

what should future
statisticians CEOs, and senators
know about the history and ethics of data?



chris.wiggins@columbia.edu
chris.wiggins@nytimes.com
chris.wiggins@hackNY.org
[@chrishwiggins](https://twitter.com/chrishwiggins)

data-ppf.github.io

what should future
statisticians CEOs, and senators
know about the history and ethics of data?



chris.wiggins@columbia.edu
chris.wiggins@nytimes.com
chris.wiggins@hackNY.org
[@chrishwiggins](https://twitter.com/chrishwiggins)

data-ppf.github.io

what should future
statisticians CEOs, and senators
know about the history and ethics of data?



chris.wiggins@columbia.edu
joint work with:

Matt Jones
Department of History, Columbia
@nescioquid

data-ppf.github.io

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

1. why history?

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

1. why history?
2. why ethics?

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

1. why history?
2. why ethics?
3. what we taught

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

1. why history?
2. why ethics?
3. what we taught
4. what we learned

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

- 0. preamble: class origin story
- 1. why history?
- 2. why ethics?
- 3. what we taught
- 4. what we learned

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

- 0. preamble: class origin story
- 1. why history?
- 2. why ethics?
- 3. what we taught
- 4. what we learned

0. preamble: class origin story

chris.wiggins@columbia.edu <chris.wiggins@columbia.edu>

to Matthew ▾

great!

On 5/11/16, Matthew L. Jones <mj340@columbia.edu> wrote:

> You beat me to the forward. Yes! Let's do it.

>

> Sent from my Amiga 1000

>

> On May 11, 2016, at 2:56 PM, "chris.wiggins@columbia.edu" <

> chris.wiggins@columbia.edu> wrote:

>

> interested in applying?

>

> ----- Forwarded message -----

> From: <collaboratory@columbia.edu>

> Date: Wed, May 11, 2016 at 2:49 PM

> Subject: Funding Opportunity from The Collaboratory@columbia.edu

0. preamble: class origin story

chris.wiggins@columbia.edu <chris.wiggins@columbia.edu>

to Matthew ▾

great!

On 5/11/16, Matthew L. Jones <mj340@columbia.edu> wrote:

> You beat me to the forward. Yes! Let's do it.

>

> Sent from my Amiga 1000

>

> On May 11, 2016, at 2:56 PM, "chris.wiggins@columbia.edu" <

> chris.wiggins@columbia.edu> wrote:

>

> interested in applying?

>

> ----- Forwarded message -----

> From: <collaboratory@columbia.edu>

> Date: Wed, May 11, 2016 at 2:49 PM

> Subject: Funding Opportunity from The Collaboratory@columbia.edu



0. preamble: class origin story

chris.wiggins@columbia.edu <chris.wiggins@columbia.edu>

to Matthew ▾

great!

On 5/11/16, Matthew L. Jones <mj340@columbia.edu> wrote:

> You beat me to the forward. Yes! Let's do it.

>

> Sent from my Amiga 1000

>

> On May 11, 2016, at 2:56 PM, "chris.wiggins@columbia.edu" <

> chris.wiggins@columbia.edu> wrote:

>

> interested in applying?

>

> ----- Forwarded message -----

> From: <collaboratory@columbia.edu>

> Date: Wed, May 11, 2016 at 2:49 PM

> Subject: Funding Opportunity from The Collaboratory@columbia.edu

(we got the grant, btw)

0. preamble: class origin story

chris.wiggins@columbia.edu <chris.wiggins@columbia.edu>

Wed, May 1, 2013, 5:55 AM

to Big ▾

i'd like to go. sounds good.

On Wed, May 1, 2013 at 12:09 AM, Rapporteur, Big Data at Columbia

<bigdata.columbia@gmail.com> wrote:

> Dear colleagues,

>

> Many thanks to those who have already RSVP'd. The University Seminar on Big

> Data and Digital Scholarship is pleased to announce our next meeting:

>

> SPEAKER:

> Prof. Matthew Jones

> James R. Barker Assoc. Professor of Contemporary Civilization

> Department of History

> Columbia University

>

> TITLE and ABSTRACT:

>

> A Tale of Two Algorithms: Toward a history of Data Mining

>

0. preamble: class origin story

chris.wiggins@columbia.edu <chris.wiggins@columbia.edu>

Wed, May 1, 2013, 5:55 AM

to Big ▾

i'd like to go. sounds good.

On Wed, May 1, 2013 at 12:09 AM, Rapporteur, Big Data at Columbia

<bigdata.columbia@gmail.com> wrote:

> Dear colleagues,

>

> Many thanks to those who have already RSVP'd. The University Seminar on Big

> Data and Digital Scholarship is pleased to announce our next meeting:

>

> SPEAKER:

> Prof. Matthew Jones

> James R. Barker Assoc. Professor of Contemporary Civilization

> Department of History

> Columbia University

>

> TITLE and ABSTRACT:

>

> A Tale of Two Algorithms: Toward a history of Data Mining

(it was a good talk, btw)

0. preamble: class origin story

The screenshot shows a GitHub organization page for "ledeprogram".

Organization Details:

- Name:** The Lede Program: An Introduction to Data Practices
- Description:** a post-bac certification program in data and data technologies, all taught in the context of the social sciences and humanities.
- Location:** 10027
- Contact:** <http://www.journalism.c...>, jrncomputation@columb...

Repository Navigation:

- Repositories:** 3 (highlighted)
- People:** 0
- Projects:** 0

Search and Filters:

- Find a repository...
- Type: All
- Language: All

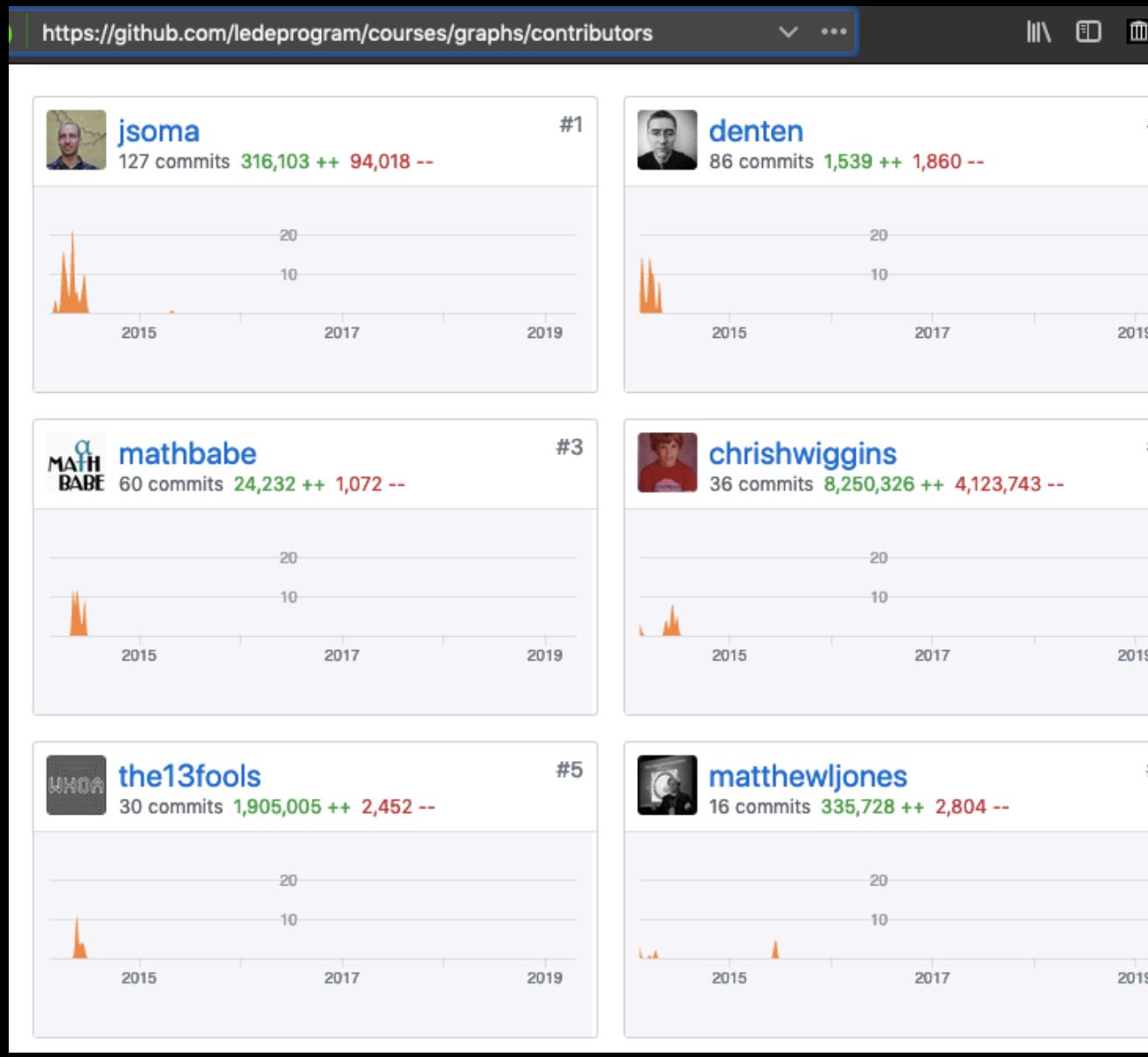
Repositories Listed:

- algorithms**
Repository for the Algorithms course
Jupyter Notebook · GPL-3.0 · 34 · 13 · 0 · 6 · Updated on Aug 17, 2016
- data-and-databases**
Repository for course materials related to Data and Databases. (Supercedes directories in "courses" repo)
Jupyter Notebook · 17 · 10 · 0 · 0 · Updated on Jul 7, 2016

Right Sidebar:

- Top languages:** Jupyter Notebook (red dot), HTML (orange dot)
- People:** 0 >
This organization has no public members. You must be a member to see who's a part of this organization.

0. preamble: class origin story



what should future
statisticians CEOs, and senators
know about the history and ethics of data?

- 0. preamble: class origin story
- 1. why history?
- 2. why ethics?
- 3. what we taught
- 4. what we learned

1. why history?

1. why history?

NEWS

Historians Politely Remind Nation To Check What's Happened In Past Before Making Any Big Decisions

9/28/11 9:00am • SEE MORE: SCIENCE & TECHNOLOGY ▾

[f](#) [t](#) [e](#) [l](#)



A photograph of a man with a white beard and glasses, wearing a brown blazer over a maroon shirt, standing at a podium and gesturing with his right hand. Behind him is a blue curtain and a black and white historical photograph of a crowd of people in hats from the early 20th century.

Trying to avoid repeating bad things we did in the past is a good idea, historians say.

1. why history?

NEWS

Historians Politely Remind Nation To Check What's Happened In Past Before Making Any Big Decisions

9/28/11 9:00am • SEE MORE: SCIENCE & TECHNOLOGY ▾

[f](#) [t](#) [e](#) [l](#)

the onion dot com

Trying to avoid repeating bad things we did in the past is a good idea, historians say.

1. why history?

Chapter 1 <i>Historical Development</i>	3
1. Energy Quantization and Heat Capacities	4
2. Blackbody Radiation	18
3. Photons	22
4. Spectra and Energy Quantization of Atoms	23
5. Matter Waves	26
6. Schrödinger's Equation	30
7. Remarks on Motion in One Dimension	41
8. Probability Interpretation	48
9. Cold Fusion*	52
10. Momentum	60
11. Expectation Values and the Momentum Operator	70
12. Many-Particle Systems	73
<i>Problems</i>	82

"Quantum Mechanics", P J E Peebles, 1992

1. why history?

Chapter 1	<i>Historical Development</i>	3
1.	Energy Quantization and Heat Capacities	4
2.	Blackbody Radiation	18
3.	Photons	22
4.	Spectra and Energy Quantization of Atoms	23
5.	Matter Waves	26
6.	Schrödinger's Equation	30
7.	Remarks on Motion in One Dimension	41
8.	Probability Interpretation	48
9.	Cold Fusion*	52
10.	Momentum	60
11.	Expectation Values and the Momentum Operator	70
12.	Many-Particle Systems	73
	<i>Problems</i>	82

"Quantum Mechanics", P J E Peebles, 1992

1. why history?

Chapter 1	<i>Historical Development</i>	3
1.	Energy Quantization and Heat Capacities	4
2.	Blackbody Radiation	18
3.	Photons	22
4.	Spectra and Energy Quantization of Atoms	23
5.	Matter Waves	26
6.	Schrödinger's Equation	30
7.	Remarks on Motion in One Dimension	41
8.	Probability Interpretation	48
9.	Cold Fusion*	52
10.	Momentum	60
11.	Expectation Values and the Momentum Operator	70
12.	Many-Particle Systems	73
	<i>Problems</i>	82

"Quantum Mechanics", P J E Peebles, 1992

1. why history?

Chapter 1 <i>Historical Development</i>	3
1. Energy Quantization and Heat Capacities	4
2. Blackbody Radiation	18
3. Photons	22
4. Spectra and Energy Quantization of Atoms	23
5. Matter Waves	26
6. Schrödinger's Equation	30
7. Remarks on Motion in One Dimension	41
8. Probability Interpretation	48
9. Cold Fusion*	52
10. Momentum	60
11. Expectation Values and the Momentum Operator	70
12. Many-Particle Systems	73
<i>Problems</i>	82

truth is contested

"Quantum Mechanics", P J E Peebles, 1992

1. why history?

The Moral Character of Cryptographic Work*

Phillip Rogaway

Department of Computer Science
University of California, Davis, USA
rogaway@cs.ucdavis.edu

December 2015
(minor revisions March 2016)

Abstract. Cryptography rearranges power: it configures who can do what, from what. This makes cryptography an inherently *political* tool, and it confers on the field an intrinsically *moral* dimension. The Snowden revelations motivate a reassessment of the political and moral positioning of cryptography. They lead one to ask if our inability to effectively address mass surveillance constitutes a failure of our field. I believe that it does. I call for a community-wide effort to develop more effective means to resist mass surveillance. I plead for a reinvention of our disciplinary culture to attend not only to puzzles and math, but, also, to the societal implications of our work.

1. why history?

The Moral Character of Cryptographic Work*

Phillip Rogaway

Abstract. Cryptography rearranges power: it configures who can do what, from what. This makes cryptography an inherently *political* tool, and it confers on the field an intrinsically *moral* dimension. The Snowden revelations motivate a reassessment of the political and moral positioning of cryptography. They lead one to ask if our inability to effectively address mass surveillance constitutes a failure of our field. I believe that it does. I call for a community-wide effort to develop more effective means

of cryptography. They lead one to ask if our inability to effectively address mass surveillance constitutes a failure of our field. I believe that it does. I call for a community-wide effort to develop more effective means to resist mass surveillance. I plead for a reinvention of our disciplinary culture to attend not only to puzzles and math, but, also, to the societal implications of our work.

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

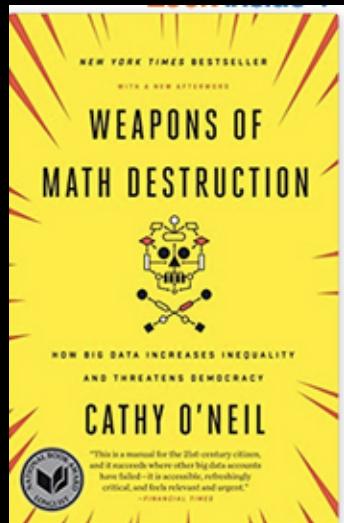
- 0. preamble: class origin story
- 1. why history?
- 2. why ethics?
- 3. what we taught
- 4. what we learned

1. why ethics?

1. why ethics?

something is wrong on the internet

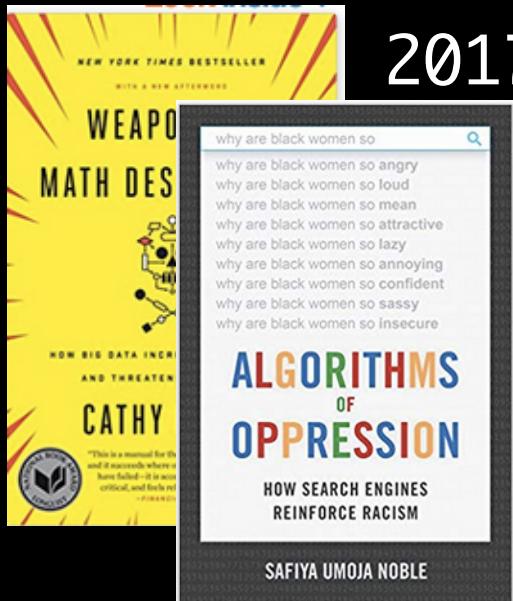
1. why ethics?



2017-09-05: cathy o'neil

something is wrong on the internet

1. why ethics?

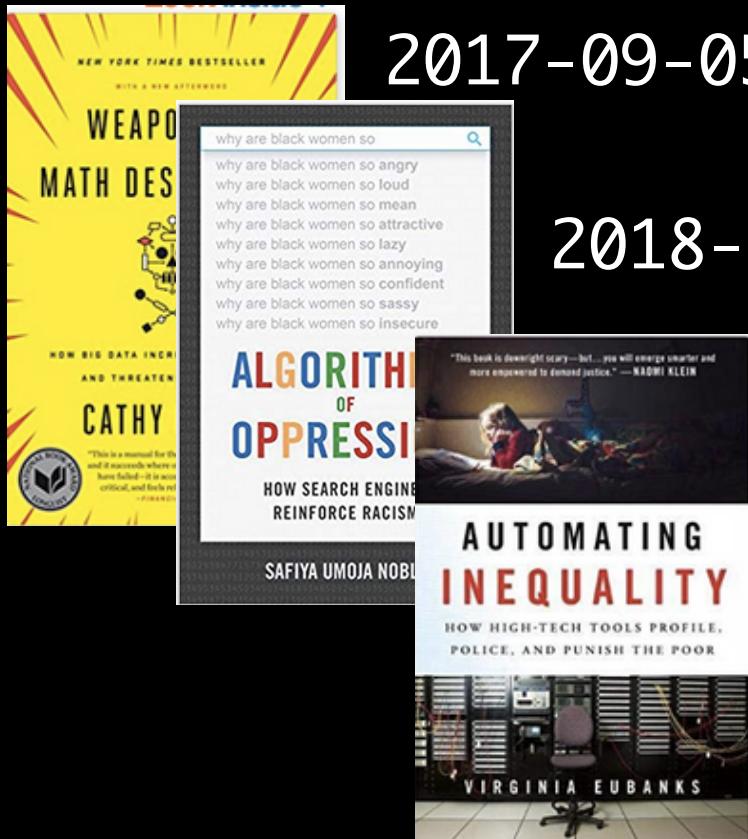


2017-09-05: cathy o'neil

2018-01-08: safiya noble

something is wrong on the internet

1. why ethics?



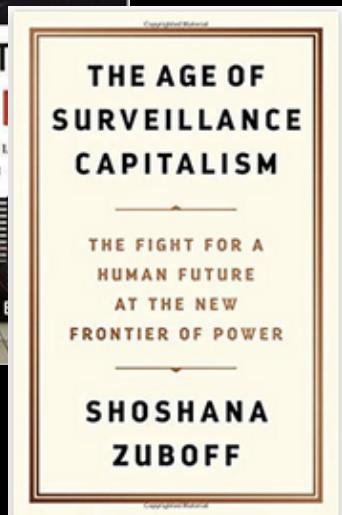
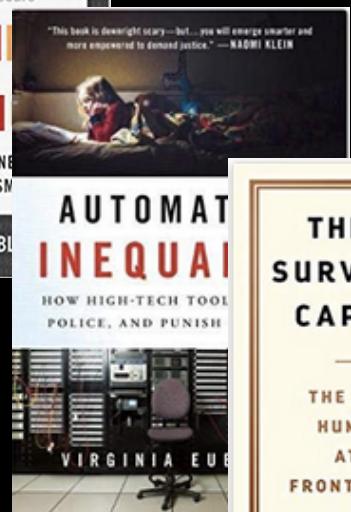
2017-09-05: cathy o'neil

2018-01-08: safiya noble

2018-01-23: virginia eubanks

something is wrong on the internet

1. why ethics?



2017-09-05: cathy o'neil

2018-01-08: safiya noble

2018-01-23: virginia eubanks

2019-01-15:
shoshana zuboff

(among increasingly many others)
something is wrong on the internet

1. why ethics?

1. fuzzies
2. techies

1. why ethics?

1. fuzzies
2. techies

FOR IMMEDIATE RELEASE

Friday, August 25, 2017

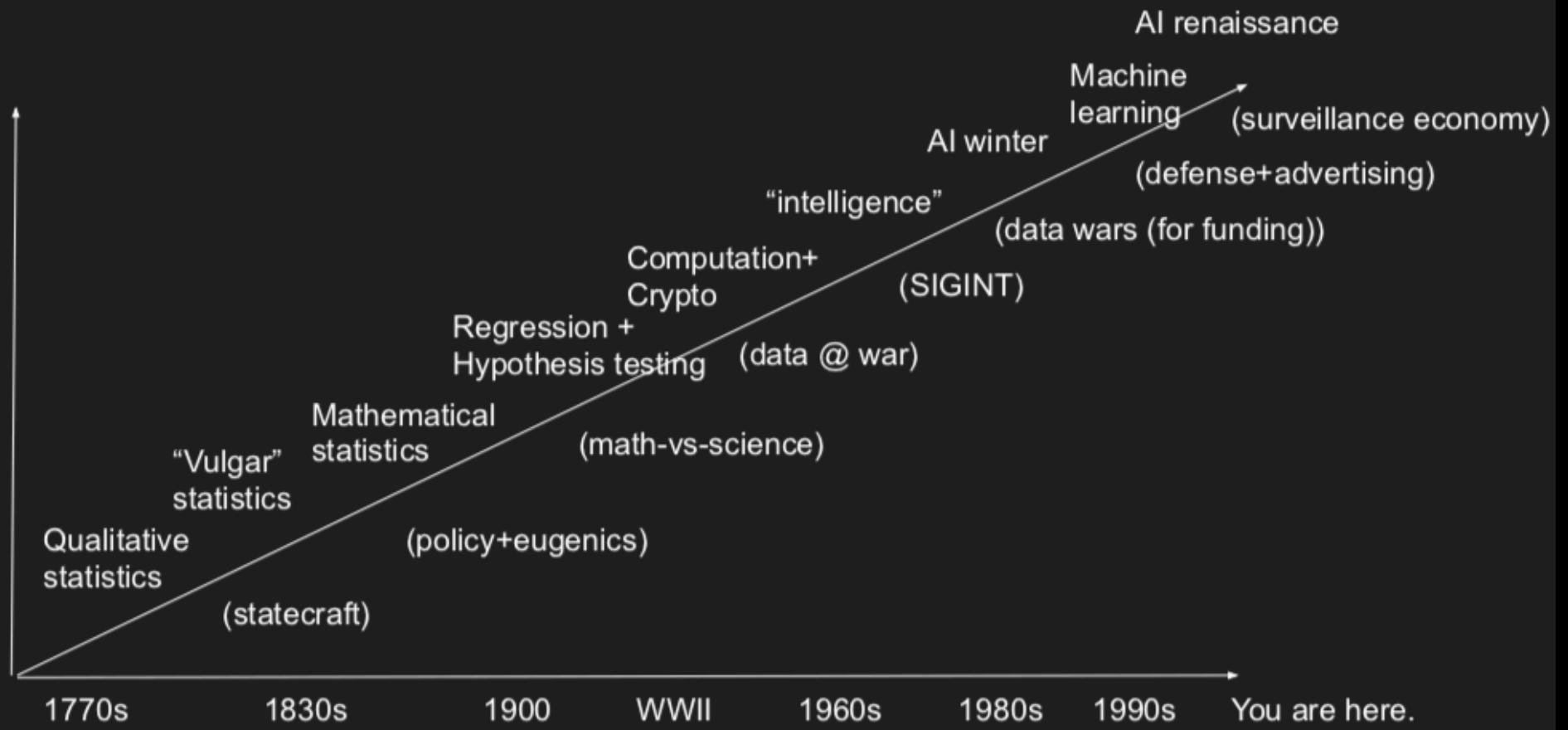
Volkswagen Engineer Sentenced for His Role in Conspiracy to Cheat U.S. Emissions Tests

A Volkswagen engineer was sentenced today by U.S. District Judge Sean F. Cox of the Eastern District of Michigan to 40 months in federal prison, and two years of supervised release, for his role in a nearly 10-year conspiracy to defraud U.S. regulators and Volkswagen customers by implementing software specifically designed to cheat emissions tests in hundreds of thousands of Volkswagen “clean diesel” vehicles sold in the U.S., the Justice Department announced today. During the hearing, the Court noted that the sentence took into consideration the defendant’s cooperation in the investigation and prosecution of the company and others.

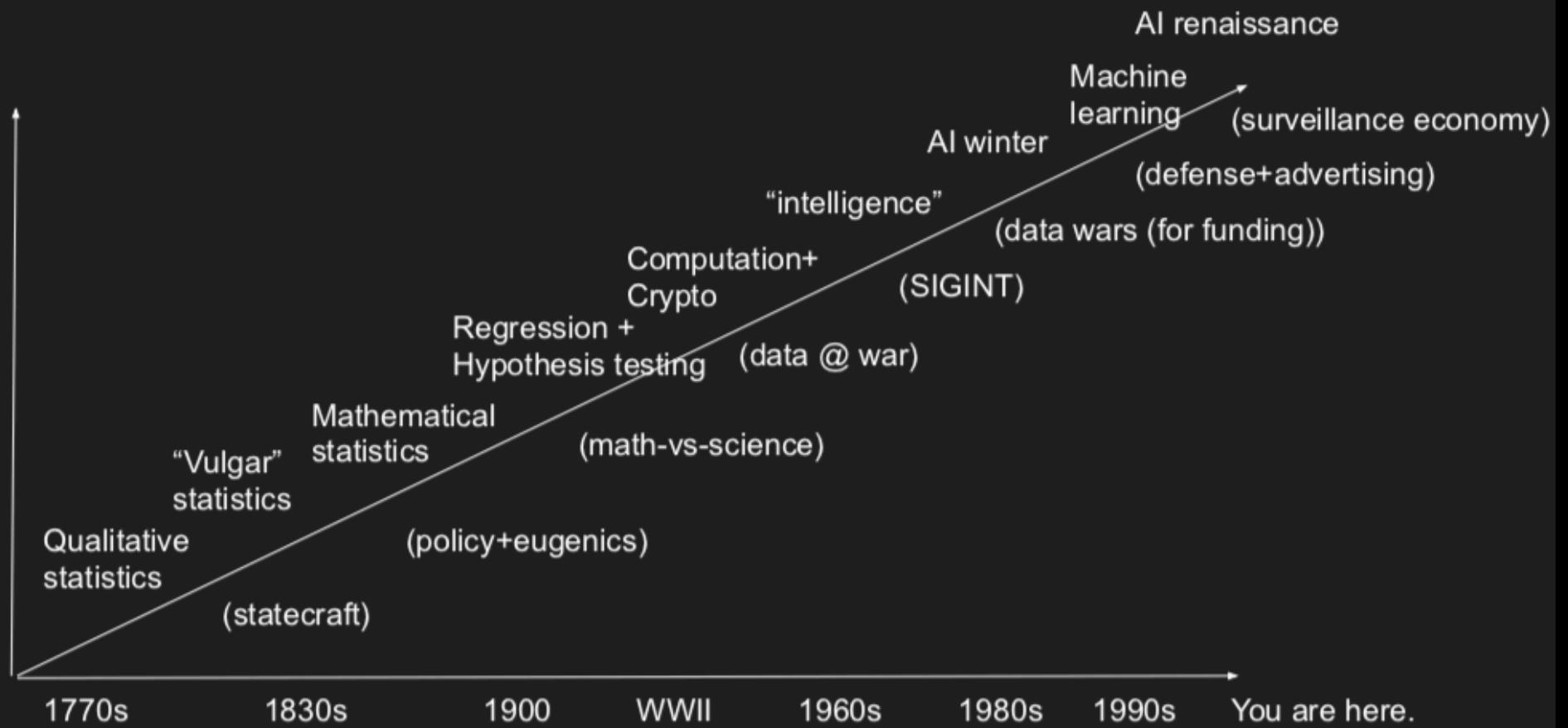
what should future
statisticians CEOs, and senators
know about the history and ethics of data?

