Coursework

Objective: *implementing a visualization with R*

- Pick a dataset of your interest
- Pose the initial questions (3 to 5) that you would like to answer
- Assess the fitness of the data
- Answer the questions by visualizing the dataset using R in an exploratory fashion
- Further refine/propose questions and produce the visualization for those refined/proposed (more exploratory) questions (<= 10 questions in total).
- It is a bonus if you can make your visualization interactive
- You can also try other visualization tools for the ultimate visualization if you want (optional, e.g., to make it more interactive). However, using R for the initial exploratory analysis is required. You should work closely with the "R Graphics Cookbook".

Written report:

Due date: **April 21st 2023 3:00 PM**
maximum 10 pages (3000 words). In addition to the 10 pages main content, you can also add additional appendix (no limitations) if you want.

- Description of your data
- The description with the initial questions
- For each question, a description of your visualization strategies, including data cleaning, transformation, visual encoding, etc.
- An explanation of the exploratory process of generating new questions and visualizations.
- Critical discussion of your visualization design (e.g. why you pick these encodings or this visualization)
- A reflection on the development process
- You need to upload your R codes as well

A Case Study: House Price Visualization

- *Pick a dataset of your interest*: The land registry Price Paid Data includes information on all property sales in England and Wales that are sold for full market value and are lodged with us for
  registration. Format: tid,price,date,postcode,type,age,tenure,paon,saon,street,locality,city,district,county,ppdcategory,recordstatus
- *Pose the initial questions (3 to 5) that you would like to answer*: **RQ1**: When do property sales generally happen? Which month is the best month to sell? **RQ2**: How do property sales change according to different property types and age? **RQ3**: How do property sale price compare across different counties?
- *Assess the fitness of the data*: Can we answer all those questions by using only the land registry data? For RQ3, although we have area data, we need the county data, in order to compare across different counties, i.e. County-postcode mapping data.
- *Answer the initial questions by visualizing the dataset using R*: Load data; Manipulate data; Visualization
- *Further refine/propose questions*: Example question (RQ4: Which types of properties and in which county suffer the most from the economical crisis?) Need to think about: How to manipulate/process data (e.g. need to calculate year (month) over year (month) change)? What visualization techniques to use? What if there are too many variables?

- The demo R codes are available. The sample land registry data and county-postcode mapping data are also available. You need to change the directory path in order to make it work.