

# All the ways to break into data science and analytics

SQL Saturday #759

Taras Kaduk

2018/05/05

# What's the crowd?

# What's the crowd?

- Some are in tech, but not in data

# What's the crowd?

- Some are in tech, but not in data
- Some are students or recent graduates

# What's the crowd?

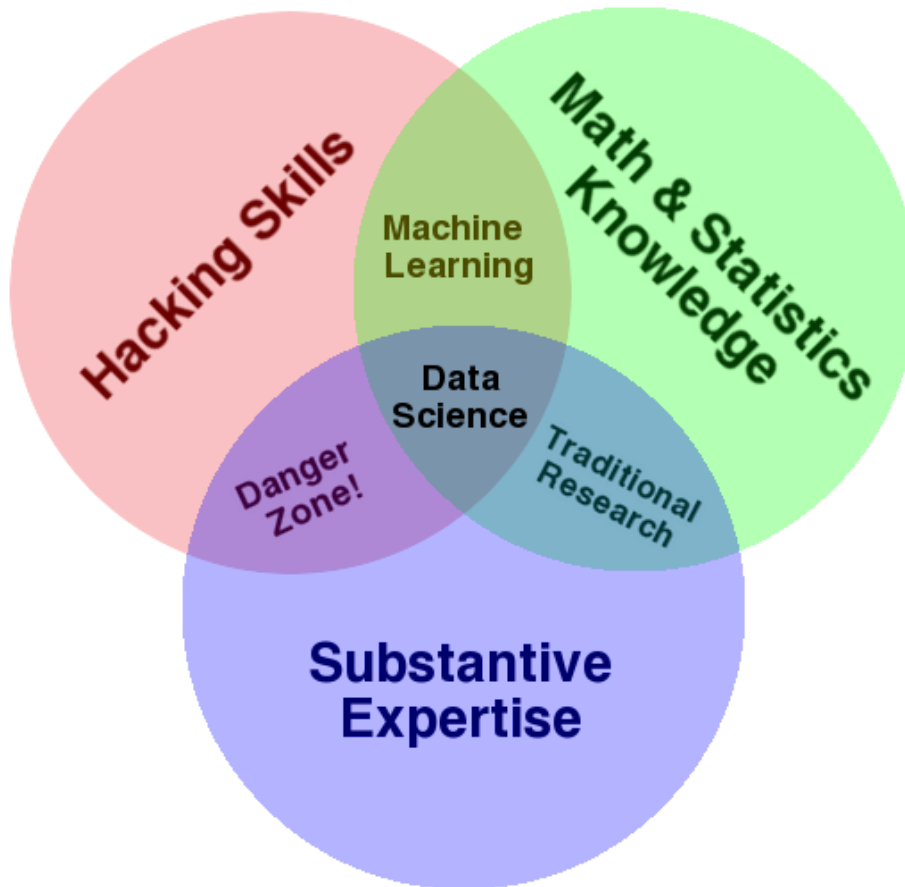
- Some are in tech, but not in data
- Some are students or recent graduates
- Some are neither

# What's the crowd?

- Some are in tech, but not in data
- Some are students or recent graduates
- Some are neither
- Those few already in the data field. *What are you doing here?*

