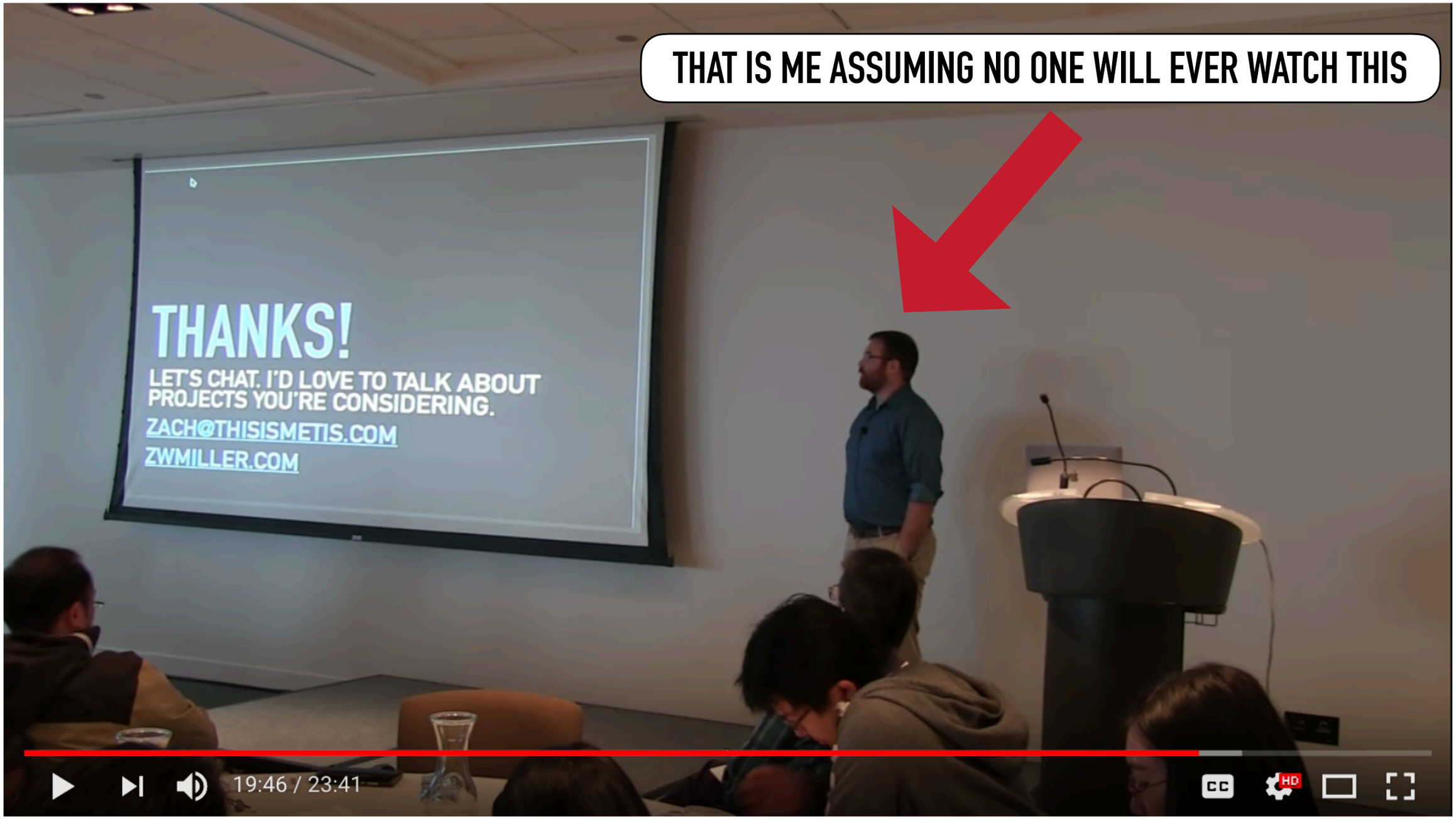
DISPELLING DATA SCIENCE MYTHS FOR BEGINNERS

Z. W. MILLER - 05.19.2018

MY TALK FROM THIS CONFERENCE LAST YEAR GOT A FEW VIEWS ON YOUTUBE. I DID SOMETHING DUMB AT THE END OF IT.



