

Good data save lives: But what are good data?

Samuel V. Scarpino, PhD

Vice President, Pathogen Surveillance
The Rockefeller Foundation

External Faculty

Santa Fe Institute

Vermont Complex Systems Center

Affiliate Faculty, Physics & CS

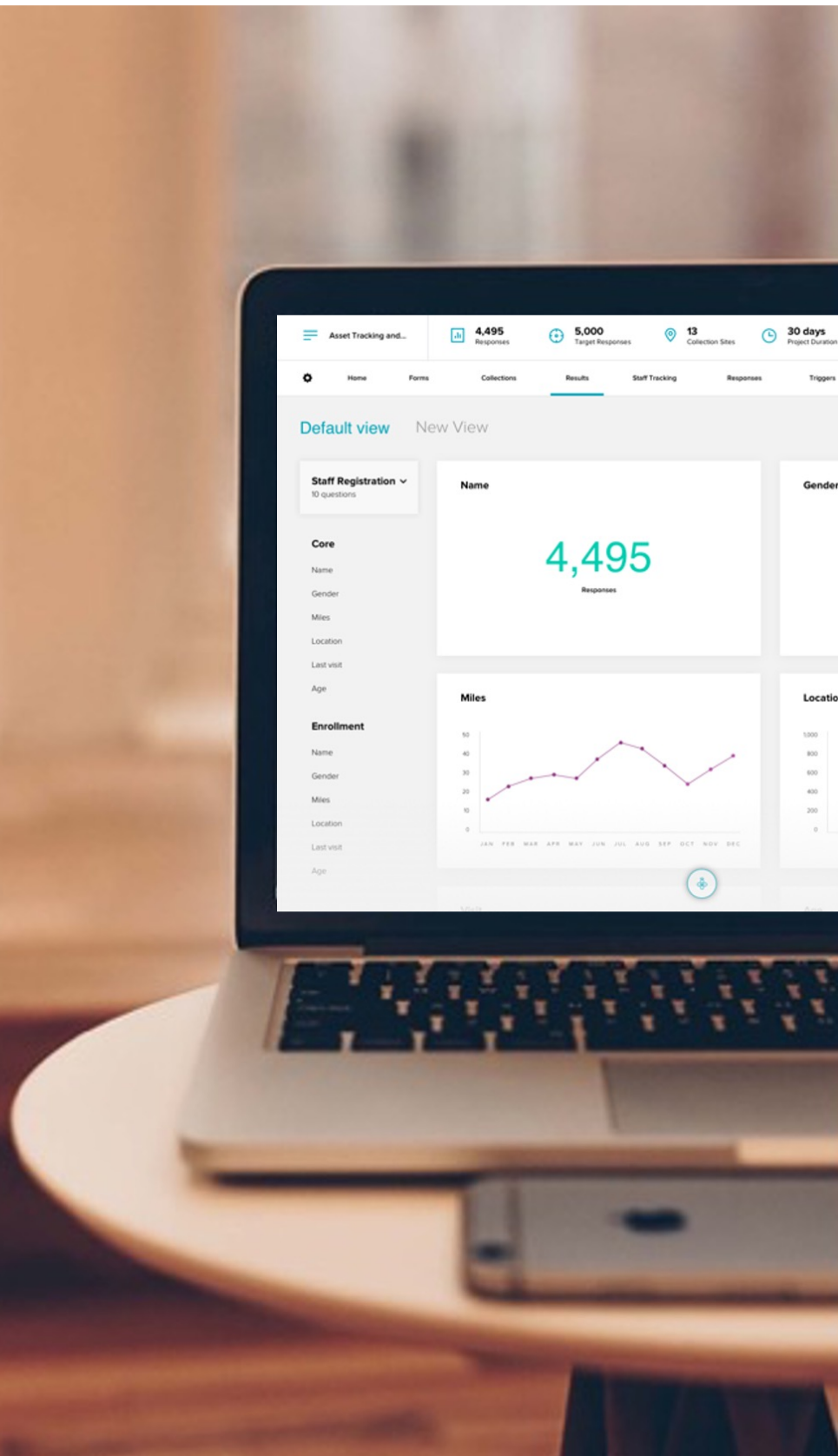
Northeastern University

Co-Founder, Global.health

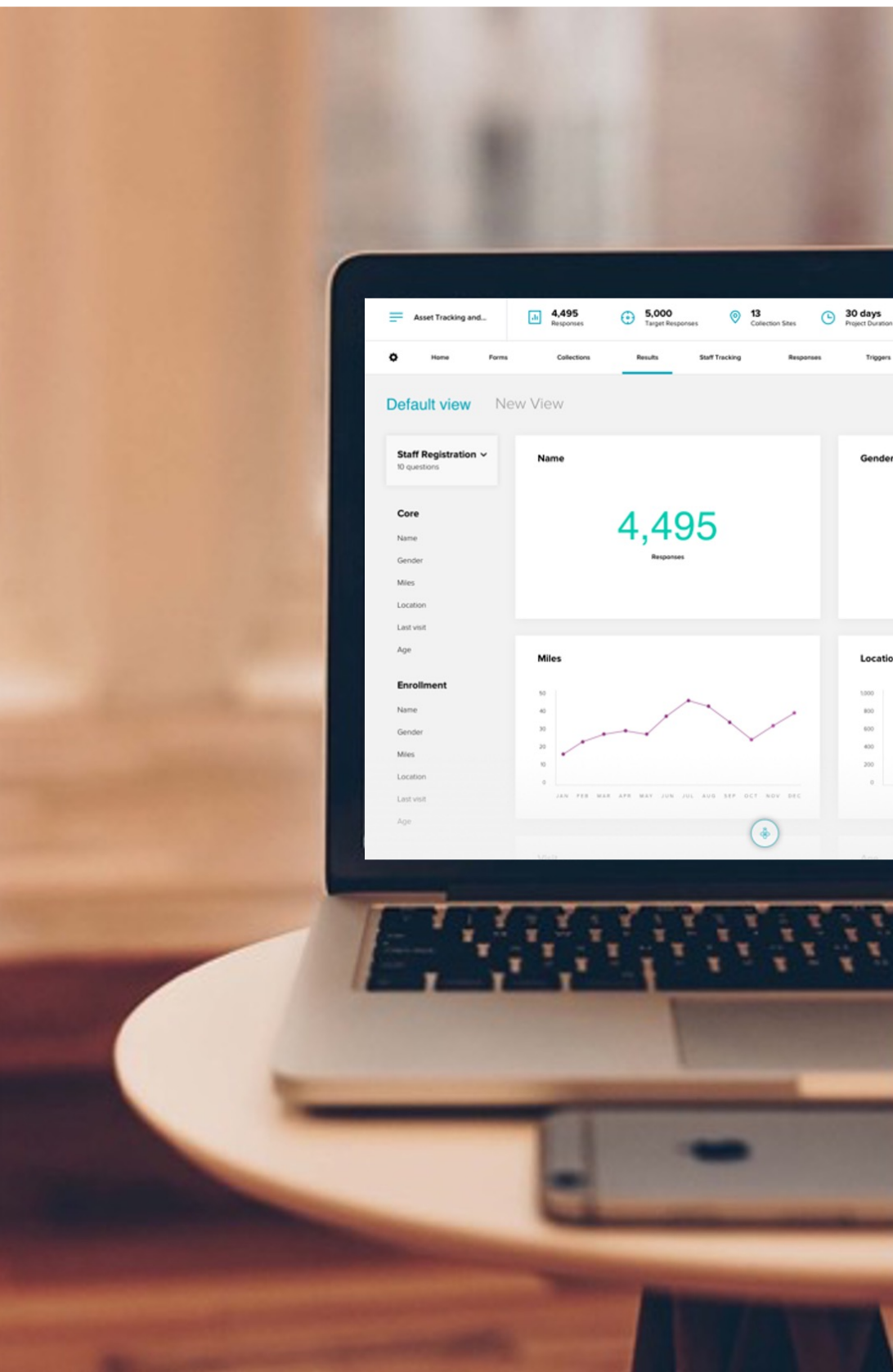
@svscarpino

scarpino.github.io





What are good data?



What are good data?

What are
good data?

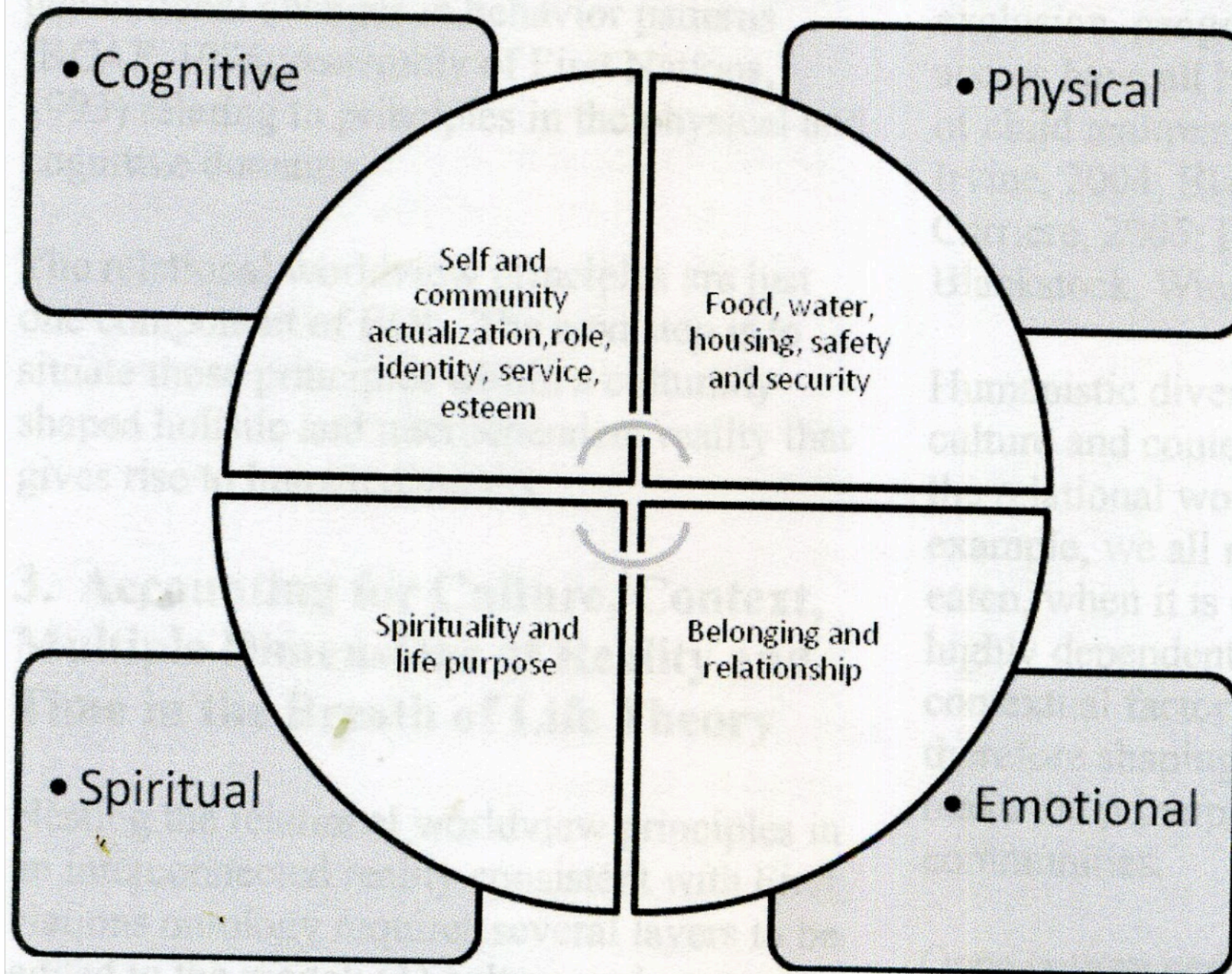


Modified from Maslow's hierarchy of needs

“Change is a constant, inevitable, cyclical, and dynamic part of the human experience”

What
good

Figure 3: Cross's worldview principles (2007) oriented in the holistic model



TA NEEDS

y of needs

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped
II	Machine-readable, but proprietary (e.g., Excel)	Data can be analyzed using proprietary software	Data are still trapped. Stuck with expensive software

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped
II	Machine-readable, but proprietary (e.g., Excel)	Data can be analyzed using proprietary software	Data are still trapped. Stuck with expensive software
III	Machine-readable and non-proprietary (e.g., CSV)	Data can be analyzed using most any software	Data are open, but can be challenging to analyze/share

