academy

Analyse exploratoire - Projet d’expansion à l’international

Analyse des datasets de « The World Bank »

Dernière MàJ
01 Décembre 2022

academy

Formation en ligne pour un public de niveau lycée et université

Projet d'expansion à l'international

Quels sont les **pays avec un fort potentiel** de clients pour nos services ?

Pour chacun de ces pays, quelle sera l'**évolution** de ce potentiel de clients ?

Dans quels pays l'entreprise doit-elle **opérer en priorité** ?

Analyse exploratoire des données sur l'éducation (The World Bank)

Informations pertinentes ? De qualité ?

Identification d'indicateurs pour aider à la prise de décisions

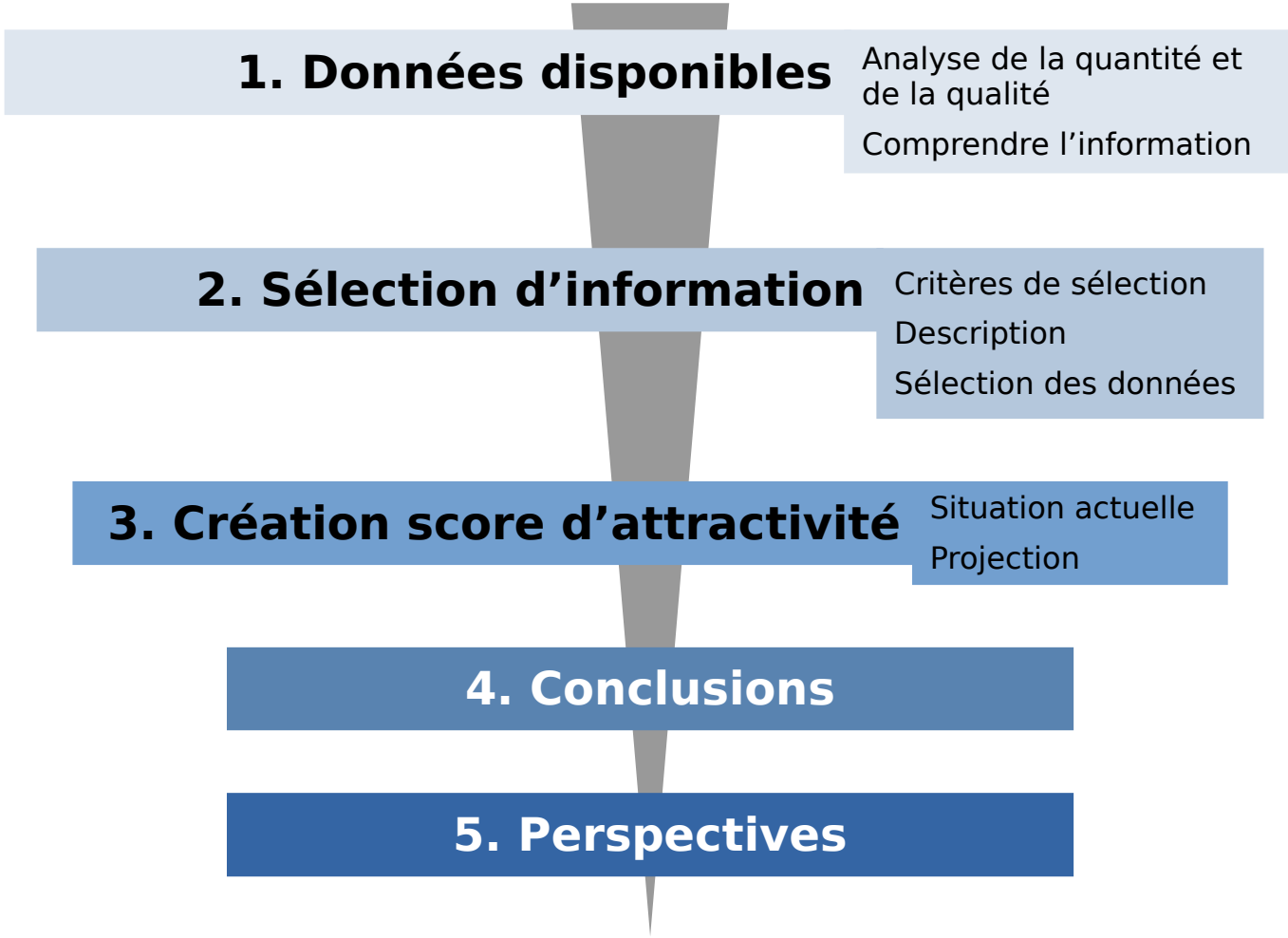
Conclusions : est-il possible de départager les pays ?

Pays à fort potentiel ?

Évolution ? Priorisation ?

THE WORLD BANK

SOMMAIRE


1. Données disponibles
Analyse de la quantité et de la qualité
Comprendre l'information
2. Sélection d'information
Critères de sélection
Description
Sélection des données
3. Création score d'attractivité
Situation actuelle
Projection
4. Conclusions
5. Perspectives

1. Données disponibles : Analyse de la quantité et de la qualité

academy

EdStatSeries

Décrit les **indicateurs** et les classe par thème.
3665 lignes et 20 colonnes. Remplissage colonnes min 0 % ; max 100 % .

THE WORLD BANK

EdStatCountry

Informations **socio-économiques** des **pays**.
241 lignes et 31 colonnes. Remplissage (colonnes autres que nom) min 13,28 % ; max 99,58 % .

EdStatData

Valeurs des **indicateurs** pour les différents **pays** entre **1970 et 2100**
886930 lignes et 69 colonnes
Colonnes 1970 à 2100 remplissage min 0,16 % ; max 27,33 %.

EdStatFootNote

Détaille des **informations** sur les valeurs de couples **pays/indicateur** pour certaines années.
643638 lignes et 4 colonnes. 100 % remplies.

EdStatCountrySeries

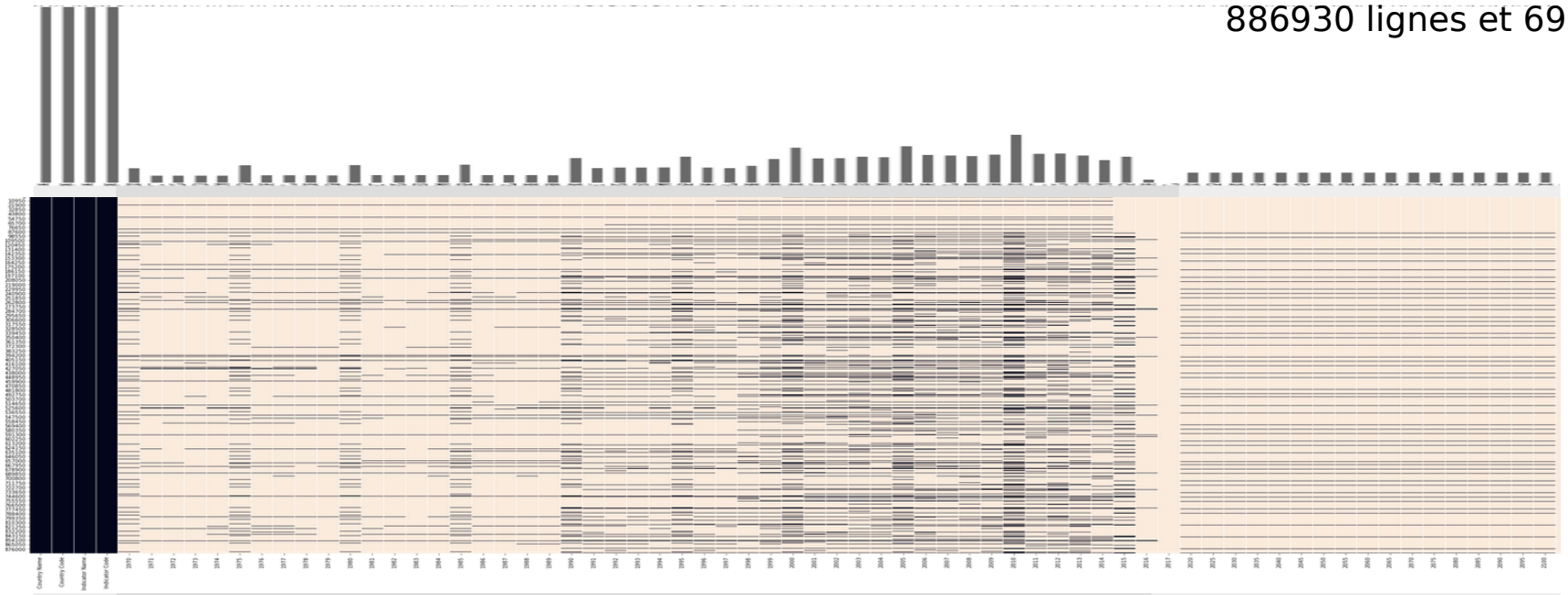
Précisions sur la **source des données** des couples Pays/Indicateur
613 lignes et 3 colonnes. 100 % remplies.

1. Données disponibles : Analyse de la quantité et de la qualité

academy

EdStatData

886930 lignes et 69 colonnes



Pays et indicateurs (Nom et Code)

Taux remplissage 100 %

Observations intervalle 1970 – 2017

Taux remplissage colonnes 1970 à 2017 entre 0,2 % et 27,3 %

Projections, intervalle 5 ans, 2020 – 2100

Taux remplissage 5,8 %

TOTAL 3665 indicateurs

3353 indicateurs renseignés 1970-2016

308 renseignés 1970-2100

4 Indicateurs jamais renseignés

THE WORLD BANK

1. Données disponibles : Comprendre l'information

academy

EdStatSeries

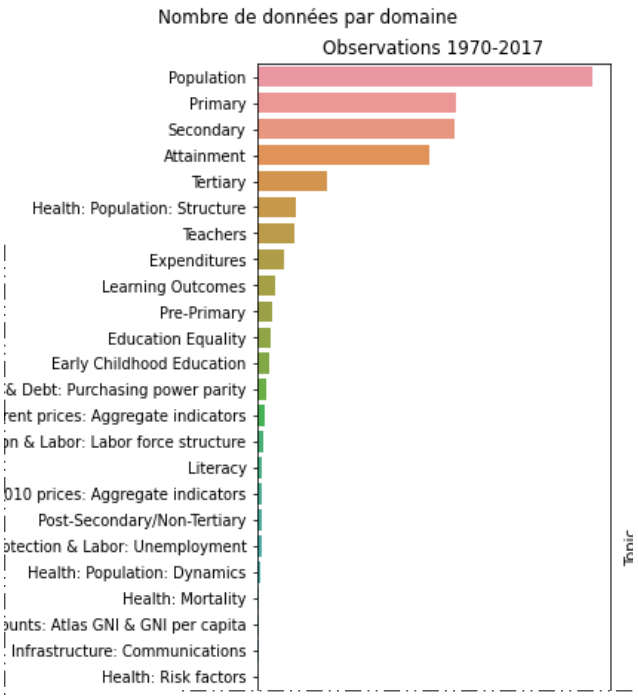
Décrit les **indicateurs** et les classe par thème

