Becoming a Data Scientist: Advice From My Podcast Guests

(and my twitter followers, and me!)

Renée Marie Parilak Teate

Presented by Renee Teate at PyData DC 10/9/2016
Shared under Creative Commons Non-Commercial ShareAlike License
(Share/remix, with attribution. Any derivative works must be similarly licensed. Not for commercial use.)

Renée Teate

· ·

Data Scientist

@becomingdatasci
on Twitter

Data Scientist at HelioCampus
Better Data. Brilliant Insights.
"Asking and answering the most pressing questions in higher education for your institution."

Host of the Becoming A Data Scientist Podcast & Learning Club
Interviewing data scientists to find out how they got to where they are today since Dec. 2015
500+ member learning club (~90 of whom have posted a thread in the forums)

Creator and Developer of DataSciGuide
Online Data Science Learning Directory (work in progress)

Database Designer & Developer, Web Developer, SQL Data Analyst
2004-2016

B.S. Integrated Science and Technology
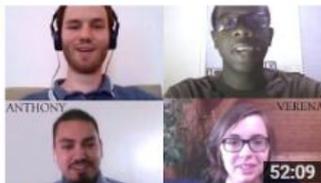from James Madison University, 2004

M.Eng. Systems Engineering
from University of Virginia, 2015

Becoming a Data Scientist Podcast Guests

Becoming a Data Scientist
Podcast: Episode 13 - Debbie...
873 views • 3 months ago

Becoming a Data Scientist
Podcast: Episode 12 - Data...
617 views • 3 months ago

Becoming a Data Scientist
Podcast: Episode 11 - Stephani...
442 views • 4 months ago

Becoming a Data Scientist
Podcast: Episode 10 - Trey...
772 views • 5 months ago

Becoming a Data Scientist
Podcast: Episode 09 - Justin...
347 views • 5 months ago

Becoming a Data Scientist
Podcast: Episode 08 - Sebastia...
1,261 views • 6 months ago

Becoming a Data Scientist
Podcast: Episode 07 - Enda Ridge
443 views • 6 months ago

Becoming a Data Scientist
Podcast: Episode 06 - Erin...
805 views • 7 months ago

Becoming a Data Scientist
Podcast: Episode 05 - Clare...
2,207 views • 7 months ago

Becoming a Data Scientist
Podcast: Episode 04 - Sherman...
532 views • 8 months ago

Becoming a Data Scientist
Podcast: Episode 03 - Shlomo...
628 views • 8 months ago

Becoming a Data Scientist
Podcast: Episode 02 - Safia...
1,177 views • 8 months ago

Becoming a Data Scientist
Podcast: Episode 01 - Will Kurt
2,273 views • 9 months ago

Becoming a Data Scientist
Podcast: Episode 00 - Renee...
2,459 views • 9 months ago

"How do I become a data scientist?"

"It depends..."

Becoming a Data Scientist

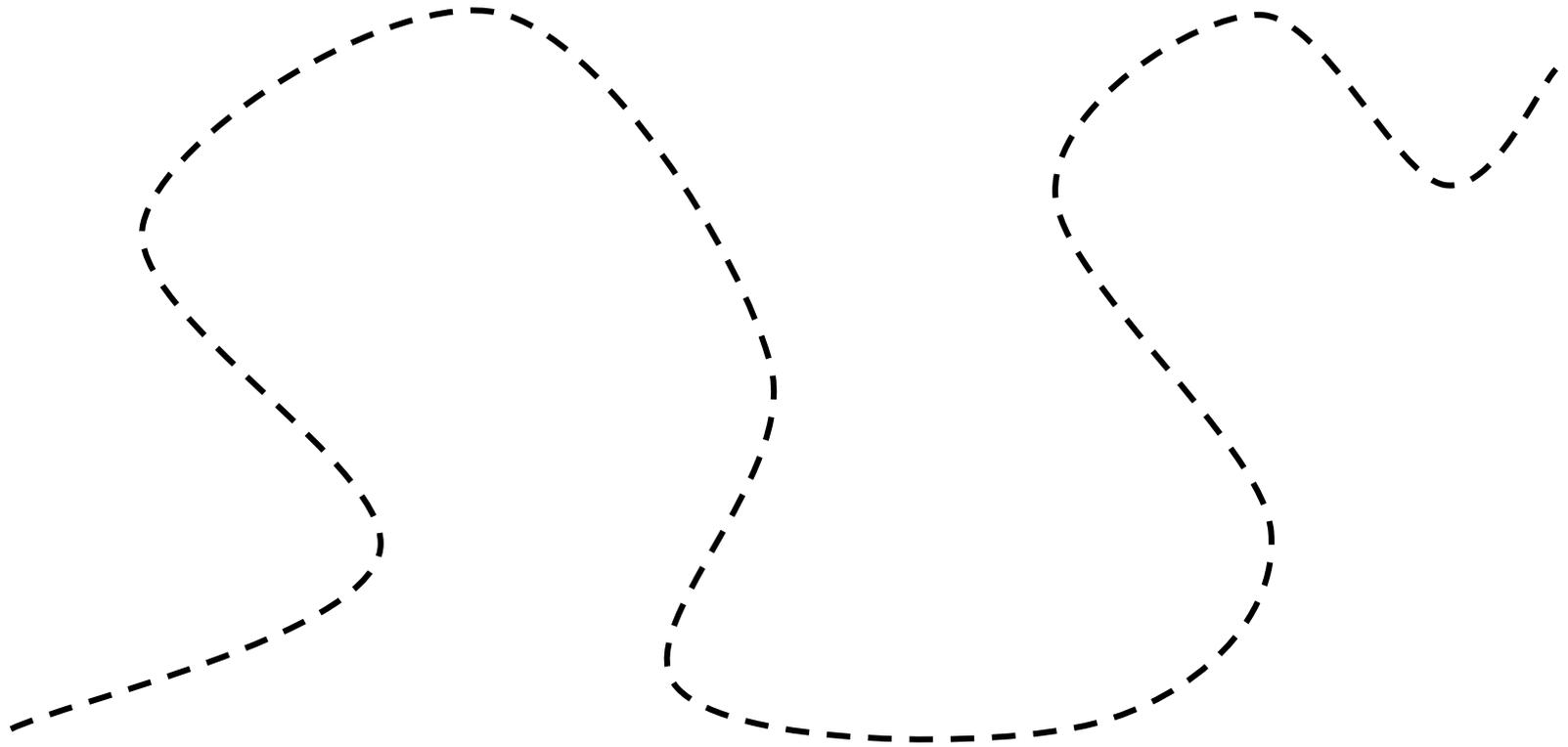|	
01 Where to Start	**02** Who to Listen to
03 What to Learn	**04** How to Learn
05 How to Practice	**06** When You're Ready

1.
Where to Start

"Where should I start?"

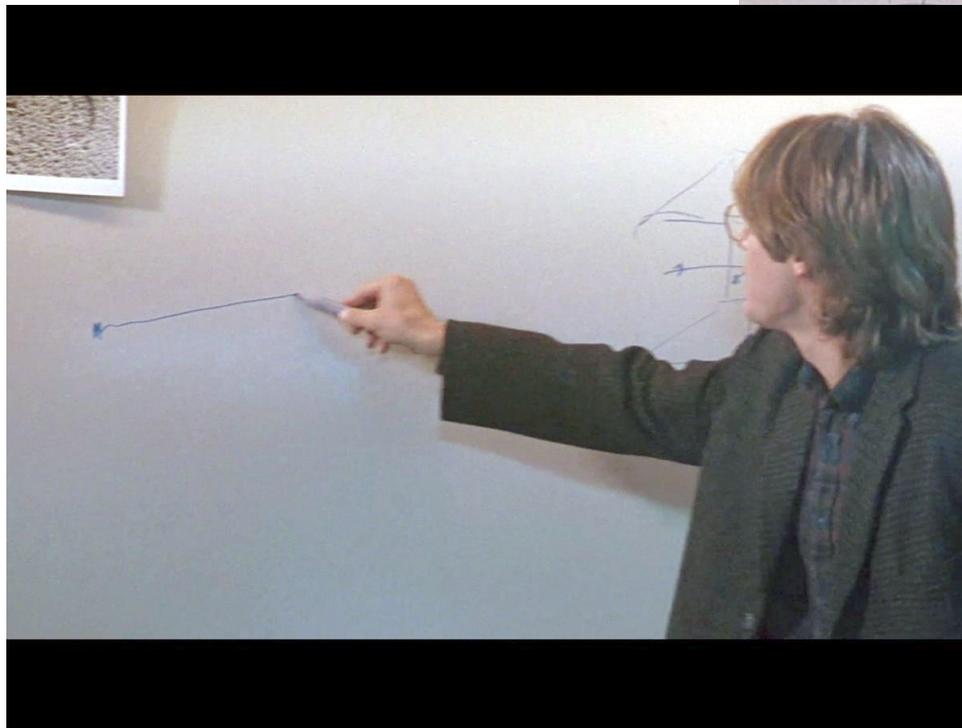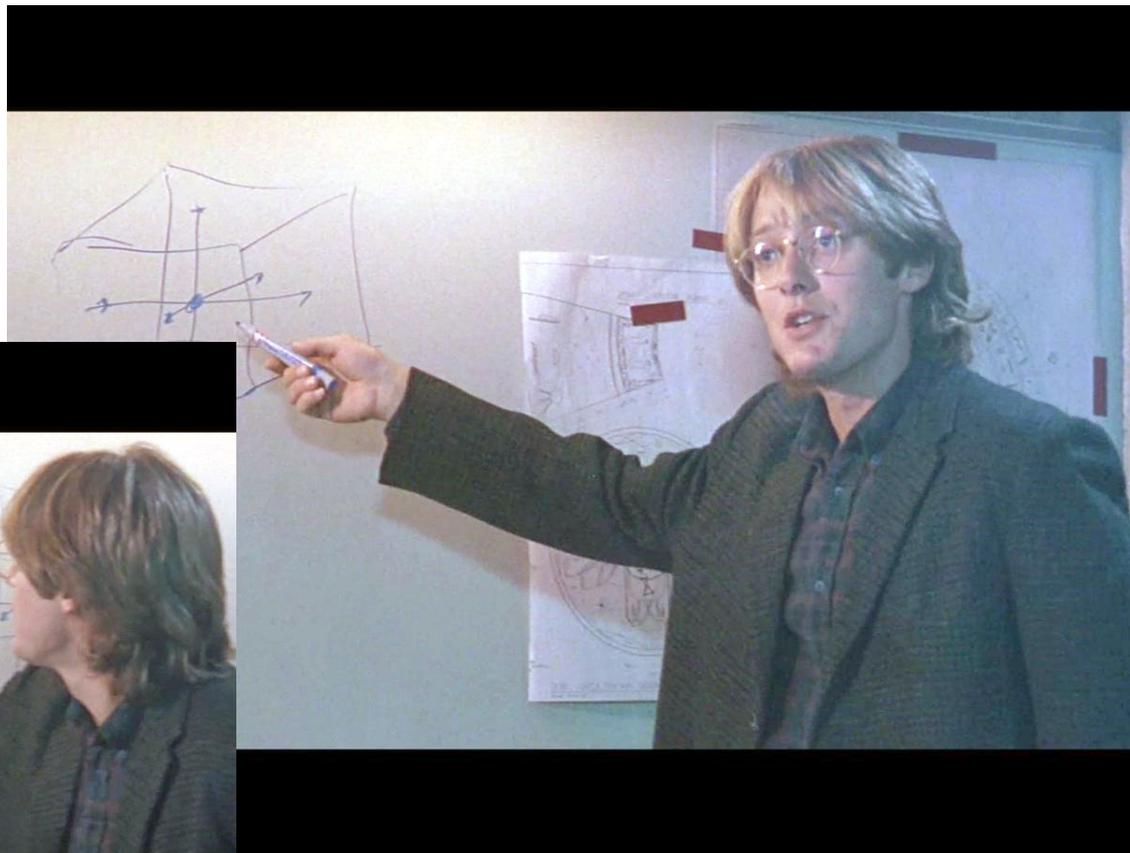# "Where do I even start?!?!"

