

Exploring the Data Structure

MIDFIELD Institute

June 12, 2024



Institute Objectives

Qualitatively, by the end of the workshop participants should be able to:

1. **Describe the data available in MIDFIELD**
2. **Describe how the MIDFIELD data are organized**
3. Describe key principles of effective data visualization
4. Draft a research question that can be addressed using MIDFIELD

Computationally, participants should be able use `midfieldr`, an R package specifically designed for use with MIDFIELD, to:

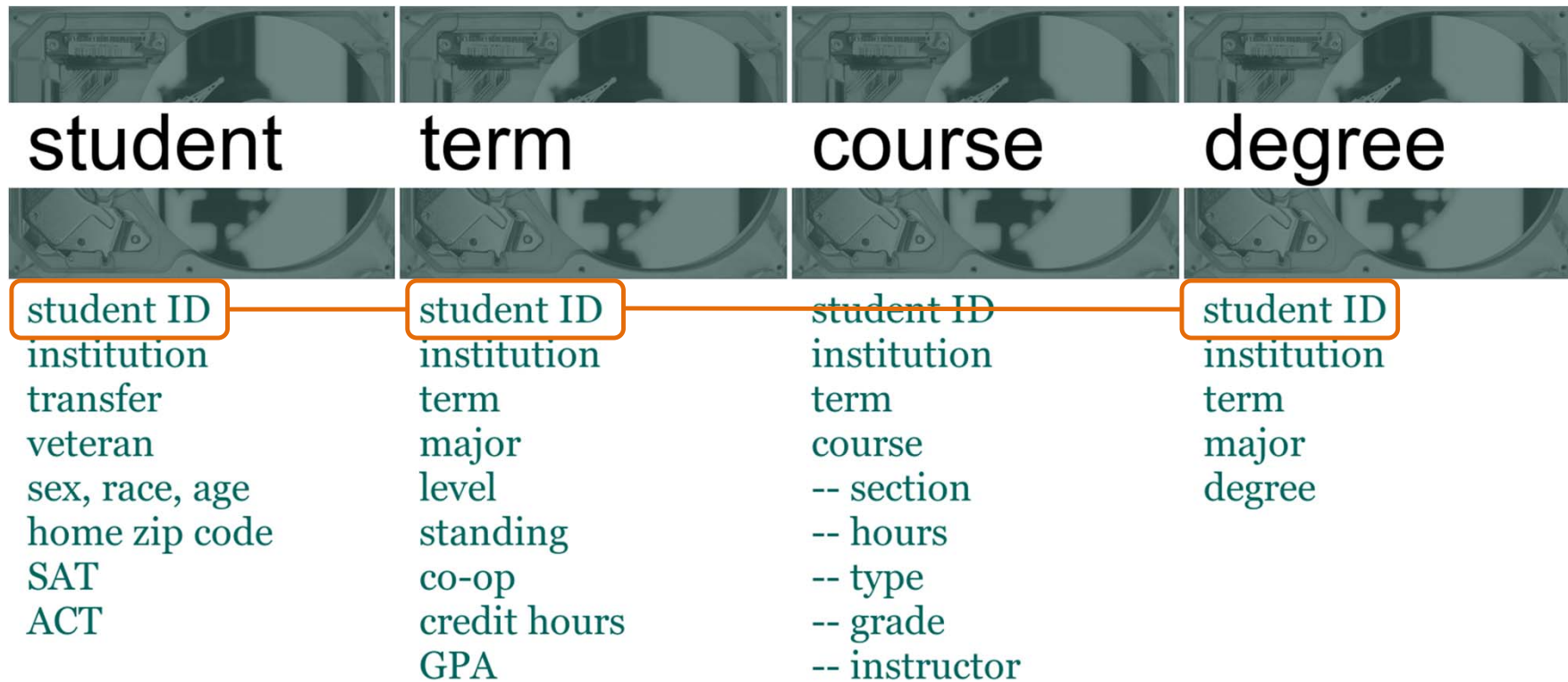
1. Subset MIDFIELD data to obtain a population to study
2. Classify the student records by desired groupings
3. Summarize the data by groups and display results

