LO1 – LO2

Examine data mining through social media tracking

Objectives

- Introduction to Data analytics
- Collecting and extracting social media data
- Data analysis, visualization and exploration

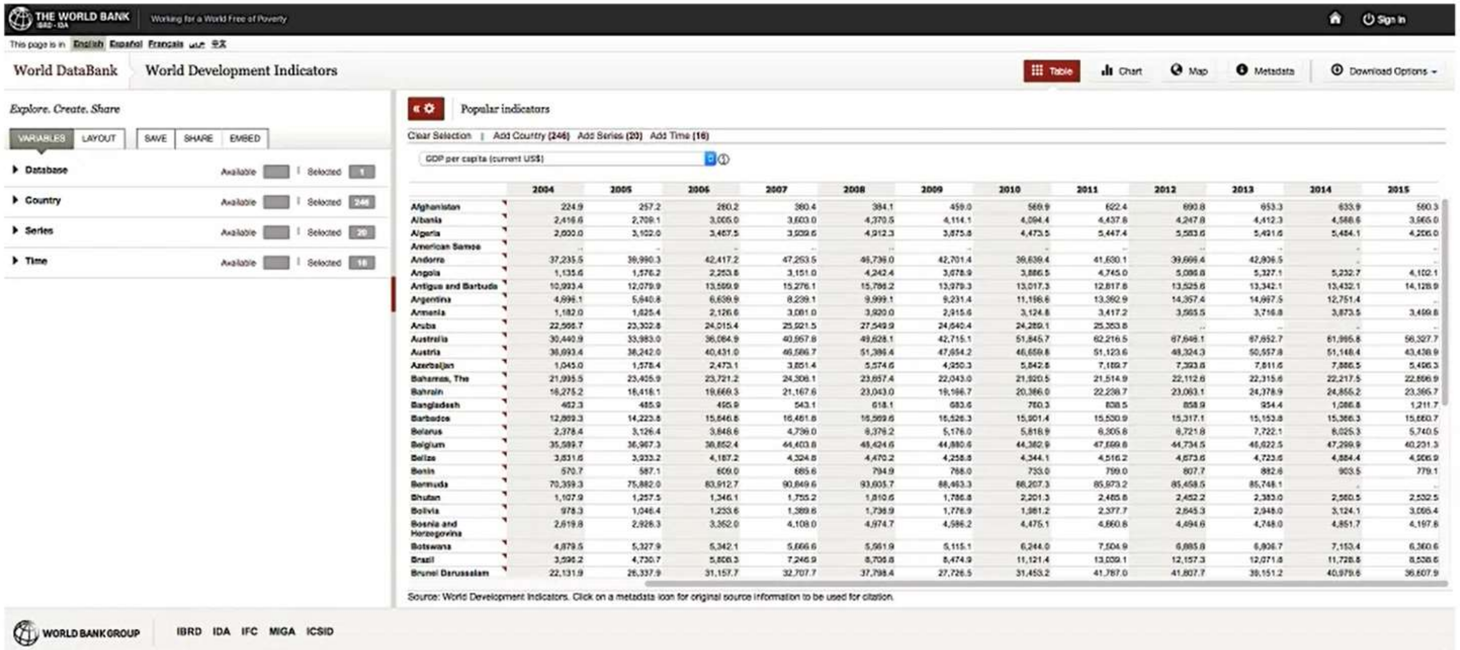
Objective of this session

After attending this session, you should be able

- Difference between structured and unstructured data
- Explain how data analysis is performed on a typical structured dataset
- List some of the techniques of quantitative data analysis

Structured vs unstructured data

- Data can be represented in various types of structures, formats, and media

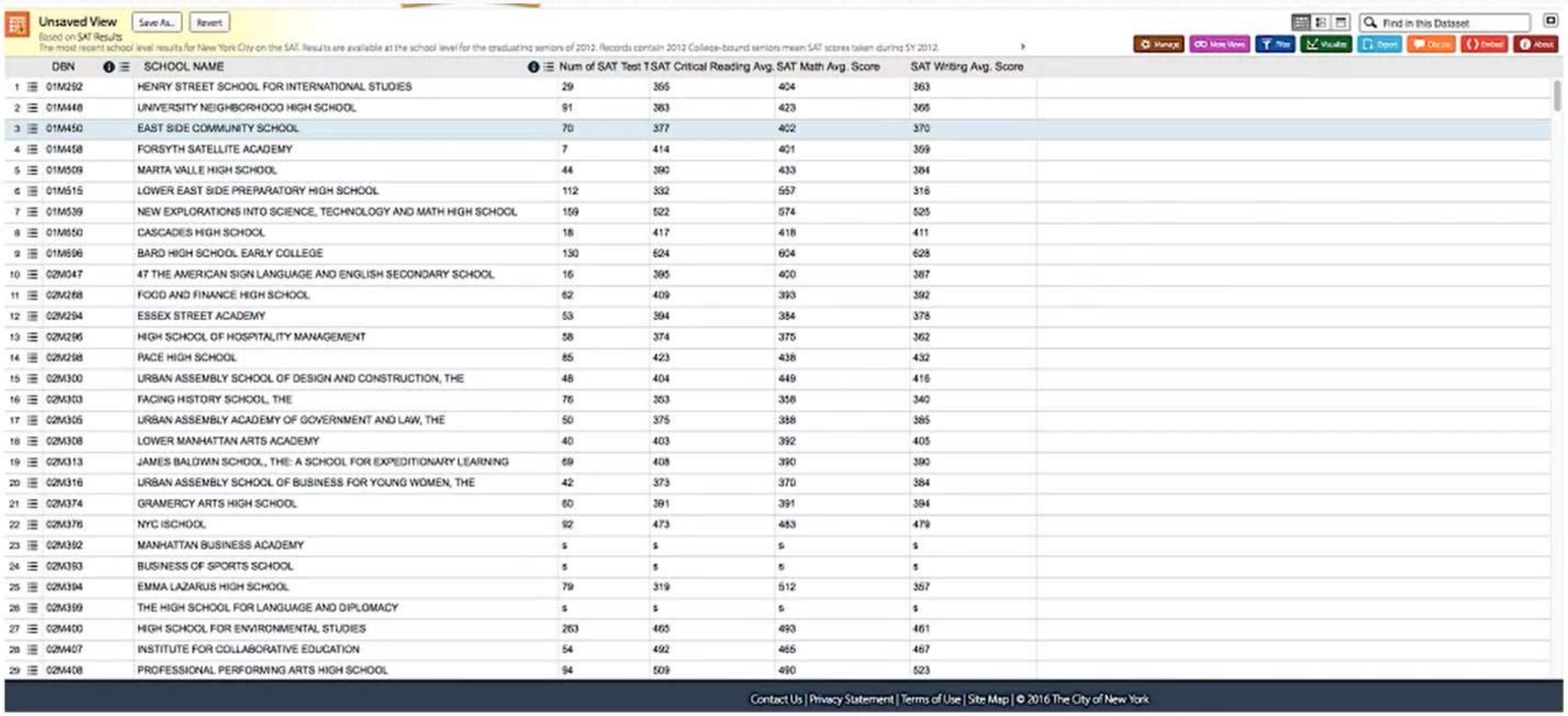


GDP per capita (current US$)