"It depends..."

First you have to figure out:
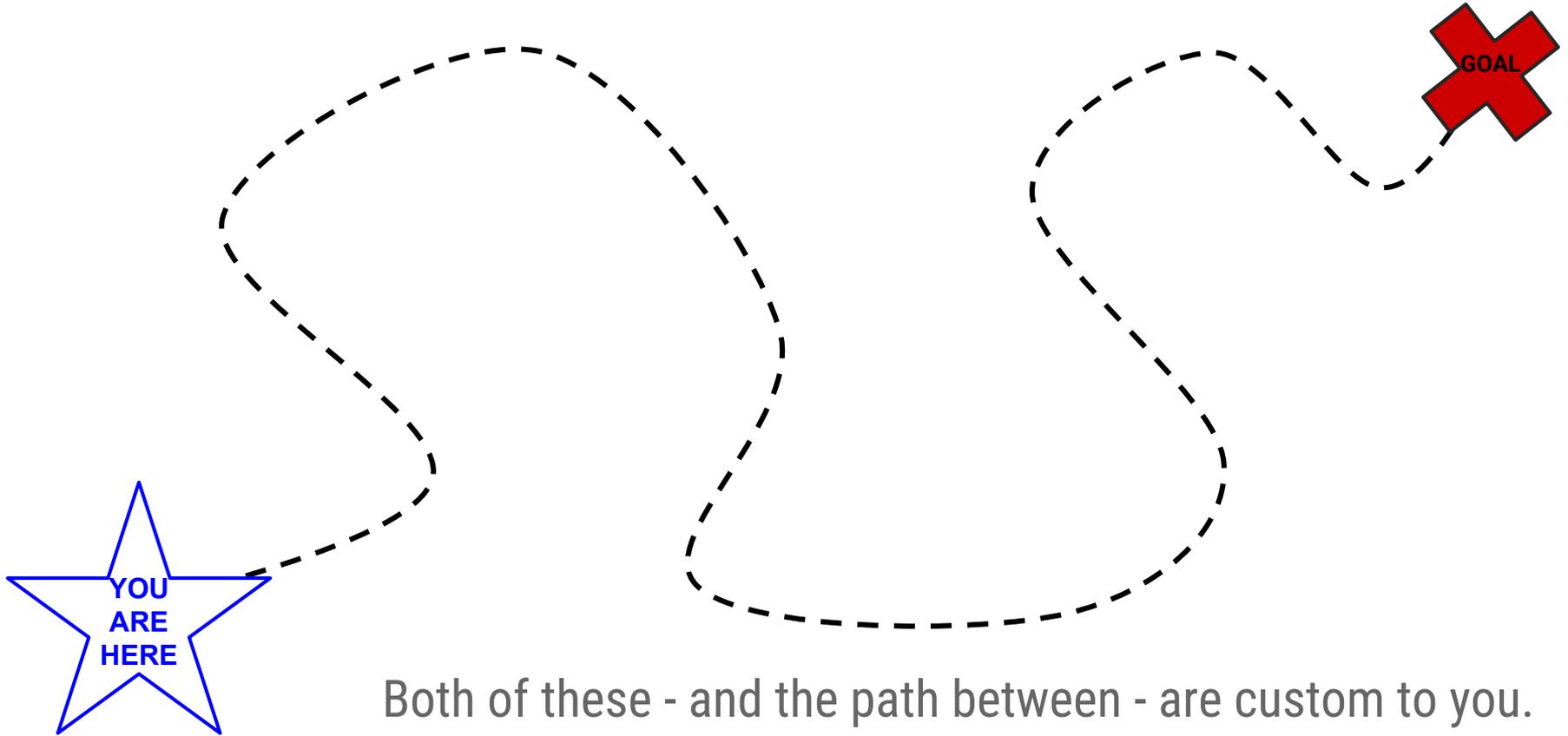
1. *Where You Are*

and

2. *Where You Want To Go*

James Spader as Daniel Jackson
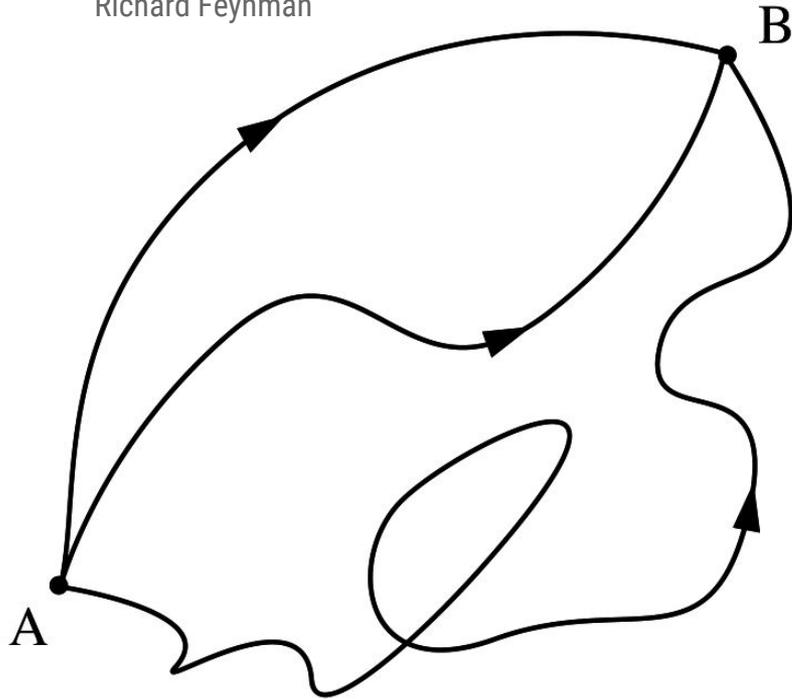in Stargate,
explaining the 7th chevron

Assess your starting point & have an ultimate goal end point to move towards.

GOAL

YOU
ARE
HERE

Both of these - and the path between - are custom to you.

Richard Feynman

$$W(q,\, t;\, q_0,\, t_0) = \int_{q_0}^{q} \mathscr{D}q \, \exp\left[\int_{t_0}^{t} -\frac{1}{i\hbar}\, L(\dot{q},\, q,\, \tau)d\tau\right]$$

$$= \lim_{N\to\infty} \int dq_1 \cdots dq_{N-1} \left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{N/2}$$

$$\times \exp\left\{-\frac{\varepsilon}{i\hbar}\sum_{k=1}^{N}\right.$$

$$\left.\times \left[\frac{m}{2}\left(\frac{q_k - q_{k-1}}{\varepsilon}\right)^2 - U(q_{k-1})\right]\right\}, \quad (8)$$

My husband is a physicist and thought I should remind you here that there are infinite paths between two points (Feynman Path Integral)

NOTE:
There will be no more gratuitous complicated equations in the remainder of this presentation.

Where are you starting from?

YOU ARE HERE

How much math have you taken in school
or learned yourself?

Have you ever programmed?

Have you done basic data analysis
on a professional level before?

