

Morphotypes du corps humain

Approche par clustering

Steve de Rose

Supervisé par :
Frédéric Bertrand (UTT)
Myriam Maumy Bertrand (UTT)
Philippe Meyer (UTT)

Table des matières

1 Contexte

- LabCom-DiTeX
- Problématique
- Science des données

2 Analyse topologique de données

- Comparer deux diagrammes
- Clustering

3 Données UCSC

- Distance de Wasserstein
 - Couple le plus éloigné
 - Couple le plus proche

4 Clustering à partir des distances

- Liaison complète
 - Détection des scans défectueux
- K-Médoïdes
 - Courbe Elbow
 - Résultats

5 Indice de classification par sexe

- Définitions

- Distance de Wasserstein

6 Vecteur de Silhouette

- Paysage de persistance
- Silhouette de persistance

7 Indice de classification par sexe (GDI)

- Vecteur de Silhouette
- Troncs
 - Distance de Wasserstein
 - Vecteur de Silhouette

8 Clustering results using Silhouette vectors

- K-Medoid clustering applied to the trunk points
 - Male dataset
 - Female dataset

9 Conclusions

- Perspectives

10 Bibliographie

Contexte : LabCom-DiTeX

Laboratoire-commun de R&D

- UTT
- Institut Français du Textile et de l'Habillement

Contexte : LabCom-DiTeX

Laboratoire-commun de R&D

- UTT
- Institut Français du Textile et de l'Habillement

Problématiques

Textile, Habillement, Mensurations → Statistique, Machine Learning

Contexte : LabCom-DiTeX

Laboratoire-commun de R&D

- UTT
- Institut Français du Textile et de l'Habillement

Problématiques

Textile, Habillement, Mensurations → Statistique, Machine Learning

Sujets

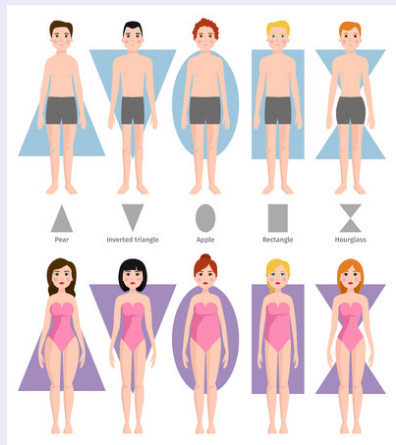
- Théorie du sondage : redressement
- Vieillissement des mensurations
- ...

→ **Types de morphologies**

Problématique : Morphotypes

La **classification des morphologies humaines** est une question importante pour l'industrie du textile et de l'habillement.

Permet une organisation appropriée des systèmes de dimensionnement des vêtements.

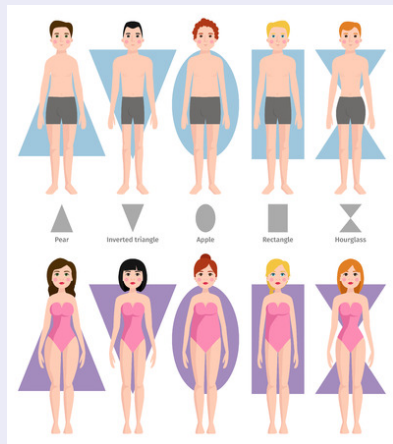


Problématique : Morphotypes

La **classification des morphologies humaines** est une question importante pour l'industrie du textile et de l'habillement.

Littérature

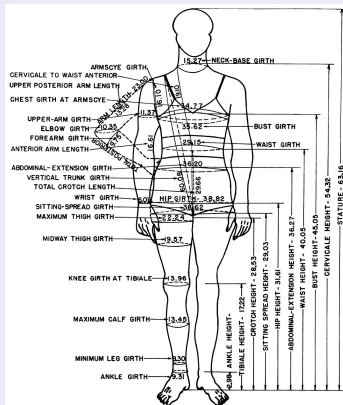
- différentes classifications
- dépend de la méthode utilisée et de la population étudiée.



Morphotypes : Science des données

Quel type de données utiliser ?

Mensurations

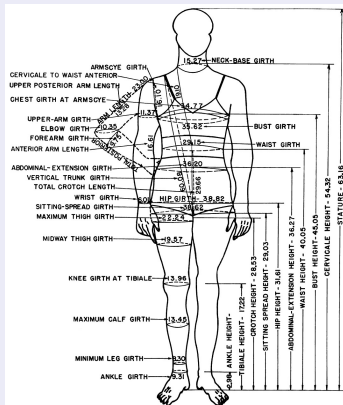


→ Longueurs

Morphotypes : Science des données

Quel type de données utiliser ?

Mensurations



→ Longueurs

Nuages de points anthropométriques



→ Points dans l'espace

Morphotypes : Clustering

Mensurations

Clustering dans un espace de grande dimension

Exemple : ANSUR II

- 4000 militaires hommes
- 4000 militaires femmes
 - ▶ 100 mensurations

Morphotypes : Clustering

Mensurations

Clustering dans un espace de grande dimension

Exemple : ANSUR II

- 4000 militaires hommes
- 4000 militaires femmes
 - ▶ 100 mensurations

Nuages de points anthropométriques

Comment les comparer ?

- ⇒ Analyse topologique des données (TDA)
analyse de données + topologie algébrique

Analyse topologique de données

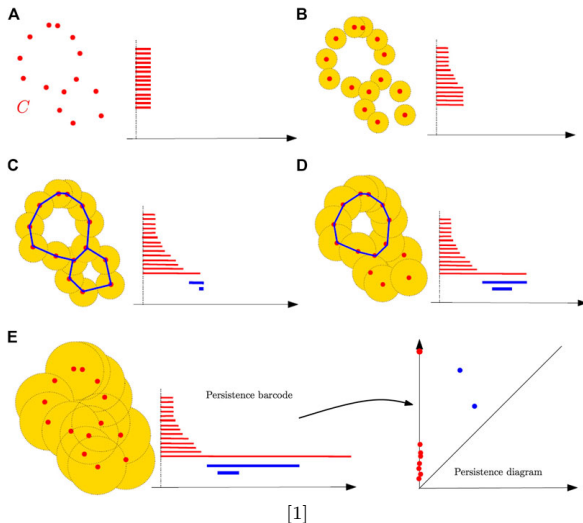
Topologie algébrique

Espace topologique X

→ **groupes d'homologie** $H_i(X)$ décrivant les “trous” de X .