**Let the data science venn diagram madness  
begin!**

# The Original. Mokay. Data Science != Data Scientist

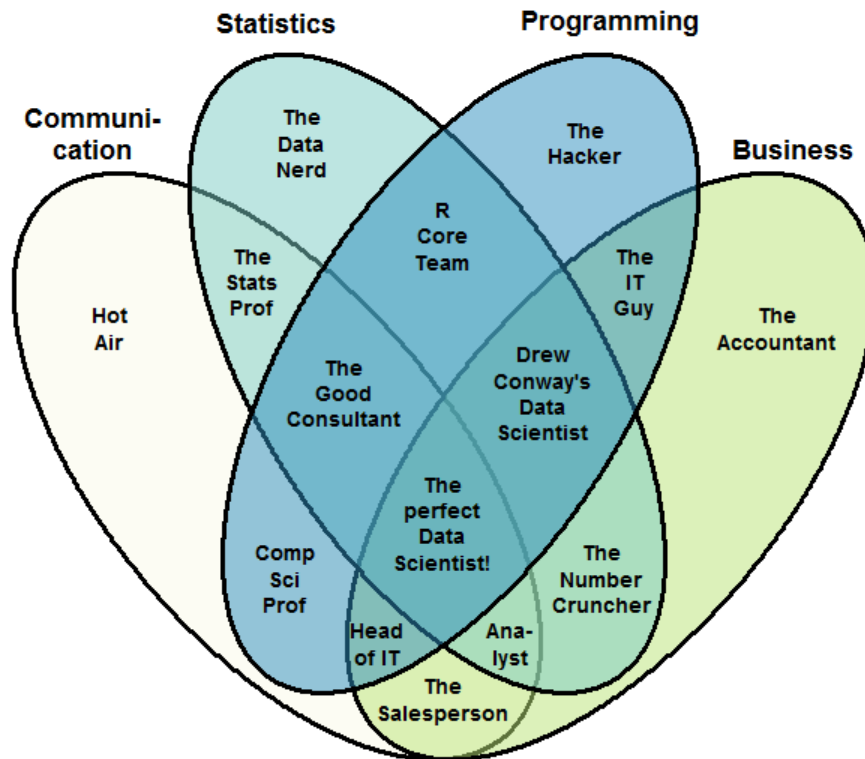


<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>



# Hmmm... Interesting puzzle

**The Data Scientist Venn Diagram**

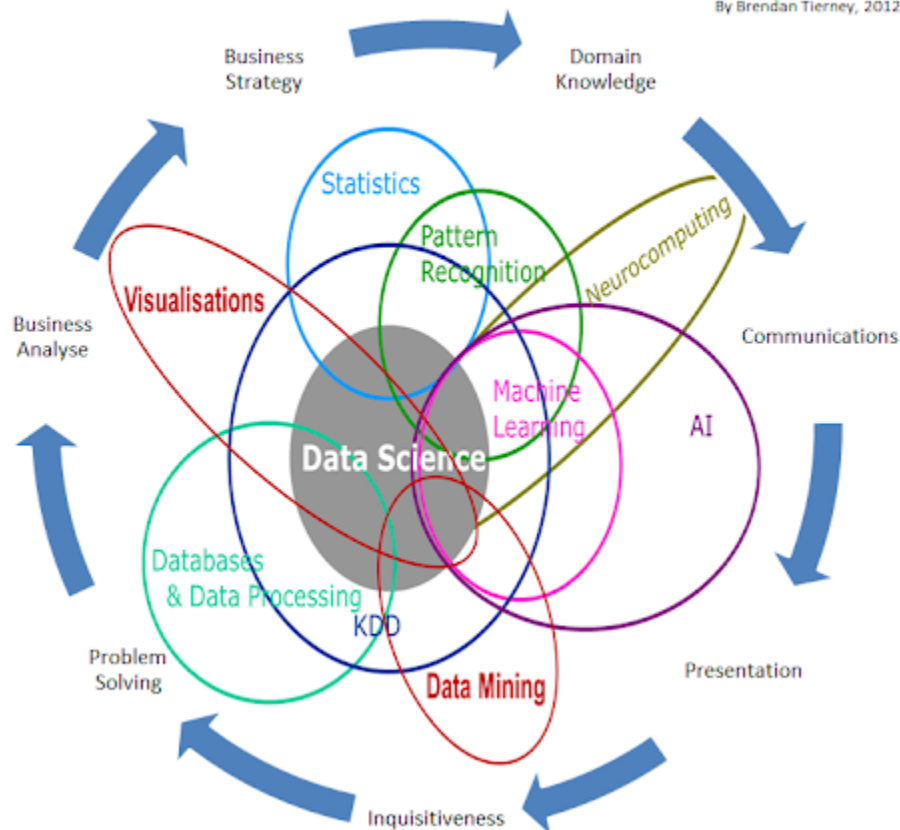


