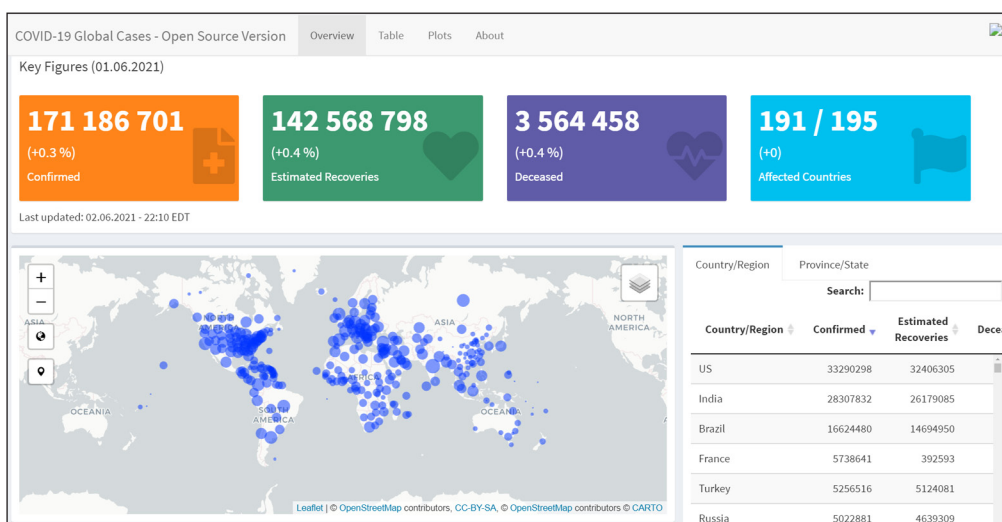


Figure 12 Global tracking dashboard (author generated).

The dashboard presents cases, recoveries, and deaths for 195 states. It utilized data from the Johns Hopkins CSSE repository (listed in Table 1). During the peak of the pandemic, the country with the highest cumulative number of confirmed case counts was the United States.

The coronavirus pandemic drew attention to the contributions of Data Science to a humanitarian crisis. Even before the WHO declared the virus as a pandemic, numerous organizations began actively collecting and distributing data on the virus and its effects on people. As a result, data visualizations and dashboards became critical tools in educating the public and supporting healthcare professionals, researchers, and decision-makers. Properly generated visualizations and dashboards help mitigate the challenges of summarizing the vast amounts of data (Volume), simplifying the different types of data (Variety), and identifying errors (Veracity), which are some of the ‘Five V’s’ Marr (2014) associated with live data. In addition, the greatest challenge is transforming the data into Value, especially when the timing of decisions is critical. Visualizations and dashboards provide value and make a significant contribution to our insight into crises. This paper reviews these contributions by demonstrating how the COVID-19 story unfolded through author-generated data visualizations and dashboards and by providing the community with open-source access to the scripts that generated these visualizations.⁷ Using publicly available datasets from multiple sources and employing R toolkits, the author demonstrates the role that Data Science can play in a pandemic, and that can be implemented by anyone with some basic knowledge of scripting languages like R.

ACKNOWLEDGEMENTS

The author wishes to thank two outstanding individuals, Abdullah Zahid, Data Analyst, ModusLink Corporation, and Dong Hyun (Veo) Chae, Georgetown University, for their technical advice and contributions to this paper.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Dennis F. X. Mathaisel  orcid.org/0000-0001-9093-9844

Research Scientist (retired) M.I.T. and Professor of Management Science, Babson College, Babson Park, Massachusetts, US

