Sequential Data Analysis
Introduction

Gilbert Ritschard

Alexis Gabadinho, Matthias Studer

Institute for Demographic and Life Course Studies, University of Geneva
and NCCR LIVES: Overcoming vulnerability, life course perspectives
http://mephisto.unige.ch/traminer

September - November, 2012

© G. Ritschard (2012), 1/67. Distributed under licence CC BY-NC-ND 3.0

Outline

Outline

Section outline

Objectives of the course

- Concepts related to (categorical) sequence data
  - Types of sequences: with or without time content, states, transitions, events, ...
- Methods for extracting knowledge from sequence data
- Principles of sequence analysis
  - exploratory approaches
  - more causal and predictive approaches
- Practice of sequence analysis (TraMineR)

© G. Ritschard (2012), 5/67. Distributed under licence CC BY-NC-ND 3.0

Objectives of the course

- Concepts related to (categorical) sequence data
  - Types of sequences: with or without time content, states, transitions, events, ...
- Methods for extracting knowledge from sequence data
- Principles of sequence analysis
  - exploratory approaches
  - more causal and predictive approaches
- Practice of sequence analysis (TraMineR)

© G. Ritschard (2012), Distributed under licence CC BY-NC-ND 3.0

Objectives of the course

- Concepts related to (categorical) sequence data
  - Types of sequences: with or without time content, states, transitions, events, ...
- Methods for extracting knowledge from sequence data
- Principles of sequence analysis
  - exploratory approaches
  - more causal and predictive approaches
- Practice of sequence analysis (TraMineR)

© G. Ritschard (2012), 5/67. Distributed under licence CC BY-NC-ND 3.0

Objectives of the course

- Concepts related to (categorical) sequence data
  - Types of sequences: with or without time content, states, transitions, events, ...
- Methods for extracting knowledge from sequence data
- Principles of sequence analysis
  - exploratory approaches
  - more causal and predictive approaches
- Practice of sequence analysis (TraMineR)

 Distributed under licence CC BY-NC-ND 3.0

Objectives of the course

- Concepts related to (categorical) sequence data
  - Types of sequences: with or without time content, states, transitions, events, ...
- Methods for extracting knowledge from sequence data
- Principles of sequence analysis
  - exploratory approaches
  - more causal and predictive approaches
- Practice of sequence analysis (TraMineR)

Objectives of the course

- Concepts related to (categorical) sequence data
  - Types of sequences: with or without time content, states, transitions, events, ...
- Methods for extracting knowledge from sequence data
- Principles of sequence analysis
  - exploratory approaches
  - more causal and predictive approaches
- Practice of sequence analysis (TraMineR)

 Distributed under licence CC BY-NC-ND 3.0

Objectives of the course

- Concepts related to (categorical) sequence data
  - Types of sequences: with or without time content, states, transitions, events, ...
- Methods for extracting knowledge from sequence data
- Principles of sequence analysis
  - exploratory approaches
  - more causal and predictive approaches
- Practice of sequence analysis (TraMineR)

Objectives of this first lesson

- Understand what kind of data we will be considering
  - State sequences and event sequences
  - How do they compare with other longitudinal data?
- Get an idea of what we can learn from sequence data?
- TraMineR: A first run

Objectives of this first lesson

- Understand what kind of data we will be considering
  - State sequences and event sequences
  - How do they compare with other longitudinal data?
- Get an idea of what we can learn from sequence data?
- TraMineR: A first run

Objectives of this first lesson

- Understand what kind of data we will be considering
  - State sequences and event sequences
  - How do they compare with other longitudinal data?
- Get an idea of what we can learn from sequence data?
- TraMineR: A first run

Objectives of this first lesson

- Understand what kind of data we will be considering
  - State sequences and event sequences
  - How do they compare with other longitudinal data?
- Get an idea of what we can learn from sequence data?
- TraMineR: A first run

Objectives of this first lesson

- Understand what kind of data we will be considering
  - State sequences and event sequences
  - How do they compare with other longitudinal data?
- Get an idea of what we can learn from sequence data?
- TraMineR: A first run

Outline

About longitudinal data: Sequence data

Sequence data

- Multiple cases (n cases)
- For each case a sorted list of (categorical) values

- Example:

$$\begin{array}{lllllll} 1: & a & a & d & d & c & \\ 2: & a & b & b & c & c & d \\ 3: & b & c & c & & & \\ . & . & . & . & . & & \end{array}$$

© G. Ritschard (2012), 8/67. Distributed under licence CC BY-NC-ND 3.0

What is longitudinal data?

Longitudinal data

- Repeated observations on units observed over time (Beck and Katz, 1995).
- "A dataset is longitudinal if it tracks the same type of information on the same subjects at multiple points in time". (http://www.caldercenter.org/whatis.cfm)
- "The defining feature of longitudinal data is that the multiple observations within subject can be ordered" (Singer and Willett, 2003)

Successive transversal data vs longitudinal data

- Successive transversal observations (same units)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

- Longitudinal observations

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

 Distributed under licence CC BY-NC-ND 3.0

Successive transversal data vs longitudinal data

- Successive transversal observations (same units)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

- Longitudinal observations

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

© G. Ritschard (2012), 10/67. Distributed under licence CC BY-NC-ND 3.0

Repeated independent cross sectional observations

- Successive independent transversal observations

id	t_1	t_2	t_3	$\cdots$
11	B	.	.	$\cdots$
12	A	.	.	$\cdots$
13	B	.	.	$\cdots$
.	.	.	.	$\cdots$
21	.	B	.	$\cdots$
22	.	B	.	$\cdots$
23	.	B	.	$\cdots$
.	.	.	.	$\cdots$
24	.	.	D	$\cdots$
25	.	.	C	$\cdots$
26	.	.	A	$\cdots$
.	.	.	.	$\cdots$

- This is not longitudinal ...
- but ... sequences of transversal (aggregated) characteristics.

Repeated independent cross sectional observations

- Successive independent transversal observations

id	t_1	t_2	t_3	$\cdots$
11	B	.	.	$\cdots$
12	A	.	.	$\cdots$
13	B	.	.	$\cdots$
.	.	.	.	$\cdots$
21	.	B	.	$\cdots$
22	.	B	.	$\cdots$
23	.	B	.	$\cdots$
.	.	.	.	$\cdots$
24	.	.	D	$\cdots$
25	.	.	C	$\cdots$
26	.	.	A	$\cdots$
.	.	.	.	$\cdots$

- This is not longitudinal ...
- but ... sequences of transversal (aggregated) characteristics.

Longitudinal data: Where do they come from?

- Individual follow-ups: Each important event is recorded as soon as it occurs (medical card, cellular phone, weblogs, ...).
- Panels: Periodic observation of same units
- Retrospective data (biography): Depends on interviewees' memory
- Matching data from different sources (successive censuses, tax data, social security, population registers, acts of marriages, acts of deaths, ...)

  Examples: Wanner and Delaporte (2001), censuses and population registers, Perroux and Oris (2005), 19th Century Geneva, censuses, acts of marriage, registers of deaths, register of migrations.

- Rotating panels: partial follow up

  e.g.; Swiss Labor Force Survey, SLFS, 5 year-rotating panel (Wernli, 2010).

Longitudinal data: Where do they come from?