<https://whatsthebigdata.com/2016/07/08/the-new-data-scientist-venn-diagram/>

# Is this a Jumanji map?

## Data Science Is Multidisciplinary

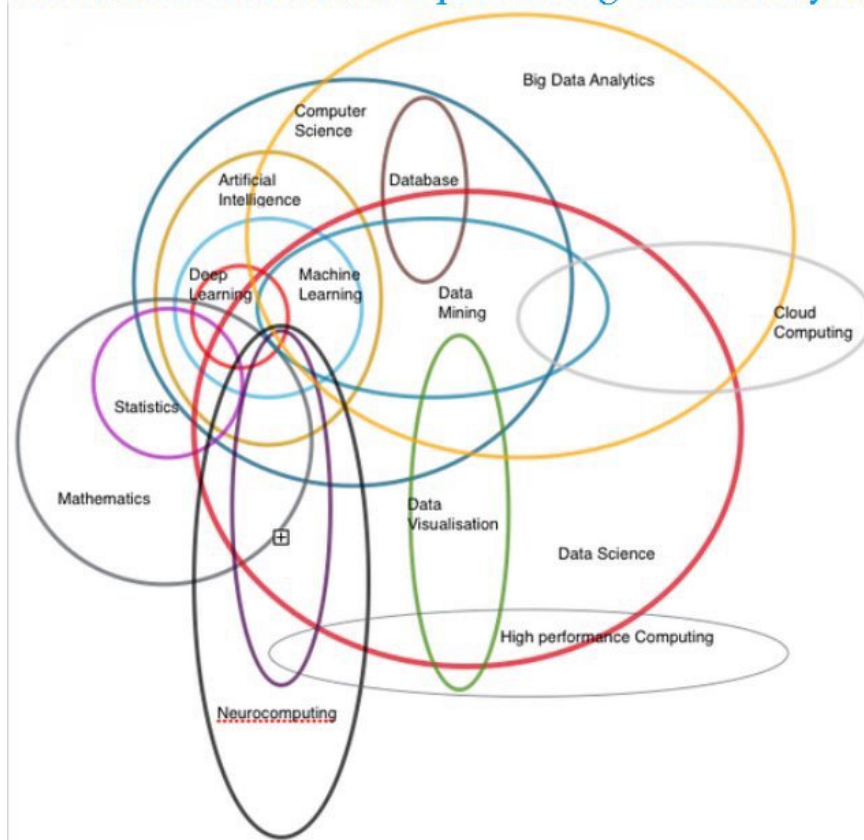
By Brendan Tierney, 2012



<http://www.oralitics.com/2012/06/data-science-is-multidisciplinary.html>

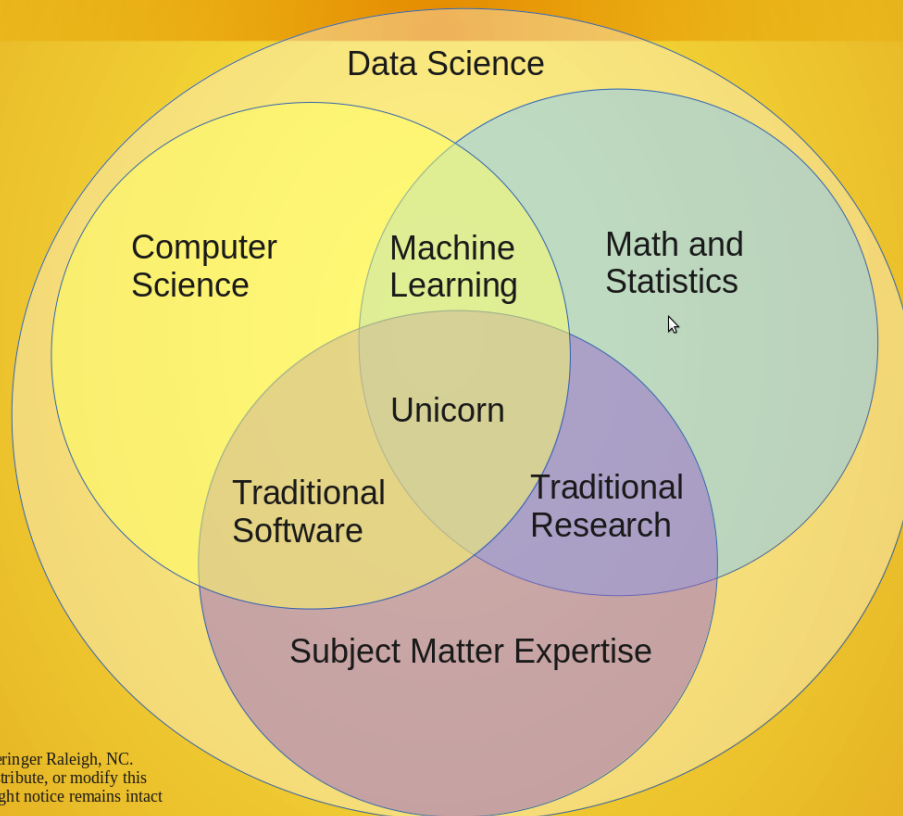
# Looks like my son's pre-school art, but worse