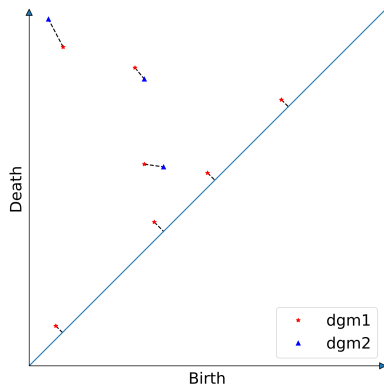
- $H_0(X)$: composantes connexes
- $H_1(X)$: lacets, *trous circulaires*
- $H_2(X)$: *vides, cavités*
- ...
- $H_i(X) = 0 \quad \forall i > \dim(X)$

Analyse topologique de données



Comment comparer deux diagrammes ?

Distances



γ : Couplage entre Dgm_1 et Dgm_2

- **p-Wasserstein :**

$$d_p(Dgm_1, Dgm_2) = \left(\sum_{x \in Dgm_1} d(x, \gamma(x))_2^p \right)^{1/p}$$

$$\downarrow p \rightarrow +\infty$$

- **Bottleneck :**

$$d_\infty(Dgm_1, Dgm_2) = \sup_{x \in Dgm_1} d(x, \gamma(x))_2$$

Clustering de diagrammes

Barycentre entre diagrammes ?

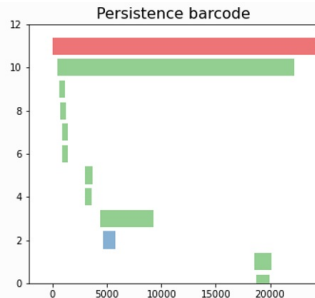
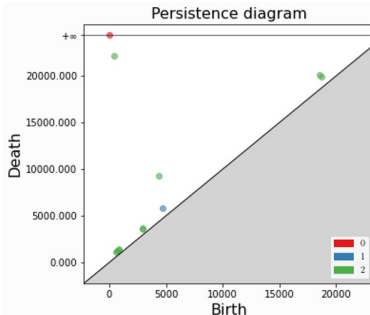
Pas évident^[2].

Méthodes ne nécessitant pas de barycentre

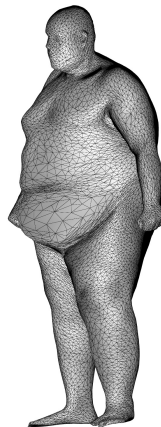
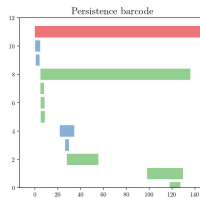
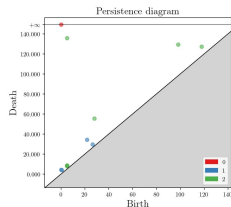
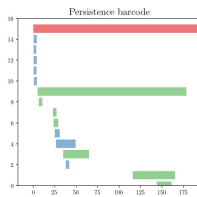
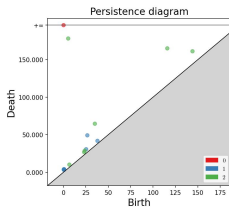
- K-Médoïdes
- Hiérarchique
- ...

Base CAESAR (Civilian American and European Surface Anthropometry Resource)

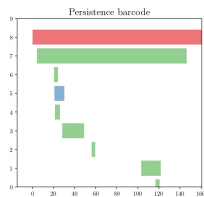
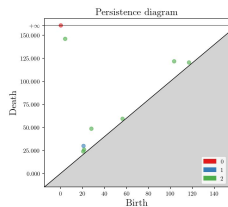
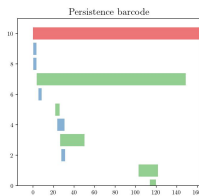
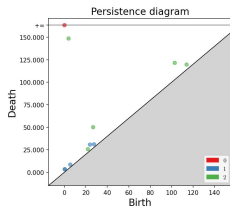
- 1500 hommes (+1500 femmes)
- Maillages triangulaires de 25000 faces et 12500 sommets
→ Nuages de 12500 points



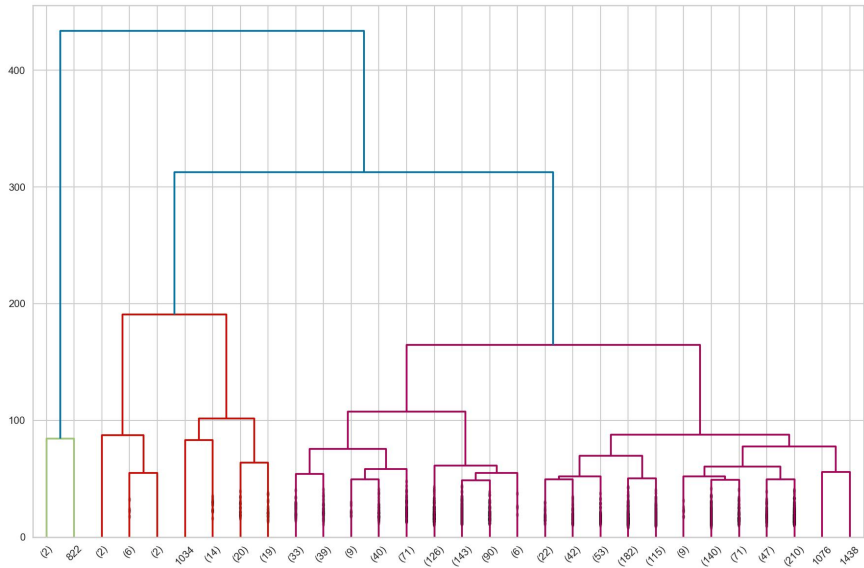
Couple le plus éloigné après normalisation



