

Quo Vadis, libre software? A look into Debian's past to see what future will bring

Linuxtag
Karlsruhe (Germany)

Alvaro Navarro , Gregorio Robles, Diego Barcelo
{anavarro,grex,dibarman}@gsync.escet.urjc.es



- Libresoft @ URJC: Research group in «Libre Software Engineering».
- CALIBRE: Coordination Action for Libre Software.

Summary

- Studying evolution by:
 - Number of contributors/developers
 - Counting source lines of code (SLOC)
 - Counting packages, files, programming languages
- Effort estimations (Basic COCOMO)
- Studies about code authorship
- Conclusions

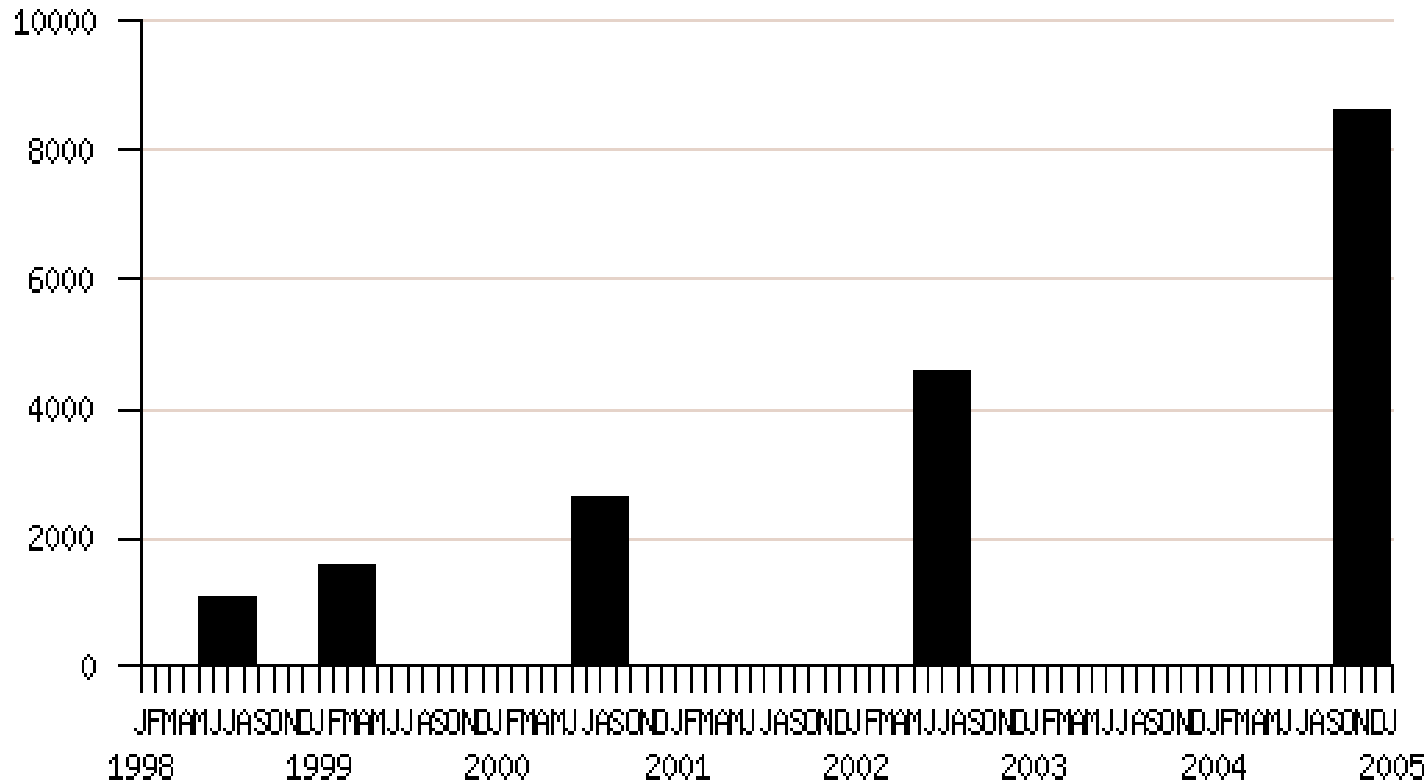
About Debian

- A Libre Software Operating System
- Based in libre kernels (Linux, HURD, *BSD) and libre applications (GNU tools, Mozilla, etc)
- Created and maintained by volunteers
- Several contemporary Debian versions (stable, testing, unstable, experimental)
- Why Debian?
 - It's a very representative libre project
 - We only study *main* branch

