THE ETHICS OF DATA ANALYTICS

Kaiser Fung

Columbia University
MS in Applied Analytics Program

Jul 2015
Ashley Madison

Sep 2015
Volkswagen

Feb 2016
Apple v. FBI

Jun 2014
Facebook Research

Aug 2015
Google Search

Oct 2015
Theranos
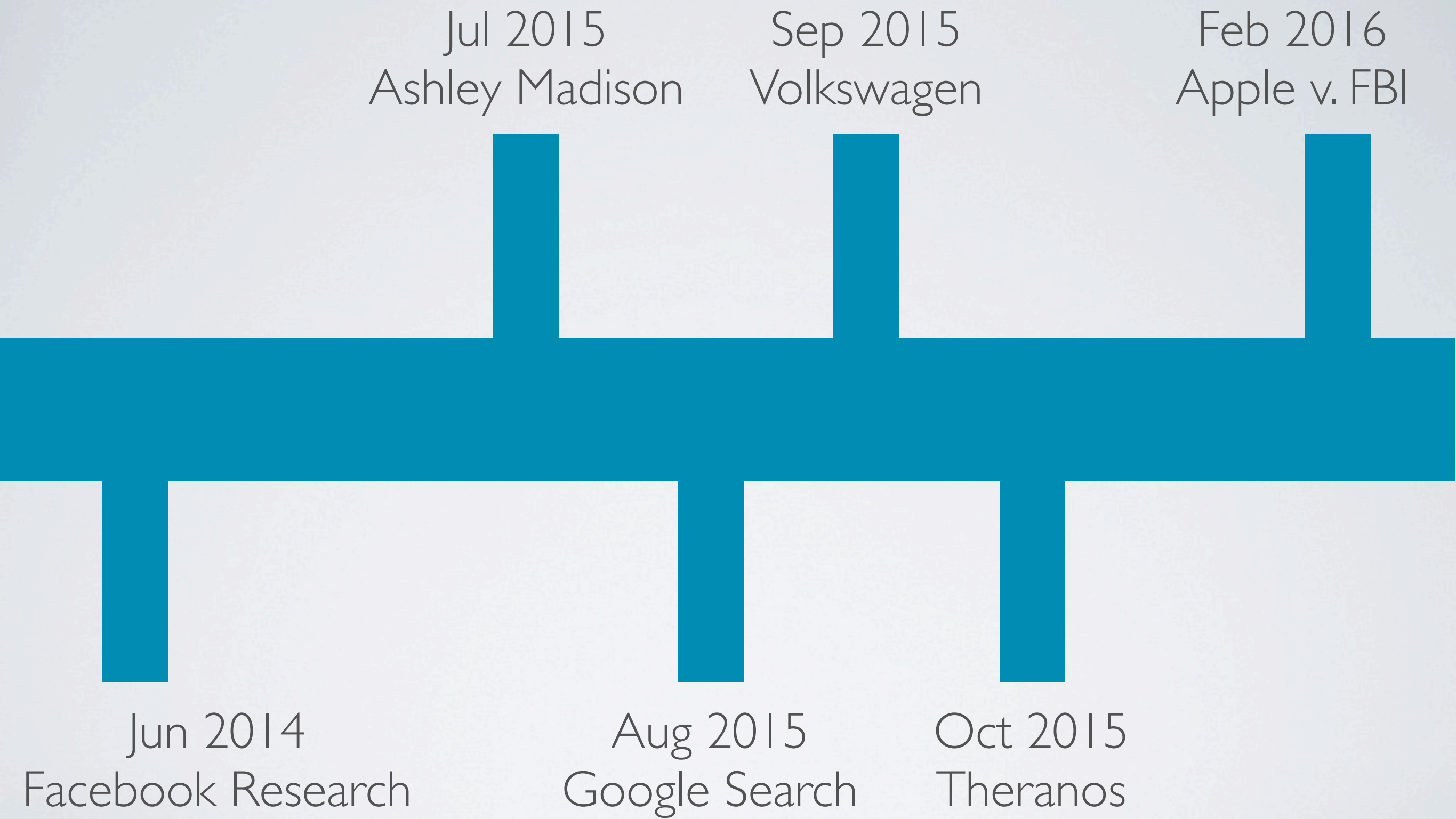
ASHLEY MADISON®

Life is short. Have an affair.®

Get started by telling us your relationship status:

Please Select ▾

See Your Matches »

Over 44,300,000 anonymous members!

