



Ouvrons la boîte à outils du Data Scientist pour analyser les données du SNDS

Martin PRODEL

« Data Scientist » (Ingé, PhD) chez HEVA

Meetup SNDS #3

Différents Concepts



DATA SCIENCE

*Sur les banc de l'école
ou la fiche de poste*



INTELLIGENCE ARTIFICIELLE

*Le maxi
buzz word*



BIG DATA

*SNDS,
what else ?*



DATA MINING

La philosophie



DEEP LEARNING

Boîte à outils n° 1



MACHINE LEARNING

*Boîte à outils n° 2
→ ma préférée*

La mise en pratique du Machine Learning

1. Profilage de patients avec la Classification
2. Les parcours de traitements avec le Clustering

Mise en œuvre de 2 approches sur un cas d'étude : données SNIIRAM

≈ 96 500 PV.VIH*
en 2013



dont 3 373 incidents

Rétrospective de 5 ans pour étudier les comorbidités et suivi 1 an pour étudier les coûts recours aux soins

→ Question : présence de sous-profils atypiques au regard des coûts ?

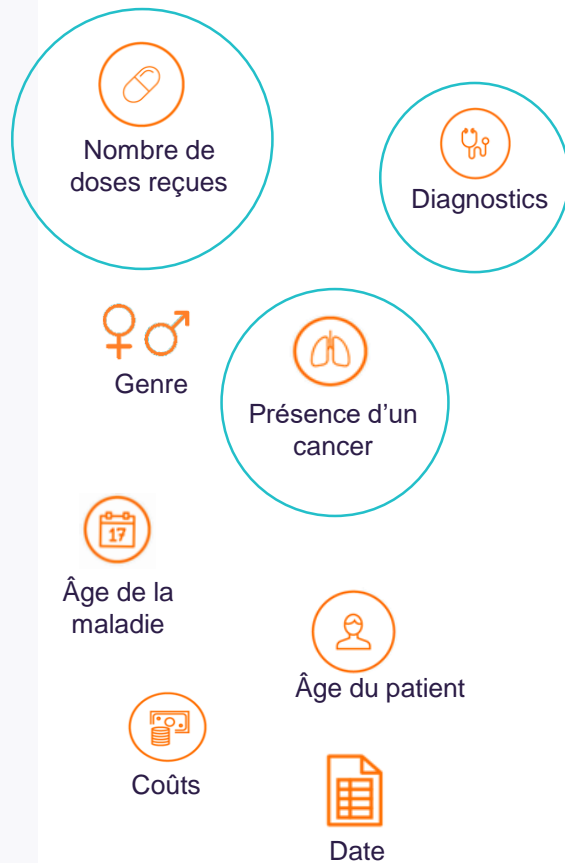
Suivi 2 ans afin de détecter leurs **délivrances d'antirétroviraux (ART)** :

