Uncovering age-gender patterns in migration: A Bayesian extension of the Rogers–Castro model for European migration corridors

Unpublished Manuscript
November 2025

Athina Anastasiadou[1], Andrea Aparicio Castro[2]

[1]Max Planck Institute for Demographic Research
Netherlands Interdisciplinary Demographic Institute (NIDI) - KNAW/University of Groningen

[2]University of Oxford, Leverhulme Centre for Demographic Science

Abstract

Migration is a highly age- and gender-selective process. While age patterns of migration have been extensively studied by demographers, most notably through the widely used Rogers–Castro schedule, systematic attention to gender differences in the age profiles of migration remains limited. In this study, we model a comprehensive dataset of bilateral migration flows among 30 European countries, disaggregated by age and gender, by developing a flexible Bayesian extension of the Rogers–Castro model. The multi-exponential specification is customized to the empirical shape of each flow's age distribution across 870 emigration corridors. From these fits, we obtain Rogers-Castro parameter estimates for every origin-destination-year-gender combination, which we then used to construct a typology of age-gender emigration profiles by implementing K-means clustering. Through this approach, we identify five distinct groups of age profiles in emigration, distinguished by the timing and intensity of life-course migration peaks. By analyzing these groups, we find that many emigration flows diverge substantially from the typical age structure proposed by Rogers and Castro. Moreover, female and male flows often display different age structures even within the same emigration flow or corridor. With our work, we reevaluate existing model schedules for migration age distributions and address a gap in the literature on how these might vary by gender. Additionally, we offer new insights into the age structure of migration flows in Europe by gender, time, and space.

1 Introduction

Understanding the interaction of gender and age in shaping migration dynamics is crucial for analyzing the demographic impact of migration on both sending and receiving countries. Yet, little is known about how international migration patterns vary simultaneously by age and sex.

Analyzing the distinct migration trajectories and outcomes for males and females requires data on both immigration and emigration, disaggregated by age and gender. While immigration-only data can reveal different entry patterns, understanding the balance of male and female migrants across corridors depends on analyzing both directions of movement. However, little attention has been given to the simultaneous disaggregation by age and gender (Raymer, Guan, Shen, Hertog, & Gerland, 2023). A widely used approach is the Rogers-Castro exponential model schedule, which provides age-specific migration patterns (Rogers & Castro, 1981). Yet, this framework considers gender as secondary, rather than integrating age and gender characteristics jointly.

As age and gender compositions have likely changed over time due to globalization, emancipation, and increased mobility, we believe it is time to re-assess international migration patterns systematically by age, gender, time, and space. Previous findings in the migration and gender literature have shown that female and male migrants exhibit different migration motivations and propensities, and experience different outcomes of migrating (Anastasiadou, Kim, Sanlitürk, de Valk, & Zagheni, 2024). Consequently, the gender compositions of corridors are owed to a complex and dynamic interplay between conditions in origin and destination countries, and they are subject to changes over time.

With this work, we aim to answer the following questions: (1) What distinct shapes of age-specific emigration profiles can be identified across European corridors using a flexible Bayesian Rogers–Castro framework? and (2) can these shapes be grouped into meaningful types, and do these groups exhibit systematic variation by gender, time, and space? By answering these questions, we also reassess the shape and validity of the long-established Rogers-Castro schedule in capturing contemporary emigration patterns that explicitly incorporate both age and gender.

2 Background

2.1 The role of age and gender in migration patterns

It has long been recognized in demography that an individual's likelihood of migrating is closely linked to their age and hence, stage in the life course (D. B. Pittenger, 1974). This is due to several reasons. First, age has been found to be a critical determinant for migration behavior due to migration decisions being closely tied to life-course events such as education, labor market, family formation and retirement (Rogers & Castro, 1981). Secondly, the origin population's age profile strongly determines the age profile of its emigrants, since the pool of potential migrants reflects the demographic composition of sending countries (Little & Rogers, 2007). In other words, variations in the age profiles of out-migrants can be linked to variations in the age composition of the population they originate from (Little & Rogers, 2007).

Migration intensity varies systematically across the life course. While under 10-year-old children mimic migration patterns from their parents, migrations due to marriage usually peak at 30. Furthermore, moves due to employment appear during the early years of labor force participation and health motivates migration of the elderly. These cause-specific age pattern can be attributed to different phases of an individual's life and hence can be interpreted within a life course framework (Raymer, Rogers, et al., 2008). Besides the motivation of the individual to move, the family has been recognized as a unit that creates dependencies among its members and consequently shapes the age profiles of migration (Raymer et al., 2008).

Among the first ones to explore dependency structures and formalize them for population projections are Castro and Rogers (1983). The authors explore the family and migration relationship, acknowledge the dependency of individual migrations mathematically, and develop a dependency ratio for family migrations. When exploring Mexican census data, the authors observe that females exhibit lower headship ratios and larger dependency ratios than male migrants (Castro & Rogers, 1983). They also notice that age profiles of dependent female migrants differ from those of the corresponding general population, while male profiles look rather similar. Moreover, the age profiles of female migrant heads are bimodal while the modes in profiles for the corresponding female heads in the general population are less pronounced. Also, the male age profiles of dependents are steeper than the profiles of the females. With their work, the authors highlight the importance of the family life cycle to explain variations in age patterns among migrants.

While the spatial, temporal, and age dimensions of migration have received increasing attention in the literature, the gender[1] dimension has been less systematically integrated into migration modeling (Collinson, 2017). Yet, it clearly intersects with the other three migration dimensions (geography, time, and age). Reasons for migration can be distinct by gender, as females and males face distinct motivations, encounter unique experiences during their journeys and at their destinations, rely on different social networks, and achieve varying levels of integration (Anastasiadou et al., 2024). The demographic implications have been shown by Guttentag and Secord (1983) , who stated that selective migration is the primary source of imbalanced sex ratios.

Probably the biggest obstacle in understanding these gendered relationships and structures is the scarce availability of gender- and age-disaggregated migration flow data. Flow data are essential for analyzing migration change over time because they capture the timing and composition of migration events. Unlike stock data, which reflect the accumulated presence of migrants shaped by fertility, mortality and aging, flow data reveal patters of age-sex selectivity and make it possible to observe dynamics in response to policy, economic and social changes. (Donato & Gabaccia, 2015; Abel & Cohen, 2019; Abel & Sander, 2014; Abel, 2013). Many countries rely on census data to report migration figures, though some use surveys or population registers. However, stock data has significant limitations. It often fails to reflect current migration trends, especially by gender, unless detailed questions about migration timing and

[1] Gender relations — rather than biological sex — shape the relative numbers of female and male migrants (Donato & Gabaccia, 2015). For this reason, we use the term *gender* rather than *sex* throughout the article. However, the dataset employed in our analysis records binary sex, not gender, so both terms will appear. We recognize that current migration data collection systems typically capture sex rather than gender, which constrains our analysis. A comprehensive understanding of gender dynamics in migration would require rethinking how migration data are conceptualized and collected.

birthplace are included. Stock data also overlooks temporary migration, which is typically male-dominated and occurs between census periods. Additionally, migrant stocks may over-represent women due to the aging of migrant populations. As a result, our understanding of gender in migration is constrained by the limits of the available data (Donato & Gabaccia, 2015).

2.2 Understanding regularities in migration patterns

Deriving typologies based on observed characteristics of destinations and origins of migration flows[2] is not new. However, most of them were developed a long time ago or focus only on age or internal migration.D. B. Pittenger (1974) developed a typology of age-specific net internal migration schedules based on place characteristics, which he later applied in D. Pittenger (1978) to produce age-specific net migration projections. The author observed that the age at which regional net migration rates peak varies between females and males. To better understand these dynamics, D. B. Pittenger (1974) developed theoretical age-specific migration profiles based on the socioeconomic characteristics of different areas and known patterns of rural-to-urban or inter-regional migration, and then tested these profiles using empirical data from the US.

Another effort that laid the groundwork for understanding age-profiles of migration was undertaken by Rogers and Castro (1981). The authors analyzed 524 observed age profiles of migrants and demonstrated that migration has strong regularities in age patterns much like fertility and mortality. Like the study by D. B. Pittenger (1974), the work by Rogers and Castro (1981) utilizes data on internal migration, albeit from 17 different countries. The most influential contribution of their work is the mathematical formulation of the discovered regularities in the age-distribution of migration, known as the Rogers-Castro model migration schedule. Moreover, the authors identified several families of such schedules. Namely, schedules with retirement peaks or post-retirement up slopes, early-peaking vs. late-peaking labor force age components, symmetrical vs. asymmetrical labor-force age profiles, child-dependent vs. labor-dominant age patterns, and parallel slopes for the child-parent age groups (typically separated by the mean age of generation "parental shift") (Rogers & Castro, 1981). These families of schedules can be further categorized by a number of key parameters (Rogers, Castro, & Lea, 2005).

The Rogers–Castro model schedules offer several advantages for analyzing migration patterns (Rogers, Little, & Raymer, 2010). First, they simplify complex age-specific migration data by reducing the 85 individual age values in a typical migration profile to just a few key parameters, depending on the version of the model used. These parameters are not only efficient but also interpretable and meaningful, representing distinct components of the migration process (such as childhood moves, labor force mobility, or retirement migration). Additionally, as the parameters capture essential features of the migration profile, they allow for easy comparison of migration patterns across different populations, regions, or time periods (Rogers et al., 2010).

[2]We refer to *flows* to describe the number of people changing their country of residence between two time periods. By *corridor*, we refer to all origin-destination pairs that have a migration connection over the study period.

While the work by Rogers and Castro (1981) has motivated analyses along similar lines, for instance by Plane and Heins (2003) who offer another approach to identify seven groups of migrant age profiles of inter-metropolitan migrants in the US, they remain the only work that assessed internal migration patterns for a larger number of countries. And their multi-exponential model migration schedules have since been used to describe, smooth, and infer age-specific migration patterns across many different contexts (Raymer et al., 2008).

Age-sex profiles of migration have been studied by Collinson (2017). The author asserts that regularities in age-sex patterns of migration are shaped by underlying biological, psychological, and socio-cultural behaviors just like mortality and fertility. However, social norms rather than biological characteristics influence the numbers of females and males migrating (Donato & Gabaccia, 2015). Collinson (2017) study is exceptional as it focuses on locations in the Global South.

Age-gender distributions of migration corridors are dynamic processes evolving over time. The data that was used by Rogers and Castro to discover age regularities in migration are internal migration data in the "developed nations" [sic!] (Raymer et al., 2008). However, international (as well as internal) migration patterns have likely changed over the past decades due to globalization and increased human mobility. And so do the profiles of the migrants which have likely changed due to emancipation and growing networks. The patterns of the Rogers-Castro multi-exponential migration schedule has rarely been assessed systematically by gender, and application to international migration in particular remain limited (see e.g. Raymer et al. (2008); Shen, Raymer, Guan, and Wiśniowski (2024)). With our work, we aim to contribute to reassessing the shape of the Rogers-Castro schedule, as well as to shed light on the understudied gender dimension of migration. To the above outlined body of literature, our work contributes in three main ways. (1) We extend the analysis of migration profiles by adding selectivity by gender alongside spatial, temporal, and age dimensions. (2) We develop a flexible Bayesian extension of the Rogers-Castro model schedule for migration, which can capture atypical age distributions and allows us to model a wide variety of observed migration shapes. (3) We identify regularities across 20,460 age-gender-specific migration flows by clustering the estimated schedules, thereby deriving a typology of European migration corridors that highlights systematic variation by gender.

3 Data

3.1 QuantMig bilateral migration flows by age and gender

To address our research questions, we employ bilateral emigration flow estimates for 33 European Union (EU) and European Free Trade Association (EFTA) countries, as well as North Macedonia (of which we use 30 for the analysis), derived as part of the *QuantMig* project (Aristotelous, Smith, & Bijak, 2022). The dataset provides annual bilateral migration flows, covering immigration, emigration, and net migration, over the eleven-year period from 2009 to 2019. We decided to focus the analysis on emigration flows as the original Rogers-Castro migration schedules were derived on emigration rates. The flows are provided as counts disaggregated into 17 five-year age groups (from $0-4$ to $85+$) and by gender (female and male).

Figure 1 shows an example of a migration flow included in the data set.

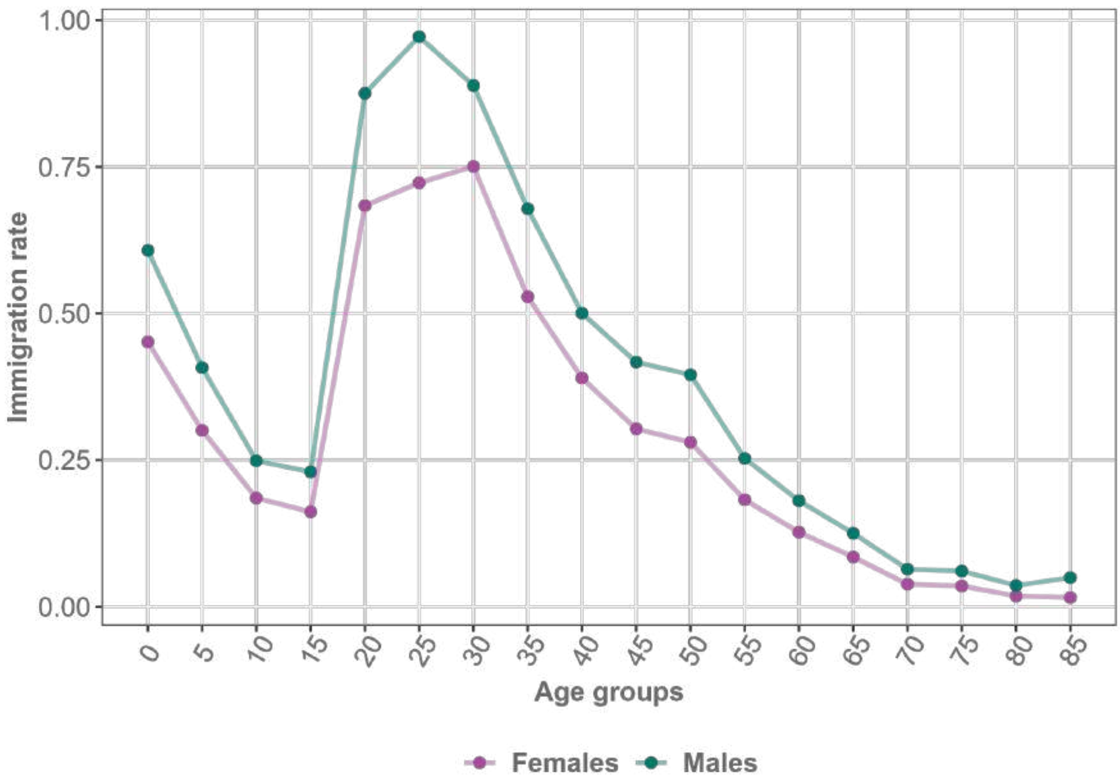

Figure 1: Flows from the United Kingdom to Malta in 2019 by five-year age group and gender reported by Malta.

