

# GENERAL ASSEMBLY

# INTRO TO DATA SCIENCE

# MEET YOUR INSTRUCTORS



---

## ABOUT YOUR INSTRUCTOR!



Ruben Naeff

[rubennaeff@gmail.com](mailto:rubennaeff@gmail.com)

Data Science Instructor

DATA SCIENTIST  
AT KNEWTON

MUSIC COMPOSER  
STRATEGY CONSULTANT  
ECONOMIC RESEARCHER  
MATH TEACHER

AMSTERDAM, NL  
BROOKLYN, NY

---

## YOUR EXPERT IN RESIDENCE



Antoine Grant

lordtwon@gmail.com

Expert In Residence

DIRECTOR  
AT ROBB VICES

BUSINESS DEV.  
STRATEGY CONSULTANT

YOUR MENTOR &  
ADVOCATE

---

## **ABOUT YOUR PRODUCER!**



Daniel Demoray

[ddemoray@google.com](mailto:ddemoray@google.com)

Education Programs Producer

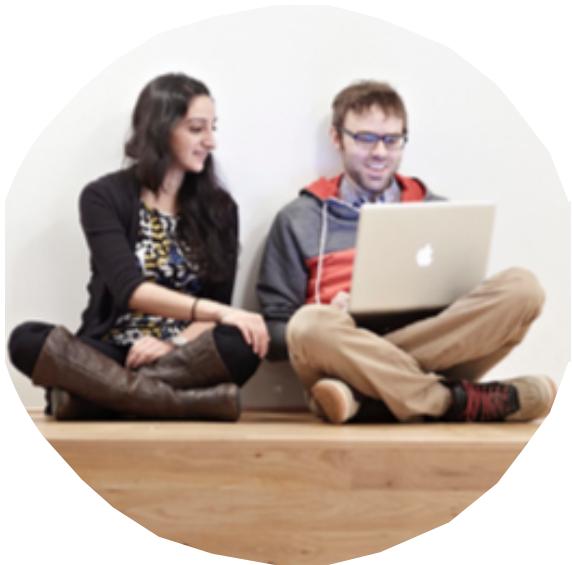
STUDENT

INSTRUCTOR

CURRICULUM

---

## ABOUT YOU!



Who  
are you

What  
do you do

Why  
are you here



# 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



```
# store the person objects in a people array
people = []
# display all the people in the array
people.each do |person|
  puts person
end
# Person is elementary and 15and years old.
# Would you like to create a (Person or Student)?
# Name: 
# Age: 
# Gender: 
# Student: 
# Grade: 
# Person(name:, age:, gender:)
# The people you have are: [people.join(", ")]
```

# 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.

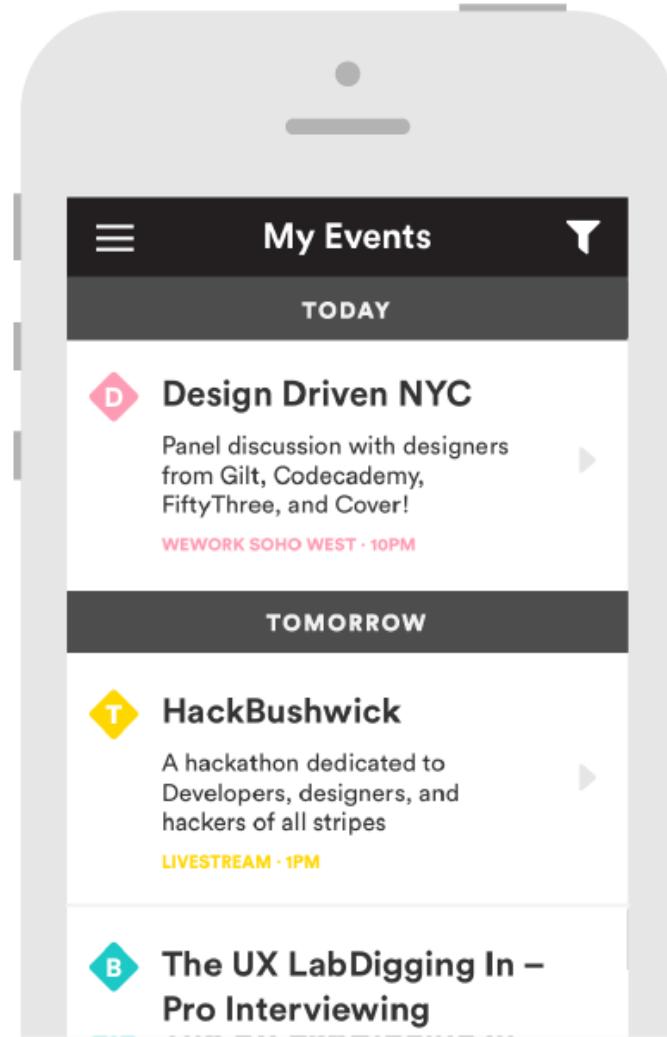


# Hub

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

Join at [hub.ga.co/ny](http://hub.ga.co/ny)

