Information(?)
~~Data~~ ^ Ethics

Dwight Barry, PhD

Principal Data Scientist

Seattle University, May 2019

Seattle Children's®
HOSPITAL • RESEARCH • FOUNDATION

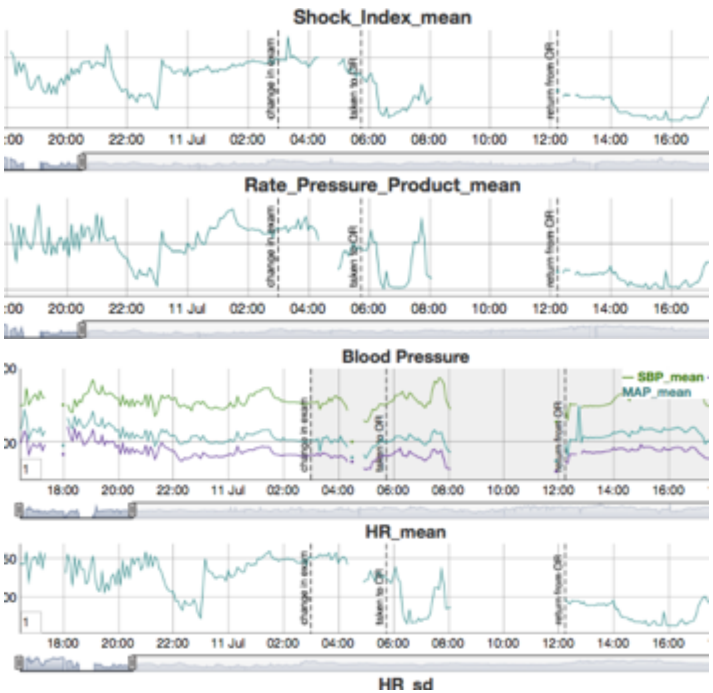
Request for survey responses (*voluntary*)

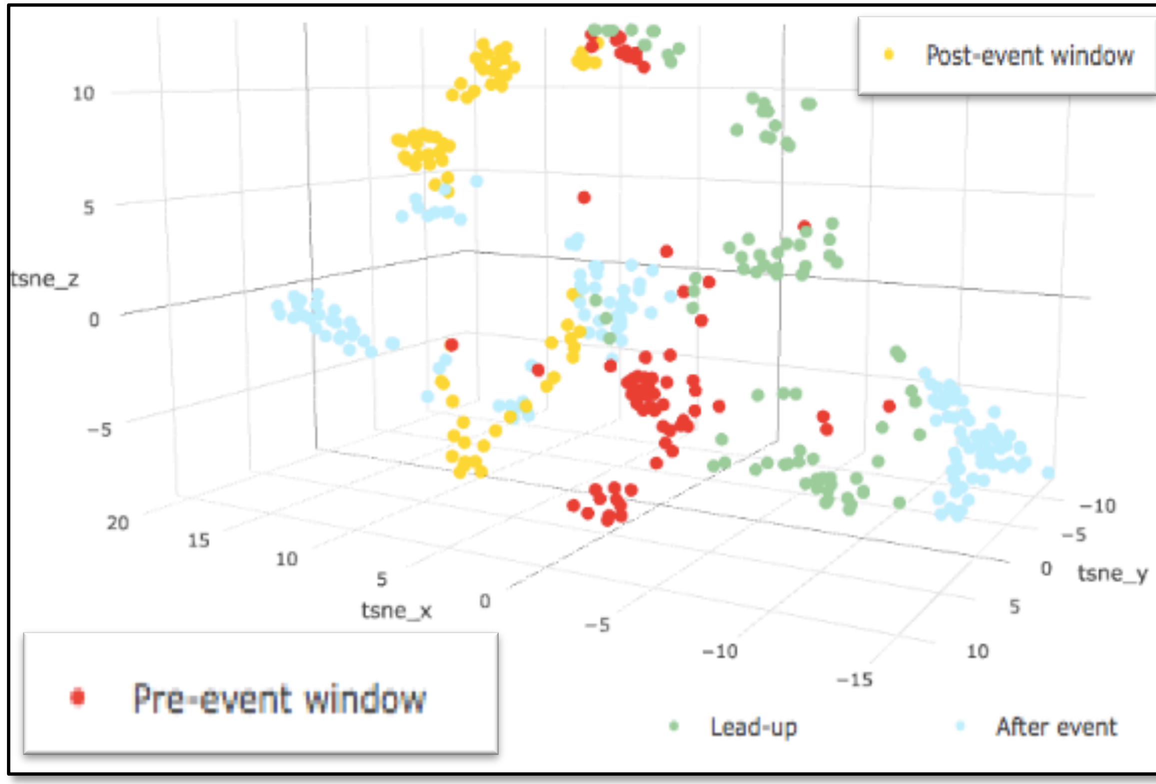
https://is.gd/techethics

Approved by Seattle Children's IRB, STUDY00001849

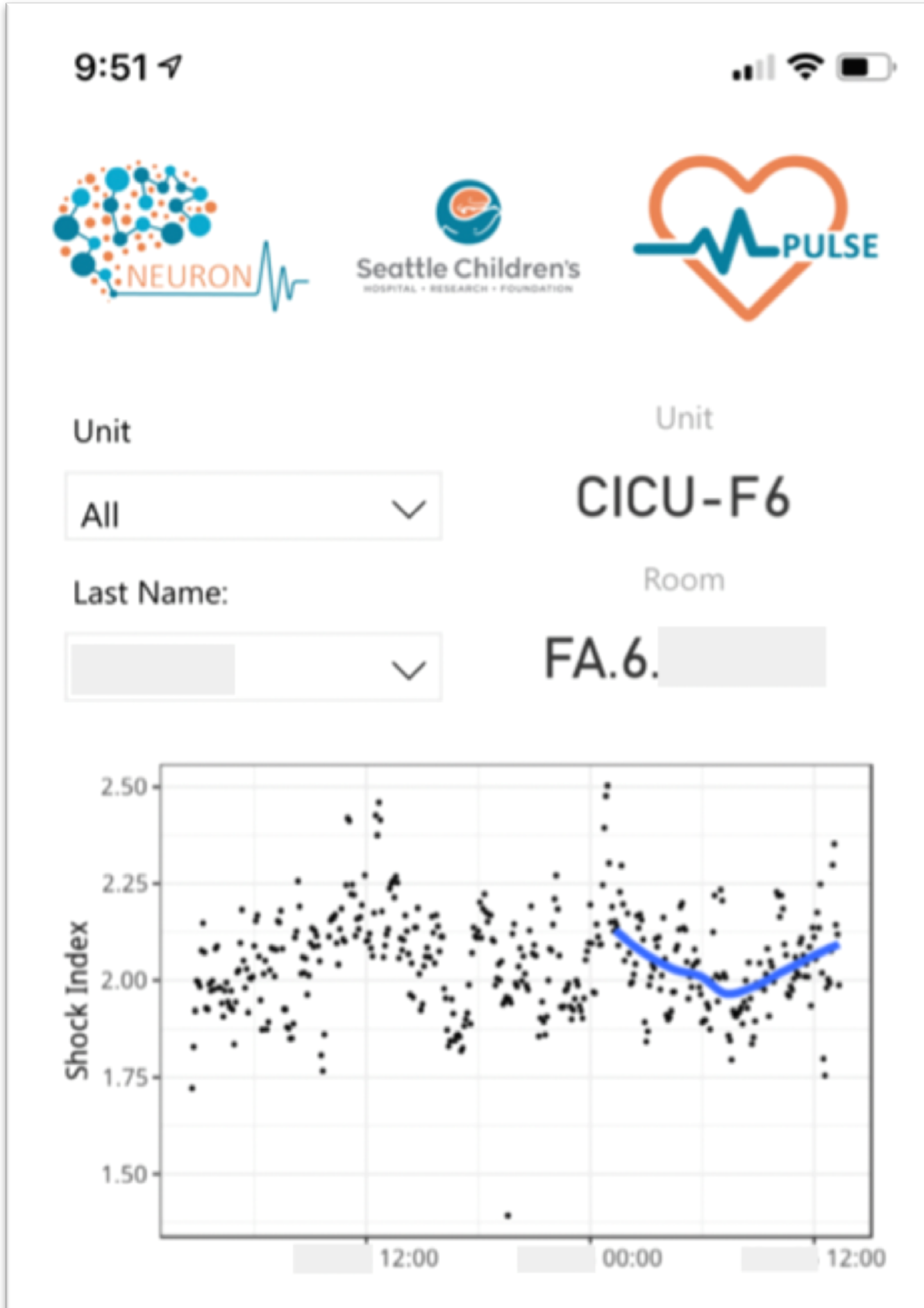
Serves a population of 2.8 million children
$2.3 billion USD annual revenue
10k clinicians & employees


1,000
data points
per patient
per minute
(188)
(170)
104/76 (83)
PHILIPS
SOMANETICS





Goal: Real-time analysis of organ function to prevent critical incidents & assist best possible long-term outcome by supporting brain health


9:51
NEURON
Seattle Children's
HOSPITAL • RESEARCH • FOUNDATION
PULSE
Unit
All
Last Name:
Unit
CICU-F6
Room
FA.6.
2.50
2.25
2.00
1.75
1.50
Shock Index
12:00
00:00
12:00

What comes to mind when you think about ethics?


RUN THIS BY THE LEGAL DEPARTMENT, BUT RUN SUPER FAST SO THE ETHICS DEPARTMENT DOESN'T SEE IT.
DIST. BY CREATORS
SPEEDBUMP.COM
©'10 COVERLY
12-21

Why *data* ethics?

PS Ops: Home

Cincinnati Children's

System State

Housewide

			Inpatient		Psychiatry		
			Staffing	Capacity	Staffing	Capacity	Mitigation
Today 4/1/2016	Night	2					
	Day	1					
		2					
	Evening	1					
		2					
	Night	1					
4/2/2016							
4/3/2016							
4/4/2016							
4/5/2016							
4/6/2016							
4/7/2016							
1-2 Weeks							
3-4 Weeks							
5-6 Weeks							

Unit: PICU

			Staffing	Capacity	State of Unit
Today 4/1/2016	Night	2			
	Day	1			
		2			
	Evening	1			
		2			
	Night	1			
4/2/2016					
4/3/2016					
4/4/2016					
4/5/2016					
4/6/2016					
4/7/2016					
1-2 Weeks					
3-4 Weeks					
5-6 Weeks					

Reports

Daily Summary | IP Microsystem | Staffing | Patient | Forecasts | Utilization | Events | System | Manuals

Events

Event Type	Impact Level	Start	End	Description
Open Overflow	Critical	3/30/2016	4/1/2016	open B6400 Pod
Open Overflow	Critical	3/31/2016	4/1/2016	Opening B6 500 Pod for critical care
Renovation	High	3/7/2016	6/13/2016	Location B Entrance Renovation

System Update Times

0230, 0430, 0600, 0930, 1030, 1130, 1230, 1430, 1530, 1830, 2030, 2230

New Version | My Version: TV0.04 | Current Version: V0.03 | Test Version: TV0.04 | System Messages: No Message.

