HIDK 4()5():

Today

- Data Wrangling Part II
- Ethics stuff we didn't get to last week

Events

Title	Date - Time	Location
Careers in Data Science in Ed	09/25 - 3:00pm	Everett Lounge
AWS: A Simplified Approach to Data <u>Driven Research</u>	09/30 - 9:00am	Webinar
Aligning Learning Analytics with Classroom Needs	10/01 - 12:00pm	NYU Kimmel 405
Columbia Curricular Innovation Fellows Info Session	10/01 - 12:00 10/03 - 4:00	Butler 203 Butler 523
Formal & Informal LA (lunch)	10/4 - 11:00am	GDH 449
Cornell Tech: Day of Data	10/15	Cornell Tech



News

Who Should Truly Have the Power in K-12 Edtech Adoption?

NewSchools Venture Fund and Gallup Release Survey Findings About Ed Tech Usage in U.S. PreK-12 Schools





Grant to Fund More Research into Ed Tech Best Practices



Millions of Americans' Medical Images and Data Are Available on the Internet. Anyone Can Take a Peek.

Wrangling Recap

- Tidy data
- Tidyverse (dplyr & tidyr)
- Wide/Long Format
- Messiness
- Subset/Variable Generation/Combine/Summarize
- Reshape: gather, spread
- Class Activity 1: answers available on Github

Data Wrangling II

Dataframes & Matrices

Matrix vs. Data Frame

Matrix

- Uses less memory
- Operations are faster
- Requires same data type (character or numeric)
- Useful for matrix algebra

Data Frame

- Convenient
- Intuitive
- Can have different data types in one format
- Useful for referring to columns individually

Create Matrix

- matrix()
- as.matrix()

Other Useful Operations

Transpose Function

- t()
- Transposes a matrix or data frame
- rows → columns, columns → rows
- Output = matrix

Diagonal Function

- diag()
- Replace or extract the diagonal of a matrix

Matrix Multiplication

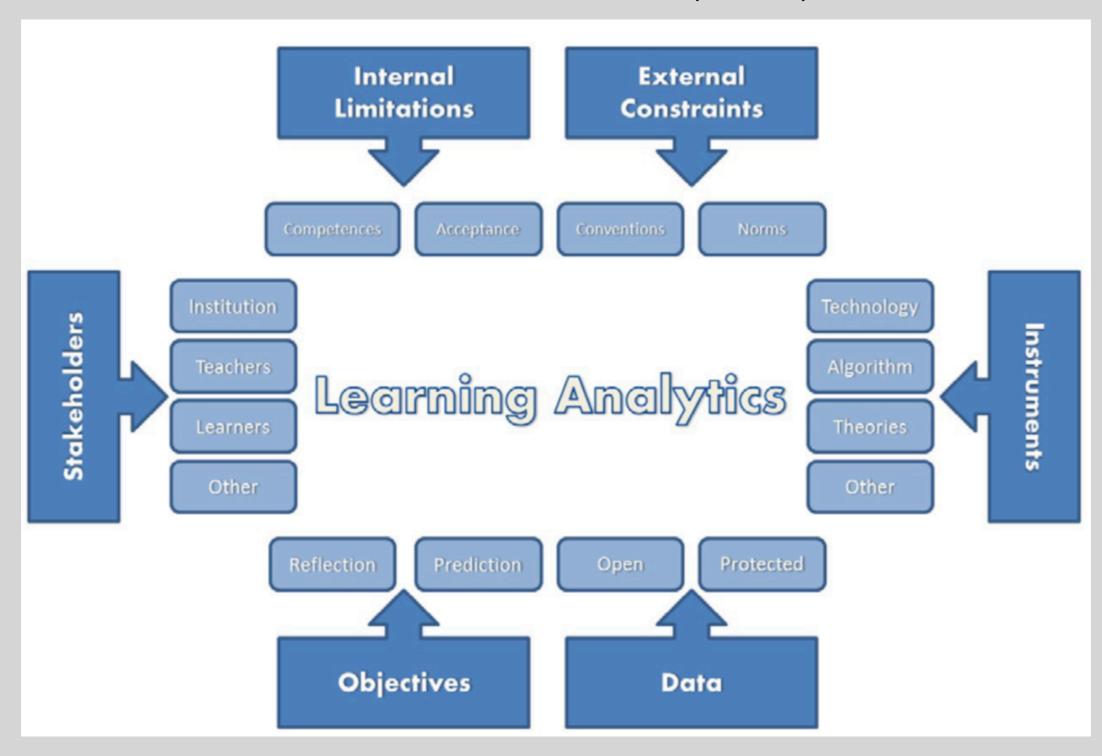
- %*%
- Two matrices must be the same size
- Multiplies the rows of the first matrix by the columns of the second
- Multiplies two matrices together
- Will become useful when we get to Social Network analysis

Activity

- Create a data frame called A of three variables, each having three values
- Convert the data frame to a matrix called B
- Create a matrix called C that is the transposition of A
- Create a matrix called **D** that is the multiplication of **C** and
 B
- Replace the diagonal values in **D** with missing values

Translating Learning Into Numbers

Greller & Draschler (2012)



Exercise 2

- Consider the fake data you generated and the diagram on page 44 of Greller & Draschler
- Work your way through each of the boxes
- Which would pose problems for you to actually acquire the data you want?
- Write a note summarizing the article in Zotero, include thoughts based on your answers to the above

Code of Ethics

- There have been several Learning Analytics Codes of Ethics drawn up for institutions:
 - Open University
 - JISC
 - American Library Association
 - Data for Good

D

DETERMINATION – Why you want to apply Learning Analytics?

- ▶ What is the added value (Organisational and data subjects)?
- What are the rights of the data subjects (e.g., EU Directive 95/46/EC)



EXPLAIN – Be open about your intentions and objectives

- What data will be collected for which purpose?
- How long will this data be stored?
- Who has access to the data?

L

LEGITIMATE – Why you are allowed to have the data?

- Which data sources you have already (aren't they enough)?
- Why are you allowed to collect additional data?

INVOLVE – Involve all stakeholders and the data subjects

- Be open about privacy concerns (of data subjects)
- Provide access to the personal data collected (about the data subjects)
- Training and qualification of staff

C

CONSENT – Make a contract with the data subjects

- Ask for a consent from the data subjects before the data collection
- Define clear and understandable consent questions (Yes / No options)
- Offer the possibility to opt-out of the data collection without consequences

A

ANONYMISE - Make the individual not retrievable

- Anonymise the data as far as possible
- Aggregate data to generate abstract metadata models (Those do not fall under EU Directive 95/46/EC)

T

TECHNICAL - Procedures to guarantee privacy

- Monitor regularly who has access to the data
- If the analytics change, update the privacy regulations (new consent needed)
- Make sure the data storage fulfills international security standards

E

EXTERNAL – If you work with external providers

- Make sure they also fulfil the national and organisational rules
- Sign a contract that clearly states responsibilities for data security
- ▶ Data should only be used for the intended services and no other purposes

Code of Ethics

bit.ly/HUDK4050COE

Exercise 3

- Read over the code
- Does it seem reasonable?
- Is there anything missing?
- Do you believe it is useful?

Anonymous Code of Ethics Survey

http://bit.ly/2w3GR51