

# INFO 201

---

Technical Foundations of Informatics

Joel Ross  
Winter 2017

# Hi, I'm Joel!

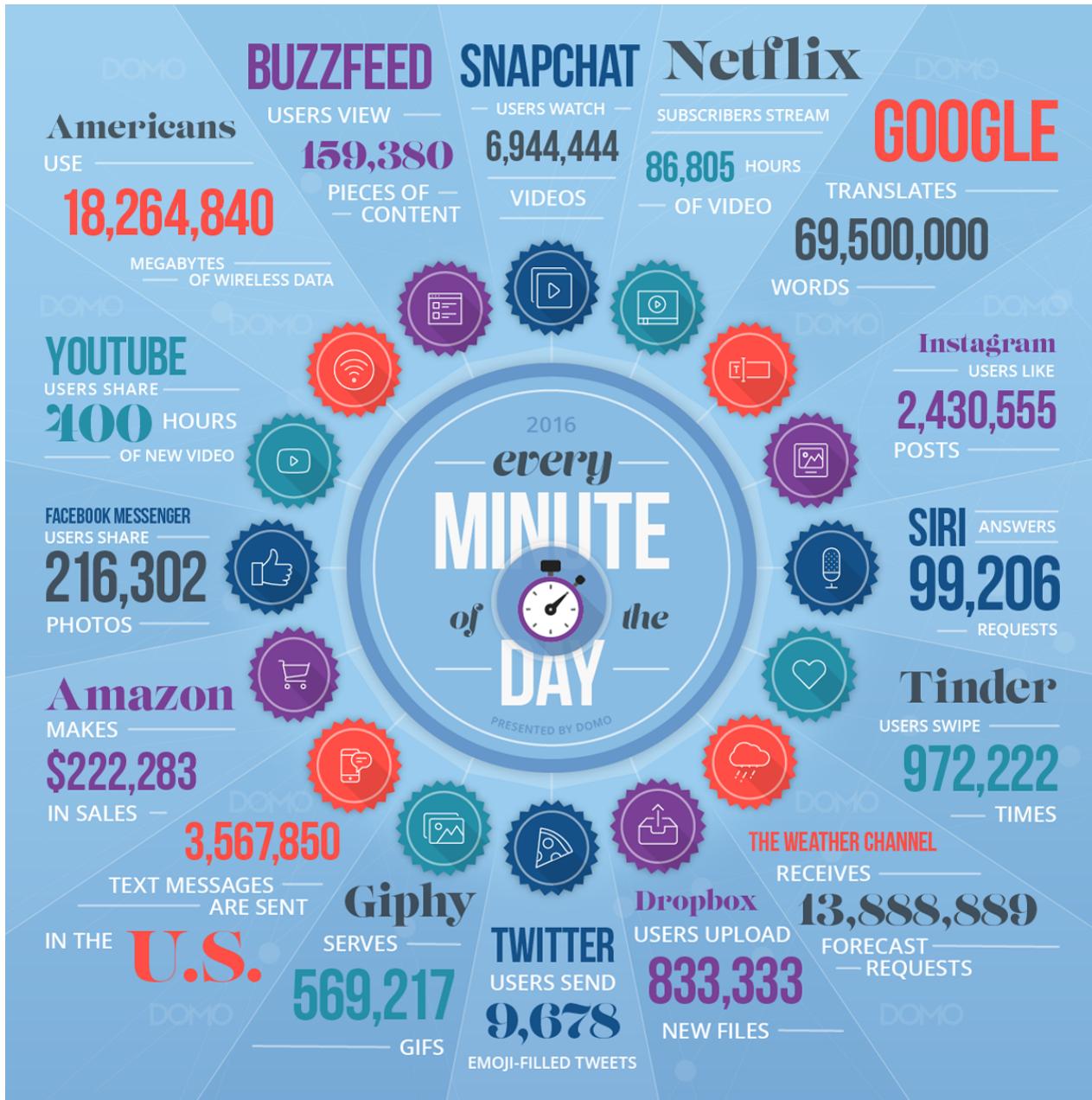
# Who are you?

# Meet someone new!

- Who are you?
- Where did you grow up?
- Re: this course
  - What are you most excited about?
  - What is your greatest fear?

# What are we doing here?

By the time I get to  
the next slide, lots of  
data will have been  
produced.



**That's a lot of data**

**some of which is interesting...**

# How Computer Scientists Are Using Twitter to Predict Gentrification

Cambridge researchers have created a way to predict a neighborhood's fortunes in coming years by analyzing social media data



(Steven Greaves/Corbis)

By [Emily Matchar](#)

