



# Using bibliometric data for demographic research

INED, June 2024

## Speaker:

Tom Theile, Max Planck Institute for  
Demographic Research (MPIDR)

## Contributors

Emilio Zagheni, MPIDR  
Aliakbar Akbaritabar, MPIDR  
Andrea Miranda Gonzalez, UC-Berkeley  
Asli Ebru Sanlitürk, MPIDR  
Xinyi Zhao, MPIDR



## AGENDA

- Bibliometric data for migration research
- The Scholarly Migration Database
- Trends of return international migration worldwide
- Economic development and international scholarly migration
- Gender perspective in international scholarly migration
- Internal migration in Mexico
- Brexit's effect on scholarly migration to and from the UK

# KEY CONTRIBUTORS TO PROJECTS ON SCHOLARLY MIGRATION AT MPIDR



**Emilio Zagheni**  
(MPIDR)



**Aliakbar Akbaritabar**  
(MPIDR)



**Andrea Miranda-Gonzalez**  
(UC Berkeley)



**Samin Aref**  
(U. of Toronto)



**Ebru Sanliturk**  
(MPIDR)



**Maciej J. Dańko**  
(MPIDR)



**Tom Theile**  
(MPIDR)

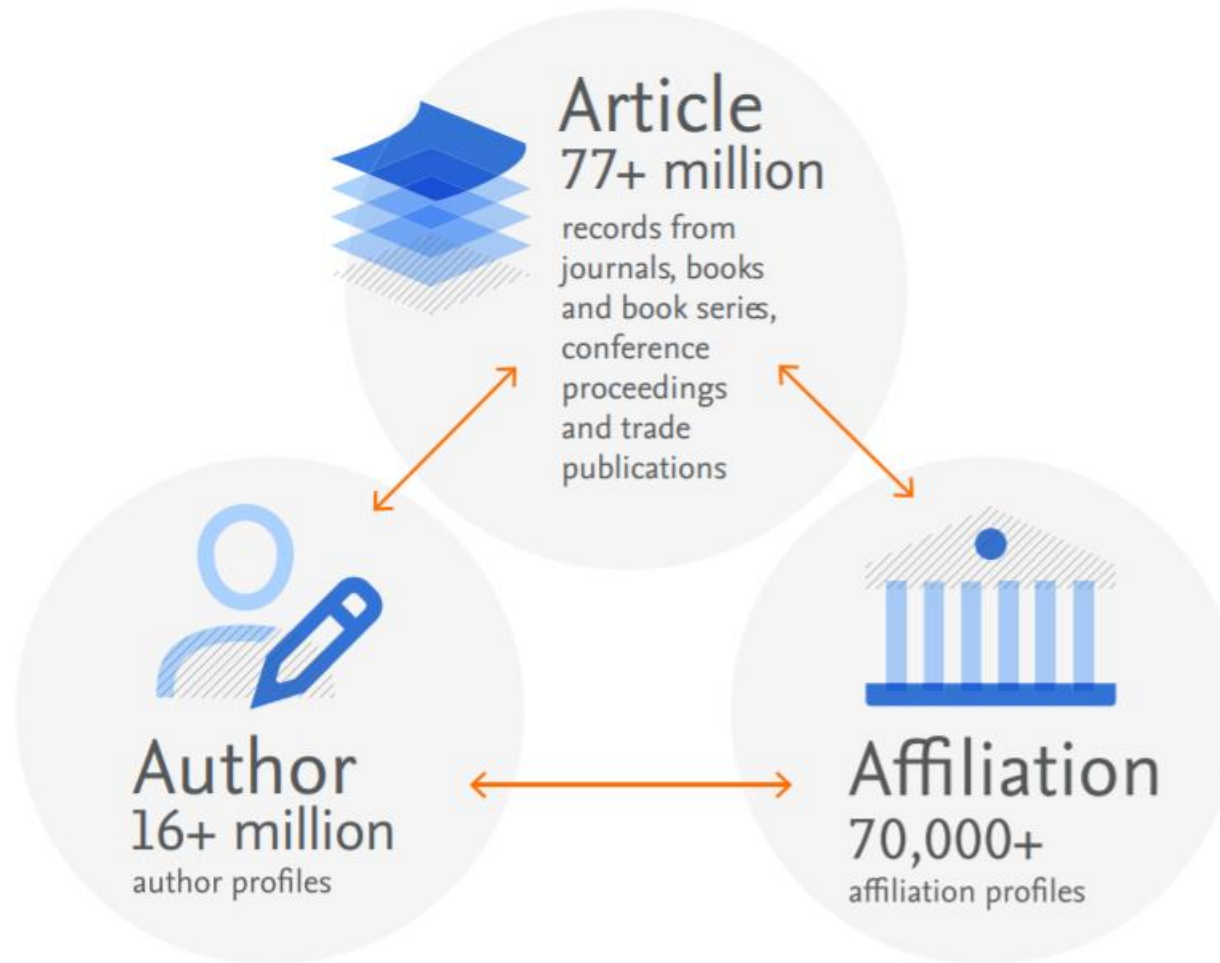


**Xinyi Zhao**  
(MPIDR)



## How can we use bibliometric data for migration research?

# Bibliometric Data



Data accessed via:



MAX PLANCK  
digital library

Source: [https://www.elsevier.com/\\_\\_\\_data/assets/pdf\\_file/0007/69451/Scopus\\_ContentCoverage\\_Guide\\_WEB.pdf](https://www.elsevier.com/___data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf)

# Frank W. Notestein



FRANK W. NOTESTEIN  
1902-1983



# HUMAN BIOLOGY

a record of research  
DECEMBER, 1934

VOL. 6



No. 4

## FERTILITY OF THE SOCIAL CLASSES IN THE NATIVE WHITE POPULATION OF COLUMBUS AND SYRACUSE<sup>1</sup>

BY FRANK W. NOTESTEIN AND CLYDE V. KISER  
*Milbank Memorial Fund*

**ONE** of the neglected aspects of the much discussed decline in the birth rate is the problem of the incidence of the decline in constituent classes of the population. We know that the birth rate has dropped more rapidly in the foreign than in the native population. But within the native group we know little enough of the nature of the changes that have taken place. Has the decline fallen with full force on the established families which have remained unbroken? Within this group are the birth rates declining with greater rapidity in some classes than in others, and if so, are these differences tending to strengthen or weaken the well-established inverse relation between social status and fertility?



# HUMAN BIOLOGY

a record of research  
DECEMBER, 1934

VOL. 6



No. 4

## FERTILITY OF THE SOCIAL CLASSES IN THE NATIVE WHITE POPULATION OF COLUMBUS AND SYRACUSE<sup>1</sup>

BY FRANK W. NOTESTEIN AND CLYDE V. KISER  
*Milbank Memorial Fund*

ONE of the neglected aspects of the much discussed decline in the birth rate is the problem of the incidence of the decline in constituent classes of the population. We know that the birth rate has dropped more rapidly in the foreign than in the native population. But within the native group we know little enough of the nature of the changes that have taken place. Has the decline fallen with full force on the established families which have remained unbroken? Within this group are the birth rates declining with greater rapidity in some classes than in others, and if so, are these differences tending to strengthen or weaken the well-established inverse relation between social status and fertility?



# HUMAN BIOLOGY

a record of research  
DECEMBER, 1934

VOL. 6



No. 4

## FERTILITY OF THE SOCIAL CLASSES IN THE NATIVE WHITE POPULATION OF COLUMBUS AND SYRACUSE<sup>1</sup>

BY FRANK W. NOTESTEIN AND CLYDE V. KISER  
*Milbank Memorial Fund*



ONE of the neglected aspects of the much discussed decline in the birth rate is the problem of the incidence of the decline in constituent classes of the population. We know that the birth rate has dropped more rapidly in the foreign than in the native population. But within the native group we know little enough of the nature of the changes that have taken place. Has the decline fallen with full force on the established families which have remained unbroken? Within this group are the birth rates declining with greater rapidity in some classes than in others, and if so, are these differences tending to strengthen or weaken the well-established inverse relation between social status and fertility?

ECONOMIC, FINANCIAL AND TRANSIT  
DEPARTMENT



## The Future Population of Europe and the Soviet Union Population Projections 1940-1970

BY

FRANK W. NOTESTEIN

IRENE B. TAEUBER

DUDLEY KIRK

ANSLEY J. COALE


LOUISE K. KISER

OF THE OFFICE OF POPULATION RESEARCH  
PRINCETON UNIVERSITY

LEAGUE OF NATIONS, GENEVA, 1944

## A simple Idea

**Changes in institutional affiliations can be used to infer changes in residence over time for individual scholars and for populations**



PNAS  
nexus

PAPER

Article title in sentence case

Janine Gonzales,<sup>a,\*</sup> Jan. Doe<sup>b</sup> and Michael Michelson<sup>b</sup>

<sup>a</sup>Laboratory for computational Science, Institute for advanced studies, Neutral-Zone-Str. 1, 10007 Berlin, MV, Germany and <sup>b</sup>Department for Bioinformatics, Max-Planck-Institute for Demography, Neutral-Zone-Str. 1, 10007 Berlin, MV, Germany

<sup>\*</sup>To whom correspondence should be addressed; email: M.Michelson@demographic.mpg.de

FOR PUBLISHER ONLY Received on Date Month Year; accepted on Date Month Year<sup>00</sup>

OXFORD GENETICS

A random paper title

Jan. Doe<sup>1,2</sup>, Tom Thelle<sup>1</sup> and Susanne Stein<sup>2</sup>

<sup>1</sup>Liverpool Centre for Demographic Science, Oxford, England

**Abstract**

The abstract should be written for people who may not read the entire paper, so it must stand on its own. The impression it makes usually determines whether the reader will go on to read the article, so the abstract must be engaging, clear, and concise. In addition, the abstract may be the only part of the article that is indexed in databases, so it must accurately reflect the content of the article. A well-written abstract is the most effective way to reach intended readers, leading to more robust search, retrieval, and usage of the article.



Please see additional guidelines notes on preparing your abstract below.

PNAS Nexus, 2008, pp. 1-7

doi: DOI HERE

Paper

**Bibliometric Research reveals hidden migration patterns**

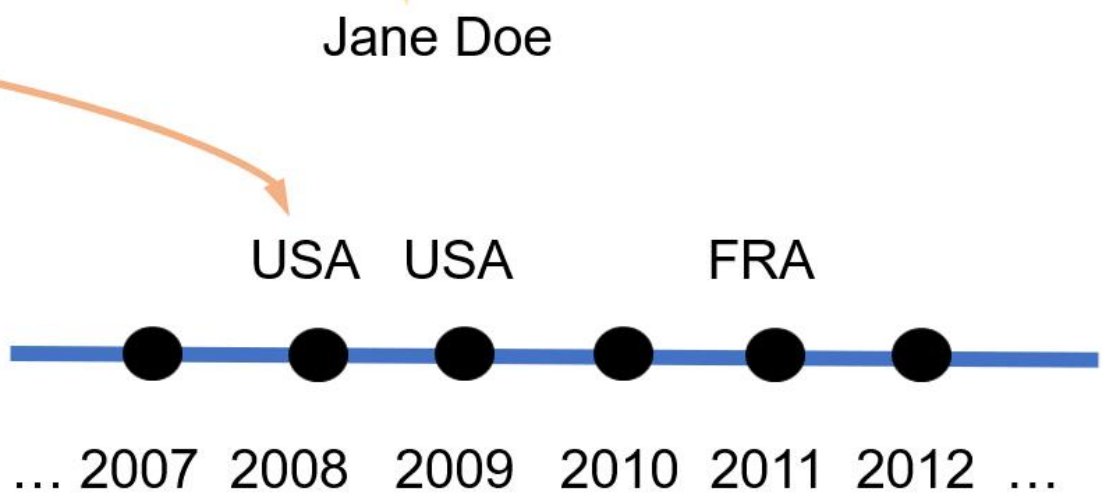
Jane Doe <sup>1</sup> e Thomas Mustermann <sup>2</sup>

<sup>1</sup>Universidade Federal de São João del Rei, DTECH, Ouro Branco, MG, Brasil.

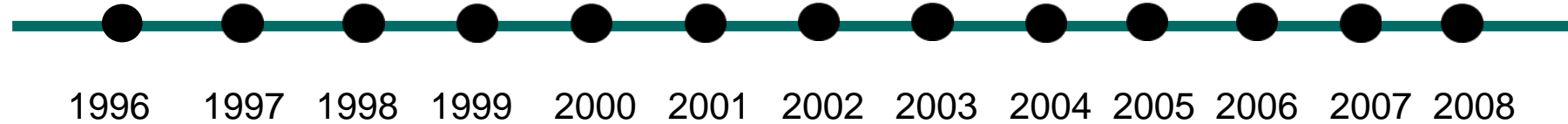
<sup>2</sup>Universidade Federal de Minas Gerais, Faculdade de Letras, Belo Horizonte, MG, Brasil.

**Abstract**

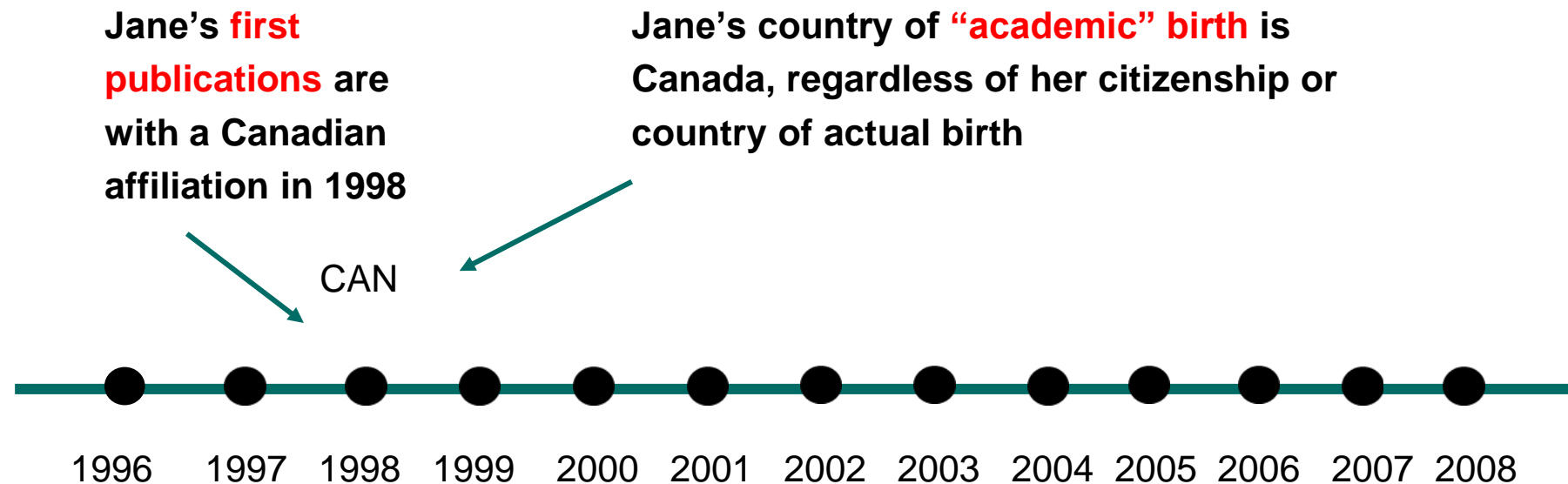
This article is a template article which aims to guide authors who will submit their papers to the *Texto Livre* journal. This template should facilitate their endeavor, viewing and configuring the text in the format already configured accordingly. Just write or paste your content in the desired places, replacing texts and figures, checking if they keep up the format. The abstract should have between 150 and 200 words.