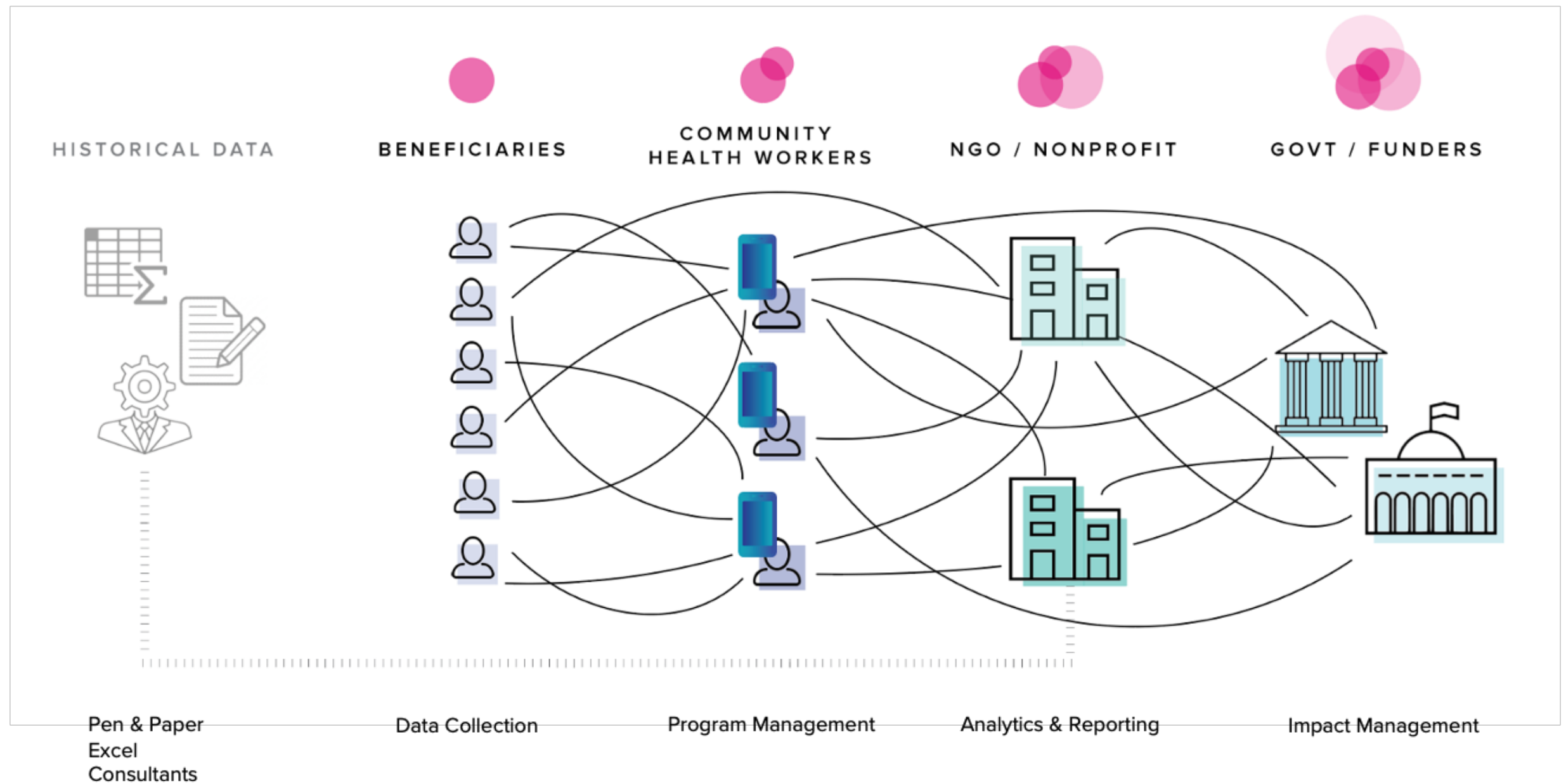
Adapted from The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences by Kitchin

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped
II	Machine-readable, but proprietary (e.g., Excel)	Data can be analyzed using proprietary software	Data are still trapped. Stuck with expensive software
III	Machine-readable and non-proprietary (e.g., CSV)	Data can be analyzed using most any software	Data are open, but can be challenging to analyze/share
IV	Machine-readable, non-proprietary, and linked to metadata/schema/etc.	Data can be readily shared and are ready for analytics/ML/AI	Requires technology

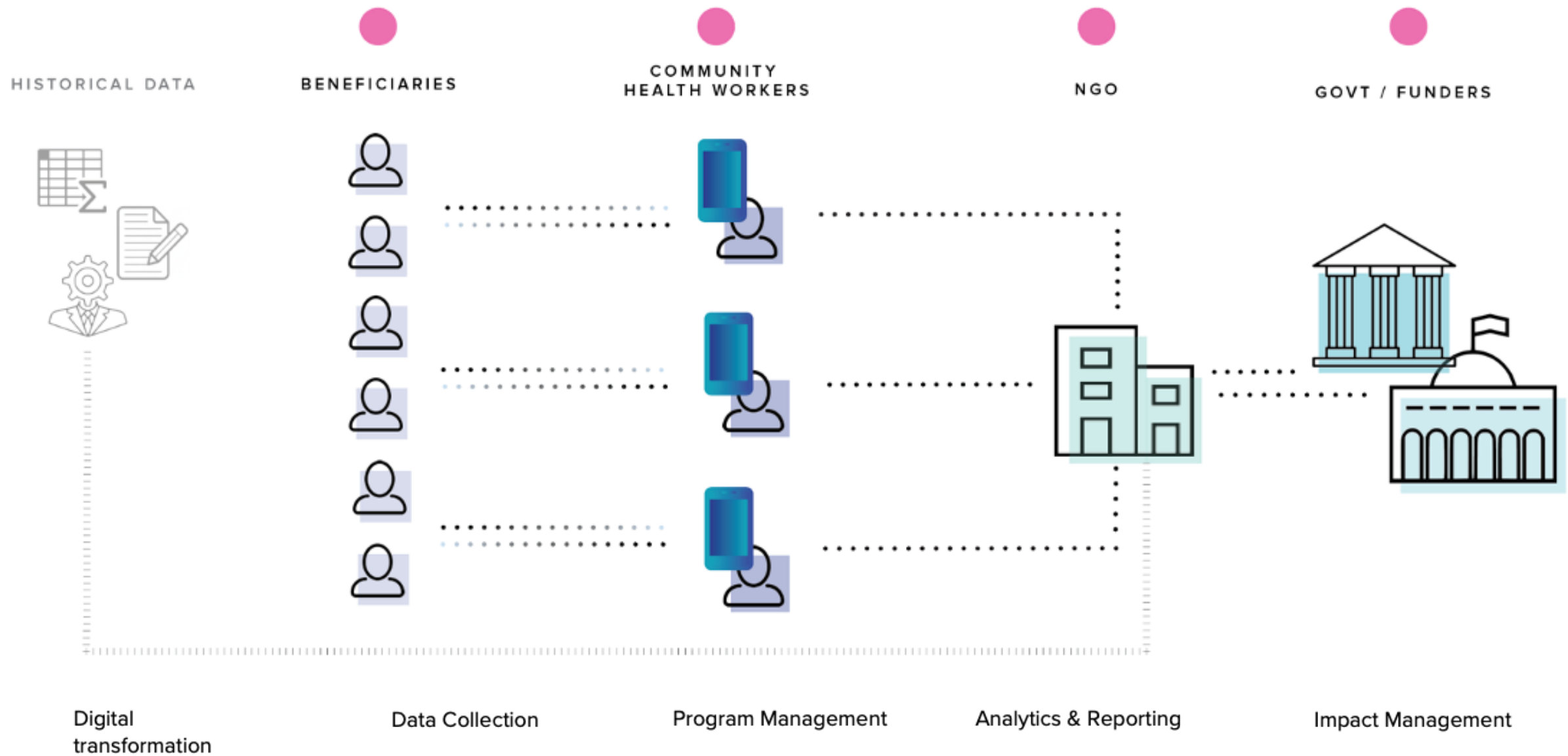
Adapted from The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences by Kitchin

