

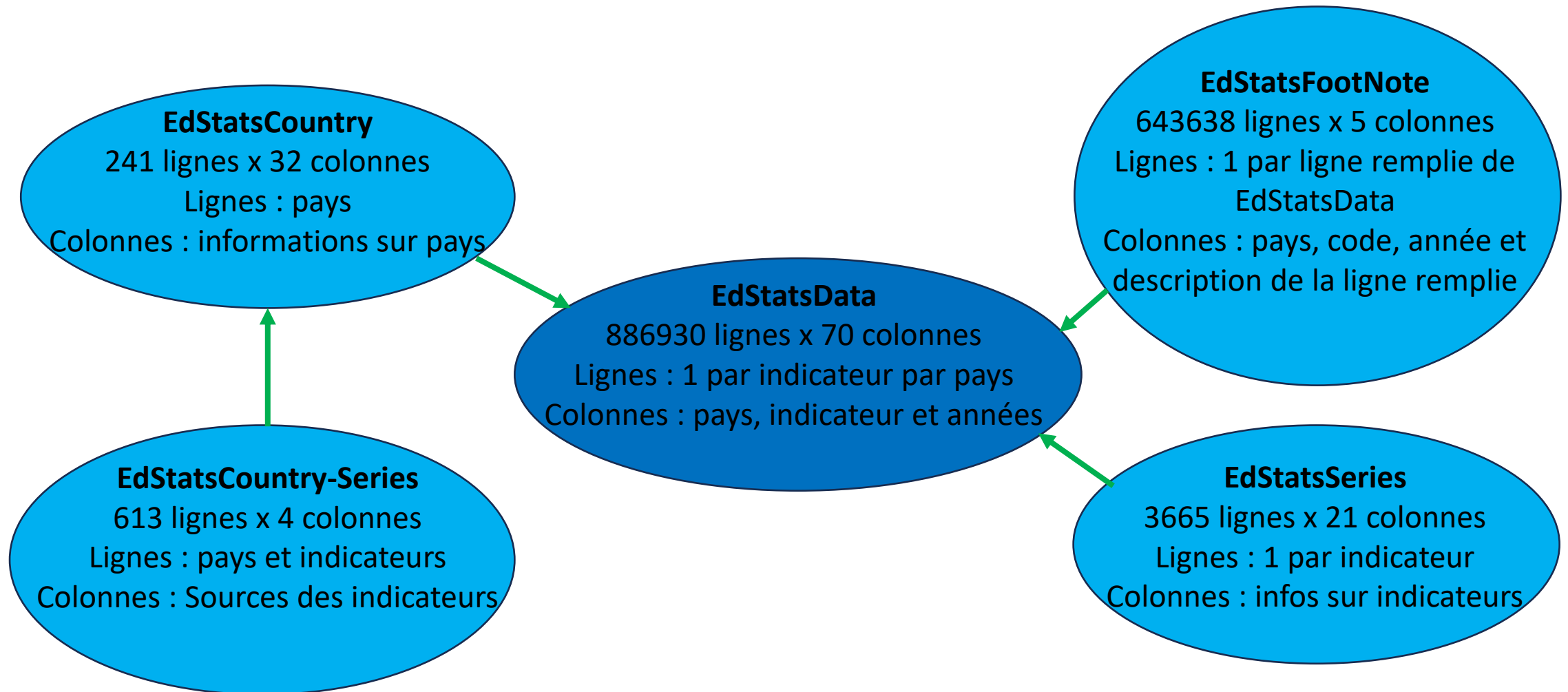
Expansion internationale Academy

Analyse exploratoire du jeu de données EdStats
de la banque mondiale

Introduction

- Description de la base de données
- Validation de la qualité des données
- Sélection des informations pertinentes
- Premières analyses
- Conclusion

Description de la base de données



Validation de la qualité des données

- Recherche des lignes dupliquées :

→ 0 lignes dupliquées

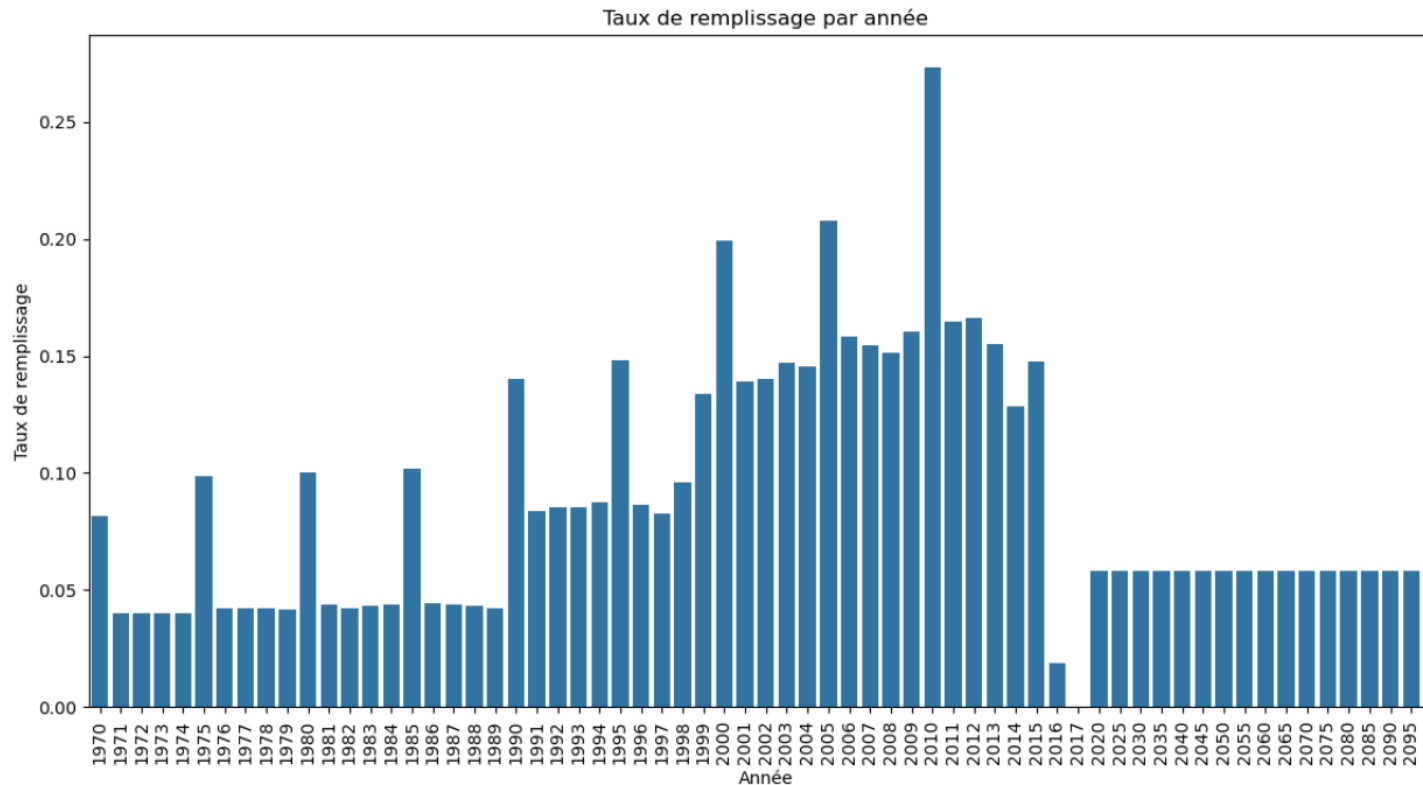
- Calcul du taux de remplissage :

