

Vendredi 2 avril
9h30-12h30
Visio



R à l'usage des
sciences sociales



Luc-Olivier Hervé
data scientist, Cour des comptes
& Doctorant-chercheur, LEMNA



Visualiser et catégoriser des trajectoires :
panorama des méthodes existantes,
application de l'analyse de séquences par appariement optimal
et démonstration d'une nouvelle application dédiée : ViCaTraj*



Application développée de manière collaborative (Départements de Loire-Atlantique et de l'Isère ; Élie Chosson, MRIE : www.mrie.org) permettant d'avoir accès à une interface utilisateur basée sur l'articulation de packages existants (TraMineR ; WeightedCluster ; ggalluvial notamment).

-
**Visualiser et catégoriser des trajectoires
(Luc-Olivier Hervé)**

Présentation personnelle & introduction

 **Méthodes de typologie de « trajectoires » ?**

 **Analyse de séquences : présentation & illustration**

 **Application *ViCaTraj* & démonstration**

Conclusion & échanges

Présentation personnelle & introduction



Luc-Olivier Hervé :
Doctorant-chercheur, LEMNA
Data scientist, Cour des comptes



Hier

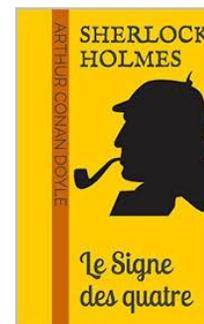
=> Chargé d'études

Travaux et entrepôt de données pour les maj du
Service Public de l'Insertion (SPI) du Département de Loire-Atlantique
(typologies pour identifier : profils-types & trajectoires ; open data etc.)

Aujourd'hui

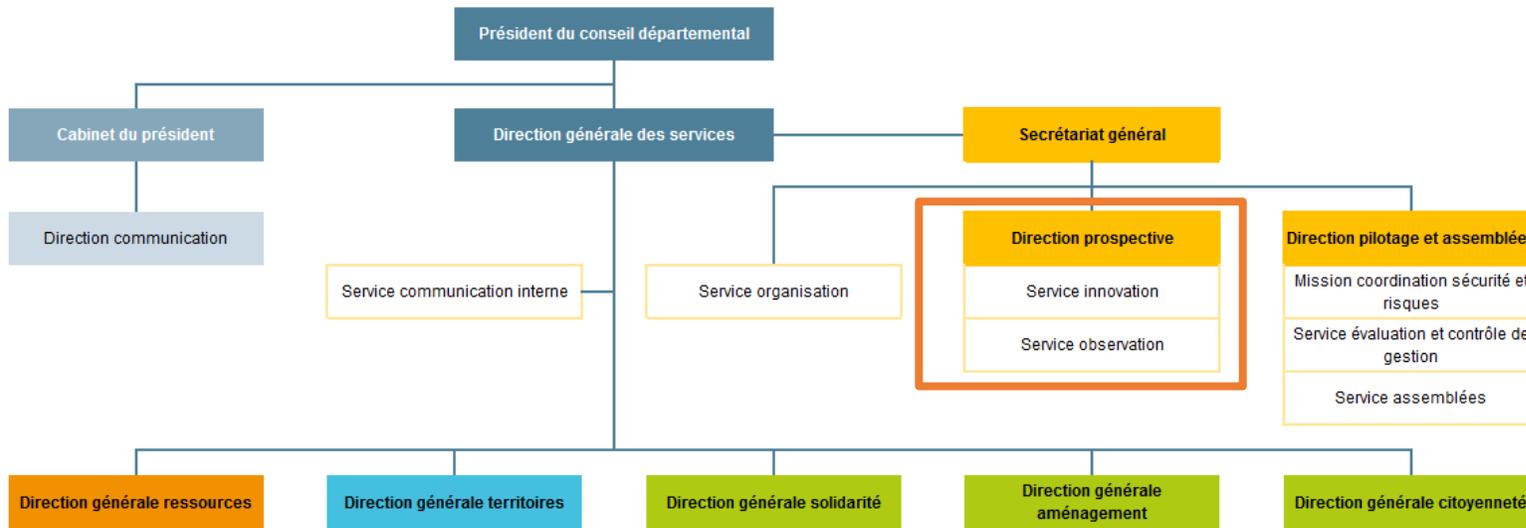
=> Doctorant-Chercheur (LEMNA) : trajectoires d'accompagnement et
de sortie du RSA en Loire-Atlantique & déterminants

=> Data scientist, 5° chambre de la Cour des comptes : poursuite et
extension des travaux (à ce jour ces travaux sont couverts par le secret
qui s'attache à tous les travaux d'instruction des juridictions financières).



«
L'observation,
les connaissances,
la déduction
sont les trois qualités indispensables
pour le détective.
C'est de l'alliance de ces trois facultés
complexes
que naît son expertise.
»

Présentation personnelle & introduction



- Depuis 2012, 21 agents**
- Cartographie / Géomatique
 - Gestion de bases de données
 - Études et analyses
 - Veilles
 - Design de service public

observatoire.loire-atlantique.fr

🏠 Les cartes Les statistiques Les études et veilles Les vues aériennes Les données ouvertes



Avril 2018 : thèse
sous la direction de Messieurs
François-Charles Wolff & Florent Sari

Présentation personnelle & introduction



Programme départemental d'insertion



Vers une stratégie départementale d'insertion
Synthèse de la concertation
Jan-décembre 2017

La mise en place d'un observatoire des parcours des allocataires du RSA

I. Analyse descriptive

→ **objectiver** la situation des allocataires du RSA en Loire-Atlantique

II. Typologies d'individus



III. Parcours

→ **faire ressortir** les différents parcours d'allocataires

IV. Analyse des trajectoires

→ **étudier** les facteurs qui influencent les trajectoires

Inédit en France



Un modèle de projection à un an du nombre de ménages allocataires du RSA et des montants à allouer



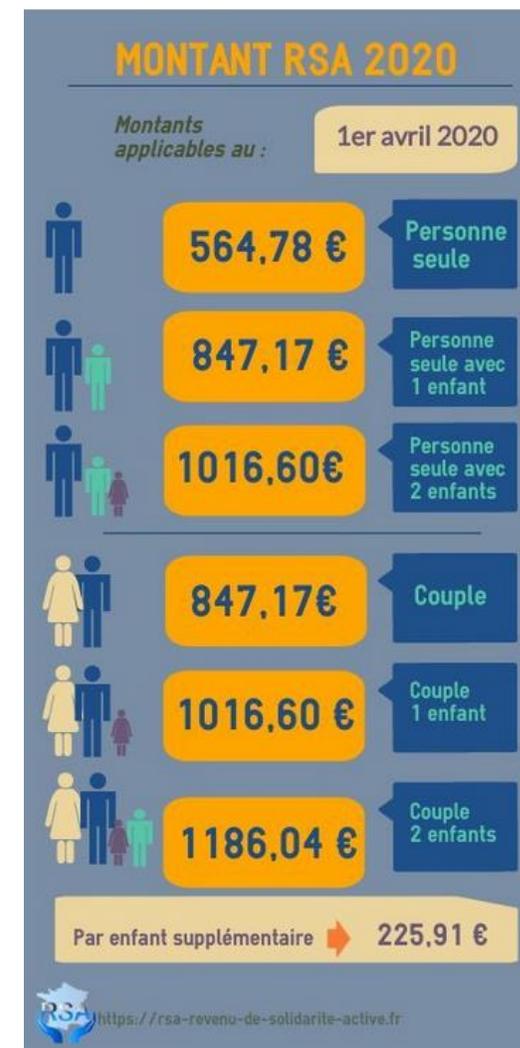
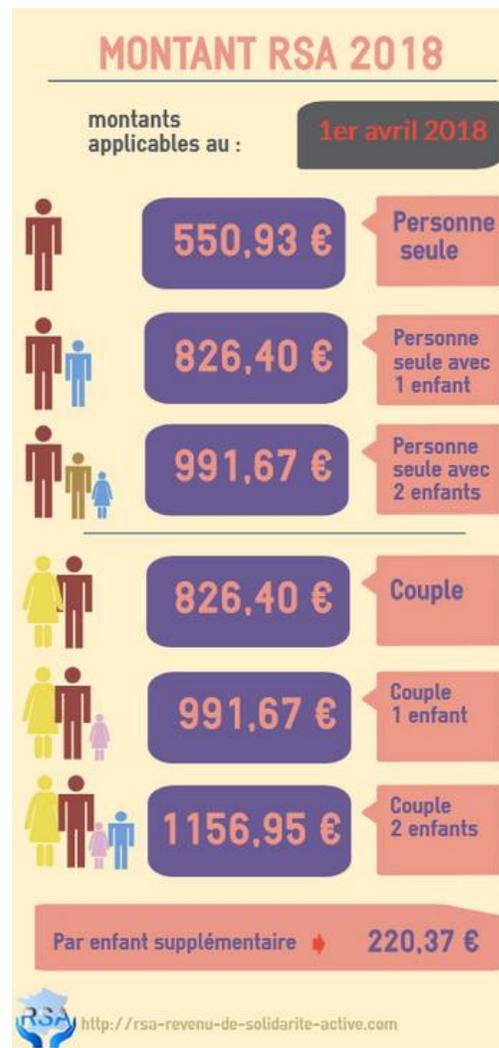
Inédit en France

Une plateforme d'accès aux données localisées sur les ménages allocataires du RSA (données ouvertes)

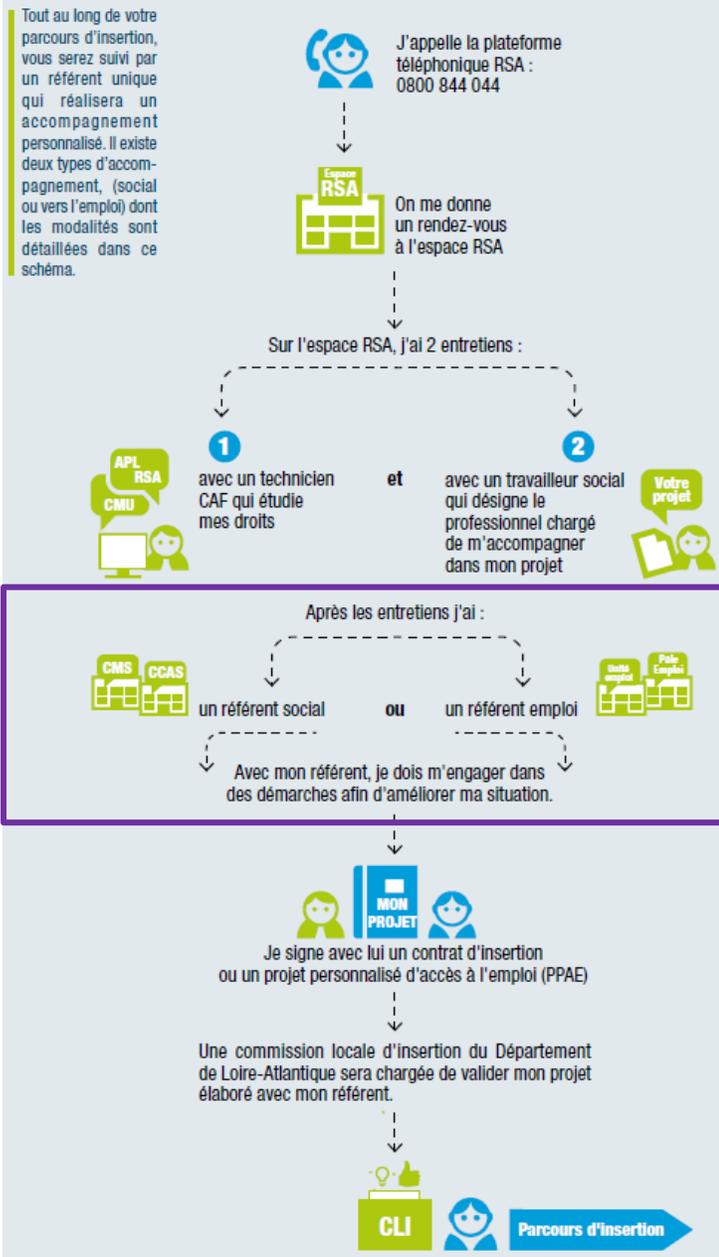
Présentation personnelle & introduction



Une allocation différentielle dont le montant est fonction de la composition du ménage



Présentation personnelle & introduction



Un service du Département
RSA
 Connaître vos droits et vos obligations

■ Pour toutes vos questions sur le RSA :
 0800 844 044
 (appel gratuit depuis un poste fixe)



La contractualisation

- Une obligation
- Cela marque l'engagement dans une « démarche d'insertion » par la contractualisation des objectifs par le biais d'un contrat d'insertion

OBJECTIF(S) FIXE(S) DANS LE CADRE DU CONTRAT D'INSERTION Rang du contrat

Objectif(s) du contrat :

<input type="checkbox"/> Accès /Maintien des droits	<input type="checkbox"/> Gestion du budget	<input type="checkbox"/> Maintien dans l'emploi
<input type="checkbox"/> Logement	<input type="checkbox"/> Aide à la parentalité	<input type="checkbox"/> Diagnostic professionnel
<input type="checkbox"/> Démarches de soins	<input type="checkbox"/> Développement du lien social	<input type="checkbox"/> Élaboration de projet professionnel
<input type="checkbox"/> Développement de l'activité professionnelle (Travailleurs indépendants)	<input type="checkbox"/> Accès à l'emploi	

Action(s) envisagée(s) :

Présentation personnelle & introduction

CONVENTION 15/12-01

AVENANT N°1 A LA CONVENTION DE GESTION
DU REVENU DE SOLIDARITE ACTIVE

CNIL
COMMISSION NATIONALE
INFORMATIQUE & LIBERTES



Convention de gestion du Revenu de Solidarité Active
entre la Caisse d'Allocations Familiales de Loire-Atlantique
et le Département de Loire-Atlantique

Décret n° 2019-341 du 19 avril 2019 relatif à la mise en œuvre de traitements comportant l'usage du numéro d'inscription au
répertoire national d'identification des personnes physiques ou nécessitant la consultation de ce répertoire

