A Global View Of Covid-19 Statistics - Data Visualization

Matthew Laws
University of Nottingham
Nottingham, United Kingdom
psxml7@nottingham.ac.uk

ABSTRACT

Data visualisations are a key tool for presenting large quantities of data clearly and concisely. During this COVID-19 pandemic, many people have been faced with large volumes of complex information that has led them to search for credible sources that can be easily understood. The World Health Organisation (WHO) is a highly respected institution from which to obtain reliable data, but the complexity of its data visualisations present a high barrier to entry for the layperson to comprehend. Thus, it is important to create simple and factual visualisations to enable a viewer to understand the threat, act, and stay safe. This paper details the process of creating such a system from reliable WHO data, TopoJSON data, Javascript, HTML, and CSS. It also assesses the success of the site through user surveys, eye-tracking, and system analysis. Of the final site, users reported ease of use, speed of understanding, and a professional appearance; particularly in comparison with existing WHO visualisations. This site is well-placed to be hosted online for access to a wider audience, although some improvements could significantly improve the site's reach.

Author Keywords

TopoJSON; HTML; CSS; Javascript; COVID-19; Data; Data Set, Data Visualisation; WHO,

ACM Classification Keywords

Information Systems → World Wide Web → Web Applications
Applied Computing → Life and Medical Sciences → Consumer Health
General and Reference → Document Types → Surveys and Overviews

INTRODUCTION

Novel COVID-19's first death was reported to the World Health Organisation (WHO) and attributed to the deadly virus in Wuhan, China, on the 31st of December 2019; approximately 490 days before this report was written. There have been at least 152,500,000 confirmed cases of COVID-19 and at least 3,200,000 confirmed COVID-19 related deaths. WHO provides accurate, ongoing and daily-evolving data for tracking the global effects of the virus but lacks concise and accessible data visualizations for the layman in one place. This report outlines the steps taken to provide a data visual representation of COVID-19's worldwide effects in a simple, yet effective, way. It also briefly outlines the work needed to improve this visualization and move it forward to maximise its efficacy.

RELATED WORK

Jacob, 2020 [1]
This study investigated the impact of data visualisations of the pandemic in two news outlets in India. It found that colour choices, especially the use of the colour red, impacted peoples' actions significantly. Importantly, the visualisations of the data impacted the levels and ways in which people reacted to the pandemic with visualisations that stressed spread and severity increasing the likelihood that a person would modify their behaviour to reduce spreading or catching the disease. This impressed upon me the human effect of the visualisations I create and to be keenly aware in a situation as important as dissemination of information during a pandemic, that the data I presented be impactful, factual, and useful.

Raineri and Molinari, 2020 [4]
This book explores the value of data visualisations to business decision making and the sway of public opinion. It stressed that when large volumes of data need to be analysed, data visualisations present the fastest method for the viewer to make a decision or form an opinion of the facts. It also stressed the manipulation of opinion that can occur, intentionally or otherwise, through colours, data selection, and presentation method, and the prevalence of this occurrence in the data visualisations seen by the public throughout the pandemic. This influenced my decisions in terms of ensuring I took all steps to prevent misinterpretation and to promote awareness of the threat.

Sun *et al.*, 2020 [5]
This letter to the editor in response to a study on machine learning in disease tracking shows the importance of simple

visualisations in understanding a complex topic. Importantly, it used a regional heatmap and graph of the disease spread in Japan which directly influenced my own decisions on the best way to present such data.

METHODS, DESIGN AND IMPLEMENTATION

Initially, the WHO dataset [8] (accurate at the date of 3rd May 2021) was routinely massaged per update to match both the TOPOJSON data used to map the choropleth visualization and the coding within the visualization itself [9, 10]. This proved to be a very lengthy process and needed remedying as the data was updated every few days throughout the development of the project. So, the TopoJSON data from github was copied and pasted into a new JSON file and added to the zip. This enabled a much more streamlined process for updating the COVID-19 data from the WHO. The massaging of data of the dataset only requires a removal of the total figures row and then matching column names to those called as variables. Most significantly, thereafter, country names were amended to match the data set in the JSON files so that values were appropriately assigned to the correct country - for example Czech Republic became Czechia. This was a two-fold solution. As mentioned, it alleviated the data massaging workload but also implemented important political correctness in the final product.

