



An Introduction to Data Profiling

27.4.2017
Felix Naumann

Overview

- 1. Profiling tasks**
2. Profiling tools
3. Visualization
4. Next generation profiling
5. Profiling challenges

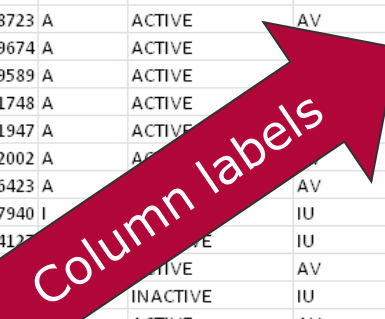


Felix Naumann
Data Profiling
Summer 2017

Microsoft Excel interface showing the ribbon with tabs: Datei, Start, Einfügen, Seitenlayout, Formeln, Daten, Überprüfen, Ansicht, Add-Ins. The ribbon includes options for font (Calibri, 11), alignment, and number formatting. The status bar at the bottom shows 'V106177' and 'O'.

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
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	VA	06/29/1998	13
105911	VERIFIED	YUAN	YUEXI			1208 RASPBERRY F	GRAHAM	NC	27253	1208 RASPBERRY RUN	GRAHAM	NC		27253		A	UN	UNA	F		63	OC	10.10.2008	06S
105912	OVERSEAS CI	YUEN	DIANE			552 MEADOWWOOD	BURLINGTON	NC	27215	552 MEADOWWOOD DR	BURLINGTON	NC		27215		U	UN	DEM	F		24		09/28/2012	03C
105916	VERIFIED	YUQILIMA	AIDA	GUILLERMIN		1204 GRACE AVE	BURLINGTON	NC	27217	1204 GRACE AVE	BURLINGTON	NC		27217	336 675 4492	U	UN	DEM	U		57	OC	10.01.2008	12E
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		05.04.2012	35
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	PA	03.12.2010	10S
105959	CONFIRMATI	ZALDIVAR	NICK			3475 S NC HWY 119	HAW RIVER	NC	27258	3475 S NC HWY 119	HAW RIVER	NC		27258	919 741 3647	O	NL	UNA	M		21	NY	08.03.2010	103
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	OC	10.08.2003	4
105971	VERIFIED	ZAMORA	J BUENAVEN	LOPEZ		4398 BLUE HERON	MEBANE	NC	27302	4398 BLUE HERON LN	MEBANE	NC		27302	336 516 1586	O	HL	REP	M		41	OC	10.07.2008	10N
105972	VERIFIED	ZAMORA	MIRA	ELIZABETH		4398 BLUE HERON	MEBANE	NC	27302	4398 BLUE HERON LN	MEBANE	NC		27302	336 290 3215	O	HL	DEM	F		19	NC	06.05.2012	10N
105973	VERIFIED	ZAMORA	MARIANA			2552 DARE ST	BURLINGTON	NC	27217	2552 DARE ST	BURLINGTON	NC		27217	336 264 8329	U	UN	UNA	U		23		07/30/2012	63
105974	VERIFIED	ZAMORA	SOFIA	CAMACHO		4398 BLUE HERON	MEBANE	NC	27302	4398 BLUE HERON LN	MEBANE	NC		27302	919 304 2412	U	HL	UNA	F		43	OC	07/27/2004	10N
105977	VERIFIED	ZAMPARELLI	EDNA	KOKO		627 W FRONT ST	BURLINGTON	NC	27215	627 W FRONT ST #A	BURLINGTON	NC		27215	704 307 8555	B	UN	DEM	F		35	OC	09.05.2008	12W
105978	VERIFIED	ZAMUDIO	MILTON	MARTIN		2723 MEADOW OAK	HAW RIVER	NC	27258	2723 MEADOW OAK DR	HAW RIVER	NC		27258	786 488 5889	O	NL	DEM	M		47	FL	03/20/2012	09N
106000	VERIFIED	ZAPATA	ANA	GONZALEZ		660 MANNING AV	ELON	NC	27244	660 MANNING AVE #73	ELON	NC		27244	336 449 9421	O	NL	DEM	F		57	NY	10.11.2000	35
106001	VERIFIED	ZAPATA	TIFFANY	MARIA		208 COLLEGE ST	GRAHAM	NC	27253	208 COLLEGE ST #G	GRAHAM	NC		27253	336 270 4116	U	UN	DEM	U		27		07/30/2012	06N
106002	VERIFIED	ZAPIEN	ABELARDO			1403 N PARK AVE	BURLINGTON	NC	27217	1403 N PARK AVE	BURLINGTON	NC		27217	336 269 5746	U	UN	DEM	M		20		10.12.2012	127
106017	CONFIRMATI	ZAVALA	ANNA	CRISTINA		101 BARNEY DORN	ELON	NC	27244	CAMPUS BOX 8188	ELON	NC		27244	412 527 0279	U	UN	DEM	F		24	PA	09.10.2008	03N
106018	VERIFIED	ZAVALA	LUZ			632 E ELM ST	GRAHAM	NC	27253	632 E ELM ST	GRAHAM	NC		27253	336 343 9932	O	HL	DEM	F		60		09/23/2008	06N
106019	VERIFIED	ZAVALA	STEPHANIE	PAOLA		632 E ELM ST	GRAHAM	NC	27253	632 E ELM ST	GRAHAM	NC		27253		O	HL	DEM	F		23		09/23/2008	06N
106023	CONFIRMATI	ZAYAS	DEBORAH	PINNIX		725 SHARPE RD	BURLINGTON	NC	27217	725 SHARPE RD	BURLINGTON	NC		27217	227 6960	B	NL	DEM	F		58	NC	01/31/1991	12N
106024	UNAVAILABL	ZAYAS	EDGARDO	LUIS		2313 HUNTINGTOI	BURLINGTON	NC	27215	2313 HUNTINGTON RD	BURLINGTON	NC		27215	910 570 0091	O	NL	REP	M		43		10.01.1996	
106027	VERIFIED	ZAYAS	NANCY	CORDERO		510 TRAIL ONE	BURLINGTON	NC	27215	510 TRAIL ONE #F-104	BURLINGTON	NC		27215	336 269 8716	O	HL	UNA	F		55	NY	09/19/1994	1210
106034	VERIFIED	ZEIGLER	EDDIE	LAMAR		1211 OLYMPIC DR	MEBANE	NC	27302	1211 OLYMPIC DR	MEBANE	NC		27302	919 568 8126	B	NL	DEM	M		50	FL	09/25/2008	10S
106036	VERIFIED	ZEIGLER	RONDA	W		1211 OLYMPIC DR	MEBANE	NC	27302	1211 OLYMPIC DR	MEBANE	NC		27302	919 568 8126	B	NL	DEM	F		41	NC	03/29/1990	10S
106037	VERIFIED	ZEIGLER	SHEILA	MOBLEY		1237 KENWOOD D	BURLINGTON	NC	27215	1237 KENWOOD DR	BURLINGTON	NC		27215	336 570 0146	B	NL	DEM	F		64	NC	10/21/1993	128
106038	VERIFIED	ZELAYA	MARLENY	ARELY		611 CHANDLER AV	BURLINGTON	NC	27217	611 CHANDLER AVE	BURLINGTON	NC		27217	336 512 7292	U	HL	DEM	F		29	MD	11.04.2008	12N
106045	VERIFIED	ZEMPEL	GINA	LOUISA		603 E DAVIS ST	BURLINGTON	NC	27215	603 E DAVIS ST	BURLINGTON	NC		27215	262 672 0902	O	HL	UNA	F		32	MI	06/23/2010	128
106057	REMOVED A	ZHAGUI	CECILIA	ISABEL		2732 WESTCHESTE	BURLINGTON	NC	27217	2732 WESTCHESTER DR	BURLINGTON	NC		27217	336 228 3620	U	HL	DEM	F		35	OC	08.02.2004	
106058	VERIFIED	ZHENG	JICHUN			1914 HILLDALE DR	BURLINGTON	NC	27215	1914 HILLDALE DR	BURLINGTON	NC		27215	917 977 1073	U	UN	UNA	U		25		04/30/2012	1210
106059	CONFIRMATI	ZHENG	KELLY			95 DENISE DR	BURLINGTON	NC	27215	95 DENISE DR #B	BURLINGTON	NC		27215		A	UN	UNA	F		26	OC	09/16/2008	1210
106060	VERIFIED	ZHENG	SAIFEN			515 BURLINGTON	GIBSONVILLE	NC	27249	515 BURLINGTON AVE	GIBSONVILLE	NC		27249		A	UN	UNA	F		33	OC	12.12.2011	03W
106066	VERIFIED	ZIEGLER	BRANDON	GERALD		2459 MORNINGSID	BURLINGTON	NC	27217	2459 MORNINGSIDE DR	BURLINGTON	NC		27217	336 212 7747	B	NL	REP	M		22	NJ	08/21/2008	63
106072	VERIFIED	ZIEGLER	VIENGSAMA	MINDY		2459 MORNINGSID	BURLINGTON	NC	27217	2459 MORNINGSIDE DR	BURLINGTON	NC		27217		U	UN	UNA	F		24		11.06.2012	63
106079	VERIFIED	ZIENTY	SCOTT	CARL		511 W HAGGARD A	ELON	NC	27244	511 W HAGGARD AVE	ELON	NC		27244		O	NL	UNA	M		23	MO	03/13/2012	03N
106100	VERIFIED	ZIMMERMAN	KATRINA	MCINTOSH		852 S MAIN ST	BURLINGTON	NC	27215	852 S MAIN ST #16	BURLINGTON	NC		27215		M	NL	UNA	F		25		10/29/2012	128
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	OC	09/26/2008	06W
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		05.04.2012	03N
106139	VERIFIED	ZMIJEASKI	SEAN			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD	MEBANE	NC		27302	336 376 1987	O	UN	REP	M		23		04.09.2008	09S
106144	VERIFIED	ZMIJEWSKI	N VIRGINIA	LOURDES		4872 THOM RD	MEBANE	NC	27302	4872 THOM RD	MEBANE	NC		27302	336 376 1987	U	UN	UNA	F		45	PR	03/24/2005	09S
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	NC	10.08.2012	03S
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	OC	08/14/2003	03S
106154	VERIFIED	ZOLLI	COFFEY	ANTONIO	MARK	108 OAKGROVE D	GRAHAM	NC	27253	108 OAKGROVE DR	GRAHAM	NC		27253	336 260 6673	B	UN	DEM	M		53	MD	10.07.2008	06E

A1	county_id																								
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X		
1	county	county_desc	voter_reg	nistatus_cd	voter_status_desc	reason_cd	voter_status	last_name	first_name	mid_name	name	res_street_address	res_city_desc	state	zip_code	mail_addr1	mail_addr2	mail_city	mail_state	mail_zipcode	full_phone	race_code	ethnic_code	party_cd	
2	1	ALAMANCE	9005990	A	ACTIVE	AV	VERIFIED	AABEL	EVELYN	LARSEN		4430 E GREENSBOR	GRAHAM	NC	27253	4430 E GREENSBORO-CHA	GRAHAM	NC	27253	000 0000	W	NL	UNA		
3	1	ALAMANCE	9048723	A	ACTIVE	AV	VERIFIED	AARON	CHRISTINA	CASTAGNA		421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177	BURLINGTON	NC	27215	229 1110	W	UN	UNA		
4	1	ALAMANCE	9019674	A	ACTIVE	AV	VERIFIED	AARON	CLAUDIA	HAYDEN		1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST	BURLINGTON	NC	27215	222 8834	W	NL	UNA		
5	1	ALAMANCE	9129589	A	ACTIVE	AV	VERIFIED	AARON	JAMES	MICHAEL		1647 SAXAPAHAW	GRAHAM	NC	27253	PO BOX 98	SAXAPAHAW	NC	27340	336 525 2484	W	UN	DEM		
6	1	ALAMANCE	9041748	A	ACTIVE	AV	VERIFIED	AARON	NATHAN	EDWARD		421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177	BURLINGTON	NC	27215	336 229 1110	W	UN	UNA		
7	1	ALAMANCE	9021947	A	ACTIVE	AV	VERIFIED	AARON	WILLIE	DALE		1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST	BURLINGTON	NC	27215	336 999 9999	W	NL	UNA		
8	1	ALAMANCE	9062002	A	ACTIVE	AV	VERIFIED	AARONSON	GENA	HOLT		107 TERRYWOOD	HAW RIVER	NC	27258	107 TERRYWOOD CT	HAW RIVER	NC	27258	336 578 9123	W	NL	REP		
9	1	ALAMANCE	9096423	A	ACTIVE	AV	VERIFIED	AARONSON	MICHAEL	CHARLES		107 TERRYWOOD	HAW RIVER	NC	27258	107 TERRYWOOD CT	HAW RIVER	NC	27258	336 266 7615	W	NL	UNA		
10	1	ALAMANCE	9117940	I	INACTIVE	IU	CONFIRMATI	ABAD	PRISCILLA	MARIE		100 COLONNADE	ELON	NC	27244	CAMPUS BOX 3008	ELON	NC	27244		O	HL	UNA		
11	1	ALAMANCE	9034127	A	ACTIVE	AV	CONFIRMATI	ABADIE	COLLEEN	MIASHEL		1097 IVEY RD #C	GRAHAM	NC	27253	1097 IVEY RD #C	GRAHAM	NC	27253		M	HL	REP		
12	1	ALAMANCE	9111111	A	ACTIVE	AV	VERIFIED	ABADIE	JACK	EDWARD JR		612 SIDEVIEW ST	GRAHAM	NC	27253	612 SIDEVIEW ST	GRAHAM	NC	27253	336 212 8140	W	NL	UNA		
13	1	ALAMANCE	9111111	A	INACTIVE	IU	CONFIRMATI	ABADIE	MYRA	HOLLIFIELD		612 SIDEVIEW ST	GRAHAM	NC	27253	617 MITCHELL ST	BURLINGTON	NC	27217	336 212 8140	W	NL	UNA		
14	1	ALAMANCE	9151111	A	ACTIVE	AV	VERIFIED	ABBAS	FALISA			707 SUMMIT RIDG	MEBANE	NC	27302	707 SUMMIT RIDGE RD #	MEBANE	NC	27302	919 568 9001	B	UN	DEM		
15	1	ALAMANCE	9068411	A	ACTIVE	AV	VERIFIED	ABBAS	RAFAT			514 WESTRIDGE D	BURLINGTON	NC	27215	514 WESTRIDGE DR	BURLINGTON	NC	27215		A	UN	DEM		
16	1	ALAMANCE	9049573	A	ACTIVE	AV	VERIFIED	ABBATECOLA	RONALD	JOSEPH JR		504 BROOKFIELD	GIBSONVILLE	NC	27249	504 BROOKFIELD DR	GIBSONVILLE	NC	27249	336 449 9029	W	UN	UNA		
17	1	ALAMANCE	9033877	A	ACTIVE	AV	VERIFIED	ABBATECOLA	TRACY	BOONE		504 BROOKFIELD	GIBSONVILLE	NC	27249	504 BROOKFIELD DR	GIBSONVILLE	NC	27249		W	NL	DEM		
18	1	ALAMANCE	9083557	I	INACTIVE	IU	CONFIRMATI	ABBETT	DAWN	LEANN		3900 JOHNS CREEK	GIBSONVILLE	NC	27249	3900 JOHNS CREEK DR	GIBSONVILLE	NC	27249	336 584 3319	W	NL	DEM		
19	1	ALAMANCE	9027554	A	ACTIVE	AV	VERIFIED	ABBEY	BRENT	DAVID		3304 GOLDEN OAK	GRAHAM	NC	27253	3304 GOLDEN OAKS DR	GRAHAM	NC	27253	919 682 6873	W	NL	REP		
20	1	ALAMANCE	9029477	A	ACTIVE	AV	VERIFIED	ABBEY	DEMETRA	AINSWORTH		3304 GOLDEN OAK	GRAHAM	NC	27253	3304 GOLDEN OAKS DR	GRAHAM	NC	27253	336 376 0673	W	NL	REP		
21	1	ALAMANCE	9022529	I	INACTIVE	IU	CONFIRMATI	ABBEY	DOROTHY	ESTELLA		1029A QUAKENBUS	SNOW CAMP	NC	27349	1029A QUAKENBUSH RD	SNOW CAMP	NC	27349	376 3663	W	NL	REP		
22	1	ALAMANCE	9113186	A	ACTIVE	AV	VERIFIED	ABBOTT	AMELIA	BETH		2876 CALLOWAY D	MEBANE	NC	27302	2876 CALLOWAY DR	MEBANE	NC	27302	919 304 6161	W	NL	UNA		
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	336 261 3357	W	NL	DEM		
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 2654	W	NL	UNA		
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	336 261 3357	W	NL	UNA		
26	1	ALAMANCE	9079257	A	ACTIVE	AV	VERIFIED	ABBOTT	BRUCE	CLEATON		188 LAKE CAMMA	BURLINGTON	NC	27217	188 LAKE CAMMACK CT	BURLINGTON	NC	27217	336 214 2703	W	NL	REP		
27	1	ALAMANCE	1389300	A	ACTIVE	AV	VERIFIED	ABBOTT	CHERYL	FAULKNER		188 LAKE CAMMA	BURLINGTON	NC	27217	188 LAKE CAMMACK CT	BURLINGTON	NC	27217	336 229 3027	W	NL	REP		
28	1	ALAMANCE	9140392	A	ACTIVE	AV	VERIFIED	ABBOTT	CHRISTOPHE	BRANDON		309 BURLINGTON	GIBSONVILLE	NC	27249	309 BURLINGTON AVE	GIBSONVILLE	NC	27249		W	NL	UNA		
29	1	ALAMANCE	9135711	A	ACTIVE	AV	VERIFIED	ABBOTT	COURTNEY	LOVE		309 BURLINGTON	GIBSONVILLE	NC	27249	309 BURLINGTON AVE	GIBSONVILLE	NC	27249		W	NL	UNA		
30	1	ALAMANCE	9028439	A	ACTIVE	AV	VERIFIED	ABBOTT	DWAYNE	ROGER		2839 LADALE LN	MEBANE	NC	27302	2839 LADALE LN	MEBANE	NC	27302	563 3956	W	NL	UNA		
31	1	ALAMANCE	9090420	A	ACTIVE	AV	VERIFIED	ABBOTT	FRANK	PATRICK		1202 JAMESTOWN	ELON	NC	27244	1202 JAMESTOWNE DR	ELON	NC	27244	336 227 4088	W	UN	UNA		
32	1	ALAMANCE	9079222	A	ACTIVE	AV	VERIFIED	ABBOTT	GLADYS	MARIE MILES		614 TUCKER ST	BURLINGTON	NC	27215	614 TUCKER ST	BURLINGTON	NC	27215	336 570 1418	B	NL	DEM		
33	1	ALAMANCE	9129722	A	ACTIVE	AV	VERIFIED	ABBOTT	HAROLD	GRANT		507 EVERETT ST	BURLINGTON	NC	27215	507 EVERETT ST #320B	BURLINGTON	NC	27215	437 3638	W	NL	REP		
34	1	ALAMANCE	9094352	A	ACTIVE	AV	VERIFIED	ABBOTT	JESSICA	NADINE		2876 CALLOWAY D	MEBANE	NC	27302	2876 CALLOWAY DR	MEBANE	NC	27302	919 304 4661	W	NL	UNA		
35	1	ALAMANCE	9023803	A	ACTIVE	AV	VERIFIED	ABBOTT	JOYCE	HODGES		1934 TUCKER ST	BURLINGTON	NC	27215	1934 TUCKER ST #A	BURLINGTON	NC	27215	336 227 4079	W	NL	DEM		
36	1	ALAMANCE	9084794	R	REMOVED	RS	MOVED FRO	ABBOTT	LATWOIA	BEREA		201 STALEY HALL	ELON	NC	27244	CAMPUS BOX 3039	ELON	NC	27244		B	NL	DEM		
37	1	ALAMANCE	9020357	A	ACTIVE	AV	VERIFIED	ABBOTT	LAWRENCE	ELMER JR		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR	ELON	NC	27244	336 563 4708	W	NL	UNA		
38	1	ALAMANCE	9108338	A	ACTIVE	AV	VERIFIED	ABBOTT	MARIA	LYNETTE		614 TUCKER ST	BURLINGTON	NC	27215	614 TUCKER ST	BURLINGTON	NC	27215	336 570 1418	B	NL	DEM		
39	1	ALAMANCE	9077192	A	ACTIVE	AV	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR	ELON	NC	27244	800 222 7566	W	NL	UNA		
40	1	ALAMANCE	9035500	A	ACTIVE	AV	VERIFIED	ABBOTT	PATTI	BELVIN		1202 JAMESTOWN	ELON	NC	27244	1202 JAMESTOWNE DR	ELON	NC	27244	336 228 0571	W	UN	REP		
41	1	ALAMANCE	9090949	R	REMOVED	RM	REMOVED A	ABBOTT	RACHEL	MARA		103 DANIELEY C	ELON	NC	27244	CAMPUS BOX 3044	ELON	NC	27244	336 278 4012	W	NL	REP		
42	1	ALAMANCE	9135295	A	ACTIVE	AV	VERIFIED	ABBOTT	SUSAN	HANKS		2876 CALLOWAY D	MEBANE	NC	27302	2876 CALLOWAY DR	MEBANE	NC	27302	919 568 8056	W	UN	UNA		
43	1	ALAMANCE	9113731	I	INACTIVE	IU	CONFIRMATI	ABBOTT	TAYLOR	RENEE		406 W LEBANON A	ELON	NC	27244	CAMPUS BOX 3077	ELON	NC	27244		W	UN	REP		



