Learning How to Learn: Data Science

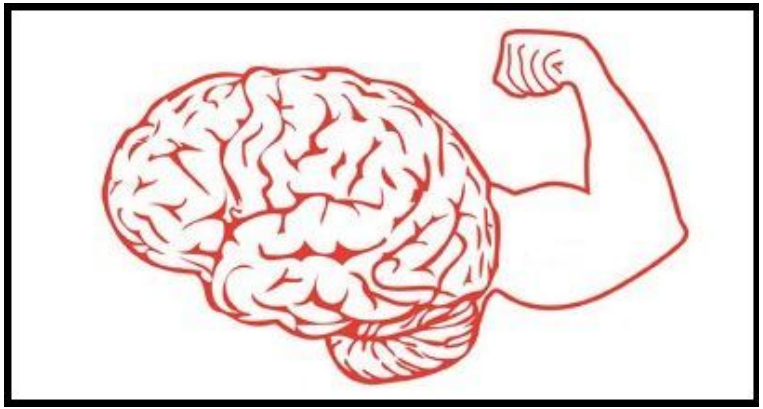
BDAA

BIG DATA & ANALYTICS ASSOCIATION

Image: Human Connectome Project

Why Teach Yourself?

- Empower yourself! Learn what ***you*** want to when and how you want to
- Be a better student
- Be more a more competitive job candidate



Pros & Cons of Self-Directed Learning:

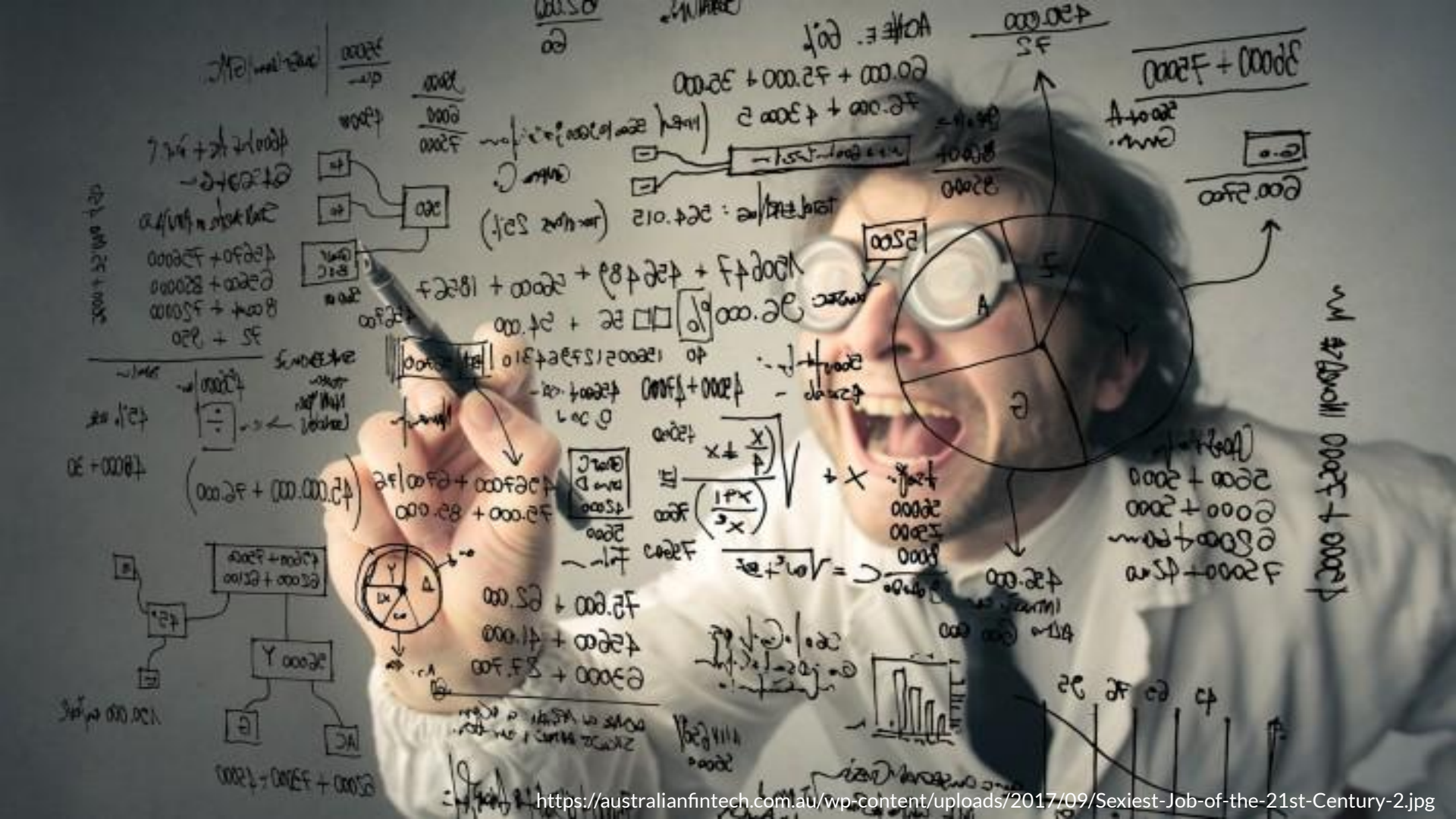
- You ***get*** to choose what you want to learn
- You ***get*** to choose how you want to learn
- You ***get*** to choose when you want to learn
- ***You are in control*** of your learning and have to take full responsibility for it

- You ***have*** to choose what you want to learn
- You ***have*** to choose how you want to learn
- You ***have*** to choose when you want to learn
- ***You are in control*** of your learning and have to take full responsibility for it

Why teach yourself Data Science?

