**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 80 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 Peng Zhao through the R package "ncovr" (see “Additional resources” below). To download the most up-to-date data, you can run the "get\_ncovr.R" script in the assignment folder. I have downloaded data current as of 5 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 5 data frames

> str(ncov, max.level=1)

List of 5

$ area :'data.frame': 8369 obs. of 23 variables:

$ overall :'data.frame': 1 obs. of 23 variables:

$ provinceName: chr [1:115]

$ news :'data.frame': 10 obs. of 7 variables:

$ rumors :'data.frame': 10 obs. of 6 variables:

Of this, we are only interested in the "area" data frame, which contains cumulative counts of confirmed coronavirus cases, cured cases, deaths, with additional variables for country and time. There are many variables in the dataset, but the ones we are interested in are:

- confirmedCount

- currentConfirmedCount

- curedCount

- deadCount

- countryEnglishName

- continentEnglishName

- updateTime

Some of the other variables are self-explanatory (giving additional city or province information for Chinese cases, for example), but others are not (e.g. "comment"). You are not expected to use any of these columns (you can if you want, but I haven't looked at them and don't know what they mean, so you are more-or-less on your own).

**Some additional notes:**

1) In most but not all cases currentConfirmedCount + curedCount + deadCount = confirmedCount.

2) In some cases there is an NA value for currentConfirmedCount. In those cases there is a value for another variable "suspectedCount". However in most cases suspectedCount + curedCount + deadCount does not = confirmedCount. There are no rows where suspectedCount > 0 AND currentConfirmedCount > 0. I suspect these numbers are for cases where precise numbers are hard to come by.

3) the "updateTime" variable is in Julian date format. To convert to a regular datetime, use the "conv\_time" function in the "get\_ncov.R" script (this step is done for you in the script)

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

**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%).

All code in the app should be clearly commented (60%). 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. The submission deadline is on or before 23:59:59 on 6 April 2020. 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>