PS Ops: Housewide Requests

Save! Exit

Cincinnati Children's

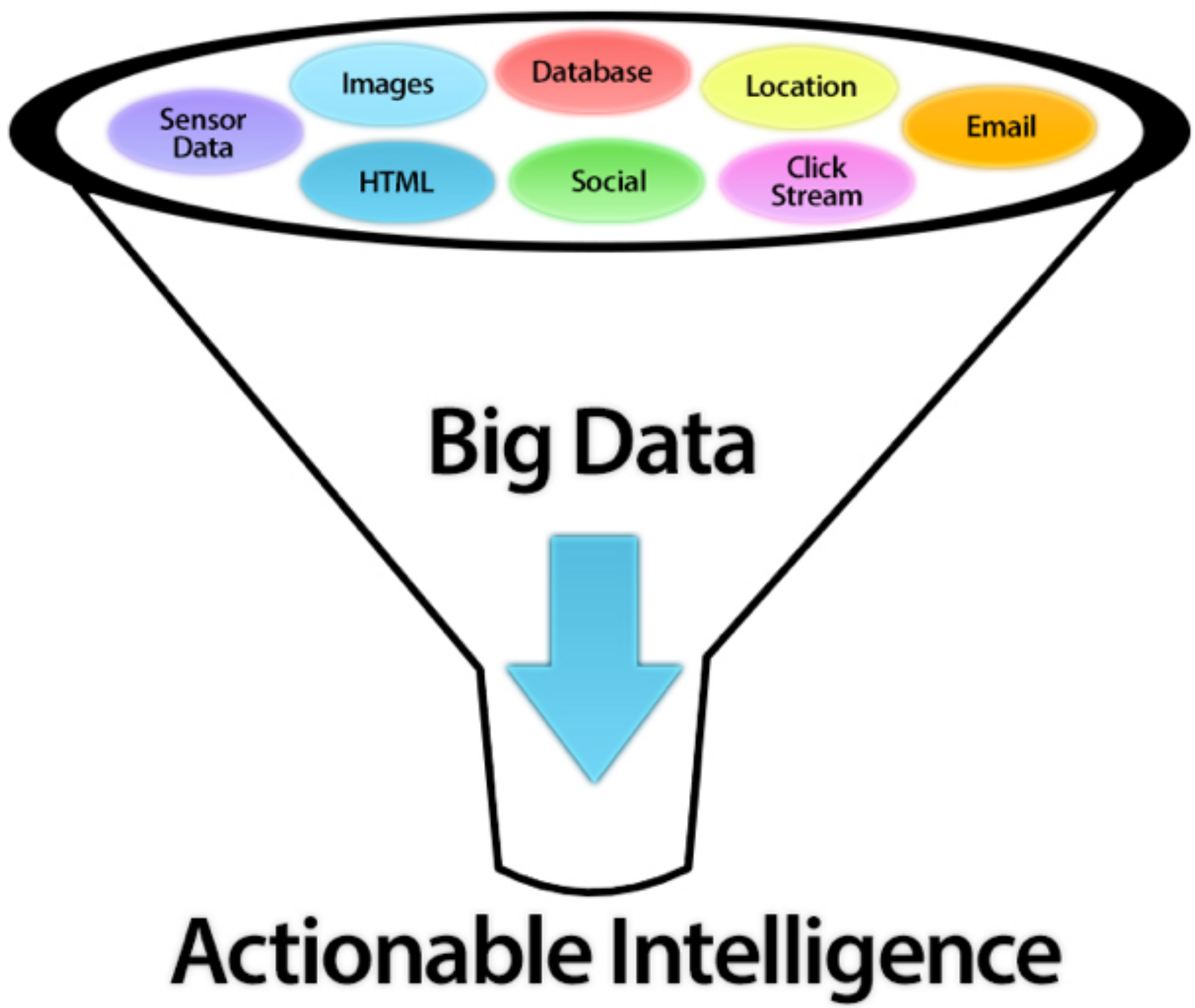
Home **Date:** 4/1/2016 **Shift:** Day 2 Refresh

RN | MHS/MHF | PCA | PCA 1:1 | HUC | Psych | PTO HL | PCA 1:1 Requests | Reports

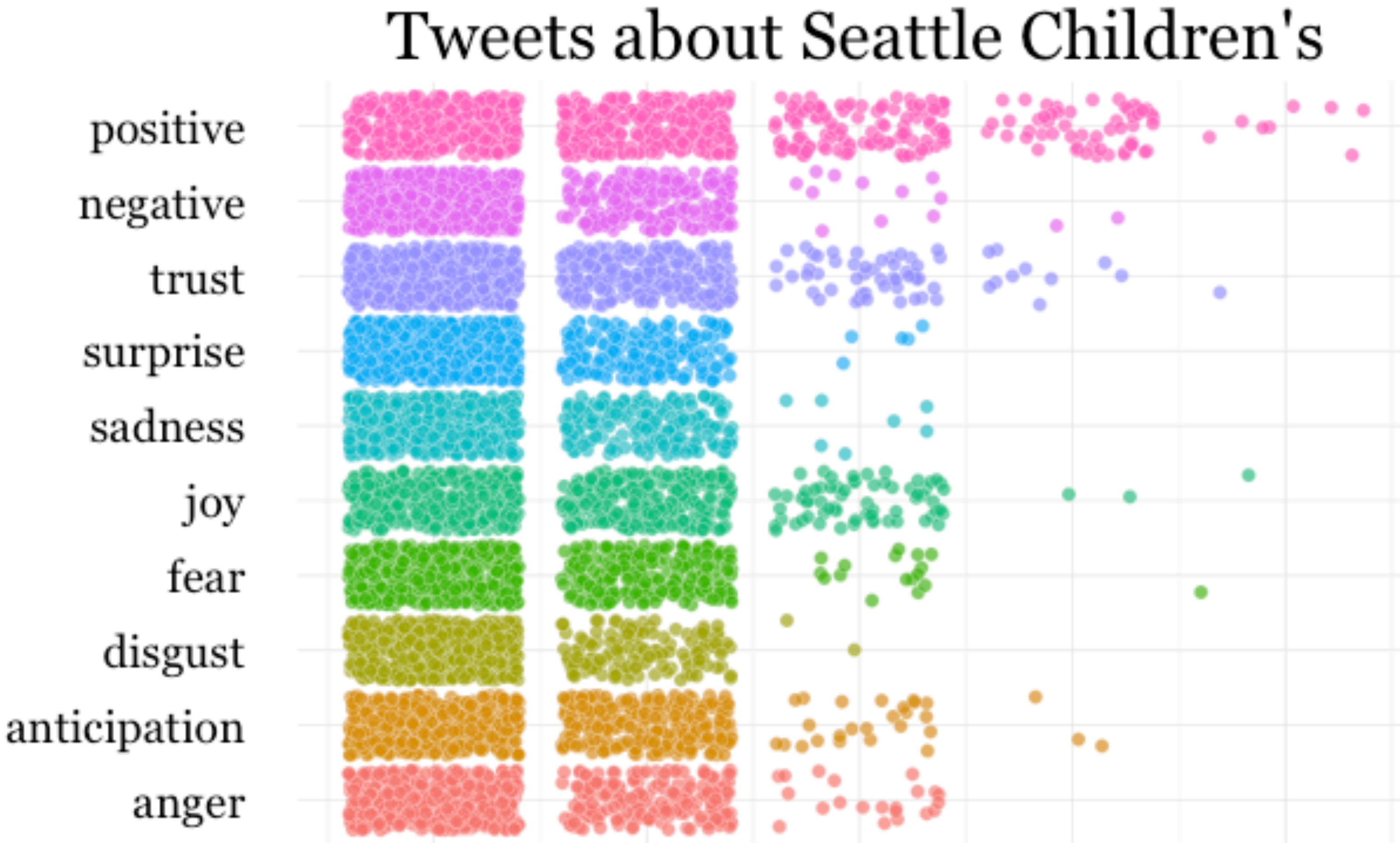
RN (POC)

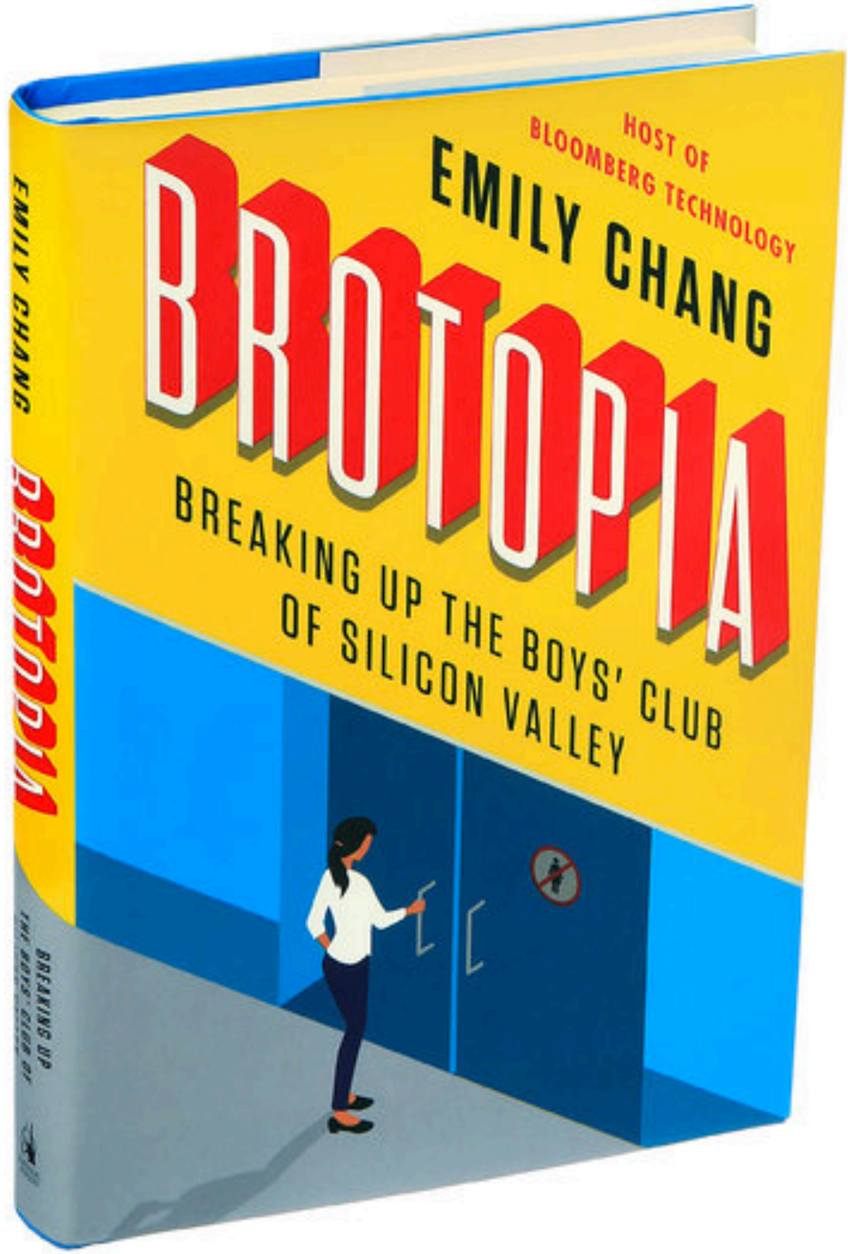
Copy Last Shift Filter: All Main Psych Liberty Sort: Unit Priority

