

# Project Proposal - NTDS 2017

Project members: Blandine Clément, Antonio Jimenez Gonzalez, Salmane Kechkar

## Health assessment at a global and country level for future evaluation.

### 1. Data acquisition

We plan to use the “Health Nutrition and Population Statistics” from Kaggle. The dataset comes from World Bank data and contains 345 indicators concerning health, nutrition, education for 263 countries and several world regions, from 1960 to 2016.

Link of the dataset:

<https://www.kaggle.com/theworldbank/health-nutrition-and-population-statistics>

### 2. Data exploration

This dataset will allow us to explore several trends in 2 dimensions: temporal and spatial.

1. time: explore the temporal dynamics of a country, or a world region.
2. geographical regions: check if there is any outliers country in one given world regions and the cause behind.
3. Determine if there are some correlations between features, for further efficient exploitation.

### 3. Data exploitation

We would like to generate similarity graphs that would have countries as nodes, whose connections would rely on the computed difference between well-chosen features. The features would be categorized according the theme that we want to study (education, sanitation, health etc). We would like to display and point out some clustering patterns between regions that would allow us to:

1. find some relationship between indicator 1 and indicator 2, and how they compare over time.
2. Focus on the evolution of one indicator in one country/one region of the world over time
3. Benefit from the known outcome of past trends to predict some trends for the next years.