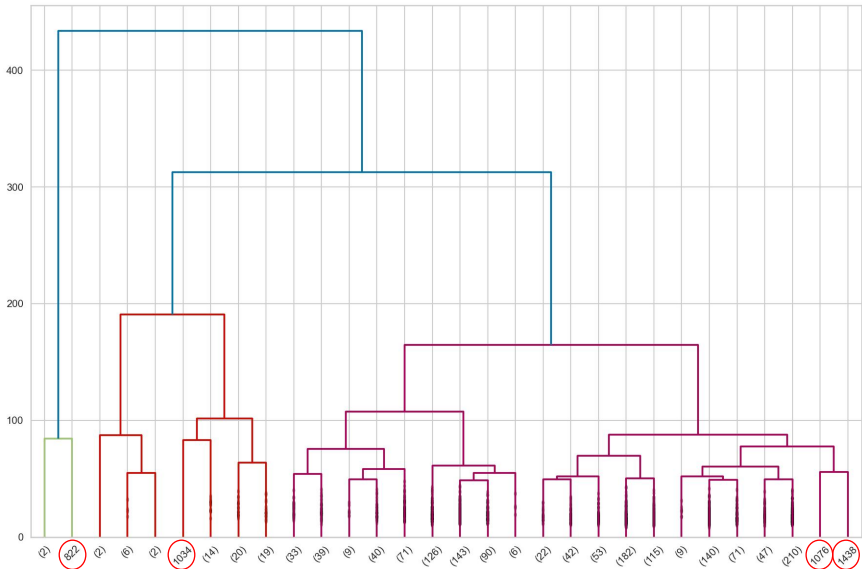
Couple le plus proche



Linkage complete

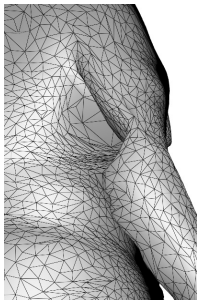


Linkage complete

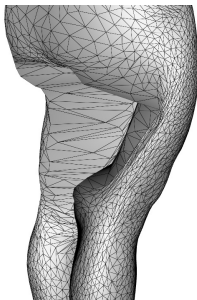


Scans défectueux

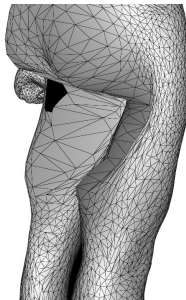
822



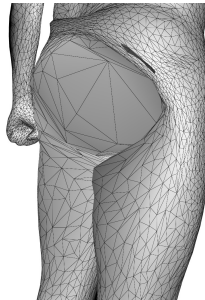
1034



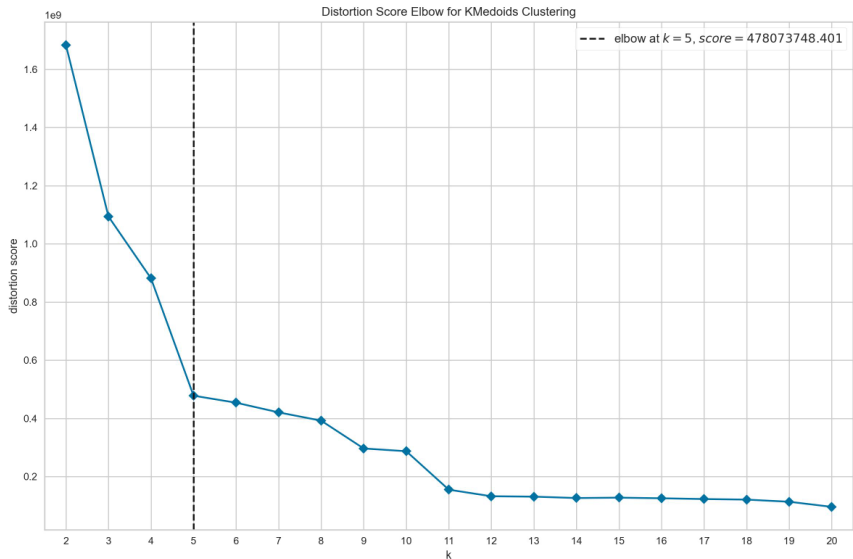
1076



1438



K-Médoides (PAM) - Courbe elbow



Résultats - K-Médoïdes

Cluster n°	1	2	3	4	5
Taille	144	575	34	417	347
Proportion	9,49%	37,9%	2,24%	27,49%	22,87%

1322

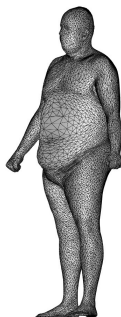
603

40

733

793

Médoïdes



Indice de classification par sexe (GDI)

Proportions

$$P_m(C) = \frac{n_m(C)}{s(C)}$$
$$P_f(C) = \frac{n_f(C)}{s(C)}$$

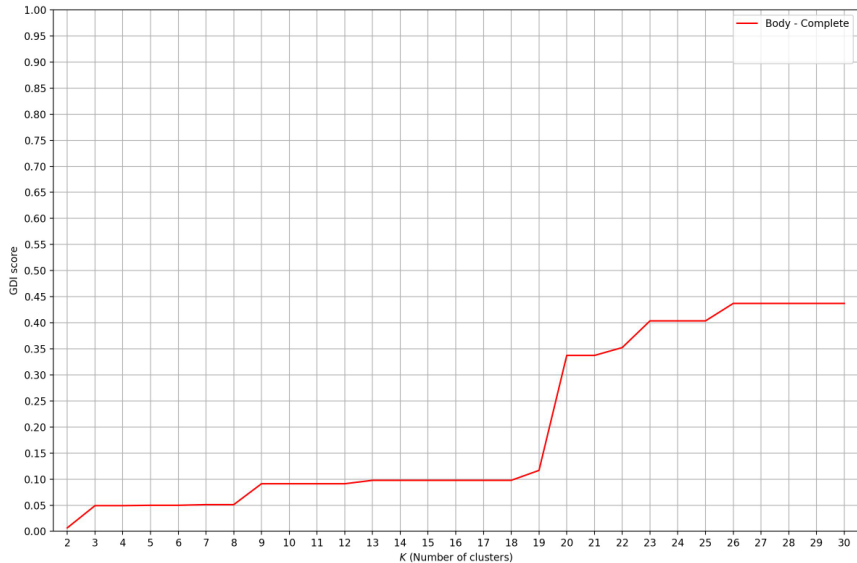
où $n_m(C)$ et $n_f(C)$ sont, respectivement, le nombre de diagrammes masculins et féminins dans C et $s(C)$ la taille de C .