MIDFIELD data are organized in four tables.

student	term	course	degree
student ID institution transfer veteran sex, race, age home zip code SAT ACT	student ID institution term major level standing co-op credit hours GPA	student ID institution term course -- section -- hours -- type -- grade -- instructor	student ID institution term major degree

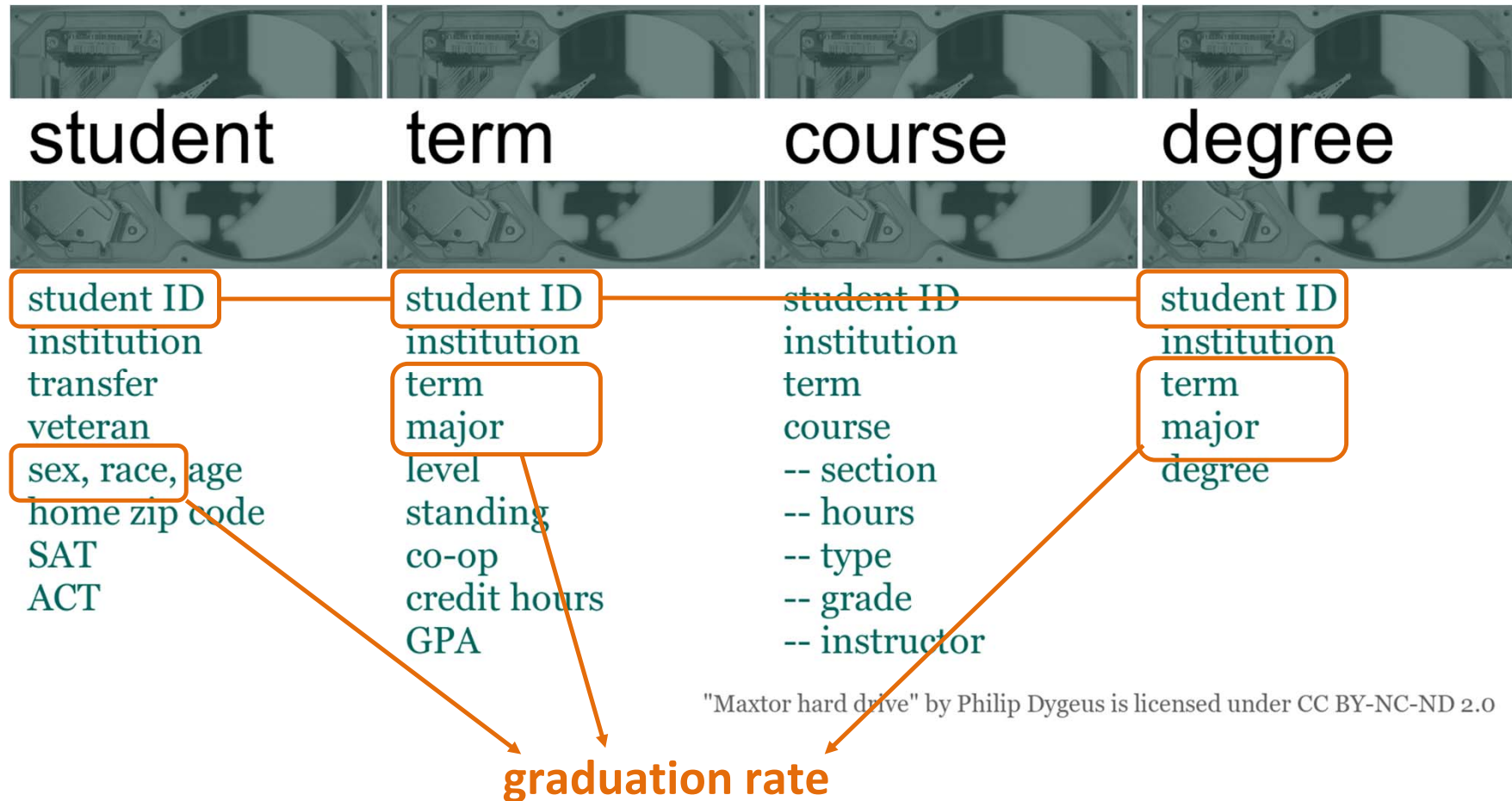
"Maxtor hard drive" by Philip Dygeus is licensed under CC BY-NC-ND 2.0

The real power of MIDFIELD is that we can link across tables.



