



**Hasso  
Plattner  
Institut**

IT Systems Engineering | Universität Potsdam

## Data Profiling Revisited

New York  
November 2013

Felix Naumann

# The Hasso Plattner Institute

2

- Founded in 1998 as a Public Private Partnership
- Hasso Plattner, co-founder of SAP, endowed over 200 Mio. Euro.
- Adjoined with the University of Potsdam
- 500 students

- BA, MA, PhD

- Enterprise Platform and Integration Concepts
- Internet Technologies and Systems
- Human Computer Interaction
- Computer Graphics Systems
- Operating Systems and Middleware
- Business Process Technology
- Software Architecture
- Information Systems
- System Engineering and Modeling





ncvoter1.txt - Microsoft Excel

Datei Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Add-Ins

Einfügen Ausschneiden Kopieren Format übertragen Zwischenablage Schriftart Ausrichtung Zahl Formatvorlagen Zellen

Calibri 11 A A Zeilenumbruch Standard Standard Gut Neutral Schlecht Ausgabe Berechnung Eingabe Erklärender ...

F K U Verbinden und zentrieren Bedingte Formatierung Als Tabelle formatieren

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
2	1	ALAMANCE	9005990	A	ACTIVE	AV	VERIFIED	AABEL	EVELYN	LARSEN		4430 E GREENSBOF	GRAHAM	NC	27253	4430 E GREENSBORO-CHA	GRAHAM	NC		27253	000 00	
3	1	ALAMANCE	9048723	A	ACTIVE	AV	VERIFIED	AARON	CHRISTINA	CASTAGNA		421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177		BURLINGTON	NC		27215	229 11
4	1	ALAMANCE	9019674	A	ACTIVE	AV	VERIFIED	AARON	CLAUDIA	HAYDEN		1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST		BURLINGTON	NC		27215	222 88
5	1	ALAMANCE	9129589	A	ACTIVE	AV	VERIFIED	AARON	JAMES	MICHAEL		1647 SAXAPAHAW	GRAHAM	NC	27253	PO BOX 98		SAXAPAHAW	NC		27340	336 52!
6	1	ALAMANCE	9041748	A	ACTIVE	AV	VERIFIED	AARON	NATHAN	EDWARD		421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177		BURLINGTON	NC		27215	336 22!
7	1	ALAMANCE	9021947	A	ACTIVE	AV	VERIFIED	AARON	WILLIE	DALE		1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST		BURLINGTON	NC		27215	336 99!
8	1	ALAMANCE	9062002	A	ACTIVE	AV	VERIFIED	AARONSON	GENA	HOLT		107 TERRYWOOD	(HAW RIVER	NC	27258	107 TERRYWOOD CT		HAW RIVER	NC		27258	336 57!
9	1	ALAMANCE	9096423	A	ACTIVE	AV	VERIFIED	AARONSON	MICHAEL	CHARLES		107 TERRYWOOD	(HAW RIVER	NC	27258	107 TERRYWOOD CT		HAW RIVER	NC		27258	336 26!
10	1	ALAMANCE	9117940	I	INACTIVE	IU	CONFIRMATI	ABAD	PRISCILLA	MARIE		100 COLONNADE	ELON	NC	27244	CAMPUS BOX 3008		ELON	NC		27244	
11	1	ALAMANCE	9034127	I	INACTIVE	IU	CONFIRMATI	ABADIE	COLLEEN	MIASHEL		1097 IVEY RD	#C GRAHAM	NC	27253	1097 IVEY RD #C		GRAHAM	NC		27253	
12	1	ALAMANCE	9121656	A	ACTIVE	AV	VERIFIED	ABADIE	JACK	EDWARD	JR	612 SIDEVIEW ST	GRAHAM	NC	27253	612 SIDEVIEW ST		GRAHAM	NC		27253	336 21!
13	1	ALAMANCE	9118154	I	INACTIVE	IU	CONFIRMATI	ABADIE	MYRA	HOLLIFIELD		612 SIDEVIEW ST	GRAHAM	NC	27253	617 MITCHELL ST		BURLINGTON	NC		27217	336 21!
14	1	ALAMANCE	9131788	A	ACTIVE	AV	VERIFIED	ABBAS	FALISA			707 SUMMIT RIDG	MEBANE	NC	27302	707 SUMMIT RIDGE RD #	MEBANE	NC		27302	919 56!	
15	1	ALAMANCE	9068460	A	ACTIVE	AV	VERIFIED	ABBAS	RAFAT			514 WESTRIDGE	DIBURLINGTON	NC	27215	514 WESTRIDGE DR		BURLINGTON	NC		27215	
16	1	ALAMANCE	9049573	A	ACTIVE	AV	VERIFIED	ABBATECOLA	RONALD	JOSEPH	JR	504 BROOKFIELD	(GIBSONVILLE	NC	27249	504 BROOKFIELD DR		GIBSONVILLE	NC		27249	336 44!
17	1	ALAMANCE	9033877	A	ACTIVE	AV	VERIFIED	ABBATECOLA	TRACY	BOONE		504 BROOKFIELD	(GIBSONVILLE	NC	27249	504 BROOKFIELD DR		GIBSONVILLE	NC		27249	
18	1	ALAMANCE	9083557	I	INACTIVE	IU	CONFIRMATI	ABBETT	DAWN	LEANN		3900 JOHNS CREEK	GIBSONVILLE	NC	27249	3900 JOHNS CREEK DR		GIBSONVILLE	NC		27249	336 58!
19	1	ALAMANCE	9027554	A	ACTIVE	AV	VERIFIED	ABBEY	BRENT	DAVID		3304 GOLDEN OAK	GRAHAM	NC	27253	3304 GOLDEN OAKS DR		GRAHAM	NC		27253	919 68!
20	1	ALAMANCE	9029477	A	ACTIVE	AV	VERIFIED	ABBEY	DEMETRA	AINSWORTH		3304 GOLDEN OAK	GRAHAM	NC	27253	3304 GOLDEN OAKS DR		GRAHAM	NC		27253	336 37!
21	1	ALAMANCE	9022529	I	INACTIVE	IU	CONFIRMATI	ABBEY	DOROTHY	ESTELLA		1029A QUAKENBU	SNOW CAMP	NC	27349	1029A QUAKENBUSH RD		SNOW CAMP	NC		27349	376 36
22	1	ALAMANCE	9113186	A	ACTIVE	AV	VERIFIED	ABBOTT	AMELIA	BETH		2876 CALLOWAY	DMEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC		27302	919 30!
23	1	ALAMANCE	9087980	A	ACTIVE	AV	VERIFIED	ABBOTT	ANGELA	MORTON		2006 WINN CREEK	HAW RIVER	NC	27258	2006 WINN CREEK DR		HAW RIVER	NC		27258	336 26!
24	1	ALAMANCE	9019273	A	ACTIVE	AV	VERIFIED	ABBOTT	BRENDA	CARMICHAEL		611 N THIRD ST	MEBANE	NC	27302	611 N THIRD ST		MEBANE	NC		27302	563 26
25	1	ALAMANCE	9102615	A	ACTIVE	AV	VERIFIED	ABBOTT	BRIAN	CHRISTOPHE		2006 WINN CREEK	HAW RIVER	NC	27258	2006 WINN CREEK DR		HAW RIVER	NC		27258	336 26!
26	1	ALAMANCE	9079257	A	ACTIVE	AV	VERIFIED	ABBOTT	BRUCE	CLEATON		188 LAKE CAMMA	BURLINGTON	NC	27217	188 LAKE CAMMACK CT		BURLINGTON	NC		27217	336 21!
27	1	ALAMANCE	1389300	A	ACTIVE	AV	VERIFIED	ABBOTT	CHERYL	FAULKNER		188 LAKE CAMMA	BURLINGTON	NC	27217	188 LAKE CAMMACK CT		BURLINGTON	NC		27217	336 22!
28	1	ALAMANCE	9140392	A	ACTIVE	AV	VERIFIED	ABBOTT	CHRISTOPHE	BRANDON		309 BURLINGTON	GIBSONVILLE	NC	27249	309 BURLINGTON AVE		GIBSONVILLE	NC		27249	
29	1	ALAMANCE	9135711	A	ACTIVE	AV	VERIFIED	ABBOTT	COURTNEY	LOVE		309 BURLINGTON	GIBSONVILLE	NC	27249	309 BURLINGTON AVE		GIBSONVILLE	NC		27249	
30	1	ALAMANCE	9028439	A	ACTIVE	AV	VERIFIED	ABBOTT	DWAYNE	ROGER		2839 LADALE LN	MEBANE	NC	27302	2839 LADALE LN		MEBANE	NC		27302	563 39
31	1	ALAMANCE	9090420	A	ACTIVE	AV	VERIFIED	ABBOTT	FRANK	PATRICK		1202 JAMESTOWN	ELON	NC	27244	1202 JAMESTOWNE DR		ELON	NC		27244	336 22!
32	1	ALAMANCE	9079222	A	ACTIVE	AV	VERIFIED	ABBOTT	GLADYS	MARIE MILES		614 TUCKER ST	BURLINGTON	NC	27215	614 TUCKER ST		BURLINGTON	NC		27215	336 57!
33	1	ALAMANCE	9129722	A	ACTIVE	AV	VERIFIED	ABBOTT	HAROLD	GRANT		507 EVERETT ST	#BURLINGTON	NC	27215	507 EVERETT ST #320B		BURLINGTON	NC		27215	336 43!
34	1	ALAMANCE	9094352	A	ACTIVE	AV	VERIFIED	ABBOTT	JESSICA	NADINE		2876 CALLOWAY	DMEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC		27302	919 30!
35	1	ALAMANCE	9023803	A	ACTIVE	AV	VERIFIED	ABBOTT	JOYCE	HODGES		1934 TUCKER ST	#BURLINGTON	NC	27215	1934 TUCKER ST #A		BURLINGTON	NC		27215	336 22!
36	1	ALAMANCE	9084794	R	REMOVED	RS	MOVED FRO	ABBOTT	LATWOIA	BEREA		201 STALEY HALL	ELON	NC	27244	CAMPUS BOX 3039		ELON	NC		27244	
37	1	ALAMANCE	9020357	A	ACTIVE	AV	VERIFIED	ABBOTT	LAWRENCE	ELMER	JR	110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC		27244	336 56!
38	1	ALAMANCE	9108338	A	ACTIVE	AV	VERIFIED	ABBOTT	MARIA	LYNETTE		614 TUCKER ST	BURLINGTON	NC	27215	614 TUCKER ST		BURLINGTON	NC		27215	336 57!
39	1	ALAMANCE	9077192	A	ACTIVE	AV	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC		27244	800 22!
40	1	ALAMANCE	9035500	A	ACTIVE	AV	VERIFIED	ABBOTT	PATTI	BELVIN		1202 JAMESTOWN	ELON	NC	27244	1202 JAMESTOWNE DR		ELON	NC		27244	336 22!
41	1	ALAMANCE	9090949	R	REMOVED	RM	REMOVED A	ABBOTT	RACHEL	MARA		103 DANIELEY	CENELON	NC	27244	CAMPUS BOX 3044		ELON	NC		27244	336 27!
42	1	ALAMANCE	9135295	A	ACTIVE	AV	VERIFIED	ABBOTT	SUSAN	HANKS		2876 CALLOWAY	DMEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC		27302	919 56!
43	1	ALAMANCE	9113731	I	INACTIVE	IU	CONFIRMATI	ABBOTT	TAYLOR	RENEE		406 W LEBANON	A ELON	NC	27244	CAMPUS BOX 3077		ELON	NC		27244	
44	1	ALAMANCE	9120825	I	INACTIVE	IN	CONFIRMATI	ABBOTT	TIFFANY	MURIEL ARLE		144 W CRESCENT	S GRAHAM	NC	27253	144 W CRESCENT SQUARE		GRAHAM	NC		27253	336 23!
45	1	ALAMANCE	9013866	I	INACTIVE	IN	CONFIRMATI	ABBOTT	VIRGINIA	SMITH		2820 BLANCHE DR	BURLINGTON	NC	27215	2820 BLANCHE DR		BURLINGTON	NC		27215	584 46
46	1	ALAMANCE	9027717	A	ACTIVE	AV	VERIFIED	ABBOTT-LUN	SHELBY	LYNN		509 FERNWAY DR	BURLINGTON	NC	27217	509 FERNWAY DR		BURLINGTON	NC		27217	336 22!
47	1	ALAMANCE	9108552	A	ACTIVE	AV	VERIFIED	ABDALLA	KHALED	ISMAIL		605 ISLEY PL	#C BURLINGTON	NC	27215	605 ISLEY PL #C		BURLINGTON	NC		27215	336 68!
48	1	ALAMANCE	9128403	A	ACTIVE	AV	VERIFIED	ABDEL-MAGI	LISA	ANN		1841 DUNBAR PL	BURLINGTON	NC	27215	1841 DUNBAR PL		BURLINGTON	NC		27215	214 43!
49	1	ALAMANCE	9117192	I	INACTIVE	IU	CONFIRMATI	ABDELKARIM	AMNA	ELHAG		1105 PROVIDENCE	ELON	NC	27244	1105 PROVIDENCE CT		ELON	NC		27244	
50	1	ALAMANCE	9099437	A	ACTIVE	AV	VERIFIED	ABDELRAHAI	ABUBAKR	MERGANI		2954 ETHAN POIN	BURLINGTON	NC	27215	2954 ETHAN POINTE DR		BURLINGTON	NC		27215	336 68!



ncvoter1.txt - Microsoft Excel

Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Add-Ins

Einfügen Ausschneiden Kopieren Format übertragen Zwischenablage

Calibri 11 A A Zeilenumbruch Standard Bedingte Formatierung Als Tabelle formatieren

Standard Gut Neutral Schlecht Ausgabe Berechnung Eingabe Erklärender...

