

# Data, Machines and ... People!

Enrico Bertini  
Associate Professor



How many of you have heard the term  
“*Data Science*” before?



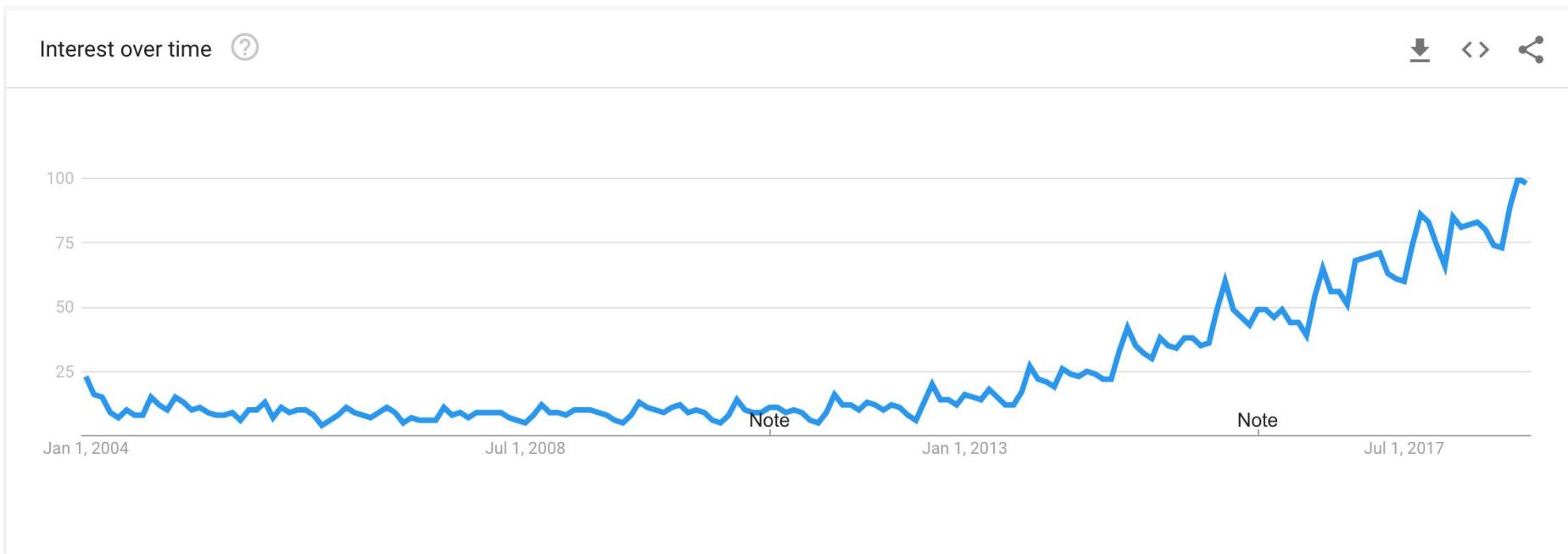
## VAC Consortium 2010 Meeting



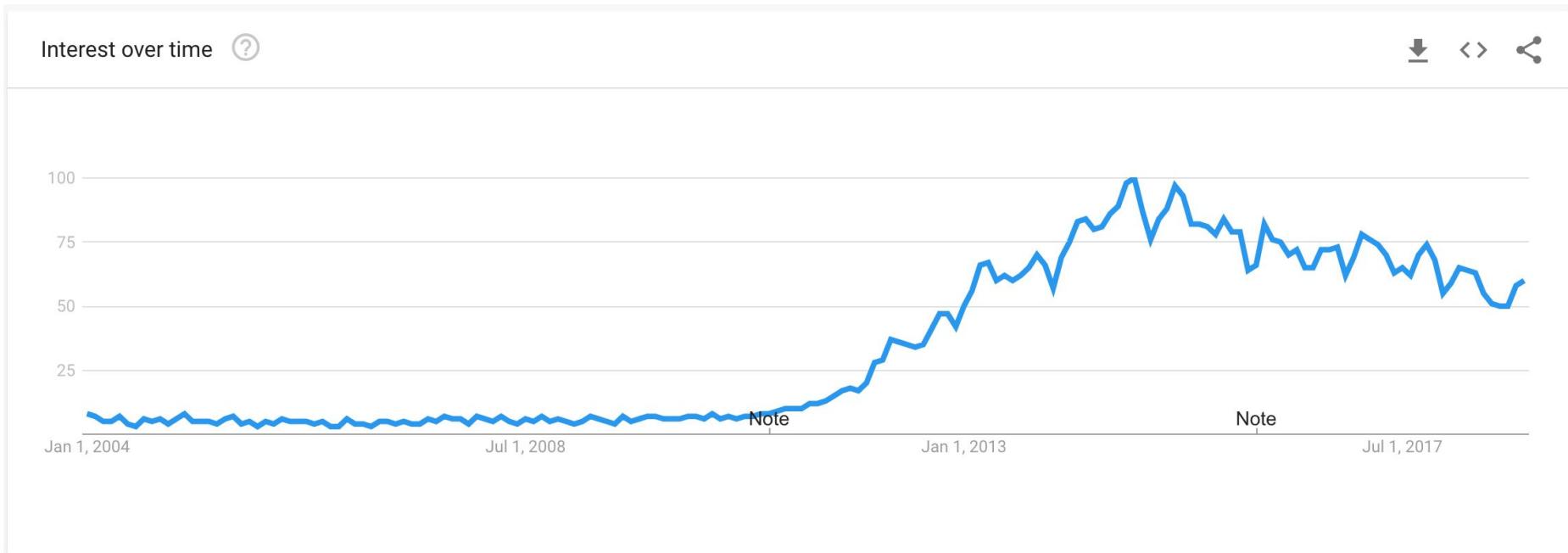
More than 150 participants from academia, government, and industry attended the Visual Analytics Community Consortium Meeting August 31-September 1, 2010 at the University of Maryland. The theme of this year's meeting was "Broadening Our Horizons." Dr. Jennifer Leaning kicked off the first day with a keynote address, which was followed by a panel discussion on Global Crisis Management that focused on the information challenges of the Haiti earthquake, the Gulf of Mexico oil spill, and other recent events. In another first-day session, NVAC leaders introduced the new community website, [vaccommunity.org](http://vaccommunity.org); described new community resources being developed; and elicited feedback from attendees.

