Course Reminders

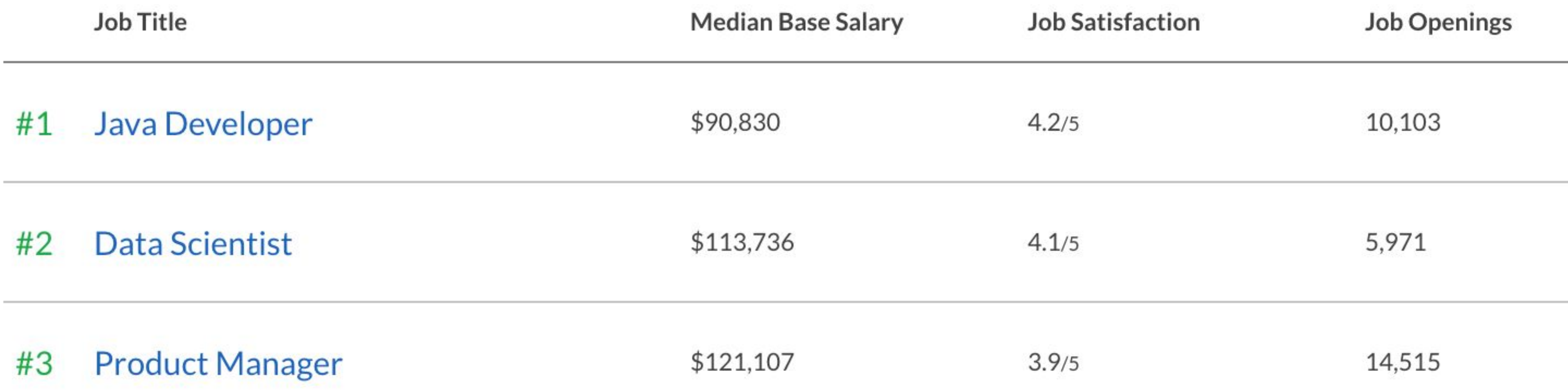
- Final Project due Wed, June 9th (11:59 PM)
  - Report (GitHub)
  - Video (one person per group submits on Canvas)
  - Team Evaluation Survey: link to survey (link also on Canvas)
- Post COGS 108 Survey: link to survey (link also on Canvas; *optional* for EC)
- CAPEs: http://www.cape.ucsd.edu/ (42%)

The Future of Data Science

Shannon E. Ellis, Ph.D
UC San Diego

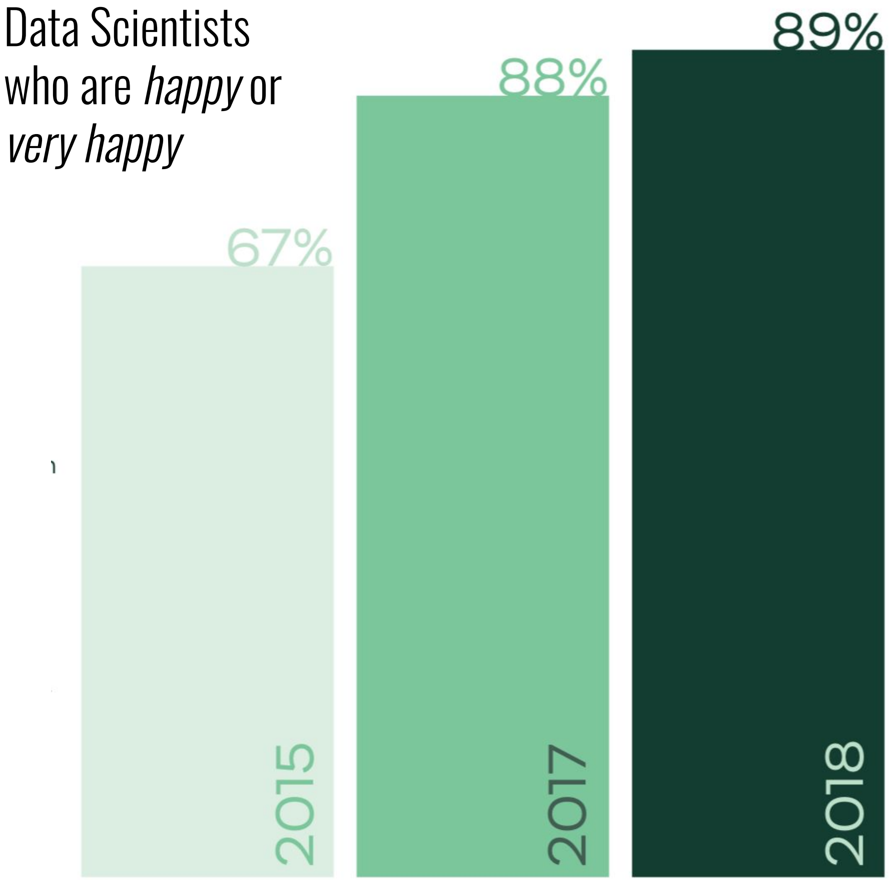
Department of Cognitive Science
sellis@ucsd.edu