GDI

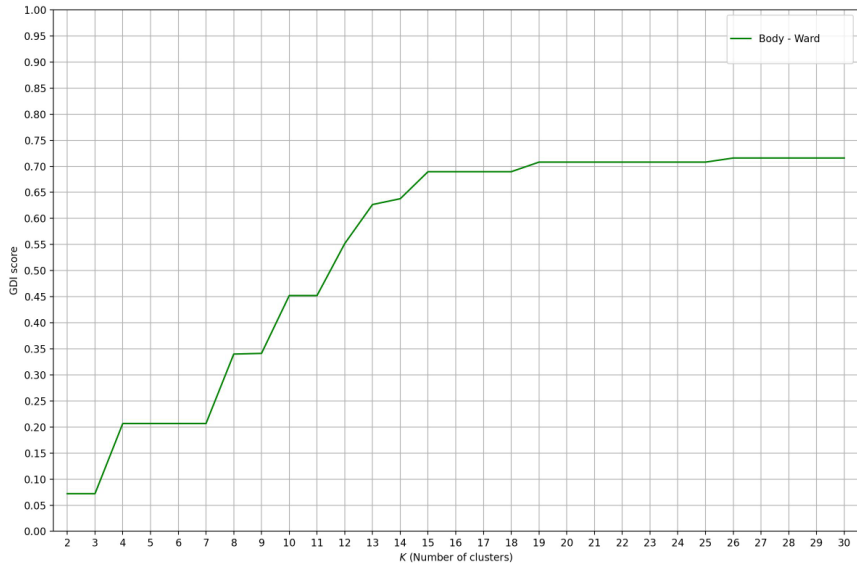
$$\text{GDI}(\mathcal{C}) = \frac{2}{s(D_{\text{MF}})} \sum_{k=1}^K s(C_k) \left| P_m(C_k) - \frac{1}{2} \right|$$

où K est le nombre de clusters de $\mathcal{C}$, C_k sont les clusters de $\mathcal{C}$ et $s(D_{\text{MF}})$ est le nombre de diagrammes dans D_{MF} .

