

**MIS 0855 Fall 2016 –
Data Science**

***Week 2 – Data, Information,
and Knowledge***

Min-Seok Pang

**Management Information Systems
Fox School of Business, Temple University
minspang@temple.edu**

Sep. 8th, 2016

Data, Information, and Knowledge

- We tend to use data and information interchangeably.
- Can you think of a difference between data, information, and knowledge?

Data, Information, and Knowledge

- Data : Raw, unorganized facts that describe the characters of an event or an object
- Information : Data that is processed and organized with meaning and value
- Knowledge : Collection of information and data that is useful for decision makings

Data – Example

Student	Exam Score	Attendance
Bob	87	90%
Sue	98	100%
Jack	41	30%
Mary	77	80%
Phillip	65	60%
Steve	88	95%
Helen	72	82%

- What *information* can we find out from this data?

Information – Example

Student	Exam Score	Attendance
Bob	87	90%
Sue	98	100%
Jack	41	30%
Mary	77	80%
Phillip	65	60%
Steve	88	95%
Helen	72	82%

Exam Analysis:

Mean	75.4
Median	77
Lowest Score	41 (Jack)
Highest Score	98 (Sue)

- What *knowledge* can we discover from this information?

Knowledge – Example

Student	Exam Score	Attendance
Bob	87	90%
Sue	98	100%
Jack	41	30%
Mary	77	80%
Phillip	65	60%
Steve	88	95%
Helen	72	82%

Exam Analysis:

Mean	75.4
Median	77
Lowest Score	41 (Jack)
Highest Score	98 (Sue)

- The students who get lower exam scores also have lower attendance.

Data – Another Example

Comcast Corporation (CMCSA) - NasdaqGS ★ Watchlist

58.03 ↑ 0.81 (1.42%) 10:53AM EST - Nasdaq Real Time Price

Prices							
Date	Open	High	Low	Close	Volume	Adj Close*	
Dec 22, 2014	57.22	57.81	57.06	57.22	19,139,100	57.22	
Dec 19, 2014	56.16	57.34	56.16	57.17	20,976,200	57.17	
Dec 18, 2014	55.40	56.54	54.98	56.29	17,210,100	56.29	
Dec 17, 2014	54.74	55.49	54.74	55.34	12,660,600	55.34	
Dec 16, 2014	54.81	55.84	54.74	54.81	15,471,600	54.81	
Dec 15, 2014	55.68	55.75	54.83	54.94	12,780,100	54.94	
Dec 12, 2014	55.00	55.98	54.98	55.35	11,322,400	55.35	
Dec 11, 2014	55.73	56.45	55.48	55.90	10,482,700	55.90	
Dec 10, 2014	56.02	56.34	55.32	55.57	12,149,600	55.57	
Dec 9, 2014	55.94	56.55	55.65	56.19	10,190,800	56.19	
			56.69	56.02	56.36	9,560,200	56.36
			56.73	56.17	56.34	7,633,500	56.34
			57.23	56.10	56.38	15,036,200	56.38
Dec 3, 2014	57.06	57.25	56.52	57.14	10,377,100	57.14	
Dec 2, 2014	56.49	57.16	56.30	57.02	9,929,100	57.02	
Dec 1, 2014	56.97	56.97	56.28	56.60	11,351,300	56.60	
Nov 28, 2014	57.28	57.28	56.78	57.04	5,535,500	57.04	
Nov 26, 2014	56.77	56.91	56.45	56.86	9,200,800	56.86	
Nov 25, 2014	55.64	56.69	55.40	56.62	17,547,400	56.62	
Nov 24, 2014	54.32	55.04	54.26	55.02	12,912,900	55.02	
Nov 21, 2014	54.85	55.19	53.86	54.08	18,891,900	54.08	
Nov 20, 2014	54.37	54.47	54.05	54.38	9,240,400	54.38	
Nov 19, 2014	54.33	54.41	53.91	54.38	16,690,500	54.38	
Nov 18, 2014	54.20	54.58	53.77	54.42	13,817,800	54.42	
Nov 17, 2014	54.01	54.35	53.69	54.15	13,055,700	54.15	
Nov 14, 2014	54.48	54.48	53.77	54.10	10,018,600	54.10	
Nov 13, 2014	53.70	54.61	53.41	54.30	20,752,800	54.30	
Nov 12, 2014	53.00	53.99	52.88	53.61	28,902,800	53.60	
Nov 11, 2014	52.88	53.09	51.56	52.96	38,985,300	52.96	
Nov 10, 2014	55.24	55.25	51.79	52.55	18,812,800	52.55	

