# DSC96 Final Lecture

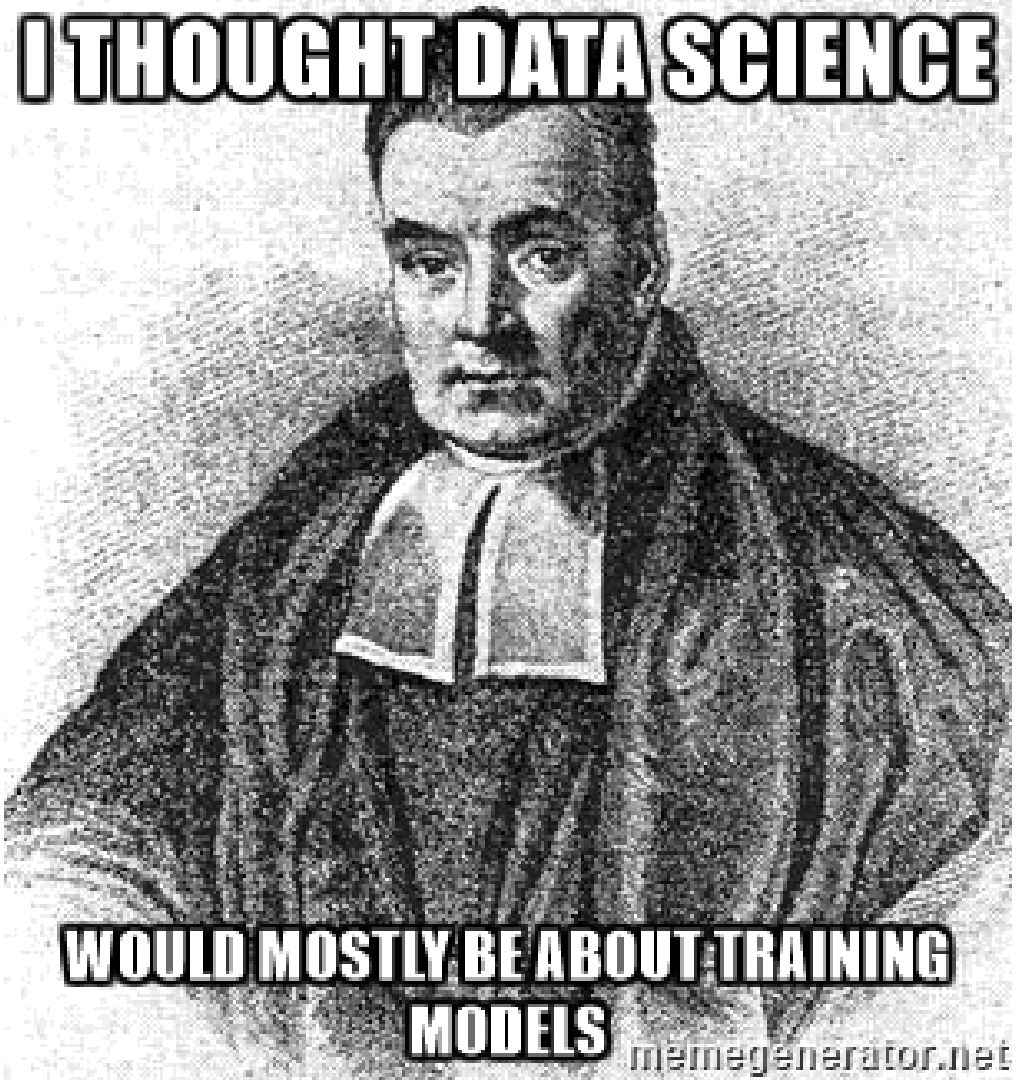
Colin Jemmott, June 2019

# Part I

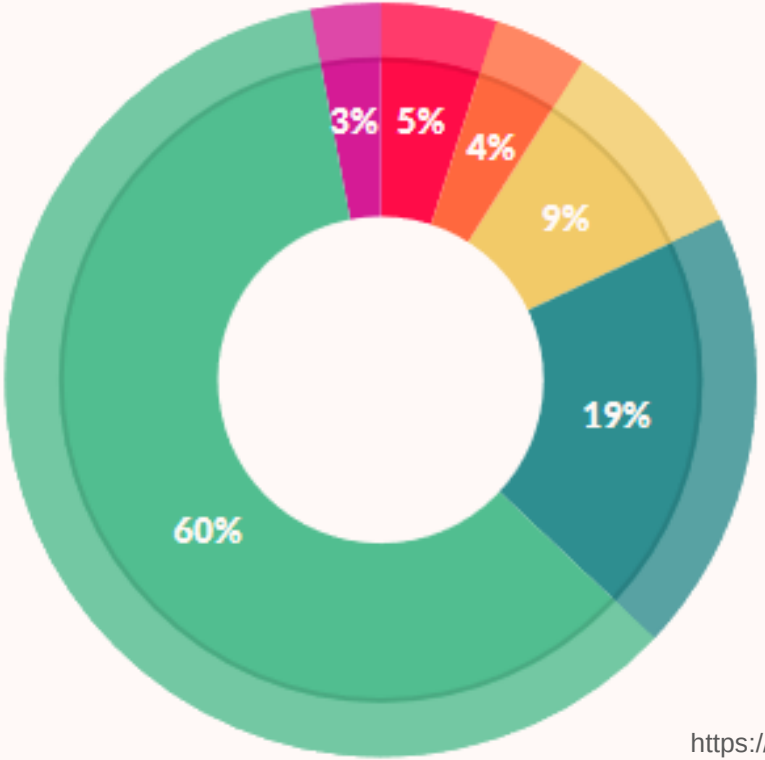
Data Science

# So now what?

aka: What the hell?  Coursera said I can be a data scientist in one week for $29.95.  Some dude donated millions of dollars and I worked my ass off for ten weeks.  So why don’t I know data science yet?


I THOUGHT DATA SCIENCE
WOULD MOSTLY BE ABOUT TRAINING MODELS
memegenerator.net

# What data scientists do all day




**What data scientists spend the most time doing**

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

# The Data Science Process

1. Identify the question
    - What is the goal?  What is the scope?
2. Prepare the data
    - Find, source, clean.  Check completeness, fix anomalies, perform QA
3. Analyze the data
    - Build models, data mine, run text analysis, etc.
4. Visualize the data
    - Complex results -> easy-to-digest visuals
5. Present your findings
    - Document, UI/UX, productize.

From Kirill Eremenko’s “Confident Data Skills” book. (highly recommended)

