

# BUILDING YOUR DATA ACADEMY

# OBJECTIVES FOR TODAY

---

- What is meant by data academy?
- How to build one
  - Selling it
  - Designing it
  - Implementing it
  - Evaluating it



# WHERE IS HE COMING FROM?

- Author and instructor
- Blogger and content developer
- Former bewildered data analyst...

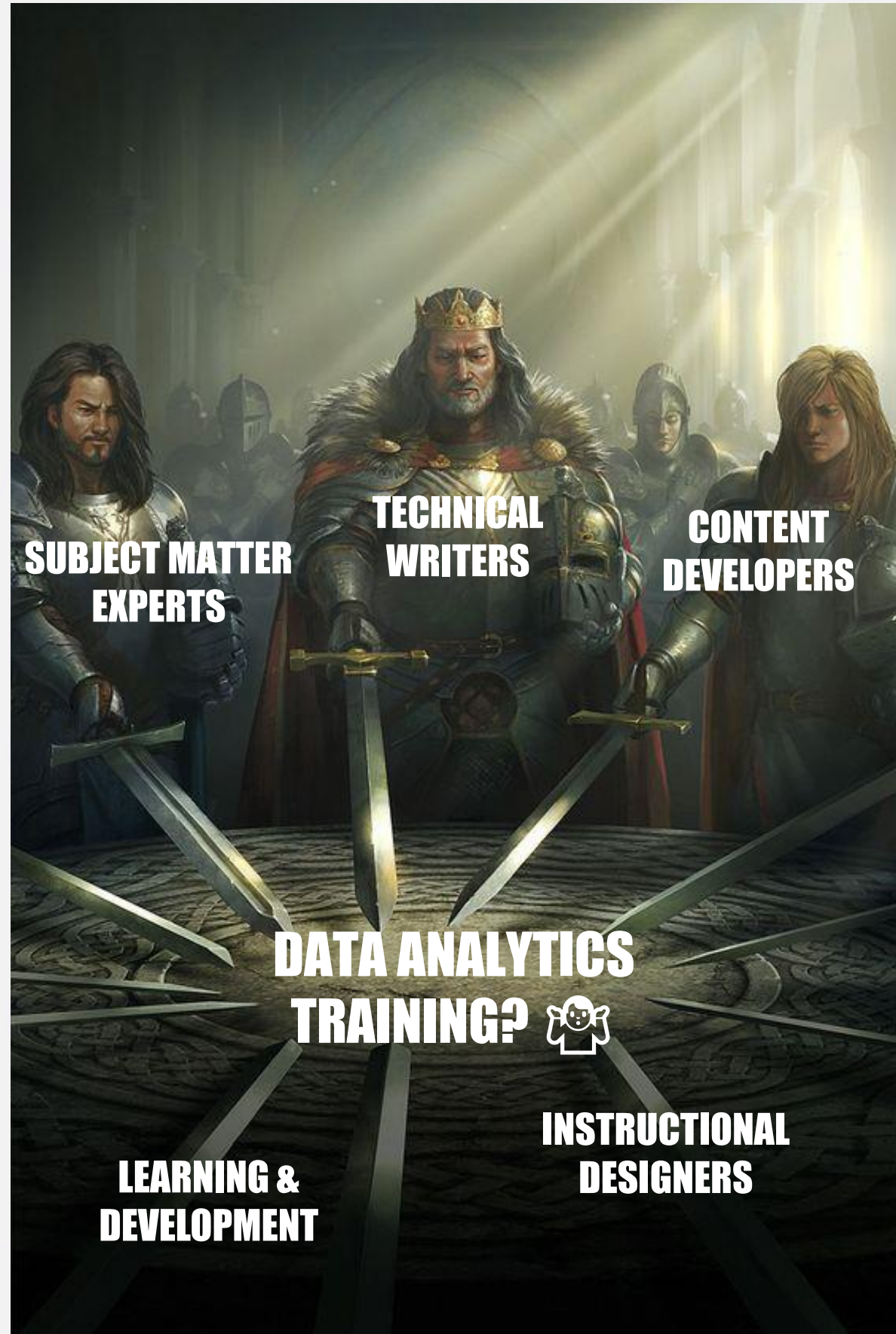
**THINKFUL**  
a Chegg® service



PLURALSIGHT



# WHERE ARE YOU COMING FROM?





# READY OR NOT...



Careers [Cc](#)  
About us > N

Zurich leads the

December 5, 2019

- Zurich develops industry's first levy fu
- Supports Zurich's global data commit
- Funded with over £500K by Governm

Zurich has launched a data academy a people with skills for the future. In an a data academy.

In order to keep up with the new skills partnership with Arch Apprenticeships

Candidates will join the academy from HR. The apprenticeship will be deliver sessions, online study, hackathons and

Zurich's data academy apprentices will to benefit customers. This includes ke innovative solutions beyond insurance

Data, as a key discipline in the insuran Data Officer growing from 12 to 120 ii

## Press release

### Amazon Pledges to Upskill 100,000 U.S. Employees for In-Demand Jobs by 2025

July 11, 2019 at 7:30 AM EDT

*Amazon will invest over \$700 million to provide upskilling training programs for one in three of its employees across the U.S.*

*Programs will help Amazonians from all backgrounds access training to move into highly skilled technical and non-technical roles across the company's corporate offices, tech hubs, fulfillment centers, retail stores, and transportation network, or pursue career paths outside of Amazon*

*Based on a review of its workforce and analysis of U.S. hiring, Amazon's fastest growing highly skilled jobs over the last five years include data mapping specialist, data scientist, solutions architect and business analyst, as well as logistics coordinator, process improvement manager and transportation specialist within our customer fulfillment network*



