

# Data Past Present Future

Chris Wiggins + Matthew Jones  
SEAS + A&S

Lecture 1: Jan 22 2019

Suggested Page

...



Data: Past, Present, and Future

Sponsored

how did this end up in my news feed? find out in new course:  
<https://data-ppf.github.io/>



Data: Past, Present, and Future

Community

3 people like this



Like Page

Pages X Power Editor - Manage Ads X

Secure | https://business.facebook.com/ads/manage/powereditor/manage/adsets

Search business Q Data: Past, Present, and Future 3 Help ?

Discard Changes Review Draft Items gear

Search Filters Add filters to narrow the data you are seeing. This month: Jan 1, 2018 – Jan 16, 2018

Account Overview Campaigns 1 selected Ad Sets for 1 Campaign Ads for 1 Campaign

+ Create Ad Set Duplicate Edit Columns: Performance Breakdown Export

	Ad Set Name	livery	Results	Reach	Impressions	Cost per Result	Budget	Amount Spent
<input type="checkbox"/>	Ad Set - 1st PPF ad	Recently Completed	7 Link Clicks	258	399	\$0.57 Per Link Click	\$4.00 Lifetime	\$3.40
<input type="checkbox"/>	Ad Set - Ad B	Recently Completed	8 Link Clicks	259	381	\$0.50 Per Link Click	\$4.00 Lifetime	\$4.00
▶ Results from 2 ad sets ⓘ			15 Link Clicks	517 People	780 Total	\$0.53 Per Link Click		\$7.93 Total Spent

f

Pages X Power Editor - Manage Ads X

Secure | https://business.facebook.com/ads/manage/powereditor/manage/adsets/edit/

Power Editor Search business Data: Past, Present, and Future 1 Help ?

**Creating Ad Set: test T-1 ad set**

Detailed Targeting i INCLUDE people who match at least ONE of the following i

Demographics > Education > Schools  
Columbia University

Interests > Additional Interests  
Flat Earth

Add demographics, interests or behaviors | Suggestions | Browse

**Audience Definition**

Your audience selection is fairly broad.

Potential Reach: 810,000 people i

**Estimated Daily Results**

Reach  
2,300 - 6,500

Link Clicks  
28 - 110

⚠ Your results are likely to differ from estimates  
We have limited data available to calculate this estimate, so estimates may be less accurate.

No Results Found

Connections i

Placement

Saving to draft

2 items to review, including 1 new campaign Close



## Ads Manager

Search business



Data: Past, Present, and Fu...



Search



Traffic-20190121T14h54

&gt; 2 Ad Sets &gt; 2 Ads

Active  
In Split Test

This month: Jan 1, 2019 – Jan 22, 2019

- Traffic-20190121T14h54 ...
- ▼  Ad Set for Ad B ...
- Ad B ...
- ▼  Ad Set for Ad A ...
- Ad A ...

Performance

Demographics

Placement

6 Results: Link Clicks ▾ 102 Reach ▾

Age

All Women

83% ( 5 )

45% ( 46 )

\$0.17

Cost per Result



All Men

17% ( 1 )

55% ( 56 )

\$0.93

Cost per Result

Split Test: Traffic-20190121T14h54 | ● Active

TEST SCHEDULE: Jan 21, 2019 – Jan 22, 2019

View By Cost per Result ▾

No Winner Yet

## Ads Manager

Search business



Data: Past, Present, and Fu...



Search



Traffic-20190121T14h54 &gt; 2 Ad Sets &gt; 2 Ads

65+

Active  
In Split Test

Traffic-20190121T14h54



Ad Set for Ad B



Ad B



Ad Set for Ad A



Ad A

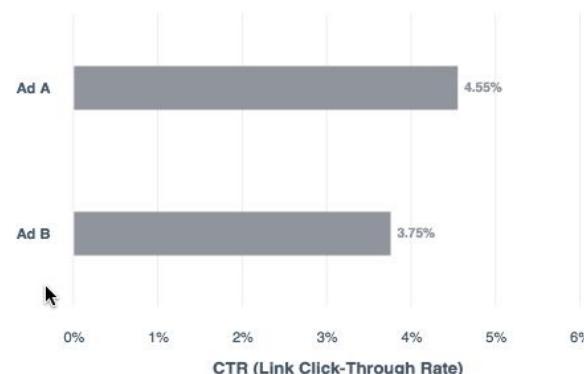
Split Test: Traffic-20190121T14h54 | Active

TEST SCHEDULE: Jan 21, 2019 – Jan 22, 2019

View By CTR (Link Click-Through Rate) ▾

Ad A had the highest link click-through rate with 4.55%.

View by Cost per Result to find out if there was a winner from the test.



0% 1% 2% 3% 4% 5% 6%

CTR (Link Click-Through Rate)

Variable: Creative | Versions: 2 Ads | Total Budget: \$10.00, Even Split (50/50) | Objective: Traffic

Give Feedback

**Ads Manager**

Search

Search business



Data: Past, Present, and Fu...

Traffic-20190121T14h54 > Ad Set for Ad B > **Ad B**

Active

...

Edit Review

(Recommended)

Show multiple images or videos for the same price. Learn more.

Collection

Feature a collection of items that open into a fullscreen mobile experience. Learn more.

**Instant Experience**

Include a mobile landing page that opens instantly when someone interacts with your ad and track activity with a Facebook pixel. Start with a template or create a custom layout. Learn more.

 Add an Instant Experience **New! Turn Images Into Videos**

Now you can create a Single Video ad when you don't have a video. Choose a template in the Video Creation Kit to get started.

Use Templates

**Now You Can Show Bigger Images****SPONSORED**

Data: Past, Present, and ...



data: past, present, and future



\*NO PREREQUISITES OR PROGRAMMING EXP...

LEARN MORE &gt;

LIKE | COMMENT | SHARE

Refresh preview • Report a problem with this preview

Close

Discard Draft

Publish

**Ads Manager**

Search

Search business



Data: Past, Present, and Fu...



Traffic-20190121T14h54

...

Ad Set for Ad B

...

Ad B

...

Ad Set for Ad A

...

Ad A

...

Traffic-20190121T14h54 &gt; Ad Set for Ad A &gt; Ad A

Active

...

Edit Review

(Recommended)

Show multiple images or videos for the same price. Learn more.

Collection

Feature a collection of items that open into a fullscreen mobile experience. Learn more.

**Instant Experience**

Include a mobile landing page that opens instantly when someone interacts with your ad and track activity with a Facebook pixel. Start with a template or create a custom layout. Learn more.

 Add an Instant Experience **New! Turn Images Into Videos**

Now you can create a Single Video ad when you don't have a video. Choose a template in the Video Creation Kit to get started.

[Use Templates](#) **Now You Can Show Bigger Images****SPONSORED**

Data: Past, Present, and ...



new class on history and ethics of data, with Python



new class on history and ethics of data, with Py...

[LEARN MORE >](#)

LIKE | COMMENT | SHARE

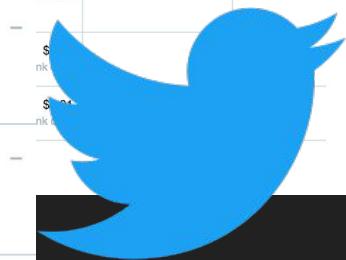
[Refresh preview](#) • [Report a problem with this preview](#)[Close](#)[Discard Draft](#)[Publish](#)

# 2018



Objective: All  Status: All

Spend  
**\$25.00**



Location <span>?</span>	Impressions	Spend	Results
Total for account	12,940	\$25.00	-
Brazil	38	\$0.06	-
Indonesia	5,638	\$3.38	-
Bosnia and Herzegovina	3	\$0.01	-
Ukraine	2	\$0.01	-
Trinidad and Tobago	1	\$0.00	-
Greece	2	\$0.00	-
Kuwait	18	\$0.00	-

Filters: Default

It's faster to monitor and optimize campaigns with our [Customization](#), [Automation](#), and [Export](#) capabilities.

Looking for? [Get help](#)

Campaigns  Ad Groups  Share

Name

Summary for 1 item

On Twitter

On Twitter Audience Platform

data:PPF spring 2018 Website clicks or conversions

On Twitter

On Twitter Audience Platform

**Spend \$25.00**

Objective: All Status: All

Location	Impressions	Spend	Results
Total for account	12,940	\$25.00	
Brazil	38	\$0.06	
Kuwait	18	\$0.00	

**3 Amplify your message by promoting your Tweets**

When you promote your Tweets, Twitter will prominently display your most engaging Tweets to your followers and those with interests similar to your followers.

Spend no more than:  per day

You will be charged between: \$0.01 and \$0.75 for each click.

We recommend a maximum bid of at least \$0.75.

How do you want to select Tweets?

- Automatically refresh to include your newest, most engaging Tweets. [More info](#)
- Manually select your Tweets

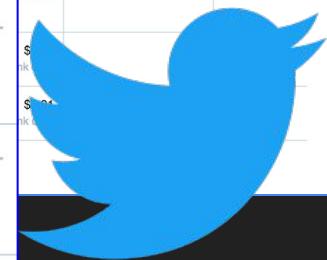
Promote a new Tweet

**Start promoting your Tweets** **Skip for now**

**Preview your 5 Promoted Tweets**

Click below to stop any Tweet from being promoted.

result	Daily budget	Re campaign
\$0.03 k click	\$5.00	
\$1.07 k click	-	
\$0.01 k click	-	
\$0.03 k click	\$5.00	
\$0.01 k click	-	





