amU Maison méditerranéenne des sciences humaines et sociales Aix Marseille Université

Plateforme universitaire de données

# Plateforme Universitaire de Données d'Aix-Marseille Université

# PUD-AMU

pud-amu@univ-amu.fr

https://pud.mmsh.fr/

# PUD-AMU

(1er avril 2018)

**Référente scientifique**
Corine Eyraud
**Ingénieur d'études**
Clément De Belsunce



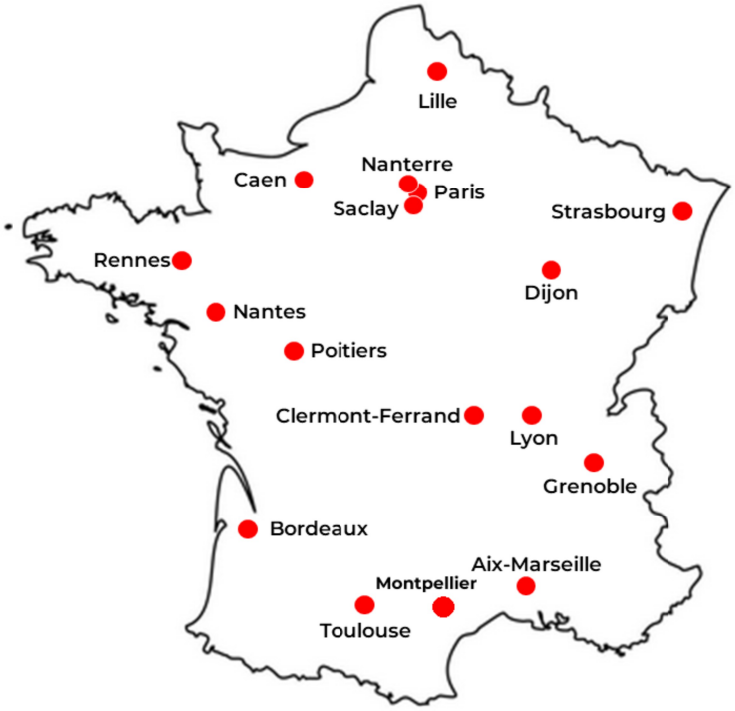
**Localisation**
Maison Méditerranéenne des Sciences de l'Homme
Bureau 107, bâtiment C.

# IR* PROGEDO

Progedo est l'**infrastructure nationale pour les données quantitatives en sciences humaines et sociales**

| **Produire** | **Gérer** | **Promouvoir** |
|---|---|---|
| Grandes enquêtes<br>Bases de données | Archiver<br>Diffuser | Faire connaître<br>Accompagner |

# IR* PROGEDO




**PRO**duction
et **GE**stion
des **DO**nnées

DATA INFRASTRUCTURE

Plateformes Universitaires de Données (PUD)

# La PUD-AMU au service de quels publics ?

| | |
|---|---|
| CHERCHEURS | Articles, montage de projets … |
| ENSEIGNANTS-CHERCHEURS | Fournir des jeux de données quantitatives supports aux cours … |
| PERSONNELS ADMINISTRATIFS ET TECHNIQUES | Formations … |
| DOCTORANTS | Thèse, articles, formations … |
| ETUDIANTS DE MASTER | Mémoires, projet de thèse … |

# Les missions de la PUD-AMU

- **Accompagnement individuel**
- Accompagnement collectif
- Promotions des bonnes pratiques autour des données en SHS
- Organisation de manifestations scientifiques

# Accompagnement individuel

## Pour connaître et accéder aux données

- Identification des bases : choix des sources, des diffuseurs/producteurs
- Exploiter les nouveaux outils numériques de recherche de données
- Démarches/procédures d'accès



