Duration of Erasmus+ mobility periods

Ulrik Bernhardt Danielsen & Alexander Johansen Ohrt

UPC

May 18, 2022

Outline

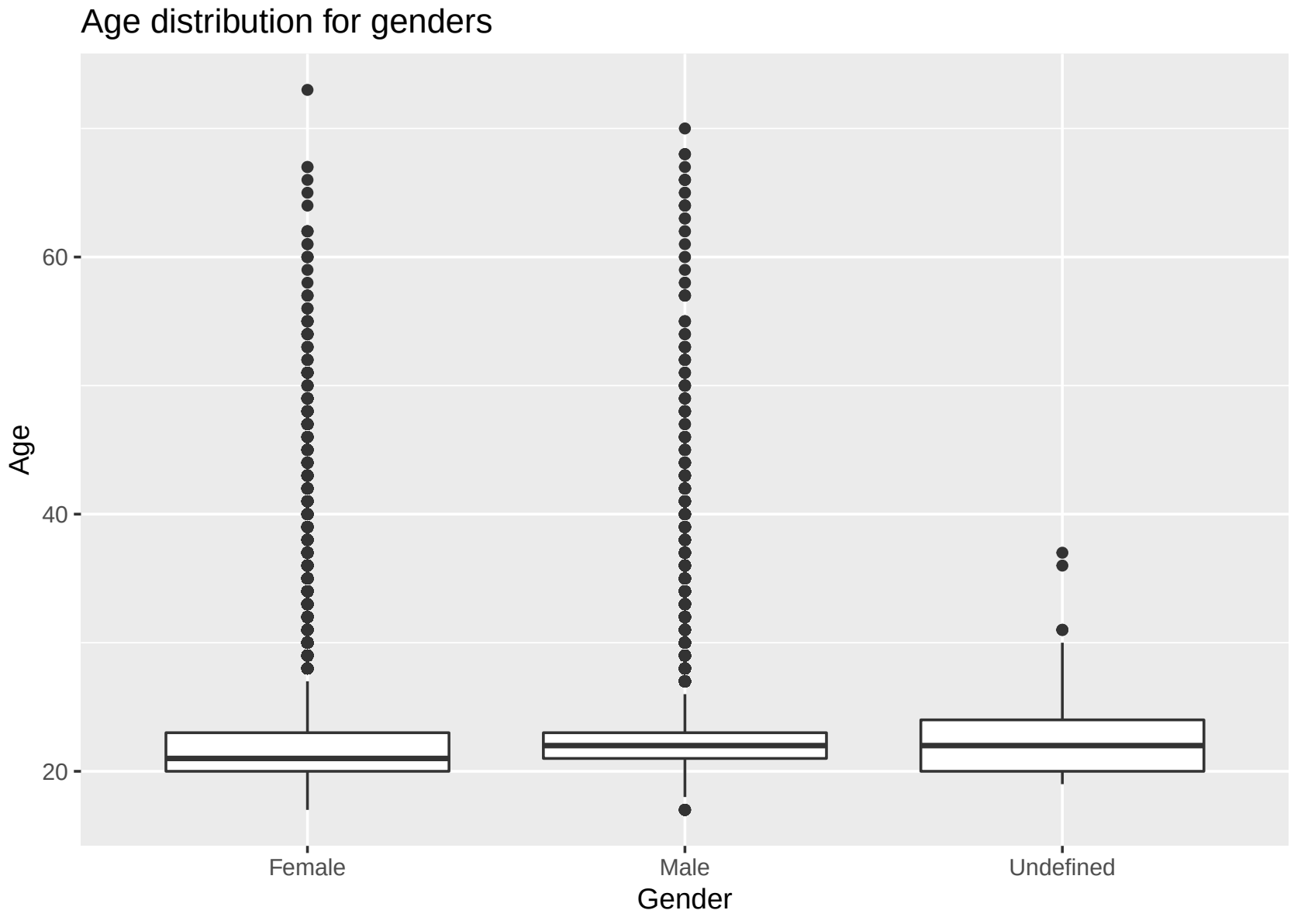
