EDCT GE2550: DATA SCIENCE IN EDUCATION

Big Data, Learning Analytics & The Information Age

2/25/16 12:48 PM

In the news

newschools ignite

A New, No-Strings Funding Opportunity for Ed Tech Startups Focused on Middle and High Schools

NMC Horizon Report > 2016 Higher Education Edition

http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf

How Data Can Protect You From Cognitive Bias

Join us for the latest DSC Webinar on March 15th, 2016

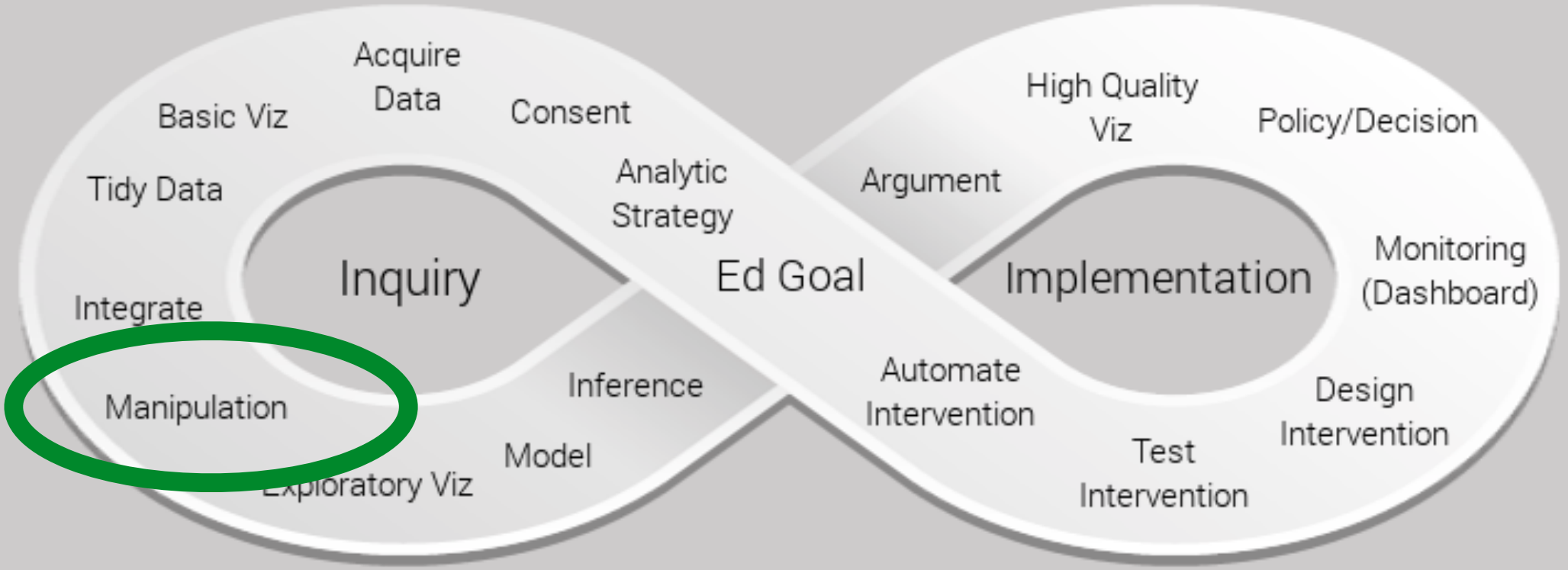
REGISTER NOW

tableau

Today

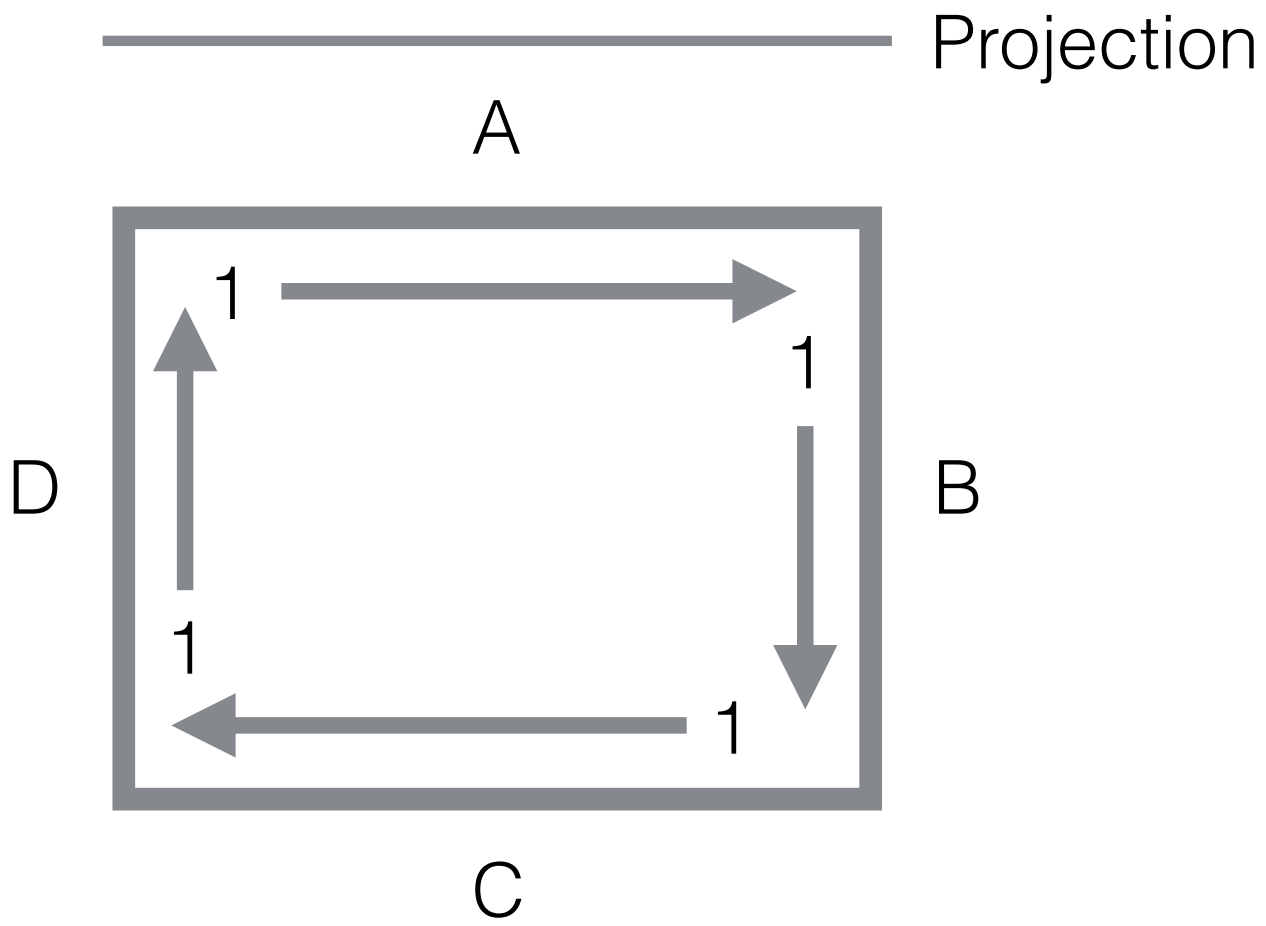
In the news	6:45 - 6:50
Quiz	6:50 - 7:00
Zotero reset	7:00 - 7:10
Zotero	7:10 - 7:20
Web traffic	7:20 - 7:30
Dimensionality	7:30 - 7:40
Twitter	7:40 - 8:10
Q & A	8:10 - 8:25

Ed Data Science Cycle



Quiz

http://bit.ly/1SzEPC3



Zotero

2/25/16 12:48 PM

Reset Library

- Turn off auto sync:  -> preferences -> sync -> uncheck “Sync automatically”
- Locate Zotero data file: preferences -> advanced -> File & Folders -> Show data directory
- Close Zotero standalone or Firefox
- Delete the Zotero folder you located
- Go to Github data-science-in-ed/Tools -> Download Zip
- Unpack the folder
- Re-open Zotero and go to  -> Import -> Locate the dse-references.df
- Now you should have a more manageable library

Index your code!

