MAX-PLANCK-INSTITUT
FÜR DEMOGRAFISCHE
FORSCHUNG

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC
RESEARCH

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG
MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

Digital Trace Data for Demographic Research

Emilio Zagheni

HMD Symposium – Berlin, May 13th, 2019

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG
MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

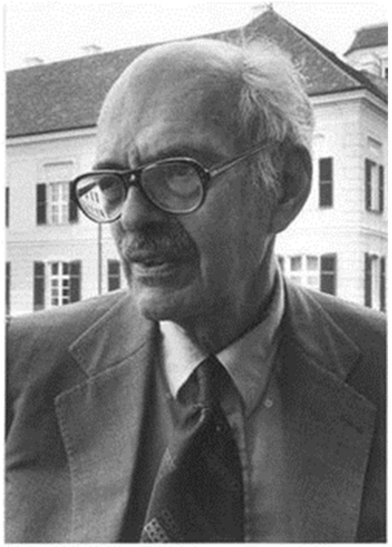
Inter-generational conversations on HMD developments

Emilio Zagheni

HMD Symposium – Berlin, May 13th, 2019

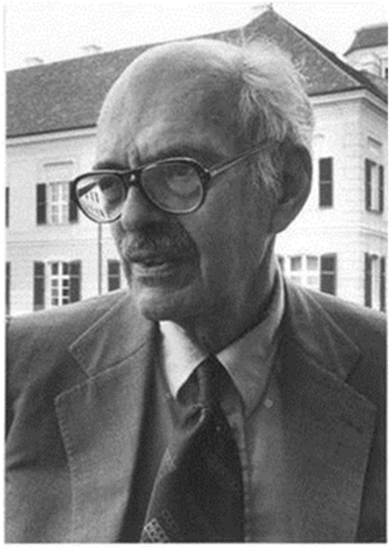
The characters for these conversations

The characters for these conversations

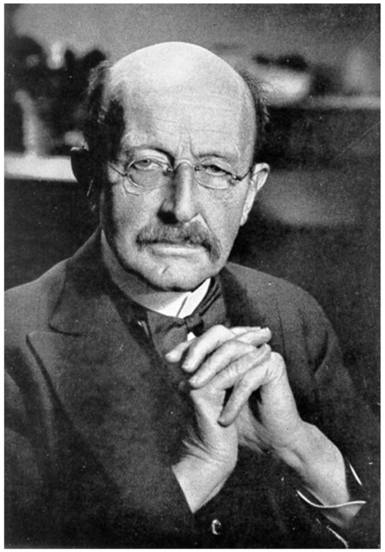


Nathan Keyfitz

The characters for these conversations

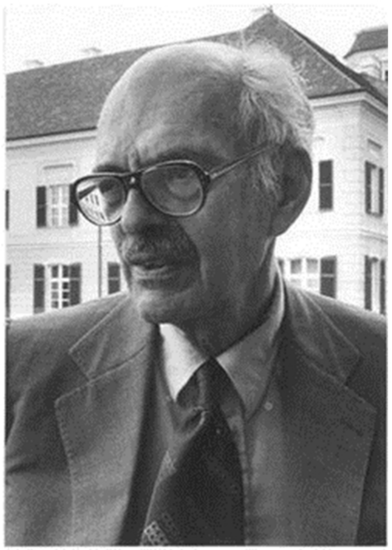


Nathan Keyfitz

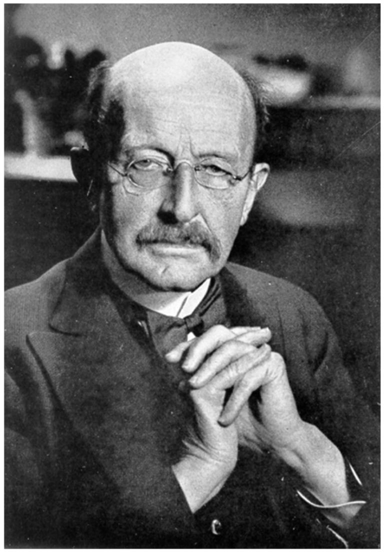


Max Planck

The characters for these conversations



Nathan Keyfitz



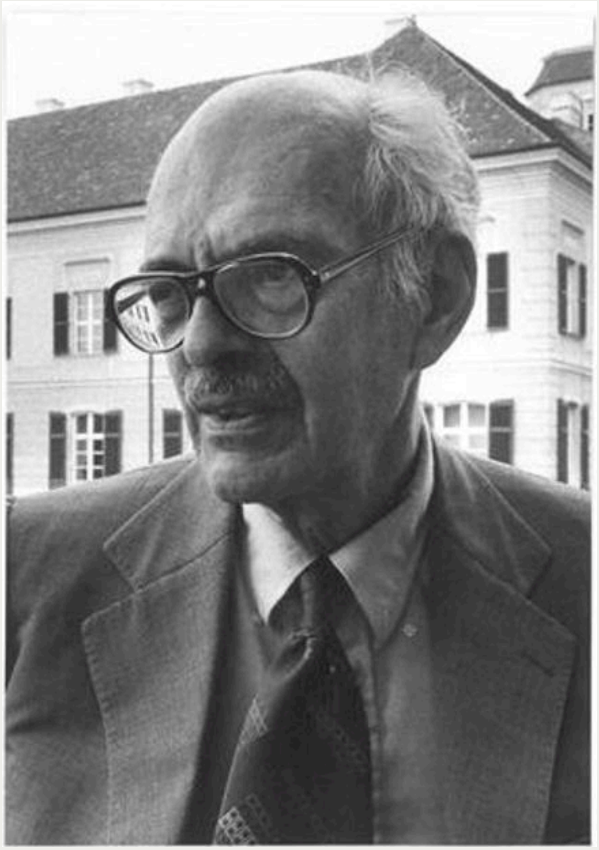
Max Planck

The “geo-location pin”

Nathan Keyfitz (1913-2010) - Mathematics and Data

“There were plenty of formulas in my assembly, but abstract formulas is what they could well have remained. [..] What saved my hard work from this fate was the advent of the computer”

Nathan Keyfitz’s Memoir



THEORETICAL POPULATION BIOLOGY 5, 1–27 (1974)

Family Formation and the Frequency of Various Kinship Relationships

LEO A. GOODMAN

The University of Chicago

NATHAN KEYFITZ AND THOMAS W. PULLUM

Harvard University

Received January 19, 1970

A set of age-specific rates of birth and death implies expected numbers of kin. An individual girl or woman chosen at random out of a population whose birth and death rates are specified can be expected to have a certain number of older sisters, younger sisters, nieces, cousins; expressions for these values are provided for both total kin and kin who are still living. Included also are the probabilities of living mother, grandmother, and great grandmother for girls and women of various ages. The methods are applicable to the size of the nuclear and the extended family. All formulas have been programmed and specimen numerical values are given.

A factor of l_{a-y} under the double integral will provide for younger sisters living to time t:

$$\int_{\alpha}^{\beta} \left[\int_{0}^{a} (l_{x+y}/l_x)\, m_{x+y} l_{a-y}\, dy \right] W(x)\, dx. \tag{3.2.b}$$

Note that the denominator l_x can cancel with the l_x in $W(x)$ to give an alternative form for (3.2.a) and (3.2.b).

