Data Exploration and Visualisation

Coursework 2

Coursework Outline

This is the second coursework in Data Exploration and Visualisation module. The coursework provides you with an opportunity to demonstrate the knowledge and skills that you have accumulated in data visualisation so far. The coursework particularly focuses on interactive data visualisation and animation. Like any real-life applications, the coursework requires preparing the data for visualisation by re-formatting the data followed by identifying any missing values and proposing solutions to deal with them.

The project has a 50% weight of the overall mark.

Please submit a zipped file that contains the following:

- A python file with the codes for all questions (answers should be numbered clearly).
- A document (word or PDF) for any other (non-code) answers + screenshot of map from Q3 and Q6 + screenshots of the final web app from Q4 and Q5 (at least one for the bubble map and one for the choropleth).

For any question, please contact the module leader at maysson.ibrahim@buckingham.ac.uk

Dataset description

In this coursework, you will be using the Oxford COVID-19 Government Response Tracker (OxCGRT) dataset [1]. In addition to recording the total confirmed and death cases on daily basis, OxCGRT provides a systematic way to track the stringency of government responses to COVID-19 across countries and time. It uses a novel index that combines various measures of government responses. The data is collected and updated in real time by a team of dozens of students and staff at Oxford University.

[1] https://www.bsg.ox.ac.uk/research/research-projects/oxford-covid-19-government-response-tracker

The dataset needed for the practical elements of this coursework is given to you in two csv files accessible via Moodle “OxCGRT_summary20200520.csv” and “country-and-continent.csv”.

The “OxCGRT_summary20200520.csv” includes data from 1st March to 20th May with the following columns:

- CountryName
- CountryCode
- Date
- School closing
- Stay at home requirements
- ConfirmedCases
- ConfirmedDeaths
- StringencyIndex

Name	Description	Coding
School closing	Record closings of schools and universities	0 - no measures 1 - recommend closing 2 - require closing (only some levels or categories, e.g. just high school, or just public schools) 3 - require closing all levels
Stay at home requirements	Record orders to "shelter-in-place" and otherwise confine to the home	0 - no measures 1 - recommend not leaving house 2 - require not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips 3 - require not leaving house with minimal exceptions (e.g. allowed to leave once a week, or only one person can leave at a time, etc)

The “country-to-continent.csv” includes the following columns:

- Continent_Name
- CountryCode

Data preparation and visualisation

Q1. As a data preparation exercise, write the code needed to produce a single dataframe from the two files in which the Continent_Name is added to the set of columns from “OxCGRT_summary20200520.csv”. If any countries in the output have a missing value for the Continent_Name, replace it with a relevant value. Note that the output of the above merge will result in multiple entries for transcontinental countries such as Turkey and Azerbaijan (no actions required).

(5 marks)

Q2. The columns of ConfirmedCases and ConfirmedDeaths have missing values. To prepare the data for visualisation, choose an appropriate strategy to handle the missing values in the data. Justify your choice and write the code needed to implement it.

Note: After preparing the data and handling the missing values, use the output data to answer **Q3**, **Q4**, and **Q5**.

(10 marks)

Q3. Create a world map with bubbles to show the total number of confirmed cases in each country in addition to animation showing how the numbers change over the given time (i.e. between 1st March and 20th May). The bubbles should be coloured by continents.

(10 marks)

Q4. Write the code needed to create a web app using Dash to run on a local server (see snapshots below). The web app should include the following components:

1. The bubble map created in Q3.

2. **Dropdown menu** to represent “**Scope**” with the following values:
   - ✓ World (the default value)
   - ✓ Asia
   - ✓ Africa
   - ✓ Europe
   - ✓ North America
   - ✓ South America

When a user changes the scope from the dropdown menu, the map should be updated automatically to show only the chosen continent(s).

3. **Radio Items** to represent **"Data Input"** with the following values:
   - ✓ Confirmed Cases (the default value)
   - ✓ Confirmed Deaths
   - ✓ Stringency Index

   The bubbles on the map of the selected scope should change based on user's inputs e.g. if a user selects "confirmed Deaths" from the **Data input** and "Europe" from the **Scope**, the output should be the map of Europe with bubbles representing the number of confirmed Deaths in each country.

4. **Radio Items** to represent the "**Policy**" with the following values:
   - ✓ Not selected (the default value)
   - ✓ School closing
   - ✓ Staying at home

   When a user selects "school closing" or "staying at home", the bubble map should turn into choropleth map. The colours on the map should change according to the selected policy in different countries and shown only in the selected scope e.g. if a user selects "school closing" from the **Policy** and "Europe" from the **Scope**, the output should be the map of Europe only coloured according to the school closing values in different countries.
   To activate the bubble map again, the user should selects the "Not selected" option from the **Policy** menu.

