

Clustering

○○  
○○○○  
○○○  
○○

CAH

○○  
○  
○○○○

Agrégation autour de centres mobiles

○  
○  
○○

Données fonctionnelles

○○○○  
○○○○○○○○

# Alykis

Data | Information | Innovation

---

## Classification non supervisée de données fonctionnelles

---



Alykis  
[www.alykis.com](http://www.alykis.com)  
 @alykisdata

## Clustering



## CAH

○○  
○○○○  
○○

## Agrégation autour de centres mobiles

○  
○  
○○

## Données fonctionnelles

○○○○  
○○○○○○○○

---

# Classification

---

## Clustering



## CAH



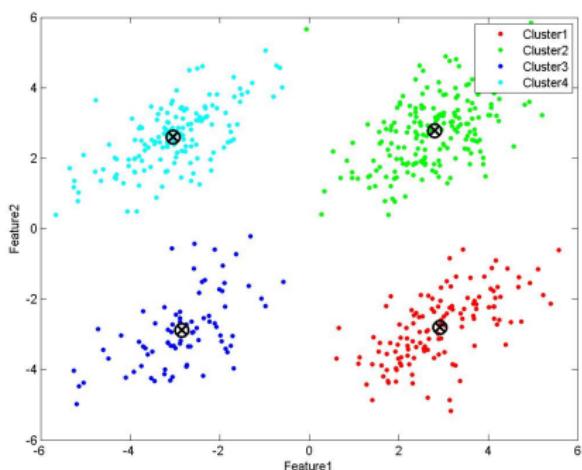
## Agrégation autour de centres mobiles



## Données fonctionnelles



# Classification



2 classifications :

- supervisée
- non supervisée

## Clustering



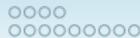
## CAH



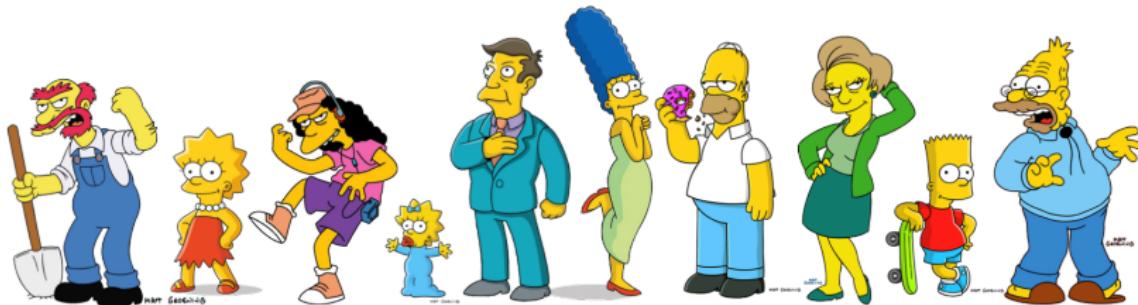
## Agrégation autour de centres mobiles



## Données fonctionnelles



Exemple : quelles sont les partitions possibles des personnages des Simpsons?



Clustering



CAH



Agrégation autour de centres mobiles



Données fonctionnelles



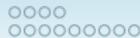
## Regrouper par sexe

Personnages Féminins



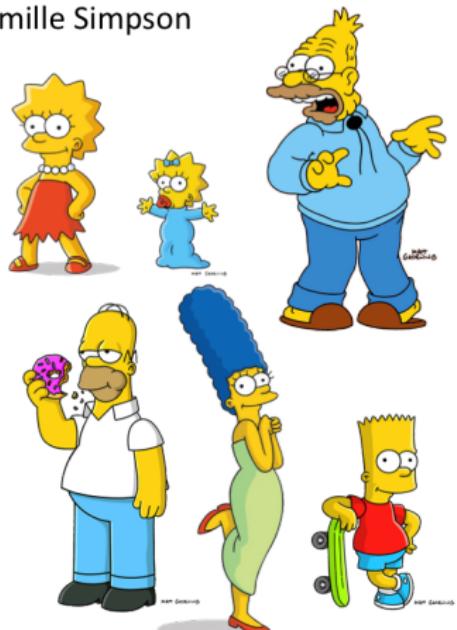
Personnages Masculins





## Regrouper par "famille"

Famille Simpson



Personnages de l'école



## Clustering



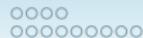
## CAH



## Agrégation autour de centres mobiles

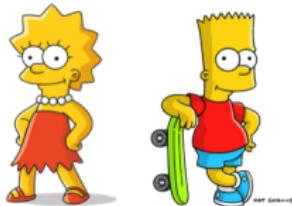


## Données fonctionnelles



# Regroupement par couleur de vêtement

Rouge



Bleu



Vert



Violet



Clustering



CAH



Agrégation autour de centres mobiles

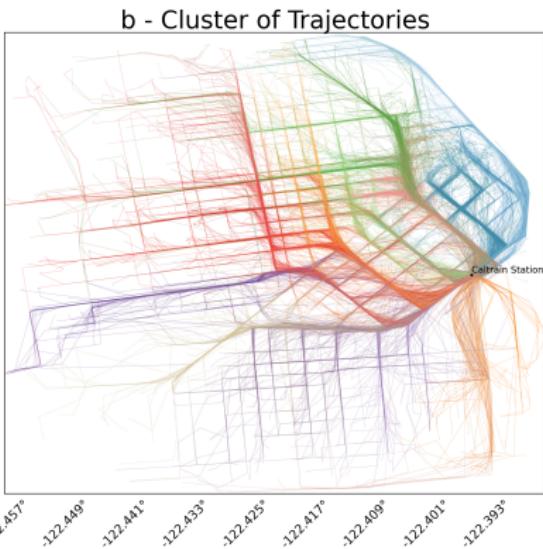
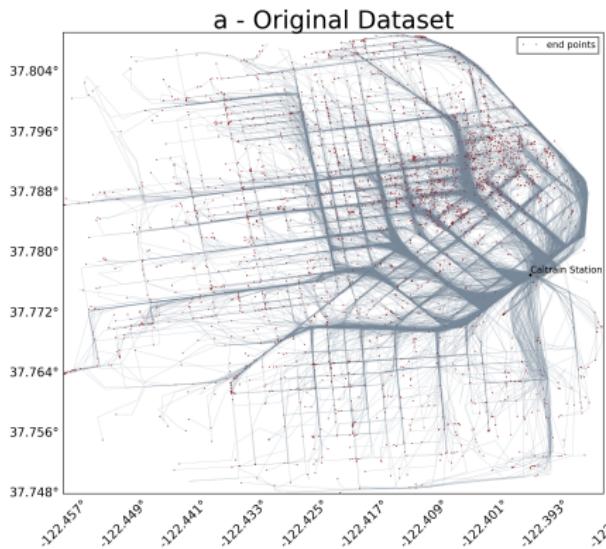


