



Data Science Jobs

What they're like and how to get one



Outline

- Brief intro about me + Civis
- The data science job landscape
- How to get a data science job
- Transitioning from academia to industry



My physics → data science path



Worked on graphene stuff in Nadya Mason's lab.
Decided I didn't want to stay in academia ~3 years in,
started doing Python stuff on the side.

Data Science For Social Good

Summer Fellowship



Did an internship with DSSG the summer before I
graduated, working on predicting adverse interactions
between police officers and the public.



Joined Civis after I graduated where I work on the
statistics, machine learning, and data-y parts of our
products.



- Founded in 2013, Civis Analytics is a 150-person data science technology and advisory company with offices in Chicago and Washington, D.C.
- We invented technology to operationalize data science for our consultants and our clients' internal data science teams.
- Today, Civis is the premier data science platform and provides world-renowned applied data science consulting.



FEMA

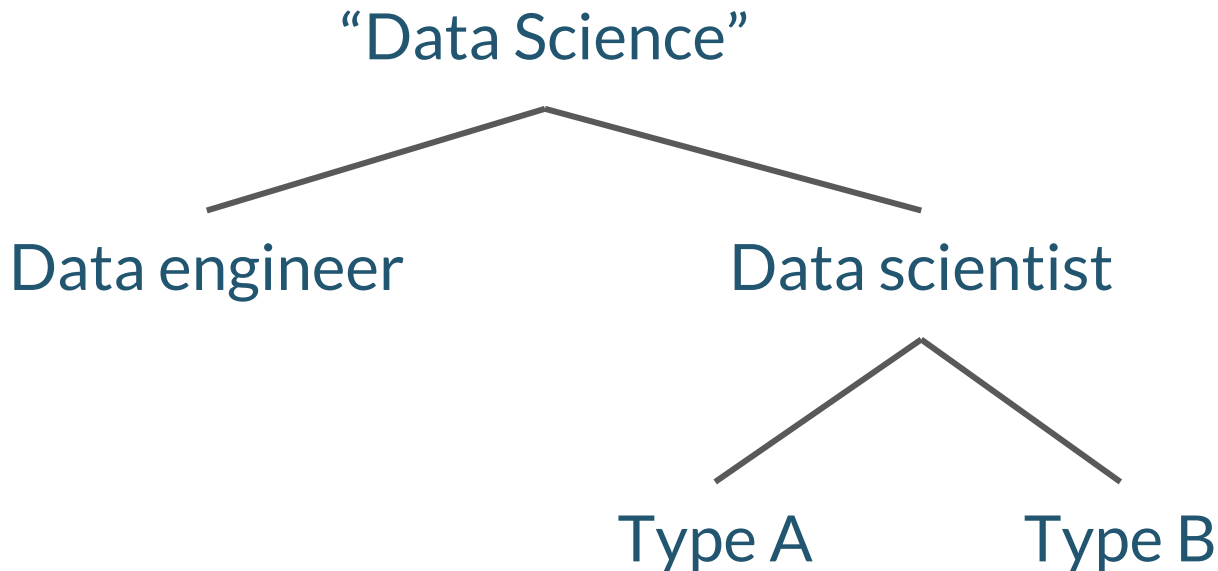


American Red Cross

The data science job landscape



The data science job landscape



source:

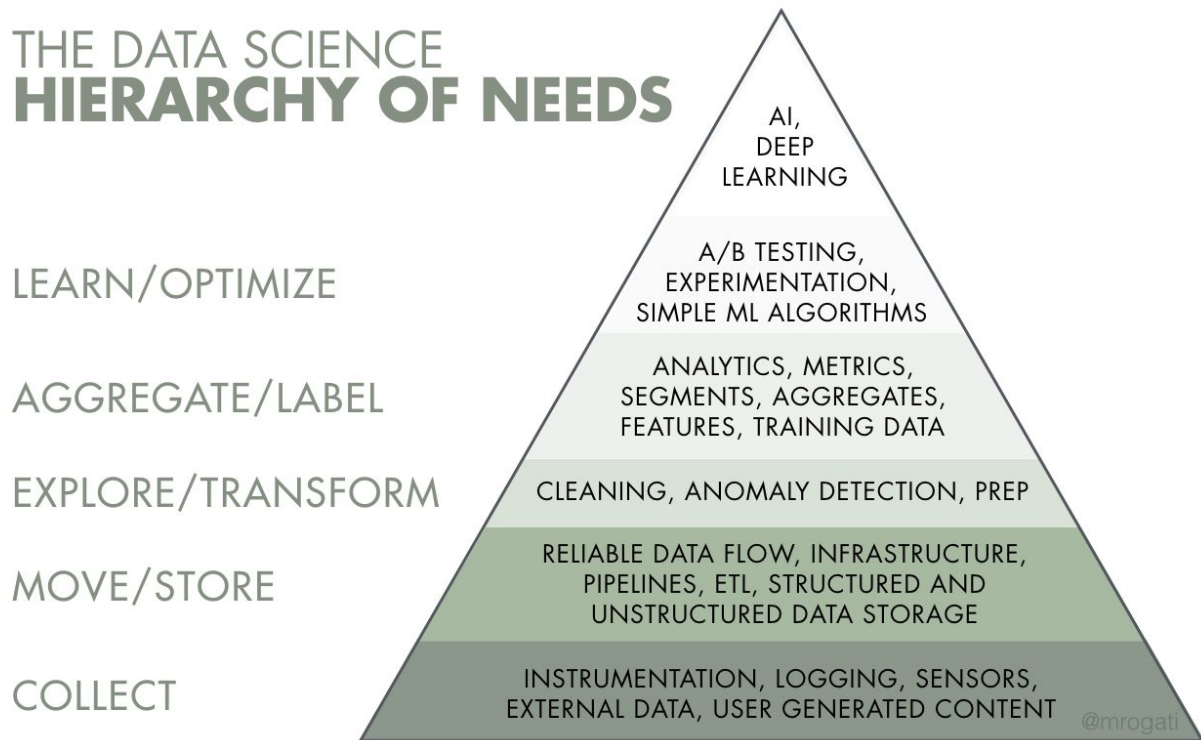
<https://medium.freecodecamp.org/the-rise-of-the-data-engineer-91be18f1e603>

<https://medium.com/@rchang/my-two-year-journey-as-a-data-scientist-at-twitter-f0c13298aee6>



