Exploring the Data Structure

MIDFIELD Institute

Aug 3, 2022

MIDFIELD

Multiple-Institution Database For Investigating Engineering Longitudinal Development

Institute Objectives

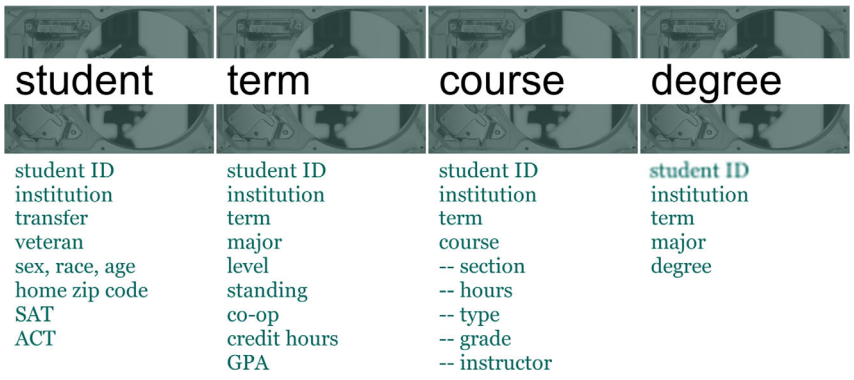
Qualitatively, by the end of the workshop participants should be able to:

1. **Describe the data available in MIDFIELD**
2. **Describe how the MIDFIELD data are organized**
3. Describe key principles of effective data visualization

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

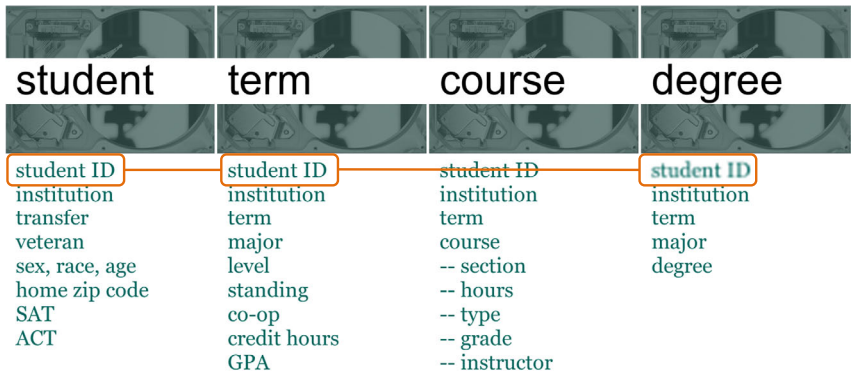
1. Subset MIDFIELD data to obtain a population to study
2. Classify the student records by desired groupings
3. Summarize the data by groups and display results

MIDFIELD data are organized in four tables.



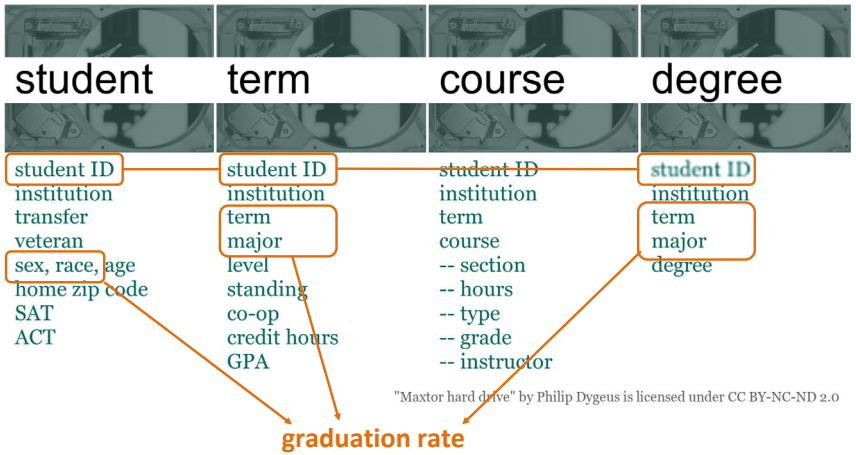
"Maxtor hard drive" by Philip Dygeus is licensed under CC BY-NC-ND 2.0

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 repeat any courses?
7. Did they **graduate**? **How long** did it take?

https://bit.ly/MIDFIELD-Institute

mcid	institution	transfer	race	sex
MID25836044	Institution M	First-Time in College	White	Female