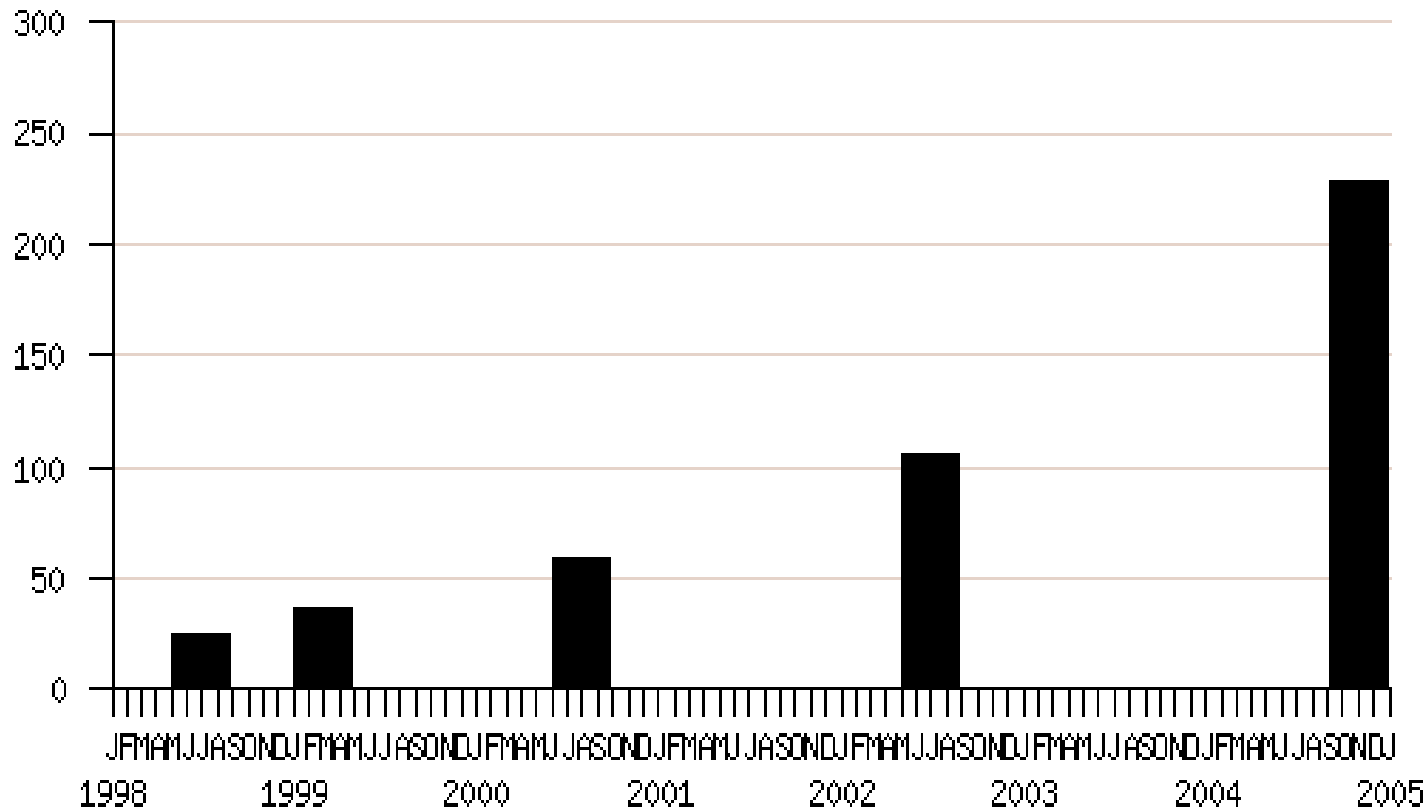
Studying Debian: Methodology

- Methodology: download, unpack, count, sum
- Tools:
 - SLOCCount: it counts LOCS and identify programm languages
 - PyTernity: heuristic analysis of source and documentation files

