HUDK 4051: ANAIYTICS: PROCESS &



First-ever online library of edtech pricing data launches for school districts

The Boston Globe

In the news



Microsoft Honors Safer Internet Day with Launch of Digital Citizenship Campaign

Beyond Retention: Colleges Expand Data Analytics to **Business and Finance Operations**



Top Programming Languages Trends: The Rise of Big Data







How colleges can use Big Data to improve IT operations

The Myth of **Immigrants'**



Those who move to the United States tend to have higher socioeconomic standing in their native countries than what they settle for when they



Technology hijacked sports

AI bests Air Force combat tactics experts in simulated dogfights

Wildflower schools bloom amid discontent with education

Vizio smart TVs tracked viewers around the clock without consent



B B C Local voting figures shed new light on EU referendum





The Prestige Gap



UNIVERSITY of WISCONSIN Online elementary education program recognized by AffordableColleges.com





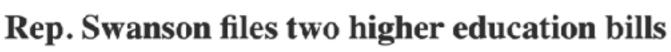
This is why we publish the World University Rankings



Virginia education officials trying to figure out why 1.5 percent fewer students enrolled in state public, private colleges and universities this academic year

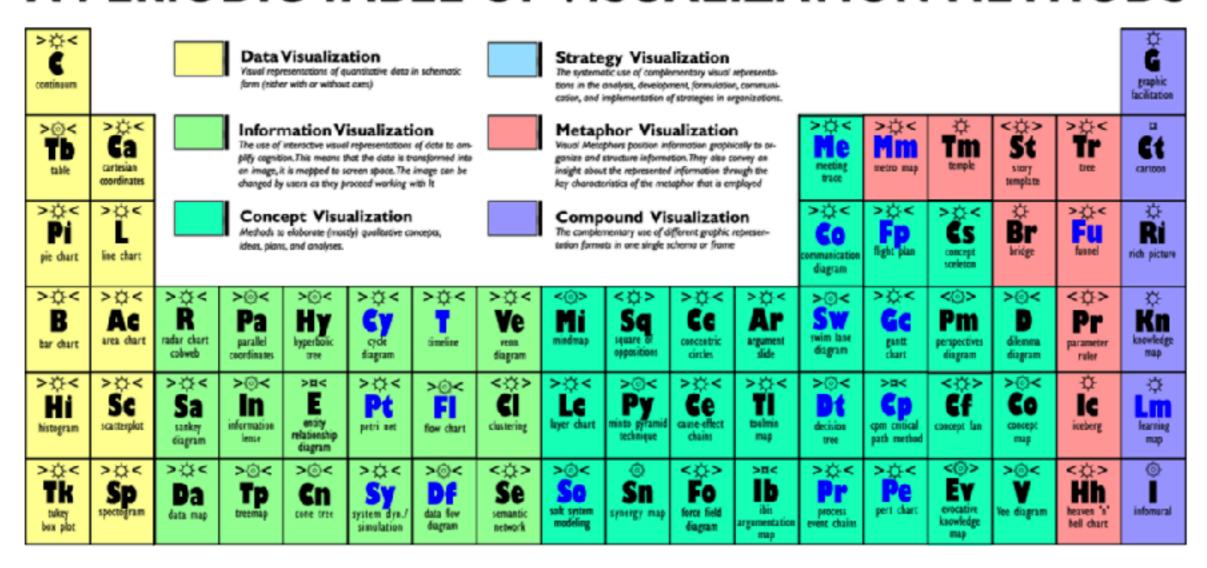


State 'school climate' ratings a disservice, says Clarke data analyst





A PERIODIC TABLE OF VISUALIZATION METHODS



http://www.visual-literacy.org/periodic_table/ periodic_table.html

Opportunities

- Application Deadline: Feb 24, 2017 Curriculum Associates Psychometric and Research Summer 2017 Internship (see attached)
- Application Deadline May 2, 2017: Simon Initiative's LearnLab Summer School (http://learnlab.org/index.php/simon-initiative-summer-school/)
- Mar 4, 2017 BetaNYC School of Data (https://schoolofdata.nyc/)
- Mar 7, 2017, Boston, AnalyticsFair (http://analyticsfair.com, FREE for Jobseekers and Early Startups)
- Feb 24, 2017 DataKind Giving-Tuesday Prep-Course (https://www.eventbrite.com/e/givingtuesday-datadive-pre-event-prep-course-with-metis-tickets-31640075317)

