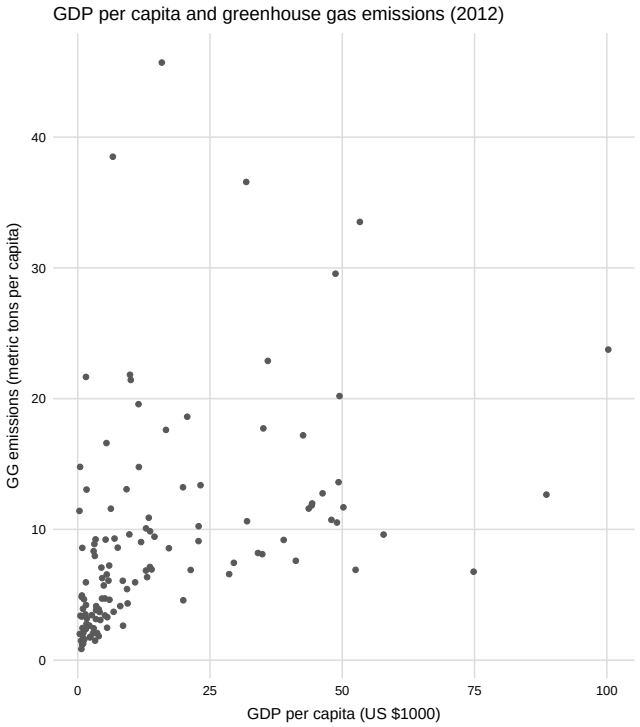
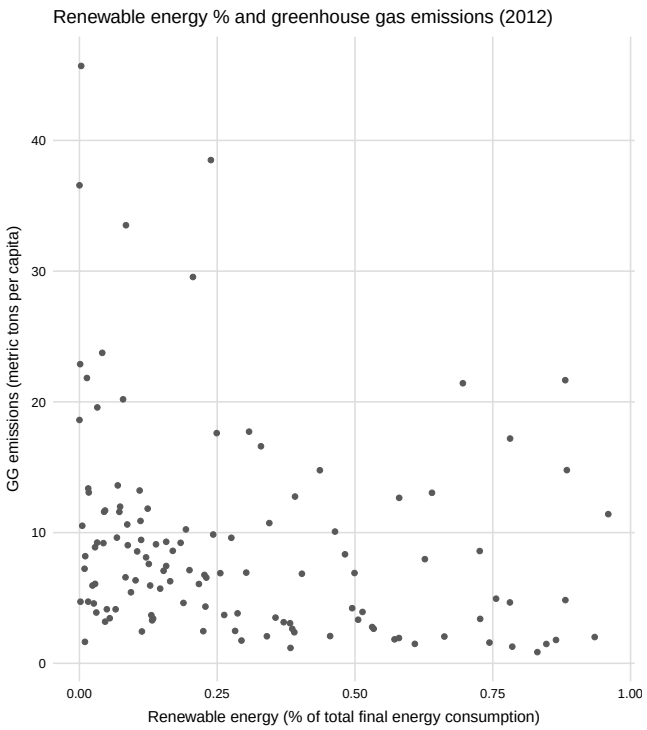
EEP/IAS 118 - Introductory Applied Econometrics
Problem Set 1
Due Monday, July 1 at the beginning of class

Submit materials (all handwritten/typed answers, Excel workbooks, and R reports) on Gradescope

Measuring greenhouse gas emissions and understanding how economic activity affects these levels can be a difficult task. For example, greenhouse gases vary in the amount of time they remain in the atmosphere and how much energy they absorb. To address this issue, total greenhouse gas emissions are often measured in units of Global Warming Potential (GWP). GWP is also known as CO_2 equivalents; CO_2 has a baseline GWP of 1 and all other values convey how much more/less global warming potential the gas has relative to CO_2. For example, while methane (CH_4) has a shorter lifespan than CO_2 it absorbs much more energy, so it is considerably more harmful and is given a GWP of 28-36. In this problem set we will explore how two measures of economic activity, GDP per capita and renewable energy consumption, are correlated with greenhouse gas emissions.

The data comes from the World Bank's Development Indicators for 2012, which are provided for you. Here we focus on three variables. First is total greenhouse gas emissions, which include CO_2, all anthropogenic sources of methane (CH_4), nitrous oxide (N_2O) and F-gases. The total emissions per capita in a given country are calculated as metric tons of CO_2 equivalents divided by the midyear population. Next is GDP per capita, calculated as gross domestic product divided by midyear population and measured in **thousands** of US dollars. Lastly, renewable energy consumption is the share of renewables in total final energy consumption. The figures below plot data from all 123 countries contained in the original dataset (the provided dataset is a subset of this original data). The values for selected countries can be found in Tables 1, 2, and 3.





Exercise 1 (Excel). Relationship between GDP per capita and greenhouse gas emissions per capita

This exercise is to be completed using `Excel`. Looking at the first graph, there appears to be an association between GDP per capita and greenhouse gas emissions per capita. We will establish a simple linear relationship on a (very small) subset of 5 countries.