- 0. preamble: class origin story
- 1. why history?
- 2. why ethics?
- 3. what we taught
- 4. what we learned

data 1770s-present: capabilities & intents

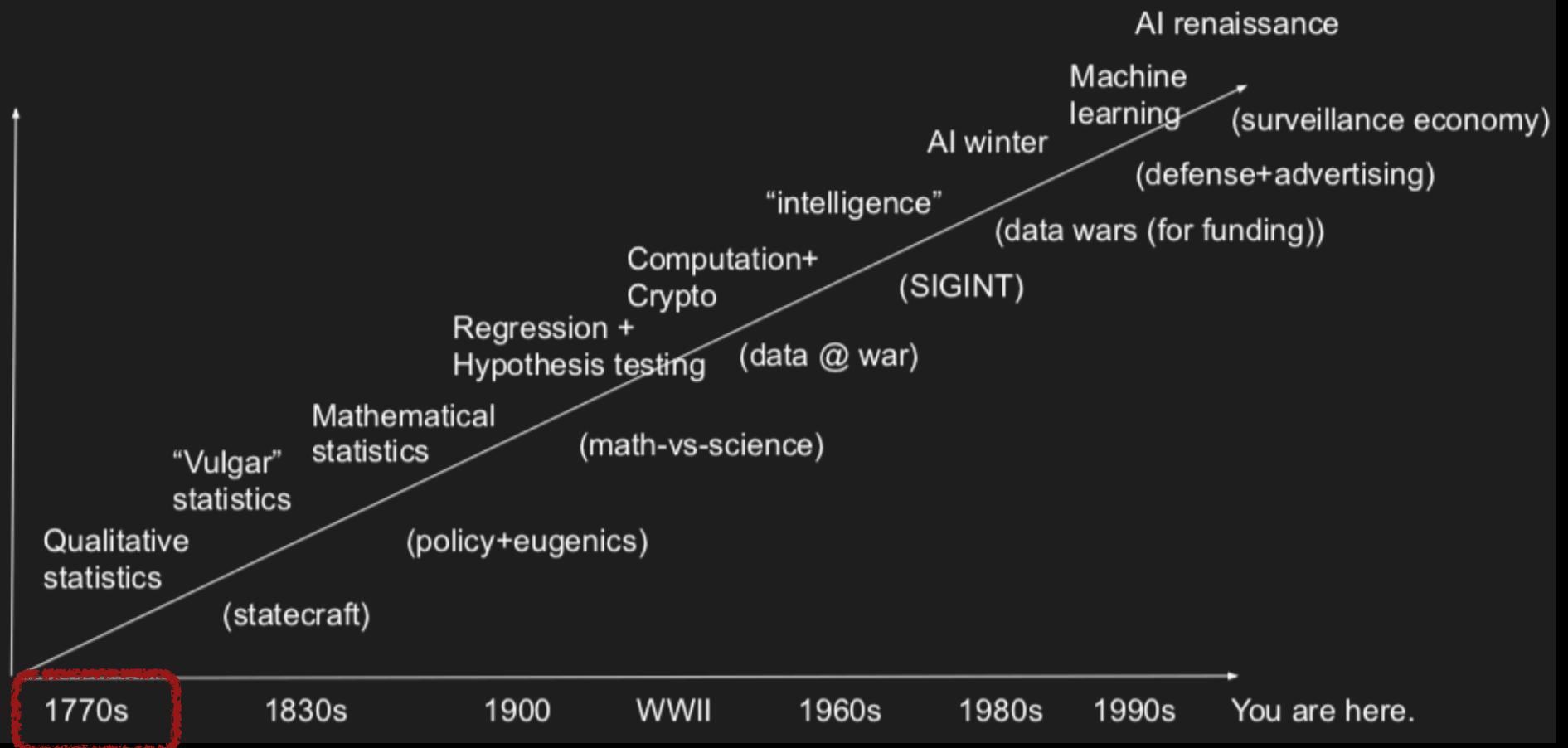


data 1770s-present: capabilities & intents



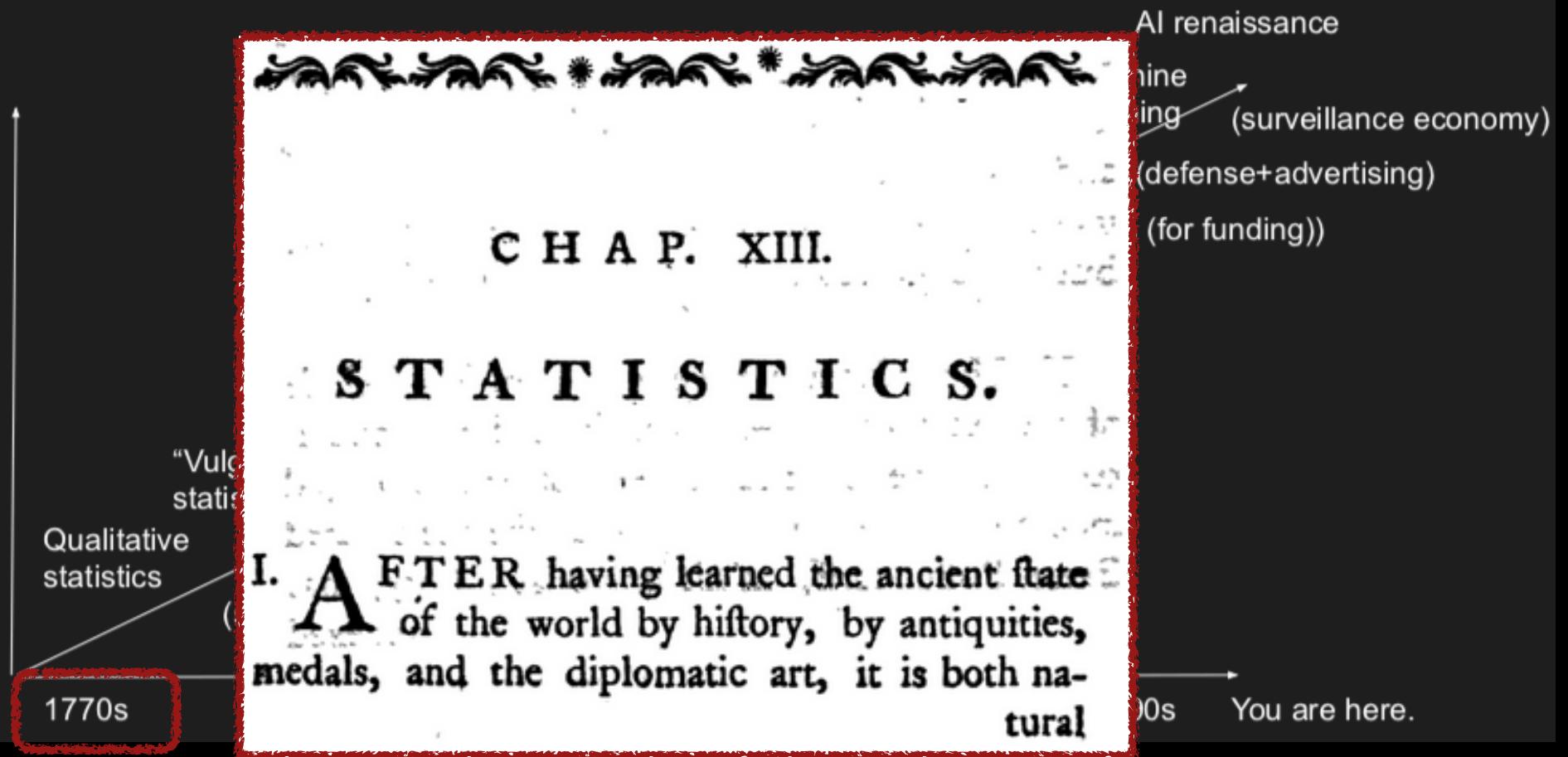
(really this is just weeks 3-11)

data 1770s-present: capabilities & intents



(really this is just weeks 3-11)

data 1770s-present: capabilities & intents



(really this is just weeks 3-11)

data 1770s-present: capabilities & intents

AI renaissance

STATISTICS.

269

tural and just, to desire to have a knowledge of the state of the present world, and of the most important occurrences of our own days; and this we learn by Statistics, by the relations of travellers, and by geography. The science, that is called Statistics, teaches us what is the political arrangement of all the modern states of the known world. This arrangement, comprehended for-

Qualitative
statistics

1770s

(really this is just weeks 3-11)

data 1770s-present: capabilities & intents

Jakob Friedrich von Bielfeld

From Wikipedia, the free encyclopedia

Jakob Friedrich von Bielfeld (31 March 1717 – 5 April 1770) was a German writer and statesman for the Kingdom of Prussia during the Age of Enlightenment.^[1] His work mainly consists of reflections on national governance, but also of comedies. His work had significant impact on both Political Science^[2] and Statistics.^[3] Bielfeld was an advisor to Frederick the Great of Prussia. His political work was translated into several languages and introduced by Joachim Georg Darjes.^[4]



Jakob Friedrich von Bielfeld
(1717-1770)

1770s

Qualitative statistics

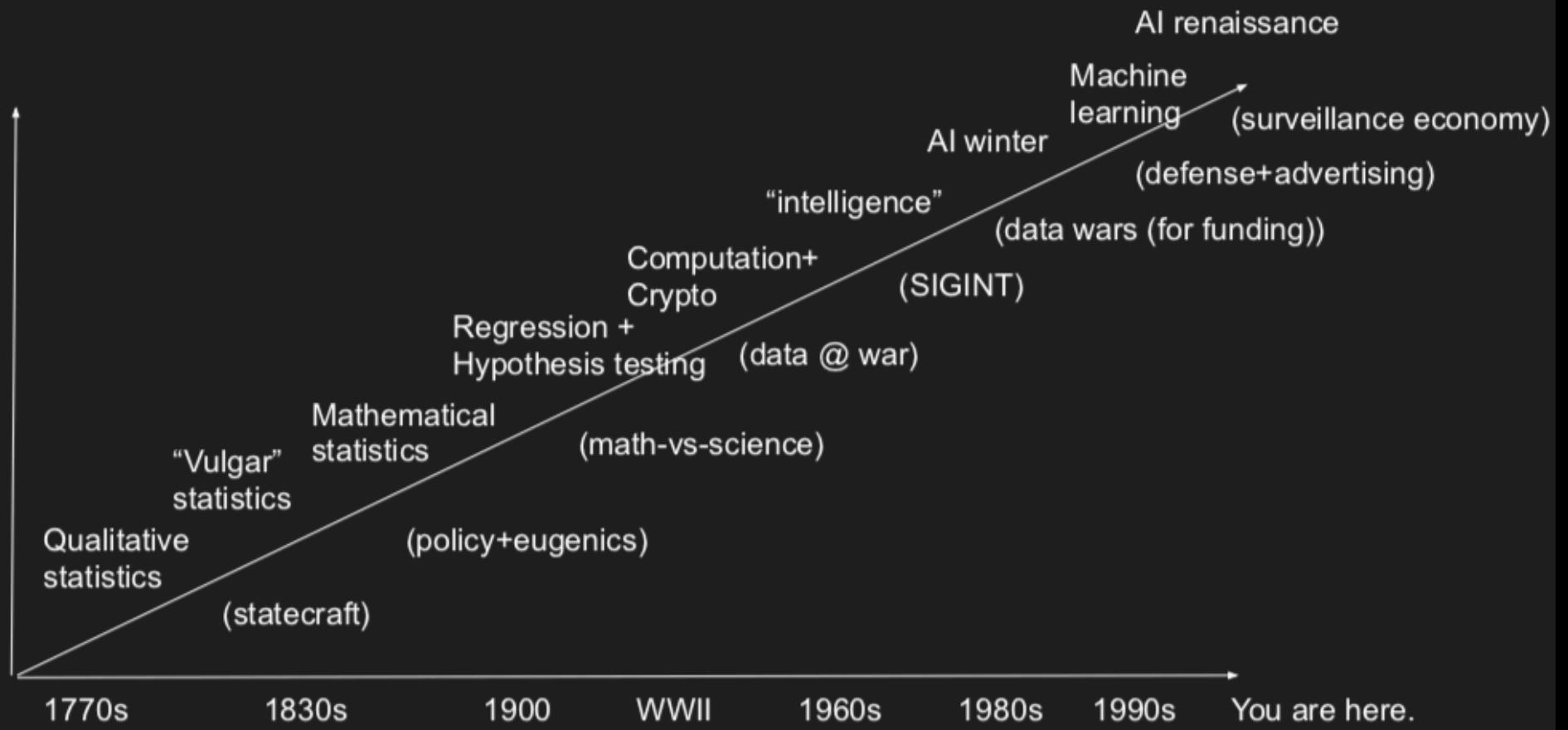
"Vulgar statistics"

AI renaissance
Machine learning (surveillance economy)
(defense+advertising)
Wars (for funding))

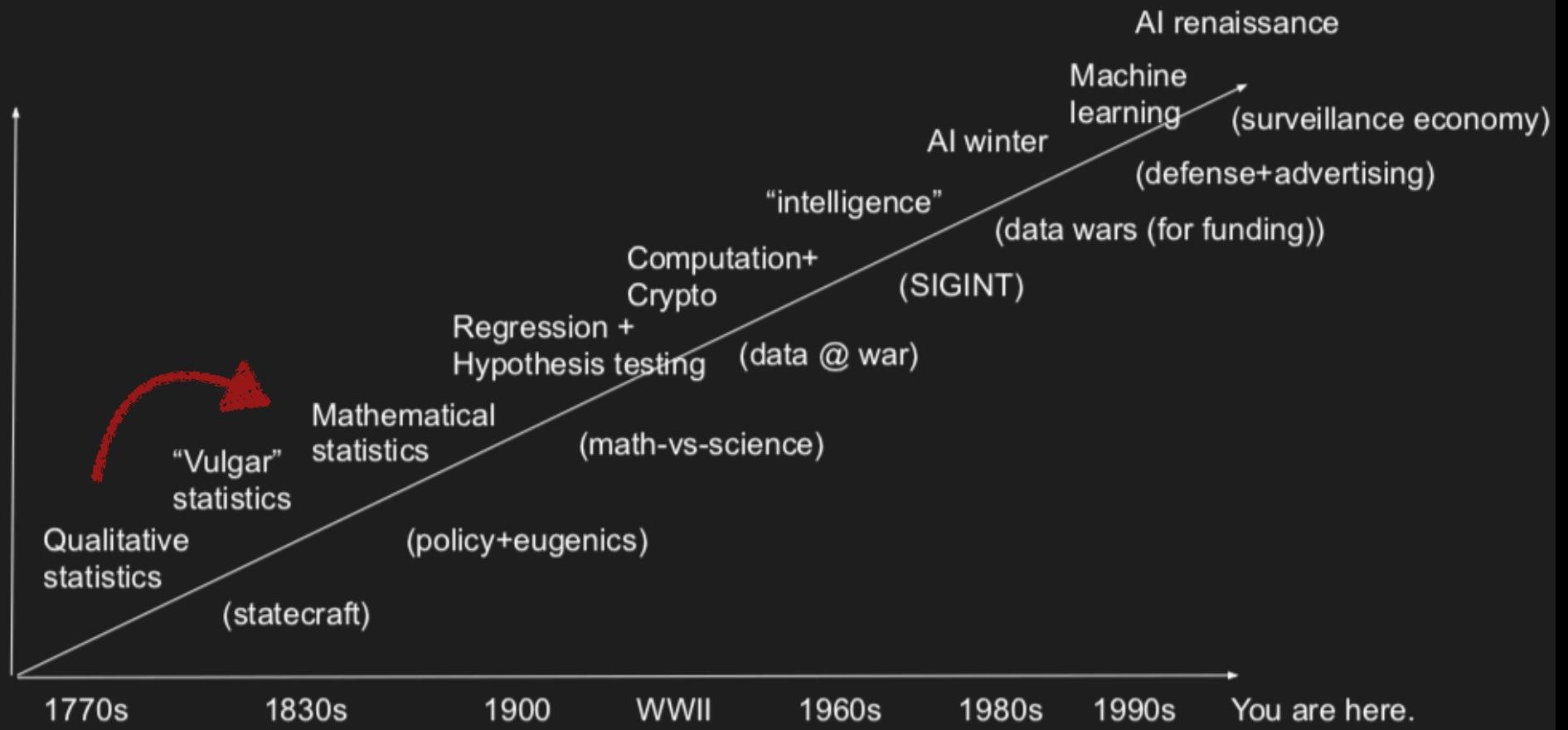
1990s You are here.

(really this is just weeks 3-11)

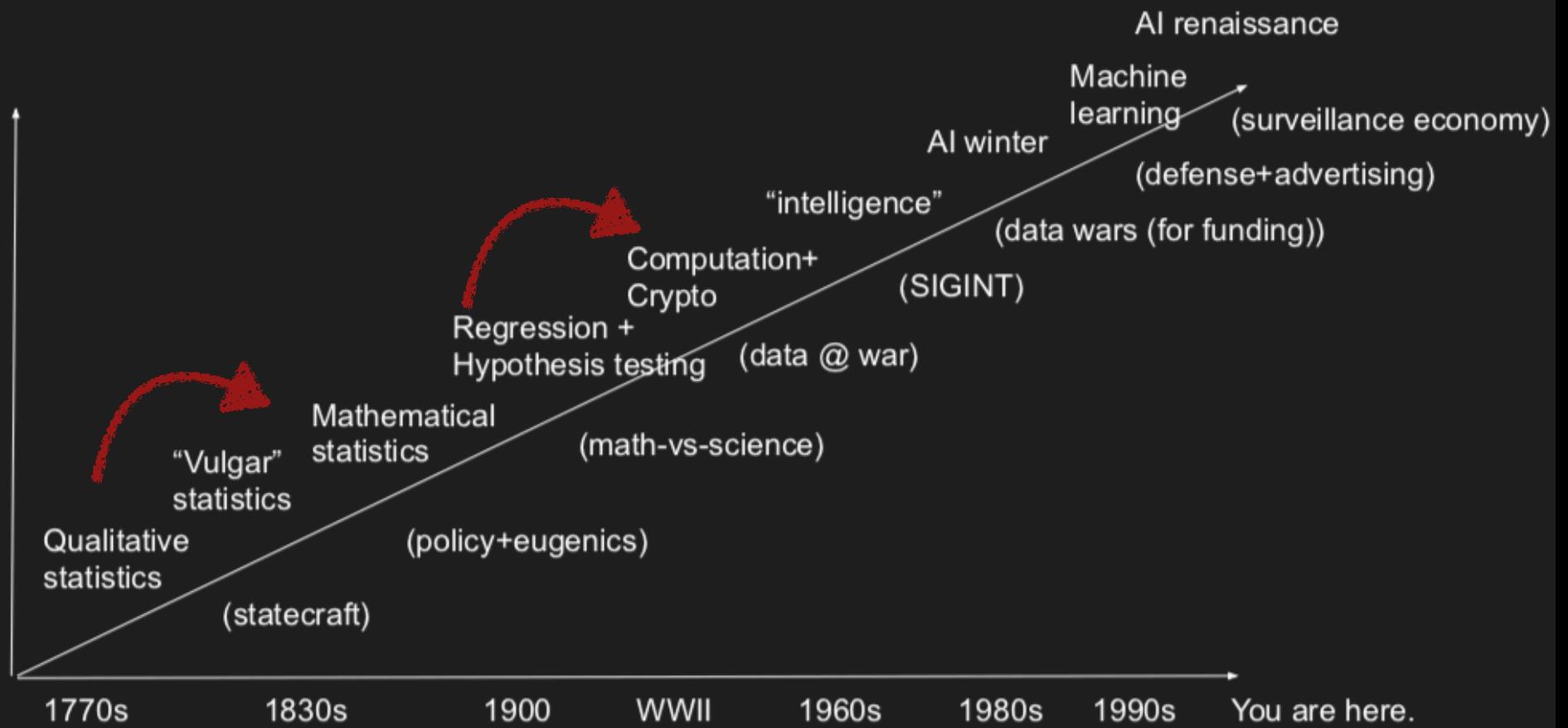
data 1770s-present: capabilities & intents



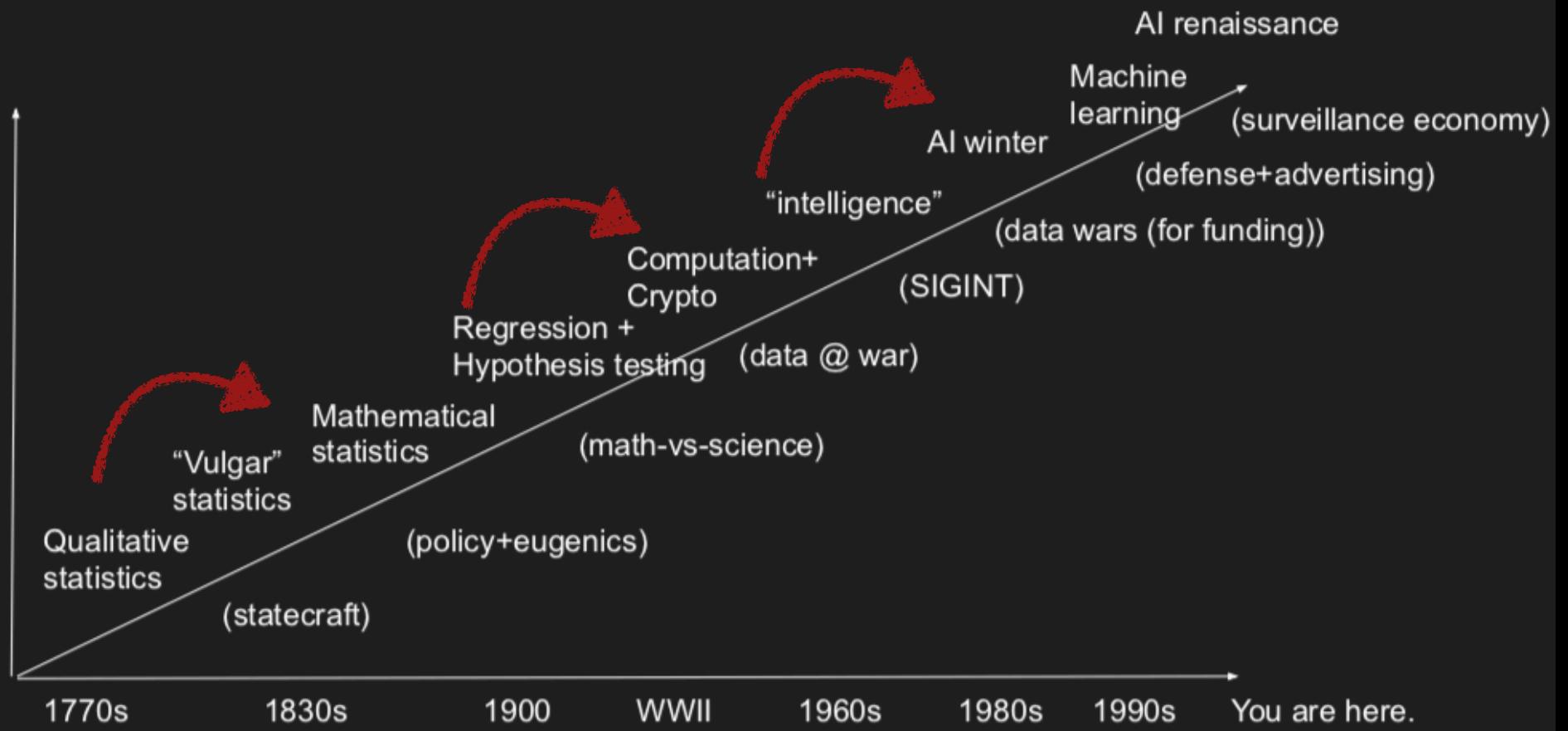
data 1770s-present: capabilities & intents



data 1770s-present: capabilities & intents



data 1770s-present: capabilities & intents



data 1770s-present: capabilities & intents

AI renaissance

close each week with:

Qua
stat

1770s

1830s

1900

WWII

1960s

1980s

1990s

You are here.

data 1770s-present: capabilities & intents

AI renaissance

close each week with:

- how did new capabilities rearrange power?
(who can now do what, from what, to whom?)

Qua
stat

1770s

1830s

1900

WWII

1960s

1980s

1990s

You are here.

data 1770s-present: capabilities & intents

AI renaissance

close each week with:

- how did new capabilities rearrange power?
(who can now do what, from what, to whom?)
- role of

Qua
stat

1770s

1830s

1900

WWII

1960s

1980s

1990s

You are here.

data 1770s-present: capabilities & intents

AI renaissance

close each week with:

- how did new capabilities rearrange power?
(who can now do what, from what, to whom?)
- role of
 - 1. rights

Qua
stat

1770s

1830s

1900

WWII

1960s

1980s

1990s

You are here.

data 1770s-present: capabilities & intents

AI renaissance

close each week with:

- how did new capabilities rearrange power?
(who can now do what, from what, to whom?)
- role of
 1. rights
 2. harms

Qua
stat

1770s

1830s

1900

WWII

1960s

1980s

1990s

You are here.

data 1770s-present: capabilities & intents

AI renaissance

close each week with:

- how did new capabilities rearrange power?
(who can now do what, from what, to whom?)
- role of
 1. rights
 2. harms
 3. justice

Quo
stat

1770s

1830s

1900

WWII

1960s

1980s

1990s

You are here.

data 1770s-present: capabilities & intents

AI renaissance

close each week with:

- how did new capabilities rearrange power?
(who can now do what, from what, to whom?)
- role of
 1. rights
 2. harms
 3. justice

Quo
stat

1770s

1830s

1900

WWII

1960s

1980s

1990s

You are here.

(week 1 & 2 had plenty of harms+injustice)

3. what we taught: 14 weeks: Tuesday discussion

github.com/data-ppf/data-ppf.github.io/wiki/Syllabus

3. what we taught: 14 weeks: Tuesday discussion

1 intro

2 setting the stakes

3 risk and social physics

4 statecraft and quantitative racism

5 intelligence, causality, and policy

6 data gets real: mathematical baptism

7 WWII, dawn of digital computation

8 birth and death of AI

9 big data, old school (1958-1980)

10 data science, 1962-2017

11 AI2.0

12 ethics

13 present problems & VC-backed attention economy

14 future solutions

github.com/data-ppf/data-ppf.github.io/wiki/Syllabus

3. what we taught: 14 weeks: Thursday Labs

github.com/data-ppf/data-ppf.github.io/wiki/Syllabus

3. what we taught: 14 weeks: Thursday Labs

1. first steps in Python interrogating the UCI dataset
2. EDA with the UCI dataset
3. Quetelet and GPAs
4. Galton
5. statistics and society; Yule, Spearman, Simpson
6. p-hacking; Fisher
7. the first data science
8. AI 1.0; Expert systems; Perceptron
9. databases and recsys; the Netflix Prize story
10. trees along with in-lab lecture on trees
11. interactive: 3 ML's; FAT 1.0 disparate impact, disparate treatment, and COMPAS
12. normative+technical approaches to defining and defending privacy; our own database of ruin: constructing and de-identifying; FAT 2.0 featuring ToS/EULAs
13. problems along with in-lab lecture on NSA history
14. solutions

github.com/data-ppf/data-ppf.github.io/wiki/Syllabus

2. why ethics:

3. what we taught: 14 weeks: Tuesdays

4. what we learned

1 intro

2 setting the stakes

3 risk and social physics

4 statecraft and quantitative racism

5 intelligence, causality, and policy

6 data gets real: mathematical baptism

7 WWII, dawn of digital computation

8 birth and death of AI

9 big data, old school (1958-1980)

10 data science, 1962-2017

11 AI2.0

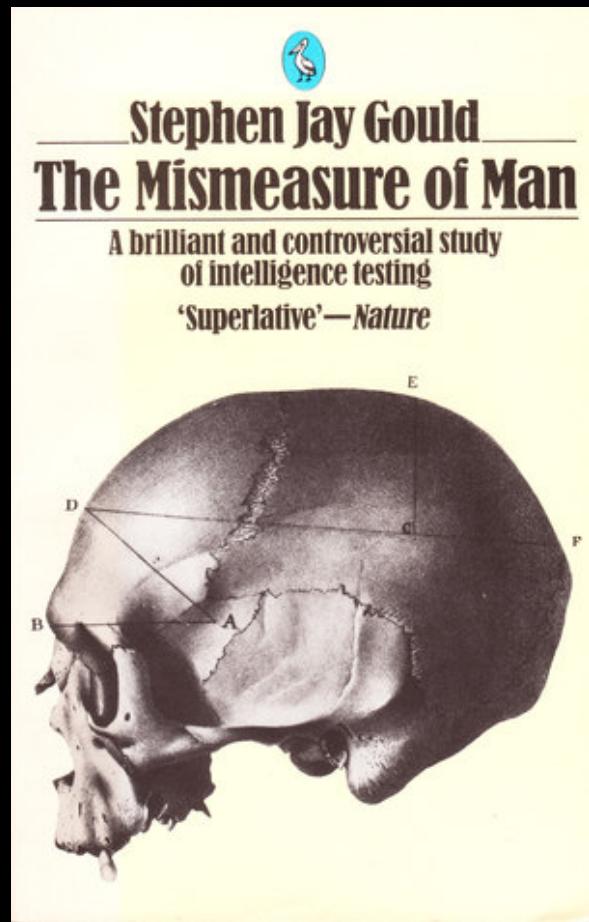
12 ethics

13 present problems & VC-backed attention economy

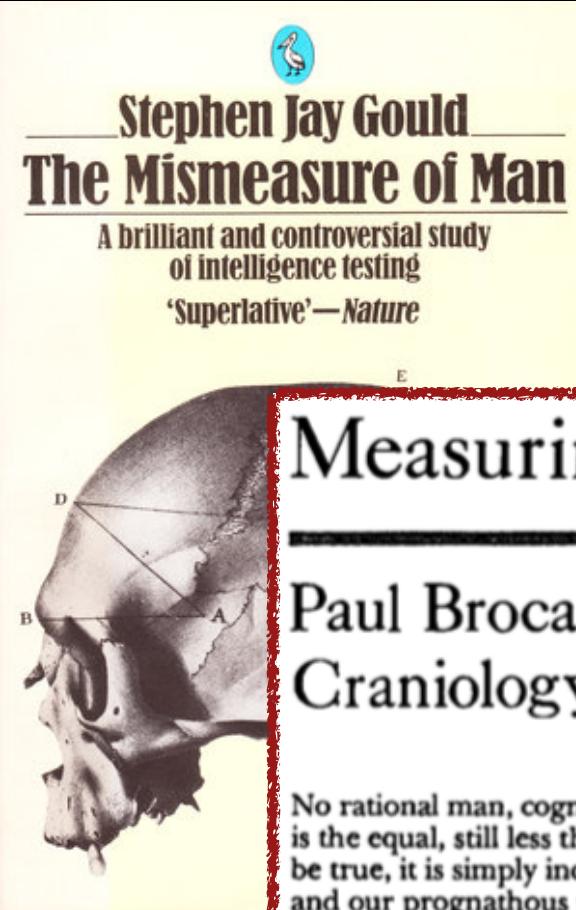
14 future solutions

e.g., week 4 regression & quantitative racism

e.g., week 4 regression & quantitative racism



e.g., week 4 regression & quantitative racism



Stephen Jay Gould
The Mismeasure of Man
A brilliant and controversial study
of intelligence testing
'Superlative'—*Nature*

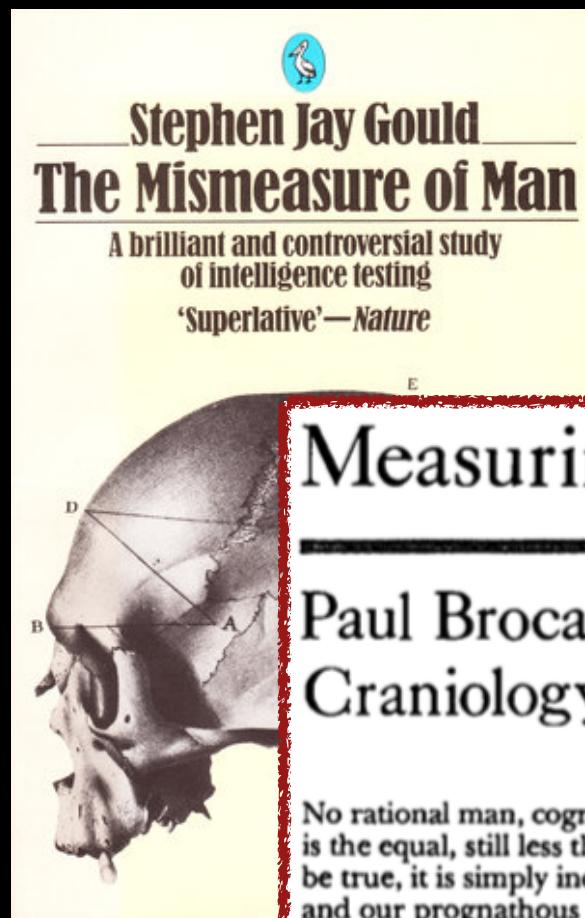
E

Measuring Heads .