## Identifying migration events: Illustrative Examples



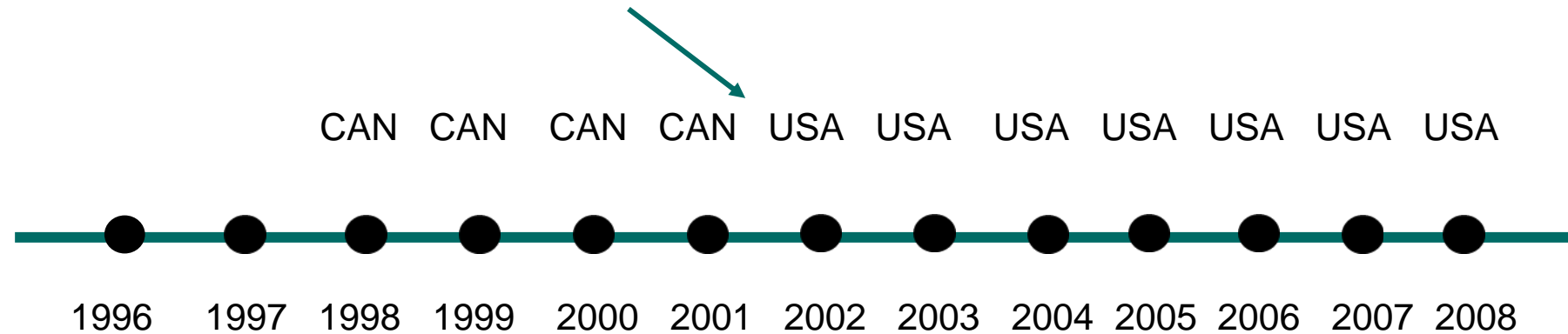
## Identifying migration events: Illustrative Examples



The country of residence is inferred as the **modal country of publications** in a given year

## Identifying migration events: Illustrative Examples

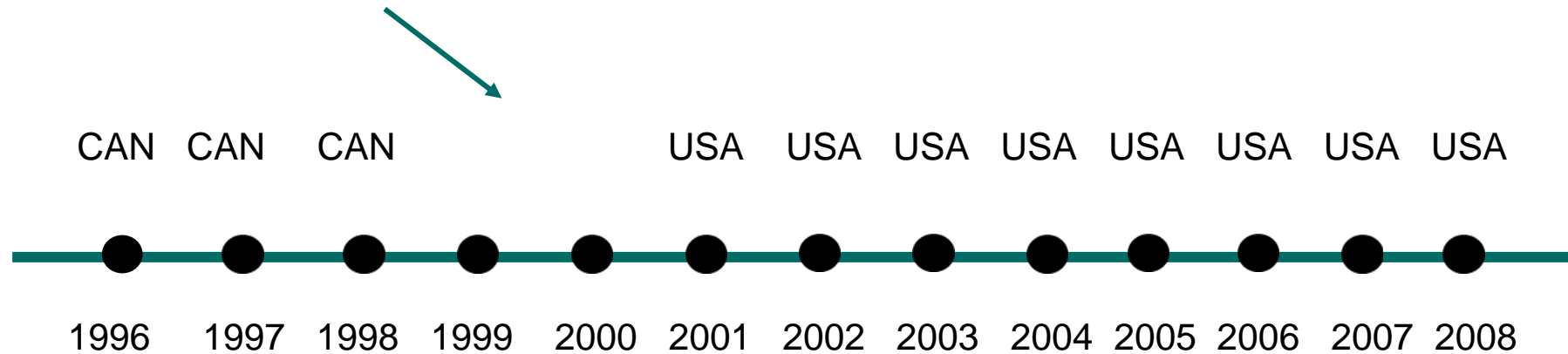
Jane changes residence  
from Canada to the US  
between 2001 and 2002



Maria's **modal countries**  
**of publications** over time

## Identifying migration events: Illustrative Examples

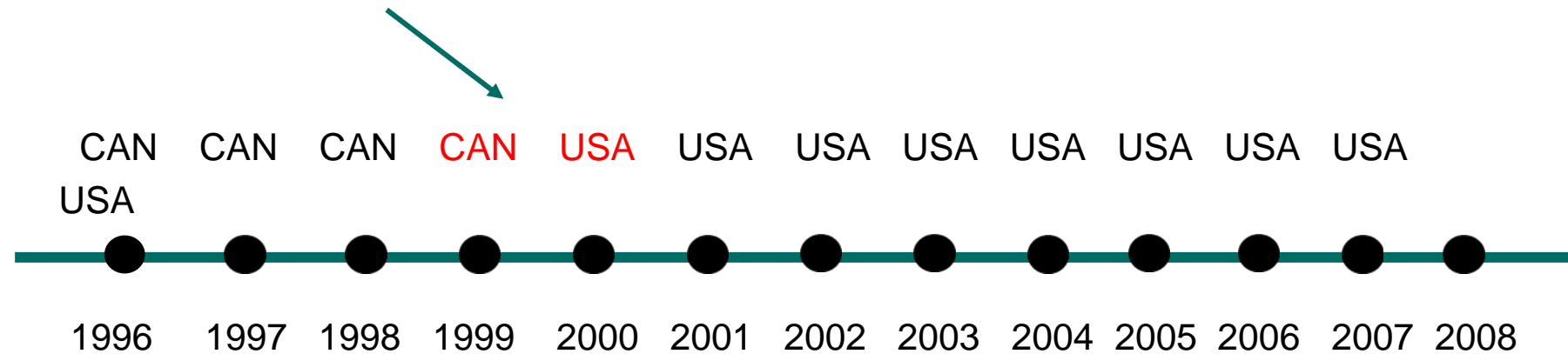
Jane has no publications in  
our database for 1999 and 2000





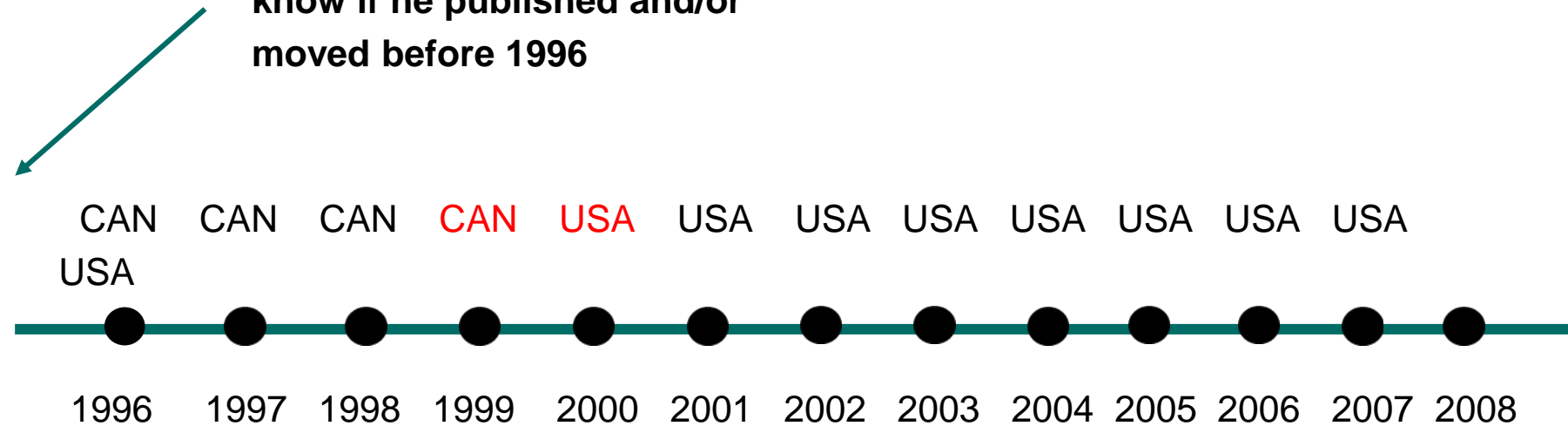
## Identifying migration events: Illustrative Examples

We impute the country values  
for up to 2 years from the  
closest observation with  
backward or forward filling

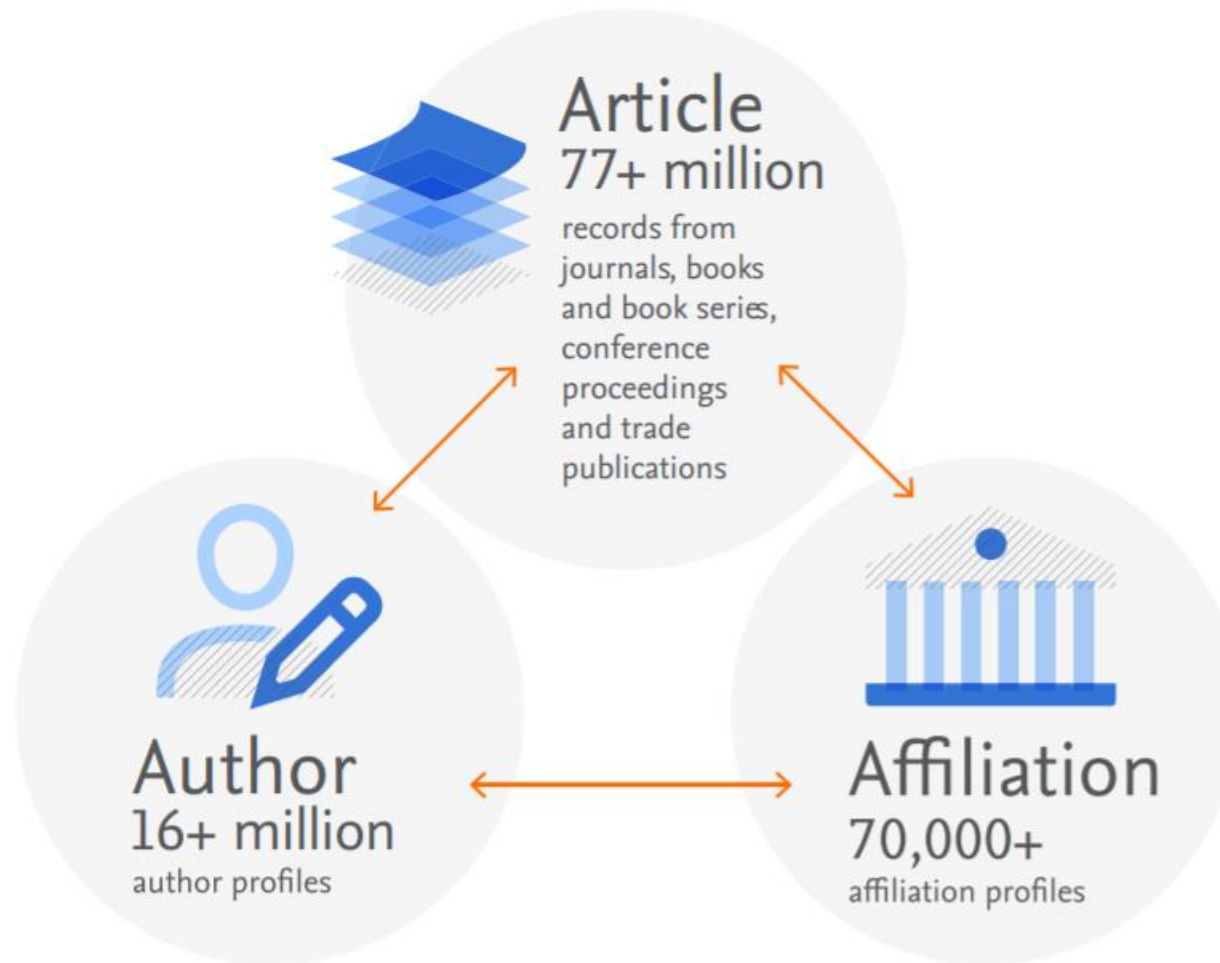


## Identifying migration events: Illustrative Examples

For Jane, we may have a left-censoring issue: we do not know if he published and/or moved before 1996



# A SCALABLE IDEA: THE SCOPUS DATABASE



Data accessed via:



MAX PLANCK  
digital library

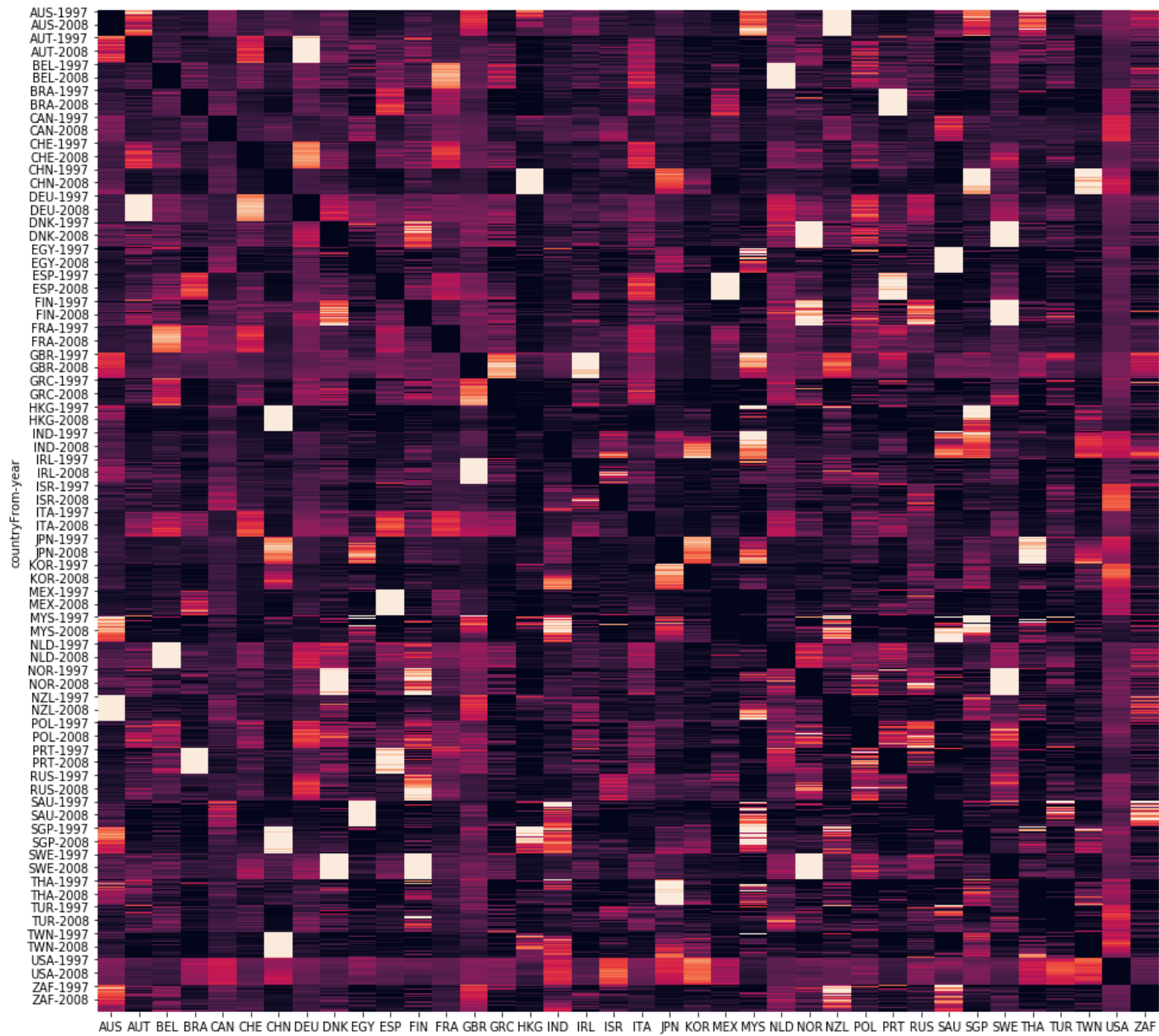
Source: [https://www.elsevier.com/\\_\\_data/assets/pdf\\_file/0007/69451/Scopus\\_ContentCoverage\\_Guide\\_WEB.pdf](https://www.elsevier.com/__data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf)

