Aashima Parvathi Poudhar
Heriot-Watt ID: H00423612
F21DV Coursework 1
Link to video- F21DV-CW1-Video.mp4

Course code and name:	F21DV Data Visualisation & Analytics
Type of assessment:	**Individual** *(delete as appropriate)*
Coursework Title:	F21DV Data Visualisation & Analytics Coursework 1
Student Name:	Aashima Parvathi Poudhar
Student ID Number:	H00423612

Declaration of authorship. By signing this form:

- **I declare** that the work I have submitted for individual assessment OR the work I have contributed to a group assessment, is entirely my own. I have NOT taken the ideas, writings or inventions of another person and used these as if they were my own. My submission or my contribution to a group submission is expressed in my own words. Any uses made within this work of the ideas, writings or inventions of others, or of any existing sources of information (books, journals, websites, etc.) are properly acknowledged and listed in the references and/or acknowledgements section.

- I confirm that I have read, understood and followed the University's Regulations on plagiarism as published on the University's website, and that I am aware of the penalties that I will face should I not adhere to the University Regulations.

- I confirm that I have read, understood and avoided the different types of plagiarism explained in the University guidance on Academic Integrity and Plagiarism

Student Signature *(type your name):* *Aashima Parvathi Poudhar*

Date: *16/03/2023*

Copy this page and insert it into your coursework file in front of your title page.
For group assessment each group member must sign a separate form and all forms must be included with the group submission.

Your work will not be marked if a signed copy of this form is not included with your submission.

Table of Contents

COVID-19 – A Proportionate or Disproportionate Impact on the World ***4***

1 Overview ***4***

1.1 Story **4**

1.2 Data **4**

1.3 Visualizations **4**

1.4 Coursework Submission Structure **5**

1.5 Browser Compatibility **5**

2 Details of Visualizations and Task Achievement ***5***

2.1 Main Tasks **5**

2.1.1 How did the COVID-19 pandemic grow over countries and continents throughout the world from its start? 6

2.1.2 Choosing some specific countries, how successfully did they manage the outbreak? 6

2.1.3 How might the geographical position of a country change how the pandemic impacted them? 7

2.1.4 What effect did vaccinations have on the spread of cases/deaths? Did booster jabs also have an impact on the spread/transmissibility of the virus? 7

2.1.5 Choosing some specific countries, is there a relationship between the relative "wealth" (e.g., GDP) of a population and the spread of the pandemic? 8

2.2 Application Requirements **8**

Requirement A1 8

Requirement A2 9

Requirement A3 9

Requirement A4 9

Requirement A5 10

Requirement A6 10

Requirement A7 11

Requirement A8 11

Requirement A9 11

Requirement A10 11

Requirement A11 11

2.3 Other Design Choices **12**

2.4 Core Requirements **13**

References ***16***

COVID-19 – A Proportionate or Disproportionate Impact on the World

1 Overview

During the COVID-19 pandemic core period between February 2020 and February 2023, countries across the world were hit very differently due to varying social, economic, and environmental factors. For this course work, I chose to approach the data based on this perspective.

1.1 Story

Title: **COVID-19 – A Proportionate or Disproportionate Impact on the World**. Based on key parameters namely, population density, stringency index, vaccination rate and GDP per capita, the visualizations aim to depict to the user, the severity of the pandemic in various countries in terms of the number of new cases per month, mortality rate calculated as number of new deaths per million / number or new cases per million * 100, total deaths per million, new cases per million. From the data visualization, the user is free to draw conclusions about whether the varying factors of each country influenced the severity of the outbreak and helped control it.

1.2 Data

https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-data.csv

The data is available for nearly every day for the COVID-19 core period. My data visualization story focuses on monthly changes and therefore, I pre-processed the owid-covid-data.csv to aggregate the values for every month before loading it in the code. The extraction file is included in the code/resources directory of my course work code submission.

1.3 Visualizations

I have used 3 main visualizations in a single page:

1. Time Series Map of the COVID-19 spread for every month from February 2020 to February 2023 depicted on the world map. (Bostock, 2012)
2. Bubble Chart depicting number of new cases against stringency index, with the size of the bubbles showing the population density and colours representing countries.
3. Line Plot comparing the mortality rate and vaccination rate for each country selected from a drop down for the entire covid period. ((Mathieu, 2022), ((COVID-19))

And 2 supporting visualizations:

1. Heat map that compares GDP per capita with – 1. Total deaths per million, 2. New cases per million.
2. A scatter plot that is created using the result of clustering the countries based on the selected parameter and number of new cases per million.

