# Course Reminders

- Final Project due Wed, June 9th (11:59 PM)
  - Report (GitHub)
  - Video (one person for group submits on Canvas)
  - Team Evaluation Survey: link to survey (link also on Canvas; required)
- Post COGS 108 Survey: link to survey (link also on Canvas; *optional* for EC)
- SETs (Complete by Sat 3/16 by 8:00AM)
- Please do not email me about rounding your grade up

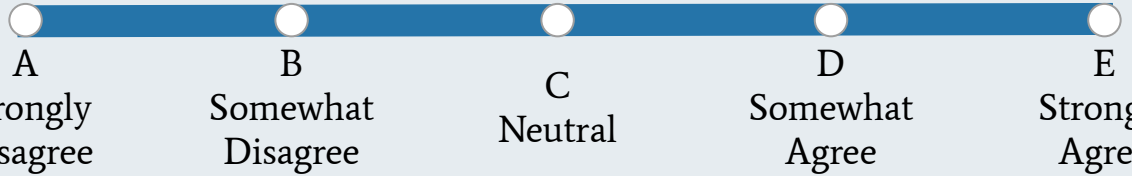
# Jobs and the current state of Data Science

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

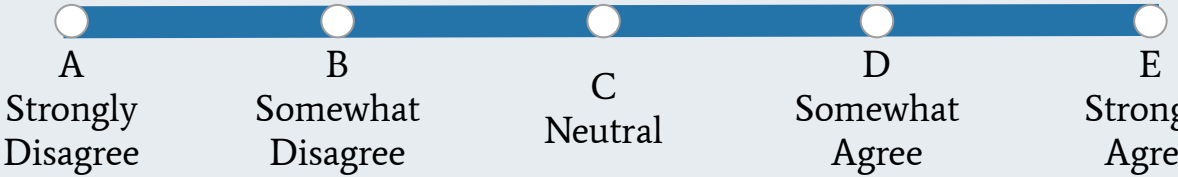
# Career Certainty

I know what career I want to pursue after graduation.



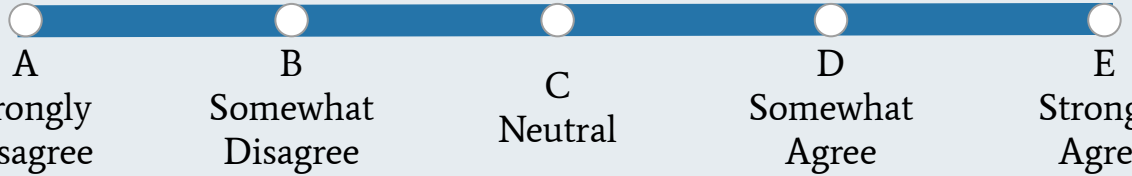
# Data Science Careers

I want to pursue a career in data science.



# Path Certainty

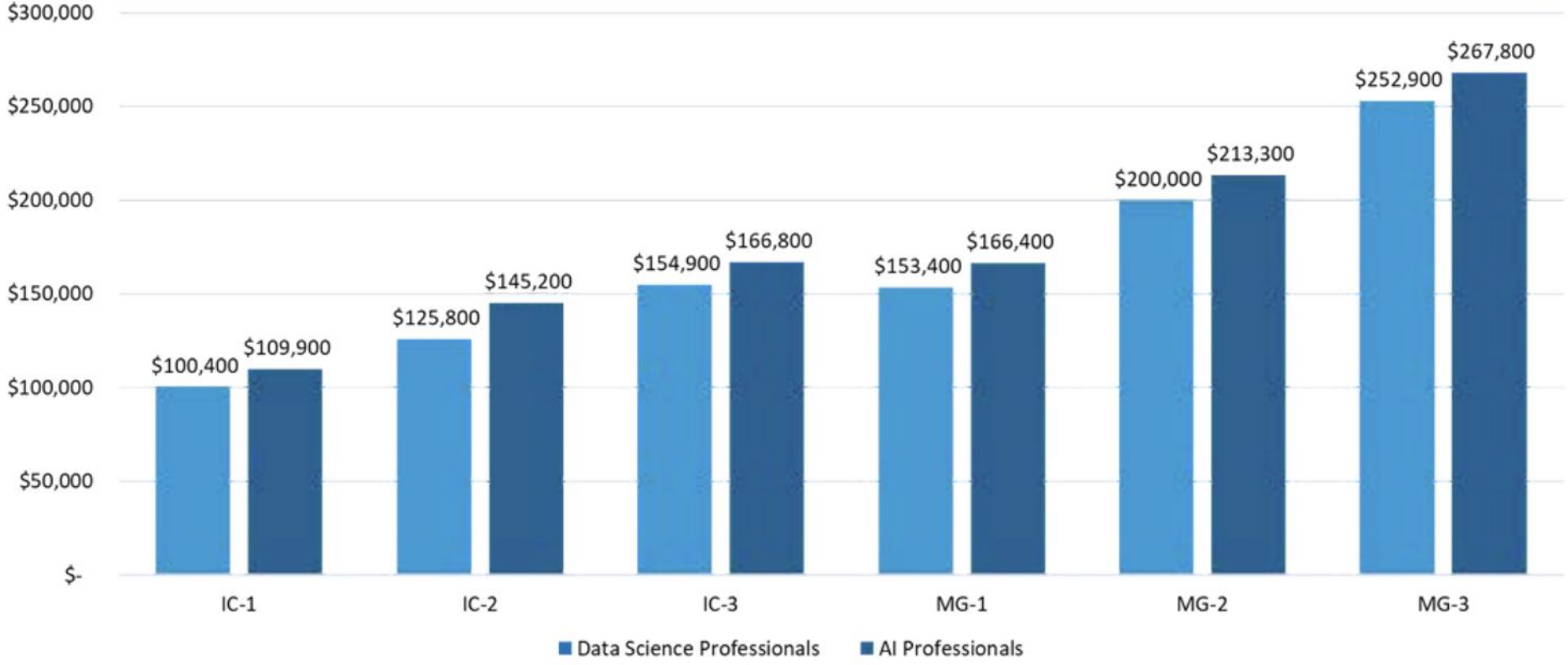
I understand the steps necessary to get a job in data science.



