# Google Public Data

Enhancing Data Discovery and Exploration

Jürgen Schwärzler

March 2011

### Users interest in public data



- education statistics by school
   unemployment
   population

   population, cities
   population, density
   population, growth

   sales tax
   salaries
- 6 exchange rates
- 7 crime statistics crime statistics, human trafficking crime statistics, homicides crime statistics, hate crime
- 8 prevalence

aids

alcohol abuse, prevalence drug abuse, prevalence

- 9 GDP
  - GDP, nominal GDP, real
- 10 minimum wage
- 11 disaster statistics disaster statistics, hurricanes disaster statistics, floods disaster statistics, storms

- 12 oil price
- 13 last names
- 14 poverty statistics
- 15 mortality
  - mortality, swine flu mortality, infant
    - mortality, teen suicides
- 16 election results
- 17 consumer price index/inflation
- 18 cost of living
- 19 accident statistics accident statistics, car/traffic accident statistics, drunk driving accident statistics, distracted driving
- 20 gas price
- 21 prison statistics
- 22 earthquake statistics
- 23 obesity statistics
- 24 solar energy solar energy, production solar energy, costs
- 25 baby names

#### Google's interest in public data



- It is part of Google's mission to organize the world's information
- Weak search results (e.g. hiring and firing statistics for santa clara county)
- There is more public data that is created and collected every day that is meaningful to users



Google

Unified metadata - DSPL format - Shared concepts (time, geo, etc.) Partnerships with official providers

# Organize the world's public data and make it universally accessible and useful.

Public data search - Onebox - Vertical search Rich visualizations - <u>Public Data Explorer</u> - <u>Motion chart</u> Organize the **world's public data** and make it universally accessible **and** useful.