1.4 Coursework Submission Structure

F21DV_CW1

> ***code***

> ***index.html***

> *README.md*

> *js*

> *css*

> *data*

> *images*

> ***report.pdf***

1.5 Browser Compatibility

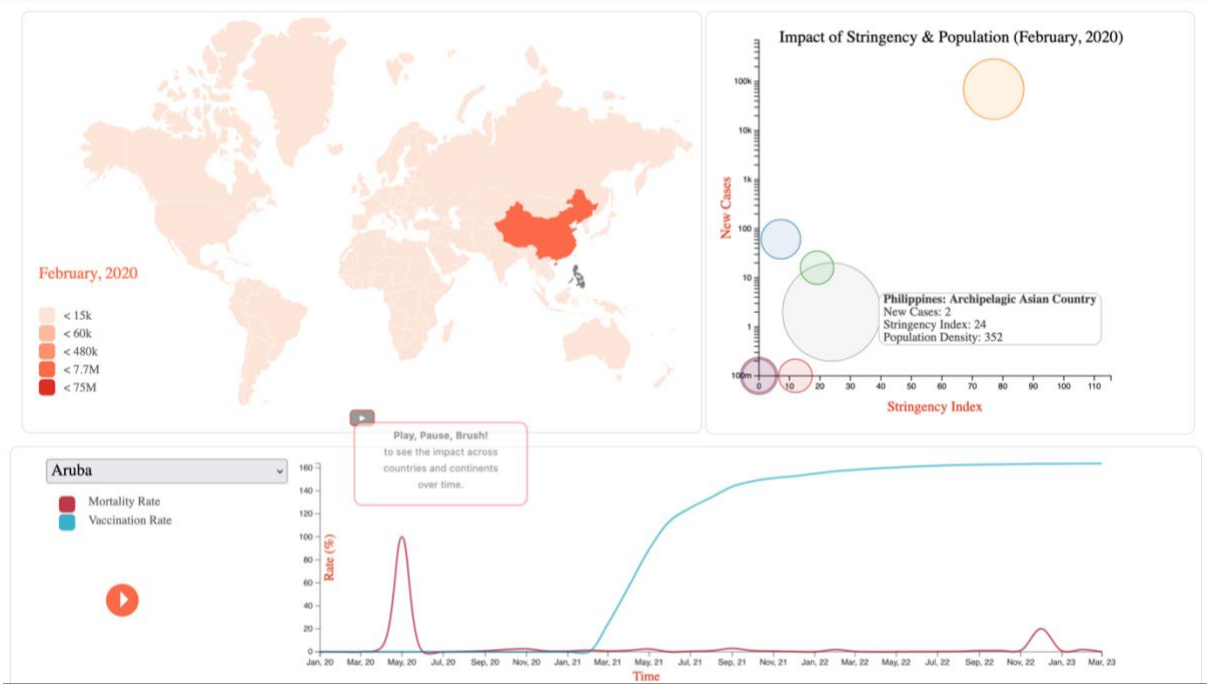
The code was tested on **Firefox**, **Safari** and **Chrome** and it functions correctly and as expected on all 3.

Test Note: Chrome has a small glitch on the introduction page when the page is refreshed.

2 Details of Visualizations and Task Achievement

2.1 Main Tasks

You must narrate the timeline of the COVID-19 pandemic around the world, and the effect the vaccines had on the scale of the outbreak. This task is achieved, including a screenshot of the main visualizations.



2.1.1 How did the COVID-19 pandemic grow over countries and continents throughout the world from its start?

Time Series Map

This task is illustrated by the heat map depicted on the world map to show the wave of COVID 19 in terms of the number of new cases per month.

- Code reference: (Gallery, 2018)
- The legend indicates the severity in terms of number of new cases where the colours are intuitive, a lighter shade indicates lower number of cases, and a darker shade indicates higher number of cases.
- The map has a play button which can be used to start viewing the time series data.
- The user can use the brushing tool to select the area of interest and compare the data of included countries on the automatically updated bubble chart and the line plot that gets updated upon clicking any of the bubbles.

Choice Reasoning: Most people are familiar with maps, and it is easy for users to view and understand the severity of cases in countries and continents on a map rather than say, a bar chart where each bar represents a country. Also, a time series with a simple play button interface takes the burden of navigating away from the user. The brushing on the other hand gives the provision to the user to narrow down on countries of their choice.