- Create a new folder called “R code”
- Create a new “computer program” item
- Give it the title “Data tidying”
- Right/Control click on the item in the list
- Go to “add attachment” -> “attach stored copy of file” -> Locate your Assignment 3 R markdown document (EG - assignment3.Rmd)
- Go back and fill in as many other descriptors as you can
- Repeat for your Twitter code and your Assignment 2 code

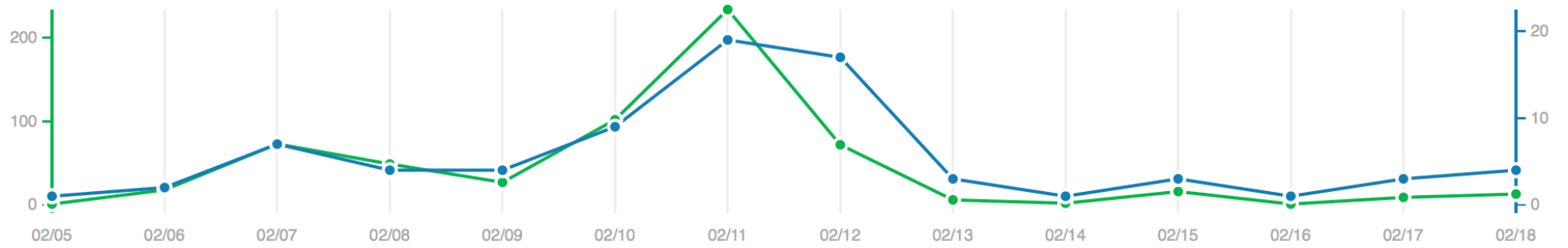
Web Traffic

Can web traffic tell us anything useful?

2/25/16 12:48 PM

Assignment 2 Web Traffic

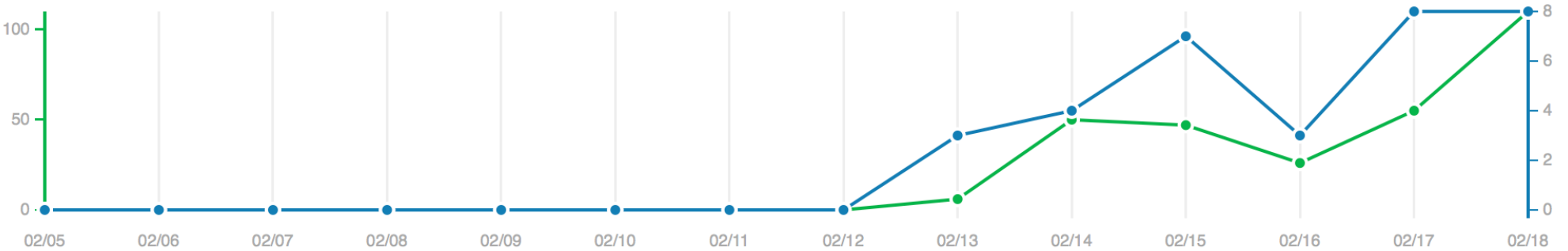
Visitors



623	23
Views	Unique visitors

Assignment 3 Web Traffic

Visitors



294	21
Views	Unique visitors

Personalization

2/25/16 12:48 PM

"Uniformity is the curse of American schools...Individual instruction is the new ideal."

Charles Eliot, 1899

Vocabulary

Individualization: learning goals are the same for all students, but students can progress through the material at different speeds.

Differentiation: learning goals are the same for all students, but the method or approach of instruction varies according to the preferences of each student.

Personalization: learning goals and content as well as the method and pace may all vary (so personalization encompasses differentiation and individualization).

