

# **A Meandering Path Towards a Career in Data Science - One Math Grad's Journey**

## Sample Probability/Statistics Question from an Interview at Indeed

- You flip a coin 10 times.
- It comes up heads twice.
- Is the coin biased?

Given time, we'll dig into this, but I doubt we'll get there.

I have put a solution in appendix to cover this eventuality, and also for those who could not attend the presentation.

## About Me

- I currently work as a Lead Data Scientist at Ascension - the largest non-profit healthcare provider group (hospitals, clinics, long-term care facilities, etc.) in the US and 2nd largest such group in the US in general.
- I started as a data scientist (predictive modeling, big data analytics, etc.), but now function more as a project leader, while also running our internal education system (my real goal and best part of my job).
- I find working at Ascension particularly satisfying as I loved being a teacher, and working for Amazon or such just didn't appeal. At Ascension, I still feel like I'm giving back to the community as we give \$2B in free healthcare to the less fortunate each year.
- I got my job by referral from a fellow student from my Data Science bootcamp - more on this later.

## Before UH Manoa

Vancouver, Canada



Dublin, Ireland



## **Before UH Manoa**

- Graduated in 1996 from Simon Fraser University in Vancouver, British Columbia
- Worked as a TA for 2 years before graduation and one year after
- Left academia for a while, random jobs (being Gen X sucked for careers!)
- Moved to Ireland in 2000, worked as a TA, tutor trainer and tutor for disabled students at University College Dublin in Dublin, Ireland
- Applied to UH Manoa and started in August 2005

## While at UH Manoa and/or in Hawaii



## **Teaching While at UH Manoa and/or in Hawaii**

- Started teaching the evening Math 140 class in my second semester and taught every summer as an instructor (2005 - 2012), up to and including Calc III
- Worked at various UH System CC's during the Ph.D. and full-time for a couple of years after graduation
- Teaching is not a prerequisite to becoming a data scientist, but, in my opinion, the real need in industry is not for the technically accomplished, but rather for those who can communicate complicated topics to all levels of listener (both skills are clearly necessary, but the latter is rarer)
- The higher you go up the power ladder, the less time people have and the more areas they lead, hence the less focus they can expend on any one topic:
  - Immediate superior - Director level, talk to them like a smart undergrad
  - Next up - VP level, talk to them like a smart high schooler
  - Next up - C suite, talk to them like a smart middle schooler (unless they're the Chief Data Science Officer or Chief Data Officer, in my experience)



## After Hawaii - Austin, TX





### After Hawaii - Austin, TX

- Moved to Austin, Texas in May 2017
- Chose Austin because it was booming (and also it used to be very cheap to live here) - this may be less important in the new age of remote work, but it really mattered back then
- Got a job at Macmillan Publishing as a Calculus Content Author/Editor, kept this until becoming a Data Scientist
- Taught at St. Edward's University for a semester at the start, but stopped to study Data Science
- Studied Data Science independently and somewhat haphazardly before realizing I needed more structure
- Attended Galvanize Data Science Immersive for 12 weeks in Spring 2018
- During this time, we had some family difficulties, so although I passed the course (barely), I spent the rest of 2018 studying the material more closely before beginning interviews in late 2018

## Michael Joyce Ph.D.

- I started at Ascension Technologies in March 2019 by referral from a fellow bootcamp student.
- This referral came with no technical interview whatsoever (quite unusual), as referrals are the strongest indicator of success in a job apparently and Ascension back then was fine with this.
- I then referred Lukasz Grabarek, another UH Ph.D. grad, who was also hired solely on the basis of my recommendation, without a technical interview and without even a bootcamp (very unusual).
- The lack of a technical interview is uncommon, but the strength of a referral is common.
- Going to meetups (virtual probably), attending a known bootcamp, or getting a Master's in Data Science are some ways to meet people and to grow a network.
- **This is as crucial as your studies**, although it can be hard for the more introverted and/or physically distanced.



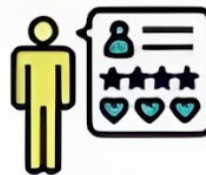
Recommendation



Networking



Resume



Word of mouth



Online database



Refer a friend



Approval



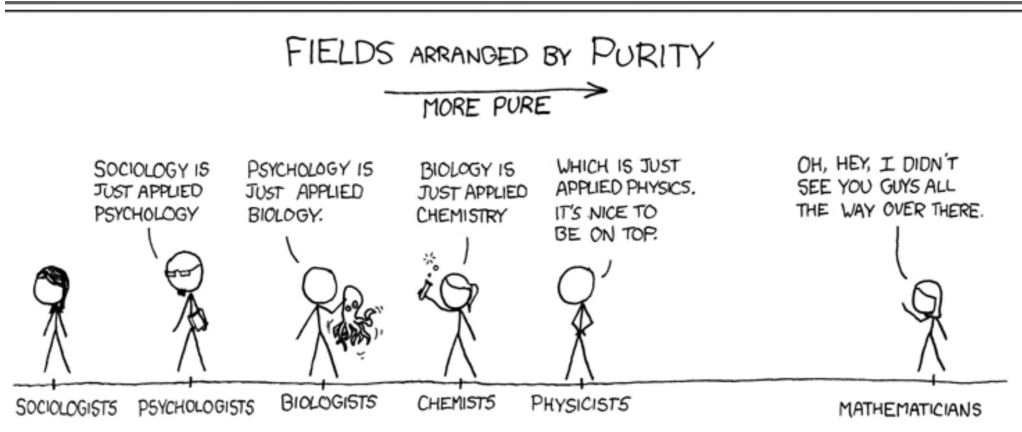
Team advocacy



Vouch for someone

## Why Math as a Path to Data Science?

- Most of the methodologies are based in Math, although you rarely use it directly.
- Coding is also important and should be studied concurrently, but computer science degrees are much more common, so standing out is tough.
- Statistics is also important and, if possible, should also be studied concurrently. However, it's not hard to pick up Stats on your own if you understand Math.
- It's a lot easier to get teaching work while studying math and, in my opinion, this really helps you stand out versus your competition.
- Because this cartoon is arguably a little over the top, but it is how others see Math and that can often be to your advantage!