Where we are now



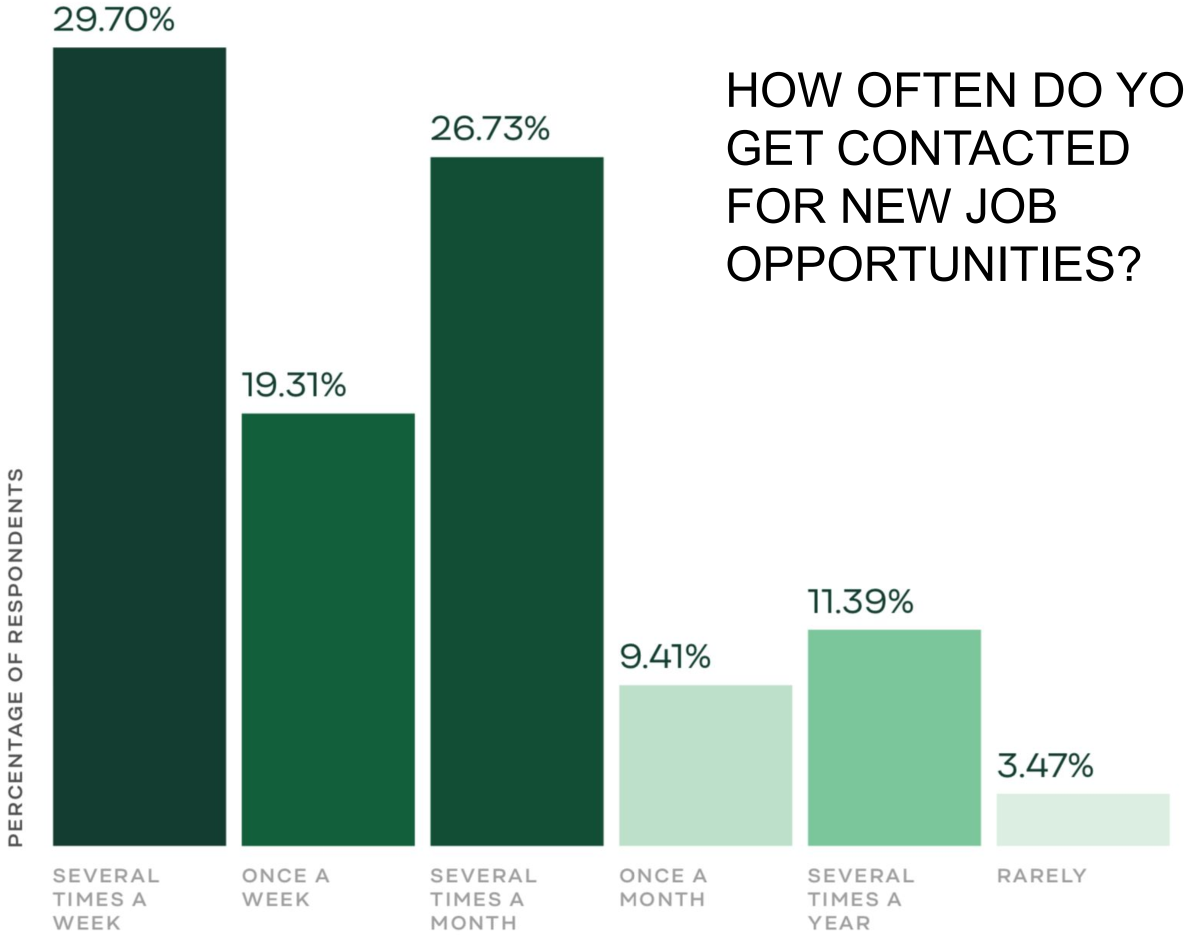
	Job Title	Median Base Salary	Job Satisfaction	Job Openings
#1	Java Developer	$90,830	4.2/5	10,103
#2	Data Scientist	$113,736	4.1/5	5,971
#3	Product Manager	$121,107	3.9/5	14,515

Source: Glassdoor (2021)

