Beyond the Buzzword

A look at data science in practice and how you can be a part of it

Diversity in Data Science and Machine Learning

October 18, 2019

bit.ly/beyond-the-buzzword

Maria Tackett
Duke University

🎓 B.S. in Mathematics

✗ Minor in Computer Science

✗ Minor in Economics

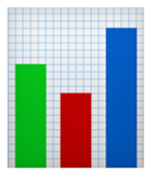
✗ Minor in German

B.S. in Mathematics

✗ Minor in Computer Science

✗ Minor in Economics

✗ Minor in German



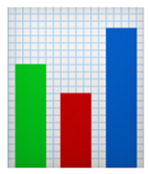
First statistics class junior year

B.S. in Mathematics

✗ Minor in Computer Science

✗ Minor in Economics

✗ Minor in German



First statistics class junior year

M.S. in Statistics

Statistician @ Capital One

Design of Experiments

Regression Modeling

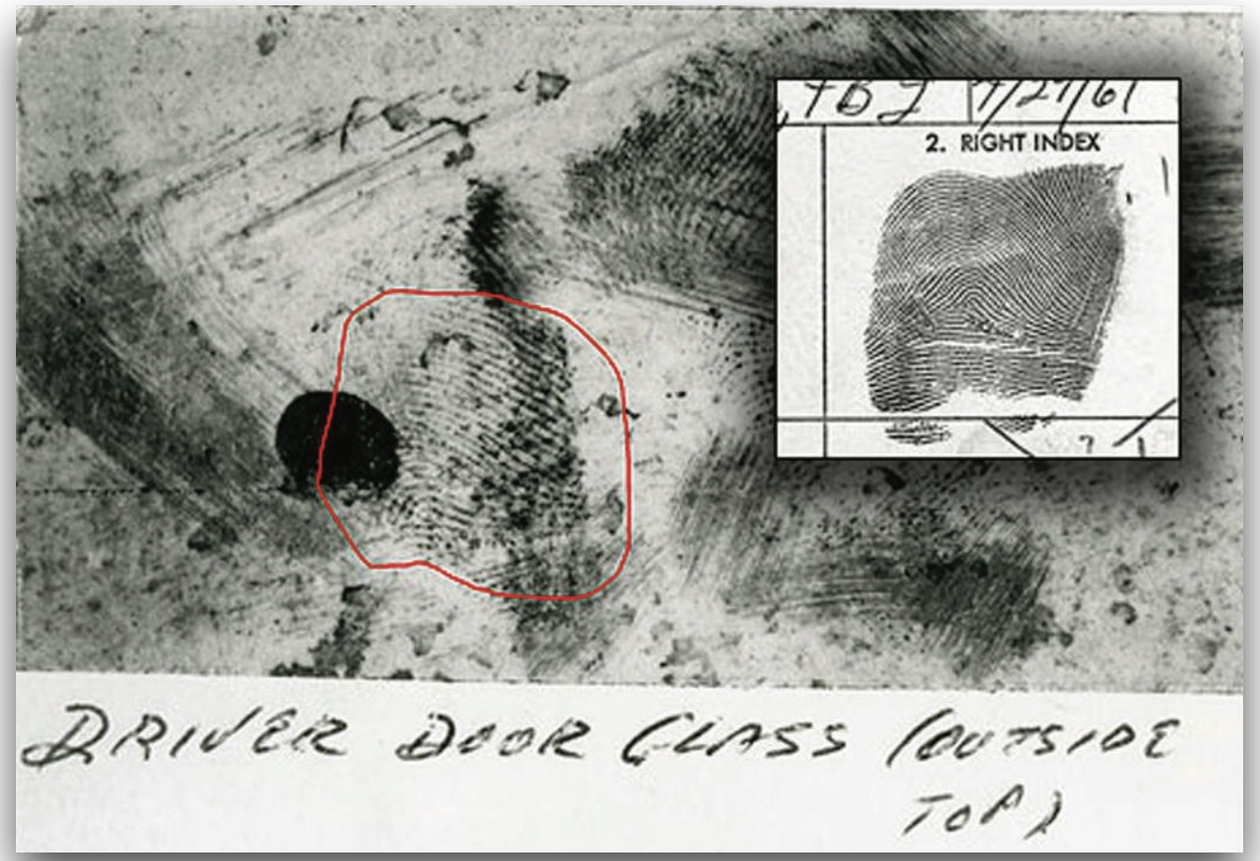
🎓 Ph.D. in Statistics

Dissertation Research

Used Bayesian methods to...

... decompose the sources of variability in two fingerprints

... develop a method to quantify certainty in fingerprint examiner's decision, adjusting for the number of fingerprints examined



