

How to Read & Represent Data

ISE:4172 Big Data Analytics
Stephen Baek

Vocabulary

- Data: noun, *plural (singular: datum)* (dā-tə; dā-)
 - Collection of entities and attributes

Vocabulary

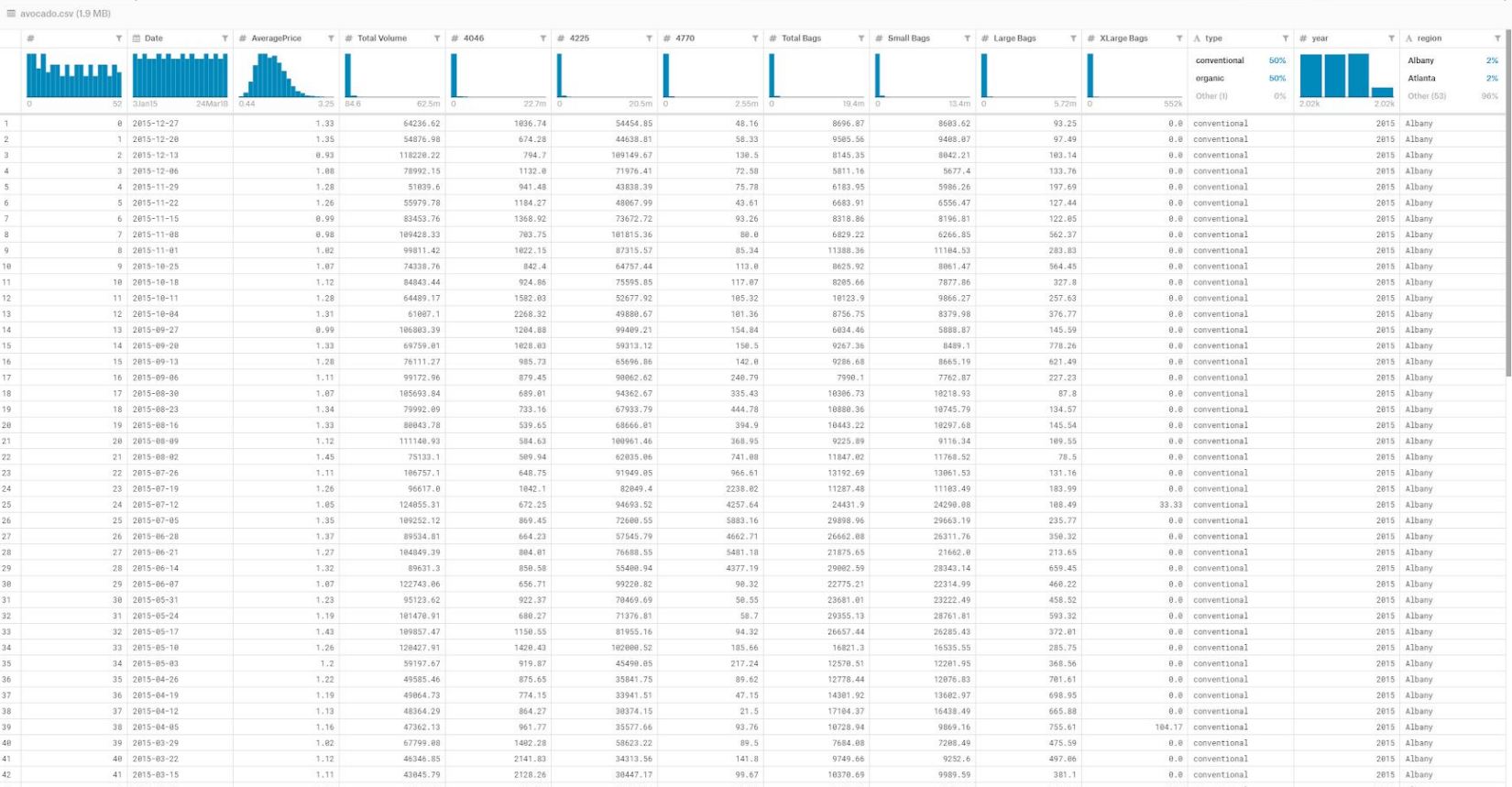
- Data: noun, *plural (singular: datum)* (dā-tə; dā-)
 - Collection of entities and attributes
- Object:
 - Also known as entity, sample, instance, data point, record, etc.
 - “Row” of the table
 - e.g. a person, a school, a tweet

Vocabulary

- Data: noun, *plural (singular: datum)* (dā-tə; dā-)
 - Collection of entities and attributes
- Object:
 - Also known as entity, sample, instance, data point, record, etc.
 - “Row” of the table
 - e.g. a person, a school, a tweet
- Attribute:
 - Also known as field, feature, parameter, variable, code, encoding, etc.
 - “Column” of the table
 - e.g. BMI of a person, student enrollment of a school, number of words in a tweet