Données fonctionnelles



# Pourquoi classifier?

## 1. La structure





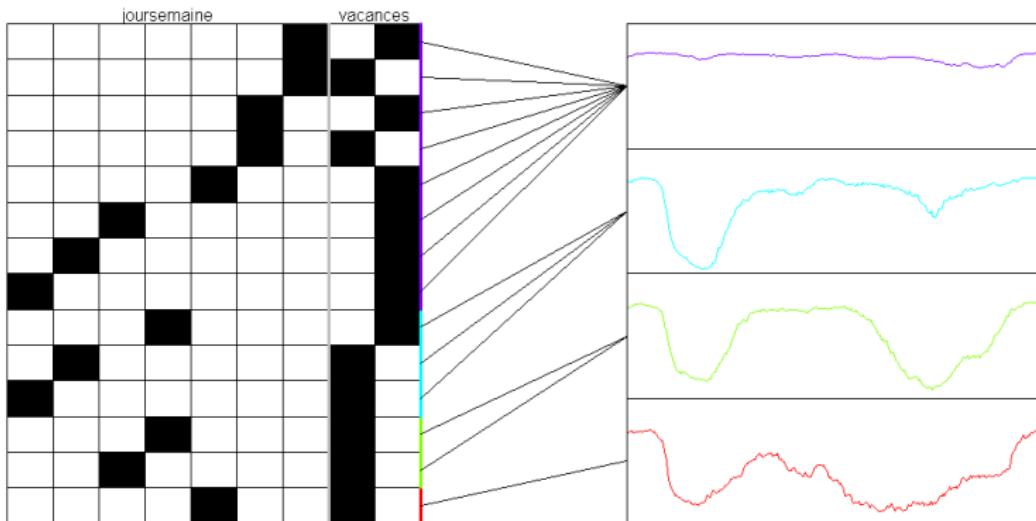
## Pourquoi classifier? 2. La segmentation