"Maxtor hard drive" by Philip Dygeus is licensed under CC BY-NC-ND 2.0

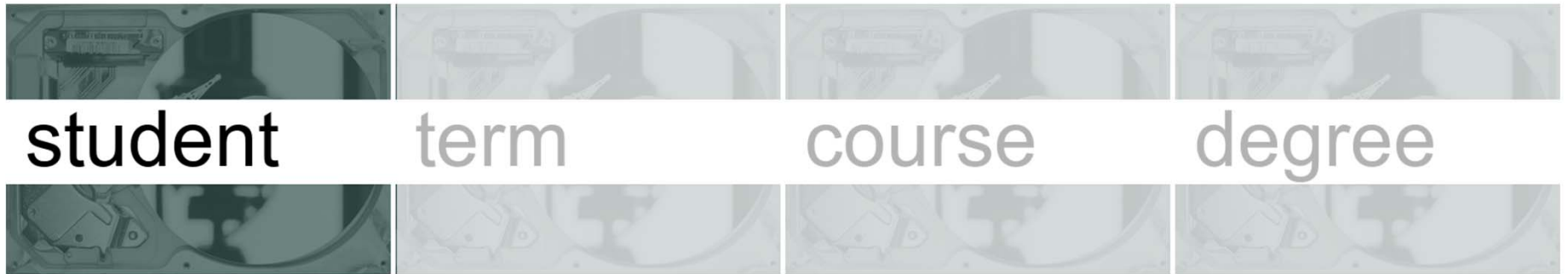
The real power of MIDFIELD is that we can link across tables.



"Maxtor hard drive" by Philip Dygeus is licensed under CC BY-NC-ND 2.0

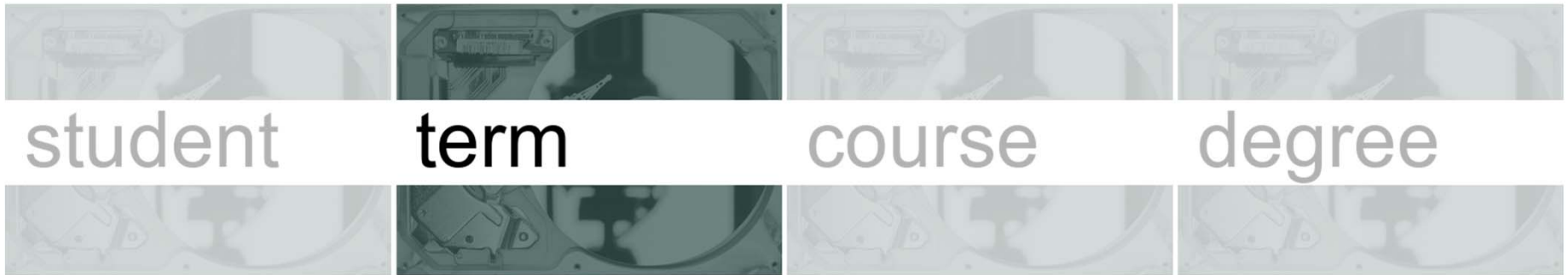
Use the tables to answer these questions about a student.