The file app.js is the main structure for the encoding of the visualization. It contains the key architecture for all the visual elements such as the map. It also contains numerous utility functions that are used throughout these elements, for example, the getColour function. This is specifically designed to normalise the values selected for in the dataset so that a smoother range of colours is visualized using the selected colour scale. Before this was implemented, countries like the USA and India were a deep red and most other countries a pale orange. App.js also contains all the necessary encodings for the tooltips to work and requesting client-side x and y position of the cursor in order to map them appropriately; that cannot otherwise be done in the HTML or CSS. Other functions such as the sortGraph function are present here. This allows the functionality of the graph to be sorted per the users input and applies the animation.

A secondary file called dataselector.js was constructed to host a series of functions designed to switch the data currently being viewed when the user clicks the buttons in the visualization. Once the functions construct the variables appropriately, these are then added to an array which the buttons access and link directly to the position within. For example, “Cumulative Total Deaths” are at position “0” in the array and this is referenced when using the button to change the data selection. This file also contains the necessary code to rename the column headers to be appropriate for the visualization, rather than reproducing the headers from the CSV which are not practically named for user understanding.

A “ReadMe” was also created to help inform the user on how to run the project once they have unzipped it. The design ethos of this project was to create and allow a more informative and streamlined experience for the layman and as it is assumed many will not have the technical knowledge to run everything on the first attempt, a ReadMe felt appropriate. In addition to this, the RunMe batch file was created to, once the ReadMe was followed, streamline the user experience to launch the visualization with the click of a single button.

Vis.html operates as the fundamental architecture that draws all the visual elements to the screen and provides the general position, z-index, and text to the screen where necessary. It also serves to provide a series of scripts for the server to run from the internet to import the necessary libraries and other functionality in order to make the visualization work. This file also serves as a link to open the page if the server is already being hosted locally, when opened from the file explorer and not an integrated development environment (IDE).

A cascading style sheet (CSS) was created in order to stylize all of the elements that are outside of, and inclusive in some cases, of the visualizations. This includes more precise positioning, colouring and size of text and many other functions such as coloring the slider or table rows.

Finally, server.js was created at the start of the project and acts as the root directory for the project and fulfills two purposes. Firstly, it allocates a name to the files in the directory which, when used in the code, calls the file and its contents. For example, it renames the csv to “data” instead of “COVIDds” as it is easier when coding. Secondly, it tells the program where the files are located in the directory in order to access them.

Another noteworthy inclusion is the dataset itself. It is necessary for the dataset CSV to be present in the directory, and be referenced correctly in server.js, in order for the project to run. Further noteworthy additions are the “node_modules”, “package.json” and “package-lock.json”, these are automatically added and generated when installing the necessary libraries in order for d3 to work and correctly use json data.

RESULTS

Map

The first visualization the user will interact with is a choropleth map (Figure 1) [9, 10]. This is a type of heatmap, displaying all the countries on a mercator projection of the world. The map was coloured using a light orange to deep red scale to infer severity in the initial user observations. While it could be argued that using reds increases the fear response of viewers [1], in this case the

implication that higher levels of infection and death are a negative thing is not a misleading conclusion to draw, and in fact, was the aim of this decision. Jacobs [1], Raineri and Molinari [4] found that visualisations that stressed the importance or severity of the pandemic through colour choices and imagery could increase public response in the prevention of spread.

Buttons are positioned above the map and coloured in the same way as for any selection (Figure 2). The buttons have a hover over function to allow the user faster comprehension of where their mouse is and what data selection they are about to make. There is also a focus selection to the button where it is made darker than all the other buttons, and a hover color so the user knows which data set they are focused on. These buttons allow a switch to a different key value pair and thus change the data presented in the choropleth map visualisation.

When the user hovers their mouse over a country a tooltip appears (Figure 3) that will give them the information of the key value pair they are selecting for. For example, the country's name and their cumulative total cases in the format "There are X cumulative cases in Y". The country that is hovered over is surrounded by a black border and all other countries in the map inherit an alpha value to put them in the background. If the country, or continent in the case of Antarctica, do not have any values within the data set the user will receive an error message in the tooltip in the format of "There is no data available for X" (Figure 4). The use of tooltips with exact information on case numbers in a country, alongside the use of colour intensities, rendered the use of an overall scale unnecessary; instead the user can infer the relative numbers by country colour and investigate using a tooltip. This was determined by the use of a rudimentary eye tracker while users were on the site that showed the scale was rarely looked at, and the survey showed no users attempted to use it to interpret the data in the map.