## Relation between techniques of Big Data Analytics



# Of course, the unicorns!

## Data Science Venn Diagram v2.0



Copyright © 2014 by Steven Geringer Raleigh, NC.  
Permission is granted to use, distribute, or modify this  
image, provided that this copyright notice remains intact

# Speaking of unicorns...



<https://imgur.com/gallery/T4Wy0>

# Please stop



# Common language



**Hadley Wickham** 

@hadleywickham



Replying to @kierisi

Also lots more jobs do data science than are called data scientists

8:06 AM - Sep 28, 2017



46



See Hadley Wickham's other Tweets



# Common language



**Hadley Wickham**

@hadleywickham



Replying to @kierisi

Also lots more jobs do data science than are called data scientists

8:06 AM - Sep 28, 2017

46 See Hadley Wickham's other Tweets



**Hadley Wickham**

@hadleywickham



Replying to @kierisi @AaronJTrent

Personally I think the difference is programming - if you're programming an analysis you're doing data science

8:21 AM - Sep 28, 2017

19 See Hadley Wickham's other Tweets





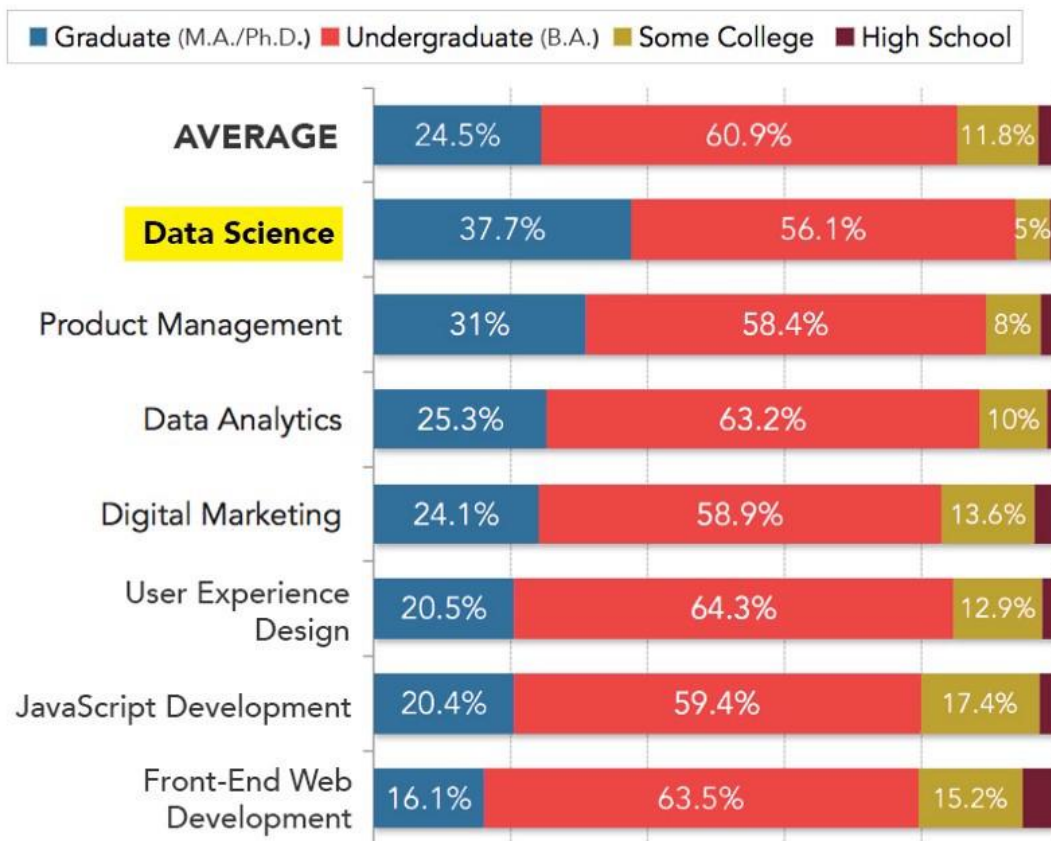
# Data science career checklist

- Get a PhD
- Wait for job offers to pour in

# Data science career checklist

- ~~Get a PhD~~
- ~~Wait for job offers to pour in~~

## Education Level of General Assembly Students by Course



Source: General Assembly part-time student data (09/2016-01/2017)