Monothérapie / bithérapie / trithérapie 

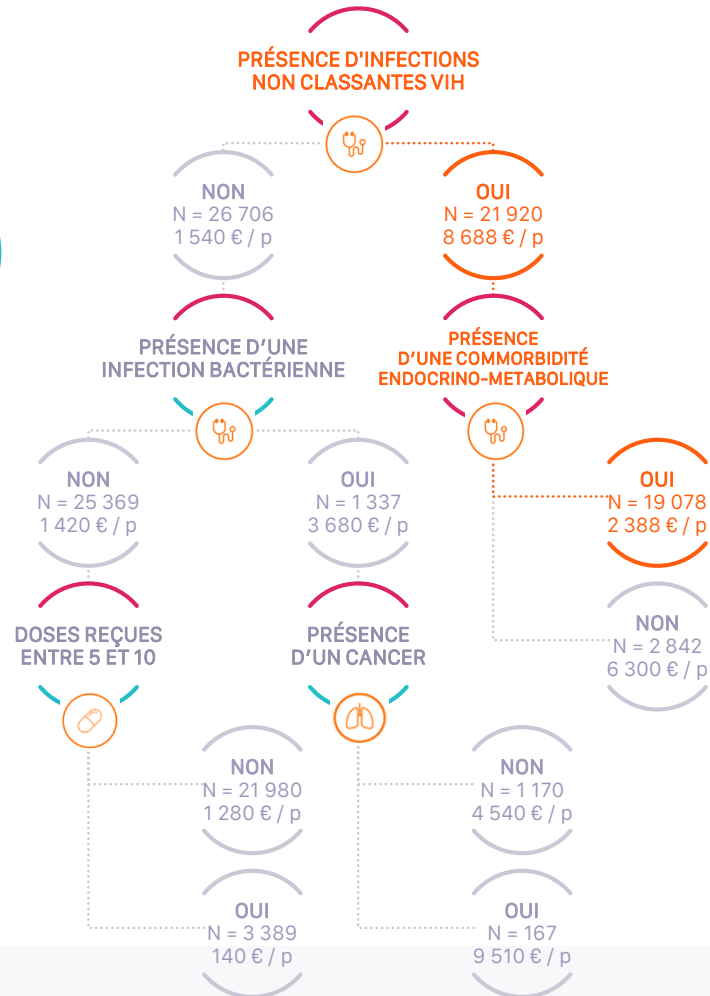
→ Question : présence de trajectoires de soins types, de switch dans les traitements, durées sous traitements ?

Question 1 : utilisons un algorithme de Machine Learning qui ...

SÉLECTIONNE AUTOMATIQUEMENT
LES **VARIABLES PERTINENTES**
PARMI LES DIZAINES DISPONIBLES



POUR CRÉER AUTANT
DE **PROFILS PATIENTS** QUE BESOIN



EN FONCTION
D'UN **OBJECTIF PRÉCIS**

Expliquer le **coût**

(Ou encore mesurer des risques :
ré-hospitalisations ou
complications)

EXEMPLE D'UN PROFIL PATIENT

 N 19 078 (63%)
PROFIL 1 COÛT MOYEN 2 388 €

CARACTÉRISTIQUES :

1. Présence d'infections non classables VIH
2. Présence d'une comorbidité endocrino-metabolique

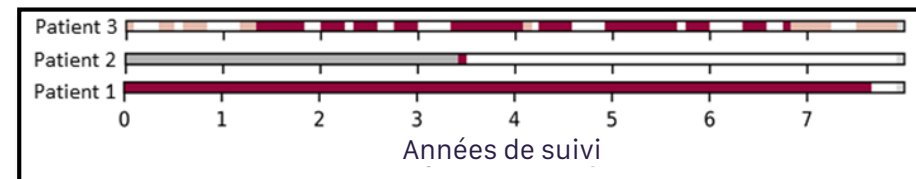
Question 2 : analyser les parcours thérapeutiques

But : Comprendre **les enchainement possibles des traitements** chez les patients d'une cohorte (les incidents) dans les 2 après leur inclusion

1. Pour **l'ensemble la cohorte**
2. En ayant une **vision temporelle précise**
3. De manière **lisible**

Evènements repérables dans le SNDS :

- Délivrances des traitements du VIH



Exemple de 3 patients vectorisés



Méthodologie : utiliser la technique de Machine Learning « TAK* »

