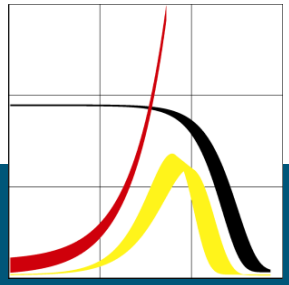
Max Planck Institute for Demographic Research



Multistate analysis of life histories with R

Frans Willekens

Pre-conference workshop

Ottawa, 18 March 2015

The overarching question

"How do events, experiences, conditions, institutions and policies affect the lives of people?"

Content

- Questions multistate models are designed to answer
- Life course data: what are the issues?
- Sequence analysis: a very brief introduction
  - Studies **observed** sequence of states
  - Disregard censoring
- Multistate life history data analysis
  - Studies transitions between states and state occupation times
  - Much attention to censoring and truncation: later entry and early departure

Research questions

- What is the effect of early-life experiences on disability status at old age?
- Is increase in longevity associated with more years with disability?
- What is the effect of smoking and obesity on life expectancy and years with disability?
- Does retirement increase the incidence of cognitive impairment?
- What is an optimal lifestyle for a long / healthy life?

Research questions

- What is the probability of being employed at 65?
- What is the probability of a divorce?
- What is the rate of divorce?
- What is the expected waiting time to a tenure?
- What is the expected sojourn time in cohabitation?

Research questions

- What are the most frequent life histories?
- Do life histories differ between cohorts or SES?
- What is the effect of a personal attribute (covariate) on the rate of transition?
- How do transition rates differ between birth cohorts and vary with socio-economic status (SES), gender, level of education, place of residence, etc.

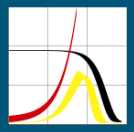
Research questions

- What is the effect of an additional year of schooling on a transition (rate and age)
- What is the effect of an additional year of schooling on the subsequent life path?
- What is the effect of age at leaving parental home on age at first birth?
- What is the effect of sequence of living arrangements on age at first birth?

Research questions

- What is the effect of marriage on employment career?
- What is the effect of postponement of marriage on age at first birth?
- What is the effect of duration of unemployment on likelihood of finding a job?

Two observations on challenges in life course research



“New directions in life course research”
Karl Ulrich Mayer
Annual Review Sociology, 2009

- “There will always be more need for modeling transitions between states.” (p. 425)
- Concept of risk: “Exposure to risk, measured by its incidence and duration, can be a powerful concept in mapping and measuring life courses” (p. 424)

“New Life for Old Ideas: The ‘Second Wave’ of Sequence Analysis Bringing the ‘Course’ Back Into the Life Course”

Aisenbrey and Fasang

Sociological Methods and Research, 2010

“To date, the majority of quantitative life course studies are based on event history applications that model the probability of single transitions and durations under certain conditions.”

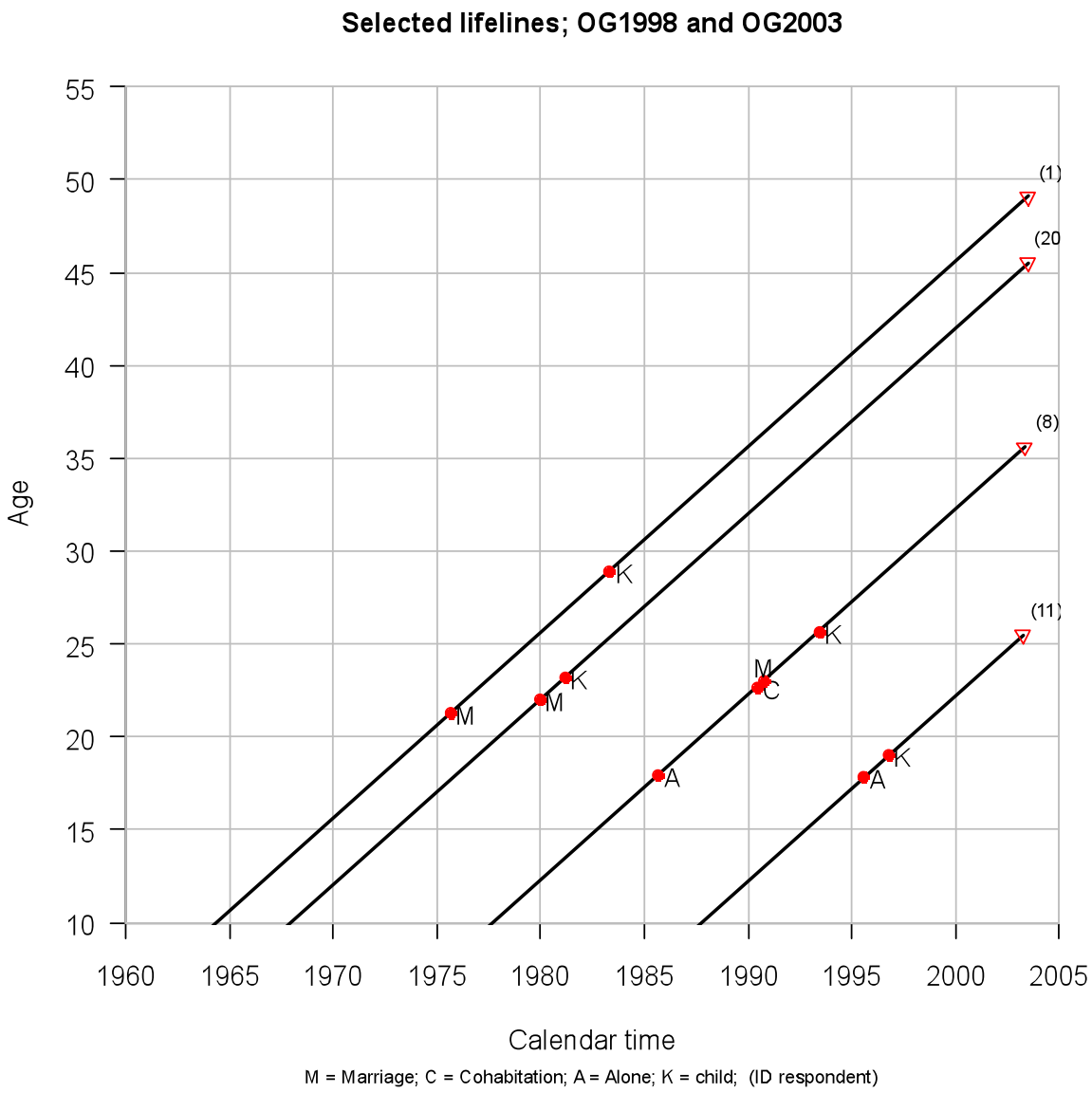
Life course data

Data on subjects **over time** (focus on processes)

- Quantitative: panel, follow-up study, cohort study
  => life histories

- Qualitative: autobiographies, oral histories
  => life stories

Lexis diagram



Events

- Events of interest
- Unrelated events (censoring: lost to observation)
- Competing events (competing risks)

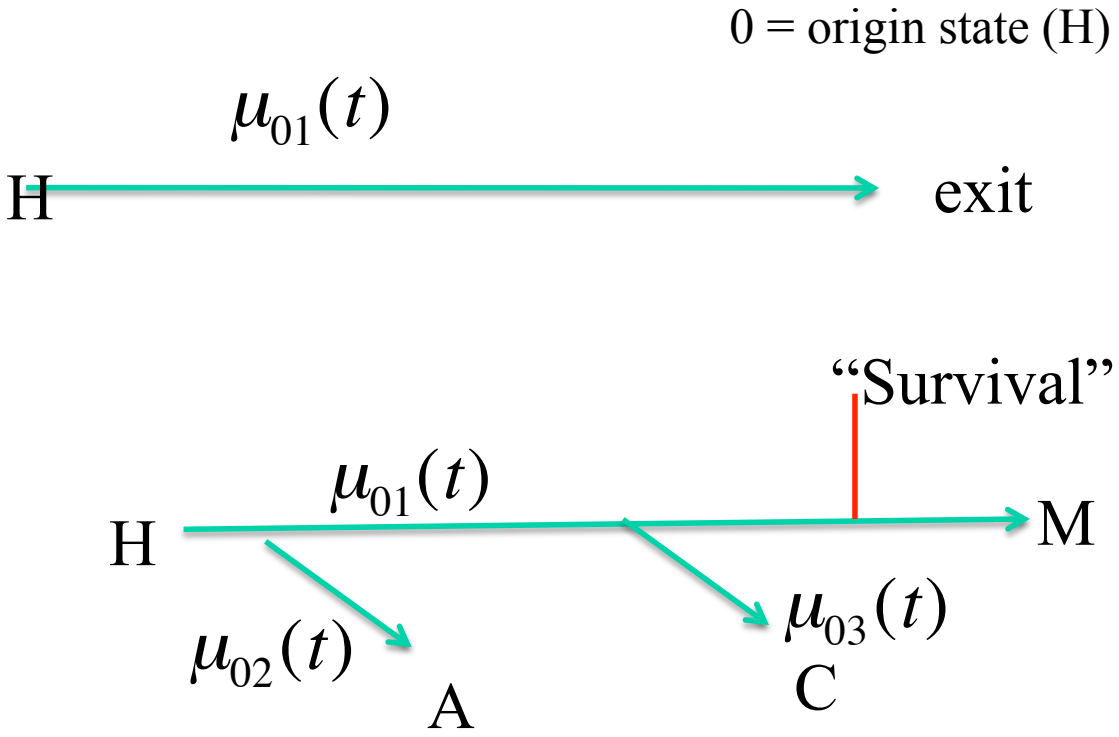
Exposures

- Episode
- Exposure
- Exposure ends in event of interest or censoring

Sequences

- Event sequence (event history)
- State sequence

Single event
(single decrement)