● What kind of information can we get from this data?

<http://finance.yahoo.com/q/hp?s=CMCSA>

Information – Another Example

Comcast Corporation (CMCSA) ★ Watchlist
57.99 +0.77(+1.35%) NASDAQ - As of 10:57AM EST

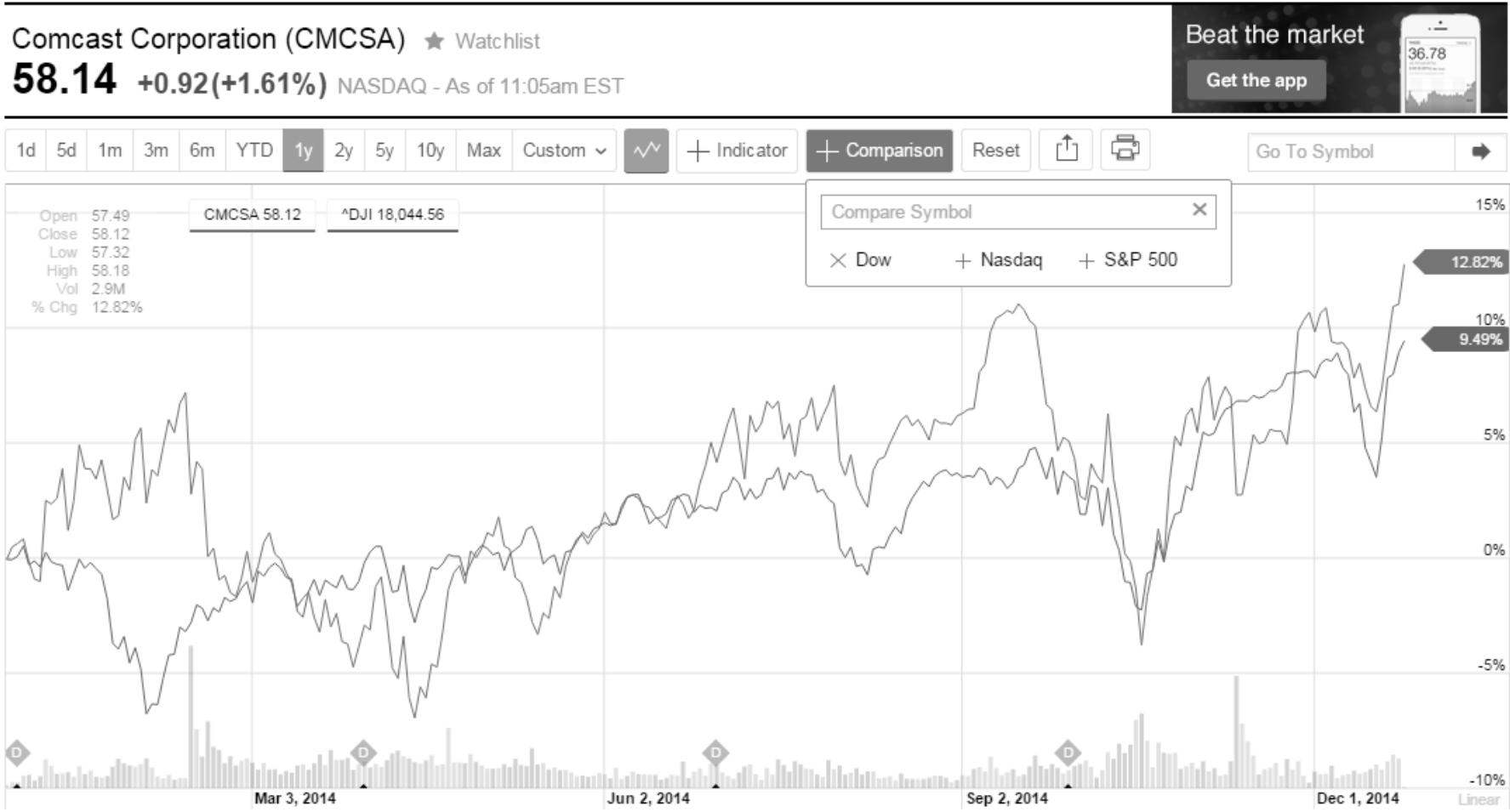
Beat the market

Get the app

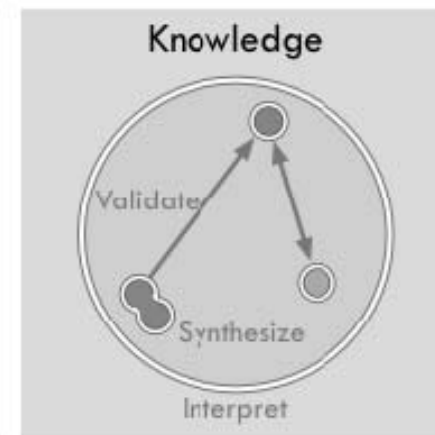
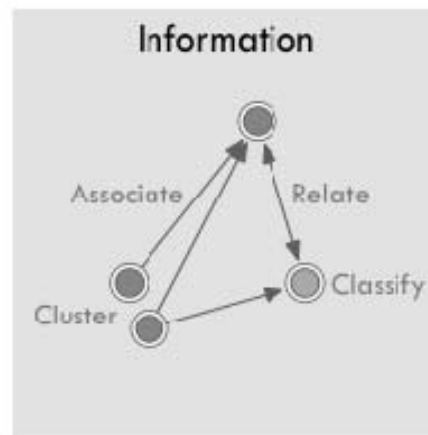
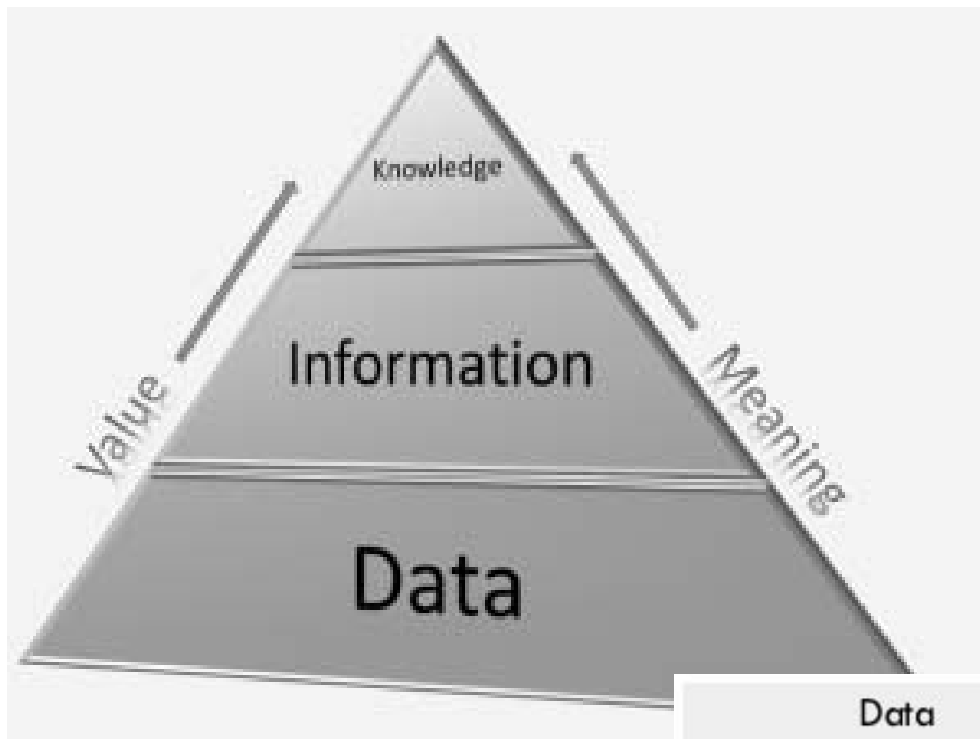


