Week 7 Video 4

Clustering Examples

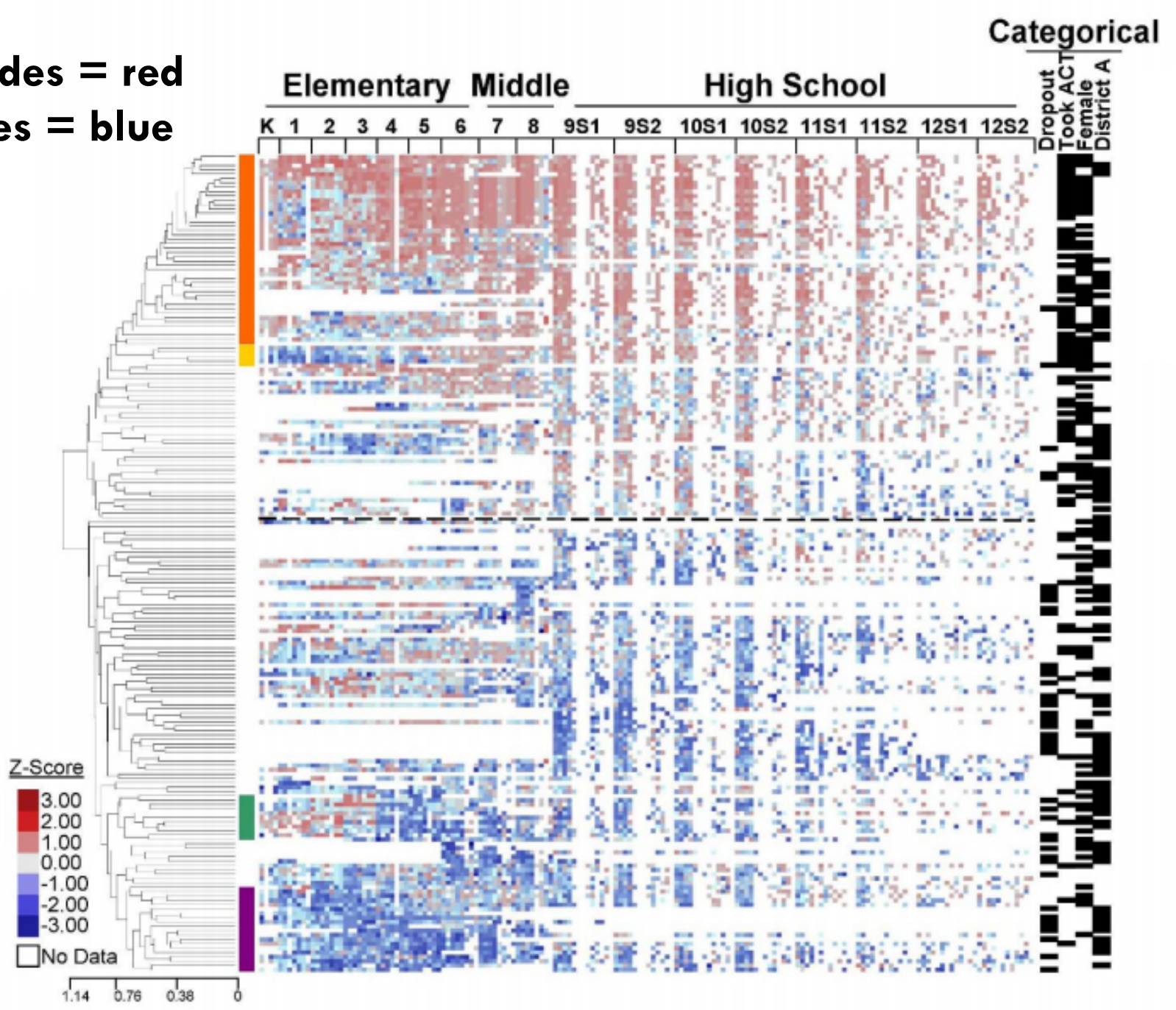
Bowers (2010)