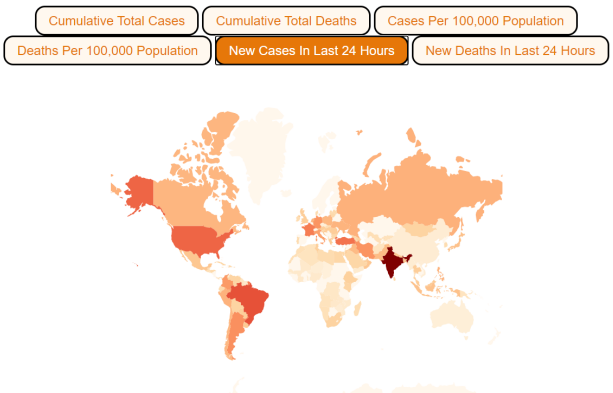

Figure 1. Screenshot of the choropleth map when 'New Cases in Last 24 Hours' is selected.

Figure 2. Screenshot of the button options available to the user to change the data selection of the choropleth map.

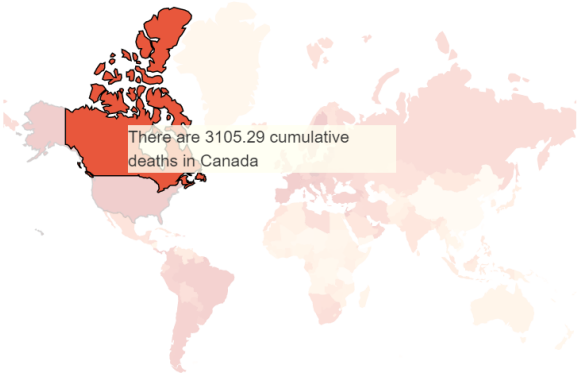

Figure 3. Screenshot of the choropleth map displaying the tooltip generated when the mouse is passed over Canada.

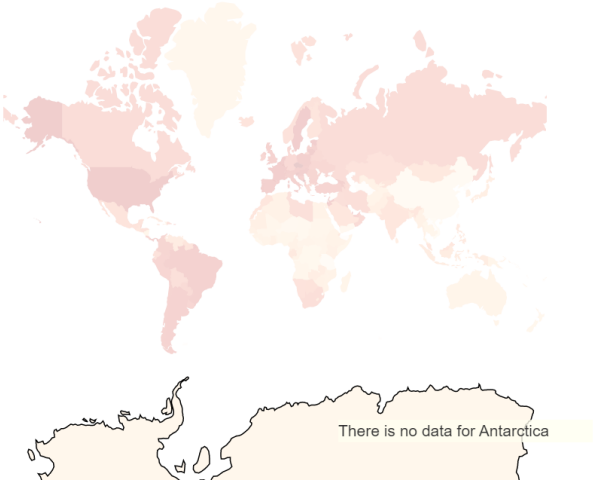

Figure 4. Screenshot of the choropleth map displaying the tooltip generated when the mouse is passed over Antarctica, a country with no available data within the dataset.

Table

The second visualization, a table of data relating to a specific, chosen country, follows below the map [7]. Before the table can be seen it prompts the user to click a country (Figure 5). The selection of a country in the map causes a table to appear showing all of the available data for that country (Figure 6) without the need to select for anything using the buttons - i.e. the table will display the same information regardless of the data selection button used on the map. If the user picks a country with no information available they are prompted with a message that replaces either the initial prompt or the table itself (Figure 7).

The first row of the table is styled boldly as a header (Figure 6) to allow for more streamlined user comprehension of what country they are viewing data for. The subsequent rows are colored by odd and even rows for a smoother visual experience, in addition to a hover function which mimics the first row's clear stand-out style to highlight clearly what the user is interested in or looking at, at the time. Orange was selected as the main colour for the table as it is a more neutral colour than the reds used in the map, while also fitting the colour scheme of the website.

Click a country from the map above for detailed information and statistics of their national COVID-19 data

Figure 5. Screenshot of the message that appears in place of a table when no country has been selected.

Name	Niger
WHO Region	Africa
Cumulative Total Cases	5158
Cases per 100,000 Population	21.31
Cases - newly reported in last 7 days	44
Cases - newly reported in last 7 days per 100000 population	0.18
Cases - Newly Reported In Last 24 Hours	17
Cumulative Total Deaths	191
Deaths per 100,000 Population	0.79
Deaths - Newly Reported In Last 7 Days	1
Deaths - newly reported in last 7 days per 100000 population	0
Deaths - Newly Reported In Last 24 Hours	0
Transmission Classification	Community transmission

Figure 6. Screenshot of the table generated when Niger has been selected on the choropleth map.

