

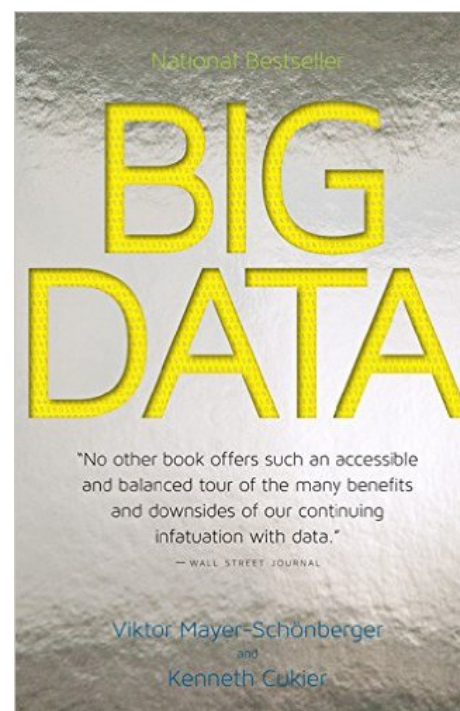
Lecture 6: Datafication & More SQL

Core 109S IDWT?, Spring 2017
Michael Hay

Instructions: Please find a *NEW* group of *FOUR* to discuss readings.

Discussion

- What is *datafication*?
- How is it different from *digitization*? (Illustrate with an example.)
- To datafy something, do you first need to digitize it? (Illustrate with an example.)
- What does Streetbump have to do with datafication?
- "The Hidden Biases in Big Data" article talks about a "signal problem." What does that mean? Illustrate with an example.



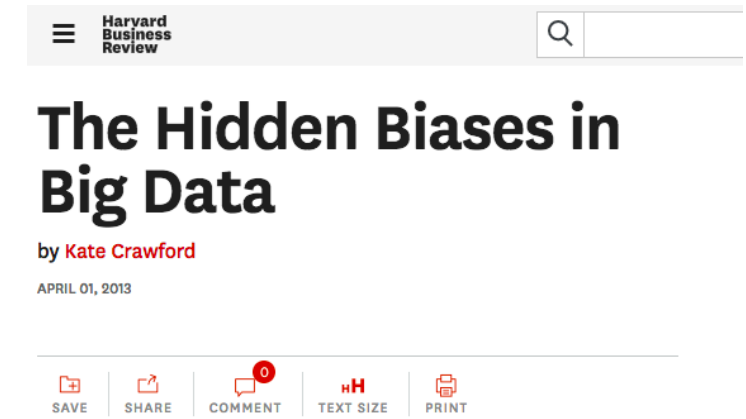
PARTNER CONTENT PHIL SIMON

POTHOLES AND BIG DATA: CROWDSOURCING OUR WAY TO BETTER GOVERNMENT



Image: get directly down/Flickr

<https://www.wired.com/insights/2014/03/potholes-big-data-crowdsourcing-way-better-government/>



<https://hbr.org/2013/04/the-hidden-biases-in-big-data>

SQL Continued...

- See today's handout

Exercise 1

Instructions: ~1 minute to think/
answer on your own; then discuss with
neighbors; then I will call on one of you

- Find applications where the major is CS and the college is either Colgate or Bucknell.
- (If you are trying this out on the computer) Double check the result... do you have any non-CS majors in the result? You shouldn't!

Exercise 2

Instructions: ~1 minute to think/
answer on your own; then discuss with
neighbors; then I will call on one of you

- Write a query that returns the students with the top three highest GPAs. (Bonus: break GPA ties by ordering by high school size, smallest to largest.)

Exercise 3

Instructions: ~1 minute to think/
answer on your own; then discuss with
neighbors; then I will call on one of you

cName	state	enrollment
Colgate	NY	2700
Bucknell	PA	3650
Williams	MA	2000
Cornell	NY	21000

- Write a query to compute total college enrollments by state. On the input College relation (shown above), it would produce this output:

state	totalEnroll
MA	2000
NY	23700
PA	3650

Exercise 4

Instructions: ~1 minute to think/
answer on your own; then discuss with
neighbors; then I will call on one of you

cName	state	enrollment
Colgate	NY	2700
Bucknell	PA	3650
Williams	MA	2000
Cornell	NY	21000

- Write a query to find states whose total college enrollment exceeds 20,000. On the input relation above, it would produce this output:

state
NY