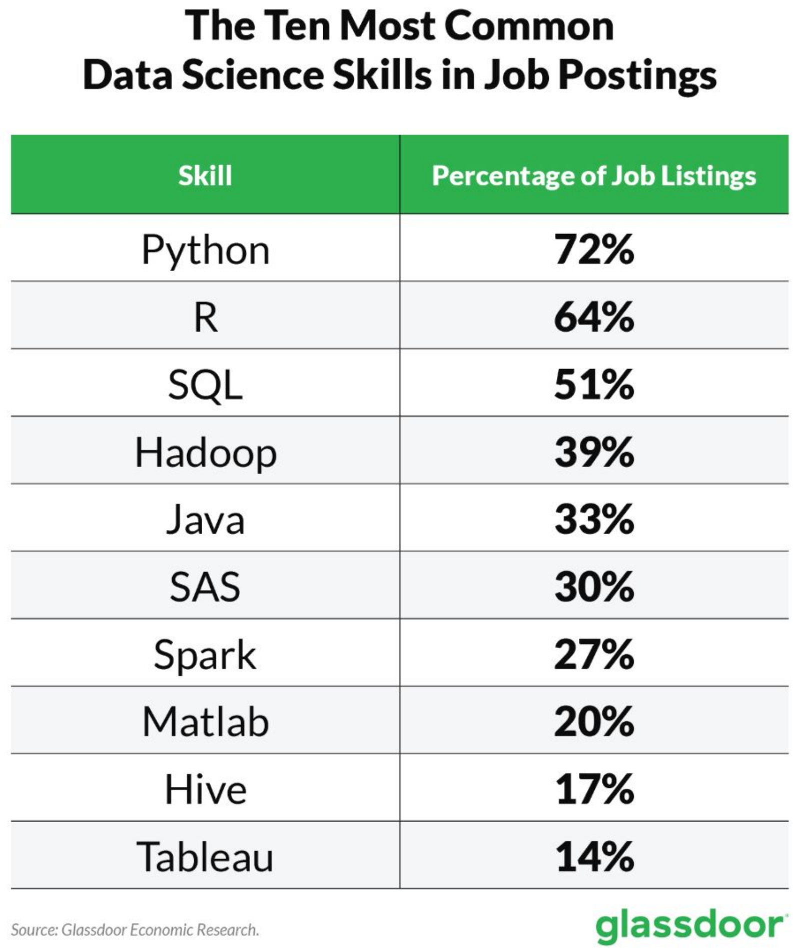


Source: Figure Eight, *Data Science Report* (2018)

HOW OFTEN DO YOU GET CONTACTED FOR NEW JOB OPPORTUNITIES?



Source: Figure Eight, *Data Science Report* (2018)

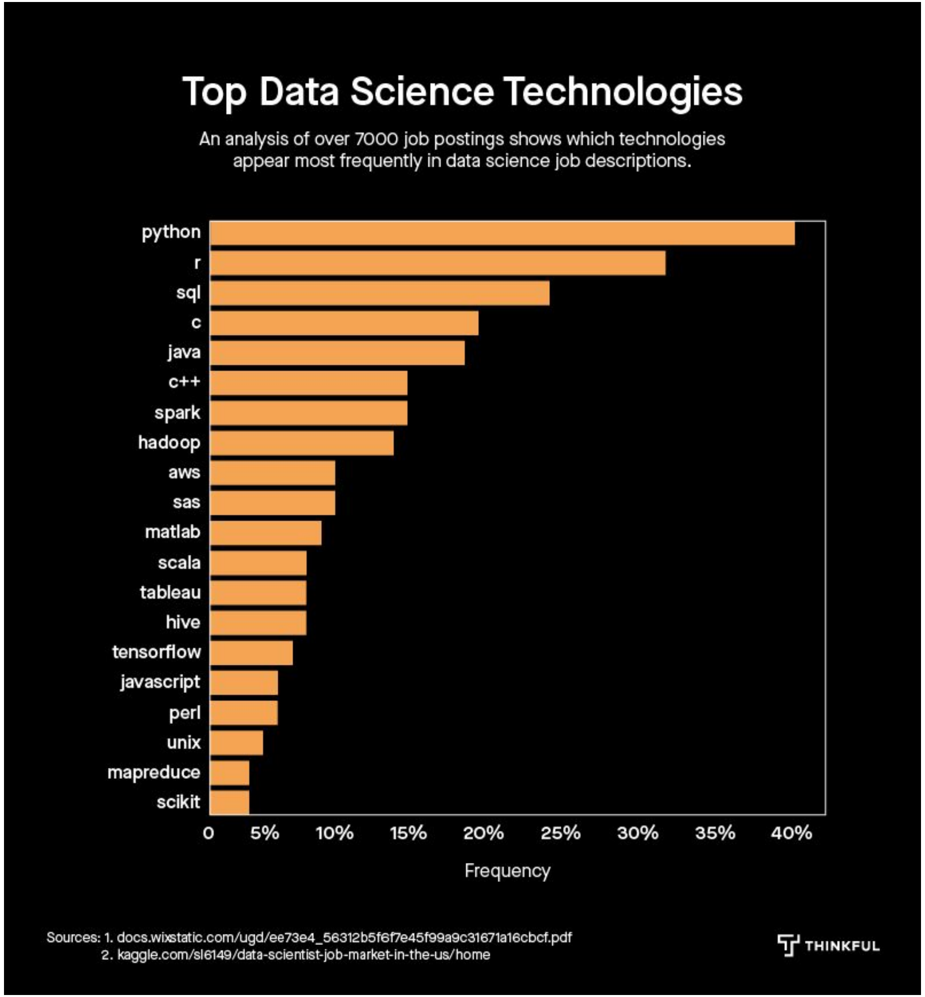


The Ten Most Common Data Science Skills in Job Postings

Skill	Percentage of Job Listings
Python	**72%**
R	**64%**
SQL	**51%**
Hadoop	**39%**
Java	**33%**
SAS	**30%**
Spark	**27%**
Matlab	**20%**
Hive	**17%**
Tableau	**14%**

Source: Glassdoor Economic Research.

glassdoor



Top Data Science Technologies

An analysis of over 7000 job postings shows which technologies appear most frequently in data science job descriptions.

