

Sequence Analysis

Liliana Patricia Calderón Bernal

Gonzalo Daniel Garcia

Ainhoa-Elena Leger

14/4/2021

Load necessary packages.

```
# Call TraMineR library
library(TraMineR)

# Call other required libraries
library(ggplot2)
library(grDevices)
library(graphics)
library(foreign)
library(cluster)
library(Hmisc)
library(TraMineRextras)
library(WeightedCluster)
library(RColorBrewer)
library(colorspace)
```

Exercise 1

- 1) Input the Dataset 2

[Sol.]

```
data2 <- read.csv("SFS2018_Data2.csv", na.strings=c(".", "a", "b"))
```

- 2) Define a sequence object with elements in data columns 2:61 and alphabet 1:6, using the following state names and labels

- 1 SNP "Single, childless",
- 2 SBP "Single, child b/separat.",
- 3 SAP "Single, child a/separat.",
- 4 UNP "Union, childless",
- 5 UBP "Union, child b/separat.",
- 6 UAP "Union, child a/separat."

[Sol.]

```
# Create a vector for the state labels
seqlab <-c("Single, childless",
           "Single, child b/separat.",
           "Single, child a/separat.",
           "Union, childless",
           "Union, child b/separat.",
           "Union, child a/separat.")
```

```

    "Union, child a/separat.")

# Create a vector of short state names (default would be alphabet labels)
sllist <- c("SNP", "SBP", "SAP", "UNP", "UBP", "UAP")

# Define Color palette
color1 <- sequential_hcl(6, palette = "SunsetDark", rev= TRUE)

### Generate sequence object
seqObj2 <- seqdef(data2,
                  var=2:61,
                  alphabet=c(1:6),
                  cpal=color1,
                  states=sllist,
                  labels=seqlab)

### Retrieve information from sequence object
summary(seqObj2)
names(seqObj2)

```

3) Display (print) the first 10 sequences in extended and compact form

[Sol.]

```

#display the first 5 sequences, and sequence elements 1-20 (STS format - default).
print(seqObj2[1:10, ], format = "STS")
#display the first 5 sequences, and sequence elements 1-20 (SPS format)
print(seqObj2[1:10, ], format = "SPS")

```

4) Plot a full representation of sequences, and order them from the first state

[Sol.]

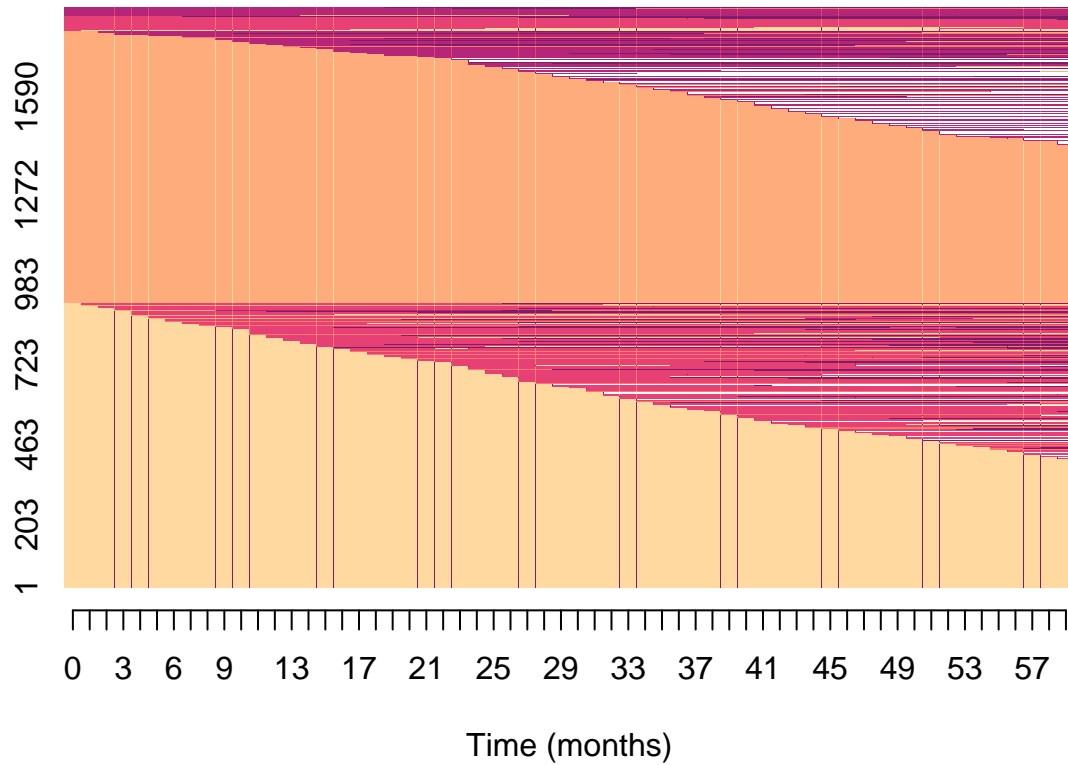
```

# X-axis for exercise
xtlab=seq(0,60, by=1)

#All sequences -sequence index plot (sorted - first state)
par(mfrow=c(2,1))
seqIplot(seqObj2, with.legend=TRUE, main= "All sequences",
         xtlab=xtlab, xlab="Time (months)", ylab=NA, yaxi=TRUE,
         border=NA, sortv="from.start")

