Assignment-5 (COVID-19) Basic Information Sheet

1. Data

1.1 Data source:

- Name: **Our World in Data (University of Oxford)**
- URL: https://ourworldindata.org/coronavirus
- CSV url for reading data: https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv

1.2 Dataset characteristics (Type of data and description):

Column Name	Data Types	Description
iso_code	Nominal data	3-letter country codes
continent	Nominal data	Continent of the geographical location
location	Nominal data	Geographical location
date	Ordinal data	Date of observation
total_cases	Quantitative data	Total confirmed cases of COVID-19
new_cases	Quantitative data	New confirmed cases of COVID-19
total_deaths	Quantitative data	Total deaths attributed to COVID-19
new_deaths	Quantitative data	New deaths attributed to COVID-19
total_cases_per_million	Quantitative data	Total confirmed cases of COVID-19 per 1,000,000 people
new_cases_per_million	Quantitative data	New confirmed cases of COVID-19 per 1,000,000 people
total_deaths_per_million	Quantitative data	Total deaths attributed to COVID-19 per 1,000,000 people
new_deaths_per_million	Quantitative data	New deaths attributed to COVID-19 per 1,000,000 people
total_tests	Quantitative data	Total tests for COVID-19

new_tests	Quantitative data	New tests for COVID-19
new_tests_smoothed	Quantitative data	New tests for COVID-19 (7-day smoothed). For countries that don't report testing data on a daily basis, we assume that testing changed equally on a daily basis over any periods in which no data was reported. This produces a complete series of daily figures, which is then averaged over a rolling 7-day window
total_tests_per_thousand	Quantitative data	Total tests for COVID-19 per 1,000 people
new_tests_per_thousand	Quantitative data	New tests for COVID-19 per 1,000 people
new_tests_smoothed_per_thousand	Quantitative data	New tests for COVID-19 (7-day smoothed) per 1,000 people
tests_units	Nominal data	Units used by the location to report its testing data
stringency_index	Quantitative data	Government Response Stringency Index
population	Quantitative data	Population in 2020
population_density	Quantitative data	Number of people divided by land area, measured in square kilometers, most recent year available
median_age	Quantitative data	Median age of the population, UN projection for 2020
aged_65_older	Quantitative data	Share of the population that is 65 years and older, most recent year available
aged_70_older	Quantitative data	Share of the population that is 70 years and older in 2015
gdp_per_capita	Quantitative data	Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available
extreme_poverty	Quantitative data	Share of the population living in extreme poverty, most recent year available since 2010

cvd_death_rate	Quantitative data	Death rate from cardiovascular disease in 2017
diabetes_prevalence	Quantitative data	Diabetes prevalence (% of population aged 20 to 79) in 2017
female_smokers	Quantitative data	Share of women who smoke, most recent year available
male_smokers	Quantitative data	Share of men who smoke, most recent year available
handwashing_facilities	Quantitative data	Share of the population with basic hand washing facilities on premises, most recent year available
hospital_beds_per_thousand	Quantitative data	Hospital beds per 1,000 people, most recent year available since 2010

1.3 More dataset characteristics:

- Brodlie model: E_{33}^{P} (Multivariate data, n=33).
- This dataset also contains a time dimension i.e. date, daily data is available for each country (Temporal data).
- Dynamic dataset: This dataset is updated daily and new data is added for each country every day.
- Data Reliability: As the dataset is dynamic some of the column values (e.g. total_tests) for the most recent date might not be available, to counter this problem the recent available data is taken into consideration.

(Note: Additional columns may be added to the dataset in the future, we are considering columns as described in the concept paper.)

2. User and Task

2.1 Potential users:

- Researchers who are studying COVID-19, Pandemics, etc.
- Multimedia users

2.2 Tasks:

1. stringency_index and date columns are plotted for each country (line graph) in Europe which will help the user to see the cases per day with the stringency measures imposed by the government (i.e. measures imposed by the government are actually helping or not?).
2. A parallel coordinates plot is created by using the columns location, hospital_beds_per_thousand, covid19_death_rate = (total_deaths/total_cases)*100 (recent date is considered for total_deaths, total_cases), median_age, population to see how hospital systems (i.e. beds) in a particular country affect covid19_death_rate and what is the pattern between median_age, population and covid19_death_rate.
3. A pie chart is created for continent Europe which will include the percentage of tests that are performed by that country as compared to the entire Europe. The interpretation is the more the tests performed in a country the more reliable the numbers (i.e. total_cases) are.
4. A map is created based on the covid19_death_rate = (total_deaths/total_cases)*100 (recent date is considered for total_deaths, total_cases). Each country map is colored based on the death rate.