★★★★★
100%
Like-minded
People

As seen on: Hannity, Howard Stern, TIME, BusinessWeek, Sports Illustrated, Maxim, USA Today

Register on Ashley Madison   Affiliate Program   P

Fo

Location

Ashley Madison is the most famous name in infidelity and married dating. As seen on Hannity, Howard Stern, TIME, BusinessWeek, Sports Illustrated, Maxim, USA Today. Ashley

- An unethical business
- Hackers stole all of their **user data**
- Stolen data included those users who paid a special fee for "full delete"
- Users' credit account details were attached to their profiles, even if they used pseudonyms
- Journalists and data scientists rushed to download the stolen data, seeking salacious stories about public figures

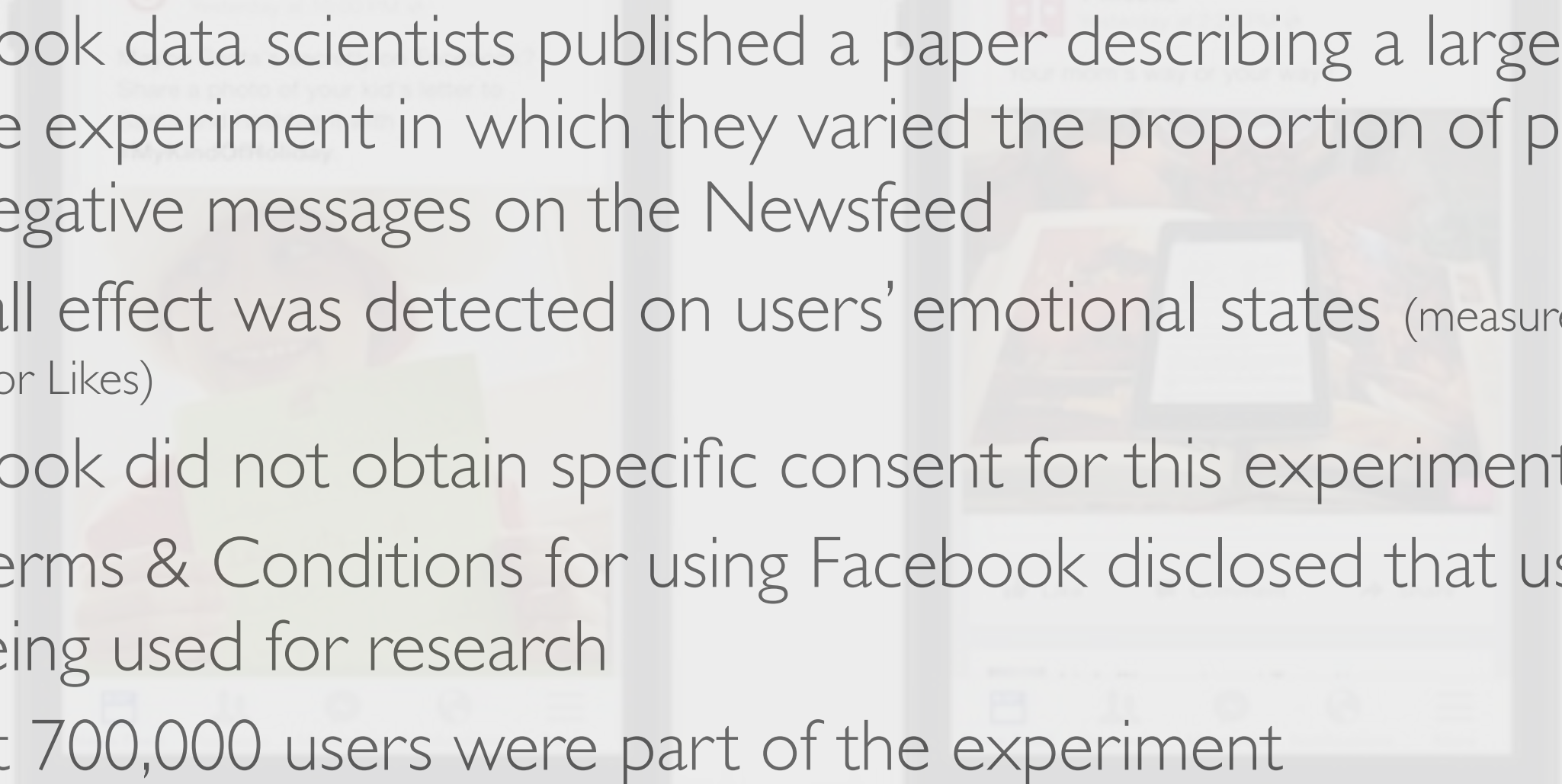# Meet the Volkswagen TDI Clean Diesel Family.