- Individual follow-ups: Each important event is recorded as soon as it occurs (medical card, cellular phone, weblogs, ...).
- Panels: Periodic observation of same units
- Retrospective data (biography): Depends on interviewees' memory
- Matching data from different sources (successive censuses, tax data, social security, population registers, acts of marriages, acts of deaths, ...)

  Examples: Wanner and Delaporte (2001), censuses and population registers, Perroux and Oris (2005), 19th Century Geneva, censuses, acts of marriage, registers of deaths, register of migrations.

- Rotating panels: partial follow up

  e.g.; Swiss Labor Force Survey, SLFS, 5 year-rotating panel (Wernli, 2010).

 Distributed under licence CC BY-NC-ND 3.0

Longitudinal data: Where do they come from?

- Individual follow-ups: Each important event is recorded as soon as it occurs (medical card, cellular phone, weblogs, ...).
- Panels: Periodic observation of same units
- Retrospective data (biography): Depends on interviewees' memory
- Matching data from different sources (successive censuses, tax data, social security, population registers, acts of marriages, acts of deaths, ...)

  Examples: Wanner and Delaporte (2001), censuses and population registers, Perroux and Oris (2005), 19th Century Geneva, censuses, acts of marriage, registers of deaths, register of migrations.
- Rotating panels: partial follow up

  e.g.; Swiss Labor Force Survey, SLFS, 5 year-rotating panel (Wernli, 2010)

Longitudinal data: Where do they come from?

- Individual follow-ups: Each important event is recorded as soon as it occurs (medical card, cellular phone, weblogs, ...).
- Panels: Periodic observation of same units
- Retrospective data (biography): Depends on interviewees' memory
- Matching data from different sources (successive censuses, tax data, social security, population registers, acts of marriages, acts of deaths, ...)

  Examples: Wanner and Delaporte (2001), censuses and population registers, Perroux and Oris (2005), 19th Century Geneva, censuses, acts of marriage, registers of deaths, register of migrations.
- Rotating panels: partial follow up

  e.g.; Swiss Labor Force Survey, SLFS, 5 year-rotating panel (Wernli, 2010)

Longitudinal data: Where do they come from?

- Individual follow-ups: Each important event is recorded as soon as it occurs (medical card, cellular phone, weblogs, ...).
- Panels: Periodic observation of same units
- Retrospective data (biography): Depends on interviewees' memory
- Matching data from different sources (successive censuses, tax data, social security, population registers, acts of marriages, acts of deaths, ...)

  Examples: Wanner and Delaporte (2001), censuses and population registers, Perroux and Oris (2005), 19th Century Geneva, censuses, acts of marriage, registers of deaths, register of migrations.

- Rotating panels: partial follow up

  e.g.; Swiss Labor Force Survey, SLFS, 5 year-rotating panel (Wernli, 2010)

 Distributed under licence CC BY-NC-ND 3.0

State sequences: an example

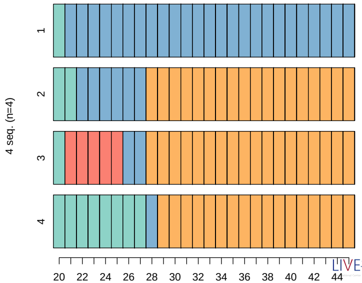
- Cohabitational state sequences (from SHP)

  2P = with 2 parents, U = with partner, C = with child, A = alone, ...

```
  Sequence
1 2P-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U
2 2P-2P-U-U-U-U-U-U-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC
3 2P-A-A-A-A-A-U-U-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC
4 2P-2P-2P-2P-2P-2P-2P-2P-U-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC
```

- Compact representation

```
    Sequence
[1] (2P,1)-(U,25)
[2] (2P,2)-(U,6)-(UC,18)
[3] (2P,1)-(A,5)-(U,2)-(UC,18)
[4] (2P,8)-(U,1)-(UC,17)
```



State sequences: an example

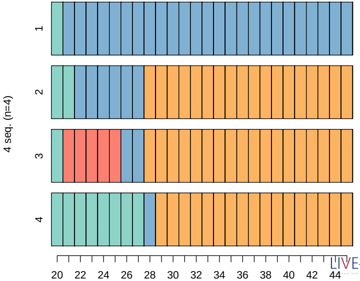
- Cohabitational state sequences (from SHP)

  2P = with 2 parents, U = with partner, C = with child, A = alone, ...

```
  Sequence
1 2P-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U-U
2 2P-2P-U-U-U-U-U-U-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC
3 2P-A-A-A-A-A-U-U-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC
4 2P-2P-2P-2P-2P-2P-2P-2P-U-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC-UC
```

- Compact representation

```
    Sequence
[1] (2P,1)-(U,25)
[2] (2P,2)-(U,6)-(UC,18)
[3] (2P,1)-(A,5)-(U,2)-(UC,18)
[4] (2P,8)-(U,1)-(UC,17)
```



Outline

 Distributed under licence CC BY-NC-ND 3.0

Section outline

3. What is sequence analysis (SA)?
   - How does SA compare with other longitudinal methods?
   - Types of categorical sequences

 Distributed under licence CC BY-NC-ND 3.0

What is sequence analysis (SA)?

- Sequence analysis (SA)
  - concerned by categorical sequences,
  - holistic: interest is in the whole sequence, not just one element in the sequence (unlike survival analysis for example)
- Aim is
  - Characterizing sets of sequences
  - Identifying typical (sequence) patterns
  - Study relationship with individual characteristics and environment

What is sequence analysis (SA)?

- Sequence analysis (SA)
  - concerned by categorical sequences,
  - holistic: interest is in the whole sequence, not just one element in the sequence (unlike survival analysis for example)
- Aim is
  - Characterizing sets of sequences
  - Identifying typical (sequence) patterns
  - Study relationship with individual characteristics and environment

What is sequence analysis (SA)?

- Sequence analysis (SA)
  - concerned by categorical sequences,
  - holistic: interest is in the whole sequence, not just one element in the sequence (unlike survival analysis for example)
- Aim is
  - Characterizing sets of sequences
  - Identifying typical (sequence) patterns
  - Study relationship with individual characteristics and environment

 Distributed under licence CC BY-NC-ND 3.0

What is sequence analysis (SA)?

- Sequence analysis (SA)
  - concerned by categorical sequences,
  - holistic: interest is in the whole sequence, not just one element in the sequence (unlike survival analysis for example)
- Aim is
  - Characterizing sets of sequences
  - Identifying typical (sequence) patterns
  - Study relationship with individual characteristics and environment

What is sequence analysis (SA)?

- Sequence analysis (SA)
  - concerned by categorical sequences,
  - holistic: interest is in the whole sequence, not just one element in the sequence (unlike survival analysis for example)
- Aim is
  - Characterizing sets of sequences
  - Identifying typical (sequence) patterns
  - Study relationship with individual characteristics and environment

What is sequence analysis (SA)?

- Sequence analysis (SA)
  - concerned by categorical sequences,
  - holistic: interest is in the whole sequence, not just one element in the sequence (unlike survival analysis for example)
- Aim is
  - Characterizing sets of sequences
  - Identifying typical (sequence) patterns
  - Study relationship with individual characteristics and environment

What is sequence analysis (SA)?