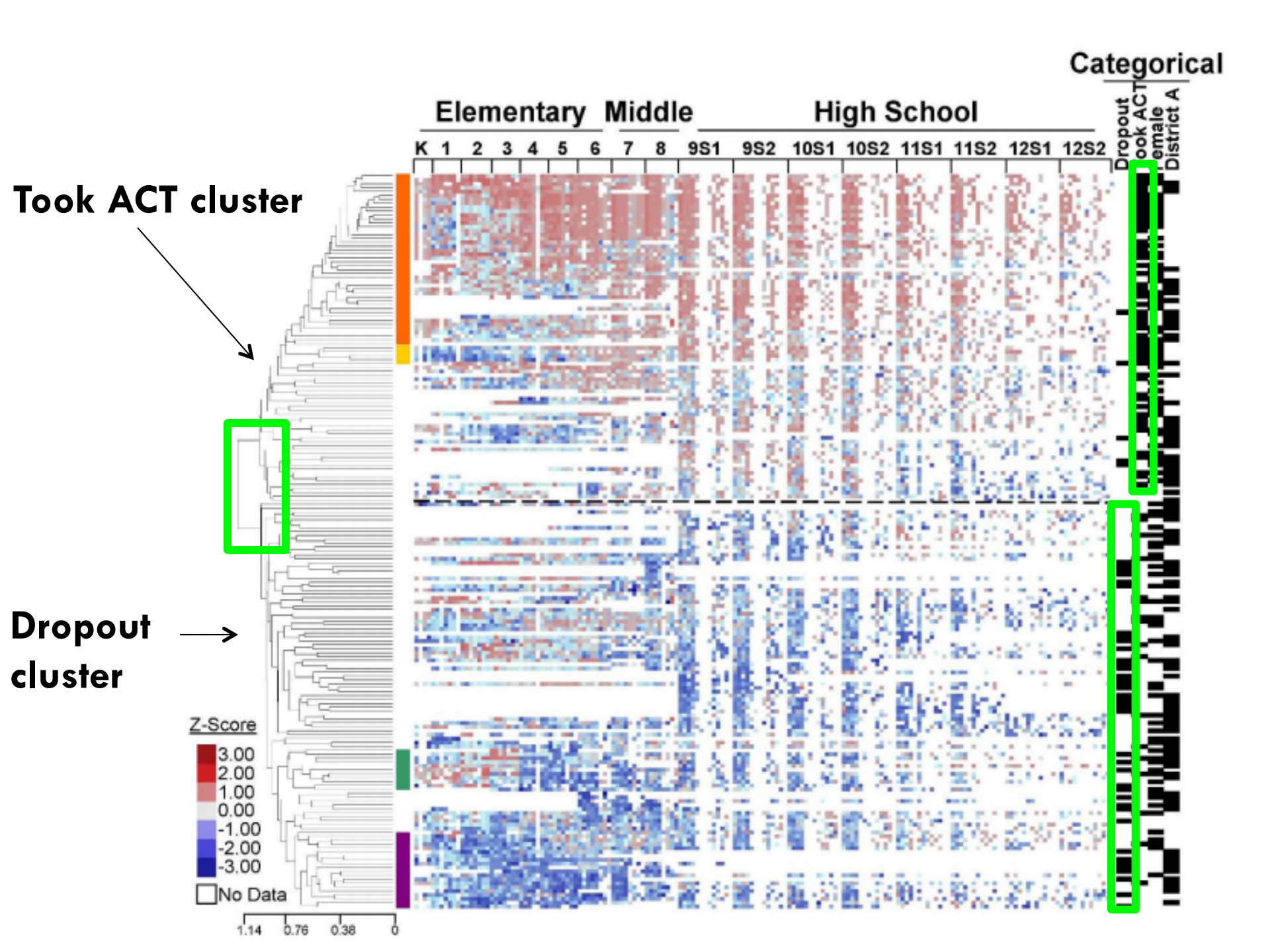
- Looked at student grades from kindergarten through high school

- Used HAC clustering to group students

Good grades = red

Bad grades = blue




Took ACT cluster
Dropout cluster
Elementary
Middle
High School
K 1 2 3 4 5 6 7 8 9S1 9S2 10S1 10S2 11S1 11S2 12S1 12S2
Categorical
Dropout
Took ACT
Female
District A
Z-Score
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
No Data
1.14 0.76 0.38 0

