

GENERAL ASSEMBLY

INTRO TO DATA SCIENCE

MEET YOUR INSTRUCTORS

A photograph of two men in an office setting. One man, wearing a brown knit beanie and a dark zip-up jacket, is seated at a desk, looking down at a white laptop. The other man, wearing a blue t-shirt with an Adidas logo and blue jeans, is leaning over him, also looking at the laptop screen. They appear to be engaged in a collaborative discussion or work session.

ABOUT YOUR INSTRUCTOR!



Ruben Naeff

rubennaeff@gmail.com

Data Science Instructor

DATA SCIENTIST
AT KNEWTON

MUSIC COMPOSER
STRATEGY CONSULTANT
ECONOMIC RESEARCHER
MATH TEACHER

AMSTERDAM, NL
BROOKLYN, NY

ABOUT YOUR PRODUCER!



Daniel Demoray

ddemoray@google.com

Education Programs Producer

STUDENT

INSTRUCTOR

CURRICULUM

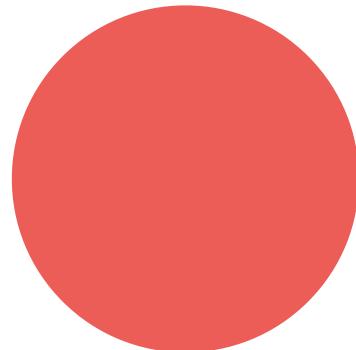
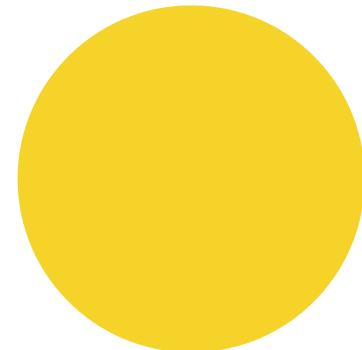
EXPERTS IN RESIDENCE



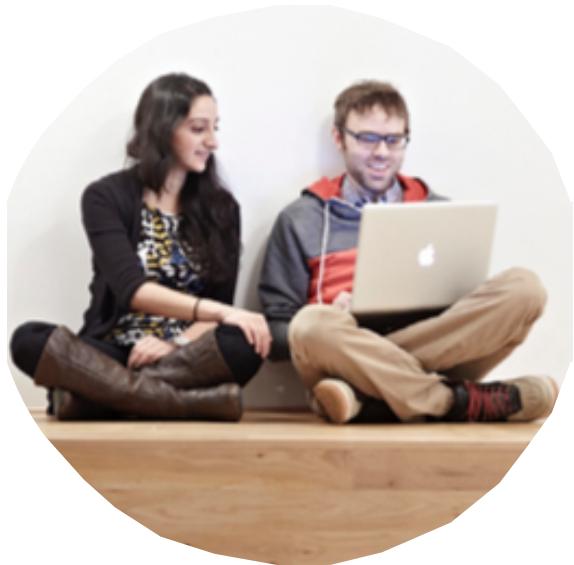
?

Expert In Residence

YOUR MENTOR
& ADVOCATE



ABOUT YOU!



Who
are you

What
do you do

Why
are you here



GA

ABOUT GENERAL ASSEMBLY



**GENERAL ASSEMBLY IS A GLOBAL
COMMUNITY OF INDIVIDUALS
EMPOWERED TO PURSUE THE
WORK WE LOVE.**



**GENERAL ASSEMBLY'S MISSION IS
TO BUILD OUR COMMUNITY BY
TRANSFORMING MILLIONS OF
THINKERS INTO CREATORS.**



TOOLS



ALL YOU CAN LEARN BUFFET

**CLASSES &
WORKSHOPS**
30% OFF GA CLASSES



**COMMUNITY
EVENTS**
HAPPY HOUR EVERY FRI



FEEDBACK / SUPPORT

- Access to EIRs: office hours, in class support
- Exit Tickets
- Mid-Course Feedback
- End of Course Feedback



Exit Tickets

On a scale of 1-5, how well did this lesson cover the learning objectives? *
Did you cover all the learning objectives mentioned at the start of the lesson.

1 2 3 4 5

Not Well Very Well

On a scale of 1-5, how easy was the content to learn? *
Was it presented well? Did you learn the concepts presented?

1 2 3 4 5

Difficult Very Easy

On a scale of 1-5, how effective was your instructor at teaching the content? *

1 2 3 4 5

Not Effective Very Effective

Anything you want to share about this lesson in particular?

Submit

Never submit passwords through Google Forms.

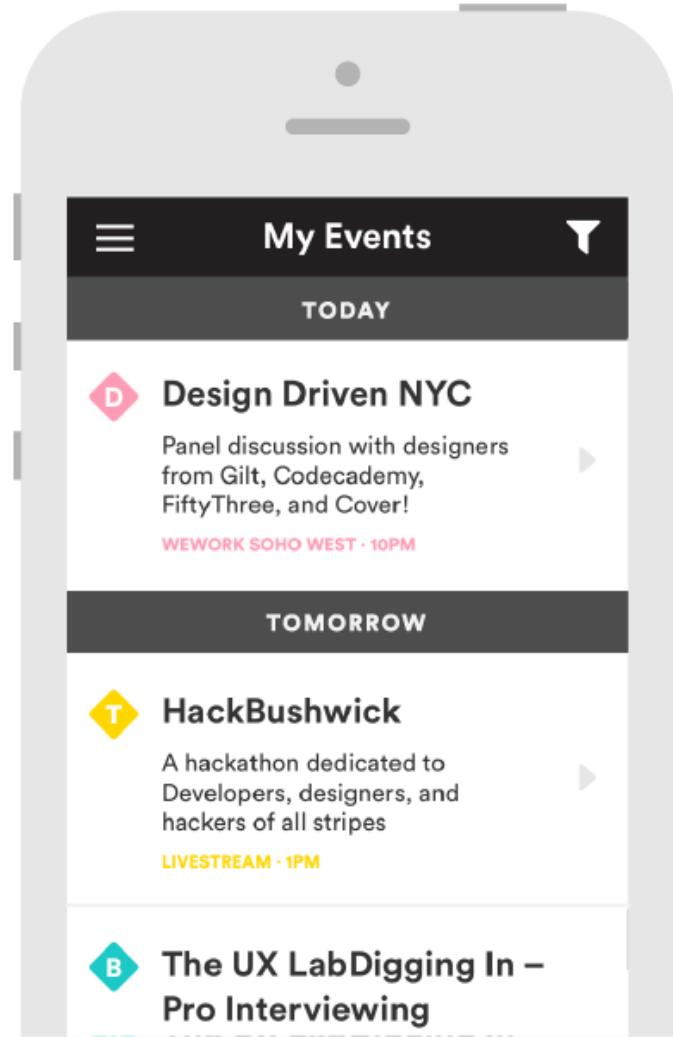


Hub

The best events in tech, business, and design — handpicked each week by our community curators.