All accounts > data-pfp on github

## All Web Site Data ▾



Search reports and help



HOME



CUSTOMIZATION



REAL-TIME



AUDIENCE



ACQUISITION



BEHAVIOR



CONVERSIONS



DISCOVER



ADMIN

## Google Analytics Home



INTELLIGENCE

Users

**476**

↑833.3%

vs last 7 days

Sessions

**505**

↑770.7%

Bounce Rate

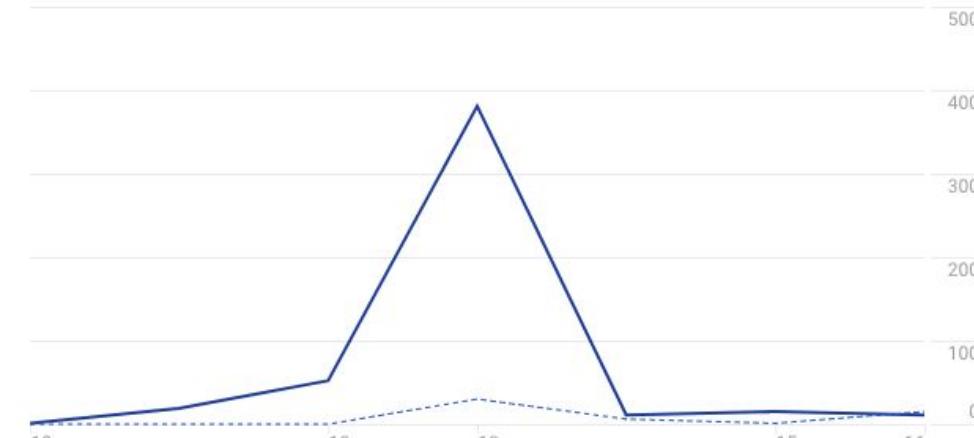
**87.92%**

↑6.2%

Session Duration

**0m 13s**

↓91.5%



Last 7 days ▾

AUT



Google  
Analytics

Country	Acquisition			Behavior			Conversions		
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration	Goal Conversion Rate	Goal Completions	Goal Value
	563 % of Total: 100.00% (563)	93.43% Avg for View: 93.43% (0.00%)	526 % of Total: 100.00% (526)	87.39% Avg for View: 87.39% (0.00%)	1.24 Avg for View: 1.24 (0.00%)	00:00:29 Avg for View: 00:00:29 (0.00%)	0.00% Avg for View: 0.00% (0.00%)	0 % of Total: 0.00% (0)	\$0.00 % of Total: 0.00% (\$0.00)
1. Indonesia	271 (48.13%)	95.20%	258 (49.05%)	92.99%	1.17	00:00:06	0.00%	0 (0.00%)	\$0.00 (0.00%)
2. United States	149 (26.47%)	85.23%	127 (24.14%)	80.54%	1.40	00:01:33	0.00%	0 (0.00%)	\$0.00 (0.00%)
3. Malaysia	81 (14.39%)	98.77%	80 (15.21%)	87.65%	1.17	<00:00:01	0.00%	0 (0.00%)	\$0.00 (0.00%)
4. Thailand	9 (1.60%)	100.00%	9 (1.71%)	100.00%	1.00	00:00:00	0.00%	0 (0.00%)	\$0.00 (0.00%)
5. Australia	7 (1.24%)	85.71%	6 (1.14%)	42.86%	2.00	00:00:04	0.00%	0 (0.00%)	\$0.00 (0.00%)
6. United Kingdom	7 (1.24%)	100.00%	7 (1.33%)	100.00%	1.00	00:00:00	0.00%	0 (0.00%)	\$0.00 (0.00%)
7. India	6 (1.07%)	100.00%	6 (1.14%)	50.00%	1.50	00:00:53	0.00%	0 (0.00%)	\$0.00 (0.00%)
8. Switzerland	4 (0.71%)	100.00%	4 (0.76%)	100.00%	1.00	00:00:00	0.00%	0 (0.00%)	\$0.00 (0.00%)
9. Peru	4 (0.71%)	100.00%	4 (0.76%)	75.00%	1.25	00:00:17			
10. Turkey	4 (0.71%)	100.00%	4 (0.76%)	50.00%	1.75	00:00:06			



Source / Medium	Acquisition			Behavior			Conversions		
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration	Goal Conversion Rate	Goal Completions	Goal Value
	563 % of Total: 100.00% (563)	93.43% Avg for View: 93.43% (0.00%)	526 % of Total: 100.00% (526)	87.39% Avg for View: 87.39% (0.00%)	1.24 Avg for View: 1.24 (0.00%)	00:00:29 Avg for View: 00:00:29 (0.00%)	0.00% Avg for View: 0.00% (0.00%)	0 % of Total: 0.00% (0)	\$0.00 % of Total: 0.00% (\$0.00)
1. (direct) / (none)	<b>393</b> (69.80%)	95.93%	377 (71.67%)	89.31%	1.21	00:00:17	0.00%	0 (0.00%)	\$0.00 (0.00%)
2. t.co / referral	<b>85</b> (15.10%)	91.76%	78 (14.83%)	77.65%	1.36	00:00:33	0.00%	0 (0.00%)	\$0.00 (0.00%)
3. ads-bidder-api.twitter.com / referral	<b>34</b> (6.04%)	82.35%	28 (5.32%)	88.24%	1.18	00:00:28	0.00%	0 (0.00%)	\$0.00 (0.00%)
4. m.facebook.com / referral	<b>20</b> (3.55%)	100.00%	20 (3.80%)	95.00%	1.05	00:00:08	0.00%	0 (0.00%)	\$0.00 (0.00%)
5. l.facebook.com / referral	<b>15</b> (2.66%)	93.33%	14 (2.66%)	80.00%	1.80	00:04:34	0.00%	0 (0.00%)	\$0.00 (0.00%)
6. datascience.columbia.edu / referral	<b>5</b> (0.89%)	20.00%	1 (0.19%)	100.00%	1.00	00:00:00	0.00%	0 (0.00%)	\$0.00 (0.00%)
7. google / organic	<b>4</b> (0.71%)	25.00%	1 (0.19%)	50.00%	1.75	00:06:15	0.00%	0 (0.00%)	\$0.00 (0.00%)
8. facebook.com / referral	<b>3</b> (0.53%)	100.00%	3 (0.57%)	100.00%	1.00	00:00:00	0.00%	0 (0.00%)	\$0.00 (0.00%)
9. ad-review-tool.twitter.biz / referral	<b>1</b> (0.18%)	100.00%	1 (0.19%)	100.00%	1.00	00:00:00			
10. adwords.google.com / referral	<b>1</b> (0.18%)	100.00%	1 (0.19%)	100.00%	1.00	00:00:00			



Country / Medium	Acquisition			Behavior			Conversions			
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration	Goal Conversion Rate	Goal Completions	Goal Value	Goal Value
	5633 % of Total: 100.00% (563)	93.43% Avg. New View: 0.00%	526 % of Bouncers: (0.00%)	87.39% Avg. Bounce: 0.00%	1.24 Avg. Pgs/View: 0.00%	00:00:29 Avg. Sess/Dur: 0.00%	0.00% Avg. Conv/ Avg. View: 0.00% (0.00%)	0 % of Total: (0.00%)	\$0.00 % of Total: 0.00% (0.00\$0.00)	\$0.00 % of Total: 0.00% (0.00\$0.00)
1. (direct) Indonesia	271 (48.19%)	how did this end up in my news feed? why?								
2. t.co / referral United States	149 (26.47%)	0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00%								
3. ads-bidder-api.twitter.com / referral Malaysia	81 (14.39%)	0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00%								
4. m.facebook.com / referral Thailand	8 (1.60%)	0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00%								
5. l.facebook.com / referral Australia	7 (1.24%)	0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00%								
6. datascience.columbia.edu / referral United Kingdom	7 (1.24%)	0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00%								
7. google / organic Indie	6 (1.07%)	0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00% 0.00%								
8. facebook.com / referral Switzerland	4 (0.71%)	100.00% (0.76%)	100.00% (0.76%)	100.00% (0.76%)	1.00	00:00:00	0.00% (0.00%)	0 (0.00%)	\$0.00 (0.00\$0.00)	\$0.00 (0.00\$0.00)
9. ad-review-tool.twitter.biz / referral Peru	4 (0.71%)	100.00% (0.76%)	4 (0.76%)	75.00% (0.76%)	1.25	00:00:17	0.00% (0.00%)	0 (0.00%)	\$0.00 (0.00\$0.00)	\$0.00 (0.00\$0.00)
10. adwords.google.com / referral Turkey	4 (0.71%)	100.00% (0.76%)	4 (0.76%)	50.00% (0.76%)	1.95	00:00:06	0.00% (0.00%)	0 (0.00%)	\$0.00 (0.00\$0.00)	\$0.00 (0.00\$0.00)

Rows 1 - 10 of 12

Country / Medium	Acquisition			Behavior			Conversions			
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration	Goal Conversion Rate	Goal Completions	Goal Value	Goal % of Total
	5633 % of Total: 100.00% (563)	93.43% Avg. New Session: 0.00%	526 % of Total: 100.00% (526)	87.39% Avg. Bounce Rate: 0.00%	1.24 Avg. Pages per View: 0.00	00:00:29 Avg. Session Duration: 0.00	0.00% Avg. Goal Conversion Rate: 0.00%	0 % of Total: 0.00% (0)	\$0.00 Avg. Goal Value: \$0.00	\$0.00 % of Total: 0.00% (0)
1. (direct) Indonesia	271 (48.19%)	<p>how did this end up in my news feed?</p> <ul style="list-style-type: none"> <li>- math</li> <li>- hardware</li> <li>- system</li> <li>- funding</li> <li>- market</li> <li>- regulation</li> <li>- data</li> </ul>								
2. t.co / referral United States	149 (26.47%)	<p>this was not possible 20 years ago.</p> <ul style="list-style-type: none"> <li>- why?</li> <li>- what did people do instead?</li> </ul>								
3. ads-bidder-api.twitter.com / referral Malaysia	81 (14.39%)									
4. m.facebook.com / referral Thailand	8 (1.46%)									
5. l.facebook.com / referral Australia	7 (1.24%)									
6. datascience.columbia.edu / referral United Kingdom	7 (1.24%)									
7. google / organic Indie	6 (1.07%)									
8. facebook.com / referral Switzerland	4 (0.71%)	100.00% (0.71%)	(0.76%)	100.00%	1.00	00:00:00	0.00% Goal Conversion Rate: 0.00%	0 Goal Completions: 0.00% (0)	\$0.00 Goal Value: \$0.00	\$0.00 % of Total: 0.00% (0)
9. ad-review-tool.twitter.biz / referral Peru	4 (0.71%)	100.00% (0.71%)	(0.76%)	75.00%	1.25	00:00:17	0.00% Goal Conversion Rate: 0.00%	0 Goal Completions: 0.00% (0)	\$0.00 Goal Value: \$0.00	\$0.00 % of Total: 0.00% (0)
10. adwords.google.com / referral Turkey	4 (0.71%)	100.00% (0.71%)	(0.76%)	50.00%	1.95	00:00:06	0.00% Goal Conversion Rate: 0.00%	0 Goal Completions: 0.00% (0)	\$0.00 Goal Value: \$0.00	\$0.00 % of Total: 0.00% (0)

Rows 1 - 10 of 12

Country / Medium	Acquisition			Behavior			Conversions			
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration	Goal Conversion Rate	Goal Completions	Goal Value	Goal % of Total
	5633 % of Total: 100.00% (563)	93.43% Avg. New Session: 0.00%	526 % of Total: 100.00% (526)	87.39% Avg. Bounce Rate: 0.00%	1.24 Avg. Pages per View: 0.00	00:00:29 Avg. Session Duration: 0.00	0.00% Avg. Goal Conver. 0.00% (0.00%)	0 % of Total: 0.00% (0)	\$0.00 Avg. Goal Value: \$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
And...										
1. indonesiæ / referral	271 (48.19%)						0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
2. t.co / referral United States	149 (26.47%)						0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
3. ads-bidder-api.twitter.com / referral Malaysia	81 (14.39%)						0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
4. m.facebook.com / referral Thailand	8 (1.46%)						0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
5. l.facebook.com / referral Australia	7 (1.24%)						0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
6. datascience.columbia.edu / referral United Kingdom	7 (1.24%)						0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
7. google / organic India	6 (1.07%)						0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
8. facebook.com / referral Switzerland	4 (0.71%)	100.00% (0.71%)	(0.76%)	100.00% (0.76%)	1.00	00:00:00	0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
9. ad-review-tool.twitter.biz / referral Peru	4 (0.71%)	100.00% (0.71%)	(0.76%)	75.00% (0.76%)	1.25	00:00:17	0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)
10. adwords.google.com / referral Turkey	4 (0.71%)	100.00% (0.71%)	(0.76%)	50.00% (0.76%)	1.95	00:00:06	0.00% (0.00%)	0 (0)	\$0.00 (\$0.00)	\$0.00 % of Total: 0.00% (0)