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
106138	1	ALAMANCE	9129972	A	ACTIVE	AV	VERIFIED	ZLUCHOWSK	AARON	MICHAEL		3551 FORESTDALE	BURLINGTON	NC	27215	3551 FORESTDALE DR	#M	BURLINGTON	NC	27215	336 270 6878	W	NL	UNA
106139	1	ALAMANCE	9106623	A	ACTIVE	AV	VERIFIED	ZMIJEASKI	SEAN			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336 376 1987	O	UN	REP
106140	1	ALAMANCE	9112148	A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	DENNIS	AL		4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302		W	UN	DEM
106141	1	ALAMANCE	9094109	I	INACTIVE	IU	CONFIRMATI	ZMIJEWSKI	DENNIS			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336 376 1987	W	UN	DEM
106142	1	ALAMANCE	9128345	A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	KEVIN	ADAM		4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336 380 5768	W	NL	UNA
106143	1	ALAMANCE	9120294	A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	SEAN	CHRISTOPHE		4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302		W	HL	UNA
106144	1	ALAMANCE	9094116	A	ACTIVE	AV	VERIFIED	ZMIJEWSKI	N VIRGINIA	LOURDES		4872 THOM RD	MEBANE	NC	27302	4872 THOM RD		MEBANE	NC	27302	336 376 1987	U	UN	UNA
106145	1	ALAMANCE	9089250	R	REMOVED	RD	DECEASED	ZOCCOLANTIENIS	PIZZOTTI			2502 S NC HWY 119	MEBANE	NC	27302	2502 S NC HWY 119		MEBANE	NC	27302		W	UN	REP
106146	1	ALAMANCE	9083629	R	REMOVED	RD	DECEASED	ZOCCOLANTIRENATO				3141 SHELLY GRAH	GRAHAM	NC	27253	3141 SHELLY GRAHAM DR		GRAHAM	NC	27253	336 227 7168	W	NL	REP
106147	1	ALAMANCE	9083630	A	ACTIVE	AV	VERIFIED	ZOCCOLANTIRITA	MARIE			3141 SHELLY GRAH	GRAHAM	NC	27253	3141 SHELLY GRAHAM DR		GRAHAM	NC	27253	336 227 7168	W	NL	REP
106148	1	ALAMANCE	9100545	I	INACTIVE	IU	CONFIRMATI	ZOGLEMANN	ANGELA	LYNNE		706 HUFFMAN MIL	BURLINGTON	NC	27215	706 HUFFMAN MILL RD	#	BURLINGTON	NC	27215	336 227 1261	W	NL	UNA
106149	1	ALAMANCE	9137285	A	ACTIVE	AV	VERIFIED	ZOLAYVAR	ERIC	WATSON		910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR		BURLINGTON	NC	27215	336 585 0248	O	NL	DEM
106150	1	ALAMANCE	9081869	A	ACTIVE	AV	VERIFIED	ZOLAYVAR	RUPERTO	BENEDICTO		910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR		BURLINGTON	NC	27215	336 585 0248	O	NL	DEM
106151	1	ALAMANCE	9109021	A	ACTIVE	AV	VERIFIED	ZOLAYVAR	STEPHANIE	WATSON		910 COLONIAL DR	BURLINGTON	NC	27215	910 COLONIAL DR		BURLINGTON	NC	27215	336 585 0248	W	NL	UNA
106152	1	ALAMANCE	9108096	A	ACTIVE	AV	VERIFIED	ZOLLARS	EVELYN	NADINE		6830 TOM WOODY	SNOW CAMP	NC	27349	6830 TOM WOODY RD		SNOW CAMFNC		27349	336 376 5754	W	NL	UNA
106153	1	ALAMANCE	9125044	A	ACTIVE	AV	VERIFIED	ZOLLARS	MATHEW	DAVID		6830 TOM WOODY	SNOW CAMP	NC	27349	6830 TOM WOODY RD		SNOW CAMFNC		27349		W	NL	UNA
106154	1	ALAMANCE	9113912	A	ACTIVE	AV	VERIFIED	ZOLLIFFE	ANTONIO	MARK		108 OAKGROVE DR	GRAHAM	NC	27253	108 OAKGROVE DR		GRAHAM	NC	27253	336 260 6673	B	UN	DEM
106155	1	ALAMANCE	9107068	A	ACTIVE	AV	VERIFIED	ZOLLIFFE	VALERIE			108 OAKGROVE DR	GRAHAM	NC	27253	108 OAKGROVE DR		GRAHAM	NC	27253		B	UN	DEM
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	336 578 1157	W	NL	UNA
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		W	NL	UNA
106158	1	ALAMANCE	9104969	I	INACTIVE	IU	CONFIRMATI	ZORNES	MICHELLE	LEE		3117 COMMERCE	BURLINGTON	NC	27215	3117 COMMERCE PL	#L	BURLINGTON	NC	27215	336 675 0520	W	UN	UNA
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		W	NL	DEM
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	570 1633	W	NL	DEM
106161	1	ALAMANCE	9110367	D	DENIED	DU	VERIFICATIO	ZORNES	TINA			801 TROLLINGWO	HAW RIVER	NC	27258	801 TROLLINGWOOD RD		HAW RIVER	NC	27258	336 578 0646	W	UN	UNA
106162	1	ALAMANCE	9132758	A	ACTIVE	AV	VERIFIED	ZORNES	TINA	MARIE		801 TROLLINGWO	HAW RIVER	NC	27258	801 TROLLINGWOOD RD		HAW RIVER	NC	27258	336 420 7630	W	NL	UNA
106163	1	ALAMANCE	9131499	A	ACTIVE	AV	VERIFIED	ZOUFALY	EVE			602 E HAGGARD A	ELON	NC	27244	CAMPUS BOX 8911		ELON	NC	27244		U	UN	UNA
106164	1	ALAMANCE	9124446	A	ACTIVE	AV	VERIFIED	ZSUPPAN	ETELKA	HALASZ		1929 HAW VILLAG	GRAHAM	NC	27253	1929 HAW VILLAGE DR		GRAHAM	NC	27253		W	NL	REP
106165	1	ALAMANCE	9121554	A	ACTIVE	AV	VERIFIED	ZSUPPAN	FERENC			1929 HAW VILLAG	GRAHAM	NC	27253	1929 HAW VILLAGE DR		GRAHAM	NC	27253		W	UN	REP
106166	1	ALAMANCE	9127457	A	ACTIVE	AV	VERIFIED	ZSUPPAN	LEVENTE	FERENC		1929 HAW VILLAG	GRAHAM	NC	27253	1929 HAW VILLAGE DR		GRAHAM	NC	27253	336 376 1365	W	NL	REP
106167	1	ALAMANCE	9131401	A	ACTIVE	AV	VERIFIED	ZUBLER	LINDSAY	BROOKE		3172 CARRIAGE CF	HAW RIVER	NC	27258	3172 CARRIAGE CREEK CT		HAW RIVER	NC	27258		U	UN	UNA
106168	1	ALAMANCE	9081728	A	ACTIVE	AV	VERIFIED	ZUBLER	TAMI	LAJEAN		3172 CARRIAGE CF	HAW RIVER	NC	27258	3172 CARRIAGE CREEK CT		HAW RIVER	NC	27258	336 578 8028	W	NL	UNA
106169	1	ALAMANCE	9089569	A	ACTIVE	AV	VERIFIED	ZUBLER	TIMOTHY	JAMES		3172 CARRIAGE CF	HAW RIVER	NC	27258	3172 CARRIAGE CREEK CT		HAW RIVER	NC	27258		W	UN	UNA
106170	1	ALAMANCE	9070674	A	ACTIVE	AV	VERIFIED	ZUBOV	ALEX			229 ENGLEMAN A	BURLINGTON	NC	27215	229 ENGLEMAN AVE		BURLINGTON	NC	27215	336 437 9776	W	NL	UNA
106171	1	ALAMANCE	9070288	A	ACTIVE	AV	VERIFIED	ZUBOV	LYNN	R		229 ENGLEMAN A	BURLINGTON	NC	27215	229 ENGLEMAN AVE		BURLINGTON	NC	27215	336 437 9776	W	NL	REP
106172	1	ALAMANCE	9008787	A	ACTIVE	AV	VERIFIED	ZUMER	FRANK	EDWARD		801 QUAKER RIDG	MEBANE	NC	27302	801 QUAKER RIDGE RD		MEBANE	NC	27302	919 563 3766	W	UN	UNA
106173	1	ALAMANCE	9008785	A	ACTIVE	AV	VERIFIED	ZUMER	LOUISE	TURNER		801 QUAKER RIDG	MEBANE	NC	27302	801 QUAKER RIDGE RD		MEBANE	NC	27302	919 563 3766	W	NL	DEM
106174	1	ALAMANCE	9141817	A	ACTIVE	AV	VERIFIED	ZUNG	PATRICK	BATE		2604 WOODS LN	GRAHAM	NC	27253	2604 WOODS LN		GRAHAM	NC	27253	919 357 3896	W	NL	DEM
106175	1	ALAMANCE	9119438	A	ACTIVE	AV	VERIFIED	ZUNIGA	JOSE	RAMON SAL		714 ROSS ST	BURLINGTON	NC	27217	714 ROSS ST		BURLINGTON	NC	27217	336 227 3108	O	HL	DEM
106176	1	ALAMANCE	9108610	A	ACTIVE	AV	VERIFIED	ZUNIGA	VANESA	ELIZABETH		512 PIEDMONT W	BURLINGTON	NC	27217	512 PIEDMONT WAY		BURLINGTON	NC	27217	336 270 0181	W	HL	DEM
106177	1	ALAMANCE	9112637	A	ACTIVE	AV	VERIFIED	ZUNIGA	YANET	SALAS		3845 MAE DOUGL	MEBANE	NC	27302	3845 MAE DOUGLAS DR		MEBANE	NC	27302		O	HL	DEM
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	330 310 0151	W	NL	REP
106179	1	ALAMANCE	9141406	A	ACTIVE	AV	VERIFIED	ZUPANCICH	RONALD	JAMES	II	2326 N NC HWY 49	BURLINGTON	NC	27217	2326 N NC HWY 49		BURLINGTON	NC	27217	757 254 3773	W	NL	REP
106180	1	ALAMANCE	9137444	A	ACTIVE	AV	VERIFIED	ZURFACE	ROSSELL	EUGENE		2074 TURNER RD	MEBANE	NC	27302	2074 TURNER RD		MEBANE	NC	27302		W	UN	UNA
106181	1	ALAMANCE	9137444	A	ACTIVE	AV	VERIFIED	ZWIER	ANDREW	MICHAEL		1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMFNC		27349	336 376 8830	W	NL	REP
106182	1	ALAMANCE	9137444	A	ACTIVE	AV	VERIFIED	ZWIER	CHRISTOPHE	ANTHONY		1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMFNC		27349	831 207 9222	W	NL	REP
106183	1	ALAMANCE	9137444	A	ACTIVE	AV	VERIFIED	ZWIER	CHRISTY	ANN		1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMFNC		27349		W	NL	REP
106184	1	ALAMANCE	9099261	A	ACTIVE	AV	VERIFIED	ZWIER	KAREN	JEAN		1497 LONGEST AC	SNOW CAMP	NC	27349	1497 LONGEST ACRES RD		SNOW CAMFNC		27349	831 207 9222	W	NL	REP
106185	1	ALAMANCE	9077804	R	REMOVED	RL	MOVED FROM	ZYLKA	MARC			1210 WILLOW BRC	MEBANE	NC	27302	1210 WILLOW BROOK CT		MEBANE	NC	27302	336 578 8580	W	UN	REP
106186																								

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	AA	AB	AC	
	voter_status	last_name	first_name	midl_name	name	res_street_address	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	birth_place	registr_dt	precinct_abtpr	
1	VERIFIED	AABEL	EVELYN	LARSEN		4430 E GREENSBOF	GRAHAM	NC	27253	4430 E GREENSBORO	CHA	GRAHAM	NC				NL	UNA	F	77	NY	10.01.1984	08N	NC
2	VERIFIED	AARON	CHRISTINA	CASTAGNA								BURLINGTON	NC				UN	UNA	F	36	NY	03/26/1996	03S	NC
3	VERIFIED	AARON	CLAUDIA	HAYDEN								BURLINGTON	NC				NL	UNA	F	68	VA	08/15/1989	124	BU
4	VERIFIED	AARON	JAMES	MICHAEL								SAXAPAHAW	NC				UN	DEM	M	65	MA	03.07.2012	09S	SC
5	VERIFIED	AARON	NATHAN	EDWARD								BURLINGTON	NC				UN	UNA	M	36	NC	10.10.1994	03S	SC
6	VERIFIED	AARON	WILLIE	DALE								BURLINGTON	NC				NL	UNA	M	68	VA	06.06.1990	124	BU
7	VERIFIED	AARONSON	GENA	HOLT								HAW RIVER	NC				NL	REP	F	41	NC	08/18/1998	13	HA
8	VERIFIED	AARONSON	MICHAEL	CHARLES								HAW RIVER	NC				NL	UNA	M	50	WI	01/19/2006	13	HA
9	CONFIRMATI	ABAD	PRISCILLA	MARIE								ELON	NC				HL	UNA	F	23		11.01.2008	35	BO
10	CONFIRMATI	ABADIE	COLLEEN	MIASHEL								GRAHAM	NC				HL	REP	F	46	AZ	09/23/1992	06S	SC
11	VERIFIED	ABADIE	JACK	EDWARD	JR							GRAHAM	NC				NL	UNA	M	27	NC	01/16/2009	06N	NC
12	CONFIRMATI	ABADIE	MYRA	HOLLIFIELD								GRAHAM	NC				NL	UNA	F	61	NC	12.02.2008	06N	NC
13	VERIFIED	ABBAS	FALISA									BURLINGTON	NC				UN	DEM	F	47	NJ	07.03.2012	10N	NC
14	VERIFIED	ABBAS	RAFAT									BURLINGTON	NC				UN	DEM	F	60	NC	03/30/2000	03S	SC
15	VERIFIED	ABBATECOLA	RONALD	JOSEPH	JR							GIBSONVILLE	NC				UN	UNA	M	37	NY	05/14/1996	03W	WI
16	VERIFIED	ABBATECOLA	TRACY	BOONE								GIBSONVILLE	NC				NL	DEM	F	45	NC	10.05.1992	03W	WI
17	CONFIRMATI	ABBETT	DAWN	LEANN								GIBSONVILLE	NC				NL	DEM	F	49	CA	01/30/2004	4	ME
18	VERIFIED	ABBEY	BRENT	DAVID								DR	GRAHAM	NC			NL	REP	M	45	NY	06.06.1991	7	AL
19	VERIFIED	ABBEY	DEMETRA	AINSWORTH								DR	GRAHAM	NC			NL	REP	F	44	SC	01/15/1992	7	AL
20	CONFIRMATI	ABBEY	DOROTHY	ESTELLA								DR	SNOW CAM	NC			NL	REP	F	91	CA	07/26/1990	08S	SC
21	VERIFIED	ABBOTT	AMELIA	BETH								MEBANE	NC				NL	UNA	F	23	NC	10.08.2008	09S	SC
22	VERIFIED	ABBOTT	ANGELA	MORTON								HAW RIVER	NC				NL	DEM	F	39	NC	09.08.2004	09S	SC
23	VERIFIED	ABBOTT	BRENDA	CARMICHAEL								MEBANE	NC				NL	UNA	F	58	NC	04.10.1989	10N	NC
24	VERIFIED	ABBOTT	BRIAN	CHRISTOPHE								HAW RIVER	NC				NL	UNA	M	40	NC	08/17/2007	09S	SC
25	VERIFIED	ABBOTT	BRUCE	CLEATON								CT	BURLINGTON	NC			NL	REP	M	63	NC	10/24/2002	5	FA
26	VERIFIED	ABBOTT	CHERYL	FAULKNER								CT	BURLINGTON	NC			NL	REP	F	59	NC	07/26/1976	5	FA
27	VERIFIED	ABBOTT	CHRISTOPHE	BRANDON								E	GIBSONVILLE	NC			NL	UNA	M	38	NC	11.01.2012	03W	WI
28	VERIFIED	ABBOTT	COURTNEY	LOVE								E	GIBSONVILLE	NC			NL	UNA	F	43		09/21/2012	03W	WI
29	VERIFIED	ABBOTT	DWAYNE	ROGER								MEBANE	NC				UN	UNA	M	53	NC	09/19/1991	09S	SC
30	VERIFIED	ABBOTT	FRANK	PATRICK								DR	ELON	NC			NL	UNA	M	46	NJ	10.05.2004	03N	NC
31	VERIFIED	ABBOTT	GLADYS	MARIE MILES								BURLINGTON	NC				NL	DEM	F	60	NC	11.05.2002	128	BU
32	VERIFIED	ABBOTT	HAROLD	GRANT								OB	BURLINGTON	NC			NL	REP	M	69	NC	03.08.2012	128	BU
33	VERIFIED	ABBOTT	JESSICA	NADINE								MEBANE	NC				NL	UNA	F	29	NC	05.11.2005	09S	SC
34	VERIFIED	ABBOTT	JOYCE	HODGES								BURLINGTON	NC				NL	DEM	F	66	VA	09/24/1990	1210	BU
35	MOVED FROM	ABBOTT	LATWOIA	BEREA								ELON	NC				NL	DEM	F	28	NC	04/20/2004		
36	VERIFIED	ABBOTT	LAWRENCE	ELMER	JR							ELON	NC				NL	UNA	M	62	NC	01.09.1990	03N	NC
37	VERIFIED	ABBOTT	MARIA	LYNETTE								BURLINGTON	NC				NL	DEM	F	27	NC	05.02.2008	128	BU
38	VERIFIED	ABBOTT	NANCY	SKIDMORE								ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
39	VERIFIED	ABBOTT	PATTI	BELVIN								DR	ELON	NC			UN	REP	F	47	NC	10.05.1992	03N	NC
40	REMOVED FROM	ABBOTT	RACHEL	MARA		103 DANIELEY	CENELON	NC	27244	CAMPUS BOX 3044		ELON	NC				NL	REP	F	28	PA	10.08.2004		
41	VERIFIED	ABBOTT	SUSAN	HANKS		2876 CALLOWAY	MEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC				UN	UNA	F	54		09/14/2012	09S	SC
42	CONFIRMATI	ABBOTT	TAYLOR	RENEE		406 W LEBANON	A	NC	27244	CAMPUS BOX 3077		ELON	NC				UN	REP	F	25	WV	10.03.2008	03N	NC
43	CONFIRMATI	ABBOTT	TIFFANY	MURIEL ARLE		144 W CRESCENT	S	GRAHAM	NC	27253	144 W CRESCENT SQUARE	GRAHAM	NC				NL	DEM	F	27	NY	08.05.2009	64	GR
44	CONFIRMATI	ABBOTT	VIRGINIA	SMITH		2820 BLANCHE	DR	BURLINGTON	NC	27215	584 4663	W					NL	REP	F	85	PA	02/22/1988	03S	SC