SMITHSONIAN.COM

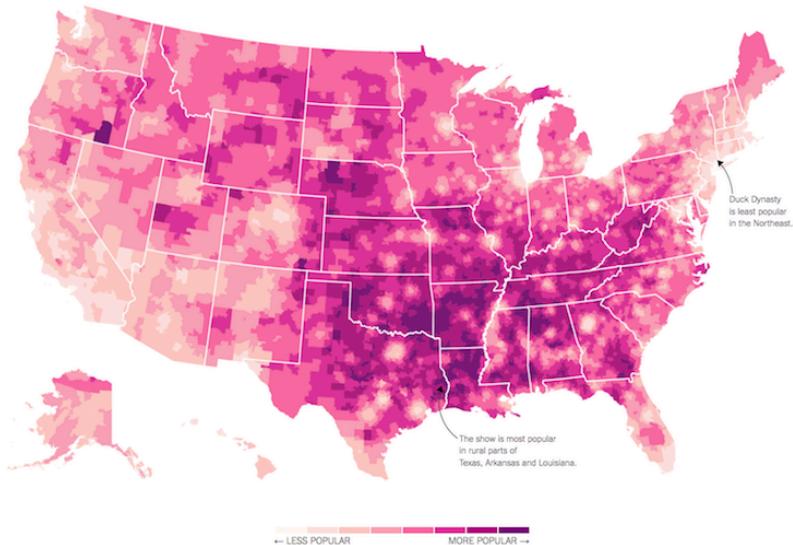
APRIL 27, 2016

Smithsonian

---

1. Duck Dynasty

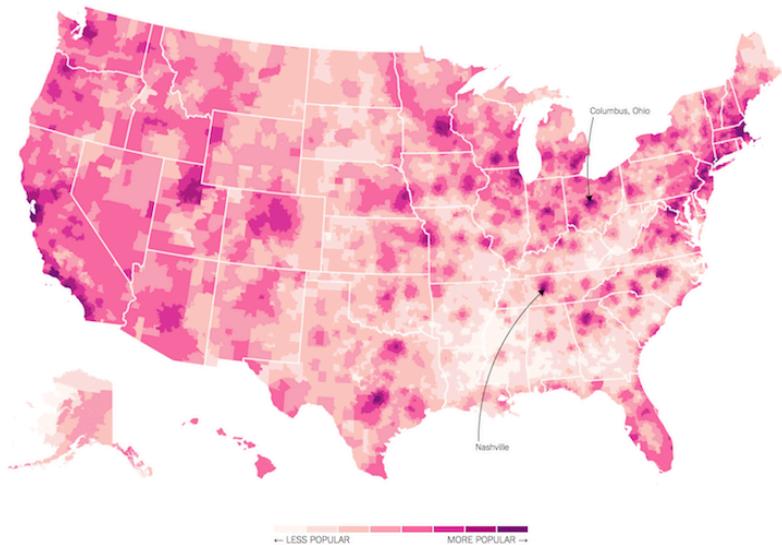
---



---

7. Modern Family

---



## 'Duck Dynasty' vs. 'Modern Family': 50 Maps of the U.S. Cultural Divide

By JOSH KATZ DEC. 27, 2016

nytimes

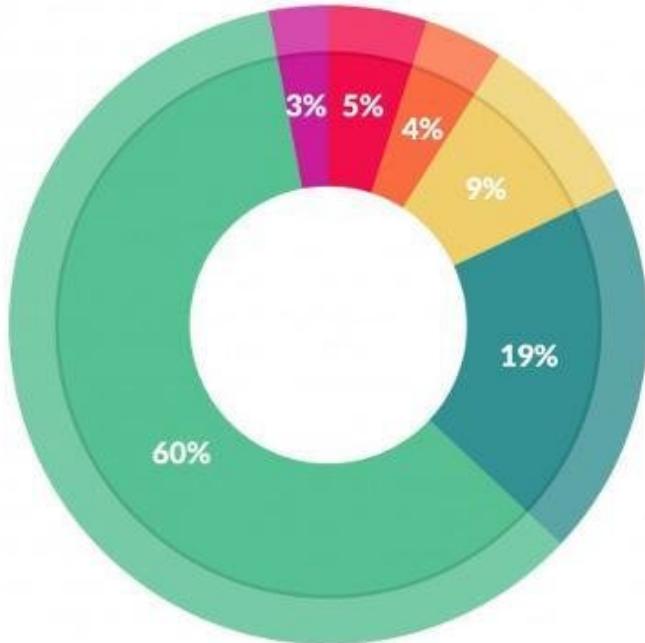
We need to convert

**data** ← raw numbers

into

**information**

↑  
interpreted data



### What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60% (highlighted)
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

NEISS-data-2014-updated-12MAY...												Search Sheet			
Home	Insert	Page Layout	Formulas	Data	Review	View									
I1	▲	X	✓	fx	product_id										

1	CPSC Case #	trmt_date	weight	age	sex	race	diagnosis	body_part	product_id	narrative			
265993	141132063	11/11/14	74.3851	18	Male	White		55	30	1205	18YO M, HX MULTIPLE SHOULDER DISL		
265994	141043714	9/21/14	15.6716	73	Male	White		50	92	841	73 YOM SLIPPED WHILE USING A TABLI		
265995	141220171	11/28/14	74.3851	90	Female	Other / Mixed Race		57	89	1807	90 YO F PT FOUND ON FLOOR AT HOM		
265996	141154621	11/22/14	15.6716	9	Female	Other / Mixed Race		64	92	1329	9 YOF REPORTS INJURING HER RT THUI		
265997	141028361	9/14/14	15.6716	28	Female	Black/African American		64	83	1466	28 YOF JUMPED OFF OF A FOUNTAIN. I		
265998	140137370	1/12/14	74.3851	7	Female	White		57	30	5040	7 YO F RIDING HER BICYCLE AND FELL C		
265999	141108521	10/30/14	15.6716	19	Male	White		64	34	1884	19 YOM GOT INTO AN ARGUMENT WIT		
266000	141043712	9/21/14	15.6716	29	Female	White		55	35	3278	29 YOF HAD KNEE PX WHILE DANCING		
266001	141146018	11/17/14	15.6716	48	Female	None listed		57	31	1807	48 YOF FELT LIGHTHEADED WHILE SITT		
266002	140636029	6/11/14	74.3851	2	Male	Other / Mixed Race		62	75	4074	2 YO M PT WAS SITTING IN A CHAIR LE.		
266003	140649577	6/16/14	14.3089	69	Female	White		66	76	1659	69 YOF REPORTS A NOSEBLEED THAT C		
266004	140804897	7/29/14	15.6716	11	Male	Other / Mixed Race		59	82	1871	11 YOM SUS A LACERATION OF THE LEI		
266005	140960904	9/16/14	74.3851	24	Female	Black/African American		62	75	1205	24YOF PAIN TO HEAD S/P FALL WHILE I		
266006	140219485	2/4/14	81.576	18	Male	Other / Mixed Race		64	37	1267	18YOM TWISTED ANKLE PLAYING SOCK		

What information might we convert this data into?

266015	140555267	5/22/14	74.3851	67	Female	White	55	70	1807	67 YOF WAS WALKING IN CHURCH AND	
266016	140510635	5/1/14	74.3851	62	Male	White	62	75	4076	62YOM PAIN TO HEAD S/P ROLLED OFF	
266017	140950682	9/9/14	74.3851	67	Female	Black/African American	59	92	464	67YOF LAC TO L 3RD FINGER WHEN US	
266018	141019429	9/30/14	74.3851	15	Female	Other / Mixed Race	64	35	1200	15YOF PAIN TO R KNFF WHFN IN GYM	

◀	▶	nss14	+
Ready	□	Average: 2112.765283	Count: 101502 Sum: 214447789

Source



Should we do this  
by hand?

Slow

Tedious

Error-prone

How to *scale*?

How to *reproduce*?

# Solution: get someone else to do the boring work for us!



# Our new assistant



We will learn the  
*technical* skills and  
tools necessary for  
working with  
information.

We need to tell the  
computer what to do!



