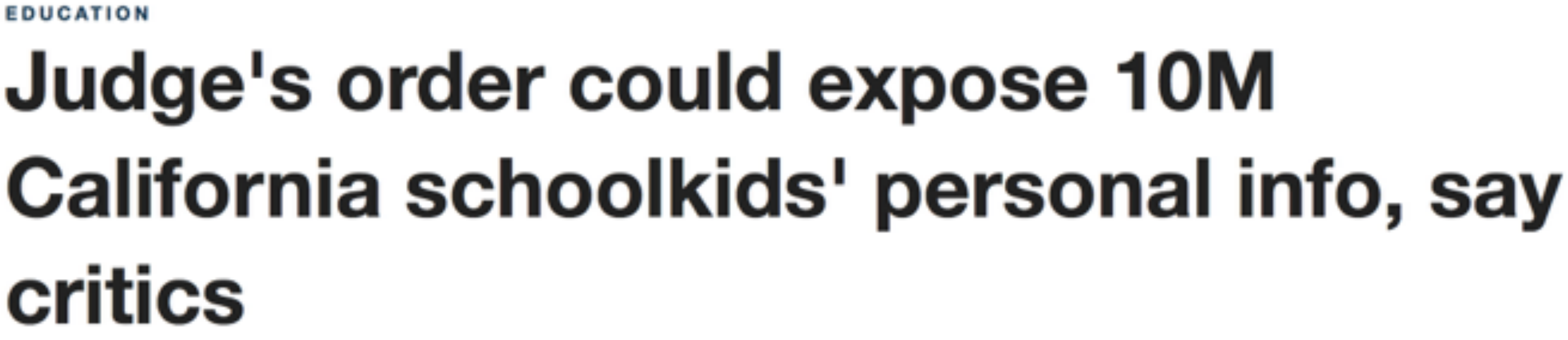
EDCT GE2550: DATA SCIENCE IN EDUCATION

Big Data, Learning Analytics & The Information Age

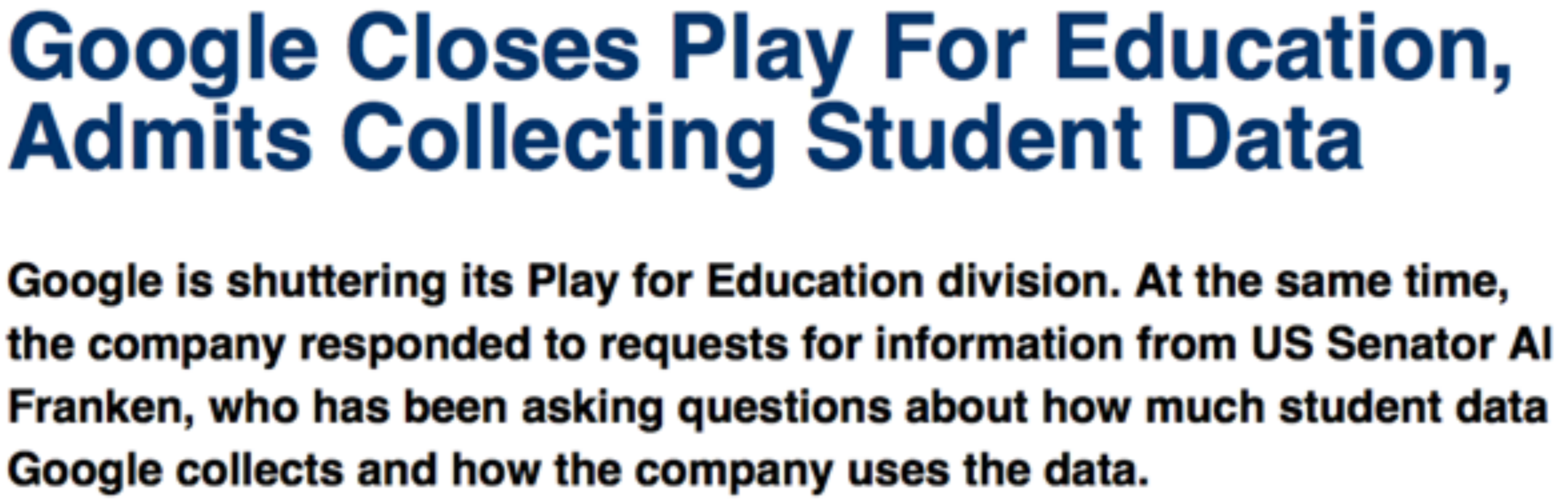
2/27/16 7:15 PM

In the news



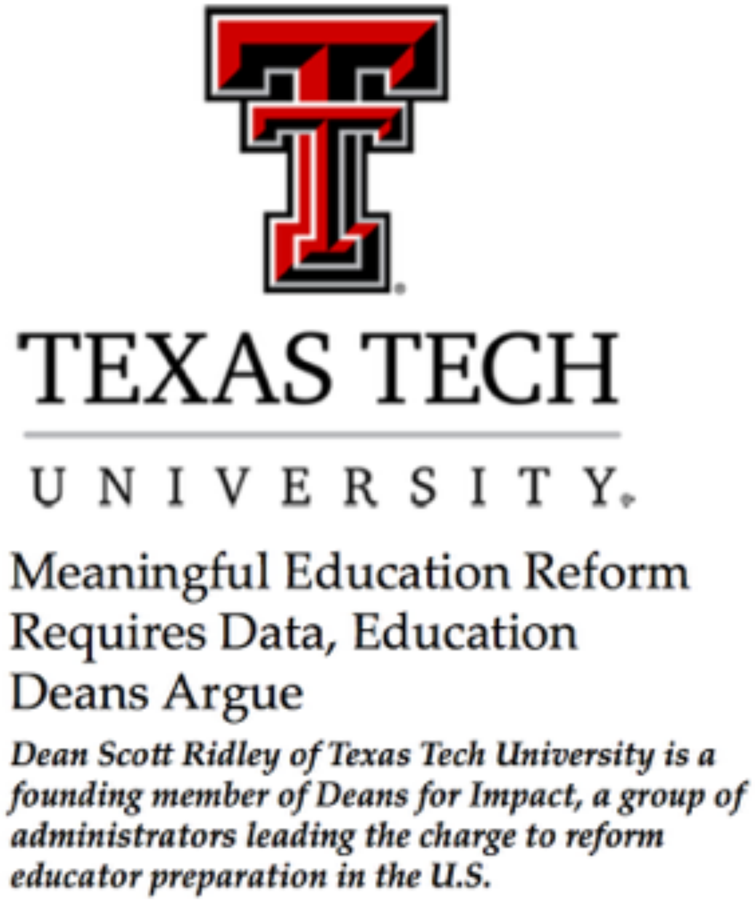
EDUCATION

Judge's order could expose 10M California schoolkids' personal info, say critics



Google Closes Play For Education, Admits Collecting Student Data

Google is shuttering its Play for Education division. At the same time, the company responded to requests for information from US Senator Al Franken, who has been asking questions about how much student data Google collects and how the company uses the data.



TEXAS TECH
UNIVERSITY.