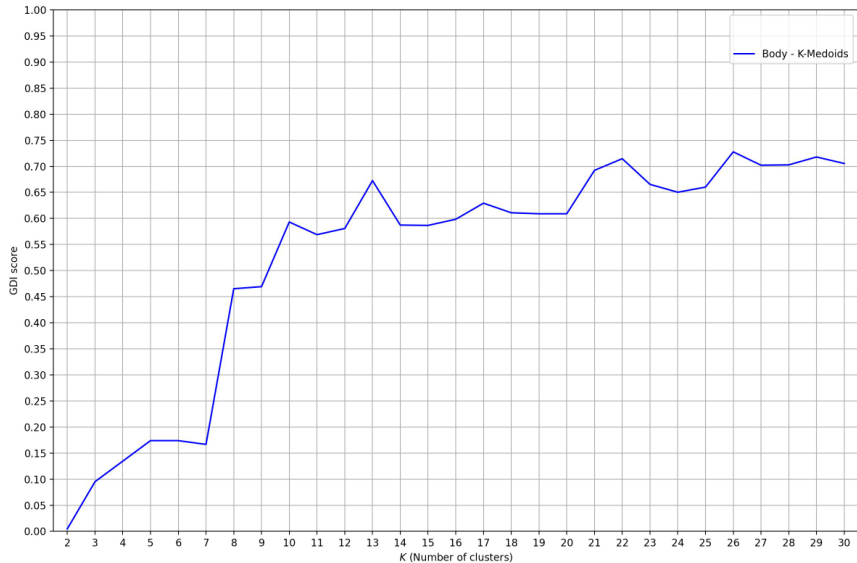
Distance de Wasserstein



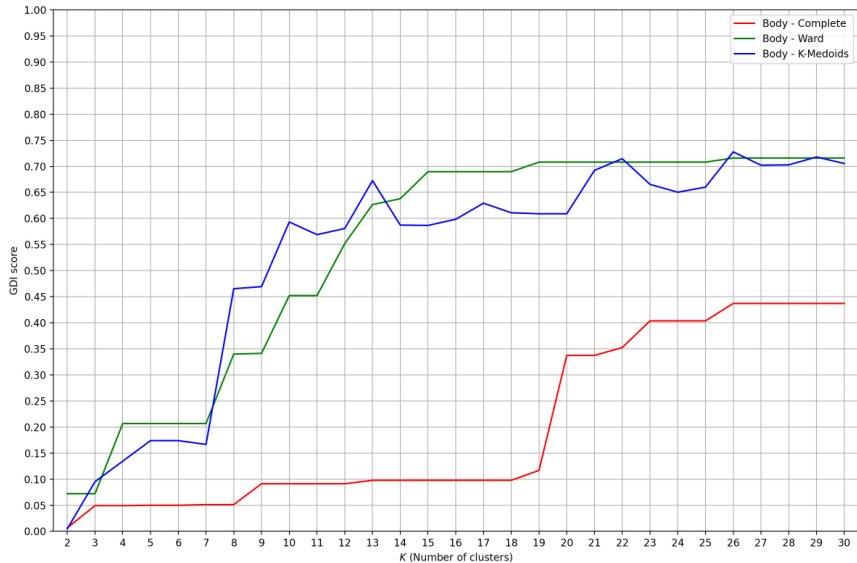
Distance de Wasserstein



Distance de Wasserstein



Distance de Wasserstein



Paysage de persistance

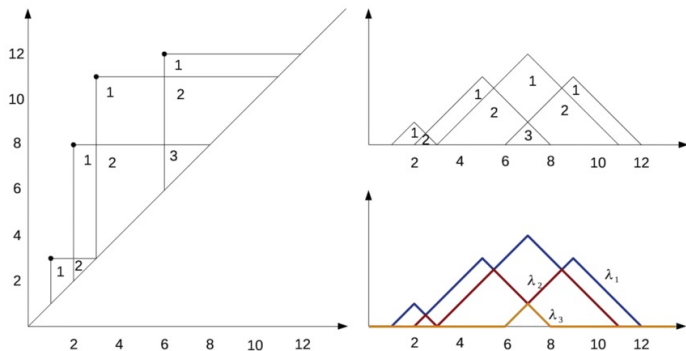
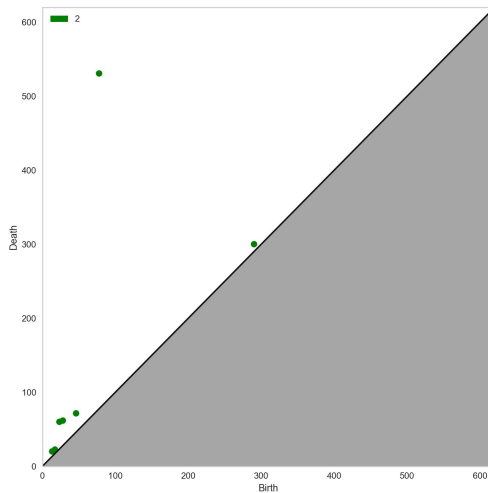


Figure – Le diagramme de persistance est incliné, de sorte que la diagonale devient le nouvel axe horizontal. Les λ_i sont les fonctions linéaires par morceaux^[3].

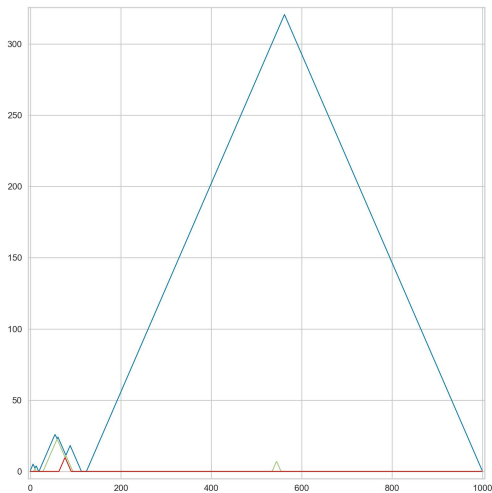
Silhouette de persistance

Figure – Diagramme de persistance.



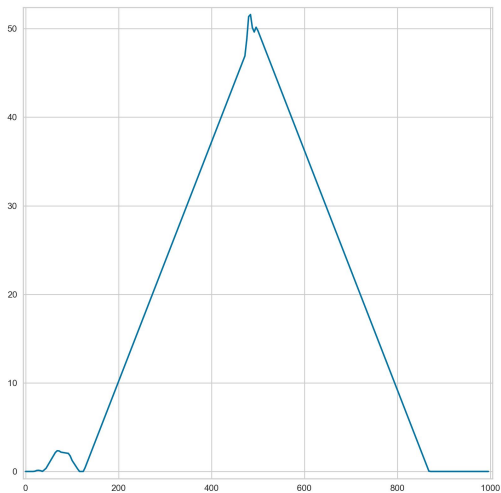
Silhouette de persistance

Figure – Paysage de persistance.

