Interactive Data Visualization

Danne Woo and Peter Darche

danne@dannewoo.com
@dannewoo
www.dannewoo.com

pdarche@gmail.com
@pdarche
www.peterdarche.com

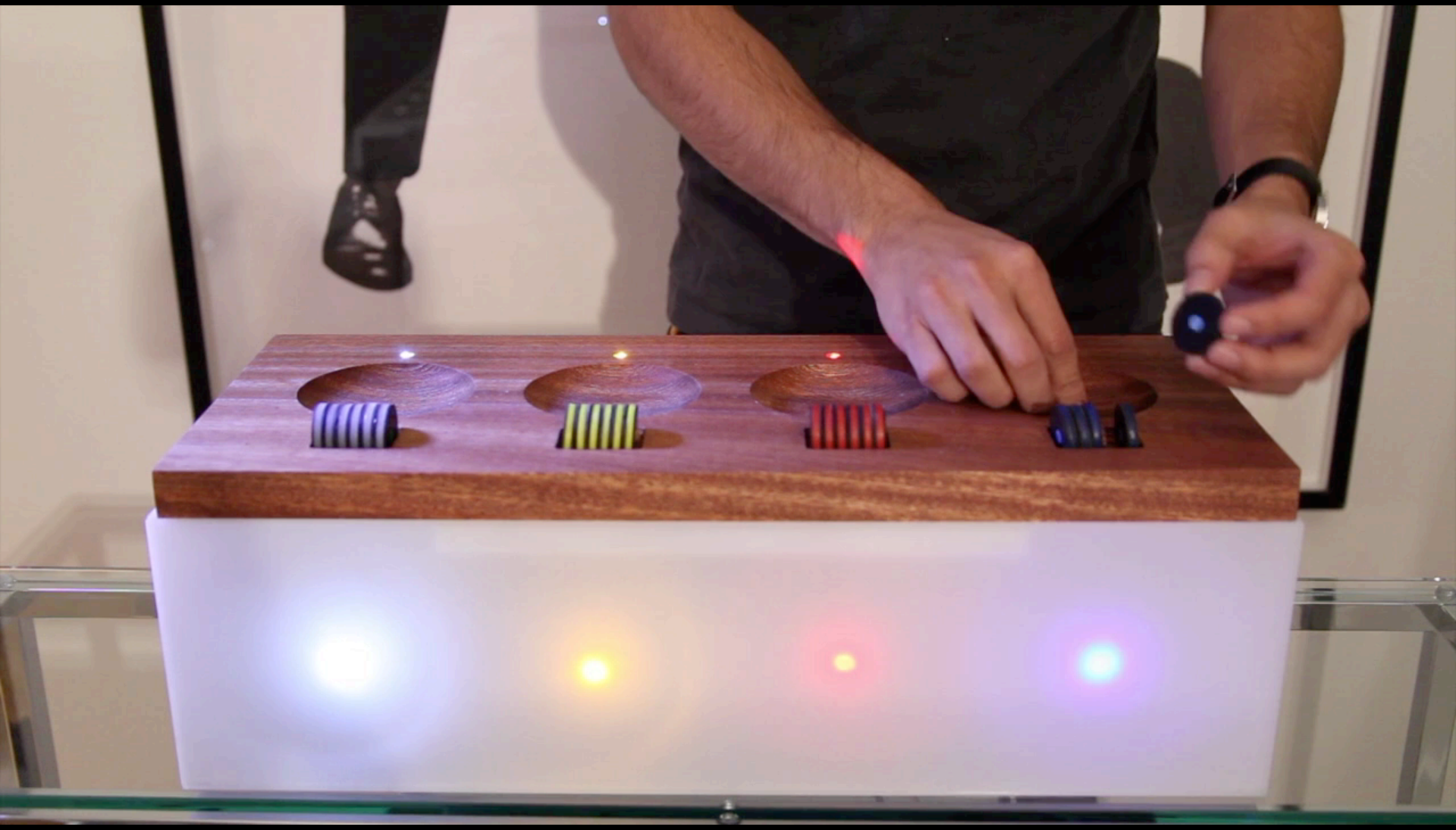
Danne Woo

Design Technologist

Interaction Design Professor

Software Designer/Developer


SPLAT
Join team two at splat.in/2
POOP



Peter Darche

Data Scientist

Programmer

Software Designer/Developer


itp
NYU
TISCH


datavisu.al
datavisu.al/beta/home.html
datavisual
Data Designer
Community Gallery
About Us
Hey danne, Log Out
Visualizing your data your way
Design
Data
Output
What is datavisual?
datavisual is a web-based tool for creating data visualizations. Perfect for designers and creatives with limited programming skills but a need to make beautiful charts and graphs. Once you design your chart style you can upload multiple data sets and output them as vector PDFs or SVGs as well as websafe PNGs.
MacBook Pro


datavisual

Your turn.

Who are you?

Where are you from?

Why did this class interest you?

What type of data interests you most?

Course Structure

Day 1 Data Sourcing, Scrubbing and Analysis

Day 2 Designing Data

Day 3 + 4 Data, d3.js and the Web

Day 5 Live Data and APIs

Project Brief

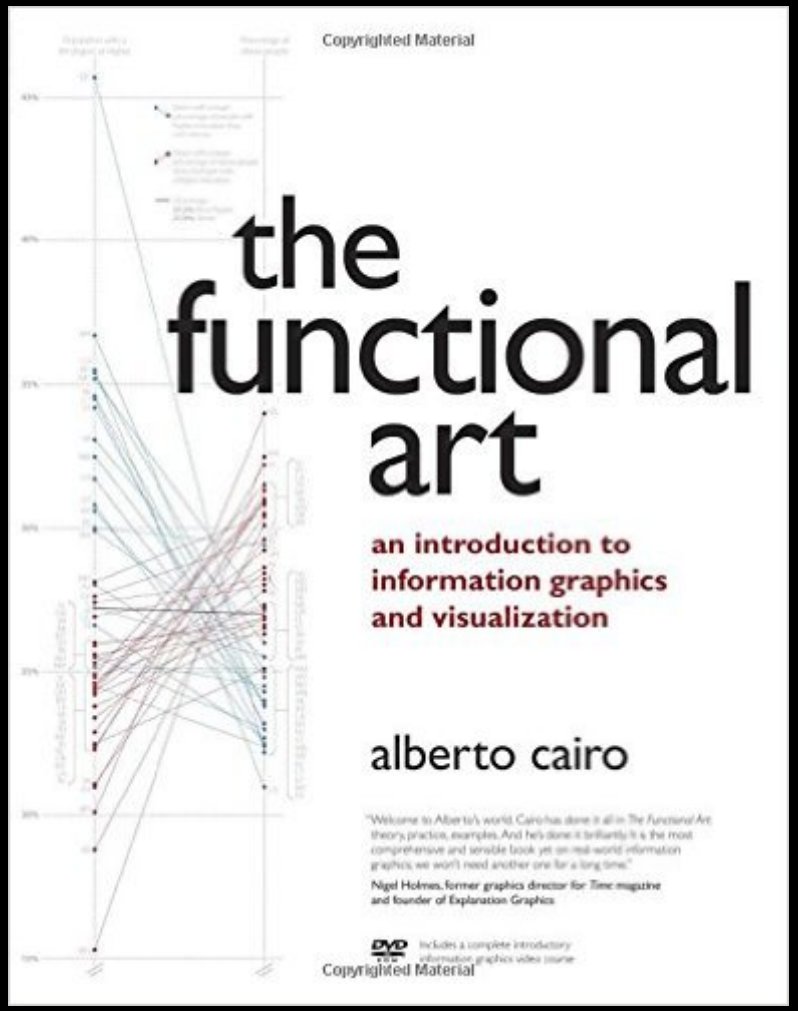
Design and build and interactive data visualization or information graphic that visually clarifies our reliance on electricity, how we generate power, how we use it and/or its effects on our planet. Formulate a question around this topic, use this question as your basis for you research, collect data from reputable sources to support your answer to this question and create a interactive visualization using this data.

Suggested Readings

The Functional Art

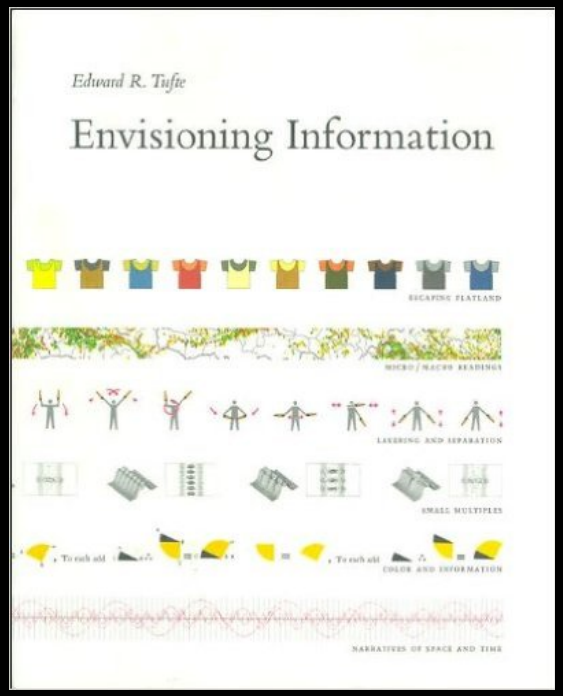
By Alberto Cairo

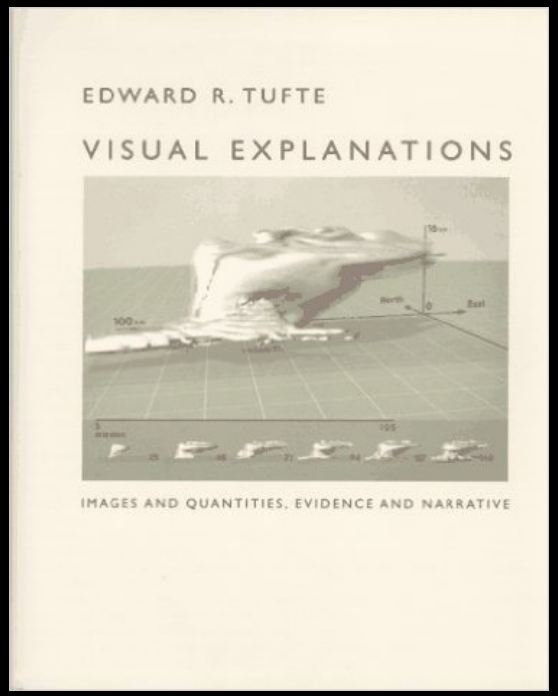
www.thefunctionalart.com

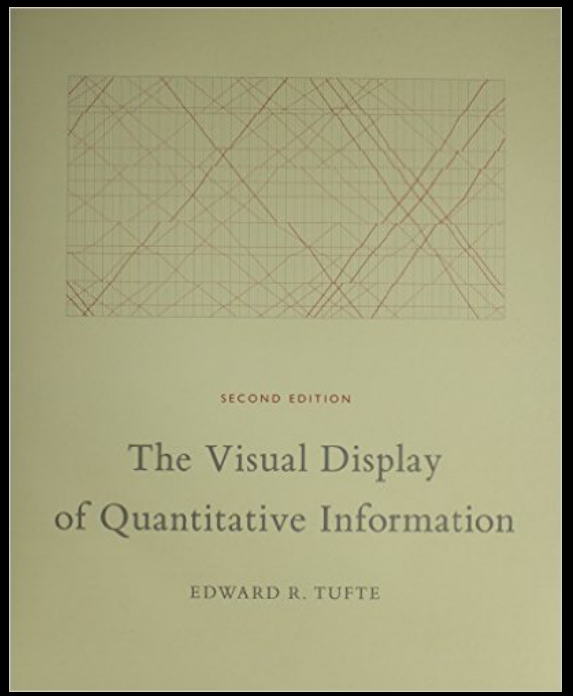


Suggested Readings

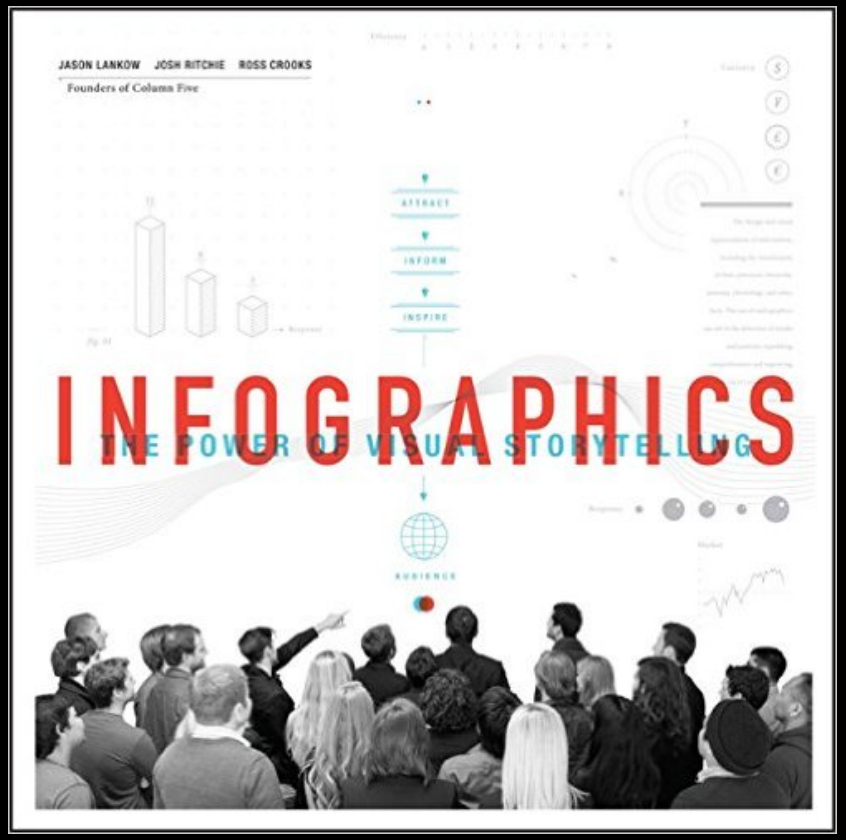
Edward Tufte Books

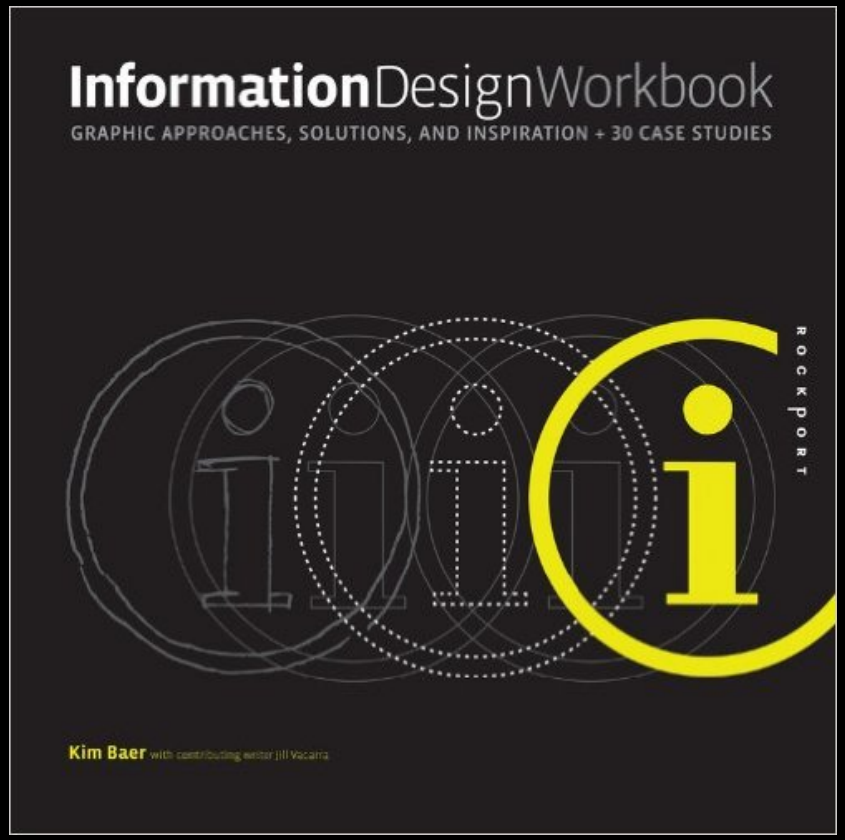






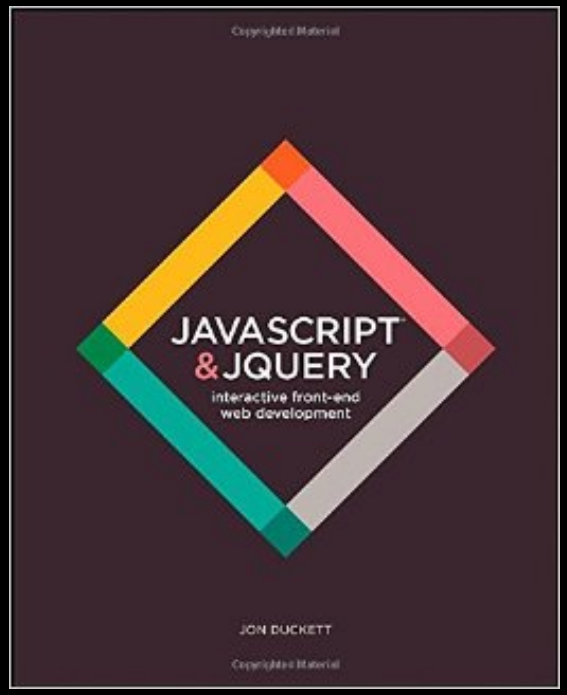
Suggested Readings

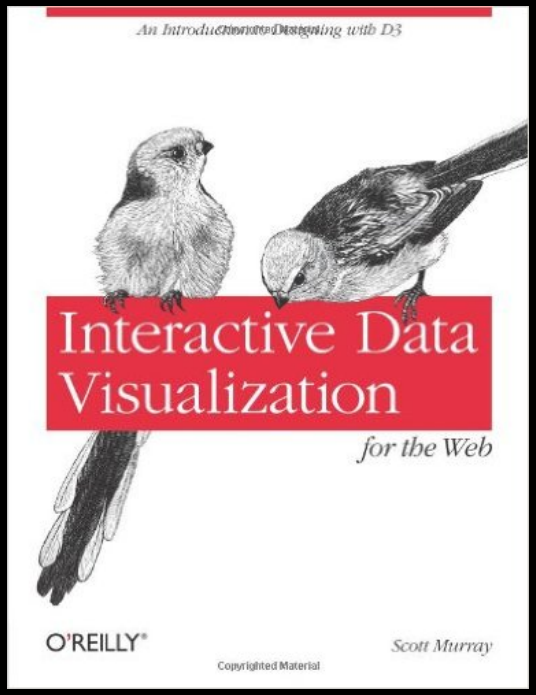




Suggested Readings

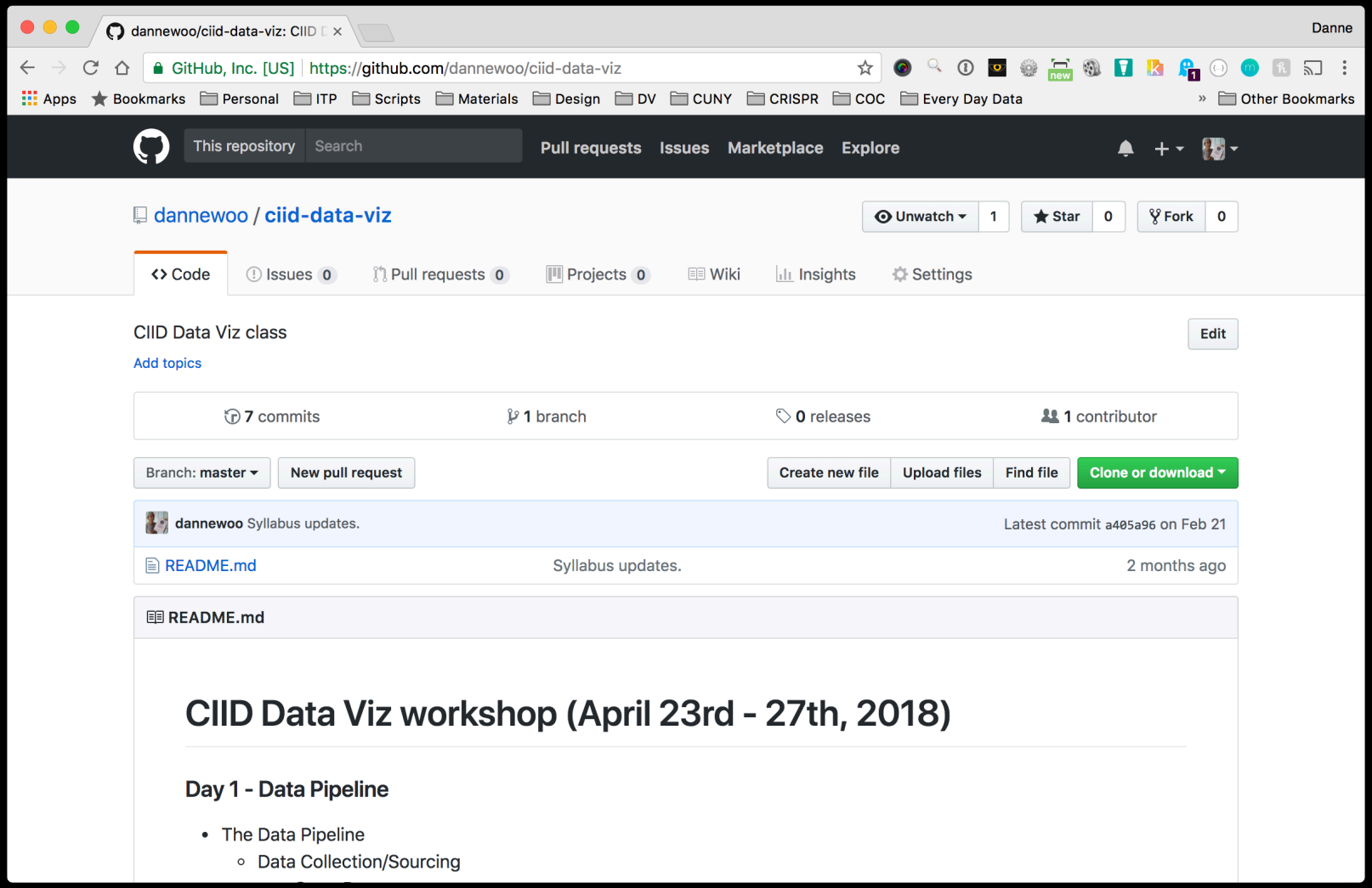
Web Development





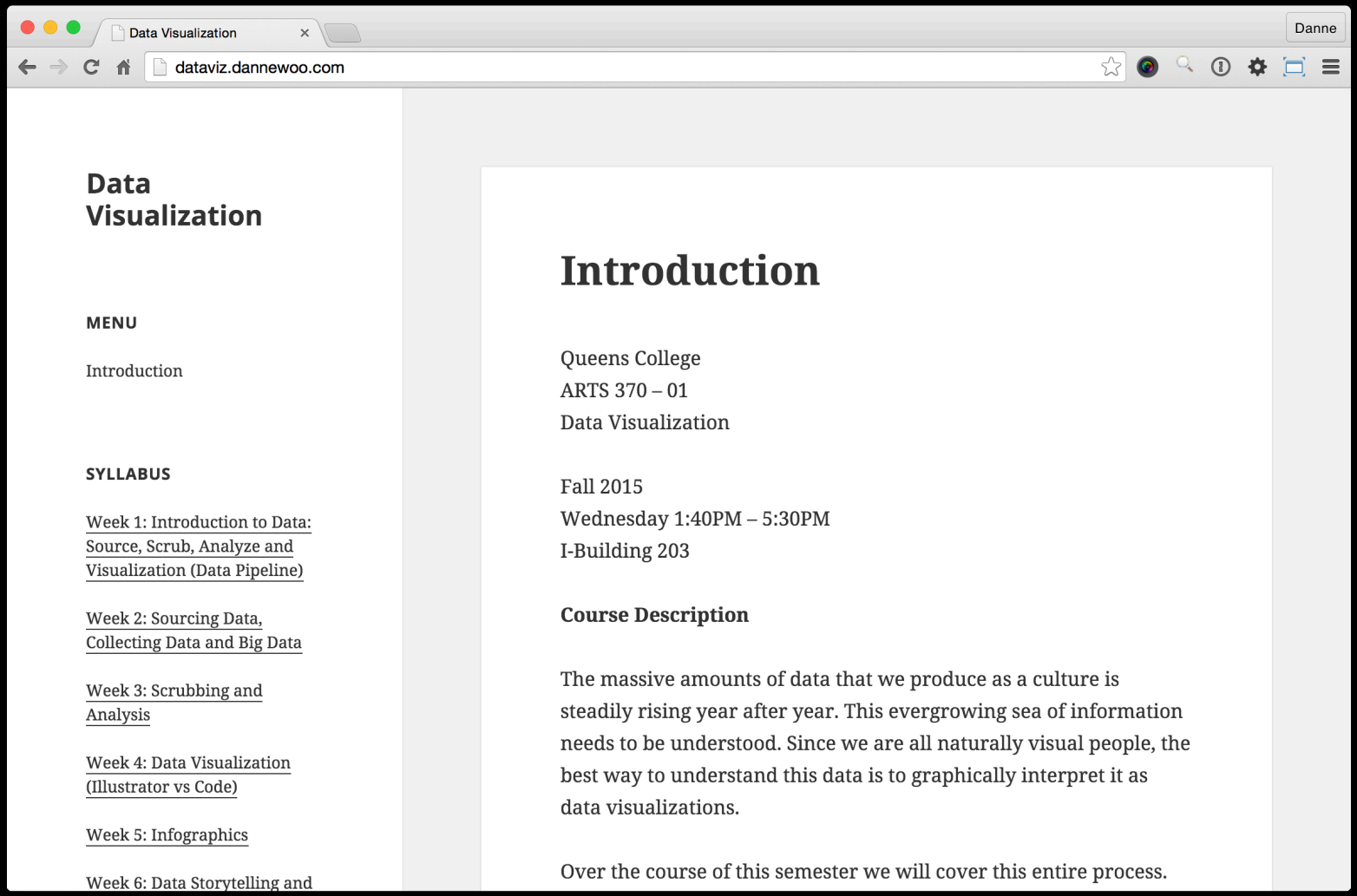
Git Repo

github.com/dannewoo/ciid-data-viz

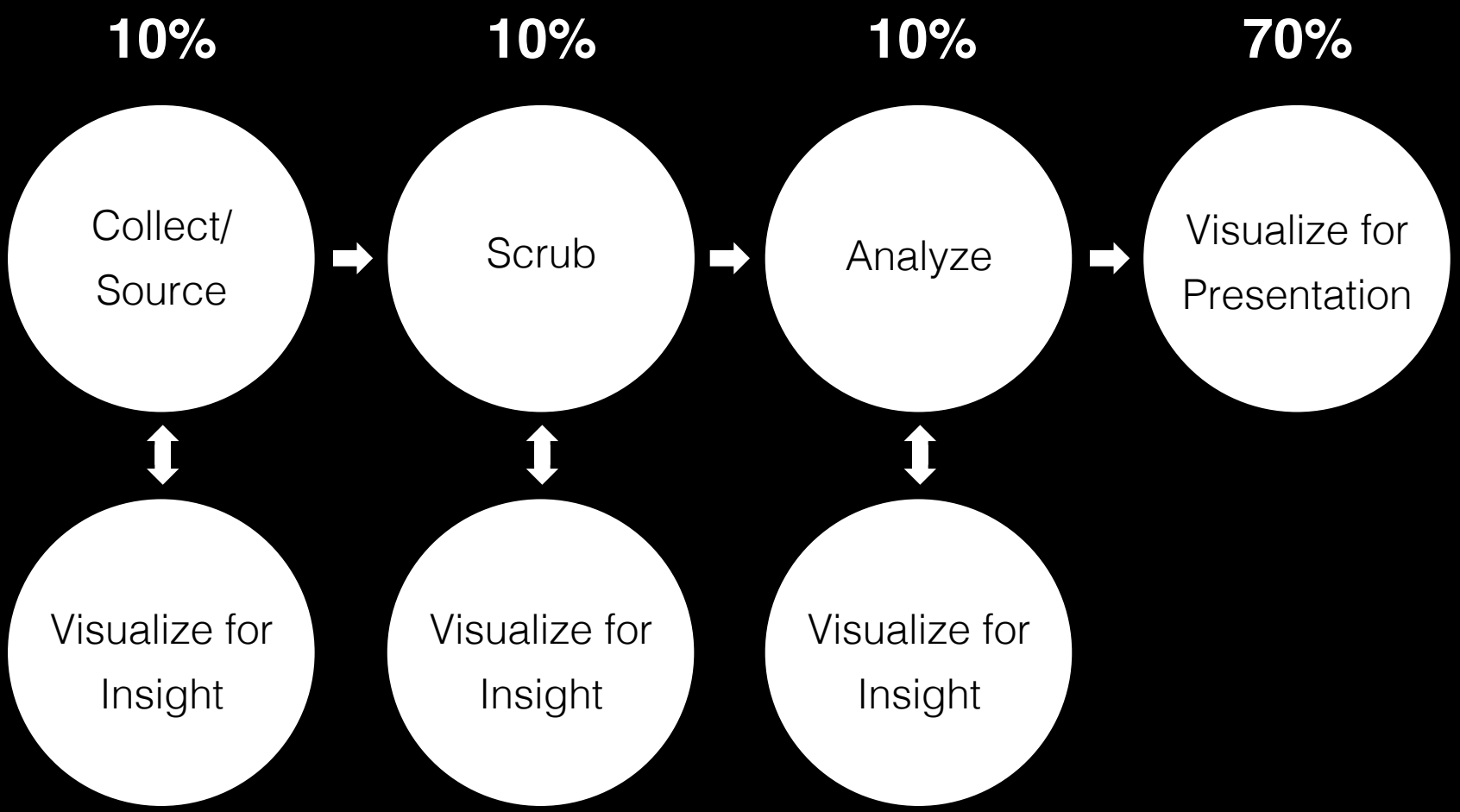


14-Week Class

dataviz.dannewoo.com



Data Pipeline



What is Data?