# Return Migration

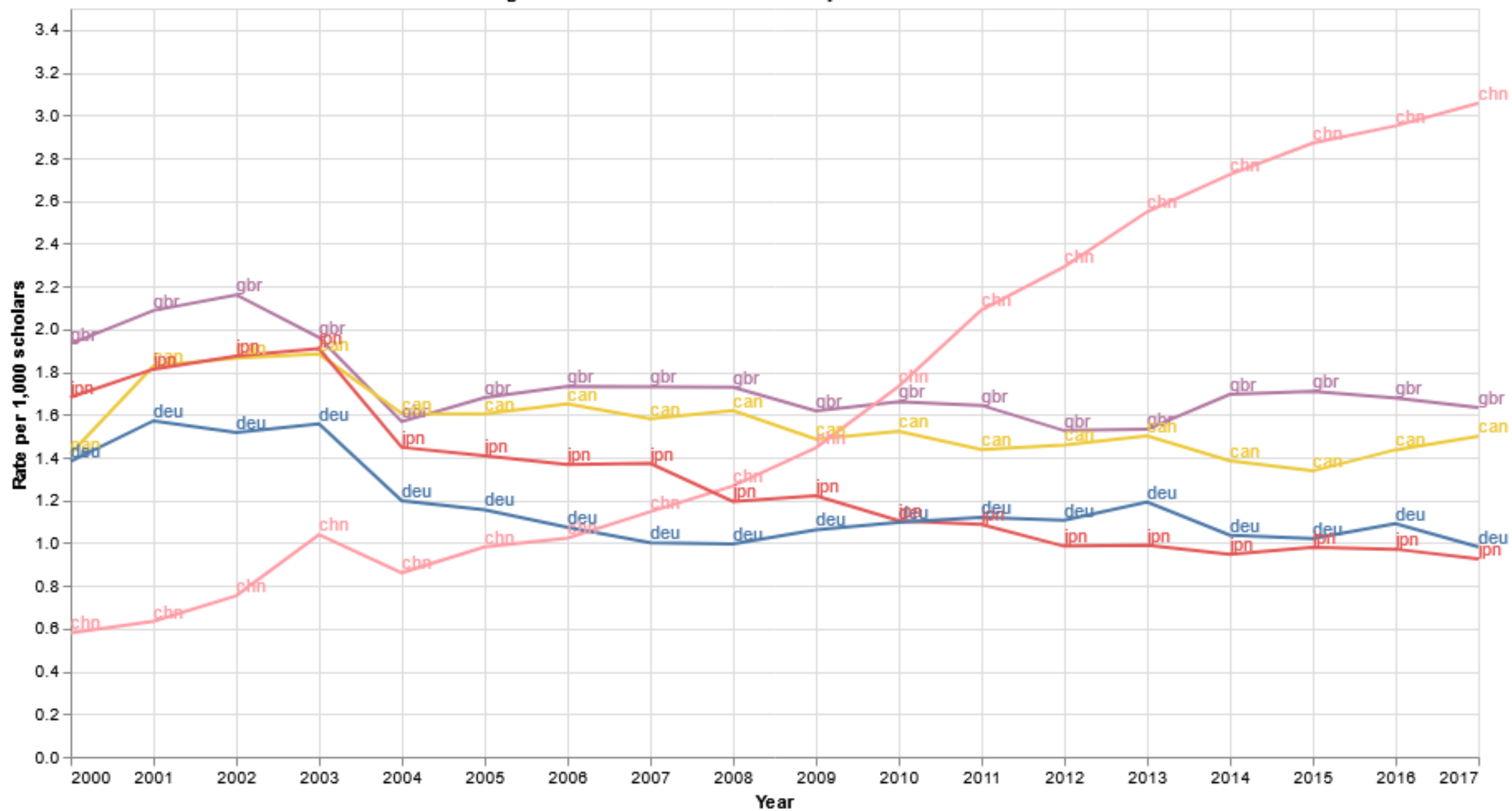
## Examples of aggregated data

Inmigration rates for selected countries





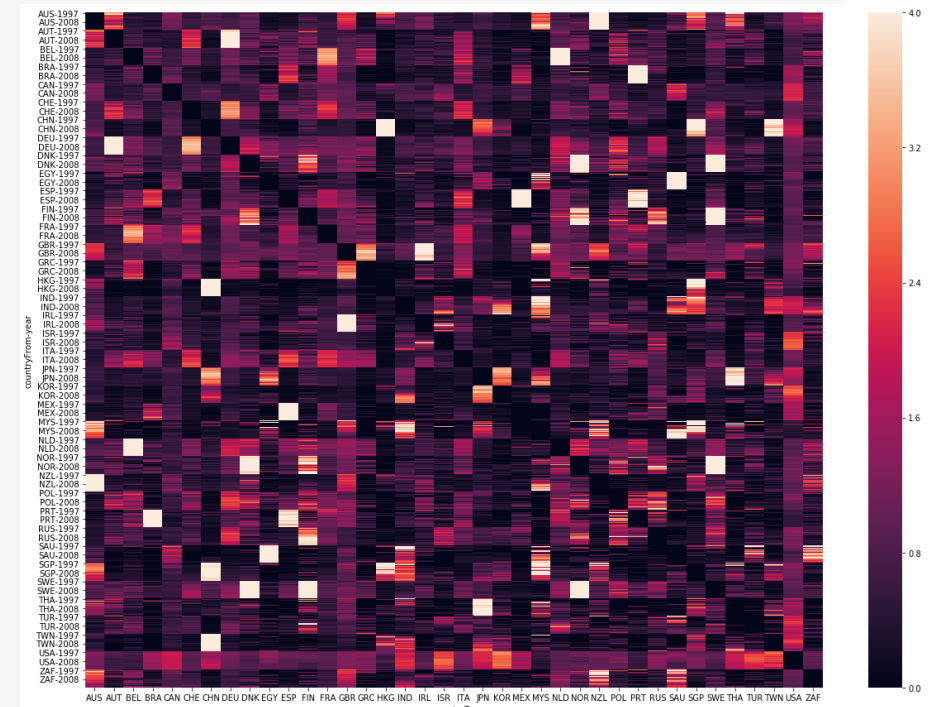
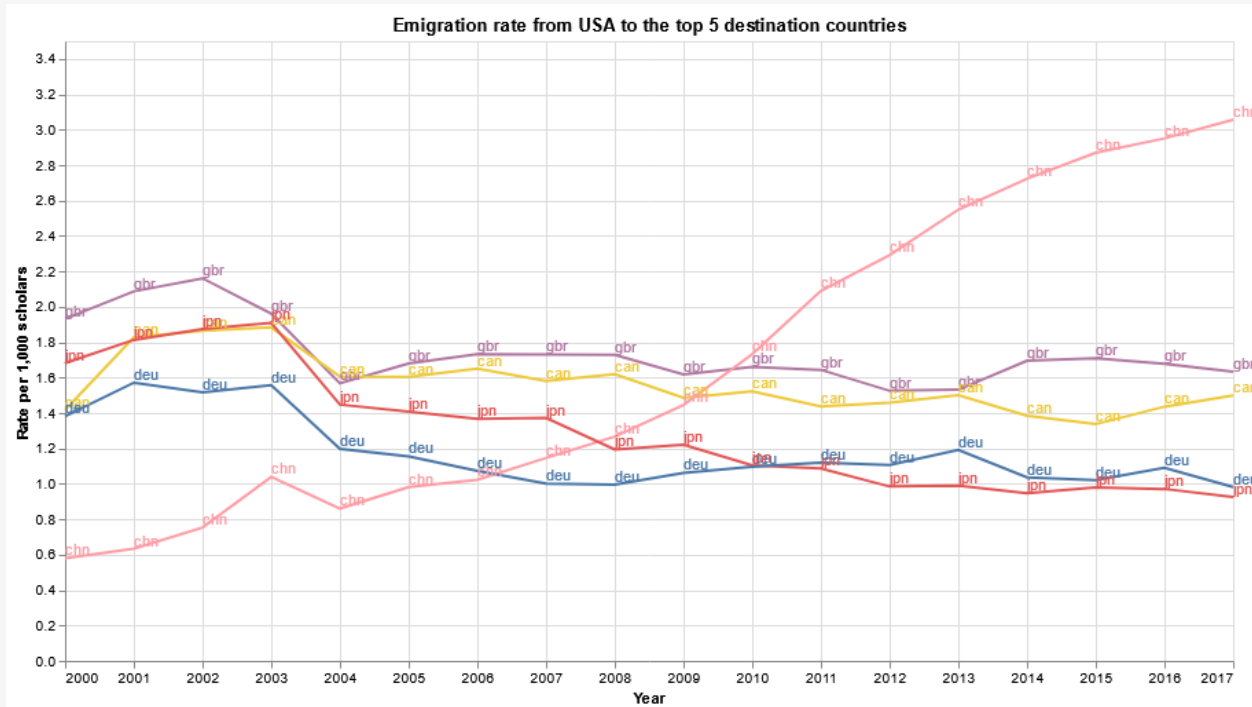
# Emigration rate from USA to the top 5 destination countries



# Return Migration

## Examples of aggregated data

### Migrationflow over time between pairs of countries



Adapted from: Author et al. (YYYY or publication status), Journal



# Return Migration

## Examples of aggregated data

### Migration flows from and to every country



*Adapted from: Author et al. (YYYY or publication status), Journal*

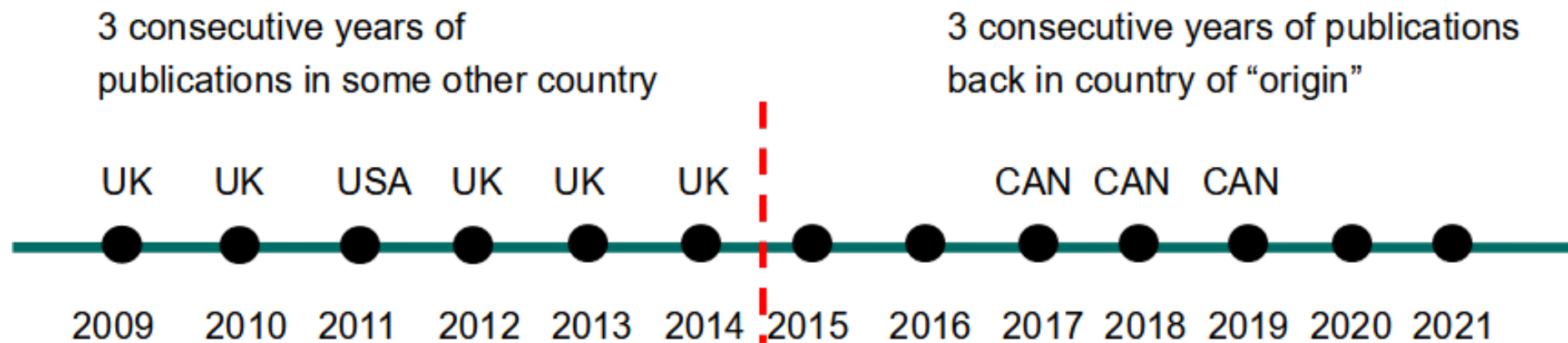
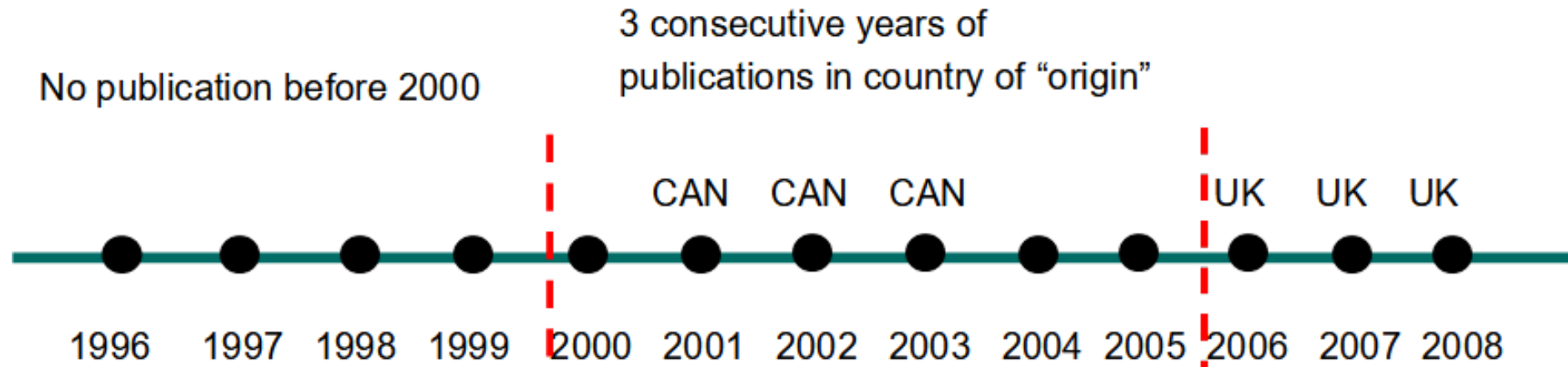


