A Dataset Is Worth >1000 Words

Whole Hog Series, Beyond Text Session Lightning Talk: January 14, 2015

Eleni Castro, Research Coordinator
Harvard Institute For Quantitative Social Science (IQSS)
Harvard Dataverse - dataverse.harvard.edu
Contact: ecastro@g.harvard.edu

Diane Sredl, Data Reference Librarian
Harvard Library, Lamont Library
Contact: govdocs@fas.harvard.edu

IQSS
The Institute for Quantitative Social Science

The Dataverse Project

HARVARD LIBRARY

What is research data?

What stories can data tell us?

Help Inform Policy

New Bike Crash Map Offers Comprehensive Look at Where and How Accidents Happen

The 'cutting edge' project breaks down the time, day, and even weather conditions of each incident.

By Steve Annear | Boston Daily | June 18, 2014, 10:30 a.m.

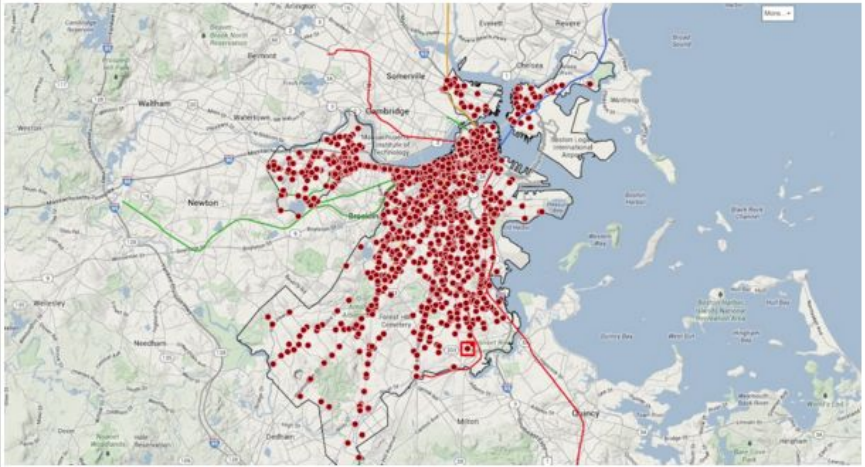
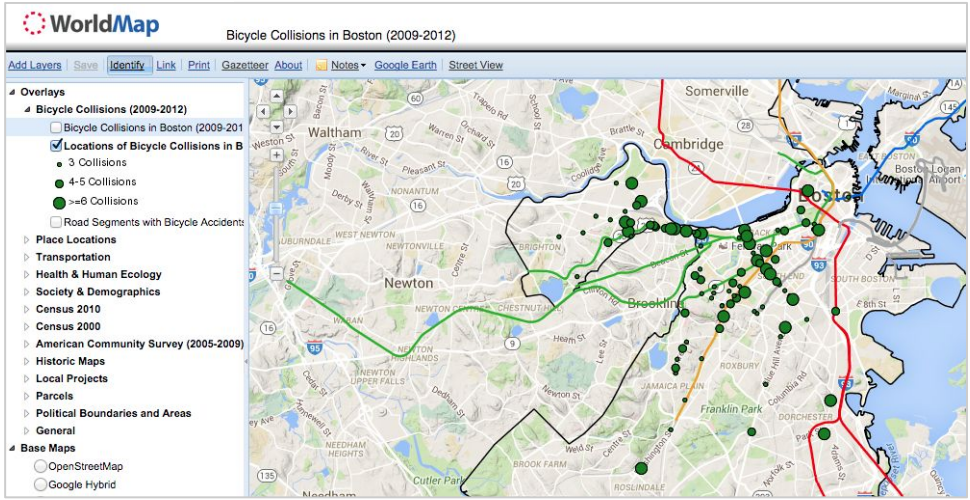


IMAGE VIA BOSTON AREA RESEARCH INITIATIVE

Source: http://www.bostonmagazine.com/news/blog/2014/06/18/bike-crash-map-boston-data