# Ethics

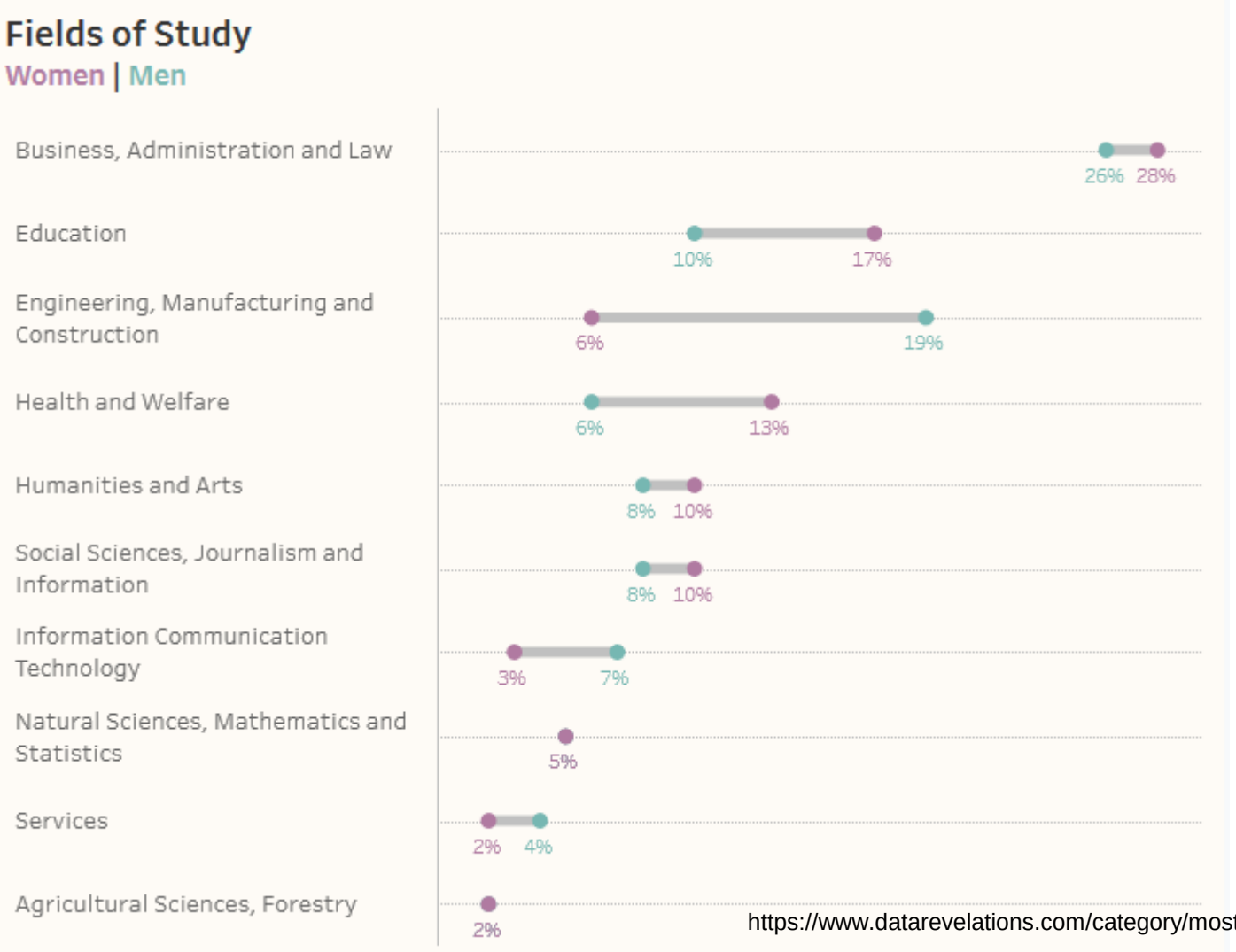
The Five Cs:

1. Consent
2. Clarity
3. Consistency and Trust
4. Control and Transparency
5. Consequences

“Ethics and Data Science” by by DJ Patil, Hilary Mason, Mike Loukides

# Visualization


https://www.datarevelations.com/category/mostly-monthly-makeovers




https://www.datarevelations.com/category/mostly-monthly-makeovers


https://haveibeenpwned.com/

https://www.google.com/maps/timeline

