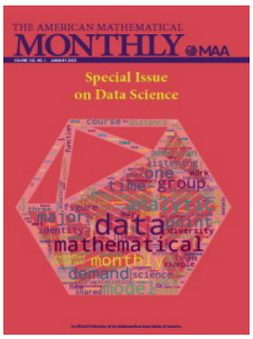


Taylor & Francis
Taylor & Francis Group

The American Mathematical Monthly

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uamm20

Becoming a Data Scientist: An Impractical Guide

Anna Haensch

To cite this article: Anna Haensch (2025) Becoming a Data Scientist: An Impractical Guide, The American Mathematical Monthly, 132:1, 6-13, DOI: 10.1080/00029890.2024.2409064

To link to this article: https://doi.org/10.1080/00029890.2024.2409064

Published online: 13 Dec 2024.

Becoming a Data Scientist: An Impractical Guide

Anna Haensch

Abstract. I remember being an undergrad math major, knowing that math was a simultaneously fun and powerful tool, but not quite understanding how I could be a "professional mathematician," or what that even meant. Sure, math is everywhere, but at the time I needed something a bit more concrete than a vague gesture in the direction of *everywhere*. Today, I still don't know everywhere that math is, but I've found a few interesting places. This is a story about those places. It's the story of life on the tenure track, around the newsroom, and in industry. It's also the story of some of the meaningful people and ideas I met with along the way. But mostly, this is the story of how I became a data scientist.

1. MISSION STATEMENT. Ever since making *the change*, people from all parts of my life and even many strangers, want to know how I did it and why. Some just want to know what it means to be a data scientist, and some want to know precisely the prerequisites to make this kind of move. Others just want to know what a nice mathematician like me is doing in a field like this.

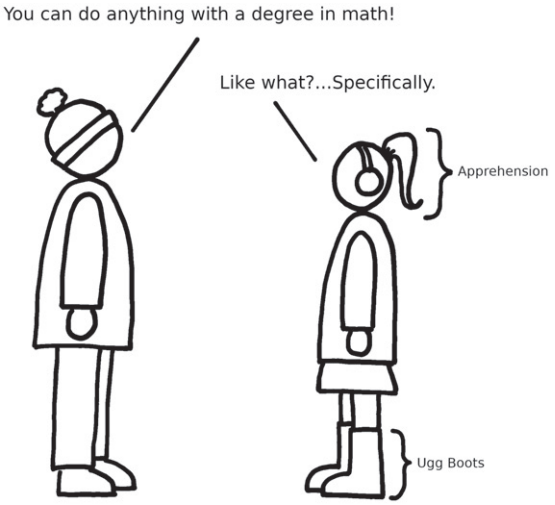
Looking back on my journey that brought me here, there were things that shaped me profoundly that you will never find on my CV: friendships and conversations that steered me in a certain direction, hobbies that turned into useful skills, fears and anxieties that sometimes forced my hand, and an overwhelming sense that I could never have planned any of this.

Maybe you're among the data sci-curious. Maybe you're ready to make a change. Or, maybe you can sense that your students are curious about *it*, whatever *it* even is. In any case, I'm here to offer you one possible way to become a data scientist. This is the way that worked for me, but it won't work for everyone. If this story feels relatable and sparks something in you, then I'm happy. If it doesn't, don't worry; there are lots of other stories out there, and this is only one.

2. A YOUNG MATHEMATICIAN. I went to college at the State University of New York at New Paltz, a small state school on the edge of the Catskill mountains. Although I went there to study fine arts, the math department beckoned from the start. I was lured into some exciting undergraduate research and before long I put down my charcoals in favor of a TI-89. In my second year, driven partially by my love of math but also by what I assumed to be more sensible employment prospects, I declared a math major. Once I became a math major something kept happening: professors, parents, and grown-ups everywhere loved telling me how *great* it was that I was studying math. "That's so great," they'd say, "you know you can do anything with a degree in math!"

