11/7/2022

Ian Wolloff

**Literature Review: Data Visualisation for COVID-19 Data Analysis**

Table of Contents

[Background 2](#_Toc118713988)

[Introduction 2](#_Toc118713989)

[Visualization History 2](#_Toc118713990)

[Data Visualization 3](#_Toc118713991)

[Data and Presentation 5](#_Toc118713992)

[Problems with dashboards. 6](#_Toc118713993)

[Conclusion 6](#_Toc118713994)

[References 7](#_Toc118713995)

**Literature Review: Data Visualisation for COVID-19 Data Analysis**

## **Background**

My current role includes working with PCTs on data analysis & Business Intelligence (BI) for this reason I have chosen COVID 19 Data Analysis as it is something I have real world experience of doing. While we should be able to do a review on any topic using the techniques in this module there is value in doing a subject that you already have some prior knowledge on or in other words.

**Without the prior knowledge, the new learning can't be understood; it can only be vaguely 'remembered'** (EBTN 2022)

## **Introduction**

We are going to look at COVID19 and how data Visualization techniques have been used through the Covid 19 Pandemic which does have its own irony as during the production of this paper I am now part of those very figures. From the now famous “Next Slide Please” used at the daily Covid briefing via custom data Visualizations used to show the spread of the virus to detailed dashboards and Visualizations that are now part of all our shared experience of the pandemic. The purpose of visualizations is to put raw data into a form that can be digested the use of Visualization is primary used to aid the Interpretation of complex data sets.

Though a new field due to the global pandemic that was a large focus on the topic because of this in the short period since the start of the pandemic there has been a significant amount of research on the use of Visualization technology in regard to Covid 19. Also relevant to this study is the wider use of Visualization techniques for the distribution of complex data to non SMEs (Subject Matter Experts).

## **Visualization History**

Before 1700 the realm of data visualization was restricted to cartography the construction of Mapping data with examples dating back to 25000 BC. Wolodtschenko, Alexander; Forner, Thomas (2007) with the first what could be described as modern map being the “Imago Mundi Babylonian Map of the World” created around the 9th Century BC. (Delnero, Paul 2017) However over time there was a requirement for more precise mapping including features such the depths and measurements required for ocean navigation. Michael Florent Van Langren born Amsterdam (Netherlands) 27th April 1598 is believed to have produced the first recorded pictorial illustration of numerical data, from which the science of data analysis and visualization is derived.



**Fig 1 [Imago Mundi. Babylonian Map]**

## **Data Visualization**

Data Visualization which is sometimes called “story telling with data” Nussbaumer Knaflic, C. (2015). as the aim is to provide a story to allow the interpretation of source data.

in the publication Nature (2020) (Helen Kennedy) says on the topic of Covid and the use of Visualization “The language of data visualization has become commonplace, but their influence on public opinion and behaviour is unclear”. The paper makes the point that “visualisations promote greater understanding of data by making them [accessible and transparent](http://www.perceptualedge.com/articles/visual_business_intelligence/what_people_need_from_infovis.pdf), or that, through visualisation” (Kennedy Helen 2020)

Although the points made appear valid based on the arguments put forward the paper it also makes the counter point that “The data on which visualisations are based are also not neutral. Human decisions influence and shape data, as well as their visual representation”. From my own experience of working with data and carrying out data analysis Data Bias is a real thing and can be problematic especially when dealing with complex data sets. In the case of Covid 19 a dataset that can also generate an emotive response. Because of this data analysts also need to factor in a Public Health messaging and Political element to any messaging presented via the use of Covid19 Datasets.

The use of data visualizations does also have potential downsides Aengus Bridgm (Harvard 2019) wrote “The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media” makes the point that “The COVID-19 pandemic has been accompanied by a so-called infodemic a global spread of misinformation that poses a serious problem for public health. Infodemic’s are concerning because the spread of false or misleading information has the capacity to change transmission patterns (Kim et al., 2019)”

The WHO (World Health Organization) as part of their Stop the spread campaign (WHO 2021) have used COVID19 infographics and data analyses to combat misinformation over the Covid 19 Vaccine they say “campaigns have included an array of social media infographics and messages across platforms to explain the safety of COVID-19 vaccines.” Or

***“The number of people who think they understand statistics dangerously dwarfs those who actually do, and maths can cause fundamental problems when badly used.”   
― Rory Sutherland,***[***Alchemy: The Dark Art and Curious Science of Creating Magic in Brands, Business, and Life***](https://www.goodreads.com/work/quotes/46190767)

Pure Information and raw data are only part of the answer and data visualisations and dashboards do not always provide the whole answer. In the paper Chasing John Snow: data analytics in the COVID-19 era an assessment of Covid 19 (Pietz, McCoy and Wilck, 2020) the paper investigated twenty plus dashboards and Visualizations used in various locations around the world the paper looked at the different data shaping methodologies when producing the Dashboard outputs. The paper makes a number of recommendations for the use of Dashboards

One of the main parts of the study was showing how by the use of data and data presentation techniques, it allows the current state of the pandemic to be shown and to act as a driver for decision makers in both the health arena and political arena to base decisions on. Problems however do exist when the dashboards that get presented do not contain the required information to make accurate decisions.