Rows 1 - 10 of 12

**Artificial intelligence  
(AI)**

# New AI can guess whether you're gay or straight from a photograph

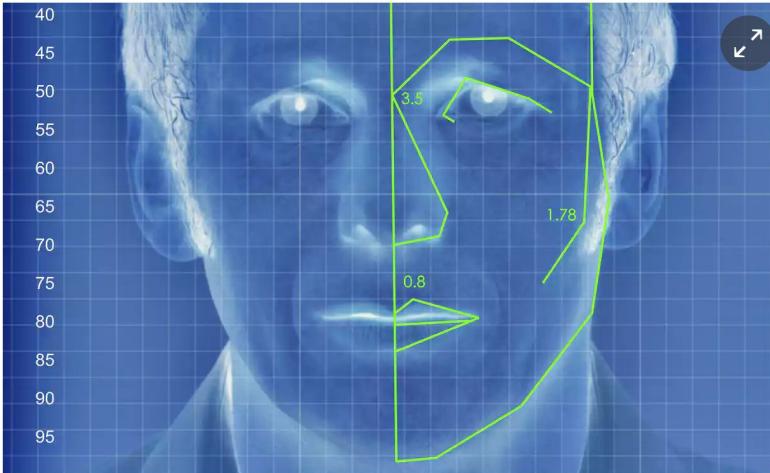
An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions

**Sam Levin** in San Francisco

 @SamTLevin

 Email

Fri 8 Sep 2017 00.46 BST



 An illustrated depiction of facial analysis technology similar to that used in the experiment. Illustration: Alamy

Artificial intelligence can accurately guess whether people are gay or straight based on photos of their faces, according to new research that suggests machines can have significantly better “gaydar” than humans.

# Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

Contributors: Yilun Wang, Michal Kosinski

Date created: 2017-02-15 08:37 AM | Last Updated: 2017-10-16 09:17 AM

Category: Project

Description: We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. Human judges achieved much lower accuracy: 61% for men and 54% for women. The accuracy of the algorithm increased to 91% and 83%, respectively, given five facial images per person. Facial features employed by the classifier included both fixed (e.g., nose shape) and transient facial features (e.g., grooming style). Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles. Prediction models aimed at gender alone allowed for detecting gay males with 57% accuracy and gay females with 58% accuracy. Those findings advance our understanding of the origins of sexual orientation and the limits of human perception. Additionally, given that companies and governments are increasingly using computer vision algorithms to detect people's intimate traits, our findings expose a threat to the privacy and safety of gay men and women.

# 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.

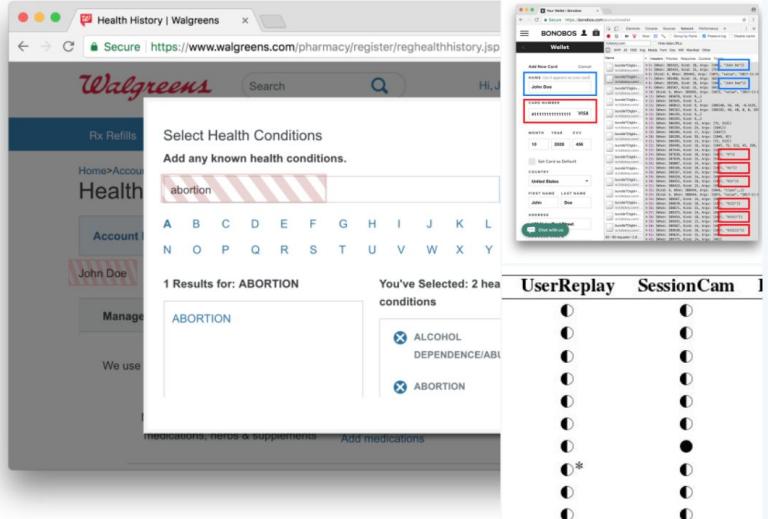
Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.

But it was too late — a neighbor who witnessed the heist had already called the police. Borden and her friend were arrested and charged with burglary and petty theft for the items, which were valued at a total of \$80.

# Session replay scripts

Arvind Narayanan @random\_walker · 15 Nov 2017

- \* 'Session replay' scripts record everything, like someone looking over your shoulder, and send it to a third party.
- \* Includes passwords, medical info, other personal info.
- \* Proper redaction is a near-impossible human & technical problem.



UserReplay	SessionCam	1
●	●	
●	●	
●	●	
●	●	
●	●	
●*	●	
●	●	
●	●	

21 677 788 1

Arvind Narayanan @random\_walker · 15 Nov 2017

Your keystrokes are recorded by third parties as soon as you type them, even if you never submit the form. If you accidentally copy-paste something, that's recorded too.



46 626 898 1

# 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

## Abstract

*We study, for the first time, automated inference on criminality based solely on still face images, which is free of any biases of subjective judgments of human observers.*

management science, criminology, etc.

In all cultures and all periods of recorded human history, people share the belief that the face alone suffices to reveal innate traits of a person. Aristotle in his famous work *Prior Analytics* asserted, "It is possible to infer character

“In all cultures and all periods of recorded human history, people share the belief that the face alone suffices to reveal innate traits of a person.”



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



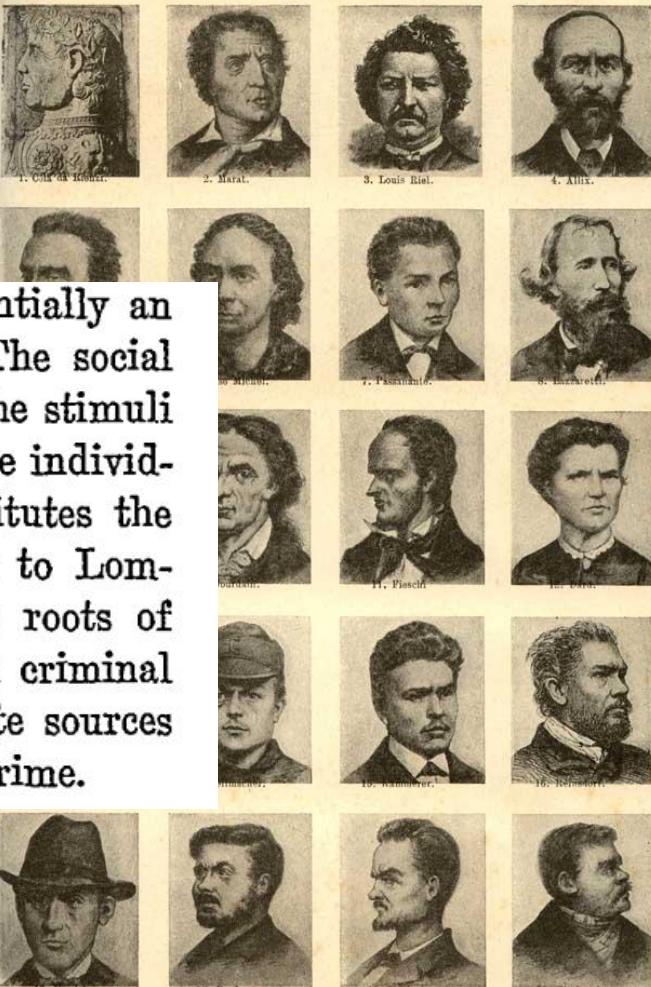
(b) Three samples in non-criminal ID photo set  $S_n$

Figure 1. Sample ID photos in our data set.

## 2. Data preparation

# We've been here before

Lombroso believed, in other words, that the criminal was essentially an organic anomaly, partly pathological and partly atavistic. The social causes of crime were at most, according to Lombroso, simply the stimuli which called forth the organic and psychical abnormalities of the individual. While the removal of the social causes of crime constitutes the immediate practical problem before criminologists, according to Lombroso, because they are the exciting causes, yet the ultimate roots of crime lie in the atavistic and degenerate heredity of the born criminal and the criminaloid, and only the extirpation of these ultimate sources of criminality can afford a final solution of the problem of crime.



# We've been here before

**Medium** [Sign in](#)



Blaise Aguera y Arcas [Follow](#)

Blaise Aguera y Arcas leads Google's AI group in Seattle. He founded Seadragon, and was one of the creators of Photosynth at Microsoft.

May 6, 2017 · 38 min read

## Physiognomy's New Clothes

by Blaise Agüera y Arcas, [Margaret Mitchell](#) and [Alexander Todorov](#)

# We've been here before

# Medium

[Sign in](#)



Blaise Aguera y Arcas [Follow](#)

Blaise Aguera y Arcas leads Google's AI group in Seattle. He founded Seadragon, and was one of the creators of Photosynth at Microsoft.

Jan 11 · 15 min read

## Do algorithms reveal sexual orientation or just expose our stereotypes?

by Blaise Aguera y Arcas, [Alexander Todorov](#) and [Margaret Mitchell](#)

# We've been here before



**Do just**

by Bla

## Medium

[Sign in](#)

This doesn't negate the privacy concerns the authors and various commentators have raised, but it emphasizes that such concerns relate less to AI per se than to mass surveillance, which is troubling regardless of the technologies used (even when, as in the days of the Stasi in East Germany, these were nothing but paper files and audiotapes). Like computers or the internal combustion engine, AI is a general-purpose technology that can be used to automate a great many tasks, including ones that should not be undertaken in the first place.

ne of the

or

We are hopeful about the confluence of new, powerful AI technologies with social science, but not because we believe in reviving the 19th century research program of inferring people's inner character from their outer appearance. Rather, we believe AI is an essential tool for understanding