# Part II

## My Principles of Data Science

Longer version with links: https://medium.com/seismic-data-science/seismics-guiding-principles-for-data-science-957ef54e064e

New data should be able to change your mind, even about strongly held beliefs.

Good understanding of data beats good algorithms every time.

Robustly and openly review others’ work, and have your work reviewed by them.

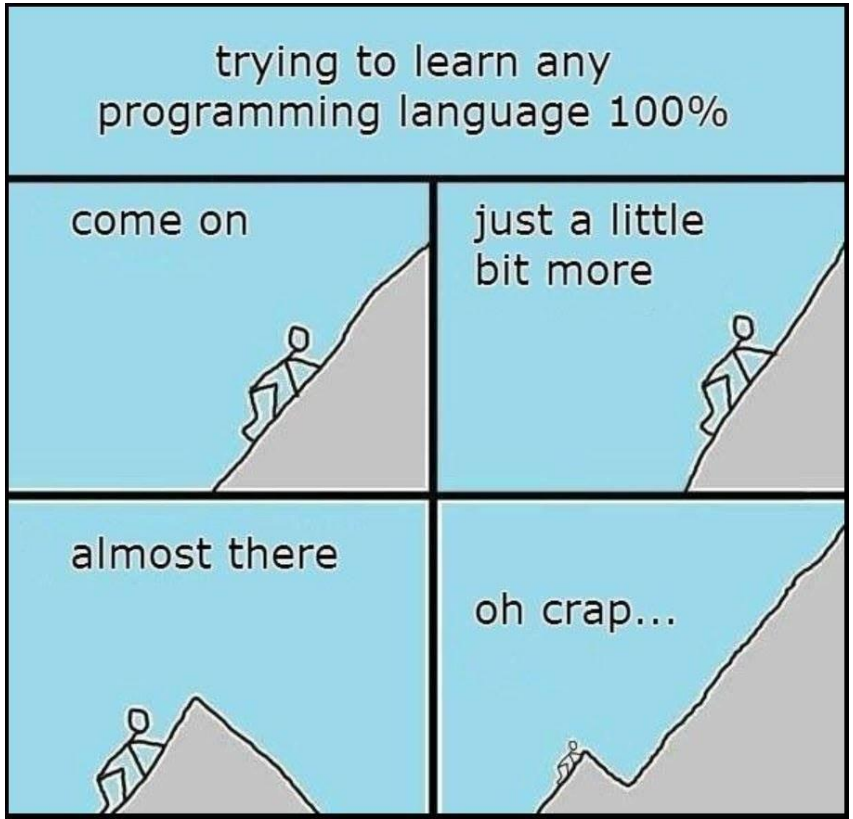
Start by solving easy problems with simple methods.

