

---

# LE DATASCAPE

## FRAMEWORK NUMÉRIQUE POUR L'ABSTRACTION ET L'EXPLORATION DE DONNÉES MULTIDIMENSIONNELLES

**Directeur de Thèse:** Christian Attiogbe  
**Co-encadrant:** Damien Eveillard  
**Co-encadrant:** Benoit Delahaye  
**Co-encadrant industriel:** Ronan Boutin

**Jury:**  
Alejandro Maass  
Xavier Lorca  
Claire Cury



BIO LOGBOOK<sup>1</sup>

Precision Healthcare Technology

# BIO LOGBOOK



BIO LOGBOOK

Precision Healthcare Technology

## **Objectif:**

Améliorer la santé de la population

## **Comment:**

- Développement d'outils d'aide à la décision à destination des médecins et biologistes
- Définir la bonne santé. Détecter lorsqu'un patient s'en éloigne: *dérive biologique*

## **Verrous:**

Définir un système complexe associé à de nombreux paramètres



# PROBLÉMATIQUE

# PROBLÉMATIQUE

---

Qu'est-ce qu'une donnée multidimensionnelle?

## **Dimension**

Caractéristique pouvant prendre différentes valeurs, décrivant un objet/évènement

# PROBLÉMATIQUE

---

Qu'est-ce qu'une donnée multidimensionnelle?

## **Dimension**

Caractéristique pouvant prendre différentes valeurs, décrivant un objet/évènement

## **Donnée multidimensionnelle**

Combinaison de valeurs des dimensions d'un objet/évènement/...

# PROBLÉMATIQUE

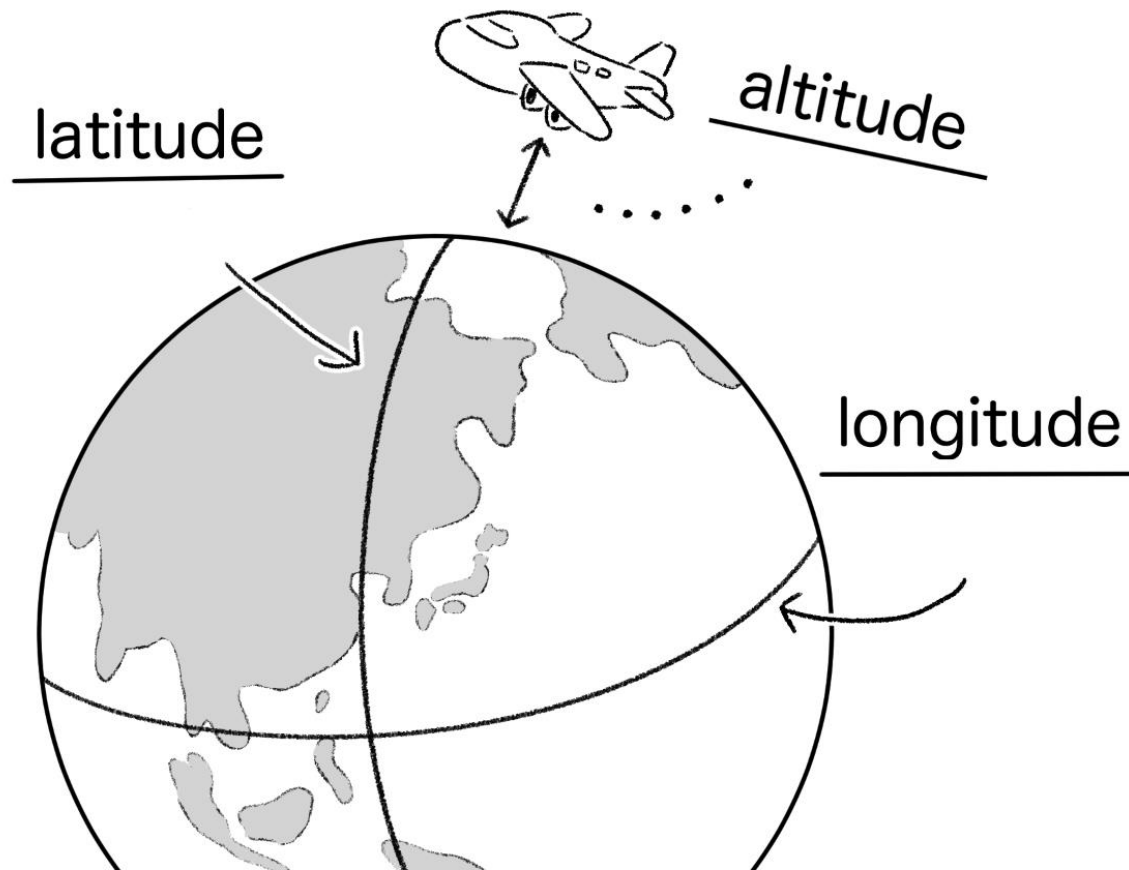
Qu'est-ce qu'une donnée multidimensionnelle?

## Dimension

Caractéristique pouvant prendre différentes valeurs, décrivant un objet/évènement

## Donnée multidimensionnelle

Combinaison de valeurs des dimensions d'un objet/évènement/...



# PROBLÉMATIQUE

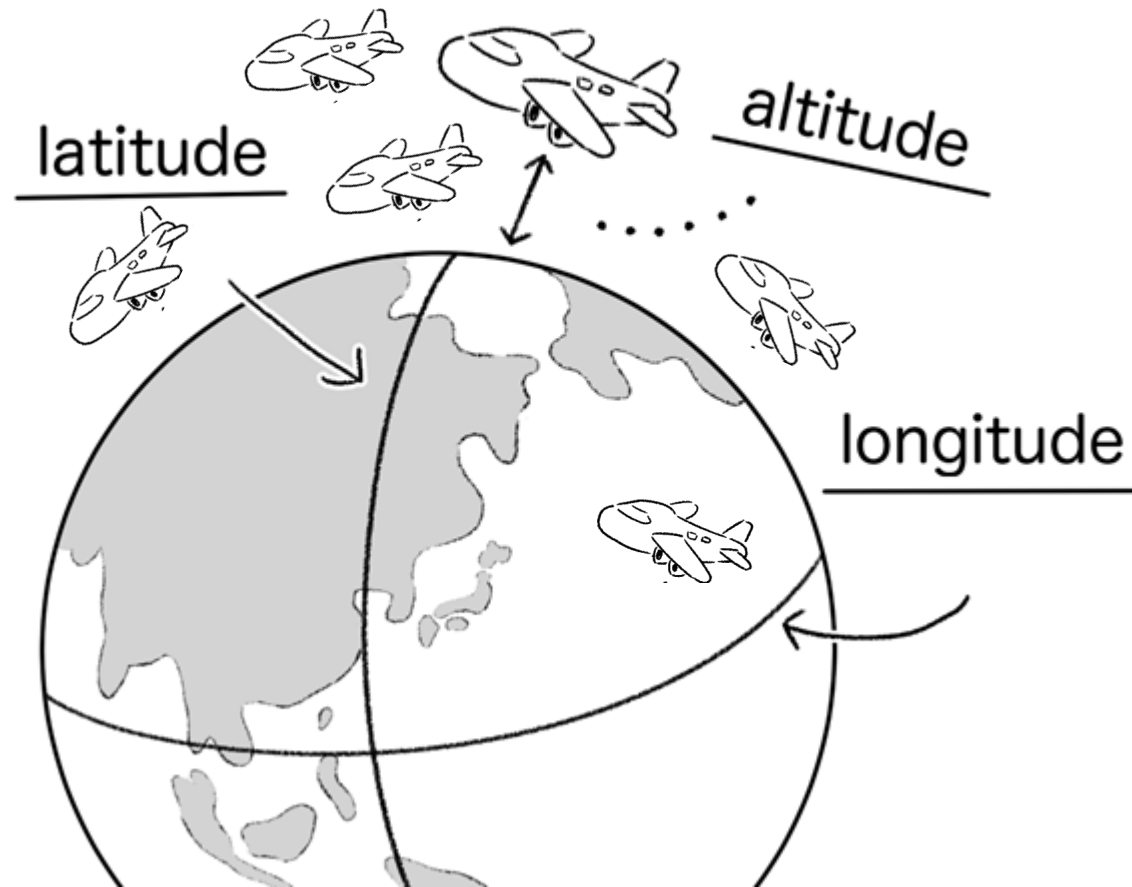
Qu'est-ce qu'une donnée multidimensionnelle?

## Dimension

Caractéristique pouvant prendre différentes valeurs, décrivant un objet/évènement

## Donnée multidimensionnelle

Combinaison de valeurs des dimensions d'un objet/évènement/...



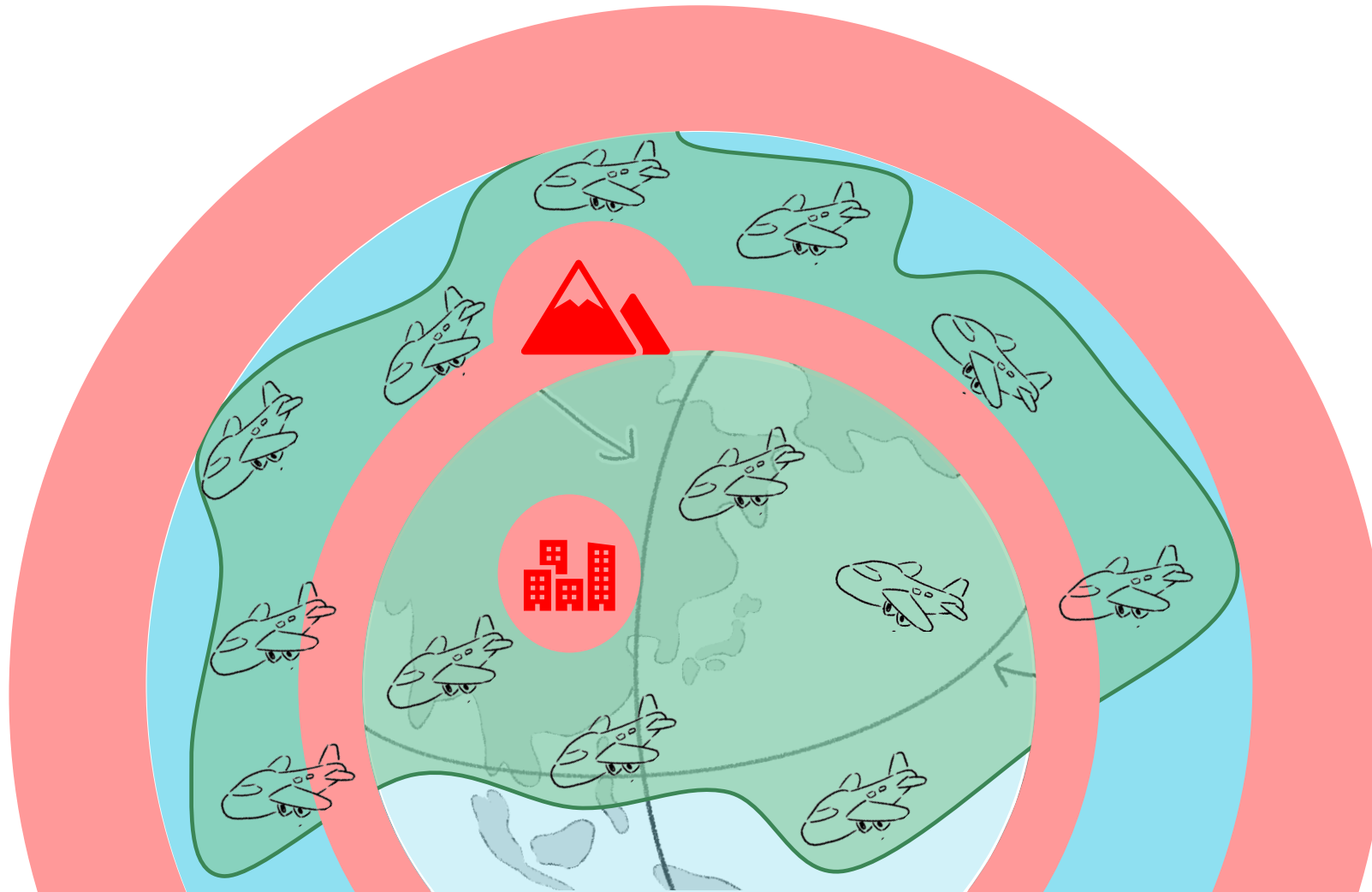
Avion n°	Altitude	Longitude	Latitude
1	3600	-1.554	47.2222
2	2650	-1.785	45.897
3	1263	19.568	13.165
4	5263	15.589	35.568
...	...	...	...

**Plusieurs données multidimensionnelles = jeu de donnée = dataset**

# PROBLÉMATIQUE

Qu'est-ce qu'une donnée multidimensionnelle?

**Qu'est-ce qu'une position possible d'un avion en vol ?**





# PROBLÉMATIQUE

Prélevement réalisé au Laboratoire le 05-11-2018 à 09:08

Compte-rendu édité le 10-11-2018 à 09:57

Compte-rendu complet



Laboratoire accrédité n°8-3155

Listes des sites accrédités et portée disponible sur [www.cofrac.fr](http://www.cofrac.fr)

Seules certaines prestations rapportées dans ce document sont couvertes par l'accréditation. Elles sont identifiées par le symbole [AC].

Résultats Valeur

REMARQUES

Pièce d'identité présentée:

Identité du patient vérifiée au Laboratoire

HEMATOLOGIE SANGUINE

## HEMOGRAMME:

(Cytométrie en flux - SYSMEX XN2000)

Leucocytes  
Hématies  
Hémoglobine  
Hématocrite  
Vol. globulaire VGM  
Concentration CCMH  
Teneur TCMH

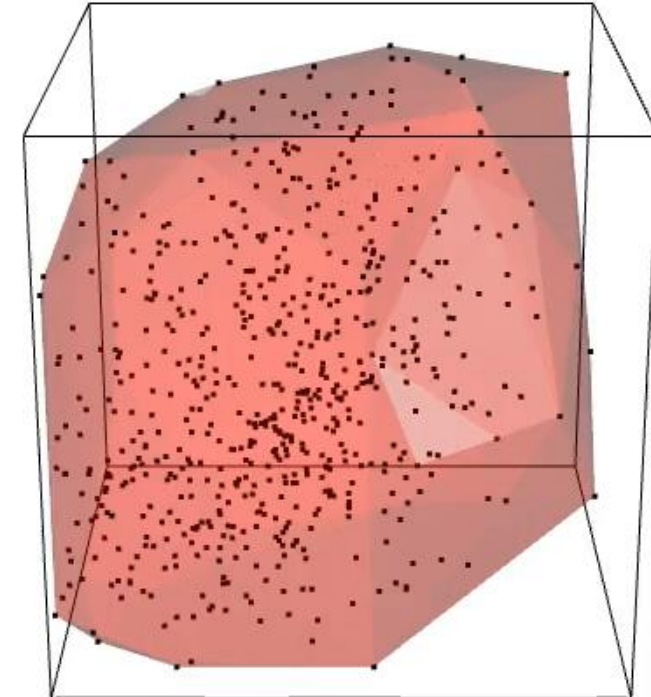
8.85	G/L	(4.0)
4.78	T/L	(4.0)
11.6	g/dL	(12)
36.4	%	(37)
76.2	fL	(80)
31.9	g/dL	(32)
24.3	pg	(27)

