

# Database & Data Mining

- **Database:** Integrated collection of data stored in records, like a filing cabinet/container.
- **DBMS:** Access data through DBMS (Database management system) is a software system used to manage the data (does all the work)

# Data Type (by which organized)

- **Date:** Calendar date
- **Char:** alphabetical characters
- **Strings:** Words and characters
- **Boolean:** True/false, 1/0, yes/no
- **Numeric:** Integers, floating-point
- **Alphanumeric:** Alphabetical and numeric combined (ASCII)

# Data Table: NASA Spitzer Observations

TargetName	RA(J2000)	Dec(J2000)	PI	ProgName	pid	AOT	min_dur	Scheduled_Time(UTC)	AOR_Key	AOR_LABEL
SDS0207+0000	2: 7:42.80	0: 0:56.00	Luhman	KLUHMANY	50013	iracmap	7.09	2009-01-30 04:58:49.9	25210624	epoch2-SDS0207+0000
WD1150-153	11:53:15.40	-15:36:37.00	von Hipp	DAZD_VAR	40369	iracmap	10.92	2009-01-30 05:14:32.0	22959104	IRAC-WD1150-153-ep4
NGC3666	11:24:23.50	11:19:51.00	Fazio	MASHBY	50128	iracmap	8.04	2009-01-30 05:25:53.7	25495808	REF_SRVY_IRAC - NGC3666
Q11267+1558	11:29:23.95	15:41:50.44	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 05:31:42.8	25497088	REF_SRVY_IRAC - Q11267+1558
IC602	10:18:19.88	7: 2:51.23	Fazio	MASHBY	50128	iracmap	8.51	2009-01-30 05:39:50.6	25487360	REF_SRVY_IRAC - IC602
IRAS1024+204	10:27:25.22	20:26:52.77	Fazio	MASHBY	50128	iracmap	8.53	2009-01-30 05:47:22.4	25488128	REF_SRVY_IRAC - IRAS1024+2042
NGC3323	10:39:38.43	25:19:22.81	Fazio	MASHBY	50128	iracmap	8.53	2009-01-30 05:53:57.0	25489664	REF_SRVY_IRAC - NGC3323
IC2598	10:39:42.63	26:43:35.71	Fazio	MASHBY	50128	iracmap	8.53	2009-01-30 05:59:45.7	25489920	REF_SRVY_IRAC - IC2598
NGC3381	10:48:24.69	34:42:38.14	Fazio	MASHBY	50128	iracmap	8.53	2009-01-30 06:06:34.7	25490944	REF_SRVY_IRAC - NGC3381
CGCG241-07	11: 6:37.31	46: 2:20.25	Fazio	MASHBY	50128	iracmap	8.51	2009-01-30 06:13:53.5	25492736	REF_SRVY_IRAC - CGCG241-07
NGC3652	11:22:38.77	37:45:53.00	Fazio	MASHBY	50128	iracmap	8.49	2009-01-30 06:20:53.9	25494784	REF_SRVY_IRAC - NGC3652
IRAS11102+30	11:12:57.50	30:10:28.30	Fazio	MASHBY	50128	iracmap	8.49	2009-01-30 06:27:34.0	27053568	REF_SRVY_IRAC - IRAS11102+3026
IRAS11069+27	11: 9:38.26	26:55: 6.14	Fazio	MASHBY	50128	iracmap	8.49	2009-01-30 06:33:37.8	25492992	REF_SRVY_IRAC - IRAS11069+271
NGC3781	11:39: 3.16	26:21:47.47	Fazio	MASHBY	50128	iracmap	8.47	2009-01-30 06:40:13.3	25499392	REF_SRVY_IRAC - NGC3781
NGC3912	11:50: 4.76	26:28:37.94	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 06:46:12.0	25501952	REF_SRVY_IRAC - NGC3912
NGC4018	11:58:41.09	25:18:53.81	Fazio	MASHBY	50128	iracmap	8.44	2009-01-30 06:52:01.5	25504512	REF_SRVY_IRAC - NGC4018
NGC4004	11:58: 5.12	27:52:36.84	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 06:57:59.6	25503744	REF_SRVY_IRAC - NGC4004
HS 1236+3937	12:39:20.14	39:21: 4.80	Galliano	FREDUS	50550	iracmap	5.90	2009-01-30 07:05:25.6	26387712	HS1236+3937-IRAC
G104.257+00.	22:23:27.92	57:51: 1.12	Carey	SMOG	50398	iracmap	168.11	2009-01-30 07:14:43.1	28963328	SMOG_IG104.257+00.447_v5
G104.333+00.	22:24:35.14	57:45:26.18	Carey	SMOG	50398	iracmap	168.11	2009-01-30 10:00:29.3	28958976	SMOG_IG104.333+00.289_v5
G104.410+00.	22:25:42.01	57:39:49.02	Carey	SMOG	50398	iracmap	168.11	2009-01-30 12:46:15.5	28963584	SMOG_IG104.410+00.130_v5
KISSR1516	13:15:49.40	43:34:30.10	Salzer	SALZER	492	iracmap	10.62	2009-01-30 17:30:42.8	28318720	K1516-IRAC
NGC4868	12:59: 7.42	37:18:47.61	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 17:39:21.6	25516544	REF_SRVY_IRAC - NGC4868
NGC5014	13:11:31.33	36:16:51.61	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 17:45:17.0	25518080	REF_SRVY_IRAC - NGC5014