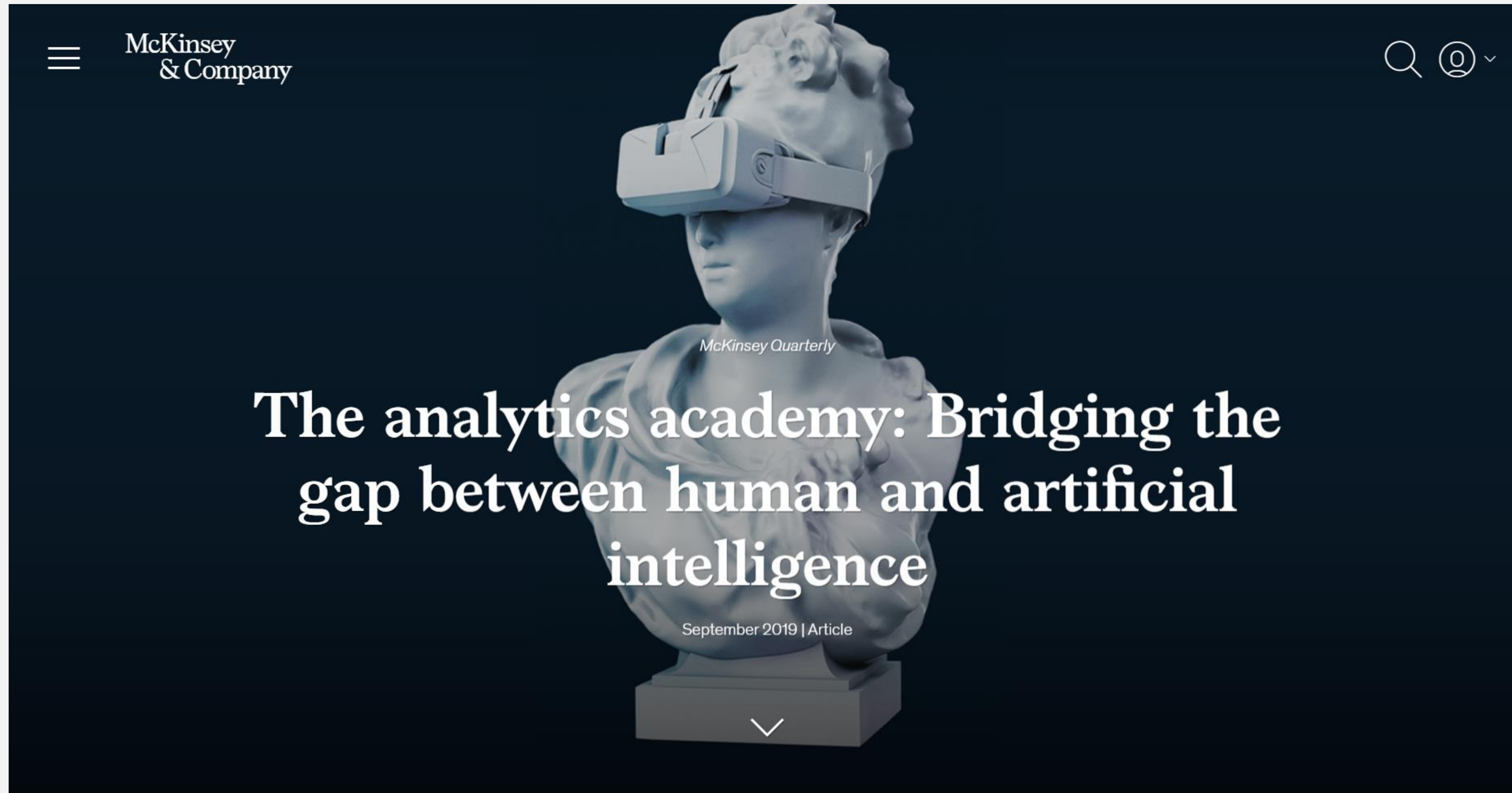
Share

Search

- <https://www.rbs.com/rbs/news/2019/01/natwest-becomes-first-uk-bank-to-launch-data-academy.html>
- <https://www.zurich.co.uk/en/about-us/media-centre/company-news/2019/zurich-leads-the-way-with-sectors-first-data-academy>
- <https://press.aboutamazon.com/news-releases/news-release-details/amazon-pledges-upskill-100000-us-employees-demand-jobs-2025>



# ROLL CREDITS? NOT SO FAST



<https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-analytics-academy-bridging-the-gap-between-human-and-artificial-intelligence>





**STEP 2: AI-  
POWERED  
ENTERPRISE**

**STEP 1: CONNECT  
COMPUTER TO  
PRINTER**





# IT'S NOT MAGIC, YET

## *For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights*



Monica Rogati, Jawbone's vice president for data science, with Brian Wilt, a senior data scientist. Peter DaSilva for The New York Times

By Steve Lohr

Aug. 17, 2014



Yet far too much handcrafted work — what data scientists call “data wrangling,” “data munging” and “data janitor work” — is still required. Data scientists, according to interviews and expert estimates, spend from 50 percent to 80 percent of their time mired in this more mundane labor of collecting and preparing unruly digital data, before it can be explored for useful nuggets.

<https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html>

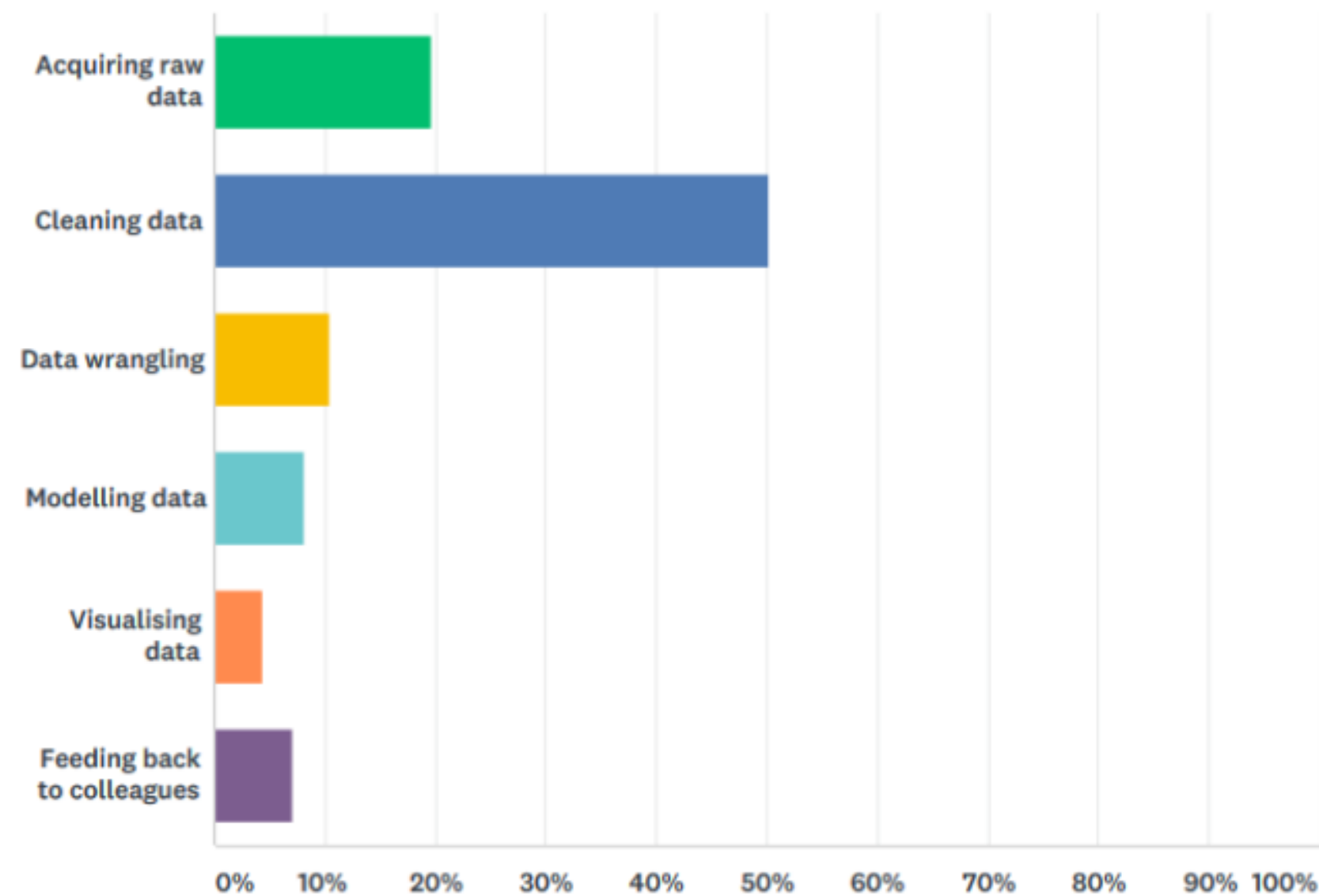




# PEOPLE DON'T LIKE IT

## Why data cleaning is a nightmare

In the recently conducted [Packt Skill-Up survey](#), we asked data professionals what the worst part of the [data analysis](#) process was, and a staggering 50% responded with data cleaning.



