

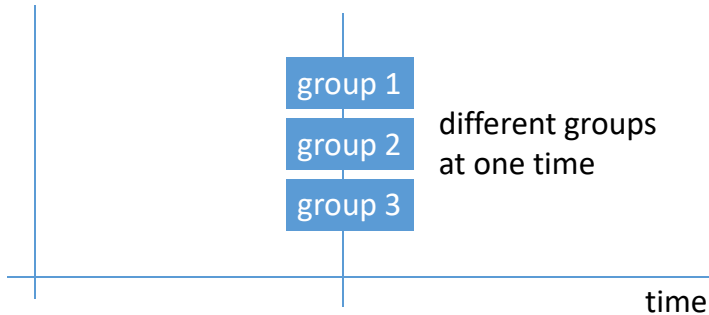
Introduction to midfieldr

2019 MIDFIELD Institute

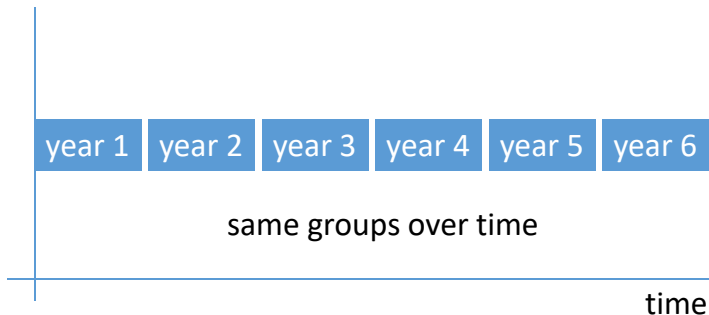
Richard Layton

June 2019

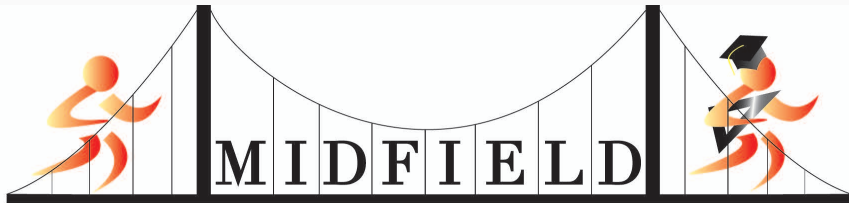
In education, cross-sectional designs are typical



Longitudinal studies offer some advantages

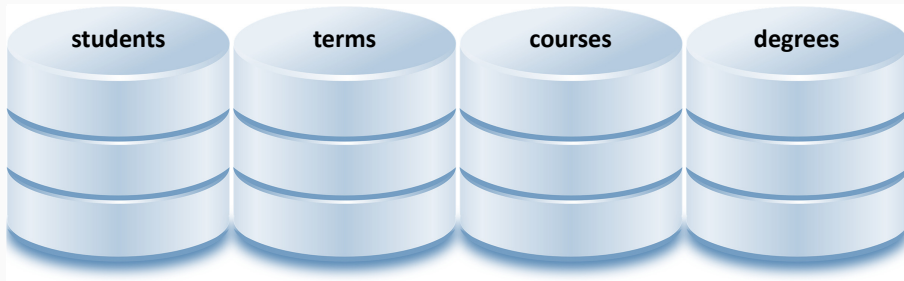


MIDFIELD is a database for longitudinal studies



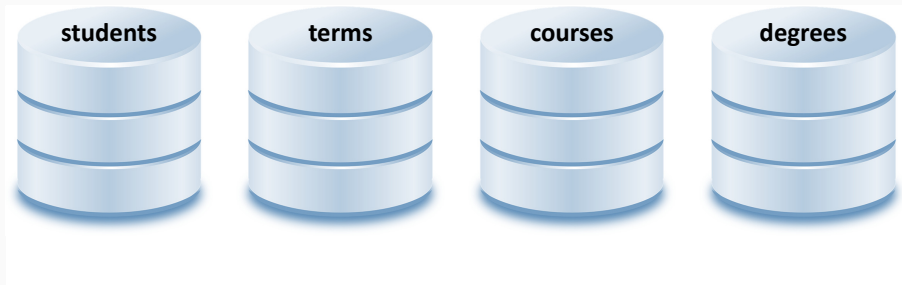
- 2.2 M undergraduate students at 23 US institutions
- Whole-population data from registrars, 1988–2017
- For more information about MIDFIELD ([link](#))

