Information Design

Professor Danne Woo

infodesign.dannewoo.com

ARTS 269

Spring 2020

Wednesday 1:40 PM – 5:30 PM

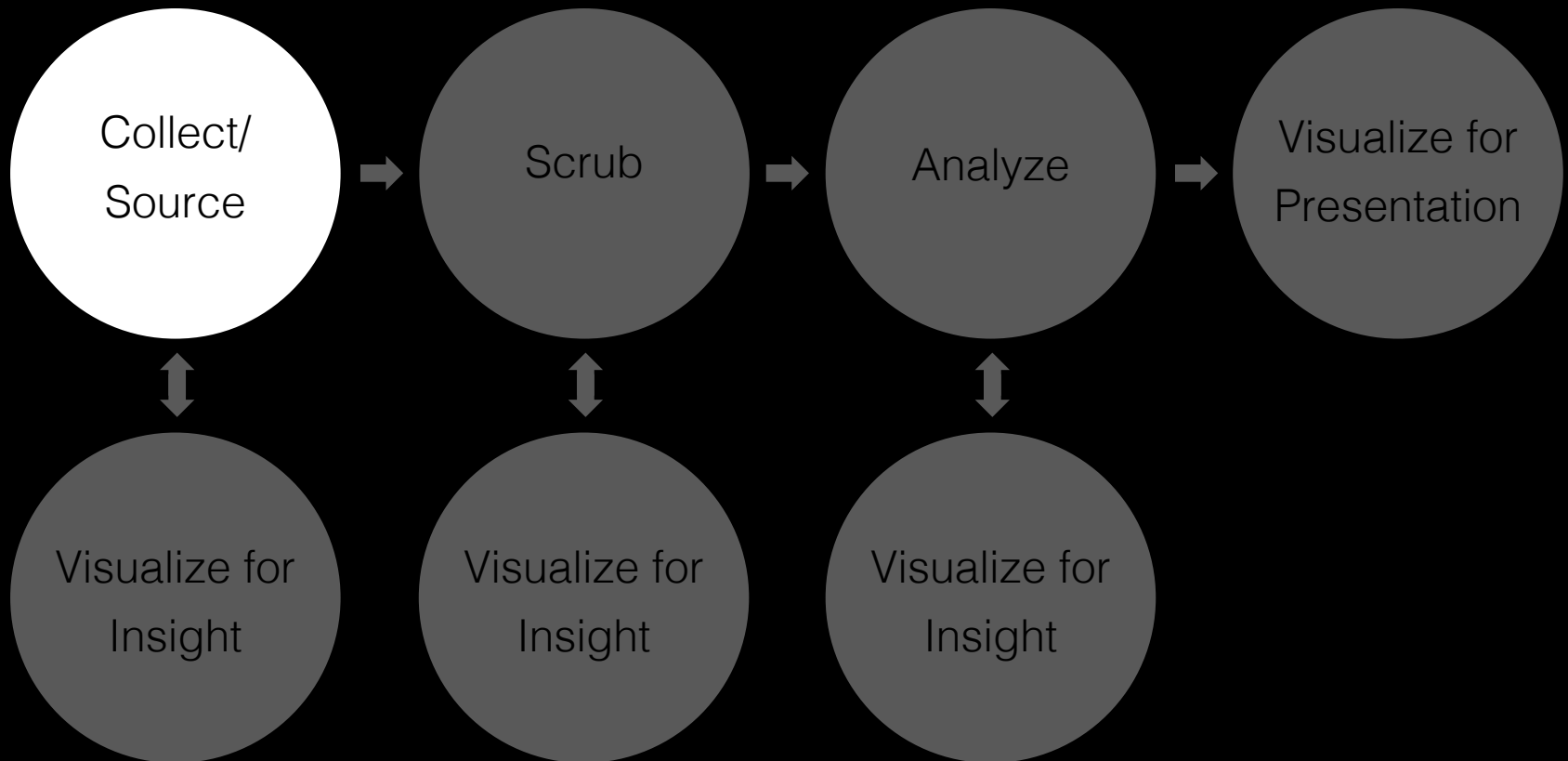Klapper 107

Source, Scrub, Analyze

Data Pipeline

What is Data?

A collection of facts.

Numbers	Observations
Words	Pictures
Measurements	Videos
Descriptions	Audio

Data

Data Types – Structured vs Unstructured

Rank This Week	Song	Artist
1	Despacito	Luis Fonsi & Daddy Yankee Featuring Justin Bieber
2	Wild Thoughts	DJ Khaled Featuring Rihanna & Bryson Tiller
3	Bodak Yellow (Money Moves)	Cardi B
4	Believer	Imagine Dragons
5	Attention	Charlie Puth
6	Unforgettable	French Montana Featuring Swae Lee
7	There's Nothing Holdin' Me Back	Shawn Mendes
8	That's What I Like	Bruno Mars
9	Shape Of You	Ed Sheeran
10	Rake It Up	Yo Gotti Featuring Nicki Minaj
11	Strip That Down	Liam Payne Featuring Quavo
12	Bank Account	21 Savage
13	Body Like A Back Road	Sam Hunt
14	Slow Hands	Niall Horan
15	Congratulations	Post Malone Featuring Quavo
16	I'm The One	DJ Khaled Featuring Justin Bieber, Quavo, Chance The Rapper & Lil Wayne
17	Redbone	Childish Gambino
18	Sorry Not Sorry	Demi Lovato
19	Humble.	Kendrick Lamar
20	Friends	Justin Bieber + BloodPop
21	Something Just Like This	The Chainsmokers & Coldplay
22	Feels	Calvin Harris Featuring Pharrell Williams, Katy Perry & Big Sean
23	Feel It Still	Portugal. The Man

Comin' over in my direction

So thankful for that, it's such a blessin', yeah

Turn every situation into heaven, yeah

Oh-oh, you are

My sunrise on the darkest day

Got me feelin' some kind of way

Data

Structured or Unstructured?

Data

Structured or Unstructured?

19:03:46, 12ZH1bPJZDGHyUThjQgYvUqXq87LqrGRKp, 9294716,
1B2YvBPW23S8v5ouPdKm4kozSzYNzmGb43, 240370,
8b7dfc4c42328512196333f2c23e1923b003e154edffc2486d071eba2578b5ba
19:03:45, 1LwtxmX13HZTDDcUEzmoW2Av2GxDBHeAMD, 399980000,
70066209f0bb4a2b0196ff3a0adff952d77cacf96890766c0d333371a7af0d70
19:03:44, 12BApnQD51QcqsjrBR8DCjcLdijHoDAs7P, 81342236,
fd53151661778c8162e4db747d38bdb9a9a0dafce4a375e5b2824291de67a782
19:03:44, 1DbCAqvSHPxKrGUc5K7E7UxYBhsAk97Dya, 6827706,
8147b14900c0587b5f5ac1e57fb1bfc17cb62fdf3896bde7021eb05f17462cb8
19:03:42, 1Dhh8W17PkRFrBYK7CZRZGHdUrGLGYfRK8, 999000000,
438dae53f595c20dc51e972ebfcb6cf96046272608e4fa1d6ef2f593a20c82bb
19:03:42, 1BAyaPneSPVheVYvX2p4Qrm9BygtWyGVx2, 306651770,

Data

Structured or Unstructured?

Data

Structured or Unstructured?

Ken: I have "heard" from two sources in this office that Enron will announce my successor today. I have not been advised of the selection and am concerned that this announcement will not treat my retirement in an appropriate and deserving manner...