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Afghanistan	224.9	257.2	280.2	380.4	384.1	459.0	569.9	622.4	690.8	653.3	633.9	590.3
Albania	2,416.6	2,709.1	3,005.0	3,603.0	4,370.5	4,114.1	4,094.4	4,437.8	4,247.8	4,412.3	4,588.6	3,965.0
Algeria	2,600.0	3,102.0	3,467.5	3,939.6	4,912.3	3,875.8	4,473.5	5,447.4	5,583.6	5,491.6	5,484.1	4,206.0
American Samoa	..	..	..	..	..	..	..	..	..	..	..	..
Andorra	37,235.5	38,990.3	42,417.2	47,253.5	46,736.0	42,701.4	39,639.4	41,630.1	39,666.4	42,806.5	..	..
Angola	1,135.6	1,576.2	2,253.8	3,151.0	4,242.4	3,678.9	3,886.5	4,745.0	5,086.8	5,327.1	5,232.7	4,102.1
Antigua and Barbuda	10,933.4	12,079.9	13,599.9	15,276.1	15,786.2	13,979.3	13,017.3	12,817.8	13,525.6	13,342.1	13,432.1	14,128.9
Argentina	4,896.1	5,640.8	6,639.9	8,239.1	9,999.1	9,231.4	11,198.6	13,392.9	14,357.4	14,997.5	12,751.4	..
Armenia	1,182.0	1,625.4	2,126.6	3,081.0	3,920.0	2,915.6	3,124.8	3,417.2	3,565.5	3,716.8	3,873.5	3,489.8
Aruba	22,566.7	23,302.8	24,015.4	25,921.5	27,549.9	24,640.4	24,289.1	25,353.8	..	..	..	..
Australia	30,440.9	33,983.0	36,084.9	40,957.8	49,628.1	42,715.1	51,845.7	62,216.5	67,646.1	67,652.7	61,995.8	56,327.7
Austria	36,893.4	38,242.0	40,431.0	46,586.7	51,386.4	47,654.2	46,659.8	51,123.6	48,324.3	50,557.8	51,148.4	43,438.9
Azerbaijan	1,045.0	1,578.4	2,473.1	3,851.4	5,574.6	4,950.3	5,842.8	7,189.7	7,393.8	7,811.6	7,886.5	5,496.3
Bahamas, The	21,935.5	23,405.9	23,721.2	24,306.1	23,657.4	22,043.0	21,920.5	21,514.9	22,112.6	22,315.6	22,217.5	22,896.9
Bahrain	16,275.2	18,418.1	19,669.3	21,167.6	23,043.0	19,166.7	20,386.0	22,238.7	23,063.1	24,378.9	24,855.2	23,395.7
Bangladesh	462.3	485.9	495.9	543.1	618.1	683.6	760.3	838.5	858.9	954.4	1,086.8	1,211.7
Barbados	12,869.3	14,223.8	15,646.8	16,461.8	16,569.6	15,526.3	15,901.4	15,530.9	15,317.1	15,153.8	15,366.3	15,660.7
Belarus	2,378.4	3,126.4	3,848.6	4,736.0	6,376.2	5,176.0	5,818.9	6,305.8	6,721.8	7,722.1	8,025.3	5,740.5
Belgium	35,589.7	36,967.3	38,852.4	44,403.8	48,424.6	44,880.6	44,362.9	47,699.8	44,734.5	46,622.5	47,299.9	40,231.3
Belize	3,831.6	3,933.2	4,187.2	4,324.8	4,470.2	4,258.8	4,344.1	4,516.2	4,673.6	4,723.6	4,884.4	4,906.9
Benin	570.7	587.1	609.0	685.6	794.9	768.0	733.0	799.0	807.7	882.6	903.5	779.1
Bermuda	70,359.3	75,882.0	83,912.7	90,849.6	93,605.7	88,463.3	88,207.3	85,973.2	85,458.5	85,748.1	..	..
Bhutan	1,107.9	1,257.5	1,346.1	1,755.2	1,810.6	1,786.8	2,201.3	2,485.8	2,452.2	2,383.0	2,560.5	2,532.5
Bolivia	978.3	1,046.4	1,233.6	1,389.6	1,736.9	1,776.9	1,981.2	2,377.7	2,645.3	2,948.0	3,124.1	3,095.4
Bosnia and Herzegovina	2,619.8	2,928.3	3,352.0	4,108.0	4,974.7	4,586.2	4,475.1	4,660.8	4,494.6	4,748.0	4,851.7	4,197.8
Botswana	4,879.5	5,327.9	5,342.1	5,666.6	5,561.9	5,115.1	6,244.0	7,504.9	6,885.8	6,806.7	7,153.4	6,360.6
Brazil	3,596.2	4,730.7	5,808.3	7,246.9	8,706.8	8,474.9	11,121.4	13,039.1	12,157.3	12,071.8	11,728.8	8,538.6
Brunei Darussalam	22,131.9	26,337.9	31,157.7	32,707.7	37,798.4	27,726.5	31,453.2	41,787.0	41,807.7	39,151.2	40,979.6	36,607.9

Source: World Development Indicators. Click on a metadata icon for original source information to be used for citation.

GDP per capita from world bank



Unsaved View

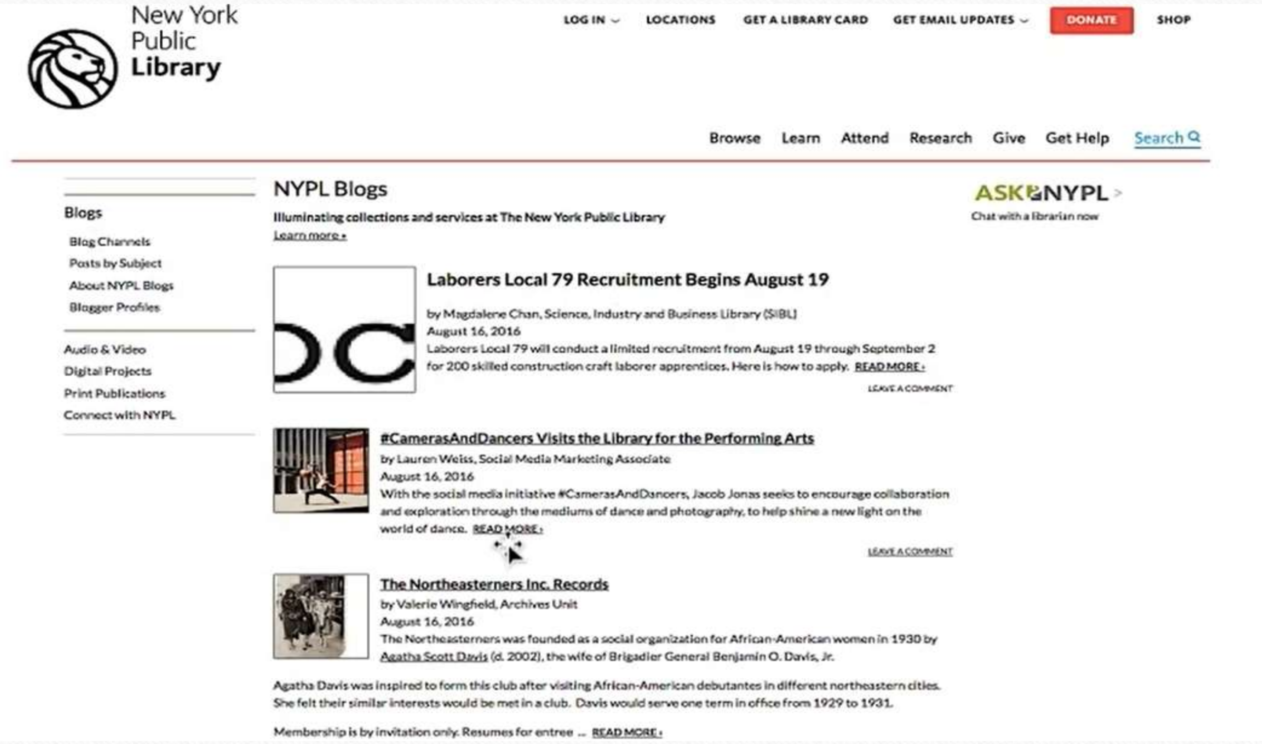
Based on SAT Results

The most recent school level results for New York City on the SAT. Results are available at the school level for the graduating seniors of 2012. Records contain 2012 College-bound seniors mean SAT scores taken during SY 2012.