## Problem

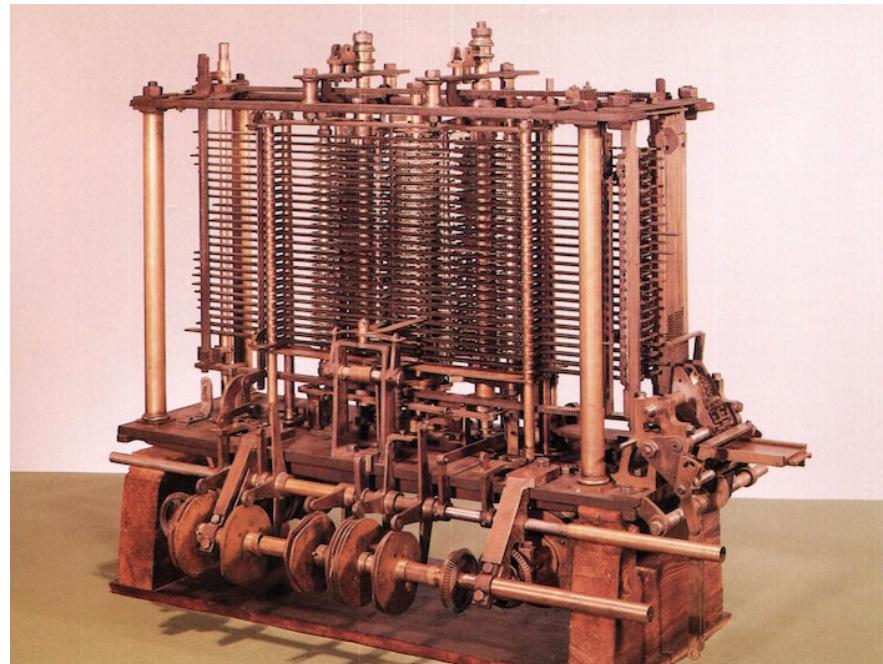
Computers don't  
speak English!



# Programming

Writing instructions for a computer in a language it understands

# First Programmer: Ada Lovelace (1815-1852)



Charles Babbage's **Analytical Engine**  
(designed 1837; never built)

# Binary

0101010111011001010101000101	LOAD Contents of A
01010111011101100101010100010101	LOAD Contents of B
01010101110110010101010001010101	LOAD Contents of C
01110111011001010101000101010101	MULTIPLY C and 4 and STORE in TMP
11011101100101010100010101010111	SUBTRACT TMP from B and STORE in TMP
01110110010101010001010101011101	ADD A to TEMP and STORE in TEMP
11011001010101000101010101110101	STORE TEMP in D

Example from David Chiu. Disclaimer: Fake example!

# Interpreter

```
# Write this instead #
LOAD A
LOAD B
LOAD C
MULTIPLY C and 4 and STORE in TMP
SUBTRACT TMP from B and STORE in TMP
ADD A to TEMP and STORE in TEMP
STORE TEMP in D
```

Have a computer program do a  
find-and-replace to change  
**LOAD** to **010101**, etc.

# Abstraction

0101010111011011001010101000101	LOAD Contents of A
01010111011101100101010100010101	LOAD Contents of B
01010101110110010101010001010101	LOAD Contents of C
01110111011001010101000101010101	MULTIPLY C and 4 and STORE in TMP
11011101100101010100010101010111	SUBTRACT TMP from B and STORE in TMP
01110110010101010001010101011101	ADD A to TEMP and STORE in TEMP
11011001010101000101010101110101	STORE TEMP in D

$$D \leftarrow A + (B - 4*C)$$

# Abstraction

The process of *generalization*;  
of working with higher-level  
representations rather than  
specific details

# Programming Language

A language that a human  
can write, and which can  
be interpreted by a  
computer

# Writing Code

```
# This is the programming language R
# It means something reasonable, I swear!

hypot <- function(a, b, c) {
  det <- sqrt(b*b - 4*a*c)
  x1 <- (-b + det)/(2*a)
  x2 <- (-b - det)/(2*a)
  return( c(x1,x2) )
}

x <- hypot(1,5,-14)
print(x)

# Any idea what this program does?
```

# Managing Code

The screenshot shows a GitHub repository page for `tidyverse / ggplot2`. The top navigation bar includes links for `This repository`, `Search`, `Pull requests`, `Issues`, and `Gist`. On the right side of the header are buttons for `Watch` (248), `Star` (2,175), `Fork` (813), and a user profile icon.

The main content area displays the repository's activity. It shows a commit from `hadley` with the message `Prepare for patch release.` followed by three dots (`...`). To the right of this message is the text `Latest commit bd7d9d6 3 days ago`.

A list of recent commits follows:

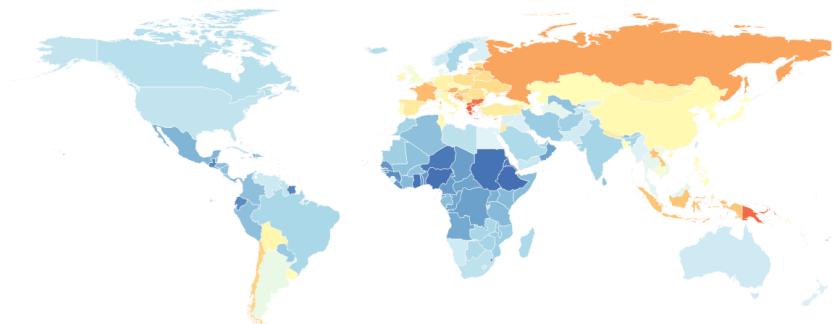
Commit	Message	Time Ago
<code>aaa.r</code>	Document and export Scale objects	a year ago
<code>aes-calculated.r</code>	Corrected style issue	2 years ago
<code>aes-colour-fill-alpha.r</code>	Alpha affects both fill and colour.	a year ago
<code>aes-group-order.r</code>	Final doc tweaks	2 months ago
<code>aes-linetype-size-shape.r</code>	Don't attach packages in examples.	a year ago
<code>aes-position.r</code>	Re-export arrow and unit from grid.	a year ago
<code>aes.r</code>	Doc tweaks for programming/extending	2 months ago
<code>annotation-custom.r</code>	Doc tweaks for annotations + general reorg	2 months ago
<code>annotation-logticks.r</code>	Add dummy data to layer to fix #1655	5 months ago
<code>annotation-map.r</code>	Doc tweaks for annotations + general reorg	2 months ago
<code>annotation-raster.r</code>	Doc tweaks for annotations + general reorg	2 months ago