avocado.csv (1.9 MB)																	
#		Date	AveragePrice	Total Volume	# 4046	# 4225	# 4770	# Total Bags	# Small Bags	# Large Bags	# XLarge Bags	A type	# year	A region			
												<div>conventional organic</div>	<div>50% 80%</div>		<div>Albany Atlanta</div>	<div>2% 2%</div>	
0	52	3Jan15	0.44	0.46	0	0	0	0	0	0	0	Other (1)	0%	2.60%	2.60%	Other (53)	96%
1	0	8 2015-12-27	1.33	64236.74	1836.74	54454.85	48.16	8696.87		8683.62	93.25	0.0	conventional	2015	Albany		
2	1	1 2015-12-20	1.35	54876.98	674.28	44638.81	58.33	5905.56		9488.87	97.49	0.0	conventional	2015	Albany		
3	2	2 2015-12-13	0.93	118228.22	794.7	189149.67	138.5	8145.35		8842.21	183.14	0.0	conventional	2015	Albany		
4	3	2015-12-06	1.08	78992.15	1132.8	71976.41	72.58	5811.16		5677.4	133.76	0.0	conventional	2015	Albany		
5	4	2015-11-29	1.28	51039.6	941.48	43838.39	75.78	6183.95		5986.26	197.69	0.0	conventional	2015	Albany		
6	5	2015-11-22	1.26	55979.78	1184.27	48067.99	43.61	6683.91		6556.47	127.44	0.0	conventional	2015	Albany		
7	6	2015-11-15	0.99	82453.76	1368.92	73672.72	92.26	8318.86		8196.81	122.65	0.0	conventional	2015	Albany		
8	7	2015-11-08	0.98	109428.33	703.75	101815.36	88.8	6829.22		6266.85	562.37	0.0	conventional	2015	Albany		
9	8	2015-11-01	1.02	99811.42	1022.15	87315.57	85.34	11388.36		11184.53	283.83	0.0	conventional	2015	Albany		
10	9	2015-10-25	1.07	74338.76	842.4	64757.44	113.0	8625.92		8861.47	564.45	0.0	conventional	2015	Albany		
11	10	2015-10-18	1.12	84843.44	924.86	75595.85	117.07	8205.66		7877.86	327.8	0.0	conventional	2015	Albany		
12	11	2015-10-11	1.28	64489.17	1582.03	52677.92	105.32	10123.9		9866.27	257.63	0.0	conventional	2015	Albany		
13	12	2015-10-04	1.31	61087.1	2266.32	49880.67	101.36	8756.75		8379.98	376.77	0.0	conventional	2015	Albany		
14	13	2015-09-27	0.99	106883.39	1204.88	99489.21	154.84	6834.46		5888.87	145.59	0.0	conventional	2015	Albany		
15	14	2015-09-20	1.33	69759.01	1028.83	59313.12	158.5	9267.36		8489.1	778.26	0.0	conventional	2015	Albany		
16	15	2015-09-13	1.28	76111.27	985.73	65696.86	142.0	9286.68		8665.19	621.49	0.0	conventional	2015	Albany		
17	16	2015-09-06	1.11	99172.96	879.45	90862.62	240.79	7990.1		7762.87	227.23	0.0	conventional	2015	Albany		
18	17	2015-08-30	1.07	105693.84	689.01	94362.67	335.43	10386.73		10218.93	87.8	0.0	conventional	2015	Albany		
19	18	2015-08-23	1.34	79992.09	733.16	67933.79	444.78	10880.36		10745.79	134.57	0.0	conventional	2015	Albany		
20	19	2015-08-16	1.33	88043.78	539.65	68666.81	394.9	10443.22		10297.68	145.54	0.0	conventional	2015	Albany		
21	20	2015-08-09	1.12	111148.93	584.63	109891.46	368.95	9225.89		9116.34	169.55	0.0	conventional	2015	Albany		
22	21	2015-08-02	1.45	75133.1	589.94	62035.06	741.08	11847.02		11768.52	78.5	0.0	conventional	2015	Albany		
23	22	2015-07-26	1.11	186757.1	648.75	91949.05	966.61	13192.69		13061.53	131.16	0.0	conventional	2015	Albany		
24	23	2015-07-19	1.26	96617.0	1042.1	82849.4	2238.02	11287.48		11183.49	183.99	0.0	conventional	2015	Albany		
25	24	2015-07-12	1.05	124055.31	672.25	94693.52	4257.64	24431.9		24290.08	108.49	33.33	conventional	2015	Albany		
26	25	2015-07-05	1.35	109252.12	869.45	72680.55	5883.16	29898.96		29663.19	235.77	0.0	conventional	2015	Albany		
27	26	2015-06-28	1.37	89534.81	664.23	57545.79	4662.71	26662.08		26311.76	350.32	0.0	conventional	2015	Albany		
28	27	2015-06-21	1.27	104849.39	884.01	76688.55	5481.18	21875.65		21662.0	213.65	0.0	conventional	2015	Albany		
29	28	2015-06-14	1.32	89631.3	858.58	55400.94	4377.19	29082.59		28343.14	659.45	0.0	conventional	2015	Albany		
30	29	2015-06-07	1.07	122743.06	656.71	99228.82	90.32	22775.21		22314.99	460.22	0.0	conventional	2015	Albany		
31	30	2015-05-31	1.23	95123.62	922.37	70469.69	50.55	23681.81		23222.49	458.52	0.0	conventional	2015	Albany		
32	31	2015-05-24	1.19	101470.91	680.27	71376.81	58.7	29355.13		28761.81	593.32	0.0	conventional	2015	Albany		
33	32	2015-05-17	1.43	115985.47	1150.55	81955.16	94.32	26657.44		26285.43	372.81	0.0	conventional	2015	Albany		
34	33	2015-05-10	1.26	128427.91	1420.43	102800.82	105.66	16821.3		16535.55	285.75	0.0	conventional	2015	Albany		
35	34	2015-05-03	1.2	59197.67	43490.85	59197.67	217.24	12570.51		12281.95	368.56	0.0	conventional	2015	Albany		
36	35	2015-04-26	1.22	45985.46	875.65	35841.75	89.62	12778.44		12076.83	781.61	0.0	conventional	2015	Albany		
37	36	2015-04-19	1.19	49864.73	774.15	33941.51	47.15	14301.92		13682.97	698.95	0.0	conventional	2015	Albany		
38	37	2015-04-12	1.13	48364.29	864.27	30374.15	21.5	17184.37		16438.49	665.88	0.0	conventional	2015	Albany		
39	38	2015-04-05	1.16	47362.13	961.77	35577.66	93.76	10728.94		9869.16	755.61	104.17	conventional	2015	Albany		
40	39	2015-03-29	1.02	67799.88	1402.28	58623.22	89.5	7684.08		7208.49	475.59	0.0	conventional	2015	Albany		
41	40	2015-03-22	1.12	46344.85	2141.83	34313.56	141.8	9749.66		9252.6	497.86	0.0	conventional	2015	Albany		
42	41	2015-03-15	1.11	43945.79	2128.26	30447.17	99.67	10370.69		9989.59	381.1	0.0	conventional	2015	Albany		

Attributes



Avocado prices dataset (<https://www.kaggle.com/neuromusic/avocado-prices>)

