1 DATA TYPES

Draw a mind map of the data types defined in the lecture and in the chapter by Andy Kirk [AK].

Give at least two examples for each data type.

[AK] *Section on Data Types from Andy Kirk, Data Visualization (see Moodle)*

2 WHY VISUALIZE - DATASAURUS

Anscombe's quartet is only the beginning. Choose a way of solving this exercise:

Hard mode:

1. Load the datasaurus dozen datasets (see Moodle, DatasaurusDozen.tsv) using the pandas library.
2. Calculate the mean, standard deviation and variance for every dataset
3. Visualize the individual datasets with a scatterplot using matplotlib.

Guides to get you started:
Python + numpy + Matplotlib:
(Highly recommended) https://cs231n.github.io/python-numpy-tutorial/
Pandas:
https://www.w3schools.com/python/pandas/pandas_intro.asp
https://pandas.pydata.org/docs/user_guide/10min.html
Matplotlib:
https://www.w3schools.com/python/matplotlib_intro.asp
https://www.w3schools.com/python/matplotlib_subplot.asp

Easy mode:

Follow the steps outlined in [DSSphere22] up to the visualization.

Tools mode:

Use Excel (LibreOffice Calc, Numbers, Google Sheets, etc.) to visualize the datasaurus dataset

Read the paper [MF17] on the generation of the datasaurus and briefly summarize how it works.

[DSSphere22] https://datasciencesphere.com/visualization/datasaurus-dozen-visualization-using-python/
[MF17] - Justin Matejka and George Fitzmaurice. 2017. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). Association for Computing Machinery, New York, NY, USA, 1290–1294. (see Moodle for pdf)
https://www.youtube.com/watch?v=It4UA75z_KQ

3 WHY VISUALIZE – REAL DATA EXAMPLE

The effect of social benefits on unemployment levels has been an ongoing debate for a long time (effect of stimulus packages and benefits on the 'great resignation' in 2021/2022 [GR22]).

In [BK79] Benjamin and Kochin analyze to what extent generous unemployment benefits were responsible for the high level of unemployment in inter-war Britain. Their result is that the persistently high level of unemployment *'was due in large part …to high unemployment benefits relative to wages'.*

1. Read the provided description of the dataset, each variable, explain its meaning in clear terms and determine the data type
2. Load the dataset using pandas
3. Experiment which type of chart is suitable for this type of data.
4. Visualize the temporal changes of unemployment during inter-war Britain.
5. Create a visualization that supports or disproves Benjamin and Kochin's hypothesis.
6. Using the conceptual language of the lecture, Actions and Targets. Which concept does your visualization follow?
7. Can you find an alternative or better explanation using visualizations?
8. Summarize your findings. Optional: Consider the historical context of [BK79].
9. Optional: Consider the strong deviation in year 1920. How can this deviation be explained? Visualize the dataset without the values of 1920.

It is recommended that you use python and a library like pandas to load the dataset and solve this exercise, but for now you may also use Excel or similar tools if you prefer.

Guides to get you started:
Python + numpy + Matplotlib:
(Highly recommended) https://cs231n.github.io/python-numpy-tutorial/
Pandas:
https://www.w3schools.com/python/pandas/pandas_intro.asp
https://pandas.pydata.org/docs/user_guide/10min.html
Matplotlib:
https://www.w3schools.com/python/matplotlib_intro.asp
https://www.w3schools.com/python/matplotlib_subplot.asp

[BK79] Daniel K. Benjamin and Levis A. Kochin, 'Searching for an explanation for unemployment in interwar Britain', Journal of Political Economy, 87, 1979, pp.441–78.

[GR22] https://www.pewresearch.org/fact-tank/2022/03/09/majority-of-workers-who-quit-a-job-in-2021-cite-low-pay-no-opportunities-for-advancement-feeling-disrespected/