state mail_zipcode full_phone race_code

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

mail_zipcode full_phone race_code

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

V1 race_code

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
	voter_status	last_name	first_name	mid_name	names	street_address	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	birth_place	registr_dt	precinct_abk	pr
1	VERIFIED	AABEL	EVELYN	LARSEN		4430 E GREENSBOR	GRAHAM	NC	27253	4430 E GREENSBORO-CHA		GRAHAM	NC				NL	UNA	F	77	NY	10.01.1984	08N	NC
2	VERIFIED	AARON	CHRISTINA	CASTAGNA		421 WHITT AVE	BURLINGTON	NC	27215	PO BOX 4177		BURLINGTON	NC				UN	UNA	F	36	NC	03/26/1996	03S	SC
3	VERIFIED	AARON	CLAUDIA	HAYDEN		1013 EDITH ST	BURLINGTON	NC	27215	1013 EDITH ST		BURLINGTON	NC				NL	UNA	F	68	VA	08/15/1989		124 BU
4	VERIFIED	AARON	JAMES	MICHAEL		1647 SAXAPAHAW	GRAHAM	NC	27253	PO BOX 98		SAXAPAHAW	NC				UN	DEM	M	65	MA	03.07.2012	09S	SC
5	VERIFIED	AARON	NATHAN	EDWARD		421 WHITT AVE	BURLINGTON	NC	27215			BURLINGTON	NC				UN	UNA	M	36	NC	10.10.1994	03S	SC
6	VERIFIED	AARON	WILLIE	DALE		1013 EDITH ST	BURLINGTON	NC	27215			BURLINGTON	NC				NL	UNA	M	68	VA	06.06.1990		124 BU
7	VERIFIED	AARONSON	GENA	HOLT		107 TERRYWOOD	HAW RIVER	NC	27215			HAW RIVER	NC				NL	REP	F	41	NC	08/18/1998		13 HA
8	VERIFIED	AARONSON	MICHAEL	CHARLES		107 TERRYWOOD	HAW RIVER	NC	27215			HAW RIVER	NC				NL	UNA	M	50	WI	01/19/2006		13 HA
9	VERIFIED	ABADIE	JACK	EDWARD	JR	612 SIDEVIEW ST	GRAHAM	NC	27215			GRAHAM	NC				NL	UNA	M	27	NC	01/16/2009	06N	NC
10	CONFIRMATION	ABADIE	MYRA	HOLLIFIELD		612 SIDEVIEW ST	GRAHAM	NC	27215			GRAHAM	NC				NL	UNA	F	61	NC	12.02.2008	06N	NC
11	VERIFIED	ABBATECOLA	RONALD	JOSEPH	JR	504 BROOKFIELD	GIBSONVILLE	NC	27215			GIBSONVILLE	NC				UN	UNA	M	37	NY	05/14/1996	03W	WI
12	VERIFIED	ABBATECOLA	TRACY	BOONE		504 BROOKFIELD	GIBSONVILLE	NC	27215			GIBSONVILLE	NC				NL	DEM	F	45	NC	10.05.1992	03W	WI
13	CONFIRMATION	ABBETT	DAWN	LEANN		3900 JOHNS CREEK	GIBSONVILLE	NC	27215			GIBSONVILLE	NC				NL	DEM	F	49	CA	01/30/2004		4 ME
14	VERIFIED	ABBEY	BRENT	DAVID		3304 GOLDEN OAK	GRAHAM	NC	27215			GRAHAM	NC				NL	REP	M	45	NY	06.06.1991		7 AL
15	VERIFIED	ABBEY	DEMETRA	AINSWORTH		3304 GOLDEN OAK	GRAHAM	NC	27215			GRAHAM	NC				NL	REP	F	44	SC	01/15/1992		7 AL
16	CONFIRMATION	ABBEY	DOROTHY	ESTELLA		1029A QUAKENBUSH	SNOW CAMP	NC	27215			SNOW CAMP	NC				NL	REP	F	91	CA	07/26/1990	08S	SC
17	VERIFIED	ABBOTT	AMELIA	BETH		2876 CALLOWAY DR	MEBANE	NC	27215			MEBANE	NC				NL	UNA	F	23	NC	10.08.2008	09S	SC
18	VERIFIED	ABBOTT	ANGELA	MORTON		2006 WINN CREEK	HAW RIVER	NC	27215			HAW RIVER	NC				NL	DEM	F	39	NC	09.08.2004	09S	SC
19	VERIFIED	ABBOTT	BRENDA	CARMICHAEL		611 N THIRD ST	MEBANE	NC	27215			MEBANE	NC				NL	UNA	F	58	NC	04.10.1989	10N	NC
20	VERIFIED	ABBOTT	BRIAN	CHRISTOPHE		2006 WINN CREEK	HAW RIVER	NC	27215			HAW RIVER	NC				NL	UNA	M	40	NC	08/17/2007	09S	SC
21	VERIFIED	ABBOTT	BRUCE	CLEATON		188 LAKE CAMMA	BURLINGTON	NC	27215			BURLINGTON	NC				NL	REP	M	63	NC	10/24/2002		5 FA
22	VERIFIED	ABBOTT	CHERYL	FAULKNER		188 LAKE CAMMA	BURLINGTON	NC	27215			BURLINGTON	NC				NL	REP	F	59	NC	07/26/1976		5 FA
23	VERIFIED	ABBOTT	CHRISTOPHE	BRANDON		309 BURLINGTON	GIBSONVILLE	NC	27215			GIBSONVILLE	NC				NL	UNA	M	38	NC	11.01.2012	03W	WI
24	VERIFIED	ABBOTT	COURTNEY	LOVE		309 BURLINGTON	GIBSONVILLE	NC	27215			GIBSONVILLE	NC				NL	UNA	M	38	NC	11.01.2012	03W	WI
25	VERIFIED	ABBOTT	DWAYNE	ROGER		2839 LADALE LN	MEBANE	NC	27215			MEBANE	NC				NL	UNA	F	43		09/21/2012	03W	WI
26	VERIFIED	ABBOTT	FRANK	PATRICK		1202 JAMESTOWN	ELON	NC	27215			ELON	NC				NL	UNA	M	53	NC	09/19/1991	09S	SC
27	VERIFIED	ABBOTT	HAROLD	GRANT		507 EVERETT ST	BURLINGTON	NC	27215			BURLINGTON	NC				UN	UNA	M	46	NJ	10.05.2004	03N	NC
28	VERIFIED	ABBOTT	JESSICA	NADINE		2876 CALLOWAY DR	MEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC				NL	UNA	F	29	NC	05.11.2005	09S	SC
29	VERIFIED	ABBOTT	JOYCE	HODGES		1934 TUCKER ST	BURLINGTON	NC	27215	1934 TUCKER ST #A		BURLINGTON	NC				NL	DEM	F	66	VA	09/24/1990		1210 BU
30	VERIFIED	ABBOTT	LAWRENCE	ELMER	JR	110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	M	62	NC	01.09.1990	03N	NC
31	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
32	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
33	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
34	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
35	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
36	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
37	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
38	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
39	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
40	VERIFIED	ABBOTT	NANCY	SKIDMORE		110 OAKVIEW DR	ELON	NC	27244	110 OAKVIEW DR		ELON	NC				NL	UNA	F	69	WV	05/17/2002	03N	NC
41	REMOVED	58	VERIFIED	ABELE		27244 CAMPUS BOX 3044	ELON	NC	27244	2966 S CHURCH APT 110		BURLINGTON	NC				NL	UNA	F	53	CA	02/28/2012		4 ME
42	REMOVED	58	VERIFIED	ABELE		27302 2876 CALLOWAY DR	MEBANE	NC	27302	2876 CALLOWAY DR		MEBANE	NC				UN	UNA	F	54		09/14/2012	09S	SC
43	CONFIRMATION	59	VERIFIED	ABELE		27244 CAMPUS BOX 3077	ELON	NC	27244	CAMPUS BOX 3077		ELON	NC				UN	REP	F	25	WV	10.03.2008	03N	NC
44	CONFIRMATION	59	VERIFIED	ABELE		27215 2820 BLANCHE DR	BURLINGTON	NC	27215	2820 BLANCHE DR		BURLINGTON	NC				NL	REP	F	85	PA	02/22/1988	03S	SC
45	CONFIRMATION	59	VERIFIED	ABELE		27215 605 ISLEY PL #C	BURLINGTON	NC	27215	605 ISLEY PL #C		BURLINGTON	NC				NL	DEM	U	41		05.02.2008	12W	WI
46	VERIFIED	60	VERIFIED	ABELE		27215 1841 DUNBAR PL	BURLINGTON	NC	27215	1841 DUNBAR PL		BURLINGTON	NC				NL	UNA	F	52	DC	11.10.2011	03S	SC
47	VERIFIED	60	VERIFIED	ABELE		27253 110 BURTON ST	GRAHAM	NC	27253	110 BURTON ST		GRAHAM	NC				NL	UNA	M	56	MD	05.01.1989	06E	EA
48	VERIFIED	61	VERIFIED	ABELE		27244 2966 S CHURCH APT 110	BURLINGTON	NC	27244	2966 S CHURCH APT 110		BURLINGTON	NC				NL	UNA	F	53	CA	02/28/2012		4 ME
49	VERIFIED	61	VERIFIED	ABELE		27215 3922 MUHLENBERG CT	BURLINGTON	NC	27215	3922 MUHLENBERG CT		BURLINGTON	NC				NL	DEM	M	101	IA	09/26/1996		
50	VERIFIED	61	VERIFIED	ABELE		27215 3905 CALVIN CT	BURLINGTON	NC	27215	3905 CALVIN CT		BURLINGTON	NC				NL	DEM	F	87	NC	10.02.1998		
51	VERIFIED	61	VERIFIED	ABELE		27215 1320 HAWTHORNE LN	BURLINGTON	NC	27215	1320 HAWTHORNE LN		BURLINGTON	NC				NL	REP	M	23	UT	05.08.2012		124 BU
52	VERIFIED	61	VERIFIED	ABELE		27215 1320 HAWTHORNE LN	BURLINGTON	NC	27215	1320 HAWTHORNE LN		BURLINGTON	NC				NL	REP	M	23	UT	05.08.2012		124 BU
53	VERIFIED	61	VERIFIED	ABELE		27215 1320 HAWTHORNE LN	BURLINGTON	NC	27215	1320 HAWTHORNE LN		BURLINGTON	NC				NL	REP	M	23	UT	05.08.2012		124 BU
54	VERIFIED	61	VERIFIED	ABELE		27215 1320 HAWTHORNE LN	BURLINGTON	NC	27215	1320 HAWTHORNE LN		BURLINGTON	NC				NL	REP	M	23	UT	05.08.2012		124 BU
55	VERIFIED	61	VERIFIED	ABELE		27215 1320 HAWTHORNE LN	BURLINGTON	NC	27215	1320 HAWTHORNE LN		BURLINGTON	NC				NL	REP	M	23	UT	05.08.2012		124 BU
56	VERIFIED	61	VERIFIED	ABELE		27215 1320 HAWTHORNE LN	BURLINGTON	NC	27215	1320 HAWTHORNE LN		BURLINGTON	NC				NL	REP	M	23	UT	05.08.2012		124 BU
57	VERIFIED	61	VERIFIED	ABELE		27215 1320 HAWTHORNE LN	BURLINGTON	NC	27215	1320 HAWTHORNE LN		BURLINGTON	NC				NL	REP	M	23	UT	05.08.2012		124 BU

Textfilter

Von A bis Z sortieren

Von Z bis A sortieren

Nach Farbe sortieren

Filter löschen aus "race_code"

Nach Farbe filtern

Suchen

(Alles auswählen)

A

B

I

M

O

U

W

OK Abbrechen

Textfilter

Von A bis Z sortieren

Von Z bis A sortieren

Nach Farbe sortieren

Filter löschen aus "race_code"

Nach Farbe filtern

Suchen

(Alles auswählen)

A

B

I

M

O

U

W

OK Abbrechen

Microsoft Excel interface showing the ribbon with tabs: Datei, Start, Einfügen, Seitenlayout, Formeln, Daten, Überprüfen, Ansicht, Add-Ins. The ribbon includes options for font (Calibri, 11), alignment (Zeilenumbruch, Verbinden und zentrieren), number formatting (Standard, % 000, .00, .00), conditional formatting (Standard, Gut, Neutral, Schlecht, Ausgabe, Berechnung, Eingabe, Erklärender...), and cell operations (Einfügen, Löschen, Format). The status bar at the bottom shows 'V106177' and 'O'.

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
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	VA	06/29/1998	13
105911	VERIFIED	YUAN	YUEXI			1208 RASPBERRY F	GRAHAM	NC	27253	1208 RASPBERRY RUN	GRAHAM	NC		27253		A	UN	UNA	F		63	OC	10.10.2008	06S
105912	OVERSEAS CI		DIANE			552 MEADOWOOD	BURLINGTON	NC	27215	552 MEADOWOOD DR	BURLINGTON	NC		27215		U	UN	DEM	F		24		09/28/2012	03C
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	OC	10.01.2008	12E
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		05.04.2012	35
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	PA	03.12.2010	10S
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	NY	08.03.2010	103
105968	VERIFIED	ZAMARRIPA	JUANA	MARIA		3668 ALTAMAHAVE	ELON	NC	27244	PO BOX 791	HAW RIVER	NC		27258	336 524 9599	O	HL	DEM	F		89	OC	10.08.2003	4
105971	VERIFIED	ZAMORA																						
105972	VERIFIED	ZAMORA																						
105973	VERIFIED	ZAMORA																						
105974	VERIFIED	ZAMORA																						
105977	VERIFIED	ZAMPARELLI																						
105978	VERIFIED	ZAMUDIO																						
106000	VERIFIED	ZAPATA																						
106001	VERIFIED	ZAPATA																						
106002	VERIFIED	ZAPIEN																						
106017	CONFIRMATI	ZAVALA																						
106018	VERIFIED	ZAVALA																						
106019	VERIFIED	ZAVALA																						
106023	CONFIRMATI	ZAYAS																						
106024	UNAVAILABL	ZAYAS																						
106027	VERIFIED	ZAYAS																						
106034	VERIFIED	ZEIGLER																						
106036	VERIFIED	ZEIGLER																						
106037	VERIFIED	ZEIGLER																						
106038	VERIFIED	ZELAYA	MARLENY	ARELY		611 CHANDLER AV	BURLINGTON	NC	27217	611 CHANDLER AVE	BURLINGTON	NC		27217	336 512 7292	U	HL	DEM	F		29	MD	11.04.2008	12N
106045	VERIFIED	ZEMPEL	GINA	LOUISA		603 E DAVIS ST	BURLINGTON	NC	27215	603 E DAVIS ST	BURLINGTON	NC		27215	262 672 0902	O	HL	UNA	F		32	MI	06/23/2010	128
106057	REMOVED A	ZHAGUI	CECILIA	ISABEL		2732 WESTCHESTE	BURLINGTON	NC	27217	2732 WESTCHESTER DR	BURLINGTON	NC		27217	336 228 3620	U	HL	DEM	F		35	OC	08.02.2004	
106058	VERIFIED	ZHENG	JICHUN			1914 HILLDALE DR	BURLINGTON	NC	27215	1914 HILLDALE DR	BURLINGTON	NC		27215	917 977 1073	U	UN	UNA	U		25		04/30/2012	1210
106059	CONFIRMATI	ZHENG	KELLY			95 DENISE DR #B	BURLINGTON	NC	27215	95 DENISE DR #B	BURLINGTON	NC		27215		A	UN	UNA	F		26	OC	09/16/2008	1210
106060	VERIFIED	ZHENG	SAIFEN			515 BURLINGTON	GIBSONVILLE	NC	27249	515 BURLINGTON AVE	GIBSONVILLE	NC		27249		A	UN	UNA	F		33	OC	12.12.2011	03W
106066	VERIFIED	ZIEGLER	BRANDON	GERALD		2459 MORNINGSI	BURLINGTON	NC	27217	2459 MORNINGSIDE DR	BURLINGTON	NC		27217	336 212 7747	B	NL	REP	M		22	NJ	08/21/2008	63
106072	VERIFIED	ZIEGLER	VIENGSA	MINDY		2459 MORNINGSI	BURLINGTON	NC	27217	2459 MORNINGSIDE DR	BURLINGTON	NC		27217		U	UN	UNA	F		24		11.06.2012	63
106079	VERIFIED	ZIENTY	SCOTT	CARL		511 W HAGGARD	ELON	NC	27244	511 W HAGGARD AVE	ELON	NC		27244		O	NL	UNA	M		23	MO	03/13/2012	03N
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		10/29/2012	128
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	OC	09/26/2008	06W
106136	VERIFIED	ZITO	ALEX	R		622 S WILLIAMSON	ELON	NC	27244	622 S WILLIAMSON AVE	ELON	NC		27244	617 276 2096	O	UN	UNA	U		22		05.04.2012	03N
106139	VERIFIED	ZMIJEASKI	SEAN			4872 THOM RD	MEBANE	NC	27302	4872 THOM RD	MEBANE	NC		27302	336 376 1987	O	UN	REP	M		23		04.09.2008	09S
106144	VERIFIED	ZMIJEWSKI	N VIRGINIA	LOURDES		4872 THOM RD	MEBANE	NC	27302	4872 THOM RD	MEBANE	NC		27302	336 376 1987	U	UN	UNA	F		45	PR	03/24/2005	09S
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	NC	10.08.2012	03S
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	OC	08/14/2003	03S
106154	VERIFIED	ZOLLI	COFFEY	ANTONIO	MARK	108 OAKGROVE D	GRAHAM	NC	27253	108 OAKGROVE DR	GRAHAM	NC		27253	336 260 6673	B	UN	DEM	M		53	MD	10.07.2008	06E

Google search results for "Excel histogram". The search bar shows "Excel histogram" and a search button. Below the search bar are navigation tabs: Web (selected), Bilder, Maps, Shopping, Apps, Mehr, and Suchoptionen. The results show approximately 1,260,000 results in 0.23 seconds. A suggestion "Meinten Sie: [Excel histogramm](#)" is displayed. The top result is "Create a histogram - Excel - Office.com" with the URL "office.microsoft.com/.../excel.../create-a-histogram-...". A link to "Diese Seite übersetzen" is also visible.