A collection of facts.

Numbers
Words
Measurements
Descriptions

Observations
Pictures
Videos
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

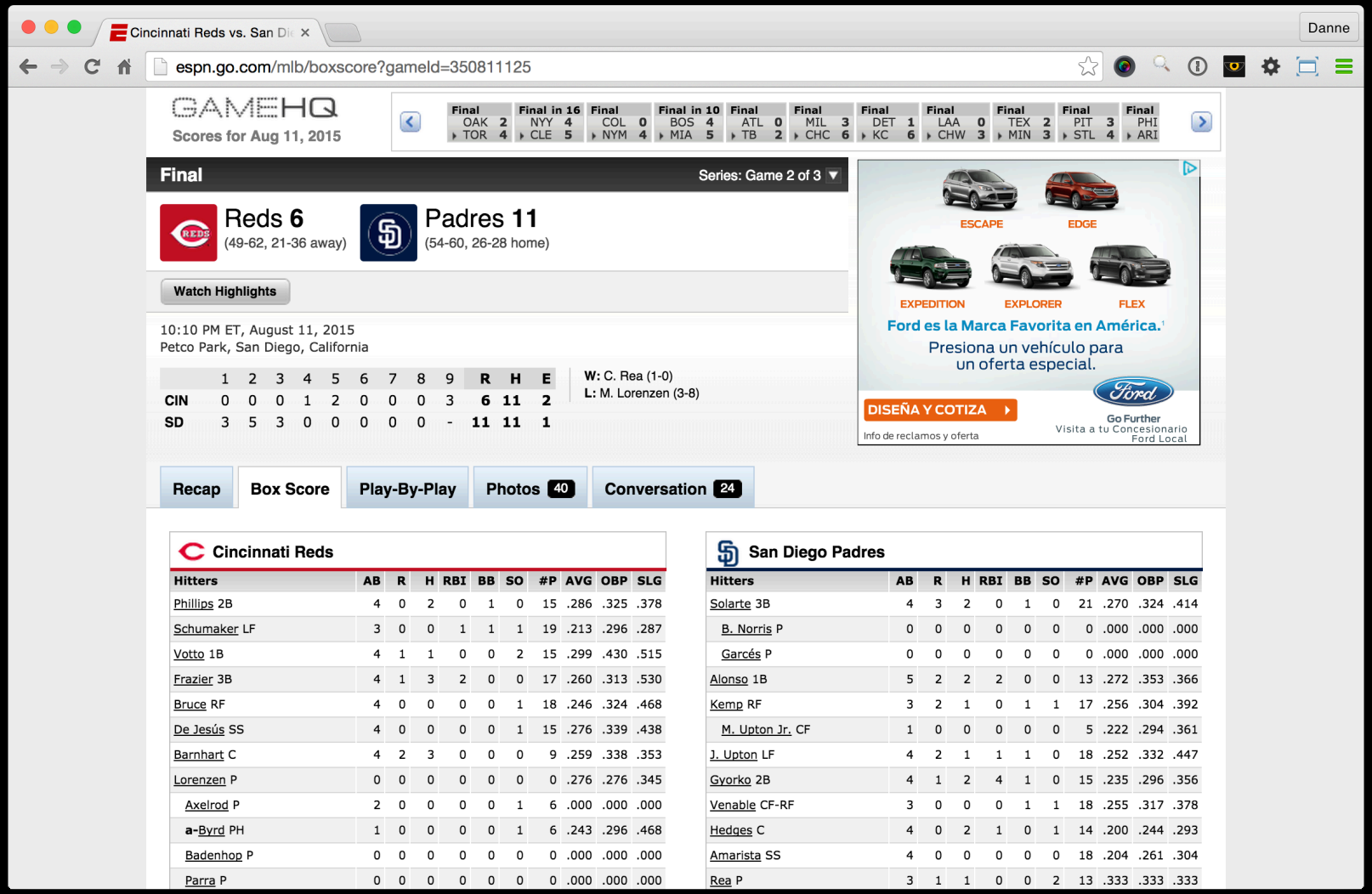
Oh-oh, you are

My sunrise on the darkest day

Got me feelin' some kind of way

Data

Structured or Unstructured?



	1	2	3	4	5	6	7	8	9	R	H	E
CIN	0	0	0	1	2	0	0	0	3	6	11	2
SD	3	5	3	0	0	0	0	0	-	11	11	1

Hitters	AB	R	H	RBI	BB	SO	#P	AVG	OBP	SLG
Phillips 2B	4	0	2	0	1	0	15	.286	.325	.378
Schumaker LF	3	0	0	1	1	1	19	.213	.296	.287
Votto 1B	4	1	1	0	0	2	15	.299	.430	.515
Frazier 3B	4	1	3	2	0	0	17	.260	.313	.530
Bruce RF	4	0	0	0	0	1	18	.246	.324	.468
De Jesús SS	4	0	0	0	0	1	15	.276	.339	.438
Barnhart C	4	2	3	0	0	0	9	.259	.338	.353
Lorenzen P	0	0	0	0	0	0	0	.276	.276	.345
Axelrod P	2	0	0	0	0	1	6	.000	.000	.000
a-Byrd PH	1	0	0	0	0	1	6	.243	.296	.468
Badenhop P	0	0	0	0	0	0	0	.000	.000	.000
Parra P	0	0	0	0	0	0	0	.000	.000	.000

Hitters	AB	R	H	RBI	BB	SO	#P	AVG	OBP	SLG
Solarte 3B	4	3	2	0	1	0	21	.270	.324	.414
B. Norris P	0	0	0	0	0	0	0	.000	.000	.000
Garcés P	0	0	0	0	0	0	0	.000	.000	.000
Alonso 1B	5	2	2	2	0	0	13	.272	.353	.366
Kemp RF	3	2	1	0	1	1	17	.256	.304	.392
M. Upton Jr. CF	1	0	0	0	0	0	5	.222	.294	.361
J. Upton LF	4	2	1	1	1	0	18	.252	.332	.447
Gyorko 2B	4	1	2	4	1	0	15	.235	.296	.356
Venable CF-RF	3	0	0	0	1	1	18	.255	.317	.378
Hedges C	4	0	2	1	0	1	14	.200	.244	.293
Amarista SS	4	0	0	0	0	0	18	.204	.261	.304
Rea P	3	1	1	0	0	2	13	.333	.333	.333

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 xv7csp2952801bb; 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

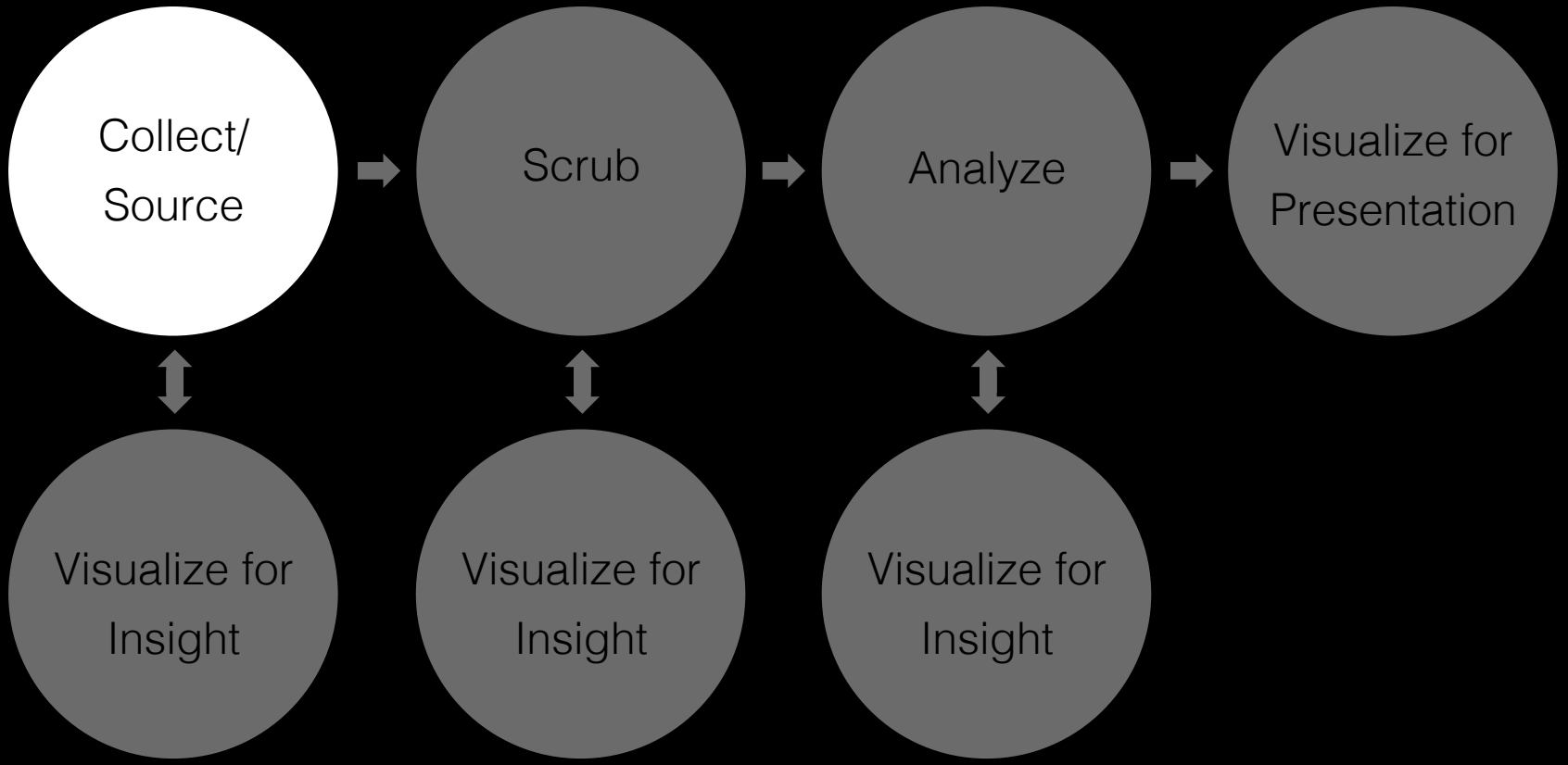
Structured or Unstructured?

Basic Image Information

Camera:	Apple iPhone 4S
Lens:	4.3 mm
Exposure:	Auto exposure, Program AE, 1/20 sec, f/2.4, ISO 125
Flash:	Off, Did not fire
Date:	**December 3, 2012** 12:26:08PM (timezone not specified) (2 hours, 44 minutes, 59 seconds ago, assuming image timezone of 8 hours behind GMT)
Location:	Latitude/longitude: **15°** 39' 29.4" North, **88°** 59' 31.8" West (15.658167, -88.992167) Photos on Jeffrey's blog that are near this location. Map via embedded coordinates at: Google, Yahoo, WikiMapia, OpenStreetMap, Bing (also see the Google Maps pane below) Altitude: 7.152159468 m Timezone guess from earthtools.org: 6 hours behind GMT
File:	**480 × 640** JPEG 132,481 bytes (0.13 megabytes) Image compression: 88% 4% crop of the 3,264 × 2,448 (8.0 megapixel) original
Color Encoding:	**WARNING:** Color space tagged as sRGB, without an embedded color profile. **Windows and Mac browsers and apps treat the colors randomly.** Images for the web are most widly viewable when in the sRGB color space and with an embedded color profile. See my Introduction to Digital-Image Color Spaces for more information.
Image URL:	http://assets.vice.com/content-images/contentimage/no-slug/8b4cd78e0012f77547dc621f5498b873.jpg Apply other tools to this image via ImgOps.com.

Main image displayed here at 70% width (49% the area of the original)

Data Pipeline

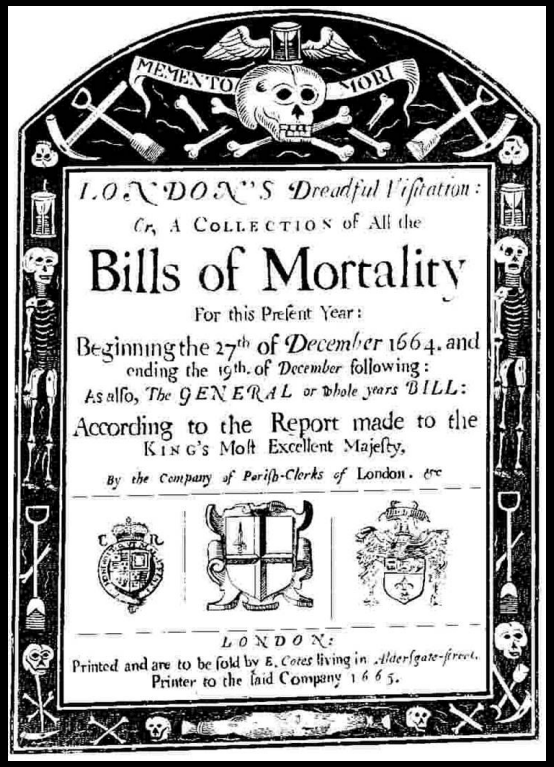


Collecting Data

Blombos Ocher Plaque – 75000BC

Collecting Data

John Graunt – 1662


MEMENTO MORI

LONDON'S Dreadful Visitation:

Or, A COLLECTION of All the

Bills of Mortality

For this Present Year:

Beginning the 27th of December 1664. and ending the 19th. of December following:

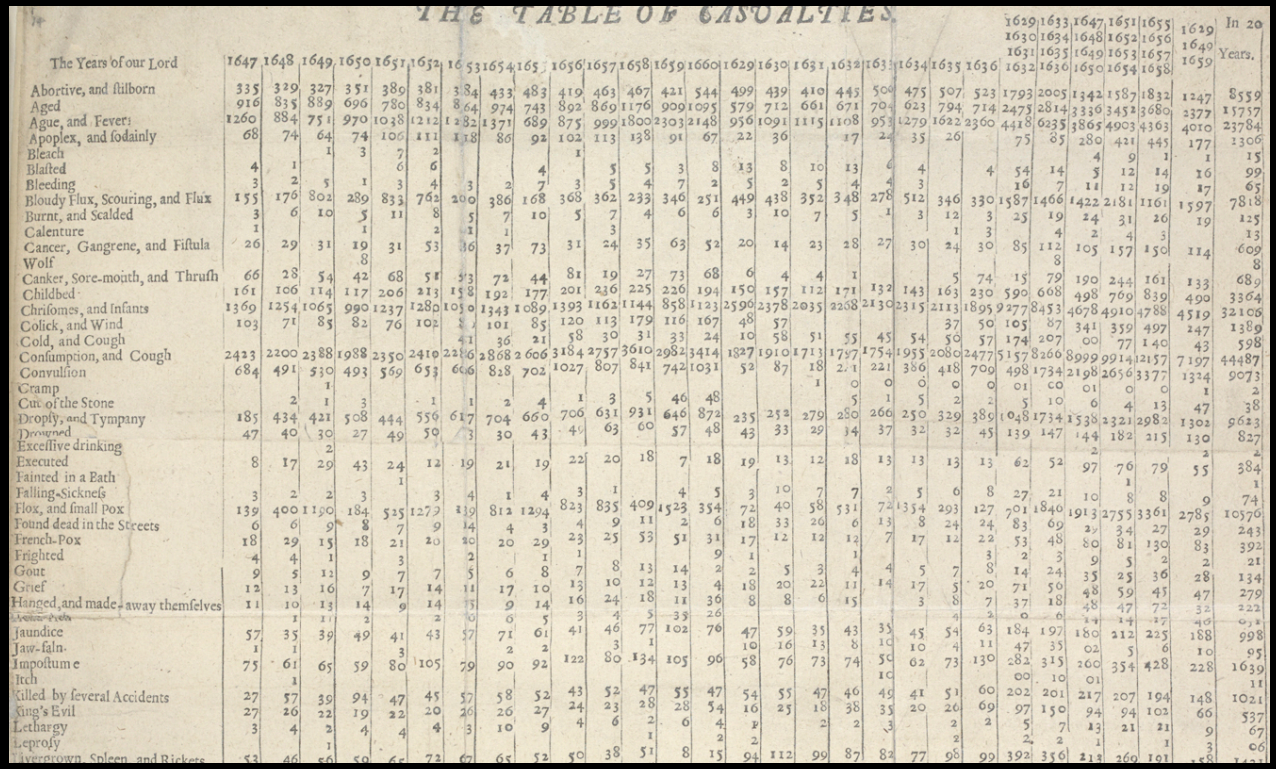
As also, The GENERAL or Whole years BILL:

According to the Report made to the KING'S Most Excellent Majesty,

By the Company of Parish-Clerks of London. &c

LONDON:

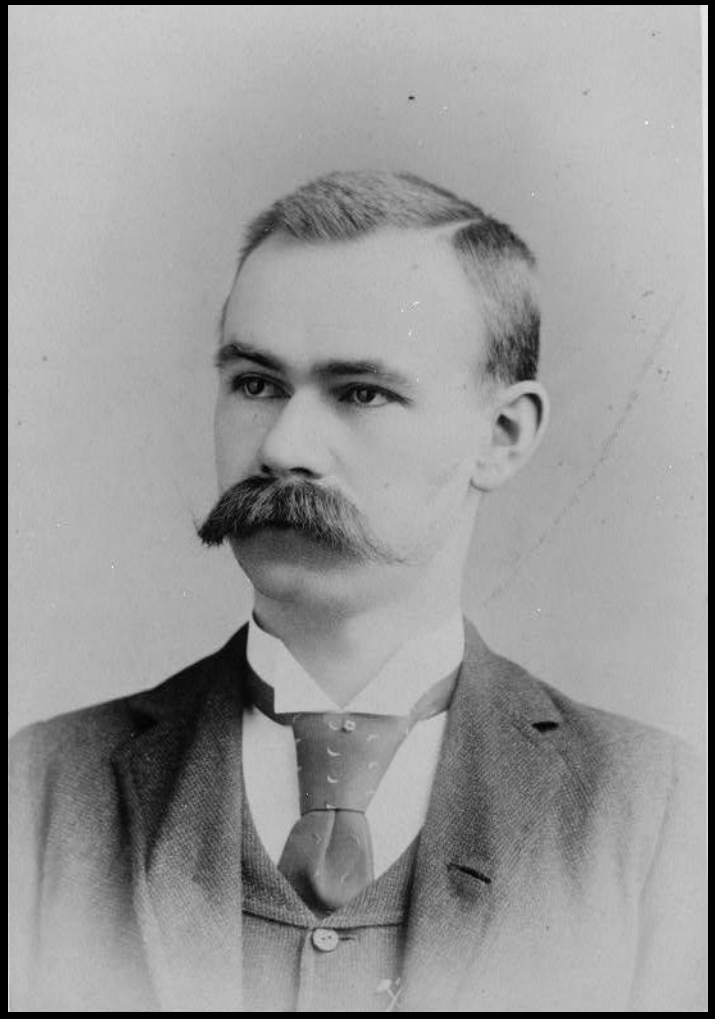
Printed and are to be sold by E. Cotes living in Aldersgate-street. Printer to the said Company 1665.


THE TABLE OF CASUALTIES

The Years of our Lord	1647	1648	1649	1650	1651	1652	1653	1654	165	1656	1657	1658	1659	1660	1629 1630	1631	1632	163	1634	1635	1636	1629 1630 1631 1632	1633 1634 1635 1636	1647 1648 1649 1650	1651 1652 1653 1654	1655 1656 1657 1658	1629 1649 1659	In 20 Years.
Abortive, and ſtillborn	335	329	327	351	389	381	384	433	483	419	463	467	421	544	499	439	410	445	500	475	507	523	1793	2005	1342	1587	1832	1247	8559
Aged	916	835	889	696	780	834	864	974	743	892	869	1176	909	1095	579	712	661	671	704	623	794	714	2475	2814	3336	3452	3680	2377	15757
Ague, and Fever:	1260	884	751	970	1038	1212	1282	1371	689	875	999	1800	2303	2148	956	1091	1115	1108	953	1279	1622	2360	4418	6235	3865	4903	4363	4010	23784
Apoplex, and ſodainly	68	74	64	74	106	111	118	86	92	102	113	138	91	67	22	36		17	24	35	26		75	85	280	421	445	177	1306
Bleach			1	3	7	2				1															4	9	1	1	15
Blaſted	4	1			6				4		5	5	3	8	13	8	10	13	6	4		4	54	14	5	12	14	16	99
Bleeding	3	2	5	1	3	4	3	2	7	3	5	4	7	2	5	2	5	4	4	3			16	7	11	12	19	17	65
Bloudy Flux, Scouring, and Flux	155	176	802	289	833	762	200	386	168	368	362	233	346	251	449	438	352	348	278	512	346	330	1587	1466	1422	2181	1161	1597	7818
Burnt, and Scalded	3	6	10	5	11	8	5	7	10	5	7	4	6	6	3	10	7	5	1	3	12	3	25	19	24	31	26	19	125
Calenture	1			1		2	1	1			3										1	3		4	2	4	3		13
Cancer, Gangrene, and Fiſtula	26	29	31	19	31	53	36	37	73	31	24	35	63	52	20	14	23	28	27	30	24	30	85	112	105	157	150	114	609
Wolf				8																				8					8
Canker, Sore-mouth, and Thruſh	66	28	54	42	68	51	53	72	44	81	19	27	73	68	6	4	4	1			5	74	15	79	190	244	161	133	689
Childbed	161	106	114	117	206	213	158	192	177	201	236	225	226	194	150	157	112	171	132	143	163	230	590	608	498	709	839	490	3364
Chriſomes, and Infants	1369	1254	1065	990	1237	1280	1030	1343	1089	1393	1162	1144	858	1123	2596	2378	2035	2268	2130	2315	2113	1895	9277	8453	4678	4910	4788	4519	32106
Colick, and Wind	103	71	85	82	76	102	82	101	85	120	113	179	116	167	48	57					37	50	105	87	341	359	497	247	1389
Cold, and Cough							41	36	21	58	30	31	33	24	10	58	51	55	45	54	50	57	174	207	00	77	140	43	598
Conſumption, and Cough	2423	2200	2388	1988	2350	2410	2216	2868	2606	3184	2757	3610	2982	3414	1827	1910	1713	1797	1754	1955	2080	2477	5157	8260	8999	9914	12157	7197	44487
Convulſion	684	491	530	493	569	653	666	828	702	1027	807	841	742	1031	52	87	18	21	221	386	418	700	498	1734	2198	2656	3377	1324	9073
Cramp			1														1	0	0	0	0	0	01	00	01	0	0	1	2
Cut of the Stone		2	1	3		1	1	2	4	1	3	5	46	48				5	1	5	2	2	5	10	6	4	13	47	38
Dropſy, and Tympany	185	434	421	508	444	556	617	704	660	706	631	931	646	872	235	252	279	280	266	250	329	389	1048	1734	1538	2321	2982	1302	9623
Drowned	47	40	30	27	49	50	53	30	43	49	63	60	57	48	43	33	29	14	37	32	32	45	139	147	144	182	215	130	827
Exceſſive drinking			2																						2			2	4
Executed	8	17	29	43	24	12	19	21	19	22	20	18	7	18	19	13	12	18	13	13	13	13	62	52	97	76	79	55	384
Fainted in a Bath					1																					1			1
Falling-Sickneſs	3	2	2	3		3	4	1	4	3	1		4	5	3	10	7	7	2	5	6	8	27	21	10	8	8	9	74
Flox, and ſmall Pox	139	400	1130	184	525	1279	239	812	1294	823	835	409	1523	354	72	40	58	531	72	1354	293	127	701	1840	1913	2755	3361	2785	10576
Found dead in the Streets	6	6	9	8	7	9	14	4	3	4	9	11	2	6	18	33	26	6	13	8	24	24	83	69	29	34	27	29	243
French-Pox	18	29	15	18	21	20	20	20	29	23	25	53	51	31	17	12	12	12	7	17	12	22	53	48	80	81	130	83	392
Frighted	4	4	1		3		2		1					9	1							3	2	3	9	5	2	2	21
Gout	9	5	12	9	7	7	5	6	8	7	8	13	14	2	2	5	3	4	4	5	7	8	14	24	35	25	36	28	134
Grief	12	13	16	7	17	14	11	17	10	13	10	12	13	4	18	20	22	11	14	17	5	20	71	50	48	59	45	47	279
Hanged, and made away themſelves	11	10	13	14	9	14	13	9	14	3	4	5	38	26	8	8	6	15		3	8	7	37	18	48	47	72	32	222
Aged Poor		1	11	2		2	10	6	5	3	4	5	38	26								2	10	6	14	14	17	46	93
Jaundice	57	35	39	49	41	43	57	71	61	41	46	77	102	76	47	59	35	43	33	45	54	63	184	197	180	212	225	188	998
Jaw-faln	1	1			3			2	2		3	1			10	16	13	8	10	10	4	11	47	33	02	5	6	10	95
Impoſtume	75	61	65	59	80	105	79	90	92	122	80	134	105	96	58	76	73	74	50	62	73	130	282	315	200	354	428	228	1639
Itch		1																	10				00	10	01				11
Killed by ſeveral Accidents	27	57	39	94	47	45	57	58	52	43	52	47	55	47	54	55	47	46	49	41	51	60	202	201	217	207	194	148	1021
King's Evil	27	26	22	19	22	20	26	26	27	24	23	28	28	54	16	25	18	38	35	20	26	69	97	150	94	94	102	66	537
Lethargy	3	4	2	4	4	4	3	10	9	4	6	2	6	4	1		2	2	3		2	2	5	7	13	21	21	9	67
Leproſy			1									1		2	2						2		2	2	1		1	3	06
Lyergrown, Spleen, and Rickets	53	46	50	58	65	72	67	65	52	50	38	51	8	15	94	112	99	87	82	77	98	99	392	356	213	260	101	158	1420