# Turn Data into Information!

	A	B	C	D	E
1	country	region	income_group	le_1960	le_2013
2	Aruba	Latin America & Caribbean	High income: nonOECD	65.56936585	75.33217073
3	Afghanistan	South Asia	Low income	31.58004878	60.93141463
4	Angola	Sub-Saharan Africa	Upper middle income	32.98482927	51.86617073
5	Albania	Europe & Central Asia	Upper middle income	62.25436585	77.5372439
6	United Arab Emirates	Middle East & North Africa	High income: nonOECD	52.24321951	77.13129268
7	Argentina	Latin America & Caribbean	High income: nonOECD	65.21553659	76.18729268
8	Armenia	Europe & Central Asia	Lower middle income	65.86346341	74.5407561
9	Antigua and Barbuda	Latin America & Caribbean	High income: nonOECD	61.78273171	75.82929268
10	Australia	East Asia & Pacific	High income: OECD	70.81707317	82.19756098
11	Austria	Europe & Central Asia	High income: OECD	68.58560976	80.8902439
12	Azerbaijan	Europe & Central Asia	Upper middle income	60.8362439	70.69314634
13	Burundi	Sub-Saharan Africa	Low income	41.23604878	54.09719512
14	Belgium	Europe & Central Asia	High income: OECD	69.70195122	80.38536585
15	Benin	Sub-Saharan Africa	Low income	37.27826829	59.28756098
16	Burkina Faso	Sub-Saharan Africa	Low income	34.47790244	56.27502439
17	Bangladesh	South Asia	Lower middle income	47.02456098	70.69339024
18	Bulgaria	Europe & Central Asia	Upper middle income	69.24756098	74.46585366
19	Bahrain	Middle East & North Africa	High income: nonOECD	57.09036595	76.66097905

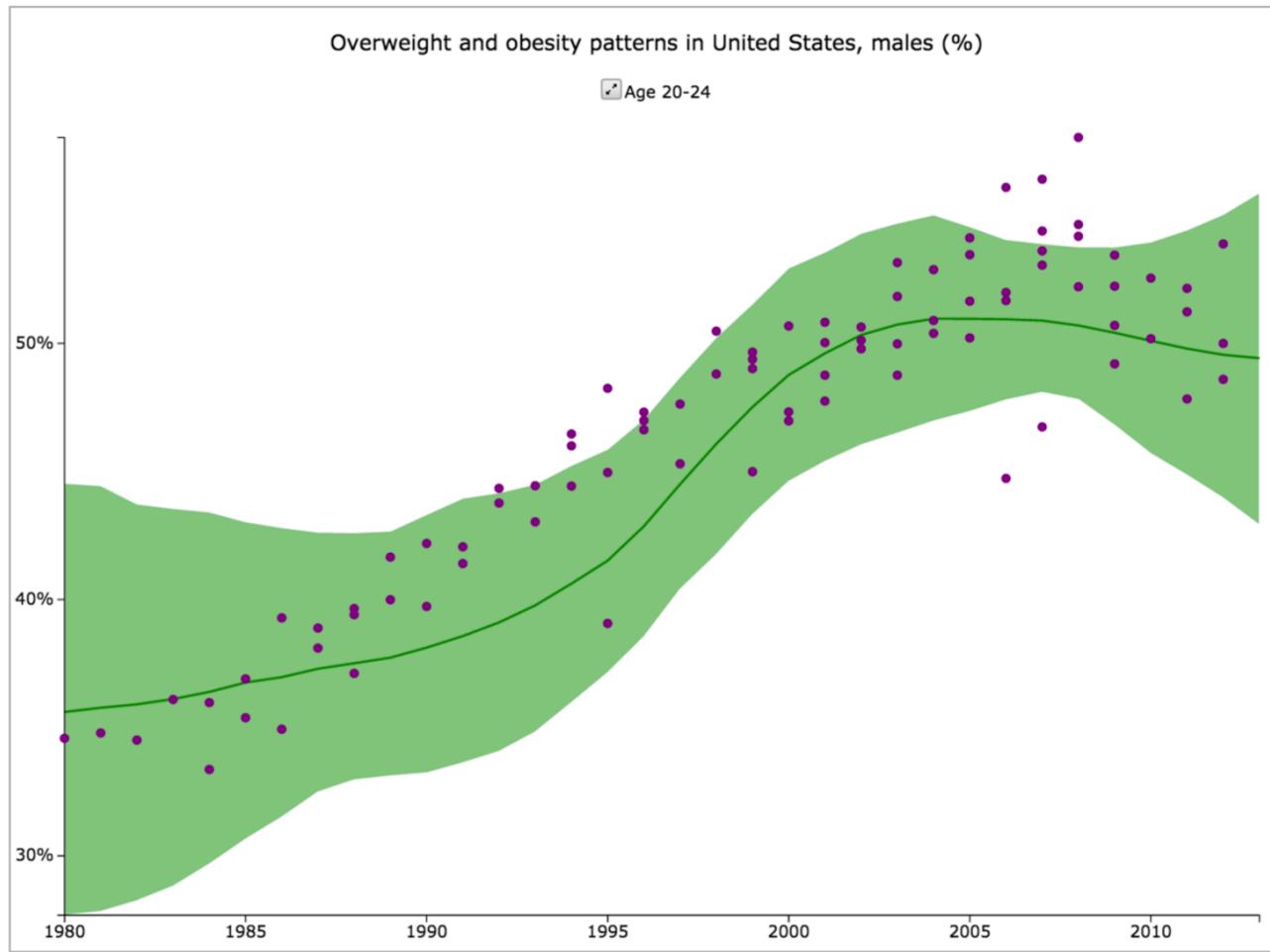
you'll write  
code to do this!