Paul Broca and the Heyday of Craniology

No rational man, cognisant of the facts, believes that the average negro is the equal, still less the superior, of the average white man. And, if this be true, it is simply incredible that, when all his disabilities are removed, and our prognathous relative has a fair field and no favor, as well as no oppressor, he will be able to compete successfully with his bigger-brained and smaller-jawed rival, in a contest which is to be carried on by thoughts and not by bites. —T. H. HUXLEY

e.g., week 4 regression & quantitative racism



Measuring Heads .

Paul Broca and the Heyday of
Craniology

No rational man, cognitively equal, will ever be less than he should be true, it is simply incorrect. And our prognathousness, the oppressor, he will be a smaller and smaller-jawed rival in thoughts and not by b

SIR FRANCIS GALTON AND THE BIRTH OF EUGENICS

Nicholas W. Gillham

DCMB Group, Department of Biology, Box 91000, Duke University, Durham,
North Carolina 27708-1000; e-mail: gillham@duke.edu

e.g., week 4 regression & quantitative racism

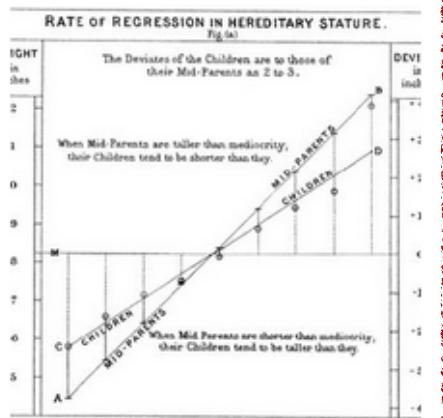
e.g., week 4 regression & quantitative racism

Data: Past, Present, Future | Lab 4 | 2/14/2019

describing and predicting: Galton, regression, inventing error, sum

Galton and regression

Galton's analysis "gives the numerical value of the regression towards mediocrity in the case of human stature, and precision [see Plate IX, fig. (a)]"



e.g., week 4 regression & quantitative racism

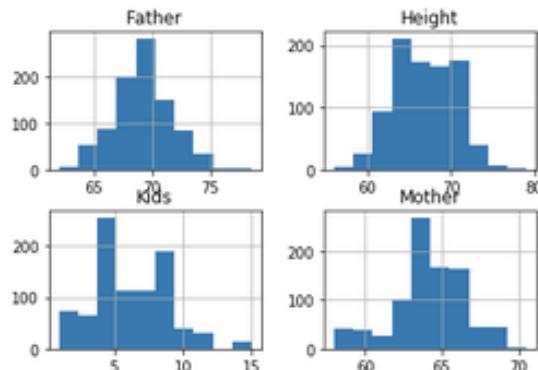
Data: Past, Present, Future | Lab 4 | 2/14/2019

describing and predicting: Galton, regression, inventing error, sum

Galton and regression:

Galton's analysis "gives us a sense of the power of statistics to describe and predict the precision of inheritance."

```
In [7]: heights.hist()  
Out[7]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>, <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>, <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>, <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>], dtype=object)
```



e.g., week 4 regression & quantitative racism

Data: Past, Present, Future | Lab 4 | 2/14/2019

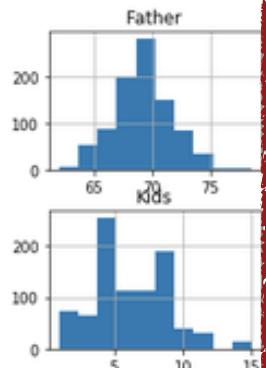
describing and predicting: Galton, regression, inventing error, sum

Galton and regression:

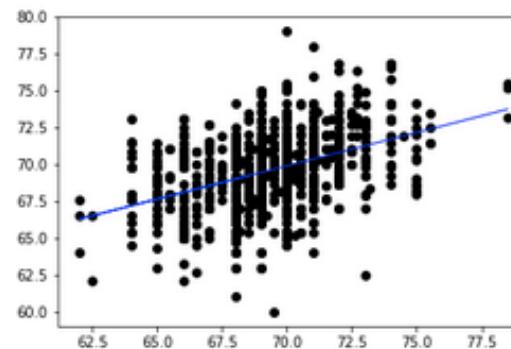
Galton's analysis "gives us a sense of the precision with which height can be predicted from the heights of parents and precision [see Plateau's Law]"

```
In [7]: heights.hist()
```

```
Out[7]: array([[[<matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>, <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>, <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>, <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>], dtype=object)
```



```
In [28]: # plot fit line  
plt.scatter(x, y, color='black')  
plt.plot(x, skl_lm.predict(x), color='blue', linewidth=1)  
plt.show()
```



Now it's your turn!

e.g., week 4 regression & quantitative racism

Data: Past, Present, Future | Lab 4 | 2/14/2019

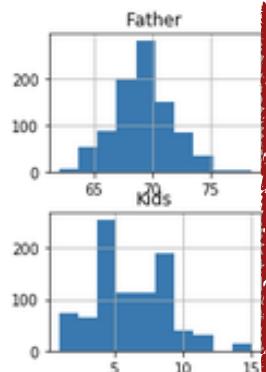
describing and predicting: Galton, regression, inventing error, sum

Galton and regression

Galton's analysis "gives
and precision [see Plate

```
In [7]: heights.hist()
```

```
Out[7]: array([[[<matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>,
   <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>,
   [<matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>,
   <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>,
   <matplotlib.axes._subplots.AxesSubplot object at 0x0000000000000000>],
  dtype=object)]
```



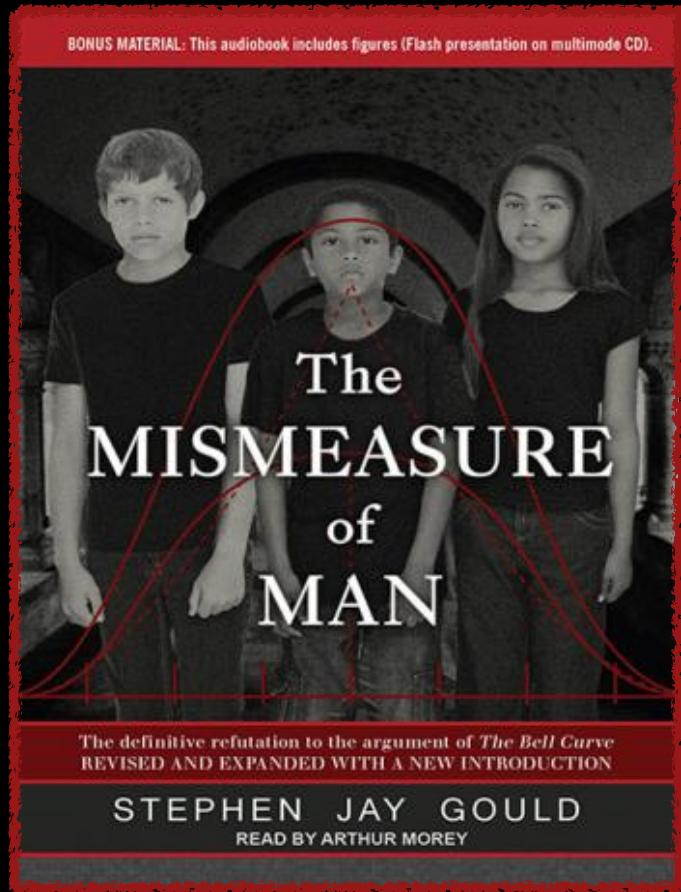
```
In [28]: # plot fit line
    plt.scatter(x, y)
    plt.plot(x, skl_
    plt.show()
```

Now it's your turn!

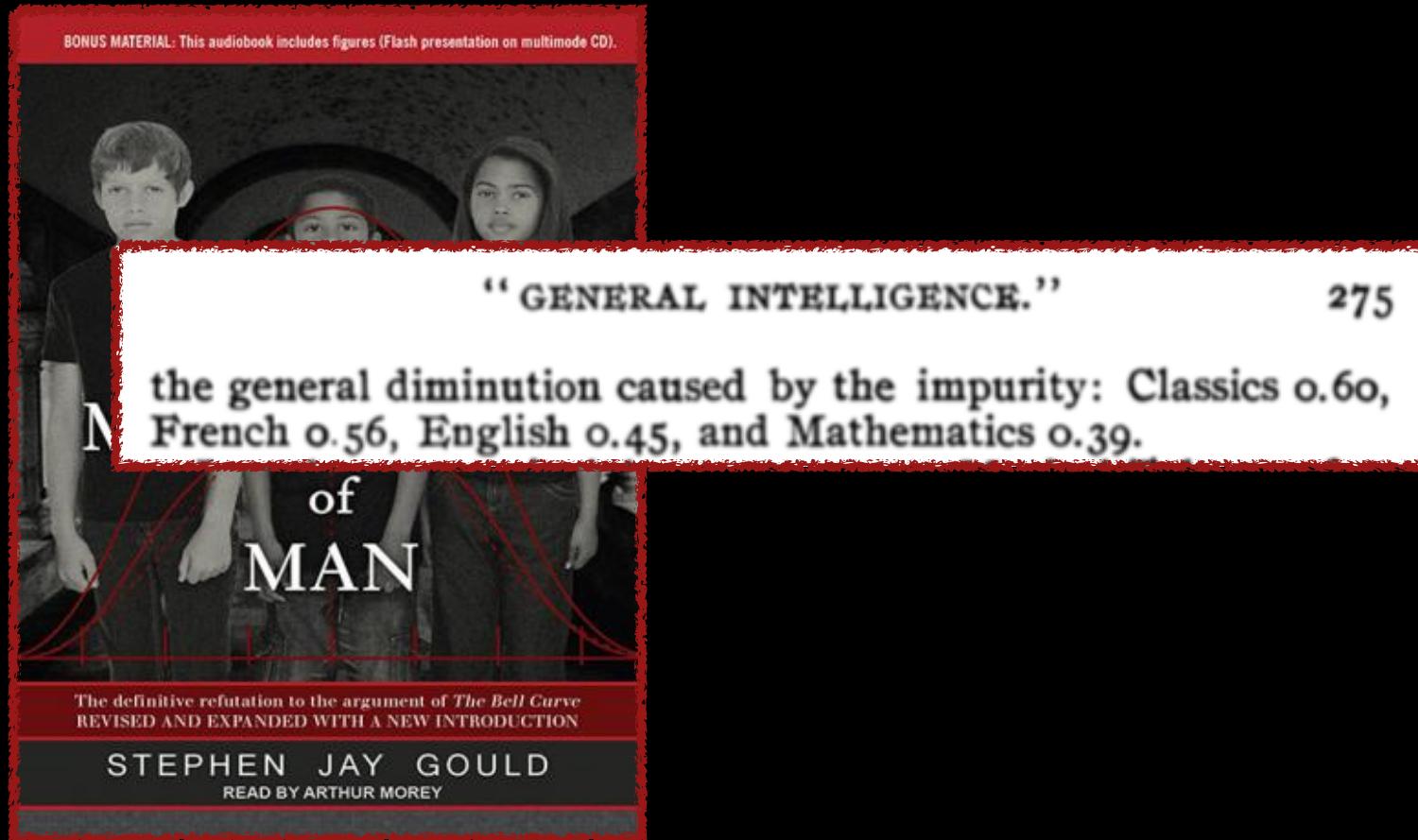
can you the regression for

1. everybody and his/her mother?
 2. males and fathers

e.g., week 5 IQ, policy and causality



e.g., week 5 IQ, policy and causality



e.g., week 5 IQ, policy and causality

BONUS MATERIAL: This audiobook includes figures (Flash presentation on multimode CD).

" GENERAL INTELLIGENCE." 275

the general diminution caused by the impurity: Classics 0.60,
French 0.56, English 0.45, and Mathematics 0.39.

M
of
MAN

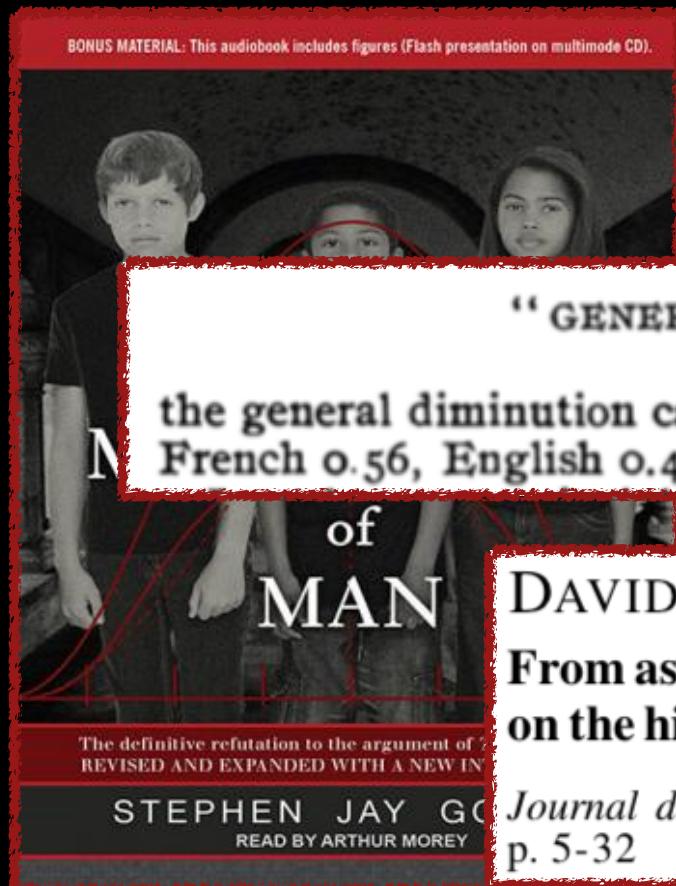
The definitive refutation to the argument of .
REVISED AND EXPANDED WITH A NEW INTRODUCTION

STEPHEN JAY GOULD
READ BY ARTHUR MOREY

DAVID FREEDMAN
**From association to causation : some remarks
on the history of statistics**

Journal de la société française de statistique, tome 140, n° 3 (1999),
p. 5-32

e.g., week 5 IQ, policy and causality



DAVID FREEMAN
**From association to causation : some remarks
on the history of statistics**

Journal de la société française de statistique, tome 140, n° 3 (1999),
p. 5-32

SOME REMARKS ON THE HISTORY OF STATISTICS

that welfare outside the poor-house creates paupers – the estimated coefficient
on the out-relief ratio is positive.

e.g., week 5 IQ, policy and causality

e.g., week 5 IQ, policy and causality

Lab05a-20190221-Yule (autosaved)

View Insert Cell Kernel Widgets Help Trusted

Yule's original paper: Yule, G. Udny. 1899. ["An Investigation into the Causes of Changes in Pauperism in England, Chiefly during the Last Two Intervening Decades."](#) *Journal of the Royal Statistical Society* 62 (Part II):249-295.

He doesn't give all his data, but gives one important example.

286 YULE—*Causes of Changes in Pauperism in England.* [June]

TABLE XIX.—1871-81, *Metropolitan Group.*

Union.	Percentage Ratios of			
	Pauperism.	Out-Relief Ratio.	Proportion of Old.	Population.
1. Kensington	27	5	104	136
2. Paddington	47	12	115	111
3. Fulham	31	21	85	174
4. Chelsea	64	21	81	124

e.g., week 5 IQ, policy and causality

Lab05a-20190221-Yule (autosaved)

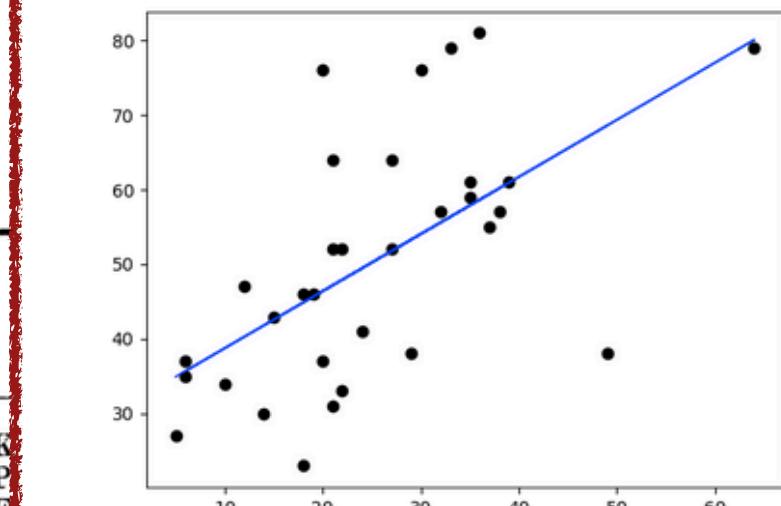
View Insert Cell Kernel Widgets Help Trusted

```
#matplotlib notebook
plt.scatter(X, y, color='black')
plt.plot(X, regression_model.predict(X), color='blue', linewidth=1)
plt.show()
```

Yule's original paper
[Decades of Journals](#)

He doesn't give us a title

286



1. K
2. P
3. F
4. C

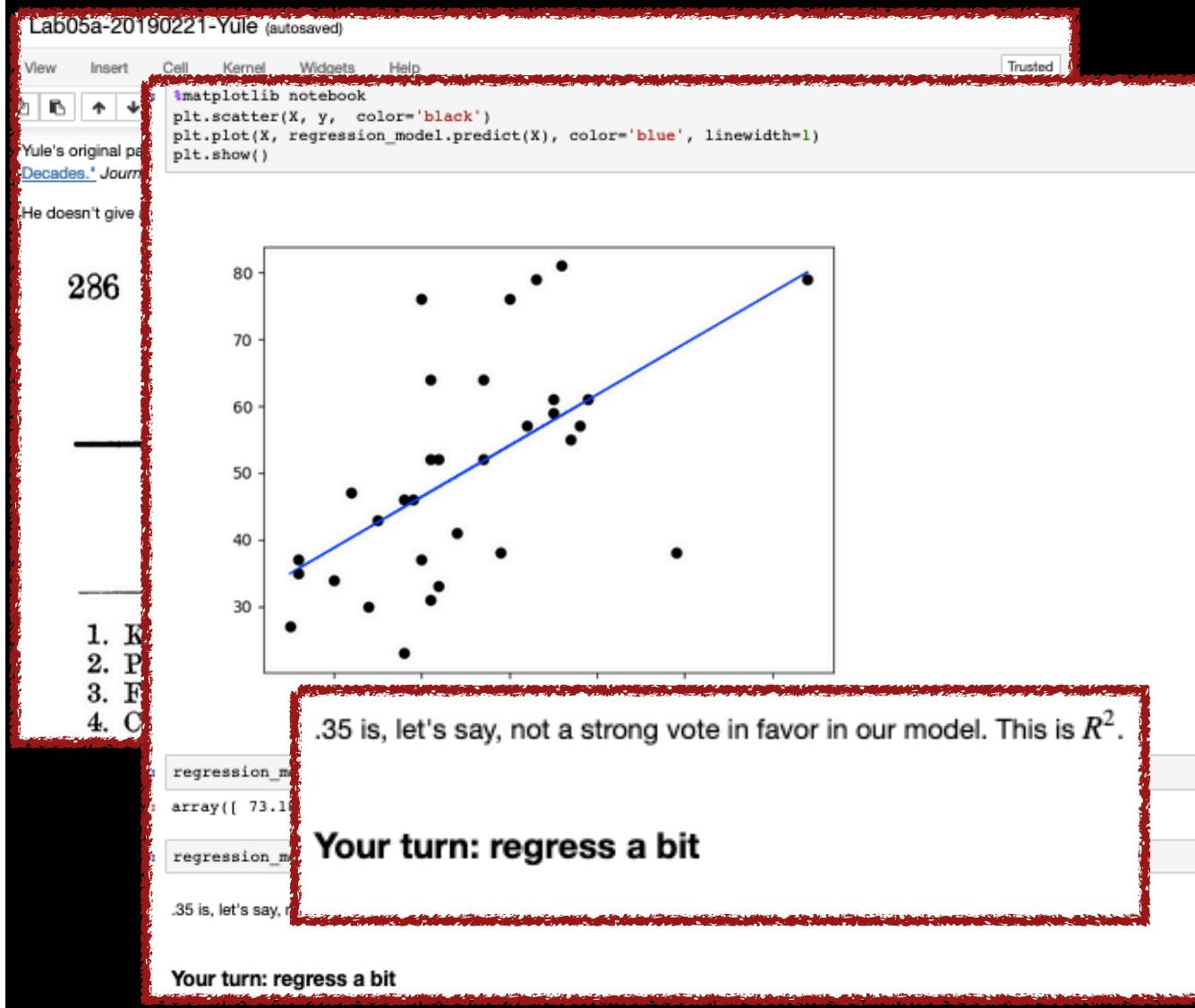
```
regression_model.predict(55)
array([ 73.1858236])

regression_model.score(X,y)
```

.35 is, let's say, not a strong vote in favor of our model. This is R^2 .

Your turn: regress a bit

e.g., week 5 IQ, policy and causality



e.g., week 5 IQ, policy and causality

Lab05a-20190221-Yule (autosaved)

View Insert Cell Kernel Widgets Help Trusted

```
#matplotlib notebook
plt.scatter(X, y, color='black')
plt.plot(X, regression_model.predict(X), color='blue', linewidth=1)
plt.show()
```

Yule's original paper
[Decades... Journal](#)

He doesn't give

286

1. K
2. P
3. F
4. C

.35 is, let's say, not

regression_m
array([73.1])
regression_m
.35 is, let's say,

Your turn: regress a bit

Lab05c-20190221-Simpsons (unsaved changes)

View Insert Cell Kernel Widgets Help

Admissions Rate and Applicants by Department and Gender

The scatter plot displays the relationship between the fraction of applicants (X-axis) and the admissions rate (Y-axis). The X-axis ranges from 0.0% to 20.0% with increments of 5.0%. The Y-axis ranges from 0.20% to 0.80% with increments of 0.10%. Data points are categorized by gender: Female (red circles) and Male (blue circles). There is a general downward trend, indicating that as the fraction of applicants increases, the admissions rate tends to decrease.

Fraction Applicants (%)	Admissions Rate (%)	Gender
2.0	0.68	Female
2.5	0.35	Male
3.0	0.28	Male
3.5	0.35	Female
4.0	0.32	Female
4.5	0.32	Male
5.0	0.32	Male
5.5	0.22	Female
6.0	0.62	Male
7.0	0.65	Male
8.0	0.35	Female
10.0	0.62	Male
20.0	0.30	Female

Admissions Rate

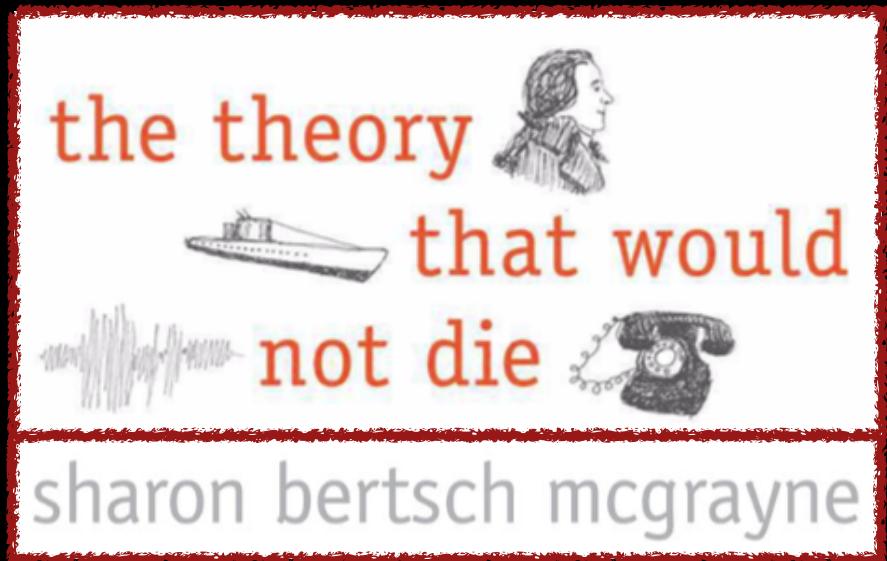
Fraction Applicants

Female

Male

for more, see "Paradoxes" chapter in Pearl 2018

e.g., week 7 “women at the dawn...” (Abbate)



e.g., week 7 “women at the dawn...” (Abbate)

The image is a collage with a red border. In the top right corner is a small black and white portrait of Alan Turing. To the left of the portrait is a slide with the title "the theory of computation" in red, a drawing of a ship on water, and the word "no" in red. Below these is a larger slide with the text "sharon bell" in grey. The main central image is a black and white photograph of the Bombe machine, a large, rectangular, metal device with many circular components, likely rotors or wheels, arranged in rows. It is mounted on a stand with wheels.

Not a computer. The Bombe, designed by Alan Turing and Gordon Welchman, found

e.g., week 7 “women at the dawn...” (Abbate)

the theoru

no

sharon be

Ada Lovelace

Breaking Codes and Finding Trajectories:
Women at the Dawn of the Digital Age

Not a computer. The Bombe, designed by Alan Turing and Gordon Welchman, found

e.g., week 7 “women at the dawn...” (Abbate)

Breaking Codes and Finding Trajectories: Women at the Dawn of the Digital Age

I was a Colossus operator, which we considered to be the crème de la crème. We felt we were “at the sharp end,” where there was a great tension and flow of adrenaline . . . operating those incredible machines.

—Jean Beech, Colossus operator¹

I don’t know if you can picture how exciting the ENIAC was to all of us. And we didn’t talk socially or any other time about anything else. It was—we discussed it almost all the time.

—Jean Jennings, ENIAC programmer²

e.g., week 7 “women at the dawn...” (Abbate)

e.g., week 7 “women at the dawn...” (Abbate)

Lab 7: Let's be Bayesian Cryptologists

It's 1959. National Security Agency. The No-Such-Agency. Fighting communism by crackin' c

There are two general mathematical methods of measurements that are used in almost every field of science. One is the [chi-squared] test, which is well known and has been used for many years. The other is Bayes' Theorem, which is less well known but has been used for many years in the cryptanalysis of our agency. The reference is to Bayes' Theorem, which in turn gives us a way to measure the probability of a hypothesis given the evidence. Mr. Bayes was, I do not know, as his name does not appear in the usual list of famous mathematicians. He was a British statistician who lived from 1702 to 1761. His work was very useful, but in fact leads to the only correct formulas for solving a large number of our problems. In fact, it is now known that most of the work done by mathematicians at N.S.A. know about all the ways that Bayes factors can be employed.

<https://www.nsa.gov/Portals/70/documents/news-features/declassified-documents/tech-journal/2012/07/03/lab-7-lets-be-bayesian-cryptologists>

e.g., week 7 “women at the dawn...” (Abbate)

Lab 7: Let's be Bayesian Cryptologists

It's 1959

Before we get started, figure out what a Vigenère cypher is. Use the google or the bing or the duckduckgo.

Now, encypher, with a Vigenère cypher, your lastname (e.g. "wiggins") using the key "lego." You can use a Vigenère sc

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

<https://www>

e.g., week 7 “women at the dawn...” (Abbate)

Lab 7: Let's be Bayesian Cryptologists

It's 1959

Before we get started, figure out what a Vigenère cypher is. Use the google or the bing or the duckduckgo.

