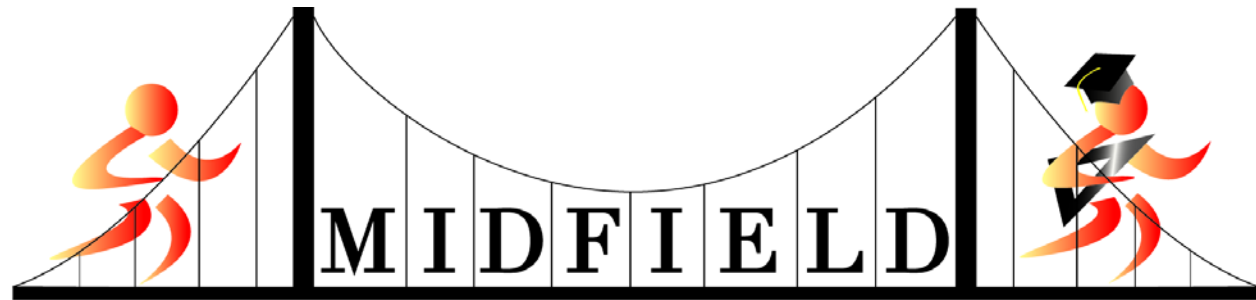# Defining your questions

MIDFIELD Institute 2019

MIDFIELD

Multiple-Institution Database For Investigating Engineering Longitudinal Development

What are you most interested in investigating?

What metric will you use?

What population(s) will you study?

Metrics

- Graduation rate*
- Stickiness*
- Yield
- Migration Yield
- Time to graduation
- Major changes

- Co-op participation
- GPA
- Credit hours
- Course grades
- …

*vignettes available to walk you through

midfieldr `0.1.0.9002`   🏠   Getting started   Vignettes ▾   Papers/Talks ▾

Identifying programs by CIP codes

Selecting groups of programs

Imputing starting majors for FYE students

Computing graduation rate

Computing stickiness

Multiway data, graphs, and tables

midfieldr

A package for investigating student record data provid   n the MIDFIELD project.

Analytical tools for research in student pathways are g   type of intersectional research for anyone with basic proficiency in R and familiarity with packages from the tidyverse.

midfieldr provides access to midfielddata, a data package containing a stratified sample of the MIDFIELD database. The sample comprises demographic, term, course, and degree information for 97,640 undergraduate students from 1987 to 2016. The purpose of creating these packages is to share our data, methods, and metrics for intersectional research in student persistence. Potential audiences include:

- Researchers interested in student persistence metrics, especially if they are R novices
- Institutional researchers responsible for reporting student persistence metrics
- Statistics instructors looking for new data for their students explore

Who will you include?

- Transfer students, or only first-time in college?
- Part-time, or only full-time?
- Spring and summer admits, or only fall?
- International students?
- Anyone ever enrolled in a given major, or just starters?
- All institutions, or only those with FYE programs?
- All start years, or only those with n years of data?

Data window

- Some students have 20+ years of data, others have 1 or 2
  - Depends on start year and institution
- Important to choose a common measurement point
  - Otherwise results are biased towards students with the most data
- Important to include/exclude by start year, not graduation status
  - Otherwise results will be biased toward those who graduate quicker

Starting major in first-year programs

If a student enters an engineering major after FYE, we impute that major as their starting major. The rest we have dealt with in 3 ways:

1. **Counts.** Partition each institution-race/ethnicity-sex group into majors proportionally to the distribution of those who did declare an engineering major. Round to the nearest integer and add that number to the discipline starting population.

2. **Top 5**. Assign them to wherever the went afterward, including unknown.

3. **Multiple Imputation in R**. Statistically assign a starting major to each student such that overall proportions are similar to case 1.
    - Vignette available to calculate
    - Standalone dataset available to use

Report out

- What are you most interested in investigating?

- This afternoon you will work with a partner to prepare the data

- Tomorrow we will talk about data visualization and give you more time to work and create a work-in-progress poster

- Tomorrow afternoon you will present your poster to the group

Recommendations

- Start with a vignette, adapt as necessary
- Check your work frequently (see data basics)
- Ask questions
- Help each other
- Have fun!

Refining your analysis

- Are your results reasonable?
- Have you carefully defined your population?
  - FTIC/transfer
  - PT/FT
  - First-year programs
  - Semesters vs. quarters
  - Are all institutions represented?
  - Are all years represented?
  - PosGrad6
- Have you carefully defined and chosen your metrics?
  - Persistence or graduation?
  - 8 semester/4 year/6 year….

HBCUs

- URM2 (Beth Myers work)

Other issues MIDFIELD or IR users have encountered? Other solutions?