



Analyse de données de systèmes éducatifs

MISSION



CONTEXTE

- **Formation en ligne** pour un public de niveau lycée et université
 - Le niveau sera interprété au sens large (pas uniquement sur la classe d'âge)
- Projet d'expansion à l'**international**

OBJECTIF

Aide à la décision pour trouver les pays à cibler en priorité

- Extraire un sous-ensemble de données à partir de l'existant fourni satisfaisant au besoin métier
- Analyse pré-exploratoire des indicateurs choisis
- Classement et évaluation de la pertinence

SOURCE DES DONNÉES



- Les données sont issues du site de la Banque Mondiale
<https://databank.worldbank.org/source/education-statistics-%5e-all-indicators>
- Ils concernent plusieurs thèmes autour de l'éducation

Secondary	Background	Pre-Primary	Education Management Information Systems (SABER)
Primary	Learning Outcomes	Infrastructure: Communications	Engaging the Private Sector (SABER)
Literacy	Early Childhood Education	Social Protection & Labor: Labor force structure	School Autonomy and Accountability (SABER)
Teachers	Tertiary	Health: Mortality	School Finance (SABER)
EMIS	Post-Secondary/Non-Tertiary	Health: Population: Dynamics	School Health and School Feeding (SABER)
Expenditures	Economic Policy & Debt: National accounts: US\$ at constant 2010 prices: Aggregate indicators	Population	Student Assessment (SABER)
Attainment	Economic Policy & Debt: National accounts: US\$ at current prices: Aggregate indicators	Health: Population: Structure	Teachers (SABER)
Education Equality	Economic Policy & Debt: Purchasing power parity	Health: Risk factors	Tertiary Education (SABER)
	Economic Policy & Debt: National accounts: Atlas GNI & GNI per capita	Early Child Development (SABER)	Workforce Development (SABER) Labor

STRUCTURE DU JEU DE DONNÉES INITIAL

Informations sur les indicateurs

	0	1
Series Code	BAR.NOED.1519.FE.ZS	BAR.NOED.1519.ZS
Topic	Attainment	Attainment
Indicator Name	Barro-Lee: Percentage of female population age ...	Barro-Lee: Percentage of population age 15-19 ...
Short definition	Percentage of female population age 15-19 with...	Percentage of population age 15-19 with no edu...
Long definition	Percentage of female population age 15-19 with...	Percentage of population age 15-19 with no edu...
Periodicity	Nan	Nan
Base Period	Nan	Nan
	0	1
Country Code	AFG	ALB
Country Name	Afghanistan	Albania
Income Group	Low income	Upper middle income
Region	South Asia	Europe & Central Asia
Other groups	HIPC	Nan
886929		
Country Name	Zimbabwe	Pakistan
Country Code	ZWE	
Indicator Name	Youth literacy rate, population 15-24 years, m...	Yc
Indicator Code	SE.ADT.1524.LT.MA.ZS	
1970	Nan	
...	...	
2080	Nan	
2085	Nan	

Données chiffrées
(pays, indicateurs,
années)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3665 entries, 0 to 3664
Data columns (total 15 columns):
 #   Column
 ---  -----
 0   Series Code
 1   Topic
 2   Indicator Name
 3   Short definition
 4   Long definition
 5   Periodicity
 6   Base Period
 7   Other notes
```

#	Column	Non-Null Count	Dtype
0	Series Code	3665 non-null	object
1	Topic	3665 non-null	object
2	Indicator Name	3665 non-null	object
3	Short definition	2156 non-null	object
4	Long definition	3665 non-null	object
5	Periodicity	99 non-null	object
6	Base Period	214 non-null	object
7	Other notes	214 non-null	object

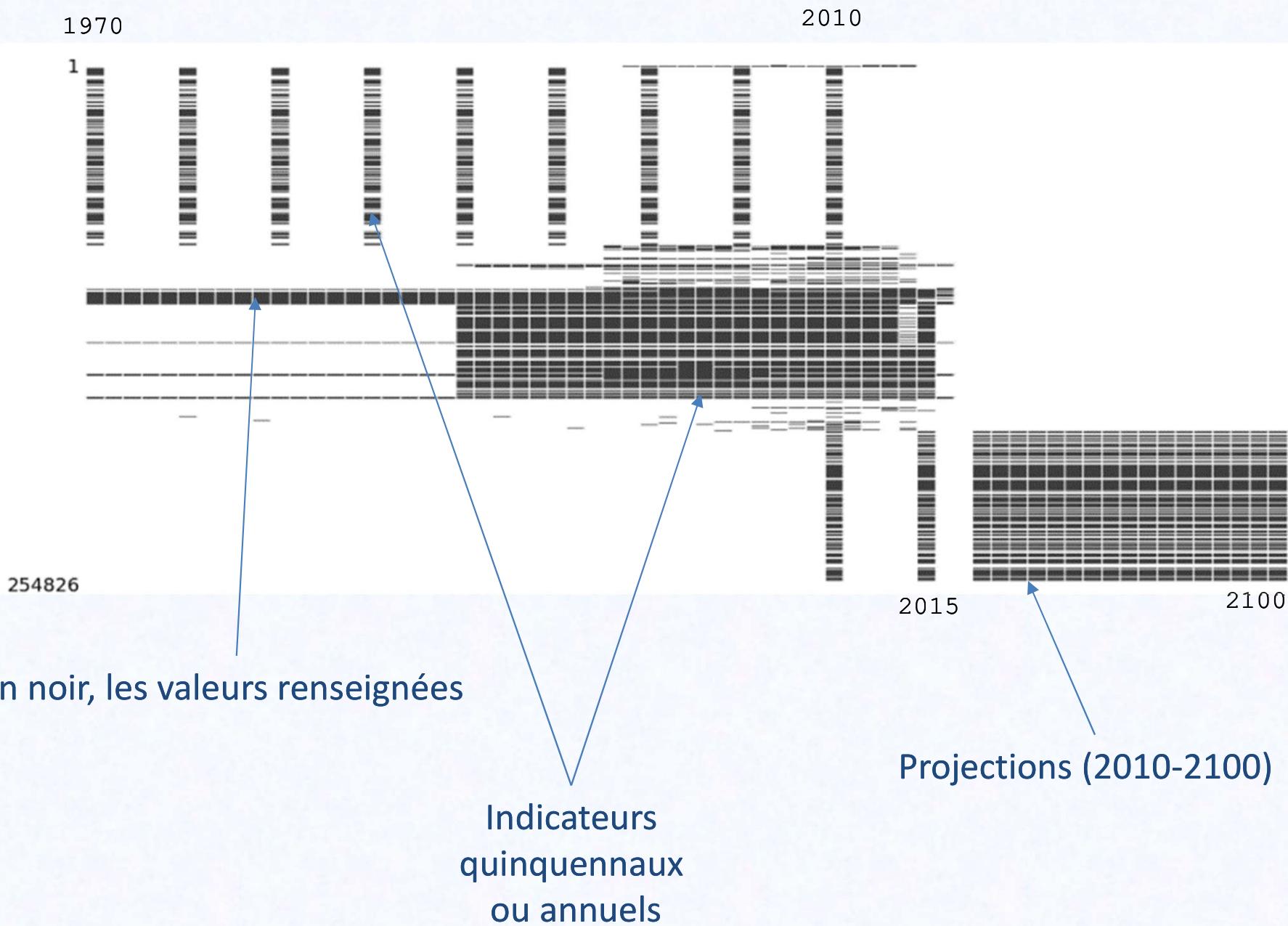
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 217 entries, 0 to 216
Data columns (total 6 columns):
 #   Column
 ---  -----
 0   Country Code
 1   Country Name
 2   Income Group
```