There is no data for Antarctica. Select a new country that provides a tooltip value to view detailed information

Figure 7. Screenshot of the message displayed when the country selected has no available data in the dataset.

Bar Chart

The interactive bar chart allows for a more specific comparison of between-country data and statistics than can be seen using the map. The bar chart is featured with a copy of the same buttons used for data selection on the map (Figure 8). This repetition is to allow the user to quickly toggle between the data they are interested in without needing to scroll up to the selection buttons above the map. In addition to this, there are animated sort buttons for sorting by descending or ascending just below the graph [2].

This visualization, by default, captures the data for all 237 countries present in the dataset and, because of that, some design choices were forced. Firstly, the bars are so thin that the line of the y-axis had to be removed in order to see the first country being displayed. Secondly, all data labels for the x-axis were removed as there would otherwise be an incomprehensible amount of data being displayed and overlapping at once. In order to compensate for this, a tooltip was added (Figure 10) that appears when the user hovers over a data bar. This provides the same information as hovering over a country in the map section and uses the same format. There is a text prompt to inform the user of this hover function (Figure 9).

In order to make this visualization more digestible, a slider bar was added to allow the user to limit the amount of countries shown on the bar chart at any given time (Figure 11) [6]. This is especially useful as there are a large number of smaller countries with very low values for cases and deaths which make up the majority of the width of the x-axis. Additionally, x-axis data label functionality is restored to the visualization when the number of data points is limited to 30 countries or fewer (Figure 12). This limited data set can then be sorted by either descending or ascending if the user wishes.

As with the table, orange was used in this visualisation, again, as a neutral colour. The use of red to infer severity wasn't necessary in a bar graph where that information is already clearly conveyed through the heights of the bars.

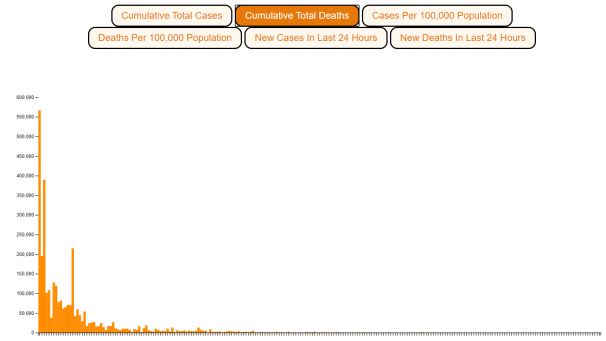

Figure 8. Screenshot of the bar chart generated when 'Cumulative Total Deaths' is selected.

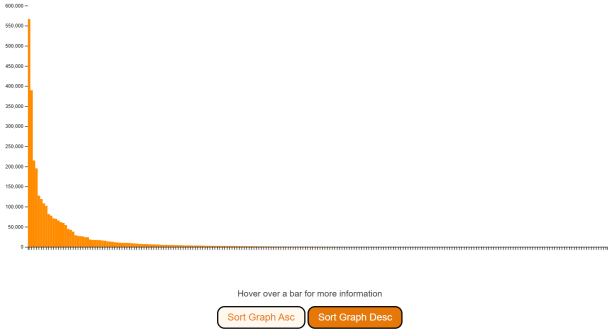

Figure 9. Screenshot of the bar chart generated when 'Cumulative Total Deaths' is selected and the data sorted by descending size using the 'Sort Graph Desc' button is used.

Figure 10. Screenshot showing the tooltip generated when the mouse is passed over a bar within the bar chart.

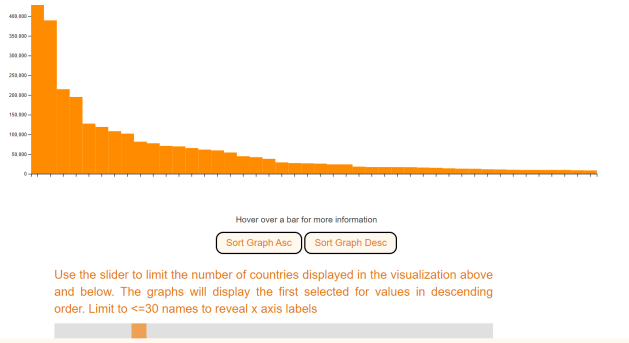

Figure 11. Screenshot of the bar chart when the number of countries has been limited to a value over 30.

