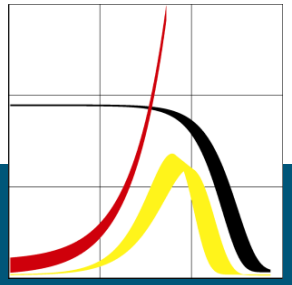
Max Planck Institute for Demographic Research



New data

Frans Willekens

Conference
Population Change and Life Course
Taking Stock and Looking to the Future
Ottawa, 19-20 March 2015

New data

- Individual level data
- Personal attributes: age, sex, martial status, lifestyle
- Activities: occupation, shopping, recreation, travel, commuting
- Expenditures: goods and services (incl. financial, health care)
- Behaviour
- Values, attitudes and opinions: voting behaviour
- Context: address, living conditions
- Social interaction and social network
- Sensors (e.g. health), surveillance cameras
- Data, byproducts of administrative processes, communication and transactions.

New data

- Longitudinal data
- Increased coverage for comparative analysis
  - e.g. IPUMS: 238 censuses of 74 countries. 545million observations (expected: 800 million from 300 censuses in 100 countries in 2018)
  - World Values Study
  - Gallup World Poll
  - Demographic and Health Surveys (DHS): 300 surveys in over 90 countries
  - Health and Retirement Studies: HRS, ELSA (UK), SHARE (Europe), Korea, China, Indonesia, Japan, Mexico, India
  - European Social Survey; US General Social Survey; International Social Survey Programme
- Spatiotemporally referenced data (GPS)

Data collection: research infrastructure

- Data archives
- Data centers
- Data linkage / record linkage
  - Data linkage centres: http://www.ipdln.org/data-linkage-centres

- ESFRI (European Strategy Forum on Research Infrastructures) (established in 2002)

  http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
- ESFRI's delegates are nominated by the Research Ministers of the Member and Associate Countries
- Roadmap for pan-European research infrastructures (first: 2006). From selection to implementation.

Data collection: research infrastructure

- **INFRAIA-1-2014/2015: Integrating and opening existing national and regional research infrastructures of European interest**

  ESFRI: Distributed, multidisciplinary European infrastructure on Big Data and social data mining. This activity should integrate large social data repositories, social data mining methods and tools, and supercomputing facilities for conducting large-scale analytical processing. The proposal should also address training in social data mining, to foster the availability of skilled data scientists

