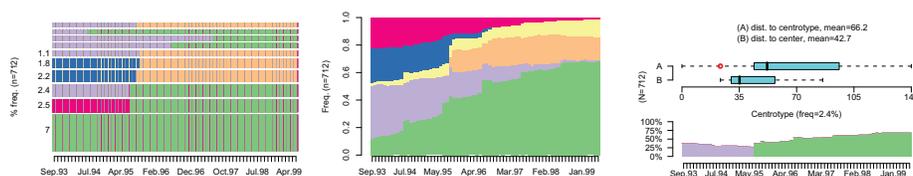


# TraMineR: une librairie R pour l'analyse de données séquentielles

Alexis Gabadinho\*, Nicolas S. Müller\*, Gilbert Ritschard\*, Matthias Studer\*

\*Département d'économétrie et Laboratoire de démographie, Université de Genève  
{alexis.gabadinho,nicolas.muller,gilbert.ritschard,matthias.studer}@unige.ch  
<http://mephisto.unige.ch/traminer/>



TraMineR est une librairie destinée à la fouille, la description et la visualisation de séquences d'états ou d'événements, principalement conçue pour l'analyse de données biographiques longitudinales issues des sciences sociales, telles que des trajectoires professionnelles ou familiales. Cependant, les méthodes sont applicables à tout type de données séquentielles catégorielles.

La librairie s'intègre dans l'environnement statistique R, un logiciel libre et multi-plateformes, et est installable depuis le CRAN (Comprehensive R Archive Network). Elle est accompagnée d'un guide de l'utilisateur détaillé et illustré par de nombreux exemples.

Les fonctions proposées par TraMineR permettent notamment de

- manipuler et transformer différents formats de données longitudinales ;
- visualiser des séquences à l'aide de plusieurs types de graphiques ;
- déterminer le centrotyp et mesurer la dispersion d'un ensemble de séquences ;
- mesurer des caractéristiques de séquences individuelles (durée par état, entropie, turbulence) ;
- déterminer des caractéristiques de l'ensemble des séquences (entropie par âge, taux de transitions, durées moyennes dans chaque état, fréquences) ;
- calculer la distance entre séquences à l'aide de différentes métriques ;
- analyser l'hétérogénéité des séquences (analyse de variance et arbre d'induction) ;
- extraire les sous-séquences typiques d'événements ;
- identifier les sous-séquences les plus discriminantes.

Ce travail a été réalisé dans le cadre d'un projet de recherche sur la fouille d'histoires d'événements financé par le Fonds national suisse pour la recherche scientifique.