

INTRO TO DATA SCIENCE

LECTURE 1: DATA EXPLORATION

Andrew Parker – adparker@gmail.com

Sam Brotherton – sbrother@gmail.com

INTRO TO DATA SCIENCE

WELCOME!

COURSE MEETING:

T/TH 7:00PM - 10:00PM

GA 1520 2ND STREET

COURSE NOTES:

[HTTPS://GITHUB.COM/ADPARKER/GADSLA_1403](https://github.com/ADPARKER/GADSLA_1403)

COURSE MAILS: SCHOOLOGY (SIGN UP!)

I. WHAT IS DATA SCIENCE?

II. THE DATA MINING WORKFLOW

EXERCISES:

III. WORKING AT THE UNIX COMMAND LINE

INTRO TO DATA SCIENCE

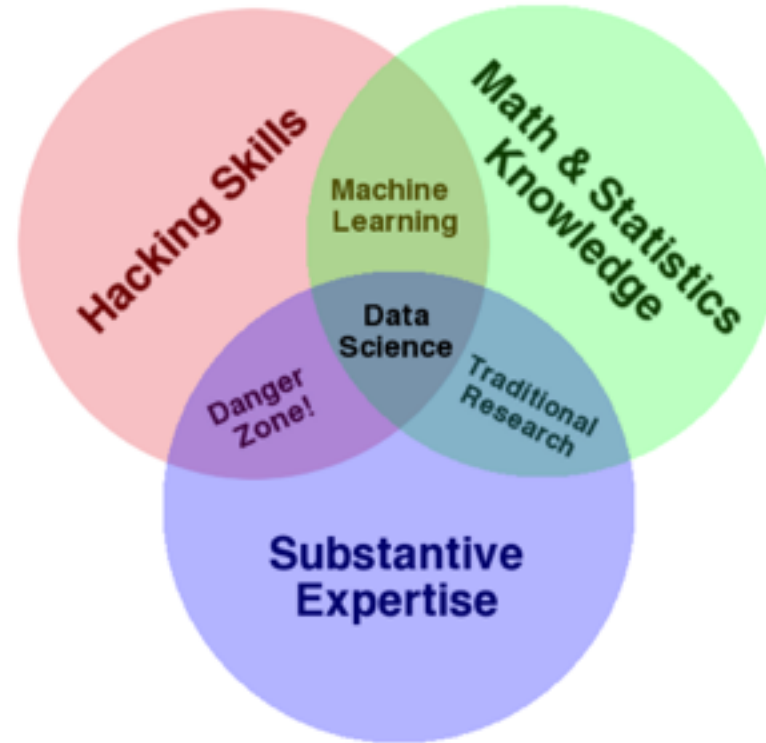
I. WHAT IS DATA SCIENCE?