```

All sequences



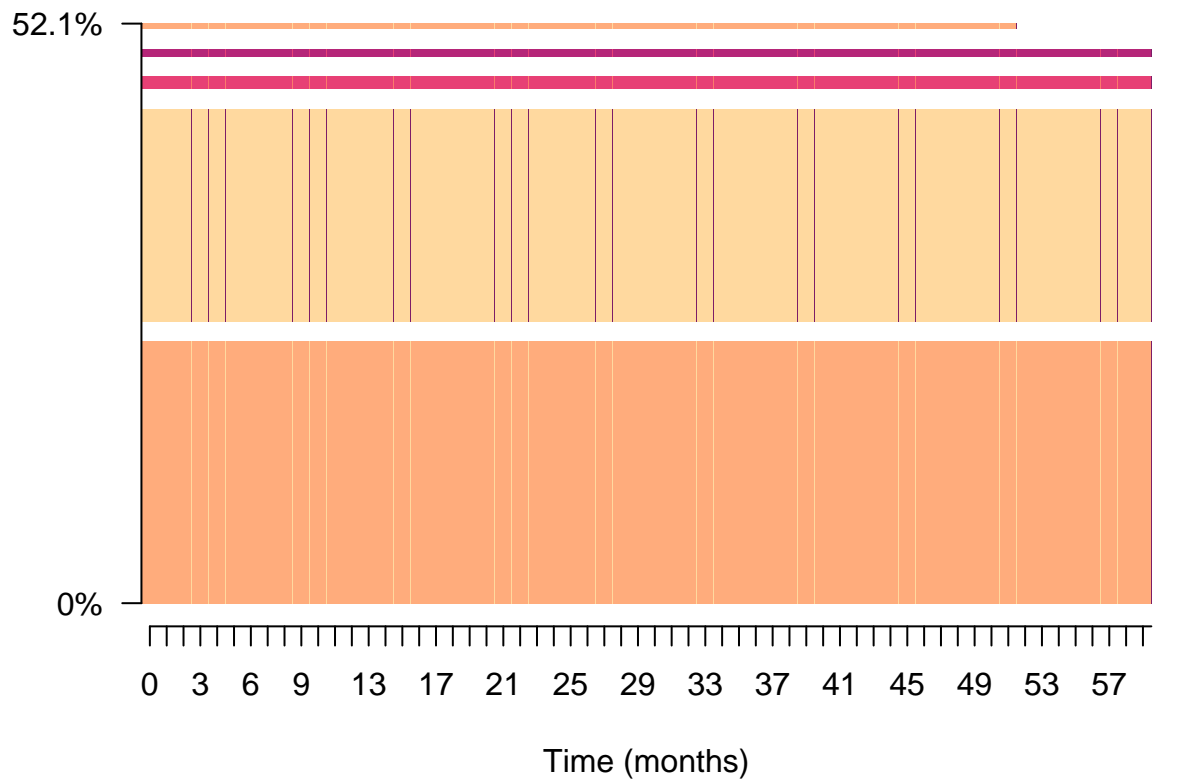
 Single, childless	 Single, child a/separat.	 Union, child b/separat
 Single, child b/separat.	 Union, childless	 Union, child a/separat