- Sequence analysis (SA)
  - concerned by categorical sequences,
  - holistic: interest is in the whole sequence, not just one element in the sequence (unlike survival analysis for example)
- Aim is
  - Characterizing sets of sequences
  - Identifying typical (sequence) patterns
  - Study relationship with individual characteristics and environment

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

 Distributed under licence CC BY-NC-ND 3.0

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

 Distributed under licence CC BY-NC-ND 3.0

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

 Distributed under licence CC BY-NC-ND 3.0

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

© G. Ritschard (2012), Distributed under licence CC BY-NC-ND 3.0

Other Longitudinal methods

- Numerical longitudinal data: Essentially modeling approaches
  - Multilevel models (Fixed and random effects) (Gelman and Hill, 2007; Frees, 2004)
    - Can handle mixed longitudinal-cross-sectional data, but do not really describe dynamics
  - Growth curve models (specialized Structural equation models) (McArdle, 2009)
- Categorical longitudinal data
  - Multilevel models for nominal and ordinal data (Hedeker, 2007; Müller, 2011)
  - Survival approaches (descriptive survival curves and hazard regression models) (Therneau and Grambsch, 2000)
  - Markov chain models and Probabilistic suffix trees (Berchtold and Raftery, 2002; Bejerano and Yona, 2001)
  - Aligning techniques (biology) (Sharma, 2008)

Characteristics of sequence analysis

- Essentially (but not exclusively) exploratory
- Focus on the sequence (evolution along the time frame)
- Holistic: sequences as unit of observation
- Looks for typical patterns (rather than at generating process)

Characteristics of sequence analysis

- Essentially (but not exclusively) exploratory
- Focus on the sequence (evolution along the time frame)
- Holistic: sequences as unit of observation
- Looks for typical patterns (rather than at generating process)

Characteristics of sequence analysis

- Essentially (but not exclusively) exploratory
- Focus on the sequence (evolution along the time frame)
- Holistic: sequences as unit of observation
- Looks for typical patterns (rather than at generating process)

Characteristics of sequence analysis

- Essentially (but not exclusively) exploratory
- Focus on the sequence (evolution along the time frame)
- Holistic: sequences as unit of observation
- Looks for typical patterns (rather than at generating process)

 Distributed under licence CC BY-NC-ND 3.0

Section outline

3 What is sequence analysis (SA)?

- How does SA compare with other longitudinal methods?
- Types of categorical sequences

Types of categorical sequences

Nature of sequences

Depends on

- Chronological order?
  - If yes, we can study timing and duration.
- Information conveyed by position j in the sequence
  - If position is a time stamp, differences between positions reflect durations.
- Nature of the elements of the alphabet
  - states, transitions or events, letters, proteins, ...

Types of categorical sequences

Nature of sequences

Depends on

- Chronological order?
  - If yes, we can study timing and duration.
- Information conveyed by position j in the sequence
  - If position is a time stamp, differences between positions reflect durations.
- Nature of the elements of the alphabet
  - states, transitions or events, letters, proteins, ...

State versus event sequences

- An important distinction for chronological sequences is between
  state sequences and event sequences
  - A State, such as 'living with a partner' or 'being unemployed', lasts the whole unit of time
  - An event, such as 'moving in with a partner' or 'ending education', does not last but provokes a state change, possibly in conjunction with other events.

 Distributed under licence CC BY-NC-ND 3.0

State versus event sequences

- An important distinction for chronological sequences is between
  state sequences and event sequences
  - A State, such as 'living with a partner' or 'being unemployed', lasts the whole unit of time
  - An event, such as 'moving in with a partner' or 'ending education', does not last but provokes a state change, possibly in conjunction with other events.

 Distributed under licence CC BY-NC-ND 3.0

State versus event sequences

- An important distinction for chronological sequences is between
  state sequences and event sequences
  - A State, such as 'living with a partner' or 'being unemployed', lasts the whole unit of time
  - An event, such as 'moving in with a partner' or 'ending education', does not last but provokes a state change, possibly in conjunction with other events.

State versus event sequences: examples

Time stamped events

Sandra	Ending education in 1980	Start working in 1980
Jack	Ending education in 1981	Start working in 1982

- There can be simultaneous events (see Sandra)
- Elements at same position do not occur at same time

State sequence view

year	1979	1980	1981	1982	1983
Sandra	Education	Education	Employed	Employed	Employed
Jack	Education	Education	Education	Unemployed	Employed

- Only one state at each observed time
- Position conveys time information: All states at position 2 are states in 1980.

© G. Ritschard (2012), 22/67. Distributed under licence CC BY-NC-ND 3.0

State versus event sequences: examples

Time stamped events

Sandra	Ending education in 1980	Start working in 1980
Jack	Ending education in 1981	Start working in 1982

- There can be simultaneous events (see Sandra)
- Elements at same position do not occur at same time

State sequence view

year	1979	1980	1981	1982	1983
Sandra	Education	Education	Employed	Employed	Employed
Jack	Education	Education	Education	Unemployed	Employed

- Only one state at each observed time
- Position conveys time information: All states at position 2 are states in 1980.

Outline

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - . . .

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - ...

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - ...

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - ...

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - ...

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - ...

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - ...

Typical questions in social sciences

- In the field of Life course analysis
  - How can we measure standardization?
  - Are there standards of life, ideal-types?
  - What are those standards, those ideal-types?
  - How are those standards linked to covariates such as sex, birth cohort, ... ?
  - More generally, how are life trajectories linked to demographic and/or socioeconomic variables?
  - How do current social statuses depend on the lived trajectories?
  - ...

Sequencing, timing and duration

- For chronological sequences (with time dimension)
- SA can answer questions about:
  - Sequencing: Order in which the different elements occur.
  - Timing: When do the different elements occur?
  - Duration: How long do we stay in the successive states?

Sequencing, timing and duration

- For chronological sequences (with time dimension)
- SA can answer questions about:
  - Sequencing: Order in which the different elements occur.
  - Timing: When do the different elements occur?
  - Duration: How long do we stay in the successive states?

© G. Ritschard (2012), 25/67. Distributed under licence CC BY-NC-ND 3.0

Sequencing, timing and duration

- For chronological sequences (with time dimension)
- SA can answer questions about:
  - Sequencing: Order in which the different elements occur.
  - Timing: When do the different elements occur?
  - Duration: How long do we stay in the successive states?

 Distributed under licence CC BY-NC-ND 3.0

Sequencing, timing and duration

- For chronological sequences (with time dimension)
- SA can answer questions about:
  - Sequencing: Order in which the different elements occur.
  - Timing: When do the different elements occur?
  - Duration: How long do we stay in the successive states?

 Distributed under licence CC BY-NC-ND 3.0

Sequencing, timing and duration

- For chronological sequences (with time dimension)
- SA can answer questions about:
  - Sequencing: Order in which the different elements occur.
  - Timing: When do the different elements occur?
  - Duration: How long do we stay in the successive states?