Debian Packages



Debian size: MSLOC



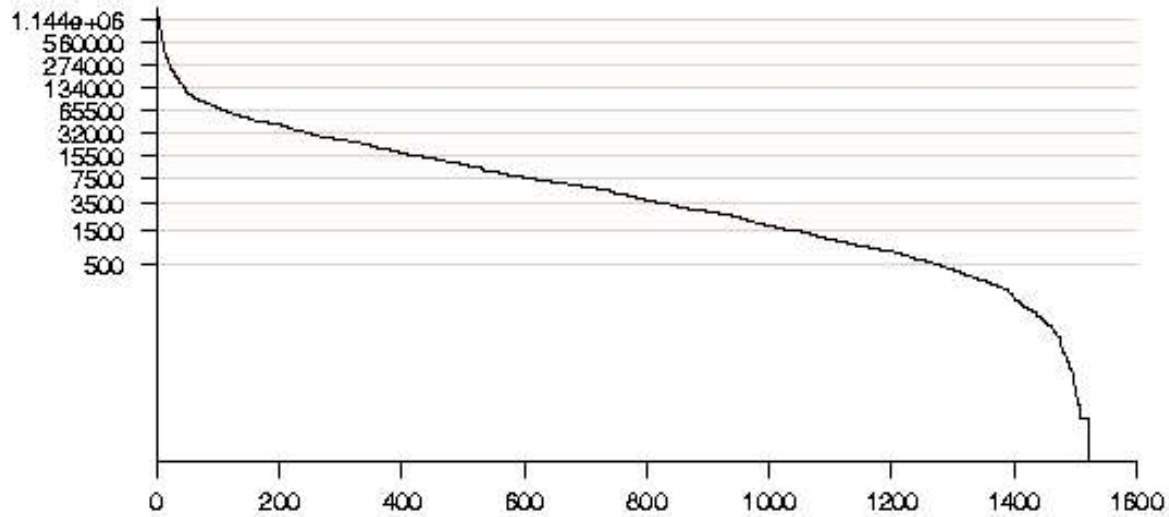
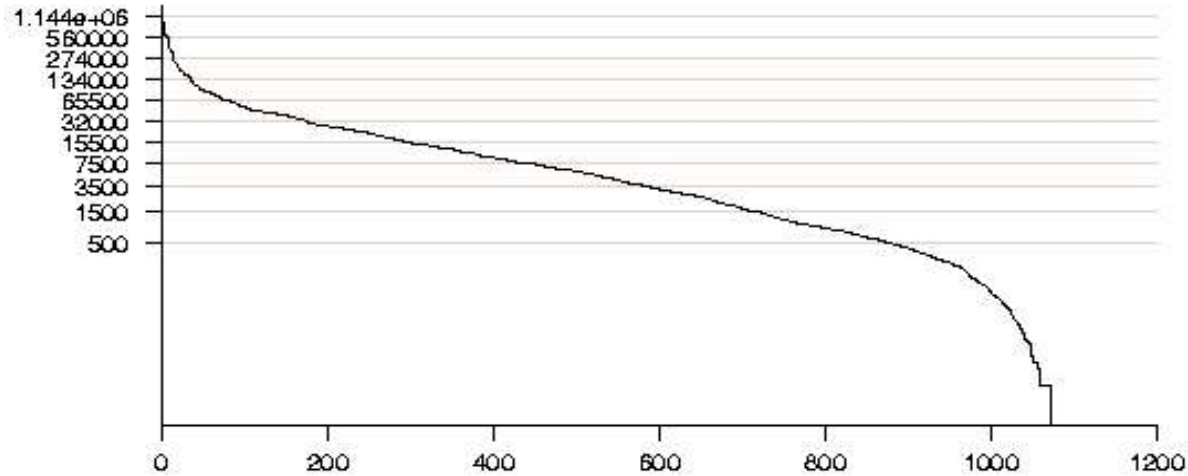
Mean package sizes

Release	Date	Packages	MSLOC	Mean SLOC
2.0	July 1998	1096	25	23050
2.1	March 1999	1551	37	23910
2.2	August 2000	2611	59	22650
3.0	July 2002	4579	105	22860
3.1	June 2005	8633	229	26584

- The mean size (in SLOC) of four releases remains almost constant in time
- Debian 3.1 has a bigger mean size