Largest Changes in Life Expectancy
Maldives (+42)
Bhutan (+35)

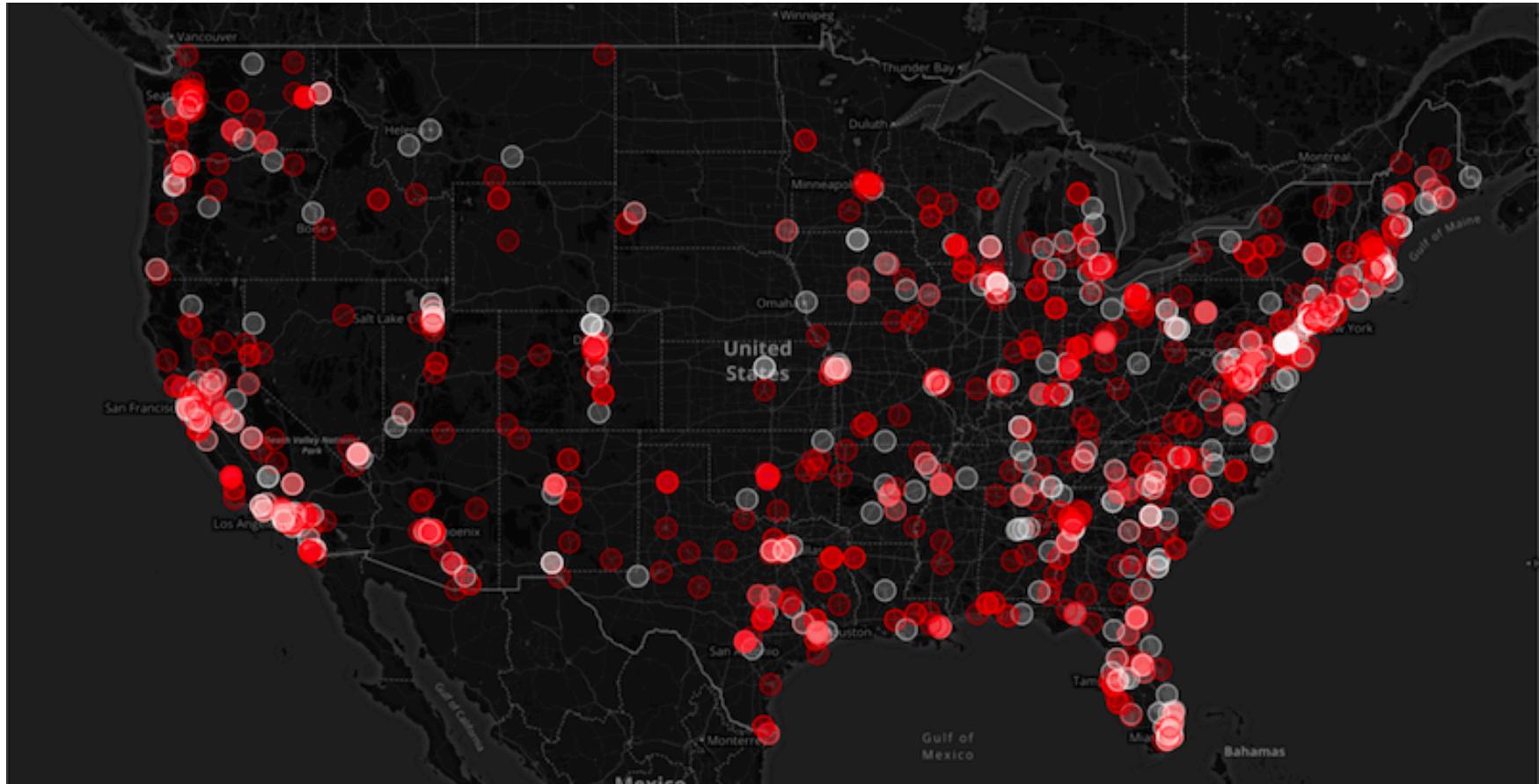


# Why convert data to information?

# Discover Patterns

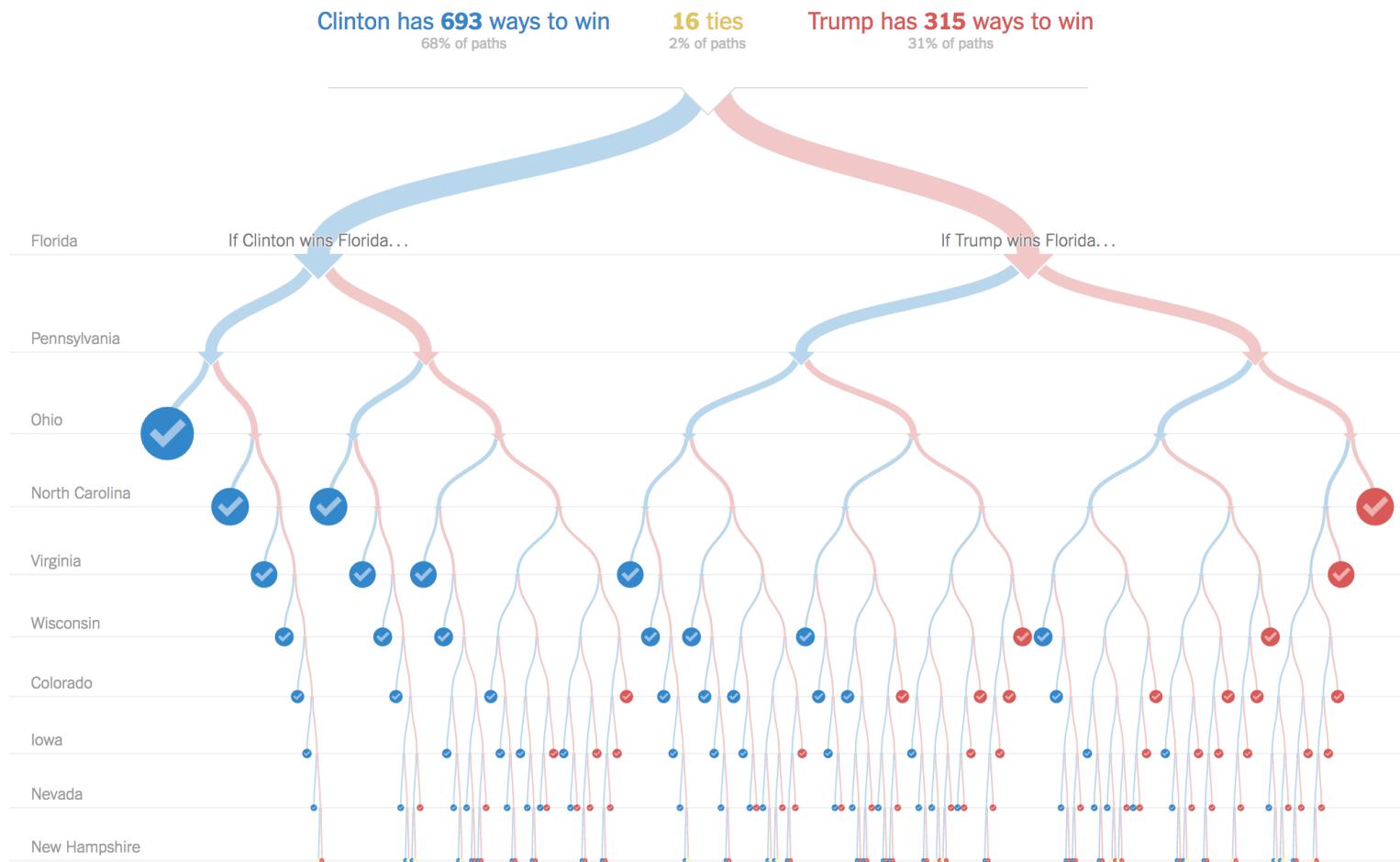