Collecting Data

Herman Hollerith – 1890



Collecting Data

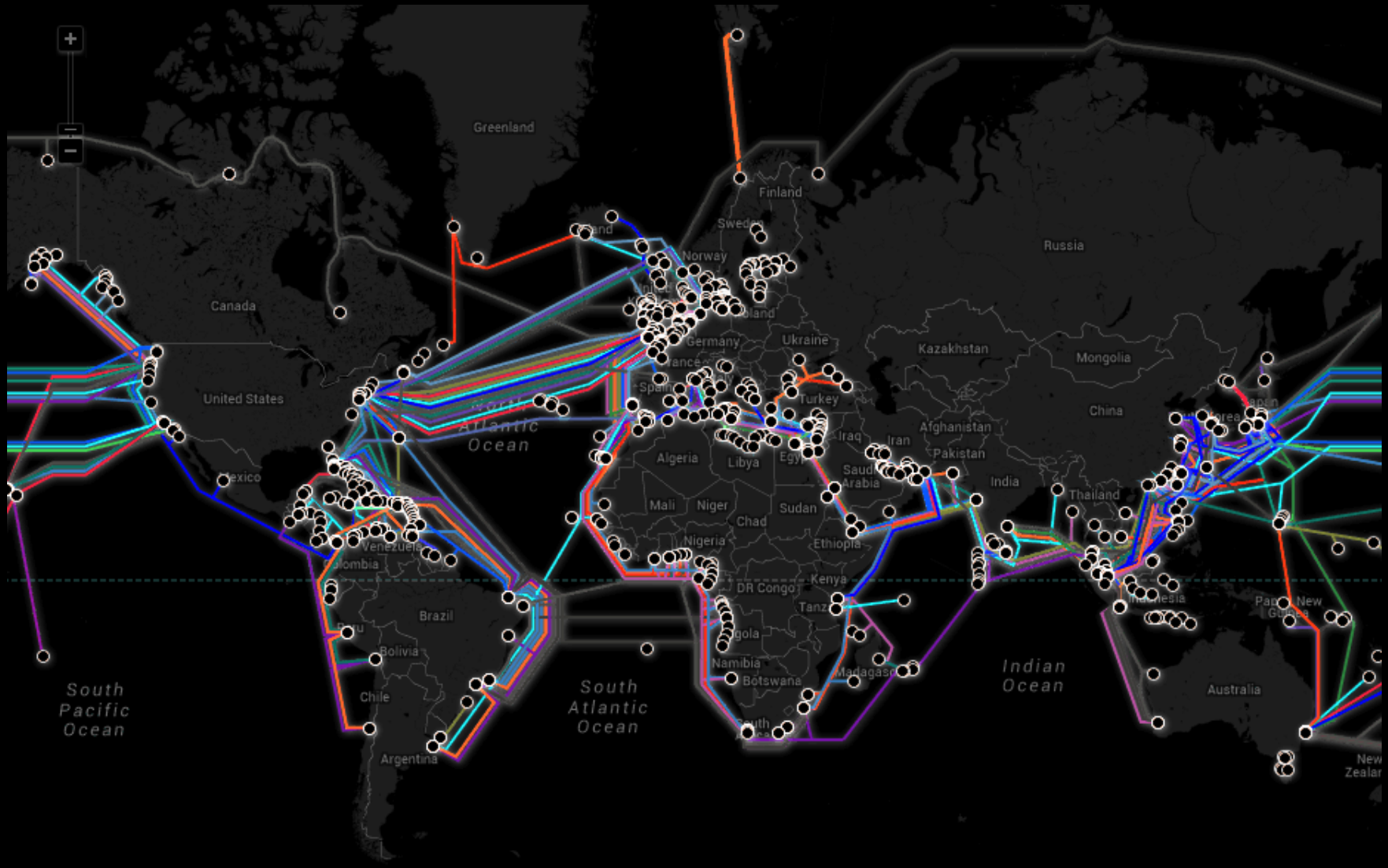
Jacquard Loom

Collecting Data

Herman Hollerith – 1890

Collecting Data

The Internet – 1982



Collecting Data

World Wide Web – 1989

Collecting Data

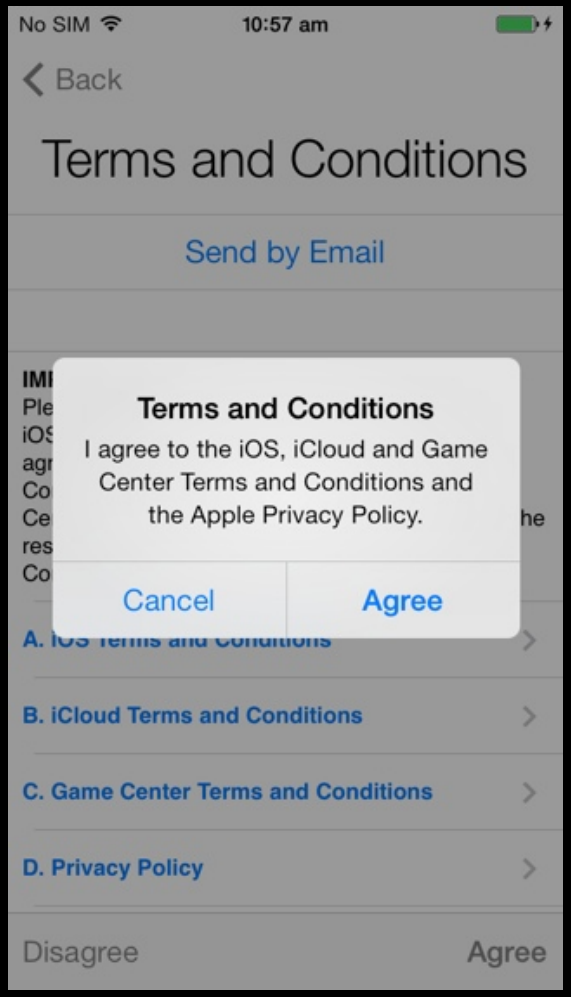
How the Web Works

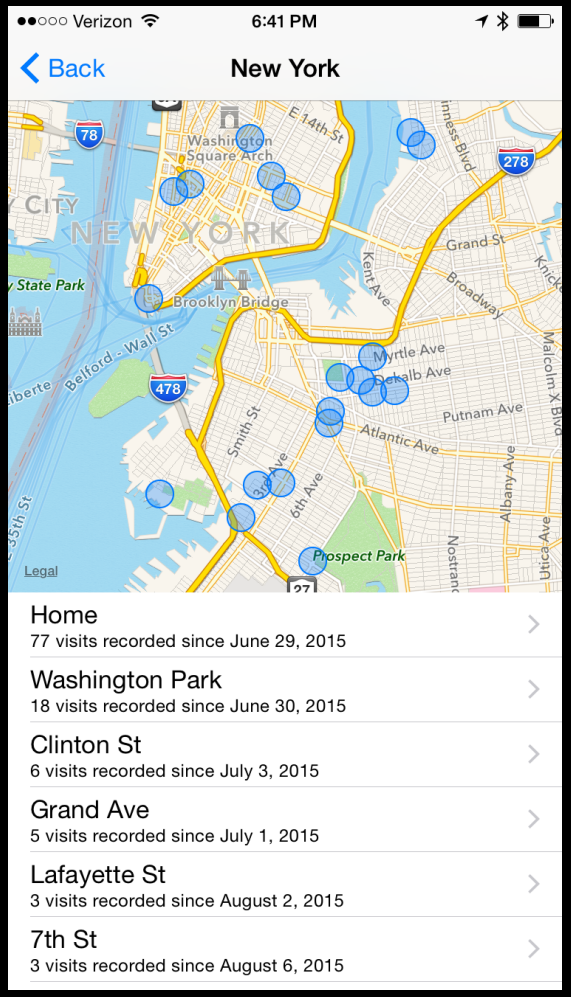
Collecting Data

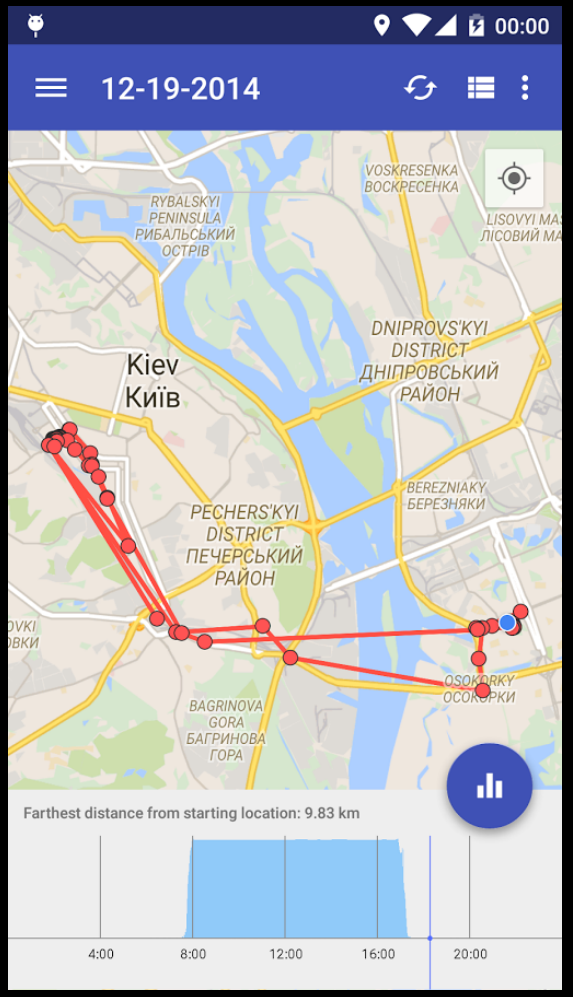
Every Company is a Data Company

Collecting Data

Every Company is a Data Company

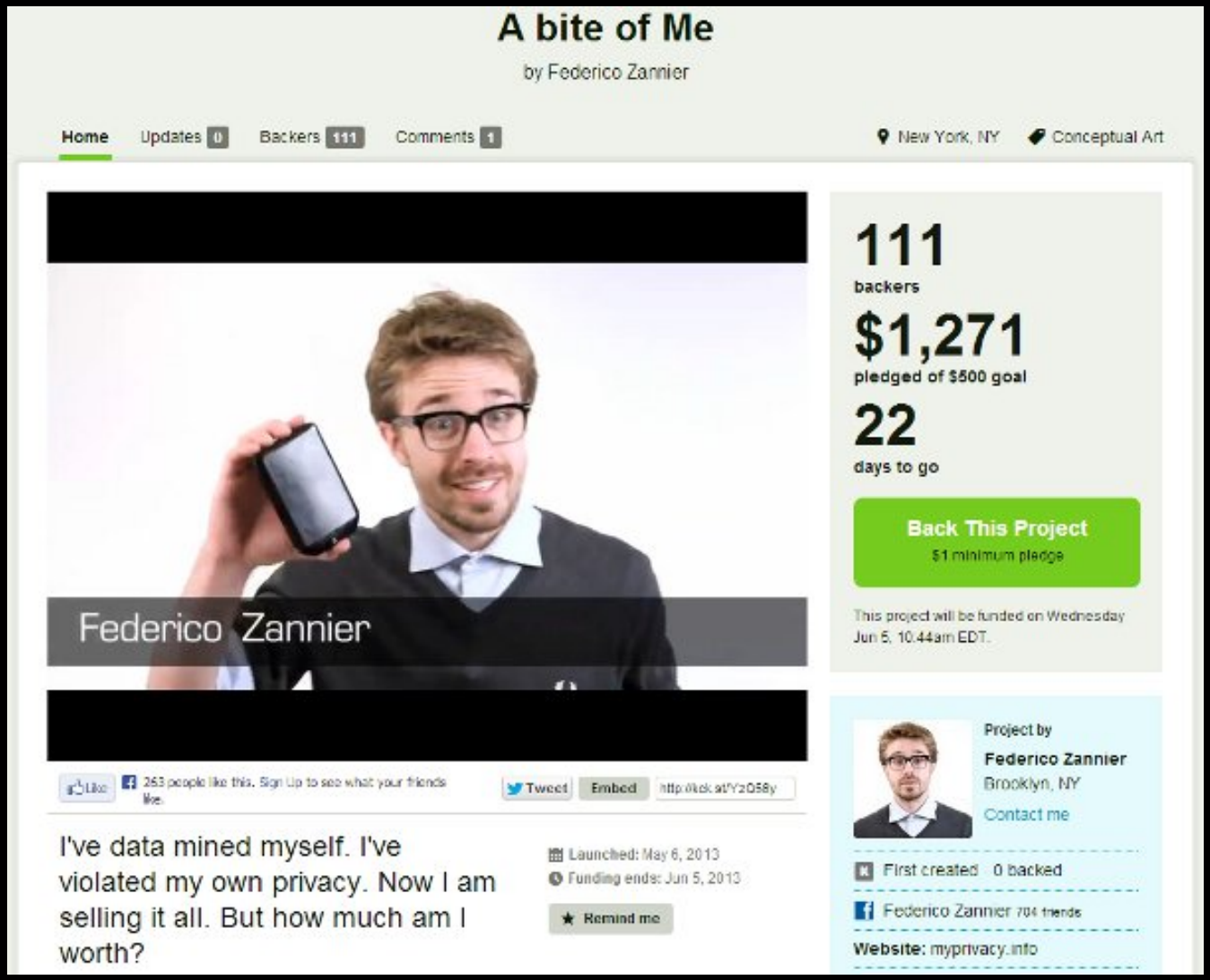






Collecting Data

Take Back Your Data

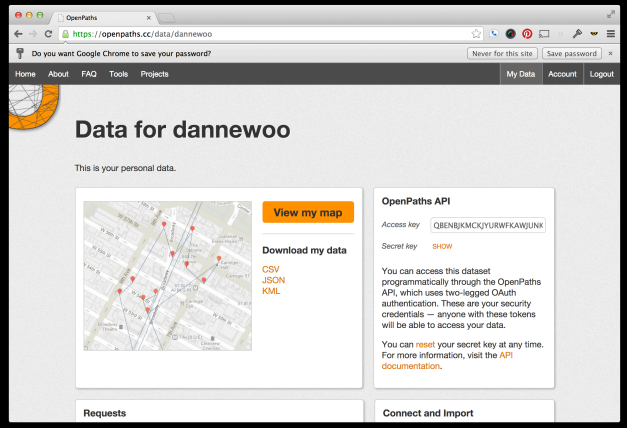


Collecting Data

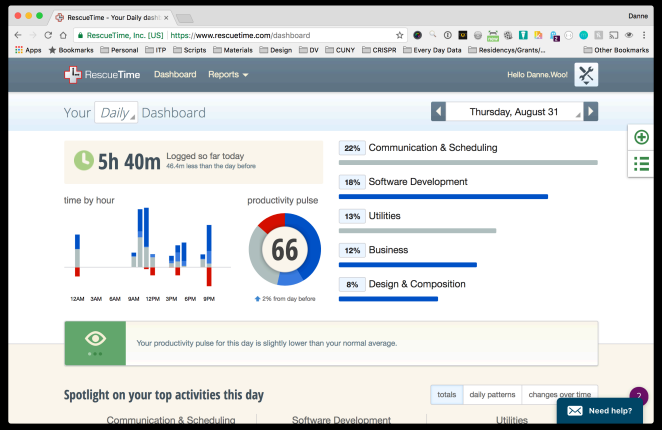
Nicholas Felton

Collecting Data

Personal Data – Software



Open Paths



Rescue Time

Feltron Reporter App

Apple iPhone Health Sensors

Collecting Data

Personal Data – Hardware

Fitbit

Jawbone

Nike Fuelband

Apple Watch

Collecting Data

Personal Data – Hardware

Muse Headband

Nest

Vessyl

Scio

Collecting Data

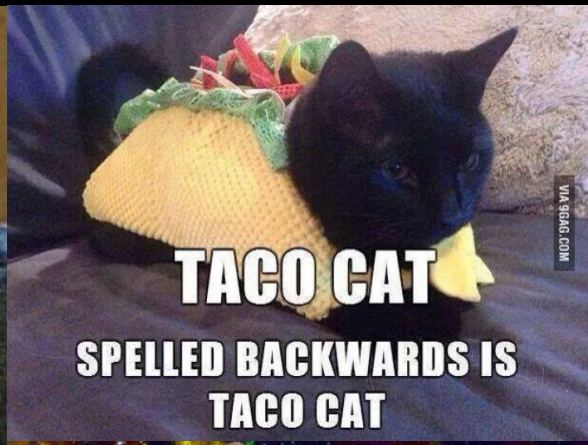
The Connected Us

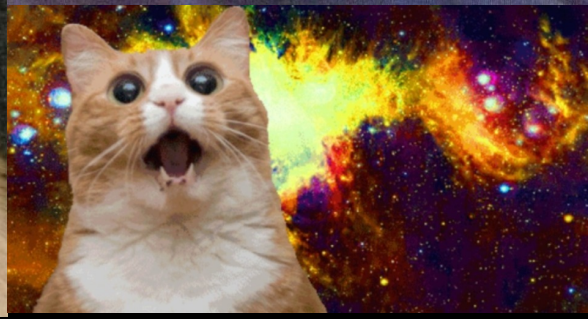
Collecting Data

Rodrigo Narciso – CH4

Sourcing Data

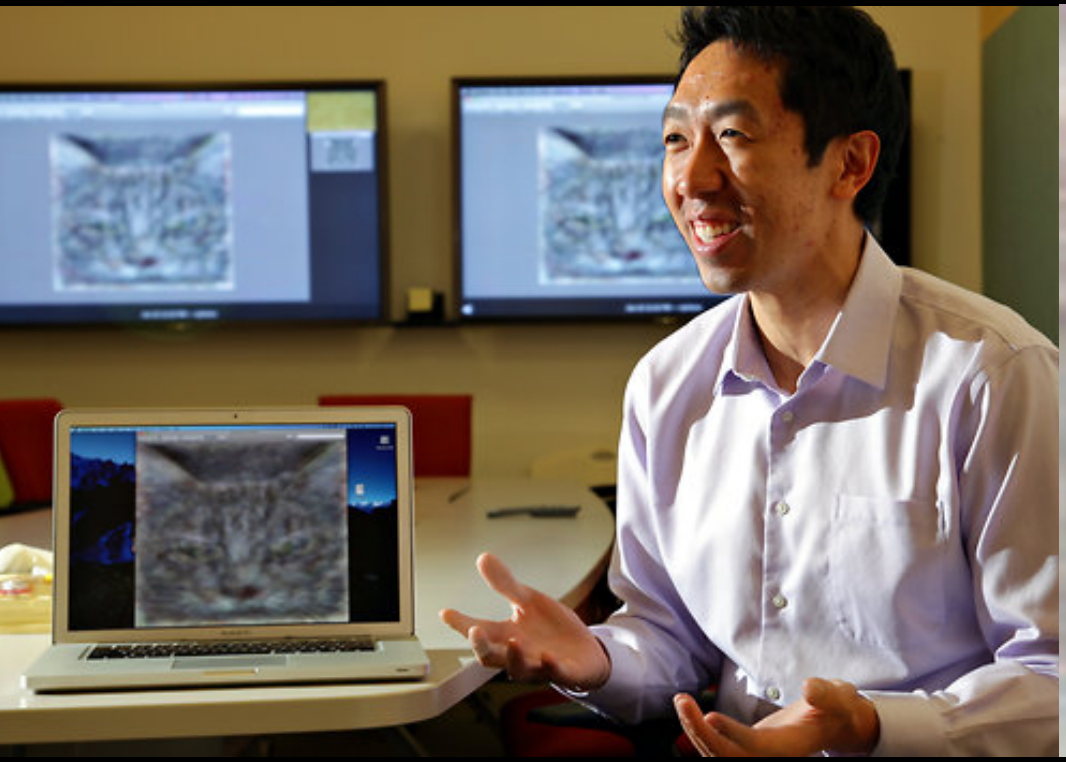
The Internet (Not Just Cat Memes)





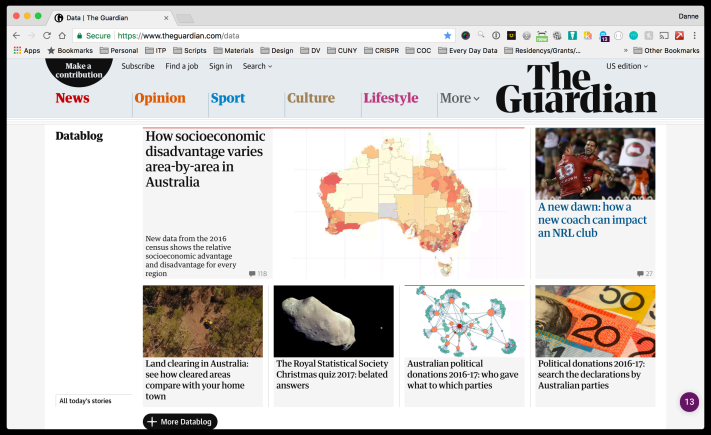
Sourcing Data

But mostly cat memes

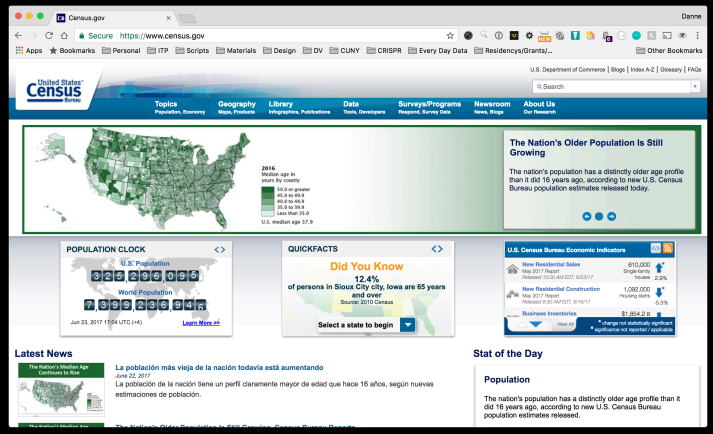


Sourcing Data

Open Datasets

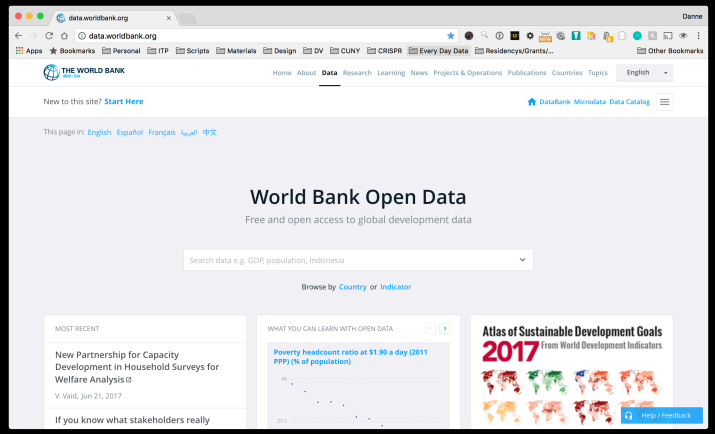


theguardian.com/data



census.gov

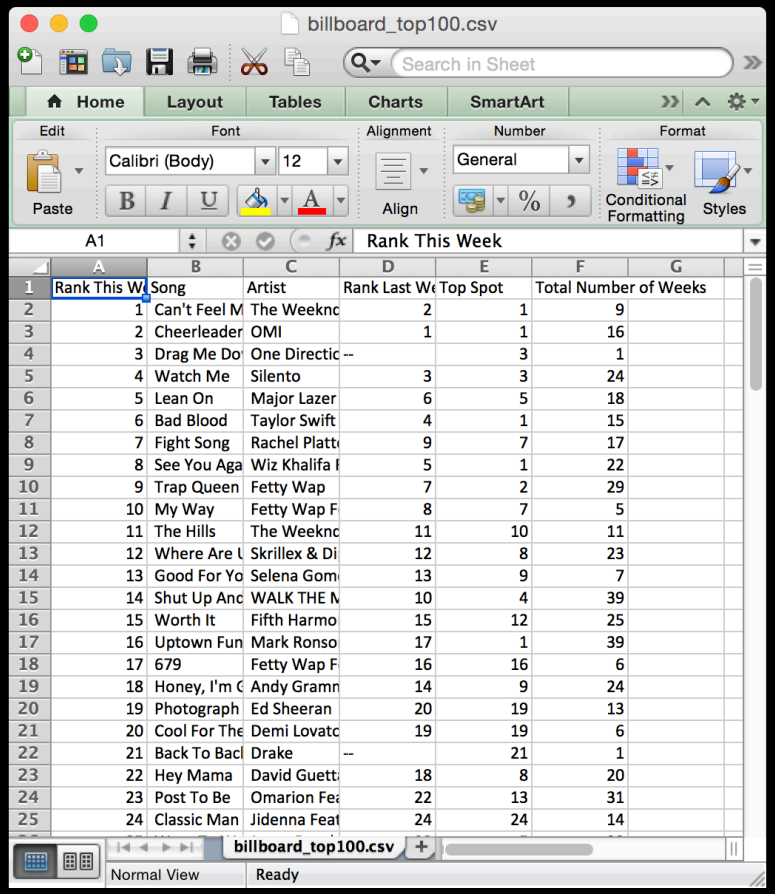
ucr.fbi.gov



data.worldbank.org

Sourcing Data

Data Formats - Tabular



Rank This We	Song	Artist	Rank Last We	Top Spot	Total Number of Weeks
1	Can't Feel M	The Weeknc	2	1	9
2	Cheerleader	OMI	1	1	16
3	Drag Me Do	One Directic	--	3	1
4	Watch Me	Silento	3	3	24
5	Lean On	Major Lazer	6	5	18
6	Bad Blood	Taylor Swift	4	1	15
7	Fight Song	Rachel Platt	9	7	17
8	See You Aga	Wiz Khalifa I	5	1	22
9	Trap Queen	Fetty Wap	7	2	29
10	My Way	Fetty Wap F	8	7	5
11	The Hills	The Weeknc	11	10	11
12	Where Are U	Skrillex & Di	12	8	23
13	Good For Yo	Selena Gom	13	9	7
14	Shut Up And	WALK THE M	10	4	39
15	Worth It	Fifth Harmo	15	12	25
16	Uptown Fun	Mark Ronso	17	1	39
17	679	Fetty Wap F	16	16	6
18	Honey, I'm G	Andy Gramn	14	9	24
19	Photograph	Ed Sheeran	20	19	13
20	Cool For The	Demi Lovatc	19	19	6
21	Back To Bac	Drake	--	21	1
22	Hey Mama	David Guett	18	8	20
23	Post To Be	Omarion Fea	22	13	31
24	Classic Man	Jidenna Feat	24	24	14