Podcast

• Last week: SC: 19, LP: 2, TW: 2

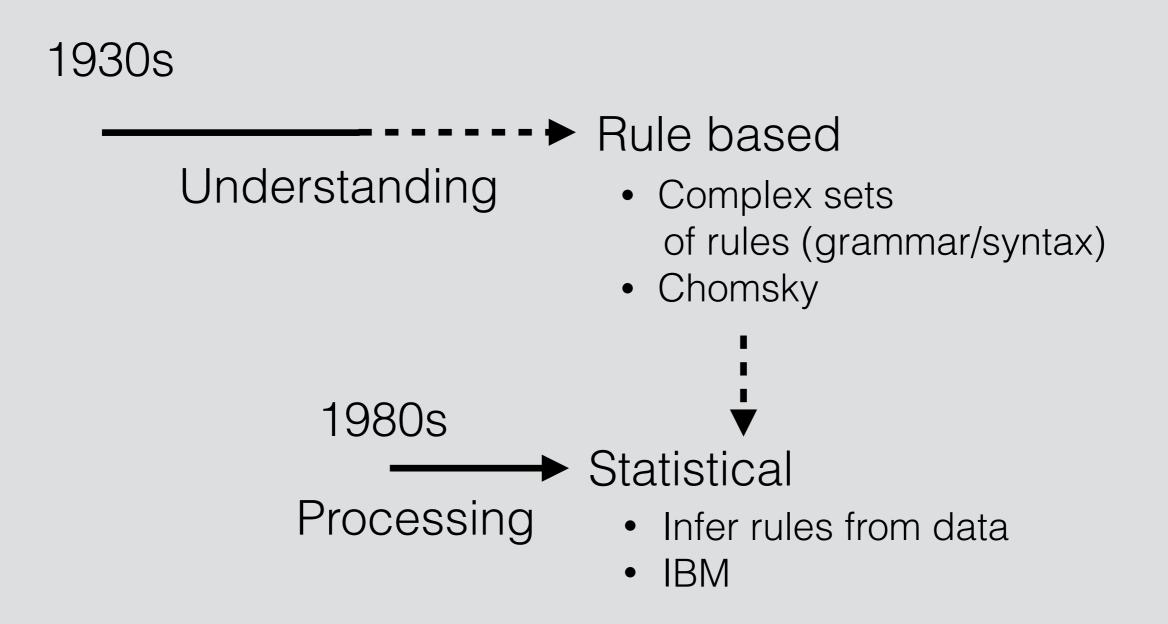
Natural Language Processing

NLP

Analyses of language produced by humans (by computers)

- Treats language as a varied pool of information sources
- In order to:
 - Understand language (Cognitive Science)
 - Respond to the speaker appropriately (AI)
- Examples
 - Translation
 - Automated feedback (education, shopping)
 - Study linguistics, cognition, development, etc.

Methodological History



Industry



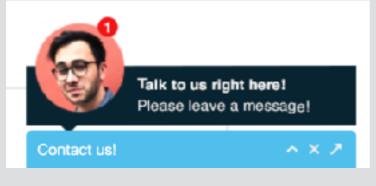












Education

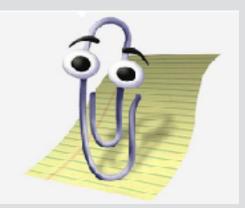












GLENCOE ONLINE ESSAY GRADER

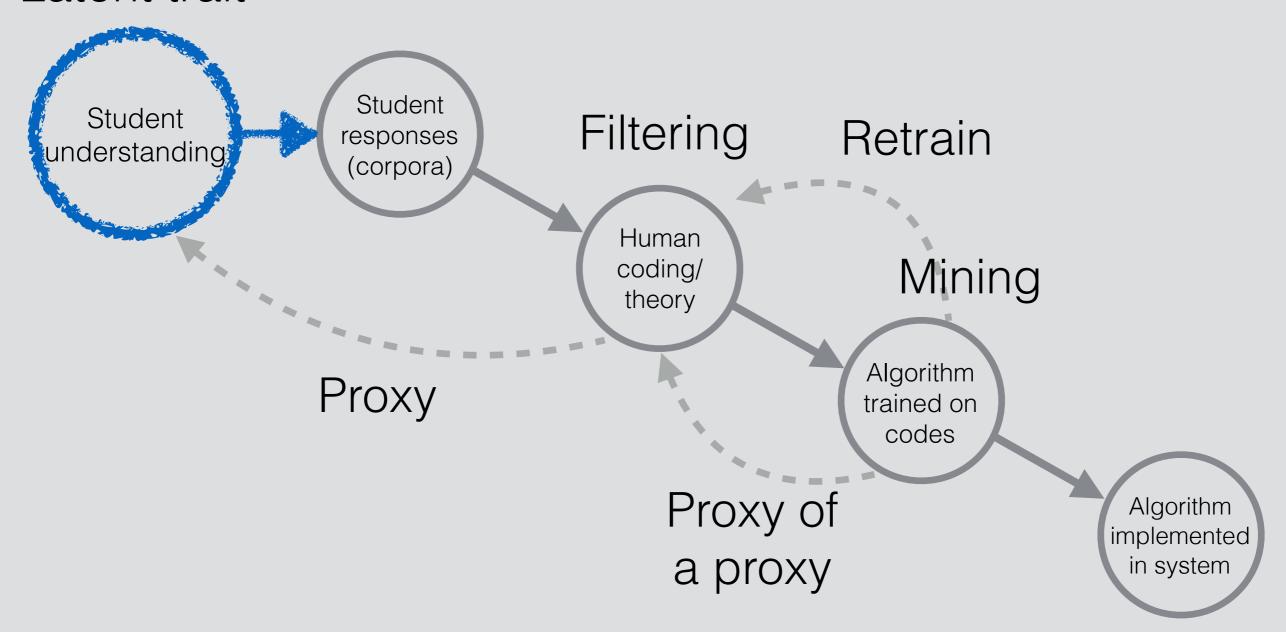


Essential Problem

- Heterogeneity
- We get rid of this by asking MCQ questions but we also throw out a lot of information when we do that
- Collect more data and more complex data through written answers

Overall Method

Latent trait



Coding

Word counting



Google books Ngram Viewer

Types of Expressions

"I don't know..."

"I dunno..."

Stemming

Take the root of the word: educate, education, educating

Tokenization (bag of words)

Chopping word/phrase into tokens

- Remove punctuation
- Find best number of letters to represent a word/meaning
- Consider all possible versions of word
- Stop word removal



Algorithms

Feature selection

 Not all tokens are useful, which ones can we scrap?

Feature extraction

Extracting features from combining tokens