	DBN	SCHOOL NAME	Num of SAT Test T	SAT Critical Reading Avg.	SAT Math Avg. Score	SAT Writing Avg. Score
1	01M292	HENRY STREET SCHOOL FOR INTERNATIONAL STUDIES	29	355	404	363
2	01M448	UNIVERSITY NEIGHBORHOOD HIGH SCHOOL	91	383	423	366
3	01M450	EAST SIDE COMMUNITY SCHOOL	70	377	402	370
4	01M458	FORSYTH SATELLITE ACADEMY	7	414	401	359
5	01M509	MARTA VALLE HIGH SCHOOL	44	390	433	384
6	01M515	LOWER EAST SIDE PREPARATORY HIGH SCHOOL	112	332	557	316
7	01M539	NEW EXPLORATIONS INTO SCIENCE, TECHNOLOGY AND MATH HIGH SCHOOL	159	522	574	525
8	01M650	CASCADES HIGH SCHOOL	18	417	418	411
9	01M696	BARD HIGH SCHOOL EARLY COLLEGE	130	624	604	628
10	02M047	47 THE AMERICAN SIGN LANGUAGE AND ENGLISH SECONDARY SCHOOL	16	395	400	387
11	02M288	FOOD AND FINANCE HIGH SCHOOL	62	409	393	392
12	02M294	ESSEX STREET ACADEMY	53	394	384	378
13	02M296	HIGH SCHOOL OF HOSPITALITY MANAGEMENT	58	374	375	362
14	02M298	PACE HIGH SCHOOL	85	423	438	432
15	02M300	URBAN ASSEMBLY SCHOOL OF DESIGN AND CONSTRUCTION, THE	48	404	449	416
16	02M303	FACING HISTORY SCHOOL, THE	76	353	358	340
17	02M305	URBAN ASSEMBLY ACADEMY OF GOVERNMENT AND LAW, THE	50	375	388	385
18	02M308	LOWER MANHATTAN ARTS ACADEMY	40	403	392	405
19	02M313	JAMES BALDWIN SCHOOL, THE: A SCHOOL FOR EXPEDITIONARY LEARNING	69	408	390	390
20	02M316	URBAN ASSEMBLY SCHOOL OF BUSINESS FOR YOUNG WOMEN, THE	42	373	370	384
21	02M374	GRAMERCY ARTS HIGH SCHOOL	60	391	391	394
22	02M376	NYC ISCHOOL	92	473	483	479
23	02M392	MANHATTAN BUSINESS ACADEMY	s	s	s	s
24	02M393	BUSINESS OF SPORTS SCHOOL	s	s	s	s
25	02M394	EMMA LAZARUS HIGH SCHOOL	79	319	512	357
26	02M399	THE HIGH SCHOOL FOR LANGUAGE AND DIPLOMACY	s	s	s	s
27	02M400	HIGH SCHOOL FOR ENVIRONMENTAL STUDIES	263	465	493	461
28	02M407	INSTITUTE FOR COLLABORATIVE EDUCATION	54	492	465	467
29	02M408	PROFESSIONAL PERFORMING ARTS HIGH SCHOOL	94	509	490	523

Contact Us | Privacy Statement | Terms of Use | Site Map | © 2016 The City of New York

SAT data from NYC open data (nycopendata.Socrata.com)



NYPL Blogs

Illuminating collections and services at The New York Public Library
Learn more ›

Laborers Local 79 Recruitment Begins August 19

by Magdalene Chan, Science, Industry and Business Library (SIBL)
August 16, 2016
Laborers Local 79 will conduct a limited recruitment from August 19 through September 2 for 200 skilled construction craft laborer apprentices. Here is how to apply. READ MORE ›

LEAVE A COMMENT

#CamerasAndDancers Visits the Library for the Performing Arts

by Lauren Weiss, Social Media Marketing Associate
August 16, 2016
With the social media initiative #CamerasAndDancers, Jacob Jonas seeks to encourage collaboration and exploration through the mediums of dance and photography, to help shine a new light on the world of dance. READ MORE ›

LEAVE A COMMENT

The Northeasterners Inc. Records

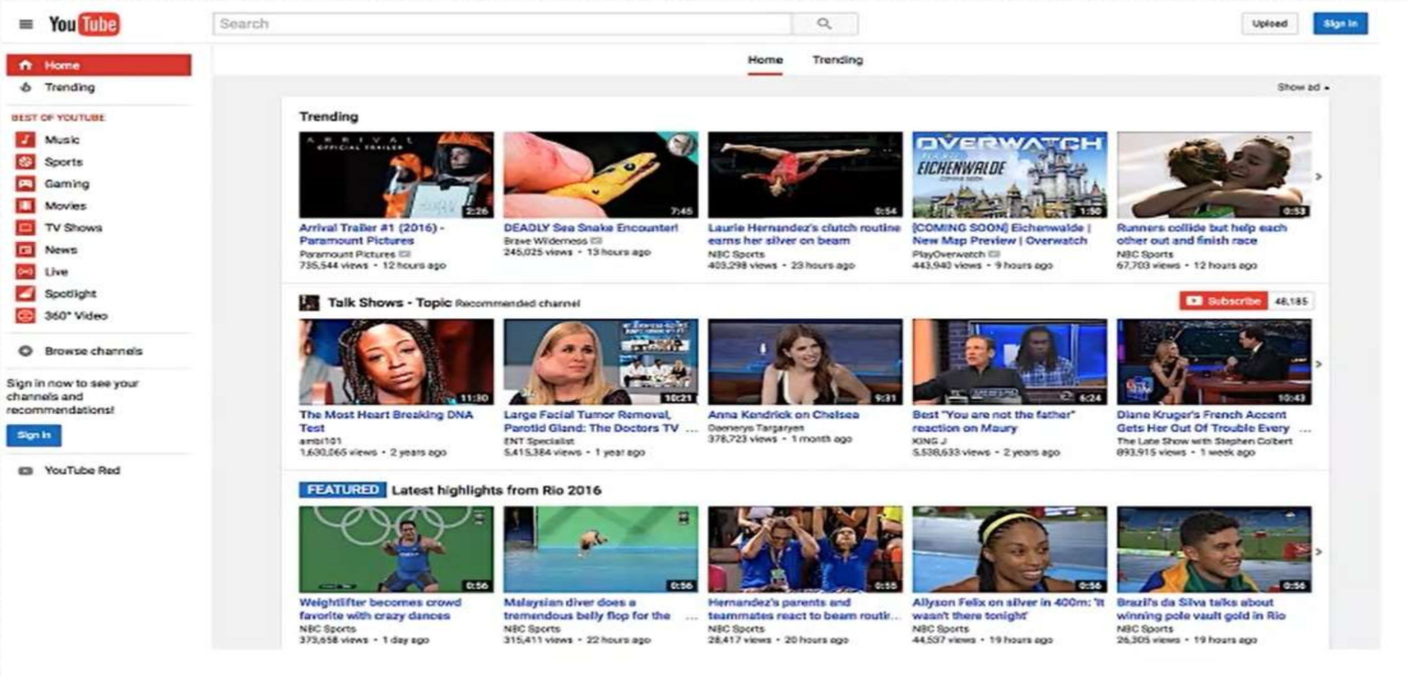
by Valerie Wingfield, Archives Unit
August 16, 2016
The Northeasterners was founded as a social organization for African-American women in 1930 by Agatha Scott Davis (d. 2002), the wife of Brigadier General Benjamin O. Davis, Jr.

Agatha Davis was inspired to form this club after visiting African-American debutantes in different northeastern cities. She felt their similar interests would be met in a club. Davis would serve one term in office from 1929 to 1931.

Membership is by invitation only. Resumes for entree ... READ MORE ›

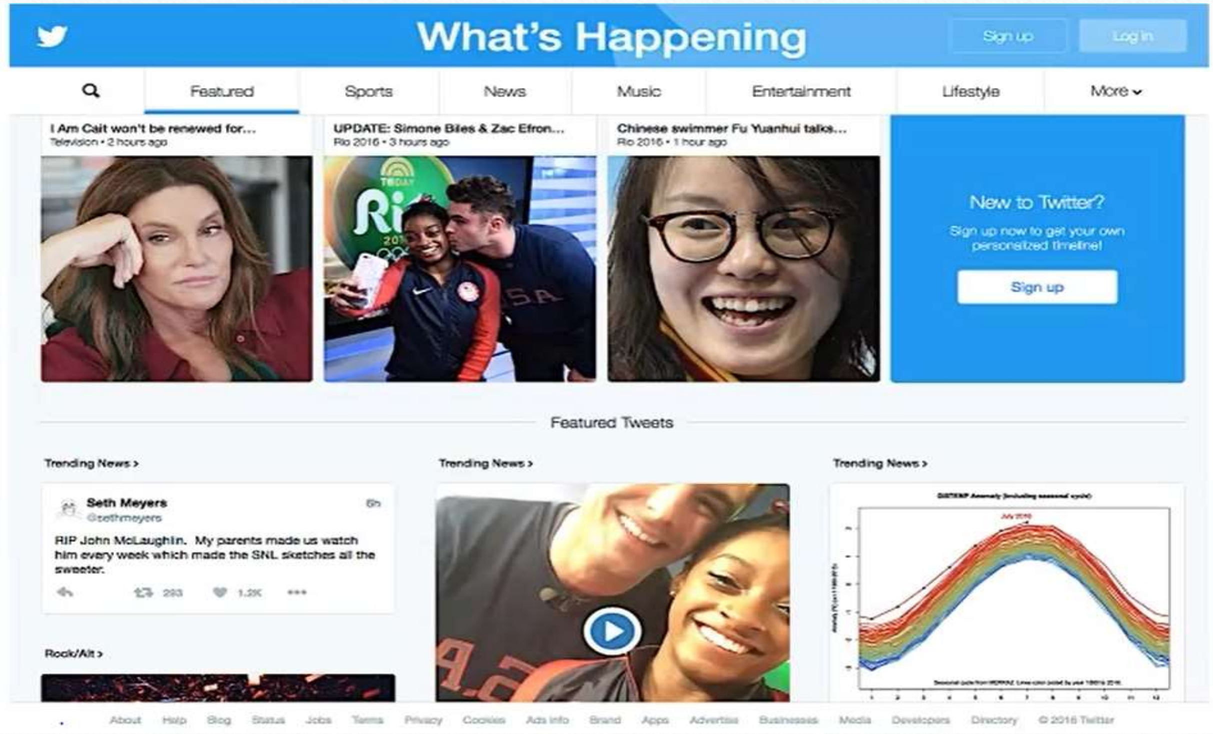
Blog of new York public library (https://www.nypl.org/blog)