EdStatData

Valeurs des **indicateurs** pour les différents **pays** entre **1970 et 2100**

EdStatCountry

Informations **socio - économiques** des **pays**

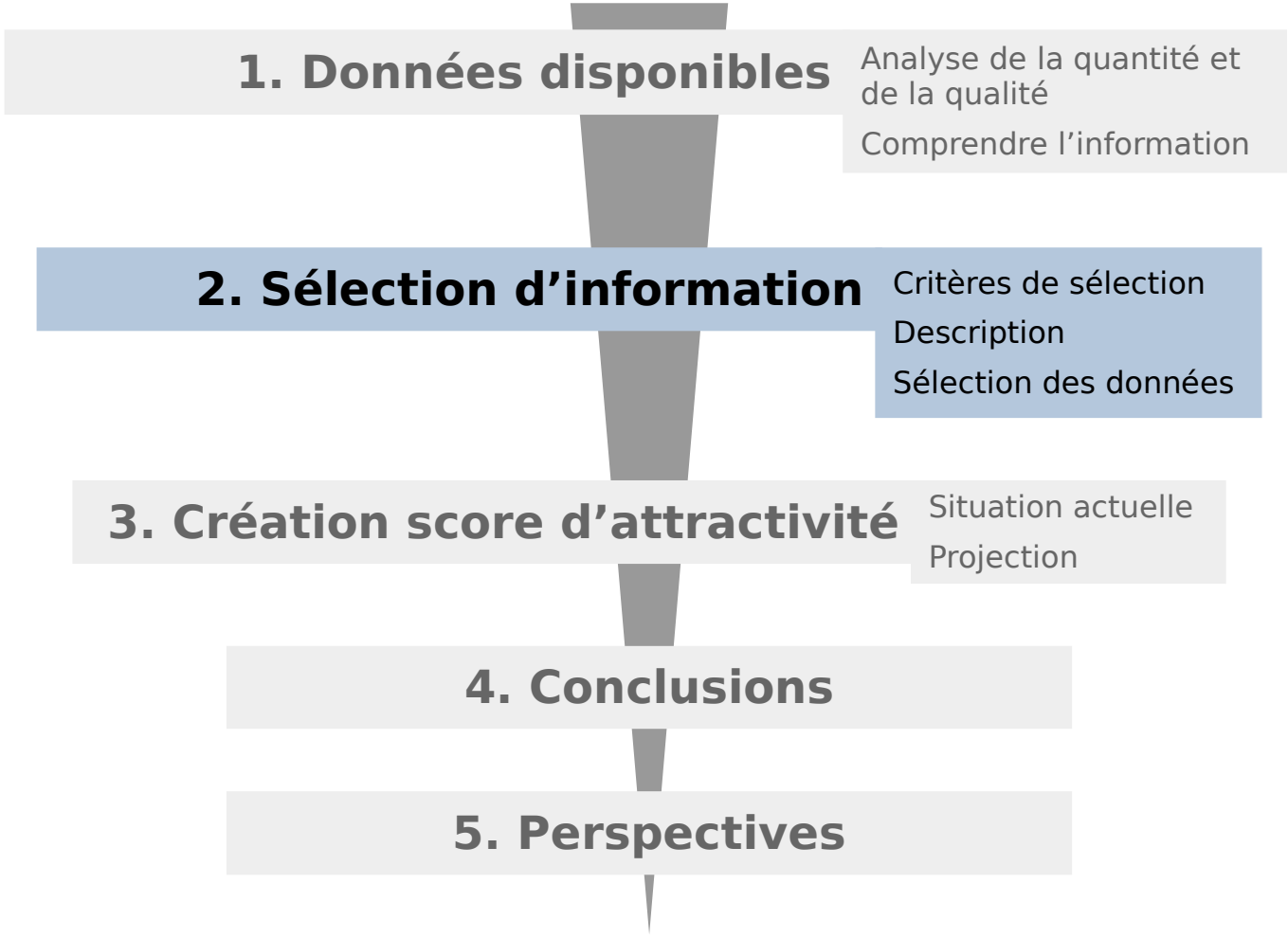


Indicateurs 3665 , 37 thèmes
2020-2100 seulement « Attainment » (niveau de scolarité)

Pays (216 pays + 26 groupements)
Toutes les régions géographiques
Tous les niveaux de revenus

THE WORLD BANK

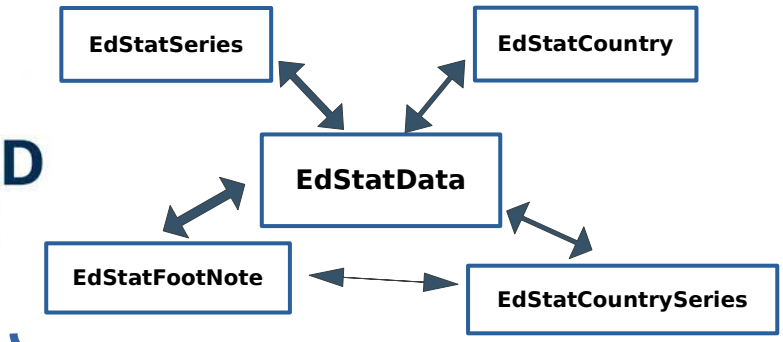
SOMMAIRE

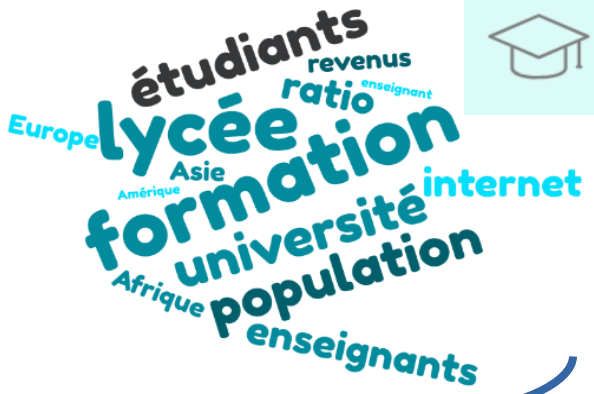

1. Données disponibles
Analyse de la quantité et de la qualité
Comprendre l’information
2. Sélection d’information
Critères de sélection
Description
Sélection des données
3. Création score d’attractivité
Situation actuelle
Projection
4. Conclusions
5. Perspectives

2. Sélection d'information : Critères de sélection

academy

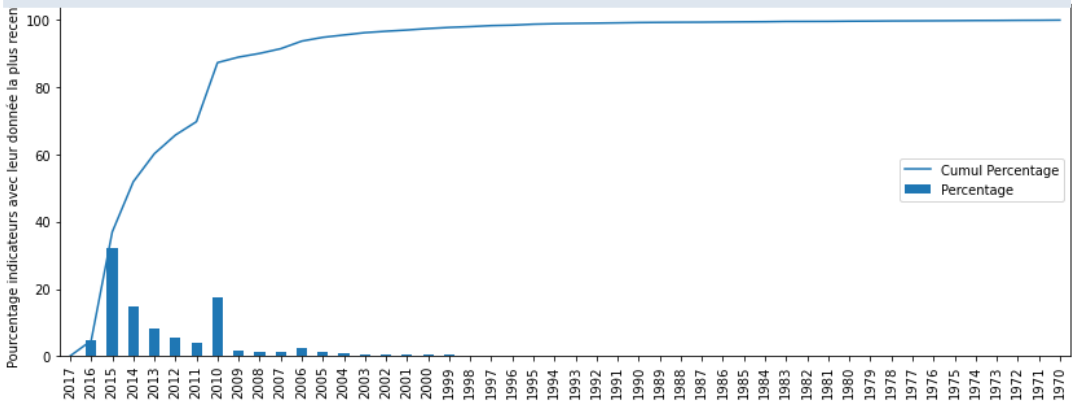
THE WORLD BANK

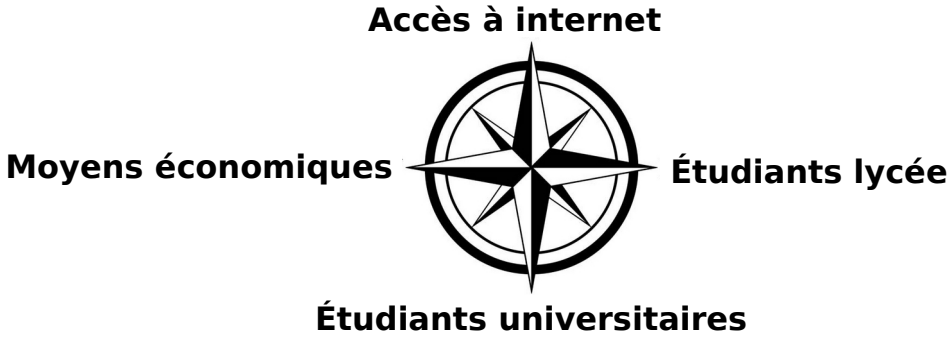




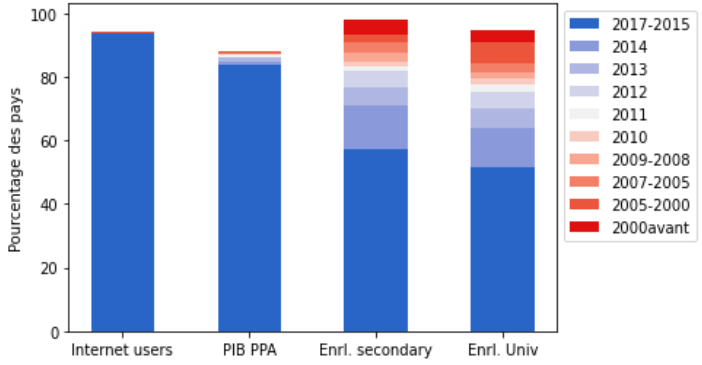
academy

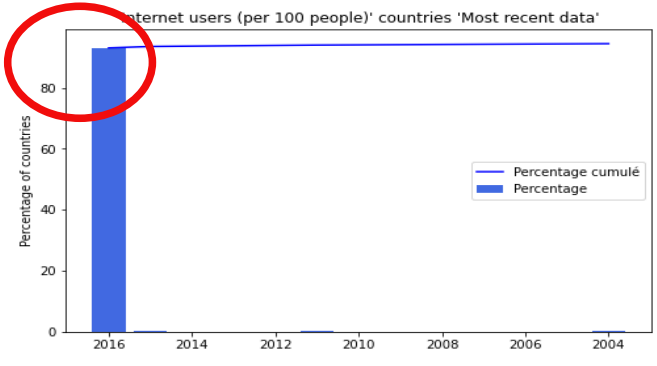
Age de la donnée la plus récente des indicateurs (pourcentage)





2. Sélection d'information : Critères de sélection



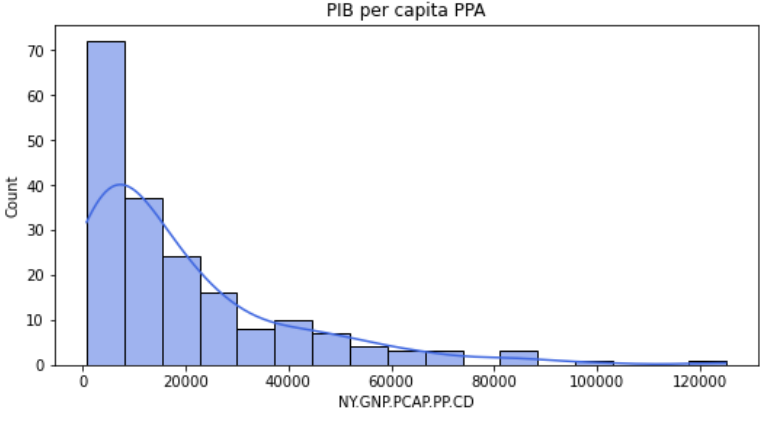


Accès à internet
Utilisateurs de l'internet (pour 100 personnes) : IT.NET.USER.P2

Moyens économiques
PIB par habitant à parité de pouvoir d'achat : NY.GNP.PCAP.PP.CD

Étudiants lycée
Étudiants dans l'éducation secondaire : SE.SEC.ENRL

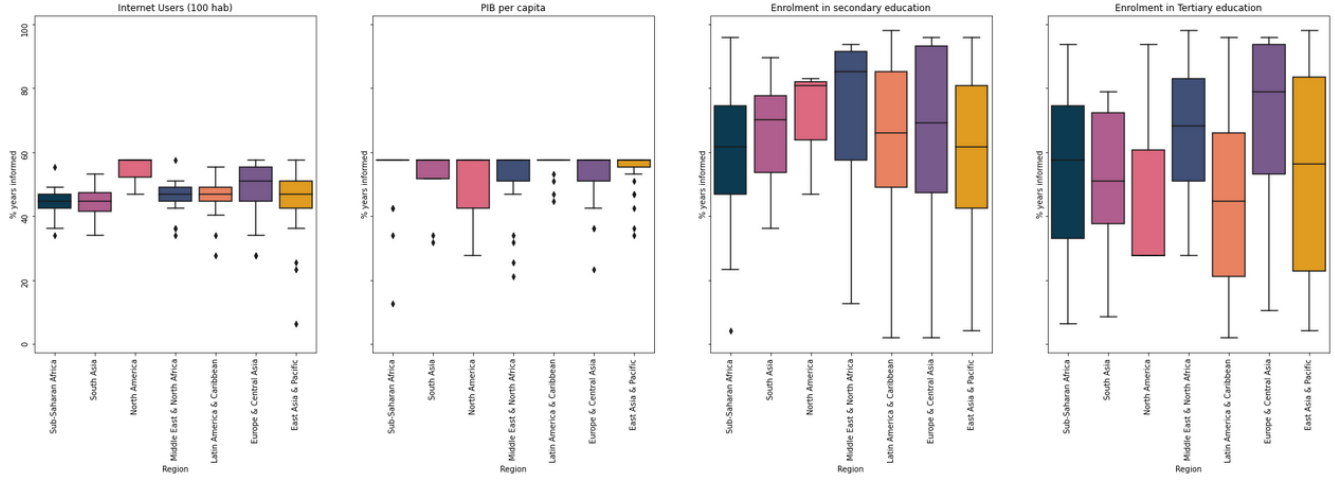
Étudiants universitaires
Étudiants dans l'éducation tertiaire (université) : SE.TER.ENRL