<https://hub.packtpub.com/data-cleaning-worst-part-of-data-analysis/>



# DATA ANALYSTS WANT BETTER

Sectors with the Highest Turnover Rates		
1.	 Technology (Software)	13.2%
2.	 Retail & Consumer Products	13.0%
3.	 Media & Entertainment	11.4%
4.	 Professional Services	11.4%
5.	 Government/Edu/Non-Profit	11.2%
6.	 Financial Services & Insurance	10.8%
7.	 Telecommunications	10.8%
8.	 Oil & Energy	9.7%
9.	 Aero/Auto/Transport	9.6%
10.	 Healthcare & Pharmaceutical	9.4%

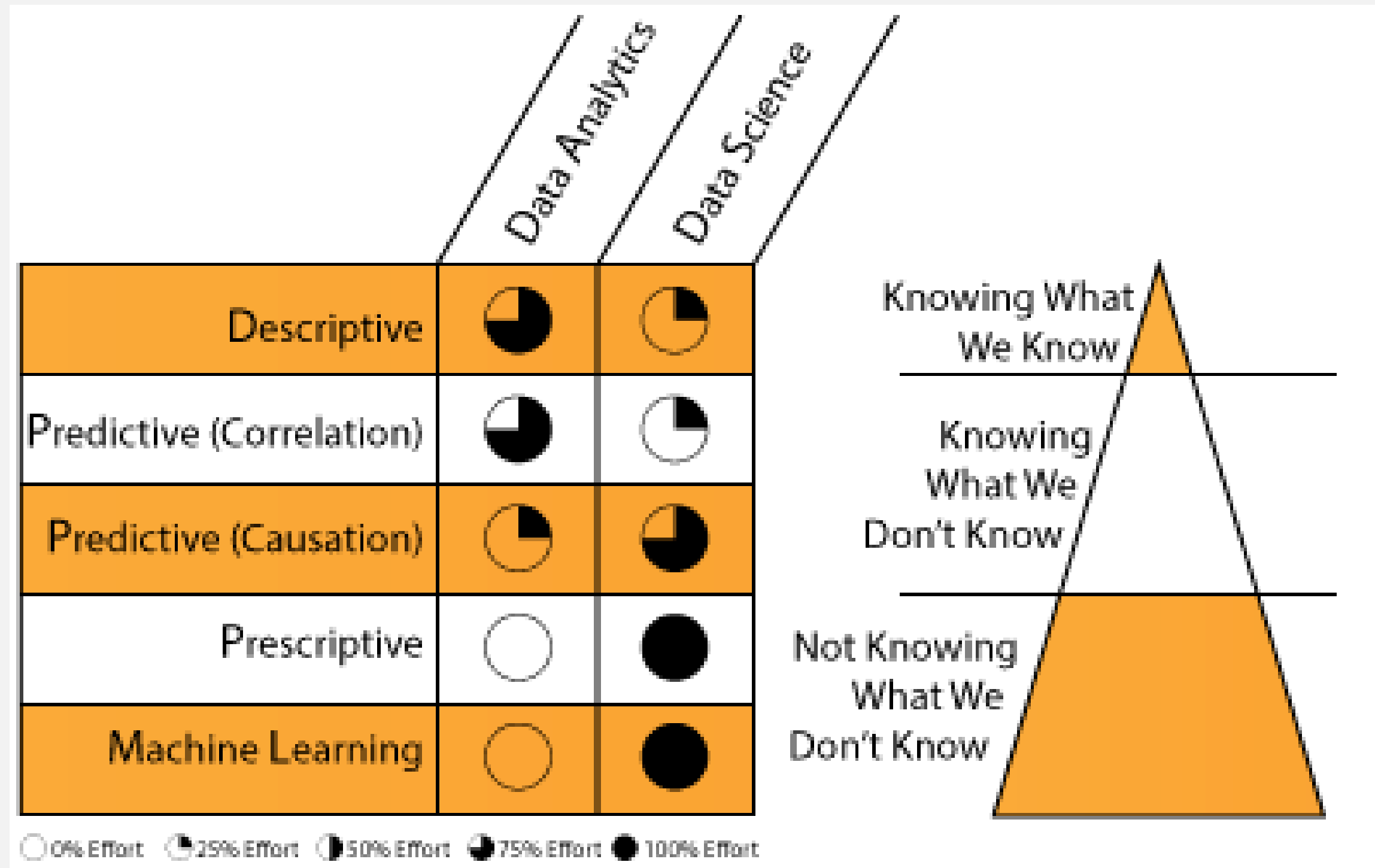
## Occupations within Technology (Software) Sector with the Highest Turnover Rates

- 1 User Experience and Designer • **23.3%**
- 2 Data Analyst • **21.7%**
- 3 Embedded Software Engineer • **21.7%**

<https://business.linkedin.com/talent-solutions/blog/trends-and-research/2018/the-3-industries-with-the-highest-turnover-rates>



# THE ANALYSIS VS THE SCIENCE



# QUESTIONS?





# SELLING THE DATA ACADEMY



# QUESTION FOR THE GALLERY

What do you think the most common objections would be to building a data academy?



# WHY DO THIS IN-HOUSE?





# WHY BUILD SOMETHING FORMAL?





# WHY NOT JUST HIRE GENIUSES?



# CUTTING THE DATA BREADLINE

- 71.7% of firms report that they have yet to forge a data culture
- 69.0% of firms report that they have not created a data-driven culture
- 53.1% of firms state they are not yet treating data as a business asset
- 52.4% of firms claim that they are not competing on data and analytics.