Assistant Professor of the Practice @ Duke UNIVERSITY

Research

Statistics Education

Statistics in Criminal Justice and Forensic Science

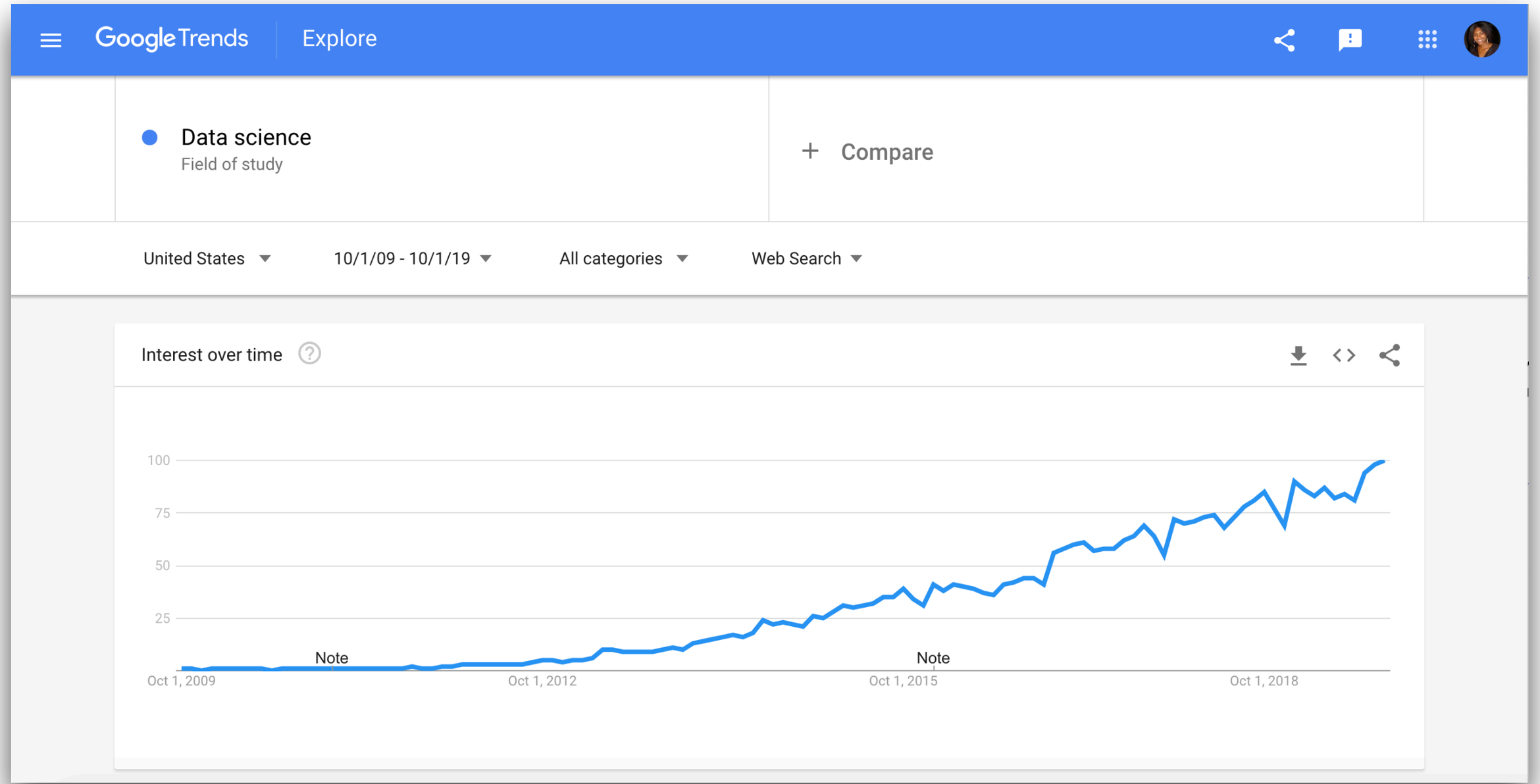
Teaching

Intro to Data Science

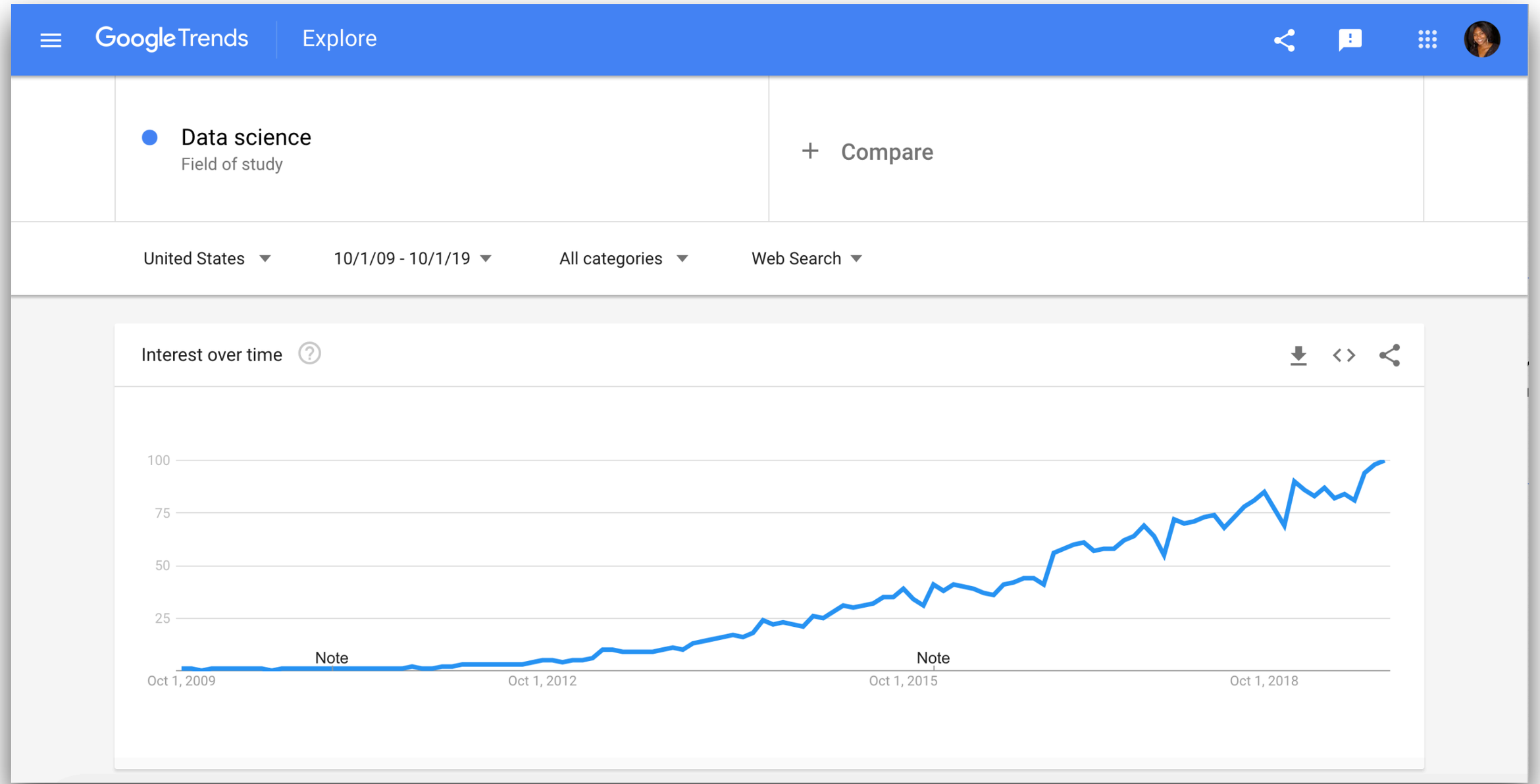
Regression Analysis

What exactly is
data science?

Data science has become increasingly popular over the past 10 years...

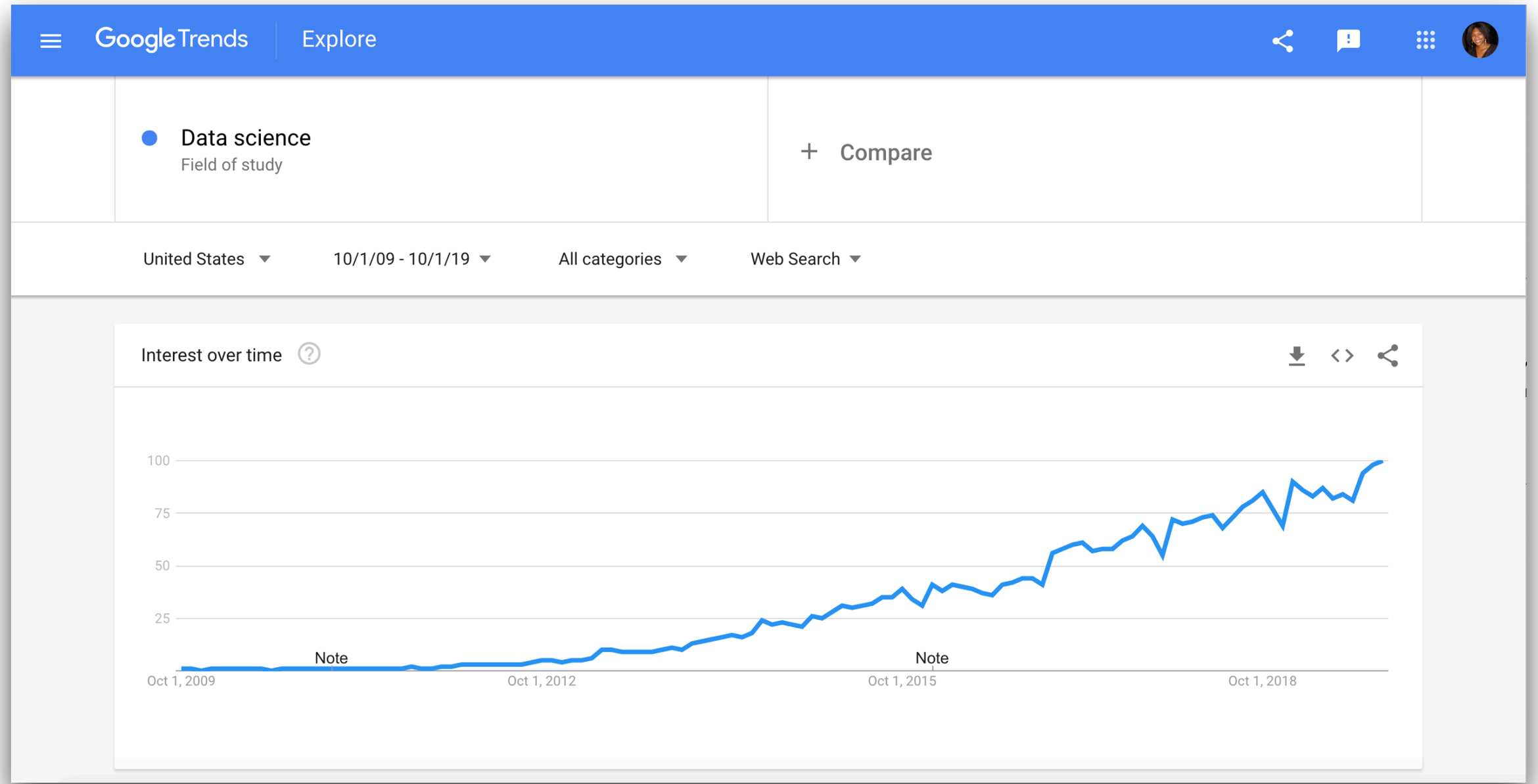


….but what exactly does “data science” mean?



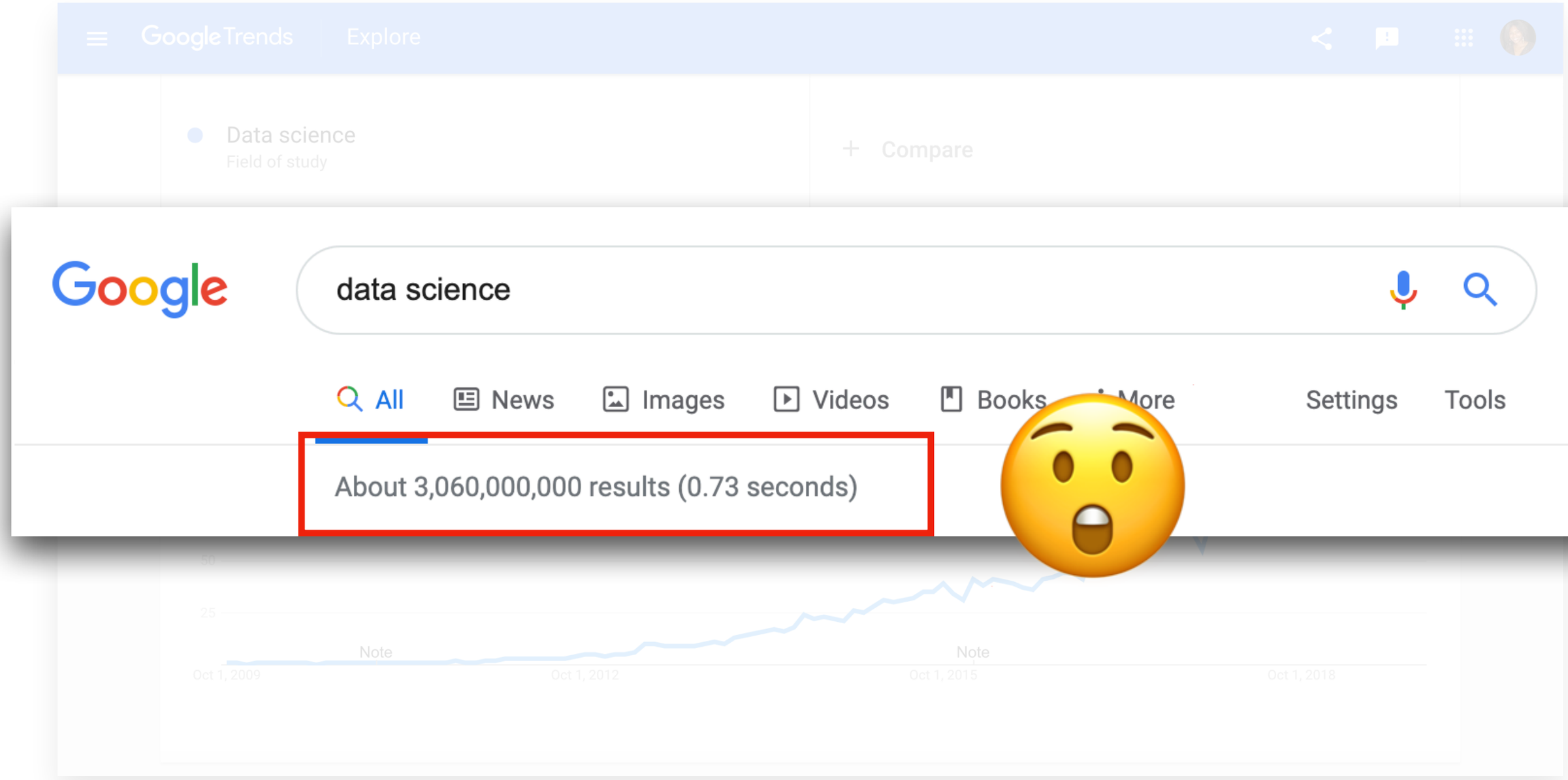
….but what exactly does “data science” mean?

