UNIVERSITY OF MICHIGAN

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Temporal and sequential analysis for learning analytics

Session 1A: Monday, June 21, 2021, 2-4pm PDT
Session 1B: Wednesday, June 23, 2021, 1-3pm PDT
Session 1C: Friday, June 25, 2021,1-3pm PDT

Please say hello and introduce yourself in the Zoom chat! ☺

Quan Nguyen, Ph.D.
Postdoctoral Fellow
School of Information

Email: quanngu@umich.edu
Twitter: @QuanNguyen3010

Agenda

W8. Temporal and sequential analysis for learning analytics ***Quan Nguyen – University of Michigan*** **Zoom**: https://us02web.zoom.us/j/85872971310 **Slack**: Invitation by email **R packages**: TraMineR, arules		
Session	**Activity**	**Resources**
1A: Monday, June 21, 2021 from 2pm to 4pm PDT	**Group discussion**: Overview of temporal analysis techniques in LA and discussion on the type of RQs and learning constructs that are suitable for temporal analysis	• Knight, S., Friend Wise, A., & Chen, B. (2017). Time for Change: Why Learning Analytics Needs Temporal Analysis. *Journal of Learning Analytics*, 4(3), 7–17. • Chen, B., Knight, S., & Wise, A. F. (2018). Critical Issues in Designing and Implementing Temporal Analytics. *Journal of Learning Analytics*, 5(1), 1–9.
	Tutorial: Exploratory data analysis for states sequences data (visualization, basic descriptive statistics for sequences), using package TraMineR	• 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 Mining sequence data in R with the TraMineR package: A user's guide. University of Geneva, 2010
1B: Wednesday, June 23, 2021 from 1pm to 3pm PDT	**Tutorial**: How can we detect common learning patterns from students' log data using sequential analysis? (Sequence similarities, optimal matching, clustering sequences)	
1C: Friday, June 25, 2021 from 1pm to 3pm PDT	**Tutorial**: How can we identify courses that are frequently taken together and provide course recommendations? (Association rule mining with apriori algorithm)	https://www.kirenz.com/post/2020-05-14-r-association-rule-mining/

Note: Slides & codes will be posted on https://github.com/quan3010/temporal_analysis

Group discussion

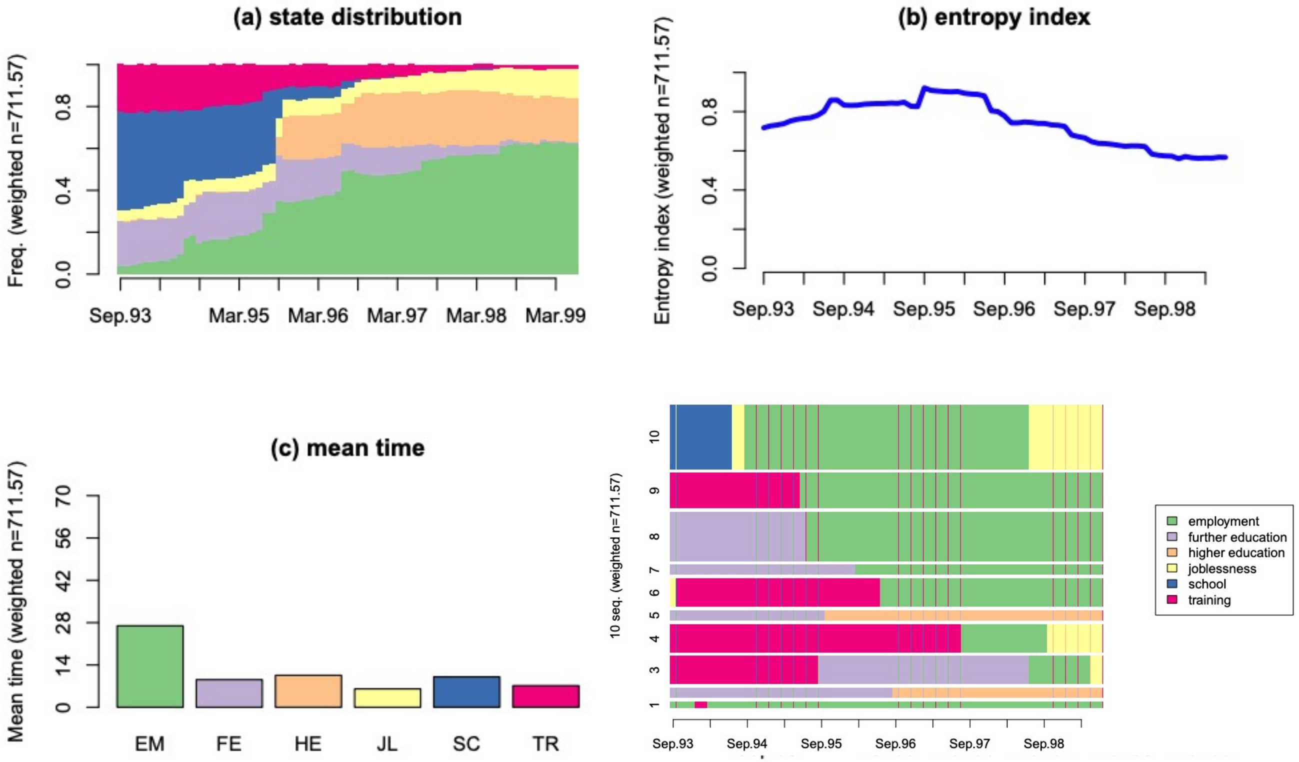
Data source	Time granularity	RQs/Learning constructs	Analysis technique
SIS/Canvas		Changes in behaviors, understanding, or knowledge	Regression MLM ANOVA Recurrence quantification Markov Growth MLM CNN transition diagnostic classification models
CLickstream		SRL calibration accuracy	
Eye-tracking		Dynamics of thinking (thought types) during learning sessions	
Mobile app		writing flow from keystroke data	
Keystroke		What is the path learners take in an online course?	
Location data		changes of learning behavior due to covid-19	
Assessment		forecasting dropout and fail risk	
Writing		Timie management	
		Time on task	
		Course design	
		Location	

