Hearing #6 on Competition and Consumer Protection in the 21st Century

American University

Washington College of Law

November 6, 2018

Welcome

We Will Be Starting Shortly

Welcome

Daniel Gilman

Federal Trade Commission

Office of Policy Planning

Welcome and Introductory Remarks

Jonathan Baker

American University

Washington College of Law

A Moment of Reckoning on Big Data

Ginger Zhe Jin, U of Maryland & NBER

What’s going on in the marketplace?

What’s going on in the marketplace?

- Who generates what data?
- Who uses which data for what?
- Where and how does data stay, flow and evolve?
- How does technology reshape data and data use?
- Who benefits, who loses from certain data practice?
- What is the aggregate consequence of data use in the short run and the long run?
- What is known and what is not known, to whom and when?

Where does the market fail?

Potential market failures:

- Market power
- Information asymmetry
- Externality
- Bounded rationality

Whether and how does big data contribute to these market failures?

How does data contribute to market power?

Market power

- Barrier to entry?
- Facilitate collusion?
- Facilitate anti-competitive contracts?
- Facilitate perfect price discrimination?
- Merger efficiency?
- Contract efficiency?

Overall impact on consumer welfare (short run and long run)

How does data contribute to information asymmetry?

Information asymmetry

- Information about data
  - Before the focal transaction
  - Right after the focal transaction
  - Long after the focal transaction
  - Content and format of data
  - Relationship across datasets

Consumer

Firm

Affiliate

Non-affiliate

Black-market players

The public

What is the harm to consumer welfare? Where and how much?

How does data contribute to externality?

Externality

- What data practice generates what spillover?
- Spillover to whom, at what magnitude?
- When is the spillover observable and consequential?
- Positive or negative spillovers?
- Does the party that generates the spillover have an incentive to internalize the spillover?

How does the spillover affect consumer welfare?

How does data contribute to bounded rationality?

Bounded rationality

- Who has bounded rationality in understanding his role in big data?
- Which party has more bounded rationality than other parties?
- Who suffers from bounded rationality?
- Who has what incentives to exploit other parties' bounded rationality?

How does this affect consumer welfare?

Potential solutions

Free market

Industry self regulation

Industry self regulation + consumer education + societal monitoring

Ex post enforcement:

driven by observed outcomes

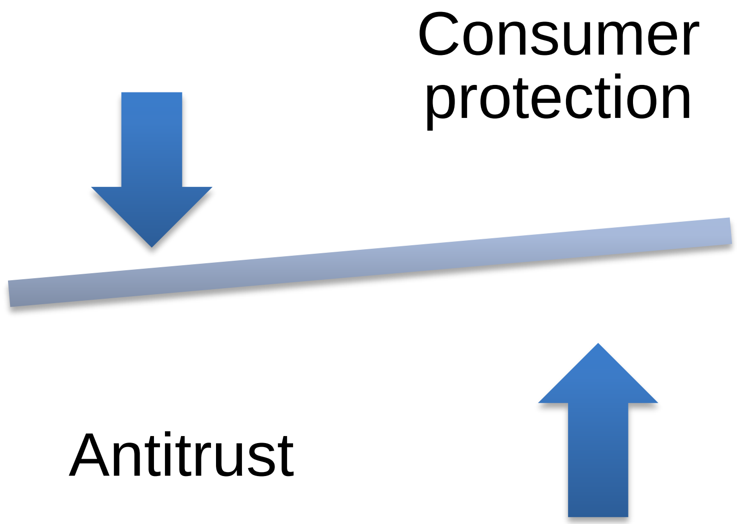
Ex ante regulation:

1. Information disclosure
2. Procedural actions
3. Mini quality standard

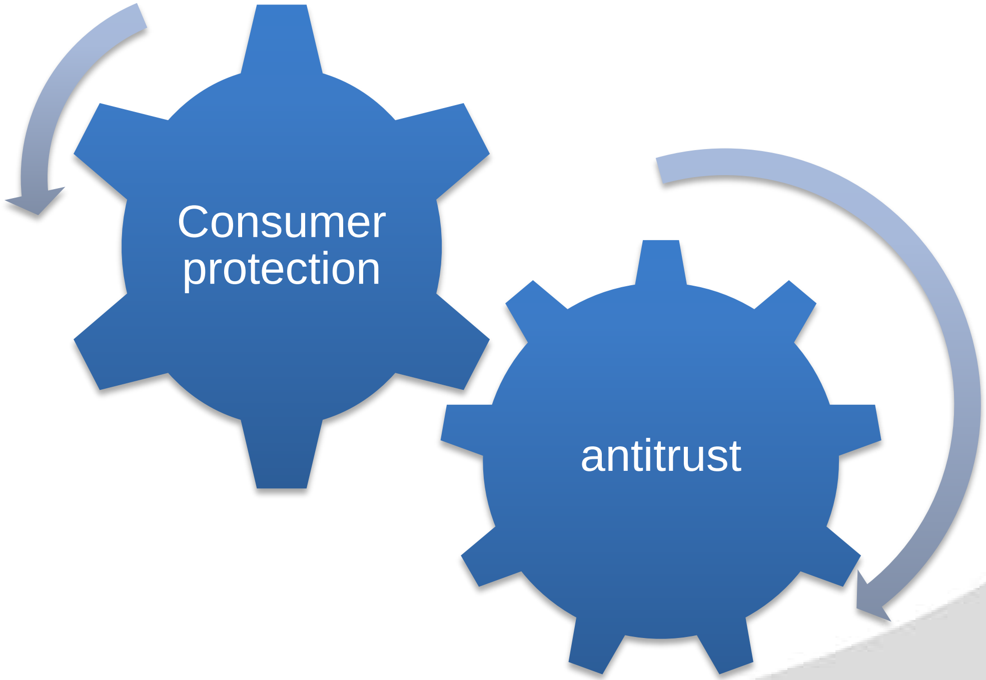
Ex ante regulation + Ex post enforcement

Existing tools:
competition and consumer protection

- How do they fit in the overall framework?
- What is the relationship between the two tools?



or



How to carry out the solution?

- Comprehensive legislation or leave details to the regulatory/enforcement agency?
- Who is the regulatory/enforcement agency?
  - One or multiple agencies?
  - One or multiple levels (federal, state, industry-specific)?
  - Degree of enforcement and/or regulatory freedom
  - Resources and expertise to fulfill the function
  - Limit the agency's power:
    - who to report to? transparency?
    - Accountability?
    - External forces to spot and correct wrongdoing?

How do <u>other parties</u> contribute to the solution in an <u>on-going</u> basis?

- Independent research institutions
- Industrial associations
- Consumer advocacy groups
- Individual firms
- Individual consumers
- Other government agencies

International complications

- EU, US and China
  - EU: GDPR + DG-comp + country-specific enforcement
  - US: patchwork of federal, state and industry-specific
  - China: nationwide laws, government censorship and surveillance

- Many data-intensive firms are global
  - Different regimes imply different compliance cost
  - Data, ideas, talents and money flow globally

Break

9:45-10:00 am

The Economics of Big Data and Personal Information

Session moderated by:

Jeremy Sandford

Federal Trade Commission

Bureau of Economics

The Economics of Big Data and Personal Information

Alessandro Acquisti

Carnegie Mellon University

Heinz College

The Economics of Big Data and Personal Information

Omri Ben-Shahar

University of Chicago

Law School

The Economics of Big Data and Personal Information: The Economics of Data Regulation

Liad Wagman

Illinois Institute of Technology

Stuart School of Business

The Short-Term Effects of GDPR on Technology Venture Investment

Liad Wagman
Stuart School of Business
Illinois Institute of Technology

Joint work with:

Jian Jia
Illinois Institute of Technology

Ginger Jin
University of Maryland

Downstream Data Trade

Consider a market with many upstream firms (e.g., banks) and upstream firms (e.g., insurance providers)

- Upstream firms screen applicants for (e.g., financial) products
- Information about applicants CAN or CANNOT be traded downstream

If data is permitted to flow downstream:

- Lower prices for upstream product (e.g., mortgages)
- More screening of applicants and subsequent fewer defaults
- Consumer surplus increases

Kim & Wagman (2015)

Downstream Data Trade

Validated using data from local opt-in/opt-out ordinances:

Opt Out	Opt In
More data is collected	Less data is collected
More efficient matching between borrowers and loans	Less efficient matching between borrowers and loans
Lower mortgage prices	Higher mortgage prices

Economic Fundamentals

1. Horizontally differentiated duopoly (Hotelling)
2. Horizontally differentiated oligopoly w/ entry (Circular City)
3. Vertically differentiated duopoly
4. Horizontally & vertically differentiated duopoly

Suppose firms have detailed records about consumer preferences. What happens when access to data is cut off?

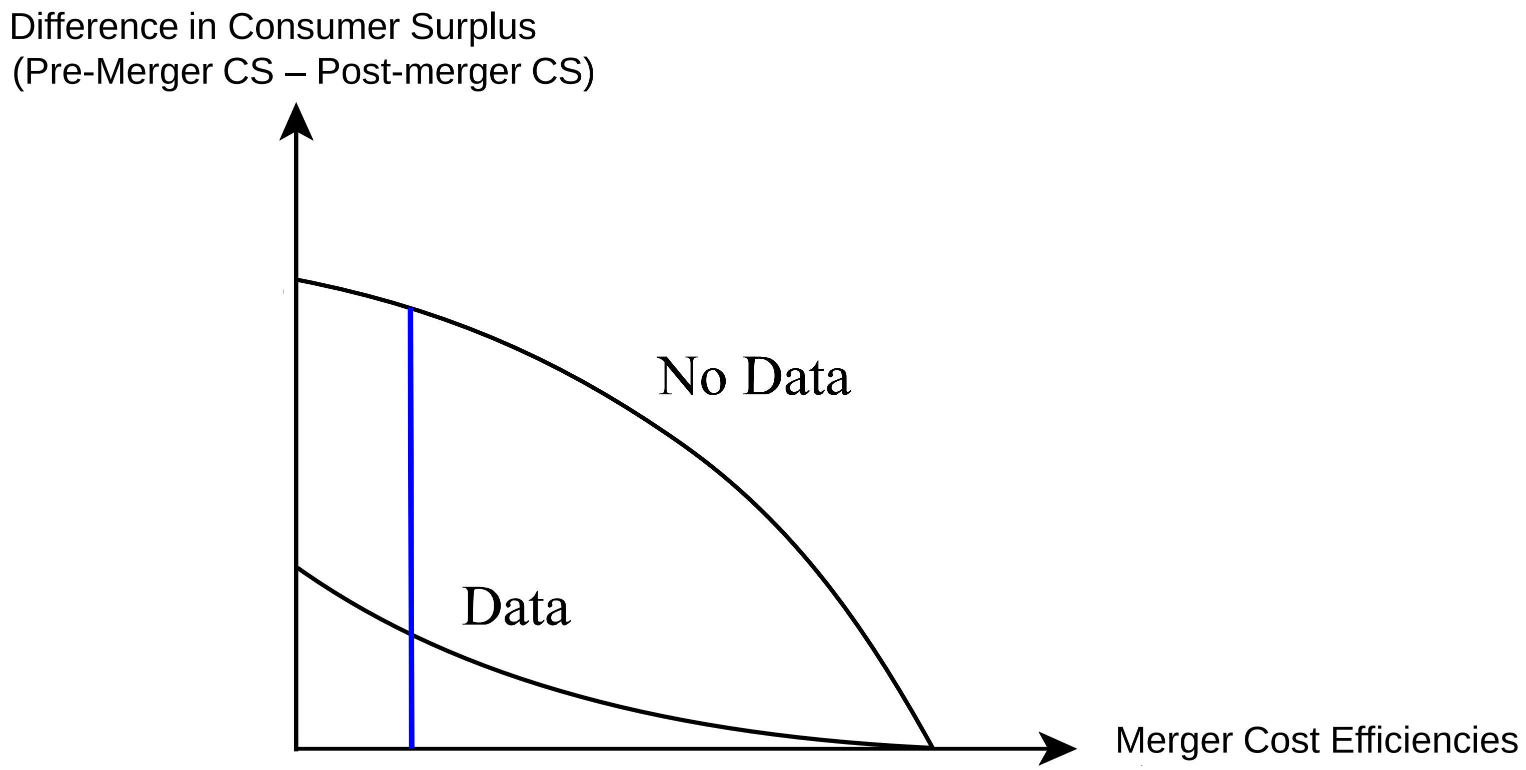
Restricting Data Access
A Price Discrimination Perspective:

Outcome of Imposing Data Restrictions	Model 1	Model 2	Model 3	Model 4
Consumer Surplus	Lower	Lower	Lower	Higher
Total Industry Profits	Higher	Same	Higher	Lower
Overall Welfare	Same	Lower	Lower	Mixed
Consumers Prefer Data Restrictions	None	Mixed	Mixed	Mixed

Taylor & Wagman (2014)

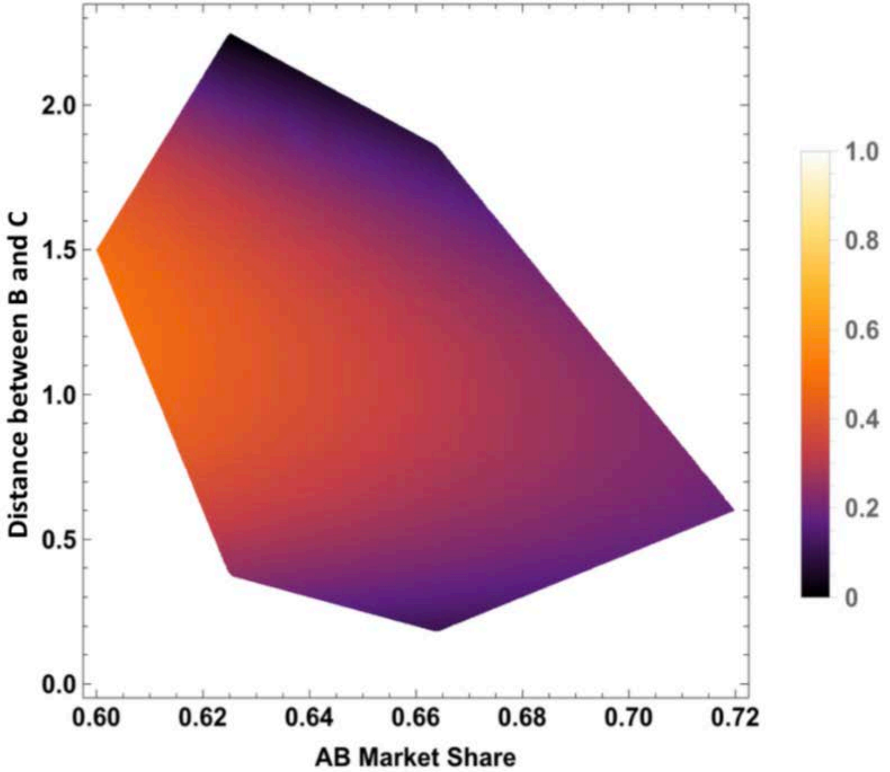
Extended:
Data restrictions impact merger considerations

Mergers when Firms Do/Do Not Have Data

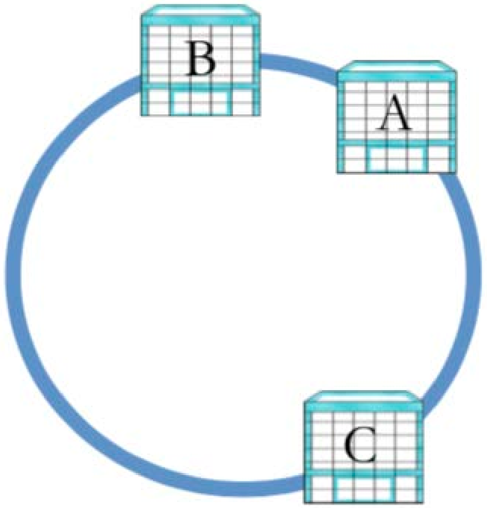


Kim, Wickelgren and Wagman (2018)

Mergers when Firms Do/Do Not Have Data



Heat map depicts $\frac{\Delta CS \text{ w/ data}}{\Delta CS \text{ w/o data}}$, darker → larger ΔCS gap

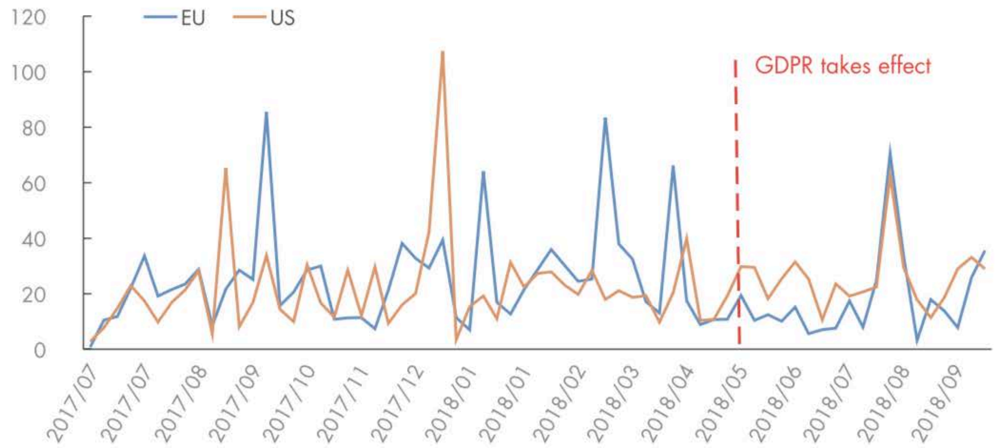


Continues to Hold even when firms have asymmetric access to data, or when upstream data brokers can sell data exclusively downstream, as long as weaker firms can survive

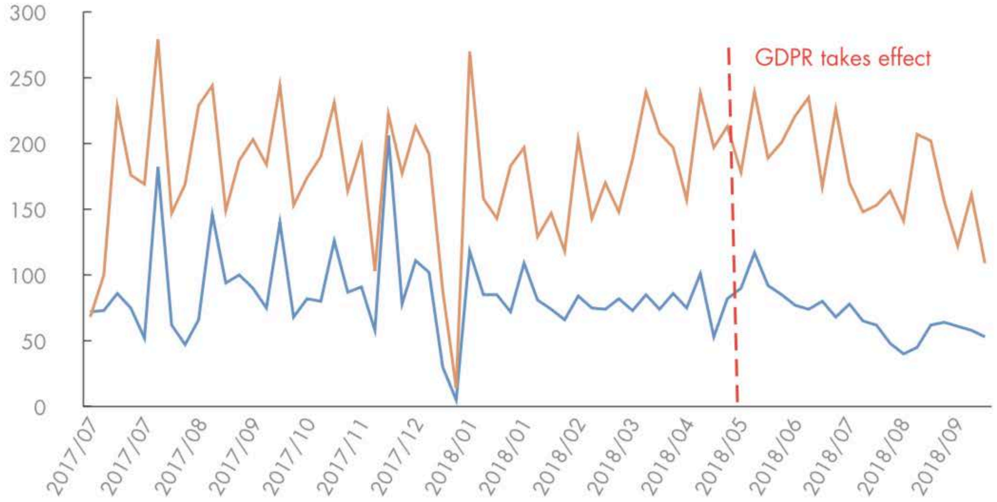
Depending on specific market structure, less restrictive data access can actually make otherwise contested mergers less contested (darker-shaded areas of figure)