Schriftart Ausrichtung Zahl Formatvorlagen Zellen

A1	county_id																					
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	county	county_desc	voter_reg_n	status_cd	voter_status_desc	reason_cd	voter_status	last_name	first_name	midl_name	name	res_street_addr	res_city_desc	state	zip_code	mail_addr1	mail_addr2	mail_city	mail_state	mail_zipcod	full_pl	
2	1	ALAMANCE	9005990	A	ACTIVE	AV	VERIFIED	AABEL	EVELYN	LARSEN		4430 E GREENSBOF	GRAHAM	NC	27253	4430 E GREENSBORO-CHA		GRAHAM	NC		27253	000 00
3	1	ALAMANCE	9048723	A	ACTIVE	AV	VERIFIED	AARON	CHRISTINA	CASTAGNA		421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177		BURLINGTON	NC		27215	229 11
4	1	ALAMANCE	9019674	A	ACTIVE	AV	VERIFIED	AARON	CLAUDIA	HAYDEN		1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST		BURLINGTON	NC		27215	222 88
5	1	ALAMANCE	9129589	A	ACTIVE	AV	VERIFIED	AARON	JAMES	MICHAEL		1647 SAXAPAHAW	GRAHAM	NC	27253	PO BOX 98		SAXAPAHAW	NC		27340	336 52
6	1	ALAMANCE	9041748	A	ACTIVE	AV	VERIFIED	AARON	NATHAN	EDWARD		421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177		BURLINGTON	NC		27215	336 22
7	1	ALAMANCE	9021947	A	ACTIVE	AV	VERIFIED	AARON	WILLIE	DALE		1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST		BURLINGTON	NC		27215	336 99
8	1	ALAMANCE	9062002	A	ACTIVE	AV	VERIFIED	AARONSON	GENA	HOLT		107 TERRYWOOD	HAW RIVER	NC	27258	107 TERRYWOOD CT		HAW RIVER	NC		27258	336 57
9	1	ALAMANCE	9096423	A	ACTIVE	AV	VERIFIED	AARONSON	MICHAEL	CHARLES		107 TERRYWOOD	HAW RIVER	NC	27258	107 TERRYWOOD CT		HAW RIVER	NC		27258	336 26
10	1	ALAMANCE	9117940	I	INACTIVE	IU	CONFIRMATI	ABAD	PRISCILLA	MARIE		100 COLONNADE	ELON	NC	27244	CAMPUS BOX 3008		ELON	NC		27244	
11	1	ALAMANCE	9034111	I	INACTIVE	IU	CONFIRMATI	ABADIE	COLLEEN	MIASHEL		1097 IVEY RD #C	GRAHAM	NC	27253	1097 IVEY RD #C		GRAHAM	NC		27253	
12	1	ALAMANCE	9034111	I	INACTIVE	AV	VERIFIED	ABADIE	JACK	EDWARD	JR	612 SIDEVIEW ST	GRAHAM	NC	27253	612 SIDEVIEW ST		GRAHAM	NC		27253	336 21
13	1	ALAMANCE	9034111	I	INACTIVE	IU	CONFIRMATI	ABADIE	MYRA	HOLLIFIELD		612 SIDEVIEW ST	GRAHAM	NC	27253	617 MITCHELL ST		BURLINGTON	NC		27217	336 21
14	1	ALAMANCE	9034111	I	INACTIVE	AV	VERIFIED	ABBAS	FALISA			707 SUMMIT RIDG	MEBANE	NC	27302	707 SUMMIT RIDGE RD #	MEBANE	NC		27302	919 56	
15	1	ALAMANCE	9034111	I	INACTIVE	AV	VERIFIED	ABBAS	RAFAT			514 WESTRIDGE D	BURLINGTON	NC	27215	514 WESTRIDGE DR		BURLINGTON	NC		27215	
16	1	ALAMANCE	9034111	I	INACTIVE	AV	VERIFIED	ABBATECOLA	RONALD	JOSEPH	JR	504 BROOKFIELD E	GIBSONVILLE	NC	27249	504 BROOKFIELD DR		GIBSONVILLE	NC		27249	336 44
17	1	ALAMANCE	9033877	A	ACTIVE	AV	VERIFIED	ABBATECOLA	TRACY	BOONE		504 BROOKFIELD E	GIBSONVILLE	NC	27249	504 BROOKFIELD DR		GIBSONVILLE	NC		27249	
18	1	ALAMANCE	9083557	I	INACTIVE	IU	CONFIRMATI	ABBETT	DAWN	LEANN		3900 JOHNS CREEK	GIBSONVILLE	NC	27249	3900 JOHNS CREEK DR		GIBSONVILLE	NC		27249	336 58
19	1	ALAMANCE	9027554	A	ACTIVE	AV	VERIFIED	ABBEY	BRENT	DAVID		3304 GOLDEN OAK	GRAHAM	NC	27253	3304 GOLDEN OAKS DR		GRAHAM	NC		27253	919 68
20	1	ALAMANCE	9029477	A	ACTIVE	AV	VERIFIED	ABBEY	DEMETRA	AINSWORTH		3304 GOLDEN OAK	GRAHAM	NC	27253	3304 GOLDEN OAKS DR		GRAHAM	NC		27253	336 37
21	1	ALAMANCE	9022529	I	INACTIVE	IU	CONFIRMATI	ABBEY	DOROTHY	ESTELLA		1029A QUAKENBU	SNOW CAMP	NC	27349	1029A QUAKENBUSH RD		SNOW CAMP	NC		27349	376 36
22	1	ALAMANCE	9113186	A	ACTIVE	AV	VERIFIED	ABBOTT	AMELIA	BETH		2876 CALLOWAY D	MEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC		27302	919 30
23	1	ALAMANCE	9087980	A	ACTIVE	AV	VERIFIED	ABBOTT	ANGELA	MORTON		2006 WINN CREEK	HAW RIVER	NC	27258	2006 WINN CREEK DR		HAW RIVER	NC		27258	336 26
24	1	ALAMANCE	9019273	A	ACTIVE	AV	VERIFIED	ABBOTT	BRENDA	CARMICHAEL		611 N THIRD ST	MEBANE	NC	27302	611 N THIRD ST		MEBANE	NC		27302	563 26
25	1	ALAMANCE	9102615	A	ACTIVE	AV	VERIFIED	ABBOTT	BRIAN	CHRISTOPHE		2006 WINN CREEK	HAW RIVER	NC	27258	2006 WINN CREEK DR		HAW RIVER	NC		27258	336 26
26	1	ALAMANCE	9079257	A	ACTIVE	AV	VERIFIED	ABBOTT	BRUCE	CLEATON		188 LAKE CAMMA	BURLINGTON	NC	27217	188 LAKE CAMMACK CT		BURLINGTON	NC		27217	336 21
27	1	ALAMANCE	1389300	A	ACTIVE	AV	VERIFIED	ABBOTT	CHERYL	FAULKNER		188 LAKE CAMMA	BURLINGTON	NC	27217	188 LAKE CAMMACK CT		BURLINGTON	NC		27217	336 22
28	1	ALAMANCE	9140392	A	ACTIVE	AV	VERIFIED	ABBOTT	CHRISTOPHE	BRANDON		309 BURLINGTON	GIBSONVILLE	NC	27249	309 BURLINGTON AVE		GIBSONVILLE	NC		27249	
29	1	ALAMANCE	9135711	A	ACTIVE	AV	VERIFIED	ABBOTT	COURTNEY	LOVE		309 BURLINGTON	GIBSONVILLE	NC	27249	309 BURLINGTON AVE		GIBSONVILLE	NC		27249	
30	1	ALAMANCE	9028439	A	ACTIVE	AV	VERIFIED	ABBOTT	DWAYNE	ROGER		2839 LADALE LN	MEBANE	NC	27302	2839 LADALE LN		MEBANE	NC		27302	563 39
31	1	ALAMANCE	9090420	A	ACTIVE	AV	VERIFIED	ABBOTT	FRANK	PATRICK		1202 JAMESTOWN	ELON	NC	27244	1202 JAMESTOWNE DR		ELON	NC		27244	336 22
32	1	ALAMANCE	9079222	A	ACTIVE	AV	VERIFIED	ABBOTT	GLADYS	MARIE MILES		614 TUCKER ST	BURLINGTON	NC	27215	614 TUCKER ST		BURLINGTON	NC		27215	336 57
33	1	ALAMANCE	9129722	A	ACTIVE	AV	VERIFIED	ABBOTT	HAROLD	GRANT		507 EVERETT ST	BURLINGTON	NC	27215	507 EVERETT ST #320B		BURLINGTON	NC		27215	336 43
34	1	ALAMANCE	9094352	A	ACTIVE	AV	VERIFIED	ABBOTT	JESSICA	NADINE		2876 CALLOWAY D	MEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC		27302	919 30
35	1	ALAMANCE	9023803	A	ACTIVE	AV	VERIFIED	ABBOTT	JOYCE	HODGES		1934 TUCKER ST	BURLINGTON	NC	27215	1934 TUCKER ST #A		BURLINGTON	NC		27215	336 22
36	1	ALAMANCE	9084794	R	REMOVED	RS	MOVED FROM	ABBOTT	LATWOIA	BEREA		201 STALEY HALL	ELON	NC	27244	CAMPUS BOX 3039		ELON	NC		27244	
37	1	ALAMANCE	9020357	A	ACTIVE	AV	VERIFIED	ABBOTT	LAWRENCE	ELMER	JR	110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC		27244	336 56
38	1	ALAMANCE	9108338	A	ACTIVE	AV	VERIFIED	ABBOTT	MARIA	LYNETTE		614 TUCKER ST	BURLINGTON	NC	27215	614 TUCKER ST		BURLINGTON	NC		27215	336 57
39	1	ALAMANCE	9077192	A	ACTIVE	AV	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC		27244	800 22
40	1	ALAMANCE	9035500	A	ACTIVE	AV	VERIFIED	ABBOTT	PATTI	BELVIN		1202 JAMESTOWN	ELON	NC	27244	1202 JAMESTOWNE DR		ELON	NC		27244	336 22
41	1	ALAMANCE	9090949	R	REMOVED	RM	REMOVED FROM	ABBOTT	RACHEL	MARA		103 DANIELEY CEN	ELON	NC	27244	CAMPUS BOX 3044		ELON	NC		27244	336 27
42	1	ALAMANCE	9135295	A	ACTIVE	AV	VERIFIED	ABBOTT	SUSAN	HANKS		2876 CALLOWAY D	MEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC		27302	919 56
43	1	ALAMANCE	9113731	I	INACTIVE	IU	CONFIRMATI	ABBOTT	TAYLOR	RENEE		406 W LEBANON A	ELON	NC	27244	CAMPUS BOX 3077		ELON	NC		27244	
44	1	ALAMANCE	9120825	I	INACTIVE	IN	CONFIRMATI	ABBOTT	TIFFANY	MURIEL ARLE		144 W CRESCENT S	GRAHAM	NC	27253	144 W CRESCENT SQUARE		GRAHAM	NC		27253	336 23
45	1	ALAMANCE	9013866	I	INACTIVE	IN	CONFIRMATI	ABBOTT	VIRGINIA	SMITH		2820 BLANCHE DR	BURLINGTON	NC	27215	2820 BLANCHE DR		BURLINGTON	NC		27215	584 46
46	1	ALAMANCE	9027717	A	ACTIVE	AV	VERIFIED	ABBOTT-LUN	SHELBY	LYNN		509 FERNWAY DR	BURLINGTON	NC	27217	509 FERNWAY DR		BURLINGTON	NC		27217	336 22
47	1	ALAMANCE	9108552	A	ACTIVE	AV	VERIFIED	ABDALLA	KHALED	ISMAIL		605 ISLEY PL #C	BURLINGTON	NC	27215	605 ISLEY PL #C		BURLINGTON	NC		27215	336 68
48	1	ALAMANCE	9128403	A	ACTIVE	AV	VERIFIED	ABDEL-MAGI	LISA	ANN		1841 DUNBAR PL	BURLINGTON	NC	27215	1841 DUNBAR PL		BURLINGTON	NC		27215	214 43
49	1	ALAMANCE	9117192	I	INACTIVE	IU	CONFIRMATI	ABDELKARIM	AMNA	ELHAG		1105 PROVIDENCE	ELON	NC	27244	1105 PROVIDENCE CT		ELON	NC		27244	



Column labels



ncvoter1.txt - Microsoft Excel

Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Add-Ins

Einfügen Ausschneiden Kopieren Format übertragen Zwischenablage

Calibri 11 A A Zeilenumbruch Standard Bedingte Formatierung Als Tabelle formatieren

F K U Verbinden und zentrieren Zahl Standard Gut Neutral Schlecht Ausgabe Berechnung Eingabe Erklärender...

Schriftart Ausrichtung Formatvorlagen Zellen

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
106138	1	ALAMANCE	9129972 A	ACTIVE	AV	VERIFIED	ZLUCHOWSK	AARON	MICHAEL			3551 FORESTDALE	BURLINGTON	NC	27215	3551 FORESTDALE DR	#M	BURLINGTON	NC	27215	336
106139	1	ALAMANCE	9106623 A	ACTIVE	AV	VERIFIED	ZMIJEASKI	SEAN				4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336
106140	1	ALAMANCE	9112148 A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	DENNIS	AL			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	
106141	1	ALAMANCE	9094109 I	INACTIVE	IU	CONFIRMATI	ZMIJEWSKI	DENNIS				4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336
106142	1	ALAMANCE	9128345 A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	KEVIN	ADAM			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336
106143	1	ALAMANCE	9120294 A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	SEAN	CHRISTOPHE			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	
106144	1	ALAMANCE	9094116 A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	N VIRGINIA	LOURDES			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336
106145	1	ALAMANCE	9089250 R	REMOVED	RD	DECEASED	ZOCCOLANTI	ENIS	PIZZOTTI			2502 S NC HWY 119	MEBANE	NC	27302	2502 S NC HWY 119		MEBANE	NC	27302	
106146	1	ALAMANCE	9083629 R	REMOVED	RD	DECEASED	ZOCCOLANTI	RENATO				3141 SHELLY GRAH	GRAHAM	NC	27253	3141 SHELLY GRAHAM DR		GRAHAM	NC	27253	336
106147	1	ALAMANCE	9083630 A	ACTIVE	AV	VERIFIED	ZOCCOLANTI	RITA	MARIE			3141 SHELLY GRAH	GRAHAM	NC	27253	3141 SHELLY GRAHAM DR		GRAHAM	NC	27253	336
106148	1	ALAMANCE	9100545 I	INACTIVE	IU	CONFIRMATI	ZOGLEMANN	ANGELA	LYNNE			706 HUFFMAN MIL	BURLINGTON	NC	27215	706 HUFFMAN MILL RD	#	BURLINGTON	NC	27215	336
106149	1	ALAMANCE	9137285 A	ACTIVE	AV	VERIFIED	ZOLAYVAR	ERIC	WATSON			910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR		BURLINGTON	NC	27215	336
106150	1	ALAMANCE	9081869 A	ACTIVE	AV	VERIFIED	ZOLAYVAR	RUPERTO	BENEDICTO			910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR		BURLINGTON	NC	27215	336
106151	1	ALAMANCE	9109021 A	ACTIVE	AV	VERIFIED	ZOLAYVAR	STEPHANIE	WATSON			910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR		BURLINGTON	NC	27215	336
106152	1	ALAMANCE	9108096 A	ACTIVE	AV	VERIFIED	ZOLLARS	EVELYN	NADINE			6830 TOM WOODY	SNOW CAMP	NC	27349	6830 TOM WOODY RD		SNOW CAMF	NC	27349	336
106153	1	ALAMANCE	9125044 A	ACTIVE	AV	VERIFIED	ZOLLARS	MATHEW	DAVID			6830 TOM WOODY	SNOW CAMP	NC	27349	6830 TOM WOODY RD		SNOW CAMF	NC	27349	
106154	1	ALAMANCE	9113912 A	ACTIVE	AV	VERIFIED	ZOLLI	COFFEY	ANTONIO	MARK		108 OAKGROVE D	GRAHAM	NC	27253	108 OAKGROVE DR		GRAHAM	NC	27253	336
106155	1	ALAMANCE	9107068 A	ACTIVE	AV	VERIFIED	ZOLLI	COFFEY	VALERIE			108 OAKGROVE D	GRAHAM	NC	27253	108 OAKGROVE DR		GRAHAM	NC	27253	
106156	1	ALAMANCE	9097324 A	ACTIVE	AV	VERIFIED	ZORNES	ASHLEY	DENICE			5556 N NC HWY 49	MEBANE	NC	27302	5556 N NC HWY 49		MEBANE	NC	27302	336
106157	1	ALAMANCE	9038407 A	ACTIVE	AV	VERIFIED	ZORNES	KENNETH	ELWOOD			5556 N NC HWY 49	MEBANE	NC	27302	5556 N NC HWY 49		MEBANE	NC	27302	
106158	1	ALAMANCE	9104969 I	INACTIVE	IU	CONFIRMATI	ZORNES	MICHELLE	LEE			3117 COMMERCE	BURLINGTON	NC	27215	3117 COMMERCE PL	#L	BURLINGTON	NC	27215	336
106159	1	ALAMANCE	9018738 A	ACTIVE	AV	VERIFIED	ZORNES	SHERRIE	AVERETTE			5556 N NC HWY 49	MEBANE	NC	27302	5556 N NC HWY 49		MEBANE	NC	27302	
106160	1	ALAMANCE	9027412 I	INACTIVE	IU	CONFIRMATI	ZORNES	TERRY	LEE			148 N STATE ST	HAW RIVER	NC	27258	148 N STATE ST		HAW RIVER	NC	27258	57
106161	1	ALAMANCE	9110367 D	DENIED	DU	VERIFICATIO	ZORNES	TINA				801 TROLLINGWO	HAW RIVER	NC	27258	801 TROLLINGWOOD RD		HAW RIVER	NC	27258	336
106162	1	ALAMANCE	9132758 A	ACTIVE	AV	VERIFIED	ZORNES	TINA	MARIE			801 TROLLINGWO	HAW RIVER	NC	27258	801 TROLLINGWOOD RD		HAW RIVER	NC	27258	336
106163	1	ALAMANCE	9131499 A	ACTIVE	AV	VERIFIED	ZOUFALY	EVE				602 E HAGGARD A	ELON	NC	27244	CAMPUS BOX 8911		ELON	NC	27244	
106164	1	ALAMANCE	9124446 A	ACTIVE	AV	VERIFIED	ZSUPPAN	ETELKA	HALASZ			1929 HAW VILLAG	GRAHAM	NC	27253	1929 HAW VILLAGE DR		GRAHAM	NC	27253	
106165	1	ALAMANCE	9121554 A	ACTIVE	AV	VERIFIED	ZSUPPAN	FERENC				1929 HAW VILLAG	GRAHAM	NC	27253	1929 HAW VILLAGE DR		GRAHAM	NC	27253	
106166	1	ALAMANCE	9127457 A	ACTIVE	AV	VERIFIED	ZSUPPAN	LEVENTE	FERENC			1929 HAW VILLAG	GRAHAM	NC	27253	1929 HAW VILLAGE DR		GRAHAM	NC	27253	336
106167	1	ALAMANCE	9131401 A	ACTIVE	AV	VERIFIED	ZUBLER	LINDSAY	BROOKE			3172 CARRIAGE C	HAW RIVER	NC	27258	3172 CARRIAGE CREEK CT		HAW RIVER	NC	27258	
106168	1	ALAMANCE	9081728 A	ACTIVE	AV	VERIFIED	ZUBLER	TAMI	LAJEAN			3172 CARRIAGE C	HAW RIVER	NC	27258	3172 CARRIAGE CREEK CT		HAW RIVER	NC	27258	336
106169	1	ALAMANCE	9089569 A	ACTIVE	AV	VERIFIED	ZUBLER	TIMOTHY	JAMES			3172 CARRIAGE C	HAW RIVER	NC	27258	3172 CARRIAGE CREEK CT		HAW RIVER	NC	27258	
106170	1	ALAMANCE	9070674 A	ACTIVE	AV	VERIFIED	ZUBOV	ALEX				229 ENGLEMAN A	BURLINGTON	NC	27215	229 ENGLEMAN AVE		BURLINGTON	NC	27215	336
106171	1	ALAMANCE	9070288 A	ACTIVE	AV	VERIFIED	ZUBOV	LYNN	R			229 ENGLEMAN A	BURLINGTON	NC	27215	229 ENGLEMAN AVE		BURLINGTON	NC	27215	336
106172	1	ALAMANCE	9008787 A	ACTIVE	AV	VERIFIED	ZUMER	FRANK	EDWARD			801 QUAKER RIDG	MEBANE	NC	27302	801 QUAKER RIDGE RD		MEBANE	NC	27302	915
106173	1	ALAMANCE	9008785 A	ACTIVE	AV	VERIFIED	ZUMER	LOUISE	TURNER			801 QUAKER RIDG	MEBANE	NC	27302	801 QUAKER RIDGE RD		MEBANE	NC	27302	915
106174	1	ALAMANCE	9141817 A	ACTIVE	AV	VERIFIED	ZUNG	PATRICK	BATE			2604 WOODS LN	GRAHAM	NC	27253	2604 WOODS LN		GRAHAM	NC	27253	915
106175	1	ALAMANCE	9119438 A	ACTIVE	AV	VERIFIED	ZUNIGA	JOSE	RAMON SAL			714 ROSS ST	BURLINGTON	NC	27217	714 ROSS ST		BURLINGTON	NC	27217	336
106176	1	ALAMANCE	9108610 A	ACTIVE	AV	VERIFIED	ZUNIGA	VANESA	ELIZABETH			512 PIEDMONT W	BURLINGTON	NC	27217	512 PIEDMONT WAY		BURLINGTON	NC	27217	336
106177	1	ALAMANCE	9112637 A	ACTIVE	AV	VERIFIED	ZUNIGA	YANET	SALAS			3845 MAE DOUGL	MEBANE	NC	27302	3845 MAE DOUGLAS DR		MEBANE	NC	27302	
106178	1	ALAMANCE	9141392 A	ACTIVE	AV	VERIFIED	ZUPANCICH	MONICA	ANITA			2326 N NC HWY 49	BURLINGTON	NC	27217	2326 N NC HWY 49		BURLINGTON	NC	27217	336
106179	1	ALAMANCE	9141400 A	ACTIVE	AV	VERIFIED	ZUPANCICH	RONALD	JAMES	II		2326 N NC HWY 49	BURLINGTON	NC	27217	2326 N NC HWY 49		BURLINGTON	NC	27217	757
106180	1	ALAMANCE	9070674 A	ACTIVE	AV	VERIFIED	ZURFACE	ROSSELL	EUGENE			2074 TURNER RD	MEBANE	NC	27302	2074 TURNER RD		MEBANE	NC	27302	
106181	1	ALAMANCE	9070674 A	ACTIVE	AV	VERIFIED	ZWIER	ANDREW	MICHAEL			1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMF	NC	27349	336
106182	1	ALAMANCE	9070674 A	ACTIVE	AV	VERIFIED	ZWIER	CHRISTOPHE	ANTHONY			1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMF	NC	27349	831
106183	1	ALAMANCE	9070674 A	ACTIVE	AV	VERIFIED	ZWIER	CHRISTY	ANN			1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMF	NC	27349	
106184	1	ALAMANCE	9099261 A	ACTIVE	AV	VERIFIED	ZWIER	KAREN	JEAN			1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMF	NC	27349	831
106185	1	ALAMANCE	9077804 R	REMOVED	RL	MOVED FROI	ZYLKA	MARC				1210 WILLOW BRC	MEBANE	NC	27302	1210 WILLOW BROOK CT		MEBANE	NC	27302	336