Data Scientist Profile Categories

From ***Doing Data Science*** by Cathy O'Neil & Rachel Schutt

Mathematics
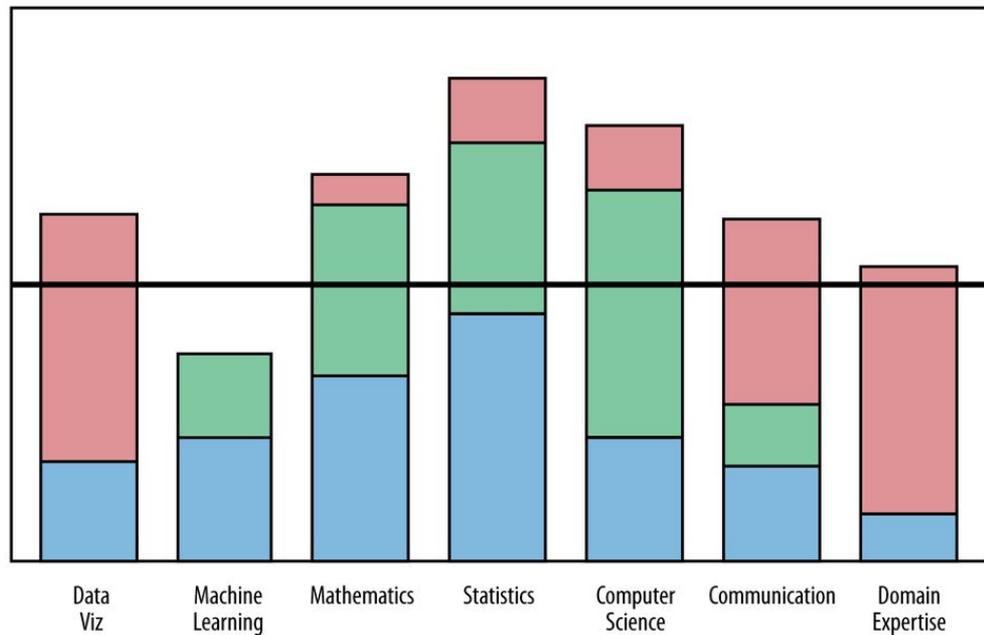
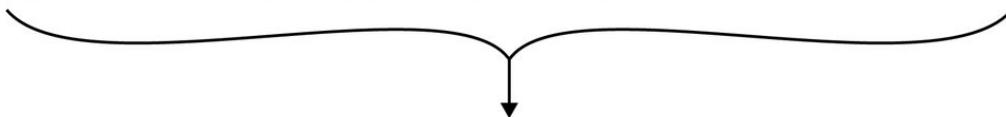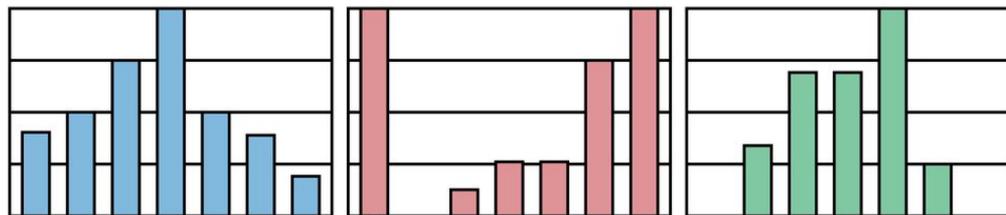Statistics

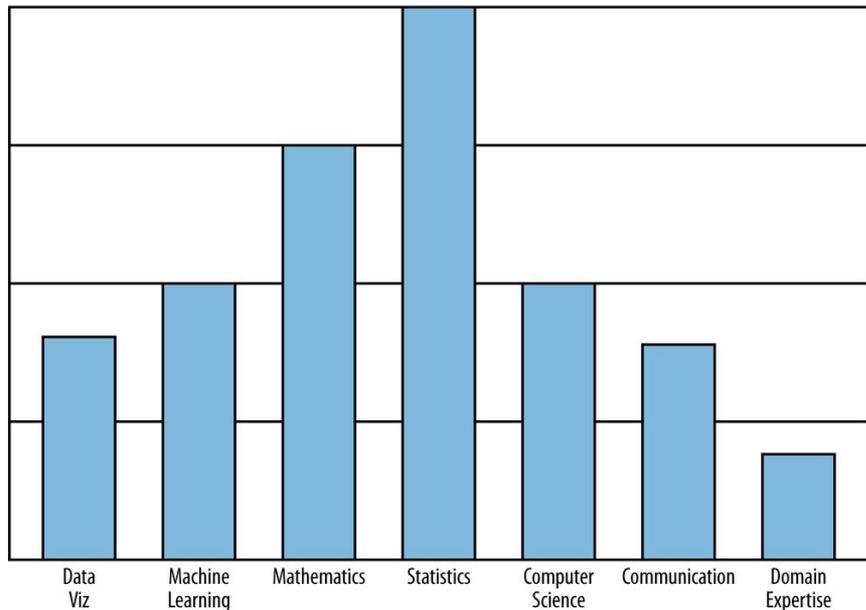Computer Science

Machine Learning

Data Visualization

Communication

Domain Expertise

No one person can be the perfect data scientist, so we need teams.

From ***Doing Data Science*** by Cathy O'Neil & Rachel Schutt

What kind of data scientist do you want to be?

GOAL

Love solving business problems with data and working with people?
Data Analyst / Data Scientist

Love writing efficient code and working with back-end "big" data systems?
Data Engineer / Data Scientist

Love doing cutting-edge research?
Machine Learning Researcher / Statistician / Data Scientist

What kind of data scientist do you want to be?

GOAL

Some jobs that now overlap with Data Scientist...

- Data Analyst
- Big Data Engineer
- Data Mining Specialist
- Machine Learning Developer
- Neuroscientist
- Computational Physicist

- Recommendation System Engineer
- Statistician
- Financial Analyst
- Biostatistician
- Social Science Researcher
- Artificial Intelligence Researcher
- Marketing Analyst

- Autonomous Vehicle Systems Developer
- GIS Analyst
- Natural Language Processing Researcher
- Sentiment Analyst
- Social Network Analyst
- Computational Biologist
- Medical Data Analyst

GOAL

What kind of data scientist do you want to be?

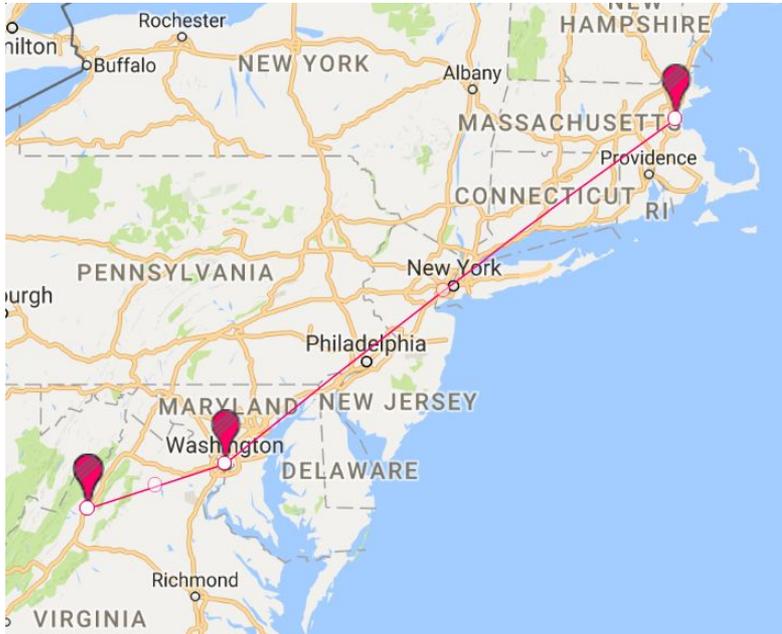**Find guidance & inspiration:**

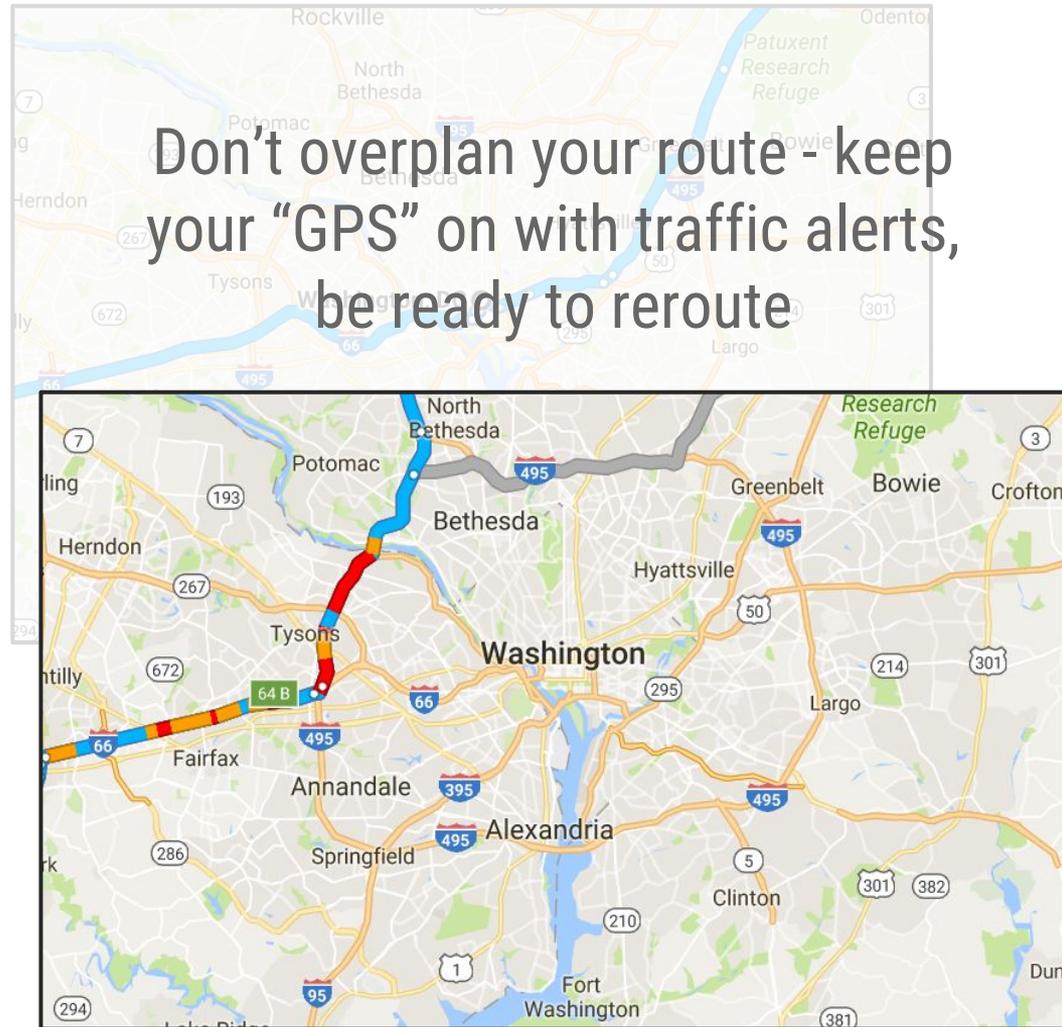Look for role models

Look at job listings online

Ask data scientists you meet

Join the conversation on Twitter

Keep your path general,
with some waypoints

Don't overplan your route - keep
your "GPS" on with traffic alerts,
be ready to reroute

There's so much to learn, how you narrow your focus? You can crowdsource your knowledge, and you definitely need a strong foundation in certain topics but…

"Part of it is being OK with not understanding everything… You're going to be doing yourself a disservice if you set out to learn all of the different topics that people say you should learn. You'll ask 10 people and get 12 answers."