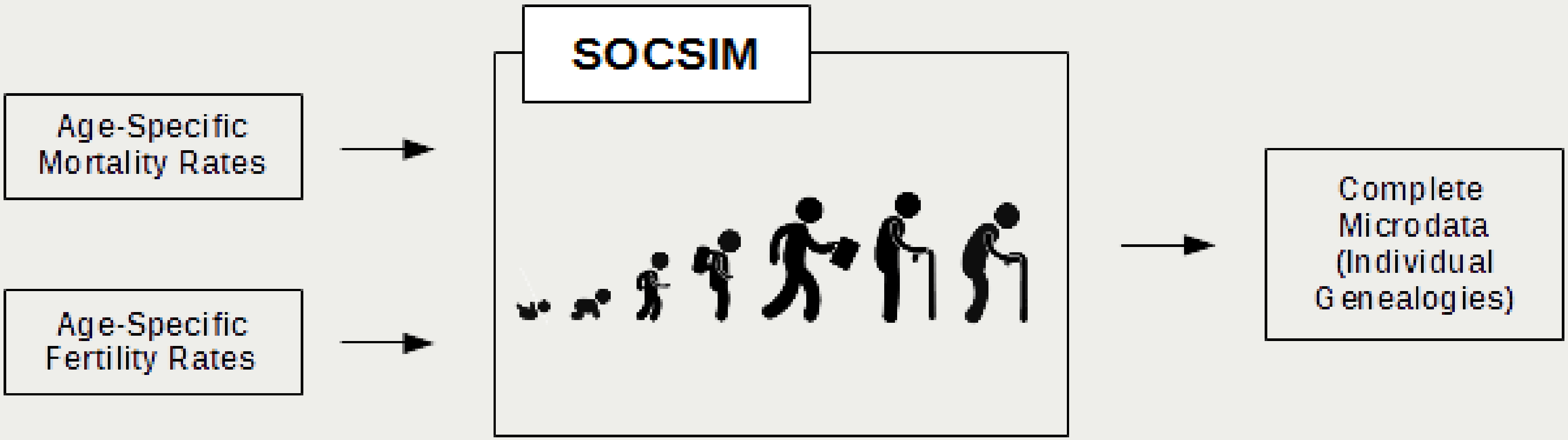
We note that in the usual model of population growth, the birth rate m_x for a female of age x can depend upon the age of the female (i.e., the rate is age specific), but it does not depend upon any other characteristic of the female, and,

TABLE IV

Expected Number of Sisters ever Born and Still Alive, by Age of Woman, for Three Selected Countries

	Sisters ever born			Sisters still alive		
Age	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
0	0.6103	1.3250	1.3804	0.5952	1.2281	0.9643
5	0.8860	1.9299	1.9484	0.8633	1.7852	1.3342
10	1.0837	2.4418	2.4147	1.0541	2.2518	1.6067
15	1.1902	2.7934	2.7351	1.1548	2.5648	1.7506
20	1.2351	2.9893	2.9172	1.1939	2.7275	1.7759
25	1.2485	3.0697	2.9999	1.2003	2.7761	1.7170
30	1.2508	3.0917	3.0296	1.1931	2.7617	1.6140
35	1.2509	3.0957	3.0378	1.1794	2.7186	1.4941
40	1.2509	3.0961	3.0393	1.1594	2.6558	1.3696

Kinship Demography 2.0 ⇒ Simulation



Credit: Paul Chung

DEMOGRAPHIC FOUNDATIONS OF FAMILY CHANGE*

SUSAN COTTS WATKINS
University of Pennsylvania

JANE A. MENKEN
Princeton University

JOHN BONGAARTS
The Population Council, New York

A longer life means that current cohorts can spend more years as members of a family in the statuses of parent, child, or spouse and in the combination of these statuses that defines the conjugal family. How much has this potential been realized? This question is addressed for the United States through a simulation of demographic conditions in 1800, 1900, 1960, and 1980. Despite declining fertility and higher divorce rates, women in the 1960 and 1980 cohorts spent more years in marriage and as parents than did earlier generations. They also spent more years as children of aged parents. But much of the potential offered by longer life spans has not been achieved. Not only did the number of years in marriage and parental statuses decline between 1960 and 1980, but current cohorts spend a smaller proportion *of their adult lives in them. On the basis of these results, we propose some scenarios of the ways that potential increases in the amount of time that people spend in family statuses may provoke social change.*

INTRODUCTION

Between 1800 and 1980, improvements in U.S. mortality nearly doubled women's expectation of life, from approximately 40 to nearly 80 years. It is reasonable to suppose that this change has affected family roles. Every indi-

been accompanied by changes in fertility and marriage. In 1800 couples on average bore about 8 children, by 1980 about 1.8. In 1800 women married at about age 20 and men 24 and then lived together until one of them died. In 1980, they married at about ages 23 and 25, and

Confessions of a Microsimulator

Problems in Modeling the Demography of Kinship

Steven Ruggles
Department of History
University of Minnesota

I could not, without effort, constrain myself to the task of either recalling, or constructing into a regular narrative, the whole burthen of horrors which lies upon my brain.
—*Thomas DeQuincey,* Confessions of an Opium Eater

Ever since Peter Laslett and John Harrison (1963) discovered that multigenerational households were rare in preindustrial northwestern Europe, historians and demographers have been trying to estimate the effects of preindustrial demographic conditions on the potential for multigenerational households. Starting with back-of-the

pendent of one another. That is, the characteristics of one member of a group of kin are assumed to be entirely uncorrelated with the characteristics of other members of the kin group. I call this the Whopper Assumption.[1]

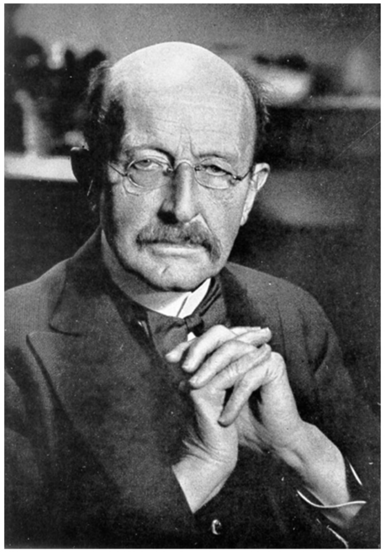
Because of the Whopper Assumption, models of kinship will produce less variation in the frequency of kin of any particular type than would occur in a real population. Our models will generally underestimate both the proportion of people with many kin and the proportion of people with few kin. As I will show, the magnitude of error is potentially large. In addition, the Whopper As-

Formal+Simulation+Crowd-sourced: Kinship Demography 3.0?

Credit: Indiegogo

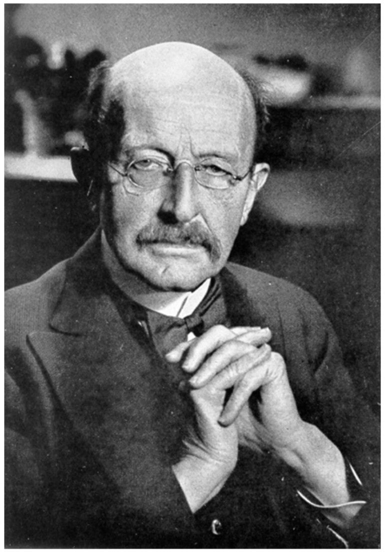
The lived experience of kin death

The lived experience of kin death



Max Planck (1858-1947)

The lived experience of kin death



Max Planck (1858-1947)

He survived his first wife Marie Merck (1861-1909) and all four children he had with her:

Karl (1888-1916)
Emma (1889-1919)
Grete (1889-1917)
Erwin (1893-1945)

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

Joint project on demographic foundations of lived experience of death
with Diego Alburez-Gutierrez and Martin Kolk

Why should we think about death of kin?

