

# Imposter syndrome

Brandon Rohrer

part of Course 121 **Navigating a Data Science Career**

[ **Portuguese version**, translated by **Marcus Oliveira da Silva**]

## **I am not a real data scientist.**

I have never used a deep learning framework, like TensorFlow or Keras.

I have never touched a GPU.

I don't have a degree in computer science or statistics. My degree is in mechanical engineering, of all things.

I don't know R.

But I haven't given up hope. After reading a bunch of job postings, I figured out that all it will take to become a real data scientist is five PhD's and 87 years of job experience.

If this sounds familiar, know that you are not alone. You are not the only one who wonders how much longer they can get away with pretending to be a data scientist. You are not the only one who has nightmares about being laughed out of your next interview.

Imposter syndrome is feeling like everyone else in your field is more qualified than you are, that you will never get hired or, if you already have been, that you are a mistake of the hiring process. Despite its statistical implausibility, most of us feel below average. Based on my conversations with colleagues, I estimate that 9 out of 10 of us suffer from imposter syndrome at one time or another. (If this sounds entirely unfamiliar to you, I recommend an introspective reading of "**Unskilled and unaware of it**" by Kruger and Dunning.)



Even Ewoks feel like imposters sometimes. (Photo courtesy of Diane Rohrer.)

## What a real data scientist looks like

“Data science” is a term that has generated a lot of excitement and, like a magnet, has pulled in lots of nearby subfields. The field we call data science is still relatively young, yet already too broad for an individual to be an expert in every corner of it. In my experience, the master-of-all-trades data science unicorn is a mythical beast. None of us can cover all the bases. So how are we to proceed?

There are two paths forward: generalist and specialist.

### A good generalist

- is superficially familiar with every part of data science,
- recognizes all the jargon and technical terms,
- has a good notion of what tools and expertise are needed to solve a given problem, and
- asks insightful questions in technical reviews.

### A good specialist

- understands one area deeply,
- can explain their area of expertise to non-experts,
- understands the tradeoffs between different approaches,
- is up to date on current research and new tools, and
- can use their tools quickly to produce high-quality results.

A generalist does not necessarily know the details of how an algorithm works and the tricks of using a tool. They will tell you that data cleaning is critical, but may not be able to enumerate the trade-offs between methods for replacing missing values. They will tell you that Spark is a good way to speed up your computations, but may not be able to advise you on the best settings to use.

A specialist does not necessarily know much about something that is outside their area. They will know the best architecture for running a linear regression on 500 million data points, but may not be able to explain a naive Bayes classifier. They will keenly grasp the trade-offs between square loss, hinge loss and logistic loss, but may be unable to query data from a Hive table.

Another way to describe generalists and specialists is “broad” versus “deep”. They are both technically savvy, but their expertise is distributed differently. We are all part generalist and part specialist. As you evolve through your career, you get to find the mixture that works best for you.

This distinction can be helpful when **hiring data scientists too**. Asking specifically for research experience in deep neural networks or a background in financial data visualization will draw applicants that fit your needs more effectively than calling for a “full-stack” data scientist.

## How to prove that you are a real data scientist

Traditionally we establish our qualification in a field with advanced degrees. Unfortunately for most of us, there are few such degrees available in data science. We have no piece of paper to use as a shield when someone questions our qualifications. So what do we do instead? How can we answer our critics, or interviewers, our colleagues, and harshest of all, the voices in our head?

Consider woodworking. Imagine that you want to install a custom cabinet in your kitchen. Three carpenters show up inquiring after the job. The first one presents you with a certificate. She says, “I apprenticed with the premier cabinet maker in the city for seven years.” The second opens her toolbox and says, “My chisels are of the latest design, and no one has a sharper plane.” The third hands you a small box, cherry-colored and perfectly smooth. When you pull the handle with a fingertip, a drawer slides out soundlessly. She says, “I made this.”

Certifications, tools and portfolio are all popular ways for establishing credentials. I won't argue that one is superior to another, but portfolios are particularly effective for data scientists. Certifications are few and not yet standardized. Listing algorithms and computer languages we have used doesn't convey our depth of familiarity with them or what we can do with them. **Building things** shows to a non-technical audience what we can do for them and demonstrates our expertise for technical interviewers and colleagues. Of course, this doesn't guarantee that you'll get a job on your first interview. But even if you don't, that's normal. **Keep interviewing.**

## How it feels to be a real data scientist

Note that both generalists and specialists have lots of things they don't know. This means that even real data scientists will spend most of their days feeling lost. Our project lead will ask us questions that we don't know the answer to. Colleagues will talk comfortably about algorithms we've never heard of. Teammates will write code that we can't begin to decipher. Articles will cite “hot” subfields that we didn't know existed. Archiv papers will throw around equations that may as well be hieroglyphic gibberish. Interns will point out fundamental flaws in our reasoning. This is OK. You're not doing it wrong. This is OK.

Our goal isn't to accumulate answers, but to ask better questions. If you are asking questions and using data to find answers, **YOU ARE A DATA SCIENTIST**. Period.

**Brandon**

August 22, 2017

All text and images CC0 except where noted.

The opinions here are wholly my own.