- EPA disclosed that Volkswagen uses "defeat devices" in diesel engines to cheat NOx emissions testing
- Software uses **sensor data** to detect emissions testing, and activate a compliant mode that lowers emissions
- In real conditions, cars emit 40x the allowed amount
- Installed in 11 million cars from 2009 to 2015 model years
- Lots of teams must have been involved

KAISER FUNG . 4/12/2016

theranos

MD CONNECT   OUR LAB   PROVIDERS   TESTS   CENTERS   COMPANY   NEWS

the lab test, reinvented.

We believe the future of health care lies in greater access for the individual. So we built a better lab experience with access in mind, making it easier than ever for you to engage with your health early and at the time it matters most.

THERANOS NEWS

- A Silicon Valley "unicorn" valued at $10 billion, selling finger-prick blood testing at a fraction of the cost
- Refused to publish data in peer-reviewed journals
- New test allegedly inaccurate
- Did not have FDA approval for 199 out of 200 tests offered
- Most of those tests were performed using standard technology so the only difference was lower prices

- Facebook data scientists published a paper describing a large-scale, on-line experiment in which they varied the proportion of positive and negative messages on the Newsfeed

- A small effect was detected on users' emotional states (measured by status changes or Likes)

- Facebook did not obtain specific consent for this experiment

- The Terms & Conditions for using Facebook disclosed that user data are being used for research