HS 1304+3529	13: 6:24.24	35:13:41.60	Galliano	FREDUS	50550	iracmap	5.92	2009-01-30 17:50:57.1	26387968	HS1304+3529-IRAC
UGC8179	13: 5:12.45	32: 0:12.60	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 17:54:28.4	25516800	REF_SRVY_IRAC - UGC8179
AXJ131021+30	13:10:24.57	30:19:28.30	Rieke	OBSBLAGN	50138	iracmap	7.90	2009-01-30 18:00:18.5	25632000	IRAC-AXJ131021
KISSR0225	13:10: 9.40	29:17:42.90	Salzer	SALZER	492	iracmap	10.64	2009-01-30 18:05:19.5	28315648	K0225-IRAC
UGC8017	12:52:53.51	28:22:16.12	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 18:13:43.0	25515008	REF_SRVY_IRAC - UGC8017
MCG5-30-069	12:44:42.13	26:25: 3.72	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 18:19:38.7	25512192	REF_SRVY_IRAC - MCG5-30-069
NGC4747	12:51:45.96	25:46:38.30	Fazio	MASHBY	50128	iracmap	10.05	2009-01-30 18:25:22.9	25514752	REF_SRVY_IRAC - NGC4747
NGC5012	13:11:37.23	22:55: 7.67	Fazio	MASHBY	50128	iracmap	10.06	2009-01-30 18:33:16.2	25518336	REF_SRVY_IRAC - NGC5012
ddo170	13:15:53.91	25:26: 7.70	McGaugh	MCGAUGH	50111	iracmap	25.70	2009-01-30 18:40:48.3	25438464	IRAC-ddo170
IC856	13:10:42.00	20:32: 7.45	Fazio	MASHBY	50128	iracmap	8.47	2009-01-30 19:04:22.1	25517568	REF_SRVY_IRAC - IC856
NGC4495	12:31:23.71	29: 8: 8.41	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 19:11:37.7	25510144	REF_SRVY_IRAC - NGC4495
NGC4509	12:33: 6.68	32: 5:33.42	Fazio	MASHBY	50128	iracmap	8.44	2009-01-30 19:17:39.0	25510400	REF_SRVY_IRAC - NGC4509
Q12468+3436	12:49:16.77	34:19:50.72	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 19:23:53.6	25513984	REF_SRVY_IRAC - Q12468+3436
NGC4062	12: 4: 3.42	31:53:40.82	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 19:30:54.8	25505792	REF_SRVY_IRAC - NGC4062
UGC7017	12: 2:22.04	29:51:40.75	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 19:36:42.3	25505024	REF_SRVY_IRAC - UGC7017
Q11571+3003	11:59:42.79	29:47:11.77	Fazio	MASHBY	50128	iracmap	8.45	2009-01-30 19:42:07.2	25504768	REF_SRVY_IRAC - Q11571+3003
GJ 436	11:42:11.09	26:42:23.70	Knutson	GJ436TRA	50051	irac	256.40	2009-01-30 19:51:18.3	28894976	p50051_IRAC_GJ436_ch2_sub1_vis
UGC5881	10:46:42.08	25:55:54.47	Fazio	MASHBY	50128	iracmap	8.51	2009-01-31 00:10:07.9	25490432	REF_SRVY_IRAC - UGC5881
NGC3758	11:36:28.80	21:35:45.62	Fazio	MASHBY	50128	iracmap	8.47	2009-01-31 00:17:16.9	25498624	REF_SRVY_IRAC - NGC3758
G104.486+00.	22:26:48.53	57:34: 9.67	Carey	SMOG	50398	iracmap	168.11	2009-01-31 00:30:39.6	28957952	SMOG_IG104.486+00.028_v5
IRAC-CLAE-19	10: 1:58.52	1:54:15.20	Capak	CAPAK	50310	iracmap	185.00	2009-01-31 04:42:48.3	25953280	IRAC-0000 - IRAC-CLAE-19
IRAC-CLAE-29	10: 2:16.77	1:55:39.10	Capak	CAPAK	50310	iracmap	184.98	2009-01-31 07:44:29.5	25955840	IRAC-0000 - IRAC-CLAE-29
G106.024+3.1	22:23:58.81	61: 2:53.31	Carey	SMOG	50398	iracmap	89.39	2009-01-31 11:00:46.8	28961280	SMOG_IG106.024+03.113_v5
G106.905+02.	22:30:47.37	61:23:45.12	Carey	SMOG	50398	iracmap	155.00	2009-01-31 12:27:54.5	28967936	SMOG_IG106.905+02.980_v5
G106.982+02.	22:32: 0.76	61:17:55.91	Carey	SMOG	50398	iracmap	155.00	2009-01-31 15:00:31.3	28967168	SMOG_IG106.982+02.822_v5
SDS1021-0304	10:21: 9.70	-3: 4:20.00	Luhman	KLUHMANY	50013	iracmap	7.09	2009-01-31 19:22:28.5	25210112	epoch2-SDS1021-0304
IC676	11:12:40.50	9: 3:16.22	Fazio	MASHBY	50128	iracmap	8.47	2009-01-31 19:29:00.8	25493248	REF_SRVY_IRAC - IC676
3c273	12:29: 6.70	2: 3: 8.60	Wehrle	ARCHETYP	50231	iracmap	10.61	2009-01-31 19:36:54.8	28975616	IRAC 1 3c273
G107.058+02.	22:33:13.71	61:12: 4.29	Carey	SMOG	50398	iracmap	155.00	2009-01-31 19:53:20.0	28967680	SMOG_IG107.058+02.664_v5