- Data
- Goal
- Models

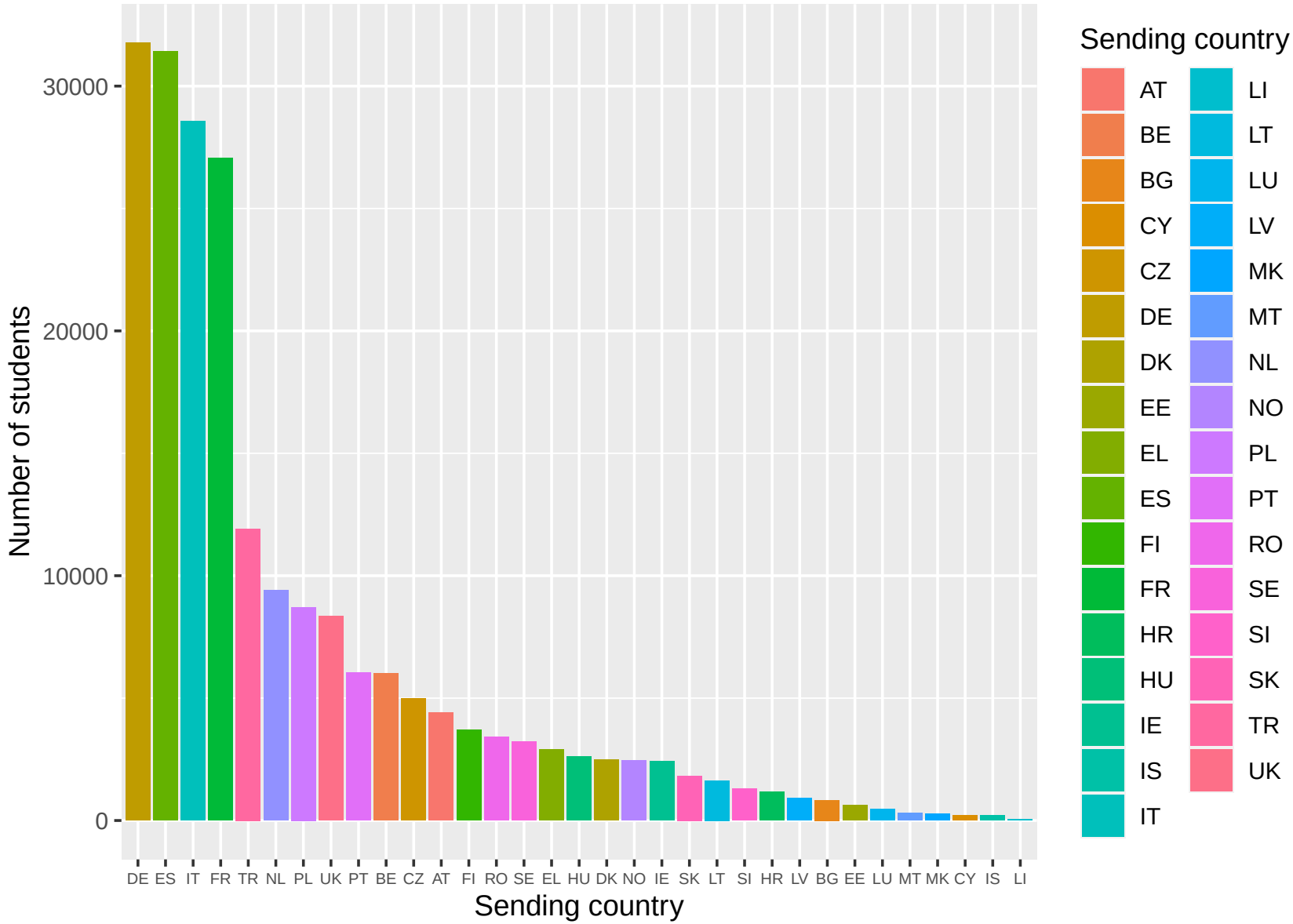
Data

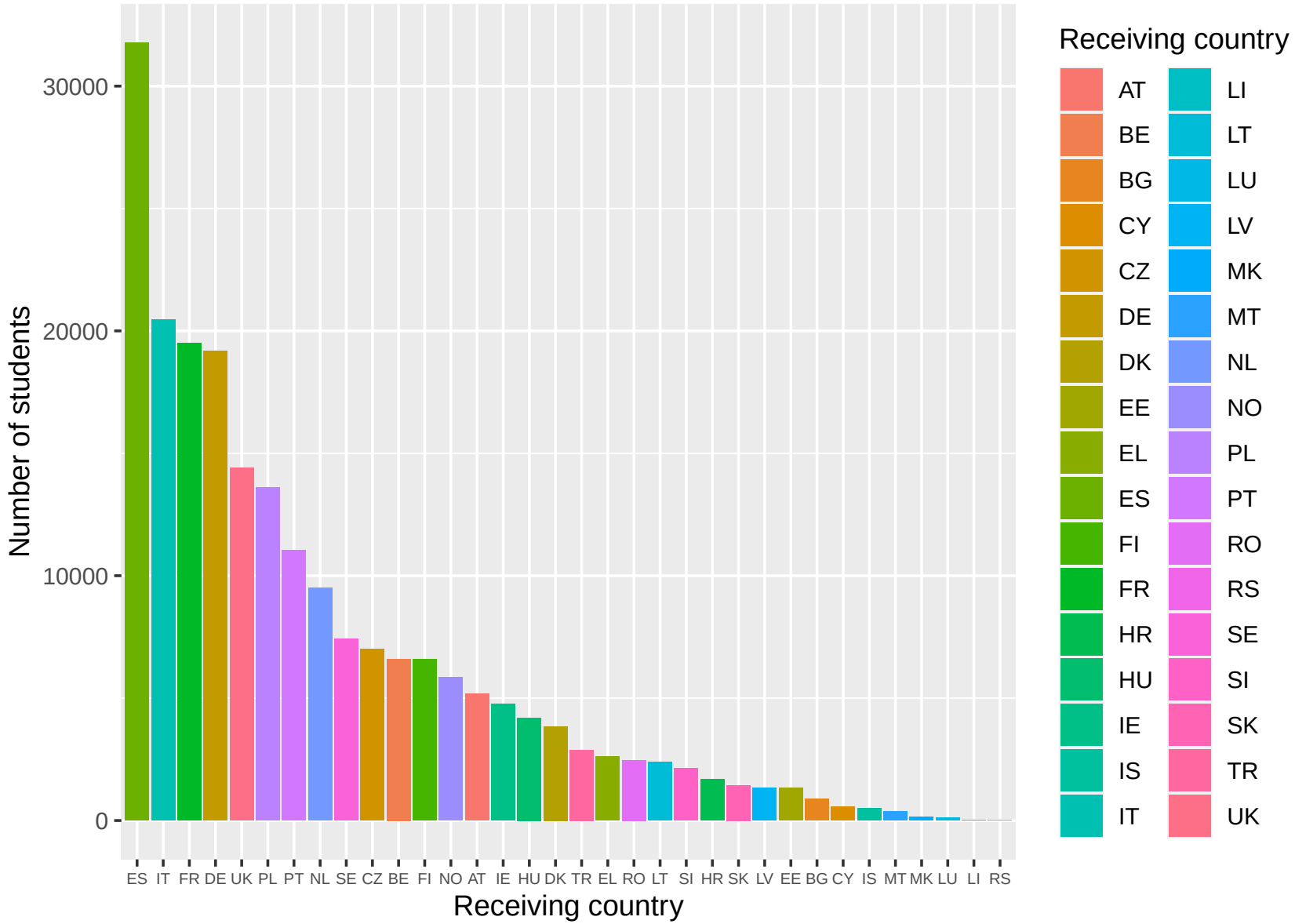
- https://data.europa.eu/data/datasets/erasmus-mobility-statistics-2014-2019-v2?locale=en
- All Erasmus+ students who finished their stay during 2019
- $\approx$ 730000 rows, 24 columns

