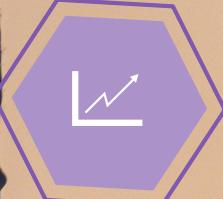


Soutenance Projet 2
OpenClassrooms
Formation Data Scientist

25 Octobre 2022

Analysez des données de systèmes éducatifs



Plan de la soutenance

1. Introduction :

- ✿ Problématique
- ✿ Présentation jeux de données

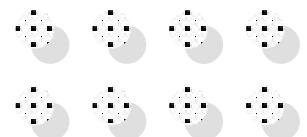
2. Analyse pré-exploratoire

- ✿ Nettoyage jeu de données
- ✿ Analyse des données

3. Conclusion



Illustrations by Pixeltrue on
[icons8](#)



Mission :



Données de
qualité suffisante ?



Photo by Gaël Gaborel - OrbisTerrae on Unsplash



3 fichiers :



Statistiques

Individu : pays et indicateur

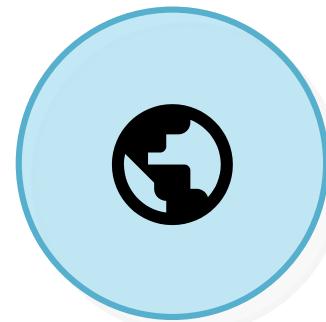
Lignes : 795 305
Colonnes: 69



Indicateurs

Individu : indicateur

Lignes : 3 665
Colonnes : 21



Pays

Individu : pays

Lignes : 241
Colonnes: 32

Plan de la soutenance

- ✓ Introduction :
 - ❖ Problématique
 - ❖ Présentation jeux de données

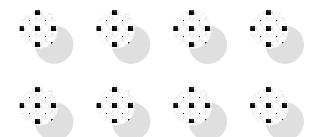
1. Analyse pré-exploratoire

- ❖ Nettoyage jeu de données
- ❖ Analyse des données

2. Conclusion



Illustrations by Pixeltrue on
icons8



01 Filtre

Indicateurs en lien
avec
problématique

02 Analyse

Qualité des
données

03 Impute

Données
manquantes

Academy



Indicateurs à filtrer : 3 665

Academy



Photo by Matthew Henry on Unsplash



Indicateurs par thème

Résultats d'apprentissage

30%

20%

Niveau de scolarité



Indicateurs à filtrer : 93



Academy

Photo by Joe Green on Unsplash

1970



Années : 65



Années : 35



Dernière année
avec valeurs

Indicateurs
(%)

2100

4 %

2009-2017

96 %

Année

Valeurs manquantes
(%)

2017

99%

2016

98%

Années : 16





« Cours en ligne »

Valeurs manquantes

Utilisateurs internet 7 %

Ordinateur individuel 64 %

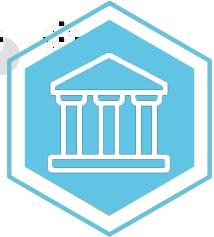


« niveau lycée + »

Valeurs manquantes

Élèves dans secondaire 30 %

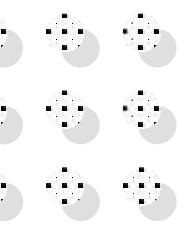
Élèves dans tertiaire 39 %



« priorité publique »

Valeurs manquantes

Dépenses publiques
dans l'éducation
(en % du PIB) 48 %



Utilisateurs internet

Ordinateur individuel

Élèves dans secondaire

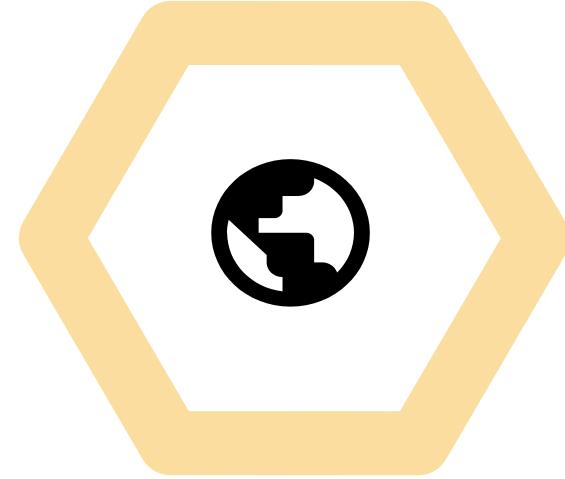
Élèves dans tertiaire

Dépenses publiques dans l'éducation
(en % du PIB)