Overview

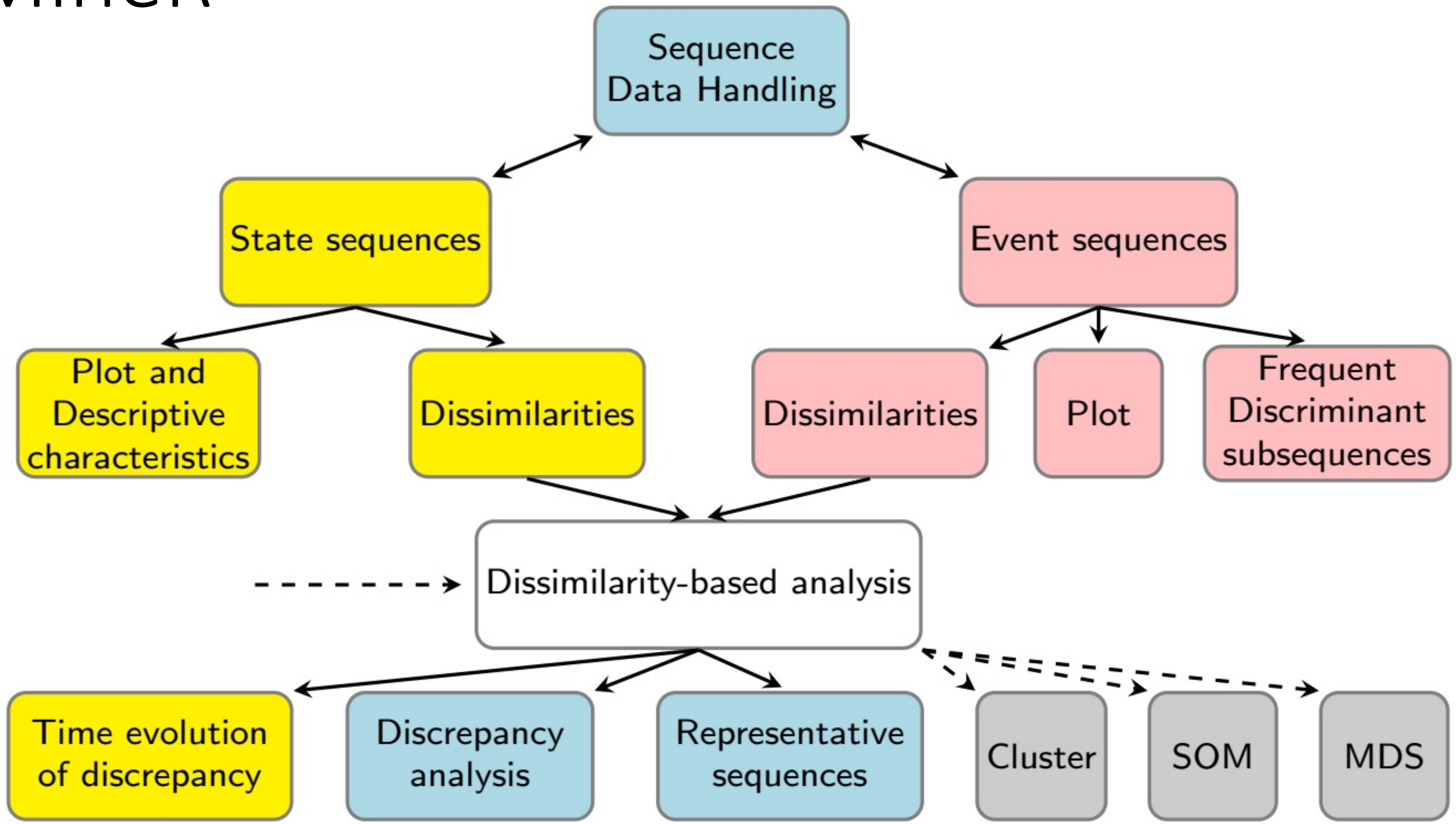
- Sequence analysis → Learning strategies, tactics → cluster by performance
- Process mining → transitional probabilities between activities
- Time-series analysis (ARIMA and the like) → I haven't seen much application in LA, partly because of the data type (panel data instead of univariate time-series)
- ML (LSTM, RNN) → Predict learner's outcome such as dropout on a weekly basis
- Temporal network analysis → SOAMS, REM, ENA
- Theory-driven → procrastination, spaced learning, SRL phases, etc…

