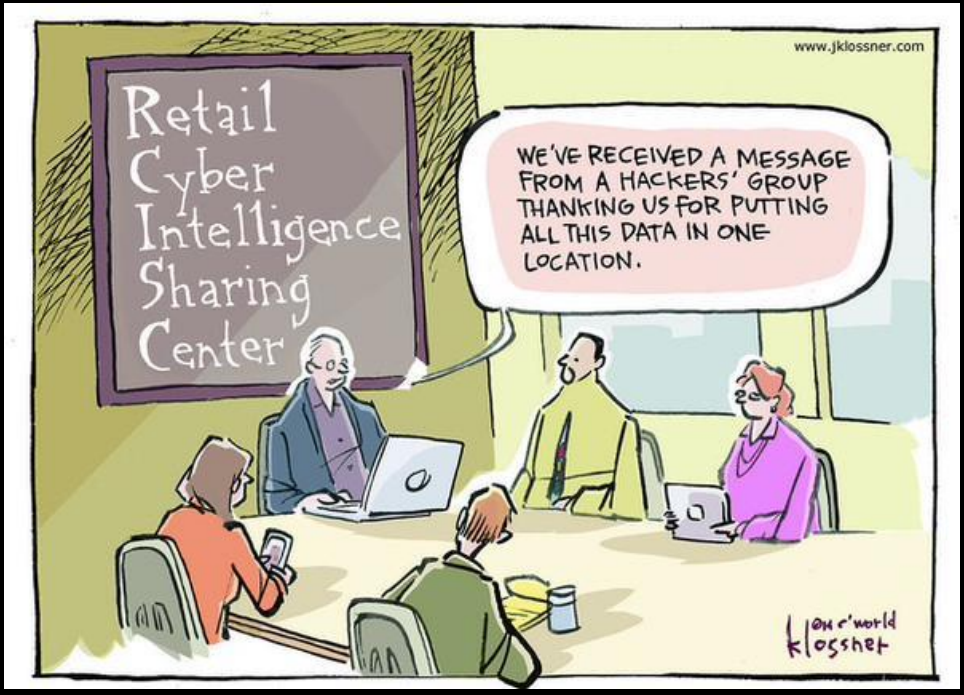
NIDDEL™

Data-Driven Threat Intelligence: Metrics on Indicator Dissemination and Sharing (#ddti)

Alex Pinto
Chief Data Scientist
MLSec Project / Niddel
@alexcpsec
@MLSecProject @NiddelCorp

Agenda

- Previously on #ddti
- Challenges at TI Sharing
- Measuring TI Sharing
- The Future of Sharing



This is a data-driven talk!

Please check your anecdotes at the door


PREVIOUSLY ON
GAME OF THRONES
Data-Driven Threat Intelligence

Previously on #ddti

- Useful Methods and Measurements for Handling Indicators
  - Analysis of Threat Intelligence Feeds
  - Indirectly, a methodology for analyzing TI Providers

- Combine (https://github.com/mlsecproject/combine)
  - Gathers TI data (ip/host) from Internet and local files

- TIQ-Test (https://github.com/mlsecproject/tiq-test)
  - Runs statistical summaries and tests on TI feeds

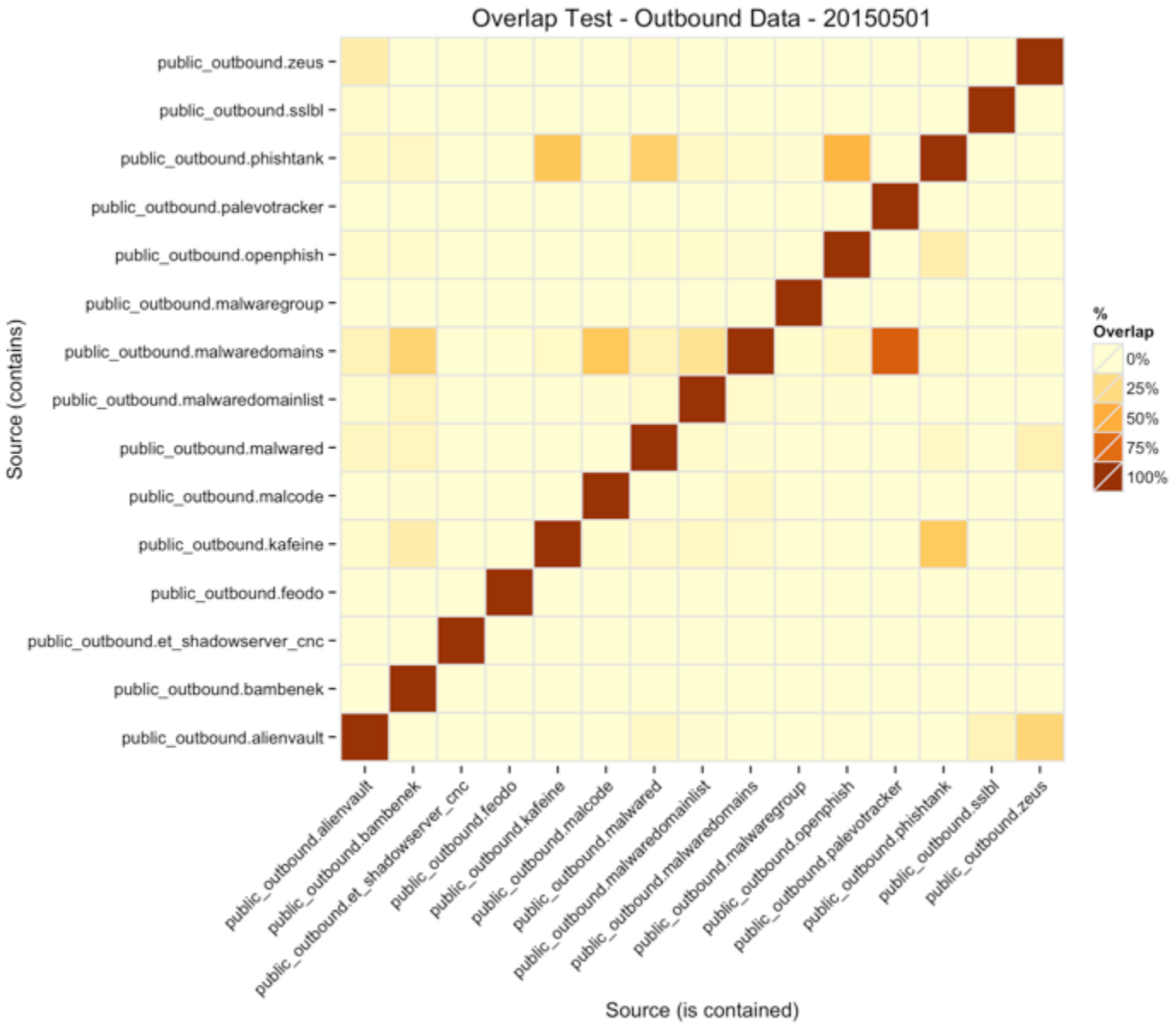
TIQ-TEST - Tons of Threat-y Tests

Putting this threat intel data to work

- ~~NOVELTY – How often do the feeds update themselves?~~
- ~~AGING – How long does an indicator sit on a feed?~~
- ~~POPULATION – How does this population distribution compare to my data?~~
- OVERLAP – How do the indicators compare to the ones you got?
- UNIQUENESS – How many indicators are found only on one feed?