Pays : 214



Pays : 191



Filtre : + 70 % valeurs manquantes



01 Quantile

0.9

Valeurs manquantes
consécutives

03

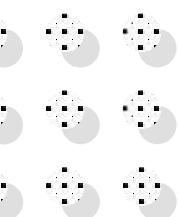
Indicateur
suivant

02 Interpolation linéaire



Academy





Valeurs manquantes consécutives

0 . . . • 1 . . . • 7 < 8 < 8 . . . 16

90 % pays

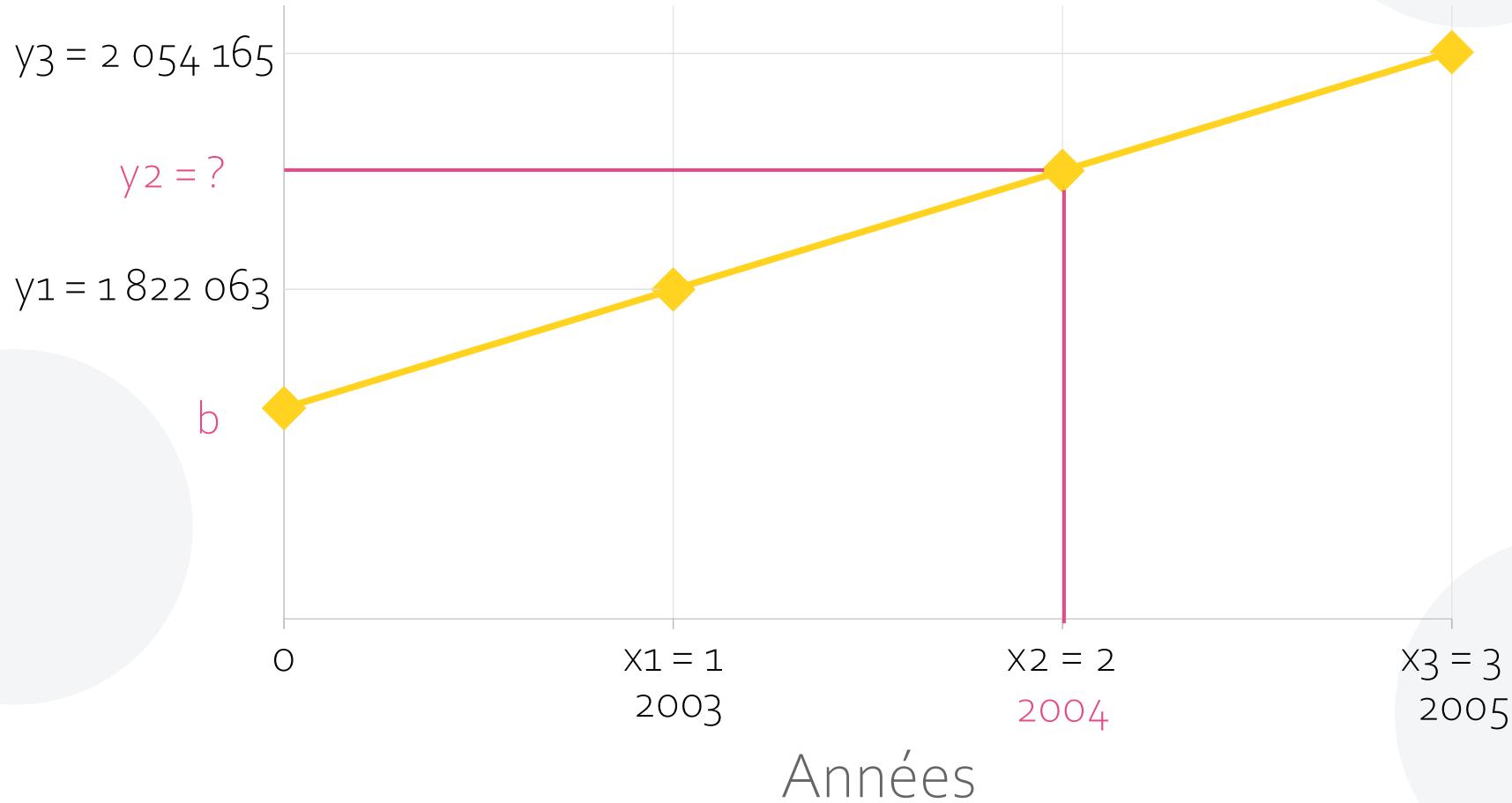
Quantile 0.9

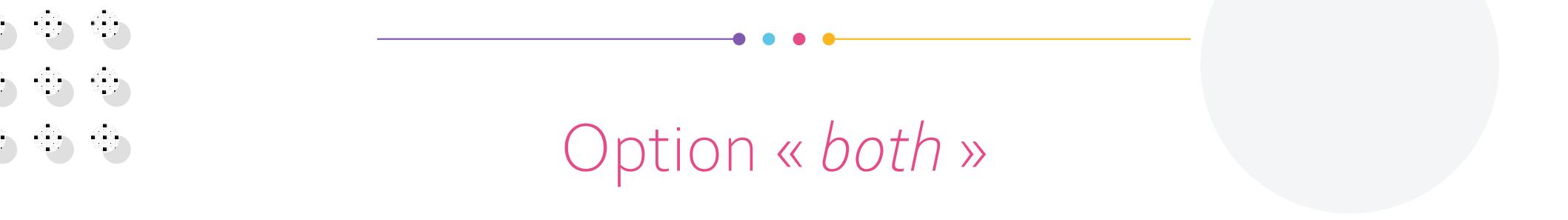
10 % pays



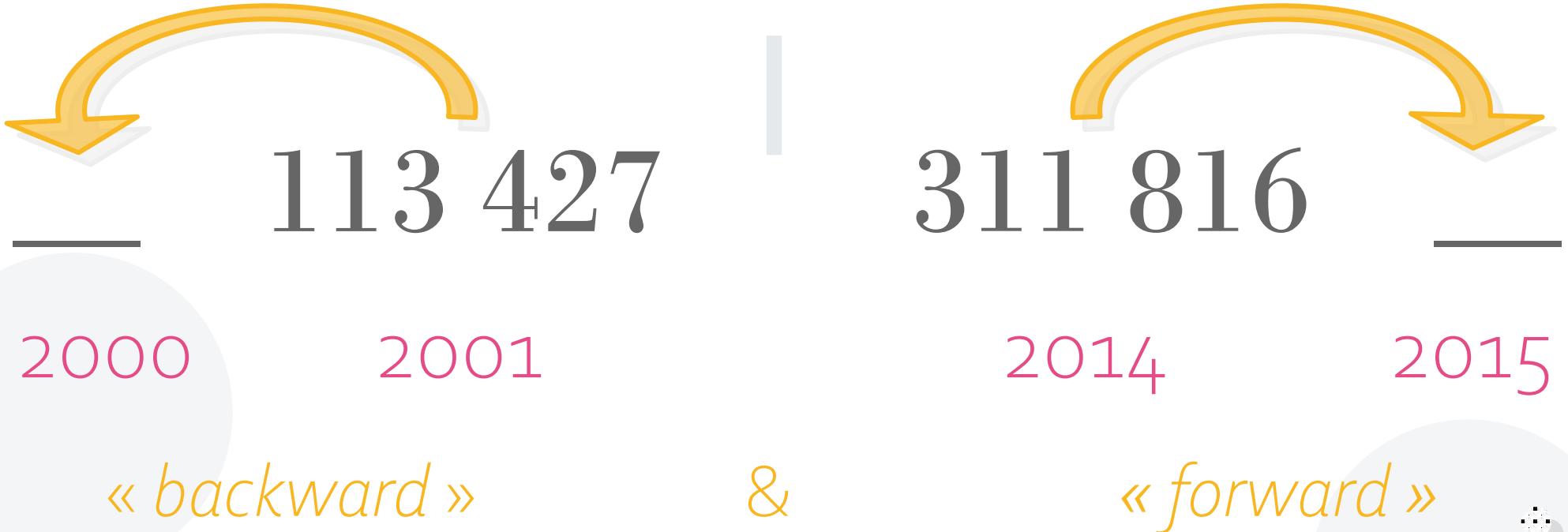
Élèves dans secondaire

$$y = a*x + b$$

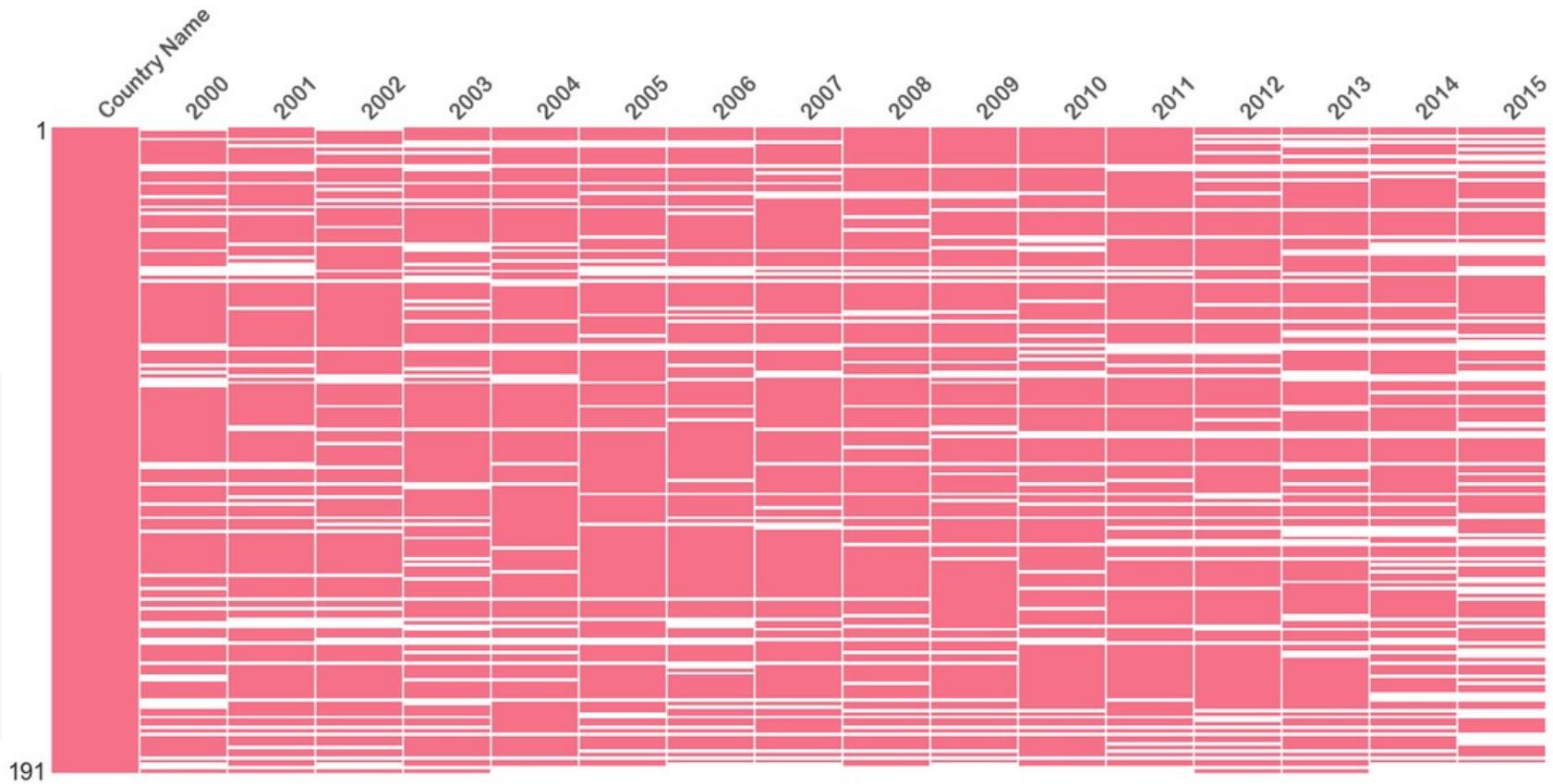