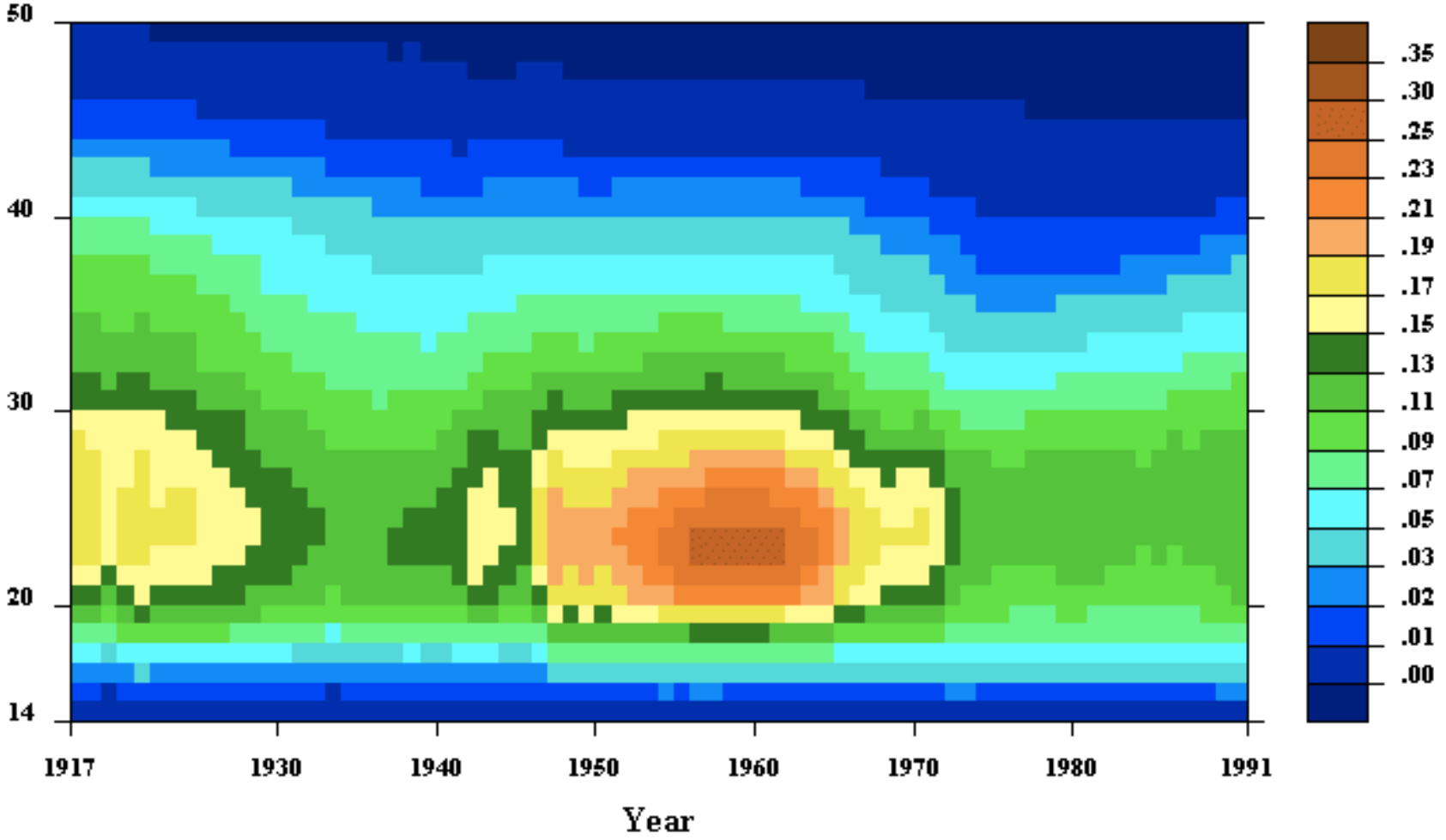
Data analysis: challenges ahead

How to get insight from large amounts of data?

- Pattern recognition
  - Visualisation