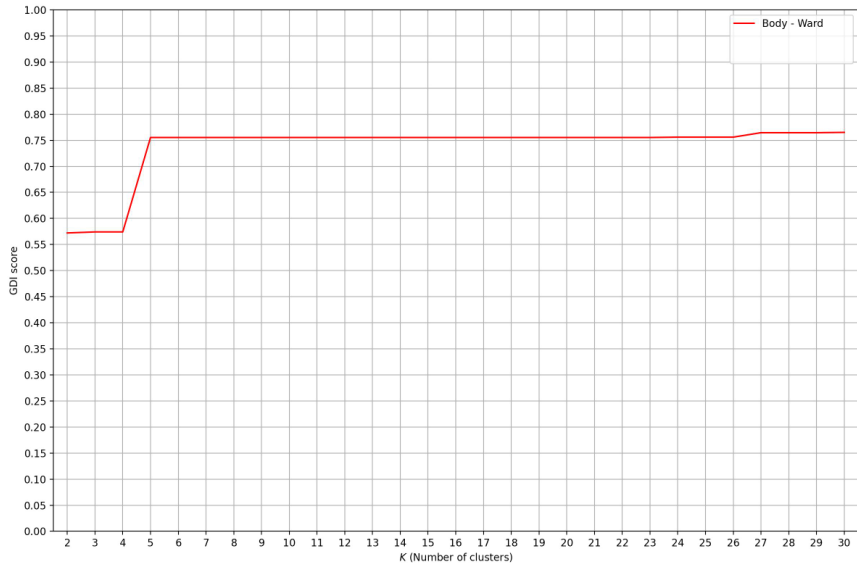


Silhouette de persistance

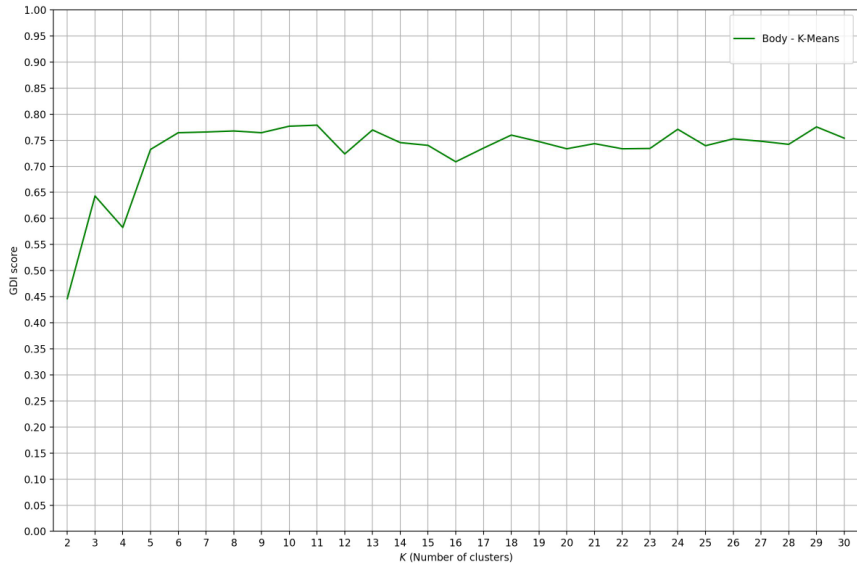
Figure – Silhouette de persistance.



