

# PROJET 2 - ANALYSEZ DES DONNÉES DE SYSTÈMES ÉDUCATIFS

RACHEDI YASSINE

# ACADEMY

- Formation en ligne, pour lycéen et universitaire en France
- 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 ?

# EDUCATION STATISTICS - DONNÉES

- 5 datasets
- EdStatsData :
  - 242 pays
  - 3665 indicateurs
  - 1970 à 2100
  - 886 930 lignes – 70 colonnes

# INDICATEURS

- GNI (current US\$)
- GNI per capita, PPP (current international \$)
- Gross enrolment ratio, secondary, both sexes (%)
- Gross enrolment ratio, tertiary, both sexes (%)
- Internet users (per 100 people)
- Population, ages 15-24, total



# NETTOYAGE

Country Code	Indicator	1970	...	2100
FRA	GNI (current US\$)	1	...	3
...	...	...	...	...
FRA	Internet users	4	...	6



Country Code	Year	GNI (current US\$)	...	Internet users
FRA	1970	1	...	4
...	...	...	...	...
FRA	2100	3	...	6

- Suppression des années pour lesquels on n'a aucune donnée (sur les indicateurs choisis)
- Suppression des données avant 1990
- Suppression des pays pour lesquels il manque plus de 50% des données

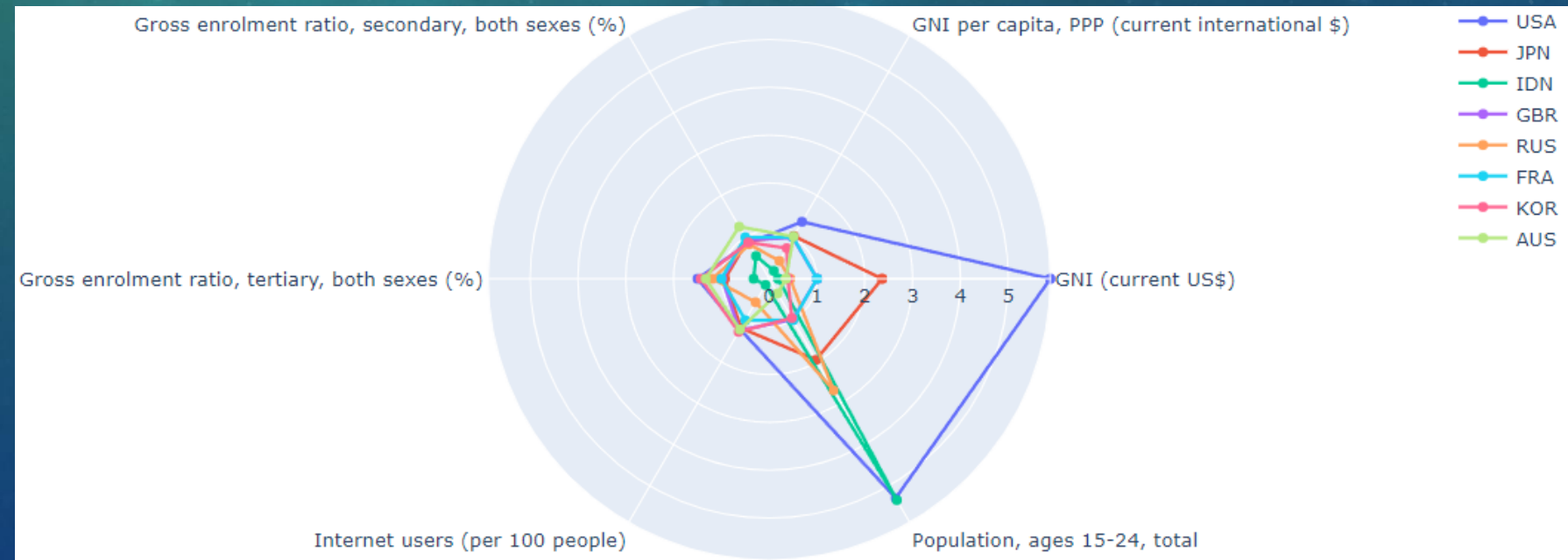
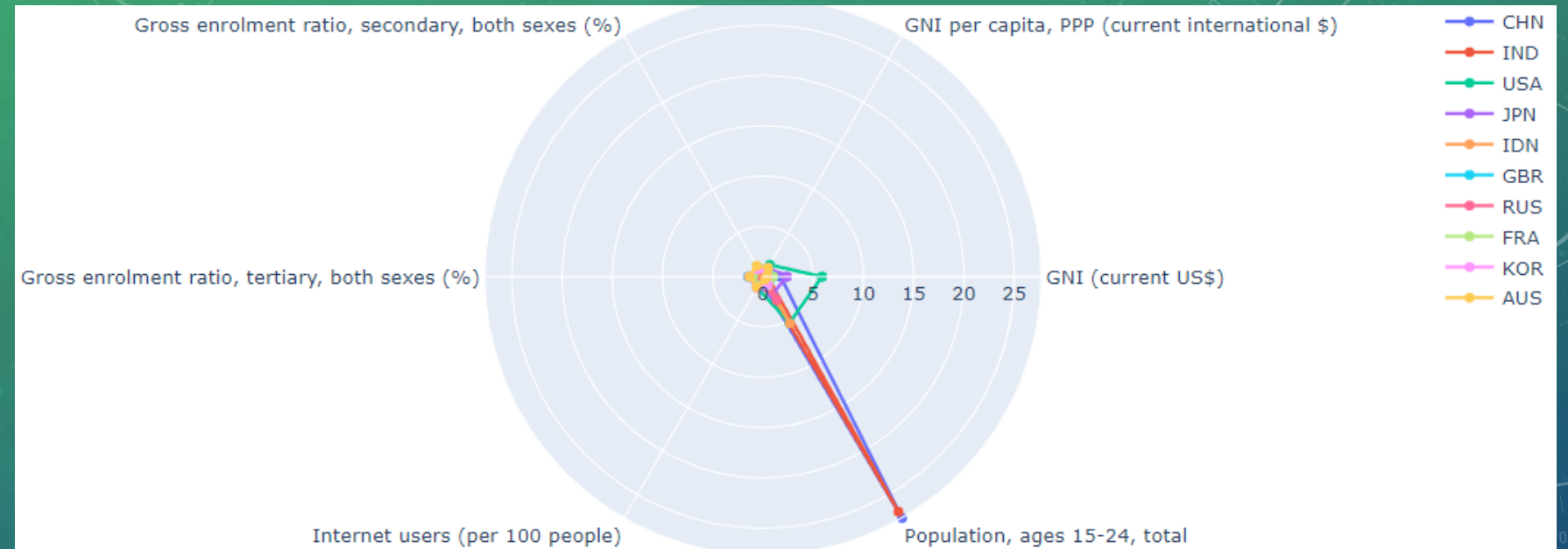
(886 930, 70) => (1 452,70) => (10 077,8) => (6209,8) => (2727,8)