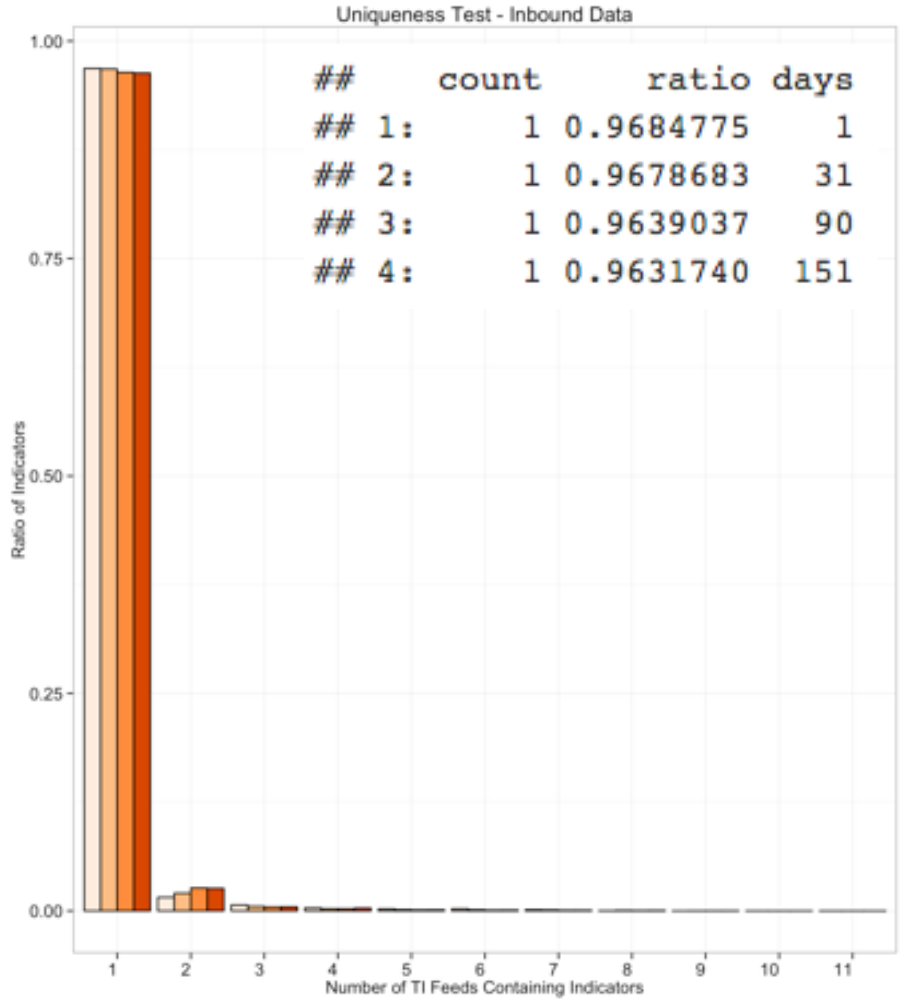
Overlap Test

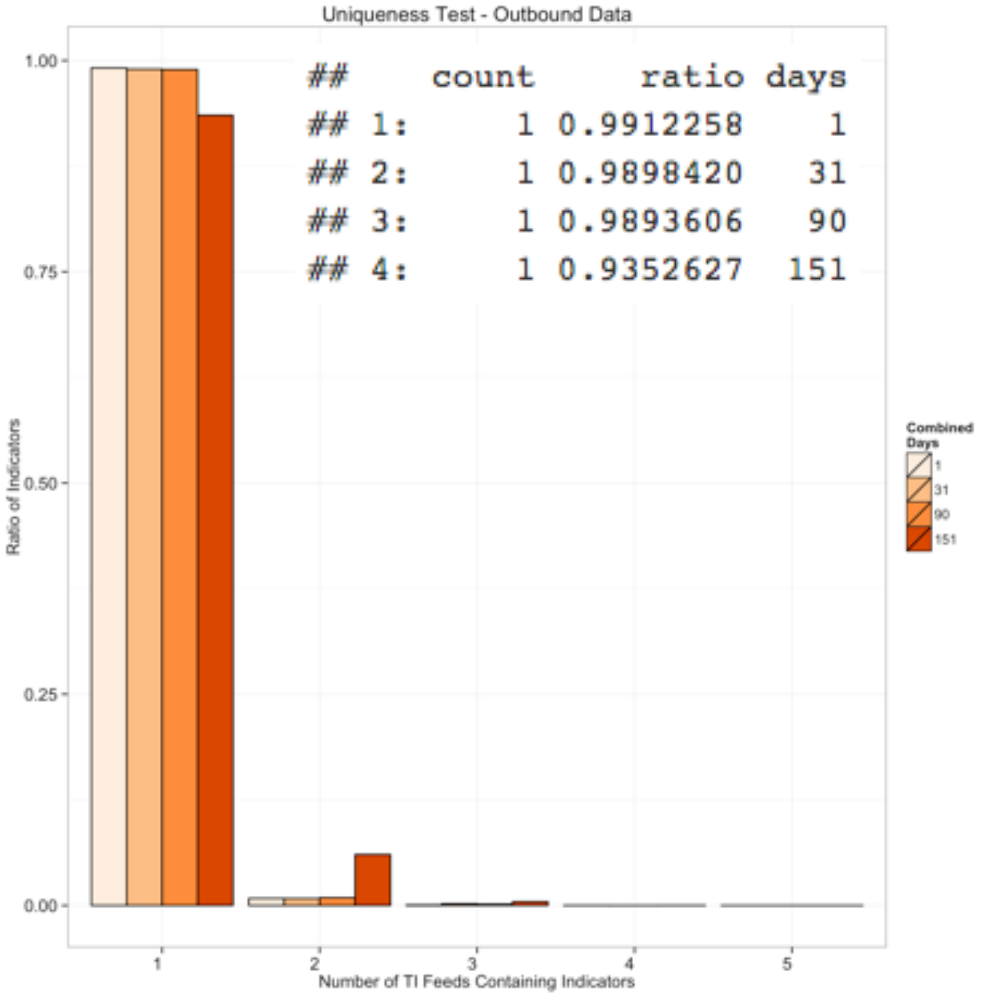
More data is fine, but make sure it is different


Overlap Test - Outbound Data - 20150501
Source (contains)
public_outbound.zeus
public_outbound.sslbl
public_outbound.phishtank
public_outbound.palevotracker
public_outbound.openphish
public_outbound.malwaregroup
public_outbound.malwaredomains
public_outbound.malwaredomainlist
public_outbound.malwared
public_outbound.malcode
public_outbound.kafeine
public_outbound.feodo
public_outbound.et_shadowserver_cnc
public_outbound.bambenek
public_outbound.alienvault
Source (is contained)
% Overlap
0%
25%
50%
75%
100%

Uniqueness Test

Can we tell if we are close to finding *all* the threats?


Uniqueness Test - Inbound Data
count ratio days
1: 1 0.9684775 1
2: 1 0.9678683 31
3: 1 0.9639037 90
4: 1 0.9631740 151
Ratio of Indicators
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11
Number of TI Feeds Containing Indicators


Uniqueness Test - Outbound Data
count ratio days
1: 1 0.9912258 1
2: 1 0.9898420 31
3: 1 0.9893606 90
4: 1 0.9352627 151
Ratio of Indicators
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5
Number of TI Feeds Containing Indicators
Combined Days
1
31
90
151

I hate quoting myself, but…

It is hard to draw a positive conclusion from these metrics, and it seems to suggest that if threat intelligence indicators were really able to help an enterprise defense strategy, one would need to have access to all of the feeds from all of the providers to be able to get the "best" possible coverage. This would be a Herculean task for any organization, and given the results of our analysis, the result would still be incomplete intelligence. There is a need for companies to be able to apply their threat intelligence to their environment in smarter ways so that even if we cannot see inside the whole lake, we can forecast which parts of it are more likely to have a lot of fish we still haven't caught.

MORE != BETTER

Threat Intelligence
Indicator Feeds

Threat Intelligence
Program

Constructive Feedback from the Internet:

"TI Sharing is TOTALLY going to solve this"

Right, folks? Right?

TI Sharing Solution Plan:

Or at least a rough straw man

1. The best Threat Intelligence is the one that you analyze from your own incidents (homegrown / organic intelligence)

2. There is strength in numbers – vertical herd immunity!

3. ????????

4. PROFIT!! (or at least SECURITY!!)

Issue 1 - BYOTI

Spotlight

Threat intelligence: only for the 1%?

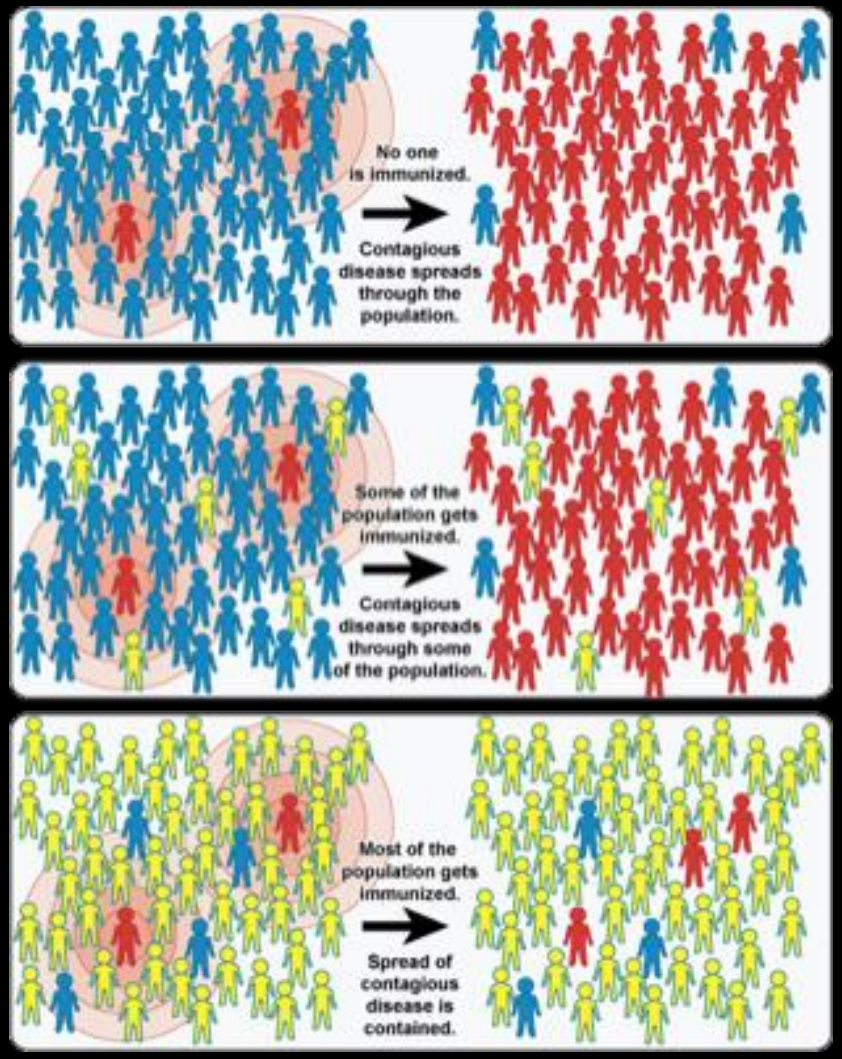
Analyst: Scott Crawford 1 Jul, 2015

Threat intelligence has become a booming area of information security, and with good reason. Attackers have the luxury of exploiting whichever weaknesses in a target best serve their intent. Defenders, on the other hand, must make the most of limited resources to defend all the most vulnerable aspects of critical information assets. Understanding the nature of current threats and adversary intent is essential to knowing how and where to place the most effective bets on defense.

If CONSUMING is for the 1%, what is the percentage of organizations able to PRODUCE?

Issue 2 - Herd Immunity

- We may be able to detect more ”virus strains” together but we are *terrible* at inoculation.
- The things we detect the most mutate too fast (Pyramid of Pain)
- Who didn’t get immunized, still gets sick (FOMO-TI)



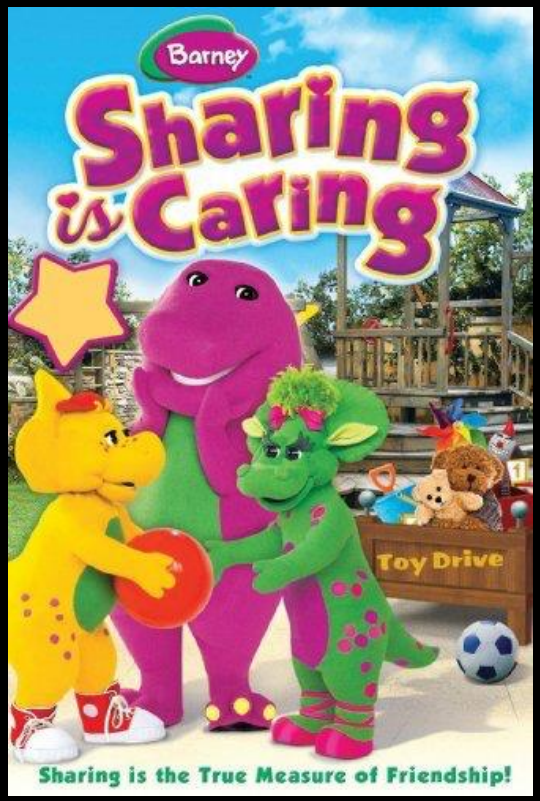
Source: www.vaccines.gov

Issue ? - What are we sharing

- AUTOMATION-DRIVEN (PLATFORMS)
  - Straight to the point IOC sharing
- ANALYST-DRIVEN (COMMUNITIES)
  - Strategic data, best practices, unstructured IOCs

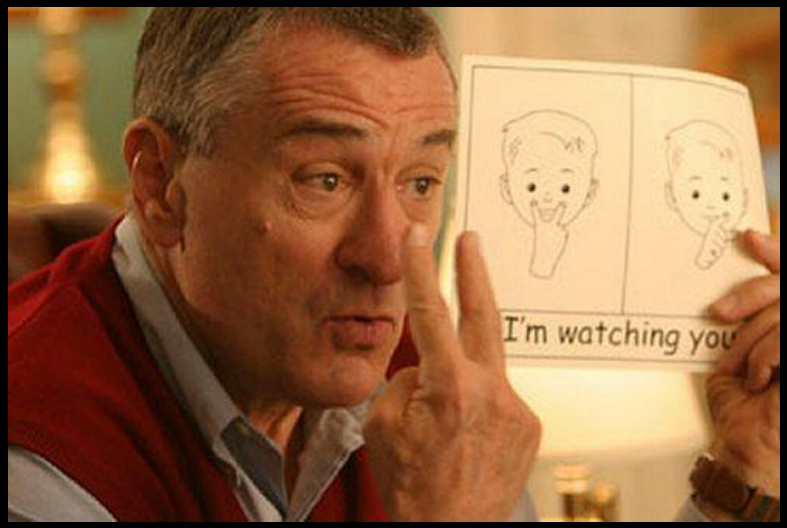
- ”Analyst-driven” has been around forever (in non-IC, at least since FS-ISAC was created)
- The same people who bash ”just IOC sharing”:
  - Bash STIX/TAXII for trying to encode complexity
  - Tells everyone it is IMPOSSIBLE to hire analysts

The Cognitive Dissonances of TI Sharing



Everybody should share!

