BCB BSc (Hons) Intro R Assignment

Due date: 11th March 2020

The emergence of the COVID-19 epidemic
This self-study component of the Intro R Workshop is about the emergence of the corona virus, or COVID-19 as it is officially called.

Please make sure that you collect data on COVID-19 up until a day or two before the report deadline. New data will become available each day, so include it daily and append it to your database. Please ensure that you use reliable sources (*i.e.* provide a justification for what is considered 'reliable' in the Methods section). These data need to be presented graphically in a way that both the spread of the disease (*i.e.* through the world as a function of time) and the rate of infections / deaths (again as a function of time, and per geographical area/country) are clear. Tables may also be used to supplement the graphical displays. Any kind of ancillary information will be valued—*i.e.*, you may want to report the infections and/or deaths also as a function of gender or age. Use your imagination.

Contrast this emerging disease to a previous epidemic or pandemic (*i.e.* provide matching graphs *etc.* so that the two diseases can clearly be compared). Since we have not yet done the BioStats component of Basic Stats, I'll not require a full statistical treatment of the data. Data summaries (Tables and Figures) are sufficient.

Please pay specific attention to the graphs that you create by ensuring that they are of publication quality. This therefore requires that you do not use the default **ggplot2** theme. Either create your own theme (more marks) or use a theme that comes with **ggplot2** or one of the other theme packages. Also ensure that strict attention is paid to following the South African Journal of Botany author guidelines. Marks will be deducted for any deviations from these author guidelines, so sticking with the 'recipe' is a good way for ensuring you don't unnecessarily lose marks. Formatting of the document, following the guidelines w.r.t. the referencing format, and ensuring publication quality Figures and Tables will count 20% of the report mark.

Ensure that the R script is fully annotated and that it will run on our computers without modification. This script together with the data in .csv format will be submitted together with the report. The R script and data will count 40% of the report mark

A report is comprised of an Introduction, Methods, Results, Discussion, Conclusion, and References. For this assignment each student is required to write a mini-report of no longer than 12 pages, including all References, Figures, and Tables. The general writing (structuring, following scientific conventions, logical flow of information, reasoning, discussion, and synthesis) of the content in these various sections will count another 40% of the report mark.

Instructions regarding content headings
Introduction
- Background information
- Importance
- Limitations
- Justification of the present work
- Rationale
- Aims, objectives, and research questions

Methods
- Map of the affected locations, covering the extent of the data gathered and reported in your paper
- Approach to collecting the data, and a listing of the sources consulted
- Data processing and analysis
- Mention which software and packages were used to do these analyses (there are specific ways of citing R and the various packages used)

Results
- Visualisations of trends, patterns, maps, group differences, *etc.*—*i.e.* graphs and tables that relate to the research questions being addressed
- Explanation (no interpretation) of the outcome of the graphs and tables, focussing only on the things that

needed testing according to the aims, objectives, and research questions

Discussion
- State the major findings
- Explain the meaning of the findings and why they are important
- Relate the findings to those of similar studies and previous pandemics/epidemics
- Synthesis within state of knowledge within the field (epidemiology)
- Mention/explain alternative explanations of the findings
- State the relevance and/or novelty of the findings
- Acknowledge the study limitations

Conclusion
- Include the summary of the main points
- Leave the reader with a strong take-home message

Bonus marks
- Using Rmarkdown
- Creating graphs that has not already been done in class
- Creating a theme for your plots
- Providing statistical support for hypotheses

Final submission
Create a folder titled with your name and surname *e.g.* “amieroh_abrahams”. In this folder you will have the following:
- A final MS Word and PDF document
- A script with all of the necessary code used
- All of the datasets used within this research assignment
- If Rmarkdown was used, please attach the Rmarkdown scripts as well as the MS Word and PDF document

Tips
- Make comments throughout the script
- Clean scripts will only have a positive influence on the person marking which is important