2.1.2 Choosing some specific countries, how successfully did they manage the outbreak?

Bubble Chart

The bubble chart next to the month compares the stringency index against number of new cases. Stringency index is an indicator of the strictness of the measures taken by countries in response to the COVID-19 pandemic. Therefore, I chose this as the key element to fulfil this task.

It also considers the population density of each country since that is a key parameter alongside stringency index which might influence the severity of a pandemic.

- The period on the bubble chart is kept in sync with the map to give the user an overall picture.
- The bubbles are clickable and would update the line plot for the specific country.

Choice Reasoning: I wanted users to get a picture of the effect of stringency index in the context of population density. Bubble chart provides a simple, easy to understand interface and supports the comparison of 3 parameters and additionally, the colour of the bubbles can distinguish the countries.

Default Countries Chosen and Why

1. United States of America – Country with the highest number of total cases based on the dataset.
2. China – Asian representative, second highest in terms of total cases and because of its rumoured role in the origin of the pandemic.
3. Australia – Isolated continent that was known to have handled the pandemic well.
4. Guyana – South American representative.
5. Central African Republic – African representative with least number of cases in Central Africa.
6. Vanuatu – Geographically isolated island country with very low number of total cases.
7. Greenland – North American representative which has the least number of cases in the continent and a geographically isolated country.
8. Philippines – Isolated archipelagic country with high population density.
9. France – European representative with highest total cases in the continent.

2.1.3 How might the geographical position of a country change how the pandemic impacted them?

To help users understand this, I leveraged the choice of countries in the bubble chart. The selected countries are either –

1. Geographically well-connected
2. Geographically isolated
3. Or somewhere in between

- When the users hover over the country bubbles on the bubble chart, they can see the corresponding country highlighted on the map. This shows where the country is and how isolated it is from others.
- Also, the tooltip help text guides the users further in this context.
- Or the users can also brush over the map to choose both isolated and well-connected countries and view the impact on the bubble chart.

2.1.4 What effect did vaccinations have on the spread of cases/deaths? Did booster jabs also have an impact on the spread/transmissibility of the virus?

Line Plot

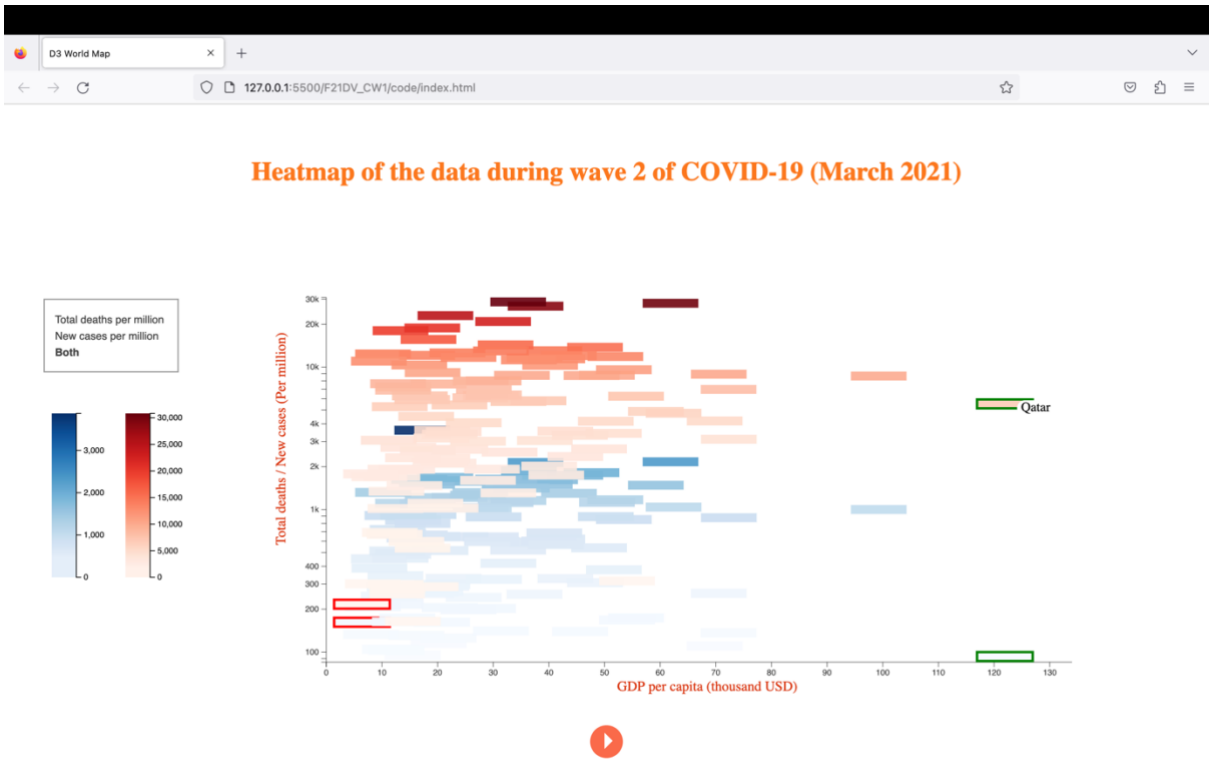
The line plot below the time series map compares the mortality rate calculated as number of new deaths per million / number or new cases per million * 100 and vaccination rate calculated as total vaccinations / total population * 100.