#	Column	Non-Null Count	Dtype
0	Country Code	217 non-null	object
1	Country Name	217 non-null	object
2	Income Group	217 non-null	object

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 886930 entries, 0 to 886929
Data columns (total 70 columns):
 #   Column
 ---  -----
 0   Country Name
 1   Country Code
 2   Indicator Name
 3   Indicator Code
 4   1970
 5   1971
 6   1972
 7   1973
 8   1974
 9   1975
```

#	Column	Non-Null Count	Dtype
0	Country Name	886930 non-null	object
1	Country Code	886930 non-null	object
2	Indicator Name	886930 non-null	object
3	Indicator Code	886930 non-null	object
4	1970	72288 non-null	float64
5	1971	35537 non-null	float64
6	1972	35619 non-null	float64
7	1973	35545 non-null	float64
8	1974	35730 non-null	float64
9	1975	87306 non-null	float64

QUALITÉ DES DONNÉES INITIALES



CHOIX DES INDICATEURS



- Sélection des thèmes
 - situation économique, investissement dans l'éducation, accès à internet, structure de la population et niveaux scolaires
- Sélection des indicateurs
 - Parmi les thèmes sélectionnés
 - En maximisant le nombre de pays renseignés
 - Tout en conservant la pertinence métier
- Pour évaluer
 - La richesse et le développement technologique des pays par rapport à leur population
 - Les efforts des pays concernant l'éducation (généraux et relativement à leurs moyens)
 - Le nombre d'individus potentiellement concernés par l'activité d'Academy
 - La combinaison de ces éléments pour effectuer un classement (score)

INDICATEURS CALCULÉS

Barro-Lee: Percentage of population age 15+ with secondary schooling. Total (Incomplete and Completed Secondary)



Percentage of population age 15+ with tertiary schooling. Total (Incomplete and Completed Tertiary)

Percentage of population age 15-19 with secondary schooling. Total (Incomplete and Completed Secondary)

Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Total (Incomplete and Completed Tertiary)

GDP per capita, PPP (current international \$)



GDP, PPP (current \$)



Government expenditure on education as % of GDP (%)

Government expenditure on secondary education as % of GDP (%)

Government expenditure on tertiary education as % of GDP (%)