Des données reçues annuellement

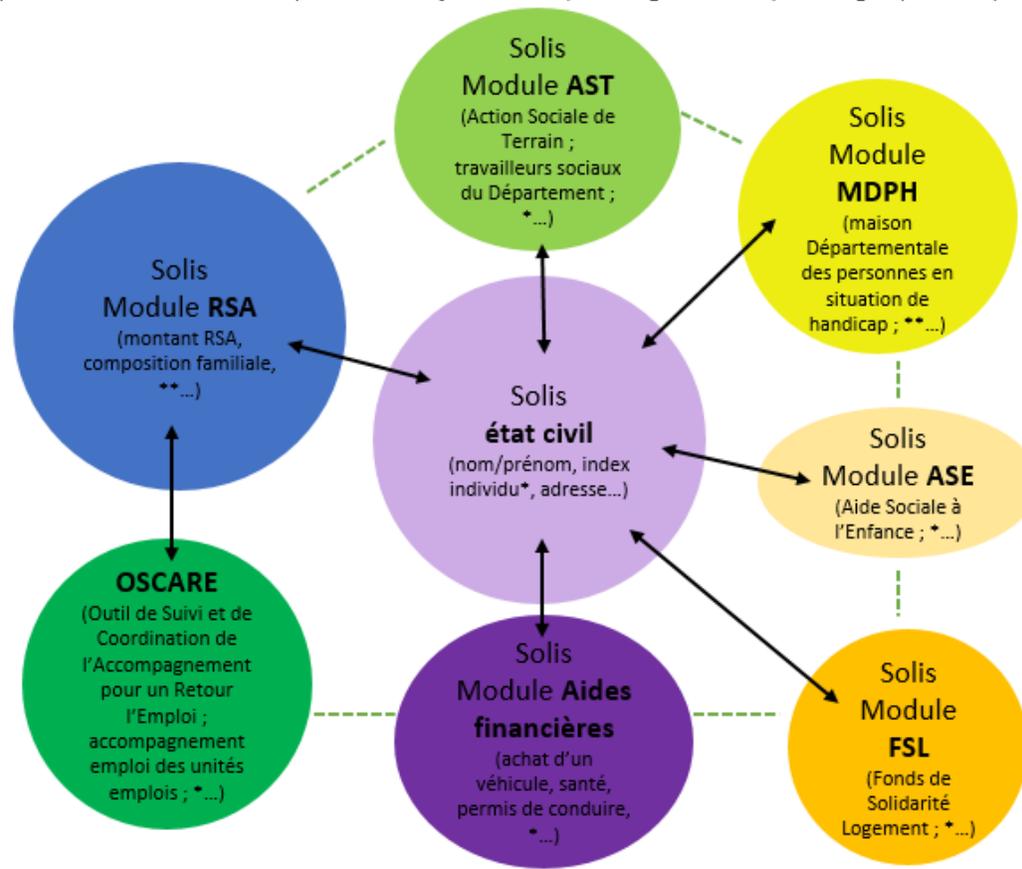
(3 fichiers/an)

Fichiers CAF / MSA

(composition familiale, montant de RSA,
index individu*, NIR**...)

Des données de flux (quotidien) requêtées mensuellement

(données sur un mois donné et/ou au dernier jour du mois ; une vingtaine de requêtes regroupées en 4)



Liens existants

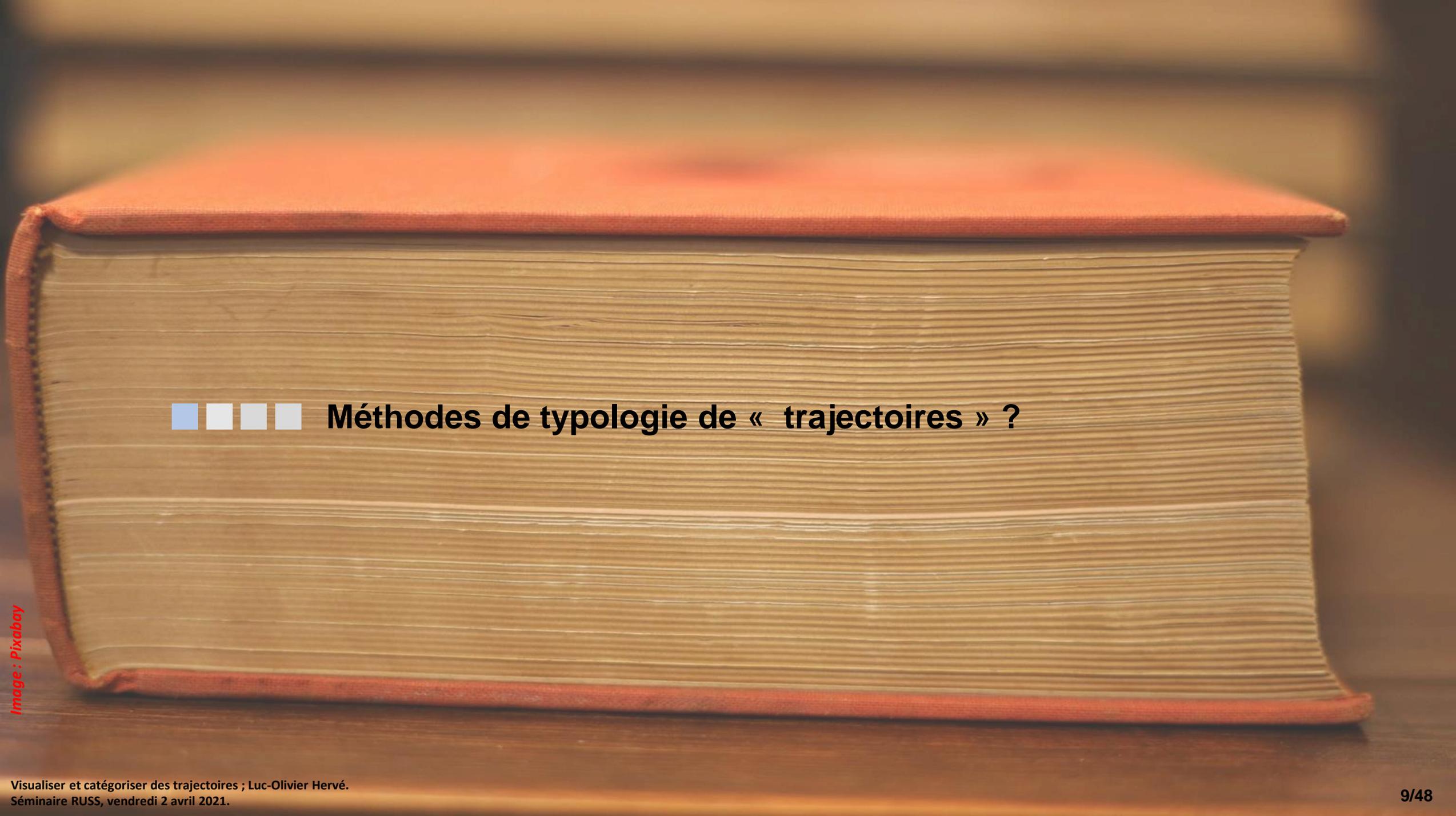
Liens à créer

(index individu*, NIR**)



Un entrepôt unique et inédit
dans la collectivité

- Une trentaine de requêtes imbriquées dans 4 distinctes aux périmètres différents
- Un apurement, croisement de différents modules automatisé et consolidé
- Un filtre sur les allocataires du RSA présents dans d'autres modules / dispositif d'aides financières et/ou d'accompagnement liés au Département
- Knime : un outil gratuit, puissant
- RGPD, secret statistique...
- Possibilité de créer des données longitudinales/biographiques ex-post !



■ ■ ■ ■ Méthodes de typologie de « trajectoires » ?

Méthodes de typologie de « trajectoires » ?

LAROUSSE

Formé de **type*** (« ensemble des traits généraux caractérisant un genre d'êtres ou de choses ») avec le suffixe **-logie*** (« parole, discours »).

Source : *Centre Nationale des Ressources Textuelles et Lexicales*



Le mot « typologie » apparaît au XIX^e siècle pour désigner les types ou les classifications des sciences de la nature, et différencier leur mode d'élaboration « scientifique » des classifications communes ou des types des philosophes et des exégètes bibliques. Il implique, comme en [philosophie](#), une explicitation raisonnée des principes « essentiels » des êtres et des choses mais aussi des méthodes expérimentales validant des hypothèses.

La sociologie adapte ce mode d'élaboration pour expliquer et comprendre les sociétés humaines. [Auguste Comte](#) la place au premier rang de sa typologie des sciences : toute pensée et toute organisation sont une construction historique et sociale. L'ambition est grande : elle soutient d'abord la mise en œuvre de macrotypologies de l'évolution des sociétés qui différencient la sociologie d'autres sciences sociales. Les typologies limiteront ensuite leurs objets tout en approfondissant leurs méthodes et en prenant distance avec le modèle « positiviste » des sciences de la nature.

Les macrotypologies des pères fondateurs posent des principes durables. [Émile Durkheim](#) en énonce dans les *Règles de la méthode sociologique* : expliquer le social par le social ; chercher les causes et les fonctions dans l'histoire ; les classer en privilégiant leurs manifestations les plus visibles et durables ; les relier logiquement entre elles (Max Weber dira « comprendre », on parlera de [causalité](#) structurale). Pour Durkheim, « la sociologie est comparative » : elle compare des hypothèses et leurs effets attendus aux faits sociaux observables, compare les variations des types sociaux ainsi définis. [Karl Marx](#) et [Max Weber](#) posent des principes analogues avec d'autres accentuations. La « domination » du mode de production est fortement affirmée par Marx ; pour Webe [...]

Système de classification des individus en types physiques et/ou psychologiques où, le plus souvent, des correspondances sont établies entre des types physiques et des types psychologiques, les premiers étant supposés prédéterminer les seconds.

Classification des langues qui s'appuie sur la comparaison de leurs caractéristiques internes que l'analyse synchronique permet de dégager. (Elle s'oppose à la classification génétique [familles de langues] ou à des classifications qui font appel à des critères extralinguistiques, comme la proximité géographique [langues amérindiennes, par exemple] ou le degré d'évolution des sociétés.)

Système de classification d'un ensemble de données empiriques concernant un phénomène social en types distincts.

Département de Loire-Atlantique

observatoire.loire-atlantique.fr

Rechercher dans le site

Les cartes

Les statistiques

Les études

Les vues aériennes

Les données ouvertes

Accueil > Les études > Action sociale de proximité et insertion > Vidéocassettes RSA

Vidéocassettes RSA

Source : Luc-Olivier Hervé, Léa Jouan et Isabelle Thoumin-Leprince (Département, service innovation)
Actualisé le : 11/06/2019

Méthodes de typologie de « trajectoires » ?

- Différentes **méthodes/algorithmes** : naïve bayésienne ; CAH ; ACM ; k-means ; PAM ; CLARA ; arbre de décision ; forêt aléatoire ; SVM etc.
- Des méthodes dites : simples ; hiérarchiques ou non ; mixtes ; supervisé (utilisation d'un échantillon d'apprentissage) ou non
- De nombreux outils disponibles : excel ; xlstat ; R ; Stata ; Sas ; Sphinx etc.



Article de recherche

Analyse comparative des méthodes de classifications

L'exemple du bien-être au travail

Jordane Creusier, Franck Biétry

DANS RIMHE : REVUE INTERDISCIPLINAIRE MANAGEMENT, HOMME &

ENTREPRISE 2014/1 (n° 10), PAGES 105 À 123

L'Usine ML, une application web de Machine Learning automatique écrite en R et Shiny

July 23, 2018

datascience R shiny

Depuis un simple navigateur web, en quelques clics et zéro code, menez un projet de Machine Learning (modélisation prédictive, classification ou régression) sur des données en 3 étapes simples :

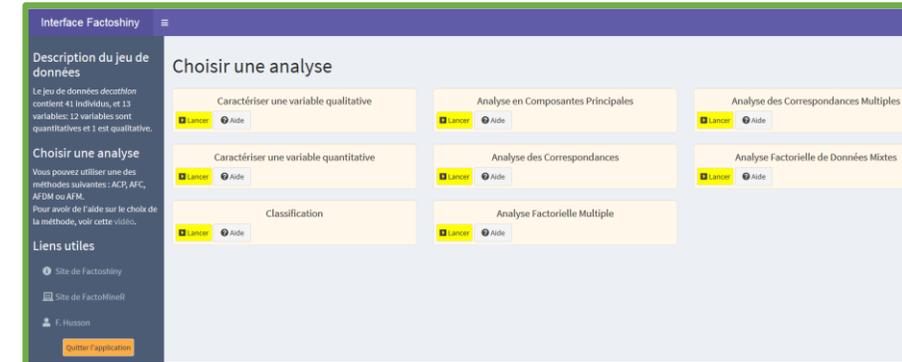
1. **D'abord, les données** ! - Choix du jeu de données (5 datasets démo dispo pour le moment), de la variable à prédire et des variables explicatives à utiliser pour la modélisation. On choisit également la taille du jeu de test (en % du total).
2. **Entraînement & choix du modèle** - En un clic, l'app entraîne et optimise automatiquement jusqu'à 56 modèles différents (SVMLinéaire, Neural Net, Random Forest, kNN, SVMPoly, Naive Bayes) et choisit le meilleur en termes de performance. Plusieurs métriques et graphiques sont affichés et permettent de comparer les modèles.
3. **Performance de modèle** - Caractéristiques du meilleur modèle.

Clustering

François Husson - 1/6



▶		Clustering (part 1/4: introduction) François Husson 21:49
2		Clustering (part 2/4: example and how to choose the number of clusters) François Husson 9:16
3		Clustering (part 3/4: the partitioning method K-means) François Husson 11:33
4		Clustering (part 4/4: characterizing clusters) François Husson 19:07
5		Clustering with FactoMineR François Husson 18:30
6		Clustering with R (FactoMineR & Factoshiny) François Husson 17:02



Appli R Shiny : ACP ; k-means

Méthodes de typologie de « trajectoires » ?

Analyser des trajectoires : de la comparaison de photographies à différentes dates à l'identification des dynamiques au fil du temps.

Comparaison de situations à deux dates

Jeu des différences (ici 2)



Des méthodes variées et riches mais des « comparaisons statiques »

évolution ; différences ; doubles différences etc.
modèle (logit...)



Analyses dynamiques

Modèles économétriques :

- modèle de durée,
fonctions :
- de distribution,
 - de survie,
 - de densité,
 - de hasard & de hasard intégré...

Analyses multivariées
avec données modifiée
pour intégrer la
dimension temporelle

Carte de Kohonen

Analyse de
séquences par
appariement optimal

Méthodes de typologie de « trajectoires » ?

Modèles économétriques :

modèle de durée,

fonctions :

- de distribution,
- de survie,
- de densité,
- de hasard & de hasard intégré...

Exemple :

La modélisation économétrique des transitions individuelles sur le marché du travail

Jean-Pierre Florens^(*)

Denis Fougère^(**)

Thierry Kamionka^(***)

Michel Mouchart^(****)

Analyses multivariées :

AFC, ACP

=> recodage disjonctif

=> analyse

harmonique

qualitative (AHQ)

Carte de Kohonen

(apprentissage

non-supervisé, itératif,

SOM)

Analyse de séquences

par appariement

optimal (OMA)

Méthodes de typologie de « trajectoires » ?