- Modeling and simulation

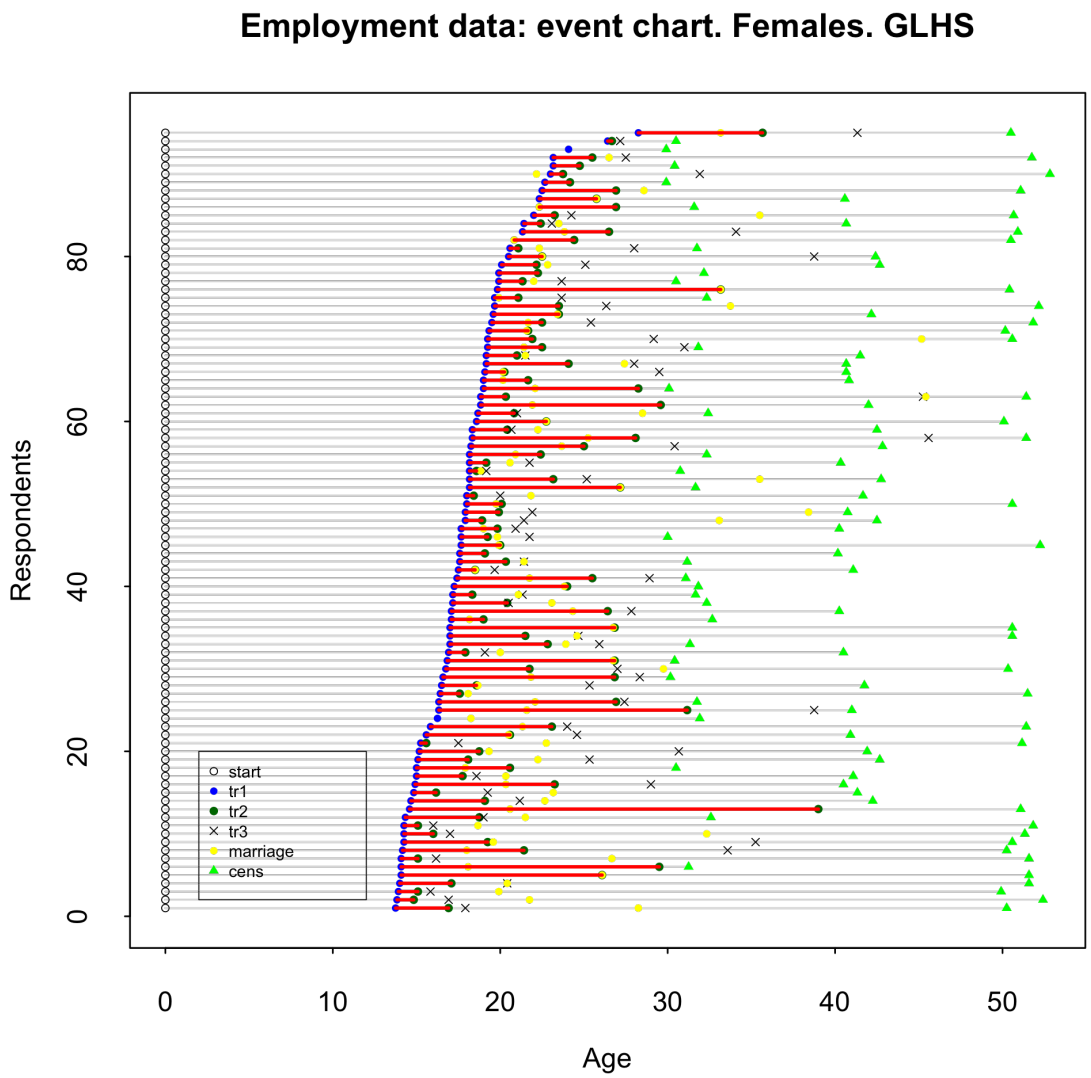
- Life histories (e.g. impact of genes and early life experiences)
- Social interaction