Day two kicked off with keynote addresses from Joe Kielman, who described new applications for visual analytics, and Ben Shneiderman, who spoke about "Understanding Innovation Trajectories for Visual Analytics." These addresses were followed by a panel discussion about the use of interactive visualization and visual analytics in consumer applications. Representatives from UKVAC, NVAC, and Defence R&D Canada described current research efforts in the afternoon. The day concluded with panel discussions on the uses and opportunities for visual analytic technologies in a variety of fields.

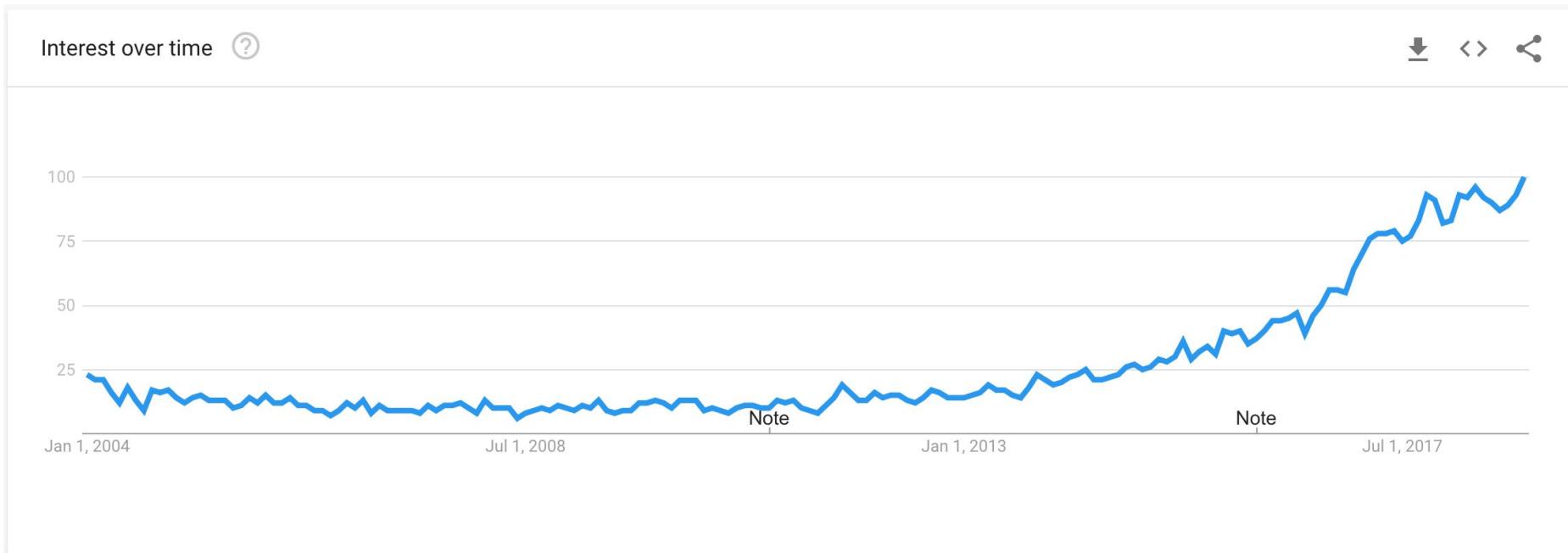
# Data Science



# Big Data



# Machine Learning

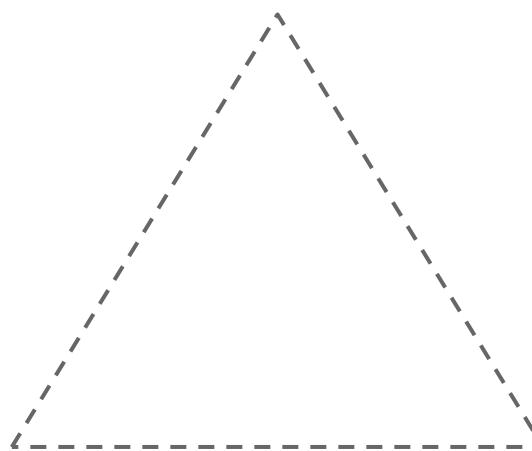


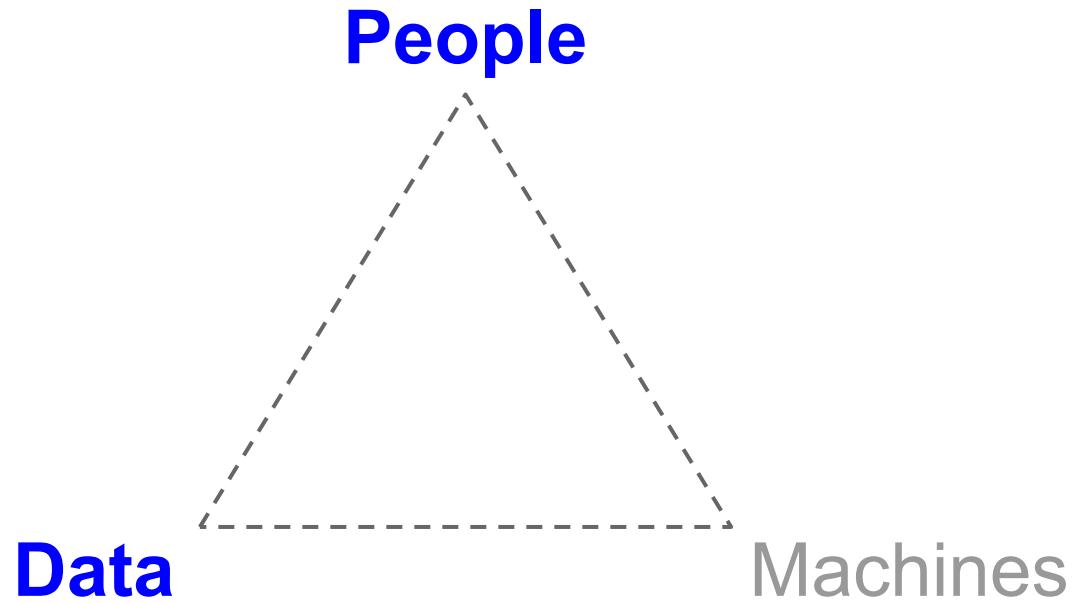
Why are we so excited?

What **can** we achieve?

What do we **want** to achieve?

**People**





# “Intelligence Amplification”

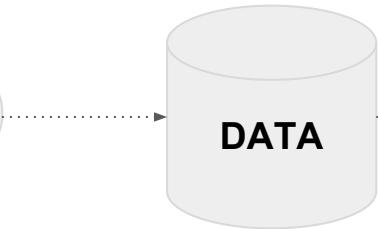