- **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-oriented subject.**

THE QUALITIES OF A DATA SCIENTIST

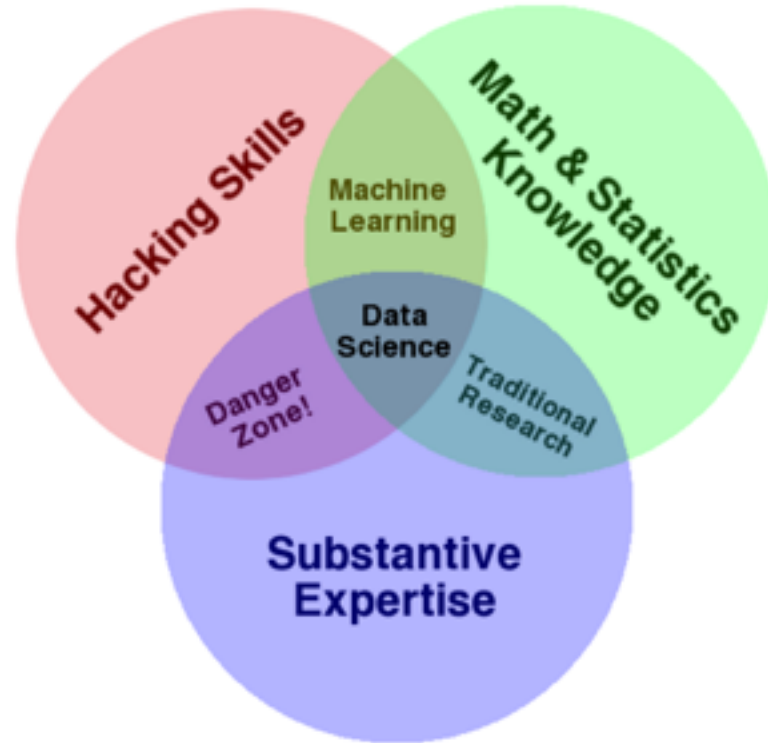
8



source: <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>

THE QUALITIES OF A DATA SCIENTIST

9



ONE MORE THING!

Communication
skills

source: <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>

- **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.**
- **A rapidly growing field.**



[Your Amazon.com](#)
[Today's Deals](#)
[Gift Cards](#)
[Sell](#)
[Help](#)

[Shop by Department](#)

[Hello, Sign in](#)
[Your Account](#)
[Join Prime](#)
[Cart](#)
[List](#)

- Unlimited Instant Videos
- MP3s & Cloud Player
- 20 million songs, play anywhere
- Amazon Cloud Drive
- A 1TB of free storage
- Kindle
- Appstore for Android
- Read what's new 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
- 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~~ **\$179**

Enter **DADDY** at checkout

Offer valid through June 6, 2013

Top Story [Outfit Trends](#) [Fast Free Shipping](#)

Amazon Fashion

Dress Shop

Our favorite wear everywhere styles from Sunday by Shelli Segal, London Tones, and more

Shop Dresses [Shop All Clothing](#)

Class of 2013

Graduation Gifts

Shop now

Try Amazon Prime

free for 30 days

[Get Started](#)

amazon Prime

[Advertise on Amazon](#)

Deal of the Day

Up to 60% Off

Select BestDisk Memory

[See Details](#)

UP by Jawbone

Measure your activity, sleep quality, and live better

[Hear More](#)

\$30 Off Instantly

[Shop Now](#)

Noncredit courses and certificates

Center for Advanced Digital Applications (CADA)

There's 422 Time to...

What Other Customers Are Looking At Right Now

SanDisk UltraFit SDHC Memory Card

★★★★★ (2,814)

Price \$42.99

Samsung Galaxy Tab 2 (7-inch, Wi-Fi)

★★★★★ (2,470)

\$719.00 ~~\$759.00~~

Kindle Fire 7", LCD Display, Wi-Fi, & Amazon Digital Navigation System

★★★★★ (8,188)

\$159.00

Kindle 7" C Ink Display, Wi-Fi, & Amazon Digital Navigation System

★★★★★ (4,478)

\$69.00

Kindle Fire HD 7", Galaxy Audio, & Amazon Digital Navigation System

★★★★★ (4,917)

\$199.00

Kindle Fire HD 8.9", Galaxy Audio, & Amazon Digital Navigation System

★★★★★ (5,470)

\$299.00

Amazon Gift Card - Email

★★★★★ (19,180)

\$50.00

Digital Cameras Real Sellers

Nikon COOLPIX W100 16 MP CMOS

★★★★★ (1,000)

Canon PowerShot SX600 HS 16.0 MP

★★★★★ (1,000)

Canon PowerShot A1000 HS 16.0 MP

★★★★★ (1,000)

Canon PowerShot Rebel T6 18 MP CMOS

★★★★★ (719)

Canon PowerShot SX700 HS 16.1 MP

★★★★★ (1,000)

Canon PowerShot SX600 HS 16.1 MP CMOS

★★★★★ (1,000)

Canon PowerShot Rebel T6 18.0 MP CMOS

★★★★★ (1,000)

[Your Amazon.com](#)
[Today's Deals](#)
[Gift Cards](#)
[Sell](#)
[Help](#)