Frankly, I think this matter has not been handled in a caring way and I have been accused of giving information on the selection to Tom DeLay, which is untrue. My office mates are all more aware of the candidates and their status than I have been.

I regret having to contact you on this matter but feel that despite my total cooperation, allegiance and dedication to Enron, I am not being treated with the courtesy and sensitivity that this matter deserves.

Data

Structured or Unstructured?

```
Delivered-To: kenneth.lay@enron.com
Received: by 10.112.161.231 with SMTP id xv7csp295280lbb; Tue, 25 Sep 2000 01:01:10 -0800 (PST)
Return-Path: <bounce+a5d1e6.0977-kenneth.lay=enron.com@enron.com>
Received: from mail-s83.mailgun.info (mail-s83.mailgun.info. [184.173.153.211]) by mx.enron.com with ESMTP id
z8si1471337yhb.134.2013.11.19.01.01.08 for <kenneth.lay@enron.com>; Tue, 25 Sep 2000 01:01:08 -0800 (PST)
Received-SPF: pass (enron.com: domain of bounce+a5d1e6.0977-kenneth.lay=enron.
com@enron.com designates 184.173.153.211 as permitted sender) client-ip=184.173.153.211;
Date: Tue, 25 Sep 2000 09:00:03 +0000
From: joe.hillings@enron.com
To: kenneth.lay@enron.com
Subject: Selection of My Successor — My Concern
Message-Id: <528b2893.J26eP57xAov4mdbI%joe.hillings@enron.com>
User-Agent: Mozilla/4.0 (compatible; Lotus-Notes/5.0; Windows-NT)
Mime-Version: 1.0
Content-Type: text/plain; charset="us-ascii"
X-Mailgun-Sid: WyI2M2M0NyIsICJhbWFkZXVzQGRlY29kZWQuY28iLCAiMDk3NyJd
Sender: joe.hillings@enron.com
Content-Transfer-Encoding: 7bit
```

Data

Structured or Unstructured?

Data

Structured or Unstructured?