Table 1: GDP per capita and greenhouse gas emissions

CountryName	GDPCapita	GGTonnesCapita
Brazil	11.56	14.77
Cambodia	0.87	8.59
Canada	48.72	29.55
China	5.28	9.22
United States	49.48	20.20

(a) Use Excel to create a scatter plot of these observations. Don't forget to (1) label the axes and their units, and (2) title your graph. **You should use the tables provided here for these calculations, not the actual observations from the .csv data file.**

(b) This question has **two parts**. First: Estimate the linear relationship between GDP per capita and greenhouse gas emissions per capita (E) by OLS, showing all intermediate calculations as we saw in the lecture 3 slides (use Excel to create the table and show all the steps).
Second: interpret the value of the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\widehat{E}_i = \hat{\beta}_0 + \hat{\beta}_1 GDP/cap_i + \epsilon_i \qquad i = \{\text{Brazil, Cambodia, Canada, China, US}\}$$

(c) In your table, compute the fitted value and the residual for each observation, and verify that the residuals (approximately) sum to 0.

(d) According to the estimated relation, what is the predicted $\widehat{E}$ for a country with a GDP per capita of $10,000? (Pay attention to units)

(e) How much of the variation in greenhouse gas emissions per capita for these 5 countries is explained by their GDP per capita? Can you provide some reasons for why emissions might be lower in less developed countries?

(f) Repeat exercise (b) for two additional sets of countries below. **You should use the tables provided below for these calculations, not the actual observations from the .csv data file.**

Table 2: GDP per capita and greenhouse gas emissions

CountryName	GDPCapita	GGTonnesCapita
Chile	13.96	6.94
Georgia	3.47	3.82
Indonesia	3.43	3.15
Malta	19.95	4.58
Philippines	2.30	1.74

Table 3: GDP per capita and greenhouse gas emissions

CountryName	GDPCapita	GGTonnesCapita
Belarus	6.26	11.59
Cambodia	0.87	8.59
Lebanon	8.05	4.14
Mozambique	0.45	14.78
Namibia	5.44	16.60

(g) How do your estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ change between Tables 1, 2, and 3? Briefly explain this variation in 3-5 sentences.

(h) Save a copy of your workbook as a pdf (OLS tables and scatter plot) to combine with the later work.

Exercise 2 (Written). Functional Forms

(a) An economist asks you whether they should estimate a linear relationship between greenhouse gas emissions per capita and % of renewable energy in total final energy. You will make your decision based on the figure on page 2. If you answer "yes", please say why. If you answer "no", can you name another functional form you might try?

(b) The economist sees your analysis of the relationship between GDP per capita and greenhouse gas emissions per capita. She asks you to explore alternative functional forms for this relationship. Using the provided World Development Indicators data, you estimate the following equations:

$$\text{A linear relationship: } \hat{E}_i = 6.464 + 0.176\, GDP/cap_i + \epsilon_i$$
$$\text{A linear-log relationship: } \hat{E}_i = 4.046 + 2.655\, log(GDP/cap)_i + \epsilon_i$$
$$\text{A log-log relationship: } log(\hat{E})_i = 1.191 + 0.364\, log(GDP/cap)_i + \epsilon_i$$

Note that it is convention to always use the natural log.

(i) Interpret the parameter on GDP per capita (or log(GDP per capita)) in each of these equations.

(ii) What are predicted greenhouse gas emissions (metric tons per capita) for a country with a GDP per capita of $10,000 with each of the models?

Exercise 3 (R). Importing data into `R`

For the purposes of this class, we will be using a cloud-based version of RStudio provided through UC Berkeley's *Datahub*.[1] The data files can be accessed directly through *Datahub* and do not require you to install anything on your computer. This exercise is designed to get you familiar with accessing the service, loading in data files, and obtaining summary statistics.

(a) Since you're reading this document, you've successfully accessed RStudio on *Datahub*. You'll access RStudio and the data files the exact same way with the same link for future problem sets.

(b) Open a new R Script – you can do this by clicking **File** > **New File** > **R Script**, by using the button just below **File** (looks like a sheet of paper with a plus symbol), or by pressing (Control + Shift + N) on Windows / (Command + Shift + N) on Mac. Write all of your code/comments in this script.

(c) In the file viewer window (bottom right), you should see a folder called "ENVECON-118." This contains the "Summer-19" folder where all our problem set subfolders will be (you can click on the folders in the viewer window and browse through the structures by hand. Click the ".." by the green arrow to return to the previous directory). Set the working directory to the "ENVECON-118/Summer-19/ProblemSet1" folder using the `setwd()` command.

(d) Save the R script by using **File** > **Save**. Since we changed the working directory, the script will be saved in the "ProblemSet1" folder.

(e) Load the data set *World_Bank_Indicators_PS1.csv* into `R` (since this is a ".csv" file, you should use the `read.csv()` command) as *indicators*.

(f) Provide basic summary statistics on the greenhouse gas emissions per capita variable (*GGTonnesCapita*) in the *indicators* dataframe. Use the `summary()` command. This command is part of base `R`, so you do not need to load any packages before using it. What is the median value of greenhouse gas emissions per capita?

(g) Next, generate custom summary statistics on *GGTonnesCapita* using the `summarize()` command provided by `dplyr`. You will need to call the `tidyverse` package to use it (`tidyverse` is a collection of packages designed for data science. It includes `dplyr` and several other packages we'll use this term). What are a) the number of observations, and b) the value of the 95th percentile of greenhouse gas emissions per capita in the sample? (For help generating the statistics, type `?summarize` in the console and hit enter, then select the `dplyr` function)

(h) Compile the report of your `R` script and save a pdf to your computer (note: this can be done by going to **File** > **Compile Report** in the menu).

[1] If you already have an installation of R/RStudio on your personal computer and prefer to work offline, you can download the data for this assignment from bCourses. Make sure to install/update all packages mentioned in the problem sets in order to prevent issues regarding deprecated or outdated packages.