- User can choose country from the drop down or click the country bubbles from the bubble chart.

- It shows the data for the entire covid period. Therefore, users can compare the mortality rate before and after vaccinations began.

Choice Reasoning: A line plot helps compare 2 things simultaneously. Additionally, the parameters I've chosen make it easy for the user to grasp the effect of vaccinations.

2.1.5 Choosing some specific countries, is there a relationship between the relative "wealth" (e.g., GDP) of a population and the spread of the pandemic?



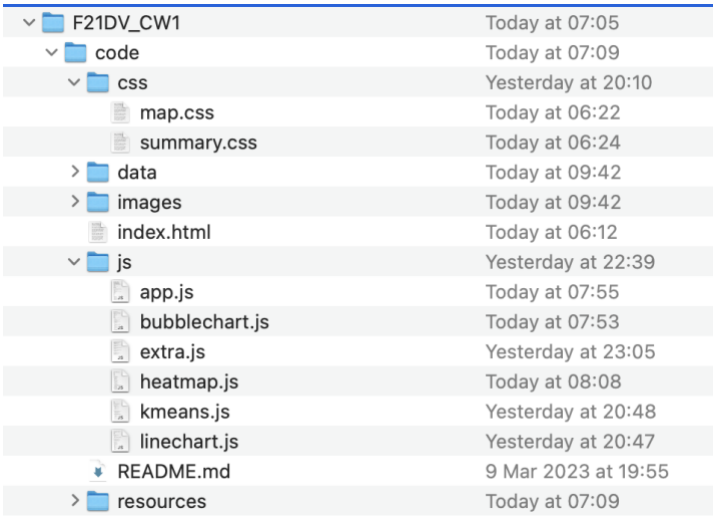
For this task, instead of choosing specific countries, I decided to go with data of all the countries during the wave 2 peak month, i.e., March 2021 since I wanted to depict a heatmap and the visualization looked beneficial with more data.

- The visualization highlights the countries with the highest and lowest GDP.
- It also gives the user the option to compare the pandemic in terms of total deaths per million and new cases per million or both using the selection box provided next to the chart.

2.2 Application Requirements

Requirement A1

Your application should only require a single HTML page called index.html, that is within the root of the project. You are free to create as many additional CSS and JS files as you feel are necessary to support your application. – **Met**.



Requirement A2

All visualisations should be loaded on the single HTML page. – **Met**.

- All visualizations are loading on a single web page – index.html
- There are buttons to assist navigating to subsequent sections of the web pages for users who do not like scrolling.