Multimedia data


YouTube
Search
Upload
Sign in
Home
Trending
BEST OF YOUTUBE
Music
Sports
Gaming
Movies
TV Shows
News
Live
Spotlight
360° Video
Browse channels
Sign in now to see your channels and recommendations!
Sign in
YouTube Red
Home
Trending
Show ad
Trending
Arrival Trailer #1 (2016) - Paramount Pictures
Paramount Pictures
735,544 views · 12 hours ago
DEADLY Sea Snake Encounter!
Brave Wilderness
245,025 views · 13 hours ago
Laurie Hernandez's clutch routine earns her silver on beam
NBC Sports
403,298 views · 23 hours ago
[COMING SOON] Eichenwalde | New Map Preview | Overwatch
PlayOverwatch
443,940 views · 9 hours ago
Runners collide but help each other out and finish race
NBC Sports
67,703 views · 12 hours ago
Talk Shows - Topic Recommended channel
Subscribe 46,185
The Most Heart Breaking DNA Test
ambi101
1,630,065 views · 2 years ago
Large Facial Tumor Removal, Parotid Gland: The Doctors TV ...
ENT Specialist
5,415,384 views · 1 year ago
Anna Kendrick on Chelsea
Daenerys Targaryen
378,723 views · 1 month ago
Best "You are not the father" reaction on Maury
KING J
5,538,633 views · 2 years ago
Diane Kruger's French Accent Gets Her Out Of Trouble Every ...
The Late Show with Stephen Colbert
893,915 views · 1 week ago
FEATURED Latest highlights from Rio 2016
Weightlifter becomes crowd favorite with crazy dances
NBC Sports
373,658 views · 1 day ago
Malaysian diver does a tremendous belly flop for the ...
NBC Sports
315,411 views · 22 hours ago
Hernandez's parents and teammates react to beam routin...
NBC Sports
28,417 views · 20 hours ago
Allyson Felix on silver in 400m: 'It wasn't there tonight'
NBC Sports
44,537 views · 19 hours ago
Brazil's da Silva talks about winning pole vault gold in Rio
NBC Sports
26,305 views · 19 hours ago

Multimedia data (YouTube)

Social media data



I Am Cait won't be renewed for...
Television • 2 hours ago

UPDATE: Simone Biles & Zac Efron...
Rio 2016 • 3 hours ago

Chinese swimmer Fu Yuanhui talks...
Rio 2016 • 1 hour ago

New to Twitter?

Sign up now to get your own personalized timeline!

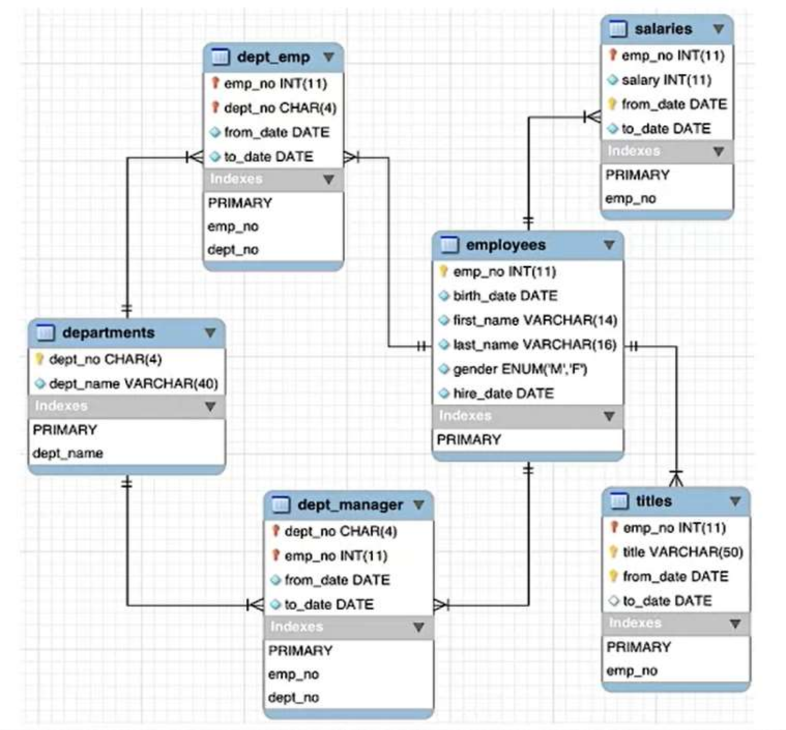
Seth Meyers @sethmeyers 6h

RIP John McLaughlin. My parents made us watch him every week which made the SNL sketches all the sweeter.

293 1.2K

Structured data

- Structure data refers to data with a high level of
- Organization, such as in relational databases and spreadsheets
- Depends on data model- a model of the data types and how they will be stored, processed and accessed
- Easily entered, stored, queried and analyzed
- Structured query language (SQL) is used for management of structured data (e.g. MySQL)



dept_emp
emp_no INT(11)
dept_no CHAR(4)
from_date DATE
to_date DATE
Indexes
PRIMARY
emp_no
dept_no

salaries
emp_no INT(11)
salary INT(11)
from_date DATE
to_date DATE
Indexes
PRIMARY
emp_no

employees
emp_no INT(11)
birth_date DATE
first_name VARCHAR(14)
last_name VARCHAR(16)
gender ENUM('M','F')
hire_date DATE
Indexes
PRIMARY

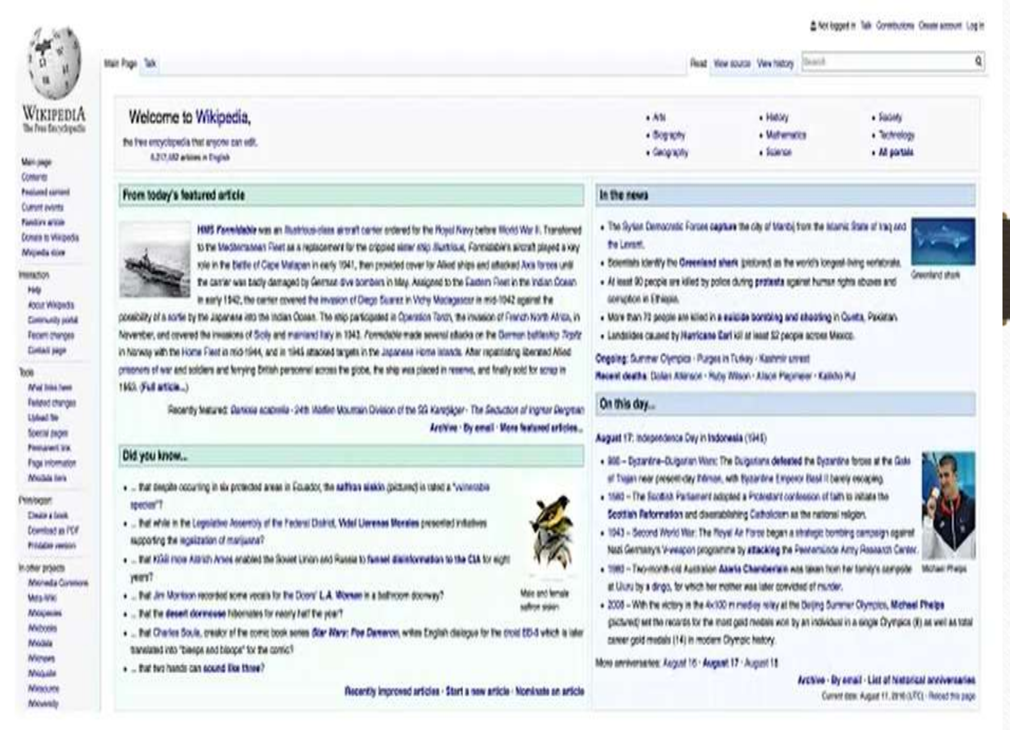
departments
dept_no CHAR(4)
dept_name VARCHAR(40)
Indexes
PRIMARY
dept_name

dept_manager
dept_no CHAR(4)
emp_no INT(11)
from_date DATE
to_date DATE
Indexes
PRIMARY
emp_no
dept_no

titles
emp_no INT(11)
title VARCHAR(50)
from_date DATE
to_date DATE
Indexes
PRIMARY
emp_no