# Statistical sciences always political

Dream of sciences of social difference

Central to development of

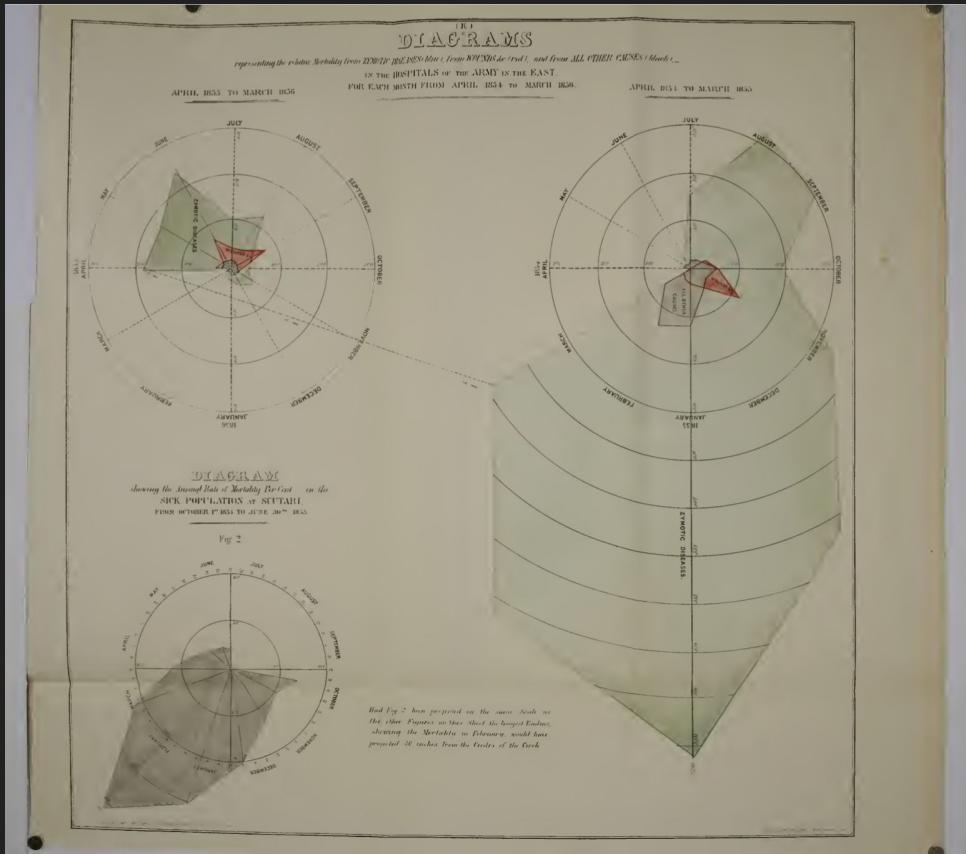
Statistics

And the

Data sciences

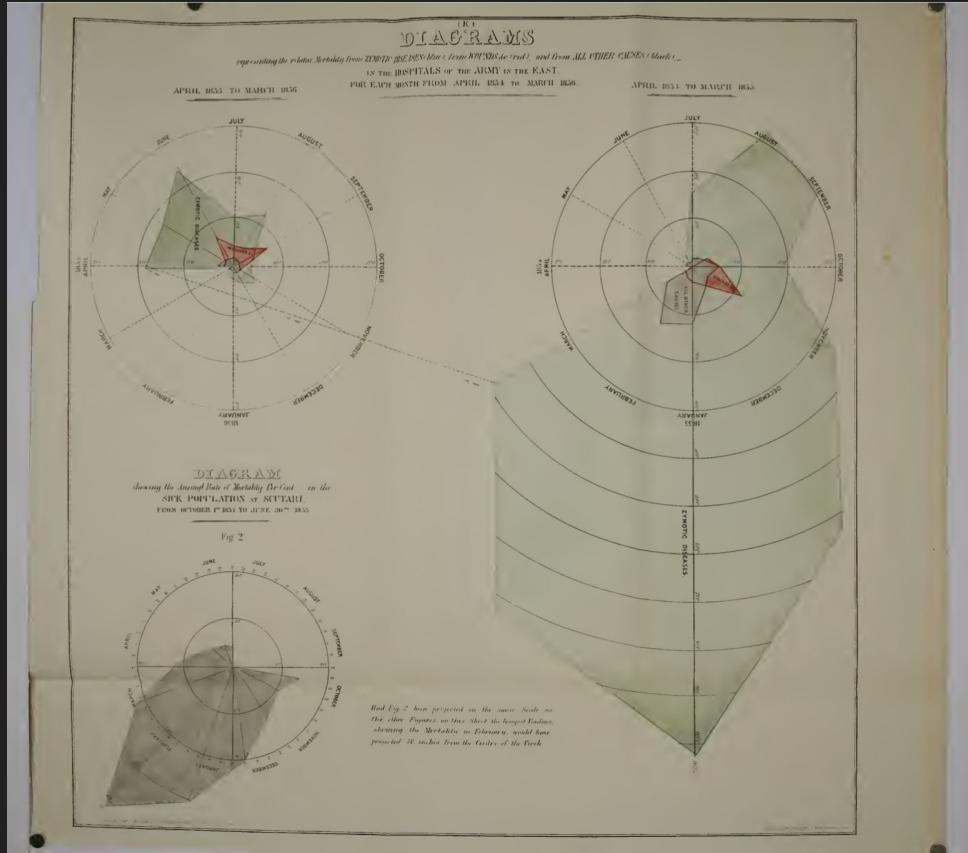
# Florence Nightingale & Data Visualization

“Experience has shown that without special information and skilful application of the resources of science in preserving health, the drain on our home population must exhaust our means. The introduction, therefore, of a proper sanitary system into the British army is of essential importance to the public interests.”



# Florence Nightingale & Data Visualization

“Upon the British race alone the integrity of that empire at this moment appears to depend. The conquering race must retain possession.”



# Every week

Scientific and mathematical development

Technologies and engineering

Driving forces: money, prestige, resources, Imperial competition

Power, ethics, and data intensive knowledge

# Tech story: three chronological stages

Data and Math

Data and Engineering

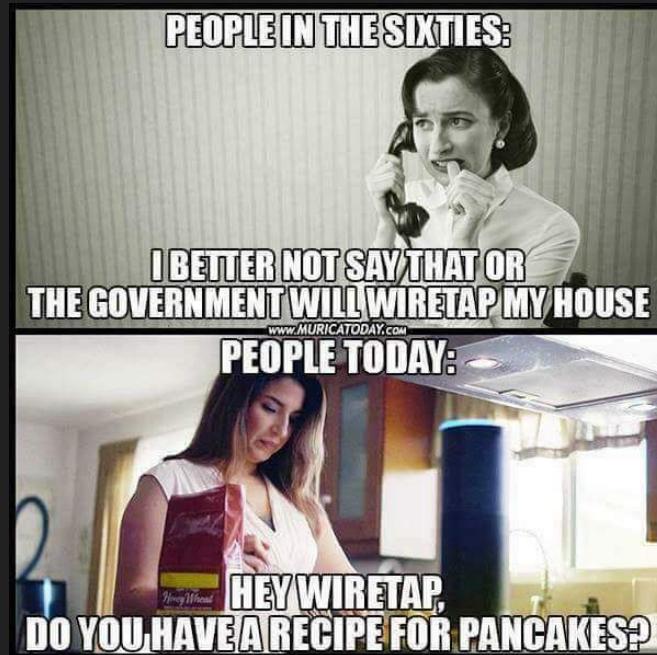
Data and Technology

# Data technologies

Census and government survey

Information processing machines and digital computers

Always on network infrastructure



# Power

How should social and political order be organized on basis  
of science and engineering?

How do technologies transform the social and political order?

How do technologies augment and diminish democratic orders? Autocratic ones?

# Power and politics\*

New technologies mean new capabilities

These capabilities are first available to those in power

(cf., “The future is already here — it's just not very evenly distributed.” --Gibson)

How does this distribution of capability reorder power?

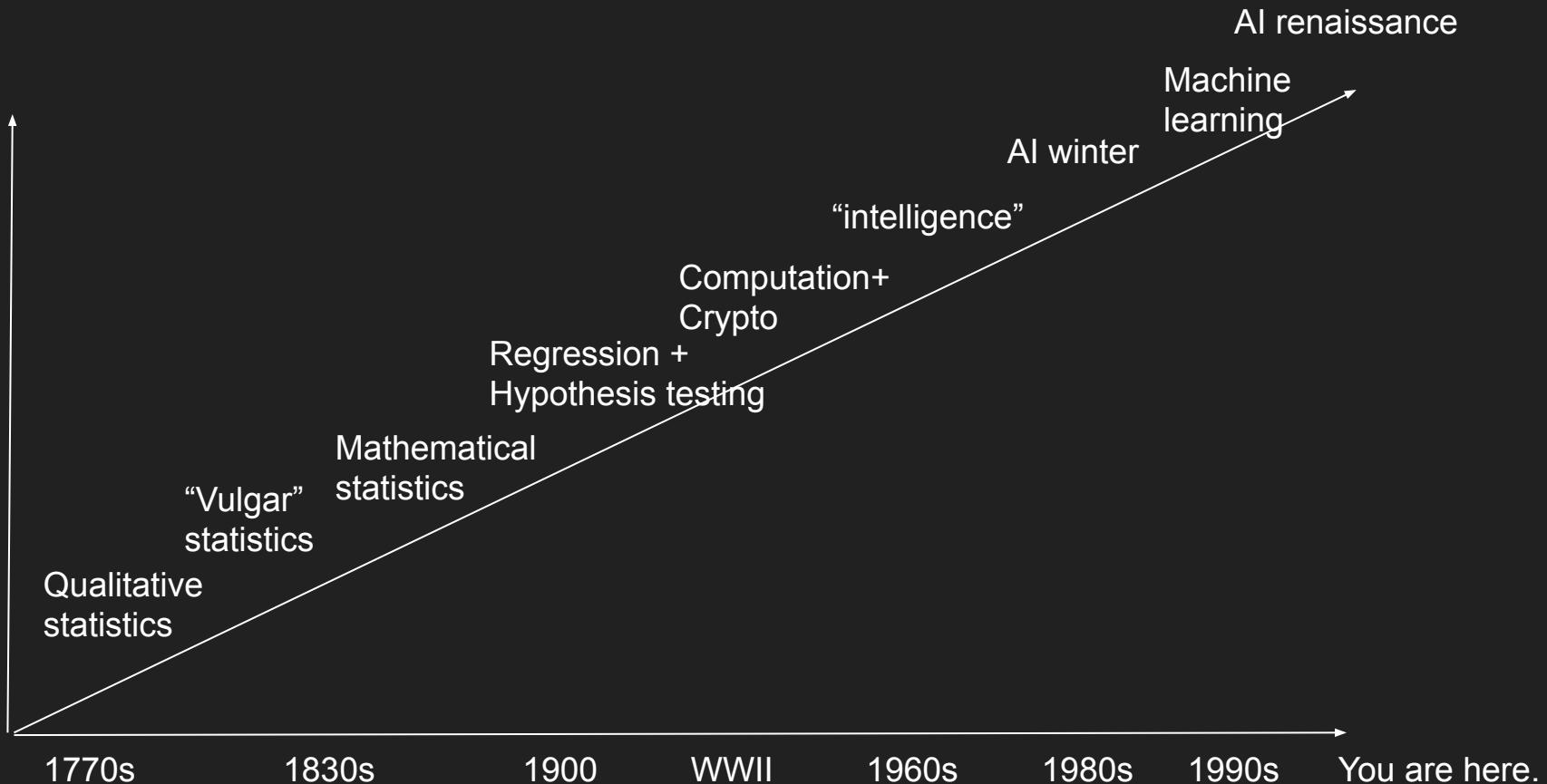
How are data-empowered algorithms an example of this dynamic

- of capability, and
- of reinforcing or distributing power?