Short-Term Effect of GDPR on Investment

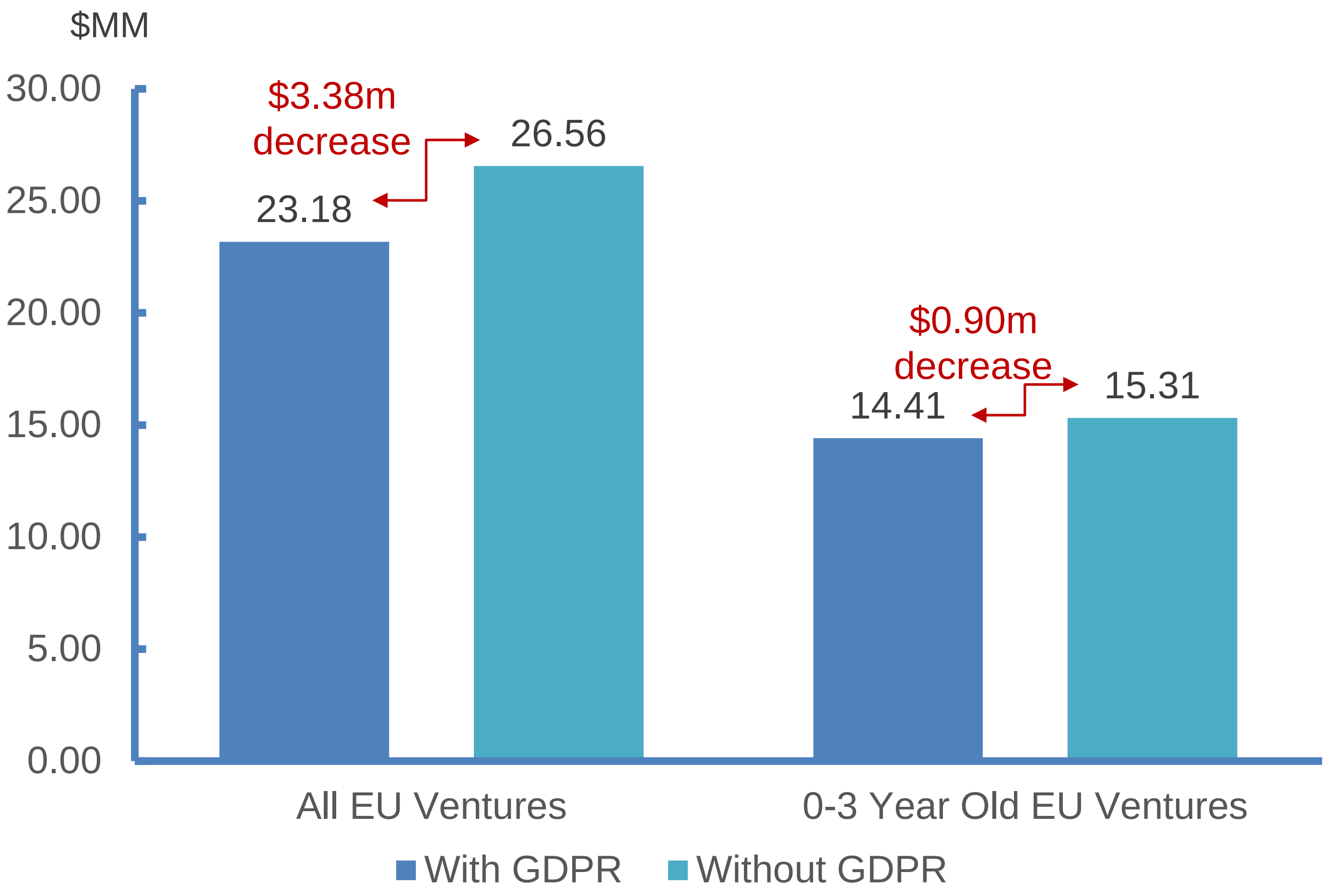
a. Average $MM amount per deal at weekly frequency



b. Average total number of weekly deals



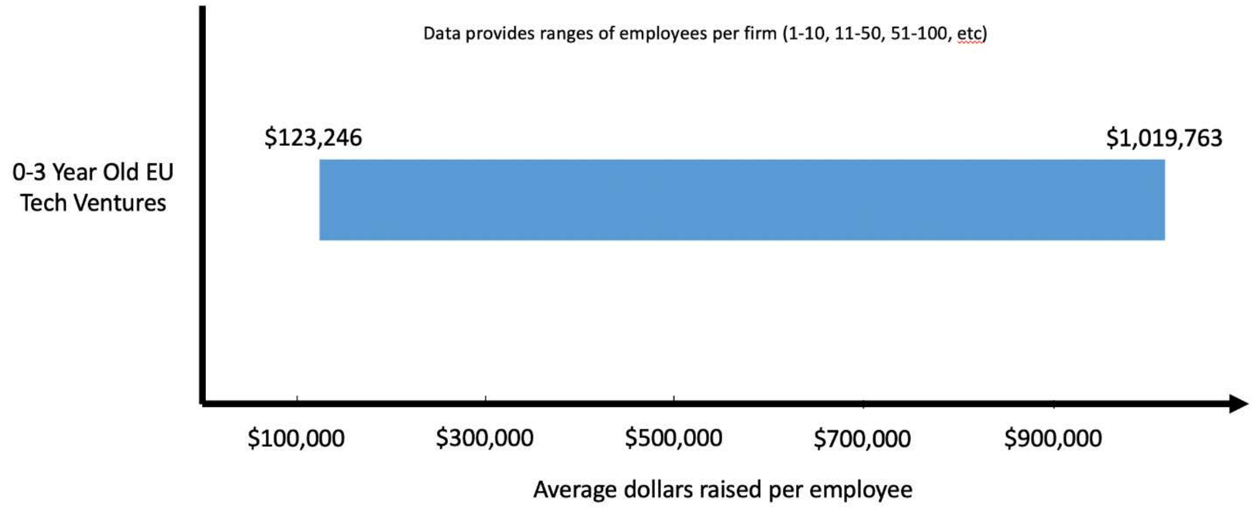
GDPR Effect on $MM Raised Per Week Per Member State Per Tech Category (Average EU)



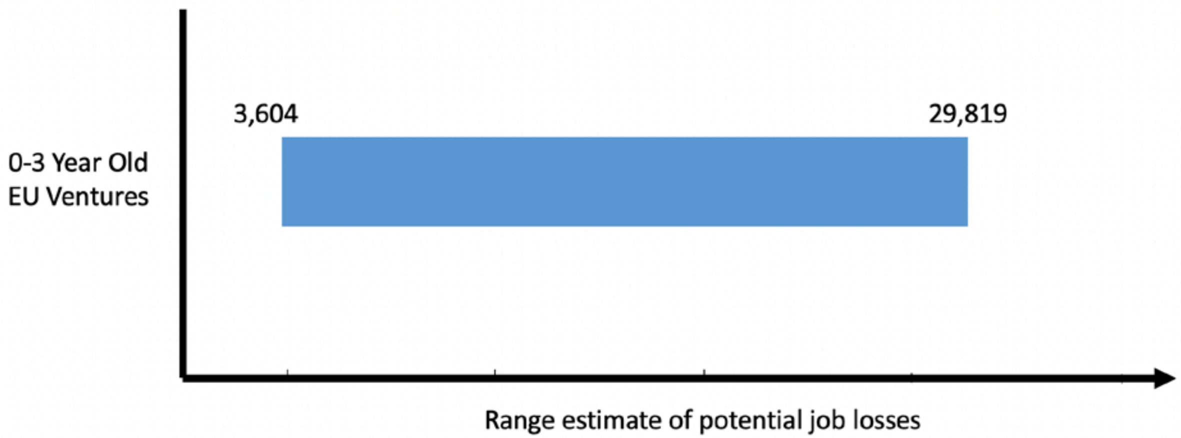
Jia, Jin and Wagman (2018)

GDPR and Technology Jobs

Average $ Raised Per EU Tech Employee



Rough Bound Estimates of EU Tech Jobs Lost



- Could be indicative of wait-and-see approach (only observe short term)
- Preliminary, back-of-the-envelope calculation
- 4.09%-11.20% of overall 0-3 year old venture tech jobs in the EU in our sample

Other Concerns

Campbell, Goldfarb and Tucker (2015): Identical compliance costs disproportionately burden smaller firms

Krasteva, Sharma and Wagman (2015): Compliance costs both deter innovation and shift some of it into established firms

Lessons Observed

- Regulatory approach should embrace nuance, dynamism, and be market specific. A blanket approach is likely to be inefficient.
- Strike a balance b/w data usability and data security
  - Data privacy as a means for data security seems intuitive but a proper balance is needed
- Incorporate data considerations into merger review
- Seek to avoid burdening smaller ventures with disproportionate costs of compliance

The Economics of Big Data and Personal Information: The Economics of Data Regulation

Florian Zettelmeyer
Northwestern University
Kellogg School of Management

KEY POINT

Firms are increasingly adopting machine learning for advertising, pricing, promotions, inventory optimization, etc.

These high-dimensional targeting algorithms create strong selection effects.

We expect the increase use of machine learning to severely limit the use of traditional non-experimental methods for measurement.

Illustration for today:

"Does great data + observational (non-experimental) methods accurately measure advertising effects?"

Source: Gordon, Zettelmeyer, Bhargava, Chapsky (2016): "A Comparison of White Paper, Kellogg School of Management, Northwestern University No data contained PII that could identify consumers or advertisers to maintain privacy. Based upon data from 12 US advertising lift studies. The studies were not chosen to be representative of all Facebook advertising.

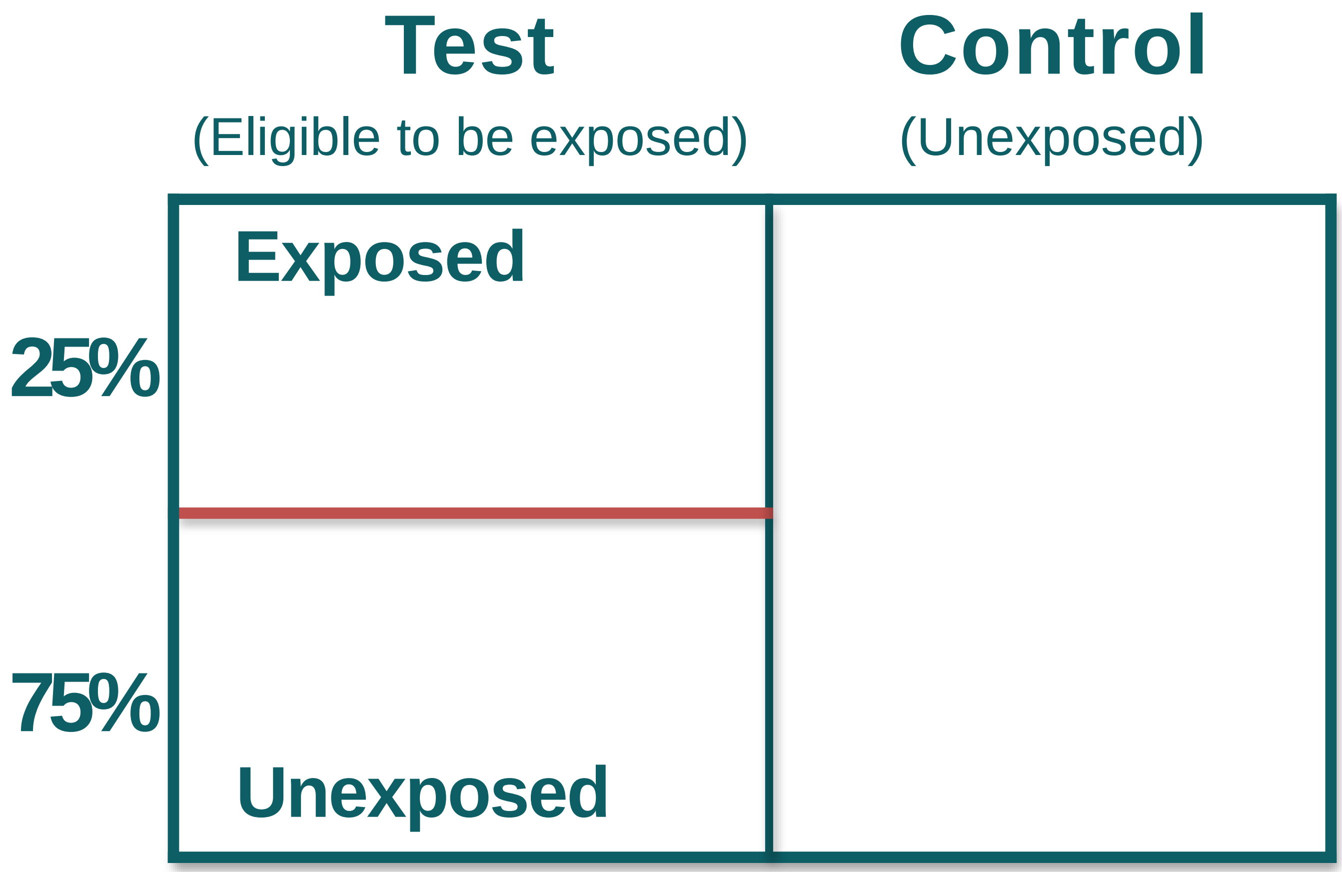
We teamed up with Facebook to answer this question

- 15 large-scale RCTs across verticals – FB "Lift Test" product
  - Note: not chosen to be representative of all Facebook advertising
- Statistical power
  - Between 2 million and 150 million users per experiment
  - 1.4 billion total ad impressions
- Single-user login
  - Captures cross-device activity
  - Eliminates issues with cookie-based measurement
- Measure outcomes (e.g., purchases, registrations) directly via conversion pixels on advertisers' websites—no ad clicks required

Study 4: An example

- Sample size: 25.5 million users
  - 30% Control
  - 70% Test
- Outcome: purchase at the digital retailer via “conversion pixel,” which the advertiser placed after the checkout page
- Duration: Study ran for 17 days in Q2, 2015

We measure the lift from the RCT



We measure the lift from the RCT

	Test (Eligible to be exposed)	Control (Unexposed)
25%	Exposed 0.104%	0.059%
75%	Unexposed	

Users who would have been exposed if they had been in the test group

We measure the lift from the RCT

Test
(Eligible to be exposed)

Control
(Unexposed)

Exposed

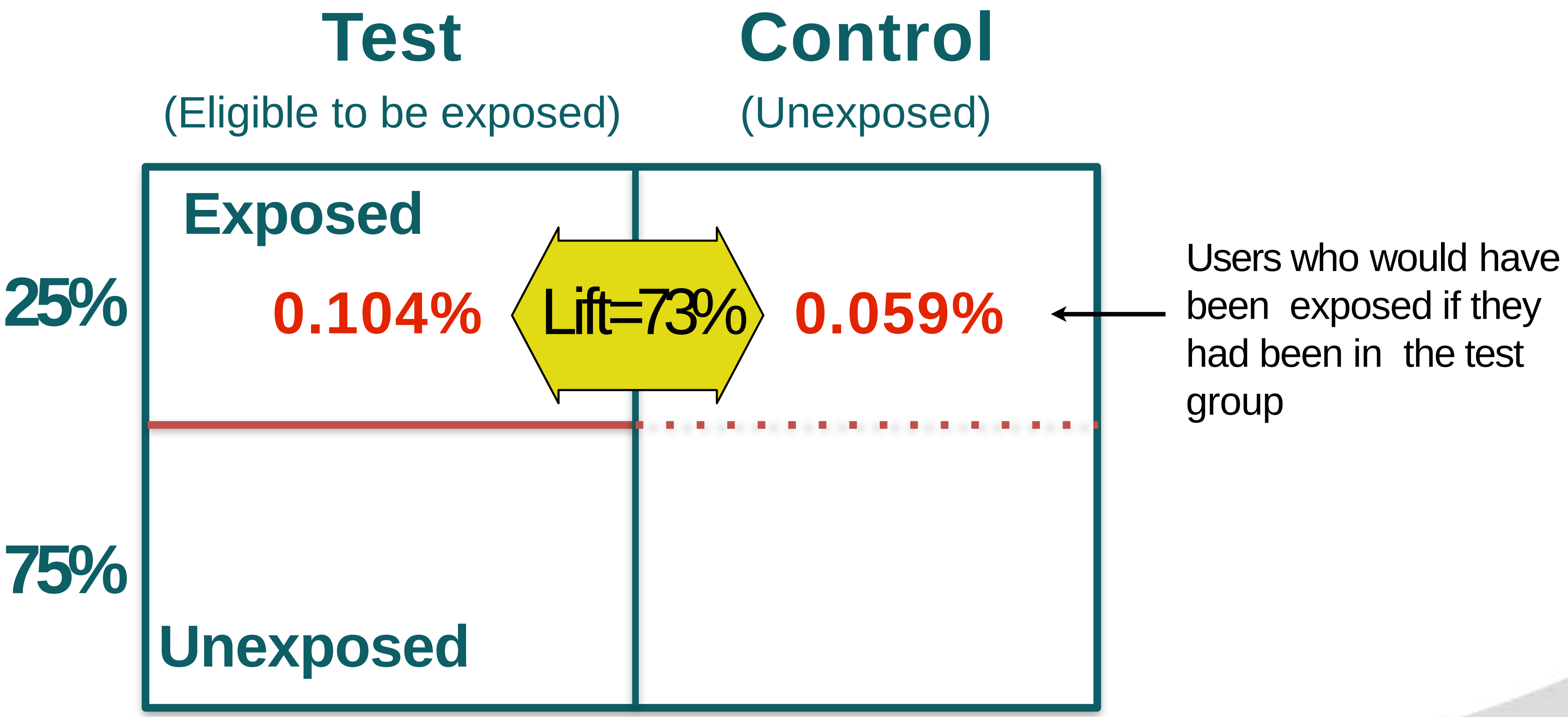
25% **0.104%** Lift=73% **0.059%**

Users who would have been exposed if they had been in the test group

75%

Unexposed

Benchmark from RCT ("Truth"): 73%



In practice, many advertisers don’t use a true control group

Test

(Eligible to be exposed)

Control

(Unexposed)

Exposed

Unexposed

The simplest measurement approach is to compare exposed to unexposed consumers

Test

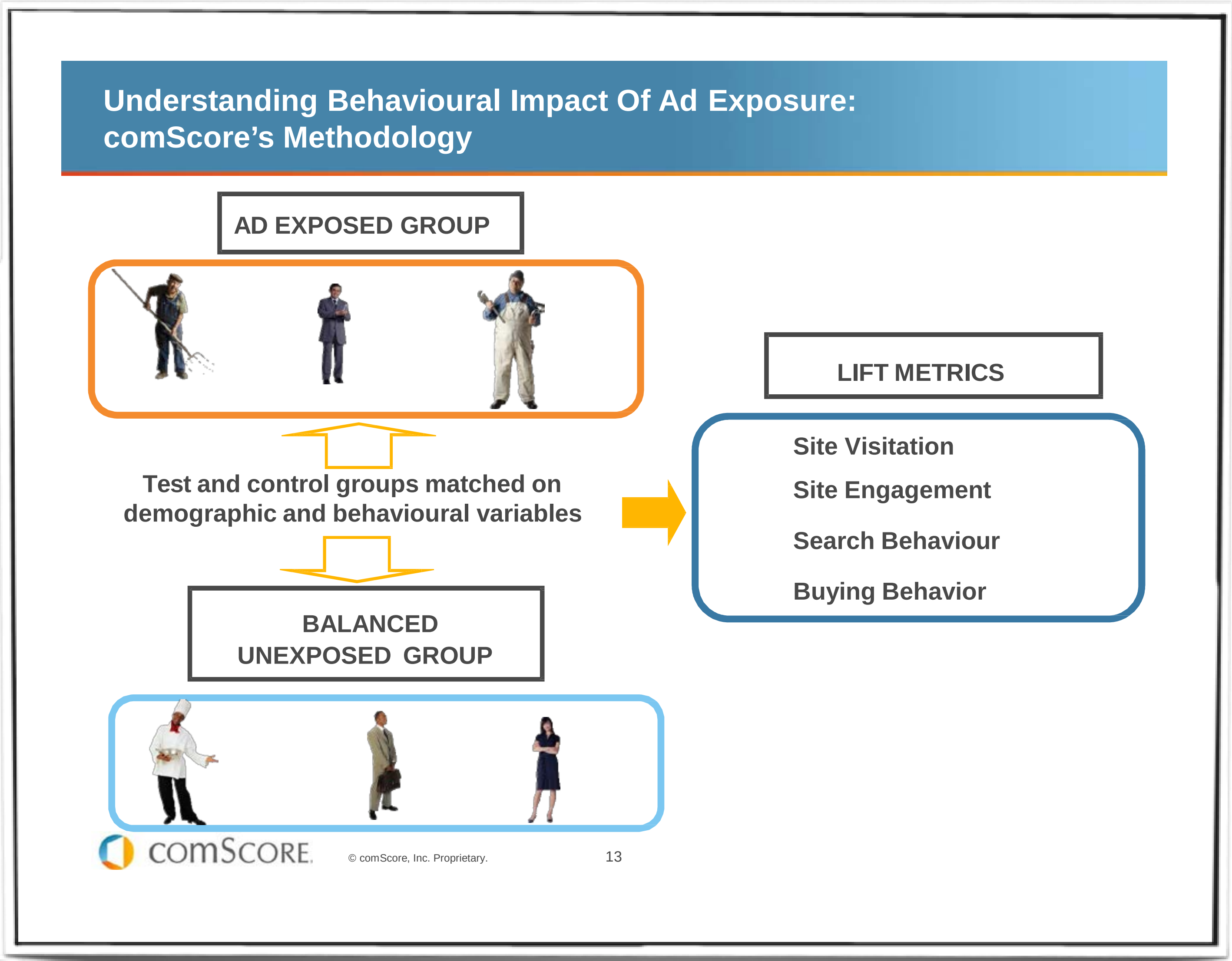
(Eligible to be exposed)

Exposed
0.104%
0.020%
Unexposed

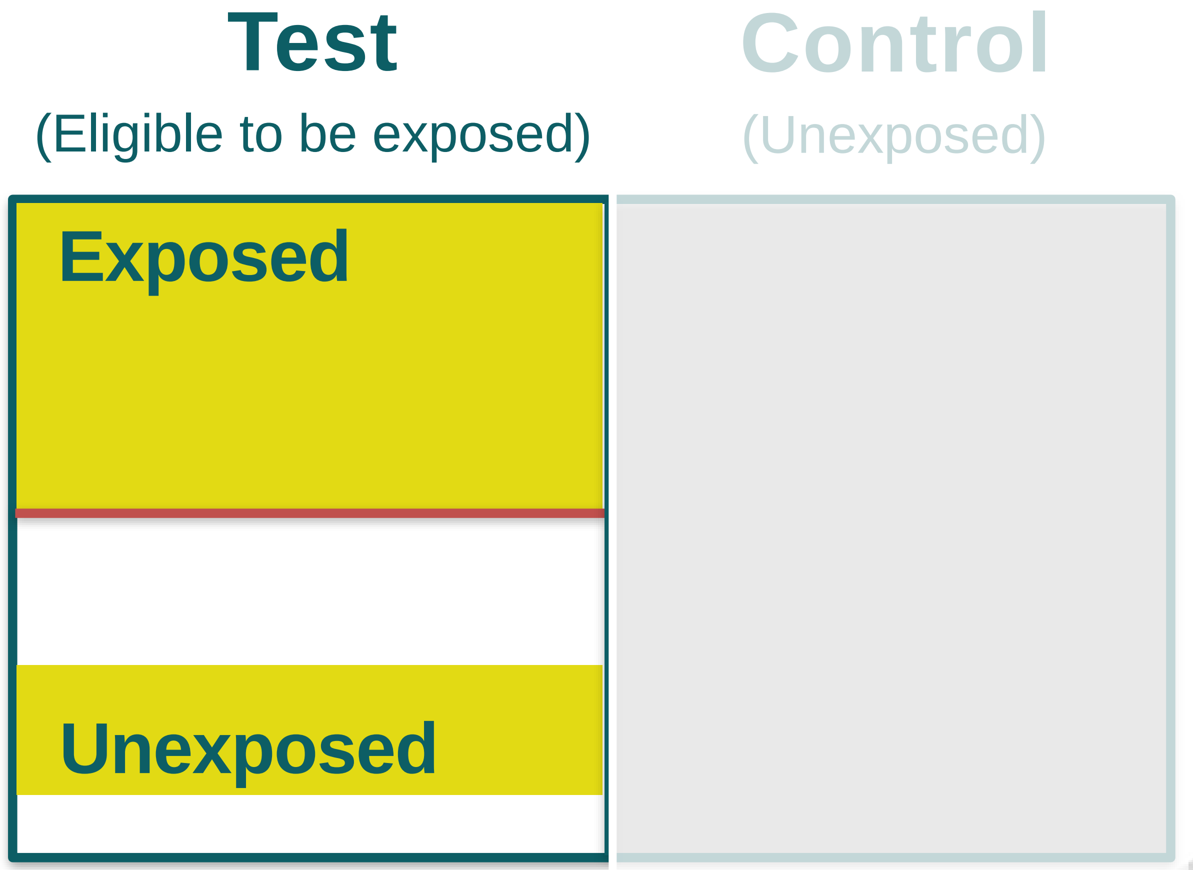
Lift = 316%

Benchmark (RCT) Lift = 73%

Ad measurement firms know about this problem



The key idea of observational methods



- Start with all exposed users
- Find unexposed users who look "similar to" exposed users based on observable variables/ features