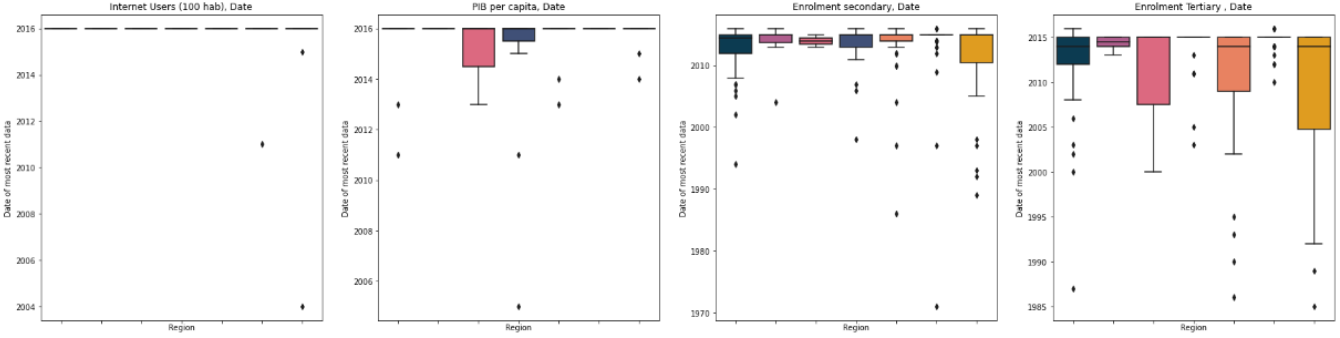
Indicator Code	Accés internet 100 hab.	PIB per capita PPA	Étudiants secondaire	Étudiants université
Type	float64	float64	float64	float64
count	204.0	189.0	210.0	203.0
mean	51.396525	19336.984127	2772515.764732	1045266.701201
std	28.453198	21012.314092	11153604.737931	4116538.575868
Median	53.613386	11990.0	476282.171875	158262.0
Variance	805.615908	439181272.915092	123810503894514.59375	16862412556722.751953
Skewness	-0.059856	1.927155	9.219282	8.088915
Kurtosis	-1.272158	4.496923	94.695131	72.628869

2. Sélection d'information : Description

Pourcentage d'**années renseignés** pour chacun des indicateurs sélectionnés par Région



Année de la donne la plus récente



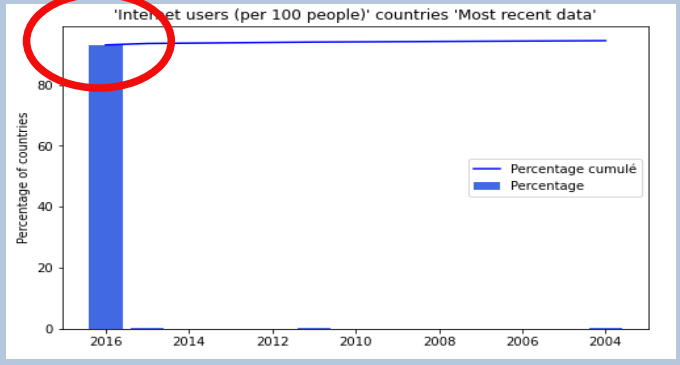
- Pas de grosses différences du taux de renseignement des indicateurs sélectionnés par Région
- Différents patrons en fonction des indicateurs

2. Sélection d'information : Sélection de données

academy

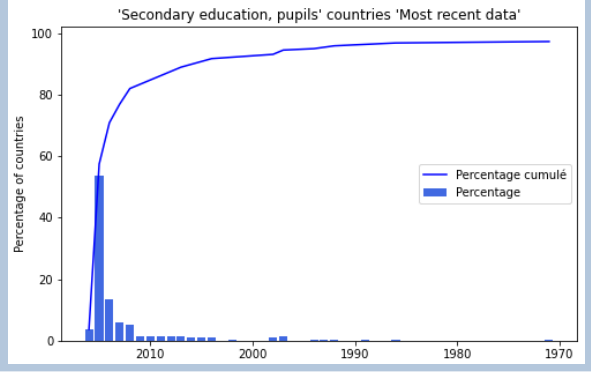
Pas d'action nécessaire

Toutes les données disponibles sont récentes

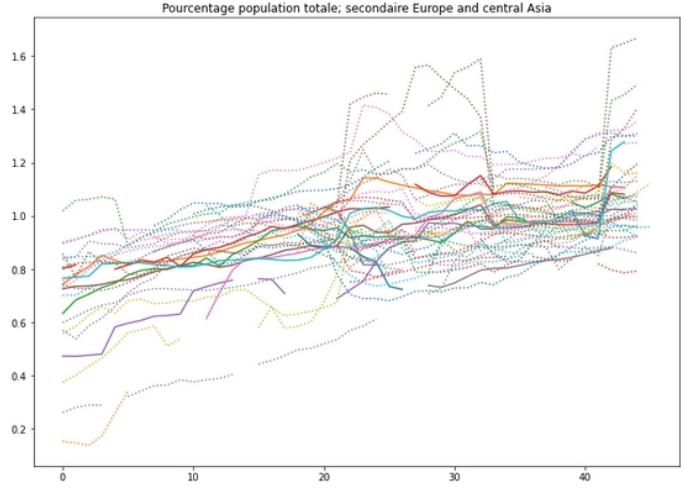


« Actualisation des données »

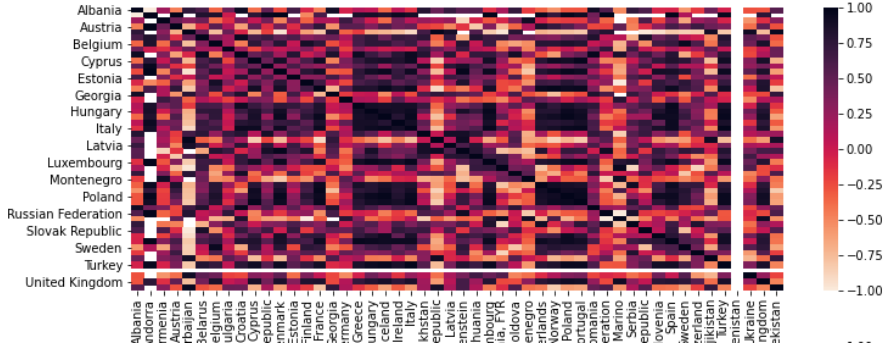
La donnée la plus récente pour certains pays est obsolète



Pour atténuer l'influence des variations de population : calcul **pourcentage population dans l'âge de du secondaire qui réalise des études**



- Identification du pays avec données récentes qui présente la meilleure corrélation (seuil 0,9)
- Calcul de la tendance de variation du pays-modèle, application de cette tendance aux derniers donnés du pays à compléter



2. Sélection d'information : Sélection de données

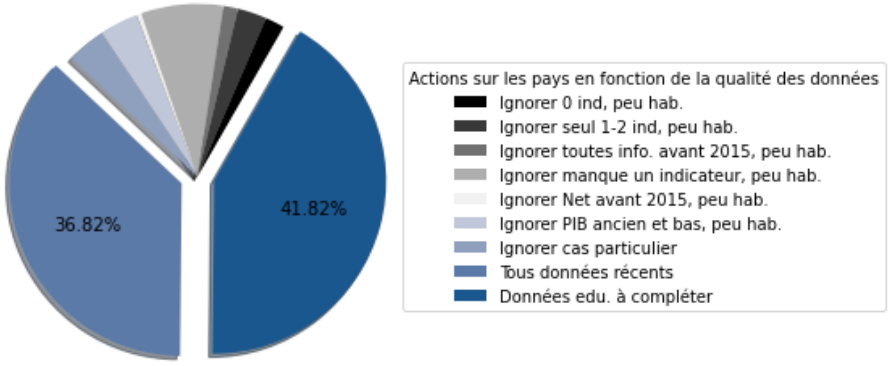
academy

Suppression des lignes sans information

Compiler les données plus récentes

Sélection indicateurs par pertinence (analyse métier)

Suppression pays sans informations récentes et moins 1,000,000 habitants

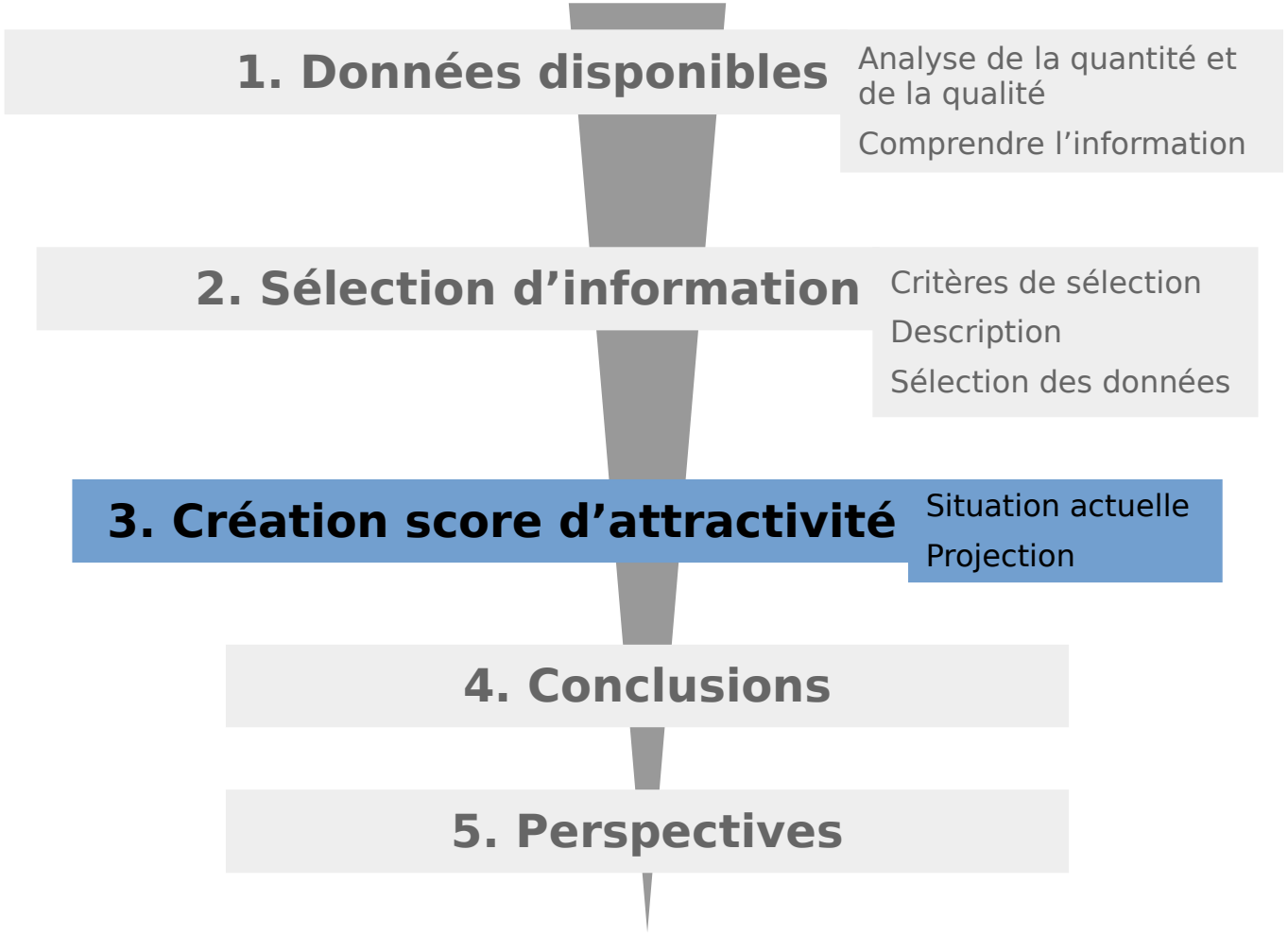


Données pour la création du score d'attractivité

	Country Name	Data_IT.NET.USER.P2	Data_NY.GNP.PCAP.PP.CD	Total Students
0	Afghanistan	10.595726	1900.0	2.970635e+06
1	Albania	66.363445	11670.0	4.756060e+05
2	Algeria	42.945527	14420.0	5.699442e+06
4	Angola	13.000000	6100.0	1.355244e+06
5	Antigua and Barbuda	73.000000	22130.0	1.022028e+04
...	...	...	...	...
200	Uzbekistan	46.791287	6650.0	4.174423e+06
203	Vietnam	46.500000	6170.0	1.342438e+07
206	Yemen, Rep.	24.579208	2500.0	2.035600e+06
207	Zambia	25.506579	3850.0	7.658013e+05
208	Zimbabwe	23.119989	1810.0	1.355750e+05