Outline

 Distributed under licence CC BY-NC-ND 3.0

Starting TraMineR

Creating occupational sequence object

- Reading SPSS data file and preparing labels

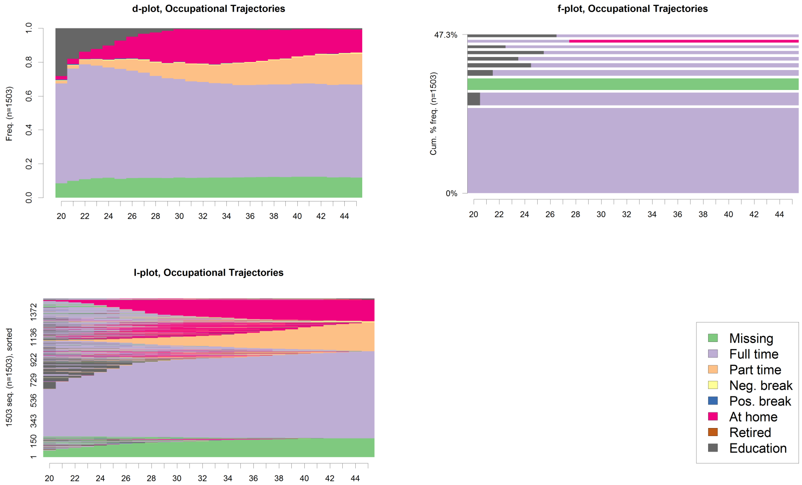
```
R> library(foreign)
R> seqs <- read.spss(file = paste(readir, "SHPbio-w.sav", sep = ""),
       to.data.frame = T)
R> labels.occ <- c("Missing", "Full time", "Part time", "Neg. break",
       "Pos. break", "At home", "Retired", "Education")
R> short.labels.occ <- c("Mi", "FT", "PT", "NB", "PB", "AH",
       "RE", "ED")
R> xtlab20 <- seq(20, 45)
```

- Loading TraMiner and creating a state sequence object

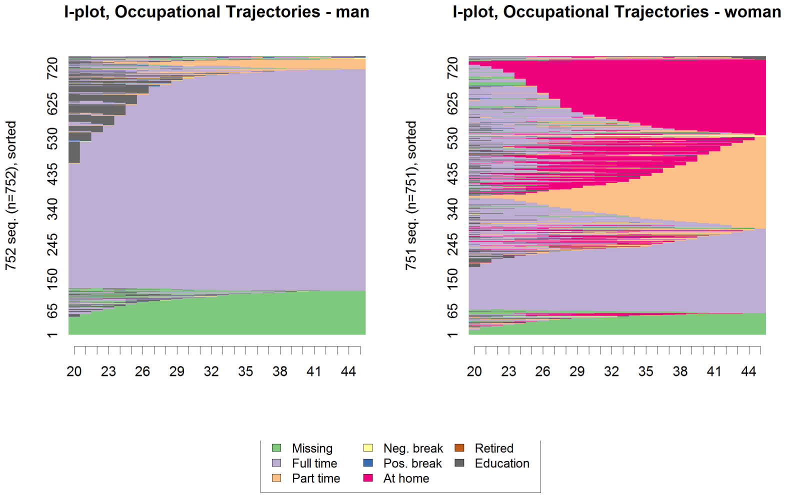
```
R> library(TraMineR)
R> seqs.occ <- seqdef(seqs[, 4:29], states = short.labels.occ,
       labels = labels.occ, cnames = xtlab20)
```


© G. Ritschard (2012), 27/67. Distributed under licence CC BY-NC-ND 3.0

Rendering sequences



Rendering sequences by group (sex)



Characterizing set of sequences

- Sequence of transversal measures (modal state, between entropy, ...)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

- Summary of longitudinal measures (within entropy, transition rates, mean duration ...)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

- Other global characteristics: sequence medoid, diversity of sequences, ...

Characterizing set of sequences

- Sequence of transversal measures (modal state, between entropy, ...)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

- Summary of longitudinal measures (within entropy, transition rates, mean duration ...)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

- Other global characteristics: sequence medoid, diversity of sequences, ...

 Distributed under licence CC BY-NC-ND 3.0

Characterizing set of sequences

- Sequence of transversal measures (modal state, between entropy, ...)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

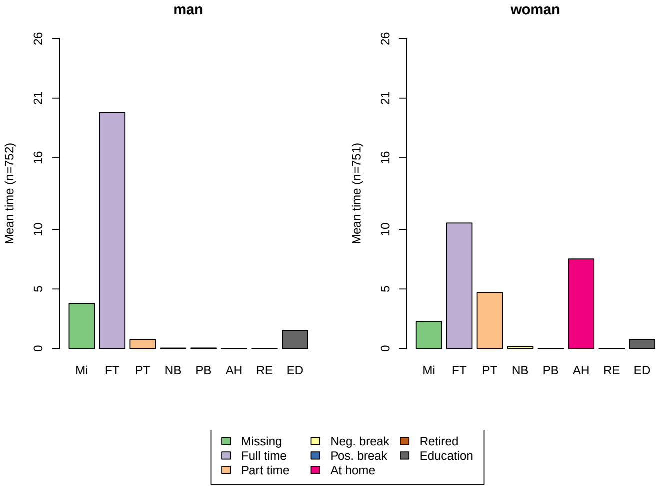
- Summary of longitudinal measures (within entropy, transition rates, mean duration ...)

id	t_1	t_2	t_3	$\cdots$
1	B	B	D	$\cdots$
2	A	B	C	$\cdots$
3	B	B	A	$\cdots$

- Other global characteristics: sequence medoid, diversity of sequences, ...

Mean time in each state

```
R> seqmtplot(seqs.occ, group = seqs$sex)
```

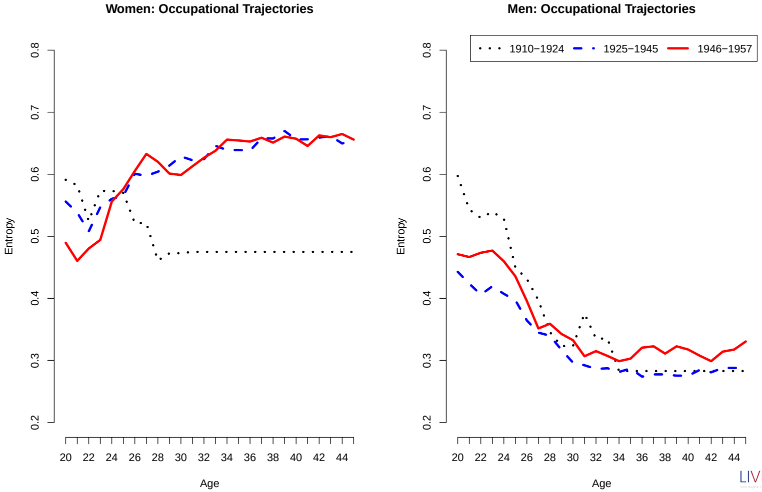


Transition rates

	[-> Mi]	[-> FT]	[-> PT]	[-> NB]	[-> PB]	[-> AH]	[-> RE]	[-> ED]
[Mi ->]	0.969	0.005	0.004	0.001	0.001	0.011	0.000	0.008
[FT ->]	0.003	0.971	0.009	0.001	0.001	0.013	0.000	0.003
[PT ->]	0.005	0.026	0.939	0.001	0.001	0.018	0.000	0.010
[NB ->]	0.040	0.047	0.027	0.880	0.000	0.007	0.000	0.000
[PB ->]	0.105	0.316	0.105	0.000	0.404	0.018	0.000	0.053
[AH ->]	0.003	0.007	0.032	0.000	0.000	0.956	0.000	0.002
[RE ->]	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
[ED ->]	0.044	0.236	0.045	0.001	0.002	0.006	0.000	0.664

Heterogeneity: Sequence of transversal entropies

Occupational, Women vs Men



Number of state transitions (longitudinal)

Pairwise dissimilarities between sequences

- Distance between sequences
  - Different metrics (LCP, LCS, OM, HAM, DHD)
- Once we have pairwise dissimilarities, we can
  - Partition a set of sequences into homogeneous clusters

