Cleaning and Wrangling data

- *All materials for the exercises below are available in the homework folder.*
- *Please submit an R script file containing the code and results.*
- *You can #commentout any sentences written answers*

*

Courses at the Graduate Institute

The Graduate Institute offers courses in spring and autumn. The datasets autumn_21.csv and spring_22.csv contain information on all courses offered by the Graduate Institute in the academic year 2021-2022.

Question 1

Your first task is to consolidate both datasets in a single dataset: academic_year. This dataset should contain the following variables:

Name	Description
`title_course`	Title of the course
`department`	Department that offers the course (MINT, EI, RISP, HPI, DI…)
`language`	Language in which the course is instructed (french or english)
`ECTS`	How many ECTS you can get for the course.
`semester`	Takes the categories autumn or spring.

Question 2

What is the share of courses offered in french during the full academic year?

Question 3

Rank the departments by number of courses offered in each semester.

Question 4

Which department offers a higher share of courses in spring? Importantly, different departments offer a different number of courses (BONUS: if you want to go next level, consider ECTs as well).

Question 5

The `faculty_n.xlsx` dataset contains the number of faculty per department. In which department faculty teaches more courses on average? Before proceeding, remove observations from DE and IA.

Leaders Assassination

One longstanding debate in the study of international relations concerns the question of whether individual political leaders can make a difference. Some emphasize that leaders with different ideologies and personalities can significantly affect the course of a nation. Others argue that political leaders are severely constrained by historical and institutional forces. Did individuals like Hitler, Mao, Roosevelt, and Churchill make a big difference? The difficulty of empirically testing these arguments stems from the fact that the change of leadership is not random and there are many confounding factors to be adjusted for. Each observation of the CSV data set `leaders.csv` contains information about an assassination attempt. The variables are:

Name	Description
`country`	The name of the country
`year`	Year of assassination
`leadername`	Name of leader who was targeted
`age`	Age of the targeted leader
`politybefore`	Average polity score during the 3 year period prior to the attempt
`polityafter`	Average polity score during the 3 year period after the attempt
`civilwarbefore`	1 if country is in civil war during the 3 year period prior to the attempt, or 0
`civilwarafter`	1 if country is in civil war during the 3 year period after the attempt, or 0
`interwarbefore`	1 if country is in international war during the 3 year period prior to the attempt, or 0
`interwarafter`	1 if country is in international war during the 3 year period after the attempt, or 0
`result`	Result of the assassination attempt, one of 10 categories described below

The `polity` variable represents the so-called *polity score* from the Polity Project. The Polity Project systematically documents and quantifies the regime types of all countries in the world from 1800. The polity score is a 21-point scale ranging from -10 (hereditary monarchy) to 10 (consolidated democracy).

The `result` variable is a 10-category factor variable describing the result of each assassination attempt.

Question 1

How many assassination attempts are recorded in the data? How many countries experience at least one leader assassination attempt? (The `unique` function, which returns a set of unique values from the input vector, may be useful here). What is the average number of such attempts (per year) among these countries?

Question 2

Create a new binary variable named `success` that is equal to 1 if a leader dies from the attack and to 0 if the leader survives. Store this new variable as part of the original data frame. What is the overall success rate of leader assassination? Does the result speak to the validity of the assumption that the success of assassination attempts is randomly determined?

Question 3

Investigate whether the average polity score over 3 years prior to an assassination attempt differs on average between successful and failed attempts. Also, examine whether there is any difference in the age of targeted leaders between successful and failed attempts. Briefly interpret the results in light of the validity of the aforementioned assumption.

Question 4

Repeat the same analysis as in the previous question, but this time using the country's experience of civil and international war. Create a new binary variable in the data frame called `warbefore`. Code the variable such that it is equal to 1 if a country is in either civil or international war during the 3 years prior to an assassination attempt. Provide a brief interpretation of the result.

Bonus Question 5: Proceed only if you wish!

Does successful leader assassination cause democratization? Does successful leader assassination lead countries to war? Answer these questions by analyzing the data. Be sure to state your assumptions and provide a brief interpretation of the results.