In order to estimate migration flows, Aristotelous et al. (2022) utilize three sources of information within a Bayesian Hierarchical framework: (1) flow data reported by the sending and receiving countries reported in Eurostat, (2) measurement features of these data, which account for differences in duration-of-stay criteria, undercount, and data accuracy across national systems, and (3) migration covariates, including economic, geographic, and demographic factors (Aristotelous et al., 2022). Eight of the 33 countries do not provide any flow data. Difficulties in producing this harmonized data set arise from inconsistent definition of migration across Europe, variation in data collection systems (e.g. registers vs. surveys), and inclusion or exclusion of specific migrant subgroups in national statistics.

The gender breakdown is already provided in Eurostat and is not estimated. Therefore, we consider the gender and age breakdown in these estimates reliable. Moreover, the input data have been carefully assessed by the authors of the data set. We calculated immigration and emigration rates based on population estimates from the World Population Prospects (WPP) (2024) (United Nations, Department of Economic and Social Affairs, Population Division, 2022). We removed Luxembourg, Liechtenstein, Malta, and the residual category 'Rest of the World' (ROW). For Luxembourg and Liechtenstein, atypical migration patterns make their inclusion problematic. In contrast, ROW is a heterogeneous aggregate. After these exclusions, the dataset encompasses $20,460$ origin-destination-year-gender combinations, each disaggregated into 17 five-year age groups.

The estimates provided by Aristotelous et al. (2022) also come with a number of limitations. Migration is typically undercounted because countries often apply different definitions of migration and may exclude certain groups, such as refugees, asylum seekers, or those without legal status . Even within the EU reporting practices vary, with some countries using the UN's 12-month residence criterion and other relying on alternative definitions. As a results, the reported values for immigration and emigration usually differ greatly. The Bayesian model corrects for under- and overcounting based on expert knowledge. Nevertheless, for countries that do not provide any data, estimates are derived from drivers of migration, triggering high uncertainty.

3.2 (Atypical) patterns in migration flows in Europe

The observed emigration profiles in the *QuantMig* dataset often deviate from the text book version of the Rogers-Castro model schedule. While most flows display the typical two to three peaks in their distribution (see Figure 2), a high number of flows exhibit additional peaks, indicating more complex age-specific emigration patterns than the standard model suggests. These complexities become even more distinctive when comparing the age-specific distributions of female and male migrants.

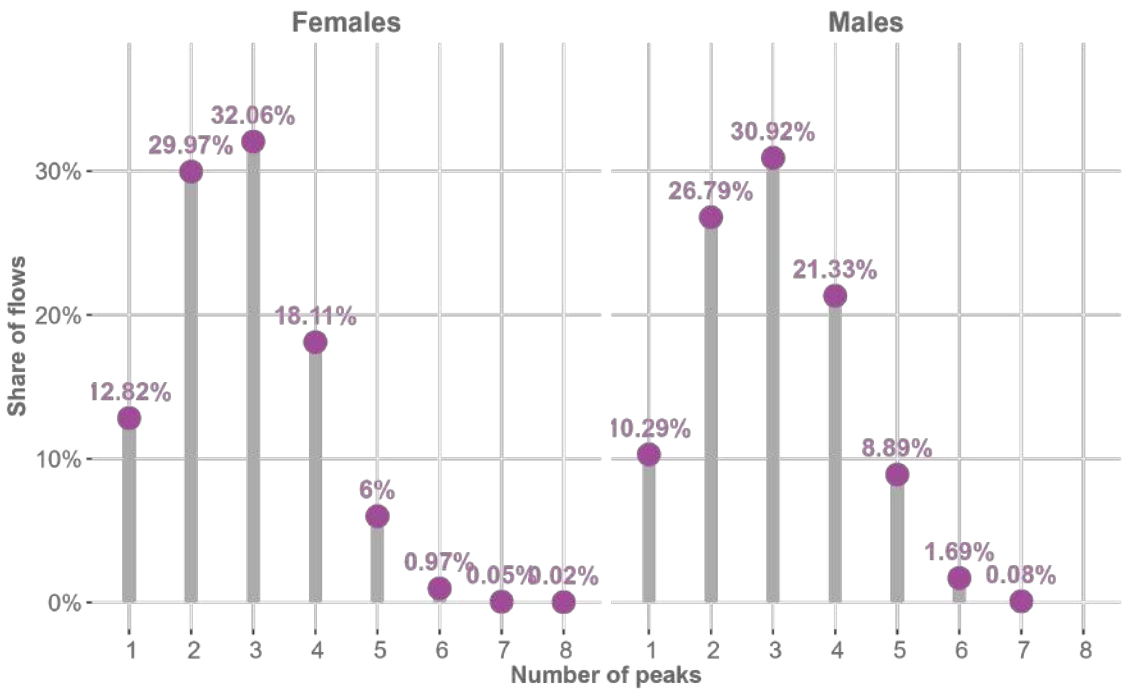

Figure 2: Frequency of number of peaks across emigration flows in percent of all flows. For all origin-destination-gender-year combinations in the data set and by gender.

Figure 3 shows that also the ages at which migration peaks occur differ from the text book Rogers-Castro examples. While most peaks are concentrated in the working ages $25-30$, many flows also display in childhood ages $5-10$ and a sizable share of flows in the student age group $20-25$. More generally, with the exception of ages 10–15 and 15–20, a considerable number of flows exhibit peaks across nearly all age groups. This suggests that the typical Rogers-Castro model migration schedule may be insufficient to capture the full range of patterns observed in the data.

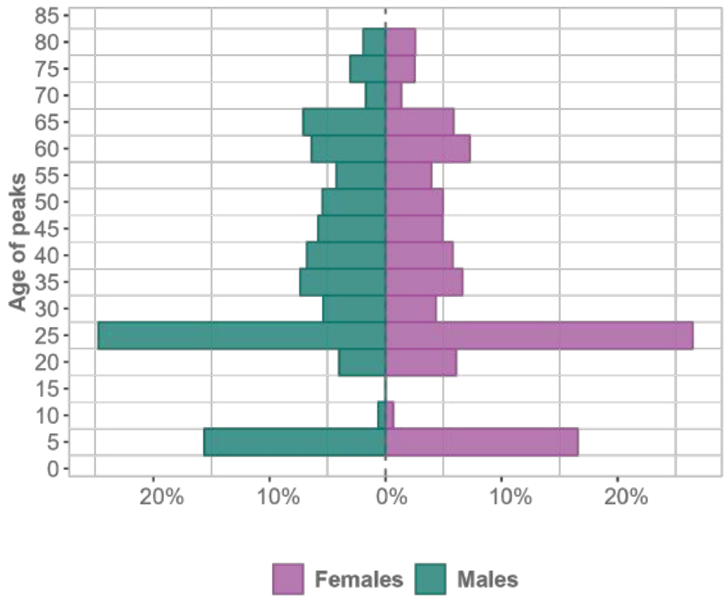

Figure 3: Age groups at which peaks occur in emigration rates by gender.

4 Methodology

4.1 Identification of individual shapes of age distributions

As shown in the previous section, the age profiles in the data deviate from the typical pattern assumed by the Rogers-Castro multi-exponential model schedule. For instance, many flows exhibit peaks during child ages, yet the schedule's childhood component does not allow such peaks to be modeled. Another deviation from the typical setup is the occurrence of peaks around student ages or early work life. Wilson (2010) proposed an extension of the Rogers-Castro model schedule that incorporated a symmetric unimodal curve centered on the average age of university entry. As of now, we are not aware of a method that proposes a flexible multi-exponential model specification based on the ex ante detected shape of the respective age distribution of migration rates.

To assess the shapes of the curves, we propose a function that takes migration rates and the corresponding age groups as input and identifies the number and location of modes in a curve. Modes are detected as local maxima as outlined in equation 1. In addition to clear peaks, many curves also contain belly-shaped concave regions that are not captured as local maxima. These are instead identified by assessing changes in slope, specifically, the point at which a relatively flat negative slope becomes distinctly steeper. The exact conditions that have to be satisfied are outlined in equation 2.

$$\begin{aligned} &returns\ x_i\ where \\ &i \in \{2, \dots, n-1\} | m_i > m_{i-1}\ and\ m_i > m_{i+1} \quad (m_i \text{ is a local maximum}) \end{aligned} \tag{1}$$

$$
\begin{aligned}
&returns\ x_i\ for i \in \{2, \ldots, n-1\}\ where \\
&(1) \quad |\Delta m_i| < |\Delta m_{i+1}| && \text{(slope is increasing in magnitude)} \\
&(2) \quad \Delta m_i < 0\ and\ \Delta m_{i+1} < 0 && \text{(both differences are negative)} \\
&(3) \quad \Delta m_{i-1} \leq 0 && \text{(previous slope is not positive)} \\
&(4) \quad |\Delta m_{i+1} - \Delta m_i| > \frac{\sigma_m}{10} && \text{(change in magnitude is significant)}
\end{aligned}
\tag{2}
$$

It should be noted, that condition (4) in equation 2 disregards non-meaningful concave patterns in the curve, i.e. changes in rates that are small relative to the standard deviation of the rates in this flow. This implies losing some accuracy but helps avoid over-parameterization and potential convergence issues.

4.2 Fitting customized RC schedules to each flow

We model emigration rates m_i using a Bayesian Normal model as proposed by Yeung, Alexander, and Riffe (2023). Their specification includes four migration components: a descending curve for child ages, unimodal curves for working ages and retirement, and a post retirement component. Our approach extends this models by allowing flexible specification of these components based on the shapes detected in the previous step (i.e. the numbers of peaks and concave regions).

The process model is defined as a sum of migration components, where its number depends on the total of detected peaks P and an additional baseline c.

$$
\begin{gathered}
for\ i \in 1, \ldots, N\ and\ p \in 1, \ldots, P \\
\textbf{Process model:} \quad \mu^{RC}(x_i) = \sum_{p=1}^{P} \mu_p^{RC}(x_i) + c
\end{gathered}
\tag{3}
$$

If no childhood peal is detected (i.e. a peak at $x < 18$), the model includes only the working ages, retirement, and post-retirement components as follows:

$$
\begin{aligned}
\mu_1^{RC}(x_i) &= a_1 exp(-\alpha_1(x_i - m_1)), \quad with\ \mu_1 = 0 \\
\mu_{p+1}^{RC} &= a_{p+1} exp(-\alpha_{p+1}(x_i - \mu_{p+1}) - exp(-\lambda_{p+1}(x_i - \mu_{p+1})))
\end{aligned}
\tag{4}
$$

If a pronounced childhood peak is detected, an additional pre-working-age components is added (i.e. $\mu_1 \neq 0$) as shown below:

$$
\mu_p^{RC} = a_p exp(-\alpha_p(x_i - \mu_p) - exp(-\lambda_j(x_i - \mu_p))) \tag{5}
$$

The parameters of the multi-exponential equation have purposes that allow for later comparison between flows and interpretation. The a_p parameters usually indicate the height of curve component p, λ_p represents the rate of ascent of a curve p, α_p indicate the rate of descent of a curve p, while μ_p denotes the position of curve p on the x-axis, and c represents the baseline migration rate.

Migration rates $\mu^{RC}(x_i)$ are assumed to follow a normal distribution around the Rogers-Castro specification with standard deviation σ as shown below:

$$m_{i,j,g,t} \sim Normal(\mu^{RC}(x_i), \sigma), \tag{6}$$

where $i = 1, \ldots, 33$ denotes the origin countries, $j = 1, \ldots, 33$ destination countries (with $i \neq j$), $g \in \{male, female\}$ denotes the two genders, and $t = 2009, \ldots, 2019$ the year in which the respective migration flow occurs.

For priors, we replace the fixed-location priors of Yeung et al. (2023) with priors informed by the detected shapes. For example, instead of the prior for the average age of migration in the working ages μ_2 centered around 25 and the prior for the parameter for the average age of retirement migration centered around 65, our approach used the values determined via the shape function as means of the prior distributions of the mode parameters. Also, for parameters α and λ we use only the positive values of a Cauchy distributions that allow to have fatter tails instead of the Normal distributions proposed by Yeung et al. (2023).

$$\begin{aligned}
&\textbf{General priors:} \\
&c \sim Normal_{+}(min(m_{i,j,g,t}), 0.1) \\
&\sigma \sim Cauchy_{+}(0, 2) \\
\textbf{Customized priors:} \quad &for\ all\ p \in \{1, \ldots, P\} \\
&a_p \sim HalfNormal(0, 0.5) \\
&\alpha_p \sim HalfCauchy(0, 1) \\
&\lambda_p \sim HalfCauchy(0, 1) \\
&\mu_p \sim Normal(x_p^{peak}, 1)
\end{aligned} \tag{7}$$

This flexible model specification was applied to all origin-destination-year-gender combinations independently by gender and the resulting parameter estimates were extracted for the next step.

For models that did not converge under the above specifications, we replaced the Normal distributions for the α and λ priors with a Student-t distribution with 5 degrees of freedom $HalfT(5, 0, 1)$. That helped a notable number of models to converge and left us with about 9.6% of outflows that did not converge under both specifications. These were removed from further analysis.

4.3 Assessing patterns in age-distributions

To identify similarities in age profiles, we applied K-means clustering to the parameter estimates from the implemented Rogers–Castro models. Each migration flow (by origin, destination, gender, and year) is represented as a vector of parameter values as follows:

$$V_{i,j,g,t} = \{a_1, \alpha_1, \lambda_1, \ldots, c\} \tag{8}$$

These vectors are derived for each origin-destination-gender-year combination. Depending on the detected shape and hence, the number of modeled peaks by life stage in the flow, the length of these vectors varies. For means of comparison, the missing parameters are padded with zeros. This approach does not impede interpretability and comparability of the estimates. Because a missing slope can be equated with a slope of zero for interpretation purposes. We adopt a life-course perspective by arranging the parameters according to the life stages in which they occur. We refer to *childhood* when the peak occurs before the age of 18 (1), to *early working ages* when the peak occurs between the ages of 18 and 39 (2), to *late working ages* when the peak occurs between the ages of 40 and 65 (3), and to *old ages* when the peak occurred in age 65 and beyond (4). We enumerated the parameters accordingly with for instance μ_{2a} indicating the first peak in the early working ages.

Clustering is a common approach for detecting similarities in the data. For migration flows, Little and Rogers (2007) developed a method to predict the most likely age-profile of a region's emigrants based on its population's age distribution. They used clustering as a tool to group schedules by age patterns and then predict model schedule cluster membership from population measures. In a similar manner, we identify groups of age patterns in emigration, which we the examine for systematic differences by gender, accounting also for variation in time and space.