NewVantage Partners, “[Big Data and AI Executive Summary 2019](#)”



# IT ALL GOES BACK TO MESSY DATA

*Happy families are all alike; every unhappy family is unhappy in its own way.*

Leo Tolstoy

*Tidy datasets are all alike, but every messy dataset is messy in its own way.*

Hadley Wickham



# QUESTIONS?





# DESIGNING THE DATA ACADEMY

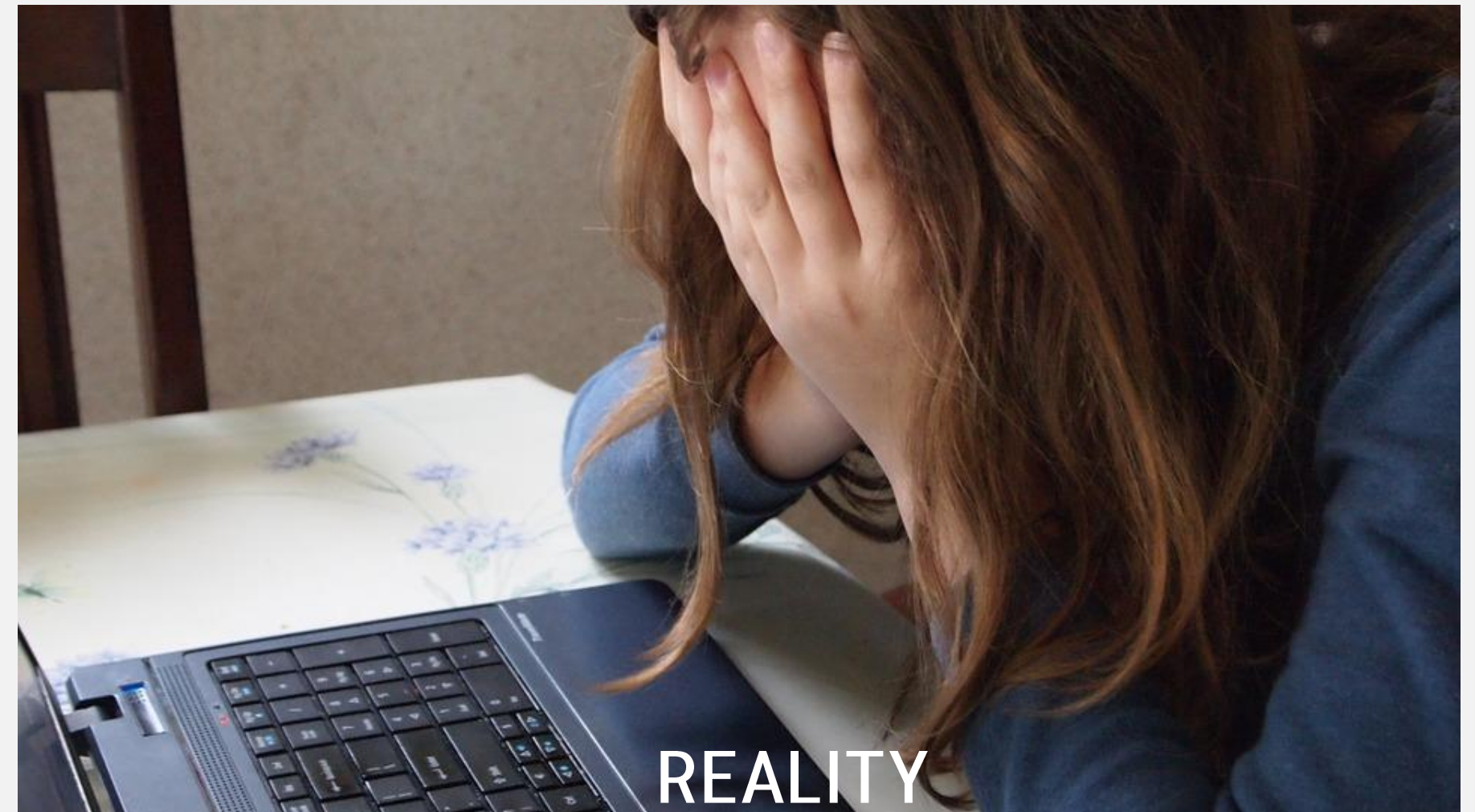


# FINDING YOUR SUBJECT MATTER EXPERTS

1. Content development trumps certifications
2. Generalization trumps specialization
3. Creative mentality trumps technical mentality



# IDENTIFYING YOUR DATA ACADEMITES



# ASSESSING THE ACADEMITES

Quantifying the data gap:

- Self-rating
- Time estimates
- Learning assessments

Qualifying the data gap:

- Experiences
- Emotions
- Goals





# SAMPLE ASSESSMENT QUESTIONS

- Why do you want to enroll?
- Approximately how many hours a week do you spend cleaning and preparing data for analysis and reporting?
- What systems do you pull your data from and to what extent do they interface?
- Rate your proficiency in Excel on a scale from 1 to 5