Raw hybridization a	rray data for	breast tuno	r cell lines	s and tumor t	tissues	10.000	1. N. 1. T. M.	21.775.835	And the Andrew Andrew Andrews		1,000		La construction de la construction	Distance and some	p.a.e		0.0.2.5		1.55				20120-010		10.000	125.25-
CR status	21PT (exp 1) Di-	21PT (exp	(R-	21PT (sap 3) [I]-	1	21NT ER-	25HT-8 (8-	21HT-2 (I)-	HDA-HD-435 (exp 1)	HDA-HD-435 (exp 2) (II-	HE	1	HEFIDA (exp 1)	HEF10A (exp 2) (I-		1-4 DP-	843 UP	ł	H 1-	PT-1 (3)=	PT-		PT-0 (01+	PT-6 (R+	#10 (21-	#16 (0+
Stage Dese #	na/cell lite Mean Intensity 1		illy \$500er 8	naforil line Mean Intensity 183	IDer Heat Infe							y \$50e			SiDev Heat Island		ran Intensity 1851						ean intensity Willier		lean intensity #3/De	
6-actin 1 3604 2	28114 46165	77 79	047 20 860 12	4147	11	12801 2		27 2217 22 25114	28 26 80 26 2141 1	4 207996	16 51 10 21051	0 2	29 897005	6 85264 ? 2901149	85 3	664 12 056 10	17704 124009	11 55 18 730	6 H	1212	17 109 70 1669	0 10	1886 8 101825 6	5022 28 144062 4	52% 29545	7 1184 BKG 6 200676 16
histona 144 3 brtubulin 4	40605	74 115	553 22 853 33	621 15421	22	12220 48	9 8105	29 9089	25 551 2 24 006 1	6 8507	045 45 30 4015	10	9 36149	29 1405430 10 1799454	21 1	10 BKG 565 12	175039	29 56 9 662	4 25	798	9 211		1094 40 6506 16	12956 7	50 B 4795	5 294052 27
pOtil cloning vector 5 unknown TA 6	2938 564425	14. 151	128 (HS 600 12	499	1 3	125 860	0 156602	10 156847	80 20 80 11 18794	5 2162904	171 40 38 36859	н	6 995429	840 21592 8 20%/805	15 14	779 25 645 6 691 82	1189062	840 2 11 6042	18 11	187022	7 8728		10269 66 8952% 9 1588 21	805947 11	12242 4	7 826685 15
5' nullectidase 7 aldose reductase 0	2731 21475 9049	6 1	392 21 147 29 931 II	5805 764 908	045	4940 8; 11096 4 8517 14	4 8092	80 414 11 3542 10 1470	12 29 80 11 964 4 12 1276 2	0 58724	44 25 18 25 91 447	2 04	168516	15 16780 6 95268 4 827080	14 1	691 82 160 19 81 8KG	4675 6370 195821	27 1 56 63 6 872	N 10	1708	18 70 21 122 15 113	12 15	20472 21 4564 16	45 800 11871 6 22216 15	999 4 463 12545	7 807 803 8 927 803 5 647994 10
alpha I-antichymotrypsin 9 alpha I-antitrypsin 10 alpha 6 integrin 11		12 59	456 12 92 BHS	65471 678	6 85	7258 15	5 6312	10 14/0 11 7116 KG 561	12 1276 2 6 4296 1 9 20 80	6 65995	27 6421 28 750	N 1	15 20000 16 2275	4 627080 3 551005 41 10694	17 2	560 21 27 BKG	100004 750	9 724		2565	15 009		250805 10 3820 14		17857 55 B	5 611294 10 7 896847 51 0 125 8K0
mys assorgene, Binl 12 subacto light shain 18		48	958 34	641 11499	883	5563 54 115662 1	6 4881	46 180 1 12 48227	8K0 556 2 28 966		70 24 17 848	6 8	15 258	6 4687	28	29 BK0		8KG 1 7 2089	19 883	1667		70 20	7956 63 455719 9		69 B 91829	
carbonic anhydrace 14 careerin 15	14419	27 2	618 16 539 20	5401 21104	85	5985 21 4970 1	0 2991	14 12%	64 30 80 31 650 1	0 880%	18 1990 6 14	6	8 85919	4 20065 7 110500	34	981 28	2829	18 2	15 810	66 6	80 102 29 57	1 1	6797 63 5218 22	282 800	8297 4800	4 2048 28 5 251 0x0
0044 16 0059 17	52491	5 54	910 11 506 41	66915- 1897	25	\$3427 ( 1202 4)	0 10402	10 13940	5 2235 1 BK0 27 BK0	7 50549	22 3404 28 1066	6 1	18 295000	6 176575 6 33915	16 2	NI2 16 408 80	474741	6 1000 83 106	8 6	2992	22 1702	75 10	107054 5	107797 9 33317 7	89828 7119	4 64515 19 6 7648 5
unicover 15e 18 condition and profig 19	12211		682 68	1531	4 1	4905 31	8 3394	9 2856 14 49053	16 28 80 8 9241 1		51 260 17 1547		9 7088 15 294958	24 0005 6 729791		144 16 150 B	11918 741451	8 87 12 3105			19 29	19 10	4948 44 563192 7	123 BOD 564950 7	3754 1 217166	1 1786 29 8 190870 8
CCR 20 connexis 26 21	7305 6277	45 1	309 15 141 54	1967 1267	883 985	8821 25 8706 21	7 10582	19 1280 94 840	80 890 8 82 27 890		17 58 30 21	5 04	69095	9 220696 7 46957	18	596 29 569 50	81909 8212	24 81			80 45 12 98	6 6	1048 37 10703 6	6498 58 2953 25	901 1970	2 280401 12 5 12669 7
osmexin 43 22 system A 25			119 BHS 779 4	029 046	0+5 0+5	1262 2' 53 80		17 000 K0 758	10 27 00 22 776 4		57 105 55 217		12 7219 15 261176	64 6500 4 170701	55 4	941 10 31 8KG	3516 554	64 34 1	H 845		45 26 11 25		1705 44 1750 26	81717 218	625 32 D	1 270 BKG 5 17981 12
dermoglein2 24 Ditik-binding protein A 25		7	811 85 275 (KD	780 1090	843 843	833 2 1381 8	1 474	13 904 44 41	22 25 80 80 26 80	0 55	28 15 845 23			45 299077 17 198	26 (KO	50 8K0 29 8K0	15144 65	4 89 BKD 1		29 1	83	10 5 19 8KD	1090 33 627 33	49 800 1880 88	722 451 1	8 200968 14 5 1221 21
elafin 26 HE1 27		17 1	708 25 972 84	948	045 045	1821 12	8 55 B	82 1805 KG 1827	20 756 8 20 709 4	4 18799	26 25 17 963	12	7 39915	3 254004 11 25702	4	748 27 599 15	48450 22180	9 87 5 86	40 4	2652	18 225		603 40 62959 7	26521 9	46 B 20563	7 0041 24
fibronectin 28 uninovn 36d 29	5405	83 9	00 045 518 10	964 28503	5	010 2× 0155 21	8 2029	12 56 80 903	840 200 3 28 1136 1	0 4515	25 2950 54 990	9 2	2 1062 20 885795	30 9050 7 149228	8 1	099 16 444 20	1697 94358	50 10 2 1175	9 2	19519	10 5159		139 BO 1454645 3	872042 5	40 D	5 5181 26 5 140620 3
fibrohestin receptor 30 067 pi 31	6115	33	405 6	1188	80	159 80	0 8021	19 45635 13 149	6 8528 1 900 20 900	6 47115	25 460	0 1	5 502992 18 742475	6 457566 9 296808	4	761 9 167 52	84994 258740	8 712	17 4		7 1070	17 5	1207% 12	14854 56	30068 24796	5 87890 15 7 46219 8