## Why Pursue a Math Ph.D. for Data Science?

- There is an automatic respect level granted to a Ph.D. in the data science sphere that is not always there for those with Master's degrees. Interviews, in particular, will be much easier to get with a Ph.D.
- This respect continues throughout your career. **You are listened to almost automatically.**
- Many companies will start you directly as a Senior Data Scientist if you have a Ph.D. This has an obvious income differential, but also is a great kickstart to a career.
- You develop an ability to analyze complicated processes on a much deeper level while writing a dissertation compared to while writing a thesis. There is also usually more original and independent research with the former rather than the latter. These facts are recognized both formally and informally in the DS space.
- If this sounds indirect, it is, but it is also widely recognized. Here is a quote from a DS article:

*People in tech understand that the value of a PhD doesn't necessarily lie in the specific research that one has done. Rather, they **know** that PhD programs churn out some of the smartest and hardest working people in the world.*

*PhD completion is a natural selection mechanism for high intelligence and high dedication to work, which in turn means that recruiters at large tech companies have to do far less work to find quality individuals.*

# Appendix

---

## Sample Probability/Statistics Question from Interviews

- You flip a coin 10 times.
- It comes up heads twice.
- Is the coin biased?

The point of this question is **not** to be able to actually answer it. It was from an interview at Indeed and they told me that few had ever answered it in the 30 minutes given.

I took about 5 minutes. This isn't a humble brag, believe it or not. I taught probability at UH and various other places, so this was in my exact wheelhouse. I GOT LUCKY!

My point is that we are MATH people. Few are and you can wow them even if you get half-way with things like this.

They are really watching to see how you reason your way through a difficult problem.



## Sample Probability/Statistics Question from Interviews

- You flip a coin 10 times.
- It comes up heads twice.
- Is the coin biased?

**NOTE: this is a tricky question on a variety of levels.**

First, set the null hypothesis to be that the coin is fair. The alternative hypothesis is that the coin is biased (towards heads OR tails, more on this later).

The chances of getting 2 heads out of 10 flips is  $p = (10C2)(1/2)^{10} = 45/1024$ . Assuming a significance level of 0.05, this is slightly less than this significance level, so we might reject the null hypothesis and assert that the coin is biased.

Apparently, this is the most common answer they got at Indeed. Unfortunately, it's not correct.

## Sample Probability/Statistics Question from Interviews

- You flip a coin 10 times.
- It comes up heads twice.
- Is the coin biased?

A statistical test hinges on the chances of getting a result at least as unlikely as that observed.

The chances of getting 2 heads out of 10 flips is  $p_1 = (10C2)(1/2)^{10} = 45/1024$ .

The chances of getting 1 heads out of 10 flips is  $p_2 = (10C1)(1/2)^{10} = 10/1024$ .

The chances of getting 0 heads out of 10 flips is  $p_3 = (10C0)(1/2)^{10} = 1/1024$ .

The combined probability of these possibilities is  $56/1024$ , which is slightly greater than 0.05. Hence, we should fail to reject the null hypothesis and we cannot conclude that the coin is biased.

(Recall, by the rules of statistics, we also can't say that the coin is unbiased, as we're not testing for that eventuality.)

## Sample Probability/Statistics Question from Interviews

- You flip a coin 10 times.
- It comes up heads twice.
- Is the coin biased?

Note that there is a more rigorous way to state this in statistical language:

**The data appears to give grounds that the coin might be biased, but this evidence is not statistically significant at the conventional 5% level of significance, i.e., a statistical test at the 5% level would not reject the null hypothesis that the coin is not biased.**

Stats language ... ugh!

## Sample Probability/Statistics Question from Interviews

- You flip a coin 10 times.
- It comes up heads twice.
- Is the coin biased?

Worse yet, a two-tailed test is appropriate in this case as it we are testing for bias in general, not towards either tails specifically (or heads specifically either), so if our significance level is 0.05, we should consider events at least as unlikely as 0.025 on each side. Hence, the first (incorrect) argument and the second (correct) argument would both lead to failing to reject the null hypothesis.

If the original significance level was 0.10, then the previous arguments would still apply, falling on both sides of the argument depending on incorrect and correct understanding. Note, however, that the go-to significance level unless stated otherwise is usually 0.05. (Why? Who knows? Stats...)

With either significance level, the proper result leads to failing to reject the null hypothesis. That is to say, we cannot determine whether the coin is biased or not given the result of the 10 coin flips.

## Sample Probability/Statistics Question from Interviews

**FINALLY, NONE OF THESE DETAILS REALLY MATTER!!!!**

The point is the ability to have the conversation.

- The real point of this is exposition.
- Talk out loud throughout the process.
- Explain your reasoning at each step.
- Ask questions even, they are often allowed to answer.

**They want to see how your brain works, not whether or not you can solve some ridiculous and arbitrary question in an arguably unreasonable amount of time.**

## Contact Information

Email: [michaeljoyce217@gmail.com](mailto:michaeljoyce217@gmail.com)

LinkedIn: <https://www.linkedin.com/in/michaeljoyce217/>

Please feel free to send questions and ask for further advice or assistance.