- Comcast stock has risen by 12.8 for the last one year.

Knowledge – Another Example



- Comcast has outperformed the Dow-Jones Index by 3.33%-point.



<http://www.open-bigdata.com/about/>
<http://www.nickdiakopoulos.com/2011/12/16/data-information-knowledge-visualization/>

What Information, Knowledge from Philadelphia Crime Records?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	DC_DIST	SECTOR	DISPATCH_DATE_TIME	DISPATCH_DATE	DISPATCH_HOUR		DC_KEY	LOCATION_BLOCK	UCR_GENERAL	OBJECTID	TEXT_GENERAL_CODE	POINT_X	POINT_Y	SHAPE
2	14	N	2010-05-05 11:34	2010-05-05	11:34:00		1.99814E+11	300 BLOCK E CLIVEDEN	300	12074596	Robbery No Firearm	-75.1772	40.05271	12028957
3	19	A	2010-03-31 12:47	2010-03-31	12:47:00		1.99819E+11	6000 BLOCK HAVERFO	600	12077790	Thefts	-75.2402	39.9705	12032151
4	22	1	2014-09-14 16:00	2014-09-14	16:00:00		1.99822E+11	1800 BLOCK W MONTG	100	12091717	Homicide - Criminal	-75.1568	39.98804	12046078
5	35	E	2007-10-15 17:37	2007-10-15	17:37:00		1.99835E+11	5200 BLOCK N WARNO	700	12075486	Recovered Stolen Mo	-75.1398	40.03251	12029847
6	19	L	2007-08-13 7:40	2007-08-13	7:40:00		2.00019E+11	0 BLOCK N.62ND ST	600	12075475	Theft from Vehicle	-75.2448	39.96251	12029836
7	8	I	2011-08-04 12:49	2011-08-04	12:49:00		2.00308E+11	3600 BLOCK RED LION F	600	12228454	Thefts	-74.9862	40.07513	12182815
8	39	U	2011-09-16 6:16	2011-09-16	6:16:00		2.00339E+11	5000 BLOCK GERMANT	600	12201847	Thefts	-75.1655	40.03021	12156208
9	2	N	2011-08-03 12:43	2011-08-03	12:43:00		2.00402E+11	1000 BLOCK ROSALIE S	600	12076028	Theft from Vehicle	-75.0897	40.03553	12030389
10	15	M	2007-09-10 7:32	2007-09-10	7:32:00		2.00415E+11	6500 BLOCK MARSDEN	700	12090673	Recovered Stolen Mo	-75.0492	40.02361	12045034
11	15	I	2011-08-05 12:57	2011-08-05	12:57:00		2.00415E+11	2100 BLOCK BETTS ST	600	12201848	Thefts	-75.0682	40.01318	12156209
12	25	D	2007-11-06 13:36	2007-11-06	13:36:00		2.00425E+11	4700 BLOCK D ST	700	12095037	Recovered Stolen Mo	-75.1161	40.02195	12049398
13	35	H	2008-04-14 8:36	2008-04-14	8:36:00		2.00435E+11	900 BLOCK W LOUDEN	700	11606771	Recovered Stolen Mo	-75.1391	40.02502	11561132
14	2	O	2007-11-27 19:13	2007-11-27	19:13:00		2.00502E+11	5900 BLOCK RISING SU	300	12281692	Robbery Firearm	-75.1022	40.04424	12236053
15	39	S	2007-08-10 11:18	2007-08-10	11:18:00		2.00539E+11	5000 BLOCK COPLEY RE	700	12070595	Motor Vehicle Theft	-75.1742	40.02045	12024956
16	1	E	2006-01-01 12:21	2006-01-01	12:21:00		2.00601E+11	2500 BLOCK S ROSEWC	700	12076261	Recovered Stolen Mo	-75.1716	39.91946	12030622
17	1	G	2006-01-01 14:08	2006-01-01	14:08:00		2.00601E+11	1500 BLOCK W MOYAM	600	12072181	Thefts	-75.1731	39.91645	12026542
18	1	J	2006-01-01 17:26	2006-01-01	17:26:00		2.00601E+11	2400 BLOCK S 24TH ST	600	12079316	Thefts	-75.187	39.92192	12033677
19	1	J	2006-01-02 10:12	2006-01-02	10:12:00		2.00601E+11	S 23RD ST /RITNER ST	600	12069753	Thefts	-75.1852	39.92258	12024114
20	1	A	2006-01-02 12:36	2006-01-02	12:36:00		2.00601E+11	1400 BLOCK MIFFLIN S	600	12067807	Theft from Vehicle	-75.1693	39.92678	12022168
21	1	G	2006-01-02 13:21	2006-01-02	13:21:00		2.00601E+11	1900 BLOCK JOHNSTON	600	12069328	Thefts	-75.1804	39.91667	12023689
22	1	F	2006-01-03 7:05	2006-01-03	7:05:00		2.00601E+11	2200 BLOCK S 20TH ST	600	12074130	Theft from Vehicle	-75.1797	39.92335	12028491
23	1	D	2006-01-03 12:31	2006-01-03	12:31:00		2.00601E+11	S 29TH ST /MC KEAN ST	300	12165554	Robbery No Firearm	-75.1939	39.92874	12119915
24	1	C	2006-01-03 15:01	2006-01-03	15:01:00		2.00601E+11	1700 BLOCK W MOYAM	300	12085171	Robbery No Firearm	-75.1750	39.91523	12040533