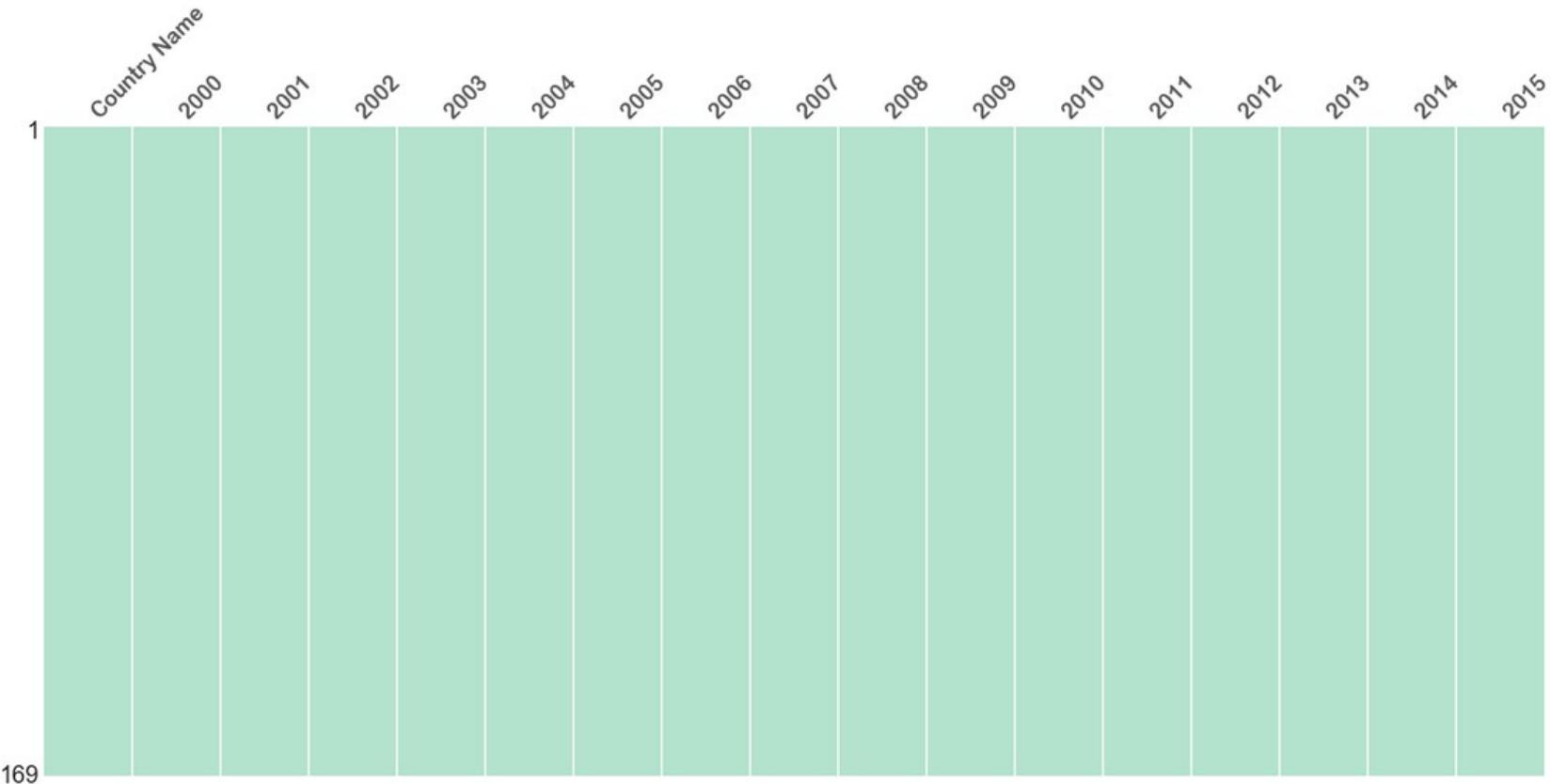
Option « both »



Valeurs manquantes : 23%



Valeurs manquantes : 0%

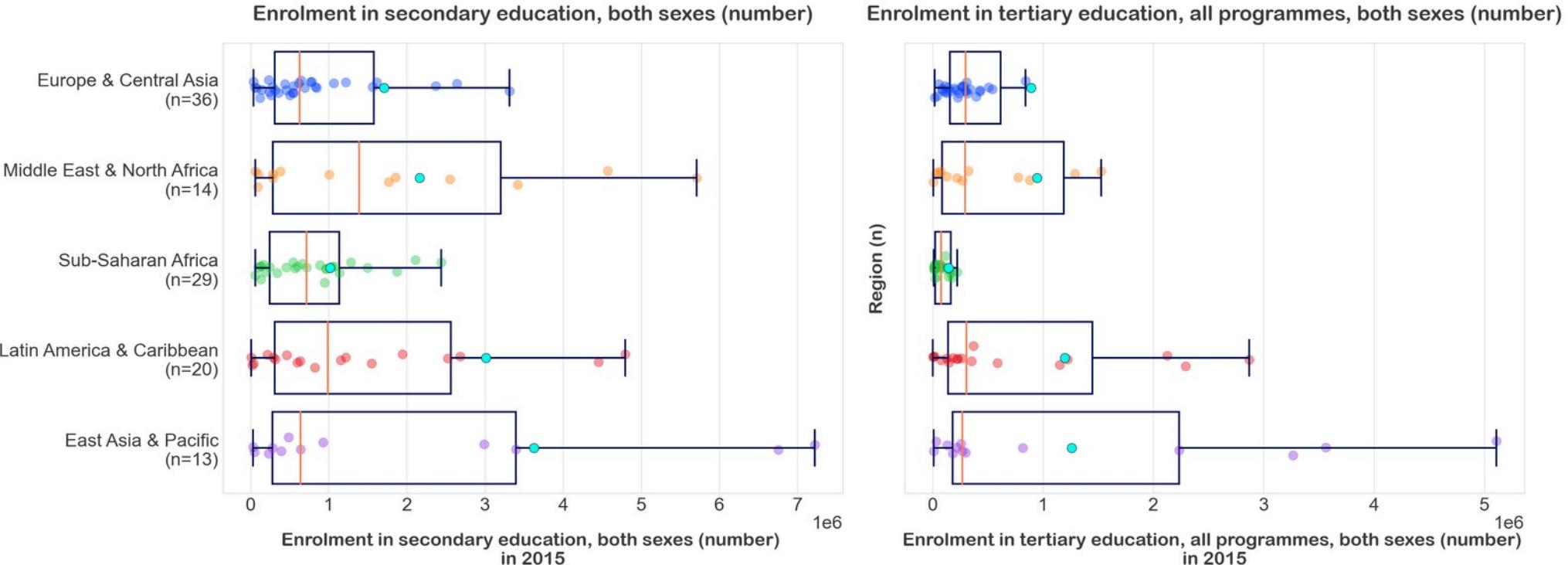


Limites démarche



2015

Élèves par région (en millions)



Classement pays



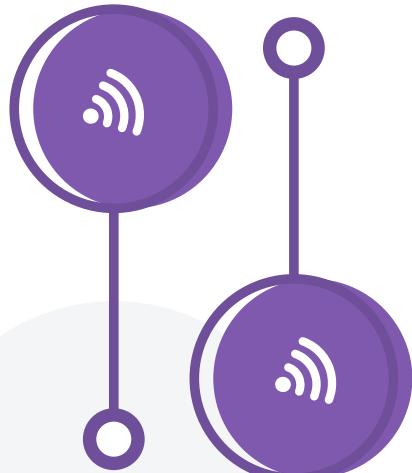
Academy



Photo by Aaron Doucett on Unsplash

Scores indicateurs

Rang centile



Rang centile

Rang centile



Rang centile



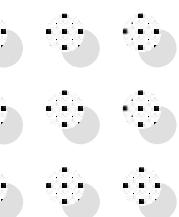
Rang centile



Score année



Moyenne



Nombre élèves secondaire (millions)