I, flattered but apprehensive, always asked, "Yeah? Like what?" and the answer was a universally hand-wavingly vague version of, "Literally *anything*!" But other than becoming a high school teacher or an actuary, nobody ever had anything concrete in mind. Since I didn't want to do either of those things, I made the only logical choice for a confused but curious 22 year old: I went to graduate school.

doi.org/10.1080/00029890.2024.2409064
MSC: 01A07, 01A70, 01A80



In 2007 I started my Ph.D. at Wesleyan University, and shortly thereafter, I started working with my Ph.D. advisor. My advisor trained me how to think deeply about one area of mathematics for prolonged periods of time. My chosen area was number theory; I was so enamored with the core ideas of a field that stretched back all the way back to antiquity: questions about sums of squares, the distribution of primes, and the geometry of numbers. As I worked my way through old theorems, I felt like I was communing with long ago forefathers like Riemann, Gauss, Fermat and Pythagoras. It felt ancient, it felt mysterious, it felt like art.[1]

There was one theorem that really struck my fancy. It's an old one, maybe you've seen it before, but I'll include it here, just in case.

Theorem 1 (Gauss, 1796). *EYPHKA!* $Num = \Delta + \Delta + \Delta$

When Gauss wrote "Δ," he was referring to what we call a *triangular number*, which is a possible number of points in an evenly spaced equilateral triangular arrangement (see Figure 1), but also have an algebraic interpretation as numbers generated by

$$T(x) = \frac{x(x+1)}{2}$$

where x is an integer.[2] Gauss's theorem means that every natural number can be written as the sum of three triangular numbers, or alternatively, $n = T(x) + T(y) + T(z)$ has a solution for every $n \in \mathbb{N}$ with x, y, and z as integers [**1**]. Eureka!

What I loved about this theorem was that it straddled the line between algebra and geometry. Not only is this a statement about the solvability of an algebraic expression, but also it is a statement about the geometry that underlies familiar numbers.

From this starting point I started to weave a path through the classical results. What could I say about other kinds of numbers with different types of geometry?

[1] Had it existed at the time, this part of me would have strongly identified with the *Dark Academia* aesthetic made popular on TikTok.

[2] You might have seen a similar definition before where x is restricted to the natural numbers. This is the difference between a triangular number ($x \in \mathbb{N}$) and a generalized triangular number ($x \in \mathbb{Z}$). This becomes important for higher order polygonal numbers, but for triangular numbers the two sets turn out to be the same. Go ahead. Try it!

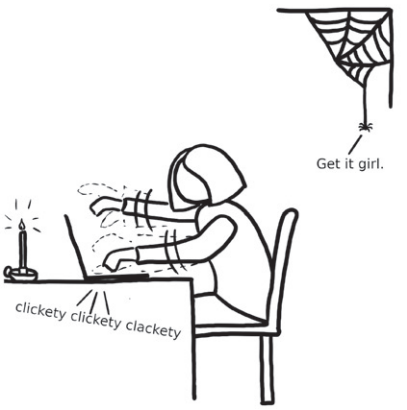
Figure 1: The triangular numbers $T(1)$, $T(2)$, $T(3)$, and $T(4)$.

I knew about Lagrange's four square theorem, and eventually learned about pentagonal numbers, hexagonal numbers, and a whole world of higher order polygonal numbers [2]. Could these ideas about connecting the algebra and geometry of polygonal numbers be generalized? And if so, how? And to what extent? This line of thinking eventually led me straight to the famous *tenth problem*, posed by David Hilbert at the 1900 International Congress of Mathematicians,

> Given a Diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: To devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers [3].

Later work by Yuri Matiyasevich, Julia Robinson, Martin Davis, and Hilary Putnam determined that no such process exists [4]. Nevertheless, this brush with determining solvability through finite processes was my first encounter with algorithmic number theory. I loved the idea of it: building concise numerical recipes that future generations of mathematicians could follow to a guaranteed outcome.

I toiled for what felt like an eternity (but in hindsight feels very short), eventually putting together a dissertation I was proud of, and along the way my advisor taught me how to be a professional mathematician. He sent me to conferences and summer schools where I met lots of different people, he introduced me to his network, and he helped me find my place in the community.



During one summer visit to MSRI I fell in with a group of Sagemath developers and found that I quite enjoyed learning Python and being part of an open source project. It felt different enough from my immediate thesis work that it gave me a bit of breathing room, but it was still on-theme enough to feel productive. Not everyone shared my enthusiasm for this work. Even though building tools to enable the important work happening in computational and algorithmic number theory had immediate and apparent value, several senior mathematicians warned me about sinking too much time into it since writing code "doesn't count" in the sense that matters most (i.e., in your tenure dossier). I understood what they were saying, but I thought it was fun, and wanted to do it anyway. So I did, and on the occasional Friday night I would turn out all the lights and pretend I wasn't home. Sitting alone in my dark kitchen with a glass of wine, a Sage notebook, and a text editor, I'd get to work on my lambda functions.