Package size distribution

Debian 2.0

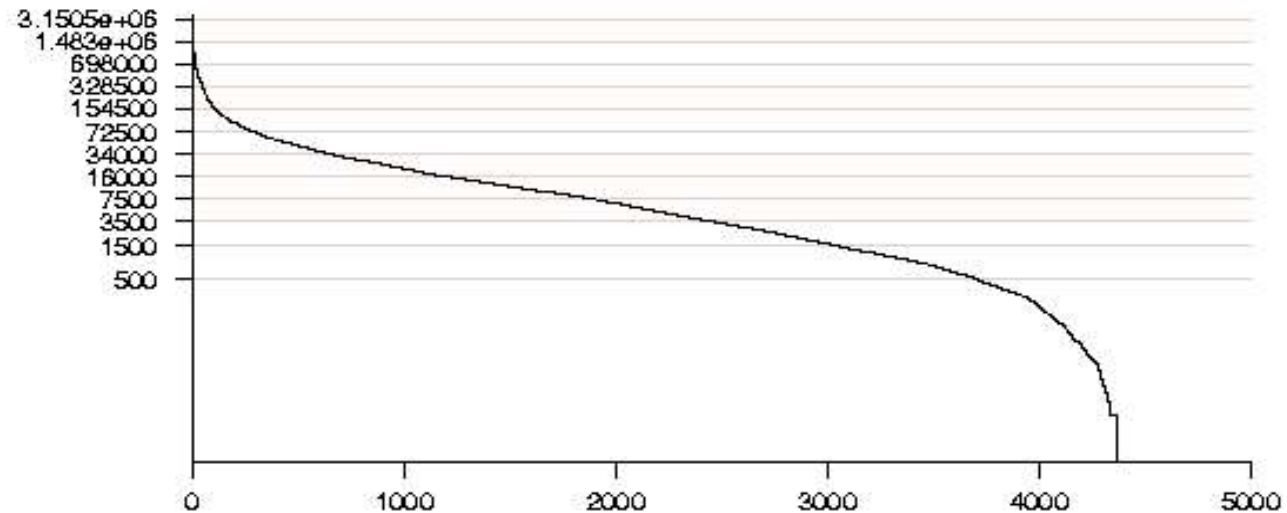
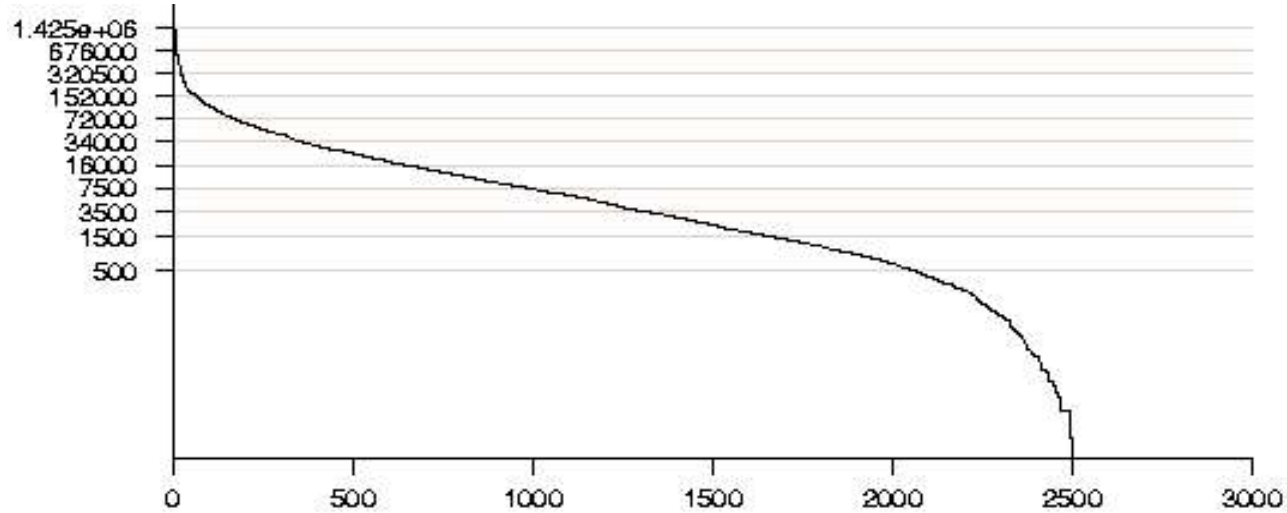