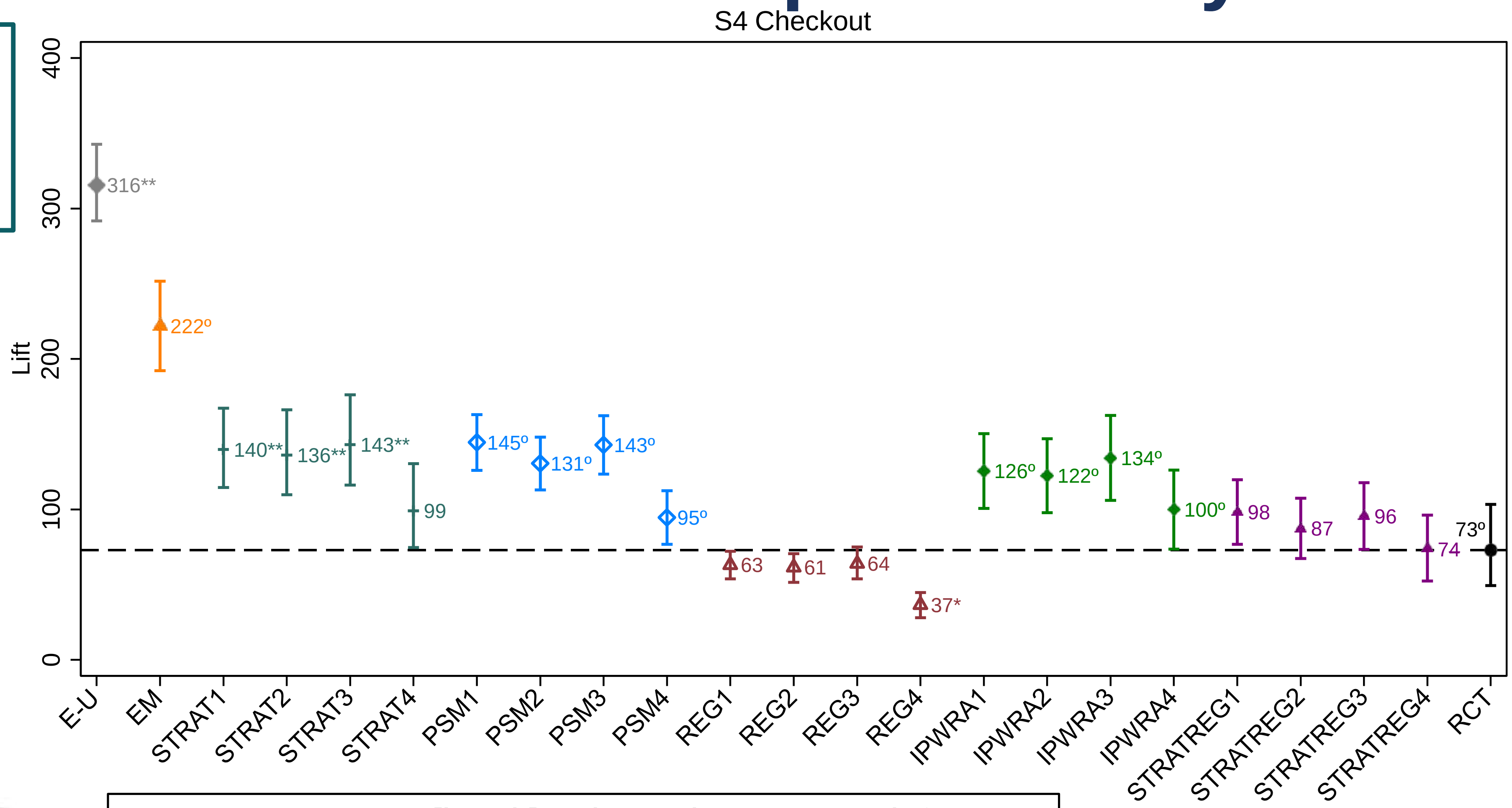
We compare RCTs with state-of-the-art methods with excellent data

- Exact Matching (EM)
- Propensity Score Matching (PSM)
- Stratification (STR)
- Regression (REG)
- Inverse Probability-Weighed Regression Adjustment (IPWRA)
- Stratification & Regression (STRATREG)

- FB variables
- FB activity variables
- FB Match score

How well do observational methods do in the example of study 4?

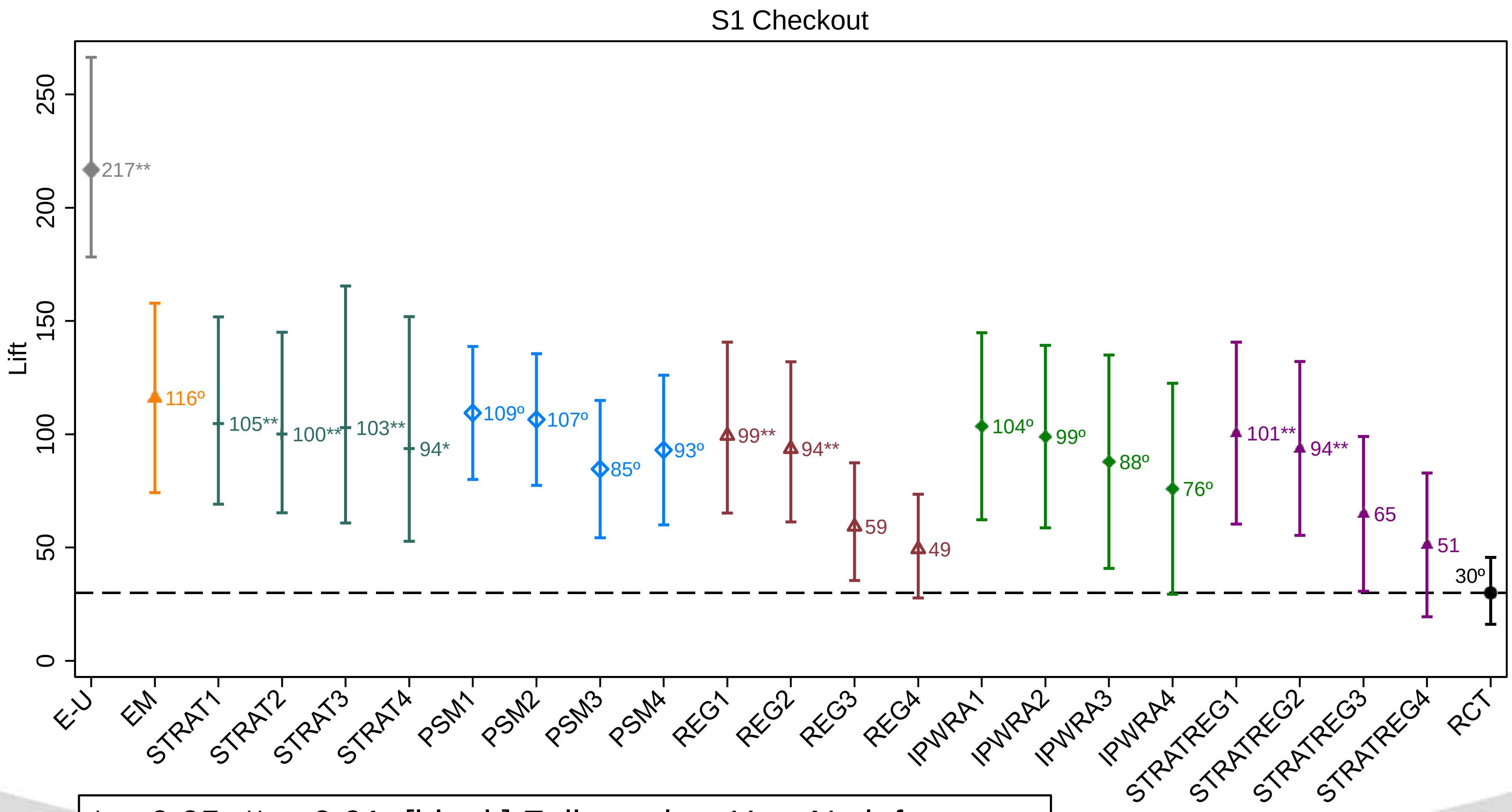
Exposed-unexposed Lift = 316%



*p<0.05, **p<0.01, [blank] Fail to reject H_0, º No inference

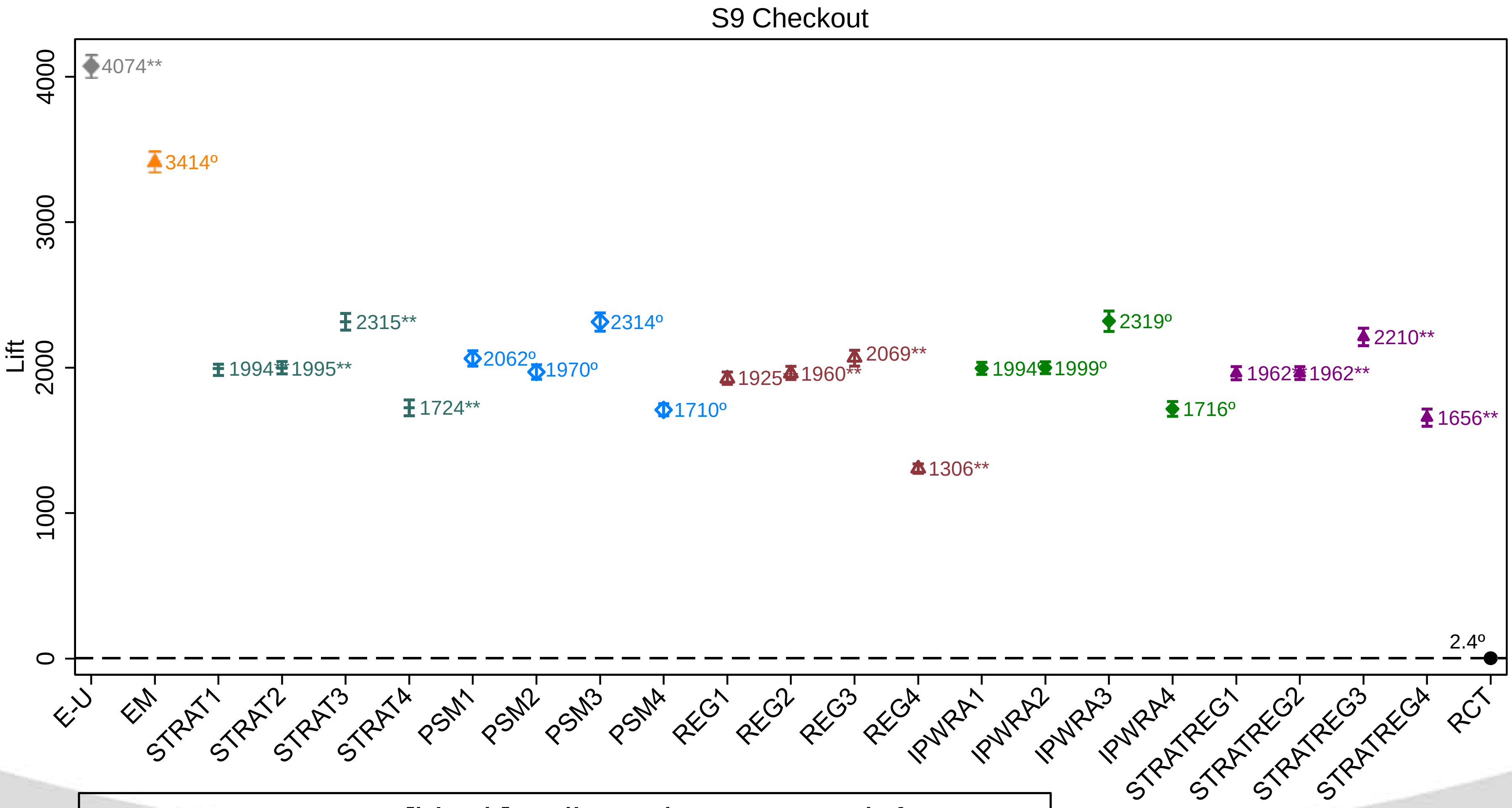
Benchmark (RCT) Lift = 73%

…and there might be a consistent pattern across methods



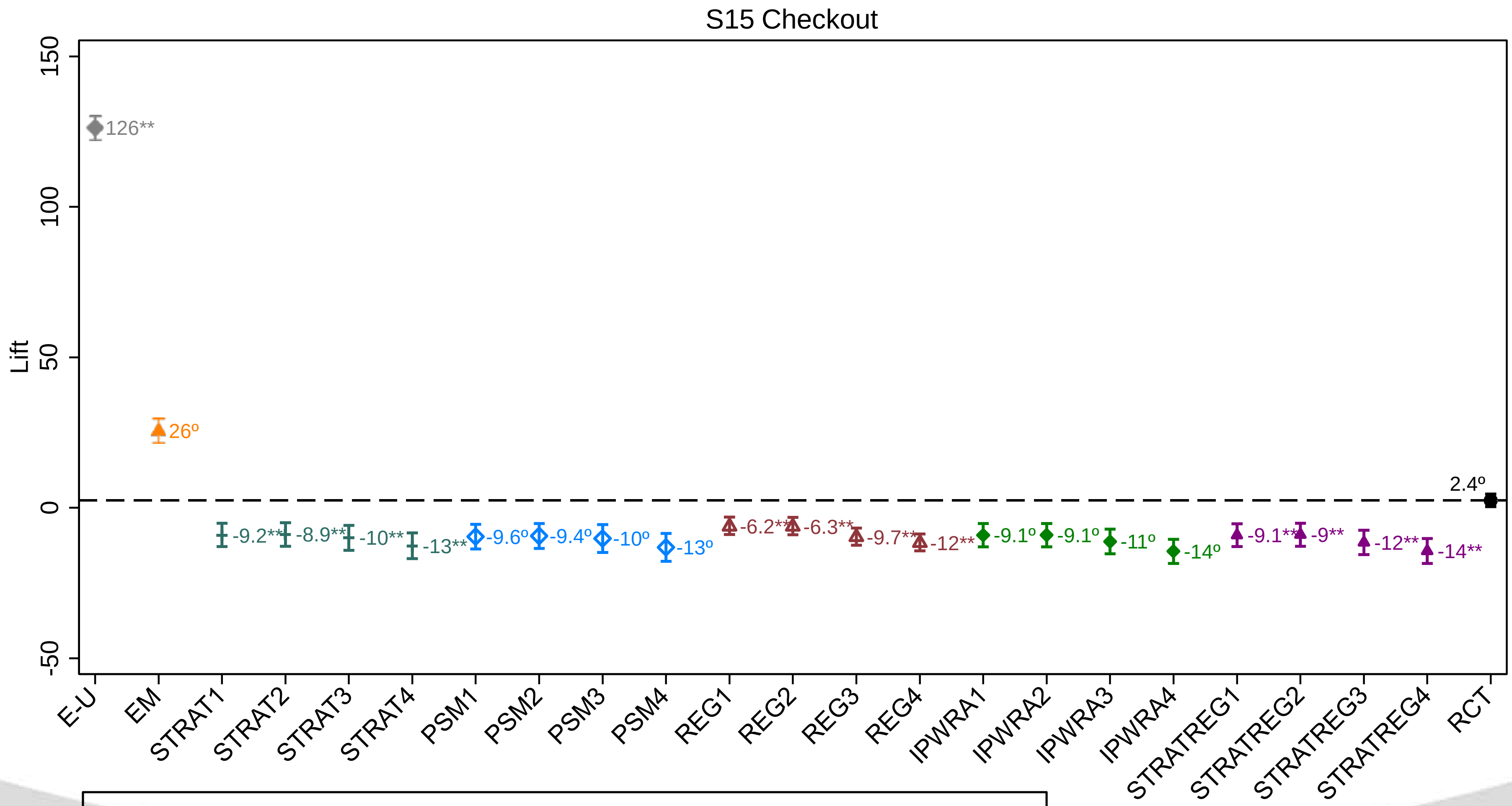
*p<0.05, **p<0.01, [blank] Fail to reject H_0, º No inference

In some of other studies lift estimates from observational methods are wildly far off



*p<0.05, **p<0.01, [blank] Fail to reject H_0, º No inference

… and some times observational methods lead to underestimates of lift



*p<0.05, **p<0.01, [blank] Fail to reject H_0, º No inference

Study	RCT Lift	E-U	EM	Stratification				Propensity Score Matching				Regression				Inv Prob. Weighted Regression Adjustment				Stratified Regression			
			Age, Gender	Age, Gender + FB	Age, Gender + FB + Census	Age, Gender + FB + Census + Activity	Age, Gender + FB + Census + Activity +FB Match	Age, Gender + FB	Age, Gender + FB + Census	Age, Gender + FB + Census + Activity	Age, Gender + FB + Census + Activity +FB Match	Age, Gender + FB	Age, Gender + FB + Census	Age, Gender + FB + Census + Activity	Age, Gender + FB + Census + Activity +FB Match	Age, Gender + FB	Age, Gender + FB + Census	Age, Gender + FB + Census + Activity	Age, Gender + FB + Census + Activity +FB Match	Age, Gender + FB	Age, Gender + FB + Census	Age, Gender + FB + Census + Activity	Age, Gender + FB + Census + Activity +FB Match
	Checkout																						
1	30%	217%	116%	105%	100%	103%	94%	109%	107%	85%	93%	99%	94%	59%	49%	104%	99%	88%	76%	101%	94%	65%	51%
2	1.3%	377%	432%	156%	148%	39%	37%	161%	149%	37%	36%	114%	103%	33%	30%	149%	140%	43%	35%	97%	98%	54%	40%
3	8.8%	198%	65%	22%	25%	46%	18%	20%	24%	41%	17%	6%	9%	21%	5%	21%	23%	38%	5%	18%	19%	30%	2%
4	73%	316%	222%	140%	136%	143%	99%	145%	131%	143%	95%	63%	61%	64%	37%	126%	122%	134%	100%	98%	87%	96%	74%
5	450%	678%	511%	427%	432%	448%	306%	418%	443%	463%	316%	409%	415%	429%	299%	428%	432%	437%	305%	447%	431%	435%	301%
7	2.7%	131%	37%	19%	20%	-34%	-35%	20%	18%	-33%	-36%	22%	23%	-19%	-21%	19%	20%	-33%	-35%	19%	19%	-31%	-33%
8	-2.9%	179%	48%	34%	39%	52%	33%	31%	36%	50%	27%	39%	45%	60%	33%	36%	41%	54%	29%	32%	37%	52%	28%
9	2.4%	4074%	3414%	1994%	1995%	2315%	1724%	2062%	1970%	2314%	1710%	1925%	1960%	2069%	1306%	1994%	1999%	2319%	1716%	1962%	1962%	2210%	1656%
10	2.0%	138%	38%	20%	20%	36%	-15%	23%	16%	43%	-7%	10%	10%	25%	-5%	20%	20%	34%	-13%	21%	21%	35%	-11%
11	9%	392%	275%	30%	30%	39%	7%	29%	31%	38%	7%	16%	16%	11%	-3%	30%	31%	35%	3%	30%	31%	34%	2%
12	1%	233%	129%	112%	110%	81%	81%	111%	110%	82%	82%	105%	107%	73%	74%	112%	111%	82%	81%	112%	111%	84%	82%
13	-15%	61%	-39%	-35%	-35%	-31%	-30%	-35%	-36%	-30%	-31%	-36%	-36%	-31%	-30%	-35%	-35%	-31%	-30%	-35%	-35%	-31%	-30%
14	62%	365%	119%	81%	86%	99%	99%	80%	85%	95%	101%	80%	83%	93%	92%	80%	83%	92%	90%	74%	77%	82%	84%
15	2%	126%	26%	-9%	-9%	-10%	-13%	-10%	-9%	-10%	-13%	-6%	-6%	-10%	-12%	-9%	-9%	-11%	-14%	-9%	-9%	-12%	-14%
	Registration																						
1	781%	1132%	1024%	976%	962%	1126%	1023%	978%	944%	1060%	977%	625%	593%	205%	155%	968%	960%	1087%	985%	824%	800%	432%	348%
5	893%	1456%	1270%	1064%	1065%	1074%	744%	1071%	1055%	1070%	765%	1204%	1189%	1196%	681%	1067%	1067%	1063%	728%	1112%	1104%	1081%	772%
8	63%	331%	180%	154%	156%	161%	135%	162%	159%	173%	167%	124%	126%	139%	99%	150%	153%	158%	114%	157%	161%	160%	125%
10	9%	136%	34%	19%	19%	32%	0%	19%	18%	34%	-3%	16%	16%	27%	3%	18%	18%	31%	0%	19%	18%	31%	2%
14	158.1%	540%	275%	219%	221%	245%	241%	215%	219%	244%	241%	234%	234%	277%	281%	219%	219%	238%	234%	219%	218%	240%	239%
	Page View																						
2	1517%	3363%	4261%	2481%	2479%	1147%	1183%	2493%	2416%	1150%	1177%	744%	747%	202%	209%	2408%	2422%	1175%	1187%	1162%	1181%	1722%	1268%
5	609%	1010%	846%	749%	747%	711%	480%	771%	731%	719%	484%	809%	803%	828%	490%	751%	748%	710%	477%	776%	769%	717%	498%
6	14%	368%	227%	103%	106%	262%	254%	103%	105%	263%	255%	66%	68%	222%	236%	103%	106%	250%	246%	111%	115%	255%	278%

* Red: RCT Lift is statistically different from 0 at 5% significance level

Observational method overestimates lift

Observational method underestimates lift

KEY FINDINGS

Firms are increasingly adopting machine learning for advertising, pricing, promotions, inventory optimization, etc.