Government expenditure per secondary student as % of GDP per capita (%)

Government expenditure per tertiary student as % of GDP per capita (%)



Internet users (per 100 people)



Population, age 15, total

Population, age 16, total

Population, age 17, total

Population, age 18, total

Population, age 19, total

Population, age 20, total

Population, age 21, total

Population, age 22, total

Population, age 23, total

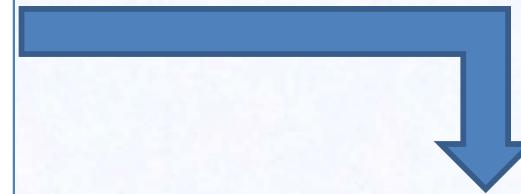
Population, age 24, total

Population, ages 0-14, total

Population, ages 15-64, total

Population, total

Des indicateurs plus ciblés sont calculés afin d'affiner les proportions et les nombres d'individus



Population age 15+ with secondary schooling (#)

Population age 15+ with tertiary schooling (#)

Population age 15-19 with secondary schooling (#)

Population age 20-24 with tertiary schooling (#)

Government expenditure on education (\$)

Government expenditure on secondary education (\$)

Government expenditure on tertiary education (\$)

Government expenditure per secondary student (\$)

Government expenditure per tertiary student (\$)

Internet users age 15+ with secondary schooling (%),

Internet users age 15+ with tertiary schooling (%),

Internet users age 15-19 with secondary schooling (%),

Internet users age 20-24 with tertiary schooling (%),

Internet users age 15+ with secondary schooling (#),

Internet users age 15+ with tertiary schooling (#),

Internet users age 15-19 with secondary schooling (#),

Internet users age 20-24 with tertiary schooling (#)



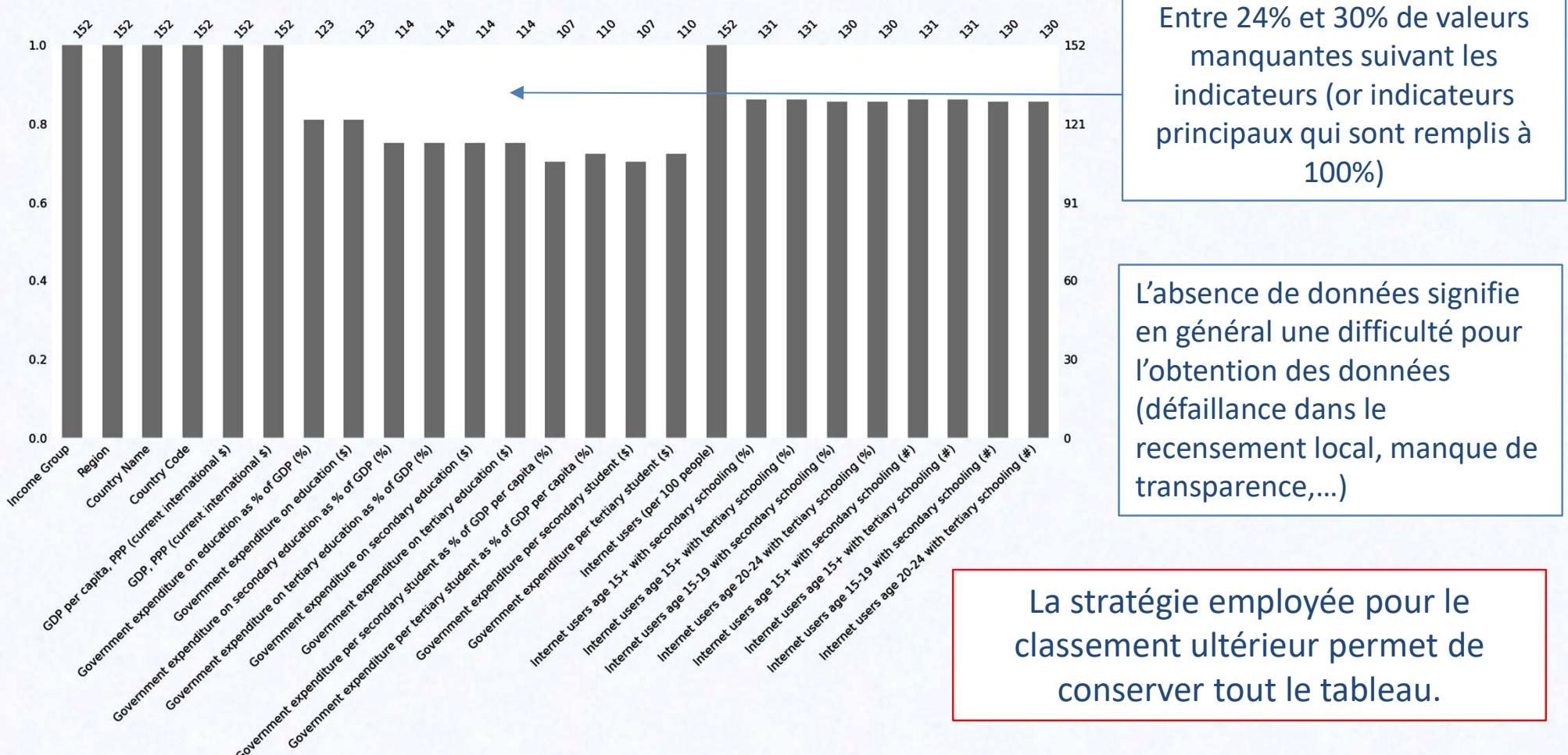
PRÉ-EXPLORATION DES DONNÉES

AIDE À LA DÉCISION
&
PERTINENCE



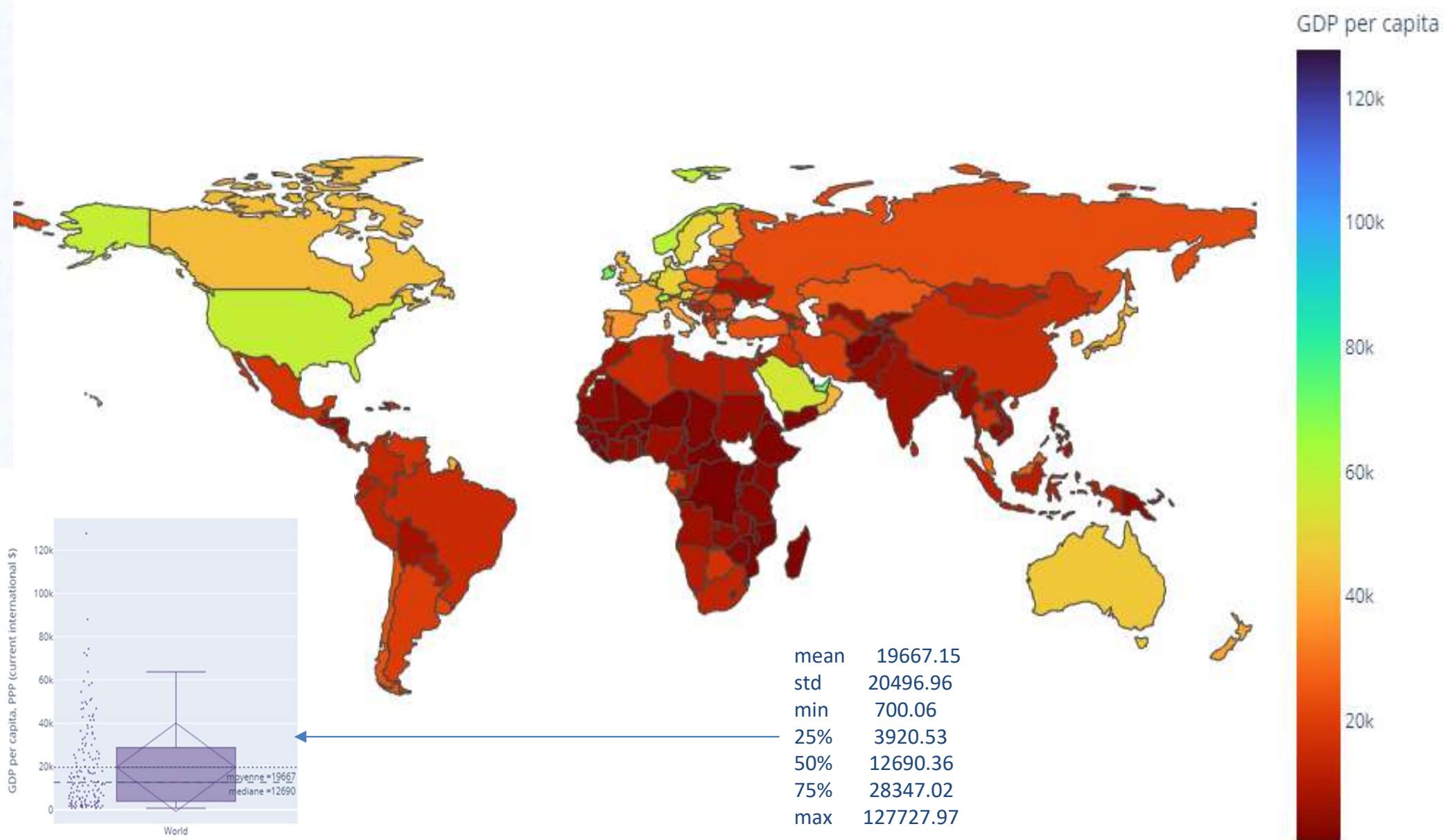
JEU DE DONNÉES FINAL

- Prise en compte de la dernière valeur renseignée pour les indicateurs choisis, avec la limite de l'année 2010 (abstraction faite des projections 2010-2100)
- Filtrage où on a éliminé les pays non pertinents
 - Par leur taille (trop petit < 1 000 000 habitants)
 - Où un indicateur fondamental n'est pas renseigné (e.g. utilisation d'Internet, situation économique)
- Sélection des indicateurs les plus précis (calculés)