If a data consumer is mislead, they may make important decisions that are based on falsehoods.

# Part III

The Future

# The near future

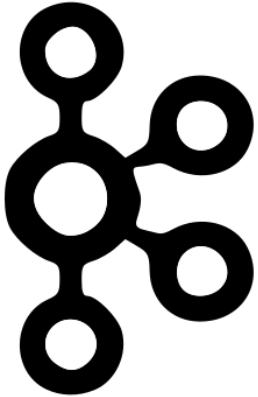
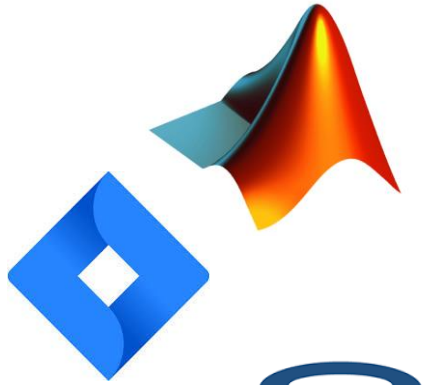


# The Future After UCSD

Get a job!

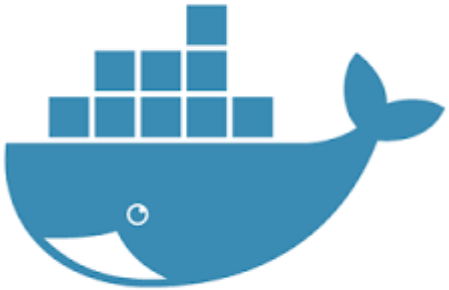
1. Skills
2. Projects
3. Portfolio
4. Interview

# Skills






# ~~Skills~~ Concepts

- Querying: SQL
- Web Visualization: d3.js
- Cloud: AWS
- Exploration: Tableau
- Collaborate: github
- Machine Learning: sklearn
- Anything weird: google for the python package

Recommendations are Colin’s perception of the best / easiest / most fun tool in 2019, but these things change fast, and he is wrong almost all the time.