5 Results

5.1 Typology of age structures in emigration flows

The K-means clustering algorithm identifies five distinct groups of age distributions among emigration flows based on their Rogers-Castro parameter estimates (see Figure 4). All groups exhibit pronounced peaks in the early working ages, while they seem to appear slightly wider for groups one, three, and four.

The first panel contains a group of flows with moderately high emigration rates which peak in the early working ages, preceded by relatively high rates of migration in the childhood ages. After that the curves seem to decline fast with only small upward swings in the older ages. In this group, the predominant country of origin is Cyprus and the predominant destination country is Bulgaria. The flows with the largest emigration rates encompass Ireland to Great Britain, followed by mainly European South East to North flows (e.g. Croatia-Germany, Cyprus-Great Britain, and Estonia-Finland). These are examples for long established student and labor migration corridors.

The second group consists of flows with mostly moderately high emigration rates as well as two outlier cases which are males emigrating from Lithuania to Great Britain in 2015 and 2016. This group includes flows with strong peaks in the early working ages. The predominant origin country is Croatia and the predominant destination Hungary. The largest emigration flows include flows from smaller European countries to larger ones (e.g. Lithuania-Great Britain, Bulgaria-Germany, Latvia-Great Britain, Slovenia-Germany). Again these are indicators for corridors of labor migration.

The third group is the smallest group in size as well as in migration intensity. The majority of the flow exhibit rather low emigration rates. Rather downward slopes than peaks are visible in the childhood ages. The most prominent origin country within this group is France and the most prominent destination country is clearly Denmark. The largest flows in this group are

Iceland-Norway, Belgium-Spain, Romania-Spain, Portugal-Spain, as well as Finland-Sweden.

Group four contains emigration flows with moderate migration intensity. Emigration is high during childhood and peaks at working ages. It slowly declines after that and remains considerably high throughout the older ages. The main country of origin is Spain, which is also the second prominent country of destination after Greece. This group contains many flows from Croatia to Germany as well as from Cyprus to Greece, as well as Estonia-Finland, Lithuania-Norway, and Cyprus-Great Britain.

Group five is the largest group and consists of flows with sharp peaks in the early working ages, peaks in the childhood ages, as well as sizable peaks in the late working ages but are flattening out in the old ages. The predominant country of origin is Portugal and the predominant country of destination is Great Britain. The largest flows include again Romania-Italy, Estonia-Finland, Croatia-Germany, and Ireland-Great Britain.

Within all groups, the years in which the flows took place are quite evenly distributed.

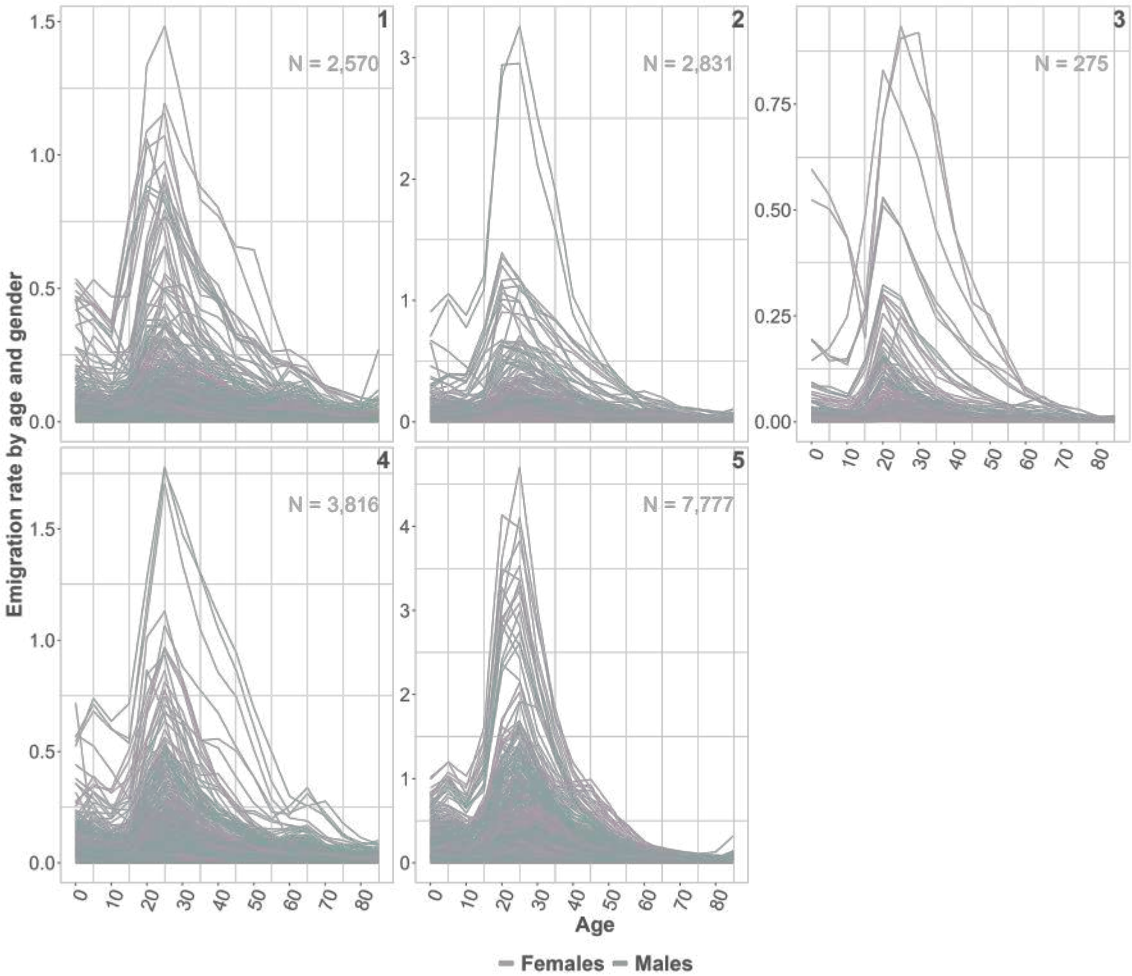

Figure 4: Group membership of emigration flows based on the K-means clustering results. Each panel displays a group of age-specific emigration rates of each origin-destination-year-gender combination in our data set. The range of the y-axis differs across panels due to the varying intensities of emigration across groups.

Figure 5 compares the number of peaks and plateaus across the different groups. Group one clearly includes flows which exhibit predominantly three to six peaks which can be classified as high multi-modal age profiles. The second group is dominated by age profiles with three to four peaks but includes as well a smaller share of profiles with two, five, or six peaks and can therefore be classified as moderately multi-modal. Group three contains the broadest variety of age profiles. With numbers of peaks ranging from one to six. However, age profiles with one or two peaks clearly dominate the group and hence it can be classified as predominantly low multi-modal. In group four, the number of peaks in the age profiles ranges from two to six but with two to four being the most frequent cases. Therefore, the profiles in group four can be classified as moderately multi-modal. And finally, group five exhibits more coherent patterns in age-profiles of one to three peaks and can be classified as lower multi-modal.

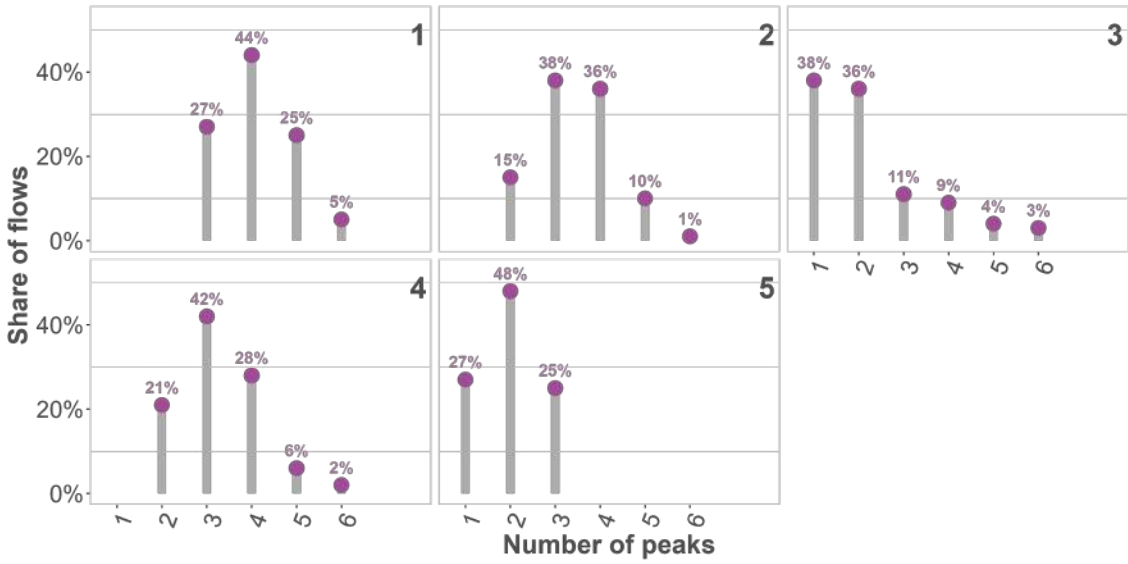

Figure 5: Share of flows within each group that display the respective number of peaks in the emigration age distribution.

As the number of peaks alone does not make the age-profile comparable, we structure the migration components according to life stages and compare them within each stage with each other. In Figure 6, the Rogers-Castro parameter estimates for α_p, λ_p, c, and a_p are plotted across groups. The points indicate the median value of the estimate within the respective group and the violins represent the distribution where it exists. We can see similar patterns for λ and α parameter estimates usually increasing with the number of peaks across the different groups.

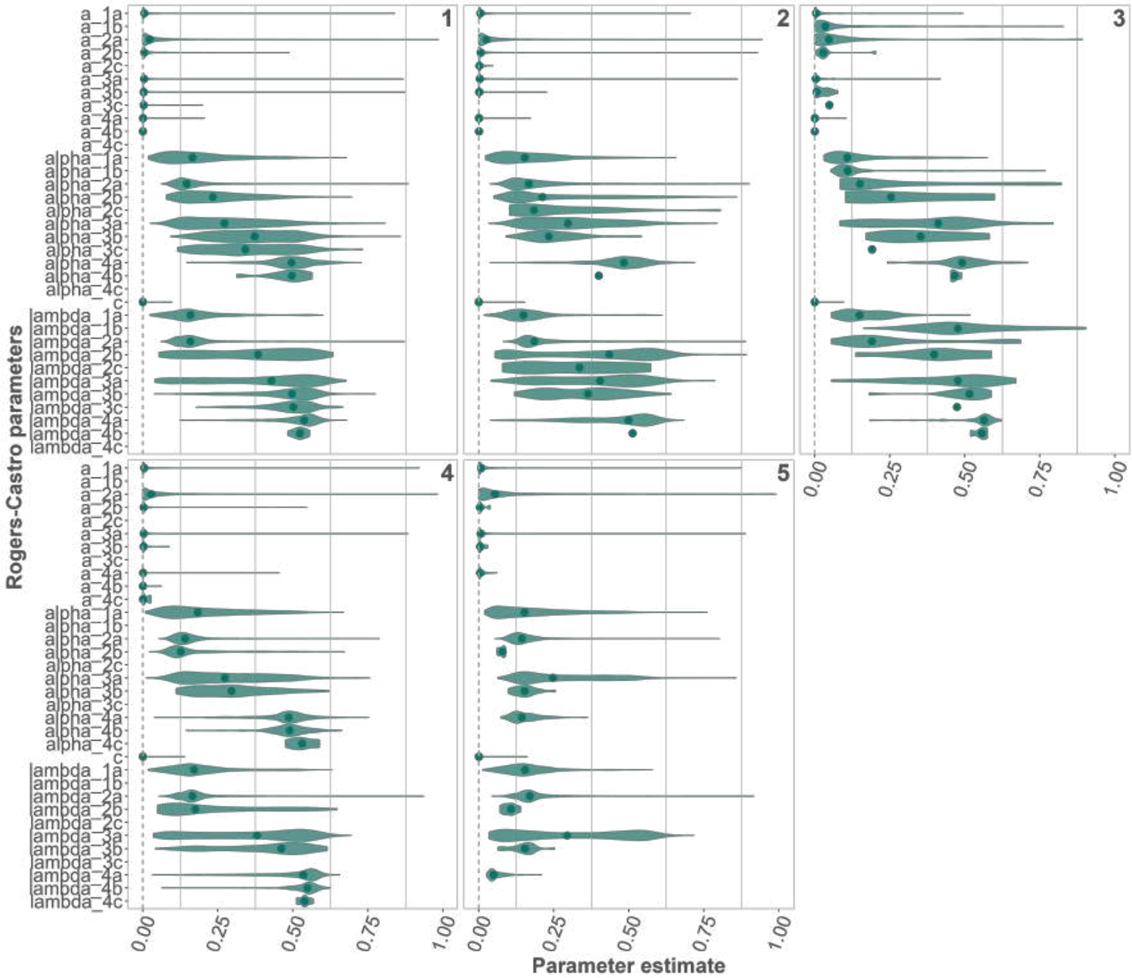

Figure 6: Rogers-Castro parameter estimates for emigration flows. Dots indicate the median values of the respective parameter estimates within each group. All estimates have been constrained to be positive due to the nature of emigration rates. The y-axis indicates the names of the estimated parameters. Note that the x-axis scale is fixed for all panels.

The parameter estimates for the location of the modes on the x-axis are illustrated in Figure 7. Peaks during the childhood ages are common in every group with group three being the only one containing profiles with two peaks in the childhood ages. The medians of the first peaks in the working ages occur at the ages $25 - 29$ for all groups. Only group two contains age profiles with a third peak in the early working ages at the ages $35 - 39$. Groups one and three exhibit profiles with a third peak in the late working ages around ages $60 - 64$. Only group four contains profiles with a third peak in the old ages. However, points surrounded by very narrow violins or no violins at all indicate isolated individual cases.