# TRAITEMENT

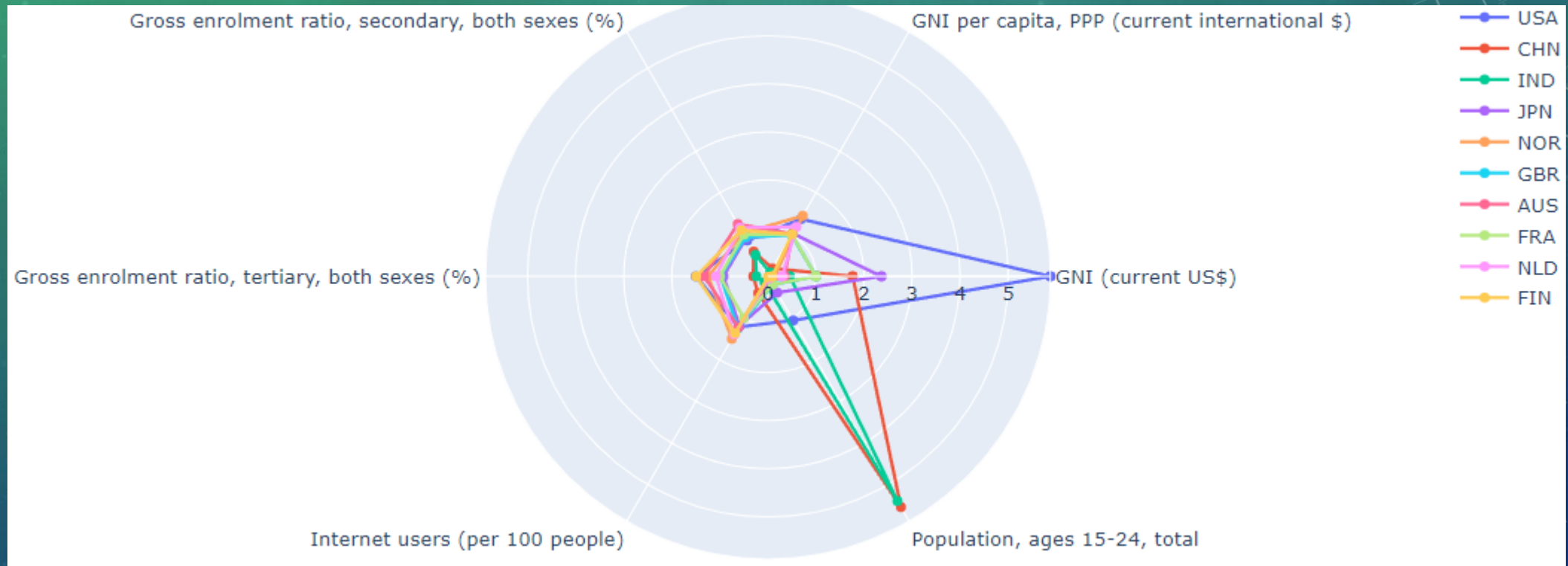
- Si première année manquante : on remplace par celle d'après
- Si valeur manquante, on remplace par la moyenne des 2 valeurs adjacentes
- Sinon, on utilise la méthode « pad » (propagation)
- Normalisation autour des donnée française ('Min – MoyFr')
- Normalisation par le log