Note1: the title of the map should change according to the chosen options e.g. the title is "*confirmed Deaths in Europe*" if the selected options are Europe for **Scope** and confirmed Deaths for **Data input**.
Note2: The animation (based on Date) should be functional in all the above options.
Note3: The **Policy** options (School closing and Stay at home requirements) override the **Data input** options (confirmed cases, confirmed deaths, and stringency index) e.g. if confirmed cases and School closing options were selected, the map will be updated according to the School closing. That means the **Data input** is considered only if the selected **Policy** is "Not selected" (see the snapshots below for examples).

Note4: The screenshots below show simple design of the app to clarify the requirements. Your app doesn't need to have the same design (be creative!).

(40 marks)

Q5. Answer the following questions:

a) Write the code needed to find the top 5 countries with the highest number of **confirmed cases** as of 20th May 2020.

b) Add a line graph in a second figure to the web app (see snapshots below) to compare the number of confirmed cases of the five countries over the given time (i.e. between 1st March and 20th May).

The line graph of the same five countries can be updated based on user's choice of

- **Data input** (confirmed cases, confirmed deaths, or stringency index).
- **Policy** (School closing or Stay at home requirements).

Note1: The y-axis and the title of the line graph should change according to the chosen option e.g. the title is "*confirmedDeaths in the top five countries*" if the selected option is confirmed Deaths.

Note2: When a user chooses to show the confirmed cases on the line graph, the log scale should be used for the y-axis.

Note3: The components in the screenshots are arranged in a specific layout. You have the choice of using the same layout or proposing a different one. More information on using layouts in Dash can be found at https://dash.plotly.com/layout

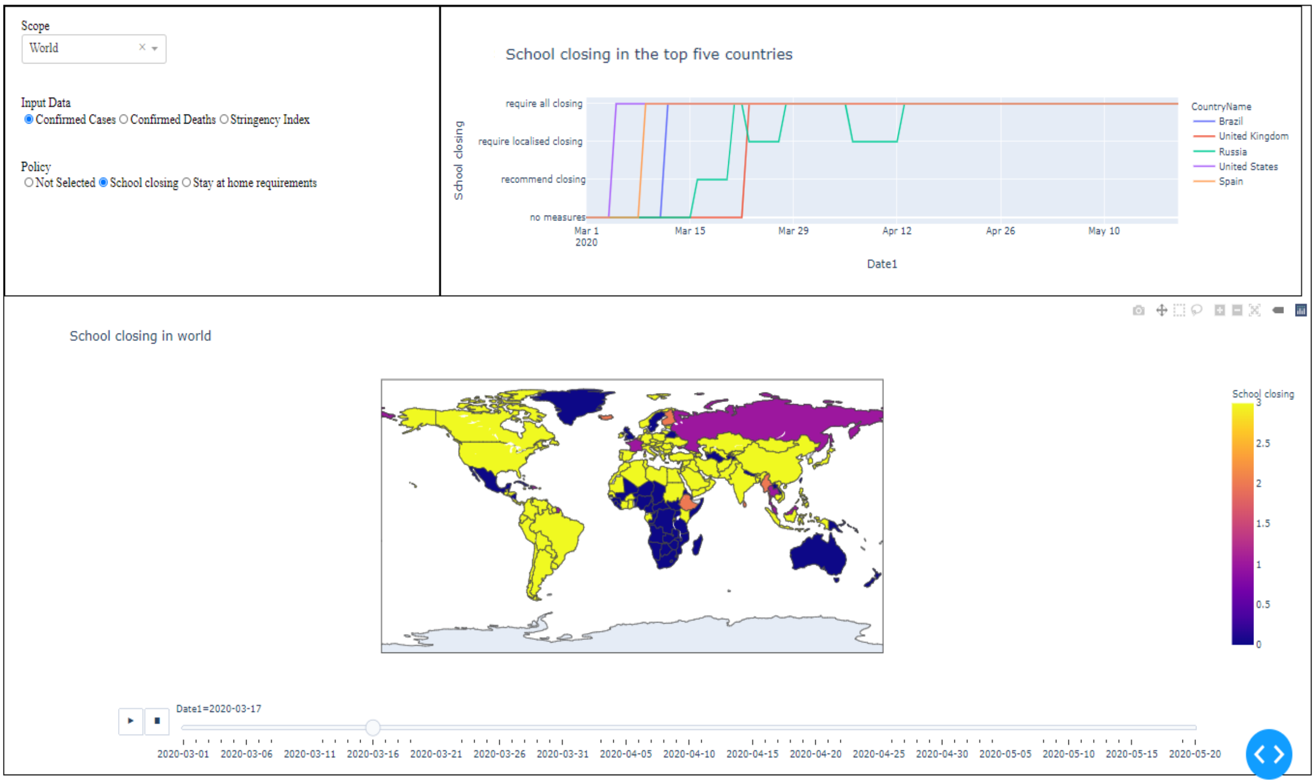
(25 marks)

Q6. A person drove from London to Dover, took a ferry from Dover to Calais, and drove from Calais to Charles de Gaulle Airport in Paris. The person then took a flight to Istanbul Airport. Use Mapbox to visualise the whole journey.