# Communicate Insights



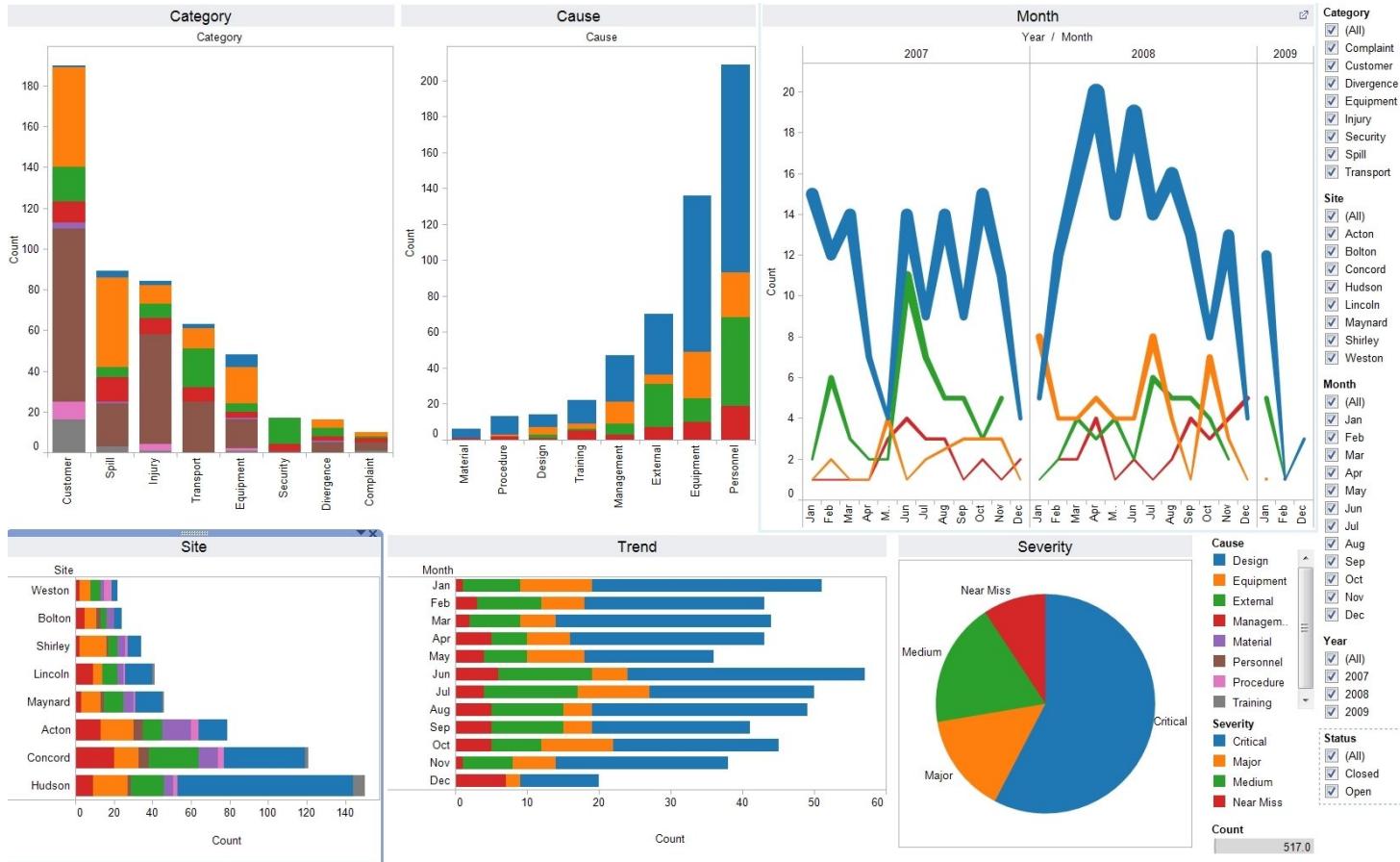
alexbbt

# Make Decisions



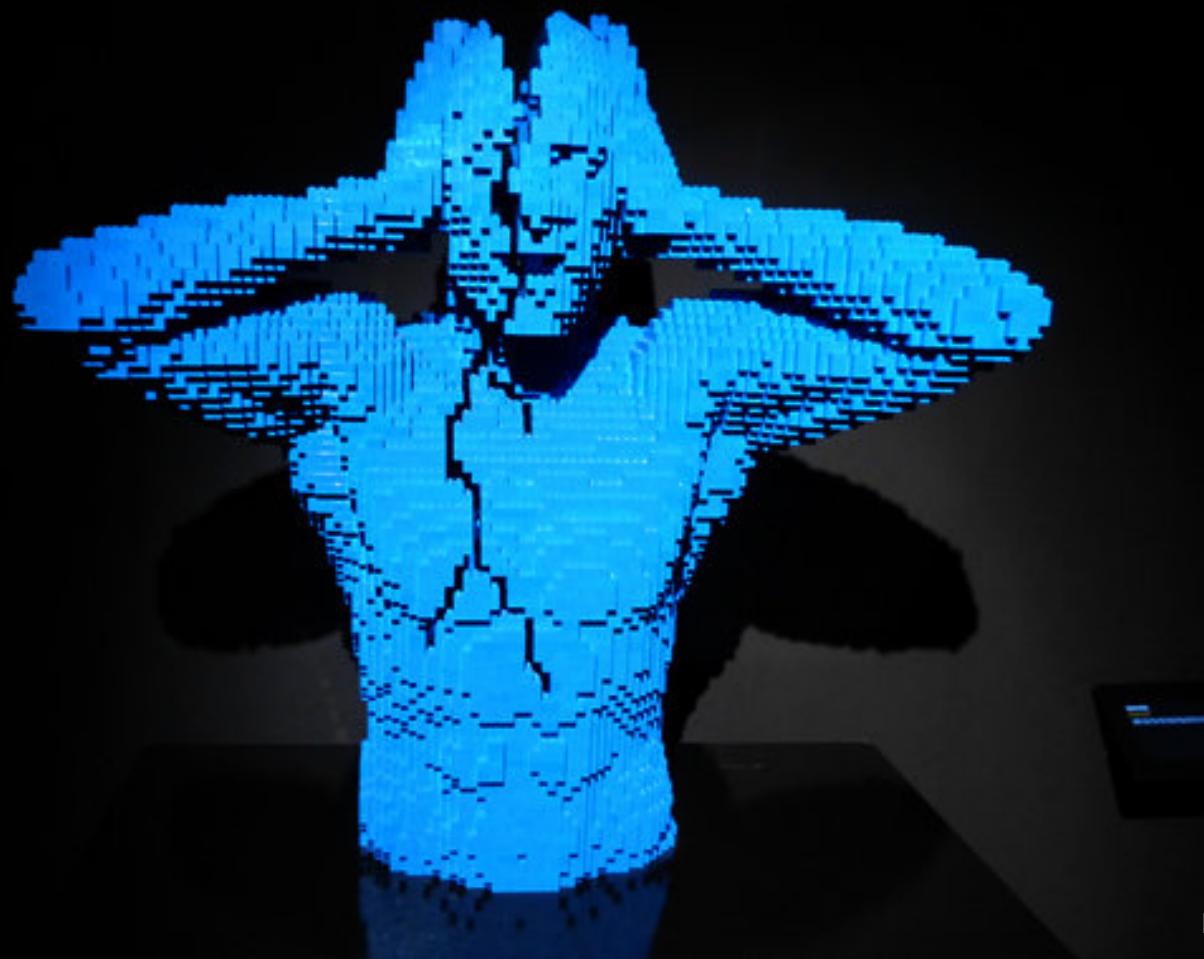
**Why is it a good idea  
to write computer  
programs to work  
with data?**