1. What can you tell about their **demographics**?
2. When did they **start**?
3. What was their **starting major**?
4. Did they **change majors**? To what and when?
5. What courses did they do well in?
6. Did they **graduate** in engineering? **How long** did it take?



mcid	institution	transfer	race	sex
MCID3111264877	Institution B	First-Time in College	White	Male

Additional variables in midfielddata
age
us_citizen
home_zip
high_school
hours_transfer
sat_math
sat_verbal
act_comp



mcid	institution	term
MCID3111264877	Institution B	19901
MCID3111264877	Institution B	19903
MCID3111264877	Institution B	19911
MCID3111264877	Institution B	19913
MCID3111264877	Institution B	19914
MCID3111264877	Institution B	19921
MCID3111264877	Institution B	19923
MCID3111264877	Institution B	19931
MCID3111264877	Institution B	19933
MCID3111264877	Institution B	19941

term is in the format YYYYT

19901 = Fall 1990

YYYY: Academic year (AY)

1990: 1990-1991

Fall 1990 – Summer 1991

T: Term of the AY:

1 : fall

2 : winter (quarter)

3 : spring

4, 5, and 6 : summer

19913 = Spring 1992



mcid	institution	term	cip6	level	hours_cumul	gpa_cumul
MCID3111264877	Institution B	19901	141001	01 First-year	15	2.49
MCID3111264877	Institution B	19903	140201	02 Second-year	32	2.64
MCID3111264877	Institution B	19911	140201	02 Second-year	47	2.64
MCID3111264877	Institution B	19913	140801	03 Third-year	62	2.45
MCID3111264877	Institution B	19914	140801	03 Third-year	69	2.36
MCID3111264877	Institution B	19921	240199	03 Third-year	81	2.50
MCID3111264877	Institution B	19923	261399	04 Fourth-year	97	2.60
MCID3111264877	Institution B	19931	261399	04 Fourth-year	111	2.54
MCID3111264877	Institution B	19933	261399	05 Fifth-year Plus	127	2.53
MCID3111264877	Institution B	19941	261399	05 Fifth-year Plus	136	2.52

Additional variables
Standing
hours_term
hours_cumul_attempt
gpa_term

CIP | THE CLASSIFICATION OF INSTRUCTIONAL PROGRAMS

CIP 2010

- Search Options ▼
- FAQs
- Resources
- Help
- Contact
- CIP Wizard

- Browse all CIP codes
- Search CIP Codes
- View Crosswalk 2010-2020

The Classification of Instructional Programs (CIP) provides a taxonomic scheme that supports the accurate tracking and reporting of fields of study. CIP was originally developed by the U.S. Department of Education's National Center for Education Statistics (NCES) in 1980, with revisions occurring in 1985, 1990, 2000, 2010 and 2020. Information on the 1985, 1990, 2000 and 2010 CIP can be accessed on the [resources](#) page under the section heading Archive and Historical. On the 2020 CIP Website, you can view both the 2020 CIP and the 2010 CIP. The default option is to view the 2020 CIP, which is the most recent version of the CIP. To view the 2010 CIP on this webpage, look for the Change Year Box, click on the down arrow and select 2010.

WHAT'S NEW AT CIP

NEW RELEASE: [CIP Website Training Video](#)

NEW RELEASE: [Updated CIP FAQs](#)

NEW RELEASE: [2020 CIP/SOC Crosswalk](#)

Ask the CIP Specialist



Single CIP Code Institutions
Exploring single CIP Code institutions, which

Search

Enter search terms if desired, filter by any of the available options, and click 'Search'. Note: Separate search terms with spaces.

[Advanced Search](#)

Search within: Title Definition Examples

2-digit series:

CIP types: 2 digit CIP code 4 digit CIP code 6 digit CIP code

Actions: No substantive changes New Deleted Moved from Moved to

CIP

term	cip6
19901	141001
19903	140201
19911	140201
19913	140801
19914	140801
19921	240199
19923	261399
19931	261399
19933	261399
19941	261399

14) ENGINEERING.

- + 14.01) Engineering, General.
- 14.02) Aerospace, Aeronautical and Astronautical Engineering.
- + 14.03) Agricultural/Biological Engineering and Bioengineering.
- + 14.04) Architectural Engineering.
- + 14.05) Biomedical/Medical Engineering.
- + 14.06) Ceramic Sciences and Engineering.
- + 14.07) Chemical Engineering.
- 14.08) Civil Engineering.
- + 14.09) Computer Engineering, General.
- 14.10) Electrical, Electronics and Communications Engineering.