Meaningful Education Reform Requires Data, Education Deans Argue

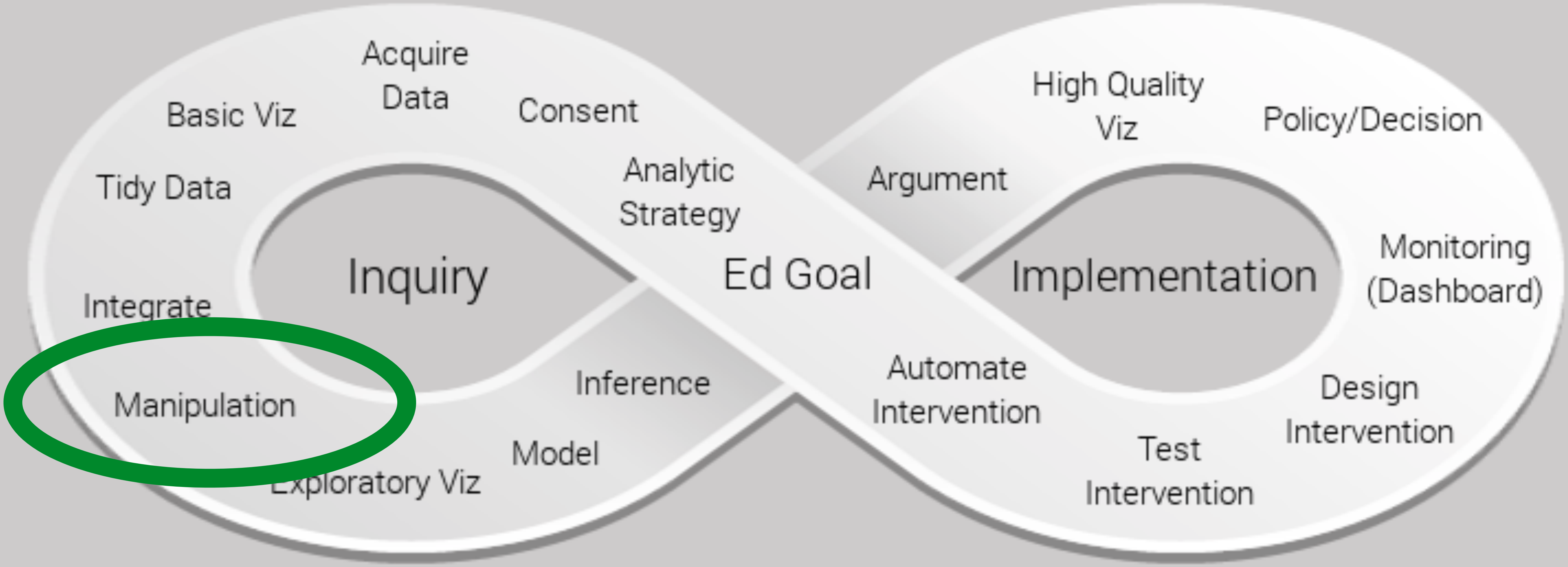
Dean Scott Ridley of Texas Tech University is a founding member of Deans for Impact, a group of administrators leading the charge to reform educator preparation in the U.S.

Today

In the news	6:45 - 6:50
Quiz	6:50 - 7:00
Open Data	7:00 - 7:10
Twitter I	7:10 - 7:40
PCA	7:40 - 7:50
Twitter II	7:50 - 8:20