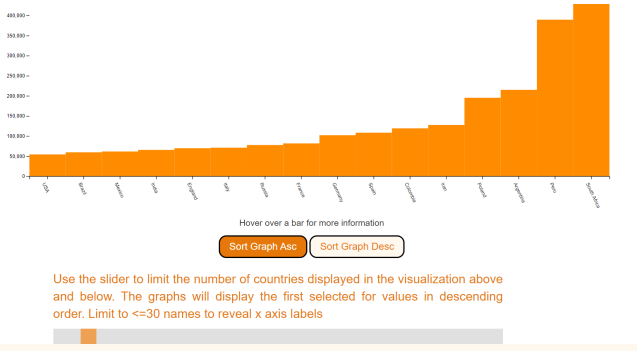

Figure 12. Screenshot of the bar chart when the number of countries has been limited to a number lower than 30 and subsequently sorted in ascending order using the 'Sort Graph Asc' button.

Pie/Doughnut

In a similar way to the bar chart, the pie chart (Figure 13) [3] allows a more precise analysis between countries than the data from the map. The pie chart serves a purpose to break down the information in the bar chart for more intimate investigation. It serves as a true comparison between countries by separating and displaying percentages taken from the whole population per country.

There are only four buttons available to use for this visualization as the per 100,000 population data are not appropriately visualized using a pie chart. If the user has an illegal data set selected for when scrolling to the pie chart they are prompted with an error message and advised to pick a data set from the buttons associated with this visualization (Figure 14) .

When the user hovers over a segment of the pie, a label appears in the center of the visualization presenting the country's name, the associated value and it's percentage from the whole population this represents. If a user clicks on a segment, the data label in the center becomes "sticky" and remains even when the mouse is moved away, whilst the selected segment is also highlighted (Figure 15). If the user clicks on a new segment, the hover labels are once again reactivated and a new sticky selection can be made. This is briefly explained to the user in a text prompt directly below the visualization (Figure 13). This was specifically designed so that users could easily compare any country to another country of interest. Most users I observed chose their home country as the 'sticky' information and then hovered over other countries of interest to see how they compared.

The slider bar for the bar chart works in the same way as for the pie chart [6]. This is prompted in the text associated with the slider bar itself (Figure 16). Again, this is especially useful as most countries have few enough cases that the size of the segment fades out of view. A limited dataset allows a more detailed view of only the top selection of countries.

As with the choropleth map, shades of red were used to infer the severity of the numbers by country and mouse overs used to find out more precise information. While the sizes of the sections in the pie chart can also convey that information, the differences in sections are not so immediately obvious to the viewer without also a colour change between segments. The colour shades were added to aid speed of understanding.

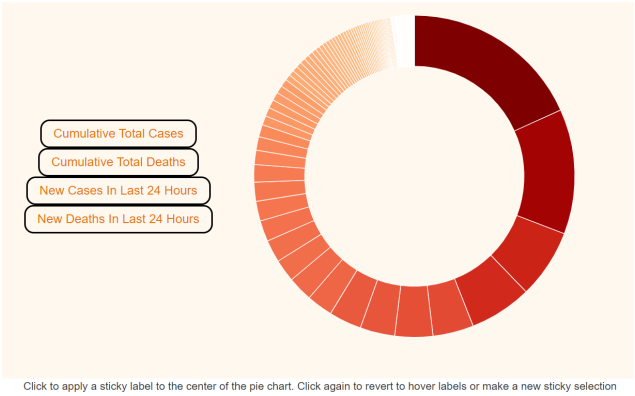

Figure 13. Screenshot of the pie chart generated when 'Cumulative Total Deaths' has been selected.

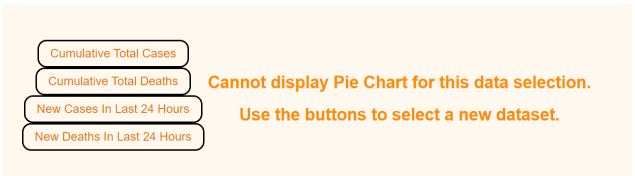

Figure 14. Screenshot of the message generated when a data selection has been made using unavailable buttons for the previous visualisations.