# MS Excel Spreadsheet

- Consists of multiple cells usually in a two-dimensional matrix or grid consisting of rows and columns
- Each cell contains texts, or numeric values, or formulas
- Can be turned into a database by organizing data into records

ckoutDateTime	ckinDateTime	collcode	itemtype	barcode	title	callNumber	deweyCla	subject1	subject2
2007-12-26 10:14:00	2008-01-...	canew	acbk	0010058061416	dead dont lie an Abe Lie...	MYSTERY KAMI...	NULL	Mystery fiction	Chicago Ill Fiction
2007-12-26 10:14:00	2008-01-...	canew	acbk	0010059801349	Fresh kills	MYSTERY NADE...	NULL	Mystery fiction	Private investig...
2007-12-26 10:14:00	2008-01-...	canew	acbk	0010056587834	Fresh disasters	MYSTERY WOO...	NULL	Suspense fiction	Mystery fiction
1970-01-01 00:00:00	2008-01-...	ccpic	jcbk	0010059898170	Polar bear puzzle	E LUMRY	NULL	Polar bear Fiction	Bears Fiction
2007-08-30 11:39:00	2008-01-...	ncef	jcbk	0010048749351	Double dragon trouble	J/E MCMULLA	NULL	Wizards Fiction	Kidnapping Fiction
2007-12-12 12:39:00	2008-01-...	nanf	acbk	0010058241208	long way gone memoirs ...	966.404 B359B...	966.404	Sierra Leone So...	Beah Ishmael 1...
2007-12-04 15:54:00	2008-01-...	nanf	acbk	0010052524096	Buddhism for beginners	294.3 T421B 2001	294.3	Buddhism Doct...	Washington Aut...
2007-12-05 17:00:00	2008-01-...	canf	acbk	0010042973049	Galloways book on running	796.42 G1384...	796.42	Running	Running Training
2007-11-28 11:42:00	2008-01-...	canf	acbk	0010045468849	Wild Seattle a celebration...	508.79777 D71...	508.79777	Natural areas W...	Natural history ...
2007-05-20 12:51:00	2008-01-...	cab	acbk	0010000794452	Fantin Latour	B F218K	NULL	Fantin Latour H...	NULL
2007-12-17 12:00:00	2008-01-...	canf	acbk	0010051040920	Sweet gratitude bake a t...	641.86 Su876S ...	641.86	Desserts	Baking
2007-10-29 13:31:00	2008-01-...	canf	acbk	0000105443394	Volkswagen Vanagon offi...	629.2873 V885...	629.2873	Vanagon van M...	NULL
2007-12-05 15:43:00	2008-01-...	nacd	accd	0010059798966	Christmas with Julie And...	CD 782.42172...	782.421...	Christmas music	Carols English
2007-12-12 17:02:00	2008-01-...	nanf	acbk	0010056488231	Cultural amnesia necess...	909.09821 J23...	909.09821	Civilization Wes...	Humanism Hist...
2007-12-02 15:17:00	2008-01-...	navid	acvhs	0010044771177	phantom train of doom	VHS PHANTOM	NULL	Adventure films	Video recording...
2007-09-27 14:39:00	2008-01-...	canf	acbk	0010056967101	secret art of Seamm Jasa...	613.09515 As1...	613.09515	Medicine Tibetan	Self care Health...
2007-11-29 09:21:00	2008-01-...	nccd	jccd	0010053504352	Kidz bop Christmas	CDJ 782.21723...	782.21723	Carols English	Songs