mai25 82 kerata 14 88	65	EKG .	641 54 648 54 247 12	970 1360 11907	843	14908 14 57 BR	0 07 8		13 216 40 000 20 000	62 62	29 560 (85 161	7 4	40 305929	28 24248 9 107889 17 19174	9	990 17 24 BK5 999 82	7975 64555	82 20 8 85 12 59	12 6	15 (	KD 51 KD 90	10 10	7171 22 1992 36	8415 26 16641 11 88822 7	2781 1	2 4991 18 1 56 BKD
kwatin 19. 34 karatin 5. 33 karatin 6a. 36		5 N	247 12 697 15 115 860	11907 155135 1545	20 12 840	1900 28 24606 0 2927 38	6 5041	60 2122 5 5685 60 39	0 1407 2 2 1579 1 840 580 4	0 12627	26 4114 11 597 41 228	N 2	0 04025 24 4050474 36 59057	17 19174 4 2502124 5 43750	0 2	099 82 422 0 22 8KG	9216 160959 95	12 B9 0 966 0KG 11	46 7	1058	15 679 84 368 85 1		42127 7 53050 5 52 840	56654 6	4000 4 36363 623 4	5 21508 5 8 10254 12 0 891 82
ianinin alpha 3 37 Ianinin leta 3 38		81	141 883 882 28	1290	843 20	41 80	0 615		800 900 8 800 27 80 10 580 3	0 84154	11 271 36 302	6 1	19 829	9 6657 9 98924	80	22 040 150 36 405 11	79 1444 X208		1 843	887	33 12 19 21	11 40	2080 61 2910 60	1106 46	1561	5 971 87 1 1118 29
laminin gamma 2 35 Systemmal playfor and fer ase 40	12929	1 14	985 12 111 845	41458	18	2560 1	9 955	87 1882 22 420	24 421 2 14 200 2	4 57	645 20 30 220	H (R	(0 145391	7 170072 12 7047	4 1	642 29 29 BK5	62177 2214	41 45	15 9	908	4 17	4 5 5 7	2067 15		577	0 21788 18 8 8011 24
marph 41 mbv-2 42	2752		742 20	054	0+5 0+5	1244 1	9 820	6 774 82 6842	28 1029 10 25 00	5 54	045 21 045 27	5 (8	0 0529	30 461409 22 2047546	18 1	056 33 27 8k0	19170	11 100	8 2	N 6	10 14	77 22 10 840	1192 4 63 800	1450 35	1556 J	105005 16
myssin light shain 45 myssin light shain 2 44	18004	14 22	404 12 795 9	78772	49	5691 18837	7 2008 7 1042	9 16/5	7 645 2		29 2607 15 40	1 1	14 64371	8 853954	4	M2 25	60768 8722	12 507	18 2	8750	8 446	19 4	129091 6	46310 13 62 BO	9210 710 1	9 282606 13
N-ras 45 meth 46	1964	82 1	716 1 587 25	2647 1208	12	1972 27 2564 1	9 1364	17 658 20 946	48 27 89		14 51E 27 IN		7 70929 81 75467	7 45458 6 21085	15	689 60 26 BKG	7121 3961	10 149		2610	8 294 85 55		49043 8 15275 12	35504 13 9693 1	2048 1 8225	1 16750 11 5 1661 0
utinovn 12x 47 utinovn 15x 48	8079	64 16	554 96 144 BHS	1109	0HS 0HS	1051 25 701 8		12 47	9x0 27 9x0 9x0 26 9x0	6 610	120 651 59 211	0 1		16 2006 4 9727		27 B(G 26 B(G	4108	33 21 32 13	55 29		85 64		26012 9 2289 82	11121 10 2460 19	1100 1 45 B	2 970 25 0 2629 24
utinovn 2702 49 utinovn 290 50	8018 8572	18 8	861 19 889 56	680 54%	883 41	1468 11		18 B02 7 1016	85 887 2 21 738 2		17 27 47 1253			16 16077 9 64065		M2 87 ND 89	15882 6712	9 10 18 81		1302 1218	9 73 34 143		7118 6 8178 14	4870 20 7804 17	2075 2340 3	0 4620 14 5 27778 16
uktovn 81g 51 uktovn 88 52	2192	31 2	035 64 276 31	1115	943 845	1730 27 43 BH	0 87 8		44 564 2 17 27 88	0 2541	14 1821 31 21	6 (8	19 14024	16 5602 9 8561	61	23 BK0 606 24	5678 2799	35 40 22 39	18 26	536	22 198 41 25	19 22	17466 5 3817 7	19341 3 2414 17	2974 2 626 1	6 6891 6 5 1879 21
unknown 30 53 unknown 40(1) 54	6196	9 5	531 22 075 38	1548	0+5 0+5	44 (90) 2930 11	2 8062	6 41 6 1209	800 26 80 23 460 1	8 2083	9 24 49 3024	0 1	14 268	22 192 045 45004	9	497 9 194 53	1036 6364	10 47		878	22 B9 87 46	N 88	1945 24 1511 24	8372 13 00 800	42 B 926 S	2 81380 6
abbox reductase-like protein 55 unknown 43(1) 56	4098	29 2	156 17 128 18	64953 887	12	7909 2	4 1061	10 4258 41 1150	9 2232 1 26 423 3	1 14574	20 16165 27 1025	6 2		6 260452 13 678	BK0	852 1.8 956 24	\$2062 3411	6 711 14 6	70 14		3 3050 11 69	10 34	300500 4 12051 19		33446 1198	6 45525 10 8 8028 81
unknown 47h 57 unknown 77 50 unknown 70(1) 59		910 1	539 34 038 41 605 6	1018 1418 1995	845 845	1020 21 491 51 730 31	7 840	23 826 43 88 1 60 811	29 624 3 000 26 00 19 24 00	0 800	19 34 10 26 26 50	0 10	12298	19 8141 12 6782 15 6805	22	778 40 28 8K5 27 8K5	3662 3342 2572	42 1 27 B1 25 10	18 21		27 21 30 B1 17 133		1124 50 5995 6 26255 2	1178 29 1821 28 7710 18	1047 1 2448 4 1480 1	2 1125 84 0 789 28 0 2768 15
unknown (1250) 60 unknown (1250) 60 unknown (1452) 61		7 0	600 6 007 19 677 045	4765	27	5430 1	2 4049	66 811 12 1872 20 426	19 24 90 2 25 90 27 27 60	0 15065	26 50 15 5272 14 32	9	2 151	15 6005 0KG 24661 54 8522	0	953 64 415 25	0011	14 103 25 11	17 5	600	87 20		2131 25 1700 13	2615 18	1502 1 35 B	0 2192 10
PAD 62 Invation 19 68	7028	8 17	204 14	47608	27	6419 B	6 4636	18 2648	18 205 1 18 EN7 1	6 50	843 6403 843 X1251	9	5 17%025 7 725810	7 96349 8 897522	12 1	201 7	29799	16 113	70 7	16298	4 1200	14 9	79514 6	61256 8	27668	4 80474 8 7 151368 6
uttorive 10(1) 64 uttorive 14 65	4040	17 1	081 8KS 151 8KS	709	845 845	4260 11	9 1105	15 1287	34 600 4 8x0 27 8x0	6 10055	80 2974 20 1184	6	2 197	BKD 4110	5	529 51	2258 81014	80 7 120	16 883			0 00	184 37 82602 7	2522 69 25563 10	904 2 83821	9 60 BK0 10310 10
uninovn 95a 66 uninovn 96(A) 67	14755	0 1	111 BHS 902 BHS	1091		10930 1	8 6891	5 5525	9 27 80	6 2953	25 4914 49 3707	4 1	18 6067	19 51825 040 9517	11 1	709 13 006 80	17258	11 166 16 70	9 8	000		40 21	8240 21 2207 5	1701 10 1506 37	3544	29317 9 1600 27
unknown 1507 68 PDAC-2 69	1595	29	774 883 290 11		8HD 16	65 BO		12 782 10 82214	18 804 2 12 8120		043 54 10 40477	2 (M		14 12888 12 680457	46	945 48 195 7	8092 676157	4 23 18 8108	72 64	35 E	45 277 6 28825		6617 67 1252710 6	11111 70 1207019 9	6947 4 828705	70 BKG 197662 12