Roadmap: How to Learn Machine Learning in 6 Months

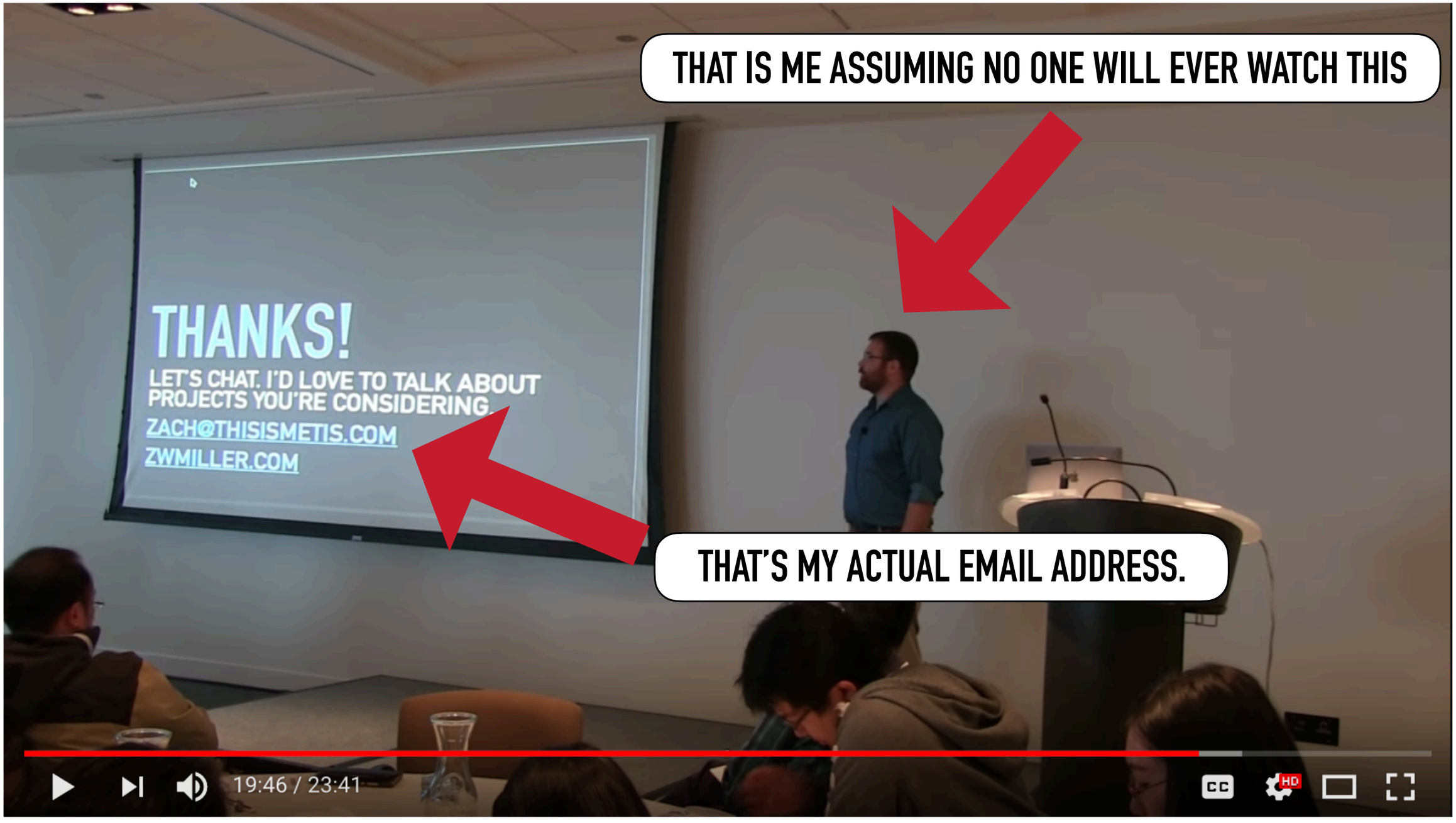
164,653 views

3.8K 62 SHARE

IDEAS
Published on May 27, 2017

SUBSCRIBED 2.6K

This talk is presented by Zach Miller, Senior Data Scientist at Metis



Roadmap: How to Learn Machine Learning in 6 Months

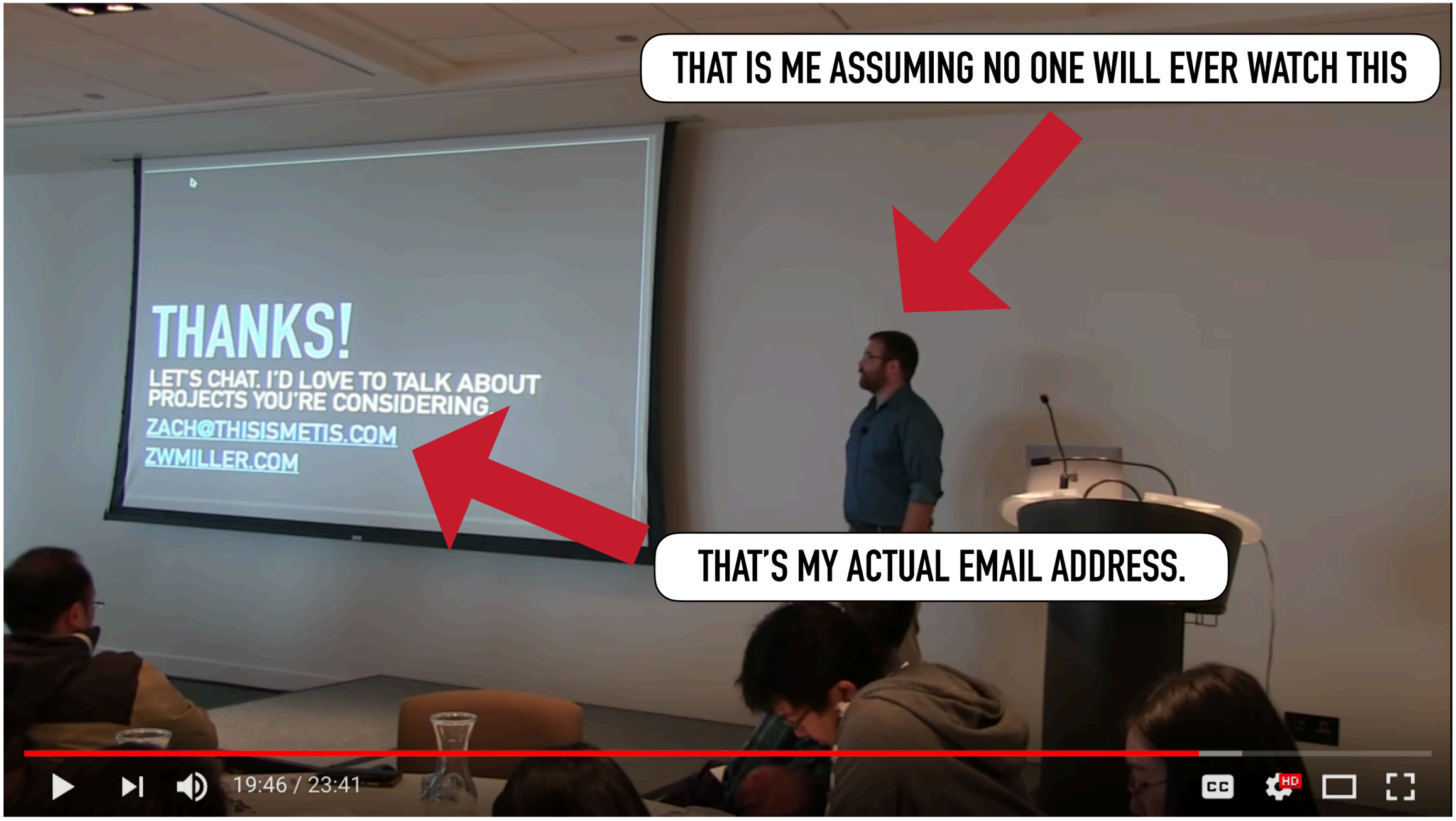
164,653 views

3.8K 62 SHARE

IDEAS
Published on May 27, 2017

SUBSCRIBED 2.6K

This talk is presented by Zach Miller, Senior Data Scientist at Metis



Roadmap: How to Learn Machine Learning in 6 Months

164,653 views

3.8K 62 SHARE

IDEAS

Published on May 27, 2017

THAT'S A NON-ZERO NUMBER

SUBSCRIBED 2.6K

This talk is presented by Zach Miller, Senior Data Scientist at Metis

FOR THE PAST YEAR, PEOPLE HAVE BEEN BARING THEIR SOULS TO ME VIA EMAILS ABOUT DATA SCIENCE

FOR THE PAST YEAR, PEOPLE HAVE BEEN BARING THEIR SOULS TO ME VIA EMAILS ABOUT DATA SCIENCE

"... ever since I finished my economics degree, I've been grinding at a horrible company. I hate getting up and going to work. Data science seems interesting and more fun than what I'm doing. Should I just keep grinding and hating my life, or do I have what it takes to become a data scientist? Oh, BTW my name is XXX."