2007-12-12 15:21:00	2008-01-...	nadvdnf	acdvd	0010058096255	No end in sight	DVD 956.7044...	956.70443	Iraq War 2003	United States Hi...
2007-11-07 19:54:00	2008-01-...	ncbot	jccas	0010036490075	cricket in Times Square	J Se48C CASSETTE	NULL	Cats Fiction	Mice Fiction
2007-12-06 15:30:00	2008-01-...	nanf	acbk	0010050303659	Hand made modern mid ...	747 OL19H 2005	747	Handicraft	Interior decoration
2007-12-16 15:50:00	2008-01-...	nadvdnf	acdvd	0010057224163	best mind since Einstein	DVD 530.092 F...	530.092	Feynman Richar...	Physicists Unite...
2007-10-29 18:18:00	2008-01-...	nanf	acbk	0010046572979	Looking for Spinoza joy ...	152.4 D18L 2003	152.4	Spinoza Benedi...	Emotions
2007-12-02 17:22:00	2008-01-...	canf	acbk	0010051467909	Dirty blonde the diaries ...	782.42166 L94...	782.42166	Rock musicians...	Love Courtney ...
2007-12-19 17:46:00	2008-01-...	nacd	accd	0010053575931	Retro	CD 782.42166 ...	782.42166	Popular music	Rock music
2007-10-28 17:59:00	2008-01-...	canf	acbk	0010036192630	Paris fashion a cultural h...	391.00944 STE...	391.00944	Fashion France ...	Paris France So...
2007-11-30 14:38:00	2008-01-...	cafic	acbk	0010058509505	Dingley Falls a novel	FIC MALONE	NULL	Domestic fiction	Anonymous lett...
2007-12-05 10:21:00	2008-01-...	nadvd	acdvd	0010056697880	good year	DVD GOOD YE	NULL	Man woman rel...	Inheritance and...
2007-11-29 09:21:00	2008-01-...	nccd	jccd	0010049182701	Lion King 1 1 2 songs fr...	CDJ 781.542 L661	781.542	Songs	Childrens song...
2007-12-11 15:21:00	2008-01-...	nadvd	acdvd	0010056685075	Mulholland Dr	DVD MULHOLL	NULL	Mystery films	Video recording...
2007-09-15 17:29:00	2008-01-...	canf	acbk	0010032059411	Taming the office tiger t...	650.1 HEMPHIL ...	650.1	Time management	Paperwork Offic...
2007-12-02 16:04:00	2008-01-...	navid	acvhs	0010039944607	Happy Texas	VHS HAPPY T	NULL	Video recording...	Feature films
2007-11-27 19:02:00	2008-01-...	nanf	acbk	0010054056113	fashion designer survival...	658.11024 G27...	658.11024	Fashion mercha...	Fashion designers
2007-11-27 14:00:00	2008-01-...	nanf	acbk	0010059478577	rest is noise listening to ...	780.904 R7331...	780.904	Music 20th cen...	NULL
2007-12-18 16:22:00	2008-01-...	nacd	accd	0010055500069	Rockin bones 1950s pun...	CD 782.42166 ...	782.42166	Rockabilly music	Rock music To ...
2007-12-05 12:09:00	2008-01-...	nanf	acbk	0010049536914	devil in the white city mu...	364.1523 L329...	364.1523	Worlds Columbi...	Mudgett Herma...
2007-12-07 12:19:00	2008-01-...	nacd	accd	0010042837871	best of Peter Paul and M...	CD 782.42166 ...	782.42166	Popular music ...	NULL
2007-11-06 15:22:00	2008-01-...	cafic	acbk	0010054550305	friend of the family	FIC WILLETT2006	NULL	England Fiction	Man woman rel...