Boston Area Research Initiative

Center for Geographic Analysis
Harvard University



Source: http://worldmap.harvard.edu/maps/boston-bikes

Or Misinform Policy

FAQ: Reinhart, Rogoff, and the Excel Error That Changed History

By Peter Coy | April 18, 2013

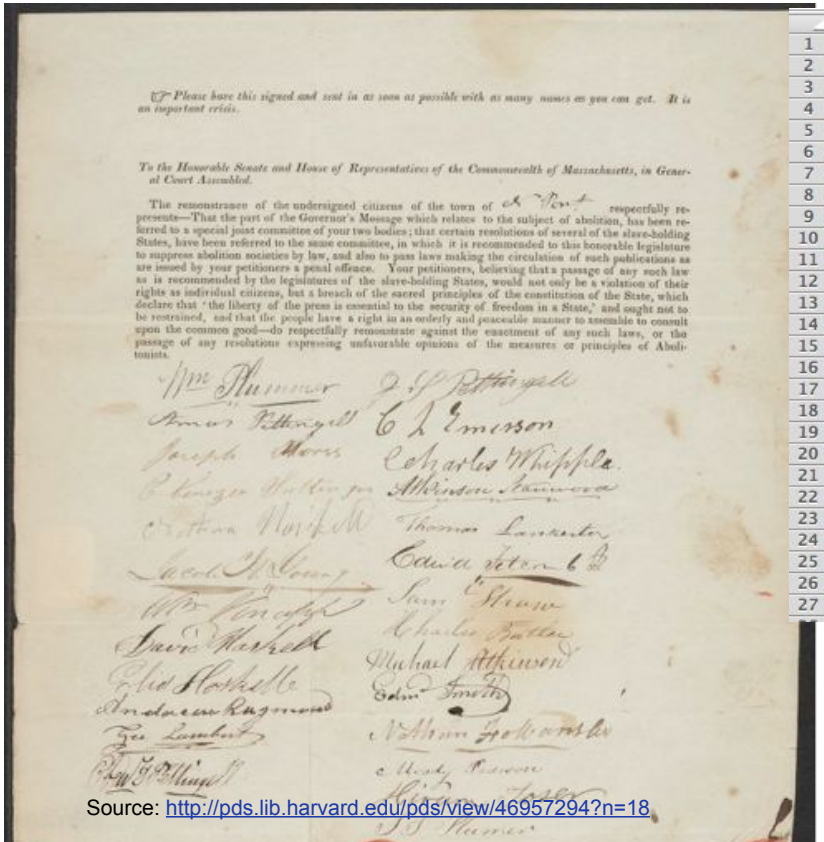
Harvard University economists Carmen Reinhart and Kenneth Rogoff have acknowledged making a spreadsheet calculation mistake in a 2010 research paper, "Growth in a Time of Debt" (PDF), which has been widely cited to justify budget-cutting. But the authors stand by their conclusion that higher government debt is associated with slower economic growth. Here's what you need to know:

Data was not publicly available!

Source: http://www.bloomberg.com/bw/articles/2013-04-18/faq-reinhart-rogoff-and-the-excel-error-that-changed-history

Understand our past

The Digital Archive of Massachusetts Anti-Slavery & Anti-Segregation Petitions



Please have this signed and sent in as soon as possible with as many names as you can get. It is an important crisis.

To the Honorable Senate and House of Representatives of the Commonwealth of Massachusetts, in General Court Assembled.

The remonstrance of the undersigned citizens of the town of [illegible] respectfully represents—That the part of the Governor's Message which relates to the subject of abolition, has been referred to a special joint committee of your two bodies; that certain resolutions of several of the slave-holding States, have been referred to the same committee, in which it is recommended to this honorable legislature to suppress abolition societies by law, and also to pass laws making the circulation of such publications as are issued by your petitioners a penal offence. Your petitioners, believing that a passage of any such law as is recommended by the legislatures of the slave-holding States, would not only be a violation of their rights as individual citizens, but a breach of the sacred principles of the constitution of the State, which declare that 'the liberty of the press is essential to the security of freedom in a State,' and ought not to be restrained, and that the people have a right in an orderly and peaceable manner to assemble to consult upon the common good—do respectfully remonstrate against the enactment of any such laws, or the passage of any resolutions expressing unfavorable opinions of the measures or principles of Abolitionists.