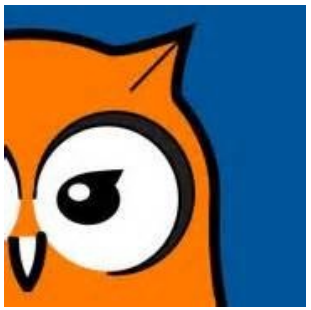
ed.gov (2010)

Adaptive

- Originally = assistive
- ~1990s = sequential estimate of aptitude (IRT)
- ~2012 = a system that adapts the educational environment according to students' learning needs
- Distinct from Intelligent Tutors in terms of methods employed

Adaptive Systems

Adaptive Engines



Recommender Systems

Collaborative filter: build a model from a user's past behavior + similar decisions made by other users

Content filter: utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties

Essential Problem

- Dimensionality Reduction
  - Feature selection: select a subset of dimensions
  - Feature extraction: transform lots of dimensions into fewer dimensions
- Why?
  - As a form of insight
  - Avoid “Curse of Dimensionality”

Curse of Dimensionality

Sparsity: The more dimensions that we add, the more comparisons we are missing

	Stats	Cog Psy
Amy	3	2
Chen	2	2
Asif	1	3

Possible Combinations

3 - 3
3 - 2
3 - 1
2 - 3
2 - 2
2 - 1
1 - 3
1 - 2
1 - 1

Curse of Dimensionality

Sparsity: The more dimensions that we add, the more comparisons we are missing

	Stats	Cog Psy	Socio-logy	Crit Theory	Wood-work	Data Sci	Music	Design
Amy	3	2	1	1	3	2	2	2
Chen	2	2	2	3	1	3	2	3
Asif	1	3	3	7	3	2	1	1

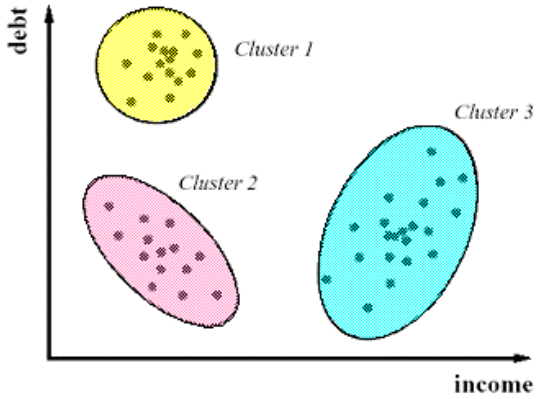
How to distill features?

Mean, median, mode

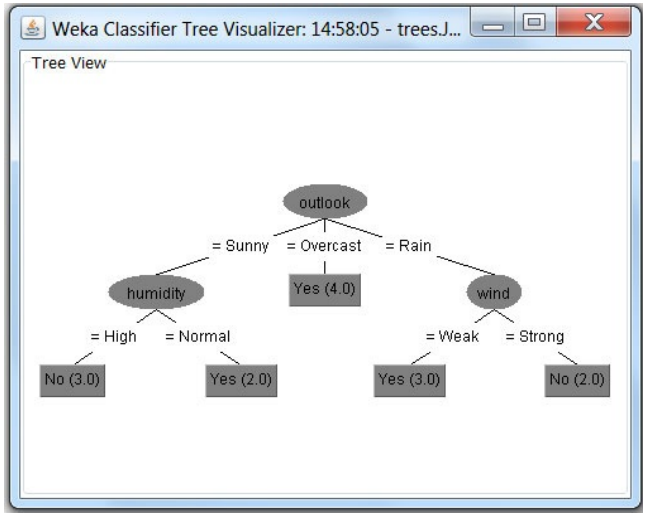
Principle Component Analysis

Google

Cluster Analysis



Decision Tree



Correlations

- Pearson Correlation Coefficient (r)
- Measure of the strength and direction of association between two variables
- Need linear association
- Doesn't equal causation

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$