Collecting Data

Blombos Ocher Plaque – 75000BC

Collecting Data

John Graunt – 1662

Collecting Data

Herman Hollerith – 1890
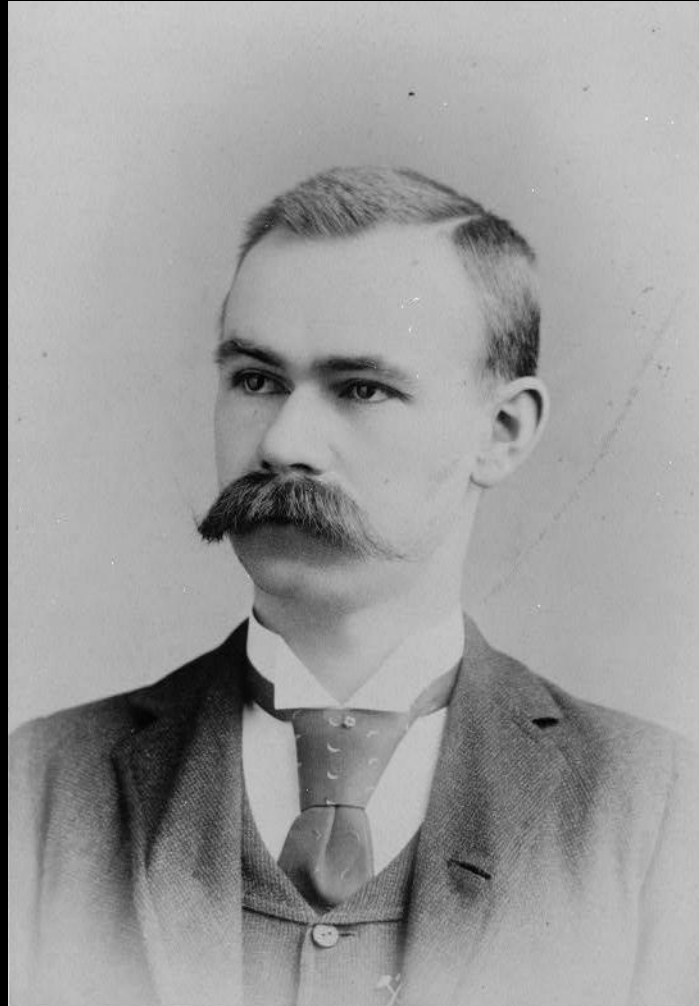
Collecting Data

Jacquard Loom
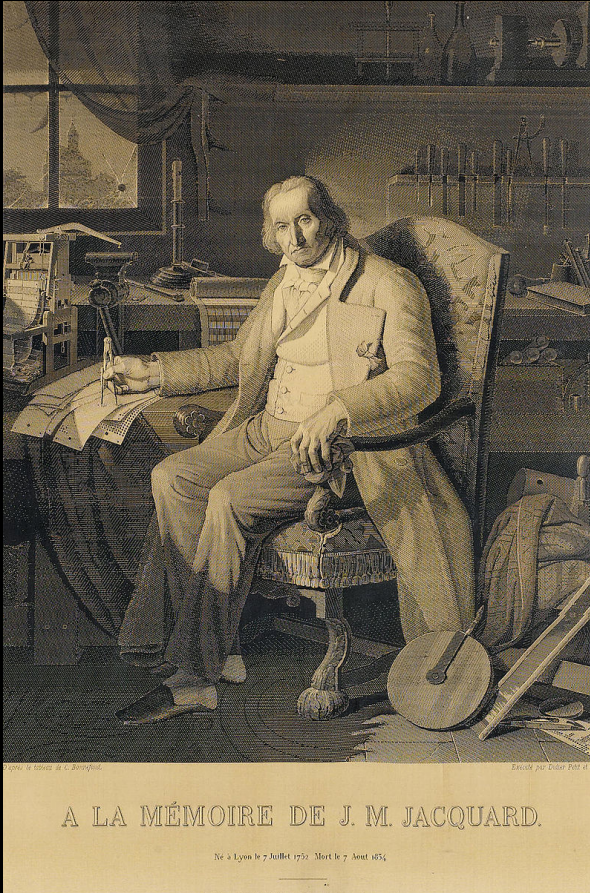
Collecting Data

Herman Hollerith – 1890

Collecting Data

The Internet – 1982
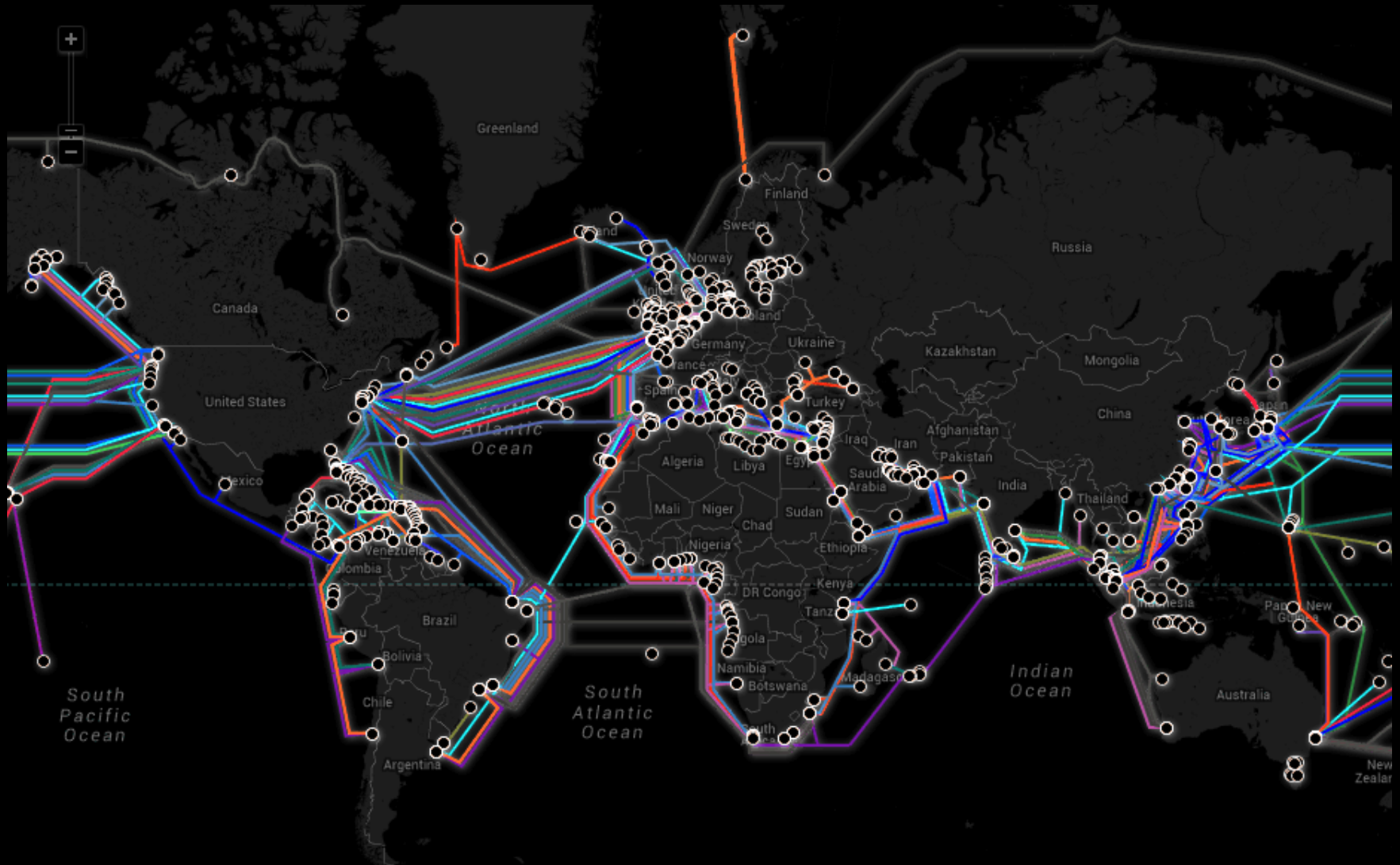
Collecting Data

World Wide Web – 1989

Collecting Data

How the Web Works

Collecting Data

Every Company is a Data Company

Collecting Data

Every Company is a Data Company
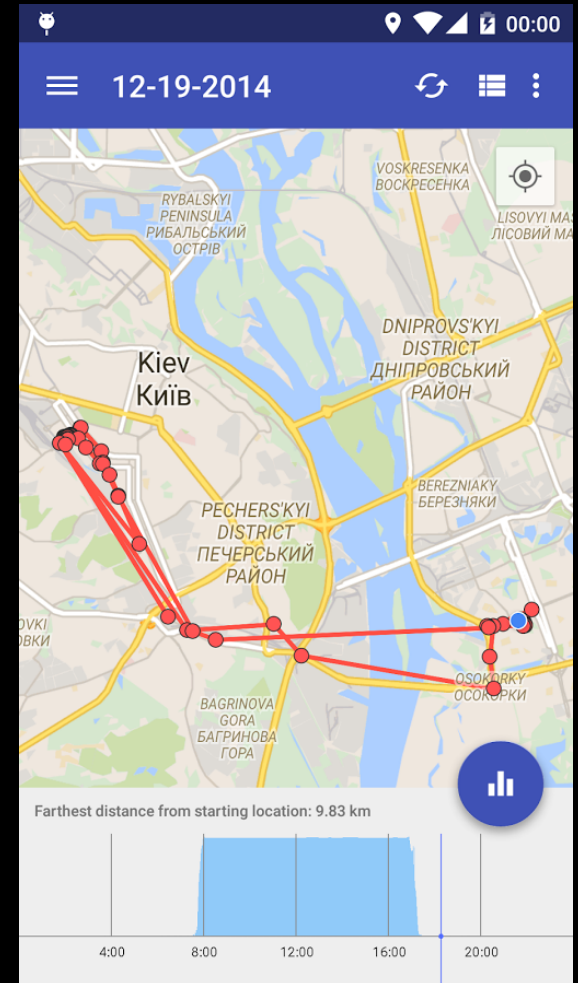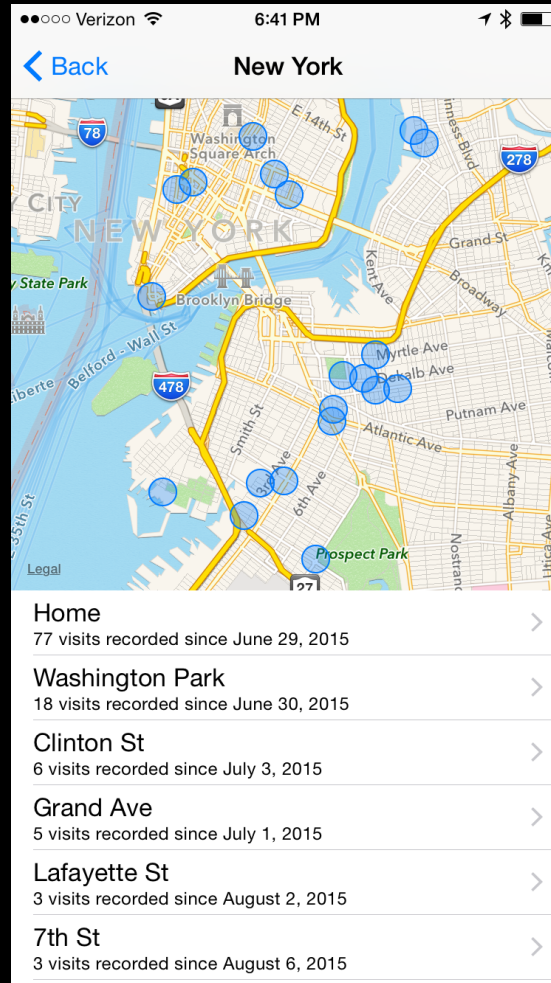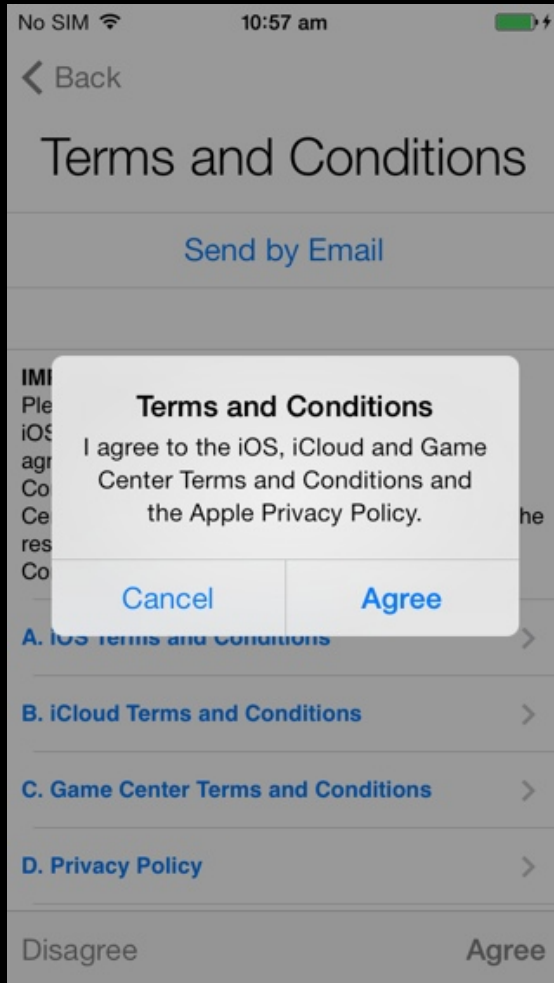
Collecting Data