→ EdStatsData : 13,9 %

➔ Tri et sélection

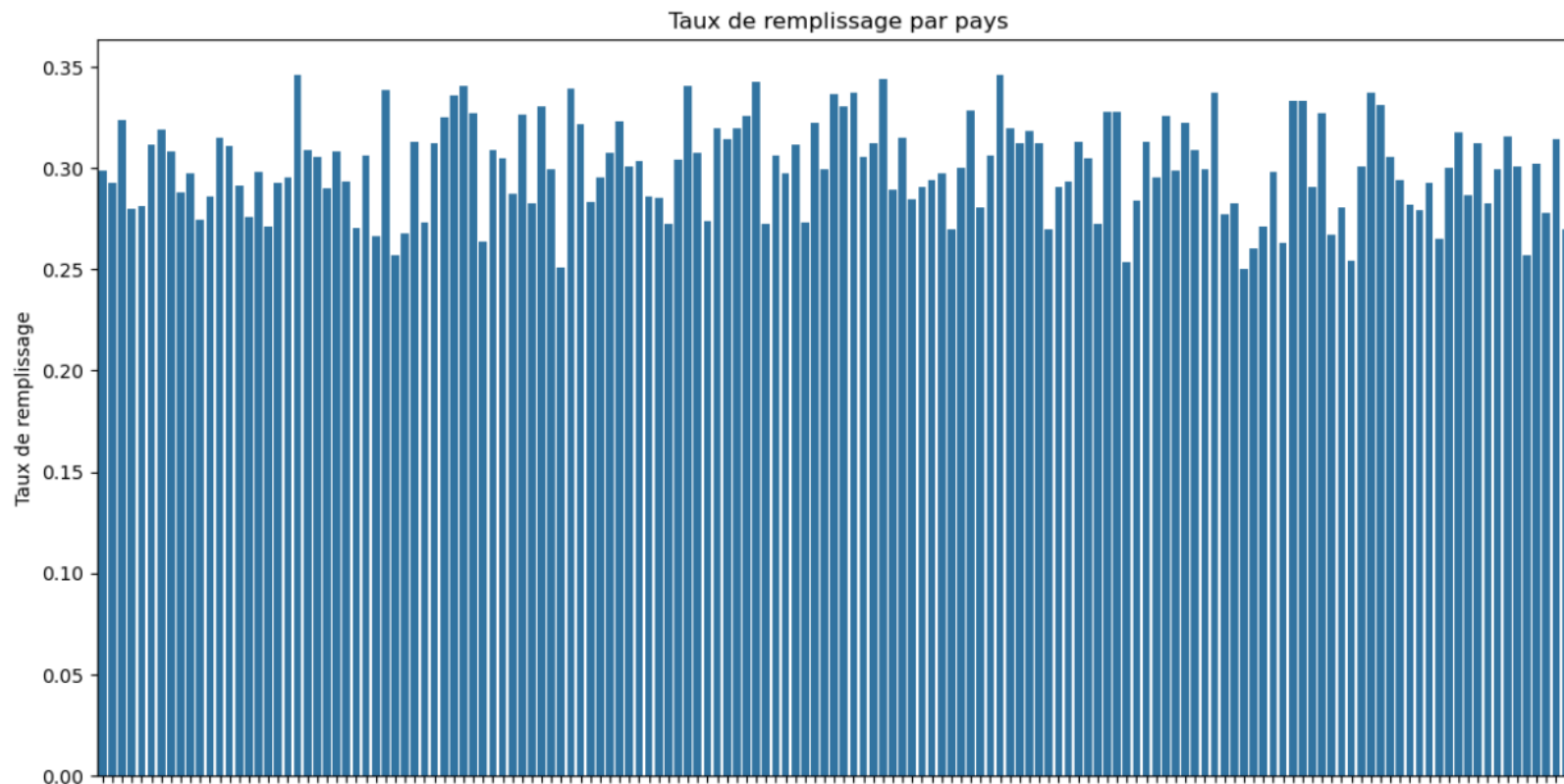
Sélection des informations pertinentes

- Approche technique : Taux de remplissage par année



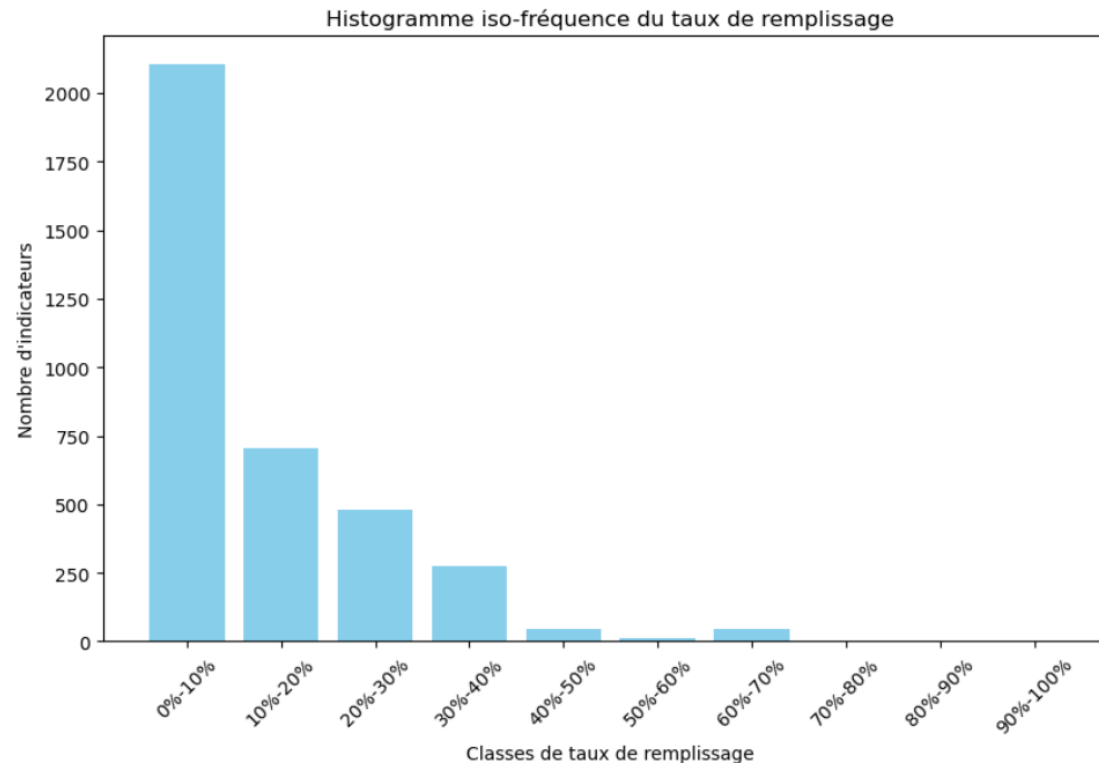
Sélection des informations pertinentes

- Approche technique : Taux de remplissage par pays (1999-2015)



Sélection des informations pertinentes

- Approche technique : Taux de remplissage par indicateurs



Sélection des informations pertinentes

- Approche technique : Sélection croisée
 - Indicateurs intéressants :
 - Population 15-24
 - Populations enseignements secondaire / supérieur
 - Utilisateurs d'internet
 - GDP per capita
 - etc

Sélection des informations pertinentes

- Approche métier : visualisation des thèmes des indicateurs (EdStatsSeries)
 - Thèmes d'intérêts : profs, éducation, financements écoles...
 - Autres indicateurs intéressants :
 - Professeurs
 - Ordinateur personnel pour 100 personnes
 - etc

Sélection des informations pertinentes