5) Plot the 5 most frequent sequences. Comment the plot

[Sol.]

```
par(mfrow=c(2,1))
seqfplot(seqObj2, idxs=1:5, main="5 most frequent sequences",
  with.legend=TRUE, border=NA,
  ylab=NA, xlab="Time (months)", xtlab=xtlab)
```

5 most frequent sequences



Single, childless	Single, child a/separat.	Union, child b/separat
Single, child b/separat.	Union, childless	Union, child a/separat

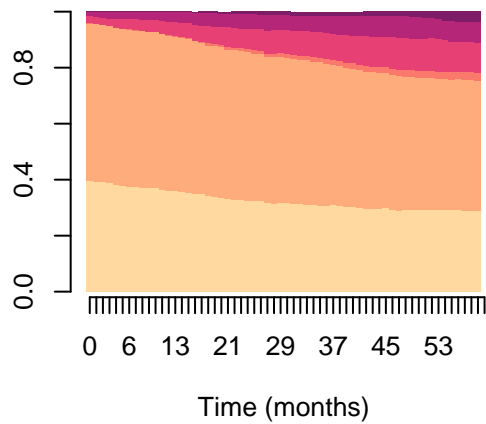
52% of the people live 5 sequences
 # The majority remains single 6 months after the separation
 # The two most frequent sequences correspond to people who stay single with or without children

6) Create a state distribution plot for each birthcohort (BIRTHCOH). What are the cross-cohort differences in the distribution of states overtime?

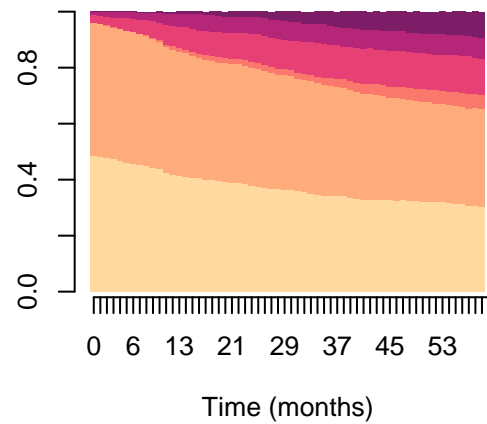
[Sol.]

```
seqdplot(seqObj2, group=data2$BIRTHCOH, with.legend=TRUE,
         main= "State distribution. Cohort", use.layout=FALSE,
         border=NA, xtlab=xtlab, ylab=NA, xlab="Time (months)")
```

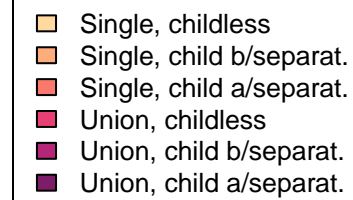
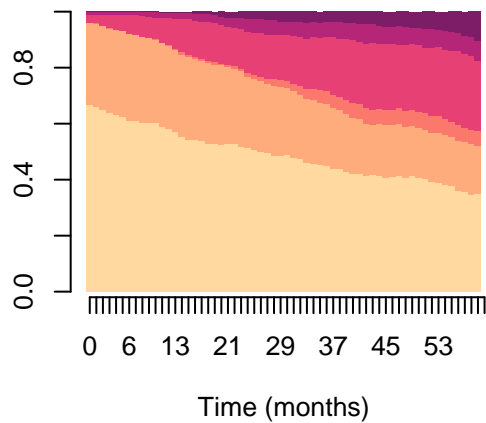
State distribution. Cohort – 1