Hypochromie

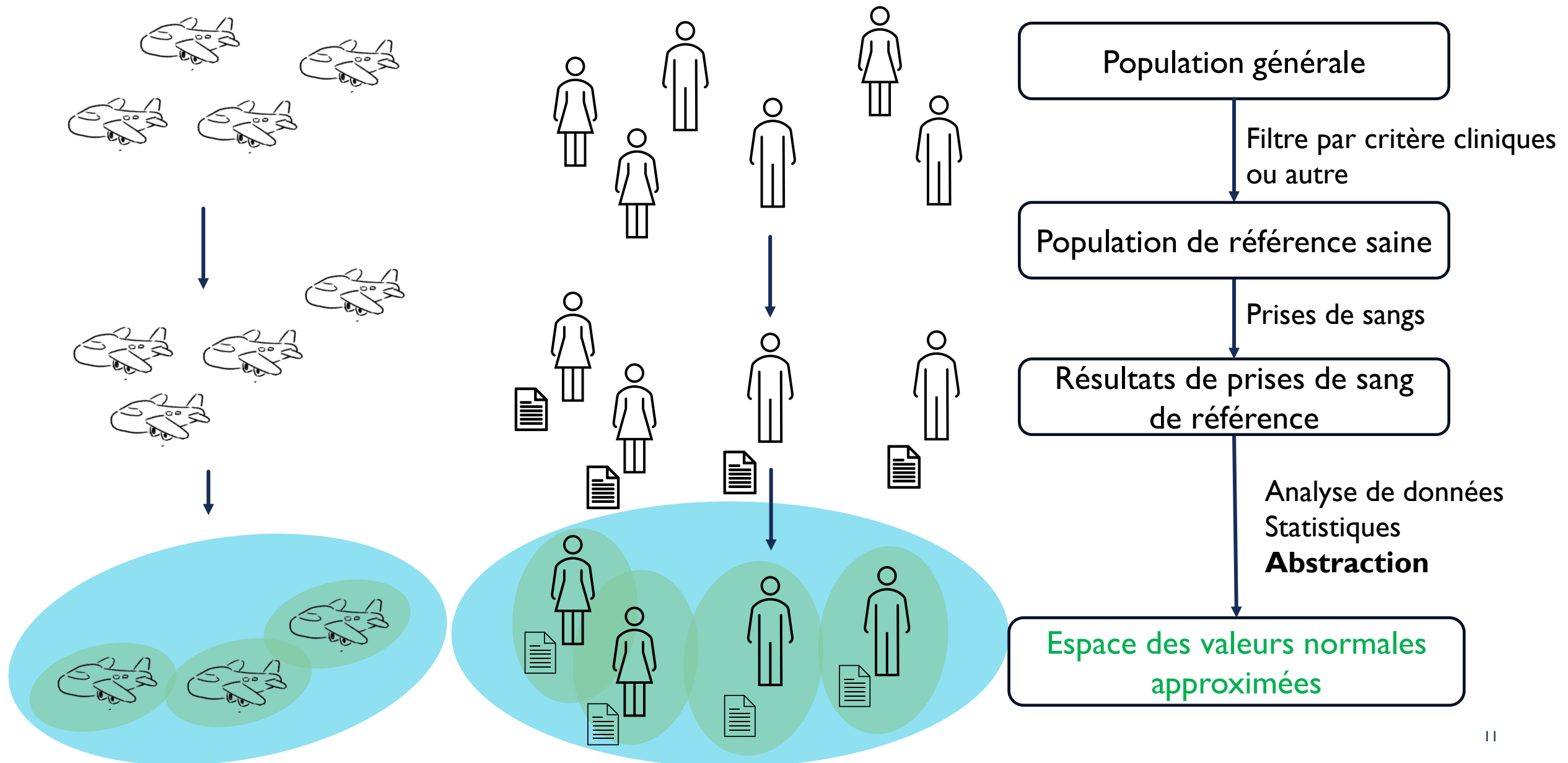
HEMOGLOBINE

LEUCOCYTES

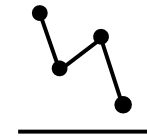
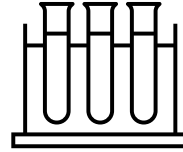
ERYTHROCYTES



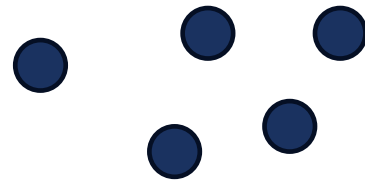
# PROBLÉMATIQUE



# PROBLÉMATIQUE

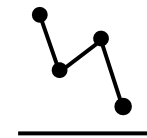
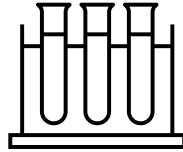


Aviation	Biologie médicale	Science des données
Position d'un avion	Résultats d'une prise de sang	Un point / une donnée

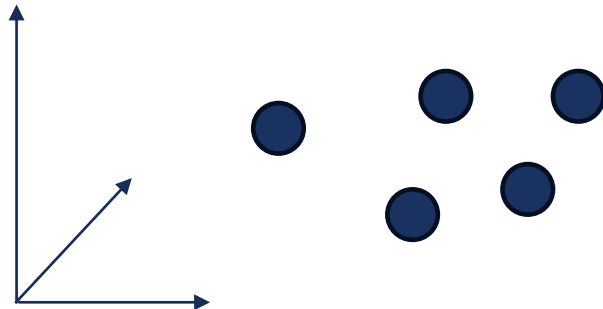


# PROBLÉMATIQUE

## VOCABULAIRE

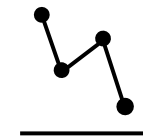
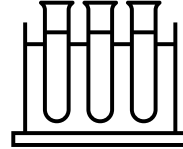


Aviation	Biologie médicale	Science des données
Position d'un avion	Résultats d'une prise de sang	Un point / une donnée
<b>Latitude / Longitude / Altitudes (3)</b>	<b>Hémoglobine / albumines ... (7)</b>	<b>Dimensions</b>

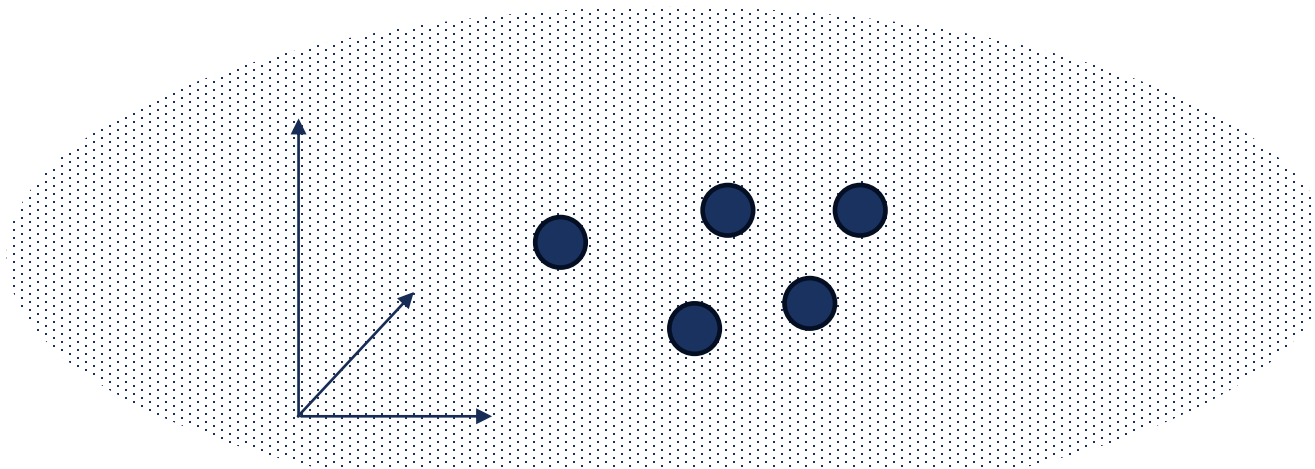


# PROBLÉMATIQUE

## VOCABULAIRE

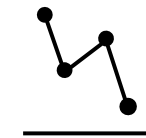
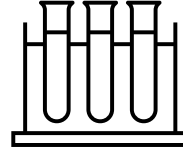


Aviation	Biologie médicale	Science des données
Position d'un avion	Résultats d'une prise de sang	Un point / une donnée
Latitude / Longitude / Altitudes (3)	Hémoglobine / albumines ... (7)	Dimensions
<b>Toutes positions imaginables (univers)</b>	<b>Tous les résultats imaginables</b>	<b>Espace ambient</b>

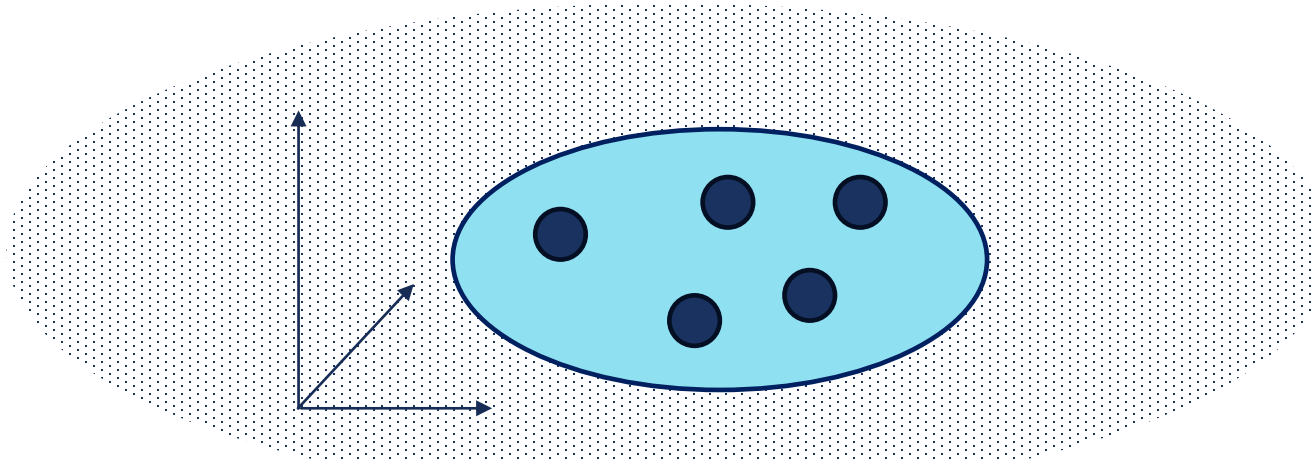


# PROBLÉMATIQUE

## VOCABULAIRE

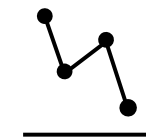
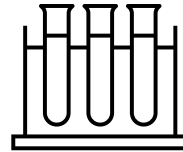


Aviation	Biologie médicale	Science des données
Position d'un avion	Résultats d'une prise de sang	Un point / une donnée
Latitude / Longitude / Altitudes (3)	Hémoglobine / albumines ... (7)	Dimensions
Toutes positions imaginables (univers)	Tous les résultats imaginables	Espace ambient
<b>Espace aérien</b>	<b>Espace des valeurs normales</b>	<b>Espace réel /sous-jacent</b>

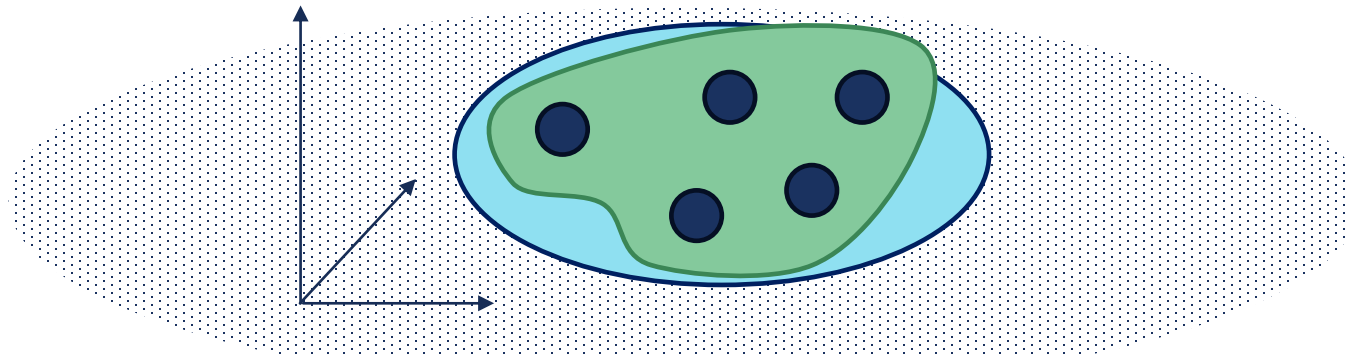


# PROBLÉMATIQUE

## VOCABULAIRE

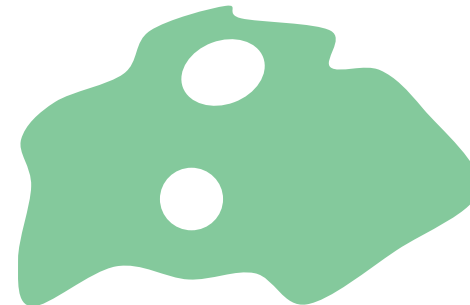
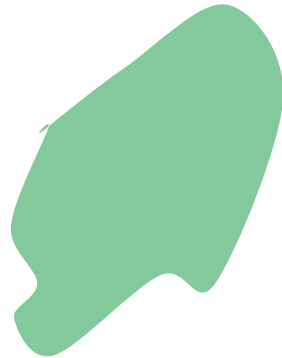


	Aviation	Biologie médicale	Science des données
●	Position d'un avion	Résultats d'une prise de sang	Un point / une donnée
↕	Latitude / Longitude / Altitudes (3)	Hémoglobine / albumines ... (7)	Dimensions
⋯	Toutes positions imaginables (univers)	Tous les résultats imaginables	Espace ambient
■	Espace aérien théorique	Espace des valeurs normales réel	Espace réel /sous-jacent
■	<b>Espace aérien observés</b>	<b>Espace des valeurs normales approximé</b>	<b>Espace approximé</b>



## Comment approximer l'espace sous-jacent d'un jeu de donnée multidimensionnel ?

- I – Comment définir et formaliser **la forme** de l'espace approximé ?
- II – Comment définir la notion de **contour/bords**?
- III – Comment calculer des **distances** dans cet espace ?





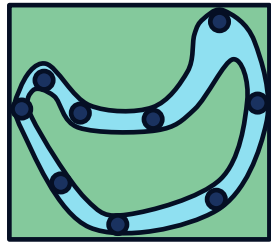


# MÉTHODES EXISTANTES

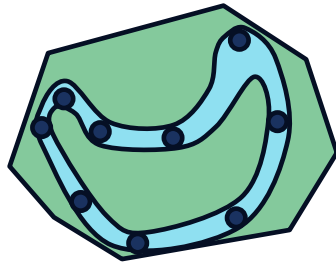
# MÉTHODES EXISTANTES: FORME / CONTOURS / DISTANCES

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?

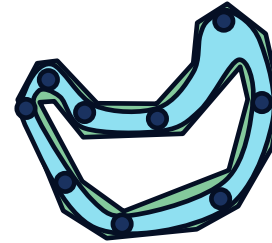
**Approches géométriques :** Trouver une forme contenant les données



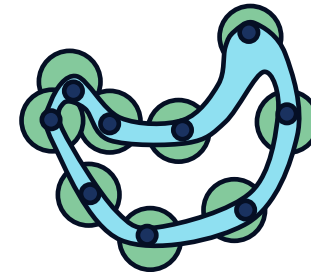
hypercube



convex hull



alpha-shape



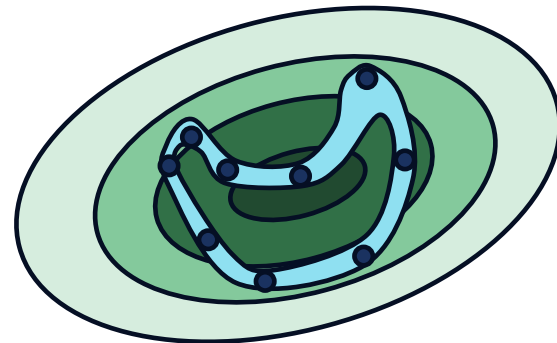
boules

- Solberg HE. Multivariate reference regions. *Scand J Clin Lab Invest Suppl.* 1995;222:3-5. doi: 10.3109/00365519509088443. PMID: 7569743.
- Boyd JC, Lacher DA. The multivariate reference range: an alternative interpretation of multi-test profiles. *Clin Chem.* 1982 Feb;28(2):259-65. PMID: 7055945.
- H. Edelsbrunner, D. Kirkpatrick and R. Seidel, "On the shape of a set of points in the plane," in *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 551-559, July 1983, doi: 10.1109/TIT.1983.1056714.
- Dimitri Konen, Davy Paindaveine "Multivariate  $\rho$ -quantiles: A spatial approach," *Bernoulli*, *Bernoulli* 28(3), 1912-1934, (August 2022)

# MÉTHODES EXISTANTES : FORME / CONTOURS / DISTANCES

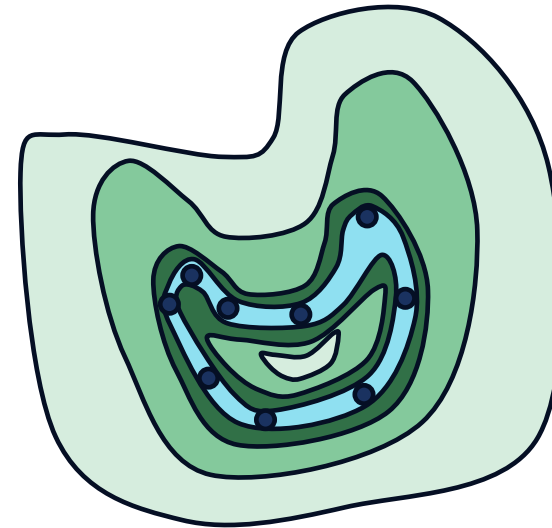
Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?