- Croisement des deux approches
- Récupération de l'information « Region » dans EdStatsCountry
→ Création du dataframe d'étude

3162 lignes x 17 colonnes

Sélection des informations pertinentes

- Indicateurs retenus :
 - GDP per capita
 - Scolarisation en secondaire / supérieur
 - Population en âge d'être en secondaire / supérieur
 - Nombre de professeurs dans le secondaire / supérieur
 - Durée du secondaire
 - % d'utilisateurs d'internet

Sélection des informations pertinentes

- Création de nouveaux indicateurs :
 - Proportion dans le secondaire (nb dans le secondaire / nb en âge d'être dans le secondaire).
 - Proportion dans le supérieur (nb dans le supérieur / nb en âge d'être dans le supérieur).
 - Professeurs par étudiant (nb prof / nb étudiants) dans le secondaire et le supérieur.

Etude des outliers

Méthode interquartile :

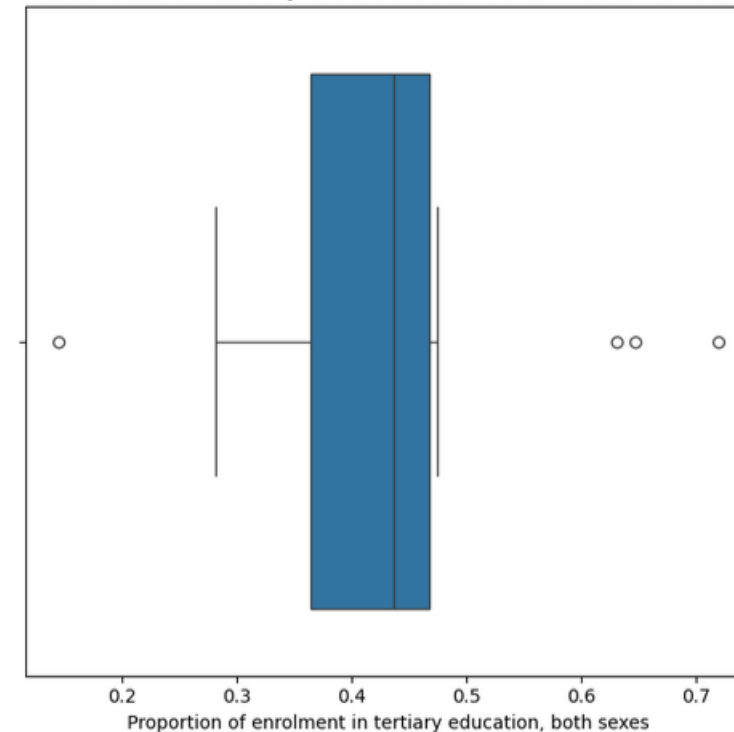
- Calcul des Q1 et Q3
- Calcul de l'IQR ($Q3 - Q1$)
- Exclusion des valeurs inférieures à