Let’s ask Google…



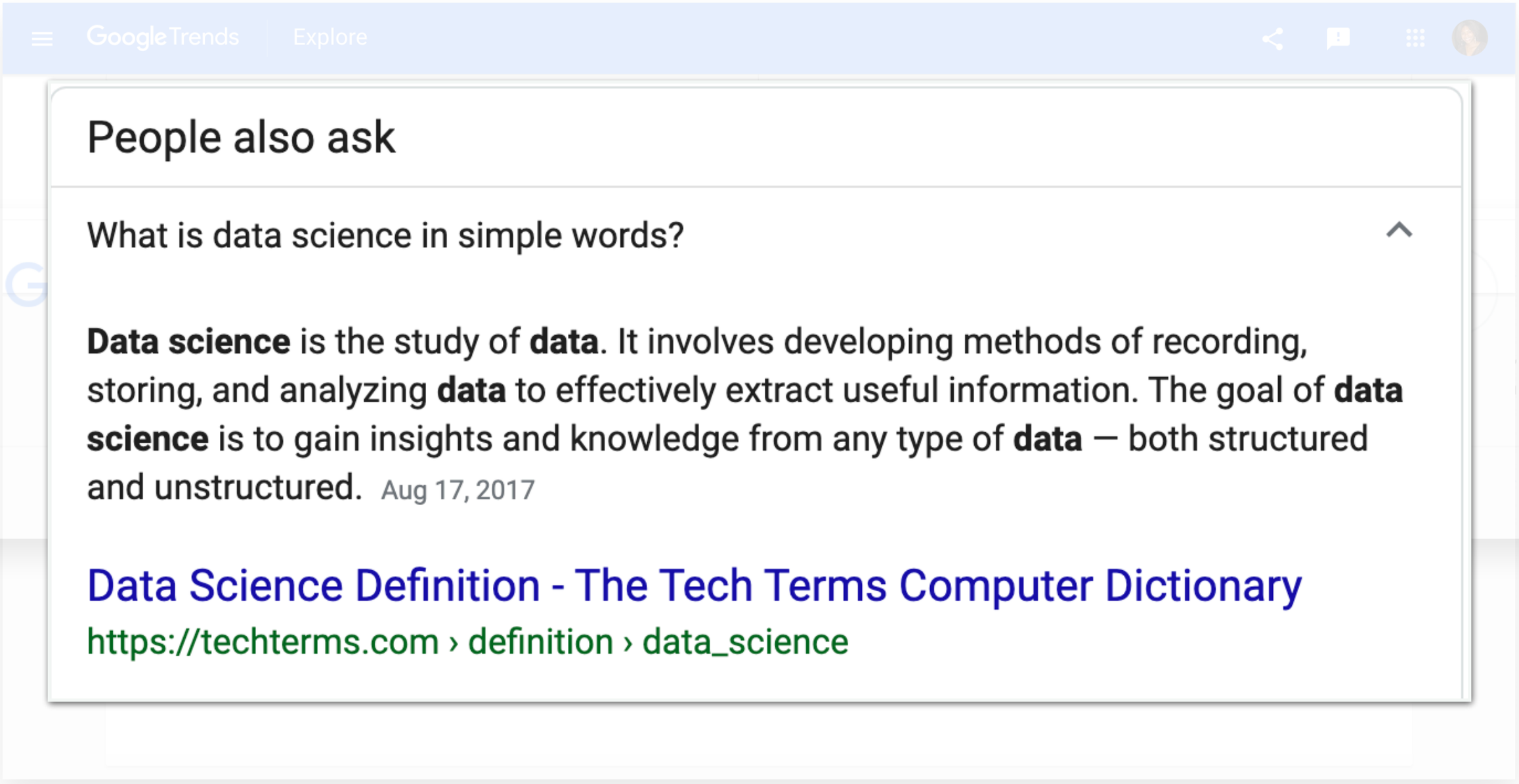
....but what exactly does “data science” mean?

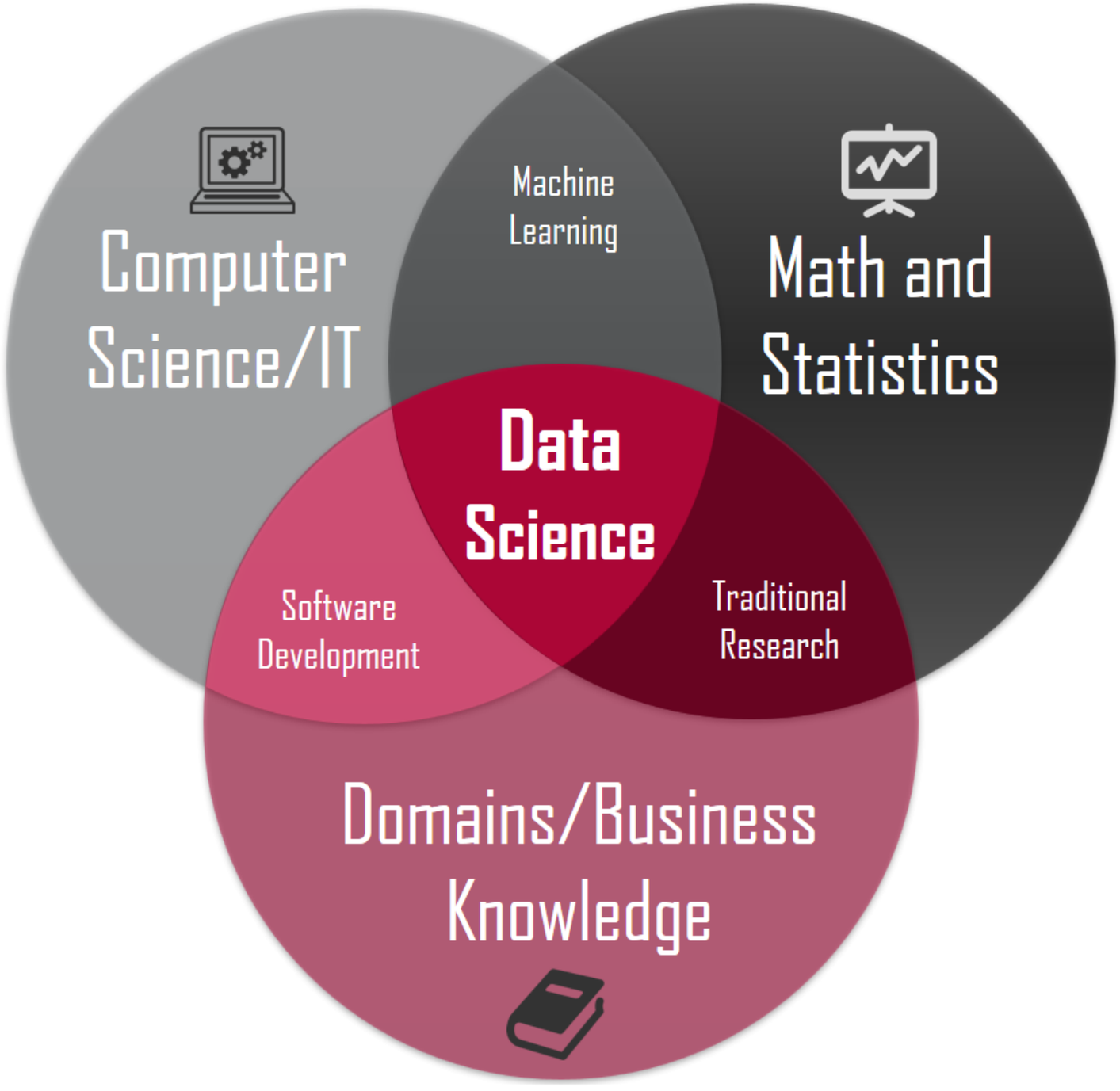
Let’s ask Google...