VS



The CIRCLE OF TRUST

The Two Sides of the Trust Coin

VS

Do you trust the group enough to share?

Do you trust the group enough to consume?

Okay, I'll bite

Can we measure our current sharing platforms communities?

Threat Intelligence Sharing

We would like to thank the kind contribution of data from the fine folks at Facebook ThreatExchange and ThreatConnect

… and also the sharing communities that chose to remain anonymous. You know who you are, and we ❤ you too.

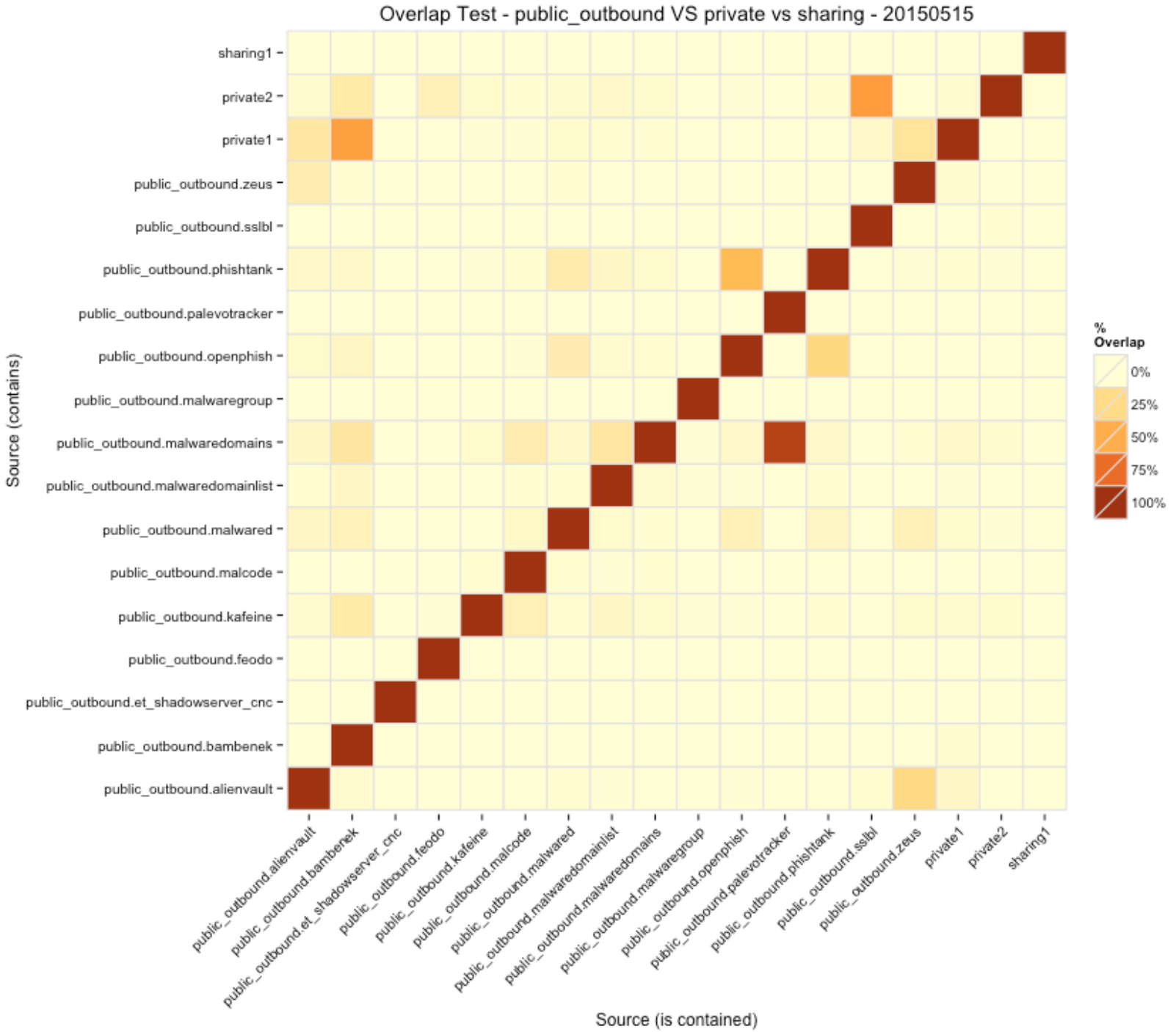
Sharing Communities ARE Social Networks

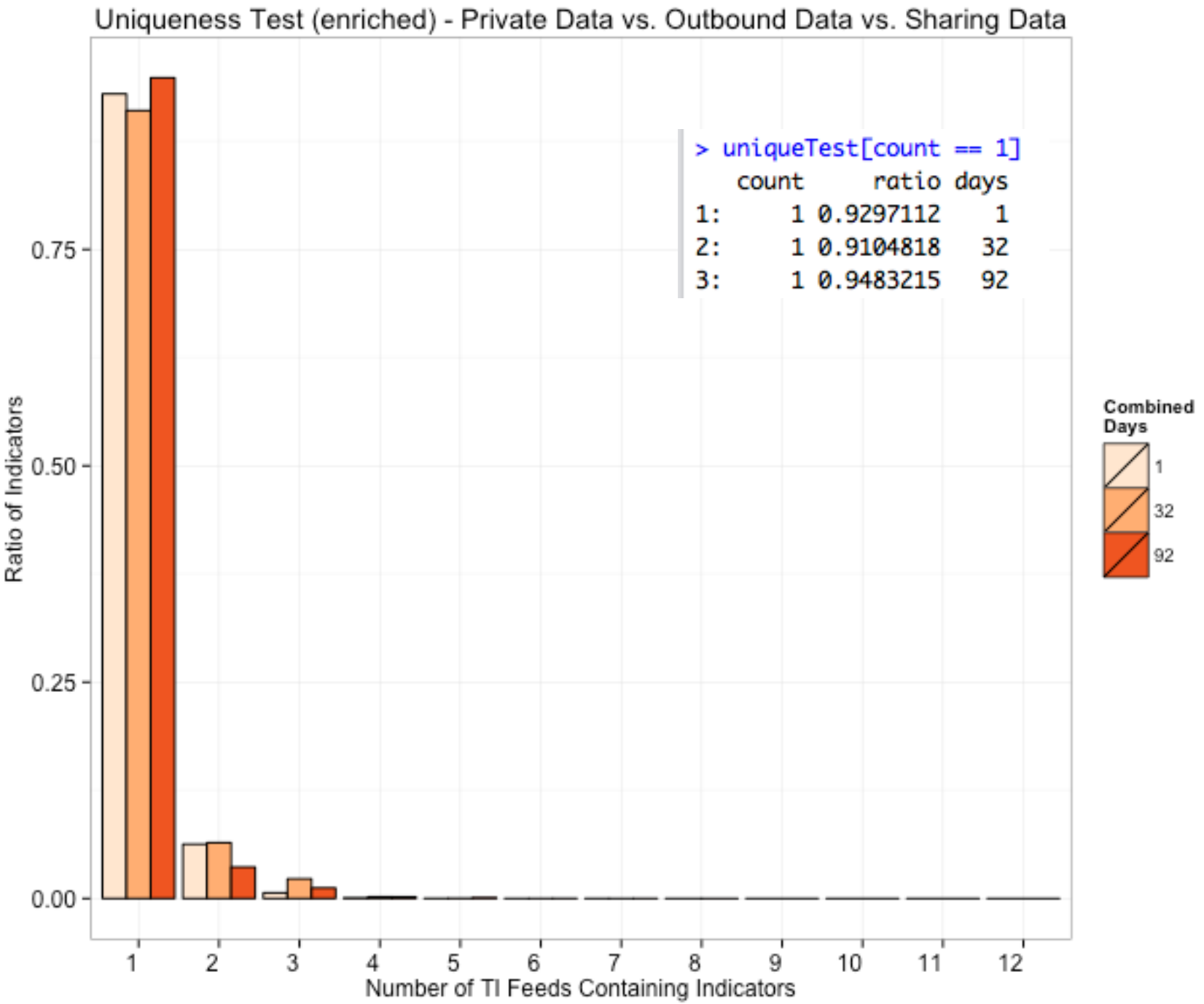
Social Network Selfie

Sharing Community Selfie

Let’s look at the indicators first

Using TIQ-TEST Overlap and Uniqueness tests


Overlap Test - public_outbound VS private vs sharing - 20150515
Source (contains)
sharing1
private2
private1
public_outbound.zeus
public_outbound.sslbl
public_outbound.phishtank
public_outbound.palevotracker
public_outbound.openphish
public_outbound.malwaregroup
public_outbound.malwaredomains
public_outbound.malwaredomainlist
public_outbound.malwared
public_outbound.malcode
public_outbound.kafeine
public_outbound.feodo
public_outbound.et_shadowserver_cnc
public_outbound.bambenek
public_outbound.alienvault
Source (is contained)
% Overlap
0%
25%
50%
75%
100%


Uniqueness Test (enriched) - Private Data vs. Outbound Data vs. Sharing Data
> uniqueTest[count == 1]
count ratio days
1: 1 0.9297112 1
2: 1 0.9104818 32
3: 1 0.9483215 92
Ratio of Indicators
0.00
0.25
0.50
0.75
1
2
3
4
5
6
7
8
9
10
11
12
Number of TI Feeds Containing Indicators
Combined Days
1
32
92

Looks like we would get similar quality on a ”good” Threat Intelligence Sharing Platform as we would on a ”paid feed"