$$Q1 - 1,5 * IQR$$

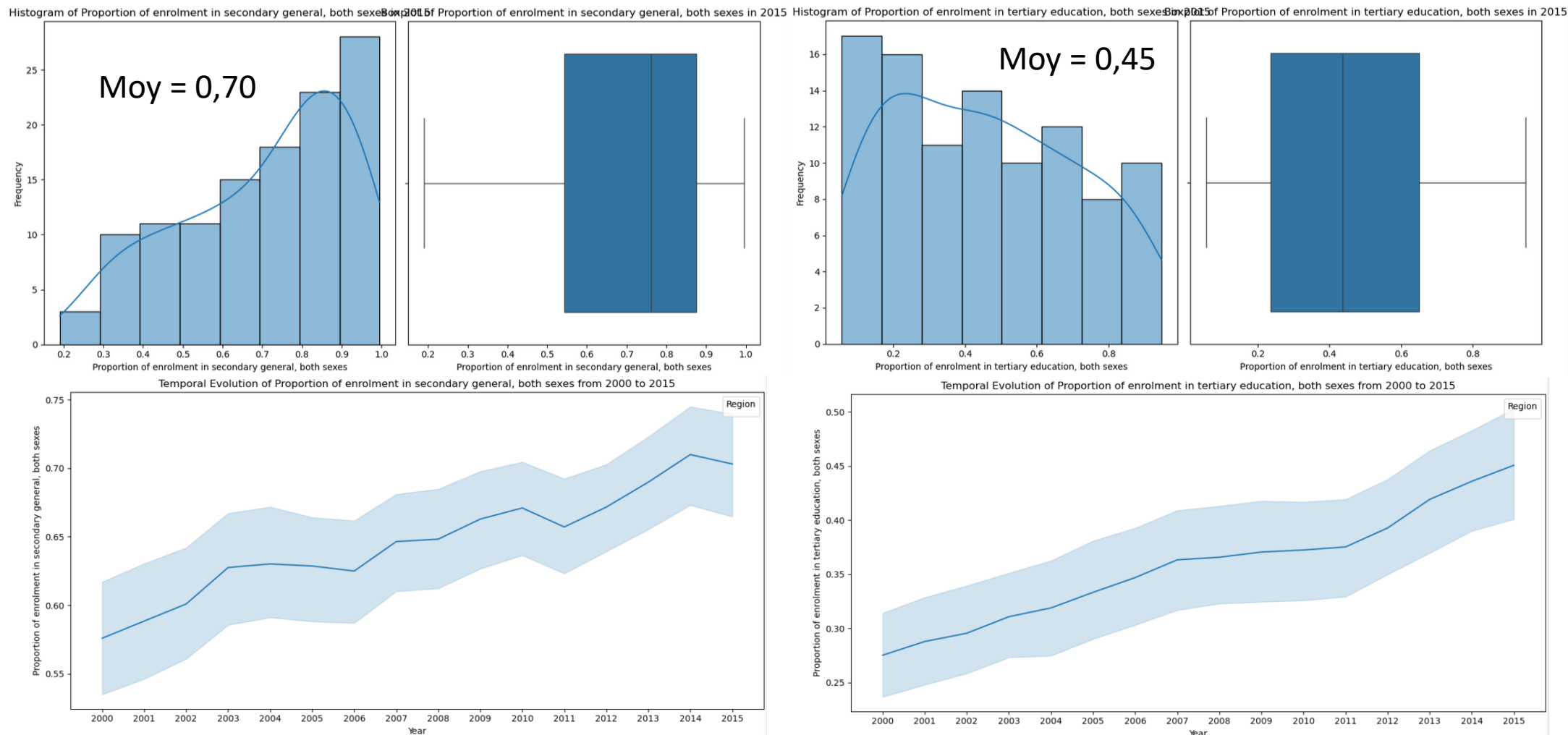
Et supérieures à

$$Q3 + 1,5 * IQR$$

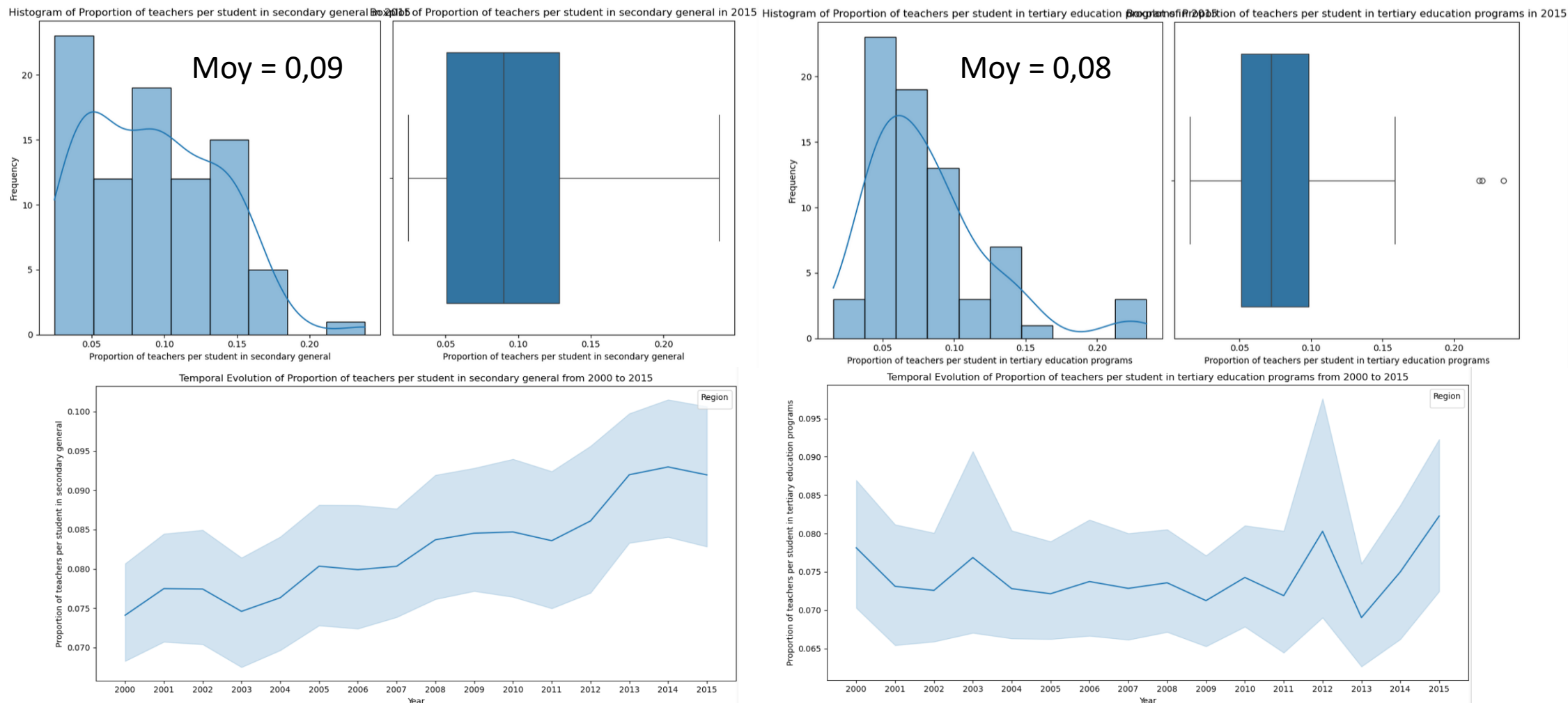
Boxplot of proportion of