

Case Study for Unit 6

Load the Gross Domestic Product data for the 190 ranked countries in this data set:

<https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv>

Load the educational data from this data set:

https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv

Original data sources (if the links above don't work):

<http://data.worldbank.org/data-catalog/GDP-ranking-table>

<http://data.worldbank.org/data-catalog/ed-stats>

Note: If you encounter NAs in your analysis, please delete those values and continue the analysis with the non-missing values. However, please include code to count the number of missing values for each variable used in the analysis.

Questions on Merged Data

- 1 Merge the data based on the country shortcode. How many of the IDs match?
- 2 Sort the data frame in ascending order by GDP (so United States is last). What is the 13th country in the resulting data frame?
- 3 What are the average GDP rankings for the "High income: OECD" and "High income: nonOECD" groups?
- 4 Plot the GDP for all of the countries. Use ggplot2 to color your plot by Income Group.
- 5 Cut the GDP ranking into 5 separate quantile groups. Make a table versus Income.Group. How many countries are Lower middle income but among the 38 nations with highest GDP?

Deliverable: Markdown file uploaded to GitHub containing the following

- Introduction to the project. The introduction should not start with "For my project I ...". The introduction needs to be written as if you are presenting the work to someone who has given you the data to analyze and wants to understand the result. In other words, pretend it's not a case study for a course. Pretend it's a presentation for a client.
- Code for downloading, tidying, and merging data in a R Markdown file.
- Brief explanations of the purpose of the code. The explanations should appear as a sentence or two before or after the code chunk. Even though you will not be hiding the code chunks (so that I can see the code), you need to pretend that the client can't see them.
- Code to answer the five questions above (plus the answers) in the same R Markdown file.
- Clear answers to the questions, including interpretation of the plots.
- Conclusion to the project. Summarize your findings from this exercise.
- The file must be readable in GitHub – 5 points off if I have to download the file to read it! In other words, don't forget to keep the md file!!

Submit the link to the file in GitHub via the space provided for Unit 6 Case Study in 2DS.

Rubric (100 points total):

- Assignment is submitted via a link to GitHub that results in a readable file (0 or 5 points)
- I can run the code on either my Mac or my PC with no modifications, except setting the working directory (0 – doesn't run to 5 – runs with no errors).
- Grammatical/spelling mistakes (0 – lots of mistakes to 5 - no mistakes)
- Correct answers for each of the five questions (10 points each, 50 points total). This piece includes a clear explanation of output and/or graphics.
- Correctly commented code (30 points)
- Correct estimate of NAs (5 points total).