Debian 2.1



Package size distribution

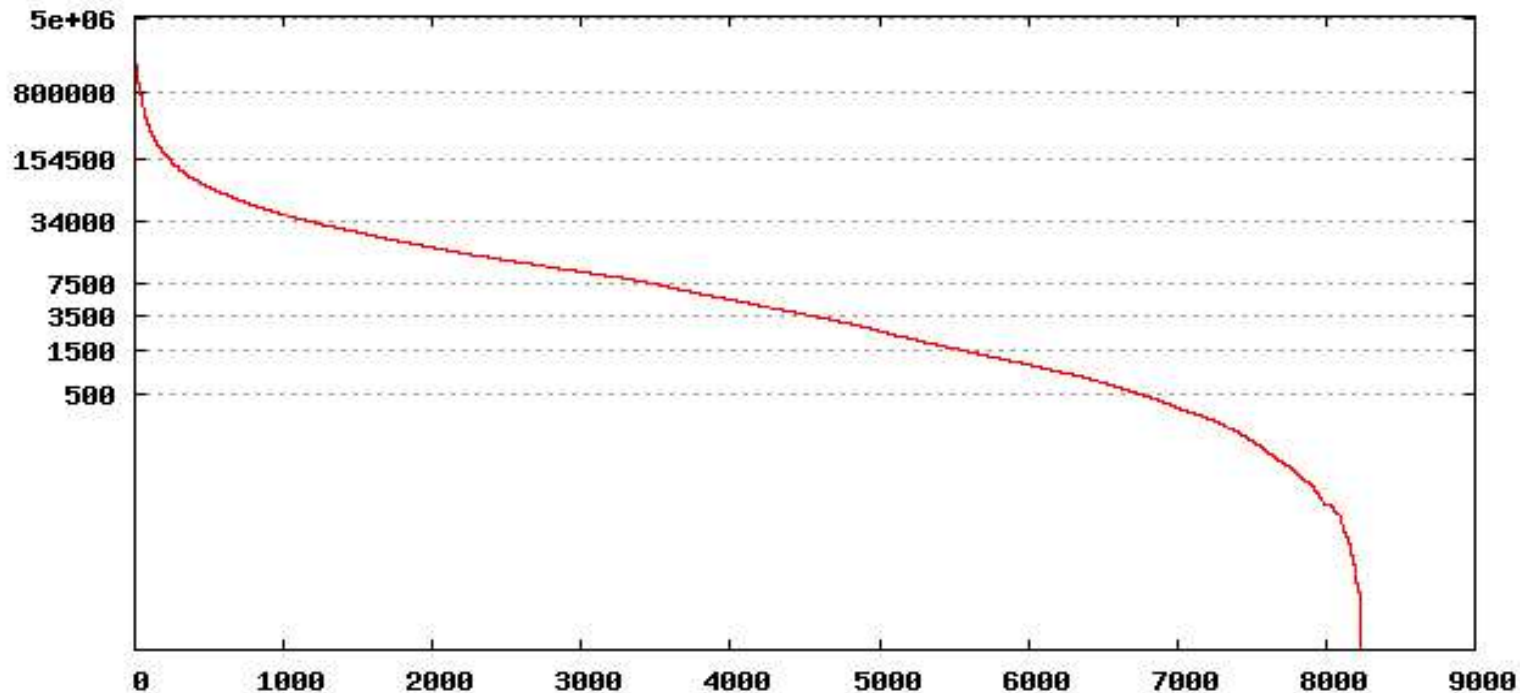
Debian 2.2



Debian 3.0



Package sizes



Debian 3.1



- Mean package size remains (almost) constant:
 - Debian 2.0: 23050
 - Debian 3.1: 26584

Top 10 packages in Debian 2.0

Rank	Package	Version	Size	Files
1	Xfree86	3.3.2.3	1200000	4100
2	Xemacs20	20.4	780000	1794
3	Egcs	1.0.3a	705000	4437
4	Gnat	3.10p	600000	1939
5	Kernel-source	2.0.34	575000	1827
6	Gdb	4.17	570000	1845
7	Emacs20	20.2	560000	1061
8	Lapack	2.0.1	400000	2387
9	Binutils	2.9.1	390000	1105
10	Gcc	2.7.2.3	350000	753