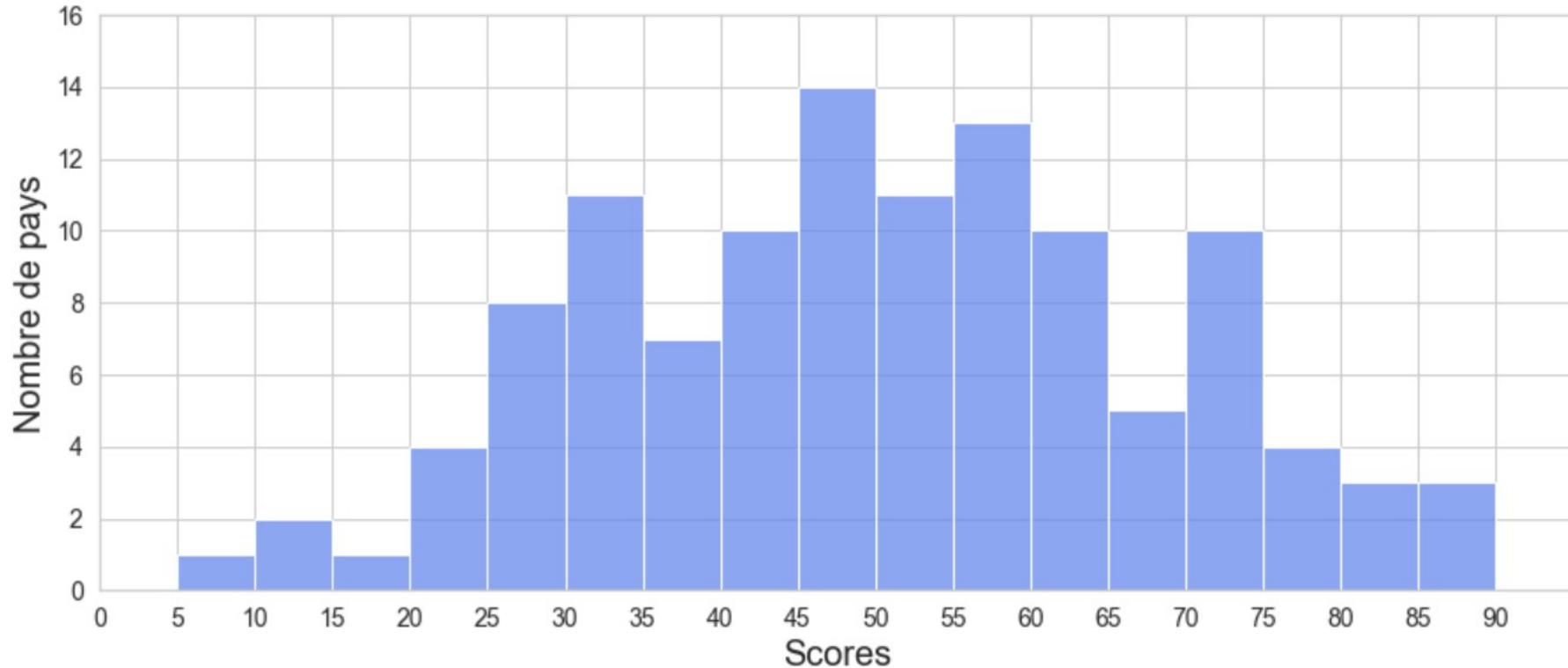
$$8 \ldots 23 < 24 \leq 129$$



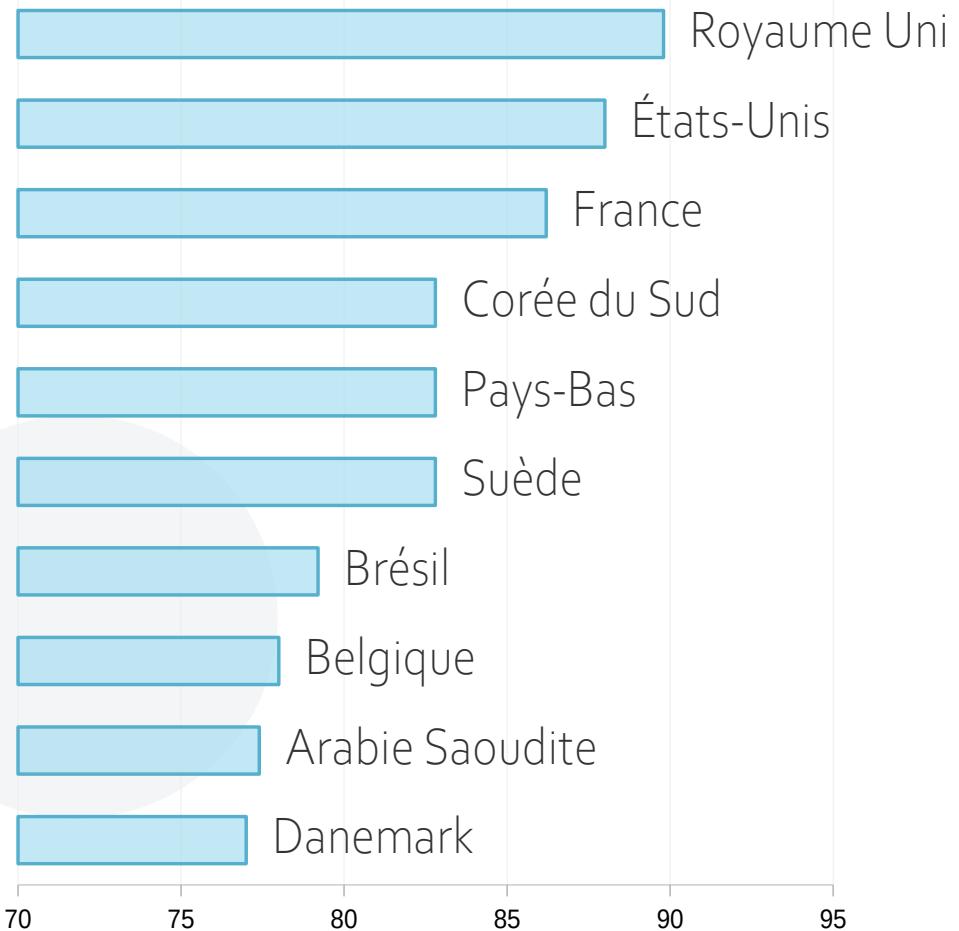
99 % pays Score : 99 1 % pays

2015

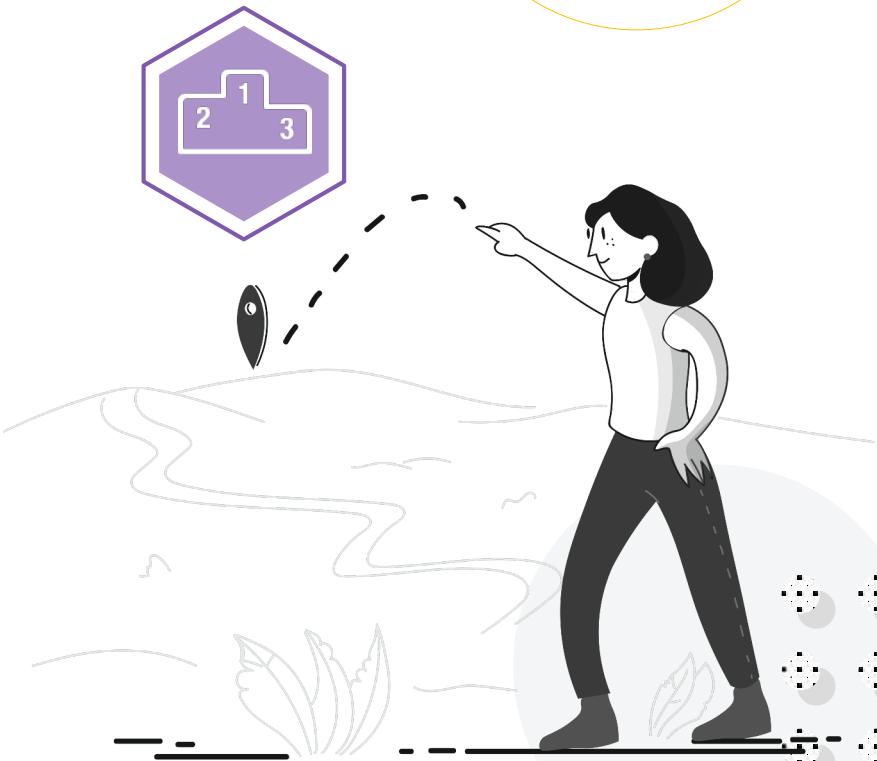
Distribution des scores



Top 10 2015

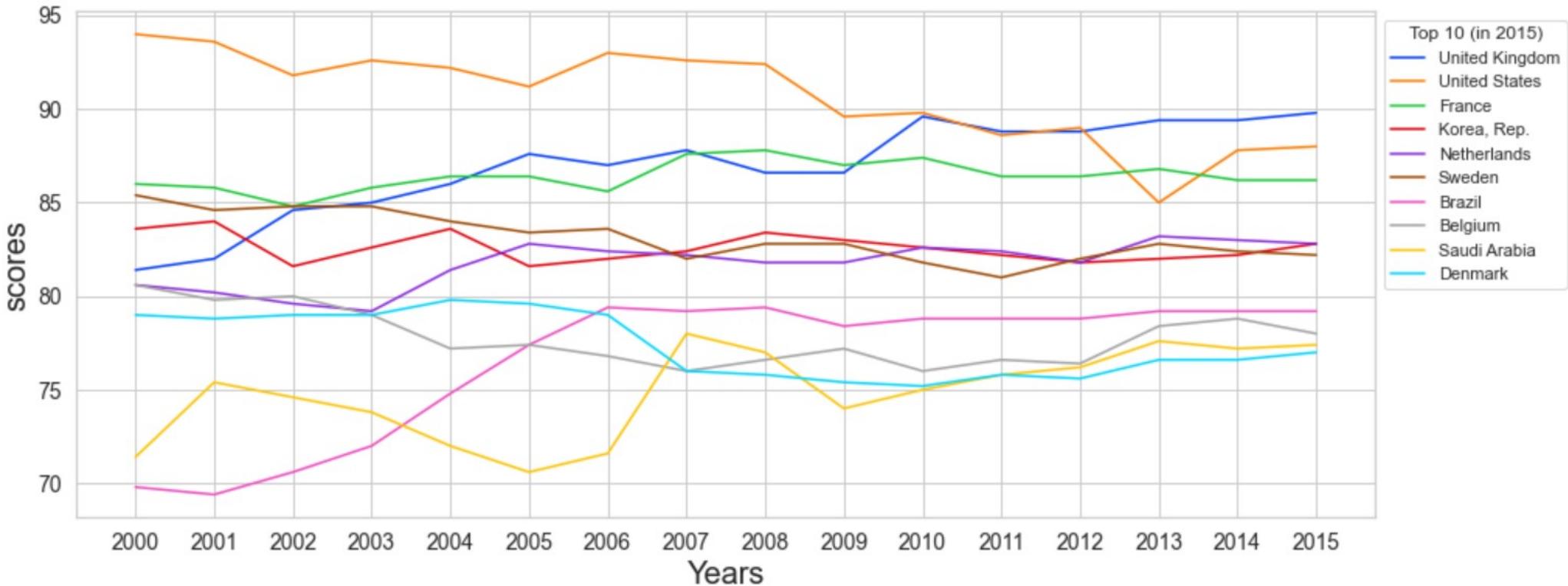