Join at hub.ga.co/ny

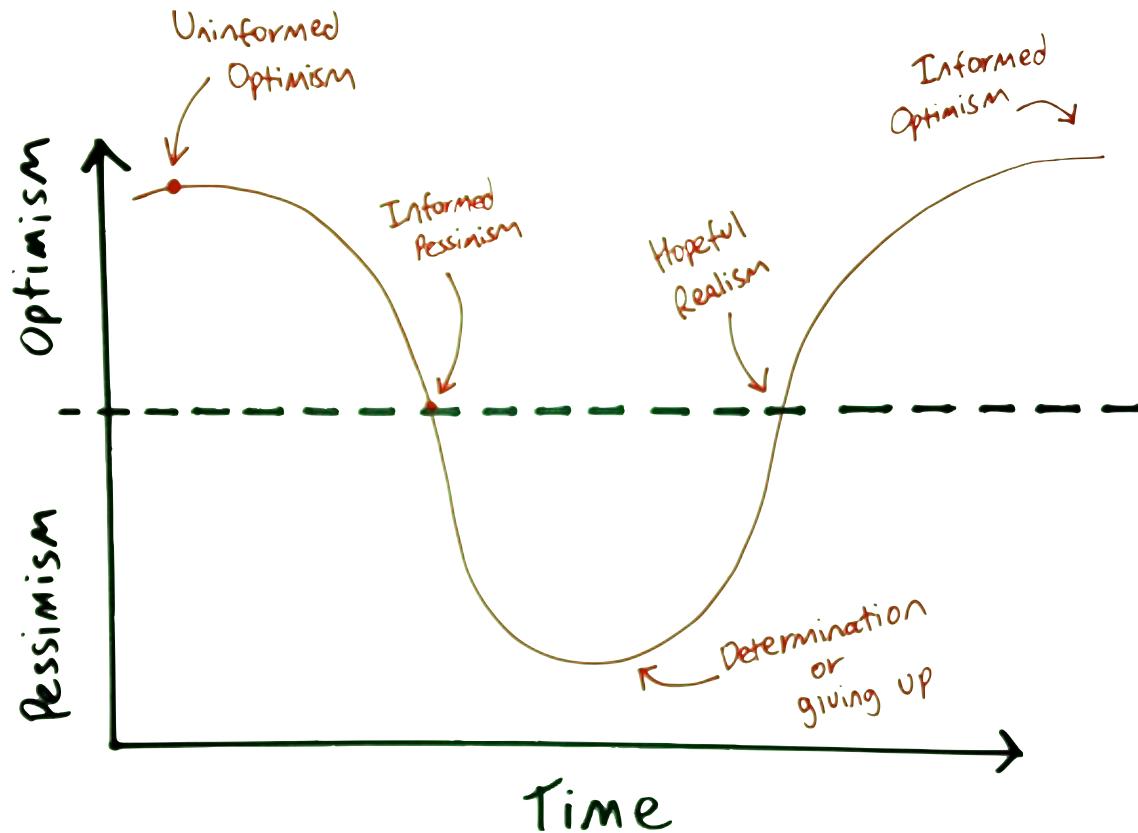
Take a sneak peek of the mobile app at
bit.ly/hubiosbeta



ROAD TO SUCCESS

A photograph of a classroom setting where several students are seated at desks, facing forward. In the foreground, a student with long blonde hair is visible from the side, looking towards the front. Behind her, other students of various ethnicities and ages are also looking forward, some with their hands on their chins in a thoughtful pose. The room has large windows in the background, and the overall atmosphere is one of focused learning.

EMOTIONAL CYCLE OF CHANGE



Student Responsibility

As a self directed program, we view students as
a crucial part of the skill acquisition process.



GA Graduation Requirements

HOMEWORK

(COMPLETE 80% OF
HOMEWORK/LABS)

ATTENDANCE

(MISS NO MORE THAN 2 CLASSES)

**FINAL
PROJECT**

**COMMUNITY
PARTICIPATION + FEEDBACK**



Forever and ever

BUILD YOUR NETWORK

It's not just about altruism, your network is your most valuable asset

FIND OPPORTUNITIES

Alumni have started companies together and recruited other alumni to join their teams

13,000+ STRONG

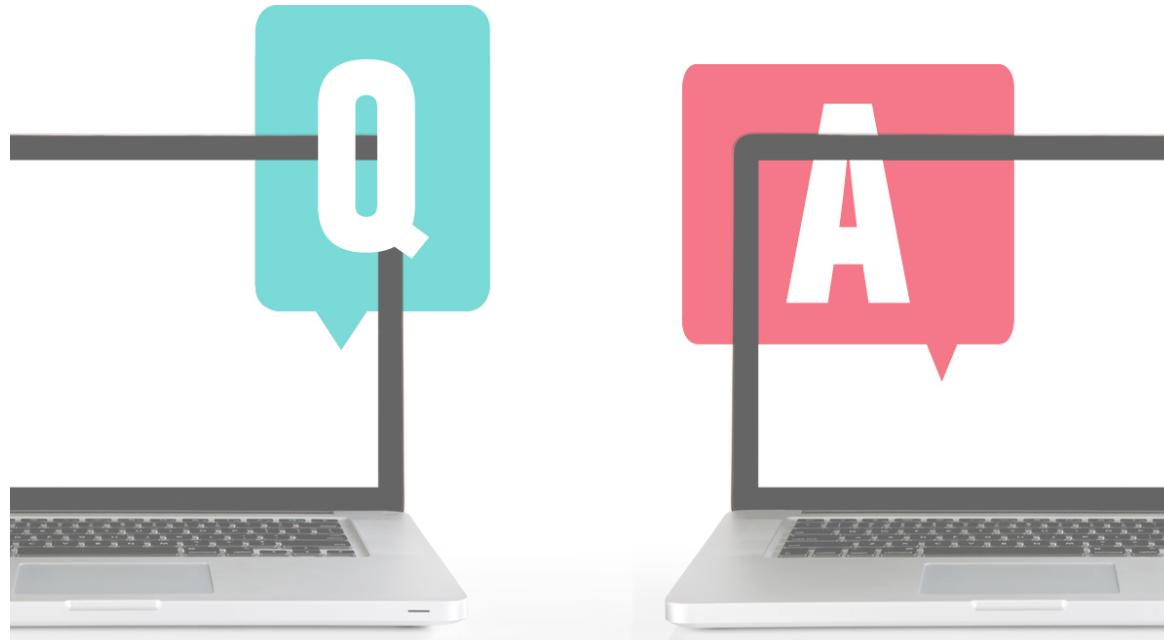
You're part of the alumni community forever

PERKS! 15% OFF CLASSES AND WORKSHOPS, \$500 TUITION CREDIT

We can't wait to have you back on campus



ASK AWAY!



INTRO TO DATA SCIENCE COURSE OVERVIEW

Daniel Demoray, producer – ddemoray@google.com

Ruben Naeff, instructor – rubennaeff@gmail.com

TBD, expert in residence – ??

Everyone – dat-nyc-23@google-groups.com (except for Karsten Kreis, somehow)

Please do not hesitate to contact any of us!

June 25 – September 10, 2015

Tuesdays and Thursdays 6.30-9.30

GA West, 10 East 21st St, room 4A (4th floor)

Office hours: TBD (led by Experts in residence)

Course notes, syllabus, slides, data, code, etc... are all available on
github.com/ga-students/DAT-23-NYC

COURSE OUTLINE:

**DATA EXPLORATION
MODELING & PREDICTING
VARIOUS
FINAL PROJECT**

COURSE OUTLINE:

DATA EXPLORATION: SQL, PYTHON, PANDAS, VISUALIZATIONS

MODELING & PREDICTING: MACHINE LEARNING AND STATISTICS

VARIOUS: SCALING, DEPLOYMENT, SPEAKERS, ETHICS, REQUESTS, ETC.

FINAL PROJECT: FIND DATA, CODE, VISUALIZE, MODEL, PREDICT, PRESENT

COURSE OUTLINE:

DATA EXPLORATION: FIRST TWO WEEKS

MODELING & PREDICTING: EARLY JULY THROUGH MID AUGUST

VARIOUS: MID AUGUST TO EARLY SEPTEMBER

FINAL PROJECT: AUGUST AND SEPTEMBER

25 27

2 4 9 11 16 18 23 25 30

JUNE

JULY

1 6 8 13 15 20 22 27 29

AUGUST

3 5 10

SEPTEMBER

COURSE OUTLINE:

DATA EXPLORATION: FIRST TWO WEEKS

MODELING & PREDICTING: EARLY JULY THROUGH MID AUGUST

VARIOUS: MID AUGUST TO EARLY SEPTEMBER

FINAL PROJECT: AUGUST AND SEPTEMBER

Detailed and up-to-date syllabus is available on github:

[github.com/
ga-students/
DAT-23-NYC](https://github.com/ga-students/DAT-23-NYC)

25 27

2 4 9 11 16 18 23 25 30

JUNE

1 6 8 13 15 20 22 27 29

AUGUST

3 5 10

SEPTEMBER

- ▶ **GO TO [GITHUB.COM/GA-STUDENTS/DAT-23-NYC](https://github.com/ga-students/dat-23-nyc)**
- ▶ SCROLL DOWN AND CLICK ON “SETTING UP YOUR ENVIRONMENT”
- ▶ INSTALL GIT
- ▶ CLONE REPO
- ▶ INSTALL ANACONDA
- ▶ HAVE A BREAK

INTRO TO DATA SCIENCE

LECTURE 1: DATA EXPLORATION

- I. WHAT IS DATA SCIENCE?**
- II. THE DATA MINING WORKFLOW**
- III. WORKING AT THE UNIX COMMAND LINE**

INTRO TO DATA SCIENCE

I. WHAT IS DATA SCIENCE?



Forbes

How Target Figured Out A Teen Girl Was
Pregnant Before Her Father Did



Forbes

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



?