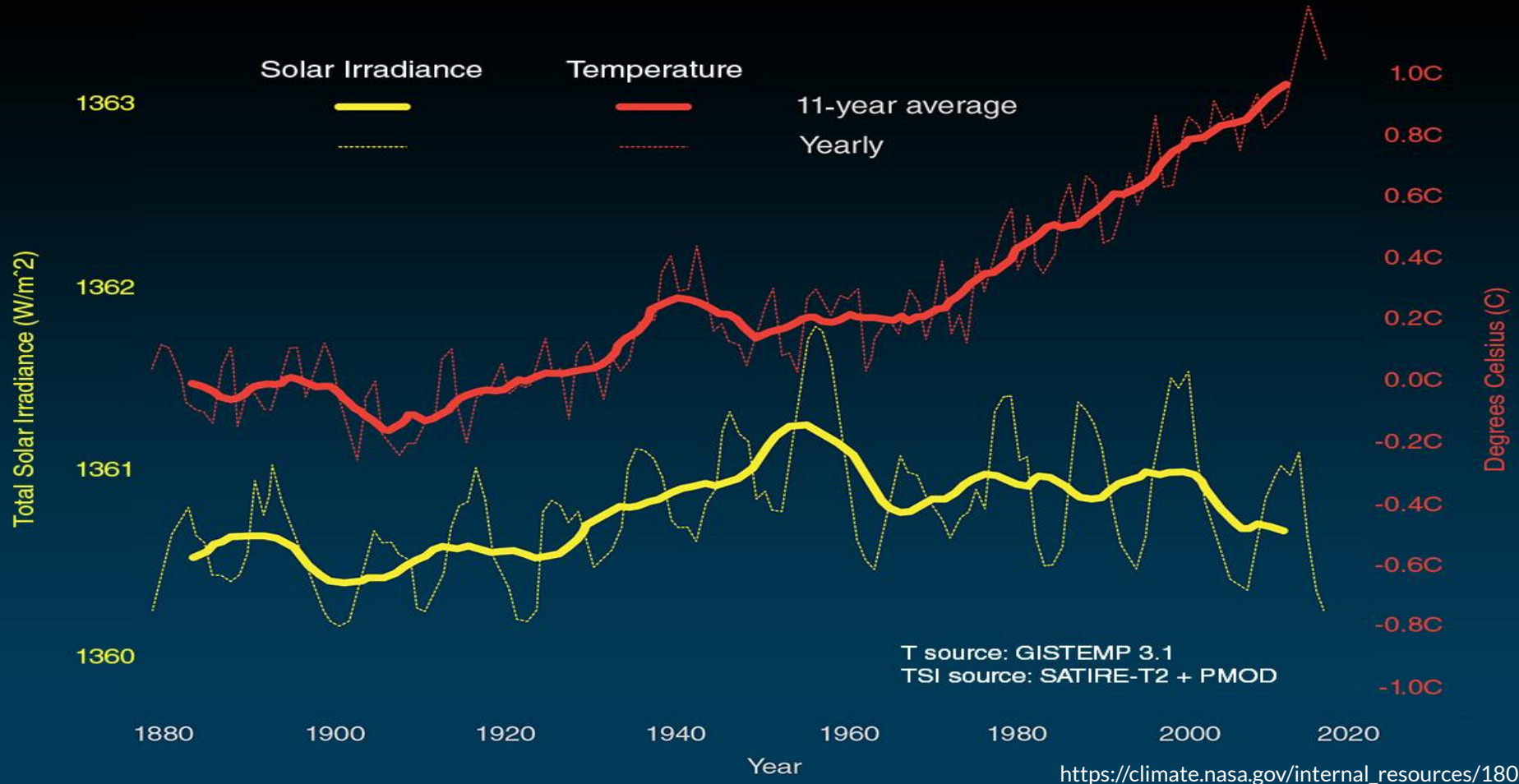
- You aren't a data analytics major, but you're interested in data science
- You are a data analytics major, but you're interested in something not covered by the major here
- Some classes just expect you to know things already
- To be successful in data analytics (and many other tech-driven fields), you must learn how to be a lifelong learner

So you want to be a Data Scientist?



https://australianfintech.com.au/wp-content/uploads/2017/09/Sexiest-Job-of-the-21st-Century-2.jpg

Temperature vs Solar Activity



https://climate.nasa.gov/internal_resources/1802

https://www.listenmoneymatters.com/wp-content/uploads/2018/05/LMM-Cover-Images-1-2.jpg

how to be a data scientist

All | Videos | News | Images | Shopping | More | Settings | Tools

About 339,000,000 results (0.50 seconds)

Ad · www.knowledgehut.com/Free_Webinar/Register

Explore Future of Data Science | Free Webinar on Trends of 2020

Register Now For **Data Science** Job Market Insights, Latest Courses & Expert Meet-N-Greets. Our Free Webinar Will Help You Choose The Best **Data Science** Program For Your Needs. 70+ Countries. 1000+ Enterprise Clients. 200+ Courses. 250K+ Trained.
Corporate Training · Free Informative Webinars

Ad · www.springboard.com/

Data Science Bootcamp Online | 100% Job Guaranteed

Curriculum by Experts. Get Hired or Money Back Guaranteed. Mentor Led program. Get Specialized in Machine Learning, Natural Processing. Download The Syllabus Now. 1:1 Personal Mentorship. Excel in **Data** Modelling. 40+ Hours Project Work.