Number of rows







V1	race_code	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	voter_status	last_name	first_name	midl_name	name	res_street_addr	res_city_desc	state	zip_code	mail_addr1	mail_addr2	mail_city	mail_state	mail_zipcode	full_phone	race_code	ethnic_code	party_cd	gender_code	birth_age	bi
2	VERIFIED	AABEL	EVELYN	LARSEN	4430 E GREENSBORO	GRAHAM	NC	27253	4430 E GREENSBORO-CHA	GRAHAM	NC						NL	UNA	F	77	N
3	VERIFIED	AARON	CHRISTINA	CASTAGNA	421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177	BURLINGTON	NC						UN	UNA	F	36	N
4	VERIFIED	AARON	CLAUDIA	HAYDEN	1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST	BURLINGTON	NC						NL	UNA	F	68	V
5	VERIFIED	AARON	JAMES	MICHAEL	1647 SAXAPAHAW	GRAHAM	NC	27253	PO BOX 98	SAXAPAHAW	NC						UN	DEM	M	65	M
6	VERIFIED	AARON	NATHAN	EDWARD	421 WHITT AVE	BURLINGTON	NC										UN	UNA	M	36	N
7	VERIFIED	AARON	WILLIE	DALE	1013 EDITH ST	BURLINGTON	NC										NL	UNA	M	68	V
8	VERIFIED	AARONSON	GENA	HOLT	107 TERRYWOOD	HAW RIVER	NC										NL	REP	F	41	N
9	VERIFIED	AARONSON	MICHAEL	CHARLES	107 TERRYWOOD	HAW RIVER	NC										NL	UNA	M	50	W
12	VERIFIED	ABADIE	JACK	EDWARD JR	612 SIDEVIEW ST	GRAHAM	NC										NL	UNA	M	27	N
13	CONFIRMATI	ABADIE	MYRA	HOLLIFIELD	612 SIDEVIEW ST	GRAHAM	NC										NL	UNA	F	61	N
16	VERIFIED	ABBATECOLA	RONALD	JOSEPH JR	504 BROOKFIELD	GIBSONVILLE	NC										UN	UNA	M	37	N
17	VERIFIED	ABBATECOLA	TRACY	BOONE	504 BROOKFIELD	GIBSONVILLE	NC										NL	DEM	F	45	N
18	CONFIRMATI	ABBETT	DAWN	LEANN	3900 JOHNS CREEK	GIBSONVILLE	NC										NL	DEM	F	49	C
19	VERIFIED	ABBEY	BRENT	DAVID	3304 GOLDEN OAK	GRAHAM	NC										NL	REP	M	45	N
20	VERIFIED	ABBEY	DEMETRA	AINSWORTH	3304 GOLDEN OAK	GRAHAM	NC										NL	REP	F	44	S
21	CONFIRMATI	ABBEY	DOROTHY	ESTELLA	1029A QUAKENBUSH	SNOW CAMP	NC										NL	REP	F	91	C
22	VERIFIED	ABBOTT	AMELIA	BETH	2876 CALLOWAY	MEBANE	NC										NL	UNA	F	23	N
23	VERIFIED	ABBOTT	ANGELA	MORTON	2006 WINN CREEK	HAW RIVER	NC										NL	DEM	F	39	N
24	VERIFIED	ABBOTT	BRENDA	CARMICHAEL	611 N THIRD ST	MEBANE	NC										NL	UNA	F	58	N
25	VERIFIED	ABBOTT	BRIAN	CHRISTOPHE	2006 WINN CREEK	HAW RIVER	NC										NL	UNA	M	40	N
26	VERIFIED	ABBOTT	BRUCE	CLEATON	188 LAKE CAMMA	BURLINGTON	NC										NL	REP	M	63	N
27	VERIFIED	ABBOTT	CHERYL	FAULKNER	188 LAKE CAMMA	BURLINGTON	NC										NL	REP	F	59	N
28	VERIFIED	ABBOTT	CHRISTOPHE	BRANDON	309 BURLINGTON	GIBSONVILLE	NC										NL	UNA	M	38	N
29	VERIFIED	ABBOTT	COURTNEY	LOVE	309 BURLINGTON	GIBSONVILLE	NC										NL	UNA	F	43	N
30	VERIFIED	ABBOTT	DWAYNE	ROGER	2839 LADALE LN	MEBANE	NC										NL	UNA	M	53	N
31	VERIFIED	ABBOTT	FRANK	PATRICK	1202 JAMESTOWN	ELON	NC										UN	UNA	M	46	N
33	VERIFIED	ABBOTT	HAROLD	GRANT	507 EVERETT ST	BURLINGTON	NC										NL	REP	M	69	N
34	VERIFIED	ABBOTT	JESSICA	NADINE	2876 CALLOWAY	MEBANE	NC										NL	UNA	F	29	N
35	VERIFIED	ABBOTT	JOYCE	HODGES	1934 TUCKER ST	BURLINGTON	NC										NL	DEM	F	66	V
37	VERIFIED	ABBOTT	LAWRENCE	ELMER JR	110 OAKVIEW DR	ELON	NC										NL	UNA	M	62	N
39	VERIFIED	ABBOTT	NANCY	SKIDMORE	110 OAKVIEW DR	ELON	NC										NL	UNA	F	69	W
40	VERIFIED	ABBOTT	PATTI	BELVIN	1202 JAMESTOWN	ELON	NC	27244	1202 JAMESTOWN DR	ELON	NC						UN	REP	F	47	N
41	REMOVED AT	ABBOTT	RACHEL	MARA	103 DANIELEY	CENELON	NC	27244	CAMPUS BOX 3044	ELON	NC						NL	REP	F	28	P
42	VERIFIED	ABBOTT	SUSAN	HANKS	2876 CALLOWAY	MEBANE	NC	27302	2876 CALLOWAY DR	MEBANE	NC						UN	UNA	F	54	N
43	CONFIRMATI	ABBOTT	TRACY	BOONE	CAMPUS BOX 3077	ELON	NC	27244	CAMPUS BOX 3077	ELON	NC						UN	REP	F	25	W
45	CONFIRMATI	58	VERIFIED	ABELE	ELISA	MARY	NC	27215	820 BLANCHE DR	BURLINGTON	NC						NL	REP	F	85	P
47	VERIFIED	59	VERIFIED	ABELE	KLAY	MARY	NC	27215	05 ISLEY PL #C	BURLINGTON	NC						NL	DEM	U	41	N
48	VERIFIED	60	VERIFIED	ABELE	KYLIE	JO	NC	27215	841 DUNBAR PL	BURLINGTON	NC						NL	UNA	F	52	D
52	VERIFIED	61	VERIFIED	ABELE	MARY	MA	NC	27253	10 BURTON ST	GRAHAM	NC						NL	UNA	M	56	M
53	VERIFIED	60	VERIFIED	ABELE	KYLIE	JO	NC	27215	966 S CHURCH APT 110	BURLINGTON	NC						NL	UNA	F	53	C
54	DI	61	VERIFIED	ABELE	MARY	MA	NC	27215	922 MUHLENBERG CT	BURLINGTON	NC						NL	DEM	M	101	I
55	DI	61	VERIFIED	ABELE	MARY	MA	NC	27215	905 CALVIN CT	BURLINGTON	NC						NL	DEM	F	87	N
56	DI	61	VERIFIED	ABELE	MARY	MA	NC	27215	320 HAWTHORNE LN	BURLINGTON	NC						NL	REP	M	23	U
57	VI	61	VERIFIED	ABELE	MARY	MA	NC	27215	320 HAWTHORNE LN	BURLINGTON	NC						NL	REP	M	64	D
58	VI	61	VERIFIED	ABELE	MARY	MA	NC	27215	320 HAWTHORNE LN	BURLINGTON	NC						NL	UNA	F	21	U
59	VI	61	VERIFIED	ABELE	MARY	MA	NC	27249	36 SLATE DR	GIBSONVILLE	NC						NL	UNA	M	26	O
60	VI	61	VERIFIED	ABELE	MARY	MA	NC	27249	36 SLATE DR	GIBSONVILLE	NC						NL	REP	F	25	W
61	VI	61	VERIFIED	ABELE	MARY	MA	NC	27215	320 HAWTHORNE LN	BURLINGTON	NC						UN	UNA	F	54	D

Von A bis Z sortieren

Von Z bis A sortieren

Nach Farbe sortieren

Filter löschen aus "race\_code"

Nach Farbe filtern

Textfilter

Suchen

(Alles auswählen)

A

B

I

M

O

U

W

OK Abbrechen

Von A bis Z sortieren

Von Z bis A sortieren

Nach Farbe sortieren

Filter löschen aus "race\_code"

Nach Farbe filtern

Textfilter

Suchen

(Alles auswählen)

A

B

I

M

O

U

W

OK Abbrechen

27249		W
27302	563 3956	W
27244	336 227 4088	W
27215	336 437 3638	W
27302	919 304 4661	W
27215	336 227 4079	W
27244	336 563 4708	W
27244	800 222 7566	W
27244	336 228 0571	W
27244	336 278 4012	W
27302	919 568 8056	W
27244		W
27215	584 4663	W
27215	336 686 0506	W
27215	214 437 8955	W
27253	336 227 6083	W
27215	5108 336 263 6215	W
27215	336 585 1926	W
27215		W
27215	336 459 8689	W
27215		W
27215	434 334 2806	W
27249	336 603 8183	W
27249	412 389 7039	W
27215	434 250 5217	W

ncvoter1.txt - Microsoft Excel

Datei Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Add-Ins

Einfügen Ausschneiden Kopieren Format übertragen

Calibri 11 A A

Zeilenumbruch Verbinden und zentrieren

Standard Bedingte Formatierung Als Tabelle formatieren

Standard Gut Neutral Schlecht

Ausgabe Berechnung Eingabe Erklärender...

Einfügen Löschen Format

Zellen

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
105859	VERIFIED	YOUNGER	LISA	CAROL		1327 STONE ST EX	MEBANE	NC	27302	1327 STONE ST EXT #A	MEBANE	NC		27302	336 578 4185	B	NL	DEM	F	52
105911	VERIFIED	YUEN	DIANE			1208 RASPBERRY F	GRAHAM	NC	27253	1208 RASPBERRY RUN	GRAHAM	NC		27253		A	UN	UNA	F	63
105912	OVERSEAS C	YUEN	DIANE			552 MEADOWWOOD	BURLINGTON	NC	27215	552 MEADOWWOOD DR	BURLINGTON	NC		27215		U	UN	DEM	F	24
105916	VERIFIED	YUQUILIMA	AIDA	GUILLERMIN		1204 GRACE AVE	BURLINGTON	NC	27217	1204 GRACE AVE	BURLINGTON	NC		27217	336 675 4492	U	UN	DEM	U	57
105953	VERIFIED	ZAGORSKI	NICOLE	LYNNE		100 PHOENIX DR	ELON	NC	27244	CAMPUS BOX 6032	ELON	NC		27244	973 349 9213	U	UN	REP	U	20
105957	VERIFIED	ZAJAC	LYNNEE	LYN		1209 BIRKDALE C	MEBANE	NC	27302	1209 BIRKDALE CIR	MEBANE	NC		27302	412 414 2947	B	NL	DEM	F	26
105959	CONFIRMATI	ZALDIVAR	NICK			3475 S NC HWY 11	HAW RIVER	NC	27258	3475 S NC HWY 119	HAW RIVER	NC		27258	919 741 3647	O	NL	UNA	M	21
105968	VERIFIED	ZAMARRIPA	JUANA	MARIA		3668 ALTAMAHAW	ELON	NC	27244	PO BOX 791	HAW RIVER	NC		27258	336 524 9599	O	HL	DEM	F	89
105971	VERIFIED	ZAMORA																		41
105972	VERIFIED	ZAMORA																		19
105973	VERIFIED	ZAMORA																		23
105974	VERIFIED	ZAMORA																		43
105977	VERIFIED	ZAMPARE																		35
105978	VERIFIED	ZAMUDIC																		47
106000	VERIFIED	ZAPATA																		57
106001	VERIFIED	ZAPATA																		27
106002	VERIFIED	ZAPIEN																		20
106017	CONFIRMATI	ZAVALA																		24
106018	VERIFIED	ZAVALA																		66
106019	VERIFIED	ZAVALA																		23
106023	CONFIRMATI	ZAYAS																		58
106024	UNAVAILABL	ZAYAS																		43
106027	VERIFIED	ZAYAS																		59
106034	VERIFIED	ZEIGLER																		50
106036	VERIFIED	ZEIGLER																		41
106037	VERIFIED	ZEIGLER																		64
106038	VERIFIED	ZELAYA																		29
106045	VERIFIED	ZEMPEL																		32
106057	REMOVED A	ZHAGUI																		39
106058	VERIFIED	ZHENG	JICHUN			1914 HILLDALE DR	BURLINGTON	NC	27215	1914 HILLDALE DR	BURLINGTON	NC		27215	917 977 1073	U	UN	UNA	U	25
106059	CONFIRMATI	ZHENG	KELLY			95 DENISE DR #B	BURLINGTON	NC	27215	95 DENISE DR #B	BURLINGTON	NC		27215		A	UN	UNA	F	26
106060	VERIFIED	ZHENG	SAIFEN			515 BURLINGTON	GIBSONVILLE	NC	27249	515 BURLINGTON AVE	GIBSONVILLE	NC		27249		A	UN	UNA	F	33
106066	VERIFIED	ZIEGLER	BRANDON	GERALD		2459 MORNINGSID	BURLINGTON	NC	27217	2459 MORNINGSIDE DR	BURLINGTON	NC		27217	336 212 7747	B	NL	REP	M	22
106072	VERIFIED	ZIEGLER	VIENGSAMA	MINDY		2459 MORNINGSID	BURLINGTON	NC	27217	2459 MORNINGSIDE DR	BURLINGTON	NC		27217		U	UN	UNA	F	24
106079	VERIFIED	ZIENTY	SCOTT	CARL		511 W HAGGARD A	ELON	NC	27244	511 W HAGGARD AVE	ELON	NC		27244		O	NL	UNA	M	23
106100	VERIFIED	ZIMMERMAN	KATRINA	MCINTOSH		852 S MAIN ST #1	BURLINGTON	NC	27215	852 S MAIN ST #16	BURLINGTON	NC		27215		M	NL	UNA	F	25
106132	VERIFIED	ZINTZUN	ABUNDIO	BAUTISTA	SR	209 SPRINGWOOD	BURLINGTON	NC	27215	209 SPRINGWOOD DR	BURLINGTON	NC		27215	336 437 1081	O	HL	UNA	M	55
106136	VERIFIED	ZITO	ALEX	R		622 S WILLIAMSON	ELON	NC	27244	622 S WILLIAMSON AVE	ELON	NC		27244	617 276 2096	U	UN	UNA	U	22
106139	VERIFIED	ZMIJEASKI	SEAN			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD	MEBANE	NC		27302	336 376 1987	O	UN	REP	M	23
106144	VERIFIED	ZMIJEWSKI	VIRGINIA	LOURDES		4872 THOM RD	MEBANE	NC	27302	4872 THOM RD	MEBANE	NC		27302	336 376 1987	U	UN	UNA	F	45
106149	VERIFIED	ZOLAYVAR	ERIC	WATSON		910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR	BURLINGTON	NC		27215	336 585 0248	O	NL	DEM	M	20
106150	VERIFIED	ZOLAYVAR	RUPERTO	BENEDICTO		910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR	BURLINGTON	NC		27215	336 585 0248	O	NL	DEM	M	73
106154	VERIFIED	ZOLLICOFFEF	ANTONIO	MARK		108 OAKGROVE D	GRAHAM	NC	27253	108 OAKGROVE DR	GRAHAM	NC		27253	336 260 6673	B	UN	DEM	M	53
106155	VERIFIED	ZOLLICOFFEF	VALERIE			108 OAKGROVE D	GRAHAM	NC	27253	108 OAKGROVE DR	GRAHAM	NC		27253		B	UN	DEM	F	47
106163	VERIFIED	ZOUFALY	EVE			602 E HAGGARD A	ELON	NC	27244	CAMPUS BOX 8911	ELON	NC		27244		U	UN	UNA	U	21
106167	VERIFIED	ZUBLER	LINDSAY	BROOKE		3172 CARRIAGE C	HAW RIVER	NC	27258	3172 CARRIAGE CREEK CT	HAW RIVER	NC		27258		U	UN	UNA	F	19
106175	VERIFIED	ZUNIGA	JOSE	RAMON SAL		714 ROSS ST	BURLINGTON	NC	27217	714 ROSS ST	BURLINGTON	NC		27217	336 227 3108	O	HL	DEM	M	49
106177	VERIFIED	ZUNIGA	YANET	SALAS		3845 MAE DOUGL	MEBANE	NC	27302	3845 MAE DOUGLAS DR	MEBANE	NC		27302		O	HL	DEM	F	34
106186																				

Google

[Web](#) [Bilder](#) [Maps](#) [Shopping](#) [Apps](#) [Mehr](#) [Suchoptionen](#)

Ungefähr 1.260.000 Ergebnisse (0,23 Sekunden)

Meinten Sie: [Excel histogramm](#)

[Create a histogram - Excel - Office.com](#)  
[office.microsoft.com/.../excel.../create-a-histogram-...](#) [Diese Seite übersetzen](#)







## Many interesting questions remain

■ What are possible keys and foreign keys?

