HUDK 4050:CORE METHODS IN EDM

10/18/16 8:59 AM

In the news

Why Mark Zuckerberg Thinks An Indian EdTech Startup Is Worth Millions

BYJU'S Forbes / Asia

Why Rafter Failed

RAFTER

EdSurgeNews

State ed tech leaders argue teachers need more training support

edscoop

New report raises questions about edtech funding in ESSA

Why Has Ed Tech Made So Little Difference?

EDUCATION WEEK

Wharton SEI Center

QS STARS AWARDS 2016 REIMAGINE EDUCATION

Edtech: Loopholes Lead to Cracks in Google's Student Privacy Policy

Real Money

McGraw-Hill Education acquires fourth edtech company since 2013

College Rankings

Stanford took the top spot in our inaugural ranking of U.S. colleges, followed by MIT and Columbia University. Analyze the data and compare schools to help find

INSIGHT

January Fellowship Applications due in 10 days

The next Insight Data Science Fellows programs start January 17th, 2017 for the New York, Silicon Valley, and the Remote Program - applications are due on Monday, October 24th.

APPLY

http://insightdatascience.com/

Attitude

SNA Solution

* - Zhuqian

Adaptive Systems

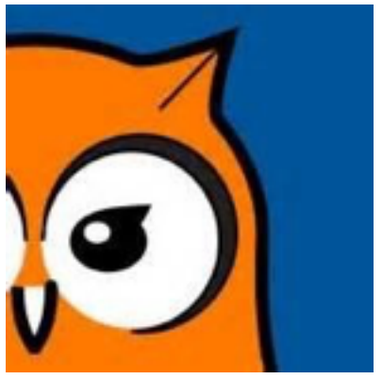
Adaptive

- 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

KNEWTON



Grockit

prepme

KAPLAN

adapt courseware

gymglish

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

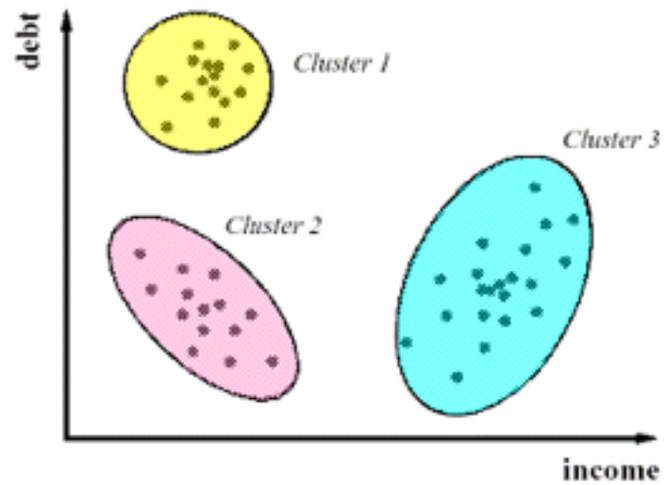
How to distill features?

