Case Study 1
Submit the link to the file in GitHub via the space provided for in the Case Study 1 page in 2DS.

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 for some cases, please 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 Show the distribution of GDP value for all the countries and color plots by income group. Use ggplot2 to create your plot.
5 Provide summary statistics of GDP by income groups.
6 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. The code should be in a make file style, meaning that the source RMD document pulls in separate files for importing data, cleaning the data, and data analysis.
- 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. Just the code to answer the questions is not enough, even if the code is correct and gives the correct answer. You must state the answer in a complete sentence outside the code chunk.
- 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 in the Case Study 1 page 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)
- Introduction to the case study provides context for the data (5 points)
- Correct answers for each of the five questions (5 points for first question, 10 points each for others, 45 points total). This piece includes a clear explanation of output and/or graphics.
- Correctly commented code in modular style (30 points)
- Conclusion summarizes findings from the exercise in paragraph form (5 points).

********** Note Well **********

R Markdown files often do not render graphics and output from R Markdown chunks in GitHub. You must do one of two things:

1. Keep the MD (markdown) file and upload it along with the R Markdown file
2. Upload the folder that is created in the working directory R Markdown is knitted into HTML. The folder contains graphics and other output.

As noted in the rubric above, the Rmd/Md file must be readable (output included) in GitHub or five points will be deducted from your final score.