CSV – Comma Separated Value

TSV – Tab Separated Value

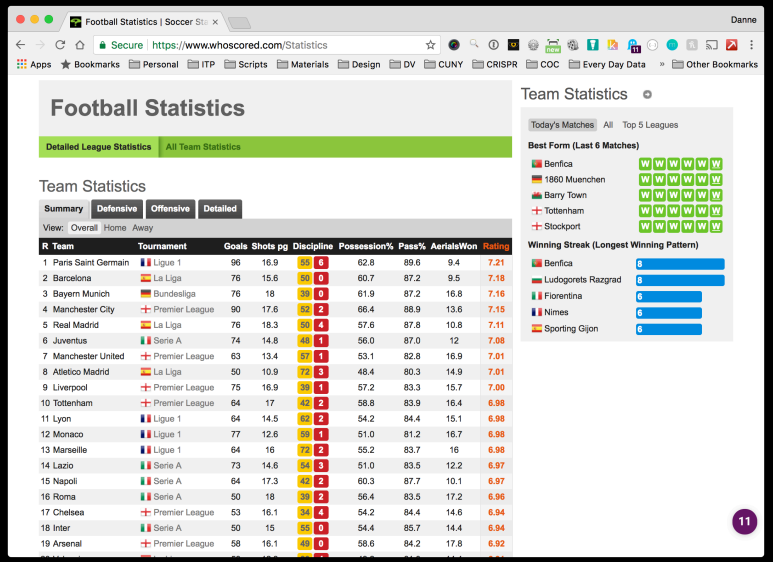
XLS – Excel

Sourcing Data

HTML

Tables

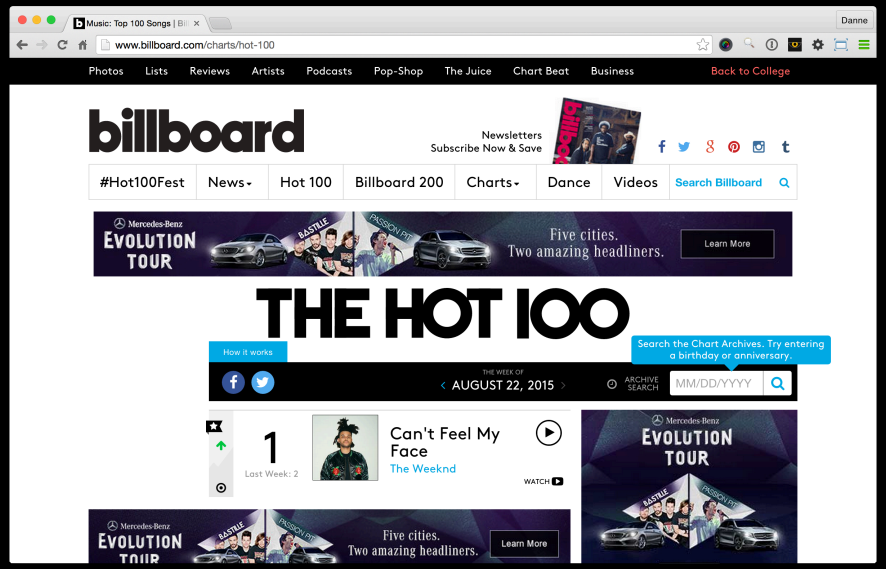
Easily Copy and Paste



whoscored.com/statistics

Complex Layout

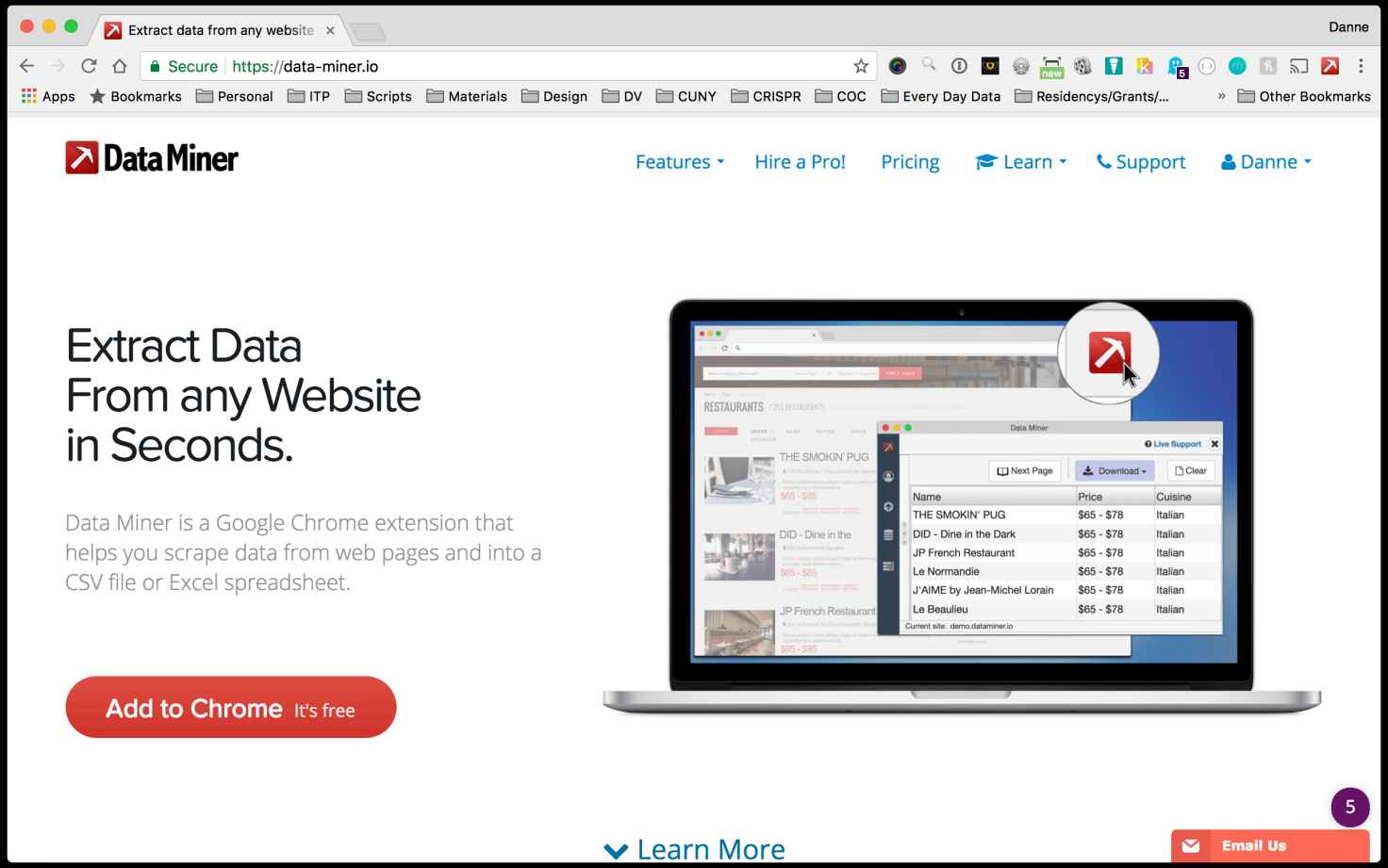
HTML Web Scrapping – Data Miner or Python



billboard.com/charts/hot-100

Sourcing Data

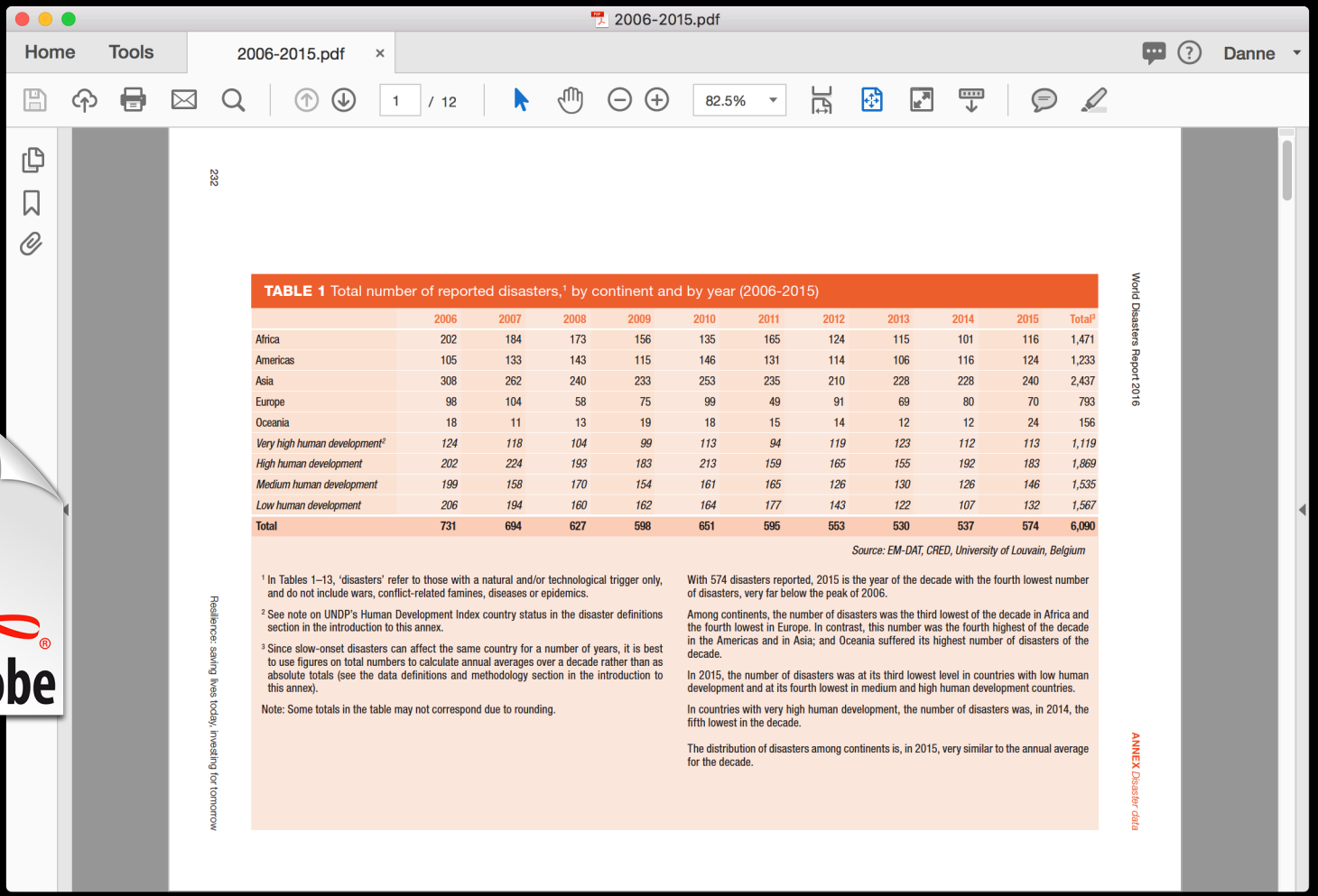
Web Scraping



www.data-miner.io

Sourcing Data

HTML based Data Tables (Tabula Free Software)



232

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.

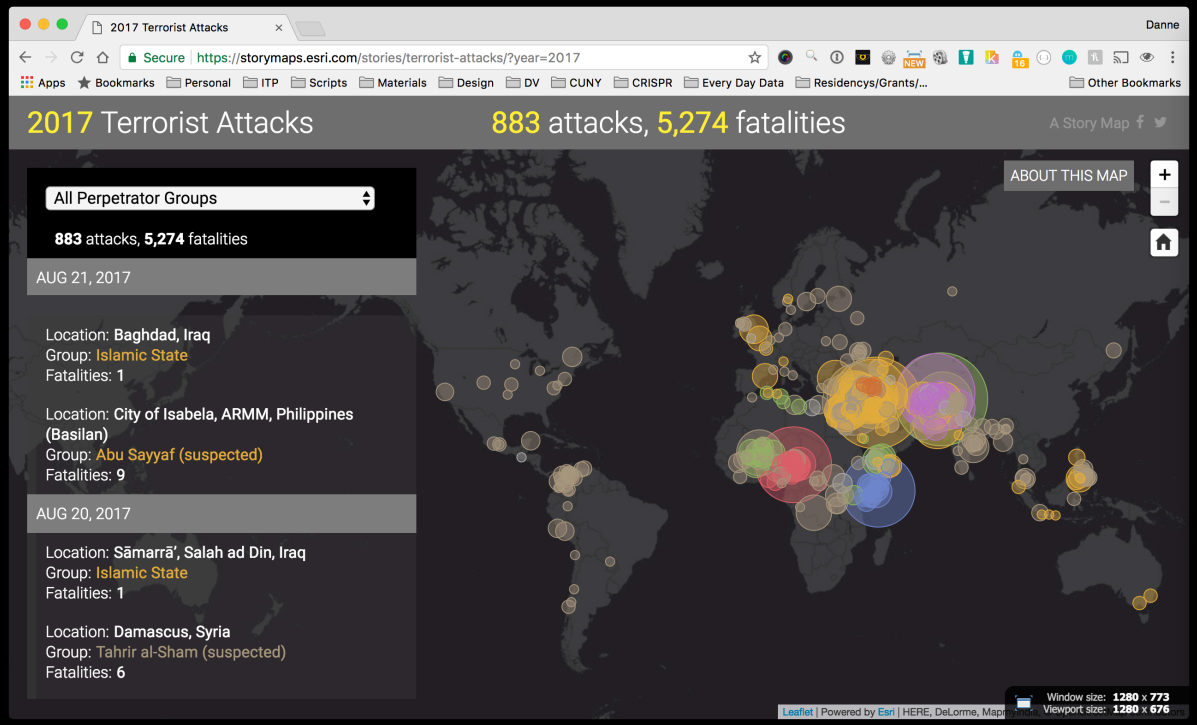
In 2015, the number of disasters was at its third lowest level in countries with low human development and at its fourth lowest in medium and high human development countries.

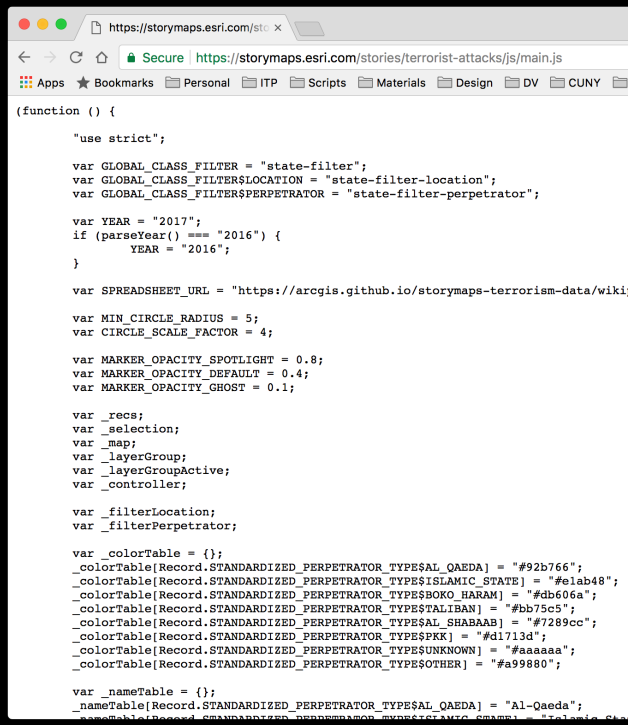
In countries with very high human development, the number of disasters was, in 2014, the fifth lowest in the decade.

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)





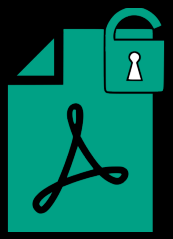
Sourcing Data

Open Dataset Formats

Download file usually as a XLS, CSV, TSV, JSON or XML format

<table>

Select, copy and paste entire table into Excel or Google Sheets



Tabula

NO <table>

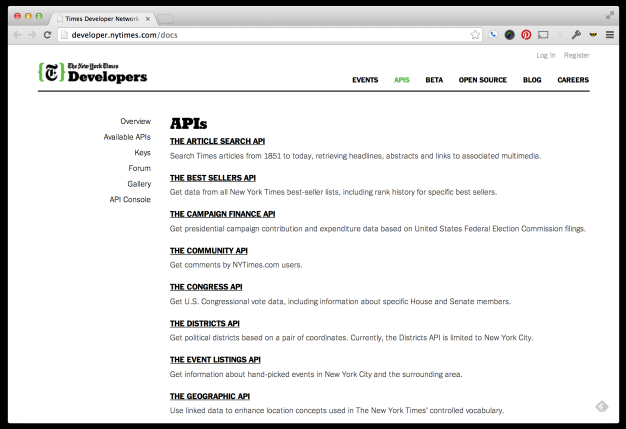
Data Miner

Web Scrapping

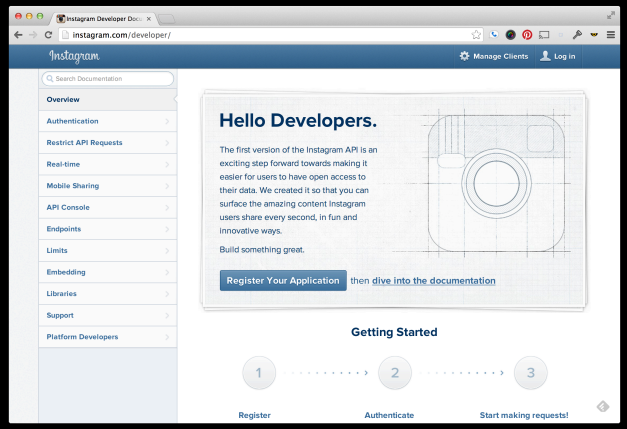
View the JavaScript source code and find any reference to a data file (CSV, JSON, XML)

Sourcing Data

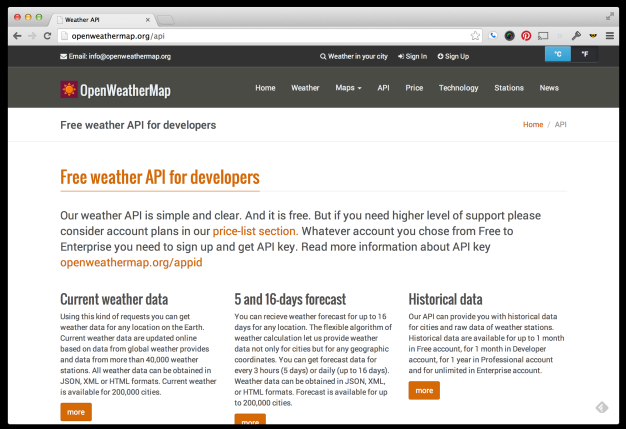
APIs (Application Programming Interfaces)



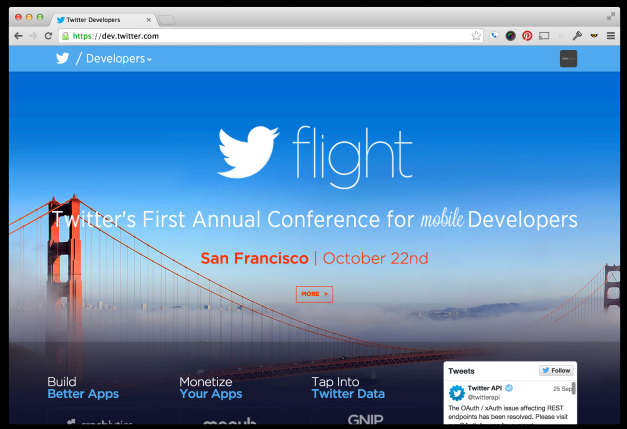
developer.nytimes.com



instagram.com/developer



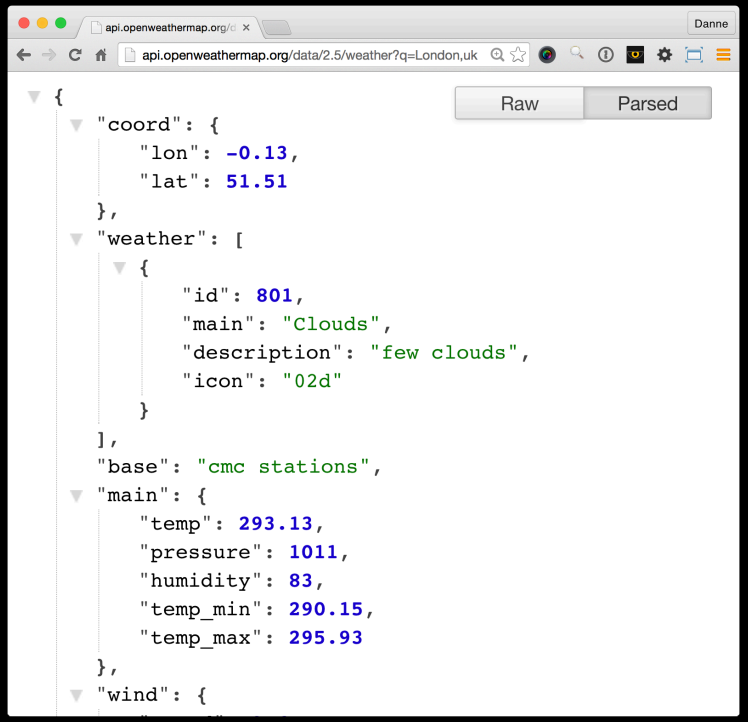
openweathermap.org/api



dev.twitter.com

Sourcing Data

Data Formats - Nested



JSON – JavaScript Object Notation

XML – Extensible Markup Language

Sourcing Data

APIs (Application Programming Interfaces)

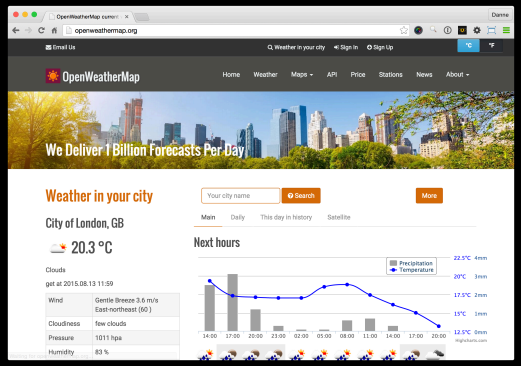
10,000+ Apps

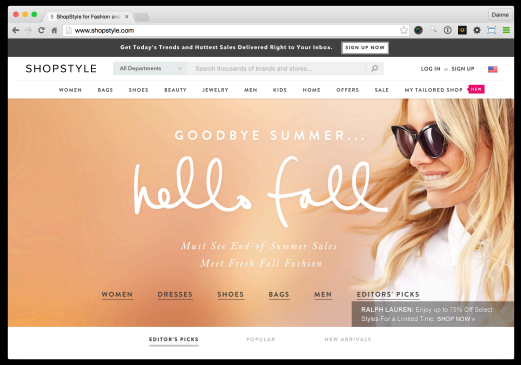
Using Foursquares

Location Data API

Sourcing Data

APIs (Application Programming Interfaces)





fashionweatherready.com

Big Data

John Mashey – 1998

sgi

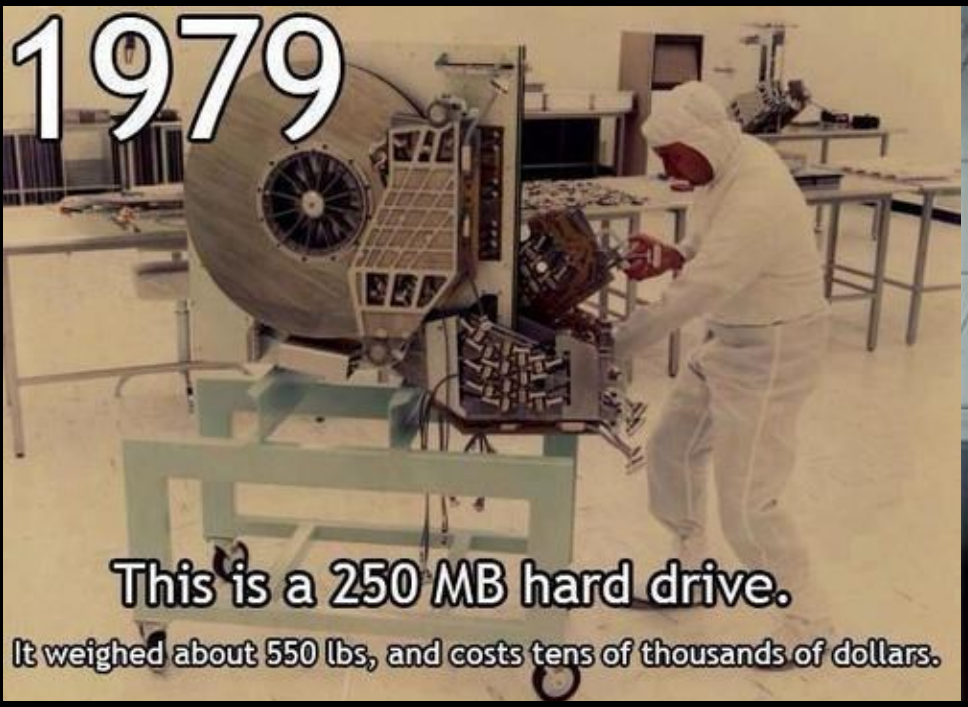
Big Data ...
and the Next Wave of InfraStress

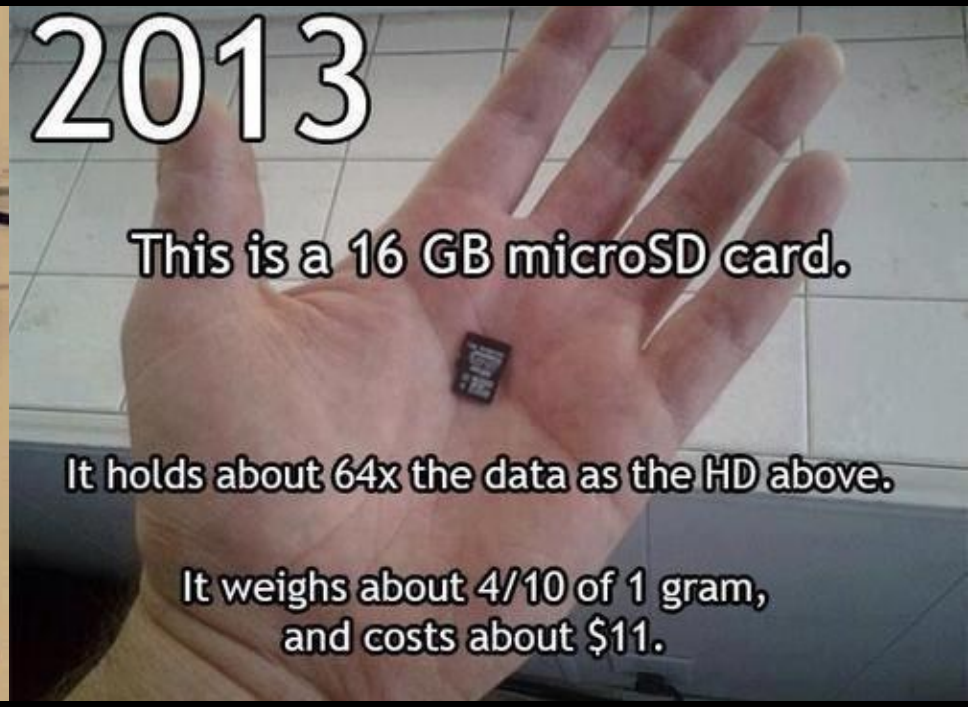
John R. Mashey
Chief Scientist, SGI

Technology Waves:
NOT technology for technology's sake
IT'S WHAT YOU DO WITH IT
But if you don't understand the trends
IT'S WHAT IT WILL DO TO YOU