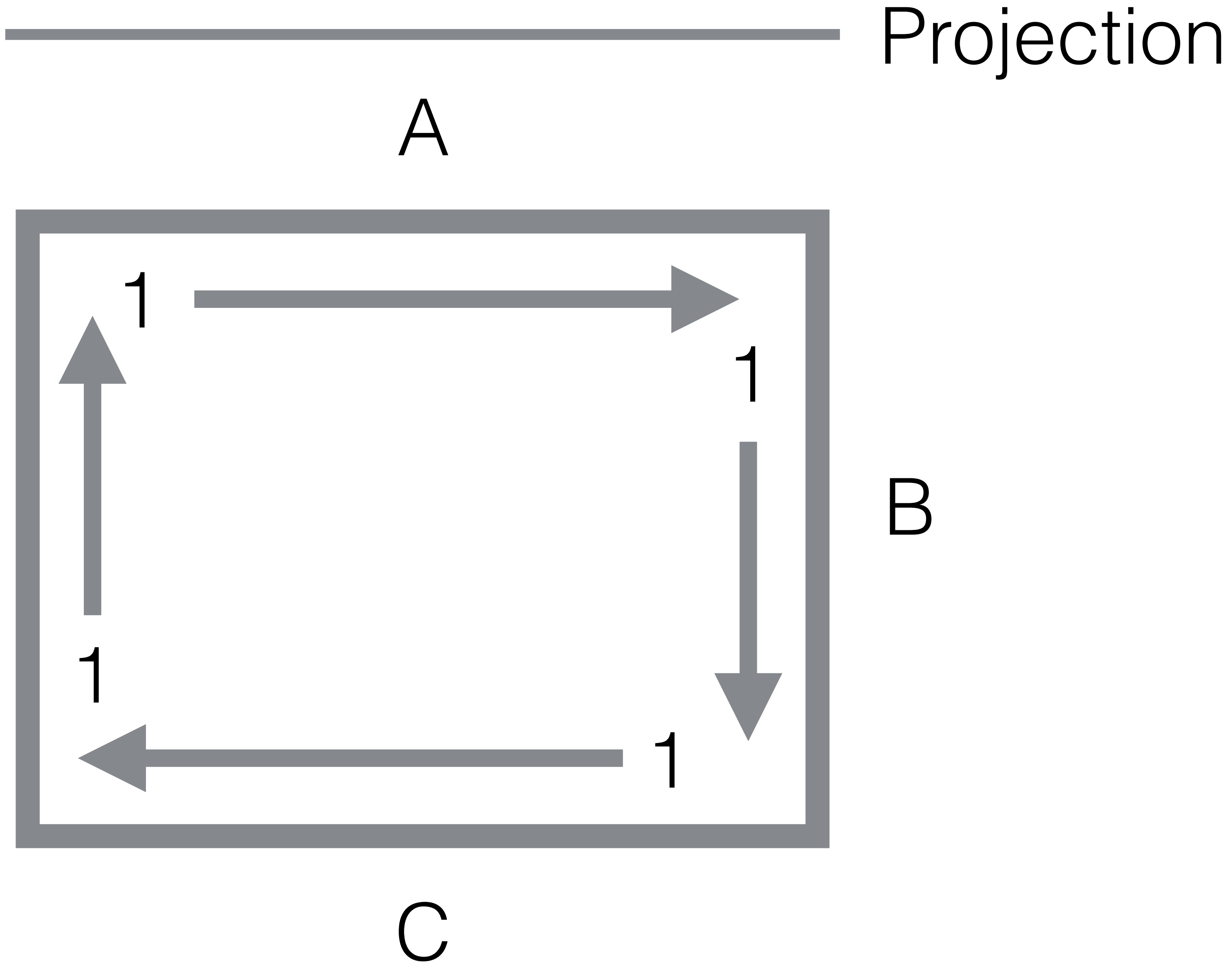
Post-Its

Ed Data Science Cycle



Quiz

http://bit.ly/1Rq6EJW



Open Data

2/27/16 7:15 PM

Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).

–opendefinition.org

Mertonian Norms (1946)

- Communalism (access)
- Universalism (contribution)
- Disinterestedness (common good)
- Organized Skepticism (critique)

Reasons for OD

- Accelerate rate of discovery
- Can’t copyright “facts”, they belong to the human race
- A lot of research is funded by the government
- “Tragedy of the anticommons” - desirable outcome is prevented by rights holders
- Prevent data rot

Open Science

- International Council for Science “World Data Centers” (1955)
- Internet decreased the cost
- Human Genome Project, Hubble Telescope, International Chemical database

Open Government

- UK (2010), Kenya (2011), Ghana (2012)
- US (data.gov) very fractured (per usual) 40 states/ 46 cities
- Government transparency, accountability and public participation
- BUT who has the means to analyze that data?