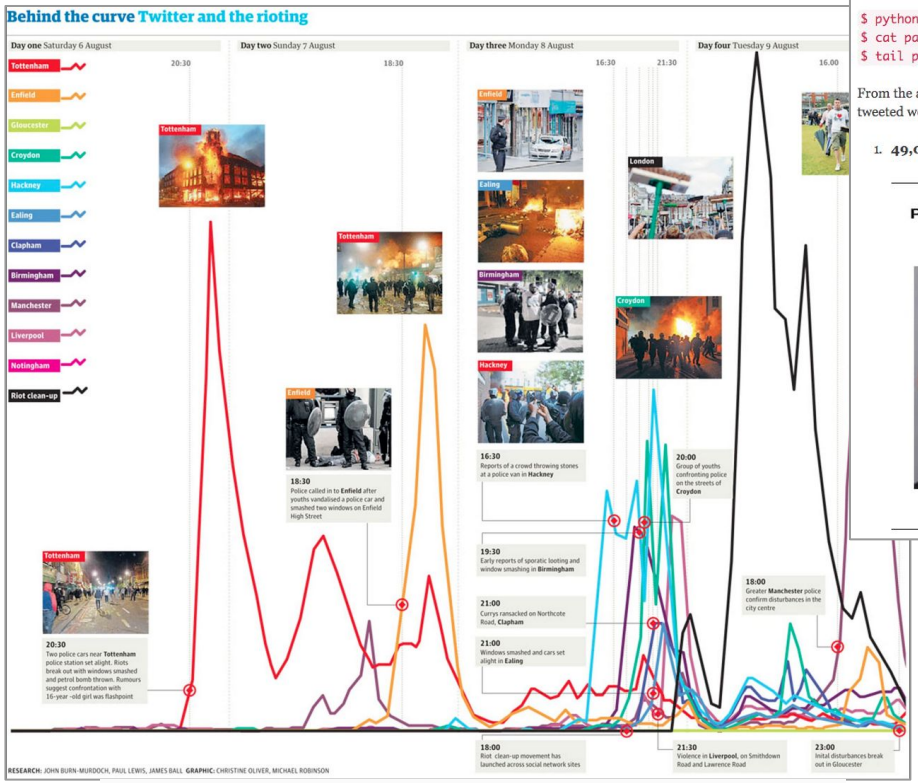
Source: http://pds.lib.harvard.edu/pds/view/46957294?n=18

	A	B	C	D	E	F	G	H
1	Location of the petition at the Massachusetts Archive	PDS link	Scholarly cita	Date of creatio	Date rece	Legislati	dateaction1	dateaction2
2	Resolves 1783, c.97, passed March 1, 1784	http://nrs.harvard.edu/urn	Digital Archiv	17840130	Not available			
3	Resolves 1784, c.175, passed March 18, 1785	http://nrs.harvard.edu/urn	Digital Archiv	17840130	Received	Receive	17840206	17840209
4	Resolves 1784, c.175, passed March 18, 1785	http://nrs.harvard.edu/urn	Digital Archiv	17840130	Received	Receive	17840206	17840209
5	Resolves 1785, c.36, passed November 5, 1785	http://nrs.harvard.edu/urn	Digital Archiv	17850610	Not available			
6	Resolves 1785, c.48, passed November 9, 1785	http://nrs.harvard.edu/urn	Digital Archiv	17851104	Not available			
7	Resolves 1785, c.17, passed February 15, 1786	http://nrs.harvard.edu/urn	Digital Archiv	Not available	Committe	Committed, concurred		
8	Resolves 1787, c.24, passed June 16, 1787	http://nrs.harvard.edu/urn	Digital Archiv	17870522	Report to	Report to concur		
9	Resolves 1787, c.2, passed October 19, 1787	http://nrs.harvard.edu/urn	Digital Archiv	17871000	Concurre	Concurred		
10	Resolves 1787, c.88, passed November 19, 1787	http://nrs.harvard.edu/urn	Digital Archiv	17871000	Not available			
11	Resolves 1787, c.142, passed November 23, 1787	http://nrs.harvard.edu/urn	Digital Archiv	Not available	Not available			
12	St. 1787, c.48, passed March 26, 1788	http://nrs.harvard.edu/urn	Digital Archiv	17880227	Actions or	Receive	17880314	17880319
13	St. 1788, c.2, passed June 12, 1788	http://nrs.harvard.edu/urn	Digital Archiv	Not available	Not available			
14	St. 1788, c.2, passed June 12, 1788	http://nrs.harvard.edu/urn	Digital Archiv	17880500	Sent to a	Received, referred		
15	St. 1788, c.38, passed January 30, 1789	http://nrs.harvard.edu/urn	Digital Archiv	17890100	Not available			
16	St. 1788, c.38, passed January 30, 1789	http://nrs.harvard.edu/urn	Digital Archiv	17881229	Not available			
17	St. 1788, c.38, passed January 30, 1789	http://nrs.harvard.edu/urn	Digital Archiv	17880730	Not available			
18	St. 1788, c.38, passed January 30, 1789	http://nrs.harvard.edu/urn	Digital Archiv	17881105	Received	Receive	17881117	17881118
19	St. 1788, c.38, passed January 30, 1789	http://nrs.harvard.edu/urn	Digital Archiv	17881031	Not available			
20	Resolves 1789, c.63, passed June 15, 1789	http://nrs.harvard.edu/urn	Digital Archiv	Not available	Received	Receive	17890528	17890528
21	Resolves 1789, c.119, passed June 26, 1789	http://nrs.harvard.edu/urn	Digital Archiv	17890617	Not available			
22	St. 1789, c.52, passed March 4, 1790	http://nrs.harvard.edu/urn	Digital Archiv	17900115	Committe	Committed and read		
23	Resolves 1790, c.12, passed June 4, 1790	http://nrs.harvard.edu/urn	Digital Archiv	17900521	Not available			
24	Resolves 1790, c.116, passed March 8, 1791	http://nrs.harvard.edu/urn	Digital Archiv	17910223	Not available			
25	Resolves 1790, c.127, passed March 9, 1791	http://nrs.harvard.edu/urn	Digital Archiv	17900610	Not available			
26	Resolves 1791, c.99, passed February 20, 1792	http://nrs.harvard.edu/urn	Digital Archiv	17920109	Not available			
27	Resolves 1792, c.67, passed June 28, 1792	http://nrs.harvard.edu/urn	Digital Archiv	17920616	Not available			