3. Visualization(s) Techniques

Types of Visualization Techniques used are as follows:

Task	Data Variable		Visual Variable
Line Graph for Multivariate Data	date	map to	X axis (Position)
	stringency_index	map to	Y axis (Position)
	location	map to	Color
Parallel Coordinates plot	location	map to	First axis (Position), Color
	hospitals_beds_per_thousand	map to	Second axis (Position)
	median_age	map to	Third axis (Position)

	population	map to	Fourth axis (Position)
	life_expectensy	map to	Fifth axis (Position)
	covid19_death_rate	map to	Sixth axis (Position)
Pie Chart	location	map to	Color
Choropleth map	iso_code	emap to	country map (Position)
	iso_code	map to	Color

4. Interaction

4.1 Interaction operators:

- **Navigation:**
  - In Task 1, hovering on each line displays date wise data containing the country name, date, government stringency index (0-100), total confirmed cases, total deaths, new confirmed cases and new deaths.
  - In Task 3, by pointing on sectors (countries) of the pie chart will display the total tests, recent data available date.
  - In Task 4, hovering on individual countries will display COVID19 death rate in %, country name, total cases, total deaths and iso_code of the recent date.
- **Filtering:**
  - In Task 2, individual records can be filtered by specifying filter criteria on each axis and pattern of that record can be isolated.
  - In Task 4, after double clicking on a particular legend then it will display that country and hide all other countries.
- **Connection Operators:**
  - In Task 2, isolating clusters in parallel coordinates.

4.2 Interaction operands:

- **Data Structure Space:**
  - In Task 1, any line of the line graph can be selected to get details of particular time span.
  - In Task 2, any record of a particular axis can be isolated to get a detailed view.

- **Data Value Space:**
  - In Task 4, zooming the maps can give a clear image of a particular country.

5. Participants

Varun Nandkumar Golani	Lalith Sagar Devagudi	Sanjay Chhataru Gupta
1. Research on data sets. 2. Input and writing concept papers. 3. Input and idea on creating tasks. 4. Dataset read, task 1 and task 3 implementation. 5. Writing information sheet.	1. Research on visualization techniques. 2. Input and writing concept paper. 3. Input and idea on creating tasks. 4. Task 2 implementation. 5. Writing information sheet.	1. Research on potential users, interaction and dash python framework. 2. Input and writing concept paper. 3. Input and idea on creating tasks. 4. Setting up Github for code sharing and version control. 5. Task 4 and dash app implementation. 6. Writing information sheet.

6. Concept

All of the tasks except one subtask have been completed from the concept paper. For subtasks that were not implemented and the additional tasks/subtasks that have been worked on, please refer to section 6.1 and 6.2 respectively.

6.1 Subtasks not implemented

- In the concept paper, subtask for Task 4 it was specified that " a summary will be shown for the top 5 countries suffering from COVID-19 in Europe" is not implemented.

6.2 Additional tasks worked on

- Our application uses the dataset that is taken directly from the github url and our visualizations are adapted in such a way to deal with the dynamic data so that the user can view the recent trends with our visualizations.
- All the visualizations are integrated in a single dash application using tabs.

7. Code Specific Information

7.1 Libraries used

1. Pandas - To read and store the datasets in a Data frame to use in coding..
2. Plotly Express - To plot Line Graph, Pie chart and Choropleth Map for Multivariate data.
3. Plotly graph objects - To plot the Parallel Coordinates.
4. Dash - To get a web application framework where all the different visualizations can be displayed in a single html file.
5. Dash core components - To create Tabs components for user interface.
6. Dash html components - To compose the html like layout using Python structure.
7. Dash dependencies - To get user input actions of selection of tabs.

7.2 Stylesheet

We are using a custom CSS stylesheet to modify the default styles of the elements in this application.
i.e.-
external_stylesheets = ['https://codepen.io/chriddyp/pen/bWLwgP.css']
app = dash.Dash(__name__, external_stylesheets=external_stylesheets)