Reasons Against OD

- The only people who can process data are private interests
- Revenue from licenses allows companies to invest in new projects
- Privacy
- Collecting, tidying, maintaining data is cost intensive - people need an incentive to do this

Restrictions

- Copyright, patents
- Encryption
- Membership/fees
- Robot-blocking
- Webstacles - permissions systems, limited queries
- Lobbying

More Dimensionality Reduction

Cluster Analysis & PCA

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	

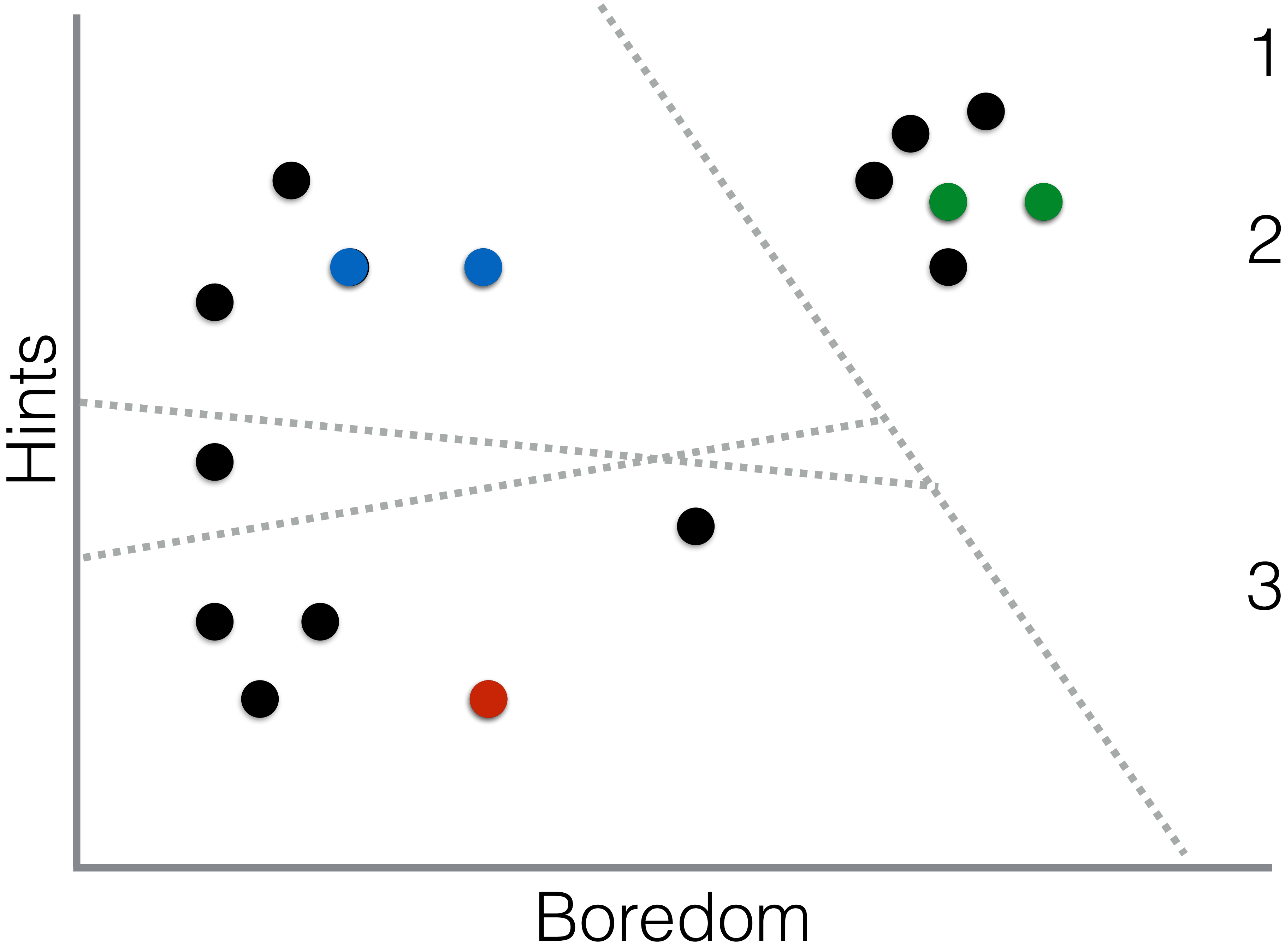
Extraction

By People

ID	Var1	Var2	Var3
A			
C			

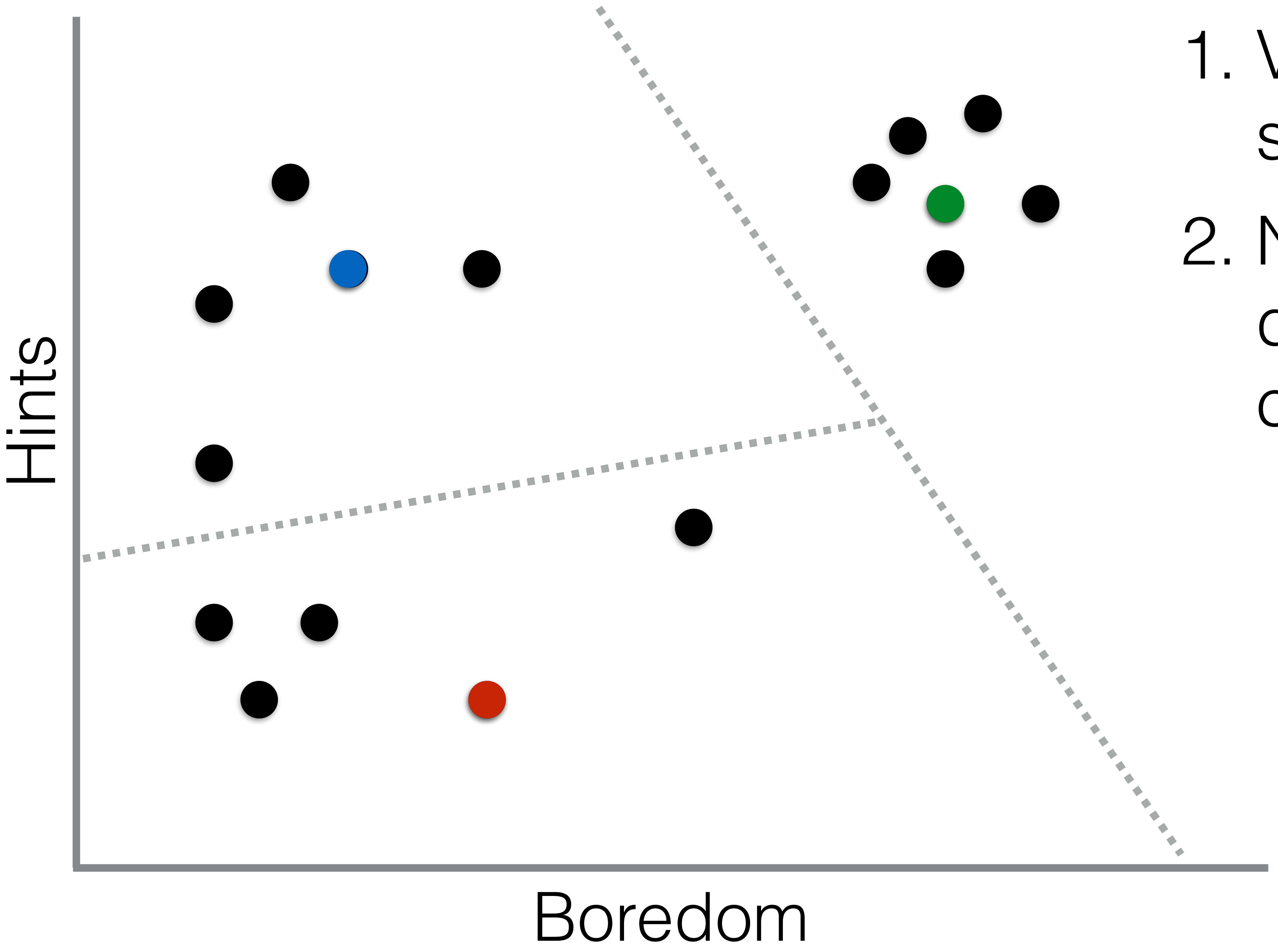
ID	Var1	Var2	Var3
B			
D			

Cluster Analysis: K-means

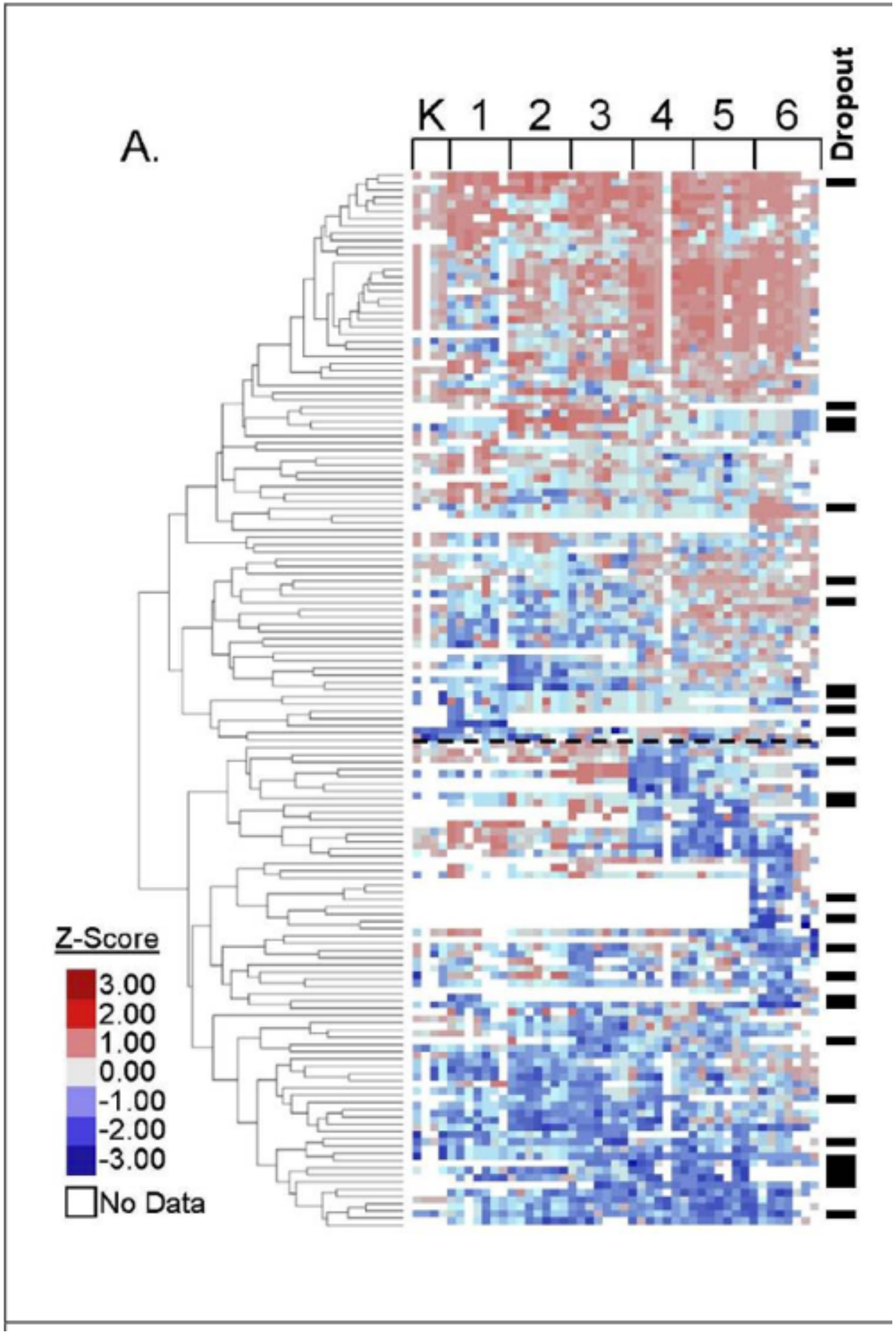