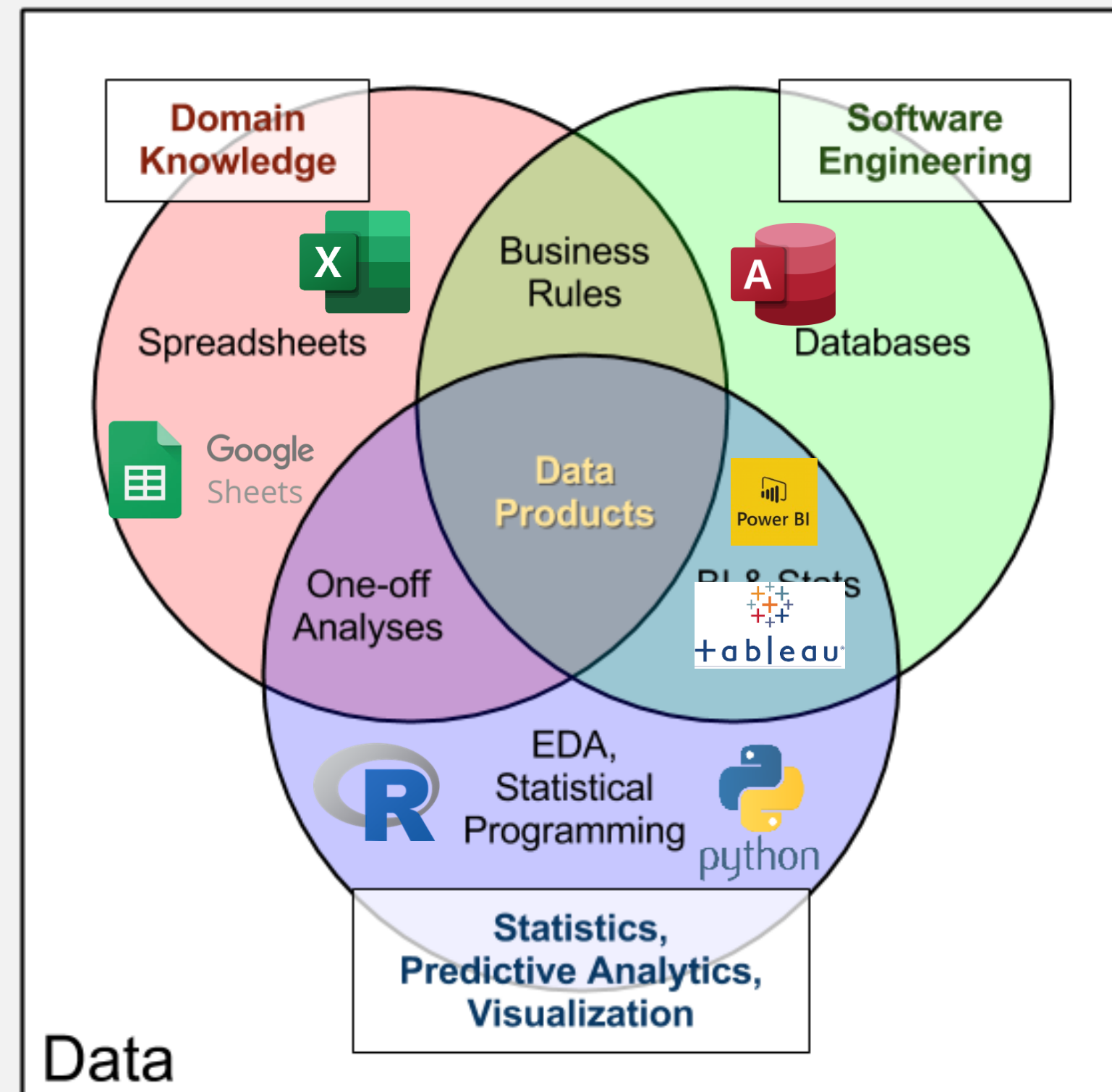


# SAMPLE ASSESSMENT QUESTIONS

- Which of these best describe how easily you could perform a left outer join in Excel Power Query?
  - a. I wouldnt know where to start
  - b. I could struggle through with trial and error and a lot of web searches
  - c. I could do it quickly with little or no use of external help.



# THE DATA ACADEMY STACK



# CREATING THE DATA ACADEMY

- What media will we use?
- How are we going to reach our learners?
- How are we going to engage them?





# DATA EDUCATION AS ROBOT-PROOFER

- 76% want to do their job faster and better
- 75% learn for their own personal development
- 60% want to increase their productivity
- 47% want to keep up with new technology
- 42% are motivated by working towards professional certification

TowardsMaturity, “[What motivates staff to learn online?](#)”



# DATA EDUCATION AS MODE OF INQUIRY

- 76% want to do their job faster and better
- 75% learn for their own personal development
- 60% want to increase their productivity
- 47% want to keep up with new technology
- 42% are motivated by working towards professional certification

TowardsMaturity, “[What motivates staff to learn online?](#)”



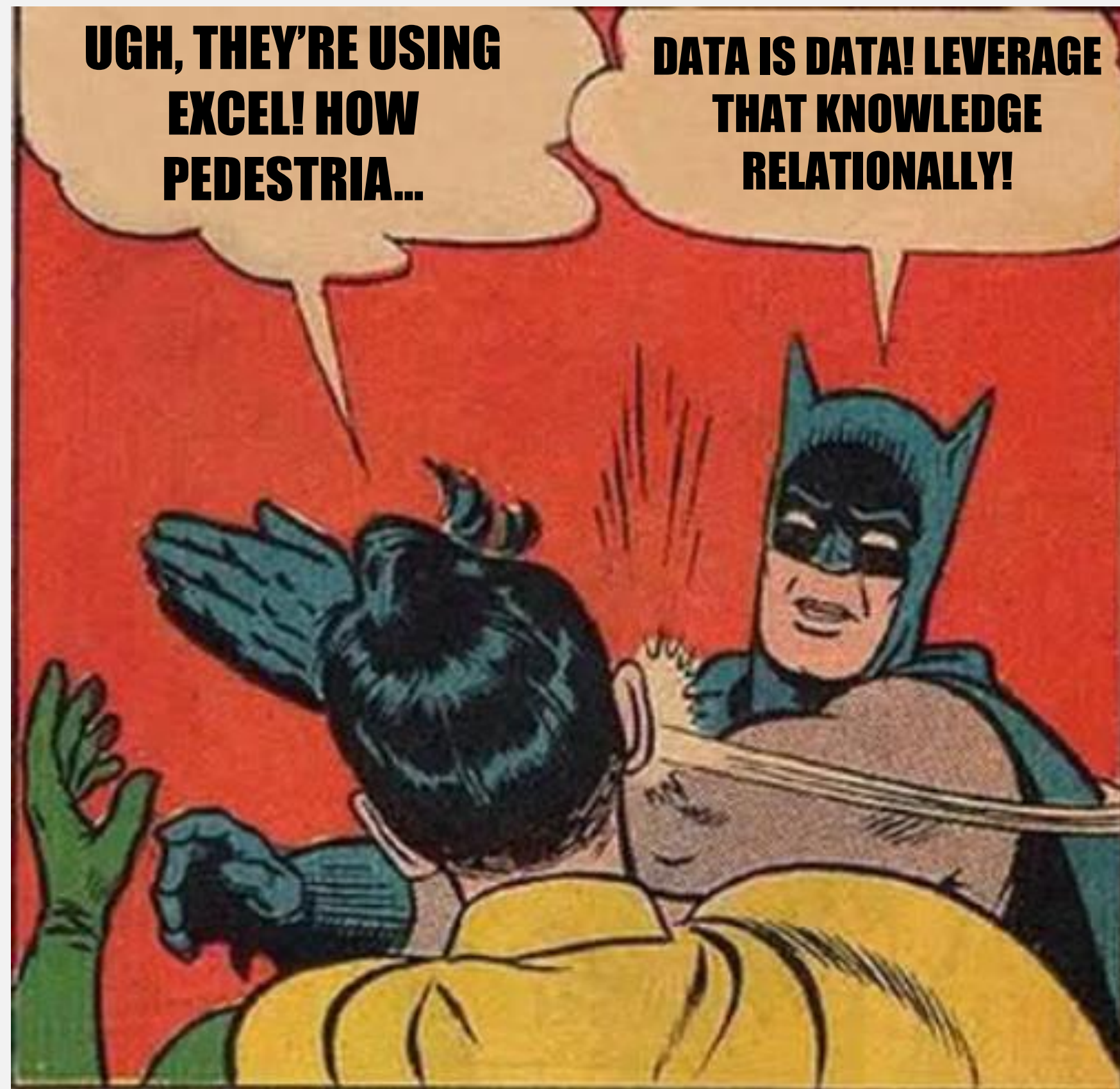
# DELIVERING A UNIFIED EXPERIENCE

## Blended learning:

- In-person sessions
- Online study
- Work-based projects
- Hackathons



# RESPECT EXTANT KNOWLEDGE





# BUY THE HOOK AND YOU'LL BUY THE...

- Proprietary data from the business
- Public data from real life
  - [Baseball records](#)
  - [Baby names](#)
  - [Movie reviews](#)




# GOT GUI'S? GET GIF'S

“To convert a range into a table, select any cell in the range and press Ctrl + T.”

	A	B	C	D	E	F	G	H
1	Day	Sales						
2	Monday	11						
3	Tuesday	56						
4	Wednesday	23						
5	Thursday	40						
6	Friday	10						
7	Saturday	88						
8	Sunday	31						
9								
10								
11								
12								

Create Table ? X

Where is the data for your table?



