# Introduction

## 1 What is data analysis?

Data analysis is a set of methods for extracting information from a data set. It's also known as statistical learning. The idea is to use statistical models to understand how data is structured and how it interacts with each other.

> Example
>
> Let's imagine you work for the United Nations (UN). Your mission is to analyze life expectancy around the world. To do this, you'll have a measure of life expectancy in each UN member country, of course, but also GDP per capita, health expenditure, fertility rate, urbanization rate, education level of the country, and so on. The aim of data analysis is to find links between these different variables and the variable of interest, life expectancy, to visualize these data, and eventually to predict life expectancy from the other variables.

## 2 Course objectives

In this course, we aim to introduce methods that allow us to study a "high-dimensional" dataset (in the sense that we can't simply graph all the variables for each observation) without having to resort to a probabilistic model. The various techniques we'll be looking at can be used to:

- visualize data
- reduce data size;
- identify certain relationships between variables;
- divide the dataset into groups/classes.

This course is not intended to be exhaustive, in the sense of presenting all possible methods. Nor is it intended to be state-of-the-art, in the sense that it will not cover the latest developments in machine learning. Nor is it a programming course.