□ Phone

□ firstname, lastname, street

■ Are there any functional dependencies?

□ zip -> city

□ race -> voting behavior

■ Which columns correlate?

□ county and first name

□ DoB and last name

■ What are frequent patterns in a column?

□ ddddd

□ dd aaaa St



# Definition Data Profiling

11

- Data profiling is the process of examining the data available in an existing data source [...] and collecting statistics and information about that data.

Wikipedia 09/2013

- Data profiling refers to the activity of creating small but informative summaries of a database.

Ted Johnson, Encyclopedia of Database Systems

- A fixed set of data profiling tasks / results

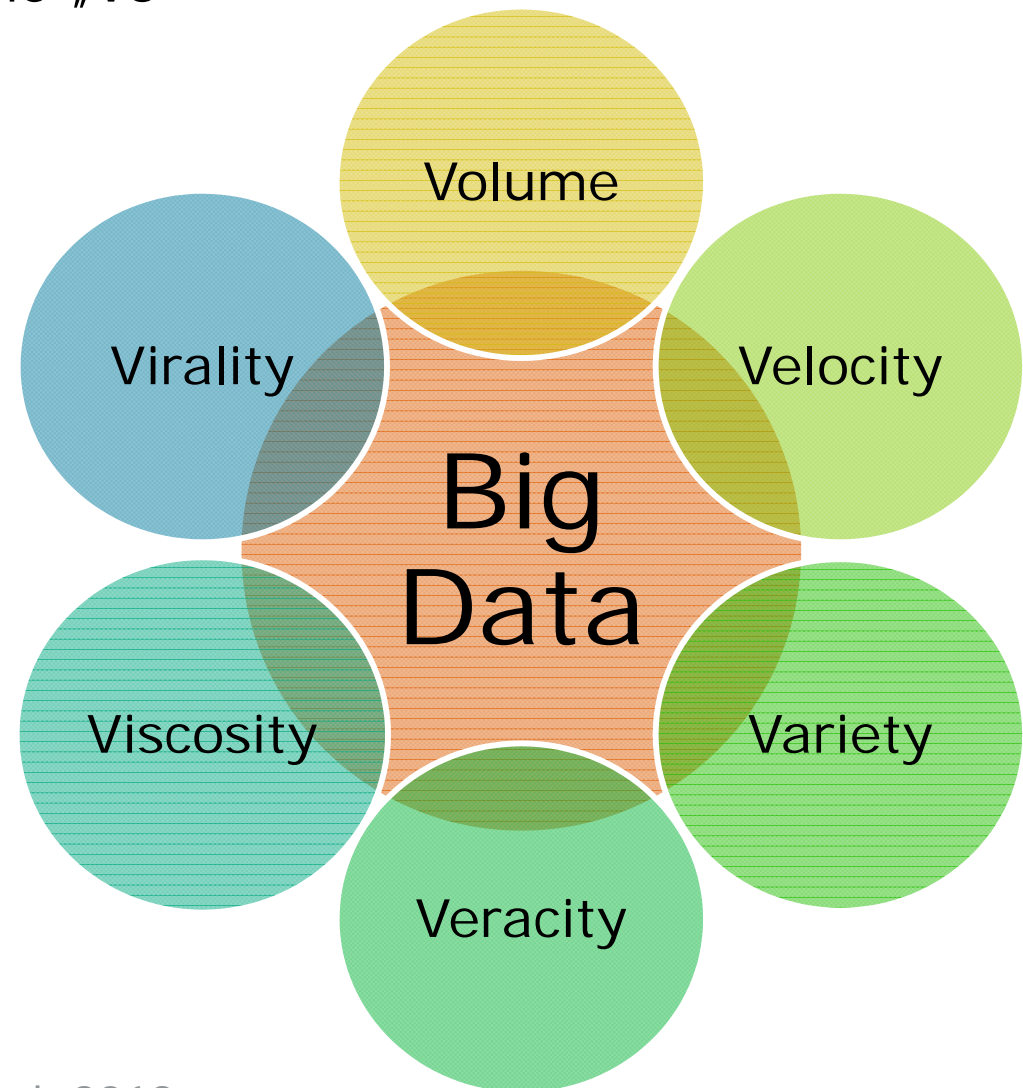


# „Big“ Data Profiling or How big is „Big“?

12

Data profiling = measuring the „Vs“

- Volume
  - Row counts, etc.
- Velocity
  - Temporal profiling
- Variability
  - How difficult to integrate and analyse
- Veracity
  - How good is it?
- ...





# Challenges of (Big) Data Profiling

13

- Computational complexity
  - Number of rows
  - Number of columns (and column combinations)
- Large solution space
  
- New data types (beyond strings and numbers)
- New data models (beyond relational): RDF, XML, etc.
- New requirements
  - User-oriented
  - Interactive
  - Streaming data



# Use Cases for Profiling

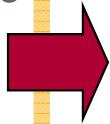
14

- Query optimization
  - Counts and histograms
- Data cleansing
  - Patterns, rules, and violations
- Data integration
  - Cross-DB inclusion dependencies
- Scientific data management
  - Handle new datasets
- Data inspection, analytics, and mining
  - Profiling as preparation to decide on models and questions
- Database reverse engineering
  
- Data profiling as preparation for any other data management task



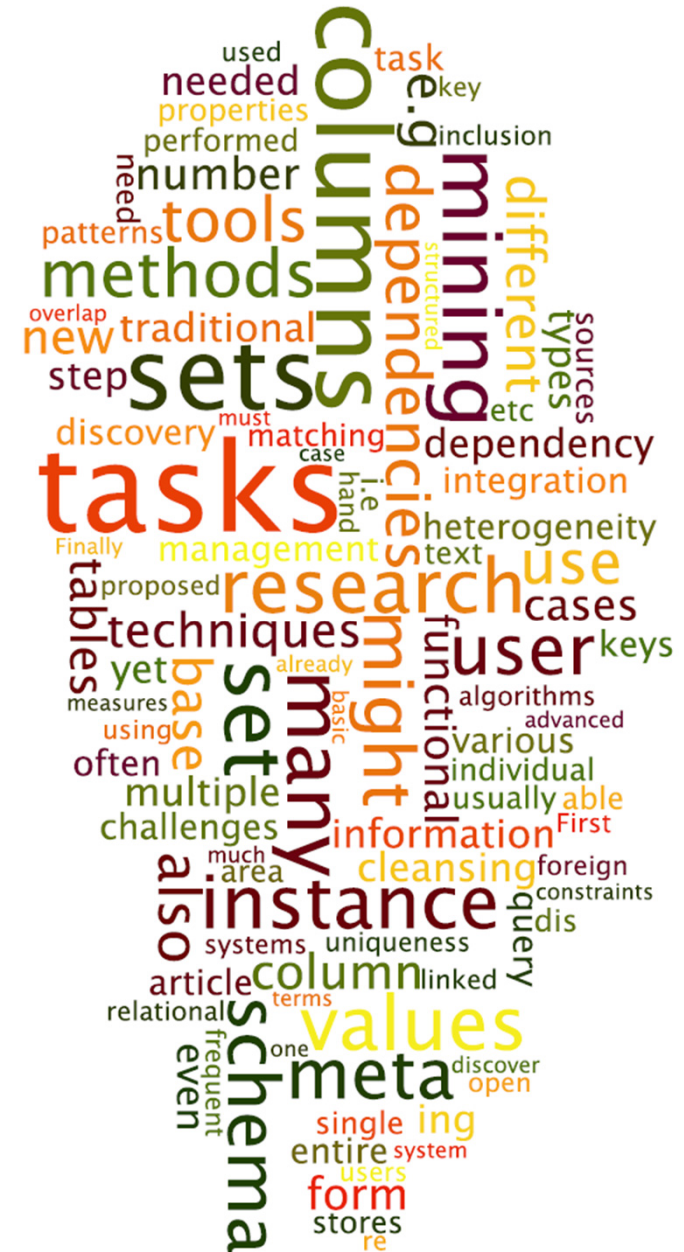
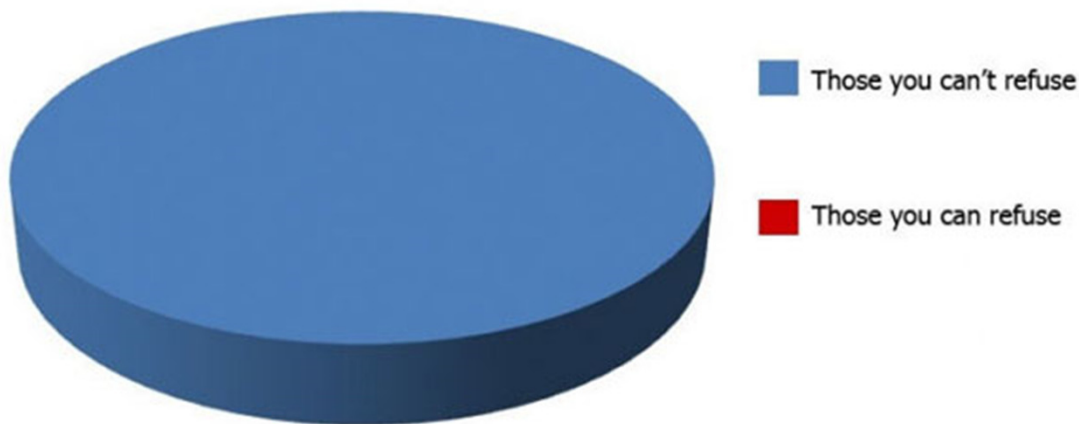
# Agenda

15



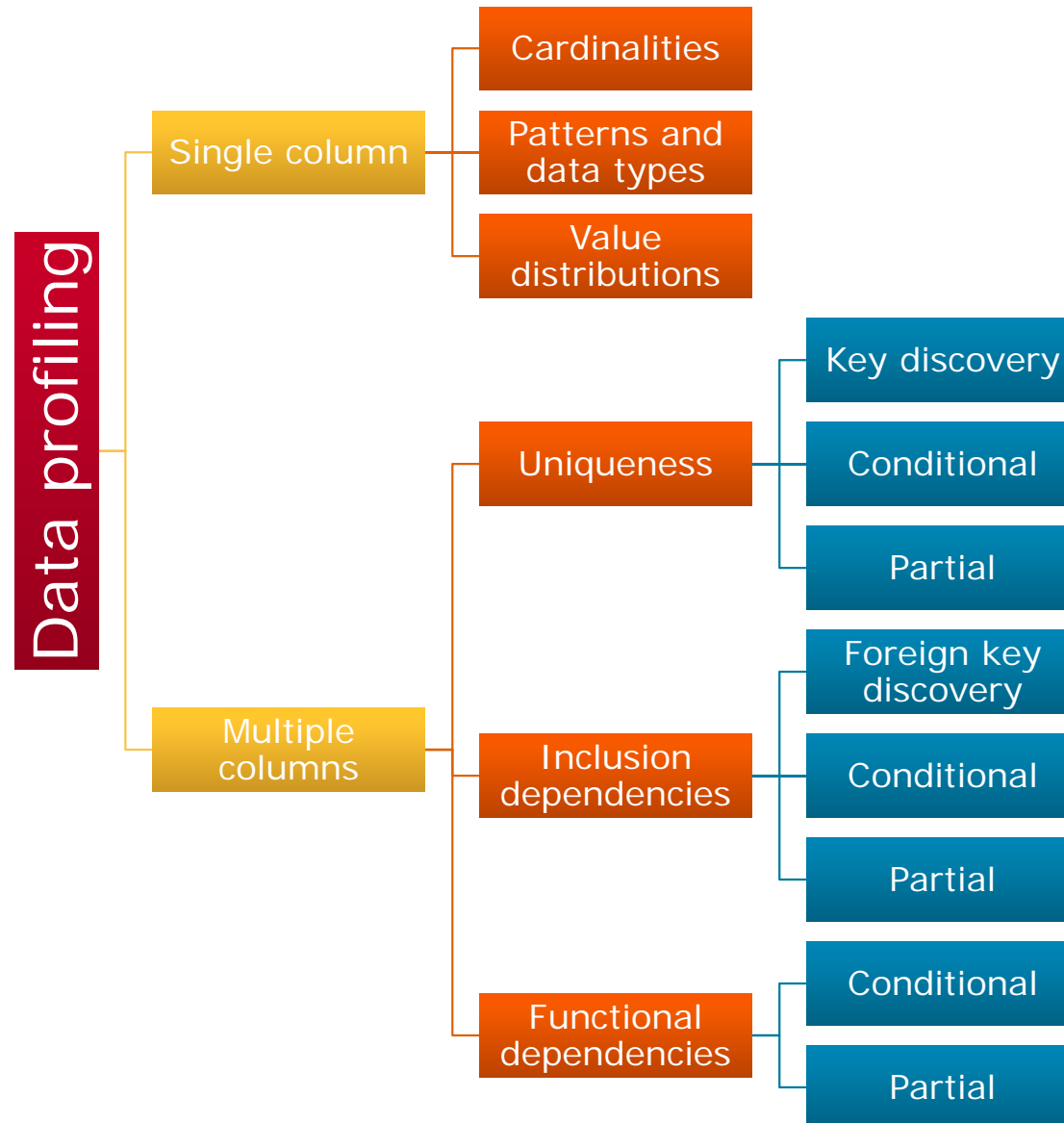
- Data profiling tasks
- Unique discovery
- Data profiling tools
- Advanced profiling
- Profiling heterogeneous data

## Offers made by Don Corleone



# Classification of Traditional Profiling Tasks

16





# Single-column vs. multi-column

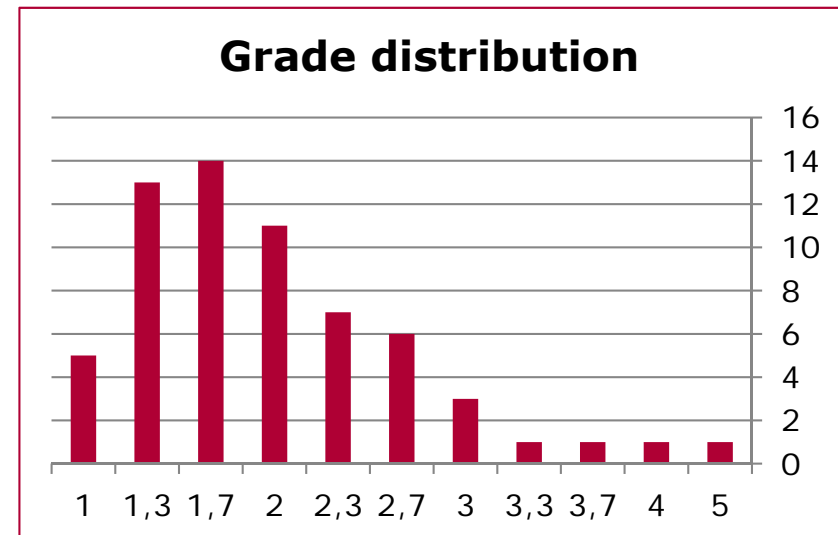
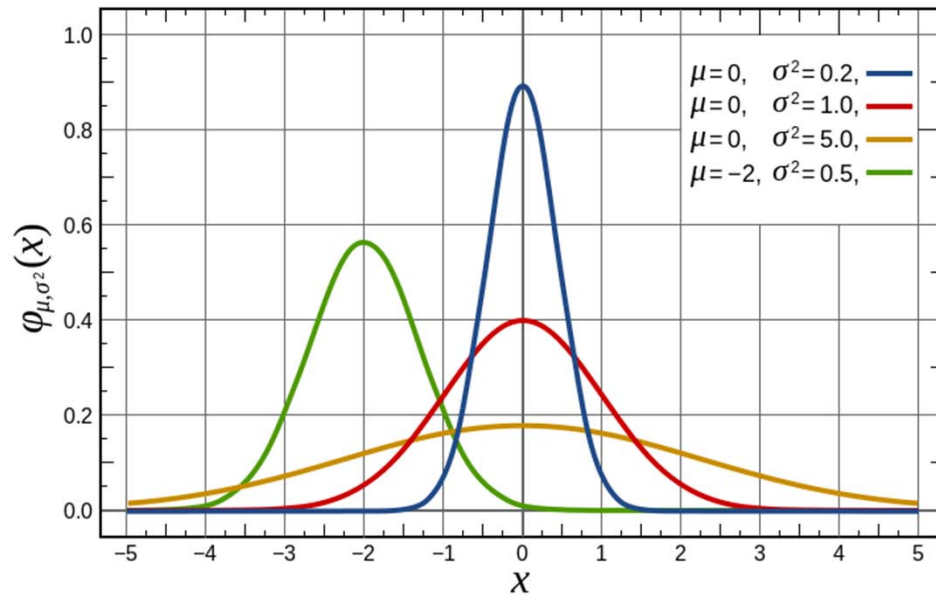
17

- Single column profiling
  - Most basic form of data profiling
  - Often part of the basic statistics gathered by DBMS
  - Discovery complexity: Number of values/rows
  
- Multicolumn profiling
  - Discover joint properties
  - Discover dependencies
  - Discovery complexity: Number of columns and number of values

# Cardinalities and distributions

18

- Number of values / distinct values
- Number of NULLs
- MIN and MAX value
- Distributions and histograms





# Data types and value patterns

19

- String vs. number
- String vs. number vs. date
- Categorical vs. continuous
- SQL data types
  - CHAR, INT, DECIMAL, TIMESTAMP, BIT, CLOB, ...
- Domains
  - VARCHAR(12) vs. VARCHAR (13)
- XML data types
  - More fine grained
- Regular expressions `(\d{3})-(\d{3})-(\d{4})-(\d+)`
- Semantic domains
  - Adress, phone, email, first name

# Uniqueness, keys, and foreign keys

20

- Uniqueness and keys
  - Unique column: Only unique values
  - Unique column combination: Only unique value combinations
    - ◇ Minimality: No subset is unique
  - Key candidate: No null values
    - ◇ Uniqueness and non-null in one instance does not imply key: Only human can specify keys (and foreign keys)
  
- Inclusion dependencies and foreign keys
  - $A \subseteq B$ : All values in A are also present in B
  - $A_1, \dots, A_i \subseteq B_1, \dots, B_i$ : All value comb. in  $A_1, \dots, A_i$  are also present in  $B_1, \dots, B_i$
  - Prerequisite for foreign key
    - ◇ Across relations and across databases
    - ◇ Again: Discovery on a given instance, only user can specify for schema



# FDs and Partial Dependencies

21

- Functional dependency „ $X \rightarrow A$ “: whenever two records have the same  $X$  values, they also have the same  $A$  values.
  