unknown 1007 70 unknown 108 71	155811 410515	18 226 4 421	720 18 627 14	899050 919772		85278 2	2 90598 4 107669	11 51688 12 75420	8 4869 9 15110 1	5 887884 0 382599	75 181790 69 40015		16 8669758 9 899050	7 1022258 10 2069484	18 5 10 20	715 7 401 5	1127769 1177618	16 5957 6 9456		149810 319493	8 62248 7 74692		2924118 14 5219149 8	8347411 10 4075897 16	524079 746695	5 614621 22 5 473742 12
unknown 1096 72 unknown 1117 75	141068	3 10	874 11 071 24	400006 399528	12	57851 44075	4 48455	15 34750 18 29414	12 1905 1 9 1960 1	9 60547	26 17451 29 22707	K 1	2 330746	14 797499 12 495150	9 2	011 9	872072 285169	15 E704 0 2907	17 18	463693 471002	2 25222 7 82400	9 9	1606029 16 2409005 7	1826855 16	817756 840587	6 246495 15 0 117711 15
unknown 112 74 unknown 20e 75	104	843	775 51 427 (83	1090	843	6536 1 824 1	0 898	4 5364 20 628	12 24 80 87 25 80	0 4496	62 4000 32 11%	и 2	25 6017	96 19769 36 3346	55	413 14 31 BK0	8271	21 134	H 37	36 0		01 880	11702 26	87 800	4695 1	
067-2 76 unknown ceitagen 77	92079	1 10	414 9 202 7 783 045	12158 97791		7948 11 18772 1	5 1895	8 5801 19 16489	9 1005 2 7 6920 1	4 67	30 54 (85 25)	2 1	18 12956	10 656756 26 6480	40 6	128 27 126 10	21757	19 179 28 61	12 18	81 6		15 10	8098 14 2891 12	1959 48	5146 1 1265 1	4 1541 19
utictovn 46a 78 pielythransferase 79 unknown 2250 80	76692	2 25	782 0x5 112 15 176 0x5	1533 53041 6666	845 5 25	1157 2- 42403 1 2004 21	7 16394	19 466 11 10120 17 1061	57 27 00 18 3057 5 5 26 80	6 56000	0HS 50 16 3313 9 750	16	4 201304	5 0673 7 262764 7 44667	5 8	30 8K0 902 0 30 8K0	5061 65495 6399	83 14 0 854 12 98	17 7	2715	15 97 8 497 50 113	8 8	47974 9 73822 6 33585 5	26490 10 43020 9 40761 7	1857 21850 7560	4 8188 12 5 17907 7 8 3999 2
unknovn 2263 81 unknovn 2263 81 aer23 82	10085	16 1	227 845		80	8978 2 15786 6	9 3901	41 2118	5 25 040 82 27 840 11 5664 1	0 11718	7 80 17 80 17 2380	8		15 14147 6 193575	21	30 8K0 NP9 80 H8 4	9488	7 43	14 80	1282	28 220 11 2599	5 7	31000 0 37174 6 202007 6	10647 12		5 0997 2 7 6406 6 2 97246 18
PAG 83 Invarian beta A inhibin 64	10842	21	300 0KS 930 0KS 095 47	52694		6777 21 506 15	8 8179	25 1720	10 NO 5 00 29 00	0 852	18 8021	4	8 146199	9 41829 18 1925	3 1	427 24 31 BKG	20510	11 12	6 12	16799	7 1967	1 1	811596 9 10940 10	66276 14	21006	6 4449 26
p58 05 PTh-r-P 06	2060	25 2	067 25 707 9	7006	19	2643 5	1 1457	16 924	44 20 00 15 5439	6 18429	19 705 21 30	8 2	29 49274	7 45117 3 36407	2 1	445 12	29619	18 215 77 576	6 6	7%	16 1136 13 625	42 10	151202 9 58103 14	70089 11		0 86819 9
FAI-2 87 professe PI 00	1909	7 12	110 BHD 609 6	78796	5	1998 N 20155 -	0 456 4 1779	21 85	800 29 80 16 183%	0 1054 7 6663	121 404 67 47	5 2 3 (K	26 4668	7 400 8 18356	8K0 19 8F	114 B9 129 6	6301 2475	81 11 28 104	14 14 10 7	80 8 1004	80 1 19 50	12 8KD 77 84	2771 43 7366 19	2156 11 7011 11	43 B	0 76 8K0 0 8022 99
jrolo-catherin 89 RAR Jaho 90	18191 92616	15 12 7 88	808 82 908 9	27852 912	13	6088 21 89775 6	0 1901 6 17144	21 3225 9 19457	14 28 80 0 7516	6 8274 4 86605	38 80 22 4/7	19 1 15 1	11 21350 10 677011	10 84754 9 487087	6 1	005 84	7289 98289	10 49	15 11 12 0	5% 1559	25 92 33 617	Ni 2 10 9	10059 8 116855 7	11092 7 61544 9	4147 39652	8 4047 40 6 25178 8
RAR beta 91 95AT 92	65987	11 1	122 33 931 045	0025 12190	23 45	\$1695 ( 1955 (	6 11656 7 1957	20 10992 82 976	14 1124 1 27 631 1	9 61292 8 36561	25 1847 20 72	12 2 17 84	20 710627	7 657662 11 66116	1 2	190 20 099 22	54589 8070	0 882 11 53	14 89 17 15	1250 3031	29 110 6 252	43 0 12 0	24700 5 59001 11	18251 15 43826 5	9527 3 3850 3	1 10016 16 8 3321 6
TOF beta 1 binding protein	8728	3 11	0% 0KS 001 14	981	843 843	44 (H) 2517 1	7 1608	28 80 16 970	(K) 27 (K) 17 26 (K)	0 17021	040 53 19 1103	8 2	28 156812	21 5500 8 122799	50 26	28 8KG 81 8KG	1639 20156	51 1 15 606	48	1966	40 25 6 202	15 13 79 8	6340 15 84027 18	22789 11	00 B	0 124911 6
ribosonal protein P2 95 c-for 96	5508	12 7	611 10 464 24	25121	9	18202 6 2569 B	6 7560 0 967		9 2845 1 27 592 8	2 14001	14 14683 38 1167	1 1	17 408797 10 58216	7 827069 12 85698	10	154 7 1992 22	124768 11959	1 1909 6 30	H 26	749	7 8094 42 525	0 0	601761 5 20115 11	28290 0	1820	7 41950 15 8 13027 6
Jun B 97 ormys 98	75	6×3	038 10 207 045	27982 912	040	911 1	9 413	10 34	11 1941 2 960 035 1	5 626	18 94 29 115	0 04	0 0540	9 140820 61 177	BHG	052 20 20 DKG		9 107 20 1	(8 (8))	34 0		71 040	31792 5 65 (80)	99 B/0	12142 55 B	0 1764 30
catheprin D 99 odk4 100	38801	10	705 29 647 (83	12198		1075 2	7 11754	17 24049	BK0 26 BK0 22 BK0 3	2 85495	15 12 28 2511	8 6	66 25555	6 17520 14 100329	20 8	82 8KG	11566 81475	7 12	17 16	1582	57 15 18 55	10 10	4940 9 9647 23	6521 0	30%6 9670	4 27171 28
eyelin E 101 H-vai 100	2 8266	62 1	628 983 618 12	981	010	1190 11 52 00	0 52 0	K0 1982	21 208 2 90 24 90	0 57	20 100 845 81	0 86	19 275	960 9661 960 129	BX0	872 16 281 17	7727 2491	85 28 M	8 86	667	85 10 28 28	10 49	1009 11 75 BK0	100 810	111 B	0 50 BKO
TA1801 103 #59(1 104	2117	26	649 7 212 045	04422 1204	845	46664 15	9 66 8		6 1824 2 46 520 4		46 10400 085 9025 33 705	A) 2	22 287	6 178297 010 8541	20	597 19 29 DiG	62027 2099	9 562	62 045	626	9 4150 31 41	14 15	888780 6 2204 12	140 810	46765 3401	6 81064 5 9 2144 17
hep=90 105 bH=2 106	5 1464 41014	75 11 4	120 845 715 18	1215 1206	945 945	417 2		43 35 14 9737	905 617 3 9 1136 1		33 709 39 71	10 BK		50 12519 19 45256	4 3 1	638 26 152 17	2751 16008	21 50 8 356			42 7 83 166	27 12 91 22	62 80 6865 11	8422 21 7564 22	1795 2542 1	0 2151 7 2 3501 18