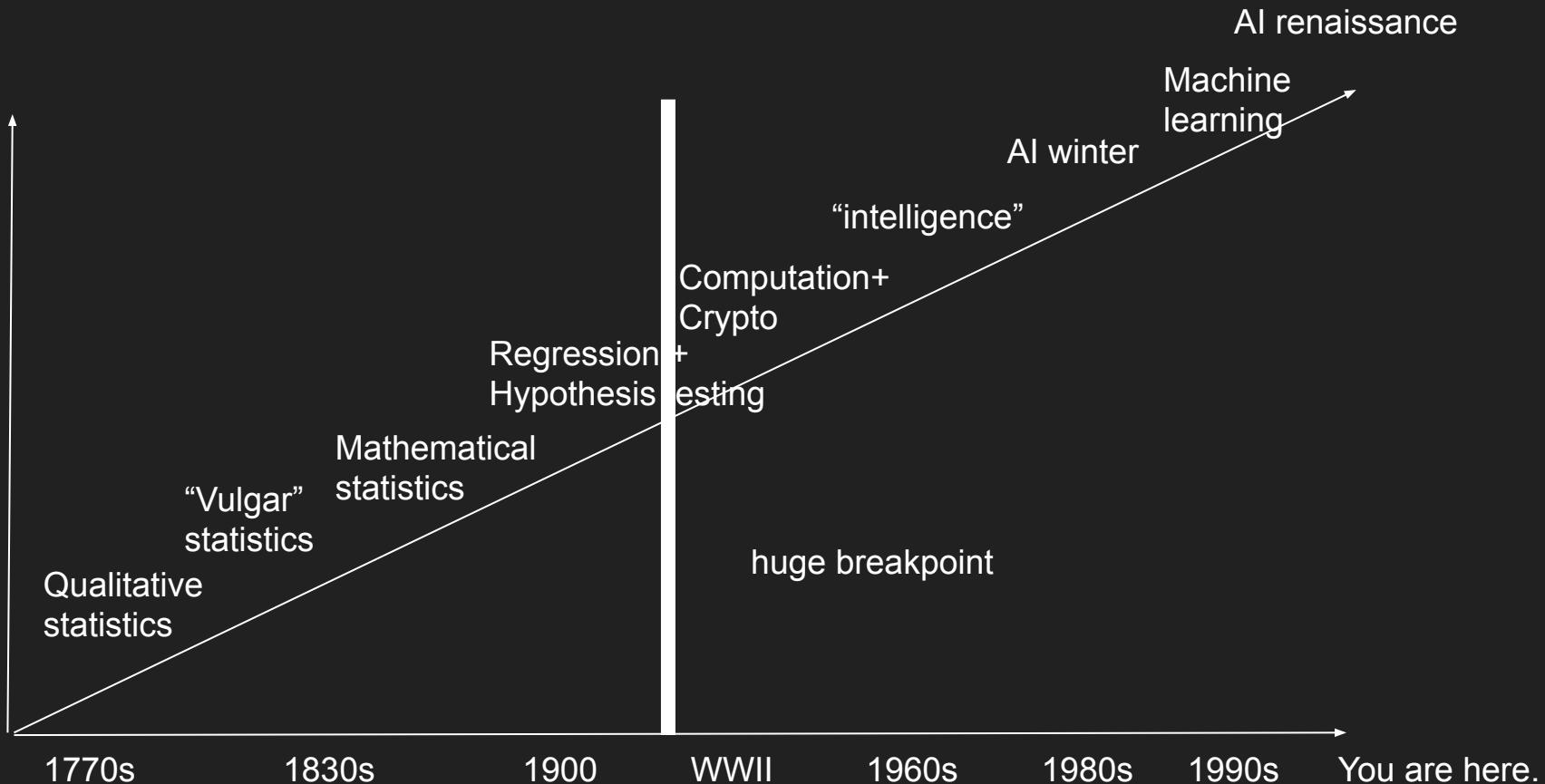
\* *politics* here meaning the dynamics of power, not to be confused with “voting”