Ad · www.simplilearn.com/

Become a Data Scientist | Work on 15+ Real-Life Projects

Learn Machine Learning, Python, R, Spark, Hadoop & Scala from Experts - Live, Online. Get...

Ad · www.indeed.com/Hire/Data-Scientist

Find a Data Scientist | Post a Job Online in Minutes | indeed.com

Post a Job for a **Data Scientist**. Reach 250M Job Seekers. Sign up & Start Today! Post Job.

There are three general steps to becoming a data scientist:

1. Earn a bachelor's degree in IT, computer **science**, math, physics, or another related field;
2. Earn a master's degree in **data** or related field;
3. Gain experience in the field you intend to work in (ex: healthcare, physics, business).

www.geteducated.com › careers › how-to-become-a-data-scientist

How to Become a Data Scientist with Online Education

towardsdatascience.com › how-to-become-a-data-scientist-3f8d6e754...

How to become a data scientist? - Towards Data Science

Jul 4, 2019 - Introduction: I am pretty sure that many of us come across the article from the Harvard Business Review back in 2012. A **data scientist** is a ...

towardsdatascience.com › how-to-become-a-data-scientist-2a02ed565...

How to Become a Data Scientist - Towards Data Science

Nov 6, 2019 - Becoming a **data scientist** pragmatically spending little money on movies, books, and courses available on the Internet.

www.discoverdatascience.org › career-information › data-scientist

How to Become a Data Scientist in 2020 ...

Learn what a **data scientist** does and the steps to becoming a **data scientist** to help determine if it is the right career path for you.
Steps to Become a Data ... · What does a Data Scientist do?

elitedatascience.com › become-a-data-scientist

How to Become a Data Scientist in 2019 (Hadouken!)

So you want to become a **data scientist**... that's fantastic! But as you may already know (or may soon find out), it's not quite that simple. In fact, you'll most likely ...

www.learnhowtobecome.org › data-scientist

Becoming a Data Scientist: Step by Step Guide

Apr 21, 2019 - **Data science** is hot right now. According to the University of Wisconsin's **data science** department, job postings in the field increased by a ...

insights.dice.com › 2019/11/20 › how-to-become-a-data-scientist

How to Become a Data Scientist - Dice Insights

Nov 20, 2019 - **Data science** is also very much a growth industry. ... Skills You Need to Become a **Data Scientist**; Getting into **Data Science**; **Data Scientist** ...

www.kdnuggets.com › 2018/05 › simplilearn-9-must-have-skills-data...

9 Must-have skills you need to become a Data Scientist, updated

Here are 9 key skills you need to be a **data scientist**. Master them and get your dream job in **data science**.

Towards Data Science
A Medium publication sharing concepts, ideas, and codes.

Essential Data Science Skills that need to be mastered:

- Programming
- Statistics
- Machine Learning
- Linear Algebra and Calculus
- Data Visualization
- Communication
- Data Wrangling
- Software Engineering
- Data Intuition

https://towardsdatascience.com/how-to-become-a-data-scientist-3f8d6e75482f

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing packages, e.g., R
- Databases: SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