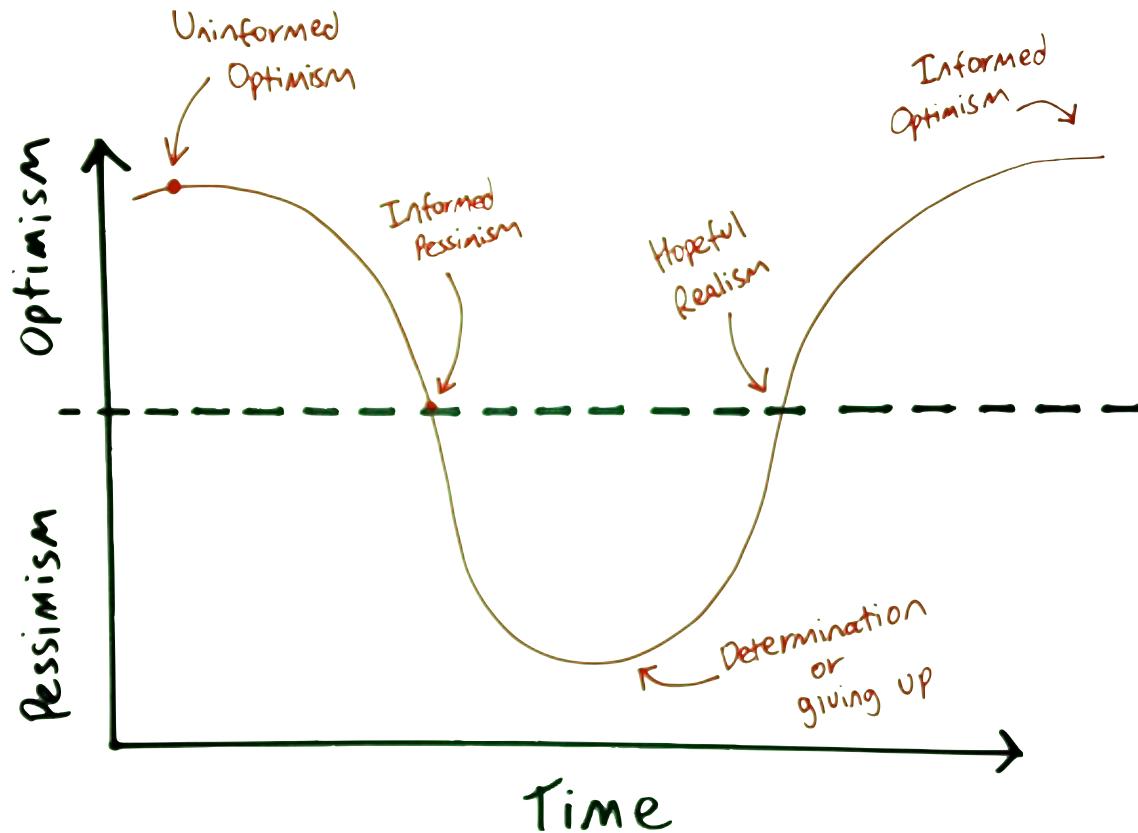
Take a sneak peek of the mobile app at  
[bit.ly/hubiosbeta](http://bit.ly/hubiosbeta)



# ROAD TO SUCCESS

A photograph of a classroom setting. Several students are seated at white desks, facing forward. In the foreground, a student with long blonde hair is seen from the side, writing in a notebook. Behind her, other students are visible, including a man in a plaid shirt and a woman with her hand near her chin. A red coffee cup sits on one of the desks. The background shows a chalkboard and more students in the distance.

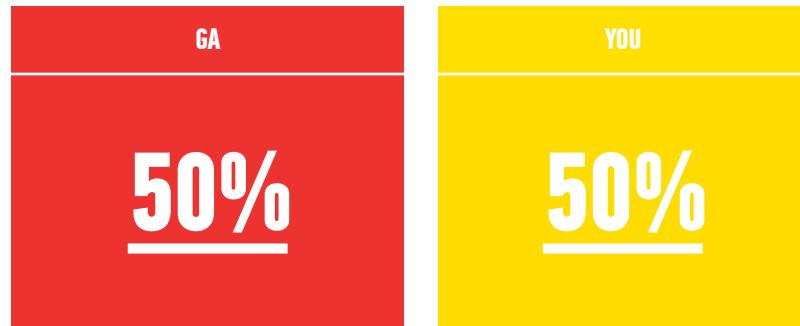
# EMOTIONAL CYCLE OF CHANGE



---

## 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**



# GA Graduation Requirements

## HOMEWORK

(COMPLETE 80% OF  
HOMEWORK/LABS)

## ATTENDANCE

(MISS NO MORE THAN 2 CLASSES)

## FINAL PROJECT

## COMMUNITY PARTICIPATION + FEEDBACK

Please let us  
know **ASAP** when  
you know that  
you'll miss a class



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

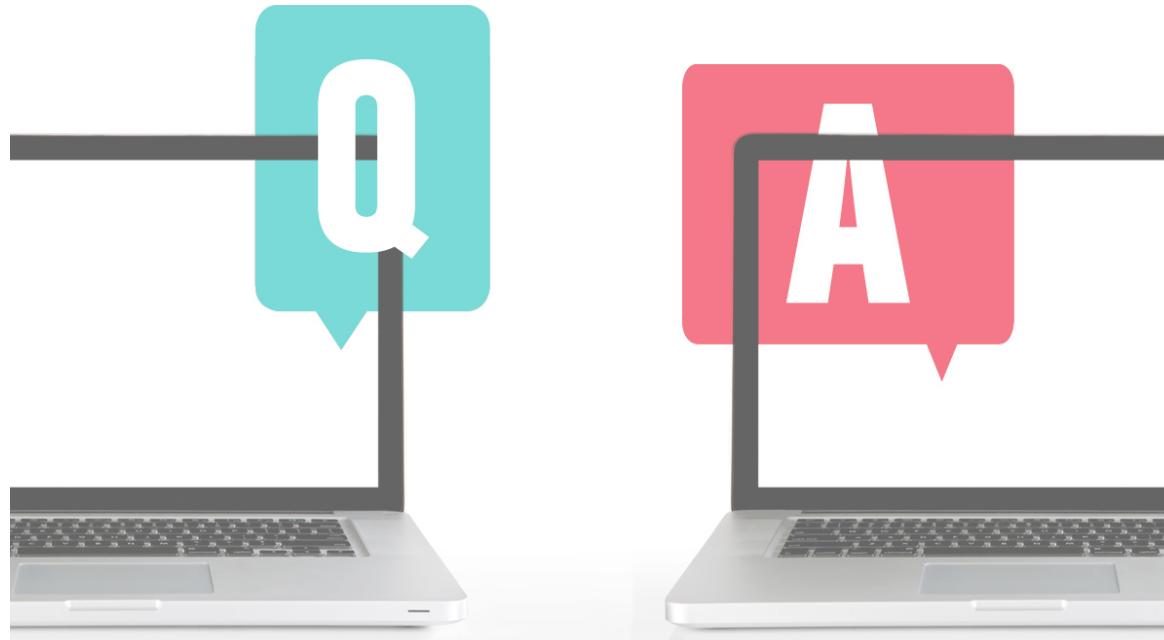
You're part of the alumni community forever