Now, encypher, with a Vigenère cypher, your lastname (e.g. "wiggins") using the key "lego." You can use a Vigenère sc

A	B	C	D	E	F	G	H	I	J	K	L
A	A	B	C	D	E	F	G	H	I	J	K
B	B	C	D	E	F	G	H	I	J	K	L
C	C	D	E	F	G	H	I	J	K	L	M
D	D	E	F	G	H	I	J	K	L	M	N
E	E	F	G	H	I	J	K	L	M	N	O
F	F	G	H	I	J	K	L	M	N	O	P
G	G	H	I	J	K	L	M	N	O	P	Q
H	H	I	J	K	L	M	N	O	P	Q	R
I	I	J	K	L	M	N	O	P	Q	R	S
J	J	K	L	M	N	O	P	Q	R	S	T

```
In [1]: from itertools import starmap, cycle
def encrypt(message, key):
    # convert to uppercase.
    # strip out non-alpha characters.
    message = filter(str.isalpha, message.upper())
    # single letter encryption.
    def enc(c,k): return chr(((ord(k) + ord(c) - 2*ord('A')) % 26) + ord('A'))
    return "".join(starmap(enc, zip(message, cycle(key))))

In [2]: encrypt('wiggins', 'lego')
Out[2]: 'NSSAZX'

In [3]: encrypt("wiggins", "optimusprime")
Out[3]: 'QDFUANQ'
```

Try your own!

e.g., week 7 “women at the dawn...” (Abbate)

Lab 7: Let's be Bayesian Cryptologists

It's 1959

Before we get started, figure out what a Vigenère cypher is. Use the google or the bing or the duckduckgo.

Now, encypher, with a Vigenère cypher, your lastname (e.g. "wiggins") using the key "lego." You can use a Vigenère sc

A	B	C	D	E	F	G	H	I	J	K	L
A	A	B	C	D	E	F	G	H	I	J	K
B	B	C	D	E	F	G	H	I	J	K	L
C	C	D	E	F	G	H	I	J	K	L	M
D	D	E	F	G	H	I	J	K	L	M	N
E	E	F	G	H	I	J	K	L	M	N	O
F	F	G	H	I	J	K	L	M	N	O	P
G	G	H	I	J	K	L	M	N	O	P	Q
H	H	I	J	K	L	M	N	O	P	Q	R
I	I	J	K	L	M	N	O	P	Q	R	S
J	J	K	L	M	N	O	P	Q	R	S	T

<https://w>

```
In [1]: from itertools import starmap, cycle  
  
def encrypt(message, key):  
  
    # convert to uppercase.  
    # strip out non-alpha characters.  
    message = filter(str.isalpha, message.upper())  
  
    # single letter encryption.  
    def enc(c,k): return chr(((ord(k) + ord(c) - 2*ord('A')) % 26) + ord('A'))  
  
    return "".join(starmap(enc, zip(message, cycle(key))))
```

In [2]:

Out[2]:

In [3]:

Out[3]:

In [3]: encrypt("wiggins", "optimusprime")

Out[3]: 'QDFUANQ'

Try your own!

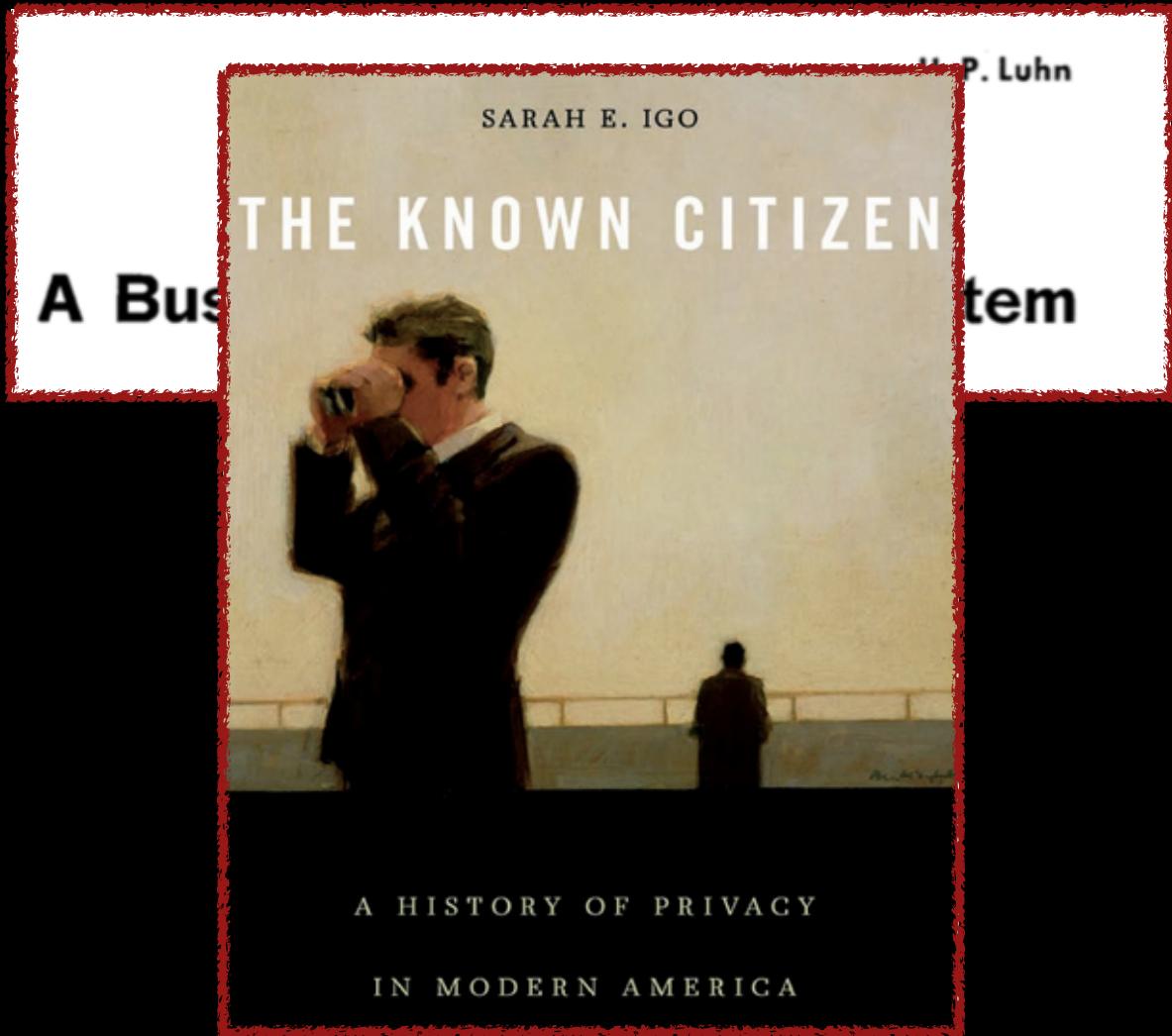
e.g., week 9 “big data” + privacy 1950-1980

e.g., week 9 “big data” + privacy 1950-1980

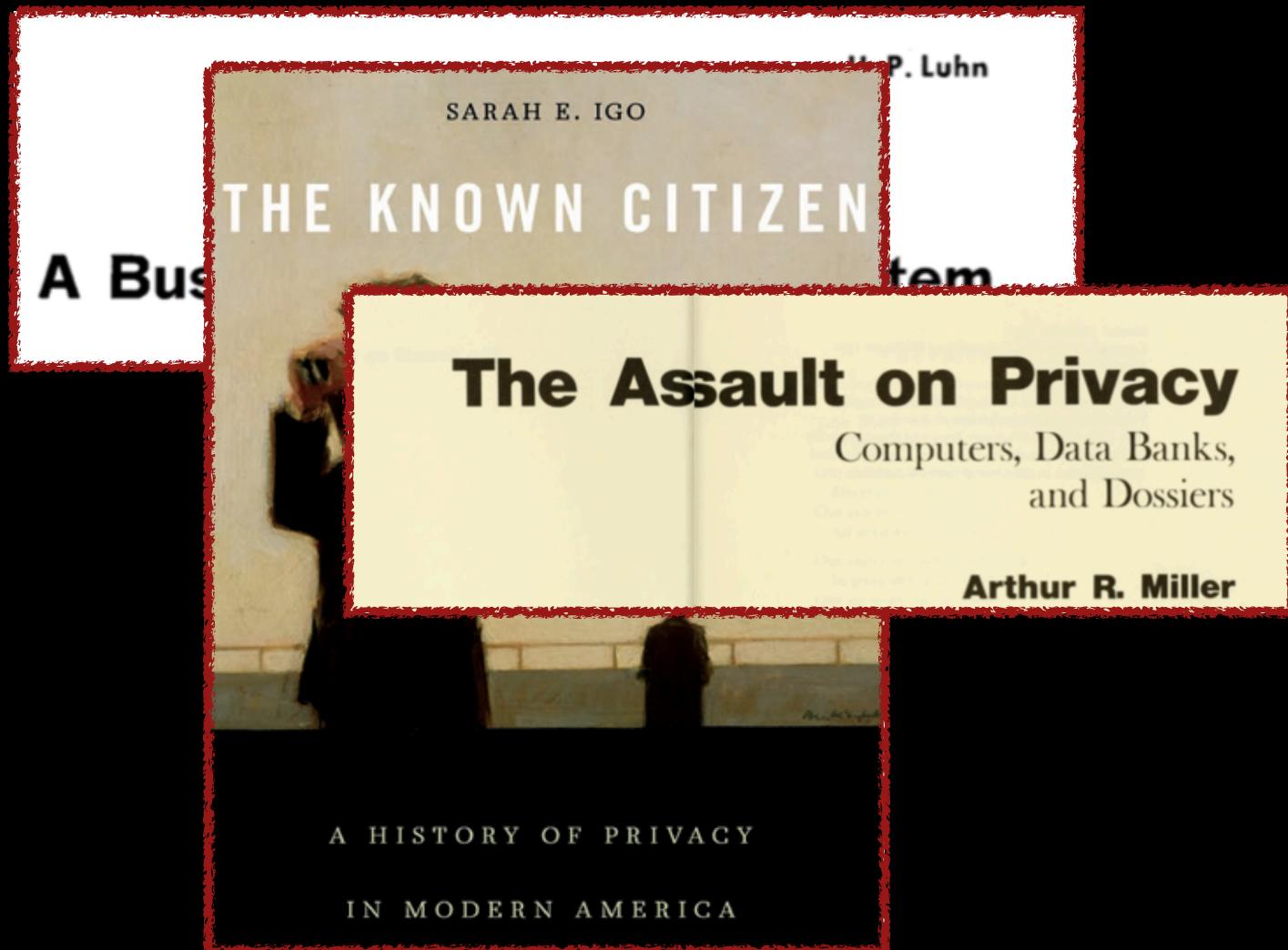
H. P. Luhn

A Business Intelligence System

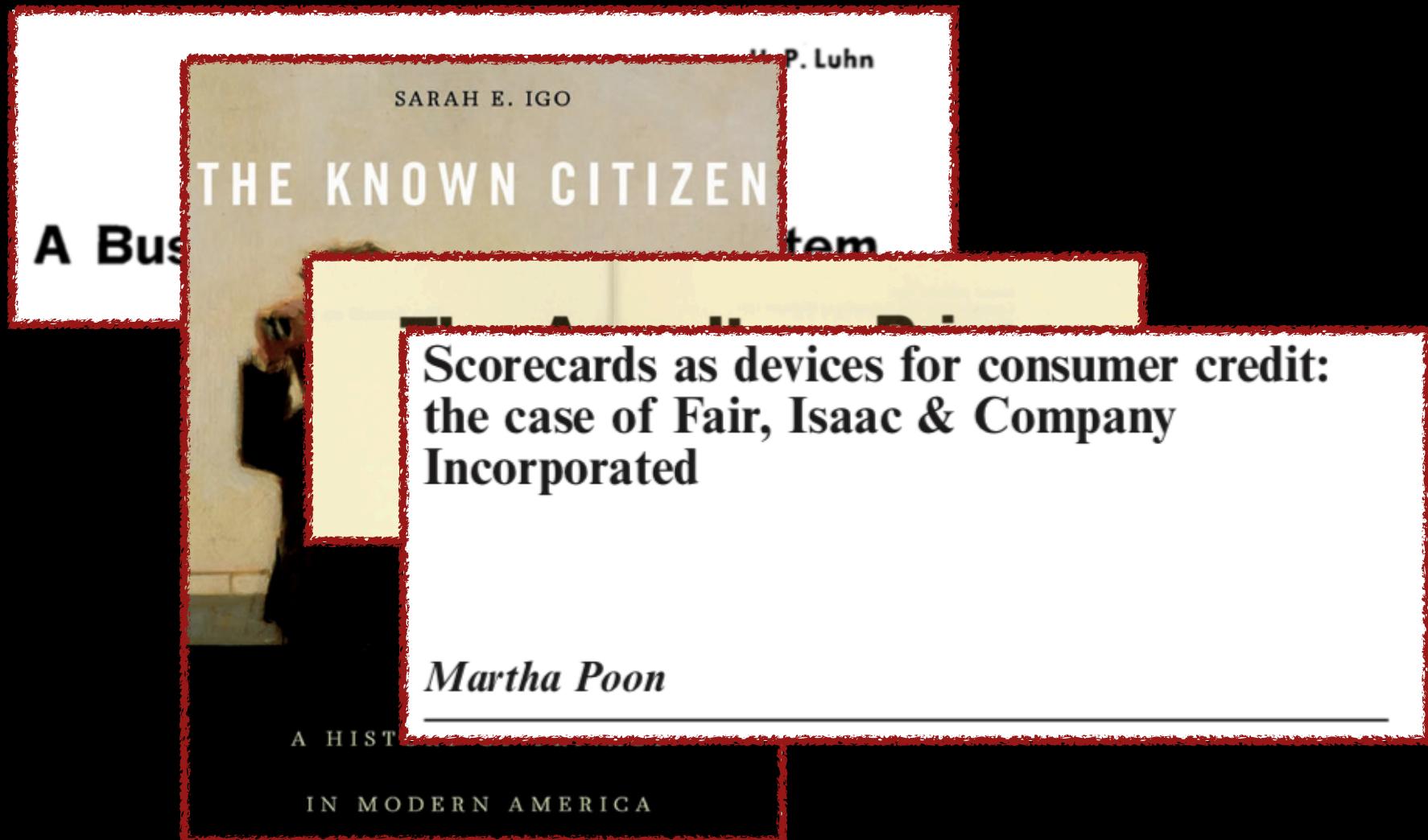
e.g., week 9 “big data” + privacy 1950-1980



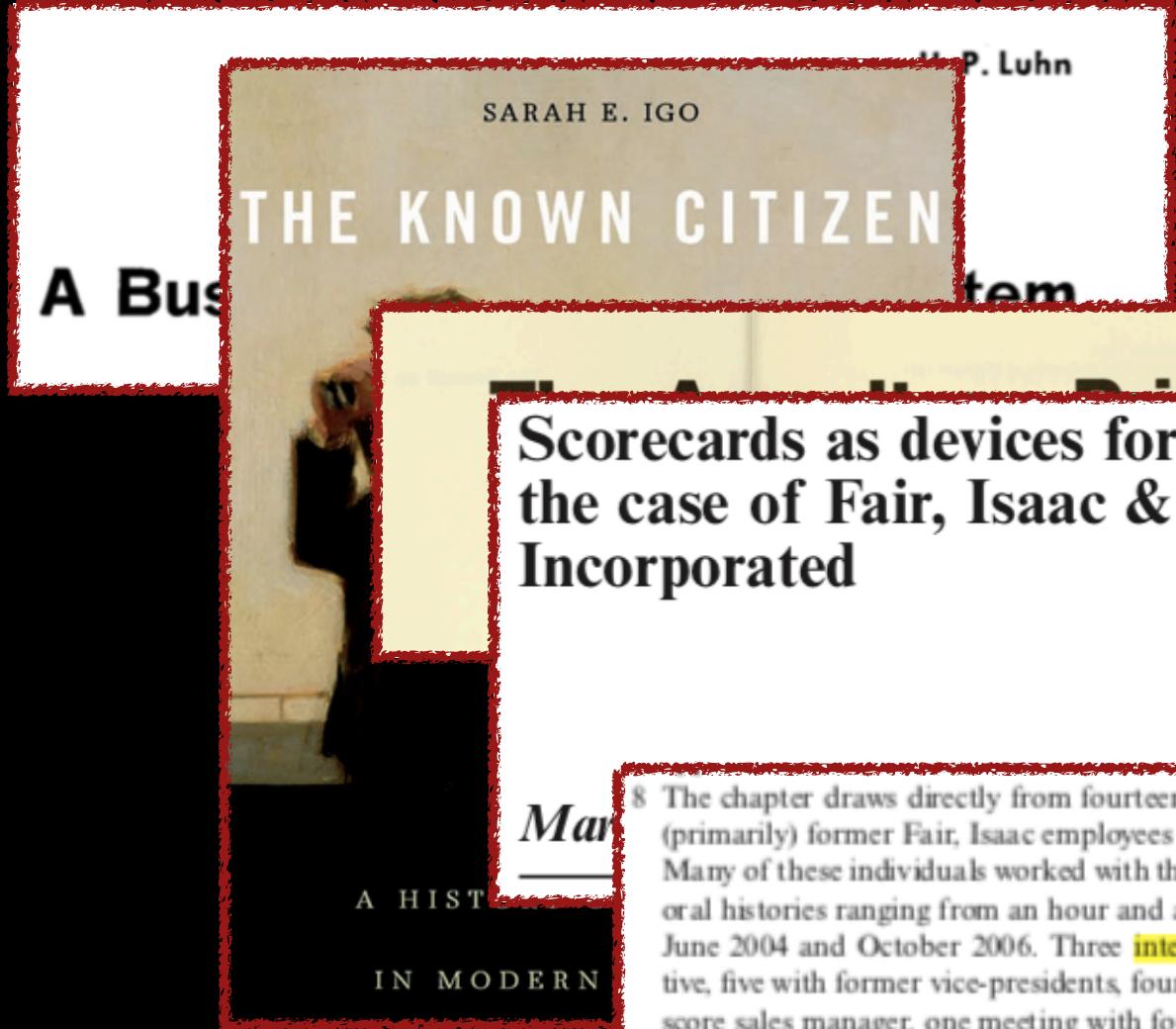
e.g., week 9 “big data” + privacy 1950-1980



e.g., week 9 “big data” + privacy 1950-1980



e.g., week 9 “big data” + privacy 1950-1980



A Bus...tem

**Scorecards as devices for consumer credit:
the case of Fair, Isaac & Company
Incorporated**

*Man...
A HIST...
IN MODERN*

^{H. P. Luhn}
SARAH E. IGO

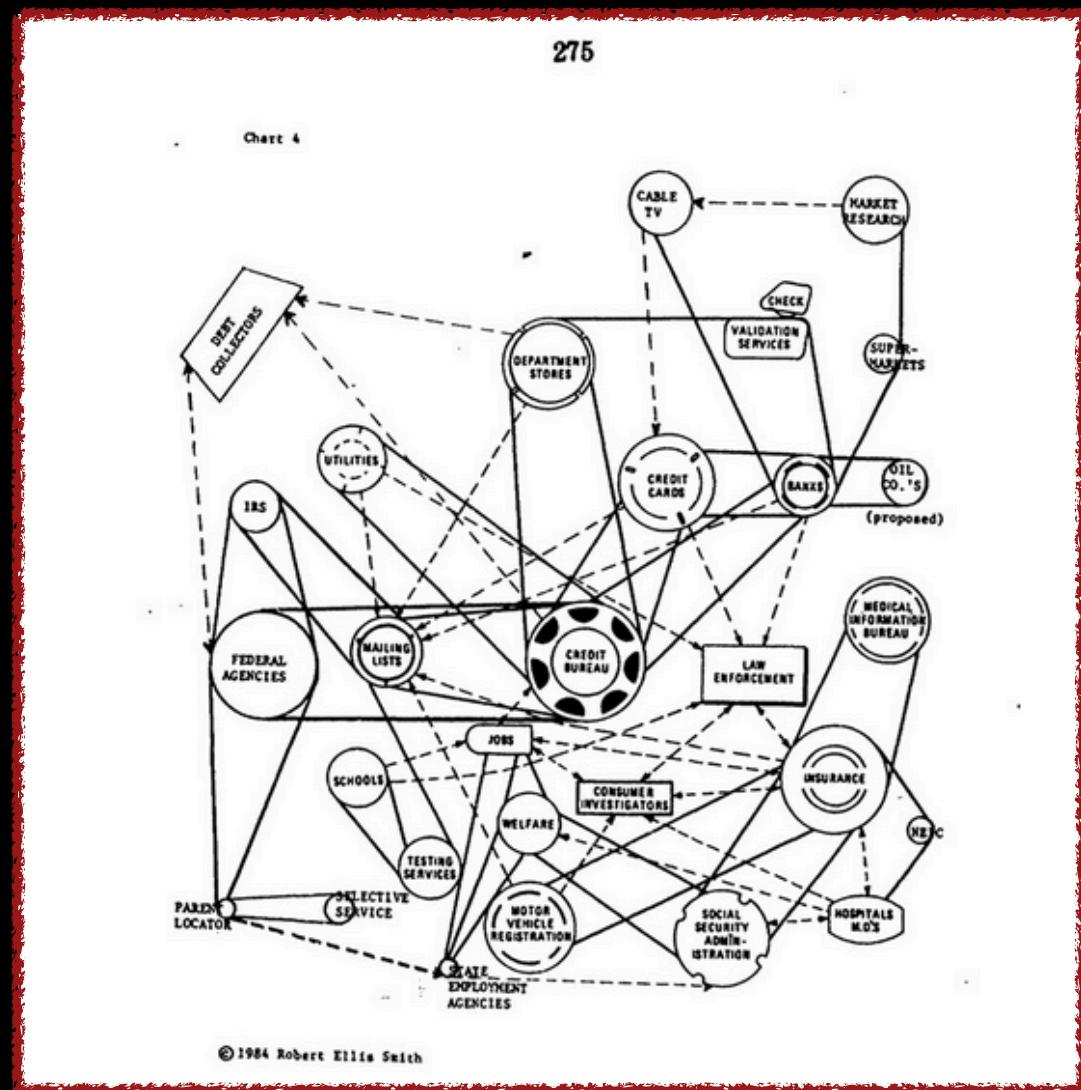
The chapter draws directly from fourteen open-ended **interviews** conducted by the author with (primarily) former Fair, Isaac employees from a number of positions in the production process. Many of these individuals worked with the company their entire careers. Conversations to collect oral histories ranging from an hour and a half to two hours each, and were carried out between June 2004 and October 2006. Three **interviews** represented here are with former senior executive, five with former vice-presidents, four with former senior analysts, one with a former bureau score sales manager, one meeting with four former data entry personnel, and one with a current member of CitiBank's global strategic analytics group. For the sake of simplicity I have indicated a position that differentiates a speakers' approximate generation within the company hierarchy.

e.g., week 9 “big data” + privacy 1950-1980

- 1967: FOIA
- 1970: Social Security Number Task Force
- 1970: Fair Credit Reporting Act
- 1973: Watergate hearings
- 1974: Privacy Act
- 1975: "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate)
- 1975: Rockefeller Commission
- 1975: Pike Committee
- 1974: The Family Educational Rights and Privacy Act

e.g., week 9 “big data” + privacy 1950-1980

e.g., week 9 “big data” + privacy 1950-1980



e.g., week 9 “big data” + privacy 1950-1980

275

Chart 4

And where would we find the names of these films?

```
In [93]: films[the_good_stuff]
```

```
Out[93]:
```

movie_id	movie_title	release_date	video_release_date	IMDb_URL	unknown	Action	Adventure	Animation	Children's	Comedy	... Fantasy
12	Usual Suspects, The (1995)	14-Aug-1995	NaN	http://us.imdb.com/title-exact?Usual%20Suspe...	0	0	0	0	0	0	0
50	Star Wars (1977)	01-Jan-1977	NaN	http://us.imdb.com/title-exact?Star%20Wars%2...	0	1	1	0	0	0	0
64	Shawshank Redemption, The (1994)	01-Jan-1994	NaN	http://us.imdb.com/title-exact?Shawshank%20R...	0	0	0	0	0	0	0
98	Silence of the Lambs, The (1991)	01-Jan-1991	NaN	http://us.imdb.com/title-exact?Silence%20of%...	0	0	0	0	0	0	0

How find the bad stuff?

```
In [ ]:
```

©1984 Robert Ellis Smith

e.g., week 9 “big data” + privacy 1950-1980

275

Chart 4

And where would we find the names of these films?

```
In [93]: films[the_good_stuff]
```

```
Out[93]:
```

movie_id	movie_title	release_date	video_release_date	IMDb_URL	...
12	Usual Suspects, The	14-Aug-1995			Action
50	Star Wars (1977)	01-Jan-1977			Adventure
64	Shawshank Redemption, The	01-Jan-1994			Animation
98	Silence of the Lambs, The	01-Jan-1991			Children's
					Comedy
					Fantasy

```
In [132]: most_similar(405, sliced)
```

movie_id	movie_title
405	Mission: Impossible (1996)
712	Tin Men (1987)
642	Grifters, The (1990)
846	To Gillian on Her 37th Birthday (1996)
686	Perfect World, A (1993)

Name: movie_title, dtype: object

Back to Netflix challenge

A photograph showing seven men in suits standing behind a large ceremonial check. The check is dated 04/16/09 21:09 and is made out to "BellKor's Pragmatic Chaos" for \$1,000,000. The check is signed by "Rod Hastings".

Very close--came down to which group submitted first!