NSA Collection of “Metadata”

- What is this report saying?
- According to this report, did the NSA listen to our actual phone conversations?
- Which information then did they collect?



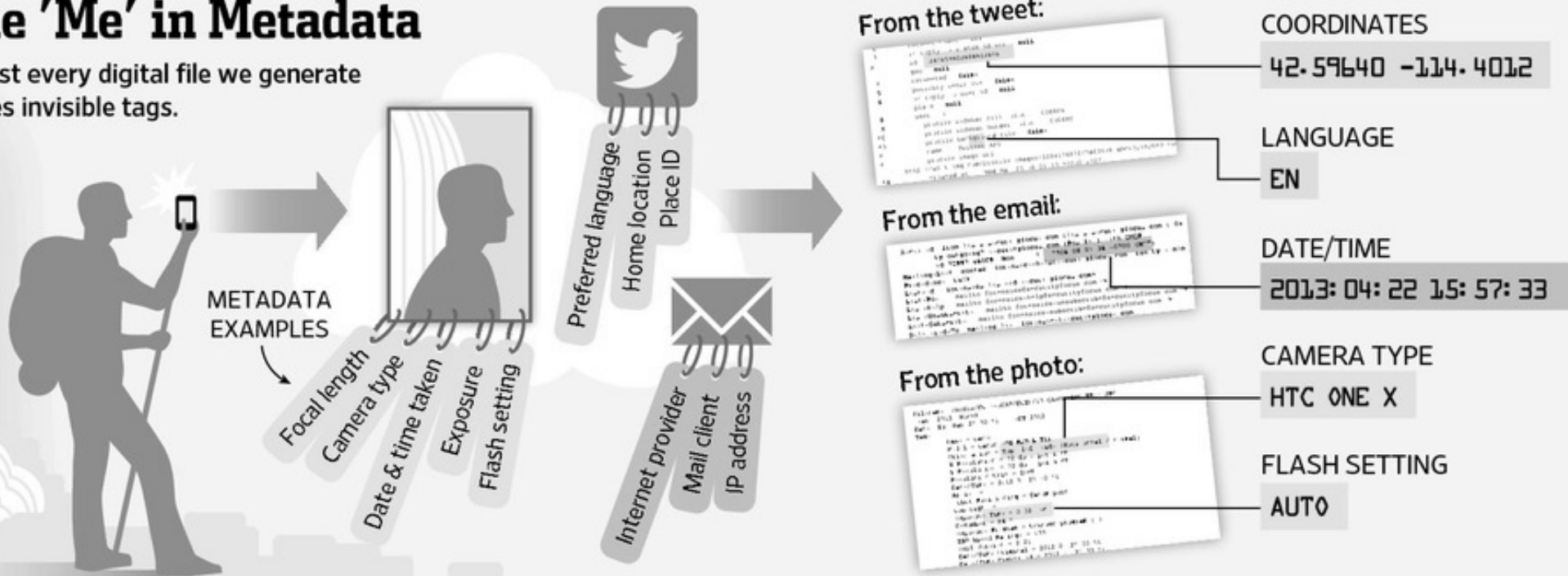
<http://mashable.com/2013/06/07/nsa-prism-faq/>