**Approches Statistiques/Probabilistes:** Décrire la densité d'un jeu de donnée



Ellipse

Lois normales multivariées



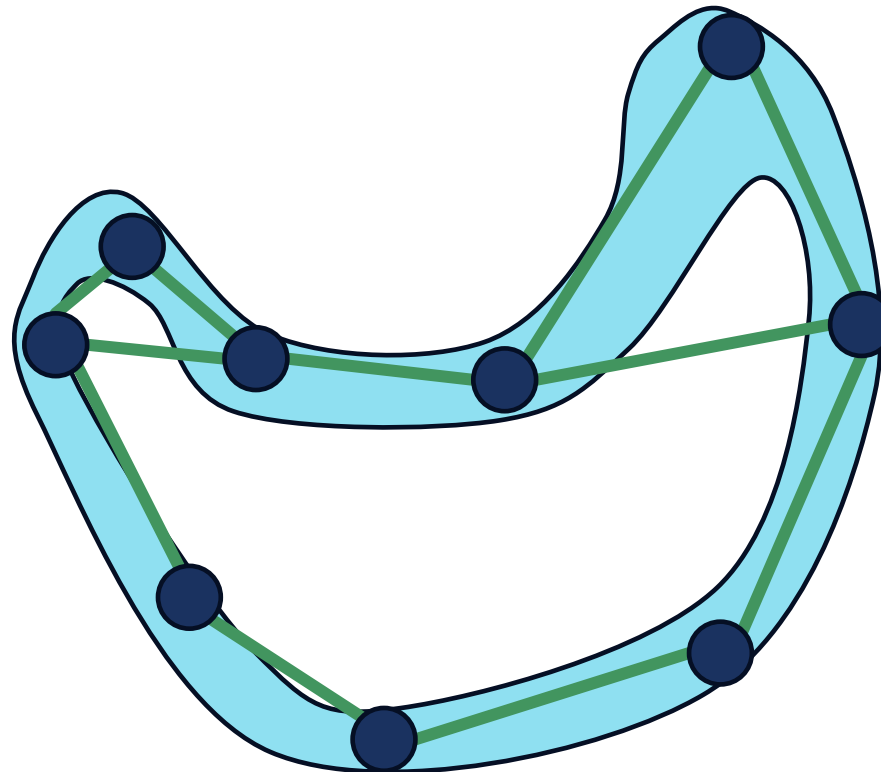
$p$ -quantiles / Density estimators

- Solberg HE. Multivariate reference regions. *Scand J Clin Lab Invest Suppl.* 1995;222:3-5. doi: 10.3109/00365519509088443. PMID: 7569743.
- Boyd JC, Lacher DA. The multivariate reference range: an alternative interpretation of multi-test profiles. *Clin Chem.* 1982 Feb;28(2):259-65. PMID: 7055945.
- H. Edelsbrunner, D. Kirkpatrick and R. Seidel, "On the shape of a set of points in the plane," in *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 551-559, July 1983, doi: 10.1109/TIT.1983.1056714.
- Dimitri Konen, Davy Paindaveine "Multivariate  $p$ -quantiles: A spatial approach," *Bernoulli*, *Bernoulli* 28(3), 1912-1934, (August 2022)

# MÉTHODES EXISTANTES : FORME / CONTOURS / DISTANCES

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?

**Approches de manifold learning:** Identifier et décrire une surface sur laquelle les données sont « posées »



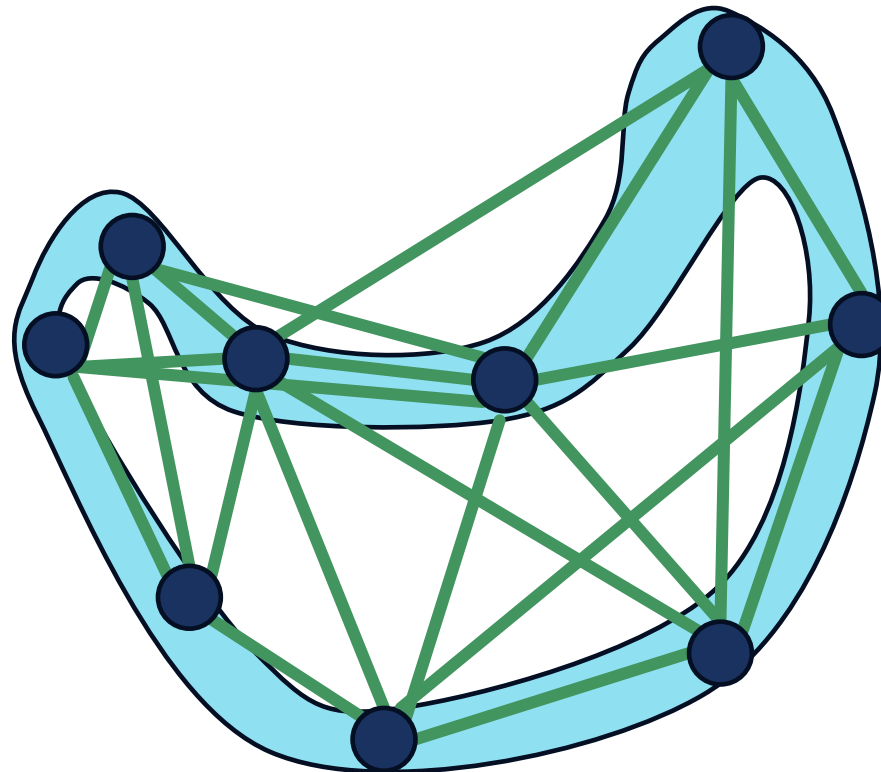
Graphe des plus proches voisins  
Nombre de voisins ( $k$ ) = 2

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

# MÉTHODES EXISTANTES : FORME / CONTOURS / DISTANCES

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?

**Approches de manifold learning:** Identifier et décrire une surface sur laquelle les données sont « posées »



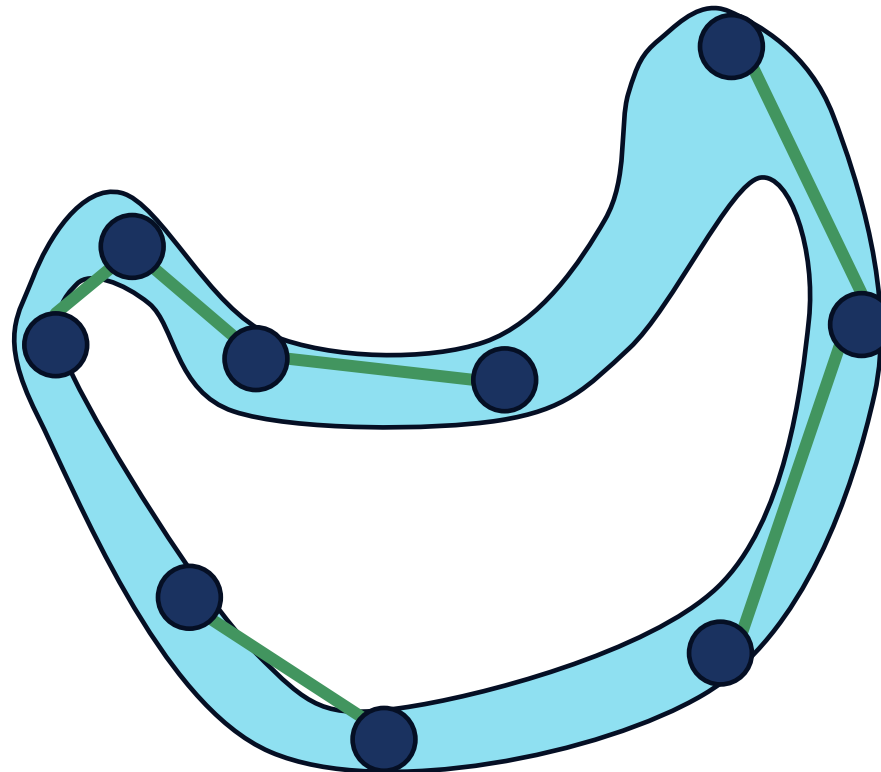
Graphe des plus proches voisins  
 $k = 4$

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

# MÉTHODES EXISTANTES : FORME / CONTOURS / DISTANCES

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?

**Approches de manifold learning:** Identifier et décrire une surface sur laquelle les données sont « posées »



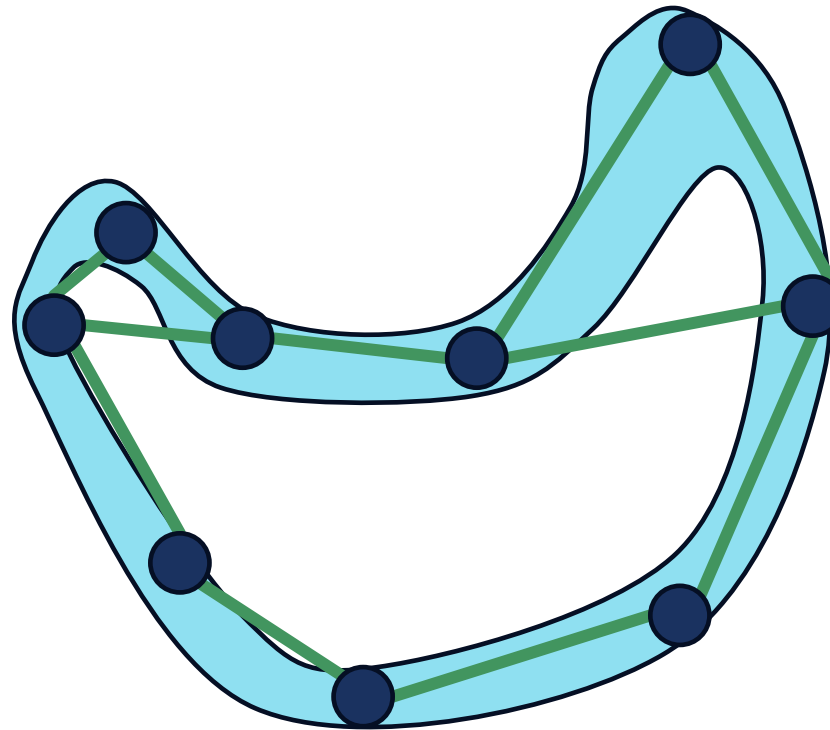
Graphe des plus proches voisins  
 $k = 1$

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

# MÉTHODES EXISTANTES : FORME / CONTOURS / DISTANCES

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?

**Approches de manifold learning:** Identifier et décrire une surface sur laquelle les données sont « posées »



Graphe des plus proches voisins  
 $k = 2$

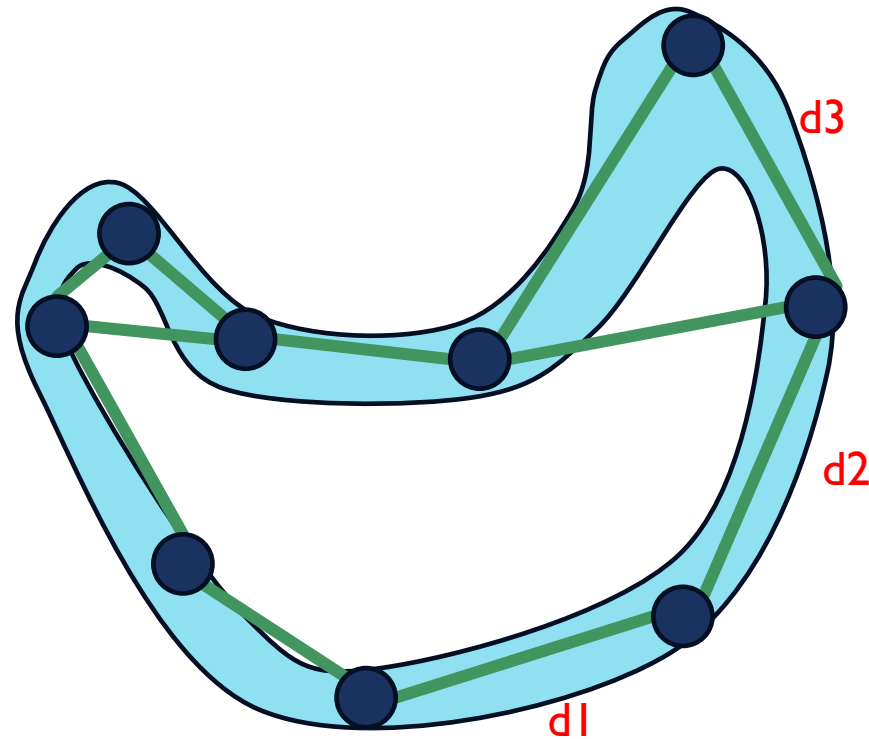
Capture la *topologie* des données (Umap 2018)

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

# MÉTHODES EXISTANTES : FORME / CONTOURS / DISTANCES

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?

**Approches de manifold learning:** Identifier et décrire une surface sur laquelle les données sont « posées »



Graphe des plus proches voisins  
 $k = 2$

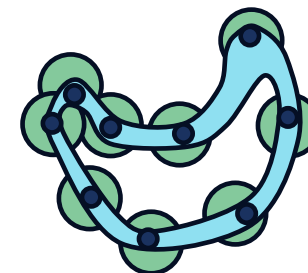
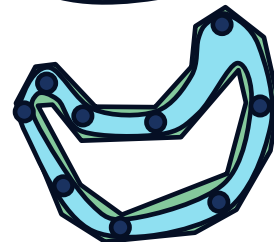
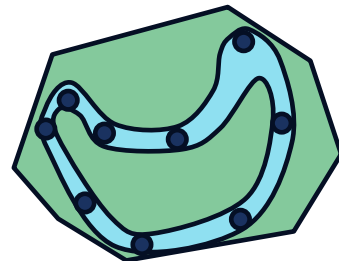
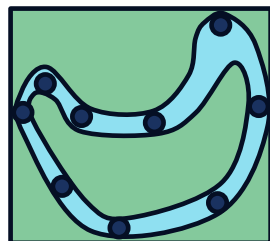
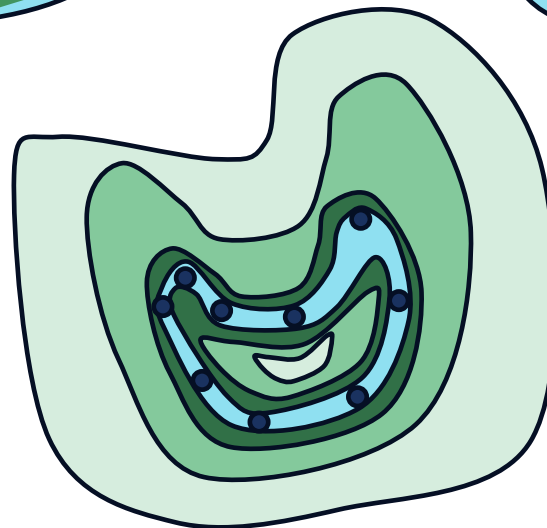
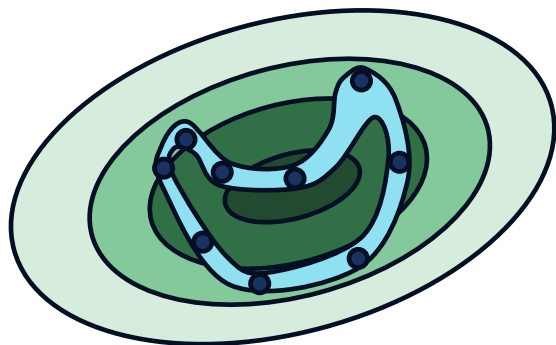
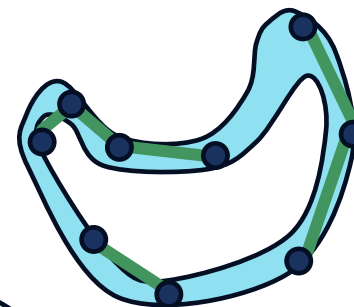
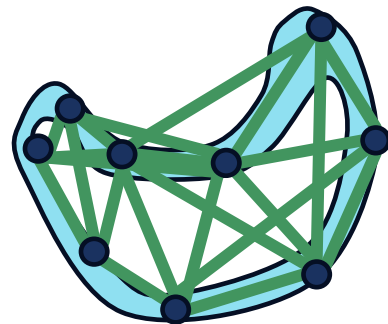
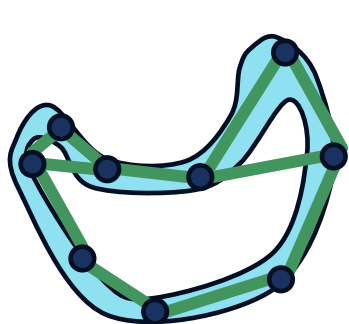
Capture la *topologie* des données (Umap 2018)

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.



# MÉTHODES EXISTANTES : FORME / CONTOURS / DISTANCES

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?



