**DATA100 PROJECT INSTRUCTION – FALL 2022**

**Due 4pm sharp on Tuesday, December 13th, 2022 (final exam is on December 11th, 2022)**

**The project has three parts: (1) Form a group; (2) Analyze data; (3) Write a final report.**
**The instructions for each part are listed below.**

## 1. PART I: FORM A GROUP

(1) Form a group using the MyLS Groups feature. The instructor will not deal with group issues except for random assignment. Minimum group size is 4, maximum is 6, no exceptions.
(2) Please put down the names of your teammates in alphabetical order by last name.
(3) Please identify the team leader using **bold font** for the team leader's **last name**.
(4) Make your own unique team name!
(5) Make a table similar to the one below at the beginning of your `.Rmd` report file, and put all the above information in it.

**TEAM NAME:** _____

| Last Name (Family) | First Name (Given) | Student ID |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

## 2. Part II: Exploratory Data Analysis

(1) You will carry out a data analysis on data provided using `RStudio` to address the questions listed in Project Questions.

(2) I expect to see you use the skills learned in this course to analyze data including data wrangling, exploratory data analysis (plots and summary statistics), modelling (regression analysis) and others.

(3) Your project does not need to analyze all variables: you may choose not to use all observations; or make new variables from the data that may be more suitable for answering your questions of interest; or make new data sets from the ones given to illustrate your point.

(4) The goal of this project is not to carry out an exhaustive analysis, nor to apply everything you have learned in the course. The goal is to demonstrate that you have learned how to use `RStudio`, that you can appropriately apply the methods and skills you learned in the course to address a question, and that you can effectively interpret and present the results.

(5) If you have further questions that you want to address, please go ahead and add them in. I am looking forward to seeing some interesting suggestions in your reports.

(6) Give complete citation of the source for the online article you choose to discuss and briefly summarize the main point(s) of the article.

(7) Relate the main point(s) in the article to your analysis of the data, as evidence either *for* or *against* it (them).

## 3. Part III: Final written report

(1) You need to submit an electronic copy of the final report to the Dropbox on MyLS by 4pm on Thursday, December 13, 2022.

(2) The submitted final report should consists of the following files

- An `.Rmd` file written using `R Markdown`, which is the main file containing all your `R` code and report. The name of the file should contain the name of your team as below:
  `<Team Name>-Report.Rmd`
- An `.html` file which is knitted from the above `.Rmd` file, which should be the report that can be put up as a webpage. It should hide the `R` code as much as possible, and the output of the `R` code that are shown in the `.html` file should be relevant to the main text of the report.
- An separate `responsibility.txt` file that describes the roles each team member played in the project.

(3) The maximum number of pages (not including the cover page) allowed is 25, excluding appendices.

(4) The report requires a title, a content list and should include at least the following

- An introduction to the data set and the problems that you want to solve
- Data description: Give a detailed description about the data including the number of variables, the variable types etc.
- Descriptive Statistics: Use both graphical and numerical methods to present the data / variables. Plots and tables are welcome. Please provide the interpretation for each plot / table.
- Discussion: Summarize the online article and discuss its relevance with the analysis of the data.
- Conclusions: Give your conclusion based on both descriptive statistics and inferential statistics.