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. |
| topic | Takes the categories “about sustainability” or “others” |
| # Question 2 |  |
| hat is the share of c | ourses offered in french during the full academic year? |
| # Question 3 |  |
| ank the departments b | y number of courses offered in each semester. |
| # Question 4 |  |
| hich department has m emember: different de | ore courses related to sustainability in their curriculum? partments offer a different quantity of courses. |
| Leaders Assassinatio | n |
| ne longstanding debat oncerns the question difference. Some em nd personalities can thers argue that poli istorical and institu ao, Roosevelt, and Ch f empirically testing hange of leadership i actors to be adjusted | e in the study of international relations of whether individual political leaders can make phasize that leaders with different ideologies significantly affect the course of a nation. tical leaders are severely constrained by tional forces. Did individuals like Hitler, urchill make a big difference? The difficulty these arguments stems from the fact that the s not random and there are many confounding for. |
| ach observation of th leaders.csv` contains ttempt. The variable | e CSV data set information about an assassination s 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.

## Question 5

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.