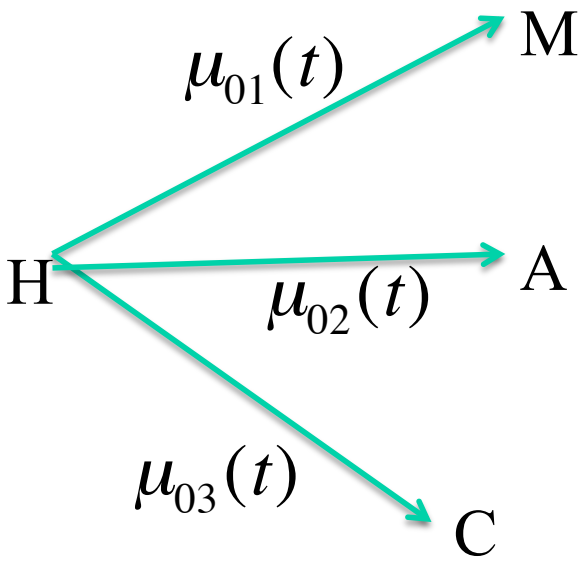


- Event of interest and unrelated events
- Exposure

Total exit rate: $\mu_{0+}(t) = \mu_{01}(t) + \mu_{02}(t) + \mu_{03}(t)$

Competing risks

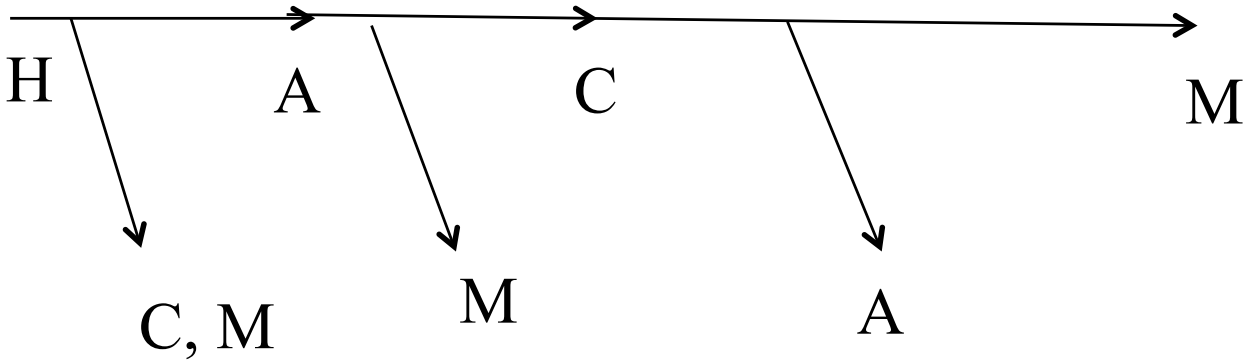
0 = origin state (H)



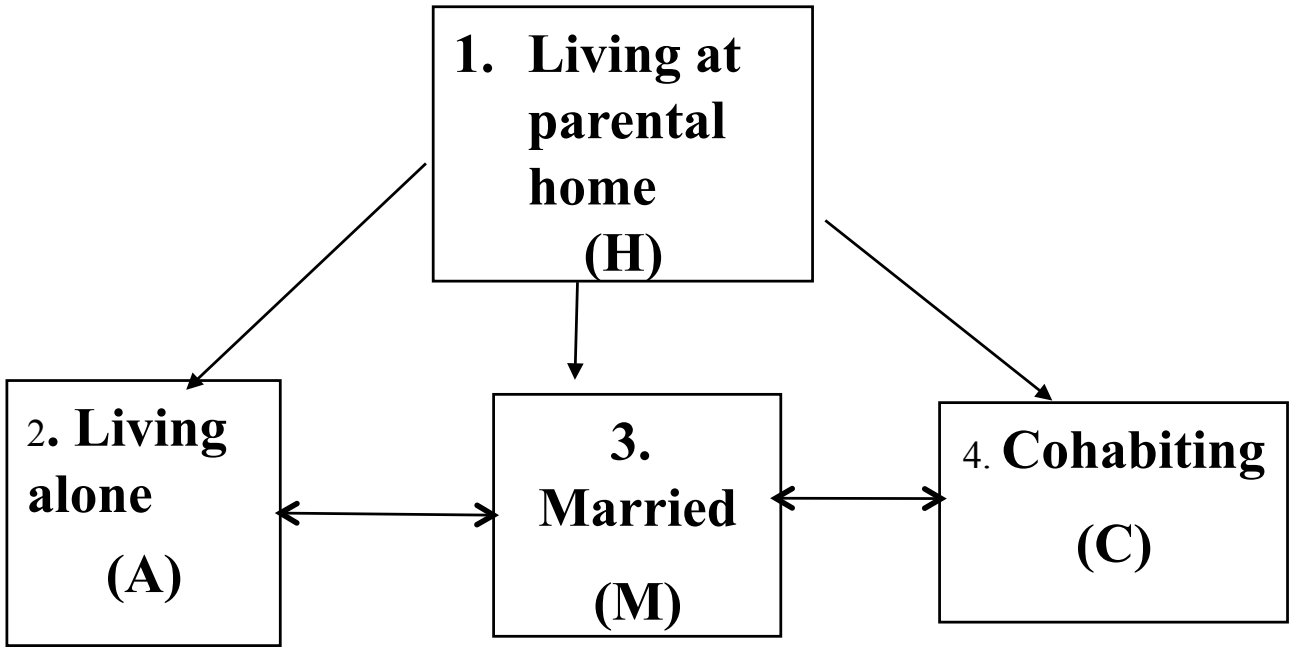
- Event of interest and unrelated events
- Exposure

Total exit rate: $\mu_{0+}(t) = \mu_{01}(t) + \mu_{02}(t) + \mu_{03}(t)$

Multistate: event sequence or state sequence



Competing risks



Multistate modeling as extension of competing risks

Multistate life history analysis

- A life domain is characterized by a finite set of possible states (*multistate; multiple states*)
  - Educational status
  - Health status
  - Employment status
  - Living arrangement
  - Marital status
- State space: a finite set of possible states
  - Alphabet (TraMineR): (short) state labels
  - State space depends on research question
- Any time, a person may leave a state and enter one of the other states
  - Multiple possible destinations: *competing risks*

Multistate life history analysis

- The life course is a sequence of states and transitions between states
- State sequence: ordered list of states (on a time axis)
- Event sequence: ordered list of events (on time axis)
- At a given time, individual occupies one state
  - Grade of membership method; mixture models
- State occupied is observable (not latent or hidden)
  - Hidden multistate models (e.g. hidden Markov model)

Sequence representation

- Sequence representation: use state space or alphabet
- Living arrangement
  - State space:
    - H Living with parents
    - A living alone
    - M married
    - C cohabiting
  - Sequence: A= {H,A,C,A,M,A}

- Employment history
  - State space (letters in alphabet)
    - N No Job
    - J Job
  - State sequence: A = {N,J,J,N,J,N,J,J,N}
  - State occupancies: {N-N-N-J-J-J-N-J-J-J-J-N-N}

Sequence data: data types

State sequence				Event sequence			
ID	time	state		ID	time	event	time
1	0	A		1	0	A	2010
1	1	B		1	0	AB	2011
1	2	A		1	1	BA	2011
2	0	C		2	0	C	2008
2	1	A		2	0	CA	2008
2	2	C		2	1	AC	2009
2	3	B		2	2	CB	2011
3	0	B		3	0	B	2005
3	1	B		3	2	CB	2007
3	2	B					
3	3	C					

Sequence data: data types

State sequence				
Year	2010	2011	2012	2013
Marital status	Single	Single	Married	Divorced
Location	Parents	Alone	Cohabitation	Alone
Child	No	No	No	Yes

Episodes					
ID	Start	End	Marital Status	Location	Child
1	1321	1333	Single	Parents	No
1	1333	1344	Single	Alone	No
1	1344	1356	Married	Cohabitation	No
1	1356	1380	Divorced	Alone	No
1	1380	1384	Divorced	Alone	Yes

Data 1. German Life History Survey (GLHS) (Blossfeld & Rohwer, 2002)

- 201 respondents; three birth cohorts: 1929-31, 1939-41, and 1949-51
- 600 job episodes; 382 out-of-job episodes