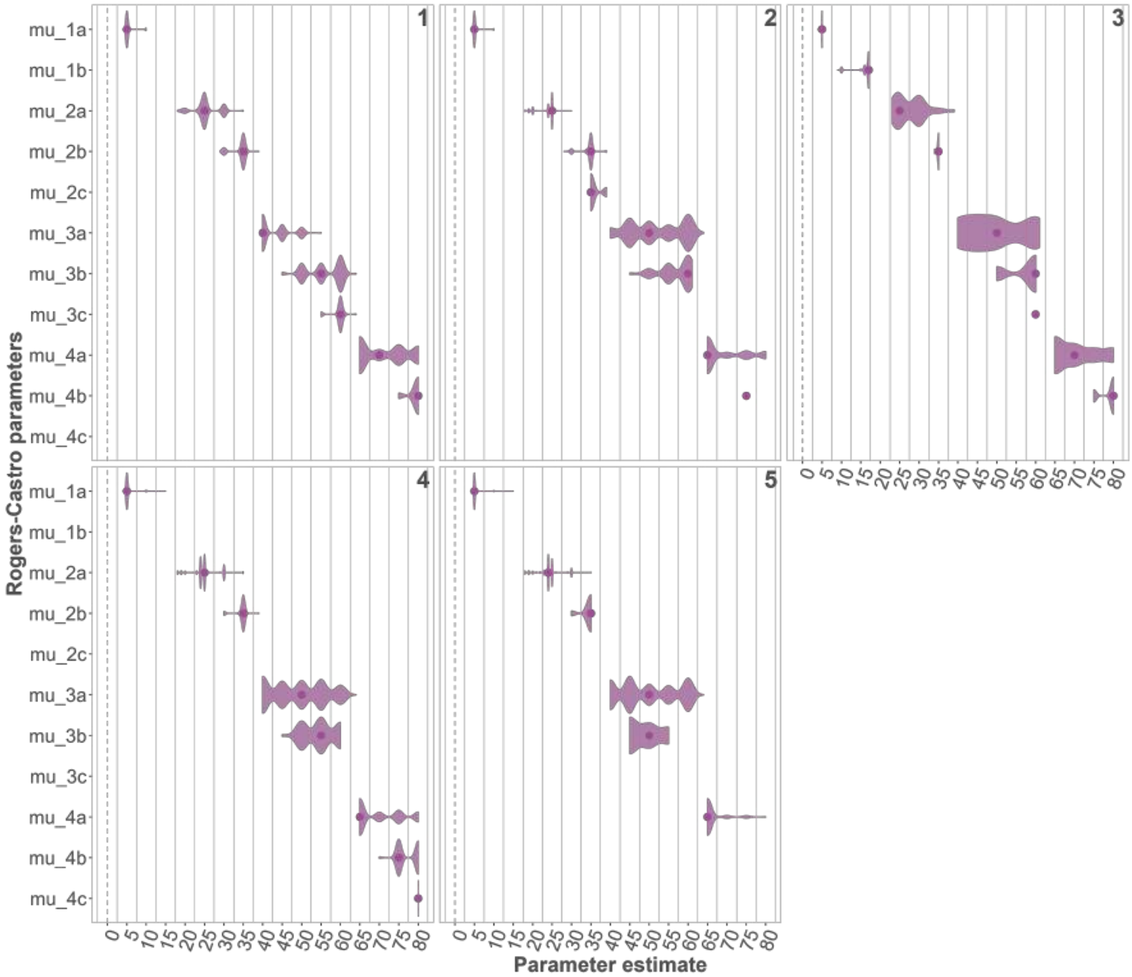

Figure 7: Rogers-Castro parameter estimates for the modes of emigration flows. Dots indicate the median values of the respective parameter estimates within each group. The y-axis indicates the names of the estimated parameters. Note that the x-axis scale is fixed for all panels.

In summary, the groups can be names and characterized as follows.

Group 1 - Across life stages migration: The flows in this group have moderate, broad working-age peaks combined with notable childhood and retirement peaks, indicating migration spread across all life stages.

Group 2 - Early work life and family migration: This group is characterized by moderately high emigration rates with peaks in early working ages, notable childhood migration, and retirement-age peaks.

Group 3 - Across life stages and low-intensity migration: In this group, emigration flows exhibit the second lowest overall rates with some outliers featuring early childhood peaks, strong working age migration, and persistently high emigration across all ages including the older ages.

Group 4 - Late work life and retirement migration: This group is characterized by wide peaks at the working ages, steeper childhood down-slopes, and elevated migration levels at retirement and older ages.

Group 5 - Early work life and childhood migration: This is the largest group with the highest emigration rates, multiple peaks in childhood, working ages, and late working ages, followed by steep declines in older ages.

5.2 Differences in emigration group membership by gender

The number of flows in each group varies slightly by gender. In group one 42.8% of flows are female flows, in group two 49.1%, in group three 63.6%, in group four 46.5%, and in group five 53.6%.

Figure 8 illustrates the overlap in group membership between female and male emigration flows within one corridor over time. The shares of overlapping years vary greatly across origin-destination combinations. The median overlap across all corridors is around 50% of overlap. 43 corridors achieve an overlap of 100%, among them are Lithuania-Finland and Cyprus-Belgium. Another 20 corridors have no overlap at all between the two genders including Denmark-Austria or Italy-Germany. It is important to note that models which did not converge in the analysis were removed from the clustering approach. Hence, some origin-destination-year-gender combinations lack the corresponding flow for the other gender. Therefore, no overlap could be calculated for them and they were excluded from the below figure.

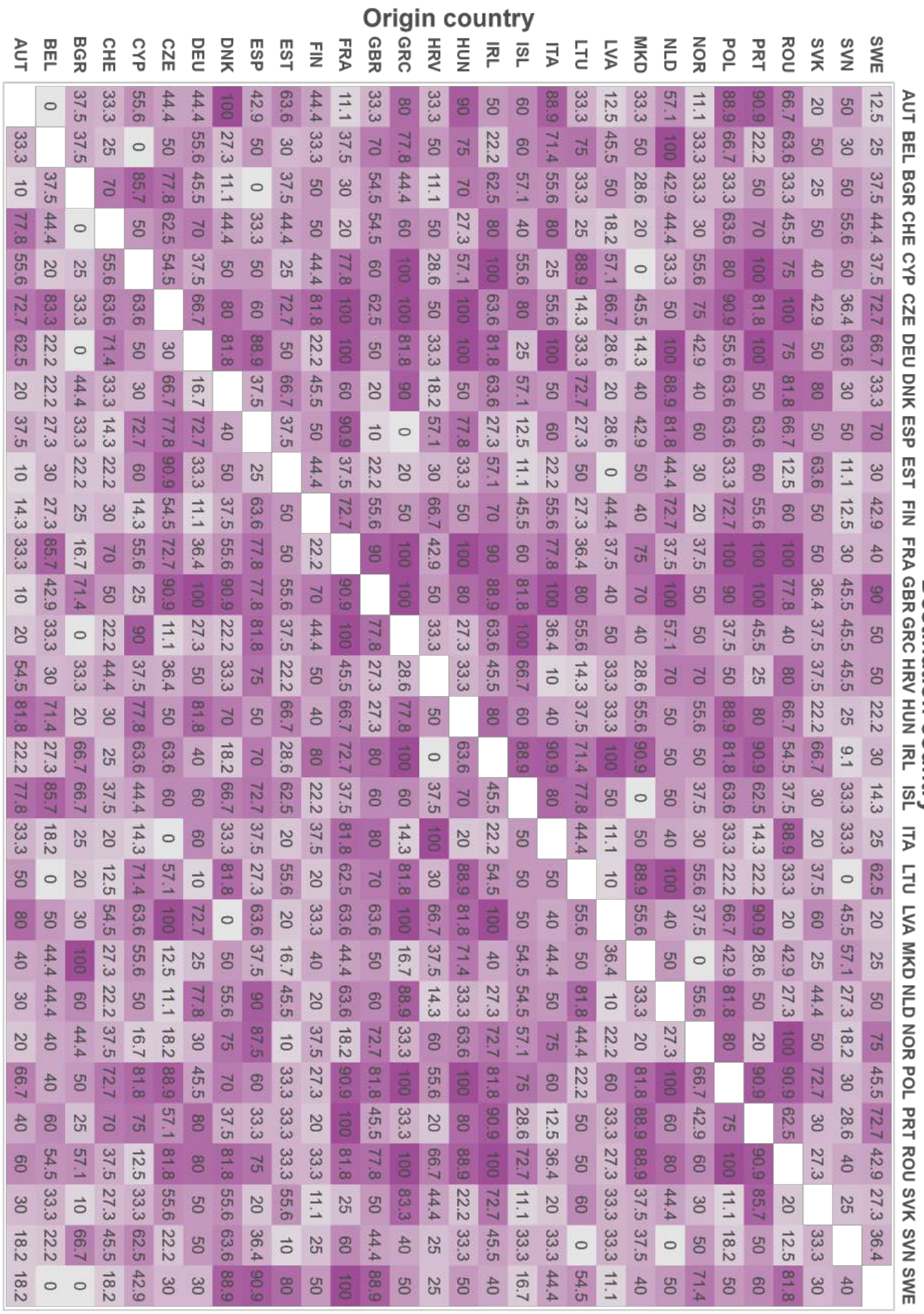

Figure 8: Overlap of emigration group memberships of females and males within the same migration corridor. Each tile represents a unilateral migration corridor. The number inside the cell indicates for how many years, out of the overall study period, female and male flows are classified in the same group.

It becomes clear that age profiles and hence group membership vary strongly by gender for most migration corridors. This finding challenges the usual assumption that female migration age profiles within the same flow mirror those of their male counterparts. Only 16.3% of all 870 corridors have an overlap that is higher than 80%.

5.3 Patterns of emigration over time

The evolution of group membership over time for each origin-destination-year-gender combination is illustrated in Figure 9. Only a small number of flows remains in the same group over the entire time period. The remaining flows fluctuate quite frequently in group membership. Even though flows seem to switch very often from one group to another over time, the group sizes remain relatively stable (see Figure S1 in the Supplementary Materials).

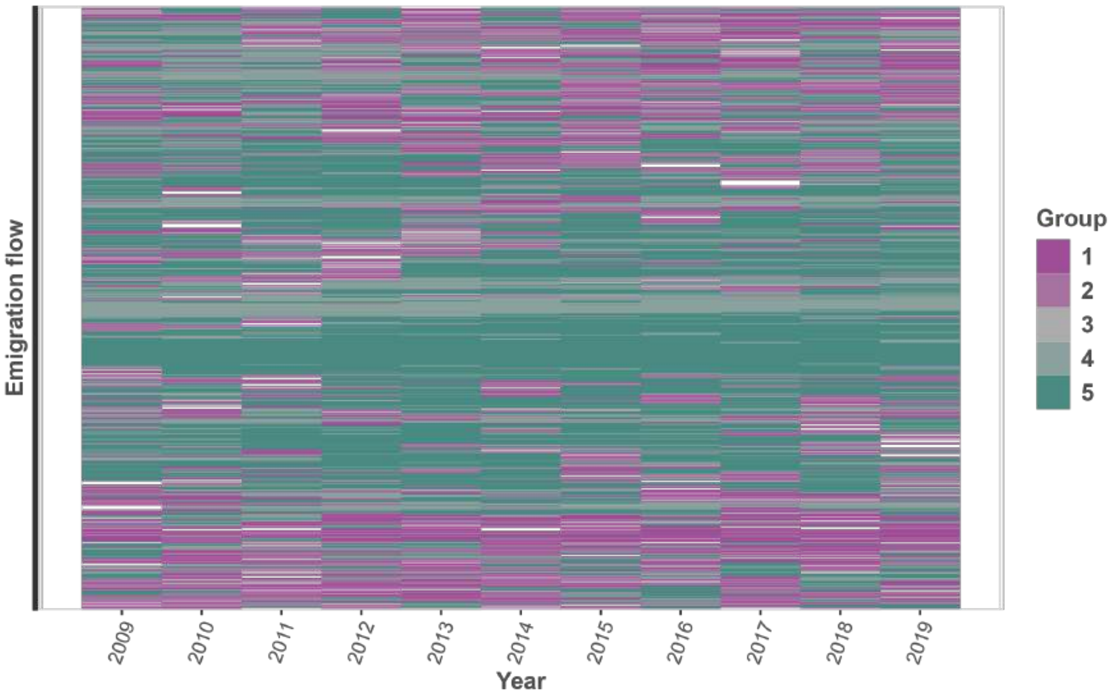

Figure 9: Group membership of each origin-country-year-gender combination over time. The rows are arranged based on similarity of sequences.

More detailed sequences for emigration flows by origin country can be found in the Supplementary Materials (see 7). These figures show that group membership varies strongly by origin country.

6 Discussion and Conclusions

We applied a novel approach to extend the typical Rogers-Castro model migration schedule. We developed a flexible extension and Bayesian implementation based on the shape of each individual migration flow. Therefore, we made use of bilateral emigration flow estimates disaggregated by age and gender and provided by QuantMig. We performed K-means clustering on the resulting parameter estimates and identified five distinct emigration age-profiles for international migration in Europe which we further analyzed by gender, time, and space.

Unlike previous research, we identify five instead of four distinct groups of age profiles for emigration. Moreover, we quantify the extent to which age-distributions differ by gender. This topic has been mentioned in previous research but has not been fully explored yet (to the best of our knowledge). Moreover, our findings indicate that female flows do not simply mirror male flows of the same corridor but differ very often in terms of their age distribution. Overall, female and male age distributions within one corridor overlap to 50% for emigration flows. Finally, our methodological approach and our findings highlight an urgent need to reevaluate the Rogers-Castro schedule and to quantify and incorporate gender differences in migration modeling. As identified by our shape detection function, age patterns oftentimes differ from those assumed and modeled by the typical Rogers-Castro model migration schedules.

Nevertheless, our approach faces several limitations. First, the data on which the study is based has to be treated with caution as it is estimated and harmonized itself. We know that the confidence in the estimates differs by reporting country. However, it is the only comprehensive data set on international migration disaggregated by age and gender we are aware of. Second, not all of our models converged when fitted to the data. About 10% of models for emigration flows did not converge even after the second round of fitting. This forced us to exclude a subset of the available migration flows. Third, we have not explored group membership in a statistical sense, meaning we do not have explored which observable characteristics might correlate with group membership of a specific migration flow. This is owed to the nature of the underlying data which was estimated with the help of a migration model that includes many country variables already. This would threaten the statistical validity of our estimation and interpretation of the results.

We believe that our work opens up many future avenues of research. It would be insightful to analyze pattern at each individual country level and to understand age and gender differences by origin and destination. Even though we do have results for each origin and destination country in our sample, analyzing them exceeds the scope of this article. When data sets on migration flows disaggregated by age and gender become available for other world regions, it would be of major interest to apply our approach to these regions and compare the results across the globe. We also hope to foster more critical reflections on our current tools in migration statistics and their decades-old assumptions in order to update them to fit the dynamic and changing nature of migration and its demographic patterns. Through this work, we aim to contribute to a better understanding of age- and gender differences in migration patterns and to lay the groundwork for theorizing about migration through the lens of socio-demographic inequalities.

7 Supplementary materials

7.1 Emigration

Figure S1: Group membership over time seems rather stable.