Finally in 2013 my dissertation reached its final form, and I was ready to defend. Upon graduating I was offered a prestigious visiting scientist position at the Max Planck Institute in Bonn Germany and a tenure track job at Duquesne University.

I had arrived. I was finally a mathematician.

3. SOMETHING ELSE WAS A'BREWING. While all of this scholarly work was going on, I was also doing some other seemingly inconsequential things in the background. I met my best friend. She was an avid runner who got me into distance running and helped me train for my first half marathon. I slowly started making friends with other runners in my academic orbit, and together we worked through many great fears and anxieties during those companionable mid-run conversations.

My best friend was, and remains, a singularly stylish woman, and she also introduced me to the early aughts world of fashion blogging, inspiring me to write my own math/fashion blog. I enjoyed writing enough that I thought it might be fun–if somewhat of a long shot–to apply for the AAAS-AMS Mass Media Fellowship. I pulled together some writing samples from my blog and on the strength of that work I was invited as the 2013 fellow to National Public Radio (NPR) in Washington D.C.

Having just spent the last six years focused single-mindedly on a niche area of mathematics, the human connections and social relevance I felt in the newsroom were a revelation. I spent my summer interviewing scientists, writing blog posts, and collaborating on the occasional piece of on-air content [**5**]. The skills I brought to the job were my ability to grasp the content of complicated scientific research papers quickly, and confidence in my ability to do arithmetic. The skill I learned on the job was how to convey sophisticated scientific ideas to a clever but not yet in-the-know audience. I also learned how to have more fun with language and use words like "poop" in a completely serious context [**6**].

While I was at NPR I started to wonder about the role of mathematicians in the newsroom. It was true that a certain level of numerical literacy is useful in all kinds of reporting, which feels completely obvious when we talk about things like political polling and uncertainty [**7**], but is also vital in the ways we talk about poverty and inequality [**8**].

I started to notice that some journalists were using data to research and report stories in a way I hadn't seen before, something called *data journalism*. The data teams at *ProPublica* and *The New York Times* were doing exciting work, *FiveThirtyEight* was expanding its coverage to include politics, science and economics, and the American public was starting to throw around words like "data viz" and "algorithm." A new way of being a mathematician was starting to take shape in my mind.

Beyond just being a useful tool, I also began to notice that in many cases the math *was* the story. This point was made very explicitly in Cathy O'Neil's influential *Weapons of Math Destruction* [**9**] and later in Safiya Umoja Noble's *Algorithms of Oppression* [**10**], both of which describe a world where data and algorithms have an outsized impact on our lived experience and access to resources both material and knowledge-based.

Bringing all of these ideas together I started to wonder if maybe the newsroom was exactly the place for a mathematician like me. My summer fellowship ended and I rejoined the ranks of academia, but this experience changed me in ways that weren't totally apparent yet. I kept on writing papers, applying for grants and teaching calculus. So much calculus. But in the back of my mind I knew mathematics wasn't my forever home. I liked the freedom of academic inquiry, but felt professionally alienated and often out of place.

This is when I started to do some strange things that I didn't understand at the time. I joined Twitter. I kept on blogging but now for the American Mathematical Society [**11**]. I started a podcast with my best friend; remember her from the previous section? The podcast was called *The Other Half*. She was a policy analyst, I was a mathematician, and we started asking questions about the big ways that math shapes how we experience the world, from racial segregation to ride-sharing. I enrolled in

some evening MBA courses, learning about business statistics and economics, and I continued to ponder these big questions about math and public good. I also started working through Coursera materials on data science. It was a field I didn't know too much about, although I knew that it had something to do with coding in Python—which I loved—and offered lots of jobs in the place I wanted to live: New England.

This was the home of my family, my friends, and my boyfriend at the time, who would eventually become my husband. He was also an academic, which complicated things.[3] I spent the summer of 2019 in the Boston area, and I was out for a run with my longtime running friend from graduate school who was now working at a local startup. I was telling her about my discontent and asked her if she had any interesting job leads. From that day, things escalated quickly and before I knew it, I had been offered a job as a research data scientist at a startup working across town.