## Pour traiter les données



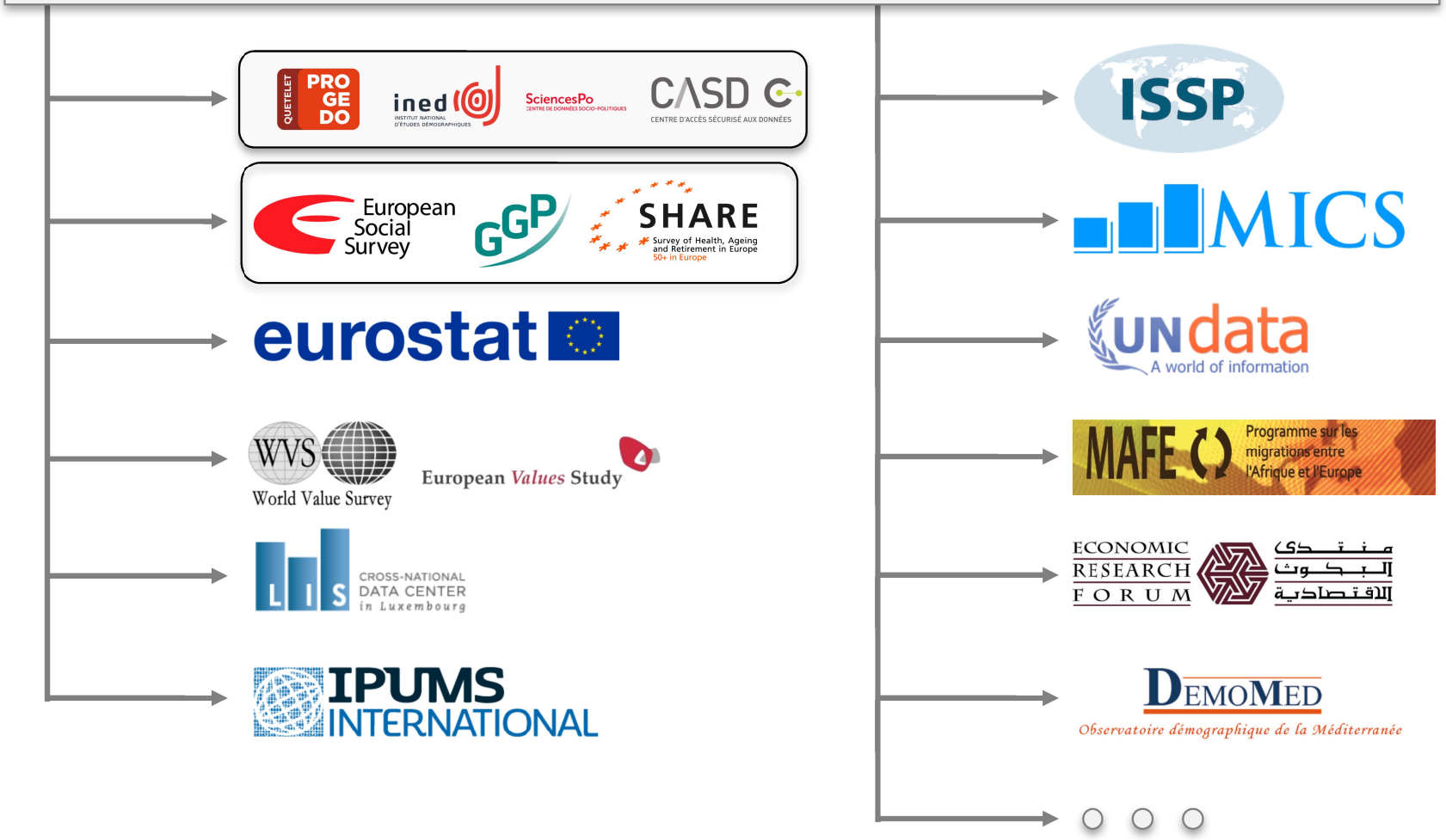
## Pour interpréter les résultats





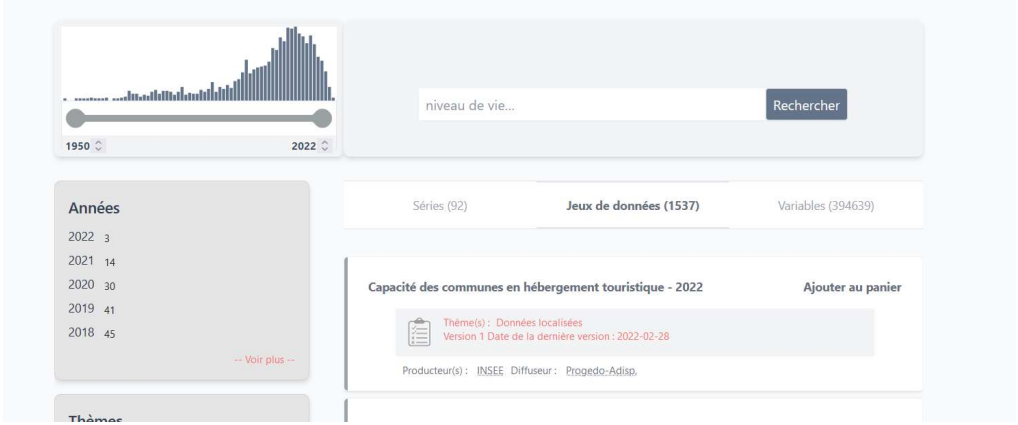
## Pour déposer ses données

# Plateforme Universitaire de Données - AMU



# Comment accéder aux données ?




https://data.progedo.fr/

# Comment accéder aux données ?




https://ec.europa.eu/eurostat/fr/

# Accompagnement individuel

## Pour connaître et accéder aux données

- Identification des bases : choix des sources, des diffuseurs/producteurs