Employment history

Time representation

Calendar time

Age

Duration in state

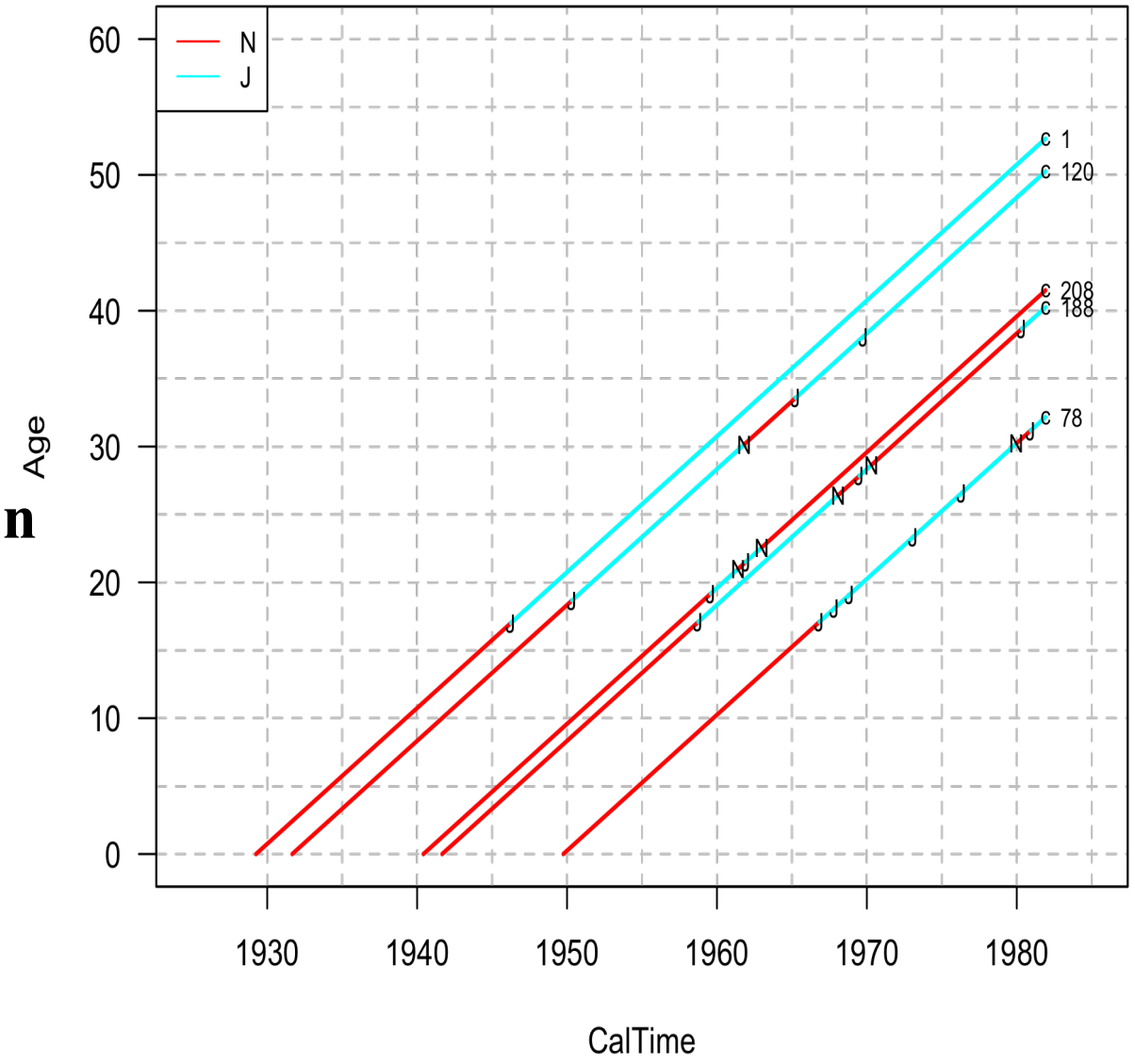
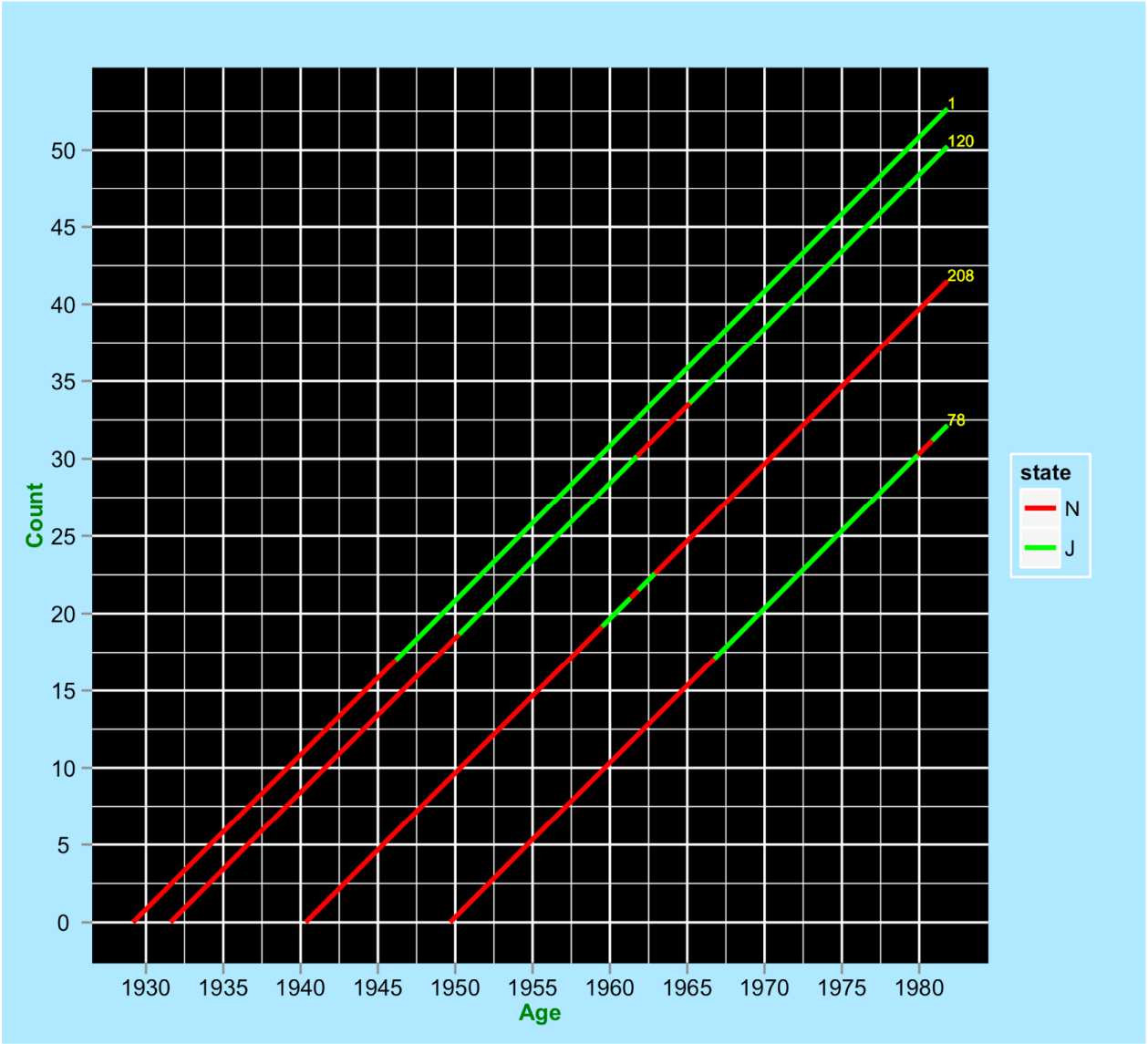
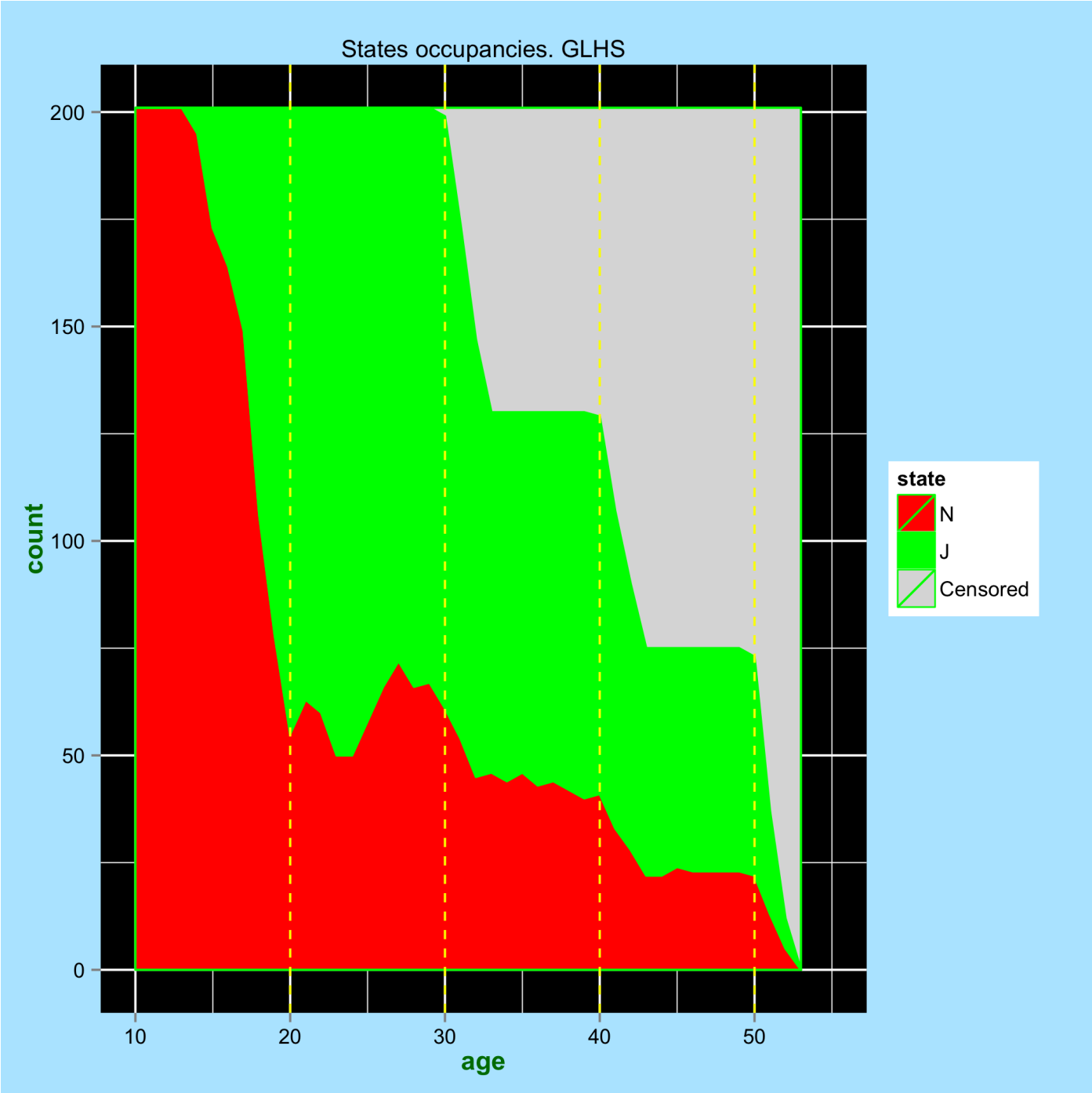


