

INTRO TO DATA SCIENCE

LECTURE 1: DATA EXPLORATION

Arun Ahuja – aahuja11@gmail.com

Aaron Schumacher -

INTRO TO DATA SCIENCE

WELCOME!

I. WHAT IS DATA SCIENCE? II. THE DATA MINING WORKFLOW

EXERCISES:

III. WORKING AT THE UNIX COMMAND LINE IV. VISUALIZING DATA WITH R & GGPLOT2

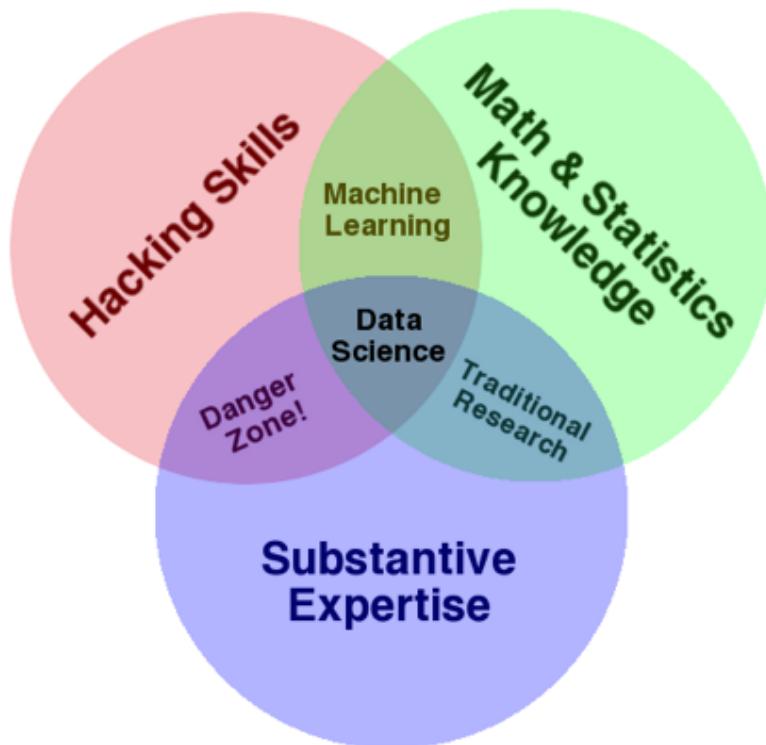
INTRO TO DATA SCIENCE

I. WHAT IS DATA SCIENCE?

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

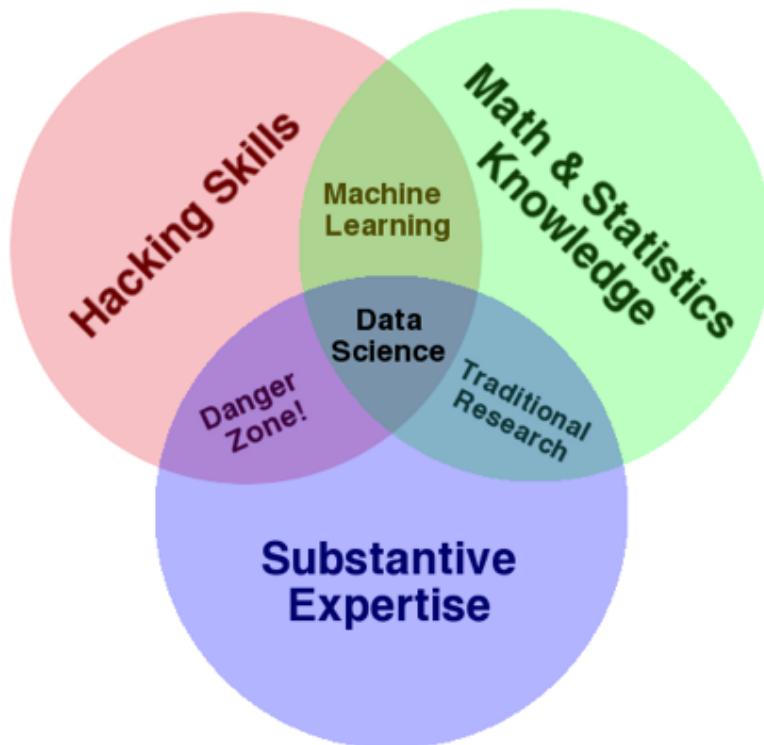
- A set of tools and techniques used to extract useful information from data.
- An interdisciplinary, problem-oriented subject.