## Pourquoi classifier?

### 3. Résumer l'information



## Clustering



## CAH

oo  
ooo  
oooo

## Agrégation autour de centres mobiles

o  
o  
oo

## Données fonctionnelles

oooo  
oooooooo

---

# Formalisme

---

## Clustering



## CAH

○○  
○○○○  
○○○○○  
○●

## Agrégation autour de centres mobiles

○  
○  
○○

## Données fonctionnelles

○○○○  
○○○○○○○○

# Distance - Homogénéité - Algorithme

## Clustering



## CAH



## Agrégation autour de centres mobiles



## Données fonctionnelles



# Distance - Homogénéité - Algorithme

## Clustering



## CAH



## Agrégation autour de centres mobiles



## Données fonctionnelles



# Distance - Homogénéité - **Algorithme**

Clustering

○○  
○○○○  
○○○○○  
○○

**CAH**

○○  
○  
○○○○

Agrégation autour de centres mobiles

○  
○  
○○

Données fonctionnelles

○○○○  
○○○○○○○○

---

## Classification Ascendante Hiérarchique

---

Clustering



CAH



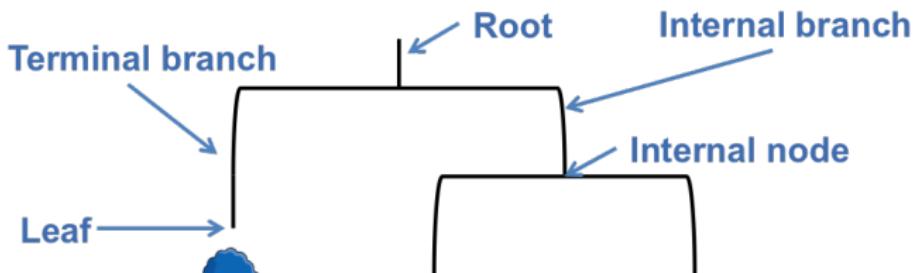
Agrégation autour de centres mobiles



Données fonctionnelles



## Dendrogramme



Clustering



CAH



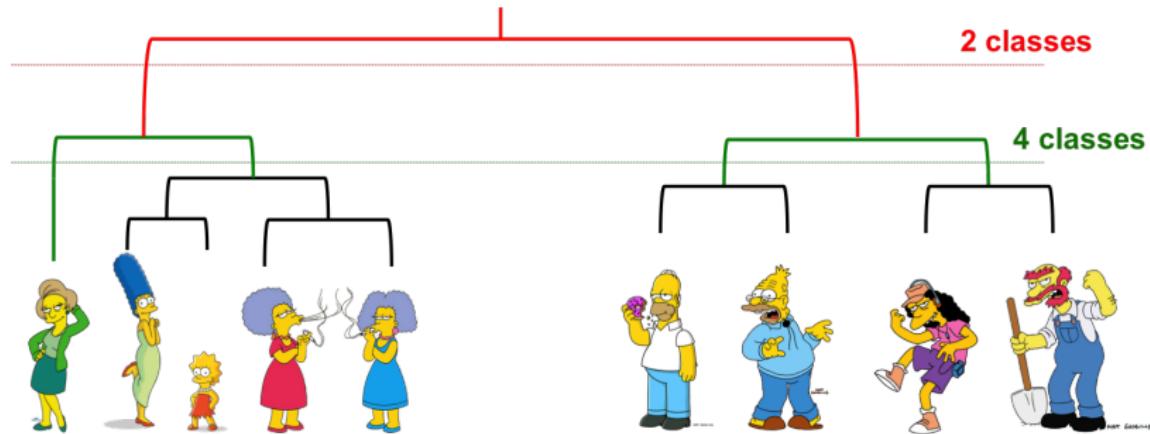
Agrégation autour de centres mobiles



Données fonctionnelles



## Exemples de troncatures



# Classification Ascendante Hiérarchique

## Algorithme

- Initialisation *Les classes initiales sont les singletons.*  
*Calculer la matrice de leurs distances deux à deux.*
- Iterer les deux étapes suivantes jusqu'à l'aggregation en une seule classe :
  - regrouper les deux classes les plus proches au sens de la "distance" entre classes choisie,
  - mettre à jour le tableau de distances en remplaçant les deux classes regroupées par la nouvelle et en calculant sa "distance" avec chacune des autres classes.

Clustering



CAH



Agrégation autour de centres mobiles



Données fonctionnelles



## Exemple : crimes violents aux USA

	Murder	Assault	UrbanPop	Rape
Alabama	13.20	236	58	21.20
Alaska	10.00	263	48	44.50

