COM6018 Assignment 1

Exploring Global Carbon Emissions Data

Due: 15:00, Friday 8th November 2024

26th October 2024.

Table of Contents

1. Introduction

The assignment provides you with a dataset that describes the carbon dioxide emissions from fossil fuels and cement production for countries around the world. It then asks you to use this data to answer a series of questions. For each question you will need to produce an illustrative figure and write a few sentences to answer the question. You will need to think carefully about how to use the data and how to design the figure to make your answer clear.

Your assignment report will take the form of a Jupyter notebook. You are provided with a template notebook that contains the questions and some code to get you started. You will need to add code to process the data and produce the figures. You will also need to add text to answer the questions. You will then submit the completed notebook to Blackboard. You will be allowed to use any modules in the Python standard library, plus the following packages: `numpy`, `pandas`, `matplotlib`, `seaborn`. No other packages are allowed.

Further details of the data, the questions, the assessment criteria and the submission process are given below.

2. The Data

The dataset for this assignment is the CO_2 and Greenhouse Gas Emissions dataset from Our World in Data. The data is available from the following GitHub repository, https://github.com/owid/co2-data. You will need to download the file `owid-co2-data.csv` from the GitHub repository linked above. Store this file in the same directory where you store your notebook.

The dataset contains records that describe the CO_2 emissions for a country (or region) in a given year. The dataset contains a very large number of fields (80 in total). There is another csv file in the repository, `owid-co2-codebook.csv`, that describes the fields in the dataset. You should read this file carefully to make sure you understand the meaning of the fields. Understanding these fields correctly will be important for answering the questions in this assignment. You can get a further understanding of the data by reading the accompanying article on the Our World in Data website, https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.

3. The Tasks

The questions are presented in the template notebook but are also listed below for convenience. For each question, select the plot type that best illustrates the point you are making, e.g., a line graph, a scatter plot, a pie chart,a box plot, a violin plot, etc.

Question 1

How have CO_2 emissions grown over time?

Consider the six continents: Africa, Asia, Europe, North America, Oceania and South America. Make a figure to show: i) how the total annual CO_2 emissions of each continent varied over the years 1950 to 2021. ii) How has the per capita CO_2 emissions (i.e., the emissions per person) of these continents varied over the same period.

Comment on how the contribution of each continent is different when measured in terms of total emissions and per capita emissions.

Question 2

Which countries have the highest CO_2 emissions?

Consider the five countries with the highest total CO_2 emissions in 2020. Make separate plots for the years 1960, 1990 and 2020 that show the share of the total world emissions that each of these countries contributed in these years.

Consider the five European countries with the highest total CO_2 emissions in 2020. Make separate plots for the years 1960, 1990 and 2020 that show the share of the total European emissions that each of these countries contributed in these years.

Do the highest global producers produce a larger or smaller proportion of the total worlds emissions in 2020 compared to 1960? Do the highest European producers produce a larger or smaller proportion of the total European emissions in 2020 compared to 1960?

Question 3

Do countries with high GDP per capita always have high per capita CO_2 emissions?

Plot the per capita CO_2 emissions against the GDP per capita for each country, adjusting the area of each marker to represent the population of the country. Only include countries with a population of at leat 10 million people in 2018. Annotate significant countries on the plot, such as large countries or outliers.

Compare these plots for the year 1978, 1998 and 2018. Comment on any patterns in the relationship between GDP per capita and CO_2 emissions per capita and how they may have changed over time.

Question 4   GDP per capital

How has wealth and wealth inequality developed over time?

Make a plot that shows the distribution of GDP per capita across countries at each decade from 1950 to 2020 (i.e., at the years 1950, 1960, 1970, etc.). Only include countries with a population of at least 5 million in 2020.

The World Bank defines an income of less than $2.15 per day as extreme poverty, corresponding to an annual GDP per capita of $785. Mark this threshold on your plots. For each year shown, calculate the number of people living in countries with a GDP per capita below this level and annotate this in your figure. Comment on whether this population has grown or shrunk as a proportion of the world's total population over time.

4. Assessment

The assignment is worth 40% of the module mark and will be marked out of 40.

For each of the 4 questions you will need to write some code to produce an appropriate figure. You will then need to write a short paragraph that describes the key points of the figure and answers the question. Each question is worth 10 marks.

The marks for each question will be awarded after considering the following criteria:

- **code**: is the code concise, clear and well documented? Does it make good use of the available libraries?
- **plots**: is the plot well-designed? Is it well labelled? Is it easy to interpret? Does it display the data in a way that addresses the question?
- **text**: Does your text draw relevant conclusions from the plot? Does it answer the question in a clear and concise way?

5. Marks

The question mark will reflect the overall quality of the answer after considering the three aspects above. Marks can be interpreted using a scale consistent with the degree's pass, merit and distinction grade boundaries.

- 0 - No attempt submitted.
- 1-4 - Unsatisfactory, an attempt has been made but it is incomplete or has serious deficiencies in one or more of the criteria.
- 5 - Satisfactory, a minimum passing mark
- 6 - Good, work consistent with a Merit level grade.
- 7 - Very good, work consistent with a Distinction level grade.
- 8-9 - Excellent work, at least one criterion is met to the highest standard.
- 10 - Exceptional work, all criteria are met to the highest standard.

6. Submission

Please name your notebook using the following convention:

```
COM6018-assignment1-<your-student-username>.ipynb
```

For example, if your username is 'ac1jpb' then your notebook should be named `COM6018-assignment1-ac1jpb.ipynb`

You must submit your completed notebook to the assignment 1 submission area on Blackboard. You may submit multiple versions up to the deadline. Only the latest submission will be marked.

When marking your work, we will be **running your notebook** to check the code runs correctly. The notebook will be run in a directory that contains a copy of the `owid-co2-data.csv` datafile. So, make sure that your data file is stored in the same directory as the notebook when you test it. If your notebook does not run correctly then you will be in danger of losing marks.

The assignment is due by 15:00 on Friday, 8th November. Standard lateness penalties will be applied.

⚠️ Academic Misconduct

The University takes academic misconduct very seriously. Academic misconduct includes, but is not limited to, plagiarism, collusion and fabrication of data. In particular, this is an individual assignment and any suspected collusion will be investigated. Do not share your code with anyone else.

If you are unsure about the University's policy on academic misconduct, please see https://www.sheffield.ac.uk/new-students/unfair-means.