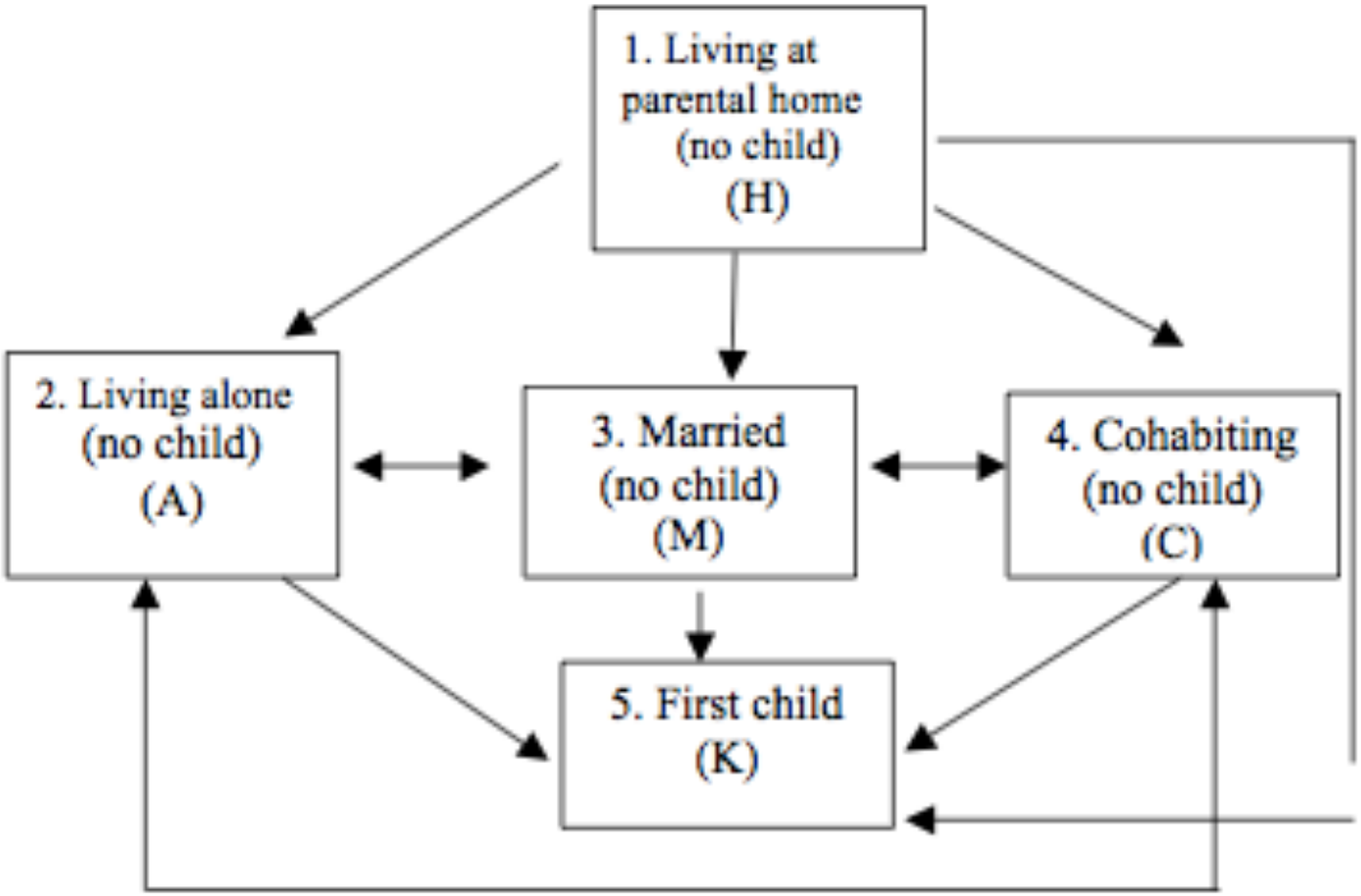
Figure 5.8 Lexis diagram: employment careers of selected GLHS respondents. Display B, using *ggplot2* package.




States occupancies. GLHS
200
150
100
50
0
count
10
20
30
40
50
age
state
N
J
Censored

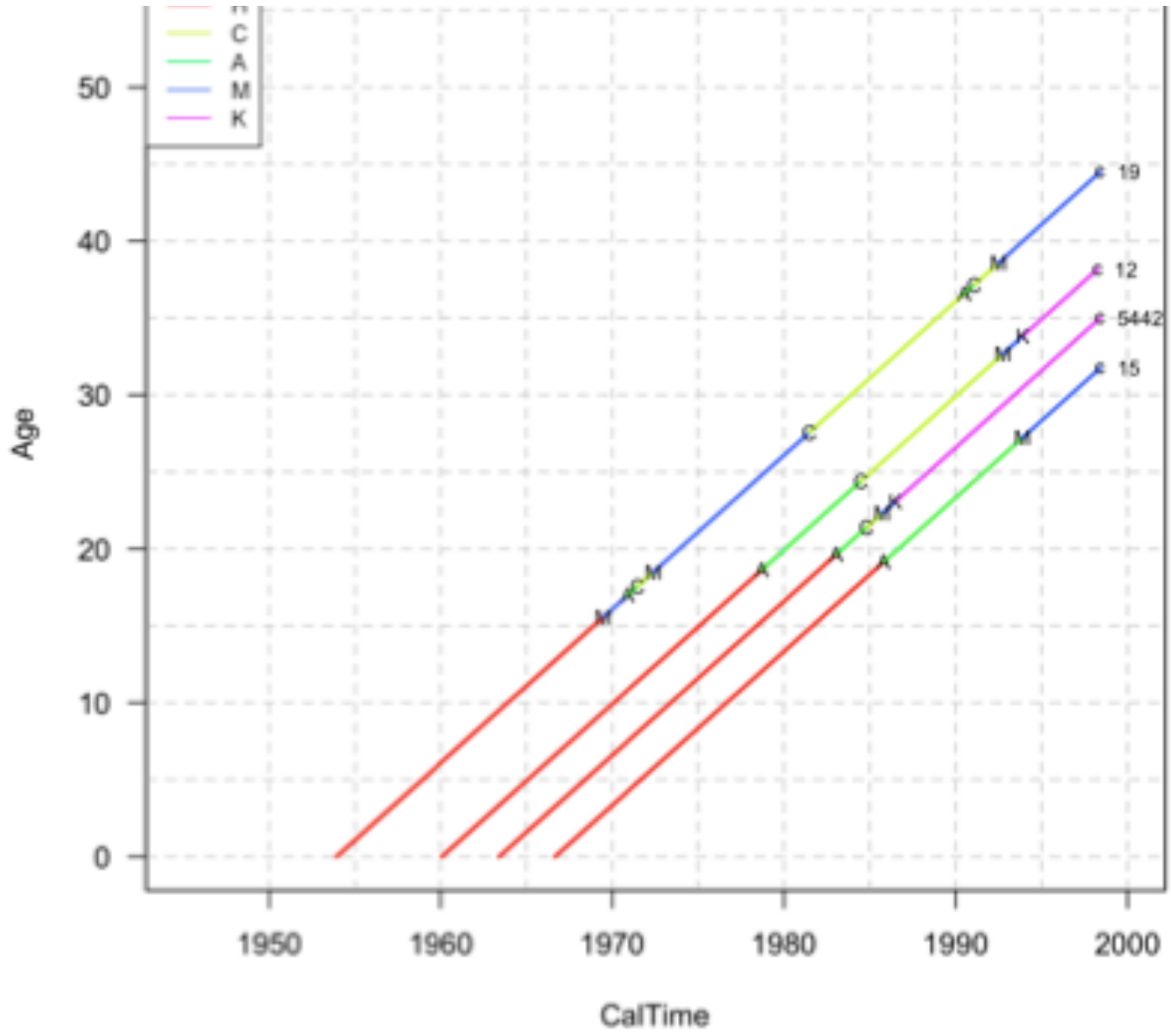
Data 2. Netherlands Family and Fertility Survey (NLOG) 1998

