1 - Introduction to Digital Demography

Diego Alburez-Gutierrez
MPIDR
European Doctoral School of Demography 2021-22

21 February 2022

Agenda

1. Course setup and final assignment
2. Introduction to digital demography
3. **Break**
4. Current topics

Hello from the Lab of Digital and Computational Demography!

https://www.demogr.mpg.de/en/research_6120/digital_and_computational_demography_5555

Course setup

Week's schedule (CEST)

- ▶ Monday 10:00–11:30 - "Introduction to digital demography"
- ▶ Tuesday 10:00–11:30 - "Digital trace data (1)"
- ▶ Wednesday 10:00–11:30 - "Digital trace data (2)"
- ▶ Thursday 10:00–11:30 - "Crowd-sourced online data"
- ▶ Friday - No class

Technical setup

- ▶ Online lectures in Zoom + breakout rooms
- ▶ Course materials (syllabus, presentations, readings, assignment)
  - ▶ https://github.com/alburezg/EDSD22_digital_demography
  - ▶ Download complete repository
- ▶ Requirements
  - ▶ Running installation of RStudio
  - ▶ packages: 'tidyverse''knitr', 'rmarkdown', 'data.table'
  - ▶ Miktex, Texlive or RStudio's 'tinytex'

Final assignment

Goal

▶ Hands-on experience using digital trace data
▶ Changes in human mobility during Covid-19 pandemic

Exercises

1. Analyse: raw mobility data from Google, Facebook, and Apple
2. Discuss: data-generating process and data quality

See full instructions in syllabus!

Evaluation

Assignments due Friday February 25 at midnight (Anywhere on Earth):

1. A written Rmarkdown report (Word document of pdf file).
2. Include R code as inline chunks
3. Introduction to Rmarkdown:
   https://rmarkdown.rstudio.com/lesson-1.html

See full instructions in syllabus!

Introduction to Digital Demography

Who's in the (Zoom breakout) room?

💡

1. Introduce yourself (if needed)
2. Describe your research interests using ONLY three keywords
3. What is 'digital demography'?
4. Have you ever worked with digital or online data? When?

Digital and computational demography

1. Apply formal demography to digital trace data[1]
   1.1 "Digital censuses" in Facebook and Linkedin
   1.2 Life tables to estimate *survival* within platforms (users are *born* and *die*)
2. Advance theory[2]
   2.1 Demographic (*holy trinity*: mortality, fertility, migration)
   2.2 Sociological theory (explain social action)

---

[1]Cesare, N., Lee, H., McCormick, T., Spiro, E., and Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research. Demography 55(5):1979–1999.

[2]Edelmann, A., Wolff, T., Montagne, D., and Bail, C.A. (2020). Computational Social Science and Sociology. Annual Review of Sociology 46(1):61–81.

What is 'digital' demography?

▶ Digital vs analogue
▶ Online vs offline
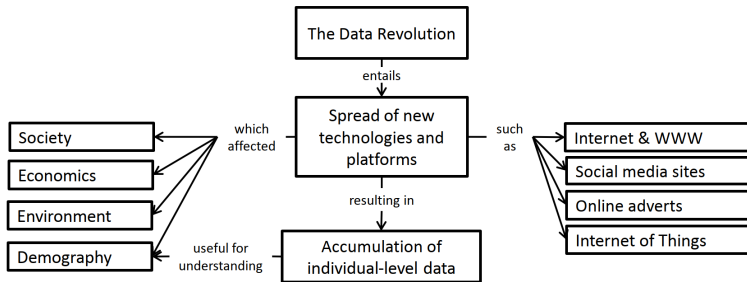▶ 'Big' vs 'small' data

Fig. 1 The Data Revolution and new sources of data for demographic analysis.

Digital data sources for demographic research

1. Digital Trace Data (online and offline)
   a. Social media
   b. Mobile phones
   c. Wearable devices, etc. . .
2. Crowd-sourced online data
   a. Wikipedia
   b. DNA and online genealogies
   c. Petitions, etc. . .
3. Simulations (made-up data)

Is 'Big data' new data?

