

---

---

**“We have the data.”**

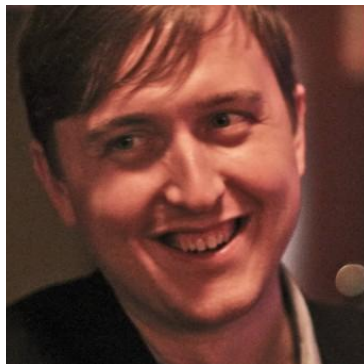
---

---

## Who are we?

### Kevin Clark

Senior Manager, Product Management  
VMware Tanzu Labs



### Jessica Peck

Senior Data Scientist, Search Relevance Team  
Zocdoc

### Victoria Nece

Principal Product Manager  
Adobe



### Dennis Clark

Data professional in finance

**This presentation has been formatted to fit your class**

---

---

**“We have the data.”**

This is a lie.

---

---

What they mean:

**“Calm down”**

---

What you should do:

**“Freak out”**

—

We | have | data

the

---

**We Have The Data**

---



We have the data.

Who is “we”?

—  
Your laptop?

Your team?

Your company?

A different company?

The government?

A different government?

—

We Have The Data

Show me a row...

...And the schema

# ...And the types



2.7k



Posted by u/RevolutionaryStrider 8 months ago



**Optimist: The glass is half full Pessimist: The glass is half empty**

**Excel: The glass is January 2nd**



105 Comments



Share



Save



Hide



Report

98% Upvoted

---

# How to handle a “we” problem:

Ask **Who is “we”?**

Ask it a *lot*.

If someone claims to be a “we,” ask them to **show you a row**.

Remember: you are going on a quest.

Don’t “start” the project until YOU have the data.

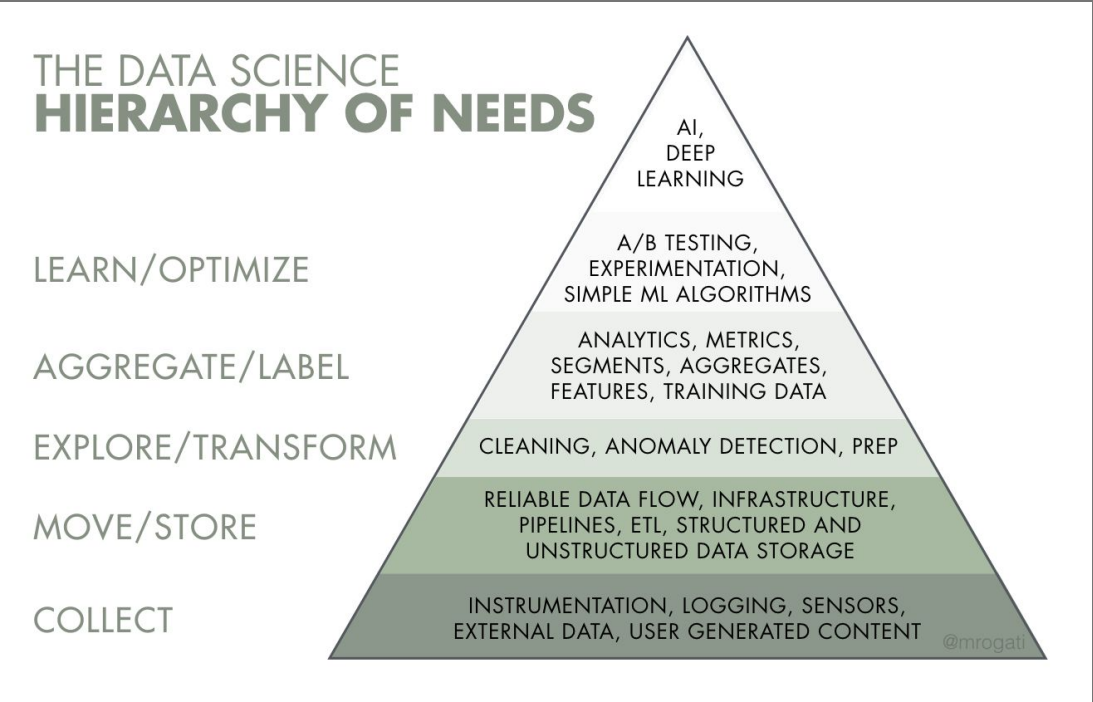
---

---

**We Have The Data**

---

# “Have” can mean a lot of things



Created by Monica Rogati



Can I have it to *use* or just  
to *look at*?