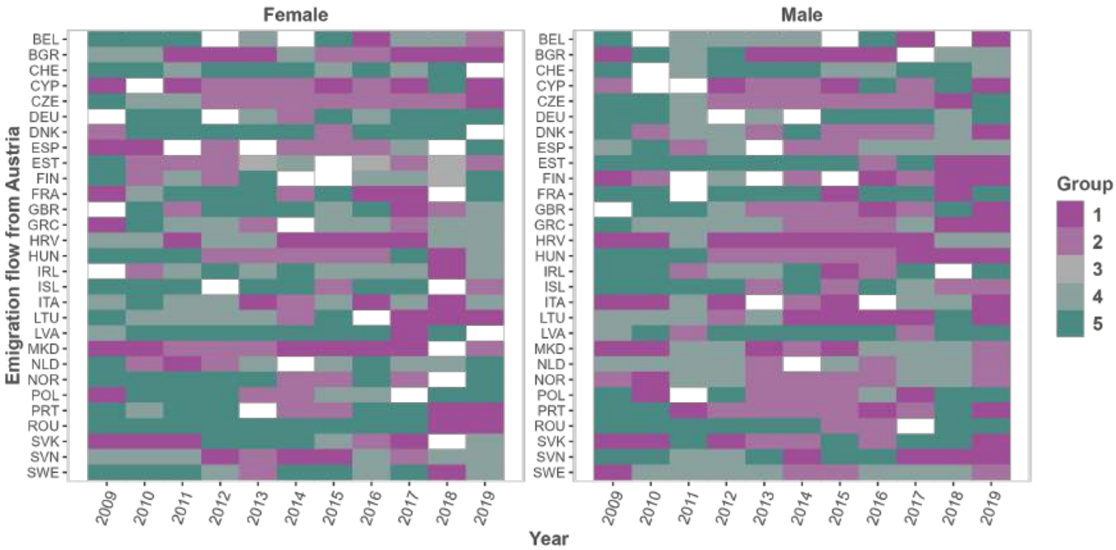

Figure S2: Emigration from Austria.

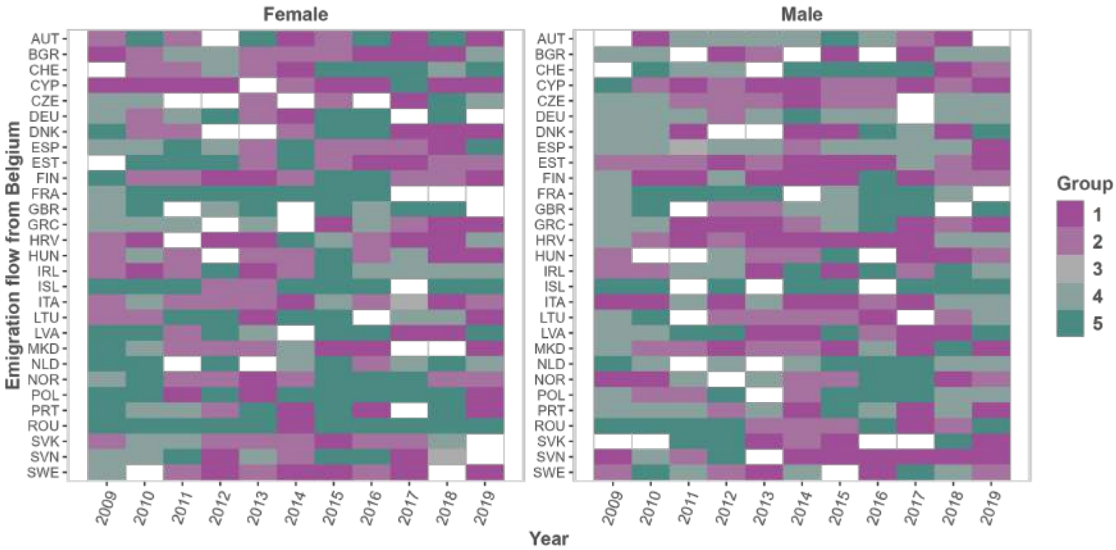

Figure S3: Emigration from Belgium

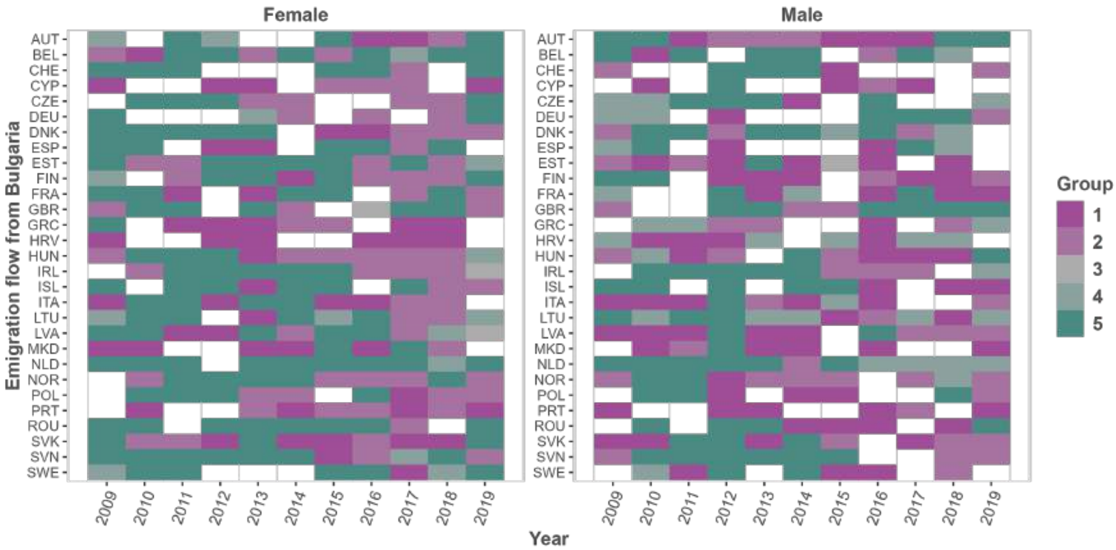

Figure S4: Emigration from Bulgaria.

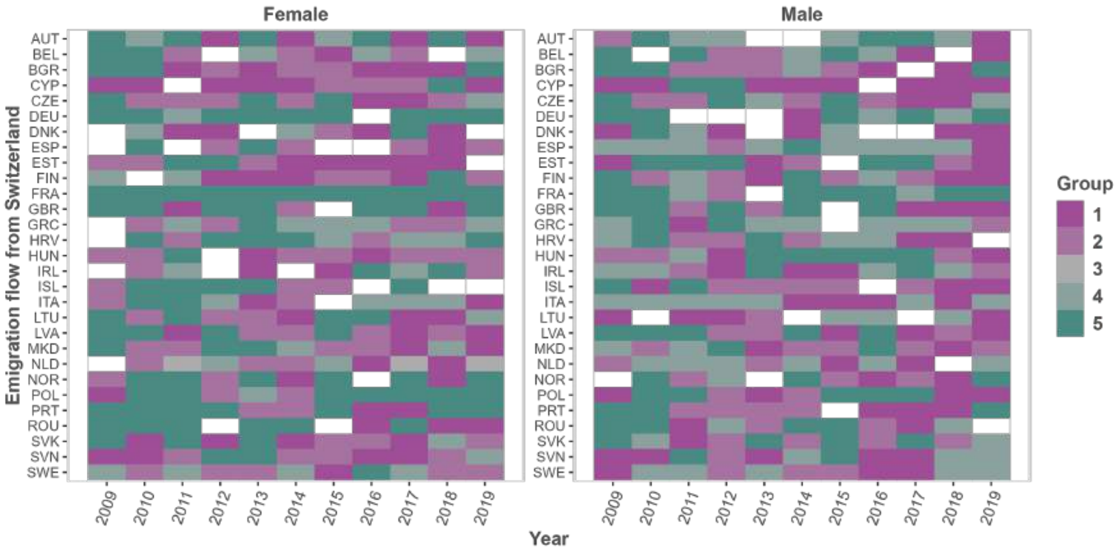

Figure S5: Emigration from Switzerland.

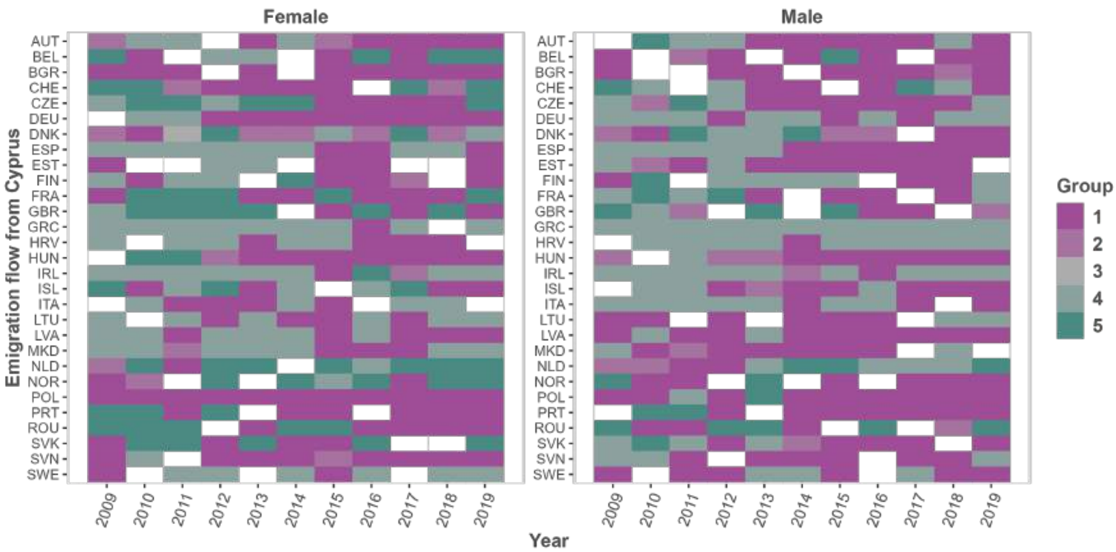

Figure S6: Emigration from Cyprus.

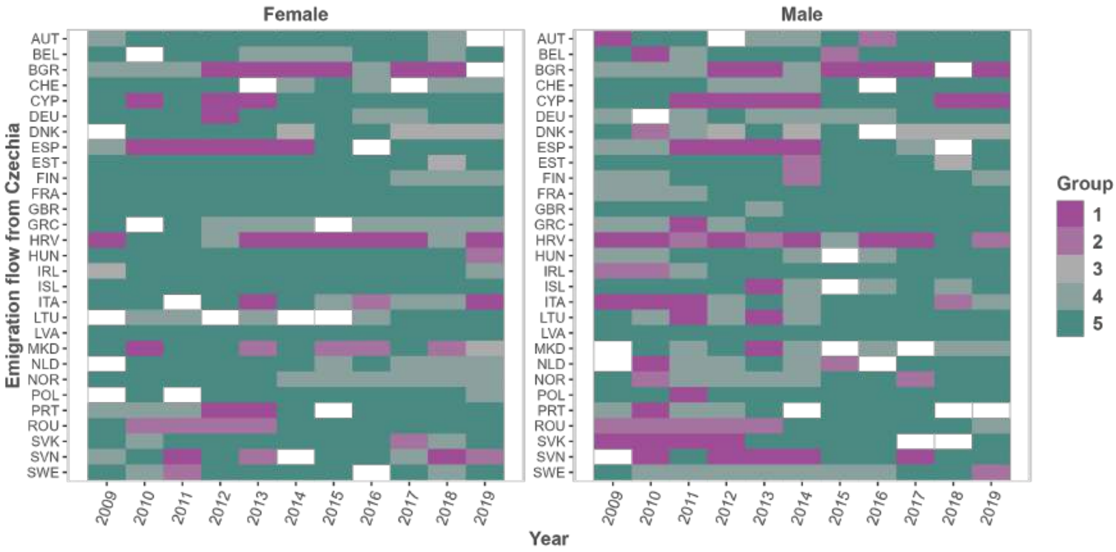

Figure S7: Emigration from Czech Republic.

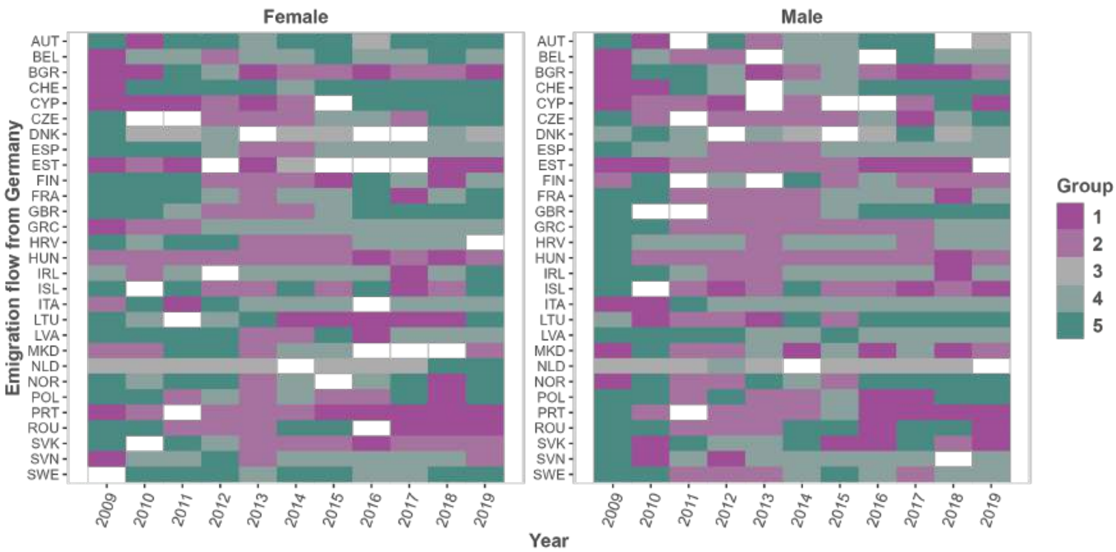

Figure S8: Emigration from Germany.

Figure S9: Emigration from Denmark.

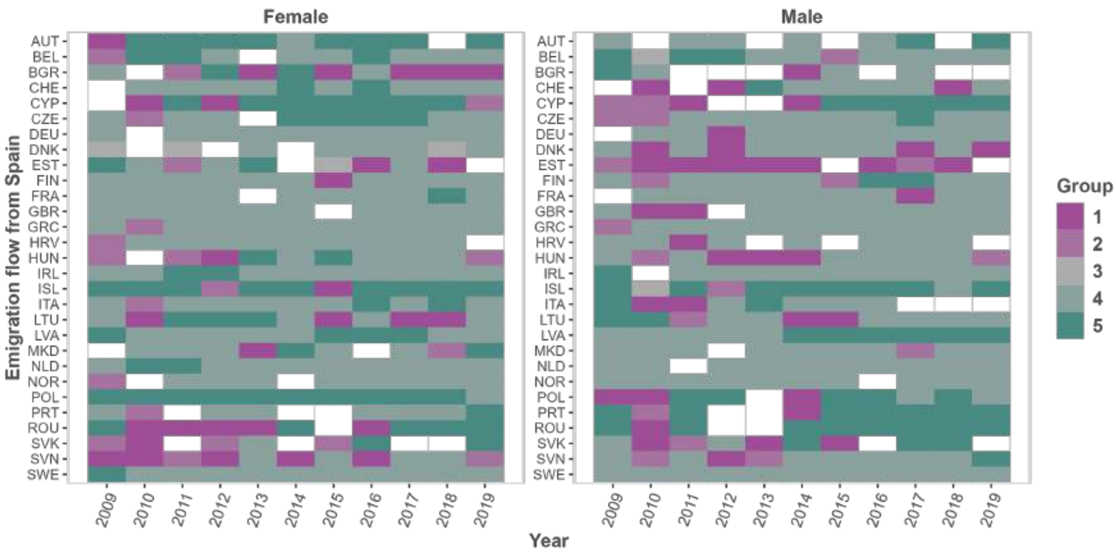

Figure S10: Emigration from Spain.

