

Cool Stuff in Data

Data can be anything!

Julie Deeke

Data can be Numbers

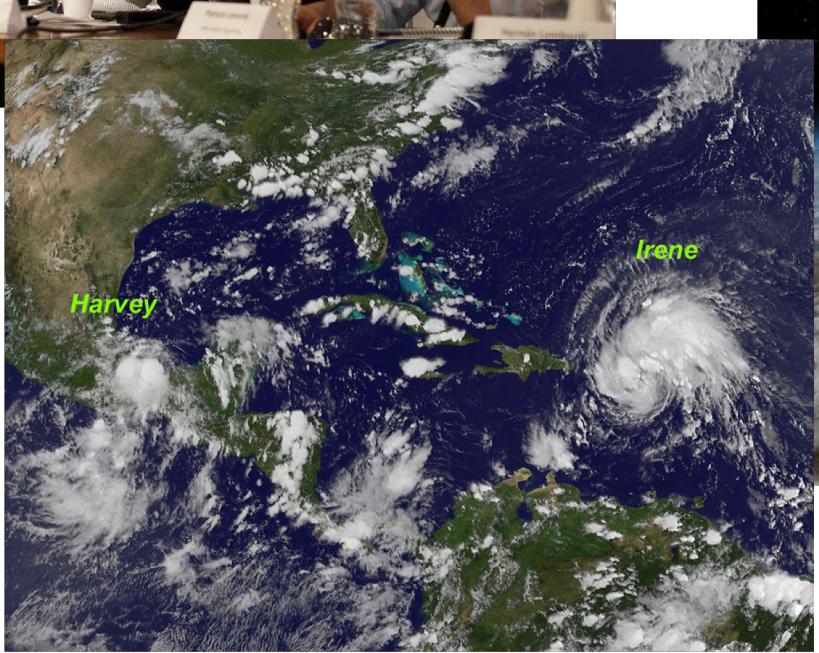
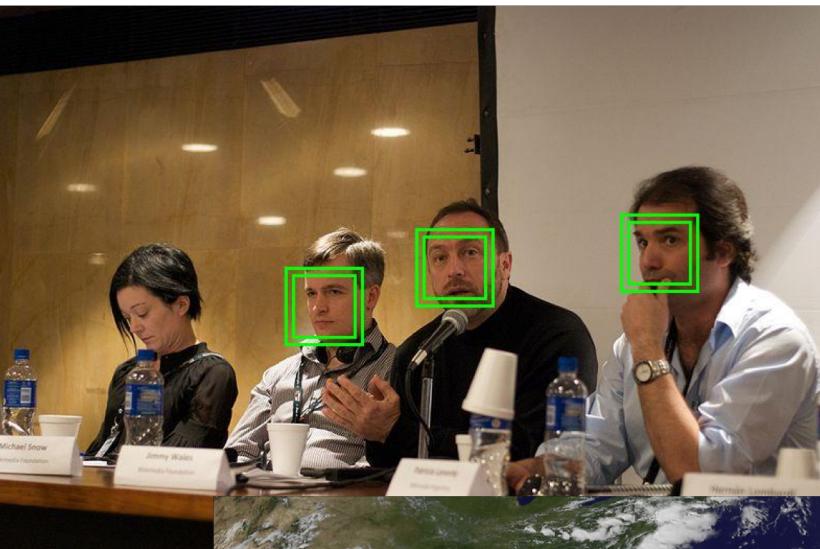


National Health and Nutrition Examination Survey

seqn	ridstatr	riagendr	RIDRETH1	dmdmartl	WTINT2YR	WTMEC2YR
62161	2	1	3	5	102641.406	104236.583
62162	2	2	1	NA	15457.737	16116.354
62163	2	1	5	NA	7397.685	7869.485
62164	2	2	3	1	127351.373	127965.226
62165	2	2	4	NA	12209.745	13384.042
62166	2	1	3	NA	60593.637	64068.123
62167	2	1	5	NA	5024.465	5303.683
62168	2	1	5	NA	5897.025	6245.044
62169	2	1	5	5	14391.778	14783.601
62170	2	1	5	NA	7794.527	8291.637