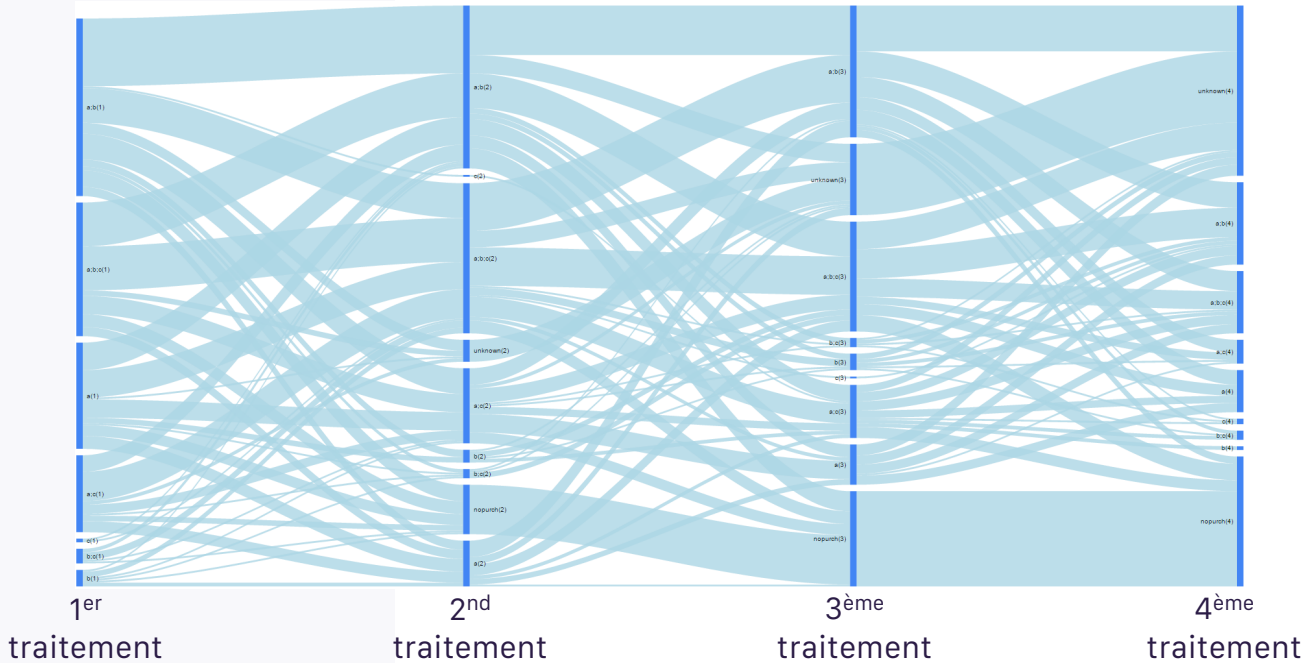


- **Modélisation vectorielle** de chaque patient et son parcours
- Recherche d'enchâînements communs et **clustering de trajectoires similaires**