Data

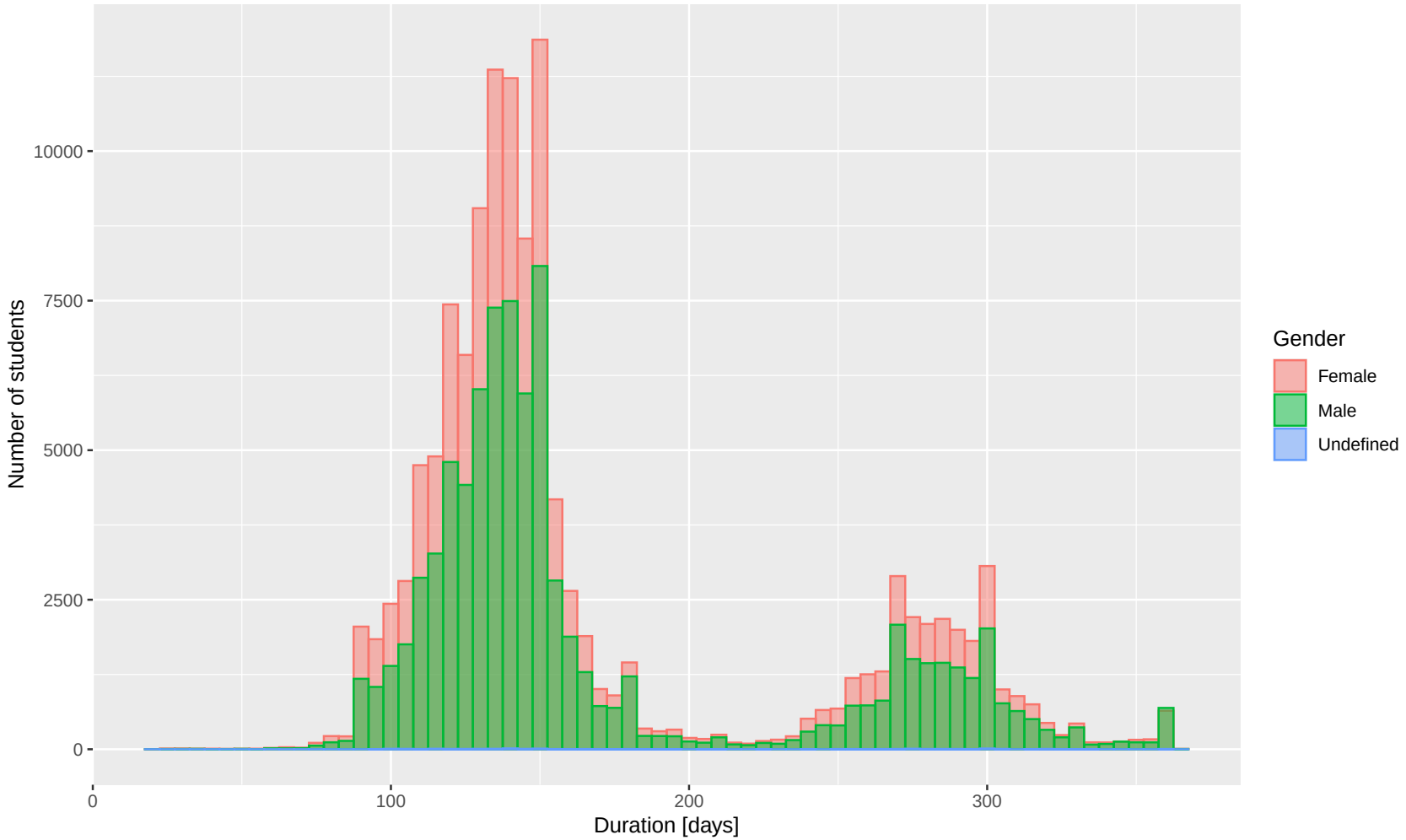
- Contains on students—age,gender,recieving country, sending country, ...
- ... and duration of mobility period
- Data cleaning $\rightarrow \approx 210000$ rows