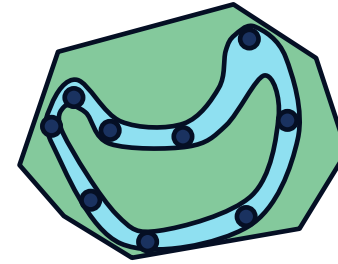
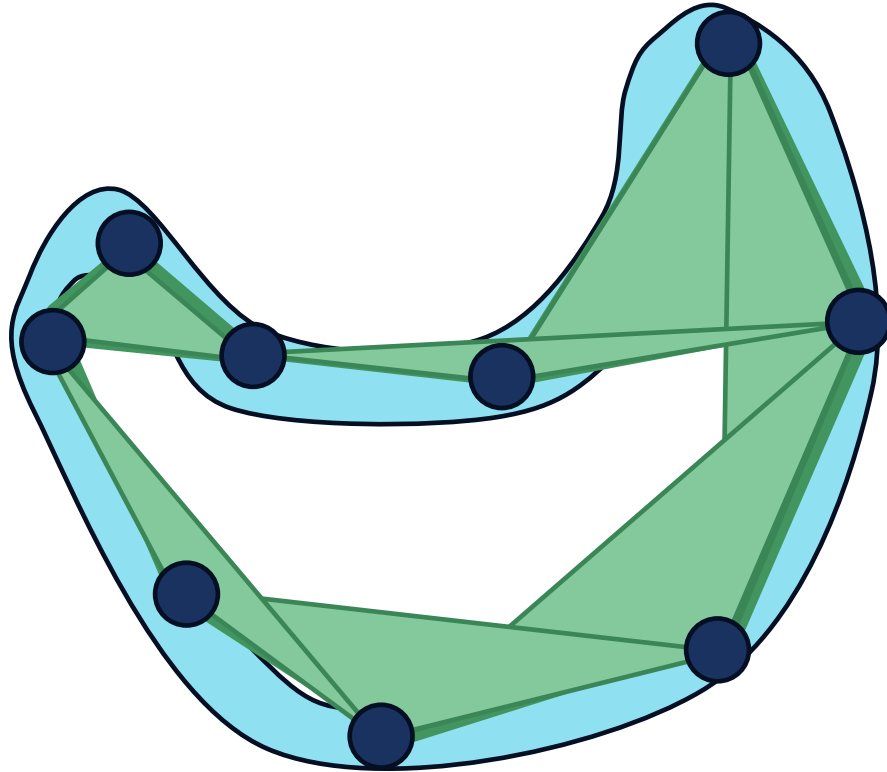
boules



# LE CONCEPT DE DATASCAPE

# LE CONCEPT DE DATASCAPE: FORMALISATION

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?



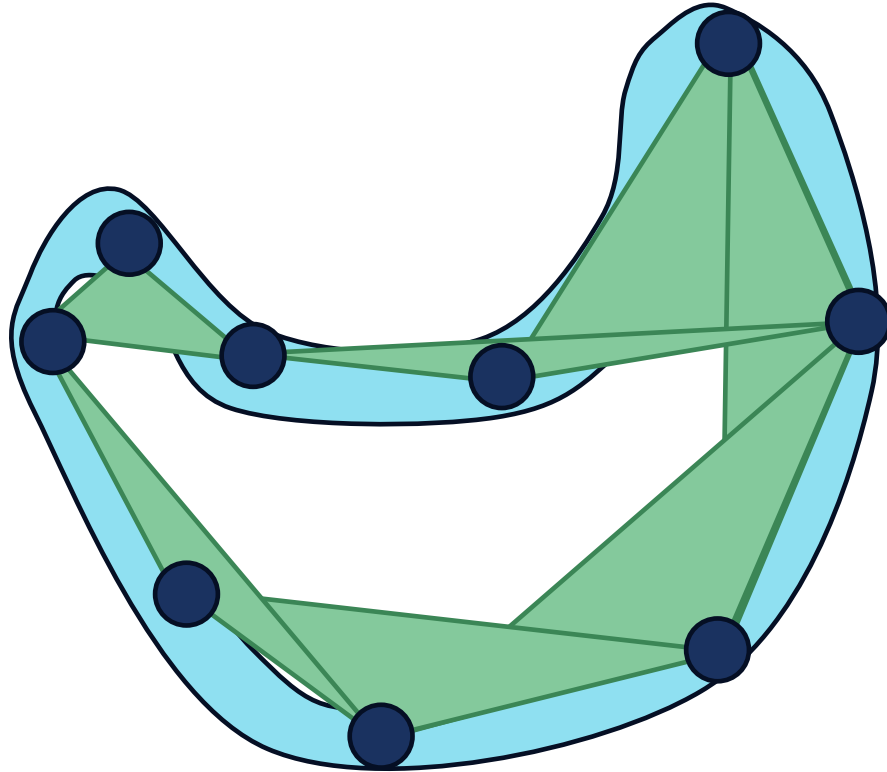
convex hull

## Construction d'enveloppes convexes sur chaque voisinage

- Briques élémentaires calculables
- S'adapte à la dimension intrinsèque des données
- Définit une notion d'intérieur, extérieur et contours
- Densité

# LE CONCEPT DE DATASCAPE: IMPLÉMENTATION

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?



I – Comment définir et formaliser **la forme** de l'espace sous-jacent?

→ graphe des plus proches voisins + enveloppes convexes

II – Comment définir la notion de **contour/bords**?

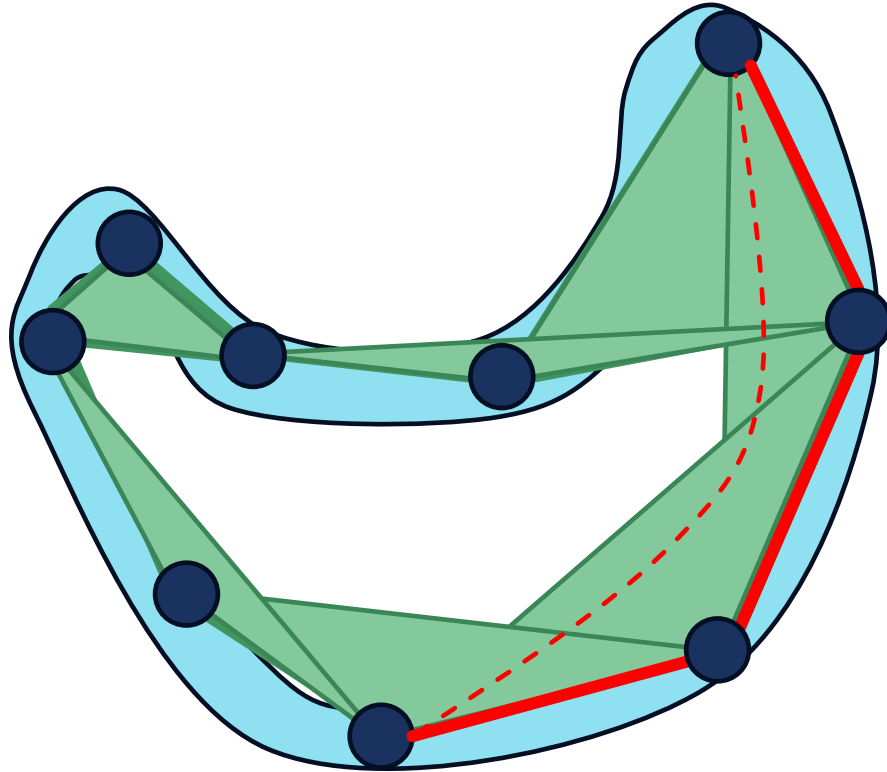
→ extrapoler la notion de point extrême d'une enveloppe convexe

III – Comment calculer des **distances** dans cet espace ?

→ approximation via le graphe

# LE CONCEPT DE DATASCAPE: IMPLÉMENTATION

Comment abstraire l'espace sous-jacent à un jeu de donnée multidimensionnel ?



## Construction du graphe

Choix du nombre de voisins  $k$

**force brute (TDA)**

**$k=10$**

Recherche des voisins

**force brute**

Choix distance

**euclidienne**

## Construction des enveloppes

**Algorithme Quickhull**

## Distance dans le Datascape

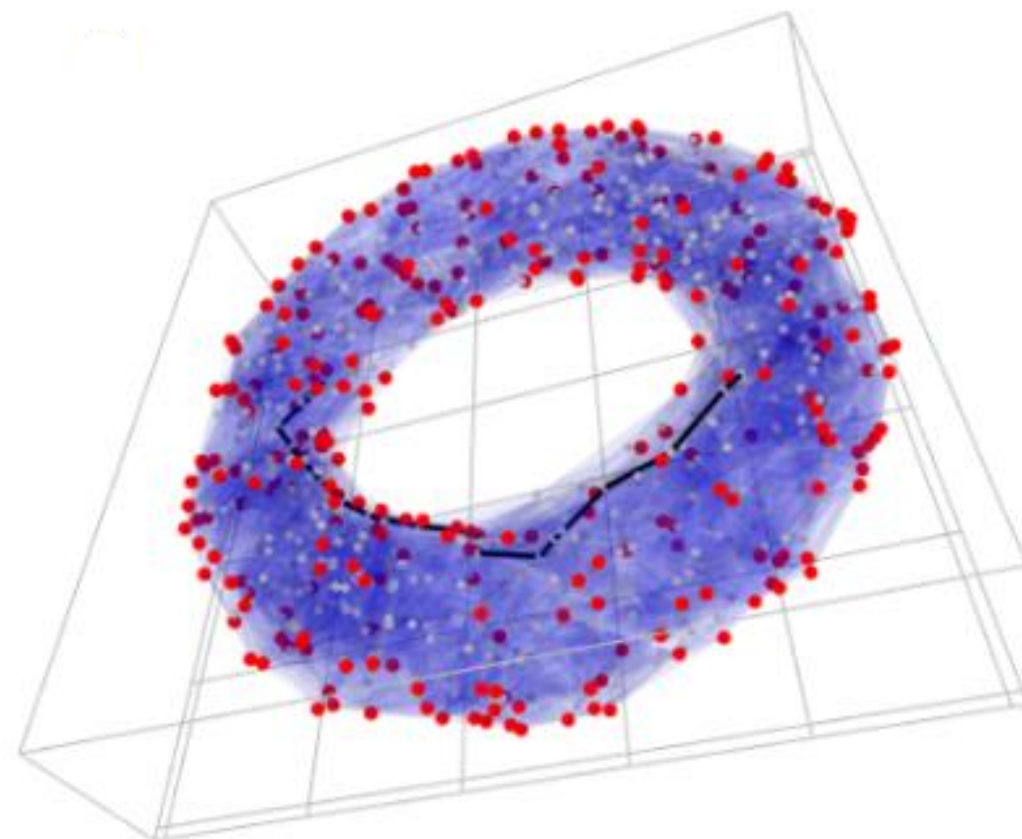
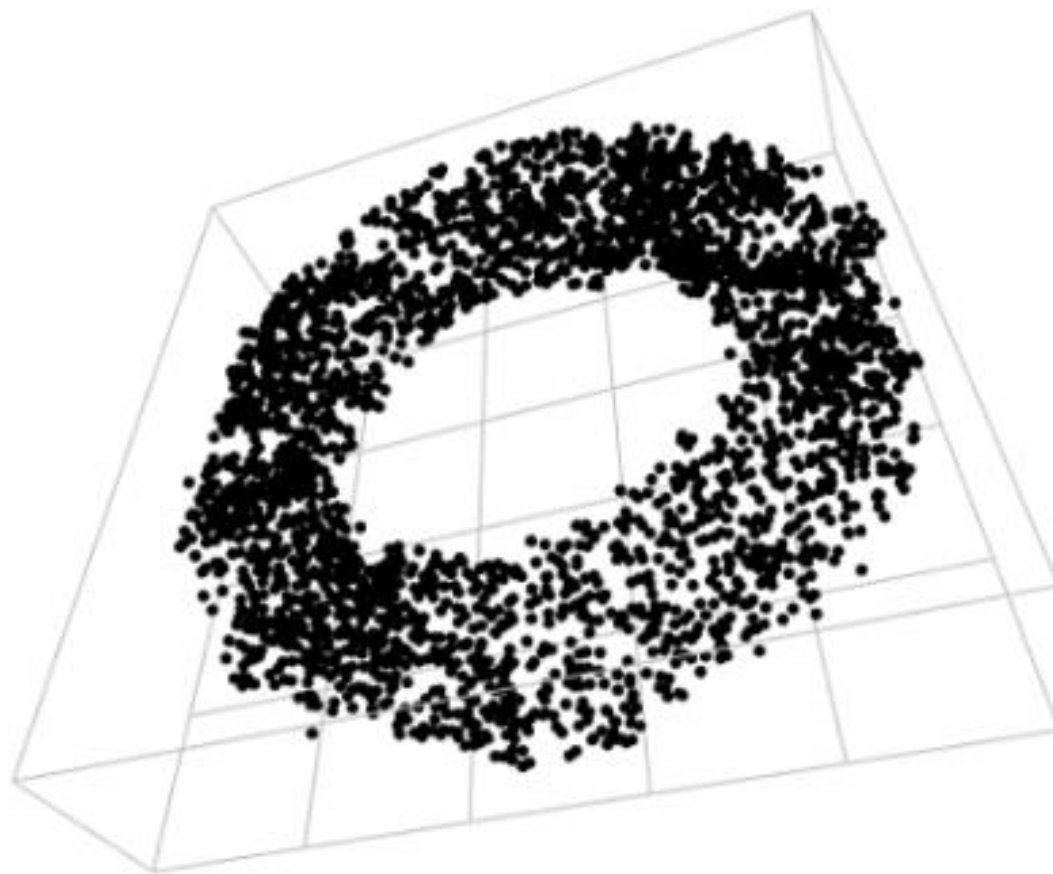
Chemin les plus court dans le graphe

**Algorithme Dijkstra**

- Bratić, B., Houle, M. E., Kurbalija, V., Oria, V., & Radovanović, M. (2018, June). NN-Descent on high-dimensional data. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics* (pp. 1-8).
- Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22(4), 469-483.
- Javaid, A. (2013). Understanding Dijkstra's algorithm. Available at SSRN 2340905.

# LE CONCEPT DE DATASCAPE: ILLUSTRATION

Applications sur des données simulées





# EXEMPLES DE DATASCAPES

**DONNÉES TARA**

PHYSIONET CARDIOLOGY CHALLENGE

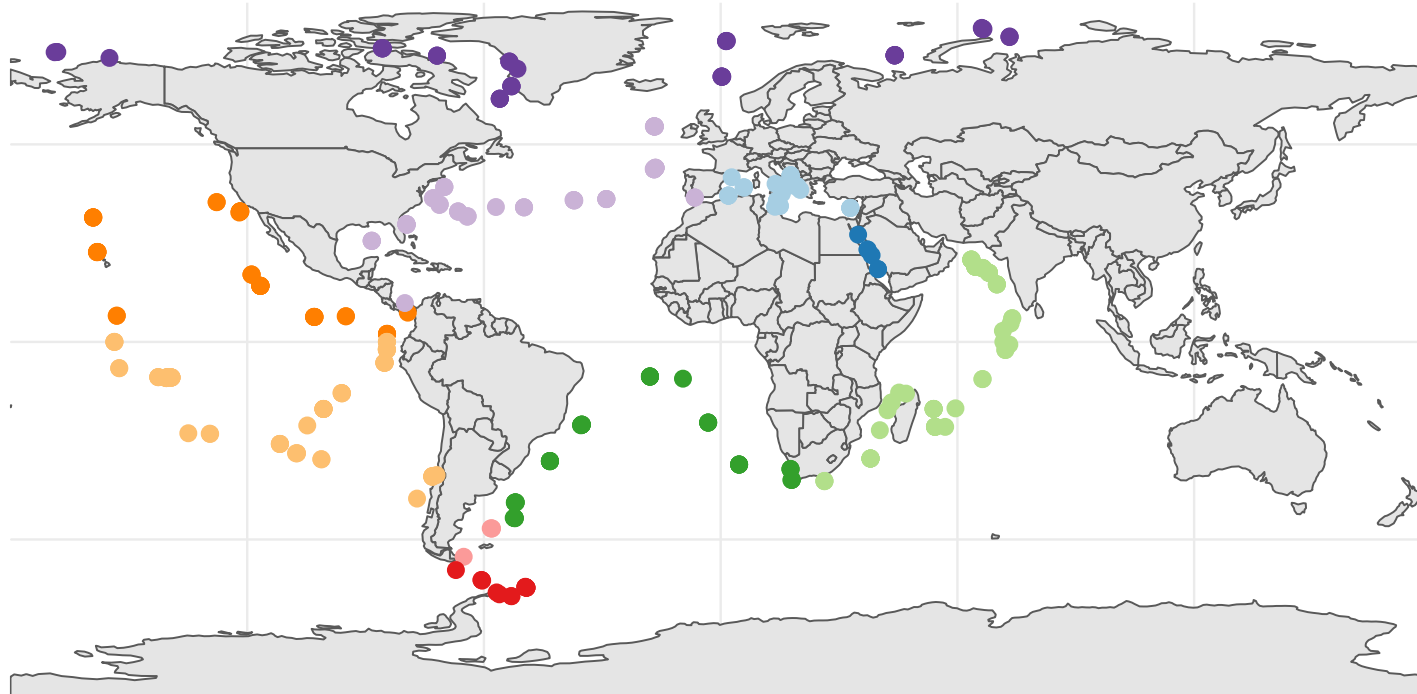
# APPLICATIONS: DONNÉES TARA

## Données

### Données:

- 900 ecological stations
- ~ 90 paramètres + omics
- Température, nitrate, concentrations fer, oxygène, salinité, chlorophyle A, lumière, le flux de carbone à 150m (8D)