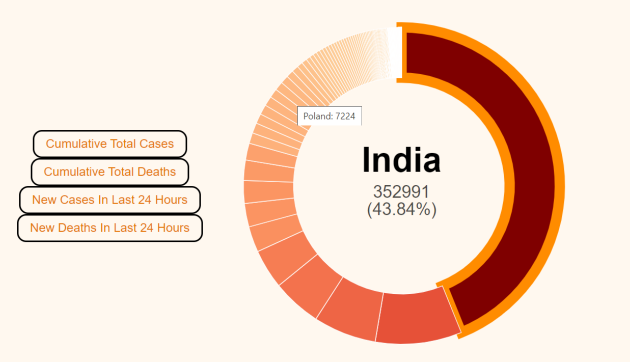

Figure 15. Screenshot of the pie chart after India has been selected and the mouse passed over a different country in the chart.

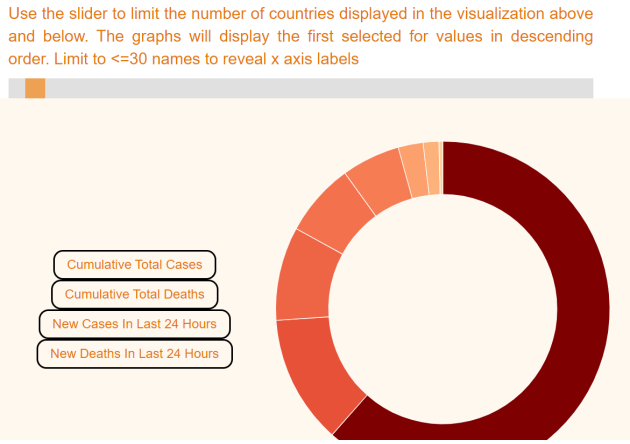

Figure 16. Screenshot of the pie chart when the number of countries displayed has been reduced.

DISCUSSION AND EVALUATION

The functionality of the website works smoothly and effectively. Visualisations load quickly and provide information chosen by the viewer, i.e. tooltips and chart selection, near-instantaneously. My user survey found that the time spent on the site averaged at 5-8 minutes and tended to vary based on the user's interest in the subject. For example, one of the users tested was a GP who spent longer on the site than another user who was an accountant. Of that time, approximately 10 seconds were used to load the site, with this time varying slightly from user to user.

The loading of the site itself was enabled through the creation of a Batch (.bat) file titled 'RunMe' which was clear and simple for the user to understand. This process is assisted by a ReadMe file for users who are not as informed on how to start this type of program or those that encounter any errors when running the batch file. It informs the user on a prerequisite software (Node JS) and a further set of short instructions and recommendations. If this was then hosted as a website online, the information would be even easier to access and so this would be the most important step in allowing the public to utilise the information and visualisations presented here.

When users of the site were surveyed it was found that rate of cognition and ease of understanding scored highly. We had users compare the visualisations and site created here against the visualisations accessible through the WHO website. On average, users relayed that the size and scale of the WHO site meant that, while its visualisations were good when they found them, finding them was difficult as lots of data were stored in different pages around the site. Multiple users were unable to find the data they were looking for on the site at all. For the casual user who just wants to see an overview of information, or to keep up with how and where the disease is spreading, using the WHO source could be overwhelming.

This is where the site I have created worked well. All data from the dataset could be seen on a single page. For a casual user, the information was immediately present without user input, and then further investigation by the user could be easily carried out. Importantly, the user could investigate further based on what interested them in the data already presented, rather than needing to come to the site with a pre-existing idea of what they wished to find out. Multiple users stated that it would be a good source to check regularly to watch the disease progress in real-time as a lot of data could be understood from only being on the site for a minute or so. To implement this, however, would require the site to be easily updated, either through a simple manual process, or automatically, to include the latest WHO data in the CSV.

I feel confident that the site has achieved the goal I had set out for it: that it would provide concise and accessible data visualisations, understandable by a layperson, in a single location.

CONCLUSION AND FURTHER WORK

While this system runs smoothly, and generates no errors in the creation and modification of the visualisations, there are improvements that can be made. In order to take this visualisation to the next stages, it would be imperative to add live updating to the CSV (Comma Separated Value) Excel sheet at the press of a button. This would version-check for whether the available dataset that is available from WHO matches that of the data being used in the visualisation. If it does not, a script would be implemented to download and replace the CSV being used in the working directory - therefore updating the visualization to the most recently published COVID-19 figures provided by the WHO.