\*Average = the courses listed above

~ How Diverse is Data Science? <https://priceconomics.com/how-diverse-is-data-science/>



ദാമുകുട്ടൻ  
@damukkuttan



Replying to @kierisi @priceconomics

it is very intimidating when all (most) of the openings for  
datascience jobs has phd as a requirement

5:54 AM - Sep 28, 2017



2



See ദാമുകുട്ടൻ's other Tweets





**ദാമു(കുട്ടൻ)**  
@damukkuttan



Replying to @kierisi @priceconomics

it is very intimidating when all (most) of the openings for datascience jobs has phd as a requirement

5:54 AM - Sep 28, 2017



2



See ദാമു(കുട്ടൻ)'s other Tweets



**Jesse Maegan**  
@kierisi



Agreed - although I think there's still room for people to enter data science (via analytics) by showing they have the skills necessary. [twitter.com/damukkuttan/st...](https://twitter.com/damukkuttan/status/914141414141414141)

6:00 AM - Sep 28, 2017



14



See Jesse Maegan's other Tweets





**ദാമു(ക്കുട്ടൻ)**  
@damukkuttan



Replying to @kierisi @priceconomics

it is very intimidating when all (most) of the openings for datascience jobs has phd as a requirement

5:54 AM - Sep 28, 2017

♡ 2 👤 See ദാമു(ക്കുട്ടൻ)'s other Tweets



**Jesse Maegan**  
@kierisi



Agreed - although I think there's still room for people to enter data science (via analytics) by showing they have the skills necessary. [twitter.com/damukkuttan/st...](https://twitter.com/damukkuttan/status/914111111111111111)

6:00 AM - Sep 28, 2017

♡ 14 👤 See Jesse Maegan's other Tweets



**Hadley Wickham** ✓  
@hadleywickham



Replying to @kierisi

Also lots more jobs do data science than are called data scientists

8:06 AM - Sep 28, 2017

♡ 46 👤 See Hadley Wickham's other Tweets



# ...Yet some folks are chill though...



**Mikhail Popov**

@bearloga



Bewildered by how many software engineers applied for the data scientist role who have no experience or training in analyzing data.

8:08 PM - Jan 30, 2018



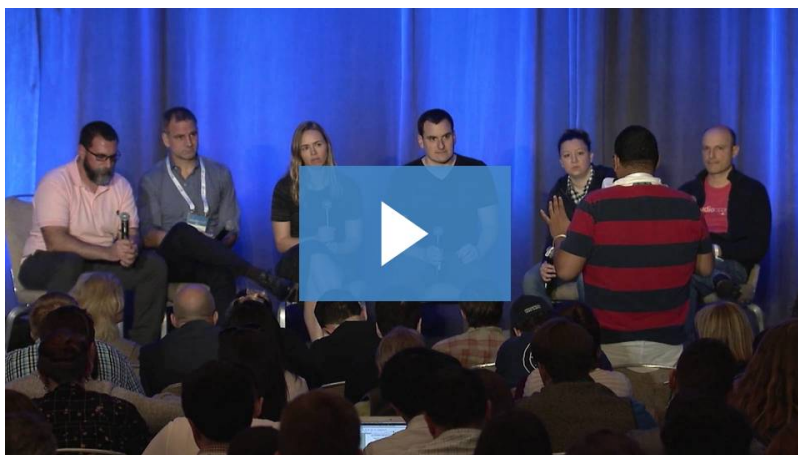
16



See Mikhail Popov's other Tweets



# Roll the tape



Fireside Chat with RStudio – R in industry discussion – RStudio (32:46)