**Objectif:** Reclassifier les stations océaniques selon des critères physico chimiques

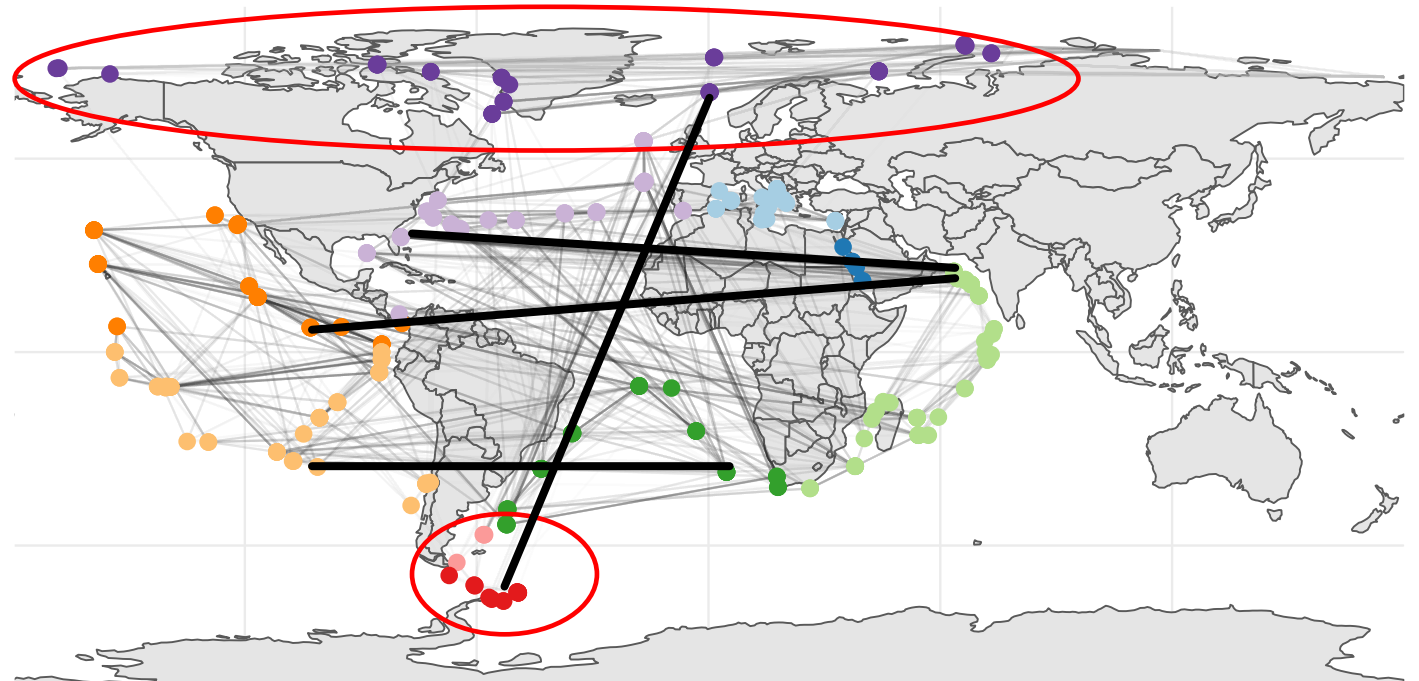
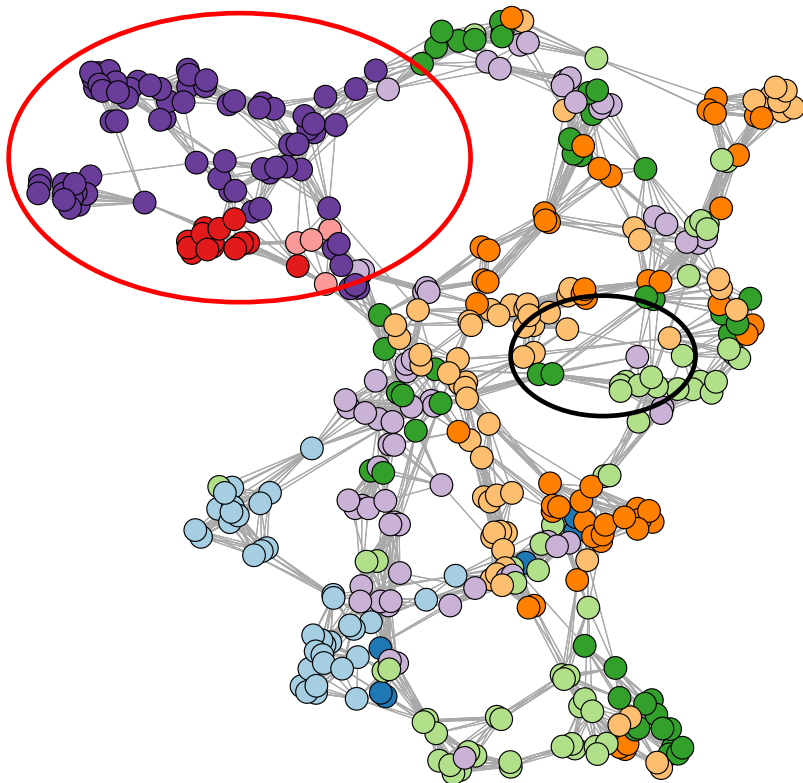




# APPLICATIONS: DONNÉES TARA

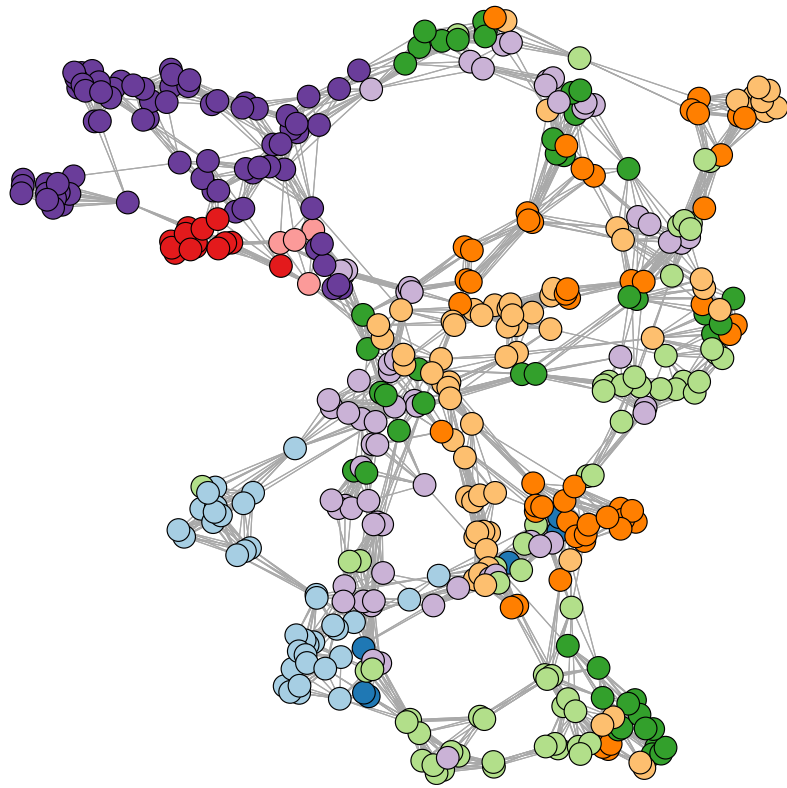
## Datascape des données Tara

- **Datascape sur les données suivantes:** Température, nitrate, concentrations fer, oxygène, salinité, chlorophyle A, lumière, le flux de carbone à 150m
- K=10 , distance euclidienne



# APPLICATIONS: DONNÉES TARA

Datascape des données Tara



*Le graphe capture la structure connue des données*

*Utilité en océanographie, de la notion des géodésiques  
Atlantification des pôles*

# APPLICATIONS: PHYSIONET CARDIOLOGY

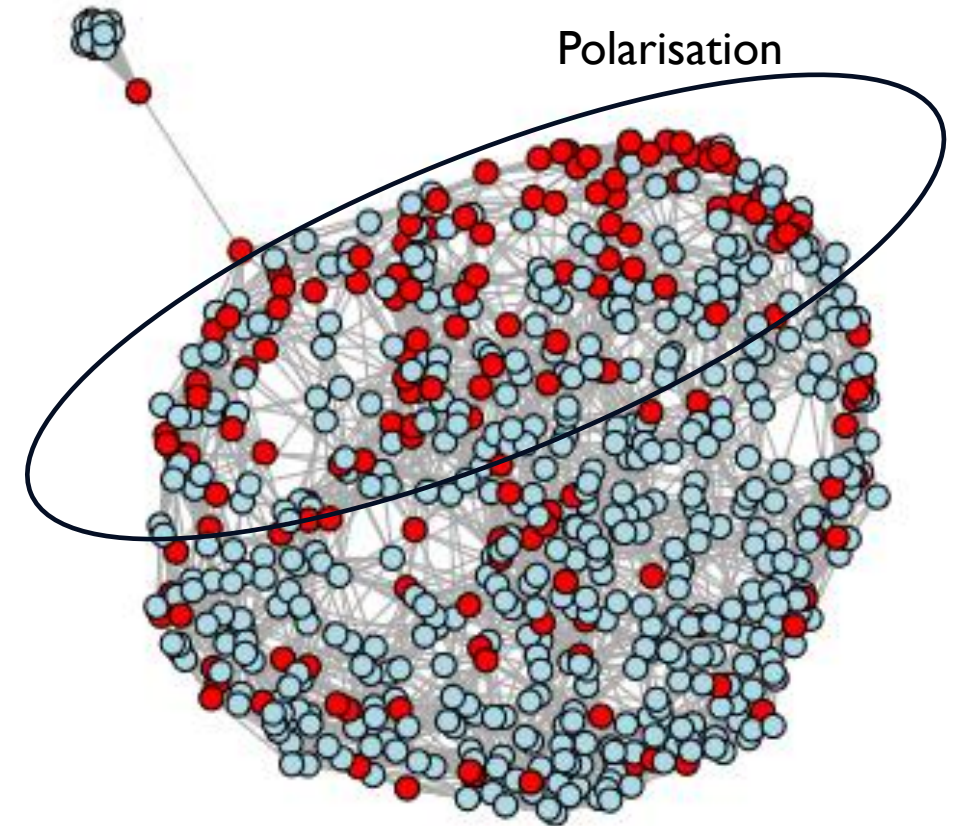
## Application au cas de trajectoires patients

### Données:

- Données cliniques et biologiques des 48h premières en soins intensif
- Données longitudinales
- Température, Tension sys/dias, Poids, NIMAP, ..., Rythme cardiaque, biologie médicale

**Objectif:** Prédire le devenir d'un patient après les 48 premières heures d'admission en soin intensif

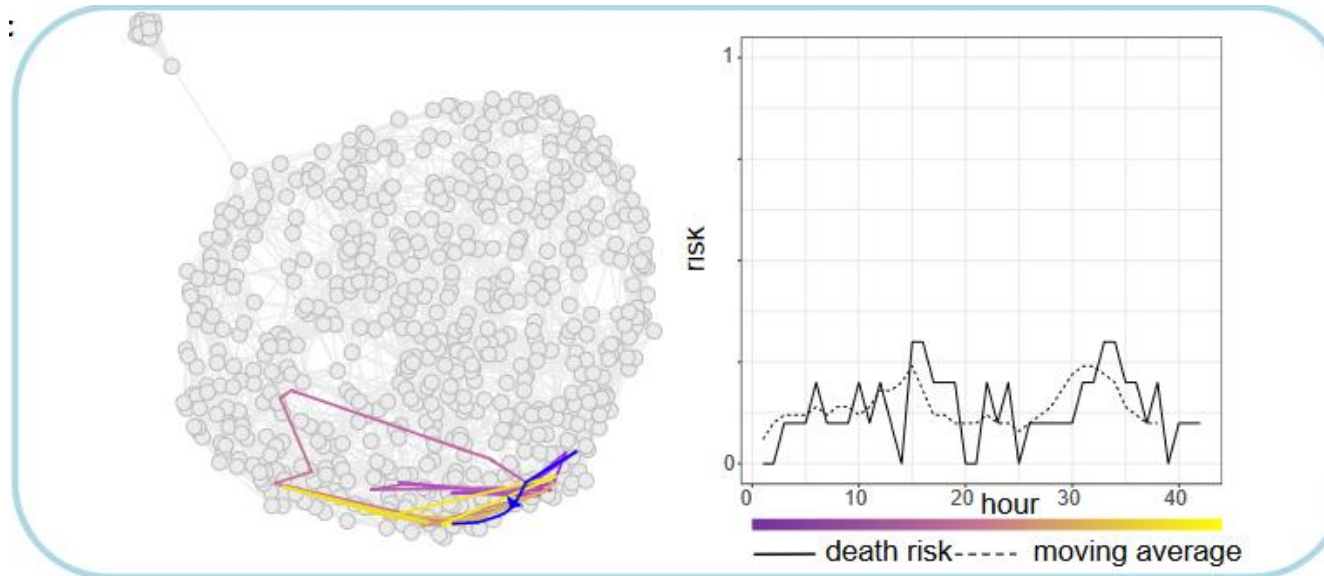
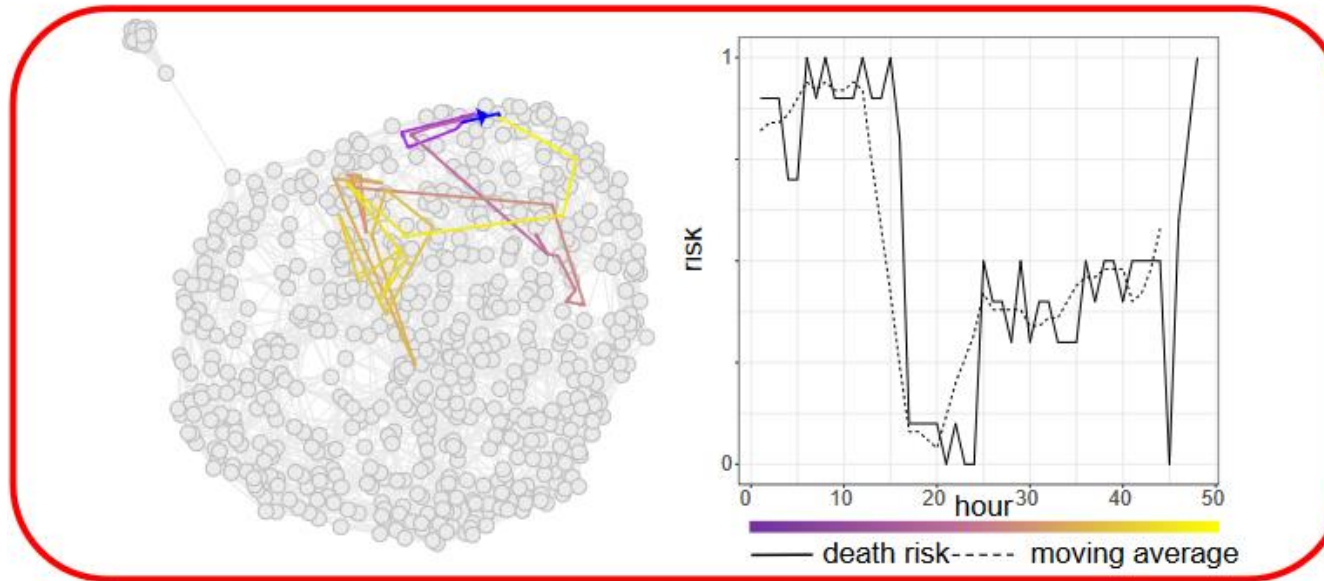
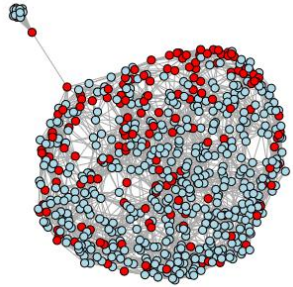
**Datascape:** Sur l'ensemble des dernières mesures de chaque patients