- Exploiter les nouveaux outils numériques de recherche de données
- Démarches/procédures d'accès



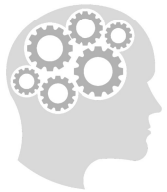


## Pour traiter les données

- Choix des méthodes adaptées
- Utilisation des logiciels informatiques pour le traitement des données



## Pour interpréter les résultats



## Pour déposer ses données

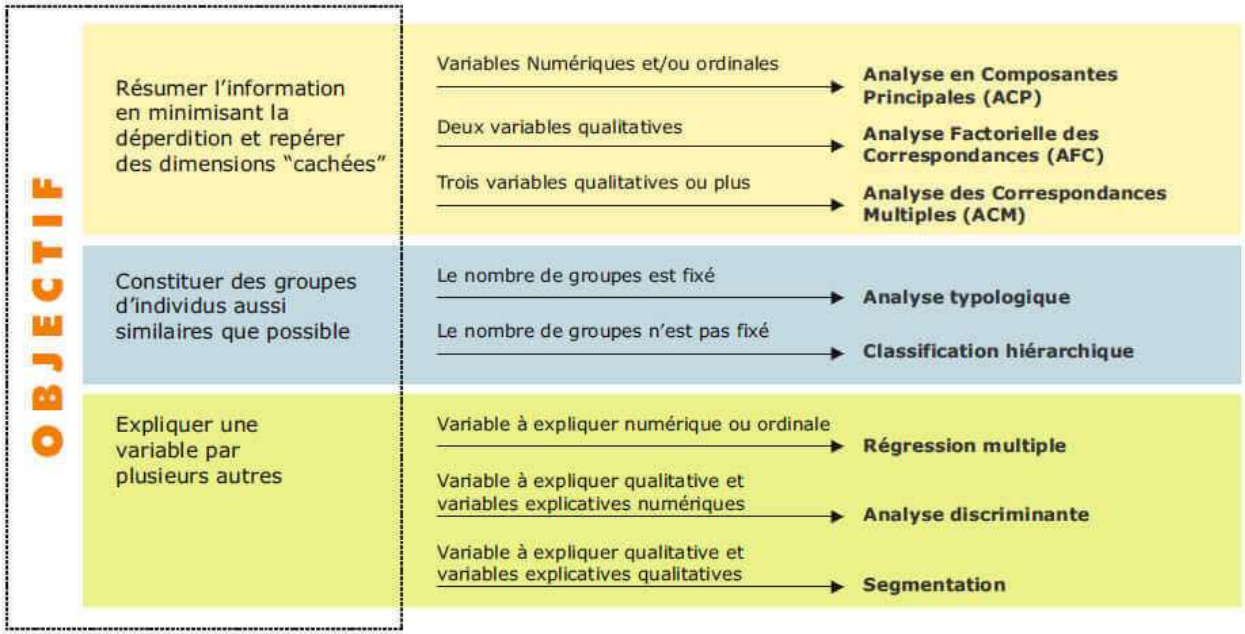
# Quelle méthode choisir ?