| | Job Title | Median Base Salary | Job Satisfaction | Job Openings |
|---|---|---|---|---|
| #1 | Enterprise Architect | $144,997 | 4.1/5 | 14,021 |
| #2 | Full Stack Engineer | $101,794 | 4.3/5 | 11,252 |
| #3 | Data Scientist | $120,000 | 4.1/5 | 10,071 |

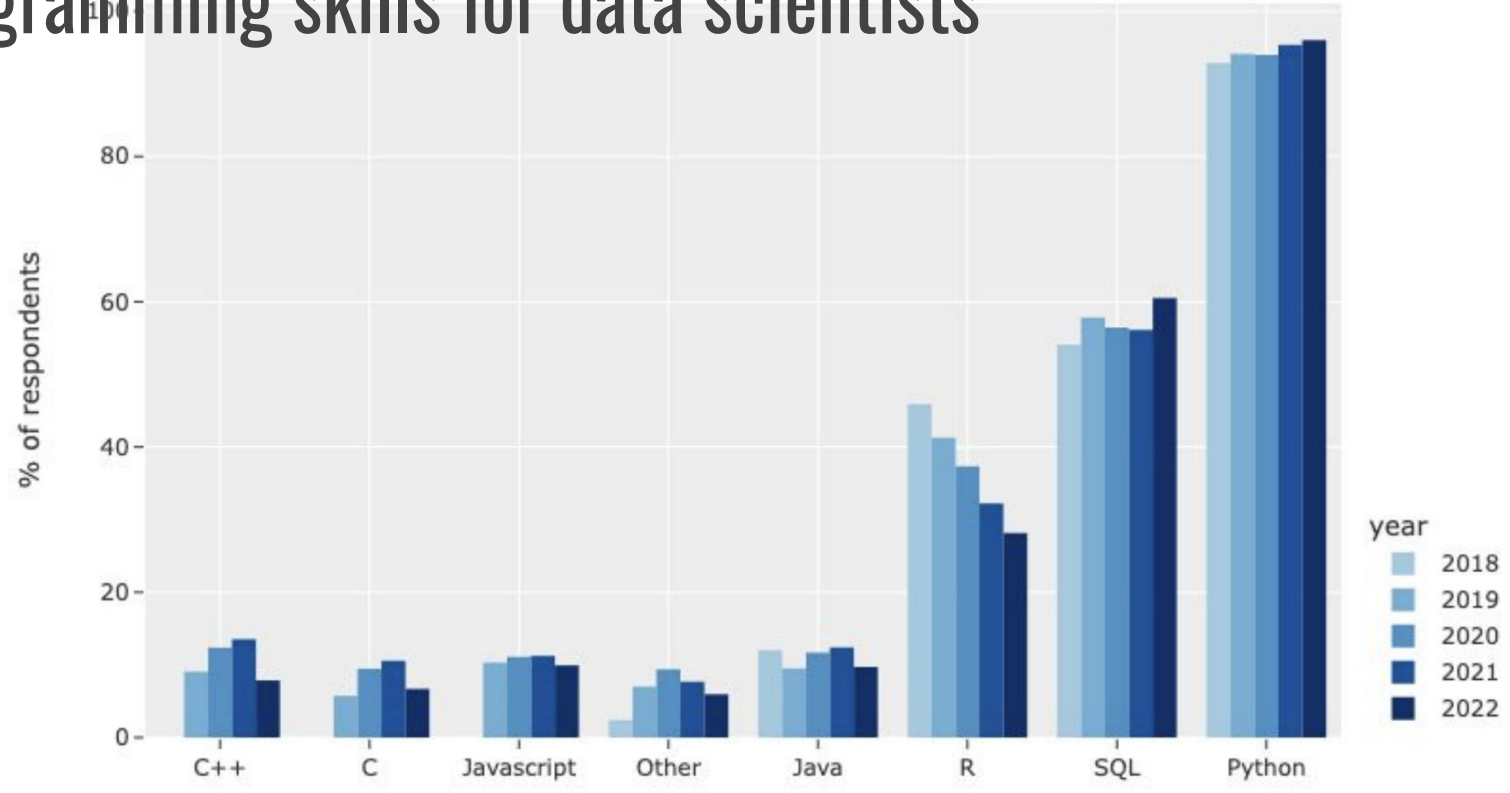
Source: Glassdoor (2021)