- Partial dependency: INDs and FDs that do not perfectly hold
  - For all but 10 of the tuples
  - Only for 90% of the tuples
  - Only for 1% of the tuples
  
- Partiality also for patterns, types, uniques, and other constraints

# Conditional dependencies

22

- Given a partial IND or FD: For **which** part does it hold?
- Expressed as a condition over the attributes of the relation
- Problems:
  - Infinite possibilities of conditions
  - Interestingness:
    - ◇ Many distinct values: less interesting
    - ◇ Few distinct values: surprising condition – high coverage
- Useful for
  - Integration: cross-source cINDs





# Uniqueness and keys

24

- Unique column
  - Only unique values
- Unique column combination
  - Only unique value combinations
  - Minimality: No subset is unique

A	B	C
a	1	x
b	2	x
c	2	y

- Uniques: {A, AB, AC, BC, ABC}
- Minimal uniques: {A, BC}
- (Maximal) Non-uniques: {B, C}

# Null values

25

- Null values have a wide range of interpretations.
  - Unknown (date of birth)
  - Non-applicable (driver license number for kids)
  - Undefined (result of integration/outer join)
- What are minimal uniques for the following data set?

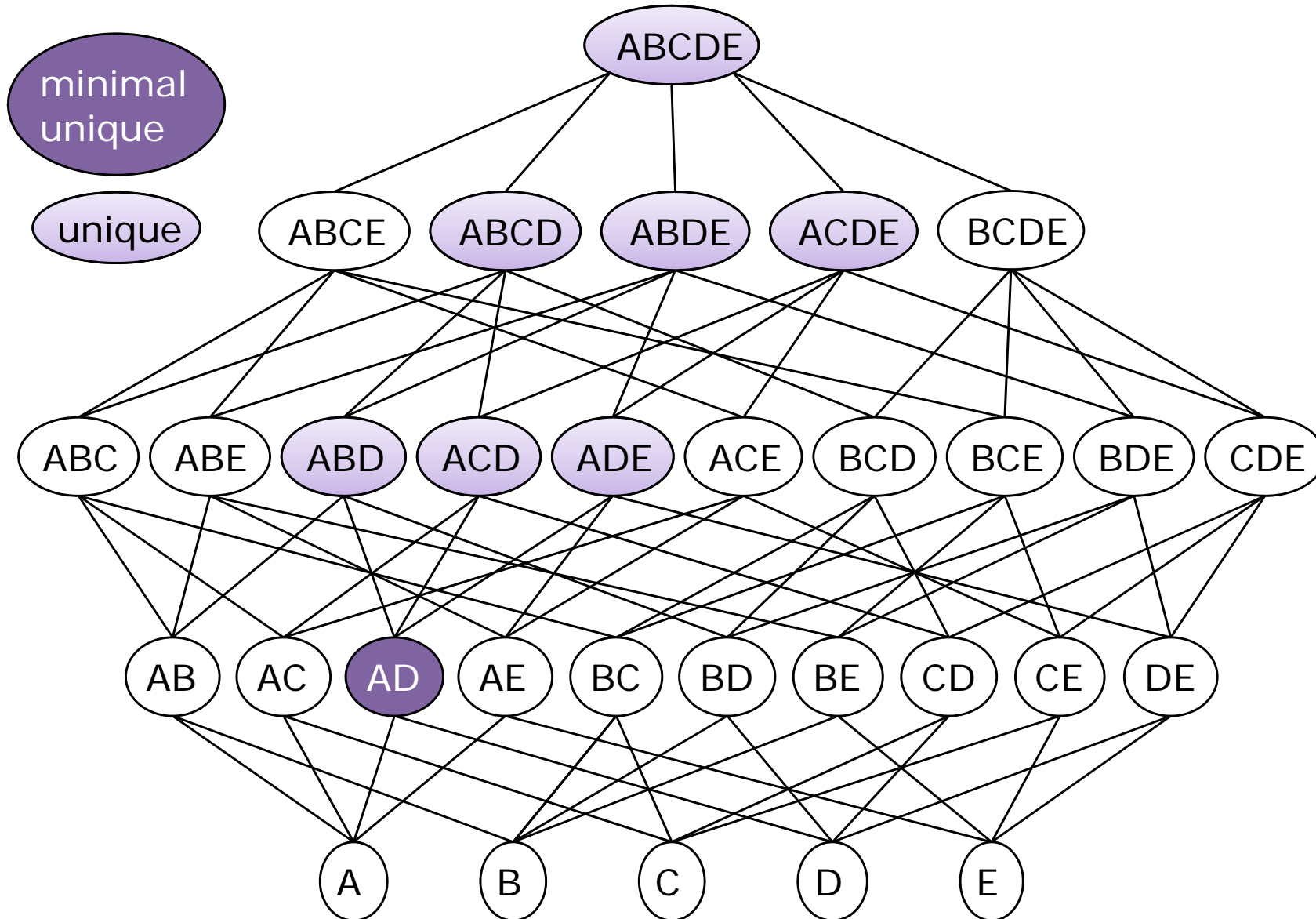
A	B	C	D
a	1	x	1
b	2	y	2
c	3	z	5
d	3	⊥	5
e	⊥	⊥	5

- Primary key {A}; Some unusual uniques: {C} and {CD}
- Distinct: {A, BC} but not {CD}



# Pruning effect of a pair

26



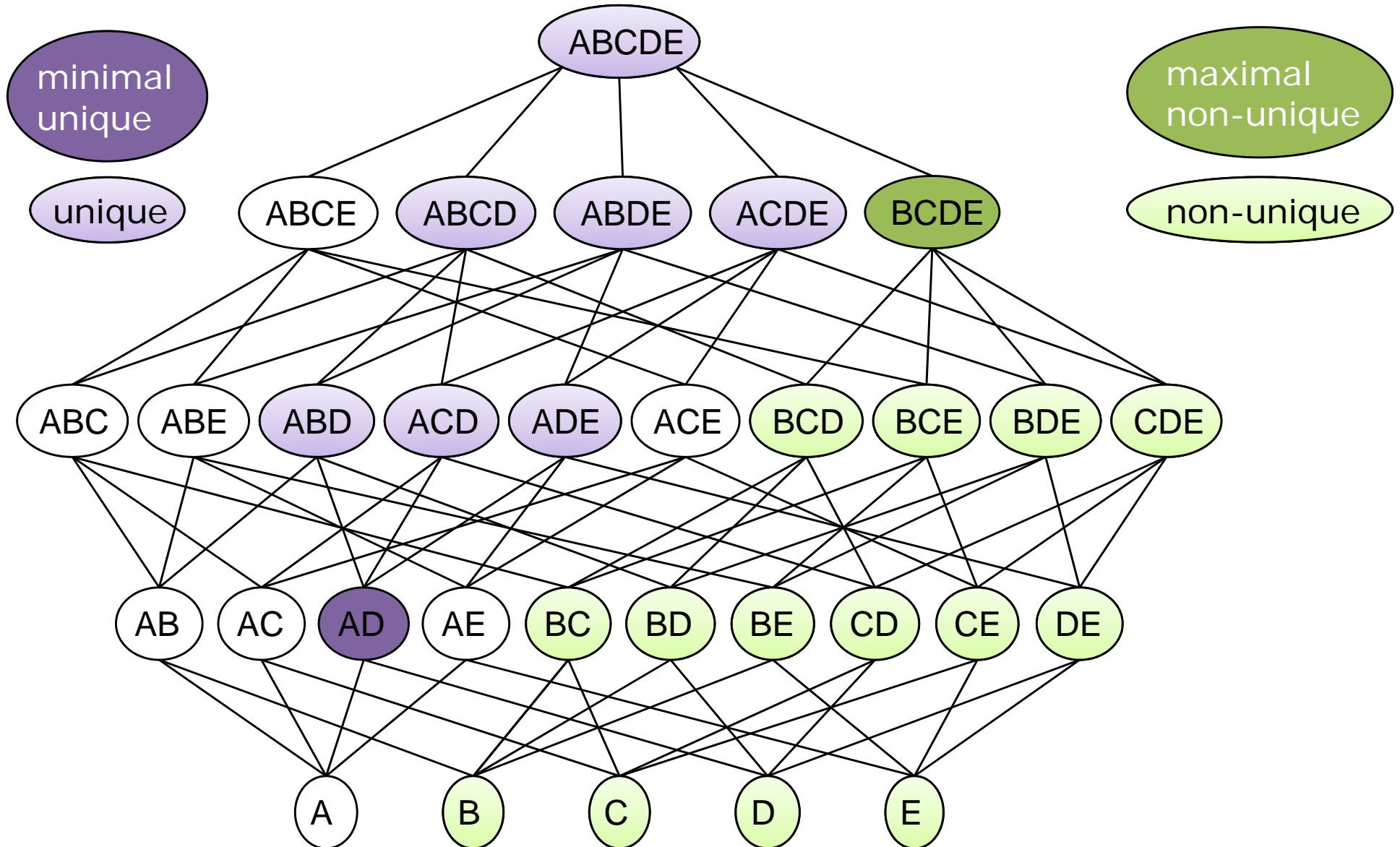
# Pruning with uniques

27

- Pruning: inferring the type of a combination without actual verification
  
- If A is unique, supersets must be unique
- Finding a unique column prunes half of the lattice
  - Remove column from initial data set and restart
- Finding a unique column pair removes a quarter of the lattice
  - In general, the lattice over the combination is removed
  
- The pruning power of a combination is reduced by prior findings
  - AB prunes a quarter
  - BC additionally prunes only one eighth
  - ABC already pruned one eights

# Pruning both ways

28

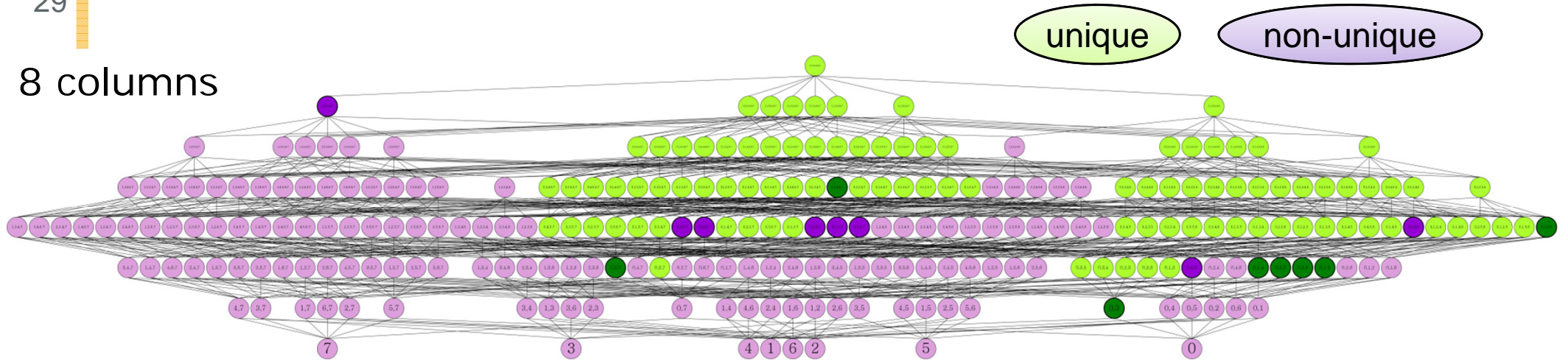




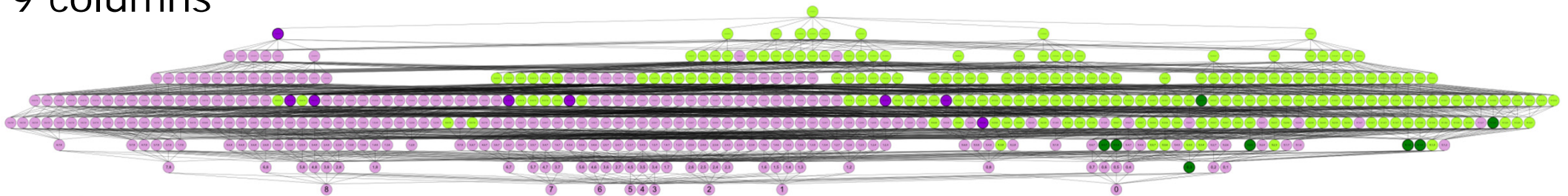
# TPCH – Uniques and Non-Uniques

29

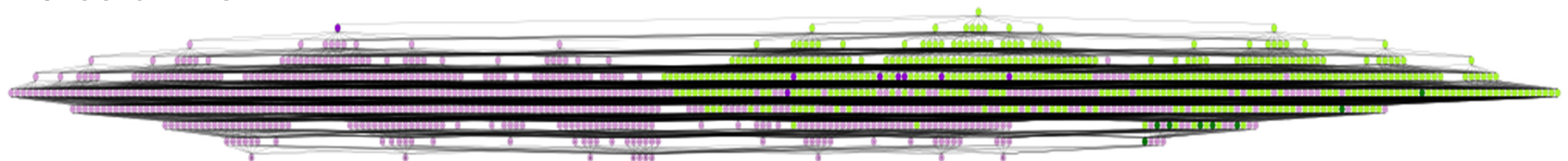
8 columns



9 columns



10 columns



# Scalable profiling

30

- Scalability in number of rows
- Scalability in number of columns
  - “Small” table with 100 columns:  
 $2^{100} - 1 = 1,267,650,600,228,229,401,496,703,205,375$   
= 1.3 nonillion column combinations
    - ◇ Impossible to check or even enumerate
- Possible solutions
  - Scale up: More RAM, faster CPUs
    - ◇ Expensive
  - Scale in: More cores
    - ◇ More complex (threading)
  - Scale out: More machines
    - ◇ Communication overhead
  - Intelligent enumeration and aggressive pruning



