But I'm not qualified to be a data scientist!

How to transition from university to the data science industry

Erika Braithwaite

2020-01-22

Welcome!

So you're thinking about leaving the comforting womb of university, and transitioning to a data science career? You may be feeling….

- Overwhelmed by the possibilies
- Confused about the terminology
- Insecure about your capacities
- And so many more emotions

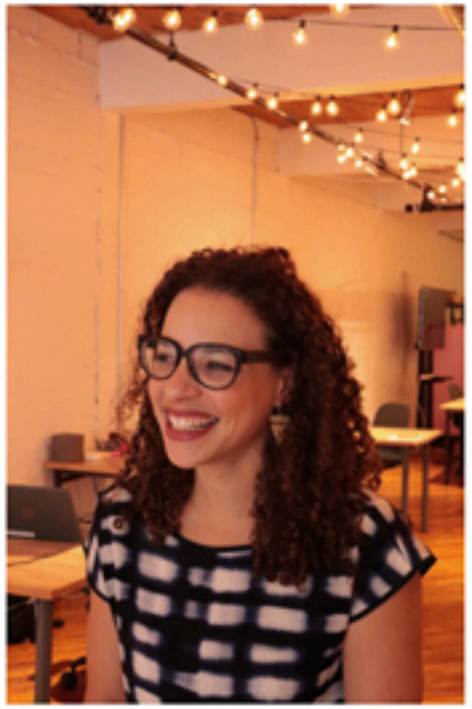
DATA SCIENCE
SO HOT RIGHT NOW

Objectives

I'll do my best to address these objectives in the most data-driven approach I can!

- Dispel some myths about the data science industry
- Tips on how to transition
- Let's talk about your CV, the job search and the interview
- Bestow some glimmer of hope ⭐

A bit about me



I'm the CEO of a health data science startup in Montreal www.precision-analytics.ca

My background

- BA Psychology, Concordia
- MSc Psychology, Université de Montréal
- PhD Epidemiology and Biostatistics, McGill

Our company

- Founded in 2017
- Began as a statistical consultancy, evolved to a blend of Software as a Service
- Work in the biopharmaceutical sector to create tools that combine data storage, analysis and communication

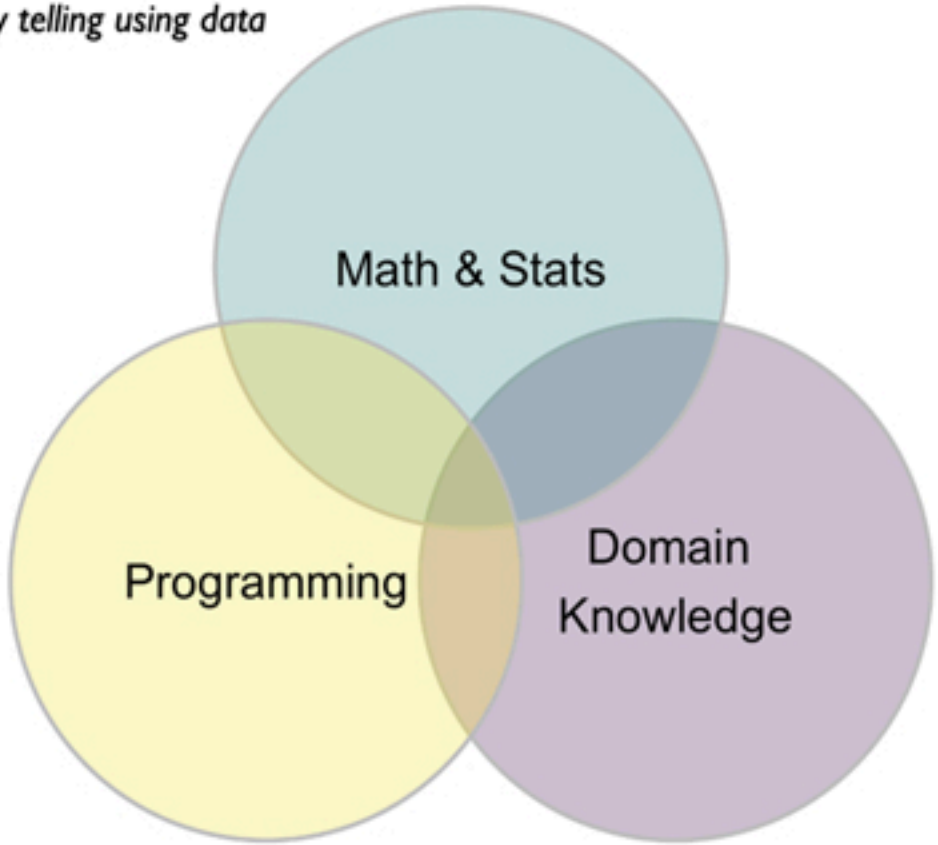
Precision Analytics

Our story

What is a data scientist

A combination of programming, statistics and domain knowledge.

I like to think of it as *story telling using data*



A very scary (and real) job posting

- Coding knowledge and experience with several languages: C, C++, Java, JavaScript, etc.
- Knowledge and experience in statistical and data mining techniques: GLM/Regression, Random Forest, Boosting, Trees, text mining, social network analysis, etc.
- Experience querying databases and using statistical computer languages: R, Python, SLQ, etc.
- Experience using web services: Redshift, S3, Spark, Digital Ocean, etc.
- Experience creating and using advanced machine learning algorithms and statistics: regression, simulation, scenario analysis, modeling, clustering, decision trees, neural networks, etc.
- Experience analyzing data from 3rd party providers: Google Analytics, Site Catalyst, Coremetrics, Adwords, Crimson Hexagon, Facebook Insights, etc.
- Experience with distributed data/computing tools: Map/Reduce, Hadoop, Hive, Spark, Gurobi, MySQL, etc.
- Experience visualizing/presenting data for stakeholders using: Periscope, Business Objects, D3, ggplot, etc.