Visualiser et catégoriser des trajectoires : analyses multivariées

Analyses multivariées :
AFC, ACP
=> recodage disjonctif
=> analyse harmonique qualitative

- **Recodage disjonctif** puis analyses multivariées (AFC, ACP)
- **Analyse harmonique qualitative (AHQ)**
=> mettre en avant sur les états dans lesquels l'individu reste plus longtemps
=> repose sur la matrice harmonique
+ : interprétation simple
+/- : classification que « les caractéristiques les plus structurantes » (analyse factorielle)
+/- : découpage de la période d'étude en intervalles (amplitude possiblement différentes)
=> insister sur période apparition événement & durée des états
- : perte de détails de séquences
- : contrôle à posteriori des facteurs non utilisés

+ loin



Ex : recodage disjonctif, AHQ

Analyse Harmonique Qualitative : une application à la comparaison des trajectoires résidentielles et géographiques de Parisiens

Elisabeth MORAND(*), Bénédicte GARNIER(*), Catherine BONVALET(*)
(*):INED, 133 boulevard Davout 75020

Visualisation graphique agrégée des trajectoires individuelles : revue de l'existant et application en géographie

Hadrien Commenges, Pierre Pistre

Les trajectoires spatiales d'activité des couples

The spatial trajectories of couples' activities

Eva Lelièvre et Nicolas Robette

qha: Qualitative Harmonic Analysis

2.2. Les étapes d'exécution de l'AHQ jusqu'à la classification des trajectoires

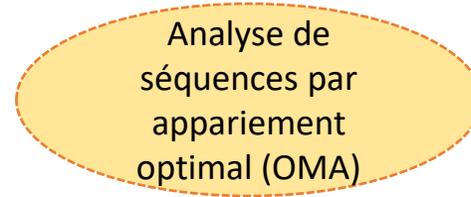
Tableau 1. Typologies en 6 classes des carrières des hommes franciliens des générations 1930 à 1950

n°	Classe (caractéristique principale)	AHQ (%)	MAO (%)
1	Professions intermédiaires	29	27
2	Ouvriers	26	26
3	Cadres	26	26
4	Employés	13	9
5	Vers artisans, commerçants et chefs d'entreprise	5	5
6	De professions intermédiaires à cadres	-	6
7	D'agriculteurs à ouvriers	1	-
	Total	100	100

Champ : 1 341 carrières professionnelles d'hommes des générations 1930-1950 résidant en Ile-de-France à la date de l'enquête.
Source : enquête *Biographies et entourage* (Ined, 2001).

Méthodes de typologie de « trajectoires » ?

Visualiser et catégoriser des trajectoires : Kohonen ; OMA



Points communs :

Définir une distance :

- Kohonen : tableau disjonctif-complet => proximité/dissimilarité entre individus fonction cooccurrences (situations proches : nombre individus avec code 1 pour une situation ; distance profils-colonnes en ACM ; matrice carrée des probabilités des cooccurrences)
- OMA : matrice de distance => choix méthode mesure de distance (OM ; HaM ; DHD...) ; nombre opérations équilibrer séquences & coûts associés aux opérations

Choix de classes : centres mobiles & CAH (critère ward) par ex

• Carte auto-ordonnées (Kohonen)

- + : informations équivalentes celles d'une ACP
- +/- : visualiser et interpréter les proximités de trajectoires (différent de la CAH par ex pour l'OMA)
- +/- : tester un nombre important de classes
- +/- : nombre de classes à retenir (choix sur une logique de proximité visuelle)

• Méthode appariement optimal (OMA) de séquences

- => rapprocher les trajectoires sur des transitions et repérer les mobilités
- => repose sur la matrice de coûts
- + : utilisation de l'ensemble des détails sur la séquences
- + : fixer les coûts (*indel*)
- => pour favoriser type de transition // moment de l'apparition d'un événement
- + : mettre davantage en avant les dynamiques : ruptures/mobilités/transitions entre états

Algorithme de Kohonen :
classification et
analyse exploratoire des données

Marie Cottrell et Patrick Letremy



La méthode

Cartes auto-organisées pour l'analyse exploratoire de données et la visualisation

Marie Cottrell, Patrick Letrémy, Patrick Rousset, Smail Ibbou

Application et comparaisons
(AHQ, OMA)



Les parcours d'insertion des jeunes : une analyse longitudinale basée sur les cartes de Kohonen

Patrick Rousset
Jean-François Cuvier
Yvette Grégoire



+ loin



Ex : cartes de Kohonen, OMA, AHQ

Comment dresser des typologies de trajectoires ?

**HORIZONS
PUBLICS**

À titre de comparaison, si la mise en place d'une typologie de ménages allocataires a consisté à élaborer des profils types de ménages au regard d'un ensemble de caractéristiques individuelles (logement, niveau de vie, composition des ménages par exemple), l'analyse de séquences vise à identifier des allocataires ayant connu, sur une période de quatre années entre 2013 et 2016, des enchaînements de situations suffisamment proches, voire identiques, pour constituer un groupe de trajectoire.

Loire-Atlantique : des outils innovants au service des allocataires du RSA (janvier 2019)

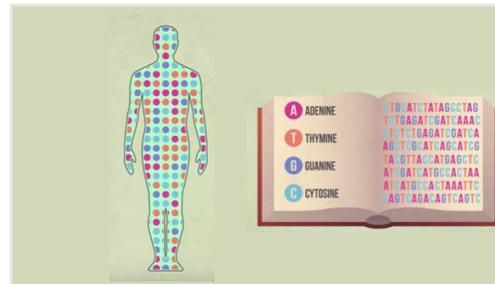
➤ ORIGINE ET DIFFUSION

- Une méthode issue du domaine médicale...
- ... appliquée aux sciences sociales

+ loin



**OMA & analyse de séquences :
de la science à la science sociale**



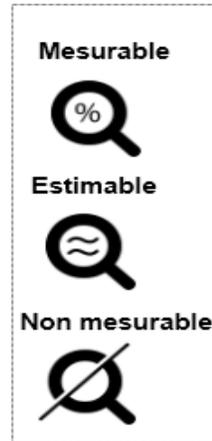
+ loin



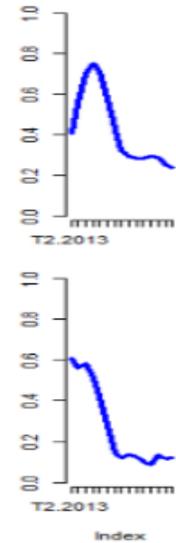
**OMA & analyse de séquences :
plusieurs outils dédiés**

Analyse de séquences : présentation & illustration

Des outils graphiques visuels et puissants



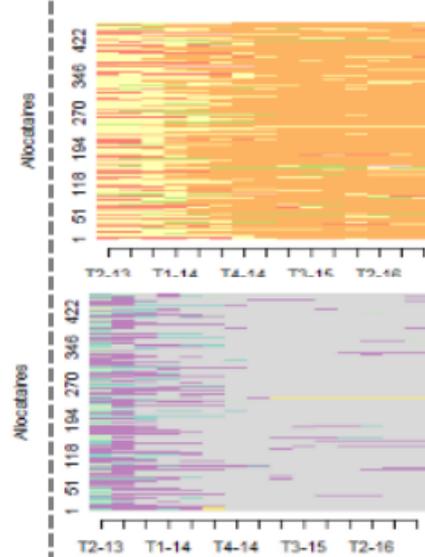
Entropie



Vers une homogénéisation des situations individuelles ?

- Homogénéisation ou hétérogénéité des situations
- Successions de situations
- Part d'individus dans chaque situation
- Durée dans chaque situation

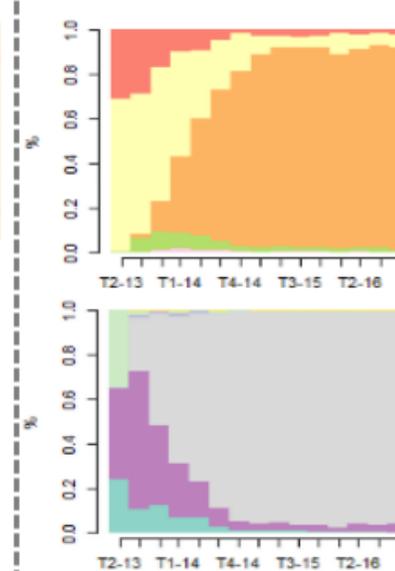
Tapis individuels



Quels parcours individuels des allocataires ?

-
-
-
-

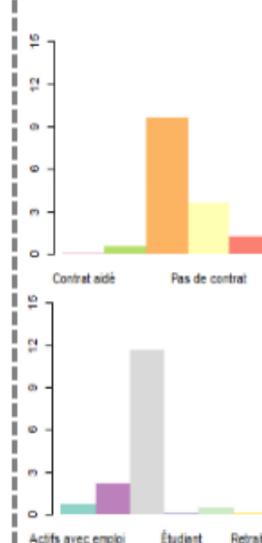
Chronogrammes



Quelles trajectoires collectives des allocataires ?

-
-
-
-

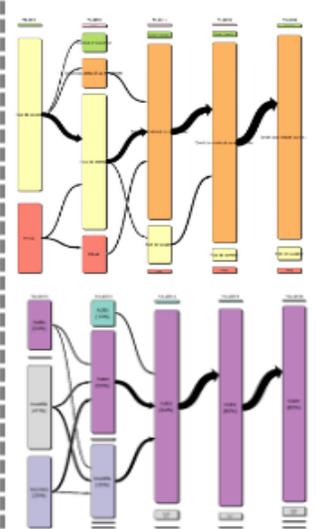
Durée moyenne



Quelles durées moyennes dans chaque situation par trajectoire ?

-
-
-
-

Transitions

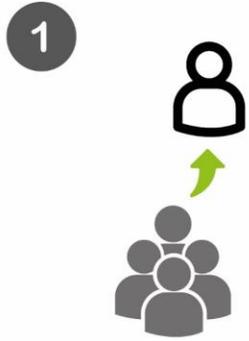


Quels enchaînements de situations successives par trajectoire ?

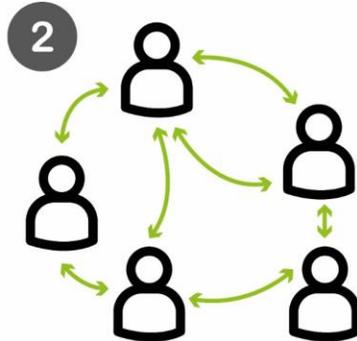
-
-
-
-

Analyse de séquences : présentation & illustration

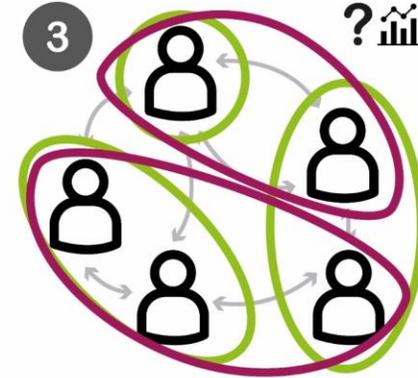
➤ Analyse de séquences : une méthode en 6 étapes



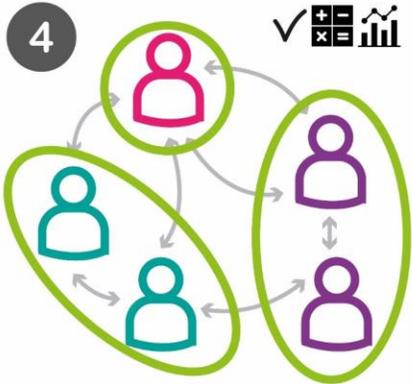
Suivre une sous-population représentative



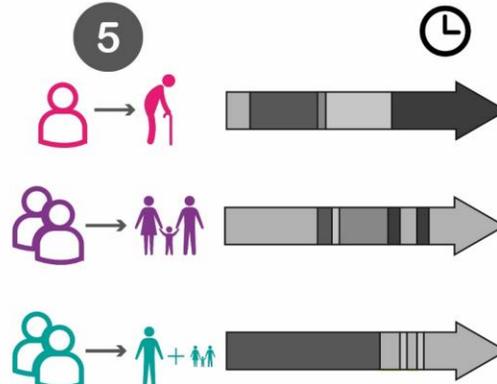
Déterminer la proximité des enchaînements de situations (trajectoires)



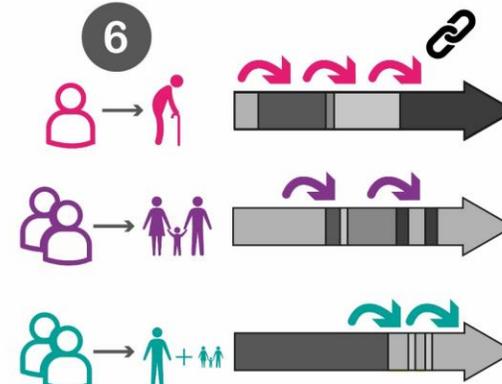
Tester des regroupements pertinents de trajectoires
2 groupes ? 3 groupes ?



Valider les regroupements pertinents (groupes)



Décrire les caractéristiques des individus et calculer les situations et leurs durées par groupe



Comprendre les enchaînements des situations pour chacun des groupes

Résultats

1



Suivre une sous-population
représentative



Choix de la cohorte ?

- **Stocks** : considérer les unités d'observation déjà présentes à la 1^{ère} date de début de période

+ : **davantage d'observation** (individus) qu'en flux

- : **sur-représentation des trajectoires « peu mobiles »**

- **Flux** : considérer les unités d'observation entrées à la 1^{ère} date de début de période

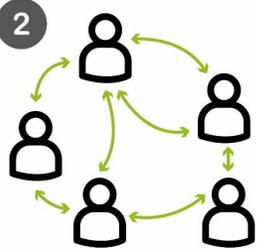
+ : **sur-représentation des « trajectoires mobiles »**

+ : **simplicité à calculer des durées dans les différents états**

- : **moins d'observation (individus) qu'en stock**

+/- : **choix usuellement retenu dans la littérature**

Analyse de séquences : présentation & illustration



Proximités/dissimilarités entre individus ?

- Proximités/dissimilarités : enchaînements des situations pour chaque individus durant l'ensemble de la période
- Calcul de la matrice de distance basée sur les coûts

Déterminer la proximité des enchaînements de situations (trajectoires)



La proximité/dissimilarité fondée sur le nombre de transformation nécessaires pour égaliser 2 séquences.

Calculs sur la base de 2 opérations : substitution ; *indel*

Calculs fonction : des situations vécues à chaque date ; des enchaînements de situations ; des durées dans chacune des situations ; de l'emplacement des enchaînements de situation

Détermination des coûts ?