# Data Tables

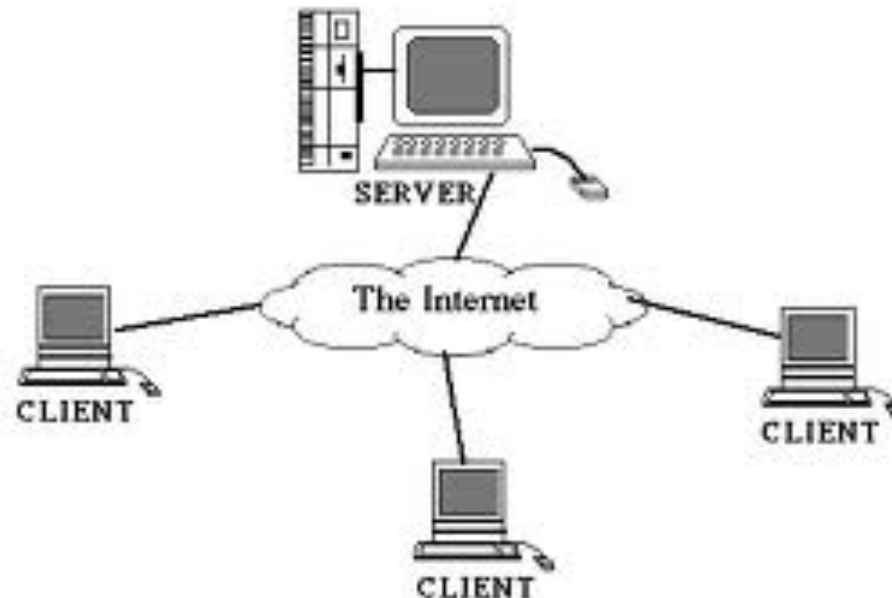
- **Table:** set of data elements (values) that is organized using a model of vertical columns (which are identified by their name) and horizontal rows.
- A table has a specified number of columns, but any number of rows

# Relational: Data into Records

- A database record is a row of data in a database table consisting of a single value from each column of data in the table.
- **Relational Database:** data is stored in tables and the relationships among the data are also stored in tables.
- The data can be accessed or reassembled in many different ways

# Client-Server

- **Client:** Submits requests to the server (Select, delete, insert, etc.)
- **Server:** processes requests, filters, discards, and sorts data – sends results to client software





# SQL (Structured Query Language)

- For Managing data in relational database systems
- **Operations:** data insert, query, update and delete, data access control
  - SELECT ... FROM ... WHERE ...
  - INSERT INTO ... VALUES ...
  - UPDATE ... SET ... WHERE ...
  - DELETE FROM ... WHERE ...

# Data Mining

- Process of discovering new patterns from large data sets
- Extract previously unknown interesting patterns such as
  - groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (Searches for relationships between variables)
- **Exploratory:** Must interact with dataset to discover interesting question
- Outcomes leads to research/projects

George Legrady

Director, Experimental Visualization Lab

Media Arts & Technology PhD program

University of California, Santa Barbara

<http://vislab.ucsb.edu>

<http://www.georgelegrady.com>