- Rendu graphique complet grâce au TAK

* *Time sequence Analysis through K-clustering, by HEVA.*

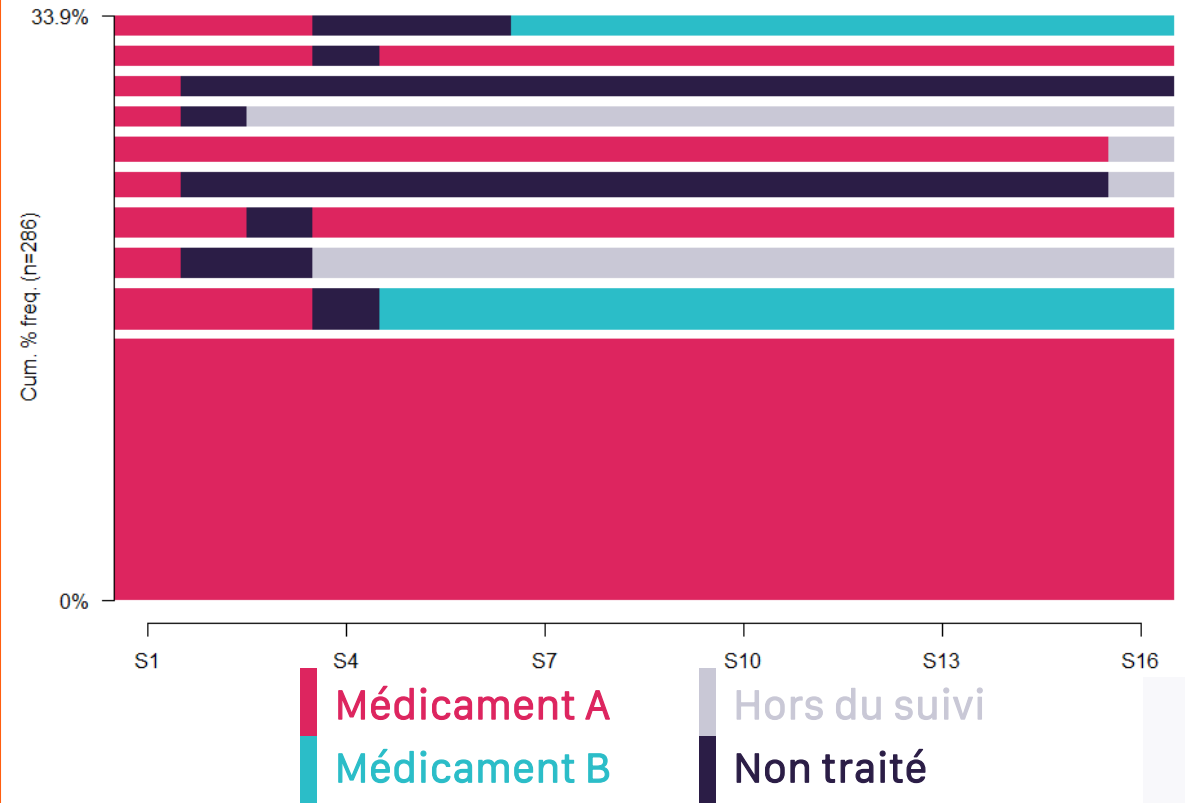
Solutions possibles :