# Trends of return international migration worldwide

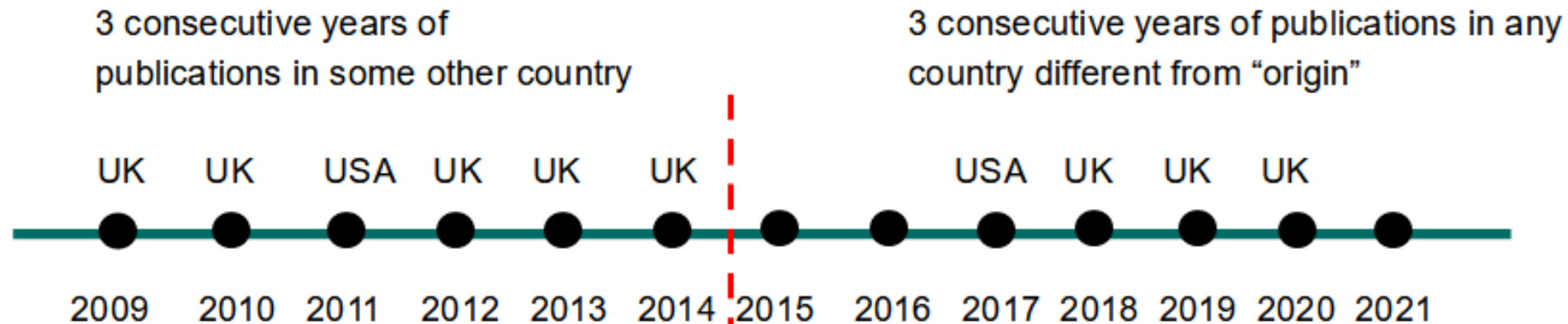
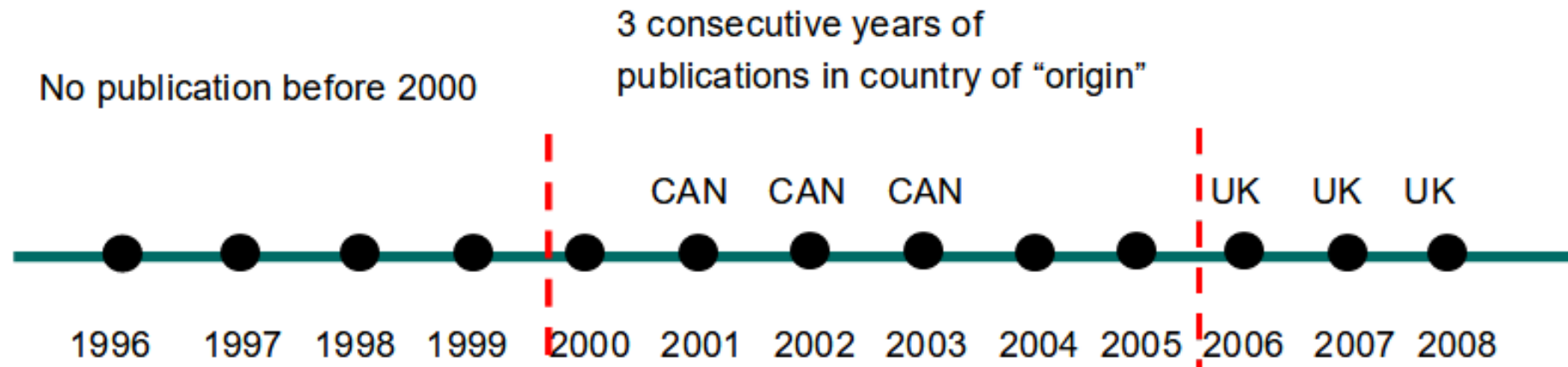
Tom Theile, MPIDR



## Return Migration - Identifying return migration - Numerator

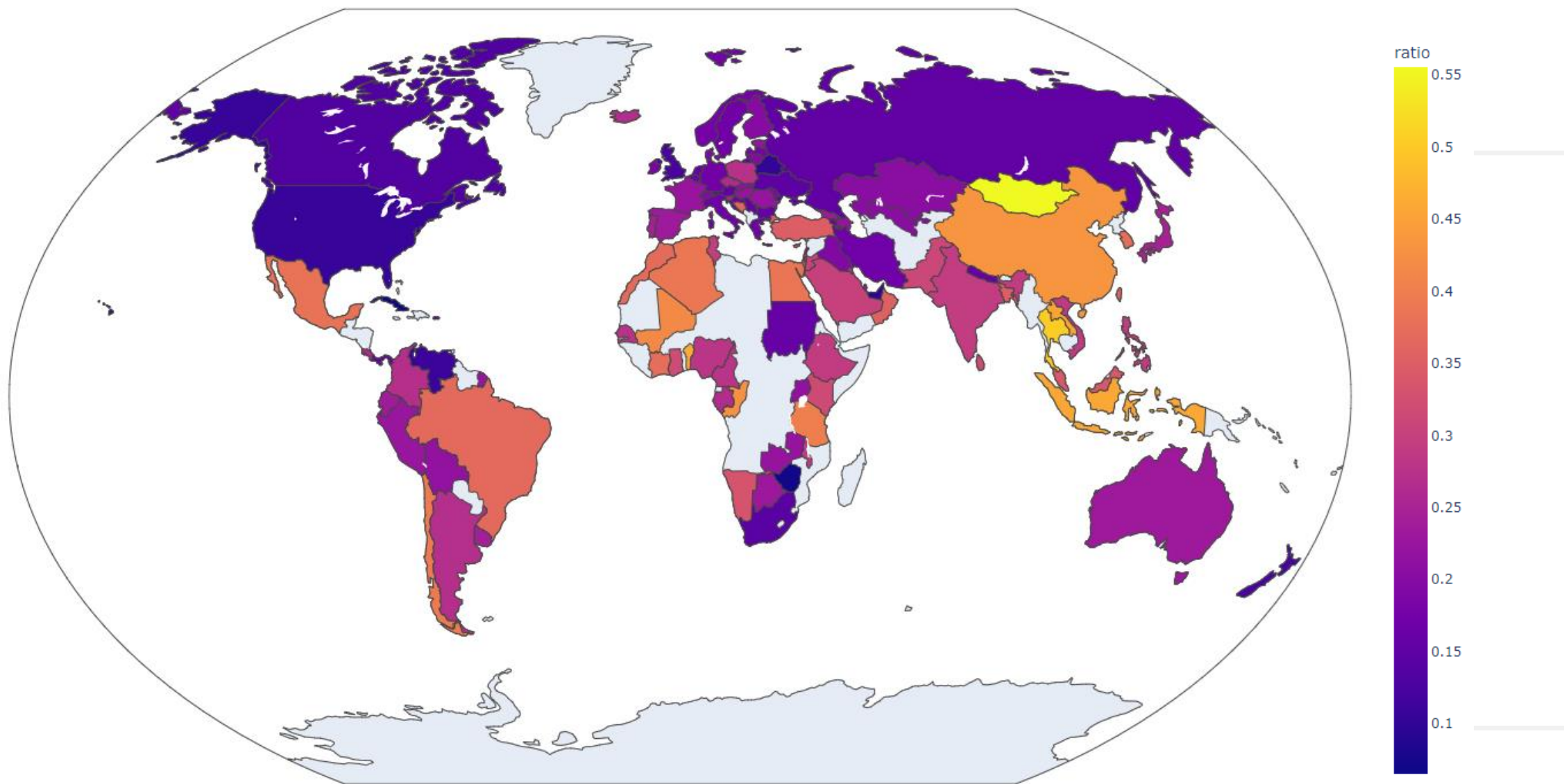


## Return Migration - Identifying return migration - Denominator

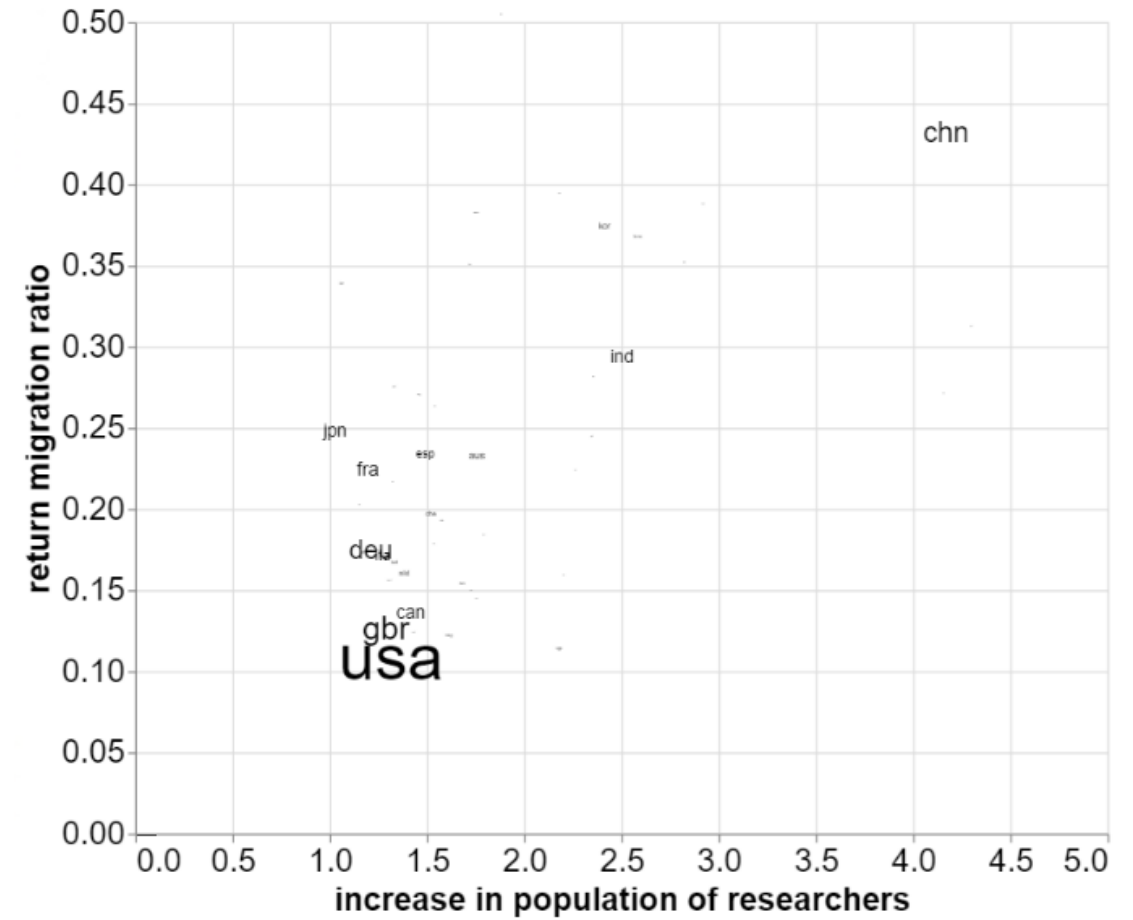
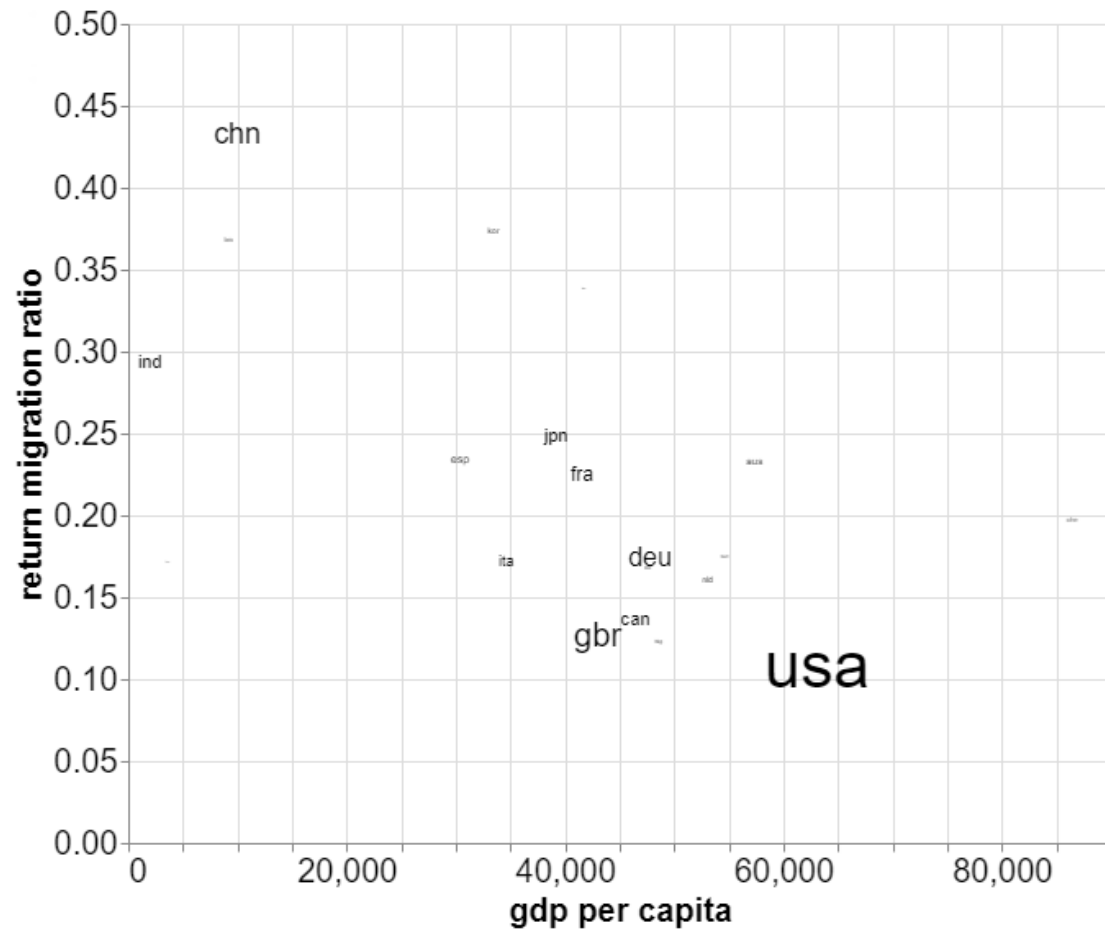




## Return Migration - Results



# Return Migration - Results & Conclusion







## Scholarly Migration Database

Quantifying the Mobility of Scholars

<https://www.scholarlymigration.org>



**Berkeley**  
UNIVERSITY OF CALIFORNIA



## Internal migration in Mexico

**Andrea Miranda Gonzalez, UC-Berkeley**  
@Andrea\_MirandaG

- **Andrea Miranda Gonzalez, UC-Berkeley:** Internal migration of scholars in Mexico

**Question:** How has migration of scholars within Mexico changed **over time** and **between regions**?

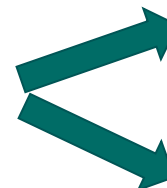
- Why internal migration?
  - Lots of research on *international* migration, but we know that *internal* migration is more common
  - Provide framework to assess sub-national estimates of migration of scholars
  - **Policy:** Understanding where people move to and from allows us to identify regional deficits and areas of progress for investment in human capital.