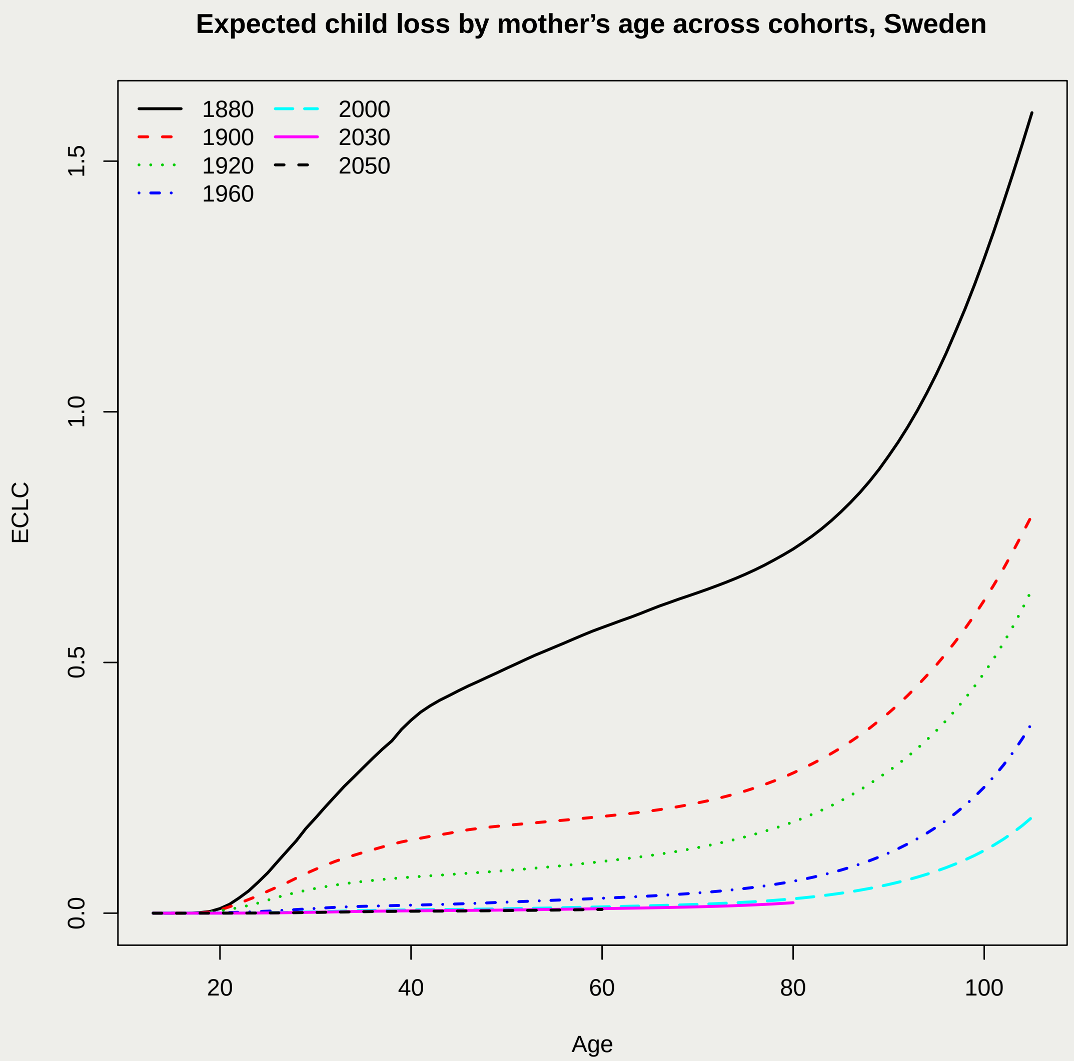
- The demographic transition theory emphasizes parents responding to reductions in child mortality by lowering their fertility. How can we quantify these reductions?
- How many children (or what fraction) can a woman expect to lose during her lifetime?
- Kin members are an important resource for individuals
- Losing a child is a traumatic life event that might affect individuals’ own survival and health

Expected Child Loss

Consider a woman born in 1900, who is still alive at age 60. How many children do we expect her to have lost in her lifetime? Her expected child loss conditional on reaching age 60 (ECLC) is:

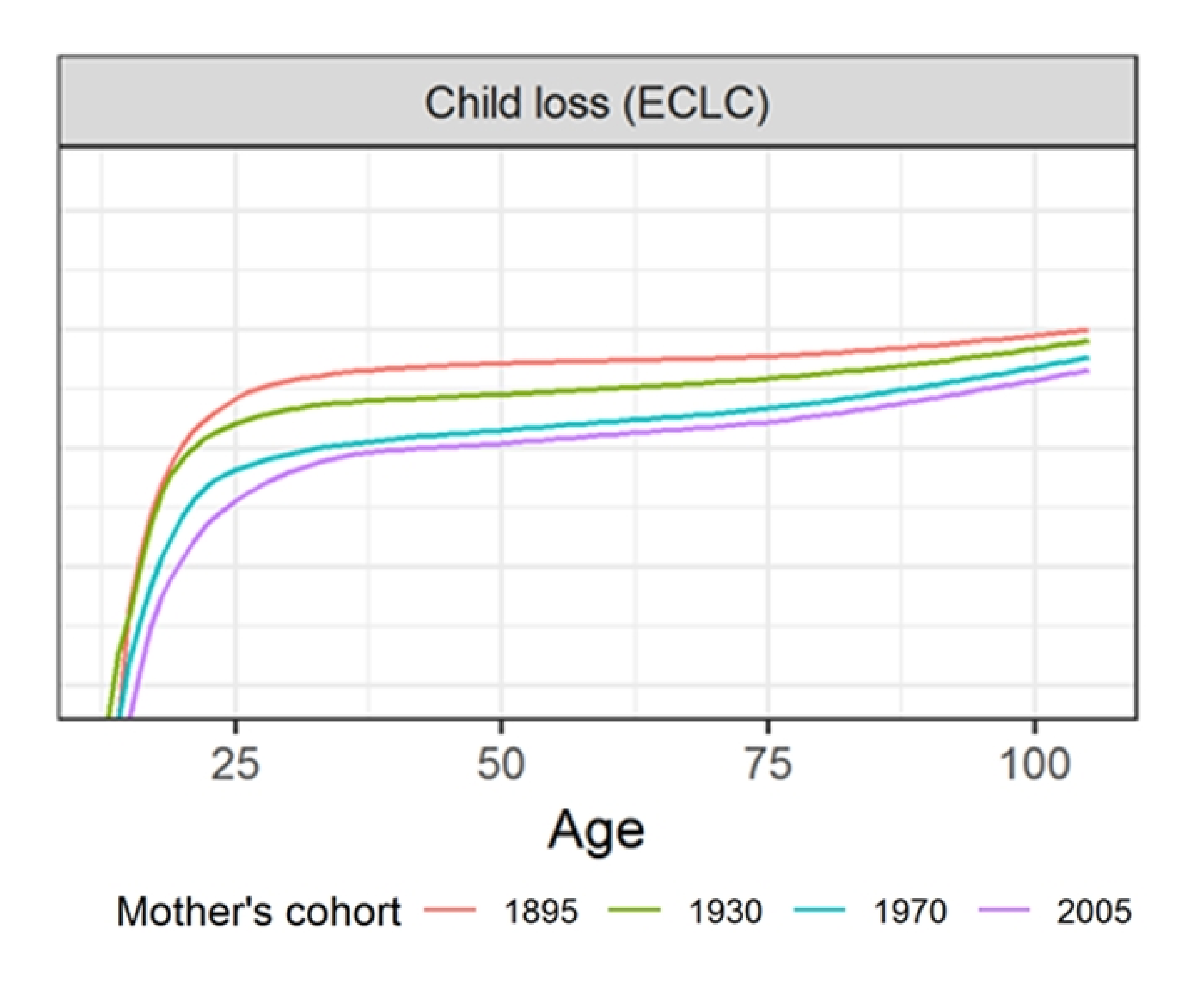
$$\underbrace{ECLC_{(60,1900)}}_{\text{children lost}} = \underbrace{TFR_{(60,1900)}}_{\text{total children}} - \underbrace{\sum_{x=15}^{x=50} {}_1F_{(x,1900)} l_{(60-x,1900+x)}}_{\text{surviving children}}$$