A yellow rectangular banner with the words "Netflix Prize" and "COMPLETED" in red and black text. There is a small graphic of a star or flower between the two words.

e.g., week 10 “data science”, 1962-present

e.g., week 10 “data science”, 1962-present

THE FUTURE OF DATA ANALYSIS¹

BY JOHN W. TUKEY

Princeton University and Bell Telephone Laboratories

e.g., week 10 “data science”, 1962-present

THE FUTURE OF DATA ANALYSIS¹

By JOHN W. TUKEY

I. GENERAL CONSIDERATIONS

1. Introduction. For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. And when I

e.g., week 10 “data science”, 1962-present

THE FUTURE OF DATA ANALYSIS¹

By JOHN W. TUKEY

I. GENERAL CONSIDERATIONS

1. Introduction. For a long time I have thought I was a statistician, interested in inference and mathematical

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

e.g., week 10 “data science”, 1962-present

THE FUTURE OF DATA ANALYSIS¹

By JOHN W. TUKEY

I. GENERAL CONSIDERATIONS

1. Introduction. For a long time I have thought I was a statistician, interested in inference and in mathematical

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistica

Leo Breiman

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS

2017, VOL. 26, NO. 4, 745–766

<https://doi.org/10.1080/10618600.2017.1384734>

50 Years of Data Science

David Donoho

e.g., week 10 “data science”, 1962-present

THE FUTURE OF DATA ANALYSIS¹

By JOHN W. TUKEY

I. GENERAL CONSIDERATIONS

1. Introduction. For a long time I have thought I was a statistician, interested in inference and mathematical

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistics

Leo Breiman

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS

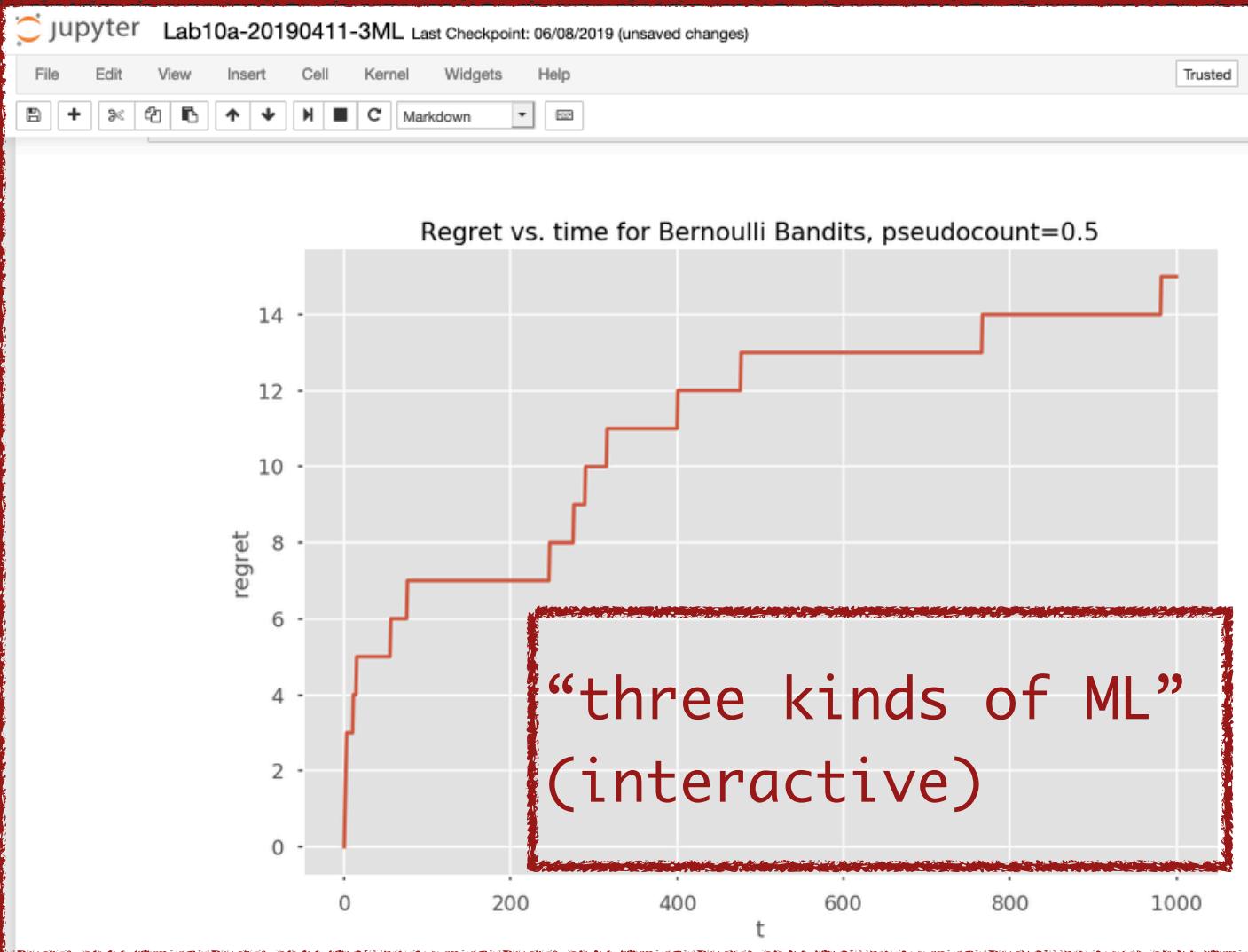
2017, VOL. 26, NO. 4, 745–766

<https://doi.org/10.1080/10618600.2017.1384734>

**Critique and Contribute:
A Practice-Based Framework for Improving
Critical Data Studies and Data Science**

Gina Neff,^{1,*} Anissa Tanweer,² Brittany Fiore-Gartland,³ and Laura Osburn⁴

David Donoho



jupyter Lab10a-20190411-3ML Last Checkpoint: 06/08/2019 (unsaved changes)

jupyter Lab10b-20190404-trees (autosaved)

```
In [17]: # show us the graph of the trees

from IPython.display import Image
import pydotplus
from sklearn import tree
dot_data = tree.export_graphviz(clf, out_file=None,
                               feature_names=names[1:19],
                               class_names=['die', 'live'],
                               filled=True, rounded=True,
                               special_characters=True)
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())
```

Out[17]:

```
graph TD
    Root["SPIDERS ≤ 1.33  
gini = 0.32  
samples = 85  
value = [17, 68]  
class = live"] -- True --> Node1["ALK_PHOSPHATE ≤ 144.0  
gini = 0.5  
samples = 30  
value = [15, 15]  
class = die"]
    Root -- False --> Node2["VARICES ≤ 1.44  
gini = 0.07  
samples = 55  
value = [2, 53]  
class = live"]
    Node1 --> Node3["BILIRUBIN ≤ 1.464  
gini = 0.444  
samples = 21  
value = [14, 7]  
class = die"]
    Node1 --> Node4["ALBUMIN ≤ 2.85  
gini = 0.198  
samples = 9  
value = [1, 8]  
class = live"]
    Node2 --> Node5["gini = 0.0  
samples = 2  
value = [2, 0]  
class = die"]
    Node2 --> Node6["gini = 0.0  
samples = 53  
value = [0, 53]  
class = live"]
```

jupyter Lab10a-20190411-3ML Last Checkpoint: 06/08/2019 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted

jupyter Lab10b-20190404-trees (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python

In [17]: # show us the code

```
from IPython.display import pydotplus
from sklearn import tree
dot_data = tree.export_graphviz(clf, feature_names=featureNames,
                                class_names=targetNames, filled=True, rounded=True,
                                special_characters=True)
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())
```

Out[17]:

Random forests

dramatically increase predictive power at the cost of interpretability

combine

A

B

In [19]: # Build a forest and compute the feature importances

```
from sklearn.ensemble import RandomForestClassifier
seed = 7
num_trees = 100
max_features = 3
#kfold = model_selection.KFold(n_splits=10, random_state=seed)
forest = RandomForestClassifier(n_estimators=num_trees, max_features=max_features)
clf=forest.fit(X_train, y_train)
```

In [20]: scores = cross_val_score(clf, X_test, y_test)

In [21]: scores

Out[21]: array([0.79166667, 0.86956522, 0.82608696])

Breiman:

...forests are A+ predictors But their mechanism for producing a prediction is difficult to understand. Trying to delve into the tangled web that generated a plurality vote from 100 trees is a Herculean task. So on interpretability they rate an F.

jupyter Lab10a-20190411-3ML Last Checkpoint: 06/08/2019 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted

jupyter Lab10b-20190404-trees (autosaved)

File Edit View Insert Not Trusted Python

In [17]: # show us the code

```
from IPython.display import pydotplus
from sklearn import datasets
dot_data = tree.export_graphviz(clf, feature_names=iris.feature_names, class_names=iris.target_names, filled=True, rounded=True, impurity=False)
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())
```

Out[17]:

Random forests

dramatically increase predictive power at the cost of interpretability

combine

A

B

In [19]: # Build a forest and compute the feature importances

```
from sklearn.ensemble import RandomForestClassifier
seed = 7
num_trees = 100
max_features = 3
#kfold = model_selection.KFold(n_splits=10, random_state=seed)
forest = RandomForestClassifier(n_estimators=num_trees, max_features=max_features)
clf=forest.fit(X_train, y_train)
```

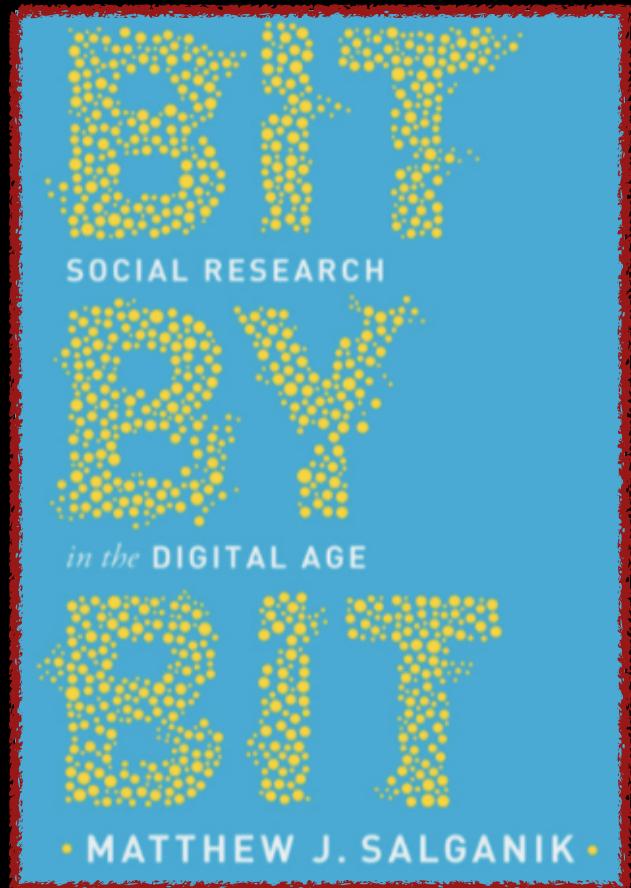
In [20]: scores = cross_val_score(clf, X_test, y_test)

...forests are A+ predictors But their mechanism for producing a prediction is difficult to understand. Trying to delve into the tangled web that generated a plurality vote from 100 trees is a Herculean task. So on interpretability they rate an F.

...forests are A+ predictors But their mechanism for producing a prediction is difficult to understand. Trying to delve into the tangled web that generated a plurality vote from 100 trees is a Herculean task. So on interpretability they rate an F.

e.g., week 12 the ethics of data

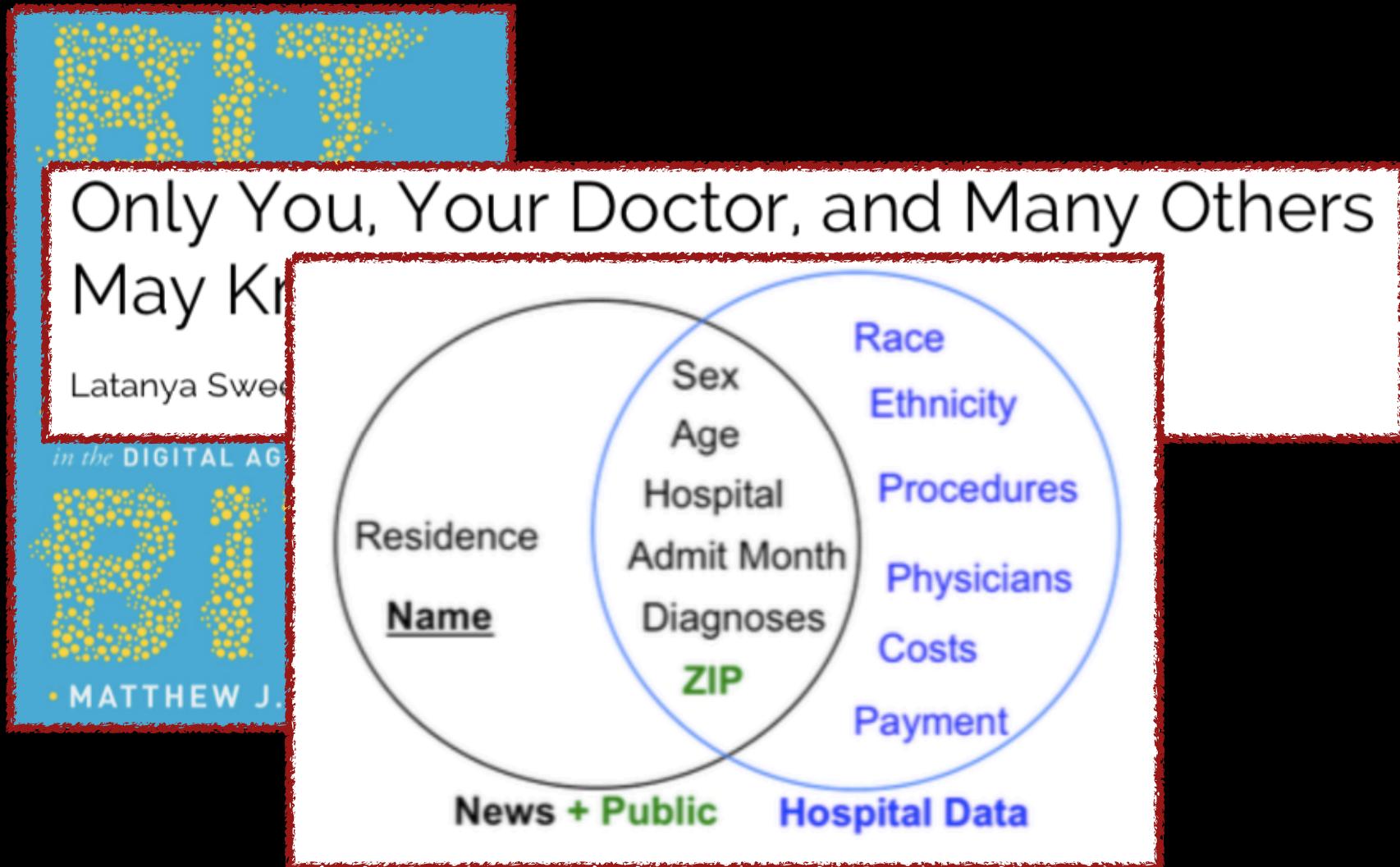
e.g., week 12 the ethics of data



e.g., week 12 the ethics of data



e.g., week 12 the ethics of data



e.g., week 12 the ethics of data

The collage includes:

- A blue book cover titled "BIG DATA in the DIGITAL AGE" by Latanya Sweeney and Matthew J. Salganik.
- A graphic showing a redacted medical record with fields like Name, Residence, New, Sex, Age, Hospital, Admit Month, Race, Ethnicity, Procedures, and a redacted section.
- A red box containing the word "podcasts" in white.
- A black and white photo of two people, one in a lab coat, with the caption "How will AI change your life? AI Now Institute founders Kate Crawford and Meredith Whittaker explain."

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

e.g., week 12 the ethics of data

The screenshot shows a dark-themed version of The New York Times website. At the top left are links for "HOME" and "SEARCH ARCHIVE". On the right is the "The New York Times" logo. Below the logo, the date "WEDNESDAY, JULY 26, 1972" is displayed. On the left side of the main content area, there is a sidebar with "SUBJECTS" and a "INDEX" button. The main article title is "Syphilis Victims in U.S. Study Went Untreated for 40 Years". Below the title is a sub-headline "SYPHILIS VICTIMS GOT NO THERAPY". The main text of the article discusses a study conducted by the United States Public Health Service where syphilis victims were untreated for 40 years. To the right of the main article, there is a sidebar with additional text and a small photo of Dr. Martin K. DuVal.

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

SYPHILIS VICTIMS GOT NO THERAPY

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

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 late effects, even though an effective therapy was eventually discovered.

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

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

Dr. THOMAS H. SPENCE, DIRECTOR OF THE U.S. PUBLIC HEALTH SERVICE, AND DR. JAMES D. RUMSFELD, SECRETARY OF DEFENSE, AT A NEWS CONFERENCE ON JULY 25, 1972, IN WASHINGTON. DR. SPENCE ANNOUNCED THAT THE U.S. PUBLIC HEALTH SERVICE WILL WITHDRAW FROM THE TUSKEGEE STUDY.

Continued on Page 8, Column 1

Continued on Page 8, Column 1

Continued on Page 8, Column 1

e.g
histo

'Never Clandestine'

The syphilis study "was never clandestine" and 15 scientific reports were published in the medical literature, Dr. Millar said in a telephone interview yesterday from Atlanta.

Officials who initiated the study in 1932 had informed the syphilis victims that they could get treatment for the infection at any time, Dr. Millar said.

"Patients were not denied drugs," Dr. Millar stressed. Rather, they were not offered drugs.

When the study began, doctors could offer only what is now regarded as poor therapy —injections of metals like bismuth, arsenic and mercury. Such treatments were known to be toxic.

Many doctors, Dr. Miller said, then thought "it better not to treat syphilis cases because of the mortality from" the metal therapies.

The critical period in ethics was in the late nineteen forties and early nineteen-fifties when antibiotics could have been but were not prescribed for the syphilis patients.

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

e.g.

histo

The experiment, called the Tuskegee Study, began in 1932 with about 600 black men mostly poor and uneducated, from Tuskegee, Ala., an area that had the highest syphilis rate in the nation at the time.

Four hundred of the group had syphilis and never received deliberate treatment for the Venereal Infection. A control group of 200 had no syphilis and did not receive any specific therapy.

Some subjects were added to the study in its early years to replace men who had dropped out of the program, but the number added is not known. At the beginning of this year, 74 of those who received no treatment were still alive.

As incentives to enter the Program, the men were promised free transportation to and from hospitals, free hot lunches, free medicine for any disease other than syphilis and free burial after autopsies were performed.

Could Have Been Helped

The Tuskegee Study began 10 years before penicillin was found to be a cure for syphilis and 15 years before the drug became widely available. Yet, even after penicillin became common, and while its use probably

e.g., week 12 the ethics of data

THE EXPERIMENT AND HEW'S ETHICAL REVIEW

Racism and Research: The Case of the Tuskegee Syphilis Study

by ALLAN M. BRANDT

ALLAN M. BRANDT is a doctoral candidate in the Department of History, Columbia University. He is presently writing a social history of venereal disease in the United States. Mr. Brandt was a student intern at The Hastings Center in 1977.

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
2. articulate tensions among them
3. articulate design to support them

what we talk about when we talk about ethics

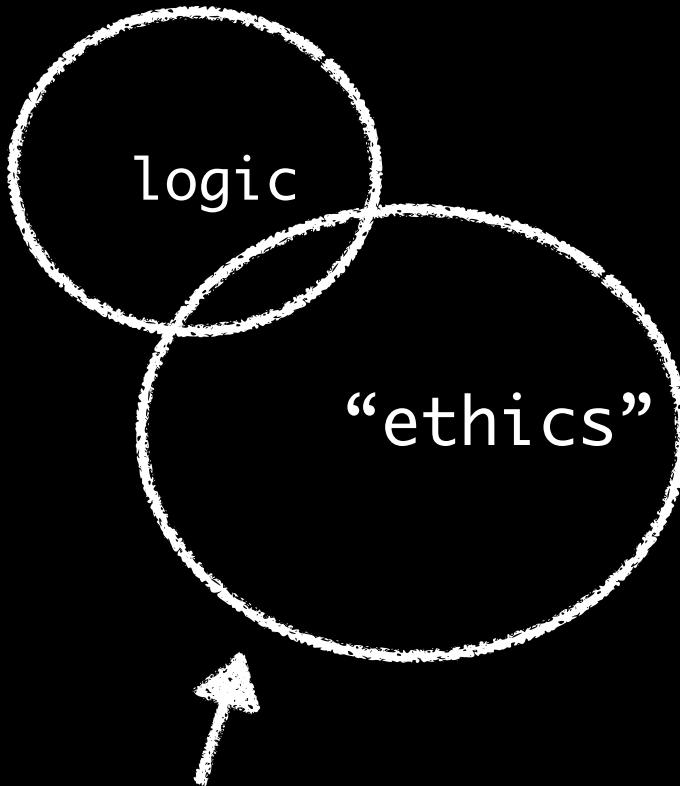
“ethics”