Myths about data science

- Myth 1. You need a PhD to become a data scientist
- Myth 2. Python is the only tool used in data science
- Myth 3. Data science == AI & machine learning
- Myth 4. Data science is all about technical skills

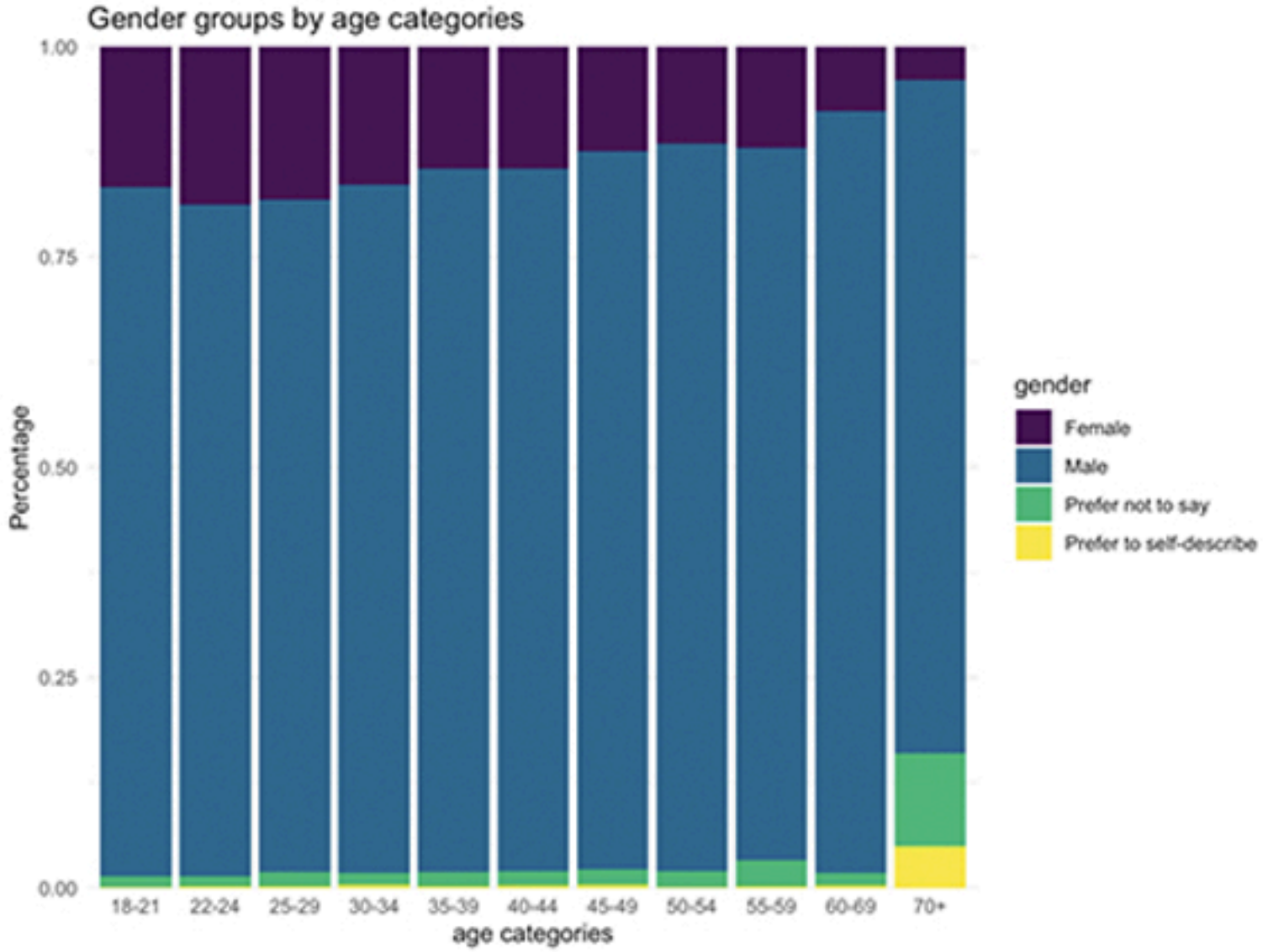
Letting the data speak!

To tackle these myths, Kaggle conducts annual survey of data scientists. Almost 20,000 respondents completed the survey from over 60 countries in 2018 and 2019.

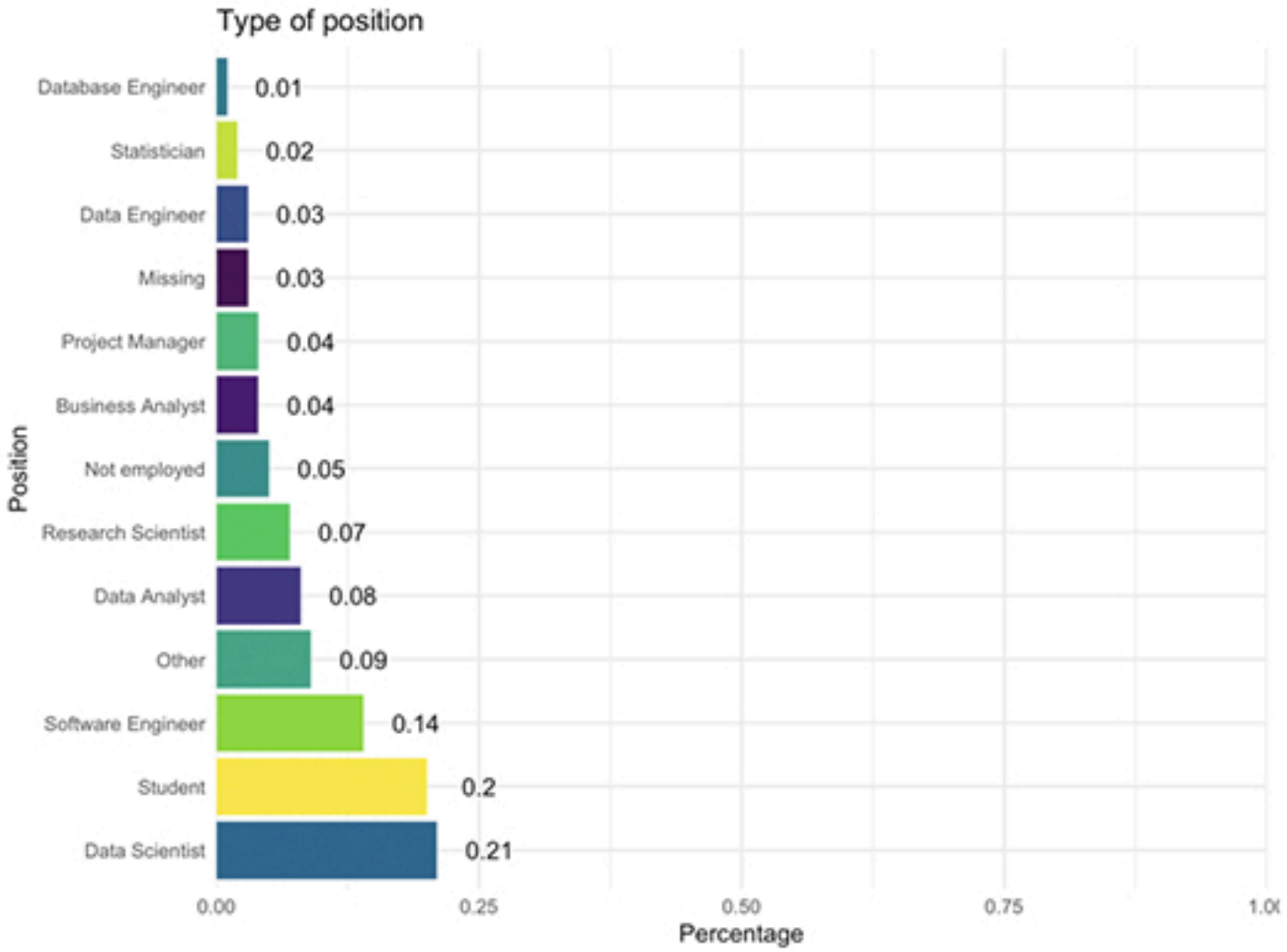
!! Kaggle is a platform that hosts machine learning competitions. It's users are not representative of the entire data science community !!