  - Identify representative sequences (medoid, densest neighborhood)
  - Measure the discrepancy between sequences
  - Run self-organizing maps (SOM) on sequences
  - MDS scatterplot representation of sequences
  - Discrepancy analysis of a set of sequences (ANOVA)
  - Grow regression trees for explaining the sequence discrepancy

Pairwise dissimilarities between sequences

- Distance between sequences
  - Different metrics (LCP, LCS, OM, HAM, DHD)
- Once we have pairwise dissimilarities, we can
  - Partition a set of sequences into homogeneous clusters
  - Identify representative sequences (medoid, densest neighborhood)
  - Measure the discrepancy between sequences
  - Run self-organizing maps (SOM) on sequences
  - MDS scatterplot representation of sequences
  - Discrepancy analysis of a set of sequences (ANOVA)
  - Grow regression trees for explaining the sequence discrepancy

 Distributed under licence CC BY-NC-ND 3.0

Pairwise dissimilarities between sequences

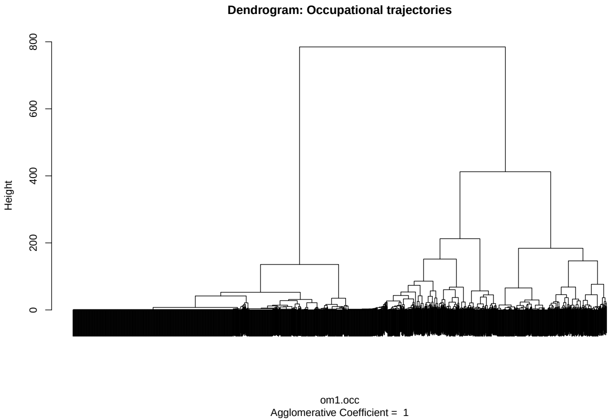
- Distance between sequences
  - Different metrics (LCP, LCS, OM, HAM, DHD)
- Once we have pairwise dissimilarities, we can
  - Partition a set of sequences into homogeneous clusters
  - Identify representative sequences (medoid, densest neighborhood)
  - Measure the discrepancy between sequences
  - Run self-organizing maps (SOM) on sequences
  - MDS scatterplot representation of sequences
  - Discrepancy analysis of a set of sequences (ANOVA)
  - Grow regression trees for explaining the sequence discrepancy

 Distributed under licence CC BY-NC-ND 3.0

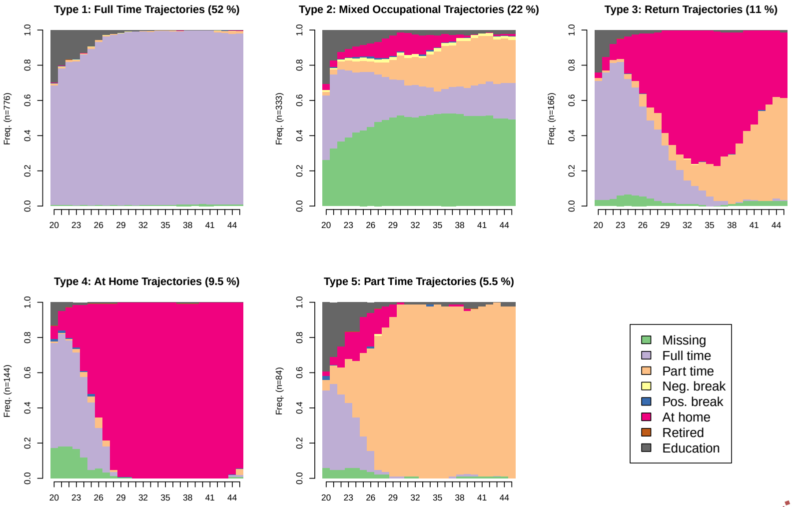
Dissimilarity matrix

```
R> print(seqs.occ[1:4, ], format = "SPS")

    Sequence
[1] (FT,26)
[2] (FT,26)
[3] (Mi,6)-(ED,3)-(Mi,17)
[4] (ED,1)-(Mi,3)-(PT,4)-(FT,18)

R> dm <- seqdist(seqs.occ[1:4, ], method = "LCS")
R> dm[1:4, 1:4]

     [,1] [,2] [,3] [,4]
[1,]    0    0   52   16
[2,]    0    0   52   16
[3,]   52   52    0   44
[4,]   16   16   44    0
```

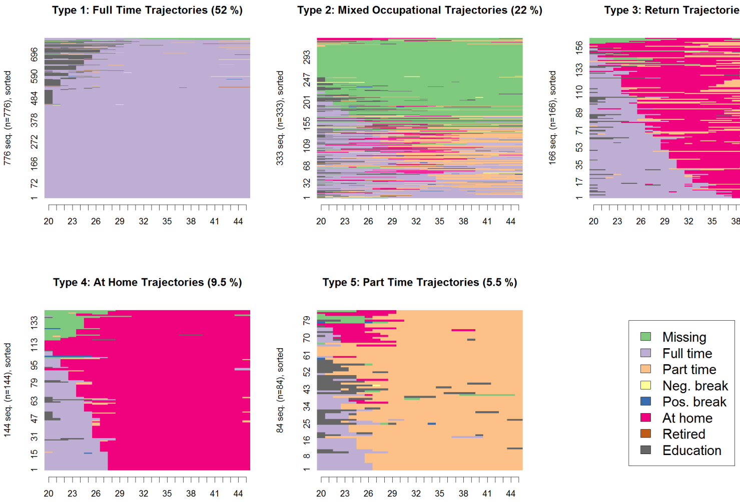
Cluster analysis: determining typologies



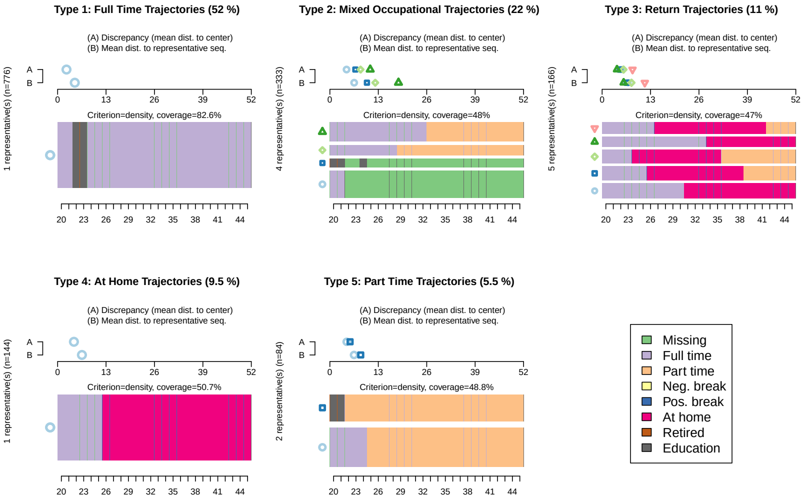
Cluster analysis: determining typologies