Objects

Set, Sequence, & Space

Ordered

⚡

Set, Sequence, & Space

⚡

Unordered

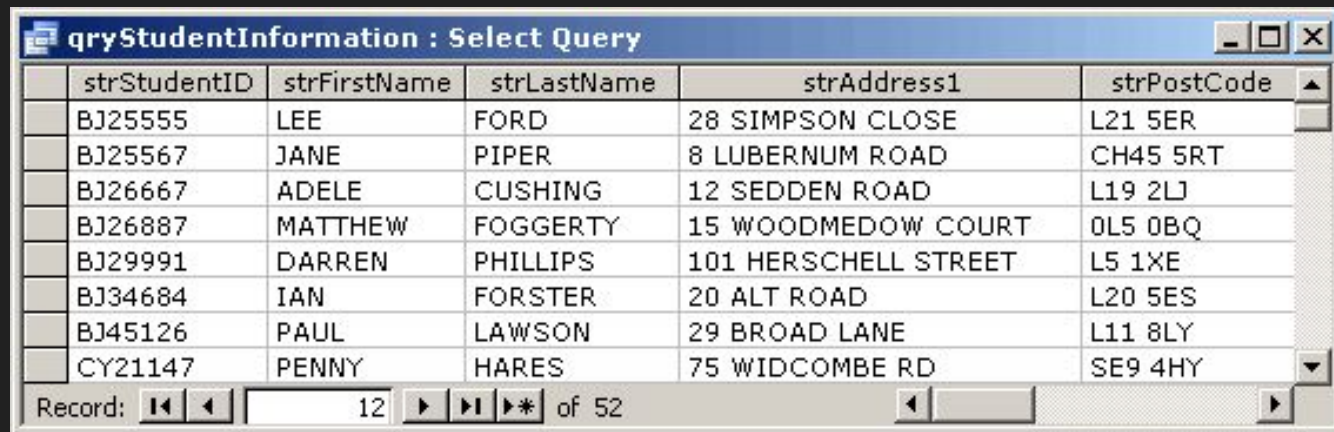
*Geometric-
Structured*

Unordered Data

- A set of attributes
- Ordering of attributes doesn't really matter $\{a, b\} = \{b, a\}$
- Examples
 - Documents, tweets, web contents
 - Demographic data
 - Employee records
 - Student records
 - Bank transaction records
 - Product inventory
 - ...

Unordered Data

- e.g. Student record table



strStudentID	strFirstName	strLastName	strAddress1	strPostCode
BJ25555	LEE	FORD	28 SIMPSON CLOSE	L21 5ER
BJ25567	JANE	PIPER	8 LUBERNUM ROAD	CH45 5RT
BJ26667	ADELE	CUSHING	12 SEDDEN ROAD	L19 2LJ
BJ26887	MATTHEW	FOGGERTY	15 WOODMEDOW COURT	OL5 0BQ
BJ29991	DARREN	PHILLIPS	101 HERSCHELL STREET	L5 1XE
BJ34684	IAN	FORSTER	20 ALT ROAD	L20 5ES
BJ45126	PAUL	LAWSON	29 BROAD LANE	L11 8LY
CY21147	PENNY	HARES	75 WIDCOMBE RD	SE9 4HY

Record: 12 of 52

Unordered Data

- e.g. Documents (bag of words)

Article ID	biolog	biopsi	biolab	biotin	almost	cancer-surviv	cancer-stage	Article Class
00001	12	1	2	10	0	1	4	breast-cancer
00002	10	1	0	3	0	6	1	breast-cancer
00014	4	1	1	1	0	28	0	breast-cancer
00063	4	0	0	0	0	18	7	breast-cancer
00319	0	1	0	9	0	20	1	breast-cancer
00847	7	2	0	14	0	11	5	breast-cancer
03042	3	1	3	1	0	19	8	lung-cancer
05267	4	4	2	6	0	14	11	lung-cancer
05970	8	0	4	9	0	9	17	lung-cancer
30261	1	0	0	11	0	21	1	prostate-cancer
41191	9	0	5	14	0	11	1	prostate-cancer
52038	6	1	1	17	0	19	0	prostate-cancer
73851	1	1	8	17	0	17	3	prostate-cancer

Ordered Data

- Ordered set of attributes
- Ordering matters! $(a, b) \neq (b, a)$
- Examples
 - Time series
 - Sequence
 - ...