## 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](mailto:ddemoray@google.com)

Ruben Naeff, instructor – [rubennaeff@gmail.com](mailto:rubennaeff@gmail.com)

Antoine Grant, expert in residence – [antoinejgrant@gmail.com](mailto:antoinejgrant@gmail.com)

E-group – [dat-nyc-23@google-groups.com](mailto:dat-nyc-23@google-groups.com)

Slack – [ganyceveningcourses.slack.com/messages/data-science-23](https://ganyceveningcourses.slack.com/messages/data-science-23)

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 Expert in residence)

Course notes, syllabus, slides, data, code, etc... are all available on

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

JULY

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?

36

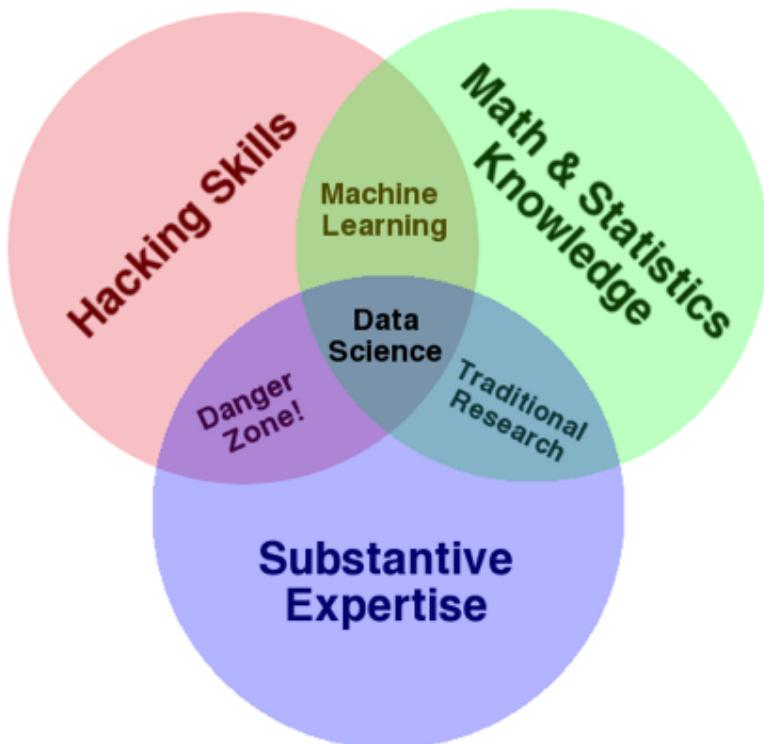


The United States alone faces a shortage of

- 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.

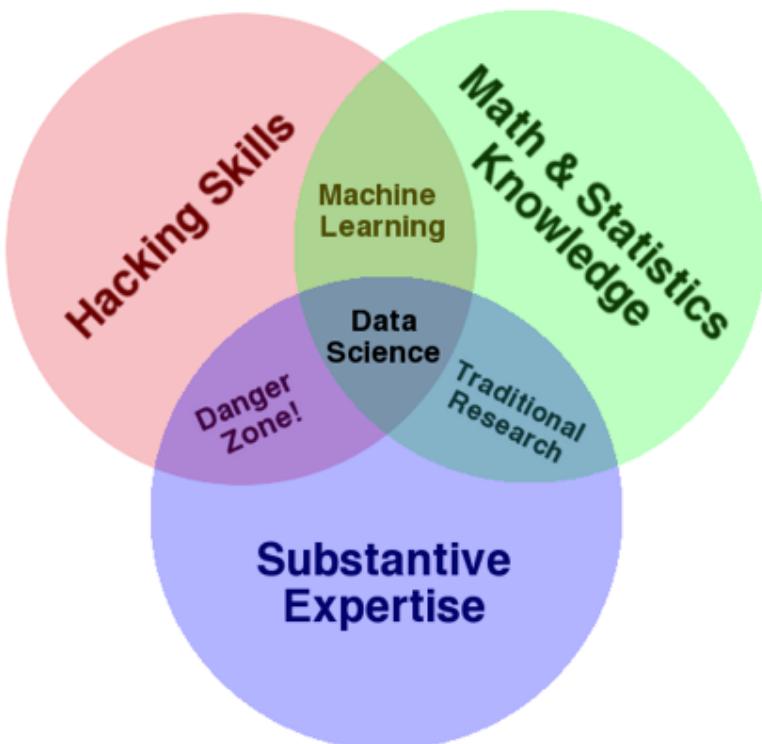
# WHAT IS DATA SCIENCE?