Take Back Your Data

Collecting Data

Nicholas Felton

Collecting Data

Personal Data – Software
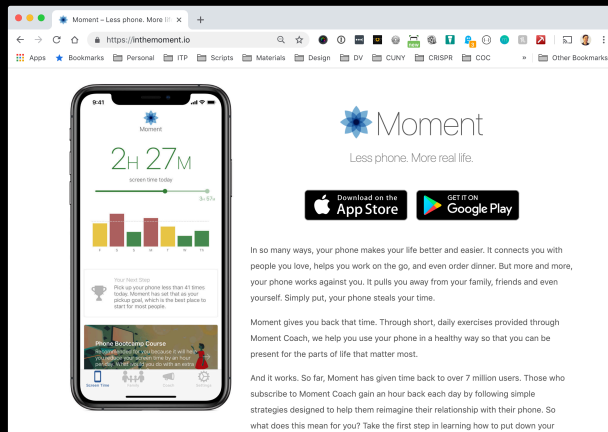
Moment App
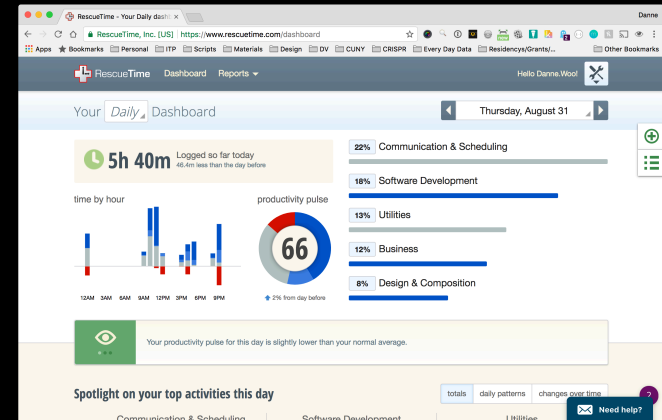
Rescue Time

Feltron Reporter App

Apple iPhone Health Sensors

Collecting Data

Personal Data – Software

Gyroscope App (www.gyrosco.pe)
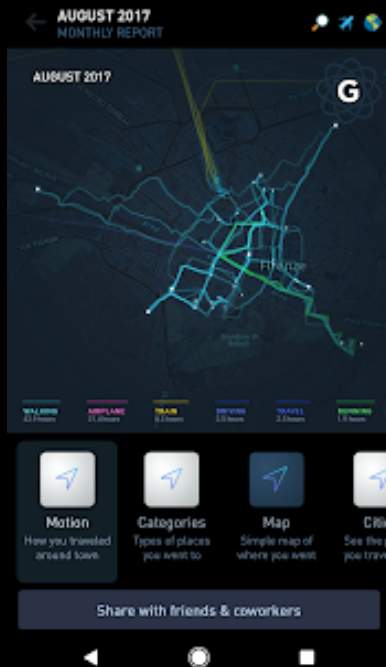
Collecting Data

Personal Data – Apple and Facebook Data
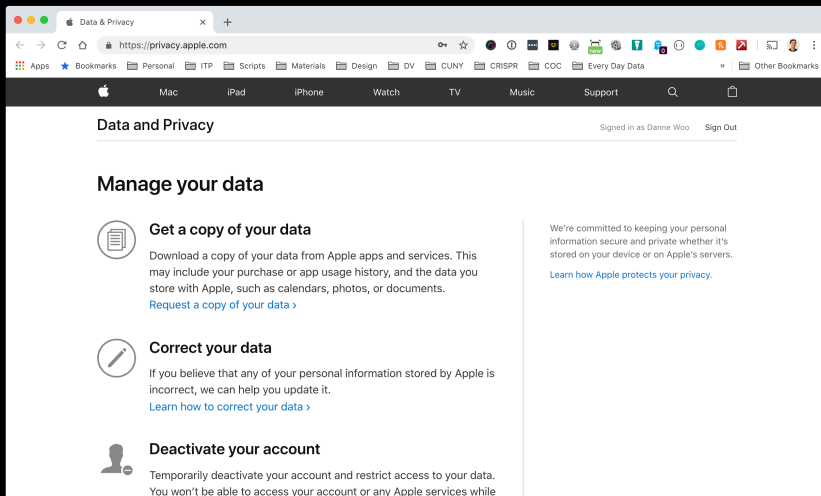
Apple Data
privacy.apple.com
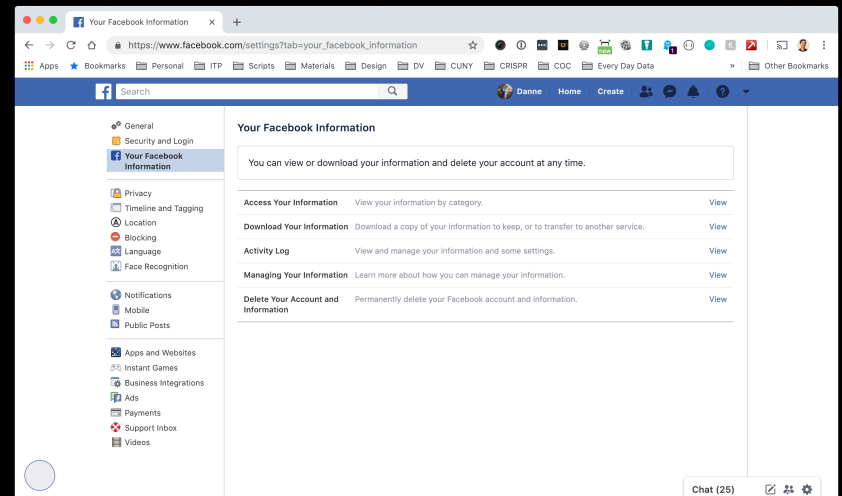
Facebook Data
https://www.facebook.com/settings
?tab=your_facebook_information