Age distribution for genders
Age
60
40
20
Female
Male
Undefined
Gender

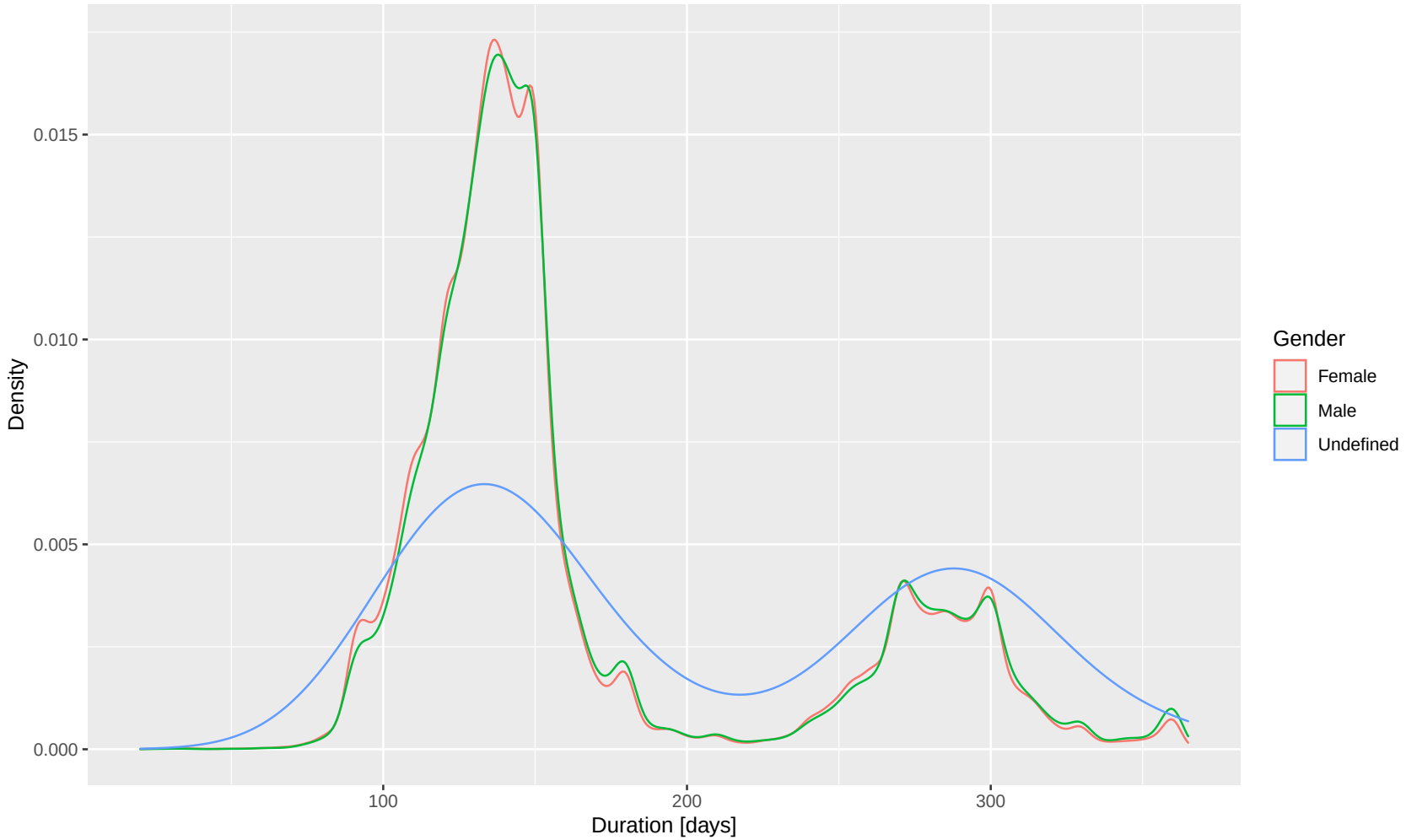

Number of students
30000
20000
10000
0
DE ES IT FR TR NL PL UK PT BE CZ AT FI RO SE EL HU DK NO IE SK LT SI HR LV BG EE LU MT MK CY IS LI
Sending country
Sending country
AT
BE
BG
CY
CZ
DE
DK
EE
EL
ES
FI
FR
HR
HU
IE
IS
IT
LI
LT
LU
LV
MK
MT
NL
NO
PL
PT
RO
SE
SI
SK
TR
UK


Number of students
30000
20000
10000
0
ES IT FR DE UK PL PT NL SE CZ BE FI NO AT IE HU DK TR EL RO LT SI HR SK LV EE BG CY IS MT MK LU LI RS
Receiving country
Receiving country
AT
BE
BG
CY
CZ
DE
DK
EE
EL
ES
FI
FR
HR
HU
IE
IS
IT
LI
LT
LU
LV
MK
MT
NL
NO
PL
PT
RO
RS
SE
SI
SK
TR
UK

