

Data Ethics and Me

What is it and why should I care as a
technologist and technology user?

Lorena Mesa

GitHub Constellation - Chicago 2018

Slides: <https://goo.gl/urR14S>



OBAMA'08

“Were it not for the Internet, Barack Obama would not be president. Were it not for the Internet, Barack Obama would not have been the nominee,” said Arianna Huffington, editor in chief of [The Huffington Post](http://www.thehuffingtonpost.com).”

<https://bits.blogs.nytimes.com/2008/11/07/how-obamas-internet-campaign-changed-politics/>

How?

1. Use of social media (e.g. YouTube)
2. GOTV drives informed by data science
3. Customized “_____ for Obama” interest groups

Slides: <https://goo.gl/urR14S>

Why do I care about this?

- I started my life in politics working on the Obama for America campaigns
- Graduate researching looking at impact of 2009 mortgage crisis on undocumented Latinxs
- Sit on the Python Software Foundation Board of Directors
- Technical Volunteer on many civic minded projects (e.d. Red Cross)

Because we should all care. It impacts us all.



Rose McGowan temporarily blocked from Twitter after Weinstein tweets

The actress settled undisclosed claims against the Hollywood mogul in 1997

by James Vincent | @jjvincent | Oct 12, 2017, 5:34am EDT

"We have been in touch with Ms. McGowan's team," Twitter said in a tweet on Thursday. "We want to explain that her account was temporarily locked because one of her Tweets included a **private phone number**, which violates of our Terms of Service." Source: [CNN](#)

ALAN WIRZBICKI

Twitter should disable Donald Trump's account

Facebook is hiring moderators. But is the job too gruesome to handle?

The social network has pledged to work harder to identify and remove disturbing content - but doing so can take a psychological toll

Meet the Humans Who Protect Your Eyes

Wednesday, June 07, 2017

People upload every kind of image (un)imaginable. Content moderators look at all the awful stuff, so you don't have to. They toil in obscurity. Why aren't they held up as heroes?

[Comments \[3\]](#)



Rose McGowan temporarily blocked from Twitter after Weinstein tweets

The actress settled undisclosed claims against the Hollywood mogul in 1997

by [James Vincent](#) | [@jjvincent](#) | Oct 12, 2017, 5:34am EDT



Twitter Safety ✓

@TwitterSafety

Following

We have been in touch with Ms. McGowan's team. We want to explain that her account was temporarily locked because one of her Tweets included a private phone number, which violates our Terms of Service. 1/3

10:44 AM - 12 Oct 2017

2,075 Retweets 3,092 Likes



3.3K 2.1K 3.1K



Twitter Safety ✓

@TwitterSafety

Following

The Tweet was removed and her account has been unlocked. We will be clearer about these policies and decisions in the future. 2/3

10:44 AM - 12 Oct 2017



Twitter Safety ✓

@TwitterSafety

Following

Twitter is proud to empower and support the voices on our platform, especially those that speak truth to power. We stand with the brave women and men who use Twitter to share their stories, and will work hard every day to improve our processes to protect those voices. 3/3

10:44 AM - 12 Oct 2017



Social works differently across platforms. So what would an algorithm for censorship look like? Should social platforms be censoring speech at all?

Slides: <https://goo.gl/urR14S>

What is data ethics?

- The language of right or wrong?
- Our rights and responsibilities?
- Something else?

What is data ethics?

1. The ethics of data (how data is generated, recorded and shared)
2. The ethics of algorithms (how artificial intelligence, machine learning and robots interpret data)
3. The ethics of practices (devising responsible innovation and professional codes to guide this emerging science)

- What is Data Ethics?, Philosophical Transactions, Luciano Floridi and Mariarosaria Taddeo

Slides: <https://goo.gl/urR14S>

Data: Big Data codifies the past

“**Big Data processes codify the past.** They do not invent the future. Doing that requires moral imagination, and that’s something only humans can provide. We have to **explicitly embed better values into our algorithms**, creating Big Data models that follow our ethical lead. Sometimes that will mean **putting fairness ahead of profit**”

- Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy

Slides: <https://goo.gl/urR14S>

Algorithms: Distribution of the sensible

“[Algorithms have] the power to enable and assign meaningfulness, managing how information is perceived by users, the ‘distribution of the sensible.’”