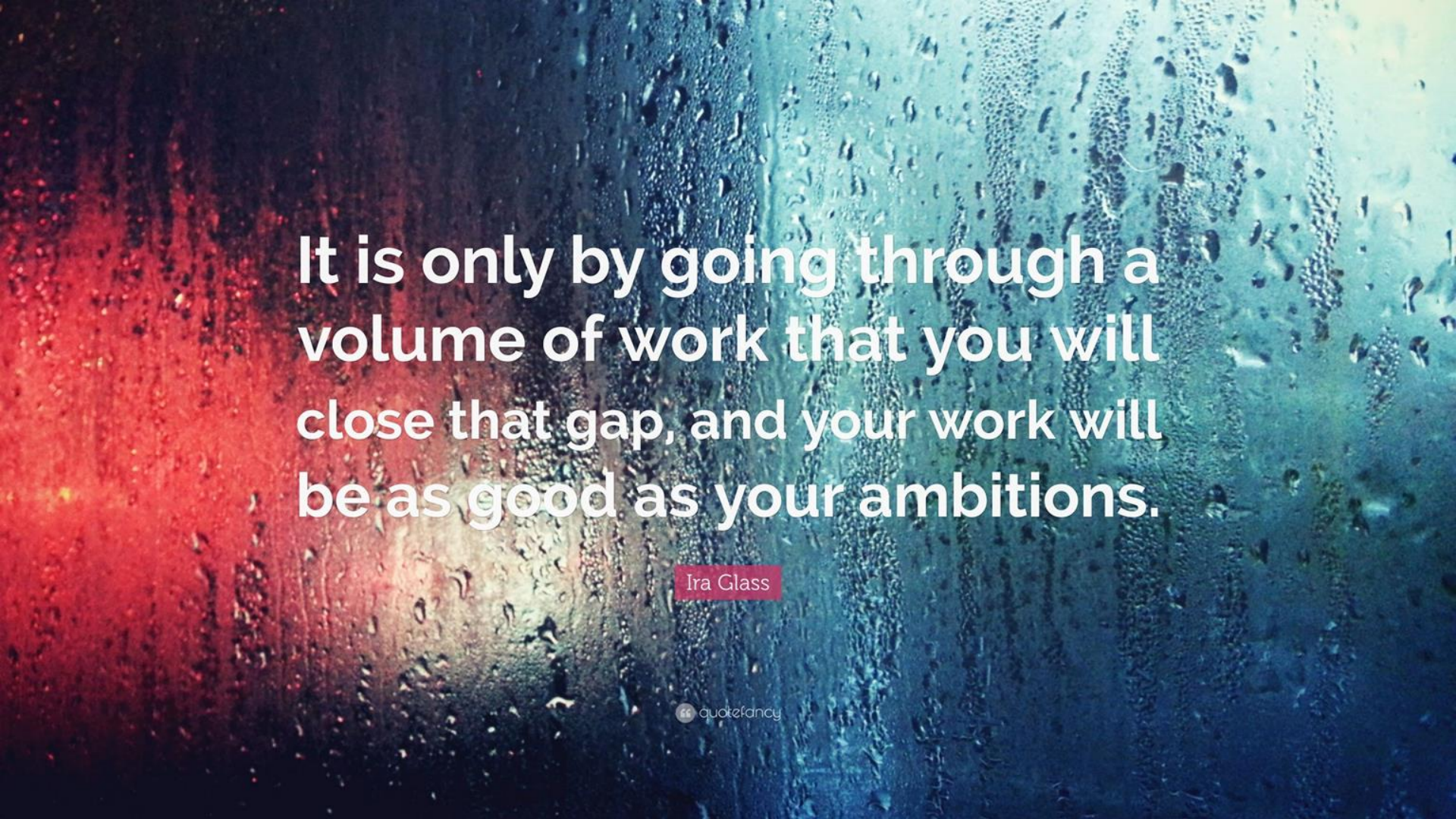
# How to get started

1. Don’t wait to get started.
2. Steal
3. Build what you want to see in the world

The Secret:

Do good work and share it with people

Stolen from the book “Steal Like an Artist”.  Really.


It is only by going through a volume of work that you will close that gap, and your work will be as good as your ambitions.
Ira Glass
quotefancy

# You need a portfolio

A resume isn’t enough.  Show off your projects.

There are a ton of options, but some I use:

- Github
- Professional social (Twitter / Medium / LinkedIn)
- Professional personal site

# Scott Cole

personal webpage

Home | Blog | Burritos of San Diego | CV | Data projects

I'm a PhD student studying neuroscience at UC San Diego and working in the Voytek lab studying brain rhythms. Specifically, I am working on a new analysis framework to extract more information from neural signals by parametrizing the waveform shapes of the brain rhythms. You can learn more about this in our recent review and methods paper.