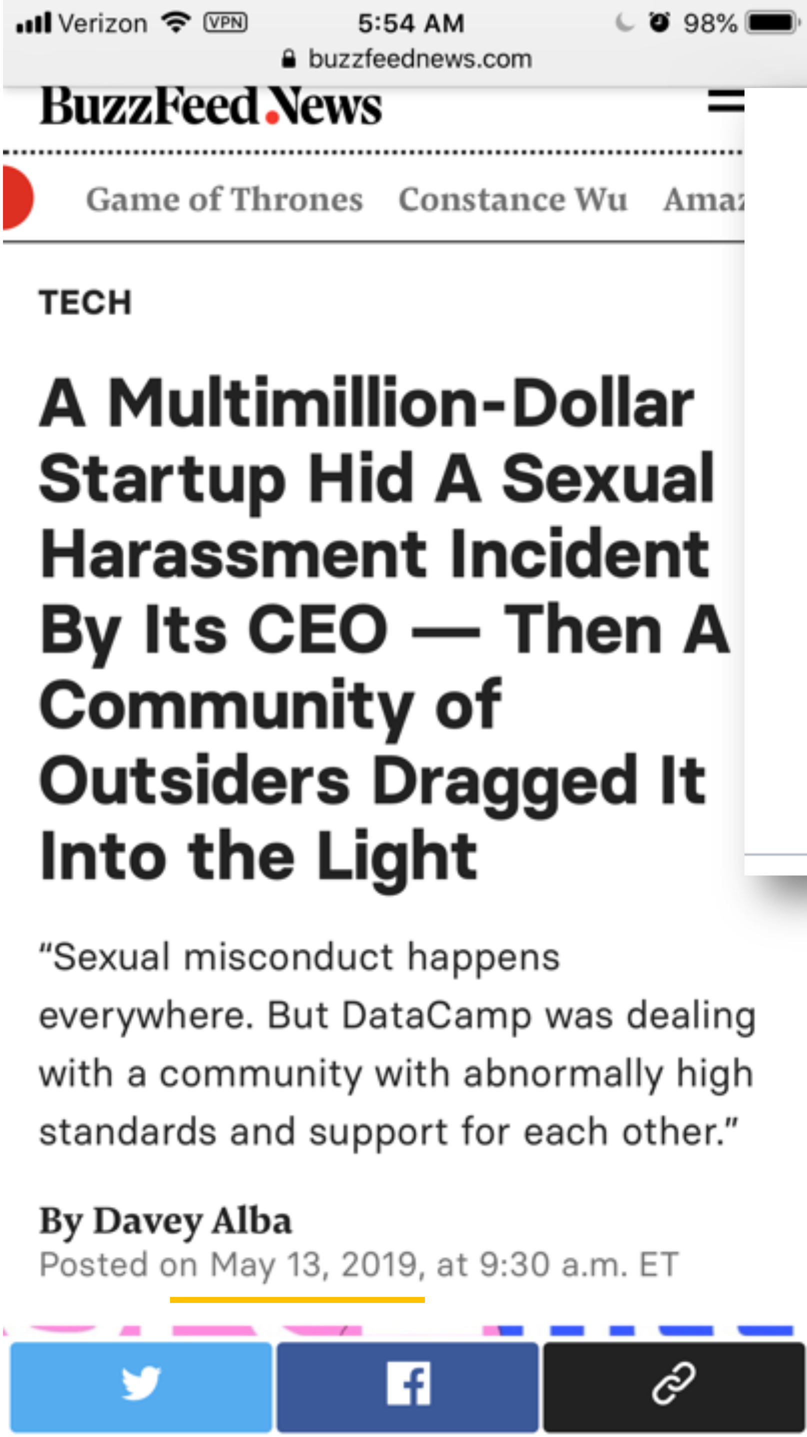
Department				Schedule				Request				Mitigate							
Req	Unit	Occupancy	Acuity	Call Offs	Kro Data	Target	Variance	Request	Extra	Comment	CN UA P A	Pickup	Sent	Rec.	Rec. Dept	Return	Canc.	Var	Comment
	Adol Med A6N	21 88%	10.8	0	1 7	8	-1	0	0			0	0	0		0	[illegible]	-1	
	BMT A5S/A5N	34 94%	16.0	2	1 17	22	-5	0	0			0	0	1		0	0	-4	
	CICU B6HI	22 88%	16.8	0	2 16	18	-2	0	0			0	0	0		0	0	-2	
	Comp Pulm A7C1	8 73%	11.3	0	1 4	3	1	0	0			0	0	0		0	0	1	
	Complex Airway B5CA	9 82%	10.8	2	1 2	4	-2	1	0	511		0	0	1		0	0	-1	
	CST CST	0	0.0	0	0 0	0	0	0	0			0	2	0	LIB A4N, B5C/	0	0	-2	
	ED B1	0	0.0	3	1 16	0	17	1	0			0	0	0		0	0	17	
	Endocrine A7C2	8 89%	10.7	0	1 2	3	-1	0	0			0	0	0		0	0	-1	
	Gen Surg A3N	22 100%	11.5	0	1 7	8	-1	0	0			0	0	0		0	0	-1	
	GI Liver A4N	23 96%	13.5	2	1 8	11	-2	0	0			0	0	1		0	0	-1	
	GI Lumen A4S	17 71%	9.9	0	1 8	6	2	0	1			0	0	0		0	1	1	
	H/O A5C/A5N	30 94%	13.0	0	2 15	15	0	0	0			0	0	0		0	0	0	
	Hosp Med A6S	20 83%	10.2	1	1 6	7	-1	0	0	PTAT 635, 64		0	0	0		0	0	-1	
	IP OF IPU2	10 100%	10.3	0	0 0	0	0	0	0			0	0	2		0	0	2	
	Neuro/Trauma A7NS	32 78%	11.0	0	2 13	12	1	0	0			0	0	0		0	0	1	
	NICU B4	52 88%	13.7	0	3 30	33	-3	2	0			0	0	2		0	0	-1	
	PICU B5CC	33 94%	16.8	2	3 23	28	-5	5	0	includes ove		0	0	5		0	0	0	
	Rehab A4C1	9 90%	10.4	0	1 3	3	0	0	0	470, 474		0	0	0		0	0	0	
	SRU-RN SRU-RN	0	0.0	0	0 9	0	9	0	0			0	9	0		0	0	0	
	TCC A3S	22 92%	21.1	0	1 10	11	-1	0	0	MHS 342; 1:		0	0	0		0	0	-1	
	Tele A6C	12 71%	12.7	0	1 8	6	2	0	0			0	1	0	B4	0	0	1	
	Total	384		12	198	198	8	9	1			0	12	12		0	1	-1	


Sensor Data
Images
Database
Location
Email
HTML
Social
Click Stream
Big Data
Actionable Intelligence

Linked in
skype
foursquare
vimeo
Tube
Insta
facebook
Blogger
my
flickr
twitter
tumblr.
Like
yelp


Tweets about Seattle Children's
positive
negative
trust
surprise
sadness
joy
fear
disgust
anticipation
anger


HOST OF
BLOOMBERG TECHNOLOGY
EMILY CHANG
BROTOPIA
BREAKING UP THE BOYS' CLUB
OF SILICON VALLEY
EMILY CHANG
BROTOPIA
BREAKING UP
THE BOYS' CLUB OF


BuzzFeed News

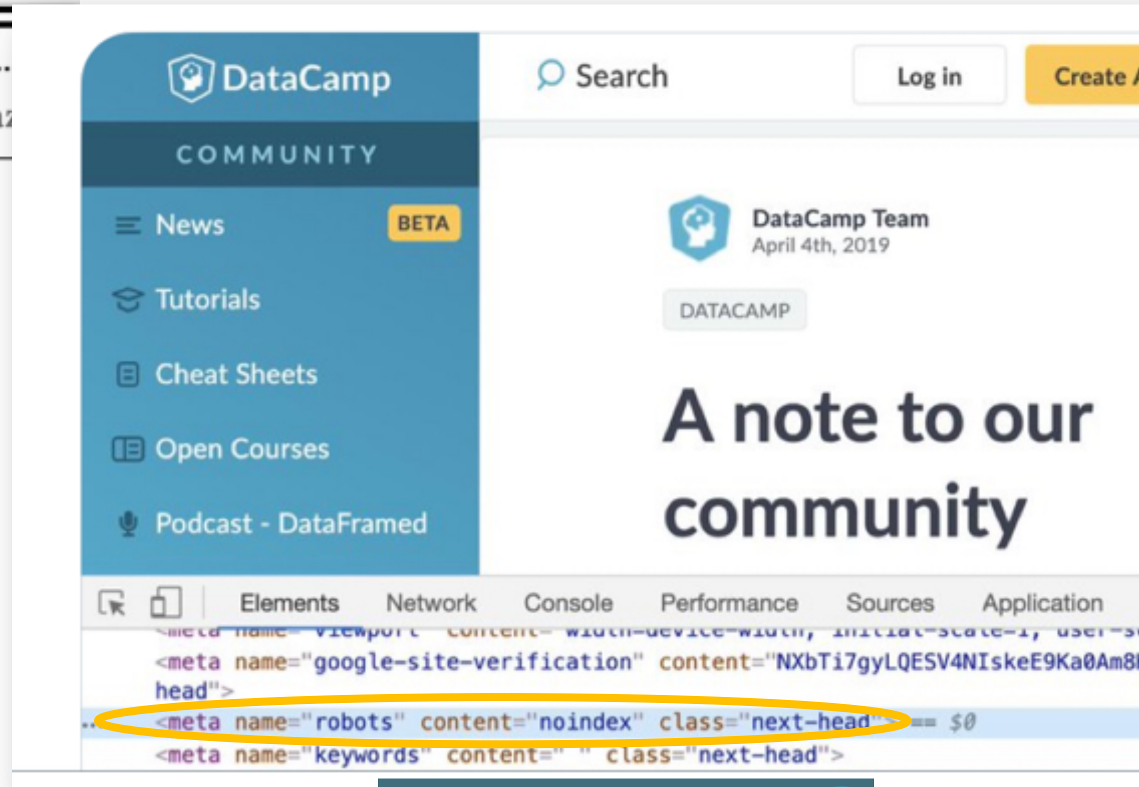
Game of Thrones Constance Wu Amaz

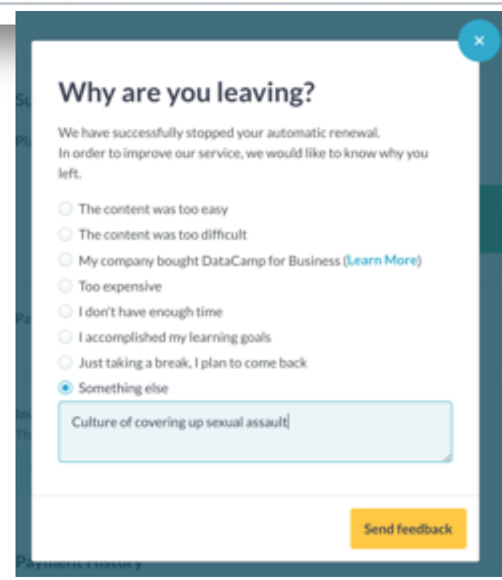
TECH

A Multimillion-Dollar Startup Hid A Sexual Harassment Incident By Its CEO — Then A Community of Outsiders Dragged It Into the Light

"Sexual misconduct happens everywhere. But DataCamp was dealing with a community with abnormally high standards and support for each other."

By Davey Alba
Posted on May 13, 2019, at 9:30 a.m. ET





Seattle Children's
HOSPITAL • RESEARCH • FOUNDATION

Treuman Katz Center for Pediatric Bioethics

Clinical Bioethics Consult Note
Big data predictive modeling on hospital staff

Patient Name:

DOB:

MR#:

Consult Date: 12/09/2016

Attending Consultant: Shah, Seema

Secondary Consultant: Tate, Tyler

Requester: Barry, Dwight

Requester's Service: Enterprise Analytics

Reason for Consult:
To address the ethics of using predictive analytics (or "big data") as an evaluative and predictive tool within a healthcare organization.

Other Issues Identified:
None

Background:
Dwight Barry, lead data scientist at Seattle Children's Hospital, requested an ethics consult to help him understand and evaluate the ethical aspects of what he calls "people analytics," or using data to predict things about individuals within the organization. A more common term for this is "big data", which is defined by the Oxford English Dictionary as "Extremely large

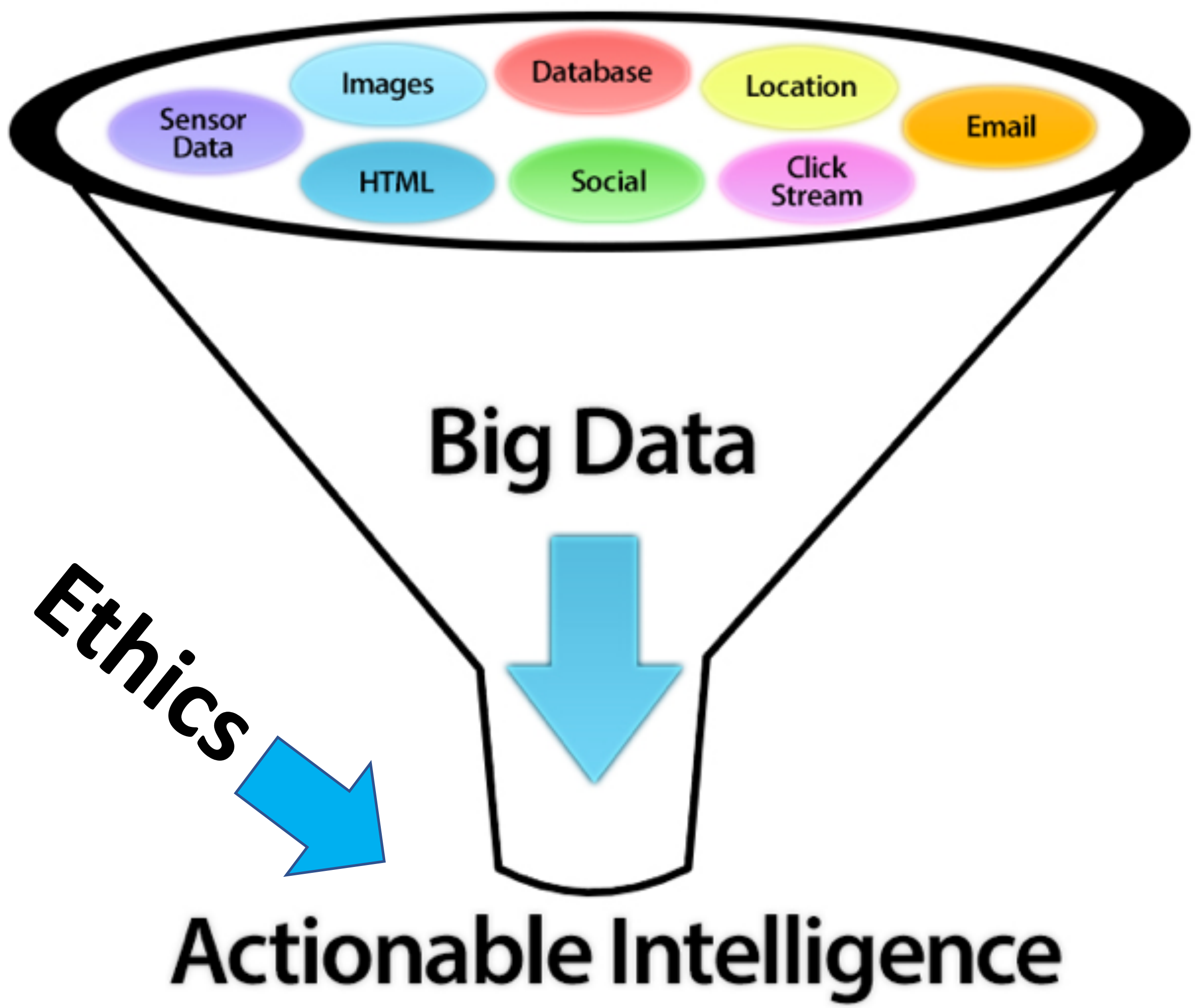
4. Training

Both the methodology/mechanics and the ethics of Big Data use are rapidly developing and constantly changing. Those working in data analysis may not have the ethics training to identify emerging issues or know when to call for a consult. Others within the organization may lack the technical training to understand what efforts are being undertaken and what potential issues might arise. It could therefore be beneficial to have planned seminars or lectures that could train large numbers of people to be able to both understand how big data actually works and to train data analysts in how to identify ethical issues that may arise in their work. One way to do this may be to have lectures or seminars open to the Seattle Children's community on the types of ethical issues that may arise, such as an ethics grand rounds presentation with an invited speaker. Another possibility is for a joint effort between an ethicist and data analyst to create a short annual training course for analysts and data-oriented employees on predictive modeling ethics. We are also happy to continue to be involved in training efforts or if further consultation is necessary.

organization.

