**Exploratory Data Analysis Assignment 4**

*You can do this assignment individually or in groups of two*

*This assignment is worth 40% of your final mark*

The coronavirus disease (COVID-19) was first reported in Wuhan, China, on 31 December 2019 and has since rapidly spread to more than 120 countries.

In this assignment your goal is:

(a) to create an **animated visualization** of the spread of the coronavirus over time. The animation should be accompanied by a 1-2 page written piece (suitable for a newspaper or magazine) describing the animation and the patterns it reveals. If you use R markdown, the animation can be embedded into the document, otherwise it can be provided separately (e.g. as a GIF).

(b) to create **a Shiny app** that provides users with a snapshot of the current coronavirus situation and allows them to interrogate the dataset in different ways. The app should be accompanied by documentation (in any format e.g. Rmd, docx, pdf) explaining how to use the app and giving a short worked example demonstrating some of the functionality.

To do this, you can make use of data collated and made available by the Core Datasets project. To download the most up-to-date data, you can run the "get\_ncovr.R" script in the assignment folder, or type:

ncov <- read.csv("https://raw.githubusercontent.com/datasets/covid-19/master/data/countries-aggregated.csv")

I have downloaded data current as of 30 March 2020, and put this data on Vula. You are welcome to use this dataset for your assignement, or you can also use more up-to-date data if you like.

If you inspect the "ncov" object you will see it consists of 8 variables:

> str(ncov)

data.frame': 25760 obs. of 8 variables:

$ Province.State: Factor w/ 316 levels "","Adams, IN",..: 1 1 1 1 1 1 1 1 1 1 ...

$ Country.Region: Factor w/ 152 levels "Afghanistan",..: 138 138 138 138 138 138 138 138 $ Lat : num 15 15 15 15 15 15 15 15 15 15 ...

$ Long : num 101 101 101 101 101 101 101 101 101 101 ...

$ Date : Factor w/ 56 levels "2020-01-22","2020-01-23",..: 1 2 3 4 5 6 7 8 9 10

$ Confirmed : int 2 3 5 7 8 8 14 14 14 19 ...

$ Recovered : int 0 0 0 0 2 2 5 5 5 5 ...

$ Deaths : int 0 0 0 0 0 0 0 0 0 0 ...

The data frame contains country ID variables, country latitude/longitude coordinates, a date variable, and cumulative counts of confirmed coronavirus cases, recovered cases, and deaths. These constitute the raw data for the assignment

**Project deliverables**

Your project consists of the following four items, with the % contribution to the final project mark:

1. A GIF of your animation, and an R file containing the code for generating your animation. (40%)
2. Your 1-2 page document describing the animation and the patterns it reveals. (10%)
3. Your R Shiny app. This should be an R script called app.R made up of a user interface object (ui) and a server function (server). All code in the app should be clearly commented (40%).
4. Documentation (in any format e.g. Rmd, docx, pdf) explaining how to use the app and giving a short worked example (10%).

This assignment is open ended. Use your imagination about what kind of animation and app you want to create.

All code in the app should be clearly commented. The key thing is that I need to be able to run your code/app and reproduce your results, so there should be clear instructions on how to use your script(s). The code itself should not be displayed in the final typeset document (use “echo = FALSE” for Rmd files) and not pasted as an appendix in your report. Please make sure that your code/app works by testing it out first on another machine before submitting it.

Assignment hand-in is via the "Assignments" tab on Vula, and you should submit a single .zip file containing all files.

**Please note that the submission deadline has been moved from 6th March while UCT is closed. Please consult Vula for the latest information – the deadline will be, roughly, a week or so after term 2 starts.**

Late submissions are penalized at a rate of 10% per day.

**Additional resources**

There are a few examples of coronavirus trackers and apps online (see links below). Feel free to use these to get ideas. You can also borrow bits and pieces of code you find online, but any resources (including any code you borrow, and any visualization ideas) that you use must be clearly acknowledged and cited, and the bulk of your code should be your own. If in doubt, rather cite someone's work, and check with me if you are uncertain about plagiarism.

- <https://github.com/pzhaonet/ncovr>

- <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (see the link to the "Situation dashboard")

- Coronavirus tracker at <https://coronavirus.john-coene.com/#/> with code at <https://github.com/JohnCoene/coronavirus>

- <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html>

- <https://www.worldometers.info/coronavirus/>