Assignment 11 Mining event sequences I

1. The `biofam` data was created from time stamped event data. Table 1 describes how the states were defined in terms of event occurrences.

Table 1: Biofam state definiton in terms of events

State	LHome	Marriage	Childbirth	Divorce
P	no	no	no	no
L	yes	no	no	no
M	no	yes	yes/no	no
LM	yes	yes	no	no
C	no	no	yes	no
LC	yes	no	yes	no
LMC	yes	yes	yes	no
D	yes/no	yes	yes/no	yes

   Using `seqecreate`, transform the `biofam.seq` state sequence object considered in the previous assignments into an event sequence object with five events: P (Starting leaving with parents), L (leaving home), M (Getting married), C (Childbirth), D (Getting divorced).

2. Plot the event sequences by sex. What is the main difference between men and women.

3. Plot the event sequences by the birth cohorts defined in Assignment 10 (Before end of word war II versus after end of WW-II). Represent non-embeddable sequence patterns and color only those with a support of at least 15%. Comment differences between the two cohorts.

4. Find the most frequent subsequences (minimum support of 10%) with at least 2 events. Among those who left home and got married, what is the proportion who did it the same year? Plot the 10 most frequent subsequences.

5. Display and plot the 10 subsequences which best discriminate (a) women from men, and (b) birth cohorts.