That's when I did the thing that had previously seemed unthinkable to me, I left my tenure track job to "go into industry." And if you've ever spent even a moment hanging around academics, you know that these scare quotes are warranted. The way academics say this phrase, it has the same intonation as "go off the deep end" or "fall off the face of the earth." It sounds amorphous yet vaguely threatening and hard to come back from.

But folks, I did it. And with that one scary jump, I became a data scientist.

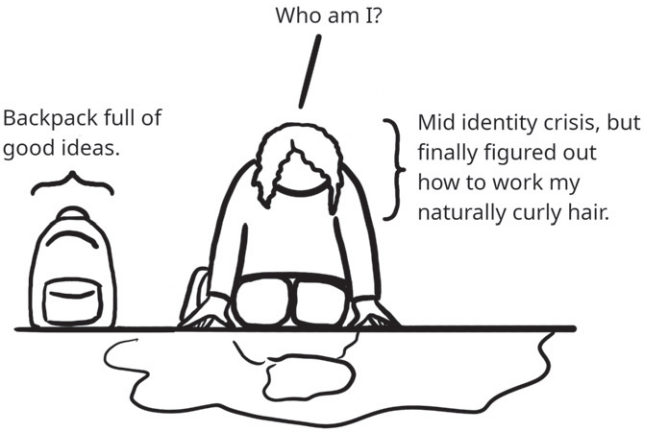
4. "GOING INTO INDUSTRY". I wasn't just any data scientist, I was a data scientist working at a tech startup at the hip and trendy Greentown Labs climatech incubator in a beautifully refurbished warehouse garage. The single-origin coffee flowed like water, the team outings were extravagant, and I think I might have been the oldest person in the building. It was ultra cool.

While I was there I got my first exposure to real industry data science problems. We were a small company that partnered with utility companies to help them understand when their equipment might fail. My favorite project involved offshore wind turbines, and I spent a lot of time learning about tools to model turbine dynamics and loads to predict the overall health of these huge pieces of equipment.

The work was hard, and after spending so much time becoming an academic mathematician and number theorist, it felt like I was starting over from the bottom again. I was overwhelmed with this sensation that I had just spent the last 15 years working toward one very narrow goal of academic success, and after getting so close, I'd just thrown it all away. All of the papers, all of the conferences, all of the professional connections I'd made, suddenly felt like they meant nothing. To say that I had a full-blown crisis of soul and self would be an understatement.

It wasn't that I didn't like the work at the startup; the technical details of it were genuinely exciting to me. Having spent so many years immersed in the extremely theoretical world of number theory, I impressed myself with how quickly I was able to pick up concepts in statistical machine learning. Having taught so much calculus and linear algebra I knew those concepts inside and out, and the basic statistics I picked up in my evening MBA courses turned out to be a lifeline. Not only that, but suddenly all those nights spent honing my Python skills were paying off. I was actually building software and tools with a team of other developers and the fact that I already knew how to "write unit tests" and "open a pull request" made me feel good. My time at NPR had taught me how to distill complicated ideas into bite-sized chunks, and I was able to do that on calls with clients and in our company blog. But I just couldn't get rid of this sad feeling deep in my core that I'd accidentally jumped into someone else's life and

[3]The "two body problem" is a cruel fact of academic life that I think we could fix if we just opened our minds and put our heads together, but that's a rant for another day.



it didn't feel right. Not only that, but if my math before felt disconnected from social good, working at a startup and selling math for profit felt really incongruous with my beliefs.

At that point I knew that I needed to think long and hard about my next move. Here's the part of the story where things double-back on themselves, so pay close attention. Remember how I started a podcast with my best friend? Well, in 2018 I was invited to Tufts University to give a colloquium talk about my experiences in science writing. During my visit I recorded a podcast interview with a math professor at Tufts whose work on wealth redistribution fascinated me. It was a great conversation. Fast forward to 2020, and I saw that Tufts was hiring data scientists in their Data Intensive Studies Center where this same professor is an affiliate faculty member. I emailed him, shared some updates on the various turns my life had taken since we spoke last, and told him about my interest in the position. It was wonderful to reconnect, and after meeting the rest of the team in the center I knew this was the place for me.