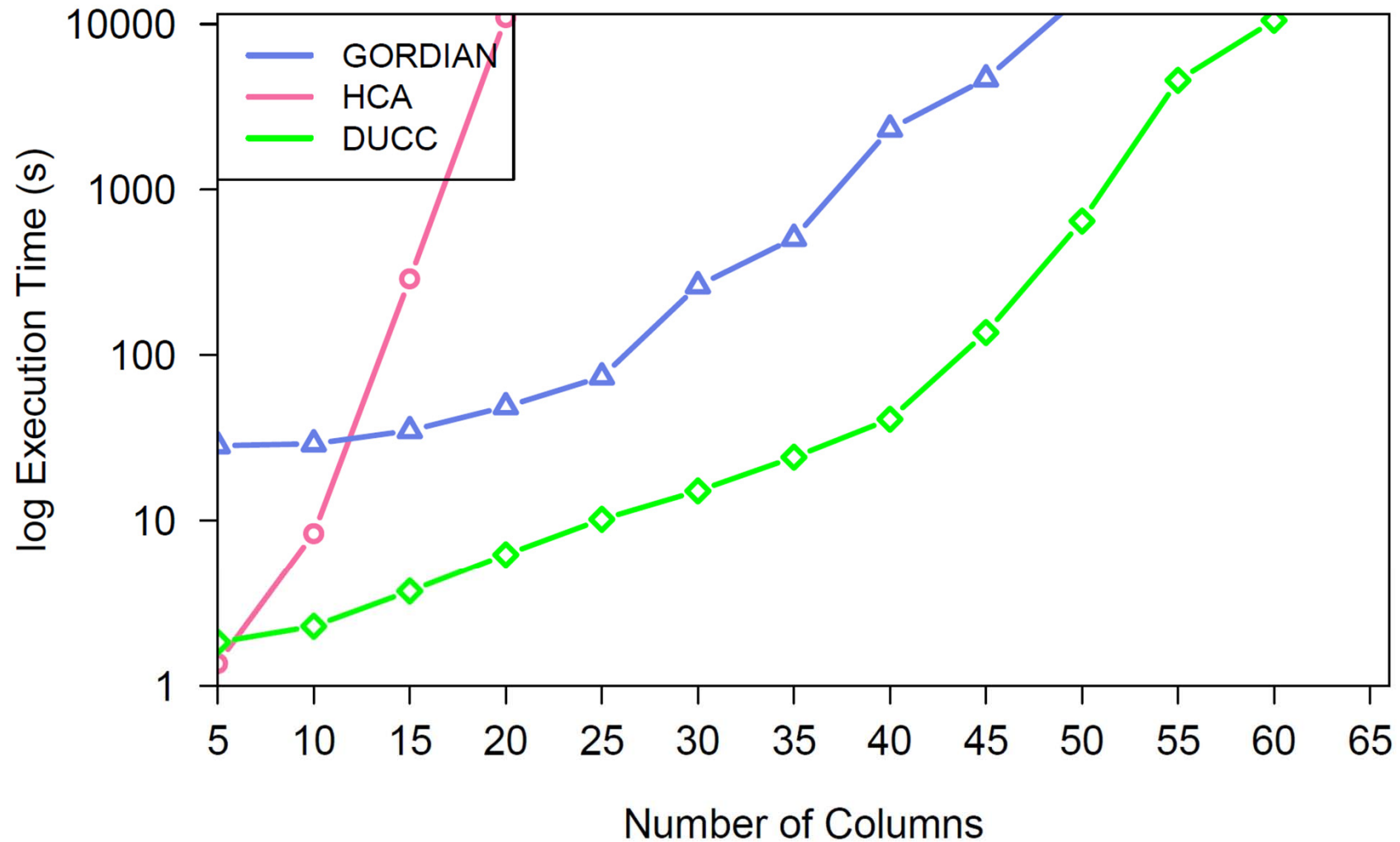




# Scaling the number of columns

35

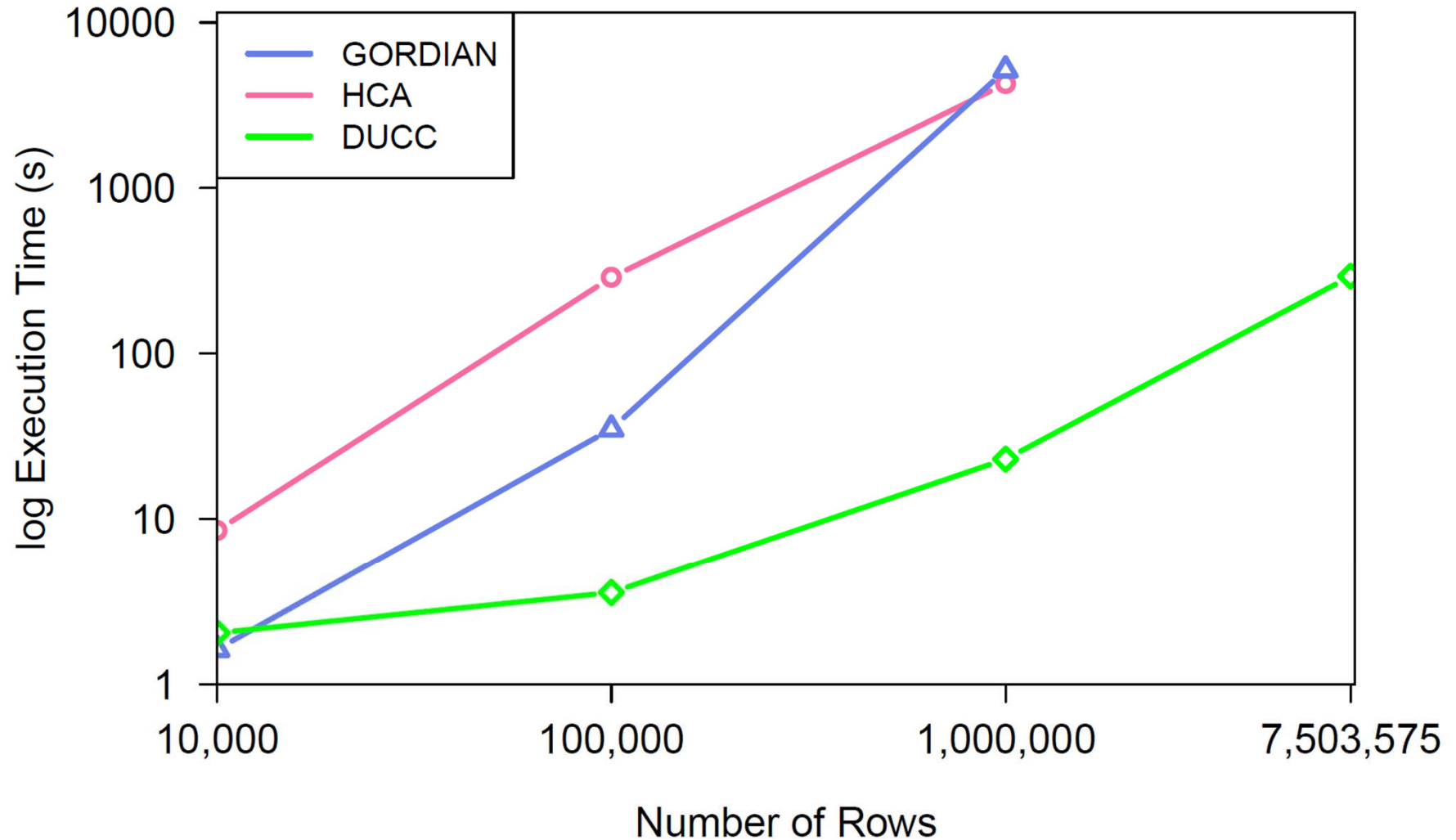
■ NCVoter, 100k rows



# Scaling the number of rows

36

■ NCVoter, 15 columns

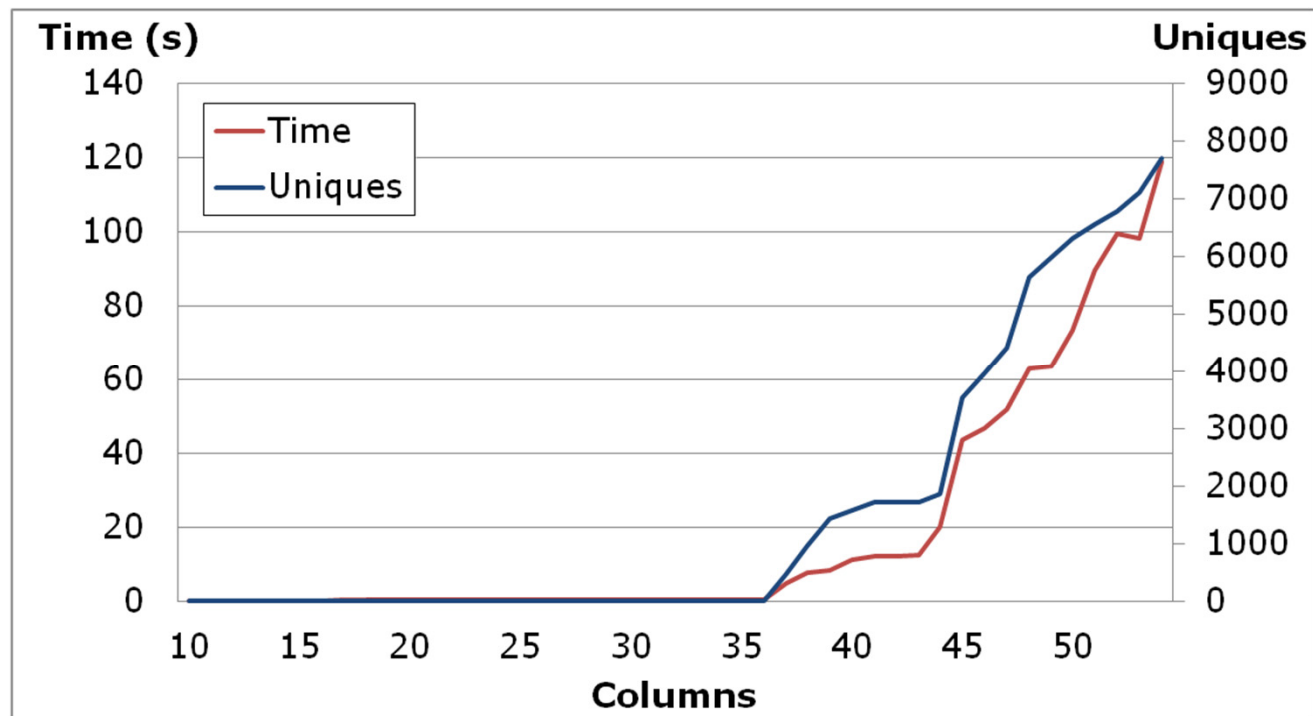




# Analysis of DUCC

37

- Runtime mainly depends on size of solution set



- Worst case: solution set in the middle

# Uniques and non-uniques in NC-voter data

38

- **A minimal unique:** voter\_reg\_num, zip\_code, race\_code
- **A maximal non-unique:** voter\_reg\_num, status\_cd, voter\_status\_desc, reason\_cd, voter\_status\_reason\_desc, absent\_ind, name\_prefix\_cd, name\_sufx\_cd, half\_code, street\_dir, street\_type\_cd, street\_sufx\_cd, unit\_designator, unit\_num, state\_cd, mail\_addr2, mail\_addr3, mail\_addr4, mail\_state, area\_cd, phone\_num, full\_phone\_number, drivers\_lic, race\_code, race\_desc, ethnic\_code, ethnic\_desc, party\_cd, party\_desc, sex\_code, sex, birth\_place, precinct\_abbrev, precinct\_desc, municipality\_abbrev, municipality\_desc, ward\_abbrev, ward\_desc, cong\_dist\_abbrev, cong\_dist\_desc, super\_court\_abbrev, super\_court\_desc, judic\_dist\_abbrev, judic\_dist\_desc, nc\_senate\_abbrev, nc\_senate\_desc, nc\_house\_abbrev, nc\_house\_desc, county\_commiss\_abbrev, county\_commiss\_desc, township\_abbrev, township\_desc, school\_dist\_abbrev, school\_dist\_desc, fire\_dist\_abbrev, fire\_dist\_desc, water\_dist\_abbrev, water\_dist\_desc, sewer\_dist\_abbrev, sewer\_dist\_desc, sanit\_dist\_abbrev, sanit\_dist\_desc, rescue\_dist\_abbrev, rescue\_dist\_desc, munic\_dist\_abbrev, munic\_dist\_desc, dist\_1\_abbrev, dist\_1\_desc, dist\_2\_abbrev, dist\_2\_desc, confidential\_ind, age, vtd\_abbrev, vtd\_desc

# Next steps

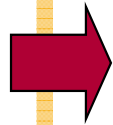
39

- Finding primary keys
  - Uniqueness is necessary criteria
  - No null values
  - Include other features
    - ◇ Name includes "id", number of columns
  
- Partial uniques
  - 99.9% of the data unique
  - Useful to detect data errors
  - Gordian, HCA, and DUCG can be easily modified
  
- Incremental discovery



# Agenda

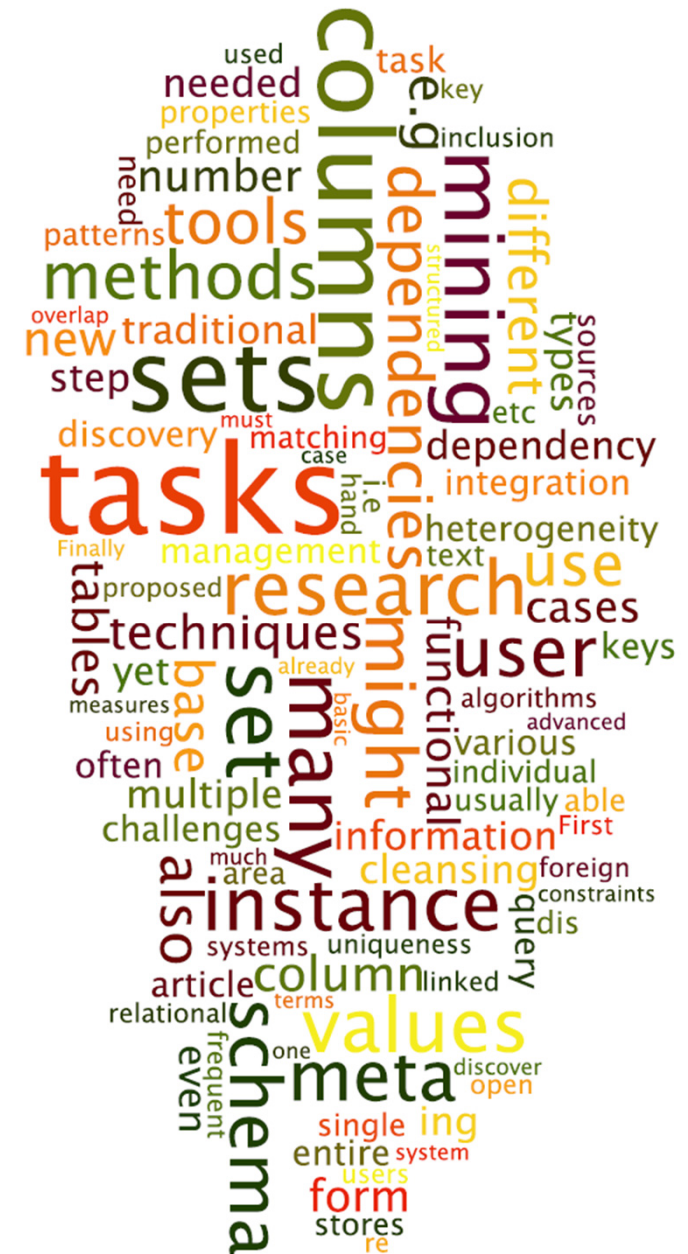
40



- Data profiling tasks
- Unique discovery
- Data profiling tools
- Advanced profiling
- Profiling heterogeneous data



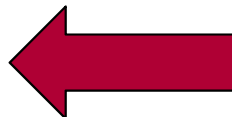
■ Pie I have eaten  
 ■ Pie I have not yet eaten



# Tools have very long feature lists

41

- Num rows
- Min value length
- Median value length
- Max value length
- Avg value length
- Precision of numeric values
- Scale of numeric values
- Quartiles
- Basic data types
- Num distinct values ("cardinality")
- Percentage null values
- Data class and data type
- Uniqueness and constancy
- Single-column frequency histogram
- Multi-column frequency histogram
- Pattern discovery (Aa9)
- Soundex frequencies
- Benford Law Frequency
- Single column primary key discovery
- Multi-column primary key discovery
- Single column IND discovery
- Inclusion percentage
- Single-column FK discovery
- Multi-column IND discovery
- Multi-column FK discovery
- Value overlap (cross domain analysis)
- Single-column FD discovery
- Multi-column FD discovery
- Text profiling

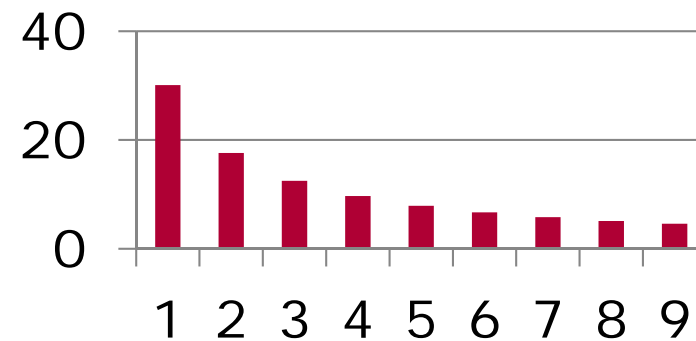
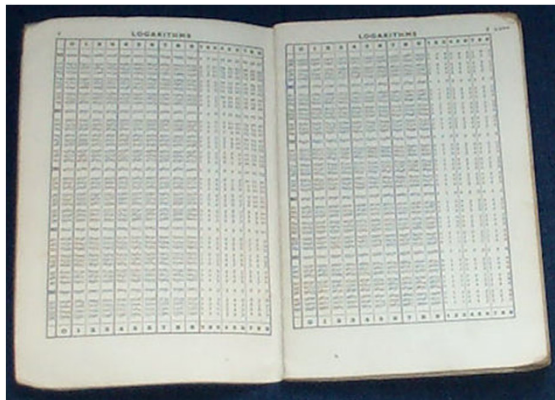


# An Aside: Benford Law Frequency ("first digit law")

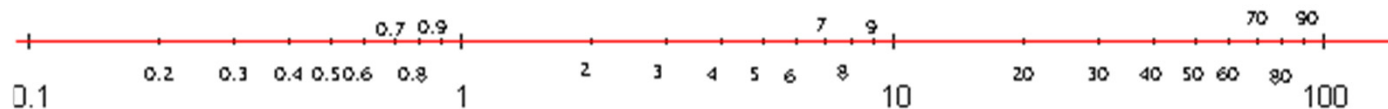
42

- Statement about the distribution of first digits  $d$  in (many) naturally occurring numbers:

- $P(d) = \log_{10}(d + 1) - \log_{10}(d) = \log_{10}(1 + 1/d)$



- Holds if  $\log(x)$  is uniformly distributed



# Examples for Benford's Law

43

- Surface areas of 335 rivers
- Sizes of 3259 US populations
- 104 physical constants
- 1800 molecular weights
- 5000 entries from a mathematical handbook
- 308 numbers contained in an issue of Reader's Digest
- Street addresses of the first 342 persons listed in American Men of Science