- About 700,000 users were part of the experiment

- Researchers claimed they proved a "search engine manipulation effect" *(but see Gelman & Fung commentary)*
- They published articles claiming that Google *could* manipulate elections by filtering search engine results
- One of their experiments involved manipulating a real election in India. They said an IRB approved the experiment on the basis that the sample size of the experiment was tiny relative to the electorate
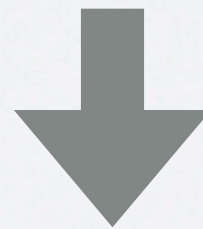
WHY SHOULD WE CARE ABOUT ETHICS?
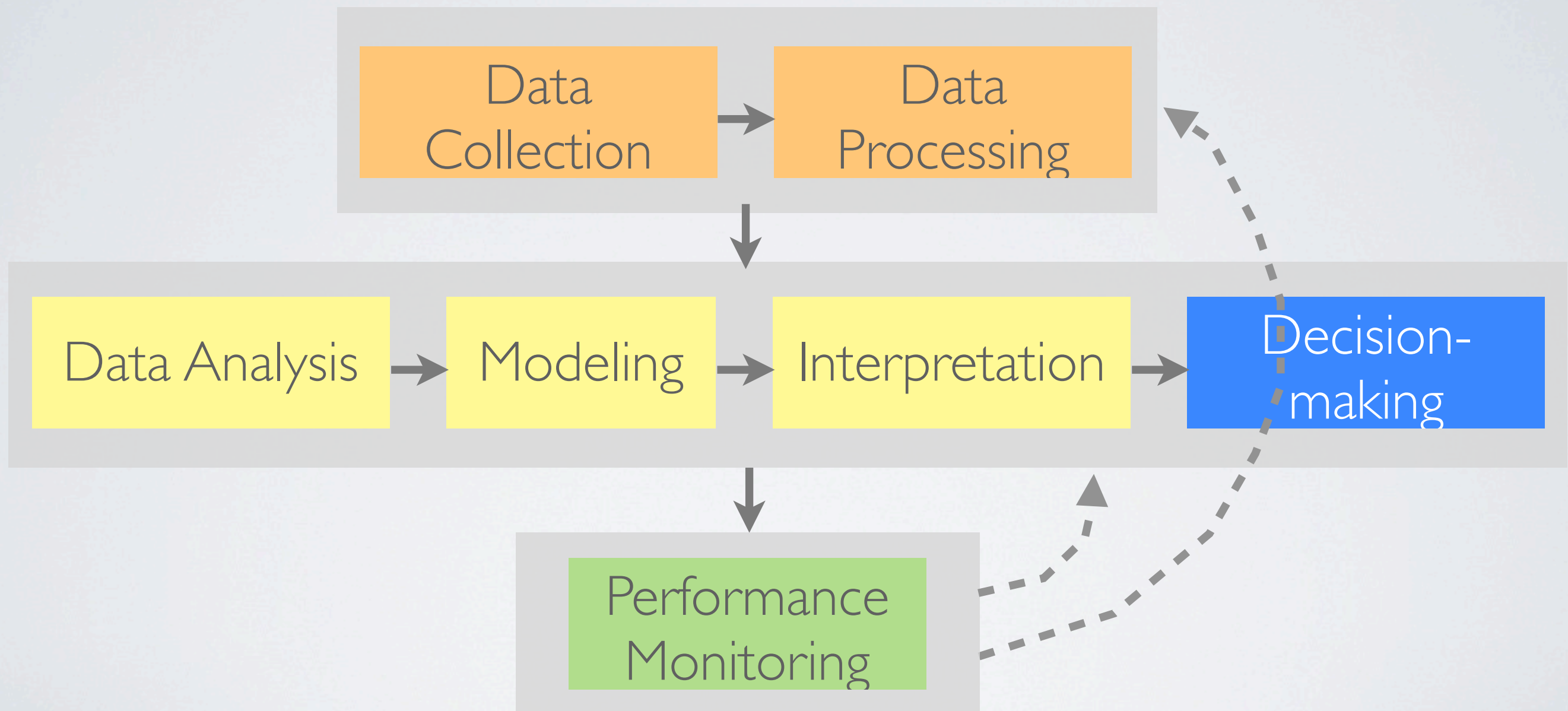
Data analytics is a form of engineering
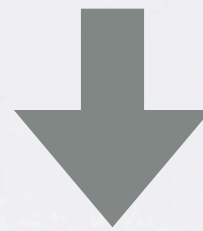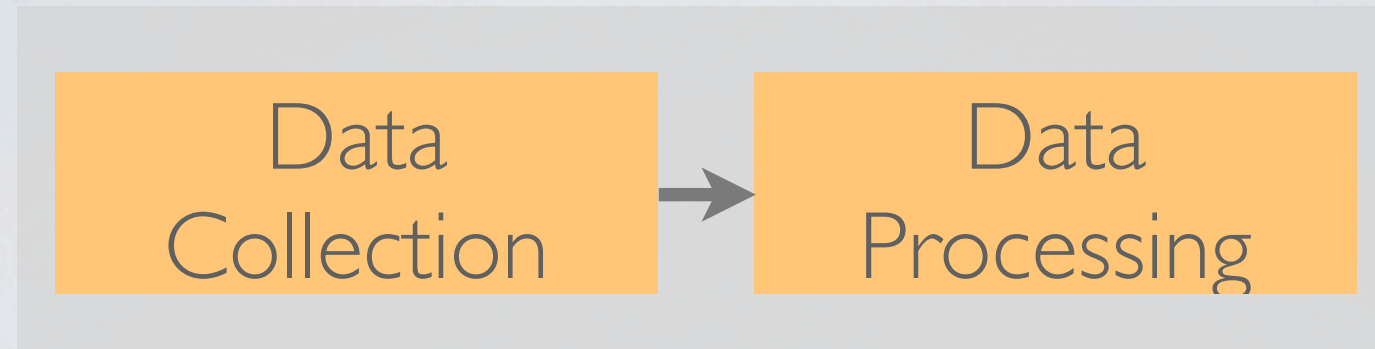
Engineers aspire to make the world better