Current state of data



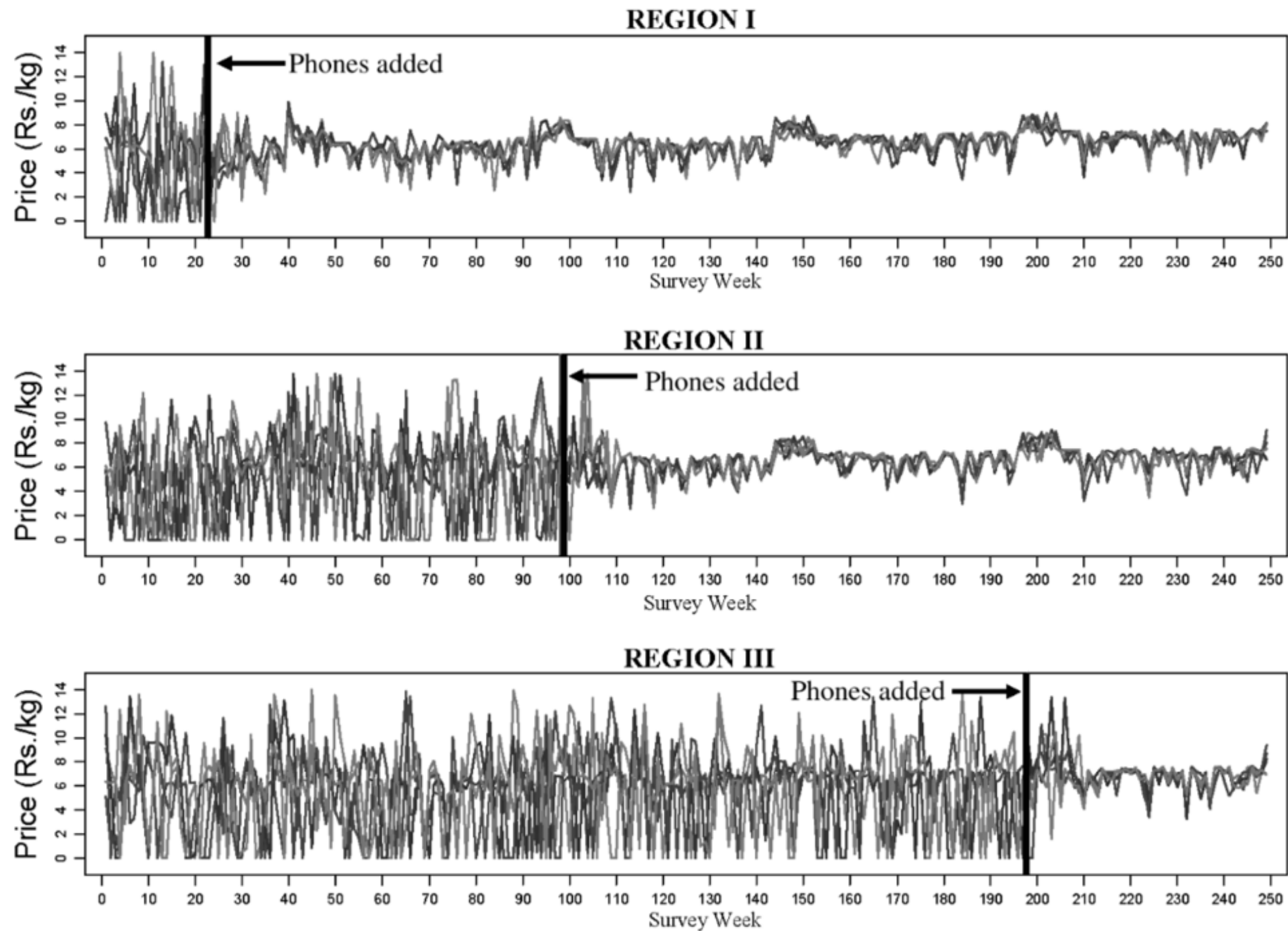
Kristen Clark, Dharma Platform 2019

Good data can improve impact measurement



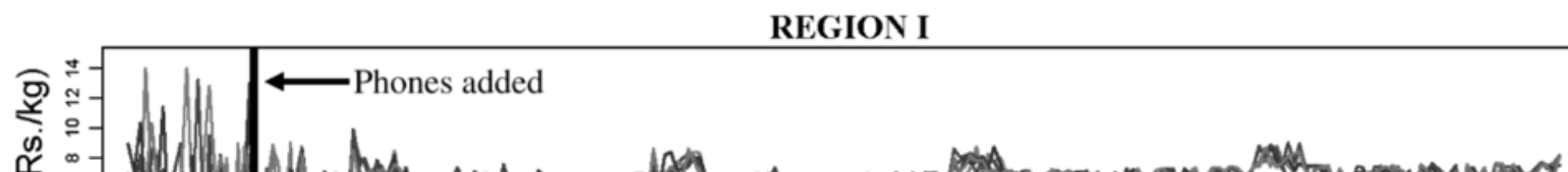
Kristen Clark, Dharma Platform 2019

We know the story of mobile phones & fishing



Jensen 2007

But why don't the results generalize?



Research Article

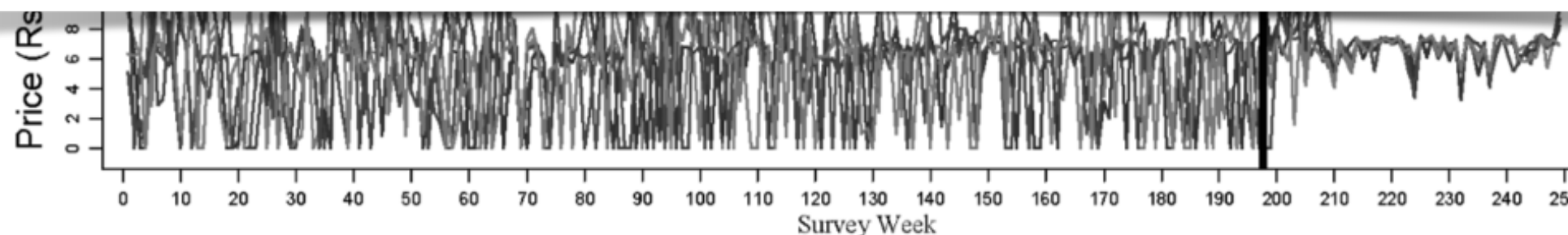
On the Importance of Price Information to Fishers and to Economists: Revisiting Mobile Phone Use Among Fishers in Kerala