Collecting Data

Personal Data – Hardware

Fitbit

Jawbone

Nike Fuelband

Apple Watch

Collecting Data

Personal Data – Hardware

Automatic
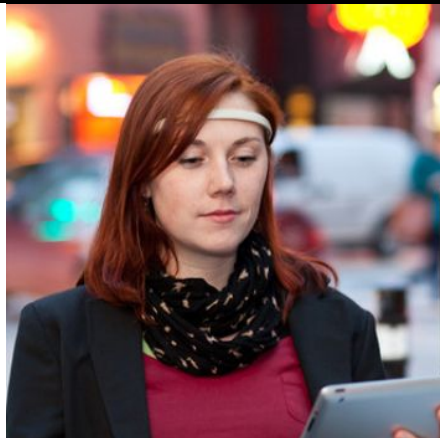
Muse Headband

Nest

Vessyl

Collecting Data

Tracking the Trackers

The House That Spied on Me

https://gizmodo.com/the-house-that-spied-on-me-1822429852
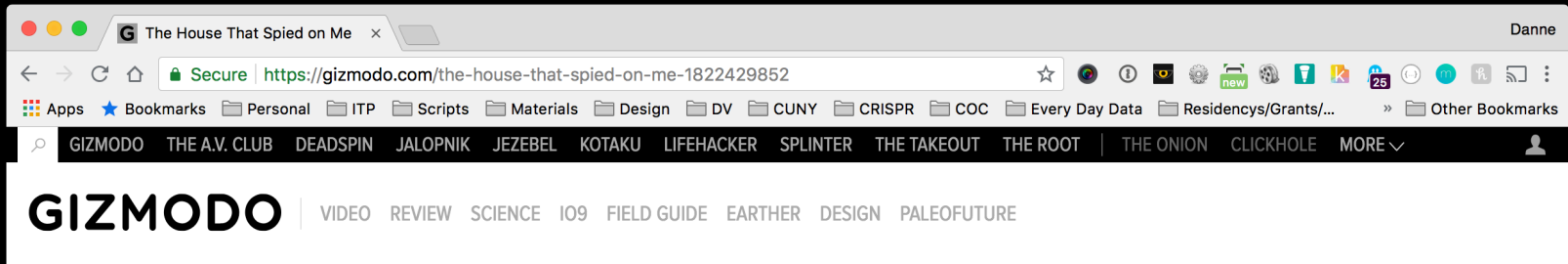
Collecting Data

The Connected Us

Collecting Data

Rodrigo Narciso – CH4

Sourcing Data

The Internet (Not Just Cat Memes)
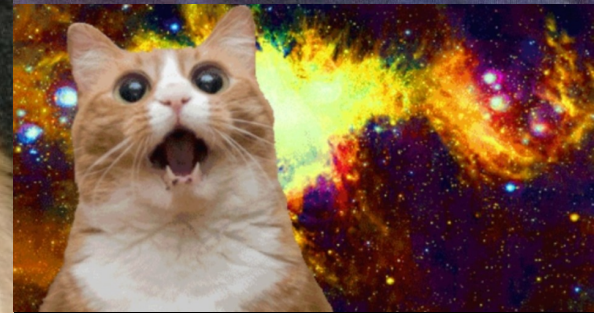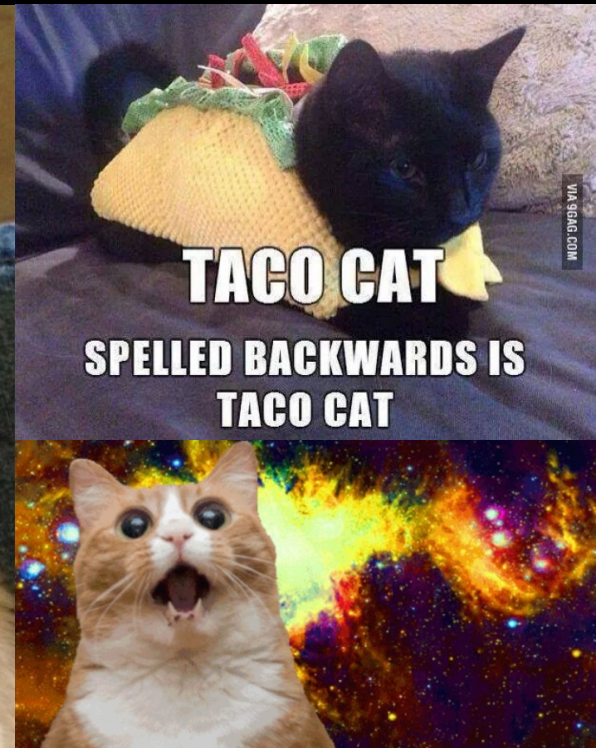
Sourcing Data