REFERENCES

- Ayyoubzadeh, SM, Ayyoubzadeh, SM, Zahedi, H, Ahmadi, M and Niakan Kalhori, SR.** 2020. Predicting COVID-19 incidence through analysis of Google trends data in Iran: Data mining and deep learning pilot study. *JMIR Public Health Surveillance*, 6(2): e18828. DOI: <https://doi.org/10.2196/18828>
- Berinato, S.** 2016. Visualizations that really work. *Harvard Business Review*, June. Available at <https://hbr.org/2016/06/visualizations-that-really-work> [Last accessed July 2018].
- Berinato, S.** 2019. Data science and the art of persuasion. *Harvard Business Review*, January–February. Available at <https://hbr.org/2019/01/data-science-and-the-art-of-persuasion> [Last accessed April 2020].
- Boost Labs.** 2015. *New digital marketing trend: Interactive data visualizations*. Available at <http://www.boostlabs.com/new-digital-marketing-trend-interactive-data-visualizations/> [Last accessed July 2018].
- Bresciani, S and Eppler, MJ.** 2009. The benefits of synchronous collaborative information visualization: Evidence from an experimental evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6): 1073–1080. DOI: <https://doi.org/10.1109/TVCG.2009.188>
- Burn-Murdoch, J.** 2020. Coronavirus tracked: The latest figures as the pandemic spreads. *Financial Times*, Available at <https://www.ft.com/coronavirus-latest> [Last accessed 19 April 2020].
- Chu, J.** 2020. 3 Questions: Catherine D’Ignazio on visualizing COVID-19 data. *MIT News*, Available at <http://news.mit.edu/2020/catherine-dignazio-visualizing-covid-19-data-0414> [Last accessed 18 April 2020].

⁷ All author-generated scripts for the visualizations are available open-source at: <https://faculty.babson.edu/mathaisel/>, then choose Data Science and Data Science Journal.

- Counts, C.** 2020. 7 Data visualizations that explain COVID-19. *EMS1*, Available at <https://www.ems1.com/ems-products/software/articles/7-data-visualizations-that-explain-covid-19-lmM1yqR0kwfnd5BE/> [Last accessed 18 April 2020].
- Dahan, E, Soukhoroukova, A and Spann, M.** 2010. New Product Development 2.0: Preference Markets - How Scalable Securities Markets Identify Winning Product Concepts and Attributes. *The Journal of Product Innovation Management*, 27(7): 937–954. DOI: <https://doi.org/10.1111/j.1540-5885.2010.00763.x>
- Danielson, M.** 2020. Notable maps visualizing COVID-19 and surrounding impacts. *Mapbox*, Available at <https://blog.mapbox.com/notable-maps-visualizing-covid-19-and-surrounding-impacts-951724cc4bd8> [Last accessed 17 April 2020].
- Dong, E, Du, H and Gardner, L.** 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infectious Diseases*, 20(5): 533–534. DOI: [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Dong, E, Ratcliff, J, Goyea, TD, et al.** 2022. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 dashboard: Data collection process, challenges faced, and lessons learned. *Lancet Infectious Diseases*, 22(12): E370–E376. DOI: [https://doi.org/10.1016/S1473-3099\(22\)00434-0](https://doi.org/10.1016/S1473-3099(22)00434-0)
- Eppler, M and Platts, K.** 2009. Visual strategizing: The systematic use of visualization in the strategic planning process, long range planning LRP. *International Journal of Strategic Management*, 42(1): 42–74. DOI: <https://doi.org/10.1016/j.lrp.2008.11.005>
- Friedman, V.** 2008. Graphics. *Monday Inspiration*, January 14. Available at <https://www.smashingmagazine.com/2008/01/monday-inspiration-data-visualization-and-infographics/> [Last accessed July 2018].
- Gardner, L.** 2022. The COVID-19 dashboard for real-time tracking of the pandemic: The Lasker-Bloomberg Public Service Award. *Journal of the American Medical Association*, 328(13): 1295–1296. DOI: <https://doi.org/10.1001/jama.2022.15590>
- Gossett, S.** 2020. Coronavirus charts are everywhere. But are they good? Visualizations can illuminate, but they can also be misleading. *BuiltIn*, April 15. Available at <https://builtin.com/data-science/visualization-lessons-pandemic> [Last accessed 18 April 2020].
- Gumbrecht, J and Howard, J.** 2020. WHO declares novel coronavirus outbreak a pandemic. *CNN*, March 11. Available at <https://www.cnn.com/2020/03/11/health/coronavirus-pandemic-world-health-organization/index.html> [Last accessed 19 April 2020].
- Hildago, C and Almossawi, A.** 2014. The data visualization revolution. *Scientific American*, March 14. <https://www.scientificamerican.com/article/the-data-visualization-revolution/> [Last accessed January 2018].
- Institute for Health Metrics and Evaluation (IHME).** 2020. *COVID-19 projections*. University of Washington, Seattle, WA. Available at <http://www.healthdata.org/data-visualization/covid-19-us-state-state-projections> [Last accessed 18 April 2020].
- Kent, J.** 2020. Understanding the COVID-19 pandemic as a big data analytics issue. *Health IT Analytics*, Available at <https://healthitanalytics.com/news/understanding-the-covid-19-pandemic-as-a-big-data-analytics-issue> [Last accessed 18 April 2020].
- Lane, A.** 2017. Data visualization top trends for 2017. *Innovation Enterprise*, Available at <https://channels.theinnovationenterprise.com/articles/data-visualization-top-trends-for-2017> [Last accessed January 2018].
- Lengler, R and Eppler, M.** 2007. Towards a periodic table of visualization methods for management. In: *IASTED Proceedings of the Conference on Graphics and Visualization in Engineering (GVE 2007)*, Clearwater, FL. Available at http://www.visual-literacy.org/periodic_table/periodic_table.html [Last accessed July 2018].
- Leonelli, S.** 2021. Data science in times of pan(dem)ic. *Harvard Data Science Review*, 3(1). DOI: <https://doi.org/10.1162/99608f92.fbb1bdd6>
- Marivate, V and Combrink, HM.** 2020. Use of available data to inform the COVID-19 outbreak in South Africa: A case study. *Data Science Journal*, 19(1): 19. DOI: <https://doi.org/10.5334/dsj-2020-019>
- Marr, B.** 2014. Big data: The 5 V's everyone must know. *LinkedIn*, Available at <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/> [Last accessed July 2018].
- Matzler, K, Grabher, C, Huber, J and Fuller, J.** 2013. Predicting New Product Success with Prediction Markets in Online Communities. *R & D Management*, 43(5): 420–432. DOI: <https://doi.org/10.1111/radm.12030>
- Mayo, M.** 2020. *COVID-19 visualized: The power of effective visualizations for pandemic storytelling*. March. Available at <https://www.kdnuggets.com/2020/03/covid-19-visualized.html> [Last accessed 18 April 2020].
- Post, FH, Nielson, G and Bonneau, G-P.** (eds.) 2003. *Data visualization: The state of the art*. New York: Kluwer Academic Publishers, Springer. DOI: <https://doi.org/10.1007/978-1-4615-1177-9>
- Roberts, S.** 2020. Flattening the coronavirus curve. *The New York Times*, March 27. Available at <https://www.nytimes.com/article/flatten-curve-coronavirus.html> [Last accessed 8 April 2020].