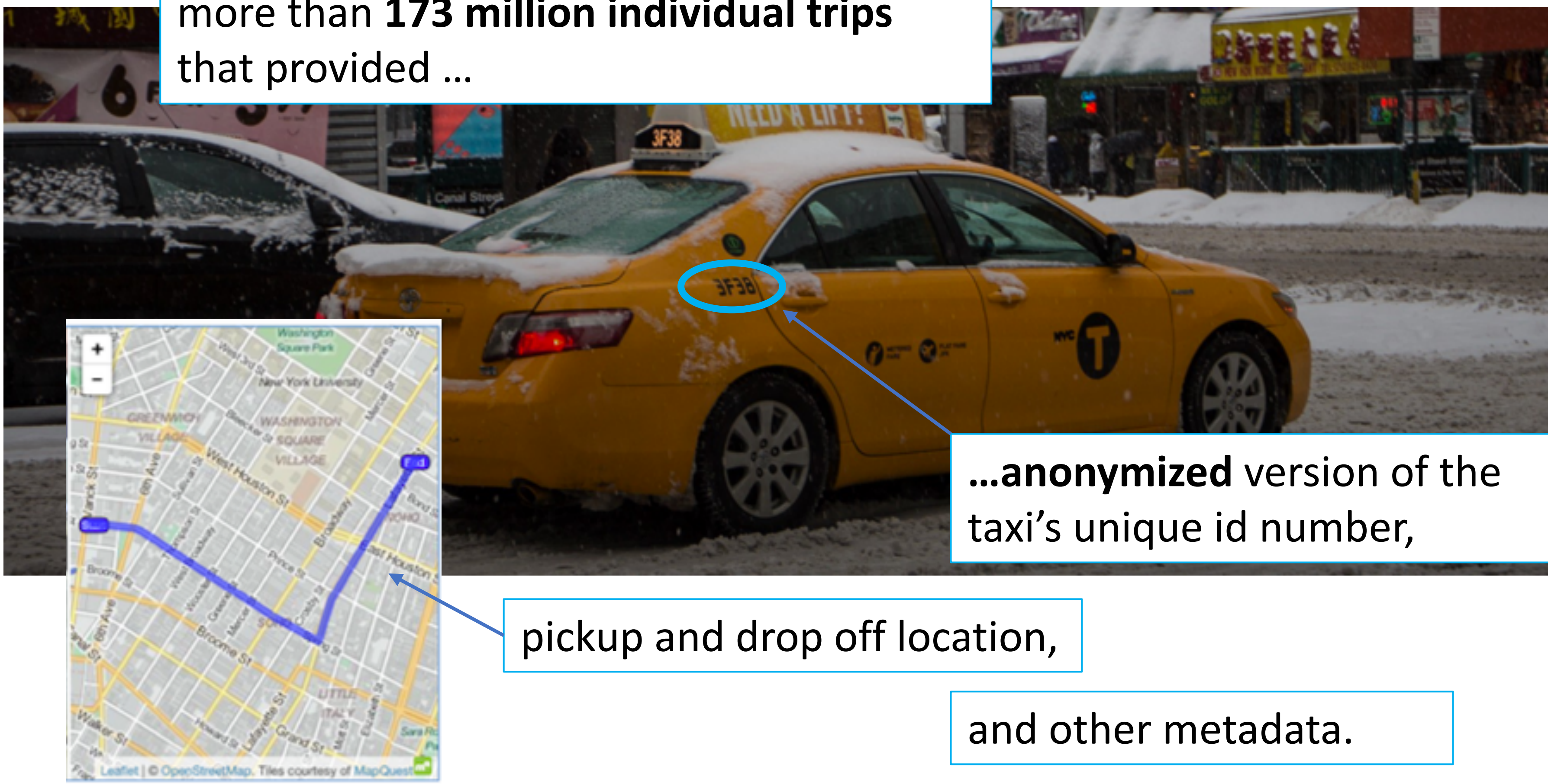
Other Issues Identified:
None

Background:
Dwight Barry, lead data scientist at Seattle Children's Hospital, requested an ethics consult to help him understand and evaluate the ethical aspects of what he calls "people analytics," or using data to predict things about individuals within the organization. A more common term for this is "big data", which is defined by the Oxford English Dictionary as "Extremely large


Images
Database
Location
Sensor Data
Email
HTML
Social
Click Stream
Big Data
Ethics
Actionable Intelligence

NYC Taxi Database

20GB+ of uncompressed data comprising more than **173 million individual trips** that provided …

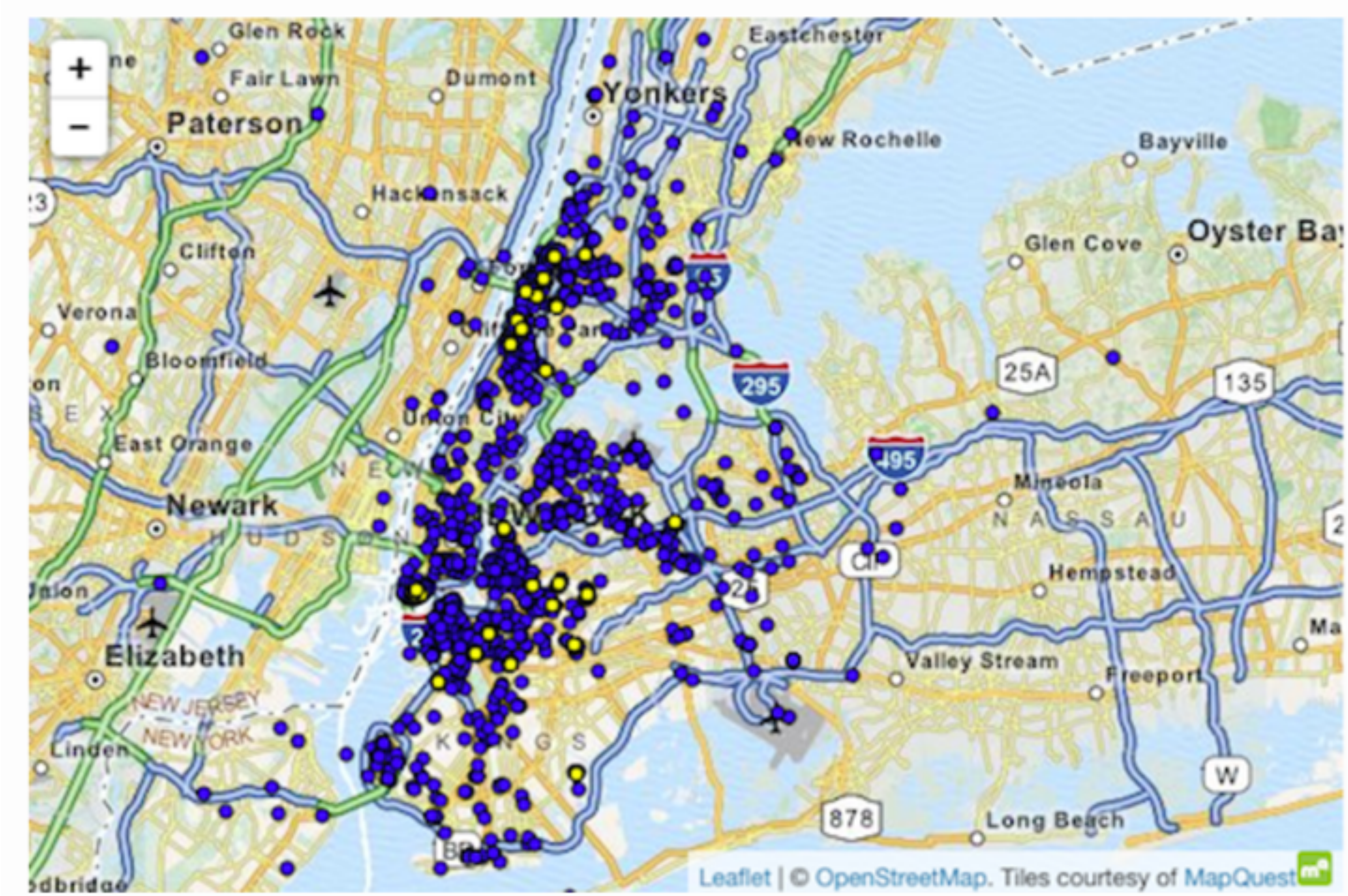


…anonymized version of the taxi’s unique id number,

pickup and drop off location,

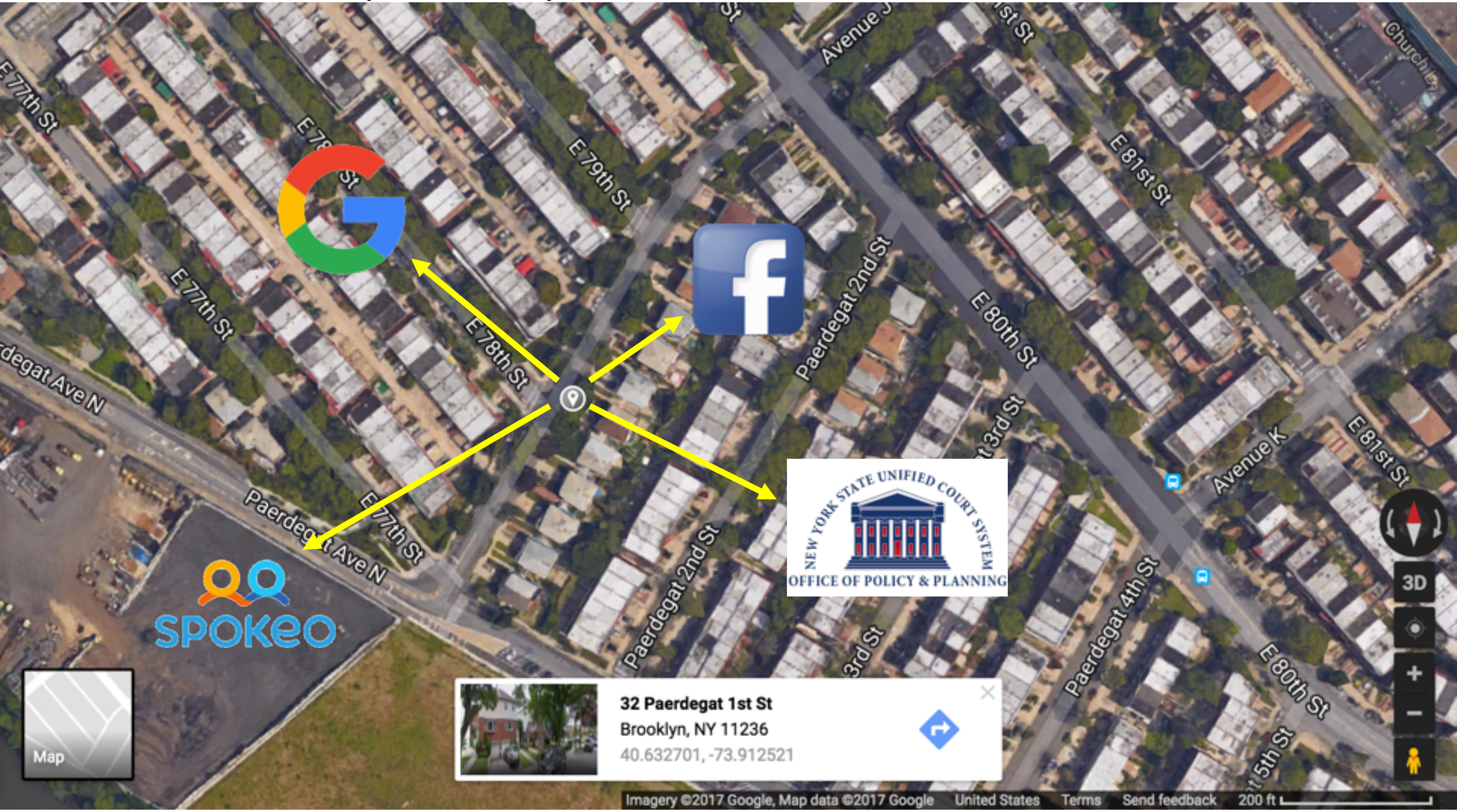
and other metadata.