These high-dimensional targeting algorithms create strong selection effects.

We expect the increase use of machine learning to severely limit the use of traditional non-experimental methods for measurement.

The Economics of Big Data and Personal Information

Panel Discussion:

Ginger Zhe Jin, Alessandro Acquisti,
Omri Ben-Shahar, Liad Wagman,
Florian Zettelmeyer

Moderator: Jeremy Sandford

Lunch Break

12:00-1:00 pm

The Business of Big Data

Session moderated by:

James Cooper
Federal Trade Commission
Bureau of Consumer Protection

The Business of Big Data

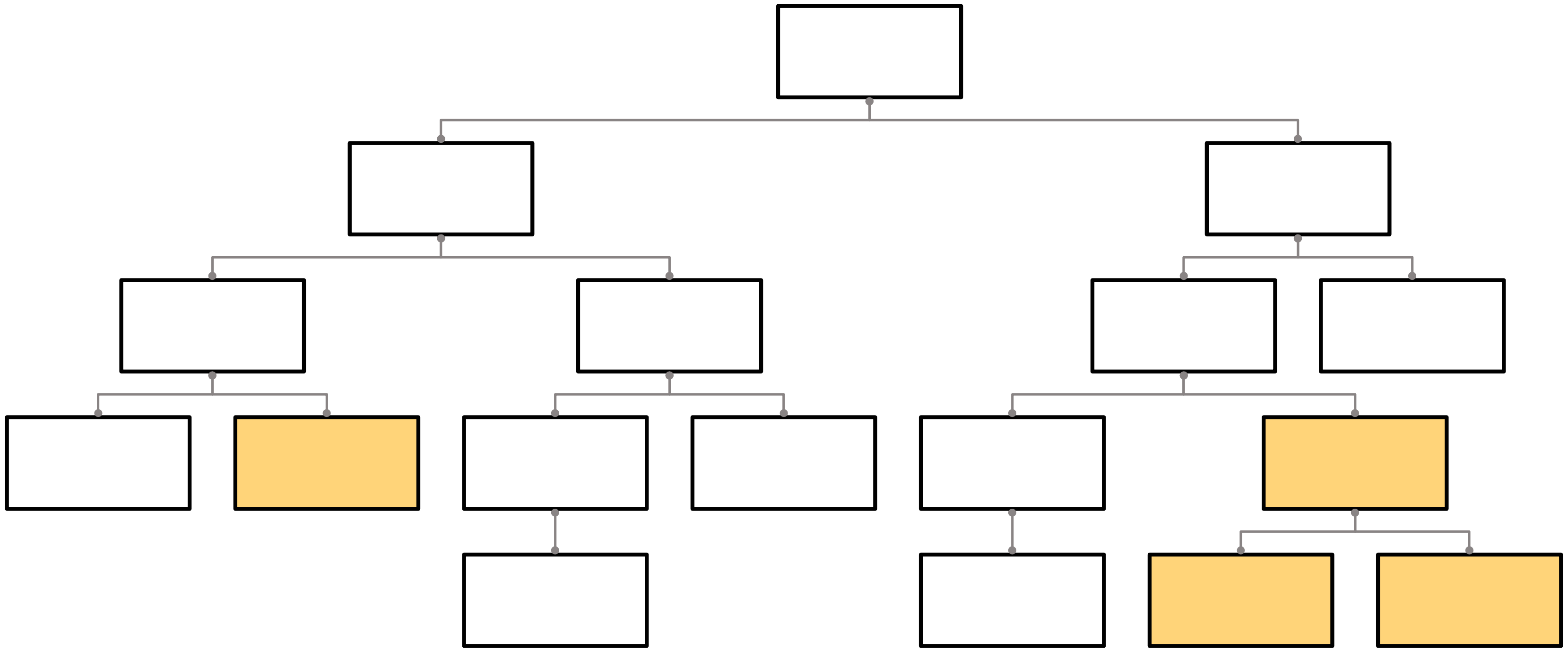
Florian Zettelmeyer

Northwestern University

Kellogg School of Management

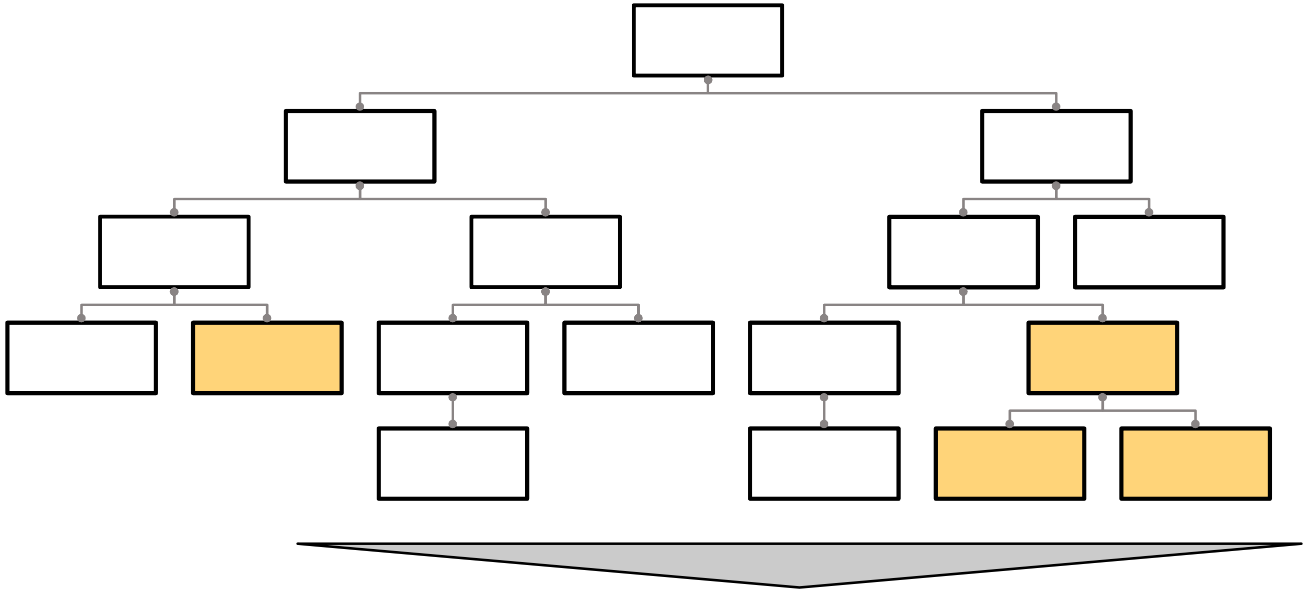
Most companies today have pockets of analytics

TYPICAL STATE OF ANALYTICS



… but they have trouble figuring out how to leverage analytics at scale

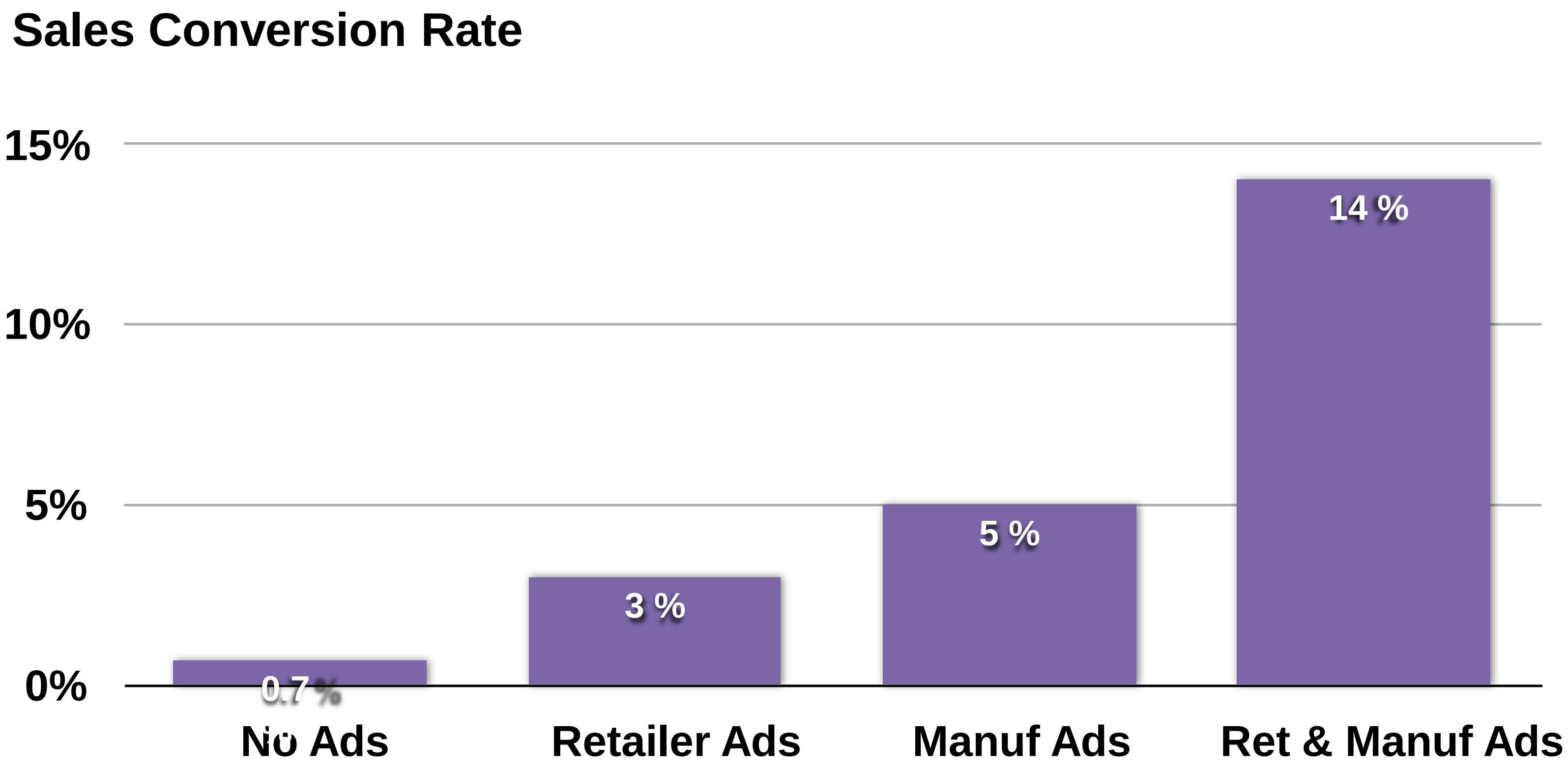
TYPICAL STATE OF ANALYTICS



My point today:

Companies are held back by a lack of data science skills at the leadership level

Consider my recent experience at an executive retreat ...



Typical view: Analytics is a problem

WHAT YOU NEED TO INVEST IN

- Analytics and Big Data technology and infrastructure
  - Hadoop, Hive, Spark, R,Python, etc.
- Cloud computing
  - Microsoft Azure, IBM Big Insights, SAP HANA, Amazon Web Services,...
- Data Scientists
  - Statistics skills
  - Computer science skills
  - Software engineering skills

Essential but nowhere close to enough ...

In reality: Analytics is mostly a problem

WHY ANALYTICS IS EVERY LEADER'S PROBLEM

(1) Analytics requires **managerial judgment**

(2) Analytics requires **organizational and incentive changes**

(3) Analytics has to be **problem-driven**

(4) Analytics has to be **planned**

Leaders need a "working knowledge of data science"

WHY A "WORKING KNOWLEDGE"?

1. Judge what good looks like

2. Identify where analytics adds value

3. Lead with confidence

Building organizational muscle with analytics requires investments in multiple areas

INVESTMENT AREAS

Integrated Data

Analytics Tools

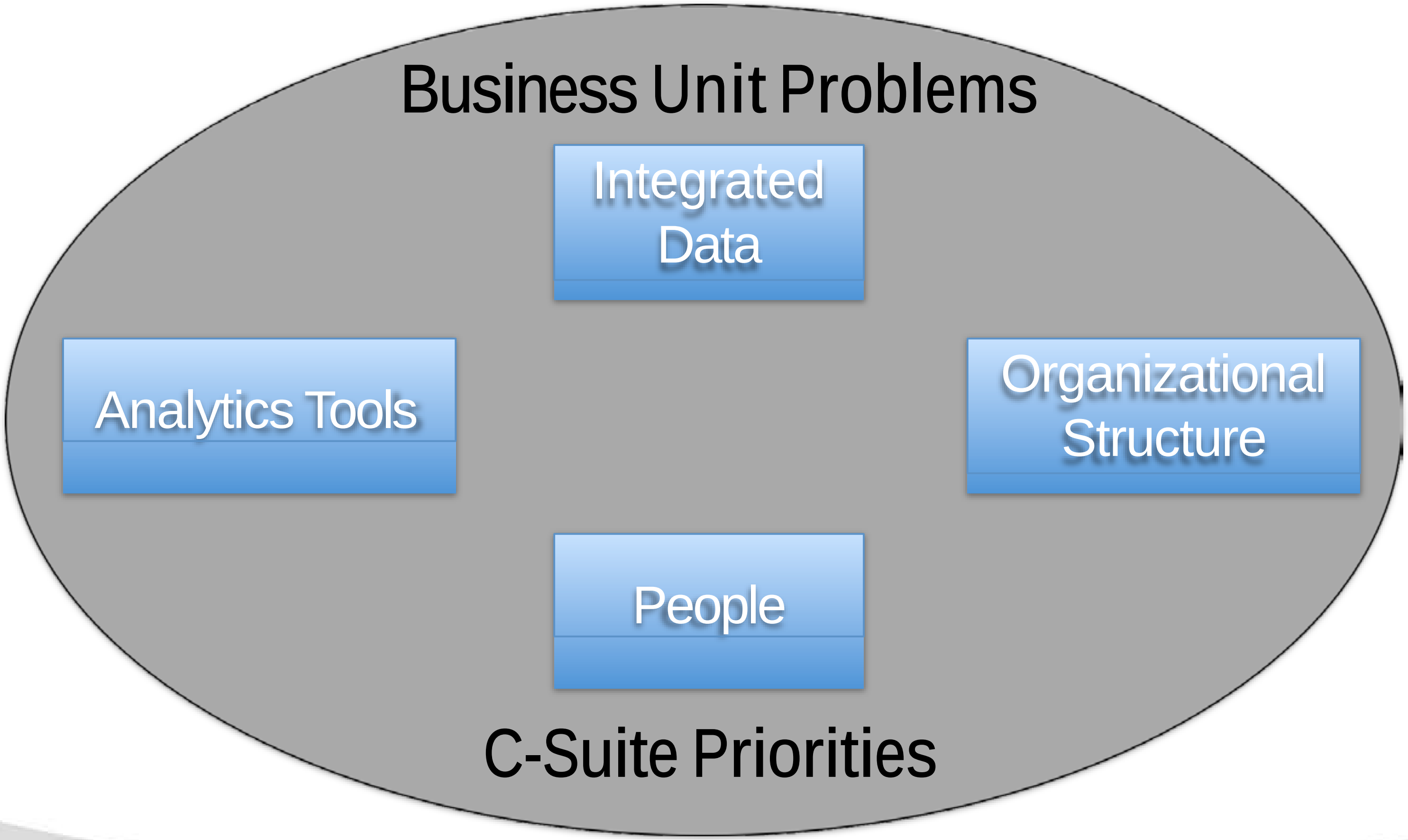
Organizational Structure

People

… but there are no generic answers to what data, tools, structure, and people are needed.

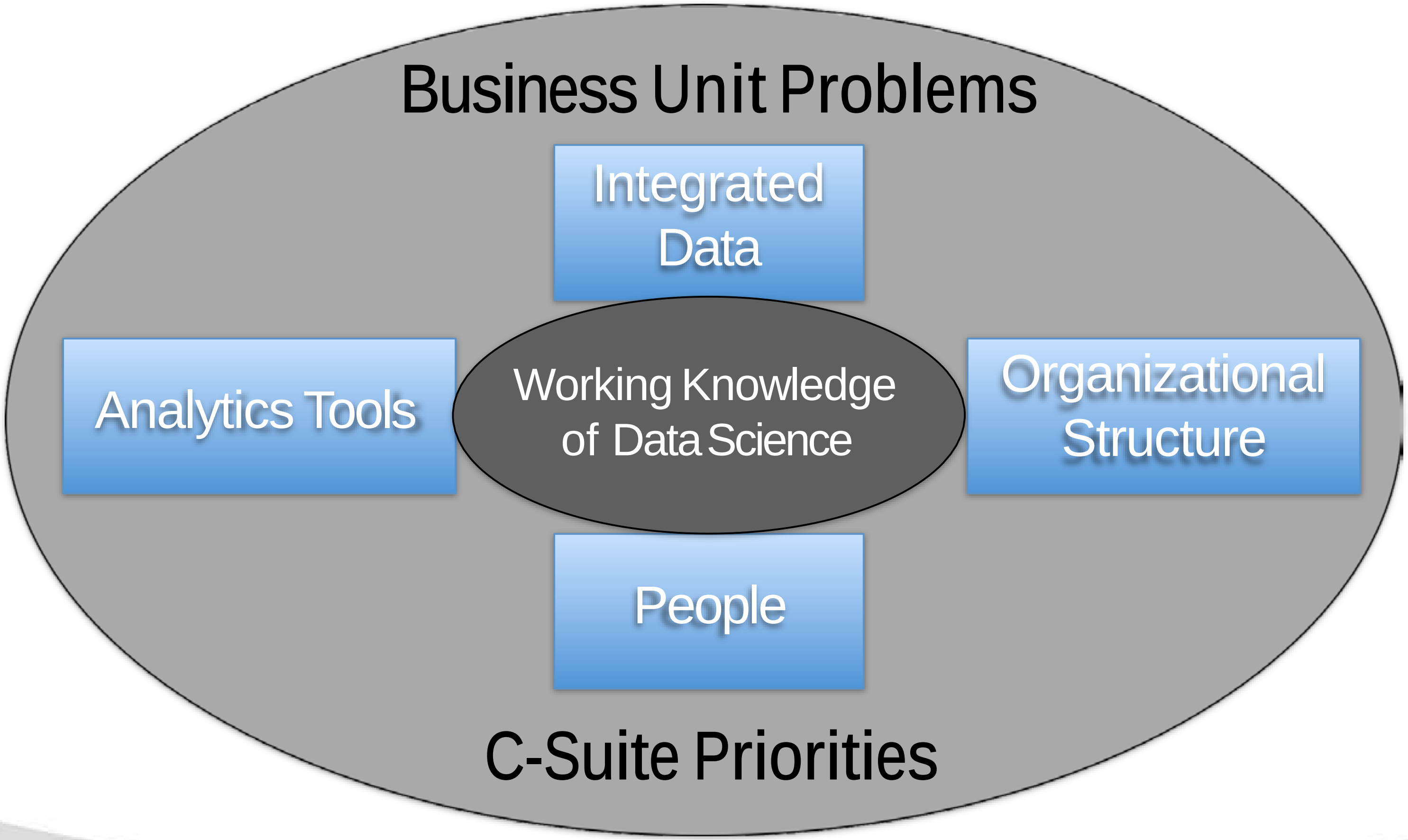
Analytics investments have to be guided by business problem and strategic priorities

INVESTMENT AREAS



Linking problems with the right investments requires a working knowledge of data science

INVESTMENT AREAS



The Business of Big Data

Christopher Boone
Pfizer

The Business of Big Data

Liz Heier
Garmin

The Business of Big Data

Marianela López-Galdos
Computer & Communications
Industry Association

The Business of Big Data

Mark MacCarthy
Software & Information Industry Association

The Business of Big Data

Morgan Reed
The App Association

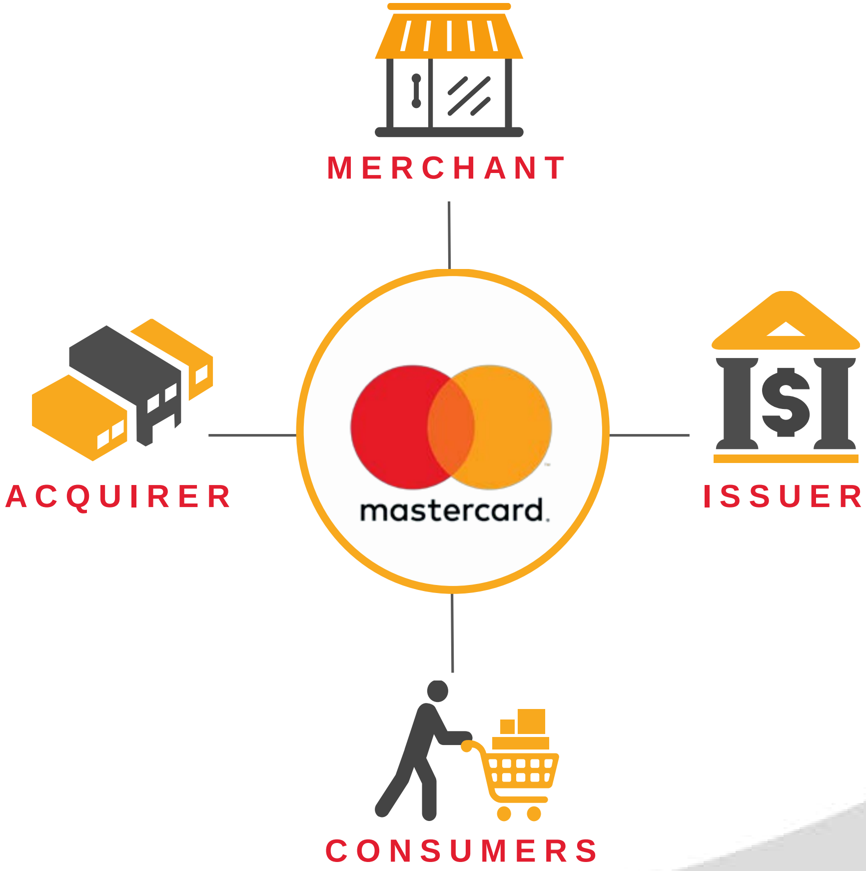
Innovating with Data at Mastercard

Andrew Reiskind
Mastercard

Who is Mastercard?