**PII, GDPR, HIPAA:  
Three acronyms in  
increasing order of  
length and also peril**

# When to mash the lawyer button

---

# How to handle a “have” problem:

Prototype as you go – build vertical slices to find problems

Talk to the experts – technical, legal, business – and share what they tell you

Turn “I can’t do X” into “I found a bug/blocker”

---

---

**We Have The Data**

---

# What data isn't

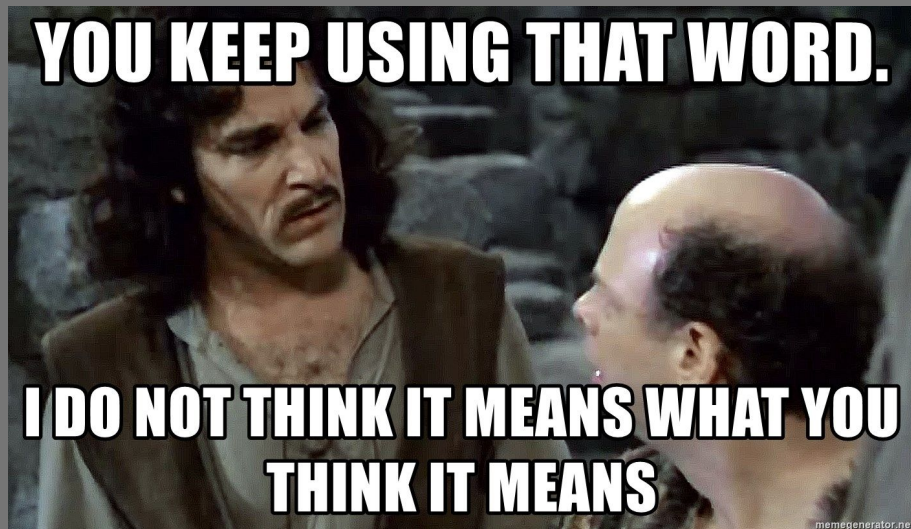


This is not data. This is a bus.

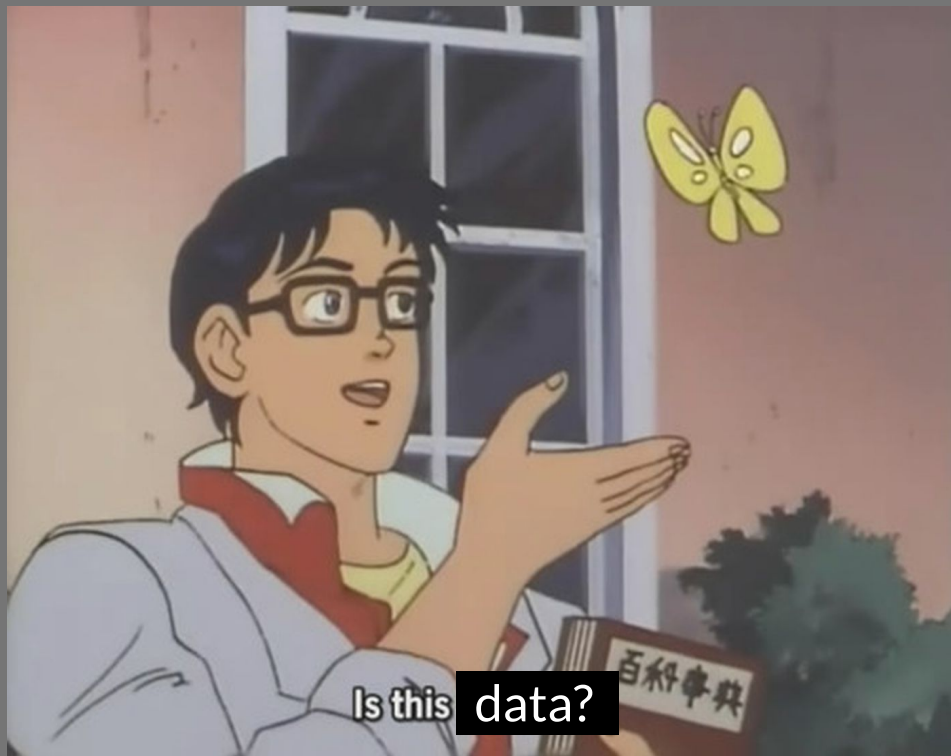
“Just Put Some  
Charts & Graphs In”



# Data-Adjacent Decision Making







## That's not data, that's...

- ...window dressing for your slide deck.
- ...qualitative user research.
- ...a statistic.
- ...test cases.
- ...that graph that's been going around.
- ...a pile of pdfs. Maybe with 1,000 hours of work it will grow up to be data.