Ordered Data

- e.g. Genetic Sequence

```
12854400 tcaaagtaagttagataaacaatgatcattcacaggtcagatgttttaaaaaaaatcattatgggtgtacatcacatgtagacaataacttcagaattcatc
12854200 tggactaccagaattgagttacgtactttctcaattctatttttaccctaacgtctaataaataacaagtaactctagcctctctcgttttatgattcctc
12854000 taggaaaagttaatgttacggcccaatcacttttttaacagcccaacaacatatattagctccaaatatcattttttcccctagaatatctcacaacct
12853800 attgtccactcaaaacgtgacaaatggaggtctaaagggagaccatacttgactcatttttagagctaggatcagacagagtagatTTTTTgccataaactc
12853600 cttgtaaatgtattcacatttcattcccaagaaaaatagactgatgaagaaatatatcagatatgacaaggccgtgtcgttttaggttacgtaactctaca
12853400 aggttttagggttctcaatataaacacacaaaagcagatagaagaagcaaaccattcacaaatcagacaATGACATCTCTCCATACGTTACTCTTCTCTCTCT
12853200 TCTTTTCTTCATCGTCTTTCCAACTTCACGTTTCCCTCCACCTTATTGTTTCAGgttcgtcttttagttttgcttcttttacatacacagactctacacac
12853000 tcacttattgggttttctttcaattgtgaaacagAGTTTCAATTGGGAGTCATGGAAGAAAGAAGGAGGATTCTACAATTCTCTCCACAACCTCCATTGACG
12852800 taccaatcttgggttactcacgcaatcttcattctcagGTTACTTACCGGGAAGCTATACGATCTAAACAGCTCCAATACGGTTCAGAGGCGGAACCTGA
12852600 AATCGTTAATCAAAGCGTTGAATCAAAAAGGAATAAAAGCTTTGGCTGATATAGTGATTAACCACAGAACAGCTGAGAGGAAAGACGATAAATGTGGATA
12852400 CTGTTATTTTCAAGGTGGGACTTCCGATGATCGTCTTGATTGGGATCCTTCTTTGCTGCGCAATGACCCTAAATTTCCTGGTACCGGAAACCTCGAC
12852200 ACCGGAGGAGATTTTGATGGAGCGCCGACATCGACCACCTTAACCCTAGAGTTTCAGAAAGAGTTGTCCGAATGGATGAATTGGCTTAAACTGAAATCG
12852000 GATTCATGGTTGGAGATTTGATTATGTTTCGAGTTATGCATCTCCATCACCAATTAACGTTACAGTtaaatcacatatgaattctcaaatatcagac
12851800 aacagtattagatatataaagaacataggttgagataatttactatttagtatataagtatcataggttgataggttttagatttagtat
ataaagaacataagtcgaatcaataaagaatatataaagaagttcactactgattatgtgataaattcctctgtttttggatacacagAATACATC
ACCGGATTTTGGGTGGGTGAGAAATGGGACGATATGAAGTACGGAGGAGACGGGAAACTAGACTATGATCAGAACGAGCATCGGTTCGGGTCTCAAACAG
TGGATCGAGGAAGCGGTGGTGGTGTGTTGACAGCTTTGATTTACCACCAAAGGGATCTTACAGTCTGCTGTCAAAGGTGAGCTTTGGAGACTAAAGG
ACTCGCAGGAAACCGCCTGGTATGATGAAGATCATGCCCGGAAACGCTGTACATTCATAGATAACCATGATACATTCAGAACGTGGGTTTTCCCTTC
TGATAAAGTCTTGCTTGGATACGTTTATATACTTACTCATCCAGGAACCTCCTTGCAATgtaagtatcatttttagtatgtagctatactatttacaactac
aatcttgttgatatgttatttttgttgagTTTTATAATCATTACATAGAATGGGGACTAAAGAGAGCATCTCAAAGCTGGTGGCTATCAGGAACAAAA
ATGGGATTTGGTAGCACAGCTCTGTAACGATAAAAGCGGACGAGGCGGATCTCTACTTGGCTATGATTGATGATAAAGTTATCATGAAGATTGGACCAA
GCAAGATGTGGGACACTTGTTCCTTCTAATTTTGCTTTAGCTTTTACGCGCTTGACTTTGCTGTCTGGGAGAAGAGTAAcgcataactcgaatcata
agaaaagtaatcgaatgtatcttcttcttttaataaaaacattttggcagtatctaaagatatgtataatgaaatataaaatgataaagaatacctaaa
taaaaagagcactagtggtgttaaaggatacaactccagtgaaagaaaagagttcaagtgaagaagtgtaactttagaataaagatttggaagtttc
catcgTTTTgttttgttgcatacaactaatatattatatttgccgactcgtataagatttggagccctactaaaatcagaattatgatgtcttaacca
cacaactactgccaaaatcagaacgaattatattttagtaagaagaaaaaaagtatggtgggaagtggaacagttagacaggtaaatcgaataaa
```

A
T
4
G
2
5
0
0
0
1

v

Ordered Data

- e.g. Stock price (candlestick chart)



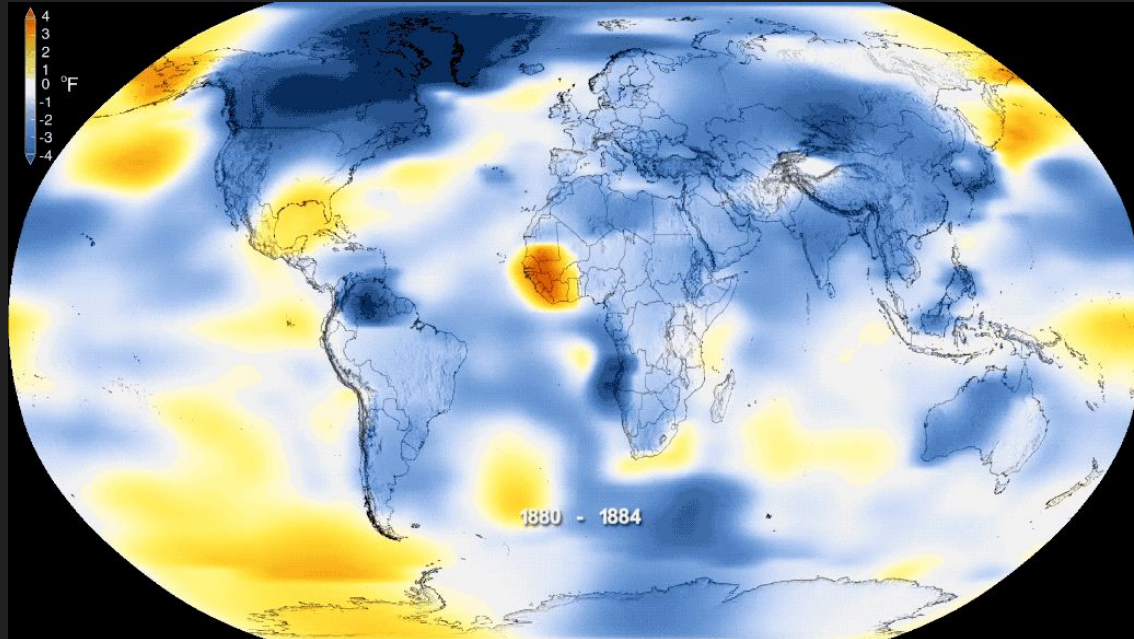
<https://www.investopedia.com/trading/candlestick-charting-what-is-it/>

Geometric/Structured Data

- Data sets that have geometric/topological/geographical structures.
- Spatial location of an object comes into play.
- Example
 - Spatio-temporal data
 - Image pixels, points in LiDAR, computer graphics models
 - Graph data