Forbes

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



WHAT IS DATA SCIENCE?

35



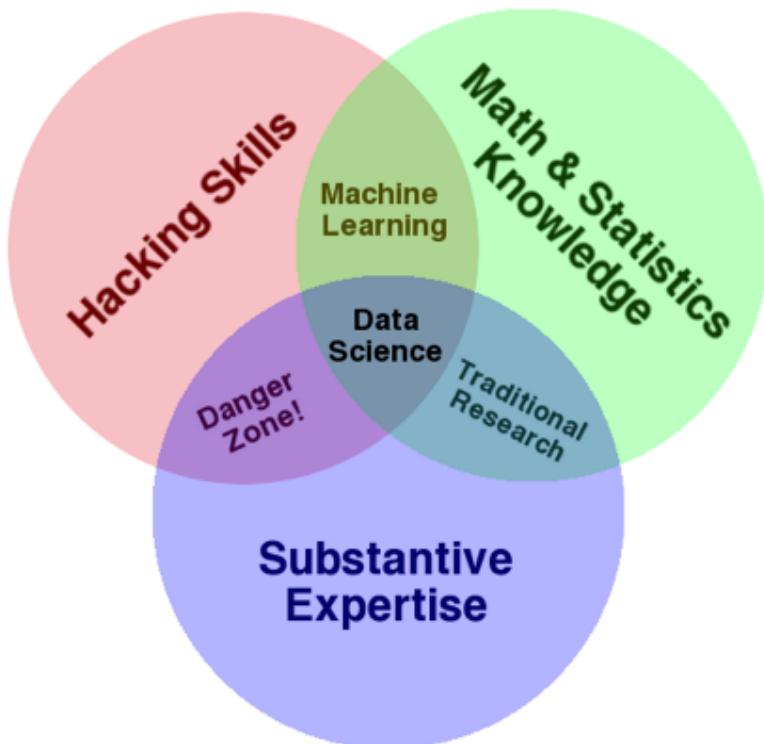
The United States alone faces a shortage of

- 140,000 to 190,000 people with analytical expertise and
- 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data.

McKinsey&Company

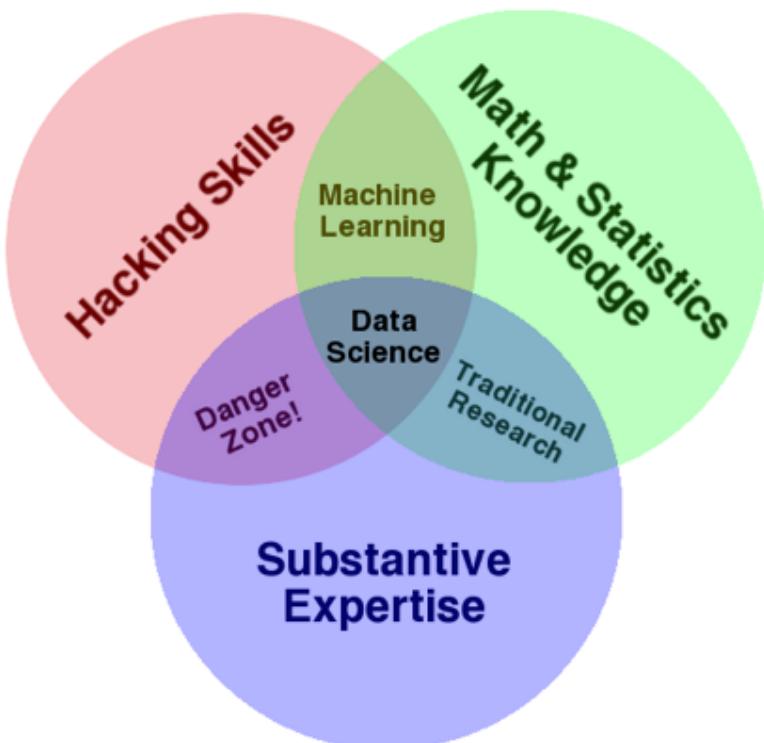
WHAT IS DATA SCIENCE?

37



WHAT IS DATA SCIENCE?

38

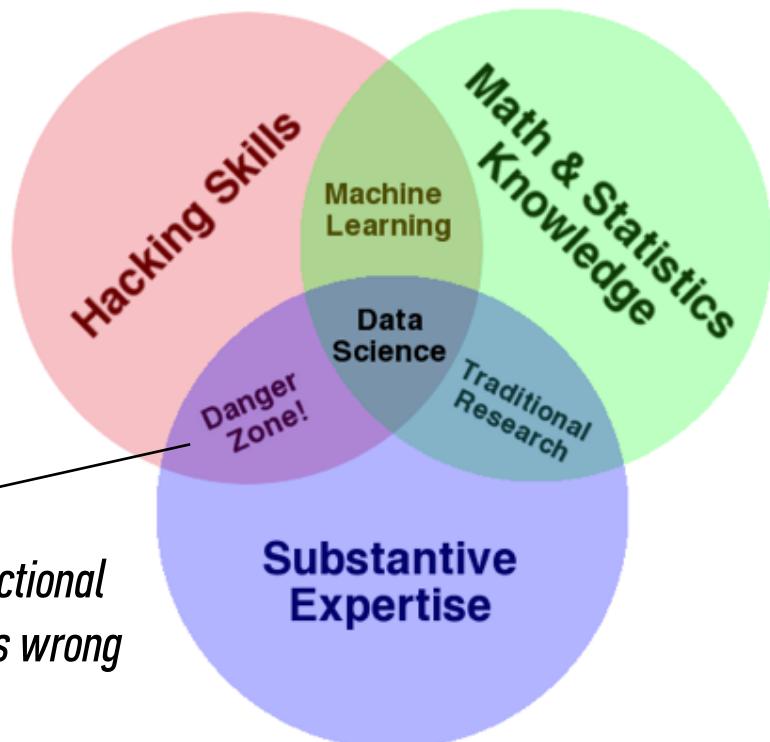


ONE MORE THING!

Communication skills

WHAT IS DATA SCIENCE? – DANGER ZONE △

39



Students who watch more instructional videos, tend to get more answers wrong

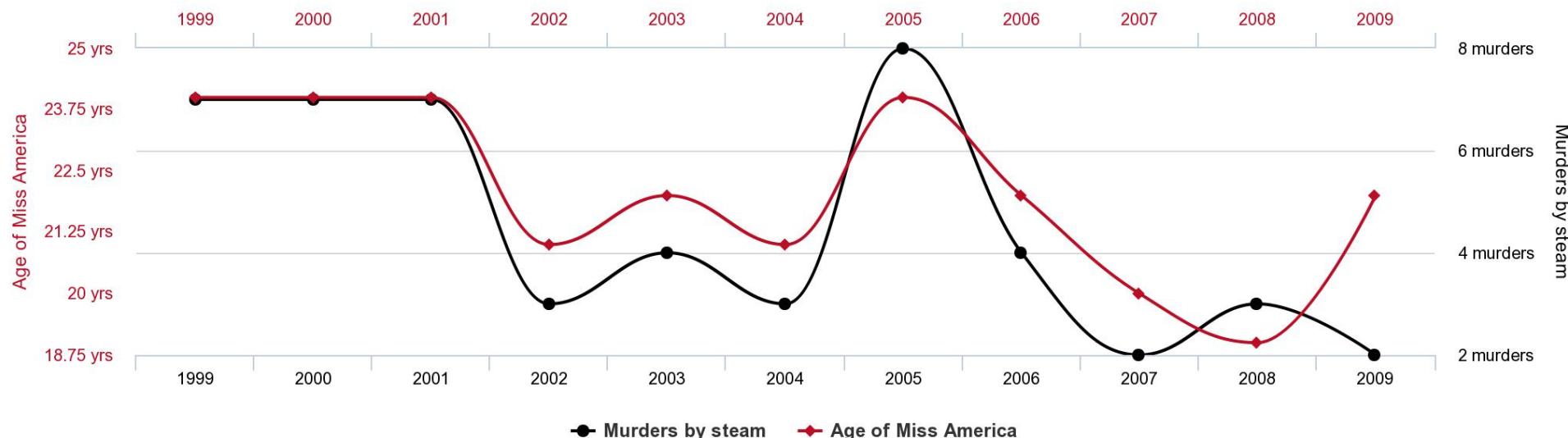
ONE MORE THING!

