



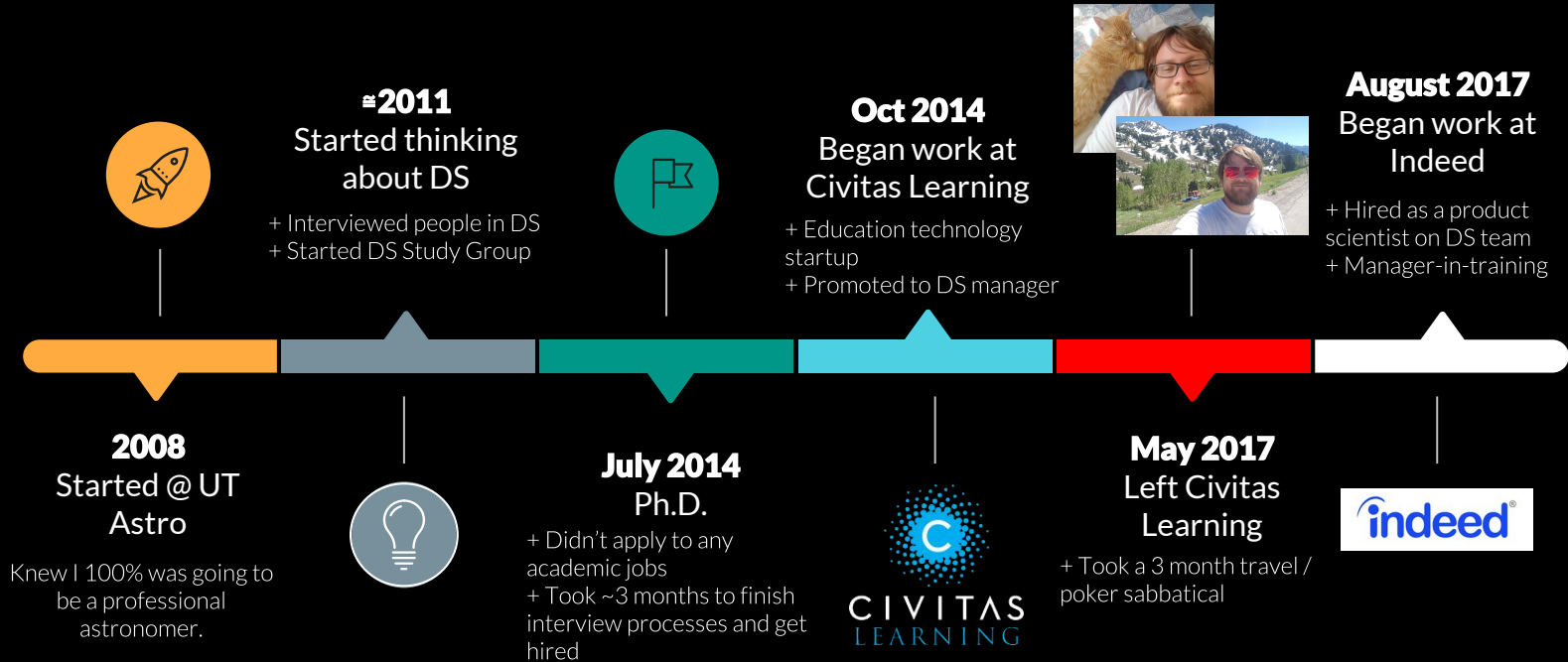
# Academia Schmacademia

## Part 4:

Data Science is *still* really awesome.

Chris Lindner, Ph.D.

# Who am I?



# I talk about this **A LOT**

≈**2011**

Started thinking about DS

- + Interviewed people in DS
- + Started DS Study Group



**Oct 2014**

Began work at Civitas Learning

- + Education technology startup
- + Promoted to DS manager 1.5 years in

Academia Schmacademia Part III:  
Having a Tech Job in Austin

**(2015)**

**Academia Shmacademia:**

A primer on non-academic jobs for astronomers



**July 2014**

Ph.D.

- + Didn't apply to any academic jobs
- + Took 3 months to find a job



**CIVITAS  
LEARNING**

**Academia Shmacademia pt. 2: It gets better!**  
**(Careers in Data Science)**

Links:

[Part 1](#)

[Part 2](#)

[Part 3](#)

And Kevin's talk:

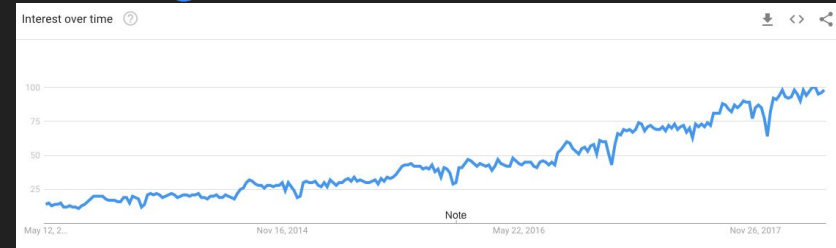
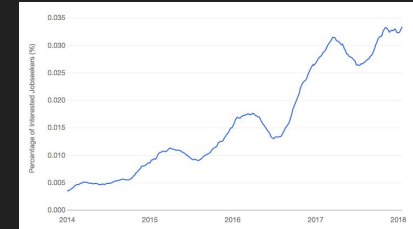
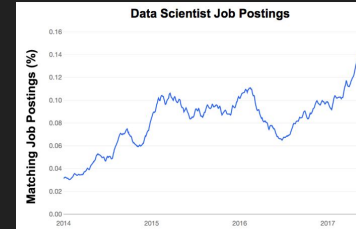
[Data Science at SparkCognition](#)

# Why are people talking about Data Science?

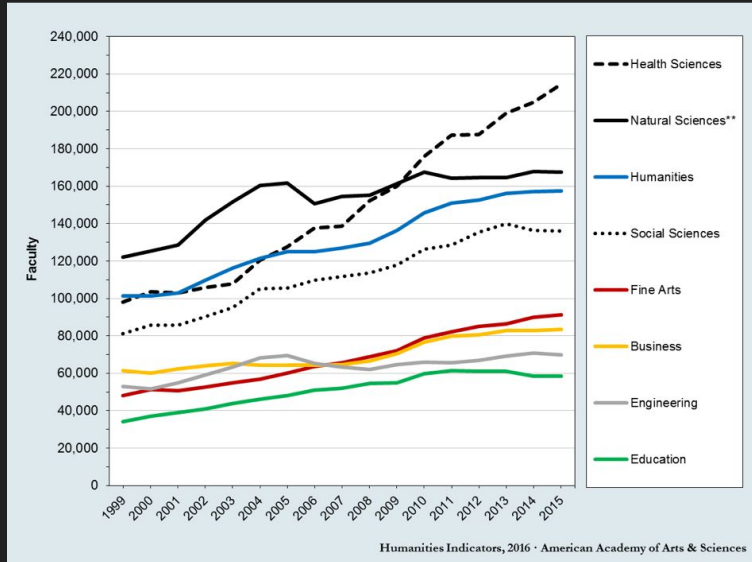
- Data Scientist tops almost every top jobs list. Why?
  - Workers are happy
  - Salaries are high
  - Demand is high
- Why is demand high?
  - **Lots of bang for buck** - Data Scientists provide large, scalable solutions
  - DS requires a unique set of skills that are difficult to learn
- **Hype?**



## DS Job Searches



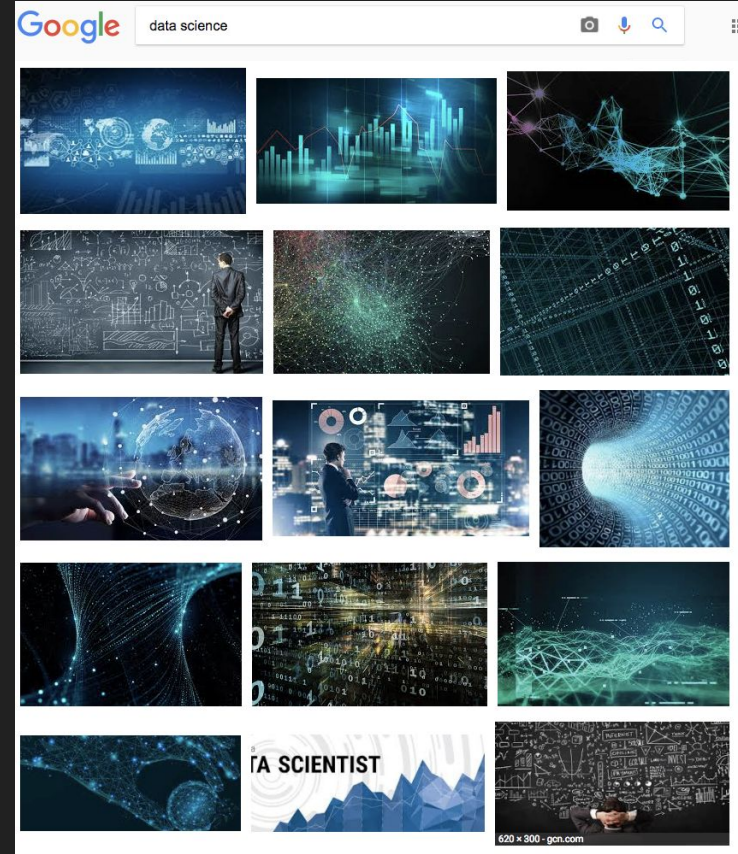
# Why should you care?