[Amazon.com](#)

[Hello, Sign in](#)
[Your Account](#)
[Join Prime](#)
[Cart](#)
[List](#)

- Unlimited Instant Videos
- MP3s & Cloud Player
- 20 million songs, play anywhere
- Amazon Cloud Drive
- A TB of free storage
- Kindle
- Apps for Android
- ReadSmart System - 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
- Full Store Directory

The Perfect Gift for Dad

Kindle Fire HD

From \$199 From \$179

Enter DADS2013 at checkout

Offer valid through June 6, 2013

Amazon Fashion

Dress Shop

Our favorite wear everywhere styles from Sunday by Shelli Segal, London Tones, and more

Shop All Clothing

Instant Video

MP3 Store

Cloud Player

Kindle

Cloud Drive

Apps for Android

Digital Games & Software

Audible Audiobooks

Class of 2013 Graduation Gifts

Shop now

Try Amazon Prime free for 30 days

Get Started

Deal of the Day

Up to 60% Off

Defiant HardDisk Memory

Shop now

What Other Customers Are Looking At Right Now

SanDisk Ultra SD 64GB MicroSDHC (Class 10)

★★★★★ (4,816)

Price \$42.99

Samsung Galaxy Tab 2 (7-inch, Wi-Fi)

★★★★★ (2,407)

\$719.99 \$174.00

Kindle Fire 7" LCD Display, Wi-Fi, & Amazon-Regd. Revolution Box

★★★★★ (5,188)

\$199.00

Kindle 8" C Ink Display, Wi-Fi, & Amazon-Regd. Revolution Box

★★★★★ (4,278)

\$69.00

Kindle Fire HD 7" Galaxy Audio, & Amazon-Regd. Revolution Box

★★★★★ (4,297)

\$199.00

Kindle Fire HD 8.9" Galaxy Audio, & Amazon-Regd. Revolution Box

★★★★★ (4,470)

\$299.00

Amazon Gift Card - Email

★★★★★ (19,180)

\$60.00

UP by Jawbone

Measure your activity, sleep quality, and live better

FREE

Digital Cameras Best Sellers

Nikon COOLPIX W300 16 MP CMOS

★★★★★ (1,000)

Canon PowerShot SX600 HS 16.0 MP

★★★★★ (1,000)

Canon PowerShot A2000 HS 16.0 MP

★★★★★ (1,000)

Canon PDS Rebel T5 18 MP CMOS

★★★★★ (1,000)

Canon PowerShot SX160 HS 16.1 MP

★★★★★ (1,000)

Canon PowerShot SX200 HS 12.1 MP CMOS

★★★★★ (1,000)

Canon PDS Rebel T5 18.0 MP CMOS

★★★★★ (1,000)

Noncredit courses and certificates

Center for Advanced Digital Applications (CADA)

There's K&S Time to...

[Your Amazon.com](#)
[Today's Deals](#)
[Gift Cards](#)
[Sell](#)
[Help](#)

[Shop by Department](#)

[Hello, Sign in to Your Account](#)
[Join Prime](#)
[Cart](#)
[List](#)