Big Data

More Storage Capabilities

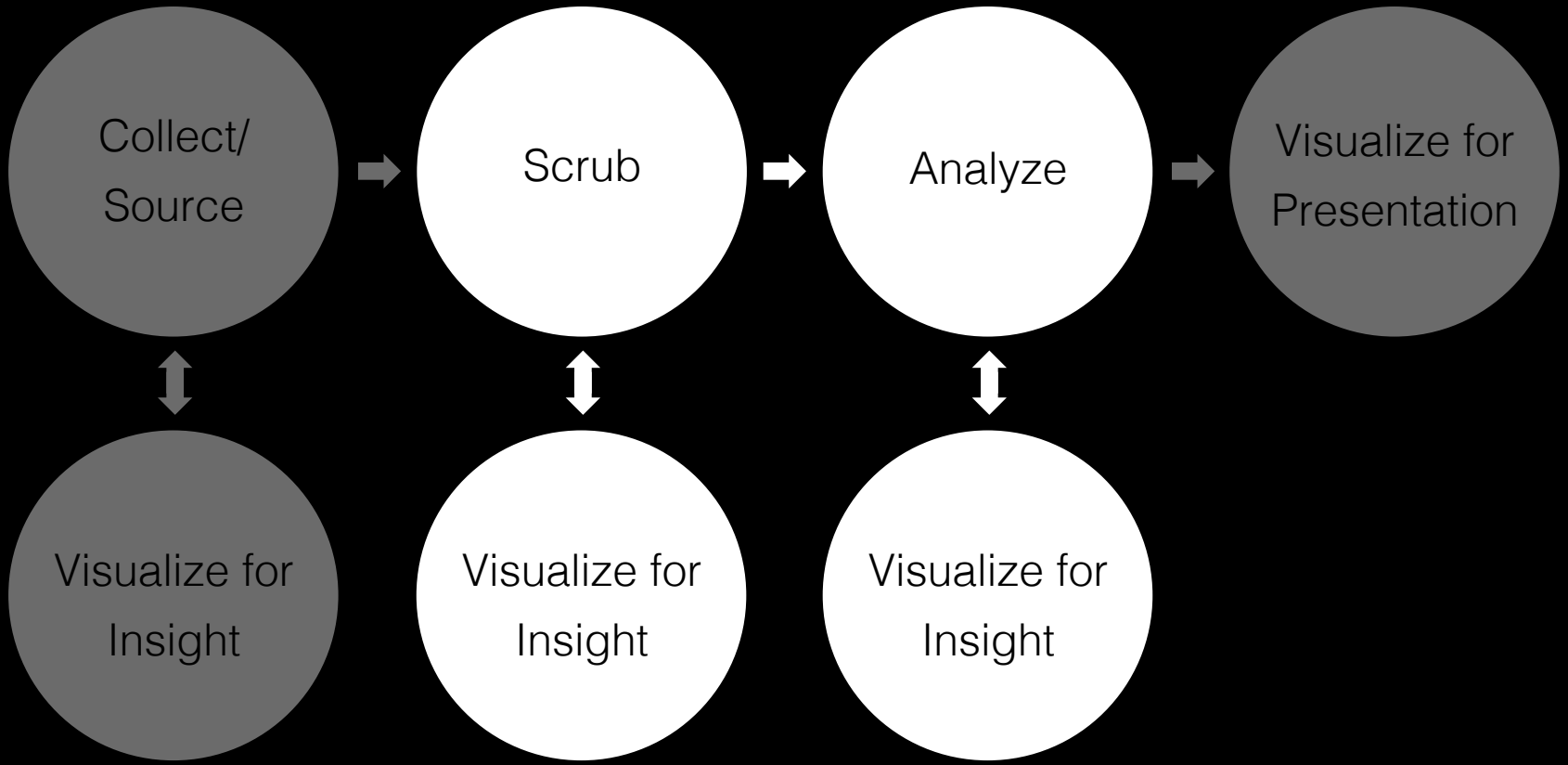




Big Data

Needs more than one computer

Data Pipeline



Data Scrubbing

Date	Country	Address	City	Min Value	Max Value
4/23/2015	United States	3245åFirståStreet	Los Angels	4	9
4/24/2015		28åOrangeåAve	Los Angeles	2	7
3/18/2014		763åMainåSt	Los Angeles	5	9
2/6/2015		72åSecondåAve	Los Angeles	6	9
5/27/2014		87åThurstonåStreet	Los Angles	3	7
6/30/2013		652åGreeneåStreet	New York	4	6
5/7/2014		887åClintonåAve	Nuw York	3	8
3/3/2015		554åWashingtonåStreet	New York	3	9
12/23/2015		92åWaverlyåAve	New York	2	5
2015/2/24	Germany	AmåFalkpl.å5	Berlin	1	7
2014/3/26		Mehringdammå32	Berlin	4	6
2014/6/8		Röntgenstraßeå7	Berlin	1	8
2014/11/12		Rudi-Dutschke-Straßeå26	Berln	2	6

Data Scrubbing

80% Time Spent

Combining Data Files

Data Formatting

Multiple Date Formats

Missing Data

Spaces Replaced with Strange Characters

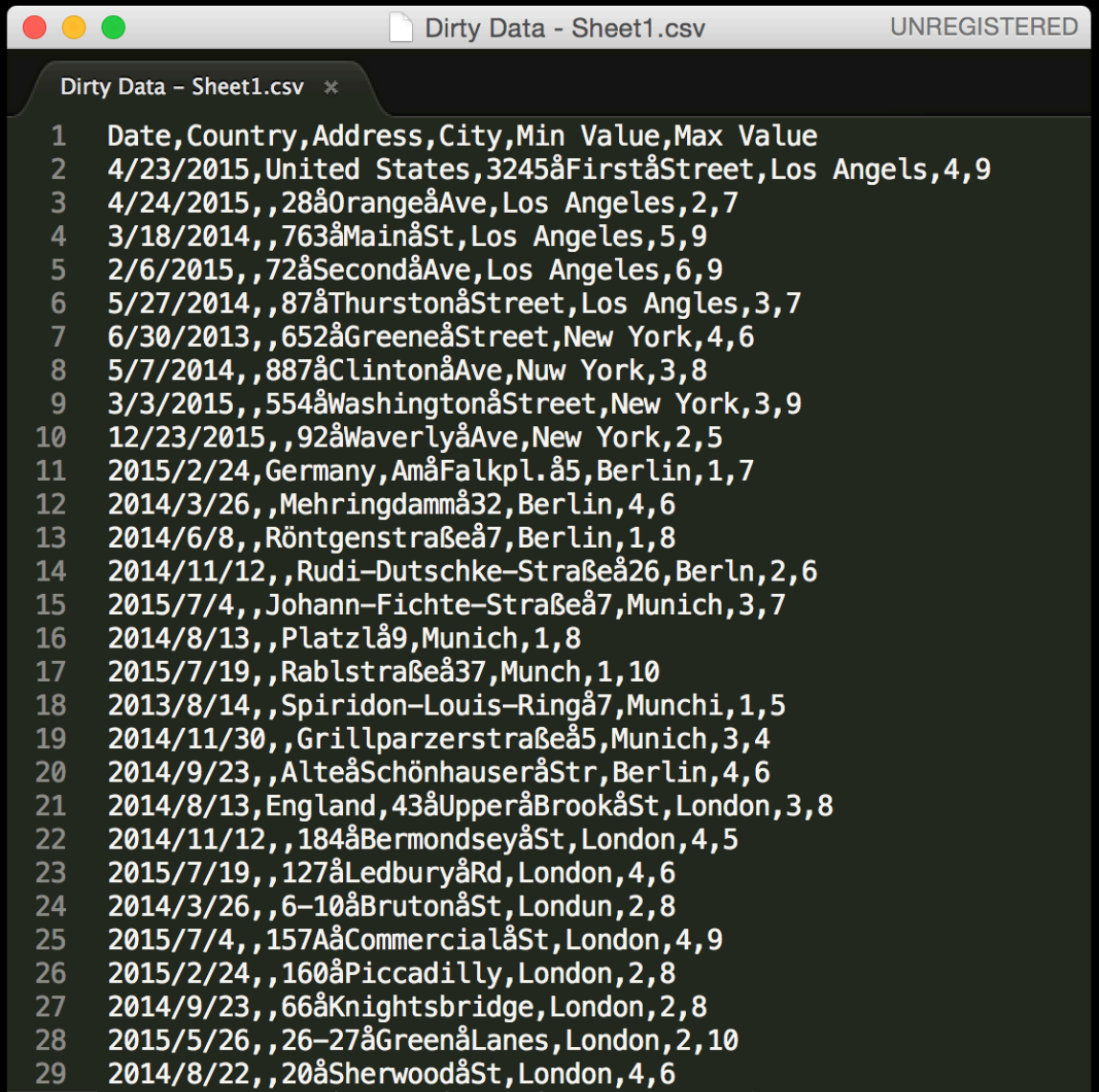
Random HTML Code

Spelling Errors

Etc.

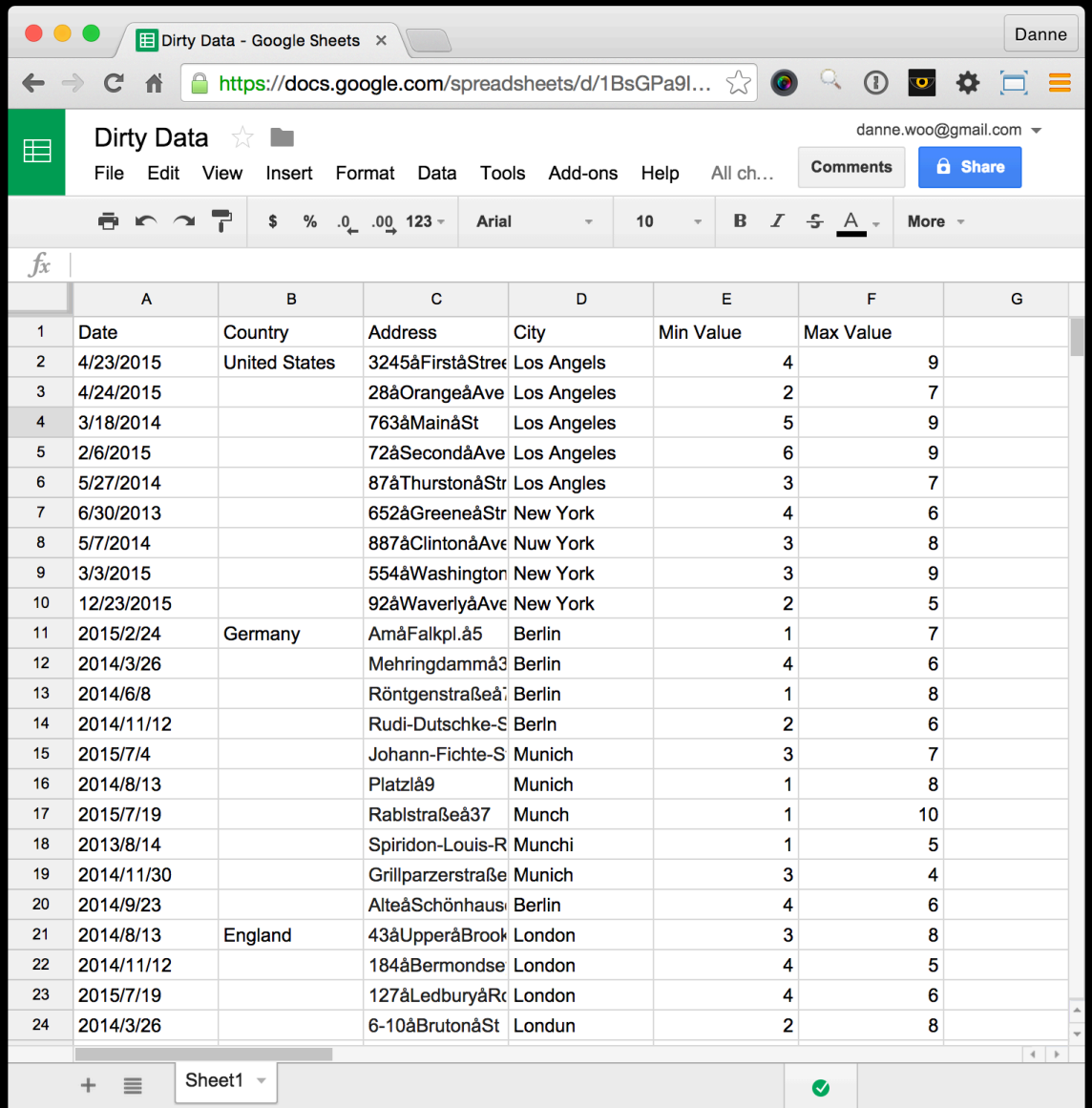
Data Scrubbing

Text Editor (Sublime Text)



Data Scrubbing

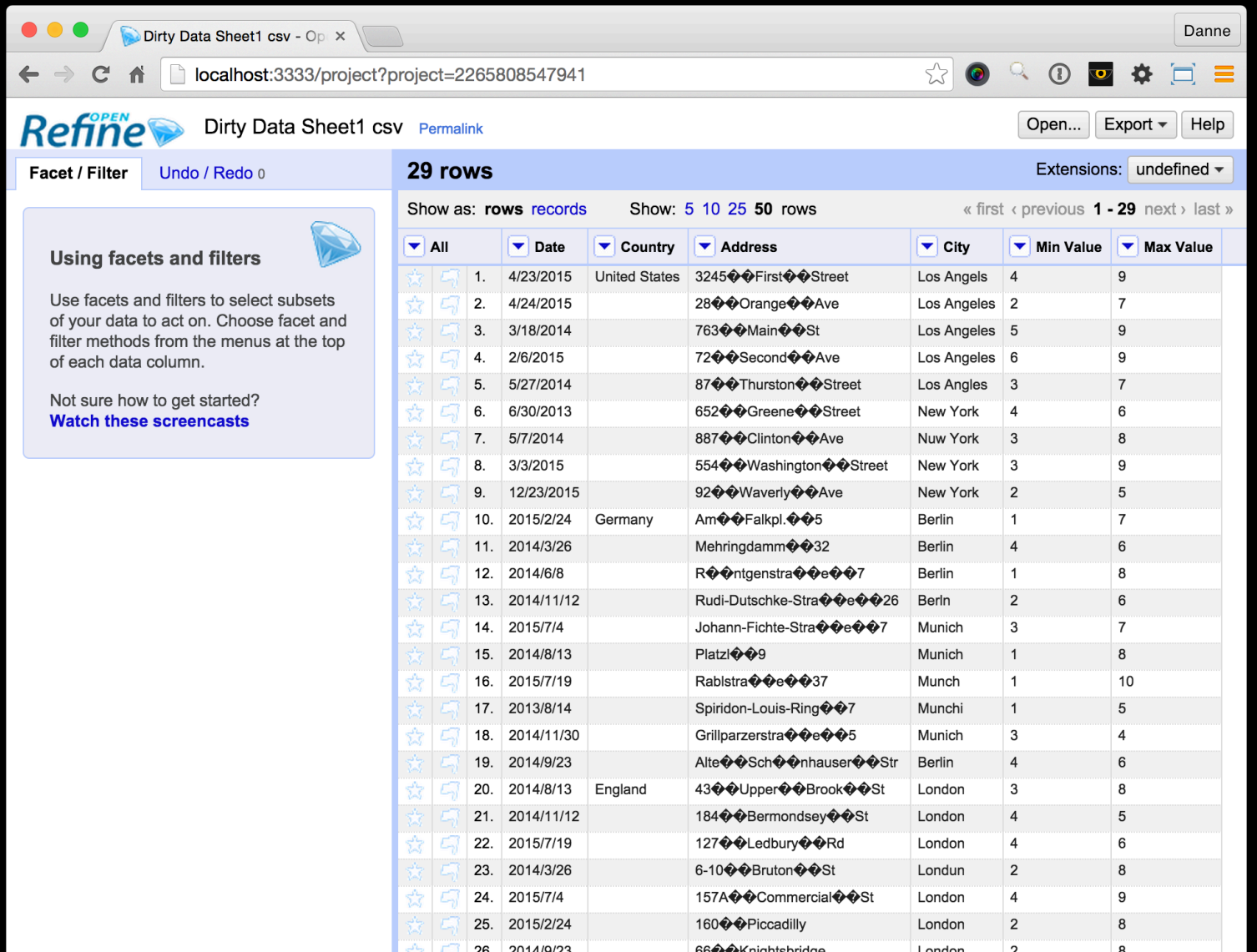
Google Sheets or Excel



	A	B	C	D	E	F	G
1	Date	Country	Address	City	Min Value	Max Value	
2	4/23/2015	United States	3245åFirståStree	Los Angels	4	9	
3	4/24/2015		28åOrangeåAve	Los Angeles	2	7	
4	3/18/2014		763åMainåSt	Los Angeles	5	9	
5	2/6/2015		72åSecondåAve	Los Angeles	6	9	
6	5/27/2014		87åThurstonåStr	Los Angles	3	7	
7	6/30/2013		652åGreeneåStr	New York	4	6	
8	5/7/2014		887åClintonåAve	Nuw York	3	8	
9	3/3/2015		554åWashington	New York	3	9	
10	12/23/2015		92åWaverlyåAve	New York	2	5	
11	2015/2/24	Germany	AmåFalkpl.å5	Berlin	1	7	
12	2014/3/26		Mehringdammå3	Berlin	4	6	
13	2014/6/8		Röntgenstraßeå7	Berlin	1	8	
14	2014/11/12		Rudi-Dutschke-S	Berln	2	6	
15	2015/7/4		Johann-Fichte-S	Munich	3	7	
16	2014/8/13		Platzlå9	Munich	1	8	
17	2015/7/19		Rablstraßeå37	Munch	1	10	
18	2013/8/14		Spiridon-Louis-R	Munchi	1	5	
19	2014/11/30		Grillparzerstraße	Munich	3	4	
20	2014/9/23		AlteåSchönhaus	Berlin	4	6	
21	2014/8/13	England	43åUpperåBrook	London	3	8	
22	2014/11/12		184åBermondse	London	4	5	
23	2015/7/19		127åLedburyåRo	London	4	6	
24	2014/3/26		6-10åBrutonåSt	Londun	2	8	

Data Scrubbing

Open Refine



All	Date	Country	Address	City	Min Value	Max Value
1.	4/23/2015	United States	3245��First��Street	Los Angels	4	9
2.	4/24/2015		28��Orange��Ave	Los Angeles	2	7
3.	3/18/2014		763��Main��St	Los Angeles	5	9
4.	2/6/2015		72��Second��Ave	Los Angeles	6	9
5.	5/27/2014		87��Thurston��Street	Los Angles	3	7
6.	6/30/2013		652��Greene��Street	New York	4	6
7.	5/7/2014		887��Clinton��Ave	Nuw York	3	8
8.	3/3/2015		554��Washington��Street	New York	3	9
9.	12/23/2015		92��Waverly��Ave	New York	2	5
10.	2015/2/24	Germany	Am��Falkpl.��5	Berlin	1	7
11.	2014/3/26		Mehringdamm��32	Berlin	4	6
12.	2014/6/8		R��ntgenstra��e��7	Berlin	1	8
13.	2014/11/12		Rudi-Dutschke-Stra��e��26	Berln	2	6
14.	2015/7/4		Johann-Fichte-Stra��e��7	Munich	3	7
15.	2014/8/13		Platzl��9	Munich	1	8
16.	2015/7/19		Rablstra��e��37	Munch	1	10
17.	2013/8/14		Spiridon-Louis-Ring��7	Munchi	1	5
18.	2014/11/30		Grillparzerstra��e��5	Munich	3	4
19.	2014/9/23		Alte��Sch��nhauser��Str	Berlin	4	6
20.	2014/8/13	England	43��Upper��Brook��St	London	3	8
21.	2014/11/12		184��Bermondsey��St	London	4	5
22.	2015/7/19		127��Ledbury��Rd	London	4	6
23.	2014/3/26		6-10��Bruton��St	Londun	2	8
24.	2015/7/4		157A��Commercial��St	London	4	9
25.	2015/2/24		160��Piccadilly	London	2	8
26.	2014/9/23		66��Knightsbridge	London	2	8

Data Scrubbing

Converting XML > CSV

export.xml — UNREGISTERED

export.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<HealthData locale="en_US">
 <ExportDate value="20150814122840-0400"/>
 <Me HKCharacteristicTypeIdentifierDateOfBirth="19820630" HKCharacteristicTypeIdentifierBiologicalSex="2"
 HKCharacteristicTypeIdentifierBloodType="0"/>
 <Record type="HKQuantityTypeIdentifierHeight" source="Danne's Apple Watch" unit="m" creationDate="20150630182606-
 0400" startDate="20150630182606-0400" endDate="20150630182606-0400" value="1.778"/>
 <Record type="HKQuantityTypeIdentifierBodyMass" source="Danne's Apple Watch" unit="kg" creationDate="20150630182606-
 0400" startDate="20150630182606-0400" endDate="20150630182606-0400" value="65.7709"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150630122800-0400" endDate="20150701122800-0400" min="0.95" max="1.85" average="1.17278" recordCount="30"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150701122800-0400" endDate="20150702122800-0400" min="1.01667" max="1.55" average="1.17302" recordCount="21"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150702122800-0400" endDate="20150703122800-0400" min="0.9" max="1.55" average="1.17742" recordCount="31"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150703122800-0400" endDate="20150704122800-0400" min="0.766667" max="1.61667" average="1.15045" recordCount="37"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150704122800-0400" endDate="20150705122800-0400" min="0.9" max="1.88333" average="1.19608" recordCount="51"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150705122800-0400" endDate="20150706122800-0400" min="0.816667" max="1.4" average="1.13947" recordCount="19"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150706122800-0400" endDate="20150707122800-0400" min="1.01667" max="1.61667" average="1.16818" recordCount="22"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150707122800-0400" endDate="20150708122800-0400" min="1.01667" max="1.53333" average="1.21759" recordCount="36"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150708122800-0400" endDate="20150709122800-0400" min="0.9" max="1.66667" average="1.14167" recordCount="28"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150709122800-0400" endDate="20150710122800-0400" min="1.01667" max="1.61667" average="1.25873" recordCount="21"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150710122800-0400" endDate="20150711122800-0400" min="0.766667" max="1.78333" average="1.22593" recordCount="36"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150711122800-0400" endDate="20150712122800-0400" min="1.01667" max="1.23333" average="1.11" recordCount="5"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150712122800-0400" endDate="20150713122800-0400" min="1.2" max="1.53333" average="1.34583" recordCount="8"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150713122800-0400" endDate="20150714122800-0400" min="1.01667" max="1.66667" average="1.22467" recordCount="25"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150714122800-0400" endDate="20150715122800-0400" min="0.683333" max="1.58333" average="1.25" recordCount="15"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
 20150715122800-0400" endDate="20150716122800-0400" min="1.03333" max="1.78333" average="1.3375" recordCount="8"/>
 <Record type="HKQuantityTypeIdentifierHeartRate" source="Danne's Apple Watch" unit="count/s" startDate="
```

Data Scrubbing

Python or JavaScript

index.html UNREGISTERED

index.html

```
<!doctype html>
<html lang="en">
<head>
  <meta charset="utf-8">
  <title>jQuery.parseXML demo</title>
  <script src="https://code.jquery.com/jquery-1.10.2.js"></script>
</head>
<body>