Expected Child Loss



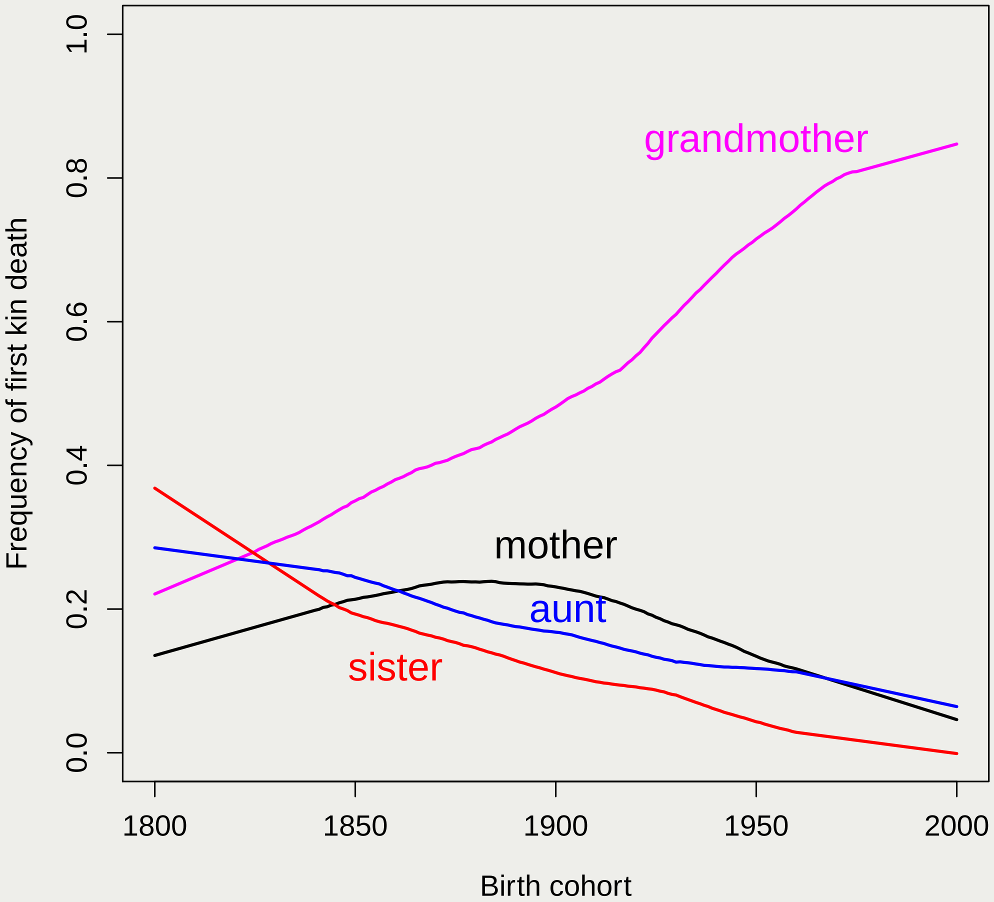
Data source: own calculations based on HMD and HFD