**Data:**  
Scopus authorship  
records with  
affiliations in Mexico  
(1.1M, 1996-2018)



**Data processing:**  
identify state of  
affiliation using rule-  
based algorithm and  
neural network



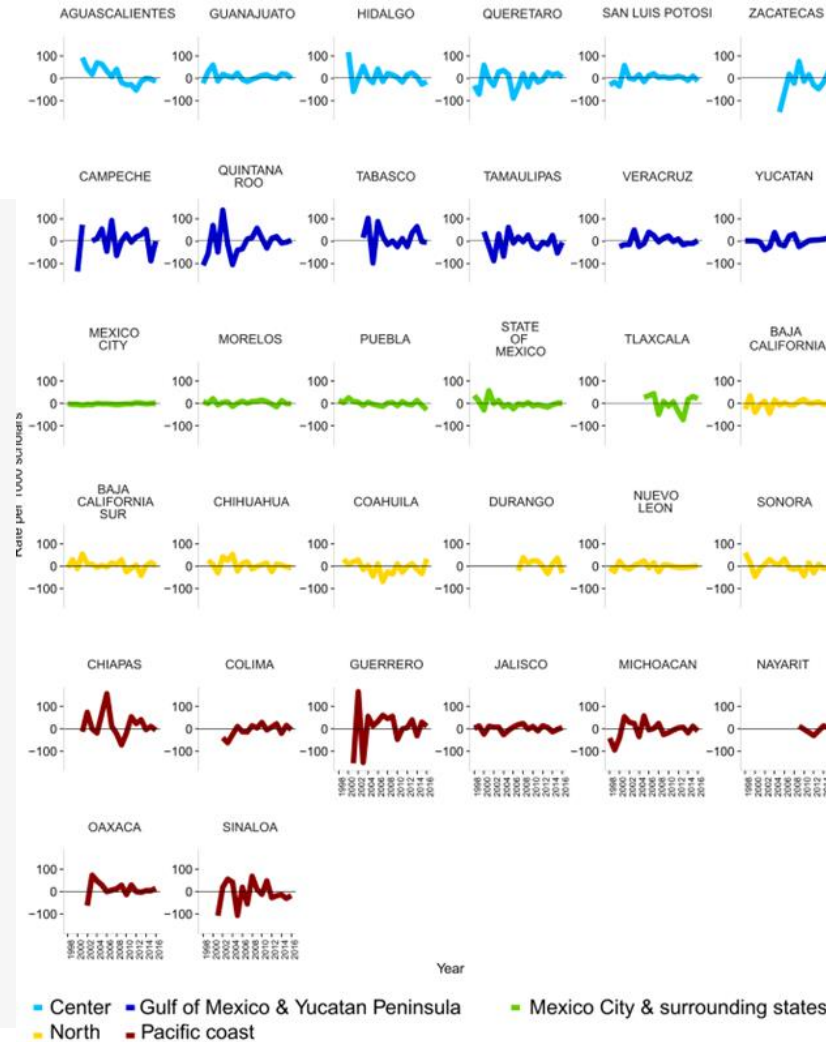
Migration rates and measures of  
redistribution

Migration networks and detection of  
communities

# Net migration rates of scholars in Mexico

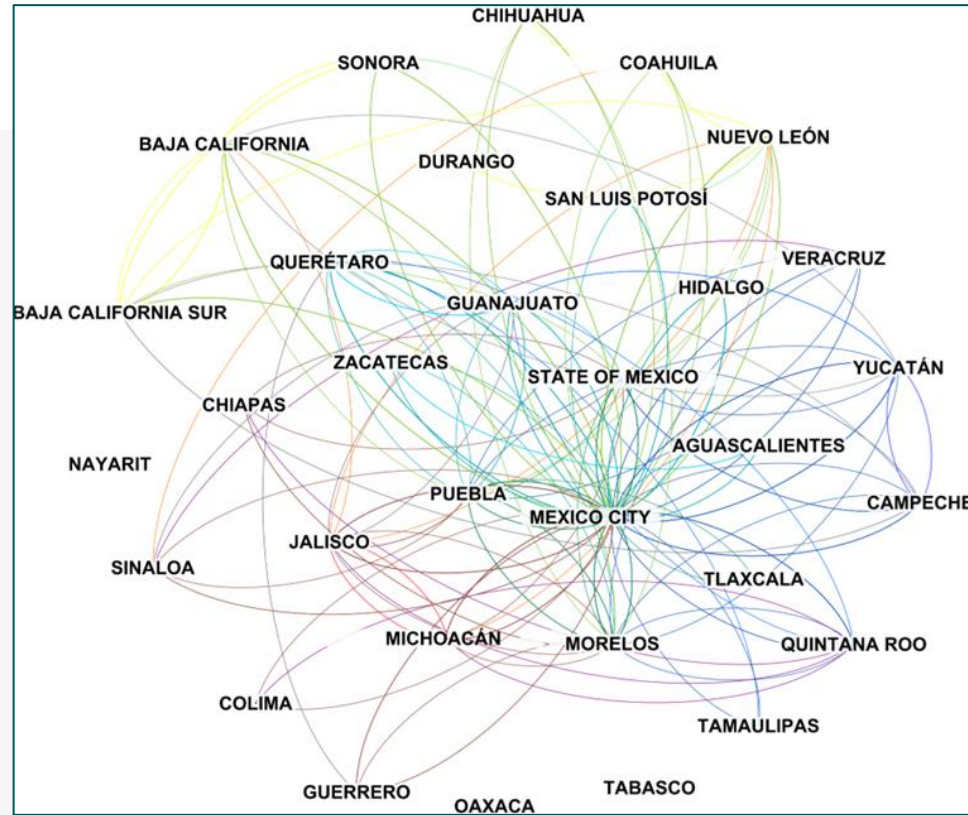
Types of net migration rates:

1. Steady oscillations around zero
2. Downward/upward period-specific trends
3. Large variation

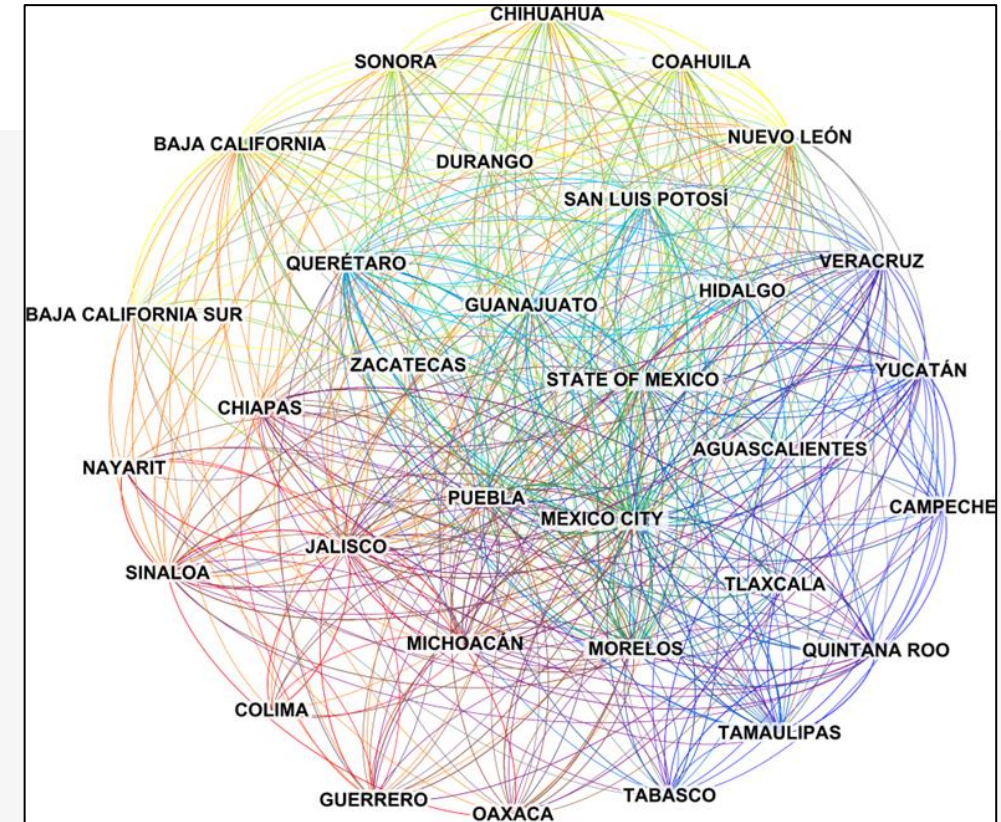


**Positive** rates: **attraction** of scholars  
**Negative** rates: **departure** of scholars

# Migration as a network



1997-1998

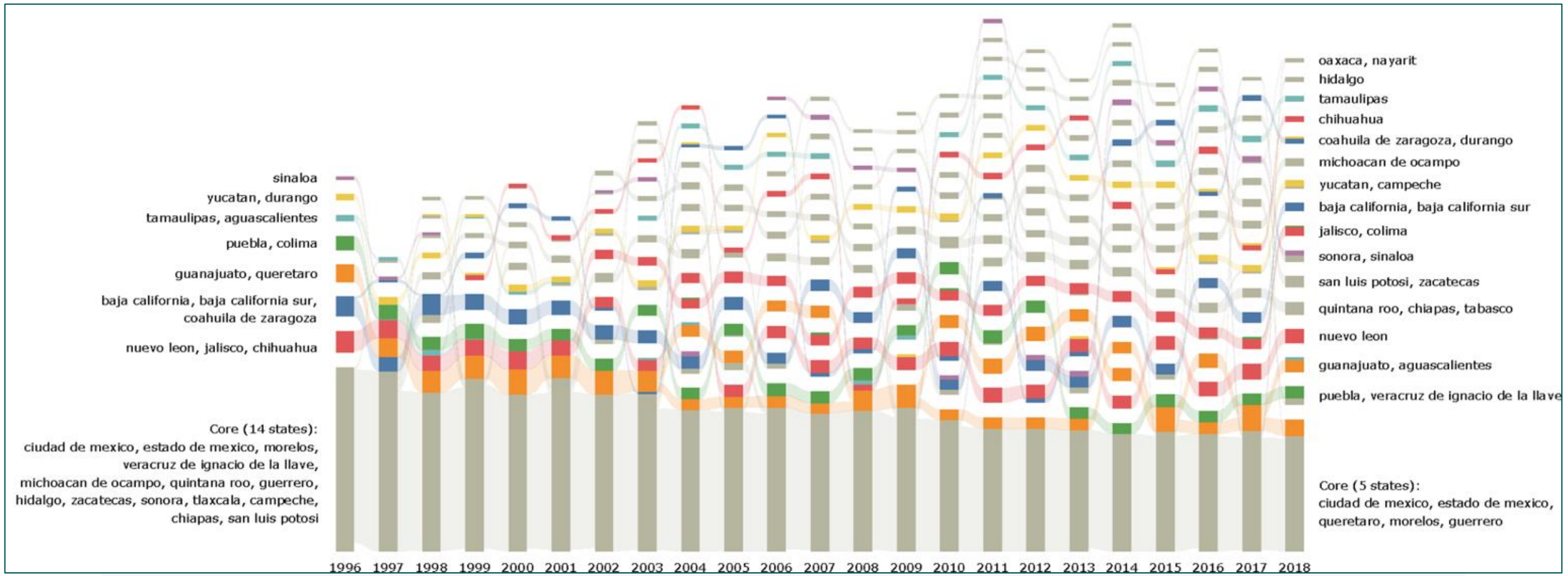


2017-2018

Adapted from: Miranda-González et al. (2020) EPJ Data Science



# Detection of communities and changes over time: core and periphery



Adapted from: Miranda-González et al. (2020) EPJ Data Science

## Conclusion

- We suggest a framework to repurpose bibliometric data and obtain migration estimates at a sub-national level
- We find that
  - There is heterogeneity in migration rates and patterns
  - Many mobile authors move to or through Mexico City during their academic life course, but there are other regions where mobility has increased
  - More communities and evidence of core-periphery structure.

**The availability of large-scale bibliometric data allows us to track the migration of researchers in a way that has not been possible with traditional sources of migration data, like censuses and surveys.**





# Brexit's effect on scholarly migration to and from the UK

Asli Ebru Sanlitürk, MPIDR

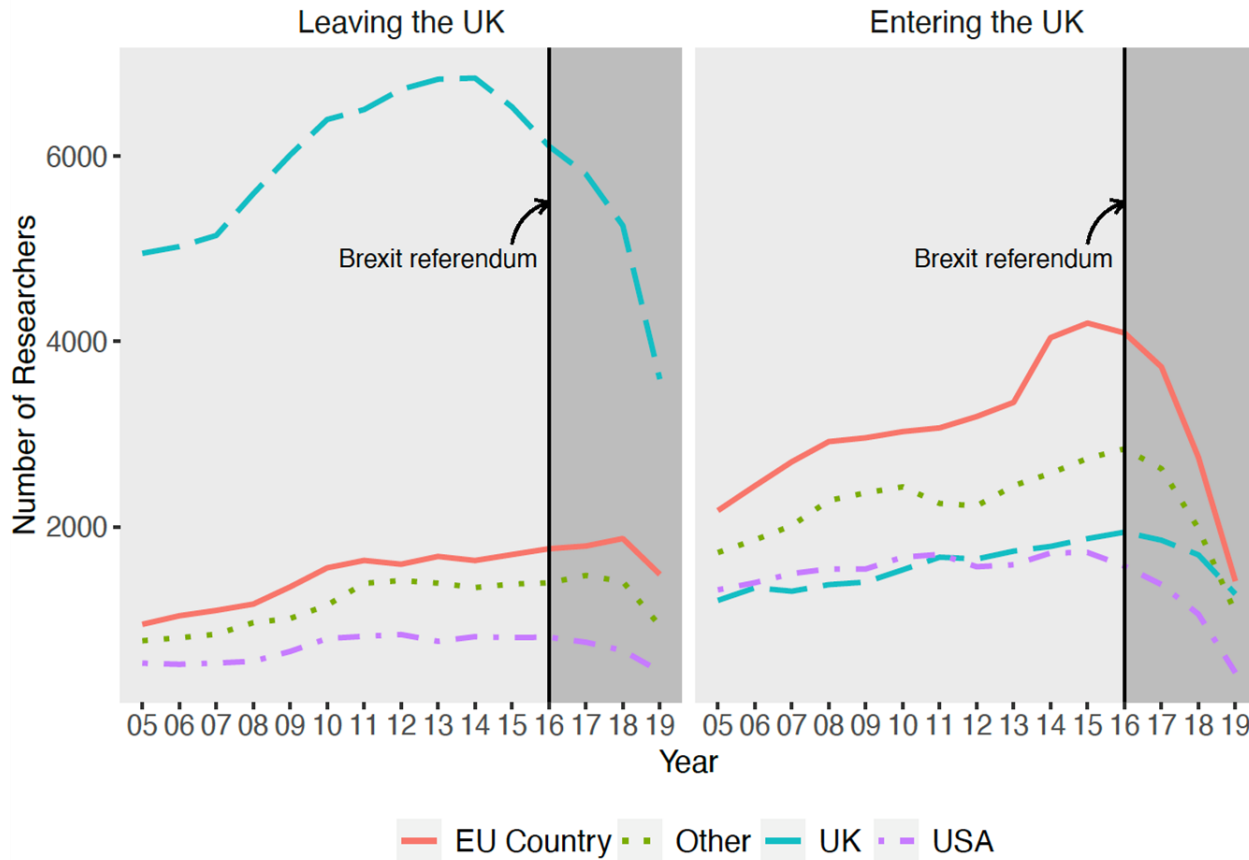
- **Asli Ebru Sanlitürk, MPIDR:** Brexit's effect on scholarly migration to and from the UK