If you do not have a PhD in Statistics and you want to be a data scientist, you need to give me the evidence that you can do data science. ~ Eduardo Ariño de la Rubia

What do you look for when hiring an entry-level data scientist? Would a master's in Data Science or a bootcamp be beneficial? ~Quora



# Unconditional advice

# Unconditional advice

- Start a blog (maintain a portfolio)
- Build, use, and give back to your network
- Keep learning

# Start a blog

# Start a blog

- Practice analyzing data and communicating about it
- Create a portfolio of your work and skills
- Get feedback and evaluation

<http://varianceexplained.org/r/start-blog/>

# Start a blog

- Practice analyzing data and communicating about it
- Create a portfolio of your work and skills
- Get feedback and evaluation

<http://varianceexplained.org/r/start-blog/>

My promise is this: **if you're early in your career as a data scientist and you start a data-related blog, tweet me a link at @drob and I'll tweet about your first post** (in fact, the offer's good for each of your first three posts) ~ *David Robinson*

# Build, use, and give back to your network

# Build, use, and give back to your network

- Volunteer at user groups, meetups, and yes, SQL Saturdays
- Ask to be a speaker
- Learn from the best and immediately share that knowledge back
- Volunteer your skills
- Contribute to open source

# Build, use, and give back to your network

- Volunteer at user groups, meetups, and yes, SQL Saturdays
- Ask to be a speaker
- Learn from the best and immediately share that knowledge back
- Volunteer your skills
- Contribute to open source



**Jesse Maegan**

@kierisi



want to build data science experience? reach out to a local non-profit you're interested in, and ask them if you can volunteer with data collection, cleaning, and basic analysis and reporting. you get experience, the NPO gets a product they desperately need, and everyone wins.

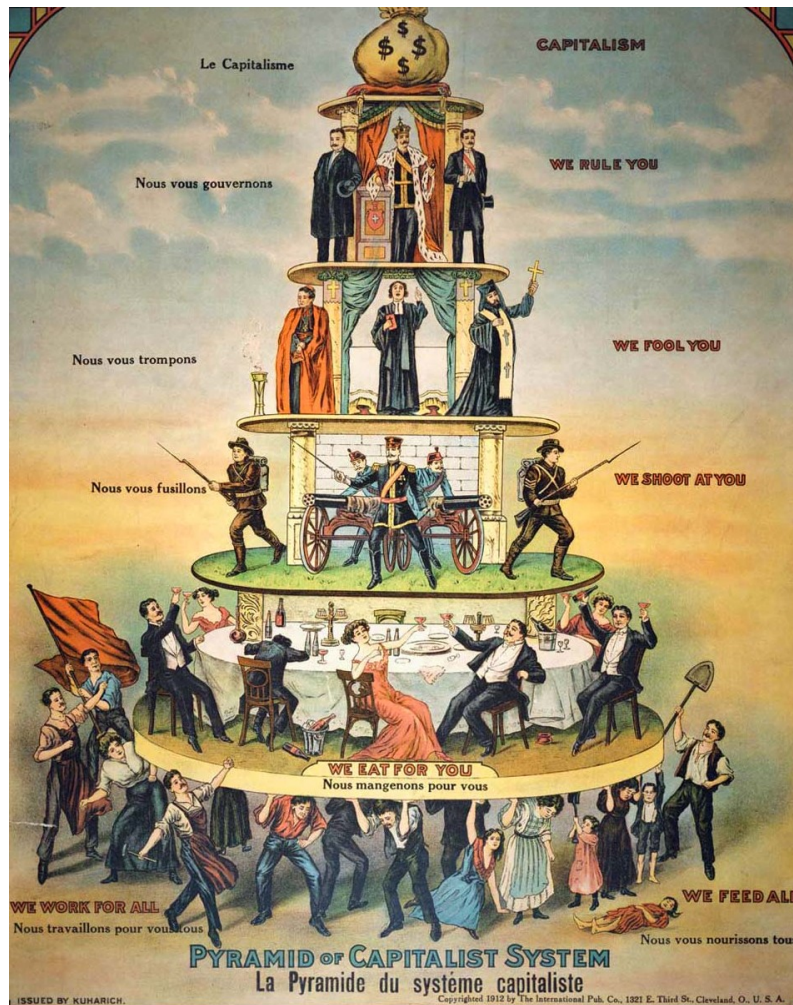
1:27 PM - Mar 30, 2018

♡ 707 💬 221 people are talking about this





# This is not a pyramid scheme



# Keep learning

# Keep learning

- MOOCs, online classes and learning platforms
- Traditional education
- Immersion

# Find your niche in the data science food chain

# Find your niche in the data science food chain

- Teach what you've learned
  - through blogging
  - through speaking
  - through message boards and Stack Overflow
- Be an aggregator and a connector
- Blog on a specific topic (e.g. politics, basketball, music)
- Contribute to open-source (even if it's correcting typos)
- Be a computer scientist for data scientists.