Strong linear relationship (on a log scale) after age 25



Data source: own calculations based on HMD and HFD for Sweden

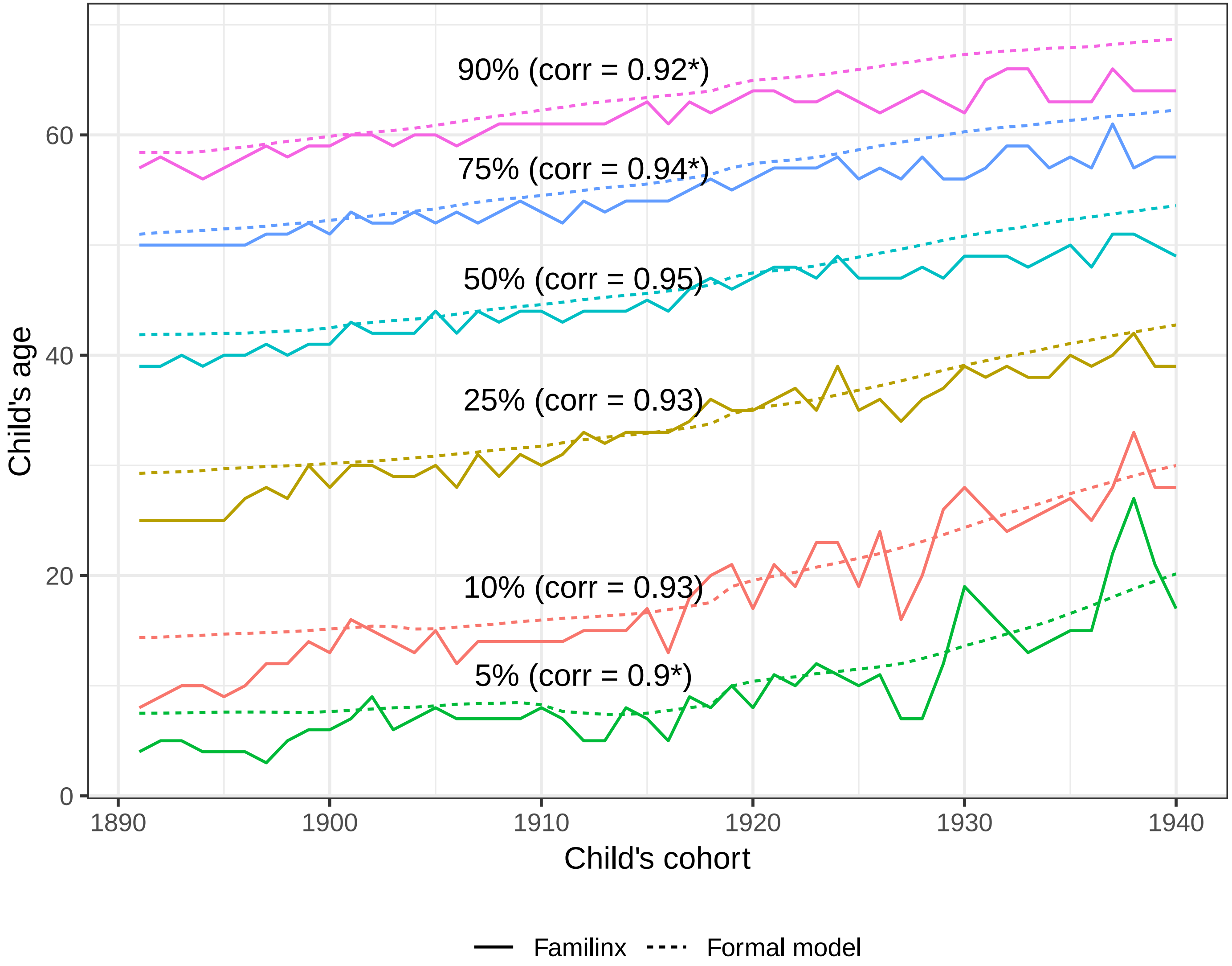
The first experience of death



SOCSIM microsimulation outcomes calibrated against HMD and HFD for Sweden

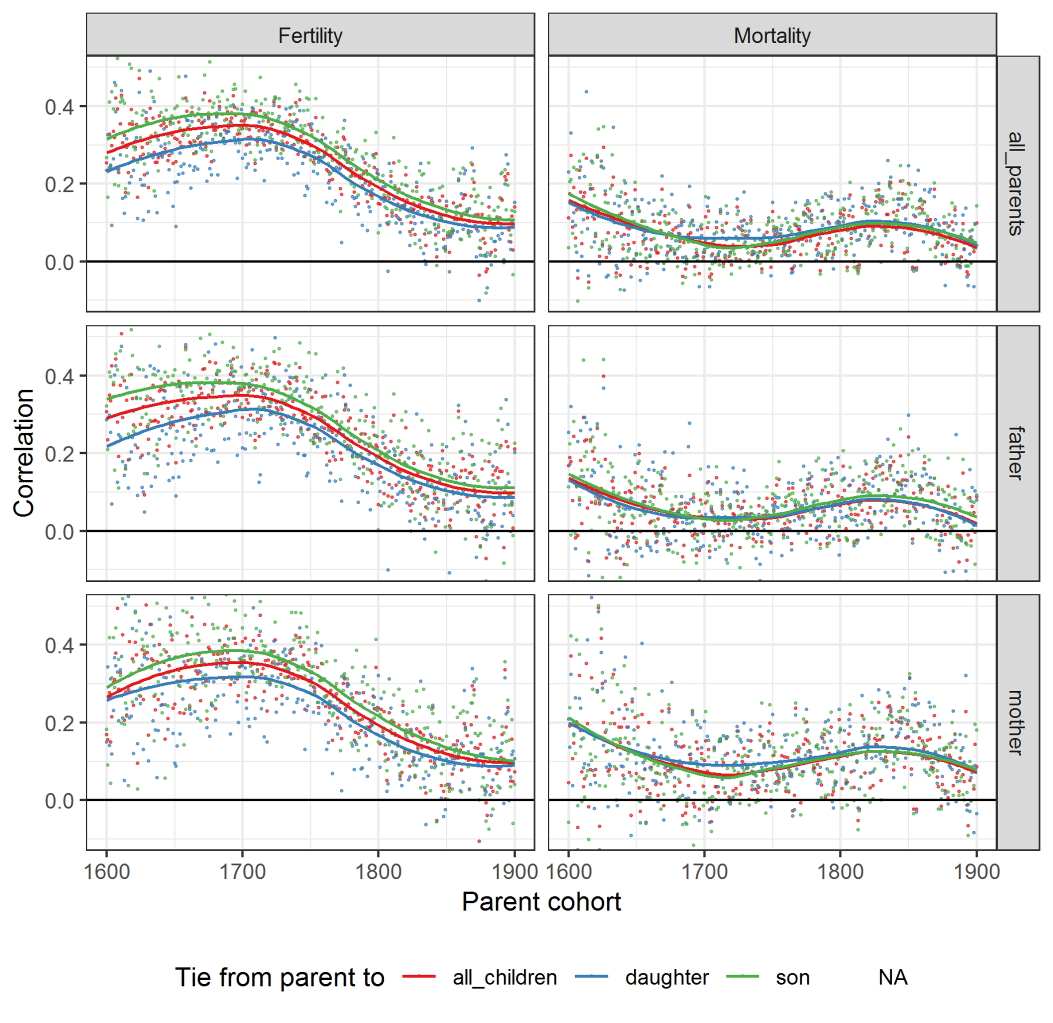
The validity of genealogical data

The validity of genealogical data



Proportion of mothers who have died by child's age

Correlations across generations revealed by genealogical data



Some reflections

- HMD is helpful not only to understand mortality patterns, but also to evaluate the lived experience of death
- HMD alone is not enough
  - Mathematical models
  - Simulation models
  - Digital trace data (like online genealogies)
- HMD is more useful when combined with HFD (and perhaps with an HMigD?)

The importance of space

Two key directions to incorporate spatial analyses

1. More statistical approaches
2. More data

More Statistics → Small area estimation

Demography (2017) 54:2025–2041
DOI 10.1007/s13524-017-0618-7

CrossMark

A Flexible Bayesian Model for Estimating Subnational Mortality

Monica Alexander[1] **· Emilio Zagheni**[2] **·**
Magali Barbieri[1,3]

Published online: 10 October 2017

Abstract Reliable subnational mortality estimates are essential in the study of health inequalities within a country. One of the difficulties in producing such estimates is the presence of small populations among which the stochastic variation in death counts is relatively high, and thus the underlying mortality levels are unclear. We present a Bayesian hierarchical model to estimate mortality at the subnational level. The model builds on characteristic age patterns in mortality curves, which are constructed using principal components from a set of reference mortality curves. Information on mortality rates are pooled across geographic space and are smoothed over time. Testing of the model shows reasonable estimates and uncertainty levels when it is applied both to simulated data that mimic U.S. counties and to real data for French *départements*. The model estimates have

More data → Accounting for mobility and migration

Denominators and numerators in mortality rates are affected by:

- Large short-term flows of people
- Changing population composition due to migration
- Selective changes in immigration and return migration

In some situations we cannot rely on existing data

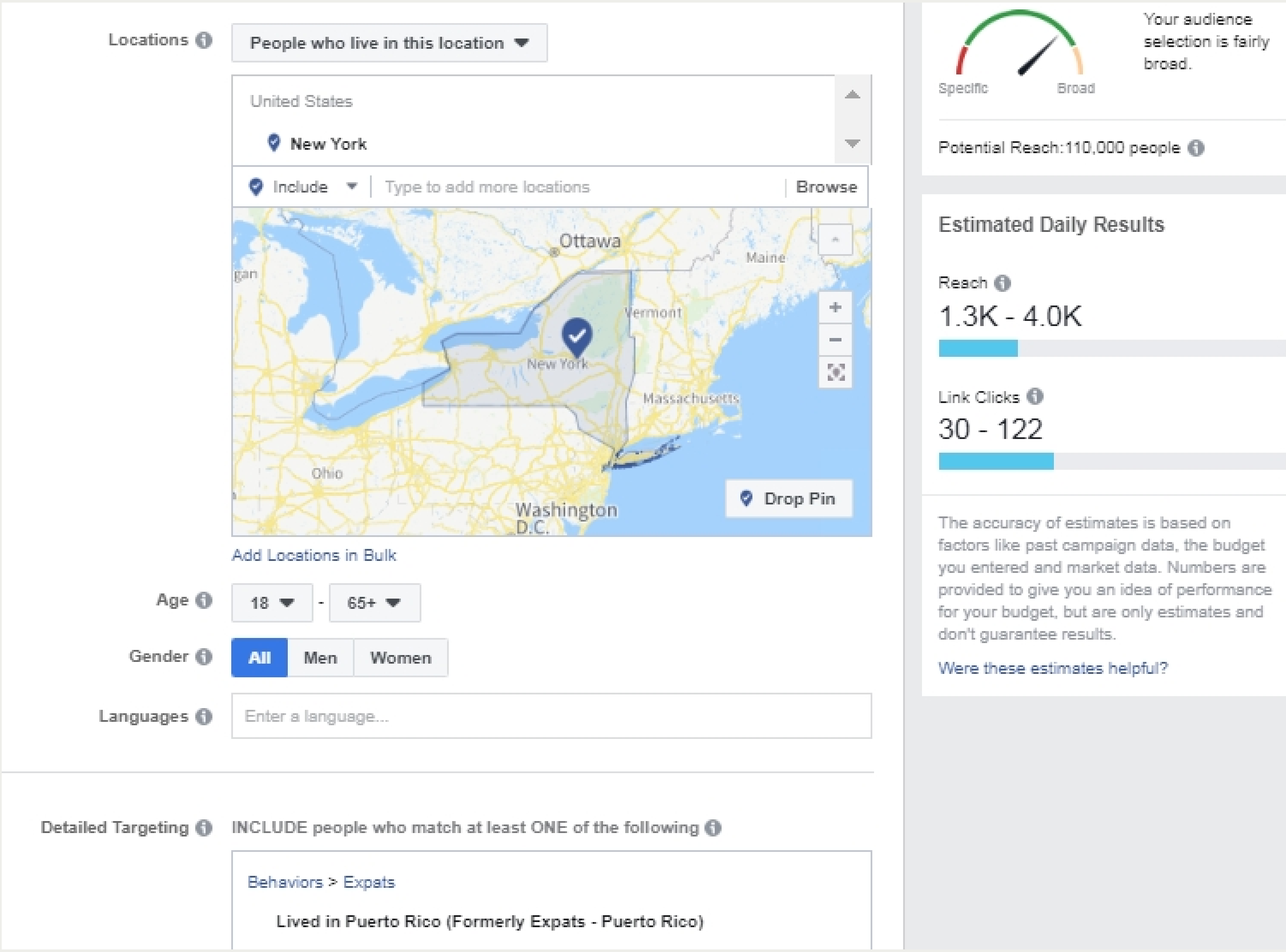
MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