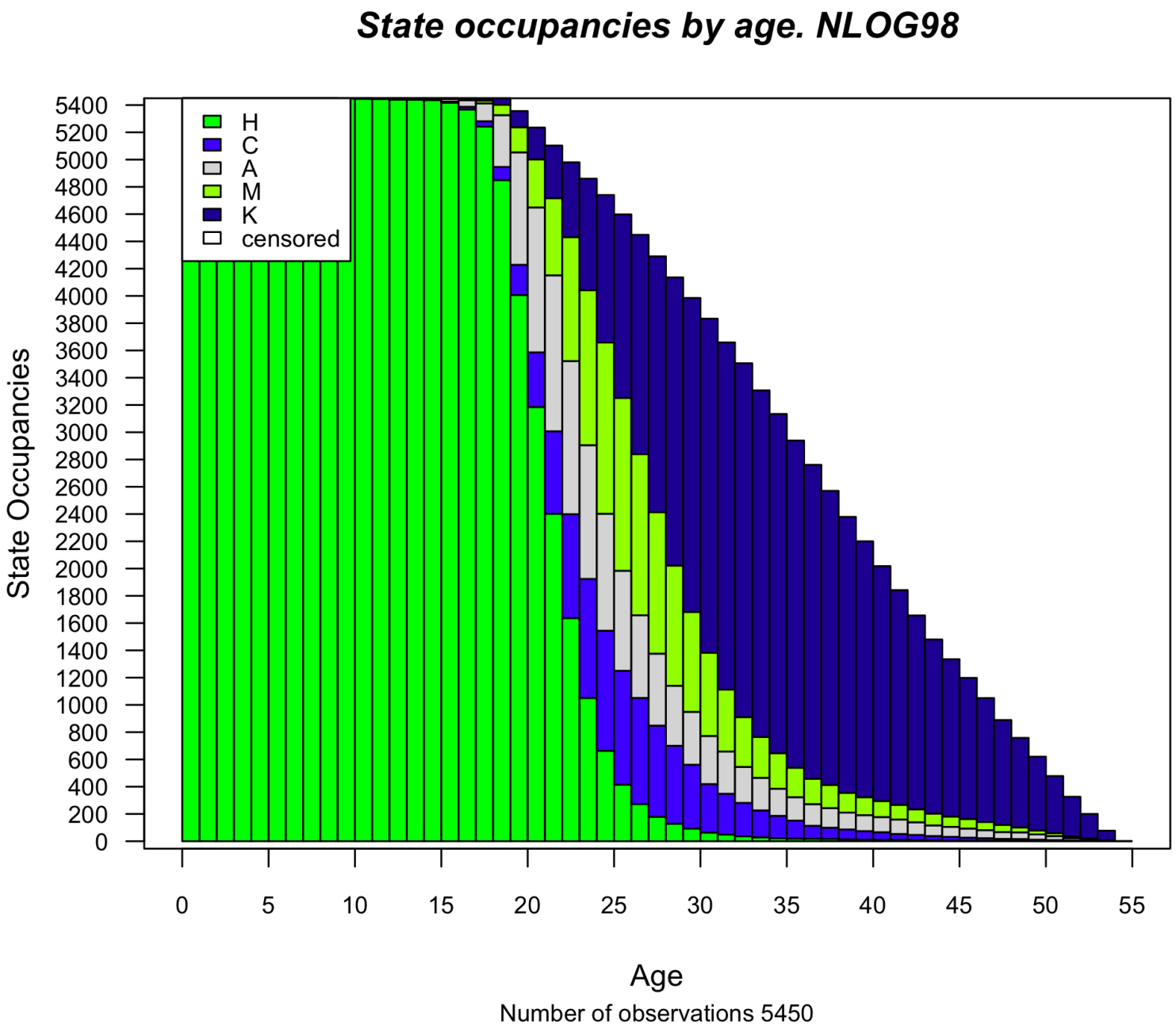
- 5450 respondents



Pathways to first child

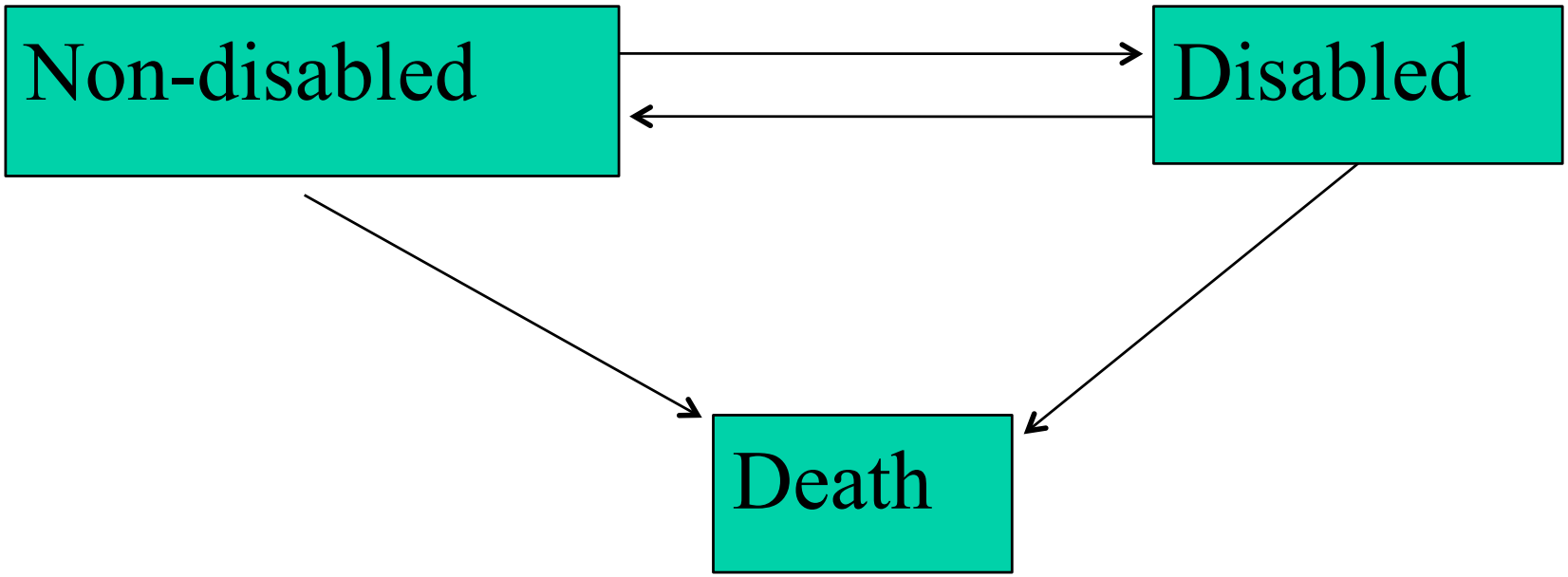
Netherland Family and Fertility Survey 1998




State occupancies by age. NLOG98
H
C
A
M
K
censored
State Occupancies
5400
5200
5000
4800
4600
4400
4200
4000
3800
3600
3400
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
0
5
10
15
20
25
30
35
40
45
50
55
Age
Number of observations 5450