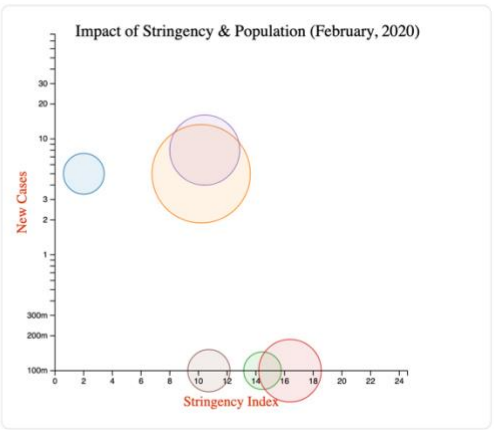
Requirement A3

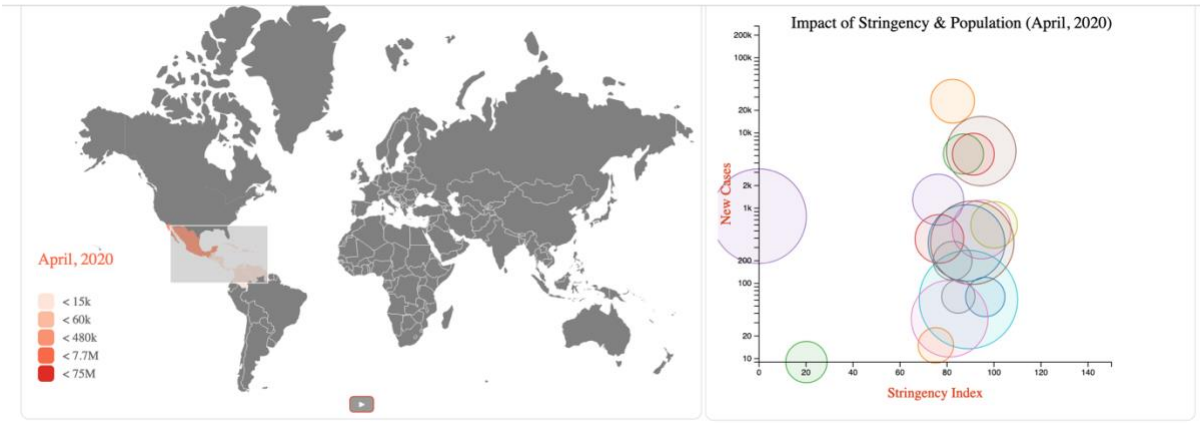
Use at least three different visualisation types. – **Met**.

- Time series map - Map
- Bubble Chart - Correlation
- Line – Evolution
- Heatmap – Correlation
- Scatter Plot - Correlation

Requirement A4

When the data of a single visualisation is updated, all axes of that visualisation should also update/rescale accordingly – **Met**.

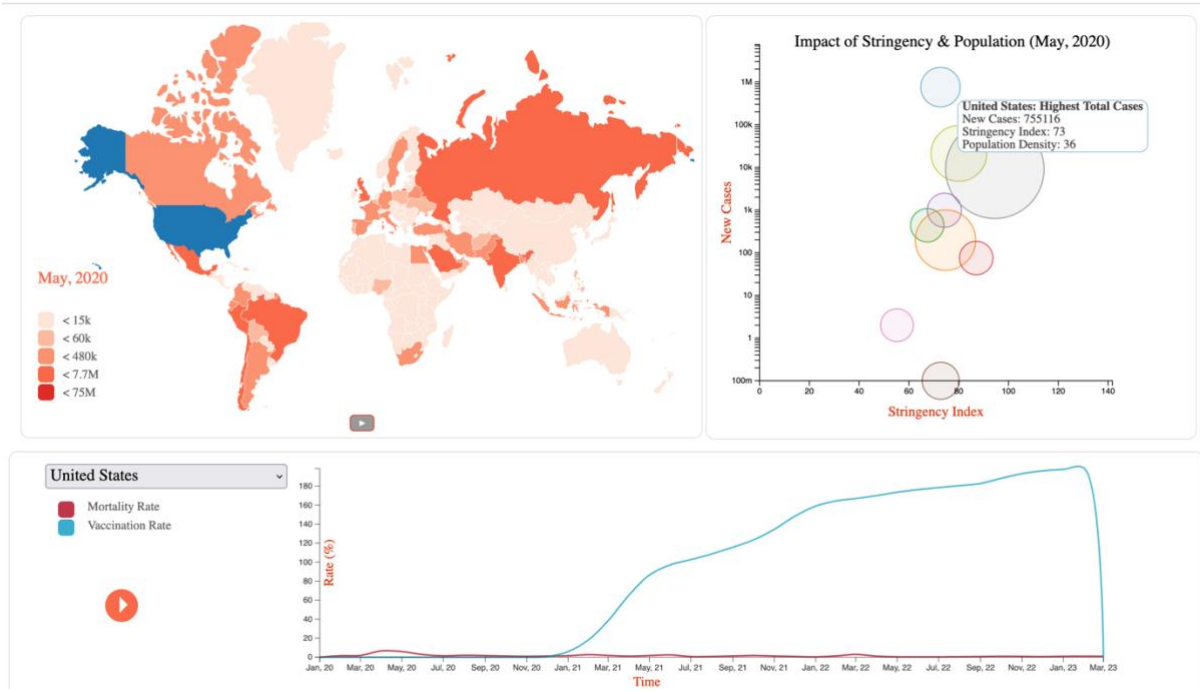




- Axes rescaling and update has been implemented for all charts on the web page.