Academy



Scores du top 10

2000- 2015

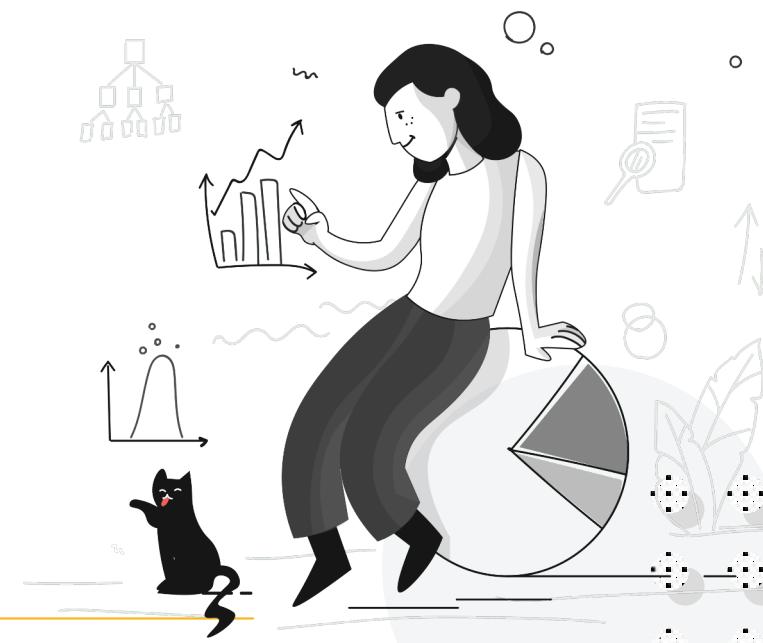


Conclusion

♦ 2022: 

♦ 2015 : 

Europe Nord
USA - Brésil
Corée Sud - Arabie Saoudite

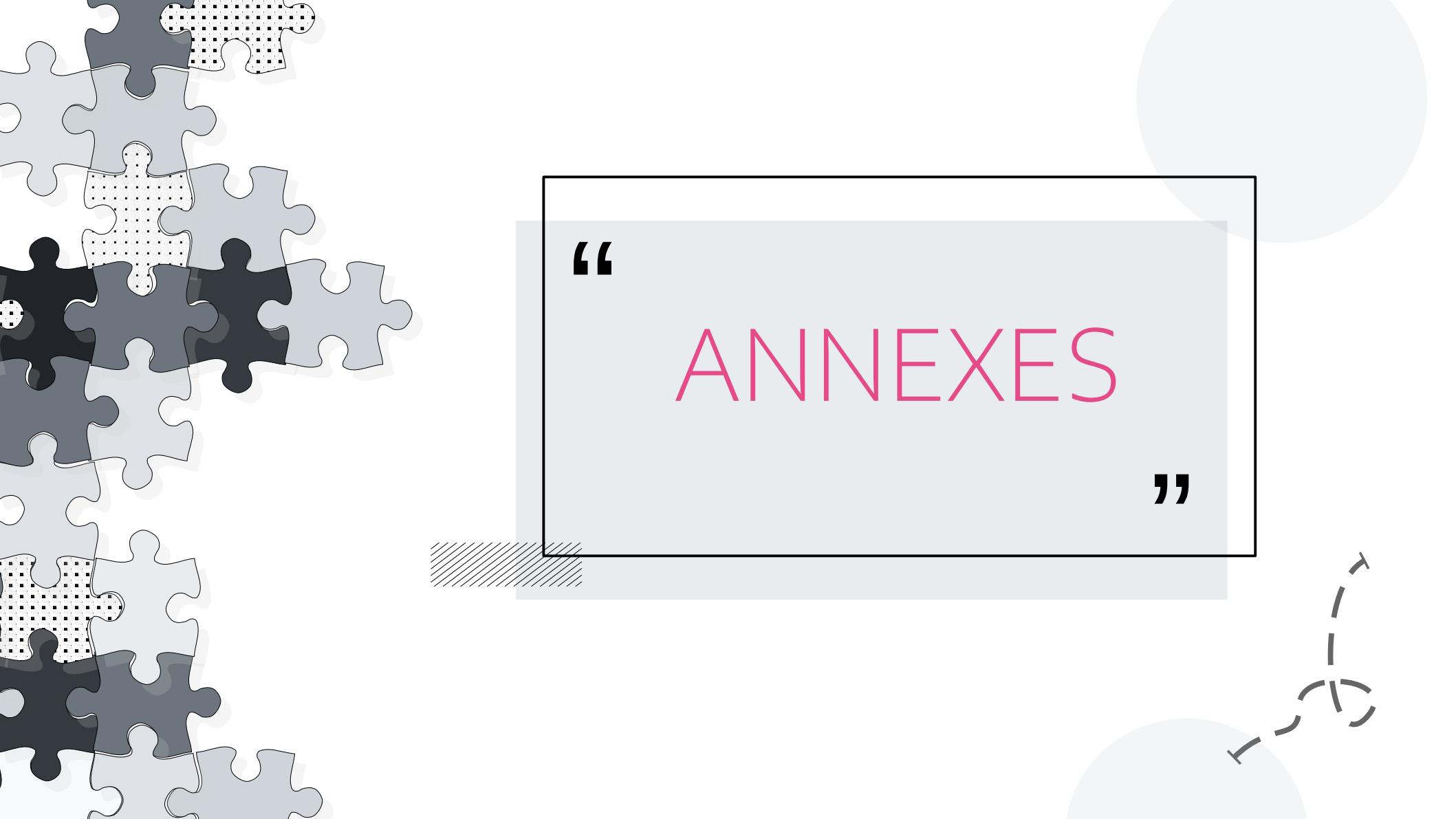


MERCI



Photo by David Clode on Unsplash

Academy



“ANNEXES”