177 rows × 4 columns

SOMMAIRE


1. Données disponibles
Analyse de la quantité et de la qualité
Comprendre l’information
2. Sélection d’information
Critères de sélection
Description
Sélection des données
3. Création score d’attractivité
Situation actuelle
Projection
4. Conclusions
5. Perspectives

3. Création score d'attractivité : Situation actuelle

Sur les 2016 pays iniciaux, **177 pays ont été classés** en fonction de leur attractivité...

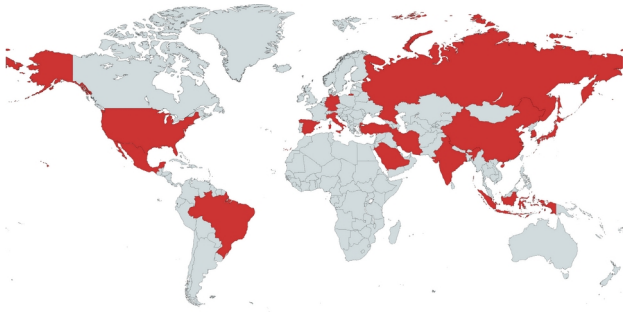
Parmi les 3665 indicateurs disponibles, **4 indicateurs** ont été sélectionnés par leur pertinence et qualité ...

Utilisateurs de l'internet (pour 100 personnes)

PIB par habitant à parité de pouvoir d'achat

Variable techno-économique

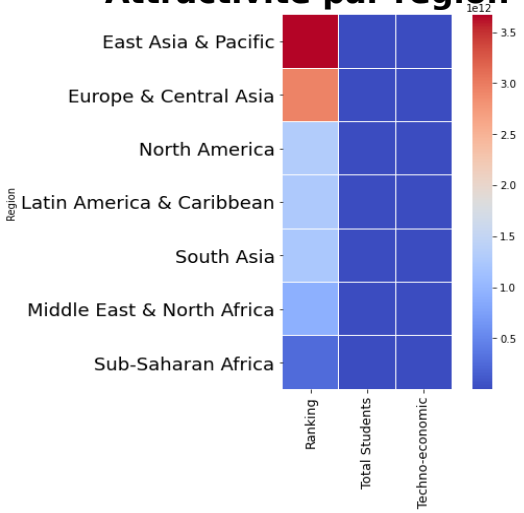
Nombre d'étudiants



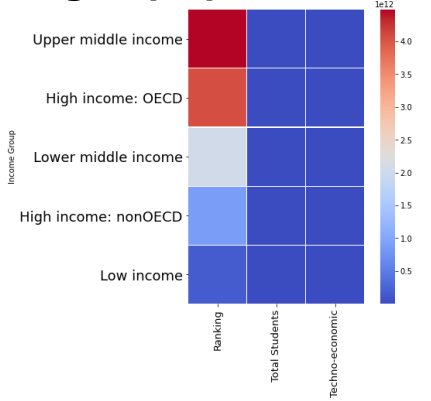
Top15 des pays

	Country Name
0	China
1	United States
2	India
3	Germany
4	Brazil
5	Japan
6	Turkey
7	Russian Federation
8	Indonesia
9	Mexico
10	Saudi Arabia
11	Italy
12	Korea, Rep.
13	Iran, Islamic Rep.
14	Spain

Attractivité par région



Attractivité par groupe par revenus

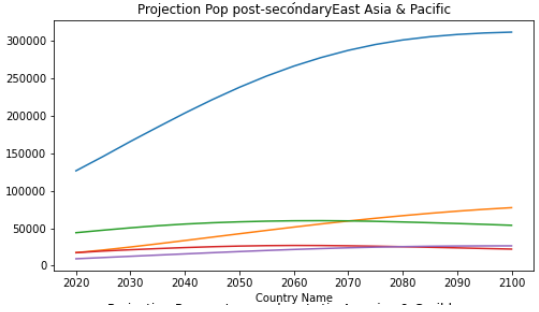


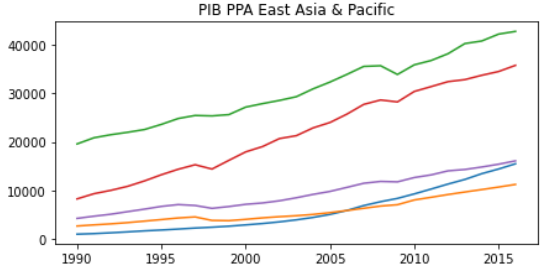
3. Création score d'attractivité : Évolution de l'attractivité

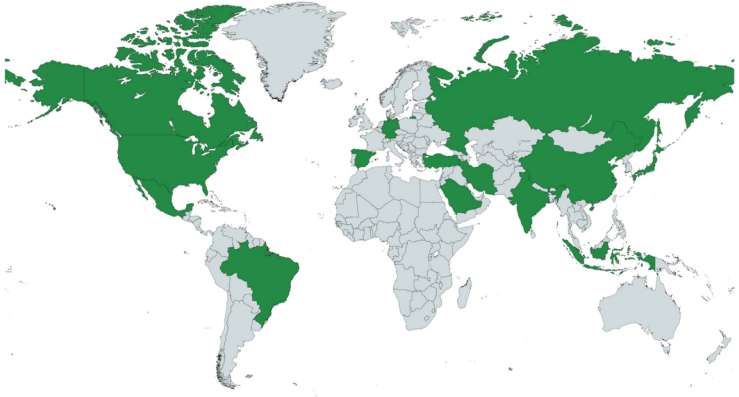
Les 20 pays les mieux classés

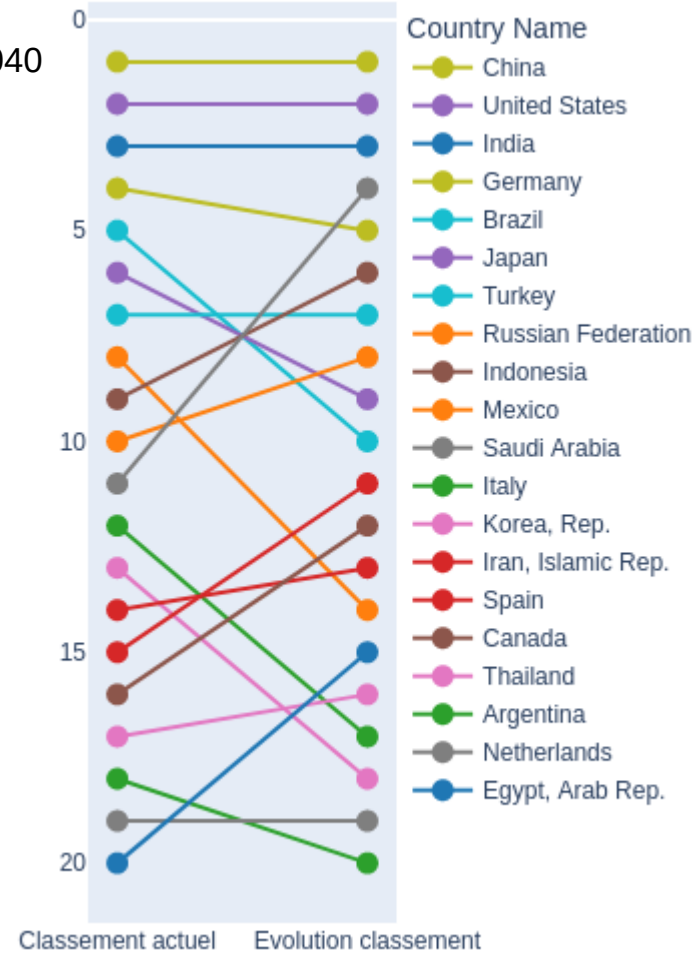
Données 2020-2100, population avec études post-secondaire → tendance 2020-2040
Internet et PIB, données observations → tendance 2010-2016

Reproduction score attractivité sur les tendances et application du score au classement à partir des observations

