# Assignment 1

## Ph.D. Applied Microeconometrics
## KDI School Fall 2024

**Due date: Thursday, September 23rd before class**

The goal of this assignment is just to get you coding in R (well, RStudio) and R Markdown. You will use data from the Penn World Tables (information about the latest release here).[1] I have uploaded an Excel file – called pwt1001 – to the assignment 1 folder with three tabs:

- Info: You can basically ignore this for our purposes.
- Legend: The variable names in the data are not very informative. This tab tells you what they mean.
- Data: This is is the actual data.

You can load Excel data into R with the `readxl` package, specifically with the function `readexcel`. Here is a handy cheatsheet on importing different types of data into R. Note that you will need to specify which tab in the Excel file you want to load into R!

I would like you to do the following:

- Create a figure with log expenditure-side GDP per capita (per person) on the y-axis and year on the x-axis. How you do this is up to you, but I'd suggest something along the lines of `geom_smooth` and `ggplot2` (you can find a handy cheatsheet for ggplot2 here).
- Create a new variable called "decade" that is equal to the decade of the year variable. For example, 1960 would be 1960-1969, 1970 would be 1970-1979, etc. You can do this with the `cut` function, but there are many other ways to do it, too. Please do 1970 through 2019 (i.e. 1970-1979, 1980-1989, etc.) only. You can drop the other years.

    - Create a table with mean expenditure-side GDP p.c., population, average hours worked, and *one variable of your choice* by decade.

Along with the figure and table, I'd like you to include a short write-up describing the results. All of this should be contained in a single document (i.e. there should be a single document with a figure, a table, and a short write-up describing the results).

---

[1]You'll likely want to use this website to figure out what all the variables are!

For full credit, you must turn in the following:

- R Markdown script
- pdf output of the R Markdown script
- R script if you did not directly do everything in the R Markdown script (I'll leave this choice up to you)

All other decisions are up to you. For example, not all countries have data for all years. Do you want to include any available data or drop countries that are missing data? It's up to you. **Please explain in the write-up what you did and why.** I'll be grading based on whether you can do what I've asked and whether you can explain what you did and why.