Homework # 3
Hypothesis Testing

This assignment covers the contents of the *Measures of Association I* lab (March 2/3, POLI 001/002 respectively). You will use the `QoG` data sets we have been using so far to select variables of your choice for this assignment, you are welcome to use the democracy and COVID examples we used in class when considering two samples, but you must use a variable different than the one we used in the lab (`wdi_wip` = % *women legislators*) when testing your hypotheses. The `QoG` codebook can be found on-line and on Canvas.

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 just add comments to your code in the corresponding section. 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, March 6, 2023
> **How**: Canvas assignment submission link
> **Files**: hw03-Lastname.R
> **Max grade**: 10 points

1. [4 pts] **Your data**
   Using the `QoG` data sets select and/or construct (and let me know in a phrase what those are):
   (a) One substantive variable of interest you want to explore and will formulate hypotheses about (i.e., something like the `wdi_wip` we used in class);
   (b) A grouping variable to later divide samples (e.g., the `br_dem` democracy var we used, other binary/dummy indicators you can find by searching for the words 'dummy' or 'binary' in the document, or another dummy variable you can construct by re-coding some other non-binary indicator);
   (c) A year you want to use to split before and after observations like in the COVID example we saw in class. You can use the same COVID example and code seen in class, you can also choose to focus in other world-wide/country based event of interest.

2. [3 pts] **One-sample hypothesis testing**
   (a) Formulate the null hypothesis you want to test about the mean of your sample/variable. What's the `mu` value you'll test on?
   (b) Perform a one-sample z-test or t-test that is adequate for your hypothesis. Choose the statistic that is appropriate for your data (justify). Do you reject the null or not? Why? What do you conclude then about the substantive hypothesis you formulated?

3. [3 pts] **Two-sample t-test hypothesis testing**
   (a) Divide your two samples using the grouping variable you selected in (1b) and conduct an unpaired/independent t-test. What do you conclude? Are the sample means equal/different? Why?
   (b) Divide your before and after samples using the time/year you selected in (1c) and conduct a paired/dependent t-test. What do you conclude? Are the sample means equal/different? Why?