## Heights of the 60 tallest structures

Leading digit	meters		In Benford's law
	Count	%	
1	26	43.3%	30.1%
2	7	11.7%	17.6%
3	9	15.0%	12.5%
4	6	10.0%	9.7%
5	4	6.7%	7.9%
6	1	1.7%	6.7%
7	2	3.3%	5.8%
8	5	8.3%	5.1%
9	0	0.0%	4.6%

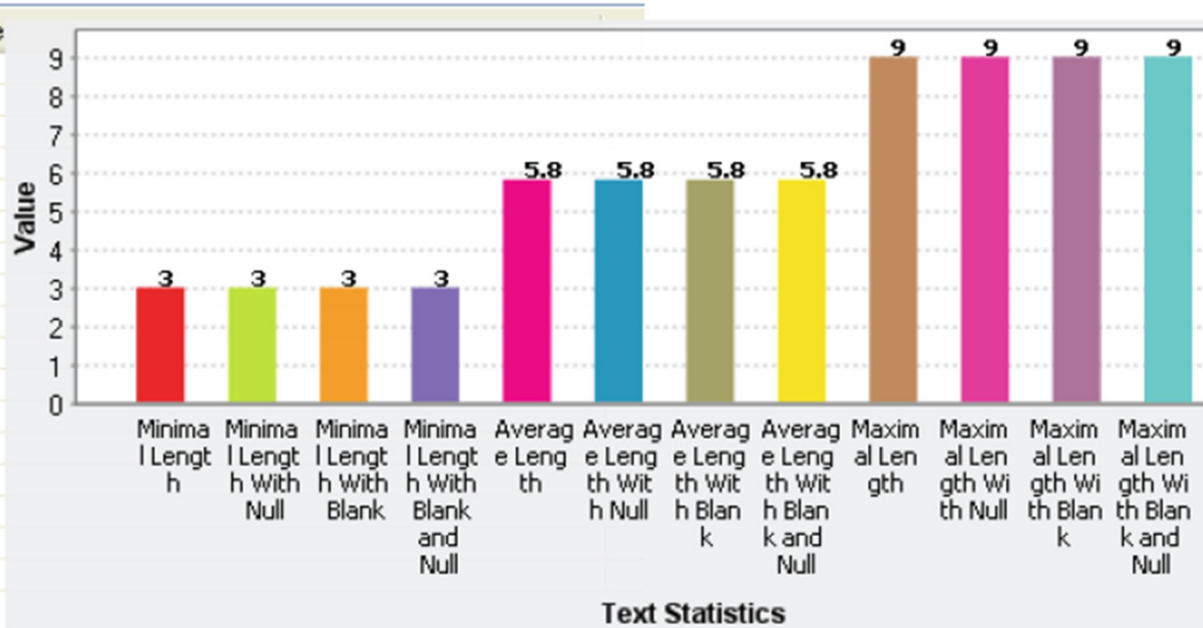


- ▼ Column:metadata.firstname
  - ▶ Pattern Matching
  - ▶ Simple Statistics
  - ▶ Soundex Frequency Table
  - ▼ Text Statistics

# Screenshots from Talend

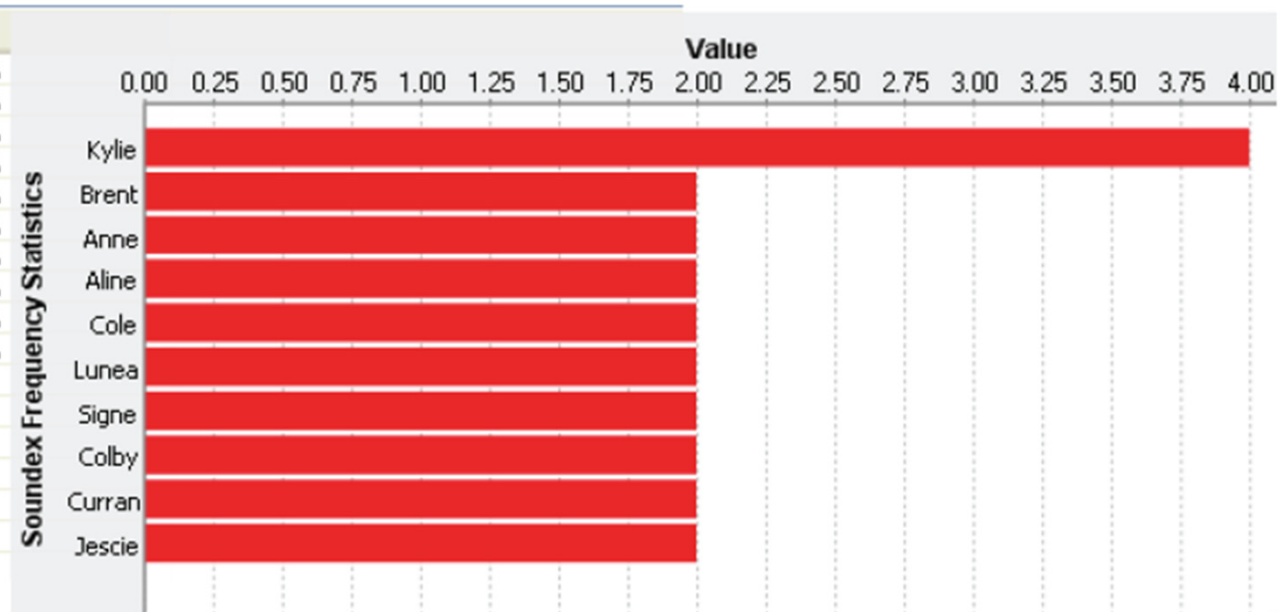
er  
ut

Label	Value
Minimal Length	3.00
Minimal Length With Null	3.00
Minimal Length With Blank	3.00
Minimal Length With Blank and Null	3.00
Average Length	5.80
Average Length With Null	5.80
Average Length With Blank	5.80
Average Length With Blank and Null	5.80
Maximal Length	9.00
Maximal Length With Null	9.00
Maximal Length With Blank	9.00
Maximal Length With Blank and Null	9.00



## ▼ Soundex Frequency Table

value	distinct count	count	%
Kylie	4.00	6	N/A
Brent	2.00	3	N/A
Anne	2.00	2	N/A
Aline	2.00	2	N/A
Cole	2.00	2	N/A
Lunea	2.00	2	N/A
Signe	2.00	2	N/A
Colby	2.00	3	N/A
Curran	2.00	2	N/A
Jescie	2.00	2	N/A



# Screenshots from IBM Information Analyzer

45

IBM. Information Server File Edit View Help

IA\_OVERVIEW\_PROJECT 9.43.86.77

INVESTIGATE Foreign Key Analysis

Select Data Source to Work With

EMPLOYEE DEPARTMENT

Open Foreign Key Analysis

View Details

You can use this pane to view analysis details about a primary key column and the foreign key column that is associated with the primary key column.

Frequency Values Analysis Details

**Foreign Key Candidate Pair**

	Base Column	Paired Column
Column	EMPNO	MGRNO
Table	EMPLOYEE	DEPARTMENT
Source	IA	IA
Primary Key	Yes	No
Foreign Key	No	Yes
<b>Data Class</b>	<b>Identifier</b>	<b>Quantity</b>
Data Type	INT32	INT8
Length	0	0
Precision	0	0
Scale	0	0
<b>Cardinality</b>	<b>48</b>	<b>9</b>
Unique	No	No
Constant	No	No
Definition	No	No

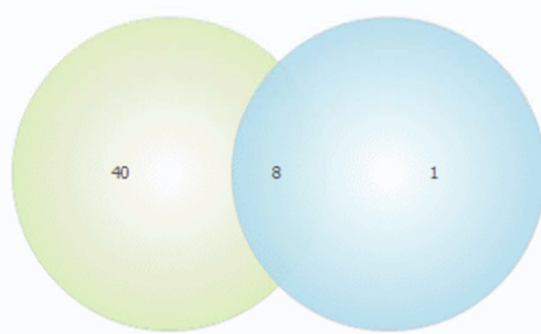
**Paired to Base:**

Common Data Values: 8 100.0000% Common Domain: Yes

**Base to Paired:**

Common Data Values: 8 16.6667% Common Domain: No

**Common Domain :**



Legend: Base Column (Green), Paired Column (Blue)

Close

# Typical shortcomings of tools (and methods from research)

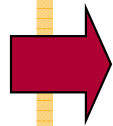
46

- Usability
  - Complex to configure
  - Results complex to view and interpret
- Scalability
  - Main-memory based
  - SQL based
- Efficiency
  - Coffee, Lunch, Overnight
- Functionality
  - Restricted to simplest tasks
  - Restricted to individual columns or small column sets
    - ◇ “Realistic” key candidates vs. further use-cases
  - „Checking” vs. „discovery”

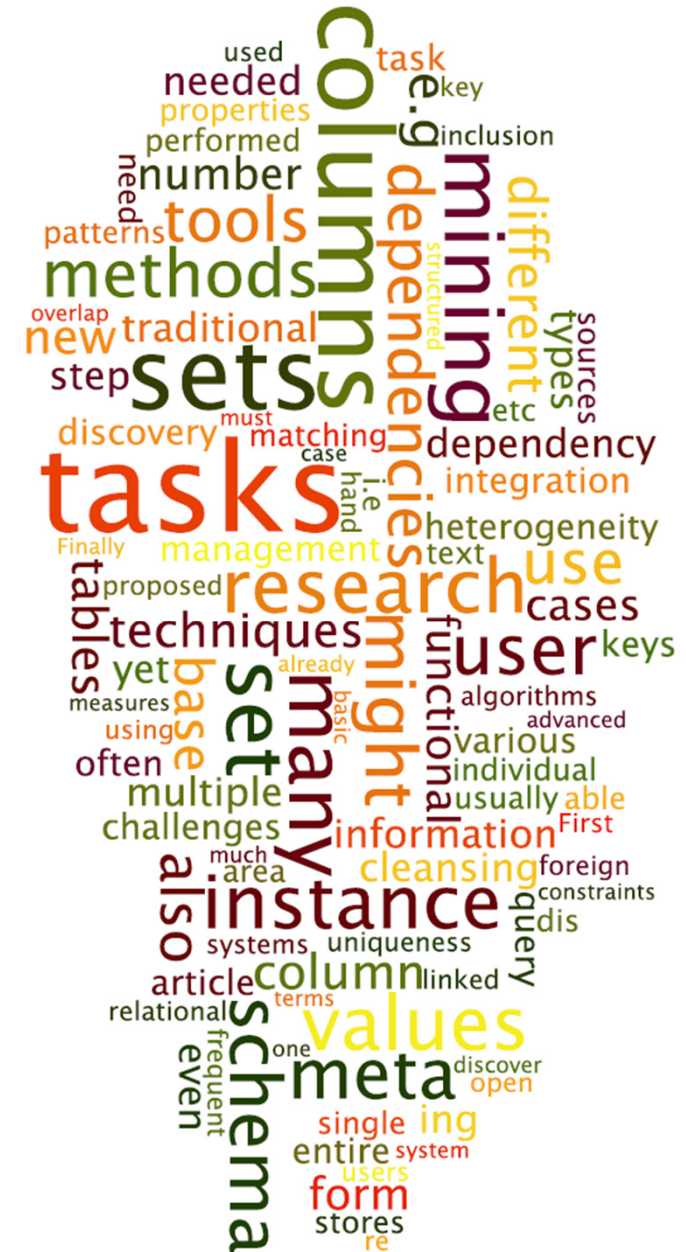
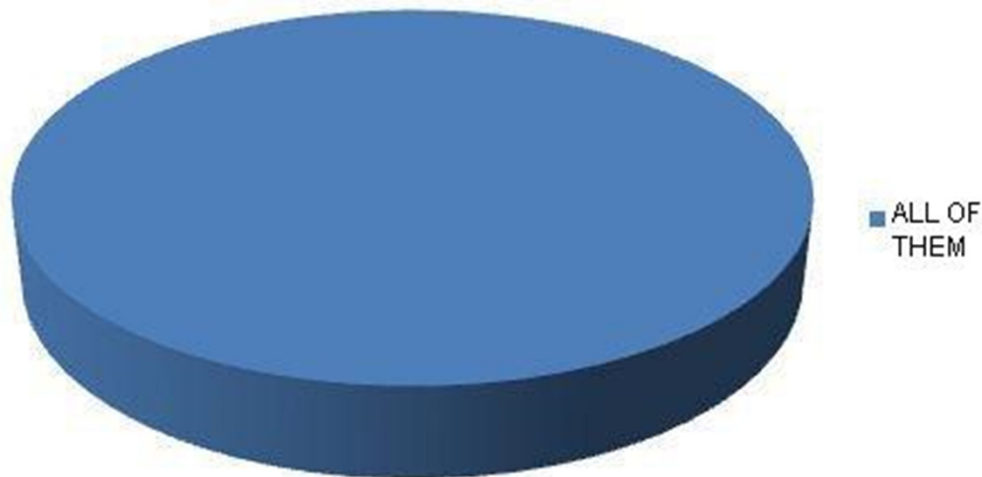
# Agenda

47

- Data profiling tasks
- Unique discovery
- Data profiling tools
- Advanced profiling
- Profiling heterogeneous data



% of ice hockey pucks that are pie chart shaped.





# Online profiling

48

- Profiling is long procedure
  - Boring for developers
  - Expensive for machines (I/O and CPU)
  
- Challenge: Display intermediate results
  - ... of improving/converging accuracy
  - Allows early abort of profiling run
  
- Gear algorithms toward that goal
  - Allow intermediate output
  - Enable early output: “progressive” profiling



# Incremental profiling

49

- Data is dynamic
  - Insert (batch or tuple-based)
  - Updates
  - Deletes
  
- Problem: Keep profiling results up-to-date...
  - ... without re-profiling the entire data set.
  - Easy examples: SUM, MIN, MAX, COUNT, AVG
  - Difficult examples: MEDIAN, uniqueness [ICDE14], dependencies

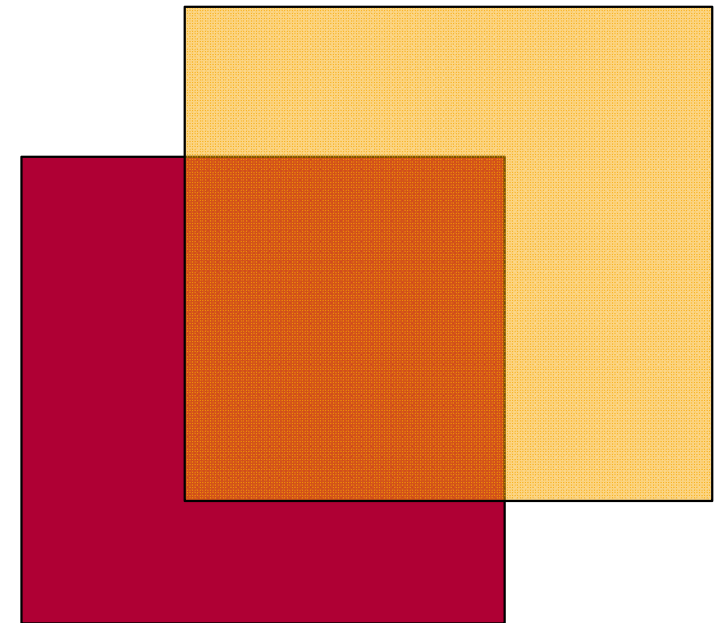




# Profiling for Integration

51

- Profile multiple sources simultaneously
- Schema matching/mapping
  - What constitutes the “difficulty” of matching/mapping?
- Duplicate detection
  - Estimate data overlap
  - Estimate fusion effort
- Create measures to estimate integration (and cleansing) effort
  - Schema and data overlap
  - Severity of heterogeneity





# Topic discovery and clustering

52

- What is a dataset **about**?
  - Domain(s), topics, entity types
- Single-topic datasets
  - GraceDB, IMDB, Sports databases, etc.
- Multi-topic datasets
  - General purpose knowledge bases: YAGO, DBpedia
  - 154 million tables crawled from the Web (2008)

Categories: [Ronald Reagan](#) | [1911 births](#) | [2004 deaths](#) | [20th-century presidents of the United States](#) | [Actors awarded British knighthoods](#) | [Actors from Illinois](#) | [American actor-politicians](#) | [American anti-communists](#) | [American Disciples of Christ](#) | [American film actors](#) | [American labor leaders](#) | [American military personnel of World War II](#) | [American politicians of Irish descent](#) | [American Presbyterians](#) | [American radio personalities](#) | [American shooting survivors](#) | [American television personalities](#) | [Actors from California](#) | [California Republicans](#) | [Chicago Cubs broadcasters](#) | [Cold War leaders](#) | [College football announcers](#) | [Colorectal cancer survivors](#) | [Congressional Gold Medal recipients](#) | [Conservatism](#) | [Deaths from Alzheimer's disease](#) | [Deaths from pneumonia](#) | [Delegates to the Republican National Convention](#) | [Eureka College](#) | [First Motion Picture Unit personnel](#) | [Governors of California](#) | [History of the United States \(1980–1991\)](#) | [Infectious disease deaths in California](#) | [Iran–Contra affair](#) | [Knights Grand Cross of the Order of the Bath](#) | [Major League Baseball announcers](#) | [New Right \(United States\)](#) | [People from Lee County, Illinois](#) | [People from Whiteside County, Illinois](#) | [People from the Greater Los Angeles Area](#) | [People involved in the Soviet war in Afghanistan](#) | [Presidential Medal of Freedom recipients](#) | [Presidents of the United States](#) | [Reagan family](#) | [Recipients of Honorary British Knighthoods](#) | [Recipients of the Order of the White Eagle \(Poland\)](#) | [Republican Party \(United States\) presidential nominees](#) | [Restoration Movement](#) | [Skin cancer survivors](#) | [Time magazine Persons of the Year](#) | [United States Army Air Forces officers](#) | [United States Army officers](#) | [United States presidential candidates, 1968](#) | [United States presidential candidates, 1976](#) | [United States presidential candidates, 1980](#) | [United States presidential candidates, 1984](#)