enrolment in tertiary education, both sexes in 2015 for Middle East and North Africa



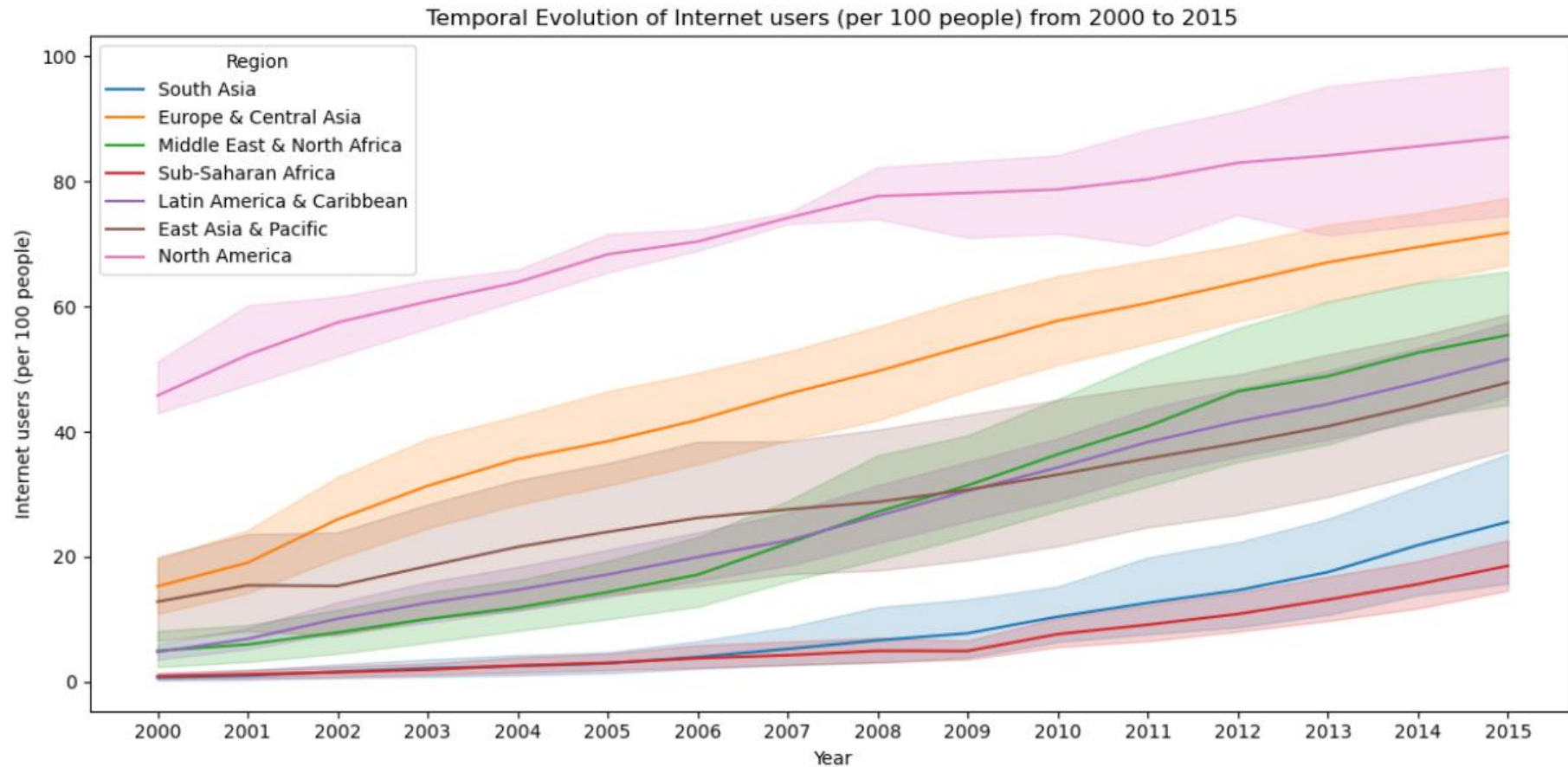
Premières analyses : taux de scolarisation