Janaki Srinivasan

International Institute of Information Technology–Bangalore

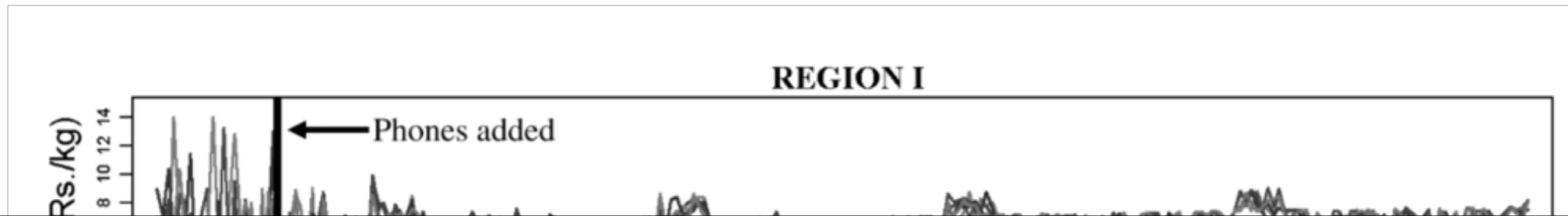
Jenna Burrell¹

University of California, Berkeley



Jensen 2007

Meta-data were under-valued



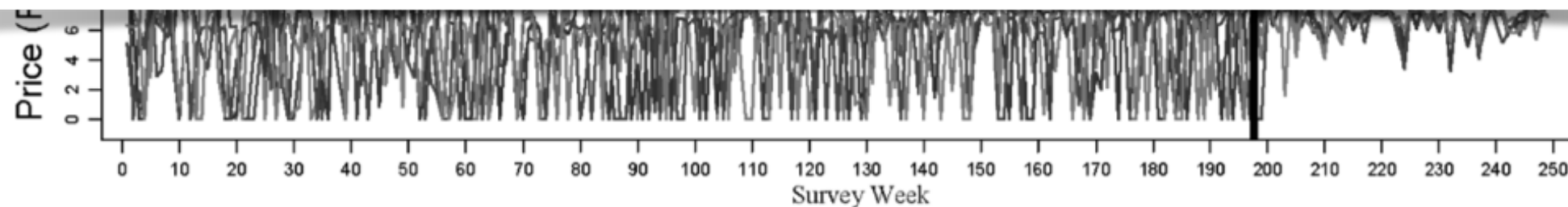
This is the extent to which the actual practice of mobile phone use among north Kerala's fishers is specified in the article. While the quantitative data that forms the substance of Jensen's argument about the reduction of price dispersion is collected systematically, details on exactly how fishers use phones are sparse in his account and lack the same kind of transparency about how such insights were acquired. Some of these details are deprioritized to footnotes. This is (broadly) a reflection of what counts as evidence in econometric analysis. Collected prices are data, but details on processes are merely background or explanatory material.

Janaki Srinivasan

International Institute of Information Technology–Bangalore

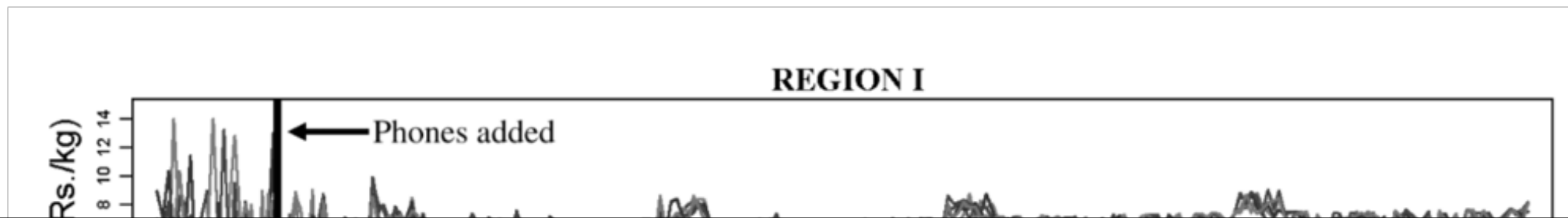
Jenna Burrell¹

University of California, Berkeley



Jensen 2007

Meta-data were under-valued



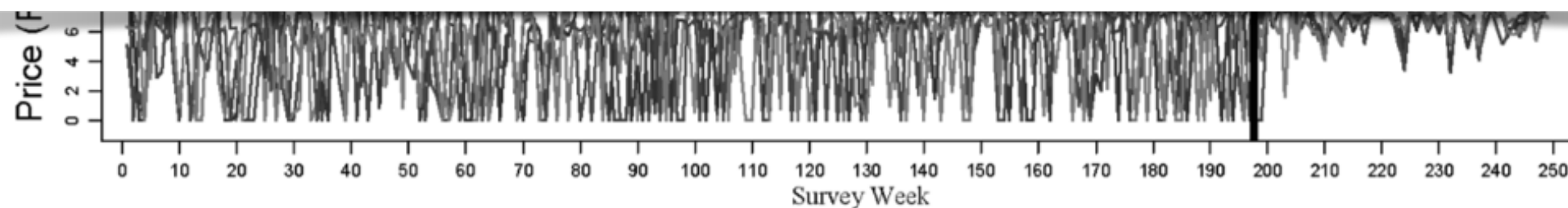
This is the extent to which the actual practice of mobile phone use among north Kerala's fishers is specified in the article. While the quantitative data that forms the substance of Jensen's argument about the reduction of price dispersion is collected systematically, details on exactly how fishers use phones are sparse in his account and lack the same kind of transparency about how such insights were acquired. Some of these details are deprioritized to footnotes. This is (broadly) a reflection of what counts as evidence in econometric analysis. Collected prices are data, but details on processes are merely background or explanatory material.