Unlimited Instant Videos

MP3s & Cloud Player

20 million songs, play anywhere

Amazon Cloud Drive

5 TB of free storage

Kindle

Appstore for Android

Read aloud Text-to-Speech

Digital Games & Software

Audible Audiobooks

Books

Movies, Music & Games

Firephones & Computers

Home, Garden & Tools

Grocery, Health & Beauty

Toys, Kids, Baby & Pet

Clothing, Shoes & Jewelry

Sports & Outdoors

Automotive & Industrial

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 DADSDAY at checkout

Offer valid through June 6, 2013

Top Story Clothing Trends Fast Free Shipping

Amazon Fashion

Dress Shop

Our favorite wear everywhere styles from

laundry by Shelli Segal, London Tones, and more

Shop Dresses Shop All Clothing

Class of 2013 Graduation Gifts

Shop now

Try Amazon Prime free for 30 days

Get Started

Amazon Prime

Activate & Renew

Deal of the Day

Up to 60% Off

Delixi HardDisk Memory

Shop now

UP by Jawbone

Measure your activity, sleep quality, and live better

HARDBONE

\$30 Off Instantly

Shop now

What Other Customers Are Looking At Right Now

<p>SanDisk Ultra SDHC 16GB Memory Card</p> <p>4.9 (1,111)</p> <p>\$42.99</p>	<p>Samsung Galaxy Tab 2 (7-inch, Wi-Fi)</p> <p>4.5 (1,111)</p> <p>\$179.00</p>	<p>Kindle Fire 7" LCD Display, Wi-Fi, & Amazon Digital Navigation System</p> <p>4.8 (1,111)</p> <p>\$159.00</p>	<p>Kindle Fire 7" C Ink Display, Wi-Fi, & Amazon Digital Navigation System</p> <p>4.8 (1,111)</p> <p>\$159.00</p>	<p>Kindle Fire HD 7" Galaxy Audio, Amazon Digital Navigation System</p> <p>4.8 (1,111)</p> <p>\$199.00</p>	<p>Kindle Fire HD 8.9" Galaxy Audio, Amazon Digital Navigation System</p> <p>4.8 (1,111)</p> <p>\$289.00</p>	<p>Amazon Gift Card - Credit</p> <p>4.9 (1,111)</p> <p>\$50.00</p>
--	--	---	---	--	--	--

Digital Cameras Best Sellers

<p>Nikon COOLPIX W300 16 MP CMOS</p> <p>4.9 (1,111)</p>	<p>Canon PowerShot SX600 HS 16.0 MP</p> <p>4.9 (1,111)</p>	<p>Canon PowerShot A1000 IS 16.0 MP</p> <p>4.9 (1,111)</p>	<p>Canon PowerShot T5 16 MP CMOS</p> <p>4.9 (1,111)</p>	<p>Canon PowerShot SX160 IS 16.1 MP</p> <p>4.9 (1,111)</p>	<p>Canon PowerShot SX280 HS 12.1 MP CMOS</p> <p>4.9 (1,111)</p>	<p>Canon PowerShot T5 16.0 MP CMOS</p> <p>4.9 (1,111)</p>
---	--	--	---	--	---	---

Noncredit courses and certificates

Center for Advanced Digital Applications (CADA)

There's Still Time to...



Roll over image to zoom in



[See 1 customer image](#)

[Share your own customer images](#)

Star Trek [Blu-ray] (2009)

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

[Reviews](#) (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 30 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

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
[Reviews](#) (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
[Reviews](#) (371)
\$53.56



Sin City (Two-Disc Theatrical & Recut, Extended, and Unrated Versions) [Blu-ray] ~ Jessica Alba Blu-ray
[Reviews](#) (933)
\$4.99

Quantity:

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

[TechShowable](#) [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](#) [Facebook](#) [Twitter](#) [Pinterest](#)



Roll over image to zoom in



See 1 customer image

Share your own customer images

Star Trek [Blu-ray] (2009)

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

4.4 out of 5 stars (2,040 customer reviews)

List Price: **\$22.99**

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 30 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

Watch instantly with amazon instant video		Rent	Buy		
Star Trek (2009)		\$2.99	\$9.99		
Other Formats & Versions		Amazon Price	New from	Used from	
(1)	Blu-ray	1-Disc Version	\$9.99	\$9.98	\$9.78
(2)	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:

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

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

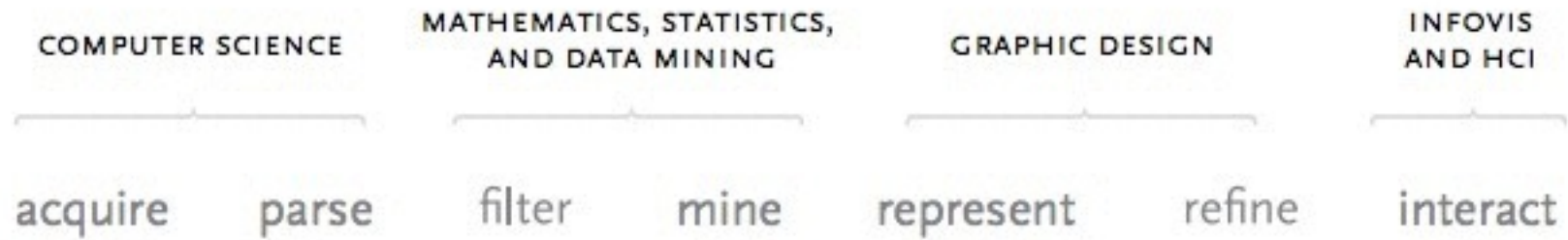
43 used & new from \$9.78

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

Share [Facebook](#) [Twitter](#) [Pinterest](#) [Google+](#)