Duration of Erasmus+ mobility periods



Duration of Erasmus+ mobility periods

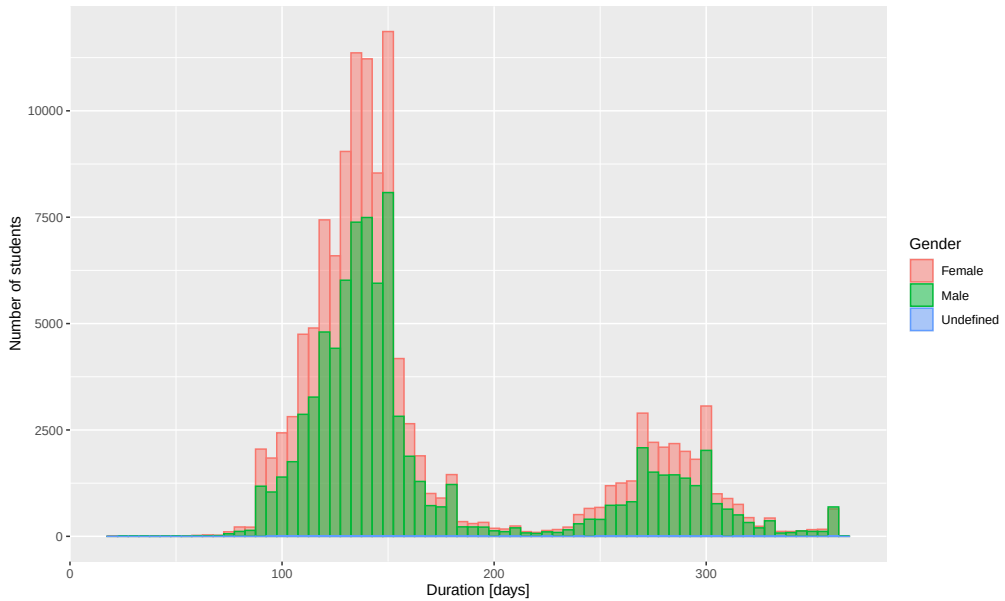


Goal

- Model the data using bayesian models
- Is there a difference for the genders?
- Does age matter?
- Is there a difference in mobility periods for countries?

Models

- Two (unknown) groups: finite mixture models
- Prior knowledge about duration length
- Regression



Bimodal mixture models

$$f(y_i|\theta, p) = p_1 f(y_i|\theta_1) + p_2 f(y_i|\theta_2).$$

Two approaches:

1. Latent variable z_{ih}
   - Multinomial with Dirichlet prior—hierarchical

     $$y_i|\theta, z_{ih} \sim f(y_i|\theta_{z_{ih}})$$

   - Does not work with Stan
2. Marginalize out z_{ih}
   - Using $\Pr(z_i = h) = p_h$ we obtain

     $$f(y_i|\theta, p) = pf(y_i|\theta_1) + (1 - p)f(y_i|\theta_2)$$

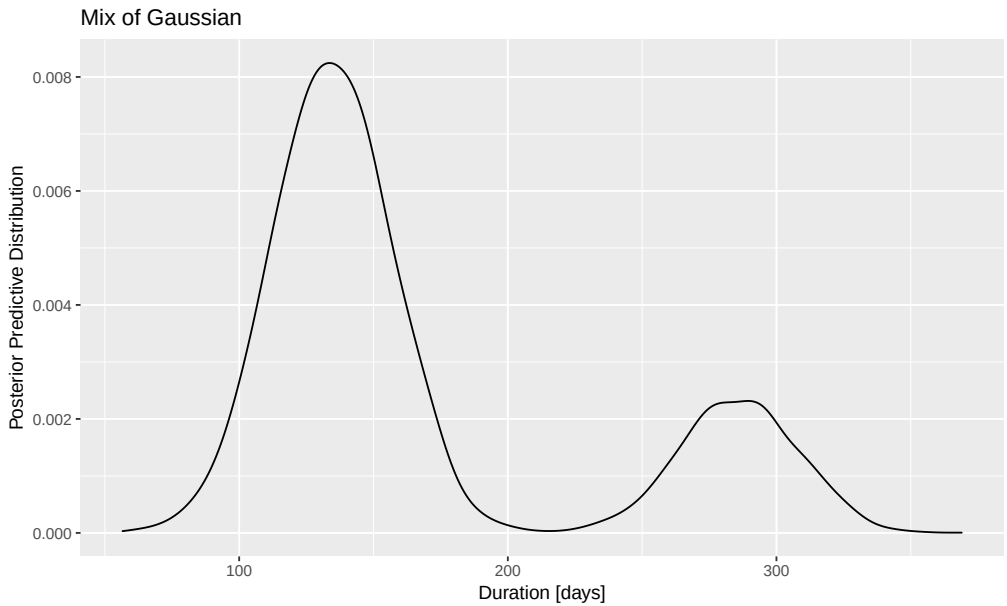
   - Here we let p have an uniform prior

Our models

- Bimodal Gaussian with equal variance
- Bimodal Gaussian with different variance
- Hierarchical bimodal Gaussian—variance is related between groups, but allowed to be different
- Regression model—age, gender, ...