Communication skills

WHAT IS DATA SCIENCE? – DANGER ZONE △

40

Age of Miss America correlates with Murders by steam, hot vapours and hot objects



- › A set of tools and techniques used to extract useful information from data.

- A set of tools and techniques used to extract useful information from data.
- An interdisciplinary, problem-solving oriented subject.

- A set of tools and techniques used to extract useful information from data.
- An interdisciplinary, problem-solving oriented subject.
- The application of scientific techniques to practical problems.

- A set of tools and techniques used to extract useful information from data.
- An interdisciplinary, problem-solving oriented subject.
- The application of scientific techniques to practical problems.
- Analysis of big data with many dimensions – often in real-time.

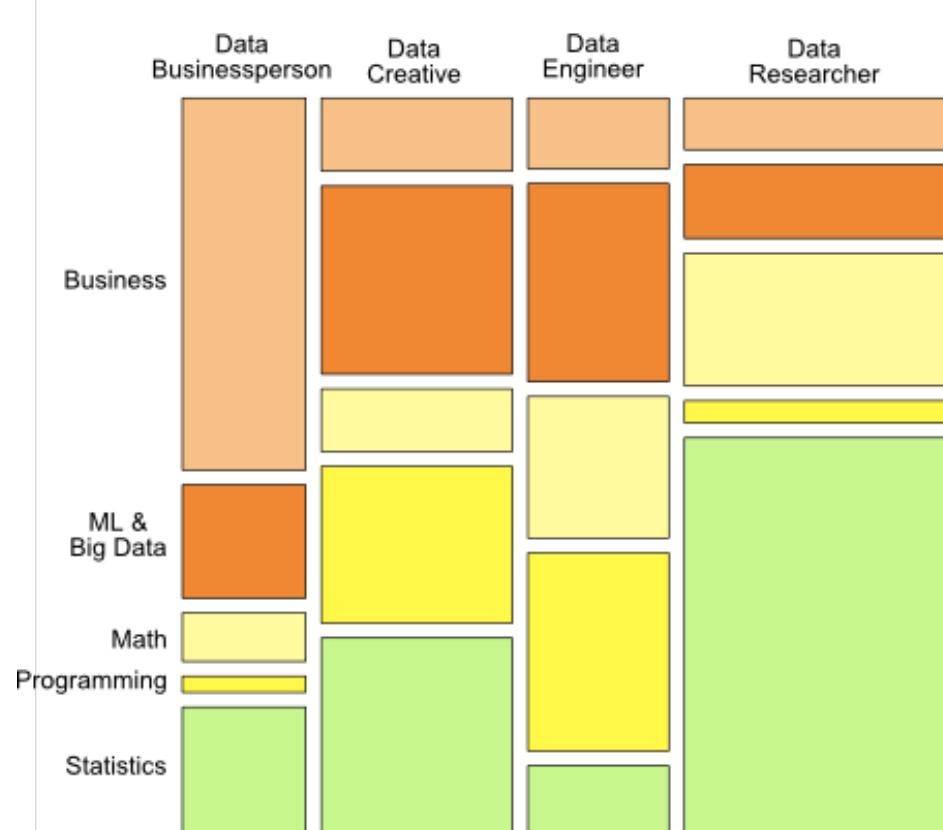
WHAT IS DATA SCIENCE?

45

| | | | |
|---------------------|--------------------|----------------|--------------|
| Data Developer | Developer | Engineer | |
| Data Researcher | Researcher | Scientist | Statistician |
| Data Creative | Jack of All Trades | Artist | Hacker |
| Data Businessperson | Leader | Businessperson | Entrepreneur |

WHAT IS DATA SCIENCE?

46



source: *Analyzing the Analyzers*

WHAT IS DATA SCIENCE?

47



Michael E. Driscoll
@medriscoll



+ Follow

Data scientists: better statisticians than most
programmers & better programmers than
most statisticians bit.ly/NHmRqu
[@peteskomoroch](https://twitter.com/peteskomoroch)



RETWEETS

35

FAVORITES

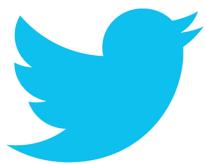
26



4:57 PM - 17 Jul 2012

WHO USES DATA SCIENCE?

48



WHO USES DATA SCIENCE?

Father's Day Savings
Sponsored by DEWALT

Go Hello, Sign in Your Account ▾ Join Prime ▾ Cart ▾ Wish List ▾

Shop by Department Search All ▾

Unlimited Instant Videos
MP3s & Cloud Player
20 million songs, play anywhere
Amazon Cloud Drive
5 GB of free storage
Kindle
Appstore for Android
Streetfood Tycoon... free today
Digital Games & Software
Audible Audiobooks
Books
Movies, Music & Games
Electronics & Computers
Home, Garden & Tools
Grocery, Health & Beauty
Toys, Kids, Baby & Pets
Clothing, Shoes & Jewelry
Sports & Outdoors
Automotive & Industrial
▶ Full Store Directory

The Perfect Gift for Dad
kindle fire HD
From \$199 From \$179
Enter DADSFIRE at checkout
Offer valid through June 8, 2013

Instant Video MP3 Store Cloud Player Kindle Cloud Drive Appstore for Android Digital Games & Software Audible Audiobooks

Toy Story Clothing Trends Fast Free Shipping

Amazon Fashion Dress Shop Our favorite wear-everywhere styles from Laundry by Shelli Segal, London Times, and more.

▶ Shop Dresses ▶ Shop All Clothing

What Other Customers Are Looking At Right Now

| | | | | | | |
|--|---|---|---|--|---|---|
|  SanDisk Ultra 64 GB MicroSDXC Card... ★★★★★ (4,513) \$99.99 \$42.99 |  Samsung Galaxy Tab 2 (7-Inch, Wi-Fi) ★★★★★ (2,401) \$249.99 \$179.00 |  Kindle Fire 7", LCD Display, Wi-Fi, 8GB Amazon Digital Services Inc. ★★★★★ (9,150) \$159.00 |  Kindle, 6" E Ink Display, Wi-Fi, 8GB Amazon Digital Services Inc. ★★★★★ (4,278) \$69.00 |  Kindle Fire HD 7", Dolby Audio... Amazon Digital Services Inc. ★★★★★ (12,817) \$199.00 |  Kindle Fire HD 8.9", Dolby Audio... Amazon Digital Services Inc. ★★★★★ (6,470) \$269.00 |  Amazon Gift Card - E-mail ★★★★★ (19,180) \$50.00 |
|--|---|---|---|--|---|---|

Digital Cameras Best Sellers

| | | | | | | |
|--|--|--|--|--|---|--|
|  Nikon COOLPIX S9200 16 MP CMOS... ★★★★★ (493) |  Canon PowerShot SX500 IS 16.0 MP... ★★★★★ (160) |  Canon PowerShot A2300 IS 16.0 MP... ★★★★★ (443) |  Canon EOS Rebel T3i 18 MP CMOS... ★★★★★ (778) |  Canon PowerShot SX150 IS 14.1 MP... ★★★★★ (283) |  Canon PowerShot SX260 HS 12.1 MP CMOS... ★★★★★ (723) |  Canon EOS Rebel T4i 18.0 MP CMOS... ★★★★★ (118) |
|--|--|--|--|--|---|--|

Class of 2013 Graduation Gifts
▶ Shop now

Try Amazon Prime free for 30 days
Get Started

amazonPrime

Advertisement

Deal of the Day
Up to 60% Off
Select SanDisk Memory
▶ Shop now

UP by Jawbone
Measure your activity, sleep quality, and live better
Learn more

\$30 Off Instantly
Learn more

NEW YORK UNIVERSITY
Noncredit courses and certificates
Center for Advanced Digital Applications (CADA)
There's Still Time to...