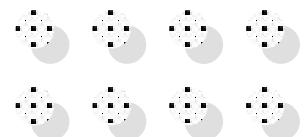
Plan des annexes

1. Boxplot

- ❖ Par région (2015)
- ❖ Par groupe de revenus (2015)
- ❖ Par région (moyenne 2000-2015)

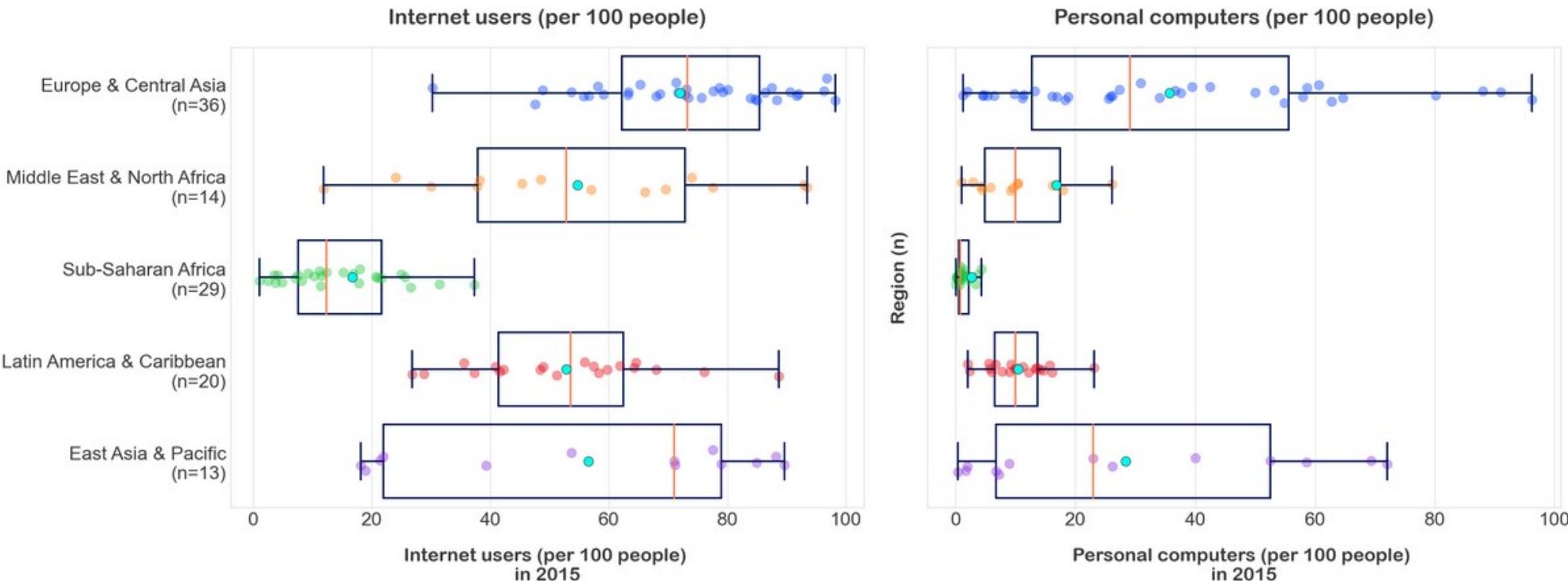


Illustrations by Pixeltrue on
[icons8](#)



