

IUSSP online training: Machine learning tools for the analysis of migration trajectories (13 and 15 of June, 2023)

In migration and residential mobility research, there is a growing interest in understanding various temporal processes, including intricate patterns of repeated and circular mobility throughout a person's life, as well as the impact of earlier moves on life outcomes in later years. Concurrently, the rise in available longitudinal data and the emergence of Big Data present methodological challenges in identifying essential properties and relationships. Consequently, there is a need for efficient methods to analyze regularities in longitudinal migration micro-data. The sequence analysis toolbox offers valuable tools to address these demands effectively.

The training consists of two sessions. The first session serves as an introduction to the sequence analysis toolbox, focusing on the examination of individual migration trajectories. It covers fundamental aspects such as managing and describing categorical time series or sequence data. Additionally, it delves into algorithmic methods used to identify regularities in migration trajectories. In the second session, a machine learning approach –specifically feature selection models– is introduced. This approach allows for the assessment of complex associations between previous migration trajectories and individual life outcomes.

The training emphasizes practical application, utilizing the free software R and a combination of real and simulated datasets. Registered participants will have access to comprehensive training materials, including handouts and R code, enabling them to reinforce their learning beyond the training sessions.

The online training is organized on behalf of the International Union for the Scientific Study of Population (IUSSP) [Scientific Panel on Lifetime Migration](#).¹

Session I: Introduction to sequence analysis

Summary: This session will provide a concise introduction to sequence analysis using the free software R. The session will introduce the notions of sequence or categorical time-series data, handling and describing sequence data in R, sequence comparison (i.e. optimal matching), and sequence grouping (i.e. cluster analysis). Real-world applications are used for exemplification, based on open-access migration data. Upon completion, participants will be familiar with measures of migration trajectories, typical methods to identify regularities in trajectory or sequence data, and efficient ways to describe and visualize migration trajectories using R.

Pre-requisites: The method will be taught assuming no prior knowledge. There are no prerequisites to attend this session, but working knowledge of the software R is desirable.

When: **June 13** (Tuesday); **CET 10:30 am -12:30 pm**

Where: online via [Zoom](#) (link will be provided upon registration)

Instructor: Sergi Vidal, Centre for Demographic Studies, Universitat Autònoma de Barcelona

¹ Note that these are hybrid sessions, where participants in the European Doctoral School in Demography will be attending in person from the Centre for Demographic Studies in Barcelona. The remote audience will have the opportunity to engage with the instructors exclusively through written questions in the chat.

Session II: Feature selection approaches

Summary: This session will introduce a machine learning approach to investigate the links between the prominent features that structure individual migration trajectories and later life outcomes using the free software R. The approach works in several steps. First, several properties of the sequences of migration episodes such as order, timing and duration are extracted. Then, feature selection algorithms (i.e. Boruta and LASSO) are used to identify the most relevant properties associated with the study outcome. Last, the directions of the associations are established using predictive regression models. The method will be illustrated using examples linking migration trajectories from early ages with life outcomes (e.g. occupational achievement).

Pre-requisites: The method will be taught assuming rudimentary knowledge of sequence analysis. Attending Session I of this training event, or having done a prior reading or basic course of sequence analysis will suffice. Working knowledge of the software R is desirable.

When: **June 15** (Thursday); **CET 10:30 am -12:30 pm**

Where: online via [Zoom](#) (link will be provided upon registration)

Instructor: Danilo Bolano, Bocconi University Carlo F. Dondena Centre for Research on Social Dynamics

About the instructors

Danilo Bolano is a research fellow and a lecturer in social demography at Bocconi University Carlo F. Dondena Centre for Research on Social Dynamics. His main research interest is in life course research and in applying and developing advanced statistical tools to improve understanding of individual behaviours and attitudes. Among others, he has developed the feature selection approach introduced in the session. Currently, he is the principal investigator on project FAST Population Change Flows, Attractiveness and Territories' resilience; a project on internal migration and resilience of the European territories.

Sergi Vidal is an associate professor at the Department of Sociology of the Universitat Autònoma de Barcelona, and a researcher at the Centre for Demographic Studies. He has expertise on migration and residential mobility research adopting longitudinal approaches, being the recipient of a Consolidator Grant award from the European Research Council to examine migration and mobility trajectories from early childhood into adulthood ([Lifelongmove project](#)).

Registration and contact

For registration, please click the following link <https://forms.gle/HRwRDPnJQG4Eip8K9>

Registration deadline is **June 5**.

For queries about the training and registration, contact us at jussptraining@ced.uab.es