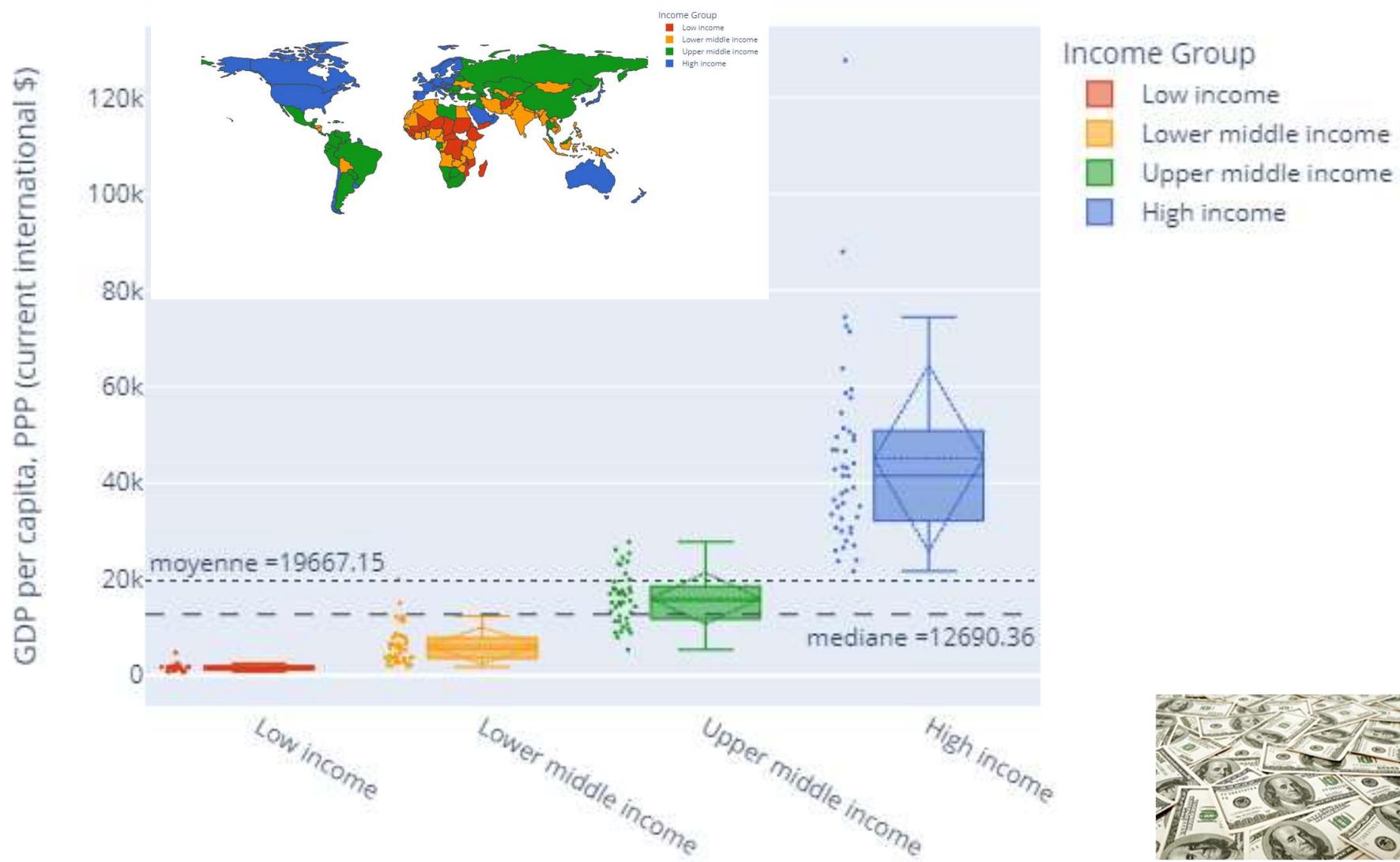




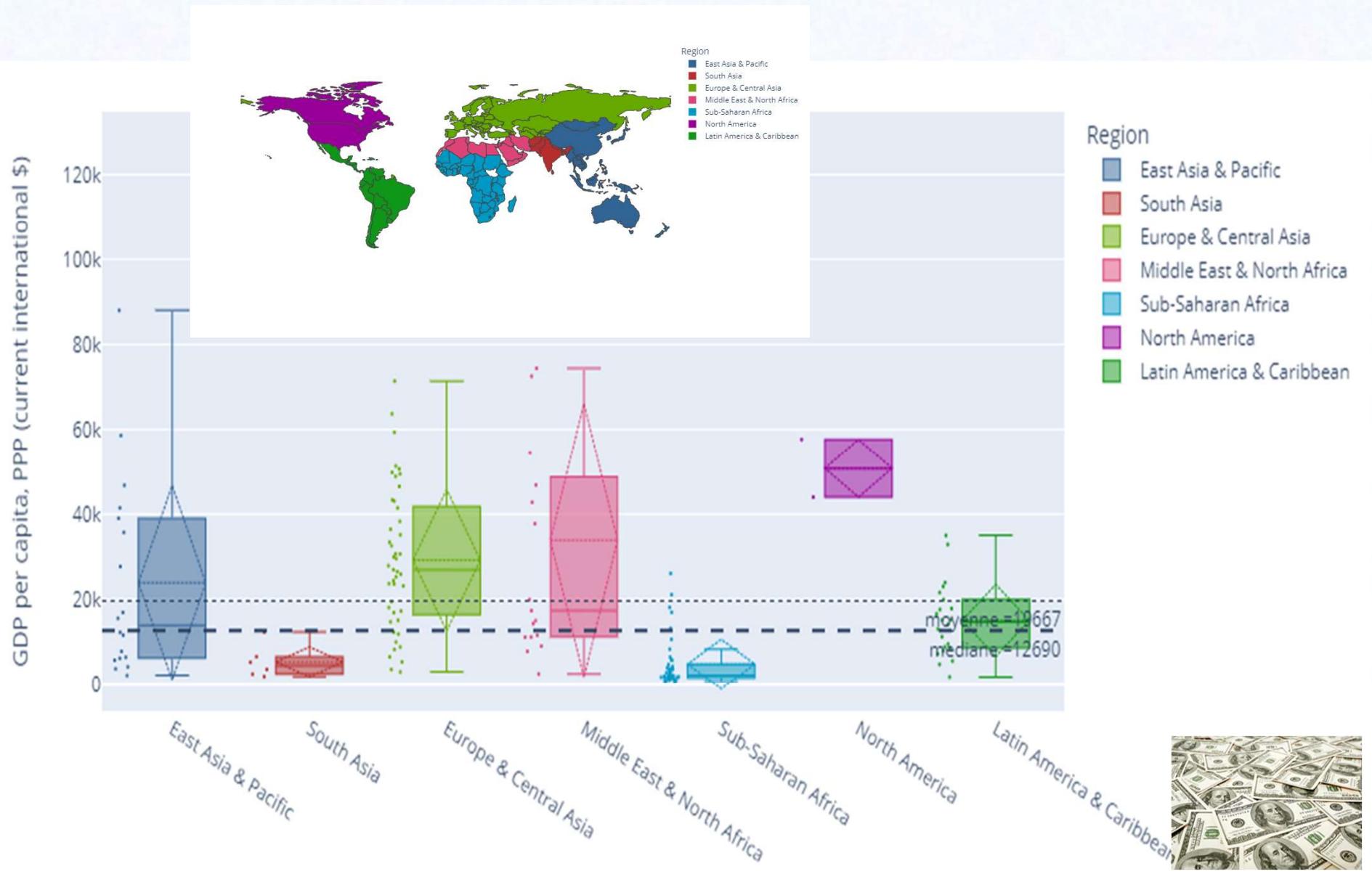
SITUATION ÉCONOMIQUE PIB/HABITANT – VUE MONDIALE



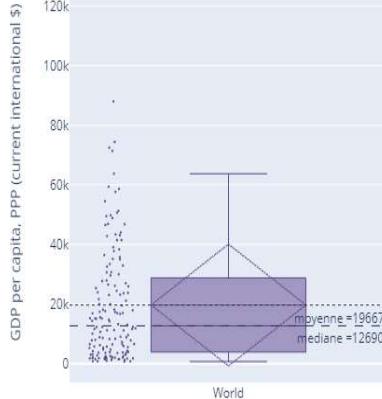
SITUATION ÉCONOMIQUE PIB/HABITANT – VUE PAR INCOME GROUP



SITUATION ÉCONOMIQUE PIB/HABITANT – VUE PAR RÉGION GÉOGRAPHIQUE

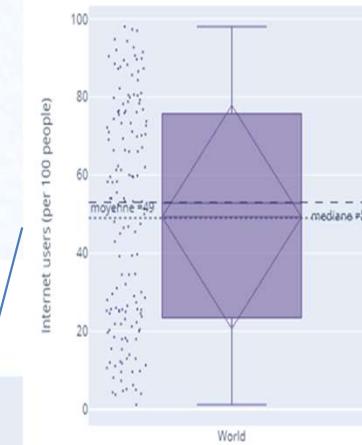


TAUX D'UTILISATION D'INTERNET ET PIB/HABITANT

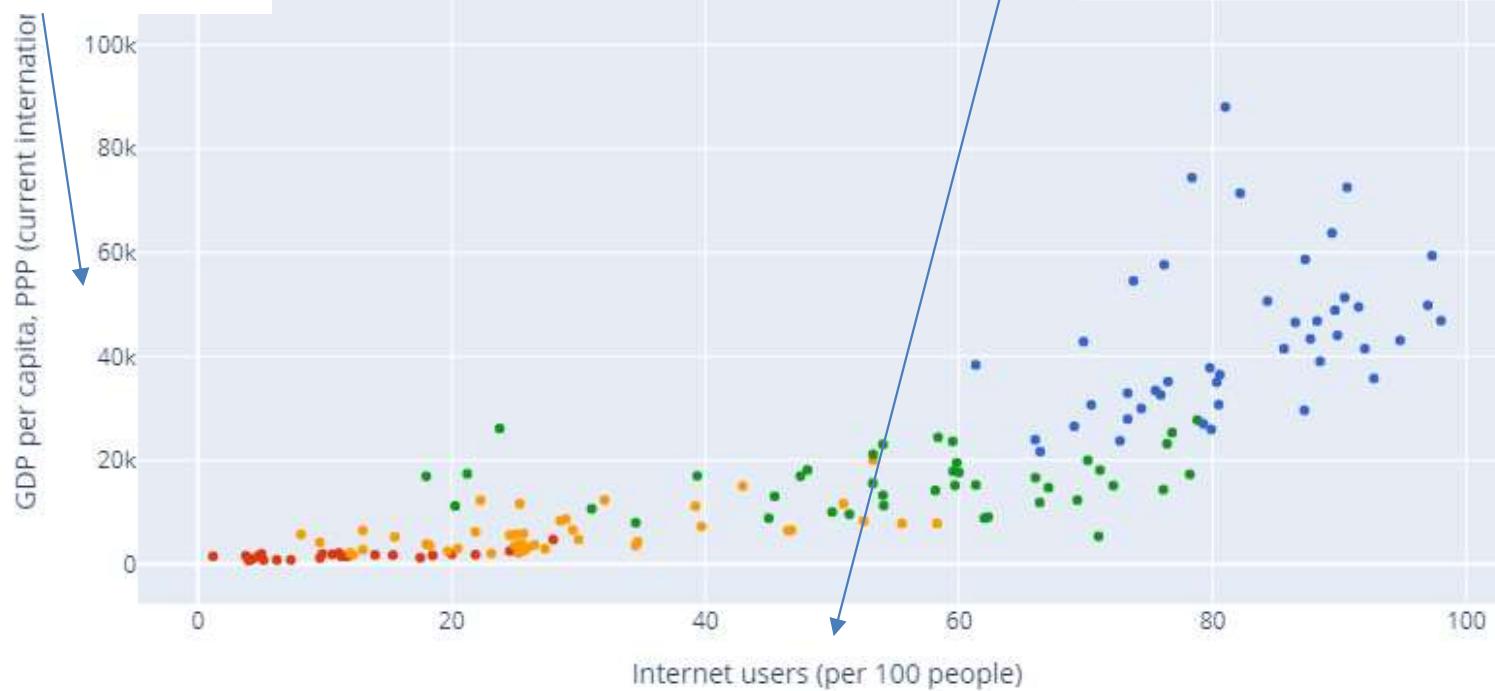


mean 19667.15
std 20496.96
min 700.06
25% 3920.53
50% 12690.36
75% 28347.02
max 127727.97

On peut observer une corrélation directe entre la richesse d'un pays et le pourcentage d'internautes



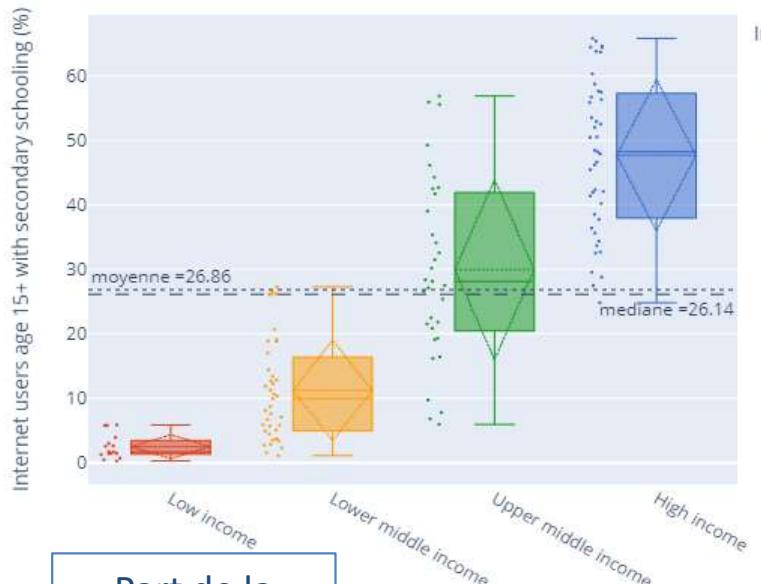
mean 49.26
std 28.64
min 1.18
25% 23.61
50% 52.84
75% 75.60
max 98.00



- Income Group
- Low income
 - Lower middle income
 - Upper middle income
 - High income



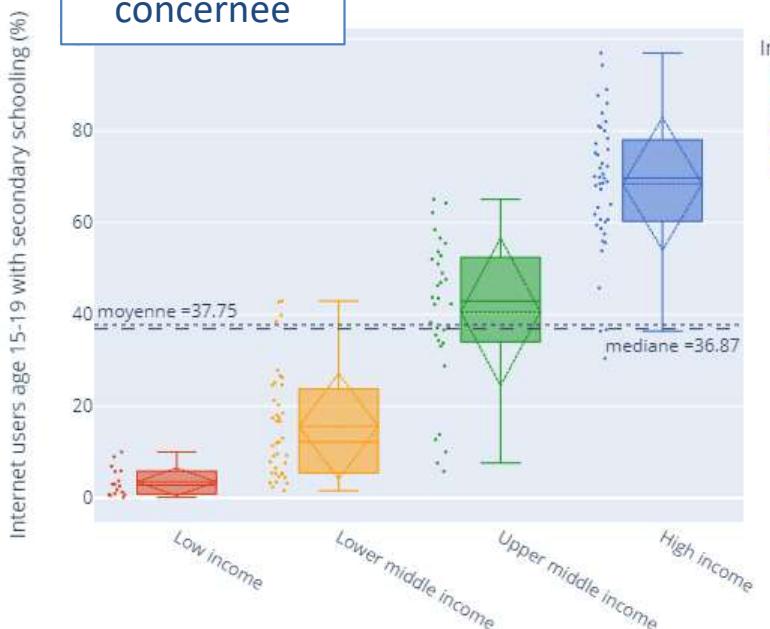
POPULATION AYANT AU MOINS DÉBUTÉ UN CYCLE SECONDAIRE ET UTILISANT INTERNET



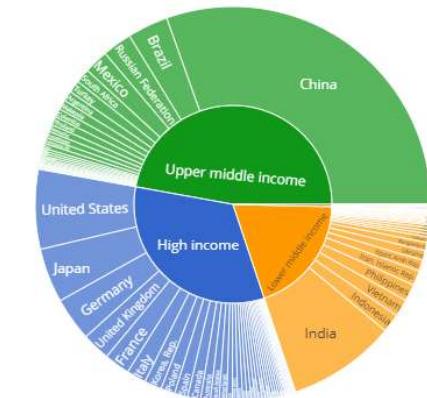
Chez les +15 ans
(tous âges)



Chez les
15-19 ans