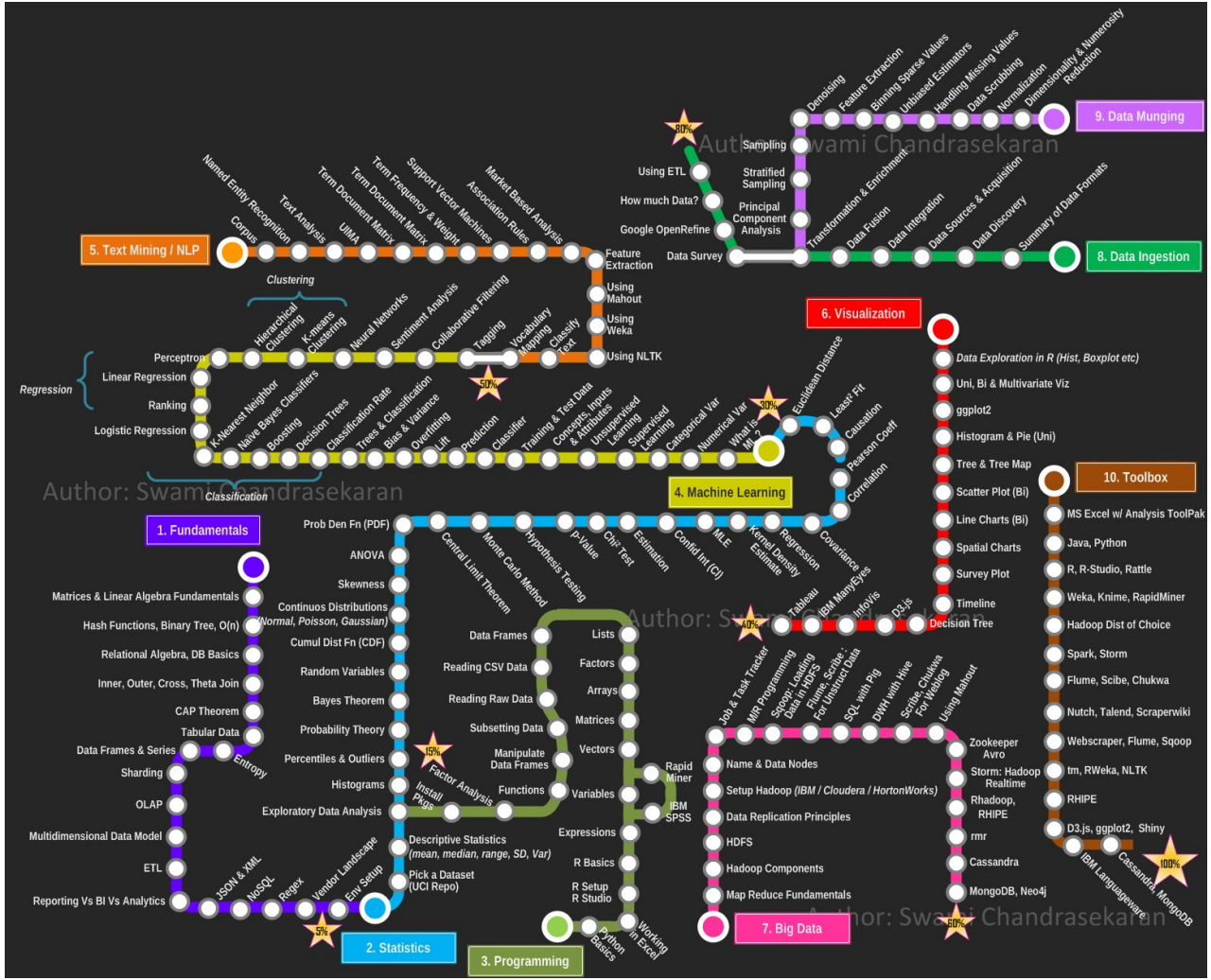
DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization: customer tracking and on-site analytics: predictive analytics and econometrics: data warehousing and big data systems: marketing channel insights in Paid Search, SEO, Social, CRM and brand.



https://towardsdatascience.com/how-to-become-a-data-scientist-2a02ed565336

https://www.listenmoneymatters.com/wp-content/uploads/2018/05/LMM-Cover-Images-1-2.jpg


My big data
too big
quickmeme.com

Data Science is a broad field, encompassing many other fields

The good news:

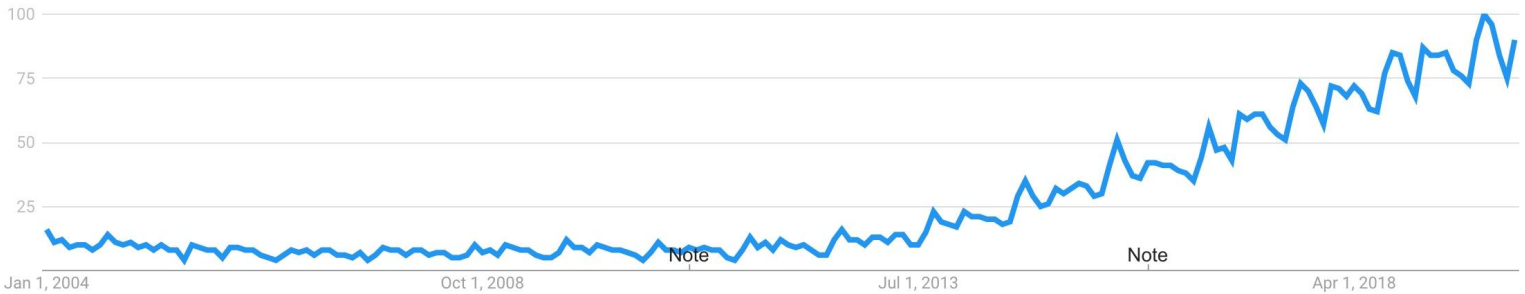
You don't have to learn EVERYTHING

data science
Search term

+ Compare

United States | 2004 - present | All categories | Web Search

Interest over time



https://trends.google.com/trends/explore?date=all&geo=US&q=data%20science

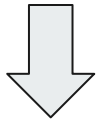
Data Science is constantly growing and evolving

The good news:

Data Science is about answering meaningful questions, NOT the tools and techniques

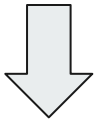
Solutions for Learning Data Science

- "Data Science" is VERY broad



- Can't learn everything overnight - must pick and choose

- Data Science is evolving



- Must be constantly learning to keep up

How to Learn Data Science?

METALEARNING

Image: Human Connectome Project

Metalearning

"learning about how knowledge is structured and acquired within this subject; in other words, learning how to learn it."