After two rounds of interviews and the usual back and forth, I accepted an offer from Tufts, and just like that, I was back in the fold. But this time it was different. Here's where we switch to the present tense. I'm research staff in a center within shouting distance of the math department. Officially, I'm called a "Senior Data Scientist." I'm not tenure-track, which I prefer. It turns out spending year after year on the One True Path wasn't the best for my mental health but I just didn't know a different way to be. This realization has been incredibly powerful for me, and if I may, I'd like you to allow for the possibility that maybe it's not the best for you—or your students—either. I encourage you to expand your mind to all of the different ways to exist inside and outside of academia. There's a whole world of wonderful and fulfilling jobs where you're allowed to leave at any time.

5. THE LIFE OF AN ACADEMIC DATA SCIENTIST. I've noticed that saying I'm a data scientist often raises lots of questions, chief among them: what does that mean? To answer that question, let me tell you more about what I do now. I teach a couple of courses in data science,[4] but mostly I do research. A large part of my work still deals with offshore wind turbines and beginning in 2023 my work is funded by the National Science Foundation. I collaborate with civil engineers who can model all of the complicated structures and strains, electrical engineers who extract data using re-

[4]Since this is the MAA *Monthly* and there are bound to be some readers who are interested in the pedagogy of data science, I coauthored a paper that includes a set of materials for a course in data science that I teach to our Masters in Law and Diplomacy students [**12**].

mote sensing and signal processing, public health researchers who study occupational hazards in the US energy industry, and policy experts who understand the stakeholders and decision makers in wind energy. I build lots of models and write optimization algorithms to do things like estimate turbine dynamics using hidden Markov models [13], and understand multi-stakeholder decision making using agent-based models. I get to work with students, postdocs, faculty, and industry partners and we all share the load in this important work.

I also do other research that has nothing to do with wind. I co-direct the highly interdisciplinary Small Town Police Accountability Research Lab where we develop tools to process policing data obtained by FOIA requests and community impacts of police behavior [14, 15]. I do work at the intersection of mathematics and political science to understand how individual choices shape population level outcomes in things like vaccine campaigns and elections [16, 17]. The common thread to all of my work is that it's backed by lots of data and it uses computational and statistical tools to explore the ways we can make a safer, more equitable, and more sustainable world.

The best thing about being an academic data scientist is getting to work with people from so many disciplines. This means that I'm always hearing new and surprising (to me) perspectives on things, but it also just means that I get to meet more kinds of people. I'm still doing plenty of math, but I don't feel much like a mathematician anymore. I rarely think about those old guys from antiquity anymore, which is just as well because they probably wouldn't have liked me very much. But perhaps best of all, I've finally found a way to use the mathematical tools that I spent so long accumulating to work on the important social issues that drive me. So to answer the question of what is data science: it's what sits at the intersection of mathematics, computer science, and domain expertise. The mathematics usually looks a bit more like statistics, the computer science usually involves Python or R, and the domain expertise can come from anywhere.

6. LESSONS LEARNED. Before I go, there are two pieces of advice that I want to share with you. The first is that it's always useful to go places and talk to people. Networking has a lot of negative baggage and it can also be uncomfortable if you're a shy or quiet person (this is where I disclose to you that I am an extrovert and always kind of enjoy chatting up strangers, in case that wasn't already abundantly clear). So to help you on your way, I'm going to share my secret no-fail trick to make those conference coffee hours and networking sessions less excruciating. When you don't have anyone to talk to and the awkward feels start to creep in, find a place toward the edge of the room, stand there, put your phone/laptop away, look forward, and smile. That's it. Just stand there and look friendly and alone. If you start slowly counting, then by the time you get to thirty some new friend will already be talking to you. Not only does this make events a little more fun, but knowing people and having people know you is incredibly useful. Not in the creepy opportunistic sense, but just in the sense that a larger network just opens your world up a bit.

The second bit of advice is to waste lots of time. Find a hobby or two that interests you, and don't be afraid to devote yourself to it for long luxurious stretches. Whether that involves developing your creative abilities or physical endurance, or spending time

with people who move in vastly different circles, this will help you find perspective on your work. Sometimes it also happens that your strange time-wasting hobby is actually preparing you for a future you can't quite conceive of just yet.