« Sankey Diagram »



- ✓ 1. Exhaustif
- ✗ 2. Manque de **précision dans la temporalité**
- ✗ 3. Manque de **lisibilité**

« TraMineR »



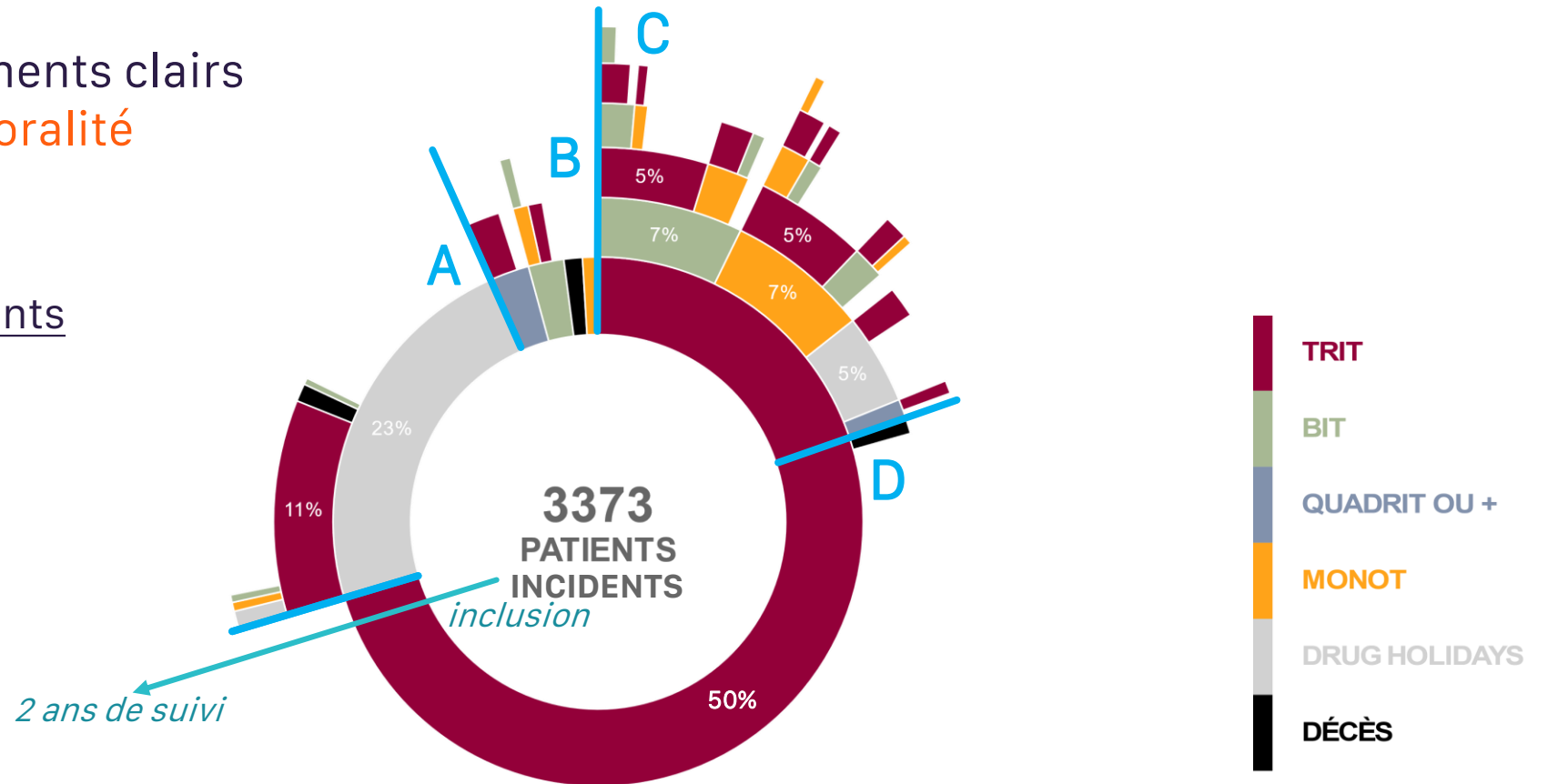
- ✗ 1. Seuls **34% de la cohorte sont représentés**
- ✓ 2. Précis sur la temporalité
- ✓ 3. Lisible