- Scott. H Young, author of *Ultralearning*

Steps to Learn Data Science

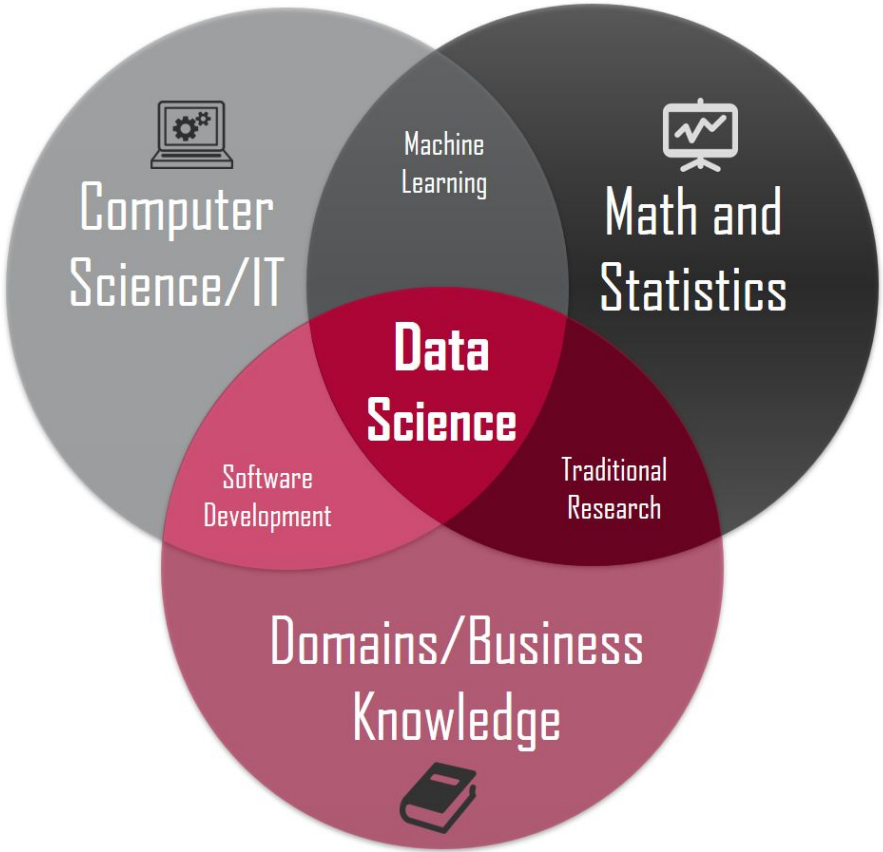
1. Figure out **WHAT** Data Science is
2. Figure out **WHY** you want to learn Data Science
3. Figure out **WHAT YOU** want to learn about Data Science
4. Figure out **HOW** you can best learn what you want
5. Make your plan
6. Overcome obstacles. Make mistakes and learn from them!

1. Figure out WHAT Data Science is

What is Data Science?

"Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to *extract knowledge and insights from* structured and unstructured *data*."

- Wikipedia (emphasis mine)

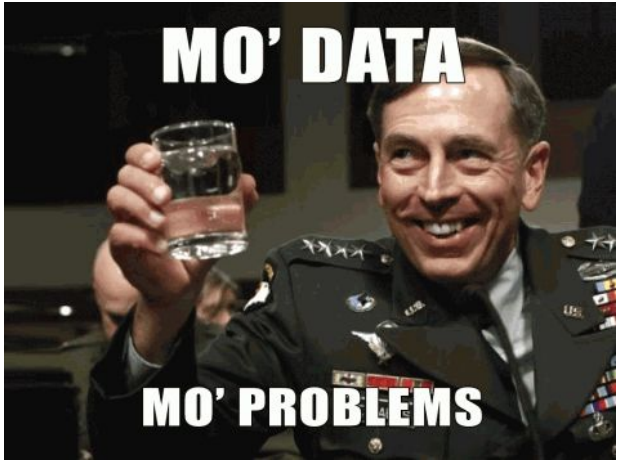


https://kainos-prod-assets.s3.amazonaws.com/uploads/2019/04/data_science.png

2. Figure out WHY you want to learn Data Science

What do you want to do with Data Science?

- Sports Analytics
- Health Science / Biomedical Informatics (Identifying & predicting disease, personalized healthcare recommendations, gene sequencing)
- Business Analytics
- Social Science Analytics
- Astronomy & Astrophysics
- … And many more!



Data Scientist vs. Data Analyst vs. Data Engineer

Data Scientist

- Senior position
- Uses advanced data techniques to derive conclusions
- Knowledge of:
  - Statistics
  - CS & ML
  - Domain Knowledge
- Tools:
  - Lots (R, Python, SQL, MapReduce, Java, etc.)

Data Analyst

- Entry-level position
- Generate reports based on organized data
- Knowledge of:
  - Statistics
  - Domain Knowledge
- Tools
  - Tableau
  - Excel
  - SPSS
  - SQL
  - Some R or Python

Data Engineer

- Intermediate position
- Organize and store structured and unstructured data on multiple machines
- Knowledge of:
  - CS
  - Domain Knowledge
- Tools:
  - Hadoop & MapReduce
  - Apache Spark
  - Java, Scala
  - SQL, NoSQL

3. Figure out WHAT YOU want to learn about Data Science

What do you need to know to do what you w to do?

- Interview those who are doing what you want to do
- Look at job descriptions and requirements

4. Figure out HOW you can *best* learn what you want

List and Prioritize Subjects/Topics

- Based on your research:
  - What are the most important topics to learn?
  - What is the natural ordering learn them in, if any? Simultaneous?