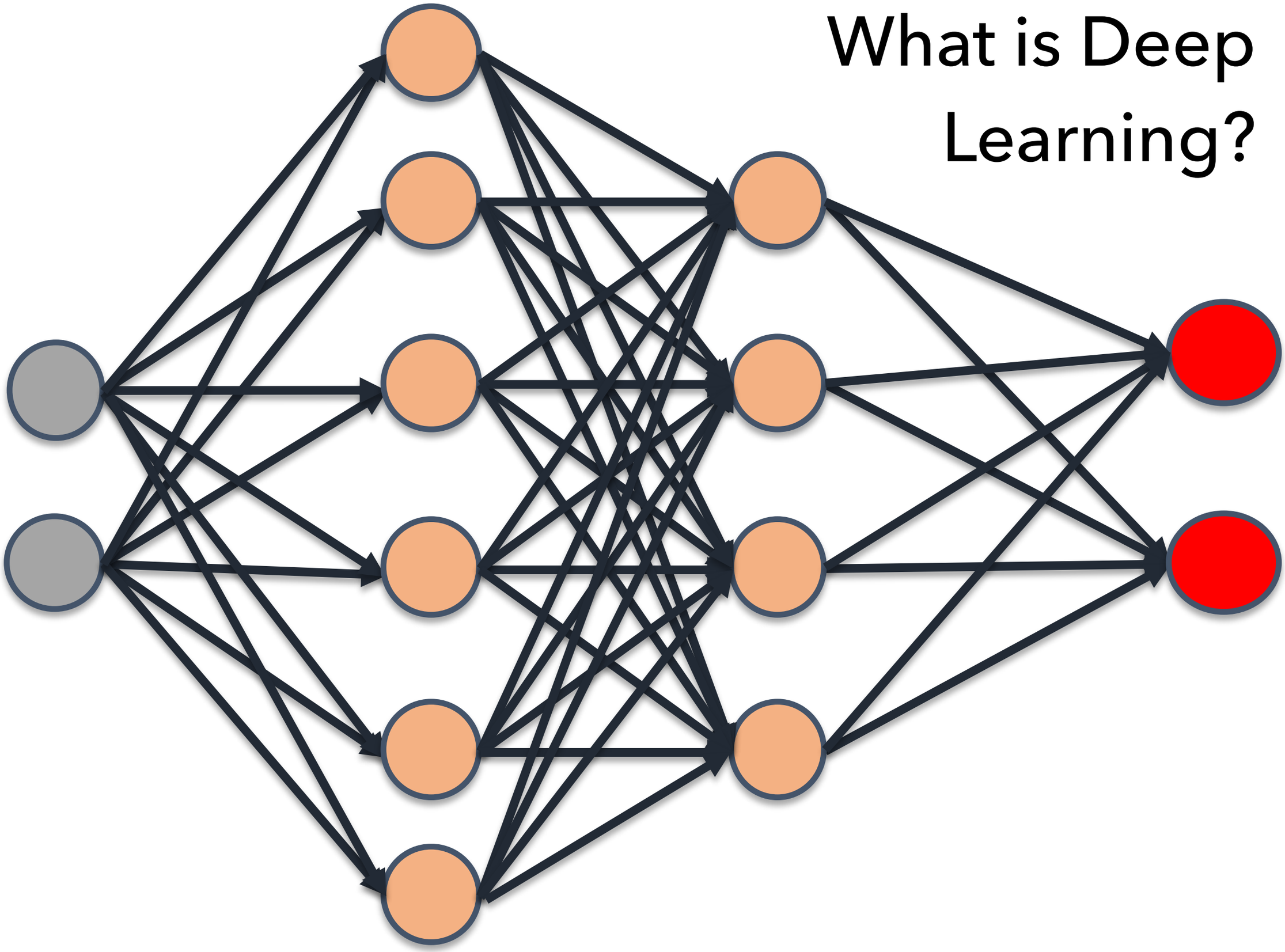
THREE REPEATED THEMES

- Should I learn Deep Learning or Machine Learning?
- Do I really need to know how to program/code?
- How science-y is data science?

1

YOU CAN PROBABLY SKIP DEEP LEARNING TO BEGIN WITH

What is Deep Learning?



WHY DO WE LIKE DEEP LEARNING?

https://github.com/Everhusk/Self-Driving-Car-Engineering/

WHY DO WE LIKE DEEP LEARNING?