Mastercard is a technology company in the global payments industry.

- We connect consumers, financial institutions (banks, issuers and acquirers), suppliers, merchants, governments and businesses worldwide.
- We facilitate the processing of payment transactions, permitting MasterCard cardholders and checking account holders to use their accounts at millions of merchants and suppliers worldwide.
- Our network provides merchants and suppliers with an efficient and secure means of receiving payments, and account holders with a convenient, quick and secure payment method.
- We make payments safe, simple and smart.



Fraud checks throughout a transaction

Mastercard delivers intelligent security decisions within milliseconds for millions of transactions every day.

Individual

Does this individual interact with this device in a familiar way?

Supplier/ Merchant

Are the channel attributes valid?

Has this individual shopped at the merchant /supplier before?

mastercard

Does the purchase fit historical transaction behaviors?

Financial Institution

How valuable is the transaction and relationship?

Authentication

Authorization

Is this out-of-pattern behavior?

The Business of Big Data

Panel Discussion:

Florian Zettelmeyer, Christopher Boone,
Liz Heier, Marianela López-Galdos,
Mark MacCarthy, Morgan Reed,
Andrew Reiskind

Moderator: James Cooper

Break

2:30-2:45 pm

The Impact of GDPR on EU Technology Venture Investment

Liad Wagman

Illinois Institute of Technology

Stuart School of Business

Moderator: Andrew Stivers

Federal Trade Commission

Bureau of Economics

The Short-Term Effects of GDPR on Technology Venture Investment

Liad Wagman
Stuart School of Business
Illinois Institute of Technology

Joint work with:

Jian Jia
Illinois Institute of Technology

Ginger Jin
University of Maryland

Hearings on Competition and Consumer Protection in the 21st Century
An FTC-AU Washington College of Law | November 6-8, 2018 | ftc.gov/ftc-hearings | #ftchearings

GDPR’s Implementation Stage

- SafeDK, 1/25/18: More than half of mobile applications are not GDPR ready
- 5/9/18, 5/23/18: Apple removes apps that share location data w/o consent, updates privacy terms
- 5/10/18: Facebook: “Businesses may want to implement code that creates a banner and requires affirmative consent… Each company is responsible for ensuring their own compliance”
- 5/24/18: Shopify updates app permissions for merchants/developers
- 5/24/18: Google releases consent SDK for developers
- 5/25/18: GDPR takes effect

Motivation

- GDPR mandates: data management, auditing and classification; data risk identification and mitigation; interfaces for users' own data + obtain granular informed opt-in consent + allow deletion; train or hire qualified staff; or face severe penalties (can be ~$23m or 4% of annual revenue)
- Bloomberg: "500 biggest corporations are on track to spend a total of $7.8 billion to comply"
- Young ventures are more susceptible to increases in compliance costs (Campbell et al., 2015; Krasteva et al., 2015)
- Who better to assess those costs than investors?
- Compliance costs were realized as new policies were rolled out
  - Reliance on larger platforms' policies (compliance, liability, compatibility)

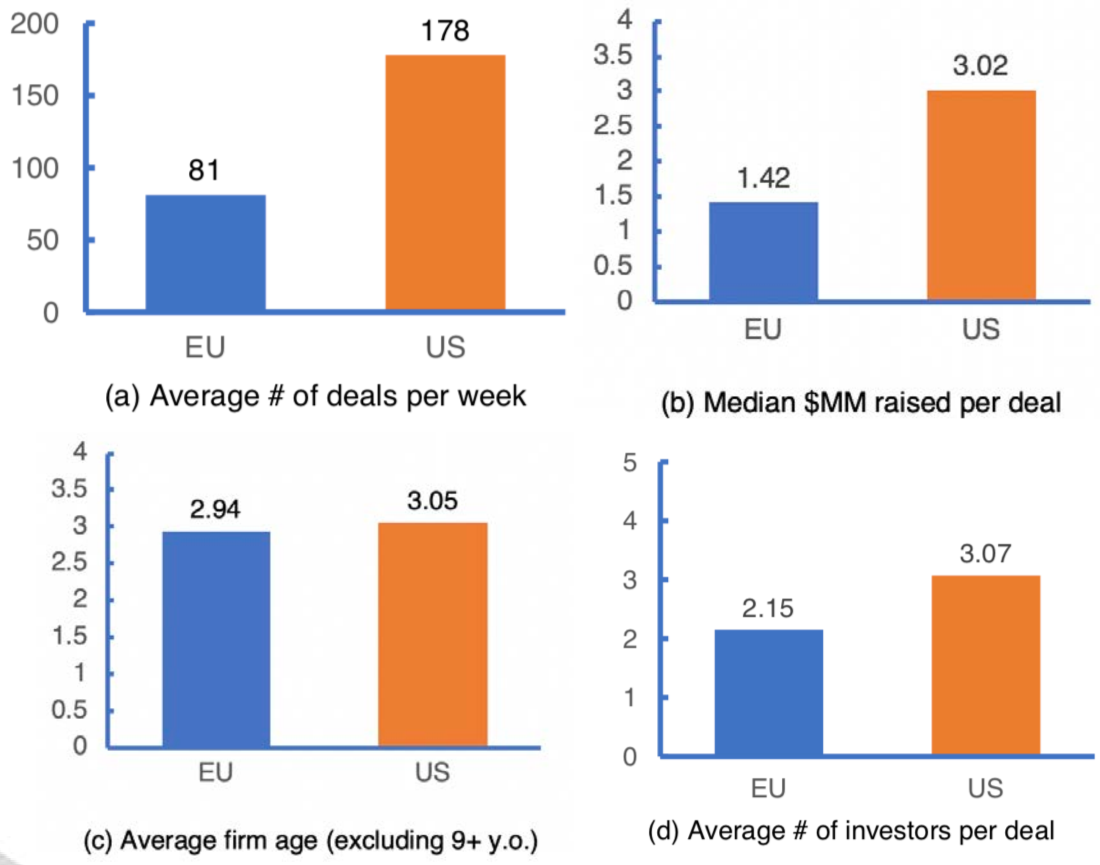
Data

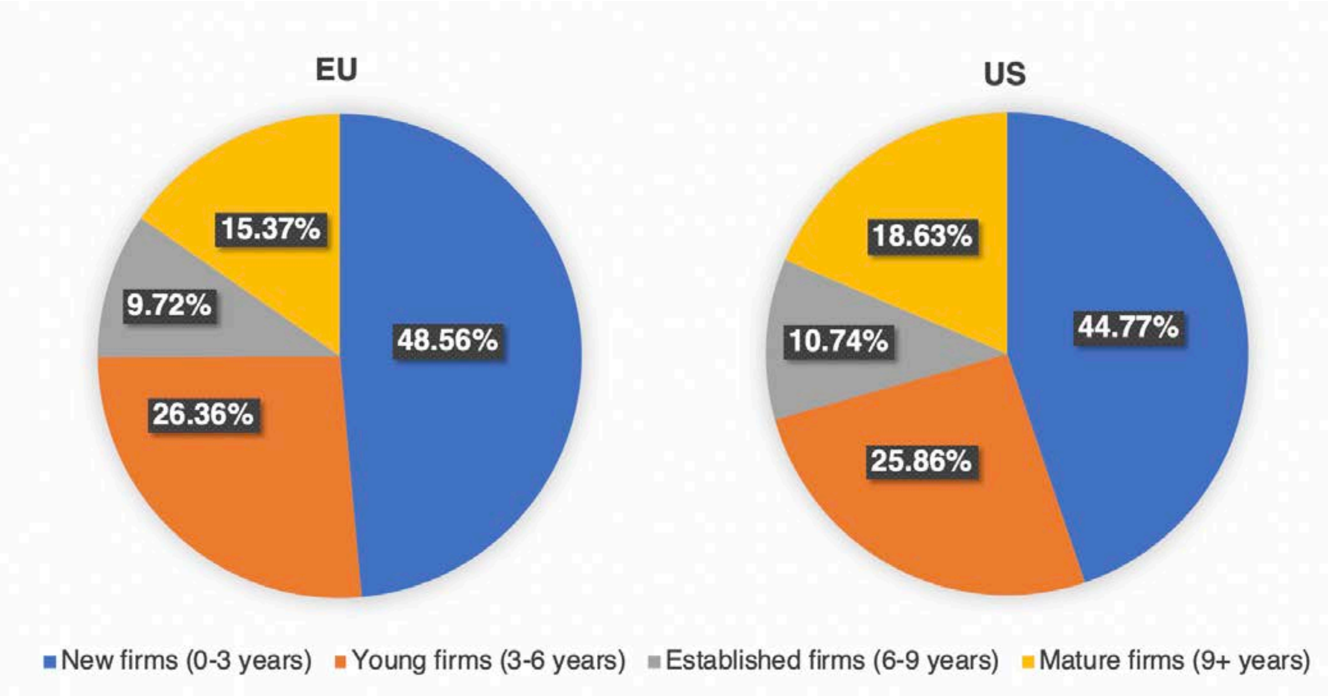
- Venture deals in EU & US taking place in July 2017 through September 2018 from Crunchbase

- Firm information (name, location, category, founding date, financing dates, employee range)

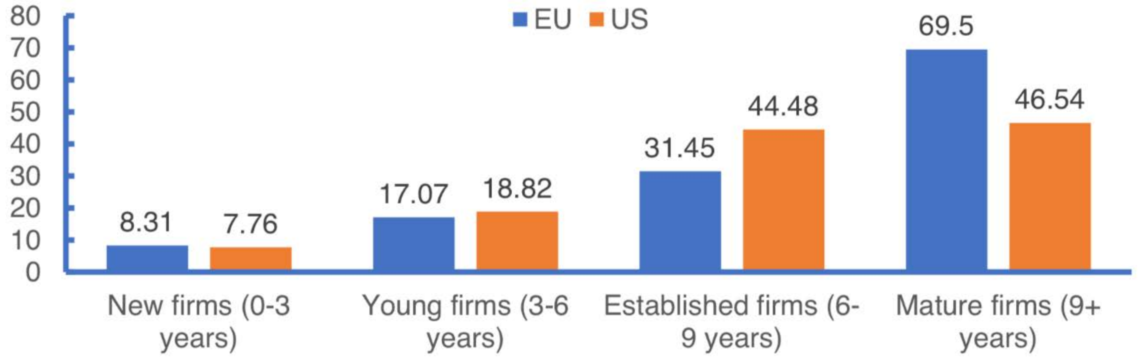
- Deal information (size & date of deal, funding stage such as Seed/Series A/etc, participating investors)

Summary Statistics

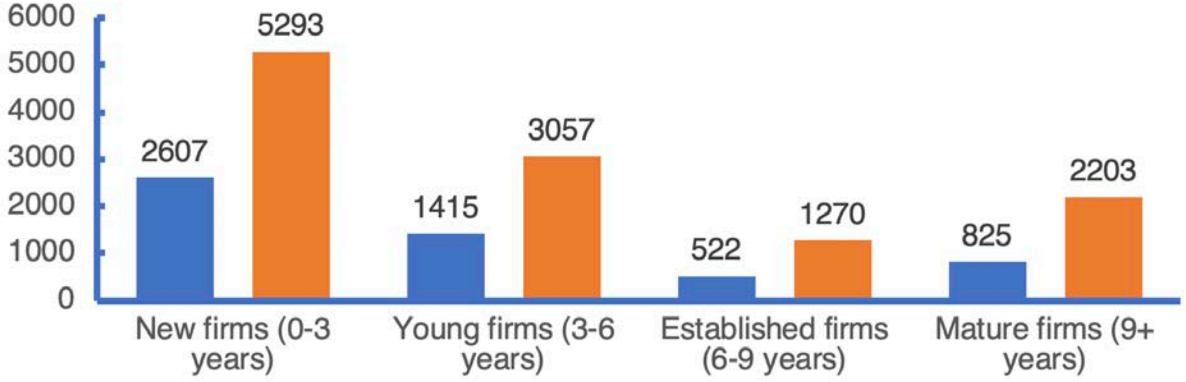




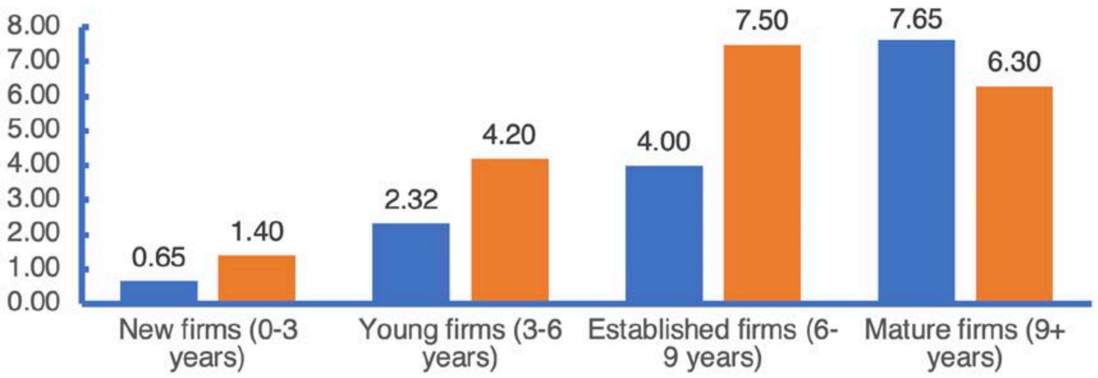
Summary by Venture Age



(a) Average $MM raised per deal

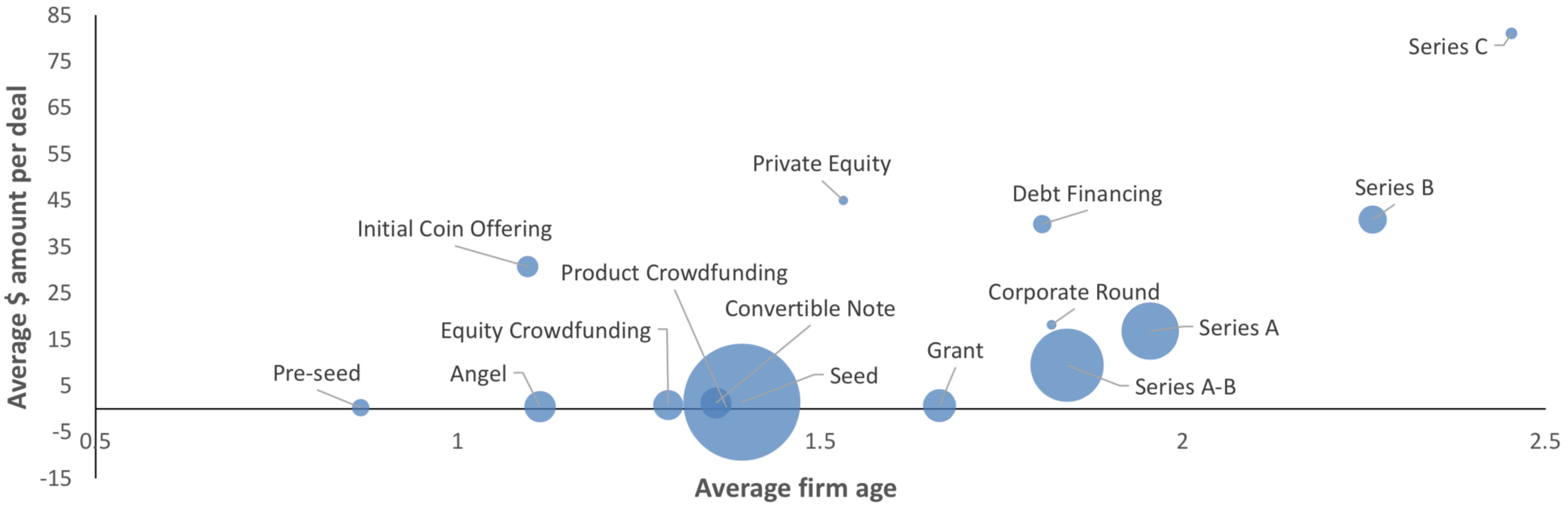


(b) Total # of deals



(c) Median $MM raised per deal

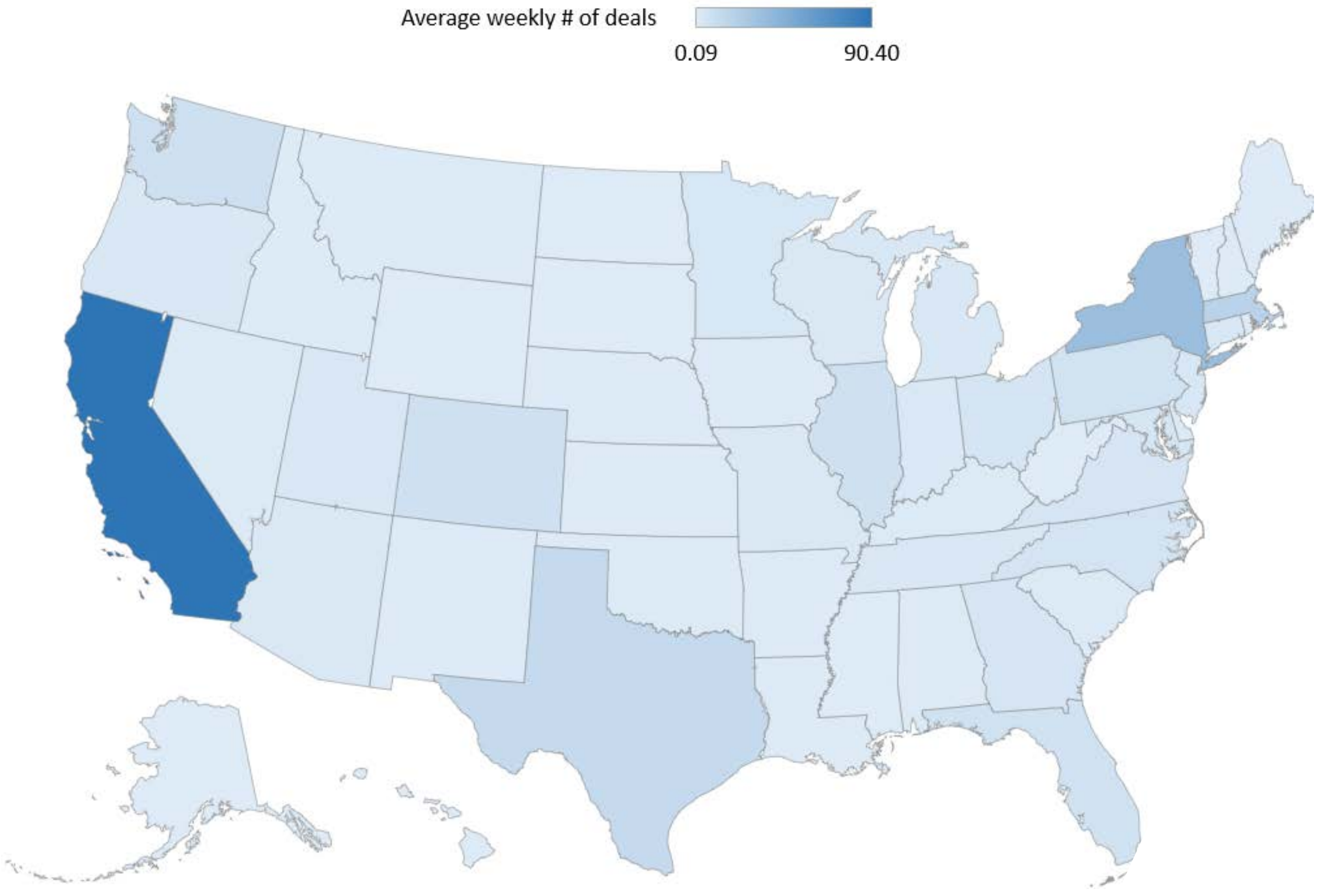
Funding Stage (Firm Age, Average $ Raised)