Suggested Metrics for Sharing

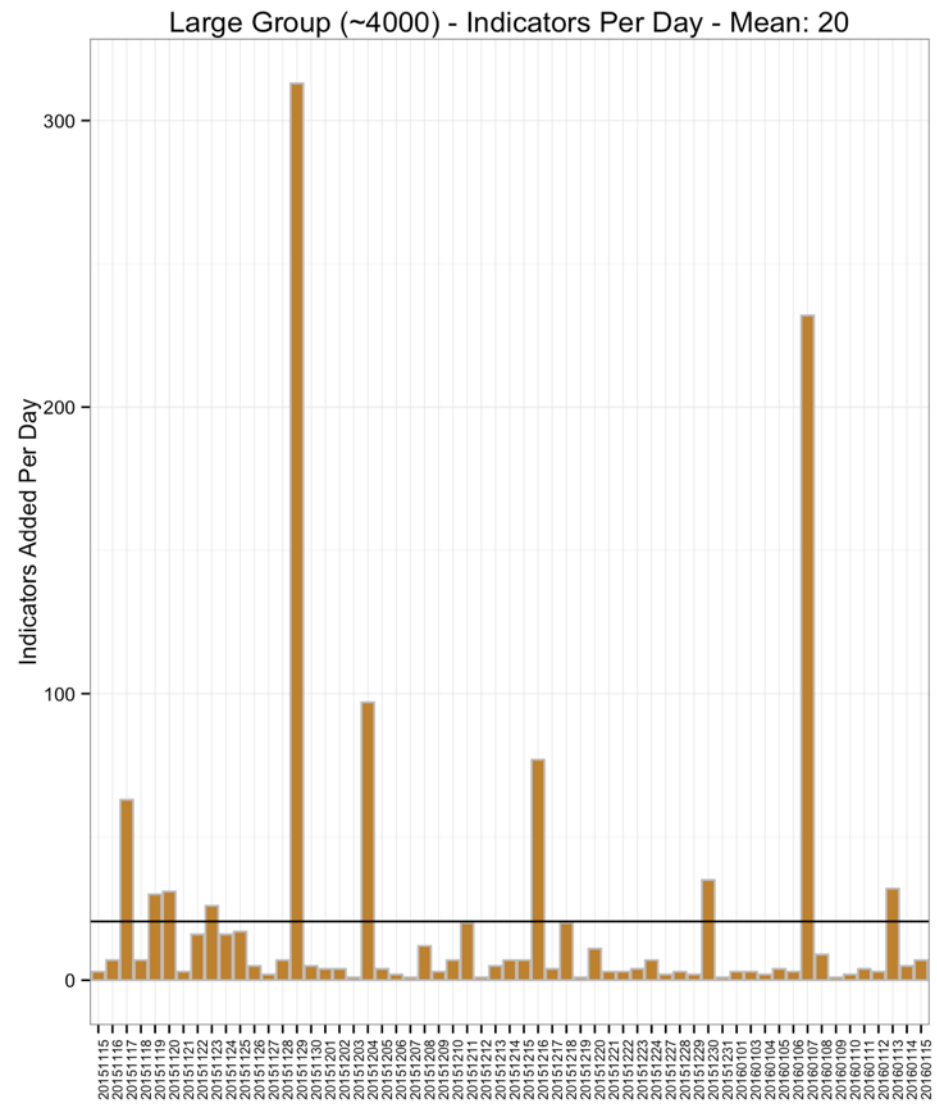
Looking for healthy dynamics

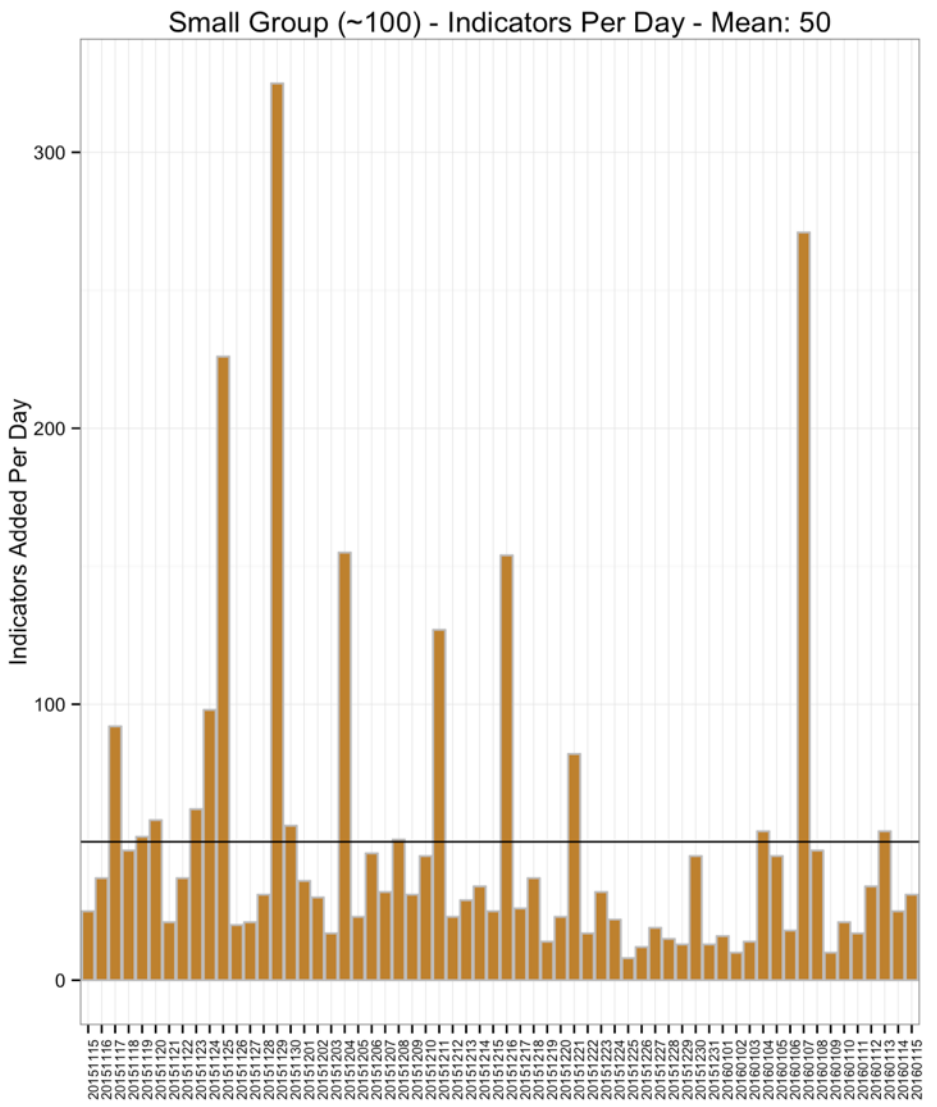
- ACTIVITY – How many indicators / posts are being shared day by day?
- DIVERSITY – What is the percentage of the population that is actively sharing?
- FEEDBACK – Are orgs collaborating on improving the knowledge in the sharing environment?
- TRUST – How much data is shared ”openly” in relation to ”privately”?

Activity Metric

Is there any actual sharing going on?

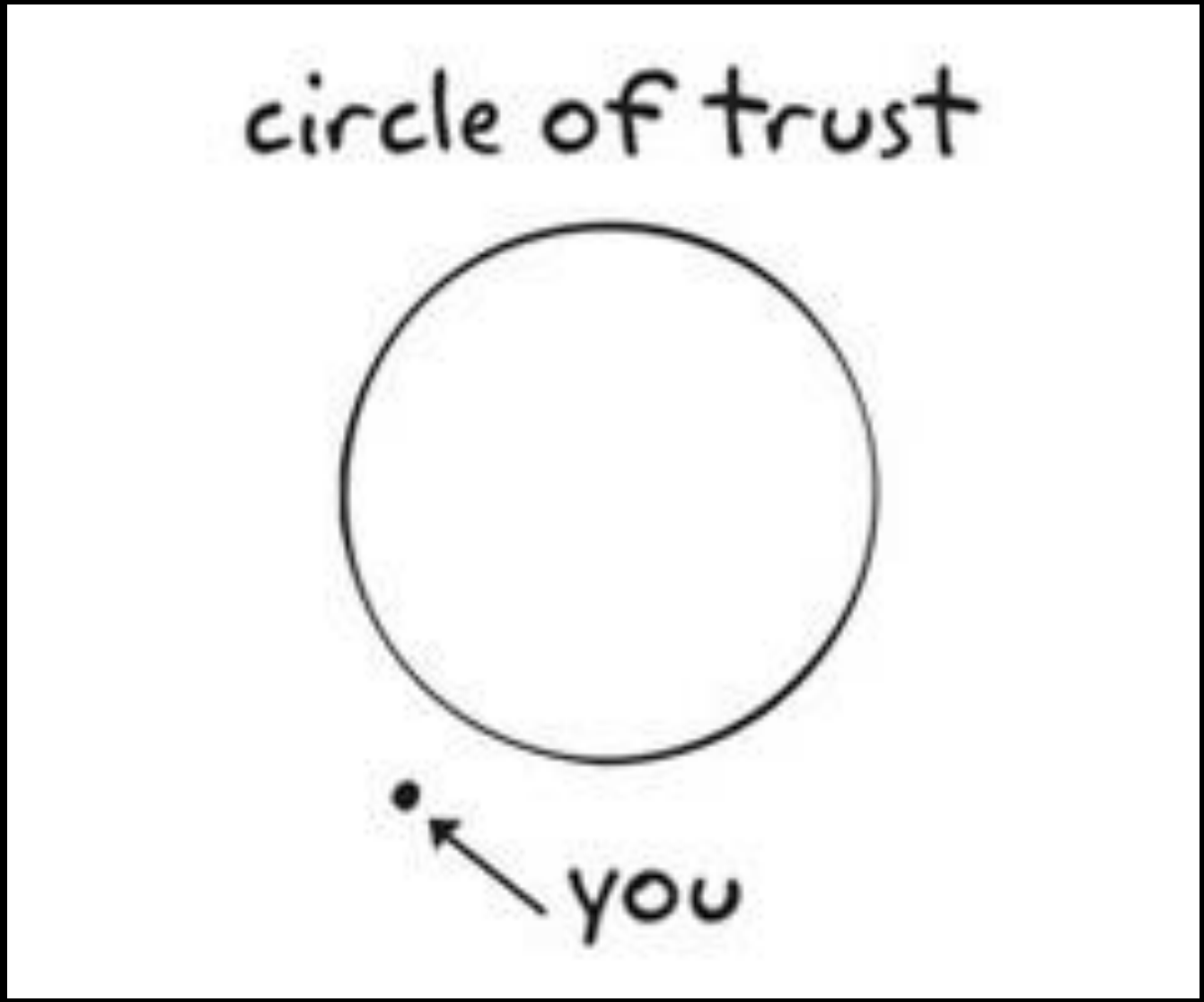
Large Group is roughly 40x bigger than Small Group