Data available for 2018 and 2019

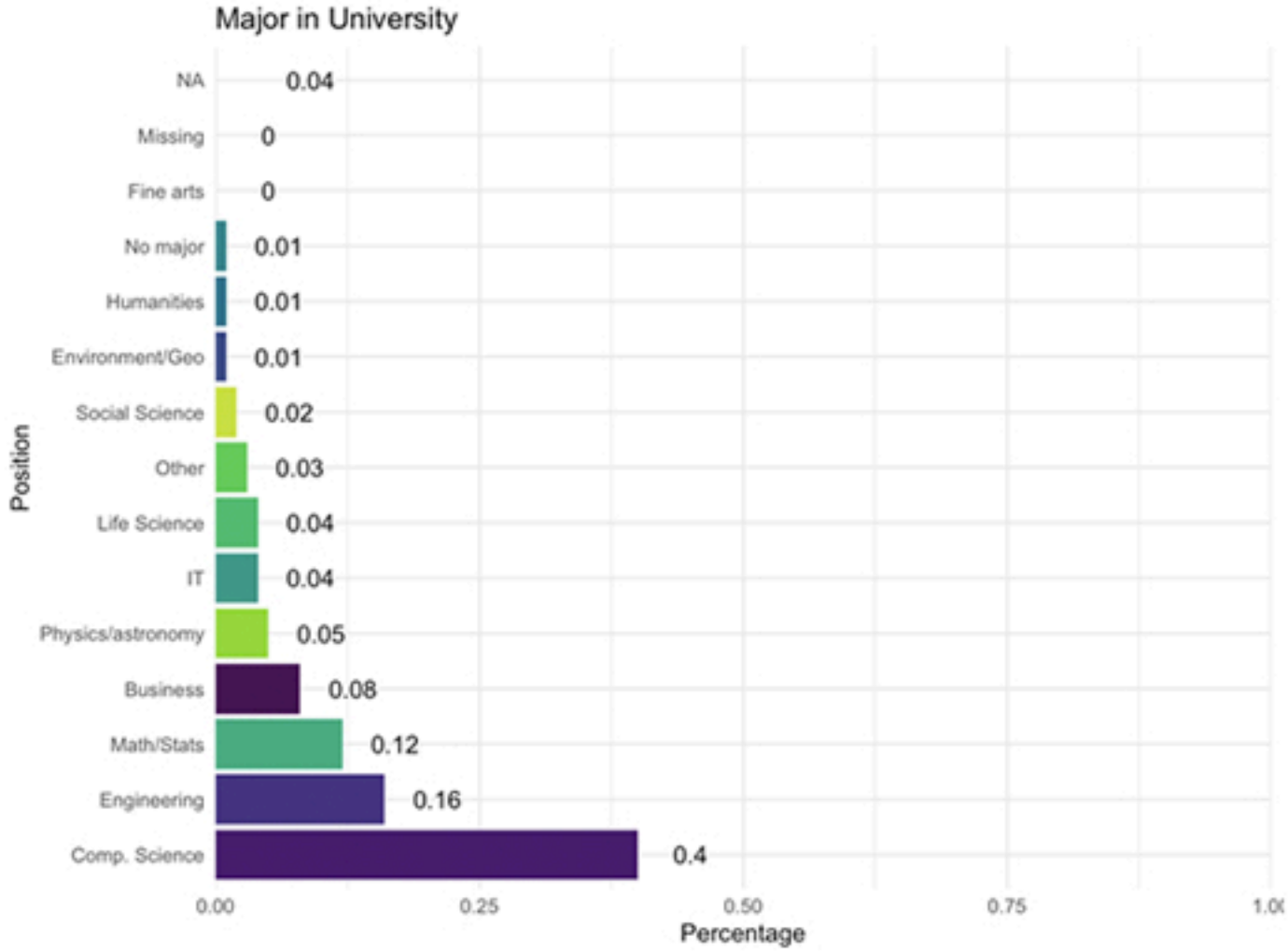
Demographics of respondents



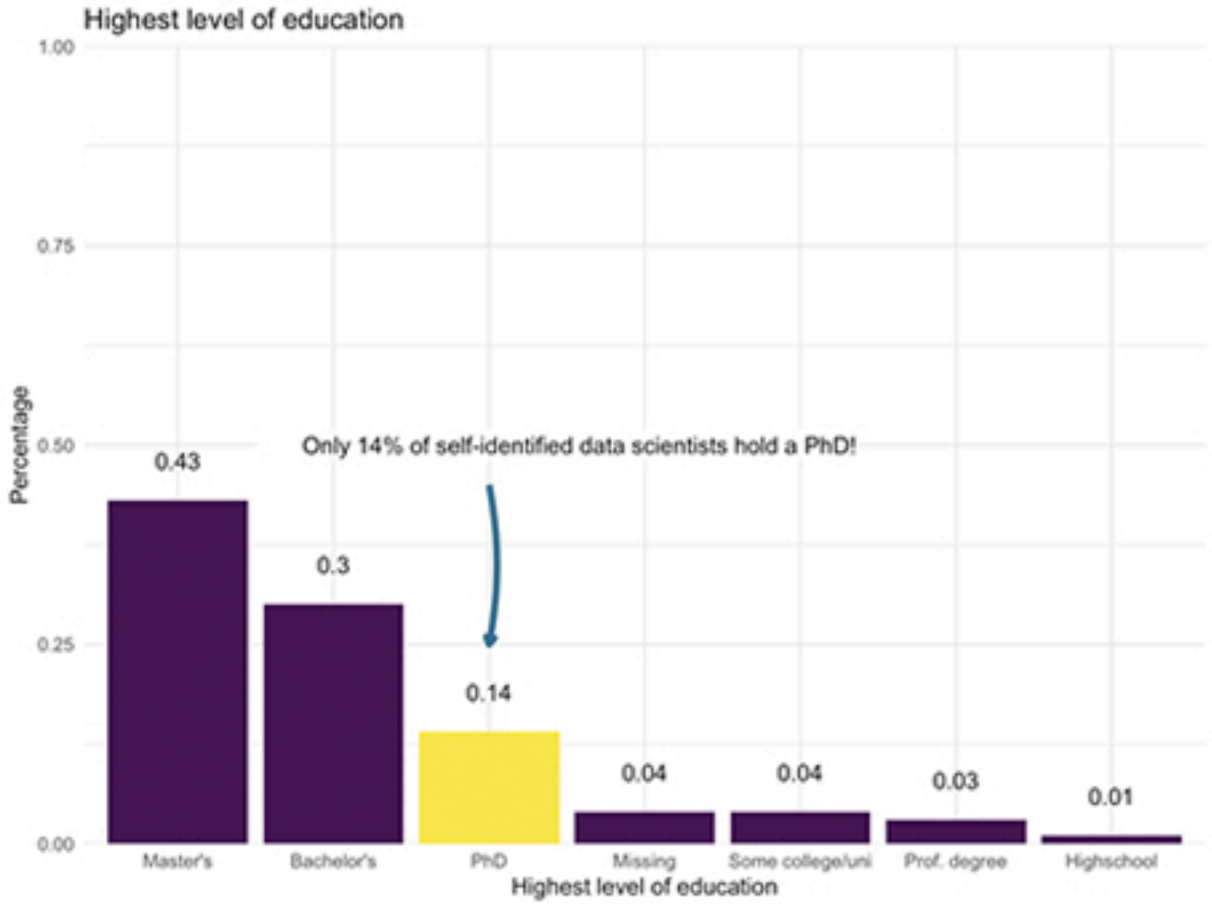
Types of data scientist positions