Joint project on estimating migration out of Puerto Rico after Hurricane Maria

with Monica Alexander and Kivan Polimis

Facebook data for Advertisers



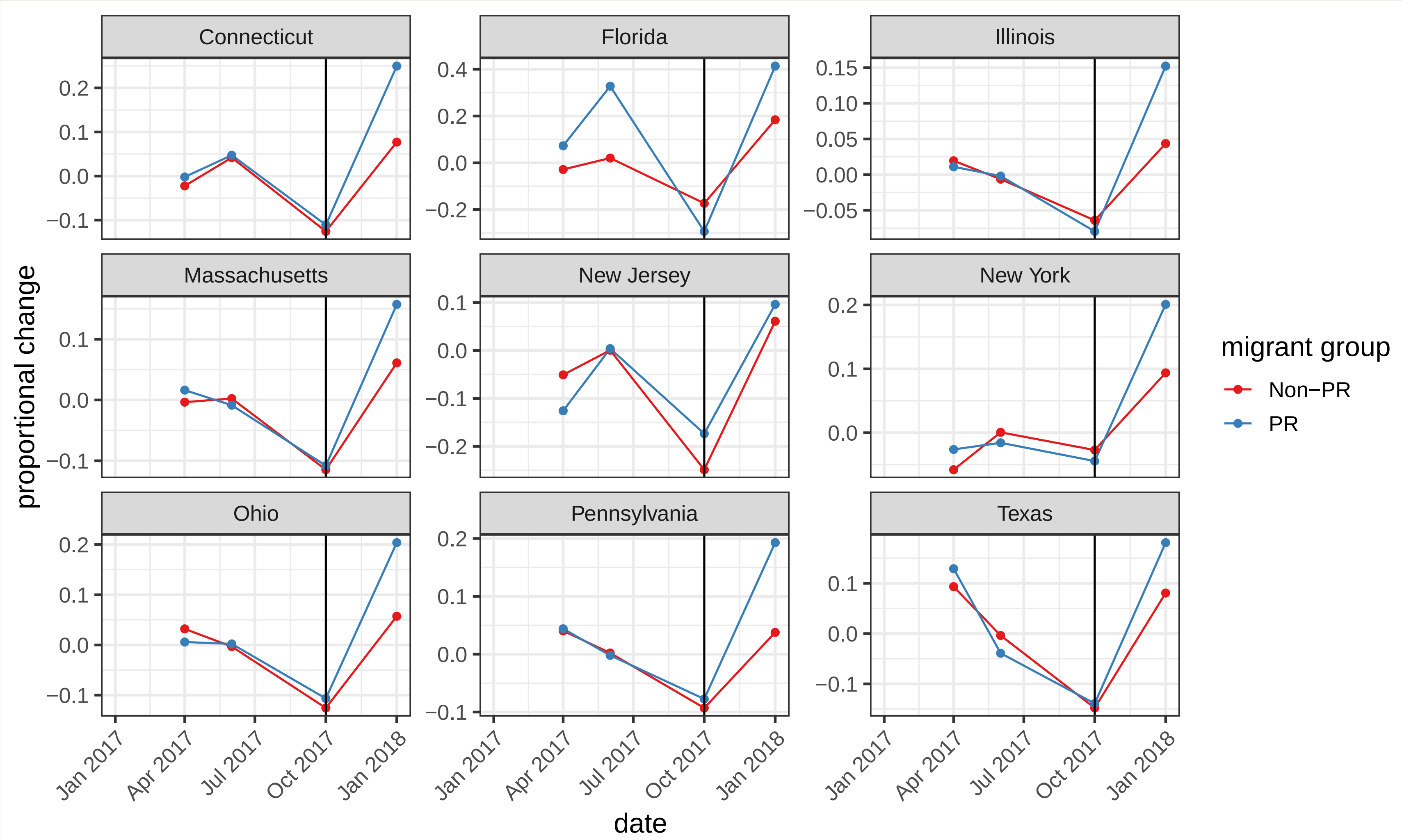
Approach: difference-in-differences in migration before and after Hurricane Maria

1. Look at the percentage change in Puerto Rican migrants vs international migrants to US states, before and after Hurricane Maria

2. Use difference-in-differences to estimate percentage change in Puerto Ricans in all US states, by age and sex

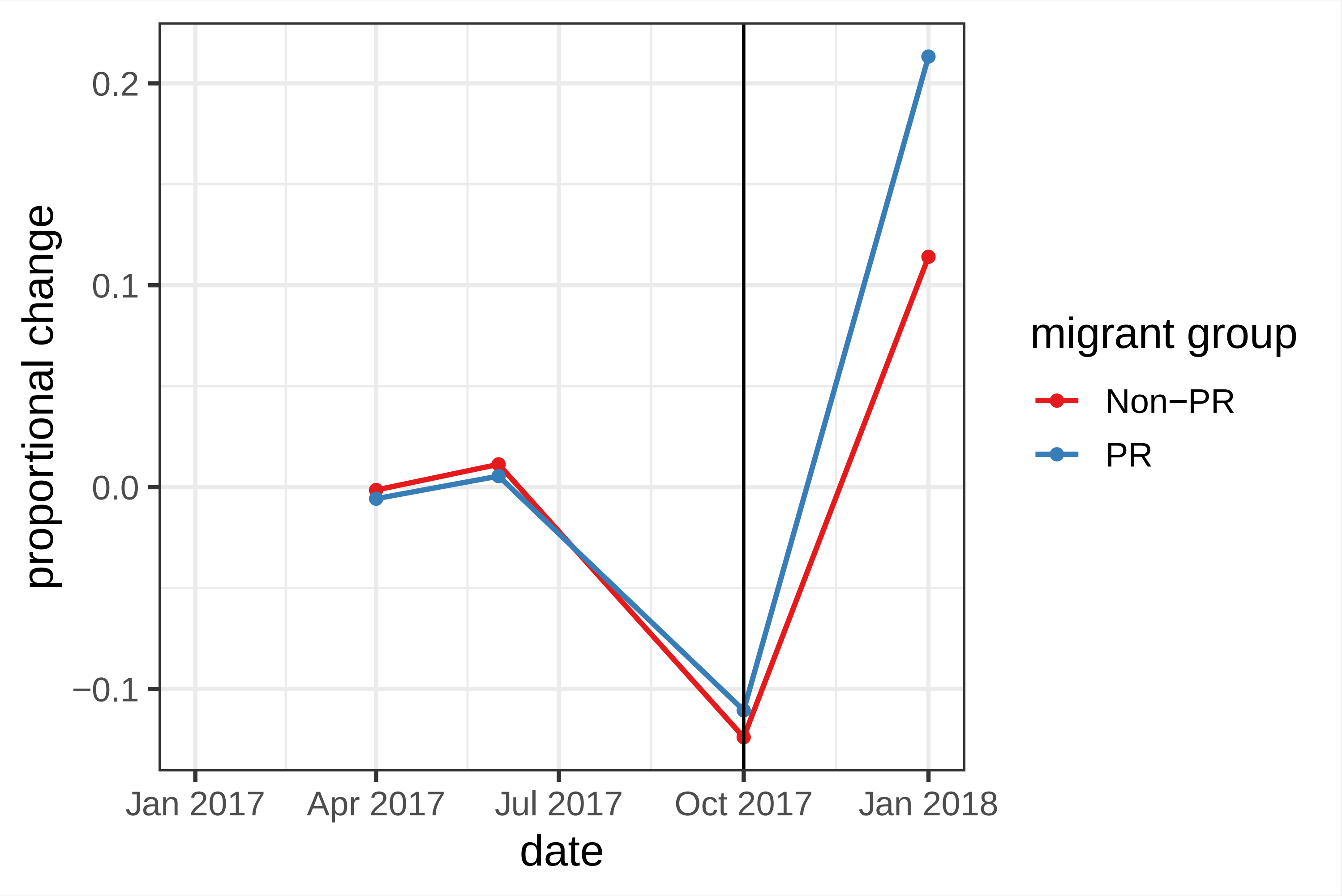
3. Transform percentage changes into population counts by multiplying for the baseline population of Puerto Ricans before the Hurricane (as reported by the ACS)