38



# WHAT IS DATA SCIENCE?

39

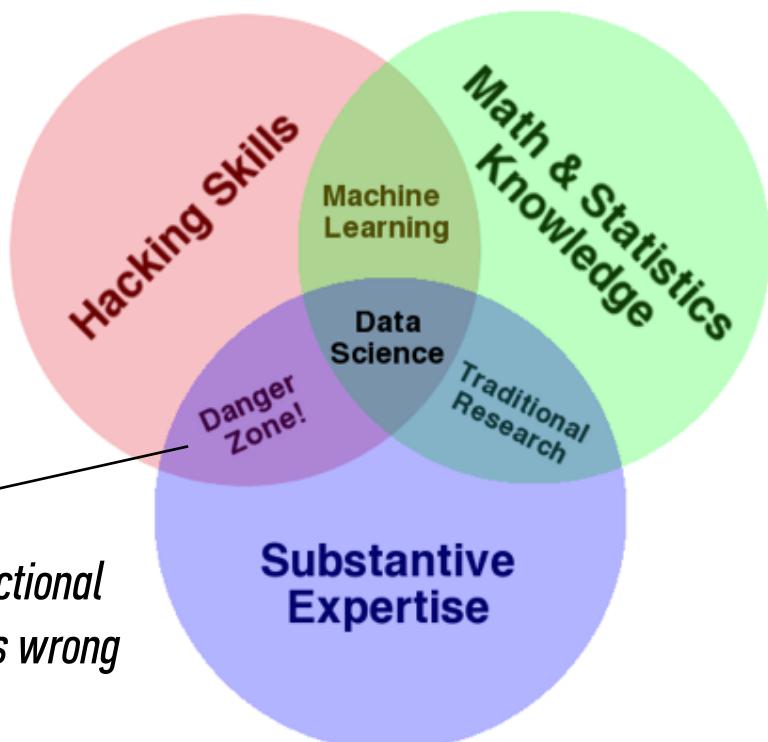


**ONE MORE THING!**

Communication skills

# WHAT IS DATA SCIENCE? – DANGER ZONE △

40

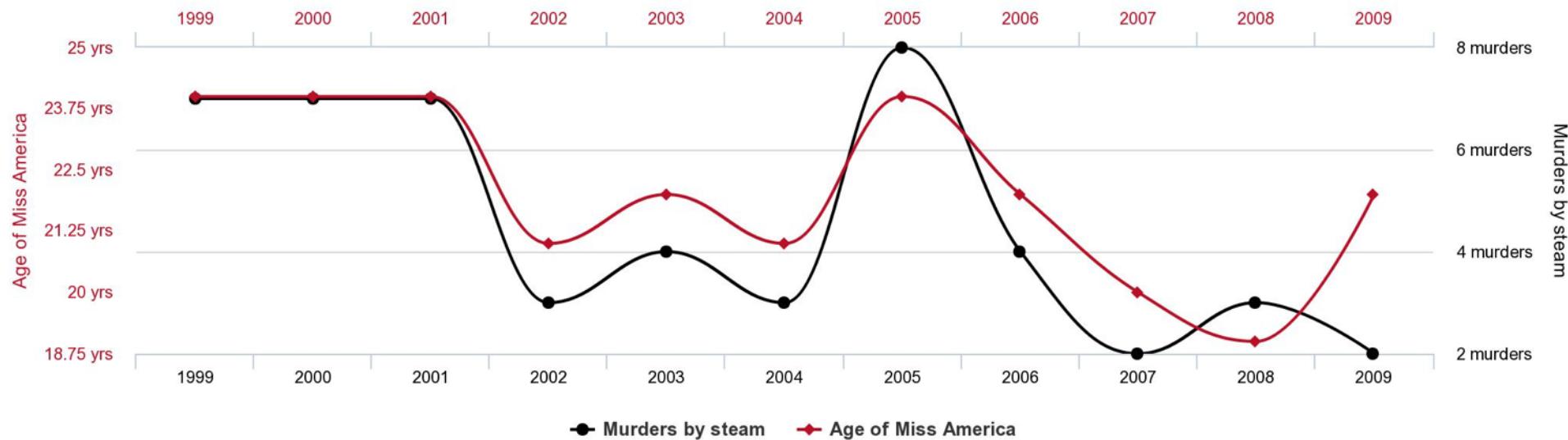


*Students who watch more instructional videos, tend to get more answers wrong*

# WHAT IS DATA SCIENCE? – DANGER ZONE △

41

## 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.

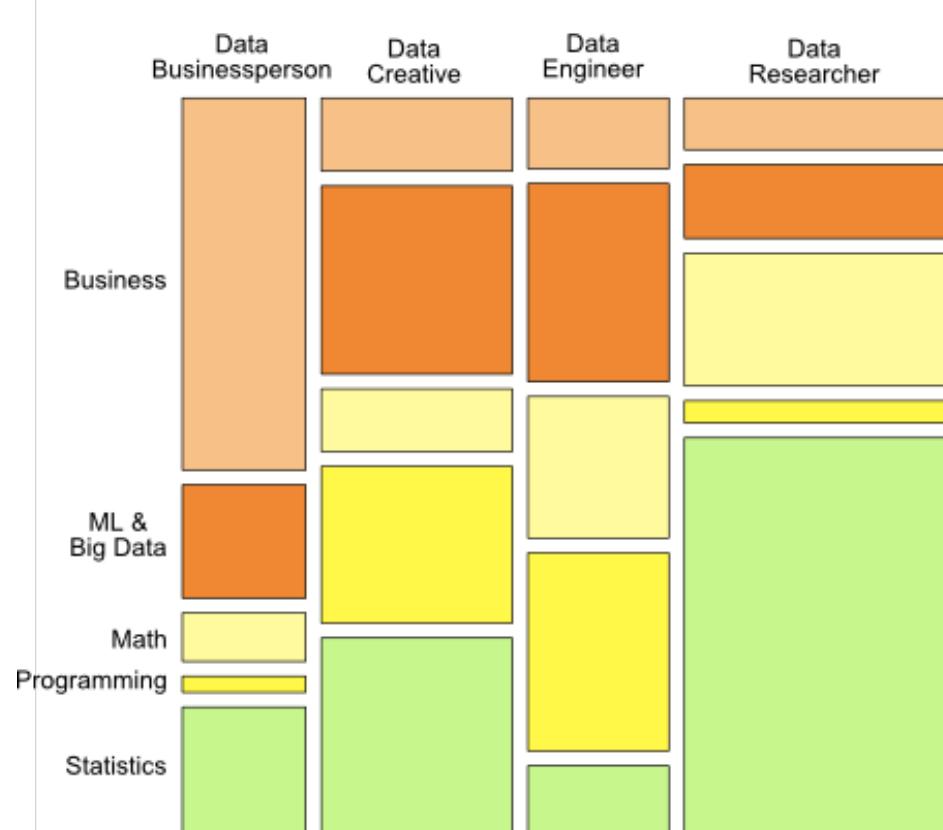
# WHAT IS DATA SCIENCE?