<script>
$(document).ready(function(){
  $.ajax({
    type: "GET",
    url: "export.xml",
    dataType: "xml",
    success: function(xml){
    console.log($(xml).find("HealthData").find("Record").size());
      $(xml).find("HealthData").each(function(){
        var record = $(this).find("Record");
        name = "start date,type,value,unit<br>"
      $(record).each(function(index, element){
        if ($(this).attr('value') != undefined) {
            name += $(this).attr('startDate') + "," + $(this).attr('type') + "," +
              $(this).attr('value') + "," + $(this).attr('unit') + "<br>";
        }
          if (index < length-1) {
            name += ', ';
          }
      });
        document.write(name);
      })
    }
   });
});
</script>
```

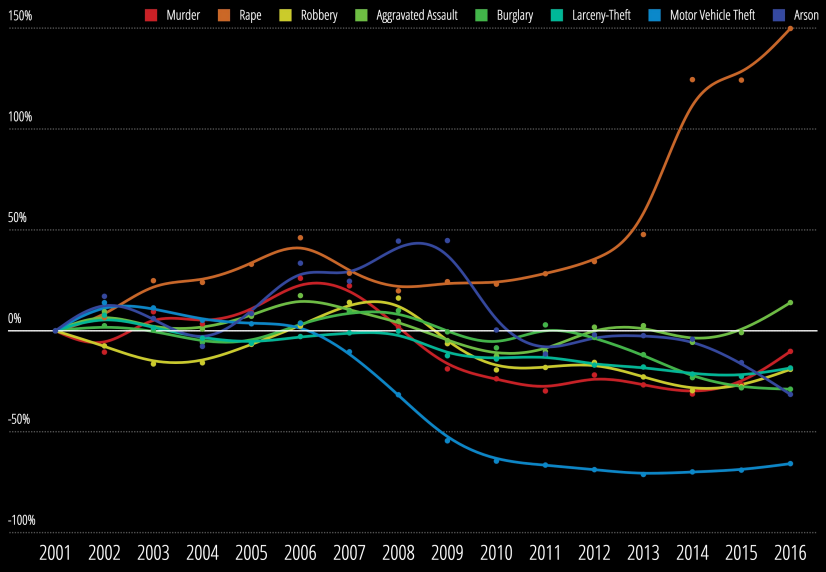
Data Scrubbing

Combining Datasets

Sheriff Joe's Crime Record

Last Friday Trump issued his first pardon for Joe Arpaio, the controversial former sheriff of Maricopa County, Arizona. Sheriff Joe first gained national recognition for his participation in the debunked Obama birther conspiracy and has since been criticized for his policing methods considered by many to be racist. After Joe refused to follow a court order to stop his racial profiling practices, he was found guilty of contempt. One aspect that makes this pardon unusual is that it happened before his sentencing.

The chart below shows the yearly change in Sheriff Joe's violent crime record since 2001. While most violent crimes either went down or stayed the same, rape cases skyrocketed.



Source: azdps.gov/about/reports/crime and city-data.com/crime/crime-Arizona.html

Created using www.datavisu.al

2013 CRIME IN ARIZONA REPORT

2012 CRIME IN ARIZONA REPORT

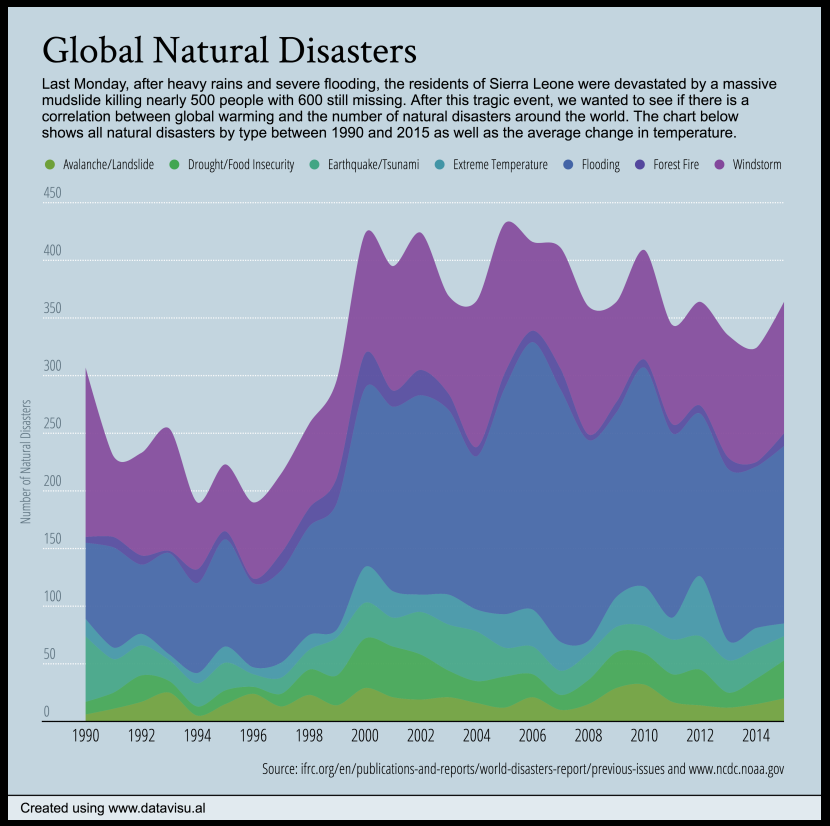
2011 CRIME IN ARIZONA REPORT

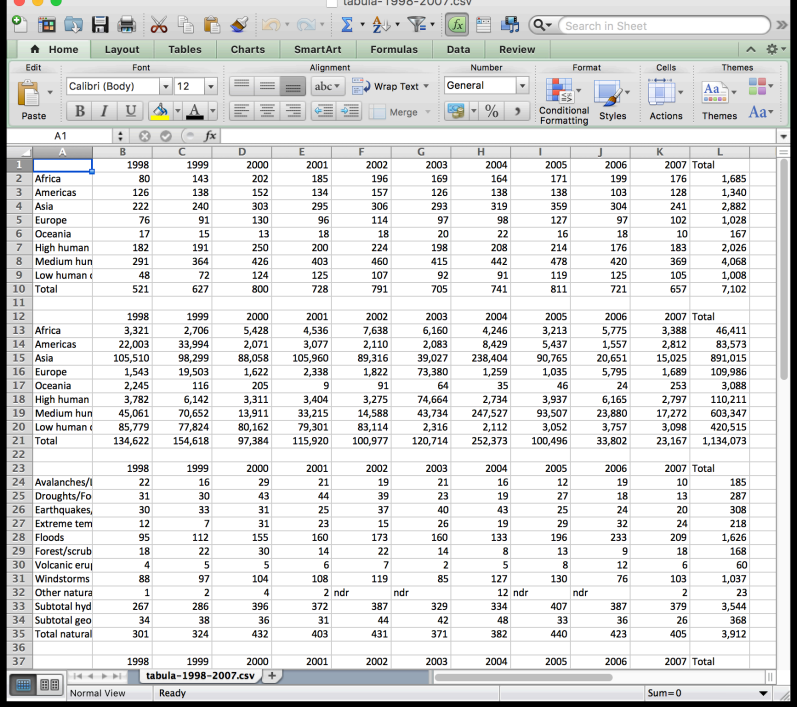
2011 CRIME IN ARIZONA REPORT

2010 CRIME IN ARIZONA REPORT

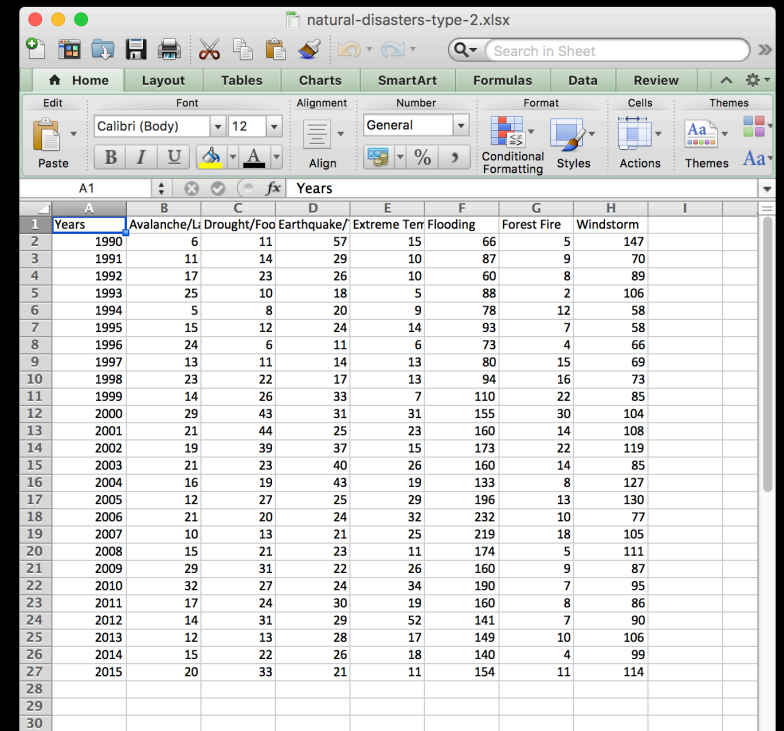
Data Scrubbing

Reformatting Datasets





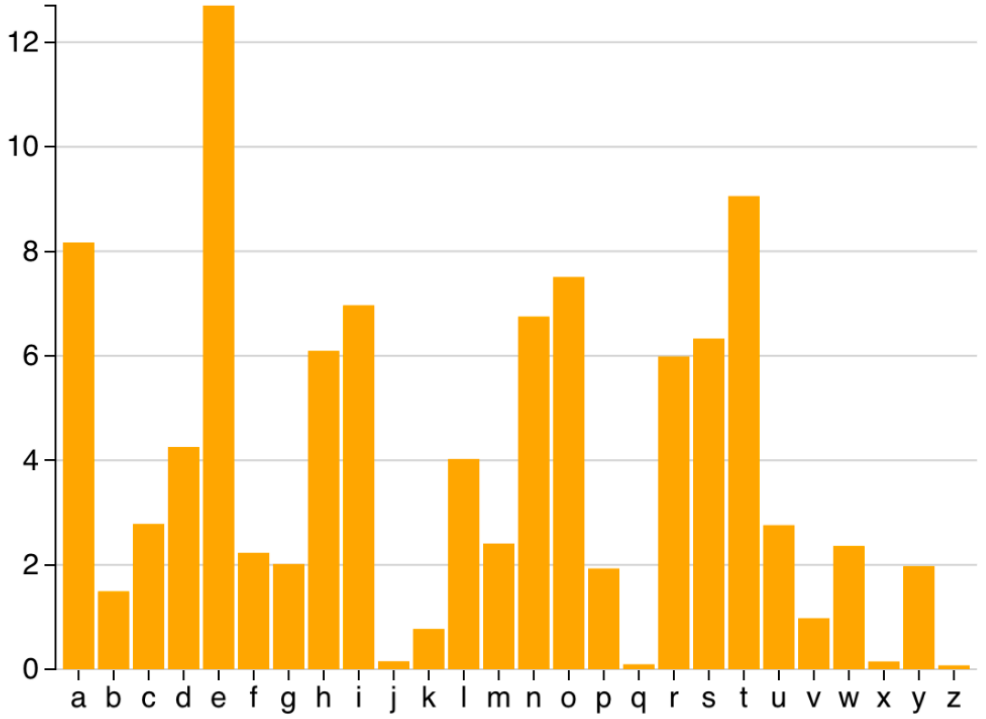
	A	B	C	D	E	F	G	H	I	J	K	L
1		1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
2	Africa	80	143	202	185	196	169	164	171	199	176	1,685
3	Americas	126	138	152	134	157	126	138	138	103	128	1,340
4	Asia	222	240	303	295	306	293	319	359	304	241	2,882
5	Europe	76	91	130	96	114	97	98	127	97	102	1,028
6	Oceania	17	15	13	18	18	20	22	16	18	10	167
7	High human	182	191	250	200	224	198	208	214	176	183	2,026
8	Medium hun	291	364	426	403	460	415	442	478	420	369	4,068
9	Low human	48	72	124	125	107	92	91	119	125	105	1,008
10	Total	521	627	800	728	791	705	741	811	721	657	7,102
11												
12		1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
13	Africa	3,321	2,706	5,428	4,536	7,638	6,160	4,246	3,213	5,775	3,388	46,411
14	Americas	22,003	33,994	2,071	3,077	2,110	2,083	8,429	5,437	1,557	2,812	83,573
15	Asia	105,510	98,299	88,058	105,960	89,316	39,027	238,404	90,765	20,651	15,025	891,015
16	Europe	1,543	19,503	1,622	2,338	1,822	73,380	1,259	1,035	5,795	1,689	109,986
17	Oceania	2,245	116	205	9	91	64	35	46	24	253	3,088
18	High human	3,782	6,142	3,311	3,404	3,275	74,664	2,734	3,937	6,165	2,797	110,211
19	Medium hun	45,061	70,652	13,911	33,215	14,588	43,734	247,527	93,507	23,880	17,272	603,347
20	Low human	85,779	77,824	80,162	79,301	83,114	2,316	2,112	3,052	3,757	3,098	420,515
21	Total	134,622	154,618	97,384	115,920	100,977	120,714	252,373	100,496	33,802	23,167	1,134,073
22												
23		1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
24	Avalanches/l	22	16	29	21	19	21	16	12	19	10	185
25	Droughts/Fo	31	30	43	44	39	23	19	27	18	13	287
26	Earthquakes,	30	33	31	25	37	40	43	25	24	20	308
27	Extreme tem	12	7	31	23	15	26	19	29	32	24	218
28	Floods	95	112	155	160	173	160	133	196	233	209	1,626
29	Forest/scrub	18	22	30	14	22	14	8	13	9	18	168
30	Volcanic eru	4	5	5	6	7	2	5	8	12	6	60
31	Windstorms	88	97	104	108	119	85	127	130	76	103	1,037
32	Other natura	1	2	4	2	ndr	ndr	12	ndr	ndr	2	23