Source: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XN9WP

For more Digital Humanities projects see: http://www.darthcrimson.org/

And the present



Source: http://www.theguardian.com/uk/2011/aug/24/twitter-study-post-riot-plans#img-1

A LOOK AT 14,939,154 #PARIS #BATACLAN #PARISATTACKS #PORTEOUVERTE TWEETS

Submitted by Nick Ruest on Sat, 12/12/2015 - 14:49

IMAGES

We are able to create a list of images tweeted in our dataset by using `image_urls.py`.

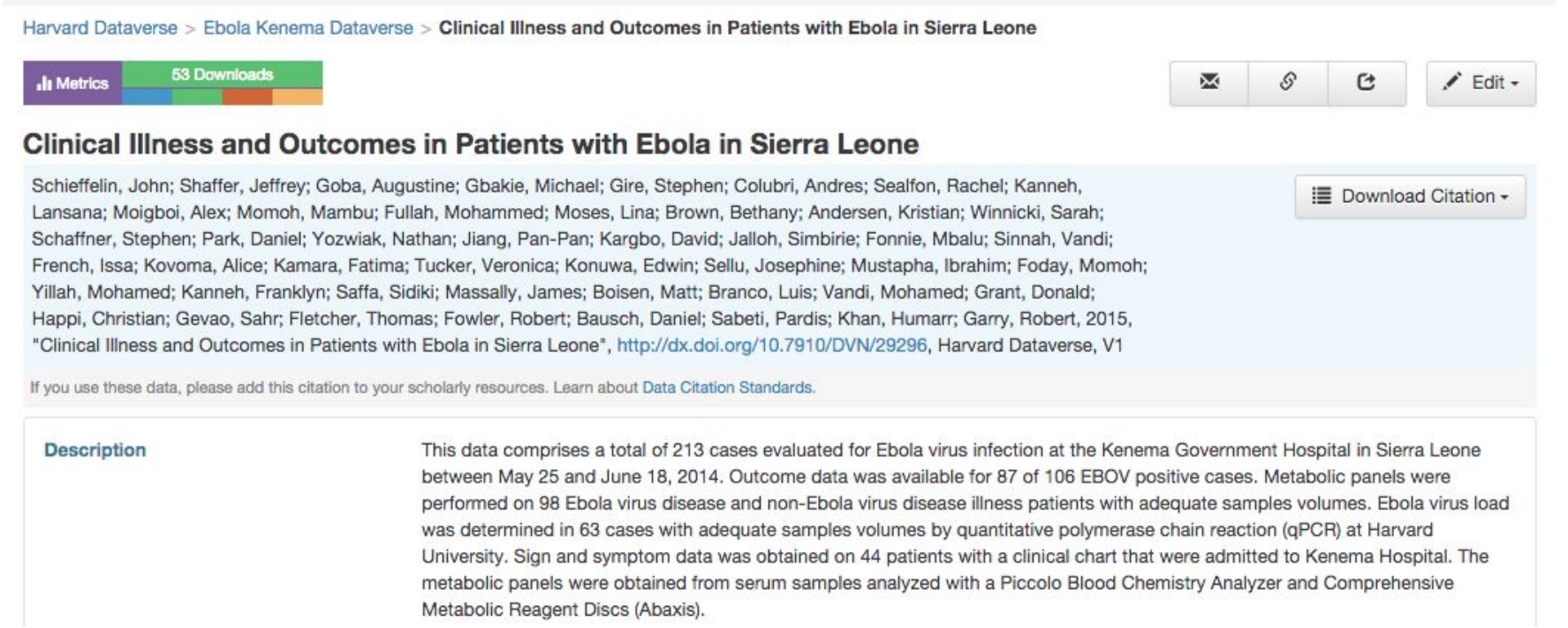
```
$ python ~/git/twarc/utils/image_urls.py paris-valid-deduplicated.json > paris-tweets-images.txt
$ cat paris-tweets-images.txt | sort | uniq -c | sort -n > paris-tweets-images-uniq.txt $ cat paris-tweets-images-uniq.txt | wc -l
$ tail paris-tweets-images-uniq.txt
```