MIDFIELD data are curated in four categories



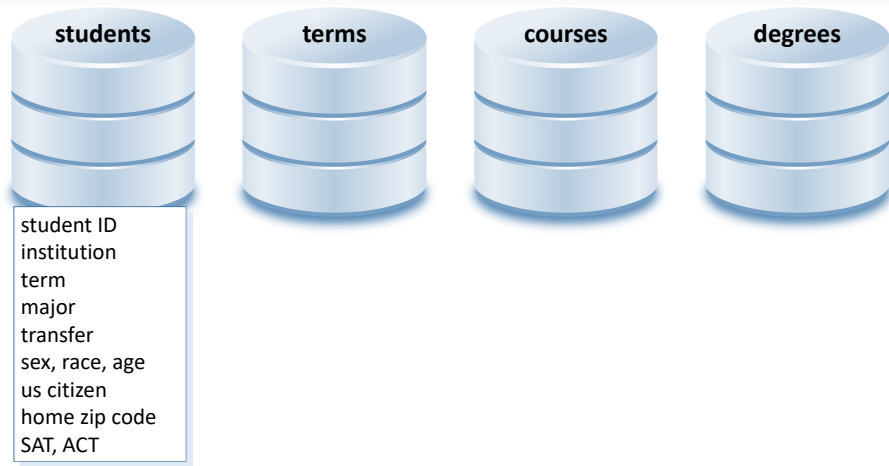
2.2M students

R package midfelddata provides a stratified sample



98,000 students

Each observation is **a unique student**

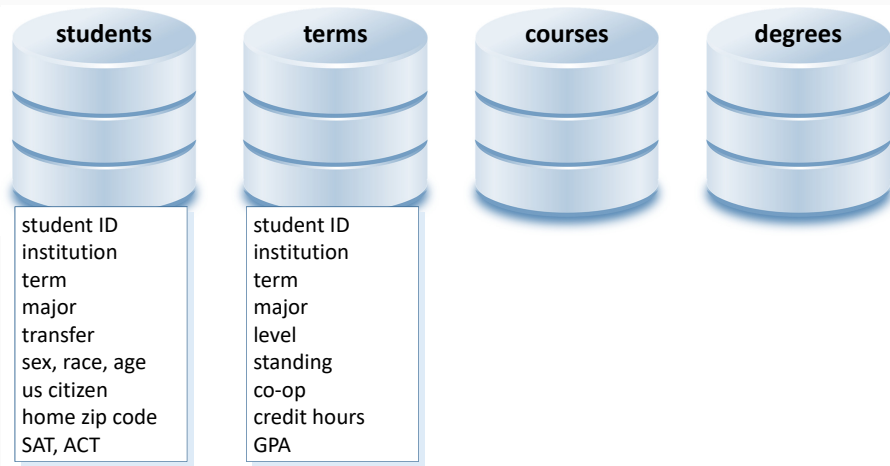


midfieldstudents

98,000 observations

19 Mb of memory

Each observation is **one term for one student**



`midfieldterms`

729,000 observations

82 Mb of memory

Each observation is **one course for one student**

students

student ID
institution
term
major
transfer
sex, race, age
us citizen
home zip code
SAT, ACT

terms

student ID
institution
term
major
level
standing
co-op
credit hours
GPA

courses

student ID
institution
term
course
--section
--hours
--type
--grade
--instructor

degrees

midfieldcourses

3.5 M observations

348 Mb of memory

Each observation is a **unique student**

students

student ID
institution
term
major
transfer
sex, race, age
us citizen
home zip code
SAT, ACT

terms

student ID
institution
term
major
level
standing
co-op
credit hours
GPA

courses

student ID
institution
term
course
--section
--hours
--type
--grade
--instructor

degrees

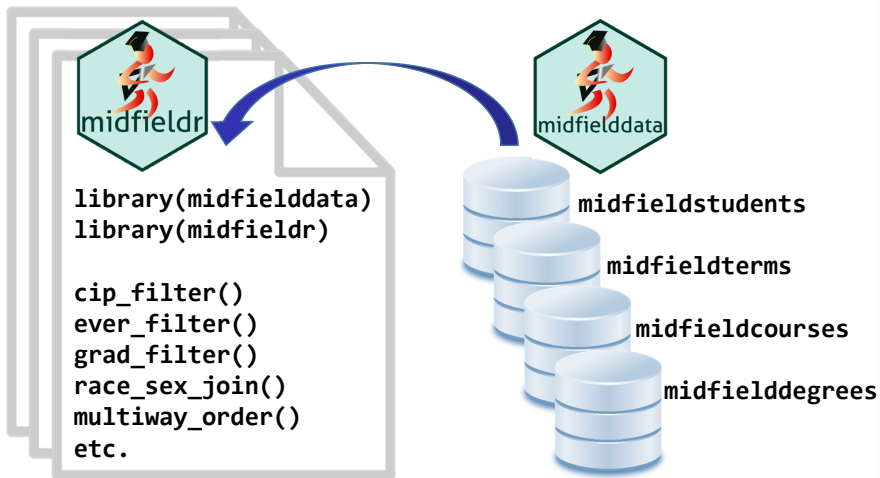
student ID
institution
term
major
degree

midfieldddegrees

98,000 observations

10 Mb of memory

midfieldr provides the functions for working with midfelddata



For more information, see these package resources online



- Package website ([link](#)) for more information and tutorials
- Package reference page ([link](#)) for descriptions of functions
- Package source code ([link](#))



- Package website ([link](#)) for more information
- Package reference page ([link](#)) for data dictionaries
- Package source code ([link](#))

References

Layton R, Long R and Ohland M (2018a) *midfielddata: Student record data for 98,000 undergraduates*. R package version 0.1.0 <https://midfieldr.github.io/midfielddata/>

Layton R, Long R and Ohland M (2018b) *midfieldr: Tools for student records research*. R package version 0.1.0.9002 <https://github.com/MIDFIELDR/midfieldr>

Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) (2019) <https://engineering.purdue.edu/MIDFIELD>

Ohland M, Zhang G, Thorndyke B and Anderson TJ (2004) The creation of the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD). *ASEE Annual Conference*. Salt Lake City