WHO USES DATA SCIENCE?

50

amazon Join Prime Your Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search All Go Hello, Sign In Your Account Join Prime Cart Wish List

Father's Day Savings Sponsored by DEWALT Shop now

Instant Video MP3 Store Cloud Player Kindle Cloud Drive Appstore for Android Digital Games & Software Audible Audiobooks

Unlimited Instant Videos
MP3s & Cloud Player
20 million songs, play anywhere
Amazon Cloud Drive
5 GB of free storage
Kindle
Appstore for Android
Streetfood Tycoon... free today
Digital Games & Software
Audible Audiobooks
Books
Movies, Music & Games
Electronics & Computers
Home, Garden & Tools
Grocery, Health & Beauty
Toys, Kids, Baby & Pets
Clothing, Shoes & Jewelry
Sports & Outdoors
Automotive & Industrial
» Full Store Directory

The Perfect Gift for Dad
kindle fire HD
From \$199 From \$179
Enter DADSFIRE at checkout
Offer valid through June 8, 2013

Class of 2013 Graduation Gifts
» Shop now

Try Amazon Prime free for 30 days
Get Started

amazon Prime

Advertisement

Deal of the Day Up to 60% Off Select SanDisk Memory
» Shop now

UP by Jawbone Measure your activity, sleep quality, and live better
Learn more

Digital Cameras Best Sellers

Nikon COOLPIX S9200 16 MP CMOS...
Canon PowerShot SX500 IS 16.0 MP...
Canon PowerShot A2300 IS 16.0 MP...
Canon EOS Rebel T3i 18 MP CMOS...
Canon PowerShot SX150 IS 14.1 MP...
Canon PowerShot SX260 HS 12.1 MP CMOS...
Canon EOS Rebel T4i 18.0 MP CMOS...

NEW YORK UNIVERSITY Noncredit courses and certificates Center for Advanced Digital Applications (CADA) There's Still Time to...

SanDisk Ultra 64 GB microSDXC Class...
★★★★★ (4,513)
\$99.00 \$42.99

Samsung Galaxy Tab 2 (7-Inch, Wi-Fi)
★★★★★ (2,401)
\$249.00 \$179.00

Kindle Fire 7", LCD Display, Wi-Fi, 8GB
Amazon Digital Services Inc.
★★★★★ (9,150)
\$159.00

Kindle, 6" E Ink Display, Wi-Fi, 8GB
Amazon Digital Services Inc.
★★★★★ (4,278)
\$69.00

Kindle Fire HD 7", Dolby Audio, 8GB
Amazon Digital Services Inc.
★★★★★ (12,817)
\$199.00

Kindle Fire HD 8.9", Dolby Audio, 16GB
Amazon Digital Services Inc.
★★★★★ (6,470)
\$269.00

Amazon Gift Card - E-mail
★★★★★ (19,180)
\$50.00

\$30 Off Instantly
Learn more

WHO USES DATA SCIENCE?

amazon Join Prime Your Amazon.com Today's Deals | Gift Cards Sell Help

Shop by Department Search All Go Hello, Sign In Your Account Join Prime Cart Wish List Father's Day Savings Sponsored by DEWALT >Shop now

Unlimited Instant Videos MP3s & Cloud Player 20 million songs, play anywhere Amazon Cloud Drive 5 GB of free storage Kindle Appstore for Android Streetfood Tycoon... free today Digital Games & Software Audible Audiobooks Books Movies, Music & Games Electronics & Computers Home, Garden & Tools Grocery, Health & Beauty Toys, Kids, Baby & Pets Clothing, Shoes & Jewelry Sports & Outdoors Automotive & Industrial > Full Store Directory

The Perfect Gift for Dad kindle fire HD From \$199 From \$179 Enter DADSFIRE at checkout Offer valid through June 8, 2013

Instant Video MP3 Store Cloud Player Kindle Cloud Drive Appstore for Android Digital Games & Software Audible Audiobooks

Toy Story Clothing Trends Fast Free Shipping Amazon Fashion Dress Shop Our favorite wear-everywhere styles from Laundry by Shelli Segal, London Times, and more. >Shop Dresses >Shop All Clothing

Play Try Amazon Prime free for 30 days Get Started

amazon Prime Advertisement

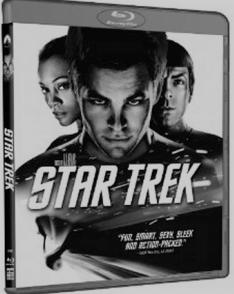
Deal of the Day Up to 60% Off Select SanDisk Memory >Shop now

SanDisk Ultra 64 GB MicroSDXC Class... ★★★★★ (4,513) \$99.99 \$42.99 Samsung Galaxy Tab 2 (7-Inch, Wi-Fi) ★★★★★ (2,401) \$249.99 \$179.00 Kindle Fire 7", LCD Display, Wi-Fi, 8... Amazon Digital Services Inc. ★★★★★ (9,150) \$159.00 Kindle, 6" E Ink Display, Wi-Fi... Amazon Digital Services Inc. ★★★★★ (4,278) \$69.00 Kindle Fire HD 7", Dolby Audio... Amazon Digital Services Inc. ★★★★★ (12,817) \$199.00 Kindle Fire HD 8.9", Dolby Audio... Amazon Digital Services Inc. ★★★★★ (6,470) \$269.00 Amazon Gift Card - E-mail ★★★★★ (19,180) \$50.00

UP by Jawbone Measure your activity, sleep quality, and live better >Learn more

\$30 Off Instantly

Nikon COOLPIX S9200 16 MP CMOS... Canon PowerShot SX500 IS 16.0 MP... Canon PowerShot A2300 IS 16.0 MP... Canon EOS Rebel T3i 18 MP CMOS... Canon PowerShot SX150 IS 14.1 MP... Canon PowerShot SX260 HS 12.1 MP CMOS... Canon EOS Rebel T4i 18.0 MP CMOS... NEW YORK UNIVERSITY Noncredit courses and certificates Center for Advanced Digital Applications (CADA) There's Still Time to...



Star Trek [Blu-ray] (2009)

Chris Pine (Actor), Zachary Quinto (Actor), J.J. Abrams (Director) | Rated: PG-13 | Format: Blu-ray

(2,040 customer reviews)

List Price: \$22.98

Price: \$9.99 & **FREE Shipping** on orders over \$25. [Details](#)

You Save: \$12.99 (57%)

In Stock.

Ships from and sold by [Amazon.com](#). Gift-wrap available.

Want it Wednesday, June 5? Order within 20 hrs 11 mins and choose **One-Day Shipping** at checkout. [Details](#)

23 new from \$9.98 **19 used** from \$9.78 **1 collectible** from \$49.99

Roll over image to zoom in



[See 1 customer image](#)

[Share your own customer images](#)

Watch Instantly with [amazon](#) instant video

Rent

Buy

Star Trek (2009)

\$2.99

\$9.99

Other Formats & Versions

Amazon Price

New from

Used from



Blu-ray 1-Disc Version

\$9.99

\$9.98

\$9.78



DVD Single-Disc Edition

\$8.49

\$5.65

\$3.29

This week only, save up to 62% on [Farscape: The Complete Series](#) in our Deal of the Week. Offer ends June 8, 2013. [Learn more](#)



Frequently Bought Together



Price for all three: **\$66.47**

Add all three to Cart

Add all three to Wish List

Some of these items ship sooner than the others. [Show details](#)

This item: Star Trek [Blu-ray] ~ Chris Pine Blu-ray **\$9.99**

Star Trek Into Darkness (Blu-ray 3D + Blu-ray + DVD + Digital Copy) ~ Chris Pine Blu-ray **\$24.99**

Iron Man 3 (Two-Disc Blu-ray / DVD + Digital Copy) ~ Robert Downey Jr. Blu-ray **\$31.49**

What Other Items Do Customers Buy After Viewing This Item?