Sources: 1. docs.wixstatic.com/ugd/ee73e4_56312b5f6f7e45f99a9c31671a16cbcf.pdf
2. kaggle.com/sl6149/data-scientist-job-market-in-the-us/home

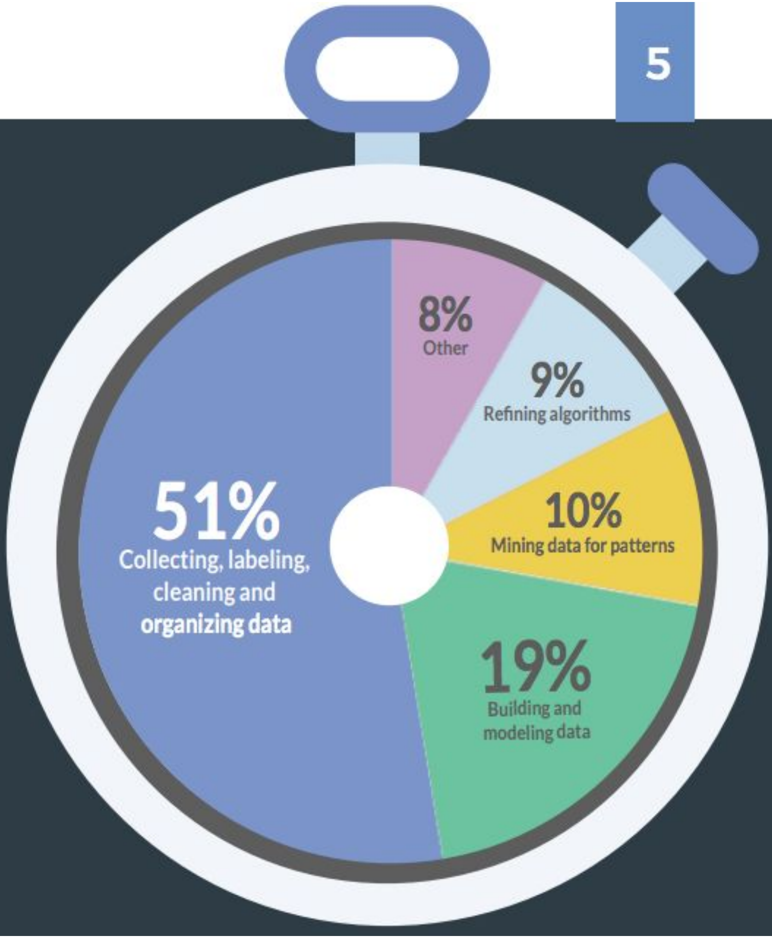
THINKFUL

Source: Glassdoor

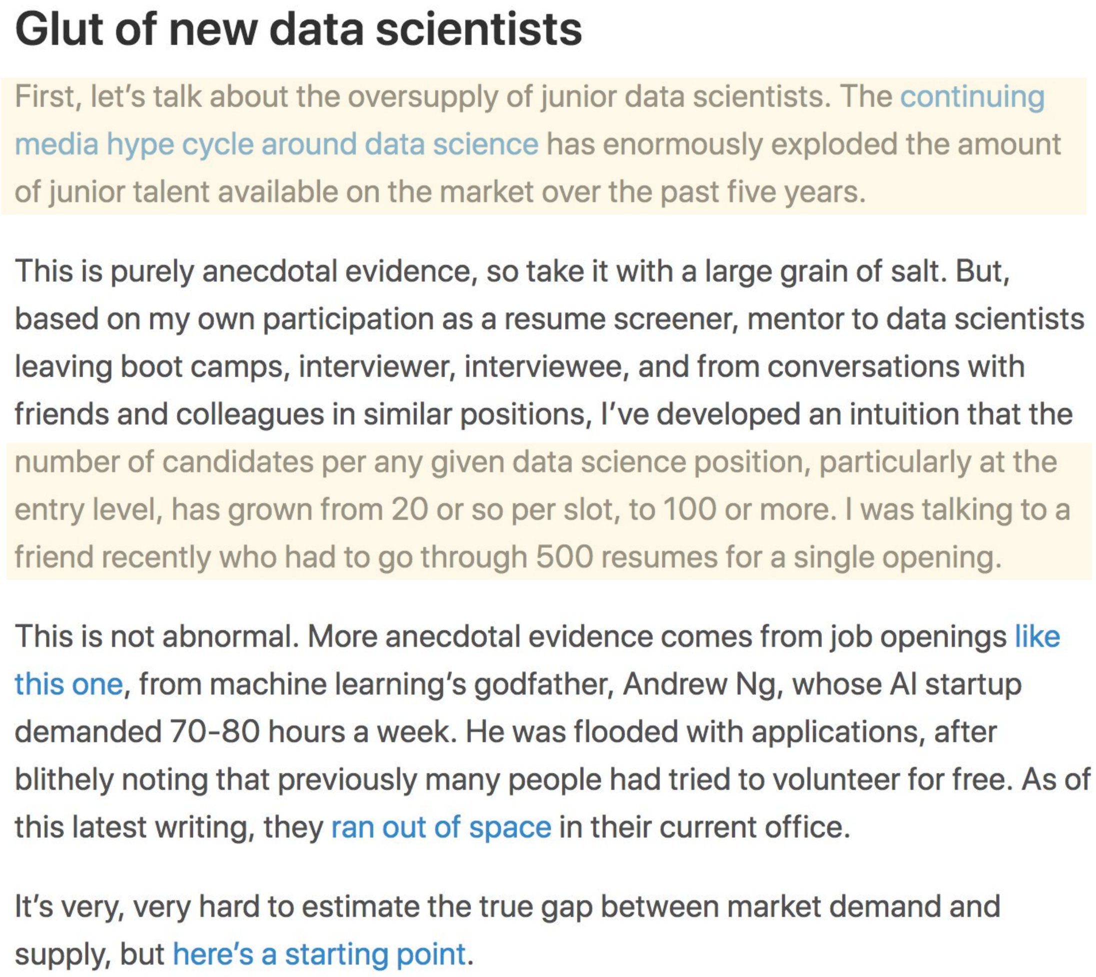
WHAT KEEPS DATA SCIENTISTS HAPPY?

(and why aren't they doing more of it?)

What activity takes up most of your time?