- décès
- survie

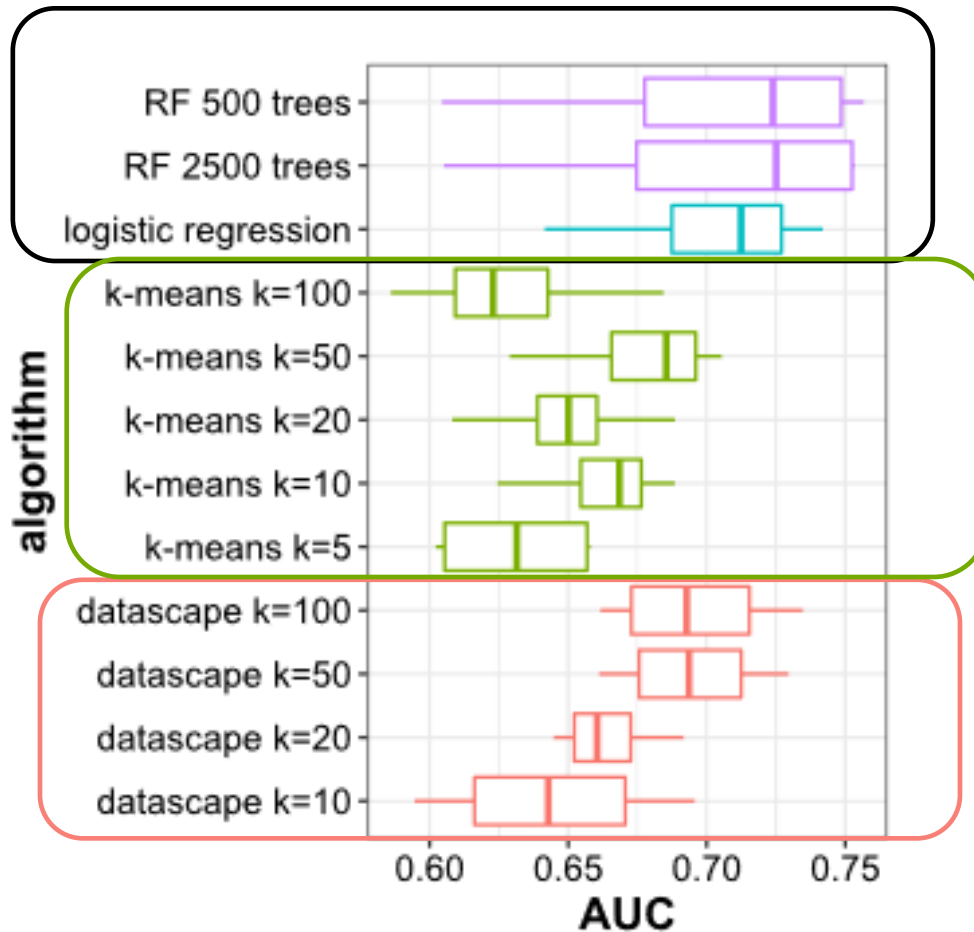
# APPLICATIONS: PHYSIONET CARDIOLOGY

## Application au cas de trajectoires patients



# APPLICATIONS: PHYSIONET CARDIOLOGY

## Application à la prédiction d'évènements



*Calcul de métriques à l'échelle des voisinages (enveloppes)*

*Etude et caractérisations de trajectoires*

*Scores prédictifs sur une structure non supervisée*



# APPLICATIONS EN DATASCIENCE

**COMPLÉTION DE DONNÉES MANQUANTES**

DONNÉES SYNTHÉTIQUES

# APPLICATIONS: DONNÉES MANQUANTES

## Applications pour imputer des données manquantes

**Donnée manquante:** Point d'un jeu de donnée avec une dimension absente

**Objectif:** On cherche à combler les caractéristiques manquantes par des valeurs possible, au regard d'une référence, d'une connaissance préalable

**Problématique:** Comment la connaissance de l'espace sous-jacent nous permet de compléter les données manquantes ? Application au travers du Datascape

Point n°	x	y
1	$x_1$	$y_1$
2	$x_2$	$y_2$
3	$x_3$	$y_3$
4	$x_4$	$y_4$
5	$x_5$	$y_5$
...	...	...

→

Point n°	x	y
1	$x_1$	$y_1$
2	<del><math>x_2</math></del>	$y_2$
3	$x_3$	<del><math>y_3</math></del>
4	$x_4$	$y_4$
5	<del><math>x_5</math></del>	$y_5$
...	...	...

→

Point n°	x	y
1	$x_1$	$y_1$
2		$y_2$
3	$x_3$	
4	$x_4$	$y_4$
5		$y_5$
...	...	...

# APPLICATIONS: DONNÉES MANQUANTES

Applications pour imputer des données manquantes

Point n°	x	y
1	$x_1$	$y_1$
2	$x_2$	$y_2$
3	$x_3$	$y_3$
4	$x_4$	$y_4$
5	$x_5$	$y_5$
...	...	...

Point n°	x	y
1	$x_1$	$y_1$
2	<del><math>x_2</math></del>	$y_2$
3	$x_3$	<del><math>y_3</math></del>
4	$x_4$	$y_4$
5	<del><math>x_5</math></del>	$y_5$
...	...	...

Données complètes

Point n°	x	y
1	$x_1$	$y_1$
4	$x_4$	$y_4$
...	...	...

imputation

Point n°	x	y
2		$y_2$
3	$x_3$	
5		$y_5$
...	...	...

Données incomplètes

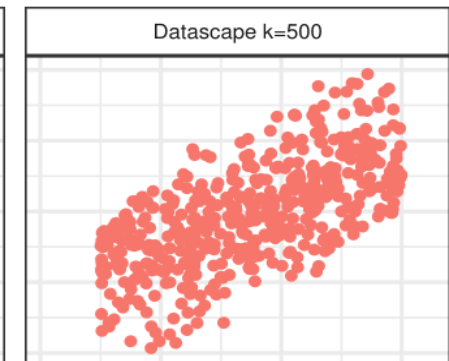
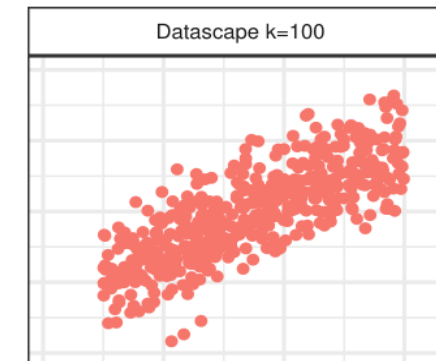
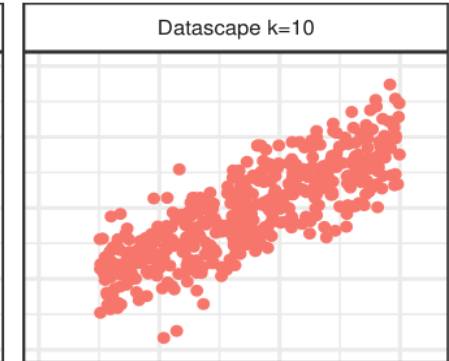
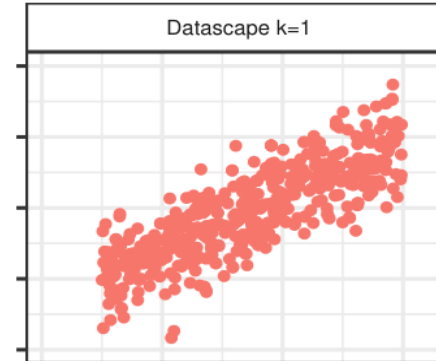
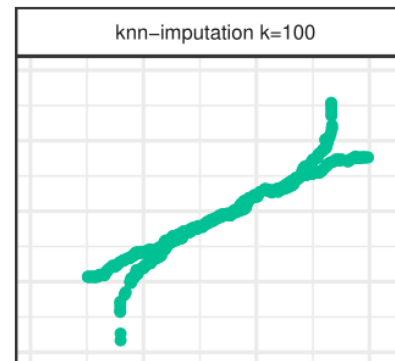
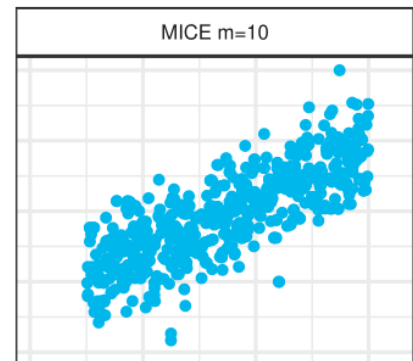
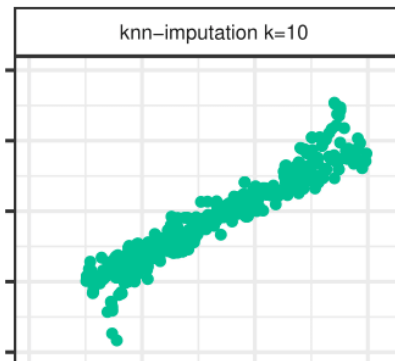
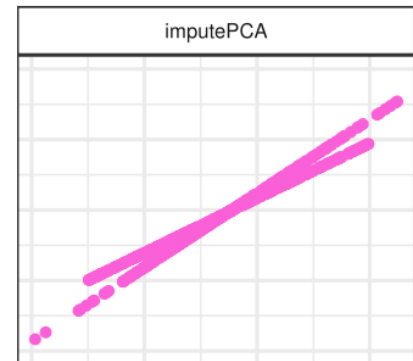
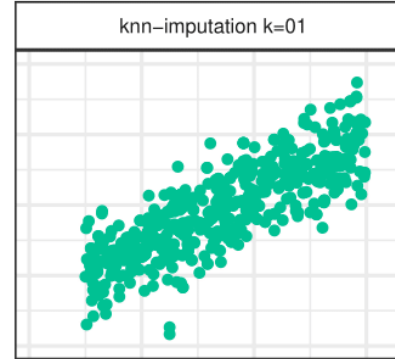
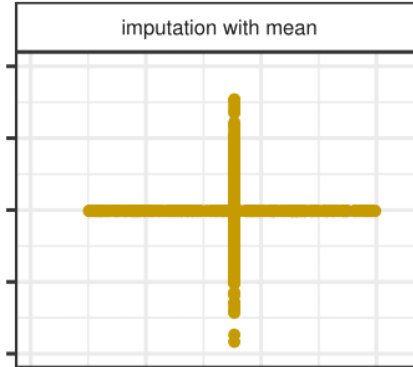
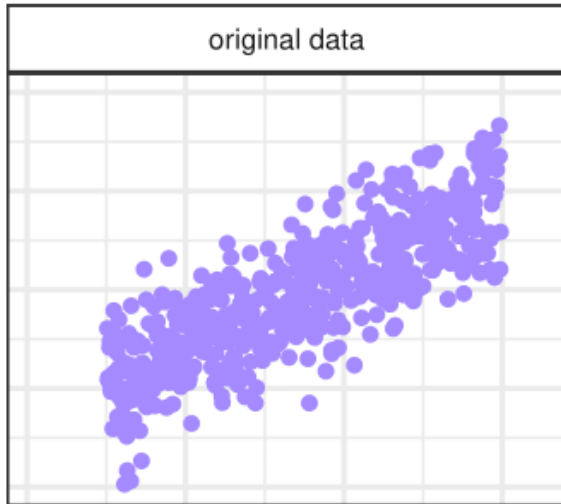
Point n°	x	y
2	$x'_2$	$y_2$
3	$x_3$	$y'_3$
5	$x'_5$	$y_5$
...	...	...



# APPLICATIONS: DONNÉES MANQUANTES

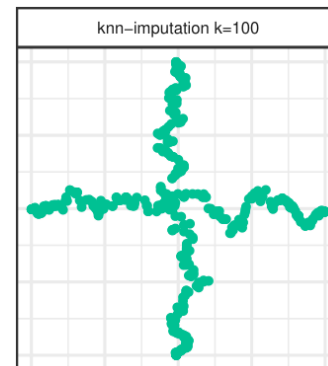
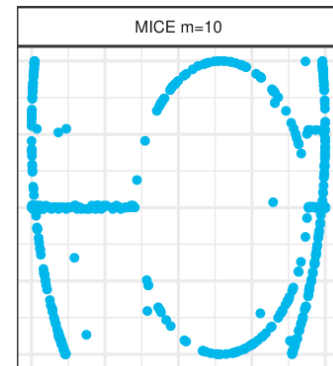
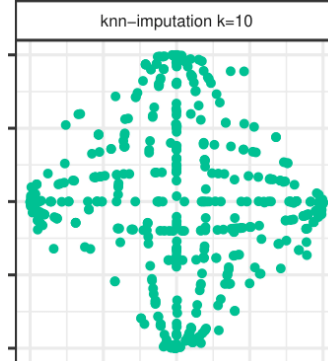
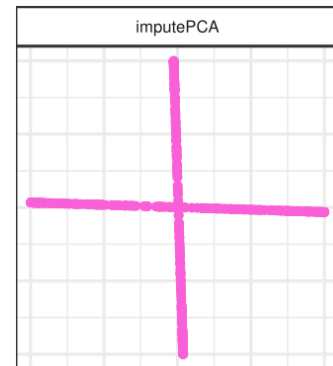
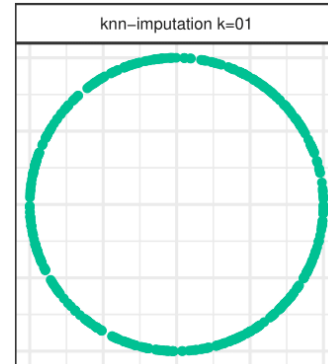
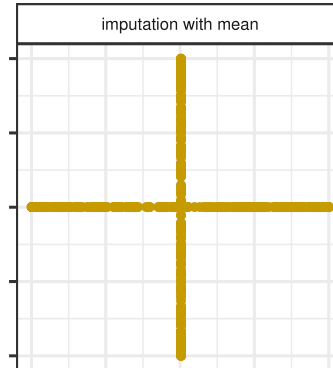
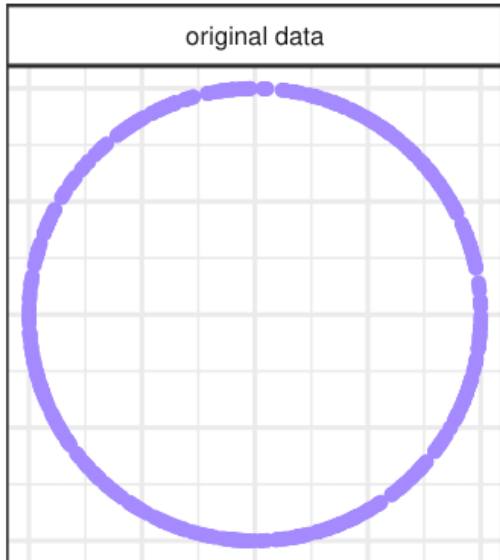
Applications pour imputer des données manquantes

- Données originales
- MICE
- moyenne
- knn-imputation
- PCA
- Datascape

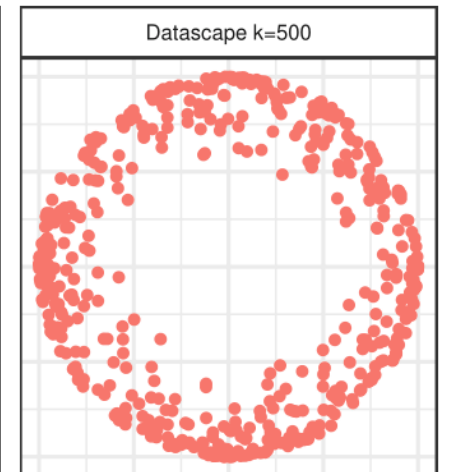
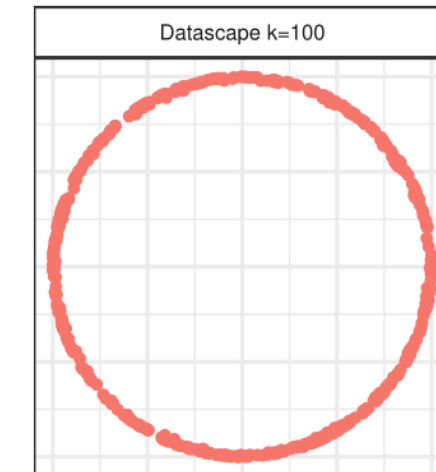
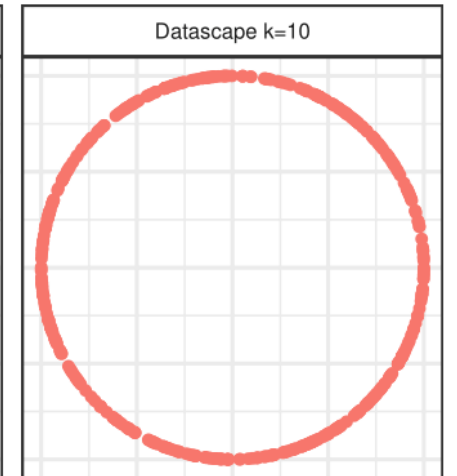
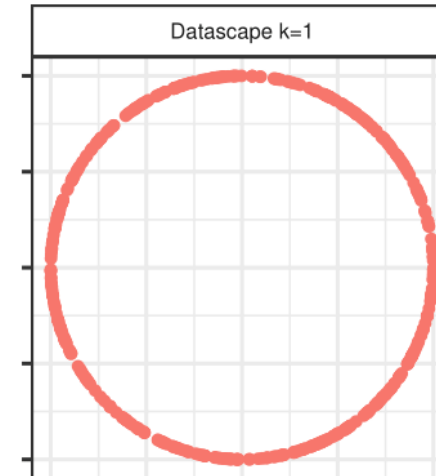