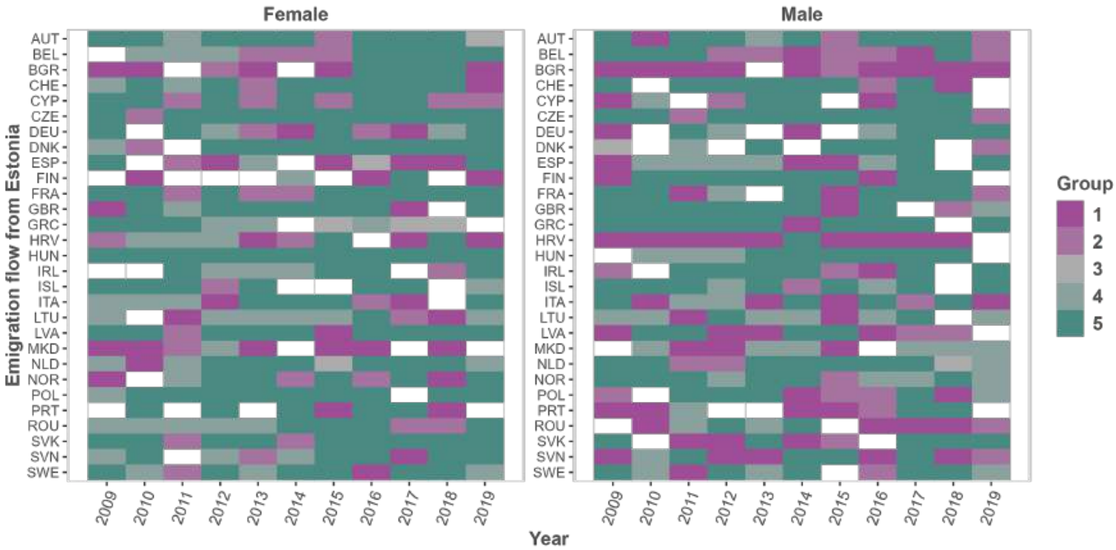

Figure S11: Emigration from Estonia.

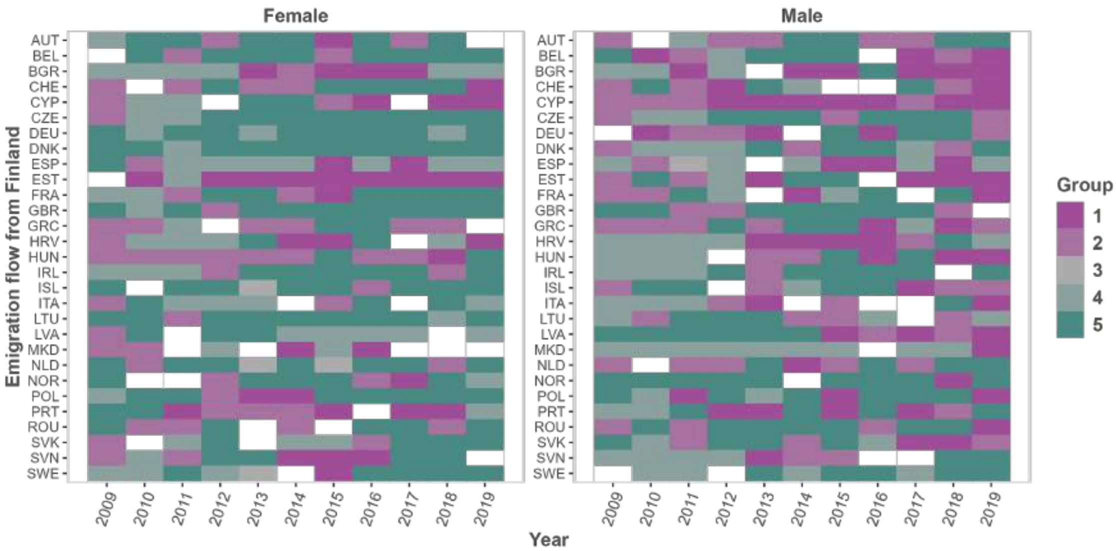

Figure S12: Emigration from Finland.

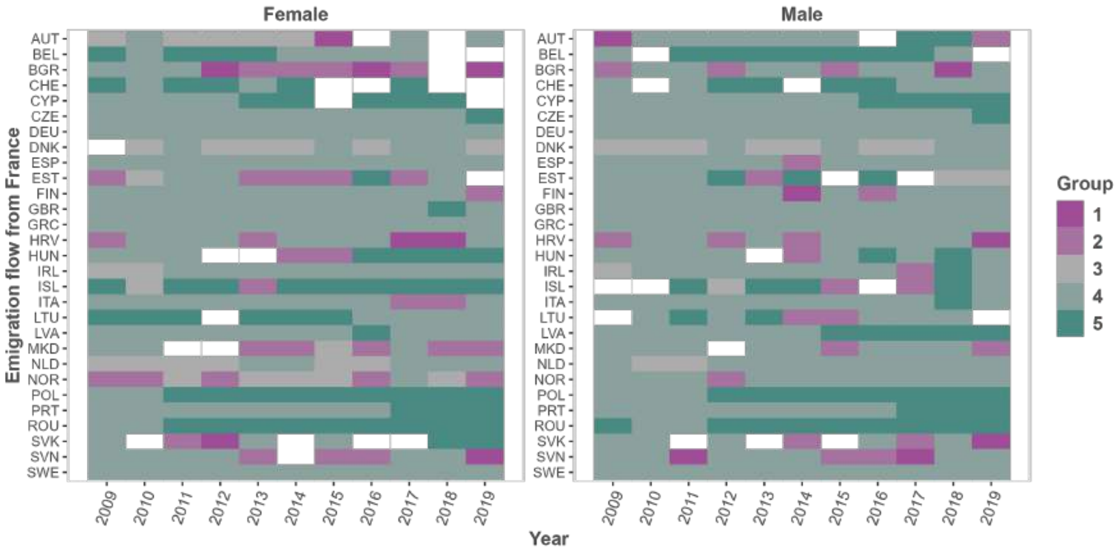

Figure S13: Emigration from France.

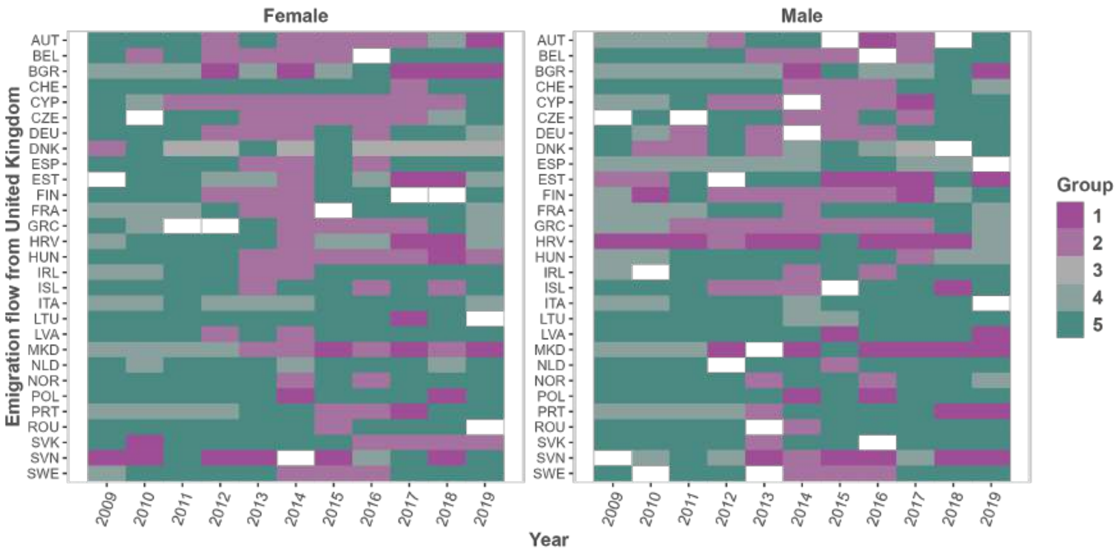

Figure S14: Emigration from Great Britain.

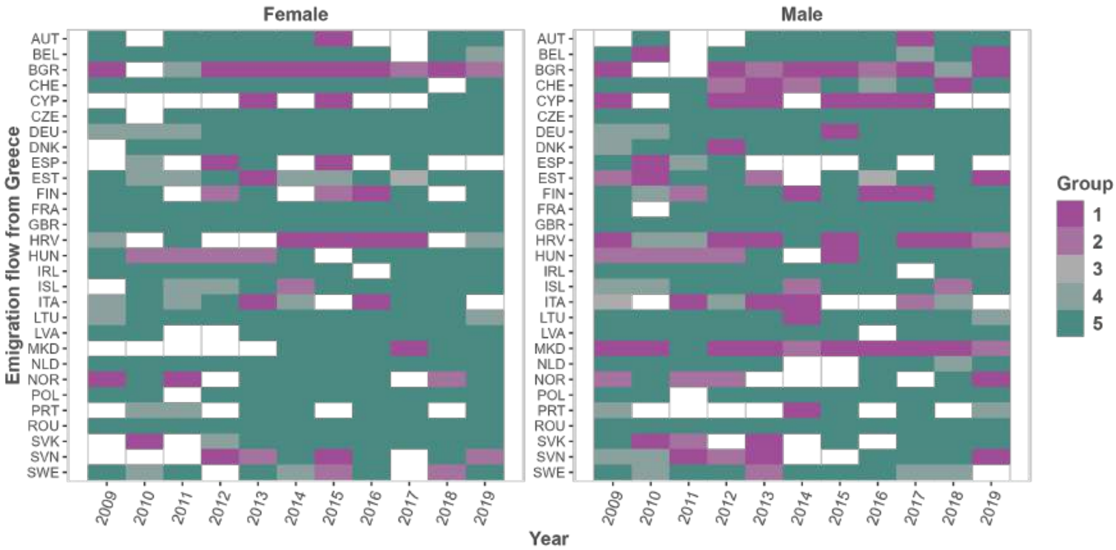

Figure S15: Emigration from Greece.

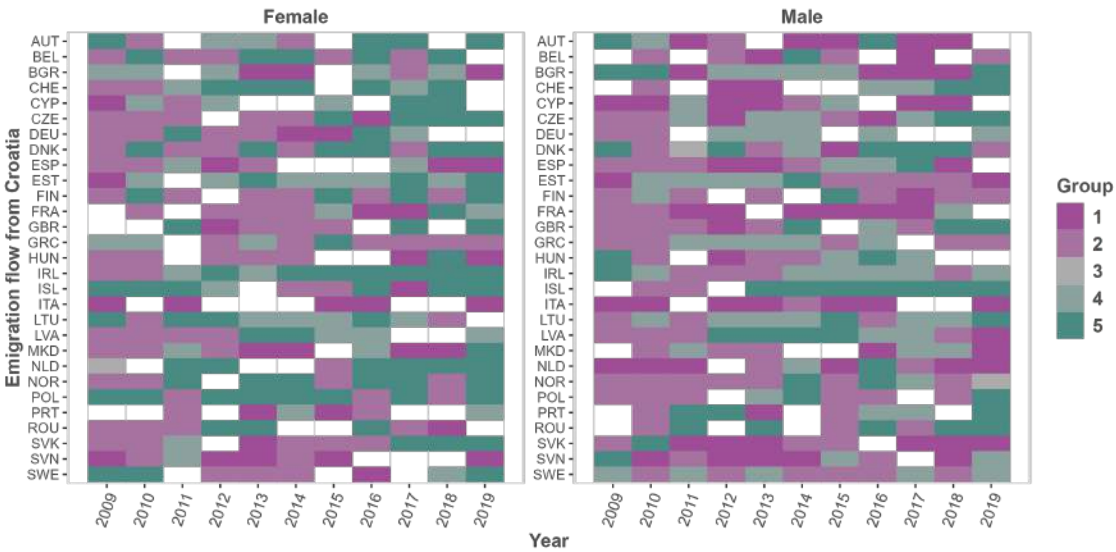

Figure S16: Emigration from Croatia.

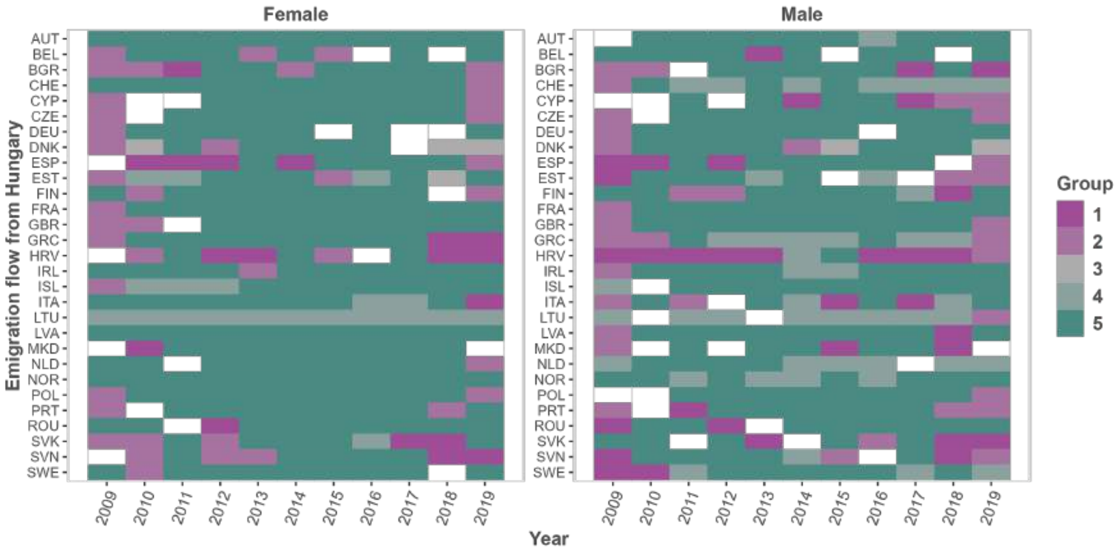

Figure S17: Emigration from Hungary.