Figure 5. US birth rates selectively placed from 0.001 to 0.35, and from age 14 to 49 and year 1917 to 1990



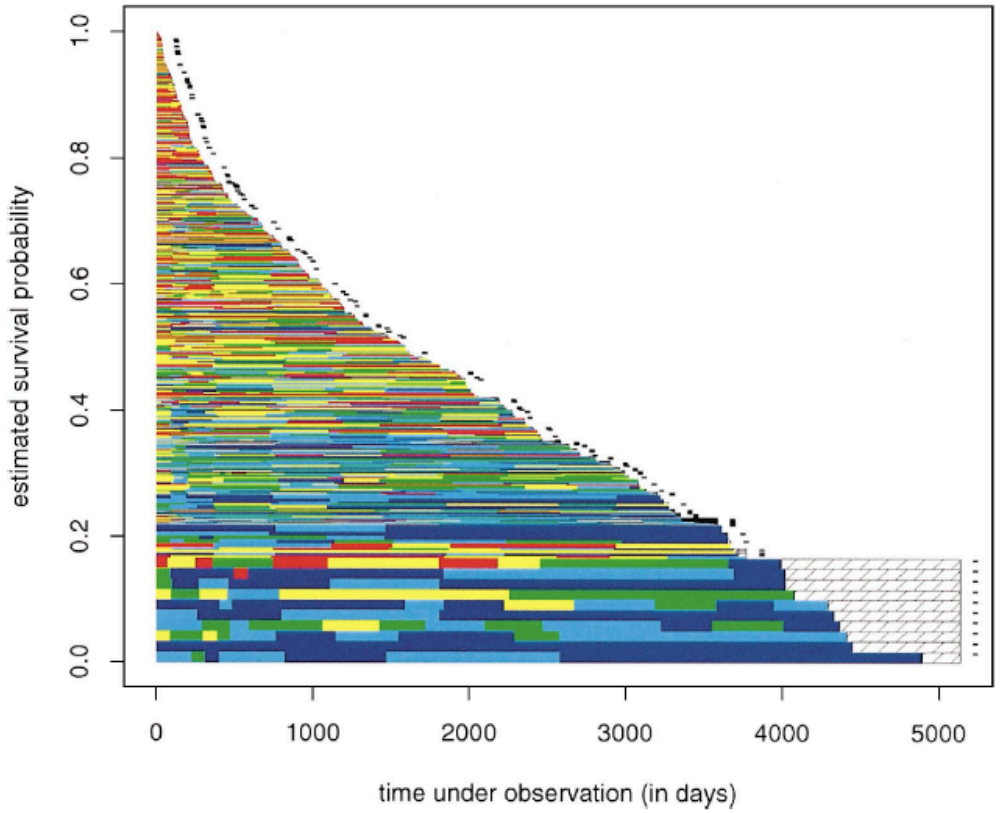
Source : VAUPEL J. W., WANG Z., ANDREEV K. F. and YASHIN A. I. (1997), *Population data at a glance. Odense Monographs on Population Aging No. 4*, Odense: Odense University Press (adresse web : http://www.demogr.mpg.de/Papers/Books/Monograph4/start.htm).