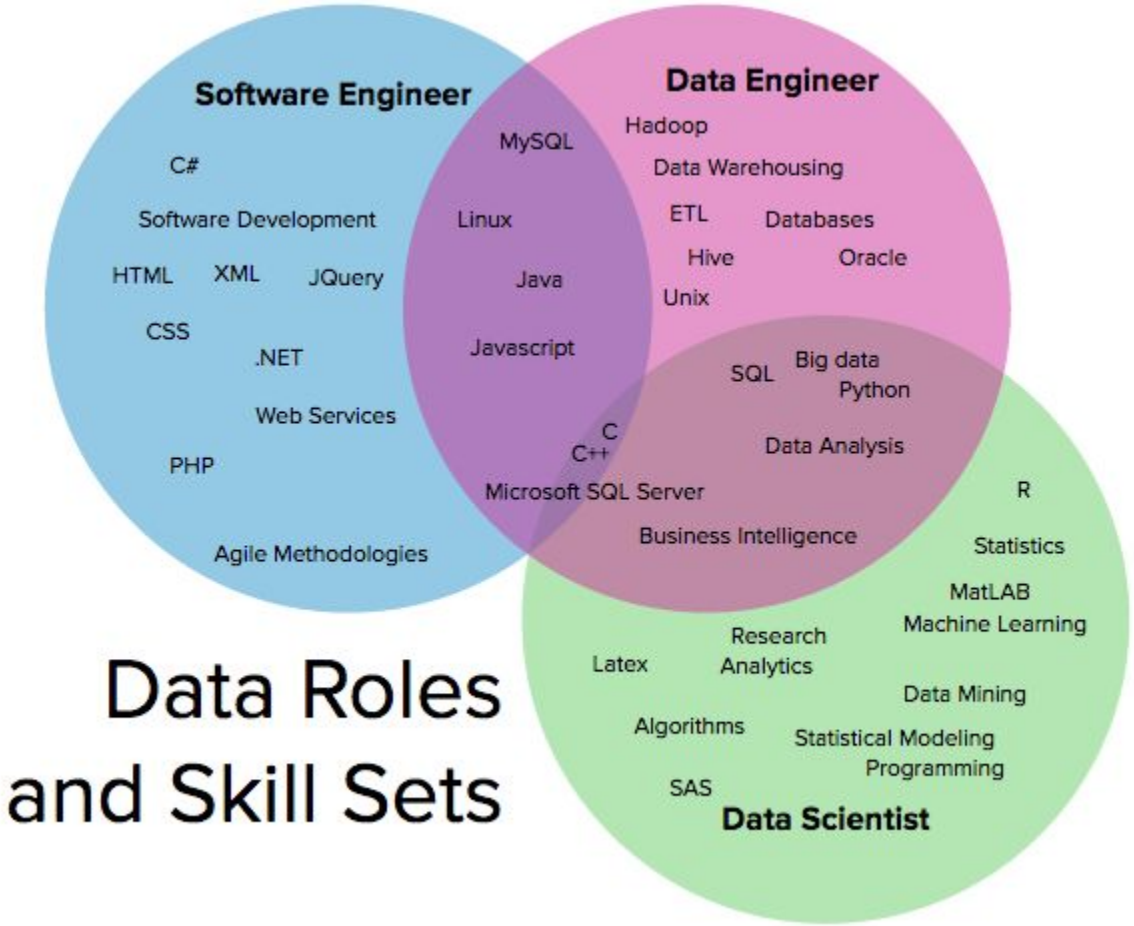
# 2023 Mean Base Salaries



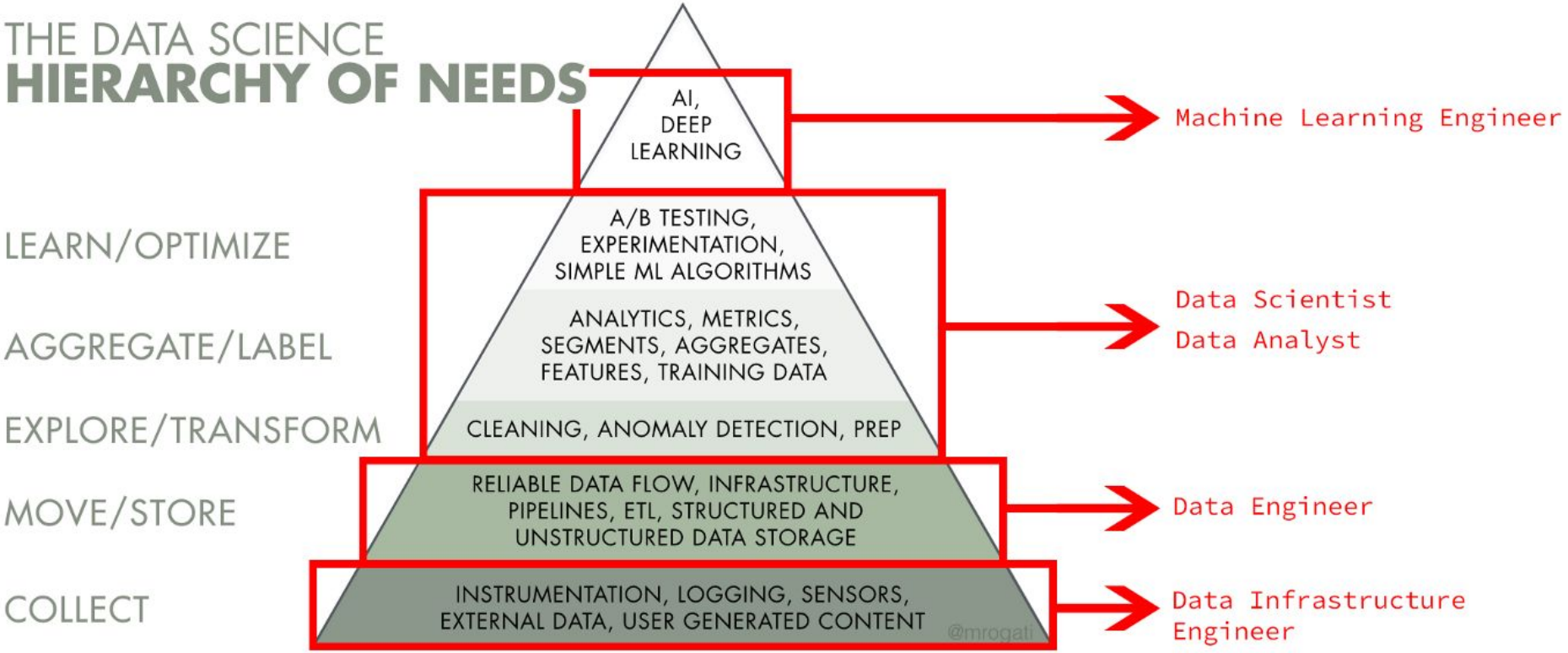

BURCH DATA SCIENCE & AI PROFESSIONALS SALARY REPORT Report (2023; N=1,331 data scientists & 506 AI professionals)