# APPLICATIONS: DONNÉES MANQUANTES

## Applications pour imputer des données manquantes

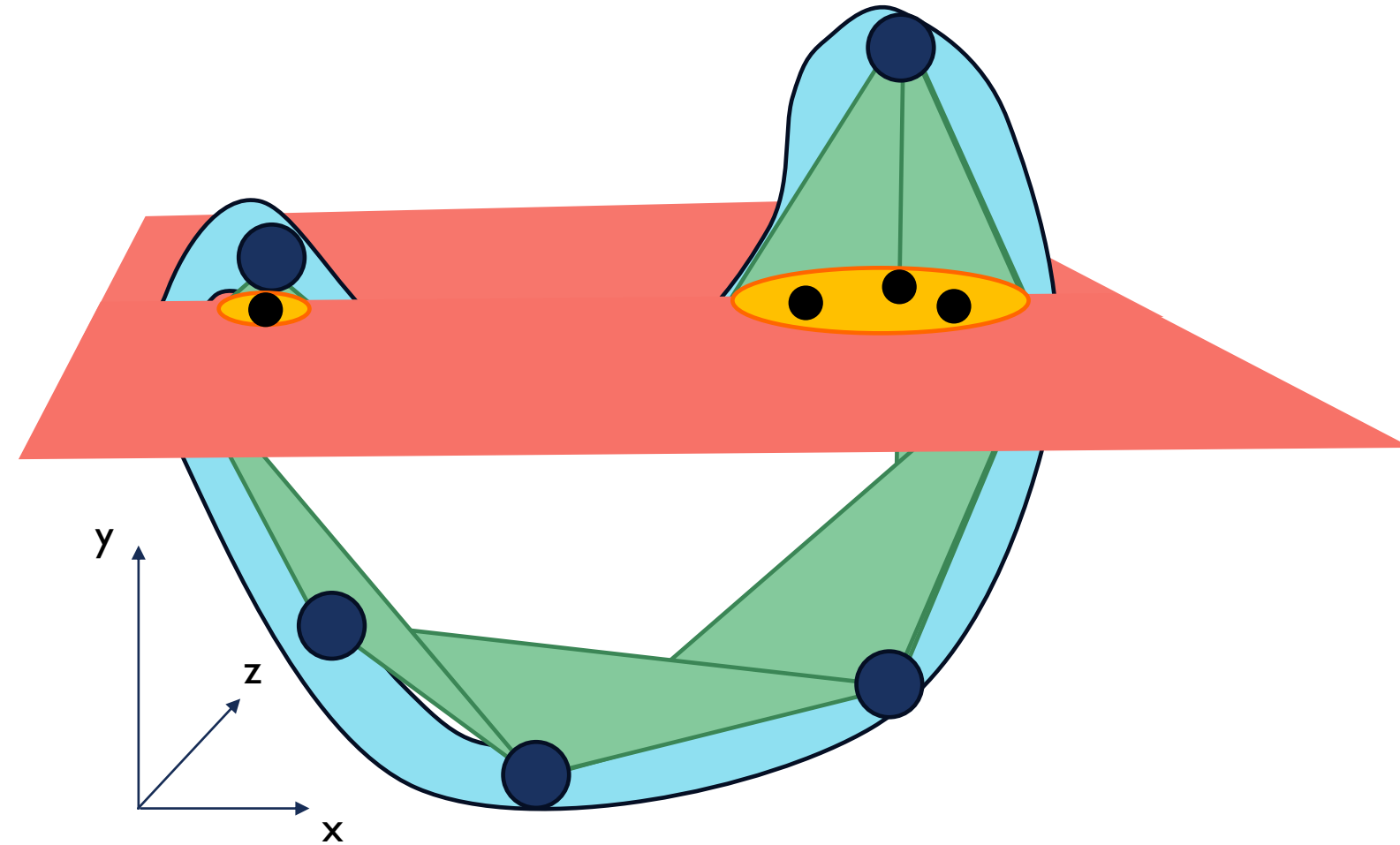


- Données originales
- MICE
- moyenne
- knn-imputation
- PCA
- Datascape



# APPLICATIONS: DONNÉES MANQUANTES

Applications pour imputer des données manquantes

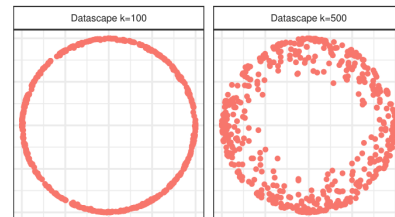
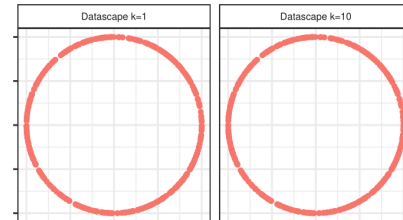
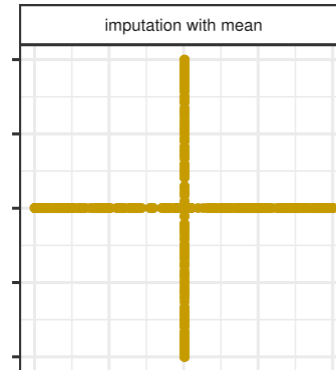
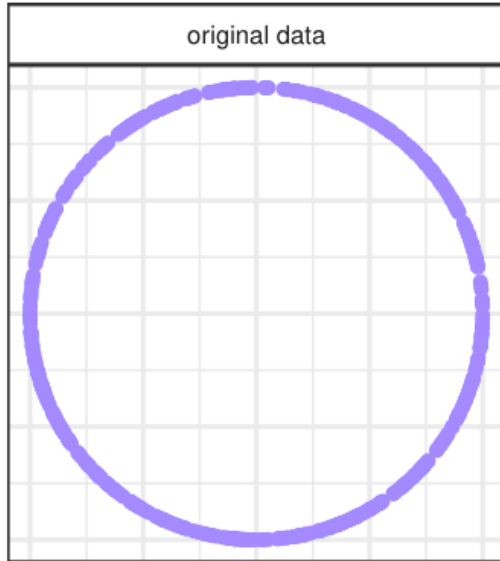


→ Un point avec des données manquantes est un hyperplan

1. Trouver les enveloppes convexes qui intersectent l'hyperplan
2. Echantillonner aléatoirement dans l'intersection

# APPLICATIONS: DONNÉES MANQUANTES

## Applications pour imputer des données manquantes



- *Méthode performante et stable*
- *Importance de prendre en compte la forme de l'espace des données*
- *Réaliser des opérations sur les données qui sont définies sur cet espace*



# APPLICATIONS

COMPLÉTION DE DONNÉES MANQUANTES

**GÉNÉRATION DE DONNÉES SYNTHÉTIQUES**

# APPLICATIONS: DONNÉES SYNTHÉTIQUES

Applications pour générer des données synthétiques

**Donnée synthétique:** Point généré à partir d'un jeu de donnée

**Objectif:** On cherche à créer des points ressemblants, mais différents

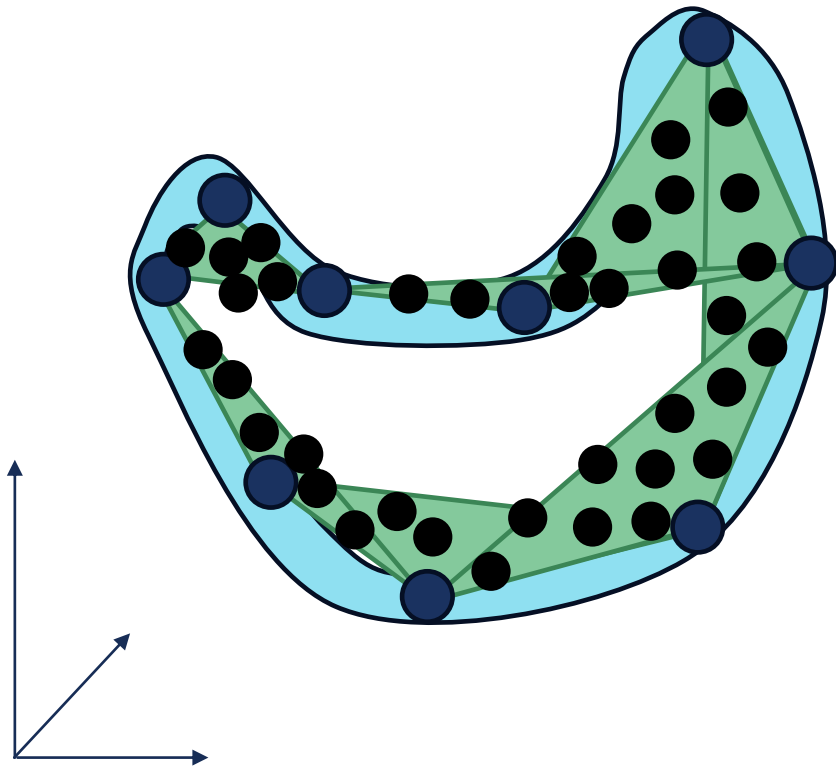
**Enjeux:**

- Manipulations de données sensibles/privées
- Enrichissement de données rares et coûteuses

**Problématique:** Comment la connaissance de l'espace sous-jacent nous permet de générer des données synthétiques ? Application au travers du Datascape

# APPLICATIONS: DONNÉES SYNTHÉTIQUES

Applications pour générer des données synthétiques



## Méthode pour générer un point dans le datascape

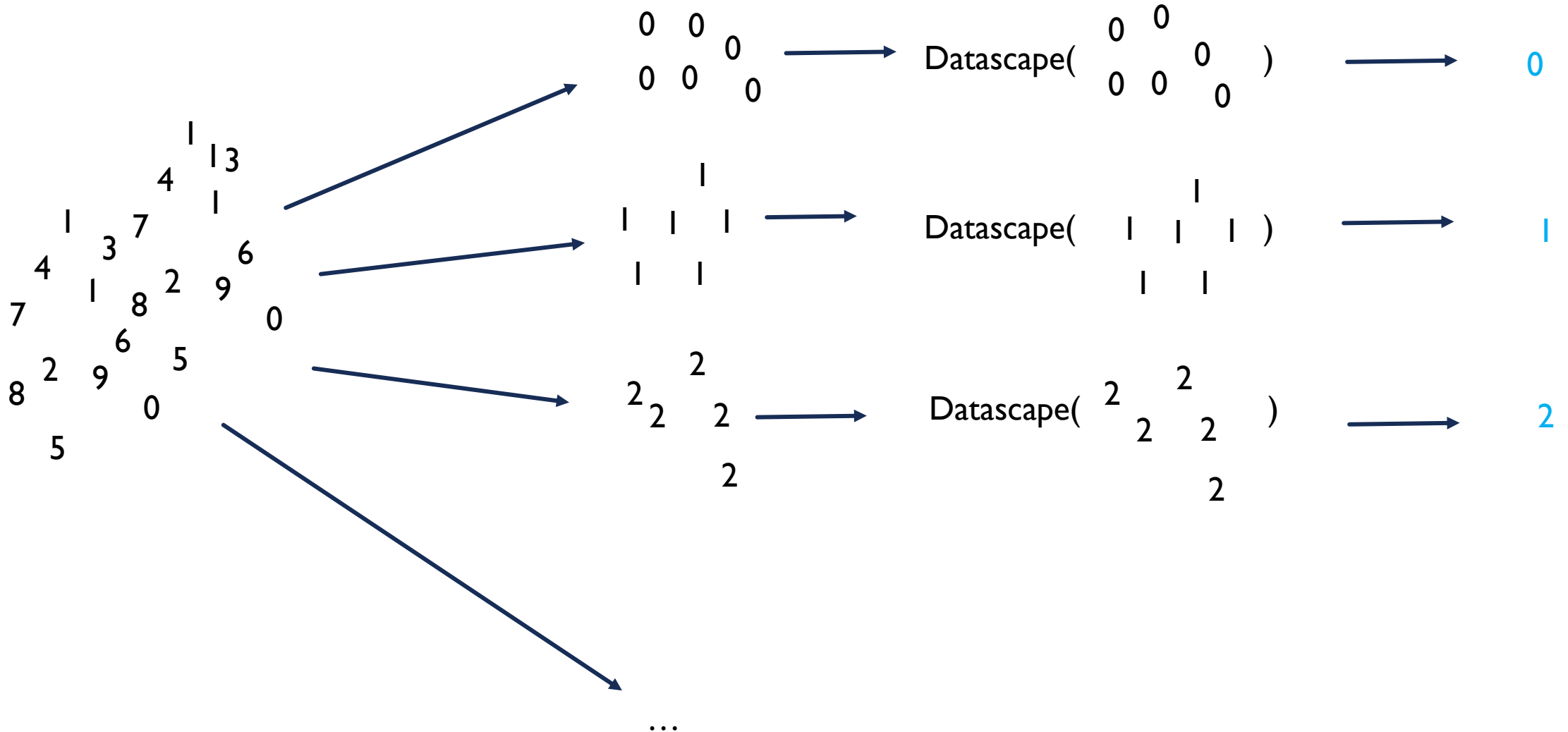
1. Sélection d'une enveloppe du Datascape au hasard
2. Echantillonnage aléatoire dans cette enveloppe





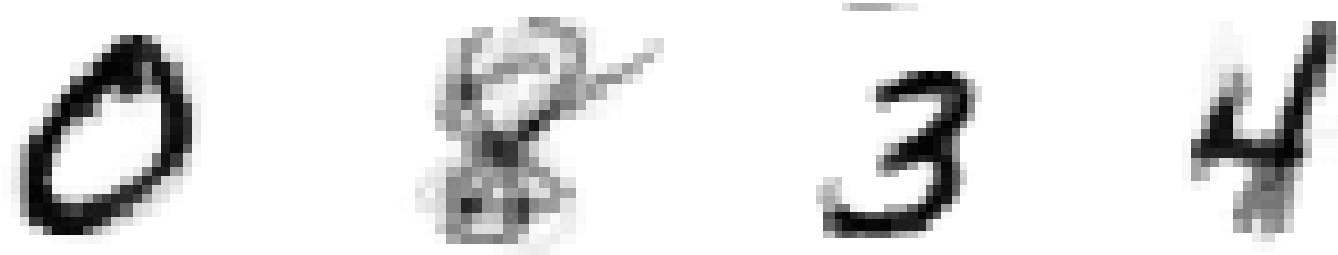
# APPLICATIONS: DONNÉES SYNTHÉTIQUES

## Applications pour générer des données synthétiques



# APPLICATIONS: DONNÉES SYNTHÉTIQUES

Applications pour générer des données synthétiques

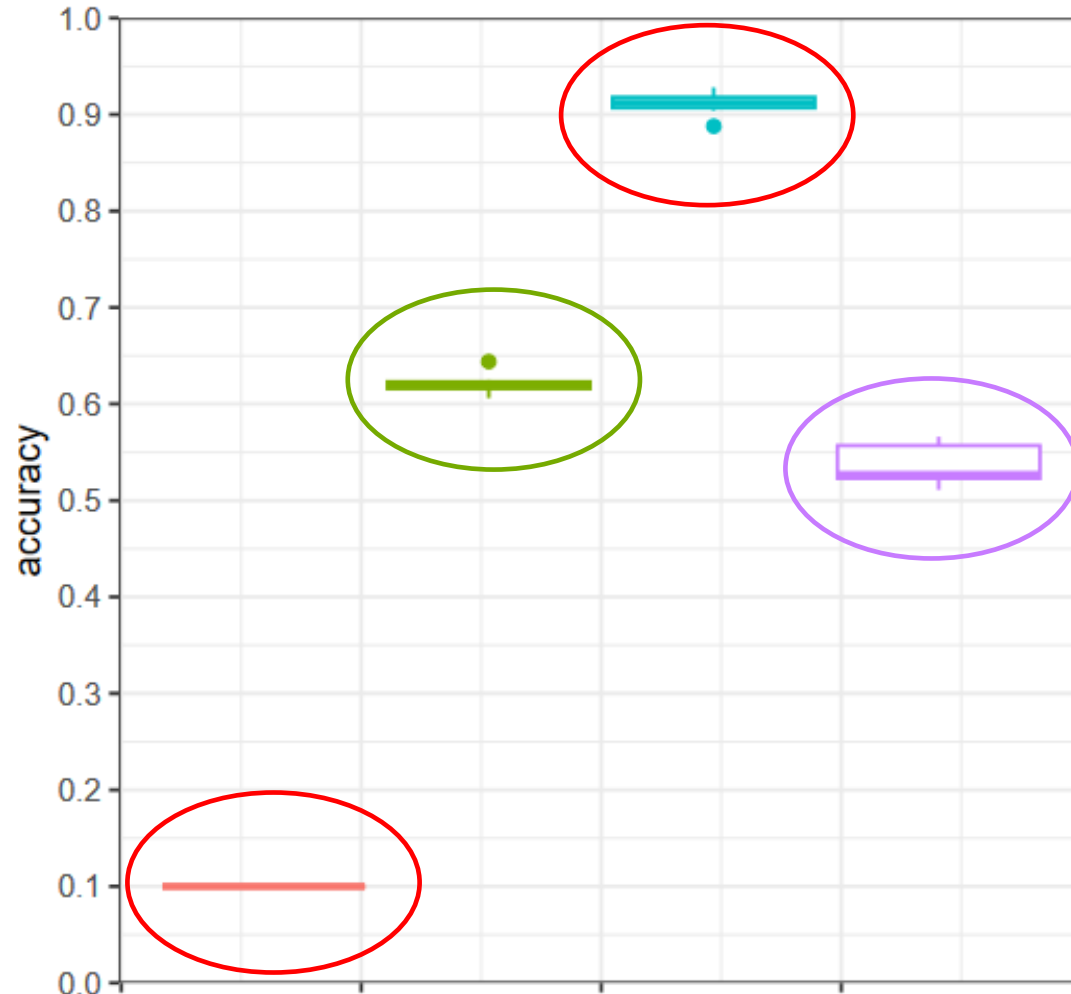


**Comment quantifier la qualité des données générées et comparer avec algorithme état de l'art**

- **Model compatibility:** Entraîner un classifieur type `random_forest` sur les données générées et comparer ses performances de classification sur de vrais données
- Algorithmes état de l'art: **VAE** (variational auto-encoder)