[14.0201\) Aerospace, Aeronautical and Astronautical Engineering, General.](#)

[14.0801\) Civil Engineering, General.](#)

[14.0802\) Geotechnical Engineering.](#)

[14.0803\) Structural Engineering.](#)

[14.0804\) Transportation and Highway Engineering.](#)

[14.0805\) Water Resources Engineering.](#)

[14.0899\) Civil Engineering, Other.](#)

[14.1001\) Electrical, Electronics and Communications Engineering, General.](#)

[14.1003\) Laser and Optical Engineering.](#)

[14.1004\) Telecommunications Engineering.](#)

24) LIBERAL ARTS AND SCIENCES, GENERAL STUDIES

- 24.01) Liberal Arts and Sciences, General Studies and Interdisciplinary Studies

[24.0101\) Liberal Arts and Sciences/Liberal Studies, General.](#)

[24.0102\) General Studies.](#)

[24.0103\) Humanities/Humanistic Studies.](#)

[24.0199\) Liberal Arts and Sciences, General Studies and Humanities, Other.](#)

26) BIOLOGICAL AND BIOMEDICAL SCIENCES.

- + 26.01) Biology, General.
- + 26.02) Biochemistry, Biophysics and Molecular Biology.
- + 26.03) Botany/Plant Biology.
- + 26.04) Cell/Cellular Biology and Anatomical Sciences.
- + 26.05) Microbiological Sciences and Immunology.
- + 26.07) Zoology/Animal Biology.
- + 26.08) Genetics.
- + 26.09) Physiology, Pathology and Related Sciences.
- + 26.10) Pharmacology and Toxicology.
- + 26.11) Biomathematics and Bioinformatics.
- + 26.12) Biotechnology.
- 26.13) Ecology, Evolution, Systematics, and Population Biology.

[26.1301\) Ecology.](#)

[26.1302\) Marine Biology and Biological Oceanography.](#)

[26.1303\) Evolutionary Biology.](#)

[26.1304\) Aquatic Biology/Limnology.](#)

[26.1305\) Environmental Biology.](#)

[26.1306\) Population Biology.](#)

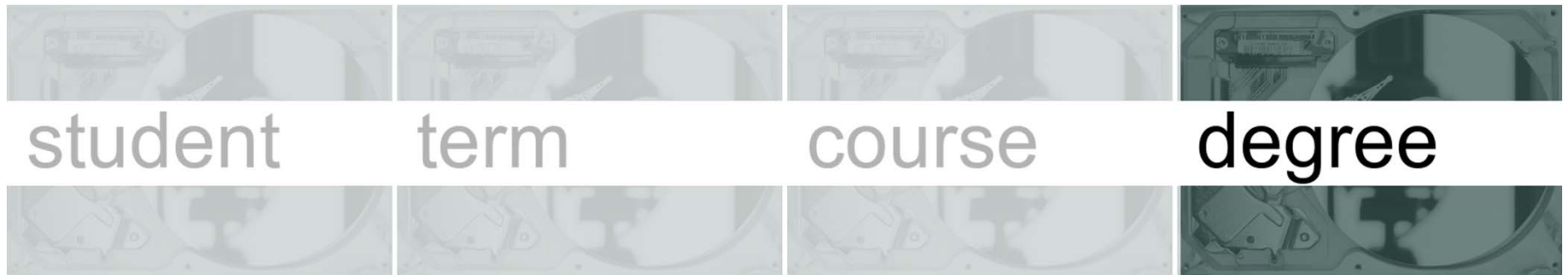
[26.1307\) Conservation Biology.](#)

[26.1308\) Systematic Biology/Biological Systematics.](#)

[26.1309\) Epidemiology.](#)

[26.1310\) Ecology and Evolutionary Biology.](#)