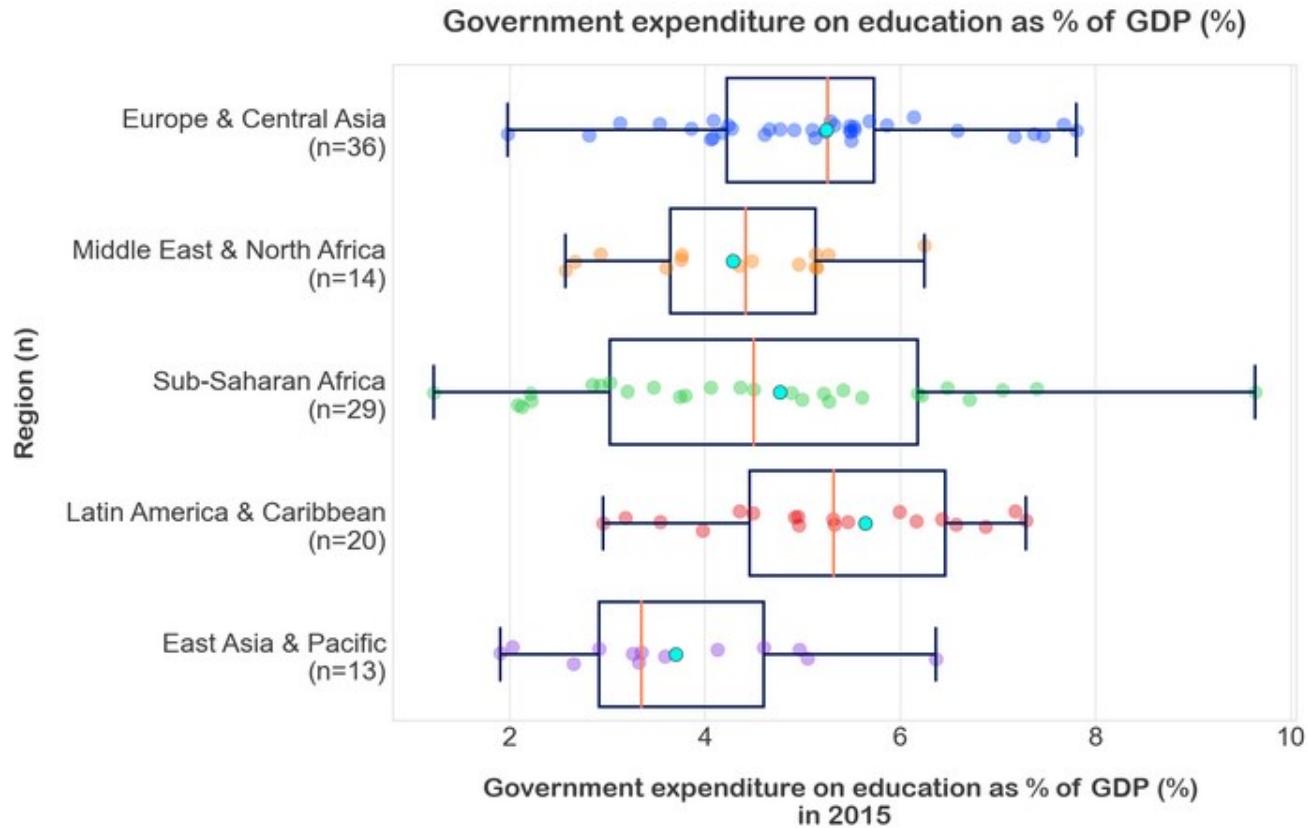
2015

Indicateurs par région

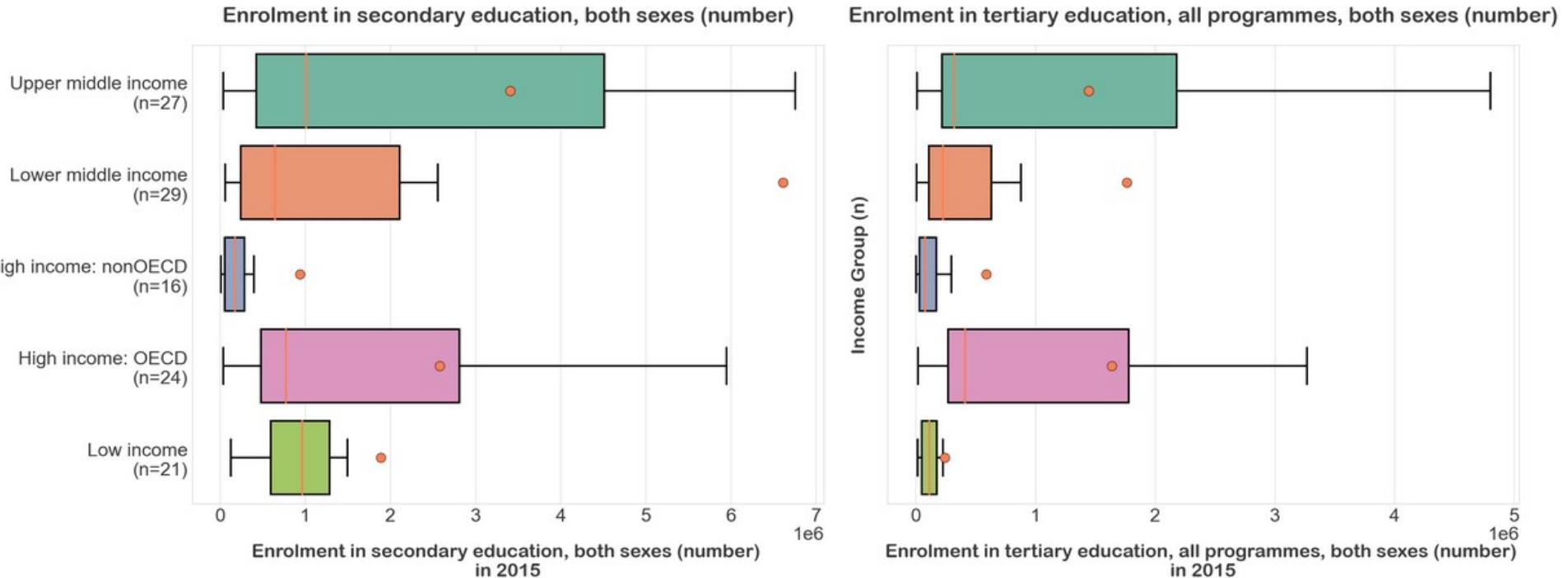


2015

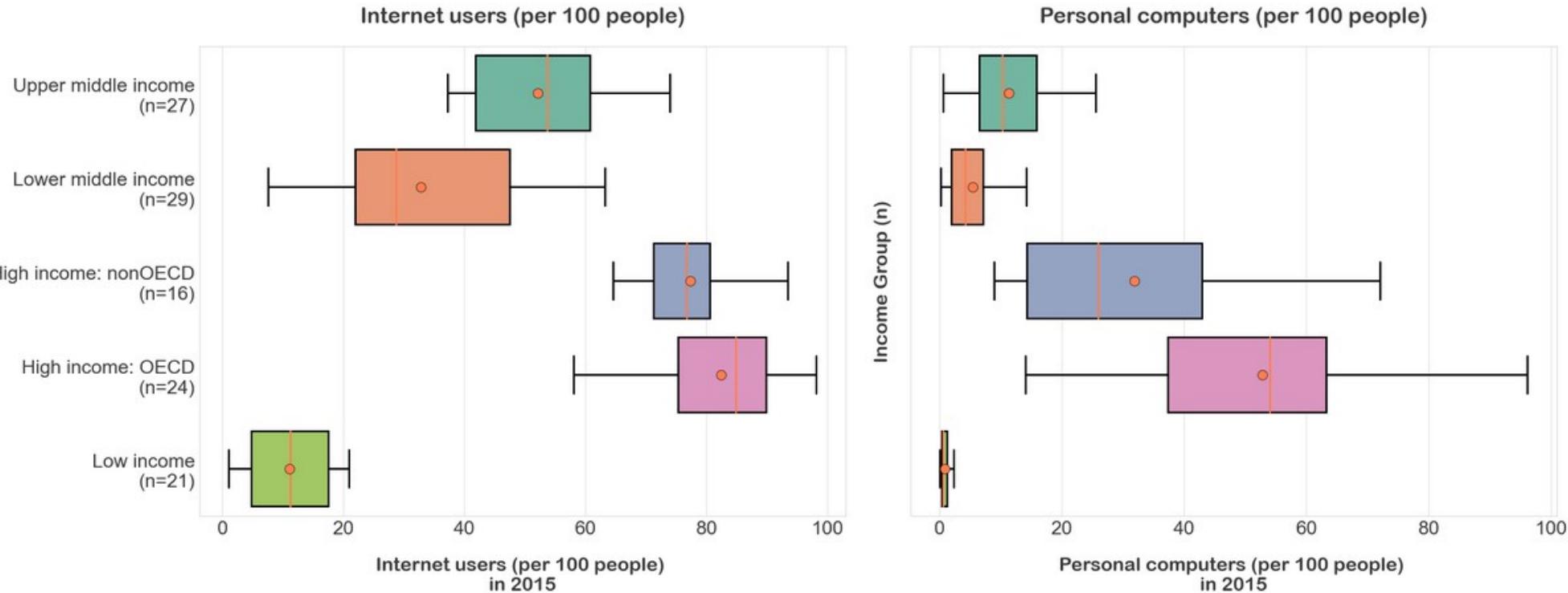
Indicateurs par région



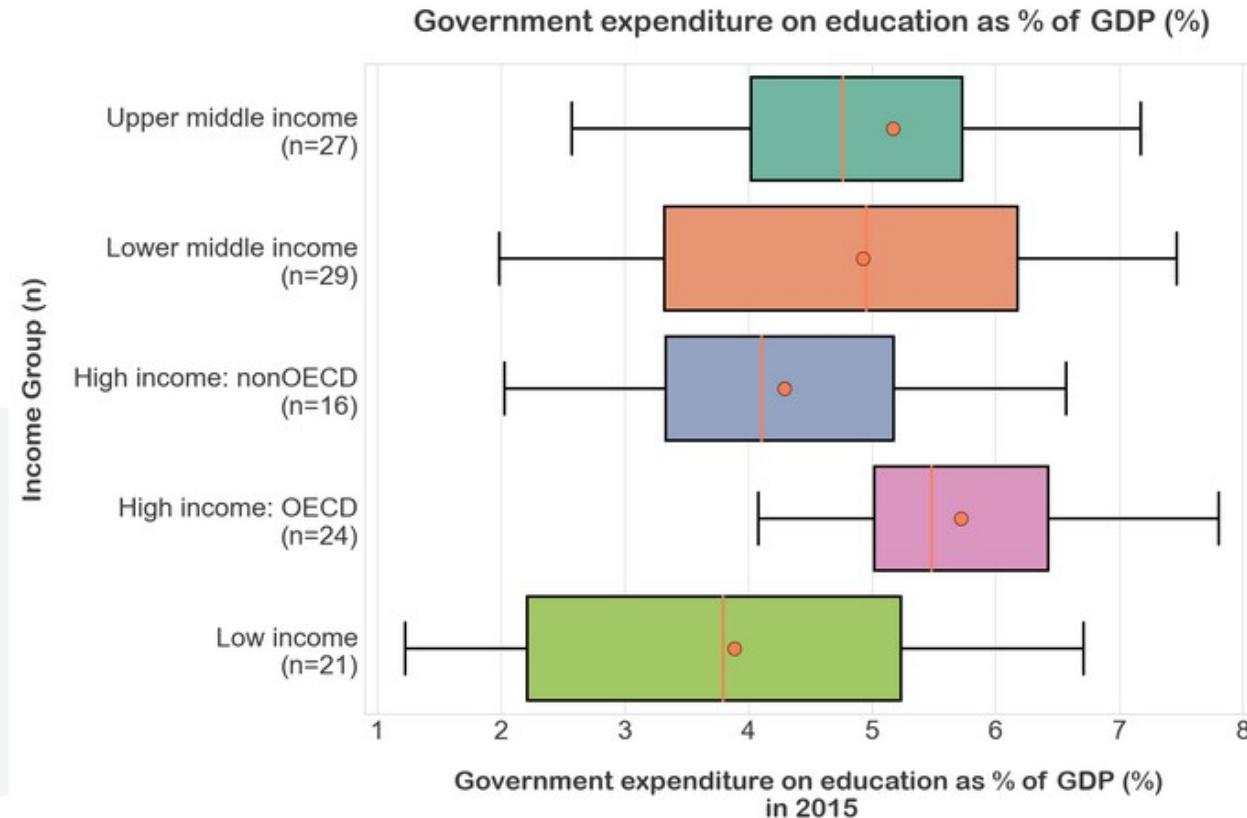
Indicateurs par groupe de revenus



Indicateurs par groupe de revenus

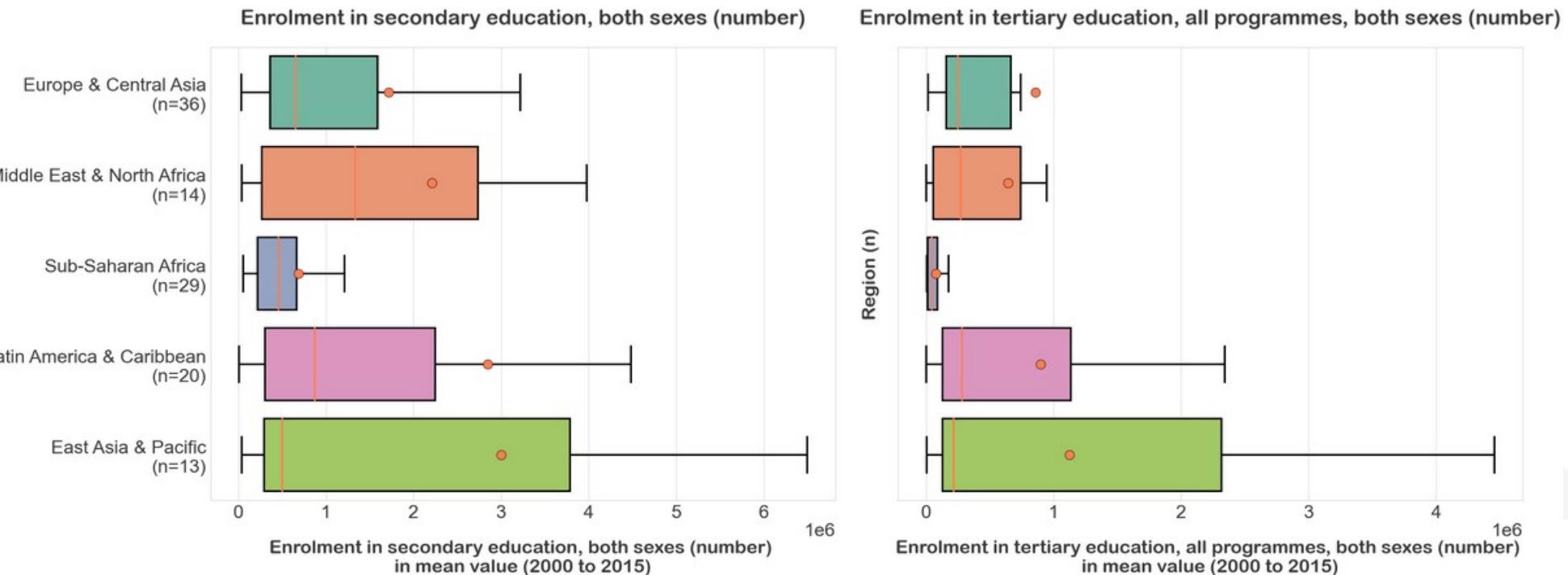


Indicateurs par groupe de revenus



Indicateurs par région