Janaki Srinivasan

International Institute of Information Technology–Bangalore

Jenna Burrell¹

University of California, Berkeley



Jensen 2007

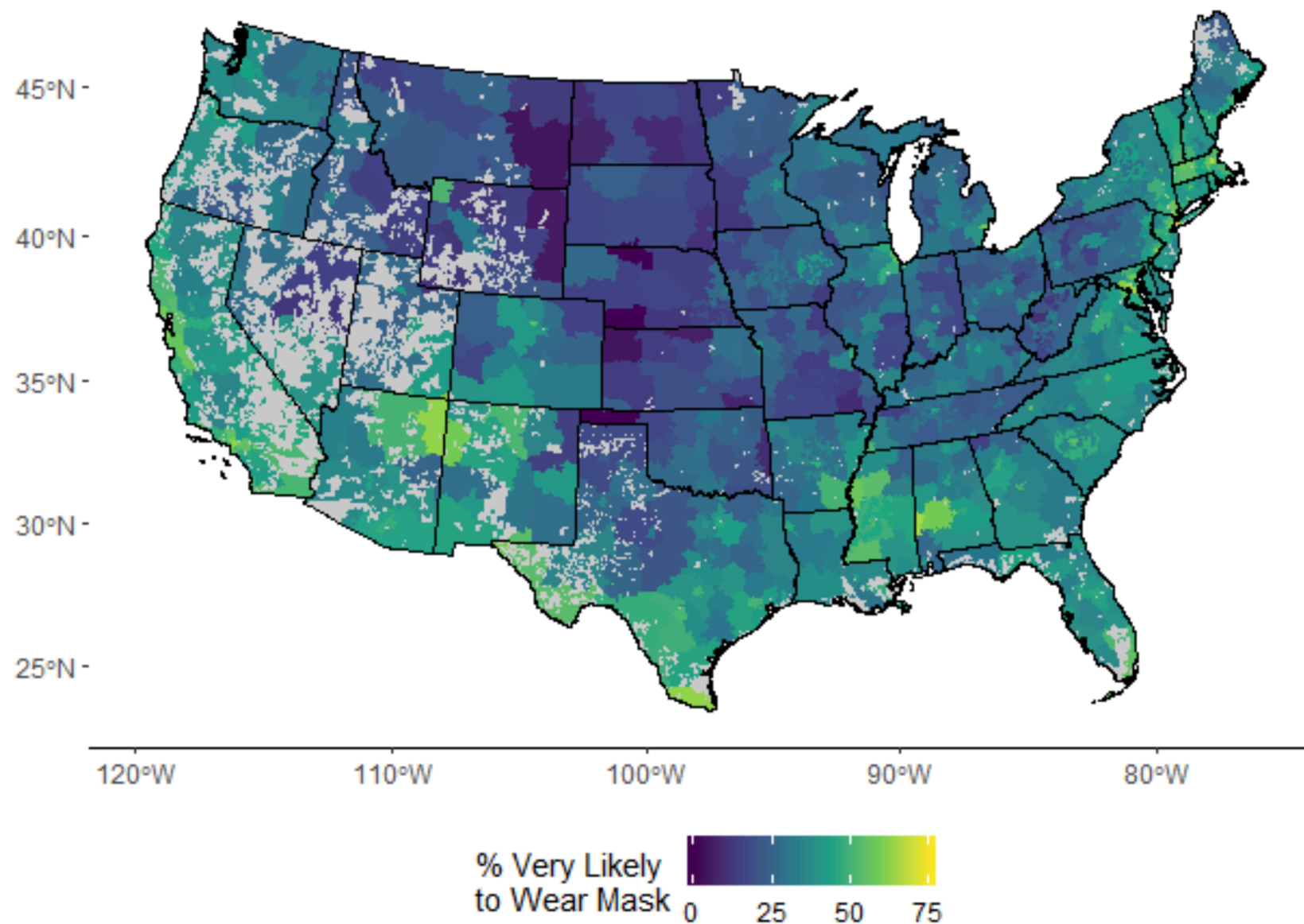
A similar effect was seen with mask mandates

Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study

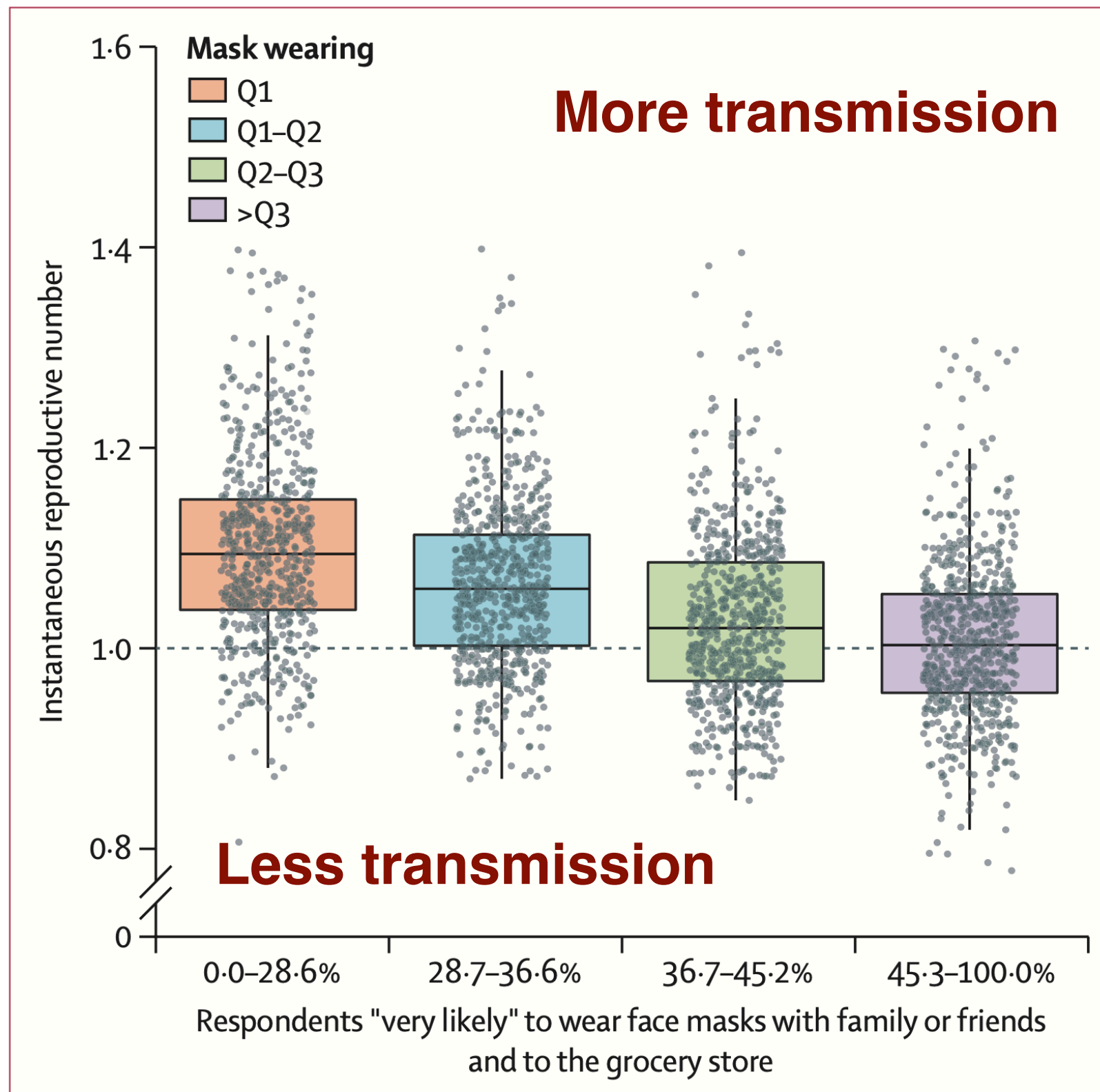
Benjamin Rader, Laura F White, Michael R Burns, Jack Chen, Joseph Brilliant, Jon Cohen, Jeffrey Shaman, Larry Brilliant, Moritz U G Kraemer, Jared B Hawkins, Samuel V Scarpino, Christina M Astley, John S Brownstein