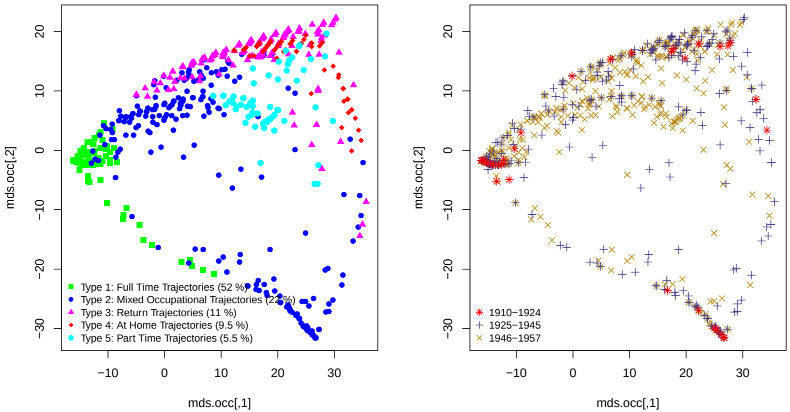
Cluster analysis: i-plots (sorted by 1st MDS factor)



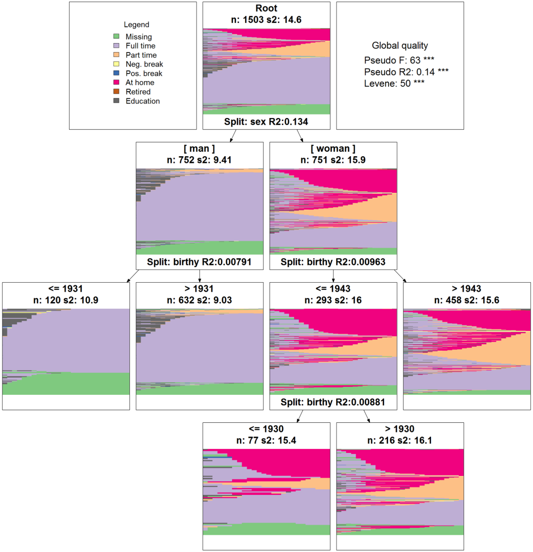
Cluster analysis: representative sequences



MDS: Scatterplot view of sequences



Regression tree



Event sequences

- Instead of the successive states, we may consider the transitions between states and more specifically the—possibly simultaneous—events that provoke the transitions.
- Event sequences are more difficult to render because they have no duration!
- Event sequences are of interest for studying the sequencing
  - What are the typical sequencing of life events?
  - Which event sequencing distinguishes men and women? younger and older cohorts?

 Distributed under licence CC BY-NC-ND 3.0

Event sequences

- Instead of the successive states, we may consider the transitions between states and more specifically the—possibly simultaneous—events that provoke the transitions.
- Event sequences are more difficult to render because they have no duration!
- Event sequences are of interest for studying the sequencing
  - What are the typical sequencing of life events?
  - Which event sequencing distinguishes men and women? younger and older cohorts?

Event sequences

- Instead of the successive states, we may consider the transitions between states and more specifically the—possibly simultaneous—events that provoke the transitions.

- Event sequences are more difficult to render because they have no duration!
- Event sequences are of interest for studying the sequencing
  - What are the typical sequencing of life events?
  - Which event sequencing distinguishes men and women? younger and older cohorts?

Event sequences

- Instead of the successive states, we may consider the transitions between states and more specifically the—possibly simultaneous—events that provoke the transitions.

- Event sequences are more difficult to render because they have no duration!
- Event sequences are of interest for studying the sequencing
  - What are the typical sequencing of life events?
  - Which event sequencing distinguishes men and women? younger and older cohorts?

LIVES UNIVERSITÉ DE GENÈVE FACULTÉ DES SCIENCES ÉCONOMIQUES ET SOCIALES

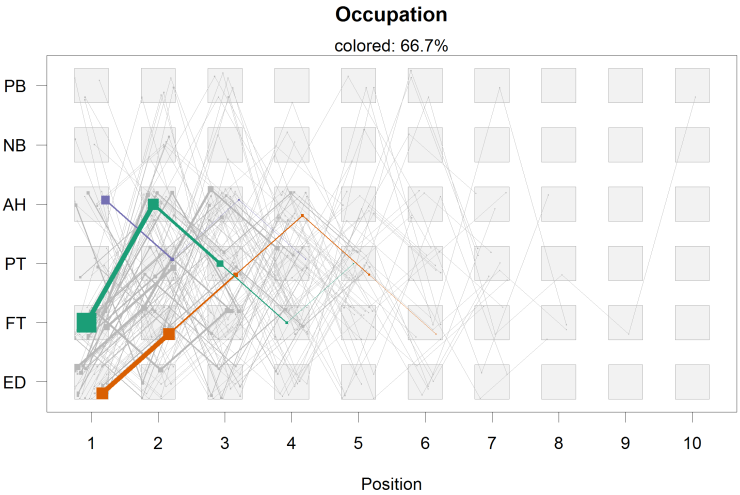
Event sequences

- Instead of the successive states, we may consider the transitions between states and more specifically the—possibly simultaneous—events that provoke the transitions.

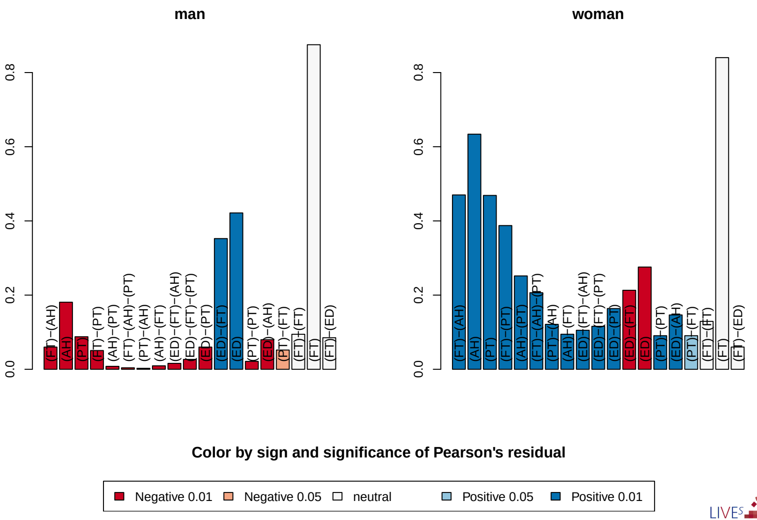
- Event sequences are more difficult to render because they have no duration!
- Event sequences are of interest for studying the sequencing
  - What are the typical sequencing of life events?
  - Which event sequencing distinguishes men and women? younger and older cohorts?

 Distributed under licence CC BY-NC-ND 3.0

Rendering event sequences



Event sequences: discriminating sub-sequences



What you will not find in this course ...

- Transition analysis by means of Markovian and other statistical models.
- for Markovian models, see for instance Berchtold and Raftery (2002)

- Survival analysis
- e.g. Hosmer and Lemeshow (1999), Hothorn et al. (2006)

© G. Ritschard (2012), 46/67. Distributed under licence CC BY-NC-ND 3.0

What you will not find in this course ...

- Transition analysis by means of Markovian and other statistical models.
- for Markovian models, see for instance Berchtold and Raftery (2002)

- Survival analysis
- e.g. Hosmer and Lemeshow (1999), Hothorn et al. (2006)

Outline

Section outline

6 TraMineR
- About TraMineR
- A first run

 Distributed under licence CC BY-NC-ND 3.0

TraMineR: What is it?

TraMineR

- Trajectory Miner in R: a toolbox for exploring, rendering and analyzing categorical sequence data

- Developed within the SNF (Swiss National Fund for Scientific Research) project Mining event histories 1/2007-1/2011
- ... development goes on within IP 14 methodological module of the NCCR LIVES: Overcoming vulnerability: Life course perspectives (http://www.lives-nccr.ch) .