Source: CrowdFlower, *Data Science Report* (2017)



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.

https://veekaybee.github.io/2019/02/13/data-science-is-different/

Hard things are hard.

https://gph.is/17VPNIG

What you all have done

COGS 108: What we’ve learned

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

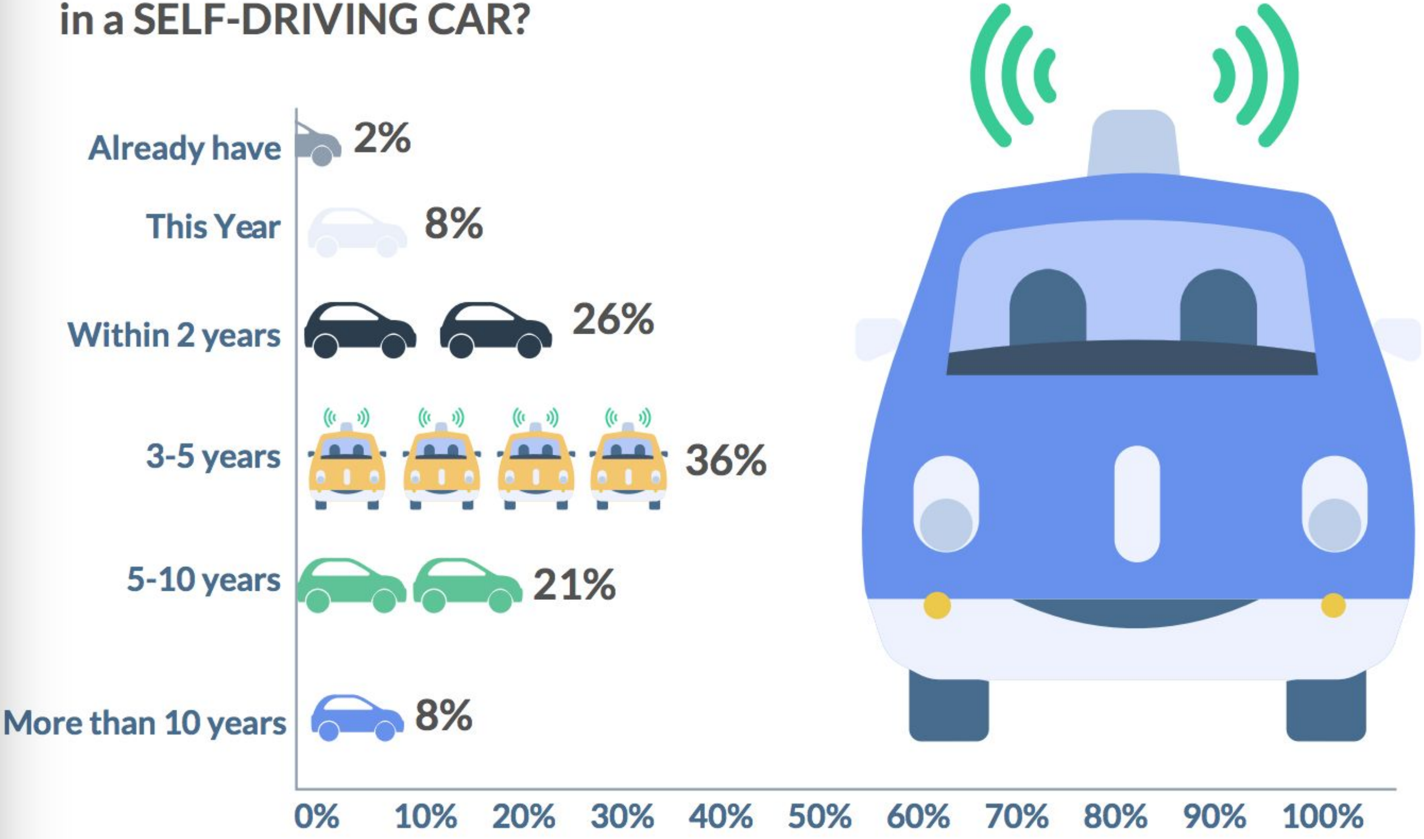
Guest Lecture: Sucheta Jawalkar Ikuenobe