But mostly cat memes
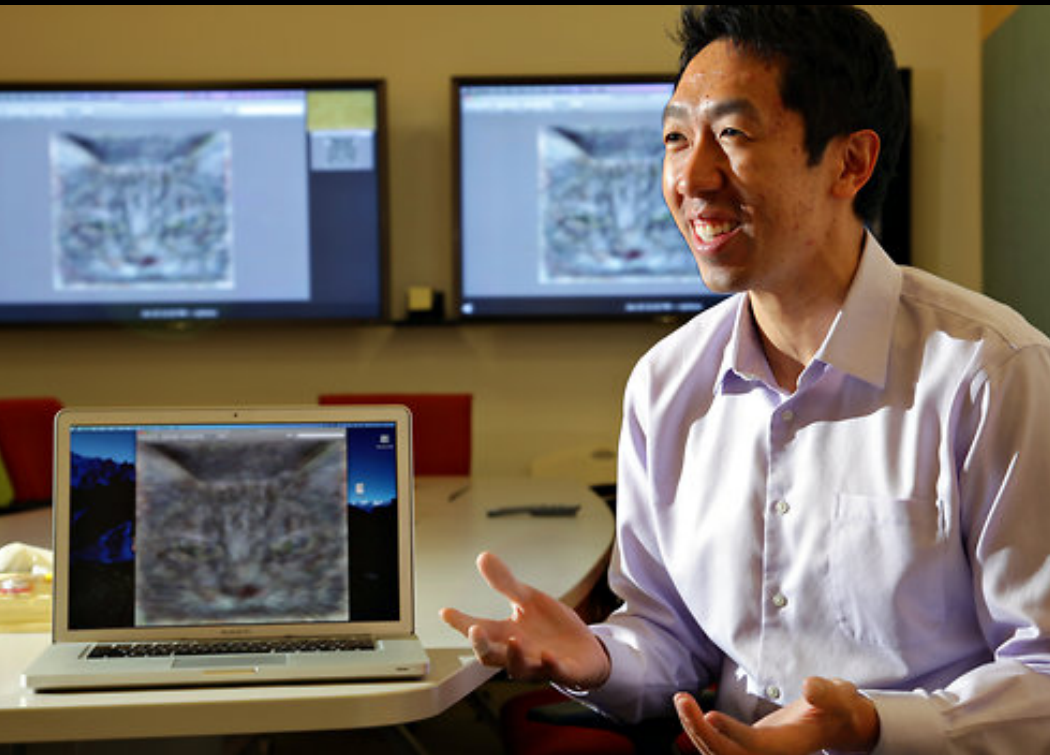
Sourcing Data

Open Datasets
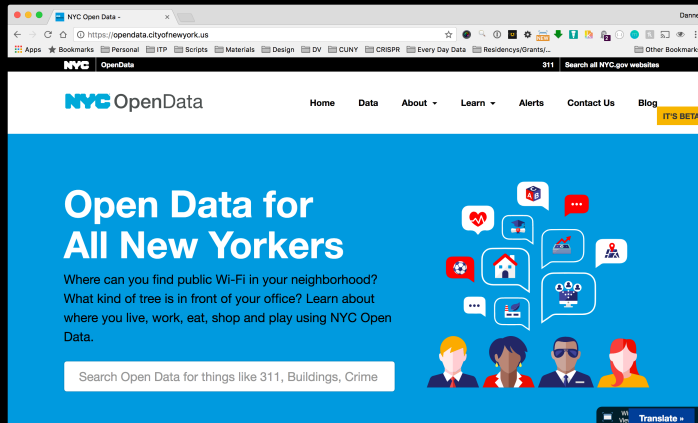
data.cityofnewyork.us
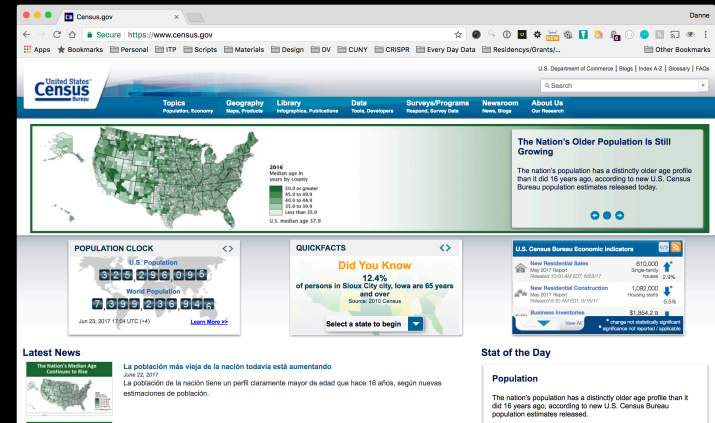
www.census.gov
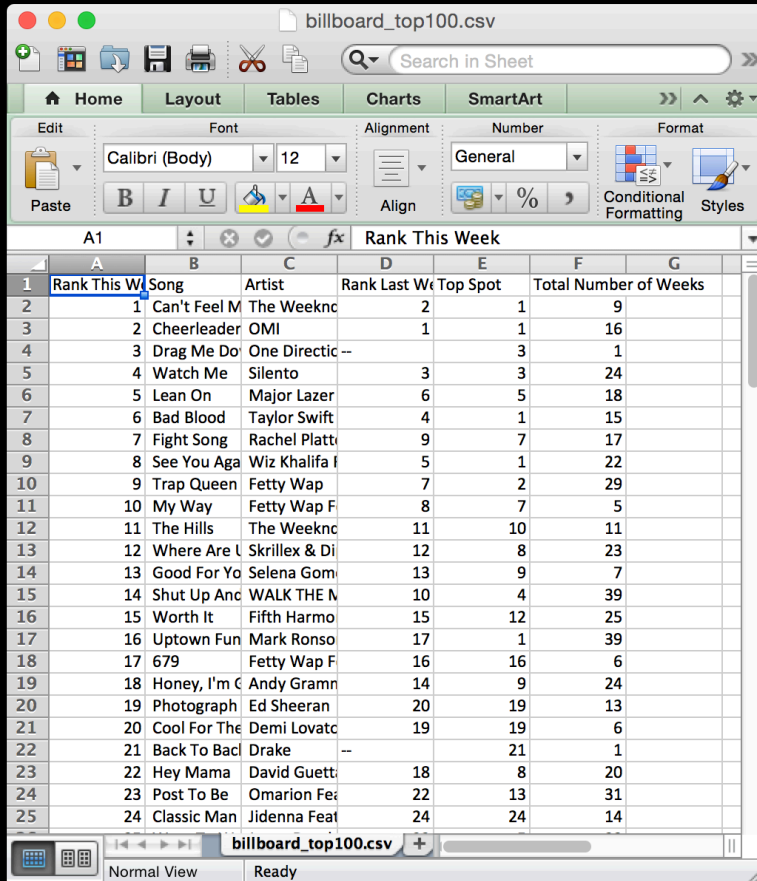
ucr.fbi.gov
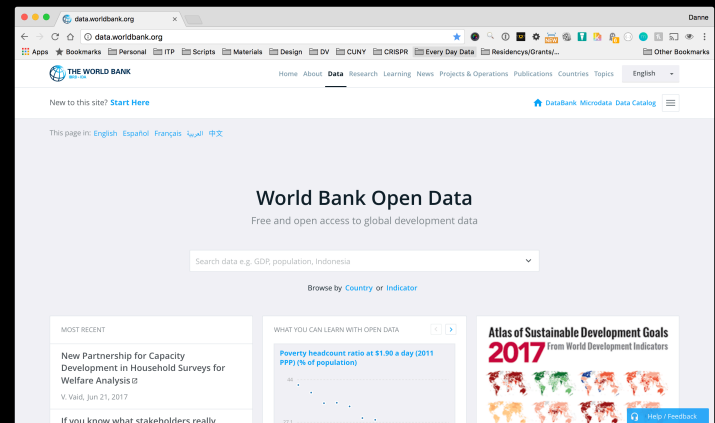
data.worldbank.org

Sourcing Data

Data Formats - Tabular

CSV – Comma Separated Value

TSV – Tab Separated Value

XLS – Excel

Sourcing Data

HTML

Tables
Easily Copy and Paste

http://www.baseball-reference.com/

Complex Layout
HTML Web Scrapping – Data-miner or Python

http://www.billboard.com/charts/hot-100

Sourcing Data

Web Scraping

www.data-miner.io

Sourcing Data

HTML based Data Tables (Tabula Free Software)

TABLE 1 Total number of reported disasters,[1] by continent and by year (2006-2015)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total[3]
Africa	202	184	173	156	135	165	124	115	101	116	1,471
Americas	105	133	143	115	146	131	114	106	116	124	1,233
Asia	308	262	240	233	253	235	210	228	228	240	2,437
Europe	98	104	58	75	99	49	91	69	80	70	793
Oceania	18	11	13	19	18	15	14	12	12	24	156
Very high human development[2]	124	118	104	99	113	94	119	123	112	113	1,119
High human development	202	224	193	183	213	159	165	155	192	183	1,869
Medium human development	199	158	170	154	161	165	126	130	126	146	1,535
Low human development	206	194	160	162	164	177	143	122	107	132	1,567
Total	731	694	627	598	651	595	553	530	537	574	6,090

Source: EM-DAT, CRED, University of Louvain, Belgium

[1] In Tables 1–13, 'disasters' refer to those with a natural and/or technological trigger only, and do not include wars, conflict-related famines, diseases or epidemics.

[2] See note on UNDP's Human Development Index country status in the disaster definitions section in the introduction to this annex.

[3] Since slow-onset disasters can affect the same country for a number of years, it is best to use figures on total numbers to calculate annual averages over a decade rather than as absolute totals (see the data definitions and methodology section in the introduction to this annex).

Note: Some totals in the table may not correspond due to rounding.

With 574 disasters reported, 2015 is the year of the decade with the fourth lowest number of disasters, very far below the peak of 2006.

Among continents, the number of disasters was the third lowest of the decade in Africa and the fourth lowest in Europe. In contrast, this number was the fourth highest of the decade in the Americas and in Asia; and Oceania suffered its highest number of disasters of the decade.

In 2015, the number of disasters was at its third lowest level in countries with low human development and at its fourth lowest in medium and high human development countries.

In countries with very high human development, the number of disasters was, in 2014, the fifth lowest in the decade.