Less data / Delays

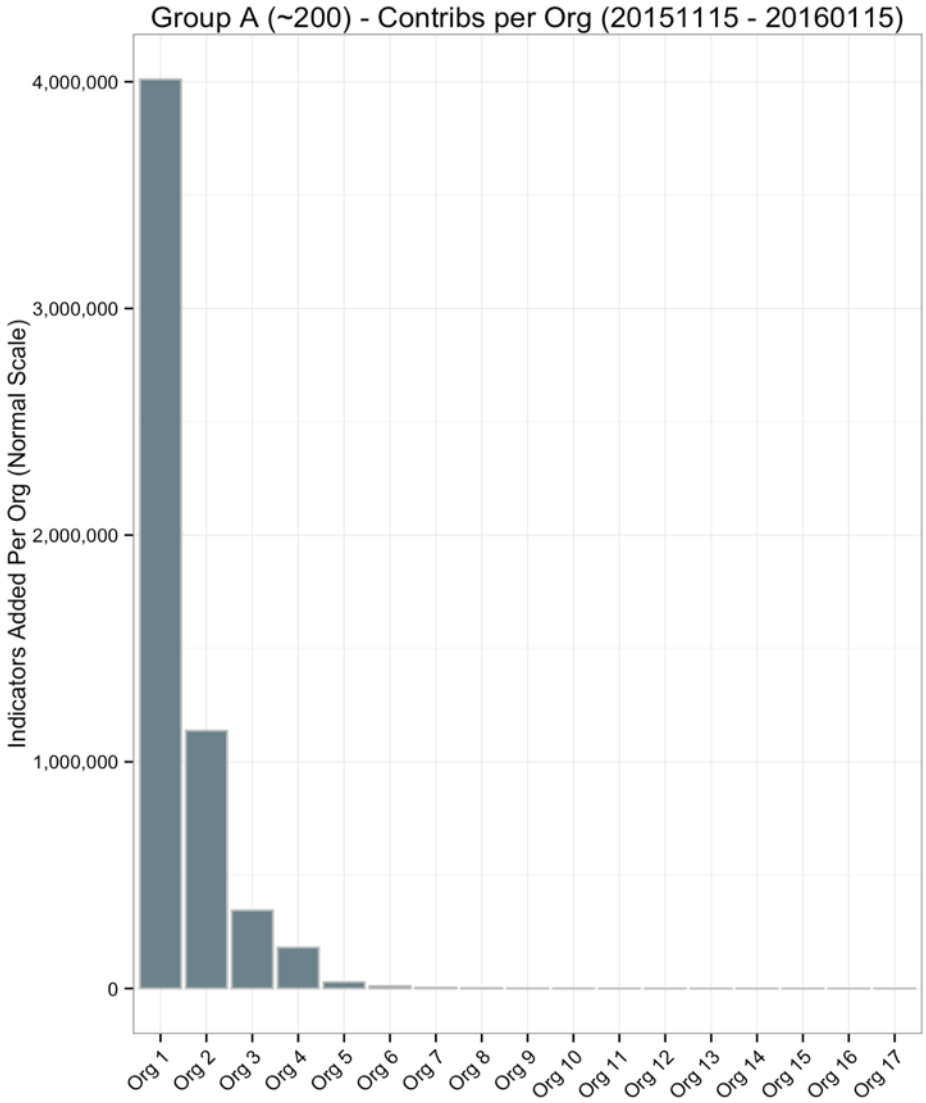
More data / Timely

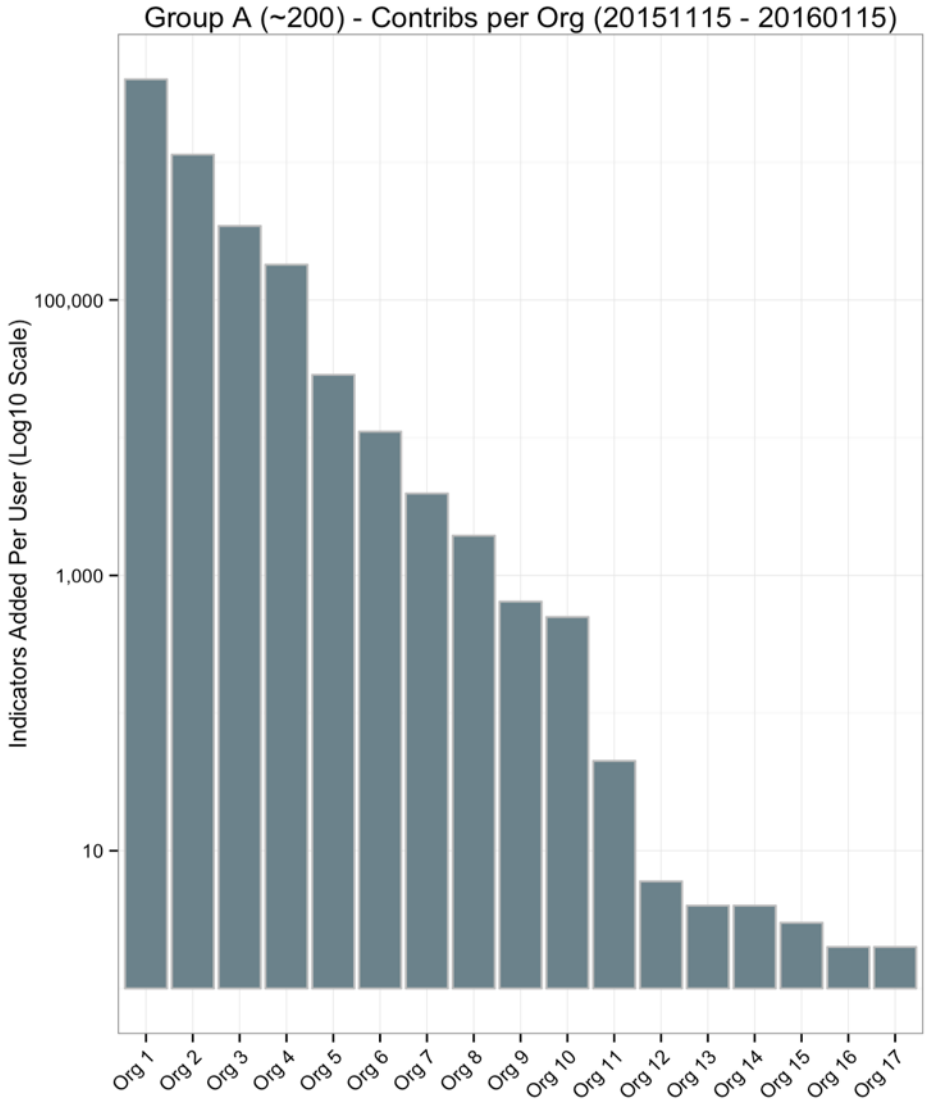


Organizations are less likely to share if they perceive they ”lost control” of who can consume.

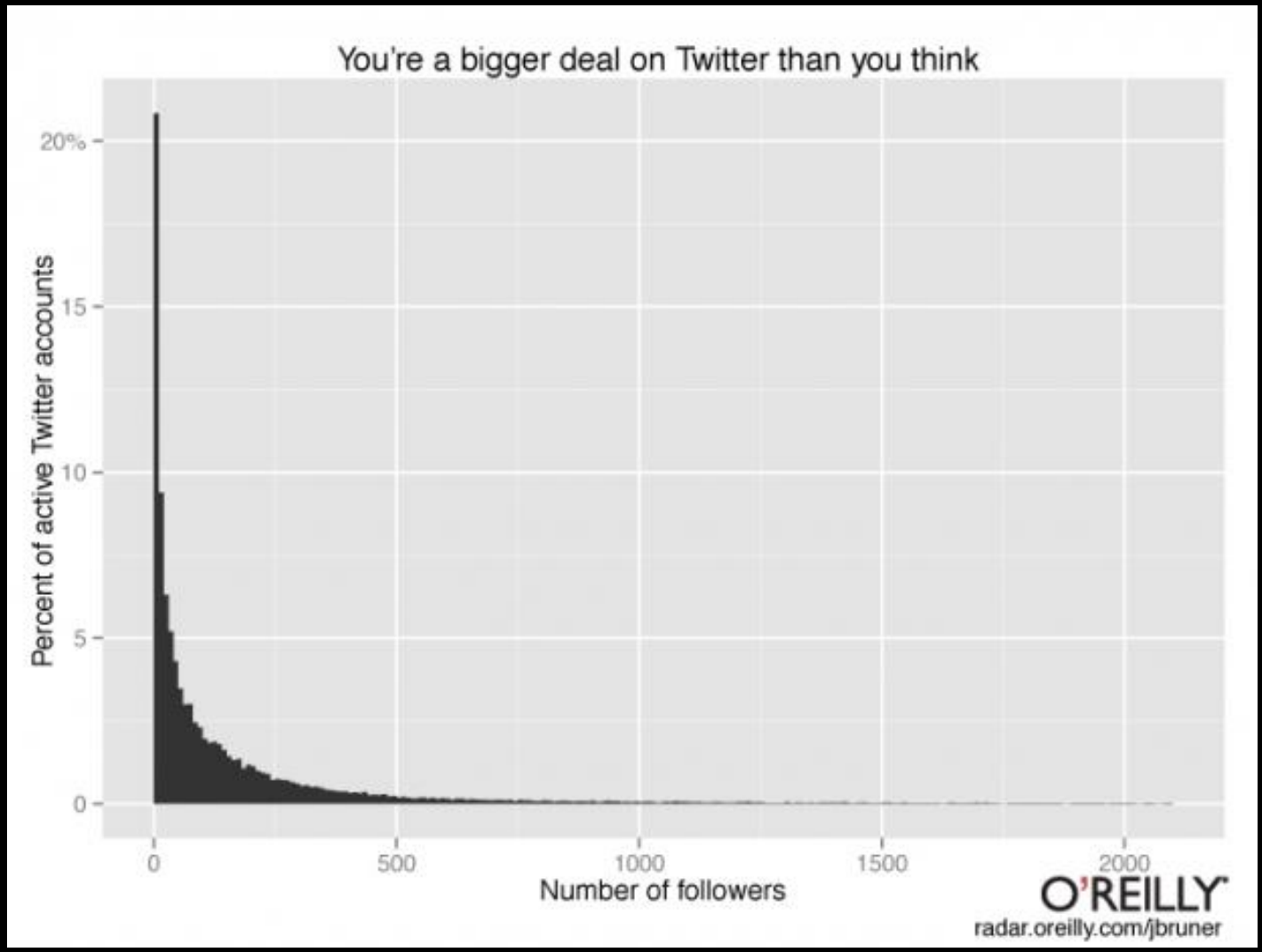
Diversity Metric

Check your sharing privilege





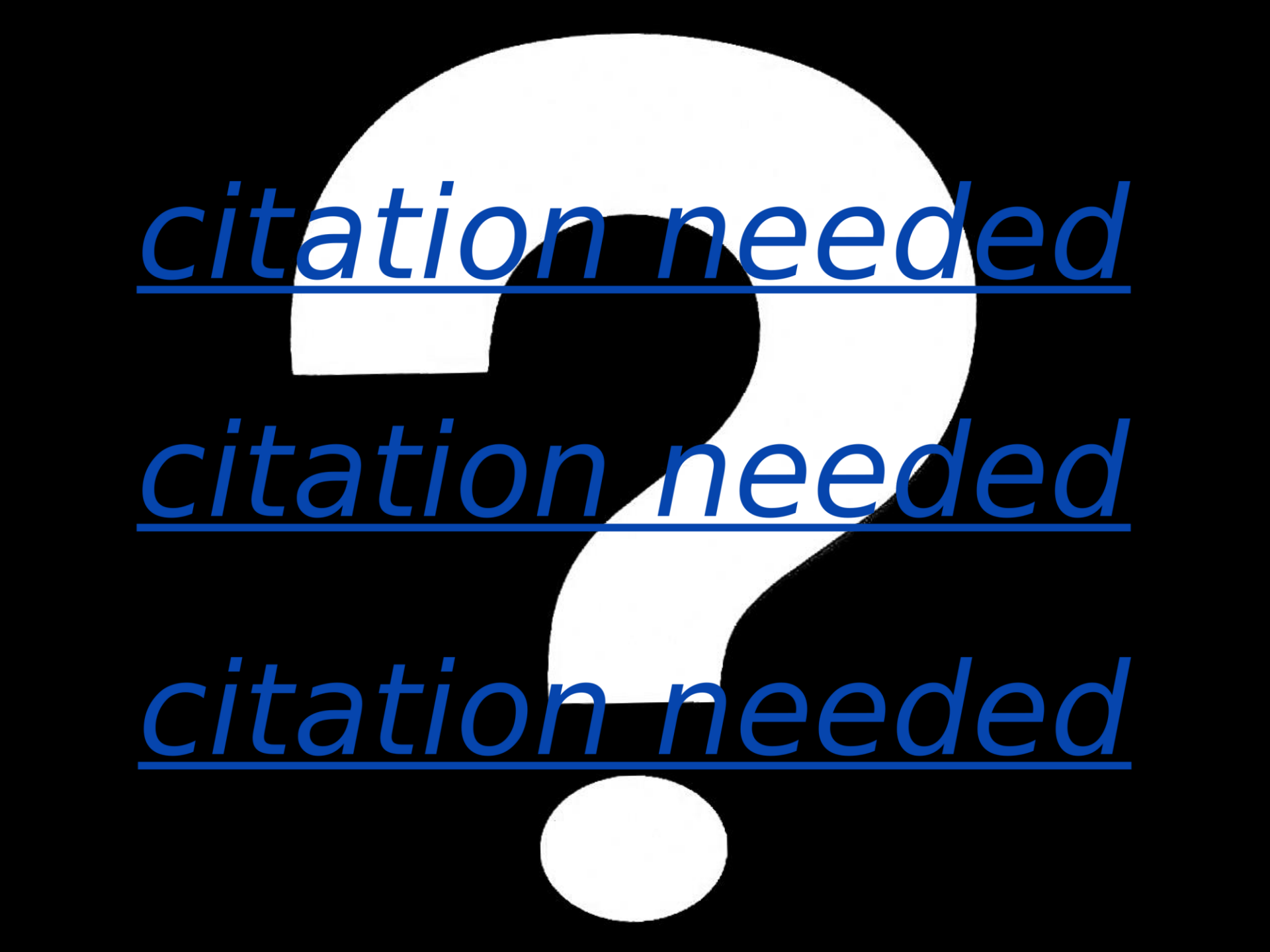
Roughly 10% of the organizations share data into the community