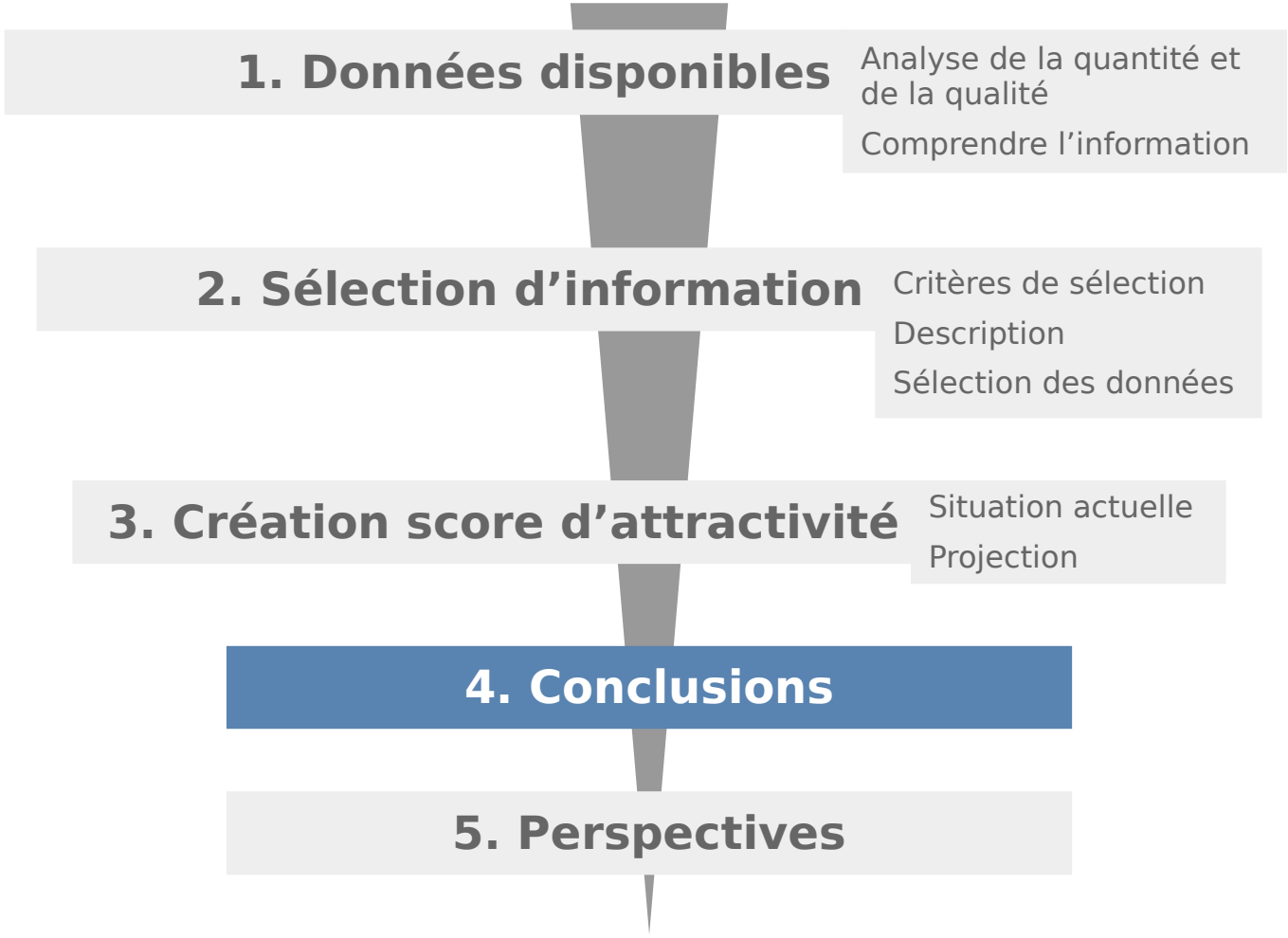




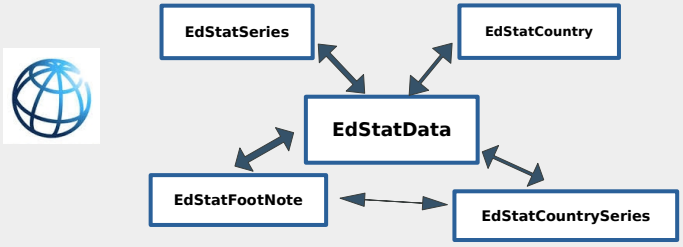




SOMMAIRE

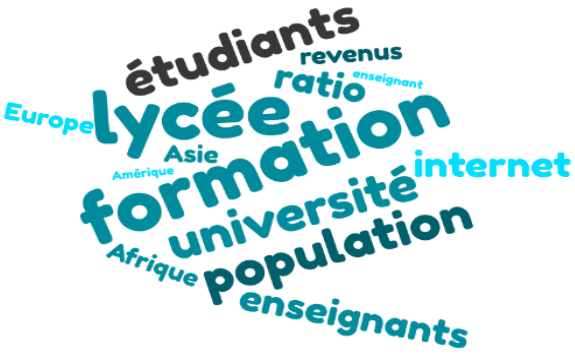

1. Données disponibles
Analyse de la quantité et de la qualité
Comprendre l’information
2. Sélection d’information
Critères de sélection
Description
Sélection des données
3. Création score d’attractivité
Situation actuelle
Projection
4. Conclusions
5. Perspectives

4. Conclusions

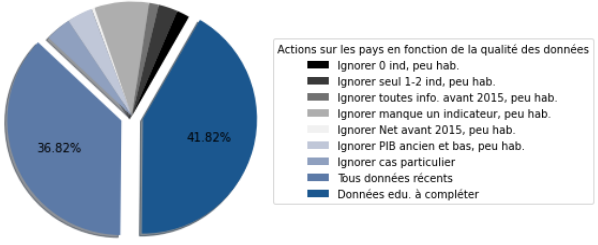


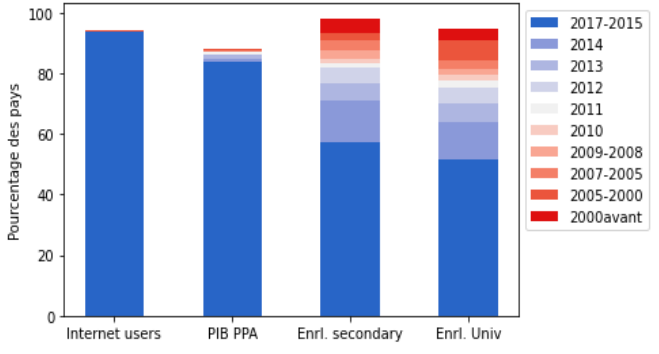
Indicateurs :
- 1970-2017 Observations : 3661 indicateurs
- 2020-2100 Projections : 308 indicateurs‘Educational attainment’

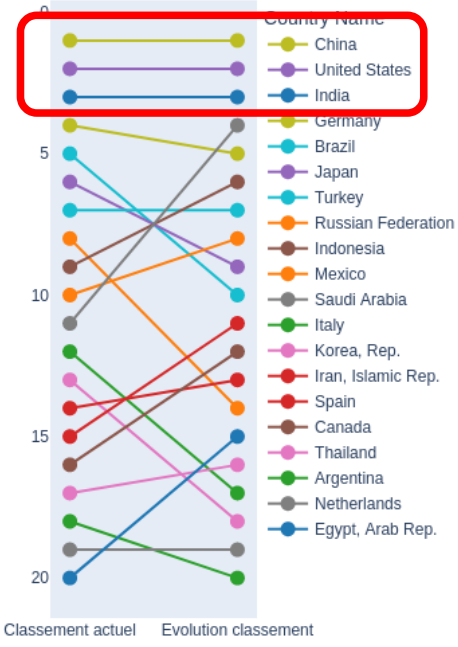
« **Countries** » :
- 216 pays
- Toutes régions géographiques renseignées
- Toutes les niveaux de revenus renseignés



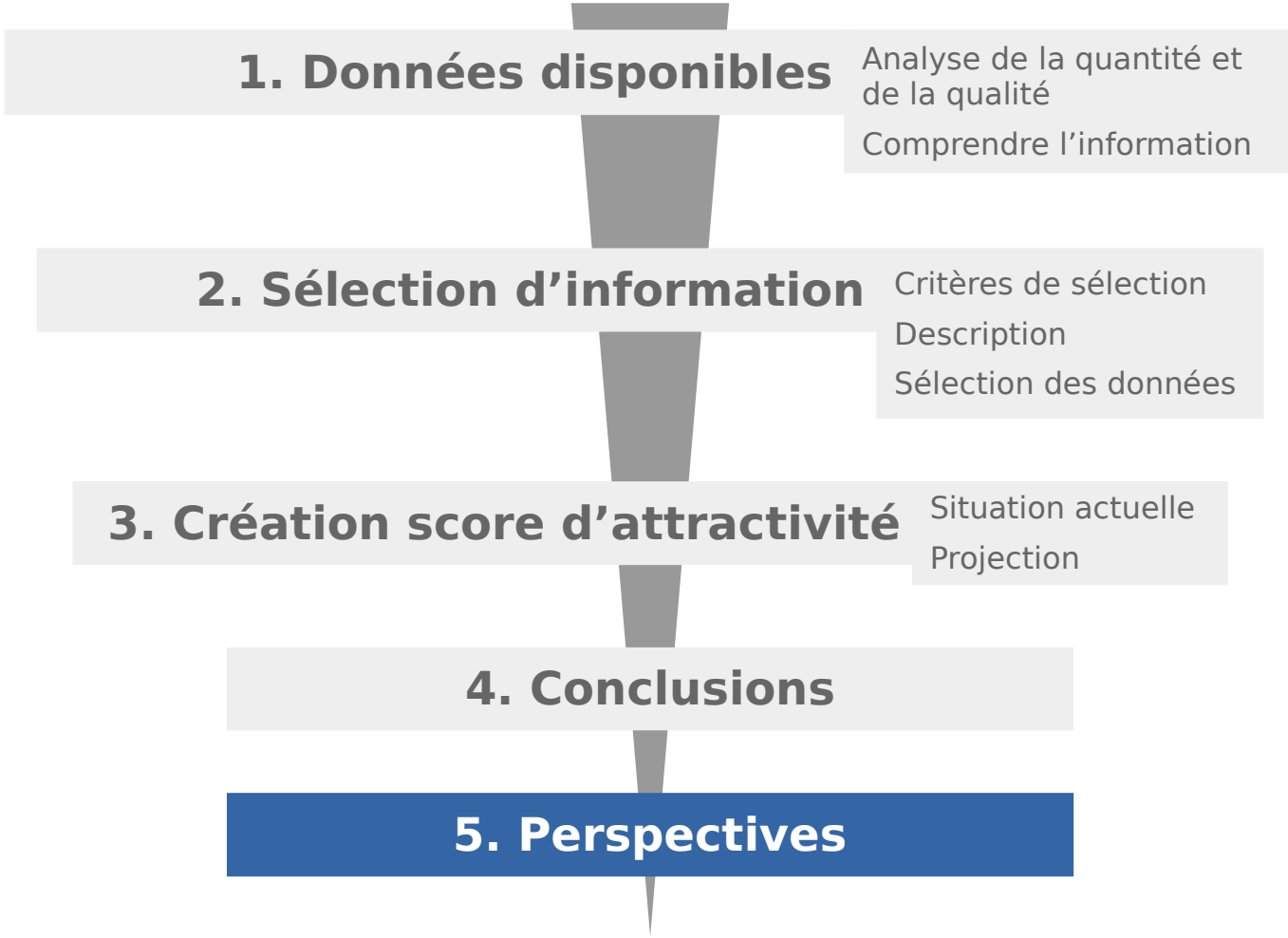
- Sur les 216 pays initiaux, **177 pays ont été classés**
- **Données 2015 ou plus récent**, recherche pays modèles pour créer projections si nécessaire
- **Ranking** prenant en compte la variable **techno-économique** et le **nombre total d’étudiants**







SOMMAIRE

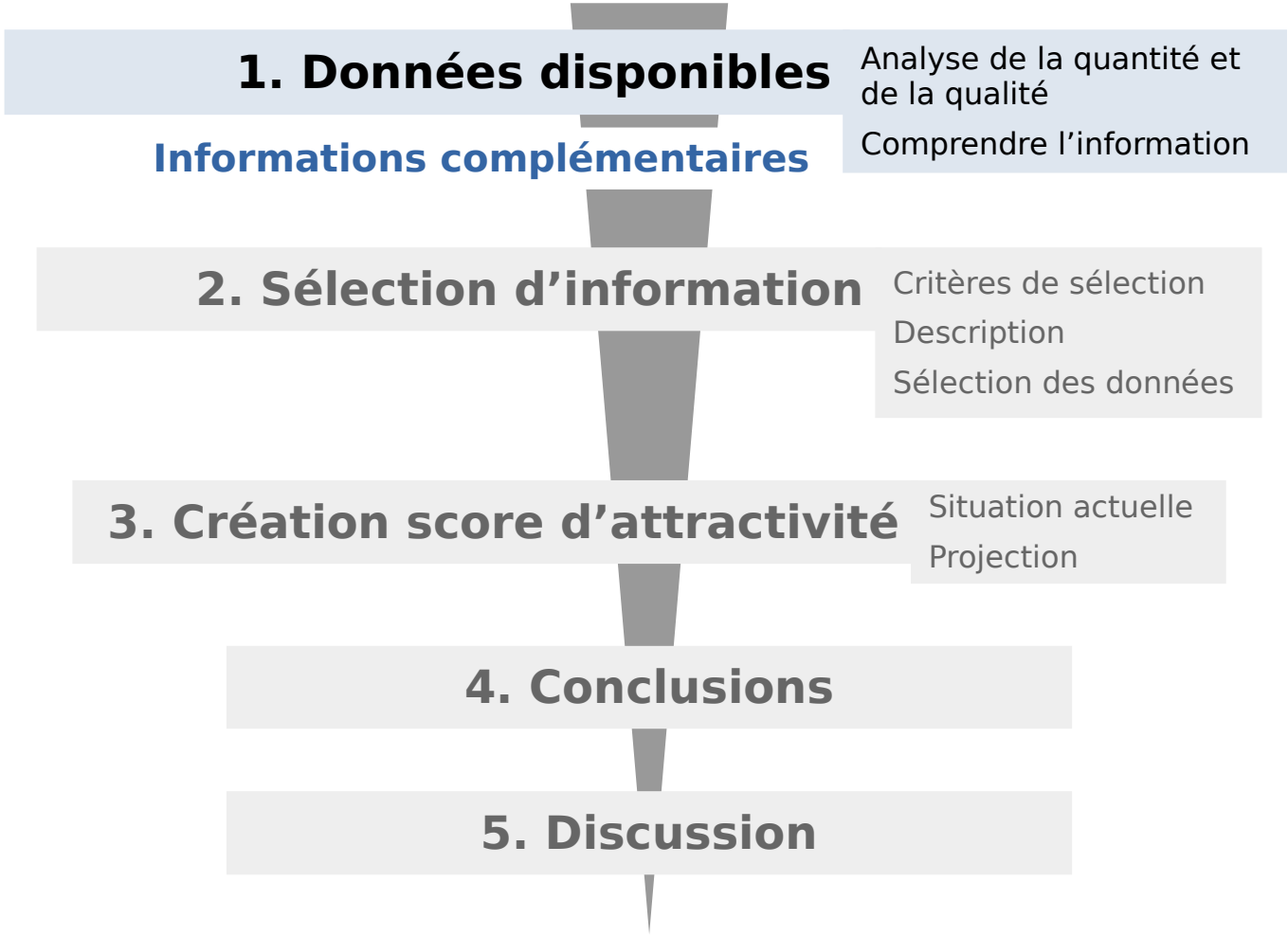

1. Données disponibles
Analyse de la quantité et de la qualité
Comprendre l’information
2. Sélection d’information
Critères de sélection
Description
Sélection des données
3. Création score d’attractivité
Situation actuelle
Projection
4. Conclusions
5. Perspectives