Metadata

- Data Dictionary
- “Data for data”
- Data that explains what this data is, what it is for, and how it is structured
 - e.g. title, description, data type

The 'Me' in Metadata

Almost every digital file we generate carries invisible tags.



1 Geoff poses by a waterfall and snaps a self-portrait, which he immediately tweets, then emails to his grandmother.

2 Geoff's text, photo, and email ascend to a series of remote servers, each dragging their own trails of metadata.

3 Once there, the metadata may be extracted and interpreted by any interested party with access.

Thanks to the above metadata, without ever having met Geoff, we know he was at Shoshone Falls near Twin Falls, Idaho, at 3:57 p.m. on April 22, that he has an HTC One X smartphone, and that he is an English language speaker.

Note: Actual metadata code modified for readability. Many metadata values, such as time, can be extracted from multiple sources, and values may differ slightly.

Source: staff reporting

The Wall Street Journal

How would you describe this data?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	DC_DIST	SECTOR	DISPATCH_DATE_TIME	DISPATCH_DATE	DISPATCH_HOUR		DC_KEY	LOCATION_BLOCK	UCR_GENERAL	OBJECTID	TEXT_GENERAL_CODE	POINT_X	POINT_Y	SHAPE
2	14	N	2010-05-05 11:34	2010-05-05	11:34:00		1.99814E+11	300 BLOCK E CLIVEDEN	300	12074596	Robbery No Firearm	-75.1772	40.05271	12028957
3	19	A	2010-03-31 12:47	2010-03-31	12:47:00		1.99819E+11	6000 BLOCK HAVERFOF	600	12077790	Thefts	-75.2402	39.9705	12032151
4	22	1	2014-09-14 16:00	2014-09-14	16:00:00		1.99822E+11	1800 BLOCK W MONTG	100	12091717	Homicide - Criminal	-75.1568	39.98804	12046078
5	35	E	2007-10-15 17:37	2007-10-15	17:37:00		1.99835E+11	5200 BLOCK N WARNO	700	12075486	Recovered Stolen Mo	-75.1398	40.03251	12029847
6	19	L	2007-08-13 7:40	2007-08-13	7:40:00		2.00019E+11	0 BLOCK N.62ND ST	600	12075475	Theft from Vehicle	-75.2448	39.96251	12029836
7	8	I	2011-08-04 12:49	2011-08-04	12:49:00		2.00308E+11	3600 BLOCK RED LION F	600	12228454	Thefts	-74.9862	40.07513	12182815
8	39	U	2011-09-16 6:16	2011-09-16	6:16:00		2.00339E+11	5000 BLOCK GERMANT	600	12201847	Thefts	-75.1655	40.03021	12156208