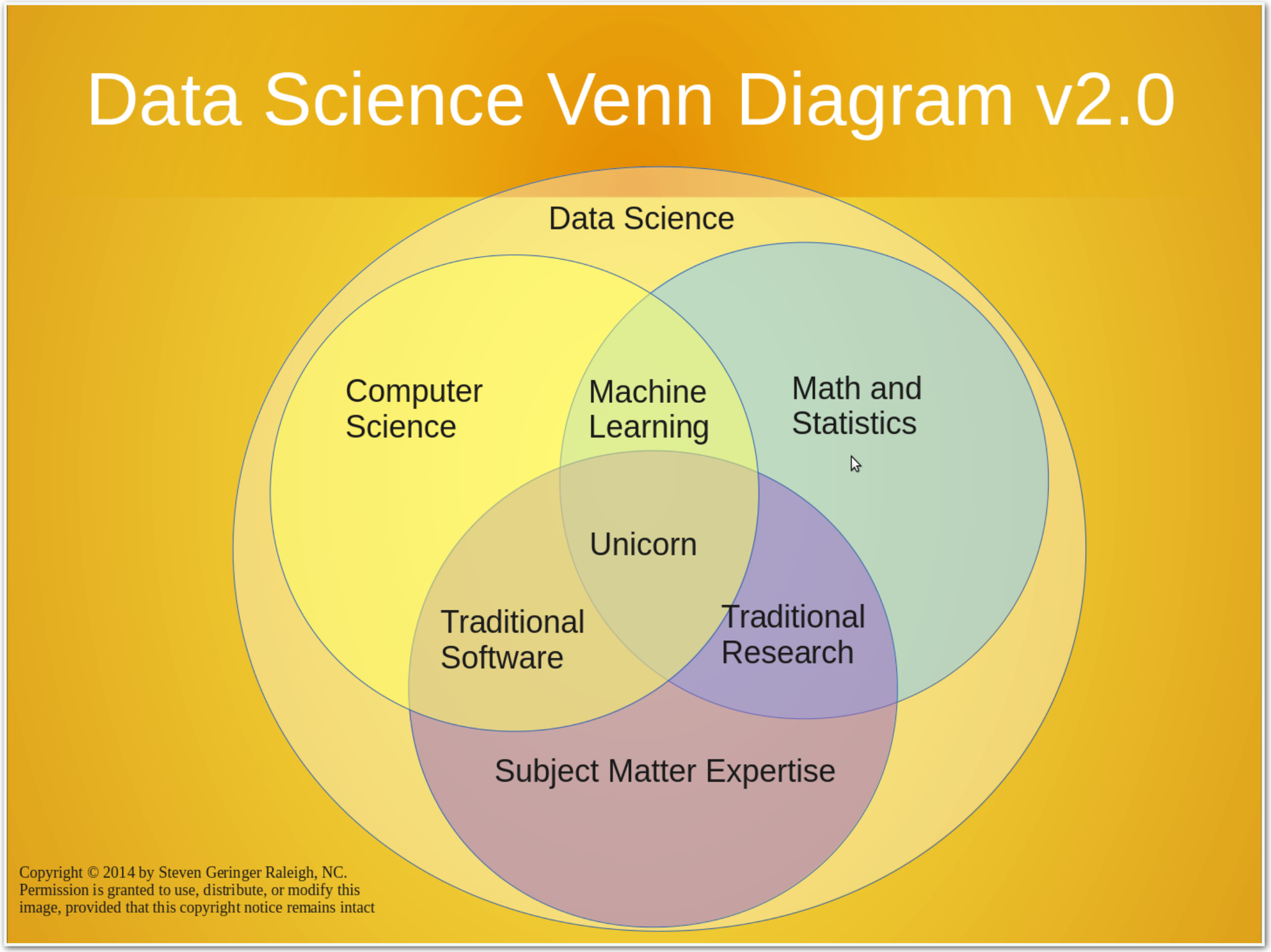
....but what exactly does “data science” mean?

Let’s ask Google...

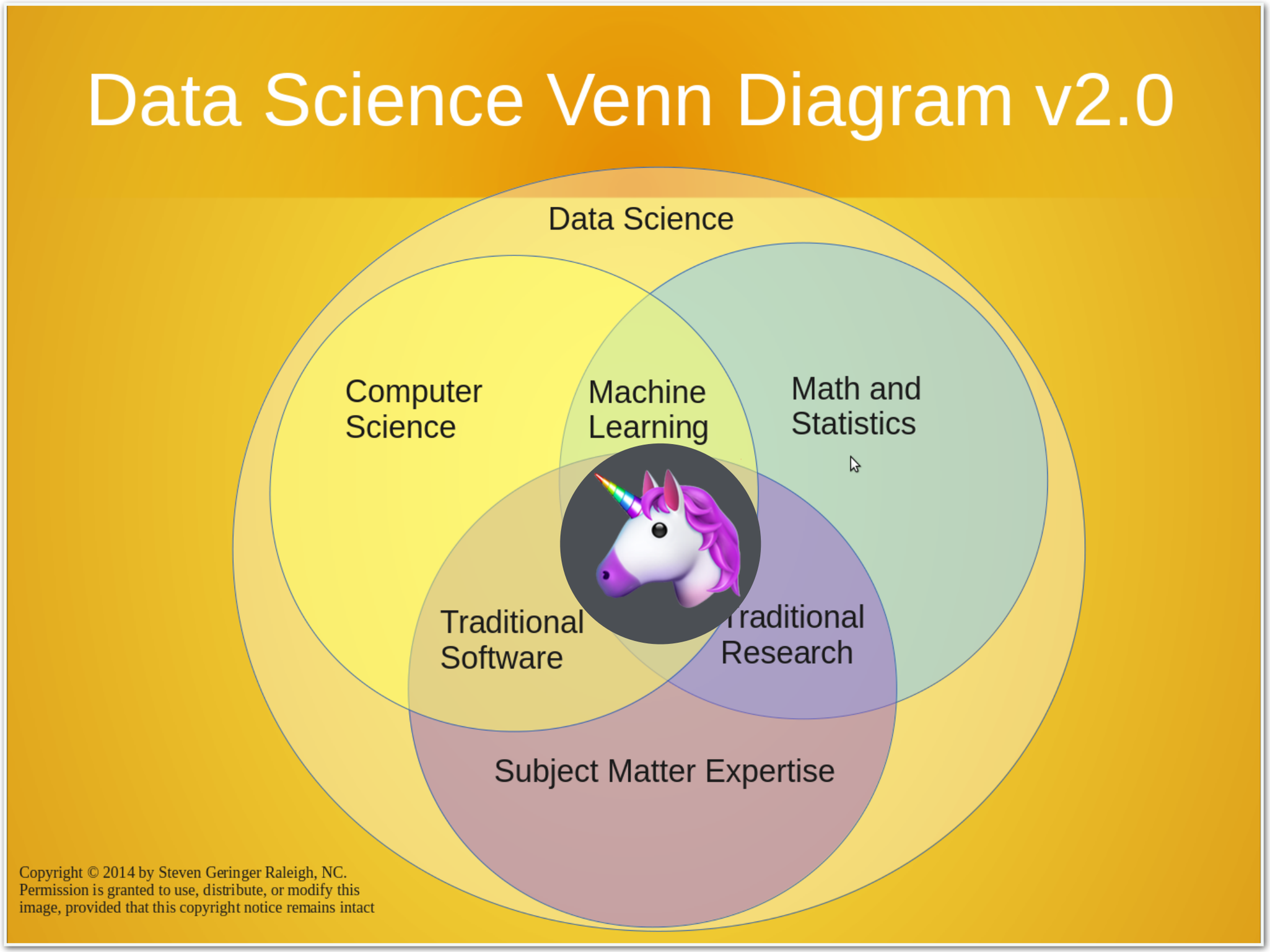




http://www.towardsdatascience.com



http://www.anlytcs.com/2014/01/data-science-venn-diagram-v20.html



http://www.anlytcs.com/2014/01/data-science-venn-diagram-v20.html

What data scientists do...

- ✓ Identify relevant questions
- ✓ Collect data from a variety of data sources
- ✓ Organize information
- ✓ Translate results to solutions
- ✓ Communicate their findings

UC Berkley - What is Data Science: https://datascience.berkeley.edu/about/what-is-data-science/

What data scientists have...

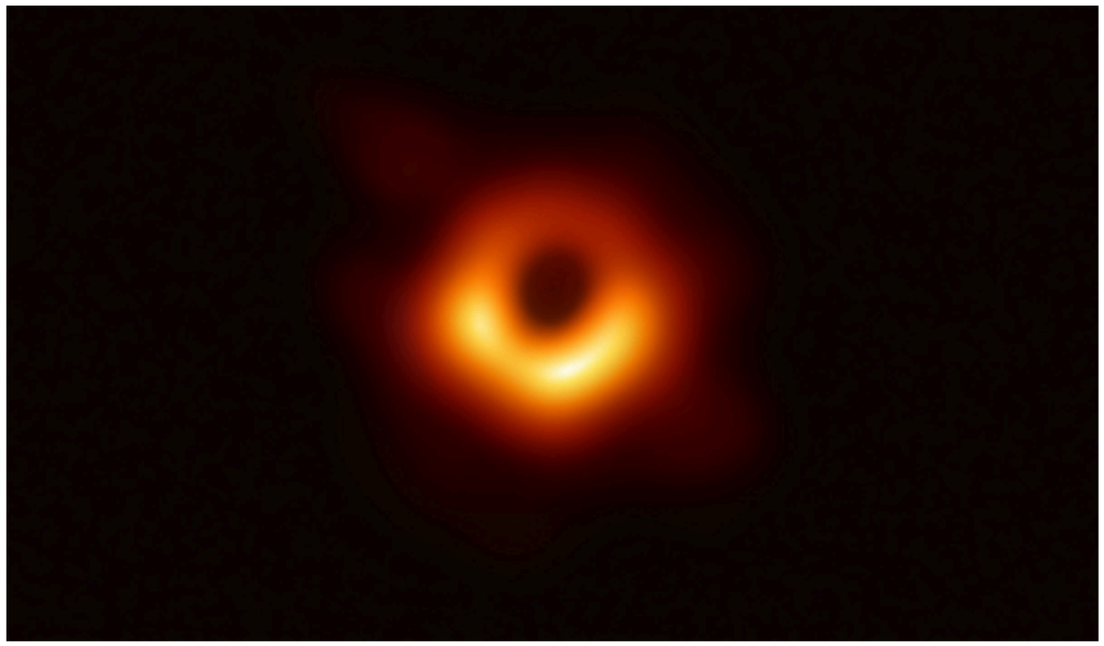
- ✓ Curiosity and results-oriented focus
- ✓ Industry-specific knowledge
- ✓ Ability to explain highly technical results to non-technical peers
- ✓ Background in statistics
- ✓ Background in computing

UC Berkley - What is Data Science: https://datascience.berkeley.edu/about/what-is-data-science/

SOPHIA CHEN SCIENCE 04.10.2019 09:45 AM

Scientists Reveal the First Picture of a Black Hole

The Event Horizon Telescope has captured a photo of a supermassive black hole at the center of M87, a galaxy 54 million light-years away.