Percentage change in Puerto Ricans vs non-Puerto Ricans migrants



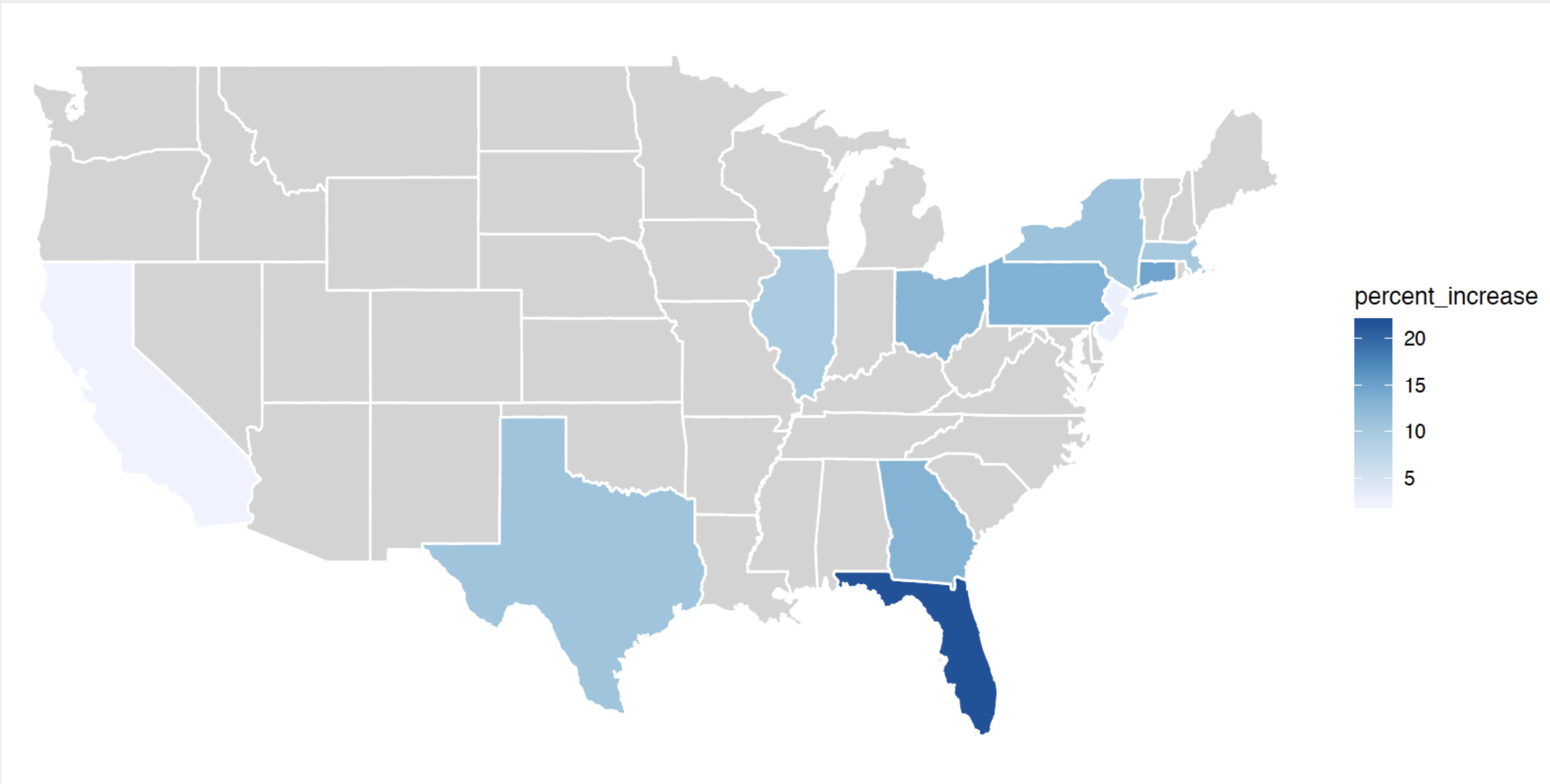
Source: Alexander, Polimis and Zagheni (forthcoming), *Population and Development Review*

All states except for Florida



Source: Alexander, Polimis and Zagheni (forthcoming), *Population and Development Review*

Percent increase in Puerto Ricans from Oct. 2017 to Jan 2018



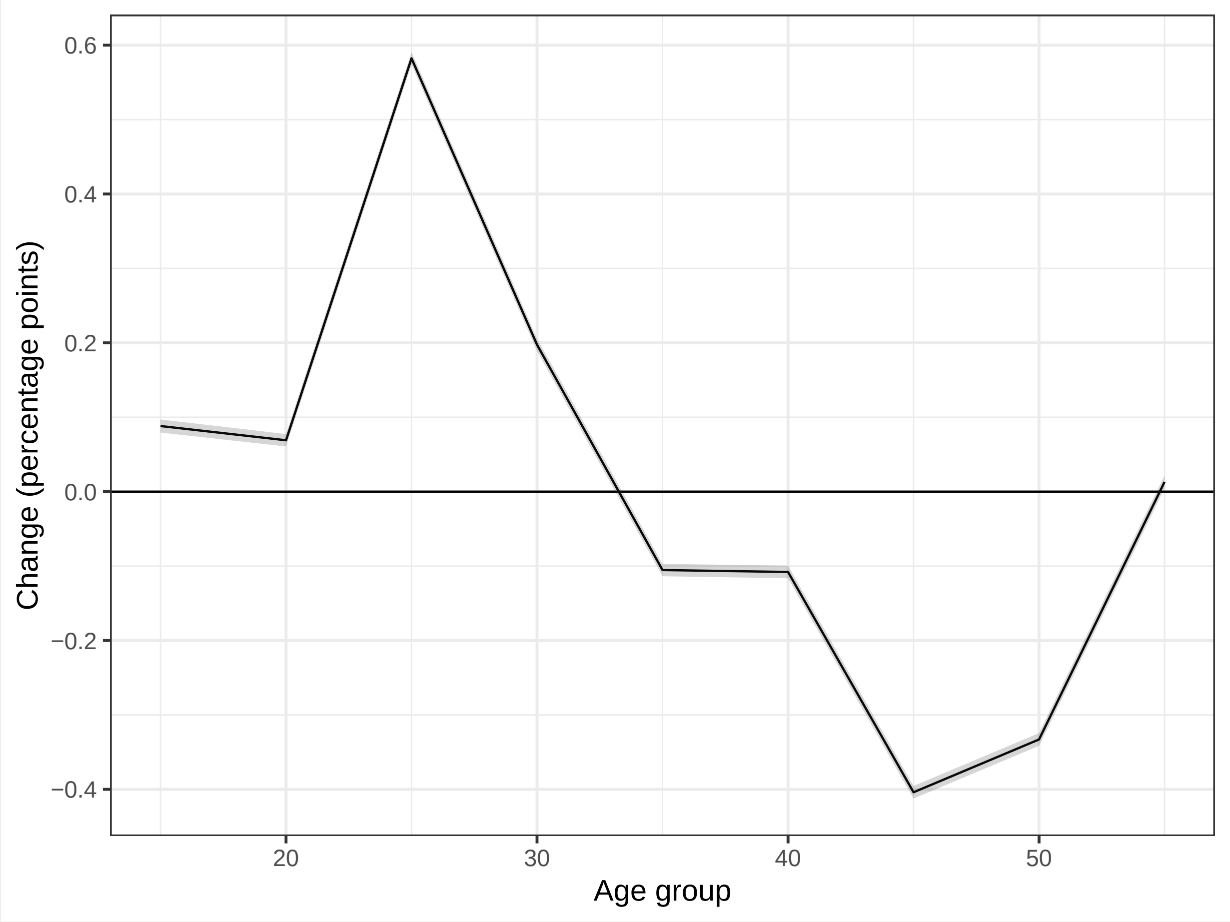
Source: Alexander, Polimis and Zagheni (forthcoming), *Population and Development Review*

Increase in stocks of Puerto Ricans from Oct. 2017 to Jan 2018

Table 1: Estimated increase in Puerto Rican migrant stocks from October 2017 to January 2018. The 95% confidence intervals are shown in parentheses.

State (95% CI)	% Increase (95% CI)	Population Increase
Florida	21.6 (20.9, 22.3)	61992 (60010, 63973)
New York	11 (10.3, 11.7)	14662 (13758, 15567)
Pennsylvania	13.4 (12.7, 14.1)	13151 (12427, 13875)
Connecticut	14.7 (12.9, 16.5)	9541 (8365, 10716)
Massachusetts	10.1 (8.82, 11.4)	8824 (7708, 9940)
Texas	10.8 (10.4, 11.2)	5394 (5179, 5609)
Ohio	12.8 (12.2, 13.4)	3002 (2865, 3139)
Illinois	9.9 (9.15, 10.6)	2684 (2482, 2887)
Georgia	13.1 (12.4, 13.8)	2600 (2464, 2736)
New Jersey	2.9 (1.56, 4.24)	2282 (1228, 3335)
California	2.4 (1.86, 2.94)	576 (446, 706)