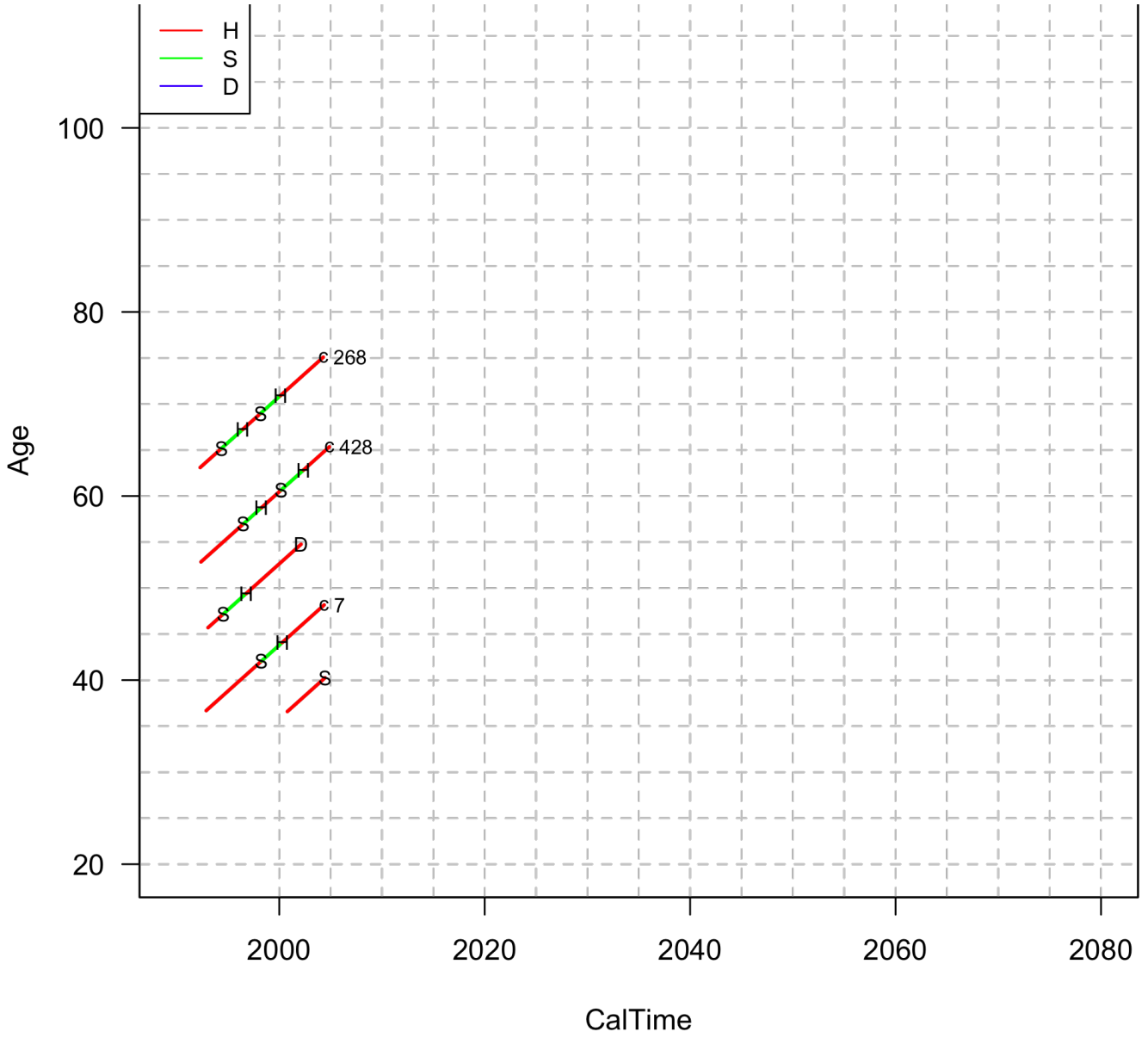
Data 3. U.S. Health and Retirement Study (1992 – 2004)

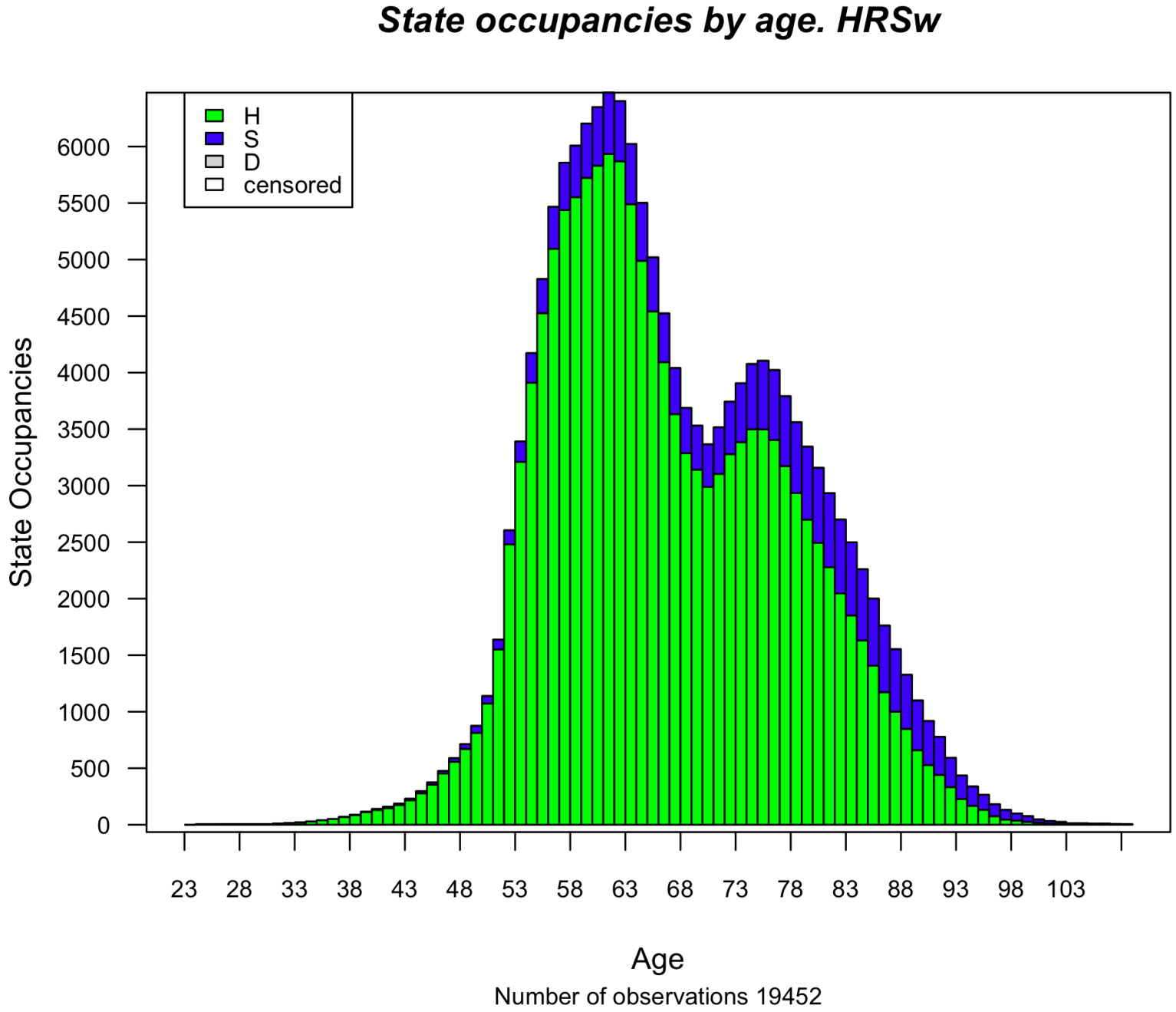
- 19425 respondents



Illness-death model with recovery
(extension of hazard model with time varying covariate)

Disability history
US HRSw 1992-2004




State occupancies by age. HRSw
H
S
D
censored
State Occupancies
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
23
28
33
38
43
48
53
58
63
68
73
78
83
88
93
98
103
Age
Number of observations 19452

Models need to consider data limitations
early exit, late entry

=> Counting process perspective: track transitions and exposures

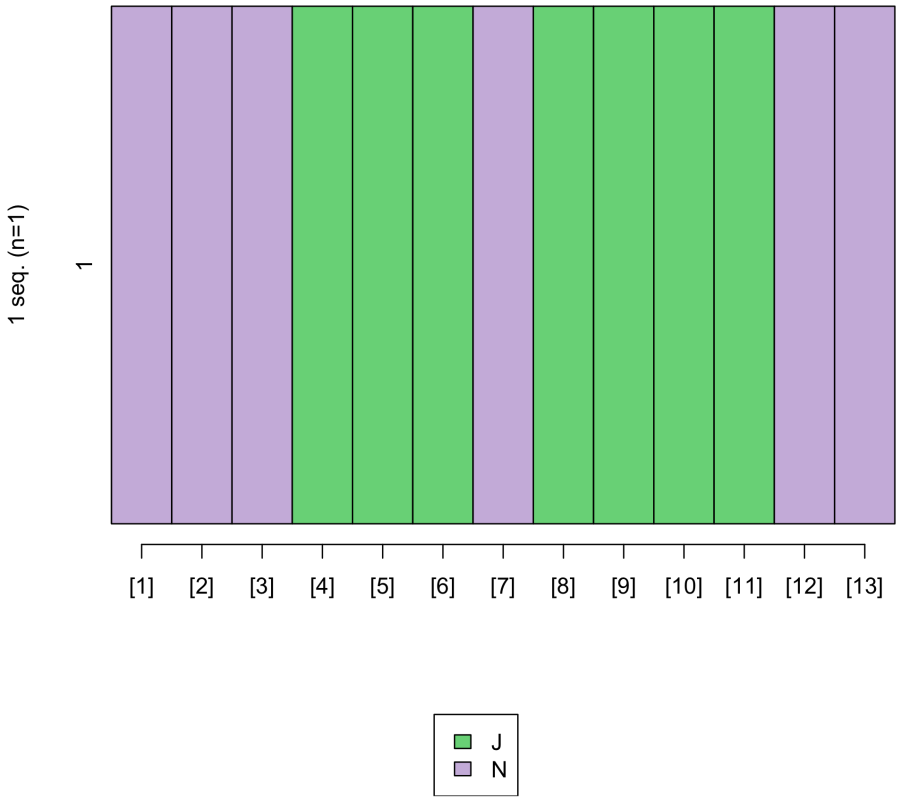
Sequence analysis

- Sequence (trajectory) is unit of analysis
- Study trajectories (fixed period)
- Compare trajectories
- 1986: Andrew Abbott and John Forrest: “Optimal matching methods for historical sequences”
- 2007: Chesa (Elzinga, Free University Amsterdam)
- 2008: TraMineR (Richards, University of Geneva)
- Research questions
  - Most frequent sequences in stage of life
  - How often do transitions occur in a cohort
  - What are general types of life stage trajectories?
  - How and why do trajectories differ between individuals?