TraMineR

- A toolbox for exploring, rendering and analyzing categorical sequence data
- Visualize
- Descriptive stats
- Clustering
- Association rules



TraMineR



Sequence Data Handling
State sequences
Event sequences
Plot and Descriptive characteristics
Dissimilarities
Dissimilarities
Plot
Frequent Discriminant subsequences
Dissimilarity-based analysis
Time evolution of discrepancy
Discrepancy analysis
Representative sequences
Cluster
SOM
MDS

TraMineR

- 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

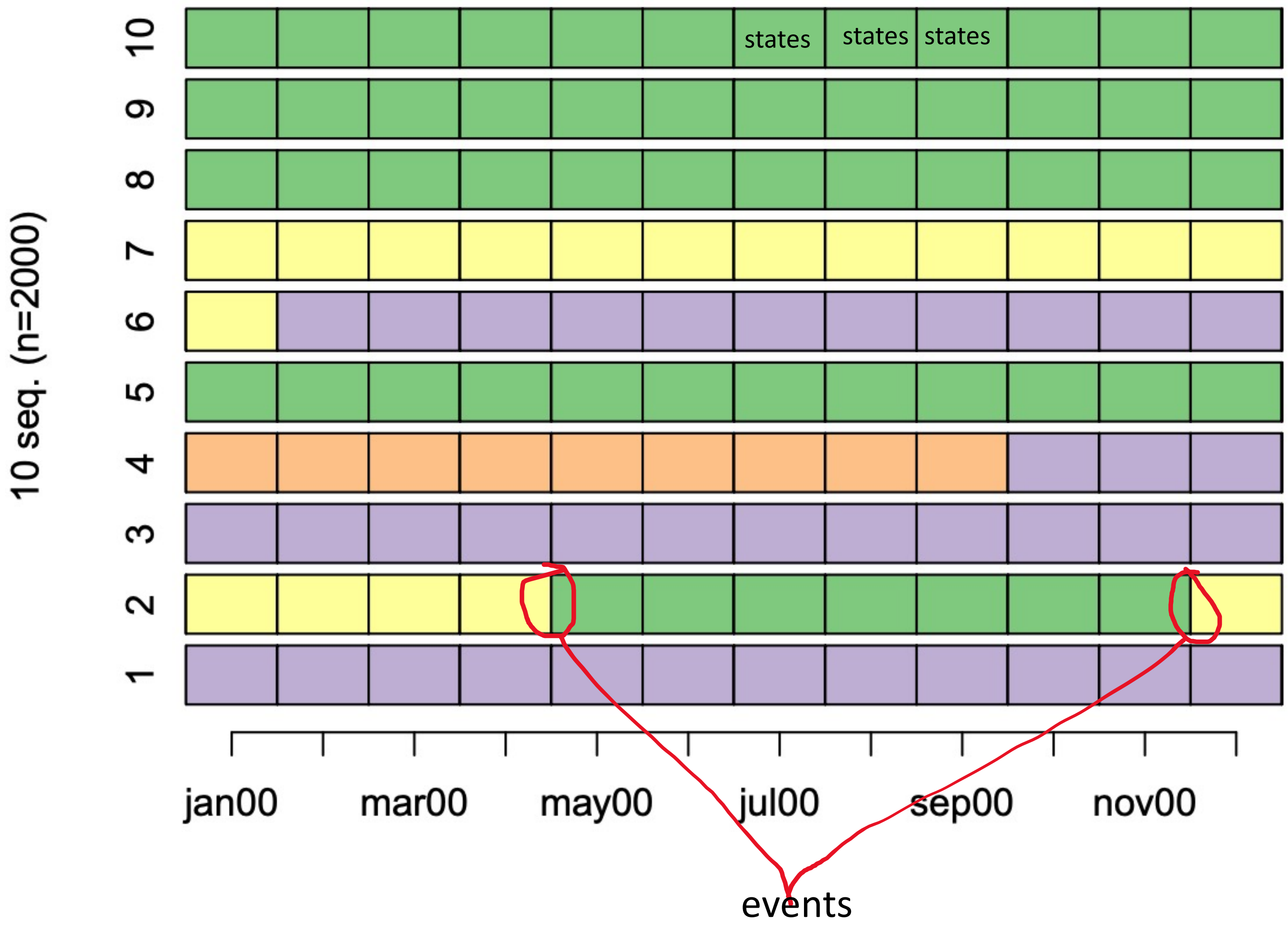
TraMineR – mvad data

- Life-trajectories

Table 3.5: List of Variables in the *MVAD* data set

id	unique individual identifier
weight	sample weights
male	binary dummy for gender, 1=male
catholic	binary dummy for community, 1=Catholic
Belfast	binary dummies for location of school, one of five Education and Library Board areas in Northern Ireland