State distribution. Cohort – 2



State distribution. Cohort – 3



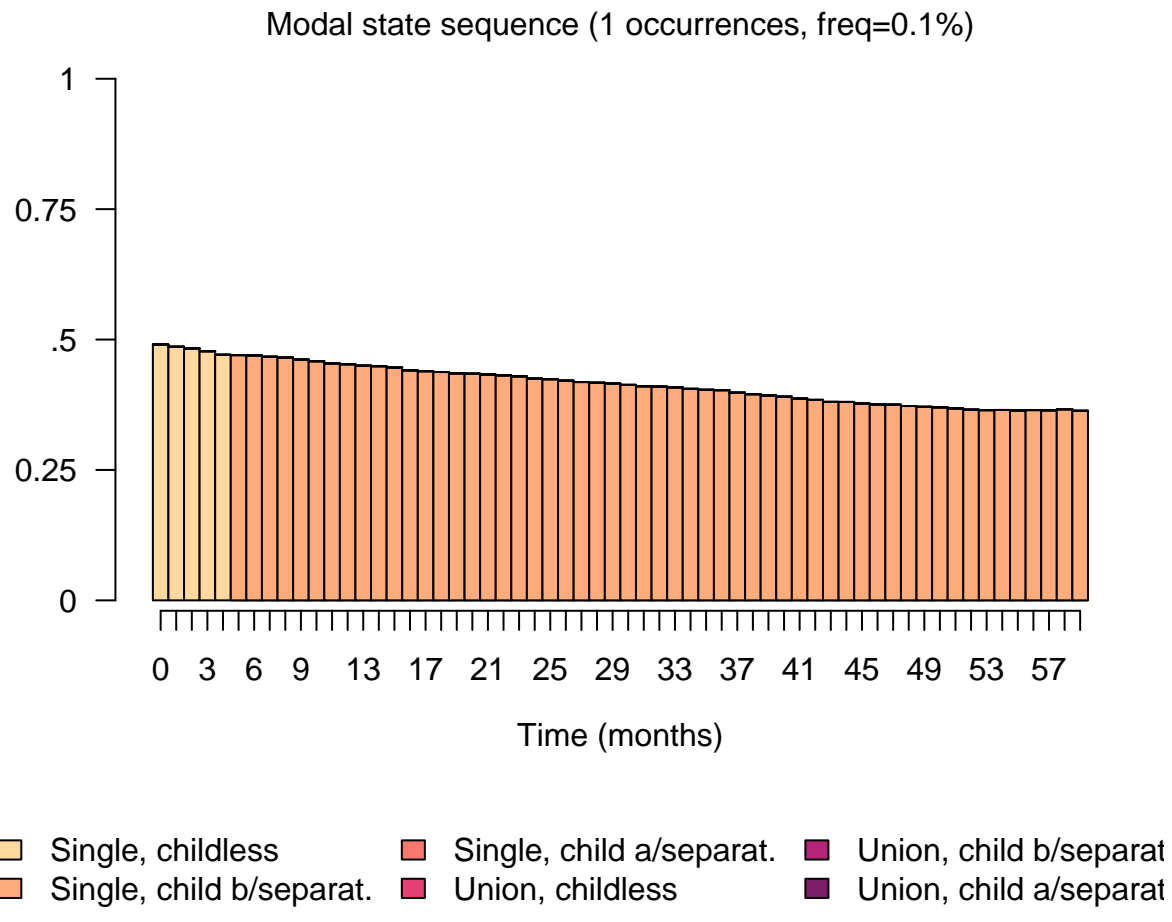
*# Cohort 1 is characterized by a lower number of people who are single and
childless just after the separation, the state remains stagnant in the
following months
For the other cohorts, there are more respondents who are single and
childless but they are less likely to remain in this state, and enter a
new union and have children after the separation*

7) What are the most frequent states one and five years after break-up? Use a modal state plot for illustration.

