

Team project assignment

Submit a descriptive analysis on your collected data by Data Analysis 1. I expect one submission for each team via ceulearning. Submission deadline is 7th of November. (I extended the original deadline with a week.)

Task:

1. Upload your data to one of your member's GitHub page as '.csv' file. (1/15p)
2. Create an RMarkdown document where:
 - a. You load the data from web.(1/15p)
 - b. You create a **max** 2 page analysis. Each additional half pages will result on 2 points reduction from the overall score.
3. Sections and sub-tasks:
 - a. Introduction:
 - i. motivating and explaining your descriptive statistics (2-3 sentence) with an embedded *url* to your data source (github). (1/15p)
 - b. Data:
 - i. short description of your main variable(s) and how it was collected. (1/15p)
 - ii. Filtering and cleaning your original data (2/15p)
 - iii. Report a descriptive statistics table on your main variable(s) with: mean, median, standard deviation, range, min, max, 5% percentile, 95% percentile, number of observations, number of missing values. Take care of caption, names and formatting! (4/15p)
 - iv. Show a graph on the distribution of your main variable. Explain why you have chosen such graph and bandwidth/number of bins. Take care of labels, placing and formatting of your graph! (4/15p)
 - c. Conclusion:
 - i. Write 1-2 sentence, what we can learn from this exercise.
4. You need to knit your RMarkdown document in pdf and in html format as well and upload it along with the .Rmd file. (1p)

You may use as a guidance the '*bpp_report.Rmd*' file that we have covered in class. Also, this is a team project, so make sure everybody gets how you carry out this project and everybody could run it on his/her computer in principle.

Good luck!

Agoston