# Python & SQL are consistently the two most common programming skills for data scientists






https://towardsdatascience.com/the-dynamics-of-data-roles-teams-6c450b27e59e




https://towardsdatascience.com/data-engineer-vs-data-scientist-bc8dab5ac124

# Some DS-related job titles

- Analytics or statistician: data handling, analysis
- Data Journalist: programming, viz, communication
- Data analyst: programming, data handling, reporting/metric tracking, viz
- Data scientist: programming, data handling, analysis
- Data engineer: programming, databases, management
- Data architect: programming, databases, design
- Data manager: databases, design, management
- *OPs (eg, devOPs, dataOPs, full stack): programming, tool development, mangagement concentrating on end to end process
- ML Engineer: programming, tool development, management of infrastructure
- ML researcher: programming, algorithm design and testing

## Glut of new data scientists

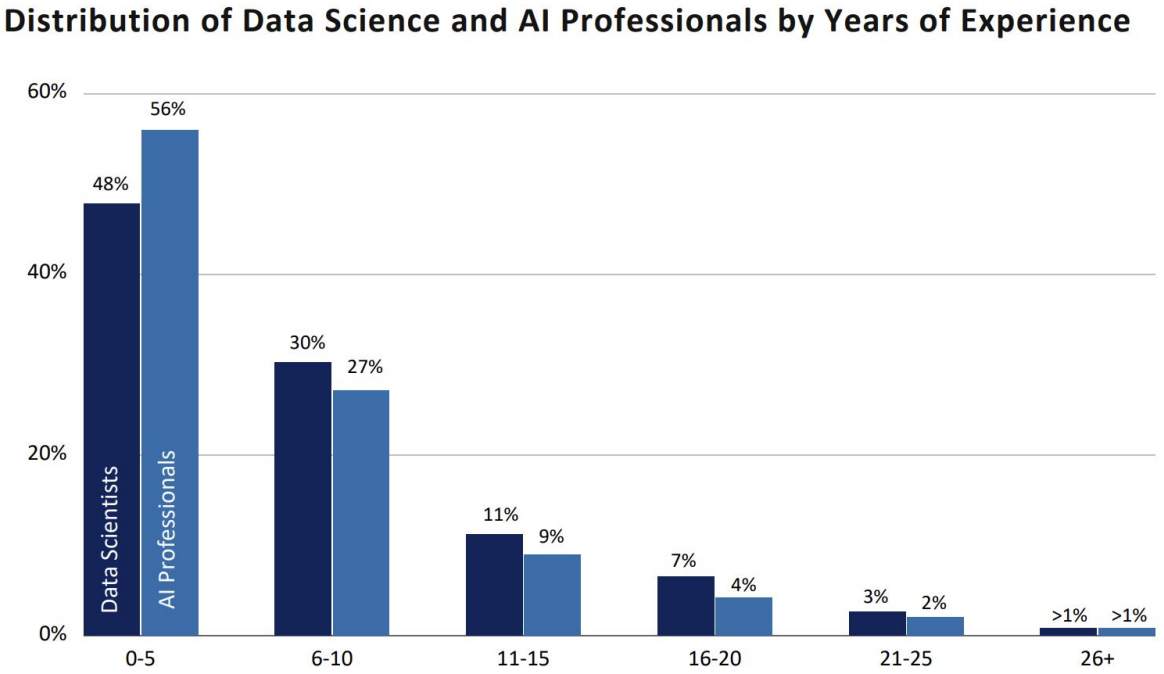
First, let's talk about the oversupply of junior data scientists. The continuing media hype cycle around data science has enormously exploded the amount of junior talent available on the market over the past five years.

This is purely anecdotal evidence, so take it with a large grain of salt. But, based on my own participation as a resume screener, mentor to data scientists leaving boot camps, interviewer, interviewee, and from conversations with friends and colleagues in similar positions, I've developed an intuition that the number of candidates per any given data science position, particularly at the entry level, has grown from 20 or so per slot, to 100 or more. I was talking to a friend recently who had to go through 500 resumes for a single opening.

This is not abnormal. More anecdotal evidence comes from job openings like this one, from machine learning's godfather, Andrew Ng, whose AI startup demanded 70-80 hours a week. He was flooded with applications, after blithely noting that previously many people had tried to volunteer for free. As of this latest writing, they ran out of space in their current office.

It's very, very hard to estimate the true gap between market demand and supply, but here's a starting point.