[Sol.]

```
par(mfrow=c(1,1))
seqmsplot(seqObj2, with.legend=TRUE, main="Modal states",
          xtlab=xtlab, ylab=NA, xlab="Time (months)")
```

Modal states



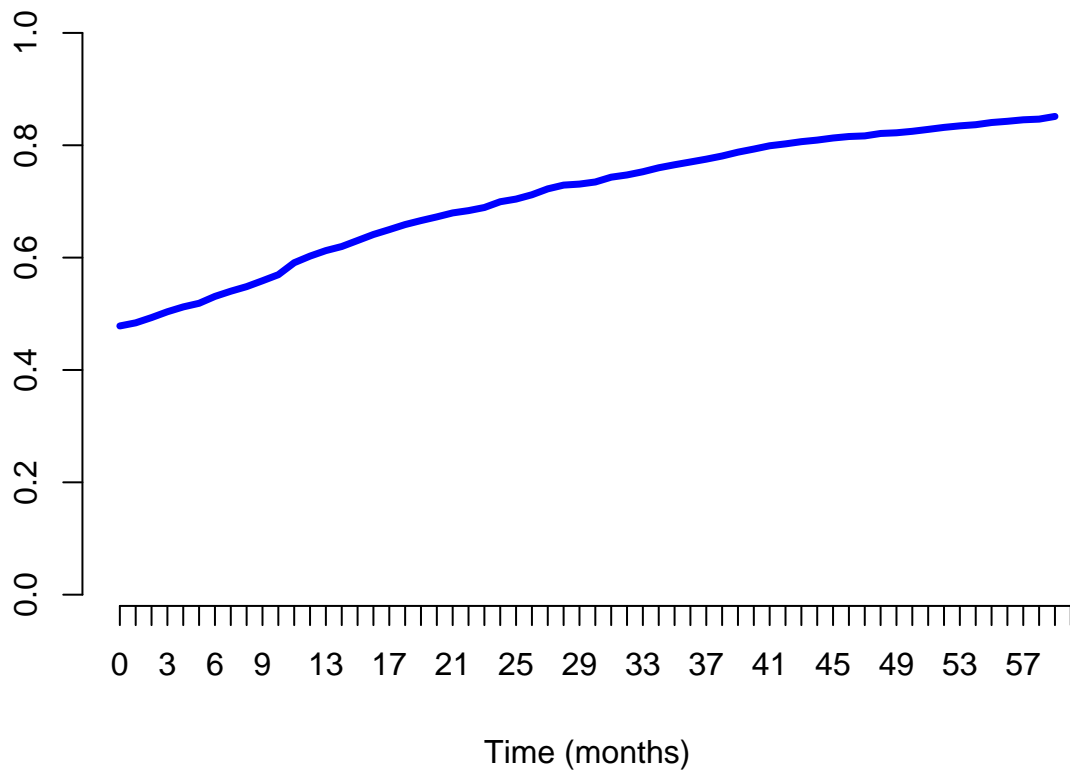
Before 4.5 months the most frequent state is single and childless
After 4.5 months the most frequent state is single and with a children
before separation

- 8) Assess the cross-sectional state diversity plotting a measure of entropy. At what time after separation is the cross-sectional diversity of the states at its highest?