# Arc of class

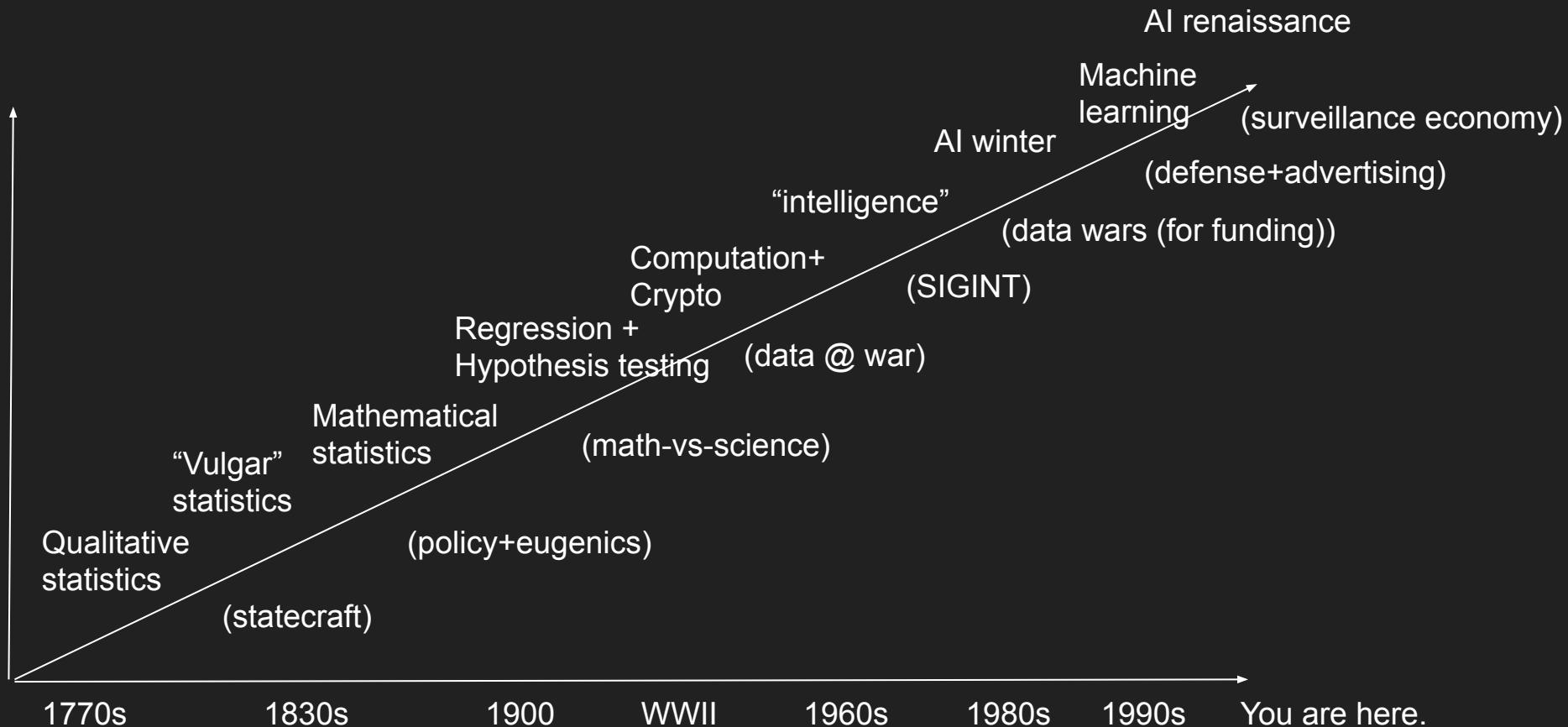
# data 1770s-present: capabilities



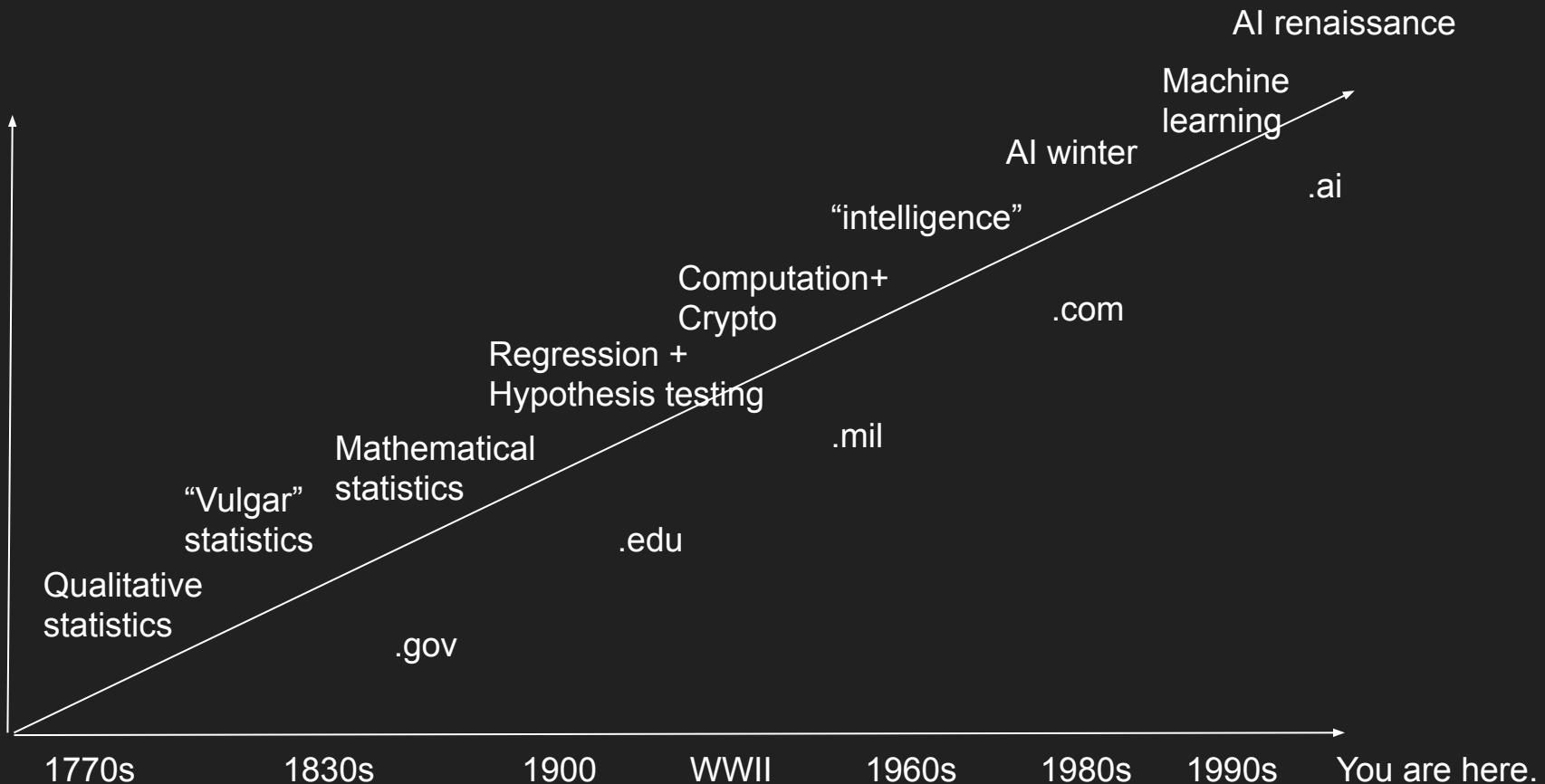
# data 1770s-present: capabilities



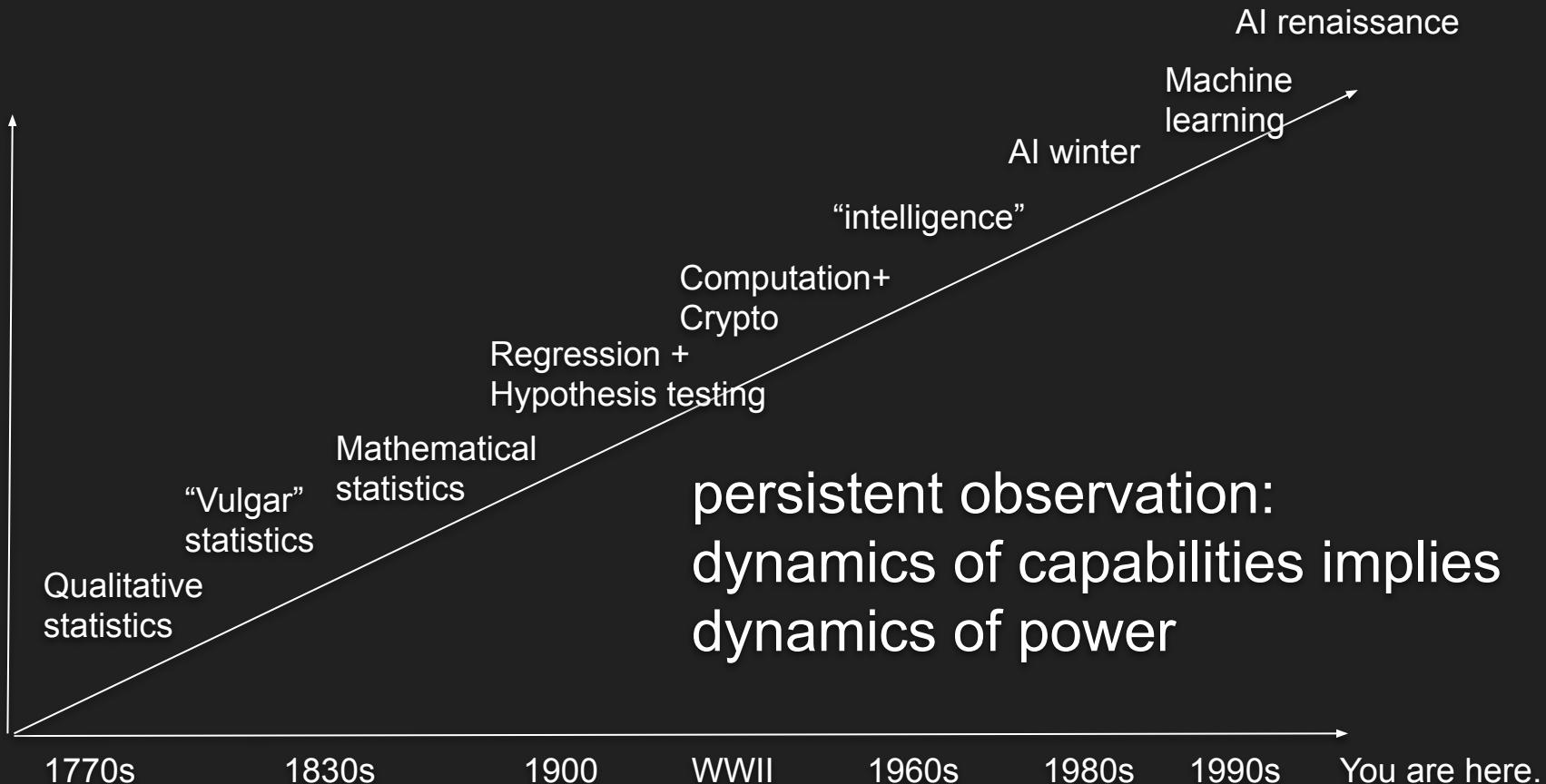
# data 1770s-present: capabilities & intents



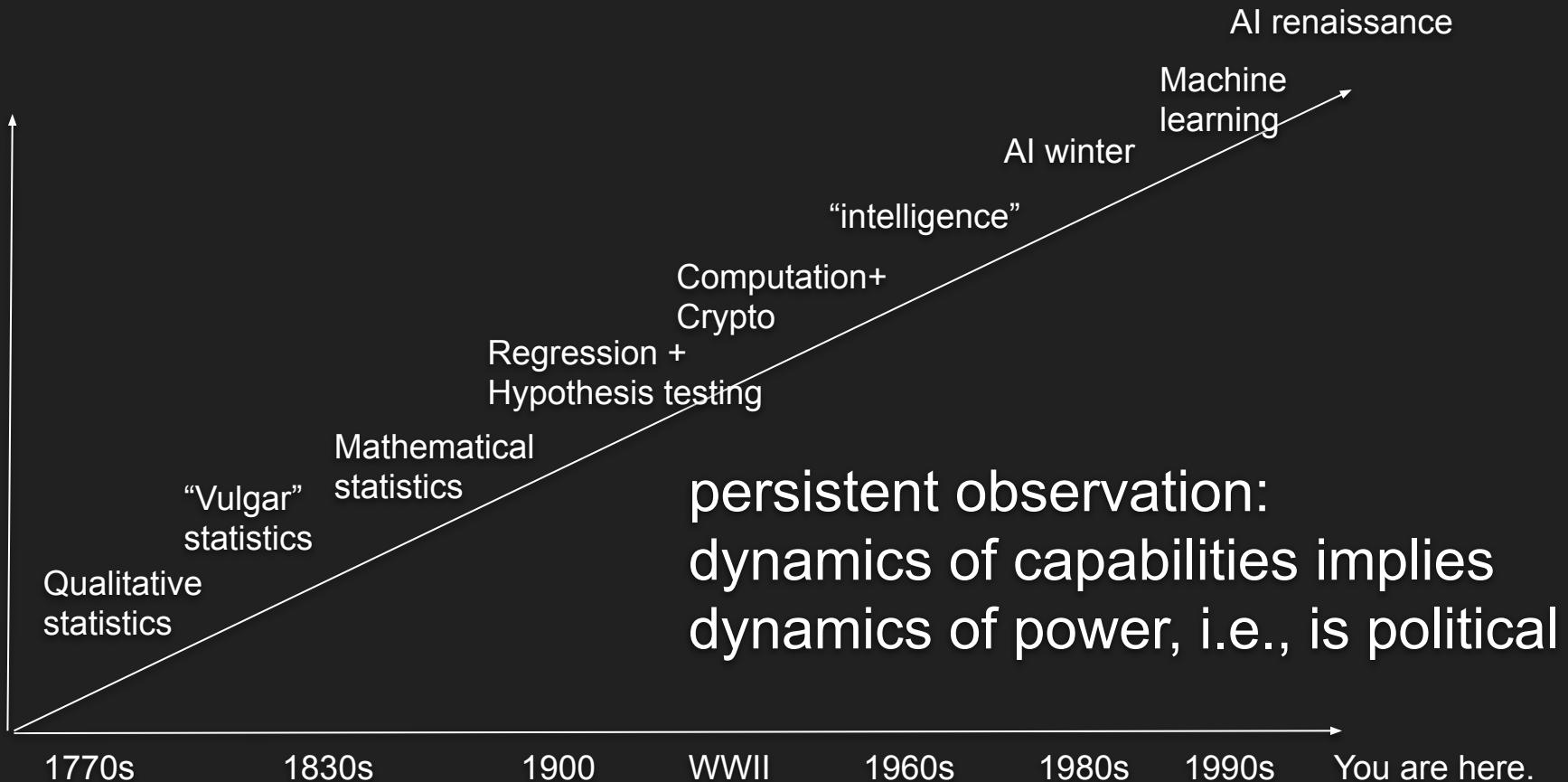
# data 1770s-present: capabilities & intents



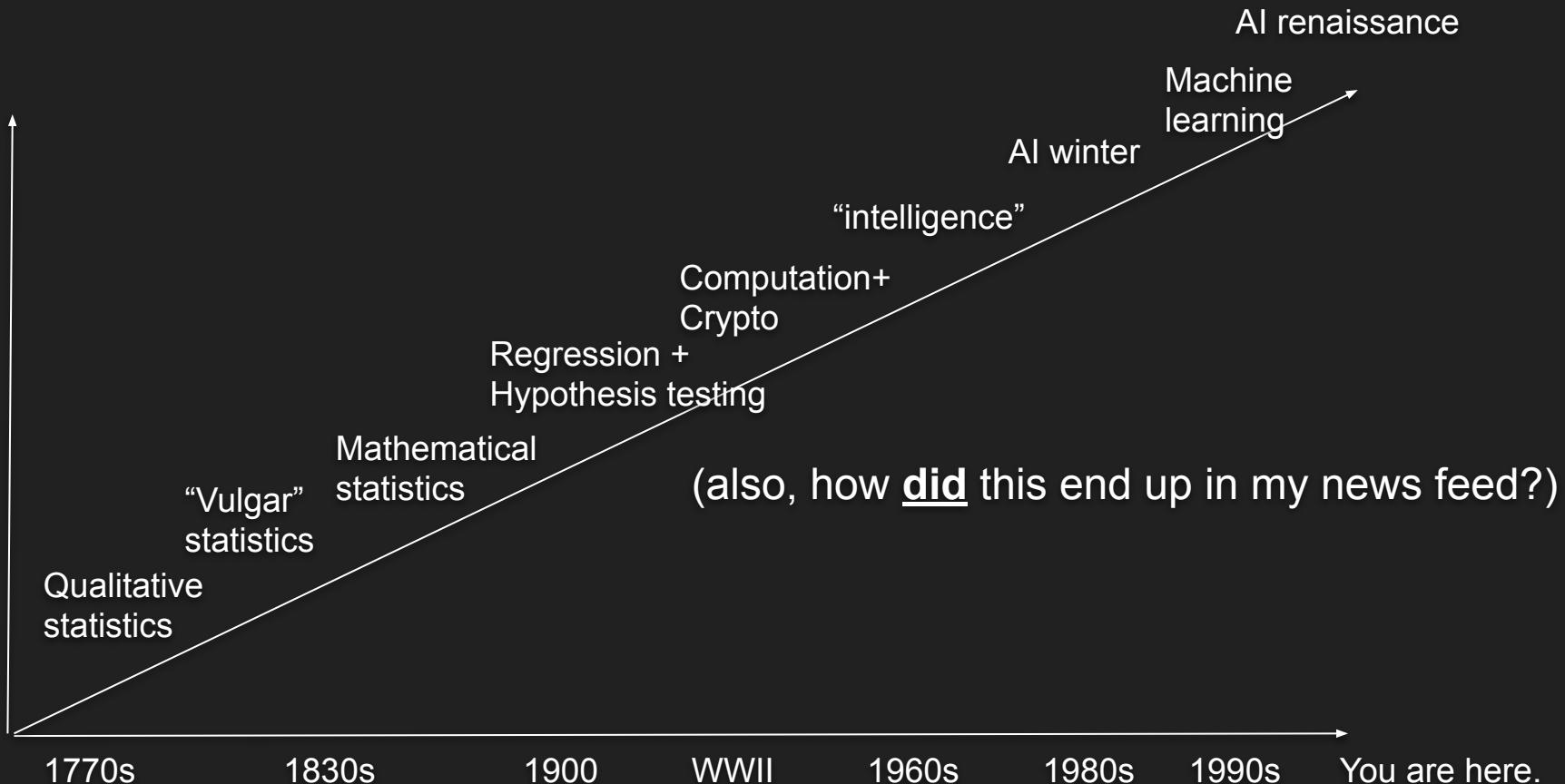
# data 1770s-present: capabilities & intents