Mean, median, mode

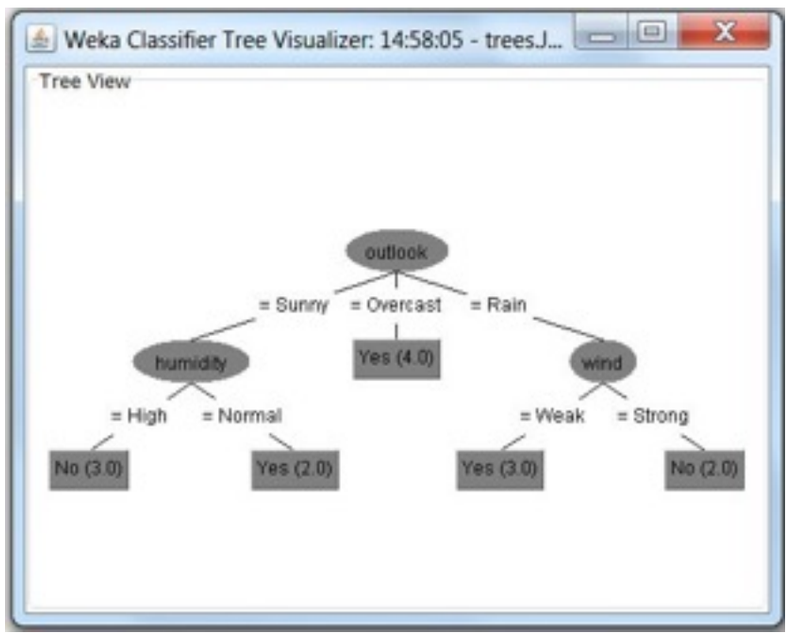
Principle Component Analysis

Google

Cluster Analysis



Decision Tree



Dimensionality Reduction

Cluster Analysis

Grouping stuff

By Variables

ID	Var1	Var2	Var3
A			
B			
C			
D			

ID	Var2
A	
B	
C	
D	

Selection

ID	Var2+3
A	
B	
C	
D	

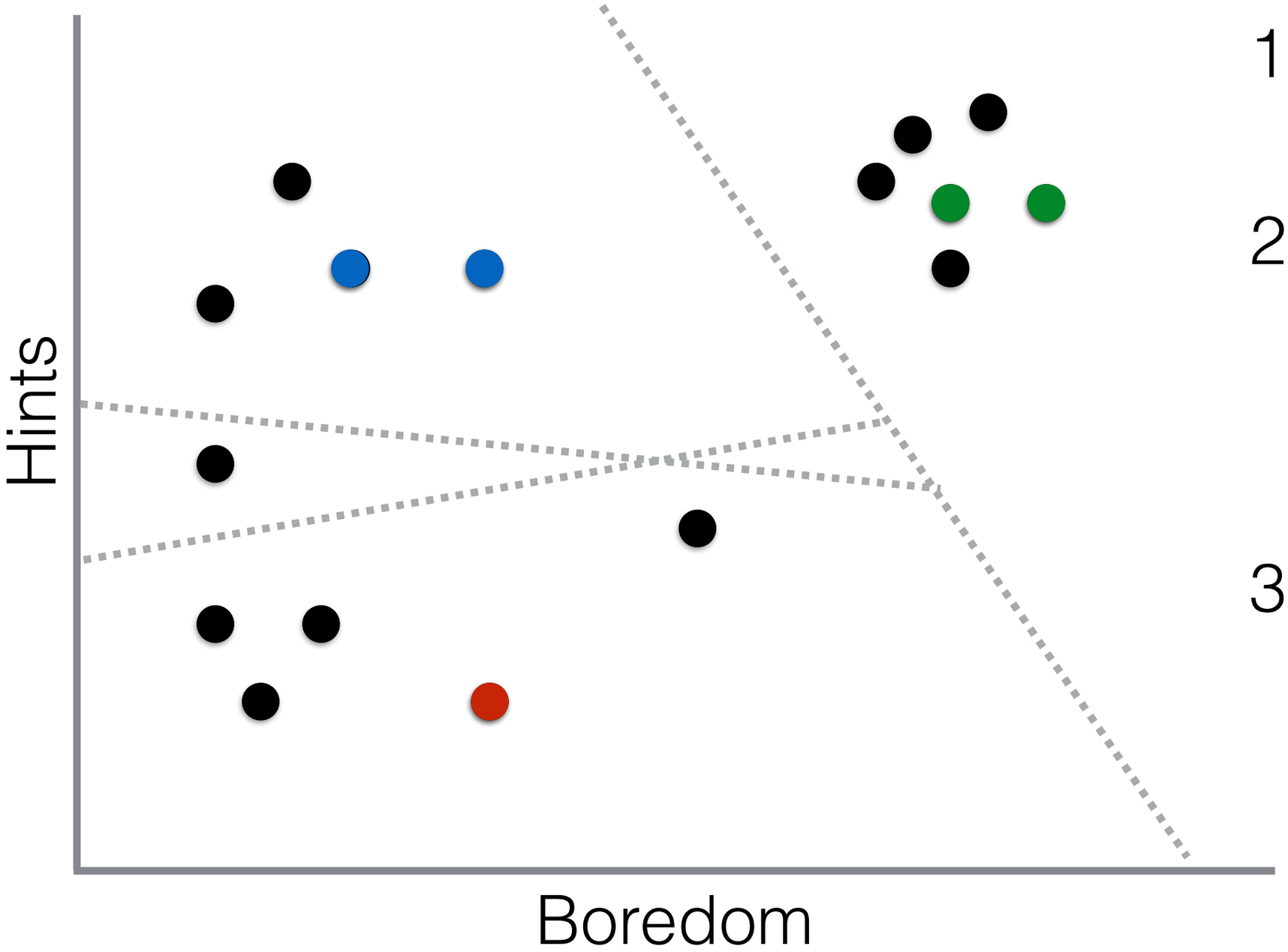
Extraction

By People

ID	Var1	Var2	Var3
A			
C			