THE QUALITIES OF A DATA SCIENTIST



THE QUALITIES OF A DATA SCIENTIST

8



ONE MORE THING!

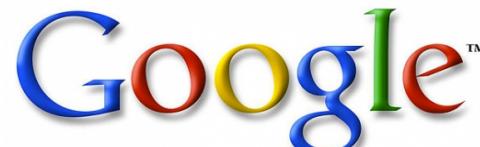
Communication skills

- 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.

WHO USES DATA SCIENCE?

11



WHO USES DATA SCIENCE?

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

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

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

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](#)

Noncredit courses and certificates
Center for Advanced Digital Applications (CADA)

There's Still Time to

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](#)

What Other Customers Are Looking At Right Now

 SanDisk Ultra 64 GB MicroSDXC Class 10 UHS-I card ★★★★★ (4,513) \$69.99 \$42.99	 Samsung Galaxy Tab 2 (7-Inch, Wi-Fi, 8GB) ★★★★★ (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, 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.com \$100.00 Gift Card - E-mail ★★★★★ (19,180) \$50.00
--	--	---	---	--	---	--

Digital Cameras Best Sellers

 Nikon COOLPIX S9200 16 MP CMOS... ★★★★★ (428)	 Canon PowerShot SX500 IS 16.0 MP... ★★★★★ (150)	 Canon PowerShot A2300 IS 16.0 MP... ★★★★★ (429)	 Canon EOS Rebel T3i 18 MP CMOS... ★★★★★ (778)	 Canon PowerShot SX150 IS 14.1 MP... ★★★★★ (281)	 Canon PowerShot SX260 HS 12.1 MP CMOS... ★★★★★ (729)	 Canon EOS Rebel T4i 18.0 MP CMOS... ★★★★★ (118)
--	--	--	--	--	---	--

WHO USES DATA SCIENCE?

Amazon.com - 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
- > Full Store Directory

The Perfect Gift for Dad
kindle fire HD
From \$199.99 From \$179.99
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

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

Father's Day Savings Sponsored by DEWALT >Shop now

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) \$69.99 \$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, 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... 1420	Canon PowerShot SX500 IS 16.0 MP... 1420	Canon PowerShot A2300 IS 16.0 MP... 1420	Canon EOS Rebel T3i 18 MP CMOS... ★★★★★ (778)	Canon PowerShot SX150 IS 14.1 MP... 1420	Canon PowerShot SX260 HS 12.1 MP CMOS... 1420	Canon EOS Rebel T4i 18.0 MP CMOS... 1420

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

» 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

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

amazonPrime

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](#)

amazonPrime

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](#)

amazon.com **\$30 Off Instantly** [Learn more](#)

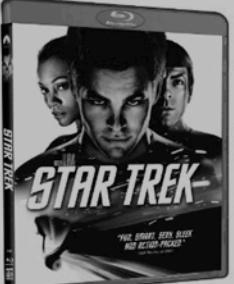
NEW YORK UNIVERSITY Noncredit courses and certificates Center for Advanced Digital Applications (CADA) There's Still Time to...

What Other Customers Are Looking At Right Now

						
SanDisk Ultra 64 GB MicroSDXC Class... ★☆☆☆☆ (4,513) \$99.00 \$42.99	Samsung Galaxy Tab 2 (7-Inch, Wi-Fi) 8... ★☆☆☆☆ (2,401) \$249.00 \$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, 8... 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... 14x Optical Zoom (420)	Canon PowerShot SX500 IS 16.0 MP... 16x Optical Zoom (420)	Canon PowerShot A2300 IS 16.0 MP... 16x Optical Zoom (420)	Canon EOS Rebel T3i 18 MP CMOS... 18x Optical Zoom (778)	Canon PowerShot SX150 IS 14.1 MP... 14x Optical Zoom (250)	Canon PowerShot SX260 HS 12.1 MP CMOS... 12x Optical Zoom (720)	Canon EOS Rebel T4i 18.0 MP CMOS... 18x Optical Zoom (418)



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
		\$9.78	\$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](#)

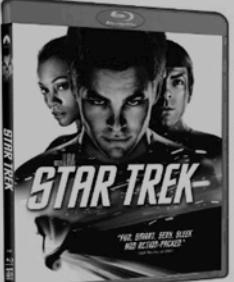
More Buying Choices

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](#)

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?