- Langlois, Ganaele. "Participatory Culture and the New Governance of Communication The Paradox of Participatory Media." *Television & New Media* 14.2 (2013): 91-105.

Slides: <https://goo.gl/urR14S>

How does this impact me as a technology user?

Disclaimer: The examples we'll address do involve some sensitive content.

Slides: <https://goo.gl/urR14S>

Case Study #1: Reporting sexual harassment on Tripadvisor

Slides: <https://goo.gl/urR14S>

TripAdvisor removed warnings about rapes and injuries at Mexico resorts, tourists say

In their place was a message from TripAdvisor that cited various reasons for the deletions: They were “determined to be inappropriate by the TripAdvisor community,” or removed by staff because they were “off-topic” or contained language or subject matter that was not “family friendly.”

The Milwaukee Journal Sentinel asked TripAdvisor to see the posts that were removed. The company refused.

<https://www.jsonline.com/story/news/investigations/2017/11/01/tripadvisor-removed-warnings-rape-and-injuries-mexico-resorts-tourists-say/817172001/> Slides: <https://goo.gl/urR14S>

“Our new email communications will clearly articulate the phrase or sentences that are in violation of our policy, inviting the reviewer to make edits and resubmit their review” TripAdvisor reports.

“These badges will remain on TripAdvisor for up to three months. However, if the issues persist we may extend the duration of the badge,” he said. “These badges are intended to be informative, not punitive.”

Message from TripAdvisor: TripAdvisor has been made aware of recent media reports or events concerning this property which may not be reflected in reviews found on this listing. Accordingly, you may wish to perform additional research for information about this property when making your travel plans.

The new TripAdvisor badge.

<https://www.nytimes.com/2017/11/08/travel/tripadvisor-sex-assault-discrimination-warnings.html>

Slides: <https://goo.gl/urR14S>

Case Study #2: Racial discrimination on Airbnb

Slides: <https://goo.gl/urR14S>

16%

“In an experiment on Airbnb, we find that applications from guests with distinctively African-American names are 16% less likely to be accepted relative to identical guests with distinctly White names.”

Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment,
Edelman et al.