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

mcid	institution	term
MID25836044	Institution M	20001
MID25836044	Institution M	20003
MID25836044	Institution M	20011
MID25836044	Institution M	20013
MID25836044	Institution M	20021
MID25836044	Institution M	20023
MID25836044	Institution M	20025
MID25836044	Institution M	20031
MID25836044	Institution M	20033

term is in the format YYYYT

20011 = Fall 2001

YYYY: Academic year (AY)
2001 : 2001-2002
Fall 2001 – Summer 2002

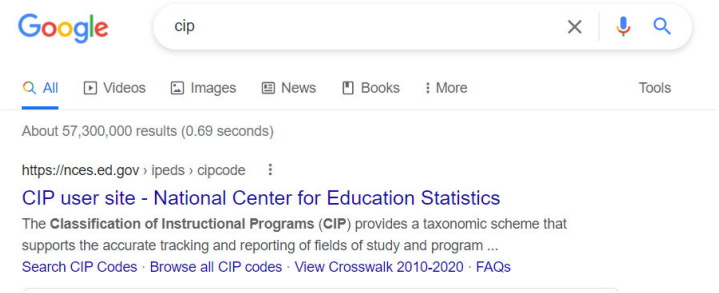
T: Term of the AY:
1 : fall
2 : winter (quarter)
3 : spring
4, 5, and 6 : summer

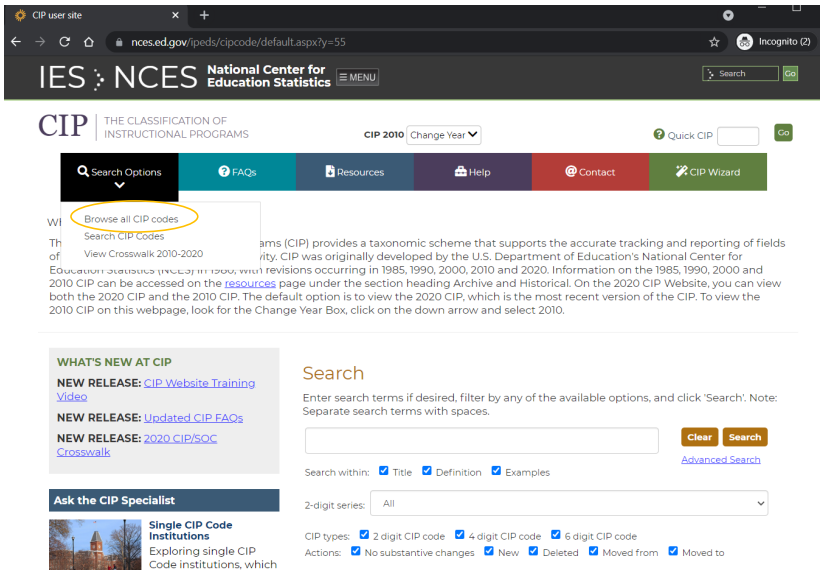
20013 = Spring 2002

mcid	institution	term	cip6	level	hours_ term
MID25836044	Institution M	20001	140102	01 Freshman	11
MID25836044	Institution M	20003	140102	01 Freshman	15
MID25836044	Institution M	20011	140801	02 Sophomore	15
MID25836044	Institution M	20013	140801	02 Sophomore	12
MID25836044	Institution M	20021	140801	03 Junior	16
MID25836044	Institution M	20023	140801	03 Junior	16
MID25836044	Institution M	20025	140801	04 Senior	7
MID25836044	Institution M	20031	140801	04 Senior	15
MID25836044	Institution M	20033	140801	04 Senior	15

Additional variables
standing
hours_cumul
hours_cumul_ attempt
gpa_term
gpa_cumul

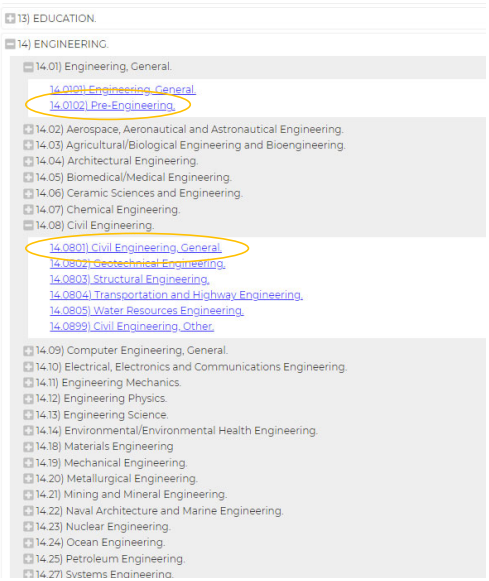
cip6





CIP

term	cip6
20001	140102
20003	140102
20011	140801
20013	140801
20021	140801
20023	140801
20025	140801
20031	140801
20033	140801