Geometric/Structured Data

- e.g. Spatio-temporal data



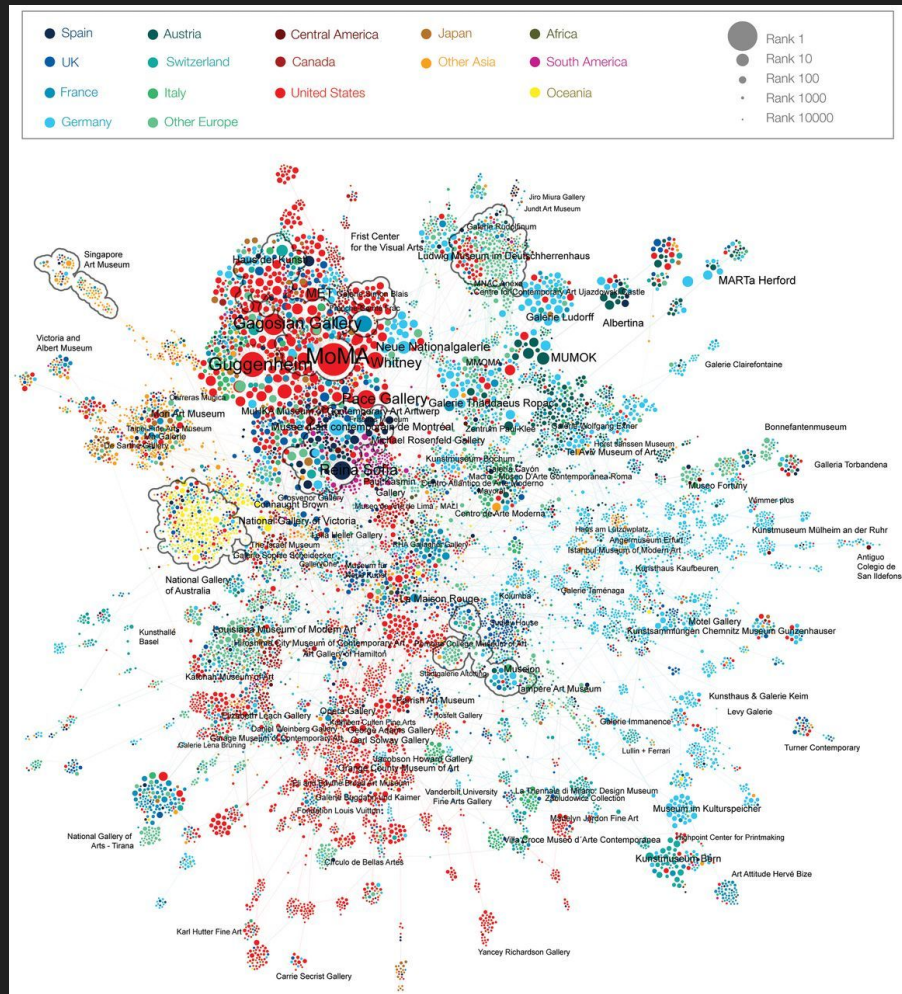
<https://climate.nasa.gov/news/2876/new-studies-increase-confidence-in-nasas-measure-of-earths-temperature/>

Geometric/Structured Data

- e.g. Graph data (social network)

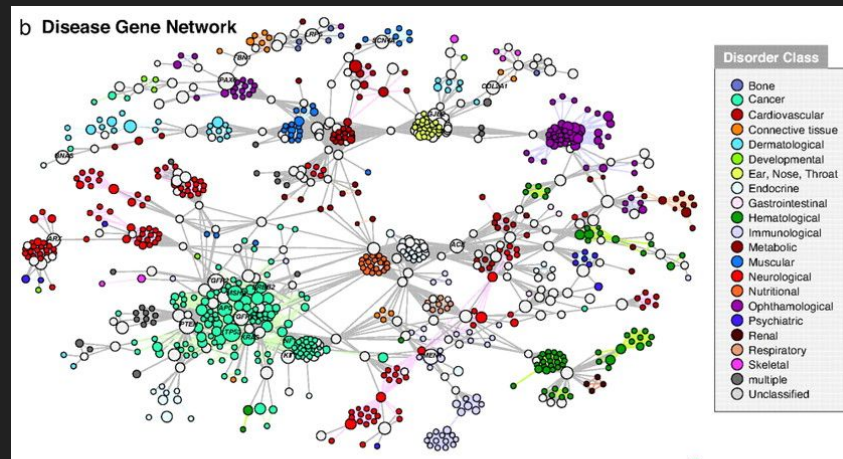
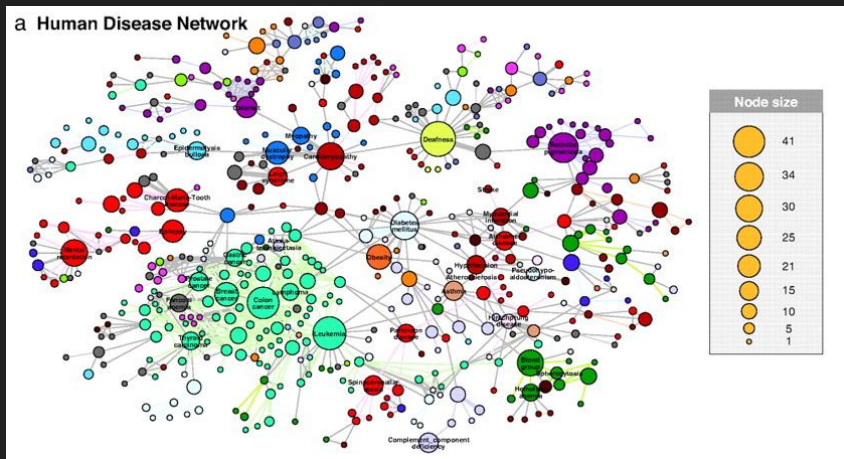
“Quantifying reputation and success in art”
Fraiberger et al. (Science, 2018)

Force-directed layout of the order $\tau = \infty$ coexhibition network, whose nodes are institutions (galleries, museums). Node size is proportional to each institution's eigenvector centrality. Nodes are connected if they both exhibited the same artist, with link weights being equal to the number of artists' coexhibitions. Node colors encode the region in which institutions are located. Links are of the same colors as their end nodes, or gray when end nodes have different colors. For visualization purposes, we only show the 12,238 nodes corresponding to institutions with more than 10 exhibits; we pruned the links by keeping the most statistically significant links (20) (supplementary text S2.2). We implemented community detection on the pruned network (21), identifying 122 communities (supplementary text S2.3). We highlighted five of them, the full community breakdown being shown in fig. S3. We also show the names of the most prestigious institution for each community.