Half of




Data Science Professionals

**Median-** **6 years**

**Mean-** **7.4 years**

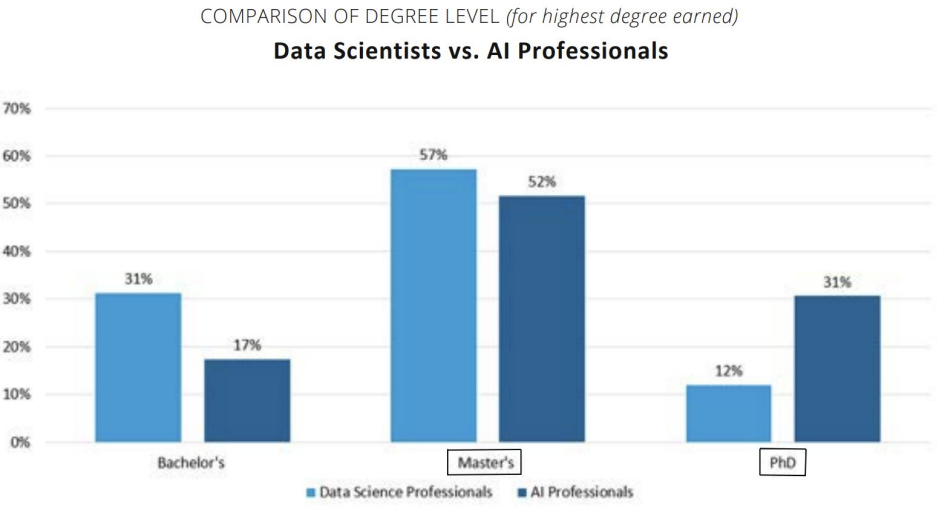
AI Professionals

**Median-** **5 years**

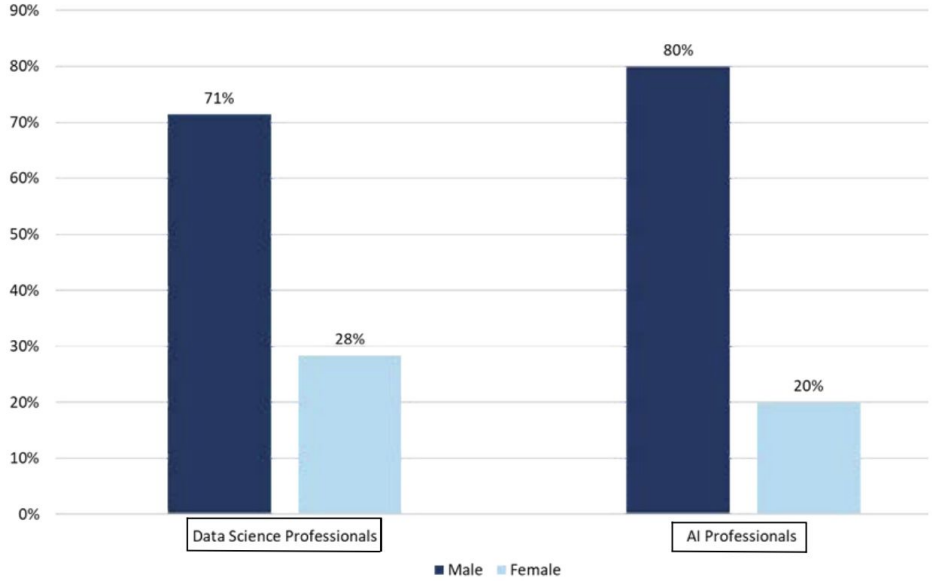
**Mean-** **6.5 years**

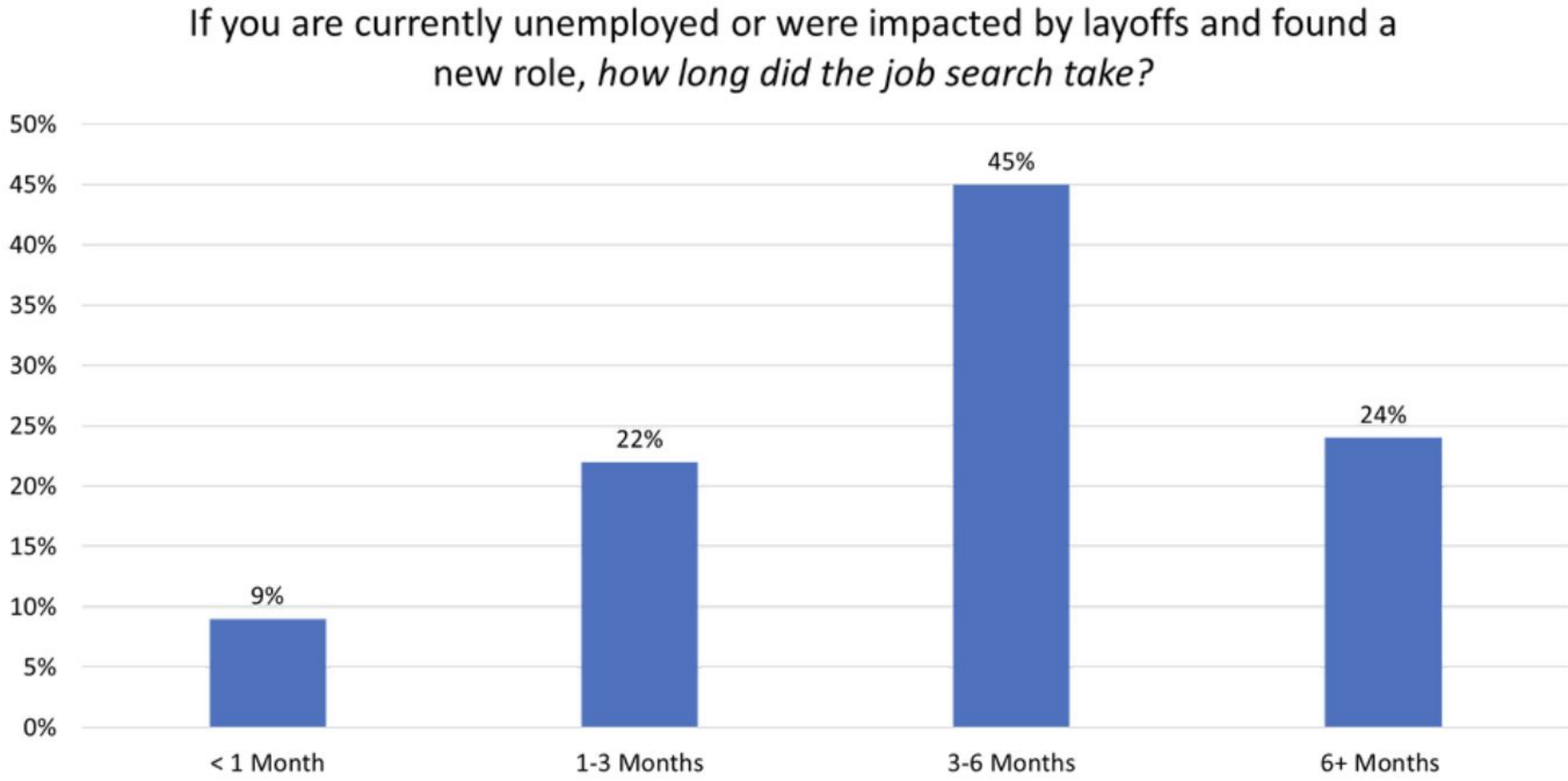
BURCH DATA SCIENCE & AI PROFESSIONALS SALARY REPORT Report (2023; N=1,331 data scientists & 506 AI professionals)