| Pays | Continent | PIB (M Mld $) | Taux de chomage | Indépendance (jour) | Indépendance (mois) | Indépendance (année) | Pop 1960 | Pop 1980 | Pop 2000 | Pop 2020 | Taux de fécondite en 2022 | Espérance de vie à la naissance Hommes | Espérance de vie à la naissance Femmes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Albanie | 3 | 13,04 | 11,8 | 29 | juillet | 1913 | 1636091 | 2682697 | 3129246 | 2877800 | 1,6 | 77 | 80 |
| Algérie | 1 | 167,55 | 12,7 | 5 | juillet | 1962 | 11057864 | 19221659 | 31042238 | 43851043 | 2,8 | 76 | 79 |
| Andorre | 3 | 3,01 | #N/A | | | 1813 | 13410 | 36063 | 65390 | 77265 | #N/A | #N/A | #N/A |
| Bosnie-Herzégovine | 3 | 18,05 | 15,2 | 5 | mars | 1992 | 3225664 | 4179850 | 3751176 | 3280815 | 1,2 | 75 | 80 |
| Bulgarie | 3 | 58,22 | 5,4 | 3 | mars | 1878 | 7885607 | 8878787 | 7997951 | 6948445 | 1,6 | 72 | 79 |
| Chypre | 3 | 22,05 | 6,1 | 16 | aout | 1960 | 572933 | 685408 | 943288 | 1207361 | 1,3 | 79 | 83 |
| Croatie | 3 | 55,21 | 8,7 | 25 | juin | 1991 | 4192641 | 4598125 | 4428075 | 4105268 | 1,4 | 76 | 82 |
| Egypte | 1 | 235,37 | 9,3 | 28 | février | 1922 | 26632891 | 43309063 | 68831561 | 102334403 | 3,2 | 70 | 75 |
| Espagne | 3 | 1314,31 | 14,7 | 21 | janvier | 1479 | 30402413 | 37698200 | 40824745 | 46754783 | 1,4 | 81 | 87 |
| France | 3 | 2583,5 | 8,1 | | aout | 843 | 45673147 | 53868006 | 59015092 | 65273512 | 1,8 | 80 | 86 |
| Grèce | 3 | 203,09 | 14,8 | 1 | janvier | 1822 | 8273628 | 9627004 | 11082103 | 10423056 | 1,3 | 80 | 85 |
| Israel | 2 | 353,27 | 5 | 14 | mai | 1948 | 2060476 | 3700916 | 5945949 | 8655541 | 2,9 | 82 | 85 |
| Italie | 3 | 1944,84 | 9,8 | 17 | mars | 1861 | 49699947 | 56349352 | 56692178 | 60461828 | 1,3 | 82 | 86 |
| Jordanie | 2 | 40,07 | 19,3 | 25 | mai | 1946 | 933102 | 2377997 | 5122495 | 10203140 | 2,6 | 73 | 77 |
| Liban | 2 | 53,58 | 14,5 | 23 | mai | 1926 | 1804935 | 2588928 | 3842774 | 6825442 | 2 | 77 | 81 |
| Libye | 1 | 38,1 | 19,6 | 24 | decembre | 1951 | 1448416 | 3219462 | 5357893 | 6871287 | 2,1 | 71 | 76 |
| Macédoine | 3 | 11,28 | 16,2 | 8 | septembre | 1991 | 1488670 | 1924203 | 2034823 | 2083380 | 1,5 | 74 | 78 |
| Malte | 3 | 12,52 | 3,5 | 21 | septembre | 1964 | 312973 | 331494 | 393649 | 441539 | 1,5 | 81 | 85 |
| Maroc | 1 | 109,7 | 11,5 | 2 | mars | 1956 | 12328532 | 19990006 | 28793672 | 36910558 | 2,3 | 76 | 79 |
| Monaco | 3 | #N/A | #N/A | | | 1861 | 22461 | 26752 | 32148 | 39244 | #N/A | #N/A | #N/A |
| Monténégro | 3 | 4,84 | 18,5 | 3 | juin | 2006 | 487416 | 580757 | 613558 | 628062 | 1,7 | 75 | 80 |
| Palestine | 2 | 14,5 | #N/A | 15 | novembre | 1988 | 1069318 | 1510717 | 3224009 | 5101416 | 3,4 | 73 | 76 |
| Portugal | 3 | 219,3 | 6,6 | 26 | juillet | 1139 | 8844638 | 9750966 | 10297117 | 10196707 | 1,3 | 80 | 85 |
| Serbie | 3 | 41,43 | 11,8 | 8 | mars | 1945 | 7556732 | 8908295 | 9487612 | 8737370 | 2,4 | 74 | 79 |
| Slovénie | 3 | 48,77 | 4,4 | 25 | juin | 1991 | 1586628 | 1835842 | 1987710 | 2078932 | 1,6 | 79 | 84 |
| Syrie | 2 | #N/A | 10,6 | 28 | septembre | 1961 | 4573514 | 8930776 | 16410847 | 17500657 | 2,7 | 79 | 84 |
| Tunisie | 1 | 39,95 | 16,8 | 25 | juillet | 1957 | 4178235 | 6374040 | 9708347 | 11818618 | 2,1 | 75 | 79 |
| Turquie | 3 | 851,55 | 13,4 | 24 | juillet | 1923 | 27472339 | 43975972 | 63240196 | 84339067 | 2 | 75 | 81 |