Premières analyses : professeurs par étudiants

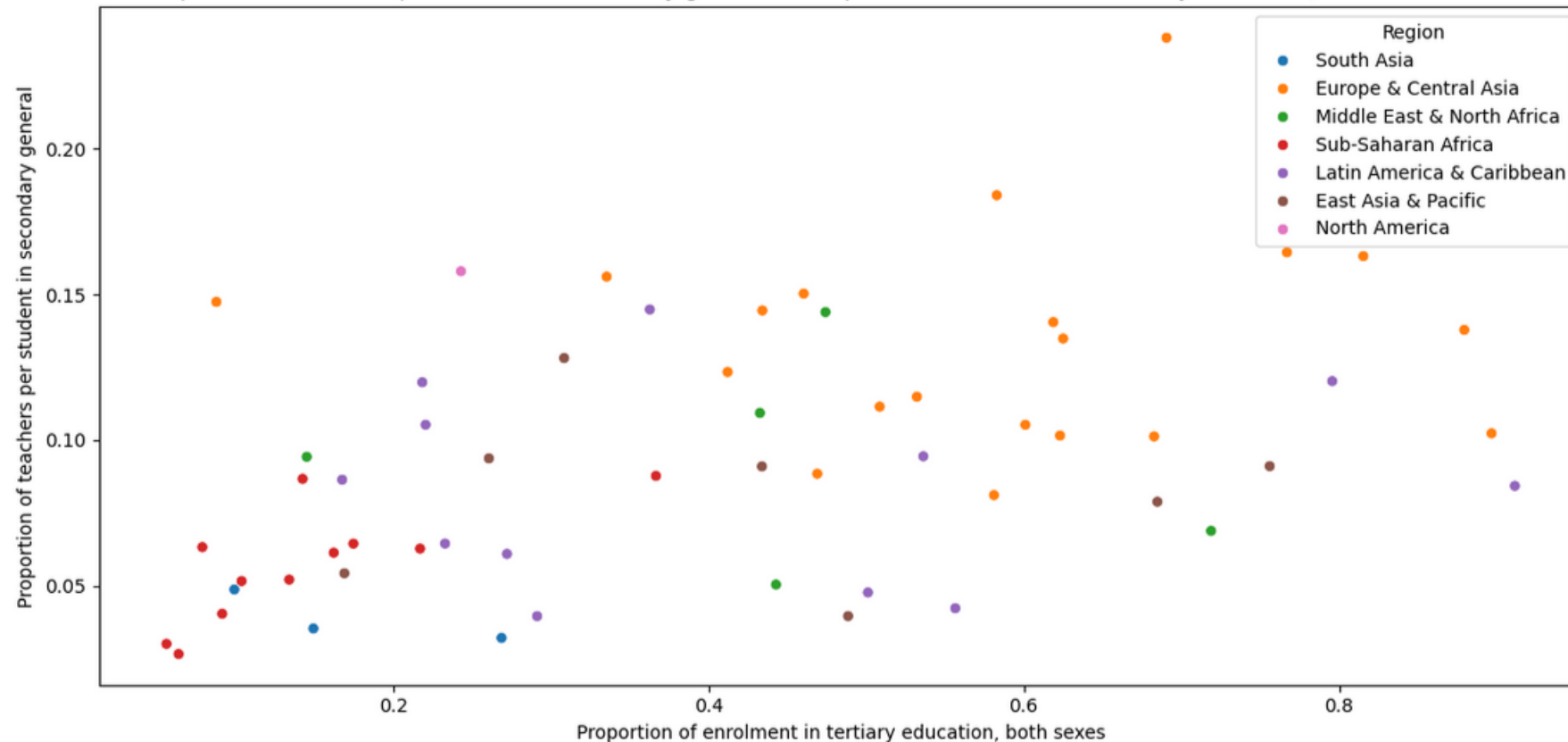


Premières analyses : analyse par région

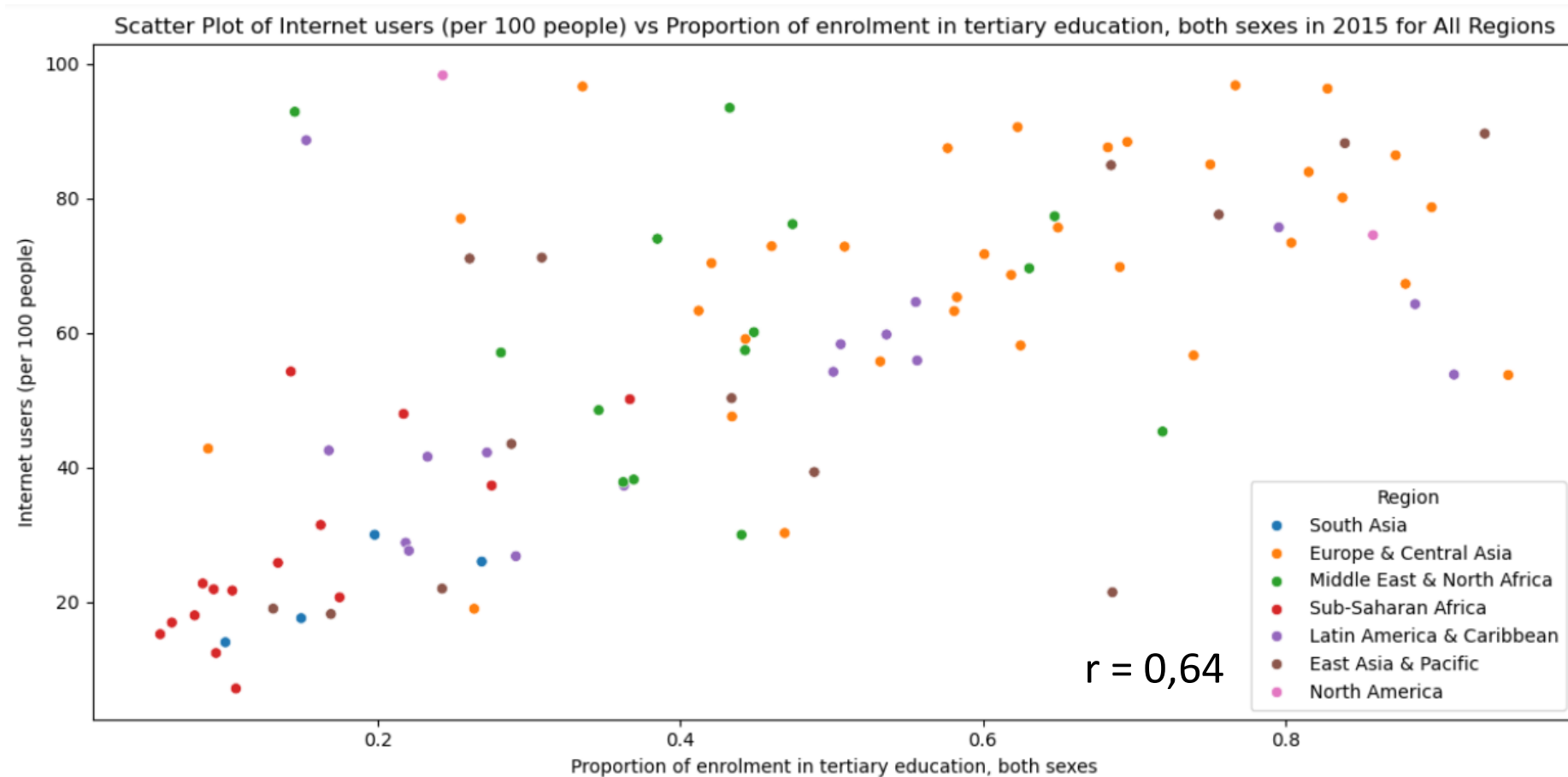


Premières analyses : scolarisation et professeurs