# Masters are most common degree




# Consistent gender imbalance across the field






BURCH DATA SCIENCE & AI PROFESSIONALS SALARY REPORT Report (2023; N=1,331 data scientists & 506 AI professionals)

# Build A Career In Data Science (2020)

## Part I: Getting Started With Data Science

- What is Data Science?
- Data Science Companies
- Getting the Skills
- Building a Portfolio

## Part II: Finding Your Data Science Job

- The Search: Identifying the Right Job for You
- The Application: Resumes and Cover Letters
- The Interview: What to Expect and How to Handle It
- The Offer: Knowing What to Accept

## Part III: Settling Into Data Science

- The First Months on the Job
- Making an Effective Analysis
- Deploying a model into production
- Working with Stakeholders

## Part IV: Growing In Your Data Science Role

- When your Data Science Project Fails
- Joining the Data Science Community
- Leaving Your Job Gracefully
- Moving up the Ladder

# 1. The Search: The Right Job

Know where to start:

- LinkedIn, Indeed, and Glassdoor
- POCIT and Tech Ladies, for people of color and women in technology respectively
- job boards for specific types of companies like startups (AngelList) and technology (Dice)
- specific company's careers page
- X (Twitter) & Meetups

Watch for red flags:

- No description of company or job
- A job that is actually three jobs

Demand Fit, not Perfection

# Advice for new grads

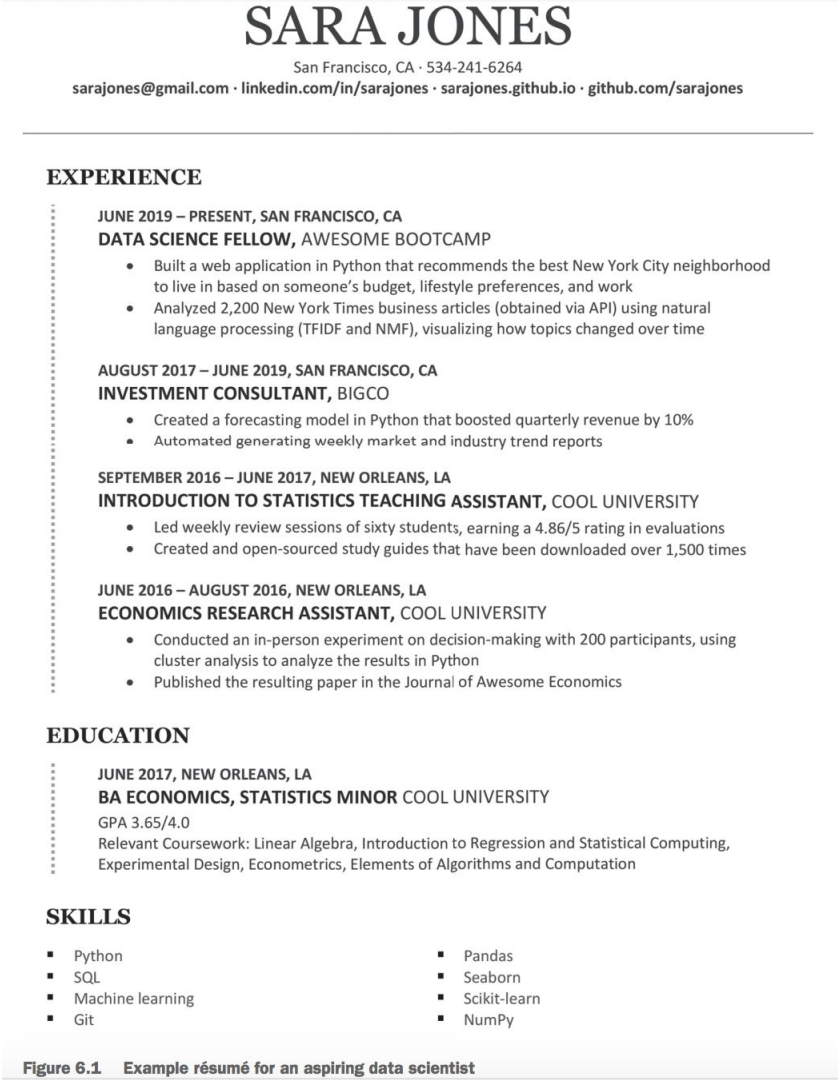
*If you're about to or just graduated college, **your most relevant skill is your education**. Your **data science portfolio** will be helpful here too. When you're searching for jobs, look for positions specifically titled **"New Grad," "Junior," "Associate," and "Entry-level."** Also look at your career center for help and go to any job fairs that happen on campus. Internships are relevant – less for what work skills you learned and more that it shows you can come to an office each day, be professional and productive.*

# On Job Descriptions

*The first thing to keep in mind is that **job descriptions are generally wish lists** with some flexibility. If you meet 60% of the requirements (e.g. you're a year short of their required work experience or haven't worked with one component of their tech stack), but are otherwise a good fit, you should still apply.*

# 2. The Application

- Resume *and* cover letter should be compelling
  - Resume:
    - goal is to get you an interview, not a job
    - Better be skimmable
    - Includes: contact info, education, experience, and skills



SARA JONES

San Francisco, CA · 534-241-6264

sarajones@gmail.com · linkedin.com/in/sarajones · sarajones.github.io · github.com/sarajones

**EXPERIENCE**

JUNE 2019 – PRESENT, SAN FRANCISCO, CA

**DATA SCIENCE FELLOW,** AWESOME BOOTCAMP

- Built a web application in Python that recommends the best New York City neighborhood to live in based on someone's budget, lifestyle preferences, and work
- Analyzed 2,200 New York Times business articles (obtained via API) using natural language processing (TFIDF and NMF), visualizing how topics changed over time