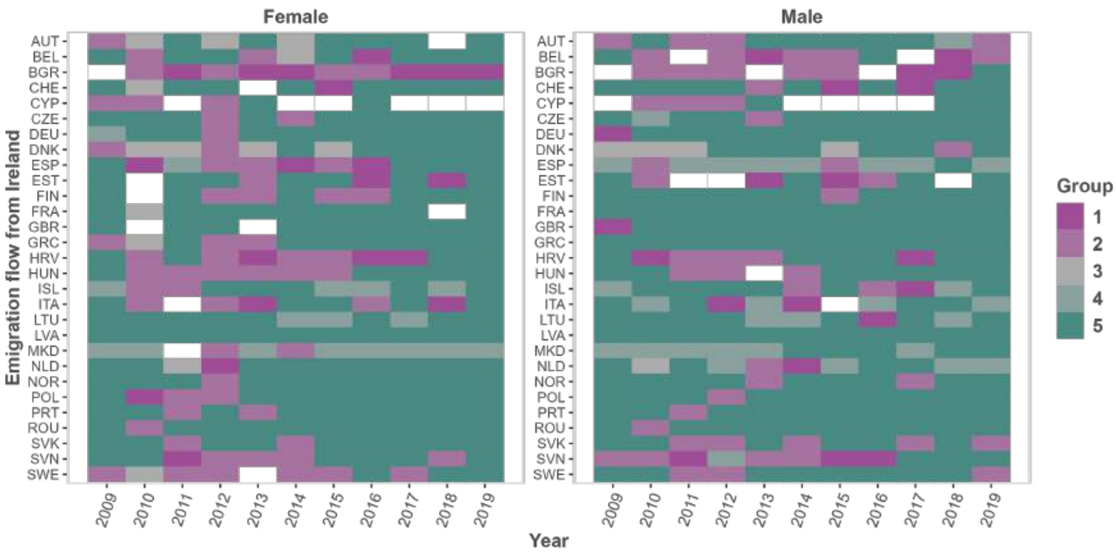

Figure S18: Emigration from Ireland.

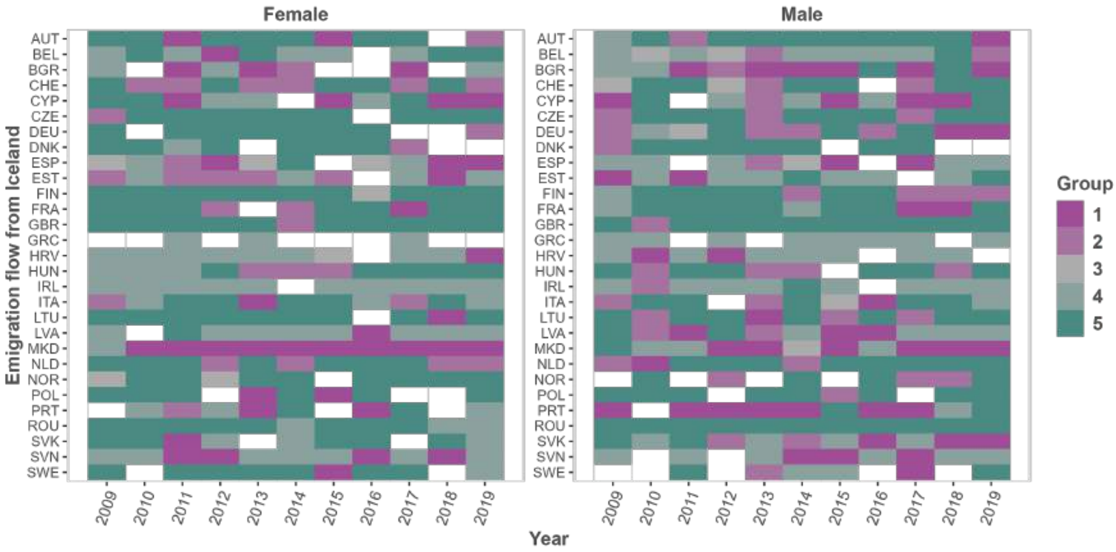

Figure S19: Emigration from Iceland.

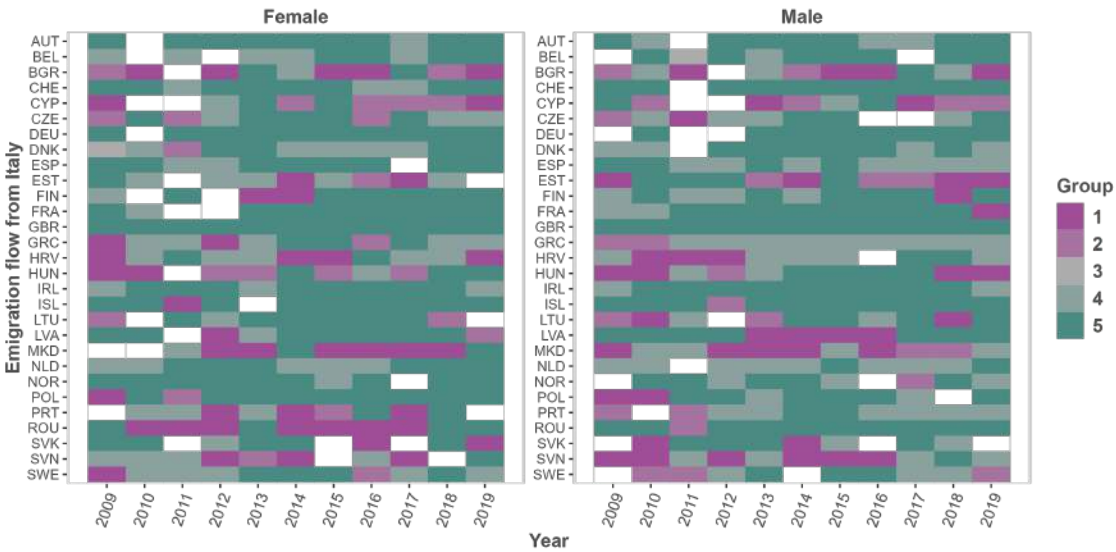

Figure S20: Emigration from Italy.

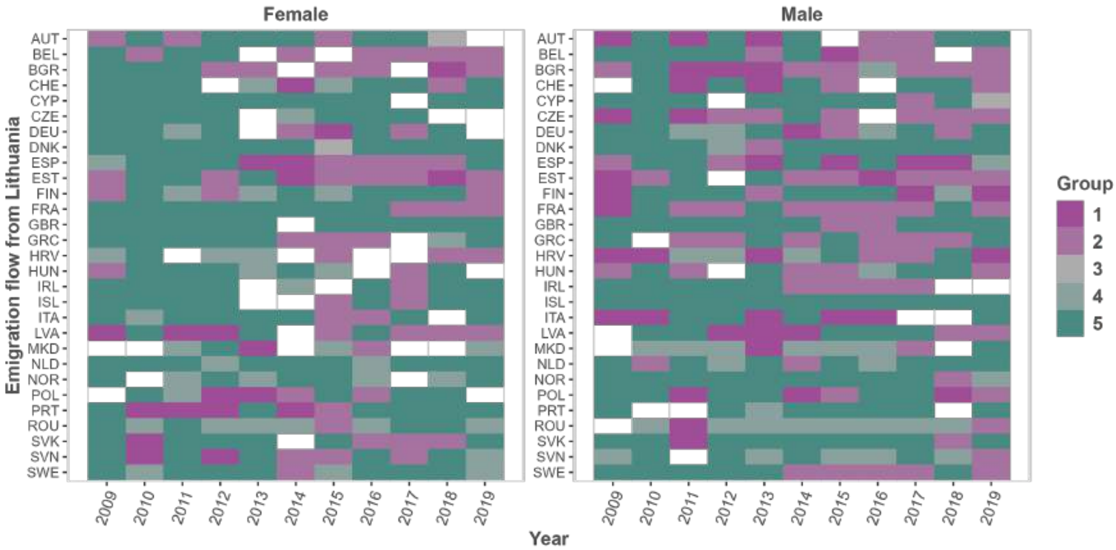

Figure S21: Emigration from Lithuania.

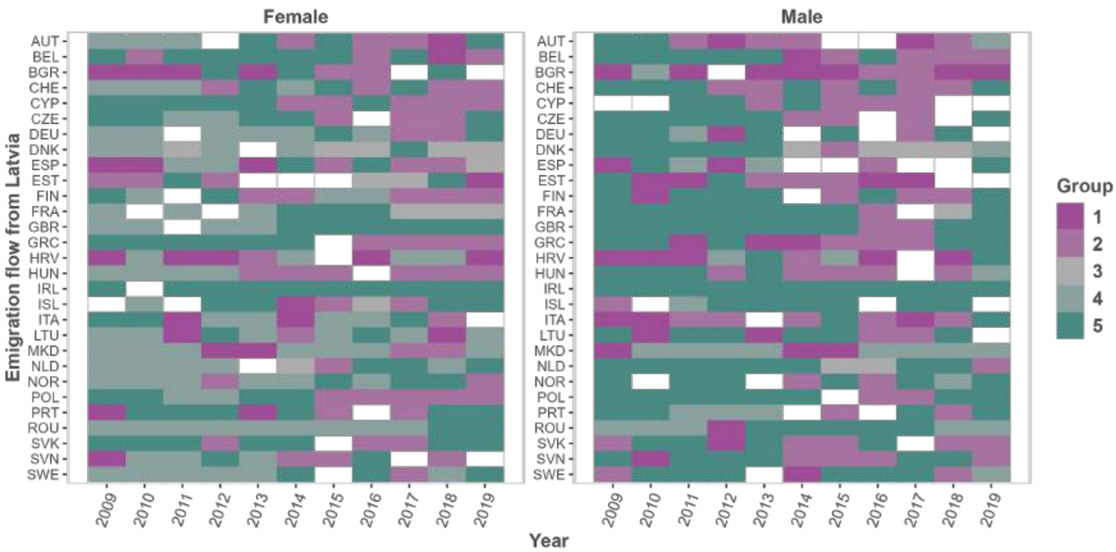

Figure S22: Emigration from Latvia.

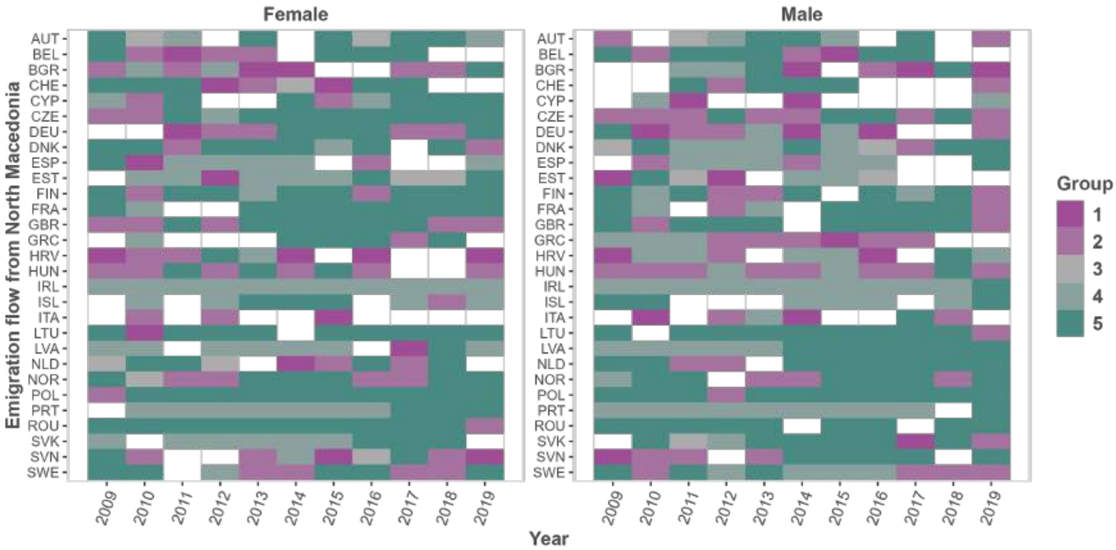

Figure S23: Emigration from North Macedonia.

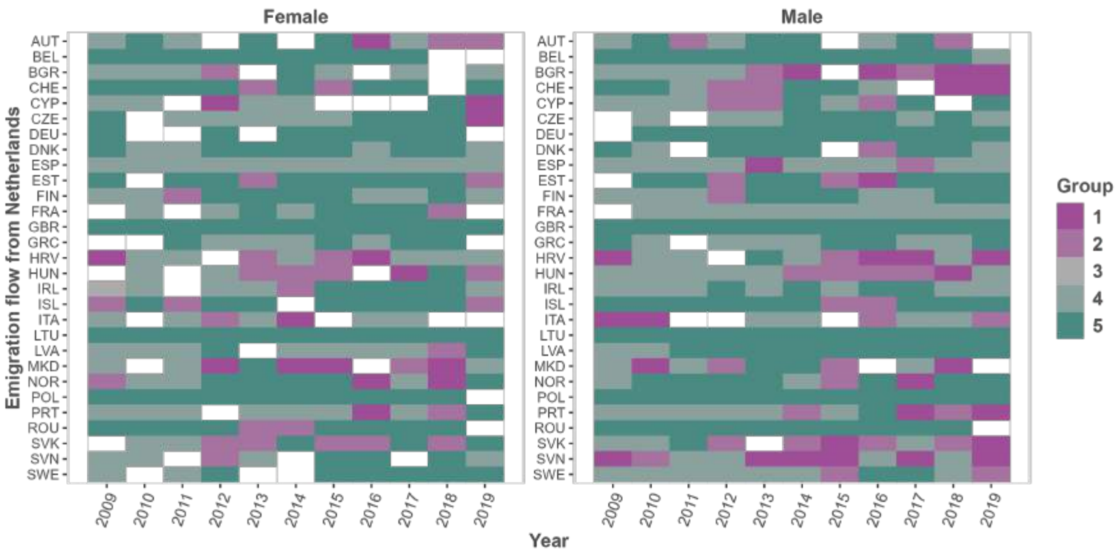

Figure S24: Emigration from the Netherlands.