Star Trek Into Darkness (Blu-ray 3D + Blu-ray + DVD + Digital Copy) ~ Chris Pine Blu-ray

(199)

\$24.99



Star Trek: Original Motion Picture Collection (Star Trek I, II, III, IV, V, VI + The Captain's Summit Bonus Disc) [Blu-ray] ~ William Shatner Blu-ray

(571)

\$53.56



Sin City (Two-Disc Theatrical & Recut, Extended, and Unrated Versions) [Blu-ray] ~ Jessica Alba Blu-ray

(933)

\$4.99

Quantity: 1

Yes, I want **FREE Two-Day Shipping** with Amazon Prime

Add to Cart

or

[Sign in](#) to turn on 1-Click ordering.

[Add to Wish List](#)

Sell Us Your Item

For up to a **\$2.60** Gift Card

Trade in

[Learn more](#)

More Buying Choices

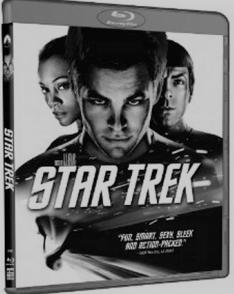
TechShowMe

Add to Cart
\$19.99 & **FREE Shipping** on orders over \$25. [Details](#)

43 used & new from \$9.78

Have one to sell? [Sell on Amazon](#)

[Share](#)



Star Trek [Blu-ray] (2009)

Chris Pine (Actor), Zachary Quinto (Actor), J.J. Abrams (Director) | Rated: PG-13 | Format: Blu-ray

(2,040 customer reviews)

List Price: \$22.98

Price: \$9.99 & **FREE Shipping** on orders over \$25. [Details](#)

You Save: \$12.99 (57%)

In Stock.

Ships from and sold by [Amazon.com](#). Gift-wrap available.

Want it Wednesday, June 5? Order within 20 hrs 11 mins and choose **One-Day Shipping** at checkout. [Details](#)

23 new from \$9.98 **19 used** from \$9.78 **1 collectible** from \$49.99

Roll over image to zoom in



[See 1 customer image](#)

[Share your own customer images](#)

| Watch Instantly with amazon instant video | | Rent | Buy |
|---|---------------------|----------|-----------|
| Star Trek (2009) | | \$2.99 | \$9.99 |
| Other Formats & Versions | Amazon Price | New from | Used from |
| Blu-ray | 1-Disc Version | \$9.99 | \$9.98 |
| DVD | Single-Disc Edition | \$8.49 | \$5.65 |
| | | \$3.29 | |

This week only, save up to 62% on [Farscape: The Complete Series](#) in our Deal of the Week. Offer ends June 8, 2013. [Learn more](#)



53% off

Frequently Bought Together



Price for all three: \$66.47

Add all three to Cart

Add all three to Wish List

Some of these items ship sooner than the others. [Show details](#)

- This item:** Star Trek [Blu-ray] ~ Chris Pine Blu-ray \$9.99
- Star Trek Into Darkness (Blu-ray 3D + Blu-ray + DVD + Digital Copy) ~ Chris Pine Blu-ray \$24.99
- Iron Man 3 (Two-Disc Blu-ray / DVD + Digital Copy) ~ Robert Downey Jr. Blu-ray \$31.49

What Other Items Do Customers Buy After Viewing This Item?



[Star Trek Into Darkness \(Blu-ray 3D + Blu-ray + DVD + Digital Copy\)](#) ~ Chris Pine Blu-ray

(199)

\$24.99



[Star Trek: Original Motion Picture Collection \(Star Trek I, II, III, IV, V, VI + The Captain's Summit Bonus Disc\) \[Blu-ray\]](#) ~ William Shatner Blu-ray

(571)

\$53.56



[Sin City \(Two-Disc Theatrical & Recut, Extended, and Unrated Versions\) \[Blu-ray\]](#) ~ Jessica Alba Blu-ray

(933)

\$4.99

Quantity: 1

Yes, I want **FREE Two-Day Shipping** with Amazon Prime

Add to Cart

or

[Sign in](#) to turn on 1-Click ordering.

[Add to Wish List](#)

Sell Us Your Item

For up to a **\$2.60** Gift Card

Trade in

[Learn more](#)

More Buying Choices

TechShowMe Add to Cart

\$19.99 & **FREE Shipping** on orders over \$25. [Details](#)

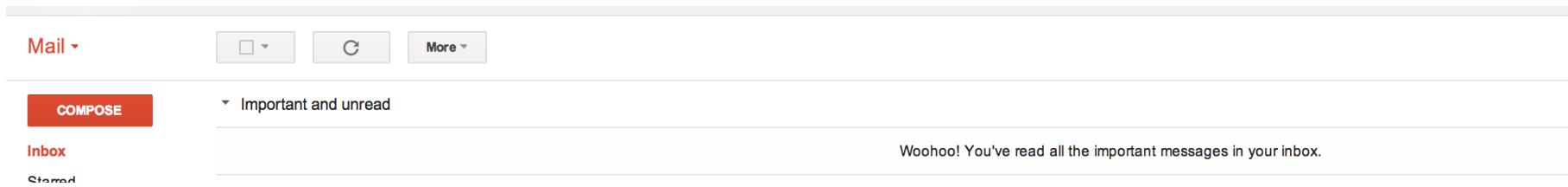
43 used & new from \$9.78

Have one to sell? [Sell on Amazon](#)

[Share](#)

WHO USES DATA SCIENCE?