How can we use machine to help people do  
remarkable things with data?

# The world is full of complexity!

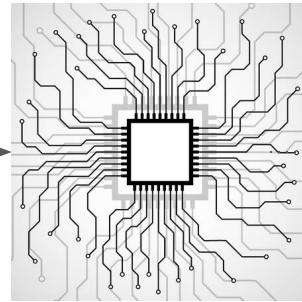
Biology  
Finance  
Social Media  
Cyber Security  
Urban Systems

...

If we are able to capture relevant information about phenomena through data, then maybe we can understand these phenomena better?



COMPUTERS



HUMANS



**We are not interested in data!**

We are interested in the **phenomena** described by data.  
And our ability to **understand** them better with data.

# Illuminating the Path

The Research and Development Agenda  
for  
Visual Analytics

Edited by James J. Thomas and Kristin A. Cook



National Visualization and Analytics Center™



## Visual Analytics

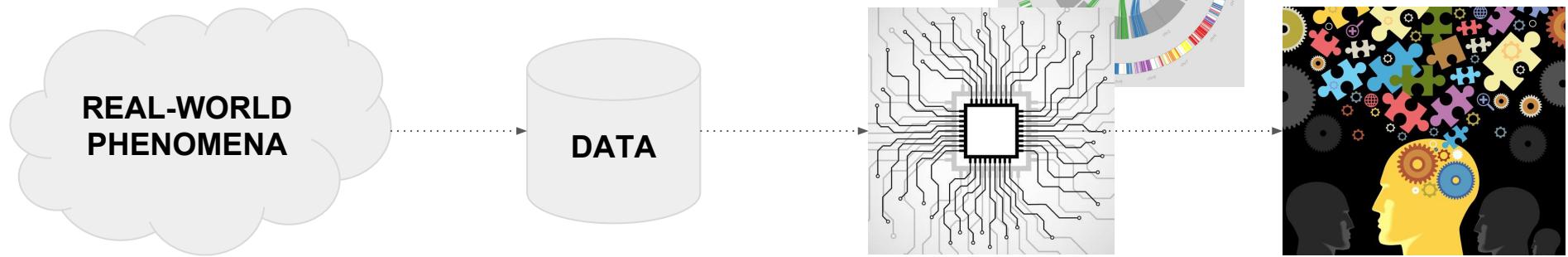
*“The science of analytical  
reasoning supported by  
interactive visual interfaces”*

