Summer School 'Harmonising and Visualising Data in Research on Health Inequalities'

Faculty of Science, Charles University Prague
Albertov 6, room Z3

This Summer School is designed for early-career researchers studying health inequalities through historical data. The program emphasizes developing advanced skills in data harmonisation, including standardizing, integrating, cleaning, and transforming, historical health datasets, and in data visualisation, enabling participants to create compelling visuals for historical trends and disparities. Through hands-on workshops, participants will gain practical experience in data manipulation and analysis, while also fostering networking and collaboration opportunities with experts and peers. Additionally, a dedicated Science Communication Day will equip researchers with strategies for effectively communicating historical health inequality findings to diverse audiences and preparing for impactful presentations.

Day	Topic	Teacher
Monday 25 August	Basics of R	Emre Sari
Tuesday 26 August	Cleaning, merging and manipulating data	Rick Mourits
Wednesday 27 August	Visualisation and structuring data for journals and conferences	Joris Kok, Liili Abuladze
Thursday 28 August	Visualisation for general public, narratives, media	Edward Morgan
Friday 29 August	Students' presentations	Věra Slováková

Basics of R: Exploring the Human Mortality Database
On Monday, August 25, we will focus on the fundamentals of R using the Human Mortality Database (HMD) as our case study. Participants will gain hands-on experience in importing, inspecting, and basic manipulation of real historical mortality data. This lecture will provide a foundation for harmonization and visualization in the following sessions.

Please visit the following online sources before the lecture:

R Programming Tutorial: https://youtu.be/_V8eKsto3Ug?si=W_JmGD_PuCmcK12y
R for Data Science (2e) (Online Book): https://r4ds.hadley.nz/
Data source: https://www.mortality.org/ (Don't forget to register)

Basics of R Day Agenda – Monday, 25 August 2025

9:00–10:15– Getting Started with R & RStudio
Introduction to R's role in demographic and historical health research.

- Why R? Advantages for reproducible data analysis in mortality studies
- RStudio Tour: Console, script editor, environment, and help panes
- Interactive Demo: Install and load the tidyverse; open the HMD CSV file

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY
Funded by the European Union
GREATLEAP
Faculty of Science Charles University

10:15– 10:30 Coffee break

10:30 – 12:00– Importing and Inspecting HMD Data
Uploading HMD data to R by students:

- Import Techniques: read.csv(), readr::read_csv(), etc., handling delimiters and encoding
- Data Inspection: str(), head(), summary(), etc., identifying missing entries
- Hands-On: Load a country's mortality series, check structure, flag any anomalies

12:00 – 13:00 Lunch break

13:00 – 14:15 – Core Data Structures & Basic Manipulation
Learn to shape HMD tables for analysis and visualisation:

- Vectors vs. Data Frames: when and how to use each
- Subsetting & Filtering: base R [,] vs. dplyr::filter() and select()
- Interactive Exercise: Extract specific ages for two selected countries and compare year ranges

14:15– 14:30 Coffee break

14:30 – 15:00 – Focused Exercises
Reflect on today's key takeaways and preview Tuesday's focus on cleaning and merging:

15:00 – 15:30 – AI for R: Prompting Your First Steps
Leveraging AI tools (e.g. ChatGPT) to support your R workflow:

- What to Ask
- Prompt Structure
- Hands-on practice

15:30 – 15:45 Coffee break

15:45 – 16:00 – Project Work Time
Dedicated time to begin working with your own dataset—or continue exploring the HMD—using today's tools.

16:00–21:00 Welcome drink and refreshments (Faculty of Science, Albertov 6, room Levá rýsovna)

Data harmonisation, Tuesday 26 August
On Tuesday, we will focus on how to work with data. We start the day with some data exploration before we discuss the concepts of ETL (Extract, Transform, Load) and data pipelines. There will be special attention for HISCO and ICD10h that categorize occupations and causes of death.

9:00 – 10:00 – Recap
Participants open a dataset and explore it using basic queries

10:00–10:15: Coffee Break

10:15 – 11:00 Concepts in Data Standardization
Participants are introduced to:

- ETL
- Data pipelines
- HISCO + ICD10h

11:00 – 12:00 Concepts and practices in Data Standardization
Participants learn:

- To work with relational databases
- Best practices in data cleaning

12:00–13:00 Lunch break

13:00–14:00 Data cleaning
Participants clean data discussed before the break.