# There is no shortage of point-and-click tools to work with data...



Why is it **NOT** a good  
idea to write computer  
programs to work with  
data?

# BREAK



Nathan Sawaya

**[WARNING]**

Programming involves a lot  
of failure and frustration

# Introductions



Information School  
UNIVERSITY *of* WASHINGTON

**WE MAKE INFORMATION WORK**

How *people* use *technology* to  
manage *information*

# About Joel

## Senior Lecturer



**Email** joelross@uw.edu

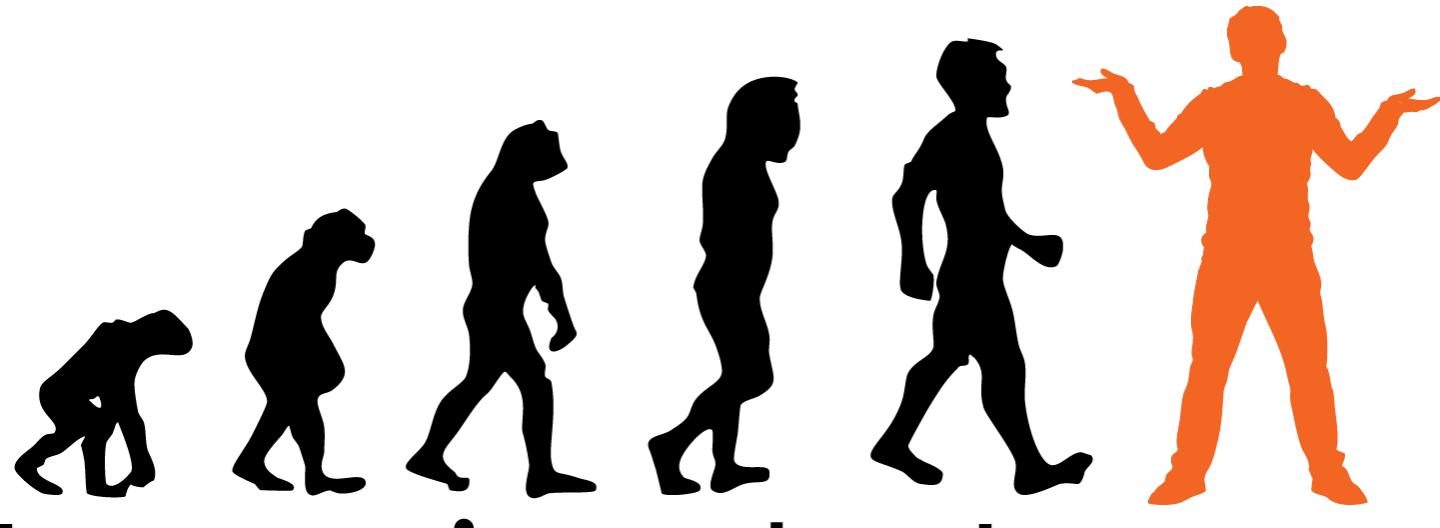
**Office** MGH 330C



# **Meet your TAs!**

# Course Resources

**Canvas has the syllabus and assignment details.**



# Learning to Learn

# Learning Modules

# Course Resources

**Canvas** has the syllabus and assignment details.

**GitHub** has all the code (including assignments and learning modules)

**Slack** is how you will ask questions, collaborate, and see announcements.

**Any questions  
so far?**

# Command Line

# Graphical User Interface (GUI)

- Windows, Icons, Menus, Pointers (e.g., WIMP)
- Easy to learn
- Not much more efficient for experts

# Command Line Interface

- Text-based interface
- Hard to learn
- Very efficient for experts
- Like programming!

# Command Shell



(command + space, search for "Terminal")

A screenshot of a macOS terminal window. The window title is "lecture — bash — 59x12". The title bar also displays "bash". The main pane shows the command "ls" output:

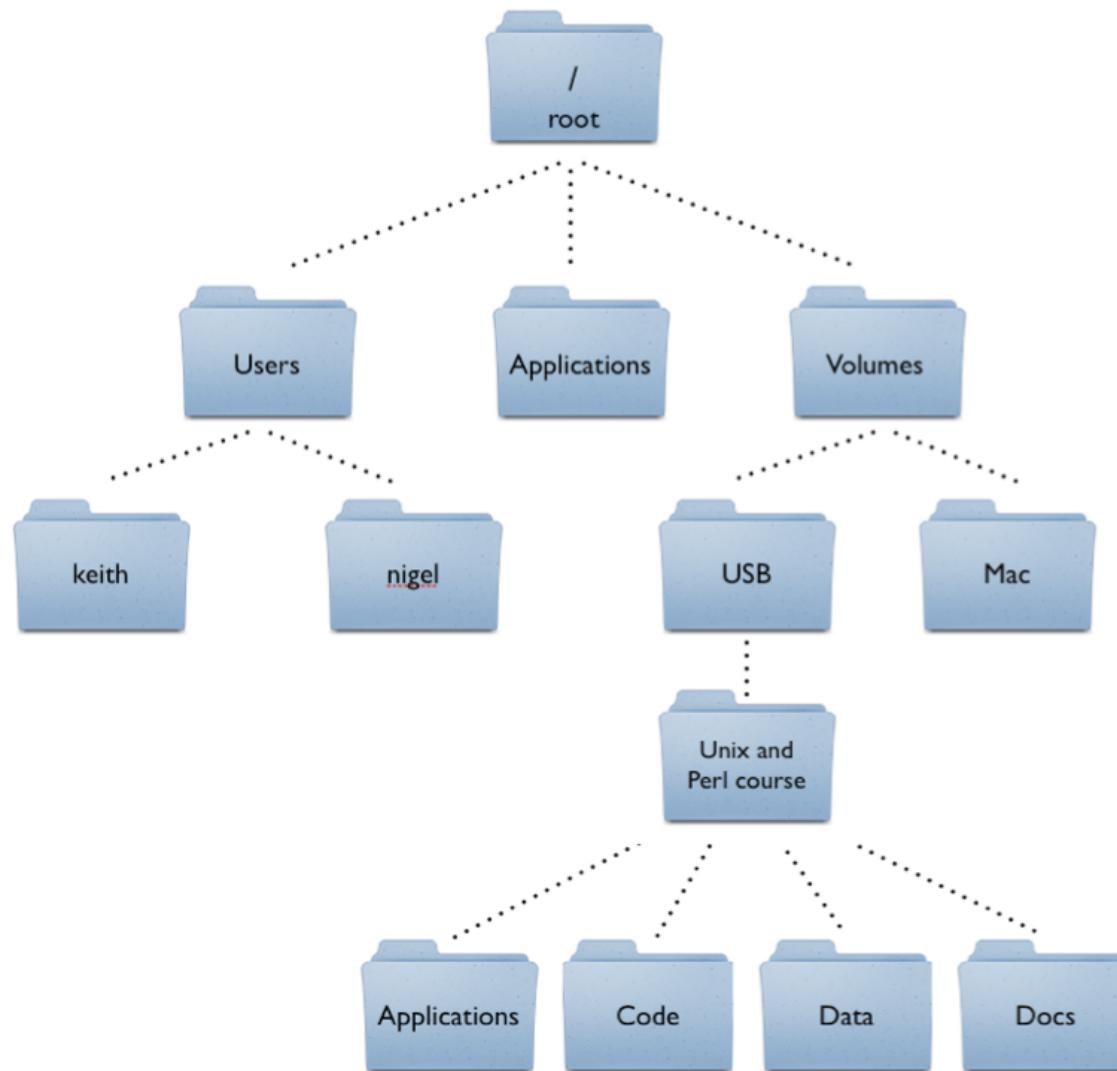
```
is-joelrossm13x:lecture joelross$ ls  
machine directory user prompt
```

# Where are we?

```
pwd
```

(print working directory)

# Directory Tree



# Change Location

```
cd folder
```

command

argument  
(required)

(change directory)

# What's here?

```
ls [ folder ]
```



optional  
argument

(list contents)

# Paths

/absolute/path/to/file

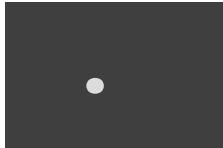
leading slash  
↑

How to get there *starting from the root*

relative/path/to/file

How to get there *starting from here*

# Path Symbols



(a period):  
the *current* directory



(two periods):  
the *parent* directory

# Path Practice

```
# if I start here:  
$ pwd  
/Users/iguest/Desktop  
  
# and then do this:  
$ cd ../../Desktop  
  
#where do I end up?
```

**ALWAYS USE  
RELATIVE PATHS  
IN CODE!**

# More File Commands

`mkdir`

(make directory)

`rm`

(remove file or folder)

`cp`

(copy file or folder)

`open`

`start`

(Mac: open a file or folder )

(Windows: open a file or folder )

# Check the Manual

```
man command
```

(manual)

joelross — less • man mkdir — 95x33  
~ — less • man mkdir

MKDIR(1)   BSD General Commands Manual                           MKDIR(1)

**NAME**    **options (flags)**

**mkdir** -- make directories

**SYNOPSIS**

**mkdir** [-pv] [-m mode] directory name ...

**DESCRIPTION**

The **mkdir** utility creates the directories named as operands, in the order specified, using mode **rwxrwxrwx** (0777) as modified by the current umask(2).

The options are as follows:

**-m mode**

    Set the file permission bits of the final created directory to the specified mode. The mode argument can be in any of the formats specified to the **chmod(1)** command. If a symbolic mode is specified, the operation characters ``+'' and ``-'' are interpreted relative to an initial mode of ``a=rwx''.

**-p**

    Create intermediate directories as required. If this option is not specified, the full path prefix of each operand must already exist. On the other hand, with this option specified, no error will be reported if a directory given as an operand already exists. Intermediate directories are created with permission bits of **rwxrwxrwx** (0777) as modified by the current umask, plus write and search permission for the owner.

**-v**

    Be verbose when creating directories, listing them as they are created.

The user must have write permission in the parent directory.

: ← Use up/down arrows to scroll. Type **q** to quit.

# Display Text

```
echo "message"
```

**(echo text back)**

When in doubt:

**ctrl-c (control and c)**

to cancel!

## Module 2 exercise-1

# Command Reference

Action	Syntax
Copy a file	<code>cp OLD_FILE NEW_FILE</code>
Move a file	<code>mv OLD_FILE NEW_FILE</code>
Delete a file (careful!)	<code>rm FILE</code>
Open a file	<code>open FILE</code> [Mac]; <code>start FILE</code> [Windows]
View file contents on command-line	<code>less FILE</code>
Output file contents	<code>cat FILE</code>
See previous commands executed	<code>history</code>

# What we did...

- Learned some command-line utilities
- Saved file versions with git
- Push and pulled data to GitHub

# Action Items!

- Attend lab this week!
- Be comfortable with **modules 0-3** by Thurs
  - Install software per **module 1**
- Recommendation: don't start until Assignment 1 until Thursday!

Lab: Setup and Markdown

Thursday: git and GitHub