Tornadoes Produced	For Each Landfall Tr	opical Cyclone							
1990-2001									
* Unable to confirm to	ornadoes due to strai	ghtline wind damage							
Year	Storm Name	Classification	Min Pressure	Max Wind Speed	FO	F1	F2	Unclassified	TOTAL
	Allison	TS			18	5			23
	Barry	TS			0			0	
TOTAL					18	5	0	0	
2000	Gordon	H1			3	2	0	1	6
	Helene	TS			5	2	0	0	6
TOTAL					8	4	0		
1999	Harvey	TS			1	1	0	0	2
	Irene	НЗ			2	2		1	
	Floyd	H4			13	2		2	
	Dennis	H2			0	0		0	1
	Bret	H4			3	0	0	0	
TOTAL					19	5			
1998	Bonnie	НЗ			4	3	0	0	7
	Charley	TS			Ō	Ō			
	Earl	H2			6	2		Ō	9
	Frances	TS			7	3		Ö	11
	Georges	H5			20	8			28
	Hermine	TS			1	1			
	Mitch	H5			3	2	1	Ö	6
TOTAL					41	19			63
1997	Danny	H1			4	3	2	0	9
TOTAL					4	3			
1996	Arthur	TS			0	0	0	0	0
	Bertha*	НЗ			1	3			9
	Fran	НЗ			1	Ō			Ō
	Josephine	TS			15	Ō		Ō	
TOTAL					17	3		5	
1995	Allison	H1			1	5	0	0	6
	Dean	TS			1		Ō	Ö	1
	Erin	H1			10				
	Jerry	TS			2				2
	Opal	H4			16	1	1		
TOTAL					30	8			39
· - · · · ·							· ·		
1994	Alberto	TS			0	0		0	0
	Beryl	TS			Ō				
	Gordon	H1			Ō				
TOTAL					Ö				
							· ·	1	1
1993	Arlene	TS			0	0	0	0	0
TOTAL					Ō			Ö	
IUIAL					<u> </u>	<u> </u>	<del>ا ا</del>	l – – – – – – – – – – – – – – – – – – –	U U