- 2 types d'opération (chacune ayant un coût) : *indel* (insertion-suppression) ; substitution
- Comment appliquer les différents coûts ?
- Plusieurs choix de coût possible : monétaire ; distances : euclidienne, Hamming, Gauthier, Holister etc.
 - Coût retenu : probabilité observée du passage d'une situation à une autre

Illustration du calcul de distance

- Comment modifier l'enchaînement des situations vécues par un individu pour l'égaliser avec celles d'un autre individu ?
Distance = nombre* de modifications nécessaires
- Autrement dit :
peu (beaucoup) de modifications = **proximité** (dissimilarité)
- **Nb*** : cela est vrai si et seulement il n'y pas de préférences, probabilités différentes... à changer une situation A par une situation B ou C
=> **sinon définir des coûts de passage d'une situation à une autre !**

Ex : une matrice de distance entre observations

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	0.00000	26.42170	42.012342	31.69349	33.329753	37.473688	32.882023	35.399130	35.74536	21.36882
2	26.42170	0.00000	32.806911	27.40776	31.692942	33.468789	29.309497	32.995448	34.42220	38.98811
3	42.01234	32.80691	0.000000	49.14747	47.508144	47.070001	27.633647	49.600000	13.98646	44.64874
4	31.69349	27.40776	49.147469	0.00000	19.649617	17.757151	26.648023	14.718690	39.32385	36.10572
5	33.32975	31.69294	47.508144	19.64962	0.000000	17.336450	22.135910	6.929154	33.52168	30.77937

Lecture : 1 est plus proche de 2 (26 de distance) que de 4 (32) ou 3 (42)

Tableau 1 – Signification des deux opérations de base des Méthodes d'Appariement Optimal

	Insertion-Suppression	Substitution
Ce qui est préservé	Événements	Temps
Ce qui est simplifié	Temps	Événements

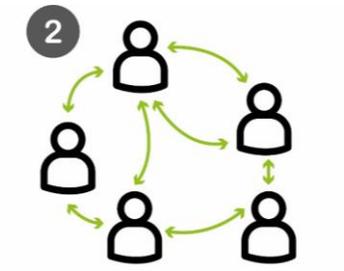
Un choix classique de coût de un pour l'insertion-suppression et de deux pour la substitution tend à réduire le temps à la seule dimension du classement (Kruskal 1983, Lesnard et Saint Pol 2006). Il est possible d'éviter les déformations du temps en favorisant les opérations de substitutions par rapport aux opérations d'insertions suppressions. Les coûts de substitutions doivent alors être faibles. On peut aussi proposer des coûts de substitutions différents pour chaque état à l'aide des taux de transitions observés sur les données.

Analyse de séquences : présentation & illustration

- Cas 1/ Coût insertion-suppression (*indel* = 1) < coûts de substitution (2) => ce sont les « coûts classiques »

	1	2	3	4	5	
Ex1	i	J	J	J	JP	JP
	j		P	P	JP	JP
Ex2	i	J	J	J	JP	JP
	j	P	P	JP	JP	JPM

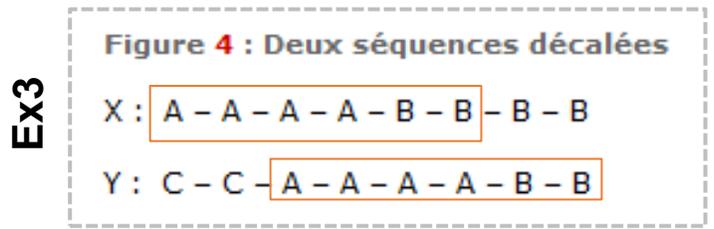
Ex1: 1 insertion, 2 substitutions, 1 deletion, Coût total = 6
 Ex2: 4 substitutions, Coût total = 8



Déterminer la proximité des enchaînements de situations (trajectoires)

- Cas 2/ Coût insertion-suppression (*indel*) = coûts de substitution
- Cas 3/ Coût insertion-suppression > coûts de substitution

À RETENIR!



OU

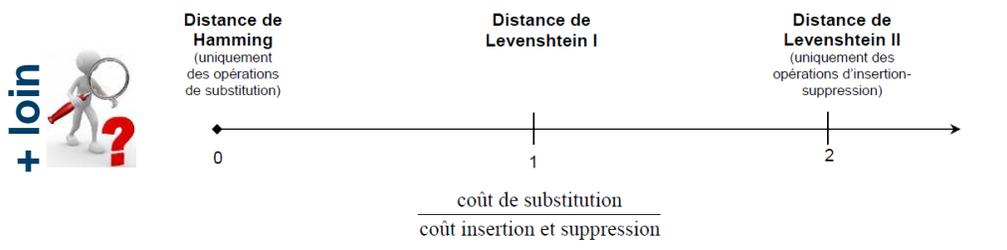
- Cas 4/ probabilité observée de passage d'une situation à l'autre => méthode usuellement retenue en science sociale (dont les travaux sur le RSA)

```

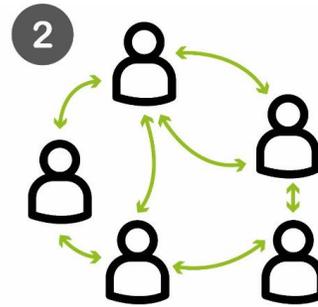
seq.contrat <- seqdef(contrat[,1:15])
couts.sc <- seqsubm(seq.contrat , method="TRATE", cval=2)

seq.emploi <- seqdef(emploi[,1:15])
couts.de <- seqsubm(seq.emploi ,method="TRATE", cval=2)
  
```

Figure 4 – Effet des coûts sur le type de régularité statistique privilégié



+ loin ? Définir des coûts ?

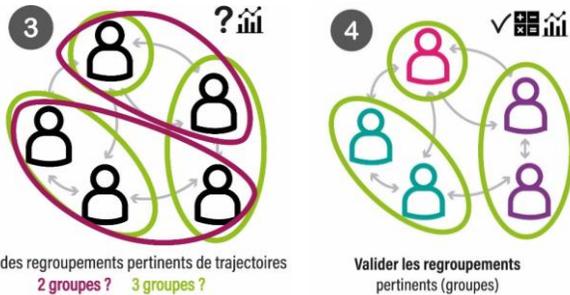


Déterminer la **proximité** des enchaînements de situations (trajectoires)

Matrice de distance & limite (*TraMineR* & *ViCaTraj*) :

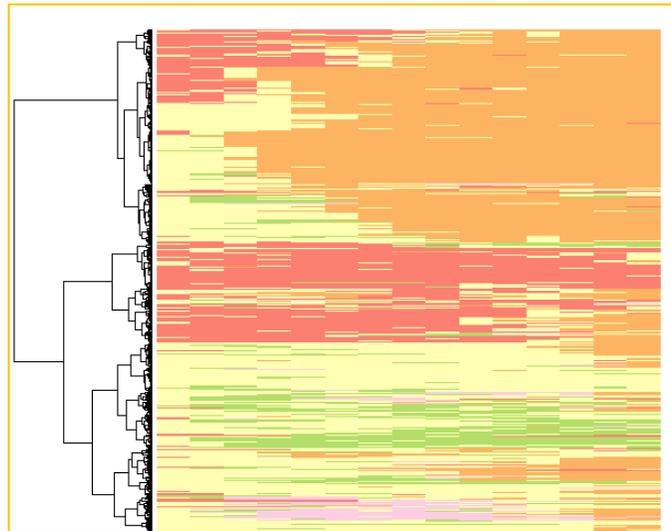
- Maximum 46 340 séquences uniques
- Des solutions existent :
=> traiter uniquement les séquences uniques :
<http://mephisto.unige.ch/weightedcluster/>
=> échantillon aléatoire => *OMA* => *medoids* des clusters => distance aux *medoids* :
<https://stackoverflow.com/questions/62888071/too-many-unique-sequences>

Analyse de séquences : présentation & illustration

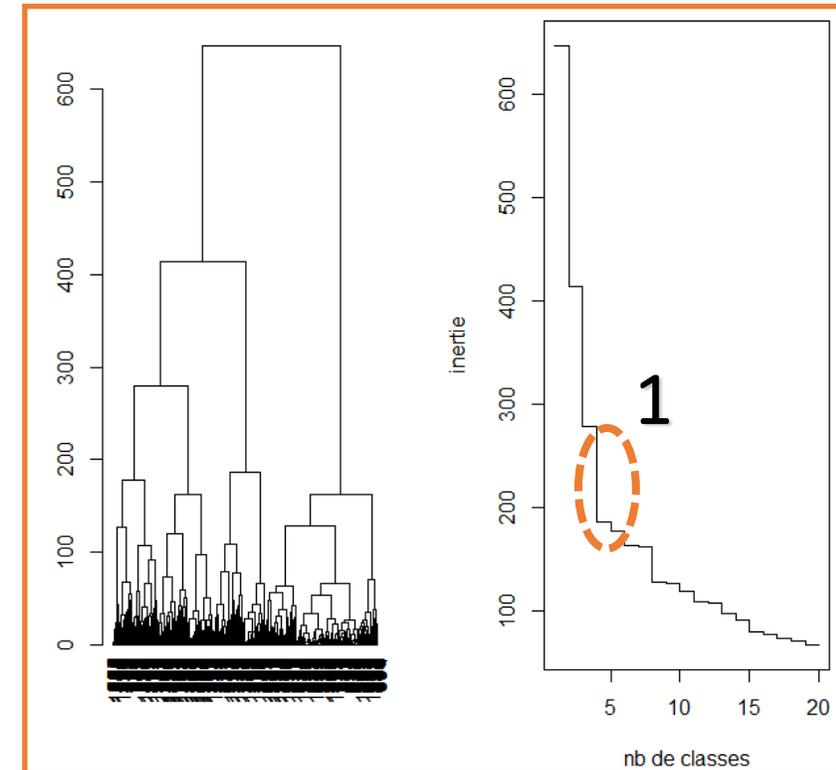


```
seq.agnes <- agnes(as.dist(multich.dist), method="ward", keep.diss=FALSE)
win.graph()
par(mfrow=c(1,2))
plot(as.dendrogram(seq.agnes))
plot(sort(seq.agnes$height, decreasing=TRUE)[1:20], type="s", xlab="nb de classes", ylab="inertie")
```

```
win.graph()
seq_heatmap(seq.contrat, seq.agnes, labcol="")
win.graph()
seq_heatmap(seq.emploi, seq.agnes, labcol="")
```

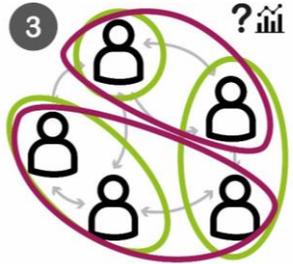


- ## 1 Test graphique 1 : cassure dans le graphique d'inertie
- Outil : dendrogramme et inertie
 - Décision : cassure dans la pente de la courbe d'inertie (= suggestion sur le nombre optimal de groupes à retenir)
 - Application RSA : **entre 4 et 6 groupes distincts**



Analyse de séquences : présentation & illustration

À RETENIR!



Tester des regroupements pertinents de trajectoires
2 groupes ? 3 groupes ?



Valider les regroupements pertinents (groupes)

1
✓

Test graphique 1 : cassure dans le graphique d'inertie

- Outil : dendrogramme et inertie
- Décision : cassure dans la pente de la courbe d'inertie (= suggestion sur le nombre optimal de groupes à retenir)
- Application RSA : **entre 4 et 6 groupes distincts**

2 Test graphique 2 : souhait d'expert, connaissance sujet

- Outil : tapis individuels et pouvoir discriminant de variables
- Décision : souhait d'expert, connaissance du sujet...
- Application RSA : **méthode non appliquée**

Le manuel de la librairie **WeightedCluster**
Un guide pratique pour la création de typologies de trajectoires
en sciences sociales avec R
3 Matthias Studer
Institute for Demographic and Life Course Studies
University of Geneva

3 Test numérique : indicateurs de qualité d'un regroupement

- Outil : 10 indicateurs à calculer
- Décision 1 : Minimisation 1 indicateur
choix pertinent = **minimisation** des indicateurs
pour un nombre de groupe donné pour l'**indicateur HC**
- Décision 2 : Maximisation 9 indicateurs
choix pertinent = **maximisation** des indicateurs
pour un nombre de groupe donné pour les **9 autres indicateurs**
(PBC, HG, HGSD, ASW, ASWw, CH, CHsq, R2, R2sq)

À RETENIR!

+ loin Qualité d'un regroupement ?

Analyse de séquences : présentation & illustration

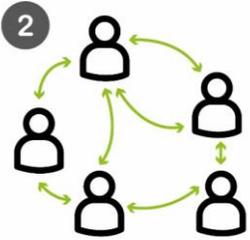
À RETENIR!



1
Suivre une sous-population représentative

Choix d'une cohorte

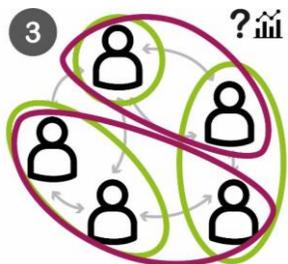
- **Stocks** (déjà présente à une date donnée) // **Flux** (entrée à une date donnée)
- + observations ; sur-représentation peu mobiles // - observations ; sur-représentation mobiles ; calcul de durée plus simples



2
Déterminer la proximité des enchaînements de situations (trajectoires)

Proximité/dissimilarité ?