14:00–14:15 Coffee break

14:15–17:00 Data manipulation
Participants perform different data manipulations using R.

Visualisation and structuring data for journals and conferences, Wednesday 27 August
On Wednesday we will focus on strategies for communicating historical health inequality findings through data visualization. The day's goal is to prepare participants for clear and interesting presentations and publications, emphasizing how visual communication can engage scientific audiences.

09:00–10:00 - Why Use Graphs and the 'Grammar of Graphics'

- Understand how data visualisations support exploratory analysis and clear communication of findings.
- Learn the key components of effective figures for different data types.
- Get introduced to the 'Grammar of Graphics' using the ggplot2 package.

10:00–10:15: Coffee Break

10:15–11:00: Visualisation Practice with Gapminder

- Implement the principles from the first session to create your own figures using the Gapminder dataset.

11:00–12:00: Common Health Inequality Visualisations and Mapping Basics

- Identify best practices for common figures in historical health inequality research.
- Explore the basics of spatial data visualisation in R.

12:00–13:00 – Lunch Break
13:00–13:45 – Finalising Your Figures + Practical Tips

- Learn how to refine your figures for clarity ready for publication.
- This session covers options for adjusting labels, axes, legends, colour schemes, and exporting high-quality outputs.

13:45–14:00 – Coffee Break

14:00–17:00 – Begin Visualising Your Own Data

- Apply your new visualisation skills to your own dataset or explore the Human Mortality Database.
- Facilitators will be available for support and guidance.

19:00 Social evening in a pub

Science Communication Day: Engaging Audiences with Historical Health Inequalities, Thursday, 28 August 2025
Participants will explore science communication strategies tailored to historical health inequality research. Drawing on expertise from high-impact ONS publications, sessions will cover crafting compelling data narratives, engaging diverse audiences, and communicating uncertainty, preparing participants for impactful presentations at Friday's feedback session.

9:00–10:15: Engaging Diverse Audiences with Health Inequality Research
Explore strategies to tailor historical health inequality research for more diverse audiences like the broader scientific community, policymakers, journalists and the general public. Using high-impact ONS publication case studies, this session focuses on audience analysis, inclusive language, and simplifying data for clarity. *Interactive exercise*: Adapt a dataset visualisation for a specific audience (e.g., policy brief or public infographic).

10:15–10:30: Coffee Break

10:30–12:00: Crafting Narratives from Historical Data
Learn to transform complex historical health datasets into compelling stories that resonate with diverse audiences. Drawing on high-profile ONS publications, this session covers narrative structure, key message identification, and contextual framing to enhance research impact. *Interactive exercise*: Draft a narrative outline for a historical health inequality dataset.

12:00–13:00: Lunch Break

13:00–14:15: Communicating Uncertainty in Historical Data
Master techniques to transparently address uncertainties in historical health datasets, avoiding misinterpretation while maintaining credibility. Informed by ONS practices, this

session covers annotating limitations and using qualifiers in visuals and narratives. *Interactive exercise*: Create an annotated visualisation highlighting data uncertainties.

14:15–14:30: Coffee Break

14:30–15:00: Wrap-Up and Preparation for Friday's Feedback Session
Synthesize key learnings and prepare for Friday's conference-style feedback session. Participants will discuss how to incorporate science communication strategies into their project presentations, drawing on session exercises.

15:00–17:00: Project Work Time
Dedicated time for participants to refine their assigned projects, applying science communication techniques from the day's sessions to enhance their outputs for Friday's feedback session.

Students´ presentations, Friday 29 August
On the last day of the training school all participants will have the opportunity to present their own research. Each student will have 10 minutes to present their research using tools gained from the training sessions. Students will be asked to present to both scientific and general audiences and show how they would approach them differently. Afterwards there will be 10 minutes for feedback from teachers and other students.

9:00–10:40 First session

10:40–11:00 Coffee break

11:00–12:20 Second session

12:20–13:20 Lunch break

13:20–16:00 Third session, final discussion