**Question:** How did the Brexit referendum affect scholarly mobility patterns?

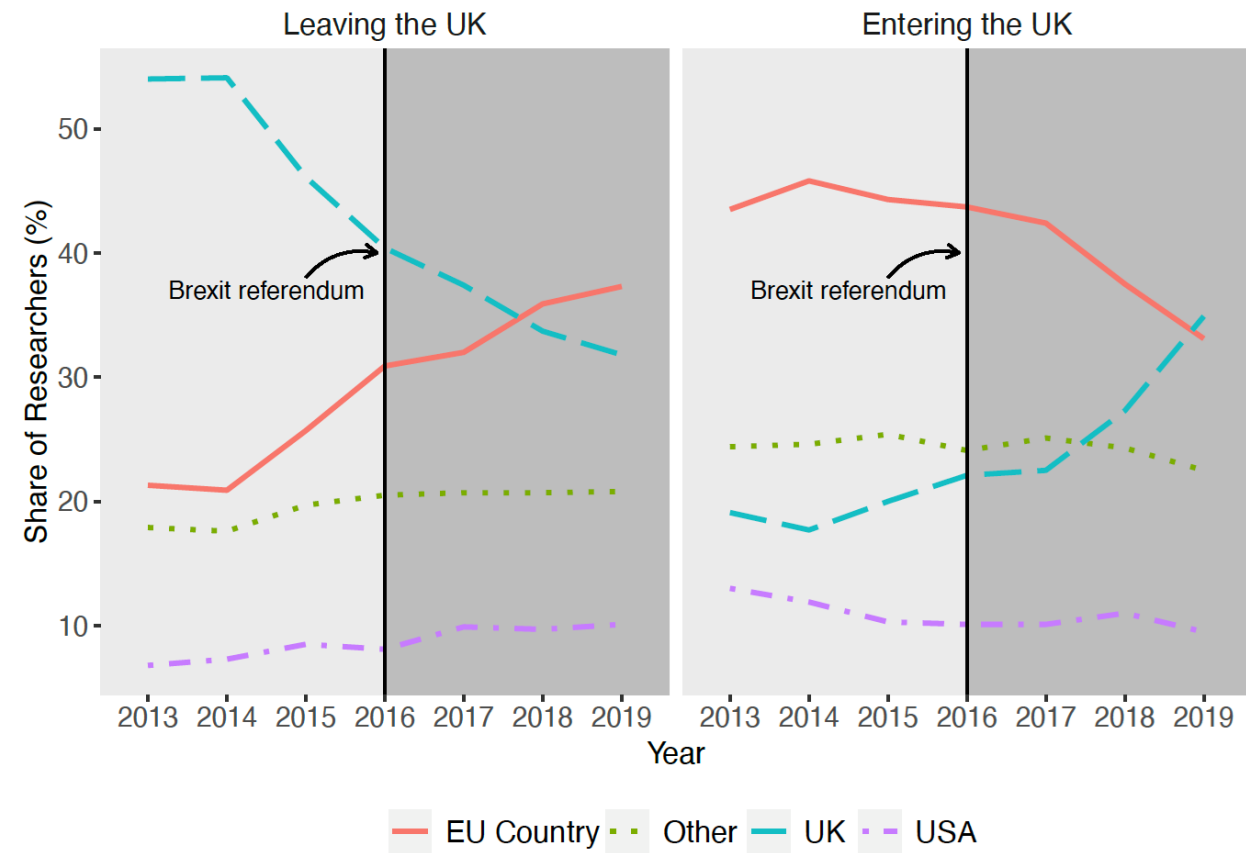
- The main aim is to analyze the scholarly migration patterns before and after the Brexit referendum (2016)
- Based on available data, we set the time frame of analysis as 2013-2019
- As we are focusing on the most recent part of data, right-censoring created an issue. To overcome this issue, we focus on *active researchers*, i.e. researchers whose residence country can be determined for each year between 2013-2019.
- N = 45,316. Selection criteria:
  - Internationally mobile
  - Ties with UK
  - Active researchers
- Random effects logistic regression model

# Descriptive analysis

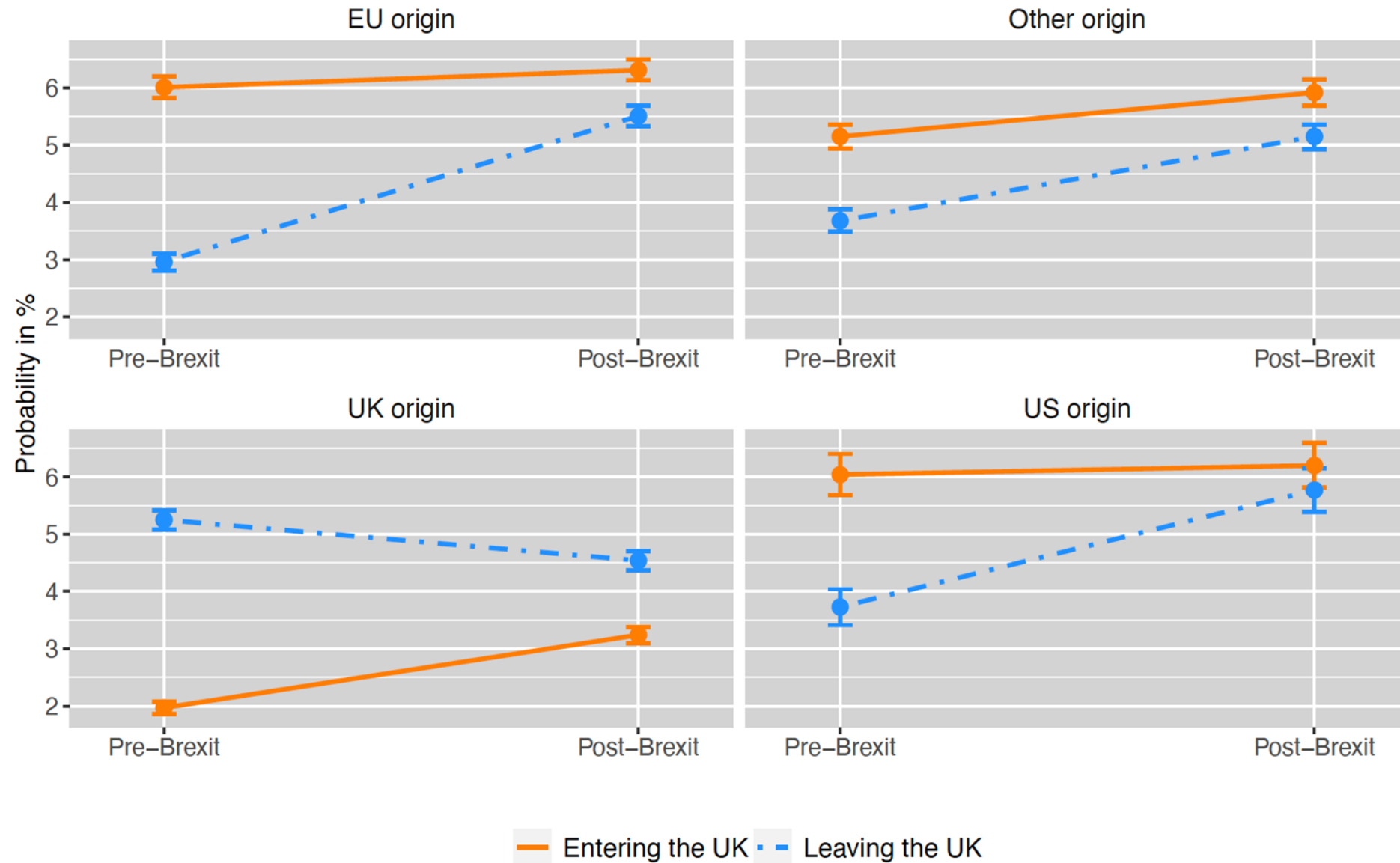
## Without restricting to active researchers



## Change in the share of active researchers



## Statistical analysis



## Conclusion

In our analysis on internationally mobile active researchers we do not observe a brain drain as an initial reaction to the result of the referendum, but a compositional change, which if continues, may diminish the academic diversity in the UK.

- Among the active researchers; the probability of leaving the UK increases after Brexit for all academic origin groups except for the UK academic origin.
  - EU academic origin: probability increases from 3% to 5,5% (increase by ~86%)
  - UK academic origin: probability decreases from 5,3% to 4,5% (decrease by ~14%)
- Probability of entering the UK increases for active researchers with UK academic origin (moving back to the UK) from near 2% before- to 3,2% after Brexit referendum (increase by ~65%).



# Development and international scholarly migration

Maciej, Ebru, Emilio, Ali and Tom

## Development and international scholarly migration

**Question:** Does economic development **increase** or **decrease** emigration rate of academic scientists?

Extensively studied relationship between level of development and emigration rate of entire populations

- First emigration rate increases, but then it decreases with income / GDP (inverse U-pattern, e.g. Clemens and Mendola 2020).

However, migration processes are heterogeneous

- Different population groups may have different patterns of emigration with income / GDP

Here we focus on one specific group

- Active academic scientists vs. different measures of GDP

## Statistical analysis

**Goal:** Estimate “baseline” pattern of emigration rate against log GDP using GAMM that “controls” for time and country-specific (random) effects.

**Method:** Generalized Additive Mixed Model (GAMM) allows to model non-linear relationships using penalized smooth functions “s()”:

$$\log( E(Y) ) = \beta_0 + s(\log GDP) + s(year) + s(country) + s(country, \log GDP) + \log exposures$$

Expected emigration counts

Smooth functions of log(GDP) and year

*Using negative-binomial distribution to account for overdispersion*

Link function

Intercept

Random intercept for countries

Random slope

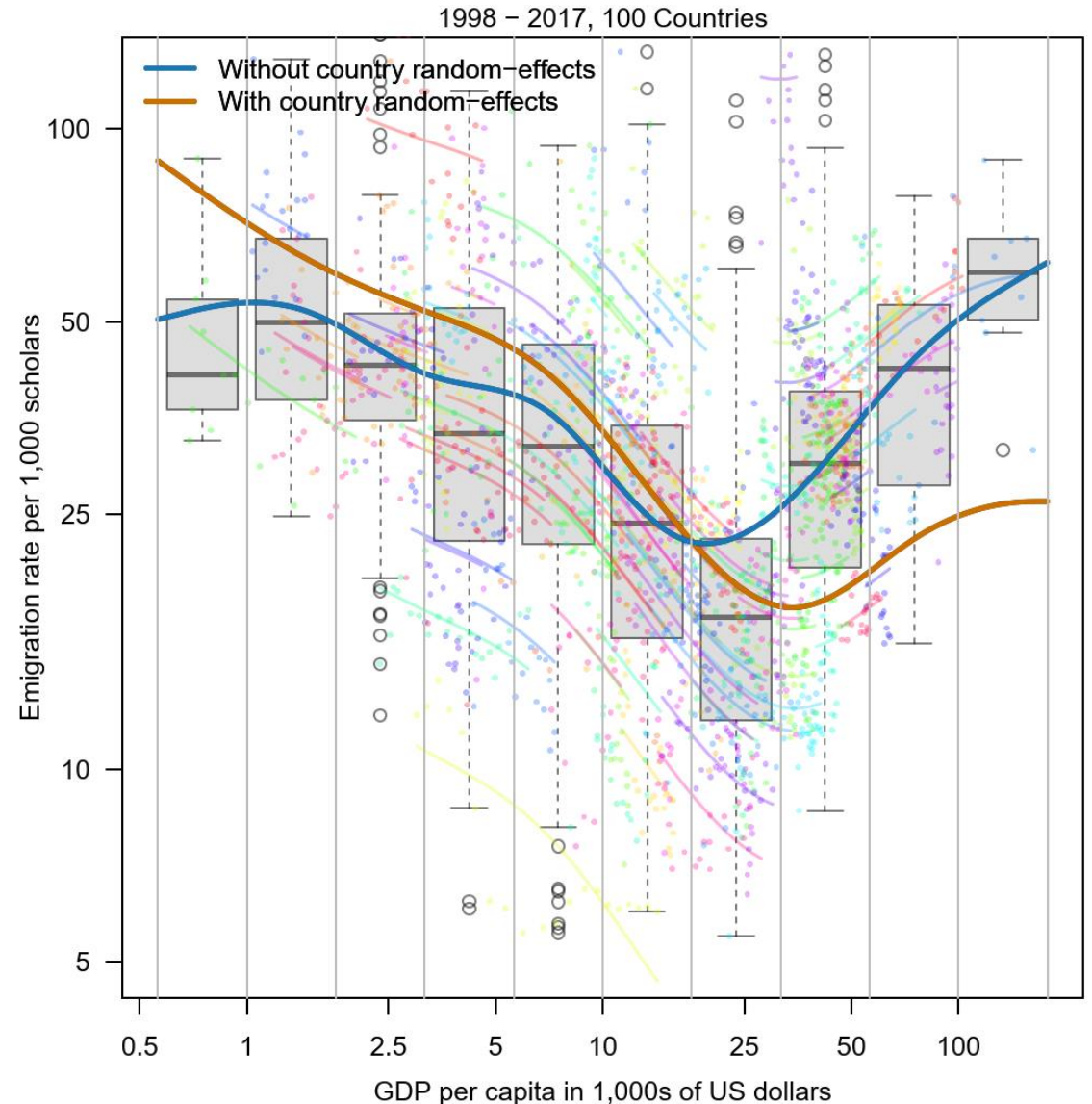
Offset used to model the rates.  
exposures = number of scholars



## Main result

1998-2017 averaged emigration rates predicted by the model as a function of **GDP per capita measured in 2017 constant international dollars at purchasing power parity**.

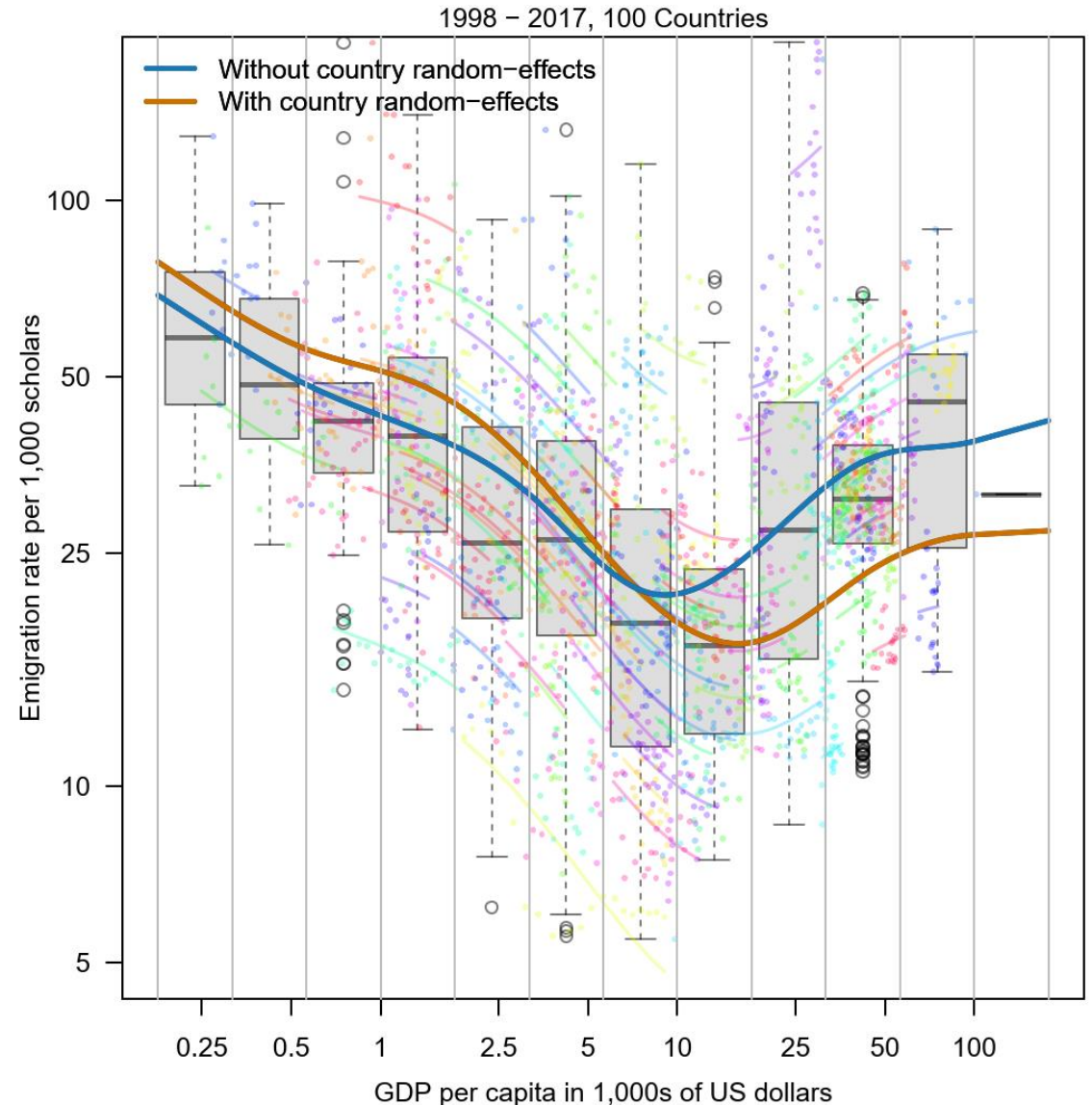
The combined data for each year and country is plotted as a **box plot**, while the country- and year- specific data are plotted using **colored dots**. Each color represents a separate country.