The Event Horizon Telescope has captured a photo of a supermassive black hole at the center of M87, a galaxy 54 million light years away. EVENT HORIZON TELESCOPE COLLABORATION ET AL.

SOPHIA CHEN SCIENCE 04.10.2019 09:45 AM

Scientists Reveal the First Picture of a Black Hole

The Event Horizon Telescope has captured a photo of a supermassive black hole at the center of M87, a galaxy 54 million light-years away.

The Event Horizon Telescope has captured a photo of a supermassive black hole at the center of M87, a galaxy 54 million light years away. EVENT HORIZON TELESCOPE COLLABORATION ET AL.

NEWS · 10 APRIL 2019

Black hole pictured for first time — in spectacular detail

The Event Horizon Telescope's global network of radio dishes has produced the first-ever direct image of a black hole and its event horizon.

Davide Castelvecchi

The Event Horizon Telescope network has obtained the first image of a black hole — at the centre of the galaxy M87. Credit: EHT Collaboration

SOPHIA CHEN SCIENCE 04.10.2019 09:45 AM

Scientists Reveal the First Picture of a Black Hole

The Event Horizon Telescope has captured a photo of a supermassive black hole at the center of M87, a galaxy 54 million light-years away.

NEWS · 10 APRIL 2019

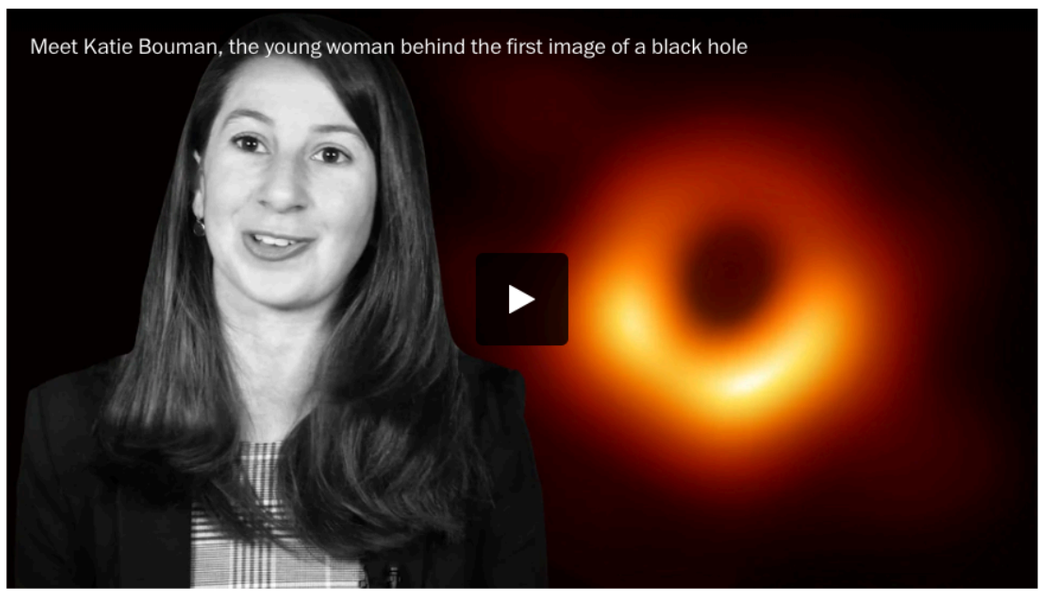
d for first time — in spectacular

bal network of radio dishes has produced the first-ever direct horizon.

of a black hole — at the centre of the galaxy M87. Credit: EHT Collaboration

Science

Algorithms gave us the black hole picture. She's the 29-year-old scientist who helped create them.



Katie Bouman was a MIT postdoctoral student when she led a team that designed one of the algorithms that led to capturing the first images of a black hole. (Adriana Usero/The Washington Post)

By **Ben Guarino**

April 10, 2019 at 3:49 p.m. EDT

How data science made it possible...

We have telescopes distributed around the world. For every two telescopes in the telescope array, we a a single spatial frequency, which tells you something about like how fast things are changing.

We get this partial information. It's almost like seeing one pixel in an image (but it's in a different kind of domain). We have to come up with methods that take this really sparse, really noisy data, and try to find the image that might have caused those measurements.

• • •

What we have to end up doing is imposing things called "regularizors" or "priors" that allow us to say, "Okay, of all of the images that possibly could fit this data, this set of images is most likely."

- Katie Bouman

Washington Post: https://www.washingtonpost.com/science/2019/04/10/algorithms-gave-us-black-hole-pic-this-scientist-helped-create-them/

How data science made it possible...

We have telescopes distributed around the world. For every two telescopes in the telescope array, we a a single spatial frequency, which tells you something about like how fast things are changing.

We get this ***partial information***. It's almost like seeing one pixel in an image (but it's in a different kind of domain). We have to come up with methods that take this really sparse, really noisy data, and try to find the image that might have caused those measurements.

• • •

What we have to end up doing is imposing things called "regularizors" or "priors" that allow us to say, "Okay, of all of the images that possibly could fit this data, this set of images is most likely."

- Katie Bouman

Washington Post: https://www.washingtonpost.com/science/2019/04/10/algorithms-gave-us-black-hole-pic-this-scientist-helped-create-them/

How data science made it possible...

We have telescopes distributed around the world. For every two telescopes in the telescope array, we a a single spatial frequency, which tells you something about like how fast things are changing.

We get this partial information. It's almost like seeing one pixel in an image (but it's in a different kind of domain). We have to come up with ***methods that take this really sparse, really noisy data, and try to find the image that might have caused those measurements.***

• • •

What we have to end up doing is imposing things called "regularizors" or "priors" that allow us to say, "Okay, of all of the images that possibly could fit this data, this set of images is most likely."

- Katie Bouman

Washington Post: https://www.washingtonpost.com/science/2019/04/10/algorithms-gave-us-black-hole-pic-this-scientist-helped-create-them/

How data science made it possible...

We have telescopes distributed around the world. For every two telescopes in the telescope array, we a a single spatial frequency, which tells you something about like how fast things are changing.

We get this partial information. It's almost like seeing one pixel in an image (but it's in a different kind of domain). We have to come up with methods that take this really sparse, really noisy data, and try to find the image that might have caused those measurements.

• • •

What we have to end up doing is ***imposing things called "regularizors" or "priors"*** that allow us to say, "Okay, of all of the images that possibly could fit this data, this set of images is most likely."

- Katie Bouman

Washington Post: https://www.washingtonpost.com/science/2019/04/10/algorithms-gave-us-black-hole-pic-this-scientist-helped-create-them/

How data science made it possible...

We have telescopes distributed around the world. For every two telescopes in the telescope array, we a a single spatial frequency, which tells you something about like how fast things are changing.

We get this partial information. It's almost like seeing one pixel in an image (but it's in a different kind of domain). We have to come up with methods that take this really sparse, really noisy data, and try to find the image that might have caused those measurements.

• • •

What we have to end up doing is imposing things called "regularizors" or "priors" that allow us to say, ***"Okay, of all of the images that possibly could fit this data, this set of images is most likely."***

- Katie Bouman

Washington Post: https://www.washingtonpost.com/science/2019/04/10/algorithms-gave-us-black-hole-pic-this-scientist-helped-create-them/

How can you be a part of the data science community?

Be a part of the data science community....