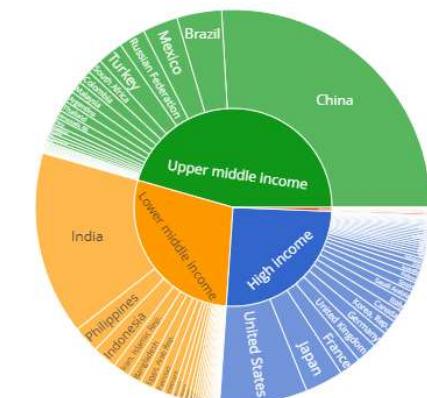


Internet users age 15+ with secondary schooling (#)

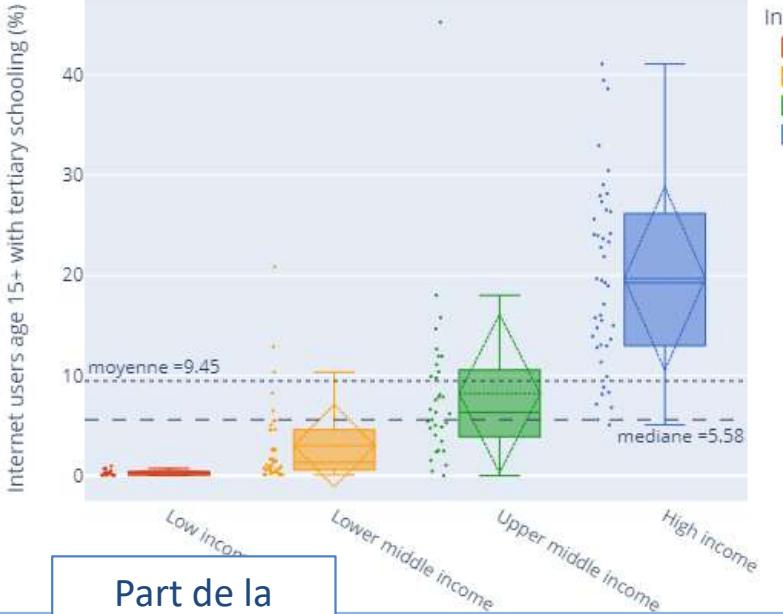


Répartition des effectifs entre les pays

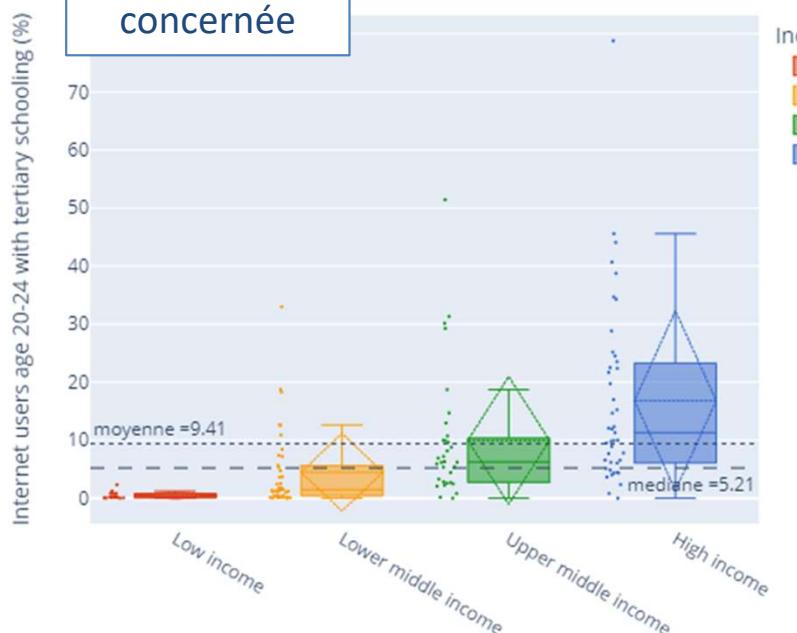
Internet users age 15-19 with secondary schooling (#)



POPULATION AYANT AU MOINS COMMENCÉ DES ÉTUDES SUPÉRIEURES ET UTILISANT INTERNET



Part de la population concernée

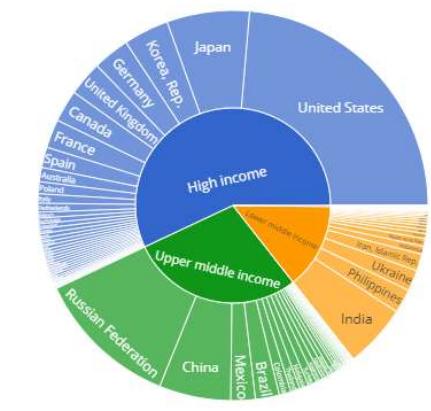


Chez les +15 ans
(tous âges)



Chez les
20-24 ans

Internet users age 15+ with tertiary schooling (#)

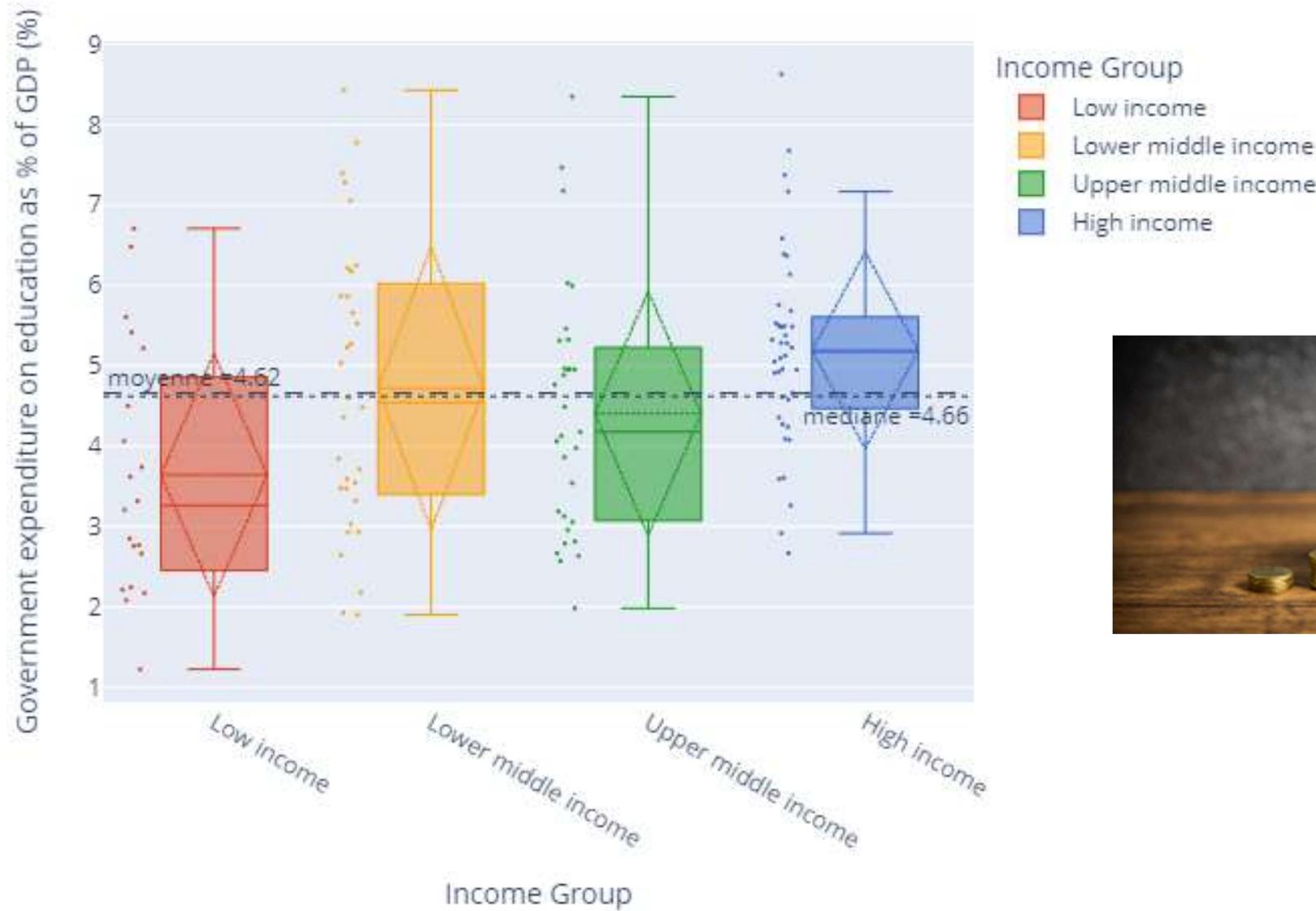


Répartition des effectifs entre les pays

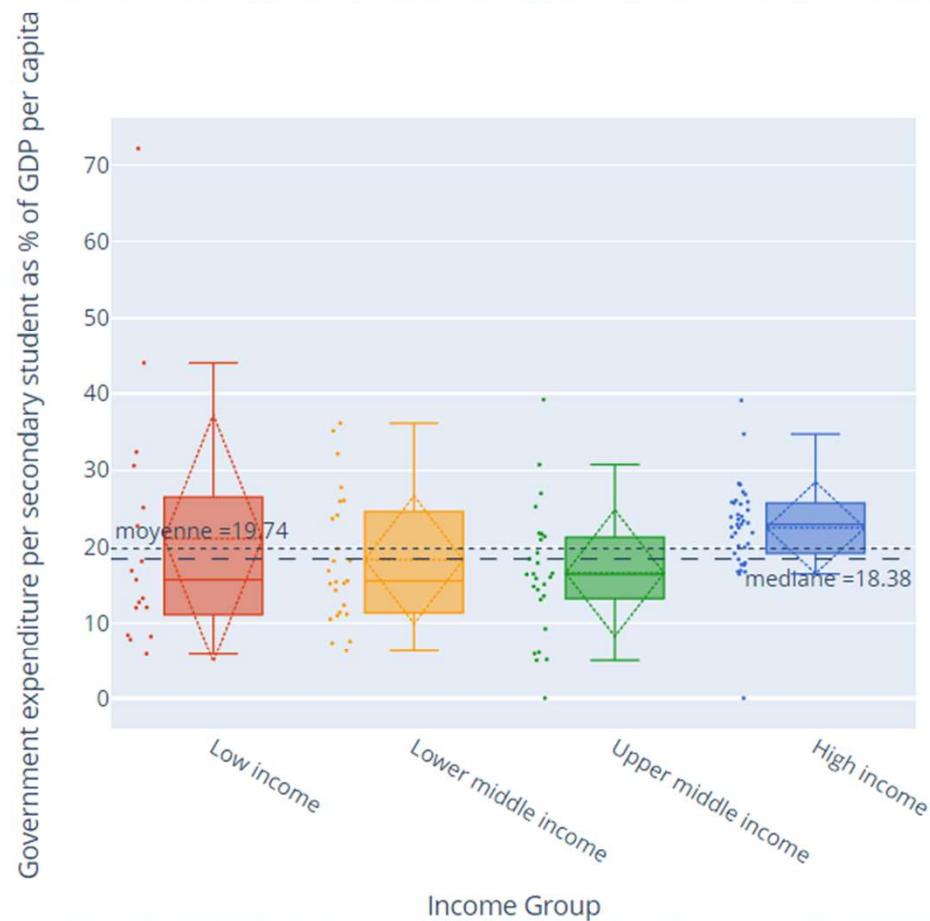
Internet users age 20-24 with tertiary schooling (#)



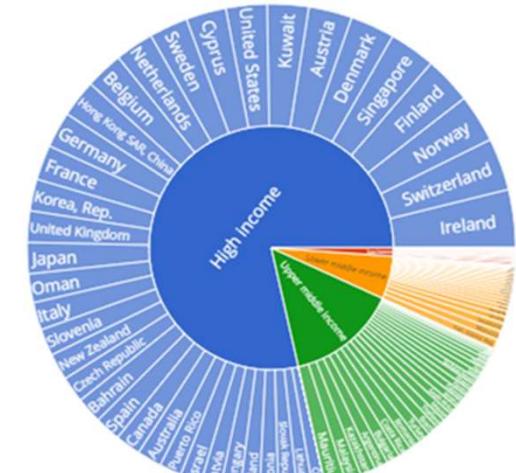
PART DU PIB DANS DÉPENSÉ DANS L'ÉDUCATION



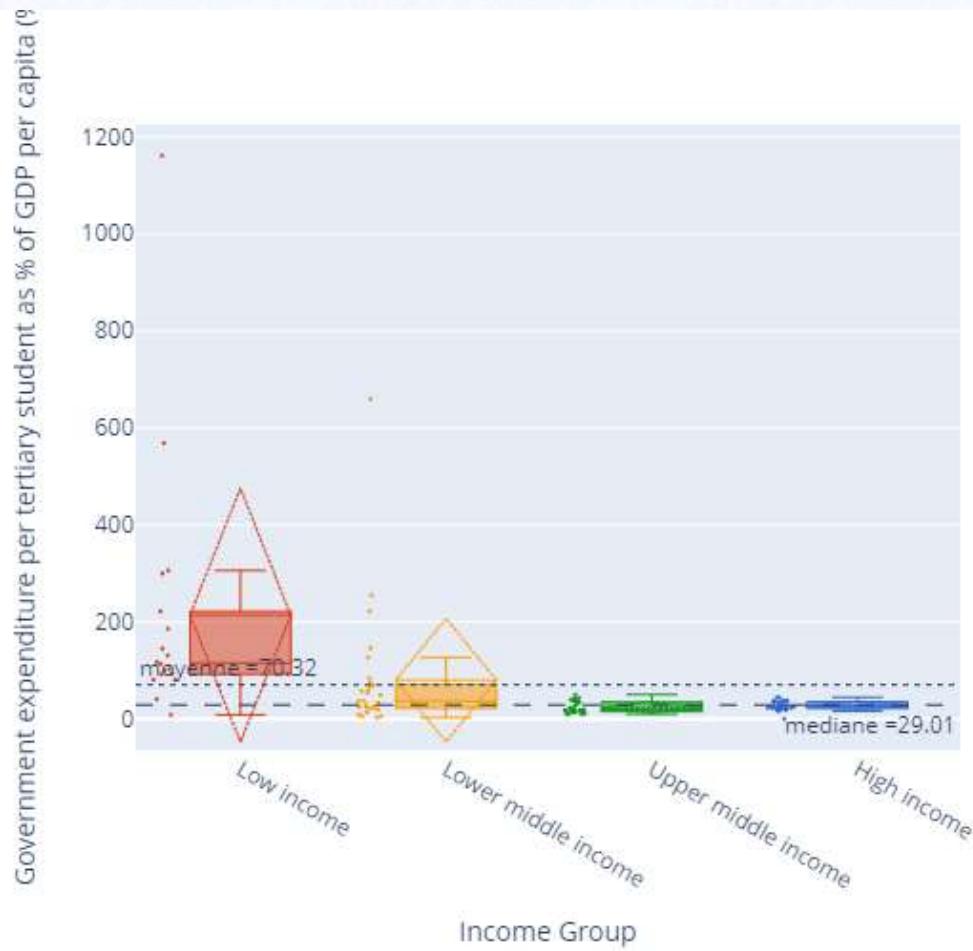
DÉPENSE PAR ÉTUDIANT DU SECONDAIRE, TOTALE ET EN PART DU PIB/HABITANT



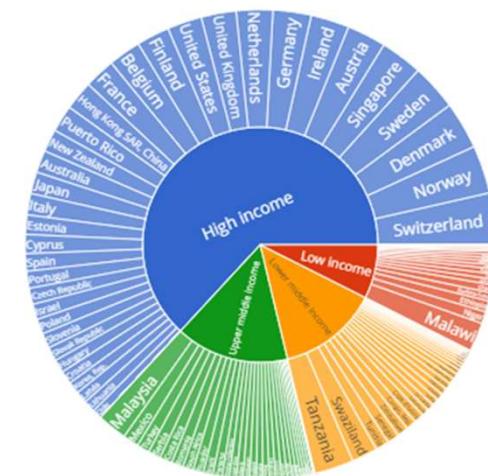
Government expenditure per secondary student (\$)



DÉPENSE PAR ÉTUDIANT DU CYCLE SUPÉRIEUR, TOTALE ET EN PART DU PIB/HABITANT