# data 1770s-present: capabilities & intents



# data 1770s-present: capabilities & intents



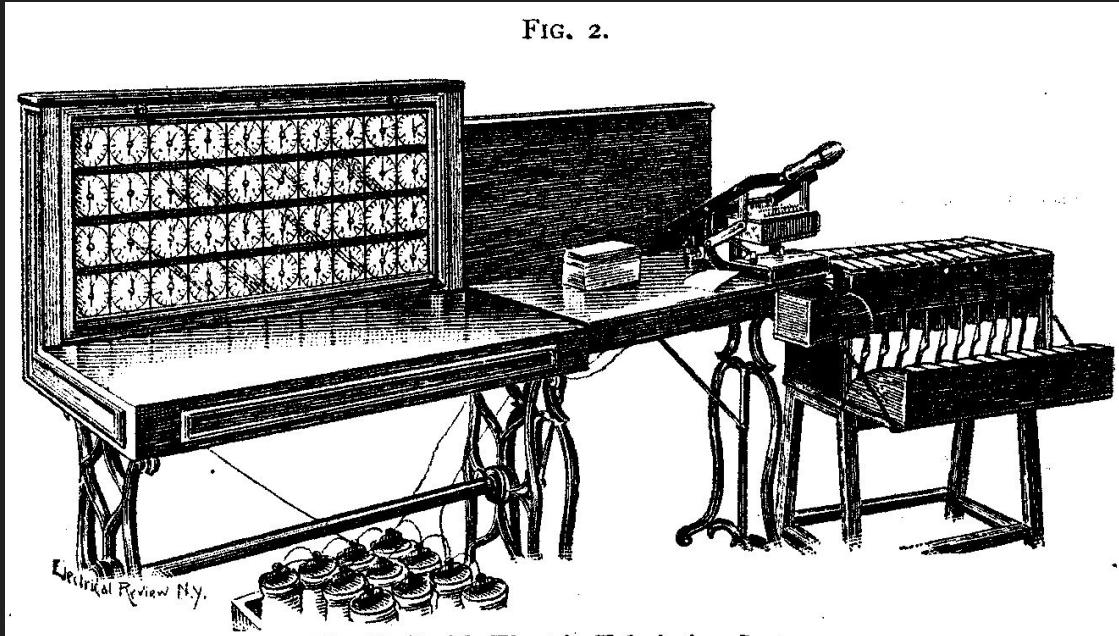
# How we know what the state of the state is

*Tabelle Schem. B.*  
Verzeichnis der Gestorbenen nach dem Alter und Jahren.  
Zusammen im Jahr 1887.

Alter und Jahre.	Männl. den Geschl.	Weibl. den Geschl.	Summa von beiden.
Erwachsene ...	266	204	470
bis zum ersten Jahr,	126	112	238
vom ersten bis zum zehn. incl.	742	724	1466
6 - 10 . . . .	218	203	421
11 - 15 . . . .	114	99	213
16 - 20 . . . .	119	118	237
21 - 25 . . . .	111	118	229
26 - 30 . . . .	116	132	248
31 - 35 . . . .	123	128	251
36 - 40 . . . .	188	199	387
41 - 45 . . . .	199	198	397
46 - 50 . . . .	222	234	456
51 - 55 . . . .	193	220	413
56 - 60 . . . .	258	301	559
61 - 65 . . . .	321	316	637
66 - 70 . . . .	298	354	652
71 - 75 . . . .	293	273	566
76 - 80 . . . .	223	266	429
81 - 85 . . . .	110	126	236
86 - 90 . . . .	76	72	148
91 - 95 . . . .	72	19	91
96 - 99 . . . .	4	1	5
100 Jahr . . . .	1	..	..
101 . . . . .	..	..	..
102 - &c. . . . .	..	..	..
<b>Summa</b>	<b>5438</b>	<b>5918</b>	<b>10356</b>

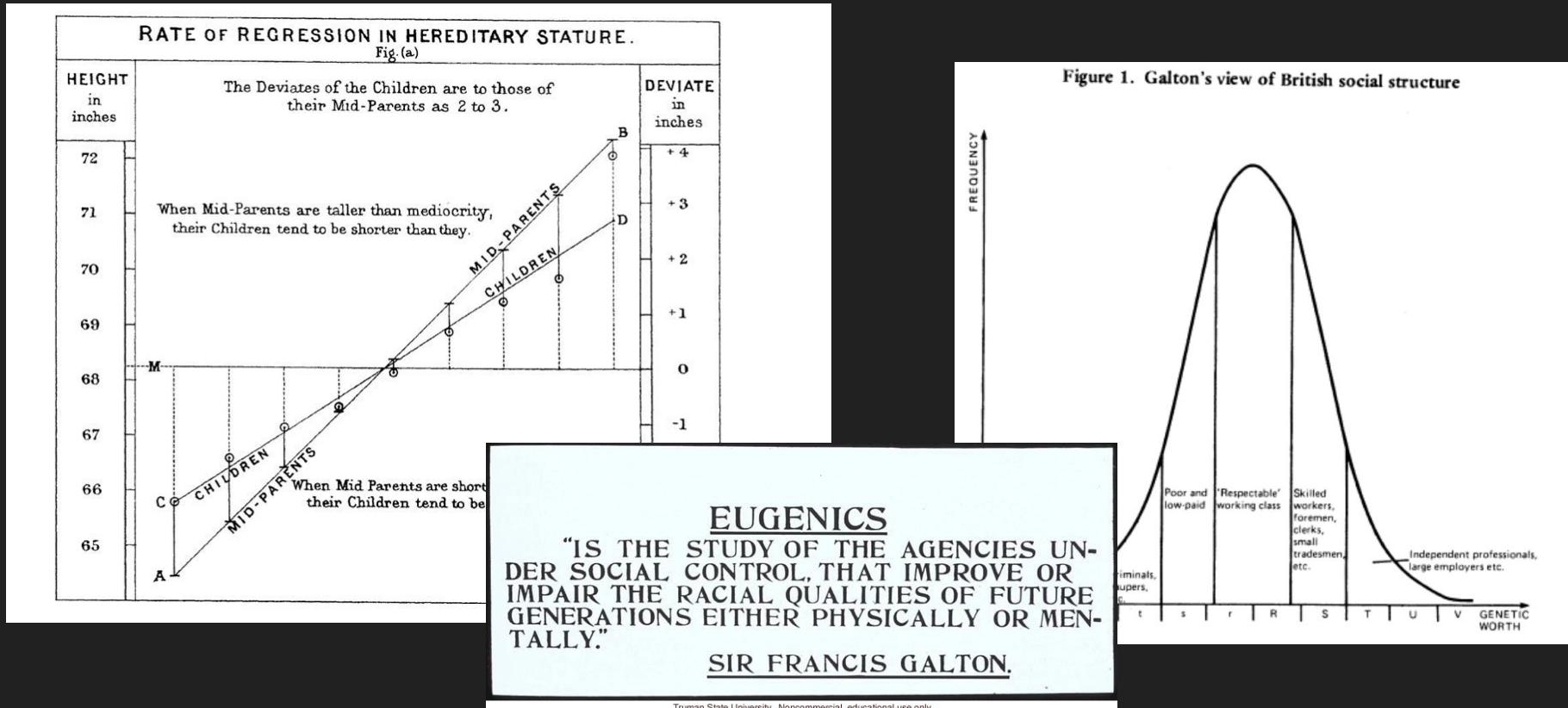
Zusammen 1. Männl. ohne Jähre ist eine Frau von ganz geistiger Größe und die sehr blöde und hässlich ist, daher von Jahr ab gesprochen.  
Zusammen 2. Ein Erwachsener ist ein Jahr als gesprochen, während das völlige Entfernen einer jungen Person ist am Ende erhalten.  
Dergleichen Zusammenfassungen geben die Zahl nicht an die Hand.

FIG. 2.



The Hollerith Electric Tabulating System.

# Regression, correlation and eugenics



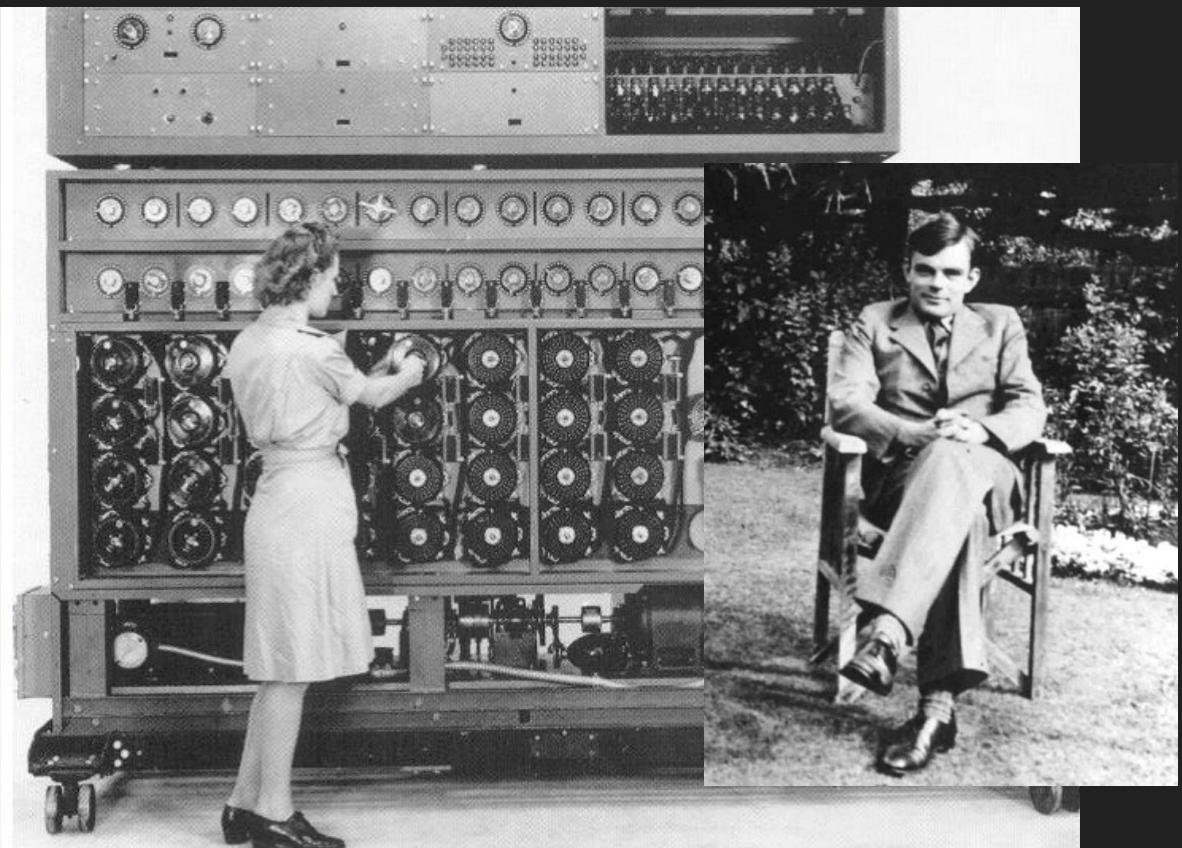
# Experimental design, hypothesis test, decision theory

To play this game with the greatest chance of success, the experimenter cannot afford to exclude the possibility of any possible arrangement of soil fertilities, and his best strategy is to equalize the chance that any treatment shall fall on any plot by determining it by chance himself.