The distribution of disasters among continents is, in 2015, very similar to the annual average for the decade.
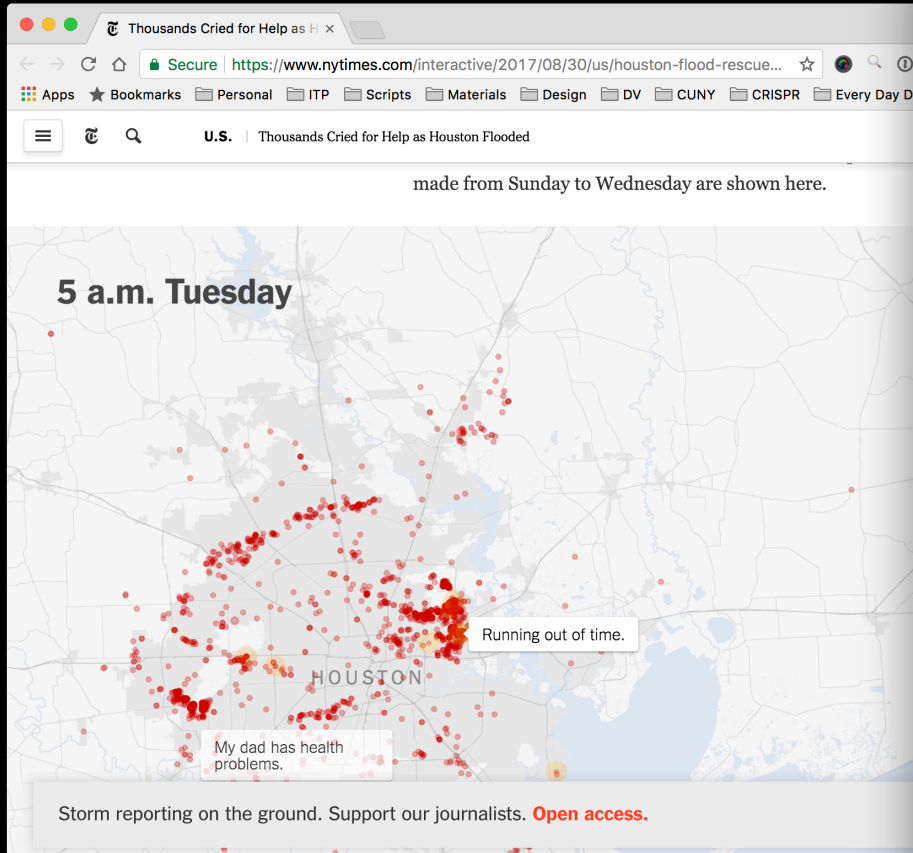
Sourcing Data

Interactive Data Visualizations (Look at the code, view source)

Sourcing Data

APIs (Application Programming Interfaces)

developer.nytimes.com
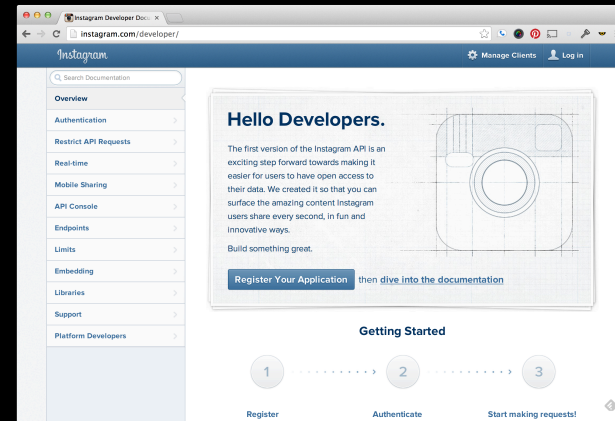
instagram.com/developer
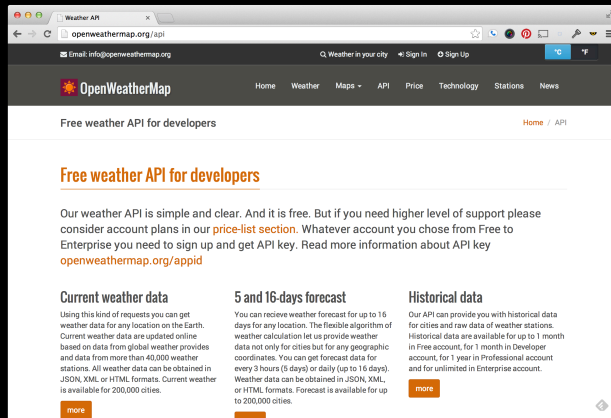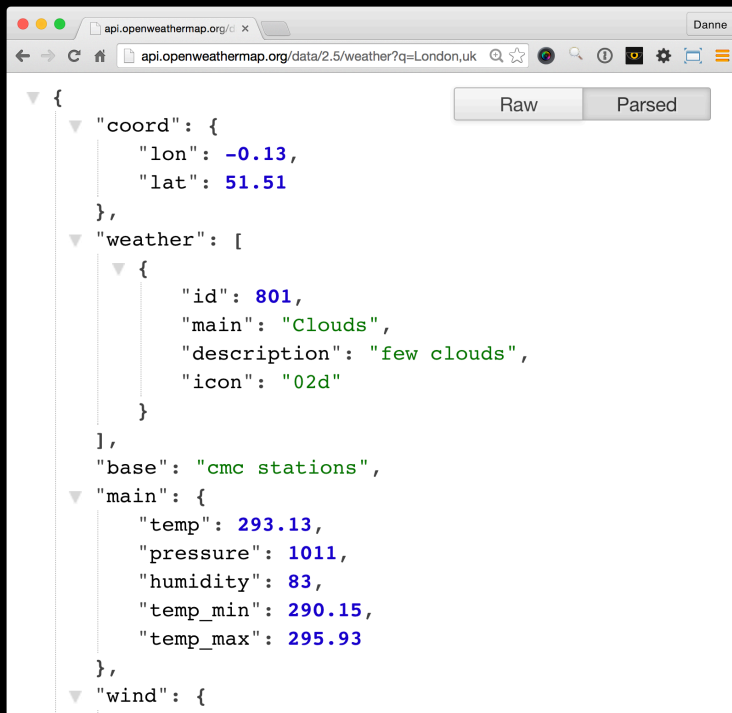
openweathermap.org/api
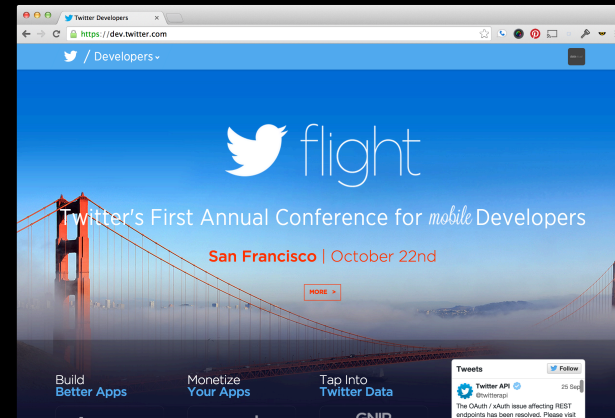
dev.twitter.com

Sourcing Data

Data Formats - Nested

JSON – JavaScript Object Notation

XML – Extensible Markup Language

Sourcing Data

APIs (Application Programming Interfaces)

10,000+ Apps

Using Foursquares

Location Data API

Sourcing Data

APIs (Application Programming Interfaces)
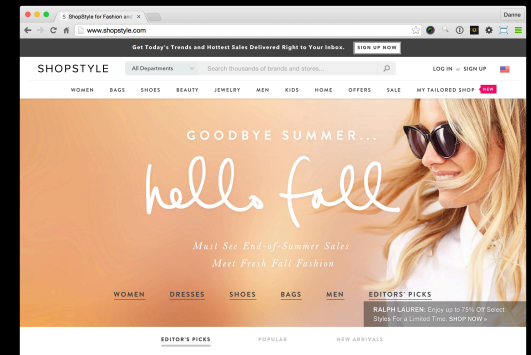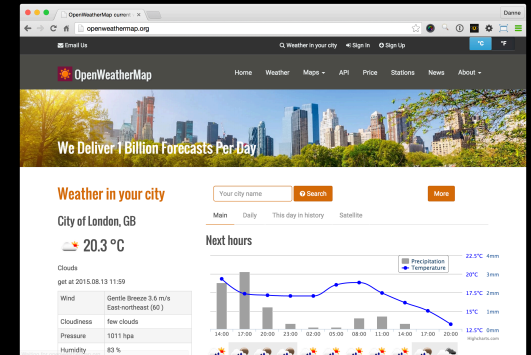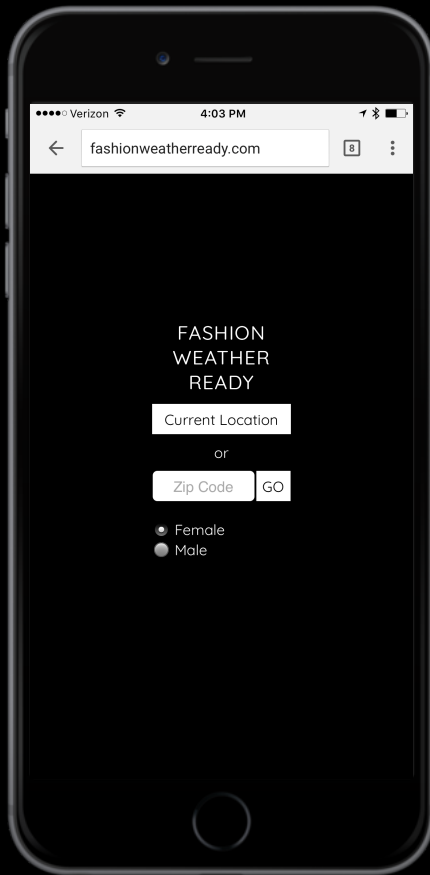
Big Data

