

So you are applying for that data scientist position? Try and impress us in 120 minutes.

Silvie Cinková

2025-08-05

Table of contents

1	What this “test” is for	1
2	Task for those that work on their own data	2
3	Task for those who want to recap on familiar data	2
3.1	Optional tasks	3

1 What this “test” is for

This test will not affect your classification if you are registered for the regular exam at the Charles University. It serves (1) your self-assessment test immediately after the course (without time to internalize what you were exposed to) and (2) as a practical feedback for the teacher about the efficiency of the course setup.

The task is a little project that encompasses virtually the entire curriculum. Test yourself whether you have an idea how to approach the task and are able to look up the corresponding workflows. If you get stuck with the coding, try to describe what you would do in prose.

Please write your project in a Quarto file and e-mail (or otherwise share) the result to the teacher as an html file. Use prose to comment what you are doing and how you interpret the results.

Please include a section of “Final rant” or “Final sigh” where you describe your learning experience with the datasets, the tools, DataCamp, the study materials, and their presentation. Be as specific as possible - which topic was particularly confusing, where did you get lost, did DataCamp help you keep on track, did you look at the teaching materials after the lectures, did that help?

Imagine you are sharing your experience with a colleague who considers attending a hypothetical next run of this course.

Also, please share your good practices from your previous studying/teaching experience for inspiration.

2 Task for those that work on their own data

Please showcase your data and your research question using R as much as possible.

3 Task for those who want to recap on familiar data

1. Read the file `datasets_ATRIUM/gapminder_metadata_filenames.tsv`.
2. Pick two files according to your liking.
3. Read each of them in and analyze it:
 1. how many countries
 2. how many years each country
3. Describe the table in words as well: is it a complete dataset, or would you have to do a lot of filtering to get the same years for each country?
4. Plot the socioeconomic indicator (most likely the third column).
5. Join the two tables that you have described.
6. Compute the correlation between the two indicators the two tables show: `cor(x,y, method = "pearson")`.
7. Render the file as html. In the unlikely case that you would need to include any images from your local computer (e.g. screenshots), replace the YAML header of your quarto file with the contents of the file `YAML_header_multiformat` from the `ATRIUM_resources` folder.

3.1 Optional tasks

If the tasks above were easy, try to level up and include in the rendered file:

1. Filter the metadata dataframe (read from `datasets_ATRIUM/gapminder_metadata_filenames.tsv`) automatically for (any number of) files whose names suggest a topic that interests you. To do that, look up the appropriate function in the `stringr` library cheatsheet and use it to find a string that indicates your favorite topic. For instance “employment”.
2. Process the files in a loop (with or without the joining)
3. Write the workflow as a function
4. Run the function in a loop and/or with the `purrr::map` or other `purrr::` function you look up.