COGS 108: Final Project Lessons

1. Asking the right question up front really helps
2. Finding the data you need <u>is</u> 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?

When do you think you'll first ride in a SELF-DRIVING CAR?



Source: CrowdFlower, *Data Science Report* (2017)

Algorithms are fragile

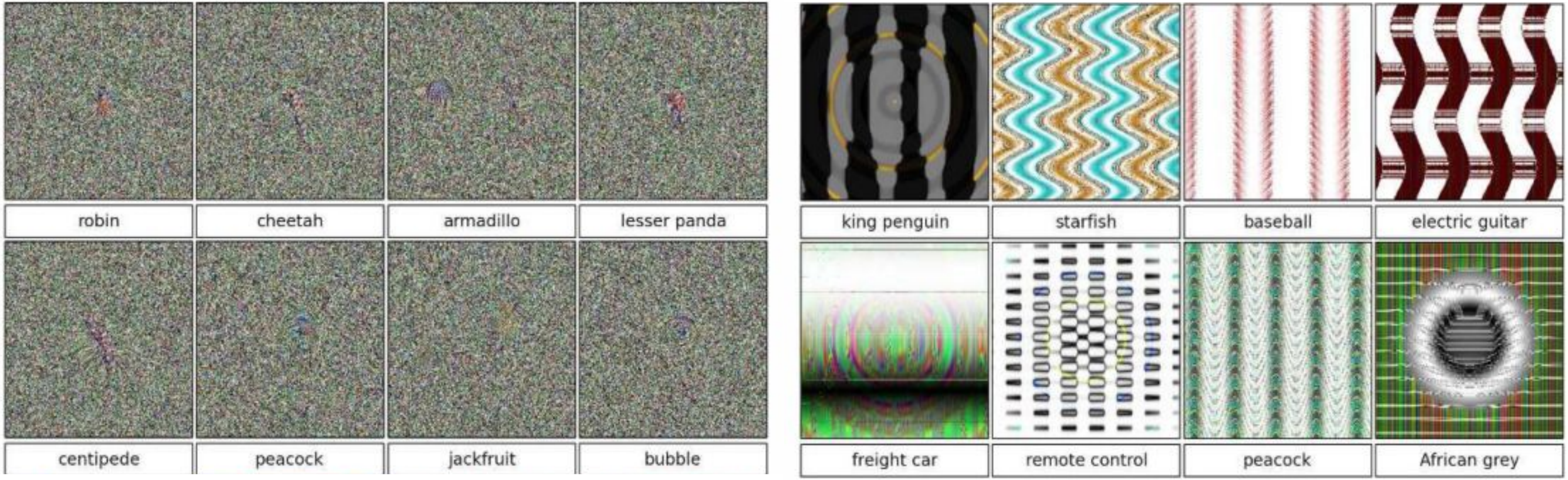
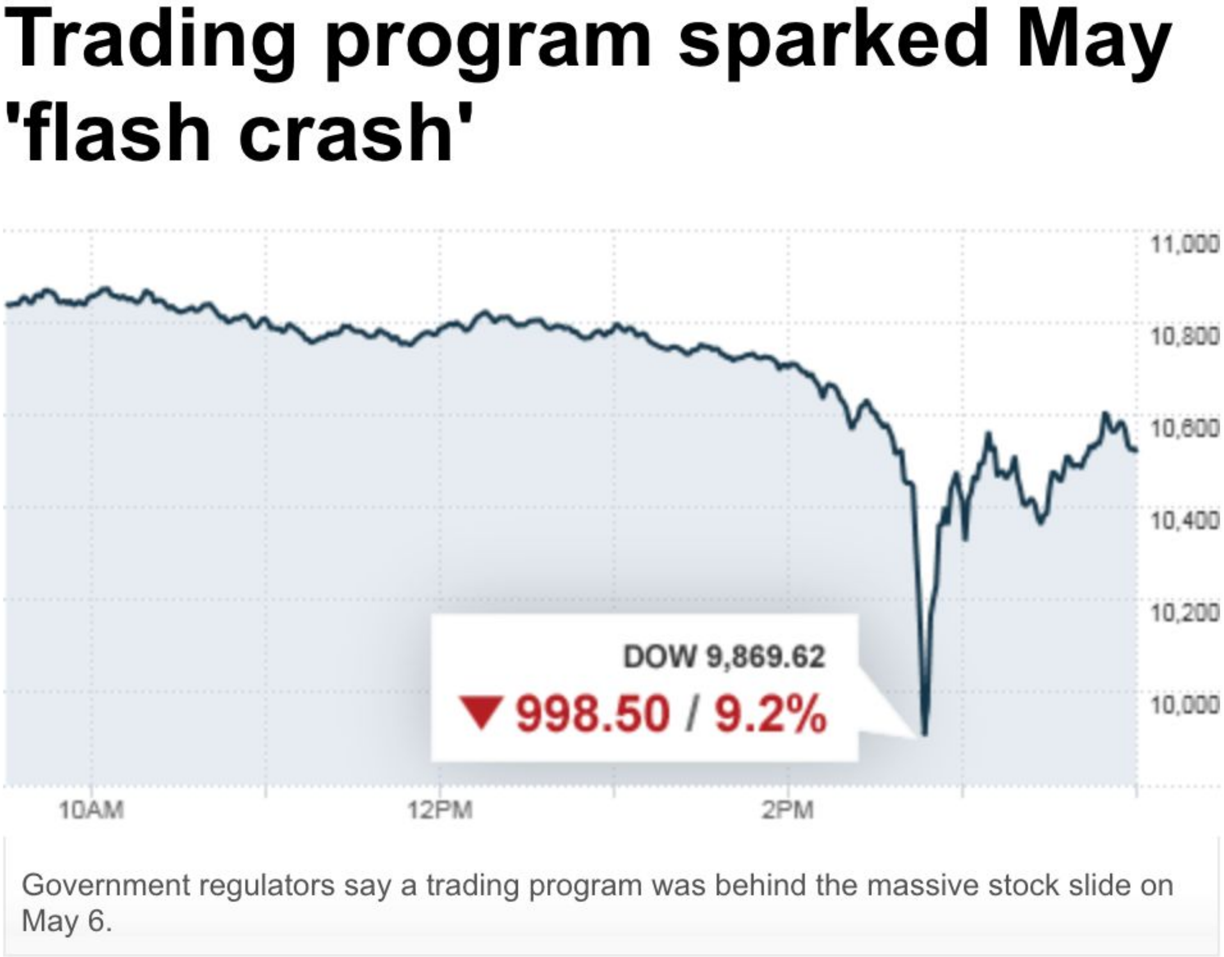


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

Deep Neural Networks are Easily Fooled (IEEE, 2015)



Trading program sparked May 'flash crash'