... even before your first internship or job

Participate in data science competitions

Present your work

Share your work online

Participate in data science competitions

This is Statistics Fall Data Challenge
https://www.causeweb.org/usproc/usresp

Oct 28

Undergraduate Statistics Class Project Competition
https://www.causeweb.org/usproc/usclap

Dec 20

Undergraduate Statistics Research Project Competition
https://www.causeweb.org/usproc/usresp

Dec 20

ASA DataFest
https://ww2.amstat.org/education/datafest/

Spring 2020

ASA DataFest™

Present your work

Present at student poster fairs

Present at meetings for data science groups (on campus and local)

R Ladies

CAUSE

Electronic Undergraduate Research Conference

Submit abstract by Oct 23

https://www.causeweb.org/usproc/eusrc/2019

Watch online Nov 1

Share your work in an online repository

Set status

Maria Tackett

matackett

PRO

Edit profile

Assistant Professor of the Practice

Duke University

maria.tackett@duke.edu

www.mariatackett.net

Organizations

Find a repository... | Type: Public | Language: All | New

12 results for **public** repositories | Clear filter

personal-website — Star

HTML Updated yesterday

cv — Star

Forked from isteves/resume

Resume/CV template using RMarkdown and LaTeX

TeX 16 Updated yesterday

talks — Star

Updated 11 days ago

intro-regression — Star

Computing Assignments for an Intro Regression Course

HTML Creative Commons Attribution Share Alike 4.0 International Updated on Jul 10