The UK Office of National Statistics (ONS) as the main agency reasonable for the data analysis of the UKs covid response carries out and provides overview graphics of covid 19 in the United Kingdom using such common data points as total covid deaths number of people in hospital number of people in ICU, number of infections recorded the type of data collection that was used by the ONS is one based on a standard statistical analysis approach along with the results of the human focused covid surveys.

* how many people across England, Wales, Northern Ireland, and Scotland would have tested positive for a COVID-19 infection, regardless of whether they report experiencing symptoms
* the average number of new cases per week
* the number of people who would have tested positive for antibodies to SARS-CoV-2 at different levels (Steel Kara & Yapp Rhiannon 2022)

From these datasets a series of visualisations are produced. The raw data is also exposed using a JSON API to allow ingestion of the raw data so it can be presented in any data visualisations tooling that supports the import of JSON formatted data.

Outside of the United Kingdom the WHO (World Health Organization) produce a series of covid dashboards focusing on deaths and vacancies, this provides a way for the performance of the health response between countries to be measured using a common dataset and a common set of dashboards and visualisations that allow a non technical people to see and understand the sometimes complex datasets that make up the worldwide response to Covid19.

The WHO however do say “Caution must be taken when interpreting all data presented, and differences between information products published by WHO, national public health authorities, and other sources using different inclusion criteria and different data cut-off times ” (WHO 2020) so when producing visualisations based on datasets produced using different inclusion criteria care must be taken to ensure parity between the datasets so that any conclusions drawn are accurate and represent the actual ground truth rather than an artifact of the data presentation.

Falisse, J.-B., & McAteer, B. (2022) say “Data visualisations are intimately connected to the emergence of public health as a discipline and policy area. Besides the mapping of cases and deaths, the COVID-19 pandemic has seen an explosion of attempts to track policy responses.” And go on to make the point that “Public health researchers and policymakers have a long history of using visuals to make sense of information about the spread and mortality of disease over time and by geographical area. For example, Florence Nightingale famously used visual diagrams to show that preventable disease was a more common cause of death among wounded soldiers in the Crimean war than the wounds themselves”

## **Data and Presentation**

If we look at the data used to generate Dashboards and visualizations F. Clement in the paper Visualising policy responses during health emergencies. Learning from the COVID-19 policy trackers. Says “Currently, there are many forms of visualization available for COVID-19 through various research centres, federal and provincial websites”

they go on to say that there are two main forms of visualizations that are produced in regard to the display and study of Covid 19. These are

1. “Visualizations that are presented in an easily consumable format. that are simple and clear, hosted on public web servers, dynamically updated, and can be viewed by the public. Visualizations in this category provide general information about the cases worldwide, as well as at the country and state/province level” (F. Clement)

Examples of these are dashboards such as the ones provided by NHS in the UK or GGD-GHOR Nederland Association of GGD's (Regional Public Health Services) in the Netherlands. Or the CDC (Centre of Disease Control) in the United States.

* <https://coronavirus.data.gov.uk>
* <https://coronadashboard.government.nl>
* <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>

In regard to best practice and visualizations, the if we look at the paper Best Practices for COVID-19 Data Visualizations (Gardner Betsy Harvard 2019) The paper makes a number of observations and recommendations. The recommendation is to use “clear language and terms” in order for non professionals to be able to consume the data being presented. The second recommendation is to show a range of data rather than a more limited point in time representation of the data to allow the data to be put in context. Thus, also allowing patterns and trends to be identified. The third recommendation was to show population totals to allow comparison of the data based on demographics. The example quoted in the paper of an example of good practice was the New York covid tracker.

1. The second category are in the form of reports generated by Business, research & Think Tanks some examples given were (IBM, Accenture, Deloitte) as the sorts of organizations producing reports. The reports rather than being for the study and understanding of Covid19 in a public health context were designed to help businesses and their leaders understand the impact of COVID-19 on the economy and businesses. These tend to use more complex visualizations as the intended audience is of a more specialised domain with greater realm knowledge.

In the paper the authors suggest a third way their dashboard while having a lot of the features of a type 1 dashboard would also contain elements of type 2 to achieve this the authors propose the use of the python development environment leveraging the following libraries



## **Problems with dashboards.**

Dashboards can have issues especially in the context of international datasets where the visualizations look at COVID19 rates between jurisdictions these were looked at in the Lancet article Dong, E., Ratcliff “Johns Hopkins University Centre for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned 2022. One of the main issues identified was the difficulty in producing an accurate high quality dataset and therefore visualizations of that data when there is not an international standard for the recording and reporting of COVID19 data.