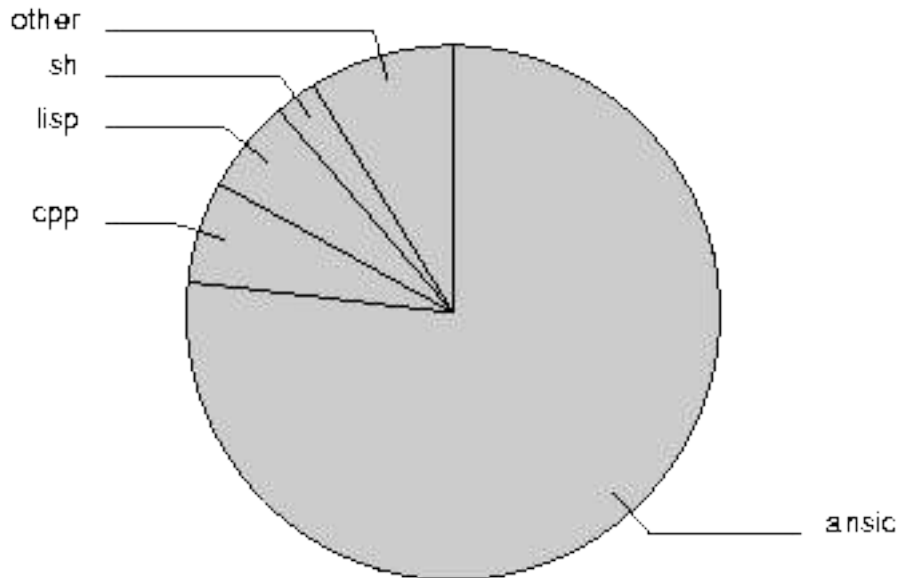
Top 10 packages in Debian 3.0

Rank	Package	Version	Size	Files
1	Kernel-source	2.4.18	2600000	8527
2	Mozilla	1.0.0	2400000	11095
3	Xfree86	4.1.0	1925000	6493
4	Pm3	1.1.15	1500000	7382
5	Mingw32	2.95.3.7	1300000	6840
6	Bigloo	2.4b	1050000	1320
7	Gdb	5.2.cvs...	1000000	2767
8	Crash	3.3	975000	2740
9	Oskit	0.97.2...	925000	5584
10	Ncbi-tools6	6.1.2001.	825000	1178

Top 10 packages in Debian 3.1

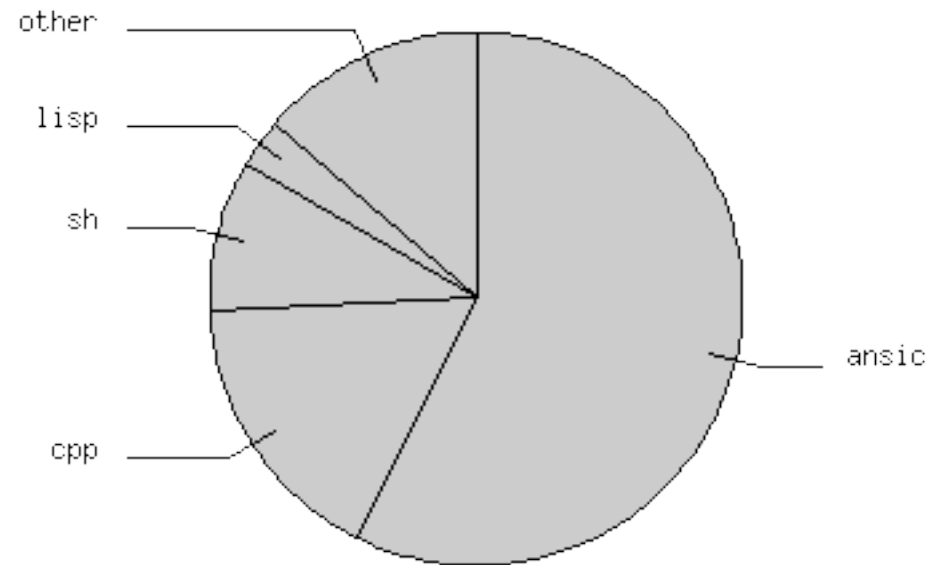
Rank	Package	Version	Size	Files
1	Openoffice.org	1.1.2...	5180000	19011
2	Kernel-source	2.6.8	4040000	13974
3	nvu	0.8	2470000	10780
4	Mozilla	2:1.7.3	2440000	10713
5	gcc-3.4	3.4.3	2420000	22992
6	xfs-xtt	1	2330000	7081
7	xfree86	4.3.0...	2320000	7216
8	vnc4	4.0	2060000	6790
9	insight	6.1+...	1690000	4063
10	kfreebsd5-src	5.3	1630000	4968