Drop off locations for …



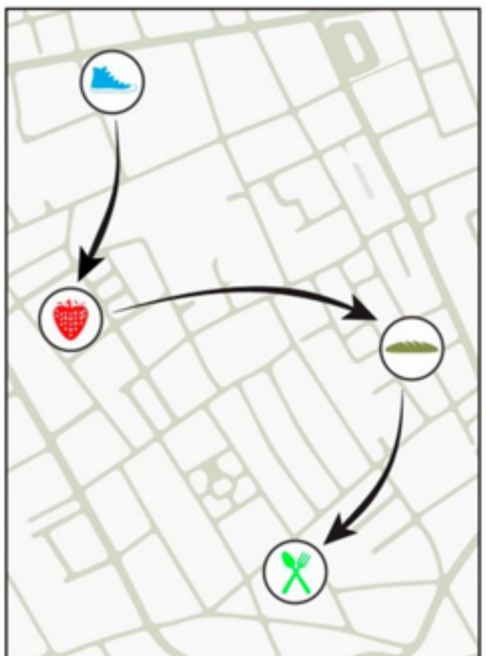
NOT AN ACTUAL DROP OFF!

(for example only)



Unique in the shopping mall: On the reidentifiability of credit card metadata

Yves-Alexandre de Montjoye,[1*] Laura Radaelli,[2] Vivek Kumar Singh,[1,3] Alex "Sandy" Pentland[1]



shop	user_id	time	price	price_bin
	7abc1a23	09/23	$97.30	$49 – $146
	7abc1a23	09/23	$15.13	$5 – $16
	3092fc10	09/23	$43.78	$16 – $49
	7abc1a23	09/23	$4.33	$2 – $5
	4c7af72a	09/23	$12.29	$5 – $16
	89c0829c	09/24	$3.66	$2 – $5
	7abc1a23	09/24	$35.81	$16 – $49

Fig. 1 Financial traces in a simply anonymized data set such as the one we use for this work.
Arrows represent the temporal sequence of transactions for user 7abc1a23 and the prices are grouped in bins of increasing size (29).

Published at the 16th ACM Workshop on Privacy in the Electronic Society (WPES 2017). More information at https://adint.cs.washington.edu/.

Exploring ADINT: Using Ad Targeting for Surveillance on a Budget — or — How Alice Can Buy Ads to Track Bob

Paul Vines, Franziska Roesner, and Tadayoshi Kohno
Paul G. Allen School of Computer Science & Engineering, University of Washington
plvines@cs.washington.edu,franzi@cs.washington.edu,yoshi@cs.washington.edu

…we find that — for $1000 USD — **we can track the location of individuals who are using apps served by that advertising network**, as well as infer whether they are using potentially sensitive applications (e.g., certain religious or sexuality-related apps)

The New York Times

What Women Know About the Internet

The digital world is not designed to keep women safe. New regulations should be.

By Emily Chang
Ms. Chang, an anchor at Bloomberg TV, is the author of “Brotopia.”

April 10, 2019

411

The Department of Justice reports that about 75 percent of the victims of stalking and cyberstalking are women.

And so women look over our shoulders online, just as we do in real life.

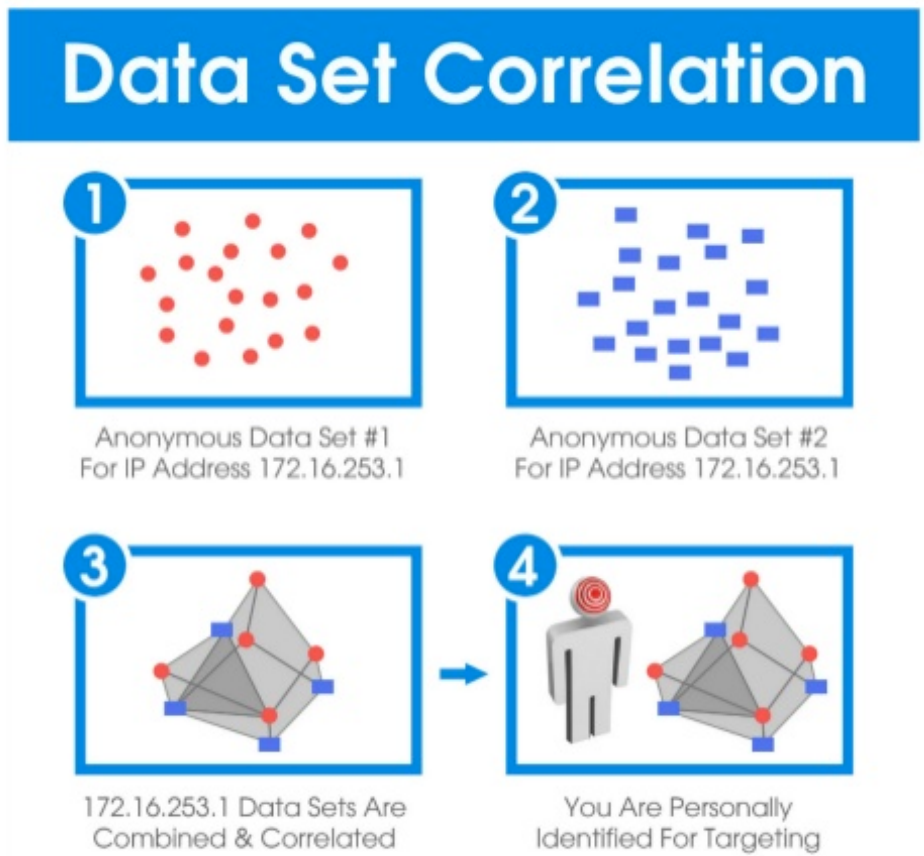
REGULATION

There's No Such Thing as Anonymous Data

by Scott Berinato

FEBRUARY 09, 2015

Harvard Business Review



FEB 16, 2012 @ 11:02 AM 3,203,944

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill, FORBES STAFF

Welcome to The Not-So Private Parts where technology & privacy collide **FULL BIO**

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target TGT +0.52%, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

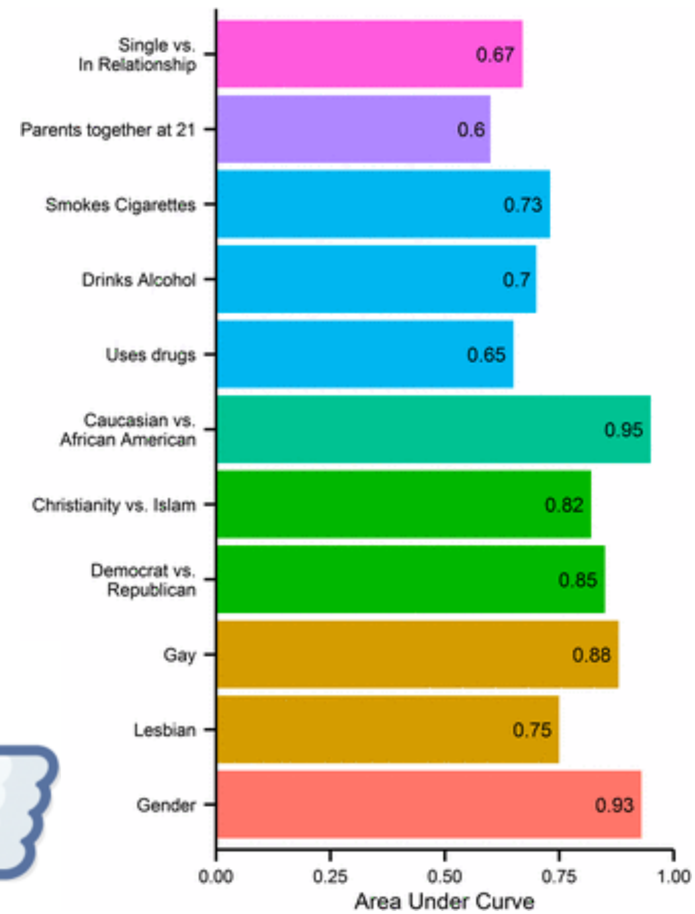
Target has got you in its aim

Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski[a,1], David Stillwell[a], and Thore Graepel[b]

[a]Free School Lane, The Psychometrics Centre, University of Cambridge, Cambridge CB2 3RQ United Kingdom; and [b]Microsoft Research, Cambridge CB1 2FB, United Kingdom

Abstract: We show that easily accessible digital records of behavior, **Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes** including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender.



Identifying Relevant Users and Groups in the Context of Credit Analysis Based on Data from Twitter

Danyllo W. A., Alisson V. B., Alexandre N. D.
Moacir L.M. J., Jansepetrus B. P.
Center of Informatics – Federal University of Paraiba.
João Pessoa, PB – Brazil.
danyllowagneralbuquerque, alisson.brito,
alexandrend, moacir.lopes.jr, jansebp}@gmail.com

Roberto Felício Oliveira
Department of Informatics – State University of Goias.
Posse, GO - Brazil
Prof.roberto.posse@gmail.com

Fig. 6. Amount of negative certificates from the neighborhood of 26 users.

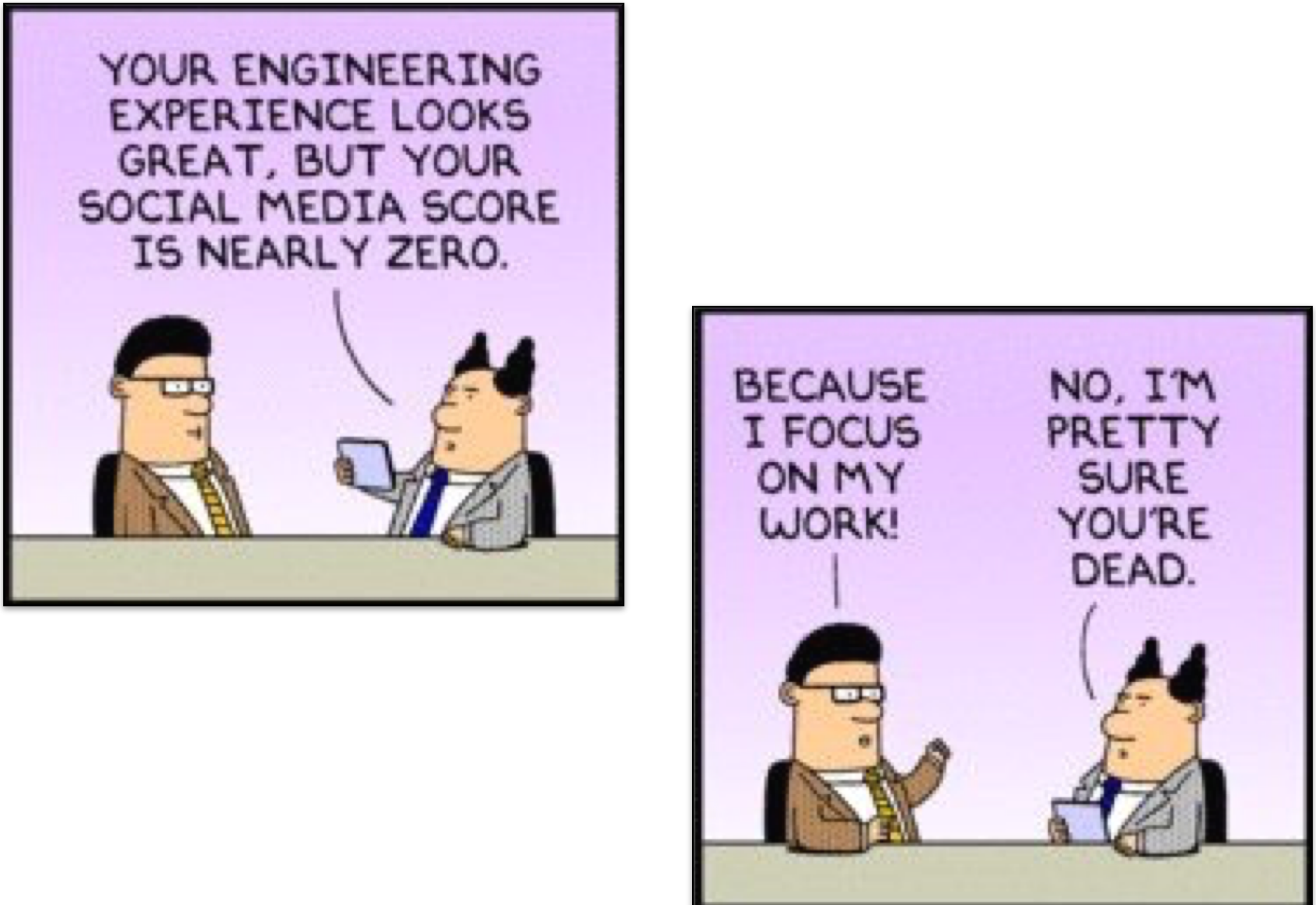
Preliminary conclusions in this study indicate that among the 26 found communities (the 26 initial users and their followers), those with the highest number of people with credit restrictions are communities where the central node has some constraint credit associated.

