

Reproducible research on child and youth development (pt. 1)

Michael Kühhirt

michael.kuehhirt@uni-koeln.de

Summer 2020

Office Hours: by appointment via email
Office: 3.323, Universitätsstr. 24

Class Hours: MON 12:00-3:30 p.m.
Class Room: 1.326a, Universitätsstr. 24

Class content

Are there differences between boys and girls in terms of behavior, verbal ability, or math skills, and if so, why? How do factors like parental divorce or poverty affect children? What are the consequences of maternal smoking during pregnancy in terms of children's health and development?

In this two-part research lab, participants carry out their own empirical project on such or similar topics in the area of child and youth development using data from the Children of the National Longitudinal Survey of Youth. During the first semester, participants develop a research question based on a previous study using these data and start with data preparation. The semester concludes with a proposal that specifies the research question, outlines the motivation and theoretical background, and describes the data and methods used to answer it. The second semester is devoted to finalizing data preparation, to carrying out the data analysis, and to writing the research report.

A key goal of the course is to establish a workflow for *transparent* and *reproducible* empirical social research (on any topic). To this end, participants practice how to prepare research data (e.g., merging data sets, recoding variables, constructing analytical samples), how to efficiently produce formatted research output (e.g., tables and figures), and, most importantly, how to rigorously document these tasks so that others (including one's future self!) can understand and repeat each step of the research process and build on the work.

Goals and learning outcomes

The class has three broad goals, which I enumerate below. Each goal is associated with a number of specific learning outcomes, which you should be able to perform as the class concludes after two semesters. Students who complete the class should be able to

1. formulate an empirical research question based on existing literature. This includes
 - providing an overview of existing research on a specific topic,
 - motivating their research question with coherent theoretical arguments,
 - systematically building on (and possibly extending) existing research on a specific topic.
2. prepare and analyze existing survey data. For this, students
 - understand different data structures,
 - explore, clean, and combine different data,
 - use statistical methods appropriate to answer their research question.
3. document their empirical research project by
 - writing a research report,
 - providing materials to reproduce their research from data acquisition through data preparation to data analysis,
 - presenting their research project in class.

Although the substantive focus of the class is on child and youth development, all learning outcomes can be readily transferred to other topics in empirical social research. Many of the skills practiced in class (e.g., summarizing existing knowledge, script-based data preparation and analysis, writing and presenting) are also valuable for jobs outside academia.

Requirements and grading

To successfully complete the class, you must meet the following requirements. In the first semester (Summer 2020), you **write a proposal** which serves as the “front-end” draft for your research report and summarizes your

- research question,
- theoretical background and
- analytical approach.

In the second semester (Winter 2020), you

- **present a progress report** on your research project in class,
- **write a research report** along with documentation to reproduce your project.

In addition, there will be practical labs and homework assignments designed to facilitate successful class completion. Although these tasks aren't graded, they are immediately related to the class requirements outlined above. Completing them in the appropriate time frame will aid you in successively putting together your proposal, data project, and research report.

Research project

In your research project, you replicate an existing empirical study on child or youth development in whole or in part. The primary goal is to competently carry out some or all of the empirical analyses in the study, to appropriately motivate the analyses, and to document your project workflow so that it can be understood and reproduced by anyone who has access to the original data. An exceptional research project additionally extends the original analysis in a meaningful way, for example, by

- looking at children at a different age,
- exploring group differences,
- investigating the mechanisms underlying some statistical relation,
- using additional or alternative control variables,
- applying alternative statistical methods.

The final report should emulate the structure of an empirical article in a typical social science journal (but without an extensive literature review). Guidelines and techniques for project documentation will be discussed and applied as the class progresses. I'll provide more details on the structure, layout, and grading of the report next semester.

Workload

Upon successful completion of the class you are awarded 15 credit points, which equals a total workload of 450 units à 45 minutes. Per semester, there are only 60 units (i.e., 3 hours/week) of class time. The remaining 165 units per semester (i.e., roughly 8 hours/week) are allotted to preparing and reviewing class contents, including labs, assignments, reading, writing etc. Therefore, please schedule sufficient time outside of class hours.

Class resources

Materials on Ilias (UoC e-learning platform)

Here you'll find all class materials and a discussion forum:

https://www.ilias.uni-koeln.de/ilias/goto_uk_crs_3183868.html

National Longitudinal Survey of Youth 1979 (NLSY79)

Data for all class projects come from the [NLSY79](#), a repeated survey of men and women in the United States born between 1957 and 1964, along with [data from the children born to NLSY79 women](#).

To access these data, use the NLS Investigator (after registration):

<https://www.nlsinfo.org/investigator/pages/login.jsp>

Following the link below you'll find a database collecting publications that have used data from *Children of the NLSY79*. You can search this database using keywords that reflect your specific research interest.

<https://www.nlsinfo.org/bibliography/search/cohort+keyword/chrt=Children+of+the+NLSY79>

Open Science Framework (OSF)

OSF is an open source software project for collaboration in scientific research. We'll use it to regularly back up our class files.

To use OSF, please create an account at <https://osf.io> and then fork the project template at <https://osf.io/nx7q8/>.

Class schedule (Subject to change)

April 6:

- Lecture: Research on Child and Youth Development
- Assignment 1 (Find An Empirical Study)

April 13: NO CLASS

April 20:

- Lecture: A Workflow for Reproducible Research
- Lab 1 (Set Up The Workflow)

April 27:

- Lecture: Selecting a Research Question
- Assignment 2 (Write A Draft Introduction)

May 4:

- Lecture: The Theoretical Background
- Assignment 3 (...But What Are The Mechanisms?)

May 11:

- Lecture: The National Longitudinal Survey of Youth (NLSY)
- Assignment 4 (Alternative Explanations)

May 18:

- Lecture: Acquiring NLSY data
- Lab 2 (Become An Investigator)

May 25:

- Lab2 Continued

June 1: NO CLASS

June 8:

- Lab 2 Continued

June 15:

- Lecture: Data Exploration and Capturing Output
- Lab 3 (Explore The Data)

June 22:

- Lecture: Data Cleaning
- Lab 4 (Clean The Data)

June 29:

- Lab 4 Continued

July 6:

- Lab 4 Continued
- Assignment 5 (Write A Proposal)

July 13:

- Buffer

August 12: *Proposal Due*