 Distributed under licence CC BY-NC-ND 3.0

TraMineR, Who?

- Under supervision of a scientific committee:
  - Gilbert Ritschard (Statistics for social sciences)
  - Alexis Gabadinho (Demography)
  - Nicolas S. Müller (Sociology, Computer science)
  - Matthias Studer (Economics, Sociology)
- Additional members of the development team:
  - Reto Bürgin (Statistics)
  - Emmanuel Rousseaux (KDD and Computer science)

both PhD students within NCCR LIVES IP-14

TraMineR, Why?

- TraMineR primary aim: Answer questions from social sciences
  - where sequences (succession of states or events) describe life trajectories
- Examples of questions:
  - Do life courses obey some social norm?
    - Which are the standard trajectories?
    - What kind of departures do we observe from those standards?
    - How do life course patterns evolve over time?
  - Why are some people more at risk to follow a chaotic trajectory or stay stuck in a state?
    - How does the trajectory complexity evolve across birth cohorts?
  - How is the life trajectory related to sex, social origin and other cultural factors?

© G. Ritschard (2012), 51/67. Distributed under licence CC BY-NC-ND 3.0

What TraMineR offers to answer those questions

- Various graphics and descriptive measures of individual sequences.
- Tools for computing pairwise dissimilarities between sequences which open access to plenty of advanced statistical and data analysis tools
  - Clustering and principal coordinate analysis (MDS)
  - Discrepancy analysis (ANOVA and regression trees)
  - Identification of representative sequences (trajectory-types)
  - ...

 Distributed under licence CC BY-NC-ND 3.0

What TraMineR offers to answer those questions

- Various graphics and descriptive measures of individual sequences.
- Tools for computing pairwise dissimilarities between sequences which open access to plenty of advanced statistical and data analysis tools
  - Clustering and principal coordinate analysis (MDS)
  - Discrepancy analysis (ANOVA and regression trees)
  - Identification of representative sequences (trajectory-types)
  - ...

What TraMineR offers to answer those questions

- Various graphics and descriptive measures of individual sequences.
- Tools for computing pairwise dissimilarities between sequences which open access to plenty of advanced statistical and data analysis tools
  - Clustering and principal coordinate analysis (MDS)
  - Discrepancy analysis (ANOVA and regression trees)
  - Identification of representative sequences (trajectory-types)
  - ...

What TraMineR offers to answer those questions

- Various graphics and descriptive measures of individual sequences.
- Tools for computing pairwise dissimilarities between sequences which open access to plenty of advanced statistical and data analysis tools
  - Clustering and principal coordinate analysis (MDS)
  - Discrepancy analysis (ANOVA and regression trees)
  - Identification of representative sequences (trajectory-types)
  - ...

© G. Ritschard (2012), 52/67. Distributed under licence CC BY-NC-ND 3.0

What TraMineR offers to answer those questions

- Various graphics and descriptive measures of individual sequences.
- Tools for computing pairwise dissimilarities between sequences which open access to plenty of advanced statistical and data analysis tools
  - Clustering and principal coordinate analysis (MDS)
  - Discrepancy analysis (ANOVA and regression trees)
  - Identification of representative sequences (trajectory-types)
  - ...

What TraMineR offers to answer those questions

- Various graphics and descriptive measures of individual sequences.
- Tools for computing pairwise dissimilarities between sequences which open access to plenty of advanced statistical and data analysis tools
    - Clustering and principal coordinate analysis (MDS)
    - Discrepancy analysis (ANOVA and regression trees)
    - Identification of representative sequences (trajectory-types)
    - ...

TraMineR: Where and why in R?

- Package for the free open source R statistical environment
  - freely available on the CRAN (Comprehensive R Archive Network) http://cran.r-project.org

```
R> install.packages("TraMineR", dependencies=TRUE)
```

- TraMineR runs in R, it can straightforwardly be combined with other R commands and libraries. For example:
  - dissimilarities obtained with TraMineR can be inputted to already optimized processes for clustering, MDS, self-organizing maps, ...
  - TraMineR 's plots can be used to render clustering results;
  - complexity indexes can be used as dependent or explanatory variables in linear and non-linear regression, ...

 Distributed under licence CC BY-NC-ND 3.0

TraMineR: Where and why in R?

- Package for the free open source R statistical environment
  - freely available on the CRAN (Comprehensive R Archive Network) http://cran.r-project.org

```
R> install.packages("TraMineR", dependencies=TRUE)
```

- TraMineR runs in R, it can straightforwardly be combined with other R commands and libraries. For example:
  - dissimilarities obtained with TraMineR can be inputted to already optimized processes for clustering, MDS, self-organizing maps, ...
  - TraMineR 's plots can be used to render clustering results;
  - complexity indexes can be used as dependent or explanatory variables in linear and non-linear regression, ...

TraMineR's features

- Handling of longitudinal data and conversion between various sequence formats
- Plotting sequences (distribution plot, frequency plot, index plot and more)
- Individual longitudinal characteristics of sequences (length, time in each state, longitudinal entropy, turbulence, complexity and more)
- Sequence of transversal characteristics by position (transversal state distribution, transversal entropy, modal state)
- Other aggregated characteristics (transition rates, average duration in each state, sequence frequency)
- Dissimilarities between pairs of sequences (Optimal matching, Longest common subsequence, Hamming, Dynamic Hamming, Multichannel and more)
- Representative sequences and discrepancy measure of a set of sequences
- ANOVA-like analysis and regression tree of sequences
- Rendering and highlighting frequent event sequences
- Extracting frequent event subsequences
- Identifying most discriminating event subsequences
- Association rules between subsequences

 Distributed under licence CC BY-NC-ND 3.0

Other programs for sequence analysis

- Optimize (Abbott, 1997)
  - Computes optimal matching distances
  - No longer supported
- TDA (Rohwer and Pötter, 2002)
  - free statistical software, computes optimal matching distances
- Stata, SQ-Ados (Brzinsky-Fay et al., 2006)
  - free, but licence required for Stata
  - optimal matching distances, visualization and a few more
  - See also the add-ons by Brenda Halpin http://teaching.sociology.ul.ie/seqanal/
- CHESA free program by Elzinga (2007)
  - Various metrics, including original ones based on non-aligning methods
  - Turbulence

 Distributed under licence CC BY-NC-ND 3.0

Section outline

6 TraMineR
- About TraMineR
- A first run

Loading the library and example data set

- Loading the library TraMineR, accessing the `mvad` dataset

```
R> library(TraMineR)
R> data(mvad)
```

- In `mvad` the sequence information starts in column 15 and ends at column 76. Here we display selected columns for the first two cases:

```
R> mvad[1:2, 14:17]

  livboth      Jul.93      Aug.93     Sep.93
1     yes    training    training employment
2     yes joblessness joblessness         FE

R> mvad[1:2, 73:76]

      May.98     Jun.98     Jul.98     Aug.98
1 employment employment employment employment
2         HE         HE         HE         HE
```

© G. Ritschard (2012), 57/67. Distributed under licence CC BY-NC-ND 3.0

Creating the state sequence object

- Provide the subset of the data frame mvad containing the sequence information

```
R> mvad.seq <- seqdef(mvad[, 15:76])
```

