**Social Data Analysis and Visualisation: Project Video Questions**

* An explanation of the central idea behind your final project (what is the idea? which datasets do you need to explore the idea?, why is it interesting?)

The 2019 - 2020 Coronavirus pandemic will be one of the most influential health crises of the 21st century. COVID-19, the strain responsible for the pandemic, is a new strain of coronavirus, causing respiratory stress and disease. The health impact of the virus continues to be great concern world-wide. However, the economic impact of the COVID-19 pandemic will continue to be felt for months after the virus has left. Over the past several weeks, the global share market has dropped, businesses have closed, and governments have been forced to pledge funds to the race to find a vaccine and to help the struggling economy. Governments worldwide have introduced totalitarian social distancing and lockdown laws.

In this project, the link between coronavirus, government and economic data will be drawn upon to compare the response of governments to the pandemic across the Europe. Through the use of coronavirus data including cases by day and country, government data including dates and measures introduced to slow the virus as well as funding pledged in the race to find a vaccine and finally economic data including country debt, GDP, etc, this investigation will analyse how various country indicators have influenced governments responses to the pandemic. Information on the number of hospital beds and healthcare workers will also be sourced to see if there is any correlation between the capacity of the healthcare system and the government measures that have been put into place. This information will be interesting to find links between adopted measures and consequences of the spread of the pandemic. It will also be interesting to see how a countries wealth has impacted the number of coronavirus cases and the government measures that have been enforced.

To explore this idea, economic indicators for example corruption index, GDP and debt will need to be sourced. This information has been retrieved from the world bank data archive. A plethora of data is available from this source and as such, indictors further to the ones listed above have also been found and identified as interesting to the analysis to be conducted. A dataset online containing the government measures introduced worldwide by date was sourced. Finally, coronavirus data, including number of cases, deaths and recovered patients by day was sourced from <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>. To utilise all of this data, each dataset was individually cleaned and the country codes and names normalised to allow for easy comparison between the datasets. Each dataset contains many attributes and observations. These will need to be filtered and narrowed down to attributes that are relevant and important for this investigation.

* A mock up of the visualization that you wish to build.

Three datasets will be utilised throughout this project, an indicators dataset consisting of economic and government indicators for Europe, a lockdown measures dataset that details information about the coronavirus lockdown measures for each European country and finally a coronavirus dataset that has the number of cases, deaths and recovered patients for each country.

The possibilities for exploring this concept are endless. Currently, we have finalised 5 different visualisations to create and explore. This project will focus on the development of the pandemic in Europe. Europe comprise of 44 different countries. While it would be interesting to analyse all 44 countries, the visualisations are limited on space. For many visualisations, all European countries will be included. However, in the presence of so much data, only selected countries will be analysed in other visualisations.

* + 1. Heat Map Movie Over Time: Before diving into the analysis of the economic and government data, it is important to first understand how the COVID-19 pandemic has evolved. This will be shown through a heat map movie over time since the first coronavirus case was confirmed in Europe. This visualisation falls into the film/video/animation genre. As the first visualisation, it is designed to introduce the audience to the story and provide context to the remaining visualisation. The use of the film genre is to provide a light introduction to the story. The visualisation is author driven and as such requires minimal audience participation which, is suitable as an introductory visualisation. With this visualisation, audiences can see how the virus has spread over time.
    2. Healthcare Expenditure: Depending on the wealth of each country, the healthcare expenditure will change. This is interesting to visualise as it shows the funds that each country spends on healthcare and can be considered an indication of how much healthcare each country can afford. When compared with the GDP and number of coronavirus cases, links can be drawn between the economic status of the country and the severity of the coronavirus pandemic experienced in that region. This visualisation is designed as an introduction to the key parameters relating to the coronavirus pandemic, government and economic data. This visualisation falls under the annotated chart genre. It is a static visualisation that is primarily author driven.

(Bar plot with common y axis, each country would have 3 bars, each colour would represent a different measure. 3 colours 3 bars)

* + 1. Interactive Plot 1: Governments worldwide have introduced totalitarian social distancing and lockdown laws in order to curb the spread of the virus. Utilising the coronavirus data, it is interesting to visualisation the number of coronavirus cases and days since the confirmation of the first case before the government went into lockdown. A bubble plot style will be used to visualise this information. An interactive measure to the plot will also be added to give audiences details on demand. On hovering over the bubble of an individual country, information about the lockdown measure introduced will be revealed. This visualisation follows the annotated chart style however, with interactive elements. This visualisation is designed to show audiences the variation in lockdown measures and the timing of their introduction. It is designed to show the variation of the government responses to provoke thought and lead on to further visualisations where the reasons for the government actions and timings are explored.

(Similar to interactive histogram but with bubbles. Bubble plot, y = no of cases, x = date since first case, colour = country, interactive element could be the measure introduced)

* + 1. Interactive Plot 2: After exploring the lockdown measures and the timing of their introduction, it is then interesting to begin to explore the reasoning behind these measures. A crucial part of this investigation is exploring how the quality of the healthcare system has influenced the government measures. Questions like have some countries been quicker to put the country into lockdown because they don’t have the health system capacity, are interesting to explore. A bubble plot display the speed of lockdown introduction (i.e. number of days after the first case before the country went into lockdown) compared to the universal healthcare coverage index and the capacity of the healthcare system classified by the number of hospital beds. Other information, including the number of healthcare personnel will be provided adding an interactive element to the plot. This visualisation is a static visualisation and falls under the annotated chart style however, does include interactive elements. The visualisation is designed to show how the capacity of the healthcare systems has influenced government decisions during the pandemic.

(Bubble plot: x = speed of lockdown, y = universal\_healthcare\_coverage\_index, size of the bubbles = capacity of the health care system classified by the hospital\_beds\_per\_1000, colour = country)

* + 1. Political State: A visualisation analysing the political system of each country and how this has affected the spread of the coronavirus is also interesting to explore. Measures including government effectiveness and political stability have been from the world bank’s data archives. It is interesting to explore how the stability and effectiveness of the government has impacted the rate of the coronavirus spread. This will be shown in a scatter plot style visualisation. The visualisation is static, author driven and falls under the annotated chart visualisation genre.

(Bubble plot: x = rate (rate of increase in coronavirus cases), y = government effectiveness, colour = country, size = political stability)

* An outline on the elements you'll need to get to your goal.

To achieve this goal the following elements will be required:

* Data collection
* Data cleaning and Reduction
* Initial data explanations
* Initial data analysis
* Visualisation planning
* Creation of video
* Manipulation of data to produce desired visualisations
* Initial production of visualisations
* Editing of visualisations
* Refinement of visualisations
* Presentation of visualisations (titles, labels, etc.)
* Creation of website
* Explanation of motivations, basic statistics, data analysis, genre, visualisations and discussion
* Final editing
* Implementation Plan

A Gantt chart is an excellent way to visualise a project schedule. The following Gantt chart shows the schedule for this project and deadlines that need to be met. The chart also illustrates which group member is responsible for completing each required element of the project.

* A walk-through of your preliminary data-analysis

When conducting preliminary data analysis, the following data statistics were found. The properties of the data were also analysed and are listed here. When analysing the fundamental distributions of the data, it is important to look at the relationships between each of the indicators. The following matrix plot shows the distribution of each of the indicators but also the relationship between each of them. Finally, the following logarithmic plots show the general distribution of the number of active, death and recovered coronavirus patients. The exponential trend in these plots are clear and we can only hope that the pandemic curve will begin to flatten.