unlike scientists

Changing the world sometimes makes it worse

DAMNED LIES & THE ANALYTIC PROCESS

| Data Collection | → | Data Processing |

What data should (not) be collected

Who owns the data

Whose data can (not) be shared

What technology for collecting, storing, managing data

Whose data can (not) be traded

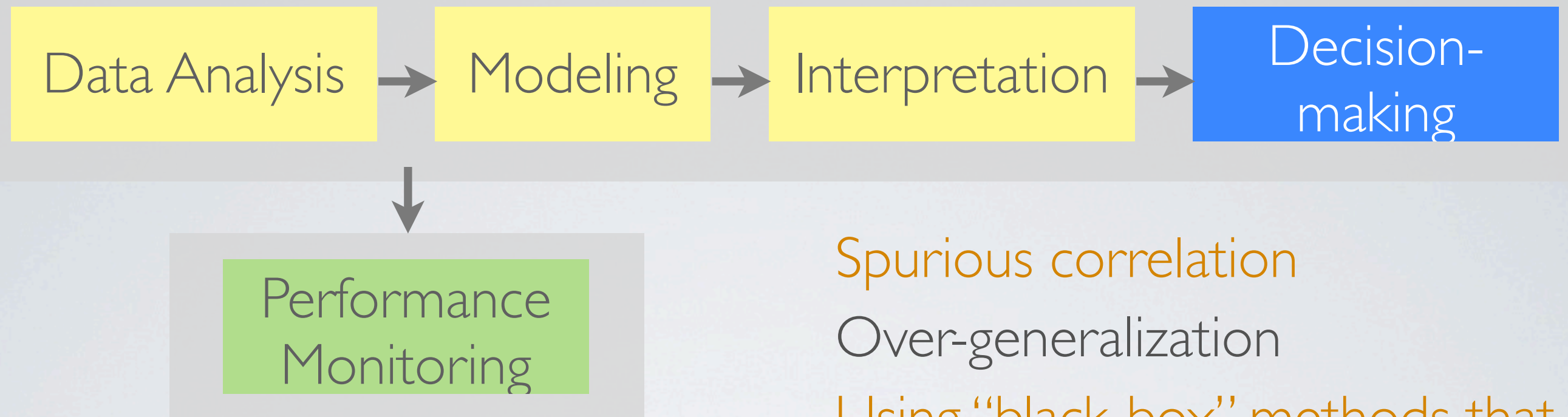What data can (not) be merged

What to do with prejudicial data

Is there a right to be forgotten

How to deal with inaccurate data

How to deal with gaming of data

Should data serve individual or business interests

Can data of friends be treated differently from data of strangers

Data Analysis → Modeling → Interpretation → Decision-making

Performance Monitoring

Data is biased (known/unknown)
Invalid assumptions
Confirmation bias
Publication bias
Misinterpreting p-values
Badly imputing or not imputing missing values

Spurious correlation
Over-generalization
Using "black-box" methods that cannot be explained
Using heuristics that are not well understood
Releasing untested code
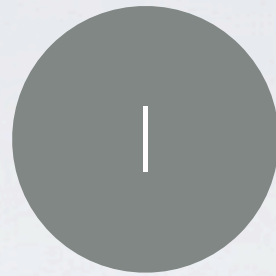Profiling applications
Extrapolating
Not back-testing
Not measuring in-market performance

THE ETHICS OF DATA ANALYTICS

When the statistics lie, is it

Ignorance v. Ignoring?

MAJOR ETHICAL PHILOSOPHIES

I

An action is ethical if it leads to good **consequences**

The only consequence that matters is
the greatest happiness for the most people
(*utilitarianism*)

Everyone is free to strive for personal happiness so long as one
does not deny/hinder the pleasure of others
(*principle of freedom*)

MAJOR ETHICAL PHILOSOPHIES

2

An action is ethical if it aligns with moral **values**

Always act out of duty: follow rules

Means cannot justify ends

MAJOR ETHICAL PHILOSOPHIES

3

An action reflects the moral character of the actor

Virtues such as creativity, honesty, professionalism, strive for quality, etc.