ACKNOWLEDGMENT. The authors wishes to thank the anonymous referees and the Editorial Board for thoughtful remarks that improved this manuscript.

DISCLOSURE STATEMENT. No potential conflict of interest was reported by the author.

ORCID
Anna Haensch http://orcid.org/0000-0002-2299-2106

REFERENCES

[1] Gauss CF. Disquisitiones arithmeticae. New Haven (CT): Yale University Press; 1966.
[2] Guy RK. Every number is expressible as the sum of how many polygonal numbers? Amer Math Monthly. 1994;101(2):169–172.
[3] Hilbert D. Mathematical problems. Bull Amer Math Soc. 2000;37(4):407–436.
[4] Davis M. Hilbert's tenth problem is unsolvable. Amer Math Monthly. 1973;80(3):233–269.
[5] Haensch A. My Summer at NPR. Not AMS. 2013;60(11):1477–1478.
[6] Haensch A. Harmful parasites in cat poop are widespread. Shots Health News from National Public Radio. 2013 July 9. Available from: https://www.npr.org/sections/health-shots/2013/07/05/199041322/harmful-paraites-in-cat-poop-are-widespread
[7] Silver N. The real story of 2016. FiveTirtyEight, 2017 Jan 19. Available from: https://fivethirtyeight.com/features/the-real-story-of-2016/
[8] Tavernise S, Feldman N, Tan S, Johnson MS, Willens P, Georges M, Lozano M, Powell D, Wood C. Why fewer American children are living in poverty. The New York Times, 2022 Sept 26. Available from: https://www.nytimes.com/2022/09/26/podcasts/the-daily/us-child-poverty-decline.html
[9] O'Neil C. Weapons of math destruction. Harlow, England: Penguin Books; 2017.
[10] Noble S. Algorithms of oppression: how search engines reinforce racism. New York (NY): New York University Press; 2018.
[11] Haensch A. The blog on math blogs. Not AMS. 2016;63(6):643–644.
[12] Haensch A, Knudson K. Python for global applications: teaching scientific Python in context to law and diplomacy students. In: Proceedings of the 21st Python in Science Conference (SciPy 2022); 2022.
[13] Haensch A, Tronci EM, Moynihan B, Moaveni B. Regularized hidden Markov modeling with applications to wind speed predictions in offshore wind. Mech Syst Signal Process. 2024;211:111229. doi: 10.1016/j.ymssp.2024.111229
[14] Haensch A, Gordon D, Knudson K, Cheng J. A multi-method data science pipeline for analyzing police service in the presence of misconduct. Am Stat. 2024 [cited 2024 Aug 26]; [18 p.]. doi: 10.1080/00031305.2024.2374275
[15] Kelling C, Haensch A, Mendible A, Brooks S, Wiedemann A, Aminian M, Hasty W, Higdon J. Data collection and analysis for small-town policing: challenges and recommendations. Stat Public policy. 2024;11(1):2379270. doi: 10.1080/2330443x.2024.2379270.
[16] Börgers C, Boghosian B, Dragovic N, Haensch A. A blue sky bifurcation in the dynamics of political candidates. Amer Math Monthly. 2023;131(3);225–238. doi: 10.1080/00029890.2023.2286221
[17] Haensch A, Dragovic N, Börgers C, Boghosian B. A geospatial bounded confidence model including mega-influencers with an application to Covid-19 Vaccine hesitancy. J Artif Soc Soc Simulat. 2023;26(1):1–8.
[18] Haensch A. A characterization of almost universal ternary inhomogeneous quadratic polynomials with conductor 2. J Number Theor. 2015;156:247–262.
[19] The Sage Developers. (2023). SageMath, the Sage Mathematics Software System (Version 10.0). Available from: https://www.sagemath.org

ANNA HAENSCH is a senior data scientist in the Tufts University Data Intensive Studies Center with a secondary appointment in the Department of Mathematics. She earned a Ph.D. in mathematics from Wesleyan University. Her research lies at the intersection of mathematics and the social sciences and it deals with the many ways that we can use data to help make a safer, more sustainable, and more equitable world. She currently serves as the 2024-2025 AAAS Legislative Fellow for Artificial Intelligence in the United States Senate.
Data Intensive Studies Center, Tufts University, Medford, MA 02155
anna.haensch@tufts.edu