Abstract: … this study collected data from the social network Twitter, and compared them with data from a financial institution in order to model the network and analyze their similarities. Three thousand users from Twitter were analyzed and 504 matched with the database from company for credit analysis. The results demonstrated that most of those users have more credit restriction than their neighbors, and users with no restrictions normally have also neighborhoods with no credit restriction as well.

TIME

93% of hiring managers will review a candidate's social profile before making a hiring decision.

October 2014


YOUR ENGINEERING EXPERIENCE LOOKS GREAT, BUT YOUR SOCIAL MEDIA SCORE IS NEARLY ZERO.
BECAUSE I FOCUS ON MY WORK!
NO, I'M PRETTY SURE YOU'RE DEAD.

Will My Next Car Be a **Libertarian** or a **Utilitarian?**

Who Will Decide?

Tom Fournier

GrendelKhan/WIKIMEDIA

Change is coming.

Experimental, self-driving cars are plying public roads in many U.S. states, heralding what some automotive industry experts and regulators see as a profound and imminent disruption in the transportation industry, and a change to our way of life. With this change, a large swath http://moralmachine.mit.edu/

MORAL MACHINE

Home Judge Classic Design Browse About Feedback En

By Grendelkhan (Own work) [CC BY-SA 4.0 (http://creativecommons.org/licenses/by-sa/4.0)], via Wikimedia Commons

Welcome to the Moral Machine! A platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars.

We show you moral dilemmas, where a driverless car must choose the lesser of two evils, such as killing two passengers or five pedestrians. As an outside observer, you **judge** which outcome you think is more acceptable. You can then see how your responses compare with those of other people.

The New York Times

Teaching A.I. Systems to Behave Themselves

Geoffrey Irving, left, and Dario Amodei demonstrate how simple video games are used to train A.I. bots.
CHRISTIE HEMM KLOK FOR THE NEW YORK TIMES

By CADE METZ

AUGUST 13, 2017

arXiv, *November 2016*

Automated Inference on Criminality using Face Images

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

Figure 1. Sample ID photos in our data set.

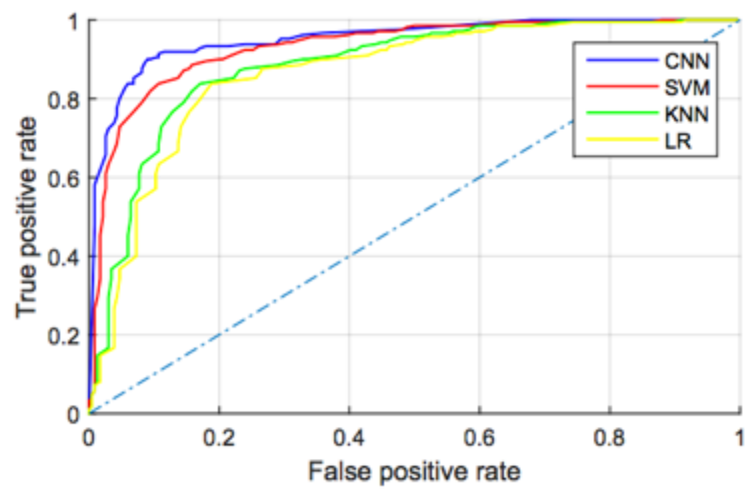


Figure 5. The ROC curves of the four tested binary face classifiers on criminality.

arXiv, *November 2016*

Automated Inference on Criminality using Face Images

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

"We are the first to study automated ***face-induced inference on criminality*** free of any biases of subjective judgments of human observers."

ssifiers on criminality.

Figure 1. Sample ID photos in our data set.

At the Kaiser Wilhelm Institute for Anthropology, Human Genetics, and Eugenics, **a racial hygienist** measures a woman's features in an attempt to determine her racial ancestry. Berlin, Germany, date uncertain.

— *National Archives and Records Administration, College Park, Md.*

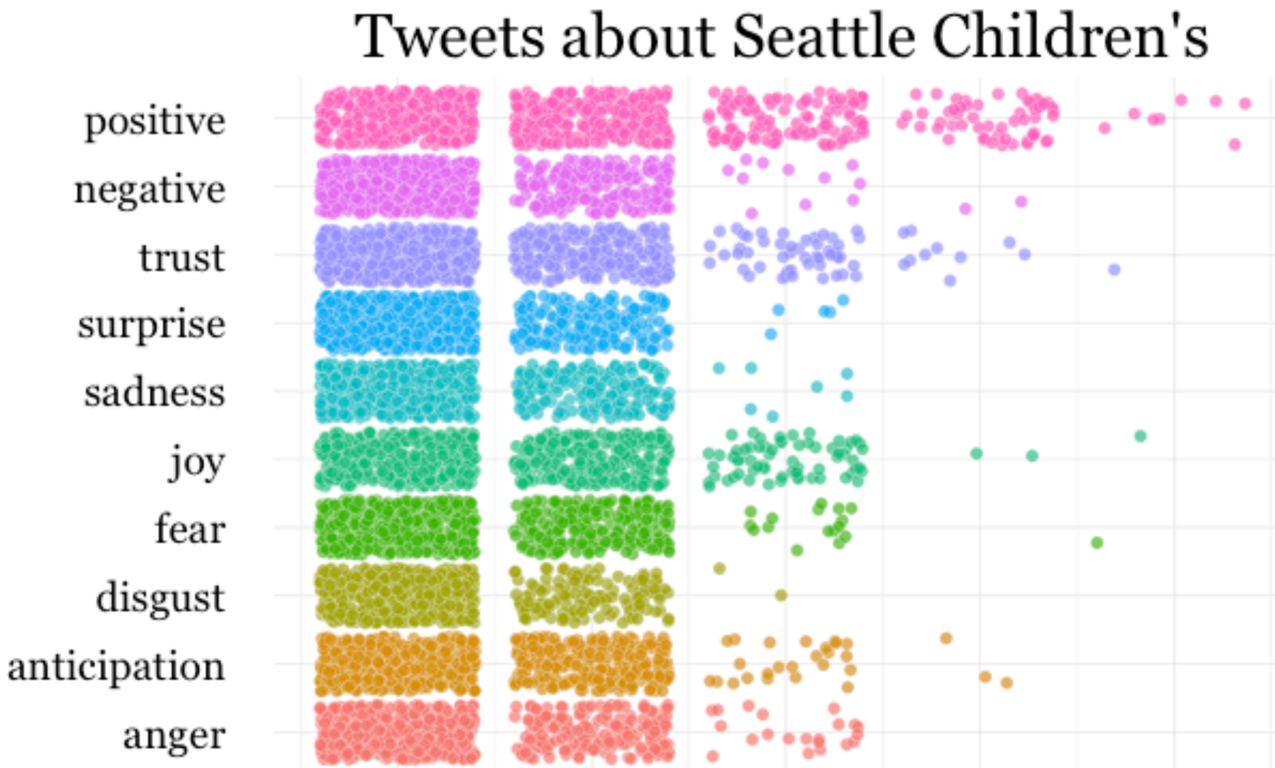
Moritz Hardt Follow
Researcher. Machine learning, optimization, privacy and social questions in computation.
Sep 26, 2014 · 8 min read

How big data is unfair

Understanding unintended sources of unfairness in data driven decision making

As we're on the cusp of using machine learning for rendering basically all kinds of consequential decisions about human beings in domains such as education, employment, advertising, health care and policing, it is important to understand why *machine learning is not, by default, fair or just in any meaningful way.*

	sentiment
fidget	-9.931679
interrupt	-9.634706
staunchly	1.466919
imaginary	-2.989215
taxing	0.468522
world-famous	6.908561
low-cost	9.237223
disapointment	-8.737182
totalitarian	-10.851580
bellicose	-8.328674
freezes	-8.456981
sin	-7.839670
fragile	-4.018289
fooled	-4.309344
undecided	-2.816172
handily	2.339609
demonizes	-2.102152
easygoing	8.747150
unpopular	-7.887475
commiserate	1.790899



```
In [12]: text_to_sentiment("this example is pretty cool")
Out[12]: 3.889968926086298

In [13]: text_to_sentiment("this example is okay")
Out[13]: 2.7997773492425186

In [14]: text_to_sentiment("meh, this example sucks")
Out[14]: -1.1774475917460698
```

```
In [15]: text_to_sentiment("Let's go get Italian food")

Out[15]: 2.0429166109408983

In [16]: text_to_sentiment("Let's go get Chinese food")

Out[16]: 1.4094033658140972

In [17]: text_to_sentiment("Let's go get Mexican food")

Out[17]: 0.38801985560121732
```

```
In [18]: text_to_sentiment("My name is Emily")
Out[18]: 2.2286179364745311

In [19]: text_to_sentiment("My name is Heather")
Out[19]: 1.3976291151079159

In [20]: text_to_sentiment("My name is Yvette")
Out[20]: 0.98463802132985556

In [21]: text_to_sentiment("My name is Shaniqua")
Out[21]: -0.47048131775890656
```

Division of Criminal Justice Services
Office of Justice Research and Performance

Criminal Justice Research Report

Andrew M. Cuomo
Governor

Michael C. Green
Executive Deputy Commissioner

September 2012

New York State COMPAS-Probation Risk and Need Assessment Study: Examining the Recidivism Scale's Effectiveness and Predictive Accuracy