Data can be images

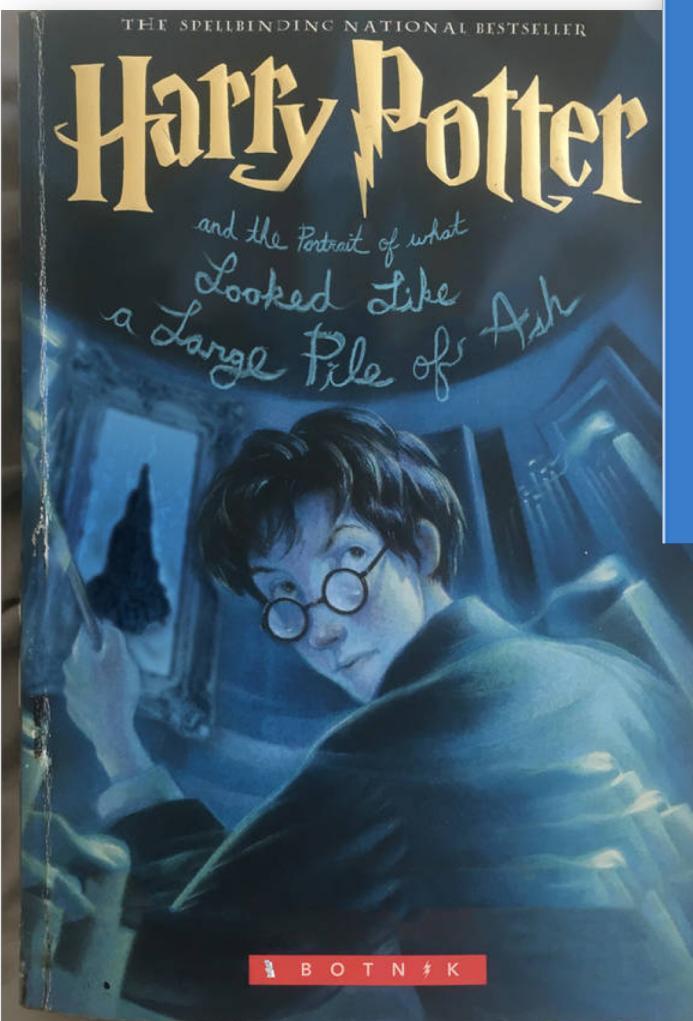
<https://quickdraw.withgoogle.com/>



"Eigenfaces" by Gunnar Grimnes is licensed under [CC BY 2.0](#)
"Face detection" by Sylenius is licensed under [CC BY 2.0](#)
"Great picture..." by Dion Hinchcliffe is licensed under [CC BY-SA 2.0](#)

Artist concept of the Orbiting Carbon Observatory. [Image credit: NASA/JPL](#)
"Satellite view of a hurricane" – [CC0 1.0](#)
"Serra do Mar forest WWF.jpg" – [CC0 1.0](#)

Data can be Words



What is an electronic health record (EHR)?

 A yellow square icon containing a stylized grid of three columns. The first column has three circles (two blue, one red), the second has three horizontal bars (blue, green, yellow), and the third has three hearts (red).

“Arts”	“Budgets”	“Children”
NEW	MILLION	CHILDREN
FILM	TAX	WOMEN
SHOW	PROGRAM	PEOPLE
MUSIC	BUDGET	CHILD
MOVIE	BILLION	YEARS
PLAY	FEDERAL	FAMILIES
MUSICAL	YEAR	WORK
BEST	SPENDING	PARENTS
ACTOR	NEW	SAYS
FIRST	STATE	FAMILY
YORK	PLAN	WELFARE
OPERA	MONEY	MEN
THEATER	PROGRAMS	PERCENT
ACTRESS	GOVERNMENT	CARE
LOVE	CONGRESS	LIFE

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

Figure 8: An example article from the AP corpus. Each color codes a different factor from which the word is putatively generated.

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Tolga Bolukbasi¹, Kai-Wei Chang², James Zou², Venkatesh Saligrama^{1,2}, Adam Kalai²