The map functioned in the way that it was intended to and with no known bugs or errors. It would however benefit from being able to dynamically resize to the user's screen size at as large a scale as would fit on the screen. The standard size of the mercator map generated using TopoJSON data has a fixed height and width. This proved difficult to manipulate whilst maintaining an appropriate size and style for the visualisation. Future improvements of

the site would allow the map to scale dynamically to a larger size, allowing smaller countries, such as Luxembourg, to be easier to select with the mouse.

Refinements to the bar chart section are also necessary. When limiting the number of countries being displayed, it will take away entries from the end of the graph, regardless of the current sort. That means that if the number of countries is limited before the sort then subsequent sorting will not encapsulate the highest, or lowest, but rather those in the first data point entries at the time or limiting. Some fed back that they enjoyed that they could switch the order of ascending or descending case numbers and then limit the dataset to investigate the countries with the least impact from COVID-19. Others, however, saw this as a bug rather than a feature and confused them in their use of the slider and organisation of data as they thought the empty bar chart was an error (Figure 17). This is mentioned in the text prompt just above the slider bar but, ideally, recoding the way the sort works to take the data from the dataset and not their positions in the graph is the most ideal solution. Additional buttons to display, for example, the top ten worst affected or top ten least affected countries per data selection would be ideal to allow for a quicker and more efficacious investigation for the user.

More work can be done into improving the pie chart visualization. The primary example would be to enable hover labels to persist whilst a sticky selection is made and a new text label would appear in an alternative color so that the user can make a quick comparison between countries. Currently, while these comparisons can be made, they are not made with the formatting of the site visible, and don't include the percentage data found otherwise on the pie chart (Figure 15) It would also help to label how the dataset is being reduced when the slider is used. For example, reducing the pie chart to the top 30 countries for COVID-19 deaths would cause a line of text explaining that 'this pie chart shows the 30 countries with the most cumulative COVID-19 deaths'.

The most important development of the project would be for it to be hosted as a publicly-accessible website. Expecting a user to download a zip folder and run a program is much more labour-inducive and less likely to attract users than simply accessing a domain. This would also make the site more accessible to users accustomed to website use, but not to running programs.

Figure 17. Screenshot of the bar chart when the data has been sorted into ascending order using the 'Sort Graph Asc' button, and then limited using the slider to less than 30 countries.

ACKNOWLEDGEMENTS

We thank all survey participants for their in-depth and informed feedback both throughout the development of the site and with assessment of the final product.

BIBLIOGRAPHY

1. Jacob, R. Visualising global pandemic: A content analysis of infographics on Covid-19. *Journal of Content, Community & Communication* 11 (2020) 116-123. DOI: 10.31620/JCCC.06.20/09
2. Mike Bostock. 2018. Sortable Bar Chart. Last Accessed May 3, 2021 at https://observablehq.com/@d3/sortable-bar-chart
3. Mike Bostock. 2018. Donut Chart. Last Accessed May 3, 2021 at https://observablehq.com/@d3/donut-chart
4. Raineri P., Molinari F. 2021. *Innovation in Data Visualisation for Public Policy Making*. (1st. ed.) SpringerBriefs in Applied Sciences and Technology. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-63693-7_4
5. Sun, G., Trung, N.V., Hoi, L.T., Hiep, P.T., Ishibashi, K., Matsui, T. Visualisation of epidemiological map using an Internet of Things infectious disease surveillance platform. *Critical Care* 24, 400 (2020) 1-2. DOI: https://doi.org/10.1186/s13054-020-03132-w
6. W3 Schools. How To Create Range Sliders. Last Accessed May 3, 2021 at https://www.w3schools.com/howto/howto_js_rangeslider.asp
7. W3 Schools. HTML Tables. Last Accessed May 3, 2021 at https://www.w3schools.com/html/html_tables.asp
8. The World Health Organisation (WHO). WHO Coronavirus (COVID-19) Dashboard: Situation by Region, Country, Territory & Area. Last Accessed May 3, 2021 at https://covid19.who.int/table
9. Yan Holtz. 2018. Choropleth map with hover effect in d3.js. Last Accessed May 3, 2021 at

https://www.d3-graph-gallery.com/graph/choroplethh_hover_effect.html
10. Yan Holtz. 2018. GeoJSON World. Last Accessed May 3, 2021 at https://www.d3-graph-gallery.com/graph/choroplethh_hover_effect.html