# Quelle méthode choisir ?

## Choisir une méthode d'analyse multivariée

Les différentes méthodes d'analyse multivariée permettent de répondre à des problématiques variées. Le choix d'une méthode dépend de l'objectif initial, des types de variables manipulées mais aussi, de la forme des résultats obtenus qui peuvent être plus ou moins faciles à présenter et à expliquer.



# Quelles fonctions utiliser ?

```r
install.packages("tidyverse")
install.packages("haven")
install.packages("FactoMineR")
install.packages("missMDA")
install.packages("Factoshiny")
install.packages("FactoInvestigate")
install.packages("openxlsx")

library(tidyverse)
library(haven)
library(FactoMineR)
library(missMDA)
library(Factoshiny)
library(FactoInvestigate)
library(openxlsx)

setwd("C:/Users/debelsunce/Downloads")

donnees1 <- read.xlsx("milieu.xlsx", sheet = 1, startRow = 1, colNames = TRUE,
                      rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
                      skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
                      namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)

res.mca1 = MCA(donnees1)

Investigate(res.mca1)

## le fichier originee ne marche pas##

donnees2 <- read.xlsx("originee2.xlsx")
res.mca2 = MCA(donnees2)
plot.MCA(res.mca2)

Investigate(res.mca2)

###################################

donnees3 <- read.xlsx("ambiance.xlsx", sheet = 1, startRow = 1, colNames = TRUE,
                      rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
                      skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
                      namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)

res.mca3 = MCA(donnees3)

Investigate(res.mca3)
```

```
> setwd("C:/Users/debelsunce/Downloads")
>
> donnees1 <- read.xlsx("milieu.xlsx", sheet = 1, startRow = 1, colNames = TRUE,
+                       rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
+                       skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
+                       namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)
>
> res.mca1 = MCA(donnees1)
>
```

# Accompagnement individuel

## Pour connaître et accéder aux données

- Identification des bases : choix des sources, des diffuseurs/producteurs
- Exploiter les nouveaux outils numériques de recherche de données
- Démarches/procédures d'accès





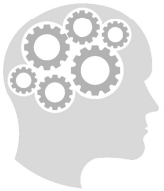
## Pour traiter les données

- Choix des méthodes adaptées
- Utilisation des logiciels informatiques pour le traitement des données



