# Information Visualization

# CHECKPOINT I: Visualization Proposal

G15

**1. Domain**

One of the first things we are taught in school is how to read. And it is this knowledge that allows us to get most of the information in our lives. When learning, most of the time the way we gather the information we need is either through reading books, or in the present times, by reading articles on the internet and searching information through websites (i.e. Wikipedia).

Not only is reading important when studying, but also, when reading books in our daily lives, and this is because reading helps us develop skills that are important to our wellbeing. It helps to improve your self-expression capabilities since you are extending your vocabulary, it can teach you how to deal with certain obstacles you find in life or help you to learn a new skill.

Our aim is to understand how different these habits are between different countries throughout Europe, and if it has an impact in the overall picture of the demographics indicators such as the level of dropout or even high achievement in education system. Can the low level of reading associated with a low investment in books be a signal that in a given country the students will underperform in science or even mathematics topics? The aim of our visualization is to answer questions like these (and all the mentioned in the question section) by correlating between reading metrics and the demographic indicators that we chose as being relevant.

**2. Dataset**

In this part of the document intend correlate the reading habit of a given country member of the European Union (EU), and whether it impacts or not in the demographics.

Our goal is to present, our visualization to try to infer whether the reading habit of a given country in European Union (EU), impacts in his demographic or what is the correlation between the country low level of reading habit and the country average income? What about the level of student dropout? What is the performance of students in this country, related to mathematics or science?

One of the challenges associated with these correlations that we try to infer is the quality (and quantity) of the datasets that are used. For that purpose, most of the dataset in use are available on Eurostat and PorData.

We assume the possibility that data from these sources may need to be clean, however most of them are already sort in some order.

For further information, we have the access links below.

<https://ec.europa.eu/eurostat>

<https://pordata.pt>

* + - 1. **3. Example Questions**

1. - Try to provide a description that allows for the understanding of the questions and their complexity: this will be reflected on your grade.
2. - Question enunciation must be as clear and concrete as possible.
3. Below each question is a brief explanation with the tasks required by the user to answer the question.
4. How many hours, in average, do the countries in EU spend reading?
5. What is the average percentage of household expenditure in reading material by country?
6. Given the reading habits of each country, what is the level of literacy of this country comparing to other EU countries?
7. Given a country’s reading habits, what is the rate of dropout?
8. Is there a correlation between low habit of reading and the high academic success given a country?
9. What is the (under)achievement of students in reading, mathematics and science?
10. What is the adult participation in learning after leaving the formal education?
    * + 1. **4. Data Sample**

Some examples of data that show that the above are possible and adequate

(from “time\_spent\_reading\_books\_Data.csv”)

unit; geo; time; sex; value

time; Belgium; 2010; 4

(from “underachive\_reading\_math\_science.csv”)

year; training\_field; country; percentage

2012; Mathematics; Belgium; 19.6

(from “early\_leavers\_edu\_and\_training.csv”)

year; country; percentage

2014; Denmark; 9.5

(from “household\_exp\_books\_newspapers.csv”)

year; country; percentage\_of\_total(value)

2015; Ireland; 0.9

(from “xpto.csv”)

year; name; cost; rating

2012; Potatoes; 12000; 4