# Some more specific advice

# Some more specific advice

- If you feel you're lacking some formal education - consider going back to school

# What happens if the explanatory and response variables are sorted independently before regression?

Suppose we have data set  $(X_i, Y_i)$  with  $n$  points. We want to perform a linear regression, but first we sort the  $X_i$  values and the  $Y_i$  values independently of each other, forming data set  $(X_i, Y_j)$ . Is there any meaningful interpretation of the regression on the new data set? Does this have a name?

I imagine this is a silly question so I apologize, I'm not formally trained in statistics. In my mind this completely destroys our data and the regression is meaningless. But my manager says he gets "better regressions most of the time" when he does this (here "better" means more predictive). I have a feeling he is deceiving himself.

<https://stats.stackexchange.com/questions/185507/what-happens-if-the-explanatory-and-response-variables-are-sorted-independently>



# Some more specific advice

- If you feel you're lacking some formal education - consider going back to school

# Some more specific advice

- If you feel you're lacking some formal education - consider going back to school
- If you're a recent graduate, consider internships

From the author *"If it's not a hell yeah - it's a no"...*

When you're earlier in your career I think the best strategy is you just say yes to everything, every piddly little gig, you just never know what are the lottery tickets. ~ **Derek Sivers**

# Some more specific advice

- If you feel you're lacking some formal education - consider going back to school
- If you're a recent graduate, consider internships

# Some more specific advice

- If you feel you're lacking some formal education - consider going back to school
- If you're a recent graduate, consider internships
- If you're a working professional, see if you can pivot within your organization.
  - business analyst
  - mentor / sponsor
  - cross-training
  - improve your work with data

# Example



**Caitlin Hudon**



@beeonaposty



Replying to @taraskaduk @drob

The last three jobs I've held have been interested in my combination of data / statistics and communication -- as judged by blogging and presentations 😊

4:54 PM - Apr 30, 2018



8



See Caitlin Hudon's other Tweets



# Example



**Nick Strayer**

@NicholasStrayer



Replying to @LucyStats and 2 others

And I got an internship at the new york times due to my website/github presence. No formal journalism education.

5:34 PM - Apr 30, 2018



9



See Nick Strayer's other Tweets



# Example



**John Coene**

@jdatap



Replying to @taraskaduk @drob

I have no formal education in data science/analytics, it's all been down to GitHub and website.

5:58 PM - Apr 30, 2018



3



See John Coene's other Tweets





# Example



**Samantha Sifleet**

@SamanthaSifleet



Replying to @taraskaduk and 2 others

I landed this job with my publication in Ecological Economics + bitbucket repo for the python module, 2 shiny Apps I built and the ability to successfully whiteboard in an interview.

2:38 PM - Apr 30, 2018



2



See Samantha Sifleet's other Tweets



# Example



**Hamed**

@HamedBH



Replying to @taraskaduk @drob

I did it: combo of MOOCs, other independent learning, and turning the bits/examples of work I'd done into something formal enough to pass an interview. Happy to give more details if you want to DM.

2:42 PM - Apr 30, 2018



1



See Hamed's other Tweets



# Example



**jeff benzos**

@jeff\_\_benzos



Replying to @taraskaduk and 2 others

i got my first data science job through blogging on my company's internal social platform which led to tech talks and eventually the job.

12:59 AM - May 1, 2018



2



See jeff benzos's other Tweets



# The end

## Follow me:

- <https://taraskaduk.com>
- <https://www.linkedin.com/in/taraskaduk/>
- <https://github.com/taraskaduk/>
- <https://twitter.com/taraskaduk>

## This presentation is here

[insert link here](#)