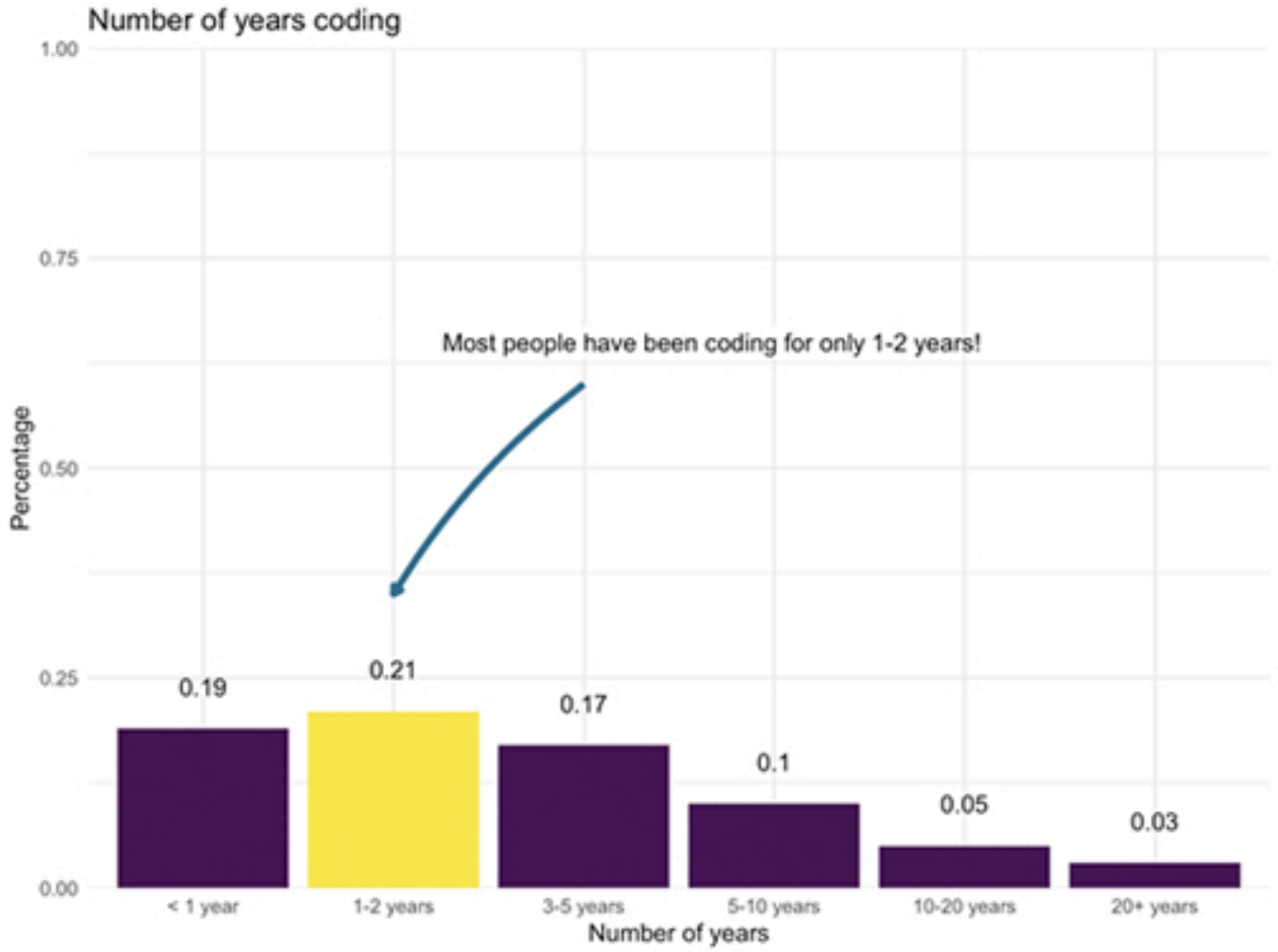
Academic background



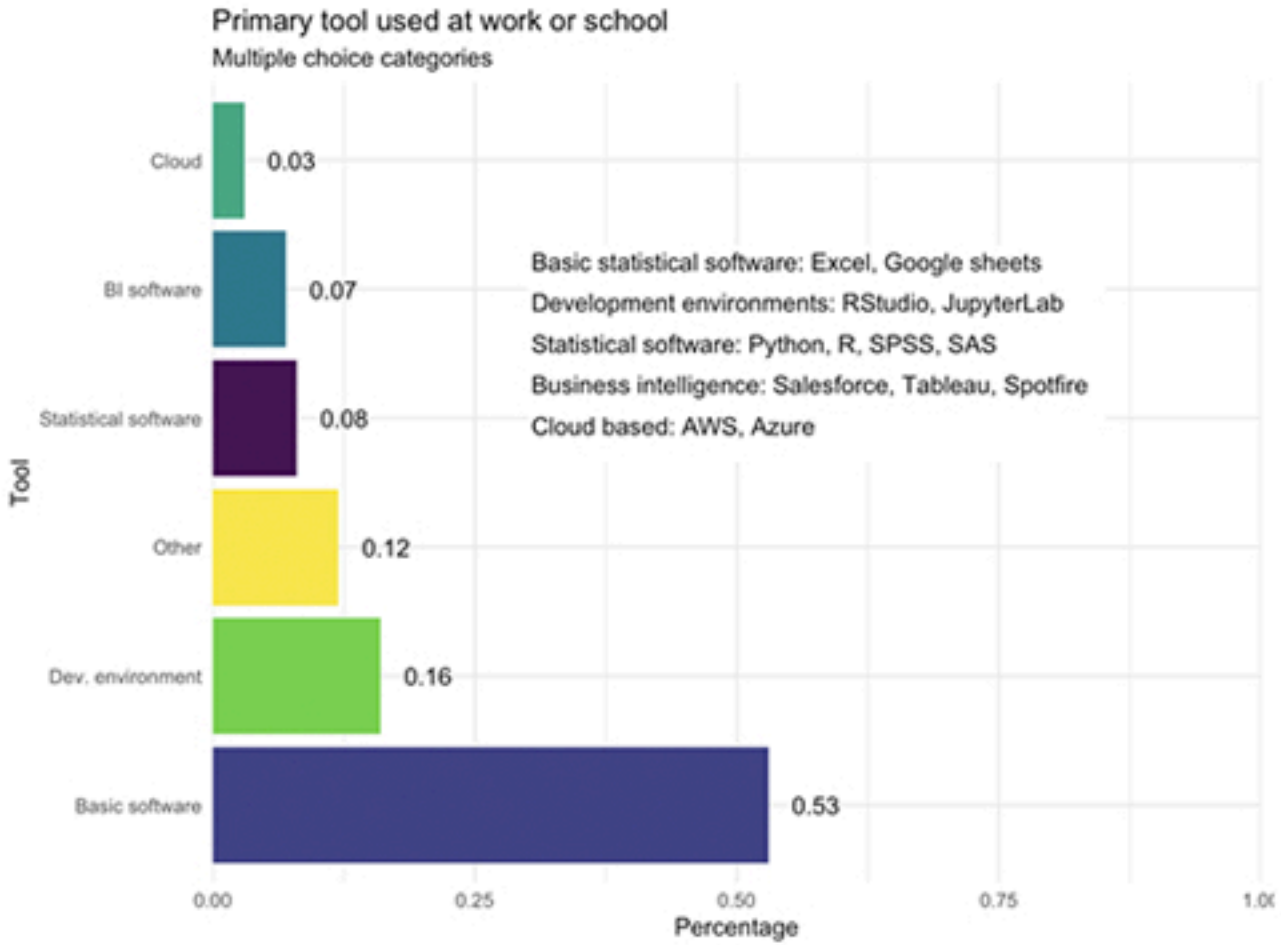
Myth 1. You need a PhD to become a data scientist



