

Licence de Mathématiques

3M248 - Projet

Emmanuel Barillot - 3370161

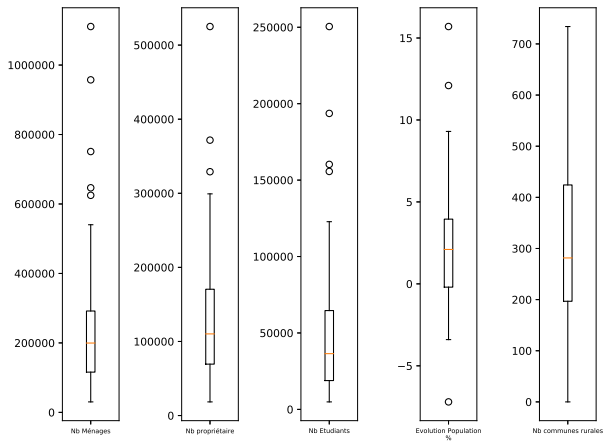
09 mai 2018

- Fichier de données INSEE de 2014 sur les 36000 communes françaises
- traitement des données brutes
 - suppression des variables non retenues
 - agrégation des données par département
 - séparation des variables en plusieurs familles
 - salaires
 - démographie
 - économie
- Idée : brosser un portrait des départements en fonction des trois aspects
- Après essais, une variable a semblée intéressante : la ruralité
 - calculée à partir du nb de communes de moins de 10000 hab dans chaque département

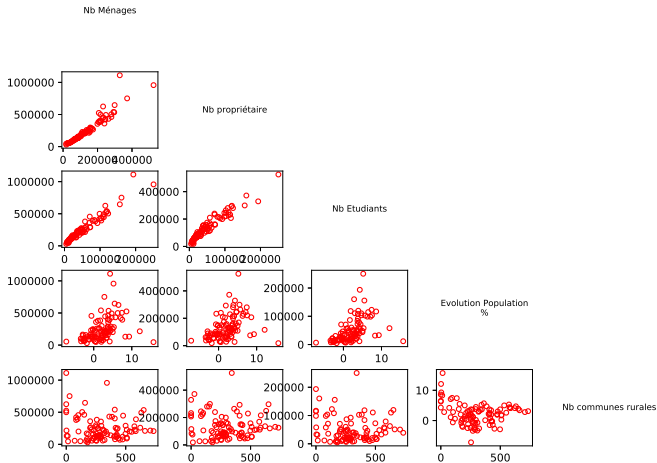
- variables relatives aux salaires :
 - Moyenne Revenus Fiscaux Départementaux,
 - Moyenne Salaires Cadre Horaires,
 - Moyenne Salaires Prof Intermédiaire Horaires,
 - Moyenne Salaires Employé Horaires,
 - Moyenne Salaires Ouvrier Horaires (avec la faute d'orthographe, dans le fichier),
- variables relatives à la population d'entreprises et d'organismes :
 - Nb Education, santé, action sociale,
 - Nb Entreprises,
 - Nb Création Entreprises,
- variables relatives à la population :
 - Population,
 - Evolution Population,
 - Nb Ménages,
 - Nb propriétaire,
 - Nb Etudiants.

Démographie

Boxplot

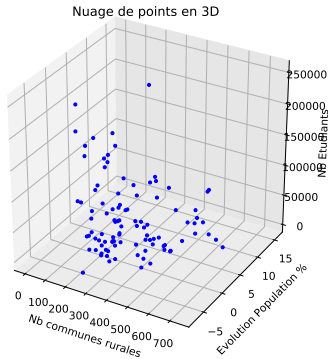


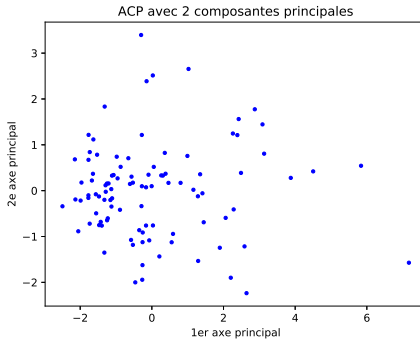
Plots démographie



Démographie

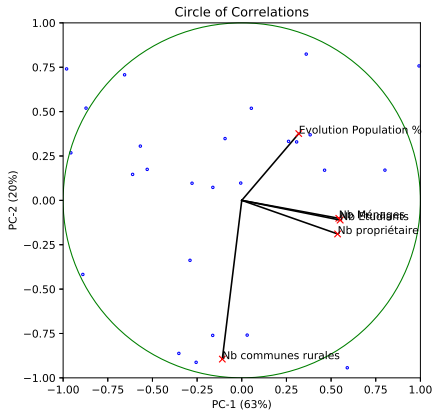
Scatter 3D





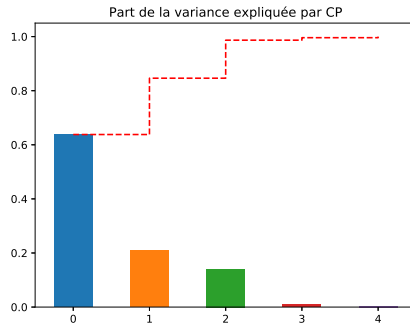
Démographie

Cerle des corrélations



Démographie

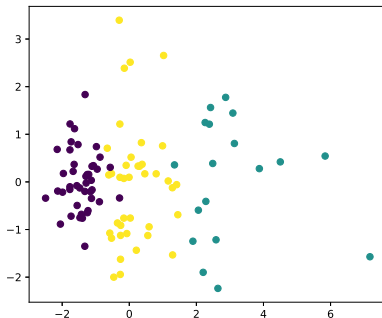
Variance



Démographie

Classification : K-means

K-Means clustering on PCA-reduced, K=3



Démographie

Regroupement hiérarchique ascendant