II. THE DATA SCIENCE WORKFLOW

acquire parse filter mine represent refine interact











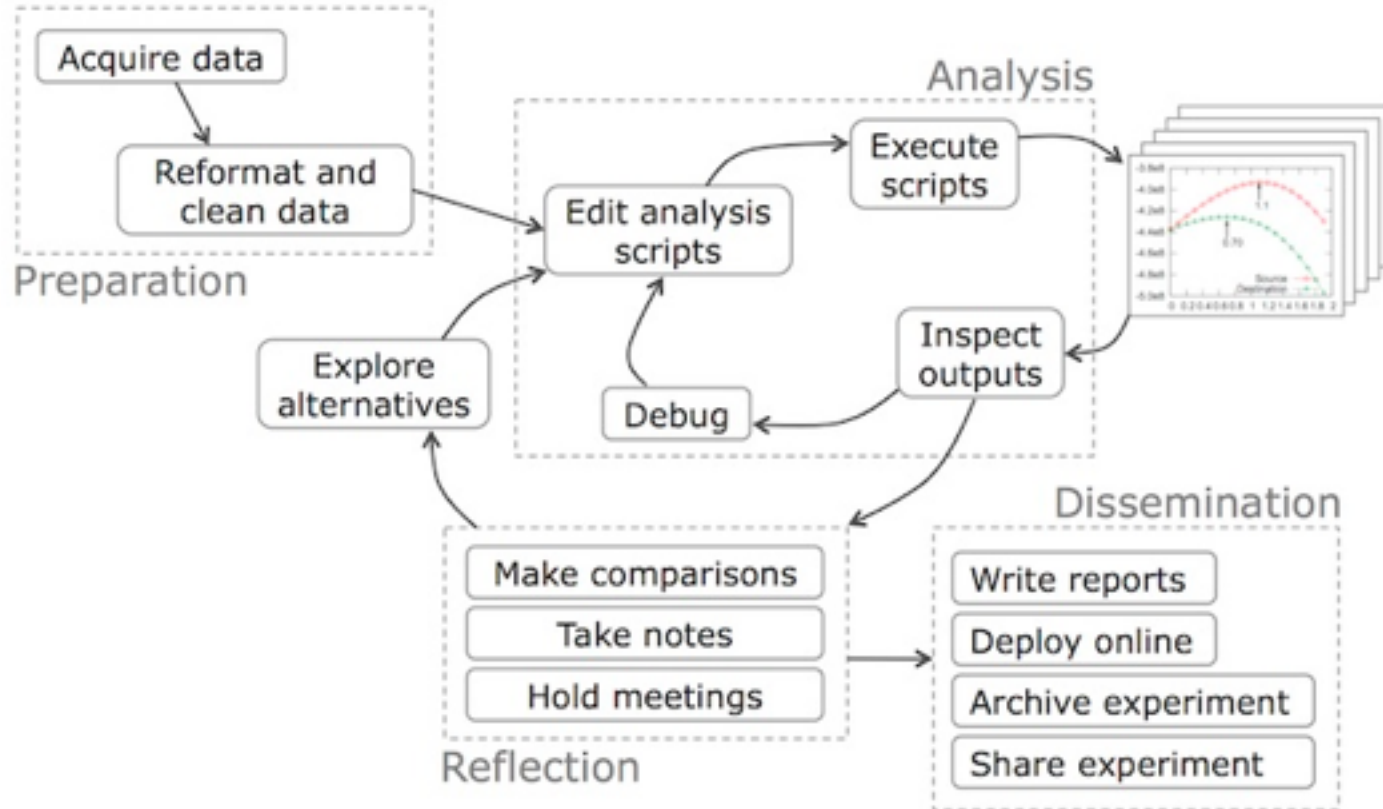


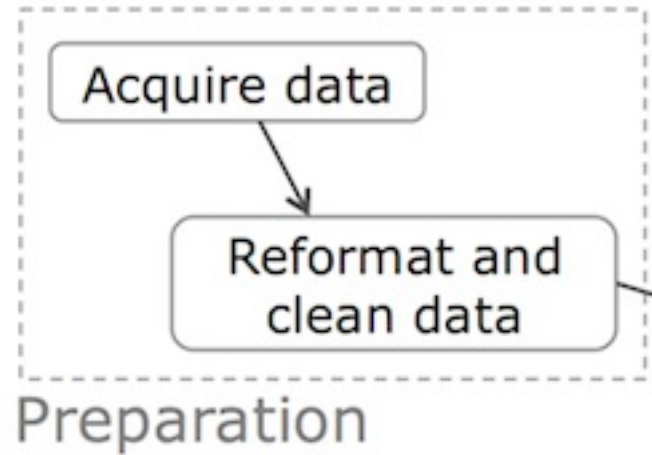
‣ From Jeff Hammerbacher:

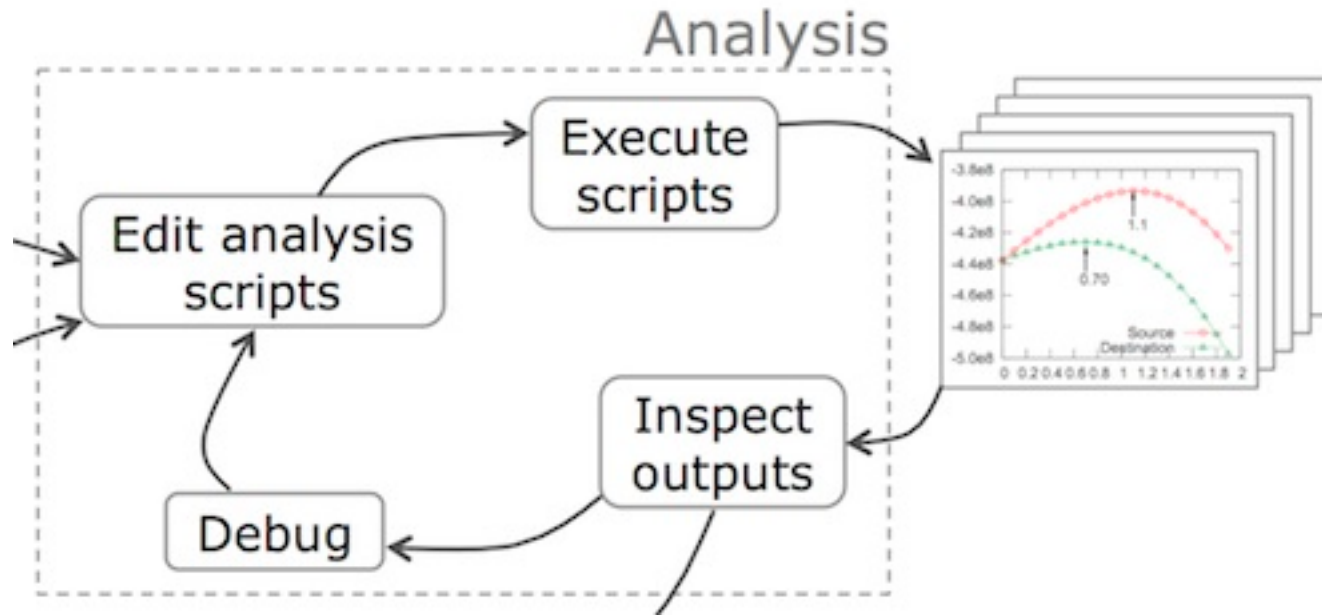
1. Identify problem
2. Instrument data sources
3. Collect data
4. Prepare data (integrate, transform, clean, impute, filter, aggregate)
5. Build model
6. Evaluate model
7. Communicate results