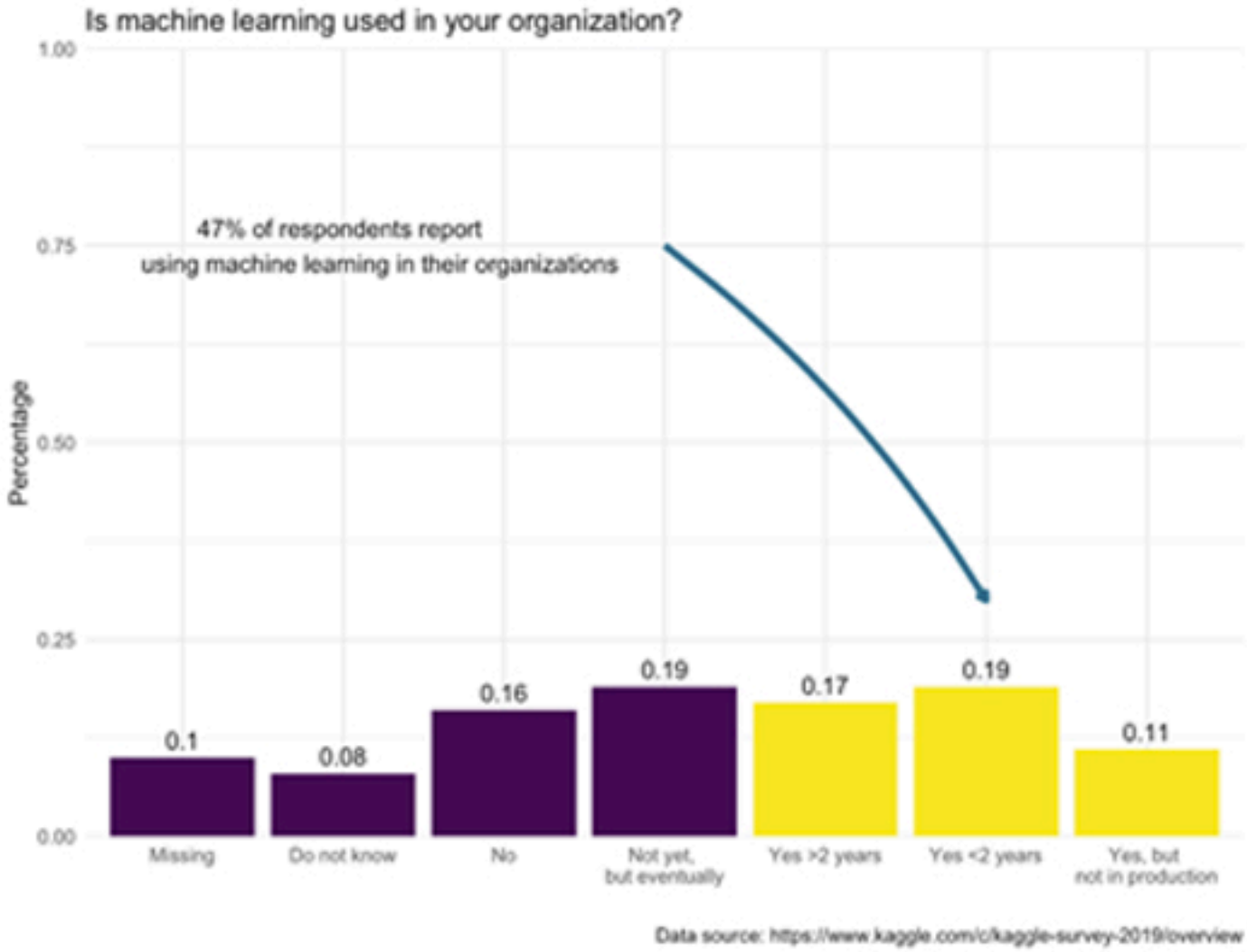
Experience of respondents



Myth 2. Python is the only tool used in data science

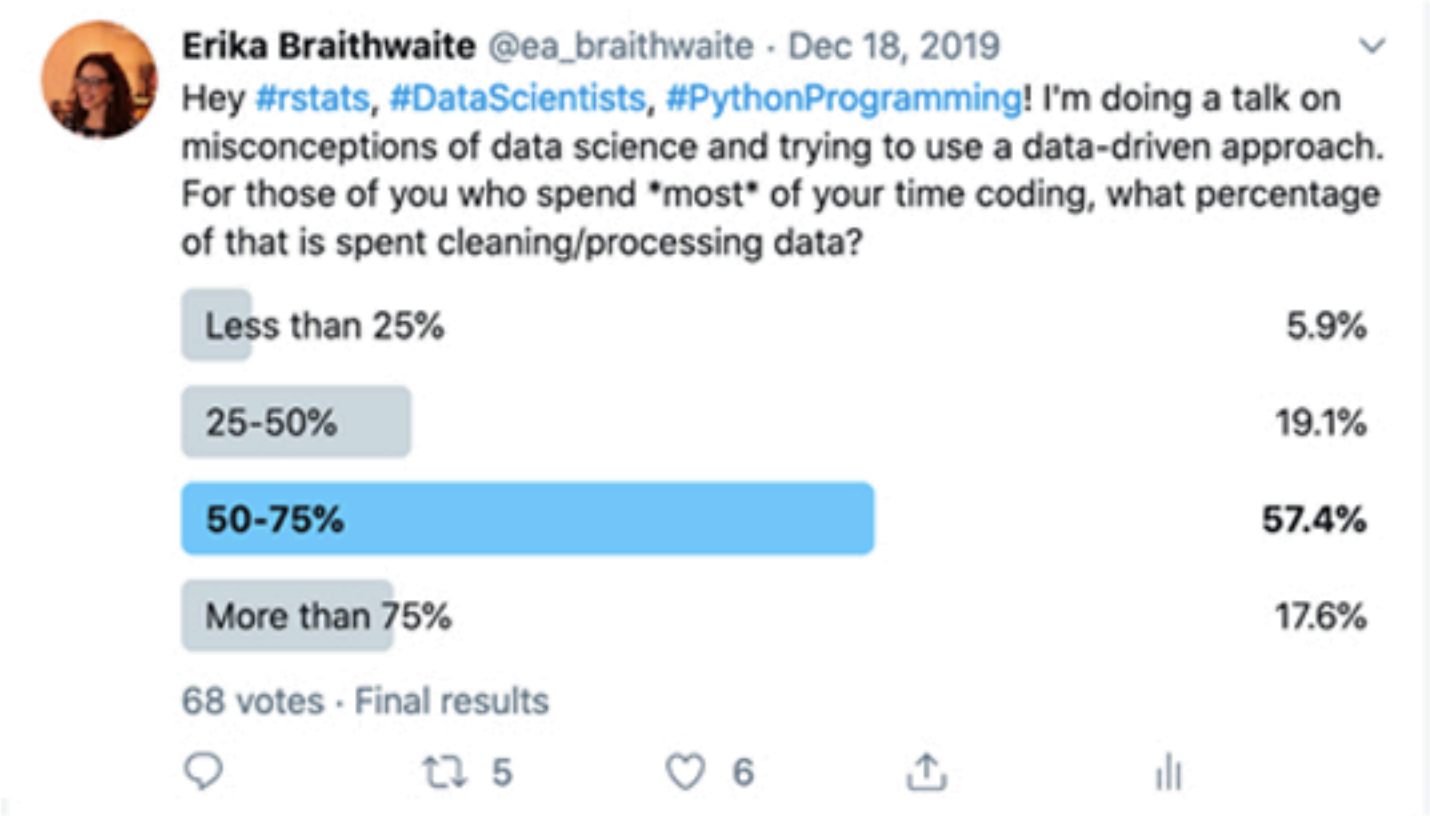


Myth 3. Data science == Machine learning and AI



Myth 4. Data science challenges are mostly analytical