-Trey Causey, @treycausey
Data Scientist & Product Strategy at ChefSteps
Episode 10

"It's going to be your job for the rest of your life, so if you're going to get really good at something, make sure it's something you enjoy, because it requires a lot of persistence and curiosity to be happy and successful... don't force yourself into something... follow what you enjoy!"

-Stephanie Rivera, @dataginjaninja
Principal Data Scientist at myStrength
Episode 11

This is your chance to become a candidate for your "dream job" and decide what YOU want to learn!

MAKE IT FUN!
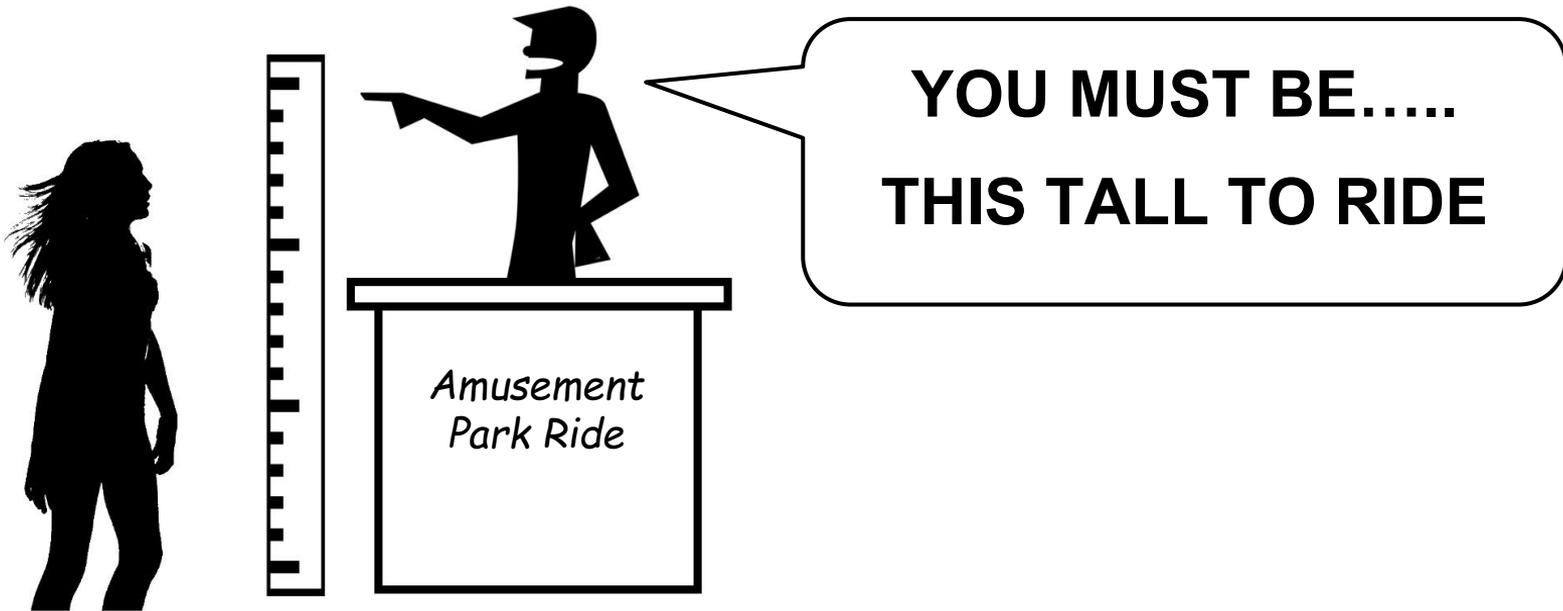You don't have to follow someone else's program or make it feel like school.

Once you have an idea about a possible path, who should you turn to for guidance?

2.
Who to Listen To
(and Who Not to Listen to)

YOU MUST BE…..
THIS TALL TO RIDE

Amusement
Park Ride

Who not to listen to:

The Gatekeeper

YOU MUST BE.....
A PHD

"Real"
Data
Scientist

Who not to listen to:

The Gatekeeper

YOU MUST BE.....
A COMPUTER SCIENTIST

"Real"
Data
Scientist
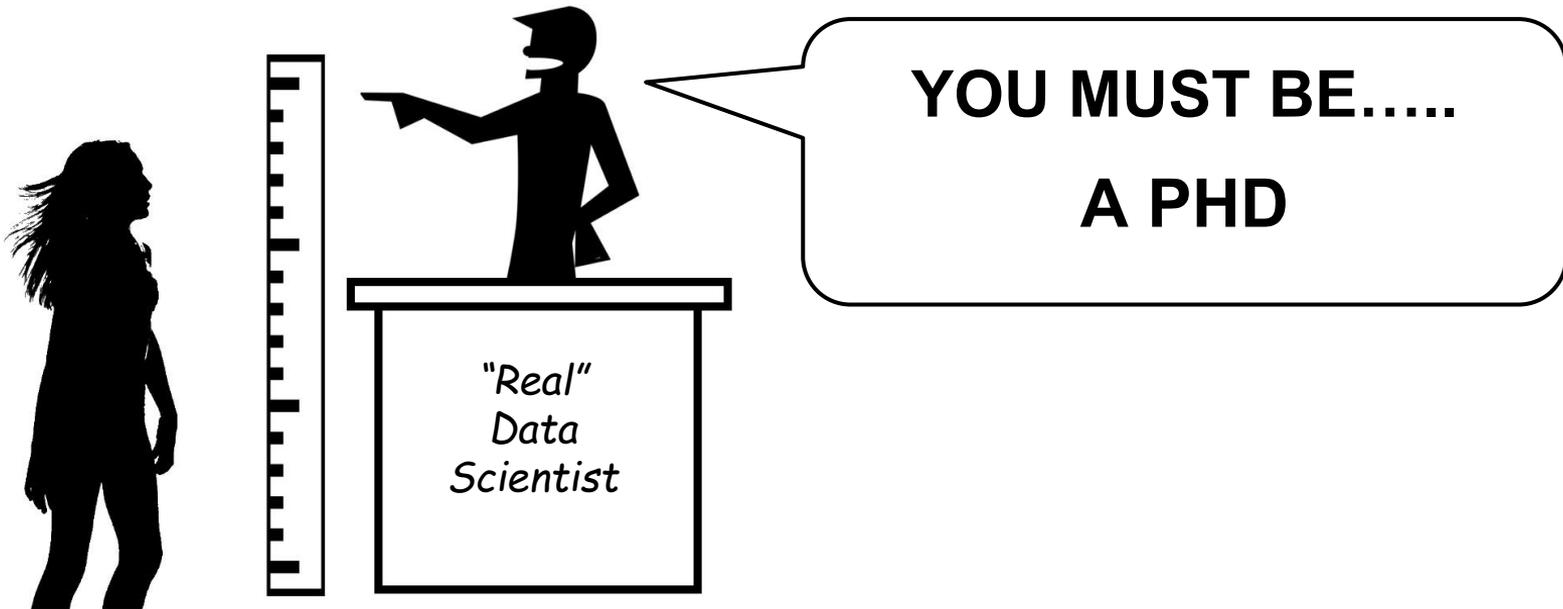
Who not to listen to:

The Gatekeeper

Who not to listen to:

The Gatekeeper

YOU MUST BE…..
TRADITIONALLY EDUCATED

"Real" Data Scientist

Who not to listen to:

The Gatekeeper

YOU MUST BE….. JUST LIKE ME AND THE PEOPLE I KNOW

"Real" Data Scientist
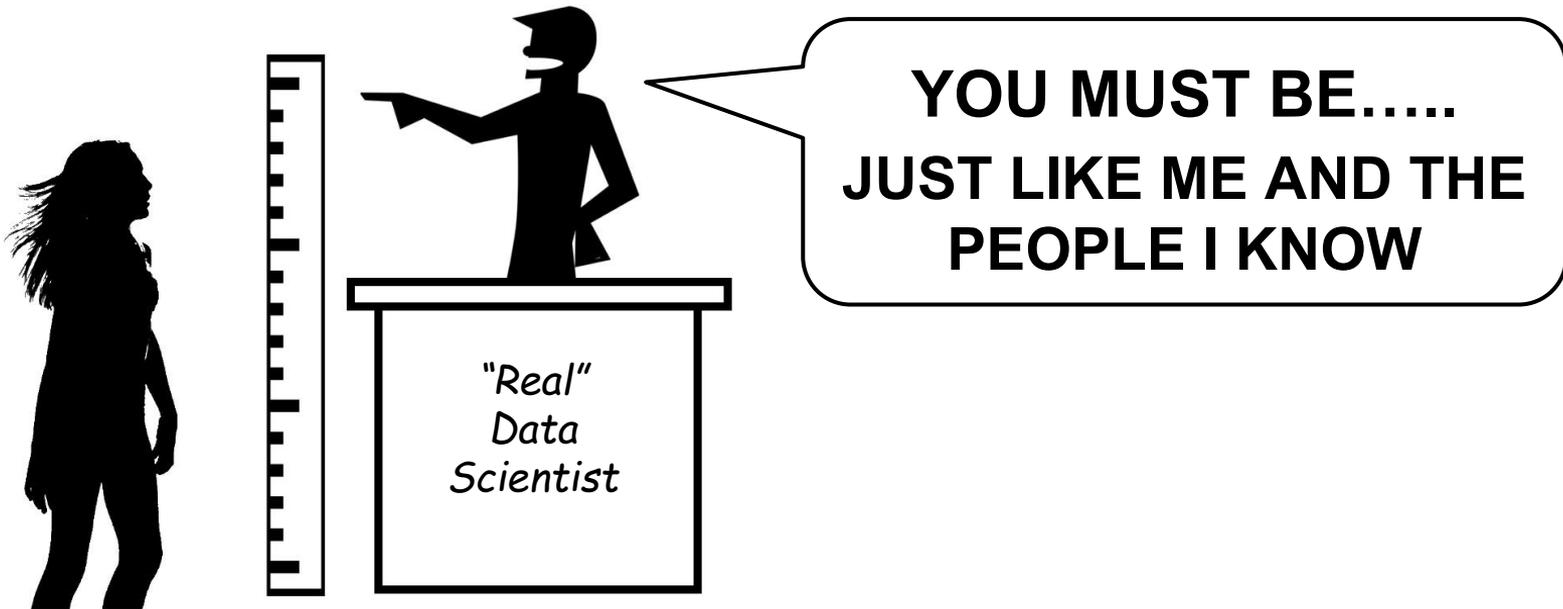
Who not to listen to:
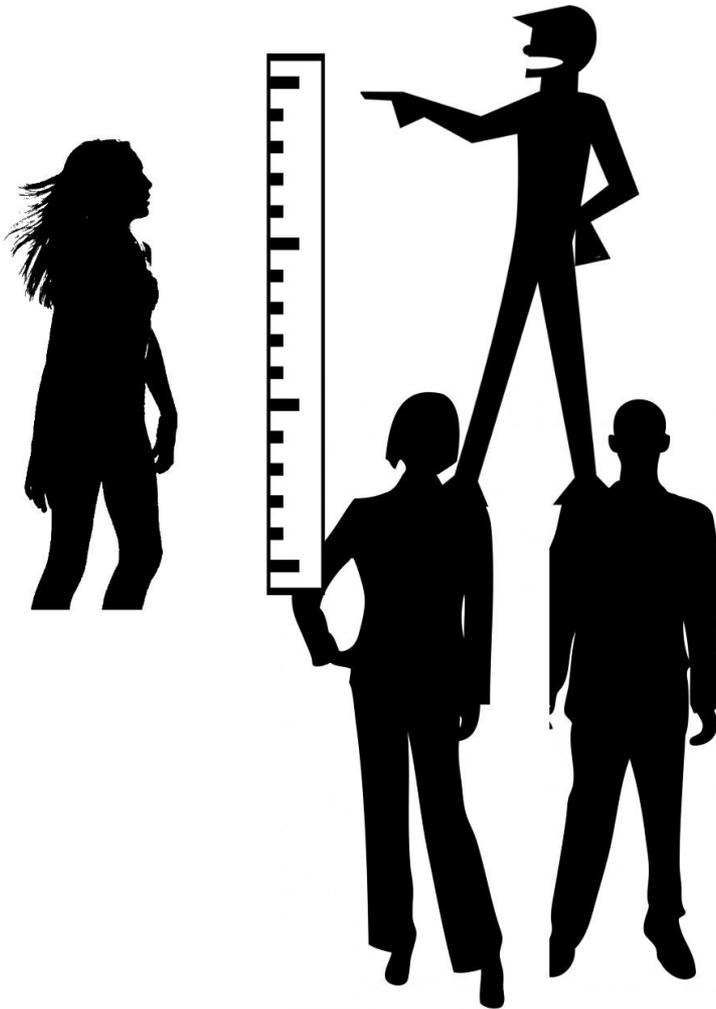
The Gatekeeper