Different jobs address different needs

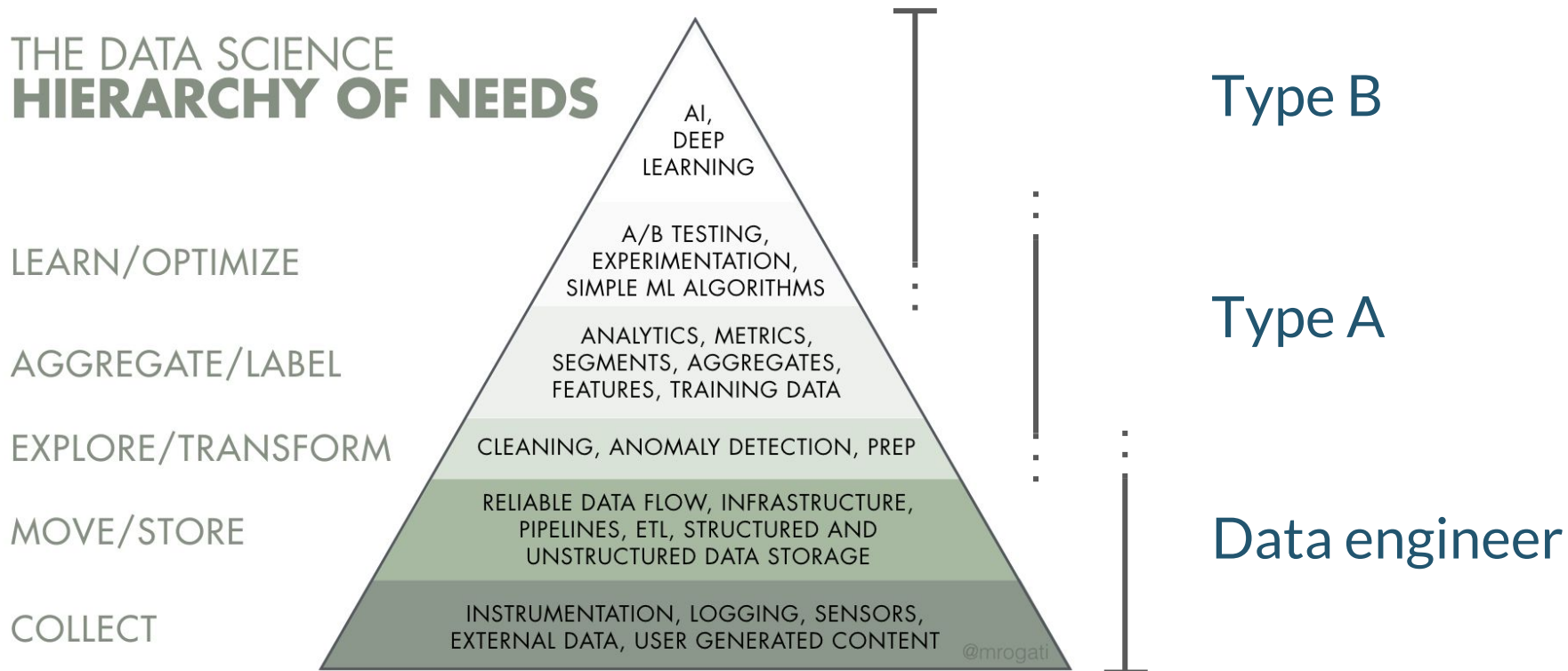
THE DATA SCIENCE **HIERARCHY OF NEEDS**





Different jobs address different needs

THE DATA SCIENCE HIERARCHY OF NEEDS





Data Engineers

Build and maintain the organization's data pipeline systems

Data pipelines encompass the journey and processes that data undergoes within a company. Data engineers are responsible for creating those pipelines.

Clean and wrangle data into a usable state

Data engineers make sure the data the organization is using is clean, reliable, and prepped for whatever use cases may present themselves. Data engineers wrangle data into a state that can then have queries run against it by data scientists.



Type A Data Scientists

***Type A Data Scientist:** The A is for Analysis. This type is primarily concerned with making sense of data or working with it in a fairly static way. The Type A Data Scientist is very similar to a statistician (and may be one) but knows all the practical details of working with data that aren't taught in the statistics curriculum: data cleaning, methods for dealing with very large data sets, visualization, deep knowledge of a particular domain, writing well about data, and so on.*



Type B Data Scientists

***Type B Data Scientist:** The B is for Building. Type B Data Scientists share some statistical background with Type A, but they are also very strong coders and may be trained software engineers. The Type B Data Scientist is mainly interested in using data “in production.” They build models which interact with users, often serving recommendations (products, people you may know, ads, movies, search results).*



These categories are more “types of data science work” than “types of data science jobs”

- Most jobs will contain elements of each type of work.
- The amount of specialization will depend on the maturity of the company.
 - Even in specialist roles it helps to be familiar with the rest of the stack.
- Different companies will call the same job by different names
 - Data Science vs Data Engineering is a fairly standard distinction, but within Data Science job titles and responsibilities vary widely.



Some example job postings for each type of job

Data Engineer (from “Lead Data Engineer” posting at Civis)

You will be leading the team responsible for:

- Building ETL pipelines between the Civis platform and outside systems
- Building Civis' internal data pipelines for various sources, including but not limited to: survey data, consumer data, public data, and client data
- Designing and building the next generation of Civis' data infrastructure
- Building and maintaining the core datasets for consulting, R&D, and deployed applications
- Working with our devops, legal, and security teams to ensure proper privacy and compliance for Civis data

Type A Data Scientist (from “Data Scientist” posting at the CTA)

Under general supervision, provides data-based insights and recommendations that lead to change. Coordinates within the team to augment performance analysis insights for the entire agency. Works with the executive office and operating departments by delivery of key performance metric relationships and performance improvement strategies to enhance organizational performance and managerial effectiveness. Manages multiple diverse and complex projects to implement performance improvement strategies resulting from data reporting and modeling and serves as liaison to address structural data needs and achieve efficiencies and process improvements throughout the Authority.



Some example job postings for each type of job

Type B Data Scientist (“Data Scientist” at Stitchfix)

This role will be responsible for the Stitch Fix algorithmic allocation efforts, which helps the company send the right amount of inventory to each warehouse. In this role, you will create, improve, and run allocation strategies that maintain a healthy inventory across Stitch Fix warehouses in collaboration with business partners. This role has more software engineering than other data scientist positions on the team and is also closely tied to inventory forecasting efforts.

Type A + B Data Scientist (“Data Scientist” at Stitchfix)

This role will be focused on quantifying behavior of stylists (humans in the loop) and building tools that augment their day-to-day tasks. Together with the data science and styling strategy and field teams teams, you will drive end-to-end deep analyses of styling behavior by partnering across the organization to frame problems and establish hypotheses. You will have an opportunity to turn your findings into algorithms and/or product features to assist stylists in helping clients find what they love.