Requirement A5

For any two visualisations, mousing over a datapoint in one visualisation highlights multiple associated datapoints in another visualisation – **Met**.



Requirement A6

There must be a bidirectional interaction between at least 3 of your visualisations — where interaction with one demonstrates a change in both of the others. This must be consistent such that interacting with any of the three visualisations in the similar manner will cause the others to react in the same way – **Not Met**.

For the visualization story and visualizations, I chose, this requirement could not be fully applied but bidirectional interaction has been implemented wherever appropriate and possible.

- Brushing over a region on the map updates the bubble chart for those countries.
- Hovering over the country bubbles highlights the countries on the map.
- Selecting a country bubble updates the line chart for that country.

Requirement A7

Inclusion of a faceted selection interaction between two visualisations — where a mouseover or a click in one visualisation results in the data of a different visualisation being filtered – **Met**.

- Brushing tool over map results in bubble chart update.
- Clicking country bubble results in line chart update.

Requirement A8

Use of a map visualisation which must interact with at least one visualisation – **Met**.

- World map with time series, interacts with bubble chart.

Requirement A9

Use of scalar data over a map to indicate the distribution of scale over a geographical area — for example, use of circles of different sizes to indicate scales – **Met**.

- Circles of different sizes to indicate population density scale are used in the bubble chart.
- The map indicates distribution of scale using the colour scheme indicated by the legend.

Requirement A10

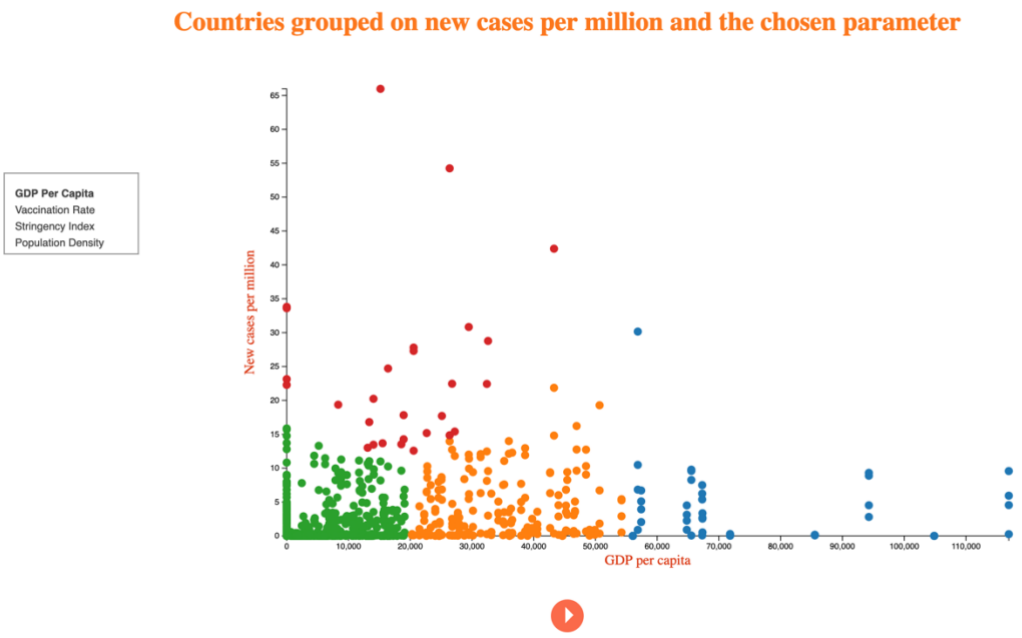
Use of cross-visualisation brushing where dragging a rectangle of several datapoints in one visualisation highlights multiple associated datapoints in another visualisation – **Met**.

- Brushing over map results in the datapoints in bubble chart getting updated.