KEY QUESTIONS

Does ethics refer to the action or the actor?

Which actors are responsible?

Who decides what is ethical?

Which parties' interests should be considered?

How to compromise between conflicting values?

FREQUENTLY RAISED OBJECTIONS

I am not doing anything illegal

Nothing bad happened

No one predicted the bad outcome

Someone else is already doing this

MORE OBJECTIONS

Customers can opt out if they don't like it

Customers have agreed to Terms & Conditions
(informed consent)

Low stakes, therefore it doesn't matter

I have to do it to keep my job

ETHICAL PROGRESS

Awareness

Appreciation

Debate

Consensus

Change

KAISER FUNG . 4/12/2016

References

Ashley Madison: Williams, "Ashley Madison hack: the depressing rise of the 'moral' hacker," The Telegraph (UK), 8/20/2015.

Volkswagen: Hotten, "Volkswagen: the scandal explained," BBC News, 12/10/2015.

Theranos: John Carreyrou, "Hot Startup Theranos has struggled with its blood-test technology," WSJ, 10/16/2015; many other articles.

Facebook: Kramer, et. al., "Experimental evidence of massive-scale emotional contagion through social networks", PNAS, 2014.

Google: Epstein & Robertson, "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections", PNAS, 2015.

Google commentary: Gelman & Fung, "Could Google rig the 2016 election? Don't believe the hype," The Daily Beast, 9/21/2015.

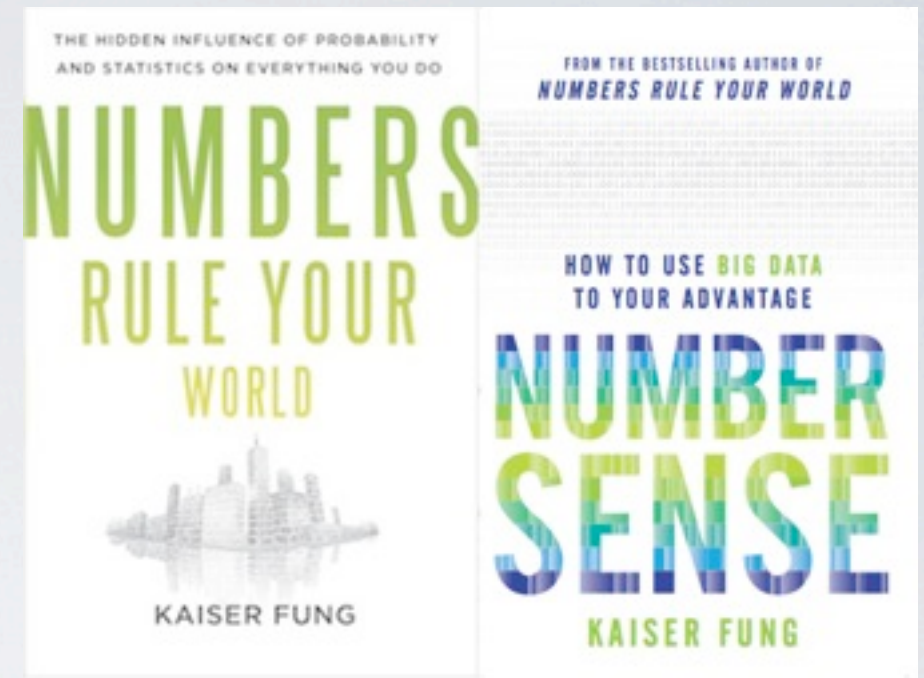http://www.thedailybeast.com/articles/2015/09/21/could-google-rig-the-2016-election-don-t-believe-the-hype.html

Kaiser Fung

http://kaiserfung.com
Twitter @junkcharts
Blog junkcharts.typepad.com
Inquiries junkcharts@gmail.com

Current courses:

Statistical Reasoning & Numbersense
rSQUAREedge
Applied Analytical Methods and Frameworks
MS in Applied Analytics, Columbia University