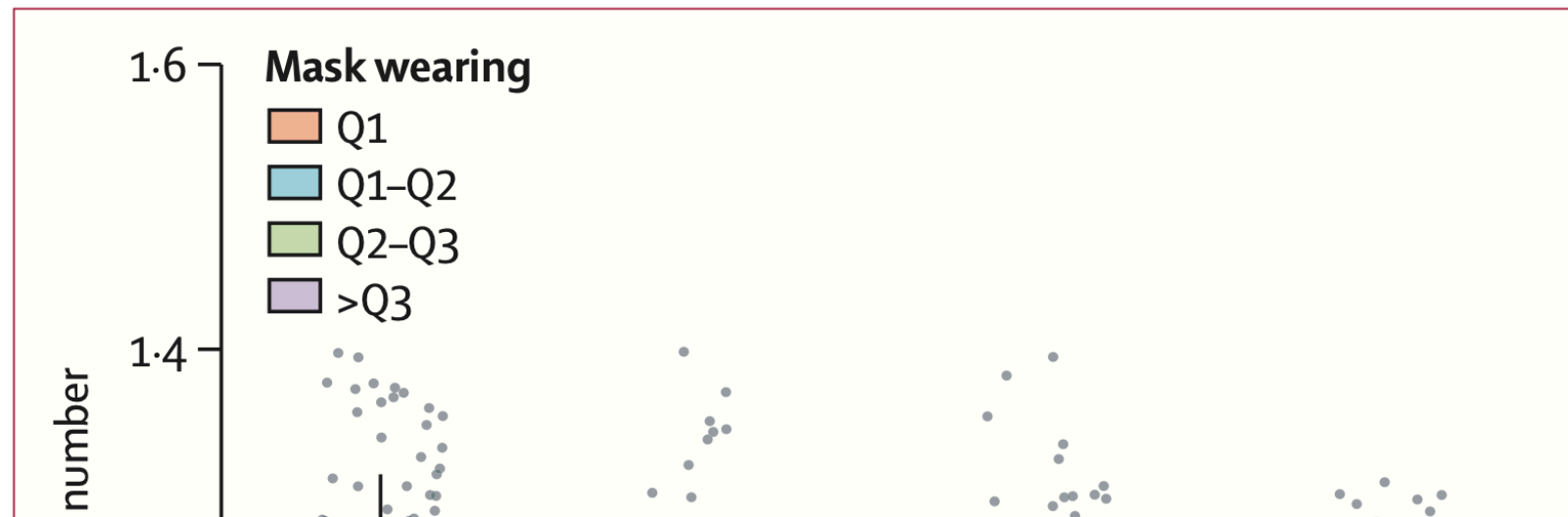
~400k responses across the USA



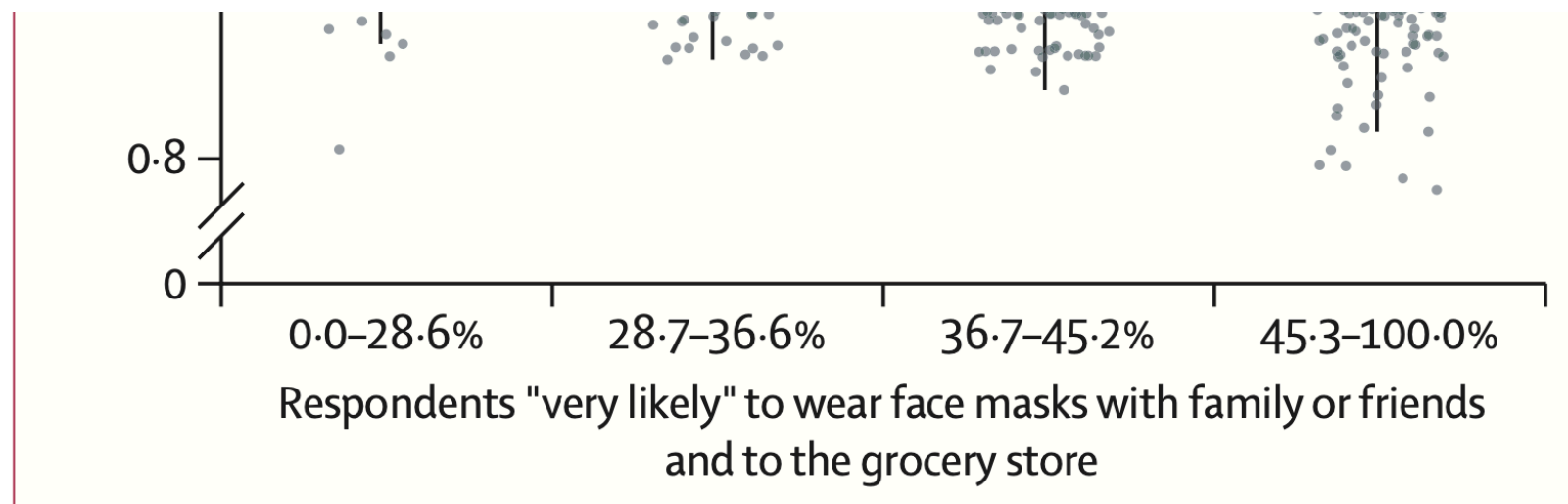
Mask wearing slowed COVID-19 transmission



But wasn't strongly influenced by mandates



Segmented regression analysis of reported mask-wearing showed no statistically significant change in the slope after mandates were introduced;



And with vaccine mandates

PNAS

RESEARCH ARTICLE

SOCIAL SCIENCE

 OPEN ACCESS



Opposition to voluntary and mandated COVID-19 vaccination as a dynamic process: Evidence and policy implications of changing beliefs

Katrin Schmelz^{a,b,c,1}  and Samuel Bowles^c 

Edited by Margaret Levi, Stanford University, Stanford, CA; received October 12, 2021; accepted January 24, 2022

What mattered most was trust in public institutions

PNAS

RESEARCH ARTICLE

SOCIAL SCIENCE

 OPEN ACCESS



Opposi
a dyna
changi

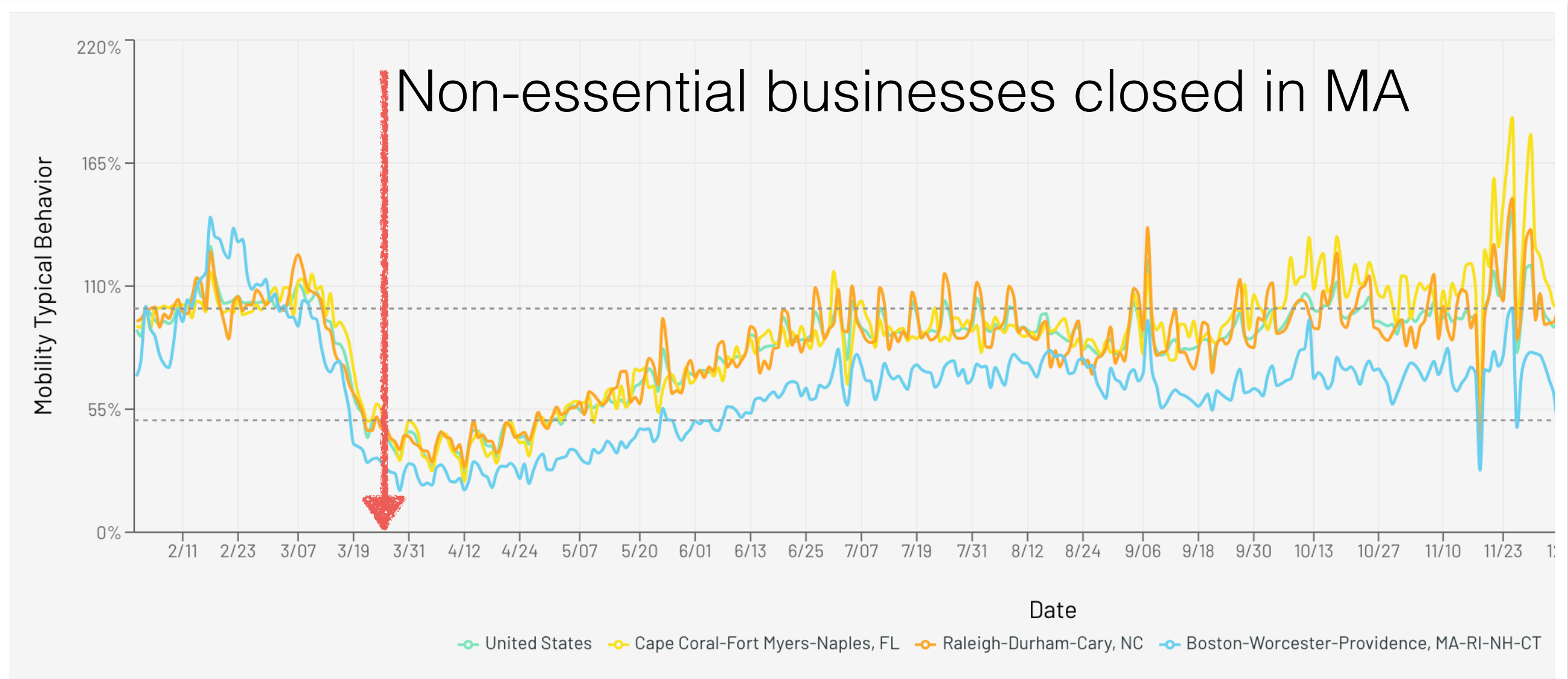
cine attitudes may apply generally across societal groups. What differentiates them from others are their beliefs about vaccination effectiveness, ~~their trust in public institutions~~, and whether they perceive enforced vaccination as a restriction on their freedom. We find that changing these beliefs is both possible and necessary to increase vaccine willingness, even in the case of mandates. An inference is that well-designed policies of persuasion and enforcement will be complementary, not alternatives.