Autre solution

Le graphe « soleil »

- ✓ 1. Exhaustif, enchaînements clairs
- ✗ 2. Imprécis sur la **temporalité**
- ✓ 3. Lisible et compact

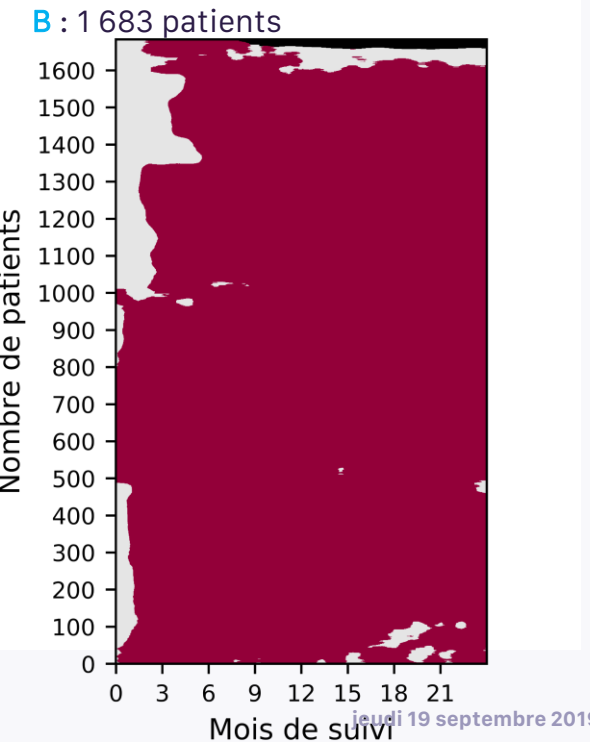
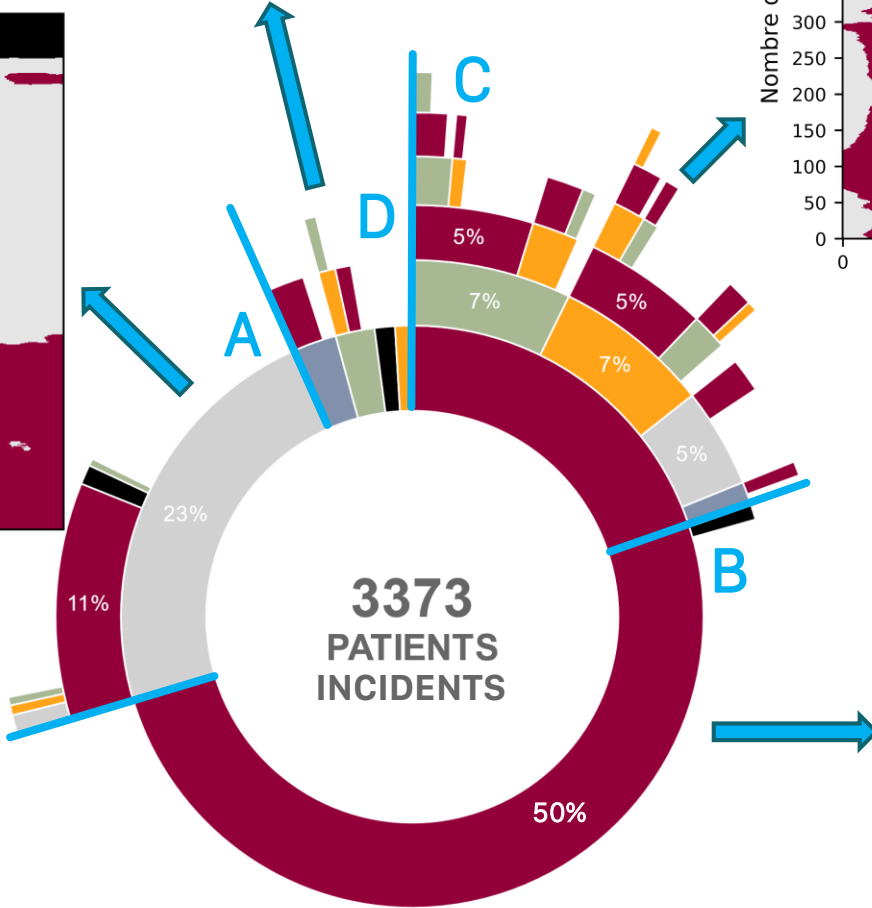