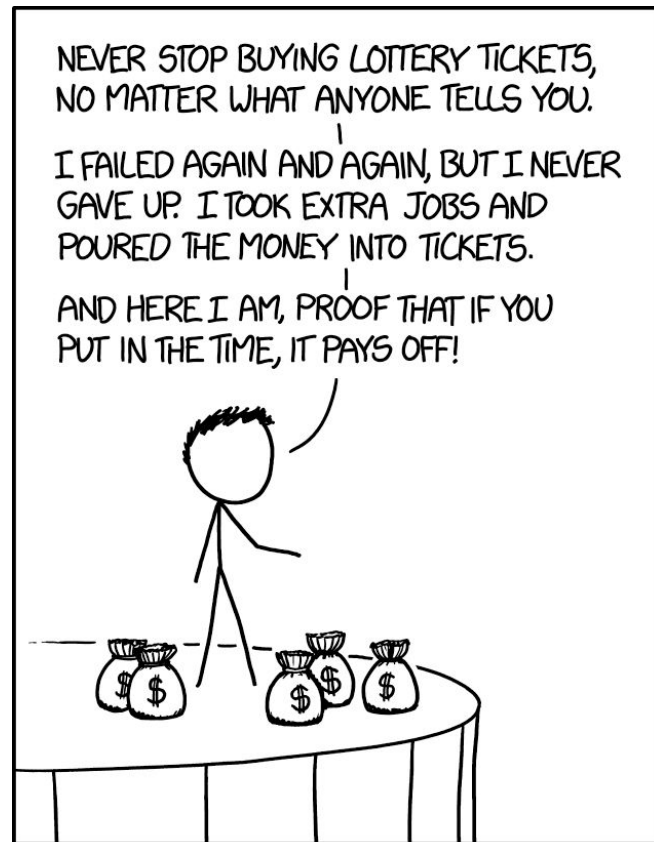
Getting a data science job

Data science hiring is a noisy process

Data science is a broad and poorly defined field, so it's hard to know ahead of time what kinds of technical questions you'll get in an interview.

My advice is to get a baseline level of knowledge (from Coursera or something) and then focus on making yourself stand out in other ways.

Caveat: this is based on my own experience getting a job and the interviews I've participated in at Civis, so N is small.



EVERY INSPIRATIONAL SPEECH BY SOMEONE SUCCESSFUL SHOULD HAVE TO START WITH A DISCLAIMER ABOUT SURVIVORSHIP BIAS.





Four recommendations

- Develop your skills with projects
- Have a public portfolio
- Go on informational interviews
- Do an internship



Four recommendations

- Develop your skills with projects
- Have a public portfolio
- Go on informational interviews
- Do an internship



ENTRY LEVEL JOB OPENING:
Hiring recent college grads

REQUIREMENTS:
5 years of experience, 6 Olympic gold medals, and superpowers.

← these are all ways to show you have experience even if you haven't had a data science job yet



Develop your skills with projects

Which skills?

- **Programming**
 - Python or R
 - SQL
- **Software engineering**
 - code style (flake8 for python, Google or tidyverse style guides for R)
 - command line basics
 - version control (git)
 - unit-testing
- **Stats / Machine Learning**
 - ~ the contents of a basic MOOC on data science



Develop your skills with projects

What kind of projects?

- Analyze an interesting dataset
 - Google dataset search: <https://toolbox.google.com/datasetsearch>
 - weekly newsletter of interesting datasets: <https://tinyletter.com/data-is-plural>
 - web-scrape something
- Build something neat with code
 - twitter bot
 - a library that helps you with your research
- Copy something cool but with different data
 - there's a list of newsletters with interesting projects at the end of the slides



A note about Kaggle

I'm always more impressed with a project somebody came up with on their own than with a high rank in a Kaggle competition. Kaggle problems aren't representative of real data science work:

- the data is too clean
- the problem is too clear cut
- in the real world it almost never makes sense to spend dozens of hours for a marginal gain in accuracy

My hot take: a Kaggle rank is like your GPA as a PhD student. It doesn't tell me anything about the part of your work that I care about and if you spend too much time worrying about it you're probably missing out on doing more valuable things.



Have a public portfolio

This is the single most important piece of advice in this talk.

Write a blog, put your code on github, submit proposals to give talks at data science meetups or conferences, get your work out in public however you can.