Event chart



Hmisc

Kaplan-Meier with supplementary life history information

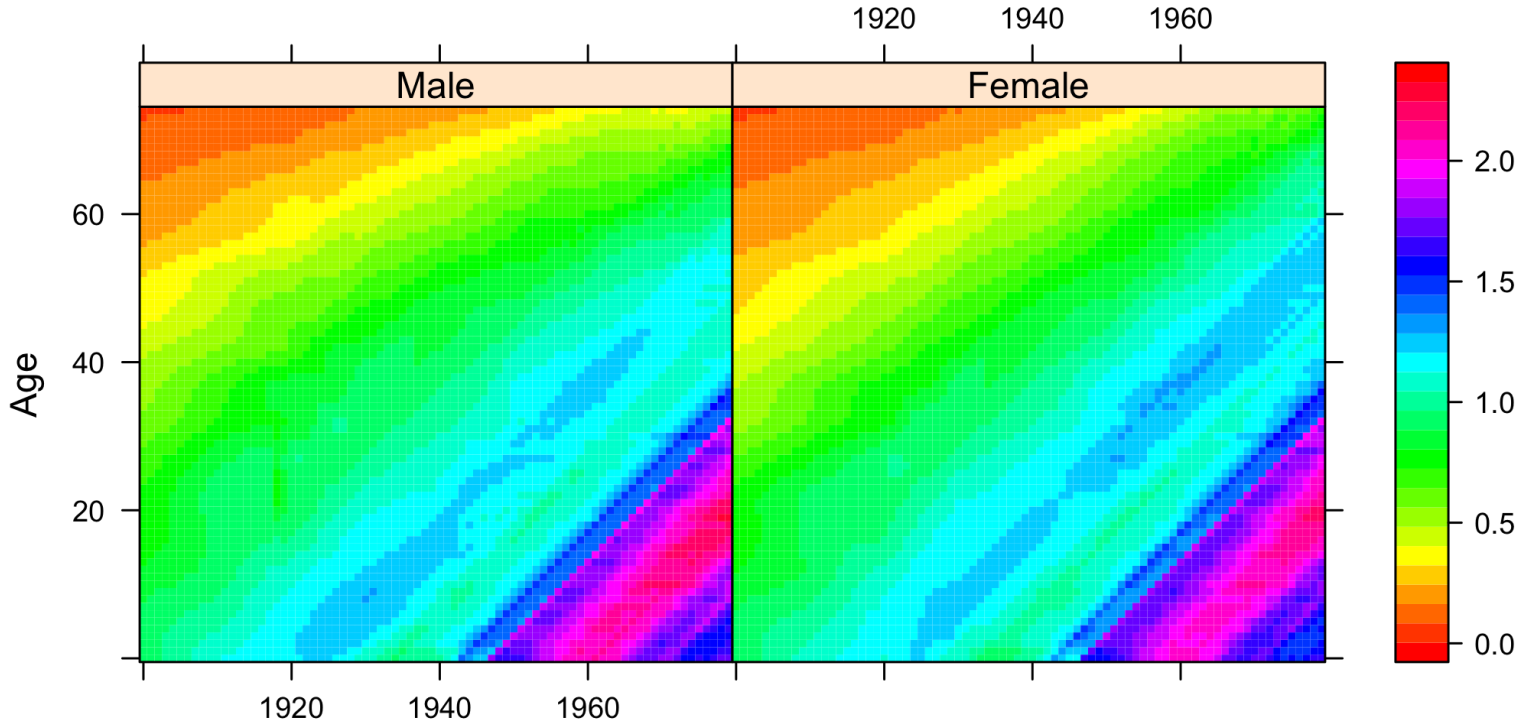


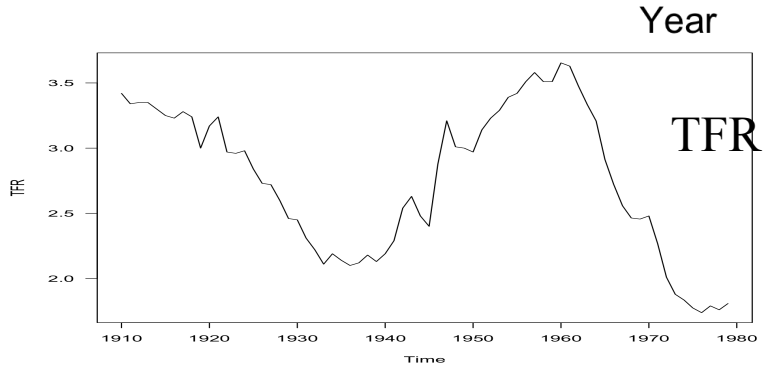
Dubin, J.A. et al. (2001) Event history graphs for censored survival data. *Statistics in Medicine*, 20:2951-2964