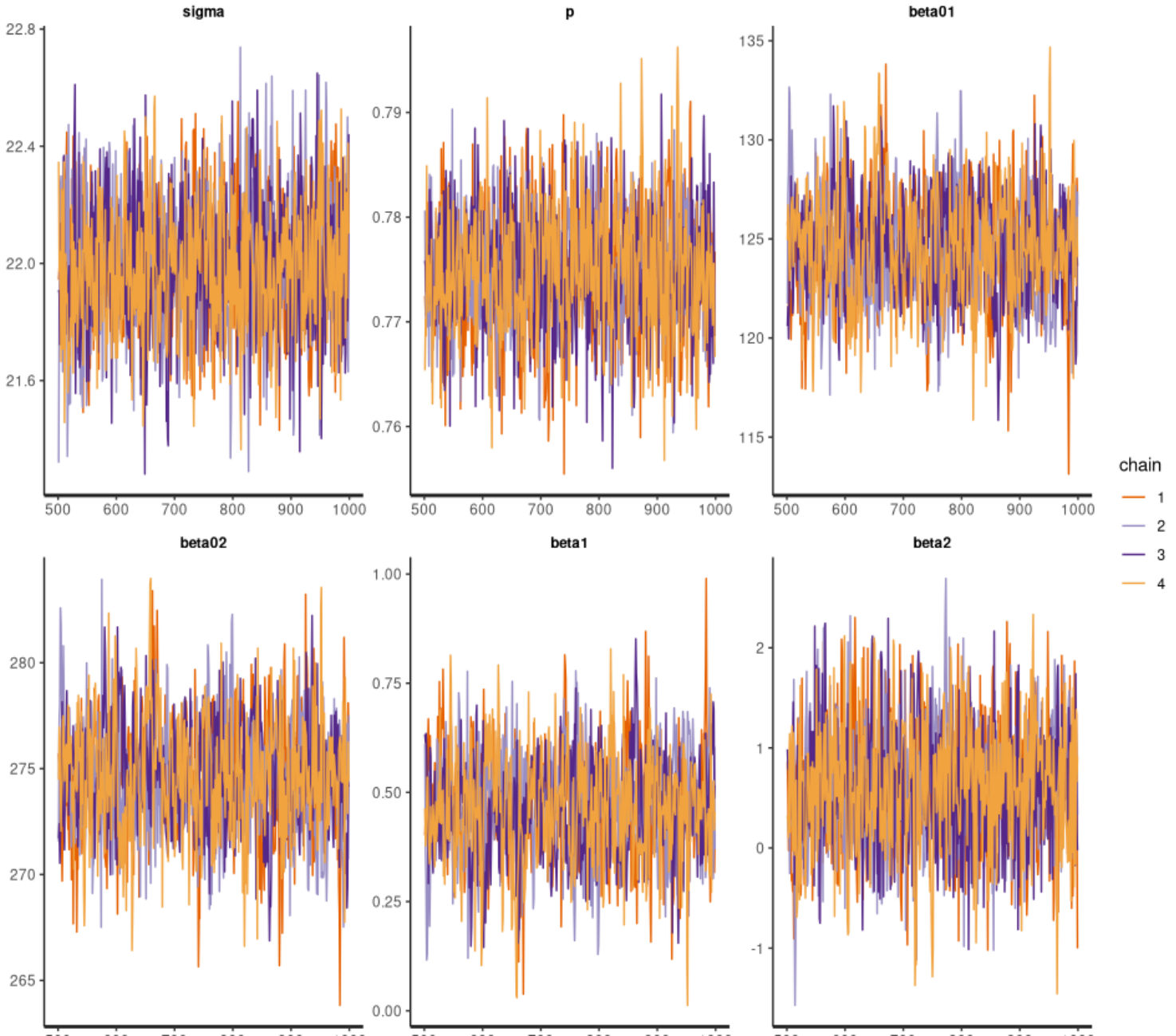
Initial model

$$\mu_1 \; N(120, \sigma)$$
$$\mu_2 \; N(280, \sigma)$$
$$\sigma \; InverseGamma(1, 1)$$
$$p \sim U[0, 1]$$



Regression model

$$y|x \sim pN(y|\beta_{01} + \beta_1 age + \beta_2 gender, \sigma_1) + (1-p)N(y|\beta_{02} + \beta_1 age + \beta_2 gender, \sigma_2)$$


sigma
p
beta01
beta02
beta1
beta2
chain
1
2
3
4