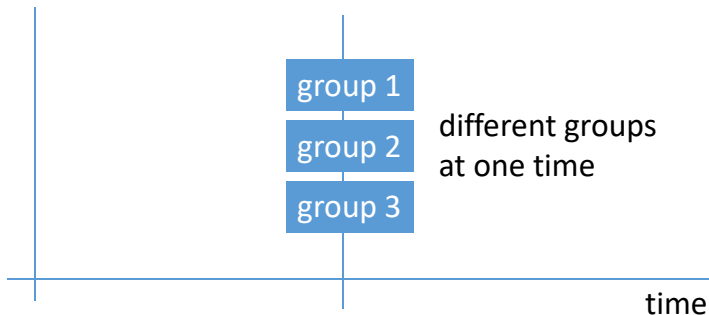# Introduction to midfieldr

2019 MIDFIELD Institute
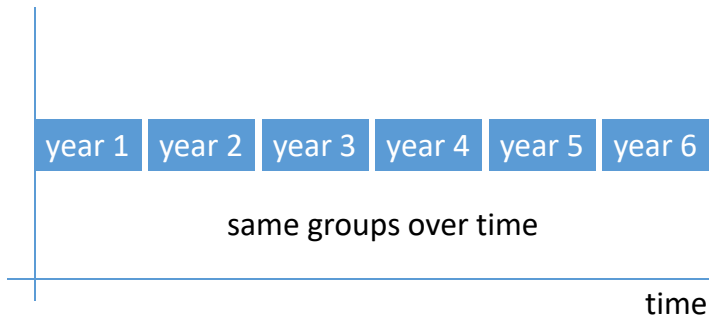

Richard Layton

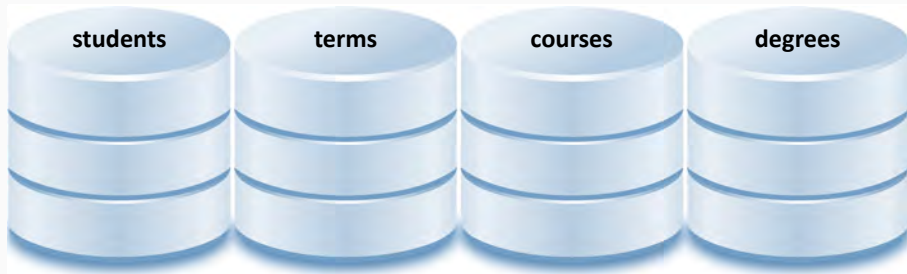June 2019

In education, cross-sectional designs are typical

Longitudinal studies offer some advantages

- 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

students   **terms**   **courses**   **degrees**

§sas 2.2M students

R package midfielddata provides a stratified sample

98,000 students

Each observation is a unique student

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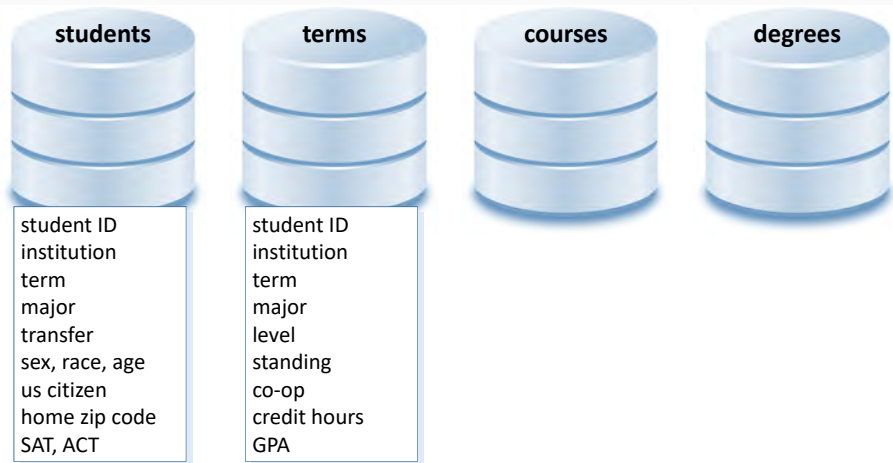
student ID
institution
term
major
transfer
sex, race, age
us citizen
home zip code
SAT, ACT