Unstructured data

- Unstructured data means all things that cannot be classified and fit into one simple model
- Photos and graphics images
- Videos
- Streaming instrument data
- Webpages, emails, blog entries, wikis
- Pdf files, PowerPoint presentations, and word processing documents



Not logged in Talk Contributions Create account Log in

Main Page Talk

Read View source View history Search

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item

Print/export
Create a book
Download as PDF
Printable version

In other projects
Wikimedia Commons
Meta-Wiki
Wikispecies
Wikibooks
Wikidata
Wikinews
Wikiquote
Wikisource
Wikiversity

Welcome to Wikipedia,
the free encyclopedia that anyone can edit.
5,217,182 articles in English

- Arts
- Biography
- Geography
- History
- Mathematics
- Science
- Society
- Technology
- All portals

From today's featured article

HMS Formidable was an Illustrious-class aircraft carrier ordered for the Royal Navy before World War II. Transferred to the Mediterranean Fleet as a replacement for the crippled sister ship Illustrious, Formidable's aircraft played a key role in the Battle of Cape Matapan in early 1941, then provided cover for Allied ships and attacked Axis forces until the carrier was badly damaged by German dive bombers in May. Assigned to the Eastern Fleet in the Indian Ocean in early 1942, the carrier covered the invasion of Diego Suarez in Vichy Madagascar in mid-1942 against the possibility of a sortie by the Japanese into the Indian Ocean. The ship participated in Operation Torch, the invasion of French North Africa, in November, and covered the invasions of Sicily and mainland Italy in 1943. Formidable made several attacks on the German battleship Tirpitz in Norway with the Home Fleet in mid-1944, and in 1945 attacked targets in the Japanese Home Islands. After repatriating liberated Allied prisoners of war and soldiers and ferrying British personnel across the globe, the ship was placed in reserve, and finally sold for scrap in 1953. (Full article...)

Recently featured: Banksia acaulis · 24th Waffen Mountain Division of the SS Karstjäger · The Seduction of Ingmar Bergman

Archive · By email · More featured articles...

Did you know...

- ... that despite occurring in six protected areas in Ecuador, the saffron siskin (pictured) is rated a "vulnerable species"?
- ... that while in the Legislative Assembly of the Federal District, Vidal Llerenas Morales presented initiatives supporting the legalization of marijuana?
- ... that KGB mole Aldrich Ames enabled the Soviet Union and Russia to funnel disinformation to the CIA for eight years?
- ... that Jim Morrison recorded some vocals for the Doors' L.A. Woman in a bathroom doorway?
- ... that the desert dormouse hibernates for nearly half the year?
- ... that Charles Soule, creator of the comic book series Star Wars: Poe Dameron, writes English dialogue for the droid BB-8 which is later translated into "bleeps and bloops" for the comic?
- ... that two hands can sound like three?

Male and female saffron siskin

Recently improved articles · Start a new article · Nominate an article

In the news

- The Syrian Democratic Forces capture the city of Manbij from the Islamic State of Iraq and the Levant.
- Scientists identify the Greenland shark (pictured) as the world's longest-living vertebrate.
- At least 90 people are killed by police during protests against human rights abuses and corruption in Ethiopia.
- More than 70 people are killed in a suicide bombing and shooting in Quetta, Pakistan.
- Landslides caused by Hurricane Earl kill at least 52 people across Mexico.

Greenland shark

Ongoing: Summer Olympics · Purges in Turkey · Kashmir unrest
Recent deaths: Dolan Atkinson · Ruby Wilson · Alison Piepmeier · Kalikho Pul

On this day...

August 17: Independence Day in Indonesia (1945)

- 986 – Byzantine–Bulgarian Wars: The Bulgarians defeated the Byzantine forces at the Gate of Trajan near present-day Ihtiman, with Byzantine Emperor Basil II barely escaping.
- 1560 – The Scottish Parliament adopted a Protestant confession of faith to initiate the Scottish Reformation and disestablishing Catholicism as the national religion.
- 1943 – Second World War: The Royal Air Force began a strategic bombing campaign against Nazi Germany's V-weapon programme by attacking the Peenemünde Army Research Center.
- 1980 – Two-month-old Australian Azaria Chamberlain was taken from her family's campsite at Uluru by a dingo, for which her mother was later convicted of murder.
- 2008 – With the victory in the 4×100 m medley relay at the Beijing Summer Olympics, Michael Phelps (pictured) set the records for the most gold medals won by an individual in a single Olympics (8) as well as total career gold medals (14) in modern Olympic history.

Michael Phelps

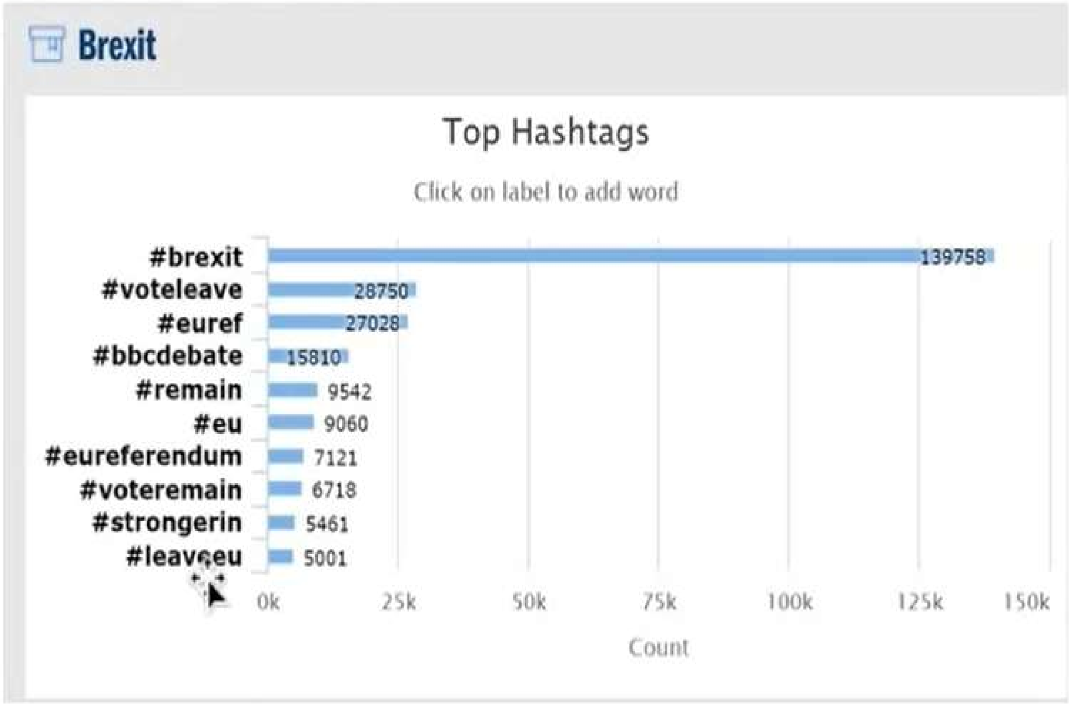
More anniversaries: August 16 · August 17 · August 18