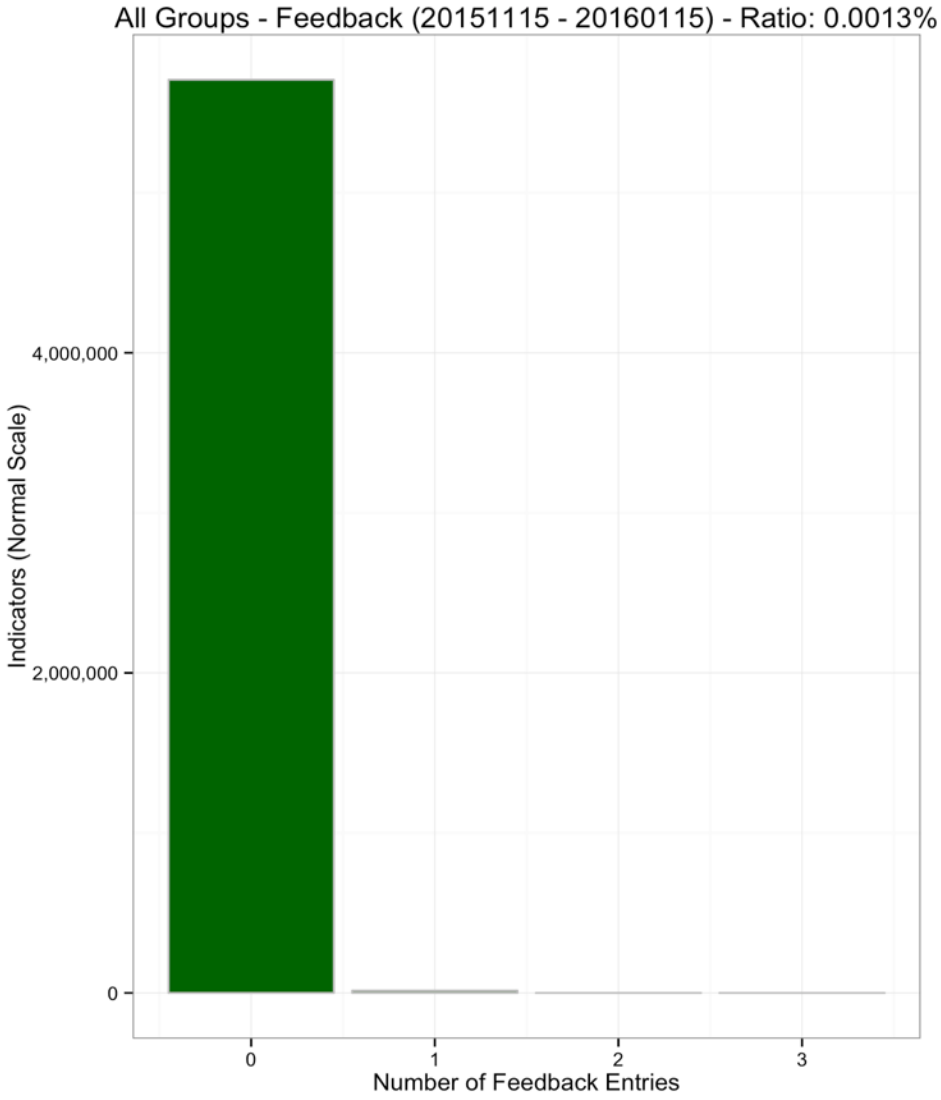


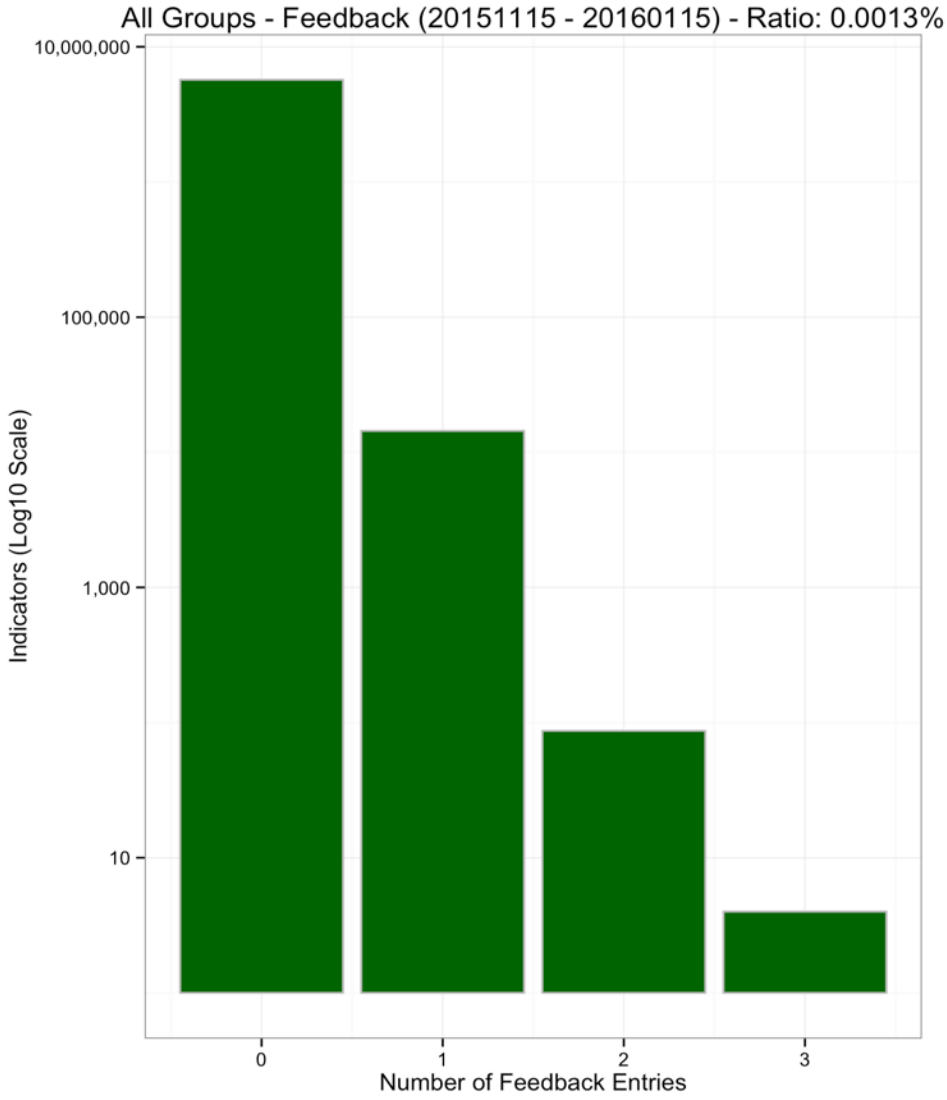
Some organizations are clearly in a better position operationally and legally to share. And that is expected due to our premises.

Feedback Metric

But is the data any good?


citation needed
citation needed
citation needed





😾 I’m sure we can do better than this 😾

Feedback Metric

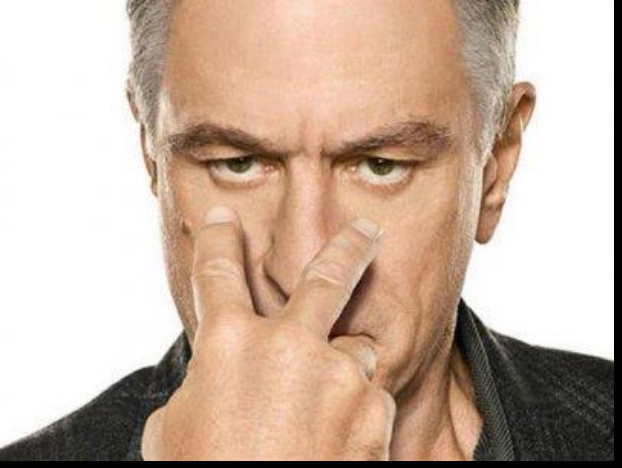
- Almost no support on automation-driven platforms
- Some allow you to leave ”comments” or ”new descriptors” for the IOCs – even by counting those very low % in relation to new shared data

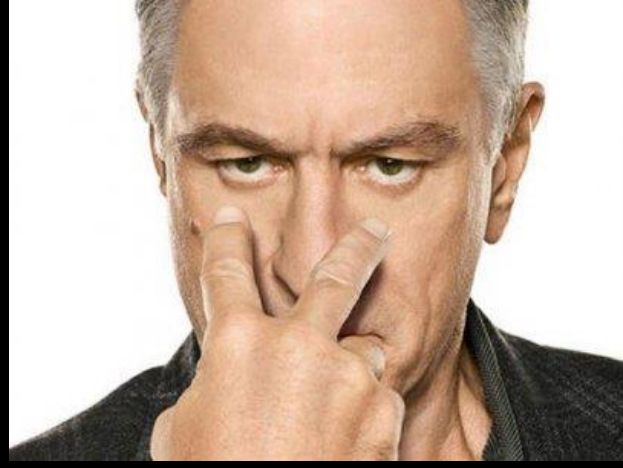
- Analyst-driven environments allow for collaboration on e-mails and forum posts to describe and refine strategies and best practices.

How can we make this collaboration work on automation-driven platforms?