terms

courses

degrees

`midfieldstudents`

98,000 observations

19 Mb of memory

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

degrees

`midfieldterms`

729,000 observations

82 Mb of memory

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	terms	courses	degrees
student ID	student ID	student ID	student ID
institution	institution	institution	institution
term	term	term	term
major	major	course	major
transfer	level	--section	degree
sex, race, age	standing	--hours	
us citizen	co-op	--type	
home zip code	credit hours	--grade	
SAT, ACT	GPA	--instructor	
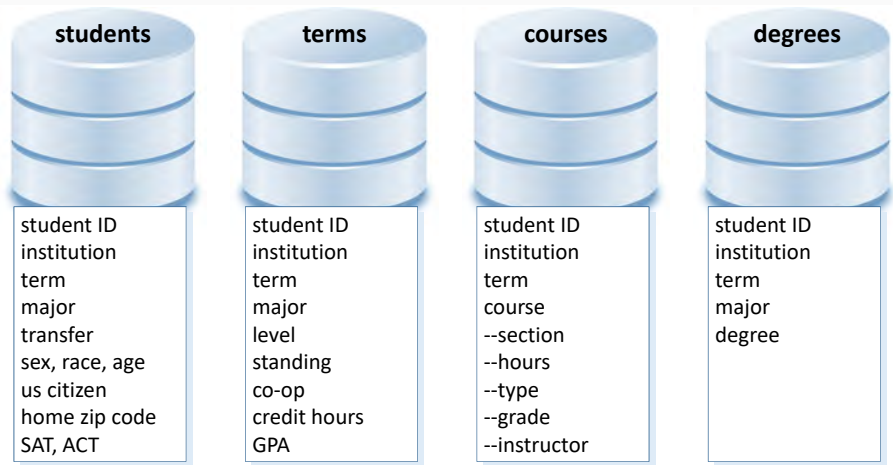
`midfielddegrees`

98,000 observations

10 Mb of memory

```
library(midfielddata)
library(midfieldr)

cip_filter()
ever_filter()
grad_filter()
race_sex_join()
multiway_order()
etc.
```

midfieldstudents

midfieldterms

midfieldcourses

midfielddegrees
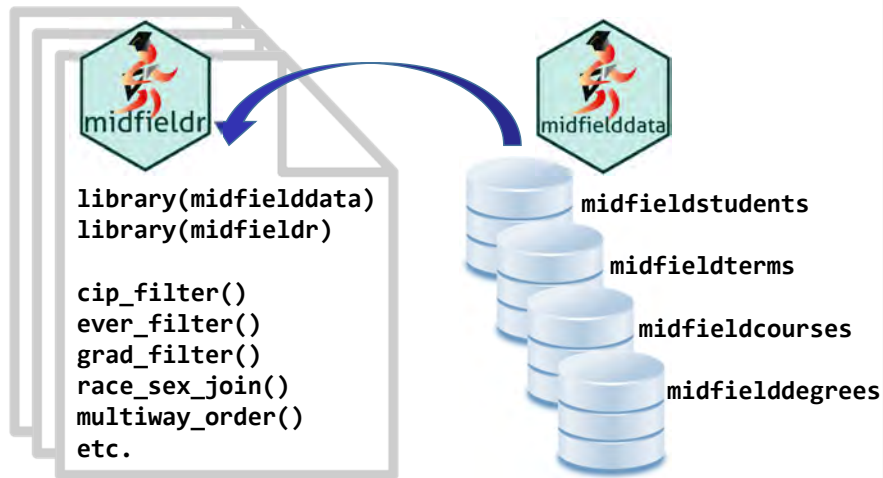
For more information, see these package resources online

- 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