John Mashey – 1998

Big Data

More Storage Capabilities
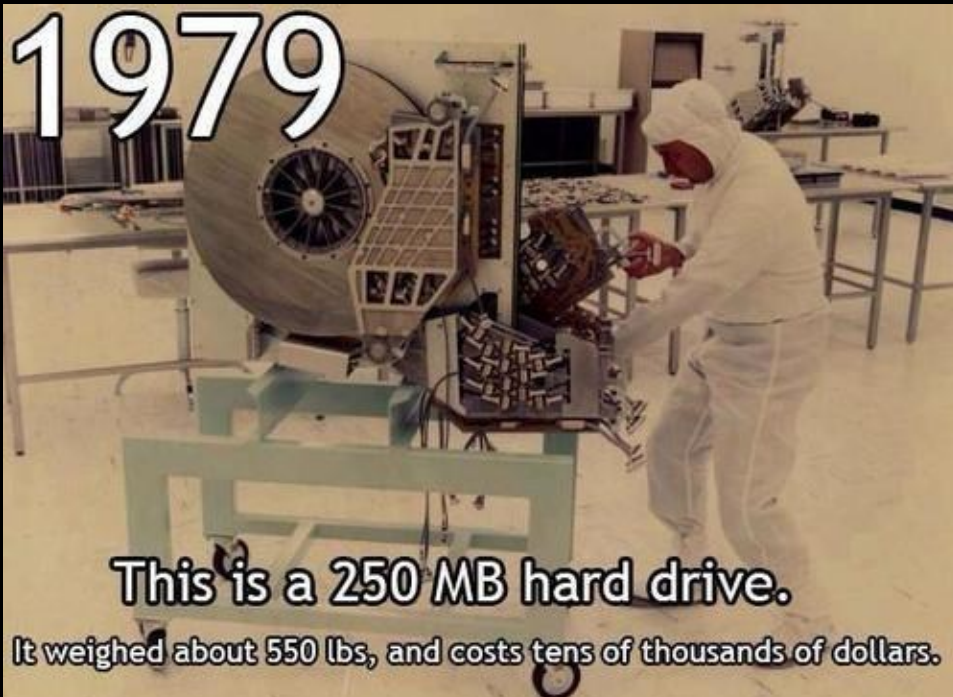
Big Data

Needs more than one computer

In Class Assignment

In groups of two, find a story in the news of interest to you and use that topic to find an interesting open data set. Present it to the class giving your reasons why you chose it and what interesting data points you noticed in it.

This can be a website or open data set, lets avoid APIs for now.

Tip: Google the topic + "statistics" or "data".

Homework

Continue collecting your own personal data which you can use in your final project.

Some examples could be location data, exercise data, type of food you eat, who you interact with, what time you wake up and go to sleep, your mood for the day, etc.

Here are a few tools to help you do this:
Gyroscope tracking location data, health, mood, etc.
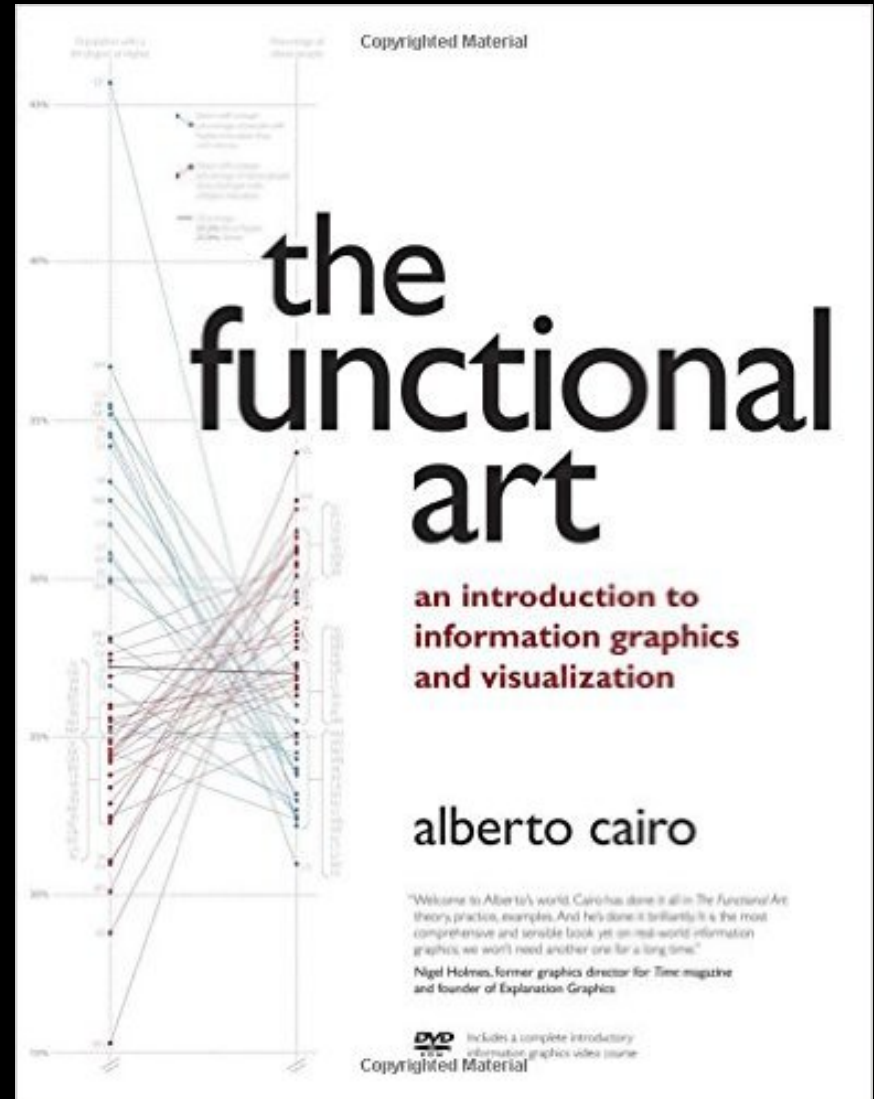**Reporter App** simple diary app created by Nicholas Felton.
iPhone Health track steps, calories burned, goals, etc.
Rescue Time tracks everything you do on your computer
And of course devices like the **Apple Watch, FitBit, Nike Fuelband**, etc.

Homework

Read pages 5-19 in
The Functional Art

Information Design

Professor Danne Woo

dwoo@qc.cuny.edu

infodesign.dannewoo.com