17



INTRO TO DATA SCIENCE

II. THE DATA SCIENCE WORKFLOW

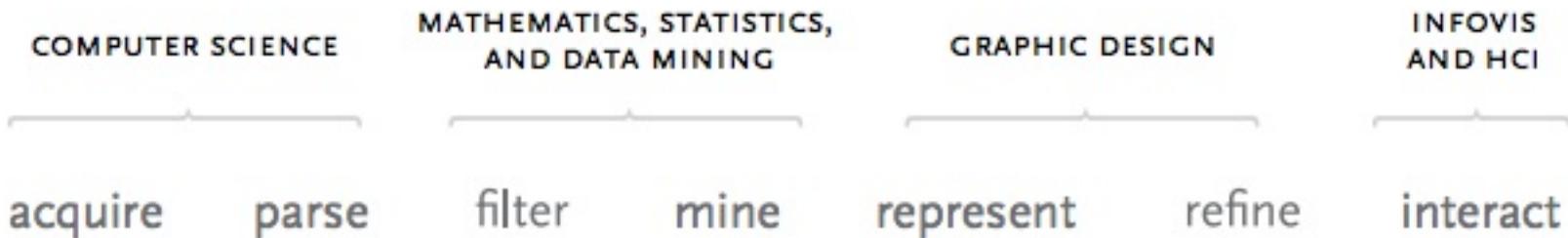
THE DATA SCIENCE WORKFLOW

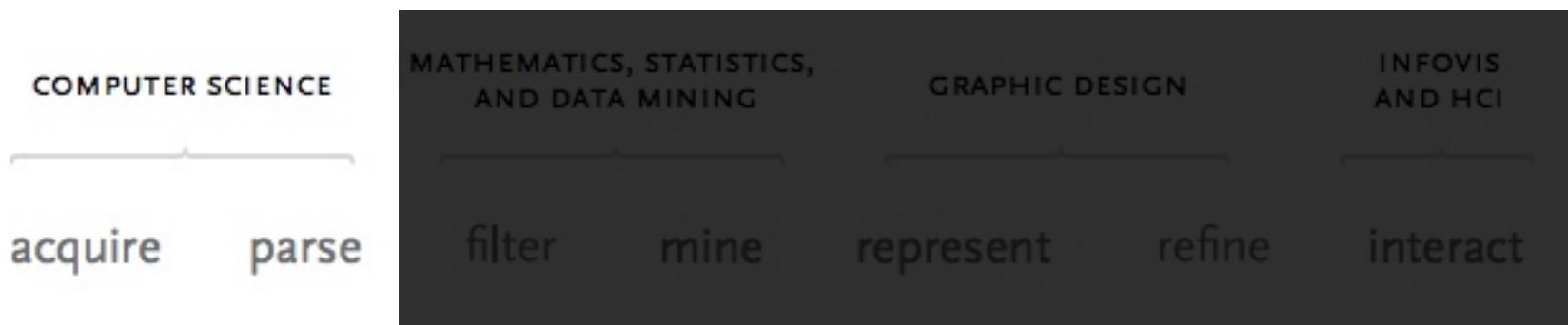
19

acquire parse filter mine represent refine interact

THE DATA SCIENCE WORKFLOW

20





THE DATA SCIENCE WORKFLOW

22



THE DATA SCIENCE WORKFLOW

23



THE DATA SCIENCE WORKFLOW

24



THE DATA SCIENCE WORKFLOW

25



III. WORKING AT THE UNIX COMMAND LINE

Download this dataset:
(insert link to pace.txt)

Let's take a look at the 538 dataset

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!

IV. VISUALIZING DATA WITH R AND GGPLOT2

KEY OBJECTIVES

- Become familiar with the R environment
- Explore data in R
- Visualize data using ggplot2
- Mathematical bonus: power laws

INTRO TO DATA SCIENCE

DISCUSSION