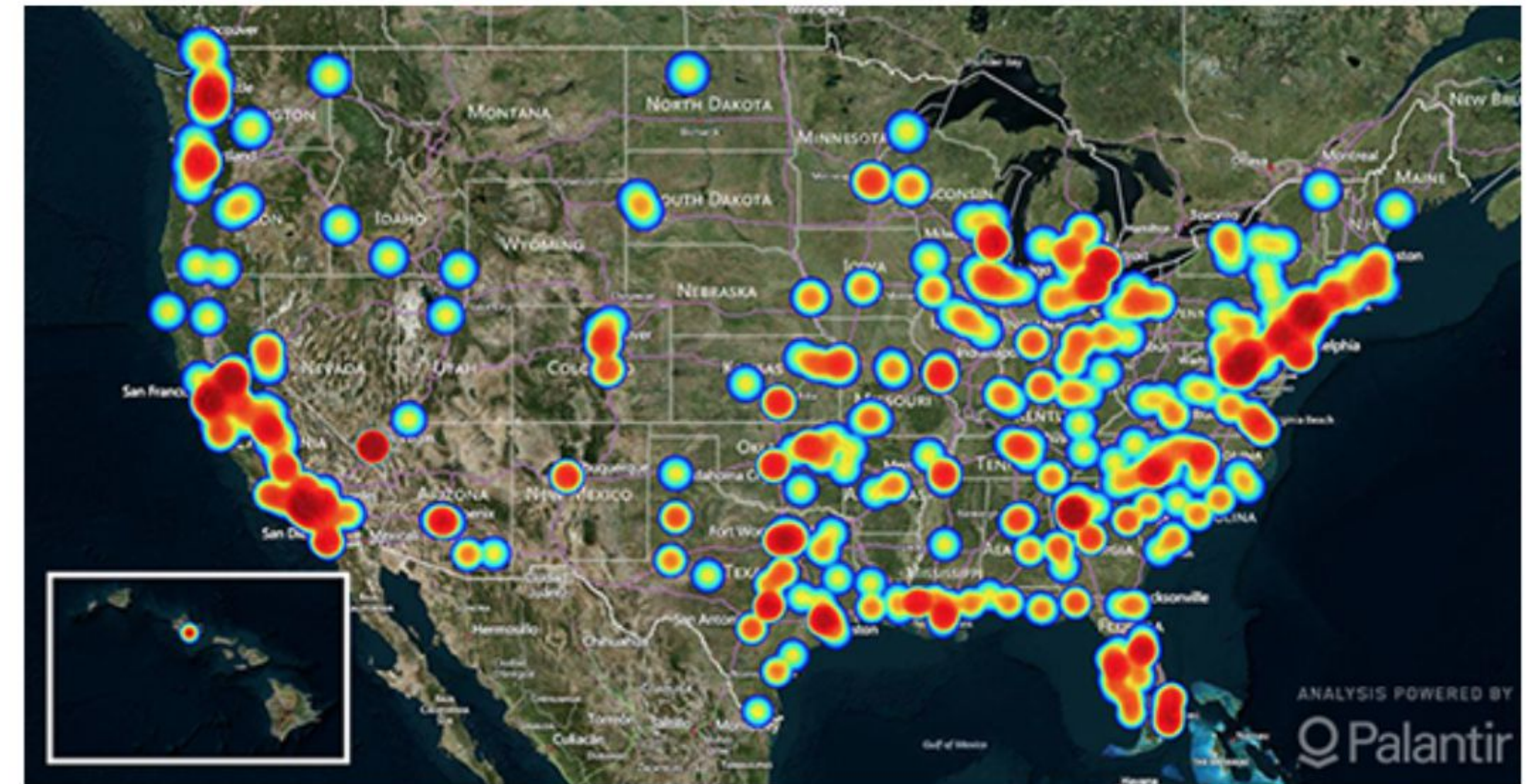
Slides: <https://goo.gl/urR14S>

Case Study #3: Predicting and Fighting Human Trafficking

Slides: <https://goo.gl/urR14S>

“By collecting and analyzing email addresses, street addresses, and phone numbers gathered off the Web, Schles’ [of Demand Abolition] tool finds common threads connecting people listed in the classified advertisements of commercial sex websites. He even used facial recognition tools to identify and track the photos of people featured in the ads, many of whom (but not all) are victims of trafficking.”

- <https://www.datanami.com/2016/10/07/data-analytics-fight-human-trafficking/>



Polaris used Palantir to visualize about 6,500 illicit massage businesses that it identified across the country

**Algorithms as well as
data are shaped and
designed by humans.**

Data scientists to develop their own version of the 'Hippocratic Oath'



Schaun Wheeler

Follow

Senior Data Scientist at Valassis Digital. Formerly at Success Academy Charter Schools, U.S. Department of the Army, and a few other places. Anthropologist.

Feb 6 · 6 min read

An ethical code can't be about ethics

This is why I can't support Data for Democracy's data science code of ethics.

Slides: <https://goo.gl/urR14S>

Ethics Principles

- Its my job to understand, mitigate and communicate the presence of bias in algorithms.
- Be responsible for maximizing social benefit and minimizing harm.
- Practice humility and openness.
- I will know my data and help future users know it as well.
- Make reasonable efforts to know and document its origins and document its transformation.
- Bias will exist. Measure it. Plan for it.
- Thou shalt document transparently, accessibly, responsibly, reproducibly, and communicate.
- Engaging the whole community. Do you have all relevant individuals engaged?
- People before data - data scientists should use a question driven approach rather than a data-driving or methods approach. Consider personal safety and treat others the way they want to be treated.
- Exercise ethical imagination.
- Open by default - use of data should be transparent and fair.
- I will not over/under represent findings.
- You are part of an ecosystem understand context and provenance.
- Respecting human dignity.
- Respect their data even more than your own. Understand where its sources and think about the consequences of your actions.
- Protecting individual and institutional privacy.
- Diversity for inclusivity.
- Attention to bias.
- Respect for others/persons.
- Be intentional as you work to create value.

Data 4 Democracy “Hippocratic Data Oath”
<https://github.com/Data4Democracy/ethics-resources>

Practices: Skin in the game?

[Technologists] don't need a list of ways to be virtuous. They need a list of ways to prove they aren't charlatans. That will do more to ensure the health and trustworthiness of the profession than anything else.