sta199-sp19-website — Star

Forked from Sta199-S18/website

Course website for Sta 199 - Intro to Data Science, Spring 18 at Duke University

JavaScript 13 Updated on Apr 28

Share your work in an online repository

Set status

Maria Tackett

matackett

PRO

Edit profile

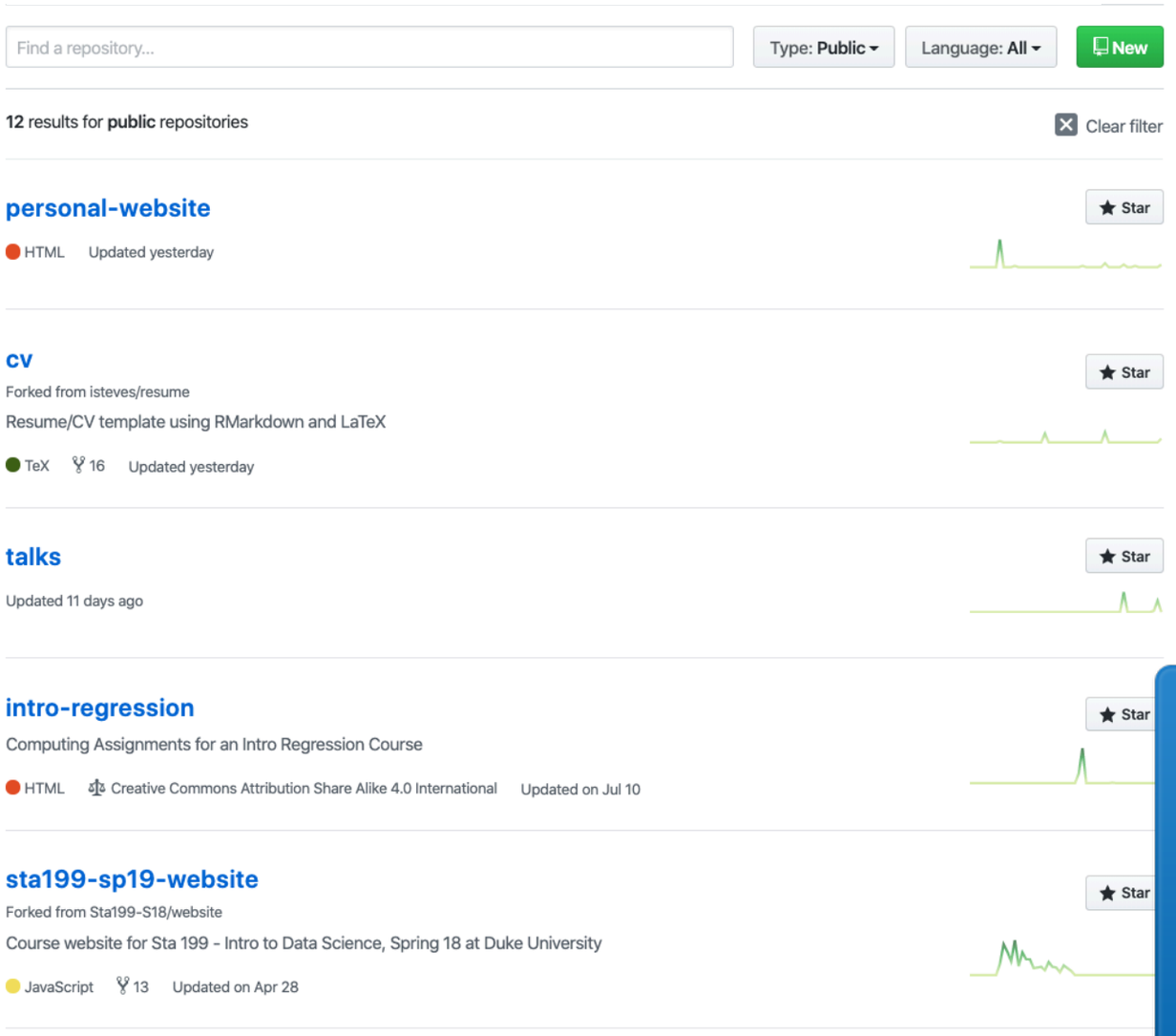
Assistant Professor of the Practice

Duke University

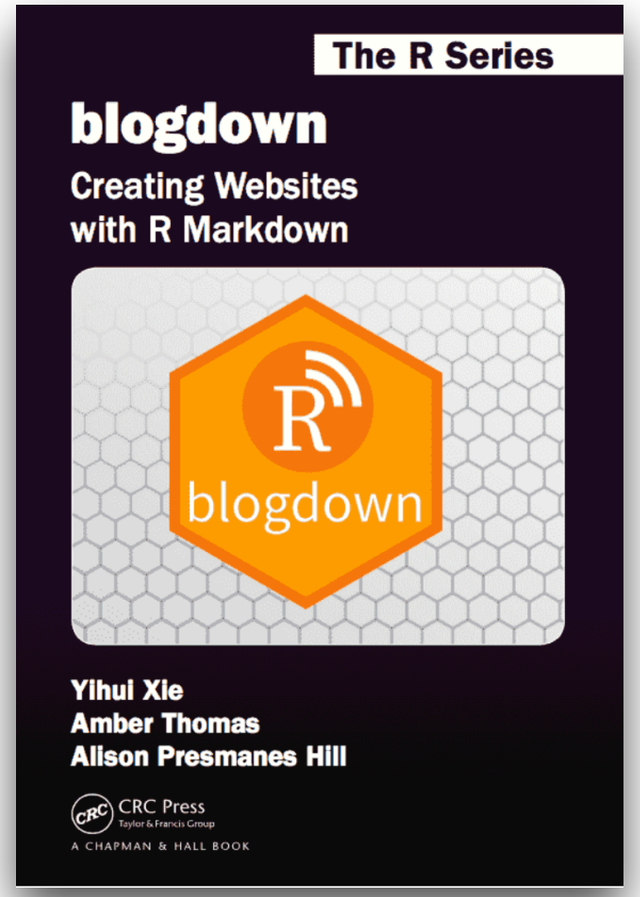
maria.tackett@duke.edu

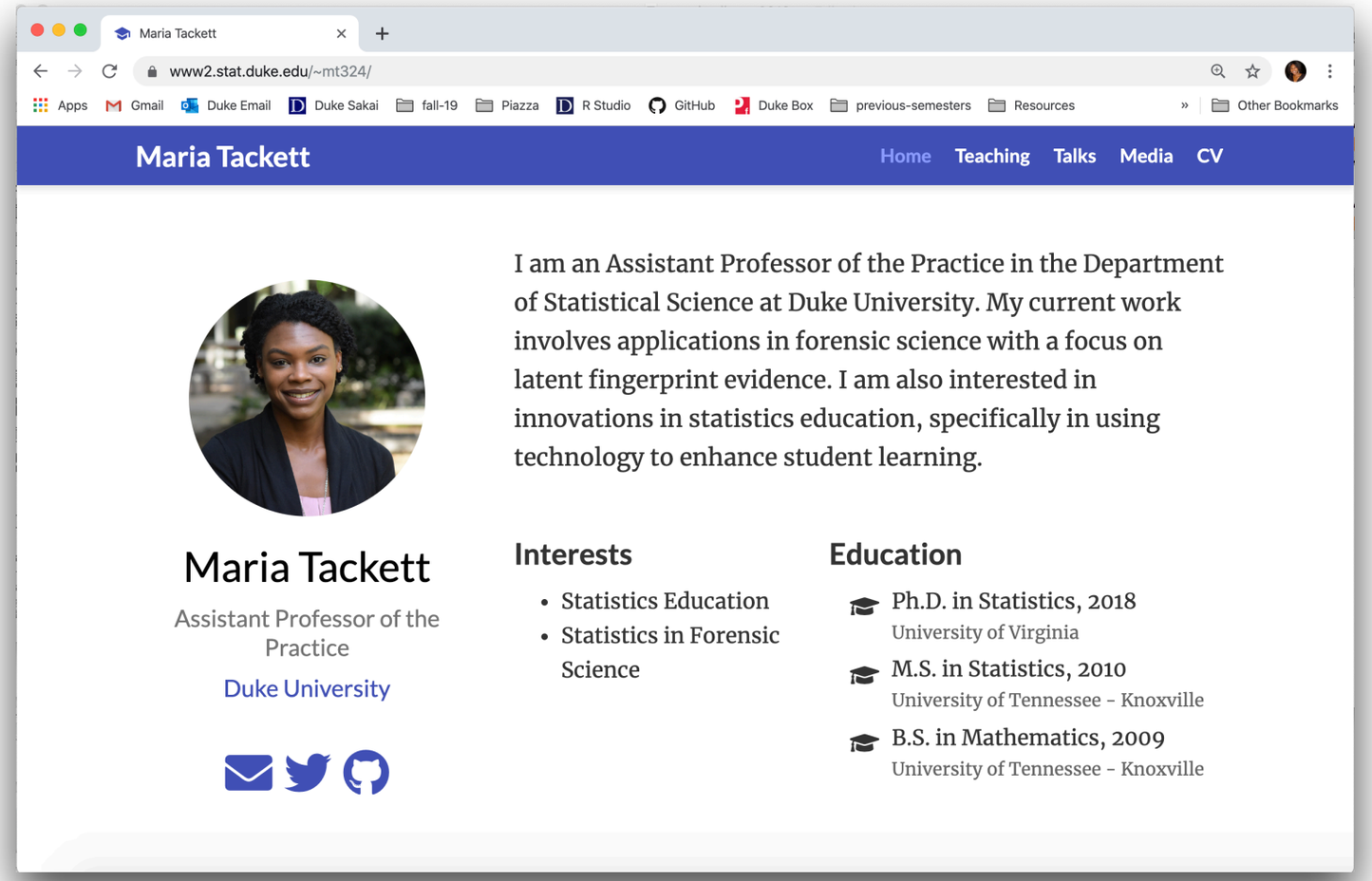
www.mariatackett.net

Organizations

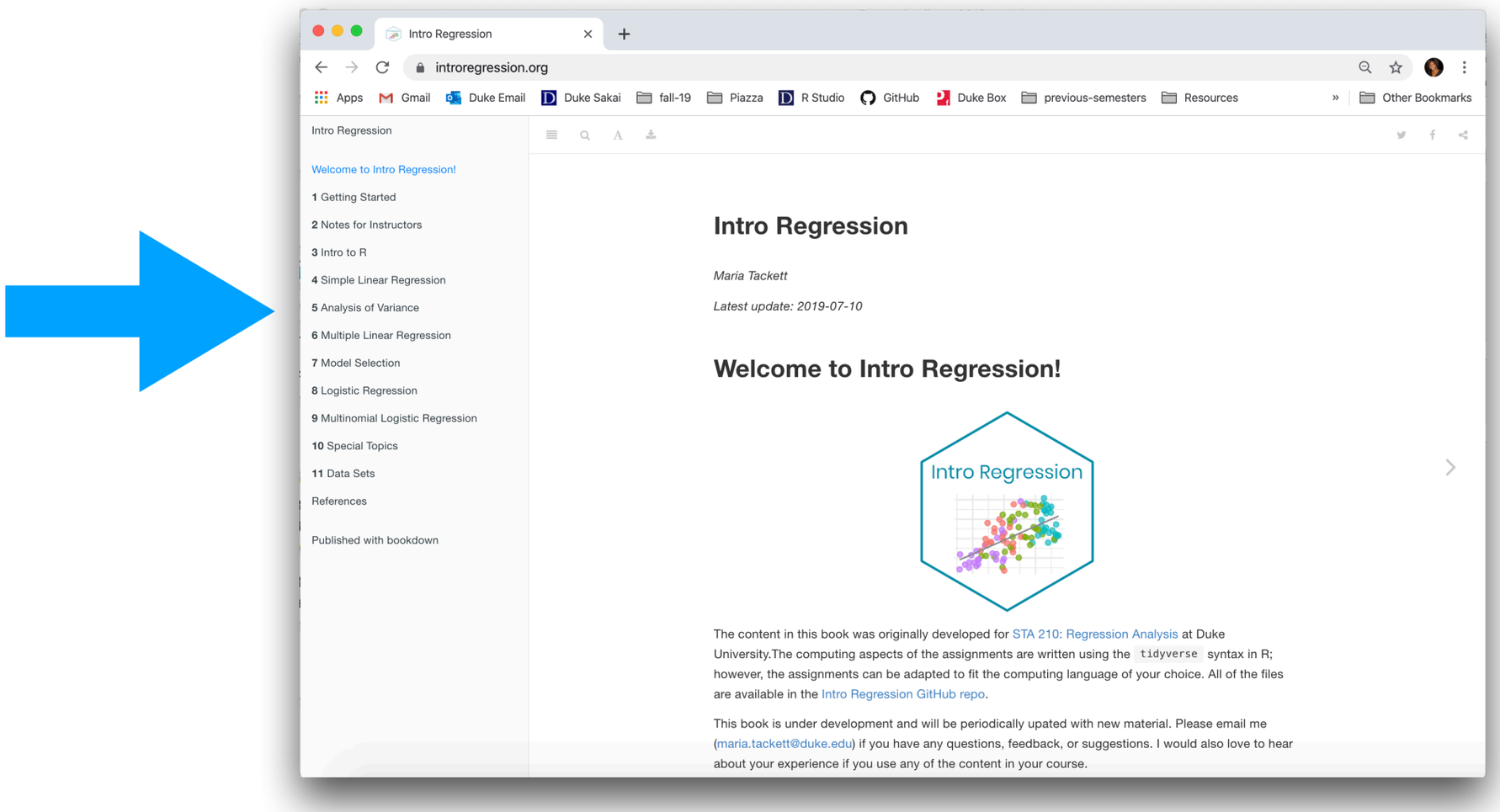


Create a personal website

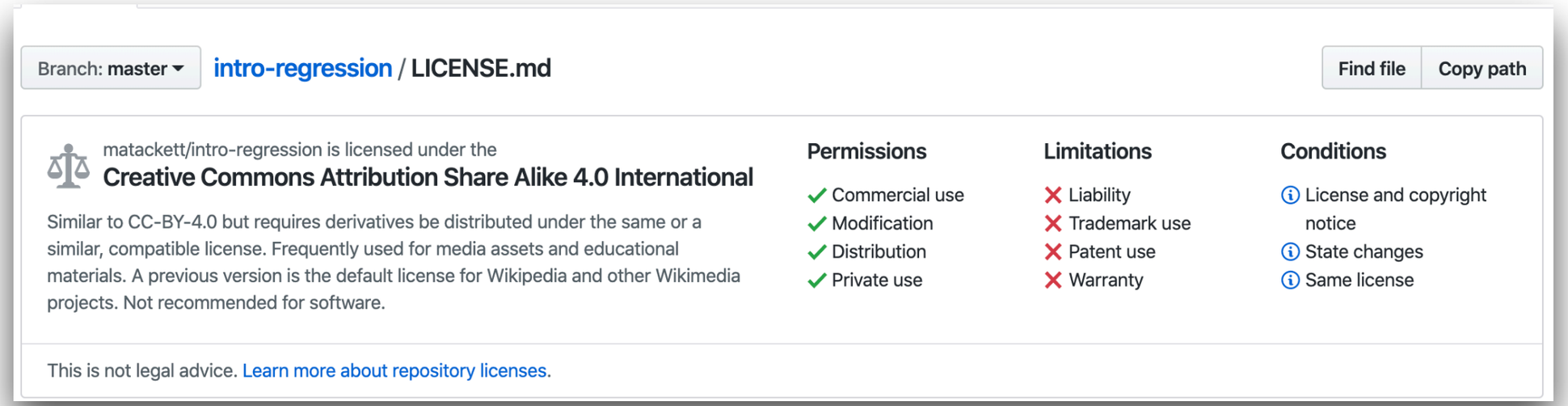




Write about your work

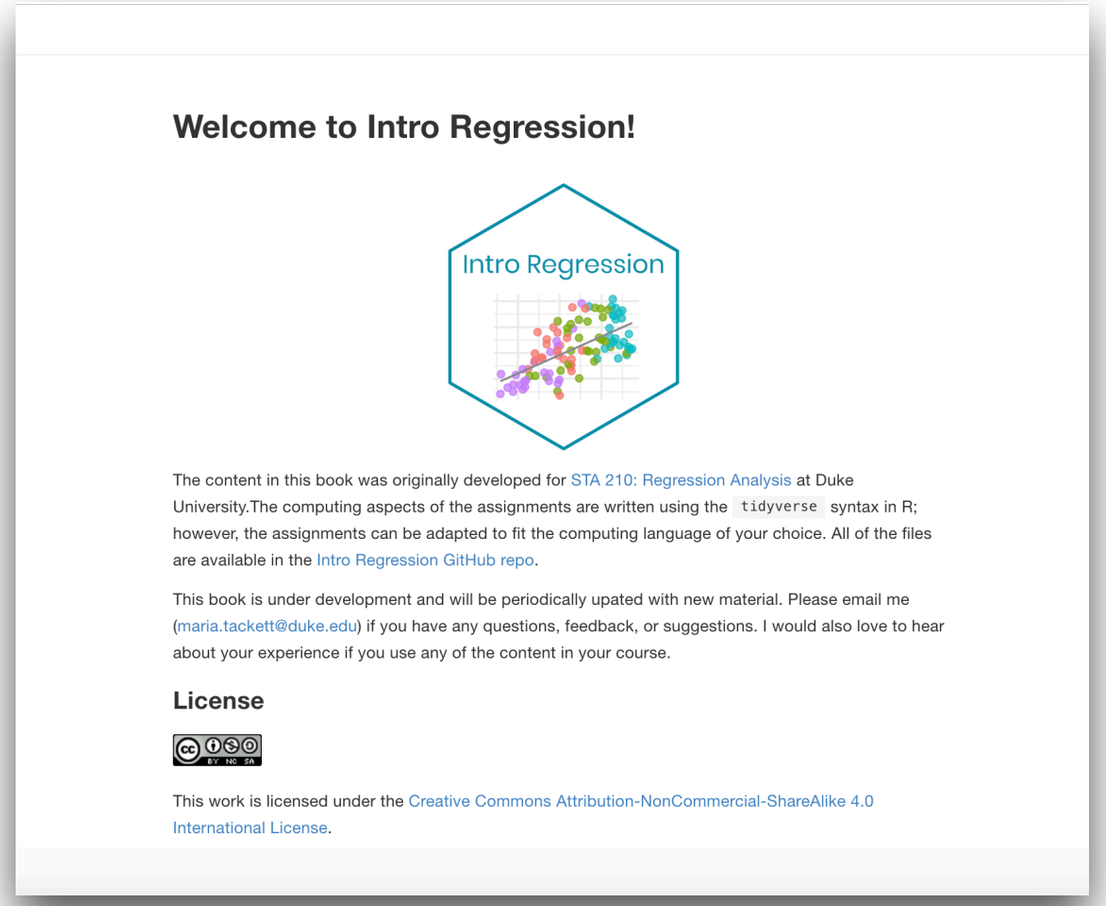


Protect your major contributions



- Use a creative commons license to protect your work
- Make the license visible on the first page of your project or the top of your online repository