# APPLICATIONS: DONNÉES SYNTHÉTIQUES

Applications pour générer des données synthétiques



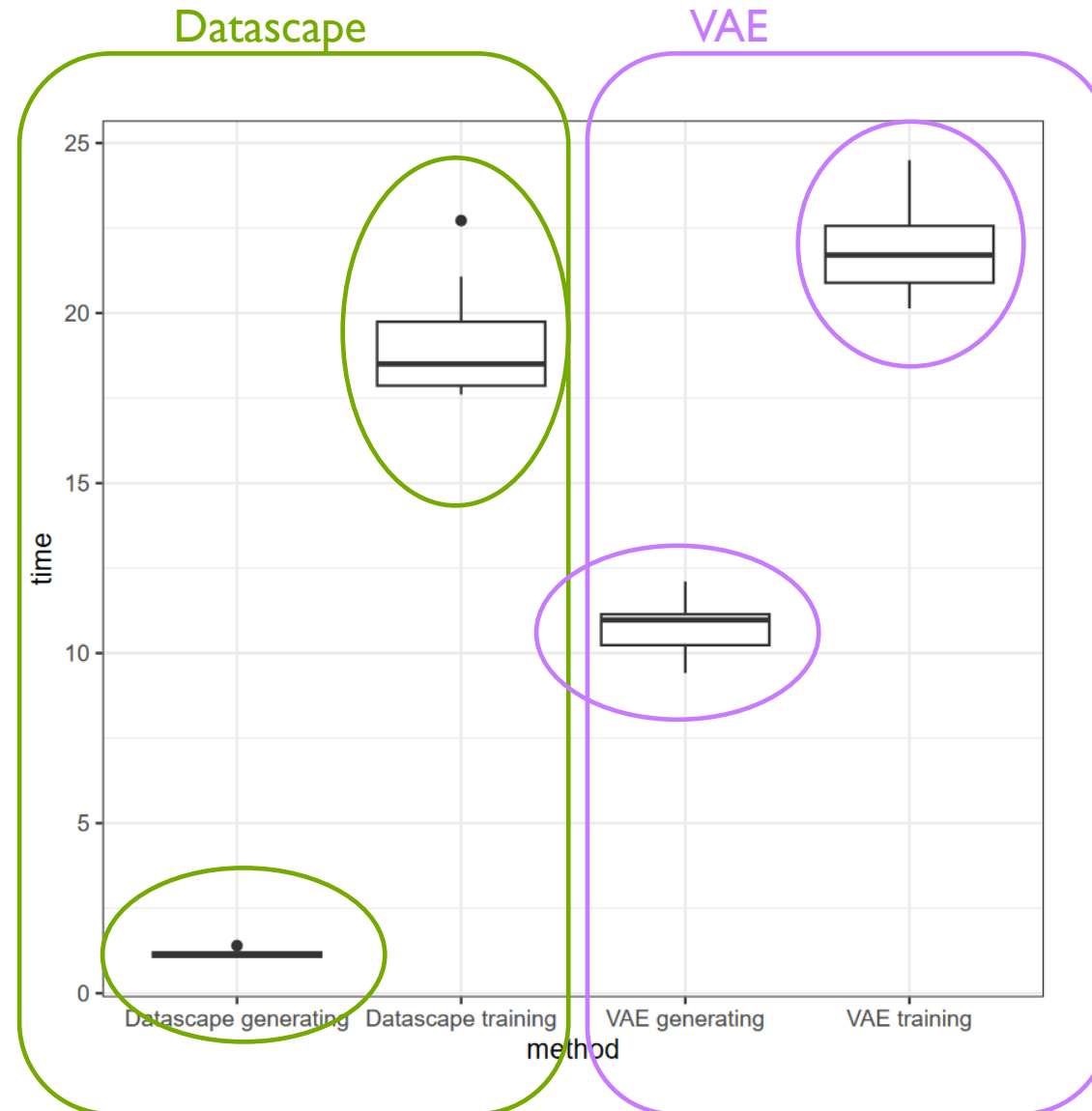
method

- random classifier
- RF on Datascape synthetic dataset
- RF on the original dataset
- RF on VAE synthetic dataset

# APPLICATIONS: DONNÉES SYNTHÉTIQUES

Applications pour générer des données synthétiques

Temps de calcul



- Entraînement
- Génération de données

# APPLICATIONS: DONNÉES SYNTHÉTIQUES

---

## Applications pour générer des données synthétiques

Le Datascape offre une méthode pour générer des données synthétiques

Le Datascape peut traiter des données de haute dimension avec une étape préliminaire (784 ici)

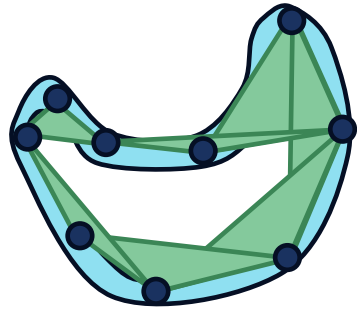
Les performances de classifieurs entraînés sur le Datascape sont meilleures que sur des algorithmes état de l'art de génération de données (VAE) basé sur des réseaux de neurones

Importance de prendre en compte l'espace dans lesquels on travail



# CONCLUSIONS ET PERSPECTIVES

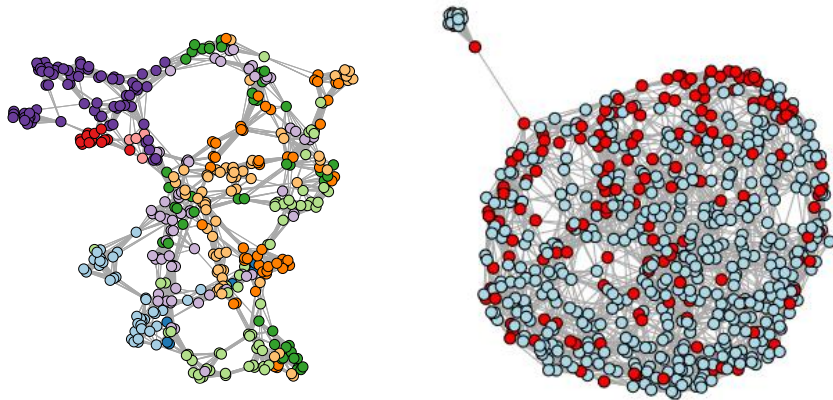
# CONCLUSIONS



## FORME / CONTOURS / DISTANCES

Graphe + ensemble d'enveloppes convexes

- *Combinaisons de différentes abstractions*
- *Implémentation*



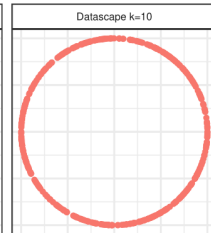
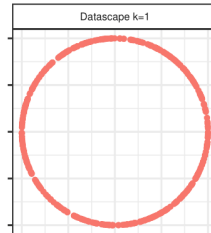
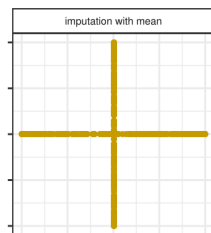
Méthode versatile utilisable dans de nombreux domaines

- *Calculer des métriques locales*
- *Etudier des trajectoires*
- *Etudier la structure sous-jacente des données*

Applications nombreuses en science de la donnée

- *Importance de prendre en compte la forme de l'espace des données*
- *Réaliser des opérations sur cet espace*
- *Un seul paradigme*

3 - 100

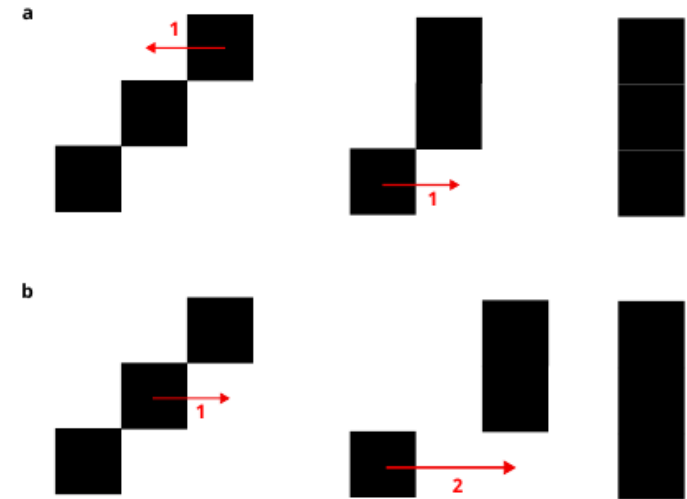


## Implémentation

Utilisation de distances autres (wasserstein)

Datascape supervisé (choix de la distance)

Optimisation (k-nn, graphes)



## Relative NN-Descent: A Fast Index Construction for Graph-Based Approximate Nearest Neighbor Search

Naoki Ono  
n\_ono@hal.t.u-tokyo.ac.jp  
The University of Tokyo  
Tokyo, Japan

Yusuke Matsui  
matsui@hal.t.u-tokyo.ac.jp  
The University of Tokyo  
Tokyo, Japan

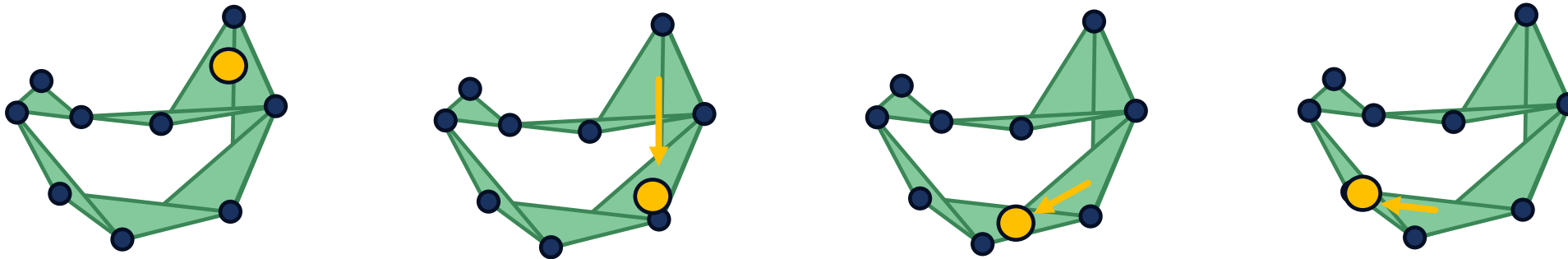
### ABSTRACT

Approximate Nearest Neighbor Search (ANNS) is the task of finding

GIST1M [15] dataset, which contains 1 million vectors of 960 dimensions. Construction of HNSW [19], a representative of graph-based

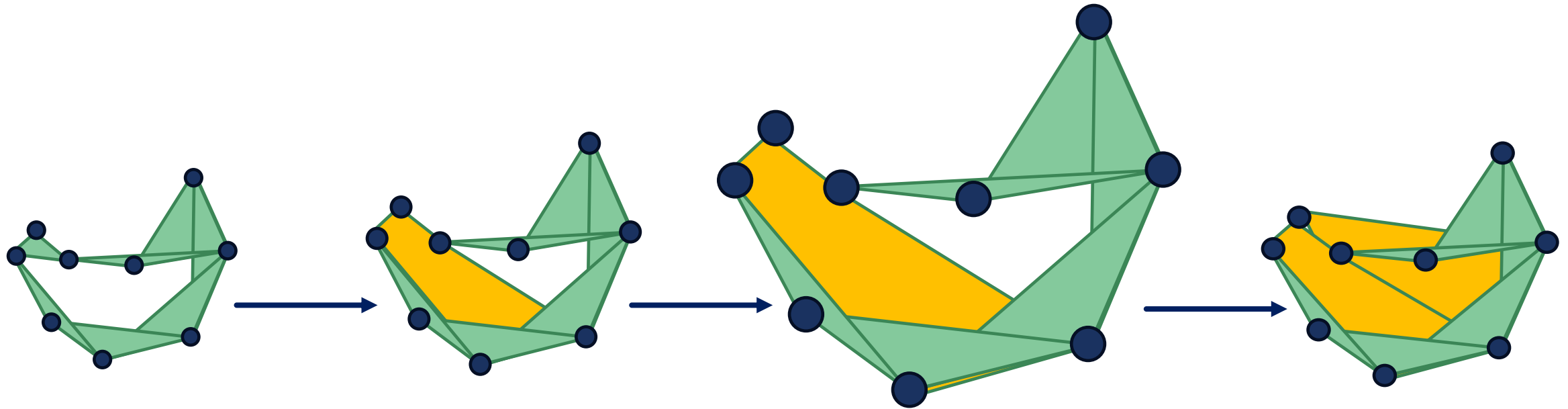


Suivi de l'évolution d'un point dans le Datascape: géodésiques, trajectoires



t

Caractériser l'évolution d'un Datascape

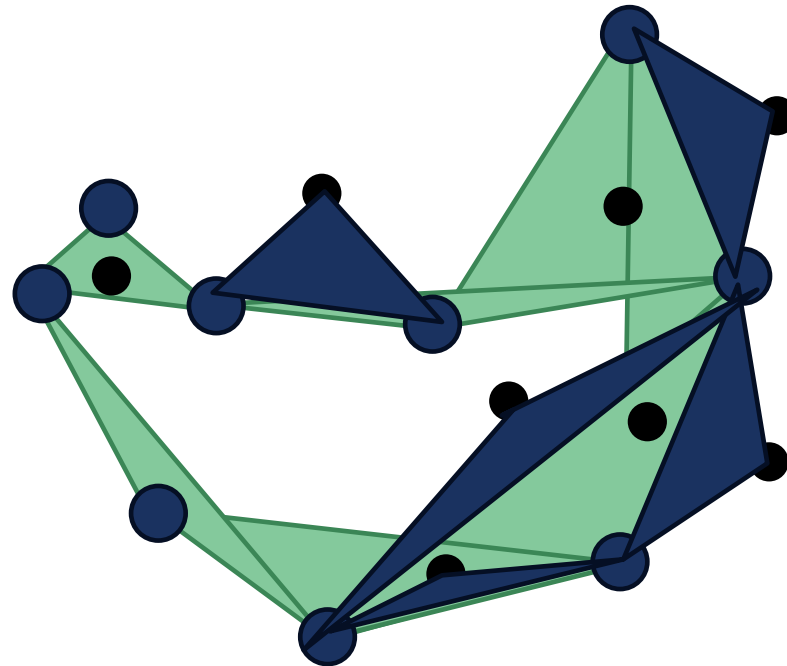


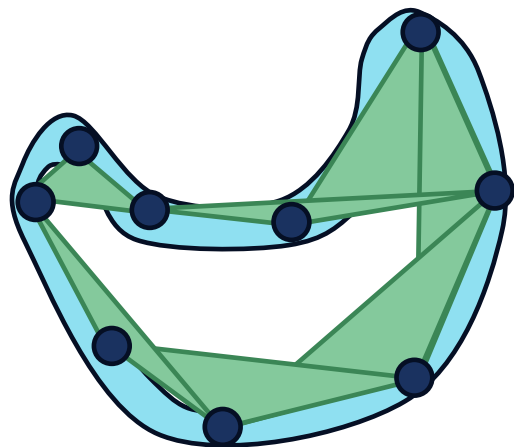
t

Mise à jour en temps réel d'un Datascape

Construction frugale

Caractériser la nouveauté d'un point (perturbation distances, forme du datascape, ...)





Merci de votre attention



BIO LOGBOOK

Precision Healthcare Technology<sup>62</sup>