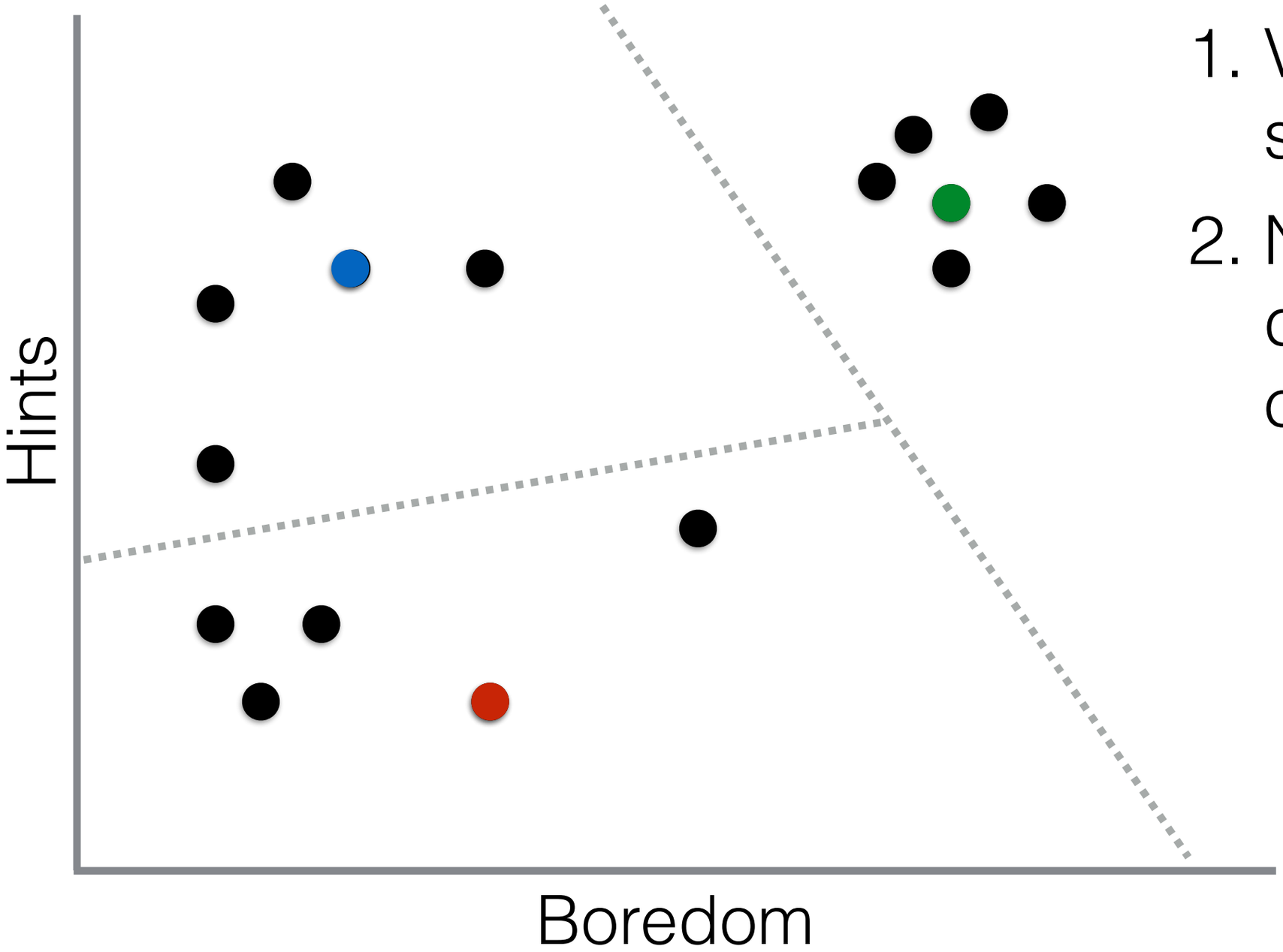
ID	Var1	Var2	Var3
B			
D			

Cluster Analysis: K-means

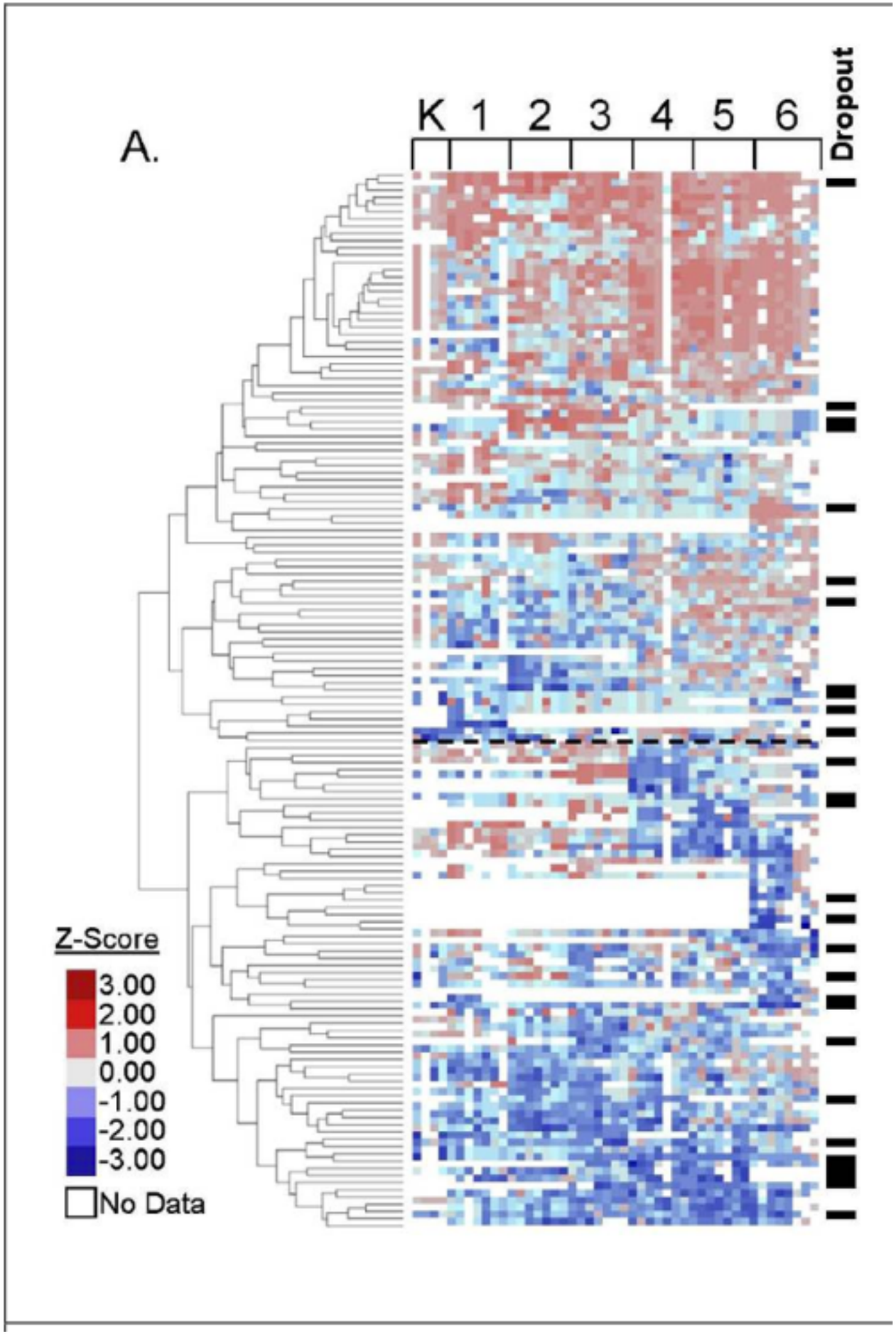


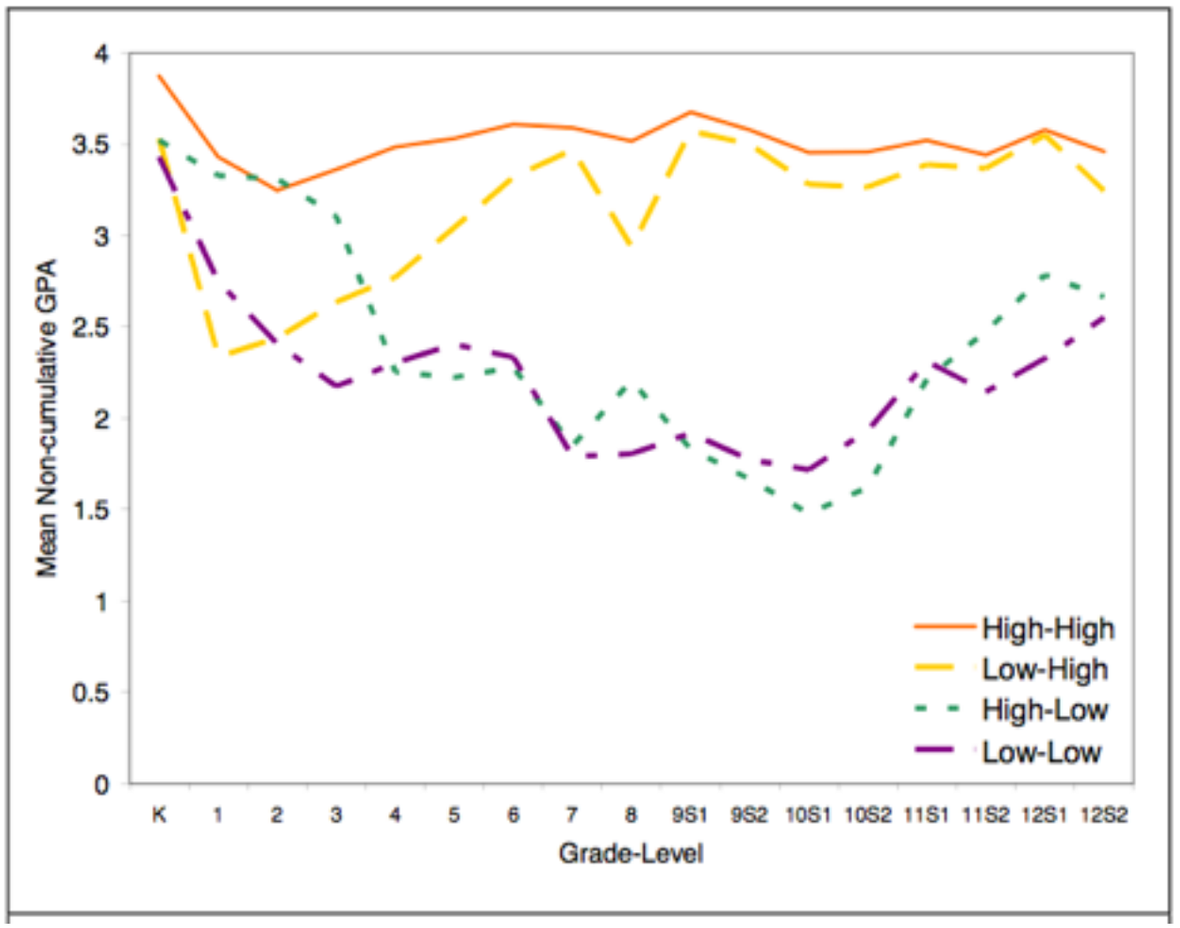
1. Select some random points
2. Associate those points with closest other points
3. Move the selected point to the mean point in the cluster

Cluster Analysis: K-means



1. Very sensitive to starting values
2. Not good at dealing with complex shapes


A.
K 1 2 3 4 5 6
Dropout
Z-Score
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
No Data


Mean Non-cumulative GPA
4
3.5
3
2.5
2
1.5
1
0.5
0
K 1 2 3 4 5 6 7 8 9S1 9S2 10S1 10S2 11S1 11S2 12S1 12S2
Grade-Level
High-High
Low-High
High-Low
Low-Low

Bowers (2010)

Grouping stuff

By Variables

ID	Var1	Var2	Var3
A			
B			
C			
D			

ID	Var2
A	
B	
C	
D	

Selection

ID	Var2+3
A	
B	
C	
D	

Extraction

By People

ID	Var1	Var2	Var3
A			
C			

ID	Var1	Var2	Var3
B			
D			

Feature Extraction

- Principal Component Analysis
- Variance
- Covariance
- Matrix algebra

Process

1. Describe data

2.Choose methodology

3.Make notes on the purpose of the method

4.Make notes on the limitations of the method (gotchas)

5.Research the code

6.Apply the code

Cluster Example