Projet 2 : Analyse des données de systèmes éducatifs

**0. Importations des modules et des données**

**1. Vérification et rectification de la qualité des données**

1.0 Description globale des tables

Bilan description globale + MPD de la base

1.1 Données dupliquées ou contradictoires

Vérification de l'unicité des clés de chaque table

Correspondance des entrées uniques des colonnes apparentées

- pays manquant dans 'Country'

- 'Country Name'/'Table Name' de "Data" et "Country"

- 'Indicator Code'/'Series Code' de "Data" et "Series"

- 'Indicator Code' et 'Country Code' dans "Country-Series"/"Footnote" et "Country"/"Series

- 'Indicator Name'/'Series Name' de "Data" et "Series"

Simplification de la base

(- élimination des colonnes redondantes)

(- fusion de 'footnote\_c' et 'cnt\_ser\_c' en 'inf\_data\_c')

- format des données de la colonne "Year" ("FootNote")

Relations bijectives entre colonnes d'une même base

- 'Indicator Name' dans "Data" et "Series

1.2 Données manquantes

- Comptage des données manquantes par table (barres)

- Elimination des colonnes sous-remplies

- Identification des pays et indicateurs sous-remplis ('Data')

- élimination des lignes vides

**2. Exploration des données**

Nombres d'indicateurs disponibles selon les années

Exploration des mots clés dans les noms d'indicateurs

Détection des outliers (boîtes à moustaches)

-------------------------

A FAIRE

finir simplification de la base

finir détection des outliers

sélection des indicateurs pertinents

calcul et tracé d'ordres de grandeurs pertinents