...just in case, you can pick up the DS skills you need while you get your Ph.D. without much effort, *and* you'll be a better scientist.

# Why is Data Science?

- Since the 90s, it's become cheap and easy for businesses to collect and store lots of data
- Businesses have been collecting this data for decades now... but haven't been sure what to do with it
- In the early 2010s, companies started nabbing large quantities of academics to make sense of the data



*According to Google, "Data Science" is blue numbers flying on black backgrounds and people staring thoughtfully at blackboards.*

# What are the requisite skills to pay one's bills?



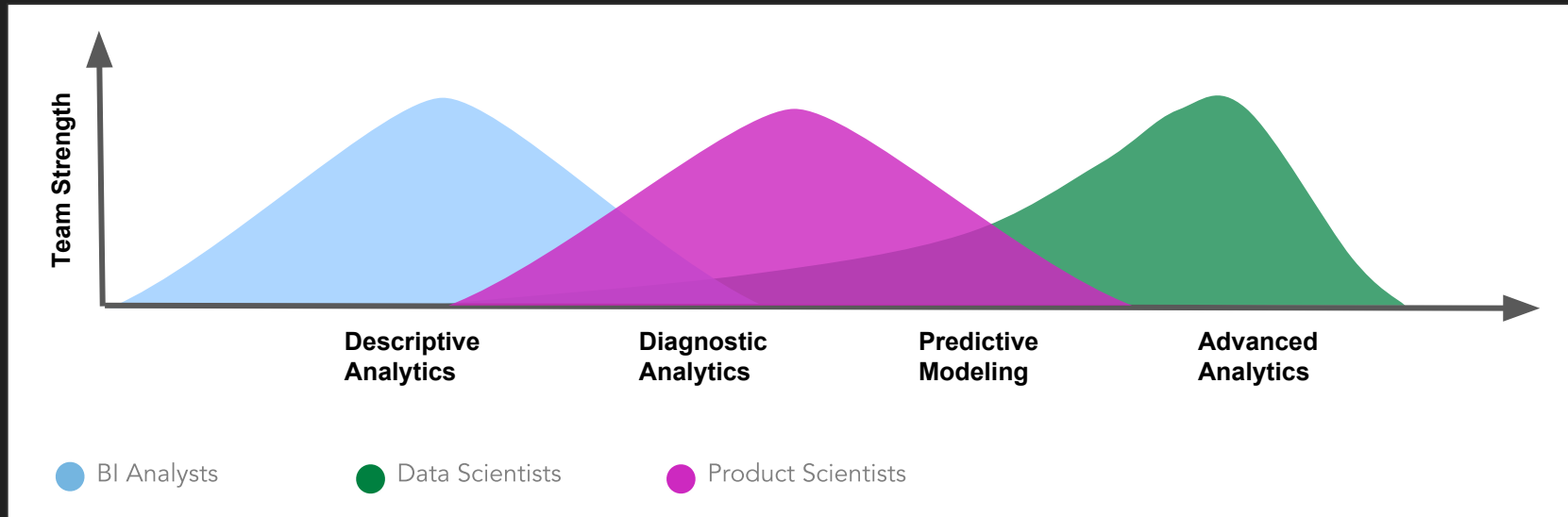
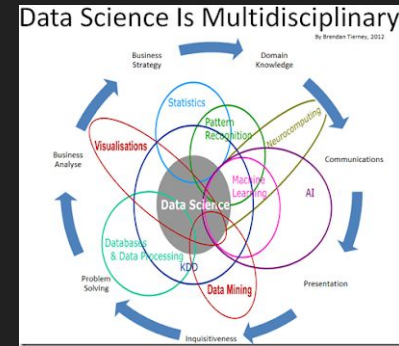
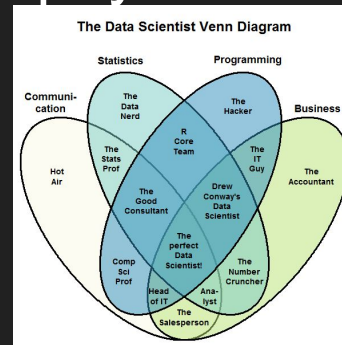
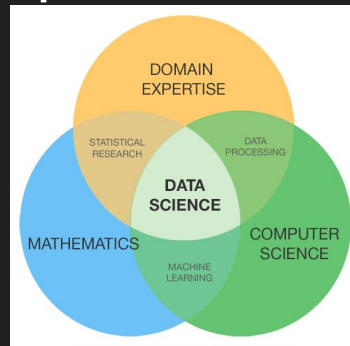
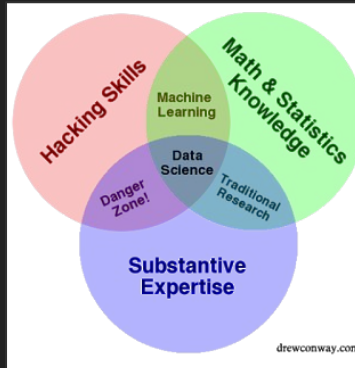
The top 10 skills listed for a Data Scientist job opening in the US from most to least common are:

1. **Python**
2. **Machine learning**
3. R
4. Data science
5. Data mining
6. Hadoop
7. Big data
8. Spark
9. SAS
10. Java

The main things I'd focus on are:

- **Python** - you don't have to be an expert, but you need to know how to hack data things together using Pandas and Scikit-Learn.
- **SQL** - very easy to learn quickly
- **Machine Learning** - 99% of the time this means drawing a straight line through some points. The best courses are still free and online.
- **Project Management** - Pick a pet project that utilizes machine learning to demonstrate you can actually accomplish something.
- **Communication** - Get comfortable with describing technical concepts to laypersons in simple, compelling ways
- **Stats** - Brush up on basic probabilities and learn what an A/B test is.

# What are the requisite skills to pay one's bills?





# What are the requisite skills to pay one's bills?

Check out my previous talks for tips  
on how and where to learn things:

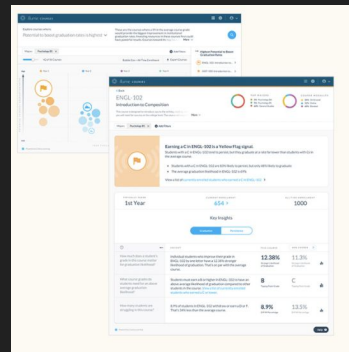
[Part 1](#) [Part 2](#) [Part 3](#)

And check out Indeed's [Data  
Science blog](#) series on DS jobs.

# What does a Data Scientist do?



- Civitas Learning uses **machine learning** to identify students who are at risk for dropping out
- My work focused on understanding how performance in specific courses affected success, and ultimately led to a new product: **Illume Courses** (I didn't get to name it)
- My work on this entailed:
  - Exploratory data analysis
  - Talking to University admins and experts
  - Prototyping plots, analysis, and insights
  - Collaborating with designers
  - Collecting and integrating feedback from users
- The main tools I used were Python, SQL, and Excel



The New York Times



## *Will You Graduate? Ask Big Data*

By Joseph B. Treaster

Feb. 2, 2017

“You could get a C or an A in that first nursing class and still be successful,” said Timothy M. Renick, the vice provost. “But if you got a low grade in your math courses, by the time you were in your junior and senior years, you were doing very poorly.”

The analysis showed that fewer than 10 percent of nursing students with a C in math graduated, compared with about 80 percent of students with at least a B+. Algebra and statistics, it seems, were providing an essential foundation for later classes in biology, microbiology, physiology and pharmacology.

# What does a Data Scientist do?



- Indeed is the #1 job site in the world
- **Data Scientists** build models to predict pricing, analyze resume text, predict user risk, etc.
- As a **Product Scientist** on the data science team, my mission is a bit more broad:
  - Talking to product managers
  - Exploratory data analysis
  - Experiment design
  - Exploratory model building
- Projects I've worked on so far:
  - Analyzing survey data
  - Figuring out how to automatically identify fraud, spam, and scams
  - Predicting when users are going to stop using our product
  - Finding what users will benefit the most from talking to customer service and sales

# What's DS life like?

- Typically a **good work/life balance**: 30-55 hrs/wk
- Liberal vacation policies
- Flexible working hours and conditions
- Travel opportunities
  -
- **Lots of office perks**
  - e.g. free breakfast, lunch, sandwich shop, coffee shop, snacks, drinks, ice cream, booze, Lyft rides, parties...
- **Great salaries**
- Project **flexibility** and work **independence**
- Really fulfilling and **interesting** work

# SIGN ME UP!



We are hiring! Talk to me or John about visiting our main office sometime for some free lunch and discuss anything career related.

- [clindner@indeed.com](mailto:clindner@indeed.com)
- [jjardel@indeed.com](mailto:jjardel@indeed.com)



## JOIN THE ASTRODATA SLACK

- Chat with new students and alums on demand.
- Details are on the grad student wiki.

**The most important thing you can do to get a job is network.**

Thankfully, you already have a huge network!

- **Chris Lindner** - Indeed
- **John Jardel** - Indeed
- **Kevin Gullikson** - SparkCognition
- **Jeff Silverman** - SambaTV
- **Julie Krugler** - Twitter
- **Chalence Timber Safranek-Shrader** - FiveStars
- **Emma Yu** - Facebook
- **Jacob Hummel** - Overstock.com
- **Emily McLinden** - Square
- **Randi Ludwig** - Dell
- **Stefano Meschiari** - Duo Security
- **Sam Harrold** - Indiana University Health