- Définition des **coûts** (associé aux coûts : *indel* ; substitution) : *Trate*, *constant*, *futur*, *features*, *indel*, *indellog*
=> en science sociales (et sur le RSA) : probabilité observée (*Trate*) de passage d'un état à l'autre souvent retenu
- Définition de la **méthode pour calculer matrice de distance** : *OM*, *Omloc*, *HAM*, *DHD*
- **Calcul** de la matrice de distance



3
Tester des regroupements pertinents de trajectoires
2 groupes ? 3 groupes ?

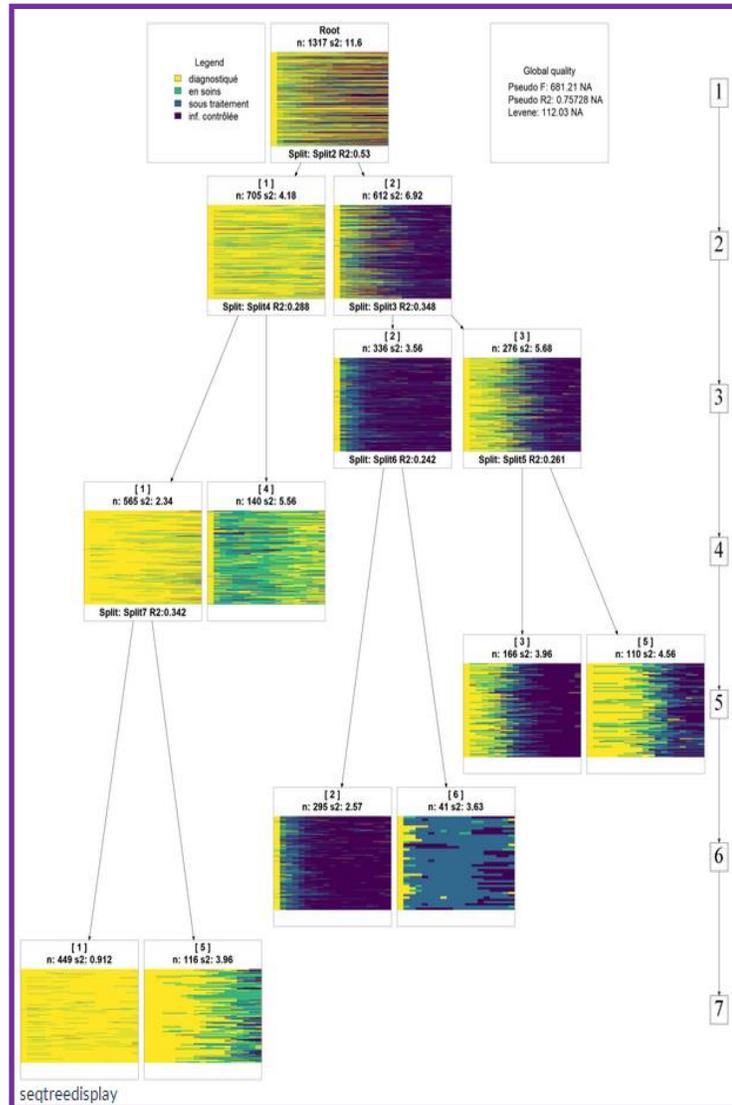
Classification ?



4
Valider les regroupements pertinents (groupes)

- Définition de la **méthode** : **CAH & PAM**, CAH, regroupement hiérarchique, FAST regroupement hiérarchique, PAM
- Définition des **paramètres de la méthode CAH** : **ward**, average, single, complete, weighted, flexible, gavage
- **Retenir** une classification et l'appliquer
Nb : aide possible avec des indicateurs de mesures de la qualité du regroupement (WeightedCluster)

Analyse de séquences : présentation & illustration



Démarche complémentaire :

- OMA : situations vécues => trajectoires => identifier caractéristiques...
//
- **Anova et séquences** pour croiser les deux directement

=> Souhait ajout dans *ViCaTraj*

Trajectoires RSA : peu de travaux sur les trajectoires des ménages allocataires

le **cnam**
ceet



99

Le recours à l'activité réduite :
Déterminants et trajectoires des demandeurs
d'emploi

Octobre
2016

Sabina Issehnane (coord.),
Fabrice Gilles,
Léonard Moulin,
Leïla Oumeddour,
Florent Sari

Rapport de recherche

Les trajectoires des bénéficiaires du RSA : enseignements d'une
classification

Élie Chosson ¹ [Détails](#)

1 CREG - Centre de recherche en économie de Grenoble



Politiques sociales et familiales

Les trajectoires de perception du RSA
Violaine Fernandez, Pauline Domingo



73

Les effets du RSA
sur le taux de retour à l'emploi
des bénéficiaires

Mars
2012

Élisabeth Danzin, Véronique Simonneau,
Danièle Trancart

Rapport de recherche

l'Observatoire
du Département

Le revenu de solidarité active
en Loire-Atlantique
-
Qui sont les allocataires du RSA en
Loire-Atlantique
et
quelles sont leurs trajectoires au sein
du dispositif ?

Février 2018

Loire
Atlantique

Analyse de séquences : présentation & illustration

❏ Épisode 3 : « Quelles trajectoires au sein du dispositif RSA ? »



Departement de Loire-Atlantique

observatoire.loire-atlantique.fr

Les cartes Les statistiques **Les études** Les vues aériennes Les données ouvertes

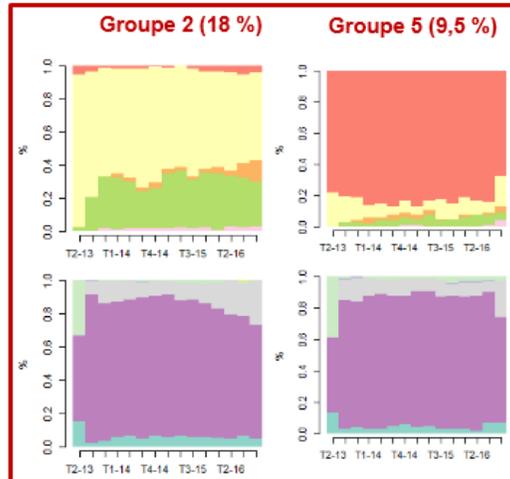
Accueil > Les études > Action sociale de proximité et insertion > Vidéocassettes RSA

 Vidéocassettes RSA

Source : Luc-Olivier Hervé, Léa Jouan et Isabelle Thoumin-Leprince (Département, service innovation)
Actualisé le : 11/06/2019

Analyse de séquences : présentation & illustration

Les trajectoires de retour à l'emploi des allocataires du Rsa de Loire-Atlantique, 2013-2016



Groupe 2 & 5 :

- Des sorties minoritaires fin 2016
- D'abord des jeunes
=> moins de 30 ans :
40 % // 25 % (département)

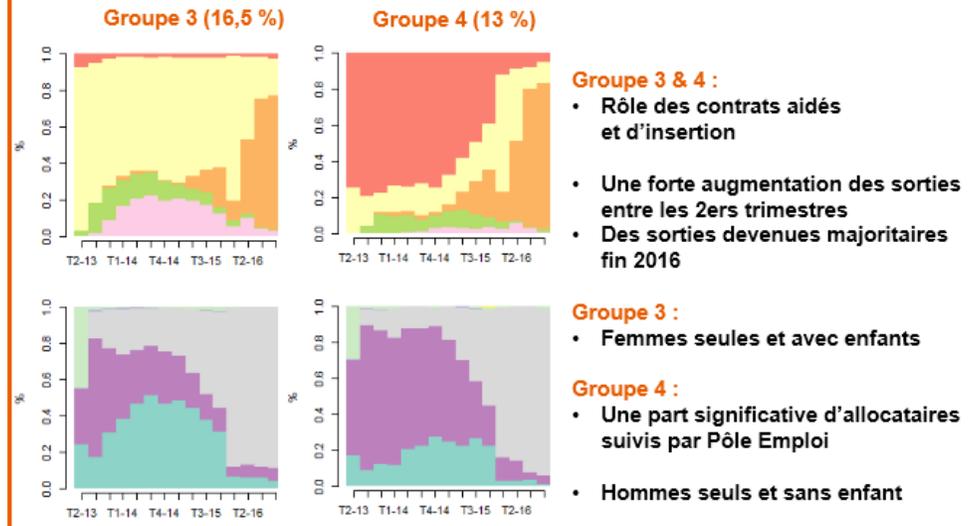
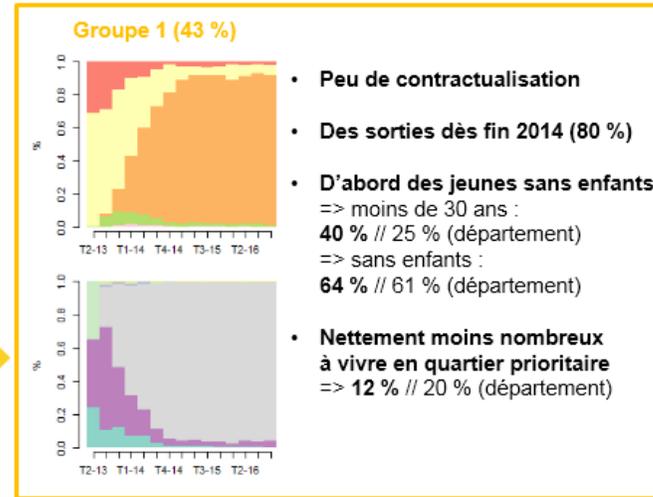
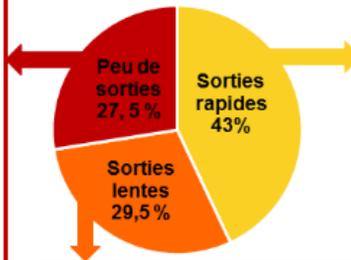
Groupe 2 :

- En majorité : Femmes seules avec enfants
- Plus de 50 ans : 14,5 % // 20 % (département)
- Une part significative vit en logement social
=> 43 % // 30 % (département)

Groupe 5 :

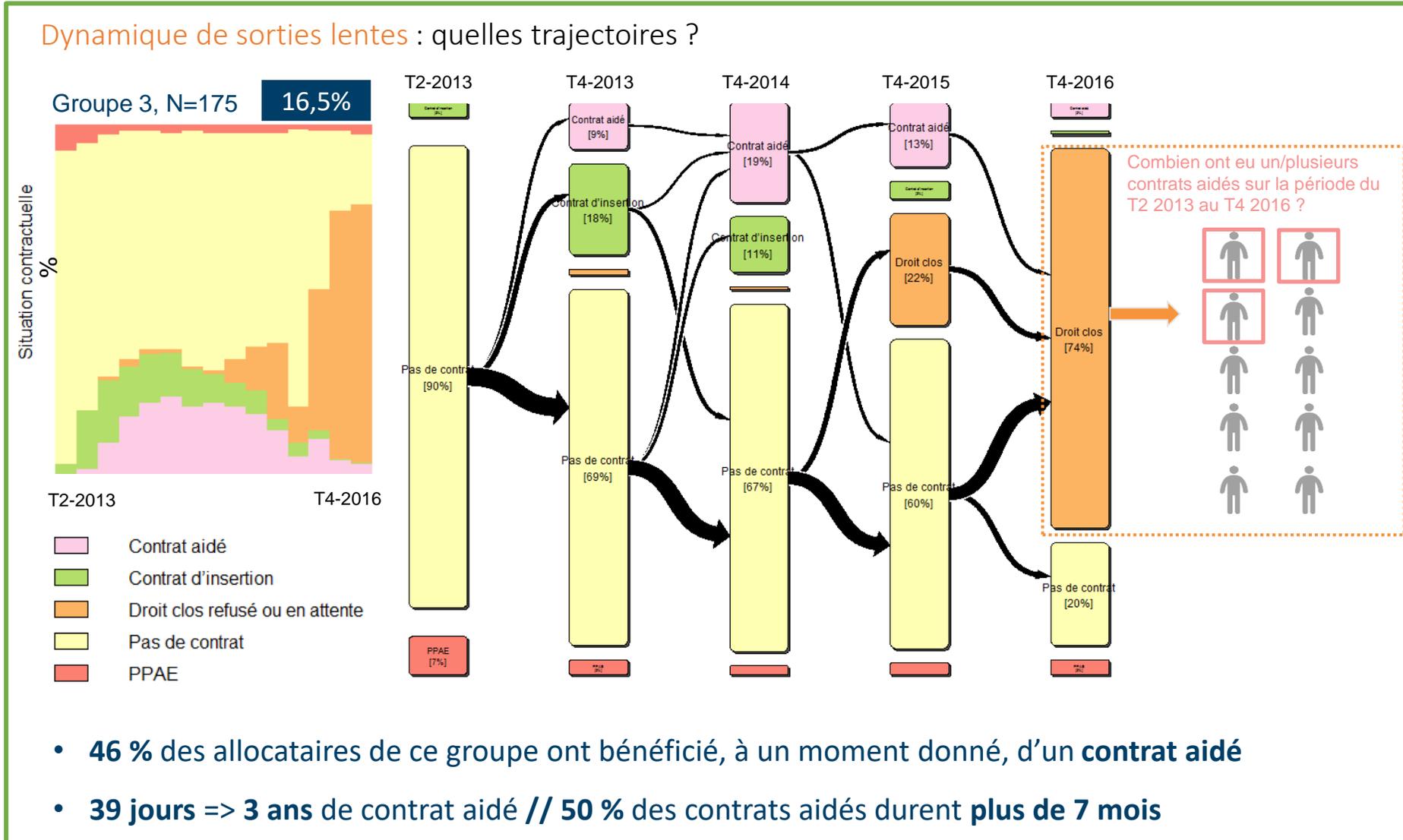
- Une part significative d'allocataires suivis par Pôle Emploi
- En majorité : Hommes seuls sans enfant
- Plus de 50 ans : 6 % // 20 % (département)
- Une part moindre vivant en logement social
=> 15 % // 30 % (département)

- ### Accompagnement
- Contrat aidé
 - Contrat d'insertion
 - Droit clos refusé ou en attente
 - Pas de contrat
 - PPAE
- ### Distance à l'emploi
- Actifs avec emploi
 - Actifs sans emploi ou inactifs
 - Autre
 - Étudiant
 - Inconnu
 - Retraité