Types of Knowledge

Concepts

Things that need to be **understood**

Ex:

- How machine learning algorithms work
- What an eigenvector is

Facts

Things that need to be **memorized**

Ex:

- Statistical formulas
- Programming syntax

Procedures

Actions that need to be **practiced**

Ex:

- Programming
- EDA
- Thinking through problems analytically

Figure out How to Learn Your Topics

- How have others done it?
- What resources are out there?
- What methods work for you?
  - Past Experience
  - Experiment w/New Ones

Methods & Resources

- Classes at THE Ohio State University ™
- BDAA
- Online Courses
- Books
- Internet (Articles, Forums, YouTube, etc.)
- People
- Projects!
  - Self driven
  - Kaggle

5. Make your plan

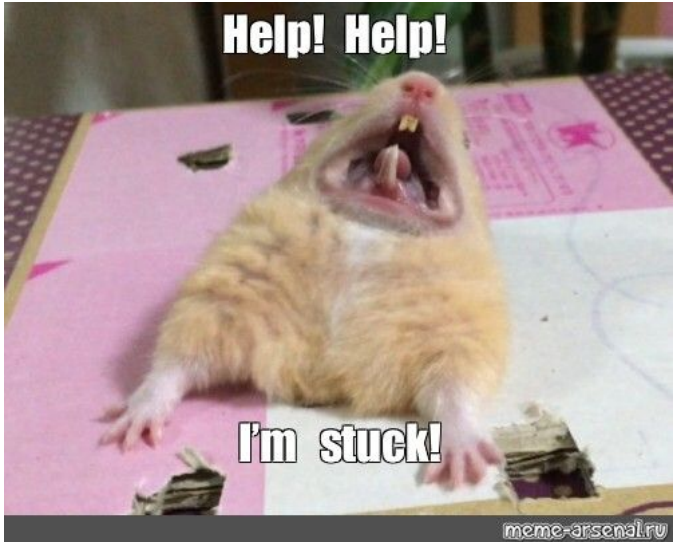
Make Your Plan & Set Goals

- What methods and resources will you use?
- When will you learn?
- Where will you learn?
- What do you want to accomplish?
- By when?

6. Overcome obstacles. Make mistakes and learn from them!

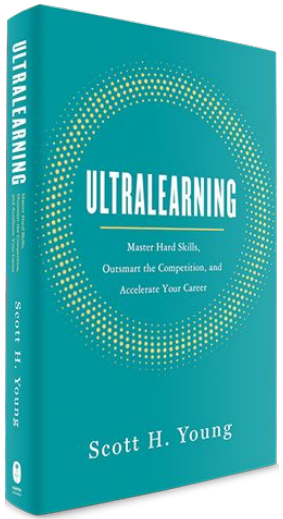
How to Overcome Obstacles

- Remember your WHY!
- Learn from Your Mistakes
  - Re-Evaluate your strategies
- Ask for help!
  - BDAA Slack
  - Me
  - BDAA Mentors & Eboard
  - Professionals on LinkedIn
  - Professors (Office hours!)



Resources

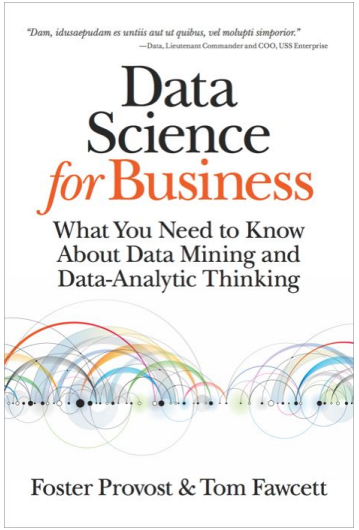
Self-Directed Learning & Studying



- Books:
  - Ultralearning - Scott H. Young
  - Deep Work - Cal Newport
- Internet
  - YouTube: Thomas Frank
  - Marty Lobdell - Study Less Study Smart
- People: Me! Happy to give tips and feedback on learning plans & strategies, BDAA slack
- Projects
  - Your own learning project helps you learn how to learn
- Classes at OSU:
  - Dennis Learning Center Courses
- Online Courses:
  - Coursera: Learning How to Learn

General Data Science

- BDAA!
- Online Courses:
  - Coursera: IBM Data Science Professional Certificate
- Books:
  - *Data Science for Business* - Foster Provost and Tom Fawcett
- Internet
  - Reddit: https://www.reddit.com/r/datascience/
- People: BDAA Slack, BDAA Eboard, Mentors
- Projects



R

- BDAA Workshops!
- Books:
  - *R for Data Science* - Hadley Wickham and Garrett Grolemund
- Internet
  - Tidyverse Website
- Online Courses:
  - Coursera: Data Science: Foundations using R Specialization
- People: BDAA Slack, BDAA Eboard, Mentors, Professors
- Projects
  - Kaggle
  - Your Own
  - Datafest @ OSU
- Classes at OSU:
  - STAT 3201 and 3202 (has prerequisites and is more stats focused, not just learning R)

Statistics & Analytical Thinking

- Classes at OSU
  - Stats Courses
- Online Courses:
  - Coursera: Introduction to Probability and Data
- Internet
  - Khan Academy Statistics
- People: BDAA Slack, Mentors, Data Analytics Learning Center
- Projects
  - Kaggle: EDA, Fitting Models, Etc.
  - Datafest @ OSU
- Books:
  - *Data Science for Business* - Foster Provost and Tom Fawcett