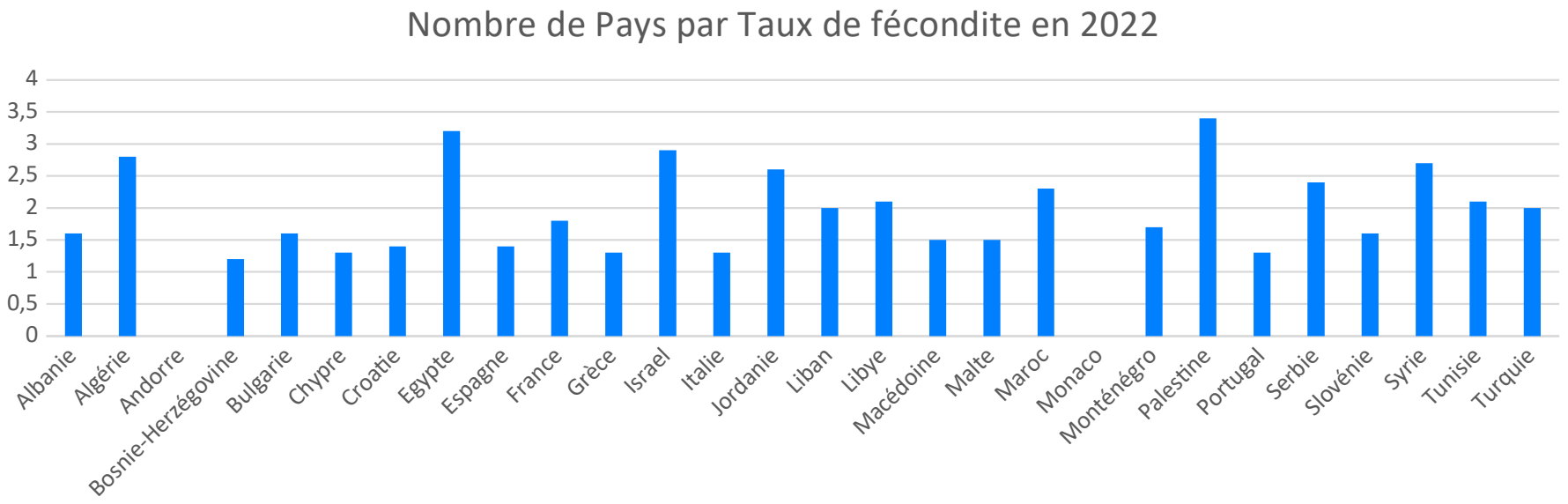
## Pour interpréter les résultats

- Interprétations des résultats pour répondre aux questions de recherche
- Valorisation graphique et mise en forme des résultats



## Pour déposer ses données

# Comment interpréter ses résultats ?



# Comment interpréter ses résultats ?

## 2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l' analyse expriment **13.66%** de l'inertie totale du jeu de données ; cela signifie que 13.66% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage très faible, et le premier plan ne représente donc seulement qu'une part de la variabilité contenue dans l'ensemble du jeu de données actif. Cette valeur est supérieure à la valeur référence de **10.45%**, la variabilité expliquée par ce plan est donc significative (cette intertie de référence est le quantile 0.95 de la distribution des pourcentages d'inertie obtenus en simulant 680 jeux de données aléatoires de dimensions comparables sur la base d'une distribution uniforme).

Du fait de ces observations, il serait alors certainement nécessaire de considérer également les dimensions supérieures ou égales à la troisième dans l'analyse.

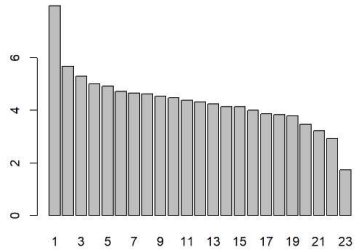



**Figure 2 - Decomposition of the total inertia on the components of the analyse**

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 5 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quantile 0.95 de distributions aléatoires (28.86% contre 25.19%). Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