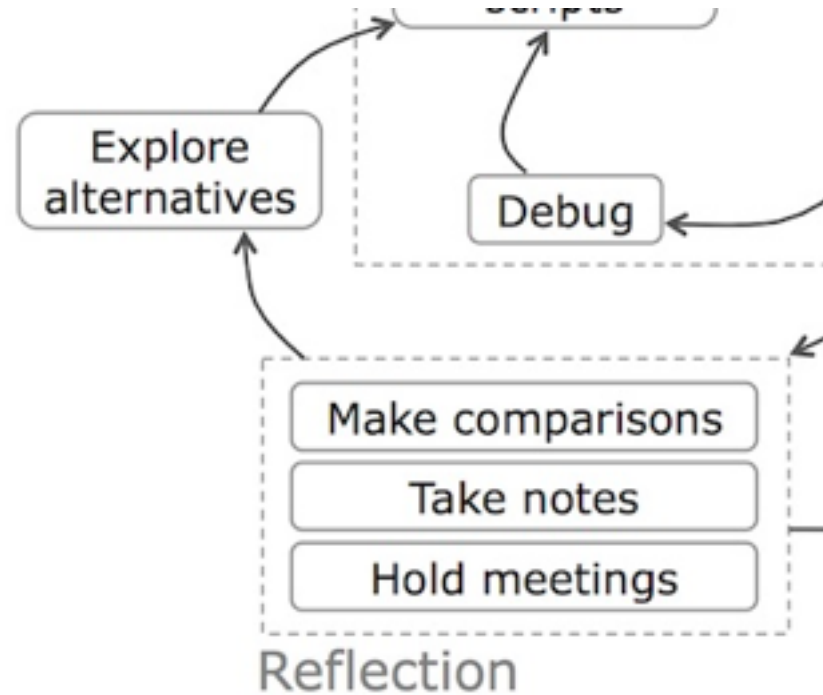
From Dataists Blog

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









- **Build an analytics team:**
 - 1. Define the top priorities of the organization**
 - 2. Determine the data you'd like to collect**
- What will your greatest challenges be?**
- What products could you build?**
- What studies could you run? How would these influence the organization?**

III. WORKING AT THE UNIX COMMAND LINE

KEY OBJECTIVES

- NAVIGATE THE FILESYSTEM
- CREATE, MOVE, COPY, AND DELETE FILES & DIRECTORIES
- VIEW & SEARCH FILES
- EDIT & INTERACT WITH FILES
- COMBINE STEPS
- LEARN MORE

TOOLS

- LS, CD
- CAT, TOUCH, MV, CP, MKDIR, RM, RMDIR
- HEAD, TAIL, LESS, CAT, GREP
- VIM, AWK, SED, TR, SORT, UNIQ, WC
- PIPE (|)
- MAN, APROPOS

NOTE

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

Let's take a look at the ZipRecruiter Data Set

Right click on “Raw” and select “Save Link As ...”

https://github.com/adparker/GADSLA_1403/blob/master/src/lesson01/jobs.tsv

Or get it from the command line:

```
curl -L 'https://github.com/adparker/GADSLA\_1403/raw/master/src/lesson01/jobs.tsv'
```

INTRO TO DATA SCIENCE

DISCUSSION