AUGUST 2017 – JUNE 2019, SAN FRANCISCO, CA

**INVESTMENT CONSULTANT,** BIGCO

- Created a forecasting model in Python that boosted quarterly revenue by 10%
- Automated generating weekly market and industry trend reports

SEPTEMBER 2016 – JUNE 2017, NEW ORLEANS, LA

**INTRODUCTION TO STATISTICS TEACHING ASSISTANT,** COOL UNIVERSITY

- Led weekly review sessions of sixty students, earning a 4.86/5 rating in evaluations
- Created and open-sourced study guides that have been downloaded over 1,500 times

JUNE 2016 – AUGUST 2016, NEW ORLEANS, LA

**ECONOMICS RESEARCH ASSISTANT,** COOL UNIVERSITY

- Conducted an in-person experiment on decision-making with 200 participants, using cluster analysis to analyze the results in Python
- Published the resulting paper in the Journal of Awesome Economics

**EDUCATION**

JUNE 2017, NEW ORLEANS, LA

**BA ECONOMICS, STATISTICS MINOR** COOL UNIVERSITY

GPA 3.65/4.0

Relevant Coursework: Linear Algebra, Introduction to Regression and Statistical Computing, Experimental Design, Econometrics, Elements of Algorithms and Computation

**SKILLS**

- Python
- SQL
- Machine learning
- Git
- Pandas
- Seaborn
- Scikit-learn
- NumPy

**Figure 6.1 Example résumé for an aspiring data scientist**

# 2. The Application

- Resume *and* cover letter should be compelling
  - Resume:
    - goal is to get you an interview, not a job
    - Better be skimmable
    - Includes: contact info, education, experience, and skills
  - Cover Letter
    - Should highlight both why you want *this* job and why *you* are a particularly good fit
    - Demonstrate your research/knowledge about the company and position
    - Tailor these for each job
    - Allow them to be machine-searchable
    - Referrals are a way to back-door past the algorithms
      - If contacting someone (LinkedIn, Twitter), give them a reason to read your message

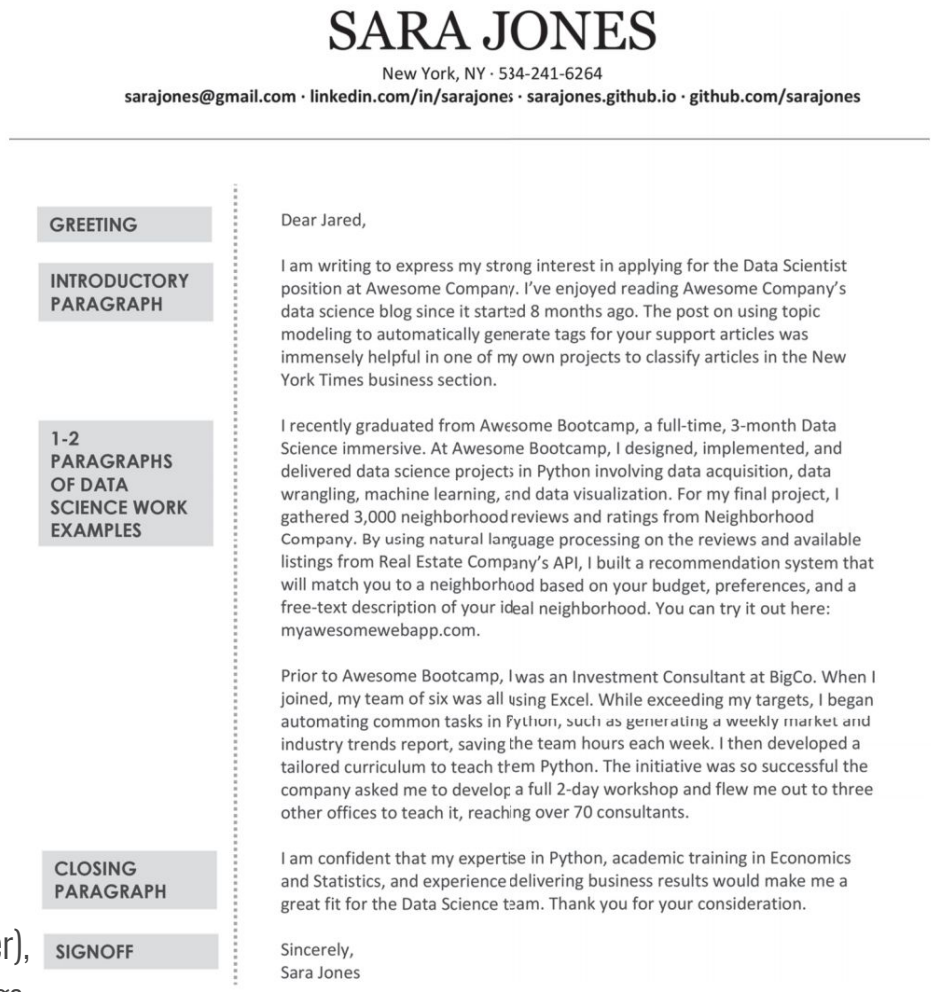
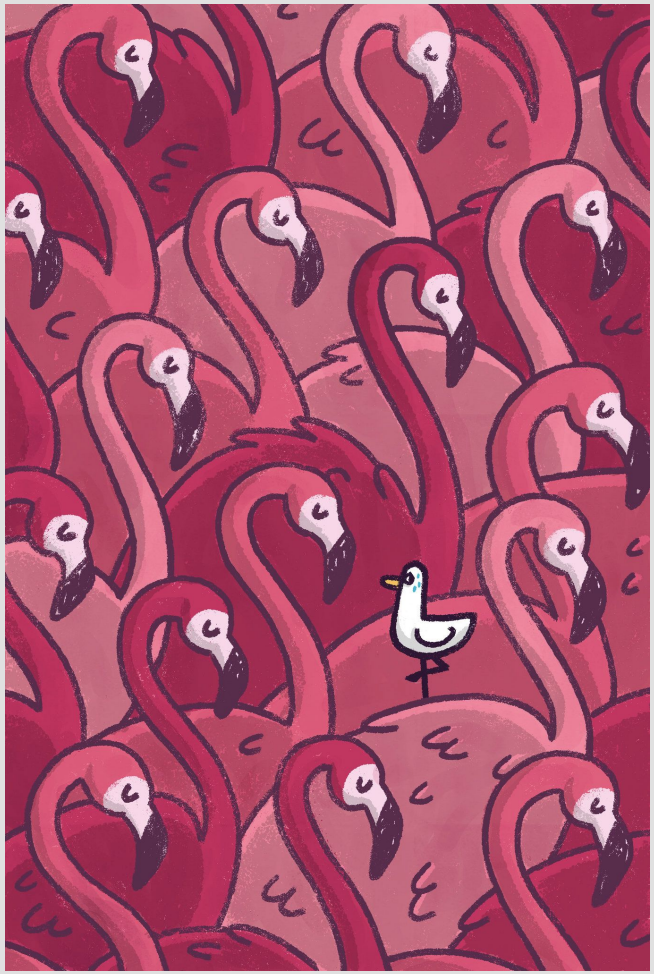