The survey did not explicitly ask the way respondents broke up their time. So I did the next best thing… ask twitter

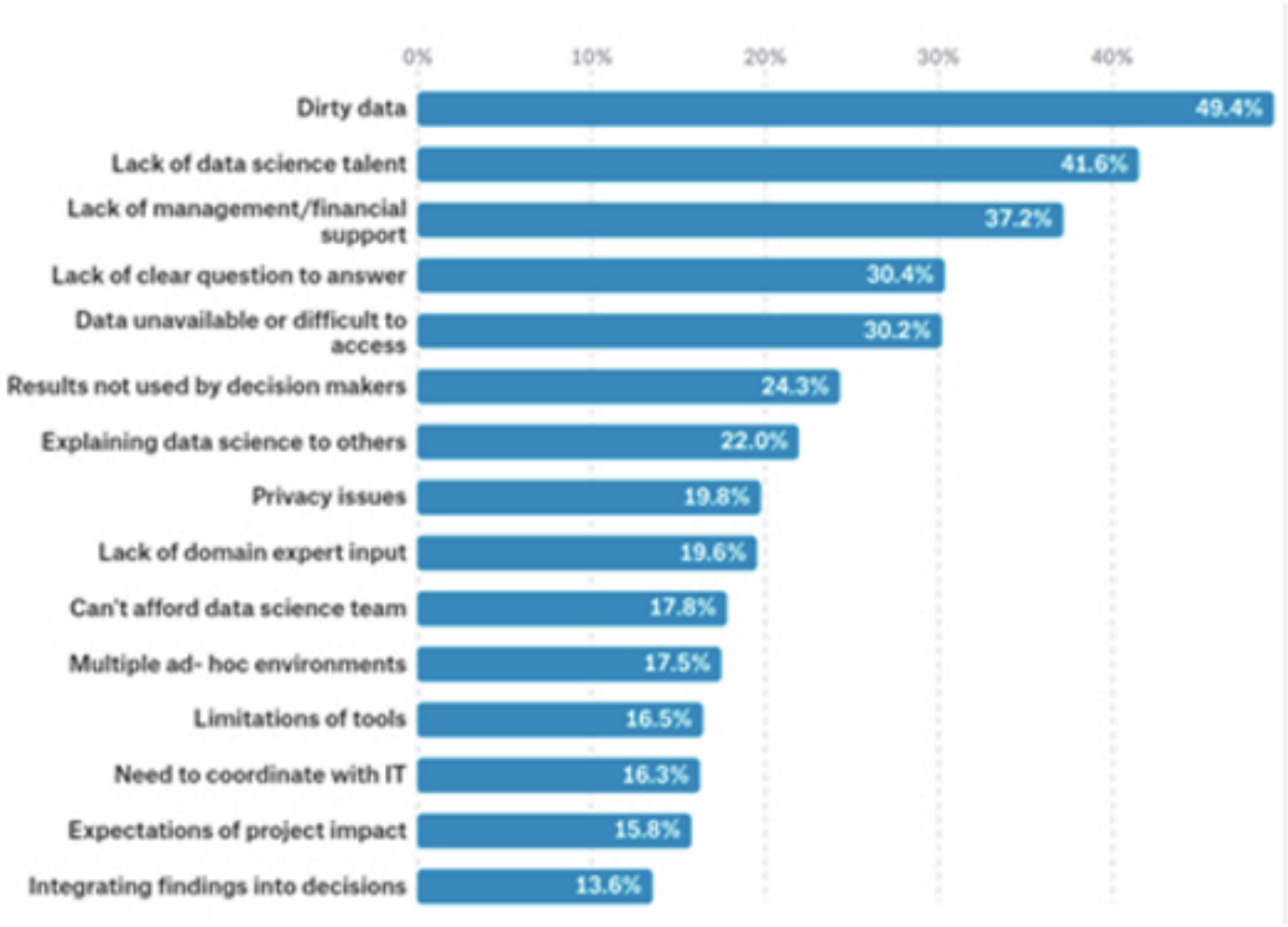


Data scientists challenges

In 2017, the Kaggle survey asked respondents about the "biggest challenges in data science".