2000-2015



Indicateurs par région

2000-2015

Europe & Central Asia
(n=36)

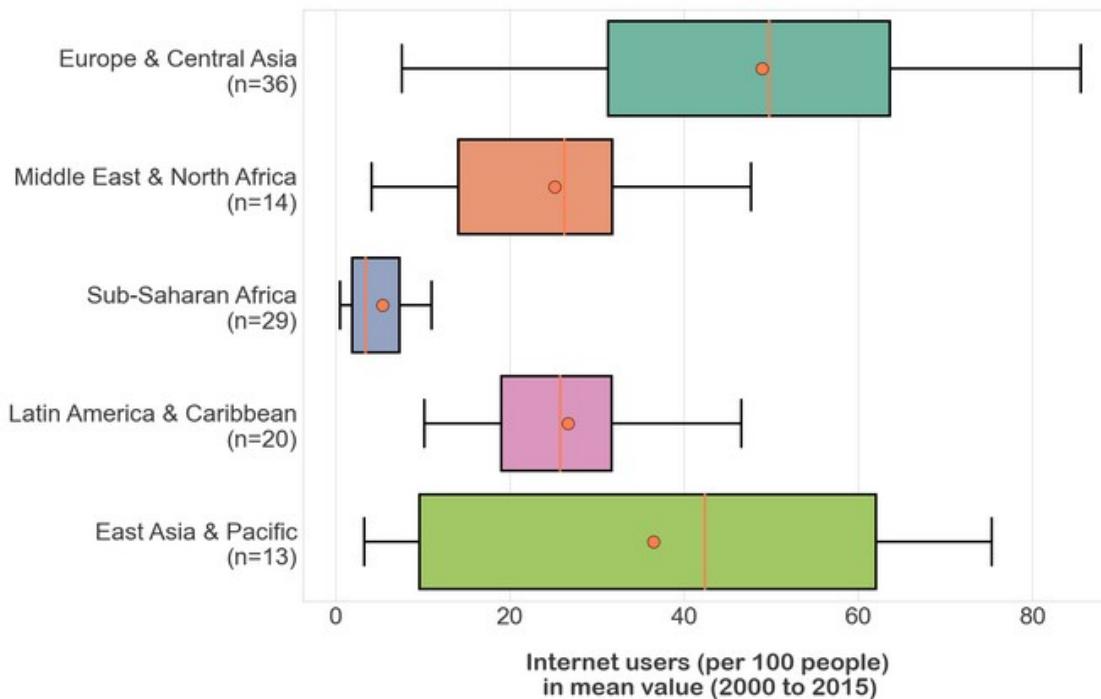
Middle East & North Africa
(n=14)

Sub-Saharan Africa
(n=29)

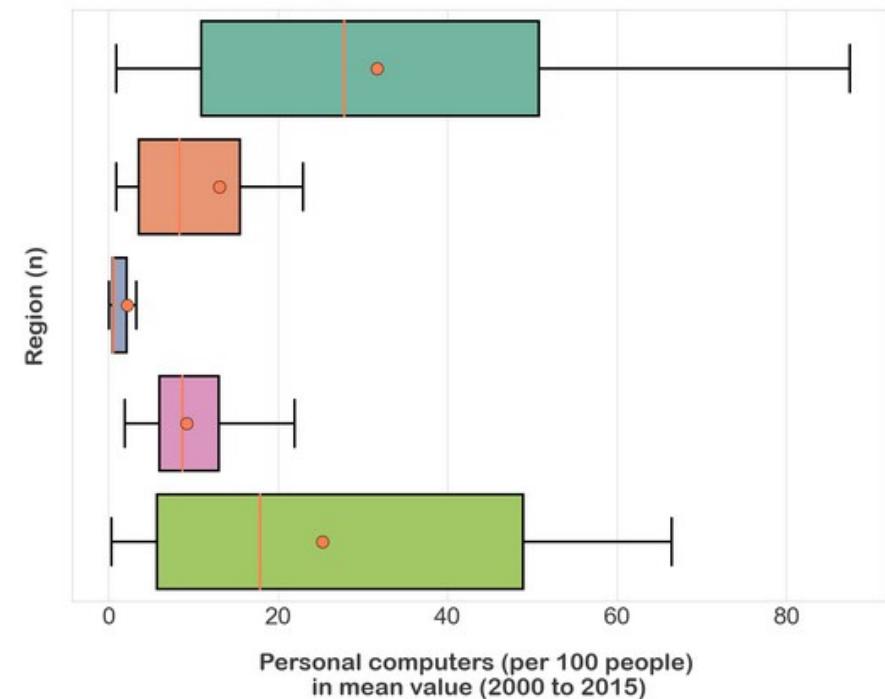
Latin America & Caribbean
(n=20)

East Asia & Pacific
(n=13)

Internet users (per 100 people)



Personal computers (per 100 people)



Indicateurs par région

2000-2015