Programming languages

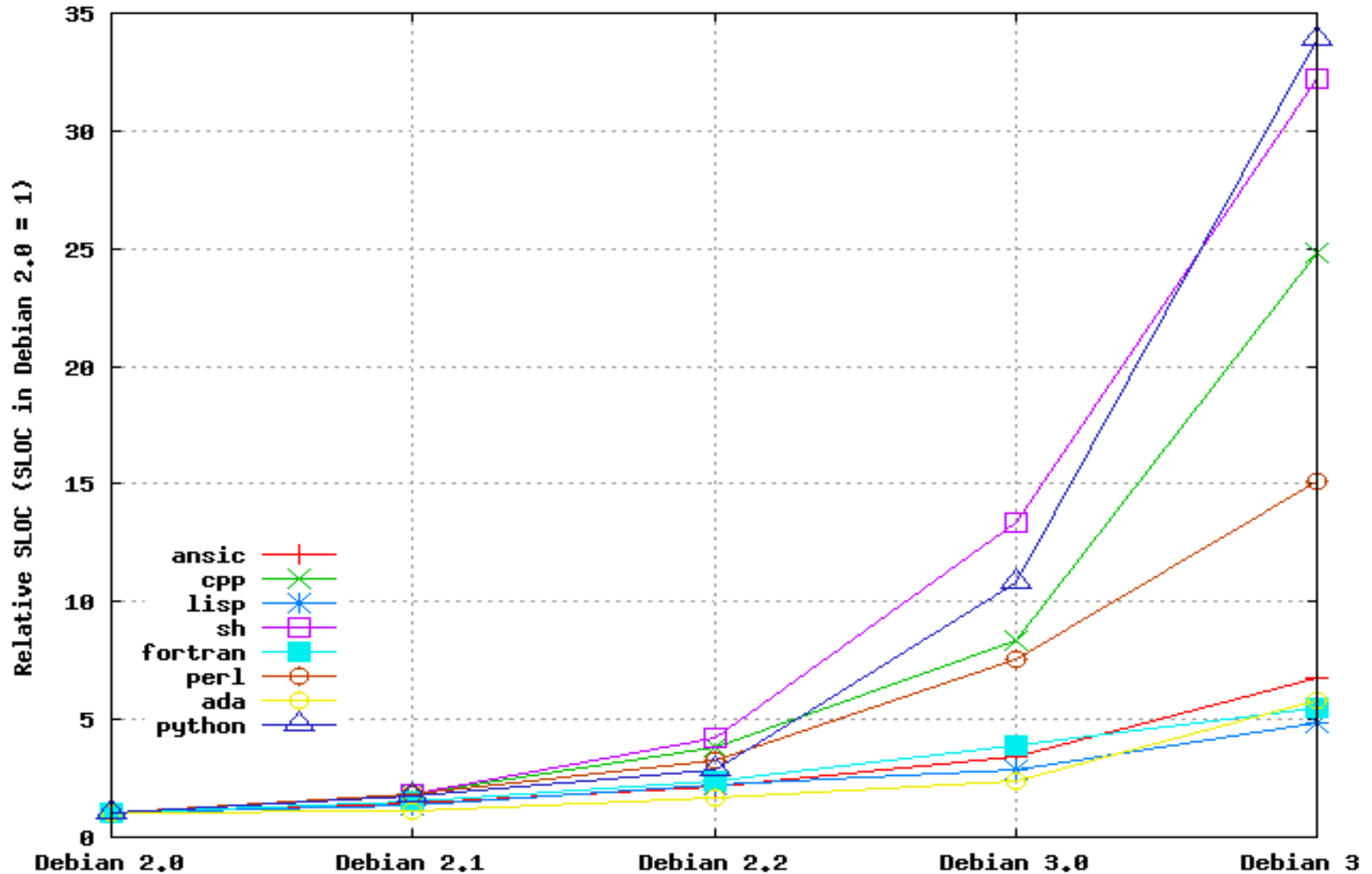