Clustering



CAH



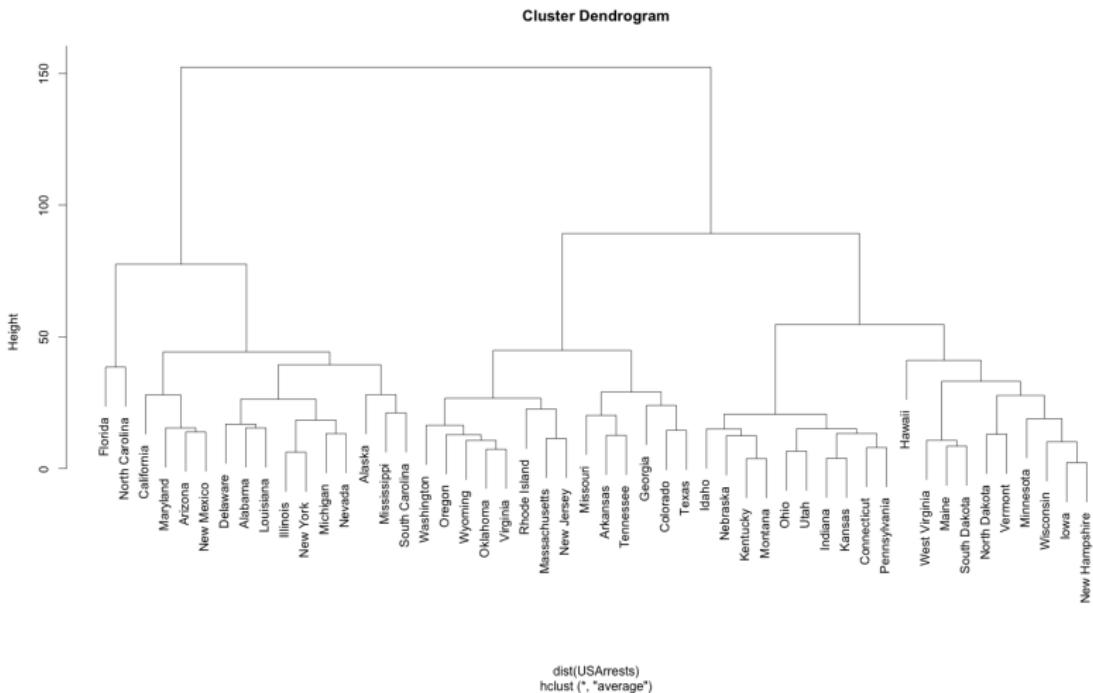
Agrégation autour de centres mobiles



Données fonctionnelles



## Exemple : crimes violents aux USA



Clustering



CAH



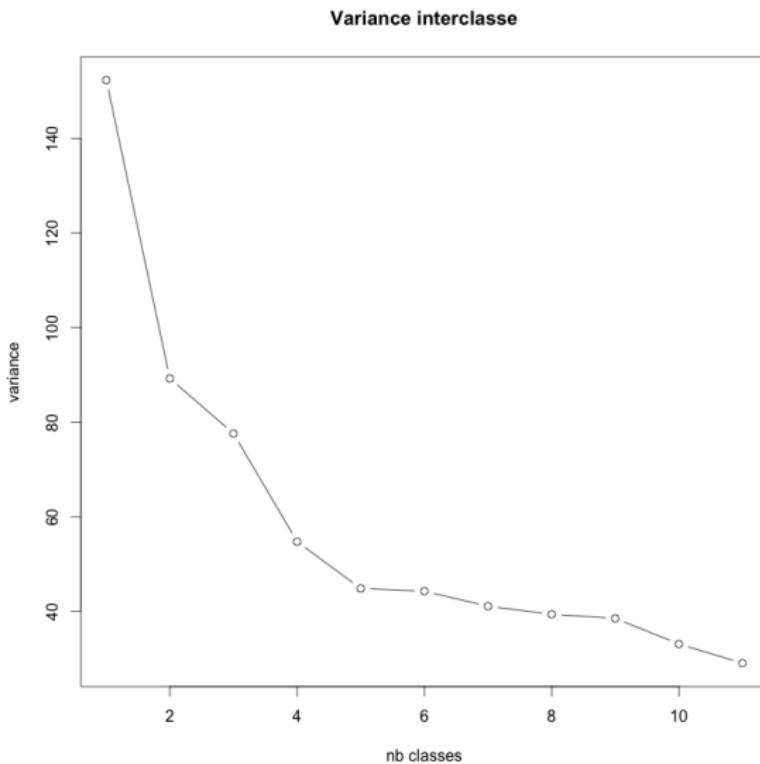
Agrégation autour de centres mobiles



Données fonctionnelles



## Exemple : crimes violents aux USA

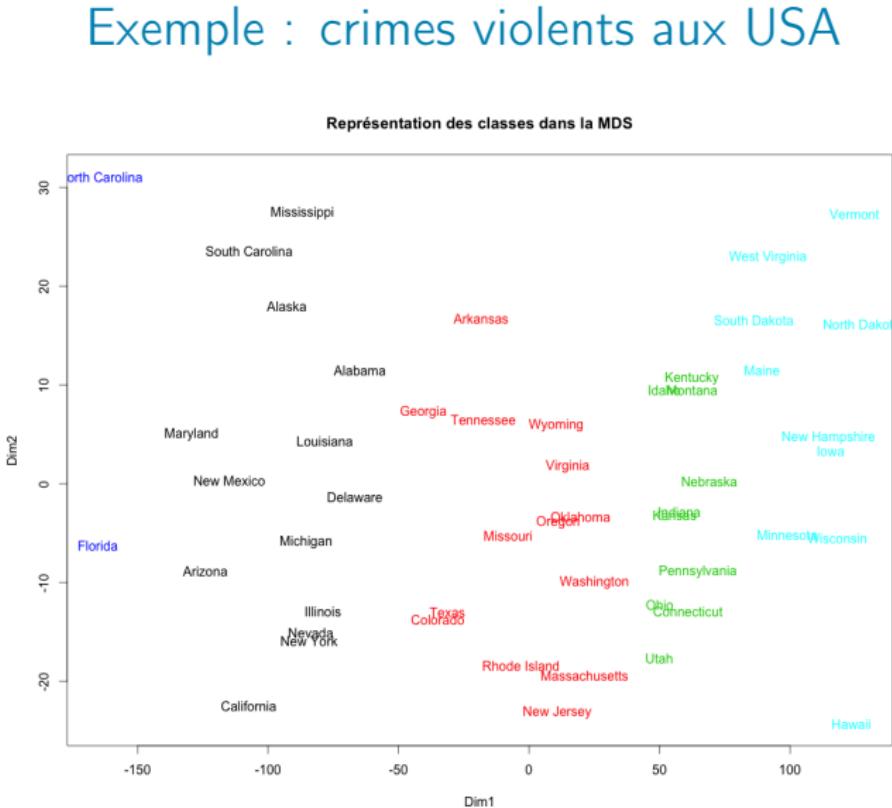


## Clustering

CAH

## Agrégation autour de centres mobiles

## Données fonctionnelles



Clustering

oo  
oooo  
ooo  
oo

CAH

oo  
o  
oooo

Agrégation autour de centres mobiles

o  
o  
oo

Données fonctionnelles

oooo  
oooooooo

---

## Agrégation autour de centres mobiles

---

Clustering



CAH



Agrégation autour de centres mobiles

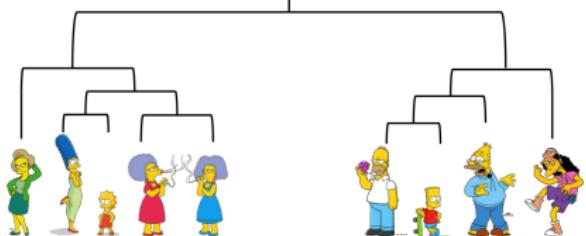


Données fonctionnelles



# Clustering

## Classification hiérarchique



## Clustering

CAH

1

## Agrégation autour de centres mobiles

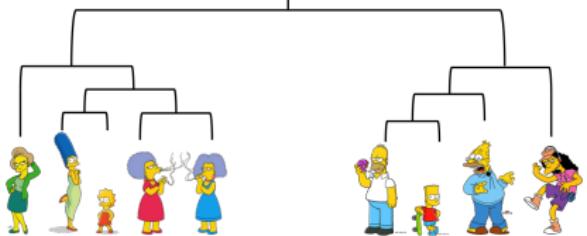
●

## Données fonctionnelles

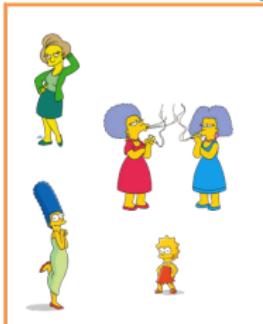
A 2x5 grid of 10 empty circles, arranged in two rows of five.

## Clustering

## Classification hiérarchique



## Classification par partition



Clustering



CAH



Agrégation autour de centres mobiles



Données fonctionnelles



## Kmeans

### Algorithme

- Initialisation *Tirer au hasard, ou sélectionner pour des raisons extérieures à la méthode, K points dans l'espace des individus, appelés centres ou noyaux.*
- Itérer les deux étapes suivantes, jusqu'à ce que le critère de variance inter-classe ne croisse plus de manière significative, ce qui signifie la stabilisation des classes.
  1. Allouer chaque individu au centre (donc à la classe) le plus proche au sens de la norme choisie ; on obtient ainsi, à chaque étape, une classification en K classes (ou moins si, finalement, une des classes devient vide).
  2. Calculer le centre de gravité de chaque classe : il devient le nouveau noyau. Si une classe s'est vidée, on peut éventuellement tirer aléatoirement un nouveau noyau complémentaire.