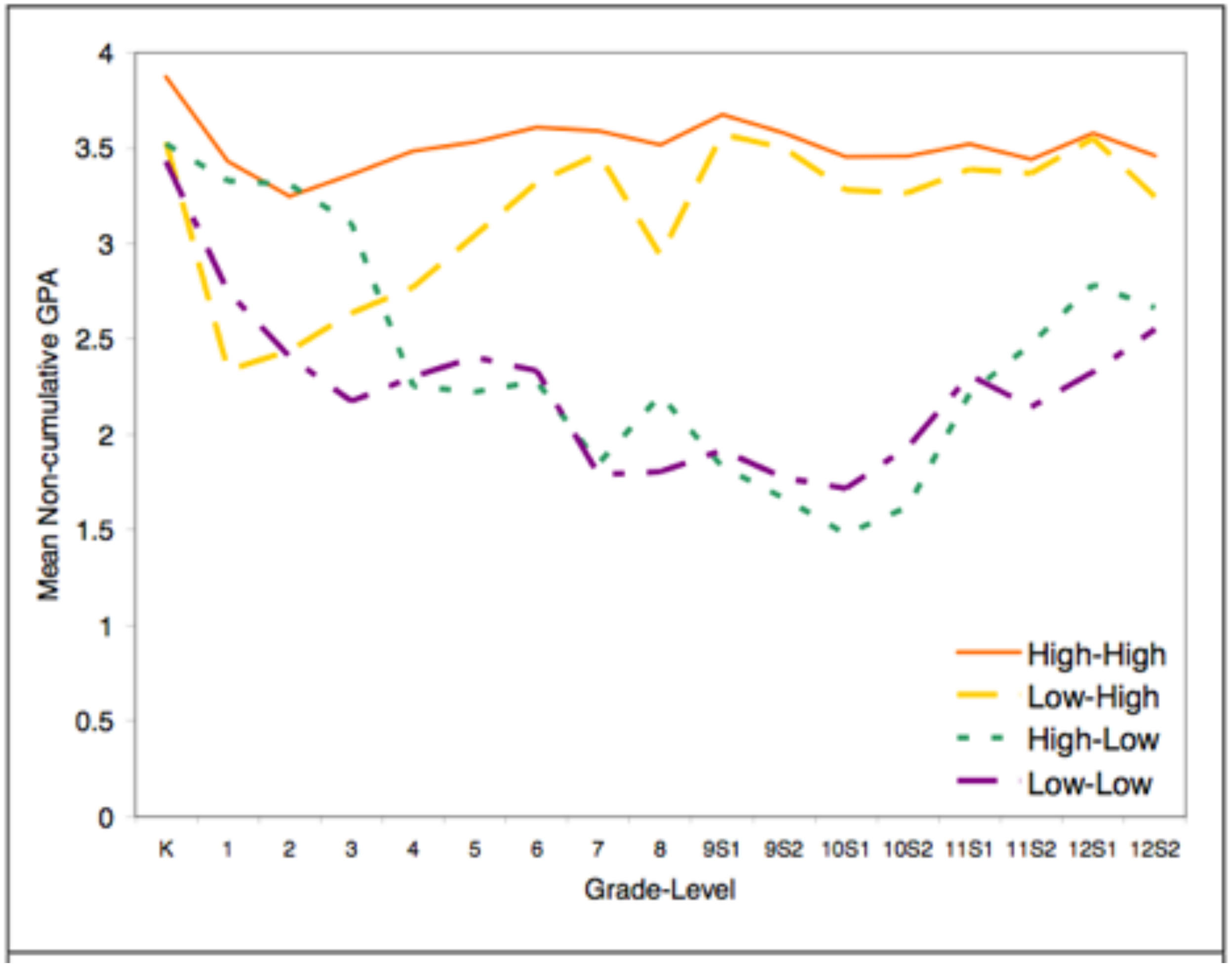
1. Select some random points
2. Associate those points with closest other points
3. Move the selected point to the mean point in the cluster

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

5.Research the code

6.Apply the code