~~*The Gatekeeper*~~

Honest Data Scientist:

"If I have seen further it is by standing on the shoulders of giants."

-Isaac Newton
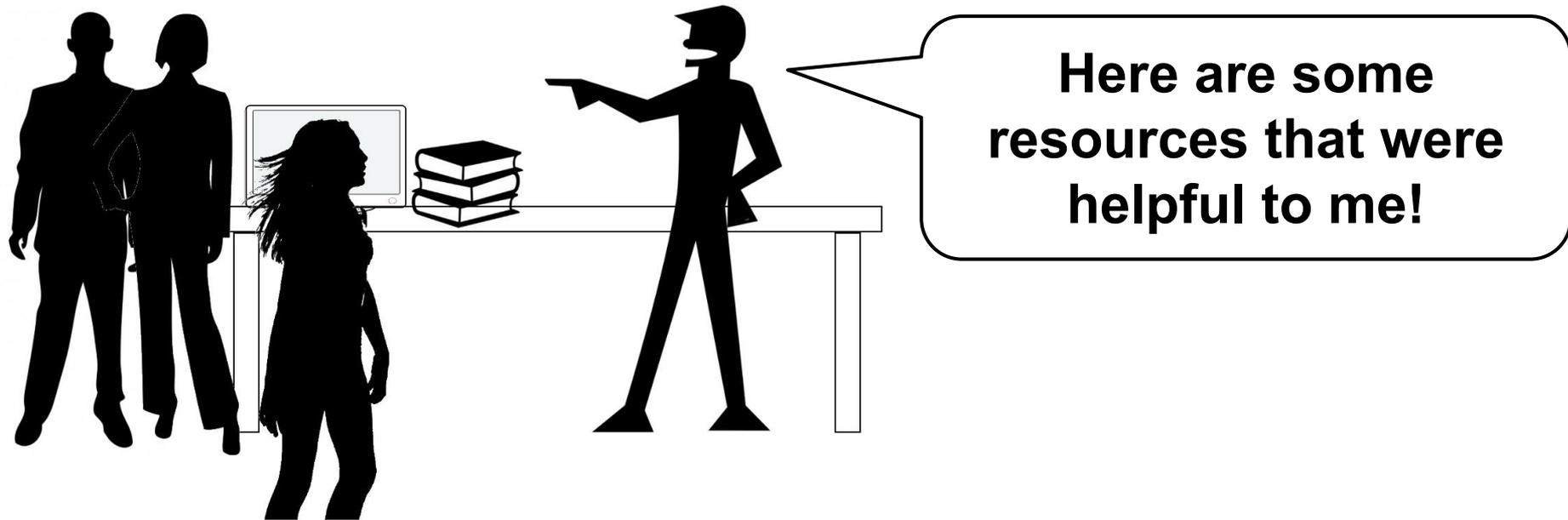
Who to listen to:

Honest and Helpful People Who Have Been There

Who Not to Listen To

- The Gatekeepers
  - "You have to be/have…."

- The Naysayers
  - "You aren't cut out to be…."
  - "In the past you…."
  - "You don't look like a…."

- The Scoffers
  - "It's SO easy…."

- Your Own Doubts
  - "If I'm already struggling…."

Who Not to Listen To:

- The Gatekeepers
  - "You have to be/have…."

- The Naysayers
  - "You aren't cut out to be…."
  - "In the past you…."
  - "You don't look like a…."

- The Scoffers
  - "It's SO easy…."

- Your Own Doubts
  - "If I'm already struggling…."

Who to Listen To:

- Honest and helpful people who have been there
  - "I got stuck at that point, too. These resources helped…."

- Your Supporters
  - "You're so awesome at…."
  - "I can't wait to see you do…."

- Experts who know how to teach beginners
  - "Let's walk through how this works…."

- Your Confident Self
  - "I can do this, I just need…."

For guidance, find….

People who are the type of data scientist you want to become

People who are experts, but know how to communicate with beginners (i.e. good teachers)

Other beginners, and people who are just ahead of you on the data science learning path

Mentors and friends who share your passion and are in the domain in which you want to work

Data Science is not "easy".
(it's not even very well-defined)

Data Science is not "too hard for you".
(though some parts might be right now)

It's up to you not to get derailed.

We'll talk more about *How to Learn* in Part 5.

3.
What to Learn

"So what do I need to learn?"

"It depends..."

Example Data Science Careers & Skills

	Mathematics & Statistics	Computer Science & Machine Learning	Data Visualization & Communication	"Business" Domain Knowledge
Data Analyst/Scientist	✓	✓	✓	✓
Big Data Engineer	✓	✓	✓	✓
Machine Learning Dev.	✓	✓	✓	✓
Marketing Analyst	✓	✓	✓	✓
Biostatistician	✓	✓	✓	✓

First, learn the basics

Basic Data Science Foundation

Source: Palmer, Shelly. Data Science for the C-Suite.
New York: Digital Living Press, 2015. Print.
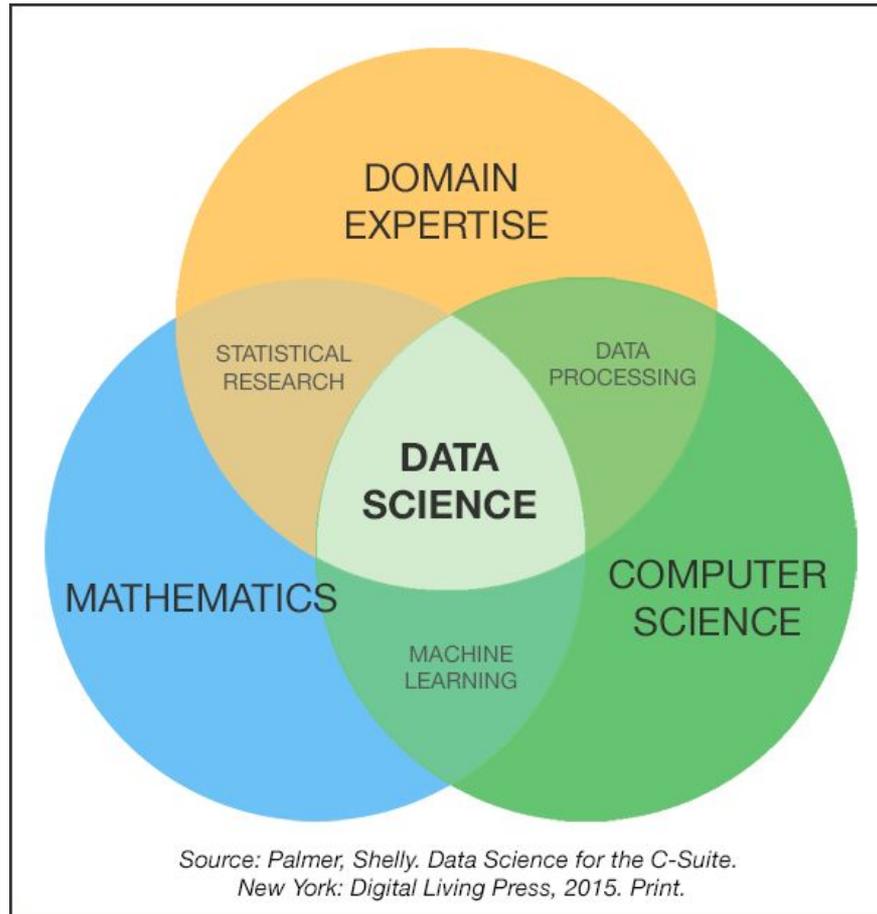
Mathematics/Statistics

Minimum

- Undergraduate-Level Descriptive Statistics

- Data Visualization Techniques

Next

- Linear Algebra
- More Advanced Statistics
- Calculus

Computer Science

Minimum

- Coding for Data Manipulation & Summarization in 1 language (like python)

- Basic "Packaged" Machine Learning Techniques

Next

- Understanding databases & SQL

- Understanding how Machine Learning Algorithms work in detail "behind the scenes"

- Additional languages (R, Julia, etc.) or tools/packages

- Additional data processing frameworks (Spark, Hadoop, etc.)