9115545	A	ACTIVE	AV	VERIFIED	HAWKINS	DERRICK	JEROME	106 TADWORTH CT	MEBANE	NC	27302	106 TADWORTH CT	MEBANE
9060012	A	ACTIVE	AV	VERIFIED	HAWKINS	DIANA	LEE	424 MEADOWOOD	BURLINGTON	NC	27215	424 MEADOWOOD DR	BURLINGTON
9118697	A	ACTIVE	AV	VERIFIED	HAWKINS	DOMINIQUE	DEVON	8 SHERRY DR	BURLINGTON	NC	27215	8 SHERRY DR	BURLINGTON
2848800	R	REMOVED	RD	DECEASED	HAWKINS	DONALD	LEE	2847 SNUG HARBO	BURLINGTON	NC	27217	2847 SNUG HARBOR RD	BURLINGTON
9025486	I	INACTIVE	IN	CONFIRMATI	HAWKINS	DONNA	KAYE	859 ROSS ST	BURLINGTON	NC	27217	859 ROSS ST	BURLINGTON
9134349	A							779 WOODY DR	GRAHAM	NC	27253	779 WOODY DR	GRAHAM
9081107	A	ACTIVE	AV	VERIFIED	HAWKINS	ELIZABETH	MOSHER	1720 OLD ST MARK	BURLINGTON	NC	27215	1720 OLD ST MARK'S CHURCH	BURLINGTON
9110146	A							5828 ANDOVER DR	GRAHAM	NC	27253	5828 ANDOVER DR	GRAHAM
9018277	A	ACTIVE	AV	VERIFIED	HAWKINS	ERIC	MICHAEL	2428 US HWY 70	MEBANE	NC	27302	2428 US HWY 70	MEBANE
9010269	A							307 N SEVENTH ST	MEBANE	NC	27302	307 N SEVENTH ST	MEBANE
9072769	A	ACTIVE	AV	VERIFIED	HAWKINS	HEATHER	ANN	7439 COBLE MILL F	SNOW CAMP	NC	27349	7439 COBLE MILL RD	SNOW CAMP
2850000	A							2912 MARLBOROU	BURLINGTON	NC	27215	2912 MARLBOROUGH RD	BURLINGTON
9139873	A	ACTIVE	AV	VERIFIED	HAWKINS	ISAIAH	FORRIESH	726 DAILEY ST	BURLINGTON	NC	27217	726 DAILEY ST	BURLINGTON
9102693	A							2111 FAIRWIND DR	GRAHAM	NC	27253	2111 FAIRWIND DR	GRAHAM
2850100	A	ACTIVE	AV	VERIFIED	HAWKINS	JACQUELINE	ISLEY	859 ROSS ST	BURLINGTON	NC	27217	859 ROSS ST	BURLINGTON
9131359	A							203 EDWARD CT	MEBANE	NC	27302	203 EDWARD CT	MEBANE
2850401	A	ACTIVE	AV	VERIFIED	HAWKINS	JAMES	HUBERT JR	1107 SOUTHERN H	BURLINGTON	NC	27215	1107 SOUTHERN HIGH SCHOOL	BURLINGTON
9034990	A							30 GRANITE CT	GIBSONVILLE	NC	27249	30 GRANITE CT	GIBSONVILLE
9102435	A	ACTIVE	AV	VERIFIED	HAWKINS	JAMES	EDWARD	1107 SOUTHERN H	BURLINGTON	NC	27215	1107 SOUTHERN HIGH SCHOOL	BURLINGTON
9083219	A							109 SLADE ST	ELON	NC	27244	109 SLADE ST	ELON
9013096	A	ACTIVE	AV	VERIFIED	HAWKINS	JERRY	MICHAEL	2730 BELLEMONT-	BURLINGTON	NC	27215	2730 BELLEMONT-ALAMA	BURLINGTON
9110147	A							5828 ANDOVER DR	GRAHAM	NC	27253	5828 ANDOVER DR	GRAHAM
9119019	A	ACTIVE	AV	VERIFIED	HAWKINS	JOHN	MATSON	3314 N NC HWY 62	BURLINGTON	NC	27217	3314 N NC HWY 62	BURLINGTON
2851100	A							613 N FOURTH ST	MEBANE	NC	27302	613 N FOURTH ST	MEBANE
9029983	A	ACTIVE	AV	VERIFIED	HAWKINS	JOHN	THOMAS	232 MONROE LN	ELON	NC	27244	232 MONROE LN	ELON
9001801	R	REMOVED	RL	MOVED FROM	HAWKINS	JOHN	DANIEL	862 ROSS ST	BURLINGTON	NC	27217	862 ROSS ST	BURLINGTON
9008655	R	REMOVED	RL	MOVED FROM	HAWKINS	JOHNNY	DEVON	862 ROSS ST	BURLINGTON	NC	27217	862 ROSS ST	BURLINGTON
9109154	I	INACTIVE	IN	CONFIRMATI	HAWKINS	JUSTIN	ANDREW	2111 FAIRWIND DR	GRAHAM	NC	27253	2111 FAIRWIND DR	GRAHAM
9063027	A	ACTIVE	AV	VERIFIED	HAWKINS	KAREN	RENEE	1717 DURHAM ST	BURLINGTON	NC	27217	1717 DURHAM ST #61	BURLINGTON
9014773	A	ACTIVE	AV	VERIFIED	HAWKINS	KAREN	COOK	716 S WILLIAMSON	ELON	NC	27244	716 S WILLIAMSON AVE	ELON
2851300	A	ACTIVE	AV	VERIFIED	HAWKINS	KATHY	ROGERS	485 PARKVIEW DR	BURLINGTON	NC	27215	485 PARKVIEW DR	BURLINGTON
9115548	A	ACTIVE	AV	VERIFIED	HAWKINS	KATHY		1107 SOUTHERN H	BURLINGTON	NC	27215	1107 SOUTHERN HIGH SCHOOL	BURLINGTON
9059505	D	DENIED	DI	UNAVAILABL	HAWKINS	KATRINA	NICOLE	2430 MARION CT	BURLINGTON	NC	27215	2430 MARION CT	BURLINGTON
9135064	A	ACTIVE	AV	VERIFIED	HAWKINS	KENNETH	WESLEY	114 W SEBASTIAN	MEBANE	NC	27302	114 W SEBASTIAN CT	MEBANE

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 last name

- DoB and first name

What are frequent patterns in a column?

- ddddd

- dd aaaa St

Felix Naumann
Data Profiling
Summer 2017

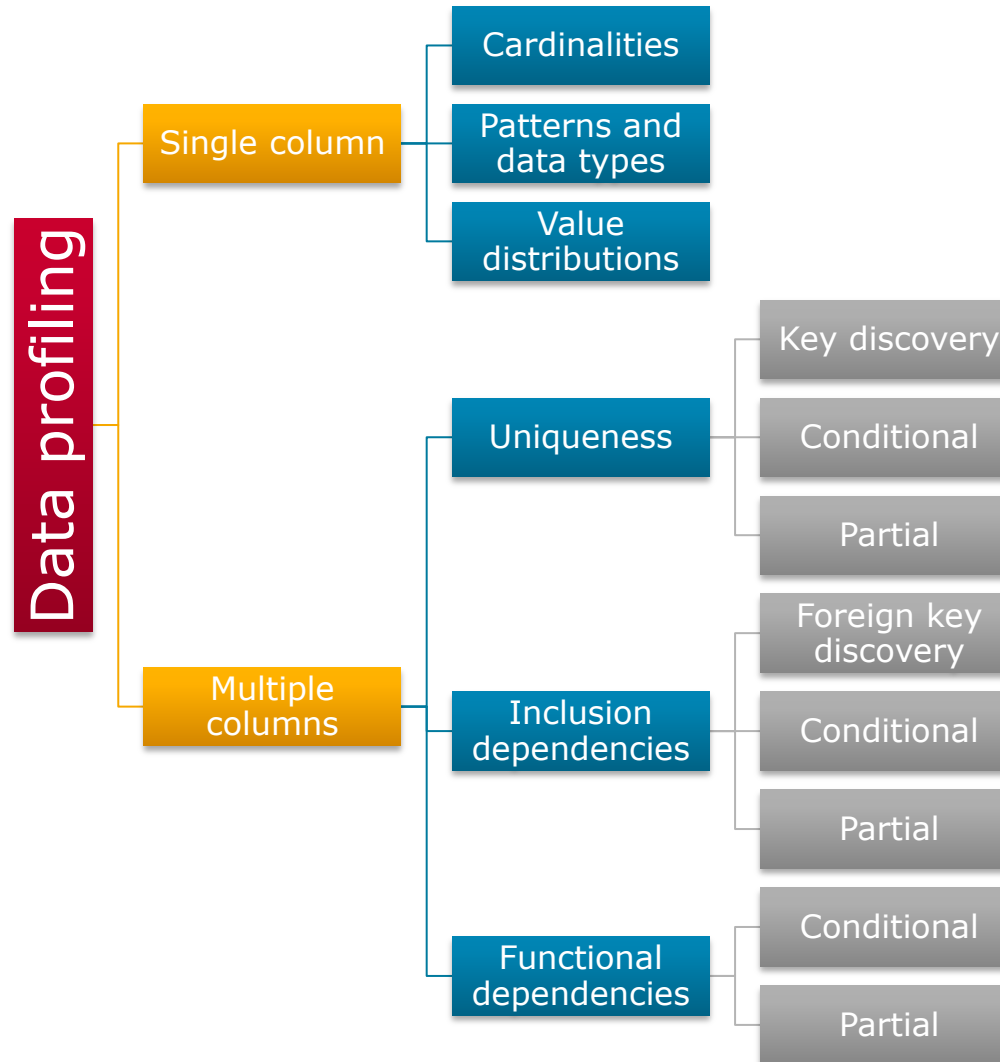
Definition Data Profiling

- 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



Felix Naumann
Data Profiling
Summer 2017

Classification of Traditional Profiling Tasks



Single-column vs. multi-column

- Single column profiling
 - Most basic form of data profiling
 - Assumption: All values are of same type
 - Assumption: All values have some common properties
 - That are to be discovered
 - Often part of the basic statistics gathered by DBMS
 - Complexity: Number of values/rows

- Multicolumn profiling
 - Discover joint properties
 - Discover dependencies
 - Complexity: Number of columns and number of values

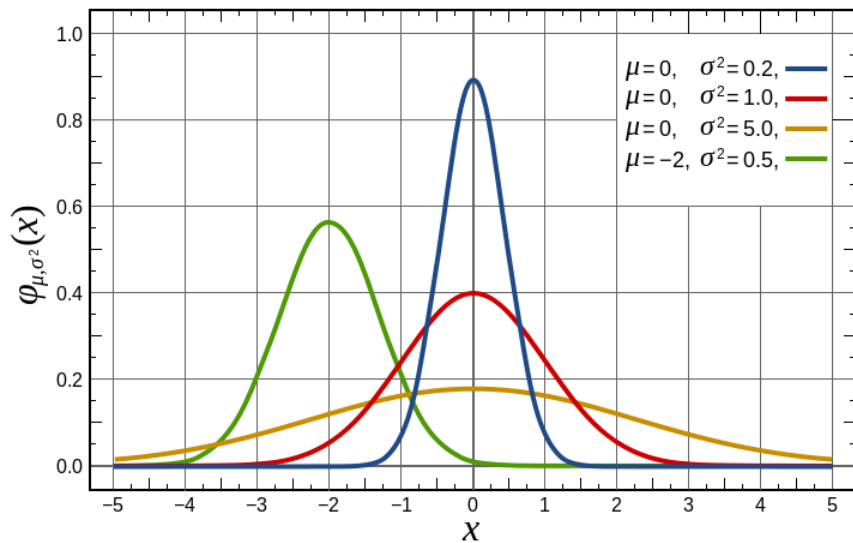
Cardinalities

- Number of values
- Number of distinct values
- Number of NULLs
- MIN and MAX value

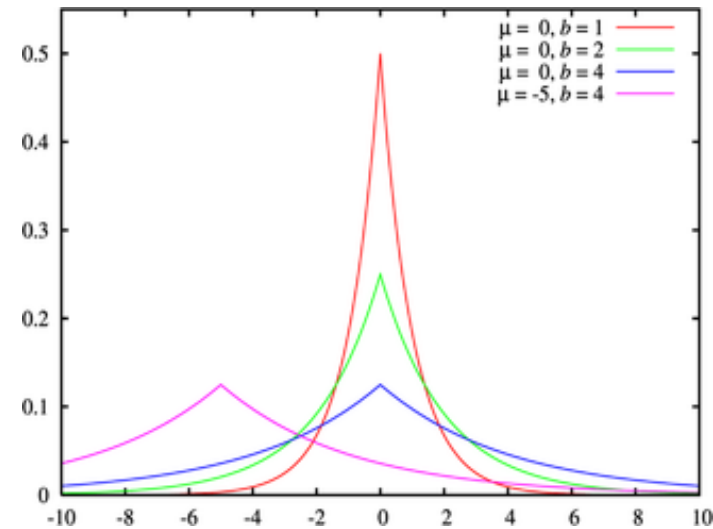
- Useful for
 - Query optimization
 - Categorization of attribute
 - Relevance of attribute

Numeric distributions

- Probability distribution for numeric values
- Detect whether data follows some well-known distribution
 - Determine that distribution function for data values
- If no specific/useful function detectable: histograms



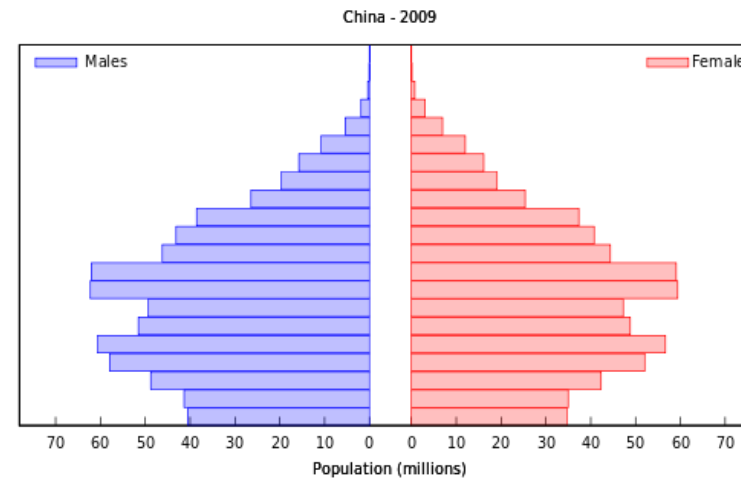
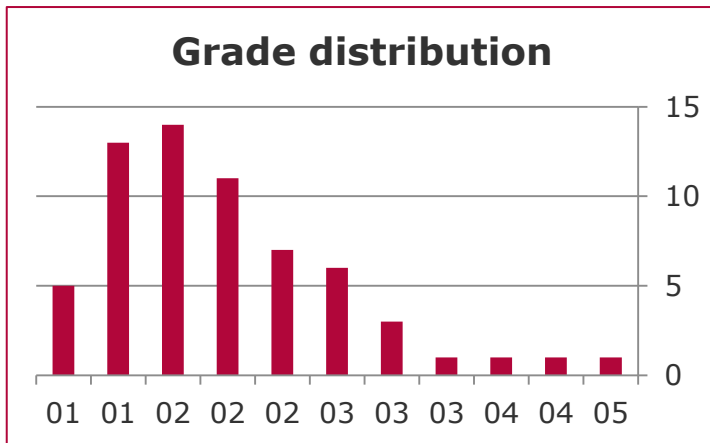
Normal distributions



Laplace distributions

Histograms

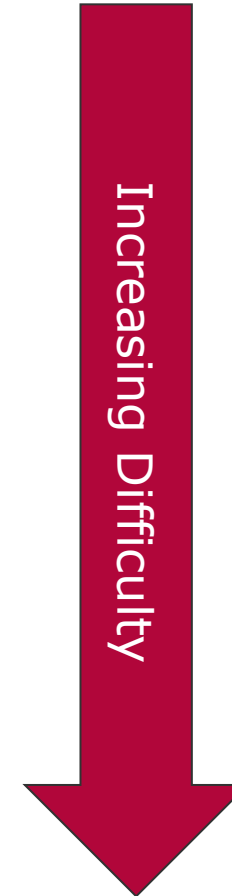
- Determine (and display) value frequencies for value intervals
 - Or for individual values
 - Estimation of probability distribution for continuous variables



- Useful for
 - Query optimization
 - Outlier detection
 - Visualizing distribution

Data types and value patterns

- 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 and keys

- Unique column
 - Only unique values
- Unique column combination
 - Only unique value combinations
 - Minimality: No subset is unique
- (Primary) key candidate
 - No null values
 - Uniqueness and non-null in one instance does not imply key: Only human can specify keys (and foreign keys)
- Meaning of NULL values?
- Useful for
 - Schema design, data integration, indexing, optimization
 - Inverse: non-uniques are duplicates

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 combinations in A_1, \dots, A_i are also present in B_1, \dots, B_i
- Prerequisite for foreign key
 - Used across relations
 - Use across databases
 - But again: Discovery on a given instance, only user can specify for schema



Game of Dependencies

Felix Naumann
Data Profiling
Summer 2017

Functional Dependencies

Person	Lineage	Hair	Religion
			New gods
			New Gods
			Old gods
			New gods
			Old gods

Some Functional Dependencies:

1. Person → Lineage
2. Person → Hair
3. Person → Religion
4. Lineage → Hair
5. Religion, Hair → Lineage
6. ...

Ned Stark: „#4 looks like a reasonable quality constraint“

Ned Stark: „I believe Joffrey violates my database constraint.“

Felix Naumann
Data Profiling
Summer 2017

Uses and Algorithms for FDs

- Schema design
 - Normalization
 - Keys
- Data cleansing
- Query optimization
- Key discovery

Naïve discovery approach

- For each column combination X
 - For each pair of tuples (t_1, t_2)
 - If $t_1[X \setminus A] = t_2[X \setminus A]$ and $t_1[A] \neq t_2[A]$: Break
- Exponential in number of attributes times number of rows squared

Partial dependencies

- INDs and FDs that do not perfectly hold
 - For all but 10 of the tuples
 - Only for 80% of the tuples
 - Only for 1% of the tuples
- Also for patterns, types, uniques, and other constraints
- Useful for
 - Data cleansing

Conditional dependencies

- Given a partial IND or FD: For **which** part do the 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

Data profiling vs. data mining

- Data profiling gathers technical metadata to support data management
- Data mining and data analytics discovers non-obvious results to support business management

- Data profiling results: information about columns and column sets
- Data mining results: information about rows or row sets
 - clustering, summarization, association rules, ...

- Rahm and Do, 2000
 - Profiling: Individual attributes
 - Mining: Multiple attributes

Overview

1. Profiling tasks
- 2. Profiling tools**
3. Visualization
4. Next generation profiling
5. Profiling challenges



Felix Naumann
Data Profiling
Summer 2017


Data profiling tools and algorithms

- **IBM InfoSphere Information Analyzer**
 - <http://www.ibm.com/software/data/infosphere/information-analyzer/>
- Oracle Enterprise Data Quality
 - <http://www.oracle.com/us/products/middleware/data-integration/enterprise-data-quality/overview/index.html>
- Talend Data Quality
 - <http://www.talend.com/products/data-quality>
- Ataccama DQ Analyzer
 - <http://www.ataccama.com/en/products/dq-analyzer.html>
- **SAP BusinessObjects Data Insight and SAP BusinessObjects Information Steward**
 - <http://www.sap.com/germany/solutions/sapbusinessobjects/large/eim/datainsight/index.epx>
 - <http://www.sap.com/germany/solutions/sapbusinessobjects/large/eim/information-steward/index.epx>
- Informatica Data Explorer
 - <http://www.informatica.com/us/products/data-quality/data-explorer/>
- Microsoft SQL Server Integration Services Data Profiling Task and Viewer
 - <http://msdn.microsoft.com/en-us/library/bb895310.aspx>
- Trillium Software Data Profiling
 - <http://www.trilliumsoftware.com/home/products/data-profiling.aspx>
- CloverETL Data Profiler
 - <http://www.cloveretl.com/products/profiler>
- OpenRefine
 - <http://www.openrefine.org>
- and many more...

Often packaged with
data quality / data
cleansing software

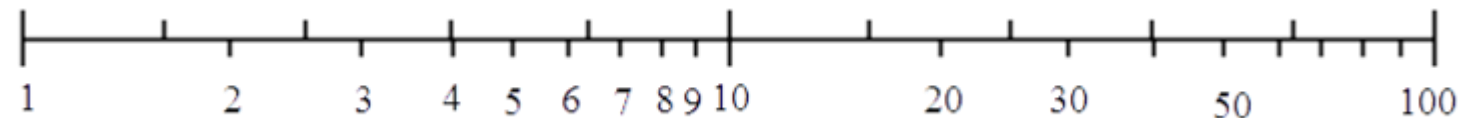
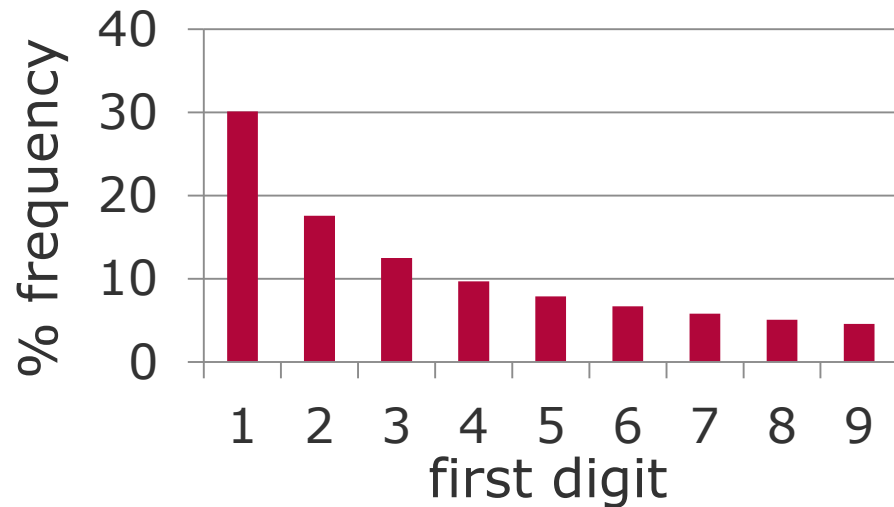
Felix Naumann
Data Profiling
Summer 2017

Very long feature lists

- 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

Benford Law Frequency , a.k.a. “first digit law”

- 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



Picking a random x position uniformly on this number line, roughly 30% of the time the first digit of the number will be 1.