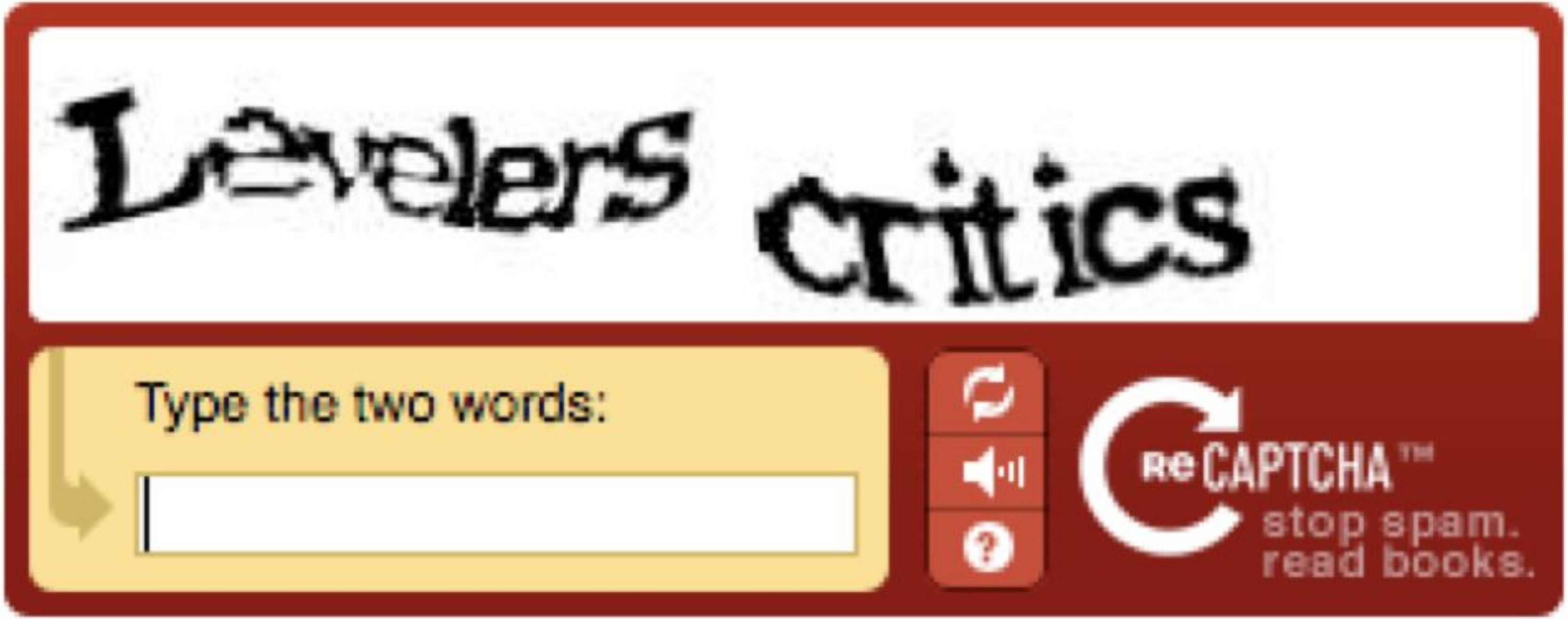
Government regulators say a trading program was behind the massive stock slide on May 6.

Automatic computerized traders on the stock market shut down as they detected the sharp rise in buying and selling. (*NYT*)

Source: CNN Money

Algorithms are fragile & powerful

Human-based computation



Levelers critics

Type the two words:

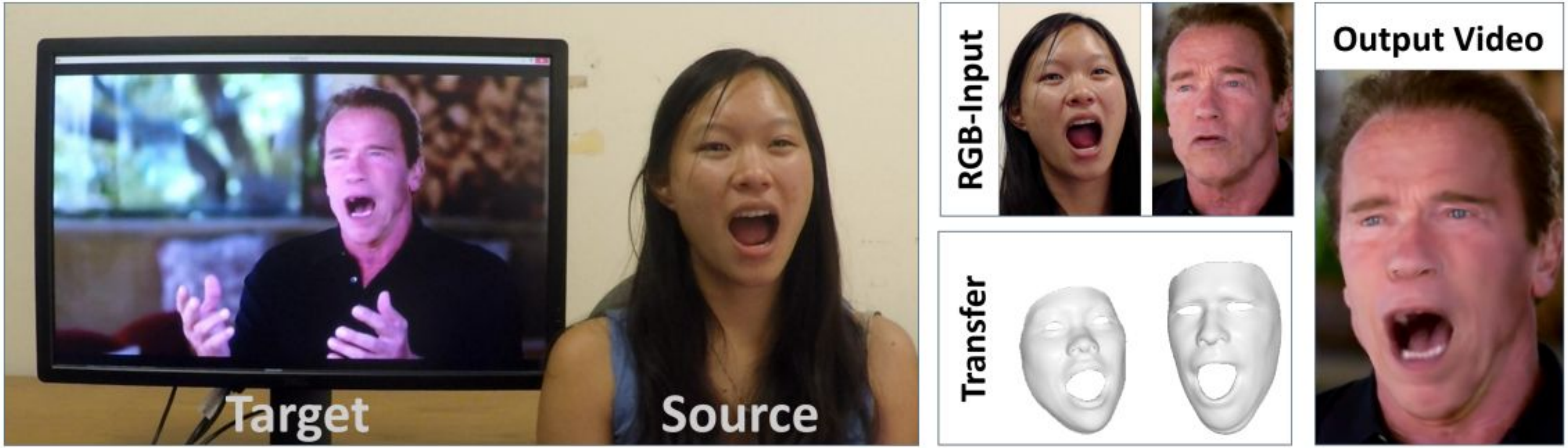
Unfortunately, it would be extremely expensive.

Source: Massive-scale online collaboration (Luis von Ahn, TEDx)

Unfortunately, it would be extremely expensive.

Source: Massive-scale online collaboration (Luis von Ahn, TEDx)

Reality manipulation



Proposed online reenactment setup: a monocular target video sequence (e.g., from Youtube) is reenacted based on the expressions of a source actor who is recorded live with a commodity webcam.

COGS 108 Thank yous!

TAs: Atman, David, Holly, Qin

IAs: Anu, Enoch, Kevin, Tiffany

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 & successful

Thanks for taking COGS 108!