- Show off your communication skills.
- Show off whatever technical skills the project used.
- You can bring up these projects in interviews to “show, don’t tell”
- A good blog is a way to show that you have “experience” even if you’ve never had a data science job.



Go on informational interviews

What's an informational interview?

- Informal 30-60 minute conversation with somebody in the field you're interested in about career advice, what their job is like, what the industry is like, etc.
- A conversation that **you** are supposed to lead.
- Not the place for you to directly ask for a job.
- A chance to hear how data scientists talk about data science so that you can talk about it the same way in job interviews.

<https://www.careercenter.illinois.edu/instructable/conduct-informational-interview>

<https://medium.com/@treycasey/do-you-have-time-for-a-quick-chat-c3f7e46de89d>



Go on informational interviews

Who should you talk to?

- People at companies you're interested in
- People in your extended network who work in the field.
- People whose work / blogs you like.

Other tips

- Be appreciative - anybody who says yes is doing you a favor.
- Send a thank you email.

<https://www.careercenter.illinois.edu/instructable/conduct-informational-interview>

<https://medium.com/@treycasey/do-you-have-time-for-a-quick-chat-c3f7e46de89d>



Do an internship

I realize this is a bit of a catch-22.

My advice: if you're sure you want to leave academia do an internship every summer. If you think you might want to leave academia do an internship the summer before you graduate.

- For PhD students, companies are more interested if you're about to graduate.
- Internships >>> bootcamps because you get real world experience instead of more coursework. I've heard good things about the Insight fellowship though.



Do an internship

Civis is hiring for summer 2019 internships

[Type A in Chicago](#)

[Type A in Washington, DC](#)

[Type B in Chicago](#)

[Software Engineer in Chicago](#)

We'll also be at the career fair tomorrow if you want to come say hi.

Transitioning from academia to industry



Adjustments I had to make once I left academia

My time has value.

In grad school it wasn't uncommon for me to bang my head against a problem for months without making progress. This doesn't fly in industry - if you're stuck for a while, ask for help. If you're still stuck, re-consider whether what you're trying to do is worth it.

The 80% solution is usually the best solution

If you can get 80% of the perfect solution with 20% of the effort (which you often can) then you should probably stop there and move on to the next thing.



Adjustments I had to make once I left academia

Work needs justification beyond being interesting

The work that is most interesting is not always the work that is most valuable. Overall the problems are still very interesting, I just can't spend a lot of time on something for no other reason than because I find it interesting.

The environment is (usually) less “academic”

I really enjoyed the atmosphere of grad school: being surrounded by super smart peers, lots of intellectual curiosity, learning for the sake of learning, etc. The environment at Civis is like this, but (from talking to friends in other DS jobs) this is not always the case.



Questions?



References

Type A vs Type B data scientists

<https://medium.com/@rchang/my-two-year-journey-as-a-data-scientist-at-twitter-f0c13298aee6>

Data engineering vs Data science

<https://medium.com/@rchang/a-beginners-guide-to-data-engineering-part-i-4227c5c457d7>

<https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007>

Applications and interviews

<http://www.erinshellman.com/crushed-it-landing-a-data-science-job/>

http://treycousey.com/data_science_interviews.html

<http://hookedondata.org/Advice-for-Applying-to-Data-Science-Jobs/>

Portfolio stuff

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

<https://medium.com/@jasonkgoodman/advice-on-building-data-portfolio-projects-c5f96d8a0627>

<https://towardsdatascience.com/how-to-build-a-data-science-portfolio-5f566517c79c> - this one has a ton of good links to other posts as well



References

Other

<https://towardsdatascience.com/red-flags-in-data-science-interviews-4f492bbed4c4>

<https://towardsdatascience.com/why-so-many-data-scientists-are-leaving-their-jobs-a1f0329d7ea4>

<https://medium.com/@rchang/advice-for-new-and-junior-data-scientists-2ab02396cf5b>

Example portfolios

<http://www.gregreda.com/blog/>

<http://www.lizsander.com/>

<http://thespread.us/>

<http://hinnefe2.github.io/> (mine)

Data science newsletters

<http://lineardigressions.com/> (podcast)

<http://flowingdata.us2.list-manage.com/subscribe>

<http://roundup.fishtownanalytics.com/>

<https://www.getrevue.co/profile/datamachina>

<https://www.datascienceweekly.org/>

<https://tinyletter.com/data-is-plural> - This one just highlights interesting datasets