					Tax	onomic R	aw Abunda	ance					Lith	ology
Locality	Diaco- dexis	Hyop- sodus	Haplo- mylus	Ecto- cion	Phena- codus	Homo- galax	Hyraco- therium	Cantius	Teton- ius	Estho- nyx	Croco- dilia indet.	Sum	Paleo- sol	Sand- stone
		300003	mytus	cion		guiun		Cuntus	140	neya	muet.	Sum	501	stone
FG 083	2	9	0	1	2	0	17	1	1	0	9	42	1.76	5.94
FG 084	11	30	14	4	9	30	81	25	2	7	8	221	2.19	11.36
FG 085	1	3	1	3	1	4	15	1	0	2	0	31	2.64	10.42
FG 090	0	1	0	0	2	4	6	0	1	1	0	15	1.74	5.48
MP 023	12	28	20	1	5	4	43	10	0	2	19	144	1.00	11.95
MP 025	1	5	1	0	3	3	14	2	1	0	9	39	1.41	18.95
MP 034	1	23	5	0	2	0	11	4	0	1	2	49	2.13	0.95
MP 035	3	1	1	0	1	0	5	0	0	1	5	17	1.40	2.50
MP 046	2	7	1	0	1	1	3	6	0	0	1	22	1.74	9.82
MP 130	2	8	5	0	2	5	15	4	1	3	4	49	1.27	0.00
MP 131	0	4	1	0	0	0	0	0	0	0	0	5	1.59	4.21
MP 133	0	11	6	2	1	1	5	4	0	0	0	30	1.39	13.79
MP 149	9	28	5	1	6	13	31	4	0	2	26	125	1.48	7.46
MP 198	0	0	0	0	0	2	5	2	0	1	1	11	1.11	0.77
MP 199	2	0	2	1	2	0	2	1	0	0	0	10	1.52	5.70
MP 200	0	1	1	1	1	3	14	0	0	0	3	24	1.34	8.18
MP 201	5	25	11	2	1	3	35	6	2	1	10	101	1.44	15.05
MP 205	2	2	1	1	0	0	6	0	0	0	0	12	1.68	9.35
MP 206	5	7	0	1	1	0	6	0	0	0	1	21	1.04	1.67
MP 208	0	0	6	0	0	3	13	0	0	1	3	26	1.06	18.92
MP 213	5	6	3	3	2	4	21	2	0	0	3	49	1.74	16.47
MP 215	0	5	2	0	0	0	4	2	0	0	2	15	1.47	9.33
MP 220	0	1	0	0	1	3	7	1	0	1	7	21	1.27	11.30
MP 222	1	3	0	2	2	0	7	3	0	0	3	21	0.98	14.00
MP 223	11	40	20	4	9	4	105	13	0	10	21	237	1.38	10.80
MP 225	3	5	2	0	4	0	14	6	0	6	2	42	1.37	12.78
MP 226	7	9	6	2	0	0	9	6	1	2	3	45	1.35	6.67
MP 227	0	6	2	0	0	0	0	2	0	1	0	11	1.06	11.04
MP 228	6	13	1	4	3	2	32	1	1	1	4	68	2.05	8.73
MP 229	5	20	11	1	0	2	14	7	0	1	1	62	1.25	25.60
MP 230	0	0	1	0	0	0	1	1	0	0	0	3	1.48	10.00
MP 231	1	6	6	0	0	4	4	3	1	0	5	30	1.44	4.94
MP 232	0	3	0	1	0	0	3	1	0	0	0	8	1.22	13.33
MP 245	9	21	14	7	4	1	51	10	8	2	34	161	1.51	4.84
MP 246	2	1	1	0	0	0	2	1	0	0	0	7	1.10	0.00
MP 247	0	4	3	0	3	1	7	0	1	1	4	24	1.61	13.91
MP 250	0	2	1	3	1	1	14	0	0	0	3	25	2.18	2.13
MP 251	10	8	4	1	2	4	18	8	0	3	0	58	2.09	5.52
MP 252	5	3	6	4	4	4	23	4	0	2	8	63	0.98	10.26
MP 267	2	0	1	0	0	0	7	2	1	0	0	13	1.60	12.94
MP 269	2	1	4	3	1	3	5	1	1	1	0	22	1.41	18.97
MP 270	2	9	8	1	0	3	9	8	1	1	0	42	1.58	1.11
MP 273	4	6	9	4	0	2	27	6	2	3	8	71	1.71	15.53
MP 276	2	0	0	0	1	1	3	0	0	2	3	12	1.33	11.19
MP 277	2	10	4	2	1	0	24	7	1	2	0	53	2.13	8.21
MD 070	-	F	0	0	-	9	7	0	0	0	0	01	0.00	4 70



