

## Homework # 2

### Data Summarization, MCT&D, Visualization II

This assignment covers the contents of the *Data Summarization, Measures of Central Tendency and Dispersion (MCT&D)*, *Data visualization II* lab (February 23/24, POLI 001/002 respectively). You will have to work with the `qog_std_cs_jan23` data set throughout, you can locate it in Canvas Module: "The Quality of Government (QoG) data sets".

Submit an R script with code that produces the tasks requested in this assignment. I will evaluate your submission by running your code. For written/explanatory answers you can either add comments to your code (I think this will make both of our lives easier), or submit a pdf/word/etc document with anything you want/need to explain. Please be as concise as possible in your written answers, if any. As long as you answer what you are supposed to, I will not grade based on the length of your explanation.

You can work in groups, consult class materials and code found online; but please write your own code and submit one assignment per student.

**Due:** End of day on Monday, February 27, 2023

**How:** Canvas assignment submission link

**File:** hw02-Lastname.R

**Max grade:** 10 points

#### 1. Describing your data:

- (a) [1 pt] Import the `qog_std_cs_jan23.csv` file to your environment. Create a new data set containing only the `cname`, `undp_hdi` and `gol_inst` variables.
- (b) [1 pt] Use appropriate commands to take a look and summarize all your data at once. Tell me something about the entire data set, how many variables and observations, type of variables, some of their statistics, etc.

#### 2. Calculating MCT&D:

- (a) [1 pt] Calculate the mean, median, and mode of the `undp_hdi` variable.
- (b) [1 pt] Calculate the standard deviation, IQR, and skewness of the `undp_hdi` variable.

#### 3. Visualizing your data:

- (a) [2 pts] Create a histogram of the `undp_hdi` variable, add lines for the mean and median values of the distribution of this variable. Explain how the skewness value you got before relates to the positions of the mean and median you observe.
- (b) [2 pts] Create a box plot of the `undp_hdi` variable, add lines for the different relevant elements of the box plot. Describe what statistics each of those are.
- (c) [2 pts] Create a plot that shows the average Human Development Index (`undp_hdi`) by regime type (`gol_inst`). Recode values of the `gol_inst` variable properly so that the plot can be read. Check the QoG data codebook for information on what each category represents. What regime types have higher and lower levels of human development? Do NA cases have something in common?