7PAM2000 Applied Data Science 1
Assignment 1: Visualisation

Your task is to apply three types of visualisation methods (graphs) to extract meaningful information. Find below a link list for good sources of data, but you are welcome to find other sources. Please check the quality. Some data sets can be quite messy and are better avoided in your first assignment.

1. Produce a line plot showing multiple lines with proper labels and legend. Describe what conclusions you can draw from this plot.

2. Produce graphs using two other visualisation methods. Explain why you picked this type of graph and describe what conclusions you can draw. Marks will be awarded for plausible choices of graphs. The choice does not need to be the perfect one, but at least one type will not a good choice if you visualise exactly the same data three times. Note, that we consider it not exactly the same, if you do a line plot of a time series and then use a pie chart to visualise relative sizes for a selected time.

A reminder: What type of graph for which task?

Line plots Good for continuous data, e.g., a time series. Not a good choice for categorical data, such as number of people being active in which sport. Bar charts are much better for this task.

Scatter plots. Good for comparing two data sets and showing correlations (often combined with line plots).

Histograms. Good for showing distributions, e.g., income distribution of the population.

Box plots. A compact form of visualising distributions. Good if several distributions should be visualised.

Pie charts. Good for comparing sizes (in absolute numbers). Good when absolute data is used, e.g., the size of countries in km^2. Not good for comparing relative numbers, e.g., fraction of urban population or GDP/head. Bar charts are far better for that.

Bar charts. Good for comparing values – absolute and fractional.

The graphs should be produced using `pyplot` functions. Data should be read using `pandas`. The code should make good use of functions. The minimum expected would be one function for each visualisation method. Best, if you write the functions in a generalised way so that they can be used more than once. Do not forget to put a short explanatory docstring at the beginning of the function (see bootcamp or `https://docs.python.org/3.10/tutorial/controlflow.html#defining-functions`).
Marks will be awarded for

- Appropriate use of visualisation techniques and quality of figures (e.g. legibility, proper labelling).

- The quality of your explanations and descriptions.

- Coding quality. Follow the guidelines in the good style guide.

- Use of your repository, best with repeated commits. (How to use a repository is explained in a lecture.)

- Have a close look at the mark scheme. It will help you to avoid losing easy marks.

What to submit?

- A word or PDF document containing the graphs and your explanations. Include your name and (a) link(s) to your data source(es) at the start.

- Your program(s) and a link to your github repository. We want the code uploaded that we can proceed with marking in case of problems with the link. Test the link and make your repository public.

- **Do not** upload the data, links suffice.

- The system will not accept .zip and .rar files.

How to optimise your marks?

- Use spyder or a similar programming environment, not notebook.

- Use pyplot. No marks will be awarded for graphs produced using other packages.

- Follow the instructions in the good coding style guide. Use `autopep8` to make your code PEP-8 conform.

- Check the rubrics and the mark spreadsheet. Are you ticking all the boxes?

- Avoid academic misconduct. See the instructions in the

- Academic misconduct and how to avoid it unit.

0.0.1 Checking your submission

We carry out and automatic check of your submission. On ADS1 we will make available an upload tab which carries out the check for you to inspect.

-

Useful data sources

Government: `https://data.gov.uk`
MET Office: `https://www.metoffice.gov.uk/research/climate/maps-and-data/data/index`
World Bank Open Data: `https://data.worldbank.org`
WHO Global Health Observatory: `https://www.who.int/data/gho`
Google Public Data Explorer: `https://www.google.com/publicdata/directory`
FiveThirtyEight: `https://data.fivethirtyeight.com`