Government expenditure per tertiary student (\$)

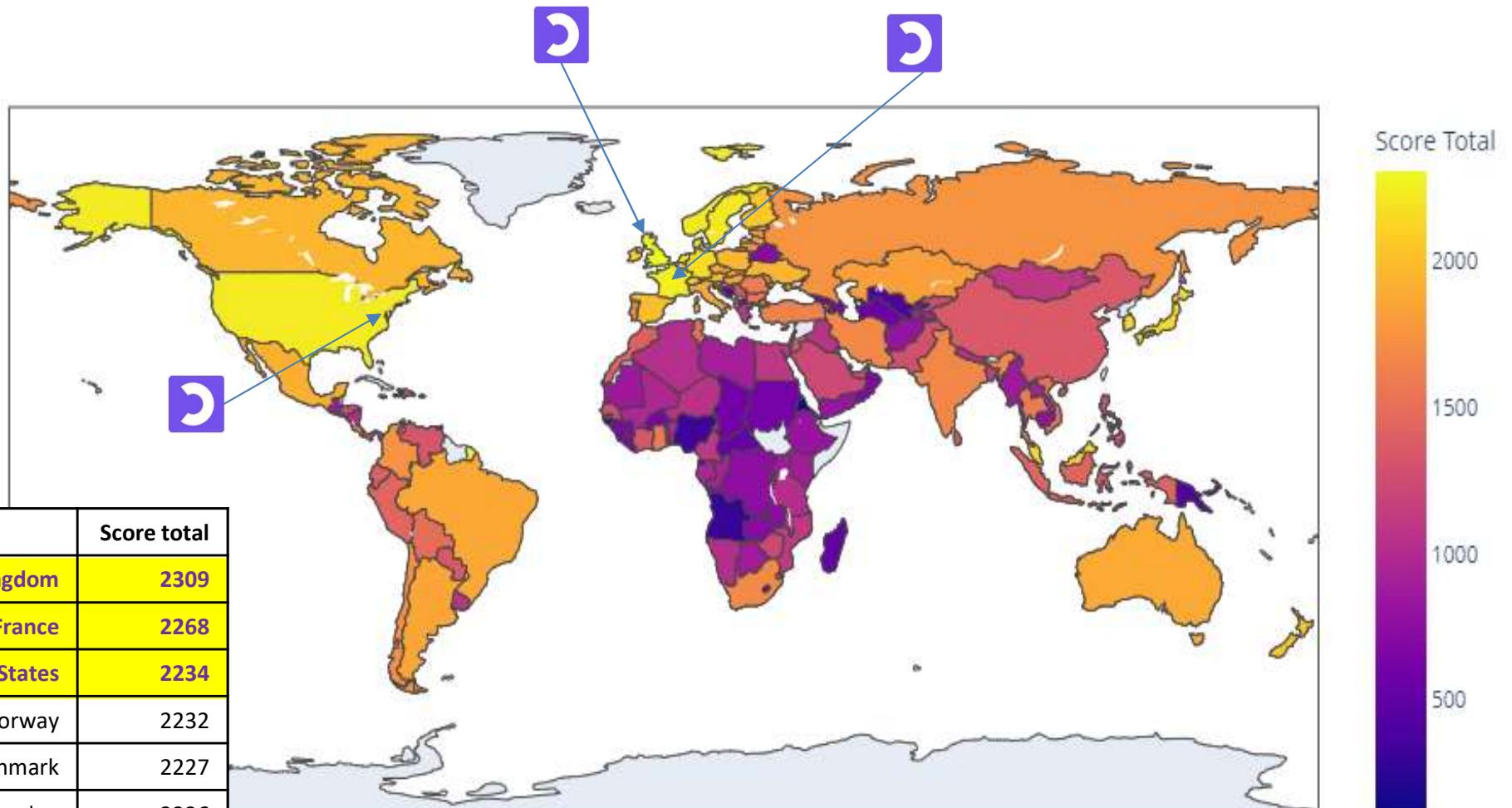


SCORE D'ATTRACTIVITÉ

- Le score du pays pour chaque indicateur est son rang dans le classement pour cet indicateur (0 si l'indicateur n'est pas renseigné)
- Les pays où des données manquent sont donc pénalisés
- Le score final est la somme des scores.
- On classe ensuite suivant le score final.

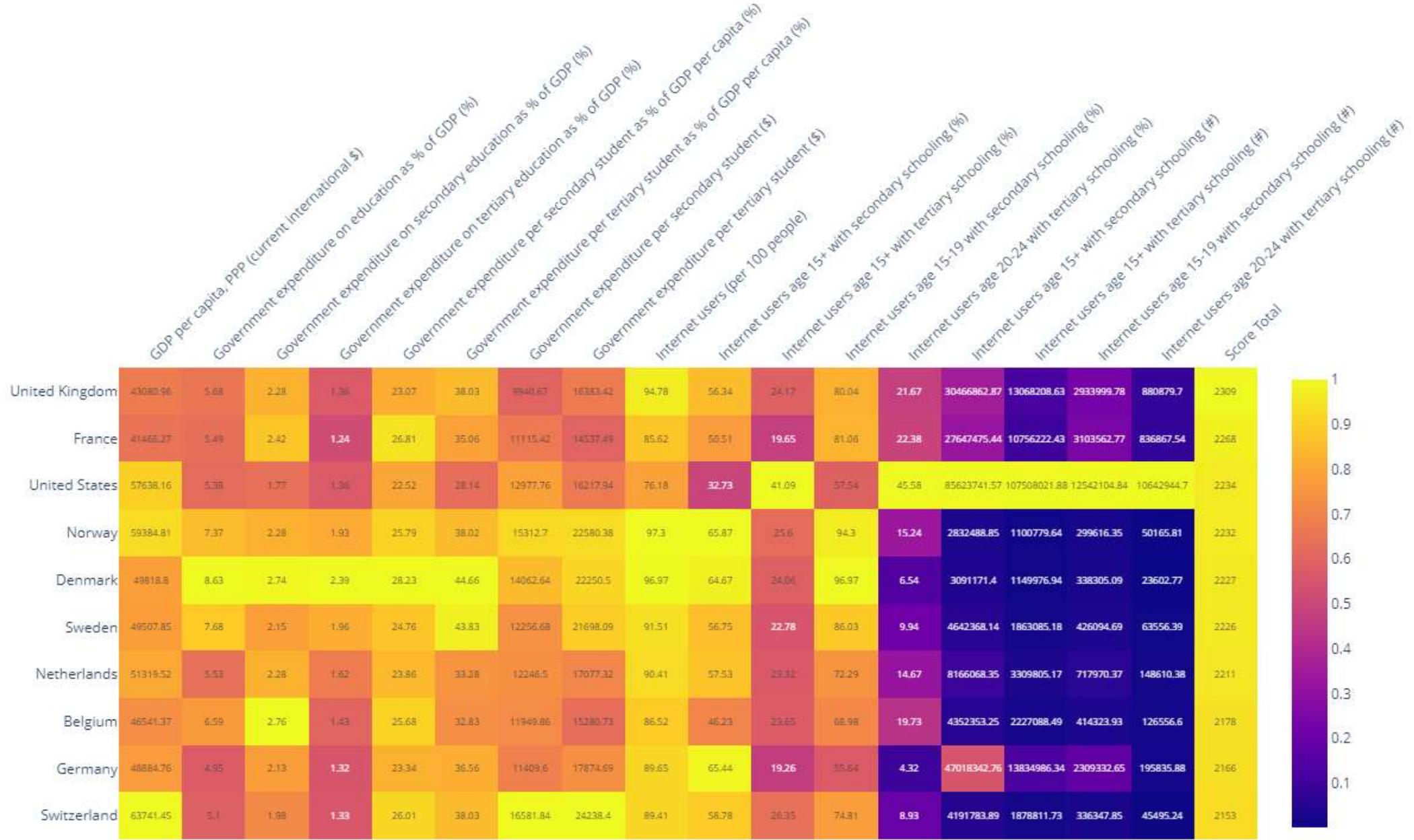
Aucun poids n'a été attribué mais c'est une possibilité qui dépend de l'objectif poursuivi.

LES PAYS LES MIEUX CLASSÉS et les implantations d'OpenClassrooms



Coïncidence ?

DÉTAIL POUR CHAQUE INDICATEUR (TOP 10)



PERTINENCE ET PERSPECTIVE D'ÉVOLUTION

- Instabilité mondiale 
- Manque d'éléments sur ...
 - L'état du marché dans le domaine de la formation en ligne 
 - L'évolution des investissements dans l'éducation
 - Les parcours scolaires et professionnels (financement des études et des reconversions)
 - Le développement des infrastructures réseaux pour les pays moins bien lotis

MERCI
DE
VOTRE ATTENTION