5. Perspectives

- Misse à jour du jeu de données
- Prise en compte de plus de précisions sur la stratégie d'entreprise (lange , proximité géographique, prix)
- Information concurrence
- Intégrer les retours des décideurs pour peaufiner l'analyse
- Améliorer la robustesse du score d'attractivité
- Améliorer la robustesse de l'estimation de l'évolution de l'attractivité

BOÎTE À OUTILS

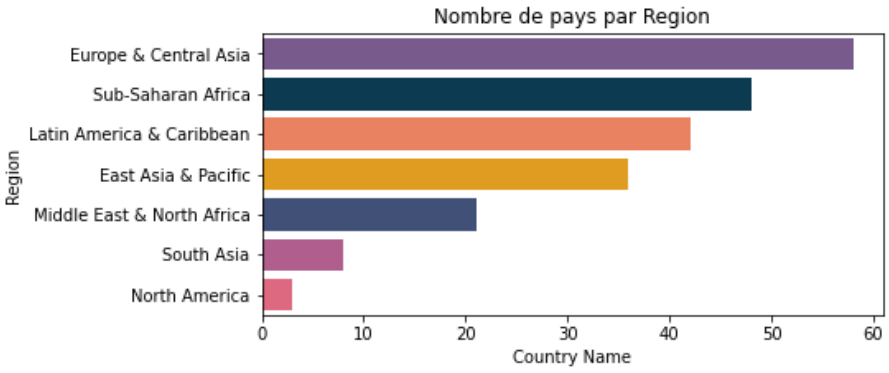
SOMMAIRE

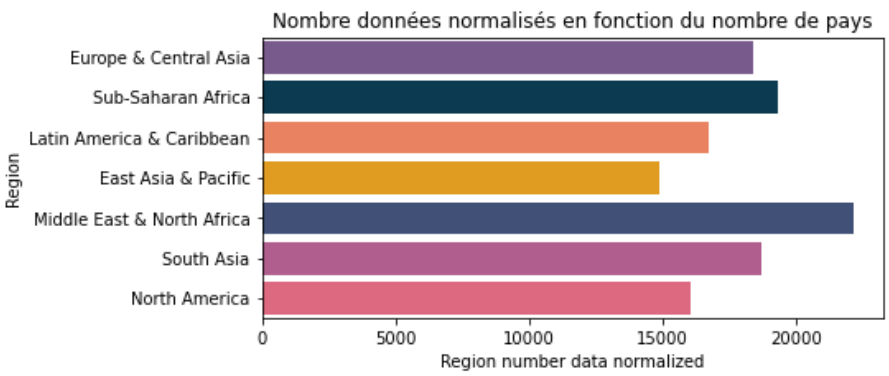

1. Données disponibles
Analyse de la quantité et de la qualité
Comprendre l'information
Informations complémentaires
2. Sélection d'information
Critères de sélection
Description
Sélection des données
3. Création score d'attractivité
Situation actuelle
Projection
4. Conclusions
5. Discussion

1. Données disponibles : Comprendre l'information

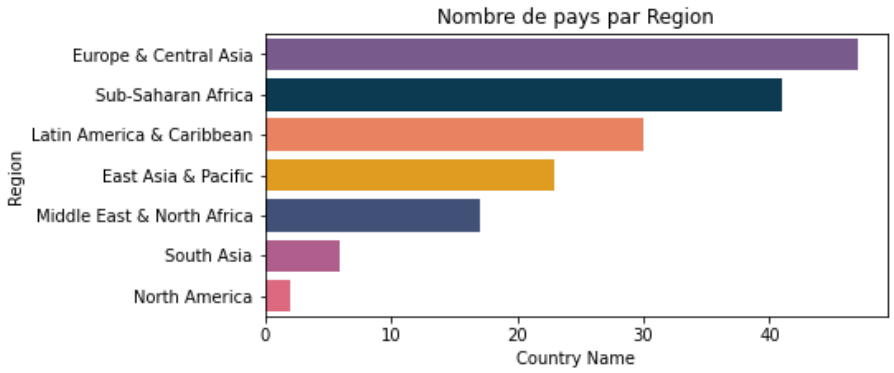
Pays (216 pays + 26 groupements) **:Indicateurs** 3665

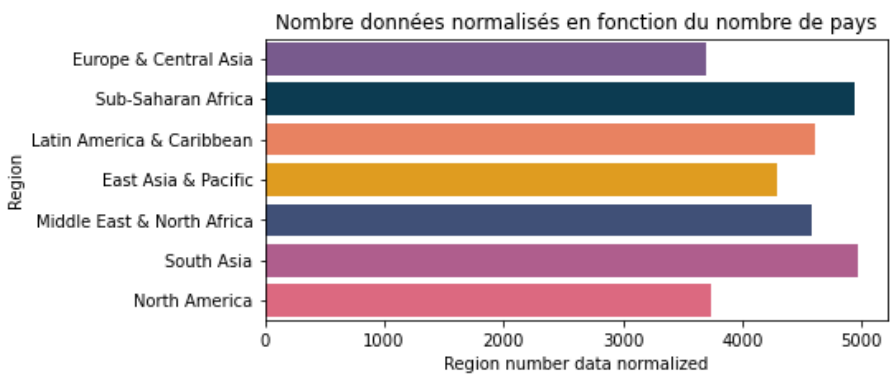
1970-2017





2020-2100

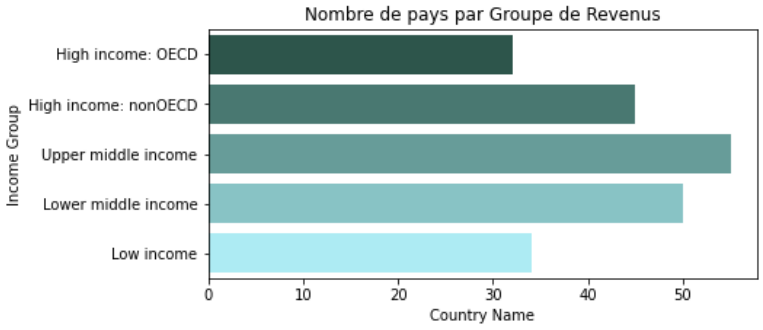


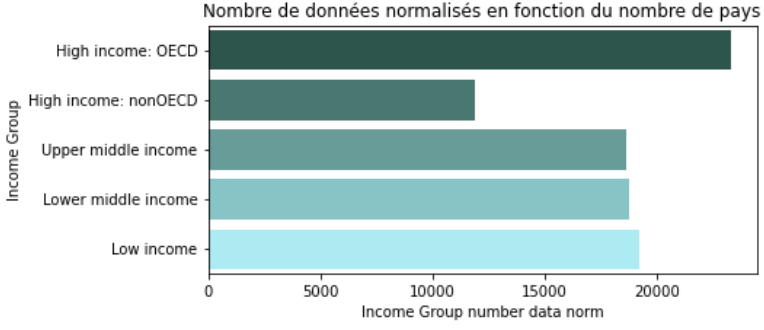


1. Données disponibles : Comprendre l'information

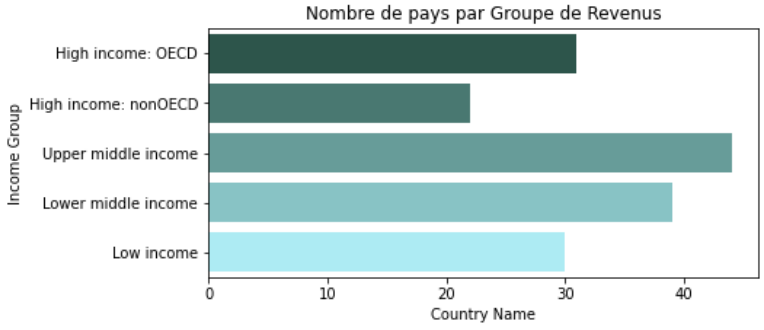
Pays (216 pays + 26 groupements) **:Indicateurs** 3665

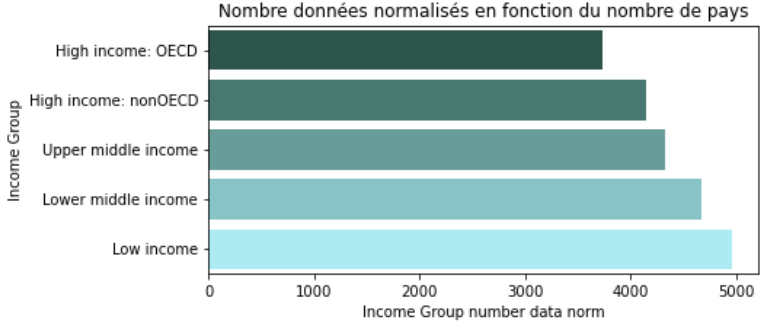
1970-2017





2020-2100

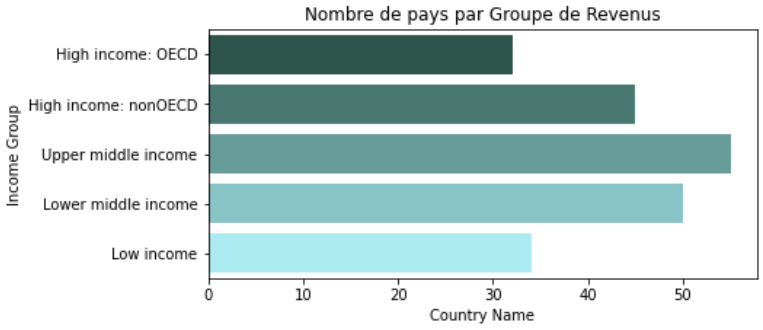


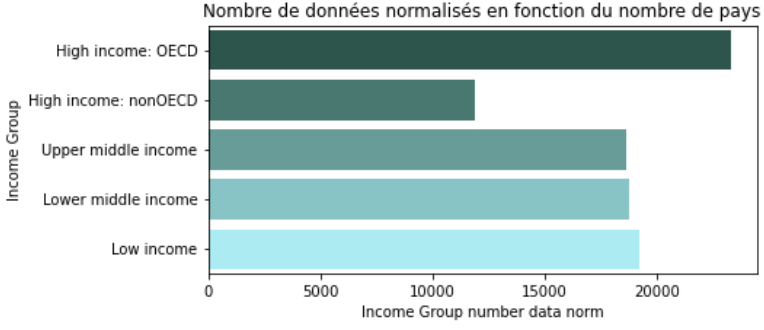


1. Données disponibles : Comprendre l'information

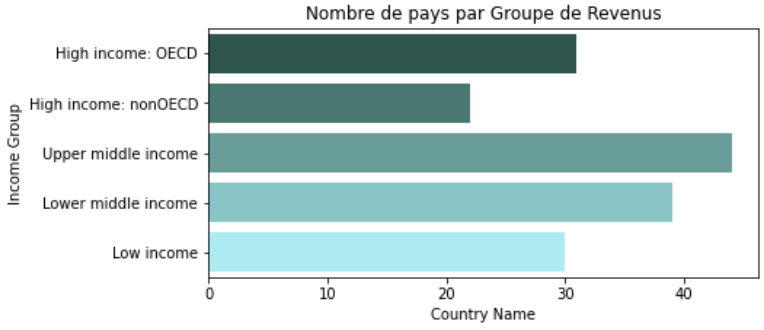
Pays (216 pays + 26 groupements) **:Indicateurs** 3665