46

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?

47



source: *Analyzing the Analyzers*



**Michael E. Driscoll**  
@medriscoll



+ Follow

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



RETWEETS

35

FAVORITES

26



4:57 PM - 17 Jul 2012

## WHO USES DATA SCIENCE?

49



# WHO USES DATA SCIENCE?

50

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

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

1 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

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

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

\$30 Off Instantly ▾ Learn more

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

What Other Customers Are Looking At Right Now

 SanDisk Ultra 64 GB MicroSDXC Class 10 ★★★★★ (4,513) \$60.00 \$42.99	 Samsung Galaxy Tab 2 (7-Inch, Wi-Fi) ★★★★★ (2,401) \$249.00 \$179.00	 Kindle Fire 7", LCD Display, Wi-Fi, 8 GB Amazon Digital Services Inc. ★★★★★ (9,150) \$159.00	 Kindle, 6" E Ink Display, Wi-Fi, 8 GB Amazon Digital Services Inc. ★★★★★ (4,278) \$69.00	 Kindle Fire HD 7", Dolby Audio, 8 GB Amazon Digital Services Inc. ★★★★★ (12,817) \$199.00	 Kindle Fire HD 8.9", Dolby Audio, 16 GB Amazon Digital Services Inc. ★★★★★ (6,470) \$269.00	 Amazon Gift Card - E-mail \$50.00
---	---	--	---	---	---	---

Digital Cameras Best Sellers

 Nikon COOLPIX S9200 16 MP CMOS... ★★★★★ (1,483)	 Canon PowerShot SX500 IS 16.0 MP... ★★★★★ (1,640)	 Canon PowerShot A2300 IS 16.0 MP... ★★★★★ (1,643)	 Canon EOS Rebel T3i 18 MP CMOS... ★★★★★ (778)	 Canon PowerShot SX150 IS 14.1 MP... ★★★★★ (1,920)	 Canon PowerShot SX260 HS 12.1 MP CMOS... ★★★★★ (1,923)	 Canon EOS Rebel T4i 18.0 MP CMOS... ★★★★★ (1,640)
--	--	--	--	---	---	--

# WHO USES DATA SCIENCE?

amazon Join Prime

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

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
- 1 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

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

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

\$30 Off Instantly > Learn more

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

What Other Customers Are Looking At Right Now

						
SanDisk Ultra 64 GB MicroSDXC Class 10... ★★★★★ (4,513) \$69.00 <b>\$42.99</b>	Samsung Galaxy Tab 2 (7-Inch, Wi-Fi) ★★★★★ (2,401) \$249.00 <b>\$179.00</b>	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) <b>\$50.00</b>

Digital Cameras Best Sellers

						
Nikon COOLPIX S9200 16 MP CMOS... ★★★★★ (4,933)	Canon PowerShot SX500 IS 16.0 MP... ★★★★★ (1,603)	Canon PowerShot A2300 IS 16.0 MP... ★★★★★ (2,423)	Canon EOS Rebel T3i 18 MP CMOS... ★★★★★ (778)	Canon PowerShot SX150 IS 14.1 MP... ★★★★★ (2,923)	Canon PowerShot SX260 HS 12.1 MP CMOS... ★★★★★ (2,923)	Canon EOS Rebel T4i 18.0 MP CMOS... ★★★★★ (1,619)

# WHO USES DATA SCIENCE?

amazon Join Prime

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

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 & Pet  
Clothing, Shoes & Jewelry  
Sports & Outdoors  
Automotive & Industrial  
1 Full Store Directory

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

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

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 Class 10 ★☆☆☆☆ (4,513) \$60.00 \$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" 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

Digital Cameras Best Sellers

						
Nikon COOLPIX S9200 16 MP CMOS... \$199.00 \$149.00	Canon PowerShot SX500 IS 16.0 MP... \$199.00 \$149.00	Canon PowerShot A2300 IS 16.0 MP... \$199.00 \$149.00	Canon EOS Rebel T3i 18 MP CMOS... \$299.00 \$199.00	Canon PowerShot SX150 IS 14.1 MP... \$199.00 \$149.00	Canon PowerShot SX260 HS 12.1 MP CMOS... \$199.00 \$149.00	Canon EOS Rebel T4i 18.0 MP CMOS... \$299.00 \$199.00

Father's Day Savings Sponsored by DEWALT [Shop now](#)

Hello, Sign In Your Account ▾ Join Prime ▾  Cart ▾  Wish List ▾

Class of 2013 Graduation Gifts [Shop now](#)

Try Amazon Prime free for 30 days [Get Started](#)

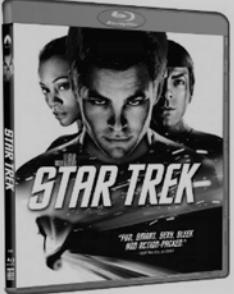
amazon Prime Advertisement

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](#)

\$30 Off Instantly [Learn more](#)

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

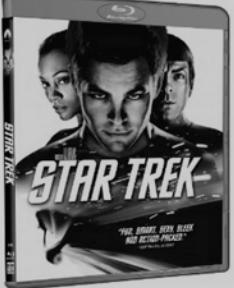
TechShowMe Add to Cart

\$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 \$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.

## Sell Us Your Item

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

Trade in

## More Buying Choices

TechShowMe

\$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?