Python

- Classes at OSU:
  - ~~CSE 4194~~ CSE 4256 now
- BDAA Workshops! This Thursday!
- Online Courses:
  - Coursera: Python for Everybody
- Books:
  - Beginner: *Python Crash Course: A Hands-On, Project-Based Introduction to Programming* - Eric Matthes
  - Beginner: *Head-First Python: A Brain-Friendly Guide* - Paul Barry
  - Intermediate: *Fluent Python* - Luciano Ramalho
- Internet
  - Codecademy - Interactive Tutorials
- People: BDAA Slack, Mentors
- Projects
  - Kaggle

Machine Learning

- BDAA Workshops! Feb 13!
- Online Classes:
  - Coursera
- Books:
  - *Data Science for Business* - Foster Provost and Tom Fawcett
  - *Hands-On Machine Learning with Scikit-Learn and TensorFlow* - Aurélien Géro
- Internet:
  - Youtube: Siraj Raval
- Classes at OSU (have prereqs):
  - CSE 5243: Introduction to Data Mining (3 cr hrs)
  - CSE 3521: Survey of Artificial Intelligence I: Basic Techniques (3 cr hrs)
  - CSE 5523: Machine Learning and Statistical Pattern Recognition (3 cr hrs)
  - CSE 5524: Computer Vision for Human-Computer Interaction (3 cr hrs)
  - CSE 5526: Introduction to Neural Networks (3 cr hrs)

Tableau

- Classes at OSU:
  - ISE 5760: Visual Analytics and Sensemaking (3 cr hrs; Prereq: Jr, Sr, or Grad standing)
- Internet:
  - Tableau Training

Databases & Data Management

- Classes at OSU (Some prereqs):
  - CSE 3241: Introduction to Database Systems (3 cr hrs)
  - CSE 3244: Data Management in the Cloud (3 cr hrs)
- Online Classes:
  - Coursera: Databases and SQL for Data Science
  - Coursera: Hadoop Platform and Application Framework
  - Coursera: Data Engineering with GCP Professional Certificate