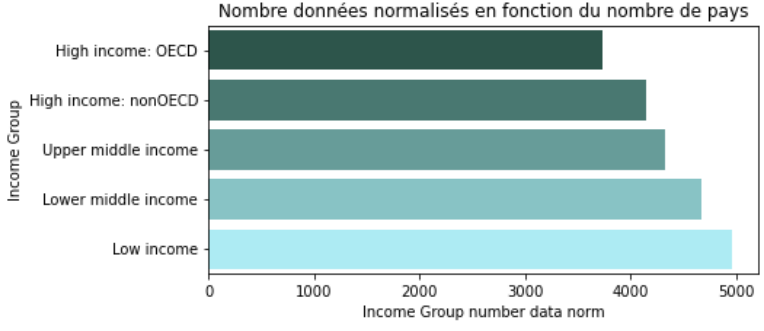
1970-2017





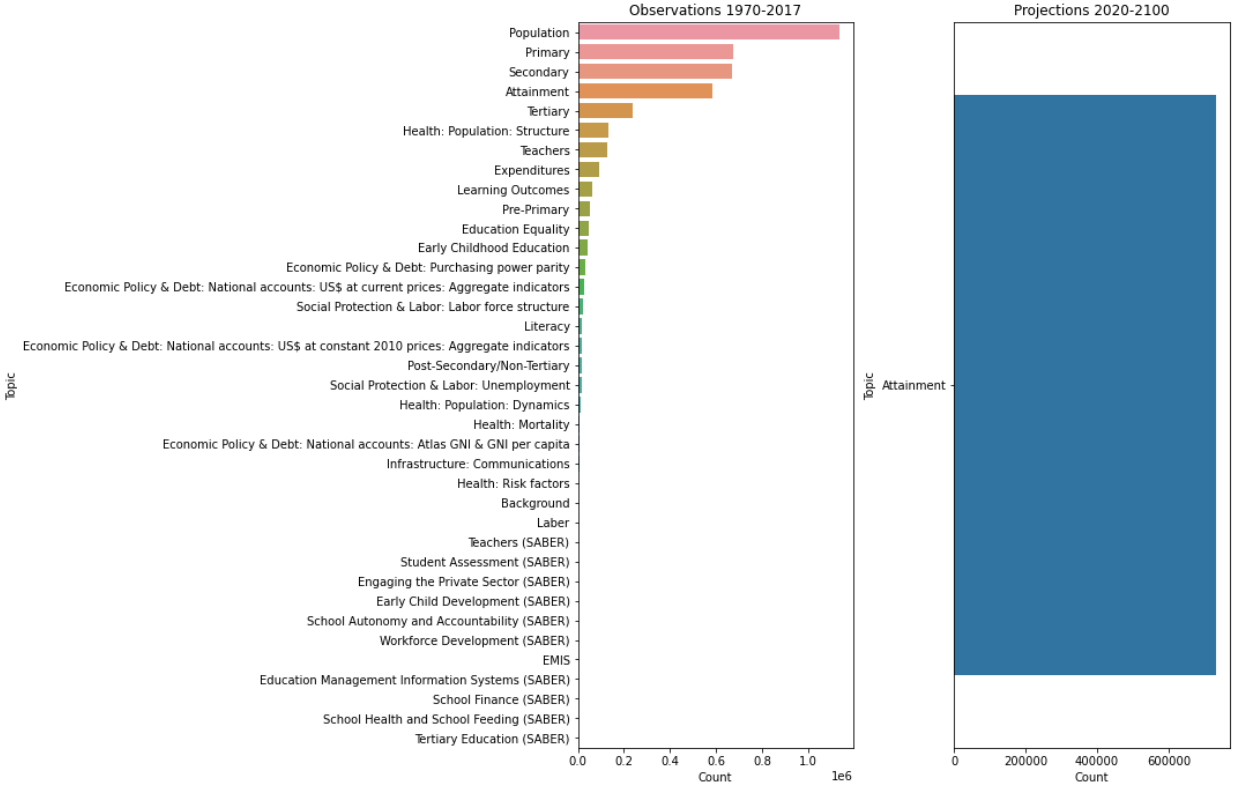
2020-2100



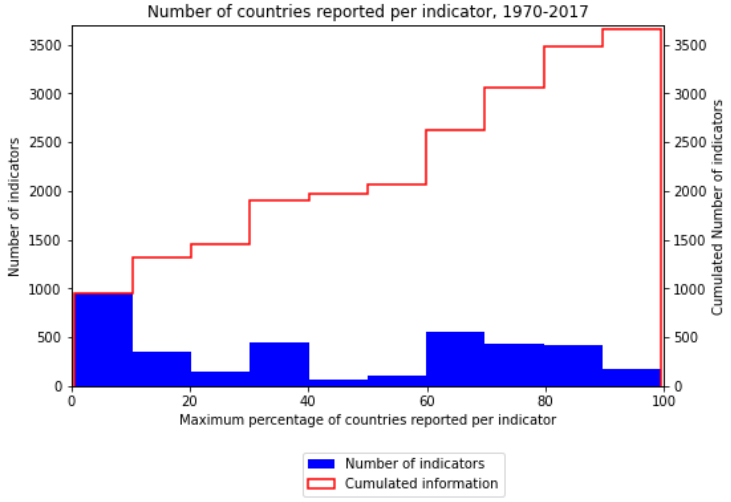


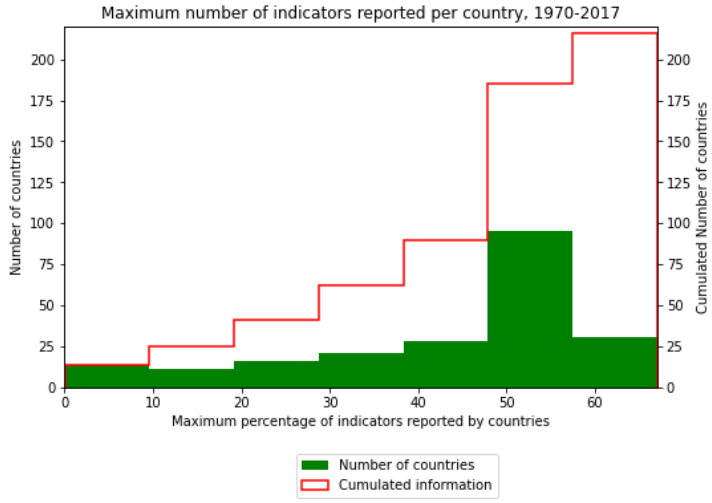
1. Données disponibles : Comprendre l'information

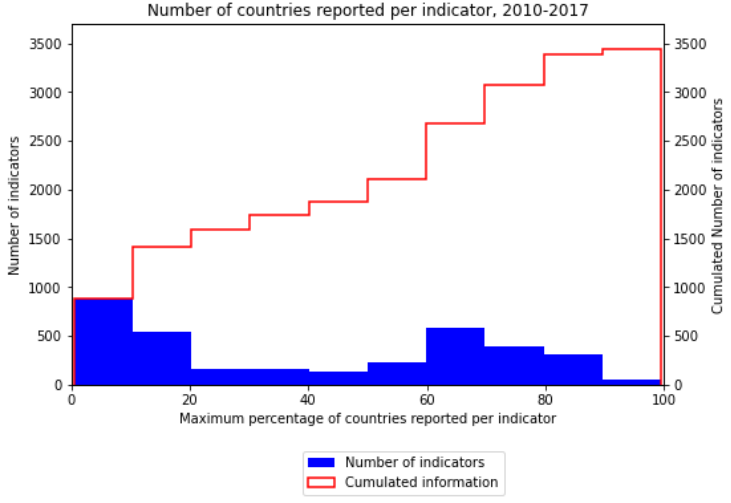
Nombre de données par thème

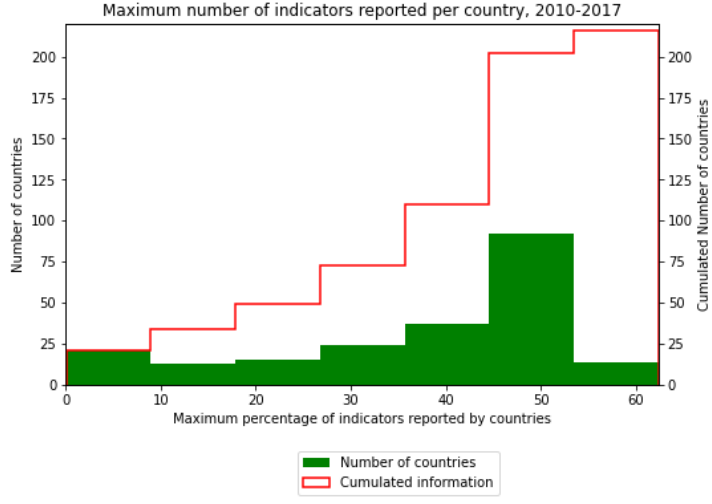


1. Données disponibles : Comprendre l'information









Création des données **« Most recent data »** et **« Date of most recent data »**

Analyse par indicateur. Pour chaque indicateur combien (numéro et%) des pays ont leur « Most Recent Data » jusqu'en… TABLEAU (DataperYearInd)

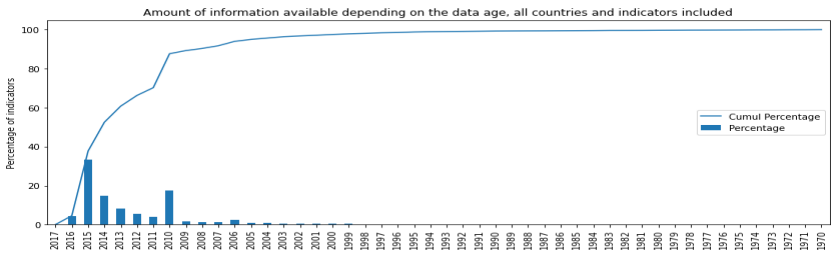
Compréhension des indicateurs.

Nombre d'indicateurs en fonction du % cumule (pays renseignés) max

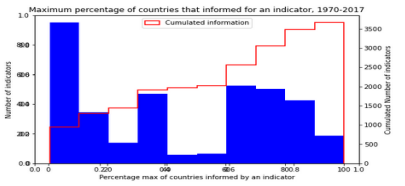
Pourcentage des indicateurs renseigné par les pays

Analyse de la distribution de l'ensemble de ces données, à niveau **global** (tous les pays et indicateurs compris) combien de « Most recent data » j'ai par année .

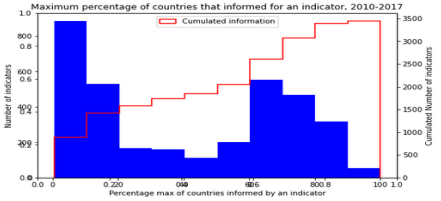
Information available all indicators and countries 1970-2017

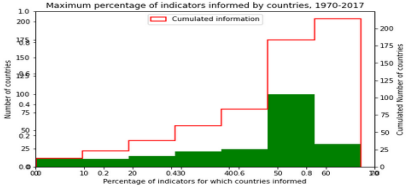


Indicators 1970-2017

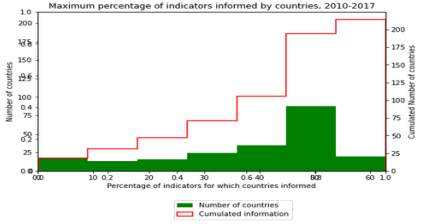


Indicators 2010-2017





Countries 1970-2017



Countries 2010-2017

D'après le graph général on conclut que la plupart des données disponibles datent de 2010 ou plus tard.