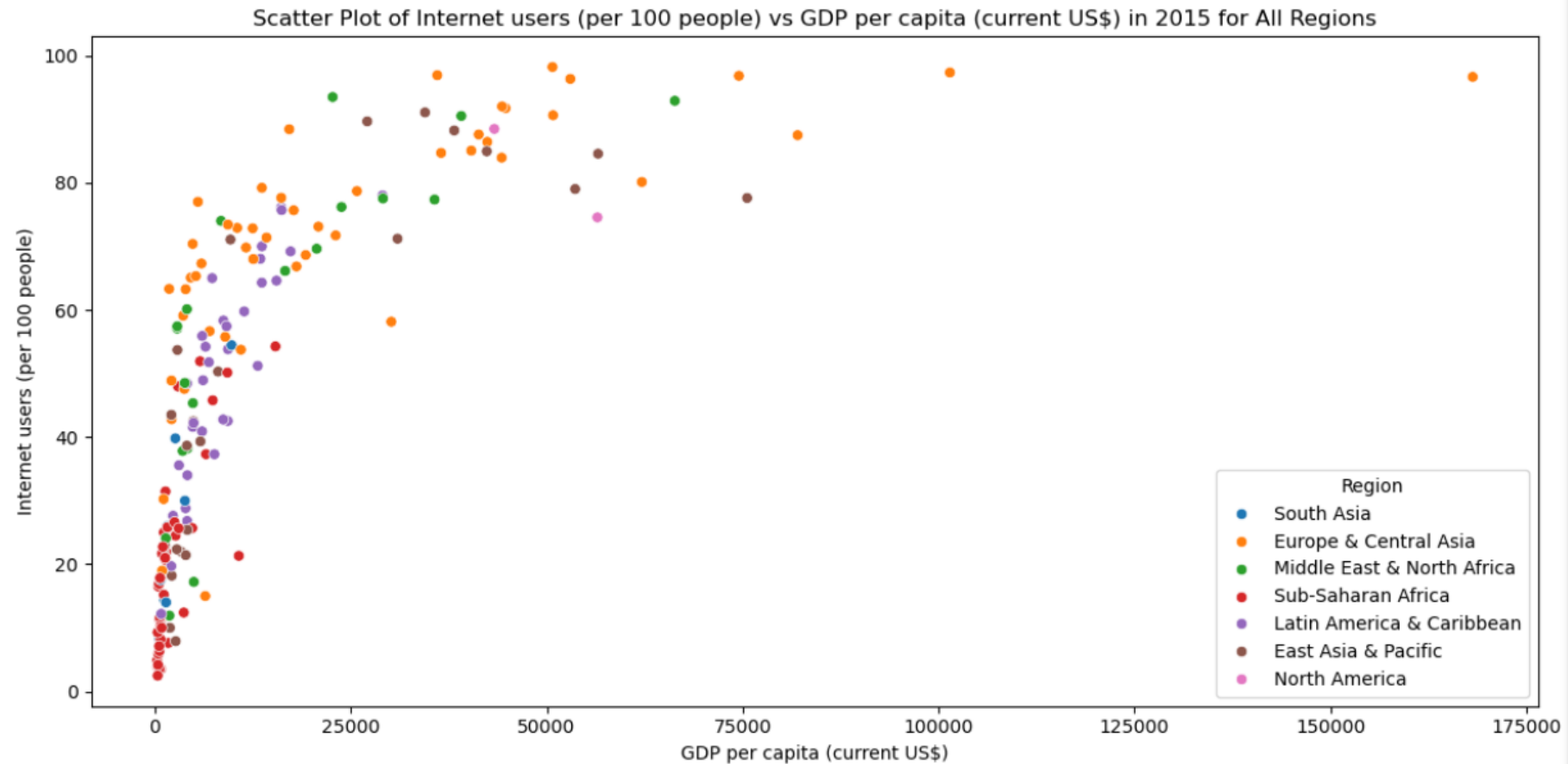
Scatter Plot of Proportion of teachers per student in secondary general vs Proportion of enrolment in tertiary education, both sexes in 2015 for All Regions