Archive · By email · List of historical anniversaries

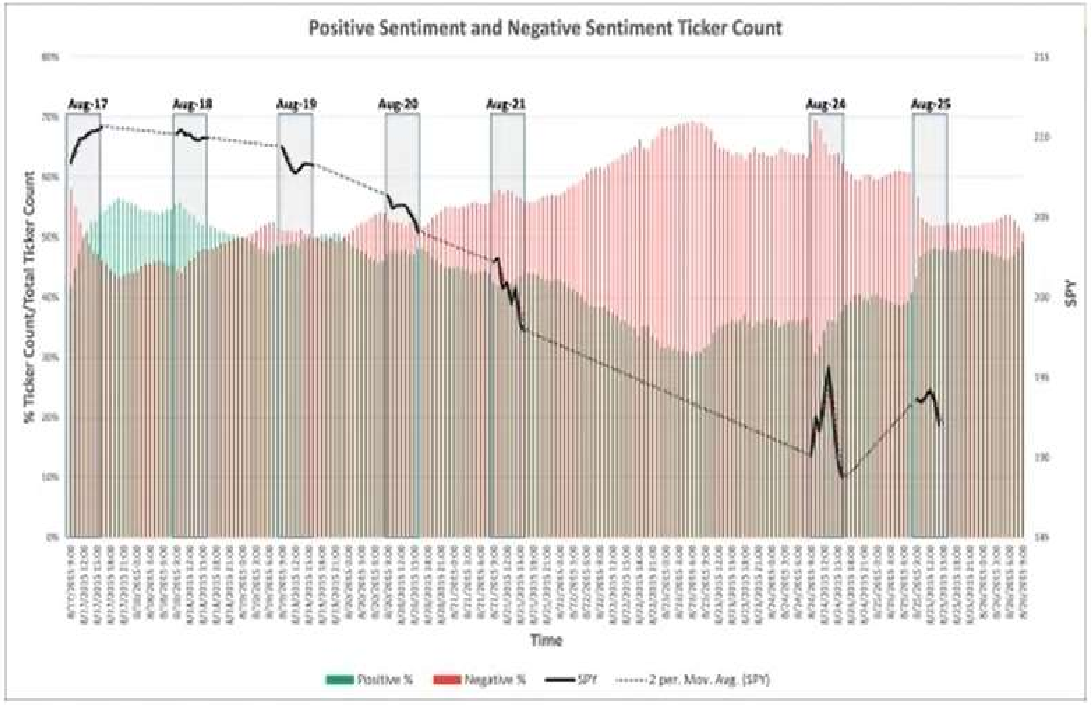
Current date: August 17, 2016 (UTC) · Reload this page

Structured and Unstructured features in social media data

- Structured data can be used for number-driven (quantitative) approaches
- Who, what, when, where, how and how many?



- Unstructured data for qualitative approaches:
- Why?
- Sentimental analysis



Class Exercise

1.1 Explain Various Kinds of Website/Social Media Data

Objective: Understand different types of social media data available on websites.

Activity 1: Brainstorming Session (20 mins)

- "What social media data can we extract from websites?"
- "How is this data useful for businesses?“

Discuss and categorize:

- **Engagement Metrics:** Likes, shares, retweets, comments
- **User Data:** Demographics, location, follower count
- **Content Data:** Post text, hashtags, media type
- **Temporal Data:** Posting times, engagement patterns
- **Sentiment Data:** Positive/negative/neutral tone in comments

Activity 2: Techniques to Analyze Social Media Data (10 mins)

Submit pdf in the Dropbox **‘Social Media Data Analysis’**

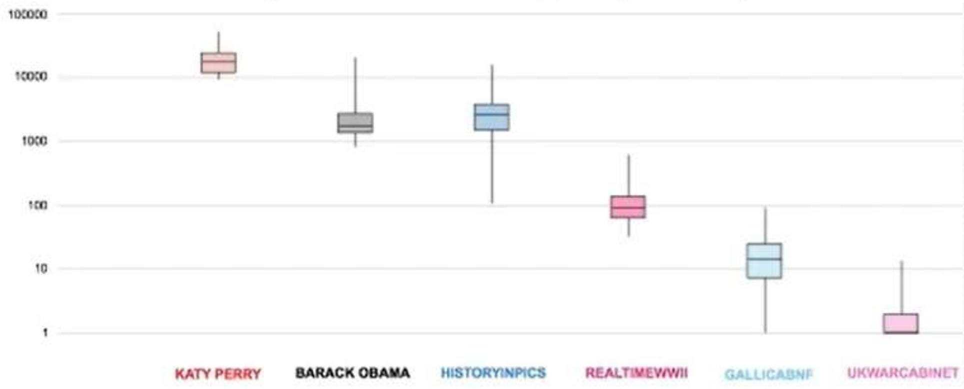
Structure – Real-World Social Media Data

- **X/Twitter post → structured**: username, follower count, retweet count, likes, timestamp, verified badge; **unstructured**: tweet text, images, emojis
- **Instagram profile → structured**: follower count, following, number of posts, location tag; **unstructured**: bio text, stories, reels video
- **YouTube video page → structured**: view count, likes/dislikes, subscriber count, upload date, comments count; **unstructured**: video itself, title, description, comments text
- **TikTok video → structured**: likes, comments, shares, play count, duration, music name; **unstructured**: video, caption, hashtags, stitched/duet content

Analyzing structured data: Descriptive statistics

- Descriptive statistics are used to describe the basic features of the data
- Summaries about the sample and the measures
  - Distribution (frequency table)
  - Central tendency (mean, median)
  - Dispersion (standard deviation)
- Simple graphics analysis
  - Bar chart, pie chart
  - Box plot

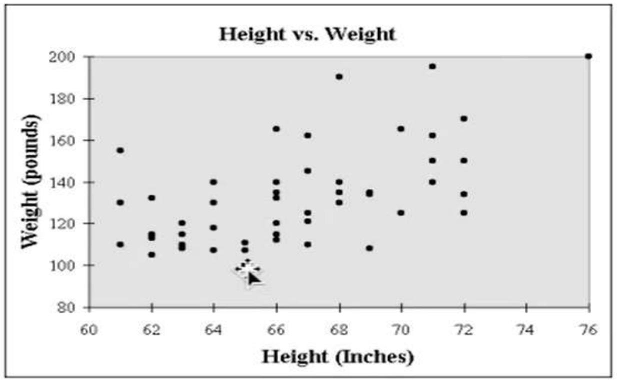
Number of retweets and favorites per tweet (log)
(http://www.martingrandjean.ch/)



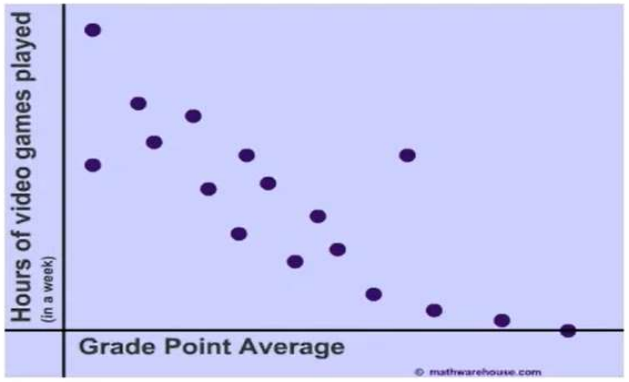
	KATY PERRY	BARACK OBAMA	HISTORYINPICS	REALTIMEWWII	GALLICABNF	UKWARCABINET
MAX	46 968	19 280	15 957	615	90	14
QUARTILE 3	23 493	2 647	3 652	136	25	2
MEDIAN	**16 755**	**1 676**	**2 567**	**90**	**14**	**1**
QUARTILE 1	11 416	1 296	1 516	63	7	0
MIN	9 424	832	108	32	0	0

Correlation

- In general, correlation refers to the extent to which two variables have a linear relationship with each other
- We will learn how to generate and test the correlation in python and R in upcoming session