## Sensitivity analysis

1998-2017 averaged emigration rates predicted by the model as a function of **GDP per capita measured in 2010 constant international dollars**.

The combined data for each year and country is plotted as a **box plot**, while the country- and year- specific data are plotted using **colored dots**. Each color represents a separate country.



## Conclusion

We observe **opposite pattern** of emigration rates with GDP per capita among scholars than it was observed in total populations.

- First emigration rate **decreases**, but then it **increases** (U-pattern).
- **Hypothesis:**
  - Researchers living in poor countries are more willing to move to richer countries.
  - Richer countries offer relatively bigger number of available positions, promote hiring researches from poor countries, but also have higher circulation of researchers (e.g., due to short term positions).
  - Researchers in poorer countries are a privileged subset of the population that does not have the same barriers for emigration

Future research: role of gender, type of institutions, etc.



# Gender perspective in international scholarly migration

**Xinyi Zhao, MPIDR**

- **Xinyi Zhao, MPIDR:** Gender perspective in international scholarly migration

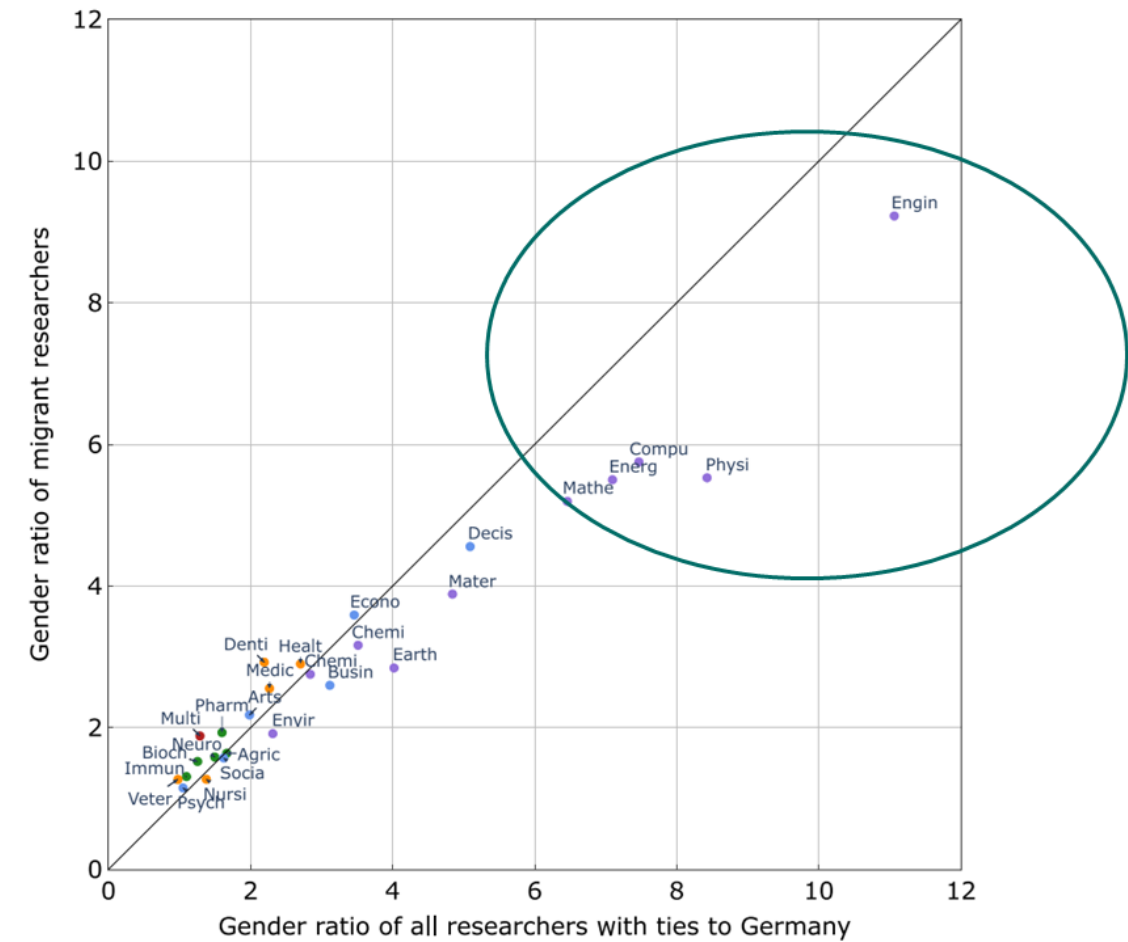
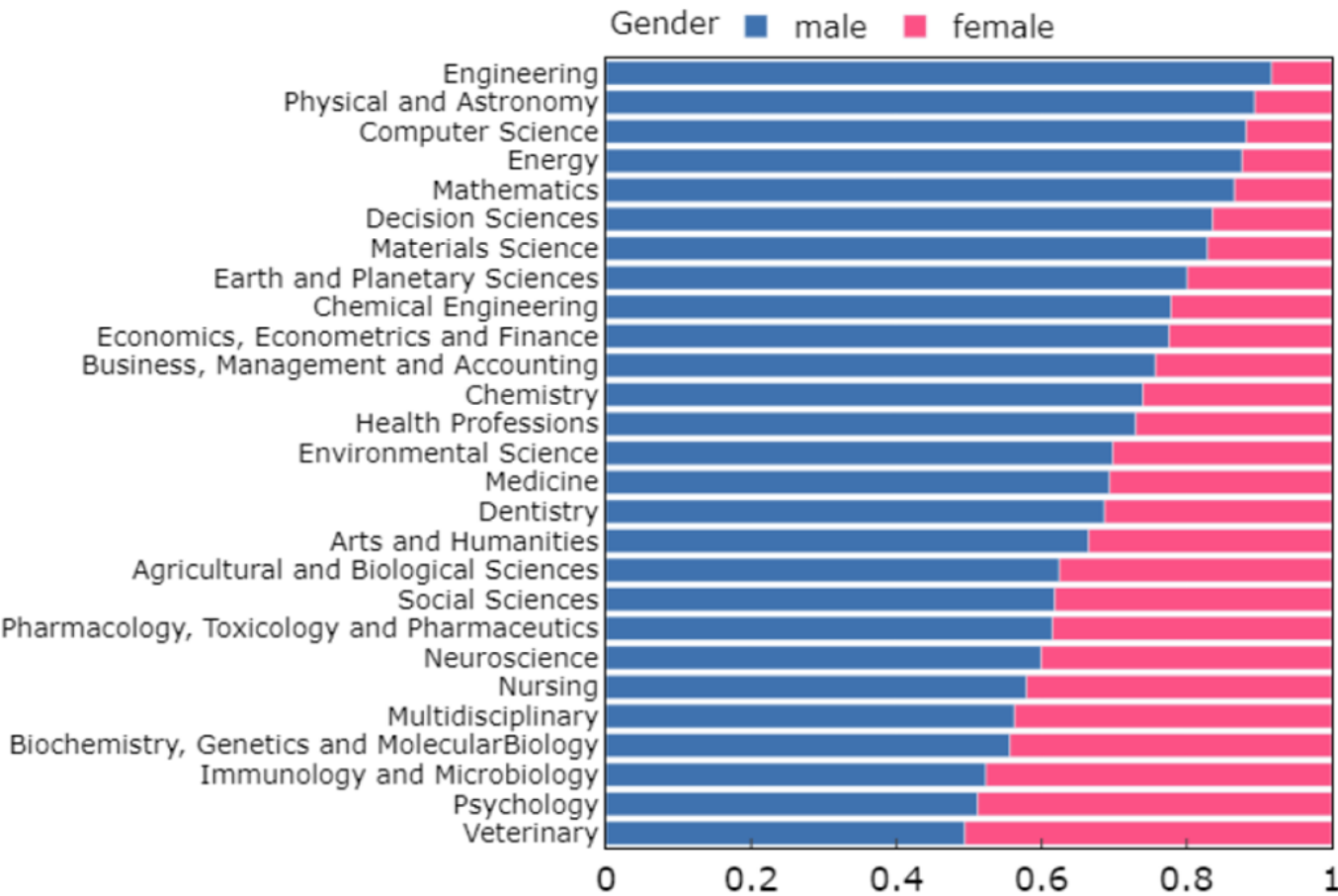
**Question: To what extent does gender inequality in international migration of German-affiliated researchers?**

Research population: German-affiliated published researchers during the 1996–2020 period

Process:

- Gender inference: *Genderize.io*
- Academic life: starting from the first publication
- International migrant: authors who have published in another country outside their country of origin
  - *Outward researchers* (origin: Germany; current country: not Germany)
  - *Returnees* (origin and current country: Germany; but with another country being the researcher's mode country at some point in time)
- Discipline inference:
  - All Science Journal Classification (ASJC) codes: 26 academic disciplines
  - Topic modeling by inferring the latent topical structure of textual bibliometric data

# Main result

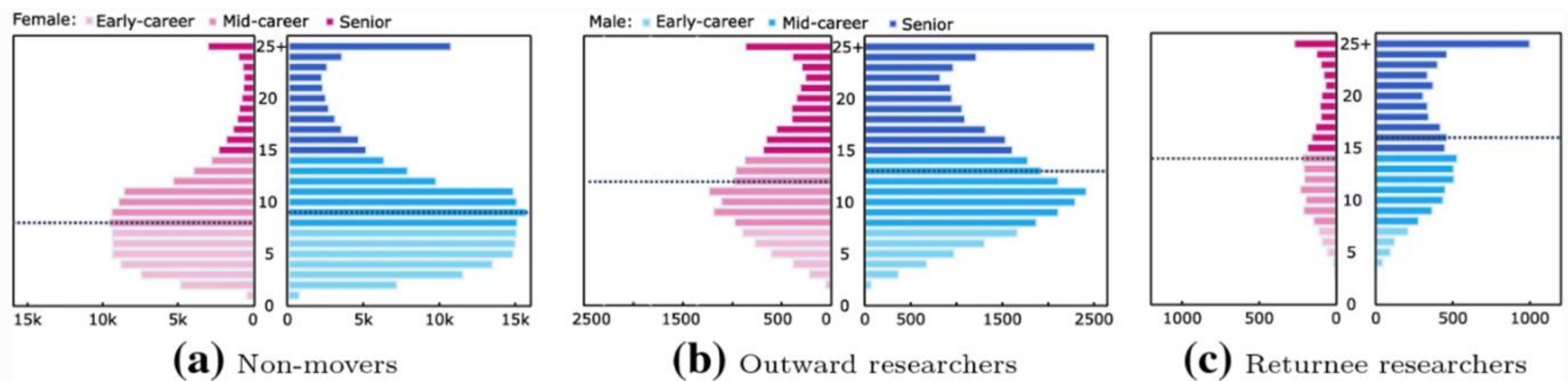


Zhao, X.Y., Aref, S., Zagheni, E., and Stecklov, G. (2021). "International migration in academia and citation performance: an analysis of German-affiliated researchers by gender and discipline using Scopus publications 1996-2020" In: *Proceedings of the 18th International Conference on Scientometrics and Informetrics (ISSI 2021)*, Leuven, Belgium, July 12-15, 2021. Leuven: ISSI. <https://arxiv.org/abs/2104.12380>

$$\text{gender ratio} = \frac{\text{Number of male researchers}}{\text{Number of female researchers}}$$



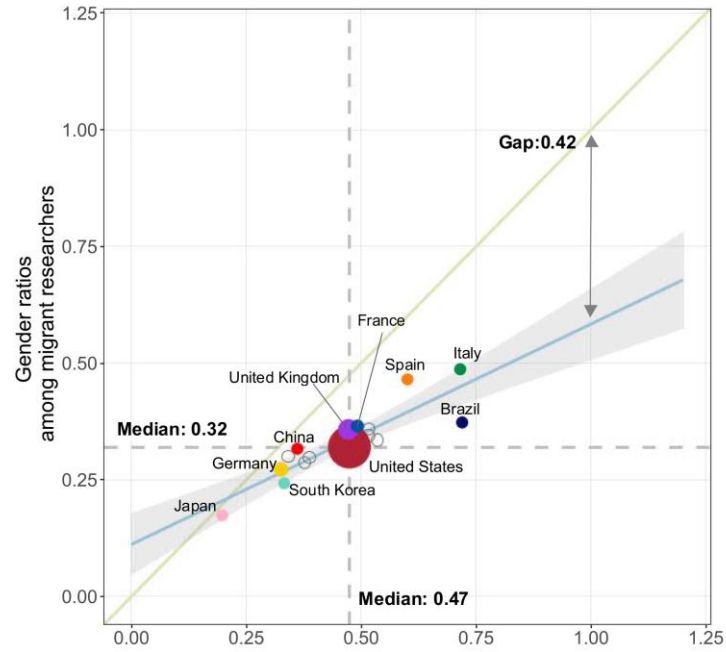
# Main result



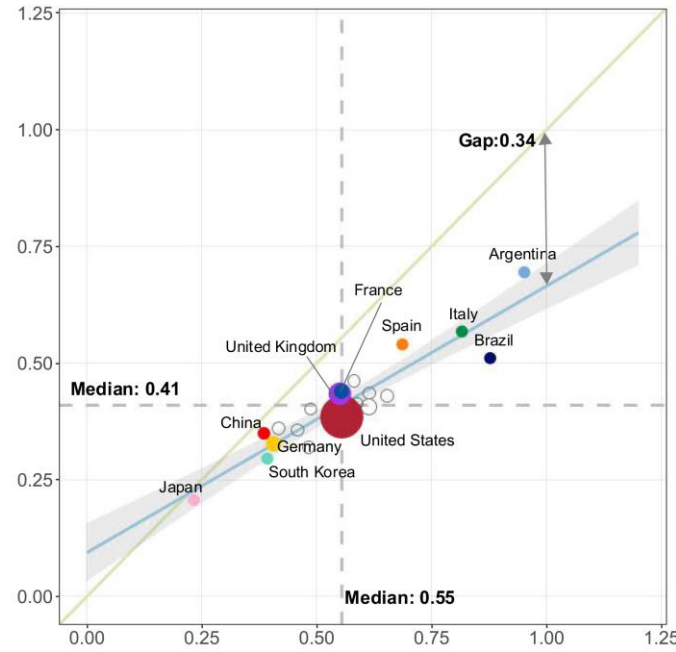
Composition of academic age and gender for non-movers, outward researchers, and returnees. Magnify all figures on the screen for higher resolution and more details