9	2	N	2011-08-03 12:43	2011-08-03	12:43:00		2.00402E+11	1000 BLOCK ROSALIE S	600	12076028	Theft from Vehicle	-75.0897	40.03553	12030389
10	15	M	2007-09-10 7:32	2007-09-10	7:32:00		2.00415E+11	6500 BLOCK MARSDEN	700	12090673	Recovered Stolen Mo	-75.0492	40.02361	12045034
11	15	I	2011-08-05 12:57	2011-08-05	12:57:00		2.00415E+11	2100 BLOCK BETTS ST	600	12201848	Thefts	-75.0682	40.01318	12156209
12	25	D	2007-11-06 13:36	2007-11-06	13:36:00		2.00425E+11	4700 BLOCK D ST	700	12095037	Recovered Stolen Mo	-75.1161	40.02195	12049398
13	35	H	2008-04-14 8:36	2008-04-14	8:36:00		2.00435E+11	900 BLOCK W LOUDEN	700	11606771	Recovered Stolen Mo	-75.1391	40.02502	11561132
14	2	O	2007-11-27 19:13	2007-11-27	19:13:00		2.00502E+11	5900 BLOCK RISING SU	300	12281692	Robbery Firearm	-75.1022	40.04424	12236053
15	39	S	2007-08-10 11:18	2007-08-10	11:18:00		2.00539E+11	5000 BLOCK COPLEY RE	700	12070595	Motor Vehicle Theft	-75.1742	40.02045	12024956
16	1	E	2006-01-01 12:21	2006-01-01	12:21:00		2.00601E+11	2500 BLOCK S ROSEWC	700	12076261	Recovered Stolen Mo	-75.1716	39.91946	12030622
17	1	G	2006-01-01 14:08	2006-01-01	14:08:00		2.00601E+11	1500 BLOCK W MOYAM	600	12072181	Thefts	-75.1731	39.91645	12026542
18	1	J	2006-01-01 17:26	2006-01-01	17:26:00		2.00601E+11	2400 BLOCK S 24TH ST	600	12079316	Thefts	-75.187	39.92192	12033677
19	1	J	2006-01-02 10:12	2006-01-02	10:12:00		2.00601E+11	S 23RD ST /RITNER ST	600	12069753	Thefts	-75.1852	39.92258	12024114
20	1	A	2006-01-02 12:36	2006-01-02	12:36:00		2.00601E+11	1400 BLOCK MIFFLIN S	600	12067807	Theft from Vehicle	-75.1693	39.92678	12022168
21	1	G	2006-01-02 13:21	2006-01-02	13:21:00		2.00601E+11	1900 BLOCK JOHNSTON	600	12069328	Thefts	-75.1804	39.91667	12023689
22	1	F	2006-01-03 7:05	2006-01-03	7:05:00		2.00601E+11	2200 BLOCK S 20TH ST	600	12074130	Theft from Vehicle	-75.1797	39.92335	12028491
23	1	D	2006-01-03 12:31	2006-01-03	12:31:00		2.00601E+11	S 29TH ST /MC KEAN ST	300	12165554	Robbery No Firearm	-75.1939	39.92874	12119915
24	1	C	2006-01-03 15:01	2006-01-03	15:01:00		2.00601E+11	1700 BLOCK W MOYAM	300	12085171	Robbery No Firearm	-75.1750	39.91523	12040533