54



INTRO TO DATA SCIENCE

II. THE DATA SCIENCE WORKFLOW

acquire parse filter mine represent refine interact

THE DATA SCIENCE WORKFLOW

57







THE DATA SCIENCE WORKFLOW

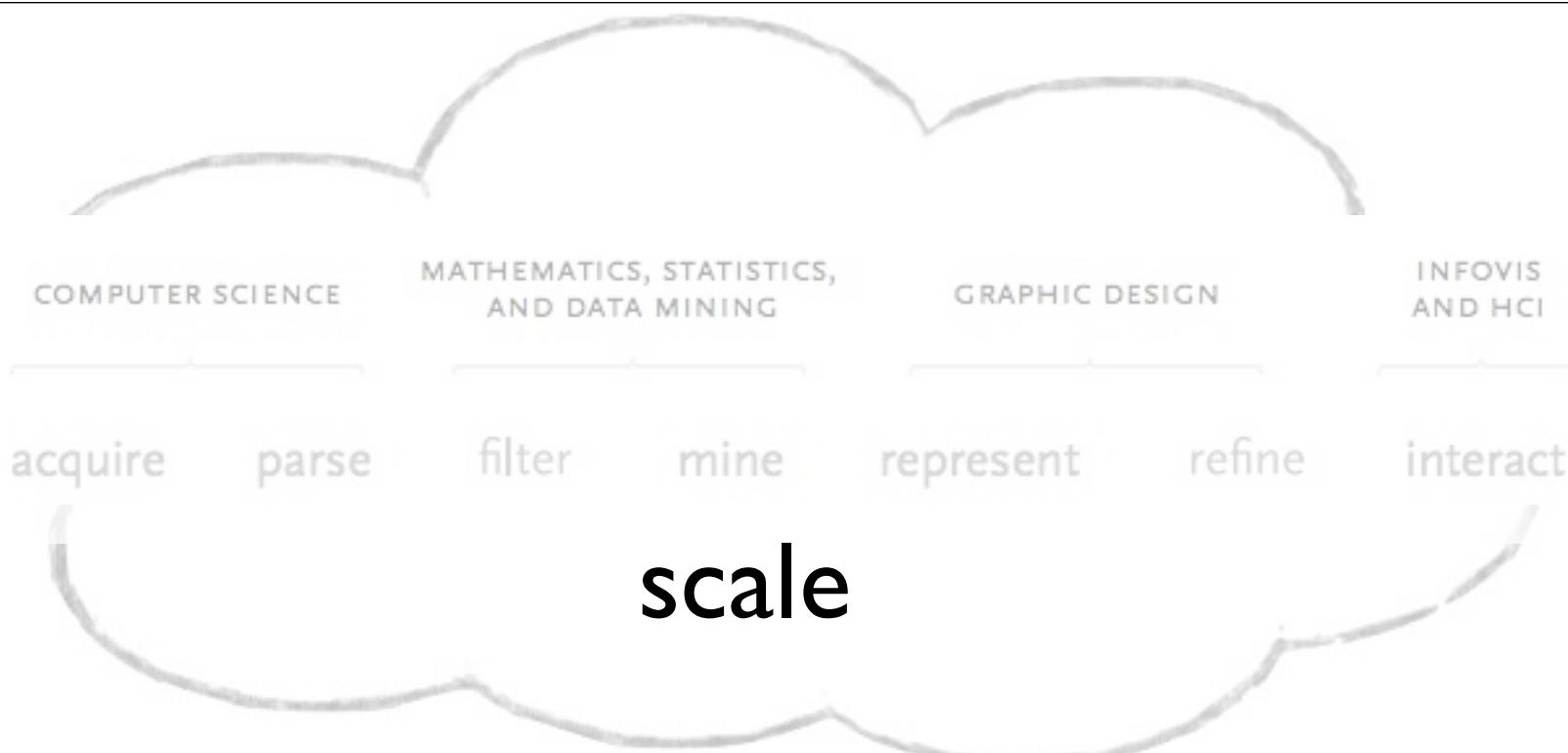
60





THE DATA SCIENCE WORKFLOW

62



source: <http://benfry.com/phd/dissertation-110323c.pdf>

THE DATA SCIENCE WORKFLOW

63



NOTE

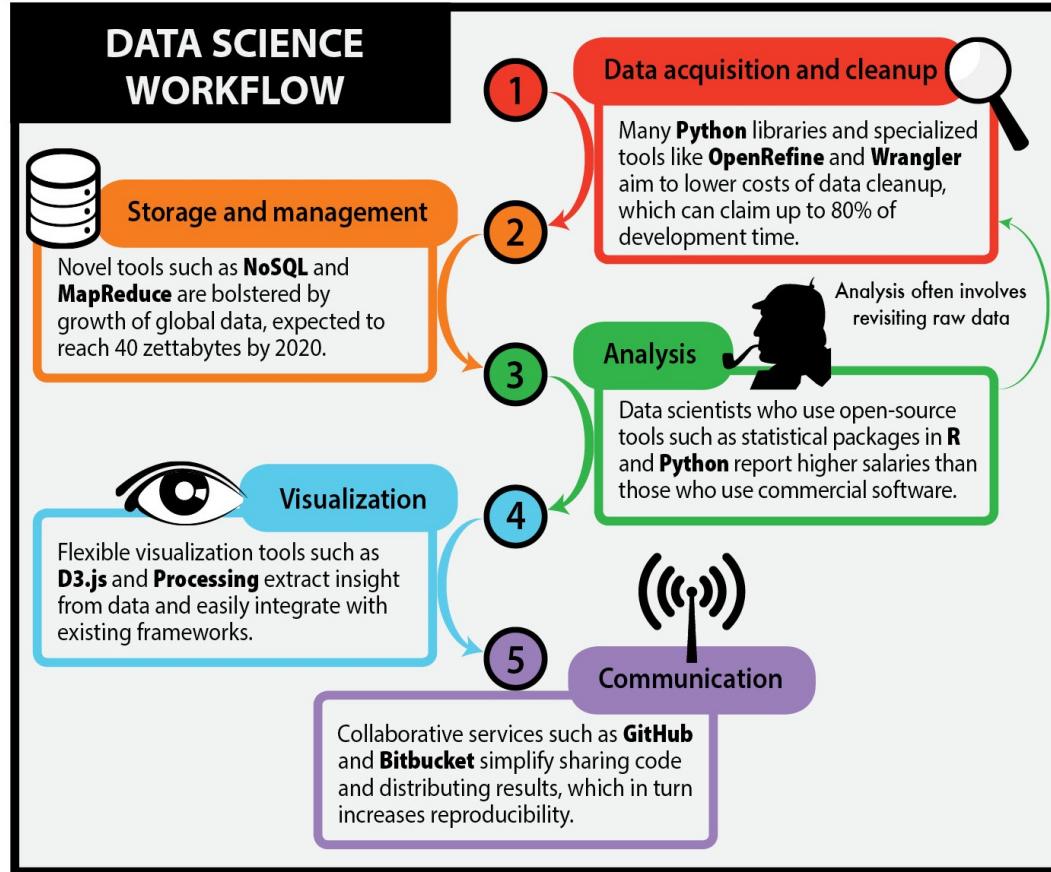
This diagram illustrates the *iterative* nature of problem solving