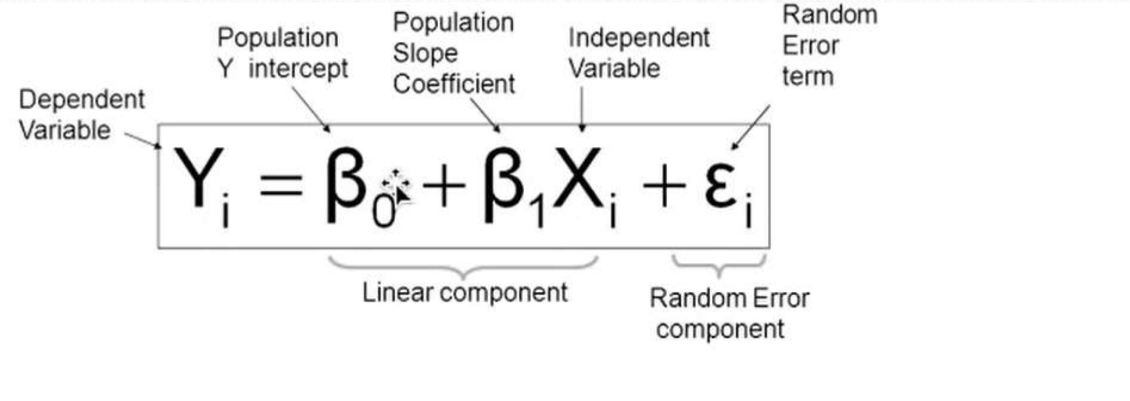
Positive correlation between height and weight



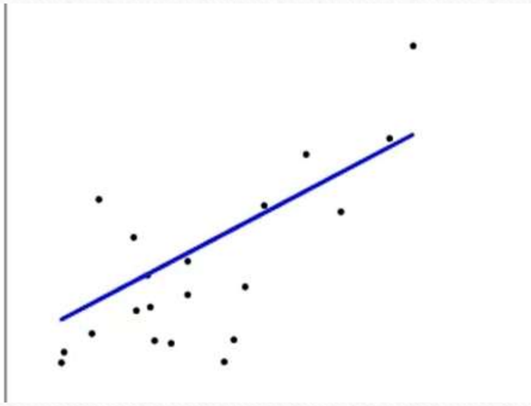
Negative correlation between GPA and video game

Regression

- Once a correlation is found, we do regression analysis to estimate the relationship among variables
- This relationship knowledge can then be used to predict one variable using others



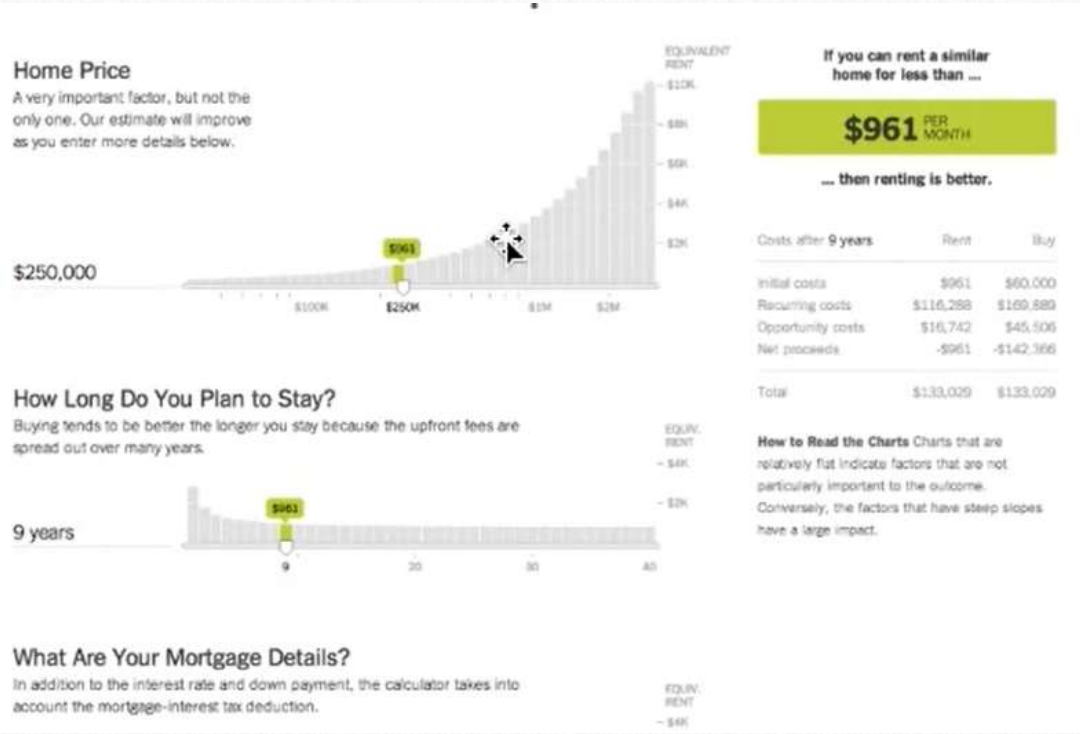
$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}} + \underbrace{\varepsilon_i}_{\text{Random Error component}}$$



Visualization of Data

- Data analytics techniques could give us insightful values, but it may take the right visualization to convey the meaning for decision-making.
- Visualization helps understand the characteristics of data and provides insights from it
- Discovery of new phenomena
- Sense-making of what data delivers to people
- Communication method between data analysts, decision-makers, service providers, etc

Renting vs Buying (from NYTimes)



Home Price

A very important factor, but not the only one. Our estimate will improve as you enter more details below.

$250,000

If you can rent a similar home for less than ...

$961 PER MONTH

... then renting is better.

Costs after **9 years**	Rent	Buy
Initial costs	$961	$60,000
Recurring costs	$116,288	$169,889
Opportunity costs	$16,742	$45,506
Net proceeds	-$961	-$142,366
Total	$133,029	$133,029

How to Read the Charts Charts that are relatively flat indicate factors that are not particularly important to the outcome. Conversely, the factors that have steep slopes have a large impact.

How Long Do You Plan to Stay?

Buying tends to be better the longer you stay because the upfront fees are spread out over many years.

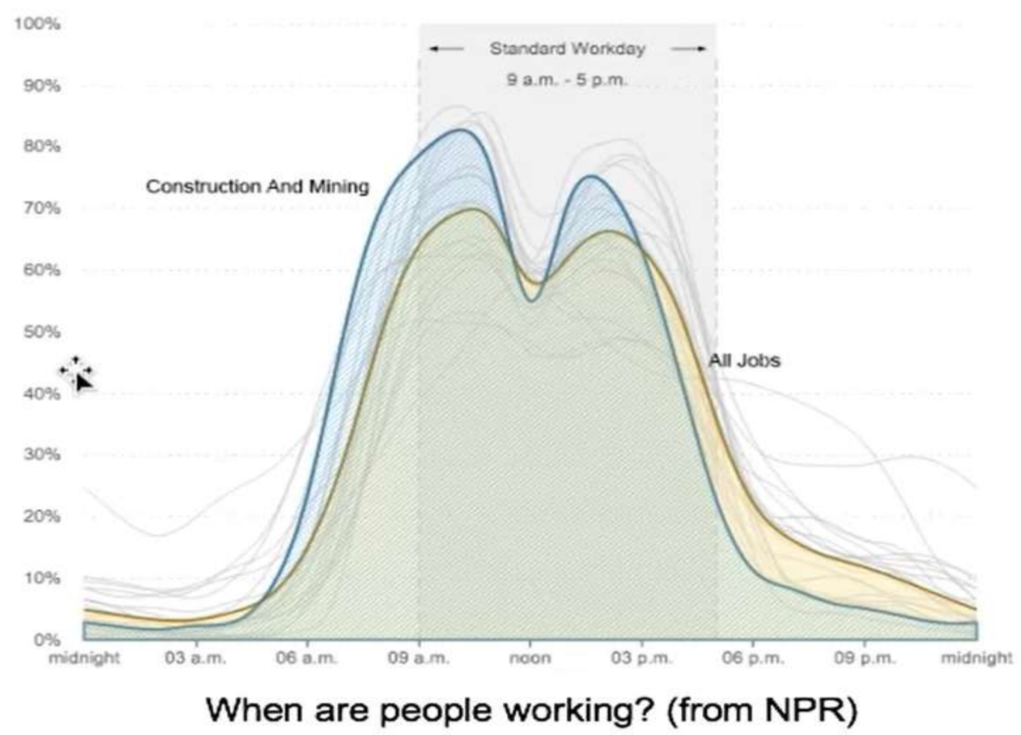
9 years

What Are Your Mortgage Details?

In addition to the interest rate and down payment, the calculator takes into account the mortgage-interest tax deduction.