Hmisc

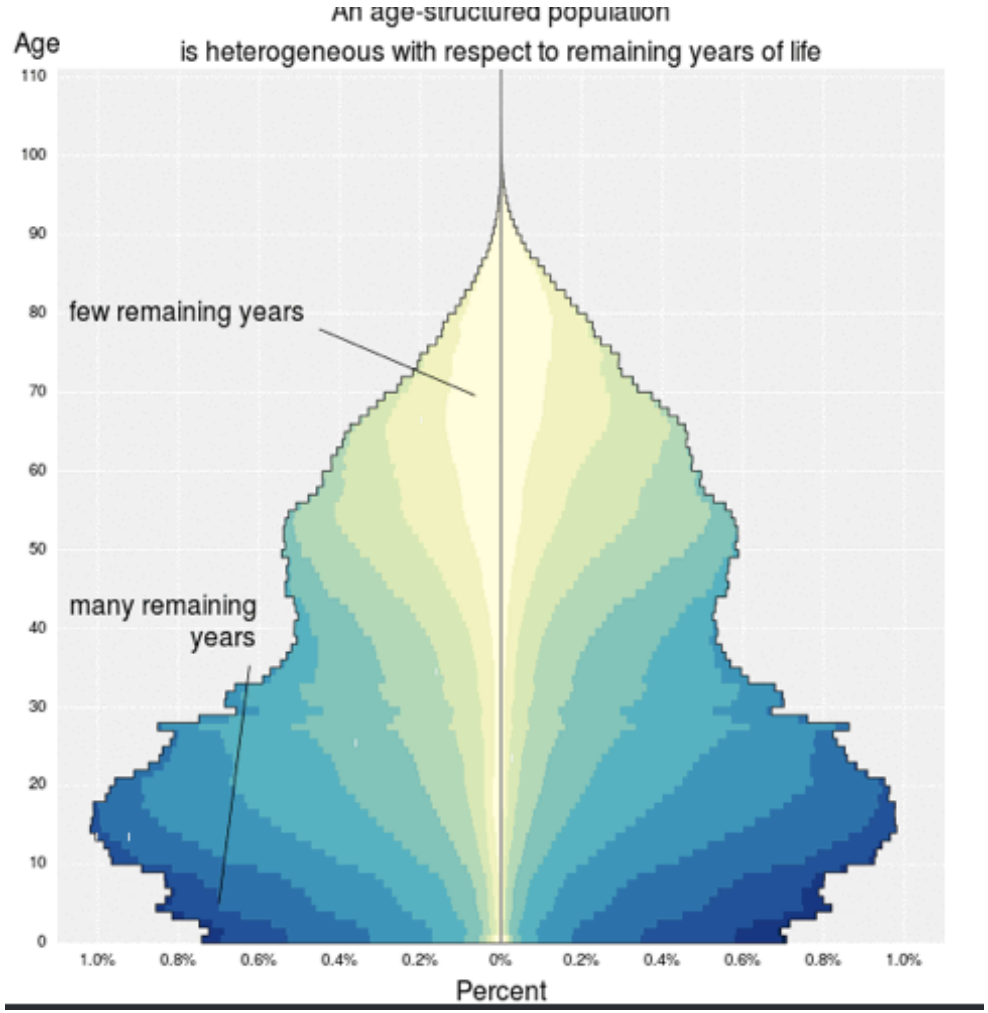
Graphics with R

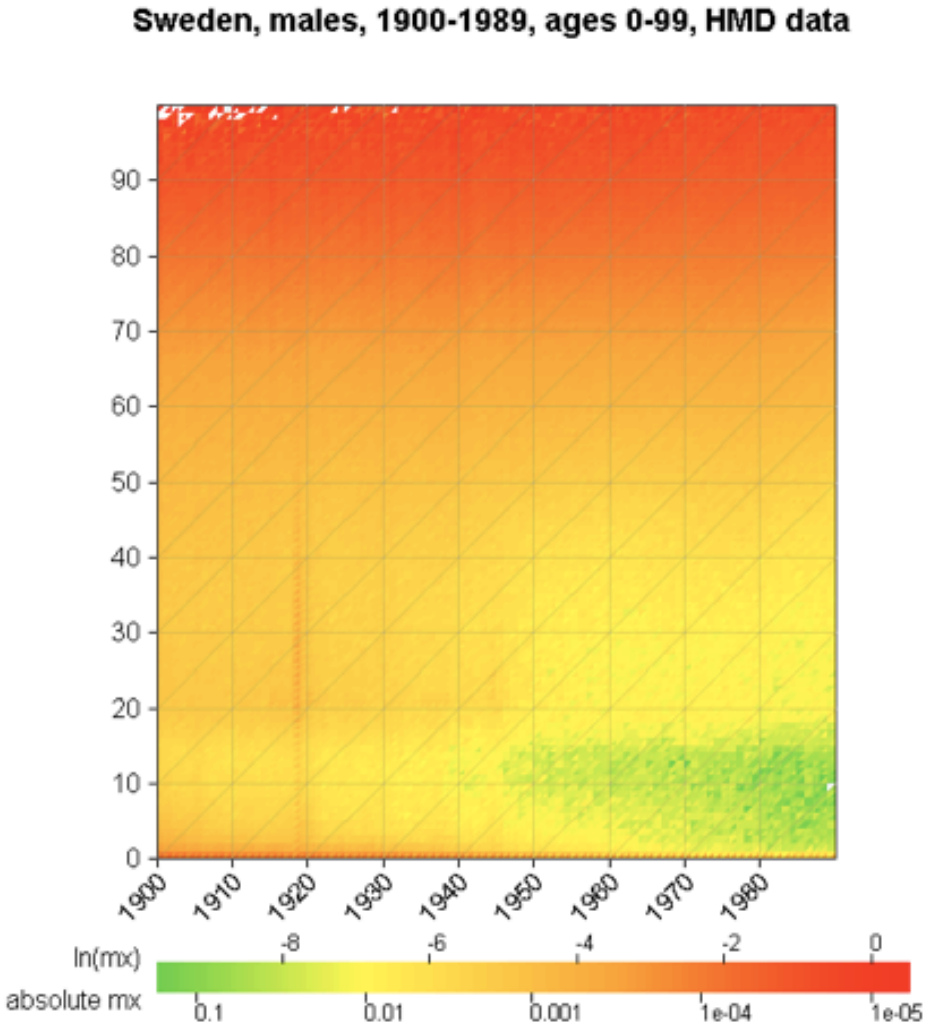
US population 1900 - 1979



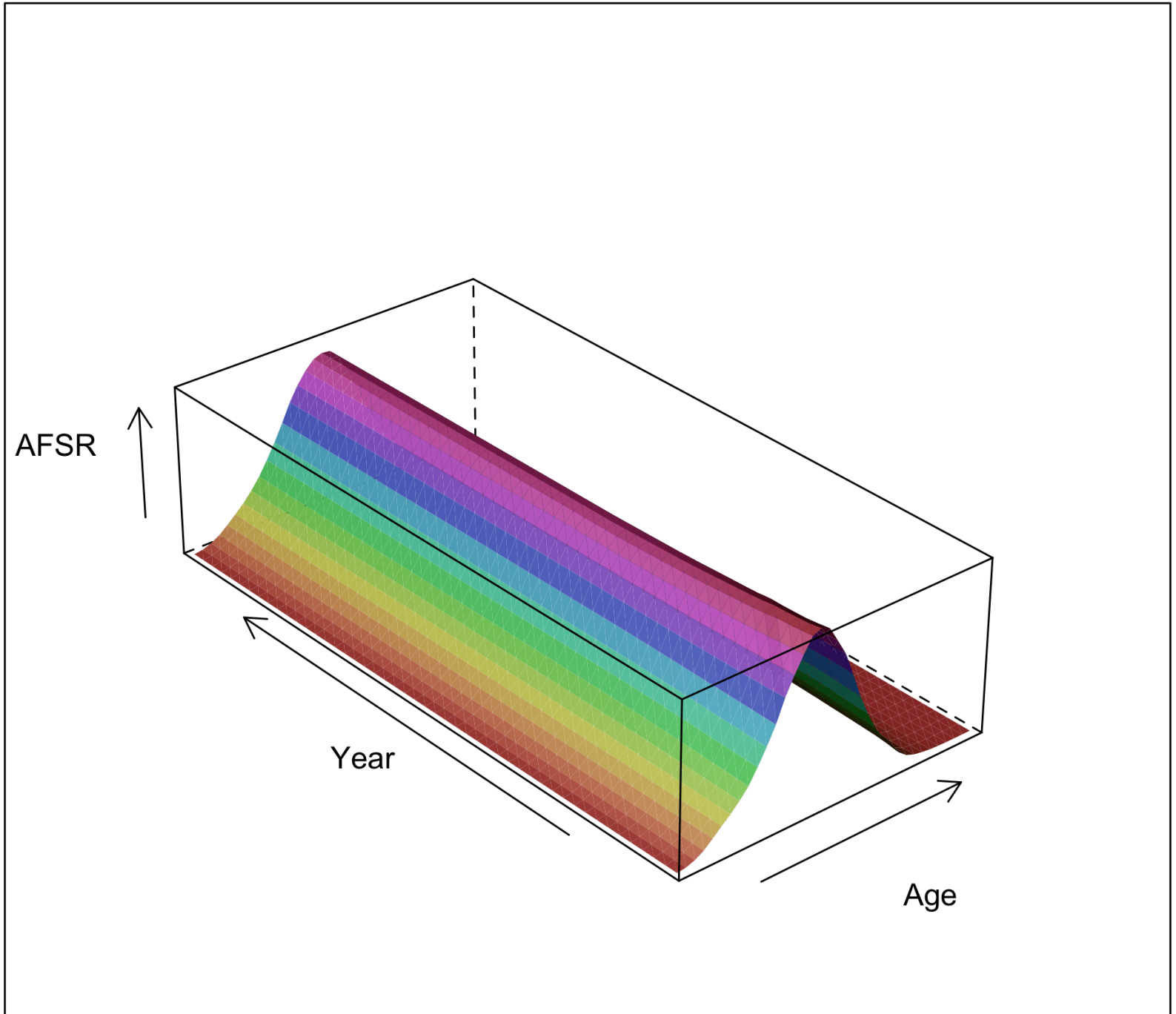


Tim Riffe, UC Berkeley





Fertility rates, Netherlands, 2004-2050 (Eurostat assumptions)



Graphic with R

googleVis

R and big data

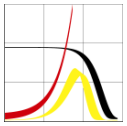
Visualize your Facebook friends network with R

http://applyr.blogspot.be/2012/01/mining-facebook-data-most-liked-status.html

r-google-analytics

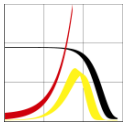
Google Analytics For The R Statistical Programming Language

Project Home | Downloads | Wiki | Issues | Source



Social interaction and emergence of social networks

- Insight from social interaction data
  - How to model social interaction?
    - Use sociological theory more effectively (e.g. emergence of social network from social interaction)
    - Stop segregation in the social sciences
    - Identify main processes (e.g. assortative mixing; homophily)
- Modeling and simulation: Agent-based modeling and computational social science



Conclusion: challenges ahead

- Visualization: produce a picture that reveals patterns and effectively communicates a main message (multimedia?)
- Process models instead of data models (implies simulation)
- Bridge the gap between theory and models: incorporate plausible causal mechanisms in demographic models
- Explore a lingua franca (unifying language) for modeling and simulation in demography
- Tranparancy in science: open source, documentation and archiving (e.g. CRAN, OpenABM)
- Training: short courses, webinars, YouTube, online courses, regular courses on modeling and simulation

Thank you