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram

$$\begin{array}{ccc} \mathcal{S} & \longrightarrow & \\ \downarrow & & \\ \xi & \longrightarrow & \mathcal{O}_{X'} \\ & & \uparrow \\ \text{gor}_s & & \searrow \\ & = \alpha' & \longrightarrow \\ & \updownarrow & \\ & = \alpha' & \longrightarrow \alpha \end{array} \qquad \begin{array}{c} X \\ \downarrow \\ \end{array}$$

$$\text{Spec}(K_\psi) \qquad \text{Mor}_{Sets} \quad \text{d}(\mathcal{O}_{\mathcal{X}_{X/k}}, \mathcal{G})$$

is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \text{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

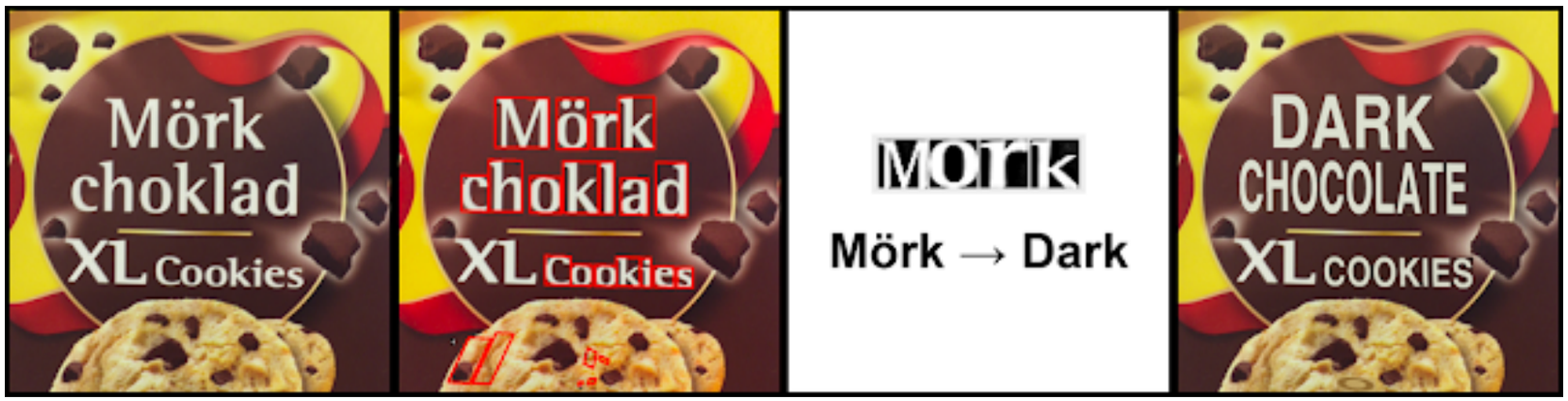
$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad \text{-1}(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

WHY DO WE LIKE DEEP LEARNING?



https://ai.googleblog.com/2015/07/how-google-translate-squeezes-deep.html

SO WHY SHOULD I SKIP IT?

- Neural nets are an extension on traditional machine learning techniques.
- If you don't understand the traditional techniques, I don't want you on my team - even if you're a neural net wizard.
- This sentiment is not just me.

SO WHY SHOULD I SKIP IT?

- Neural nets are extremely complicated models. A “real” model may have hundreds of thousands of parameters.
- Rules of thumb for how to design your architecture are hard to come by.

SO WHY SHOULD I SKIP IT?

- Complicated models need a TON of data and computational effort to work well.
- Most businesses don't have the resources to effectively weaponize deep learning.

SO WHY SHOULD I SKIP IT?

- Most regulated industries still don’t allow deep learning because it’s too “black box.”
  - e.g. How can I tell which part of my model is targeting minorities in loan rejections?

SO WHY SHOULD I SKIP IT?