Analyse de séquences : présentation & illustration

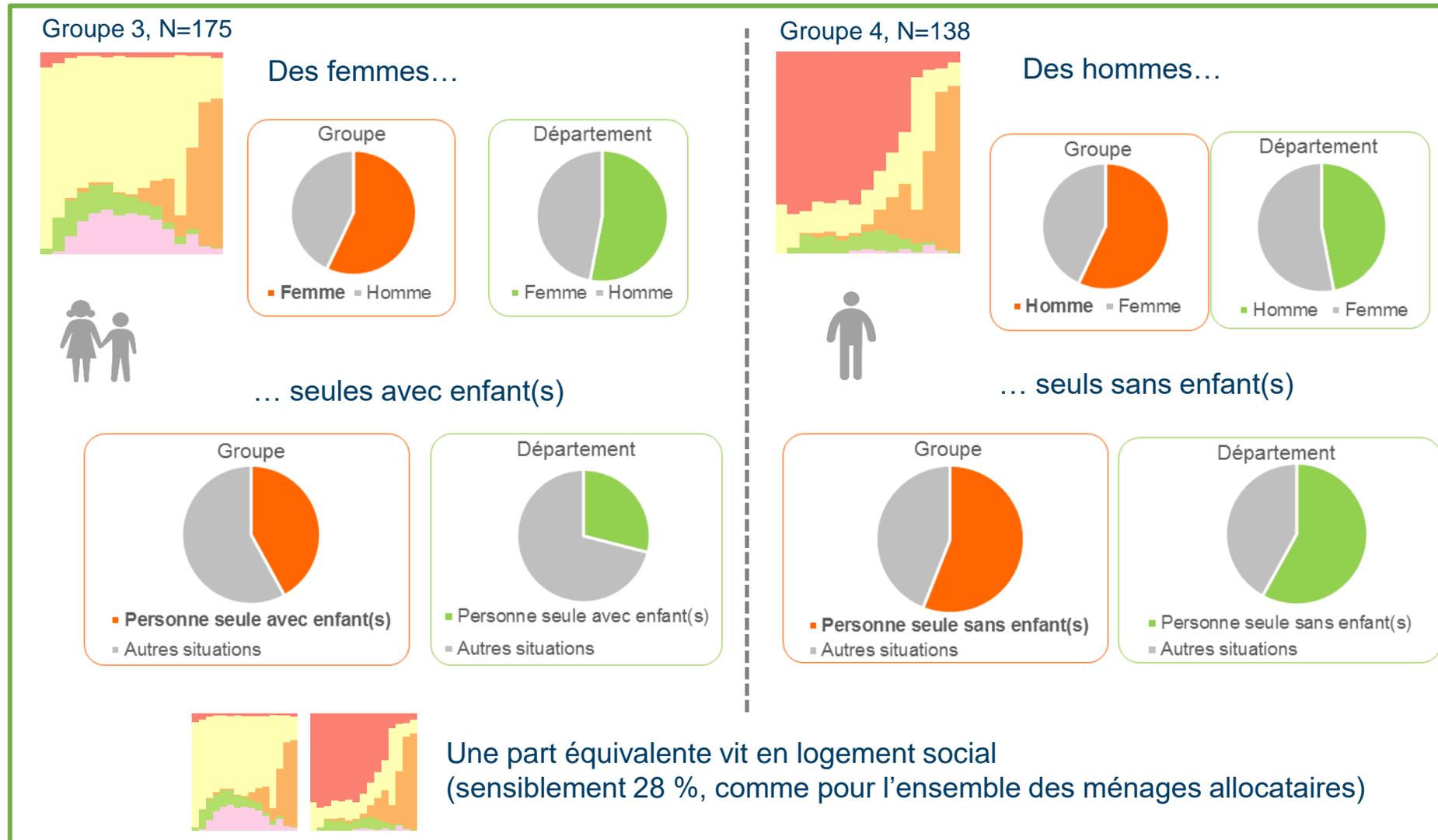
- L'exemple des trajectoires : « sorties lentes »



- 46 %** des allocataires de ce groupe ont bénéficié, à un moment donné, d'un **contrat aidé**
- 39 jours** => **3 ans** de contrat aidé // **50 %** des contrats aidés durent **plus de 7 mois**

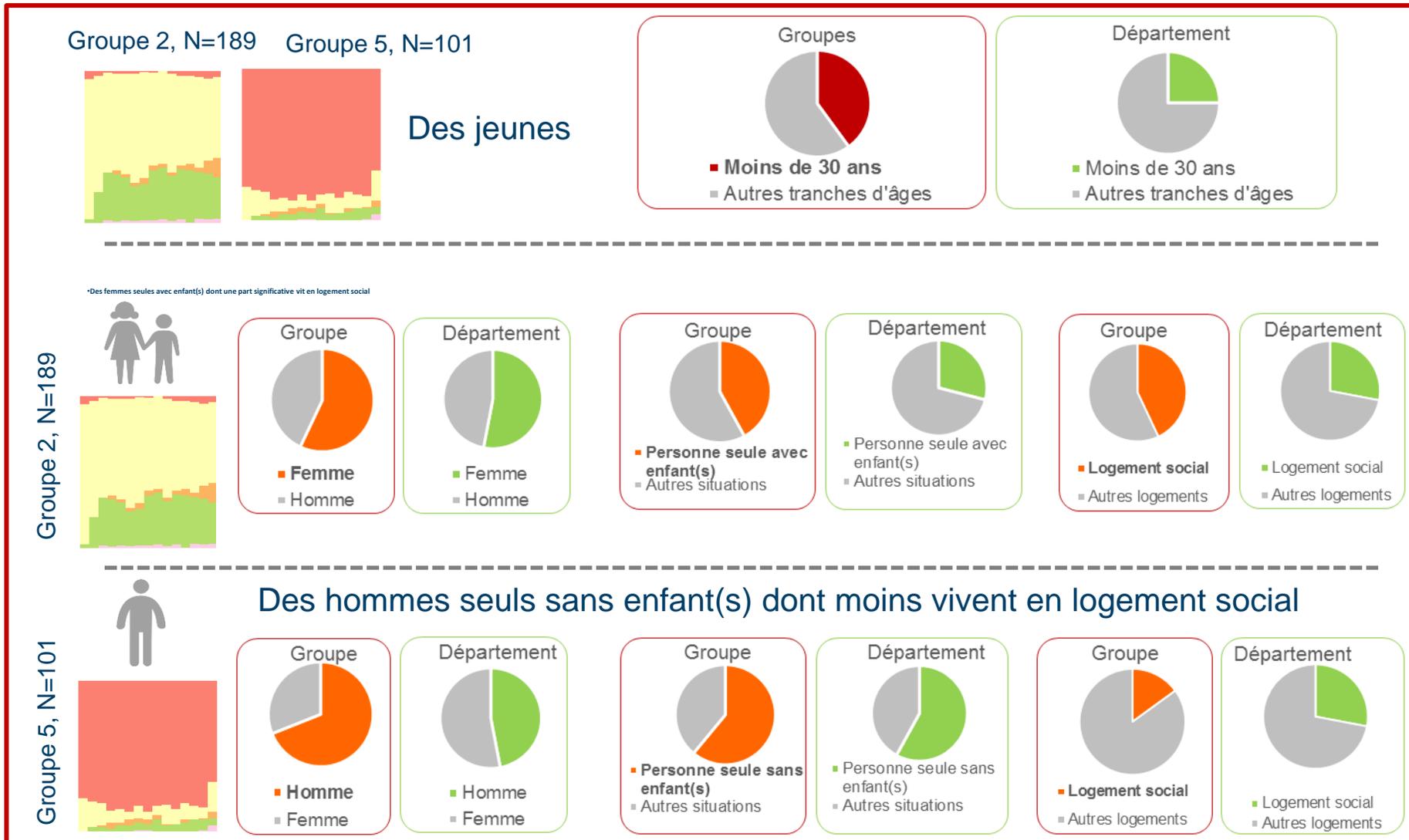
Analyse de séquences : présentation & illustration

- L'exemple des trajectoires : « sorties lentes »



Analyse de séquences : présentation & illustration

- L'exemple des trajectoires : « pas ou peu de sorties »



Analyse de séquences : présentation & illustration

Méthode ?

Analyse de séquences par appariement optimal

Utilisation de l'application **ViCaTraj**

Échantillon « **exhaustif** » (20 %)

449 300 situations individuelles suivies durant **72 mois**

- **Sorties : 14 %** (janv 2014) ; 11 % (janv 2015) ; **38 %** (déc 2018)
- Allocataires **quasi exclusivement en accompagnement « emploi »** : 98 %

PPAE et sorties lentes
11%

Peu de contrats, sorties lentes
28%

Sorties rapides
61%

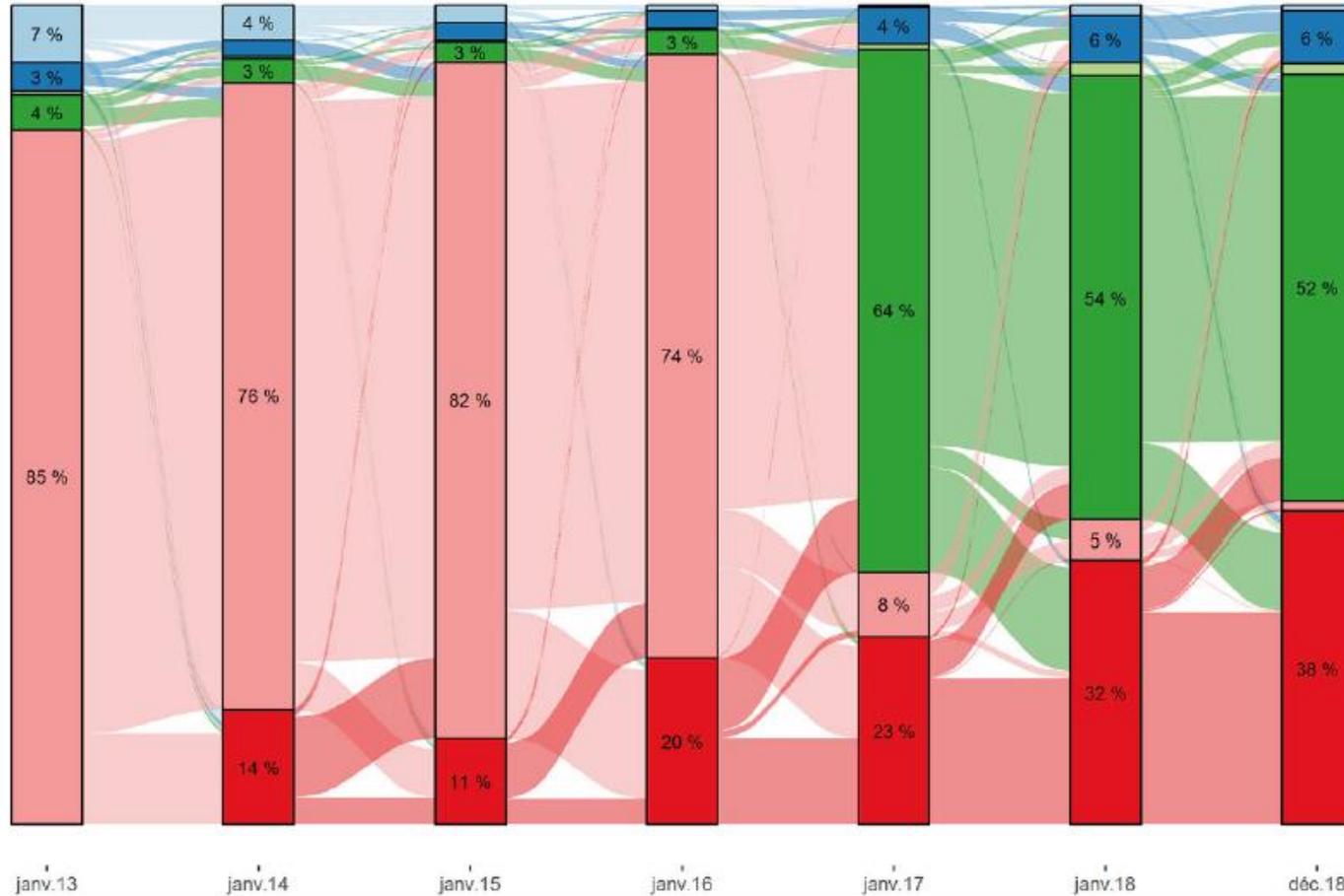
- **Sorties : 8 %** (janv 2014) ; 6 % (janv 2015) ; **29 %** (déc 2018)
- Allocataires **davantage en accompagnement « social »** (CMS ; CCAS) : 77 % contre 46 %

- **Sorties : 42 %** (janv 2014) ; 58 % (janv 2015) ; **93 %** (déc 2018)
- Allocataires **davantage en accompagnement « emploi »** (unité emploi ; PPAE) : 63 % contre 54 %



Analyse de séquences : présentation & illustration

L'analyse des dynamiques au sein du dispositif : exemple du groupe : « PPAE et sorties lentes » (11 %)

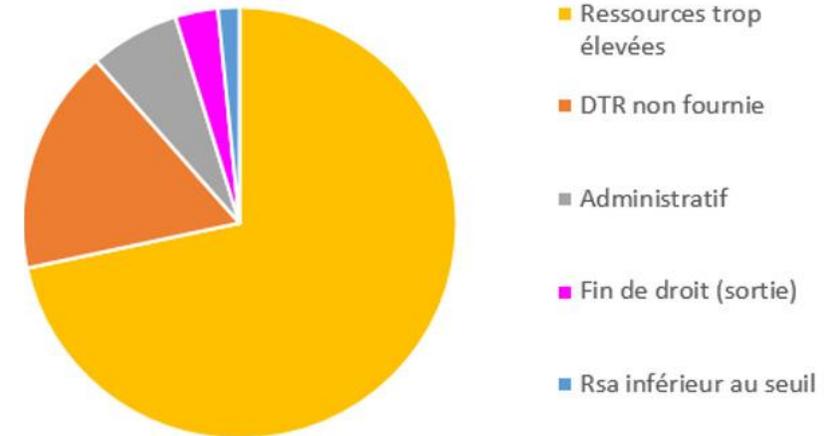


- Entrées et sorties successives du RSA après un PPAE (Pôle emploi)
- Impact de la réforme du chômage ? => Augmentation supposée du poids de ce groupe

- Contrat aidé
- Contrat d'insertion
- Pas de contrat - Non Soumis à Droits et Devoirs
- Pas de contrat - Soumis à Droits et Devoirs
- PPAE
- Sortie du dispositif (droit clos, refusé ou en attente)



Dépassement des ressources, principal motif de sortie (moyenne mensuelle 2013-2018)



Freins à l'emploi & mobilité ?

- Freins : mobilité (25 %) ; formation/qualification (20 %) ; santé (16 %) ; manque de maîtrise de la langue Française (12 %)
- Mobilité : 59 % se déplacent en transports en commun (aucun véhicule)

ViCaTraj : Visualiser et CATégoriser des TRAjectoires



Un développement collaboratif



Élie Chosson



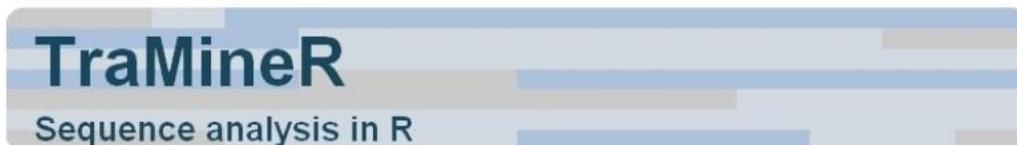
Luc-Olivier Hervé

Mélanie Boscher

Deici Alejandra Giraldo



Croisements de packages & fonctionnalités



ggalluvial

This is a [ggplot2](#) extension for alluvial diagrams.

Package ‘WeightedCluster’

Application ViCaTraj & démonstration

- Un outil gratuit développé en interne
- Charger ses propres données (après le RSA, nouvelle application à venir : MDPH)
- Paramétrer
- Innover avec l'analyse de trajectoires (par analyse de séquences avec appariement optimal)
- Exporter ses résultats
- Manipuler l'outil sans coder : boutons & menus déroulants
- Un code qui s'exécute en arrière fond



ViCaTraj Paramètres de la session Statistiques descriptives Classification des trajectoires

Chargement des fichiers de données

Choix du type de données
Un seul fichier csv contenant des données prêtes à l'emploi

Sélectionnez votre fichier source:
Browse... RSA_44.csv

Séparateur de colonnes
Point Virgule

Séparateur décimal
Point

Comment est codé le fichier ? Les accents sont-ils correctement lus ?
Latin1

La première ligne correspond-elle aux noms des variables ?
 Une variable correspond-elle à un identifiant des individus ?