Geometric/Structured Data

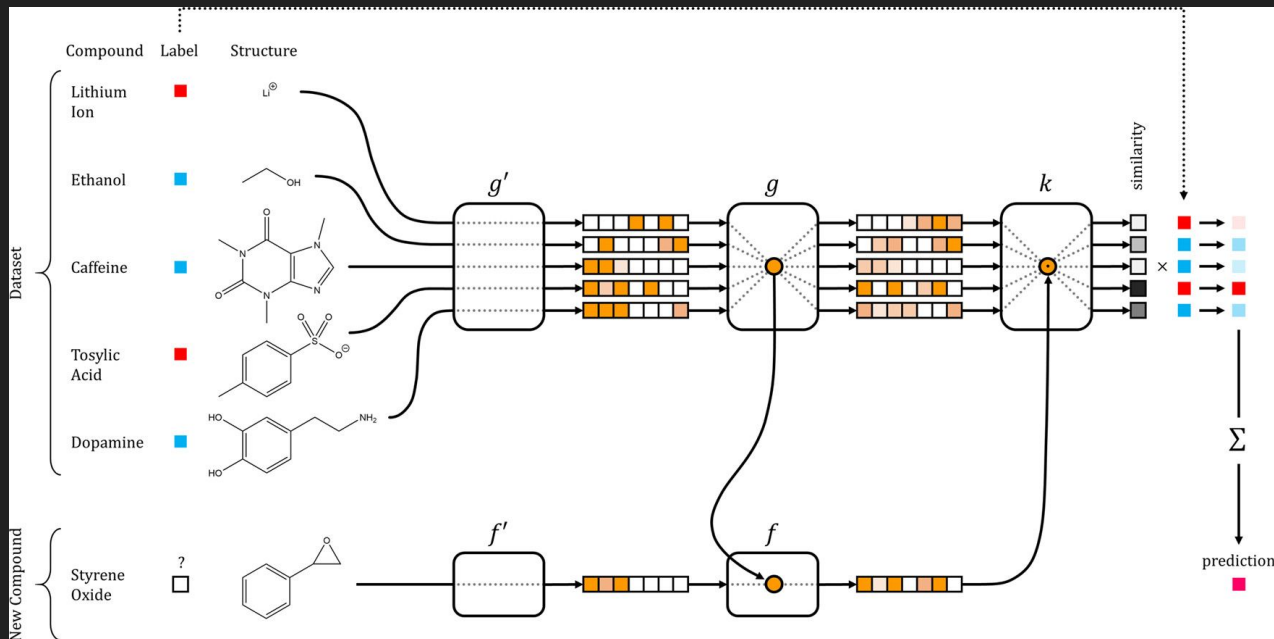
- e.g. Graph data (disease network)



“The Human Disease Network”
Goh et al. (PNAS, 2007)

Geometric/Structured Data

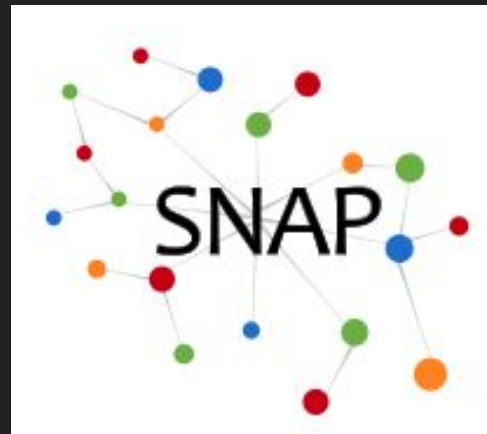
- e.g. Graph data (chemical compounds)



“Low Data Drug Discovery with One-Shot Learning”
Altae-Tran et al. (ACS Central Science, 2017)

Geometric/Structured Data

- e.g. Graph data (Stanford Large Network Dataset Collection)
 - <http://snap.stanford.edu/>
 - Social Networks
 - Communication Networks
 - Citation Networks
 - Collaboration Networks
 - Road Networks
 - Temporal Networks
 - ...



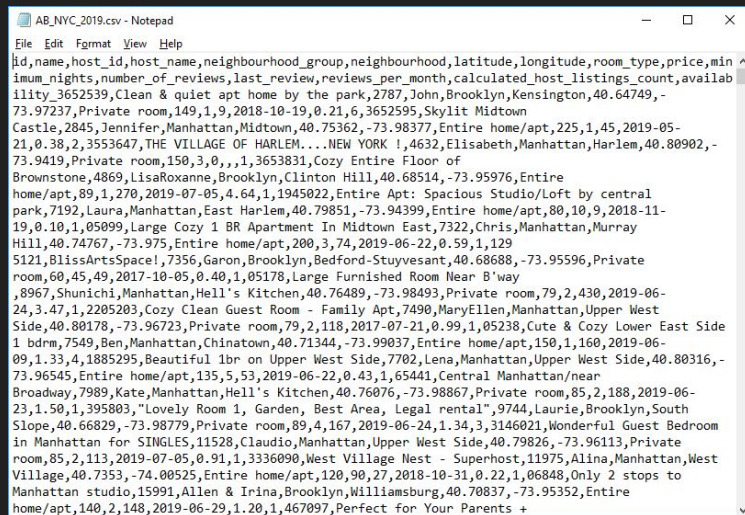
Data Formats

- There are MANY data representations in computers!
 - Comma Separated Values (CSV)
 - JavaScript Object Notation (JSON)
 - eXtensible Markup Language (XML)
 - ...