From the above, we can see that there were 6,872,441 total images tweets, representing 46.00% of total tweets, and 660,470 unique images. The top 10 images tweeted were as follows:

1. **49,051 Occurrences**

Source: http://ruebot.net/post/look-14939154-paris-bataclan-parisattacks-porteouverte-tweets

Understand outcomes of deadly viruses



Harvard Dataverse > Ebola Kenema Dataverse > **Clinical Illness and Outcomes in Patients with Ebola in Sierra Leone**

Metrics | 53 Downloads

Edit

Clinical Illness and Outcomes in Patients with Ebola in Sierra Leone

Schieffelin, John; Shaffer, Jeffrey; Goba, Augustine; Gbakie, Michael; Gire, Stephen; Colubri, Andres; Sealfon, Rachel; Kanneh, Lansana; Moigboi, Alex; Momoh, Mambu; Fullah, Mohammed; Moses, Lina; Brown, Bethany; Andersen, Kristian; Winnicki, Sarah; Schaffner, Stephen; Park, Daniel; Yozwiak, Nathan; Jiang, Pan-Pan; Kargbo, David; Jalloh, Simbirie; Fonnie, Mbalu; Sinnah, Vandi; French, Issa; Kovoma, Alice; Kamara, Fatima; Tucker, Veronica; Konuwa, Edwin; Sellu, Josephine; Mustapha, Ibrahim; Foday, Momoh; Yillah, Mohamed; Kanneh, Franklyn; Saffa, Sidiki; Massally, James; Boisen, Matt; Branco, Luis; Vandi, Mohamed; Grant, Donald; Happi, Christian; Gevao, Sahr; Fletcher, Thomas; Fowler, Robert; Bausch, Daniel; Sabeti, Pardis; Khan, Humarr; Garry, Robert, 2015, "Clinical Illness and Outcomes in Patients with Ebola in Sierra Leone", http://dx.doi.org/10.7910/DVN/29296, Harvard Dataverse, V1