https://creativecommons.org/licenses/

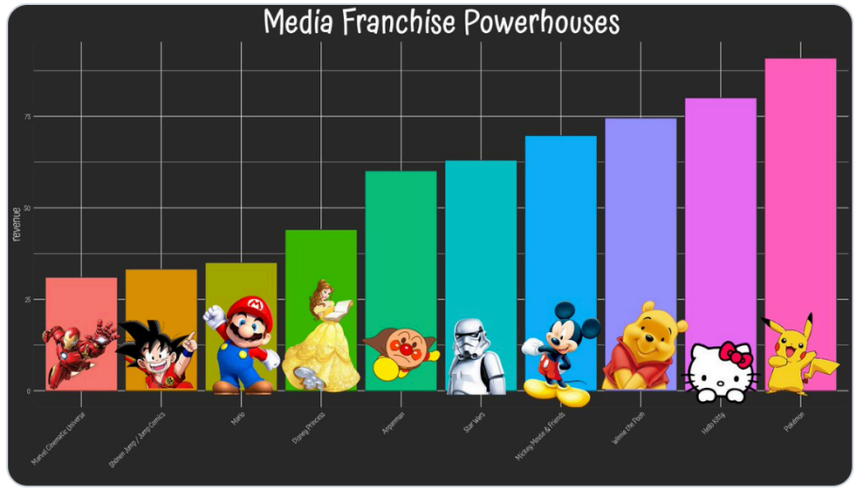


Try new things and share your experience

https://github.com/rfordatascience/tidytuesday

Romina Elena Mendez
@r0mymendez

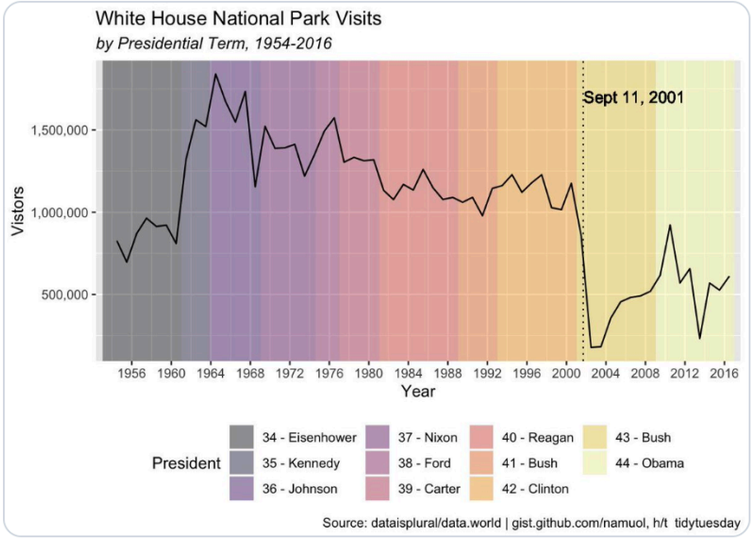
My #TidyTuesday about Media Franchise Powerhouses
the code is 👉bit.ly/2Xep8oL
#r4ds #datascience #Franchise #data #r #ggplot #rstats #dataviz #rstudio #Disney #MarioBros #winniePooh #princess #pokemon #starWars #Marvel



1:21 AM · Jul 6, 2019 · Twitter Web Client

K.G. Jewell
@KG_Jewell

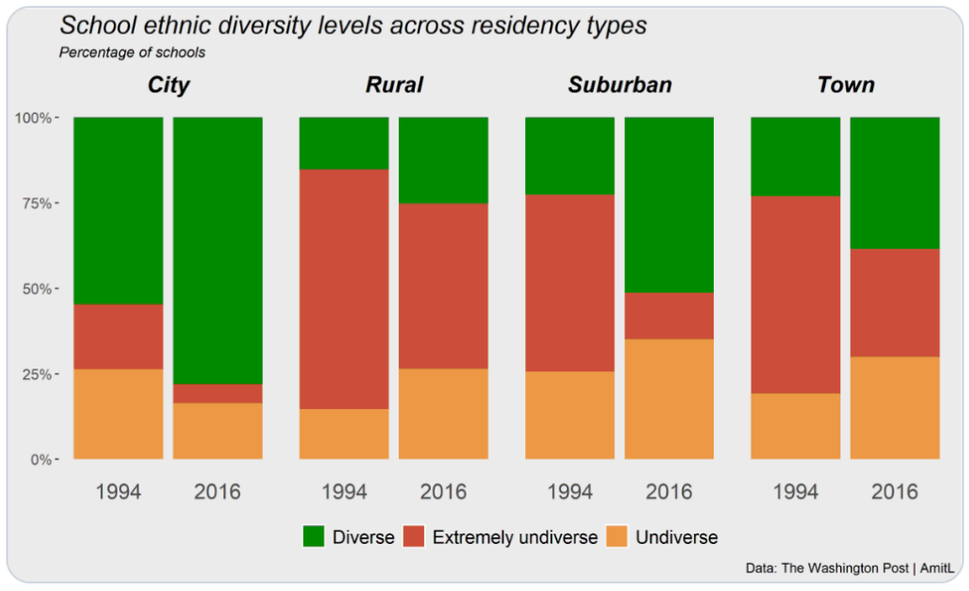
A #TidyTuesday look a visits to the White House National Park by presidential term. A chance to experiment with ggplot geom_rects delineating eras in an xy plot, and add a dropline for the discontinuity after 9/11 security changes.
github.com/kjewell/nat_pa...



3:35 PM · Sep 17, 2019 · Twitter Web App

Amit Levinson
@Amit_Levinson

#TidyTuesday plot of school ethnic diversity levels across residency types for 1994 & 2016.
I love seeing how it fits with some sociological thoughts on minorities' residency: schools are more diverse in cities > suburb/town > rural.
code: bit.ly/2nuiv51



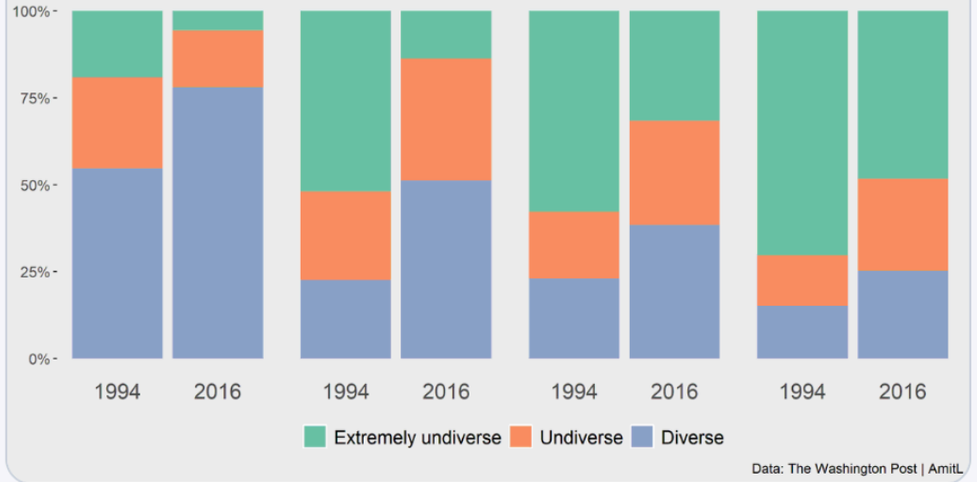
3:44 PM · Sep 26, 2019 · Twitter Web App

1

Amit Levinson @Amit_Levinson · Sep 26
Replying to @Amit_Levinson
I learned a lot this week!
Analysis: using stringr package instead of base r to change strings.
Visualization: used a stacked bar with 3 variables and various levels. I initially tried using facet of year ~ residency with regular bar plot, but i think stacked is neater.

1 1

Amit Levinson @Amit_Levinson · Sep 28
Replying to @Amit_Levinson
some minor changes after some suggestions :)



Rika Gorn 🎃
@RikaGorn

🗣️ Friends looking to get into analytics and #DataScience, participating in #TidyTuesday will give you such a huge payoff including: a portfolio, take home assignment prep, #rstats & #dataviz learning, comfort with cleaning messy datasets, a community!

9:22 AM · May 9, 2019 · Twitter for Android

Final thoughts

- Find what interests you and “follow” the people doing that work
- Find opportunities outside of class to continue developing your skills
- Stay current with what's happening in data science

Final thoughts

- Find what interests you and “follow” the people doing that work
- Find opportunities outside of class to continue developing your skills
- Stay current with what's happening in data science

Have fun! 🙂

Thank You!

maria.tackett@duke.edu

@MT_statistics

bit.ly/beyond-the-buzzword

Maria Tackett
Duke University

References

“What is Data Science"
https://datascience.berkeley.edu/about/what-is-data-science/

“Algorithms Gave Us the Black Hole Pictures. She’s the 29-year-old Scientist Who Created Them”,
https://www.washingtonpost.com/science/2019/04/10/algorithms-gave-us-black-hole-pic-this-scientist-helped-create-them/

"Share, Contribute, Collaborate, Broadcast”.
https://speakerdeck.com/minecr/share-contribute-collaborate-broadcast