

# An Automated Solution to the Financial Aid Section of the Common Data Set

Ryan M. Johnson, Ph.D.  
Bryce Mason, Ph.D.

# Goals for Today

- Covered in today's presentation:
  - A brief overview of our CDS section H1/H2 script
  - A brief overview of how to set up the data to run that script
  - A detailed description of the algorithm that determines non-need and need-based (+ used-to-meet-need) aid and creates the table for H1
- Covered in handout (available through e-mail):
  - Code we use to complete ALL questions of CDS section H
  - Leave your card after the talk or send us an e-mail to receive a copy of this file

# Our Systems and Process

- Systems/Software
  - Banner Database
  - SQL Developer
  - Stata version 11.2
    - Even if you use a different statistical package, you may still benefit from seeing our algorithm
- Our Process
  - Pull financial aid data from relevant Banner tables using SQL Developer and process into two master data files:
    - A file for CDS questions H1-H2, H6
    - A file for CDS questions H4-H5 (we will not cover this today)
    - We will not cover SQL or data merging today
  - Run the CDS script in Stata for section H to process the master data and generate results for all questions.

- By creating the data set below, one can answer CDS questions H1, H2, and H6 in Section H at the push of a button.
- Data set 1 (scope):
  - Each row is a payment for a degree-seeking undergraduate present on Fall census day
  - Students who received no aid in that year have a single row with no aid in it.

# Defining Variables for Data Set 1

- **Pidm** – unique id for each student
- **Key Indicators** – 0/1 indicator variables with 1's representing new freshmen, full-time status, non-resident aliens, and applied for/paid need-based aid
- **Paid** – amount actually paid to a student for that specific fund
- **Fund code** – an identifier in each row in which paid>0 that is unique to specific funds.
- **Type** – a code assigned by our financial aid office to break up funds into meaningful categories (e.g., LOAN, WORK, GRANT)
- **Source** – a code assigned by our financial aid office to identify who is actually paying the money for each fund (e.g., FED, STATE, ENDOWMENT, etc.)
- **Need** – we define it as the maximum need for each student based on the following possible calculations of need:
  - Federal method
  - Institutional method
  - Actual paid aid that is purely need based

# Data Structure

- Data set 1 example:

pidm	new fresh.	full time	nra	need- based aid app.	need	fundcode	type	source	paid
150042	0	0	0	0	\$0	REM	REM	REM	\$4,775
635689	0	1	0	1	\$22,801	W11400	WORK	FED	\$1,680
635689	0	1	0	1	\$22,801	L11522	LOAN	FED	\$2,709
635689	0	1	0	1	\$22,801	L11519	PLUS	FED	\$8,160
635689	0	1	0	1	\$22,801	L11523	NLOA	FED	\$985
635689	0	1	0	1	\$22,801	G31013	GRNT	INST	\$5,305
580122	0	1	1	0	\$0				\$0

- Every row must contain an ID for someone who was present on census day of that fall

# Data Structure

- Data set 1 example:

pidm	new fresh.	full time	nra	need- based aid app.	need	fundcode	type	source	paid
150042	0	0	0	0	\$0	REM	REM	REM	\$4,775
635689	0	1	0	1	\$22,801	W11400	WORK	FED	\$1,680
635689	0	1	0	1	\$22,801	L11522	LOAN	FED	\$2,709
635689	0	1	0	1	\$22,801	L11519	PLUS	FED	\$8,160
635689	0	1	0	1	\$22,801	L11523	NLOA	FED	\$985
635689	0	1	0	1	\$22,801	G31013	GRNT	INST	\$5,305
580122	0	1	1	0	\$0				\$0

- Every row must contain an ID for someone who was present on census day of that fall

# Data Structure

- Data set 1 example:

pidm	new fresh.	full time	nra	need- based aid app.	need	fundcode	type	source	paid
150042	0	0	0	0	\$0	REM	REM	REM	\$4,775
635689	0	1	0	1	\$22,801	W11400	WORK	FED	\$1,680
635689	0	1	0	1	\$22,801	L11522	LOAN	FED	\$2,709
635689	0	1	0	1	\$22,801	L11519	PLUS	FED	\$8,160
635689	0	1	0	1	\$22,801	L11523	NLOA	FED	\$985
635689	0	1	0	1	\$22,801	G31013	GRNT	INST	\$5,305
580122	0	1	1	0	\$0				\$0