- Most regulated industries still don’t allow deep learning because it’s too “black box.”
  - e.g. How can I tell which part of my model is targeting minorities in loan rejections?
- Sidenote: It’s not a blackbox, it’s just the chain rule from calculus

2

YES, YOU DO NEED TO KNOW HOW TO PROGRAM LIKE YOU AREN'T AN IDIOT

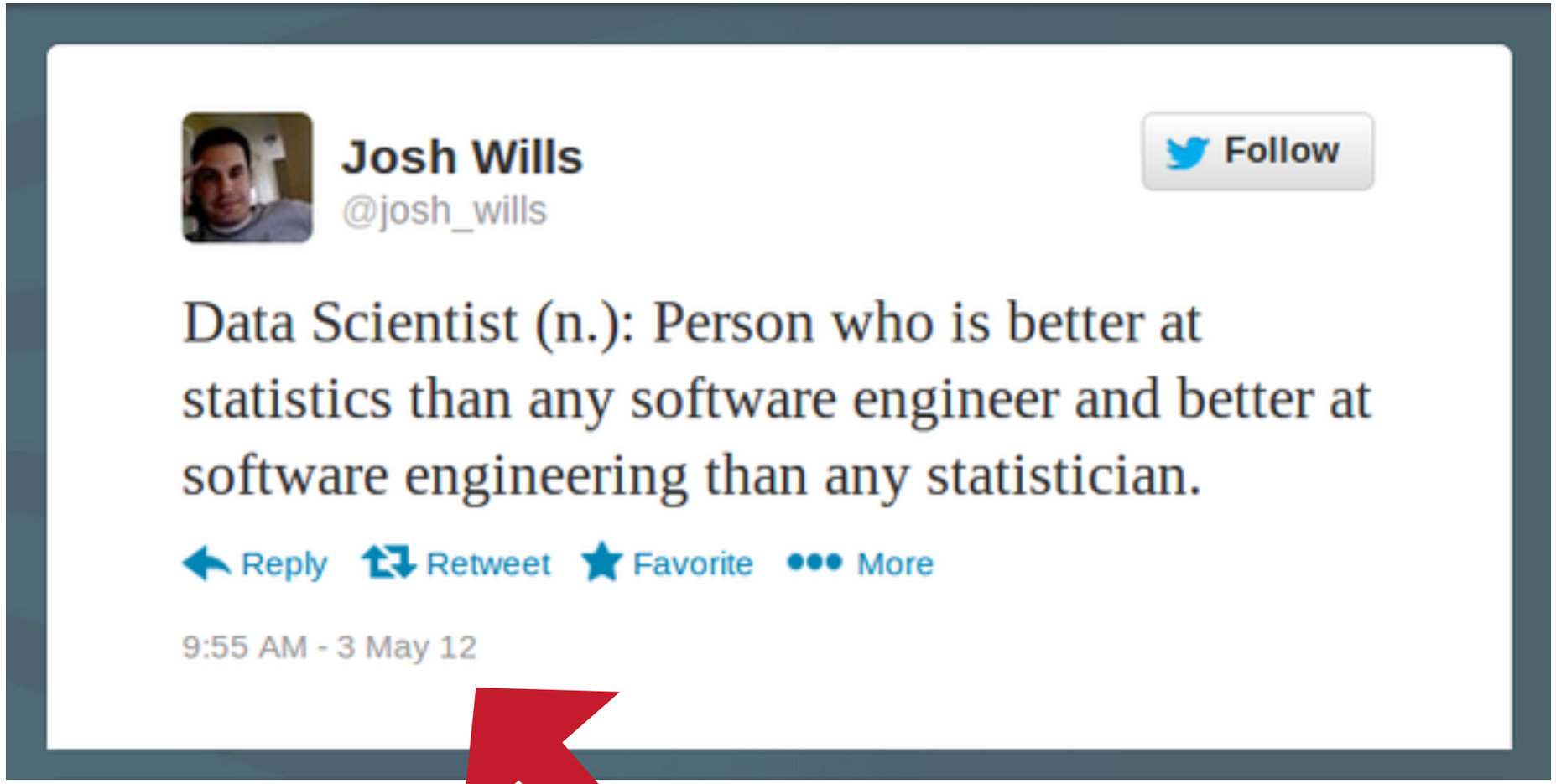
Josh Wills
@josh_wills

Follow

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

Reply Retweet Favorite More

9:55 AM - 3 May 12



2012 IS NOT 2018

THE MODERN DATA SCIENTIST

- Data is in a data warehouse or cluster.
- Prototyping still happens - but it has to be productionizable at the end.
- Data science teams are no longer just that one weird physicist guy that celebrates Pi day. Your code is part of a stack that your team relies on.

THE MODERN DATA SCIENTIST

- Jupyter Notebooks are not acceptable unless you're on a team of 1 or 2. You have to know how to write modules, link code together, and optimize your code.

THE MODERN DATA SCIENTIST

- Jupyter Notebooks are not acceptable unless you're on a team of 1 or 2. You have to know how to write modules, link code together, and optimize your code.
- Even if you're on a team of 1, you really shouldn't be treating notebooks as the end point.