Clustering

○○  
○○○○  
○○○  
○○

CAH

○○  
○  
○○○○

Agrégation autour de centres mobiles

○  
○  
●○

Données fonctionnelles

○○○○  
○○○○○○○○

---

## Propriétés

---

Clustering



CAH



Agrégation autour de centres mobiles



Données fonctionnelles



## Convergence - Optimum local - Variantes

Clustering



CAH



Agrégation autour de centres mobiles



Données fonctionnelles



## Convergence - **Optimum local** - Variantes

Clustering



CAH



Agrégation autour de centres mobiles



Données fonctionnelles



## Convergence - Optimum local - **Variantes**

Clustering



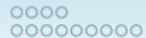
CAH



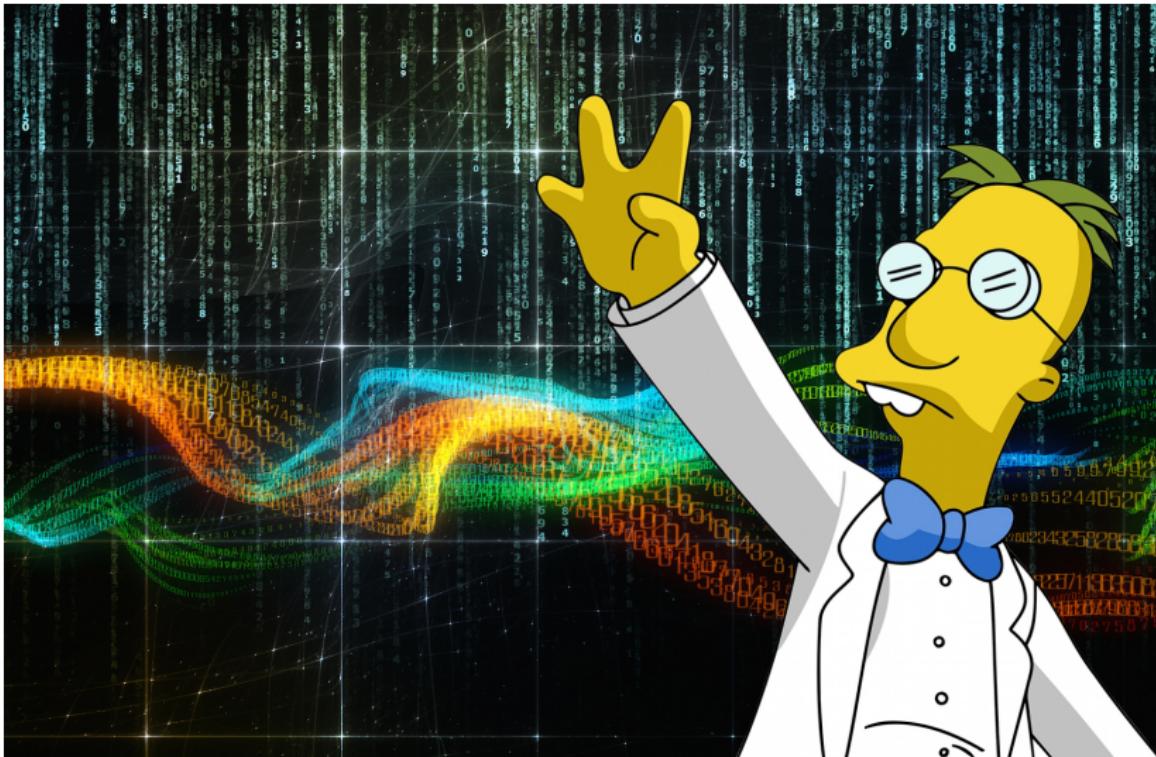
Agrégation autour de centres mobiles



Données fonctionnelles



## Données fonctionnelles



Clustering

```
oo
oooo
ooo
oo
```

CAH

```
oo
oo
o
oooo
```

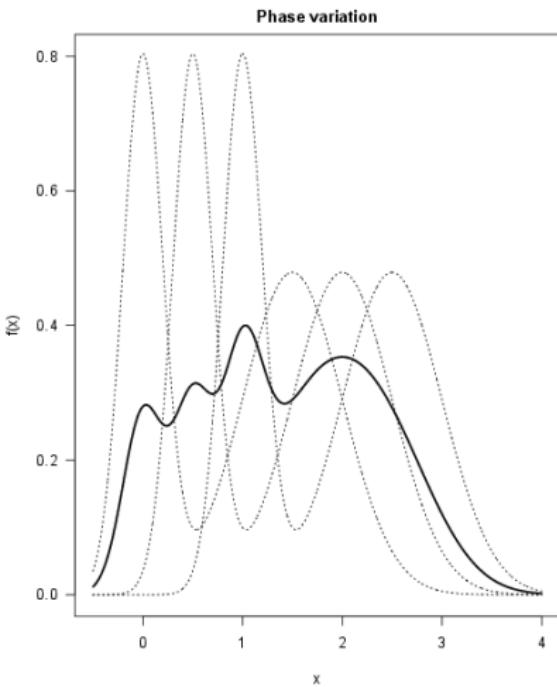
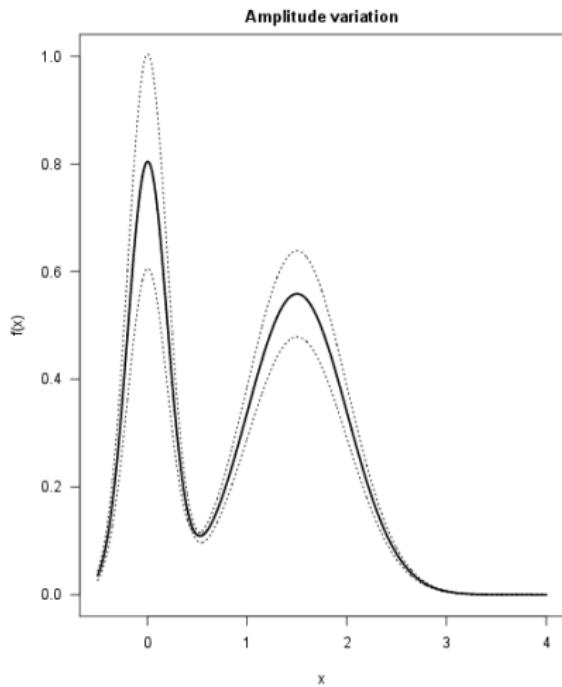
Agrégation autour de centres mobiles

```
o
o
oo
```