Short course of social sequence analysis: http://www.youtube.com/watch?v=9WJPook9Qsc

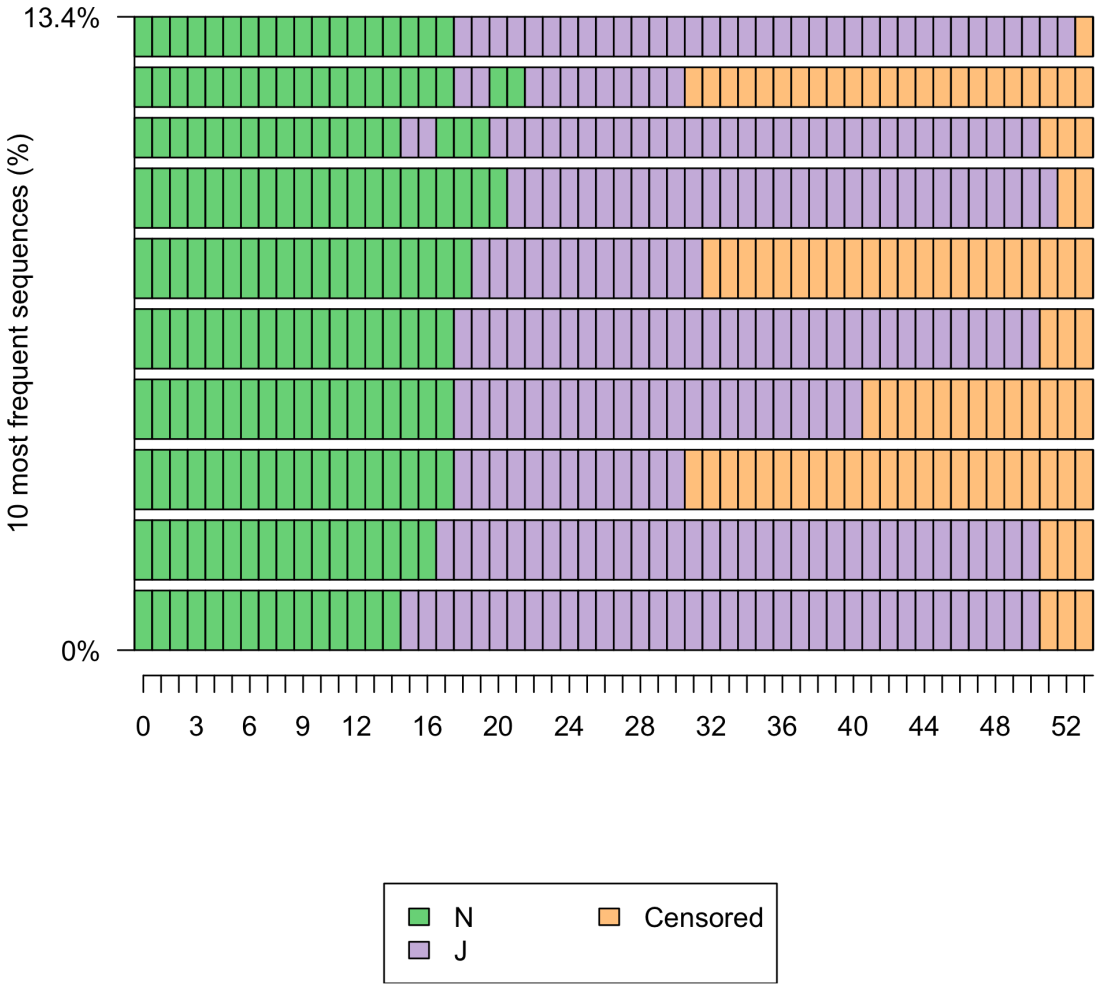
Sequence analysis

- State sequence: {N-N-N-J-J-J-N-J-J-J-J-N-N}



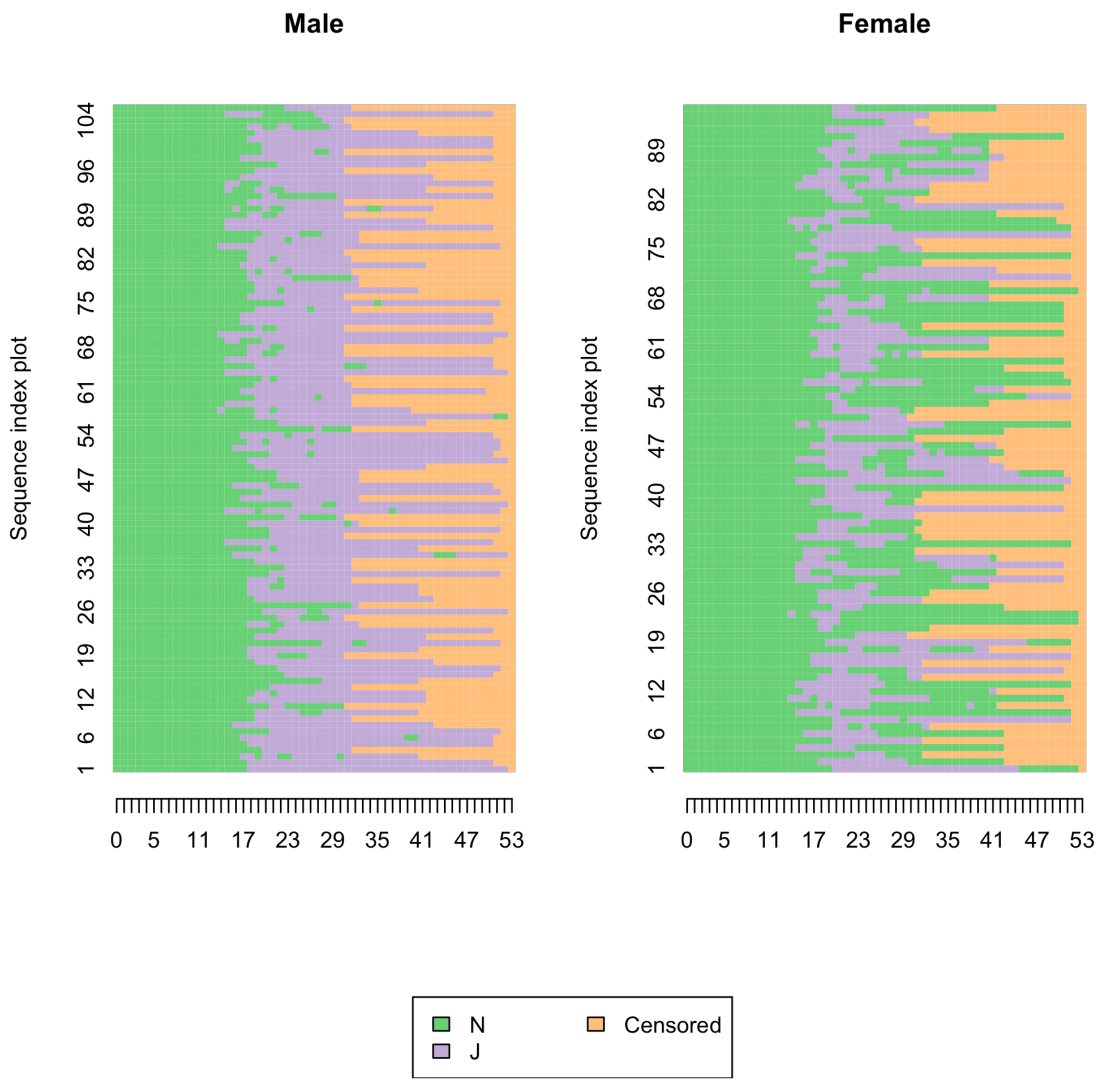

```
library (TraMineR)
dd <- "N-N-N-J-J-J-N-J-J-J-J-N-N”
ddd <- seqdef(dd)
seqplot(ddd, type="i")
```