- Shaun Wheeler, Senior Data Scientist @ Valassis Digital

Slides: <https://goo.gl/urR14S>

Practices: How to be an ethical company?

A data-ethical company sustains ethical values relating to data, asking: **Is this something I myself would accept as a consumer? Is this something I want my children to grow up with?** A company's degree of "data ethics awareness" is not only crucial for survival in a market where consumers progressively set the bar, it's also necessary for society as a whole.

- Gary Hasselbach & Pernille Tranberg, Tech Crunch, Data Ethics — The New Competitive Advantage (2016)

Slides: <https://goo.gl/urR14S>

Data Lifecycle: What is your organization's policy on data management?

Example: Mozilla's Data Collection Policy

1. How does the company use Big Data, and to what extent is it integrated into strategic planning?
2. Does the organisation send a privacy notice when personal data are collected?
3. Does my organisation assess the risks linked to the specific type of data my organisation uses?
4. Does my organisation have safeguards in place to mitigate these risks?
5. Do we make sure that the tools to manage these risks are effective and measure outcomes?
6. Do we conduct appropriate due diligence when sharing or acquiring data from third parties?

Source: 6 Ethical Questions about Big Data
<https://www.fm-magazine.com/news/2016/jun/ethical-questions-about-big-data.html>



Slides: <https://goo.gl/urR14S>

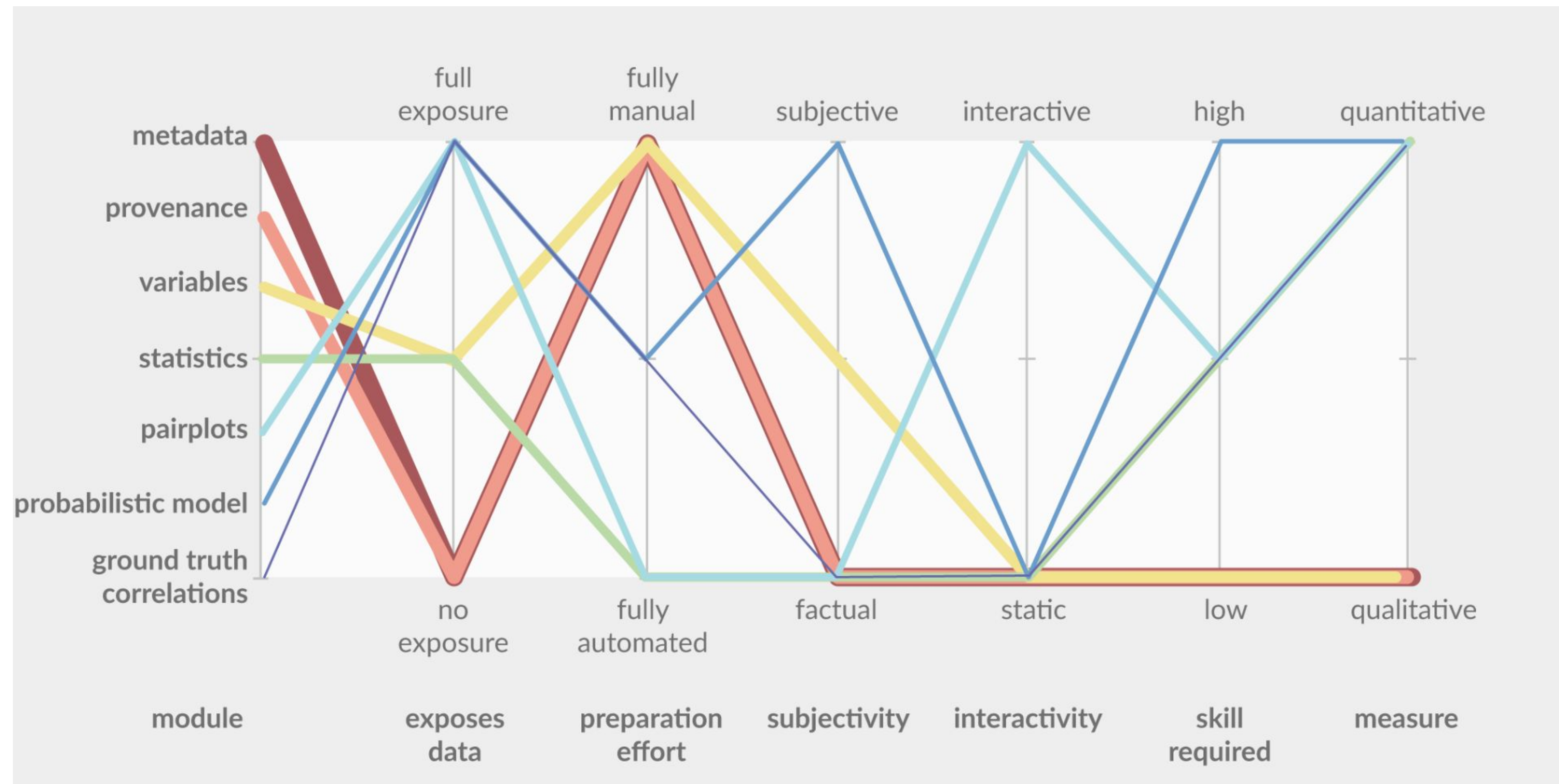
Practices: How can we improve machine learning?