☒ My table has headers

OK Cancel

AutoSave Off book1 - Excel

File Home Insert Page Layout Formulas Data Review View Developer Help ACROBAT

Clipboard Font Alignment Number

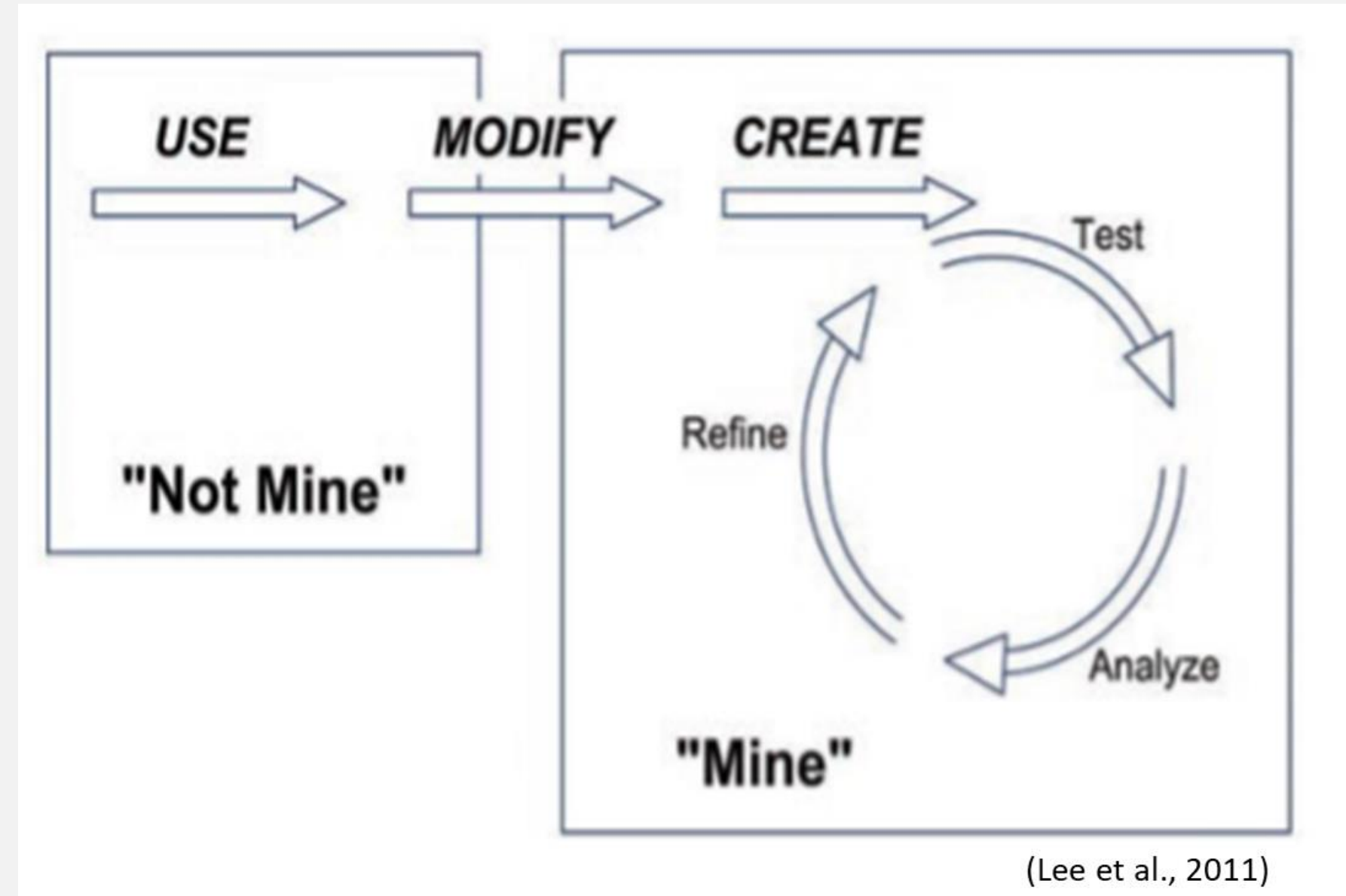
P25

	A	B	C	D	E	F	G	H	I	J	K	L
1	Day	Sales										
2	Monday	11										
3	Tuesday	56										
4	Wednesday	23										
5	Thursday	40										
6	Friday	10										
7	Saturday	88										
8	Sunday	31										
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												



# PRIMM IS PROPER

- Predict
- Run
- Investigate
- Modify
- Make



# **AVOID THE DREAD BLANK PAGE**





# PARSONS PROBLEMS

Arrange these blocks so that you have a function that returns the beginning and end of a list:

*Drag blocks from here*

```
return [a_list.first, a_list.last]
```

```
def list_ends(a_list):
```

```
def list ends(a_list):
```

```
return [a_list[0], a_list[-1]]
```

*Drop blocks here*

# FADED EXAMPLES (SCAFFOLDING)

Import the file `iris.csv` as a DataFrame and print the first ten rows.

```
import ____ as pd

# Read in iris.csv
____ = pd.____(____)

# Print first ten rows
iris.____
```

```
import pandas as pd

# Read in iris.csv
iris = pd.read_csv('iris.csv')

# Print first ten rows
iris.head(10)
```



# EMPHASIZE DEBUGGING



"Sometimes it pays to stay in bed on Monday, rather than spending the rest of the week debugging Monday's code." - Christopher Thompson



# PAIR PROGRAMMING: DOUBLE YOUR FUN WITH DATA





# QUESTIONS?



# MEASURING THE DATA ACADEMY'S EFFECTIVENESS



# IT'S YOUR WORLD... THE DATA IS LIVING IN IT

1. Time savings
2. Data democratization as employee benefit
  - Retention
  - Morale
  - Lower barriers to entry for data analytics