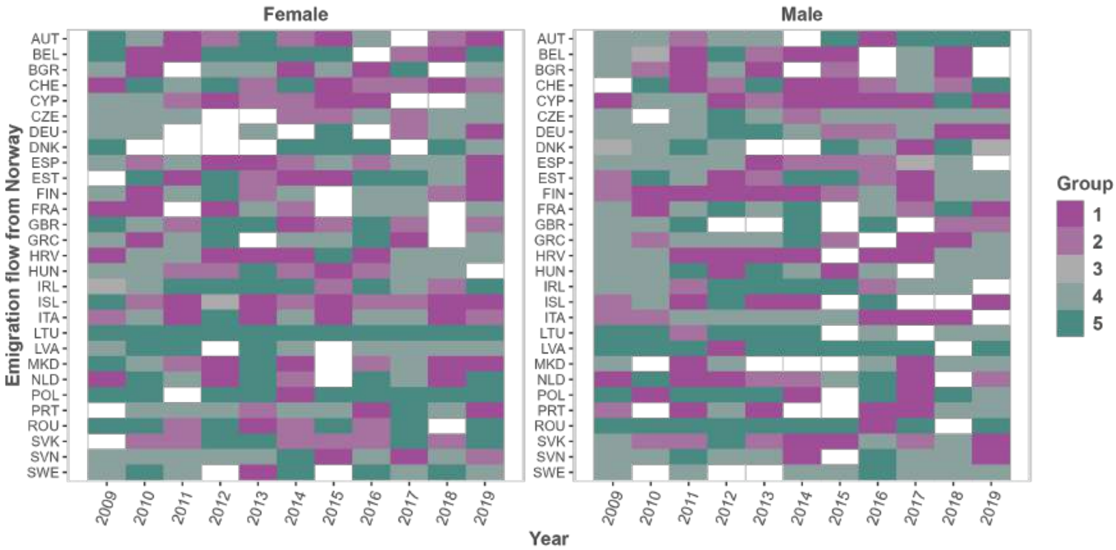

Figure S25: Emigration from Norway.

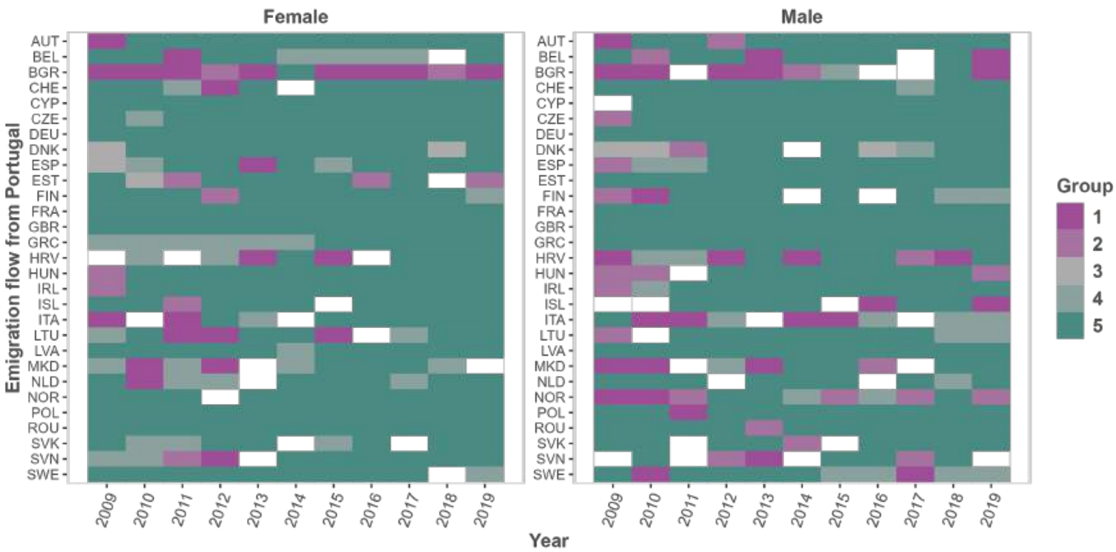

Figure S26: Emigration from Portugal.

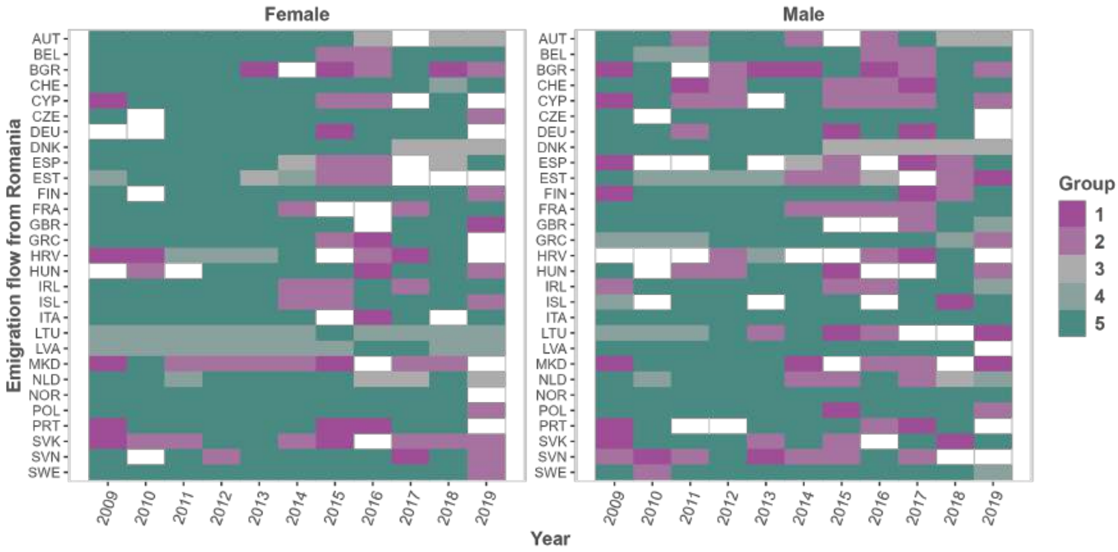

Figure S27: Emigration from Romania.

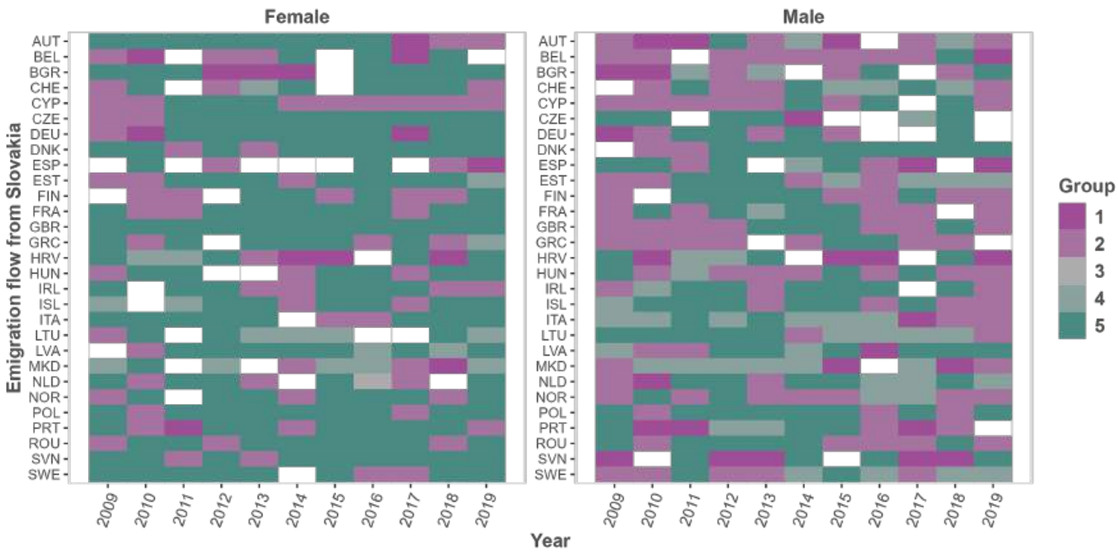

Figure S28: Emigration from Slovakia.

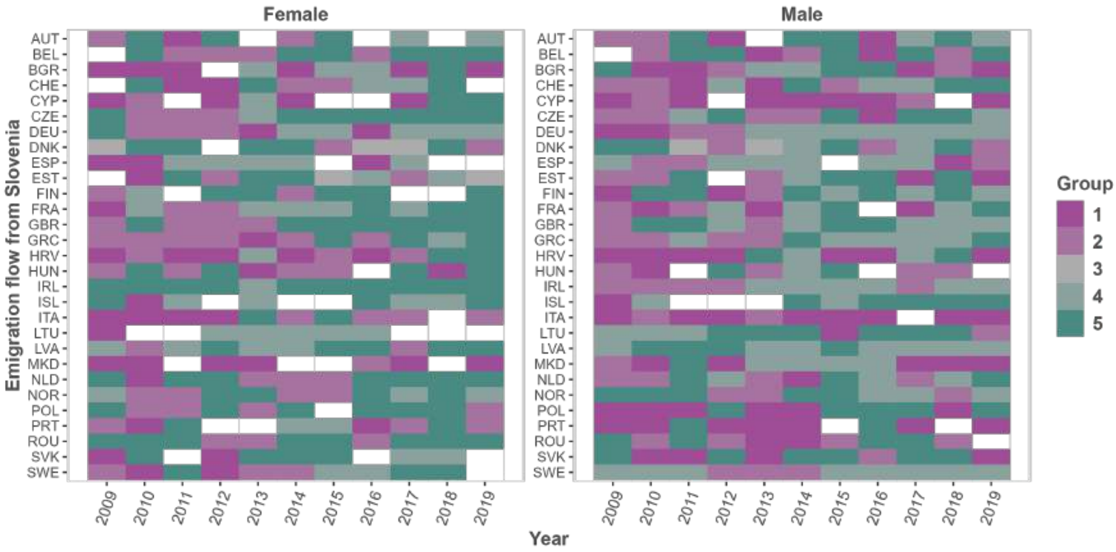

Figure S29: Emigration from Slovenia.

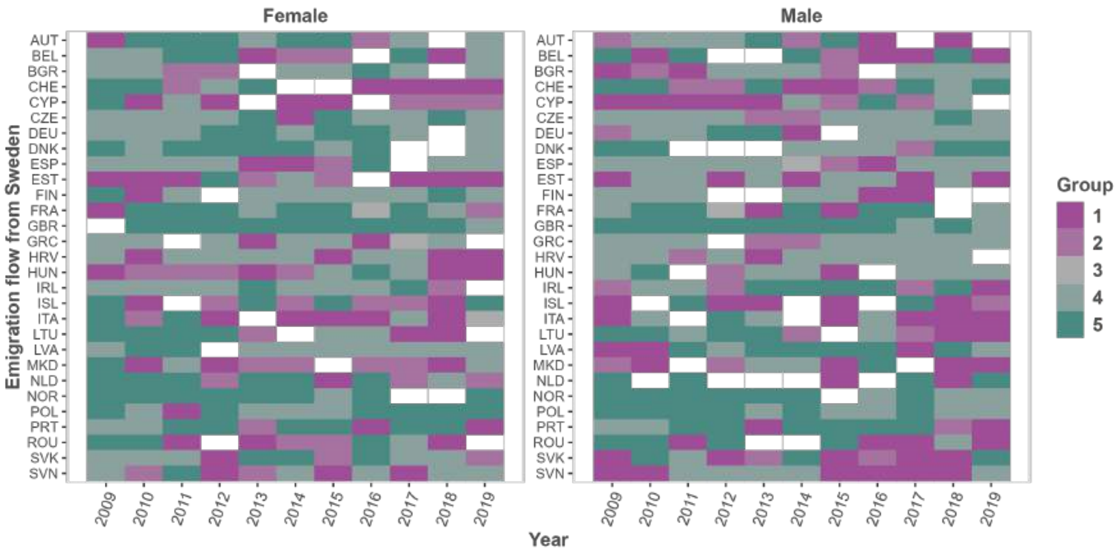

Figure S30: Emigration from Sweden.

References

Abel, G. J. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research*, *28*, 505–546.

Abel, G. J., & Cohen, J. E. (2019, June). Bilateral international migration flow estimates for 200 countries. *Scientific Data*, *6*(1), 82. Retrieved 2021-12-21, from `https://www.nature.com/articles/s41597-019-0089-3` (Bandiera_abtest: a Cc_license_type: cc_publicdomain Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Geography;Social sciences Subject_term_id: geography;social-sciences) doi: 10.1038/s41597-019-0089-3

Abel, G. J., & Sander, N. (2014). Quantifying global international migration flows. *Science*, *343*(6178), 1520–1522.

Anastasiadou, A., Kim, J., Sanlitürk, E., de Valk, H. A., & Zagheni, E. (2024). Gender differences in the migration process: A narrative literature review. *Population and Development Review*, *50*(4), 961–996.

Aristotelous, G., Smith, P. W., & Bijak, J. (2022). Technical report: Estimation methodology. *Changes*, *1*, 13.

Castro, L. J., & Rogers, A. (1983). Patterns of family migration: two methodological approaches. *Environment and Planning A*, *15*(2), 237–254.

Collinson, M. A. (2017). Age—sex profiles of migration: Who is a migrant? In *The dynamics of migration, health and livelihoods* (pp. 49–62). Routledge.

Donato, K. M., & Gabaccia, D. (2015). *Gender and International Migration*. Russell Sage Foundation. (Google-Books-ID: D8iZBgAAQBAJ)

Guttentag, M., & Secord, P. F. (1983). *Too many women?: The sex ratio question*. Sage Publications.

Little, J. S., & Rogers, A. (2007). What can the age composition of a population tell us about the age composition of its out-migrants? *Population, space and place*, *13*(1), 23–39.

Pittenger, D. (1978). On making flexible projections of age-specific net migration. *Environment and Planning A*, *10*(11), 1253–1272.

Pittenger, D. B. (1974). A typology of age-specific net migration rate distributions. *Journal of the American institute of planners*, *40*(4), 278–283.

Plane, D. A., & Heins, F. (2003). Age articulation of us inter-metropolitan migration flows. *The annals of regional science*, *37*, 107–130.

Raymer, J., Guan, Q., Shen, T., Hertog, S., & Gerland, P. (2023). Modelling the age and sex profiles of net international migration. (7).

Raymer, J., Rogers, A., et al. (2008). Applying model migration schedules to represent age-specific migration flows. In *International migration in europe: Data, models and estimates* (pp. 175–192). Wiley Chichester.

Rogers, A., & Castro, L. J. (1981). *Model migration schedules* (Tech. Rep. No. RR-81-030). International Institute for Applied Systems Analysis.

Rogers, A., Castro, L. J., & Lea, M. (2005). Model migration schedules: Three alternative linear parameter estimation methods. *Mathematical Population Studies*.

Rogers, A., Little, J., & Raymer, J. (2010). *The indirect estimation of migration: Methods for dealing with irregular, inadequate, and missing data* (Vol. 26). Springer Science & Business Media.

Shen, T., Raymer, J., Guan, Q., & Wiśniowski, A. (2024). The estimation of age and sex profiles for international migration amongst countries in the asia-pacific region. *Population, Space and Place*, *30*(1), e2716.

United Nations, Department of Economic and Social Affairs, Population Division. (2022). *World population prospects 2022: Methodology of the united nations population estimates and projections* (Tech. Rep. No. (UN DESA/POP/2022/TR/NO.4)).

Wilson, T. (2010). Model migration schedules incorporating student migration peaks. *Demographic research*, *23*, 191–222.

Yeung, J., Alexander, M., & Riffe, T. (2023). Bayesian implementation of rogers–castro model migration schedules. *Demographic Research*, *49*, 1201–1228.