- Display the first two sequences in mvad.seq

```
R> mvad.seq[1:2, ]

  Sequence
1 training-training-employment-employment-employment-employment-training-
2 joblessness-joblessness-FE-FE-FE-FE-FE-FE-FE-FE-FE-FE-FE-FE-FE-FE-FE-FE
```

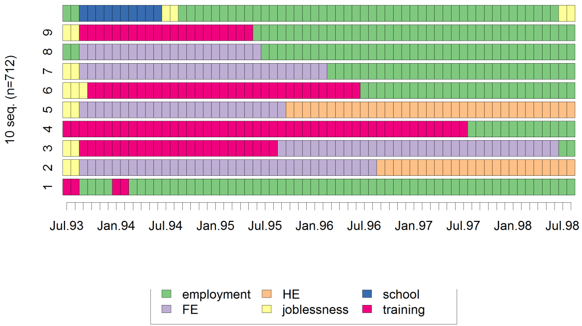
- Display the first two sequences in mvad.seq in compact form

```
R> print(mvad.seq[1:2, ], format = "SPS")

    Sequence
[1] (training,2)-(employment,4)-(training,2)-(employment,54)
[2] (joblessness,2)-(FE,36)-(HE,24)
```

Rendering the sequences

- First ten sequences

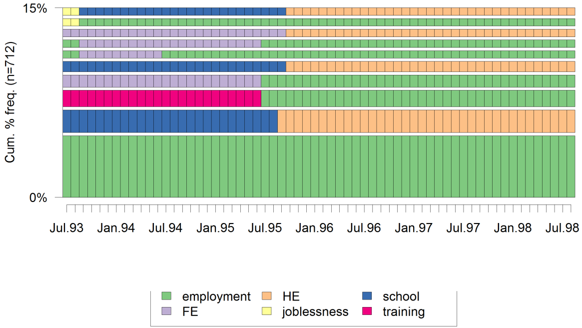
```
R> seqiplot(mvad.seq)
```



Rendering the sequences

- Ten most frequent

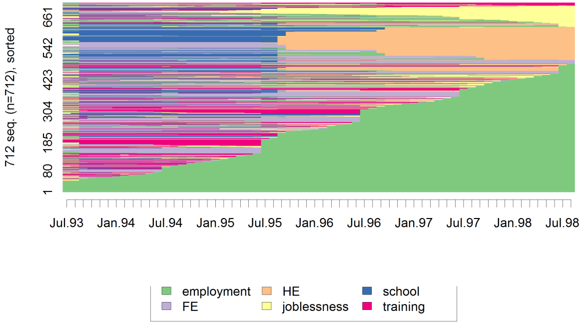
```
R> seqfplot(mvad.seq)
```



Rendering the sequences

- All sequences

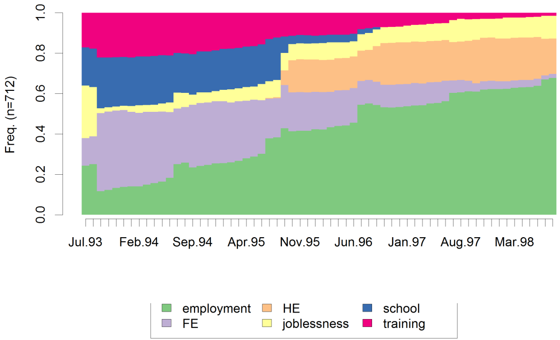
```
R> seqIplot(mvad.seq, sortv = "from.end")
```



Rendering the sequences

- Sequence of transversal distributions (chronogram)

```
R> seqdplot(mvad.seq, border = NA)
```



 Distributed under licence CC BY-NC-ND 3.0

Thank you!
See you next week.

References I

Abbott, A. (1997). Optimize. http://home.uchicago.edu/~aabbott/om.html.

Beck, N. and J. N. Katz (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review 89*, 634–647.

Bejerano, G. and G. Yona (2001). Variations on probabilistic suffix trees: statistical modeling and prediction of protein families. *Bioinformatics 17*(1), 23–43.

Berchtold, A. and A. E. Raftery (2002). The mixture transition distribution model for high-order Markov chains and non-gaussian time series. *Statistical Science 17*(3), 328–356.

Billari, F. C. (2001). The analysis of early life courses: Complex description of the transition to adulthood. *Journal of Population Research 18*(2), 119–142.

Brzinsky-Fay, C., U. Kohler, and M. Luniak (2006). Sequence analysis with Stata. *The Stata Journal 6*(4), 435–460.

Elzinga, C. H. (2007). CHESA 2.1 User manual. User guide, Dept of Social Science Research Methods, Vrije Universiteit, Amsterdam.

References II

Frees, E. W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. New York: Cambridge University Press.

Gabadinho, A., G. Ritschard, N. S. Müller, and M. Studer (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software 40*(4), 1–37.

Gabadinho, A., G. Ritschard, M. Studer, and N. S. Müller (2009). Mining sequence data in R with the TraMineR package: A user's guide. Technical report, Department of Econometrics and Laboratory of Demography, University of Geneva, Geneva.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Hedeker, D. (2007). Multilevel models for ordinal and nominal variables. In J. de Leeuw and E. Meijer (Eds.), *Multilevel Models for Ordinal and Nominal Variables*, Chapter 6, pp. 239–276. Springer.

Hosmer, D. W. and S. Lemeshow (1999). *Applied Survival Analysis, Regression Modeling of Time to Event Data*. New York: John Wiley & Sons.

 Distributed under licence CC BY-NC-ND 3.0

References III

Hothorn, T., K. Hornik, and A. Zeileis (2006). party: A laboratory for recursive part(y)itioning. User's manual.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology 60*, 577–605.

Müller, N. S. (2011). *Inégalités sociales et effets cumulés au cours de la vie: concepts et méthodes*, Volume SES-764 of *Collection des thèses*. Université de Genève, Faculté des sciences économiques et sociales.

Perroux, O. et M. Oris (2005). Présentation de la base de données de la population de Genève de 1816 à 1843. Séminaire statistique sciences sociales, Université de Genève.

Ritschard, G., A. Gabadinho, N. S. Müller, and M. Studer (2008). Mining event histories: A social science perspective. *International Journal of Data Mining, Modelling and Management 1*(1), 68–90.

Rohwer, G. and U. Pötter (2002). TDA user's manual. Software, Ruhr-Universität Bochum, Fakultät für Sozialwissenschaften, Bochum.

Sharma, K. R. (2008). *Bioinformatics – Sequence Alignment and Markov Models*. New York: McGraw-Hill.

© G. Ritschard (2012), 66/67. Distributed under licence CC BY-NC-ND 3.0

References IV

Singer, J. D. and J. B. Willett (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.

Therneau, T. M. and P. M. Grambsch (2000). *Modeling Survival Data*. New York: Springer.

Wanner, P. et E. Delaporte (2001). Reconstitution de trajectoires de vie à partir des données de l'état civil (BEVNAT). une étude de faisabilité. Rapport de recherche, Forum Suisse des Migrations.

Wernli, B. (2010). A Swiss survey landscape for communication research. In *Università della Svizzera Italiana, USI, Lugano, 2010, June 15, Institute of Communication and Health*.

Widmer, E. and G. Ritschard (2009). The de-standardization of the life course: Are men and women equal? *Advances in Life Course Research 14*(1-2), 28–39.

© G. Ritschard (2012), 67/67. Distributed under licence CC BY-NC-ND 3.0