- Public statistics is the most expensive content produced
   o tax funded
  - curated by professionals
- Users can't find it
  - numerous formats (pdf, html, db's, behind interfaces, subscriptions, etc.) --> hard to index
  - hard to use: costly to use, expertise required
- Work with data providers to curate, visualize and publish statistical data

# Organize the world's public data and make it universally accessible and useful.

## Accessible (1)



#### Example: <u>unemployment rate illinois</u>

ogle	unemployment rate illinois	×	Search
/thing es	unemployment rate illinois unemployment <b>tax</b> rate illinois unemployment rate illinois <b>by county</b> <b>current</b> unemployment rate illinois		
)S	About 897,000 results (0.33 seconds)	Ac	lvanced search
ping View, CA cation	15% 8.8% of the labor force - Not seasonally adjusted - Dec 2010 Source: U.S. Bureau of Labor Statistics Disclaimer www.google.com/publicdata		
images heel earches	[PDF] Illinois Unemployment Rate in December 2010 2 Q File Format: PDF/Adobe Acrobat - Quick View Jan 21, 2011 Illinois monthly labor force, unemployed and unemployment rates for years 1976-2009 have been revised as required by the www.ides.state.il.us/economy/cps.pdf - Similar		
ch tools	Local Area Unemployment Statistics (LAUS) - LMI Source: Your 22 Q Local Area Unemployment Statistics: LAUS. Illinois and Chicago Metropolitan Imi.ides.state.il.us/laus/lausmenu.htm - Cached - Similar		
	Welcome to IDES 2 Q The Illinois Income Tax rate for individuals has increased from 3 percent to Apply for Unemployment Insurance Online UI Services - Teleserve www.ides.state.il.us/ - Cached - Similar		
	Illinois County Unemployment rates - LMI Source: Your Source for 2 9 Illinois County Monthly Unemployment Rate Ranking Report. The report below Imi.ides.state.il.us/rank.htm - Cached - Similar		

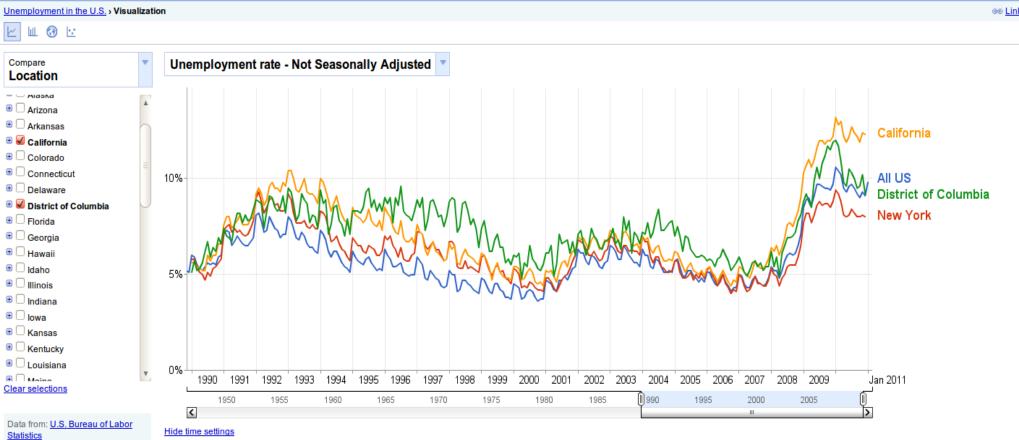
# Organize the world's public data and make it universally accessible and

"The greatest value of a picture is when it forces us to notice what we never expected to see."

- John W. Tukey (1977)

## Useful (1)

#### Unemployment in the USA



Last updated: February 4, 2011

©2011 Google - Terms of Service - Privacy - Disclaimer - Discuss

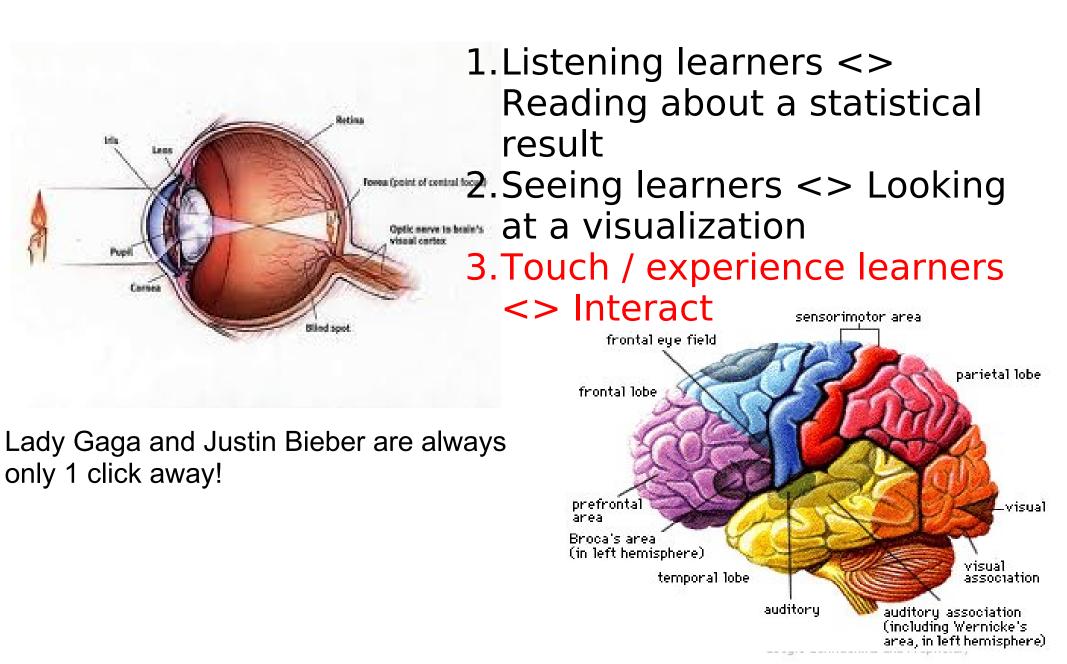
#### Map visualization **Embedding**, Sharing