Requirement A11

Use of an appropriate clustering analysis technique. Note that a scatter plot alone is not a viable form of clustering analysis – **Met.**

- K-Means clustering technique is used to group countries based on number of new cases per million and the parameter chosen by the user which may be one of GDP per capita, Vaccination rate, Stringency Index, Population Density.
- Corresponding data is plotted on a scatter plot.
- Data chosen for clustering: Data of all countries between the first and second covid wave and the second and third covid wave.
    - The data was filtered from a usability point of view because the entire data for 3 years being used for clustering was causing Chrome to slow down while clustering based on a different parameter each time.
    - The data between the peak covid times was chosen as a better indicator of the relation between these parameters and the spread of covid.
- The number of clusters and max iterations are static.



Choice Reasoning: I compared K-Means clustering and hierarchical clustering but as I could not directly think of a hierarchical relationship between the countries and the parameters under consideration, I decided to go with K-Means clustering.

2.3 Other Design Choices

- GeoMercator projection is used for the map because that's the projection most people are familiar with and despite the difference in scale between the real sizes of countries and the size on the projection, it would help users connect better with the map, the countries and continents.
- Bubble chart does not have a legend since the parameters are obvious and directly indicated by the position of the bubble. The supporting tooltip text gives additional information.
- The application contains 3 additional pages which I thought might add continuity and rest cycle between 2 graph sections. Since users are generally from a non-technical

background, the summary text pages may be beneficial for them to process the data until that point and also prepare for the next.
    - **Initial introduction page**: that sets the context for the visualizations with a title, helpful, brief guiding text and a navigation button.
    - **Added a static summary page**: to depict the situation of COVID across continents and to introduce the next section.
    - **A conclusions page**: to ease to the end of the web page. The conclusions are simply there to support the coursework objectives and are not placed to bias users.
- Testing and debugging:
    - Used $0__data__ and console.log messages for debugging.
    - Tested the visualization with different inputs and combinations.
    - Tested on Firefox, Safari, and Chrome.

2.4 Core Requirements

C1. Create a web-based application written in d3.js using version 7+. No PHP or server-side code should be used. JavaScript compiled from other languages (e.g. TypeScript) is not allowed – **Met.**

- Used d3 version 7 (Bostock, 2021).
- All visualizations are created and rendered using only d3 version 7.
- Web page is organized and styled using HTML and CSS.
- For users who do not like to scroll, buttons are added and implemented using basic JavaScript.
- The owid-covid-data.csv is processed to aggregate and get data for each month using python. Given data is not manipulated in any way. It is simply filtered and aggregated to get sums and averages for every month.

C2. Transitions and/or animations must be used to indicate what data are new, changing, or exiting – **Met.**

- Axes transformations on all charts
- Bubble chart has transitions to show new bubble formations when countries are selected by brushing over the world map.
- Heatmap and scatter plot charts and legends update based on user selections.

C3. Your user interface must be intuitive to use – **Met.**

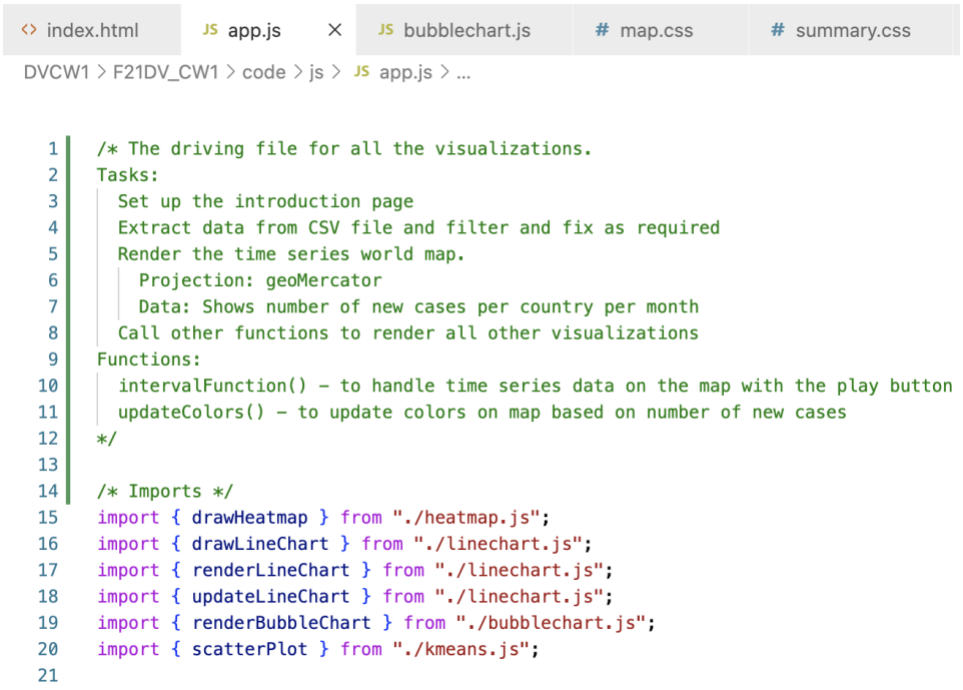
- Gave the web page to 2 people to use and obtained feedback that they were able to navigate through the visualizations with little or no help from the assistive text and tooltips.