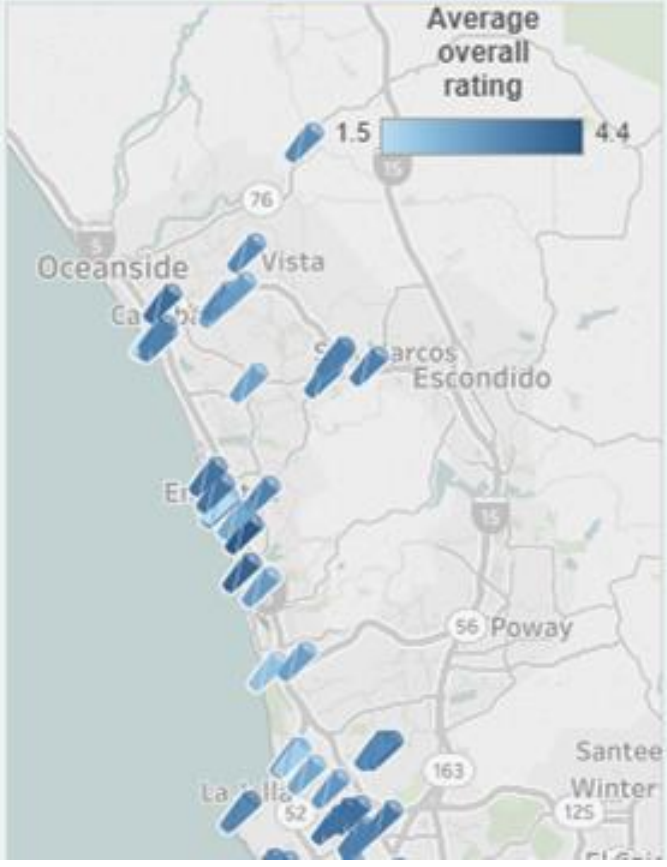
Otherwise, my main passion is in uncovering trends in data. Though I hope I'll be able to apply these skills to improving public good in the future, you can see a list of my projects here ranging from burritos to text mining to police misconduct (CV, resume).



# San Diego Burrito Ratings



350 burritos were rated at 77 taco shops across San Diego using the 10 official burrito dimensions (for more information, go to https://srcole.github.io/100burritos). Use the sliders on the right to filter the map and locate burritos based on your standards.




*Click a burrito on the map to see the details of that restaurant below.*

**Restaurant information**

Neighborhood

(All)

This visualization was based on a template by Will Griffiths, @wilsooon

Overall rating
1.5 — 4.4

Tortilla quality
1.0 — 5.0

Yelp rating
2.5 — 4.5

Temperature
1.0 — 5.0

Cost
$4.00 — $11.00

Meat quality
1.0 — 5.0

# burritos rated
1 — 28

Nonmeat quality
1.5 — 5.0

Volume (liters)
0.50 — 1.20

Meat-to-filling ratio
1.0 — 5.0

# Stretch Goal

Contribute to an open source project.

*“If I was interviewing someone and found out they had a pull request accepted for code we use, I would hire them on the spot.”*

- Ben Taylor, Chief AI Officer at ZIFF

# Prepare for Interviews

## General

- Be ready for coding interviews
  - Cracking the Coding Interview
- Read about how to interview
  - But tech companies can be, ummm, *different*

## Company specific

- Do research about the company
- Do research about how they interview

85,526 views | Mar 1, 2019, 07:45am

# Radical Change Is Coming To Data Science Jobs


**Nate Oostendorp** Forbes Councils
**Forbes Technology Council** CommunityVoice

https://www.forbes.com/sites/forbestechcouncil/2019/03/01/radical-change-is-coming-to-data-science-jobs/#172f57fddfcc

# Radical Change Is Coming To Data Science Jobs

A calculator was once a person. Webmaster was once a hot career.

[A]dvancements in hardware and software took specialized skills and put them into the hands of generalists. While specialist jobs were lost, the democratization of these technologies unleashed waves of innovation, commerce and job creation.