-R.A. Fisher



# World War 2: Turing and statistical cryptography



# AI and its many winters

## The Turing Test

1950: Alan Turing's "Computing Machinery and Intelligence"  
(the "Turing Test")

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind* 49: 433–460.

### COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

#### 1. The Imitation Game



#### *Can machines think?*

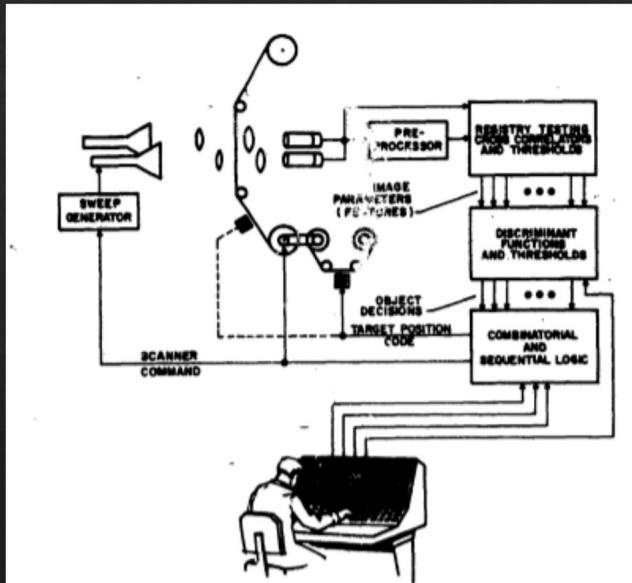
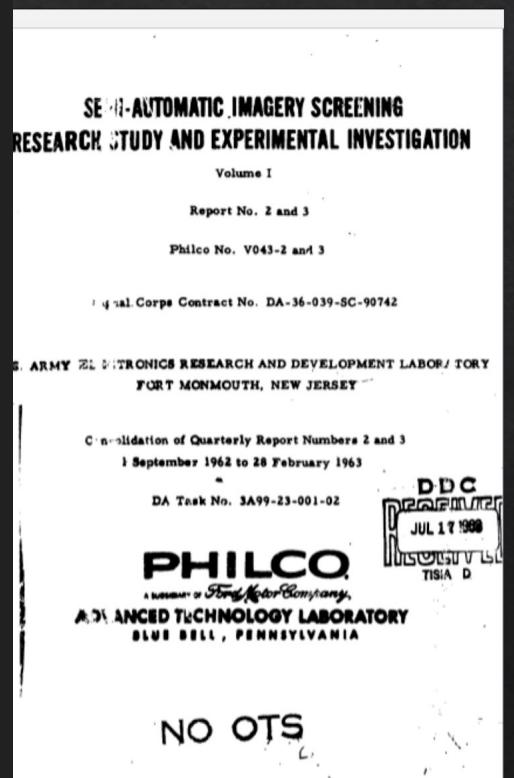
I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

# BRACE YOURSELF!



# AI WINTER IS COMING!!

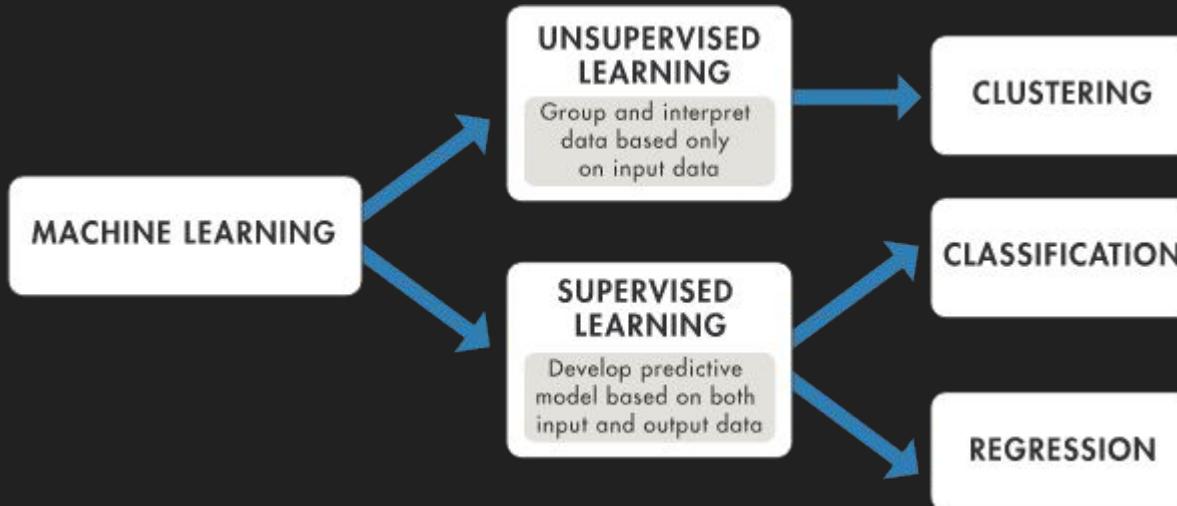
# Pattern Recognition to . . .



Conceptual Block Diagram of Image Screening System



# ... Machine Learning



... but also ethics & politics:

how did “ethics” meet “data”?

1. widely-condemned dumpster fires (1940s-1970s)
2. defining “practical” ethics (1970s)
3. enforcing / designing for ethics (1980s-present)

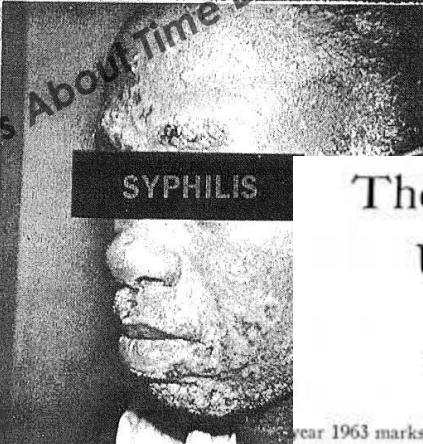
# Research ethics (and the lack thereof)

**GERM WARFARE  
DECLARED  
AGAINST BLACKS!**

HUNDREDS OF  
BLACK MEN  
DISCOVERED  
MASSACRED  
IN SYPHILIS  
"EXPERIMENT".

SEE ARTICLE INSIDE PAGE 2

*Reprinted from the front page of the New York Times, July 26, 1963.*



**SYphilis**

## The Tuskegee Study of Untreated Syphilis

*The 30th Year of Observation*

DONALD H. ROCKWELL, MD; ANNE ROOF YOUNG, MD;  
AND M. BRITTAINE MOORE, JR., MD, ATLANTA

year 1963 marks the 30th year of the evaluation of the effect of untreated syphilis in the male Negro conducted

tion such as this offered an opportunity to follow and study the disease over a long period of time. In 1932, a total

## The New York Times

**Syphilis Victims in U.S. Study Went Untreated for 40 Years**

By JEAN HELLER  
The Associated Press

WASHINGTON, July 25—For 40 years the United States Public Health Service has conducted a study in which human beings with syphilis, who were induced to serve as guinea pigs, have gone without medical treatment for the disease and a few have died of its effects, even though an effective therapy was eventually developed.

The study was conducted to determine from autopsies what disease does to the human body.

Officials of the health service, who initiated the experiment, have long since retired. Recent officials, who say they

have serious doubts about the morality of the study, also say that it is too late to treat the syphilis in any surviving participants.

Doctors in the service say they are now rendering whatever other medical services they can give to the survivors while the study of the disease's effects continues.

Dr. Merlin K. DuVal, Assistant Secretary of Health, Education and Welfare for Health and Scientific Affairs, expressed shock on learning of the study. He said that he was making an immediate investigation.

The experiment, called the Tuskegee Study, began in 1932 with about 600 black men,

# Research ethics (and the lack thereof)



**"I was going to write an angry post about Facebook's emotional manipulation study, but then I got distracted by all the happy cat pictures they showed me."**

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer<sup>a,1</sup>, Jamie E. Guillory<sup>b,2</sup>, and Jeffrey T. Hancock<sup>b,c</sup>

<sup>a</sup>Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and

Departments of <sup>b</sup>Communication and

<sup>c</sup>Information Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

## Significance

We show, via a massive ( $N = 689,003$ ) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues.

# Silicon Valley and the Attention Economy



Don't make the mistake of thinking you're Facebook's customer, you're not - you're the product,

— Bruce Schneier —

AZ QUOTES

# De-anonymization



# Weekly structure

**Tuesday**

Lecture and discussion

Expectation

arrive having done the week's readings

**Thursdsay**

Laboratory

Expectation

arrive with laptop ready to collaborate

# Weekly structure

## Tuesday

- Lessons from Slack
- “political” precondition
- intellectual advance
  - secondary readings
  - primary readings
- “political” postcondition
- what is contemporary analogue?
  - in capability
  - in politics
  - in data...

## Thursdsay

- Laboratory
  - ...motivated by readings
  - ...and presaging homeworks!
- Expectation
  - arrive with laptop ready to collaborate

“political” meaning impact on *human rights, justice, harms and benefits --- and on power*

No prerequisites

Instructors' permission  
NOT  
required

# Two tracks

more technical background track (60%)

- pursue a semester long project culminating in a 15pp paper and any associated code
- 
- complete 3 problem sets
- 
- short final presentation on paper

more humanistic background track (60%)

- write a 10 pp paper on a topic of their choice
- 
- complete 5 problem sets, these problem sets will involve both computational work and writing work
- 
- short final presentation on paper

# Jupyter notebooks: code without any coding

Using

Codio platform

With everything you'll need installed

USING JUPYTER NOTEBOOKS

## Data: Past, Present, Future

### Lab 6: Gould 2.0: IQ and Brains, again

In this lab, you will work with some data sets concerning physical attributes of brains. You will undertake a series of statistical examinations AND you will also reflect critically on the nature of these data sets, and the lictiness of different computational operations on data sets.

```
In [ 1]: # usual preliminaries  
%matplotlib inline  
  
import pandas as pd  
import matplotlib.pyplot as plt  
import numpy as np  
plt.style.use('fivethirtyeight')  
plt.rcParams['figure.figsize'] = (15, 5)
```

Let's start with some data. This data set is used in boatloads of intro stats courses. It's a mechanical sort.

```
In [ 1]: data = pd.read_csv('http://www.scipy-lectures.org/_downloads/  
    ".*")
```

Using standard pandas and scipy functions, start describing this data set. Examples go forth. What and where are the outliers?

```
In [ 1]:
```

Now try doing some comparisons of male and female brains. Divide and conquer.

# Required work

Postings on readings in slack

Problem sets (extensions of lab work, done in Jupyter)

Final paper

Your presence

# Registration

**History-APMA UN2901 section 001**

**Call Number    16947**

- **The course satisfies one science requirement for CC and GS students and counts towards "nontech elective" credit for SEAS students.**
- **This course satisfies Barnard's T&D general education requirement.**
- **This course counts as a history course for the history major.**
- **NO PREREQUISITES AND NO PROGRAMMING EXPERIENCE REQUIRED!**