N.Eastern	"
Southern	"
S.Eastern	"
Western	"
Grammar	binary dummy indicating type of secondary education, 1=grammar school
funemp	binary dummy indicating father's employment status at time of survey, 1=father unemployed
gcse5eq	binary dummy indicating qualifications gained by the end of compulsory education, 1=5+ GCSEs at grades A-C, or equivalent
fmpr	binary dummy indicating SOC code of father?s current or most recent job,1=SOC1 (professional, managerial or related)
livboth	binary dummy indicating living arrangements at time of first sweep of survey (June 1995), 1=living with both parents
jul93	Monthly Activity Variables are coded 1-6, 1=school, 2=FE, 3=employment, 4=training, 5=joblessness, 6=HE
:	"
jun99	"

States vs events



Define states sequences

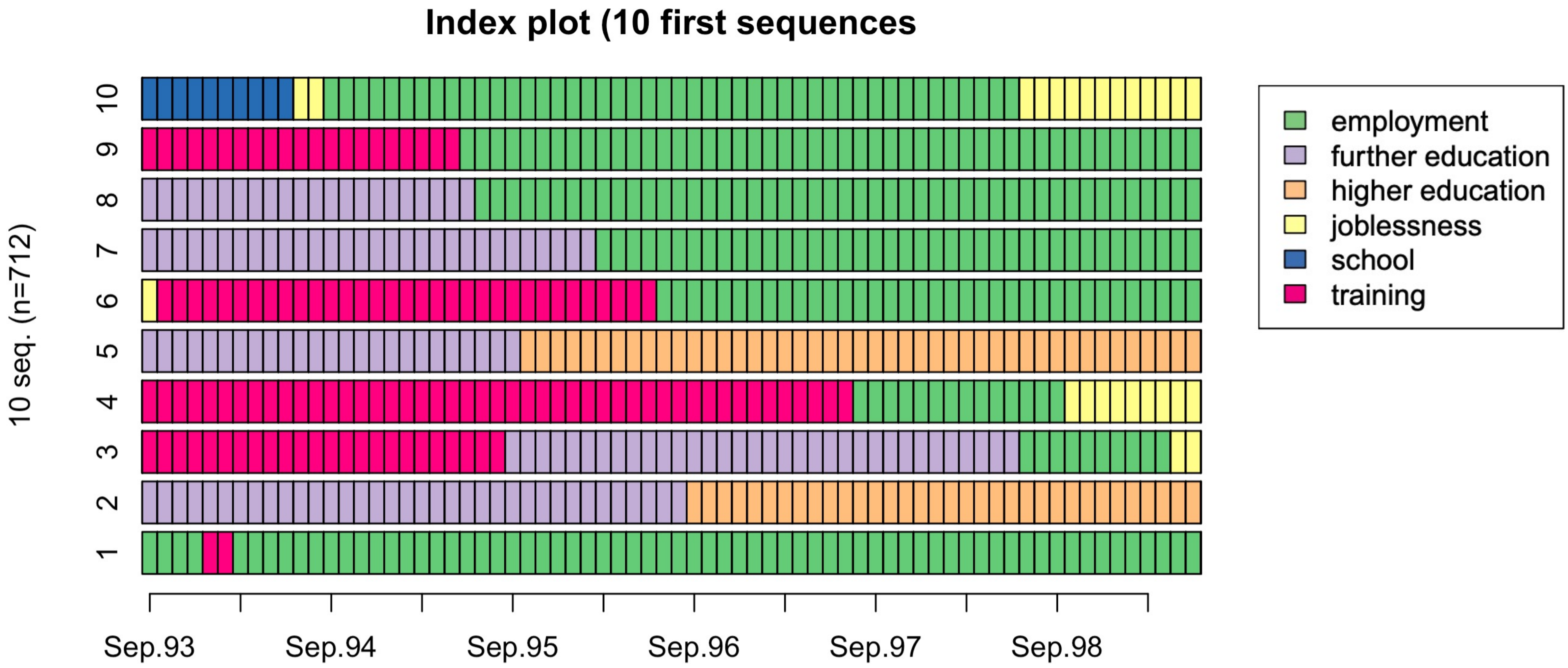
View(mvad.seq)

