Deliverable #1 - Proposal

Ava Oberle & Hieu Tran

Central Question: Do countries with better environmental quality generally have stronger economies?

Data sets:

1. List of countries by GDP (nominal) per capita: https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita
2. List of countries by ecological footprint: https://en.wikipedia.org/wiki/List_of_countries_by_ecological_footprint

Description:

- List of countries by GDP (nominal) per capita:
    - This table contains latest estimates on GDP per capita from 3 different sources (IMF, World Bank, United Nations)
    - Column names: Country/Territory (string), UN Region (string), IMF, World Bank, United Nations (each of the 3 source has subcolumns for their estimate (integer) and year of data (integer))
    - There are 223 different rows in this dataset, containing countries (ranked) and territories (unranked). Not all countries and territories are covered by every source.
- List of countries by ecological footprint:
    - This dataset describes the ecological footprint for each country
    - Column names: Rank (integer), Country/Region (string), Ecological Footprint (float), biocapacity (float), biocapacity deficit or reserve (float), population (float), total biocapacity deficit or reserve (float), population for biocapacity to equal ecological footprint (float)
    - There are 188 different countries (rows) represented in this dataset

Outline:

1. We will scrape both websites (using .xpath)
    a. We will first find the node with the tag “table” (with Wikipedia’s sortable function)
    b. Will extract the column names and the data
    c. We will then use pd.DataFrame to create a df
2. We will then use the table to answer the central question
    a. We will use only the ecological footprint and IMF estimate for GDP
    b. We might create a scatter plot to see if there’s a correlation between the two