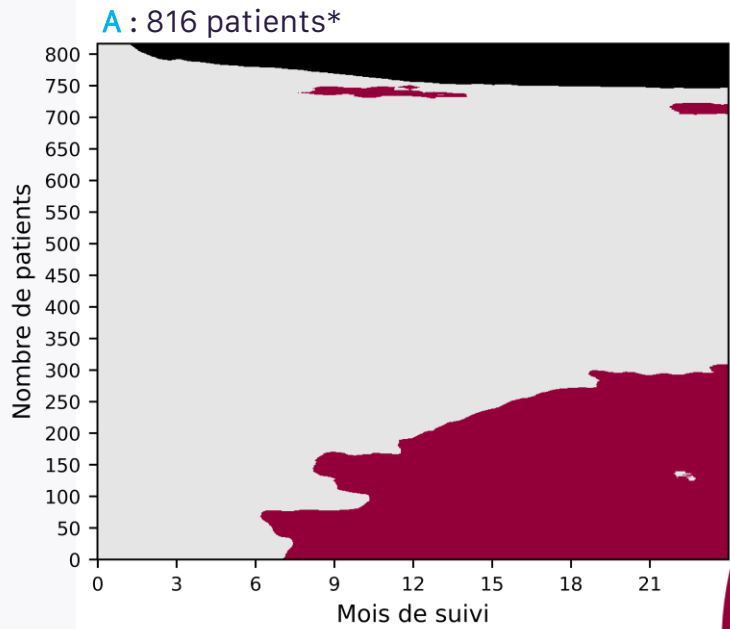
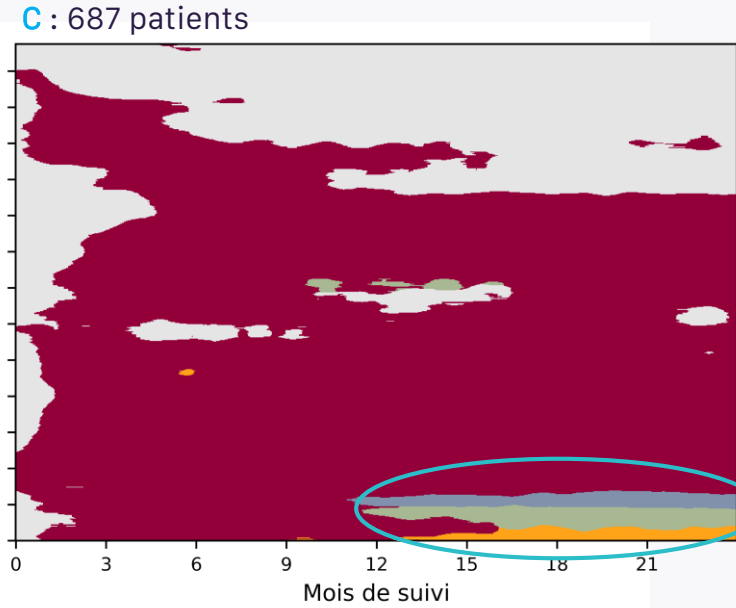
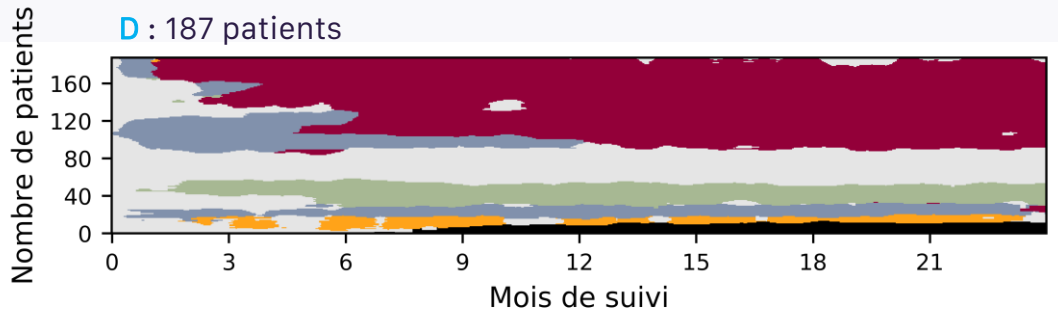
De plus, des patients différents dans le même bloc (ex : A)