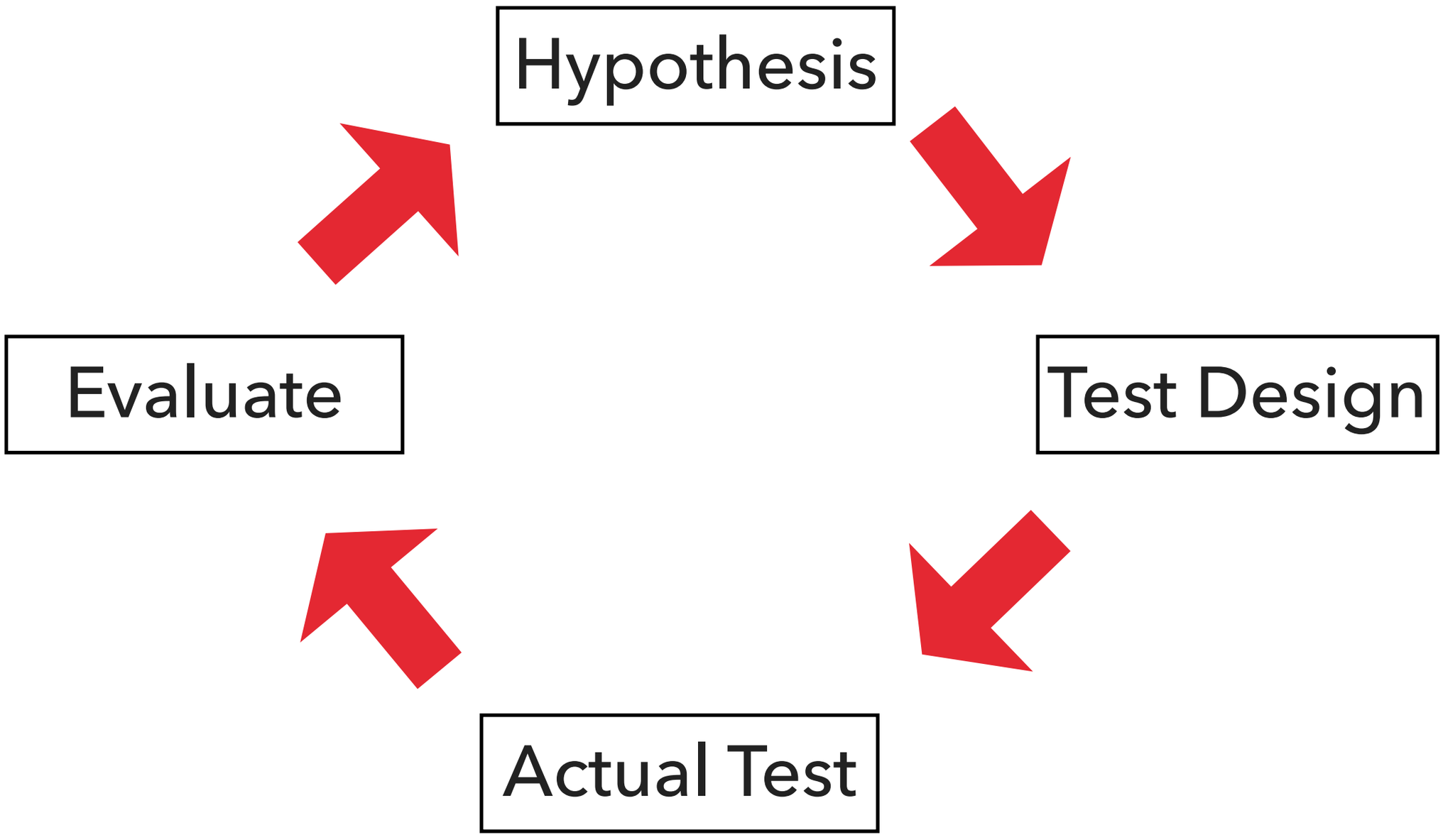
DOCUMENT YOUR CODE.

YOU ARE NOT AN ANIMAL.