student term **course** degree

mcid	institution	term	course	abbrev	number	grade
MID25836044	Institution M	20001		CH	101	B
MID25836044	Institution M	20001		ENGL	102	B
MID25836044	Institution M	20001		ENGR	101	A
MID25836044	Institution M	20001		MTHSC	106	B
MID25836044	Institution M	20001		SOC	201	B
MID25836044	Institution M	20003		CH	102	CR
MID25836044	Institution M	**20003**		**CP SC**	**120**	**A**
MID25836044	Institution M	20003		CU	101	A
MID25836044	Institution M	20003		ENGR	120	B
MID25836044	Institution M	20003		MTHSC	108	A
MID25836044	Institution M	20003		PHYS	122	A
MID25836044	Institution M	20011		C E	251	A
MID25836044	Institution M	20011		C E	253	A
MID25836044	Institution M	20011		C E	255	B
MID25836044	Institution M	**20011**		**E M**	**201**	**D**
MID25836044	Institution M	20011		MTHSC	206	B
MID25836044	Institution M	20013		A A H	210	B
MID25836044	Institution M	20013		E G	209	C
MID25836044	Institution M	20013		E M	201	B
MID25836044	Institution M	20013		E M	202	CR
MID25836044	Institution M	20013		MTHSC	208	B
MID25836044	Institution M	20013		PHYS	221	B

mcid	institution	term	course	abbrev	number	grade
MID25836044	Institution M	20021		C E	200	C
MID25836044	Institution M	20021		C E	351	B
MID25836044	Institution M	20021		ECON	211	B
MID25836044	Institution M	20021		ENGL	314	A
MID25836044	Institution M	20021		EX ST	301	A
MID25836044	Institution M	20023		**C E**	301	B
MID25836044	Institution M	20023		**C E**	311	A
MID25836044	Institution M	20023		**C E**	331	C
MID25836044	Institution M	20023		**C E**	341	B
MID25836044	Institution M	20023		**C E**	352	B
MID25836044	Institution M	20023		**C E**	353	A
MID25836044	Institution M	20025		C E	321	B
MID25836044	Institution M	20025		SPCH	250	B
MID25836044	Institution M	20031		C E	342	B
MID25836044	Institution M	20031		C E	406	C
MID25836044	Institution M	20031		C E	421	A
MID25836044	Institution M	20031		ECON	212	B
MID25836044	Institution M	20031		EE&S	401	B
MID25836044	Institution M	20033		C E	402	B
MID25836044	Institution M	20033		C E	433	B
MID25836044	Institution M	20033		C E	459	A
MID25836044	Institution M	20033		ENGL	206	B

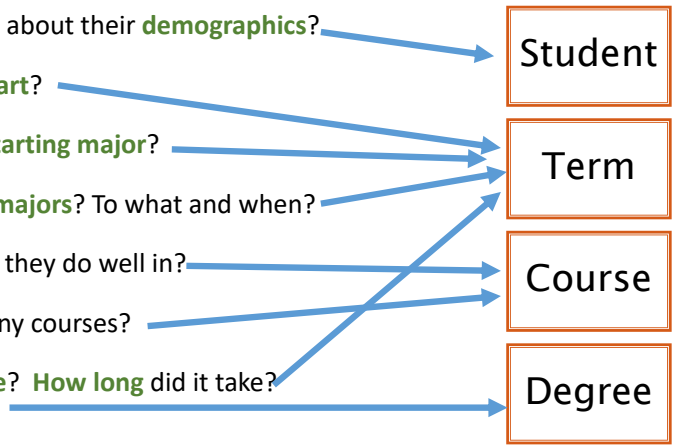
Additional variables
section
type
faculty_rank
hours_course
pass_fail

student term course **degree**

mcid	institution	term	cip6	degree
MID25836044	Institution M	20033	140801	Bachelor of Science

Take a few minutes to explore Example 2 (MID26035311). 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 repeat any courses?
7. Did they **graduate**? **How long** did it take?



Check your results for Example 2.

1. What can you tell about their **demographics**? **Hispanic/Latinx male, FTIC**
2. When did they **start**? **20041 -> Fall 2004**
3. What was their **starting major**? **140701 -> Chemical Engineering**
4. Did they **change majors**? If so, to what and when? **Yes. 141001 -> Electrical Engineering in Spring of sophomore year**
5. What courses did they do well in? **Senior Design, microecon, intro to CS**
6. Did they repeat any courses? **Calculus III , digital circuits, phys 1**
7. Did they **graduate**? **How long** did it take them? **Yes. 20084 -> Summer 2009, so 5 years (Fall to Summer is one year). Side note- the student did not take any undergrad courses in the last year**

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/2022-midfield-institute/agenda.html#wednesday-august-3)
- Please check in with us before leaving the virtual meeting.
  - How far did you get?
  - What are you thinking of using MIDFIELD for?