REAL-WORLD PHENOMENA

DATA

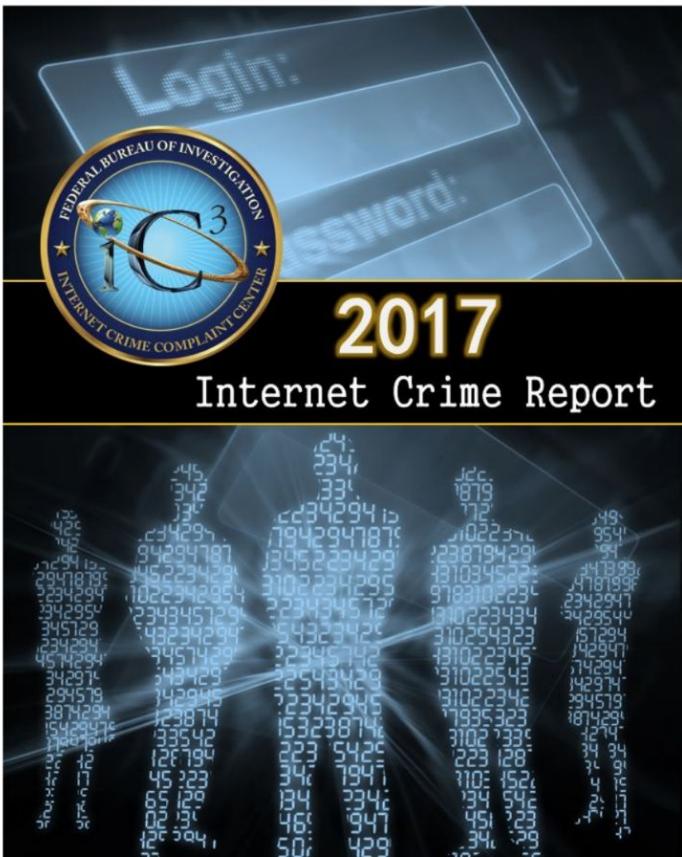
COMPUTERS  
+  
INTERACTIVE VISUALIZATIONS

HUMANS



**Visual perception** is the most effective channels we have to obtain information from the world.

# Example: Help Investigators Fight Scammers



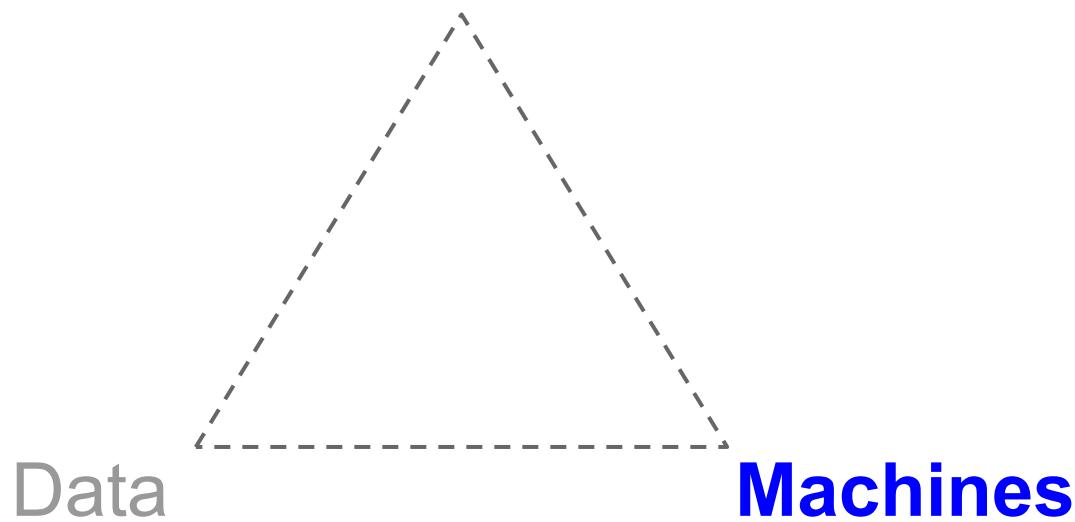
## Top 10 Crime Types Reported to IC3 in 2017 (by Victim Loss)



From the 2017 Internet Crime Report

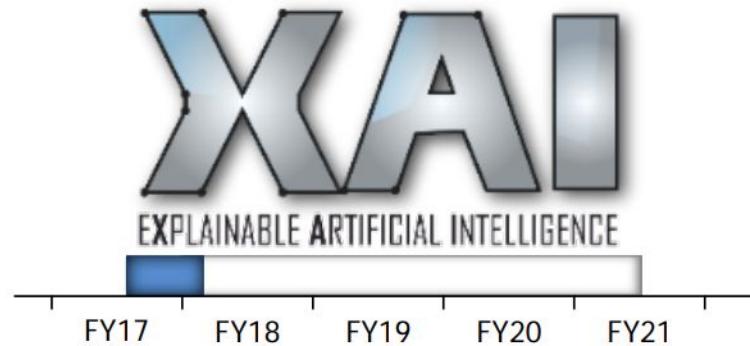


**People**



As more automated decision affect our lives, how do we retain the power to **influence, verify and understand** these decisions?