C4. You must demonstrate consideration of accessibility when designing your user interface – **Met**.

- Used dark or black text against white background.
- Provided alt text for images.
- Set up fonts for text and fall-back font styles.
- Number and time formats have been used to help users deal with large numbers and to make dates look simpler to say just month and year.
- Use of appropriate pointers:
    - to indicate an element that is clickable.
    - to show the brushing tool.
- Standard tags such as <h1>, <h2> are used for headings.

C5. Source code must be comprehensively documented - **Met**.

- Code has been logically modularized into files and functions.
- Context for each file has been documented with comments.
- Each file has been described at the top of the file.
- Added comments wherever appropriate.



DVCW1 > F21DV_CW1 > code > js > JS kmeans.js > clusterAndRenderScatterPlot

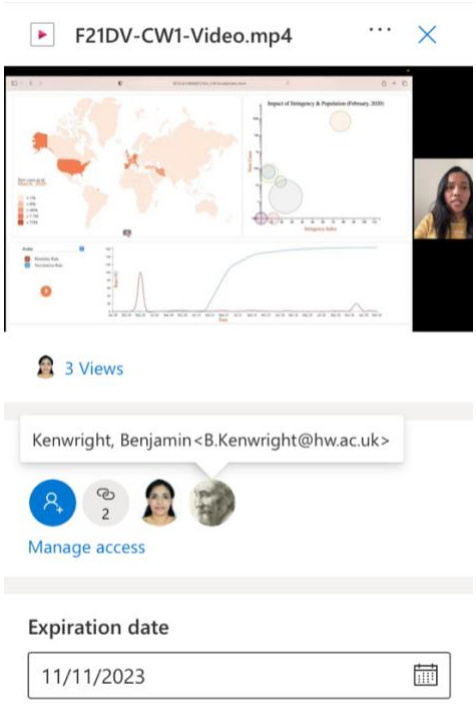
```
/* The driving file for all the visualizations.
Tasks:
  Implement the K-Means algorithm from scratch
  Data - between wave 1 and wave 2, and wave 2 and wave 3 of covid.
  Perform clustering based on user selection.
Functions:
  kMeans() - perform clustering
  scatterPlot() - set up the base for scatter plot
  clusterAndRenderScatterPlot() - call kmeans and render scatter plot with results
*/

import { fixData } from "./heatmap.js";
import {
  wave1year,
  wave2year,
  wave3year,
  wave1month,
  wave2month,
  wave3month,
} from "./app.js";

const margin = { top: 10, bottom: 45, left: 90, right: 30 },
  width = 1000 - margin.left - margin.right,
```

C6. You must explain your design and implementation choices in your report – **Met**.

- All visualizations and the reasoning behind their implementation choice has been explained in previous sections.

C7. You must demonstrate your application with a submitted video – **Met**.

- Recorded video, uploaded to OneDrive and shared with course tutor, B.Kenwright.
- [F21DV-CW1-Video.mp4]



References

(COVID-19), W.H.O.W. *Russian Federation Situation*. Available at: https://covid19.who.int/region/euro/country/ru (Accessed: 16 March 2023.

Bostock, M. (2012) *Let's Make a Map*. Available at: https://bost.ocks.org/mike/map/ (Accessed: 16 March 2023.

Bostock, M. (2021) 'D3.js - Data-Driven Documents'. Available at: https://d3js.org/.

Gallery, T.D.G. (2018) *The D3 Graph Gallery*. Available at: https://d3-graph-gallery.com/ (Accessed: 16 March 2023.

Mathieu, E. (2022) *How do key COVID-19 metrics compare to previous waves*. Available at: https://ourworldindata.org/covid-metrics-previous-waves (Accessed: 16 March 2023.