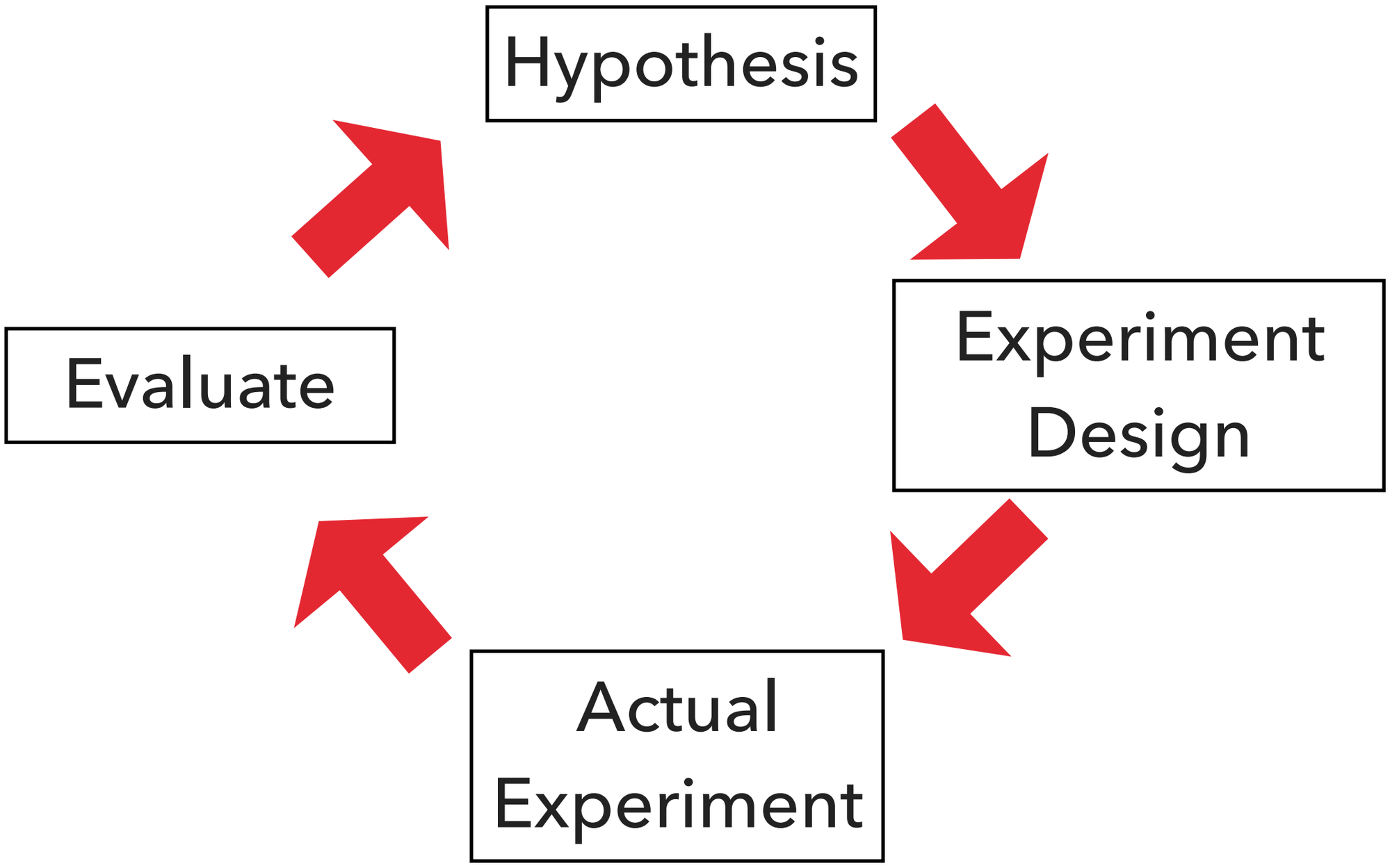
3

DATA SCIENCE SHOULD BE SCIENCE DONE WITH DATA

THE DATA SCIENCE CYCLE



THE SCIENCE CYCLE



DATA SCIENTISTS AREN'T JUST "DATA STORY TELLERS."

WE'RE INVESTIGATIVE JOURNALISTS/INTERPRETERS.

DATA STORY TELLERS?

Story Teller	Interpreter
▸ Leaves out the boring parts ▸ Changes facts for the sake of the story ▸ Don’t need to prove the story, just make it entertaining.	▸ Finds a way to convey the confusing parts in the needed language ▸ Sticks to the data and the facts ▸ Uses statistics to justify any and all choices.

THANKS!

LET'S CHAT. I'D LOVE TO TALK ABOUT PROJECTS YOU'RE CONSIDERING.

ZACH@THISISMETIS.COM

ZWMILLER.COM

Unlisted Sources

http://www.interactive-biology.com/3247/the-neuron-external-structure-and-classification/