Données fonctionnelles

```
● ooo
oooooooooooo
```

## Problématique de décalages



Clustering



CAH



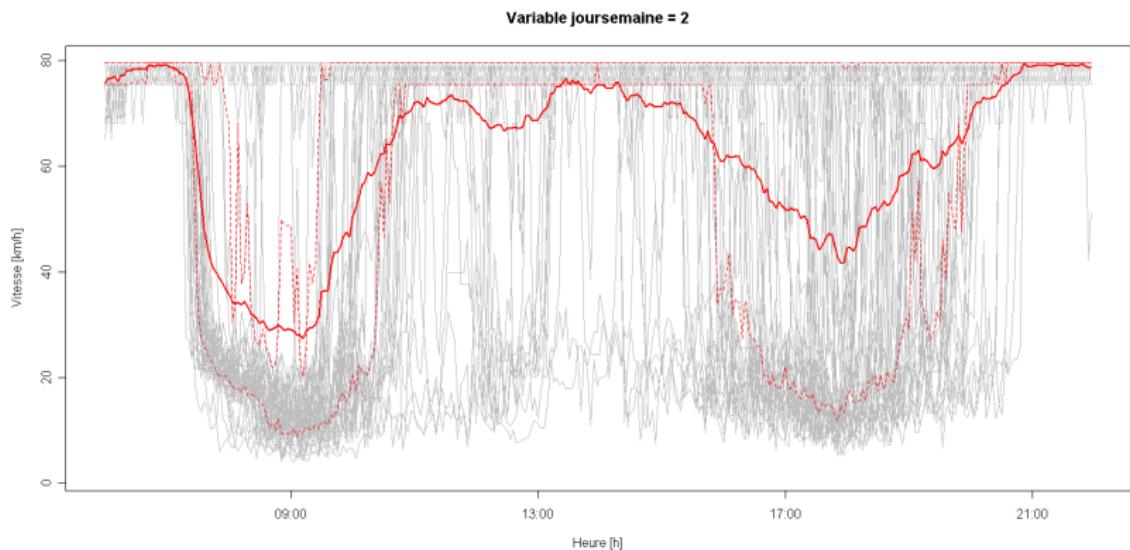
Agrégation autour de centres mobiles



Données fonctionnelles



## Exemple : trafic routier



Clustering



CAH



Agrégation autour de centres mobiles

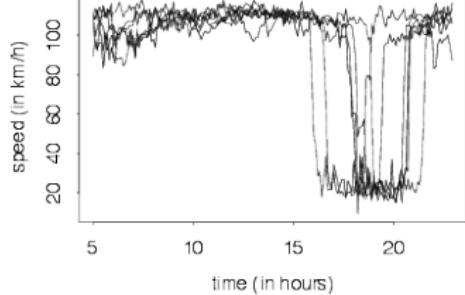


Données fonctionnelles

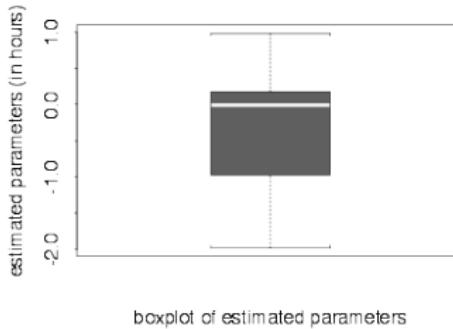


## La structure réelle d'un bouchon

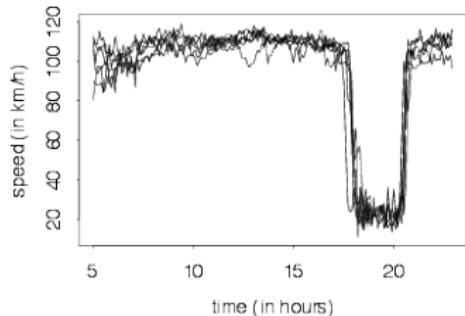
(a)



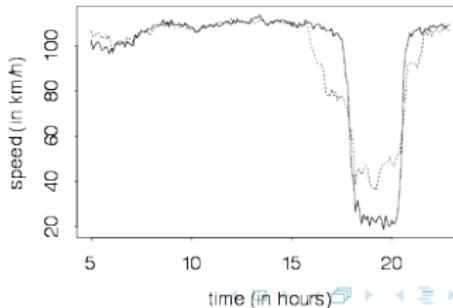
(b)



(c)



(d)