Four sub-clusters

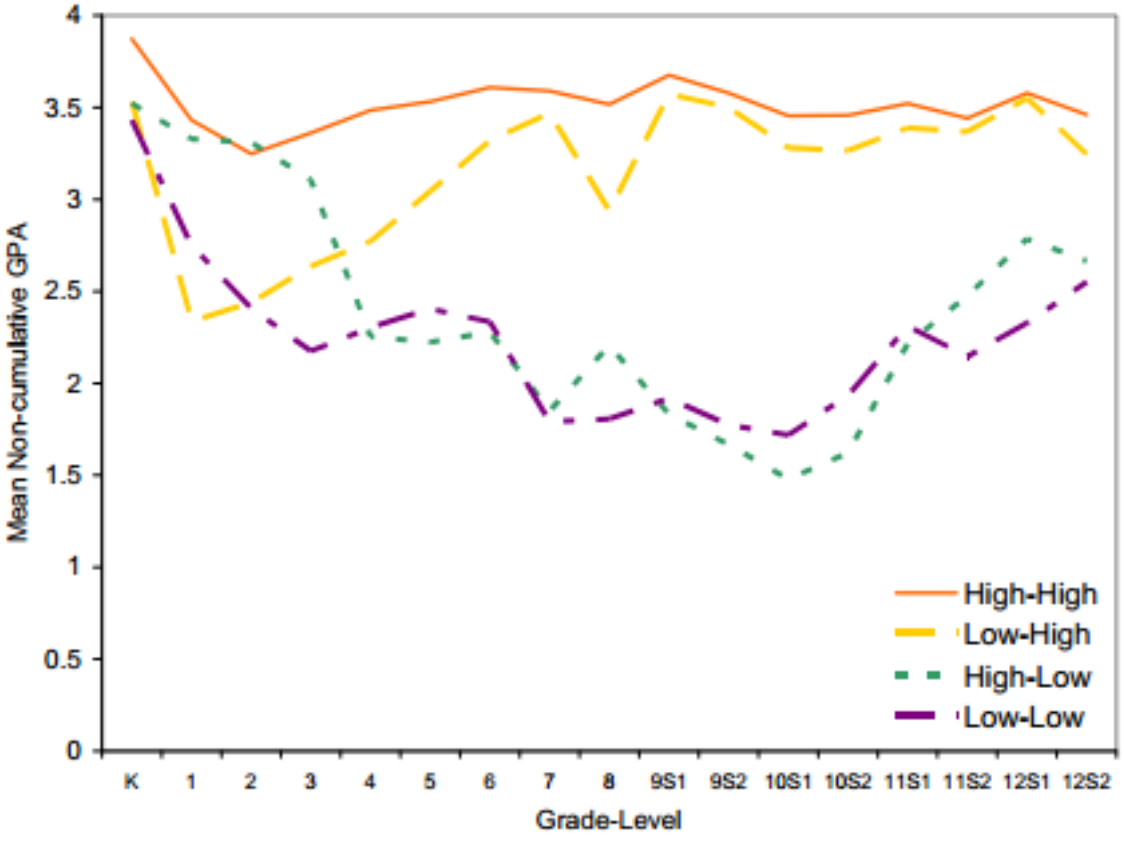


Figure 5: Mean non-cumulative GPA trends, K-12, for four sub-clusters from the hierarchical cluster analysis

Bowers (2010)

- There is a distinct group of students who did well until third or fourth grade and then started doing much poorer
  - These students were more likely to drop out of high school than students who did poorly from the start!
  - Hypothesis: Switch from "learning to read" to "reading to learn"
  - This switch was known, but perhaps not how big an impact it has…

Amershi & Conati (2009)

- Split students by their behaviors in an exploratory learning environment where students explore AI algorithms by stepping through the algorithm and making alternate execution choices in real time

- Then looked at learning outcomes

Amershi & Conati (2009)

- Three groups
  - Two low learning
  - One high learning

Amershi & Conati (2009)

- Students with lower learning (both groups) tended to move through the environment faster, and did not pause to reflect after moves
  - Failure to Self-Explain

Amershi & Conati (2009)

- The two groups of students with lower learning differed in other aspects
  - Amount of pause after back-tracking
  - Degree to which students adjusted graph
- But this (apparently) did not affect learning – both groups had poor learning

Beal, Qu, & Lee (2006)

- Clustered students in terms of five behaviors in an intelligent tutor
  - Solved problem correctly in over 10 seconds
  - Produced incorrect answer in over 10 seconds; then solved correctly in over 10 seconds
  - Answered in under 10 seconds
  - Clicked on hints in under 10 seconds per hint
  - Paused for over 10 seconds on hints

Resultant Clusters

1. Lots of fast responses; also lots of viewing hints after making an incorrect response
2. Read lots of help
3. Answered problems accurately without using help

General Point

- Were these results surprising?
- Not necessarily surprising that some students make lots of fast responses and ask for help after making an incorrect response
- A faster way to get a model of this phenomenon than conducting field observations and developing a prediction model
  - But the resultant model is by nature somewhat less validated

General Point

- The greatest power of clustering (in my opinion) is to discover completely unknown patterns
- Clustering is most useful in areas where you don't know very much…

Next lecture

- Factor Analysis