Sep.93	Oct.93	Nov.93	Dec.93	Jan.94	Feb.94
EM	EM	EM	EM	TR	TR
FE	FE	FE	FE	FE	FE
TR	TR	TR	TR	TR	TR
TR	TR	TR	TR	TR	TR
FE	FE	FE	FE	FE	FE
JL	TR	TR	TR	TR	TR
FE	FE	FE	FE	FE	FE
FE	FE	FE	FE	FE	FE
TR	TR	TR	TR	TR	TR
SC	SC	SC	SC	SC	SC
FE	FE	FE	FE	FE	FE
SC	SC	SC	SC	SC	SC
SC	SC	SC	SC	SC	SC

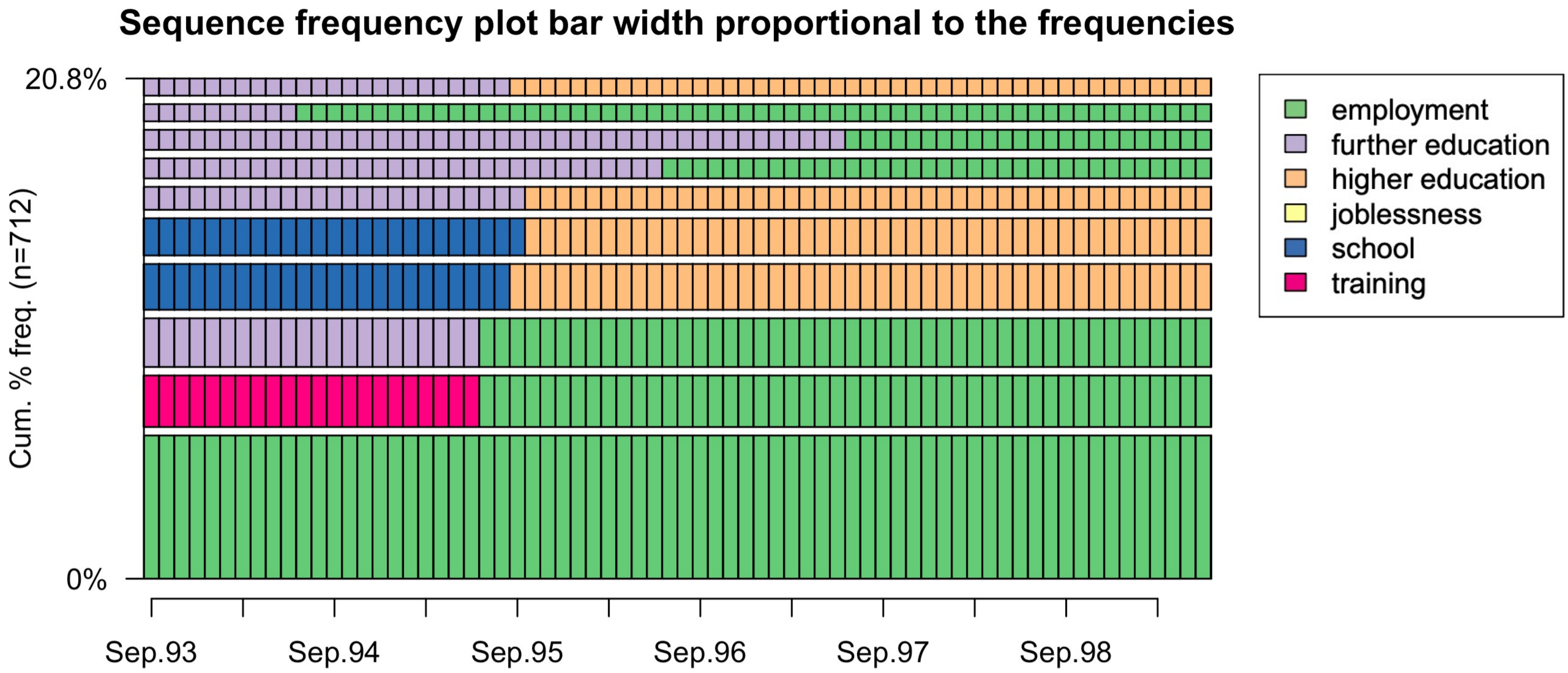
Table 4.2: Sequence data representations: Examples