There are also problems affecting dashboards in general including COVID19 ones.

* Poor User Experience (UX)
* Static experience not allowing users to focus on data they are interested in
* Data Quality & Data Accuracy
* Independent Validation of data Assumptions

## **Conclusion**

We looked at some significant issues pertinent to the use of Data visualizations in relation to Covid 19. The opinion on the use of visualizations is mixed but as a new emerging area of study this is not uncommon the important question to ask is does the use of visualizations when taking about Covid19 enhance the understanding and decision making around COVID19 and from the majority of papers in this study and the ones that did not make this study my personal opinion is the direction of travel in this area is that the use of visualizations does indeed aid understanding of COVID19 but to truly have a definitive answer we need more distance between the emergence of COVID and the use of visualizations to see it value in a historic context.

## **References**

Coronavirus (COVID-19) latest insights: “Retrieved October 10, 2022, from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19/latestinsights>

Coronavirus (COVID-19) Infection Survey “Retrieved October 11, 2022, from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation#coronavirus-covid-19-infection-survey>

Covid 19 API Developers Guide “Retrieved October 11, 2022, from <https://coronavirus.data.gov.uk/details/developers-guide>

Prior Knowledge Evidence Based Teachers Network “Retrieved October 11, 2022, from <https://ebtn.org.uk/prior-knowledge/>

Wolodtschenko, Alexander; Forner, Thomas (2007). ["Prehistoric and Early Historic Maps in Europe: Conception of Cd-Atlas"](http://www.e-perimetron.org/Vol_2_2/Wolodchenko_Forner.pdf) (PDF). E-perimetron. 2 (2). [ISSN](https://en.wikipedia.org/wiki/ISSN_(identifier)) [1790-3769](https://www.worldcat.org/issn/1790-3769). Retrieved 2015-01-24.

Delnero, Paul. "A Land with No Borders: A New Interpretation of the Babylonian “Map of the World”." Journal of Ancient Near Eastern History 4.1-2 (2017): 19-37

[Cole Nussbaumer Knaflic](https://www.amazon.co.uk/Cole-Nussbaumer-Knaflic/e/B013RK32CW/ref=dp_byline_cont_book_1)   Storytelling with Data: A Data Visualization Guide for Business Professionals (November 2015) ISBN 1119-002257

Desmon Stephanie January 2021 How to Convince the COVID-19 Vaccine Hesitant “Retrieved October 11, 2022, from <https://ccp.jhu.edu/2021/01/21/covid-19-vaccine-hesitant/>

Dong, E., Ratcliff, J., Goyea, T. D., Katz, A., Lau, R., Ng, T. K., Garcia, B., Bolt, E., Prata, S., Zhang, D., Murray, R. C., Blake, M. R., Du, H., Ganjkhanloo, F., Ahmadi, F., Williams, J., Choudhury, S., & Gardner, L. M. (n.d.). The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. The Lancet Infectious Diseases, 0(0). https://doi.org/10.1016/S1473-3099(22)00434-0

Betsy Gardner July 2020 Best Practices for COVID-19 Data Visualizations “Retrieved October 11, 2022, <https://datasmart.ash.harvard.edu/news/article/best-practices-covid-19-data-visualizations>

Falisse, J.-B., & McAteer, B. (2022). Visualising policy responses during health emergencies. Learning from the COVID-19 policy trackers. Convergence, 28(1), 35–51. <https://doi.org/10.1177/13548565211048972>

Yang, A., Shin, J., Zhou, A., Huang-Isherwood, K. M., Lee, E., Dong, C., Kim, H. M., Zhang, Y., Sun, J., Li, Y. Nan, Y., Zhen, L., & Liu, W. (2021). The battleground of COVID-19 vaccine misinformation on Facebook: Fact checkers vs. misinformation spreaders. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-78

Krishnan, N., Gu, J., Tromble, R., & Abroms, L. C. (2021). Research note: Examining how various social media platforms have responded to COVID-19 misinformation. *Harvard Kennedy School (HKS) Misinformation Review*. <https://doi.org/10.37016/mr-2020-85>

*Sutherland, R. (2021). Alchemy: The Dark Art and Curious Science of Creating Magic in Brands, Business, and Life. HarperCollins Publishers. https://books.google.co.uk/books?id=DjCtyAEACAAJ*

*Bridgman, A., Merkley, E., Loewen, P. J., Owen, T., Ruths, D., Teichmann, L., & Zhilin, O. (2020). The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-028*

*Nussbaumer Knaflic, C. (2015). Storytelling with data (C. N. Knaflic, Ed.). John Wiley & Sons.*