Percent change in Puerto Rican migrants by age group

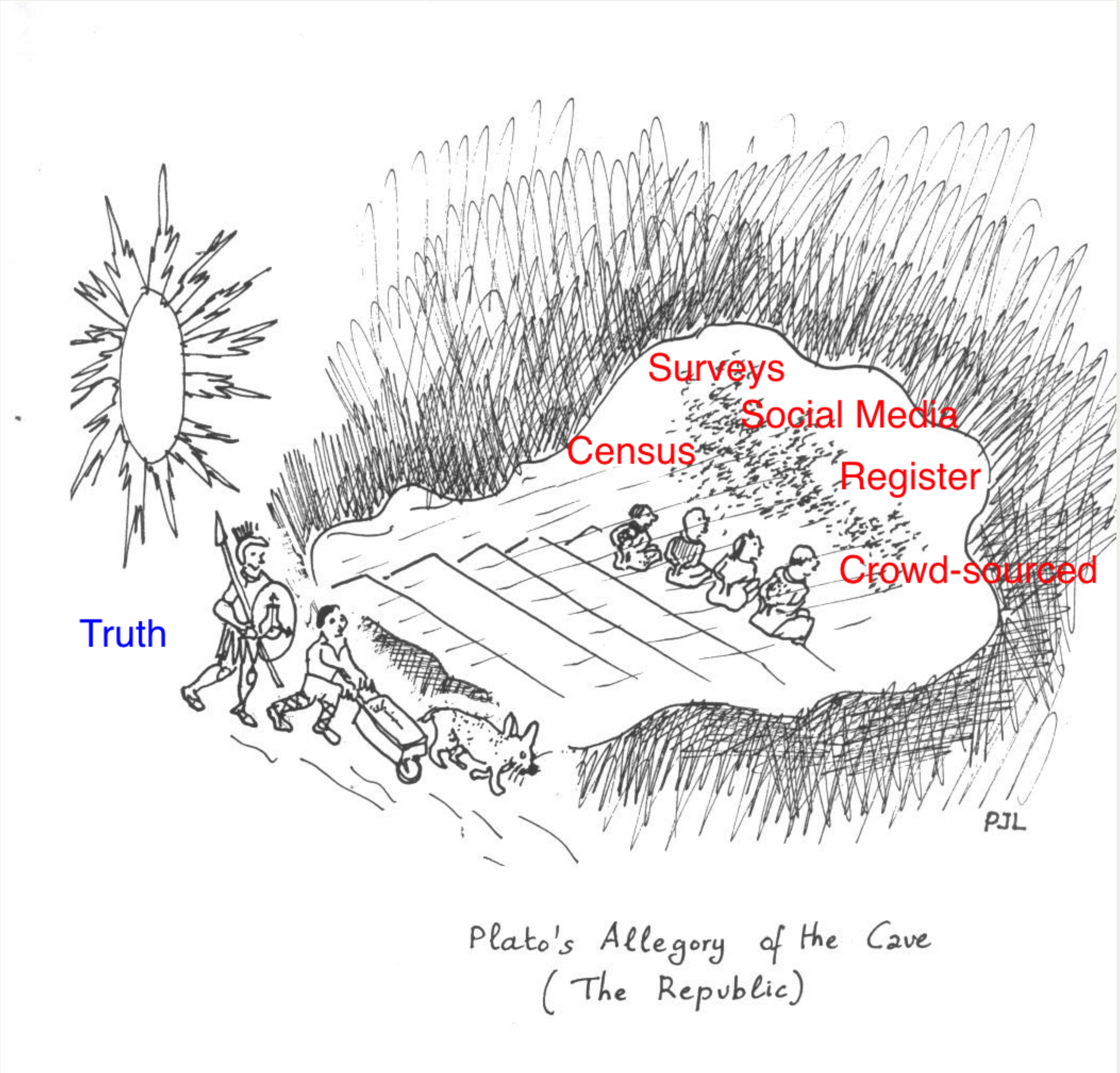


‘Return migration’ to Puerto Rico January-March 2018

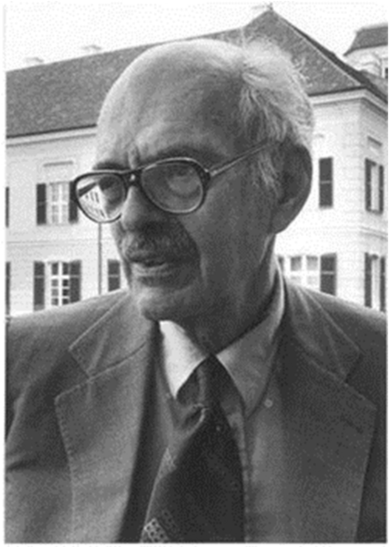
Table 3: Return migration: Estimated change in Puerto Rican migrant stocks from January 2018 to March 2018. The 95% confidence intervals are shown in parentheses.

State	% Change (95% CI)	Population Change
Florida	-7.1 (-7.77, -6.43)	-20377 (-22289, -18465)
Massachusetts	-4.5 (-5.52, -3.48)	-3931 (-4826, -3037)
Connecticut	-3.6 (-5.04, -2.16)	-2336 (-3273, -1400)
Texas	-3.5 (-3.91, -3.09)	-1748 (-1952, -1544)
Pennsylvania	-1.4 (-2.03, -0.773)	-1374 (-1989, -759)
Ohio	-1.7 (-2.14, -1.26)	-399 (-503, -295)
New York	0.4 (-0.241, 1.04)	533 (-322, 1388)
Illinois	2.8 (2.15, 3.45)	759 (583, 935)
Georgia	6.4 (5.86, 6.94)	1270 (1163, 1377)
New Jersey	2.3 (1.15, 3.45)	1810 (908, 2711)
California	8.5 (7.97, 9.03)	2039 (1911, 2167)

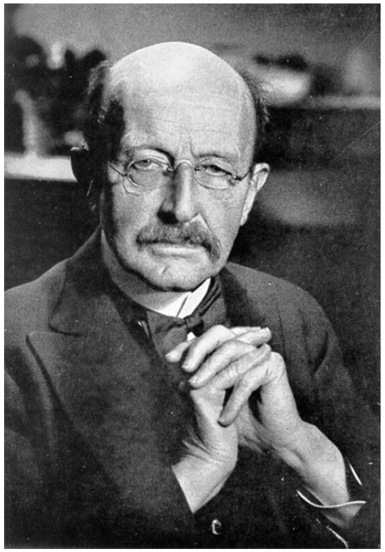
Plato’s allegory of the cave



The characters for this talk



Nathan Keyfitz



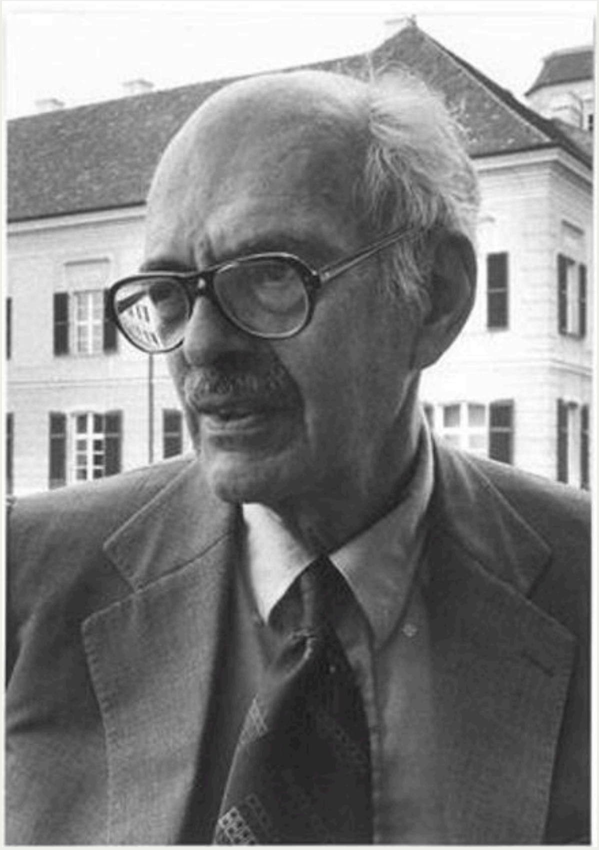
Max Planck

The “geo-location pin”

“Abstract formulas” are what can save social media research...

“There were plenty of formulas in my assembly, but abstract formulas is what they could well have remained. [..] What saved my hard work from this fate was the advent of the computer”

Nathan Keyfitz’s Memoir



MAX-PLANCK-INSTITUT
FÜR DEMOGRAFISCHE
FORSCHUNG

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC
RESEARCH

Thank you for your attention!

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

Comments or Questions?

www.demogr.mpg.de