55



---

## INTRO TO DATA SCIENCE

---

# II. THE DATA SCIENCE WORKFLOW

acquire      parse      filter      mine      represent      refine      interact





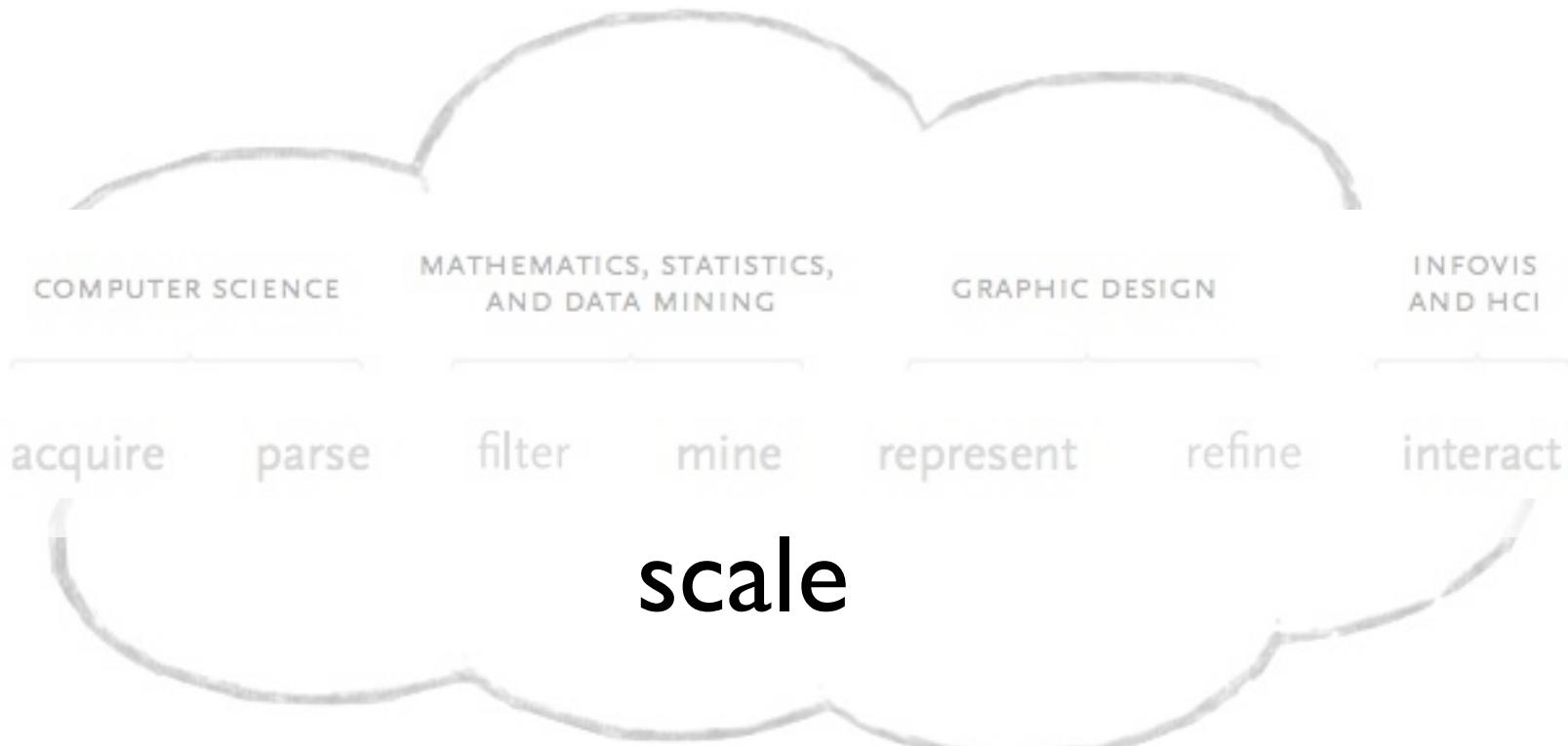


# THE DATA SCIENCE WORKFLOW

61







# THE DATA SCIENCE WORKFLOW

64



## NOTE

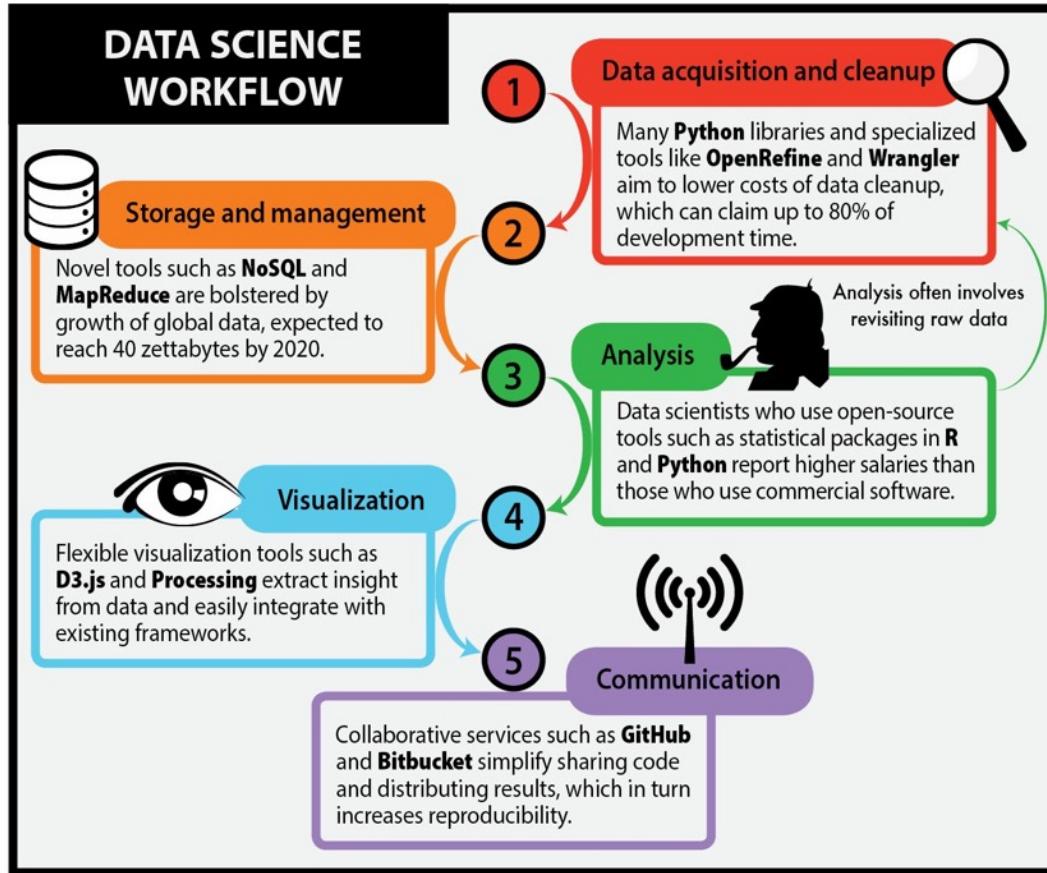
This diagram illustrates the *iterative* nature of problem solving