# Should that be data?

<b>Form 990-T</b> Department of the Treasury Internal Revenue Service		<b>Exempt Organization Business Income Tax Return</b> <b>(and proxy tax under section 6033(e))</b> <i>2004</i>		OMB No 1545-0047  <b>2019</b> Open to Public Inspection for 501(c)(3) Organizations Only
For calendar year 2019 or other tax year beginning <u>07/01</u> , 2019, and ending <u>06/30</u> , 20 <u>20</u>		▶ Go to <a href="http://www.irs.gov/Form990T">www.irs.gov/Form990T</a> for instructions and the latest information. ▶ Do not enter SSN numbers on this form as it may be made public if your organization is a 501(c)(3).		
<b>A</b> <input type="checkbox"/> Check box if address changed  <b>B</b> Exempt under section <input checked="" type="checkbox"/> 501(c)(3) <i>03</i> <input type="checkbox"/> 408(e) <input checked="" type="checkbox"/> 220(e) <input type="checkbox"/> 408A <input type="checkbox"/> 530(a) <input type="checkbox"/> 529(a)	<b>Print or Type</b>	Name of organization ( <input type="checkbox"/> Check box if name changed and see instructions ) <b>TRUSTEES OF BOSTON UNIVERSITY</b>  Number, street, and room or suite no. If a P.O. box, see instructions <b>881 COMMONWEALTH AVENUE, 4TH FL</b>  City or town, state or province, country, and ZIP or foreign postal code <b>BOSTON, MA 02215-1303</b>	<b>D</b> Employer identification number (Employees' trust, see instructions)  <b>04-2103547</b>  <b>E</b> Unrelated business activity code (See instructions)  <b>520000</b>	<b>2939309201500</b>
<b>C</b> Book value of all assets at end of year <b>7,207,997,851</b>	<b>F</b> Group exemption number (See instructions.) ▶ <b>G</b> Check organization type ▶ <input checked="" type="checkbox"/> 501(c) corporation <input type="checkbox"/> 501(c) trust <input type="checkbox"/> 401(a) trust <input type="checkbox"/> Other trust			

2021

---

# How to recognize data from a long way away:

- It has rows
- It might have columns
- It might have values
- It is stored in a format
- Hope to god it's machine readable

It is probably really awkward in:

- Size
  - Access method
  - Type/Encoding
-

---

# How to handle a “data” problem

When people don't want data, don't give them data.

Check if you actually need that data at all!

**Show me a row** is magic!

If they can't do it, they probably don't have any data.

Plan for friction

---

---

**We Have The Data**

---

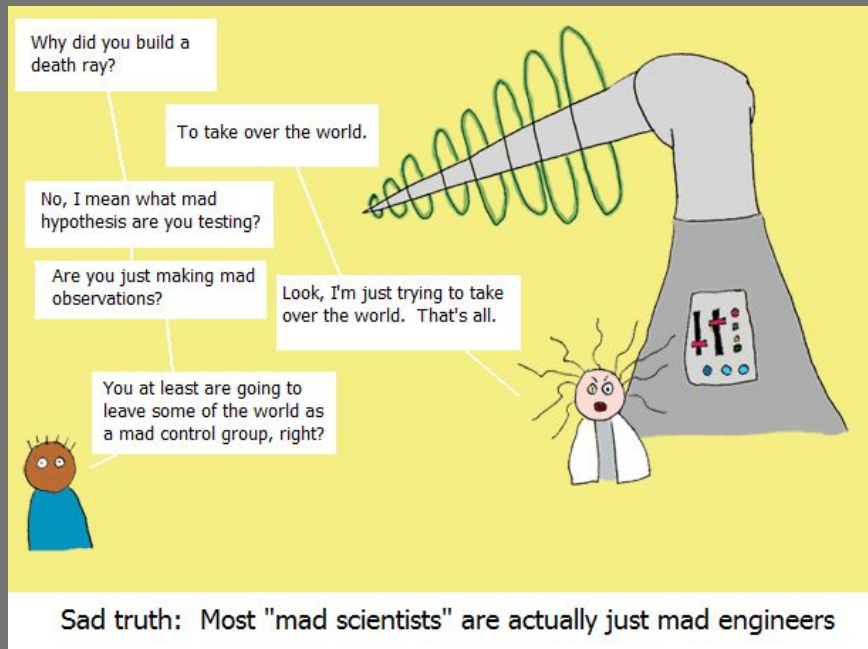
Maybe more than  
one row:  
The importance of  
scrolling down

Users lie to you  
(and only sometimes  
on purpose)

**Be kind to your  
anthropomorphized  
search algorithm**



# Observation Hypothesis Experiment



Cowbirds in Love #46 by Sanjay Kulkaček

# Jess's A/B Testing Rant™

1. Didn't log
2. Didn't plan
3. Didn't wait
4. Changed your mind

# Unpopular Opinion: Don't Always A/B Test

---

# How to handle a “the” problem

Use your spidey sense.  
Something seems off? Trend doesn't make sense? Dig in.

Let's try **SCIENCE**!  
Science takes time. Science is uncertain.  
Log your hypotheses before you start!

Write down the ideal data you wish you had before you take stock  
of what exists.

Plug everything in before you turn it on.

---

---

# We Have The Data

---

—  
**We  
Have  
The  
Data**

Who has it?

What can I do with it?

What does it show?

Is it actually data?

# Contact info and obligatory we're hiring slide



Obligatory cow image

## **Jessica**

jess.peck@zocdoc.com  
zocdoc.com/about/careers-list/

## **Kevin**

keclark@vmware.com  
careers.vmware.com

## **Victoria**

nece@adobe.com  
adobe.com/careers.html

## **Dennis**

dbfclark@gmail.com

---



---

---

# Thank you!

This is *not* a lie.

---