ISM Honours
Information and Knowledge in Organisations
Data Analysis
Assignment 1

An analysis of COVID-19 spread in South Africa

For this assignment you are required to produce an Rmarkdown notebook that reproduces as many of the analyses and visualisations displayed on the following webpage: https://mediahack.co.za/datastories/coronavirus/dashboard/ as you can.

Your notebook needs to include the code used to wrangle, and process the data, as well as the outputs and visualisations produced. Additionally, there needs to be a logical and coherent flow to your document, with headings and descriptions.

For this assignment, you do not need to programmatically collect the data (although you *can*). You should be able to find various openly accessible datasets to use for your analyses.

For data sources I recommend the following: Department of Health, National Institute for Communicable Diseases, Bhekisisa. European Centre for Disease Prevention and Control via Our World in Data, Johns Hopkins University CSSE, Worldometer, https://github.com/dsfsi/covid19za as places to start, but you can find your own datasets to use too.

For support on collecting webdata with R (if necessary, most likely you'll be able to download a dataset or datasets) see:
- https://compsocialscience.github.io/summer-institute/2019/materials/day2-digital-trace-data/screenscraping/rmarkdown/Screenscraping.html
- https://www.datacamp.com/community/tutorials/r-web-scraping-rvest

This task primarily focuses on data cleaning, wrangling, transformation, and visualisation with R and it is expected that it should take you upwards of 8 hours per day to complete.

You will be assessed on:
- The proportion of the expected visualisations you can produce and the extent to which your figures correspond to the expected figures (note, you are free to follow a different colour scheme [see: https://ggplot2-book.org/scale-colour.html or https://www.shanelynn.ie/themes-and-colours-for-r-ggplots-with-ggthemr/]). (**15%**)
- The organisation, coherence, and quality of your code for cleaning and wrangling the data. (**30%**)
- The organisation, coherence, and quality of your code for producing the figures. (**40%**)
- The structure and presentation (and writeup) of the R markdown notebook. (**15%**)

The assignment is to be completed in pairs.

Due Date: Tuesday 29 June 2021: 17h00.