Felix Naumann
Data Profiling
Summer 2017

Explanation for Benford's Law

■ Exponential growth

- Year 1: 1-100
- Year 2: 101-200
- Year 3: 201-400
- Year 4: 401-800
- Year 5: 801-1600
- Year 6: 3200

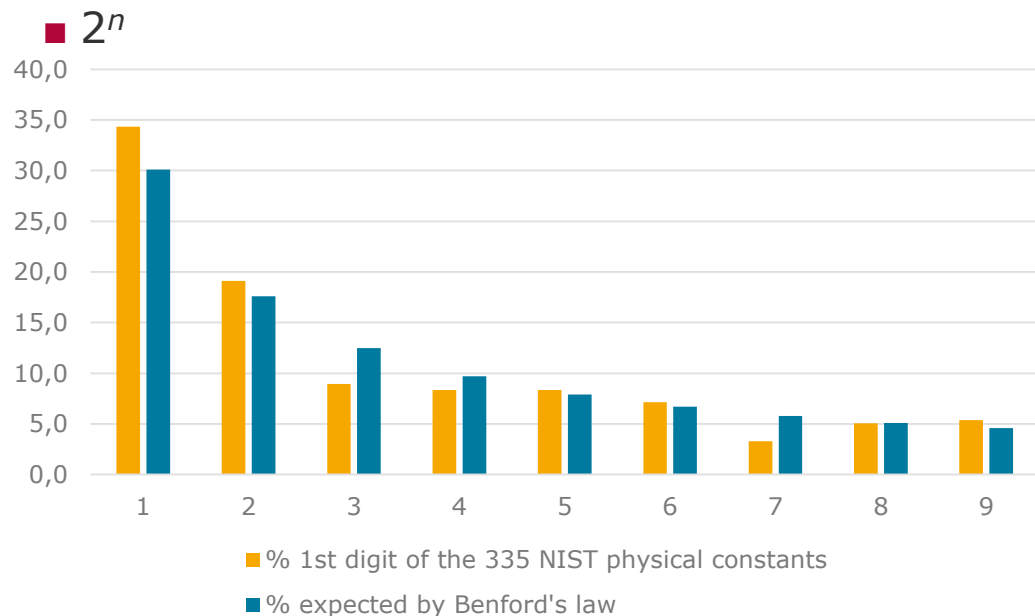
■ Function: $2^{(year-1)} \cdot 100$

■ First digit

- **1**: Some of year 1, all of year 2
- **2**: 7 months in year 3
 - $(\log_2(300/100)+1 - 2) * 12$
- **3**: remaining 5 months
- **4 – 9**: few months each
- **1**: again after under 4 months in year 5 for 1 whole year
 - $(\log_2(10)+1 - 4) * 12$

Examples for Benford's Law

- Surface areas of 335 rivers
- Sizes of 3259 US populations
- 1800 molecular weights
- 5000 entries from a mathematical handbook
- 308 numbers 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	
	Count	%
1	26	43.3%
2	7	11.7%
3	9	15.0%
4	6	10.0%
5	4	6.7%
6	1	1.7%
7	2	3.3%
8	5	8.3%
9	0	0.0%

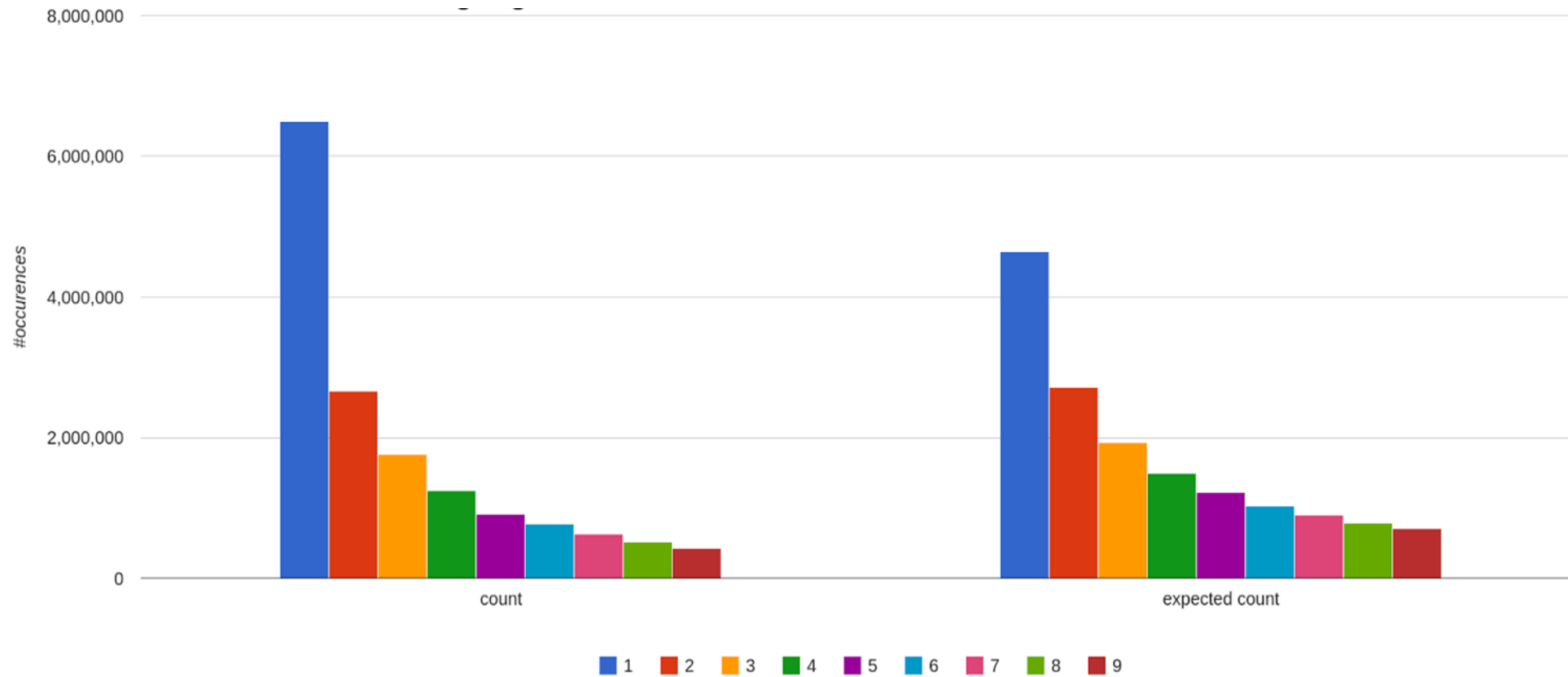
In Benford's law
30.1%
17.6%
12.5%
9.7%
7.9%
6.7%
5.8%
5.1%
4.6%

http://en.wikipedia.org/wiki/List_of_tallest_buildings_and_structures_in_the_world#Tallest_structure_by_category



Felix Naumann
Data Profiling
Summer 2017

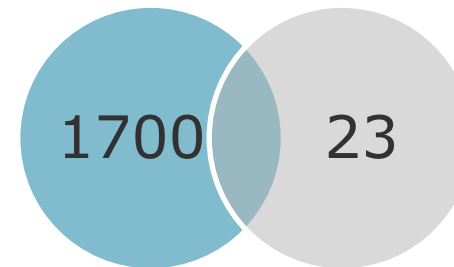
Occurrences of leading digits in WikiTable numbers



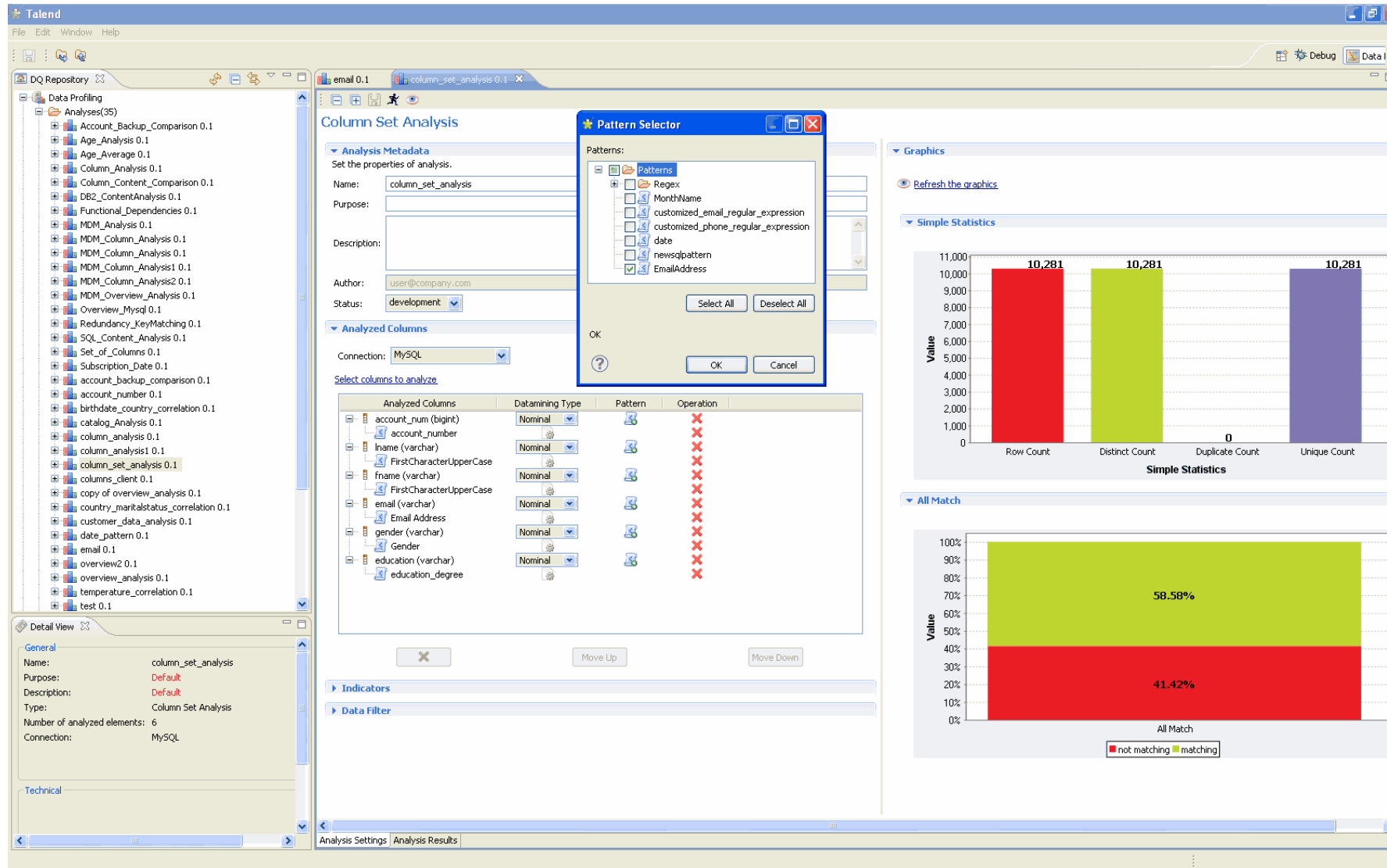
Felix Naumann
Data Profiling
Summer 2017

„Checking“ vs. „discovery“

- An anonymous tool:
 - “Cross-table analysis enables you to identify matching or orphaned records between two tables, based on a fully-customizable join condition and optional filter on either table.”
- Corresponds to
 - ```
SELECT COUNT(*)
FROM A, B
WHERE A.x = B.y
AND cond
```
  - Plus some fancy visualization:



# Screenshots from Talend Data Quality



The screenshot displays the Talend Data Quality interface. On the left, a tree view shows the 'Data Profiling' repository with various analyses. The main window is titled 'Column Set Analysis' and shows configuration for an analysis named 'column\_set\_analysis'. A 'Pattern Selector' dialog is open, showing a list of patterns with 'EmailAddress' selected. Below the configuration, a table lists analyzed columns with their datatypes and patterns. The results section includes two charts: 'Simple Statistics' and 'All Match'.

**Simple Statistics**

| Statistic       | Value  |
|-----------------|--------|
| Row Count       | 10,281 |
| Distinct Count  | 10,281 |
| Duplicate Count | 0      |
| Unique Count    | 10,281 |

**All Match**

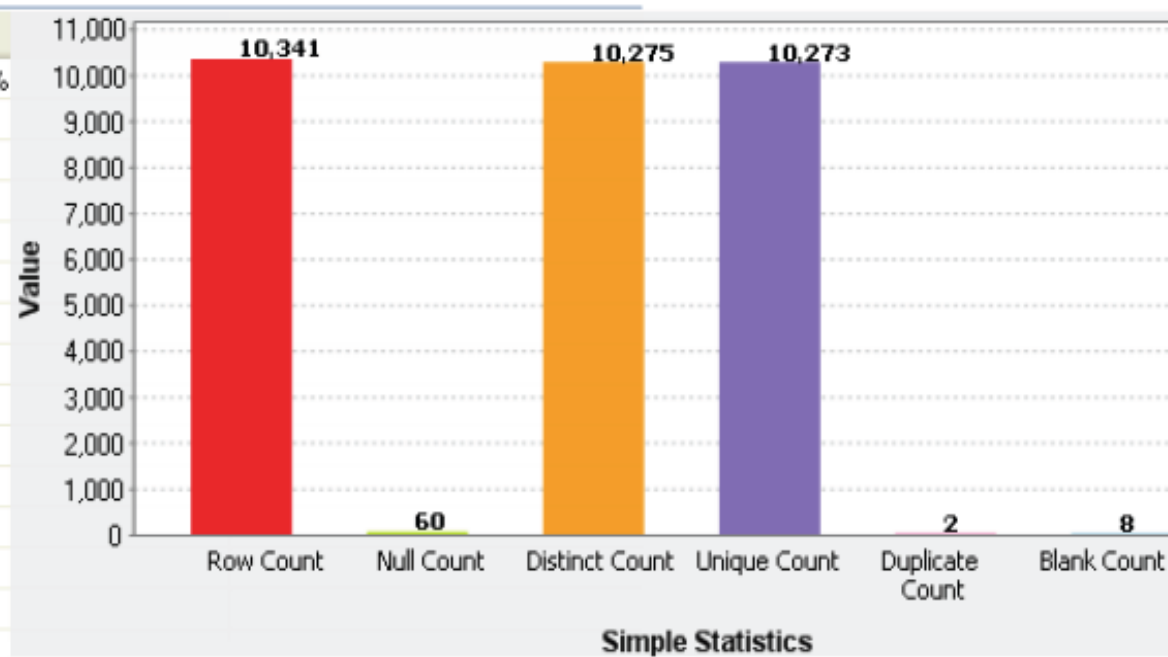
| Category     | Percentage |
|--------------|------------|
| not matching | 41.42%     |
| matching     | 58.58%     |

Felix Naumann  
Data Profiling  
Summer 2017

▼ Column:customer.email

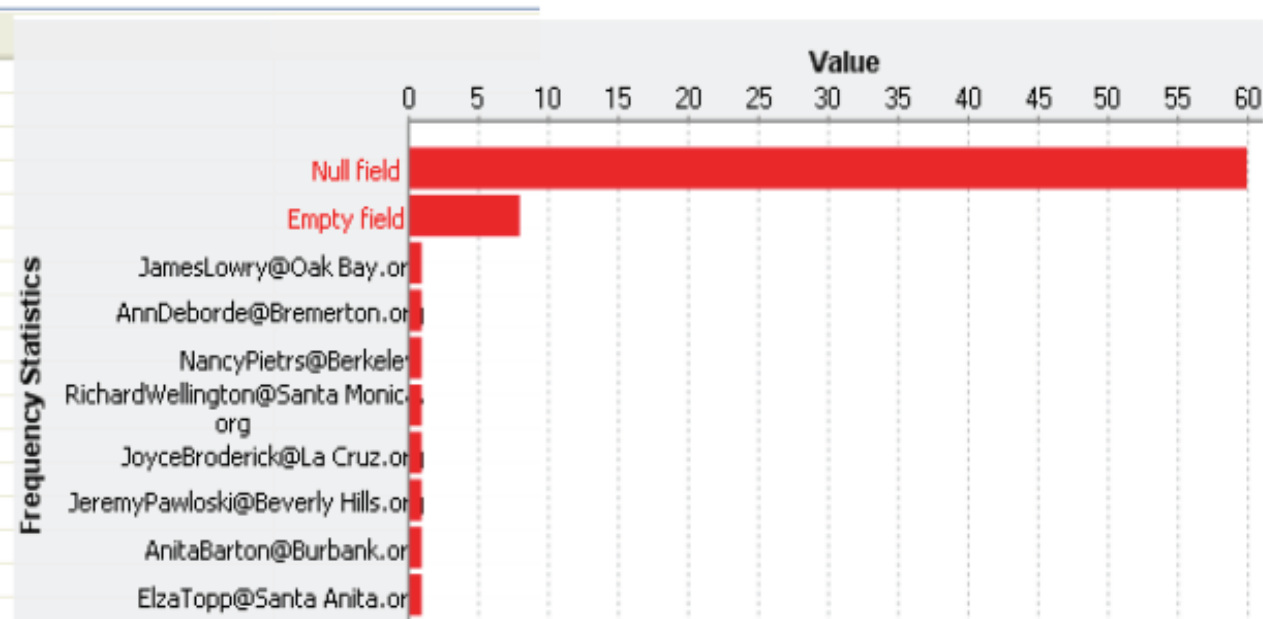
▼ Simple Statistics

| Label           | Count    | %       |
|-----------------|----------|---------|
| Row Count       | 10341.00 | 100.00% |
| Null Count      | ⚠ 60.00  | N/A     |
| Distinct Count  | 10275.00 | N/A     |
| Unique Count    | 10273.00 | N/A     |
| Duplicate Count | 2.00     | N/A     |
| Blank Count     | 8.00     | N/A     |



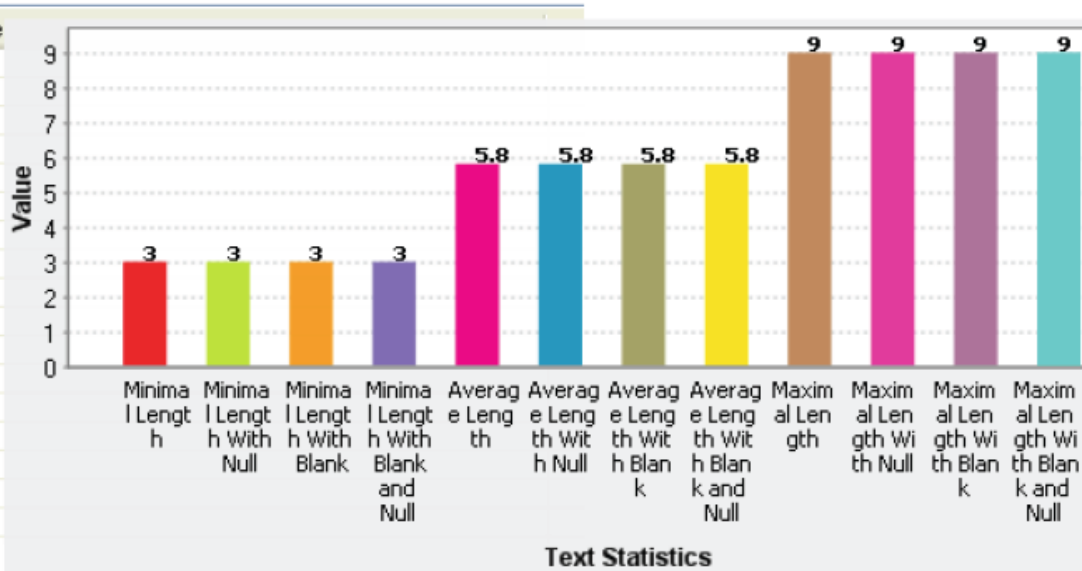
▼ Frequency Statistics

| value                              | count | %   |
|------------------------------------|-------|-----|
| Null field                         | 60.00 | N/A |
| Empty field                        | 8.00  | N/A |
| JamesLowry@Oak Bay.org             | 1.00  | N/A |
| AnnDeborde@Bremerton.org           | 1.00  | N/A |
| NancyPietrs@Berkeley               | 1.00  | N/A |
| RichardWellington@Santa Monica.org | 1.00  | N/A |
| JoyceBroderick@La Cruz.org         | 1.00  | N/A |
| JeremyPawloski@Beverly Hills.org   | 1.00  | N/A |
| AnitaBarton@Burbank.org            | 1.00  | N/A |
| ElzaTopp@Santa Anita.org           | 1.00  | N/A |



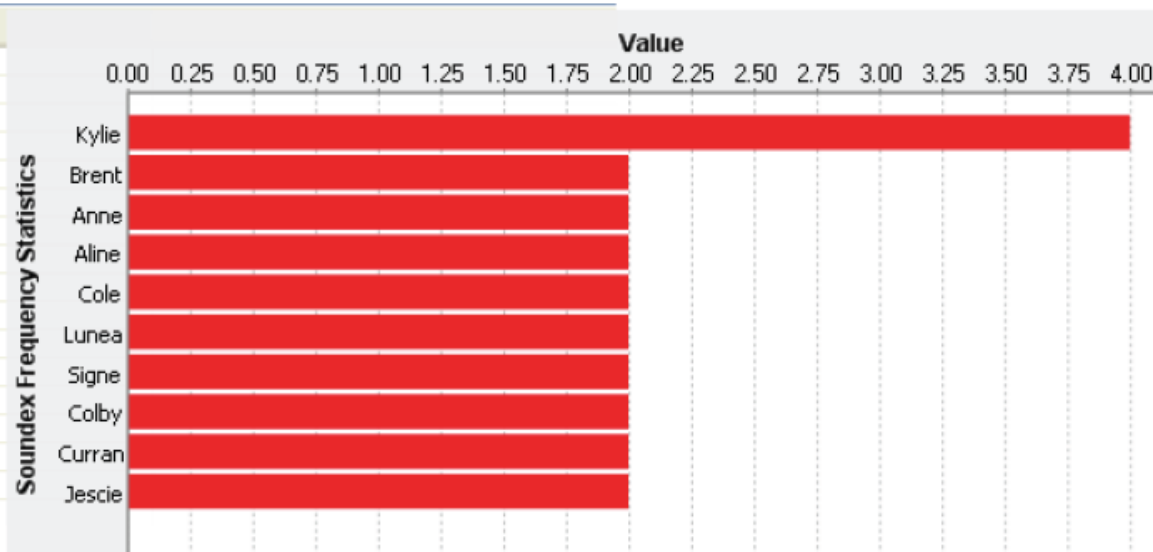
- ▼ Column:metadata.firstname
- ▶ Pattern Matching
- ▶ Simple Statistics
- ▶ Soundex Frequency Table
- ▼ Text Statistics

| 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 for IBM Information Analyzer

The screenshot displays the IBM Information Analyzer interface. The main window title is "IBM Information Server" with a menu bar (File, Edit, View, Help) and a toolbar. The current project is "IA\_OVERVIEW\_PROJECT" and the date is "9.43.86.77". The "INVESTIGATE" tab is active, and the "Column Analysis" sub-tab is selected. The "Select Data Sources to Work With" section shows the "EMPLOYEE" table. Below this, there are options to "View Analysis Summary" and "View Details". An information icon indicates that the user can view frequency distribution, data classes, properties, domain, and completeness information for the column. The "Select View:" panel on the left lists various columns, with "SALARY" highlighted. The main analysis area has tabs for "Overview", "Frequency Distribution", "Data Class", "Properties", "Domain & Completeness", and "Format". The "Properties" tab is active, showing inferred and defined structural properties. The "Data Type" section shows "Defined: DECIMAL", "Inferred: DECIMAL", and "Selected: DECIMAL". Below this is an "Inferred Summary" table:

| Inferred Data Type |       |         |  |
|--------------------|-------|---------|--|
| Data Type          | Count | Percent |  |
| DECIMAL            | 46    | 100     |  |
|                    |       |         |  |
|                    |       |         |  |
|                    |       |         |  |

The "Length" section shows "Defined: 9", "Inferred: 9", and "Selected: 9". Below this is another "Inferred Summary" section with a histogram and the following statistics:

- Minimum: 8
- Median: 8
- Average: 8.0217
- Maximum: 9
- Range: 1

At the bottom of the interface, there are buttons for "Close", "Rebuild Inferences", "Reference Tables", and "Save".