[Sol.]

```
# Plot the transversal entropies in each position of the sequence
seqHtplot(seqObj2, with.legend=FALSE, main= "Transversal entropies",
          use.layout=FALSE, border=NA, xtlab=xtlab, ylab=NA, xlab="Time (months)")
```

Transversal entropies

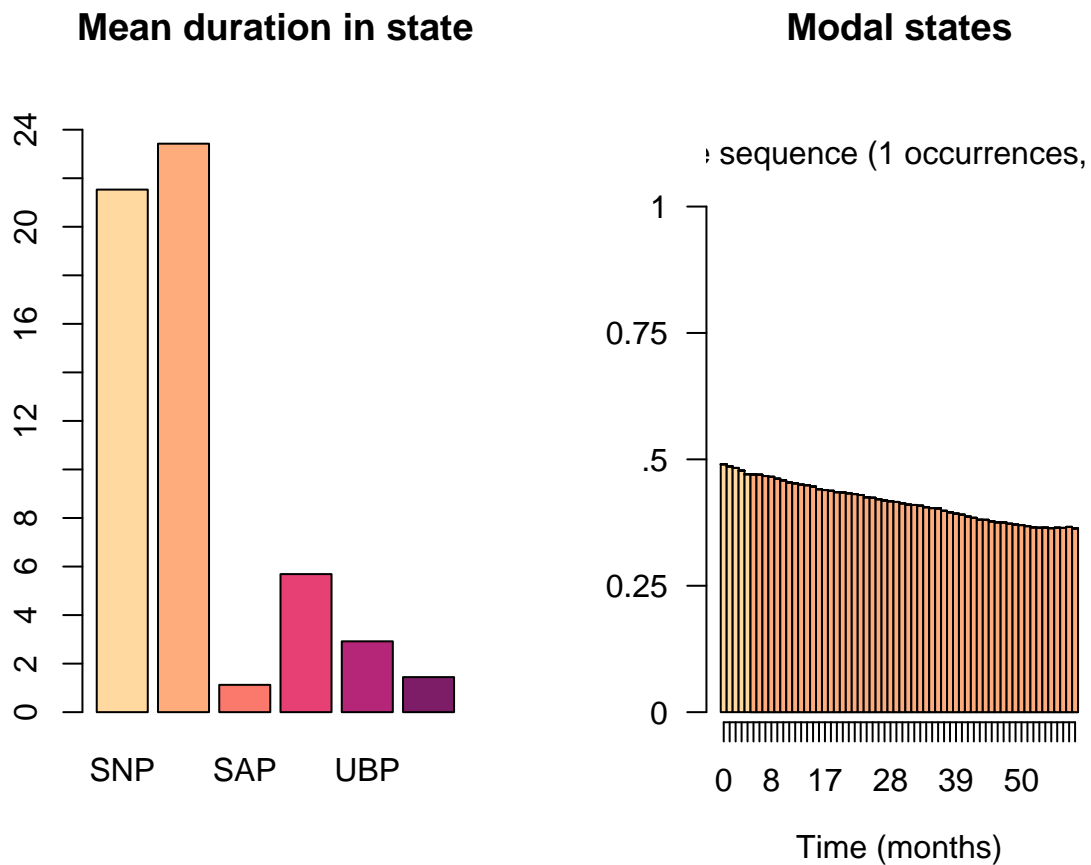


*# The diversity between states keep increasing and reaches its maximum
5 years after separation*

- 9) Display side by side in a same plot area the mean times spent in each of the states and the sequence of modal states.

[Sol.]

```
par(mfrow = c(1, 2))
# Plot the mean time spent in each state
seqmtpplot(seqObj2, with.legend=FALSE, main= "Mean duration in state",
            ylab=NA, ylim=c(0,25), yaxis=F)
axis(2, at=seq(from=0, to=25, by=2))
# Plot modal states in each position of the sequence
seqmsplot(seqObj2, with.legend=FALSE, main="Modal states", xtlab=xtlab,
            ylab=NA, xlab="Time (months)")
```



10) Compute the (overall) transition rate matrix. What is the largest transition rate between two different states?