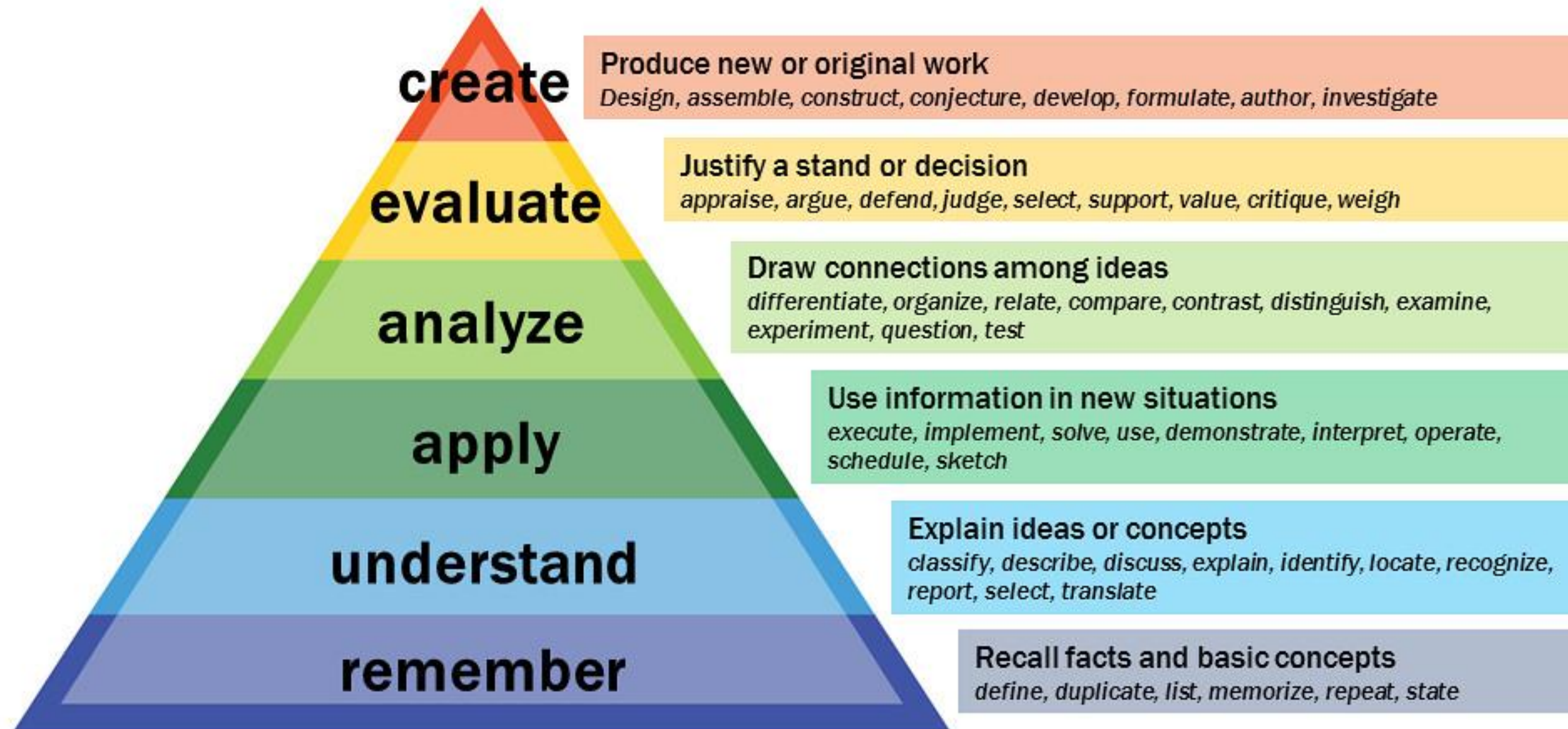


# **FUTURE-PROOFING THE DATA ACADEMY**



# YOU REALLY DIDN'T THINK YOU WOULDN'T SEE THIS, RIGHT?

## Bloom's Taxonomy





# OPEN-SOURCE YOUR DATA PROCESSES

FINANCE

## Goldman Sachs is planning on giving some of its most valuable software to Wall Street for free

PUBLISHED WED, NOV 20 2019•9:25 AM EST | UPDATED WED, NOV 20 2019•1:09 PM EST



Hugh Son  
@HUGH\_SON

SHARE



### KEY POINTS

- Goldman is taking the unusual step of making its data modeling program Alloy, as well as the language underlying it, available to the rest of Wall Street for free as open-source software.
- By giving it away, Goldman hopes to reduce its own costs when dealing with trading counterparties among banks and asset managers and compliance requests from regulators.
- “We’re using Alloy because it radically reduces the cost of wrangling disparate datasets and disparate sources of data together” said Goldman’s co-chief data officer, Jeff Wecker.

### How digital thread can speed manufacturing throughput

A digital thread quickly moves a siloed applications environment suffering from an unwieldy “spaghetti architecture” to a dynamic, interactive ecosystem with appropriate data in play across the enterprise and its supply chain.