1. Explore and understand your data
2. Explore your errors
3. Make your results interpretable (e.g. Build things that people can understand!)
4. If you do not know something, find someone that does.

- Deborah Hanus, PyCon Colombia 2018, Harvard PhD Machine Learning + Computer Science

The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards

Sarah Holland^{1*}, Ahmed Hosny^{2*}, Sarah Newman³, Joshua Joseph⁴, and Kasia Chmielinski¹
April 2018 | nutrition@media.mit.edu | datanutrition.media.mit.edu



Read white paper at: datanutrition.media.mit.edu

Slides: <https://goo.gl/urR14S>

Dataset Facts

ProPublica's Dollars
for Docs Data

Metadata

Filename	201612v1-docdollars- product_payments
----------	--

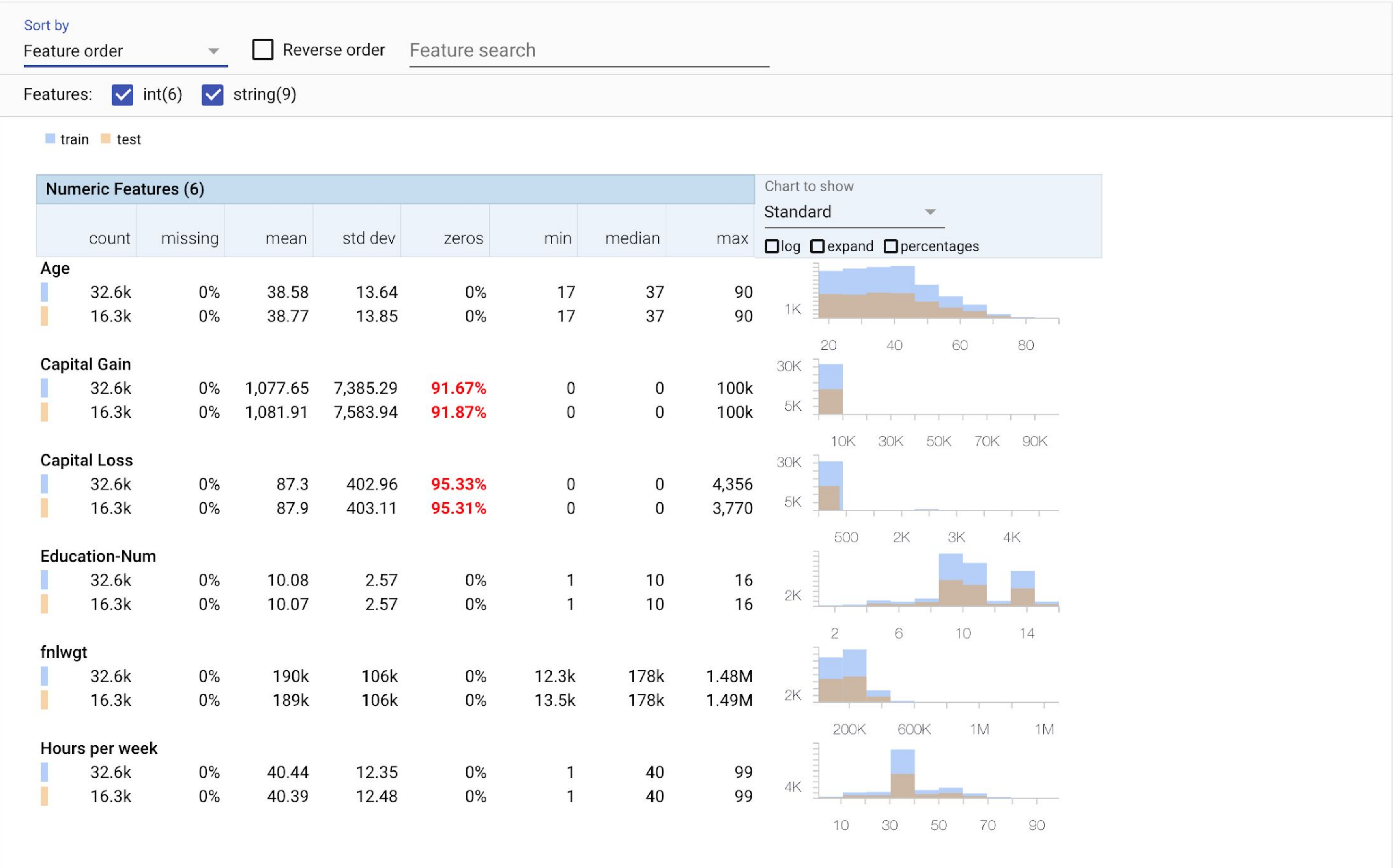
Format	csv
--------	-----

Url	https://projects.propublica.org/docdollars/
-----	---

Domain	healthcare
--------	------------

Keywords	Physicians, drugs, medicine, pharmaceutical, transactions
----------	--

Prototype on ProPublica's Dollars for Docs (2013-2015) dataset - data on payments made from pharmaceutical companies to doctors.



Contains two visualizations for understanding and analyzing machine learning datasets:
Facets Overview and Facets Dive. <https://pair-code.github.io/facets/>



Guiding Principles

Code of Conduct

Party By...

Location

Upcoming Dates

Past Dates

Learn

Tools

How-Tos

Links: Videos, Articles,
Books, etc

Resources

Connect

Contact

Resources

Media coverage

Organize

How to organize a
CryptoParty

About this Wiki

Help on editing

Chicago, Illinois

