## General Instructions:

Setup:

1 x github repository to host the code that you are writing.

An [Anaconda](https://conda.io/docs/user-guide/install/download.html) installation on your machine.

A [pandas tutorial](https://www.youtube.com/watch?v=9d5-Ti6onew).

Remember that this is your exhibition, so feel free to go wild with any fancy visualizations that you want to try or data structures you want to show off.

## Task 1:

<https://www.kaggle.com/ophi/mpi>

This task is formulated to judge your comfort level with fetching, manipulating and cleaning data, specifically to answer statistical questions. In a broader context, this serves as a convenient introduction to the subject matter that EcoEnergy deals with on a daily basis - consumption under scarcity constraints.

The dataset at the kaggle link above has been released by the Oxford Poverty and Human Development Initiative. It consists of a rural/urban breakdown of deprivation from which a “multidimensional” poverty indicator is  [developed by the OPHI](http://ophi.org.uk/research/multidimensional-poverty/).

Your task is to load the dataset in the environment of your choice (I suggest pandas. ), clean it and develop procedures to answer the following high-level questions and express them as graphs:

* What is the interquartile range for MPI in this dataset and which countries fall within it?
* What is the correlation between the Intensity of Deprivation Rural and the Continent the country is on?
* What is the correlation between the Intensity of Deprivation Urban and the Population of a country?
* What is the correlation between the Intensity of Deprivation Urban and the Intensity of Deprivation Rural?
* Which countries exhibit the largest subnational disparities in MPI?
* Which countries have high per-capita incomes yet still rank highly in MPI?

## Task 2:

Using the dataset you have prepared in task one, train a machine learning model which takes **as input the continent that a country is in, its population, the intensity of deprivation rural and the intensity of deprivation urban**, and produces as output its **MPI.**

Do an analysis of the performance of this model through visualizations of the residual errors. Summarize the approach that you took to select the model architecture, and detail your loss and evaluation functions selection process.