- Every row must contain an ID for someone who was present on census day of that fall



- Data set 1 example:

pidm	new fresh.	full time	nra	need- based aid app.	need	fundcode	type	source	paid
150042	0	0	0	0	\$0	REM	REM	REM	\$4,775
635689	0	1	0	1	\$22,801	W11400	WORK	FED	\$1,680
635689	0	1	0	1	\$22,801	L11522	LOAN	FED	\$2,709
635689	0	1	0	1	\$22,801	L11519	PLUS	FED	\$8,160
635689	0	1	0	1	\$22,801	L11523	NLOA	FED	\$985
635689	0	1	0	1	\$22,801	G31013	GRNT	INST	\$5,305
580122	0	1	1	0	\$0				\$0

- Every row must contain an ID for someone who was present on census day of that fall

## Need Algorithm (CDS H1)

- The CDS requires that aid used to meet need be counted as need-based aid.
- This can be time consuming to determine by hand for each student, so our algorithm handles all records at once.
- Before the algorithm can function, it is necessary to code type and source for all funds that are paid to students. This should be done at the database level.

## Need Algorithm: Steps the Script Takes

- The following steps are executed by the script, but a careful eye must be used in proofing the script for the step below to ensure that funds are categorized correctly.
- Assign a priority value to every row in data set 1 based on the concepts below.
  1. purely need-based funds
  2. non-need-based institutional scholarships
  3. non-need-based tuition waivers
  4. non-need-based athletic awards
  5. non-need-based external scholarships
  6. non-need-based student loans
  7. non-need-based parent loans
  8. non-need-based work

## Need Algorithm: Steps the Script Takes

- Ensure that every row in the data set has a priority value attached
- Sort the data by pidm and priority
- Generate a running total of amount paid to each person starting with the first record
- Generate and replace a new variable called paid\_need for each person one row at a time until the running total meets need, then fill in the rest of the paid amounts in a new variable called paid\_non\_need

# Need Algorithm: Steps the Script Takes

pidm

469214
469214
469214
469214
469214
469214
469214
469214
469214
469214
469214
469214
469214
469214
469214

# Need Algorithm: Steps the Script Takes

pidm      need

469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079
469214	\$43,079

# Need Algorithm: Steps the Script Takes

pidm      need      type

469214	\$43,079	LOAN
469214	\$43,079	GRNT
469214	\$43,079	LOAN
469214	\$43,079	WORK
469214	\$43,079	LOAN
469214	\$43,079	WORK
469214	\$43,079	LOAN
469214	\$43,079	GRNT
469214	\$43,079	NLOA
469214	\$43,079	NLOA
469214	\$43,079	PLUS
469214	\$43,079	PLUS
469214	\$43,079	NWOR
469214	\$43,079	NWOR

# Need Algorithm: Steps the Script Takes

pidm	need	type	source
469214	\$43,079	LOAN	FED
469214	\$43,079	GRNT	INST
469214	\$43,079	LOAN	FED
469214	\$43,079	WORK	FED
469214	\$43,079	LOAN	FED
469214	\$43,079	WORK	FED
469214	\$43,079	LOAN	FED
469214	\$43,079	GRNT	INST
469214	\$43,079	NLOA	FED
469214	\$43,079	NLOA	FED
469214	\$43,079	PLUS	FED
469214	\$43,079	PLUS	FED
469214	\$43,079	NWOR	INST
469214	\$43,079	NWOR	INST



## Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid
469214	\$43,079	LOAN	FED	\$1,000
469214	\$43,079	GRNT	INST	\$7,600
469214	\$43,079	LOAN	FED	\$1,000
469214	\$43,079	WORK	FED	\$1,900
469214	\$43,079	LOAN	FED	\$2,032
469214	\$43,079	WORK	FED	\$1,502
469214	\$43,079	LOAN	FED	\$2,032
469214	\$43,079	GRNT	INST	\$7,600
469214	\$43,079	NLOA	FED	\$1,662
469214	\$43,079	NLOA	FED	\$1,662
469214	\$43,079	PLUS	FED	\$11,688
469214	\$43,079	PLUS	FED	\$11,688
469214	\$43,079	NWOR	INST	\$1,591
469214	\$43,079	NWOR	INST	\$3,937

## Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority
469214	\$43,079	LOAN	FED	\$1,000	1
469214	\$43,079	GRNT	INST	\$7,600	1
469214	\$43,079	LOAN	FED	\$1,000	1
469214	\$43,079	WORK	FED	\$1,900	1
469214	\$43,079	LOAN	FED	\$2,032	1
469214	\$43,079	WORK	FED	\$1,502	1
469214	\$43,079	LOAN	FED	\$2,032	1
469214	\$43,079	GRNT	INST	\$7,600	1
469214	\$43,079	NLOA	FED	\$1,662	6
469214	\$43,079	NLOA	FED	\$1,662	6
469214	\$43,079	PLUS	FED	\$11,688	7
469214	\$43,079	PLUS	FED	\$11,688	7
469214	\$43,079	NWOR	INST	\$1,591	8
469214	\$43,079	NWOR	INST	\$3,937	8

## Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531
469214	\$43,079	WORK	FED	\$1,502	1	\$15,033
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677
469214	\$43,079	PLUS	FED	\$11,688	7	\$51,365
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894

# Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total	paid need
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000	\$1,000
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600	\$7,600
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600	\$1,000
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500	\$1,900
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531	\$2,032
469214	\$43,079	WORK	FED	\$1,502	1	\$15,033	\$1,502
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065	\$2,032
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665	\$7,600
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327	\$1,662
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989	\$1,662
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677	\$11,688
469214	\$43,079	PLUS	FED	\$11,688	7	\$51,365	\$3,402
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956	\$0
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894	\$0

## Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total	paid need
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000	\$1,000
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600	\$7,600
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600	\$1,000
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500	\$1,900
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531	\$2,032
469214	<b>\$43,079</b>	<b>WORK</b>	<b>FED</b>	<b>\$1,502</b>	<b>1</b>	<b>\$15,033</b>	<b>\$1,502</b>
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065	\$2,032
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665	\$7,600
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327	\$1,662
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989	\$1,662
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677	\$11,688
469214	\$43,079	PLUS	FED	\$11,688	7	\$51,365	\$3,402
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956	\$0
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894	\$0

# Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total	paid need
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000	\$1,000
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600	\$7,600
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600	\$1,000
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500	\$1,900
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531	\$2,032
<b>469214</b>	<b>\$43,079</b>	<b>WORK</b>	<b>FED</b>	<b>\$1,502</b>	<b>1</b>	<b>\$15,033</b>	<b>\$1,502</b>
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065	\$2,032
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665	\$7,600
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327	\$1,662
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989	\$1,662
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677	\$11,688
469214	\$43,079	PLUS	FED	\$11,688	7	\$51,365	\$3,402
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956	\$0
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894	\$0

## Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total	paid need
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000	\$1,000
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600	\$7,600
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600	\$1,000
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500	\$1,900
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531	\$2,032
469214	\$43,079	WORK	FED	\$1,502	1	\$15,033	\$1,502
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065	\$2,032
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665	\$7,600
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327	\$1,662
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989	\$1,662
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677	\$11,688
469214	<b>\$43,079</b>	<b>PLUS</b>	<b>FED</b>	<b>\$11,688</b>	<b>7</b>	<b>\$51,365</b>	<b>\$3,402</b>
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956	\$0
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894	\$0

## Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total	paid need
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000	\$1,000
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600	\$7,600
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600	\$1,000
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500	\$1,900
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531	\$2,032
469214	\$43,079	WORK	FED	\$1,502	1	\$15,033	\$1,502
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065	\$2,032
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665	\$7,600
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327	\$1,662
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989	\$1,662
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677	\$11,688
<b>469214</b>	<b>\$43,079</b>	<b>PLUS</b>	<b>FED</b>	<b>\$11,688</b>	<b>7</b>	<b>\$51,365</b>	<b>\$3,402</b>
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956	\$0
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894	\$0



# Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total	paid need	paid non need
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000	\$1,000	\$0
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600	\$7,600	\$0
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600	\$1,000	\$0
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500	\$1,900	\$0
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531	\$2,032	\$0
469214	\$43,079	WORK	FED	\$1,502	1	\$15,033	\$1,502	\$0
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065	\$2,032	\$0
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665	\$7,600	\$0
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327	\$1,662	\$0
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989	\$1,662	\$0
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677	\$11,688	\$0
469214	<b>\$43,079</b>	<b>PLUS</b>	<b>FED</b>	<b>\$11,688</b>	<b>7</b>	<b>\$51,365</b>	<b>\$3,402</b>	<b>\$8,286</b>
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956	\$0	\$1,591
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894	\$0	\$3,937