33	Subtotal hyd	267	286	396	372	387	329	334	407	387	379	3,544
34	Subtotal geo	34	38	36	31	44	42	48	33	36	26	368
35	Total natural	301	324	432	403	431	371	382	440	423	405	3,912
36												
37		1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total



	A	B	C	D	E	F	G	H	I
1	Years	Avalanche/L	Drought/Foo	Earthquake/	Extreme Tem	Flooding	Forest Fire	Windstorm	
2	1990	6	11	57	15	66	5	147	
3	1991	11	14	29	10	87	9	70	
4	1992	17	23	26	10	60	8	89	
5	1993	25	10	18	5	88	2	106	
6	1994	5	8	20	9	78	12	58	
7	1995	15	12	24	14	93	7	58	
8	1996	24	6	11	6	73	4	66	
9	1997	13	11	14	13	80	15	69	
10	1998	23	22	17	13	94	16	73	
11	1999	14	26	33	7	110	22	85	
12	2000	29	43	31	31	155	30	104	
13	2001	21	44	25	23	160	14	108	
14	2002	19	39	37	15	173	22	119	
15	2003	21	23	40	26	160	14	85	
16	2004	16	19	43	19	133	8	127	
17	2005	12	27	25	29	196	13	130	
18	2006	21	20	24	32	232	10	77	
19	2007	10	13	21	25	219	18	105	
20	2008	15	21	23	11	174	5	111	
21	2009	29	31	22	26	160	9	87	
22	2010	32	27	24	34	190	7	95	
23	2011	17	24	30	19	160	8	86	
24	2012	14	31	29	52	141	7	90	
25	2013	12	13	28	17	149	10	106	
26	2014	15	22	26	18	140	4	99	
27	2015	20	33	21	11	154	11	114	
28									
29									
30									

Data Analysis

Al-Kindi, 850



Data Analysis

John Tukey, 1977

Data Analysis

John Tukey, 1977

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

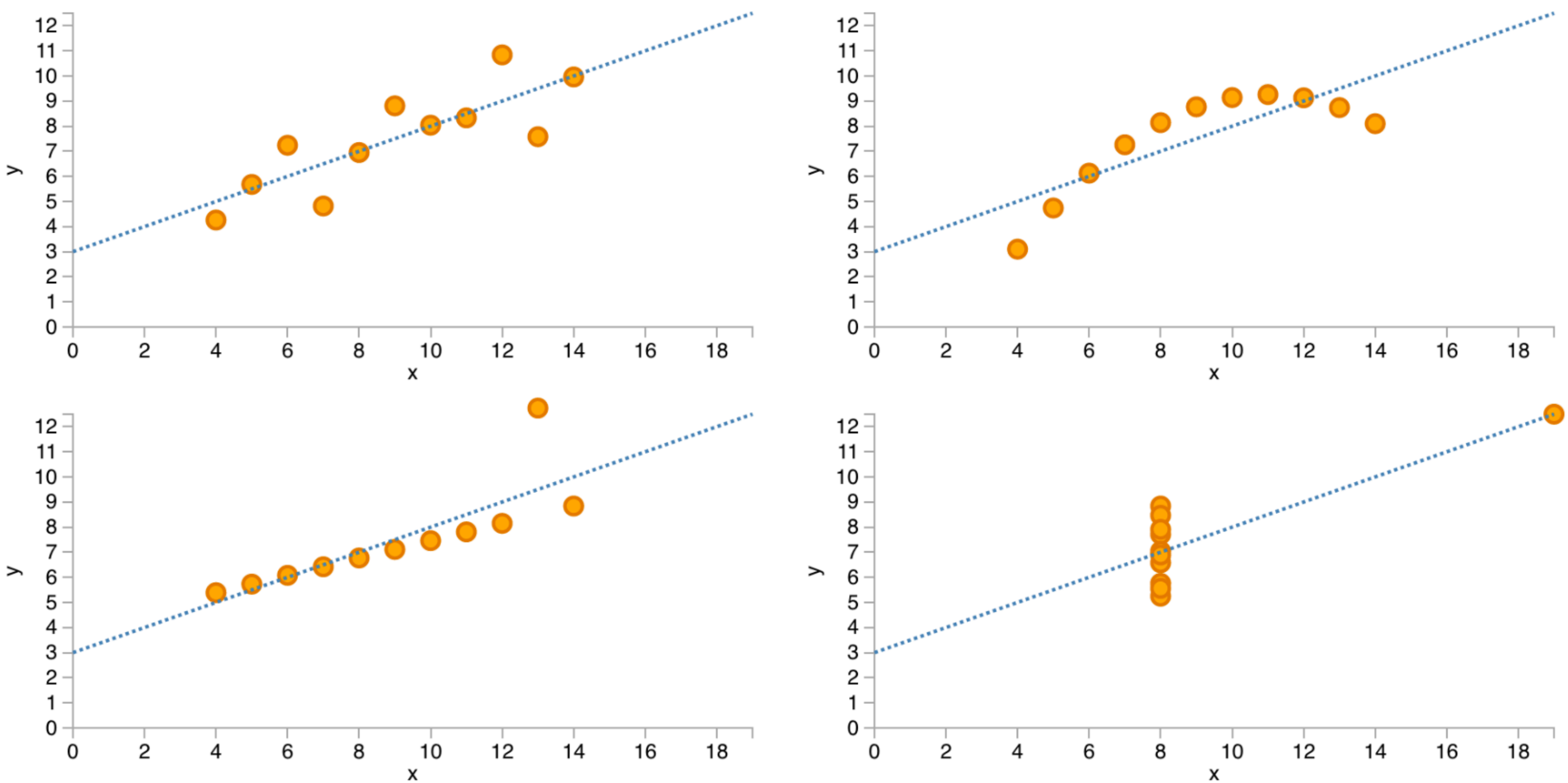
Mean of x	9	Variance of x	11
Mean of y	7.50	Variance of y	4.122

Correlation between x and y: 0.816

Linear regression line: $y = 3.00 + 0.50x$

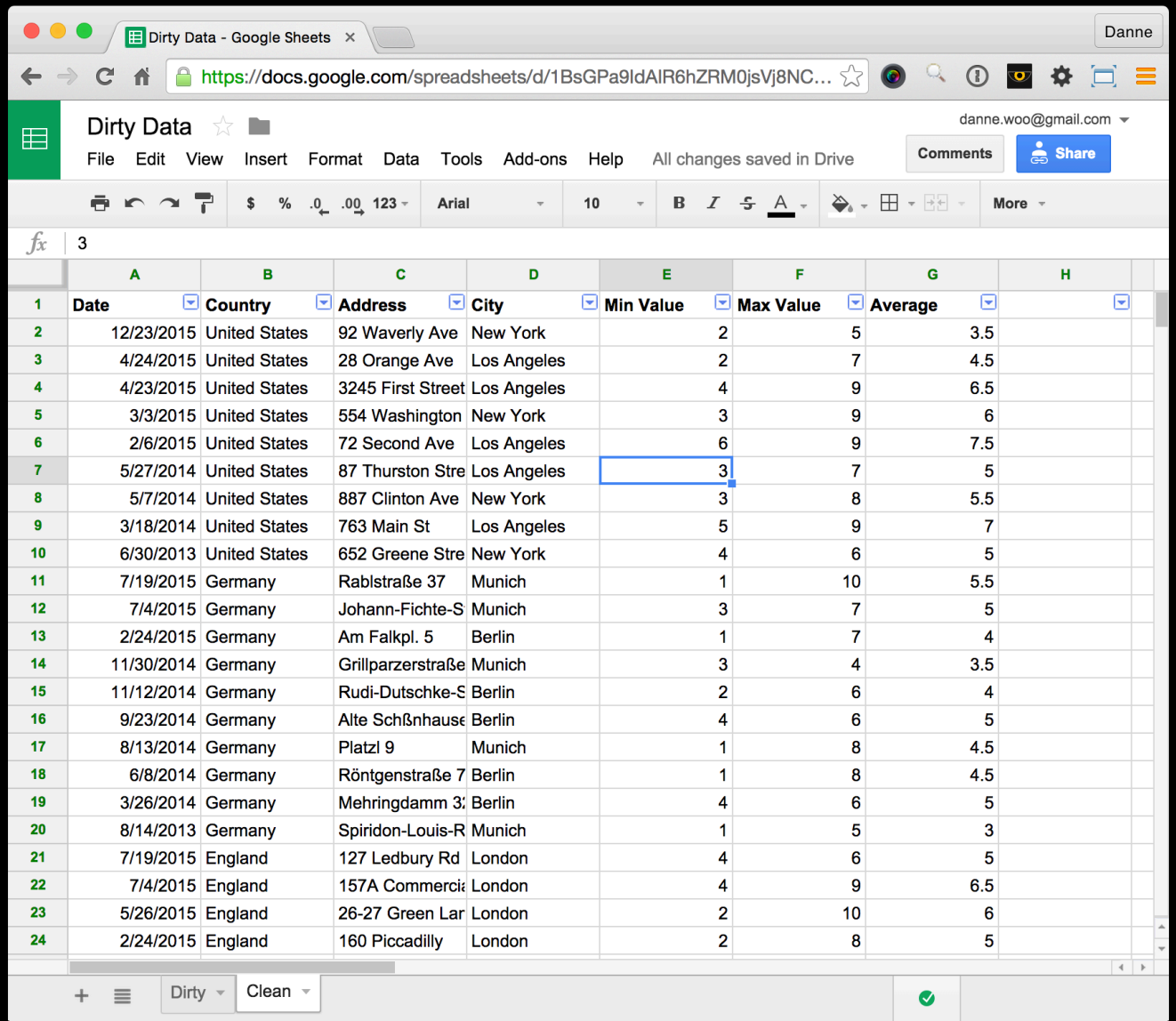
Data Analysis

John Tukey, 1977



Data Analysis

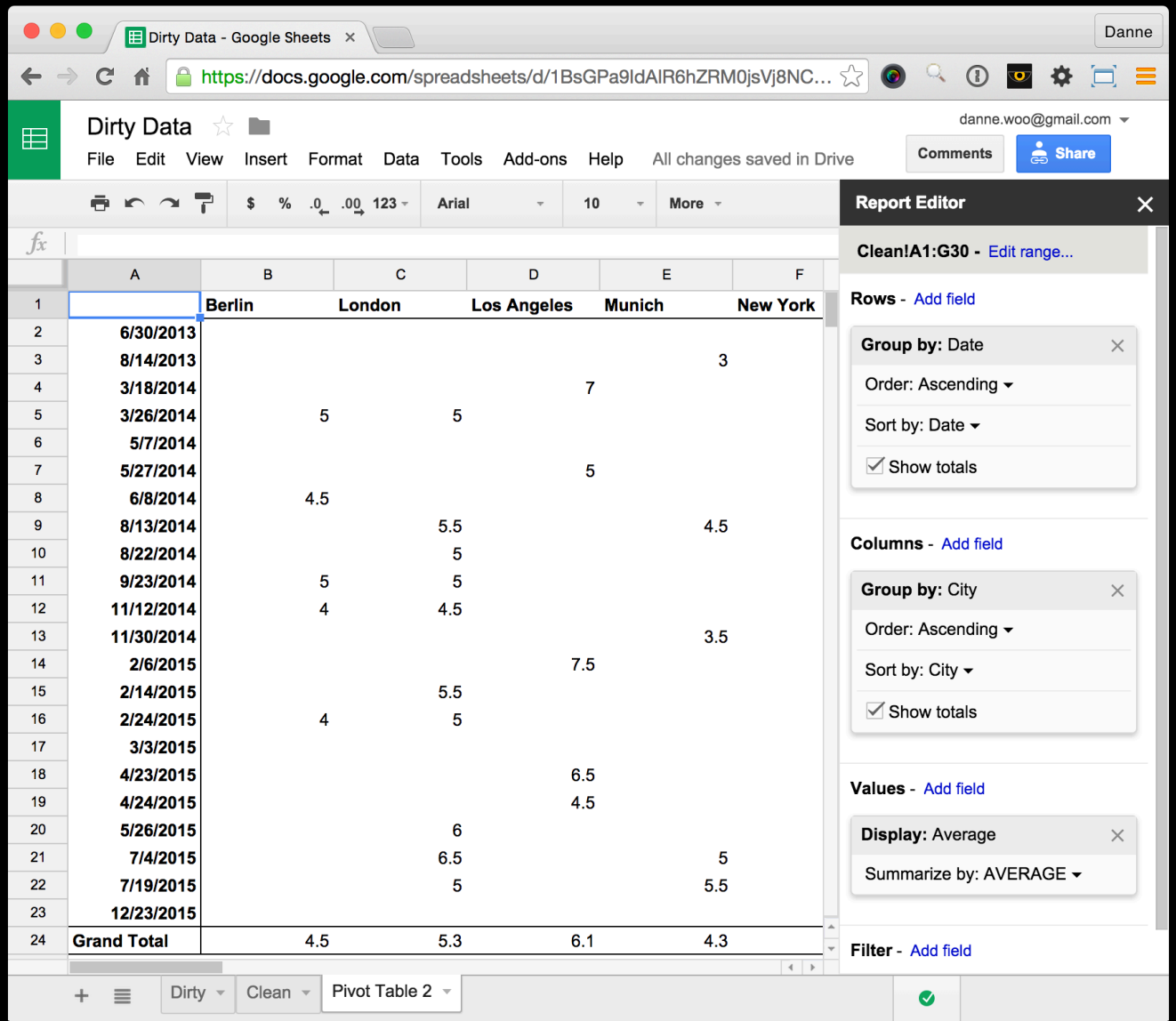
Google Sheets and Excel



	A	B	C	D	E	F	G	H
1	Date	Country	Address	City	Min Value	Max Value	Average	
2	12/23/2015	United States	92 Waverly Ave	New York	2	5	3.5	
3	4/24/2015	United States	28 Orange Ave	Los Angeles	2	7	4.5	
4	4/23/2015	United States	3245 First Street	Los Angeles	4	9	6.5	
5	3/3/2015	United States	554 Washington	New York	3	9	6	
6	2/6/2015	United States	72 Second Ave	Los Angeles	6	9	7.5	
7	5/27/2014	United States	87 Thurston Stre	Los Angeles	3	7	5	
8	5/7/2014	United States	887 Clinton Ave	New York	3	8	5.5	
9	3/18/2014	United States	763 Main St	Los Angeles	5	9	7	
10	6/30/2013	United States	652 Greene Stre	New York	4	6	5	
11	7/19/2015	Germany	Rablstraße 37	Munich	1	10	5.5	
12	7/4/2015	Germany	Johann-Fichte-S	Munich	3	7	5	
13	2/24/2015	Germany	Am Falkpl. 5	Berlin	1	7	4	
14	11/30/2014	Germany	Grillparzerstraße	Munich	3	4	3.5	
15	11/12/2014	Germany	Rudi-Dutschke-S	Berlin	2	6	4	
16	9/23/2014	Germany	Alte Schßnhause	Berlin	4	6	5	
17	8/13/2014	Germany	Platzl 9	Munich	1	8	4.5	
18	6/8/2014	Germany	Röntgenstraße 7	Berlin	1	8	4.5	
19	3/26/2014	Germany	Mehringdamm 3	Berlin	4	6	5	
20	8/14/2013	Germany	Spiridon-Louis-R	Munich	1	5	3	
21	7/19/2015	England	127 Ledbury Rd	London	4	6	5	
22	7/4/2015	England	157A Commercia	London	4	9	6.5	
23	5/26/2015	England	26-27 Green Lar	London	2	10	6	
24	2/24/2015	England	160 Piccadilly	London	2	8	5	

Data Analysis

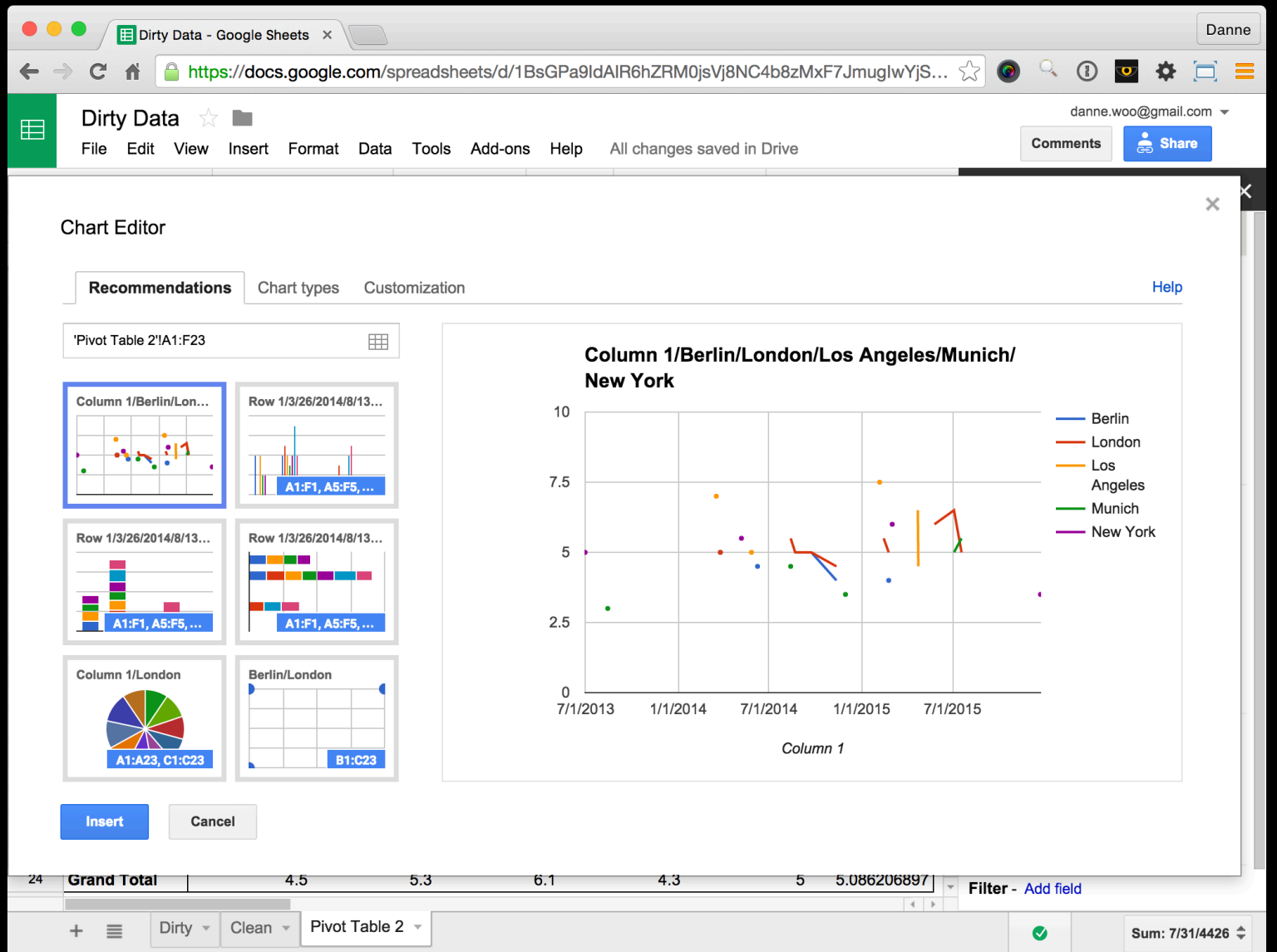
Pivot Tables



	A	B	C	D	E	F
1		Berlin	London	Los Angeles	Munich	New York
2	6/30/2013					
3	8/14/2013					3
4	3/18/2014				7	
5	3/26/2014	5	5			
6	5/7/2014					
7	5/27/2014				5	
8	6/8/2014	4.5				
9	8/13/2014		5.5			4.5
10	8/22/2014		5			
11	9/23/2014	5	5			
12	11/12/2014	4	4.5			
13	11/30/2014					3.5
14	2/6/2015				7.5	
15	2/14/2015		5.5			
16	2/24/2015	4	5			
17	3/3/2015					
18	4/23/2015				6.5	
19	4/24/2015				4.5	
20	5/26/2015		6			
21	7/4/2015		6.5			5
22	7/19/2015		5			5.5
23	12/23/2015					
24	Grand Total	4.5	5.3	6.1	4.3	

Data Analysis

Visualize for Insight

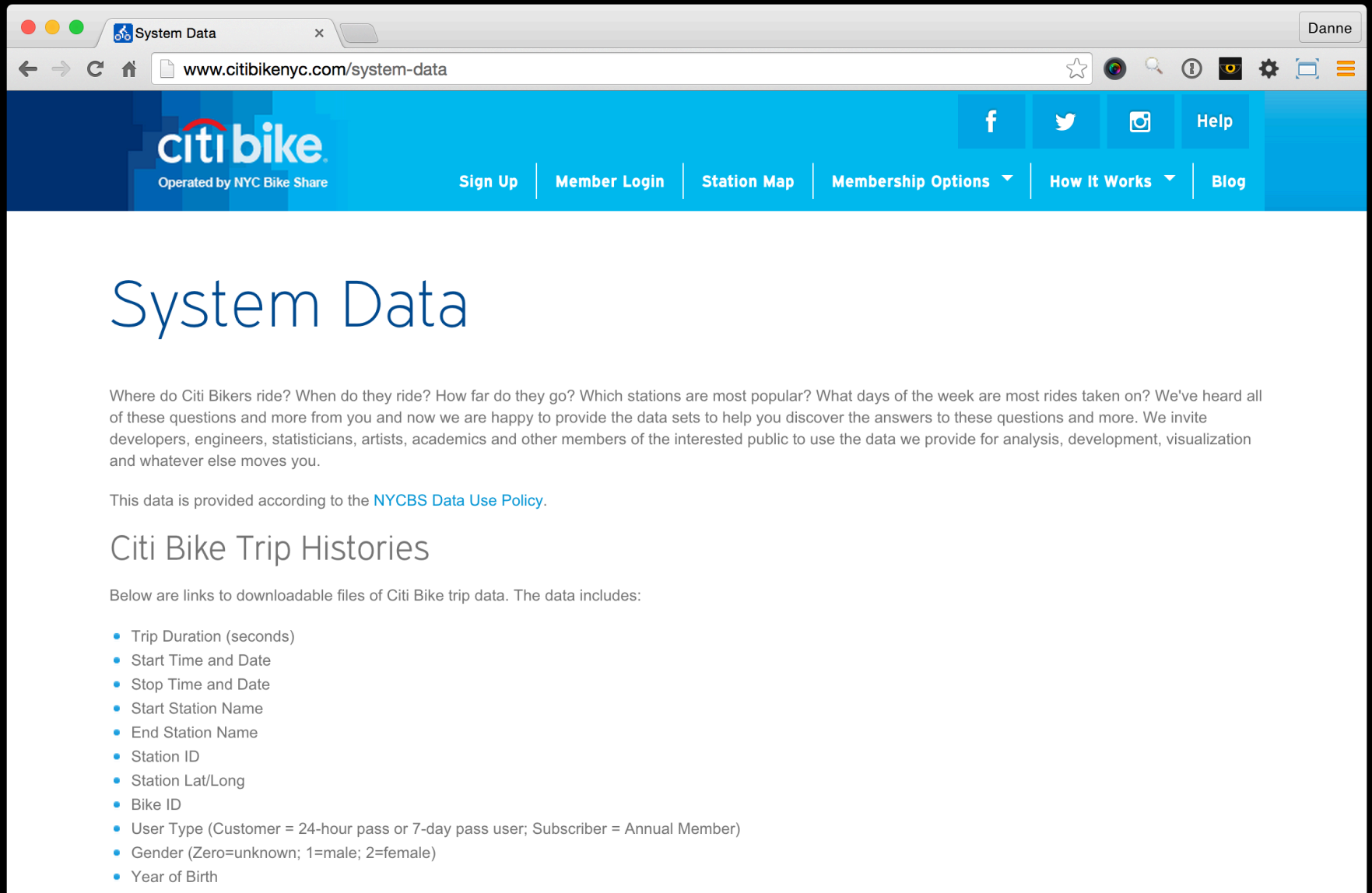


Data Analysis

CitiBike Data / bit.ly/qc-data-citibike

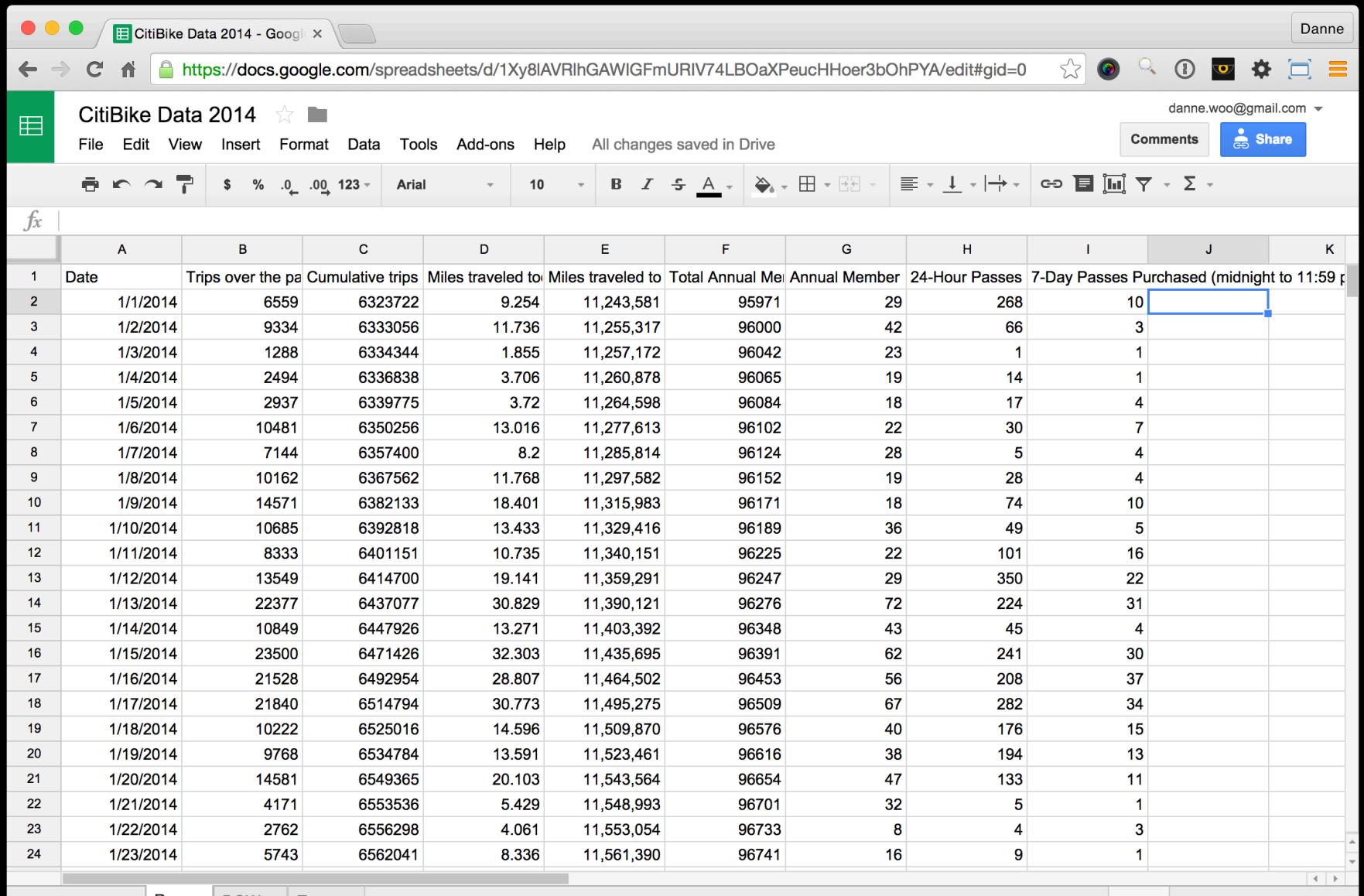
Data Analysis

CitiBike Data - citibikenyc.com/system-data



Data Analysis

CitiBike Data – bit.ly/qc-data-citibike



CitiBike Data 2014 - Google Sheets

