Concept Paper

1. Data

1.1 Data source:

- Name: **Our World in Data (University of Oxford)**
- URL: https://ourworldindata.org/coronavirus

1.2 Data overview:

- This dataset contains data on a daily basis for all the countries.
- For data visualization, we will be considering all the countries from the continent Europe from this dataset.

1.3 Reason for using this dataset:

- This dataset contains records for all countries starting from the day that particular country started reporting data regarding COVID-19 and is updated daily.
- It includes data on confirmed cases, deaths, and testing, as well as other variables of potential interest.
- This data is collected from a variety of sources (United Nations, World Bank, Global Burden of Disease, Blavatnik School of Government, etc.)

1.4 Dataset characteristics:

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

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 will be 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?). Each country will be shown in different colors with a legend.
2. A parallel coordinates plot will be created by using the columns location, hospital_beds_per_thousand, covid19_death_rate = total_deaths/total_cases (recent date will be 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 will be created for continent Europe which will include the percentage of tests that are performed by that country as compared to the entire Europe.
4. A map will be created based on the covid19_death_rate = total_deaths/total_cases (recent date will be considered for total_deaths, total_cases). Each country map will be colored based on the death rate and a summary will be shown for the top 5 countries suffering from COVID-19 in Europe.

3. Visualization(s) Techniques

1. Line Graphs for Multivariate Data (Other Visualization Techniques) (Task 1)
2. Parallel Coordinates (Other Visualization Techniques) (Task 2)
3. Pie Chart (Other Visualization Techniques) (Task 3)
4. Choropleth map: Colored Area Maps (Geospatial Data) (Task 4)

4. Interaction

4.1 Interaction operators:

- Navigation (e.g. In Task 4, hovering on individual countries will display extra information like total_cases, total_deaths etc.)
- Filtering (e.g. In Task 2, individual records can filtered by specifying filter criteria on each axis)

4.2 Interaction operands:

- Data Structure Space (e.g. In Task 4, the data structure space is graph)