Edit

Why CryptoParty?

Privacy is the space in which ideas are developed, to retreat into whenever you want. This space is not only physical but digital as well. Governments and companies don't want to respect that so we become active ourselves.

Edit

What is a CryptoParty?

CryptoParties are free and open for everyone, but especially those without prior knowledge, who haven't yet attended one.

CryptoParty is a decentralized movement with events happening all over the world. The goal is to pass on knowledge about protecting yourself in the digital space. This can include encrypted communication, preventing being tracked while browsing the web and general security advice for computers and smartphones.

To try the programs and apps at the CryptoParty bring your laptop or smartphone.

Edit

Next Dates

Day	Time	Location	Address	Additional Infos
Nov 20th, 2017	5:30-8 pm	TechNexus	20 N Wacker Dr, 12th floor	RSVP: https://www.eventbrite.com/e/november-cryptoparty-chicago-registration-39873296102

Table of Contents

Chicago, Illinois

Why CryptoParty?

What is a CryptoParty?

Next Dates

Contact



Know your attack vector! Attend or host a CryptoParty!
<https://www.cryptoparty.in>

IDNYC Data Not Shredded Yet as New Lawsuit Keeps Personal Info in Limbo



By [Nicholas Rizzi](#) | May 2, 2017 4:34pm
[@nickr15](#)



NEW YORK CITY — Personal information used to apply for the [IDNYC](#) program — [which a judge ruled nearly a month ago the city could delete](#) in 10 days — has instead been preserved due to another lawsuit.