what we talk about when we talk about ethics



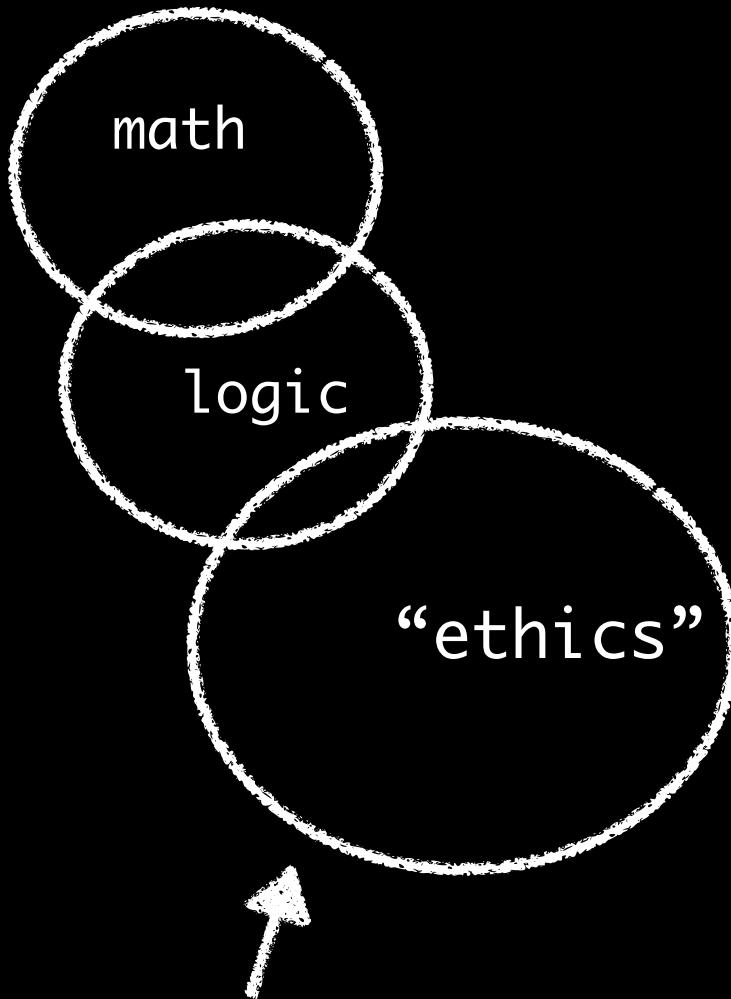
philosophy

what we talk about when we talk about ethics



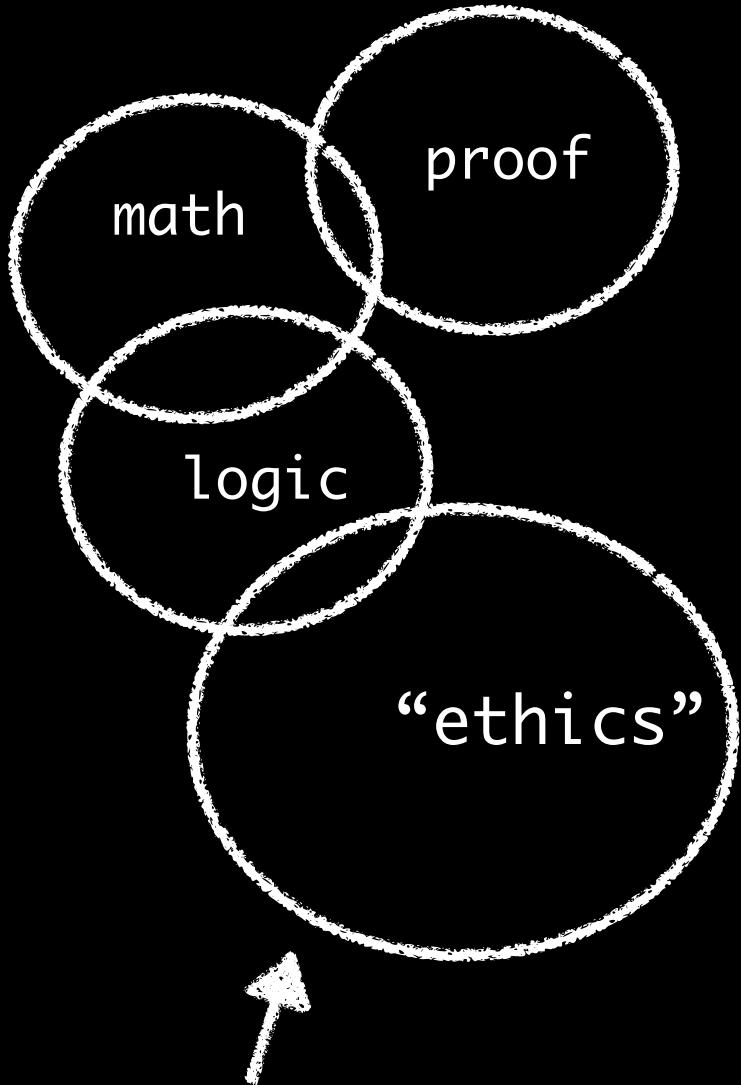
philosophy

what we talk about when we talk about ethics



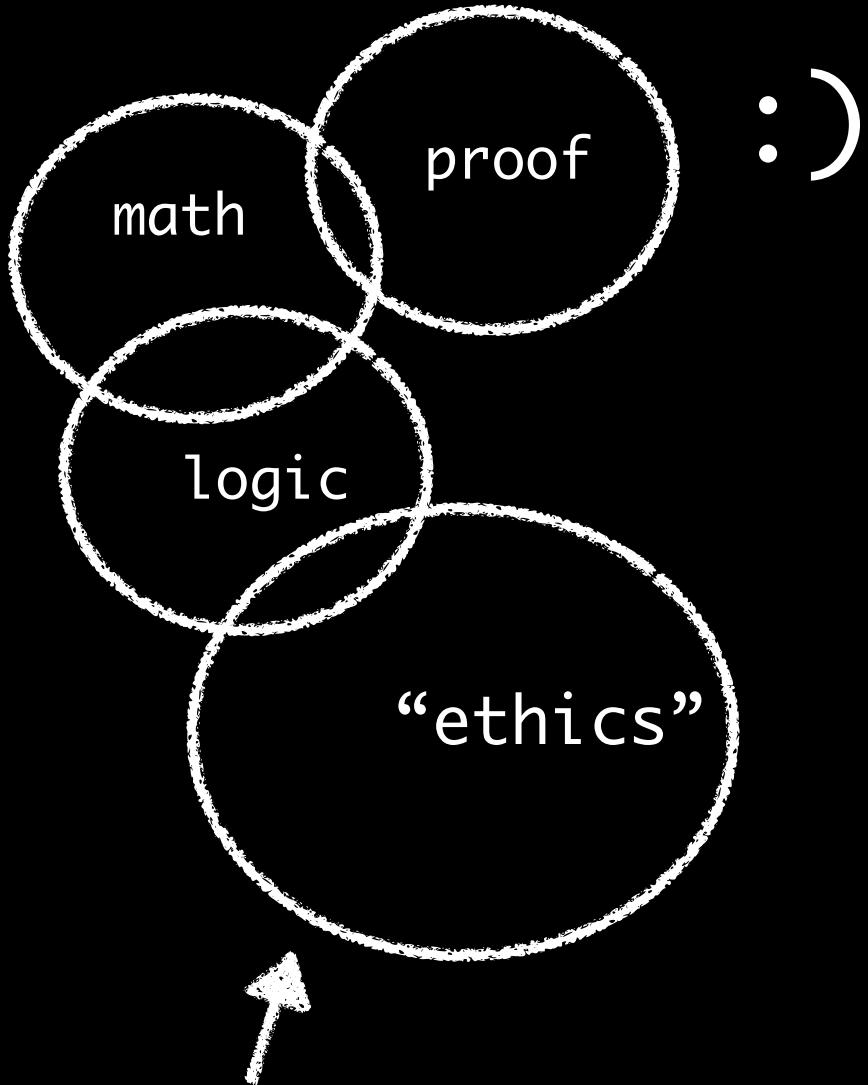
philosophy

what we talk about when we talk about ethics



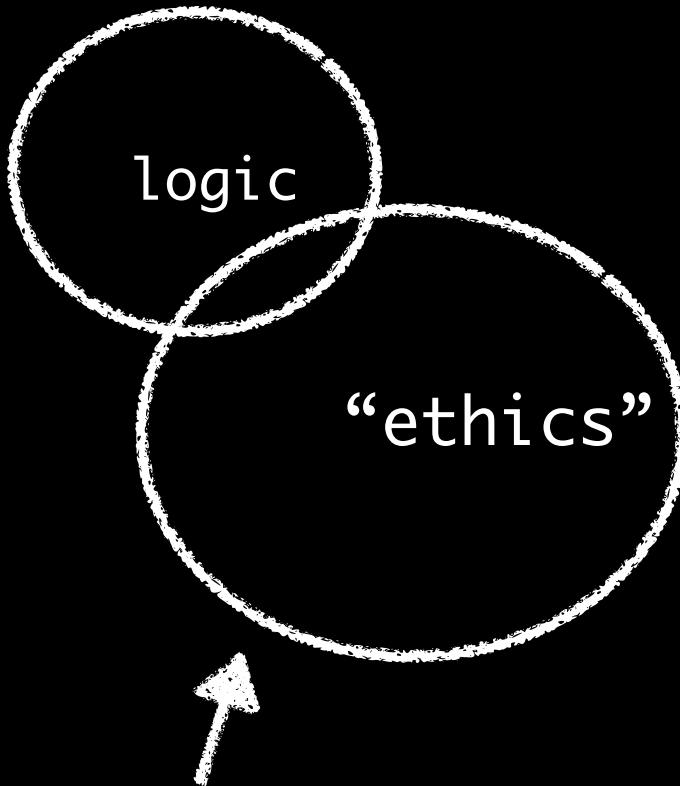
philosophy

what we talk about when we talk about ethics



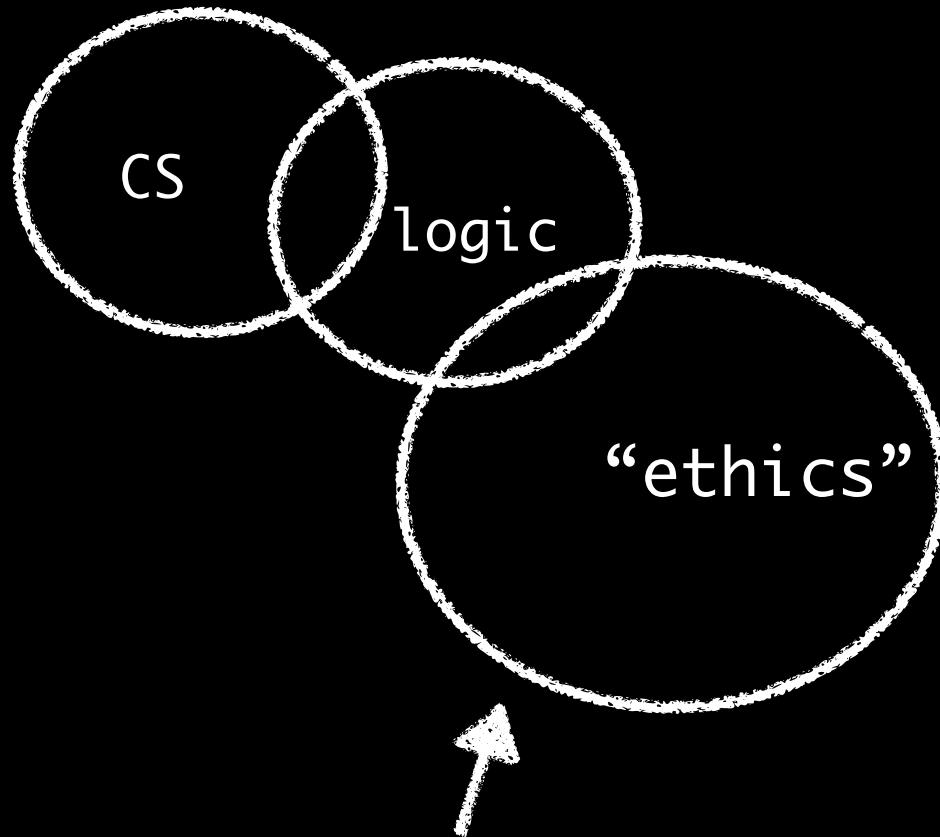
philosophy

what we talk about when we talk about ethics



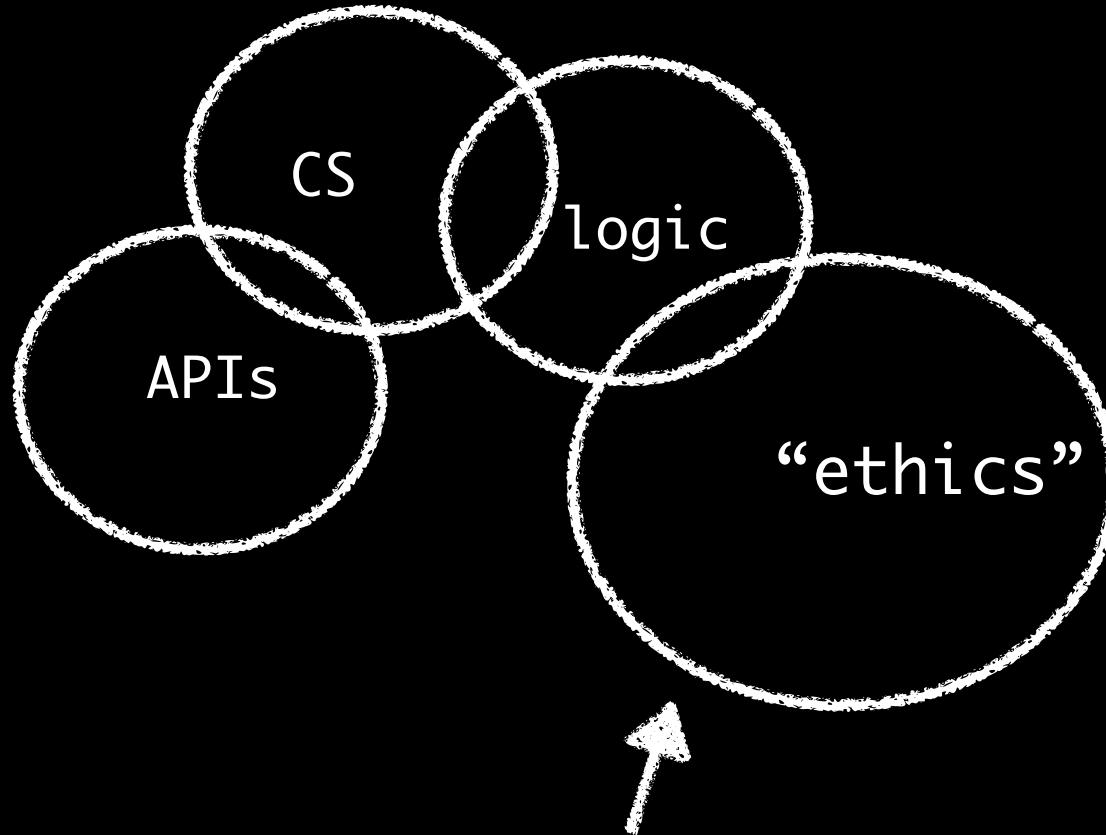
philosophy

what we talk about when we talk about ethics



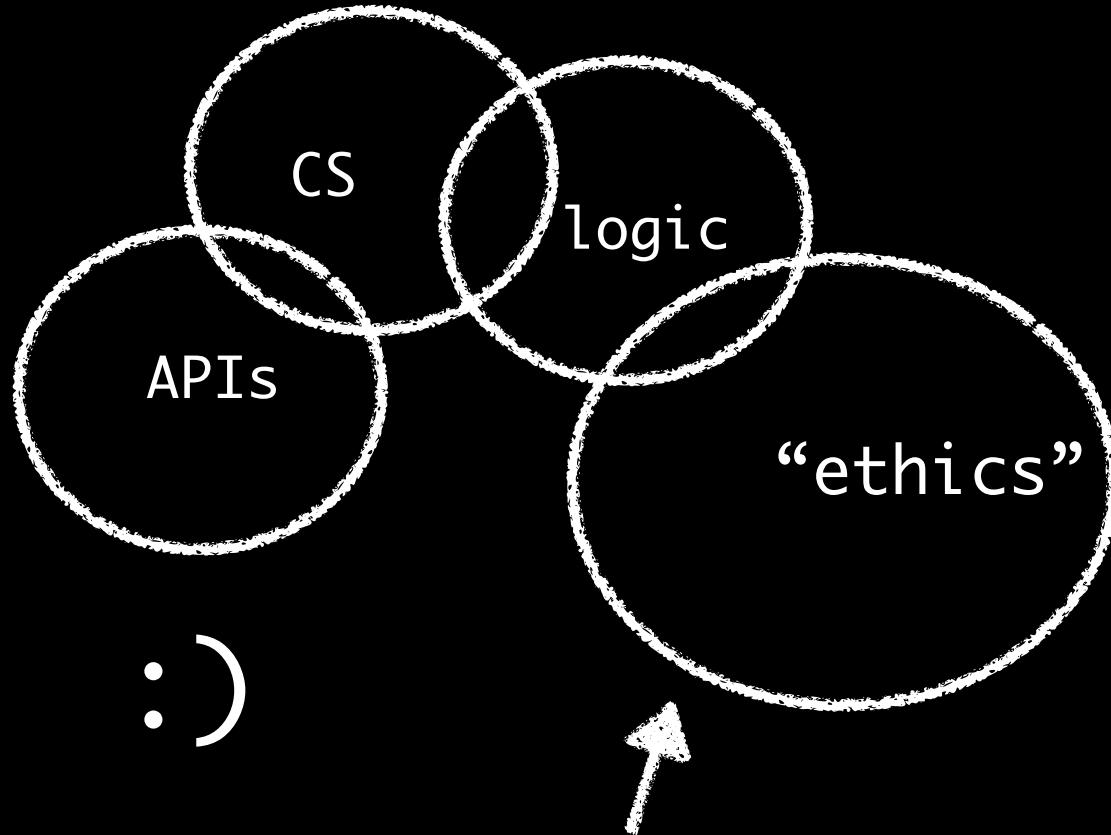
philosophy

what we talk about when we talk about ethics



philosophy

what we talk about when we talk about ethics



what we talk about when we talk about ethics



philosophy

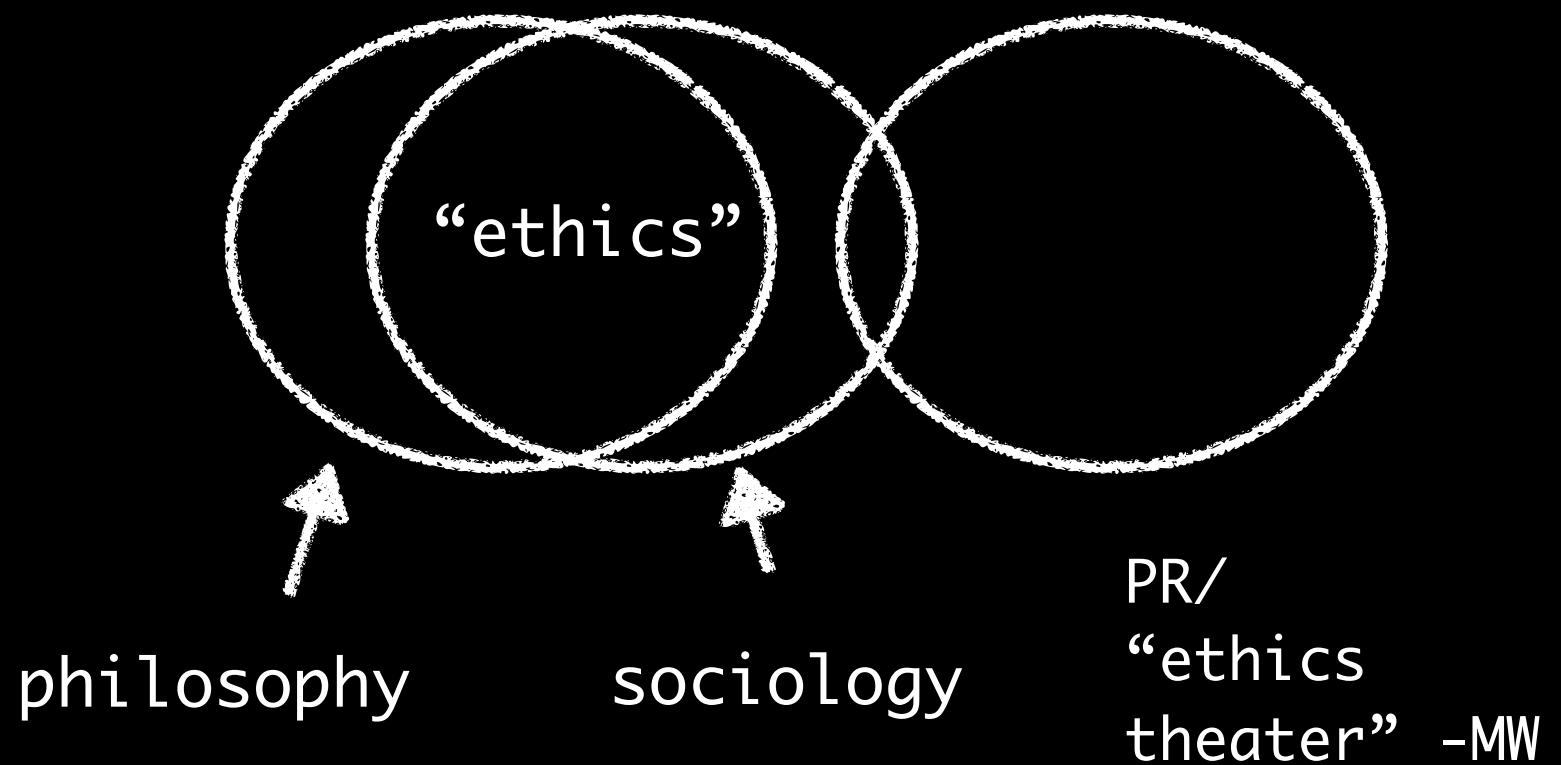
what we talk about when we talk about ethics



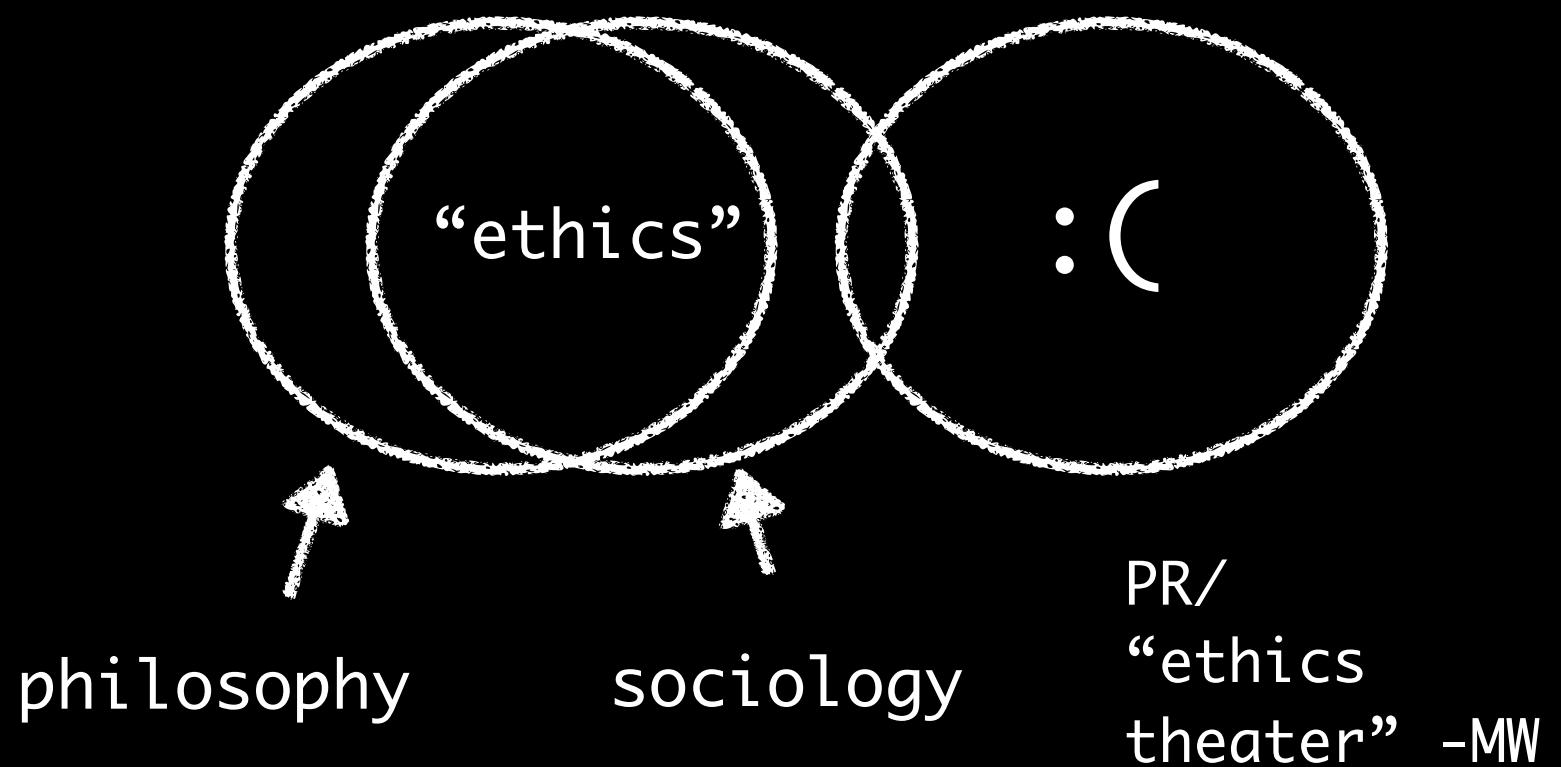
philosophy

sociology

what we talk about when we talk about ethics



what we talk about when we talk about ethics



what we talk about when we talk about ethics



philosophy

(define)

sociology

(design)



philosophy

(define)

sociology

(design)

e.g., week 12 the ethics of data

Belmont principles

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood
 - ~~informed consent~~ -> autonomy

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood
 - ~~informed consent~~ -> autonomy
2. beneficence

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood
 - ~~informed consent~~ -> autonomy
2. beneficence
 - ~~do no harm~~ -> balance risk+benefit

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood
 - ~~informed consent~~ -> autonomy
2. beneficence
 - ~~do no harm~~ -> balance risk+benefit
3. justice

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood
 - ~~informed consent~~ -> autonomy
2. beneficence
 - ~~do no harm~~ -> balance risk+benefit
3. justice
 - ~~legal~~-> fair, e.g., “veil of ignorance”

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood
 - ~~informed consent~~ -> autonomy
2. beneficence
 - ~~do no harm~~ -> balance risk+benefit
3. justice
 - ~~legal~~-> fair, e.g., “veil of ignorance”

e.g., week 12 the ethics of data

Belmont principles

1. respect for personhood
 - ~~informed consent~~ -> autonomy
2. beneficence
 - ~~do no harm~~ -> balance risk+benefit
3. justice
 - ~~legal~~-> fair, e.g., “veil of ignorance”

gives analytical, hierarchical, durable framework
for ethical audit of decisions,
from which rules, code, “design” should derive

e.g., week 12 the ethics of data

from: Wagner, Ben. "Ethics as an Escape from Regulation:
From ethics-washing to ethics-shopping?." (2018).

e.g., week 12 the ethics of data

1. “External Participation: early and regular engagement with all relevant stakeholders.

from: Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." (2018).

e.g., week 12 the ethics of data

1. “External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.

from: Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." (2018).

e.g., week 12 the ethics of data

1. “External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.
3. Ensure transparent decision-making procedures on why decisions were taken.

from: Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." (2018).

e.g., week 12 the ethics of data

1. “External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.
3. Ensure transparent decision-making procedures on why decisions were taken.
4. Develop a stable list of non-arbitrary of standards where the selection of certain values, ethics and rights over others can be plausibly justified.

from: Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." (2018).

e.g., week 12 the ethics of data

1. “External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.
3. Ensure transparent decision-making procedures on why decisions were taken.
4. Develop a stable list of non-arbitrary of standards where the selection of certain values, ethics and rights over others can be plausibly justified.
5. Ensure that ethics do not substitute fundamental rights or human rights.

from: Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." (2018).

e.g., week 12 the ethics of data

1. “External Participation: early and regular engagement with all relevant stakeholders.
2. Provide a mechanism for external independent oversight.
3. Ensure transparent decision-making procedures on why decisions were taken.
4. Develop a stable list of non-arbitrary of standards where the selection of certain values, ethics and rights over others can be plausibly justified.
5. Ensure that ethics do not substitute fundamental rights or human rights.
6. Provide a clear statement on the relationship between the commitments made and existing legal or regulatory frameworks, in particular on what happens when the two are in conflict.”

from: Wagner, Ben. "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?." (2018).

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
 - consistent w/norms, rights, philosophy

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
 - consistent w/norms, rights, philosophy
2. articulate tensions among them

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
 - consistent w/norms, rights, philosophy
2. articulate tensions among them
 - e.g., “ends” vs “means”

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
 - consistent w/norms, rights, philosophy
2. articulate tensions among them
 - e.g., “ends” vs “means”
3. articulate design to support them

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
 - consistent w/norms, rights, philosophy
2. articulate tensions among them
 - e.g., “ends” vs “means”
3. articulate design to support them
 - interaction design

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
 - consistent w/norms, rights, philosophy
2. articulate tensions among them
 - e.g., “ends” vs “means”
3. articulate design to support them
 - interaction design
 - process design

e.g., week 12 the ethics of data

history: Tuskegee -> Belmont

1. articulate ethics as principles
 - consistent w/norms, rights, philosophy
2. articulate tensions among them
 - e.g., “ends” vs “means”
3. articulate design to support them
 - interaction design
 - process design
 - (in this case, the IRB process)

e.g., week 12 define + design for ethics

“The Commission’s deliberations on Institutional Review Boards began with the premise that investigators should not have sole responsibility for determining whether research involving human subjects fulfills ethical standards. Others who are independent of the research must share this responsibility, because investigators have a potential conflict by virtue of their concern with the pursuit of knowledge as well as the welfare of the human subjects of their research.”

1978-09-01 IRB recommendation

e.g., week 12 define + design for ethics

“The Commission’s deliberations on Institutional Review Boards began with the premise that investigators should not have sole responsibility for determining whether research involving human subjects fulfills ethical standards. Others who are independent of the research must share this responsibility, because investigators have a potential conflict by virtue of their concern with the pursuit of knowledge as well as the welfare of the human subjects of their research.”

1978-09-01 IRB recommendation

reminder: “design is the intentional solution to a problem within a set of constraints.” – Mike Monteiro

e.g., week 12 define + design for ethics

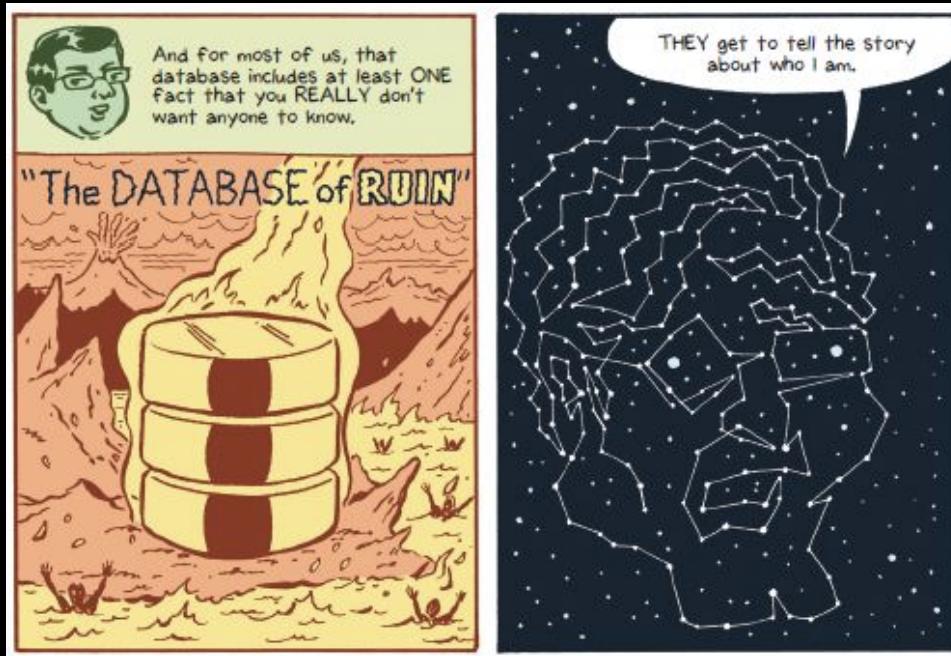
“The Commission’s deliberations on Institutional Review Boards began with the premise that investigators should not have sole responsibility for determining whether research involving human subjects fulfills ethical standards. Others who are independent of the research must share this responsibility, because investigators have a potential conflict by virtue of their concern with the pursuit of knowledge as well as the welfare of the human subjects of their research.”

1978-09-01 IRB recommendation

reminder: “design is the intentional solution to a problem within a set of constraints.” – Mike Monteiro

e.g., week 12 ethics lab:

- database of ruin
- k-anonymity
- terms of service



kdnuggets.com (2014):
“Big Data Comic Explains the Current State of Privacy”

e.g., week 13 (present) problems



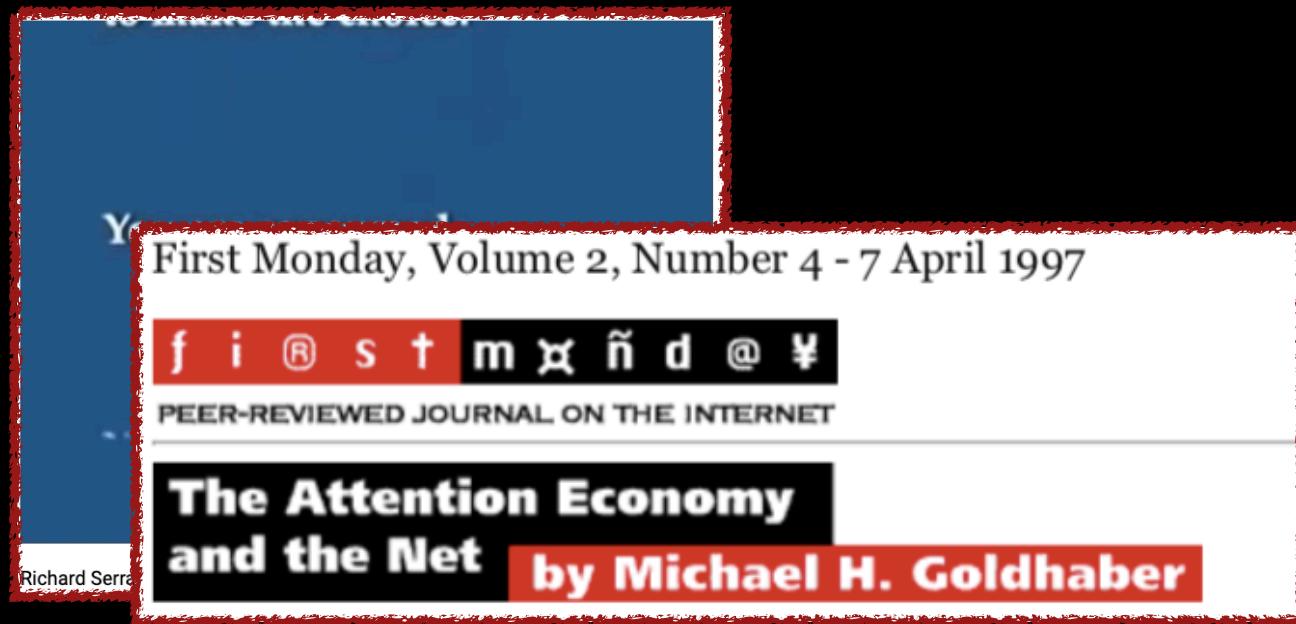
Richard Serra "Television Delivers People" (1973)

e.g., week 13 (present) problems

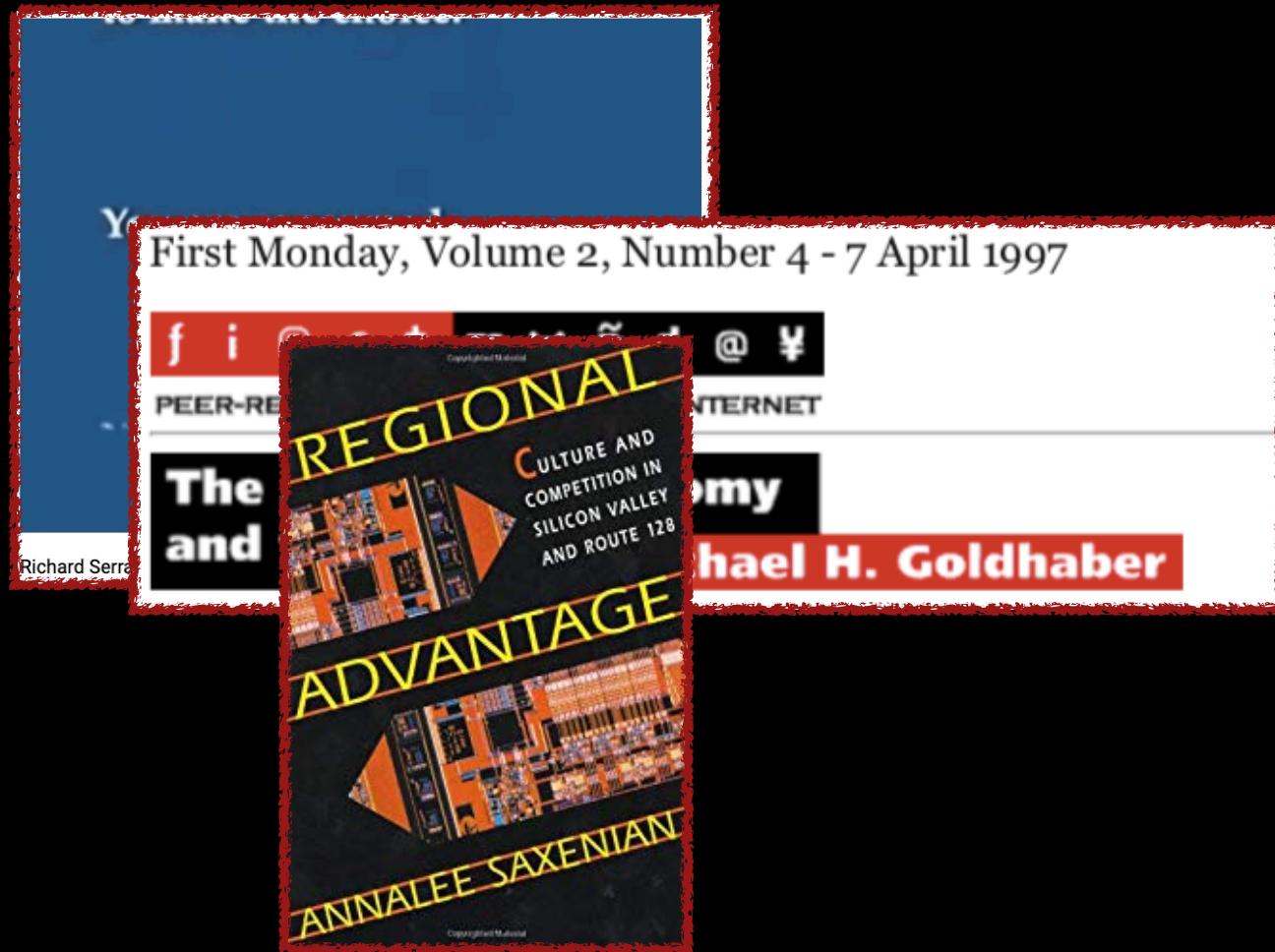


Richard Serra "Television Delivers People" (1973)

e.g., week 13 (present) problems



e.g., week 13 (present) problems



e.g., week 13 (present) problems

First Monday, Volume 2, Number 4 - 7 April 1997

f i C Copyright Material @ ¥
PEER-REVIEWED JOURNAL INTERNET

The and

Richard Serra

AN AT

REGIONAL CULTURE AND COMPETITION IN THE SAN VALLEY Economy

The Platform is the Message
James Grimmelmann*

I. TIDE PODS	2
II. PARODY	5
III. VIRALITY	8
IV. FAKE NEWS.....	14

e.g., week 13 (present) problems

A collage of images illustrating various tech and culture topics from the late 20th century:

- A blue book cover for "First Monday, Volume 2, Number 4 - 7 April 1997".
- A red book cover for "The Platform is the Message" by James Grimmelmann*.
- A black book cover for "The fundamental problem with Silicon Valley's favorite growth strategy" by Tim O'Reilly.
- A white book cover for "Blitzscaling" by Tim O'Reilly.
- A red book cover for "Regional Culture and Competition in Silicon Valley".
- A white book cover for "The Art of War" by Sun Tzu.
- A blue book cover for "Richard Serra: The Work".
- A red book cover for "Peer-to-Peer".
- A white book cover for "The Internet Book".

