<u>Instructions</u>

Open "Documentation/ddi-documentation-english-29.pdf". Read the abstract in the overview section on page 2 to understand what the Economic Census data is about.

Please answer the following by writing code exclusively in either python or R for each question. Make sure to write clean, clear code and comment properly.

Include a separate word document for answers to some of the questions below – name the document "RA_Task_FirstName_LastName.docx". Include any explanations for critical decisions made in the code in this document. If you must load any extra libraries to process the data, mention these in the document as well.

This whole exercise should take approximately 1-2 hours to complete.

<u>Questions</u>:

1. Read ec05st02.txt file based on the data dictionary in Documentation/Record layout.doc. Make sure to read correctly based on rural/urban status. Let's call this the EC datafile.
   a. Were there any rows that you could not read correctly?
   b. Code non-numeric variables correctly. For example, factor variables are not numeric variables.
2. We will now add a geographic identifier for a village/town (or groups of villages/towns) that can be used to include Economic Census data from other years on the same geographical unit. This identifier is called "shrid" and is available from the Development Data Lab[1]. You do not need to download these datafiles as they have been placed in the folder already.
   a. In the EC datafile you read in part 1, which group of variables identify a village? Which variables identify a town?
   b. Use these variables to merge the EC data file with shrug_ec05_rural and shrug_ec05_urban data files.

---

[1] https://www.devdatalab.org/ - this is also a great website to explore to understand what economists do with data science