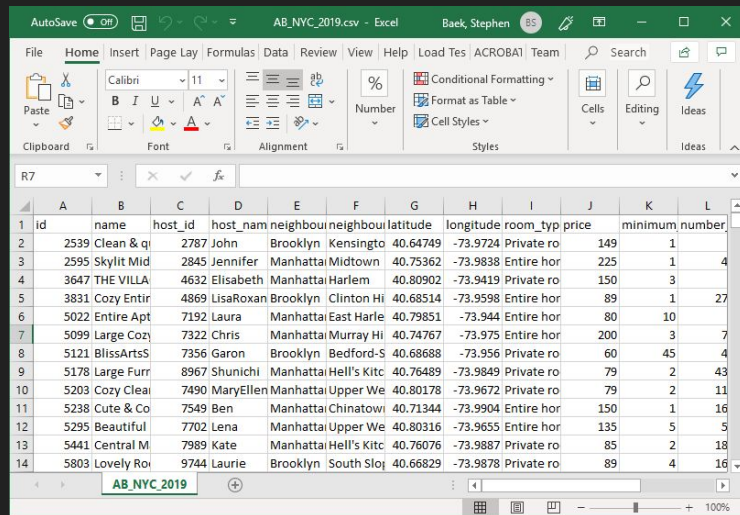
Data Formats

- Comma Separated Values (CSV)

- Delimited text file that uses comma (,) to separate values
- Some ~~weird~~ people use something else, like semicolon (;), instead
- Tabular data is stored in plain text → large file size



```
AB_NYC_2019.csv - Notepad
File Edit Format View Help
id,name,host_id,host_name,neighbourhood_group,neighbourhood,latitude,longitude,room_type,price,minimum_nights,number_of_reviews,last_review,reviews_per_month,calculated_host_listings_count,availability_365,2539,Clean & quiet apt home by the park,2787,John,Brooklyn,Kensington,40.64749,-73.97237,Private room,149,1,9,2018-10-19,0.21,6,3652595,Skyli Midtown Castle,2845,Jennifer,Manhattan,Midtown,40.75362,-73.98377,Entire home/apt,225,1,45,2019-05-21,0.38,2,3553647,THE VILLAGE OF HARLEM...NEW YORK,1,4632,Elisabeth,Manhattan,Harlem,40.80902,-73.9419,Private room,150,3,0,,1,3653831,Cozy Entire Floor of Brownstone,4869,LisaRoxanne,Brooklyn,Clinton Hill,40.68514,-73.95976,Entire home/apt,89,1,270,2019-07-05,4.64,1,1945022,Entire Apt: Spacious Studio/Loft by central park,7192,Laura,Manhattan,East Harlem,40.79851,-73.94399,Entire home/apt,80,10,9,2018-11-19,0.10,1,05099,Large Cozy 1 BR Apartment In Midtown East,7322,Chris,Manhattan,Murray Hill,40.74767,-73.975,Entire home/apt,200,3,74,2019-06-22,0.59,1,129,5121,BlissArtsSpace,7356,Garon,Brooklyn,Bedford-Stuyvesant,40.68688,-73.95596,Private room,60,45,49,2017-10-05,0.40,1,05178,Large Furnished Room Near B'way,8967,Shunichi,Manhattan,Hell's Kitchen,40.76489,-73.98493,Private room,79,2,430,2019-06-23,4.47,1,2205203,Cozy Clean Guest Room - Family Apt,7490,MaryEllen,Manhattan,Upper West Side,40.80178,-73.96723,Private room,79,2,118,2017-07-21,0.99,1,05238,Cute & Cozy Lower East Side 1 bdrm,7549,Ben,Manhattan,Chinatown,40.71344,-73.99037,Entire home/apt,150,1,160,2019-06-09,1.33,4,1885295,Beautiful 1br on Upper West Side,7702,Lena,Manhattan,Upper West Side,40.80316,-73.96545,Entire home/apt,135,5,53,2019-06-22,0.43,1,65441,Central Manhattan/near Broadway,7989,Kate,Manhattan,Hell's Kitchen,40.76076,-73.98867,Private room,85,2,188,2019-06-23,1.50,1,395803,"Lovely Room 1, Garden, Best Area, Legal rental",9744,Laurie,Brooklyn,South Slope,40.66829,-73.98779,Private room,89,4,167,2019-06-24,1.34,3,3146021,Wonderful Guest Bedroom In Manhattan for SINGLES,11528,Claudio,Manhattan,Upper West Side,40.79826,-73.96113,Private room,85,2,113,2019-07-05,0.91,1,3336090,West Village Nest - Superhost,11975,Alina,Manhattan,West Village,40.7353,-74.00525,Entire home/apt,120,90,27,2018-10-31,0.22,1,06848,Only 2 stops to Manhattan studio,15991,Allen & Irina,Brooklyn,Williamsburg,40.70837,-73.95352,Entire home/apt,140,2,148,2019-06-29,1.20,1,467097,Perfect for Your Parents +
```



	A	B	C	D	E	F	G	H	I	J	K	L
1	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews
2	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.9724	Private room	149	1	9
3	2595	Skyli Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.9838	Entire home/apt	225	1	45
4	3647	THE VILLAGE OF HARLEM...NEW YORK	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.9419	Private room	150	3	0
5	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.9598	Entire home/apt	89	1	270
6	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.944	Entire home/apt	80	10	9
7	5099	Large Cozy 1 BR Apartment In Midtown East	7322	Chris	Manhattan	Murray Hill	40.74767	-73.975	Entire home/apt	200	3	74
8	5121	BlissArtsSpace	7356	Garon	Brooklyn	Bedford-Stuyvesant	40.68688	-73.956	Private room	60	45	49
9	5178	Large Furnished Room Near B'way	8967	Shunichi	Manhattan	Hell's Kitchen	40.76489	-73.9849	Private room	79	2	430
10	5203	Cozy Clean Guest Room - Family Apt	7490	MaryEllen	Manhattan	Upper West Side	40.80178	-73.9672	Private room	79	2	118
11	5238	Cute & Cozy Lower East Side 1 bdrm	7549	Ben	Manhattan	Chinatown	40.71344	-73.9904	Entire home/apt	150	1	160
12	5295	Beautiful 1br on Upper West Side	7702	Lena	Manhattan	Upper West Side	40.80316	-73.9655	Entire home/apt	135	5	53
13	5441	Central Manhattan/near Broadway	7989	Kate	Manhattan	Hell's Kitchen	40.76076	-73.9887	Private room	85	2	188
14	5803	Lovely Room 1, Garden, Best Area, Legal rental	9744	Laurie	Brooklyn	South Slope	40.66829	-73.9878	Private room	89	4	167