Domain Knowledge

Minimum

- Ability to do simple "business intelligence" type analysis in domain

- Basic communication - understanding the terminology & general concepts

Next

- In-depth understanding of domain

- Ability to communicate with people in various roles within domain (stakeholders)

@becomingdatasci   slide 49

Specialties ("Electives")

- Machine Learning and AI Specialties
  - Deep Learning
  - Autonomous Systems (robots)
  - Computer Vision
  - Natural Language Processing
- Advanced (or prettier) Data Visualization for publication

- Big Data Engineering
- Geographic Data & Mapping Techniques (GIS)
- Specific Domains
  - Research Science
  - Education
  - Politics
  - Marketing/Retail
  - etc etc...

On whether people should go through the entire Open Source Data Science Masters program she developed….

"Absolutely not! You should cherry-pick and be very strict about what you include, because you could spend your whole life on one of the topics that's under the umbrella [of data science]."

As a generalist, you need to figure out which of those topics is relevant to your path.

-Clare Corthell, @clarecorthell
Founder at Luminant Data
Episode 05

Find a niche

(experts are usually experts in one thing)

"It's a dangerous ideology,
the notion of a Super Genius."

-Safia Abdalla, @captainsafia, Data Scientist, Maintainer of
@nteractio, and PyData Chicago organizer, Episode 2

"Get experience. A lot of the data science
skills I've developed are out of interest… If I
have a problem that's in front of me, I'll figure
out what I need to learn to solve that problem"

-Justin Kiggins, @neuromusic, Neuroscientist/Data Scientist,
Episode 9

Total beginner?

Start by creating a report:
A basic data analysis that answers a
"business" question.

Renee's expectation for any data analyst or data scientist:

Business Question
⬇
Data Question
⬇
Data Answer
⬇
Business Answer

Don't just focus on learning technical
tools & techniques.

The non-technical skills are just as important.

What do you look for when you hire a data scientists and analysts?

"The #1 thing that I look for is curiosity... the value added is by saying 'I found this and I'm curious: how did that happen?'" And taking the analysis to another level.

-Sherman Distin, @ShermanDistin
Digital Marketing Executive & Consumer Analytics Consultant at Querybridge
Episode 4

"Soft Skills" That Came Up Often in Podcast Interviews...

- Curiosity

- Ability to communicate with a variety of people

- Ability to "hear" the actual problem that needs to be solved

- Desire to continue to learn

- Creativity

- Tenacity

  - "A critical part of the [data scientist] mindset" - Enda Ridge
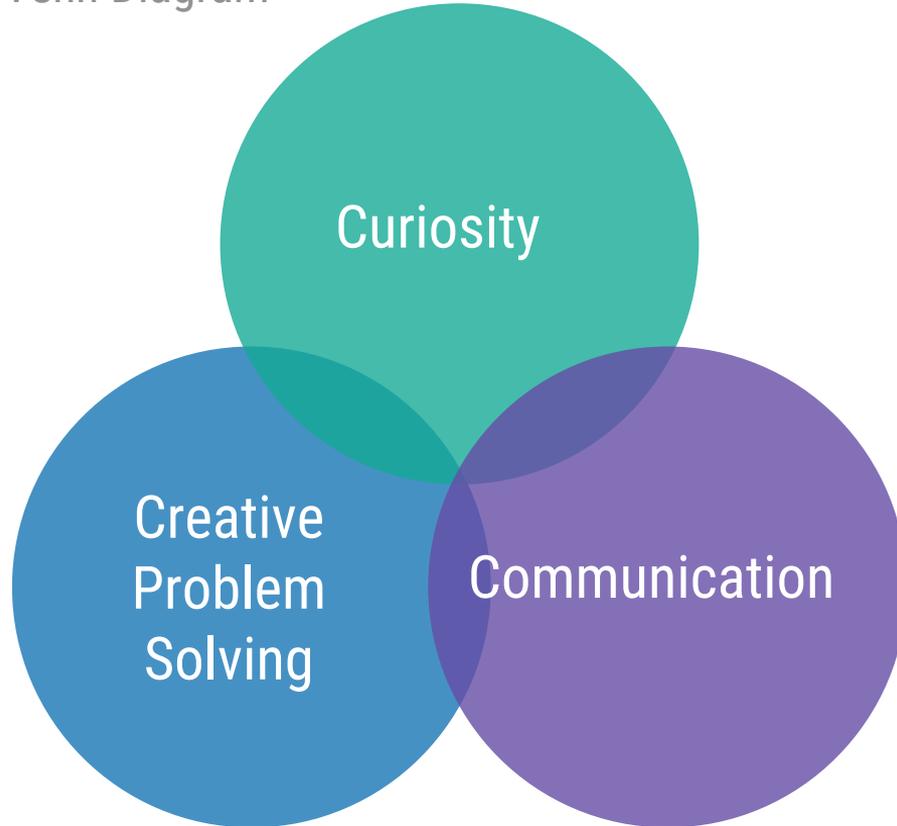   Head of Data Science, FMCG UK, Author of Guerrilla Analytics, @enda_ridge, Ep. 7

- Adaptability

  - "Overcoming unexpected difficulties" -Shlomo Argamon
    Director, Master of Data Science program at IIT, @ShlomoArgamon, Ep. 3

- Recognition of wider issues that come into play
  - Systems Thinking

Good Data Science Practice
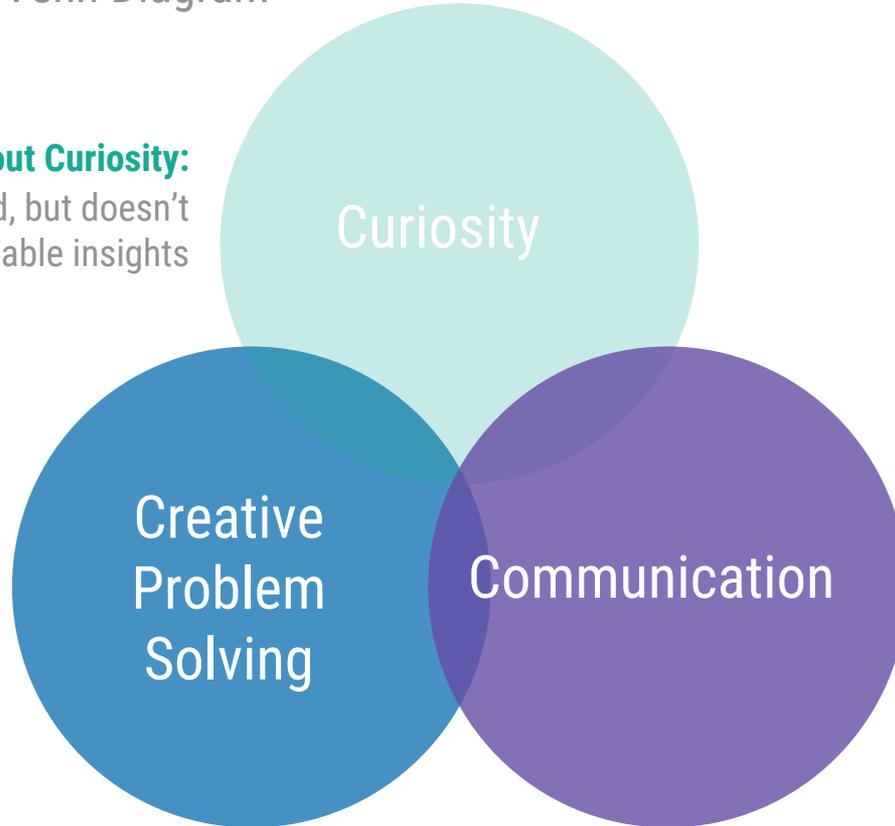
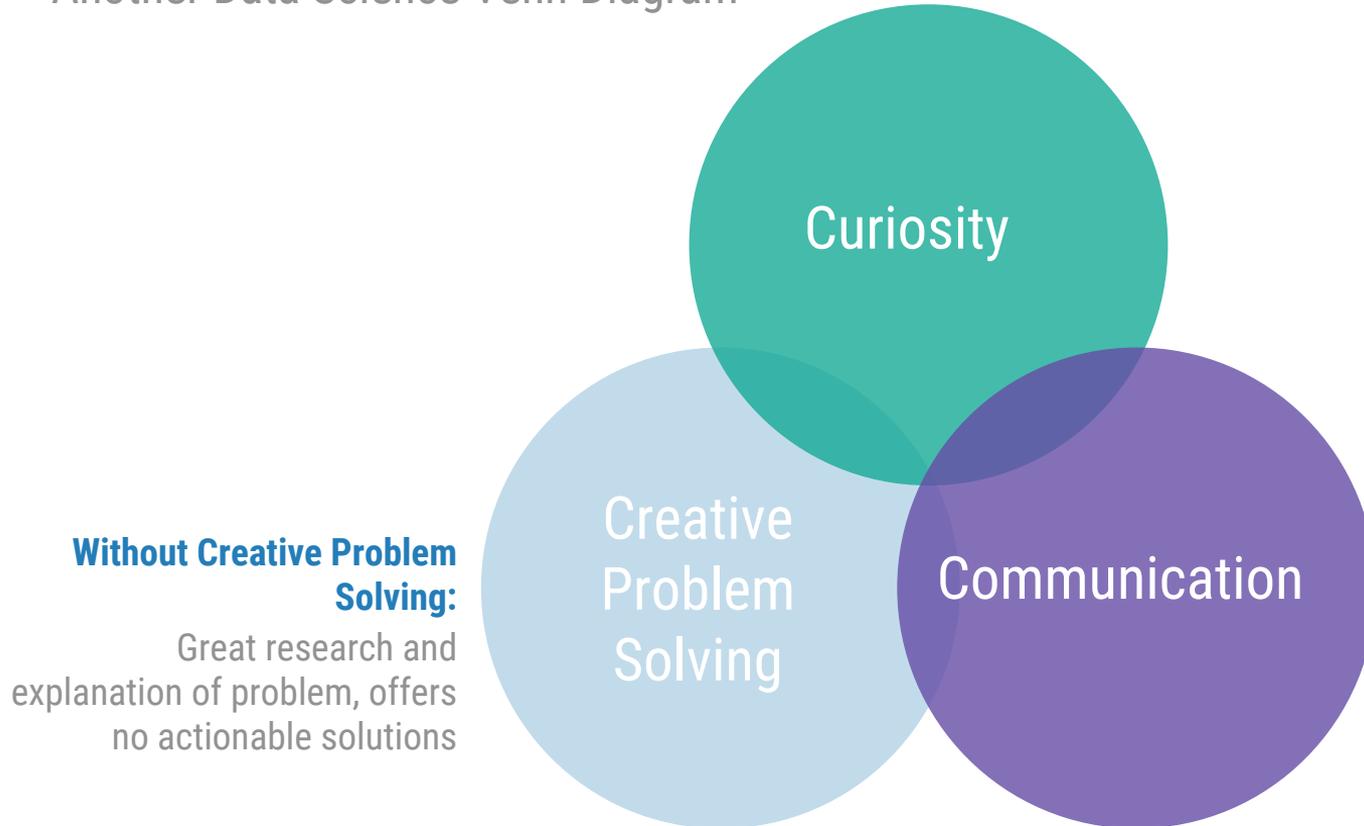Another Data Science Venn Diagram