# Explainable Artificial Intelligence (XAI)



David Gunning

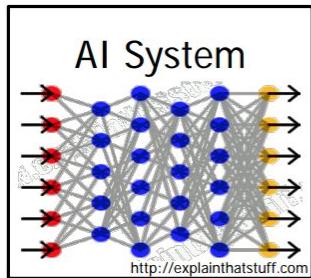
DARPA/I2O

Program Update November 2017

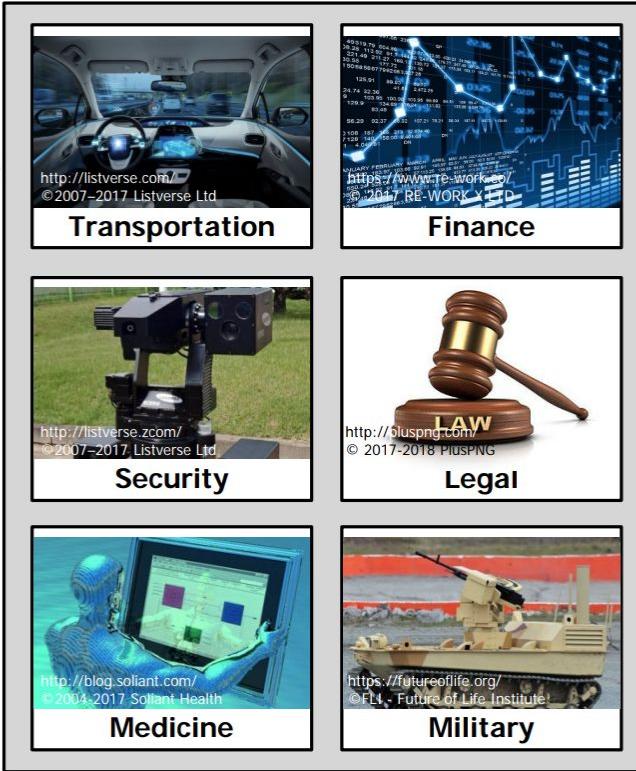




# The Need for Explainable AI



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand



- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Machine learning is only as good as the  
data we feed to them ...

# Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

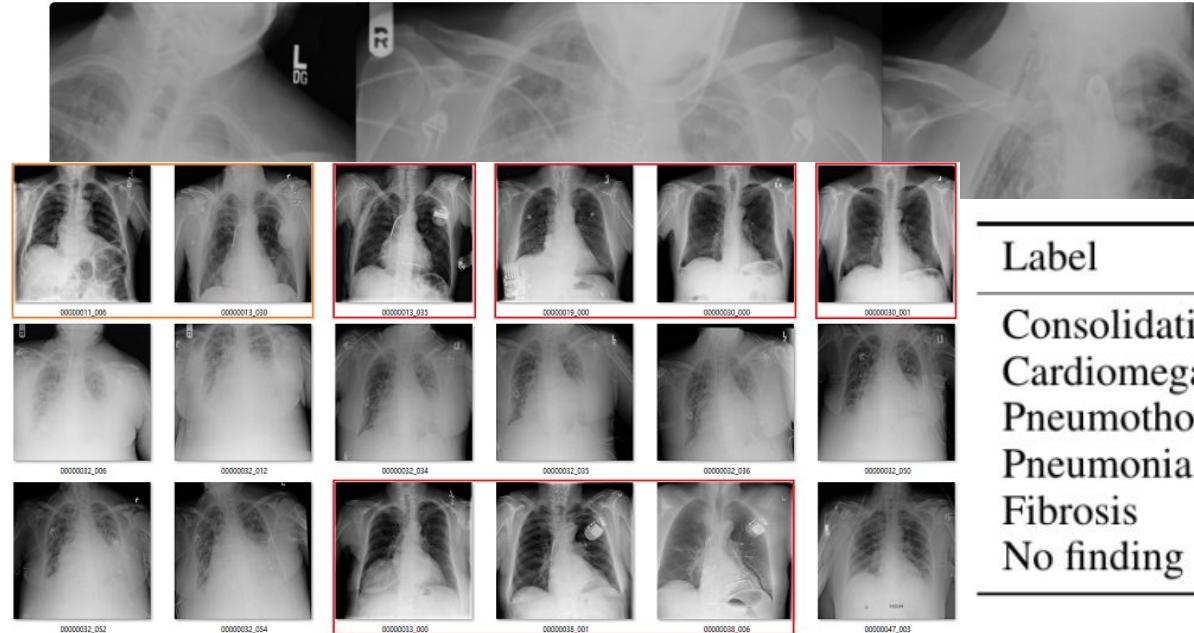
May 23, 2016

**O**N A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.



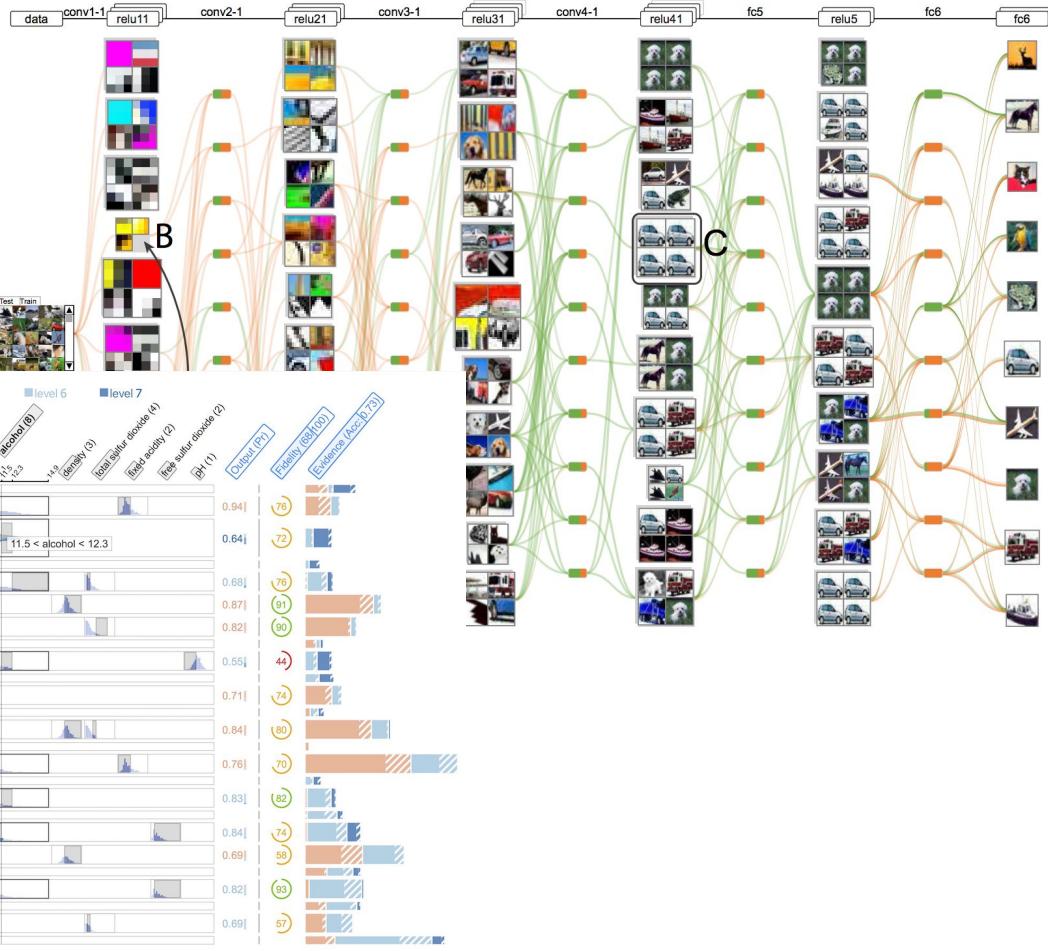
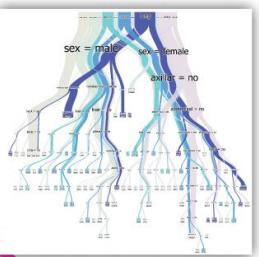
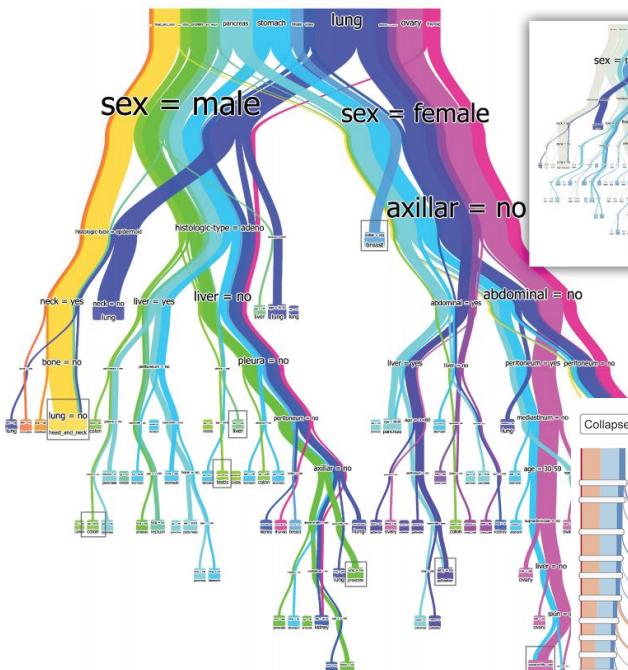
# Exploring the ChestXray14 dataset: problems



Label	PPV (visual)	PPV (text mining)
Consolidation	35%	97%
Cardiomegaly	80%	97%
Pneumothorax	60%	89%
Pneumonia	35%	89%
Fibrosis	24%	92%
No finding	60%	NA

A couple of weeks ago, I mentioned I had some concerns about the [ChestXray14 dataset](#). I said I would come back when I had more info, and since then I have been digging into the data. I've talked with Dr Summers via email a few times as well. Unfortunately, this exploration has only increased my concerns about the dataset.

We need methods and tools to help us  
**understand** and **verify** AI systems!