Figure 1: Demographer collecting Big Data for the 1925 US census

https://commons.wikimedia.org/wiki/File:Volkstelling_1925_Census.jpg

Big data for social research: the good

- big
- always-on
- nonreactive

Salganik, M. (n.d.). Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press.

Big data for social research: the bad

- incomplete
- inaccessible
- nonrepresentative (within- and out-of sample)
- drifting (population, behavioural, system)
- algorithmically confounded (observing a casino)
- inaccessible
- dirty
- sensitive

Break

Current Topics in Digital Demography

Make yourself heard!

1. What are current topics in digital and computational demography?
2. Do you have any particular papers in mind?

Current topics in digital demography

1. Methodological developments
2. Understand internet users and online use
3. Migration (internal and external)
4. Mortality and morbidity
5. Online and offline fertility dynamics

1. Methodological developments

1. Inference from non-representative samples
2. Understand and address online bias
3. Nowcasting demographic processes

1. Mobility and Covid-19 dynamics
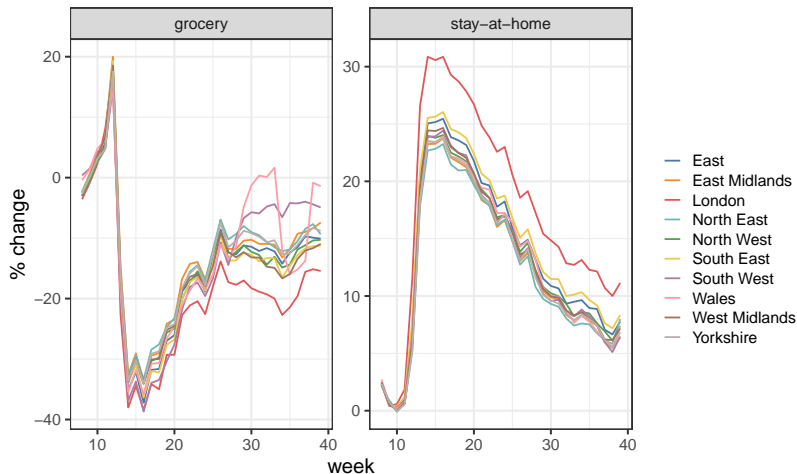
Basellini, U., Alburez-Gutierrez, D., Del Fava, E., Perrotta, D., Bonetti, M., Camarda, C. and Zagheni, E. (2021). "Linking excess mortality to Google mobility data during the COVID-19 pandemic in England and Wales". SSM - Population Health, 14. DOI: j.ssmph.2021.100799.

2. Understand internet users and online use

1. Infer demographics (age, sex, location, SE status, etc) from image and text
2. Track inequalities in online access
3. Consequences of platform use for users

2. The digital gender gap
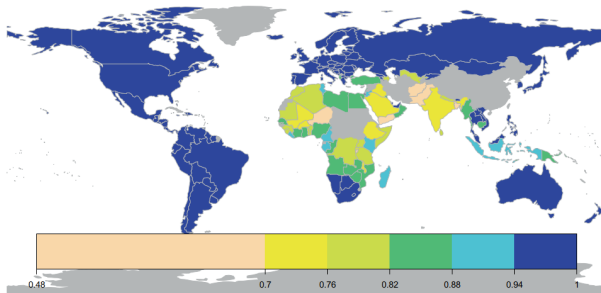
(b) Mobile Phone Gender Gap Index predicted from Online model using the Facebook age 25–29 user Gender Gap Index

Figure 2: Inequalities in digital access

Fatehkia, M., Kashyap, R., and Weber, I. (2018). Using Facebook ad data to track the global digital gender gap. World Development 107:189–209.