Code	Example										
STS	**Id**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
	101	S	S	S	M	M	MC	MC	MC	MC	D
	102	S	S	S	MC	MC	MC	MC	MC	MC	MC
SPS (1)	**Id**	**State 1**	**State 2**	**State 3**	**State 4**	**State 5**					
	101	(S,3)	(M,2)	(MC,4)	(D,1)						
	102	(S,3)	(MC,7)								
SPS (2)	**Id**	**State 1**	**State 2**	**State 3**	**State 4**	**State 5**					
	101	S/3	M/2	MC/4	D/1						
	102	S/3	MC/7								
DSS	**Id**	**State 1**	**State 2**	**State 3**	**State 4**	**State 5**					
	101	S	M	MC	D						
	102	S	MC								
TSE	**id**	**time**	**event**								
	101	21	Marriage								
	101	23	Child								
	101	27	Divorce								
	102	21	Marriage								
	102	21	Child								
SPELL	**id**	**index**	**from**	**to**	**status**						
	101	1	18	20	Single						
	101	2	21	22	Married						
	101	3	23	26	Married w Children						
	101	4	27	..	Divorced						
	102	1	18	20	Single						
	102	2	21	27	Married w Children						

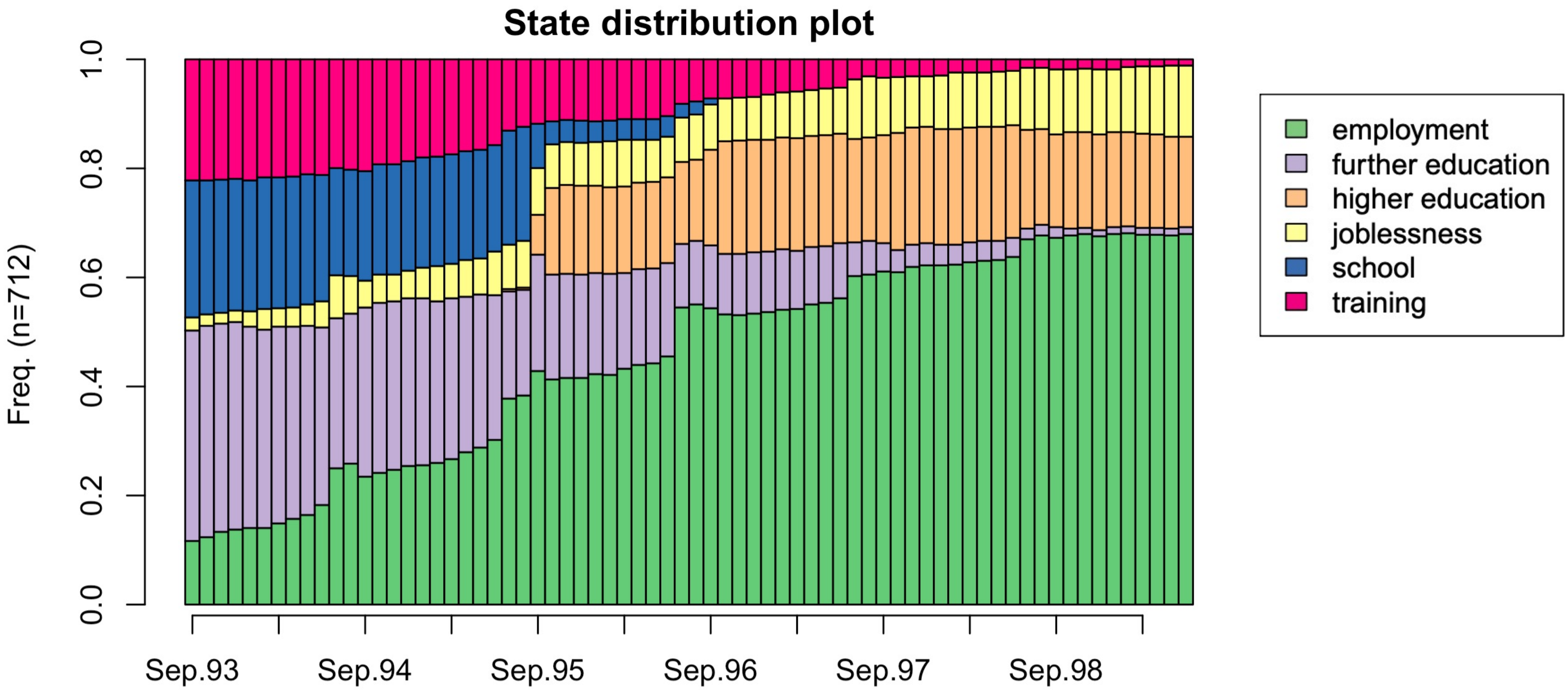
What do people's life trajectories look like?