Felix Naumann  
Data Profiling  
Summer 2017

# Screenshots for IBM Information Analyzer

The screenshot displays the IBM Information Analyzer interface for Foreign Key Analysis. The main window shows the 'Foreign Key Analysis' pane with a table of 'Foreign Key Candidate Pairs' and a 'Common Domain' Venn diagram.

| Column      | Base Column | Paired Column |
|-------------|-------------|---------------|
| Table       | EMPLOYEE    | DEPARTMENT    |
| Source      | IA          | IA            |
| Primary Key | Yes         | No            |
| Foreign Key | No          | Yes           |
| Data Class  | Identifier  | Quantity      |
| Data Type   | INT32       | INT8          |
| Length      | 0           | 0             |
| Precision   | 0           | 0             |
| Scale       | 0           | 0             |
| Cardinality | -48         | 9             |
| Unique      | No          | No            |
| Constant    | No          | No            |
| Definition  | No          | No            |

**Paired to Base:**  
Common Data Values: 8  
Common Domain: 100.0000% Yes

**Base to Paired:**  
Common Data Values: 8  
Common Domain: 16.6667% No

**Common Domain:**  
Venn diagram showing the intersection of Base Column (40) and Paired Column (1). The intersection contains 8 elements.

## Typical Shortcomings of Tools (and research methods)

---

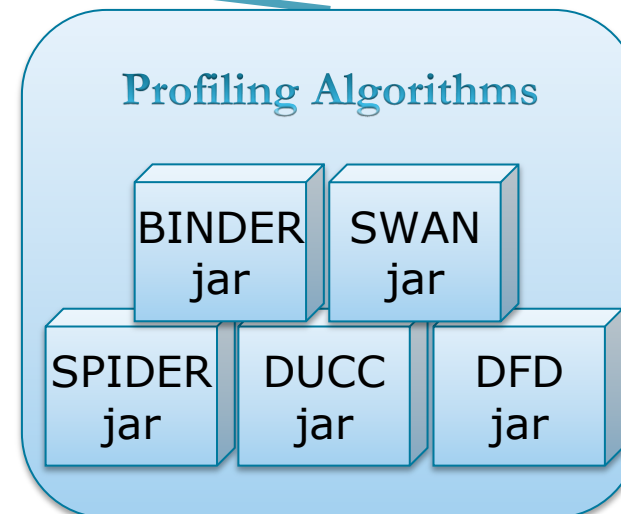
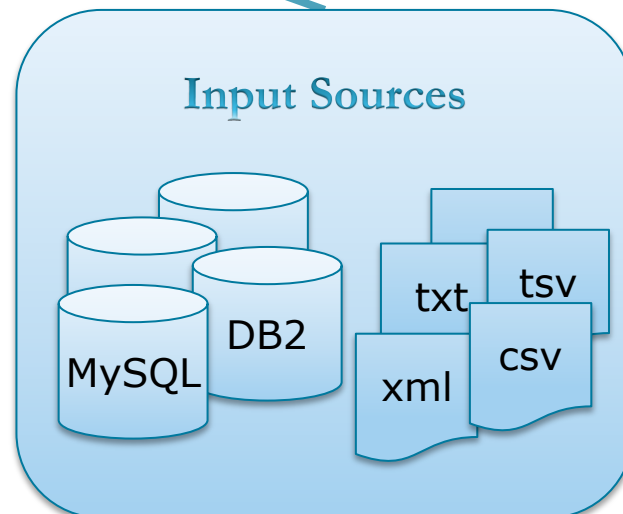
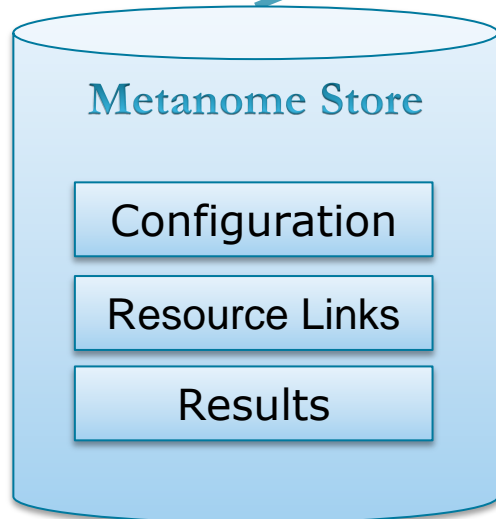
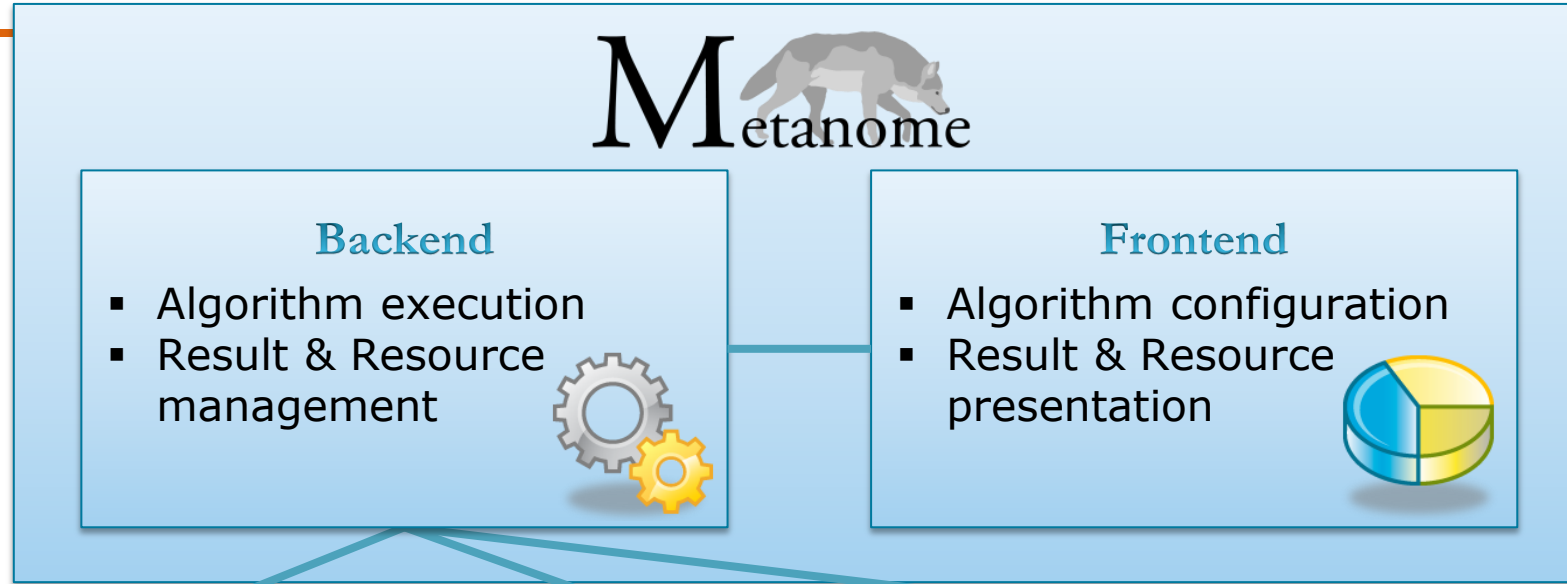
- Usability
  - Complex to configure
  - Results complex to view and interpret
- Scalability
  - Main-memory based
  - SQL based DBMS
- Efficiency
  - Coffee, Lunch, Overnight
- Functionality
  - Restricted to simplest tasks
  - Restricted to individual columns or small column sets
    - “Realistic”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

Why are DBMS a poor choice here?

That's the big one

# Metanome

An extensible architecture



Felix Naumann  
Data Profiling  
Summer 2017



## Overview

---

1. Profiling tasks
2. Profiling tools
- 3. Visualization**
4. Next generation profiling
5. Profiling challenges



Felix Naumann  
Data Profiling  
Summer 2017

## Motivation

---

- Human in the loop for data profiling and data cleansing.
- Advanced visualization techniques
  - Beyond bar-charts and pie-charts
- Interactive visualization
  - Support users in visualizing data, profiling results
  - Support any action taken upon the results
    - Cleansing, sorting, ...
    - Re-profile and visualize immediately
- Further reading:  
<http://vgc.poly.edu/~juliana/courses/cs9223/Lectures/intro-to-visualization.pdf>



# Profiler: Integrated Statistical Analysis and Visualization for Data Quality Assessment

<http://vis.stanford.edu/files/2012-Profiler-AVI.pdf>

## Schema Browser

- Creative Type
- Distributor
- IMDB Rating
- IMDB Votes
- MPAA Rating
- Major Genre
- Production Budget

Transform:

**MPAA Rating**

- R
- PG-13
- PG
- Not Rated
- G
- NC-17
- Open

**Creative Type**

- Contemporary Fiction
- Historical Fiction
- Fantasy
- Science Fiction
- Dramatization
- Kids Fiction
- Factual
- Super Hero
- Multiple Creative Typ

**Major Genre**

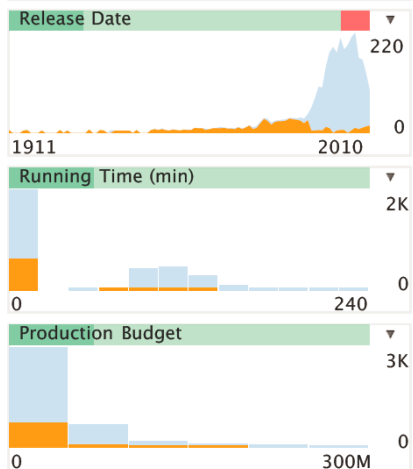
- Drama
- Comedy
- Action
- Adventure
- Thriller/Suspense
- Horror
- Romantic Comedy

Related Views: Anomalies

## Anomaly Browser

### Missing (6)

- MPAA Rating**
- Creative Type
- Source
- Major Genre
- Distributor
- Release Location
- Error (2)**
- Extreme (7)**
- Inconsistent (3)**
- Distributor (Levenshtein)
- Source (Levenshtein)



- Source**
- Original Screenplay
  - Based on Book/Short
  - Based on Real Life Ev
  - Remake
  - Based on TV
  - Based on Comic/Gra
  - Based on Play
  - Based on Game
  - Traditional/Legend/F
  - Based on Magazine A
  - Based on Musical/Op
  - Based on Short Film
  - Spin-Off
  - Based on Factual Boc
  - Disney Ride
  - Compilation
  - Based on Toy
  - Musical Group Movie

| Type         | Issue                    | Detection Method(s)                                              | Visualization           |
|--------------|--------------------------|------------------------------------------------------------------|-------------------------|
| Missing      | Missing record           | Outlier Detection   Residuals then Moving Average w/ Hampel X84  | Histogram, Area Chart   |
|              |                          | Frequency Outlier Detection   Hampel X84                         | Histogram, Area Chart   |
|              | Missing value            | Find NULL/empty values                                           | Quality Bar             |
| Inconsistent | Measurement units        | Clustering   Euclidean Distance                                  | Histogram, Scatter Plot |
|              |                          | Outlier Detection   z-score, Hampel X84                          | Histogram, Scatter Plot |
|              | Misspelling              | Clustering   Levenshtein Distance                                | Grouped Bar Chart       |
| Incorrect    | Ordering                 | Clustering   Atomic Strings                                      | Grouped Bar Chart       |
|              | Representation           | Clustering   Structure Extraction                                | Grouped Bar Chart       |
|              | Special characters       | Clustering   Structure Extraction                                | Grouped Bar Chart       |
|              | Erroneous entry          | Outlier Detection   z-score, Hampel X84                          | Histogram               |
|              | Extraneous data          | Type Verification Function                                       | Quality Bar             |
|              | Misfiled                 | Type Verification Function                                       | Quality Bar             |
|              | Wrong physical data type | Type Verification Function                                       | Quality Bar             |
| Extreme      | Numeric outliers         | Outlier Detection   z-score, Hampel X84, Mahalanobis distance    | Histogram, Scatter Plot |
|              | Time-series outliers     | Outlier Detection   Residuals vs. Moving Average then Hampel X84 | Area Chart              |
| Schema       | Primary key violation    | Frequency Outlier Detection   Unique Value Ratio                 | Bar Chart               |

Figure 1: The Profiler User Interface. The UI contains (clockwise from top-left): (a) linked summary visualizations, and (d) anomaly browser. Profiler generates a set of we investigate possible causes of missing MPAA movie ratings. The grey bar above we select it to highlight matching records. The Release Date chart shows that missing ratings correlate with earlier release dates.

Figure 6: Taxonomy of Data Quality Issues. We list classes of methods for detecting each issue, example routines used in Profiler, and visualizations for assessing their output.

Felix Naumann  
Data Profiling  
Summer 2017

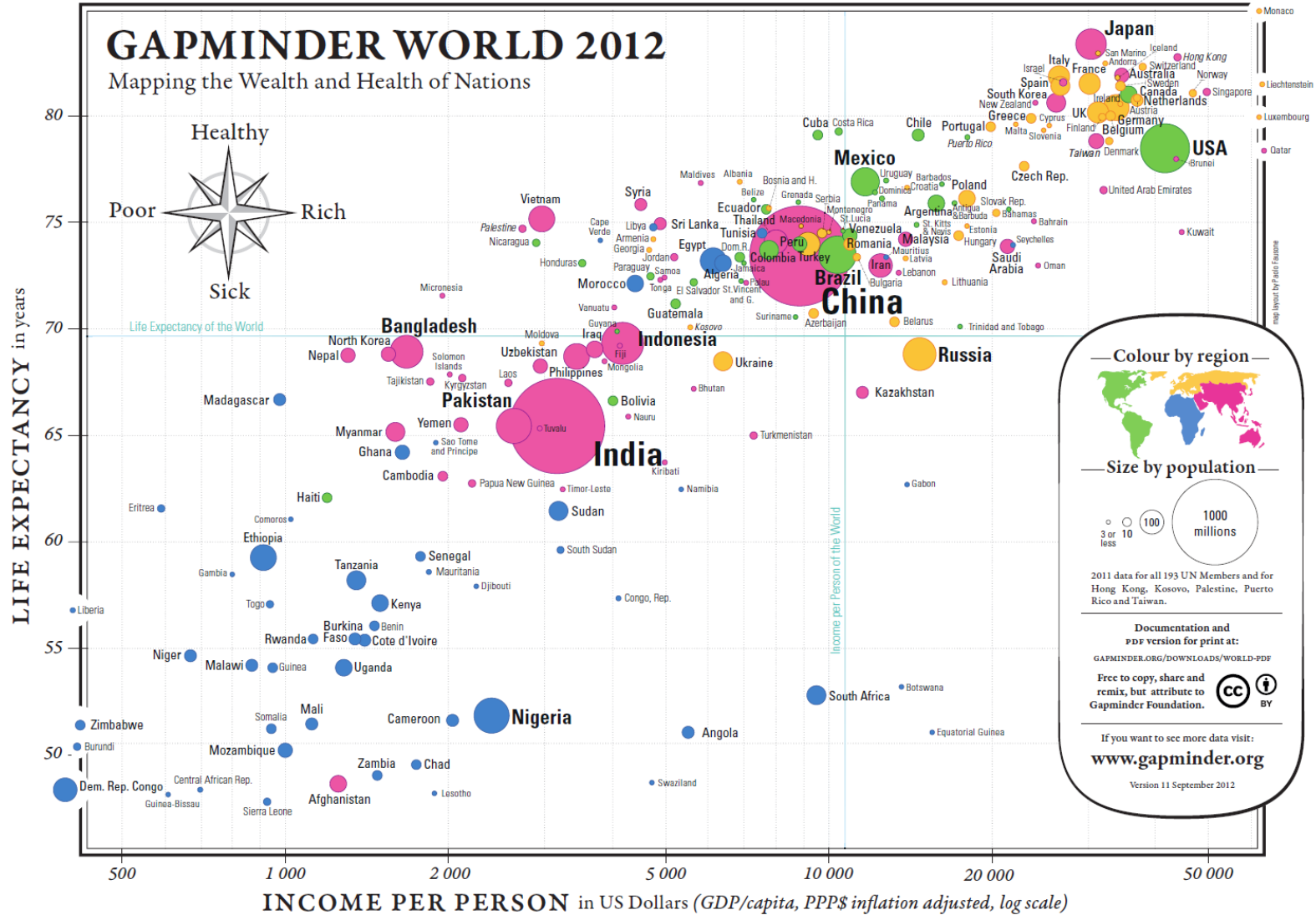
## Massive screens for massive data

---

- Powerwall, University Konstanz
- 5.20 m x 2.15 m; almost nine million pixels



Felix Naumann  
Data Profiling  
Summer 2017



<http://www.gapminder.org/>

Felix Naumann  
Data Profiling  
Summer 2017

## Overview

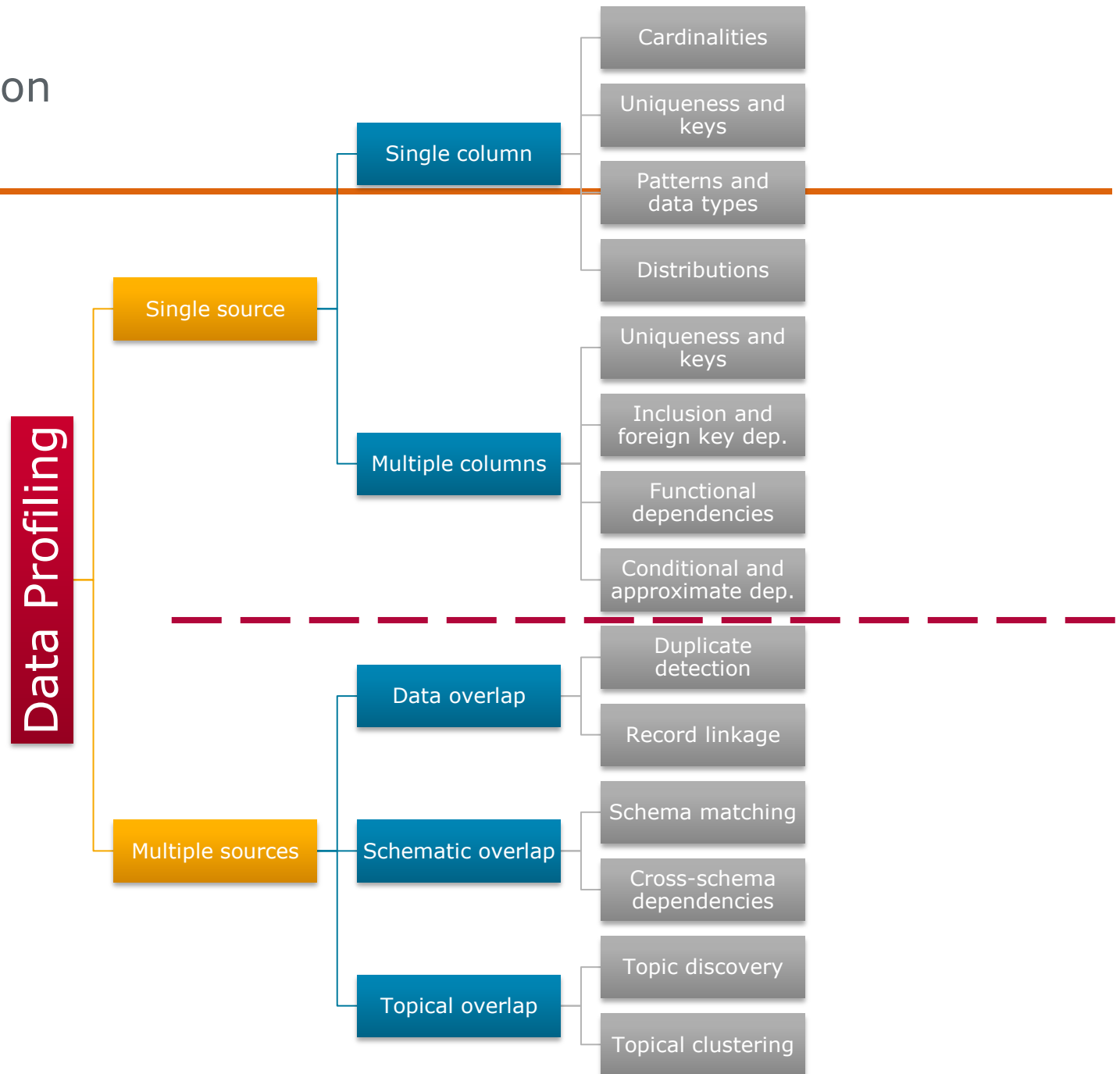
---

1. Profiling tasks
2. Profiling tools
3. Visualization
4. **Next generation profiling**
5. Profiling challenges



Felix Naumann  
Data Profiling  
Summer 2017

# Extended Classification of Profiling Tasks



## Schematic Overlap: Schema Matching

---

- (Semi-)Automatically determine cross-schema value correspondences between attributes
  