Variable servant d'identifiant
NIR

Codage des valeurs manquantes
NA Vide Espace

Paramétrage des trajectoires

Variables temporelles (mettre dans l'ordre chronologique)
T2.13 X13.jul X13.oct X14.janv X14.avr X14.jul X14.oct X15.janv X15.avr X15.jul X15.oct X16.janv X16.avr X16.jul X16.oct

Dates de début et de fin
09.2019 to 09.2019

Mode de travail:
Flux en continu

Critère de sortie : nombre de mois consécutifs
3

ViCaTraj Paramètres de la session Statistiques descriptives Classification des trajectoires

Taux de transition (et taux de sortie)

Représentation des trajectoires

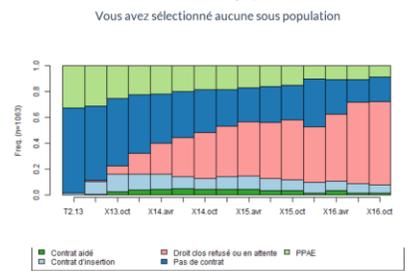
Quel graphique voulez-vous représenter ?

Chronogramme

Sous-Population

Aucune

Pour visualiser les mêmes graphiques pour des groupes, allez dans Classification des trajectoires puis faire la Matrice de distance et la Classification. Les graphiques seront dans l'onglet Visualisation des groupes.



Le chronogramme représente la proportion d'individus (ou autres unités statistiques) à chaque pas de temps dans les différentes situations.

ViCaTraj Paramètres de la session Statistiques descriptives Classification des trajectoires

Matrice de distance Classification Visualisation des groupes Statistiques descriptives

Quelle méthode voulez-vous utiliser pour regrouper les séquences ?
Combinaison de la CAH et de PAM

Choix de la méthode (CAH) : ward

Nombre de groupes
4.00 6.00

Clusters	PBC	HG	HGSD	ASW	ASWw	CH	CHsq	HC	Nb_indicateurs
4	0.7407711	0.9027100	0.9025737	0.4575516	0.4597130	304.7221	793.6640	0.05301155	8
5	0.6467977	0.8452757	0.8450241	0.3595627	0.3426197	262.7388	688.4322	0.09126858	0
6	0.6157163	0.8358317	0.8355827	0.3231264	0.3268998	233.1486	612.9941	0.09803222	0

ViCaTraj Paramètres de la session Statistiques descriptives Classification des trajectoires

Matrice de distance Classification Visualisation des groupes Statistiques descriptives

Quel graphique voulez-vous représenter ?

Graphique de flux

Sous-population

Aucune

Pas de temps

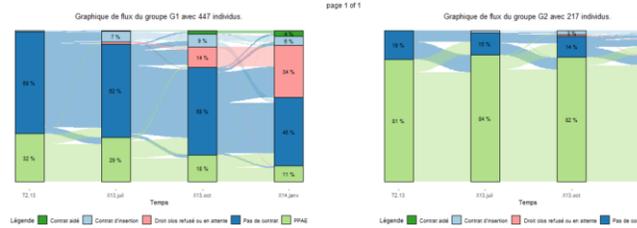
T2.13 X13.jul X13.oct X14.janv

Variable Groupe

G1 G2 G3 G4

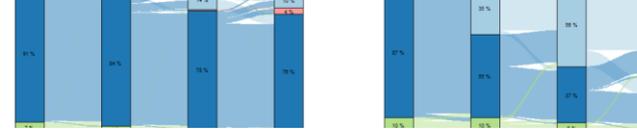
Pour visualiser les mêmes graphiques pour le global, allez dans la partie Statistiques descriptives puis dans l'onglet Représentation des trajectoires

Vous avez sélectionné les groupes G1, G2, G3, G4 et aucune sous population



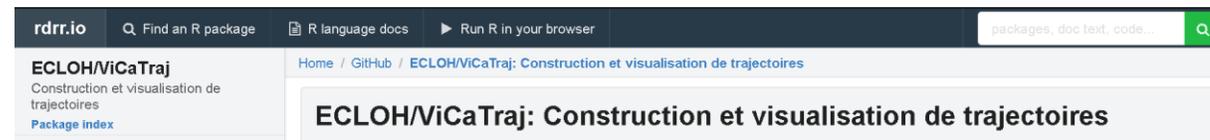
Graphique de flux du groupe G3 avec 286 individus.

Graphique de flux du groupe G4 avec 113 individus.



- Une version zippée (v1)
- Une version en ligne (v2)
- Une version GitHub & une page rdrv.io (maj en continu)
- Un réseau d'utilisateur naissant

Application ViCaTraj* : Visualiser et CATégoriser des TRAJectoires à l'aide de l'appariement optimal de séquences (OMA)



Conclusion & échanges



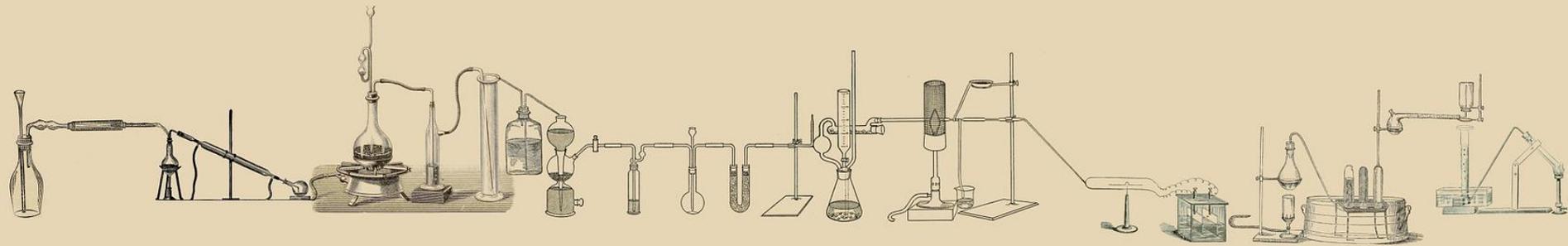
Créer ex-post des bases de données de types longitudinales/biographiques, c'est possible !



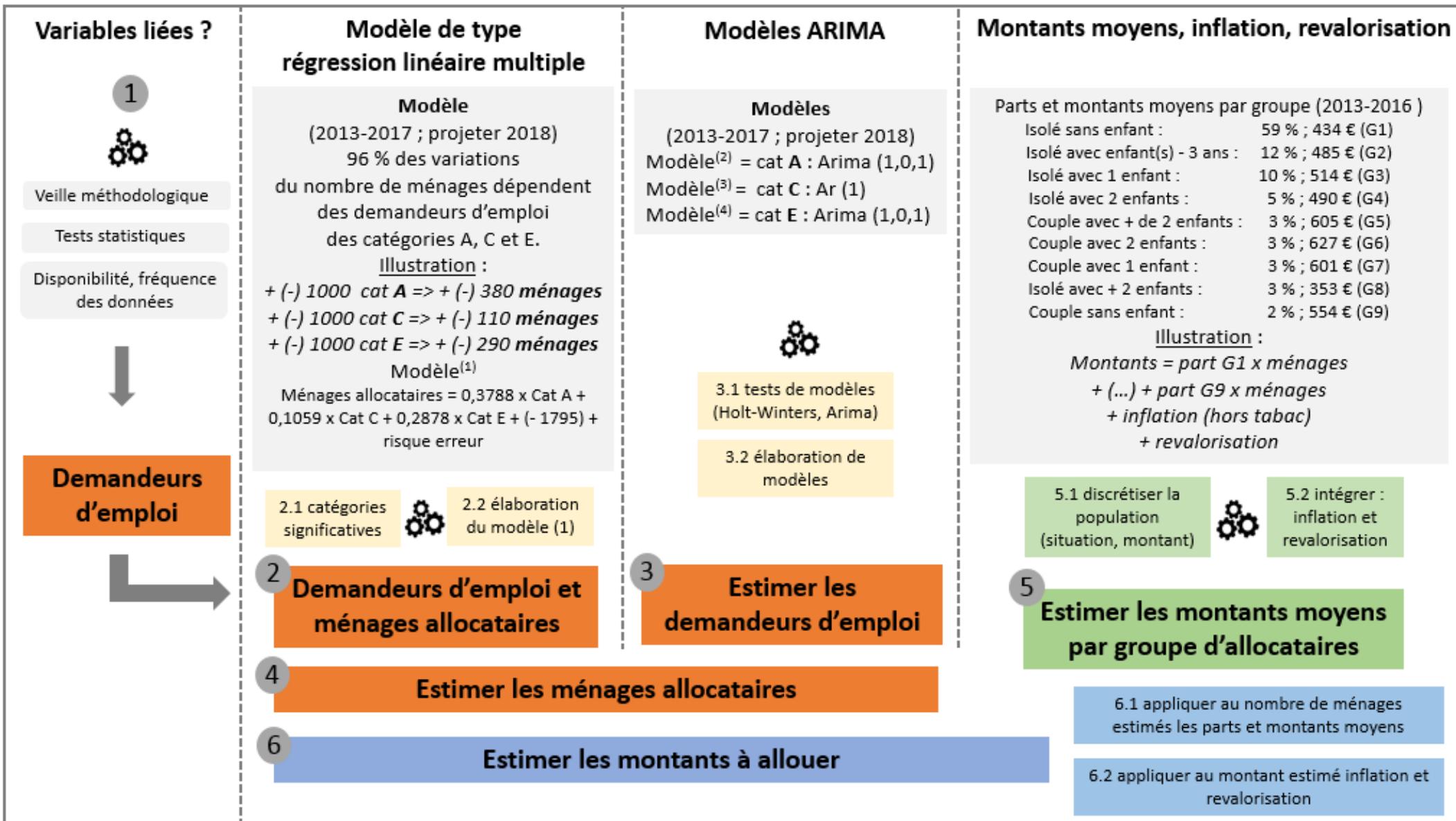
Merci !

ANNEXE

Image : Pixabay



Présentation personnelle & introduction



Méthodes de typologie de « trajectoires » ?



Tableau 3 – Les parcours professionnels de Calvin et Hobbes

	18	19	20	21	22	23	24	25	26	27	28	29
Calvin	E	E	I	I	I	A	A	A	A	A	A	A
Hobbes	E	E	E	I	A	I	A	A	A	A	A	A

- Recodage du calendrier sous forme **disjonctive complète**

Tableau 2 – Tableau disjonctif-complet du parcours de Calvin

18E	18I	18A	19E	19I	19A	20E	20I	20A	...	29E	29I	29A
1	0	0	1	0	0	0	1	0	...	0	0	1

Lecture : À 18 ans, Calvin est en études ou formation et non au chômage, ni en inactivité ou en emploi, de même à 19 ans...

Puis :

- Analyse Factorielle des Correspondances (AFC) => la distance utilisée est celle du khi-deux

OU

- Analyse en Composante Principale (ACP) non normée => la distance euclidienne est utilisée (méthode *Lhire*)

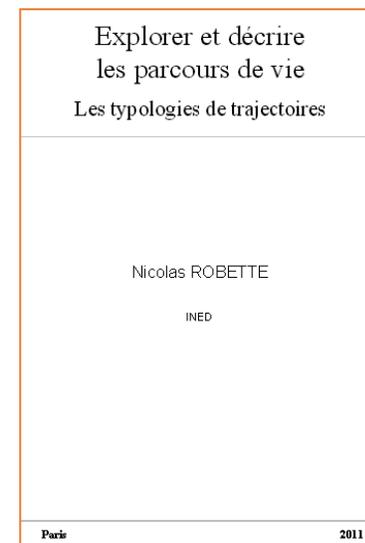
- L'analyse harmonique qualitative (**AHQ**)

Tableau 4 – Codage de l'AHQ du parcours de Calvin

1E	1I	1A	2E	2I	2A	3E	3I	3A
0.5	0.5	0	0	0.25	0.75	0	0	1

Lecture : Calvin passe la moitié de la première sous-période (de 18 à 21 ans) en étude ou formation, l'autre moitié au chômage ou en inactivité...

- 3 états : E, I, A
- 3 sous-périodes (amplitude égale) : 18-21 ans, 22-25, 26-29
- Codage : proportion de la durée dans chaque état par sous-périodes

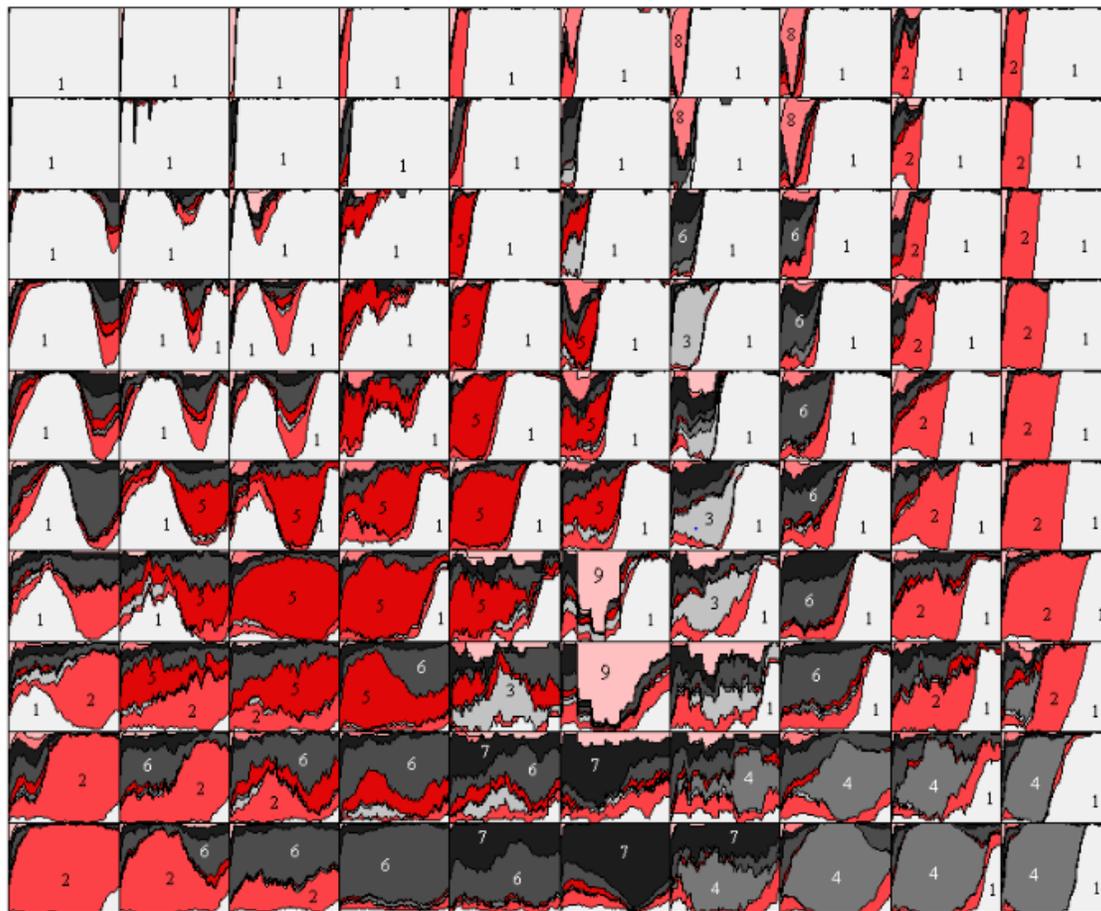


Méthodes de typologie de « trajectoires » ?