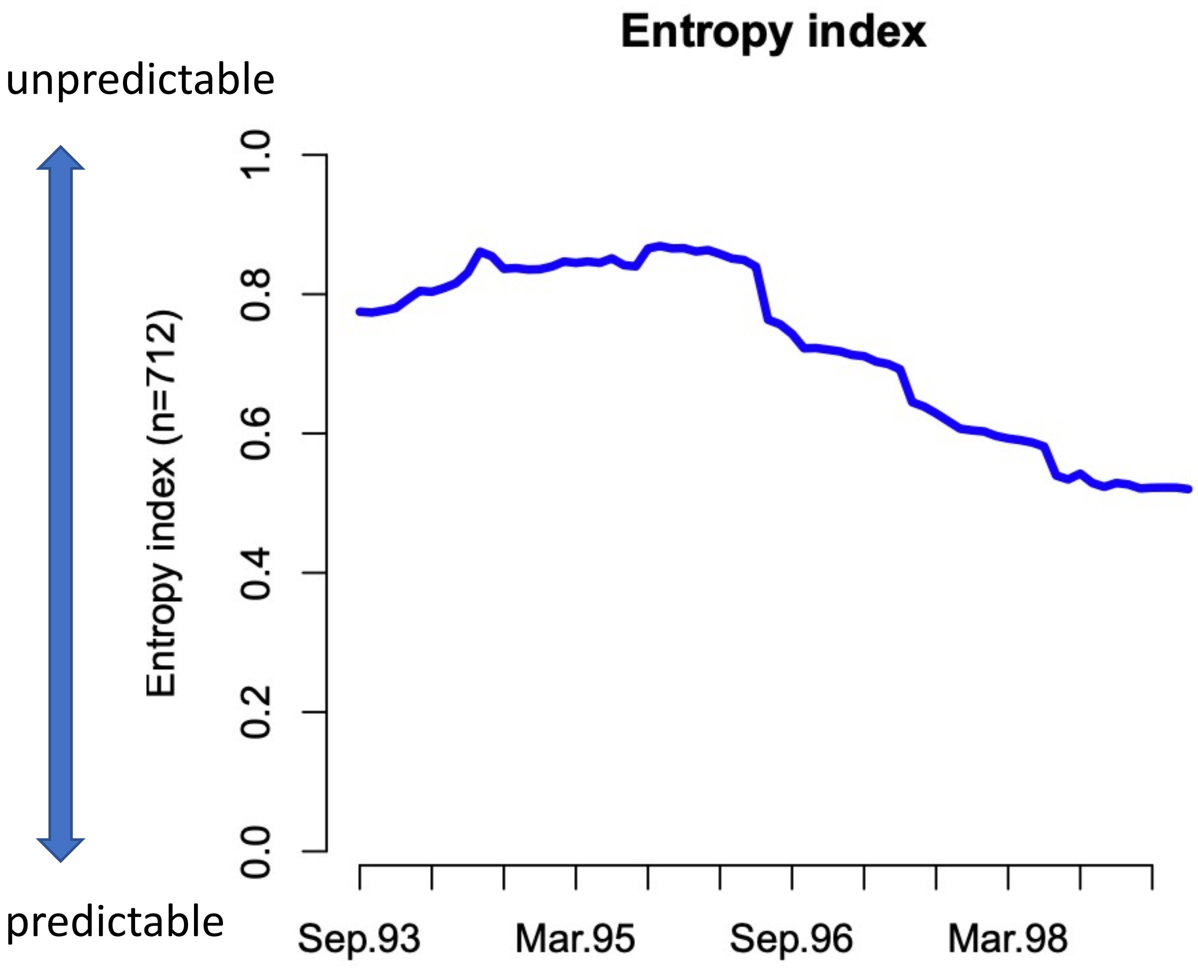
Which sequences are the most common?