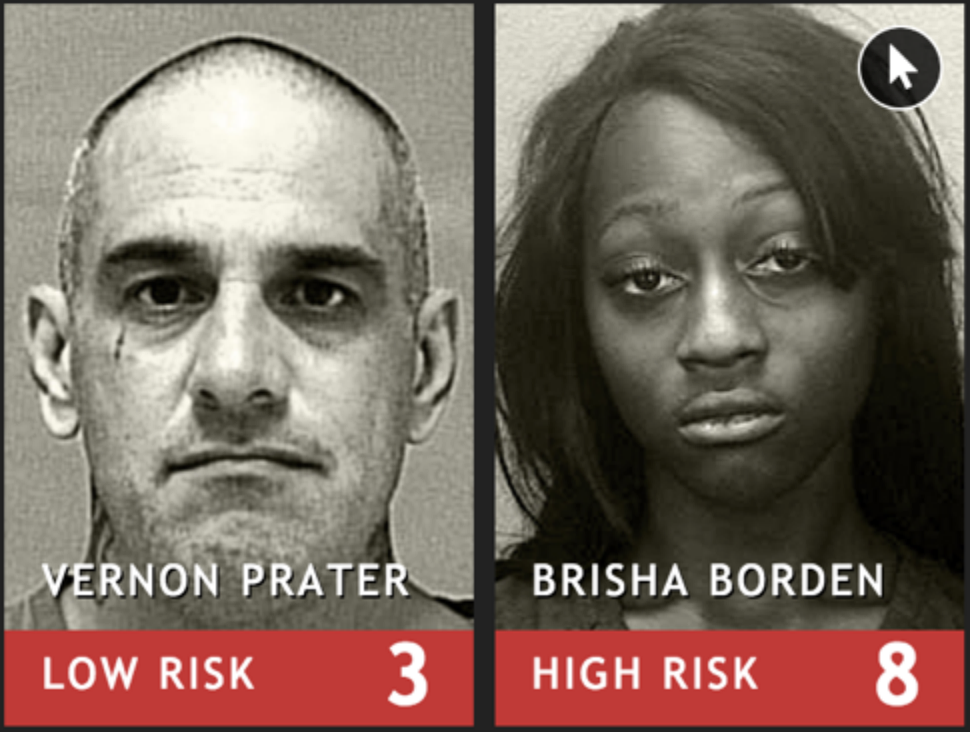
Machine Bias

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

Two Petty Theft Arrests

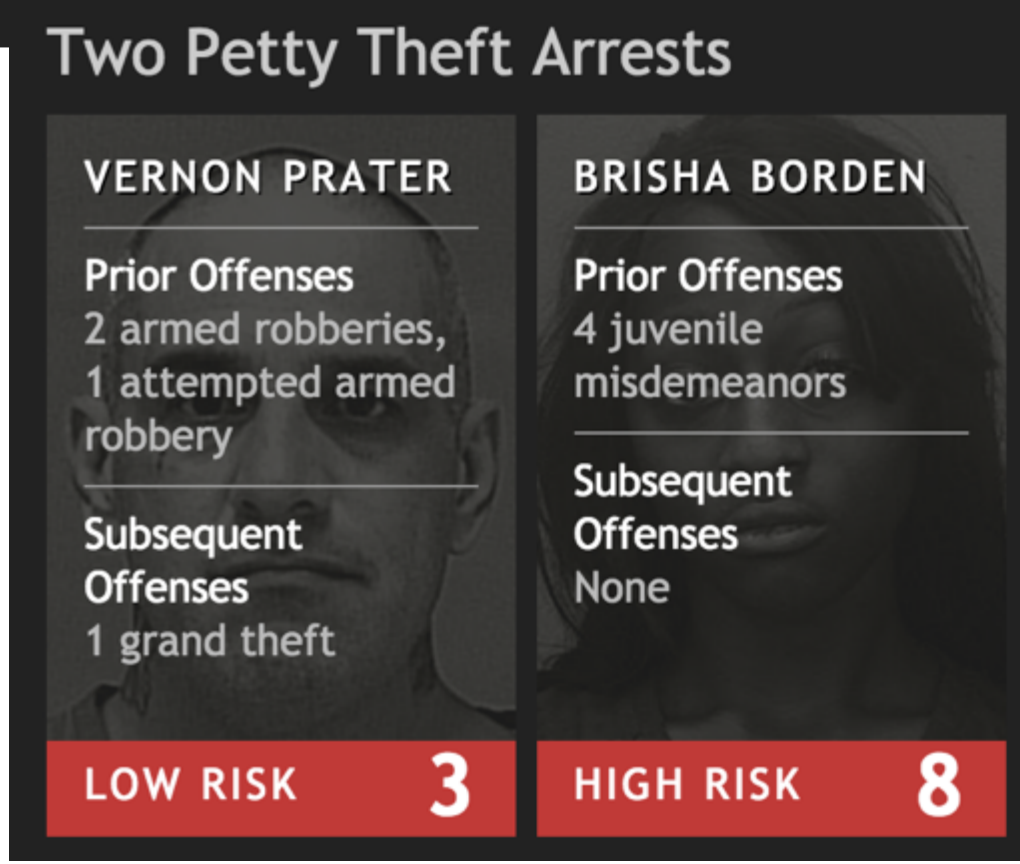


Machine Bias

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



HOME / BLOG / RESPO

RESPONSE TO PROPUBLICA: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY

The website ProPublica recently published a story that focused on the scientific validity of COMPAS, raising questions about racial bias. As a result of the article and the subsequent national attention that it garnered, Northpointe launched an in-depth analysis of the data samples used by ProPublica. Drawing from the results of our analysis of ProPublica's data, Northpointe unequivocally rejects the ProPublica conclusion of racial bias in the COMPAS risk scales.

https://arxiv.org/abs/1609.05807

Inherent Trade-Offs in the Fair Determination of Risk Scores

Jon Kleinberg * Sendhil Mullainathan † Manish Raghavan ‡

Recent discussion in the public sphere about algorithmic classification has involved tension between competing notions of what it means for a probabilistic classification to be fair to different groups. **We formalize three fairness conditions** that lie at the heart of these debates, and we prove that except in highly constrained special cases, **there is no method that can satisfy these three conditions simultaneously**.

https://research.google.com/bigpicture/attacking-discrimination-in-ml/

Attacking discrimination with smarter machine learning

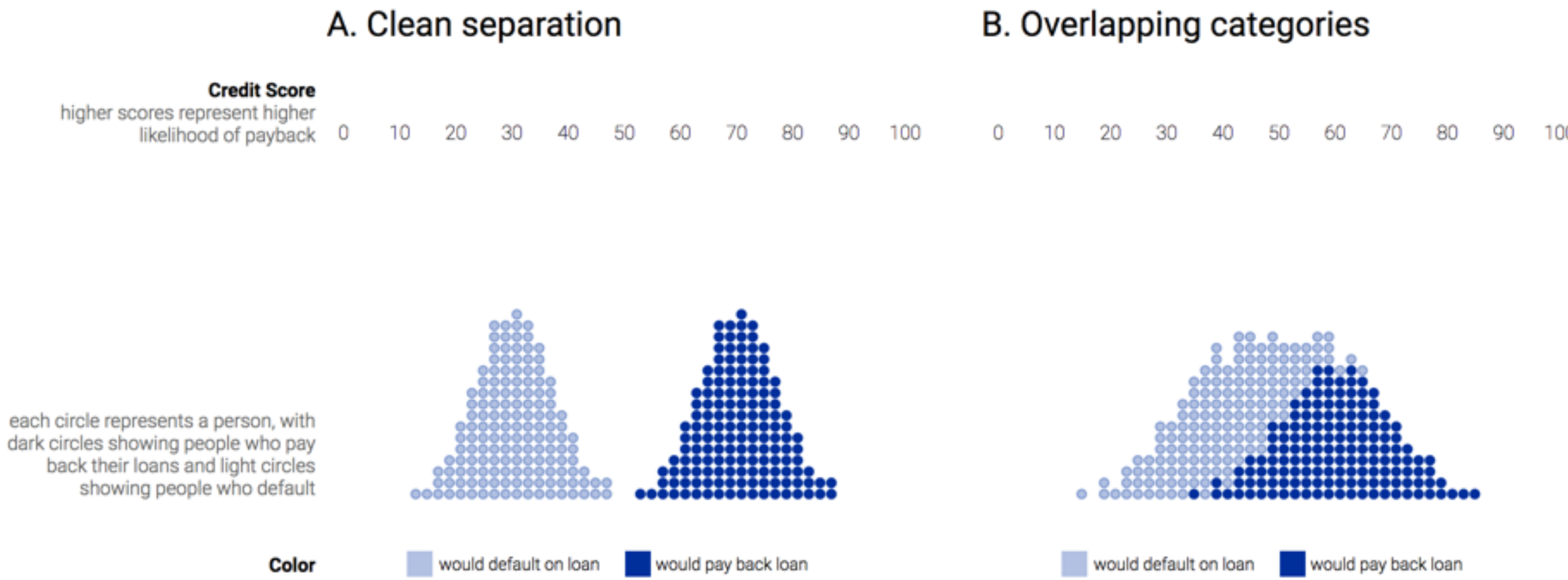
As machine learning is increasingly used to make important decisions across core social domains, the work of ensuring that these decisions aren't discriminatory becomes crucial.

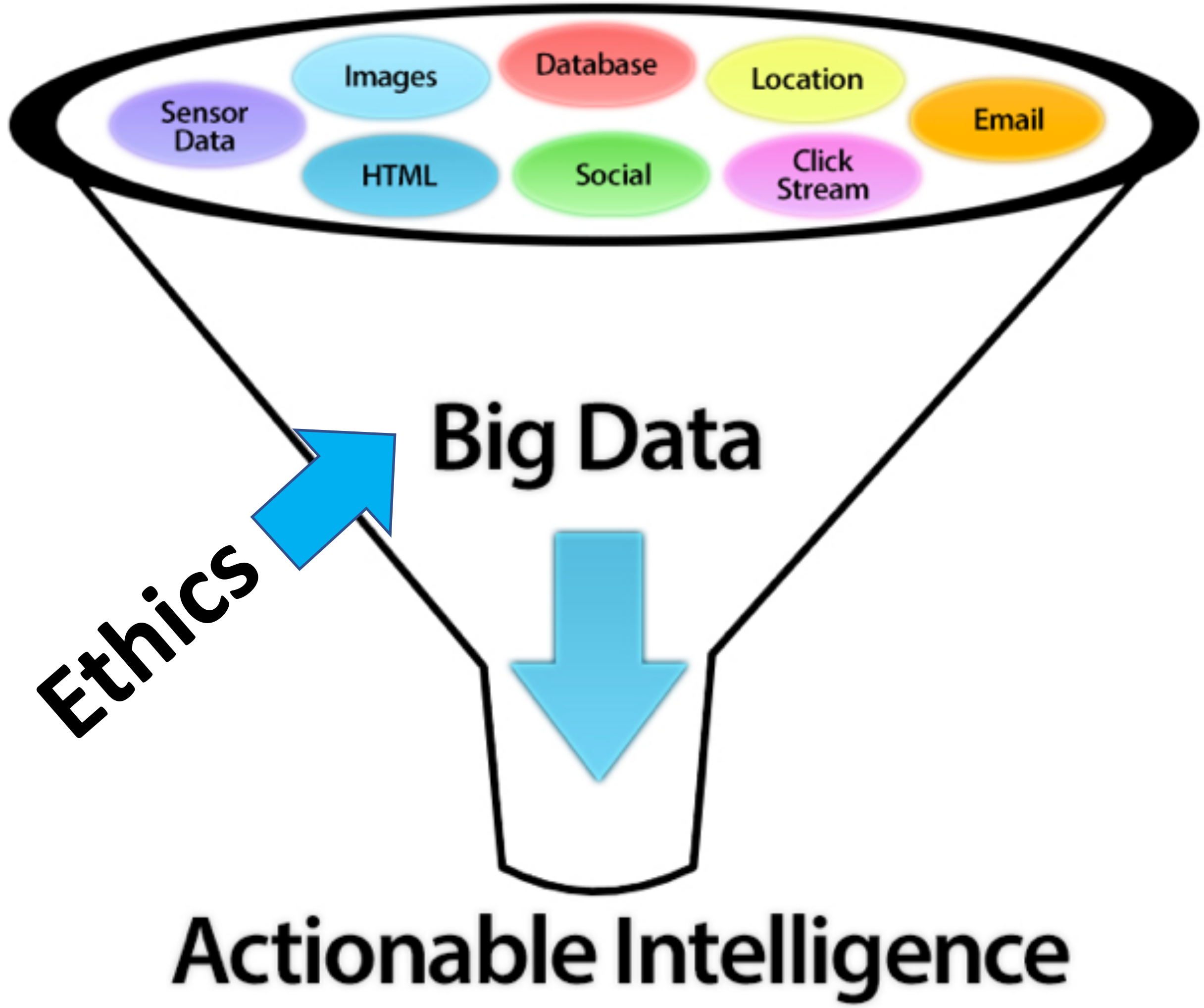
Here we discuss "threshold classifiers," a part of some machine learning systems that is critical to issues of discrimination. A threshold classifier essentially makes a yes/no decision, putting things in one category or another. We look at how these classifiers work, ways they can potentially be unfair, and how you might turn an unfair classifier into a fairer one. As an illustrative example, we focus on loan granting scenarios where a bank may grant or deny a loan based on a single, automatically computed number such as a credit score.

By Martin Wattenberg, Fernanda Viégas, and Moritz Hardt.

This page is a companion to a recent paper by Hardt, Price, Srebro, which discusses ways to define and remove discrimination by improving machine learning systems.

