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

Simplification de la base

Relations bijectives entre colonnes d'une même base

Ajout à la table ‘Data’

Elimination des colonnes sous-remplies

Groupes de pays en régions

1.2 Données manquantes

- Comptage des données manquantes par table (barres)

Nombre de pays par région, et d'indicateurs par topic

- 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 en vue de la suppression de colonnes (années) et lignes (pays) sous-remplis ou inutiles

Suppression de lignes et de colonnes

Exploration en vue de la suppression de lignes (indicateurs/pays) sous-remplis

Suppression des lignes indicateur/pays calculés (1 ou 2 valeurs) dans la plage [1990,2015]

Multiindexage de la dataframe data\_c et permutation des axes

Corrélation entre indicateurs d'un même topic

Détection des outliers

Evaluation du nombre d'outliers selon le seuil

Elimination des outliers

Evaluation du degré de distinctivité des indicateurs entre pays

**3. Tracé de certains indicateurs pertinents**

3.0 Overview rapide des indicateurs : exploration des mots clés dans les noms d'indicateurs

3.1 Analyse de la problématique¶

3.2 Sélection grossière d'indicateurs pertinents

3.3 Sélection minimale d'indicateurs à tracer

3.3 Graphiques

1. Public visé :

Enrolment in upper secondary education (number) UIS.E.3

Enrolment in post-secondary non-tertiary education (number) UIS.E.4

Population dans la tranche 15-24 ans (number) SP.POP.1524.TO.UN

2. Possibilités de déploiement dans le pays :

Internet users per 100 people IT.NET.USER.P2

3. Besoins en matière d'éducation

Pupil/teacher ratio in upper secondary UIS.PTRHC.3

Pupil/teacher ratio in tertiary UIS.PTRHC.56

4. Possibilités de financement

GDP per capita (current USD) NY.GDP.PCAP.CD