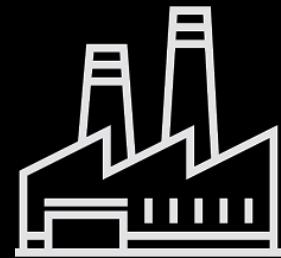
e.g., week 14 (future) solutions

e.g., week 14 (future) solutions

e.g., week 14 (future) solutions

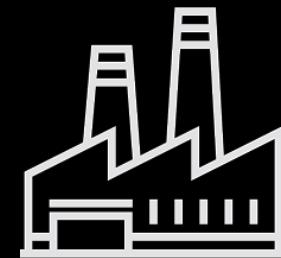
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



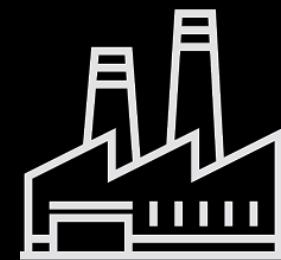
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



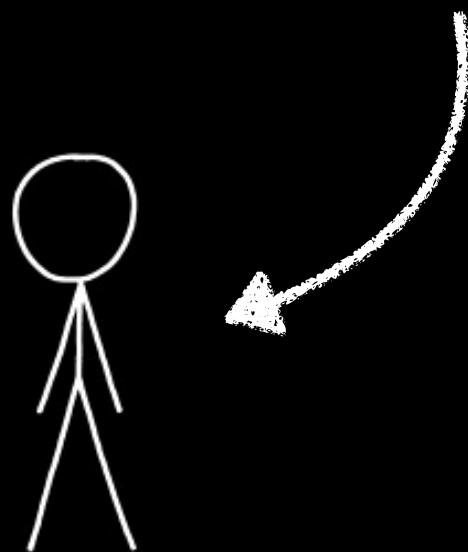
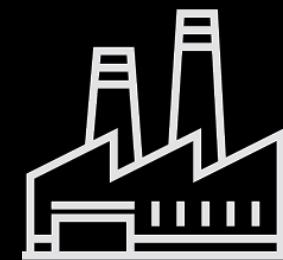
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



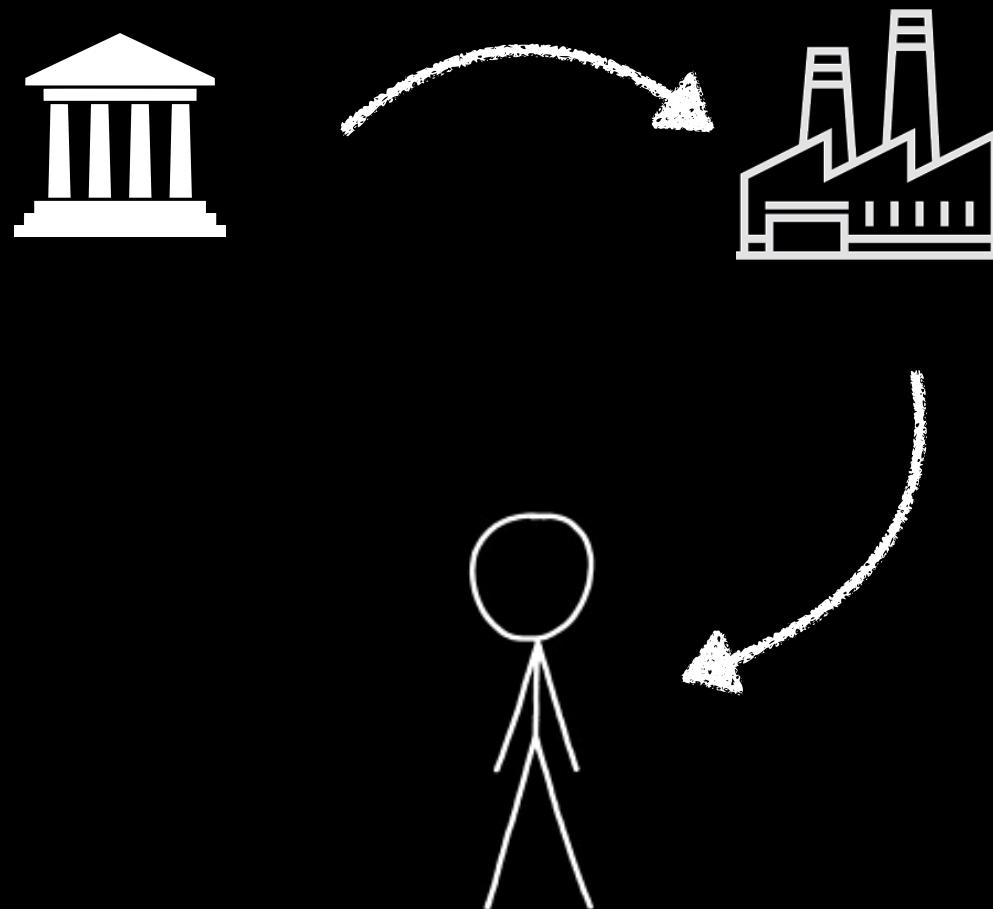
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



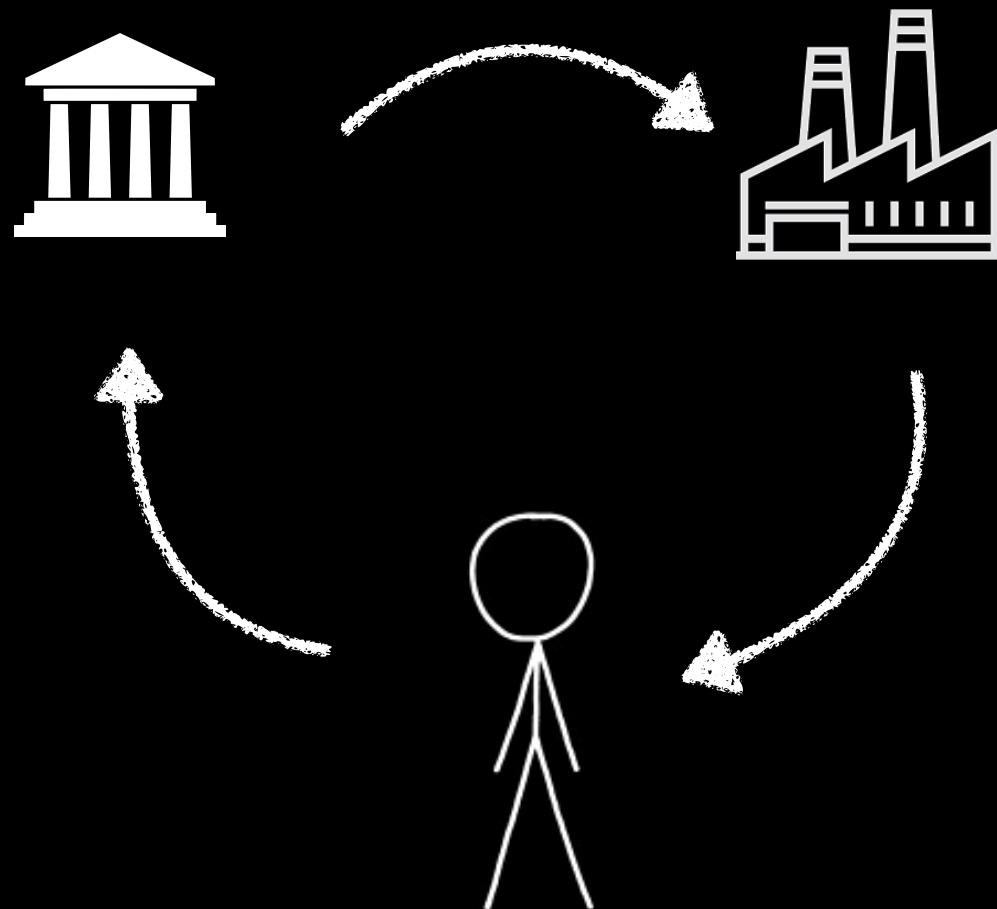
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



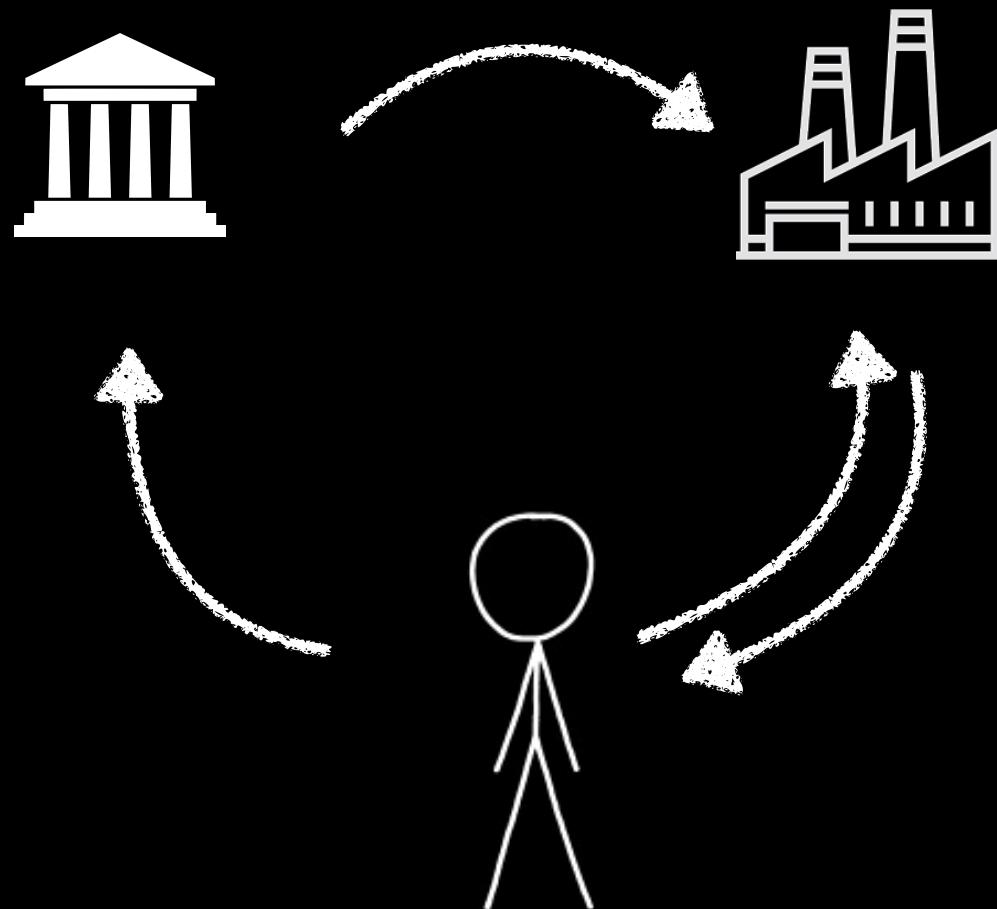
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



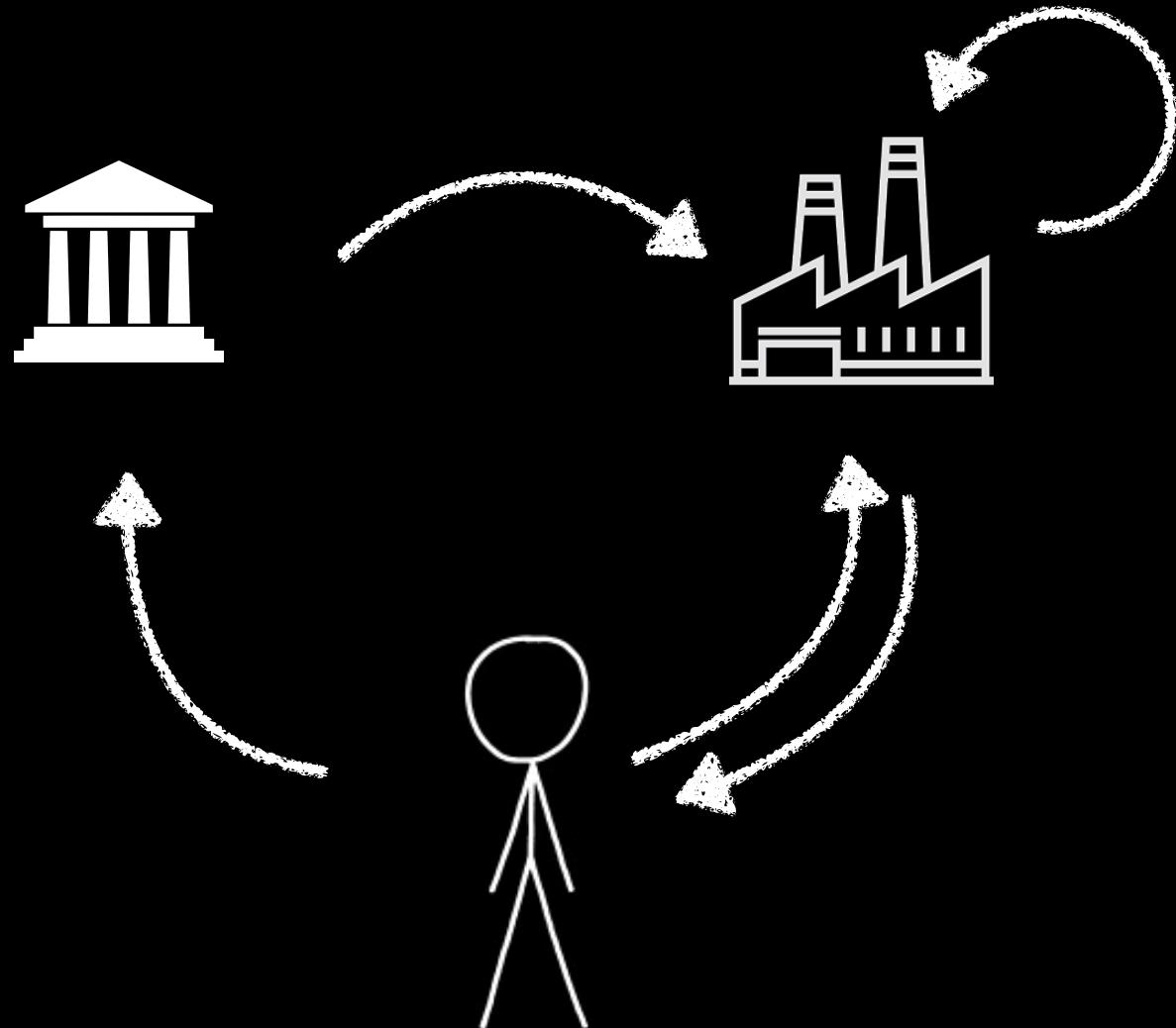
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



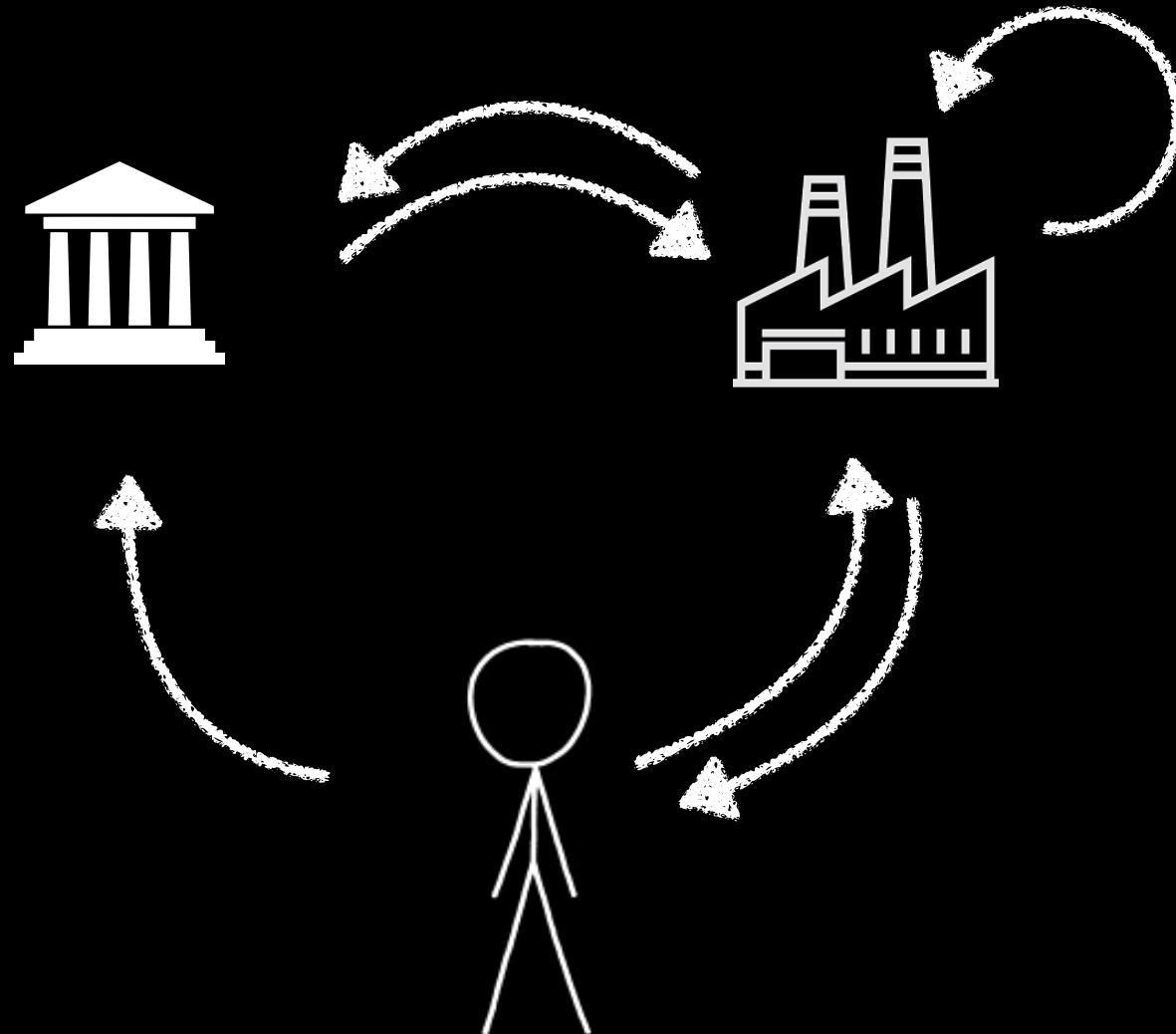
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



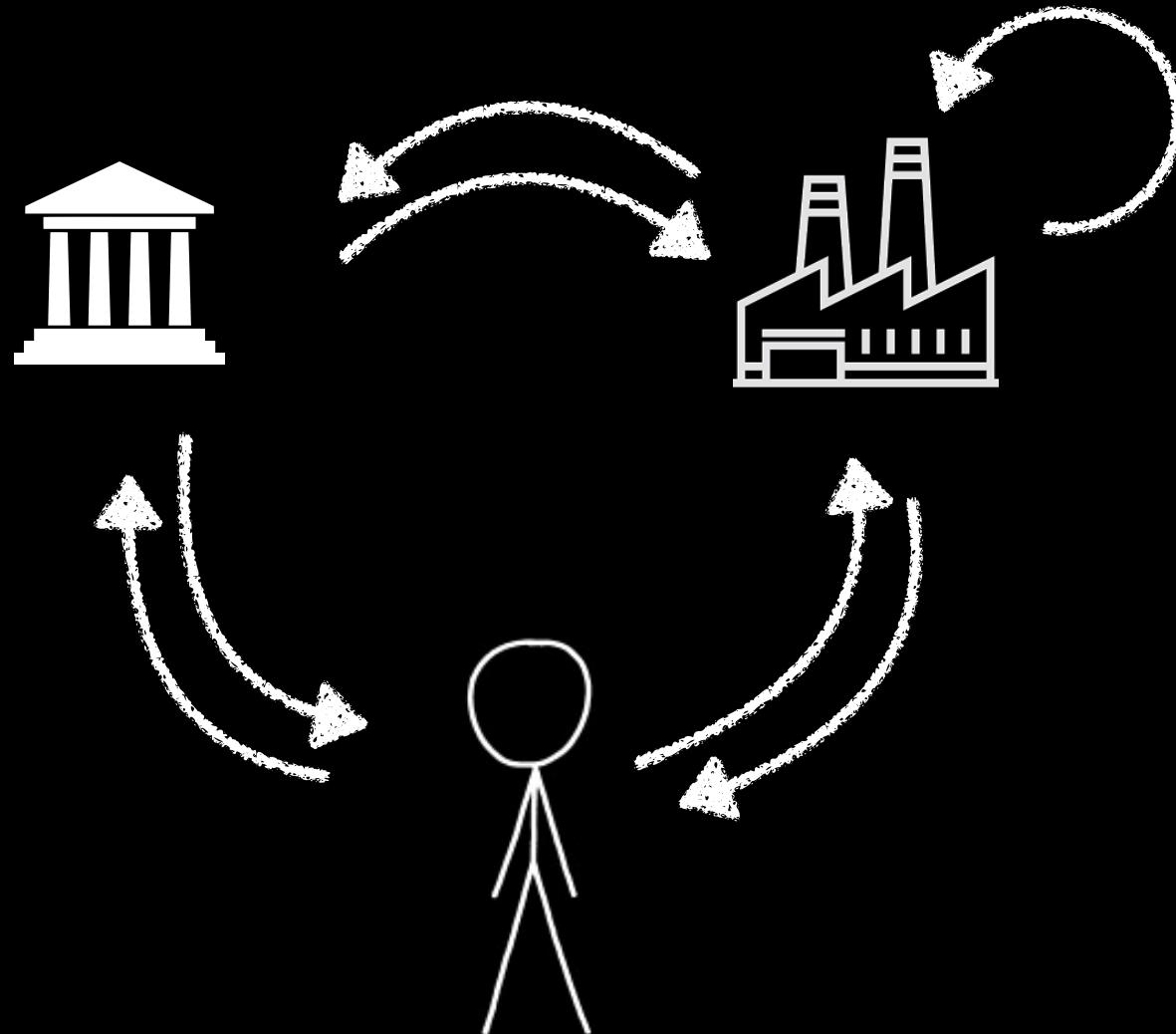
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



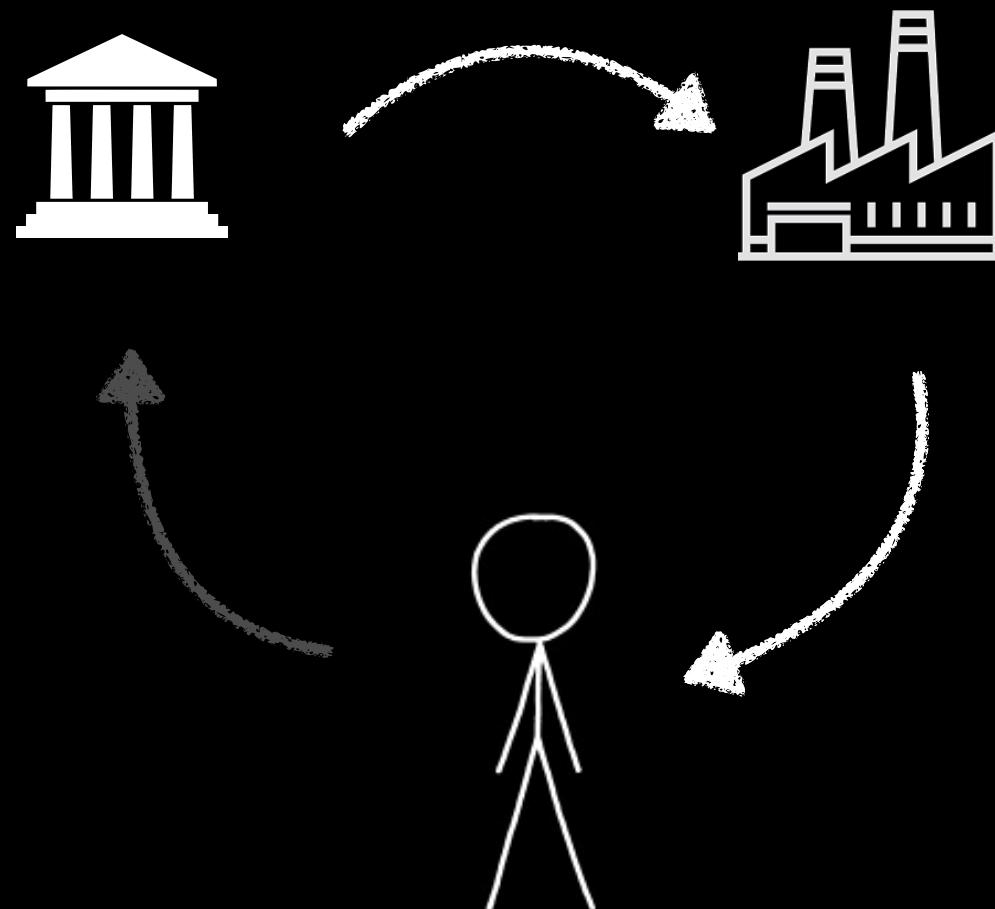
3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions



3-player unstable game (adapted from Janeway)

e.g., example: IRB



3-player unstable game (adapted from Janeway)

e.g., week 14 (future) solutions

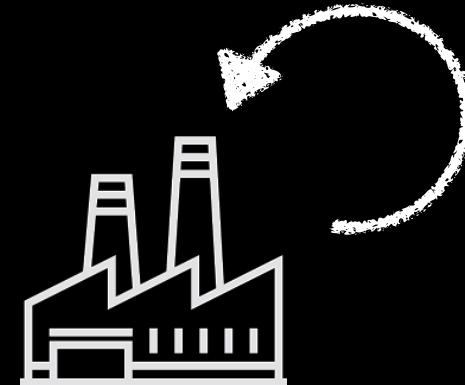
Opinion Digital economy

Privacy is a competitive advantage

Technology companies may have to say whether they are data peddlers or data stewards

RANA FOROOHAR

+ Add to myFT



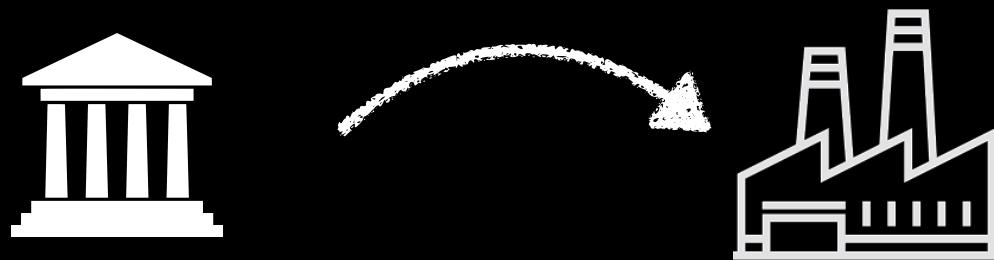
2015-10-01, APPL: “privacy is a fundamental human right”

2017-10-15 (FT) “privacy has become a competitive advantage.”

2019-02-07 CSC0: “privacy is a fundamental human right”

2019-04-28 FB: “The future is private”

e.g., week 14 (future) solutions



e.g., week 14 (future) solutions



DEFUSING
DISINFO

ESSAYS ABOUT CONTACT

WE'VE DIAGNOSED THE DISINFORMATION PROBLEM. NOW, WHAT'S THE PRESCRIPTION?

essays

january 23, 2019

By Renée DiResta

e.g., week 14 (future) solutions

The image is a collage of four distinct elements arranged horizontally. At the top left is a white outline of a classical temple or government building. To its right is a white outline of a factory with three smokestacks, connected by a large, hand-drawn style curved arrow pointing from the building towards the factory. Below these icons is a screenshot of a website titled 'DEFUSING DISINFO'. The main headline on the site reads 'WE'VE DIAGNOSED THE DISINFORMATION PROBLEM. NOW, WHAT'S THE PRESCRIPTION?'. Below the headline are two small blue text links: 'essays' and 'january 23, 2019'. At the bottom left of the collage, the author's name 'By Renée DiResta' is written in blue. To the right of the website screenshot is a photograph of a young woman, Lina Khan, smiling. She is holding a stack of several thick books in her arms. She is positioned in front of bookshelves filled with books.

DEFUSING
DISINFO

ESSAYS ABOUT CONTACT

WE'VE DIAGNOSED THE
DISINFORMATION PROBLEM.
NOW, WHAT'S THE
PRESCRIPTION?

essays january 23, 2019

By Renée DiResta

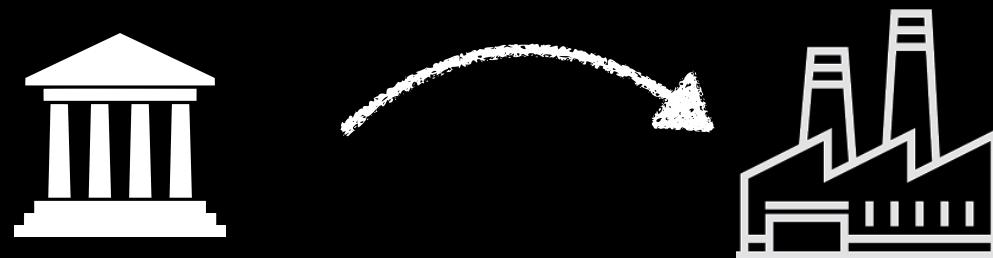
The New York Times

*Amazon's Antitrust
Antagonist Has a
Breakthrough Idea*

With a single scholarly article, Lina Khan, 29, has reframed decades of monopoly law.