Finding Your Own

- Google!
- Coursera (Free Audit)
- YouTube
- Books - Goodreads
- Kaggle

Projects - Kaggle

https://www.kaggle.com/

Food for Thought

Our Brains are complex and POWERFUL

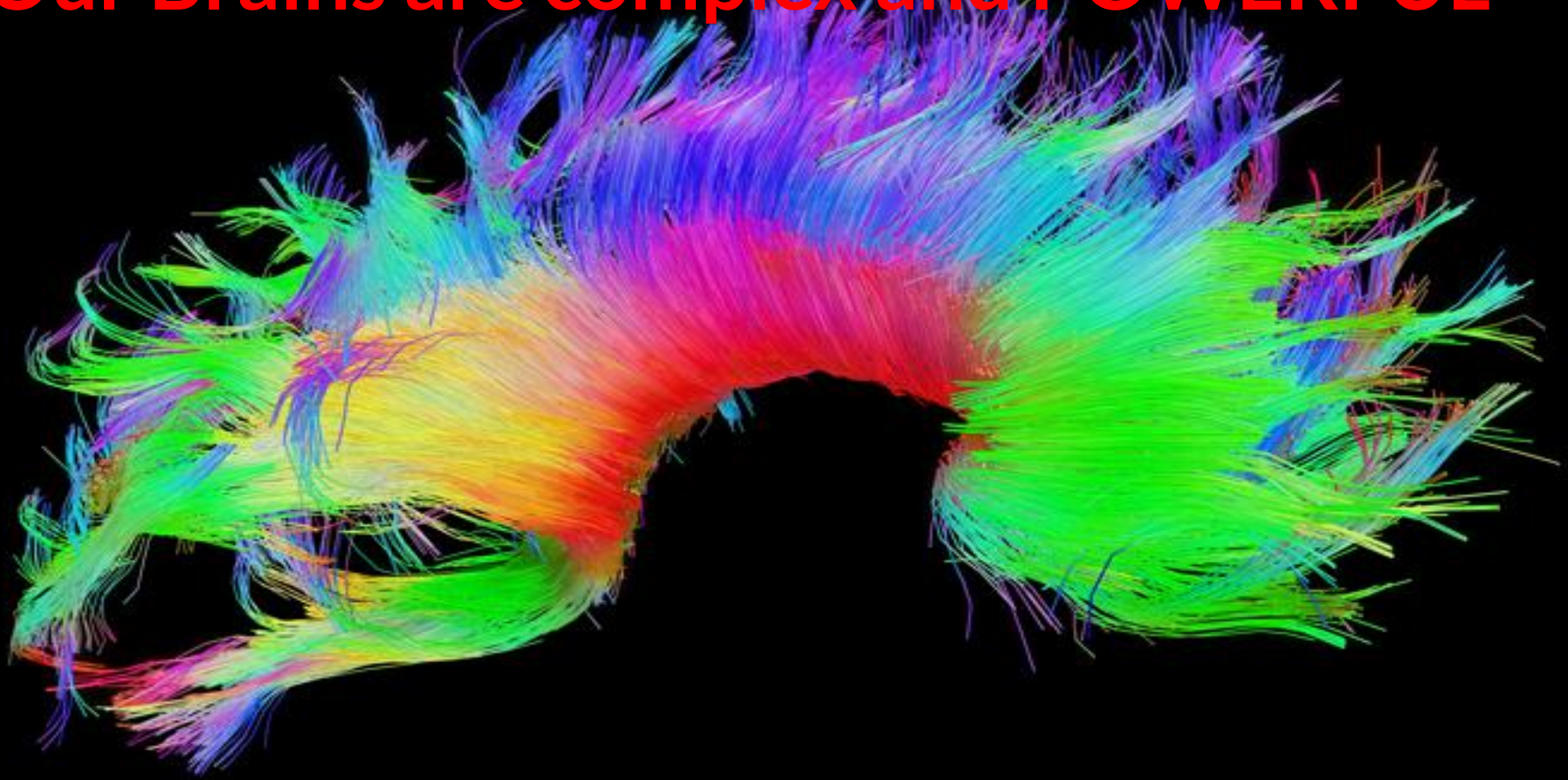
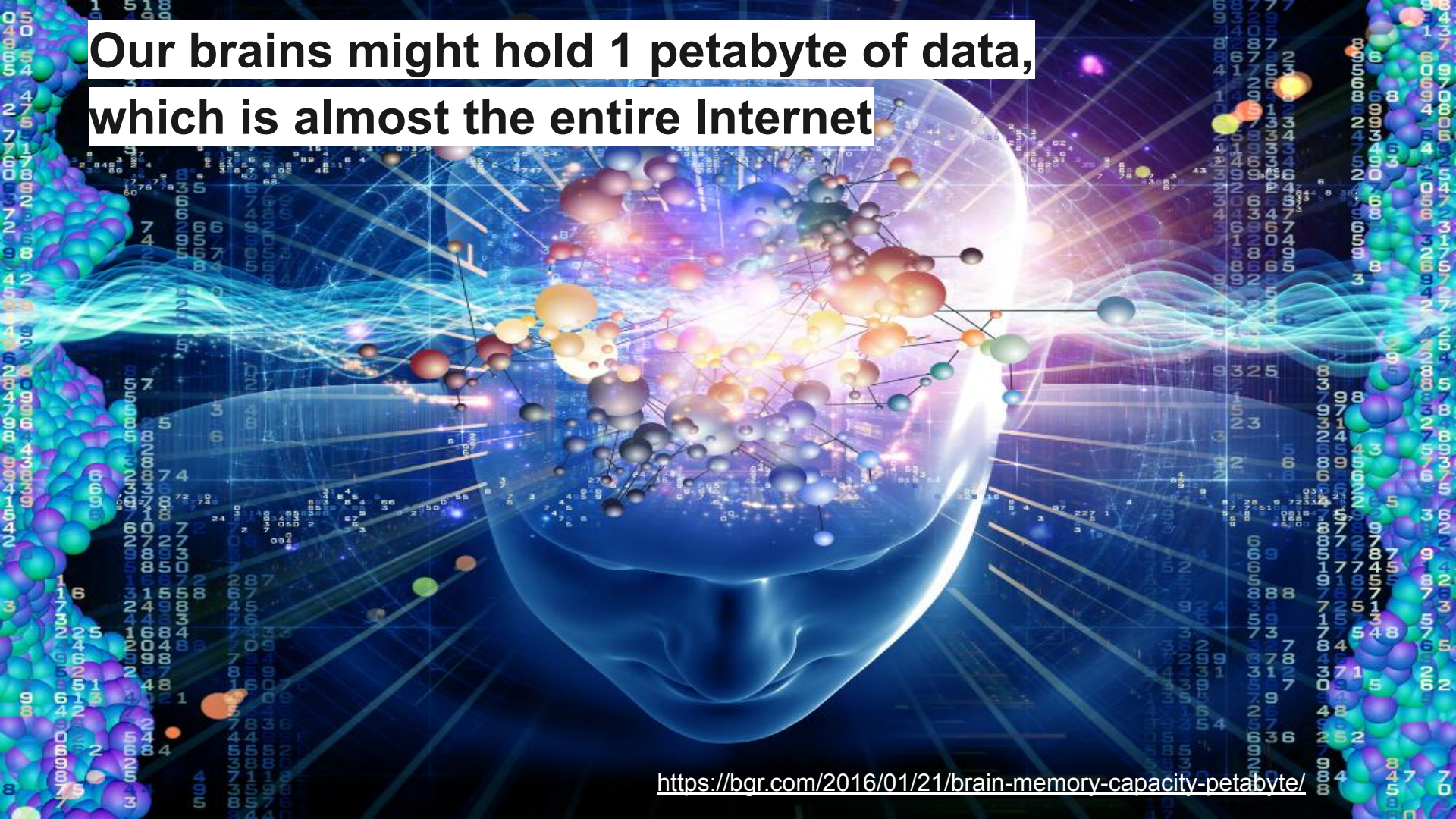


Image: Human Connectome Project


Our brains might hold 1 petabyte of data,
which is almost the entire Internet
https://bgr.com/2016/01/21/brain-memory-capacity-petabyte/


"Failure is an opportunity to grow"
GROWTH MINDSET
"I can learn to do anything I want"
"Challenges help me to grow"
"My effort and attitude determine my abilities"
"Feedback is constructive"
"I am inspired by the success of others"
"I like to try new things"
"Failure is the limit of my abilities"
FIXED MINDSET
"I'm either good at it or I'm not"
"My abilities are unchanging"
"I don't like to be challenged"
"I can either do it, or I can't"
"My potential is predetermined"
"When I'm frustrated, I give up"
"Feedback and criticism are personal"
"I stick to what I know"

Embrace the productive struggle.
Good luck!

Questions?

- Any Questions?

- My Contact Info:
  - Leo Glowacki
  - Message me on Slack!
  - www.leoglowacki.com