- Traditionally: Input for data transformation and exchange tasks
- With data profiling
  - Extract features from schema **and data**
    - Instance-based schema matching
  - Compare features
    - Similarity measure, machine learning
- For data profiling
  - Determine *closeness* of two schemata
  - Determine „schema fit“
    - Complement (few matches) or union (many matches)



# Schema Complement: Cross-Schema Dependencies

- Inclusion dependencies across schemata
  - Join paths between data sources
- Conditional INDs
  - Typical pattern among crossreferencing sources

### Catalog

| Unit cost | DBName    | ProdID |
|-----------|-----------|--------|
| 200 USD   | ToyDB     | 17     |
| 50 EUR    | ToyDB     | 18     |
| 1000 QAR  | FashionDB | 18     |

### ToyDB

| EntityID | further data |
|----------|--------------|
| 17       | abcd...      |
| 18       | efgh...      |

### FashionDB

| EntityID | further data |
|----------|--------------|
| 18       | abcd...      |
| 19       | efgh...      |

## Duplicate Detection and Record Linkage

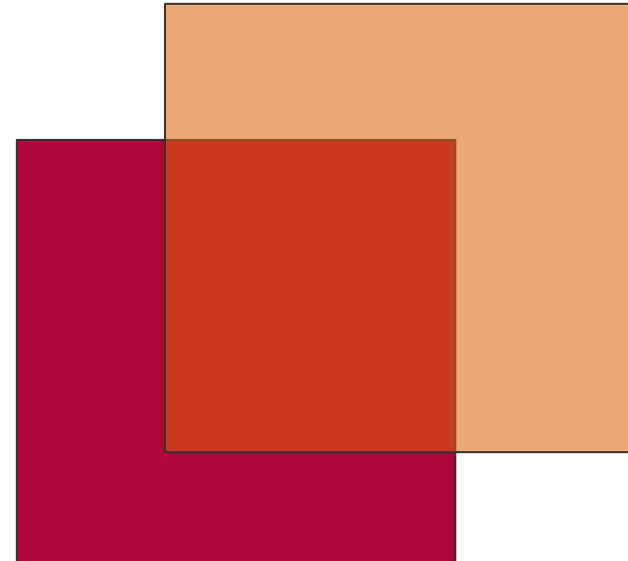
---

- Detect multiple (different) representations of the same real-world entity
  
- Traditionally: Input for data cleansing and data fusion tasks
  - See second part of course
  
- For data profiling
  - Intra-source: Determine duplicity/cleanliness
  - Inter-source: Determine „data fit“
    - Complement (few duplicates) or union (many duplicates)

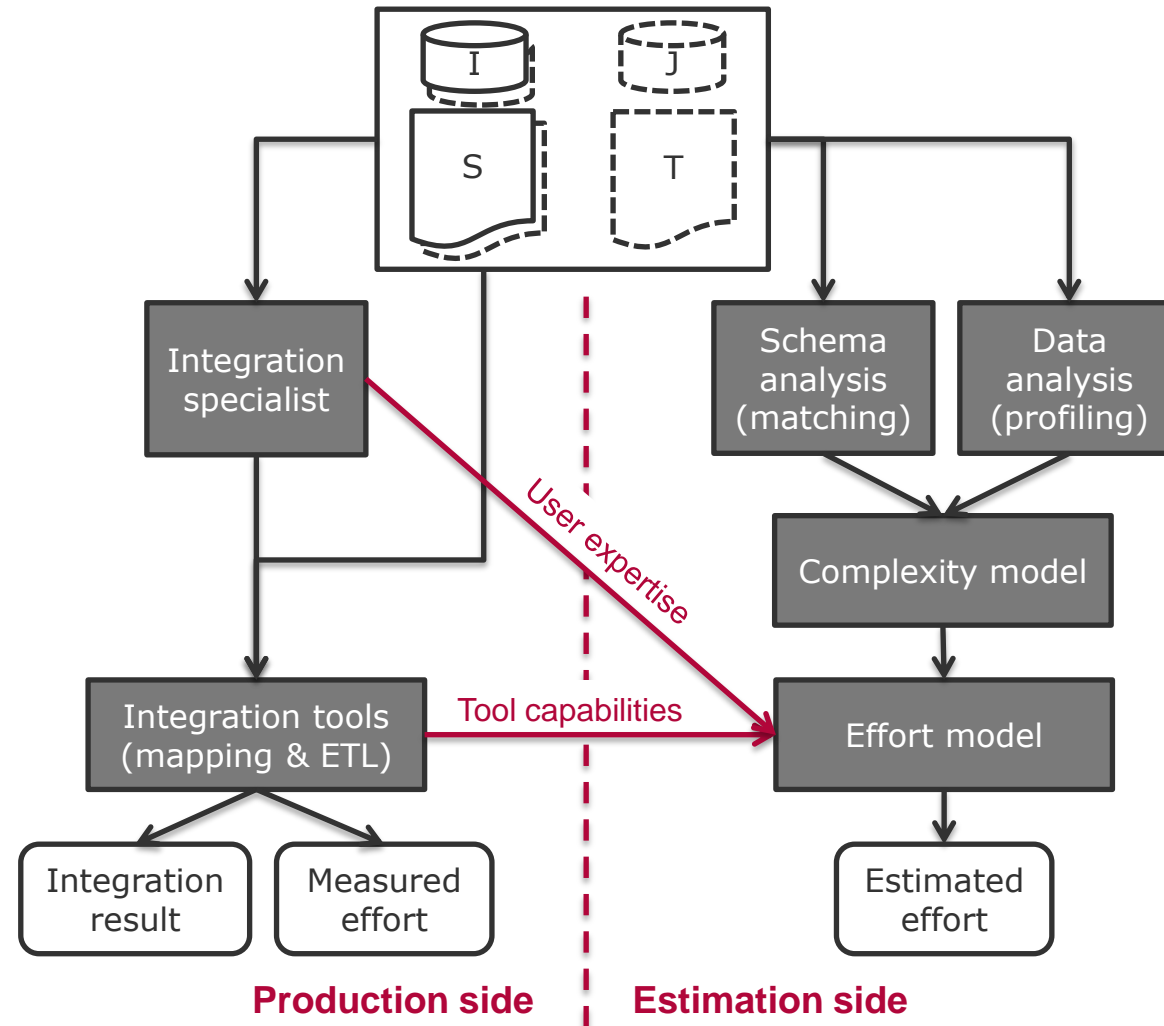
## Together: Profiling for Integration

---

- Create measures to estimate integration (and cleansing) effort
  - Schema and data overlap
  - Severity of heterogeneity
- Schema matching/mapping
  - What constitutes the “difficulty” of matching/mapping?
- Duplicate detection
  - Estimate data overlap
  - Estimate fusion effort
  
- Overall: Determine integration complexity and integration effort
  - Intrinsic complexity: Schema and data
  - Extrinsic complexity: Tools and expertise



# Integration effort estimation



Felix Naumann  
Data Profiling  
Summer 2017

## Topic discovery and clustering

---

- What is a data set **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)
  
- Topical clustering for
  - Source selection, query processing
  
- What is a topic?
  - Wikipedia categories?

# Wikipedia Categories

Kategorien: [Robotikhersteller](#) | [Volkswagen](#) | [Unternehmen \(Wolfsburg\)](#) | [Automobilhersteller](#) | [Motorenhersteller](#) | [Mitglied der Europäischen Bewegung Deutschland](#)

Kategorien: [Amtierender Außenminister](#) | [Außenminister \(Bundesrepublik Deutschland\)](#) | [Vizekanzler \(Deutschland\)](#) | [Bundestagsabgeordneter](#) | [Bundesvorsitzender der FDP](#) | [Vorsitzender der FDP-Bundestagsfraktion](#) | [Generalsekretär der FDP](#) | [Bundesvorsitzender der Jungen Liberalen](#) | [Kanzlerkandidat \(Deutschland\)](#) | [Ritter des Ordens wider den tierischen Ernst](#) | [Deutscher](#) | [Geboren 1961](#) | [Mann](#)

Categories: [Companies listed on the New York Stock Exchange](#) | [Dow Jones Industrial Average](#) | [IBM](#) | [Cloud computing providers](#) | [Cloud computing vendors](#) | [Companies based in Westchester County, New York](#) | [Companies established in 1896](#) | [Computer companies of the United States](#) | [Computer hardware companies](#) | [Computer storage companies](#) | [Display technology companies](#) | [Electronics companies of the United States](#) | [Initialisms](#) | [Multinational companies](#) | [National Medal of Technology recipients](#) | [Point of sale companies](#) | [Semiconductor companies](#) | [Software companies of the United States](#) | [UML Partners](#)

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](#)

**Felix Naumann**  
**Data Profiling**  
**Summer 2017**

## Overview

---

1. Profiling tasks
2. Profiling tools
3. Visualization
4. Next generation profiling
- 5. Profiling challenges**



Felix Naumann  
Data Profiling  
Summer 2017

## Profiling Challenges

---

- Efficient profiling
- Scalable profiling
- Holistic profiling
- Incremental profiling
- Online profiling
- Profiling query results
- Profiling new types of data
- Data generation and testing
- Data profiling benchmark

Felix Naumann  
Data Profiling  
Summer 2017

Each of these is worth one (or more) master's theses!



## Efficient and Scalable profiling

---

- Scalable in number of rows
- Scalable in number of columns
  - Number of column combinations is exponential
  - Small table with 100 columns:  $2^{100} - 1 = 1,267,650,600,228,229,401,496,703,205,375 = 1.3$  nonillion combinations
    - Impossible to check
    - Impossible even to enumerate
- Solutions
  - Scale up: More RAM, faster CPUs
    - Expensive
  - Scale in: More cores
    - More complex (threads)
  - Scale out: More machines
    - Communication overhead

## Holistic Profiling

---

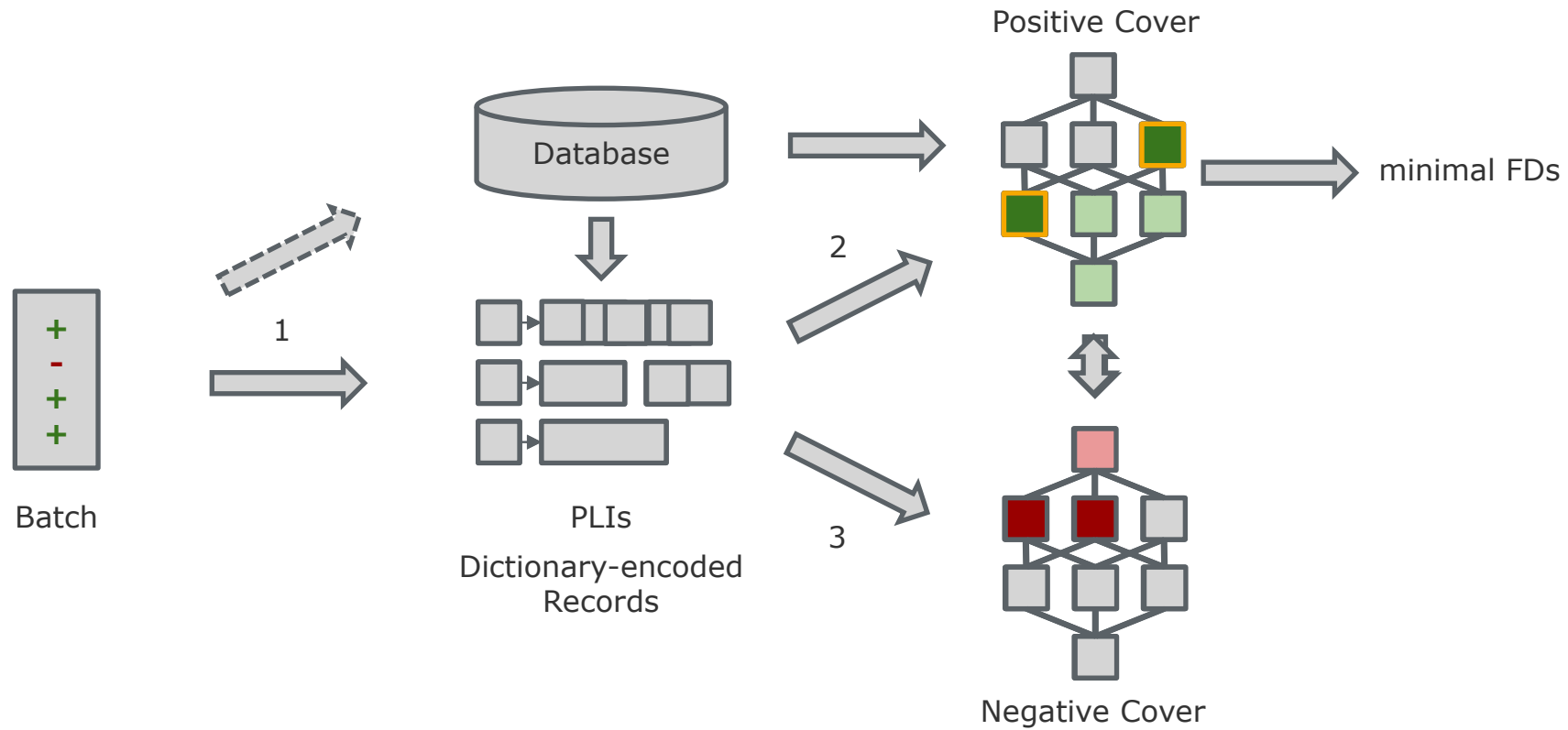
- Various profiling methods for various profiling tasks
- Commonalities/similarities
  - Search space: All column combinations (or pairs thereof)
  - I/O: Read all data at least once
  - Data structure: Some index or hash table
  - Pruning and candidate generation: based on subset/superset relationships
  - Sortation: Benefit from sorted sets
- Challenge: Develop single method to output all/most profiling results
- Masterprojekt – paper submission

## Incremental profiling

---

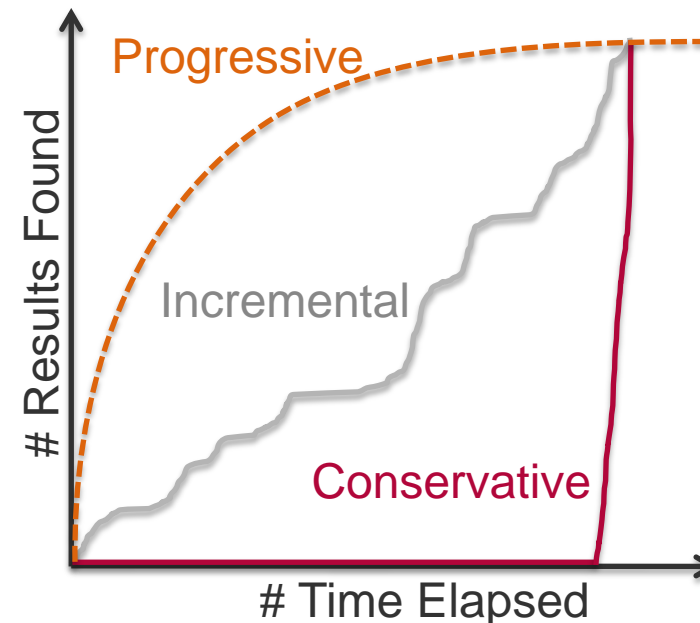
- Data is dynamic
  - Insert (batch or tuple-based)
    - Data streams
  - Updates
  - Deletes
  
- Problem: Keep profiling results up-to-date without reprofiling the entire data set
  - Easy examples: SUM, COUNT, AVG
  - Difficult examples: MIN, MAX, MEDIAN, uniqueness, FDs, etc.