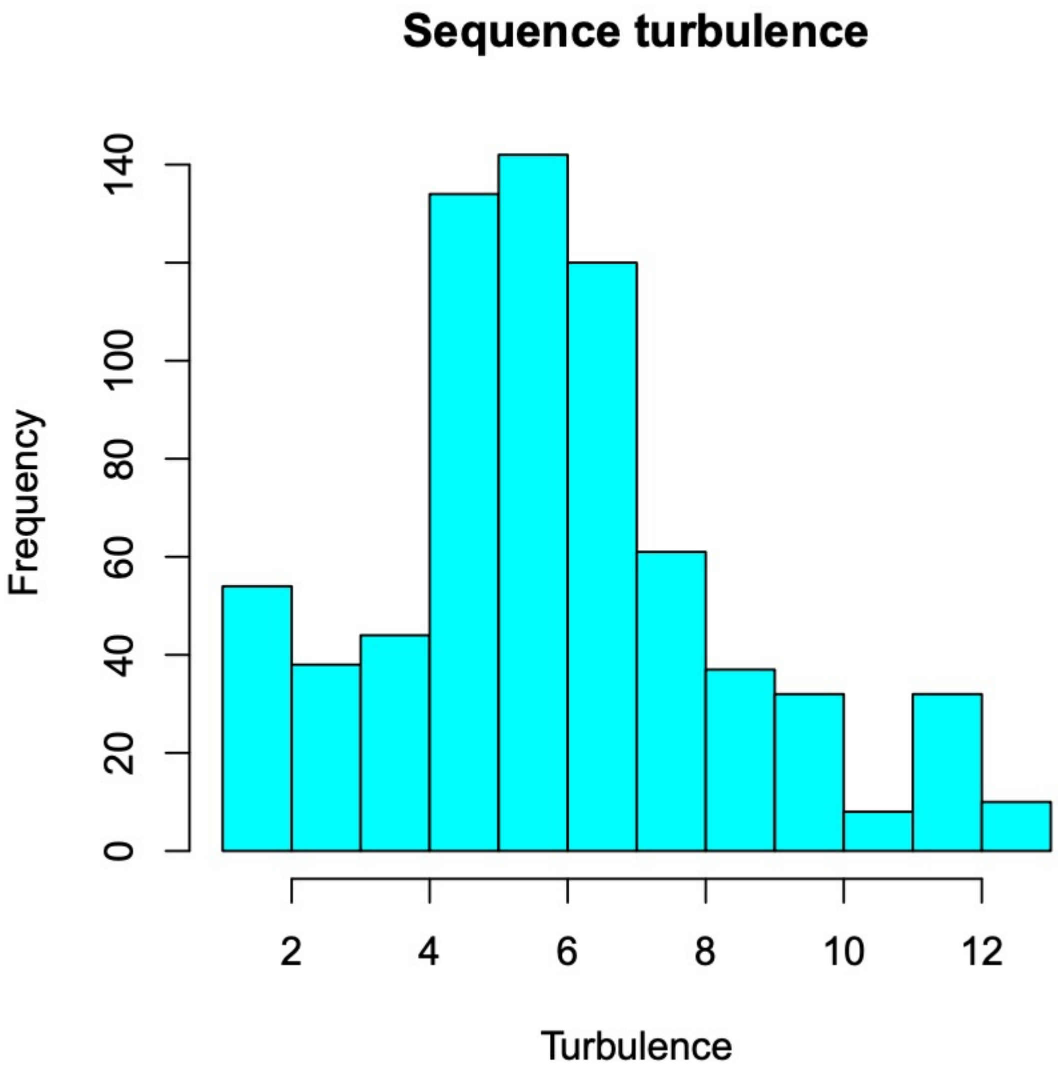


Sequence distribution over time



How stable are these sequences over time?





Case study 1

AN 8 MONTH LONG ONLINE COURSE

LOG DATA

GRADES

RQ: WHAT ARE SOME COMMON LEARNING PATTERNS AND HOW DO THEY RELATE TO ACADEMIC PERFORMANCE?

Analysis path

- Define unit of analysis (e.g., a learning session where consecutive logs are less than 30 mins)
- Define time unit (e.g., seconds, minutes, 15m, etc...)
- Label sequences
- Format sequences
- Visualize
- Descriptive stats

Take-home challenge

- Create a Rmd file that showcase how you would analyze this dataset
- Share your output in Slack