Carte 1

Typologie de l'insertion des jeunes sortis de la formation initiale en 1998 sur 7 ans
Classification obtenue à l'aide d'une cartes d'auto-organisation 10x10



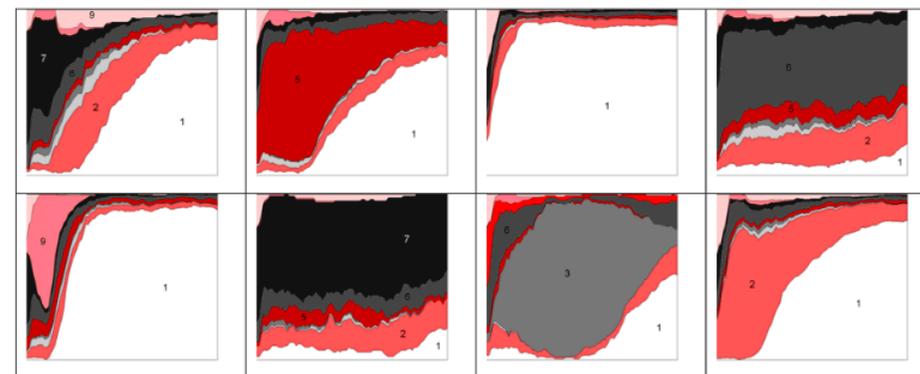
Chaque cellule correspond au chronogramme d'un groupe de trajectoires où est décrite la situation mensuelle de l'ensemble des individus de ce groupe. Le code couleur est le suivant :

- 9 Etudes
- 8 Service national
- 7 Inactivité
- 6 Chômage
- 5 Intérim
- 4 Emploi aidé
- 3 Contrat de qualification
- 2 CDD
- 1 CDI



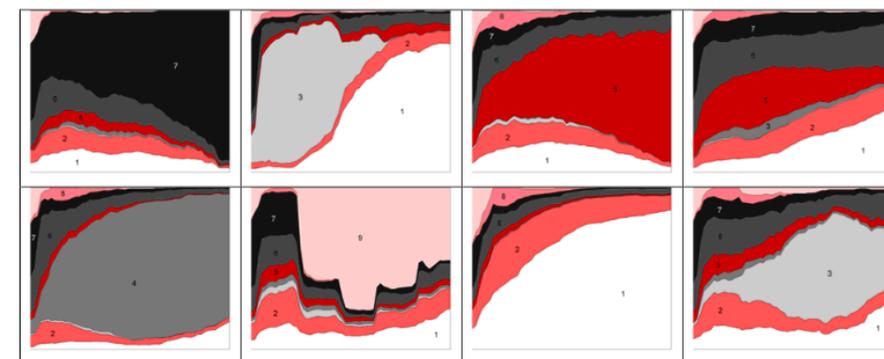
Carte 2

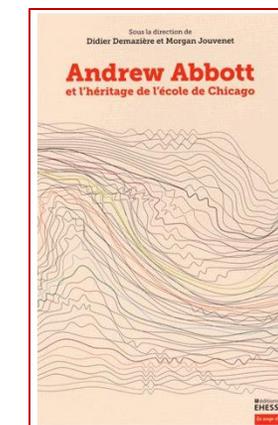
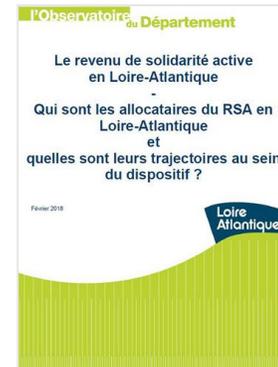
Typologie de l'insertion des jeunes sortis de la formation initiale en 1998 sur 7 ans
Partition en 8 classes à partir de la méthode de l'optimal matching



Carte 3

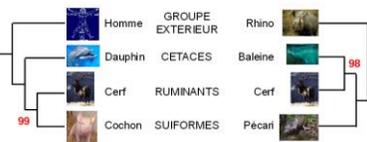
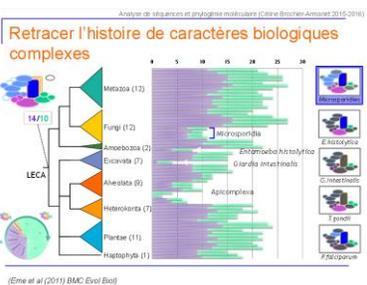
Typologie de l'insertion des jeunes sortis de la formation initiale en 1998 sur 7 ans. Partition en 8 classes à partir de la méthode de l'analyse harmonique.





Une méthode issue du domaine médical...

... appliquée aux sciences sociales...



Du prosélytisme à la sécularisation. Le processus de diffusion de l'« Optimal Matching Analysis ».



Nicolas Robette. Printemps (CNRS-UVSQ, UMR 8085)

JOURNAL ARTICLE. Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers

Analyse séquentielle et carrières militantes

Philippe Blanchard

Les trajectoires spatiales d'activité des couples

The spatial trajectories of couples' activities

Eva Lelièvre et Nicolas Robette



Journal of Statistical Software

April 2011, Volume 40, Issue 4. http://www.jstatsoft.org/

DECRIRE DES DONNEES SEQUENTIELLES EN SCIENCES SOCIALES : PANORAMA DES METHODES EXISTANTES

Laurent LESNARD (*), Thibaut DE SAINT POL (**)

(* Sciences Po, Centre de données socio-politiques Crest, Laboratoire de sociologie quantitative. (**) Insee, Conditions de vie des ménages Crest, Laboratoire de sociologie quantitative

TRAVAIL / EMPLOI

Organisation du travail dans la semaine des individus et des couples actifs : le poids des déterminants économiques et sociaux

Laurent Lesnard*, Thibaut de Saint Pol**



"Trajectoires de monoparentalité à Bruxelles : les femmes face aux épreuves de la parentalité"

Wagener, Martin

Appariement optimal d'analyse de séquences : quels outils aujourd'hui ?

TraMineR
Sequence analysis in R

Introduction Aperçu des possibilités de TraMineR Documentation et communauté d'utilisateurs Références

Logiciels pour l'analyse de séquences

Etat des lieux au début du projet

- **Optimize** le logiciel d'Abbott (Abbott, 1997)
 - Calcul des distances d'optimal matching
 - Plus maintenu
- **TDA** (Rohwer and Pötter, 2002)
 - logiciel statistique gratuit, calcul des distances d'optimal matching
- **SQ-adoss**, macros Stata, (Brzinsky-Fay et al., 2006)
 - gratuit si on a une licence Stata
 - distances optimal matching, visualisation
- **CHESA** logiciel gratuit de Elzinga (2007)
 - Nombreuses métriques, dont plusieurs non fondées sur l'alignement
 - Turbulence
- **MARCH** (Berchtold and Berchtold, 2004)
 - Modèles des transitions, chaînes de Markov cachées, ...



Introduction Aperçu des possibilités de TraMineR Documentation et communauté d'utilisateurs Références

Ce que TraMineR permet de faire

- Prise en charge et conversion de **différents types** de données longitudinales
- Gestion des **poids** et des **données manquantes**
- **Visualisation** d'un ensemble de séquences (index plot, séquences fréquentes, distributions transversales, et plus...)
- **Caractéristiques longitudinales** de séquences individuelles (complexité, durées de séjour dans chaque état, entropie longitudinale, turbulence, et plus ...)
- Séquence de **caractéristiques transversales** (distribution des états, entropie transversale, état modal)
- Autres **caractéristiques agrégées** (taux de transition, durées moyennes de séjour dans chaque état)
- **Dissimilarités** entre paires de séquences (Optimal matching, Longest Common Subsequence, Hamming, Dynamic Hamming, Multichannel et plus ...)
- Mesure de **dispersion** d'un ensemble de séquences
- Séquences **représentatives**
- **ANOVA** et **arbres de régression** à partir de matrices de dissimilarités
- Extraction de **séquences d'événements** fréquents
- Identification de séquences d'événements discriminantes



Analyse de séquences : présentation & illustration

Méthode : 4 cas principaux pour définir les coûts en pratique

- **Cas 1/ Coût insertion-suppression (*indel* = 1) < coûts de substitution (2)**

=> *Choix usuellement retenu par défaut*

=> utiliser uniquement les coûts *indel* et favoriser la comparaison uniquement sur le rapprochement temporel d'événements identiques (nombre d'unités temporelles)

OU

- **Cas 2/ Coût insertion-suppression (*indel*) = coûts de substitution**

=> absence de justifications théoriques

OU

- **Cas 3/ Coût insertion-suppression > coûts de substitution**

=> utiliser uniquement les coûts de **substitution** et favoriser la comparaison uniquement sur le rapprochement temporel d'événements identiques (nombre d'unités temporelles)

=> **rapprochement** (éloignement) = longueur la plus **courte** (longue) de **sous-séquences communes**

OU

- **Cas 4/ sur la probabilité observée de passage d'un état à l'autre**

=> *méthode retenue pour l'application sur le RSA*

Table 10. Summary of dissimilarity measures between state sequences

Measure	Type		Description	Properties					Parameters		
	Dis	AttEdt		Metric	Eucl	T.warp	S.dep	Ctxt	Subst.	Indels	Others
CHI2, EUCLID	x		Distance between per period state distributions	x	x	x					Number of periods K
CHI2fut (Rousset)	x		Position-wise state distances based on shared future	x	x			x			Time-lag weighting function
NMS (Elzinga)	x		Based on number of matching subsequences	x	x	x		x			
SVRspell	x		Based on number of matching spell subsequences with spell-length weights	x	x	x	x	x	User		Subsequence length weight a , spell duration weight b
HAM (Hamming) generalized	x	x	Number of mismatches	x	x^b						
		x	Sum of mismatches with state-dependent weights	x^a	$x^{b,c}$		x		User		
DHD (Lesnard)	x		Sum of mismatches with position-wise state-dependent weights				x	x	Data		
OM	x		Minimum cost for turning x into y using theoretically defined costs	x^a		x	x		User	Mult	
LCS / OM(1,2) / Levenshtein-II	x	x	Based on length of LCS / Number of indels	x		x					
	feature	x	Costs based on state features	x		x	x		Features	Single	State features
future	x		Costs based on similarity between conditional state distributions q periods ahead	x		x	x		Data	Single	Forward lag q
trate	x		Costs based on transition rates			x	x		Data	Single	Transition lag q
opt ^{na} (Gauthier)	x		Costs adjusted to increase similarity between similar sequences	ⁿ		x	x		Data	Single	Similarity rate
indels, indelslog	x		State dependent indels based on inverse or log inverse state frequencies.	x		x				Auto	
OMloc (Holister)	x		Context dependent indel costs			x	x	x	User	Auto	Expansion cost e , Context g
OMslen (Halpin)	x		Costs weighted by spell length	x		x	x	x	User	Mult ^{na}	Spell length weight h
OMspell (new)	x		OM between sequences of spells	x^a		x	x	x	User	Mult ^{na}	Expansion cost e
OMstran (new)	x		OM between sequences of transitions	x^a		x	x	x	User	Mult	Origin-transition trade-off w , Transition indel cost function

^a If costs fulfil the triangle inequality. ^b Squared Euclidean distance. ^c If costs are squared Euclidean distances. ^{na} Not available in TraMineR. ⁿ Can generate negative dissimilarities.

TABLE 2 – Mesures de la qualité d'un regroupement.

Nom	Abrv.	Étendue	Min/Max	Interprétation
Point Biserial Correlation	PBC	$[-1; 1]$	Max	Mesure de la capacité du clustering à reproduire les distances.
Hubert's Gamma	HG	$[-1; 1]$	Max	Mesure de la capacité du clustering à reproduire les distances (ordre de grandeur).
Hubert's Somers D	HGSD	$[-1; 1]$	Max	Mesure de la capacité du clustering à reproduire les distances (ordre de grandeur) avec prise en compte des égalités sur les distances.
Hubert's C	HC	$[0; 1]$	Min	Écart entre la partition obtenue et la meilleure partition qu'il serait théoriquement possible d'obtenir avec ce nombre de groupes et ces distances.
Average Silhouette Width	ASW	$[-1; 1]$	Max	Cohérence des assignations. Une cohérence élevée indique des distances inter-groupes élevées et une forte homogénéité intragroupe.
Average Silhouette Width (weighted)	ASWw	$[-1; 1]$	Max	Idem que précédant, si l'unité des poids n'a pas un sens explicite.
Calinski-Harabasz index	CH	$[0; +\infty[$	Max	Pseudo F calculé à partir des distances.
Calinski-Harabasz index	CHsq	$[0; +\infty[$	Max	Idem que précédant, mais en utilisant les distances <i>au carré</i> .
Pseudo R^2	R2	$[0; 1]$	Max	Part de la dispersion expliquée par la solution de clustering (uniquement pour comparer des partitions avec nombre de groupes identiques).
Pseudo R^2	R2sq	$[0; 1]$	Max	Idem que précédant, mais en utilisant les distances <i>au carré</i> .

- Abrv : abréviation utilisée dans la librairie `WeightedCluster`.
- Étendu : intervalle des valeurs possibles.
- Min/Max : Est-ce qu'une bonne partition minimise ou maximise cette mesure ?
- Interprétation de la valeur.

Vendredi 2 avril
9h30-12h30
Visio



R à l'usage des
sciences sociales



Luc-Olivier Hervé
data scientist, Cour des comptes
& Doctorant-chercheur, LEMNA



Visualiser et catégoriser des trajectoires :
panorama des méthodes existantes,
application de l'analyse de séquences par appariement optimal
et démonstration d'une nouvelle application dédiée : ViCaTraj*



Application développée de manière collaborative (Départements de Loire-Atlantique et de l'Isère ; Élie Chosson, MRIE : www.mrie.org) permettant d'avoir accès à une interface utilisateur basée sur l'articulation de packages existants (TraMineR ; WeightedCluster ; ggalluvial notamment).