## Need Algorithm: Steps the Script Takes

pidm	need	type	source	paid	priority	running total	paid need	paid non need
469214	\$43,079	LOAN	FED	\$1,000	1	\$1,000	\$1,000	\$0
469214	\$43,079	GRNT	INST	\$7,600	1	\$8,600	\$7,600	\$0
469214	\$43,079	LOAN	FED	\$1,000	1	\$9,600	\$1,000	\$0
469214	\$43,079	WORK	FED	\$1,900	1	\$11,500	\$1,900	\$0
469214	\$43,079	LOAN	FED	\$2,032	1	\$13,531	\$2,032	\$0
469214	\$43,079	WORK	FED	\$1,502	1	\$15,033	\$1,502	\$0
469214	\$43,079	LOAN	FED	\$2,032	1	\$17,065	\$2,032	\$0
469214	\$43,079	GRNT	INST	\$7,600	1	\$24,665	\$7,600	\$0
469214	\$43,079	NLOA	FED	\$1,662	6	\$26,327	\$1,662	\$0
469214	\$43,079	NLOA	FED	\$1,662	6	\$27,989	\$1,662	\$0
469214	\$43,079	PLUS	FED	\$11,688	7	\$39,677	\$11,688	\$0
469214	\$43,079	PLUS	FED	\$11,688	7	\$51,365	\$3,402	\$8,286
469214	\$43,079	NWOR	INST	\$1,591	8	\$52,956	\$0	\$1,591
469214	\$43,079	NWOR	INST	\$3,937	8	\$56,894	\$0	\$3,937

# Need Algorithm: Steps the Script Takes

- Create a string variable that contains information about the category of aid corresponding to the categories in CDS H1.
  - Scholarships/Grants (S&G)
    - Federal
    - State
    - Institutional
    - External
  - Self-help (SH)
    - Student loans
    - Federal work study
    - Other work
  - Parent Loans (O)
  - Tuition Waivers (O)
  - Athletic Awards (O)

# Need Algorithm: Steps the Script Takes

- Map of CDS categories to priority categories

		Priority Category						
		Need-Based Funds (1)	Inst. Sch. (2)	Tuition Waivers (3)	Athletic Awards (4)	External Sch. (5)	Student Loans (6)	Parent Loans (7)
CDS Category	Fed. S&G	X						
	State S&G	X						
	Fed. Work Study	X						
	Inst. S&G	X	X					
	Student Loans	X					X	
	External S&G					X		
	Other Work							X
	Parent Loans						X	
	Tuition Waivers			X				
	Athletic Awards				X			

- After completing those steps, Stata has enough information in the data set in memory to display answers to all CDS questions except H4-H5
- The script finishes by loading data set 2 into memory and completing H4-H5 (not covered today).
- Now, let's run the script on real data and examine the code.

For a copy of the code, please e-mail  
ryan.johnson@lmu.edu or leave your business card with  
your e-mail address at the front.

Thank you.

LMU|LA

## Appendix for H4/H5: Data Structure

- The data set below is necessary for the script that completes CDS questions H4/H5
- Data set 2:
  - At least one row for each student who graduated that year
  - Each row is a different paid amount representing a type and source of fund received in **any year** for that student's financial aid history.
  - Students who received no aid during any year before graduation have a single row with no aid in it.

## Appendix for H4/H5: Defining Variables for Data Set 2

- Data set 2 contains a subset of variables from data set 1, and they are defined the same way.
- The critical difference between the two data sets is the records that are included. Data set 2 contains all financial aid records for the history of all graduating undergraduates in the current year, whereas data set 1 contains only financial aid records for the current year for degree-seeking undergraduates present on census day.
- For both data sets, there should be one and only one row for each student who is part of the overall list but was not paid any financial aid during the relevant time period. Paid should be set to 0 for these rows.



## Appendix for H4/H5: Data Structure

- Data set 2 example:

pidm	aidyear	term	fundcode	type	source	paid
593554	809	200830	W32103	NWOR	INST	610.5
593554	809	200910	W32103	NWOR	INST	1012
593554	910	200930	G31013	GRNT	INST	2280
593554	910	200930	W11400	WORK	FED	1551.2
593554	910	201010	G31013	GRNT	INST	2280
593554	910	201010	W11400	WORK	FED	1461.6
512714						

- Every row must contain an ID for someone who graduated in that year
- All financial aid records for those graduates must be included