# Predicting Admission from ER

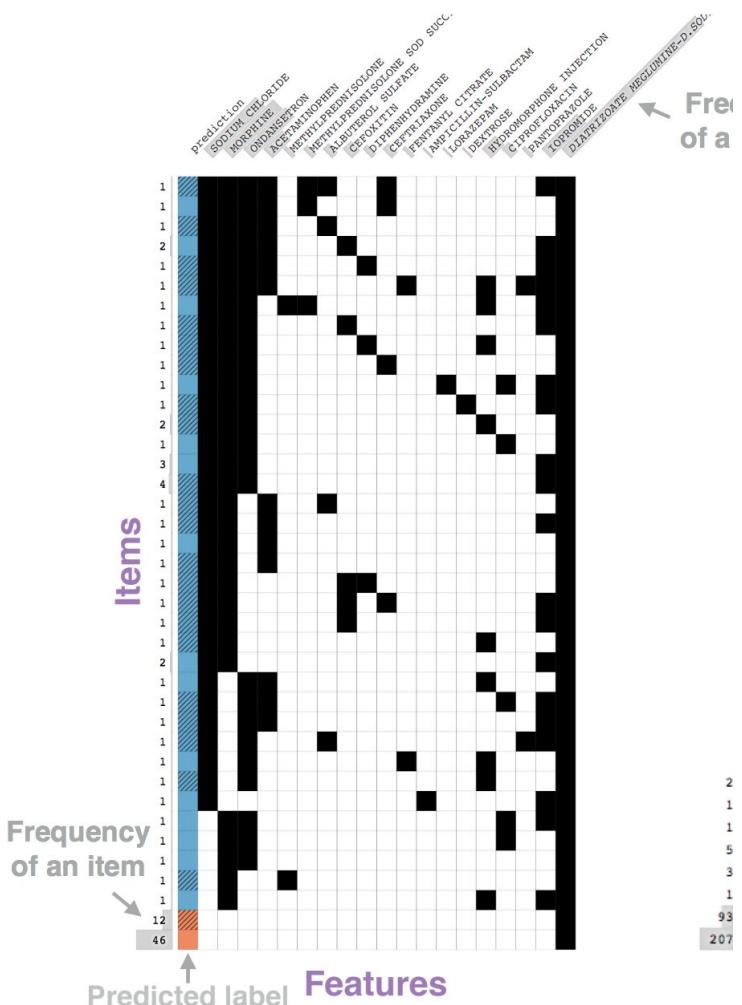


Will she/he be  
admitted?



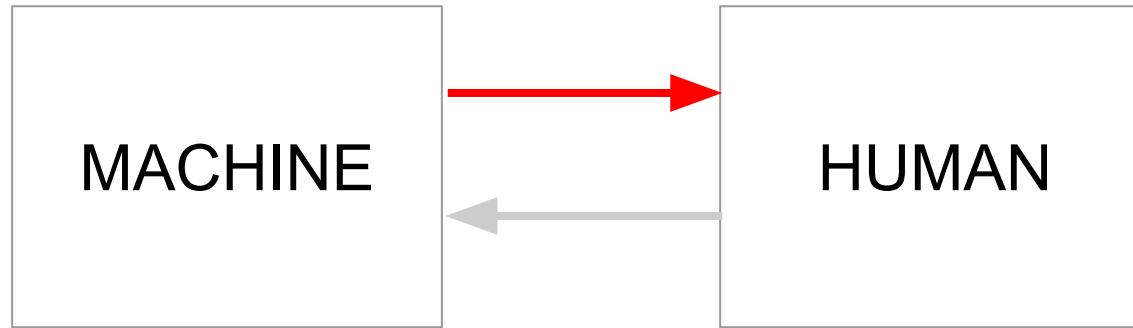
# Explainer: visual analytics to understand black-box ML models.

		tp	fp	fn	tn	pos / total	odds ratio
		3187				1328 / 4784	0.1 1 10 100
SODIUM CHLORIDE		93	207			99 / 313	
IBUPROFEN		218				13 / 231	
VANCOMYCIN		136	49			136 / 185	
ACETAMINOPHEN		137				22 / 159	
ASPIRIN		63	68			63 / 131	
OXYCODONE-ACET.		102				14 / 116	
ONDANSETRON		101				14 / 115	
DIA. MEGLUMINE.		19	26	46		31 / 103	
KETOROLAC		84				13 / 97	
MORPHINE		29	45			29 / 74	
SODIUM CHLORIDE	ACETAMINOPHEN	13	40			13 / 53	
SODIUM CHLORIDE	DIA. MEGLUMINE.	19	27			19 / 46	
	FAMOTIDINE	9	36			9 / 45	
	FUROSEMIDE	36	8			36 / 44	
ONDANSETRON	MORPHINE	25	18			25 / 43	
	LIDOCAINE	40				3 / 43	
	ONDANSETRON	13	28			13 / 41	
SODIUM CHLORIDE	MORPHINE	15	24			15 / 39	
SODIUM CHLORIDE	ONDANSETRON	10	27			10 / 37	
	DIPH	3	31			3 / 34	
SODIUM CHLORIDE	ONDANSETRON	4	27			4 / 31	
SODIUM CHLORIDE	KETOROLAC	27				2 / 29	
SODIUM CHLORIDE	ONDANSETRON	10	15			10 / 25	
	ALBUTE. SULFATE	10	15			10 / 25	
IPRATR. BROMIDE	ALBUTE. SULFATE	7	18			7 / 25	
SODIUM CHLORIDE	METOCLOPRAMIDE	25				0 / 25	
OXYCODONE-ACET.	IBUPROFEN	25				0 / 25	