@ Link

Google

#### Useful (2)

Google



**Organize** the world's public data and make it universally accessible and useful.





#### Dataset Publishing Language (DSPL)

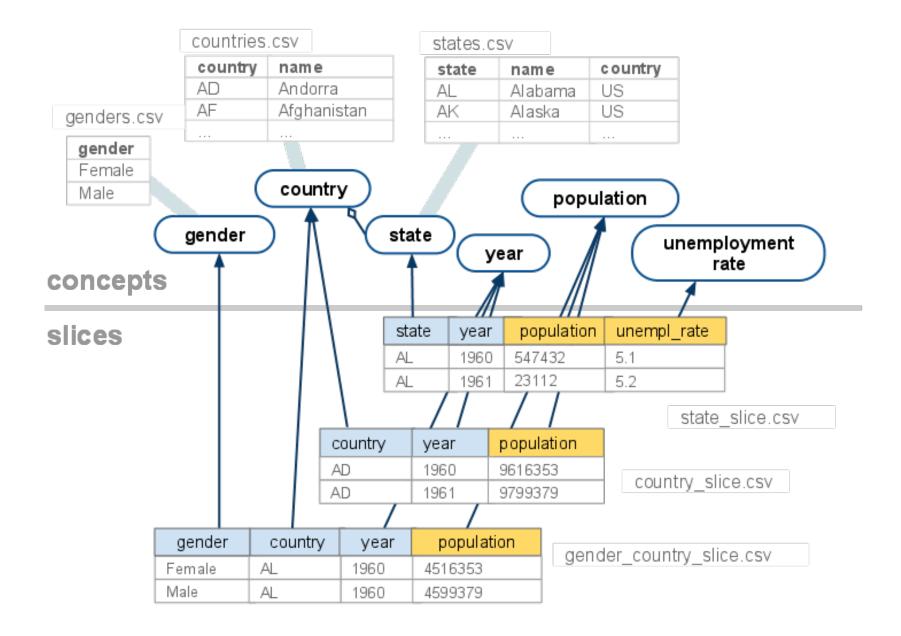
- Designed for interactive exploration and visualization
- Released under BSD, open source license
- Combines data tables (CSV) with metadata (XML)
- Works best with categorical, time series data ...
   ... but can represent generic collections of tables too

#### Metadata contents

- Dataset info (name, description, URL, etc.)
- Provider info (name, description, URL, etc.)
- Concepts
  - Dimensions (e.g., "time", "country", "gender", "state") --> canonical concepts
  - Metrics (e.g., "population", "unemployment")
- Slices
- Tables
- Topics

Organize (2)

Google



## Organize (3)



```
<?xml version="1.0" encoding="UTF-8"?>
<dspl xmlns="http://schemas.google.com/dspl/2010" ...>
 <info>
  <name>
   <value>My statistics</value>
  </name>
  <description>
   <value>Some very interesting statistics about countries</value>
  </description>
  <url>
   <value>http://www.stats-bureau.com/mystats/info.html</value>
  </url>
 </info>
 <provider>
  <name>
   <value>Bureau of Statistics</value>
  </name>
  <url>
   <value>http://www.stats-bureau.com</value>
```

</url>

. . .

</provider> ....

- Bundle (zip) xml + csv files
- Upload on http://www.google.com/publicdata/admin
- --> Instant visualization!

See http://code.google.com/apis/publicdata for more details



#### Dataset



#### Local Employment Dynamics for the State of Maine

#### All data

By Age Group (24) By County (24) By Gender (24) By Industry (24) By Metro (24) By WIA (24) US Census LED Quarterly Workforce Indicator report. Prepared by the US Census Bureau LED Program/Google based on data downloaded from the US Census Bureau LED Program.

Data from U.S. Census Bureau LED Program - Last updated: Jan 11, 2011

#### Data

A1 - Beginning of Quarter Employment: Counts - 2 - 11 - 11 A2 - End of Quarter Employment: Counts - 2 - 11 - 11 A3 - Full-Quarter Employment(Stable):Counts - 2 - 11 - 11 A4 - Employment - Reference Quarter: Counts - 2 - 11 - 11 B1 - Hires All:Counts - 2 - 11 - 11 B2 - Hires New:Counts - 2 - 11 - 11 B3 - Hires Recalls:Counts - 2 - 11 - 11 B4 - Separations:Counts - 2 - 11 - 11 B5 - Hires All (Stable):Counts - 2 - 11 - 11 B6 - Hires New (Stable):Counts - 2 -B7 - Separations (Stable):Counts - 2 - 11 - 11 B8 - Turnover (Stable): Ratio - V -C1 - Firm Job Gains:Counts - 🗠 - 🔟 - 🔛 C2 - Firm Job Loss:Counts - 📈 - 🔟 - 🔛 C3 - Firm Job Change: Net Change - V -C4 - Firm Job Gains (Stable):Counts - 2 - 10 - 10 C5 - Firm Job Loss (Stable):Counts - 2 - 11 - 11 C6 - Firm Job Change (Stable):Net Change - 2 - 11 - 11 D1 - Full-Quarter Employment (Stable): Average Monthly Earnings - 1/2 - 1/1/2 - 1/2 D2 - End of Quarter Employment: Average Monthly Earnings - 1/2 - 1/1/2 - 1/2 D3 - Hires All (Stable): Average Monthly Earnings - 2 - 11 - 11 D4 - Hires New (Stable): Average Monthly Earnings - 2 - 11 - 11 D5 - Separations (Stable): Average Monthly Earnings - 2 - 11 - 11 D6 - Hires All: Average Change-in Monthly Earnings - 2 - 11 - 11





#### Improving public data benefits providers and users Google

- Broader distribution and scale
- Compelling visualizations of data (application embed)
- Google manages traffic

# Public data providers

- Language support
- Enabled data feed (synchronization)
- Internal use: visual quality control
- Low development costs

#### **Users**

- Enhanced discovery
- Greater ability to explore data
- Share and embed visualizations

School comparisons



Where should I send my daughter to school?

Poverty



Are our policy and economic decisions improving people's lives?

**Global warming** 



What impact are we having on the environment?

Thank you! jurgen@google.com