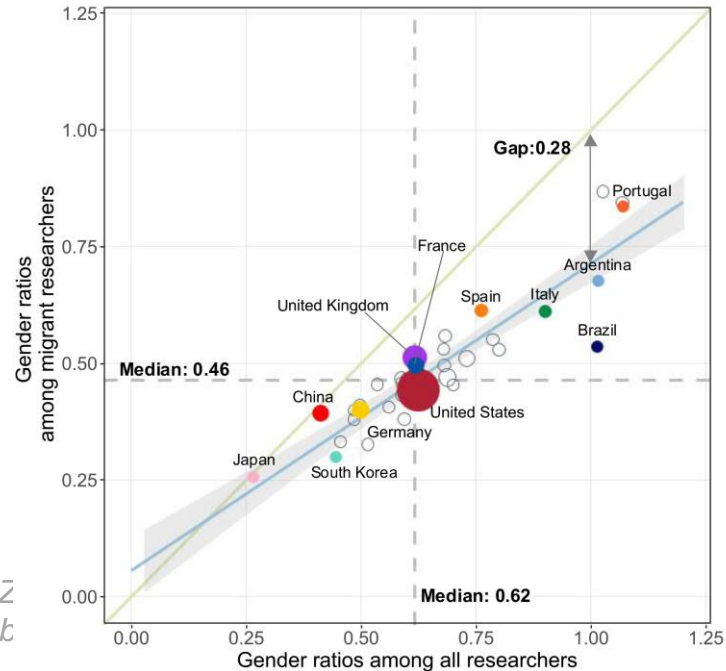
Period 1: 1998-2002



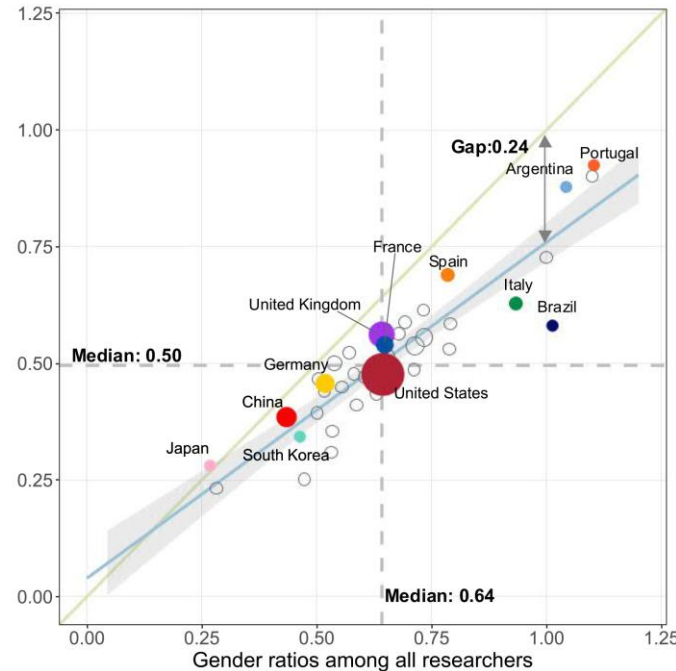
Period 2: 2003-2007



Period 3: 2008-2012



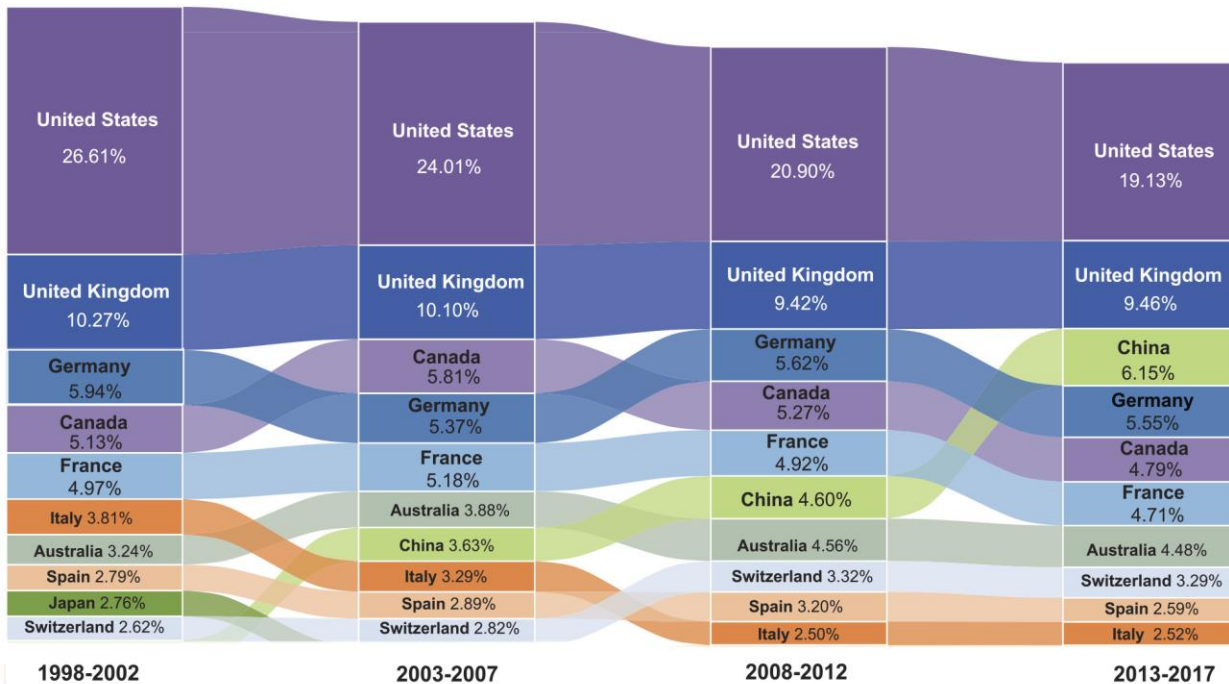
Period 4: 2013-2017



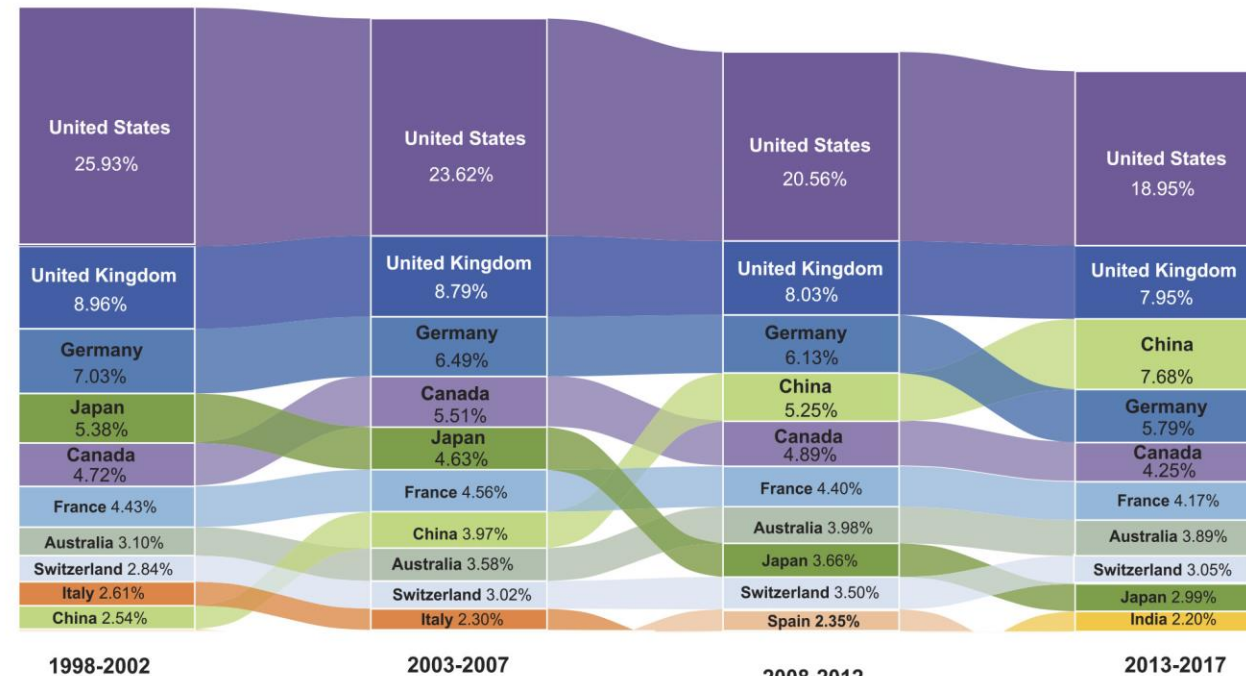
Gender ratios among all published researchers (X-axis) and migrant researchers (Y-axis). In the subfigure for each period, only the countries with over 500 female mobile researchers are shown, as including the countries with small populations of mobile researchers may give rise to bias in the ratio measurements. The size of each country's circle is proportional to the number of female researchers who migrated from and to this country. Notably, to increase readability, the countries with no more than 2,000 female migrant researchers are set to the minimum size. The vertical and horizontal dashed lines indicate the median gender ratios of all published researchers and of mobile researchers in each period. The 45° line in each subfigure is used to help compare the gender ratios of these two categories, with another double-headed arrow underlining the distance between it and the fitted regression line at the X value of one. This helps us to track the convergence tendency of female representation in the group of mobile researchers versus that in the total researcher population and how it changed over the four time periods.



# Main result

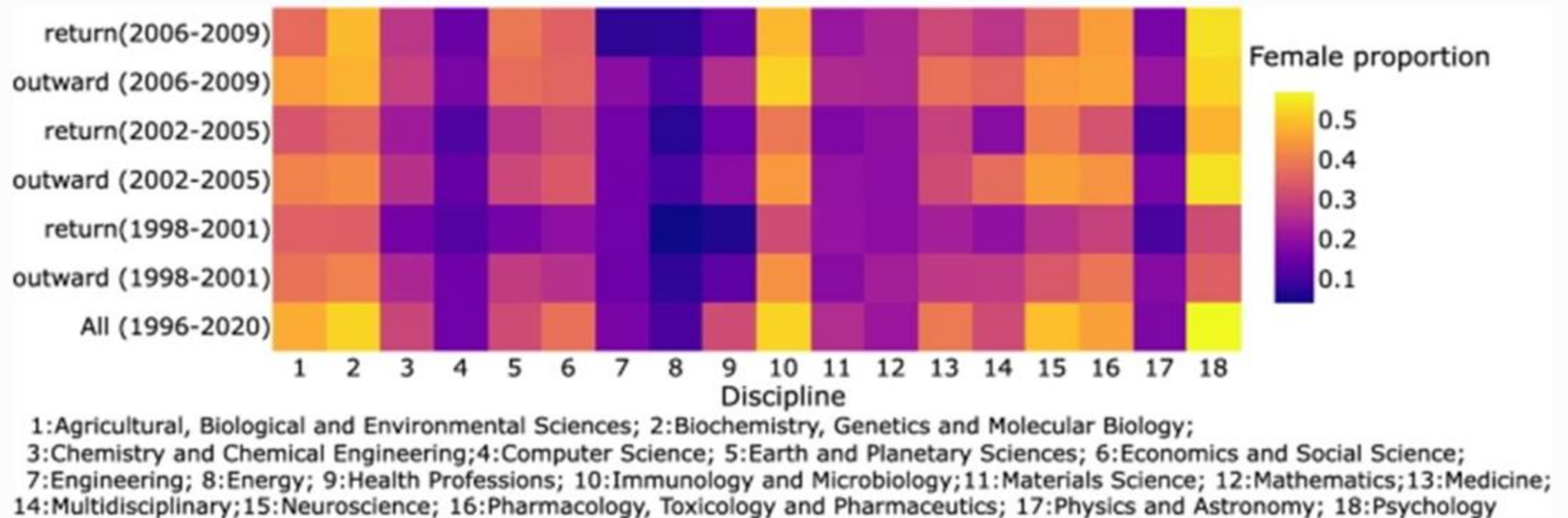


Female migrant researchers



Male migrant researchers

## Main result



Proportion of female researchers in different groups by discipline and cohort. The colored version of the figure is available online in high resolution

## Conclusion

- We observed a majority of disciplines are dominated by male researchers in Germany, **however**, international mobility may play in helping to moderate some of the most extreme gender disparities.
- Male researchers always have longer academic life than females, **however**, international migration, especially returning after the experience of international migration will help prolong the academic life of female researchers in Germany.
- Both the outward and the returnee subpopulations in most disciplines were more male-dominated, female migrant researchers are less likely to return to Germany relative to their male counterparts.



## Open discussion, Q&A

Please raise any comments, [clarification or else] questions, or points on the slides and content that were presented!

## IN THE PIPELINE...

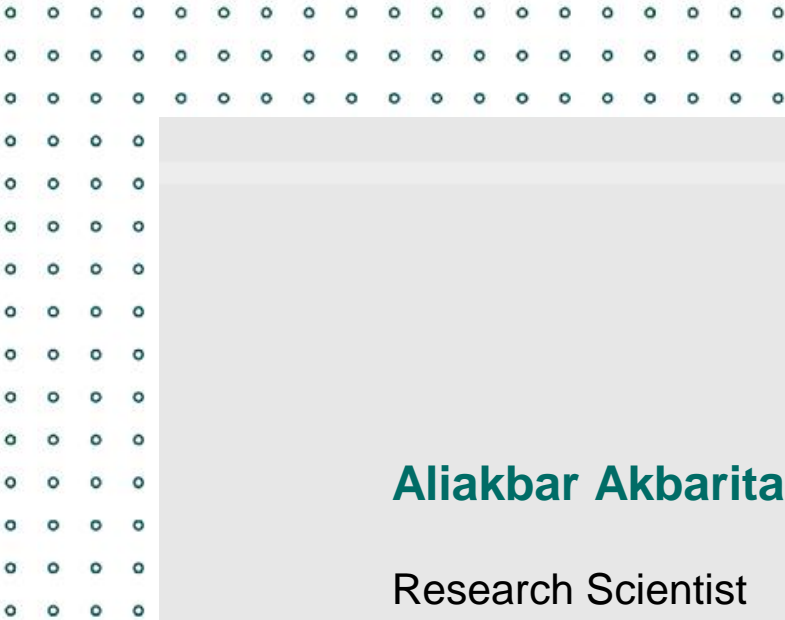
- We will announce SMD's beta website using participants' emails
  - So please share your email address with us (write to: [scholarlymigration@demogr.mpg.de](mailto:scholarlymigration@demogr.mpg.de) and include "[sign-up for SMDatabase]" in subject line), if we do not have it yet!



**Scholarly Migration Database**

Quantifying the Mobility of Scholars

Please tweet with hashtag  
#BiblioDemography,  
a tribute to James W. Vaupel (1945-2022).  
Thanks Ilya and Jonas for bringing up Jim's  
labeling idea!  
and #EPC2022



**Aliakbar Akbaritabar**

Research Scientist

[Akbaritabar@demogr.mpg.de](mailto:Akbaritabar@demogr.mpg.de)

<https://akbaritabar.github.io/>

<https://twitter.com/Akbaritabar>

**Xinyi Zhao**

PhD candidate

[Zhao@demogr.mpg.de](mailto:Zhao@demogr.mpg.de)

[https://www.demogr.mpg.de/en/about\\_us\\_6113/  
staff\\_directory\\_1899/xinyi\\_zhao\\_4083/](https://www.demogr.mpg.de/en/about_us_6113/staff_directory_1899/xinyi_zhao_4083/)

<https://twitter.com/XinyiZhao16>





THANK YOU!