**Figure 6.2 An example cover letter with highlights showing the different components**

# How are you going to stand out in the crowd?



adapted from Sean Kross

# 3. The Interview

Basic understanding of you/position; assessment of fit

Are you able to do the job? Are you a good fit?

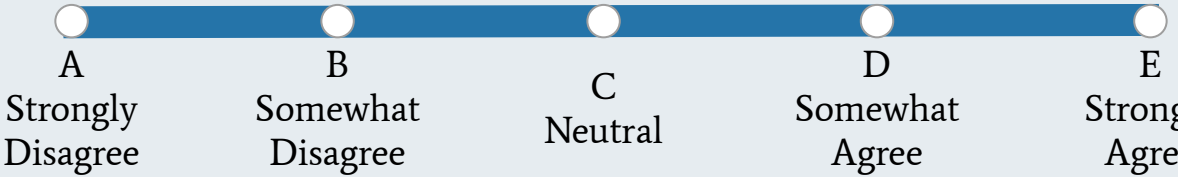
Take-home assignment to determine your problem-solving and technical skills

Tie up loose ends, presentation,

| | |
|---|---|
|  | 1. Phone interview |
|  | 2. In-person interview |
|  | 3. Case study |
|  | 4. Leadership interview and offer |

# Negotiations

I understand the process of negotiating a job offer.



# 4. The Offer

1. Offer is coming - general outline of offer coming your way
2. Company makes an offer - often in email; get it in writing; includes salary, start date,
3. You respond - thank them for offer and let them know you're excited to look it over in detail
4. You negotiate - lay out what you want/need to accept the offer; best for you
    a. What is negotiable? Salary (5%), start date, vacation, flexibility, earlier review (earlier raise), educational benefits, budget for travel/conferences, benefits (less often), options
    b. Best leverage: a competing offer
5. You decide - communicate final decision

# Minimal Advice: Getting Your First Job in Data Science (Summary)

- Learn one programming language extraordinarily well (Python, R).
- Learn SQL well.
- Learn how to set up and interact with cloud computing services.
- Know how to *think* and *communicate* about data
- Create a resume and have a few people with relevant knowledge help you revise it.
- Establish a professional web presence.
- Be prepared to apply to many dozens of jobs.

# What you all have done...

# COGS 108: What we've learned

01: Data Science Python, & Version Control
02: Version Control, Data Intuition,
03: Data wrangling w/ pandas & Ethics
04: Formulating Data Science ?s & Dataviz
05: Data Analysis: Descriptive & EDA
06: Inference
07: Text Analysis
08:  Machine Learning
09: Geospatial Analysis
10: DS Communication &  DS Jobs

Guest Lectures: Brendan Tomoschuk & Michael Baluja

# COGS 108: Final Project Lessons

1. Asking the right question up front really helps
2. Finding the data you need is a skill
    a. ...so is knowing if the data are reliable
    b. ...and if they can answer your question
    c. ....and recognizing what information you don't have
3. Data Visualization and storytelling are important skills.
4. Determining which analytical approach is best is HARD.
5. Programming is merely a piece of the puzzle for data scientists.

# ...so where are we going?

- **Data aren't going anywhere |** continued utilization (exploitation?) of data across products, apps, & platforms; remember: if you don't pay for the product, you and your data are the payment
- **(continued) Rapid development in AI & ML |** expansion beyond text, decrease in hallucinations, ability to cite sources well, hopefully increased interpretability and importance placed on ethics
- **Automation beyond now |** self-driving cars, more targeted ads, self-serve all the things
- **Increased regulation |** I don't see it staying how it is much longer, but Congress is slow

# COGS 108 Thank yous!

TAs: Matthew Feigelis, Samyak Mehta, Yueyan Tang, Yueying Dong, Kunal Rustagi, James Michaelov

IAs: Eric Song, Shenova Davis, Nate del Rosario, Siddhant (Sid) Joshi, Devin Chen, Zhicheng Wang, Steven Xie

All of you for your patience, feedback, and time!

# You all are the future of data science!

So, if you remember anything from this course...


Ethics should always be a priority in your work.


Data wrangling is a puzzle and a big part of the job. When done well, it's not boring!


Data science is a competitive, but rewarding field. You have a chance to make a big difference!


Your grade in this course is probably *not* predictive of future success.

My hope is that all of you go on to (continue to) be good people who are happy, balanced, curious, engaged with the world, fulfilled & successful

# Thanks for taking COGS 108!