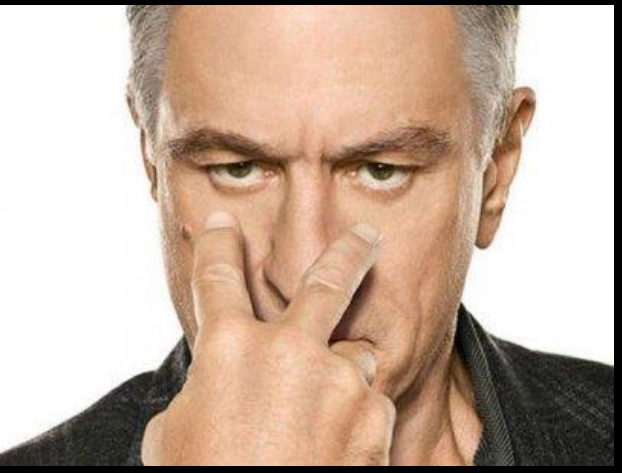
Trust Metric

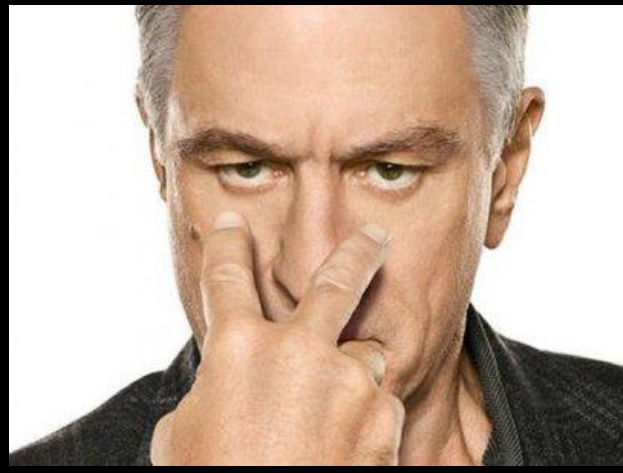
Are we helping all the community or just a few orgs at a time?





80% of data across all groups is shared privately (per the sample of data)

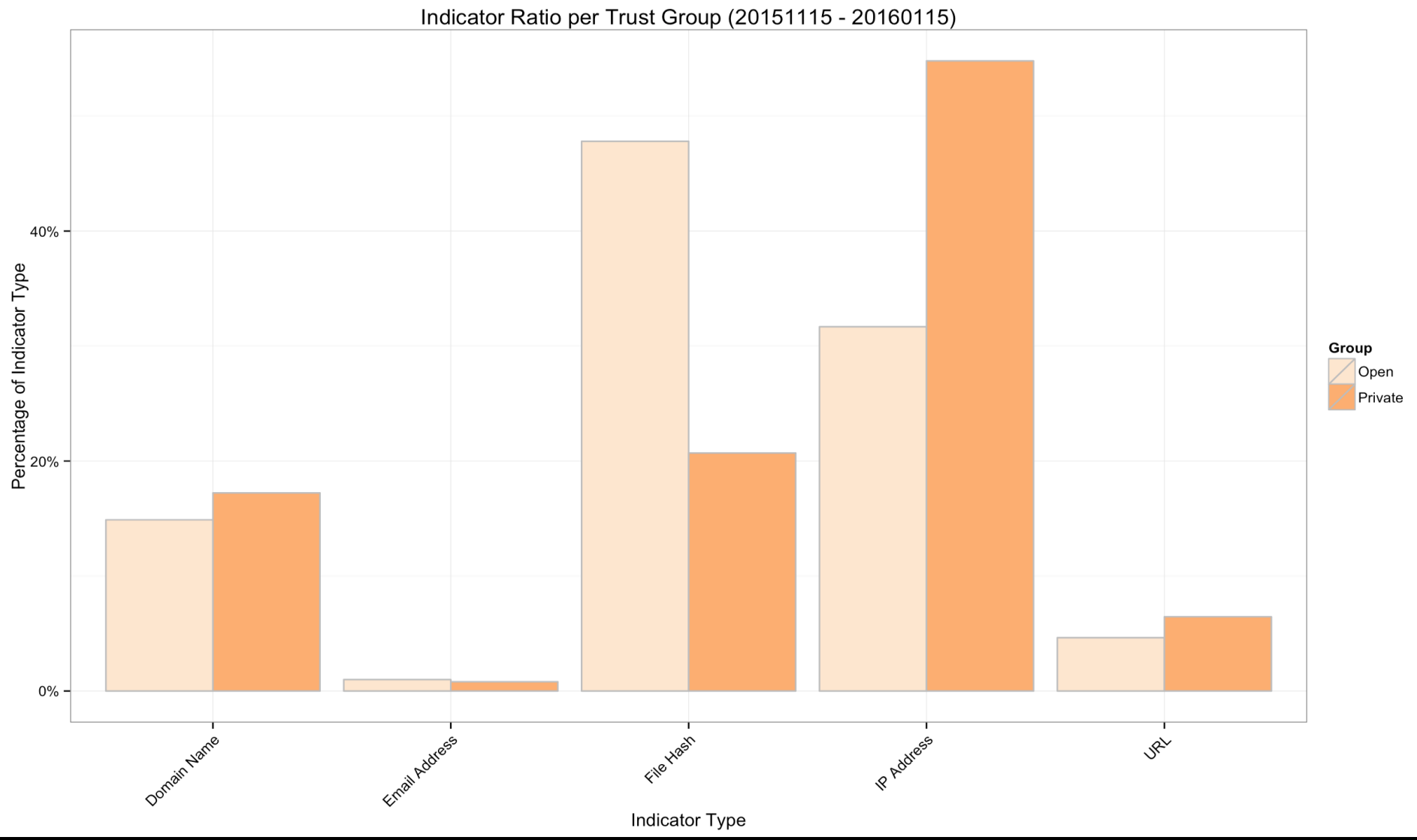




Hope you are having a good weekend! Here's a summary of what happened on your team last week:

Your team sent a total of **2,985 messages** last week (that's 132 more than the week before). Of those, **24% were in public channels**, **8% were in private channels** and **68% were direct messages**. Your team also uploaded **37 files** (that's 2 more than the week before).

76%. Again, sounds about right



Overall ”quality” of data goes up too!

Trust Metric

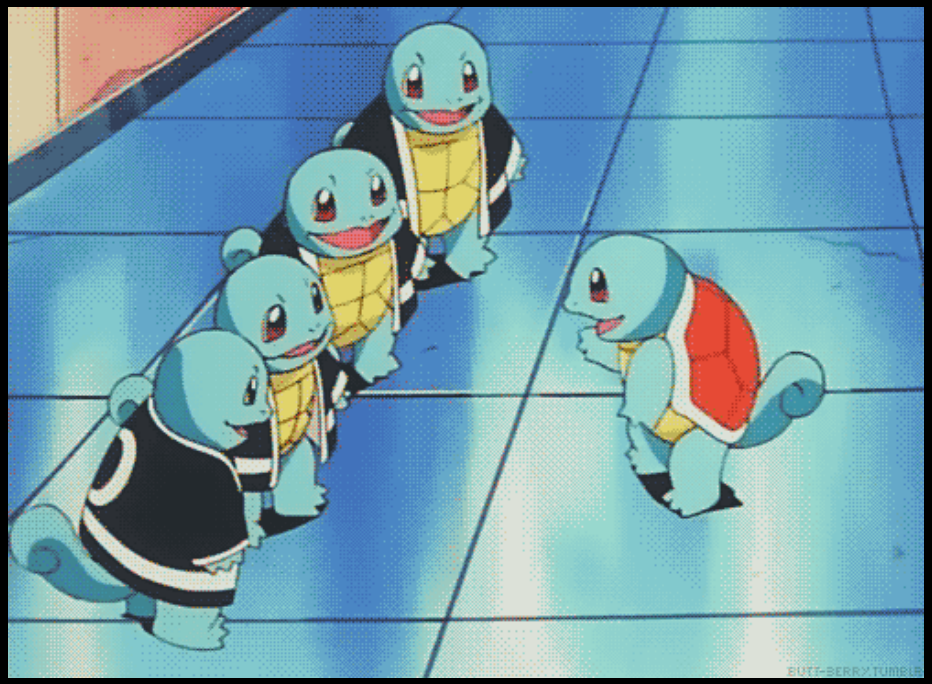
- The rough estimate seems to be that more than 60% of ”sharing” (IOCs, messages, etc) happens in ”private groups” inside the infrastructure of the sharing platform
- All communities have them:
  - Part of the DNA of the IC / cleared community
  - Offsets the trust equation, but defeats the ”herd immunity” argument
  - Usually MANDATORY on collaboration with LEA

But then the ”good” data is not helping ”the community”! Is there any way we can reconcile?