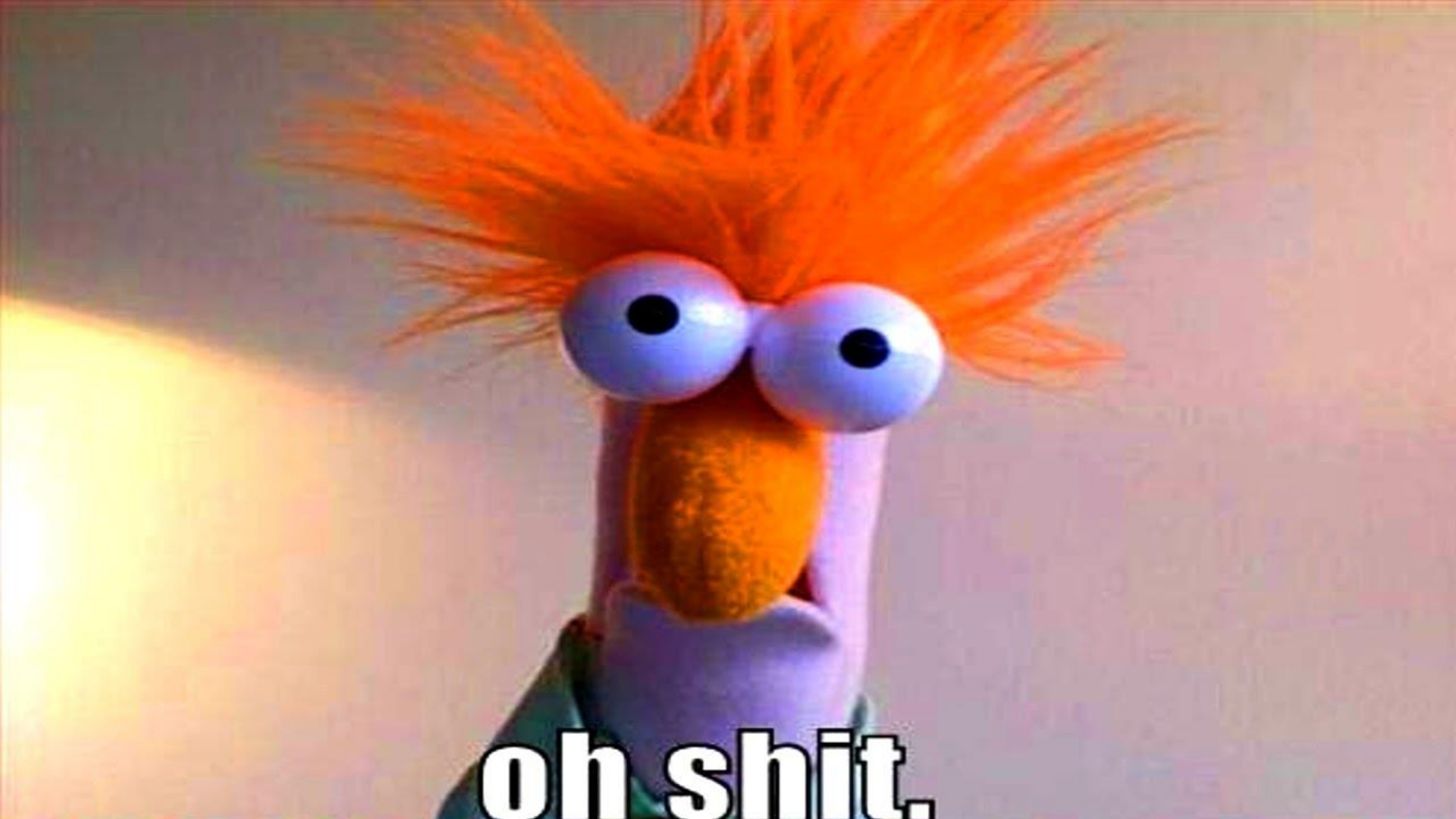
Similarly, I believe the job of data scientist as we know it today will be barely recognizable in five to 10 years….While their data science skills will be a strong career asset, a surprisingly small proportion of them will likely to be working as straight data scientists.

# Radical Change Is Coming To Data Science Jobs

When I studied computer science back in the way-back-when, compiler design was a required course. ... It was common to write pieces of commercial applications in machine language for faster performance.

Over the past few decades, successive layers of software functions have been abstracted into higher-level development tools.

Data science is quickly following the same progression.


oh shit.

# The Good News:

Drawing conclusions from data is always a good career.

Four paths (none of which is “data scientist”):

- Industry Specialist
  - Bring data science to new industries
- Analytics & Data Visualization
  - Bring data science to less technical people
- Data Engineering
  - Remember what all those data scientists are spending their time on?
- Go Deep (get a PhD)
  - Push the boundaries

# Stay in Contact

https://twitter.com/colin_jemmott

https://www.linkedin.com/in/cjemmott/

cjemmott@ucsd.edu

cjemmott@seismic.com

https://www.cjemmott.com/