[26.1399\) Ecology, Evolution, Systematics and Population Biology, Other.](#)



mcid	institution	term	cip6	degree
MCID3111264877	Institution B	19941	261399	Bachelor of Science in Evolution and Population Biology

Year	Fall	Spring	Summer
1	19901	19903	
2	19911	19913	19914
3	19921	19923	
4	19931	19933	
5	19941		



student

term

course

dearee

mcid	institution	term	course	abbrev	Number	section	hours_course	grade
MCID3111264877	Institution B	19901	Calculus 1 For Engineers	APPM	1350	020	4	C
MCID3111264877	Institution B	19901	General Physics 1	PHYS	1110	001	4	C+
MCID3111264877	Institution B	19901	Hist Of England To 1660	HIST	1113	001	3	B-
MCID3111264877	Institution B	19901	Logarith/Exponent Funct	MATH	1020	001	1	A
MCID3111264877	Institution B	19901	Science Of Flight	ASEN	1016	001	3	B-
MCID3111264877	Institution B	19903	Calculus 2 For Engineers	APPM	1360	020	4	C-
MCID3111264877	Institution B	19903	Engineering General Chem	CHEM	1211	001	3	B
MCID3111264877	Institution B	19903	Engr Gen Chemistry Lab	CHEN	1221	351	2	B
MCID3111264877	Institution B	19903	Intro Space Science	ASEN	1020	001	2	A-
MCID3111264877	Institution B	19903	Intro To Human Geography	GEOG	1992	001	3	C
MCID3111264877	Institution B	19903	Intro To Scientific Prog	CSCI	1700	010	3	A
MCID3111264877	Institution B	19911	Calculus 2 For Engineers	APPM	1360	020	4	C
MCID3111264877	Institution B	19911	Experimental Physics 1	PHYS	1140	002	1	C+
MCID3111264877	Institution B	19911	General Physics 2	PHYS	1120	002	4	B
MCID3111264877	Institution B	19911	Intro To Civil Engr	CVEN	1306	001	3	B-
MCID3111264877	Institution B	19911	Mechanics 1	ASEN	2010	001	3	B
MCID3111264877	Institution B	19913	Biological Psych 1	PSYC	2012	001	3	B-
MCID3111264877	Institution B	19913	Calculus 3 For Engineers	APPM	2350	020	4	D
MCID3111264877	Institution B	19913	Engineering Drawing	GEEN	1017	001	2	D



Additional variables
section
type
faculty_rank
pass_fail

Take a few minutes to explore Example #4 (MCID3112470255). Use the following guiding questions

- 1. What can you tell about their **demographics**?
- 2. When did they **start**?
- 3. What was their **starting major**?
- 4. Did they **change majors**? To what and when?
- 5. What courses did they do well in?
- 6. Did they **graduate**? **How long** did it take?

Student

Term

Course

Degree

Check your results for Example #4.

1. What can you tell about their **demographics**? **White male, FTIC**
2. When did they **start**? **20101 -> Fall 2010**
3. What was their **starting major**? **140801 -> Civil Engineering**
4. Did they **change majors**? If so, to what and when? **Yes. 141001 -> Electrical Engineering in Spring of first year, then 141901-> Mechanical**
5. In what courses did they do well? **Grand Challenges, Design, mechanics courses,**
6. Did they **graduate**? **How long** did it take them? **Yes. 20143 -> Spring 2015, so just under 5 years.**

How do we see the story of many students?

- Utilize data manipulation and analysis software such as R
- Excel is insufficient to open the full dataset
 - 1,048,000 row limit
- Some questions will require advanced quantitative methods
- *midfieldr* has tools for working with *midfielddata*



Next up:

- Break
- 3:00-5:00 ET Guided Practice
 - Back to Institute website (<https://midfieldr.github.io/2024-midfield-institute/agenda.html#wednesday-june-12>)
- Please check in with us before leaving the virtual meeting.
 - How far did you get?
 - What are you thinking of using MIDFIELD for?