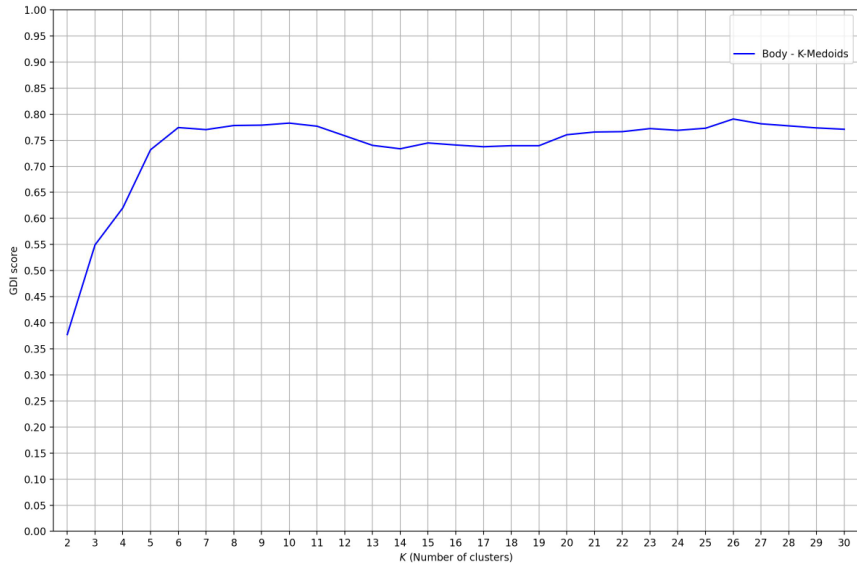
Vecteur de Silhouette



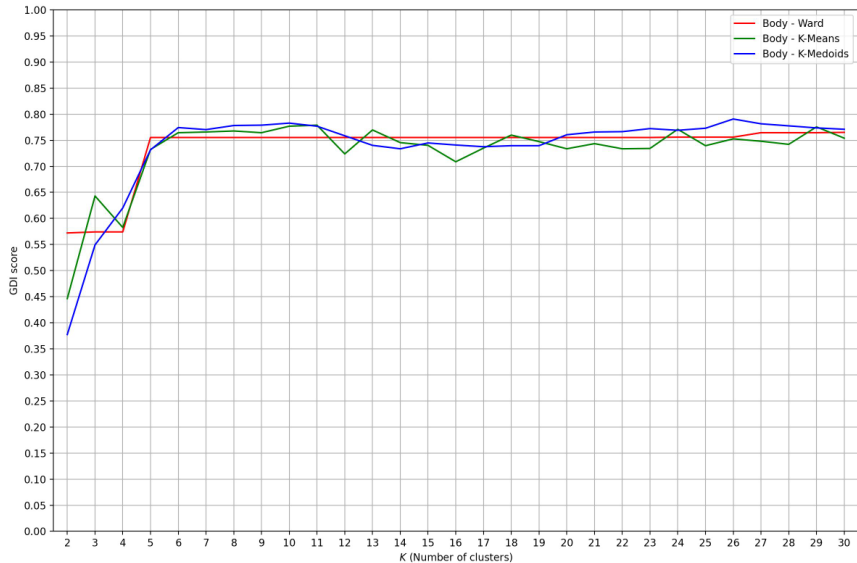
Vecteur de Silhouette



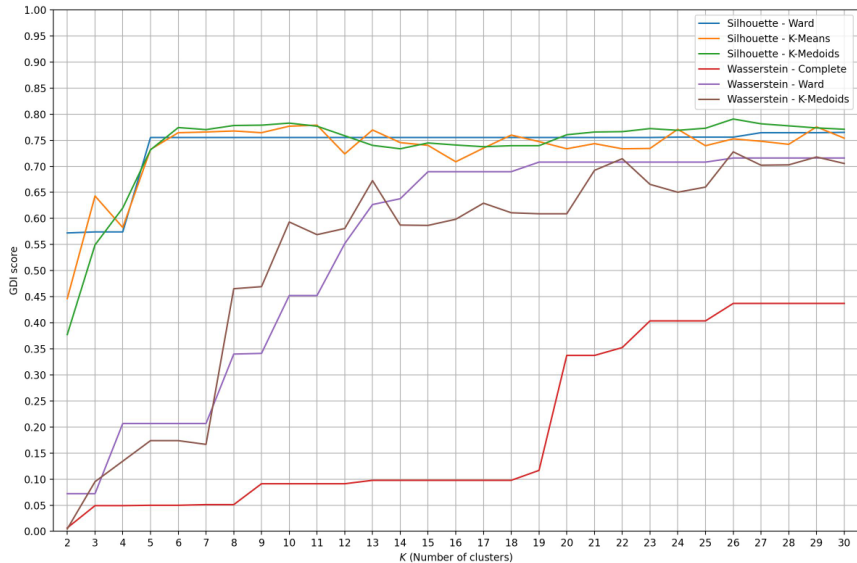
Vecteur de Silhouette



Vecteur de Silhouette



Vecteur de Silhouette



Points du tronc

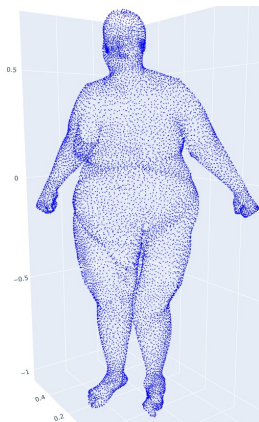


Figure – Full body

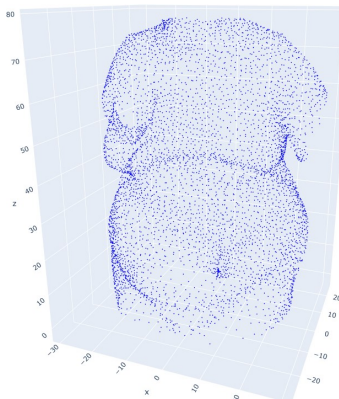
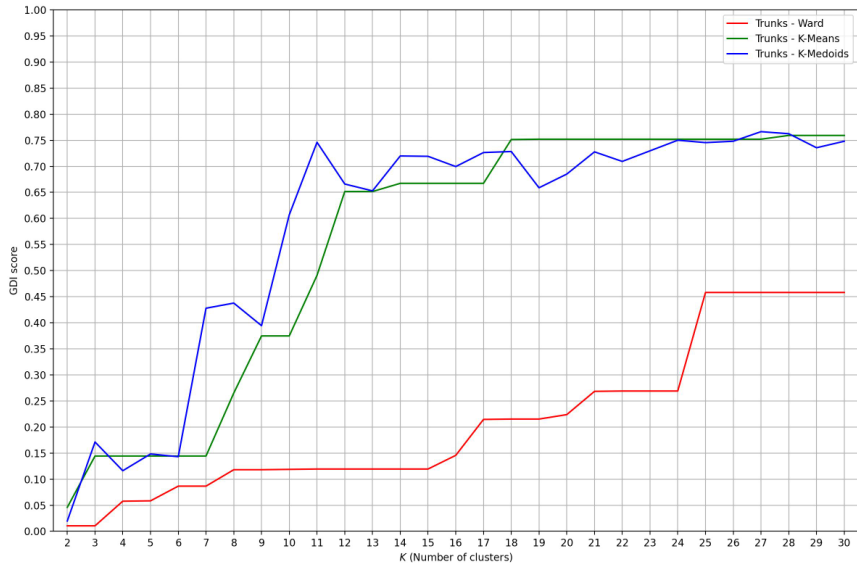
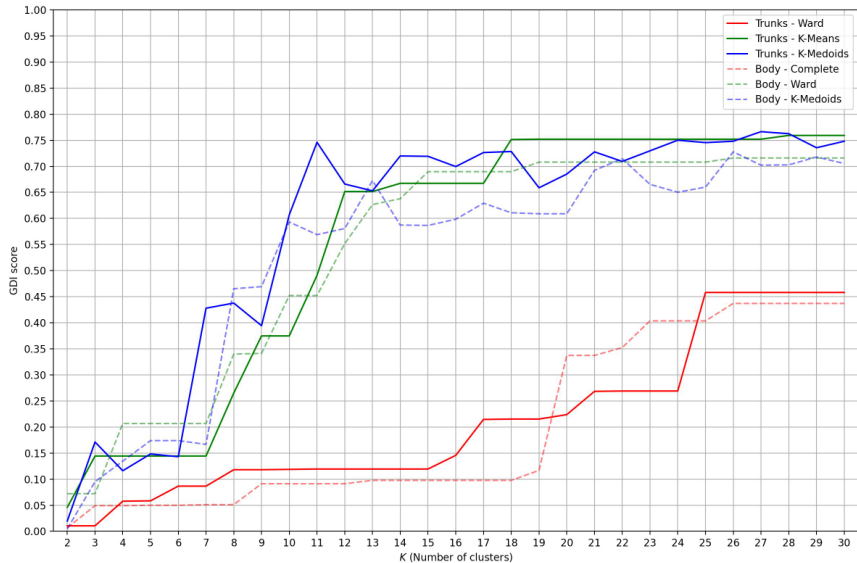