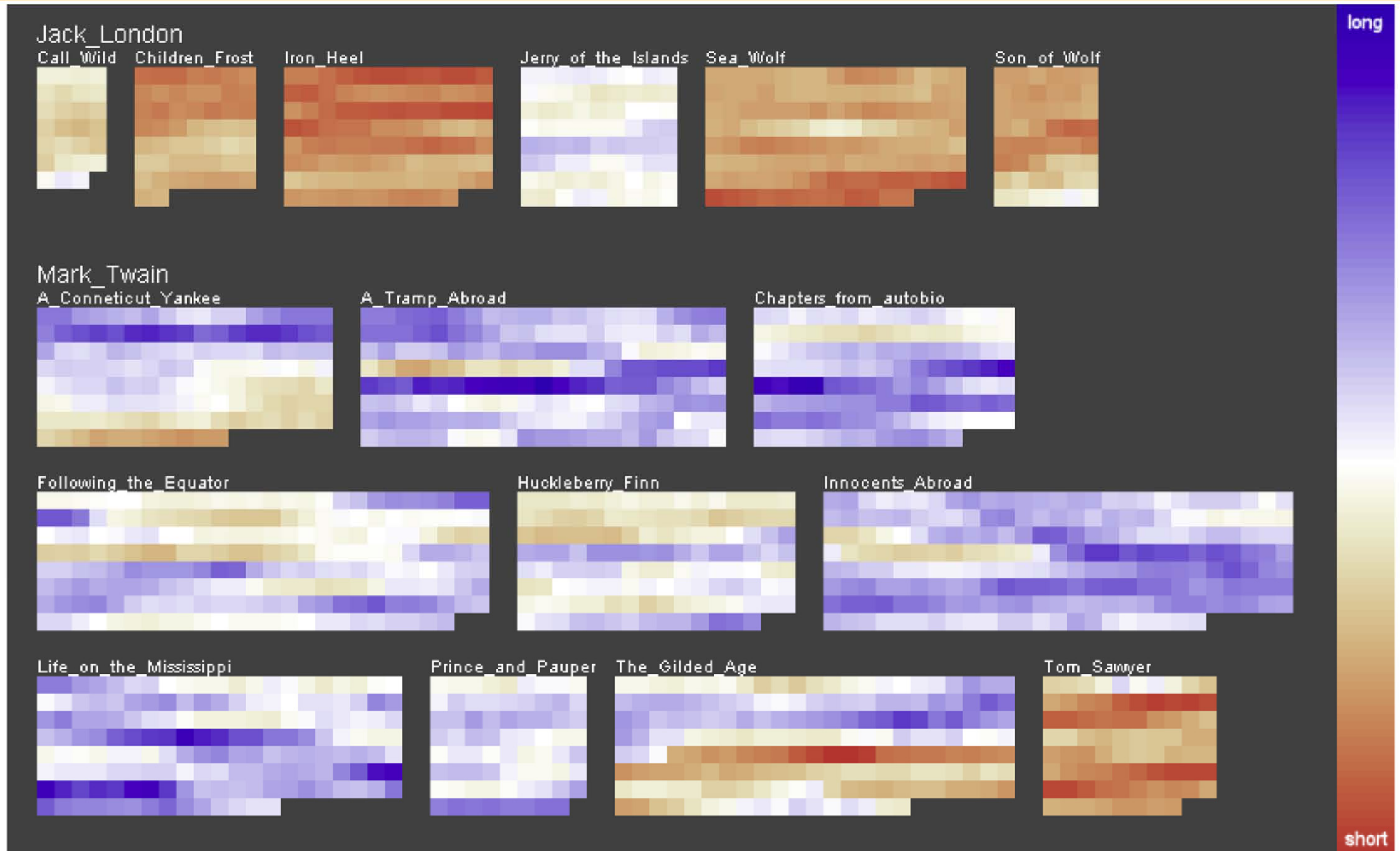
# Profiling new types of data

53

- Traditional data profiling: Single table or multiple tables
- More and more data in other models
  - XML / nested relational / JSON
  - RDF triples
  - Textual data: Blogs, Tweets, News
  - Multimedia data
- Different models offer new dimensions to profile
  - XML: Nestedness, measures at different nesting levels
  - RDF: Graph structure, in- and outdegrees
  - Multimedia: Color, video-length, volume, etc.
  - **Text**: Sentiment, sentence structure, complexity, and other linguistic measures

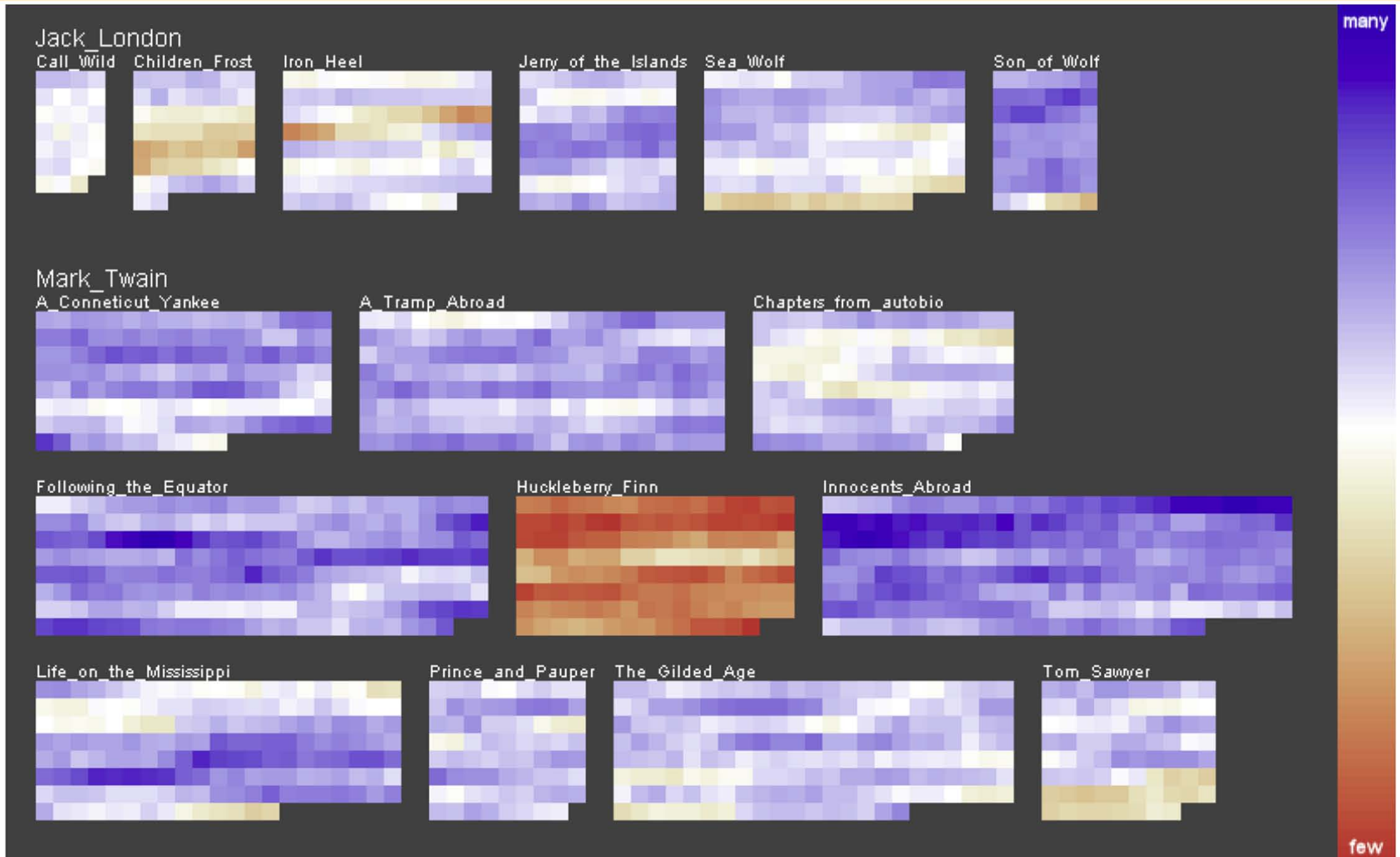
# Average sentence length

54



# Hapax Legomena

55



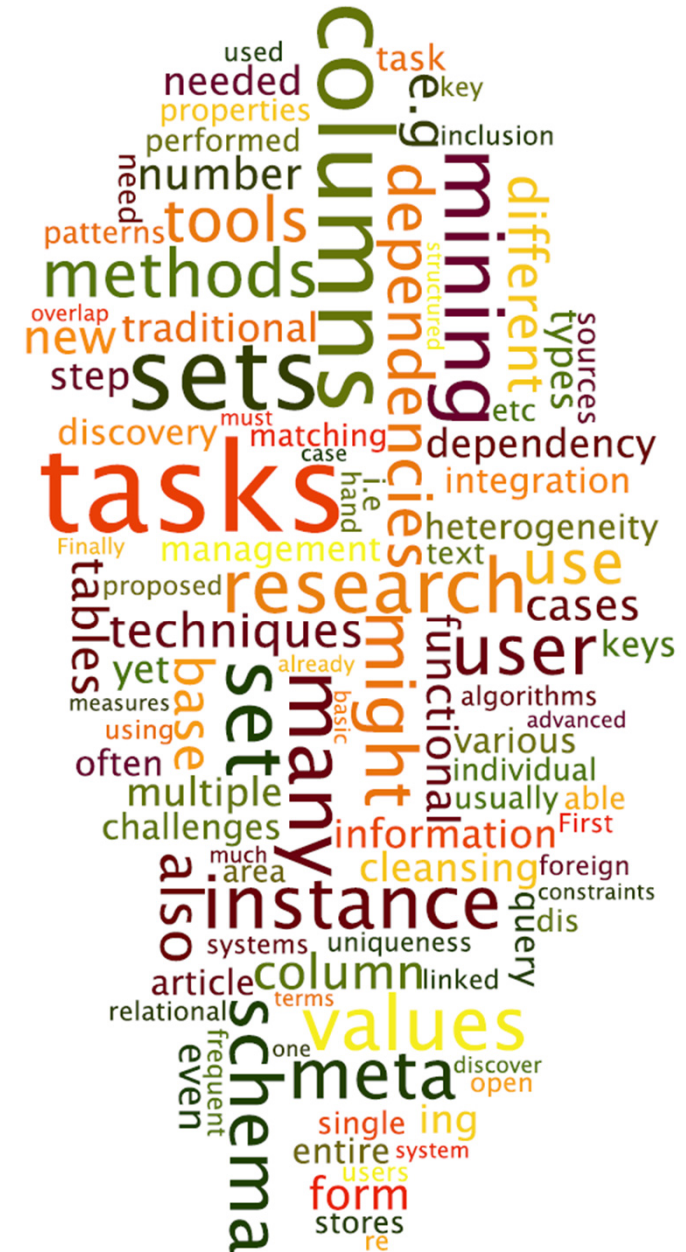
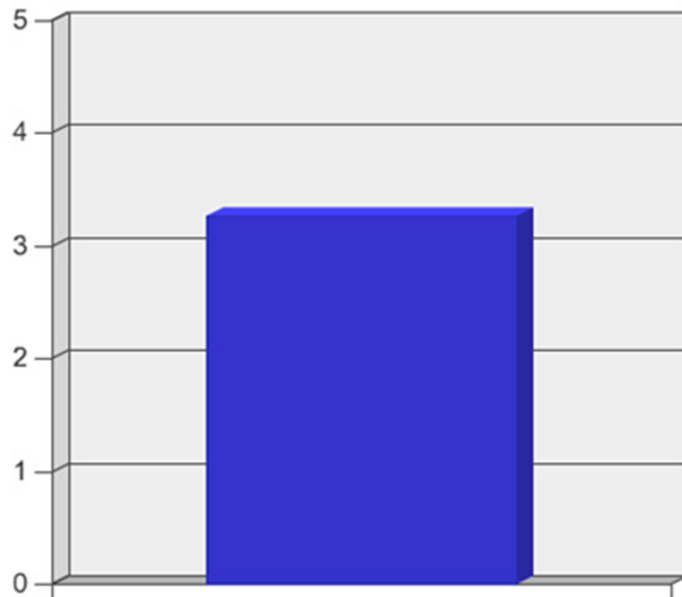


# Summary

56

- Data profiling tasks
- Unique discovery
- Data profiling tools
- Advanced profiling
- Profiling heterogeneous data

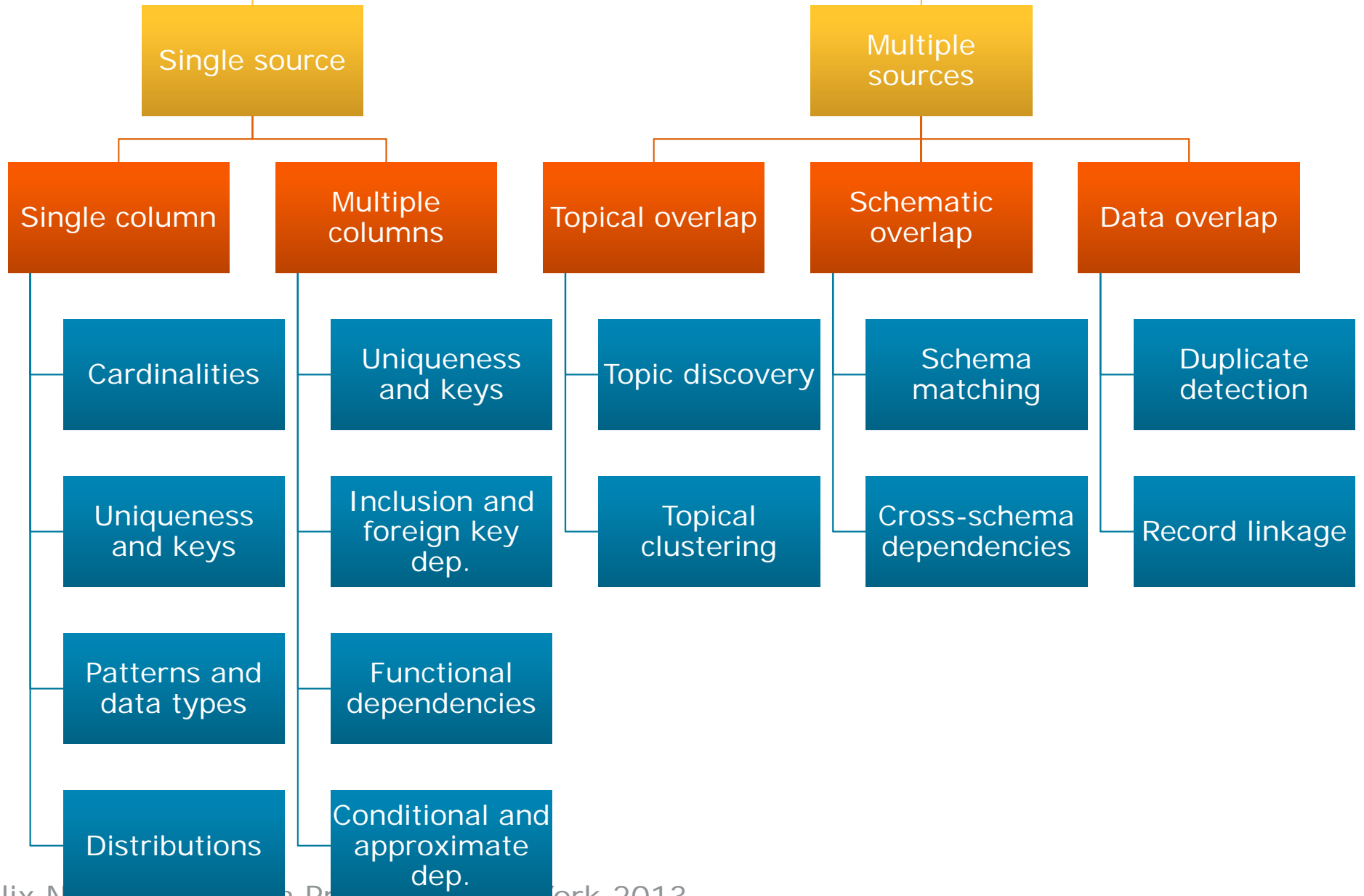
**Pi Chart**



# Summary

## Data Profiling

57



# Bibliography

58

- Slides are (in part) based on
  - [SIGREC] F. Naumann. *Data Profiling Revisited*. SIGMOD Record, 42, 2013 (to appear)
- [CIKM11] Z. Abedjan and F. Naumann. *Advancing the discovery of unique column combinations*. CIKM 2011.
- [ICDE14] Z. Abedjan, J. Quanie-Ruiz, F. Naumann. *Detecting Unique Column Combinations on Dynamic Data*. ICDE 2014.
- [PVLDB14] A. Heise, J. Quiane-Ruiz, Z. Abedjan, A. Jentzsch, F. Naumann. *Scalable Discovery of Unique Column Combinations*. PVLDB 2014.
- [Gia99] C. Giannella and C. Wyss. *Finding minimal keys in a relation instance*. TR 1999.
- [Sis06] Y. Sismanis, P. Brown, P. J. Haas, and B. Reinwald. *Gordian: efficient and scalable discovery of composite keys*. VLDB 2006.