<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

Data Formats

- JavaScript Object Notation (JSON)
 - Easy for humans to read and write, easy for machines to parse and generate
 - Attribute-value pairs + arrays
 - Good for structured data

```
{
  "firstName": "John",
  "lastName": "Smith",
  "isAlive": true,
  "age": 27,
  "address": {
    "streetAddress": "21 2nd Street",
    "city": "New York",
    "state": "NY",
    "postalCode": "10021-3100"
  },
  "phoneNumbers": [
    {
      "type": "home",
      "number": "212 555-1234"
    },
    {
      "type": "office",
      "number": "646 555-4567"
    },
    {
      "type": "mobile",
      "number": "123 456-7890"
    }
  ],
  "children": [],
  "spouse": null
}
```

Data Formats

- eXtensible Markup Language (XML)
 - Easy for humans, easy for machines
 - Made of tags
 - Start-tag: <tagname>
 - End-tag: </tagname>
 - Empty-element-tag: <tagname />
 - Attributes
 - Name-value pair that exists in a tag.
 - For example,

 - Tag: img (an empty tag)
 - Attributes: src, alt

```
<?xml version="1.0" encoding="UTF-8" ?>
<root>
  <listing>
    <seller_info>
      <seller_name> cubsfantony</seller_name>
      <seller_rating> 848</seller_rating>
    </seller_info>
    <payment_types>
      Visa/MasterCard, Money Order/Cashiers Checks, Personal Checks, See item description for details
    </payment_types>
    <shipping_info>
      Buyer pays fixed shipping charges, Will ship to United States only
    </shipping_info>
    <buyer_protection_info> </buyer_protection_info>
    <auction_info>
      <current_bid>$620.00 </current_bid>
      <time_left> 4 days, 14 hours + </time_left>
      <high_bidder>
        <bidder_name> gosha555@excite.com </bidder_name>
        <bidder_rating>-2 </bidder_rating>
      </high_bidder>
      <num_items>1 </num_items>
      <num_bids> 12</num_bids>
      <started_at>$1.00 </started_at>
      <bid_increment> </bid_increment>
      <location> USA/Chicago</location>
      <opened> Nov-27-00 04:57:50 PST</opened>
      <closed> Dec-02-00 04:57:50 PST</closed>
      <id_num> 511601118</id_num>
      <notes> </notes>
    </auction_info>
    <bid_history>
      <highest_bid_amount>$620.00 </highest_bid_amount>
      <quantity> 1</quantity>
    </bid_history>
    <item_info>
      <memory> 256MB PC133 SDRAM</memory>
      <hard_drive> 30 GB 7200 RPM IDE Hard Drive</hard_drive>
      <cpu>Pentium III 933 System </cpu>
      <brand> </brand>
    </description>
      NEW Pentium III 933 System - 133 MHz BUS Speed Pentium Motherboard, Intel Pentium II
      Panasonic CD-RW 8x4x32 - ATI All-In-Wonder 128 PRO 32MB AGP Video Card with TV tuner
      V90 US Robotics Fax/Modem, 10/100 Network Card, Microsoft Internet Keyboard, Microso
      Windows 98 2nd Edition is installed for configuration purposes only and then removed
      options. 1 Year warranty on parts and labor (3 years on monitor from mfg). Buyer agr
      through eBay's Billpointand PayPal. SHIPMENT GUARANTEED WITHIN 10 BUSINESS DAYS FROM
      insurance. NO RESERVE PRICE.. Bid with confidence with one of eBay's ID VERIFIED Pow
      cost of the system and possible upgrades the system will be built once payment is re
      Please allow 3 to 5 additional business days for shipping VIA UPS Ground Service. DO
    </description>
  </item_info>
</listing>
</root>
```


**In-class Assignment*

(ICA) Let's Play with Examples!

- Python & Pandas

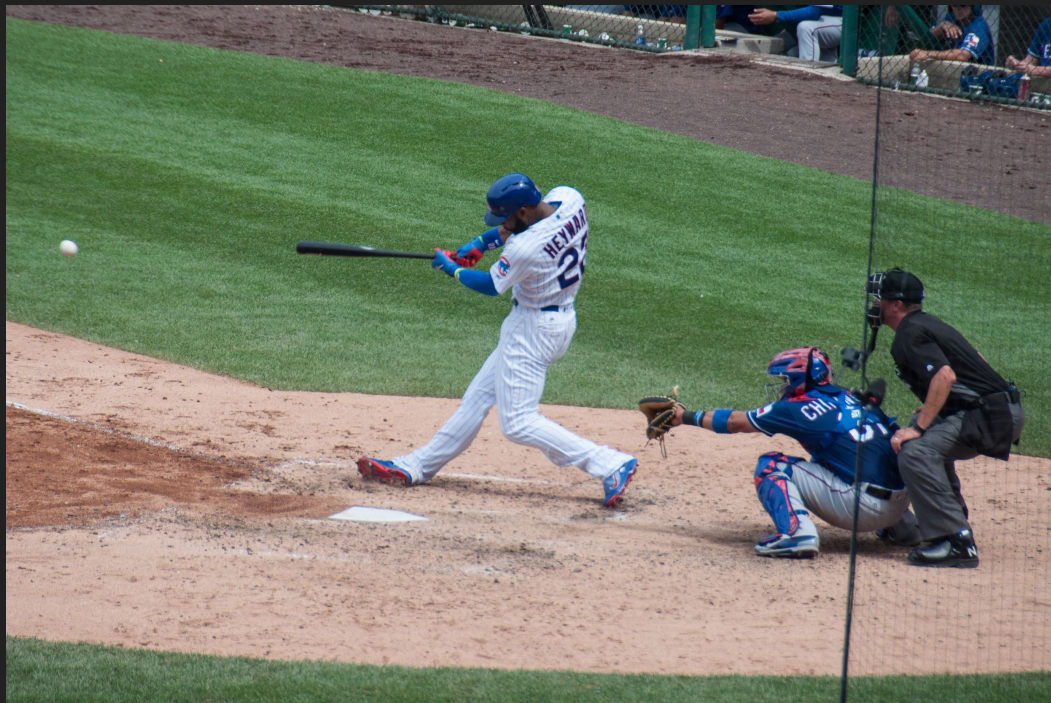


Image Source: Wikipedia