ADS2 Group exercise

MI Stefan

24/03/2020

Understanding COVID-19

Due to recent events, this data analysis exercise uses data about the 2019-20 Coronavirus pandemic.

You will be asked to work with a real-life dataset about worldwide infections. This dataset comes from the following resource:

Max Roser, Hannah Ritchie and Esteban Ortiz-Ospina (2020) - "Coronavirus Disease (COVID-19) – Statistics and Research". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]

The data file you are given is from 17 March 2020.

You are asked to work in groups. This exercise has two parts. In part 1, you are asked to answer a few well-defined questions. Do make sure that the methods you use to answer those questions are well explained and documented.

In part 2, you are asked to formulate your own question and answer it based on the dataset.

What we look for

As graders, we will look for clarity of the explanations, completeness and correctness of the code, and thoughtfulness in the interpretation of results. A simple question, thoroughly answered, can get you more points than a very ambitious project that is poorly executed.

You are asked to submit an .Rmd file with code and with text explaining the code and interpreting the ersults. This file should fulfill the following:

- Markers should be able to knit it into a pdf file that shows code, code outputs and explanatory texts and is well formatted.
- The file should serve as a template for future reports. In particular, if a marker changes the data file to a more updated data file from the same resource, your code should run just the same without needing to be manually updated. (An updated data file will contain the same columns, but will have more rows, because it includes additional dates). Of course, you need to draw the conclusions you draw based on the dataset you have been given. But the code itself should run with updated versions of the dataset as well.

Data import

The data is provided in file COVID19_full_data.csv - this contains all cases reported up to 15 March 2020.

Every row is one day, from 21 January to 17 March 2020. The first column is for the date, the second for the total number of (cumulative) cases around the world. After that, there is a column for each country or territory.

You may have to clean or re-arrange the dataset for your purposes. Do whatever it is you need to do, but make sure you include the code for everything you do in this assignment. Code for importing data should

assume that the data file is in the user's working directory. Please do not include file paths that may reveal the identity of one of your group members.

Part 1: Extent of the pandemic

- As of 17 March, there are 15 countries with more than 1000 cases. Which 15 countries are there? Plot the number of total cases over time for each of those 15 countries.
- For the last 2 weeks (4 March to 17 March), look at the total number of deaths and the total number of cases worldwide. For every day during that period, provide an estimate of the death rate (total deaths divided by total cases), including a confindence interval, for each day in those two weeks. Plot that data.
- Plot the case numbers over time for both Italy and the UK. Italy has been the first country in Europe to see a major outbreak of COVID-19. In the UK, media have reported that the UK is around 14 days "behind" Italy. Can you use a plot to check whether this is plausible?

Part 2: Ask your own question

- Ask a question that can be answered using this dataset and choose a suitable method to answer it. Explain (briefly!) why you are interested in this particular qusetion. Explain your method in words, and show and interpret your results.
- The "method" you use could be, for instance, an inferential hypothesis test, a bootstrap-based analysis, or maybe even an especially clever way to visualise data - it is up to you. There is no need to look for methods outside of what you have learnt in your course so far - see what you can do with the toolkit you have available!