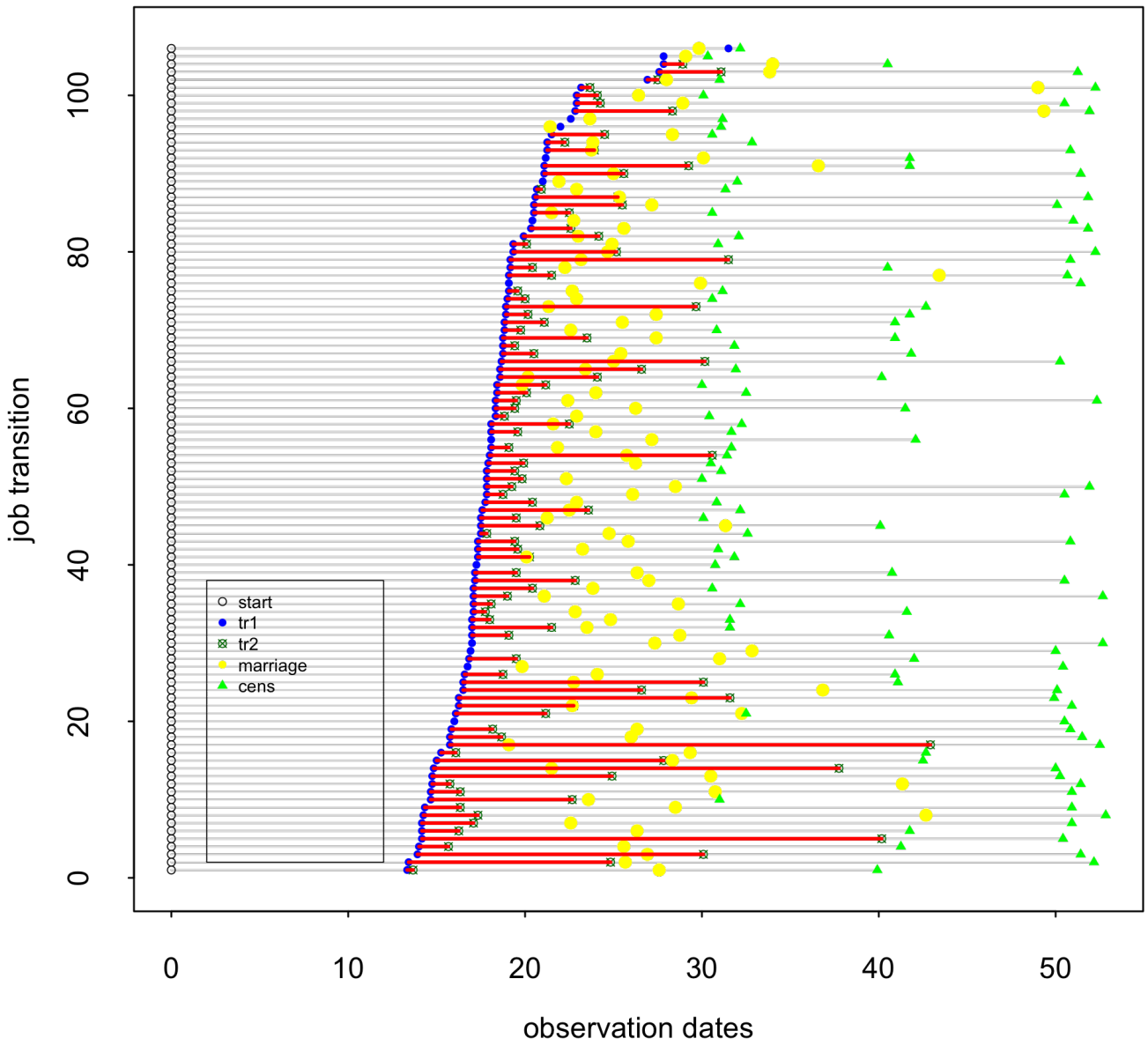
Sequence frequency plot




```
library (Biograph)
data (GLHS)
parameters <- Parameters (GLHS)
occup <- Occup (GLHS)
DTraMineR <- seqconc (occup
$st_age_1,sep="-")namst <- c(parameters
$namstates,"+")
D.seq <- seqdef
(DTraMineR,labels=namst,alphabet=c("N","J","+"
))alphabet (D.seq)
n <- 10
seqfplot (D.seq,tlim=1:n,title="Sequence
frequency
plot",xtlab=c(0:54),ltext=c("N","J","Censored"),la
s=1,ylab=paste(n," most frequent sequences (%)",
```

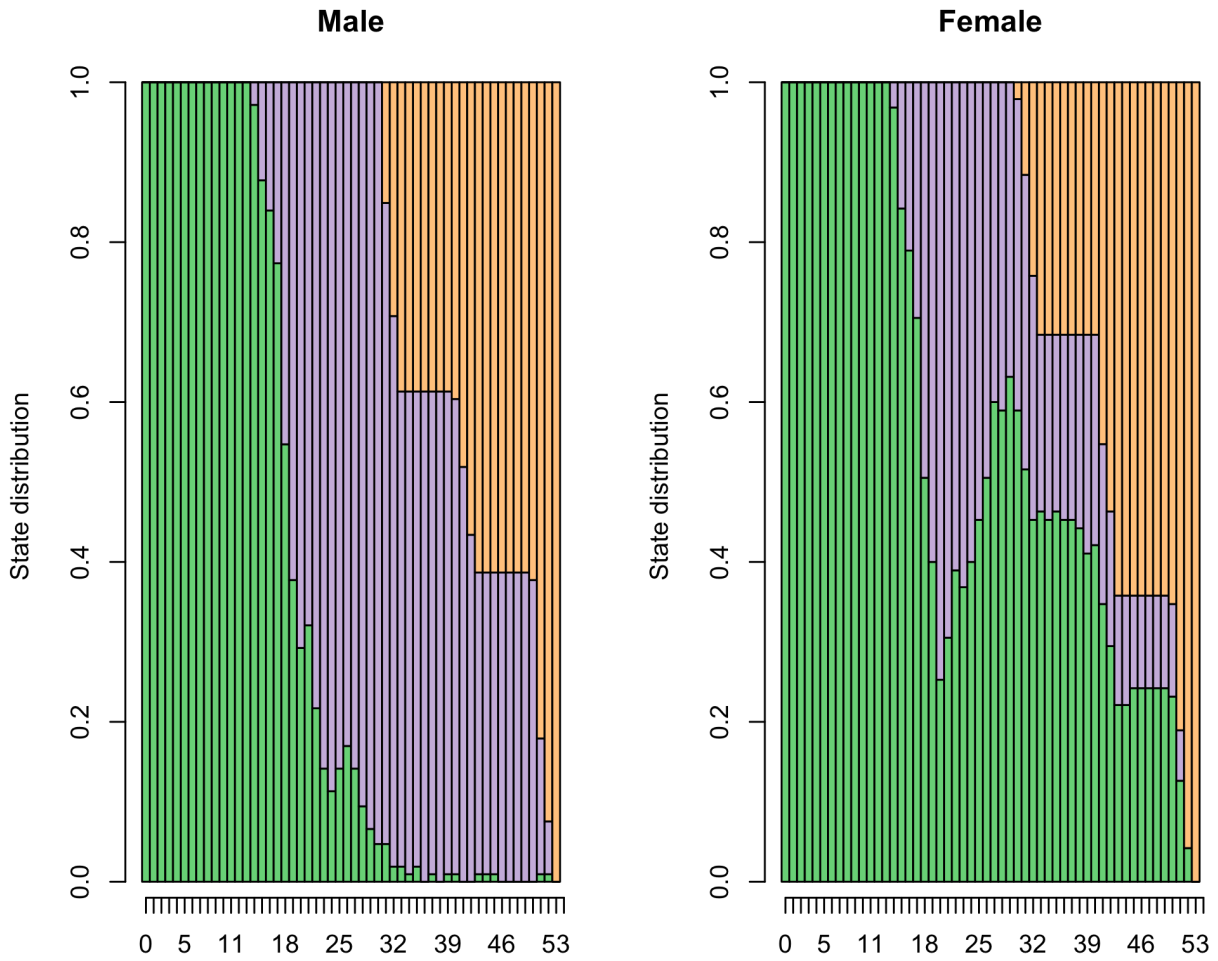
Sequence index plot, 201 respondents

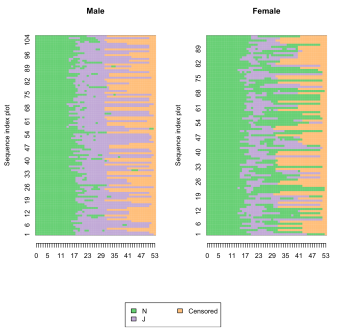




R\TEACH\MDPR2013\graphics\1 Eventchart

State distribution plot, 201 respondents





Sequence analysis: compare trajectories

- Pairwise comparison of sequences
  - Reference sequence
  - How many sequences need to be inserted or deleted to make two sequences equal
  - Dissimilarity measures
  - Optimal matching (OM)
  - Problem: censoring

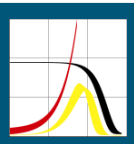
Multistate life history analysis

- The life course is a sequence of states and transitions between states
- Does not focus on description and comparison of trajectories (patterns) but on processes generating the trajectories
- Model transitions between states in a state space
- The process governing transitions is a multistate process in continuous time (e.g. Markov, semi-Markov, non-Markov multistate model)

Multistate life history analysis

- Life course dynamics (trajectory) is governed by rates of transition that vary with age and depend on history and context
- Rate = events / exposure (duration)
  - Hazard rate
  - Occurrence-exposure rate (occurrences / exposure)
  - Incidence rate
  - Transition rate

Ulrich Mayer (2009): "Exposure to risk, measured by its incidence and duration, can be a powerful concept in mapping and measuring life courses" (p. 424)



Estimation of transition rates a counting process perspective

- Events: A counting process counts event occurrences and people at risk (exposed).
  - At each occurrence of a transition
  - During an interval (observation window): event count and exposure time

- Exposure: Numbers of individuals at risk
  - Just before event time -> Nelson-Aalen estimator
  - During the observation period: number at risk AND duration at risk -> occurrence-exposure rate

Thank you