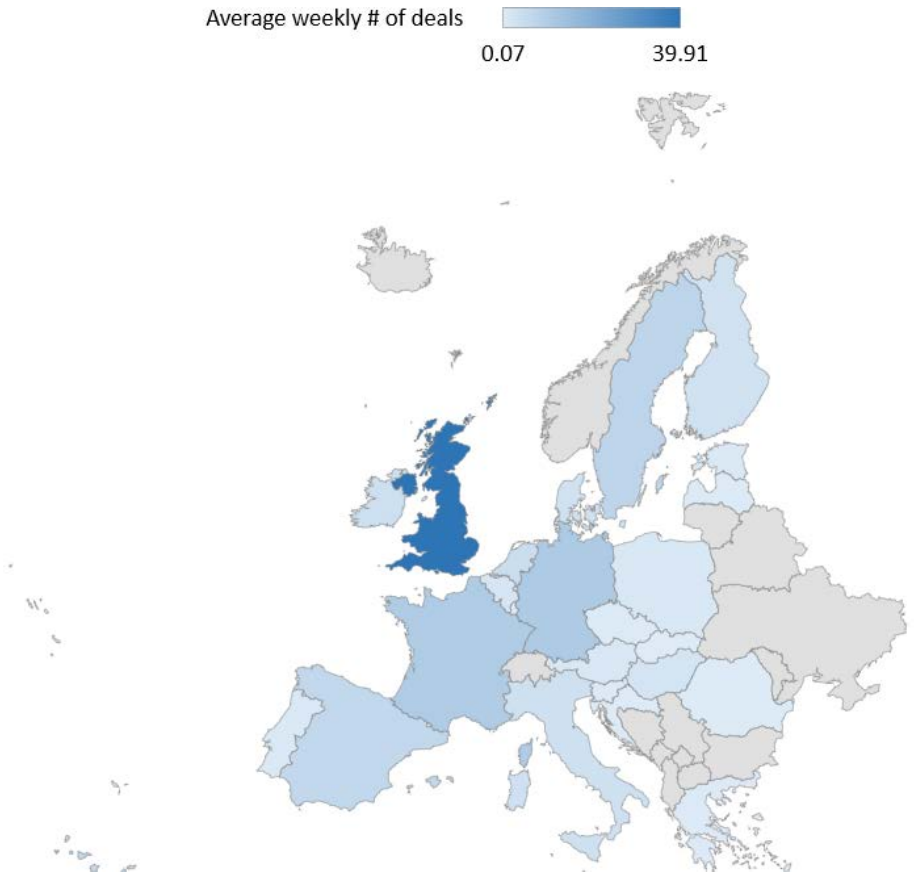


Larger circles: higher # of deals

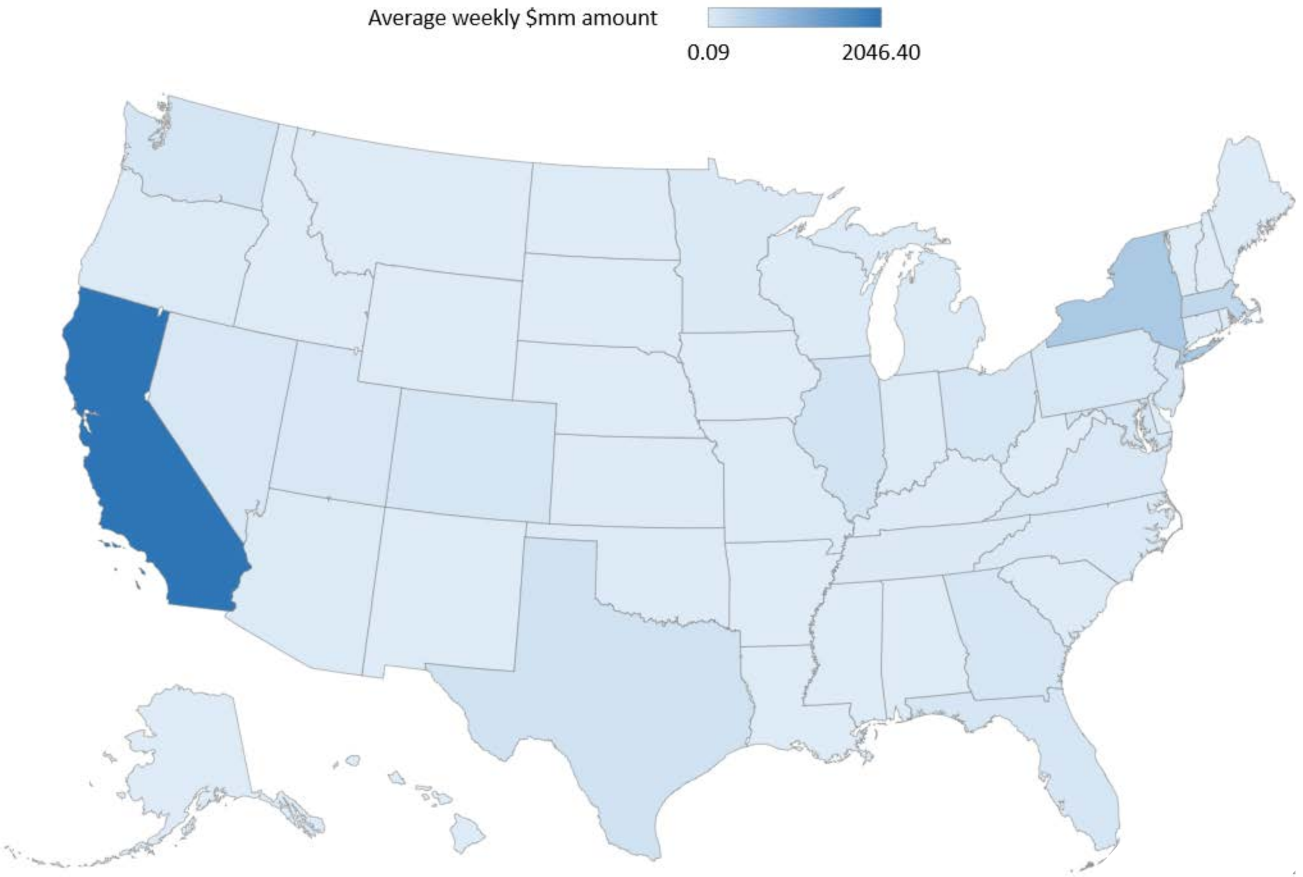
New (0-3 year old) firms

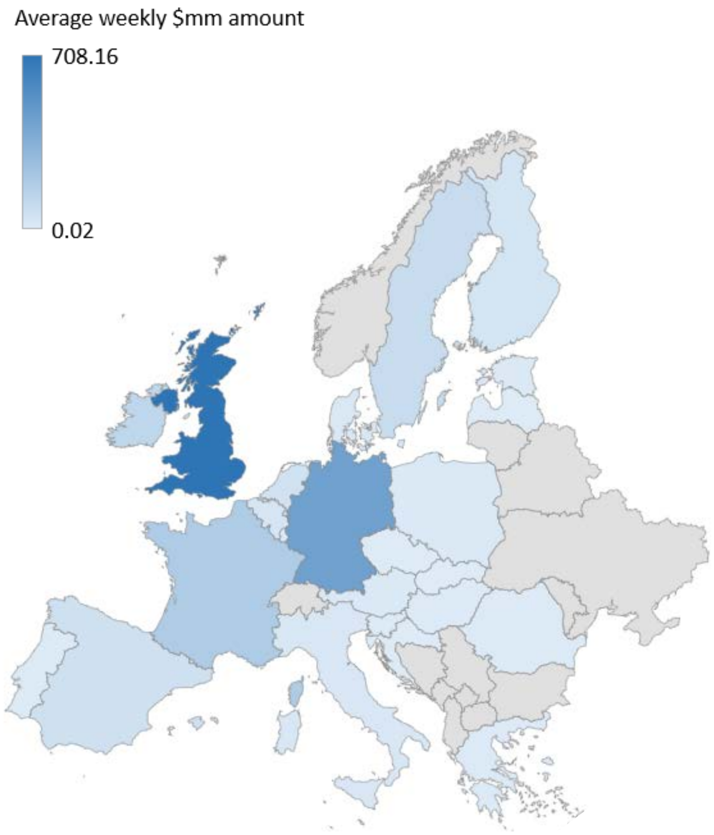
Summary by Location
Average # of deals per category (5 categories)



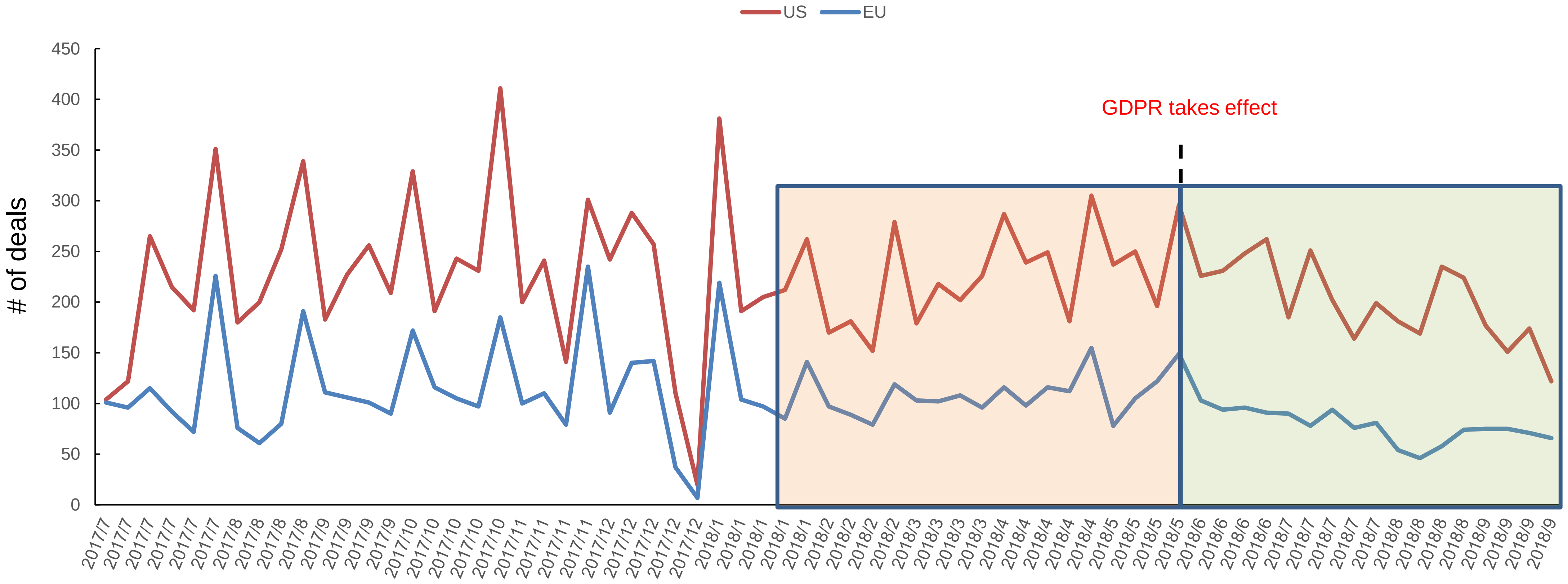


Summary by Location ($ amount)

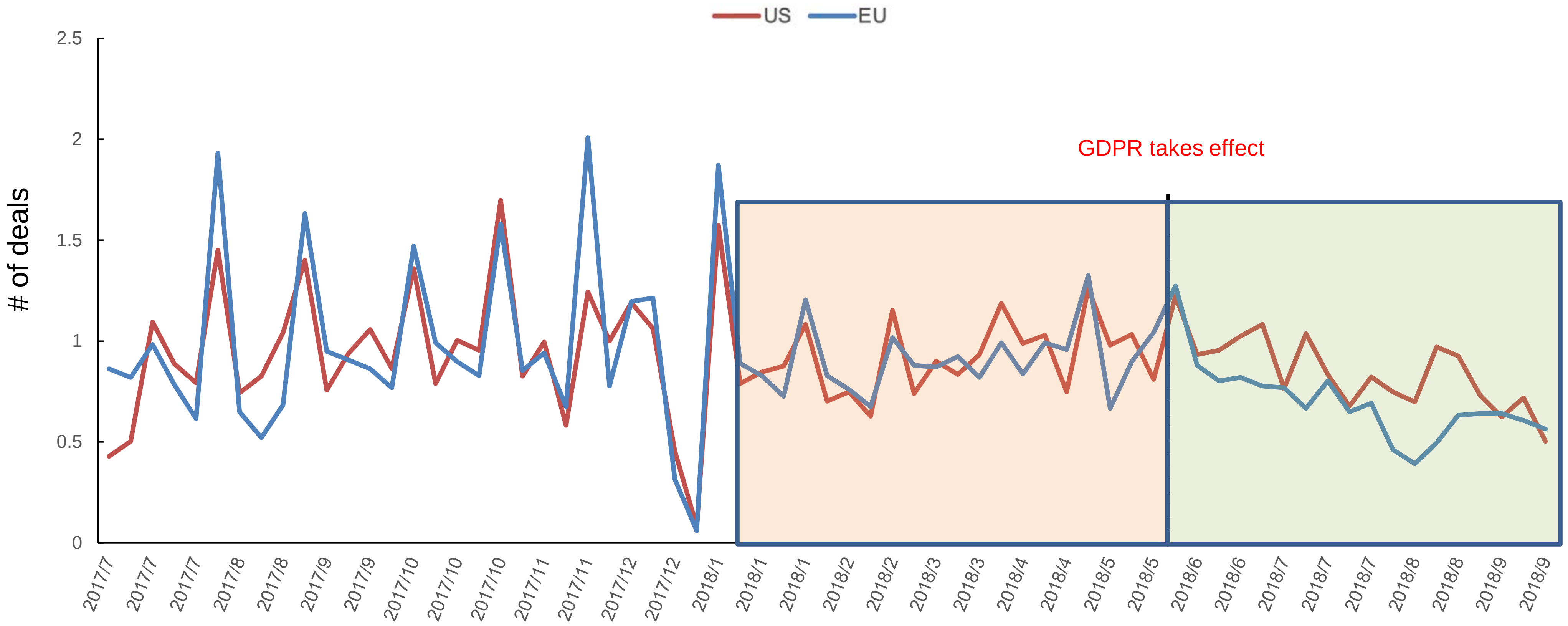




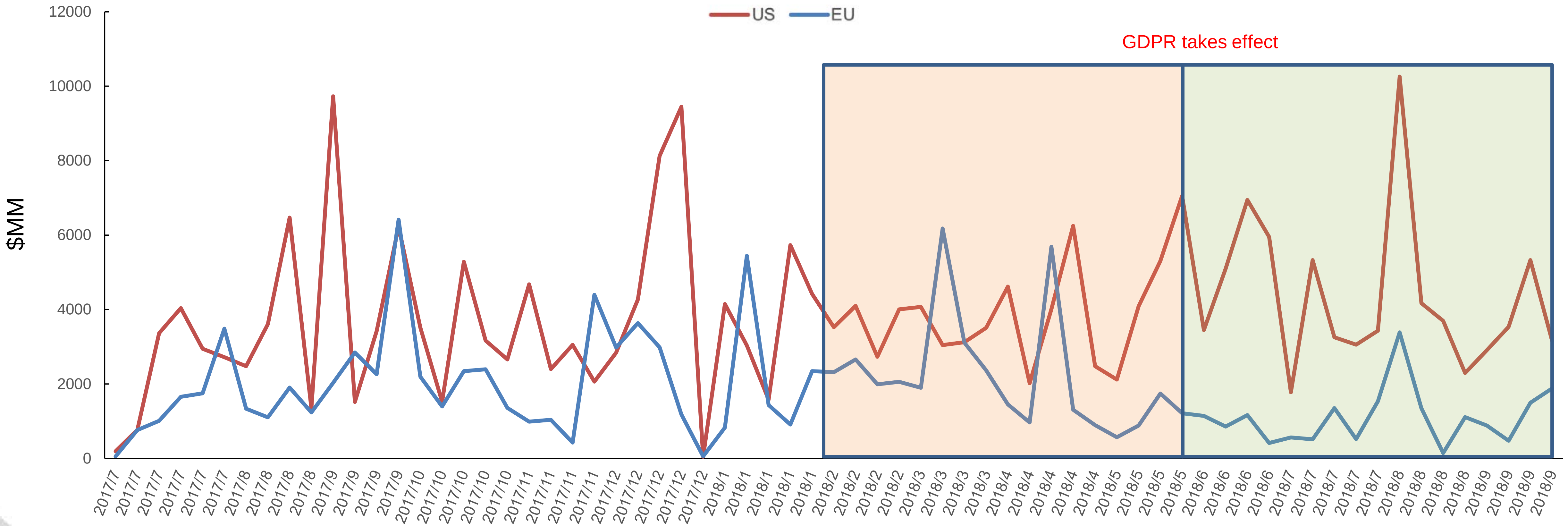
<u># Deals Per Week</u>, EU & US:



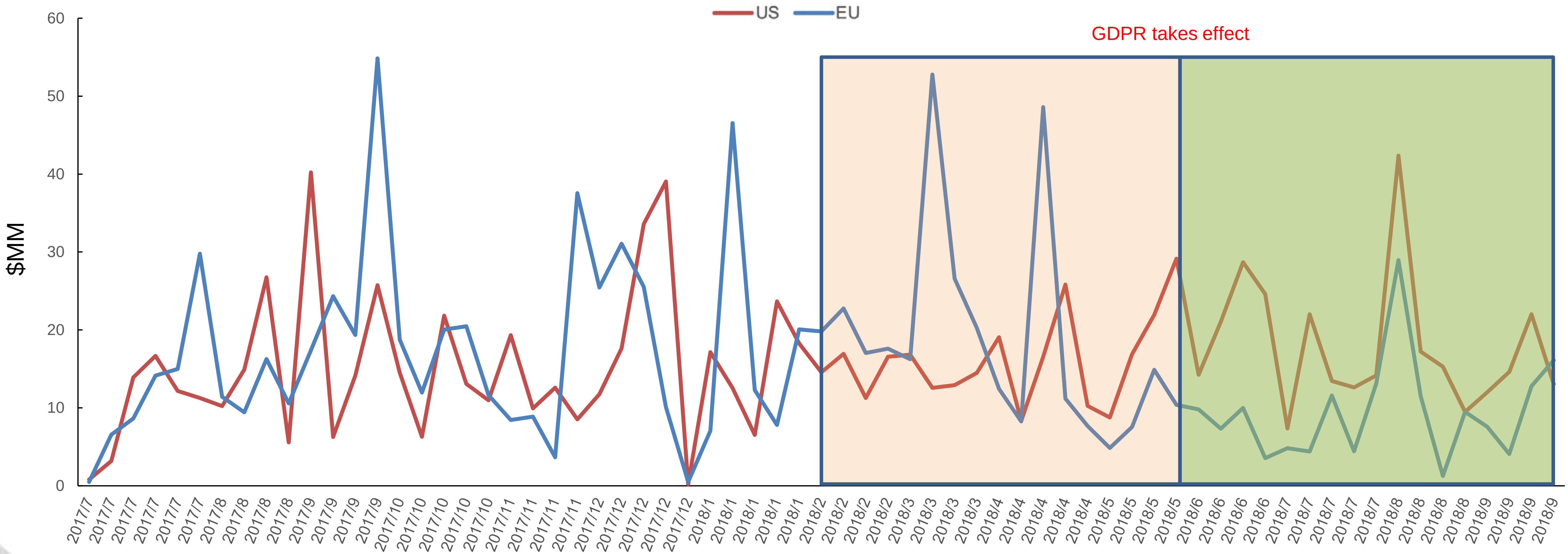
Deals/Week/State/Category, EU & US:



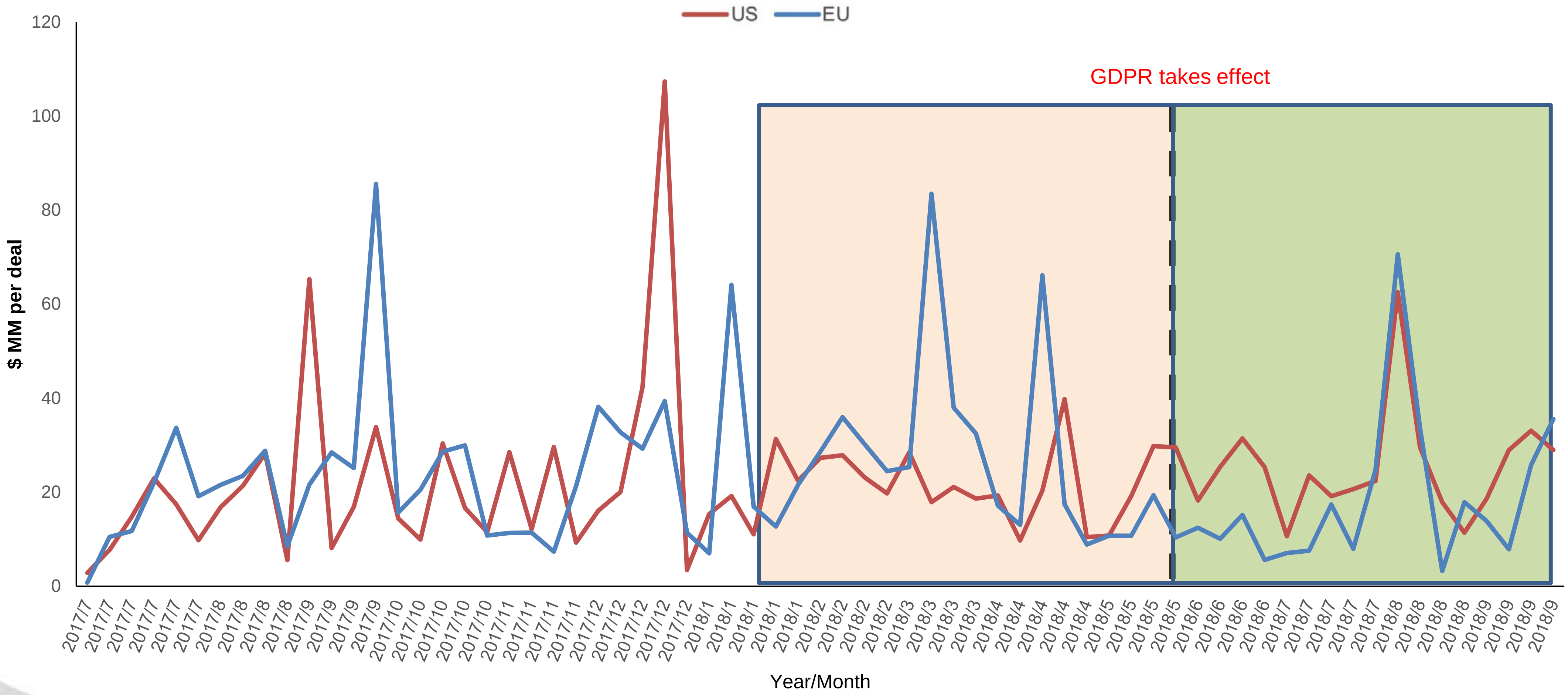
Total $ Raised Per Week, EU & US



Total $ Raised Per Week/State/Category



Average Weekly $ Raised Per Deal, EU & US

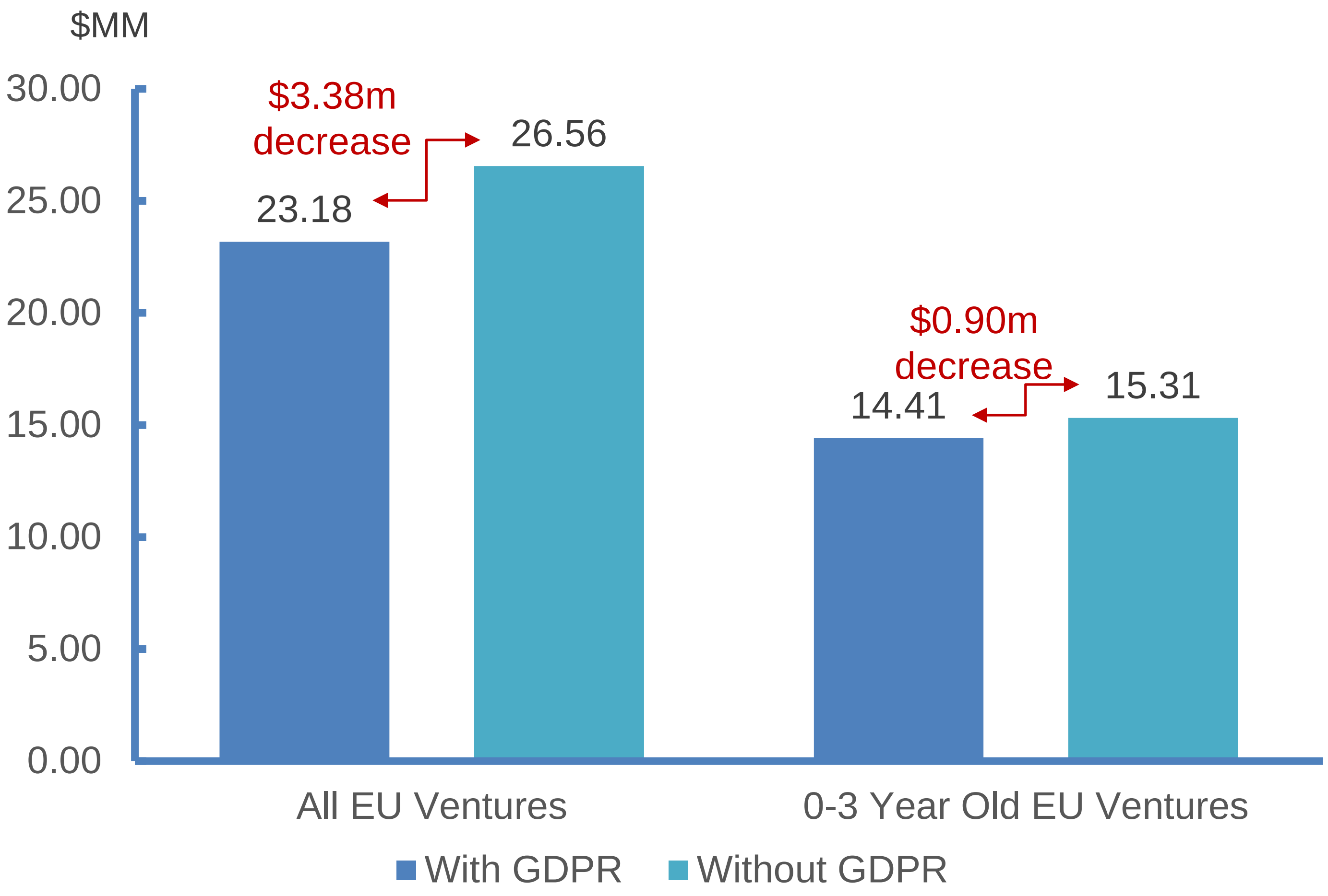


Empirical Methodology

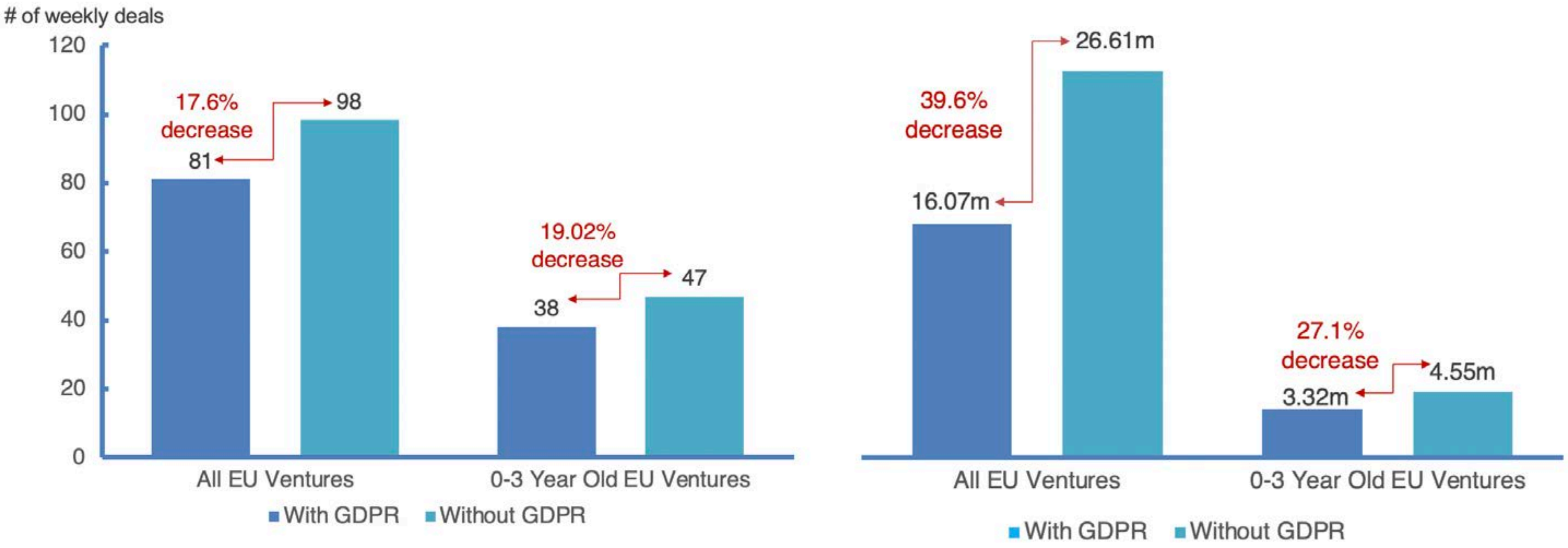
- Difference-in-difference framework
  - EU ventures after May 25 2018 as treatment, US ventures as control group
- Tobit for $ amount (0 censored), Poisson for # of deals (count data)
- Macroeconomic controls (unemployment, CPI, interest rate, GDP, exchange rate)
- Time (week) and state (US) /country (EU) fixed effects
- Log linear at deal level, control for investor type, firm age, funding stage, category

$$y_{jkt} = \alpha_t + \alpha_k + \delta X_{jkt} + \beta GDPR_{kt} + \varepsilon_{jkt}$$

GDPR Effect on $MM Raised Per Week Per Member State Per Category (Average EU)



GDPR Effect on Number of Deals and $MM Per Deal (Average EU)

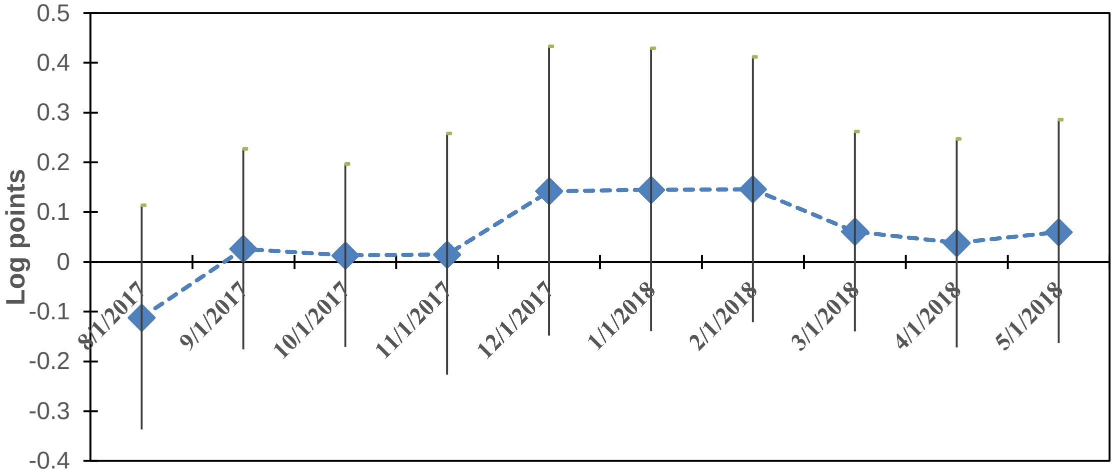


Combined findings indicate negative effects in both the **extensive** margin (# of deals) and **intensive** margin ($ per deal)

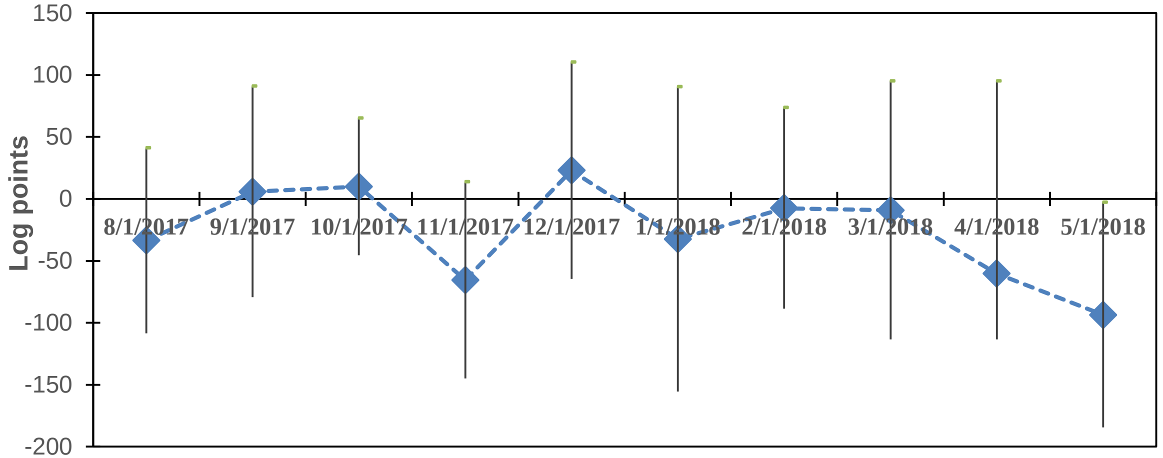
Group-Level Results

Category or Age Group	Percentage change in # of deals	Aggregate $mm per week per state change	$mm amount per deal change
Healthcare & Financial	-18.86%	-5.22m ($30.1m avg)	-56.6% ($24.79m avg)
All Other Categories	-16.69%	–	-28.4% ($20.39m avg)
0-3 Year-Old Firms	-19.02%	-0.9m ($14.82m avg)	-27.1% ($7.94m avg)

Robustness



Weekly # of deals pre-treatment test

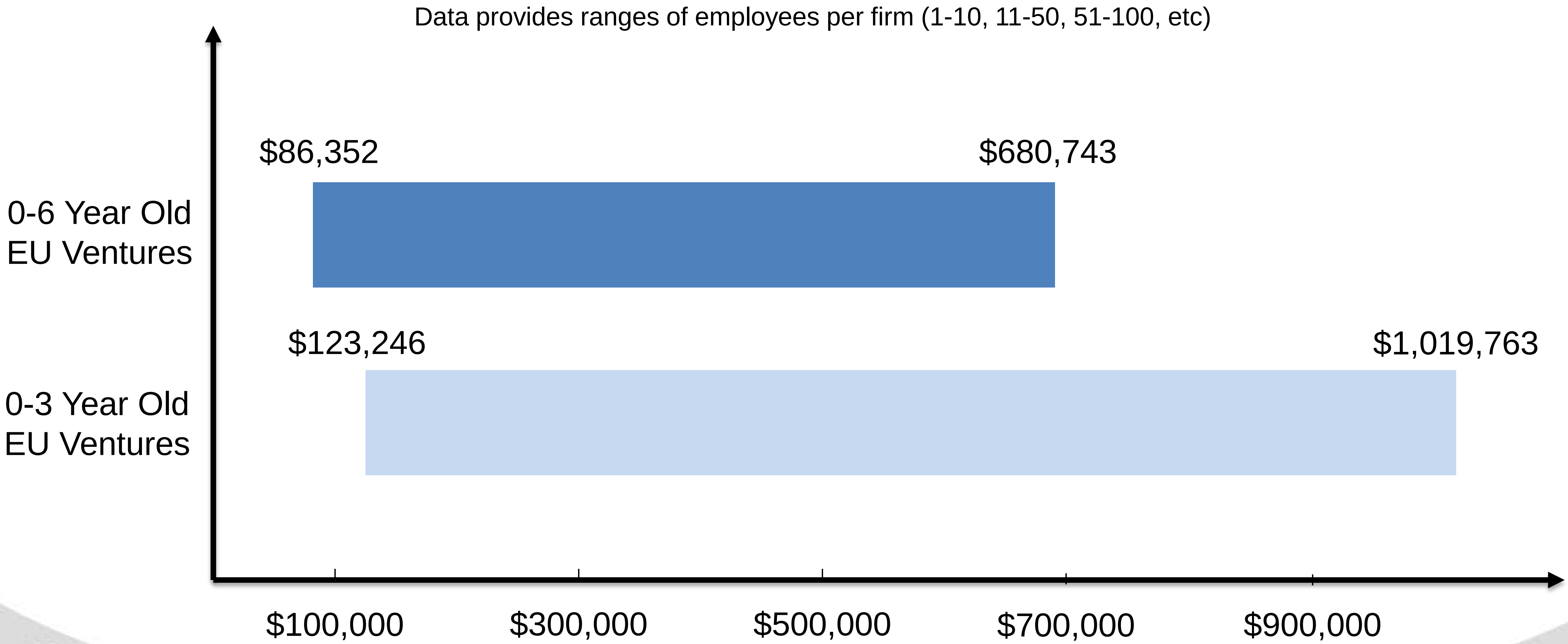


Weekly $ raised pre-treatment test

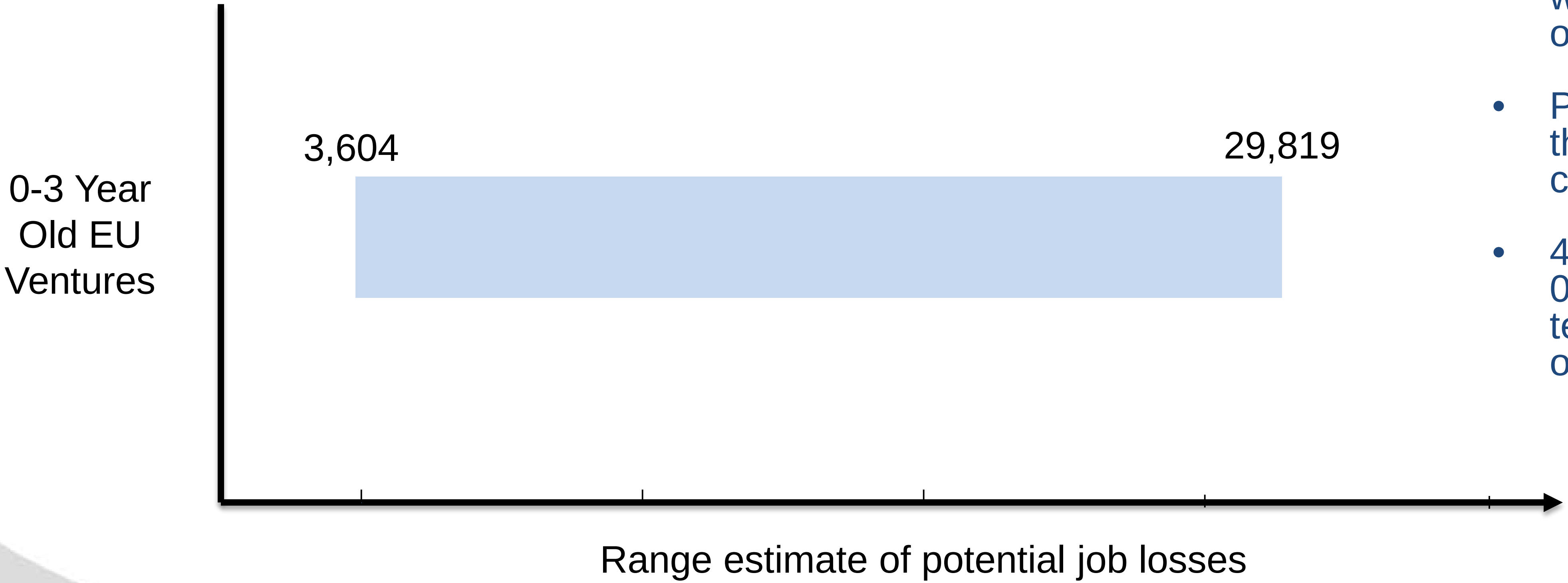
Vertical bands represent ± 1.96 times the standard error of each point estimate

- Dropped the month of May, tried other start weeks
- Top-coded observations to reduce influence of outliers
- Used unsupervised industry categorization
- Used other specifications including OLS

Average $ Raised Per EU Tech Employee



Rough Bound Estimates of Annual EU Tech Jobs Lost



- Could be indicative of wait-and-see approach, only observe short term
- Preliminary, back-of-the-envelope calculation
- 4.09-11.20% of overall 0-3 year old venture tech jobs in the EU in our sample

Preliminary Conclusions

- In the short run, GDPR has a pronounced negative effect on new EU venture financing, both on # of deals and amount per deal. More study is needed:
    - Post-GDPR sample is relatively short
    - Some investment dollars may be flowing to the US, could overstate results
    - Did not examine non-EU countries that serve EU, could understate results
    - Investors may fear rising costs / business obstacles / uncertainty – we can't distinguish
    - Small part of the bigger investment/venture picture (Crunchbase is not a complete universe)

- Ventures in the health and finance categories appear to be susceptible
    - Counterintuitive, US already has HIPPA (but at doctor's office, consent for service)
    - Calls for further study across categories when more data is available (e.g., with GDPR, service must be provided without consent, different penalties)

- Potential for technology and related job losses

Big Data Fails: Recent Research into the Surprising Ineffectiveness of Black-Box AI

Catherine Tucker

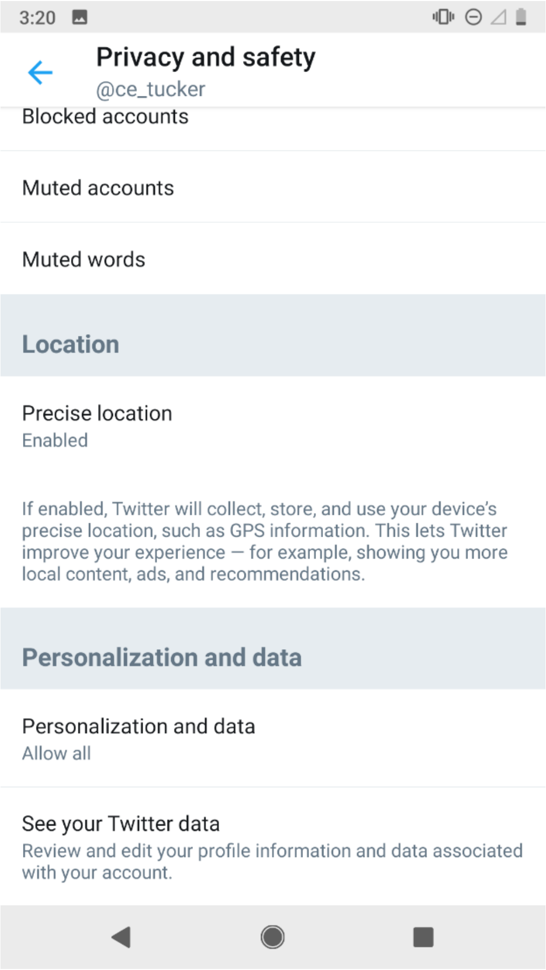
Massachusetts Institute of Technology

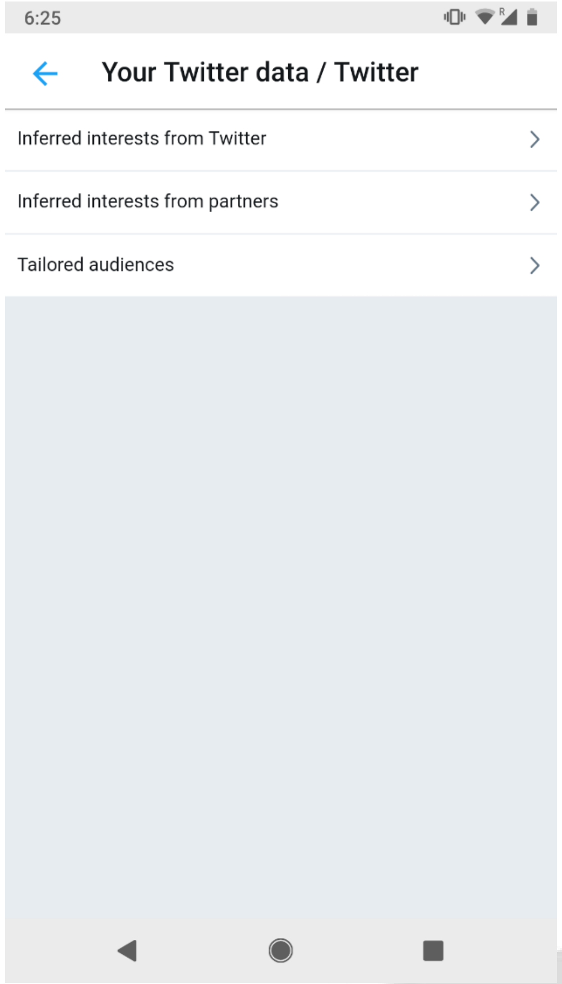
Sloan School of Management

What Kinds of Data Could A Website Use to Target an Ad?