on as

Katrin Schmelz^{a,b}

Edited by Margaret Levi, Stanford University, Stanford, CA; received October 12, 2021; accepted January 24, 2022

And with mobility



So how do we get to good data?



So how do we get to good data?





People

Process

Tech

Good data

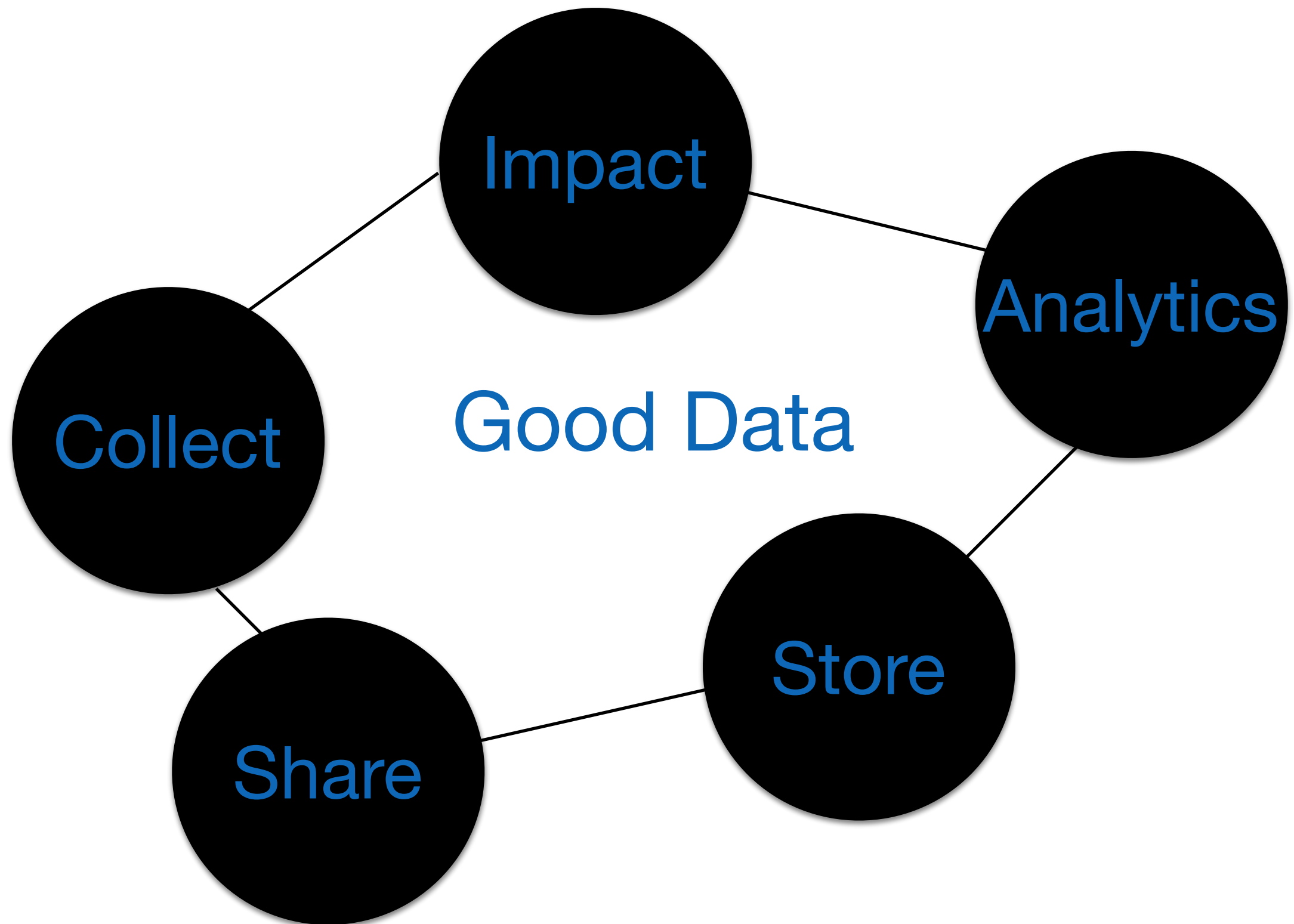


Process

People

Tech

Towards a “network data science”



Never bring data to a story fight

-Prof. Peter Dodds



Thank you!

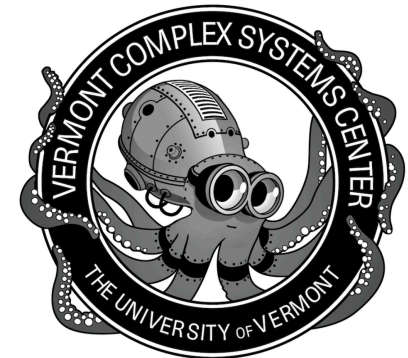
Samuel V. Scarpino, PhD
Vice President, Pathogen Surveillance
The Rockefeller Foundation

External Faculty
Santa Fe Institute
Vermont Complex Systems Center

Affiliate Assistant Professor
Northeastern University

Co-Founder, Global.health

@svscarpino
scarpino.github.io
sscarpin@rockfound.org



SANTA FE
INSTITUTE