## 3. Description du plan 1:2

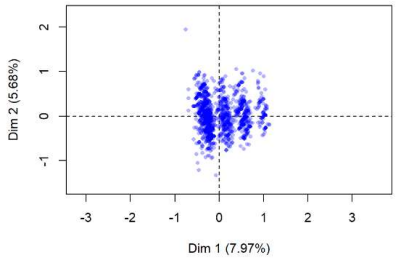
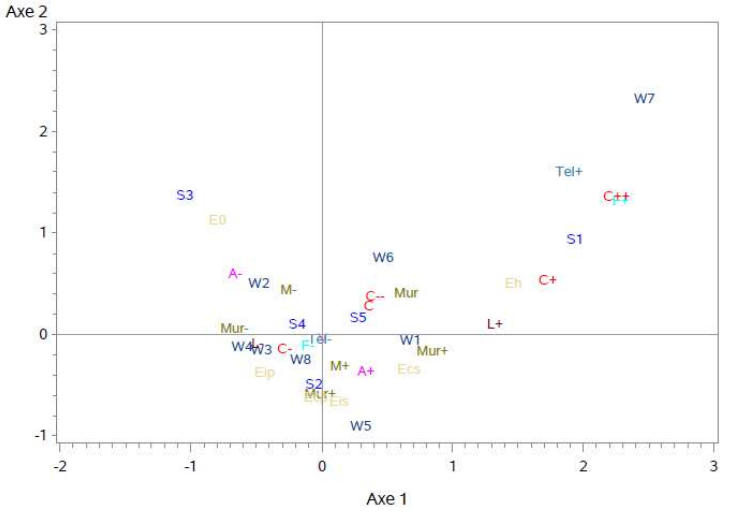



**Figure 3.1 - Graphe des individus (ACM)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.*



# Accompagnement individuel

## Pour connaître et accéder aux données

- Identification des bases : choix des sources, des diffuseurs/producteurs
- Exploiter les nouveaux outils numériques de recherche de données
- Démarches/procédures d'accès





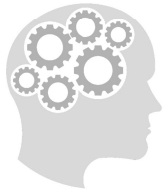
## Pour traiter les données

- Choix des méthodes adaptées
- Utilisation des logiciels informatiques pour le traitement des données



## Pour interpréter les résultats

- Interprétations des résultats pour répondre aux questions de recherche
- Valorisation graphique et mise en forme des résultats



## Pour déposer ses données

- Connaitre les principes de dépôt de données
- Adapter ses données
- Choisir le bon entrepôt

# Comment déposer ses données ?




Données quantitatives et **réutilisables**

QUETELET PRO GE DO

Tout type de données / simple **dépôt**

nakala

Autre solution, local ou national

Entrepôt de données recherche.data.gouv.fr

# Les missions de la PUD-AMU

- Accompagnement individuel
- **Accompagnement collectif**
- Promotions des bonnes pratiques autour des données en SHS
- Organisation de manifestations scientifiques

# Accompagnement collectif

- **Interventions personnalisées :**
  - Recherche de données selon une thématique
  - Initiation au traitement statistique selon une discipline
- **Formations** :
  - Traitement statistique : **R** (3 niveaux)
  - Gestion de données : **Excel**
  - Cartographie : **Magrit**
  - Questionnaire : **Limesurvey**
  - Analyse textuelle : **Iramuteq**
  - Codage **: langage Python**

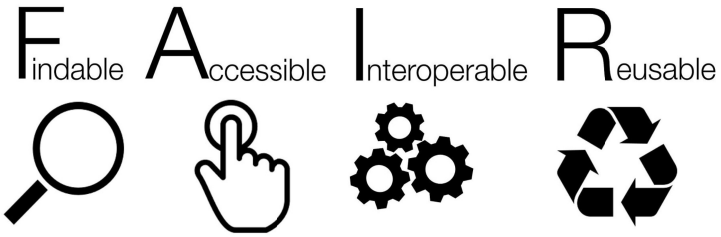
**La PUD-AMU peut aussi accompagner vos projets avec un suivi au plus long terme et des interventions techniques à différentes étapes dans l'approche autour de la donnée.**



# Les missions de la PUD-AMU

- Accompagnement individuel
- Accompagnement collectif
- **Promotions des bonnes pratiques autour des données en SHS**
- Organisation de manifestations scientifiques

# Promotions des bonnes pratiques autour des données en SHS




- Ouvrir la science
- Relais informatif dans le domaine des données
- Sensibilisation à la culture des données
- Aiguillage vers les services traitant des sujets précis (PGD, archivage, stockage…)

Aix Marseille Université

Huma-Num

# Les missions de la PUD-AMU

- Accompagnement individuel
- Accompagnement collectif
- Promotions des bonnes pratiques autour des données en SHS
- **Organisation de manifestations scientifiques**

amU Maison méditerranéenne des sciences humaines et sociales
Aix Marseille Université

Plateforme universitaire de données

# Merci de votre attention !

**pud-amu@univ-amu.fr**

**http://pud.mmsh.fr**