TAK

Time sequence
Analysis through
K-clustering

by HEVA



- TRIT
- BIT
- QUADRIT OU +
- MONOT
- DRUG HOLIDAYS
- DÉCÈS

* Les 816 patients incluent à la fois les patients du groupe A issus du sunburst et les patients décédés du groupe D du sunburst.

TAK : La technique

1

- Patients non ordonnés

2

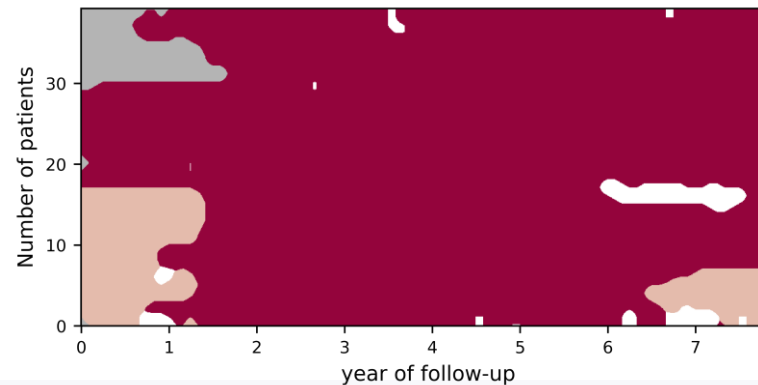
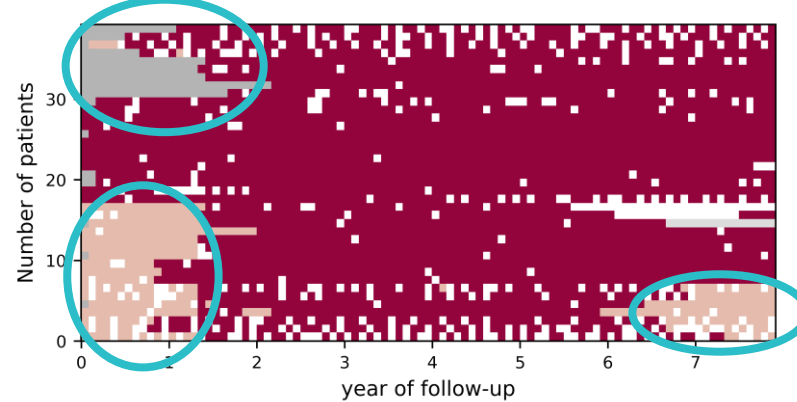
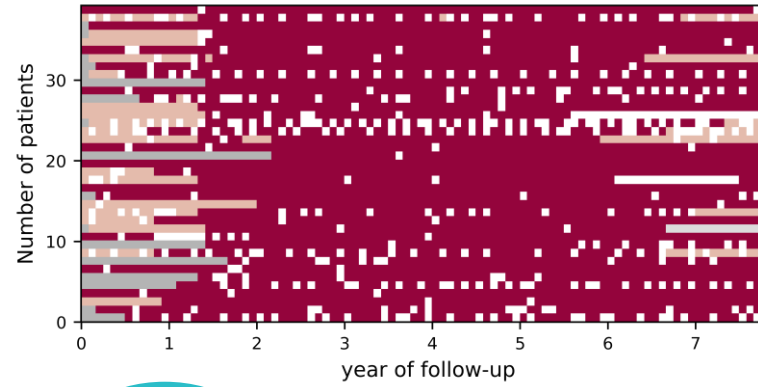
- Patients ordonnés selon leur similarités

→ Dégage les tendances

3

- Image lissée

→ Améliore la lisibilité



Technique :

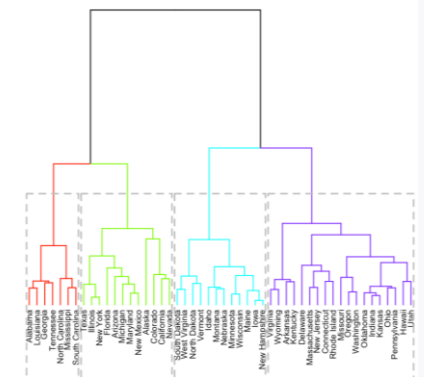
- *Modélisation vectorielle*
- *Discretisation du temps*

HEVA

Technique :
Hierarchical Agglomerative Clustering

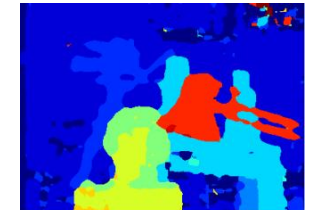
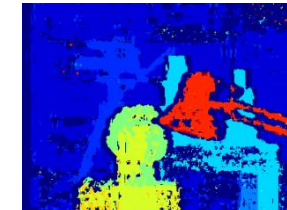
→ *Machine Learning*

Cluster Dendrogram



Technique : *ModeFilter*

→ *Image processing*



Le retour d'expérience d'un data scientist SNDS

- Très important : poser une question d'étude précise
- Les données sont là : SNDS
- La technologie « Machine Learning » est disponible pour tous
 - Les outils sont là, il manque le data scientist (l'orfèvre des temps modernes)
 - Plus de monde = plus de publications = démocratisation
- La technologie seule ne vaut rien → comité scientifique et médical



Merci de votre attention

Des questions ?

Tak : merci (Danois)

Tack : merci (Suédois)

Takk : merci (Norvégien)

Tusen takk : Milles mercis (Norvégien)