Republican Assemblymembers Ron Castorina Jr. and Nicole Malliotakis filed a second lawsuit April 13



▲ View Full Caption

DNAinfo/Trevor Kapp

Recommended

[PARK SLOPE »](#)
4 Injured When Car Jumps Curb in Park Slope, FDNY Says

[PROSPECT HEIGHTS »](#)
'Affordable' Apartments End Up on StreetEasy After Housing Lottery Flops

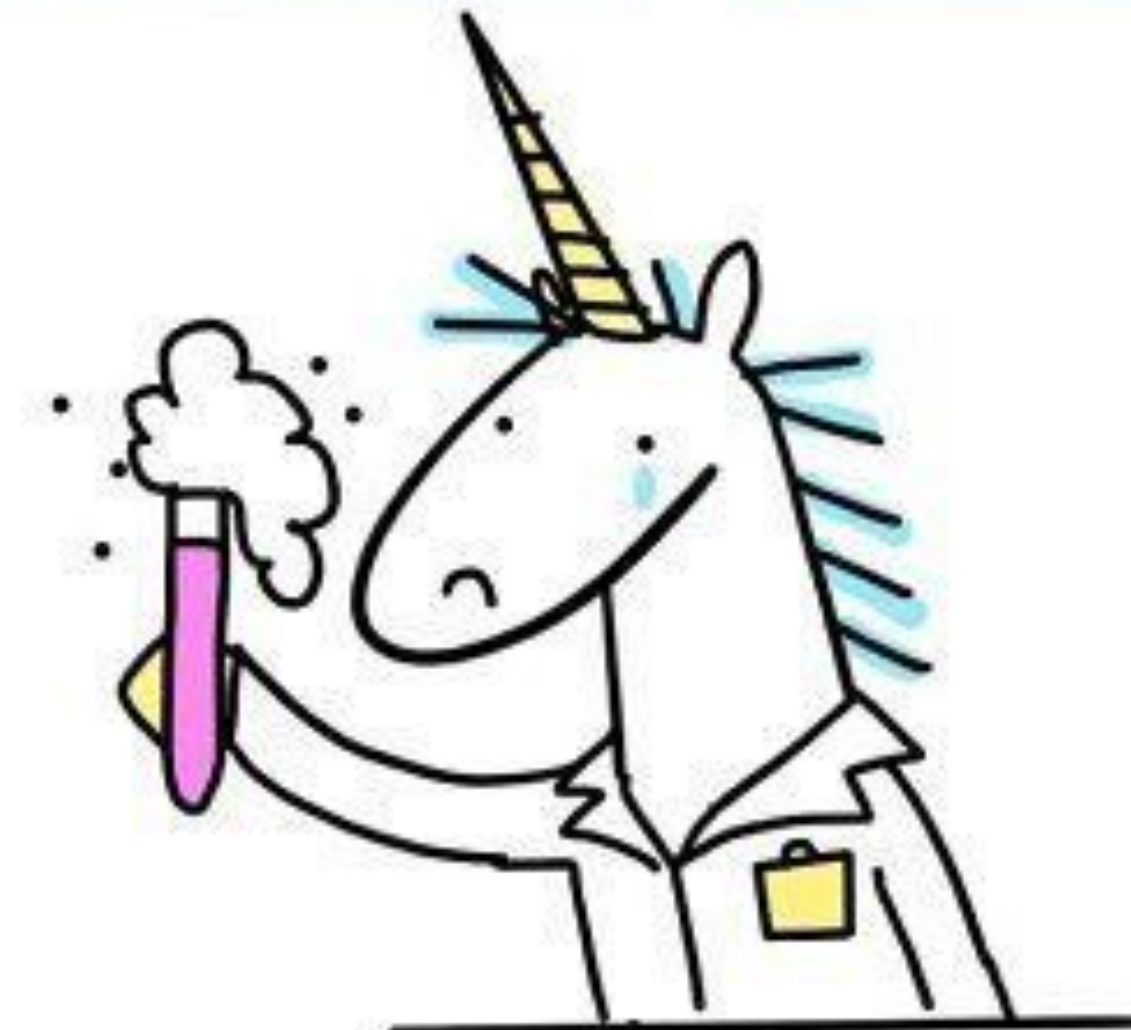
[SOHO »](#)
6 Private Clubs That Will

The decisions we make with data has a very real and very dangerous impact on the world we live in.

Thanks!

Let's continue chatting.
You can find me on social
most anywhere as
[@loooorenanicole](#).

SCIENCE UNICORN



JUST DISCOVERED
THAT
HE DOESN'T EXIST

@twisteddoodles

Slides: <https://goo.gl/urR14S>