From here we can see that most people identify "dirty data" the toughest part of the job.

The rest of the issues seem be be organizational.

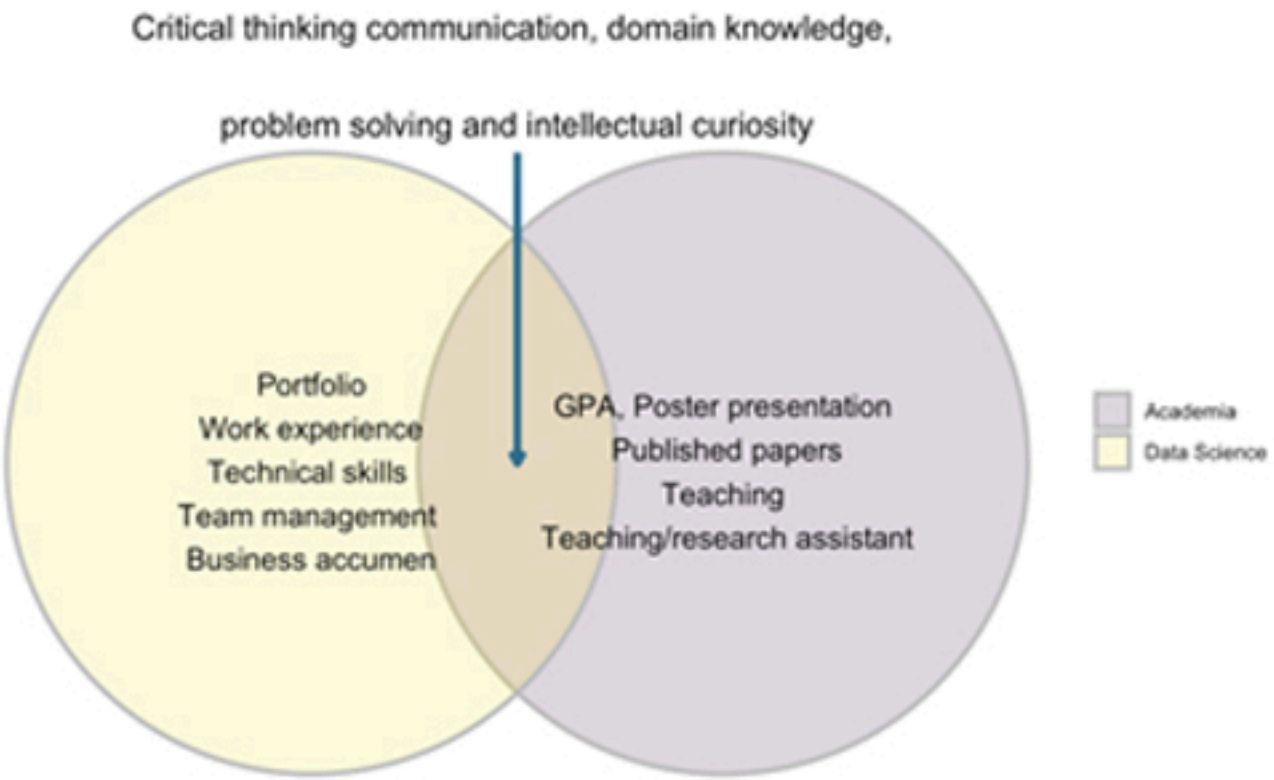


On transitioning

A JOB IN
DATA SCIENCE
ME
LET ME IN
[adult swim]
ME

Understand your new “audience”

- 💰 The currency in academia versus the currency in the private sector 💰



Gaining more experience: Courses

The American Statistical Association held a two-day Data Science summit. 72 educators, researchers and practitioners in statistics, mathematics, computer science, and data dcience from academia, industry, government and nonprofit gathered to put forth recommendations for future data science programs.

Data science courses should expose students to:

Introduction to statistics
Data analysis in the real world
Math and algorithms
Answering real problems
Expose students to modern tools
Teach data ethics
Active learning

I'd add to this list: study design. Knowing how data was generated/collected will always inform the types of questions you can ask, and appropriate method of analysis.

How to pick a learning strategy

Type	Pros	Cons
MOOC	Low cost and commitment	Self-guided learning may be less efficient than other methods
	Good for beginners	Lack credibility
Bootcamp	Short and intense	Can be expensive
	Connect to employers	Not guaranteed to be recognized by employer
Certificate	Supplement existing degree	Very general
	Shorter and less expensive than a degree	
MSc	Gain research experience & skills	Siloed depts
	Leverage University name and degree recognition	Slow to update curriculums

Note:

Source: Data Science Careers, Training, and Hiring: A Comprehensive Guide to the Data Ecosystem:How to Build a Successful Data Science Career, Program, or Unit. (2019) Rawlings-Goss, R. In Springer Briefs in Computer Science. https://doi.org/10.1007/978-3-030-22407- 3

Gaining more experience: MOOCs, bootcamps and certificates

MOOC	Bootcamps	Certificates
Data Camp	Brainstation (Toronto, ON)	McGill's Certificate in Data Science and Business Analytics
Udemy	General Assembly (Toronto and worldwide)	Concordia's Diploma in Big Data
Coursera	WeCloudData	Ryerson University Certificate in Data Analytics
MIT edX	Dataquest	York University Certificate in Big Data Analytics
Metis	NYC Data Science Academy	University of Toronto Data Science Certificate

Masters
HEC (UdeM) Data Science
University of British Columbia Data Science
Waterloo Data Science and AI
University of Alberta Computer science with Statistical Machine Learning
Trent Big Data Analytics

Build a portfolio

Showcase your work

- Github
- a simple personal website
- Kaggle competitions

Internship

- Get relevant experience, learn about new industry, create a network
- Check out your university coop programs
- Internship programs in industry IVADO labs, Plotly

Don't forget to reach out to your network and tell people you're looking to get into a new industry

Preparing your cover letter and CV

- Use the cover letter to explain *why* you're applying for the job even if your profile doesn't perfectly match the job description
- Instead of focusing on all the highlights of your academic career, explain WHY you want to work for that company
- In general, keep it short and sweet (max 2 pages)
- Academic CV != Professional CV
- Only provide extensive details about academic experience if it is relevant
- Tailor your CV and cover letter to each application - use key words from posting