- First Party: The website's knowledge of the consumer
- Second Party: Explicit sharing of data between partner websites
- Third Party: Data purchased from a third party source
  - Sometimes referred to as a "data broker"
  - This talk is about third-party data
  - Tackles whether the quantity of "data" drives accuracy of targeting in online advertising.

Let us start with an example: Twitter





6:28

Your Twitter data / Twitter

- [x] Demographics > Number of children: 1
- [x] Demographics > Presence in household: yes
- [x] Demographics > Presence of children: yes
- [x] Demographics > Residence: 6 - 8 years
- [x] Demographics > Seniors
- [ ] Demographics > Single
- [x] Demographics > Single parent
- [ ] Demographics > Soccer moms
- [x] Dining > Likely to dine at Chipotle Mexican Grill
- [x] Dining > Likely to dine at Starbucks

How effective is this profiling?

- This is a question I explore in a new called `How Effective Is Black-Box Digital Consumer Profiling And Audience Delivery?: Evidence from Field Studies'
- Joint Work with Nico Neumann and Tim Whitfield

In general consumer profiling online is surprisingly inaccurate

- In these studies, we focus on how well the consumer profiling and data broker ecosystem do in terms of identifying gender and age.
  - These are the most popular forms of data that advertisers use for targeting according to lotame survey (76% age, 61% gender)
- In our first study we asked ad platforms to show our ad to men between 25-54. They did this on average 59% of the time. Improvement of 184% relative to chance but is the ROI there?

In general consumer profiling online is surprisingly inaccurate

- In our second study, we asked for measurement of the audience of a particular website. We got a variety of answers about the proportion of men- 58%, 55%, 85% & 63%.

In general consumer profiling online is surprisingly inaccurate

- In our third study, we made the task easier by asking the gender of a particular cookie (or set of eyeballs).
  - Our source of truth was a survey which asked that cookie what gender they were.

Table: Study Three: Data Broker Accuracy at Profiling a Cookie They Have Data For

Data Broker	Number of Cookies	Gender Accuracy
A	1396	27.5
B	408	25.7
C	1777	35.2
D	495	56.4
E	527	48.8
F	480	47.9
G	562	46.8
H	1016	33.2
I	2336	33.6
J	14342	42.4
K	346	30.6
L	547	51.9
M	456	49.1
N	5099	62.7

Summary of Findings

- Algorithms that use data in online advertising do not appear very accurate
  - How can you really tell gender from browsing behavior?
  - Multiple people using a computer
- Little link between number of eyeballs that the data owners have data for and profiling accuracy
  - Quality of algorithms may matter more

Thank you

- cetucker@mit.edu

Corporate Data Ethics:
Risk Management for the Big Data Economy

Dennis Hirsch

The Ohio State University

Moritz College of Law

Free Speech and Data Privacy

Jane Bambauer

University of Arizona

James E. Rogers College of Law

Potential Conflicts with First Amendment Law

Interesting, but out of scope:

Restrictions on Commercial Speech

- Limitations on the use of personal data to craft marketing messages
- False or misleading assurances of privacy

Rebecca Tushnet, “The Meaning of Misleading”

Jane Bambauer, “Snake Oil Speech”

Potential Conflicts with First Amendment Law

Interesting, but out of scope:

Compelled Speech

- Just-in-time privacy notices

  Mandated disclosures that are not “purely factual and uncontroversial information”

  Jane Bambauer, *Are Privacy Policies Informational or Ideological?,* 66 DePaul Law Review 503 (2017)

Potential Conflicts with First Amendment Law

- Restrictions on Noncommercial Speech
  - Limitations on the transfer or sale of personal data
  - Limitations on the collection of personal data

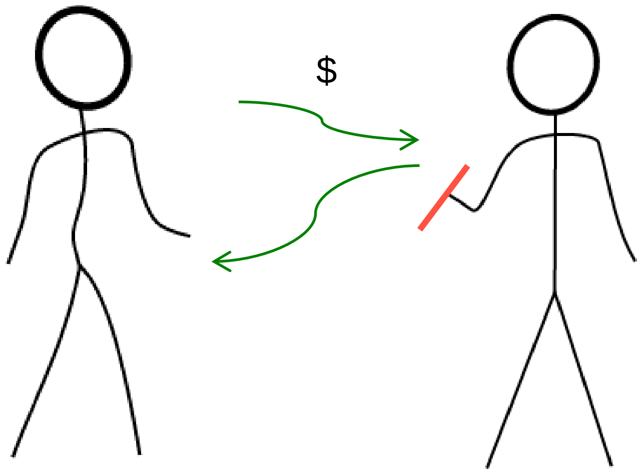
Jane Bambauer, *Is Data Speech?*,

Case Study:
California Consumer Privacy Act

- (1) The right of Californians to know what personal information is being collected about them.
- (2) The right of Californians to know whether their personal information is sold or disclosed and to whom.
- (3) The right of Californians to say no to the sale of personal information.
- (4) The right of Californians to access their personal information.
- (5) The right of Californians to equal service and price, even if they exercise their privacy rights.

Data in Transmission

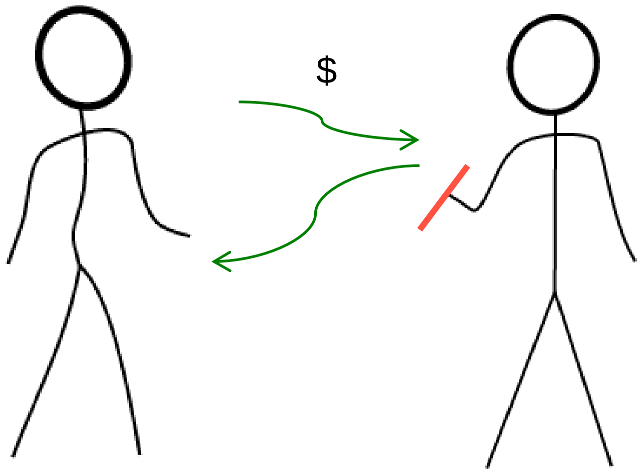
- First Amendment coverage
  - *Trans Union, U.S. West, Rubins v. Coors*
  - *Sorrell v. IMS*



"This Court has held that the creation and dissemination of information are speech within the meaning of the First Amendment. Facts, after all, are the beginning point for much of the speech that is most essential to advance human knowledge and to conduct human affairs."

Data in Transmission

- First Amendment coverage
  - *Trans Union, U.S. West, Rubins v. Coors*
  - *Sorrell v. IMS*

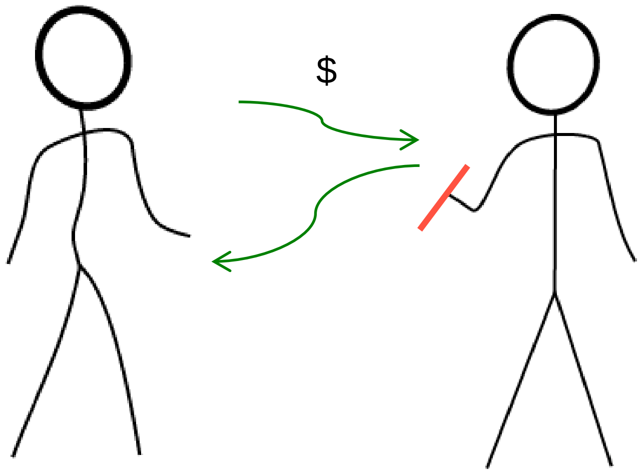


Speaker- and viewpoint-based discrimination

“Under Vermont’s law, pharmacies may share prescriber-identifying information with anyone for any reason save one: They must not allow the information to be used for marketing.”

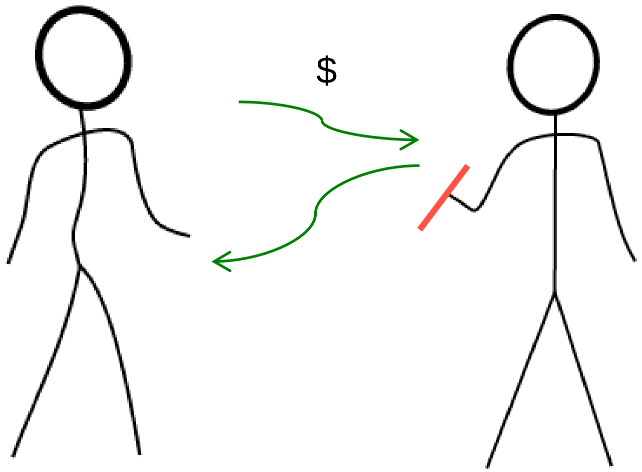
Data in Transmission

- First Amendment level of scrutiny
  - *Dun & Bradstreet v. Greenmoss Builders*
    - Intermediate scrutiny for speech of "purely private concern" like credit reports
  - *Reed v. Town of Gilbert*
    - Strict scrutiny for restrictions that make any distinction whatsoever based on the content of the speech



Data in Transmission

- First Amendment application



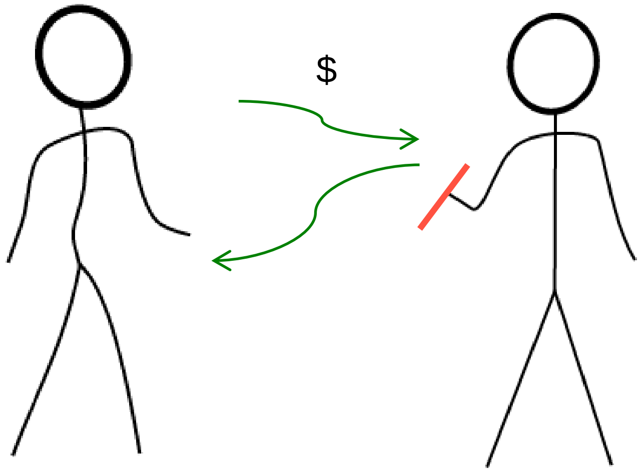
  - *Sorrell v. IMS*

    “Perhaps the State could have addressed physician confidentiality through ‘a more coherent policy.’”

    (using HIPAA as an example.)

Data in Transmission

- First Amendment application



  - *Trans Union v. FTC*

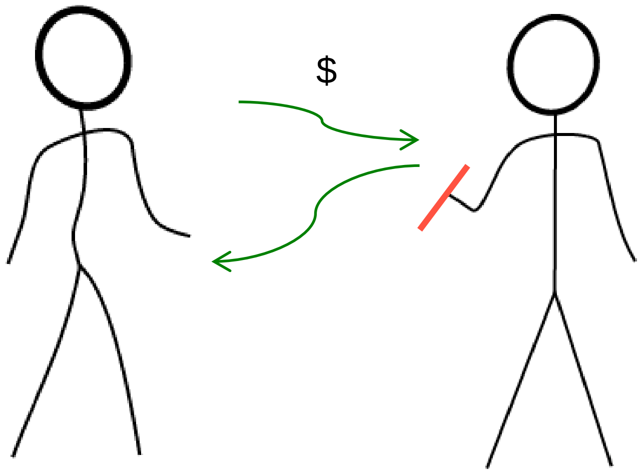
“the government cannot promote its interest (protection of personal financial data) except by regulating speech because the speech itself (dissemination of financial data) causes the very harm the government seeks to prevent. Thus, the FCRA unquestionably advances the identified state interest.

Data in Transmission

- First Amendment application

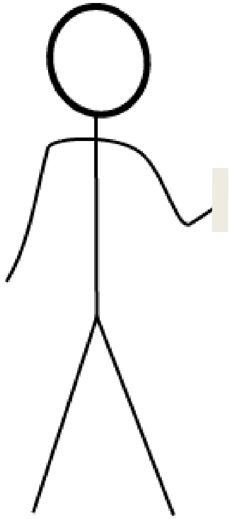
  Increasing pressure for proof of harm
  and careful tailoring:

  - Brown v. Entertainment Merchants Assoc.
  - United States v. Alvarez
  - United States v. Stevens



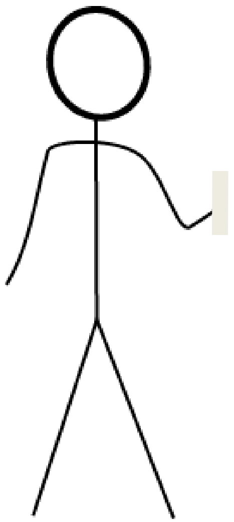
Data Collection

- *Dietemann v. Time, Inc.*



- *Bartnicki v. Vopper*

Data Collection



- However, the reason to limit data collection is to restrict knowledge-creation and downstream communications

- Right to record police cases
- Right to record and ag-gag cases

Avoiding Constitutional Conflict

- Define and protect interests in seclusion and confidentiality
- Prohibit disfavored uses of information

FTC Experience with Data Markets

Haidee L. Schwartz*

Bureau of Competition

Federal Trade Commission

* This presentation and my remarks are my own and do not necessarily represent those of the Commission or any individual Commissioner

The Many Dimensions of Data

- So how do staff at the FTC think about data?
- Investigations and cases are always very fact-specific and we will look at all aspects of data
  - Is data a product or an input?
  - Is the competition with the data or for the data?
  - Is the data unique, broadly available, or replicable?
    - Especially important for assessing likely entry

Mergers Involving Data

- In a merger, how are the companies using data?
  - Data can be a product, such as in the case of two database companies that compete to sell data products
  - Data can be an input for firms that provide analysis, verification, or other analytics
  - Data can affect entry conditions, making it more or less difficult for a firm to enter and compete
- Focus of merger analysis: is the data of the merging firms a key differentiator in how they compete? If so, are there other firms (in the market or likely to enter) that also have access to data and could replace the competition lost due to the merger?

FTC History of Database Merger Cases

- FTC has a long history of enforcement on database cases, *see Automatic Data Processing, Inc.*, Dkt. No. 9282 (complaint issued in 1996), settlement with divestiture buyer receiving an unrestricted license to proprietary database for auto parts
- FTC has investigated, and often challenged, mergers involving database assets across a wide range of industries

CoreLogic / DataQuick (2014)
Data as a Product

Product: National Assessor and Recorder Bulk Data

- CoreLogic: vast database of reformatted public record data and information on properties in the U.S.
- DataQuick: significant historical data; unique rights to relicense CoreLogic's ongoing data in bulk

- Sold to companies for different uses (risk and fraud management tools, valuation models, customer-facing websites such as Zillow)
- **Order created new competitor**
  - Required CoreLogic to license bulk data to RealtyTrac for relicensing
- **Order modified in 2018**
  - CoreLogic supplied insufficient data

CoreLogic / DataQuick (2014)

Key Takeaways

- Data as a product and as a divestiture asset
- Scope of historical database created barrier to entry
  - Ongoing data an easier hurdle than historical data
- Remedy challenges
  - Identifying the precise data for divestiture
  - Buyer due diligence may not be enough
  - Reliance on parties' representations

Verisk / EagleView (2014)
Data as a key input

Product: Rooftop Aerial Measurement Products (RAMPs) for Insurance Purposes

- Verisk: leading provider of claims estimation software that integrated with RAMPs, and a recent entrant into RAMPs
- EagleView: dominant provider of RAMPs

- **Aerial image libraries are a key input** to RAMPs
  - Verisk image library was much smaller than EagleView's

- **Transaction abandoned** after FTC challenge

Verisk / EagleView (2014)

Key Takeaways:

- Data as a necessary input into relevant product
  - But: scope of Verisk database not dispositive of Verisk's competitive significance
  - Verisk's recent success in relevant market (RAMPs) more probative
- Position in adjacent market provided Verisk with a unique ability to overcome data-related entry barriers
- Complaint alleged innovation effects related to data coverage and quality

CCC / Mitchell (2009)
Access to data as an entry barrier

Products:

(1) Estimatics, databases used to generate repair estimates for cars

(2) Total loss valuation systems, used to determine when a vehicle is totaled

- At the time of the merger, Big Three – CCC, Audatex, and Mitchell – held ~ 99% of estimatics market; Web-Est and Applied Computer Resources were fringe players
- Total loss valuation systems (TLV); Big Three accounted for more than 90% of market; Mitchell entered in 2005 and had a significantly smaller share
- Customers were insurers (estimatics and TLV) and repair facilities, such as service stations (estimatics)

CCC / Mitchell: Database Dynamics

- Primary component of Estimatics and TLV offerings: databases (parts and labor, and data from dealerships and publications respectively) and software
- CCC obtained an exclusive license to the Hearst Business Publishing, Inc. “Motor” database in 1998
- Audatex and Mitchell each had their own proprietary databases they developed over many years
- Web-Est licensed Mitchell database, but under restrictive conditions that limited its ability to compete

CCC / Mitchell: The Proposed Fix

CCC offered to relinquish its exclusive rights to Hearst's Motor database, giving any new entrant access to a comprehensive, fully updated database; Mitchell would remove restrictions on Web-Est and continue database license

- Judge found that availability of database would reduce most critical barrier to entry, but significant barriers still remain
  - In addition to database, competitor would need to develop software
  - Other barriers included existing customer relationships (large insurers were relatively sticky); need to establish a track record; and sufficient scale (including technical and customer service employees)
- Judge noted Web-Est leader had great entrepreneurship and experience, but company had only 10-12 employees and modest project revenues; growth curve too long and steep

Microsoft / LinkedIn (2016)
Is the data unique?

Microsoft: strong position in operating systems for personal computers and productivity software

LinkedIn: strong position in professional social networks; database of individual professional information

- US investigated but did not take action
- EC concluded merger did not raise competitive concerns related to data used in online advertising market because:
  - Other sources existed for similar data
  - Pre-merger each company provided limited to no access to their respective user data (in full) to third parties for advertising purposes
  - The parties had relatively low combined share of online advertising

Microsoft / LinkedIn: EC Action

EC required several commitments, but not related to data/database assets:

- Preserve ability of OEMs and distributors not to install LinkedIn with Windows, and ability of users to remove it
- Interoperability with Microsoft suite of products for competing professional social network service providers
- Provide access to competing professional social network service providers to Microsoft gateway for software developers
- Commitments apply in the EEA for five years, with a monitor

Takeaways: Competition Analysis

- **Current antitrust analysis accounts for how firms compete using data as a product or input, or as a tool for making decisions**
  - May have additional complexity if data is proprietary or subject to copyright/IP protection
- Data sets can be highly differentiated; non-price factors of competition important (e.g., quality, innovation)
- Data is often combined with analytics to make information useful in a business setting (i.e., data is an input)
- Dynamic nature of data markets requires attention to other sources of reliable information – the data feed is critical
  - Who owns the data?
  - Is the data unique? What other sources are available at similar cost? Is it difficult or costly to replicate, or are there other barriers to replication?
  - Do incumbents have a data advantage?
- Can be challenging to balance competition to reduce prices and competition to improve products
- At the FTC, we recognize that data sets that include consumer data require special attention.
  - Companies need to keep commitments to protect consumer data.

Takeaways: Remedies

- Preference for structural remedies in merger cases
  - Divest (or clone) database vs. license
  - Continuing supply of new data
  - Defining terms – what is the key data?
  - Ability to manipulate data going forward (freedom of operation)
- Need to handle IP/copyright issues
- May need ongoing support and other behavioral conditions to give new company opportunities to compete with incumbents
  - Overcome customers' reluctance to switch

Thank You,

Join Us Tomorrow