Note: this question is not related to the OxCGRT data or the previous questions.

(10 marks)

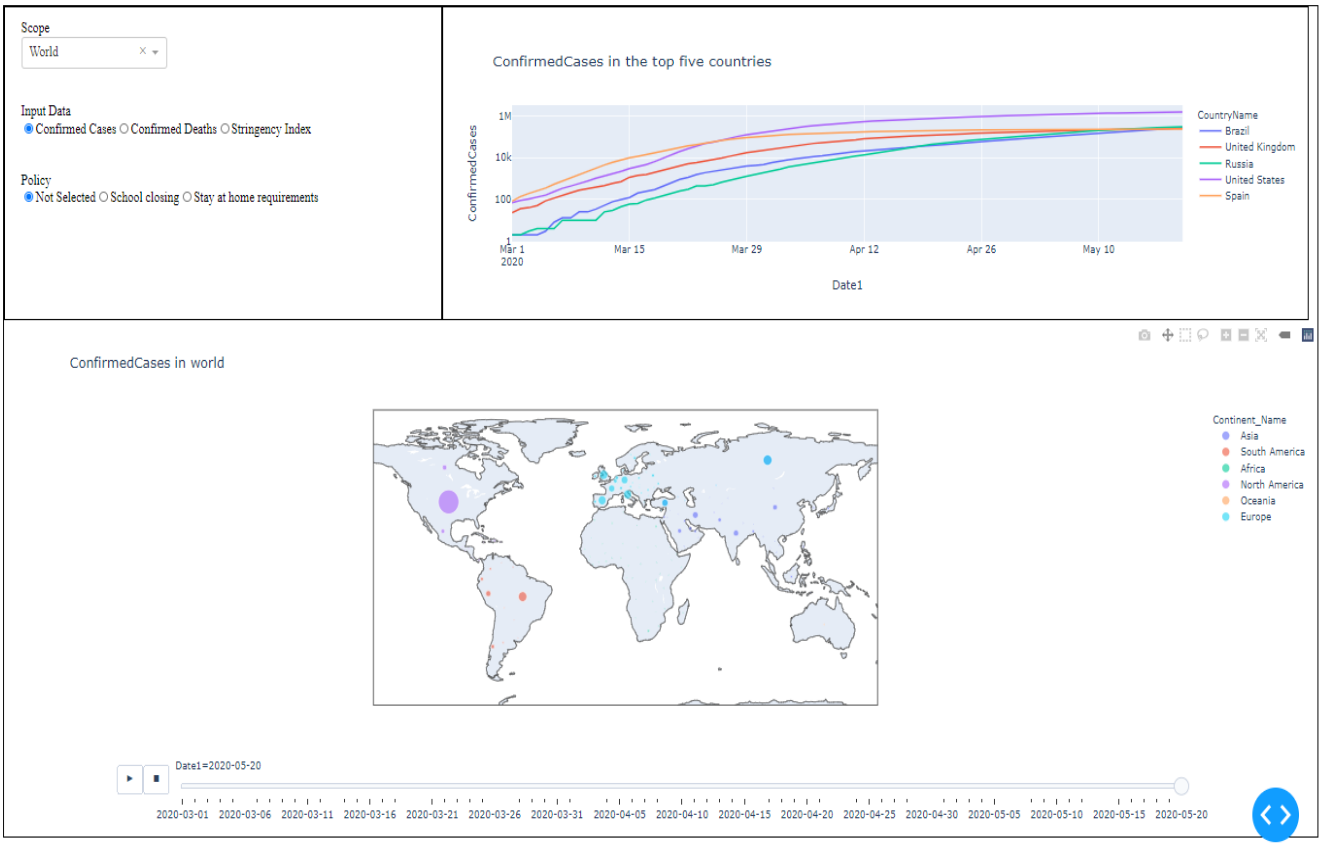
Snapshots of the Dash web app



Scope
World

Input Data
● Confirmed Cases ○ Confirmed Deaths ○ Stringency Index

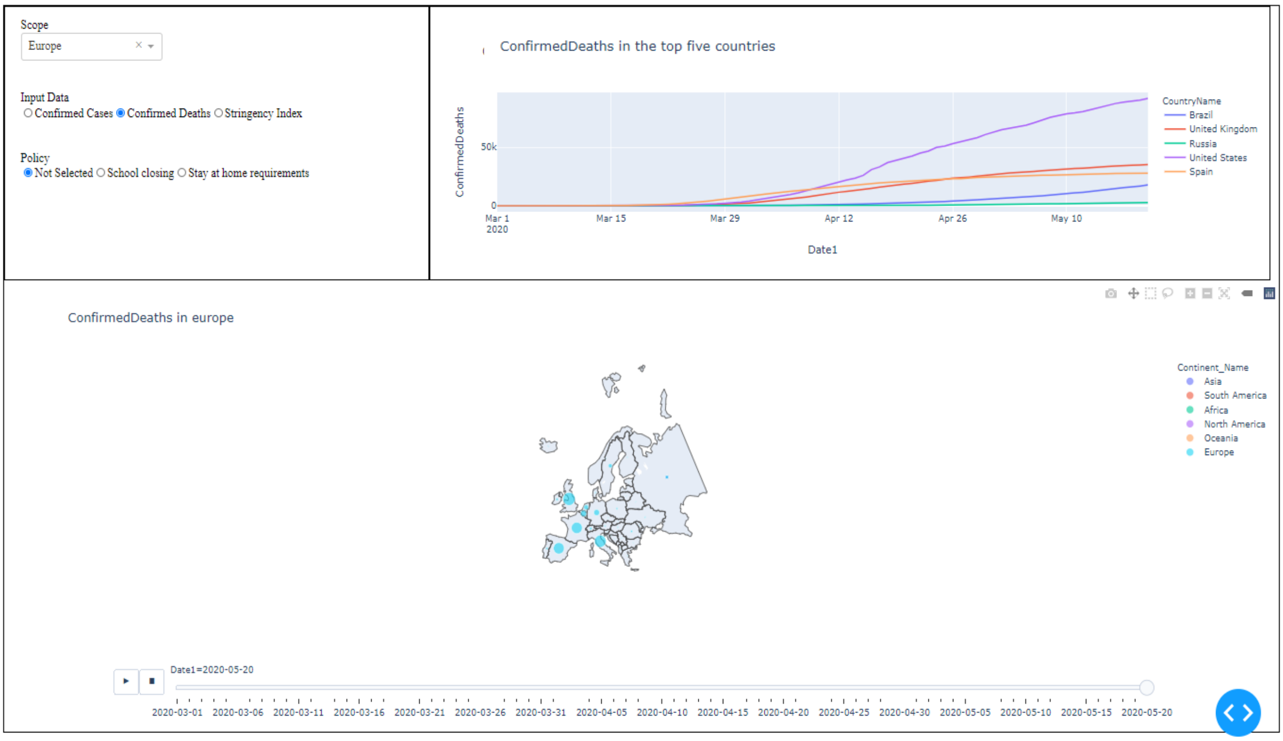
Policy
○ Not Selected ● School closing ○ Stay at home requirements



Scope
World

Input Data
● Confirmed Cases ○ Confirmed Deaths ○ Stringency Index

Policy
● Not Selected ○ School closing ○ Stay at home requirements



Scope

Europe

Input Data

○ Confirmed Cases ● Confirmed Deaths ○ Stringency Index

Policy

● Not Selected ○ School closing ○ Stay at home requirements

Marking Matrix for this coursework

Work Aspects	**Distinction** **≥ 70%**	**Merit** **≥ 60%**	**Pass** **≥ 50%**	**Fail**
Accuracy	Precise and correct terms used	Sufficiently precise. Most terms used correctly	Reasonably accurate in context but not in words	Severe lack of precision and misunderstanding
Validity	Argument consistent and logical. Show strong critical reasoning	Good logical argument. Show limited critical thinking & reasoning	Sufficiently valid argument, but may not with proper reasoning	Little valid argument. Opinionated decisions
Completeness	All required elements covered	Majority elements covered	Sufficient elements covered	Severely incomplete work
Objectivity	Factual not opinionated	Mainly factual	Limited or not well argued	No objectivity
Clarity and Professionalism	Statements clearly made, diagrams, figures and references professionally presented	Statements easy to follow, but may not be carefully built. reasonable use of figure reference, diagrams	Sufficiently clear to follow. Use of figures, diagrams and references is present	Severely lack of clarity. Extremely limited in content. No sign of professional “look and feel”.