Clustering

○○  
○○○○  
○○○○○  
○○○○○○

CAH

○○  
○○○○  
○○○○○○

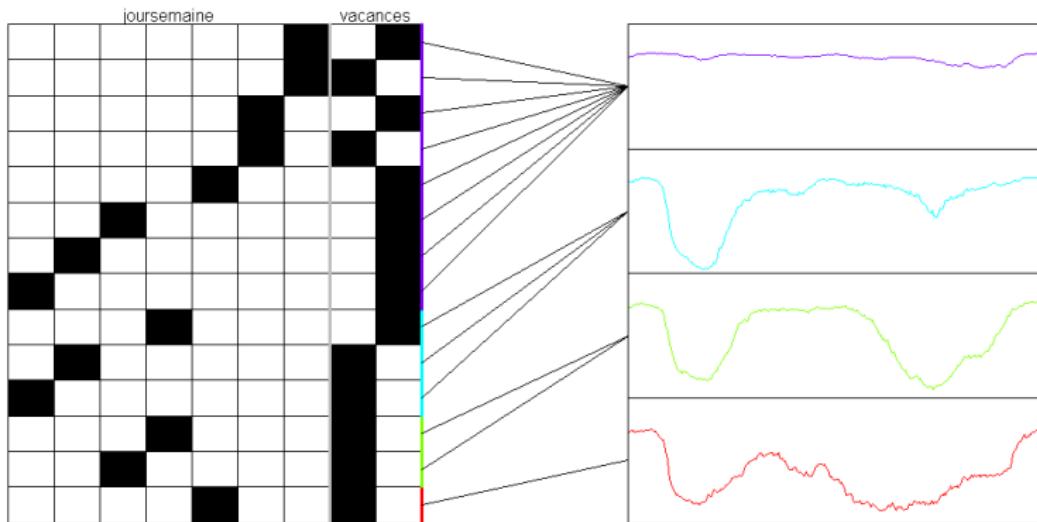
Agrégation autour de centres mobiles

○  
○○  
○○○

Données fonctionnelles

○○○●  
○○○○○○○○○

## Résultat d'un bon clustering sur cet exemple



Clustering



CAH



Agrégation autour de centres mobiles



Données fonctionnelles



---

# Recalage et Déformation Temporelle Dynamique

---

Clustering



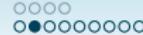
CAH



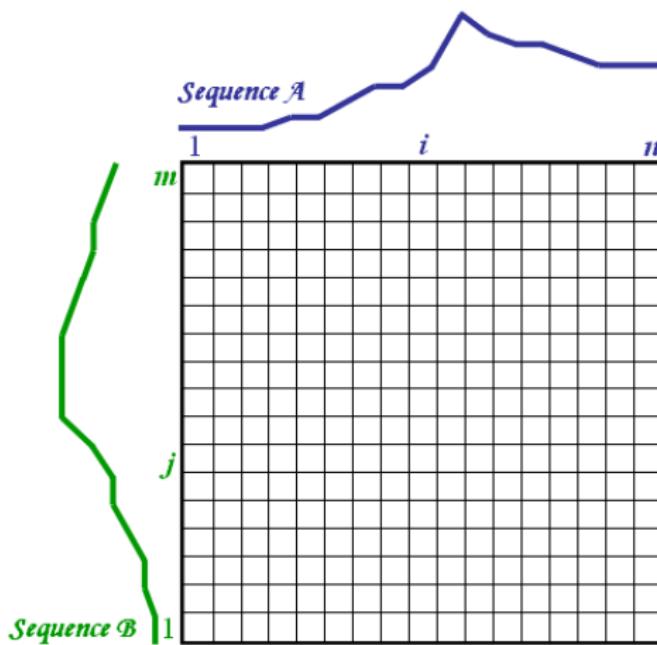
Agrégation autour de centres mobiles



Données fonctionnelles



## DTW : matrice de distances locales



Clustering



CAH



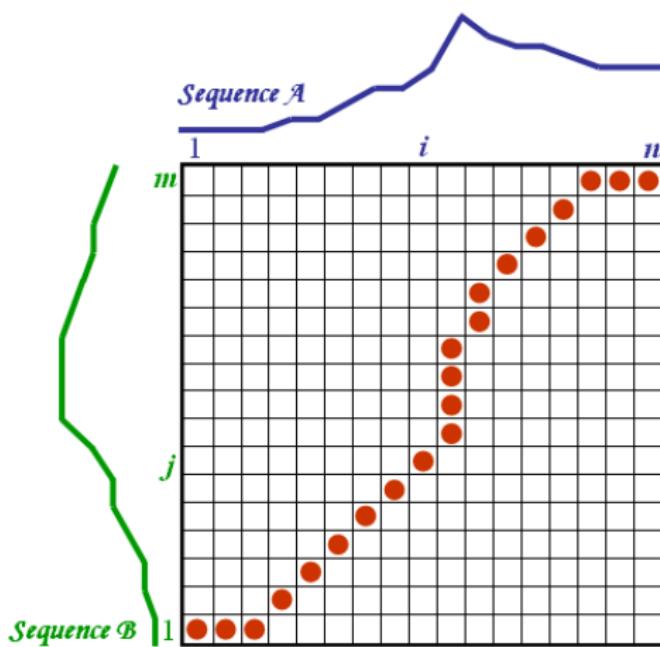
Agrégation autour de centres mobiles



Données fonctionnelles



## DTW : chemin optimal



## Clustering

CAH

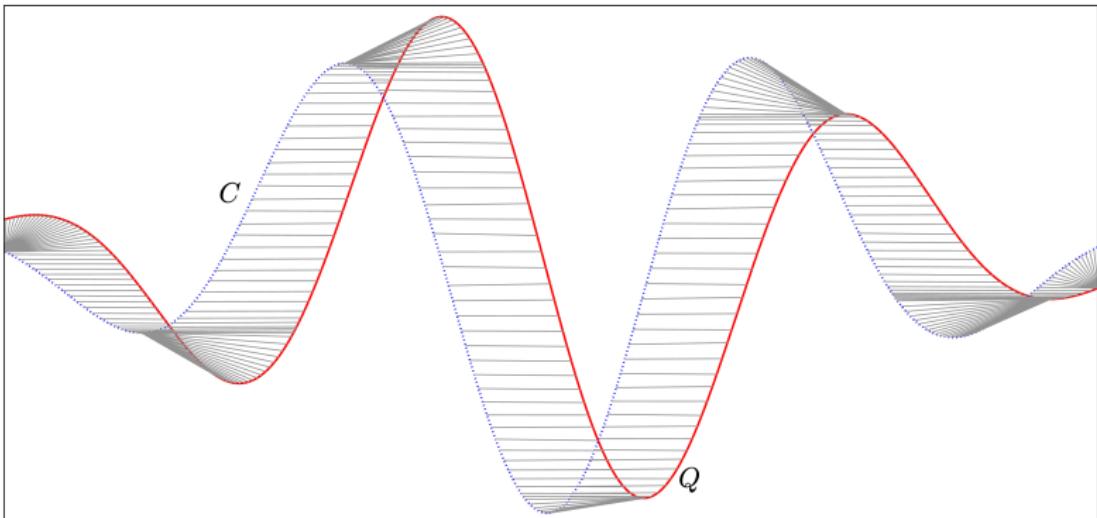
Agrégation autour de centres mobiles

## Données fonctionnelles

A 4x2 grid of eight circles, arranged in four rows and two columns.