A partir des graphs indicateur et pays pour 1970-2017 :

* au tour de 2000 des indicateurs ne sont renseignées que par 60 % des pays ou moins, 1000 indicateurs sont renseignées par moins d'un 10 % des pays
* un peu plus de la moitié des pays renseignent au moins le 50 % des indicateurs.

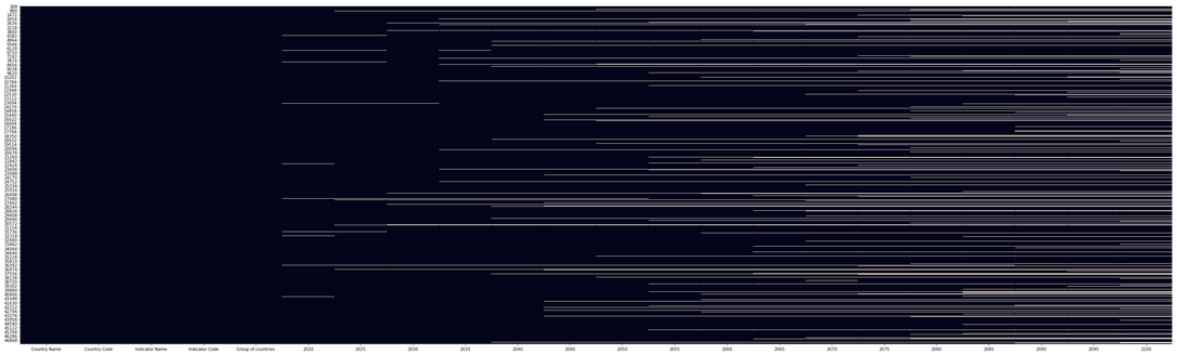
Si on compare les graphs 1970-2017 et 2010-2017 : profil général similaire (c'est normale parce que la plupart des données appartient à l'intervalle 2010-2017, on a enlevé peu d'information).

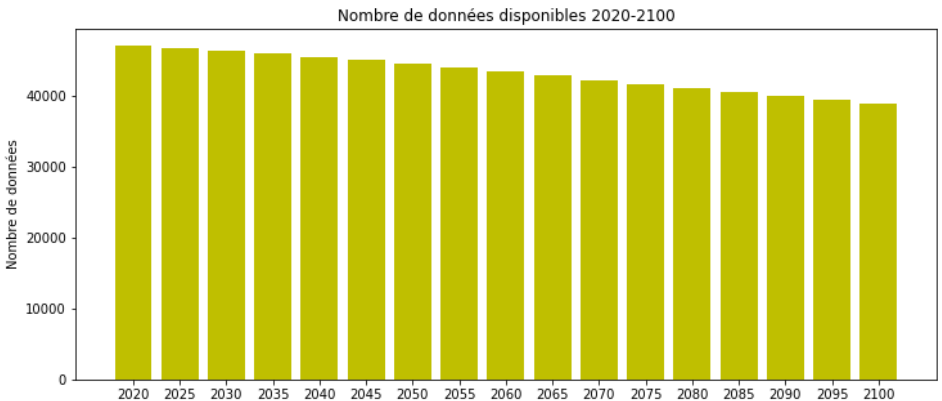
1. Données disponibles : Comprendre l'information

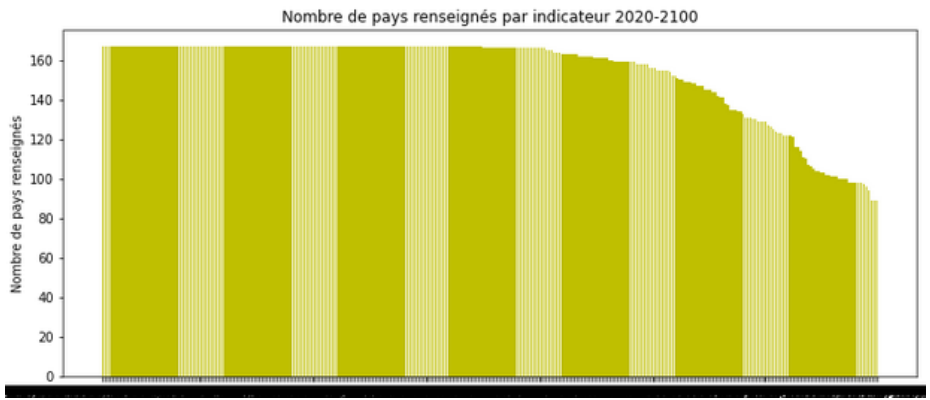
Taux remplissage 2020-2100

Indicateurs jamais renseignés éliminés

Groupements de pays éliminés

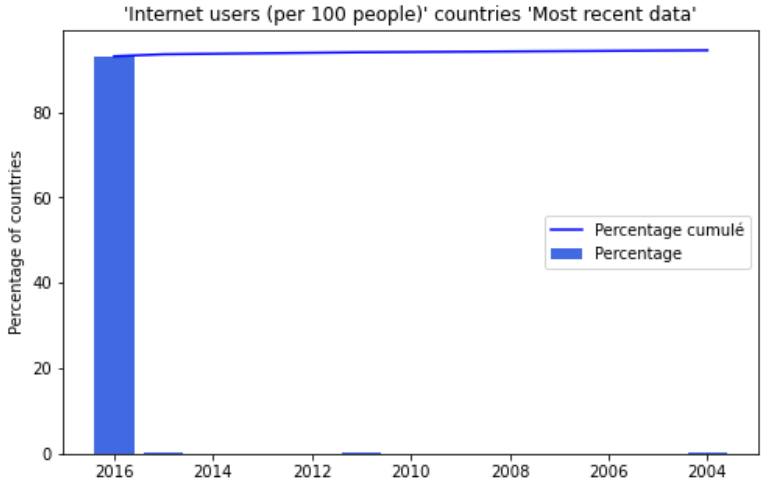


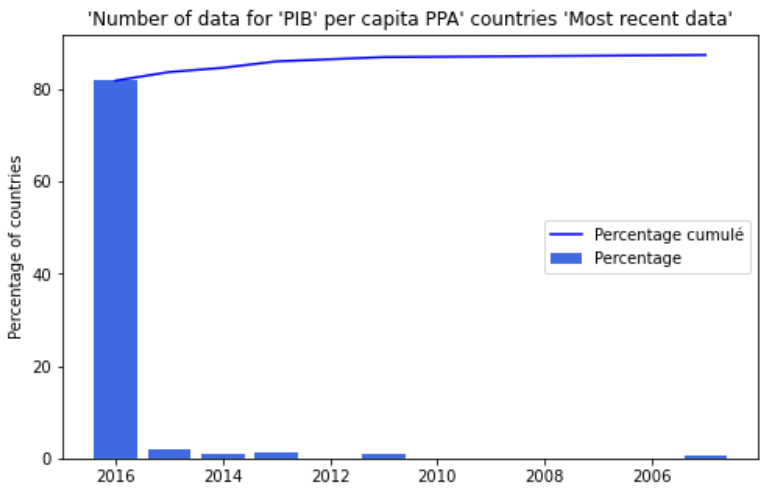


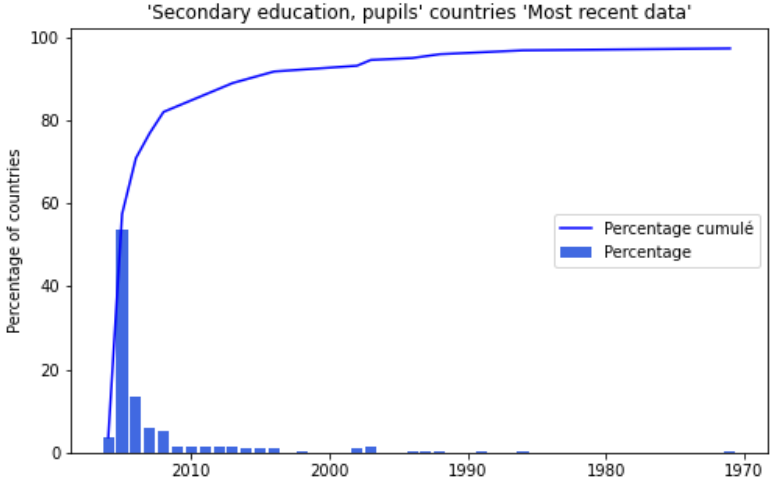


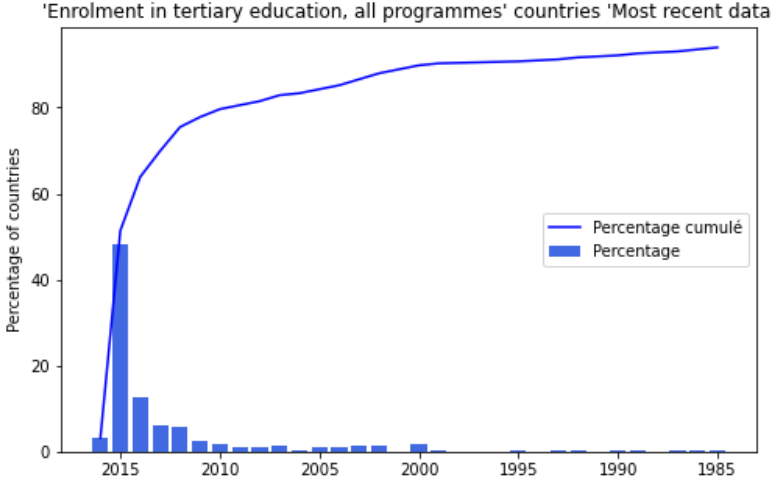
SOMMAIRE

2. Sélection d'information : Description



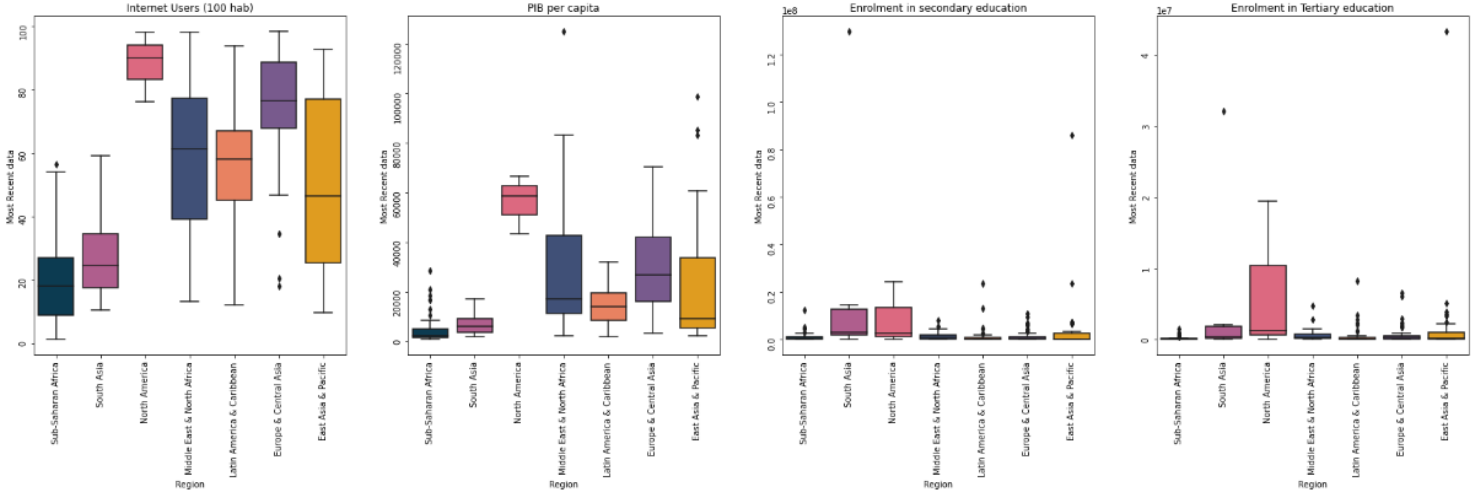






2. Sélection d'information : Description

Valeur de la donne la plus récente



Les **différences** les plus importantes entre les régions sont observées dans les valeurs des informations sur l'**accès à internet** et le **PIB per capita PPA**