From Dataists Blog

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

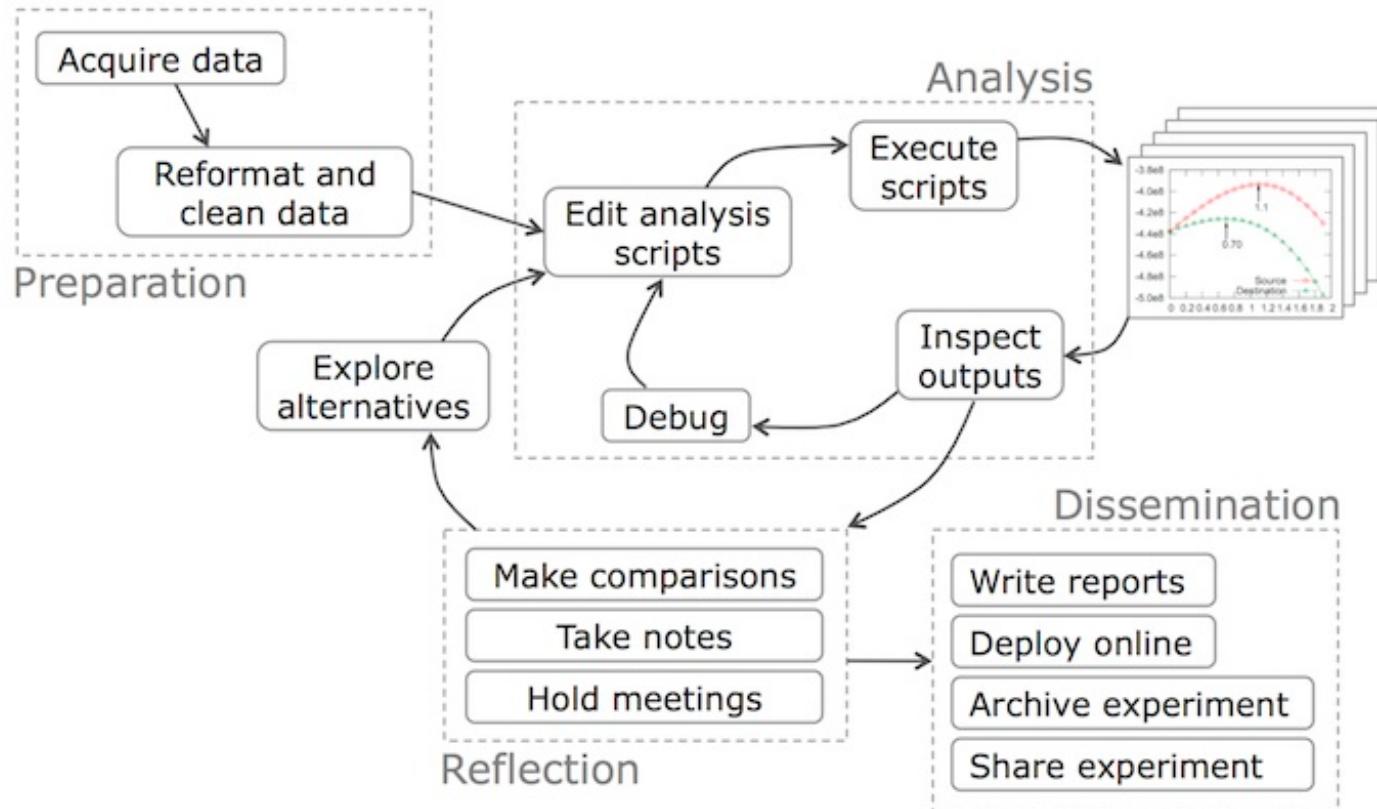
# THE DATA SCIENCE WORKFLOW

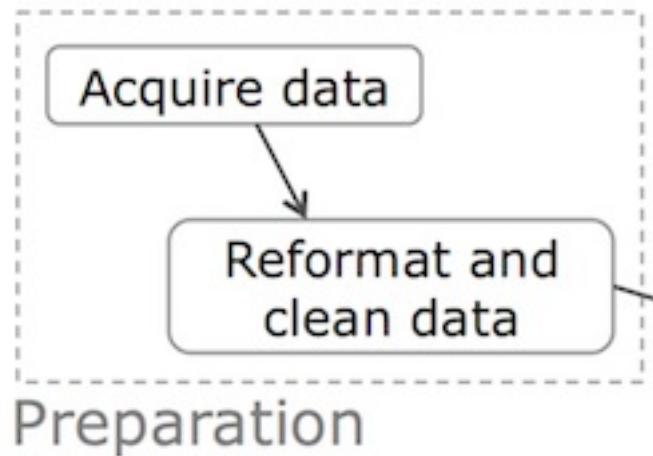
66

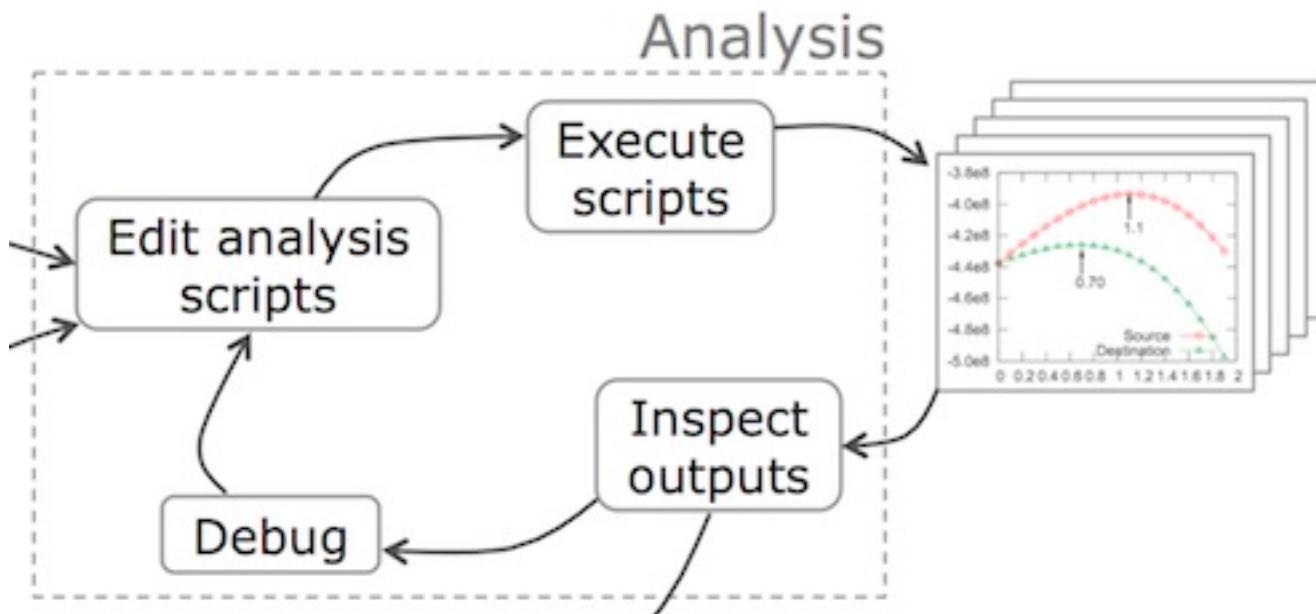


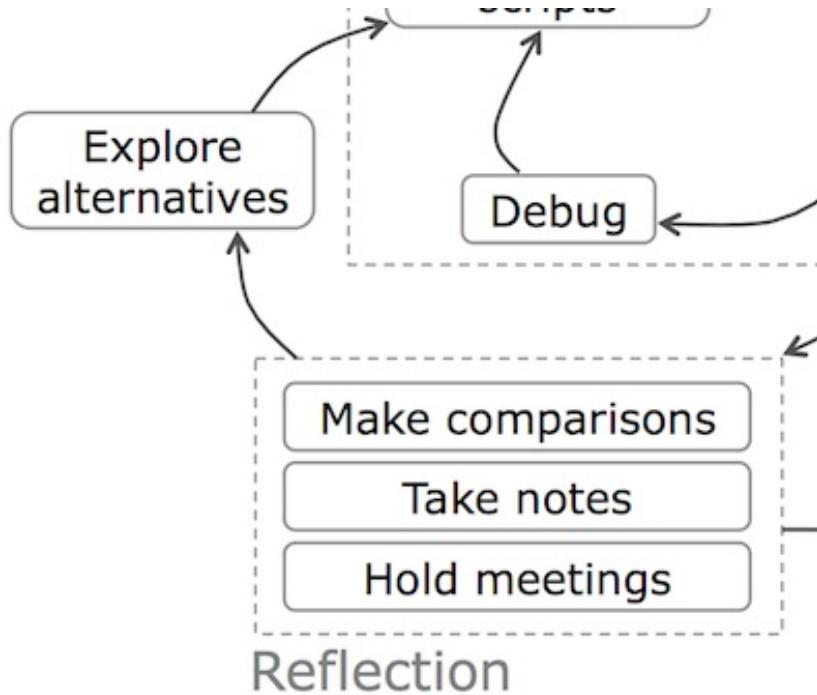
# THE DATA SCIENCE WORKFLOW

67









### PROBLEM: WHICH CUSTOMERS ARE LIKELY TO MAKE ANOTHER PURCHASE?

### PROBLEM: WHICH CUSTOMERS ARE LIKELY TO MAKE ANOTHER PURCHASE?

If we know this, then we might want to

- target these existing customers more specifically, and
- find indicators in our products or channels and focus on those.

### PROBLEM: WHICH CUSTOMERS ARE LIKELY TO MAKE ANOTHER PURCHASE?

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

### PROBLEM: WHICH CUSTOMERS ARE LIKELY TO MAKE ANOTHER 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: WHICH CUSTOMERS ARE LIKELY TO MAKE ANOTHER 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: WHICH CUSTOMERS ARE LIKELY TO MAKE ANOTHER 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: WHICH CUSTOMERS ARE LIKELY TO MAKE ANOTHER 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

- › Go to [github.com/ga-students/DAT-23-NYC](https://github.com/ga-students/DAT-23-NYC)
- › Clone the repo if you haven't done so already
- › Scroll down to lesson #1 and click on [Data Science at the Command Line](#)

## 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**