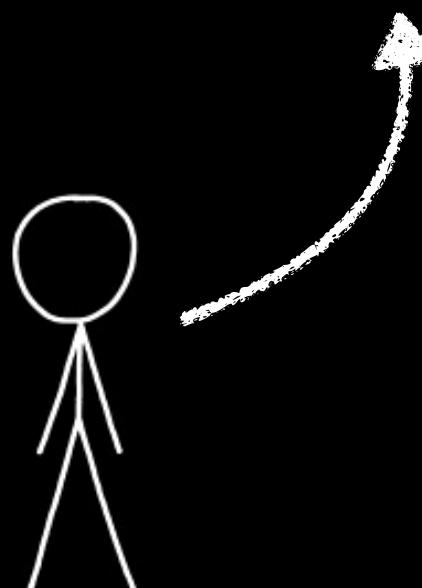
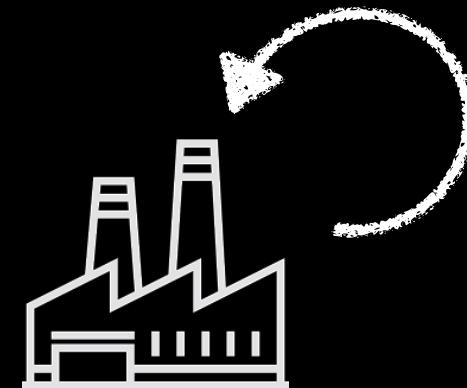
Lina Khan, 29, is a law professor at the University of Michigan who has proposed a new way to regulate Amazon. Her article, published in the Harvard Law Review, challenges the traditional legal framework foritrust law, which focuses on whether a company has "market power" and "abuses it." Khan argues that this approach is too narrow and fails to address the specific ways in which Amazon uses its power to harm competition. Instead, she proposes a "functional" approach that looks at how a company actually operates in the market. This approach could lead to more effective antitrust enforcement against Amazon and other large tech companies.

e.g., week 14 (future) solutions



- GDPR
- California Consumer Privacy Act (CCPA)
- rise of “Hipster Antitrust”
- CDA 230
- FTC “Do Not Track Me Online Act of 2011”
- FEC “Honest Ads Act”
- “SEC for the technology industry” - DiResta
- ...

e.g., week 14 (future) solutions

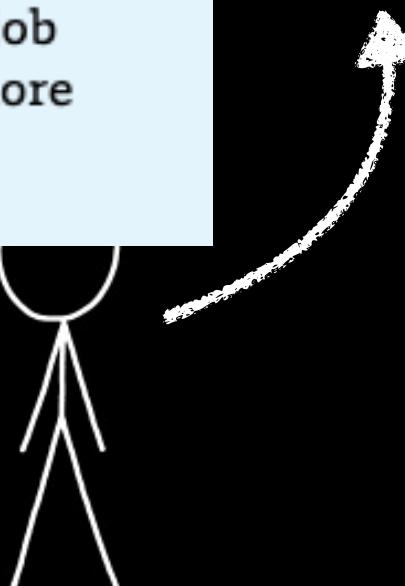
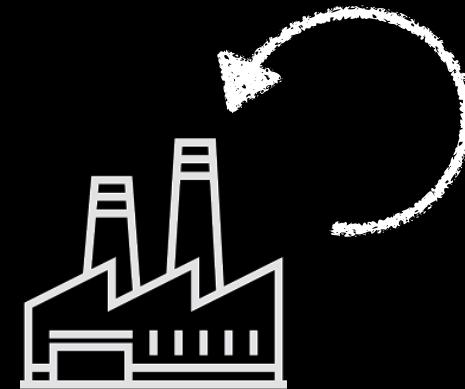


e.g., week 14 (future) solutions

JANUARY 23, 2019

The Tech Revolt

A sometimes pointed, sometimes resigned conversation with engineers, designers, research scientists, and job candidates who are pushing for a more ethical Silicon Valley



e.g., week 14 (future) solutions

://story.californiasunday.com/tech-revolt 67% ...

The Tech Revolt

Politics Comes to Silicon Valley: A Timeline

DECEMBER 2016: NEVERAGAIN.TECH

A group of engineers and activists brainstorm about how best to oppose incoming President Donald Trump's policies. They launch neveragain.tech, an online pledge whose signers promise never to cooperate with attempts to collect data on immigrants and Muslims for the purpose of racial profiling.

JANUARY 2017: SILICON VALLEY SECURITY GUARDS UNIONIZE

More than 3,000 security guards at Cisco, Facebook, Genentech, and other tech companies win union representation.

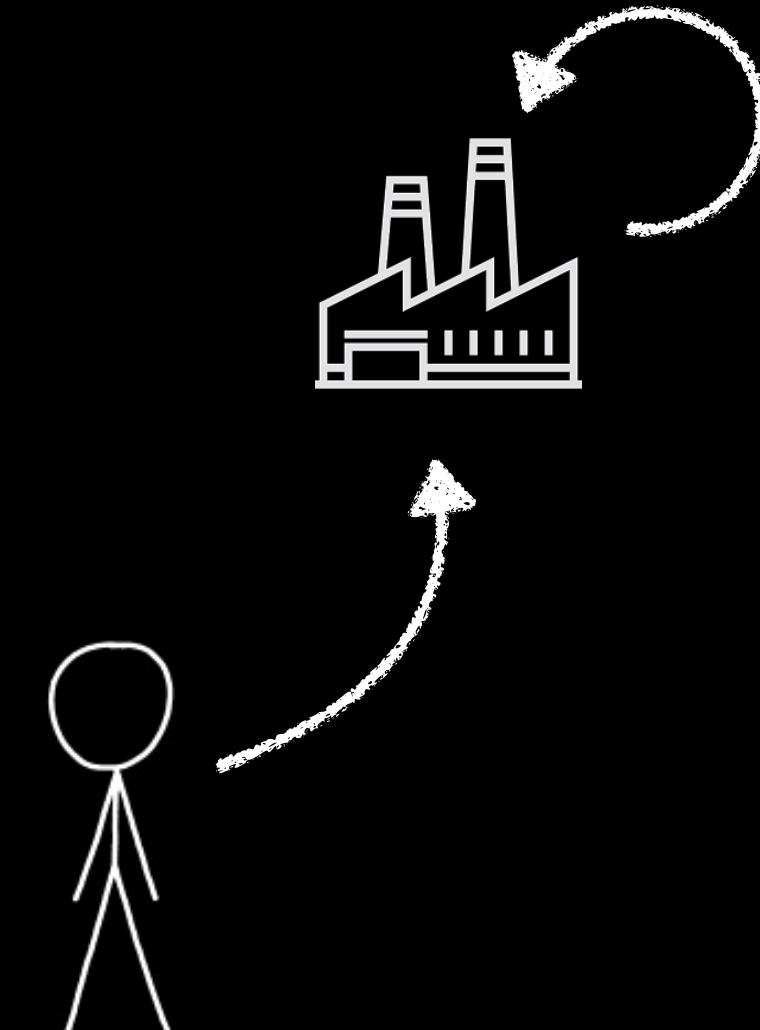
FEBRUARY 2017: GOGLERS LEARN ABOUT PROJECT DRAGONFLY

Members of the security and privacy team at Google are briefed on a secret project to build a state-censored search engine for China.

APRIL 2017: PROJECT MAVEN ANNOUNCEMENT

The U.S. Department of Defense announces the creation of an Algorithmic Warfare Cross-Functional Team, known as Project Maven, to develop AI technologies for drones.

JULY 2017: THE JAMES DAMORE MEMO



e.g., week 14 (future) solutions

://story.californiasunday.com/tech-revolt 67% ...

The Tech Revolt

Politics Comes to Silicon Valley: A Timeline

DECEMBER 2016: NEVERAGAIN.TECH

A group of engineers and activists brainstorm about how best to oppose incoming President Donald Trump's policies. They launch neveragain.tech, an online pledge whose signers promise never to cooperate with attempts to collect data on immigrants and Muslims for the purpose of racial profiling.

JANUARY 2017: SILICON VALLEY SECURITY GUARDS UNIONIZE

More than 3,000 security guards at Cisco, Facebook, Genentech, and other tech companies win union representation.

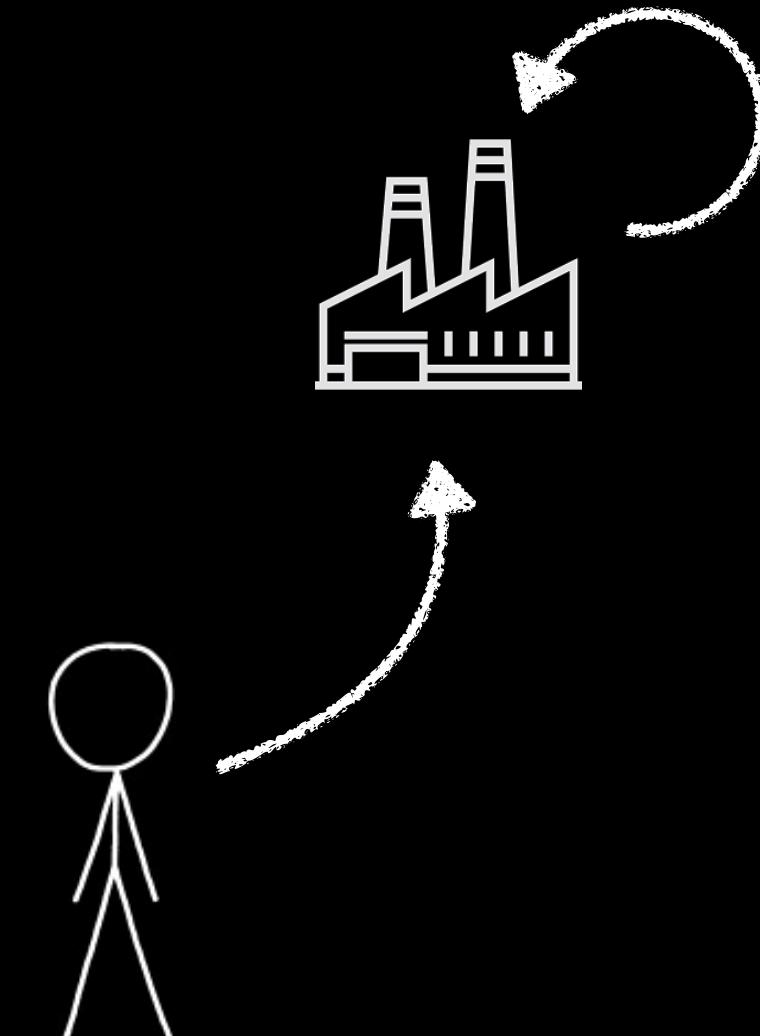
FEBRUARY 2017: GOGLERS LEARN ABOUT PROJECT DRAGONFLY

Members of the security and privacy team at Google are briefed on a secret project to build a state-censored search engine for China.

APRIL 2017: PROJECT MAVEN ANNOUNCEMENT

The U.S. Department of Defense announces the creation of an Algorithmic Warfare Cross-Functional Team, known as Project Maven, to develop AI technologies for drones.

JULY 2017: THE JAMES DAMORE MEMO



e.g., week 14 (future) solutions

AI Now Report 2018

Meredith Whittaker, AI Now Institute, New York University, Google Open Research

Kate Crawford, AI Now Institute, New York University, Microsoft Research

Roel Dobbe, AI Now Institute, New York University

Genevieve Fried, AI Now Institute, New York University

Elizabeth Kaziunas, AI Now Institute, New York University

Varoon Mathur, AI Now Institute, New York University

Sarah Myers West, AI Now Institute, New York University

Rashida Richardson, AI Now Institute, New York University

Jason Schultz, AI Now Institute, New York University School of Law

Oscar Schwartz, AI Now Institute, New York University

With research assistance from Alex Campolo and Gretchen Krueger (AI Now Institute, New York University)

DECEMBER 2018

e.g., week 14 (future) solutions

AI Now Report 2018

2. EMERGING SOLUTIONS IN 2018

2.1 Bias Busting and Formulas for Fairness: the *Limits* of Technological “Fixes”

Over the past year, we have seen growing consensus that AI systems perpetuate and amplify bias, and that computational methods are not inherently neutral and objective. This recognition comes in the wake of a string of examples, including evidence of bias in algorithmic pretrial risk assessments and hiring algorithms, and has been aided by the work of the Fairness, Accountability, and Transparency in Machine Learning community.¹¹⁷ The community has been at the center of an emerging body of academic research on AI-related bias and fairness, producing insights into the nature of these issues, along with methods aimed at remediating bias. These approaches are now being operationalized in industrial settings.

Meredith Broussard
Kate Crawford
Roel Goossens
Genevieve Hawn
Eliza Orlitzky
Varon Tilman
Sarah Walsh
Rashad Wright
Jason Zimring
Oscar Harzing

With
University
of
Massachusetts
Amherst

DECEMBER 2018

what should future
statisticians CEOs, and senators
know about the history and ethics of data?

- 0. preamble: class origin story
- 1. why history?
- 2. why ethics?
- 3. what we taught
- 4. what we learned

4. what we learned

4. what we learned

1. history + ethics:

4. what we learned

1. history + ethics:

how to integrate throughout a “tech” education

4. what we learned

1. history + ethics:

how to integrate throughout a “tech” education

2. draw parallels to today

4. what we learned

1. history + ethics:

how to integrate throughout a “tech” education

2. draw parallels to today

3. capabilities rearrange power

4. what we learned

1. history + ethics:

how to integrate throughout a “tech” education

2. draw parallels to today

3. capabilities rearrange power

4. story of “data” is story of truth+power

4. what we learned

1. history + ethics:

how to integrate throughout a “tech” education

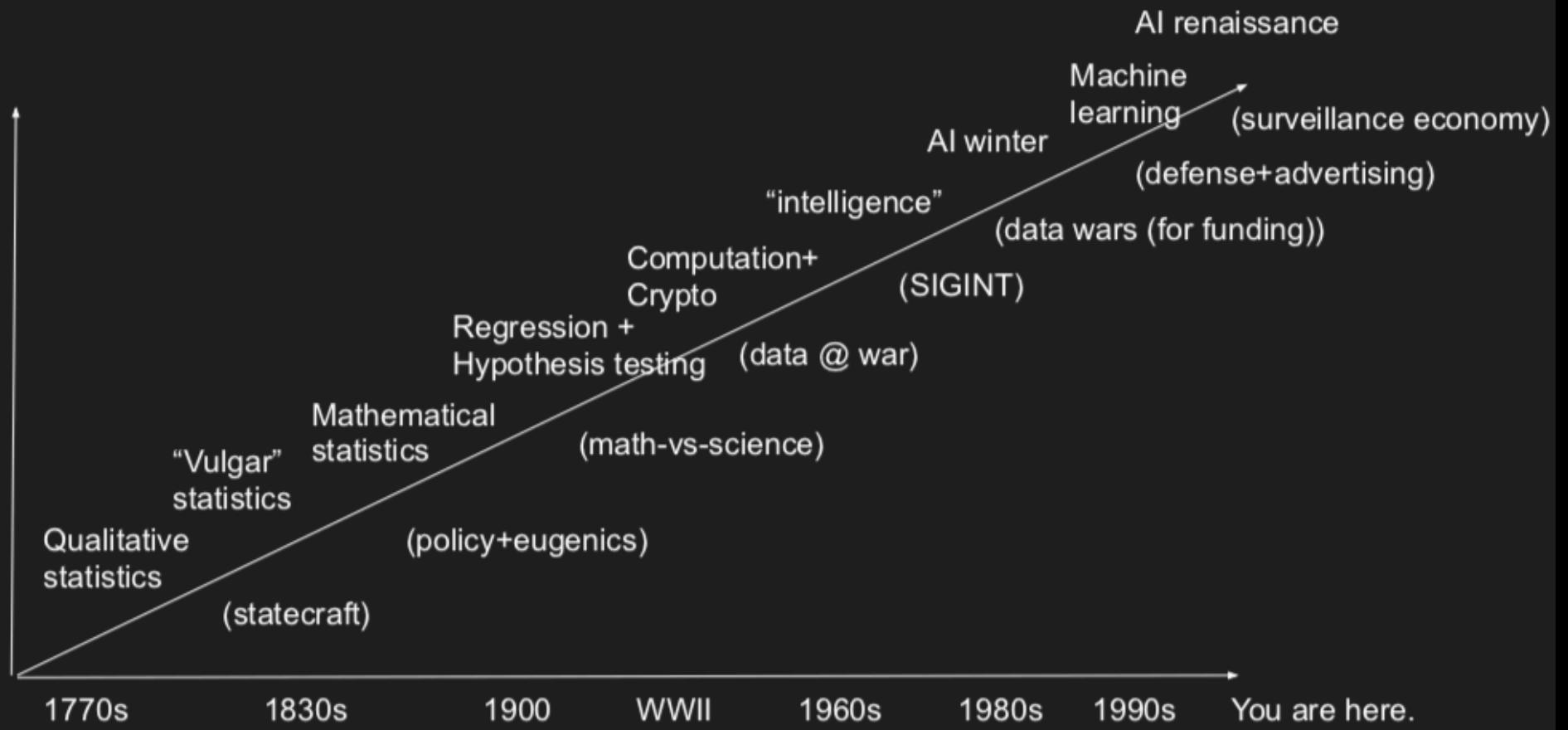
2. draw parallels to today

3. capabilities rearrange power

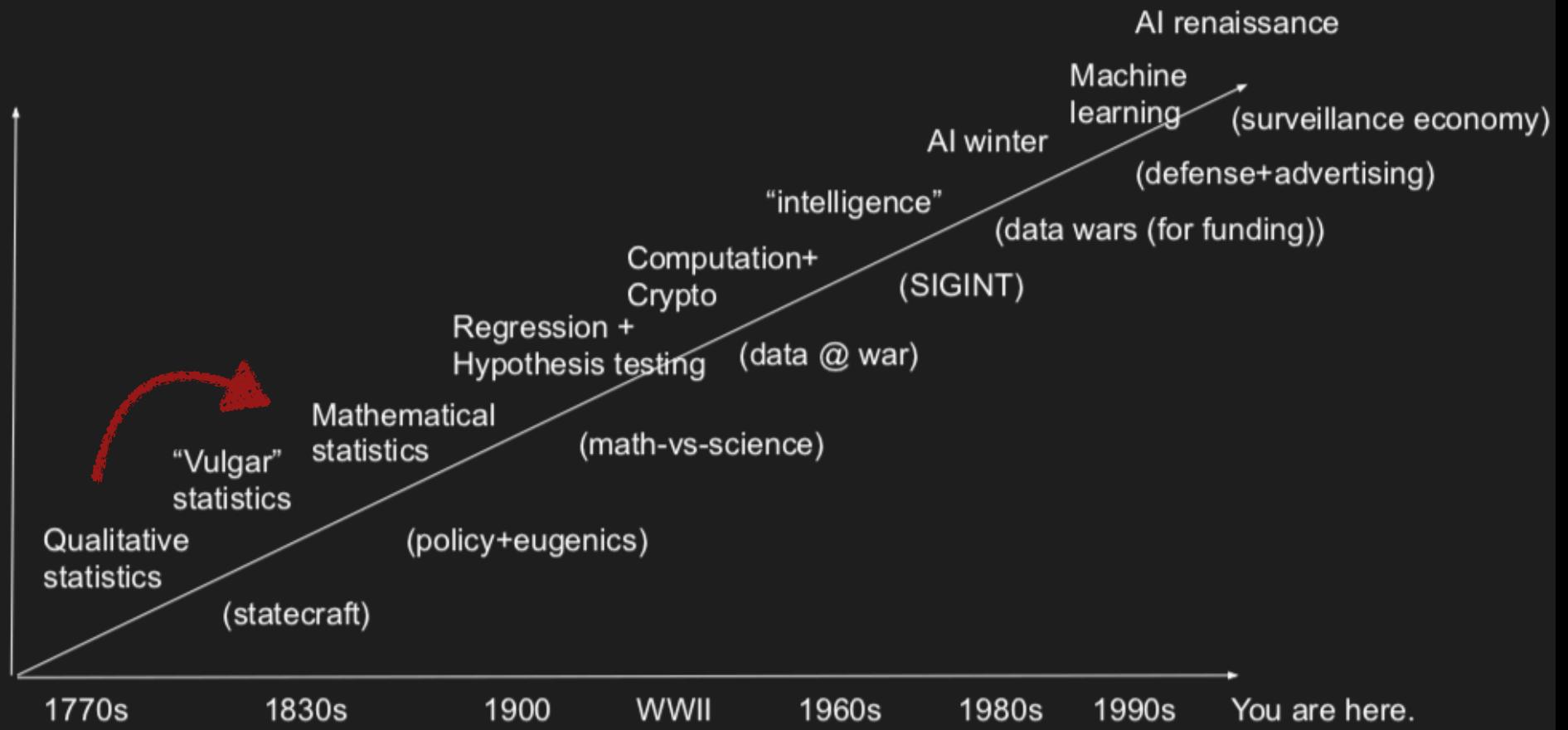
4. story of “data” is story of truth+power

- contested

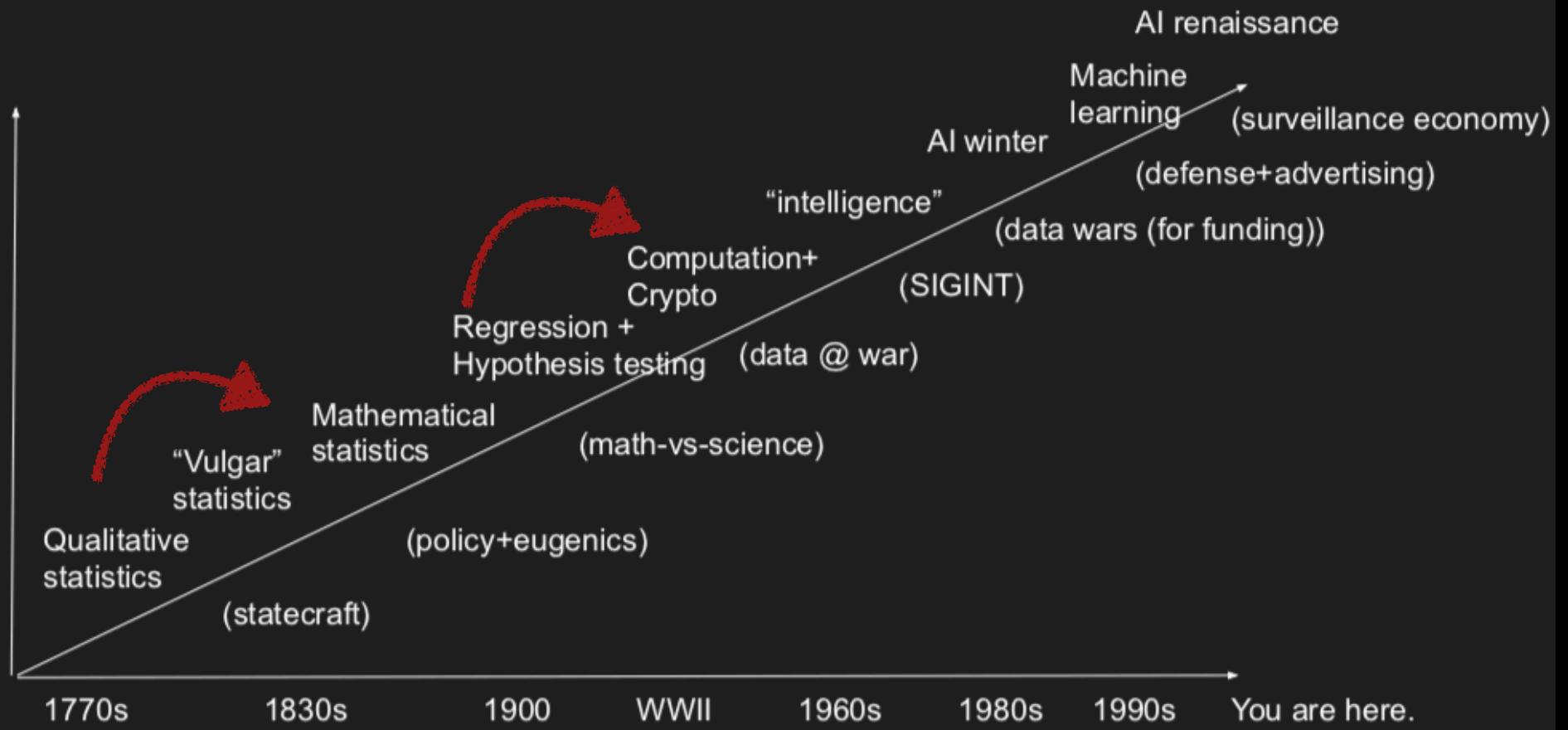
data 1770s-present: capabilities & intents



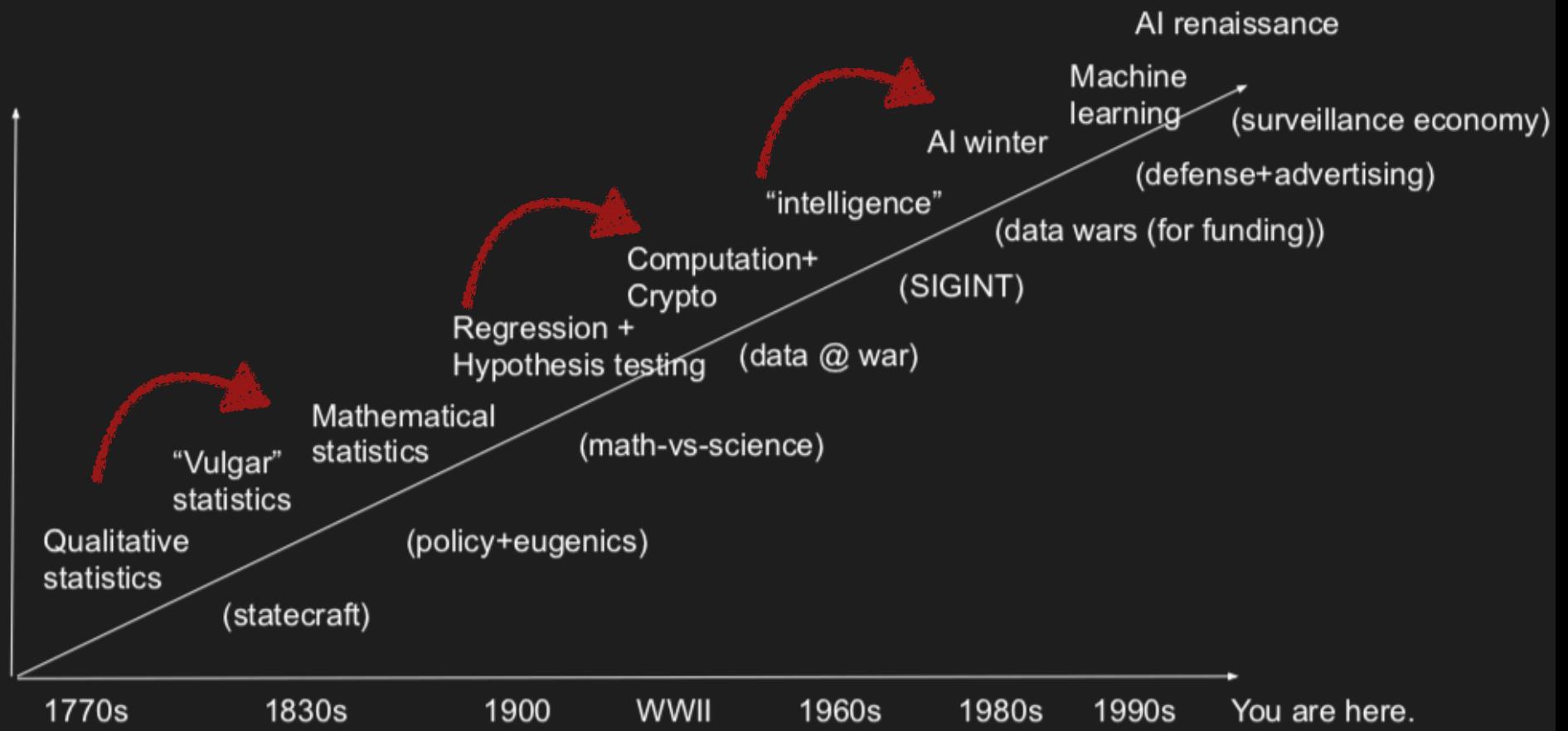
data 1770s-present: capabilities & intents



data 1770s-present: capabilities & intents



data 1770s-present: capabilities & intents



4. what we learned

1. history + ethics:

how to integrate throughout a “tech” education

2. draw parallels to today

3. capabilities rearrange power

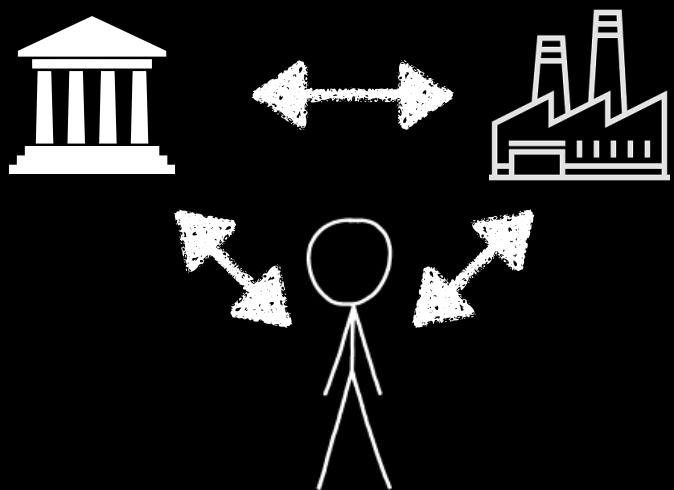
4. story of “data” is story of truth+power - contested

5. find the future by analyzing

- present contests

- present powers

what should future
statisticians CEOs, and senators
know about the history and ethics of data?



define & design

chris.wiggins@columbia.edu
@chrishwiggins
data-ppf.github.io