Metadata for Philadelphia Crime Logs

Field Name	Alias	Description	Type
DC_Dist	District	A two character field that names the District boundary.	Text
DC_Key	DC Number	The unique identifier of the crime that consists of Year + District + Unique ID.	Text
Dispatch_Date_Time	Dispatch Date/Time	The date and time that the officer was dispatched to the scene.	Date/Time
Hour		The generalized hour of the dispatched time.	Date/Time
Location_Block	Location Block	The location of crime generalized by street block.	Text
Sector	PSA	A single character field that names the Police Service Area boundary.	Text
Text_General_Code	General Crime Category	The generalized text for the crime code.	Text
UCR_General	UCR Code	The rounded crime code, i.e. 614 to 600. Part 1 100 - HOMICIDE - CRIMINAL (Includes: 111-116)	Numeric

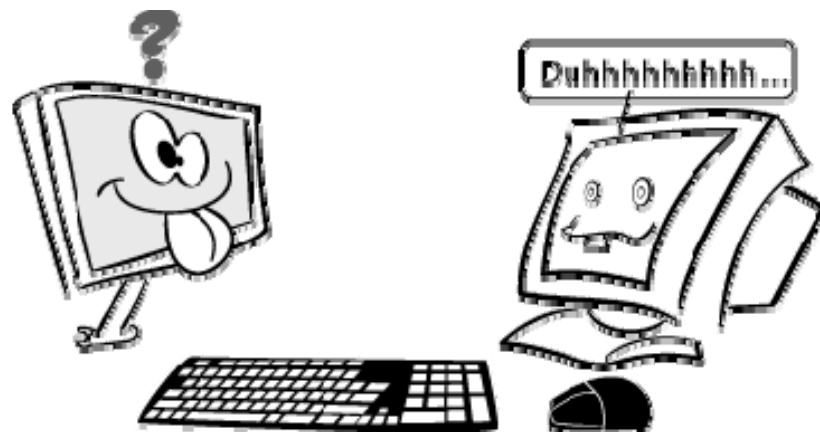
<http://metadata.phila.gov/#home/datasetdetails/5543868920583086178c4f8e/representationdetails/570e7621c03327dc14f4b68d/>

Data Types

Type	Description	Examples	What it's called in Excel	What it's called in Tableau
Integer	Whole numbers	35, 102, -40, 0	Number	Number (whole)
Floating point	Fractional values	3.56, -1.0, 10.123	Number	Number (decimal)
Boolean	Binary (2) values	True/False, Male/Female	N/A	Boolean
String	Numeric and non-numeric characters	Bob, I like cheese, hello123	Text	String
Date/Time	Calendar date and time	8/31/2014, 10:05 AM, 8/31/2014 10:05 AM	Date	Date or Date & time

Why do we need Metadata?

- Because computers are dumb!
- It is getting there, but computers are still not able to understand what the data means and what it is for.
- So, we have to describe it for them.



http://www.manning.com/sande/excerpt_preface.html

Metadata is everywhere



MIS0855 Syllabus [Schuff] - Fall 2014 Properties

Property	Value
Content	
Content status	
Content type	
Pages	1
Word count	4122
Character count	23498
Line count	195
Paragraph count	55
Template	Normal
Scale	No
Links dirty?	No
Language	
File	
Size	53.2 KB
Date created	7/14/2014 2:54 PM
Date modified	9/1/2014 10:19 AM
Date accessed	9/1/2014 10:19 AM
Offline availability	

Kindle Store > Kindle eBooks > Business & Money > Management & Leadership > Leadership

How Google Works [Kindle Edition]
 Eric Schmidt (Author), Jonathan Rosenberg (Author)
 ★★★★★ (162 customer reviews)

Print List Price: ~~\$30.00~~
 Kindle Price: **\$10.99**
 You Save: \$19.01 (63%)
 Sold by: Hachette Book Group

- Word Wise: **Enabled**
- Don't have a Kindle? [Get your Kindle here.](#)
- Whispersync for Voice: **Ready**

Free Kindle Reading App
 Anybody can read Kindle books—even without a Kindle device

The New York Times | <http://nyti.ms/1sSSAQD> **GET TICKET**

N.Y. / REGION

Cuomo to Propose 2 New Minimum Wages, for New York City and State

By **JESSE MCKINLEY** JAN. 18, 2015