# SCORING

- Moyenne des indicateurs, normalisation autour de FR
- Suppression de CHN et IND pour la visualisation



# SCORING

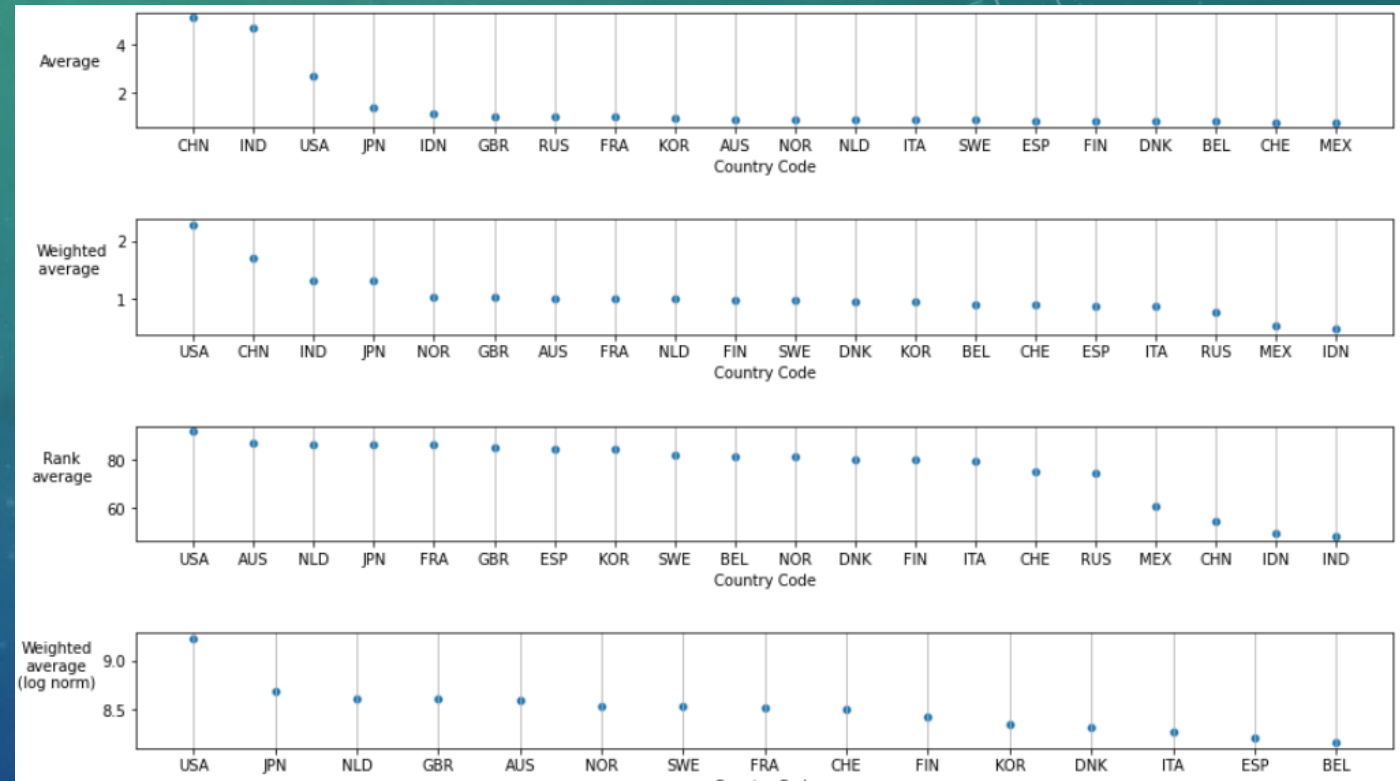




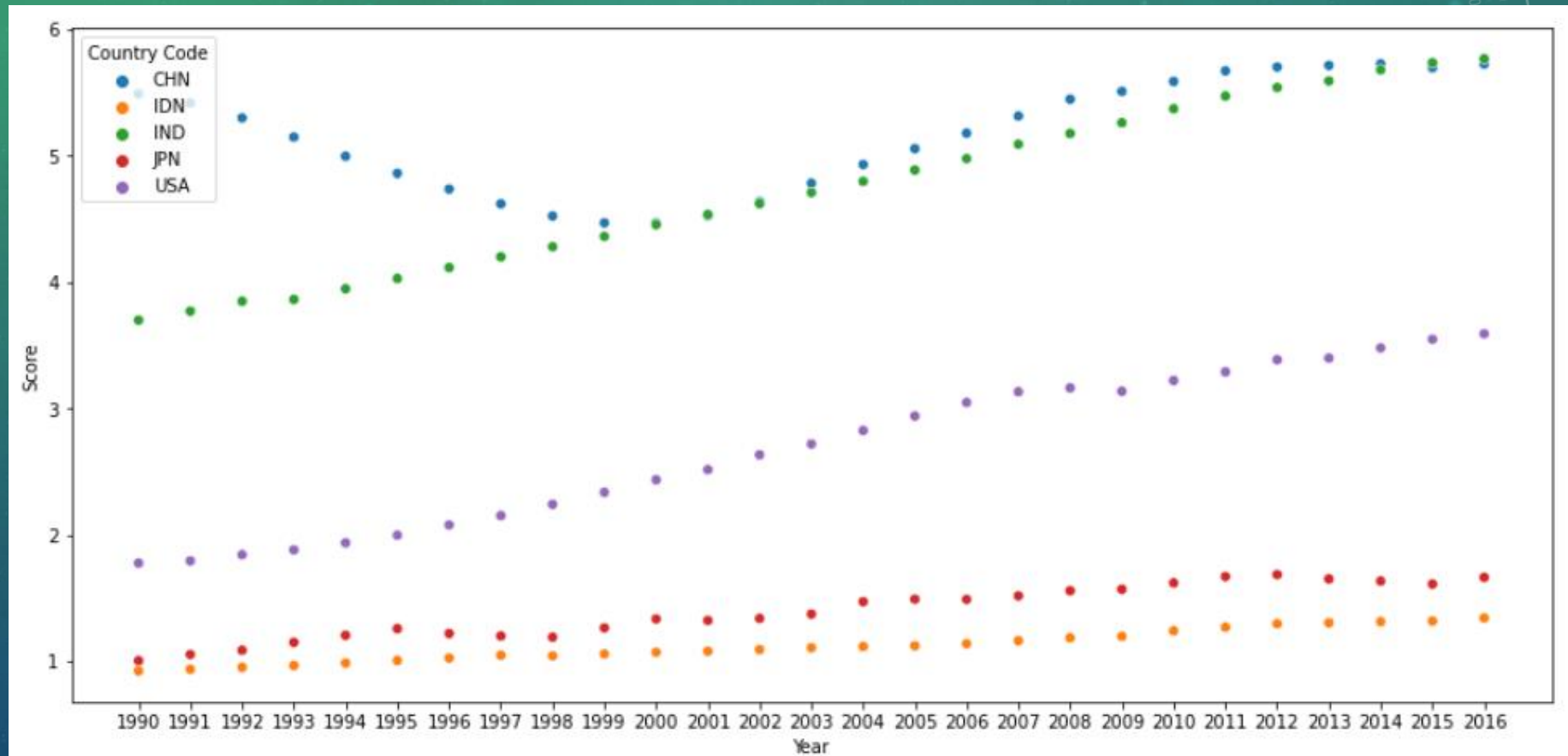
# SCORING

## 4 méthodes :

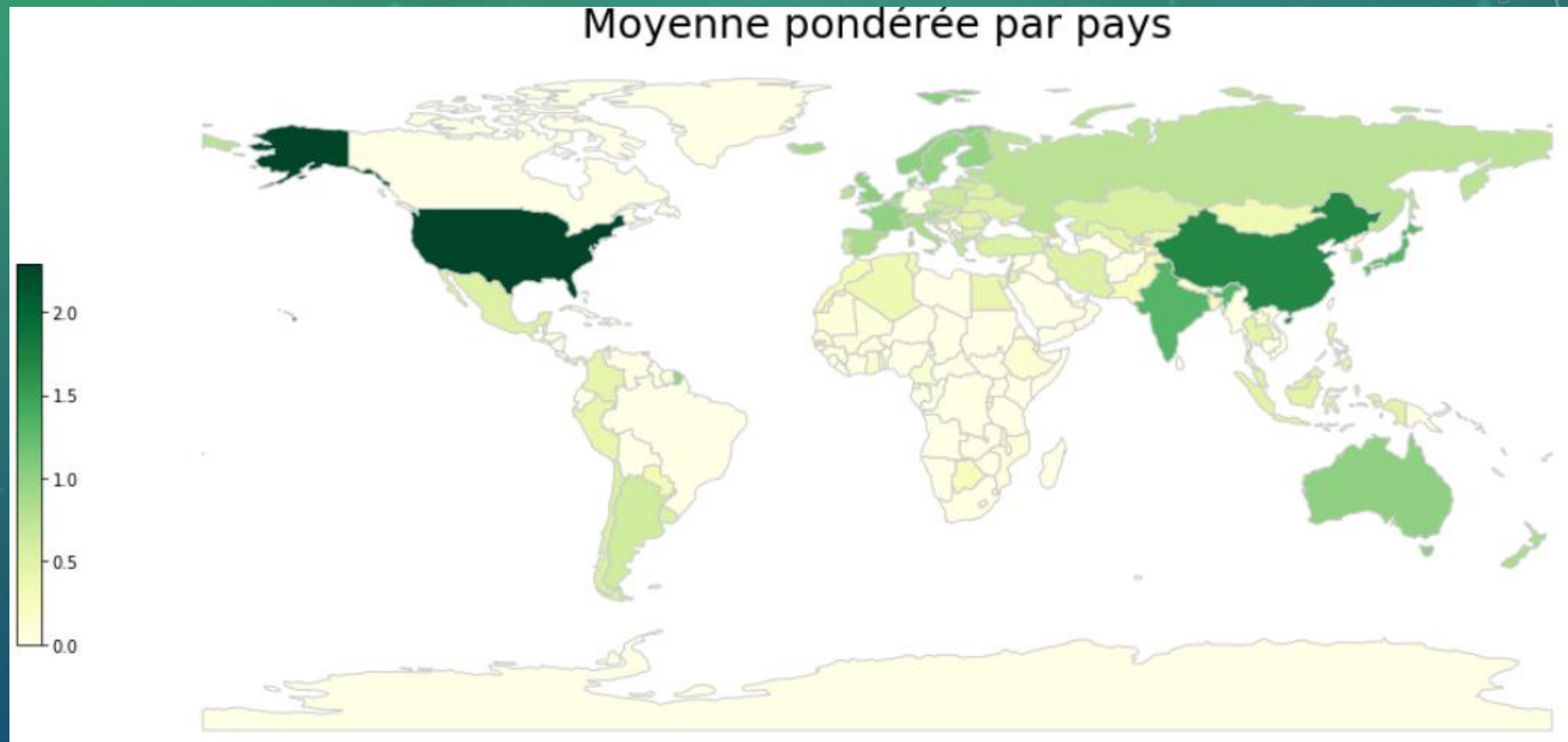
- Moyenne des indicateurs puis moyenne par année
- Moyenne pondérée des indicateurs, puis moyenne par année : coef 1/5 pour la population
- Moyenne des rangs des indicateurs (par pays), puis moyenne par année
- Moyenne pondérée des indicateurs puis moyenne par année avec la normalisation log: coef 3 pour le PIB/h, coef 4 pour internet par 100/h



# TENDANCES : MOYENNE



# CARTOGRAPHIE



# CONCLUSION

- 4 scoring différents
- Quelques pays redondant (USA, Japon)
- Evolution positive





# QUESTIONS

RACHEDI YASSINE