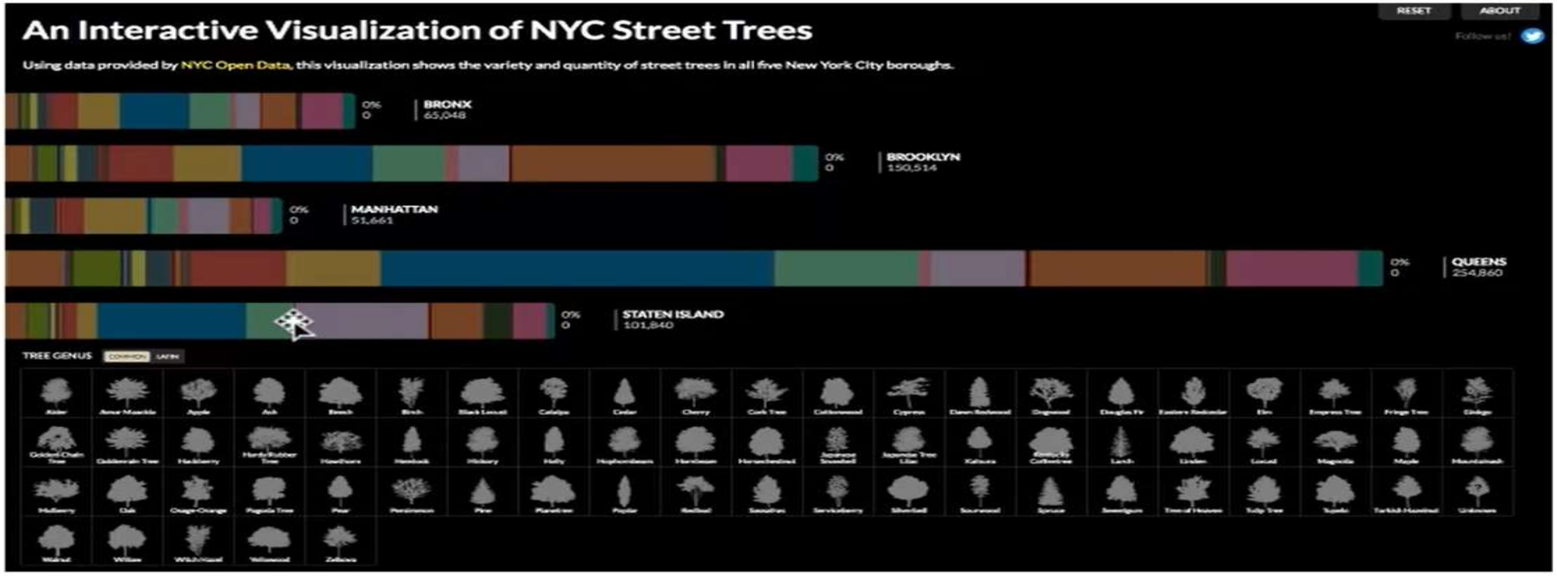
Histogram

- A graphical representation of the distribution of numerical data
- Give a rough sense of the density of the underlying distribution of the data



Bar chart

- Present grouped data with rectangular bard with lengths proportional to the values that they represent.



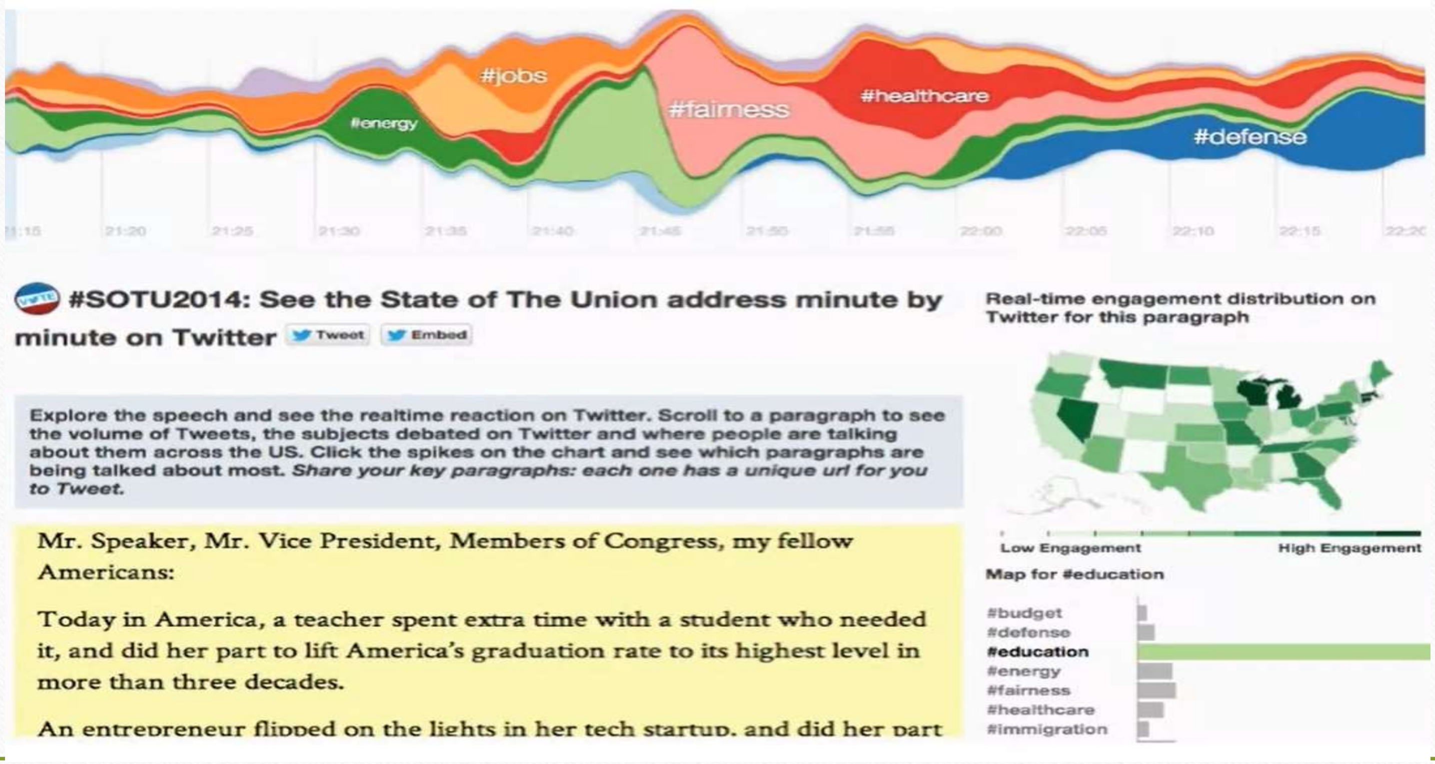
An Interactive Visualization of NYC Street Trees

Using data provided by NYC Open Data, this visualization shows the variety and quantity of street trees in all five New York City boroughs.

Interactive bar chart of NYC street trees

Time plot

- Displays values against time
- Helps understand trends



#SOTU2014: See the State of The Union address minute by minute on Twitter

Explore the speech and see the realtime reaction on Twitter. Scroll to a paragraph to see the volume of Tweets, the subjects debated on Twitter and where people are talking about them across the US. Click the spikes on the chart and see which paragraphs are being talked about most. *Share your key paragraphs: each one has a unique url for you to Tweet.*

Mr. Speaker, Mr. Vice President, Members of Congress, my fellow Americans:

Today in America, a teacher spent extra time with a student who needed it, and did her part to lift America's graduation rate to its highest level in more than three decades.

An entrepreneur flipped on the lights in her tech startup, and did her part

Real-time engagement distribution on Twitter for this paragraph

Map for #education

summary

- Social media data = mix of structured (numbers, counts, timestamps) and unstructured (text, images, video)
- Structured data → perfect for fast, quantitative analysis (“how many?”, “when?”, “who?”)
- Unstructured data → needs NLP, sentiment analysis, etc. (we’ll cover later)
- Focused only on the structured part → easier & immediately actionable

- **Descriptive Statistics**

  Central Tendency → Mean, Median, Mode

  Dispersion → Range, Standard Deviation, Variance

  Distribution → Frequency tables, Histograms

  Relationships → Correlation → Regression (prediction)

- **Visualization Toolbox**

  Histogram → see distribution shape

  Bar Chart → compare categories

  Time Plot / Line Chart → spot trends over time

  Pie Chart → show proportions

  Scatter plots, Box plots, Heatmaps

Himanshu Patel, Instructor
Saskatchewan Polytechnic
email: patelh@saskpolytech.ca
Mining building, Saskatoon