From Dataists Blog

1. Obtain
2. Scrub
3. Explore
4. Model
5. Interpret

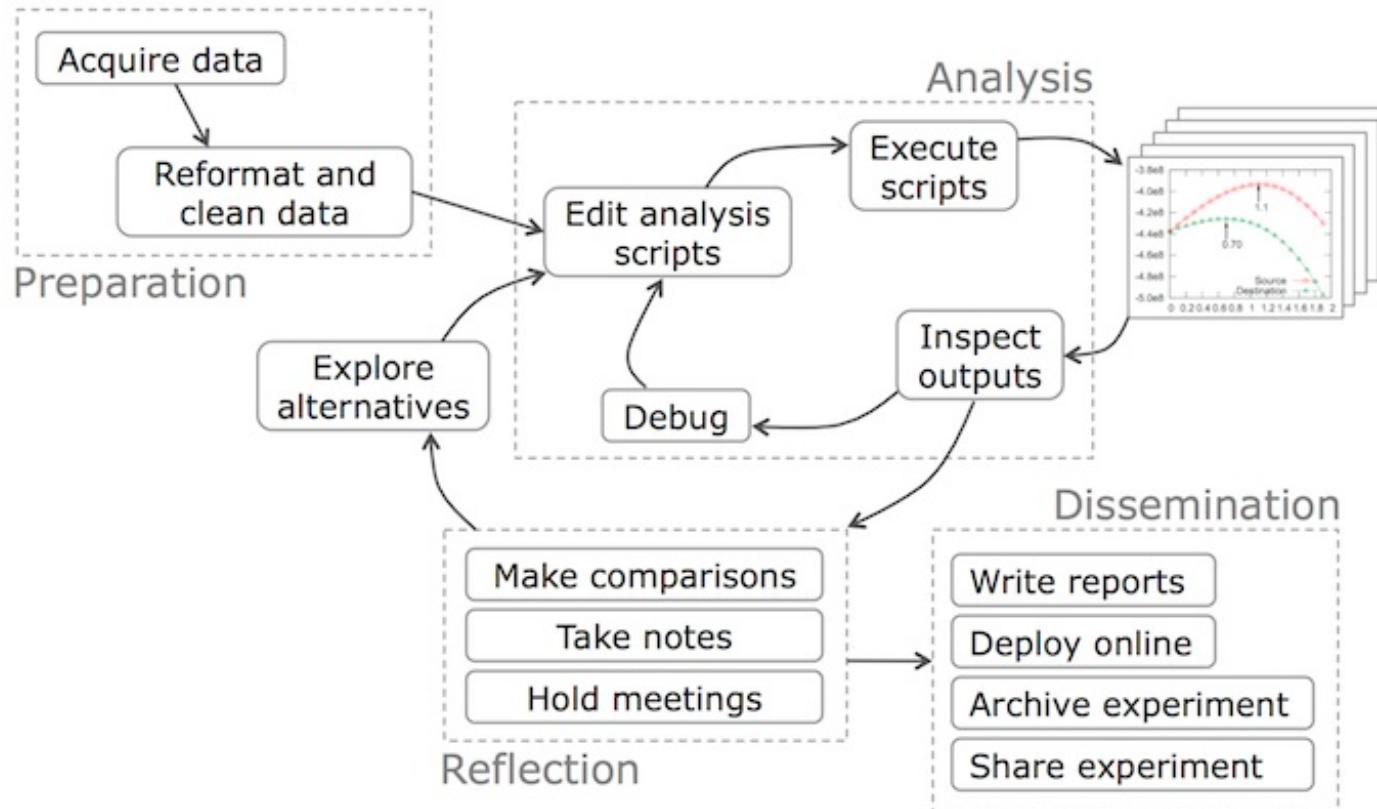
THE DATA SCIENCE WORKFLOW

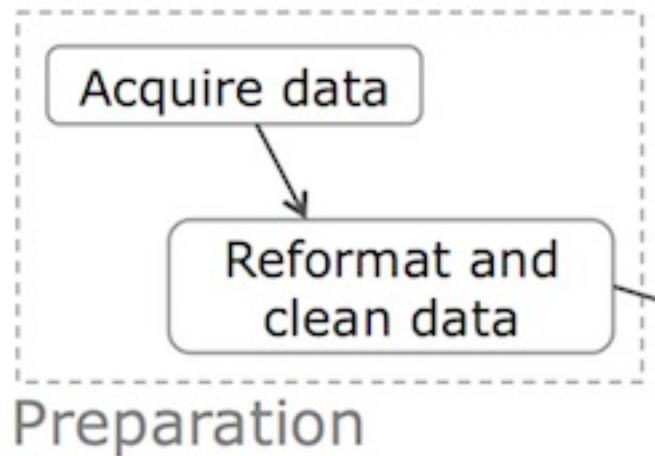
65

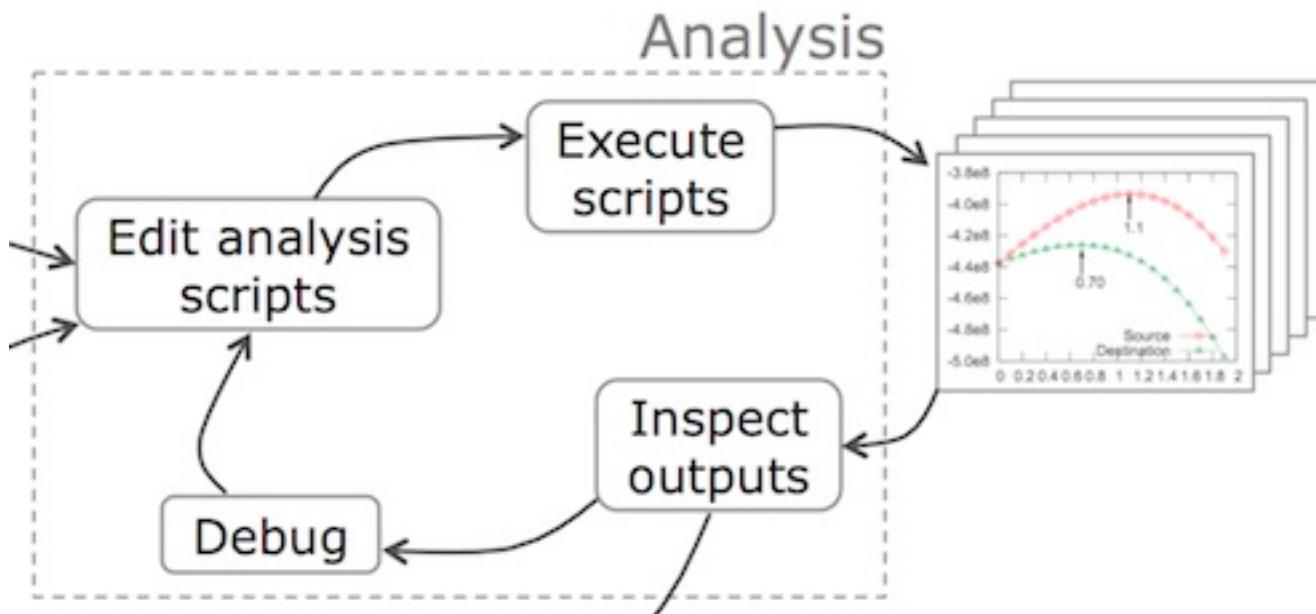


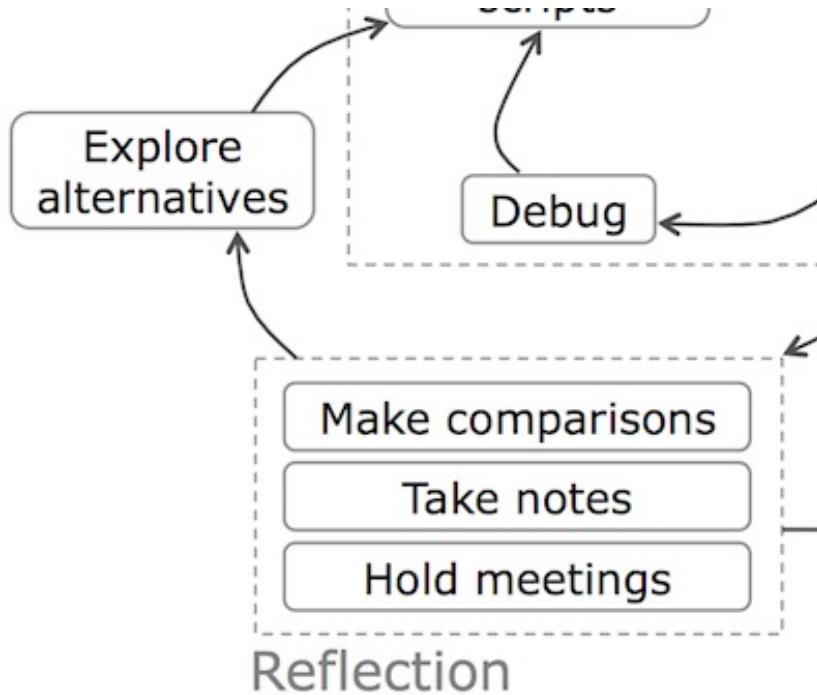
THE DATA SCIENCE WORKFLOW

66









PROBLEM: WHAT ARE THE LEADING INDICATORS THAT A USER WILL MAKE A NEW PURCHASE?

PROBLEM: WHAT ARE THE LEADING INDICATORS THAT A USER WILL MAKE A NEW PURCHASE?

1. Collect data around user retention, user actions within the product, potentially find data outside of company

PROBLEM: WHAT ARE THE LEADING INDICATORS THAT A USER WILL MAKE A NEW PURCHASE?

1. Collect data around user retention, user actions within the product, potentially find data outside of company
2. Extract aggregated values from raw data
 1. How many times did a user share through Facebook within a week? A month?
 2. How often did they open up our emails?

PROBLEM: WHAT ARE THE LEADING INDICATORS THAT A USER WILL MAKE A NEW PURCHASE?

1. Collect data around user retention, user actions within the product, potentially find data outside of company
2. Extract aggregated values from raw data
 1. How many times did a user share through Facebook within a week? A month?
 2. How often did they open up our emails?
3. Examine data to find common distributions and correlations

PROBLEM: WHAT ARE THE LEADING INDICATORS THAT A USER WILL MAKE A NEW PURCHASE?

1. Collect data around user retention, user actions within the product, potentially find data outside of company
2. Extract aggregated values from raw data
 1. How many times did a user share through Facebook within a week? A month?
 2. How often did they open up our emails?
3. Examine data to find common distributions and correlations
4. Extract new meaning to predict if a user would purchase again or not

PROBLEM: WHAT ARE THE LEADING INDICATORS THAT A USER WILL MAKE A NEW PURCHASE?

1. Collect data around user retention, user actions within the product, potentially find data outside of company
2. Extract aggregated values from raw data
 1. How many times did a user share through Facebook within a week? A month?
 2. How often did they open up our emails?
3. Examine data to find common distributions and correlations
4. Extract new meaning to predict if user would purchase again
5. Share results (and probably also go back to the drawing board)

Let's build an analytics team:

You are a data scientist of the following organizations:

GA, Pinterest, FourSquare, Bank of America, Think Coffee

Answer the following questions:

1. Define the top priorities of the organization
2. What products could you build or studies could you run?
3. What data would you need to collect?
4. What will your top challenges be?

III. WORKING AT THE UNIX COMMAND LINE

LET'S TAKE A LOOK AT THE 538 DATASET

```
cd ~  
mkdir repo  
cd repo  
git clone https://github.com/ga-students/DAT-23-NYC  
cd data/538
```

KEY OBJECTIVES

- Navigate the filesystem
- Move, copy, and delete files & directories
- View & search files
- Edit & interact with files
- Catch output & combine steps
- Learn more
- Panic

TOOLS

- ls, cd, ~ (tilde), clear, tab completion
- mv, cp, rm, mkdir, rmdir
- cat, head, tail, less, grep
- vim, emacs, tr, sort, uniq, wc, touch
- >, | (pipe), echo
- man, apropos
- ^C, ^Z (control)

NOTE

Being comfortable at the command line makes your life much easier!

INTRO TO DATA SCIENCE

DISCUSSION