The Future of Sharing 

At the very least my humble opinion

#squadgoals

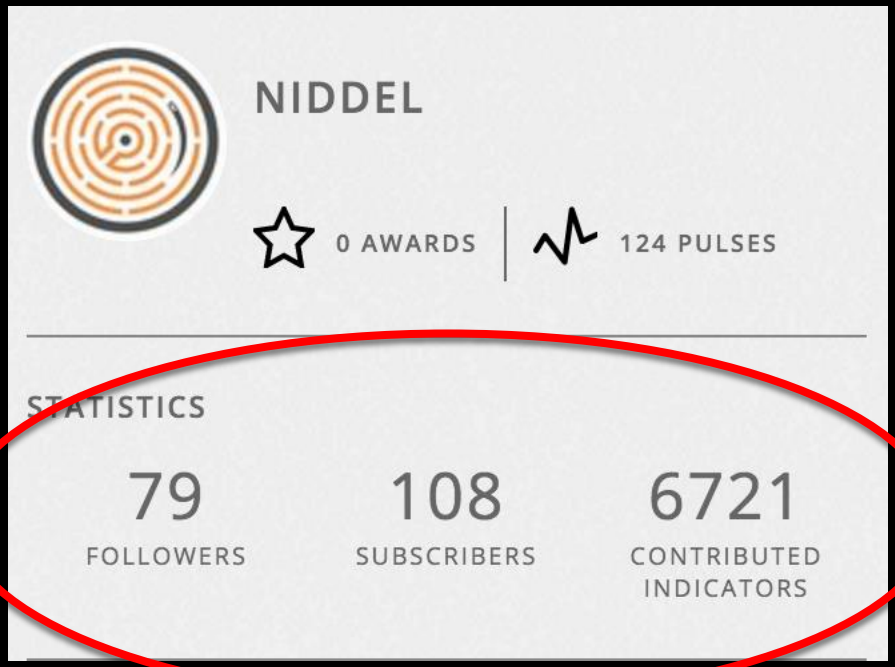


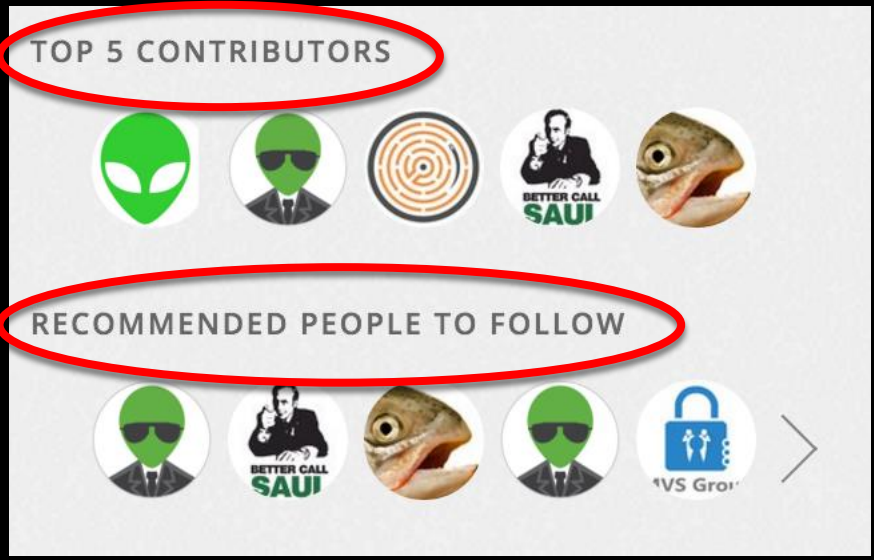
Increase the TRUST among peers

Reduce the TECHNICAL BARRIER for sharing useful information

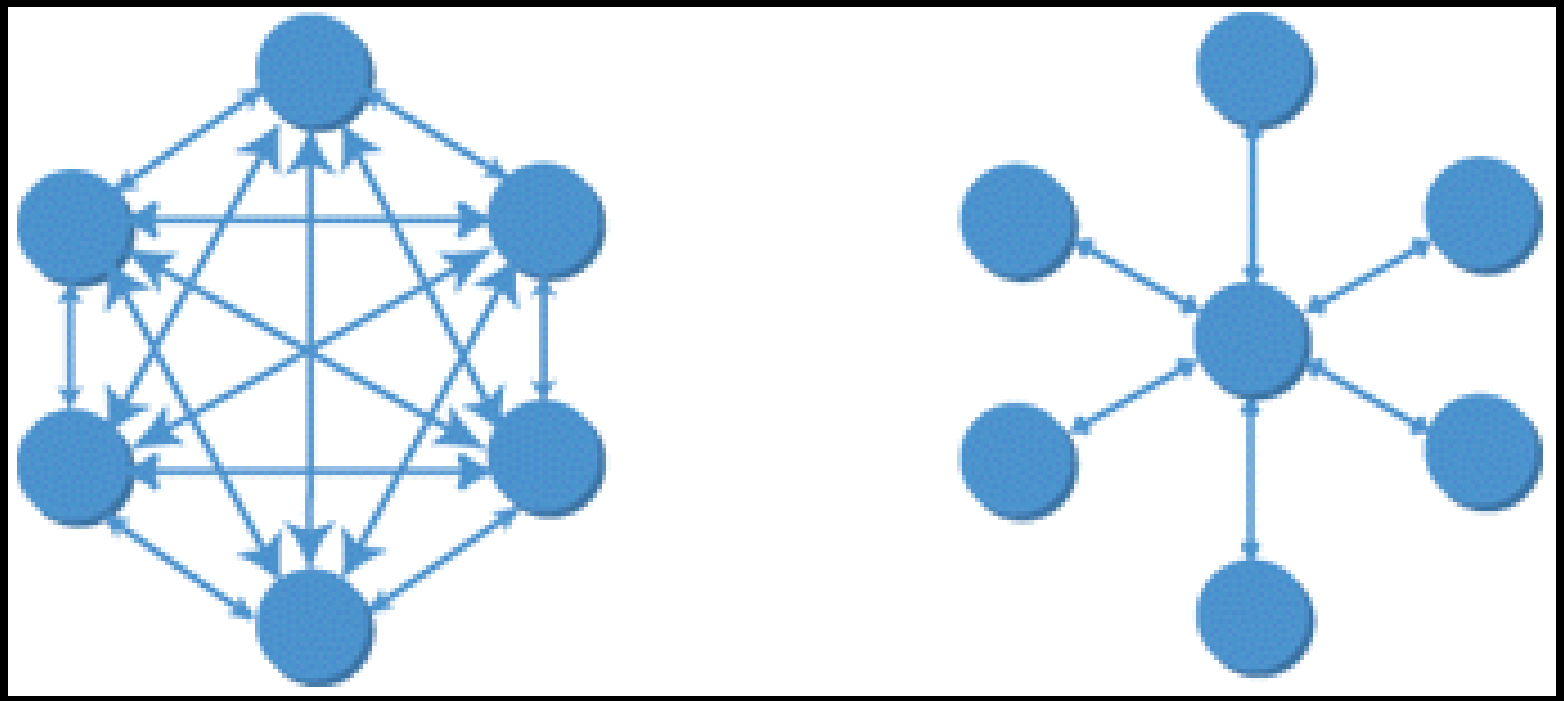
TRUST: Reputation and Anonymity

AlienVault OTX clearly got the memo



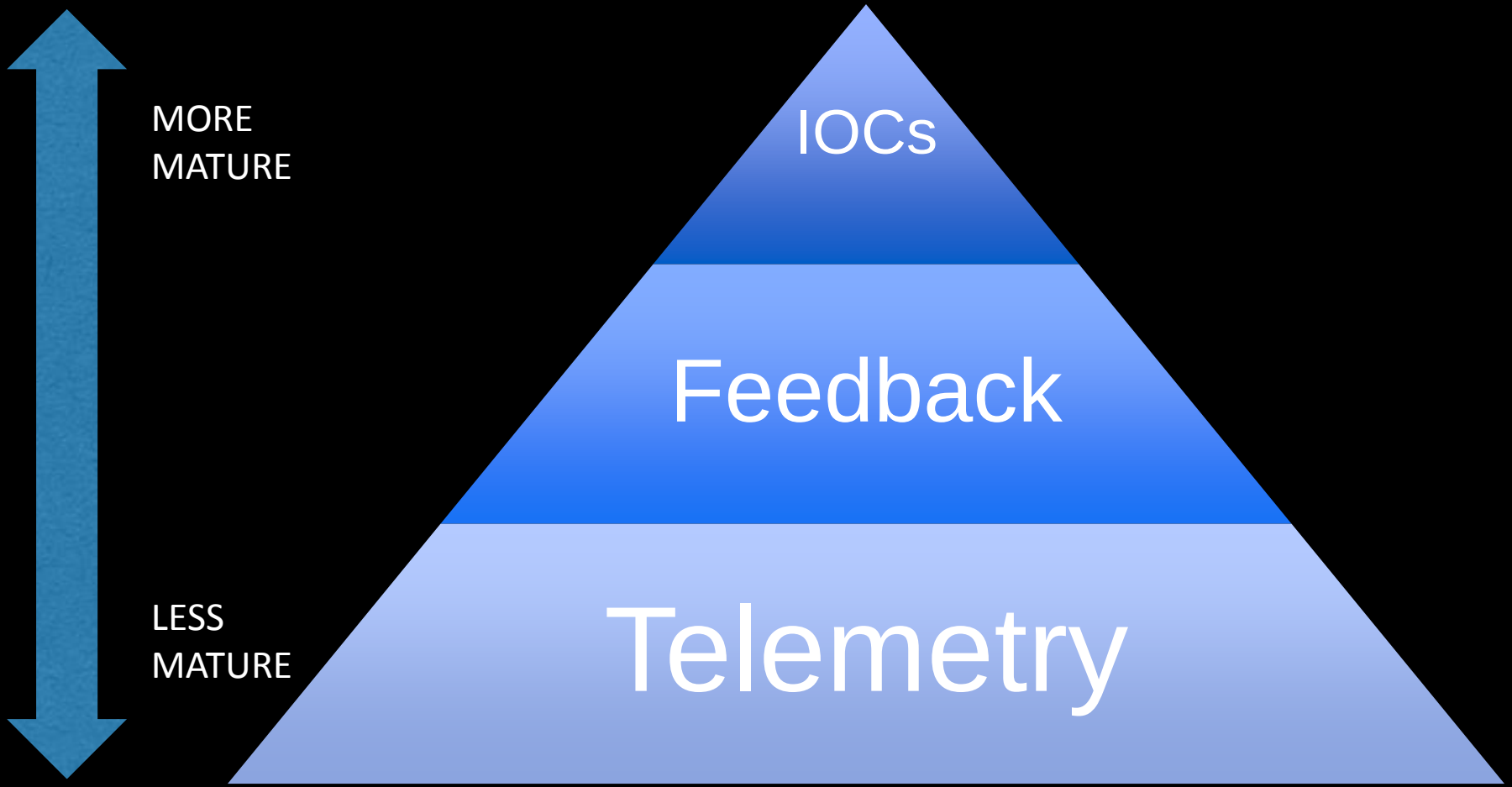


TRUST: Anonymity + Good Curation



Some sharing communities accept anonymous submissions that they then curate and disseminate to all organizations

TECHNICAL BARRIER:
"Pyramid of Sharing"



With ❤ and apologies to @DavidJBianco

Takeaways

- Intelligence Sharing is a very analyst-centric activity that we have been tasked with scaling out with automation. No wonder it seems so hard.

- Data can be as good as a paid feed, but you have to be in the right circles of trust

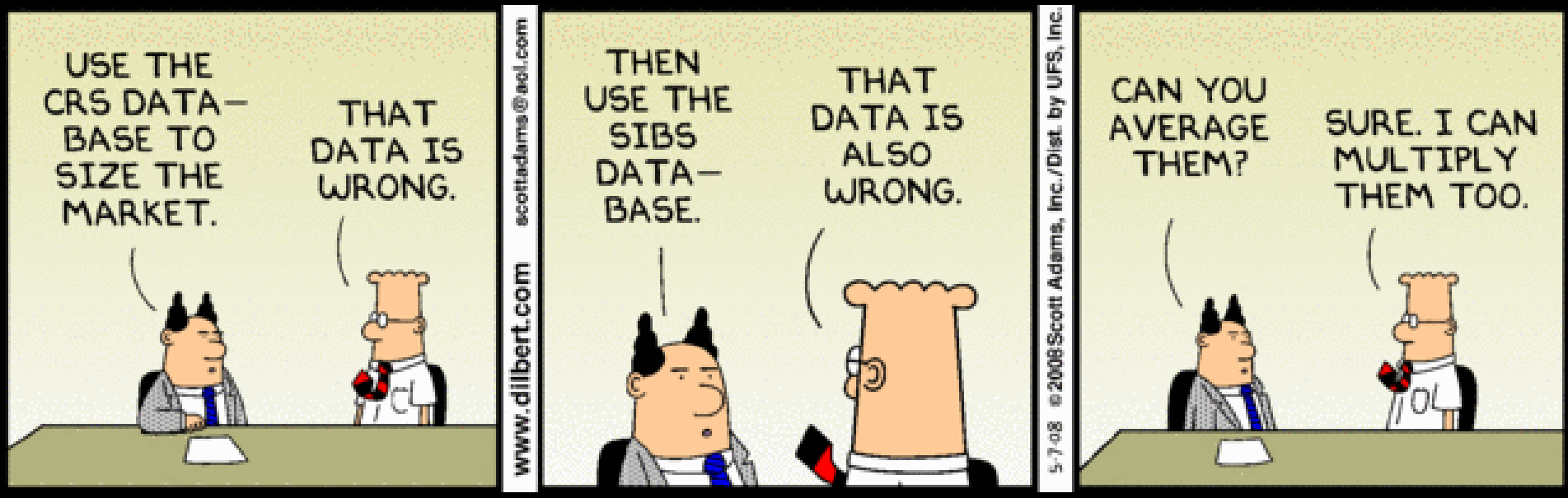
- Does not solve analyst shortage and making the indicators / strategies operational into your environment


Educational (61)
Retail (44-45)
Healthcare (62)
Public (92)
Finance (52)
Information (51)
Your gift of a few contributions
Can help a starving data scientist.

Thanks!

- Q&A?
- Feedback!

Alex Pinto
@alexcpsec
@MLSecProject / @NiddelCorp



”The measure of intelligence is the ability to change."

- Albert Einstein