Download Citation

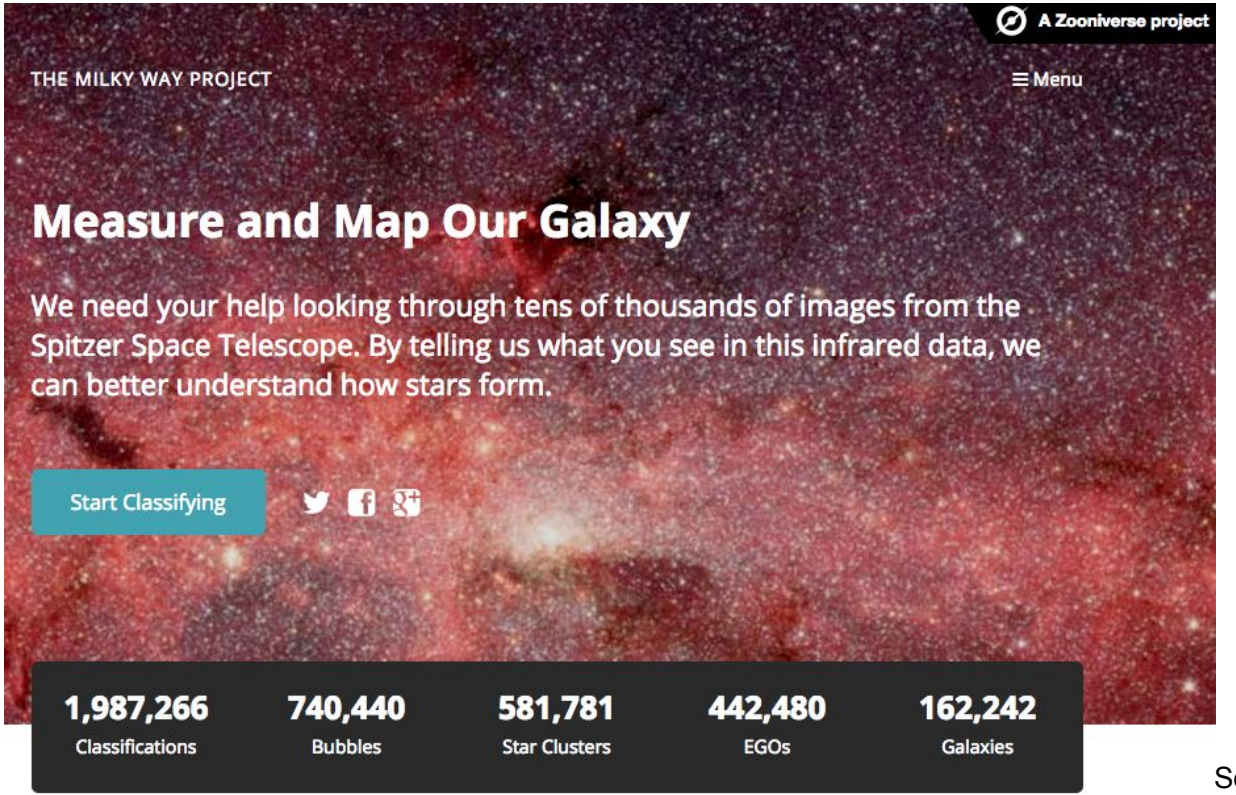
If you use these data, please add this citation to your scholarly resources. Learn about Data Citation Standards.

Description

This data comprises a total of 213 cases evaluated for Ebola virus infection at the Kenema Government Hospital in Sierra Leone between May 25 and June 18, 2014. Outcome data was available for 87 of 106 EBOV positive cases. Metabolic panels were performed on 98 Ebola virus disease and non-Ebola virus disease illness patients with adequate samples volumes. Ebola virus load was determined in 63 cases with adequate samples volumes by quantitative polymerase chain reaction (qPCR) at Harvard University. Sign and symptom data was obtained on 44 patients with a clinical chart that were admitted to Kenema Hospital. The metabolic panels were obtained from serum samples analyzed with a Piccolo Blood Chemistry Analyzer and Comprehensive Metabolic Reagent Discs (Abaxis).