3

## DTW : Distortion



Clustering

○○  
○○○○  
○○○○○  
○○

CAH

○○  
○  
○○○○

Agrégation autour de centres mobiles

○  
○  
○○

Données fonctionnelles

○○○○  
○○○○●○○○○

---

## Astuce du noyau

---

## Clustering

CAH

Agrégation autour de centres mobiles

## Données fonctionnelles

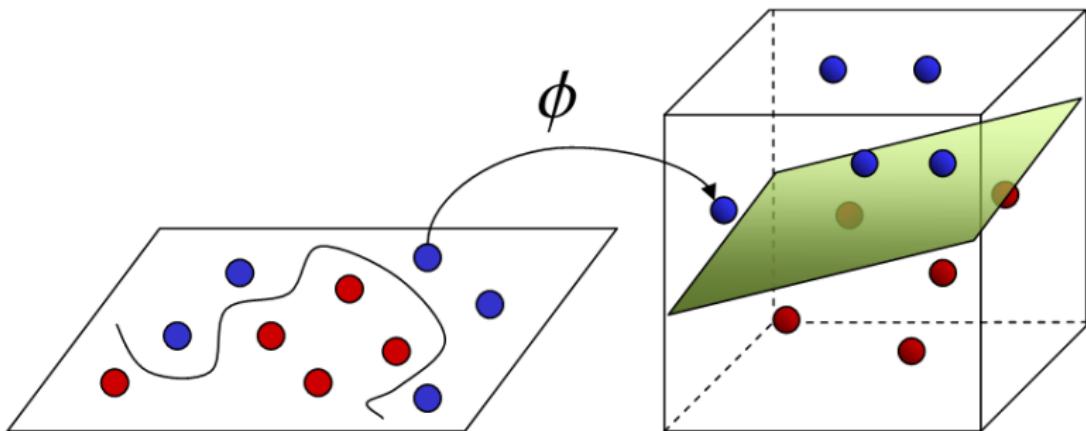
A 4x3 grid of 12 small circles, arranged in four rows and three columns.

1

100

10

## Kernel trick



## Input Space

# Feature Space

## Kmeans à noyaux

### Algorithmme

- *Initialisation : K clusters  $C_1, \dots, C_K$*
- *Pour tous les points  $X_i, i = 1, \dots, n$ , pour tous les clusters  $j = 1, \dots, K$ , on calcule  $\|\phi(X_i) - m_j\|^2$  et on trouve l'allocation de chaque observation :*  

$$\hat{q}(X_i) = \arg \min_{j=1, \dots, k} \|\phi(X_i) - m_j\|^2$$
- *On change la partition en modifiant tous les  $C_j, j = 1, \dots, K$  en  $C_j := \{x_i, \hat{q}(X_i) = j\}$ .*
- *Si les clusters n'évoluent plus, la solution est obtenue par les  $C_j$  ainsi obtenus et les barycentres  $m_j$ , sinon on recommence à l'étape 1.*

Clustering



CAH



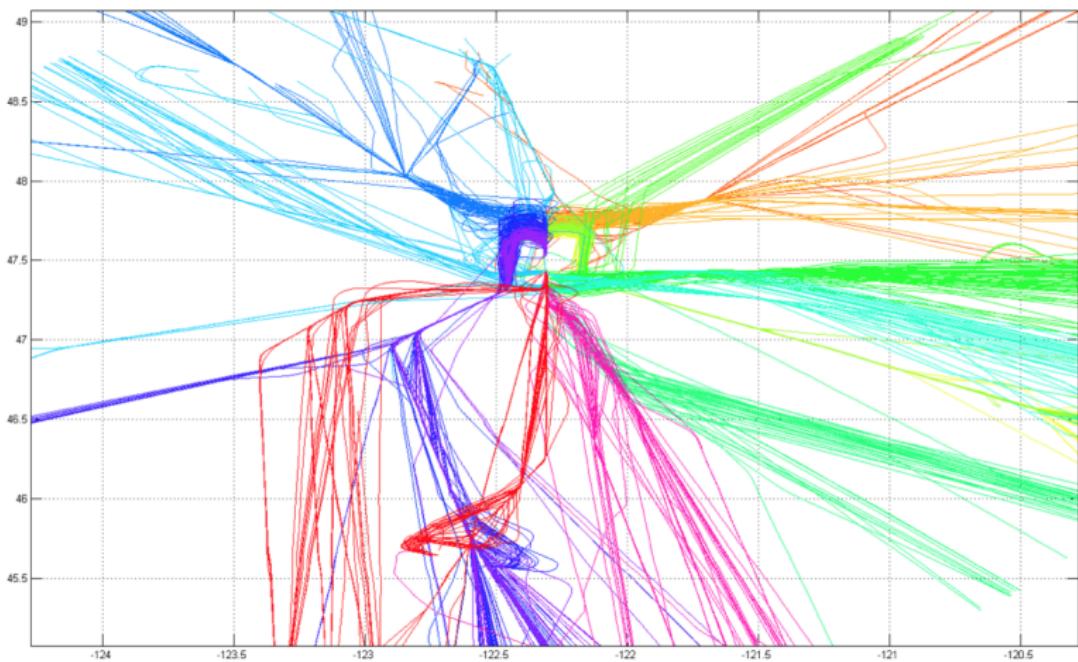
Agrégation autour de centres mobiles



Données fonctionnelles



## Clustering spectral



Clustering

○○  
○○○○  
○○○  
○○

CAH

○○  
○  
○○○○

Agrégation autour de centres mobiles

○  
○  
○○

Données fonctionnelles

○○○○  
○○○○○○○●

## Contact

Anna Choury

[anna.choury@alykis.com](mailto:anna.choury@alykis.com)

[www.alykis.com](http://www.alykis.com)

 @alykisdata