Figure – Trunk

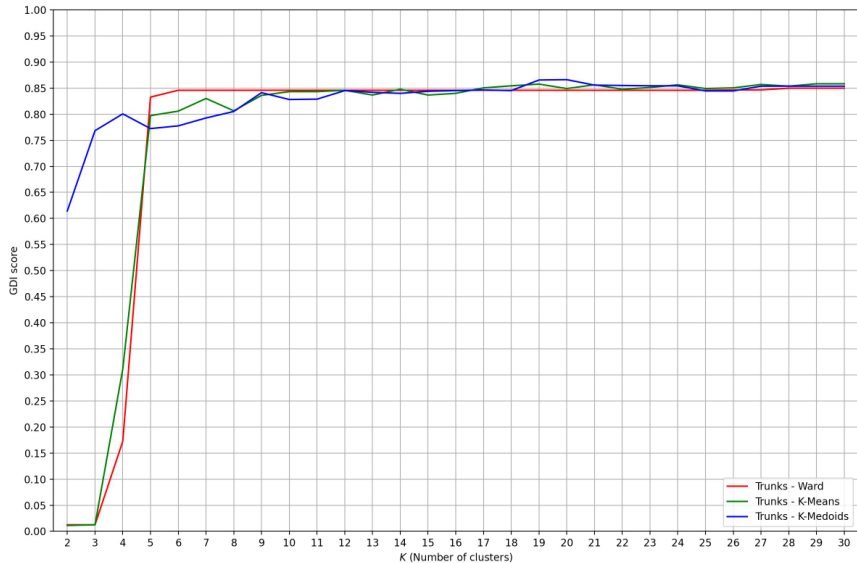
Distance de Wasserstein



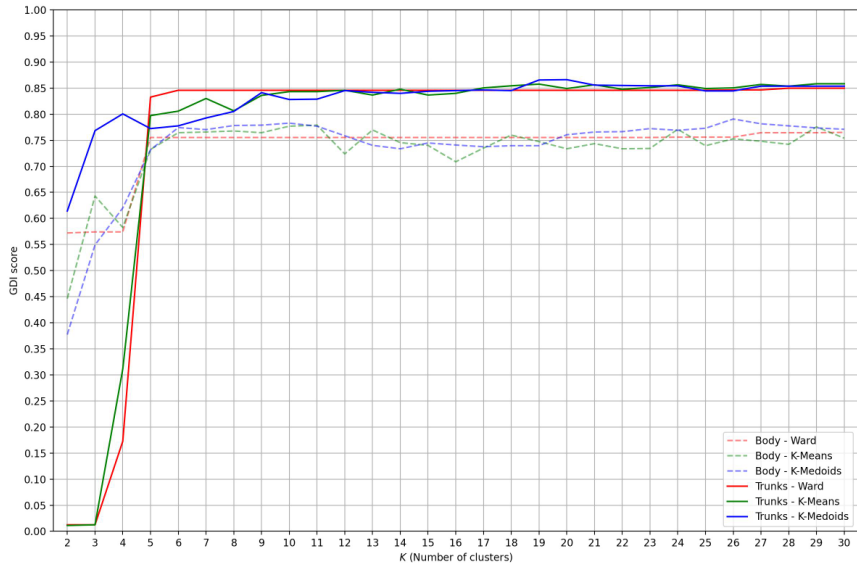
Distance de Wasserstein



Vecteur de Silhouette



Vecteur de Silhouette



Scores GDI moyens

Distance de Wasserstein

	Complete	Ward	K-Medoids
Body	0.2	0.54	0.526
Trunk	0.21	0.553	0.582

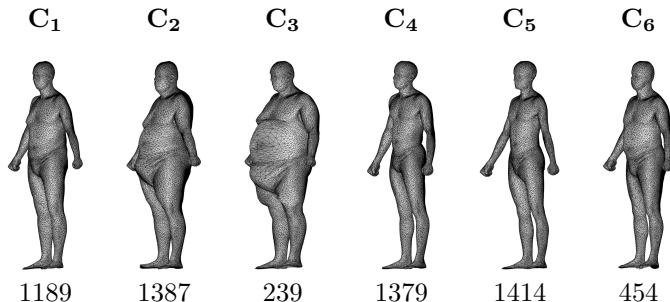
Vecteurs de Silhouette

	Ward	K-Means	K-Medoids
Body	0.738	0.73	0.737
Trunk	0.765	0.767	0.827

Male dataset

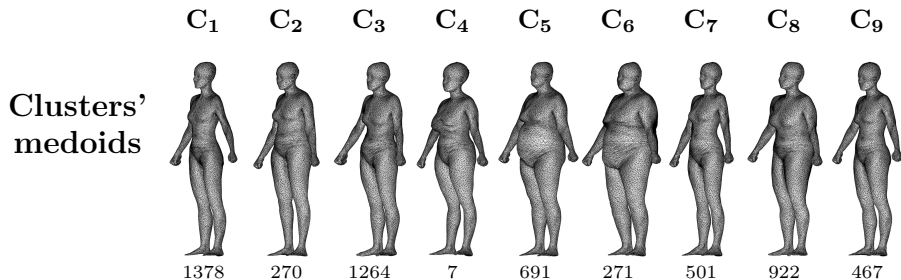
Cluster n°	1	2	3	4	5	6
Size	22	9	3	618	441	424
Proportion	1.45%	0.59%	0.2%	40.74%	29.07%	27.95%

Clusters'
medoids



Female dataset

Cluster n°	1	2	3	4	5	6	7	8	9
Size	253	148	196	154	115	2	287	144	232
Proportion	16.53%	9.67%	12.8%	10.06%	7.51%	0.13%	18.75%	9.41%	15.15%



Conclusions

TDA

- **Avec la distance de Wasserstein** on obtient une classification principalement par volume et on peut détecter les anomalies les plus importantes.
- **Avec la vectorisation des silhouettes** on peut, grâce à la pondération, mettre l'accent sur certaines homologues et classer les individus sur d'autres caractéristiques, au risque de trop se spécialiser.
- **La meilleure méthode** semble être un clustering hiérarchique avec liaison Ward des vecteurs obtenus à partir des silhouettes de persistances.

Perspectives

- Faire la moyenne de deux scans :
 - ▶ Comparaison des diagrammes de persistance.
 - ▶ Distances de Wasserstein.
- Tester d'autres méthodes de classification, sans passer par le TDA.
- Utiliser les données de l'IFTH.

Outils utilisés

-  GitHub
-  LaTeX
-  Python 3.10
 - ▶  GUDHI - Geometry Understanding in Higher Dimension
 - ▶  Scipy
 - ▶  machine learning in Python

Bibliographie

-  F. Chazal and B. Michel, “An Introduction to Topological Data Analysis : Fundamental and Practical Aspects for Data Scientists,” *Frontiers in Artificial Intelligence*, vol. 4, p. 667963, Sept. 2021.
-  Turner, Mileyko, and Mukherjee, “Fréchet means for distributions of persistence diagrams,” *Discrete Comput Geom*, vol. 52, p. 44–70, 2014.
-  M. Pirashvili, L. Steinberg, F. Belchi Guillamon, M. Niranjana, J. G. Frey, and J. Brodzki, “Improved understanding of aqueous solubility modeling through topological data analysis,” *Journal of Cheminformatics*, vol. 10, p. 54, Dec. 2018.