Source: Harvard Dataverse http://dx.doi.org/10.7910/DVN/29296

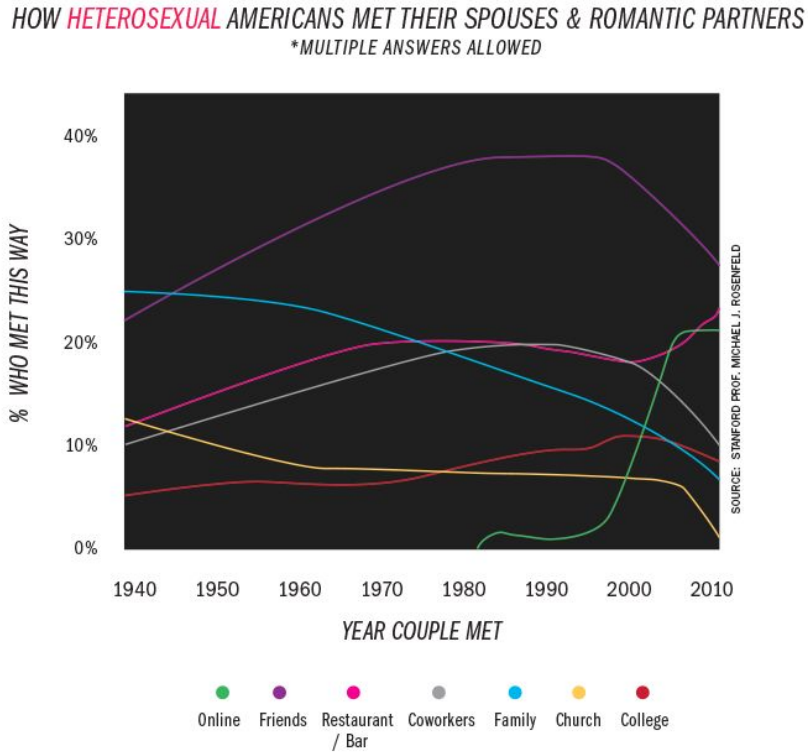
The mysteries held in the stars above

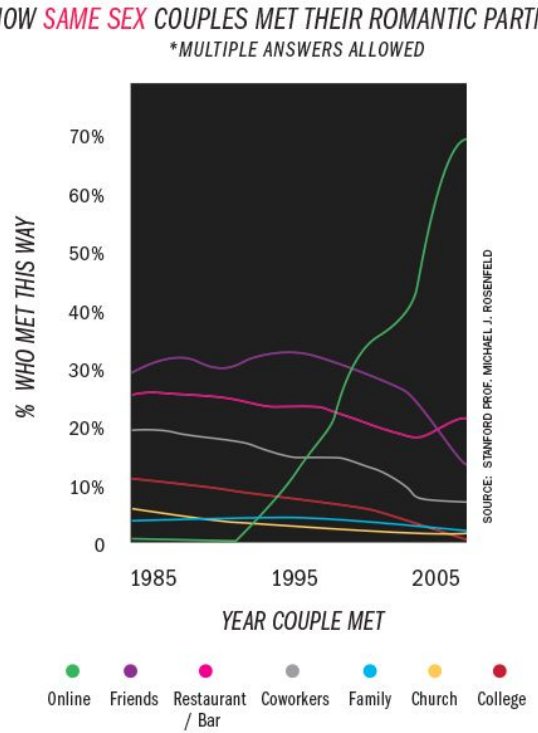


See also: theastrodata.org for Harvard-Smithsonian datasets

Source: http://www.milkywayproject.org/

And most important: Modern Romance





Sources: http://time.com/aziz-ansari-modern-romance/ & How Couples Meet and Stay Together (HCMST) dataset (http://data.stanford.edu/hcmst)

Where can data be found?

Start at➔

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Only a sample of what is out there!

The Dataverse Project

HUBBLESITE

ads ARCHAEOLOGY DATA SERVICE

zenodo

figshare

RCSB PDB
PROTEIN DATA BANK

biosharing

DRYAD

ICPSR A PARTNER IN SOCIAL SCIENCE RESEARCH | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

DATA.GOV

American Mineralogist Crystal Structure Database

What tools can I use to analyze data?

Some examples of tools:

Best Practices to Remember:

- use open or common file formats
- describe your data for discovery and reuse (metadata)
- have a data management plan (DMPTool at Harvard)
- archive & share your data for reuse (e.g., Harvard Dataverse)

Further Reading & Resources at Harvard

- Finding Data Resources at Harvard Library: http://guides.library.harvard.edu; or contact: govdocs@fas.harvard.edu
- Harvard Open Data Assistance Program Weekly Office Hours (Weds 11am-1pm) http://projects.iq.harvard.edu/odap/open-hours
- Data Science Services: Free research and training services for Harvard affiliates http://rtc.iq.harvard.edu/

Books

- Dataclysm: Who We Are When We Think No One Is Looking by Christian Rudder http://id.lib.harvard.edu/aleph/014155996/catalog
- Big data, little data, no data : scholarship in the networked world by Christine L. Borgman http://id.lib.harvard.edu/aleph/014297454/catalog