Good Data Science Practice

Another Data Science Venn Diagram

Without Curiosity:
Answers questions asked, but doesn't go beyond to add valuable insights

Curiosity

Creative Problem Solving

Communication

Good Data Science Practice

Another Data Science Venn Diagram

Curiosity

Creative Problem Solving

Communication

Without Creative Problem Solving:
Great research and explanation of problem, offers no actionable solutions

Good Data Science Practice

Another Data Science Venn Diagram

Curiosity

Creative Problem Solving

Communication

Without Good Communication:
Unintelligible presentation of otherwise good results

At Intersection:
Good Data Science Practice

These 3 things were mentioned consistently by podcast guests as either traits they themselves possess, or as desirable factors that set candidates apart when they're hiring data scientists.

Curiosity

Creative Problem Solving

Communication
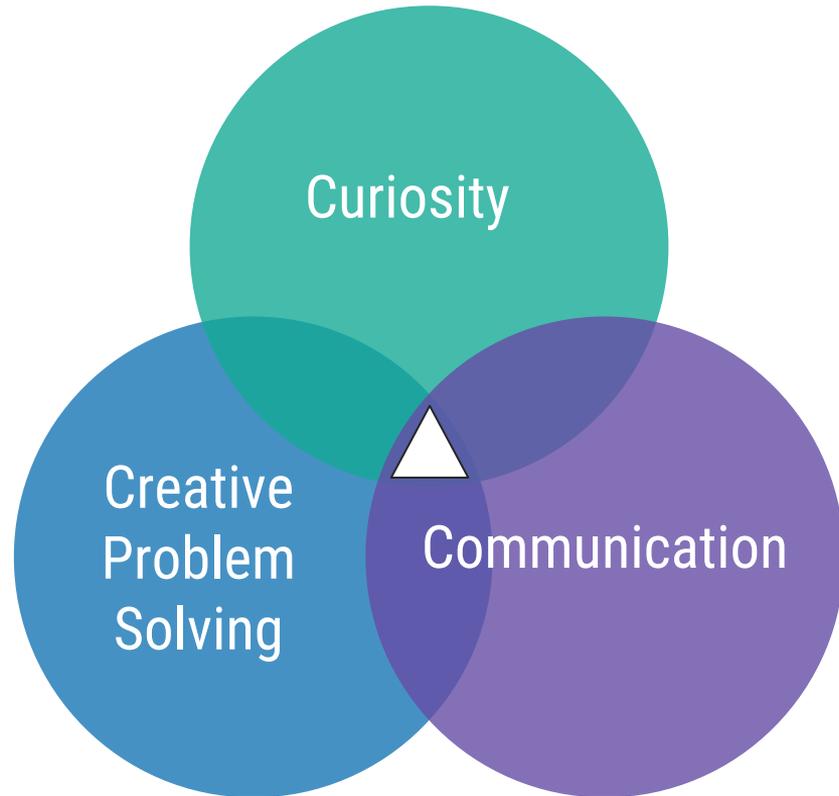
4.
How to Learn

So, once you have a plan for what to learn, how do you go about learning it?
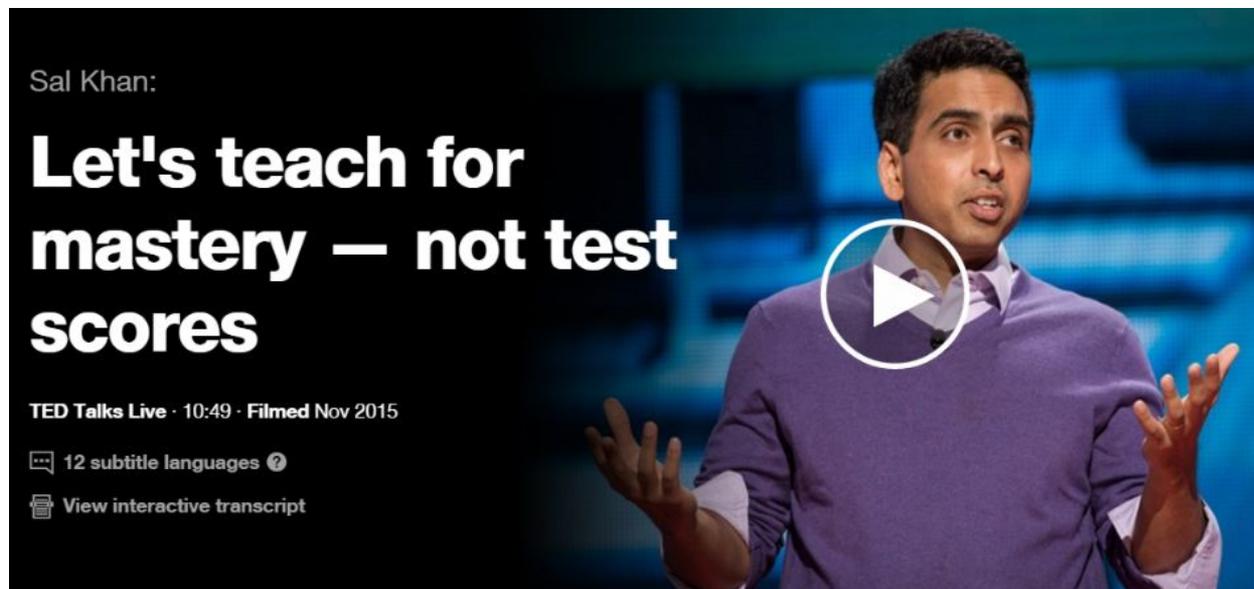
"It depends..."

Choose resources that work with
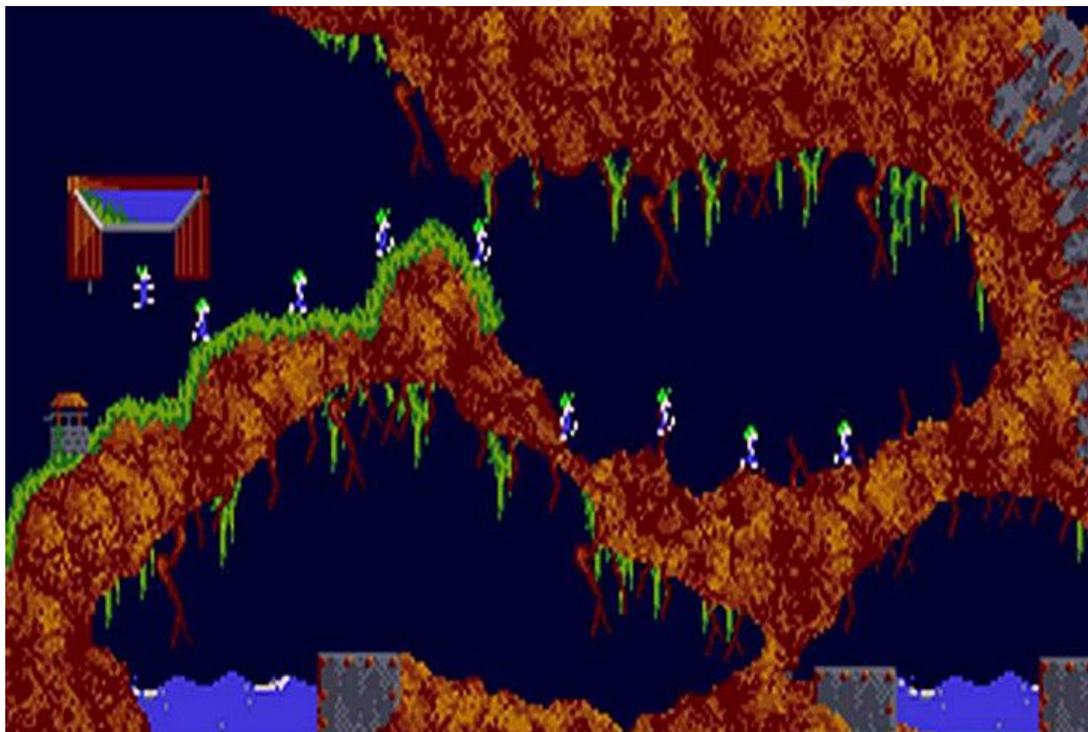your learning style, and customize accordingly.

Overcome your
Math & Programming Aversions

You don't have to love 'em, you just have to not hate 'em.

Chances are, you developed an aversion because of how they are taught in school.

You can try a different way now.

Sal Khan:

Let's teach for mastery — not test scores

TED Talks Live · 10:49 · Filmed Nov 2015

⌕ 12 subtitle languages ❓

🖶 View interactive transcript

If you have the patience, persistence, and problem-solving skills to play a level-based puzzle game, you can learn to program that game.

ANGRY BIRDS

The classic bird-flinging pig-popping game that started it all. Also the best-selling app in history.

5.
How to Practice
and Apply What You Have Learned

The most common advice is to start with a project related to something that matters to you.

Find a real-world dataset. Ask some questions that could be answered by analyzing it.