# DynFD – Masterproject 2016/17



# Online Profiling

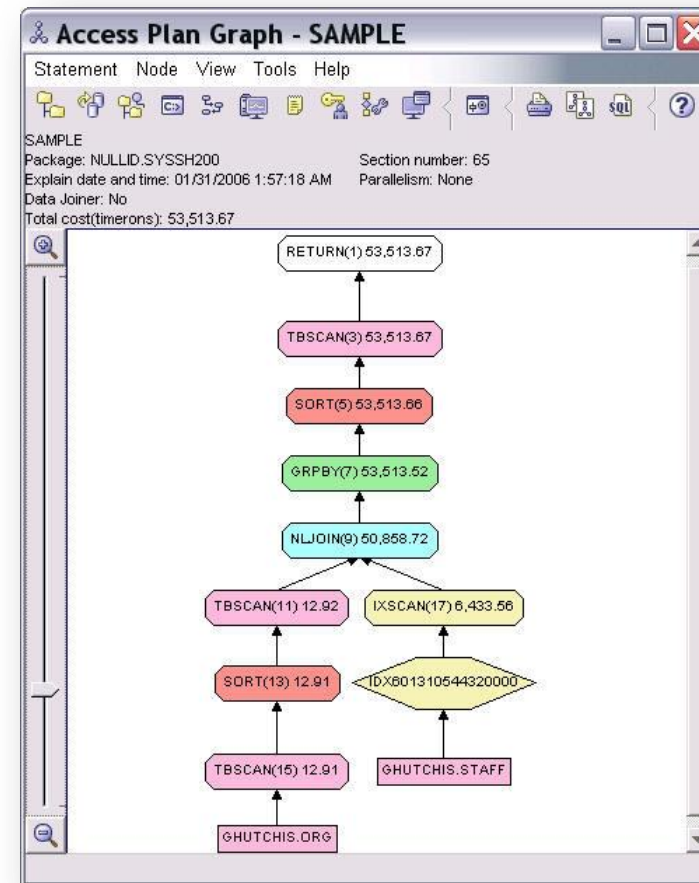
- Profiling is long procedure
  - Boring for developers
  - Expensive for machines (I/O and CPU)
- Challenge: Display intermediate results
  - ... of improving/converging accuracy
  - Allows early abort of profiling run
- Gear algorithms toward that goal
  - Allow intermediate output
  - Enable early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



Felix Naumann  
Data Profiling  
Summer 2017

# Profiling query results

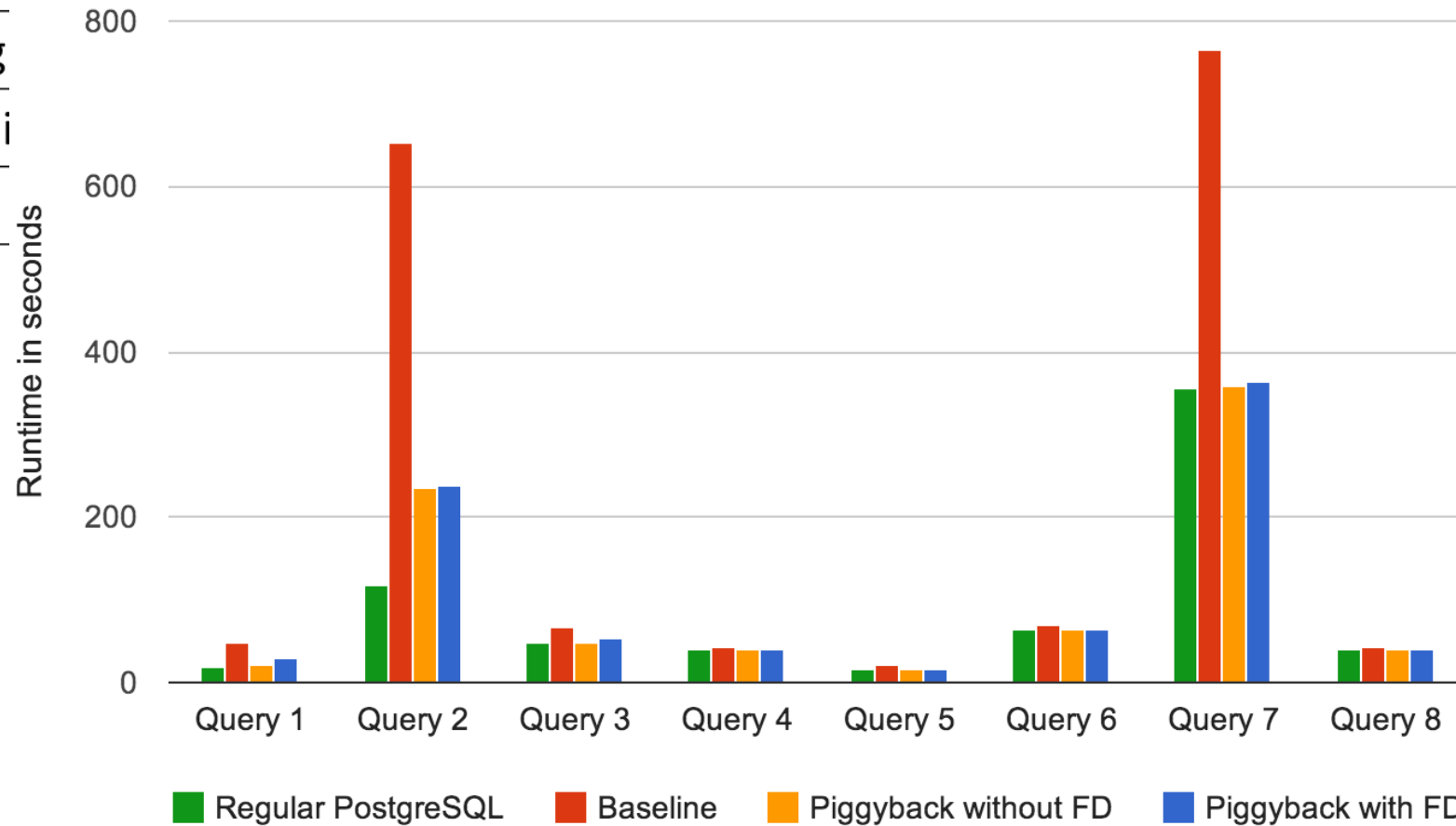
- Query results are boring: Spruce them up with some metadata
  - Usually only: Row count
  - For each column, give some statistics
    - Uniqueness, histogram, AVG, etc.
    - Show FDs
  
- Idea: Piggy-back profiling on query execution
  - Re-use sortations, hash tables, etc.



Felix Naumann  
Data Profiling  
Summer 2017

# Masterproject "Piggyback Profiling" (2013/14)

| rec.id | rec.name      | ldiff | tr.id | tr.name | tr.position |
|--------|---------------|-------|-------|---------|-------------|
| 5      | She loves you | 0     | 350   | cl...   | 12          |
| 7      | Hig           |       |       |         |             |
| 8      | Cali          |       |       |         |             |
| ...    | ...           |       |       |         |             |



Felix Naumann  
Data Profiling  
Summer 2017

## Profiling New Types of Data

---

- 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

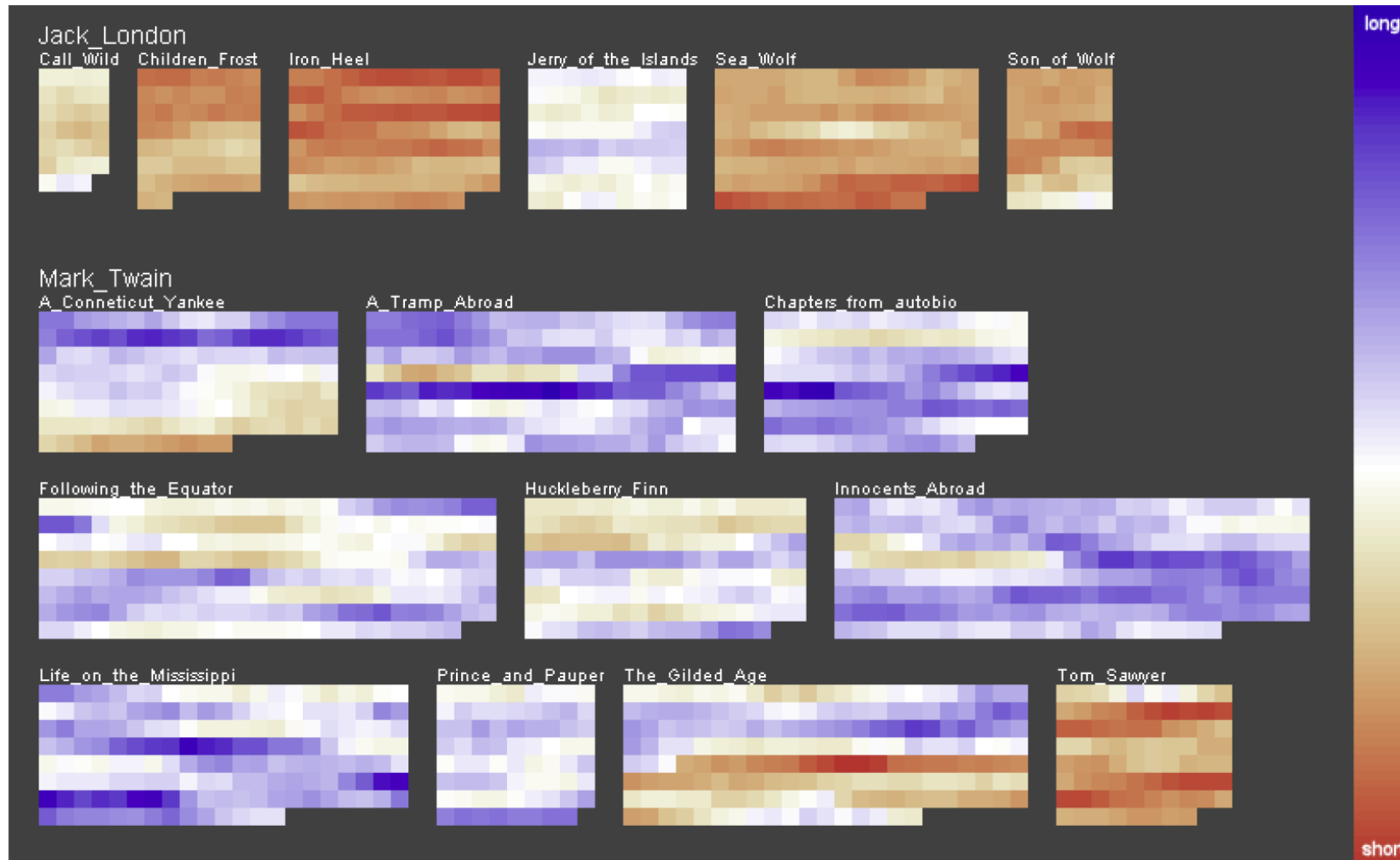


## Example: Text profiling

---

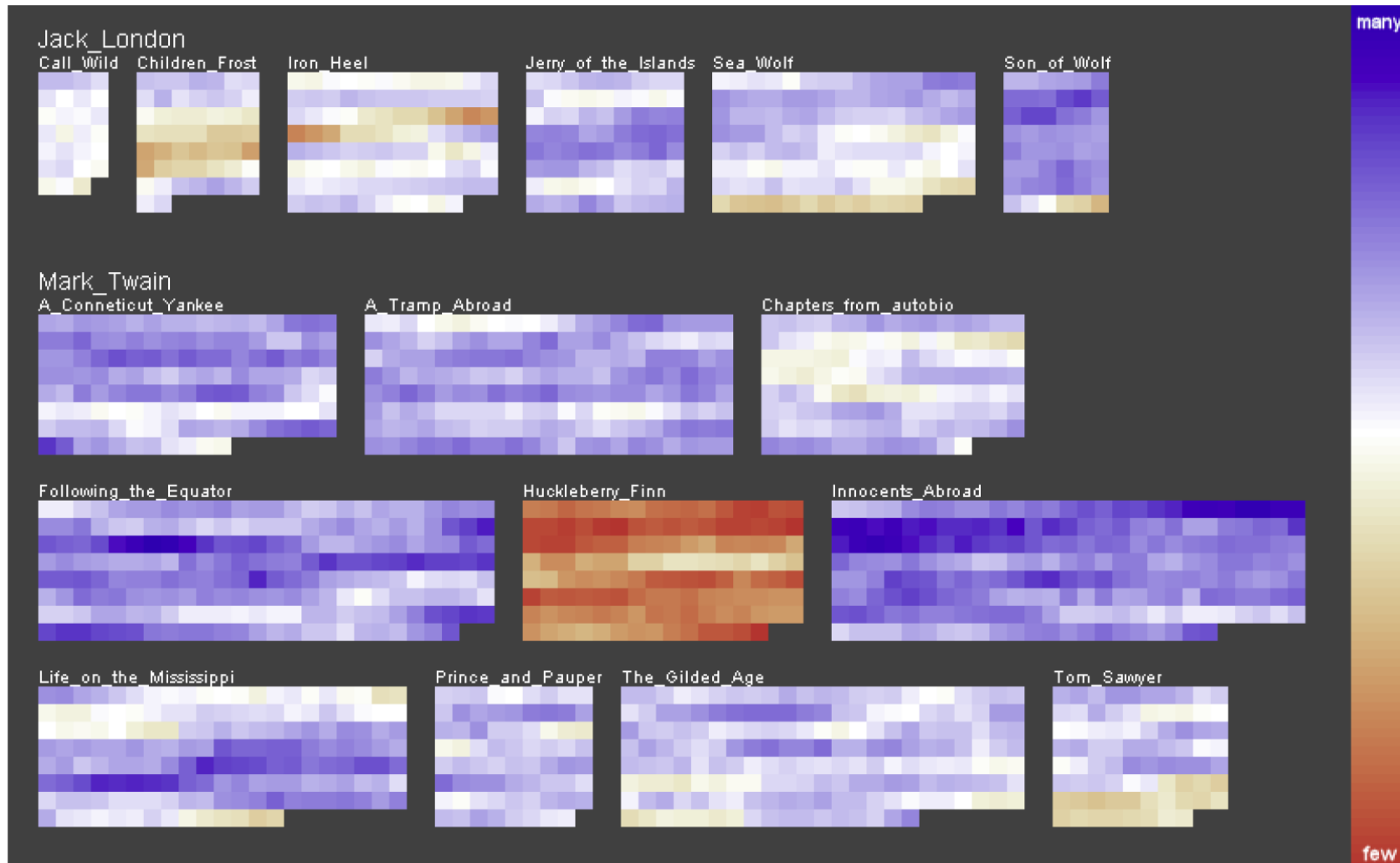
- Statistical measures
  - Syllables per word
  - Sentence length
  - Proportions of parts of speech
- Vocabulary measures
  - Frequencies of specific words
  - Type-token ratio
  - Simpson's index (vocabulary richness)
  - Number of hapax (dis)legomena
    - Token that occurs exactly once (twice) in the corpus
    - Characterize style of an author
- Idea and following figures based on:
  - „*Literature Fingerprinting: A New Method for Visual Literary Analysis*“ by Daniel A. Keim and Daniela Oelke (IEEE Symposium on Visual Analytics Science and Technology 2007)

# Average sentence length



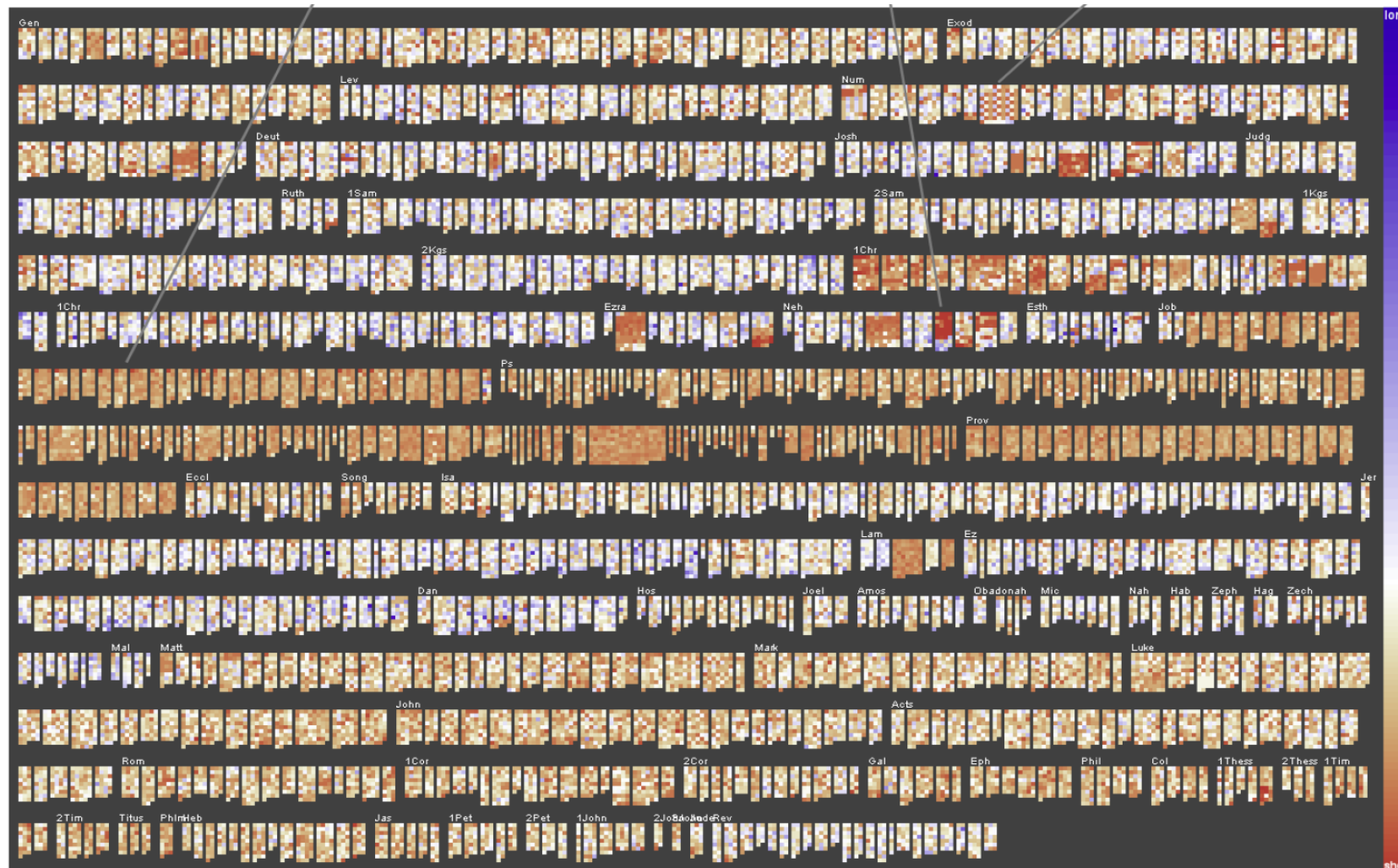
Felix Naumann  
Data Profiling  
Summer 2017

# Hapax Legomena



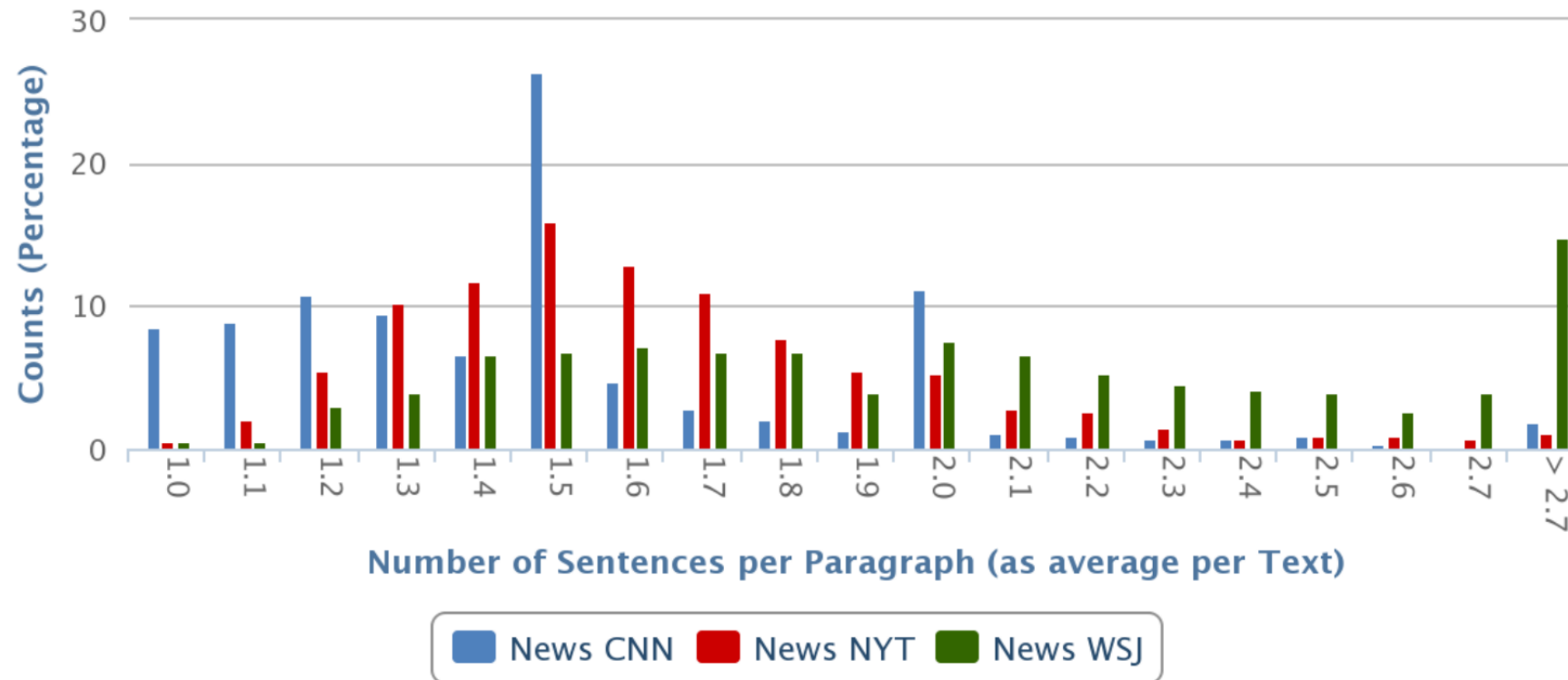
Felix Naumann  
Data Profiling  
Summer 2017

# Verse length



Felix Naumann  
Data Profiling  
Summer 2017

# News Statistics



Felix Naumann  
Data Profiling  
Summer 2017