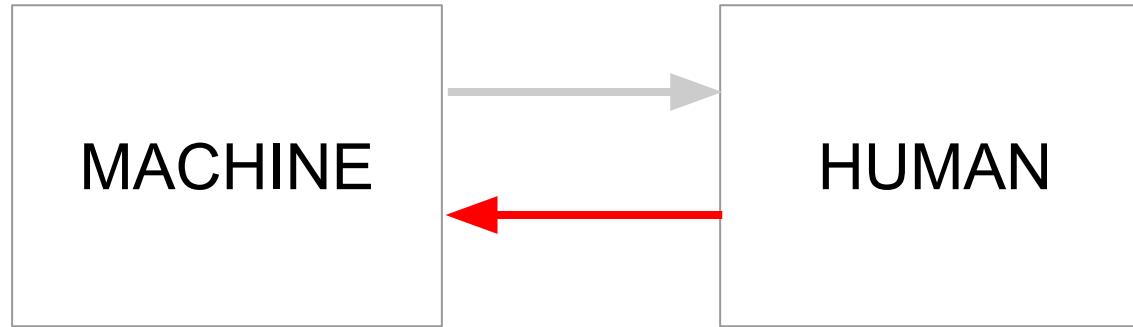


How can human + machine intelligence  
**cooperate** to solve problems together?

Machines provide the information  
we need



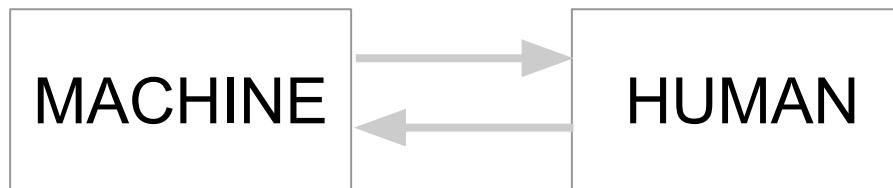
We communicate our knowledge  
and intentions to the machine



# Very Hard!

Output: hard to interpret/verify, uncertain.

Input: complex, ambiguous, small bandwidth.



SearchCoverage:  13966 19348

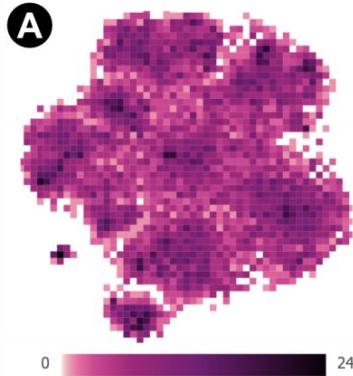
Text

Clear Filters

+ Add Label

Save

Projection LDA Word Si &gt;



car, sale, dealership

car sale dealership manager price  
told vehicle asked buy day

car, time, told

car time told change work will  
day oil tire service

car, great, repair

car great repair guy auto brought  
cost drive mechanic vehicle

car, repair, cost

&lt;

law

**B**

Positive

lawyer	<span style="color: blue;">●</span>	224 / 224
loan	<span style="color: blue;">●</span>	89 / 89
legal	<span style="color: blue;">●</span>	154 / 154
divorce	<span style="color: blue;">●</span>	61 / 61
law	<span style="color: blue;">●</span>	340 / 340
case	<span style="color: blue;">●</span>	652 / 652
firm	<span style="color: blue;">●</span>	266 / 266
court	<span style="color: blue;">●</span>	258 / 258
accident	<span style="color: blue;">●</span>	216 / 216

Documents: 1467 Exclusive: 38%

Info

Intersection

Show:

ALL:  All

ANY:  All

NOT:  All

Keywords

Discriminant

time	recommend
place	experience
great	best
service	friendly
will	going
good	year
staff	work
nice	i'm
day	people
well	price

spa

Documents: 1664 Exclusive: 47%

nail

spa

pedicure

massage

facial

hotel

Documents: 4492 Exclusive: 41%

room

stay

stayed

night

resort

hotel

apartment

Documents: 772 Exclusive: 42%

apartment

rent

tenant

bedroom

resident

haircuts

Documents: 2048 Exclusive: 48%

haircut

salon

hair

cut

barber

cutting

tattoo

Documents: 319 Exclusive: 47%

artist

tattoo

talented

ink

doctor (medical)

Documents: 308 Exclusive: 37%

(doctor OR dr) AND (practice OR prescription OR health

Documents: 19348

im going to give this hotel casino a 2 star. the reason for my 2 star is... the hotel and casino and amenities and restaurants are all good. i had no complaints about that at all. my only gripe...

This hotel has beautiful, lush, green grounds; there were palm trees and other vegetation, some of which had lights that on at night - so pretty. The grounds also had some wildlife.

Absolute worst service department. The only reason I gave them a second chance was purely out of convenience since I live down the street. They dropped my 2014 Honda off of a lif...

&lt; 1 2 3 4 5 ... 194 &gt;

&gt;

**C****D**

Concluding ...

It's crucial to keep in mind the **role humans play**  
on the road to more automation.

# Challenges

Big Data → Big Understanding

How to make sense of machines.

How to cooperate with machines.

# Big Data