The term digital thread might

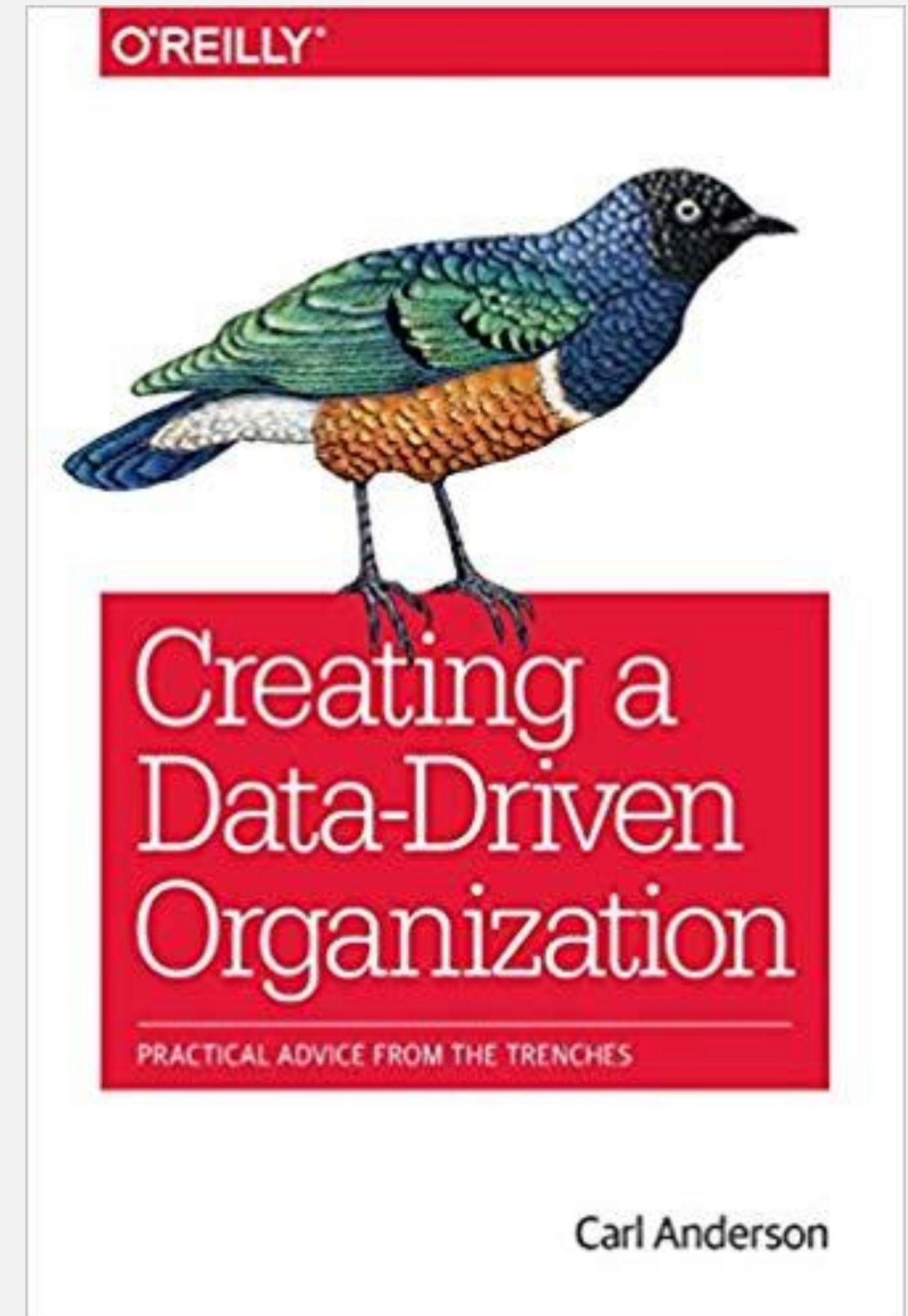


# CONCLUSION



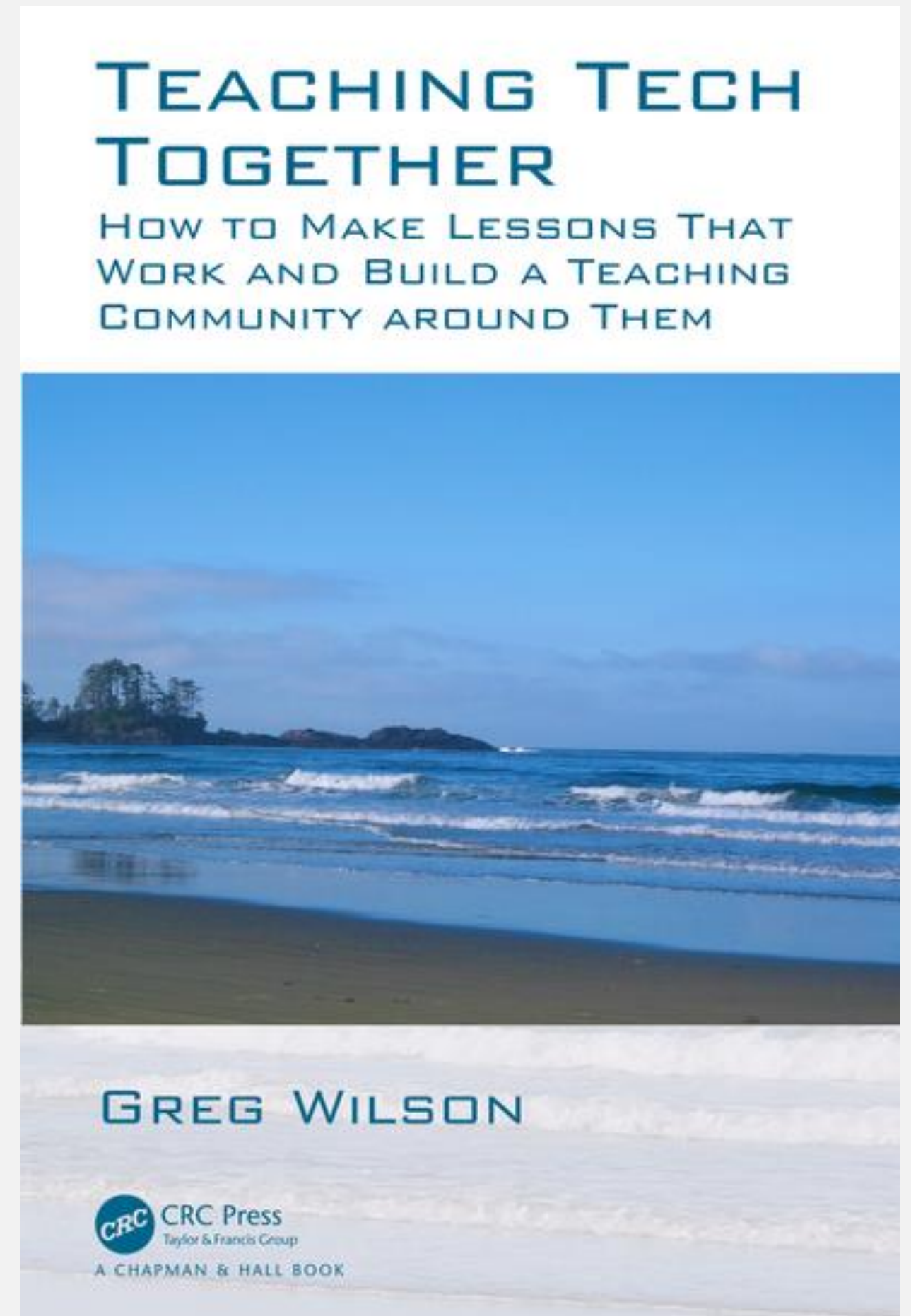
# *Creating a Data-Driven Organization* by Carl Anderson

- <https://learning.oreilly.com/library/view/creating-a-data-driven/9781491916902/>



# *Teaching Tech Together* by Greg Wilson

- Read for free at <https://teachtogether.tech>





# “Let’s start a data academy!” by Johns Hopkins/GovEx

- <https://govex.jhu.edu/wiki/start-data-academy/>

**GovEx**  **JOHNS HOPKINS UNIVERSITY**

## LET’S START A DATA ACADEMY!

“If you build it, they will come.” But when it comes to creating a data academy, it may not be that easy. While more governments are creating data academies to help their employees learn how to use data more effectively, there are a variety of choices to make them successful. Having a strategy is absolutely essential to build your data academy, reach the people you want, and achieve the outcomes you need. The simple steps in this worksheet can help you get started.

1

**Define your audience** - You may have an idea of who you want to reach, but these questions may help define them even more.

Who in your organization has data needs?

At what levels of the organization are they?

How will you assess the needs of different people and roles?

2

**Potential Participants** - Planning outreach to your audience is critical to launching your academy successfully.

How will you identify and reach potential participants?

Do you have contacts within that audience or know someone who does?

How will you select and engage participants?

3

**Engagement Strategy** - It’s important to know prior to starting, whether you will engage across the organization or within departments.

What kind of problems will you try to address in your academy?

Do those problems reach across departmental lines?

Are the data for these problems located within one department?





# Stringfest Analytics Resource Library

- [george@stringfestanalytics.com](mailto:george@stringfestanalytics.com)



## PYTHON: HALF-DAY WORKSHOP

Lesson 1: Up and running with Python + Jupyter

Lesson 2: Introduction to Python programming

Lesson 3: Working with lists

Lesson 4: Working with functions and methods

Lesson 5: Working with modules

Lesson 6: Capstone

## Learning Objectives

- Student can create, navigate and download Jupyter notebooks for Python
- Student can assign variables and perform basic operations on variables
- Student can create, inspect and modify lists
- Student can pass lists into functions and methods
- Student can install, explore and implement elements of a module
- Student can create and analyze lists using Python modules, methods and functions

Lesson plan developed by George Mount. For more resources like this, visit [stringfestanalytics.com](http://stringfestanalytics.com)



# LET'S TALK

## LINKEDIN

[linkedin.com/in/gjmount](https://www.linkedin.com/in/gjmount)

## EMAIL ADDRESS

[george@stringfestanalytics.com](mailto:george@stringfestanalytics.com)

## WEBSITE

[stringfestanalytics.com](https://stringfestanalytics.com)



# QUESTIONS?

