7PAM2000 Applied Data Science 2

Assignment 2

This second assignment will focus on exploring statistics and trends in more detail. As before, you are expected to produce a 2 page report conforming to the same guidelines as in Assessment 1.

This time you will be exploring public data from the World Bank, and specifically country-by-country indicators related to climate change: `https://data.worldbank.org/topic/climate-change`. There are a range of indicators relevant to climate change, for example access to electricity, agricultural activity, urban population, etc.

Your goal is to:

- Ingest and manipulate the data using Pandas dataframes. Your program should include a function which takes a filename as argument, reads a dataframe in Worldbank format and returns two dataframes: one with years as columns and one with countries as columns.
- Explore the statistical properties of a few indicators, that are of interest to you, and cross-compare between individual countries and the whole world (you don't have to do all the countries, just a few will do) and produce appropriate summary statistics.
- Explore and understand any correlations (or lack of) between indicators (e.g. population growth and energy consumption). Does this vary between country, have any correlations or trends changed with time?
- Again, you are expected to use your initiative and "tell a story" with the data. You should use appropriate visualisation (hint: time series could be useful) and provide a text narrative to communicate and explain your findings.
- You will be assessed on the overall quality of the report, good use of visualisation tools and good use of the methods and tools available for dataframes. See mark scheme for details.

An overall coding quality mark will be awarded combined from your three submissions. Part marks for this assignment and advice for improvement will be part of your feedback. Criteria

- Adherence to the PEP-8 guidelines.
- Well structured and commented program, good use of functions. No spaghetti code please.
- Good use of your repository with an appropriate level of commitments.

  You can add data from other sources, but you should analyse at least one of the Worldbank datasets.

This assignment does intentionally not specify which data sets to choose. Some ideas, definitely not exhaustive. You may find more interesting combinations.

- CO_2 production vs. GDP (energy efficiency)
- Arable land vs. land covered by forests (deforestation)

- Electric power consumption, access to electricity, overall energy use and CO2
- agricultural and non-agricultural methane production. How does it link to other parameters like poverty headcount, energy consumption, access to electricity
- How does this look for countries in different phases of developement? Countries in different parts of the world?