Loan applicants: two scenarios




Sensor Data
Images
Database
Location
Email
HTML
Social
Click Stream
Big Data
Ethics
Actionable Intelligence

Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission

Rich Caruana
Microsoft Research
rcaruana@microsoft.com

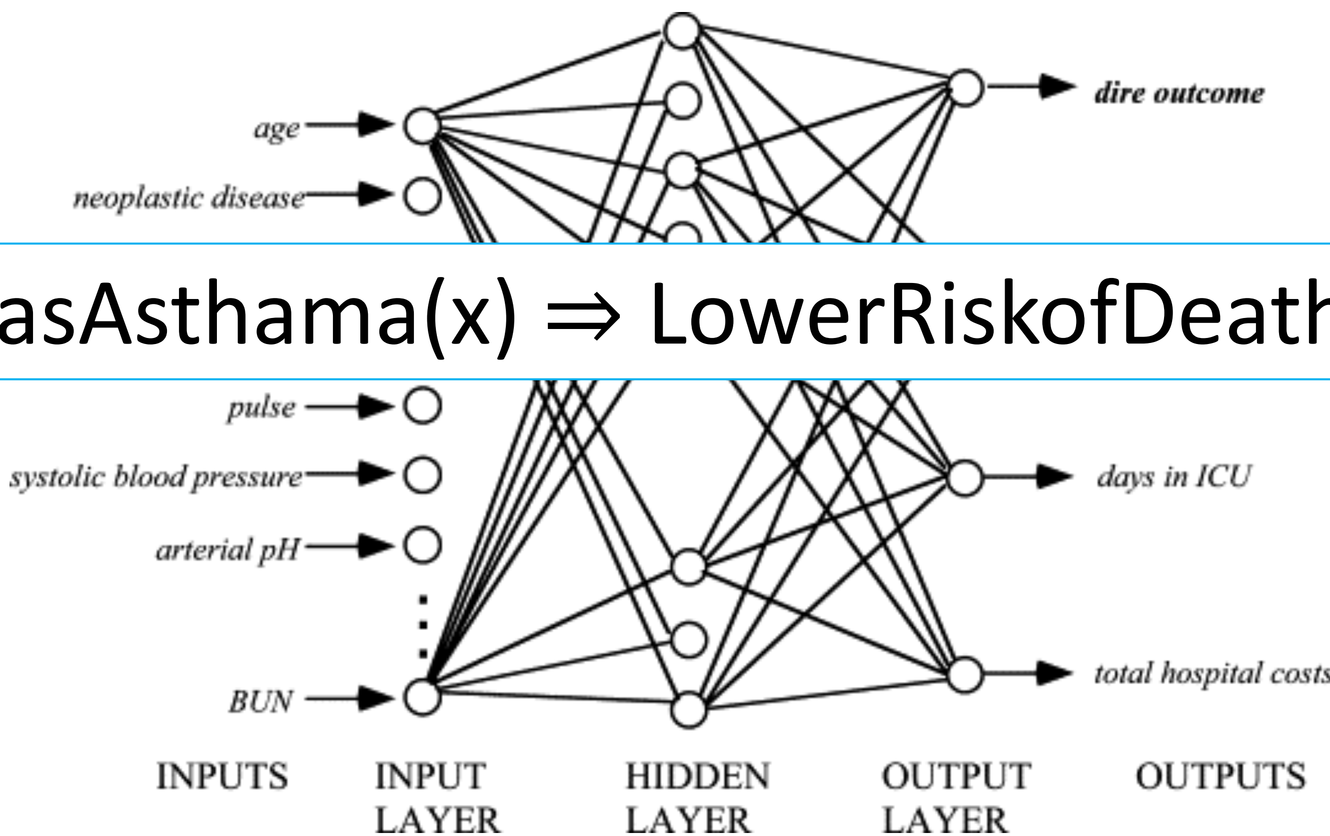
Yin Lou
LinkedIn Corporation
ylou@linkedin.com

Johannes Gehrke
Microsoft
johannes@microsoft.com

Paul Koch
Microsoft Research
paulkoch@microsoft.com

Marc Sturm
NewYork-Presbyterian Hospital
mas9161@nyp.org

Noémie Elhadad
Columbia University
noemie.elhadad@columbia.edu



HasAsthama(x) ⇒ LowerRiskofDeath(x)

Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission

Opinion

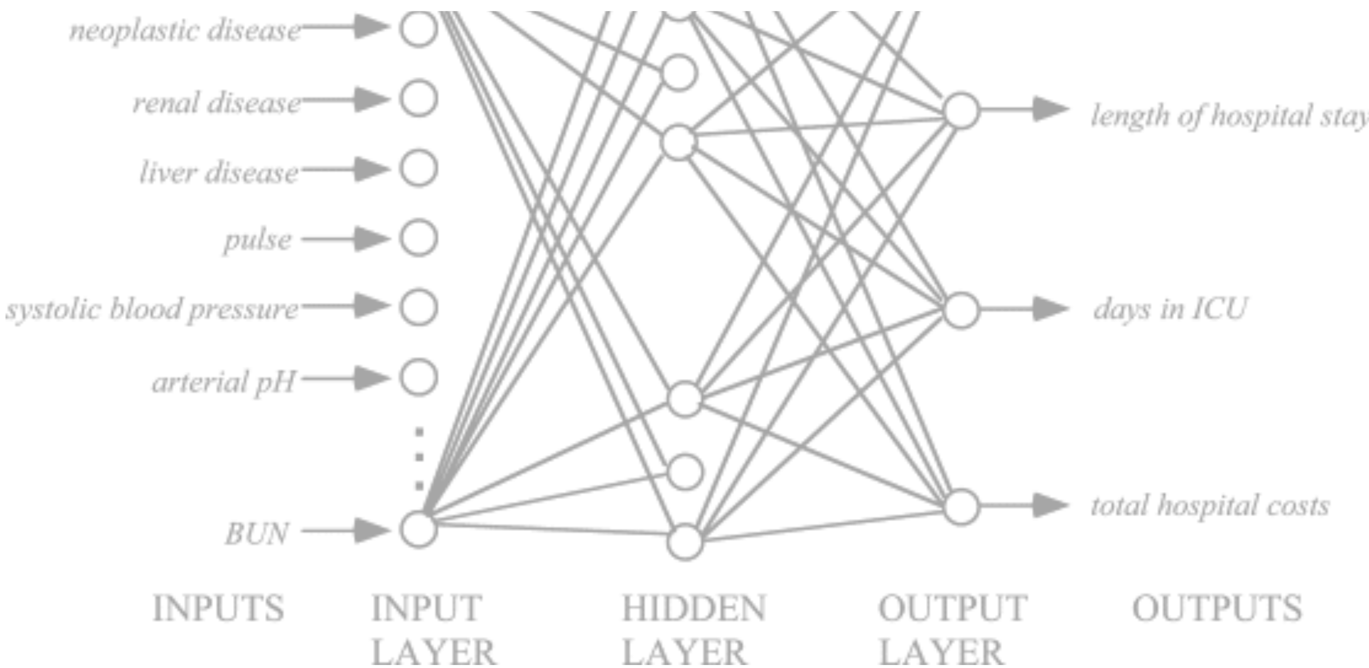
VIEWPOINT

Unintended Consequences of Machine Learning in Medicine

Federico Cabitza, PhD
Department of Informatics, University of Milano-Bicocca,

Over the past decade, machine learning techniques have made substantial advances in many domains. In health care, global interest in the potential of machine learning has increased; for example, a deep learning algorithm has

the expense of other elements that are more difficult or impossible to easily describe. Relying on ML-DSS requires considering digital data as reliable and complete representations of the phenomena that these data are



General Article

ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Psychological Science
22(11) 1359–1366
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings ($\leq .05$), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy

Hack Your Way To Scientific Glory

You're a social scientist with a hunch: **The U.S. economy is affected by whether Republicans or Democrats are in office.** Try to show that a connection exists, using real data going back to 1948. For your results to be publishable in an academic journal, you'll need to prove that they are "statistically significant" by achieving a low enough p-value.

1 CHOOSE A POLITICAL PARTY

Republicans | Democrats

2 DEFINE TERMS

Which politicians do you want to include?

- [x] **Presidents**
- [x] **Governors**
- [x] **Senators**
- [x] **Representatives**

How do you want to measure economic performance?

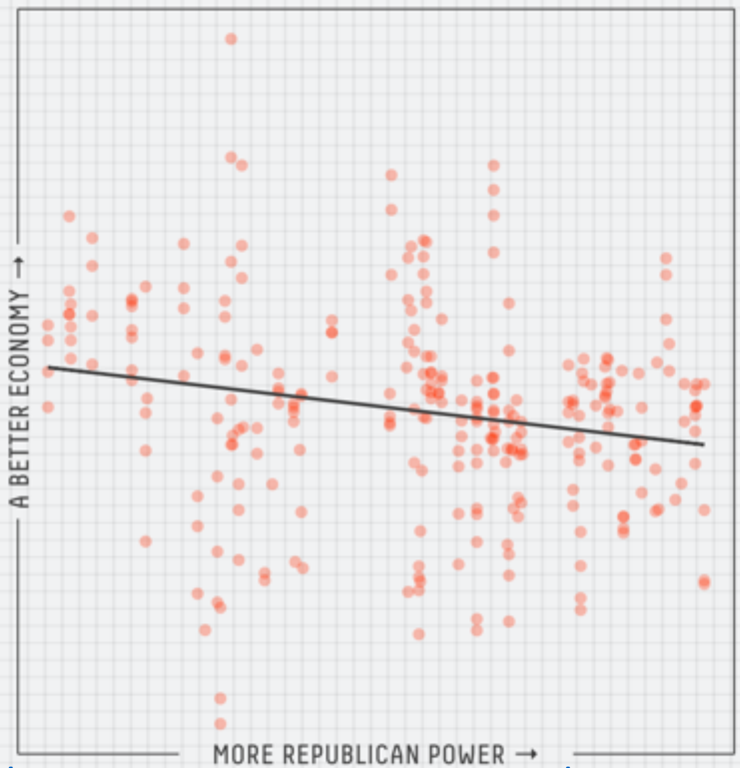
- [ ] **Employment**
- [x] **Inflation**
- [x] **GDP**
- [ ] **Stock prices**

Other options

- [x] **Factor in power**
  Weight more powerful positions more heavily
- [ ] **Exclude recessions**
  Don't include economic recessions

3 IS THERE A RELATIONSHIP?

Given how you've defined your terms, does the economy do better, worse or about the same when more Republicans are in power? Each dot below represents one month of data.



4 IS YOUR RESULT SIGNIFICANT?

If there were no connection between the economy and politics, what is the probability that you'd get results at least as strong as yours? That probability is your p-value, and by convention, you need a **p-value of 0.05 or less** to get published.

1.00 0.50 **0.05**

Result: Publishable

You achieved a p-value of **less than 0.01** and showed that **Republicans** have a **negative** effect on the economy. Get ready to be published!

If you're interested in reading real (and more rigorous) studies on the connection between politics and the economy, see the work of Larry Bartels and Alan Blinder and Mark Watson.

Data from The @unitedstates Project, National Governors Association, Bureau of Labor Statistics, Federal Reserve Bank of St. Louis and Yahoo Finance.

https://projects.fivethirtyeight.com/p-hacking/

Misuse of statistics is unethical.

Douglas Altman, 1980

Biostatistician

British Medical Journal

Ethical Guidelines for Statistical Practice

Prepared by the Committee on Professional Ethics of the American Statistical Association
Approved by the ASA Board in April 2016

[4] Running multiple tests on the same data set at the same stage of an analysis increases the chance of obtaining at least one invalid result. Selecting the one "significant" result from a multiplicity of parallel tests poses a grave risk of an incorrect conclusion. Failure to disclose the full extent of tests and their results in such a case would be highly misleading.

A mistake in the operating room can threaten the life of one patient; **a mistake in statistical analysis or interpretation can lead to hundreds of early deaths**.

Andrew Vickers

Biostatistician

Memorial Sloan Kettering Cancer Center

FDA Estimates Vioxx Caused 27,785 Deaths

The Food and Drug Administration (FDA) estimates that Vioxx may have contributed to 27,785 heart attacks

By Unknown Author

The Food and Drug Administration (FDA) estimates that Vioxx may have contributed to 27,785 heart attacks and sudden cardiac deaths between 1999 and 2003. The estimate is based on the number of prescriptions issued for Vioxx between 1999 and 2003.

Raw data is both an oxymoron and a bad idea… data should be cooked with care.

– Geoffrey Bowker 2005


GOOGLE STALKING
FOR BUSINESS
(IT'S A REAL THING, WE PROMISE)
MAXIMIZEDIGITALMEDIA.COM

Miles McBain
@MilesMcBain

Writing a formalish request for data to another gov department today and there was a para which was effectively a 'No AI' declaration. I am guessing this will become increasingly common.

9:51 PM · 5/13/19 · Twitter Web App

“Children’s is thoughtful about how we use data.”

Data Ethics Checklist Tool

Consider a medium to large data product or project that you're currently working on, or have worked on recently. Fill out the 8-question data ethics checklist tool below to score your opinion about the extent to which ethical dimensions may be involved with this data project. Some examples have been provided that may provide more insight into the meanings of the questions.

After that, we would appreciate if you could complete the optional 8-question survey that asks your thoughts related to your use of this tool.

Finally, press "Submit" when you are ready to consent to having your results included in this research.

Potential Privacy Issues

Examples: People are frequently unaware that data can be aggregated across multiple sources, used to infer seemingly unrelated information, or otherwise go beyond "mere" records. For instance, an ETL process may integrate data that the family may not be aware is connected (demographics and health outcomes). These questions aim to help determine if those inferences are potentially problematic.

P1. Does this process/method/analysis use information that patient/families/employees are probably unaware that Seattle Children's has collected (or could integrate)?

* must provide value

- 0: Definitely not
- 1: In some situations
- 2: Definitely yes

reset

Roughly, did they "consent" to have their data used in this way, though think more broadly than "mere" legal consent.


What would *you* do?