¹Boston University, 8 Saint Mary's Street, Boston, MA

²Microsoft Research New England, 1 Memorial Drive, Cambridge, MA

tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

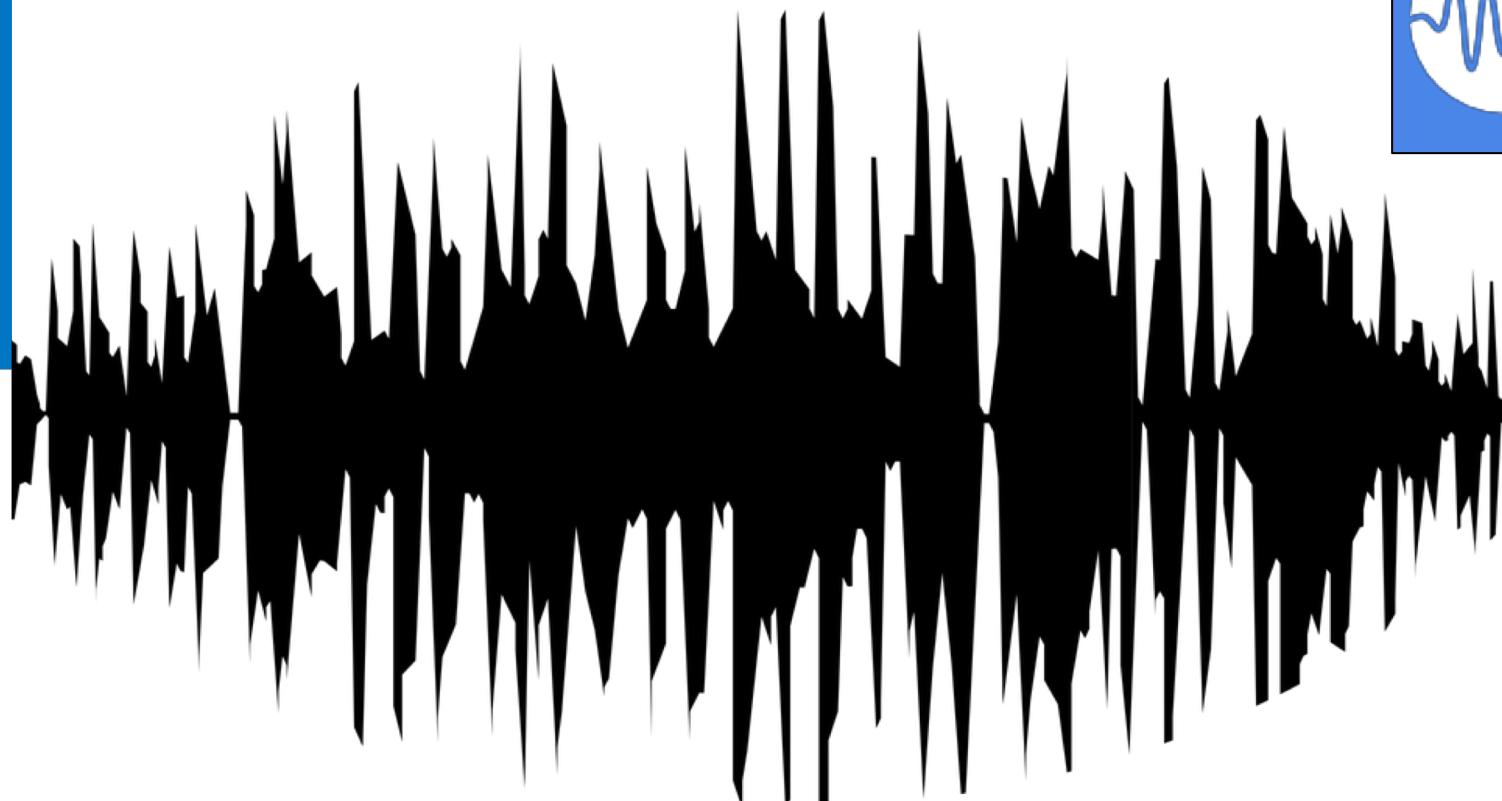
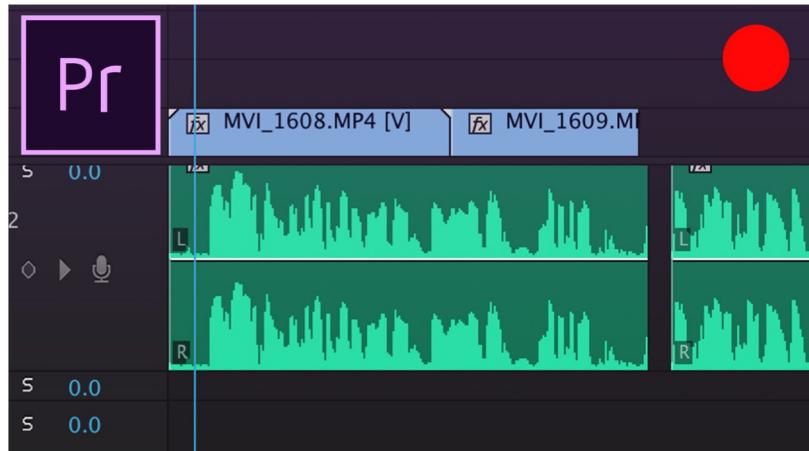
Image source: <https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/health-it-and-health-information-exchange>

Image source: <http://botnik.org/content/harry-potter.html>

Image source: Page 17 <http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>

Image source: <https://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf>

Data can be Audio



Historical Example

