Data Mining, Sheet 4

In this exercise sheet we will work with real world data. You are free to use any technology that you deem fit. However, please make your results reproducible.

Exercise 1

Gather statistical data about education from `https://data.un.org/`:

1. Primary education (ISCED 1): contains the numbers of female and/or total primary school students for selected countries and years.
2. Secondary education: contains the percentage enrollment rates of female, male and/or total secondary school students for selected countries and years.
3. Tertiary education (ISCED 5 and 6): contains the enrollment number of female and/or total students in tertiary education for selected countries and years. It also contains the female student enrollment as a percentage of total students.
4. Public expenditure of education as a percentage of GNI (Gross National Income).
5. Public expenditure of education as a percentage of total government expenditure.

Exercise 2

Examine the data sets for discrepancies as discussed in class. Describe all problems that you find.

Exercise 3

Clean and simplify the data.

Exercise 4

Report correlations in the data and support your findings by appropriate plots.

Exercise 5

Write a short report in LaTeX and describe how you obtained your findings.