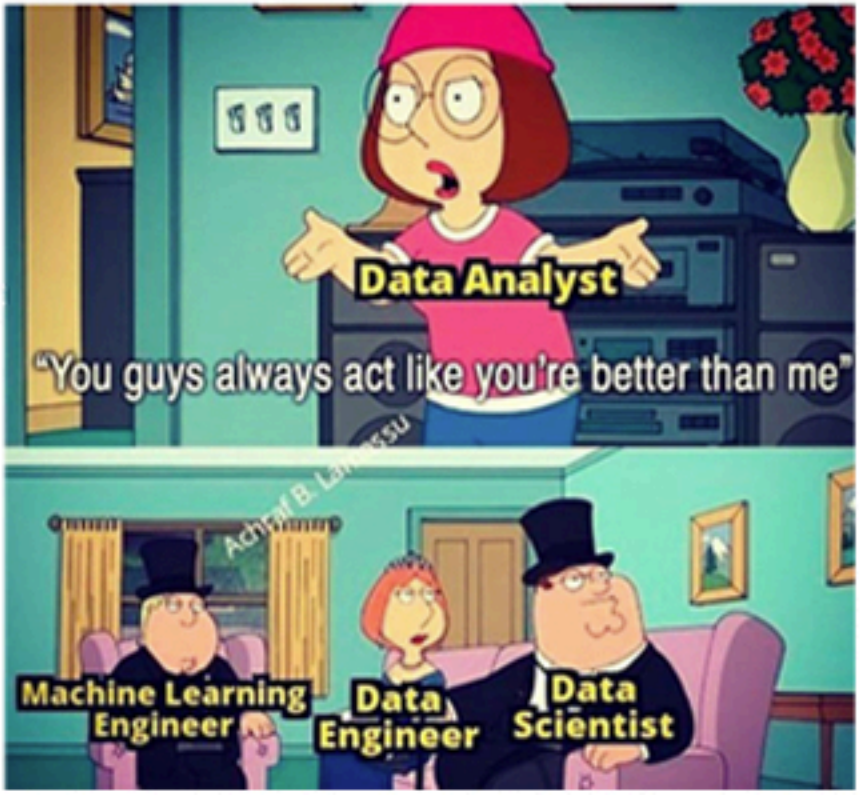
More on CV's

A few tips (based on my personal pet peeves)

- Use consistent formatting
- If using a job aggregator site, upload your actual CV (the printout we see is very ugly)
- Don't include abbreviations or technical jargon
- Don't indicate that you have high proficiency in 22 programming languages
- Don't list 30 MOOC's I *know* you didn't finish

Searching for a position

- Focus on the skills listed in the job positing, not the title *but be weary of descriptions*
- Don't expect that you will possess 100% of the stated skills; it's often a wishlist
- Try to talk to people at those companies or seasoned data scientists to get insider perspective
- Be aware that many companies are jumping on the "data" bandwagon without having any infrastructure to support a junior hire
- Many companies's job description is aspirational; often doesn't reflect the day-to-day



During the interview

- Expect technical questions and/or coding challenge
- Be honest about your skills
- Be mindful about expressing your desired to be paid to learn
- Highlight the value you bring to the employer, rather than the value you accrued in university
- Be ready to explain technical concepts to non-technical people

Jacqueline Nolis's book Build a Career in Data Science

> *"I want to hear about a project they've worked on recently. I ask them about how the project started, how they determined it was worth time and effort, their process, and their results. I also ask them about what they learned from the project. I gain a lot from answers to this question: if they can tell a narrative, how the problem related to the bigger picture, and how they tackled the hard work of doing something".*

Asking for feedback after your interview can help highlight the interviewer's perceptions of your strenghs and weaknesses

Employer fears

In general, employers are concerned that new hires without job experience will struggle to see the bigger picture and will struggle with the pace.

Undergrad & Masters	PhD
No revelant experience	Experience is hyperfocused
Will expect a lot of hand holding	Unaccustomed to working collaboratively
Training is expensive and time consuming	Both under- and over-qualified

Try to address these fears by addressing them directly in your cover letters, CV's and during the interview

Conclusion

Pivoting to data science from humanities, social sciences or any other discipling doesn't mean you need to leave all of your training behind.

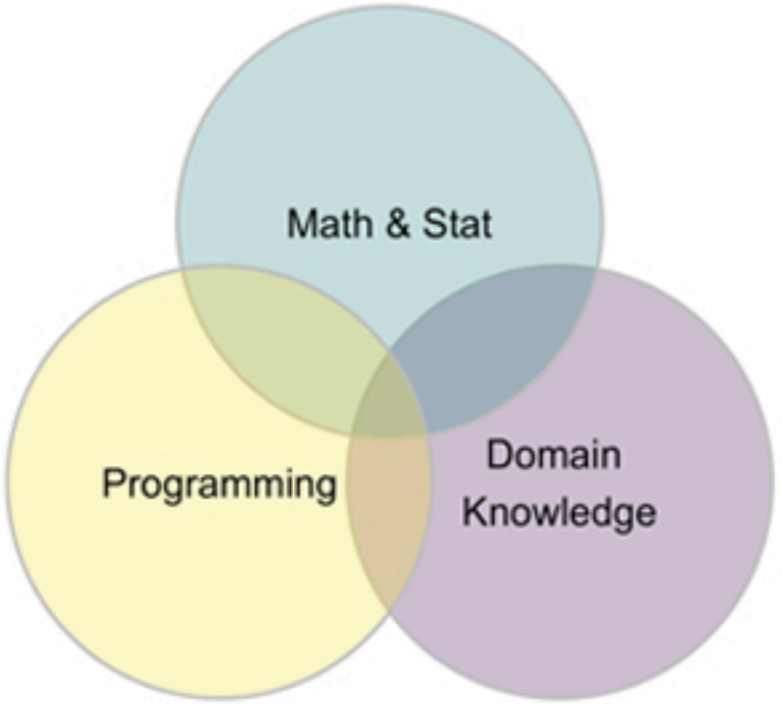
✅ You have knowledge and/or experience in at least one of these circles

✅ Look for people in your substantive area who are doing some more quantitative/tech driven

✅ Be confident that *you* know how to learn

✅ Be ready to start at the bottom

✅ Your path is the right path



A big thanks!

Please come visit us in our booth! We're happy to chat

We're looking for summer interns and other potential academic collaborators!!

A big thank you to CSCDS for inviting to me speak today and reddit.com/r/datascience for all the memes

I can be found on twitter

Come check out our website www.precision-analytics.ca

This presentation & R code can be found on Github

Happy to answer any questions

Unless it's about the R vs Python debate