https://docs.google.com/spreadsheets/d/1Xy8lAVRlhGAWIGFmURIV74LBOaXPeucHHoer3bOhPYA/edit#gid=0

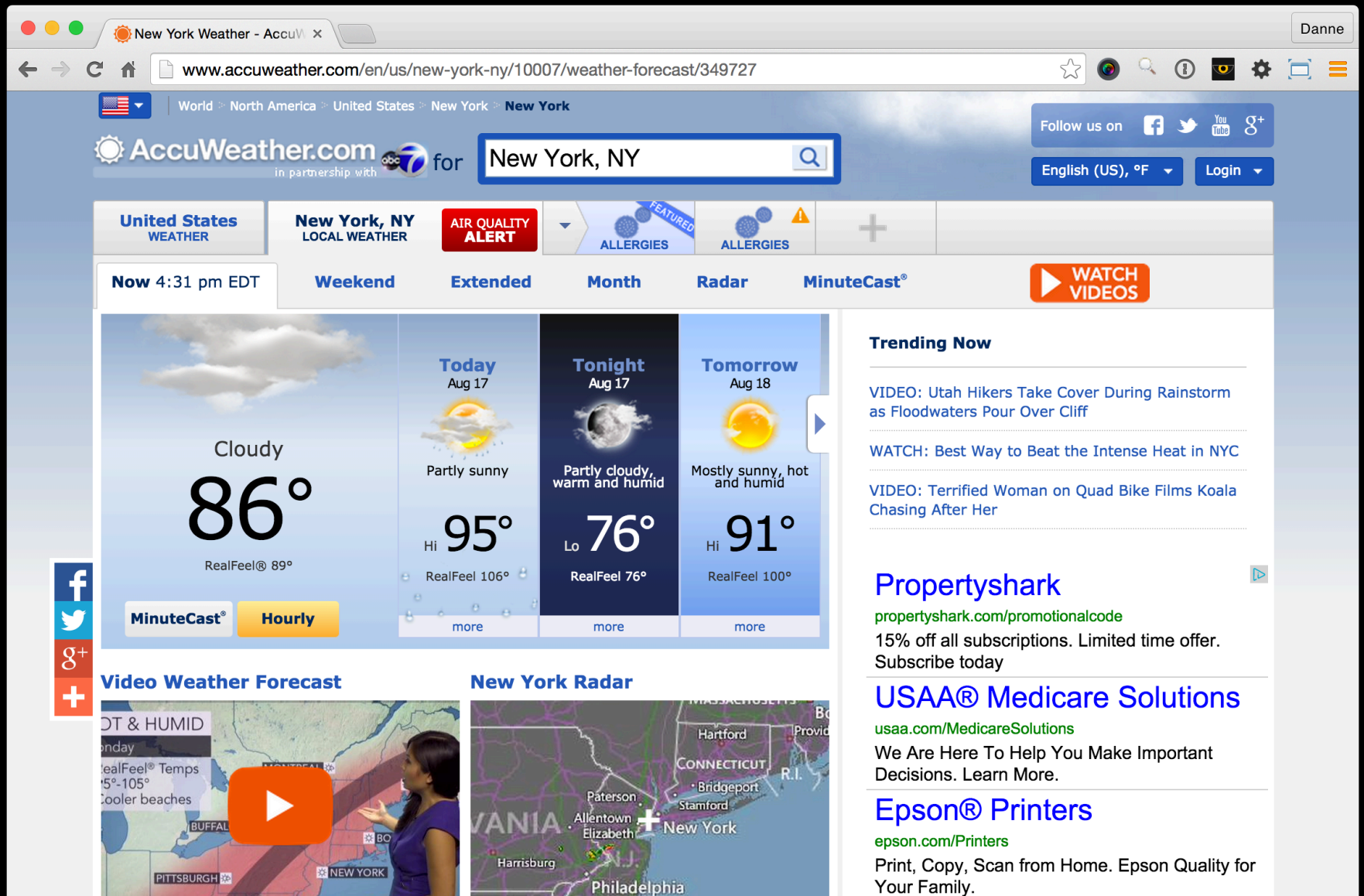
CitiBike Data 2014

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

	A	B	C	D	E	F	G	H	I	J	K
1	Date	Trips over the pa	Cumulative trips	Miles traveled to	Miles traveled to	Total Annual Me	Annual Member	24-Hour Passes	7-Day Passes Purchased (midnight to 11:59 p		
2	1/1/2014	6559	6323722	9.254	11,243,581	95971	29	268	10		
3	1/2/2014	9334	6333056	11.736	11,255,317	96000	42	66	3		
4	1/3/2014	1288	6334344	1.855	11,257,172	96042	23	1	1		
5	1/4/2014	2494	6336838	3.706	11,260,878	96065	19	14	1		
6	1/5/2014	2937	6339775	3.72	11,264,598	96084	18	17	4		
7	1/6/2014	10481	6350256	13.016	11,277,613	96102	22	30	7		
8	1/7/2014	7144	6357400	8.2	11,285,814	96124	28	5	4		
9	1/8/2014	10162	6367562	11.768	11,297,582	96152	19	28	4		
10	1/9/2014	14571	6382133	18.401	11,315,983	96171	18	74	10		
11	1/10/2014	10685	6392818	13.433	11,329,416	96189	36	49	5		
12	1/11/2014	8333	6401151	10.735	11,340,151	96225	22	101	16		
13	1/12/2014	13549	6414700	19.141	11,359,291	96247	29	350	22		
14	1/13/2014	22377	6437077	30.829	11,390,121	96276	72	224	31		
15	1/14/2014	10849	6447926	13.271	11,403,392	96348	43	45	4		
16	1/15/2014	23500	6471426	32.303	11,435,695	96391	62	241	30		
17	1/16/2014	21528	6492954	28.807	11,464,502	96453	56	208	37		
18	1/17/2014	21840	6514794	30.773	11,495,275	96509	67	282	34		
19	1/18/2014	10222	6525016	14.596	11,509,870	96576	40	176	15		
20	1/19/2014	9768	6534784	13.591	11,523,461	96616	38	194	13		
21	1/20/2014	14581	6549365	20.103	11,543,564	96654	47	133	11		
22	1/21/2014	4171	6553536	5.429	11,548,993	96701	32	5	1		
23	1/22/2014	2762	6556298	4.061	11,553,054	96733	8	4	3		
24	1/23/2014	5743	6562041	8.336	11,561,390	96741	16	9	1		

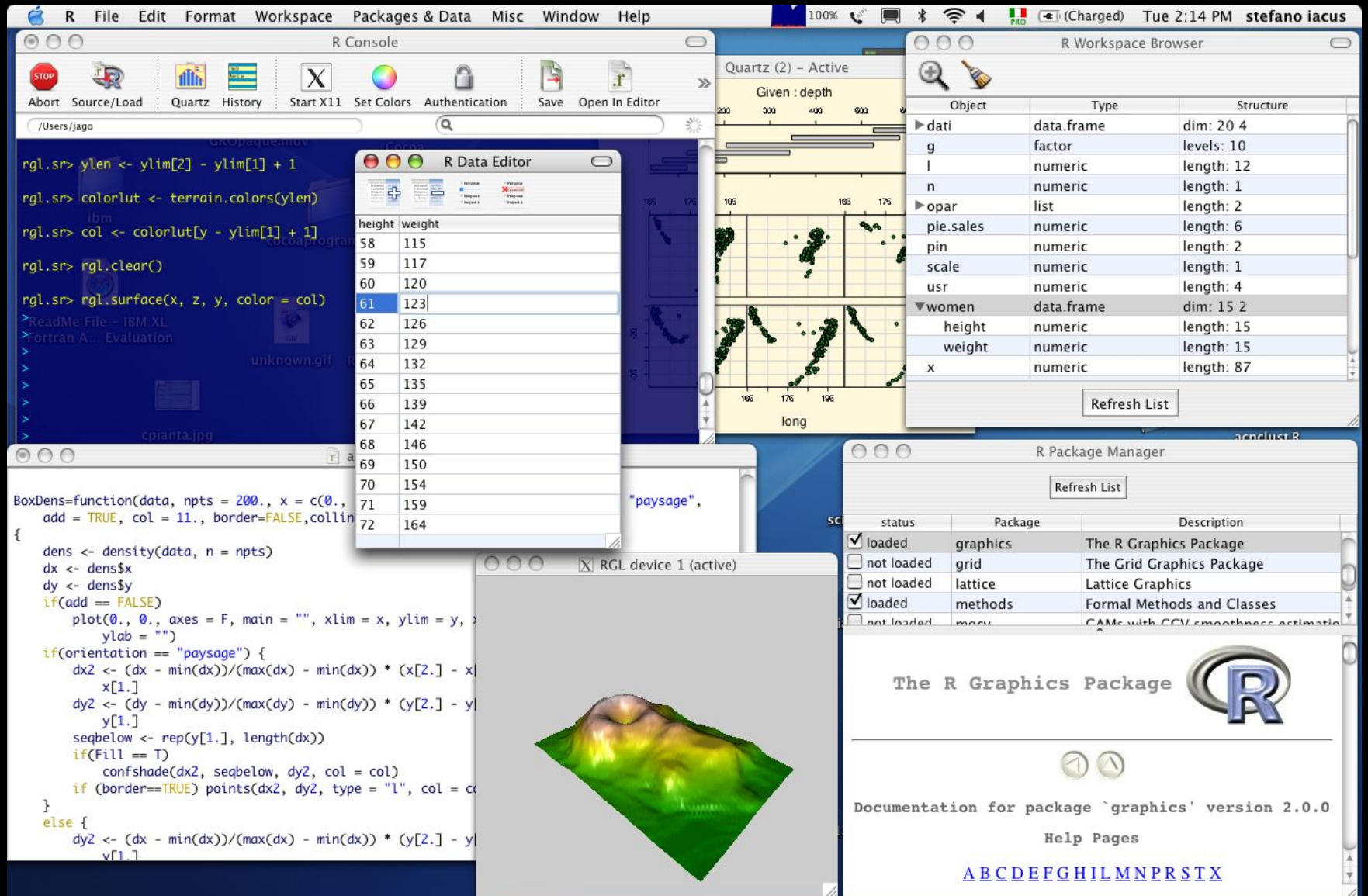
Data Analysis

CitiBike Usage based on temperature?



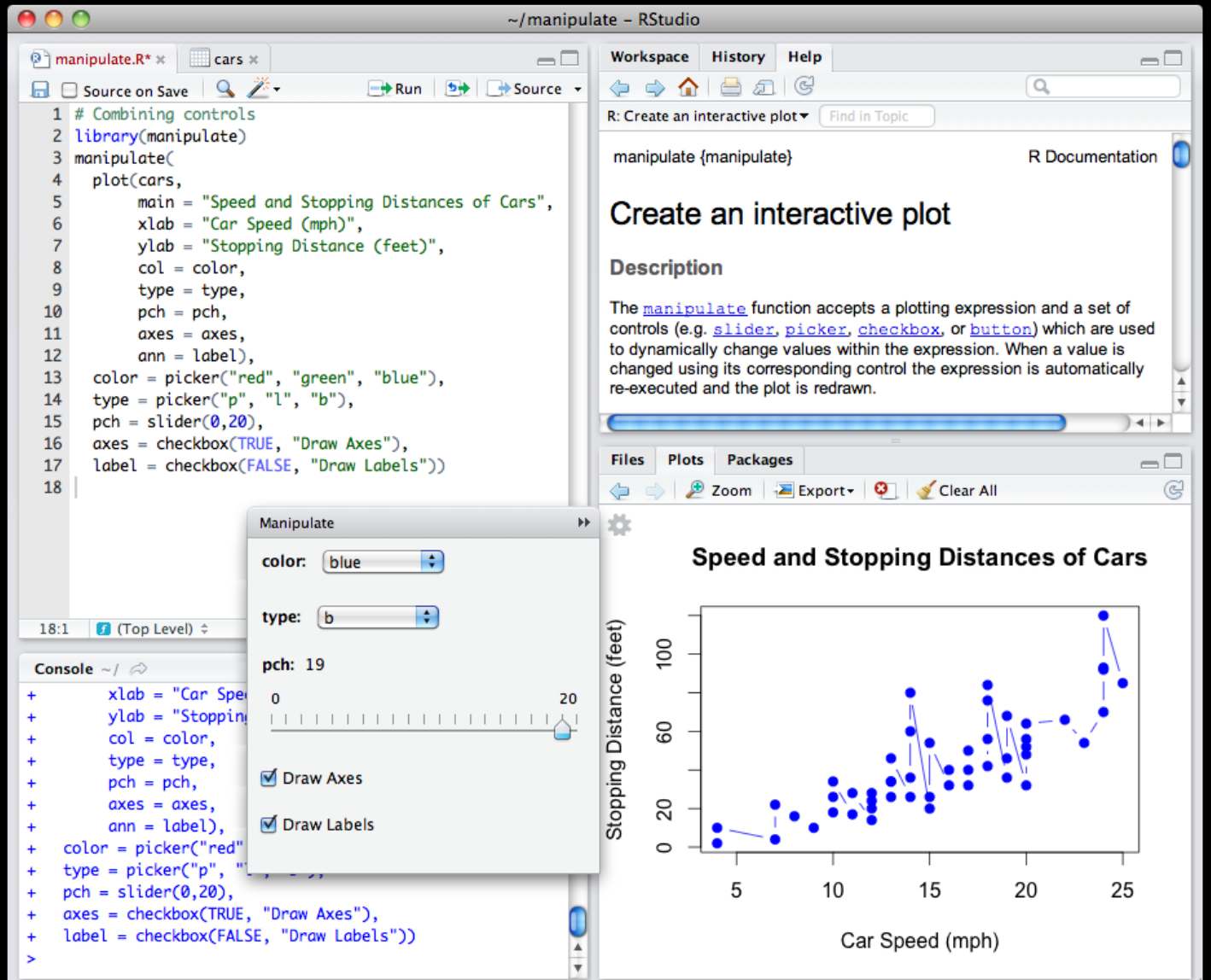
Data Analysis

R



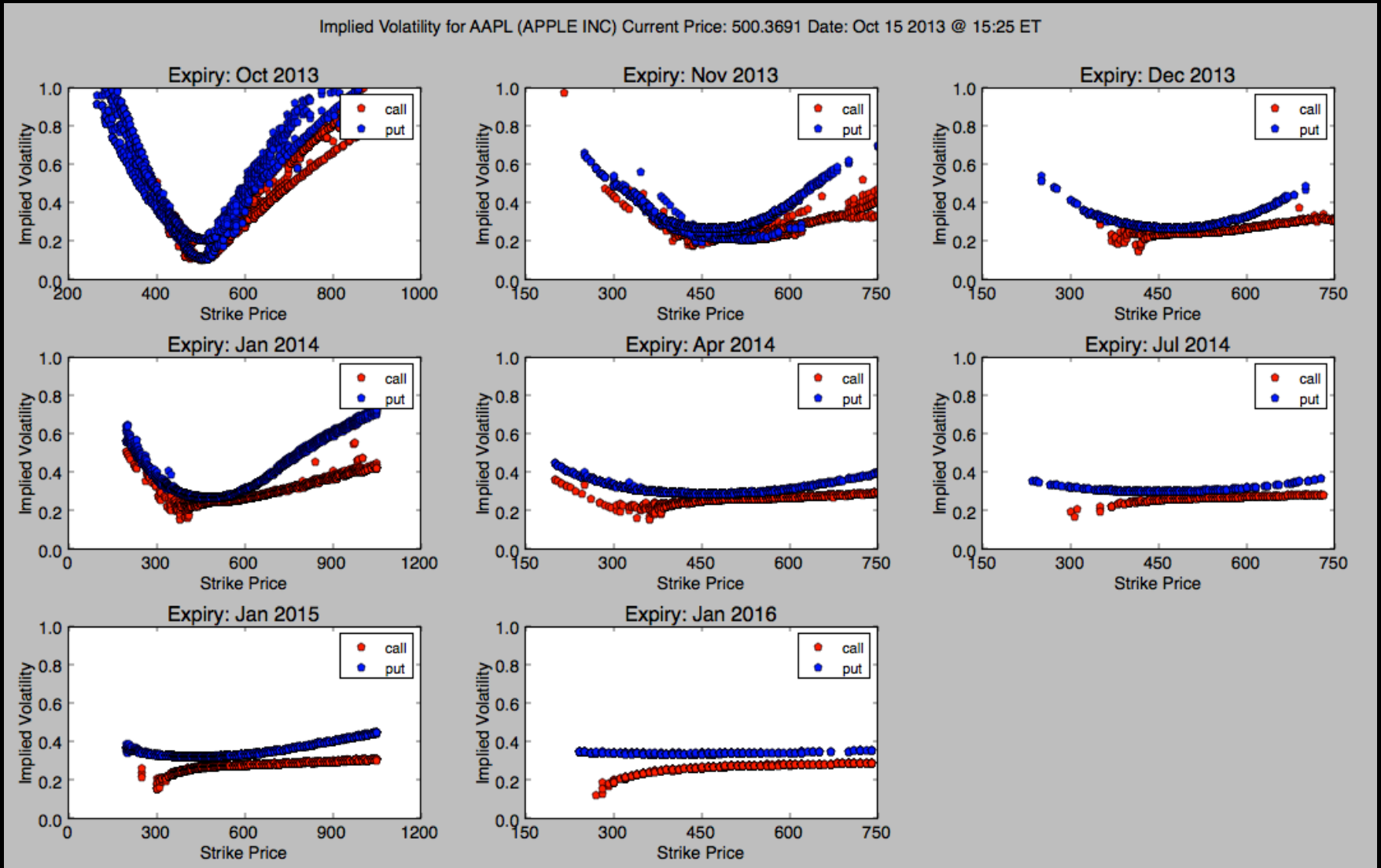
Data Analysis

R Studio



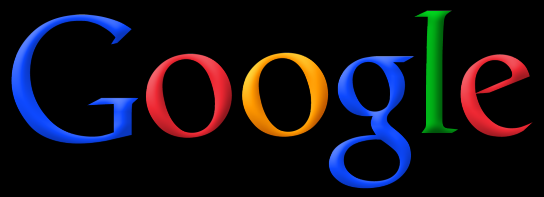
Data Analysis

Python and Pandas



Data Analysis

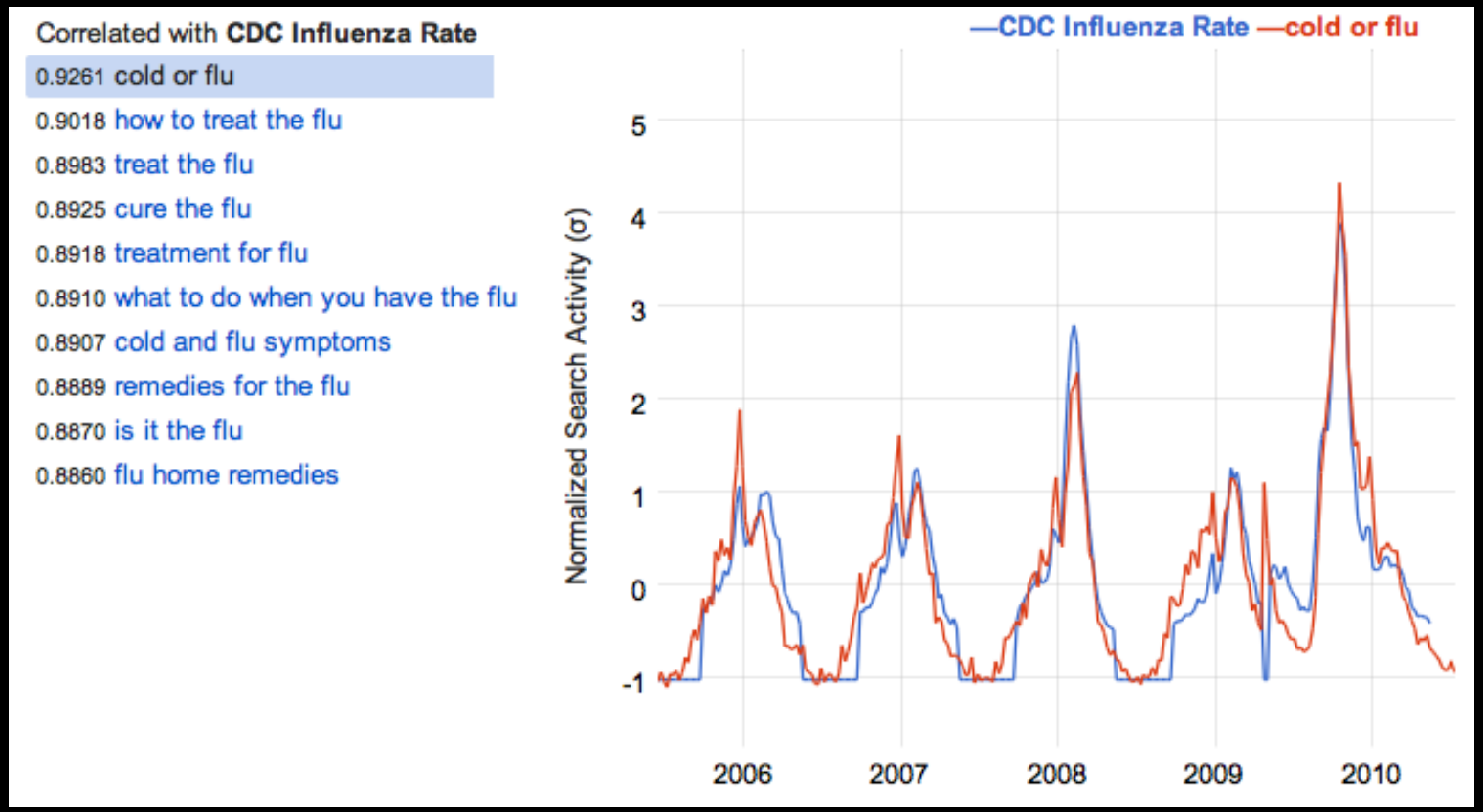
Companies Predicting the Future



Data Analysis

Companies Predicting the Future

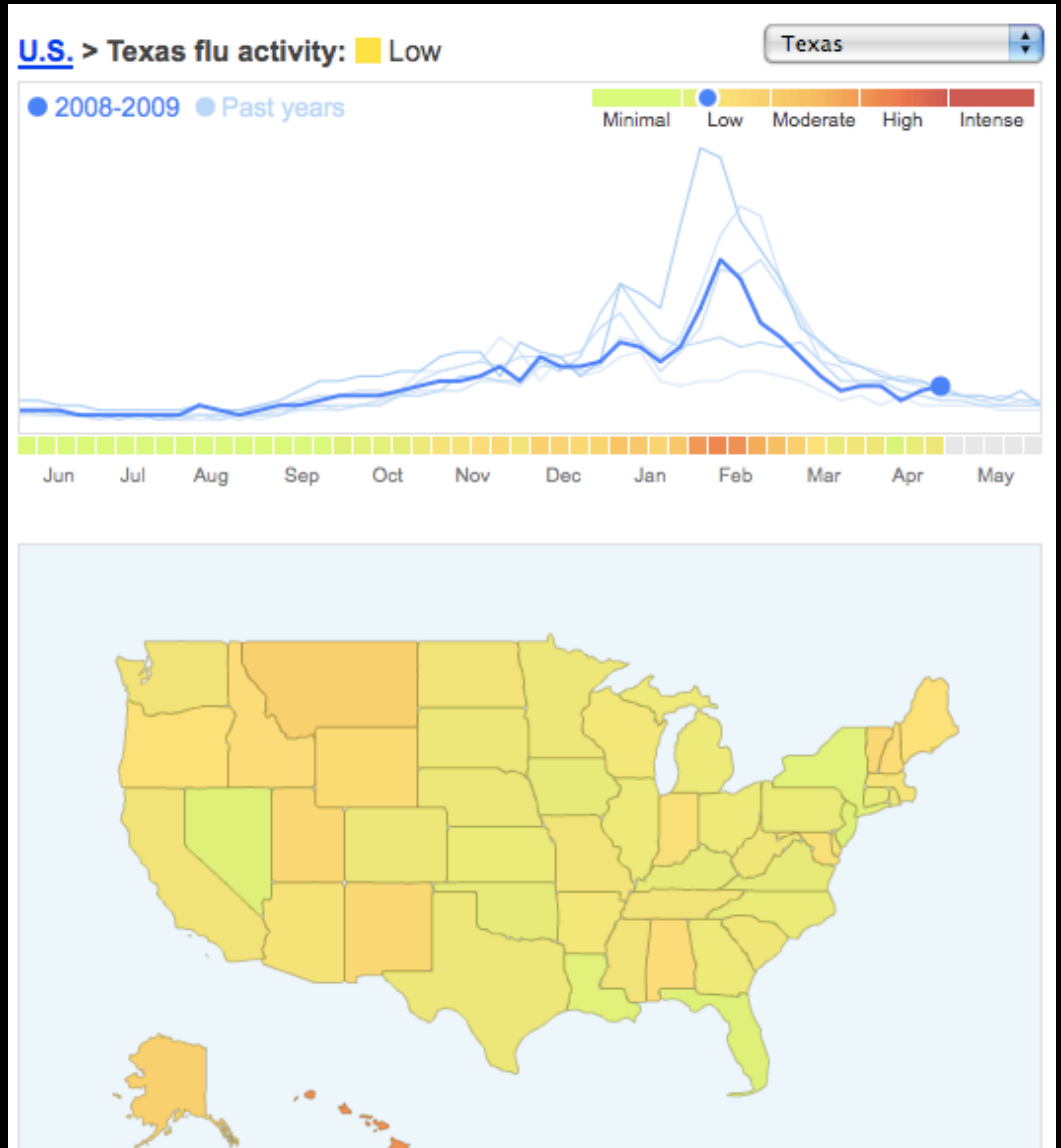
Google



Data Analysis

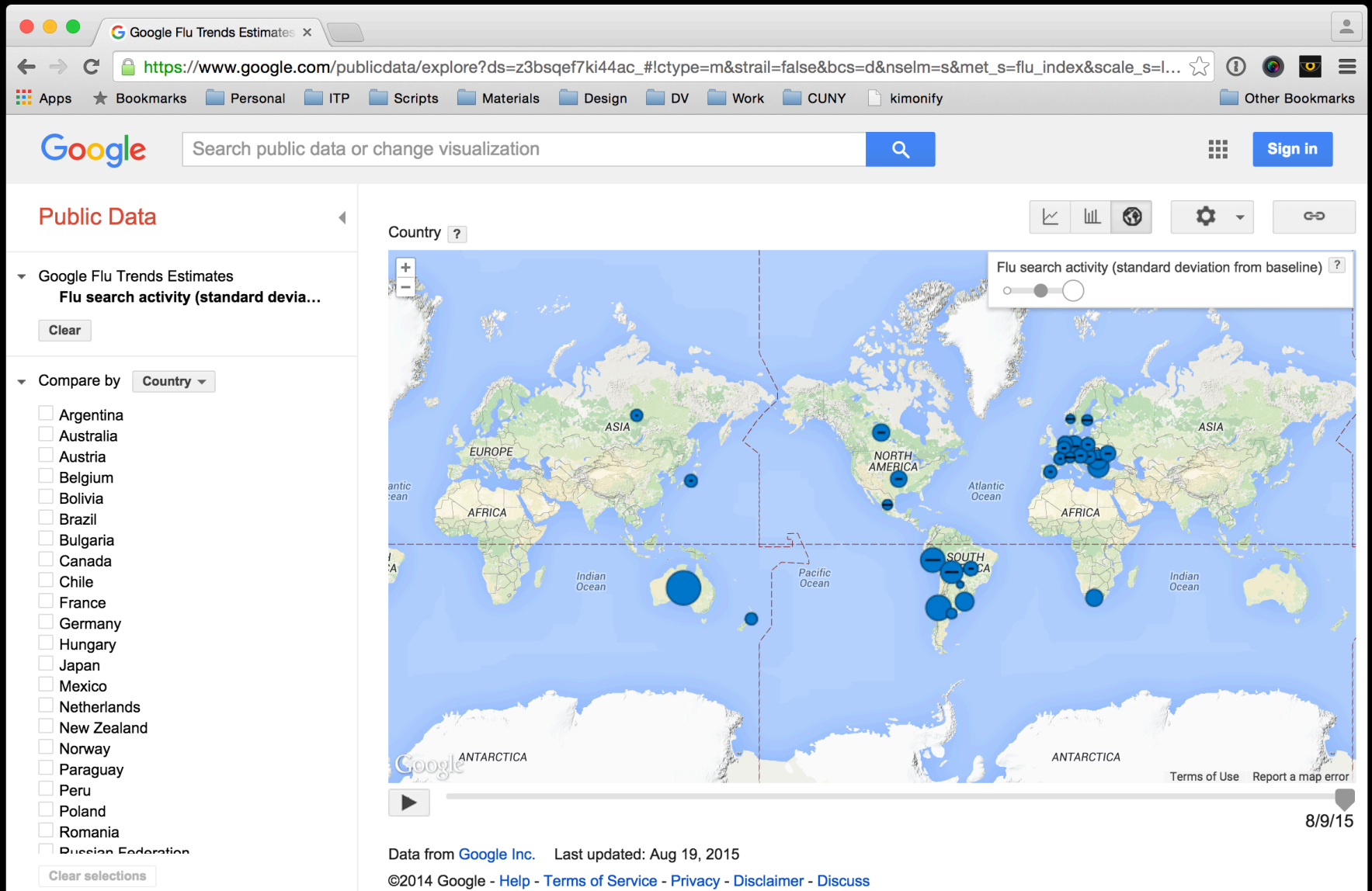
Companies Predicting the Future

Google



Data Analysis

Companies Predicting the Future

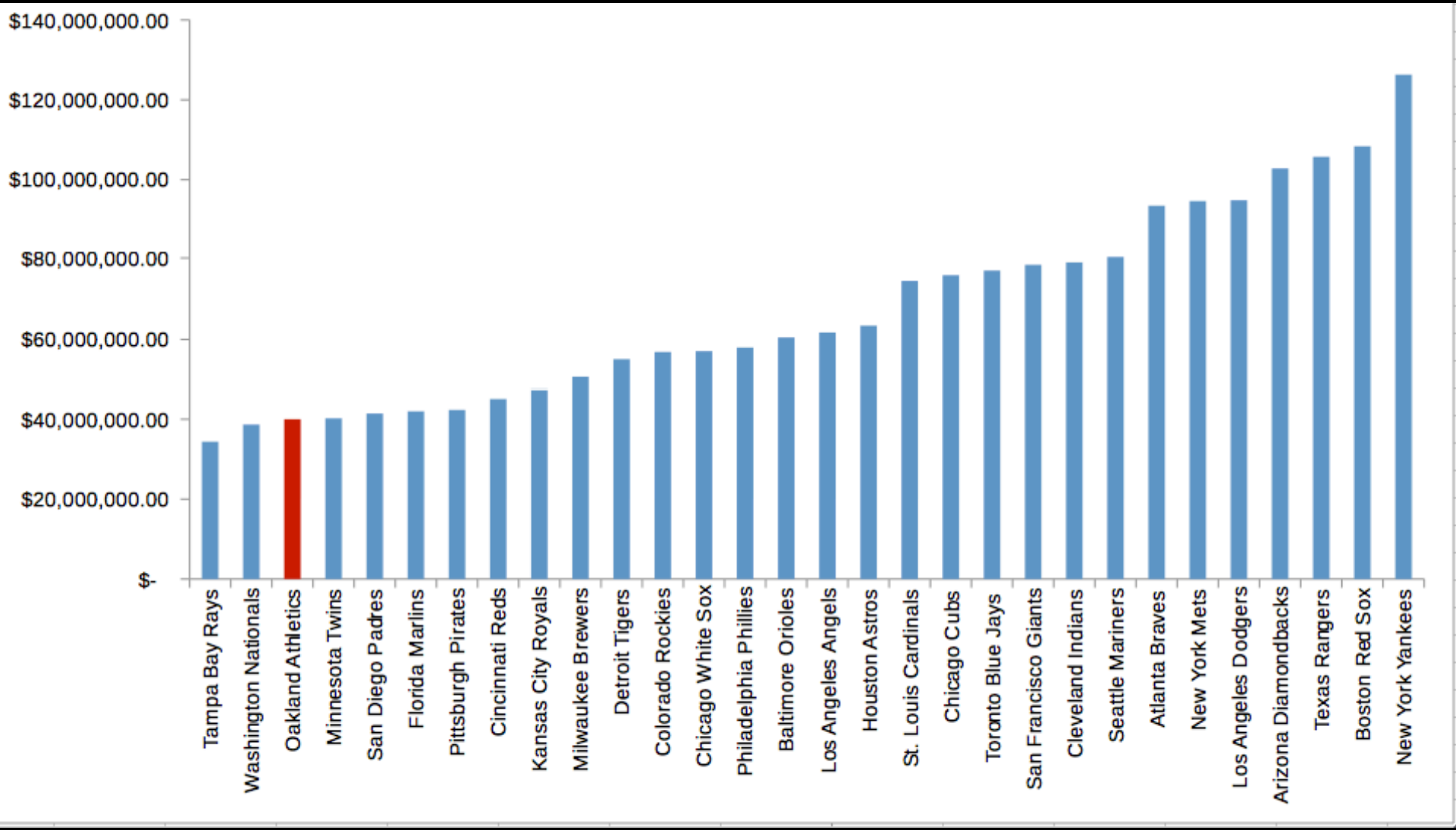


Data Analysis

Companies Predicting the Future

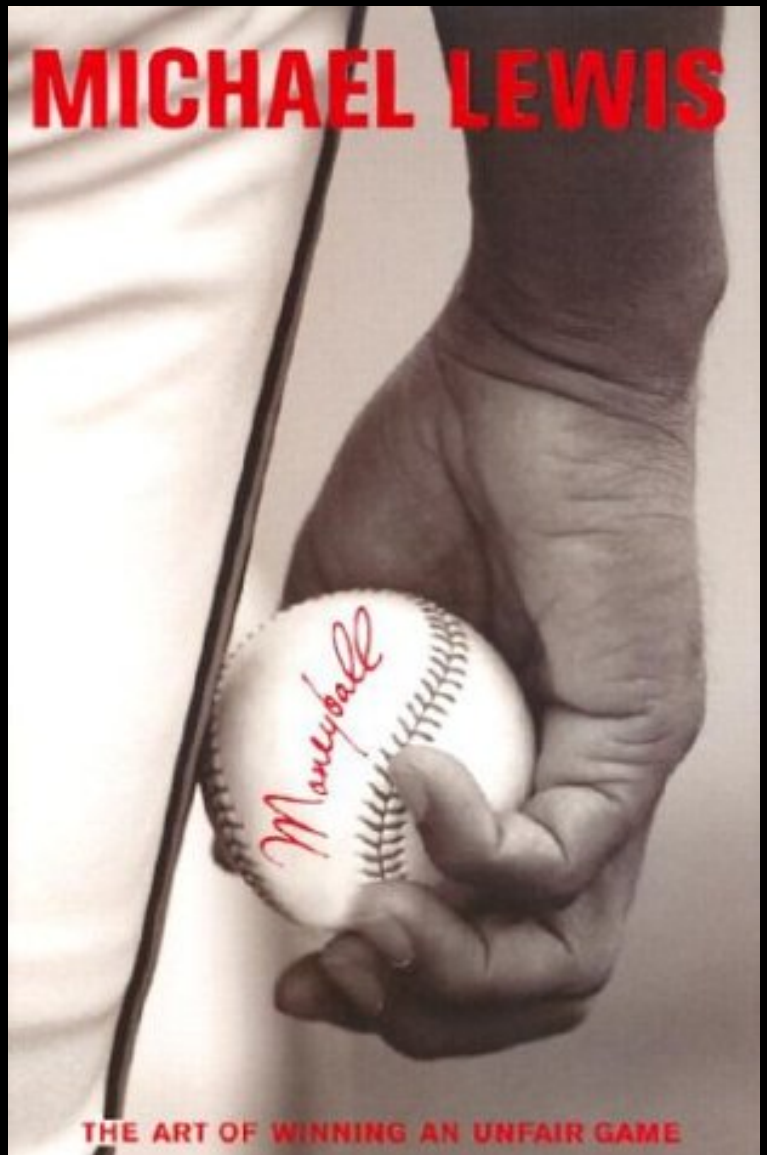
Data Analysis

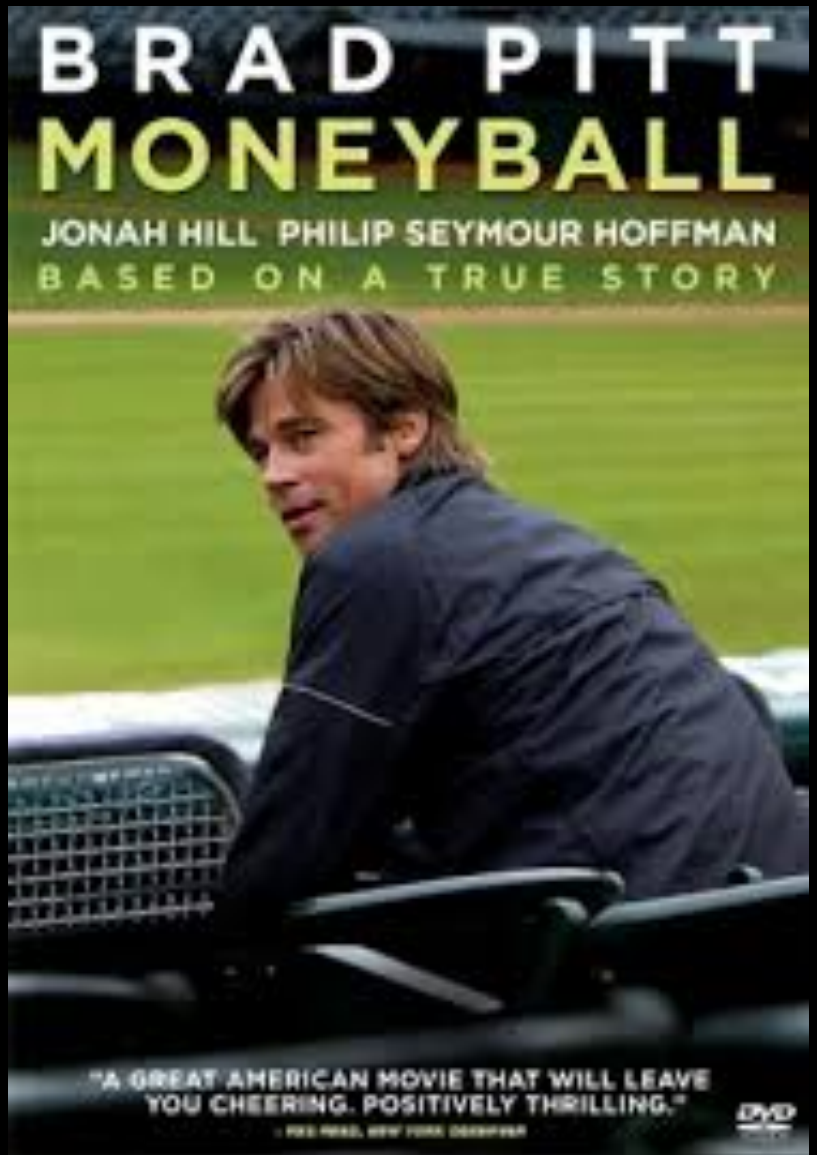
Companies Predicting the Future



Data Analysis

Companies Predicting the Future





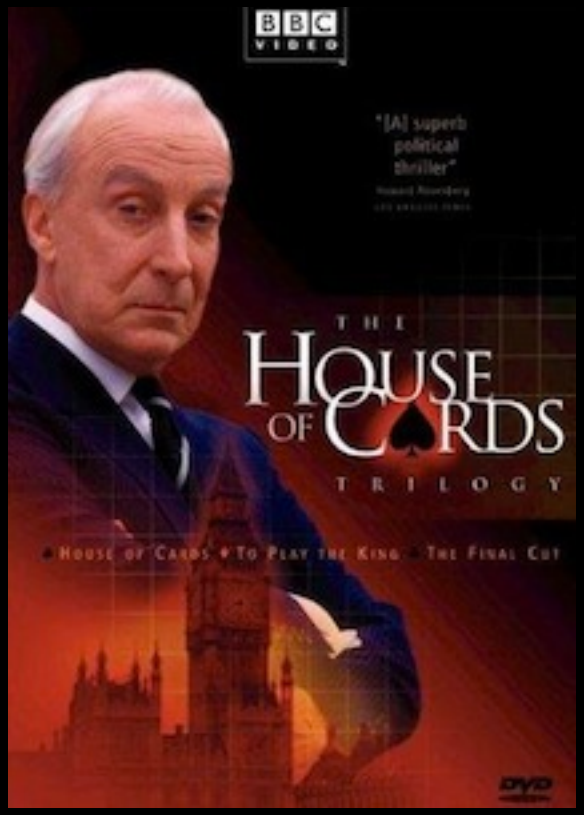
Data Analysis

Companies Predicting the Future

- More than 25 million users
- About 30 million plays per day (and it tracks every time you rewind, fast forward and pause a movie)
- More than 2 billion hours of streaming video watched during the last three months of 2011 alone
- About 4 million ratings per day
- About 3 million searches per day
- Geo-location data
- Device information
- Time of day and week (it now can verify that users watch more TV shows during the week and more movies during the weekend)
- Metadata from third parties such as Nielsen
- Social media data from Facebook and Twitter

Data Analysis

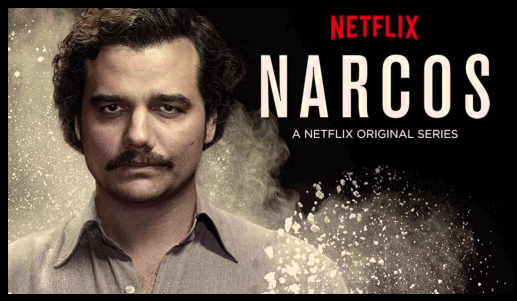
Companies Predicting the Future



- Data of their subscriber viewing preferences showed they would enjoy a remake of the 1990 BBC miniseries, House of Cards.
- Same users who loved this miniseries also loved movies starring Kevin Spacey or directed by David Fincher.
- Because of this they financed a remake of the BBC drama with Spacey and Fincher.
- Targeted advertising before and after launch.

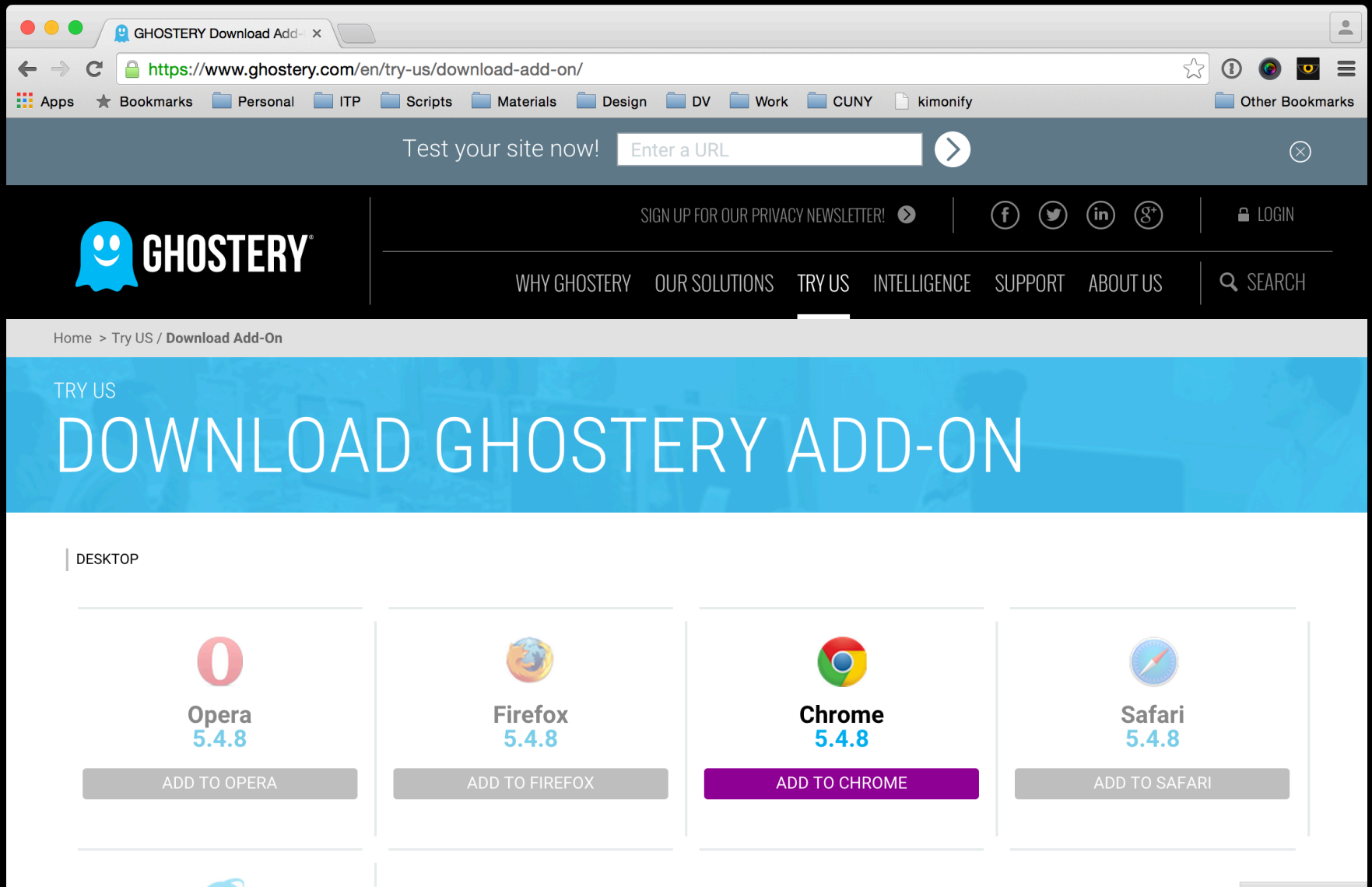
Data Analysis

Companies Predicting the Future



Data Analysis

Know Who is Tracking You / Ghostery



Data Analysis

Know Who is Tracking You / Data Selfie