Each step of the project will mean learning a new topic, or practicing a new skill or technique.

Blog about your experiences to practice communicating (or record yourself talking about it). Get feedback

"My answer is always read lots of books, and study every night. And they ask, 'How else?!?' No one likes that answer!"

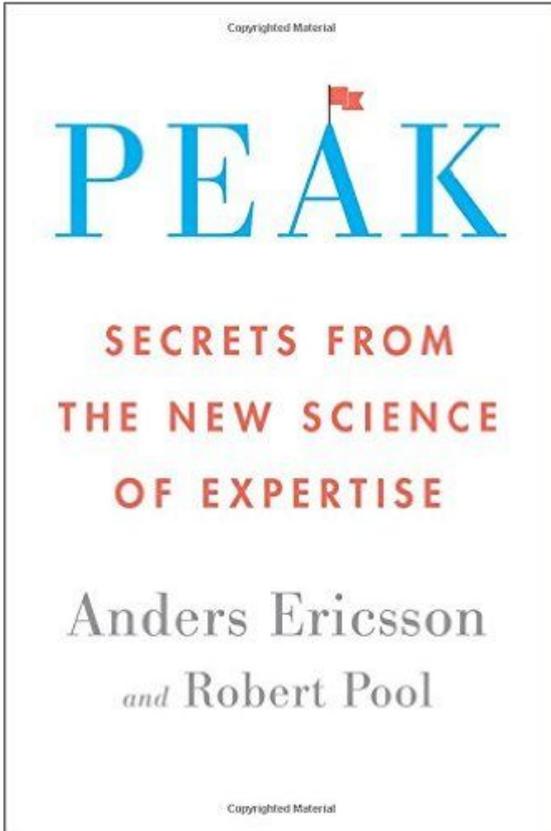-Will Kurt, @willkurt, Data Scientist at Quick Sprout
Episode 1

"How I learn best is really by struggling… If you are lifting weights, and you only lift the small weights, you will never become any stronger."

-Sebastian Raschka, @rasbt, Computational Biologist, Data Scientist, Author of Python Machine Learning
Episode 8

Deliberate Practice Concepts

Peak: Secrets from the New Science of Expertise

- "Inherent talent" doesn't determine who becomes an expert

- Requires consistent not-very-fun practice with a focus on pushing just beyond where you are

- You need feedback and guidance - get a good teacher or data scientists to learn along with

PEAK

SECRETS FROM
THE NEW SCIENCE
OF EXPERTISE

Anders Ericsson
and Robert Pool

6.
How to Know When You're Ready

(for your next career step)

Whoah…. I know Data Science!

"I think that people overestimate what you need to know to get started in this field."

-Erin Shellman, @erinshellman, Senior Data Scientist at Zymergen, Episode 6

I (Renee) felt ready when I had the foundational skills, was comfortable with the terminology, had accomplished a few difficult tasks, and realized I could learn anything else on the job.

And there's nothing wrong with doing a few "stretch" job interviews for practice. And who knows - you might get hired anyway!

In Summary

You've probably learned today that

"It depends..."

means you have to customize advice.

Hopefully I've given you some guidance on
how to do that, and therefore how to get started on the path
to your data science dream job!

1. Figure out where you're starting from and what kind of data scientist you want to be, and map out a basic learning path

2. Learn the "required" baseline topics: descriptive statistics, coding for data manipulation and simple machine learning, solid analytical understanding in domain of your choosing

3. Get good at a specific niche beyond the basics

4. Deliberately practice both technical skills and "soft skills" with real-world data (don't forget the "get feedback" part)

5. You'll never be done learning, so don't wait too long to try applying for data-science-related jobs. Go for it!

Most importantly: **Have fun!**

What is your advice to data science learners?

"To not let anyone take your dreams away, to know that they're there for a reason, and you have them because they're part of your essence. And if you choose to not listen to that voice that's telling you, one day you're going to pay for it (and be unhappy)... pursue your dreams no matter how hard you think they are in the moment.
And find a mentor."

-Debbie Berebichez, @debbiebere
Physicist, Chief Data Scientist at Metis
Episode 13
(pictured with Bill Nye the Science Guy)

@becomingdatasci   slide 80

Questions?

Renée Teate

@becomingdatasci on Twitter

Email: renee@becomingadatascientist.com

BecomingADataScientist.com (blog, learning club, podcast)

DataSciGuide.com (learning resource directory under development)

*See the slide notes
for more info + reference links.
Look for them at
becomingadatascientist.com*

*See the slide notes
for more info + reference links.
Look for them at
becomingadatascientist.com*

Presented by Renee Teate at PyData DC 10/9/2016
<u>Shared under Creative Commons Non-Commercial ShareAlike License</u>
(Share/remix, with attribution. Any derivative works must be similarly licensed. Not for commercial use.)

@becomingdatasci   slide 82

How to find
"Becoming a Data Scientist" Podcast

- Podcast & Show Notes on my blog

- RSS feed for podcast players

- YouTube playlist (subscribe to channel)

- iTunes

- Stitcher

Learning Resources

General Resources for Data Science Learning

- Data Science Learning Club

- DataSciGuide

- Open Source Data Science Masters (Clare)

- Guerrilla Analytics (Enda)

- Harvard CS109 Data Science

- Metis - Explore Data Science

- Metis - Intro to Data Science

- Lynda

- DataCamp

- Data School

- DataQuest.io

- Podcasts
  - Becoming a Data Scientist
  - Partially Derivative
  - O'Reilly Data Show
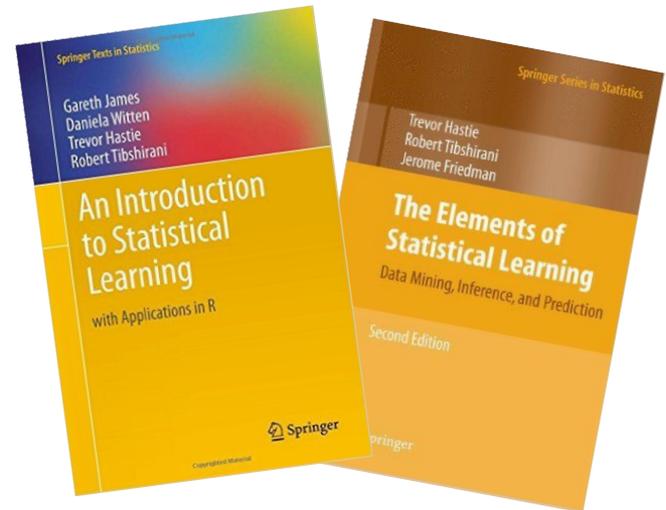 - What's the Point
  - (these & more on DataSciGuide)

Stats Learning Resources for Beginners

- Khan Academy

- Mathematical Monk (Safia rec)
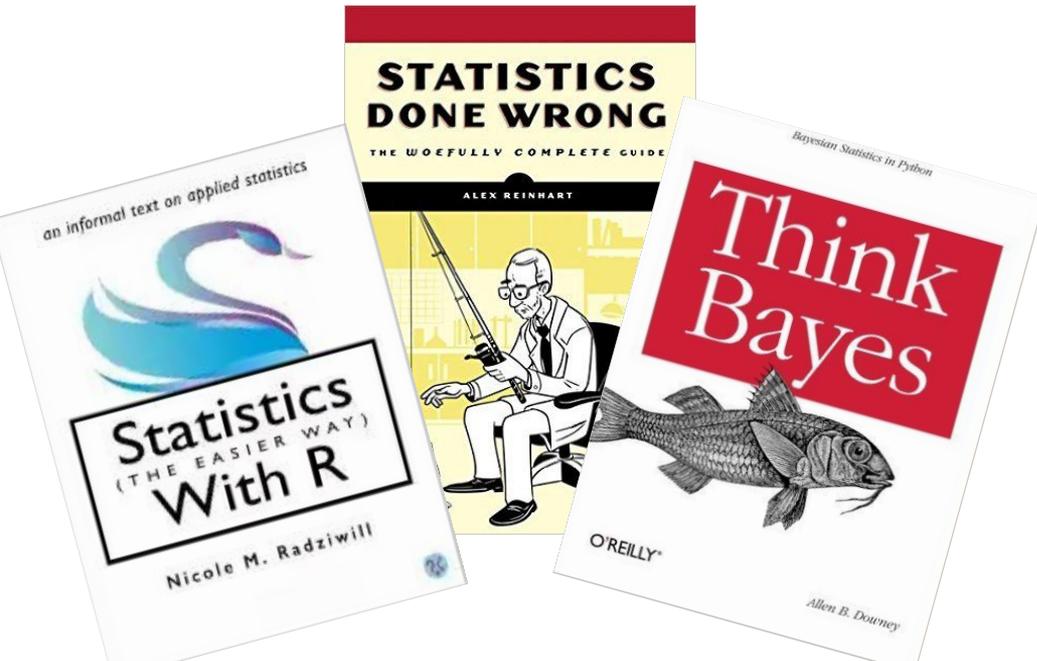
Stats Learning Resources for Data Science

- Count Bayesie Blog (Will's blog)

- Coursera

- More on DataSciGuide

Data Visualization Resources for Beginners

- Flowing Data blog & books

- Storytelling with Data

- The Truthful Art

- Stephen Few books

- Data Stories podcast

- More on DataSciGuide

Machine Learning Resources for Beginners

- Data Smart (Excel)

- Interactive Coding Courses
  (see DataCamp, etc. on Data Science slide)

- Coursera

- SciKit-Learn Docs

- Podcasts
  - Data Skeptic
  - Linear Digression
  - Talking Machines
  - These and more on DataSciGuide

Python Learning Resources for Beginners

- Talk Python to Me Podcast

- Learn Python the Hard Way
  (Safia recommendation)

- Anaconda / Jupyter

- Udacity

- Codecademy

- More on DataSciGuide

Python Learning Resources for Data Science

- DataCamp

- Scikit-Learn

- Python Machine Learning book & Jupyter Notebooks
  (Sebastian)

- Python for Data Analysis
  (Pandas)

- Fluent Python
  (not intended for beginners)

Keep trying different resources until you find the ones that work best for your current level and learning style.

This is only a sampling of what's available!

I'll continue to add more content (that you can rate!) to the searchable Data Science Learning Directory at
DataSciGuide.com