HUDK 5053: FEATURE ENGINEERING STUDIO

10/4/16 8:52 AM

In the news

School districts question TEA's special education 'cap'

SEGUIN GAZETTE
BRINGING LIFE TO YOUR DOORSTEP SINCE 1888

California Governor Signs Bill to Disaggregate Asian-American Health Data

NBC NEWS

Variety of data considered when accrediting institutions

The Council of Regional Accrediting Commissions,

Physical Education Technology Market Posts an $8 Billion Revenue

UNIVERSITY HERALD

Why Women in Tech Might Consider Just Using Their Initials Online

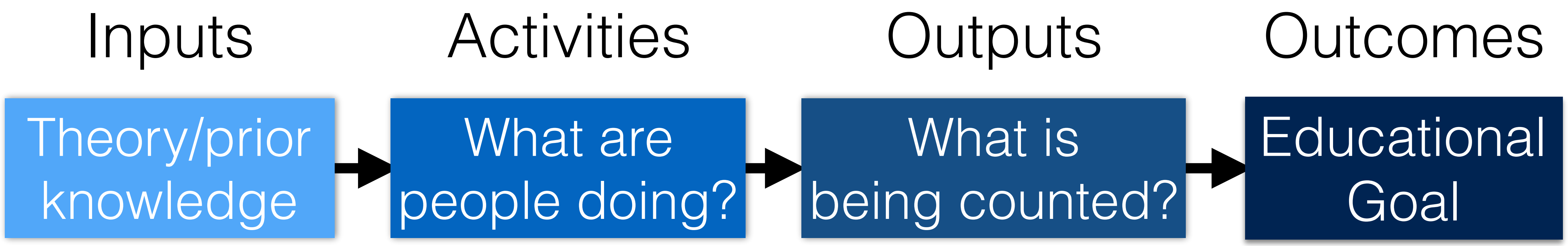
THE WALL STREET JOURNAL.

The University of Texas system is teaming up with Salesforce to make college courses more like Netflix

BUSINESS INSIDER

DML
DIGITAL MEDIA AND LEARNING CONFERENCE

Logic Model/Theory of Action



Inputs

Theory/prior knowledge

- Stake holders
- Domain
- Own perspective
- Participants perspectives

Activities

What are people doing?

- Environment
- Incentives

Outputs

What is being counted?

- What can be captured
- How granular
- What meaning is ascribed to it

Outcomes

Educational Goal

- Level of action
- Rules of inference

Topic in 2min

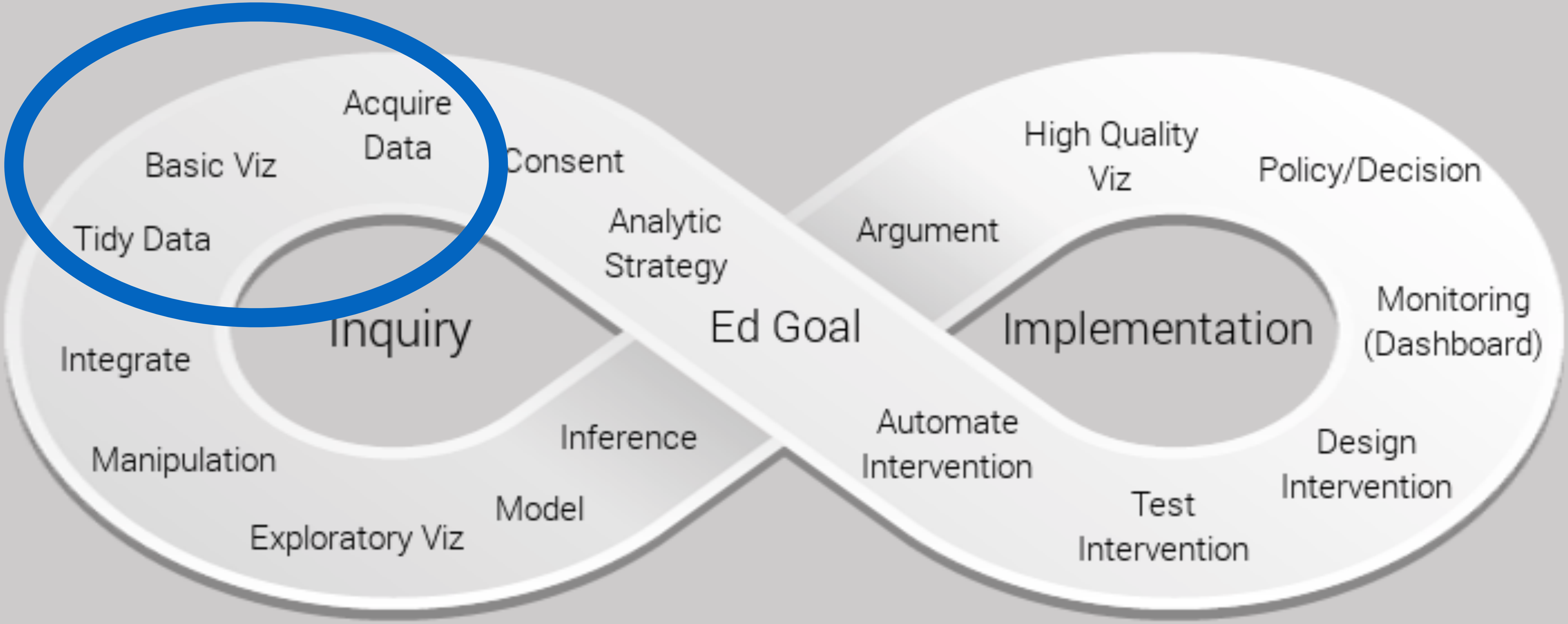
Did you make progress last Thursday?

Thursday Pairs

Tiantian Jin	Sunmin Li
Chad Coleman	Yaoli Mao
Shan Ding	Yan Xu
John Soul	Jiayi Zhang
Shiyue Zheng	Josh Coleman

Thursday Data Collection

Ed Data Science Cycle



Why is tidy data?

- Difference between “clean” and “tidy”
- Data comes in a lot different structures, some which are difficult to analyze
- We want to make them manageable
- We want them to be “intuitive” to R (vectorized)
- BUT we want to keep a very precise record of how we did that

What is tidy data?

1. Observations are in rows
2. Variables are in columns
3. In a single data set

But…?

- What is a variable?
- What is an observation?
- What goes where in a data matrix?

Generalize

Male	Female
4	10
7	10

How many variables are in the above matrix?

1. Male
2. Female
3. Count

Types of Messiness

- Column headers are values, not variable names
- Multiple variables are stored in one column
- Variables are stored in both rows and columns
- Multiple types of experimental unit stored in the same table
- One type of experimental unit stored in multiple tables