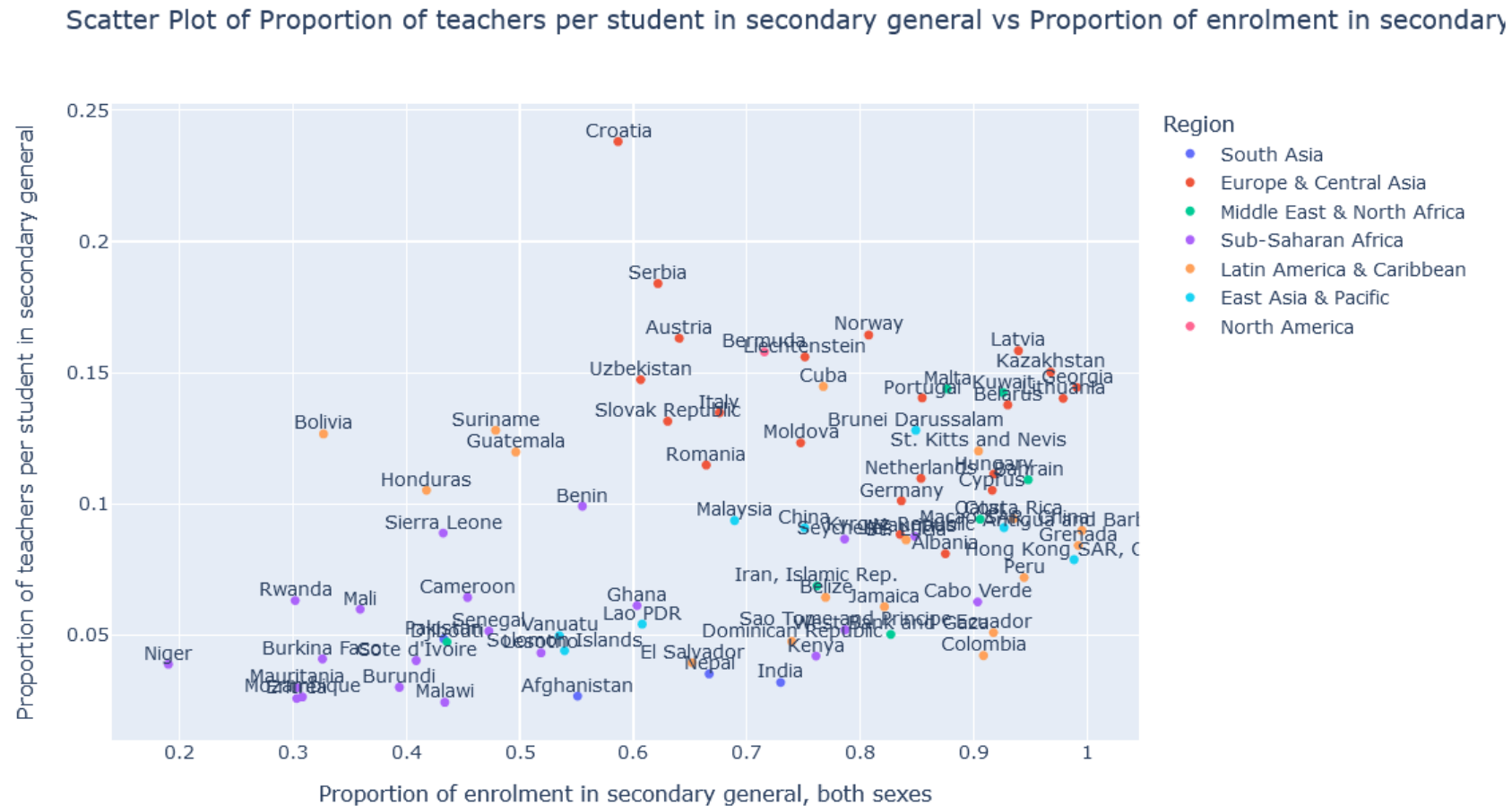
Premières analyses : scolarisation et internet



Premières analyses : internet et PIB/habitant



Premières analyses : visualisation pays



Conclusion

- Base de données EdStats bien pour pré-analyse.
- Peu de surprises, pays les plus intéressants sont globalement les pays « riches »
- Manques données essentielles :
 - Langue
 - Perception des formations en ligne dans milieux professionnels