[Sol.]

```
seqtrate(seqObj2)
```

```
# SNP -> UNP
```

```
# People who are single have more chances to become in union
```

11) Compute the sequence length, the number of transitions, the number of subsequences and the longitudinal entropy

[Sol.]

```
# Sequence length - number of elements with valid cases (print results for first five sequences)
```

```
length <- seqlength(seqObj2)
```

```
length[1:5]
```

```
# Number of transitions between state episodes in each sequence (print results for first five sequences)
```

```
transn <- seqtransn(seqObj2)
```

```
transn[1:5]
```

```
# Number of subsequences contained in a sequence
```

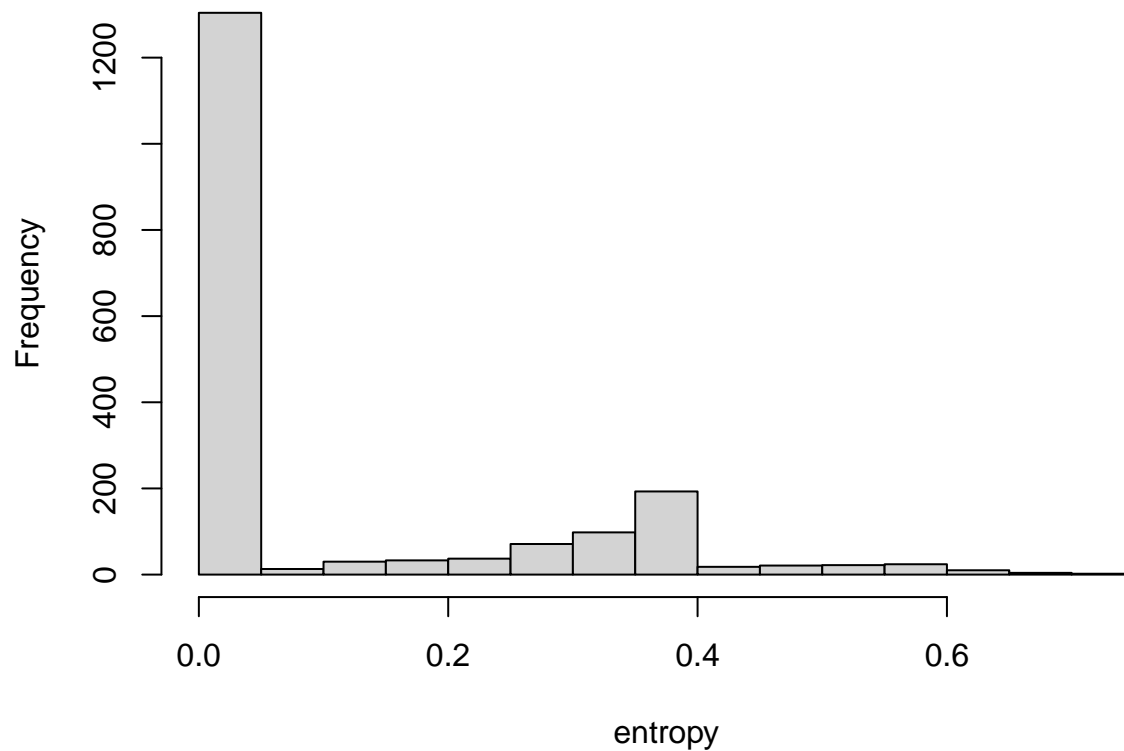
```
subseq <- seqsubsn(seqObj2)
```



```
table(subseq)

# Longitudinal or within-sequence entropy
entropy <- seqient(seqobj2)
par(mfrow=c(1,1))
hist(entropy)
```

Histogram of entropy



- 12) Using `summary()`, look at the min, max, mean, median and quartiles of the distribution of each of the computed longitudinal characteristics.

[Sol.]

```
summary(length)
summary(transn)
summary(subseq)
summary(entropy)
```

Exercise 2

[Sol.]

[Sol.]

References