3. Migration (internal and external)

1. Estimate flows and stocks
2. Mobility by subgroup (eg. undocumented, highly-skilled)
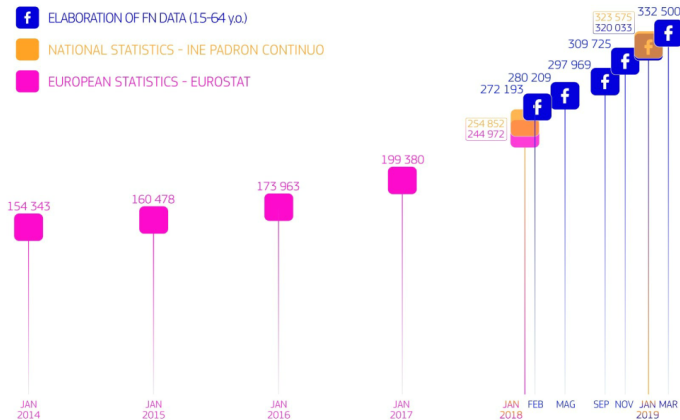3. Cultural assimilation of immigrants

Fig 8. Stocks of international migrants and FN migrants from Venezuela in Spain. The national and Eurostat statistics refer to all ages, while the FN-derived estimates refer to ages 15–64.

https://doi.org/10.1371/journal.pone.0224134.g008

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., and Rango, M. (2019). Quantifying international human mobility patterns using Facebook Network data. PLOS ONE 14(10):e0224134. doi:10.1371/journal.pone.0224134.

4. Mortality and morbidity

1. "Historical" mortality estimates (obituaries, genealogies)
2. Enhanced data collection with IOT and online surveys
3. Morbidity from online behaviour (digital traces)

4. Long-term mortality patterns
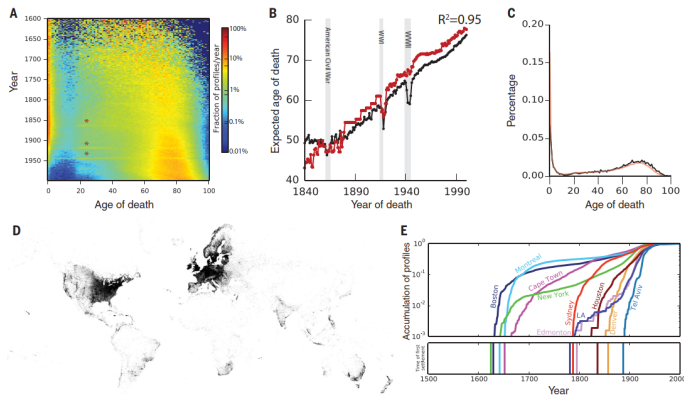
Kaplanis, J., et al. (2018). Quantitative analysis of population-scale family trees with millions of relatives. Science 360(6385):171–175.

5. Online and offline fertility dynamics

1. Estimate fertility from friendship networks
2. Online discourse around reproduction
3. Partnership formation and assortative mating (online dating)

5. Google searches and future fertility

▶ Google keyword search data
▶ 15% monthly US birth drop between November 2020 and February 2021

Figure 6: Seasonality-Adjusted Keyword Search Volumes
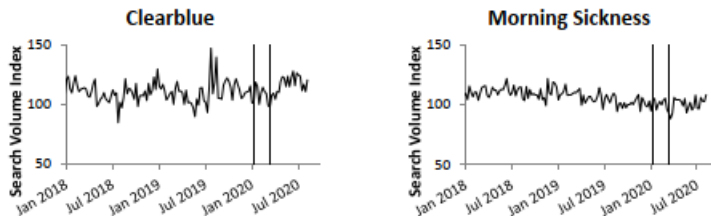
Group discussion

1. How could you appliy digital data or methods to study a topic you care for?
2. What type of data/expertise would you need?

Homework

1. Download the course materials from
   https://github.com/alburezg/EDSD22_digital_demography
2. Review the assignment instructions in the syllabus
3. Use the script in the `Assignment/R` directory to knit an
   Rmarkdown document