Homework: cleaning and wrangling data

October 2022

# 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. |
| type | Type of course(compulsory, elective, or workshop) |
| topic A | broad category that summarizes the topic |

## Question 2

What is the share of courses offered in french during the full academic year? How many courses were offered in autumn? How many courses were offered in spring?

## Question 3

How many departments are there? Rank the departments by number of courses offered in each semester.

## Question 4

Which department offers a higher share of courses in spring? Remember: different departments offer a different quantity of courses.

## 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.

## Question 6

List the three favorite topics overall. List the three favorite topics of each department.

## Question 7

One of the categories of type is workshops. Workshops are normally about skills, but in the dataset, workshops are missing values for topic. Assign the category skills at the topic variable for all workshops.

## Question 8

Are compulsory courses mostly about theory and methods in all departments? Now, create a new variable that that takes the value of 1 if the course is compulsory and about theories or methods. The variable should take the value of 0 if it compulsory and about something else.