



Rapport

Project 2: Unsupervised Analysis of the Human Development Report

Auteur:

Siéwé Kouéta ANICET : 00364245

Professeur:

Benoît FRÉNEY



UNIVERSITÉ LIBRE DE BRUXELLES **ULB**

November 22, 2019

1 Clustering and Visualization

1.1 task 1

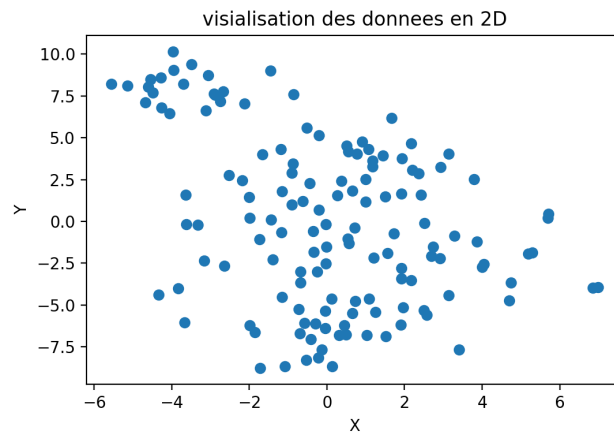


Figure 1: repartition global des données dans l'espace 2D

Après avoir charger et normalisation les donnees du dataset **hdr_data.dat** nous avons un jeu de données constitué de 138 points visualisées via un diagramme de dispersion bidimensionnel. en observant la figure 1 on peut estimer que les point sont regroupés grossièrement en cinq(5) régions. nous avons donc decider de regrouper les pays en 5 cluster(**K=5**) dans l'algorithme K-means,

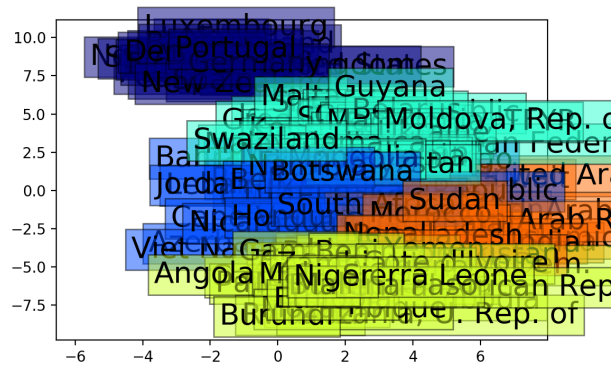


Figure 2: cluster K-means, avec k=4

par code de couleur, les centroïdes associer a chaque Cluter sont :

Austria, , , Syrian Arab Republic

1.2 task 2

NOTE : les reponses données sur cette partie se base les donnees recoller et stoker dans les fichier *.txt* se trouvant dans `/doc/Graphes/ *.txt` .

1.2.1 pas assez de clusters (k=2)

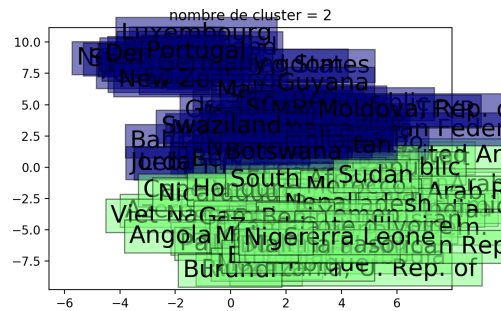


Figure 3: 2_Means

1.2.2 pas assez de clusters (k=6)

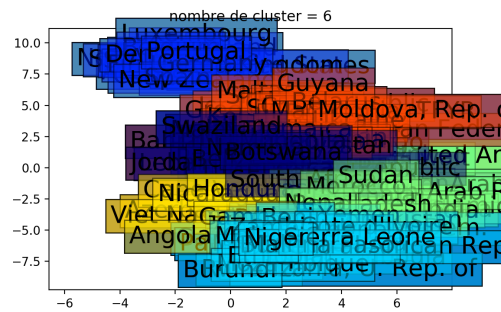


Figure 4: 2_Means

1.2.3 pas assez de clusters (k=10)

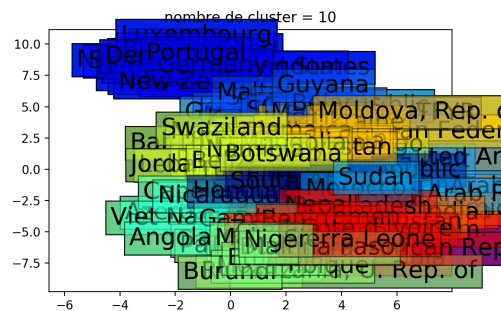


Figure 5: 2_Means

1.3 cluster optimal

afin de déterminer le nombre optimal de cluster \mathbf{k} , nous avons utiliser la méthode du *coude*. la courbe des coudes est réprésenter a figure 6.

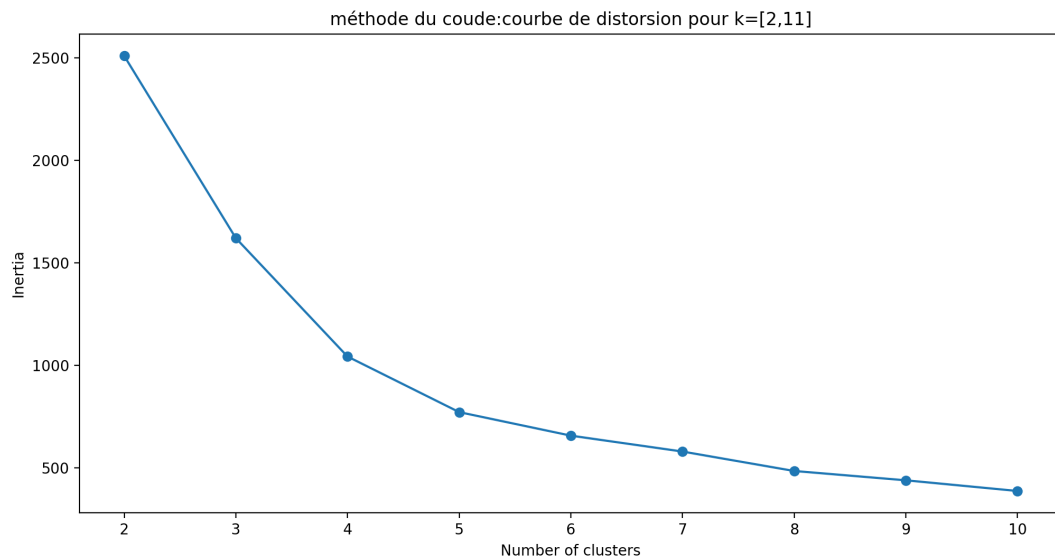


Figure 6: courbe de distorsion : methode du *coude*

on constate dans le graphique obtenu (figure 6), que le coude est situé à $\mathbf{k} = 4$, ce qui montre que $k = 4$ est effectivement un bon choix pour cet ensemble de données.