- Romer, B.** 2015. A picture is worth more than a thousand words. *Exchange Magazine*, Available at http://exchangemagazine.financial.thomsonreuters.com/articles/a-picture-is-worth-more-than-a-thousandwords?adbid=619572992046043136&adbpl=tw&adbpr=14412844&cid=social_20150710_48979916 [Last accessed 18 April 2020].
- Roser, M and Ritchie, H.** 2020. *Coronavirus disease (COVID-10) – The data. Our world in data.* Available at <https://ourworldindata.org/coronavirus-data> [Last accessed 18 April 2020].
- Rouse, M.** 2018. Definition: Data visualization. *TechTarget*, With contributions from Ed Burns. Available at <https://searchbusinessanalytics.techtarget.com/definition/data-visualization> [Last accessed August 2018].
- Saxena, N, Gupta, P, Raman, R and Rathore, AS.** 2020. Role of data science in managing COVID-19 pandemic. *Indian Chemical Engineer*, 62(4): 385–395. DOI: <https://doi.org/10.1080/00194506.2020.1855085>
- Schönenberger, C.** 2020. COVID-19 dashboard. *GitHub*, Available at https://github.com/chschoenenberger/COVID19_dashboard [Last accessed 1 April 2020].
- Slamka, C, Jank, W and Skiera, B.** 2012. Second-Generation Prediction Markets for Information Aggregation: A Comparison of Payoff Mechanisms. *Journal of Forecasting*, 31(6): 469–489. DOI: <https://doi.org/10.1002/for.1225>
- Stevens, H.** 2020. Why outbreaks like coronavirus spread exponentially, and how to ‘flatten the curve’. *Washington Post*, March 14. Available at <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/> [Last accessed 18 April 2020].
- Umesh, UN and Kagan, M.** 2015. Data visualization in marketing. *Journal of Marketing Management*, 3(2): 39–46.

TO CITE THIS ARTICLE:

Mathaisel, DFX. 2023. Data Science in a Pandemic. *Data Science Journal*, 22: 41, pp. 1–12. DOI: <https://doi.org/10.5334/dsj-2023-041>

Submitted: 02 March 2023

Accepted: 03 October 2023

Published: 30 October 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.