Debian 3.1

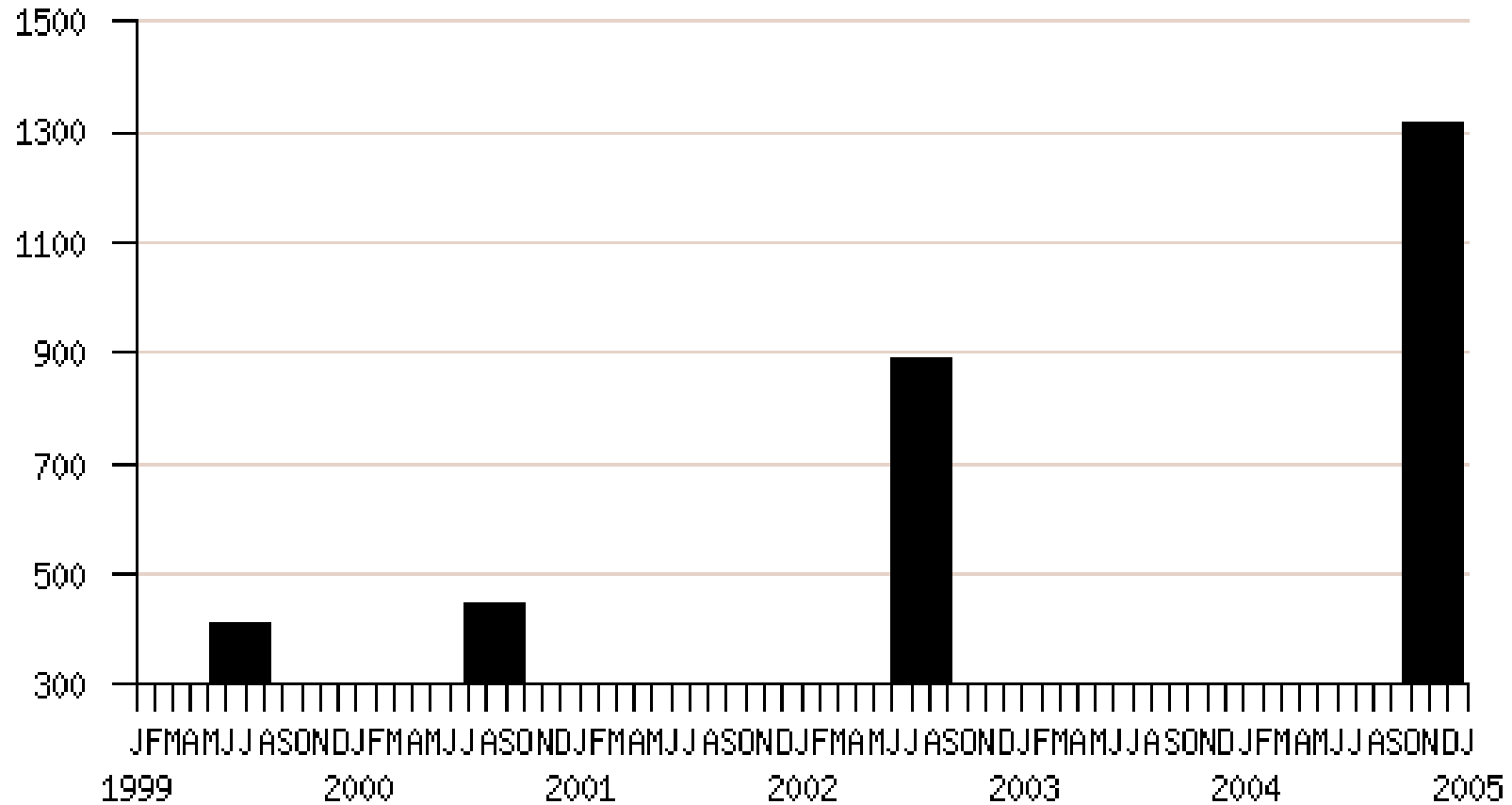
← Debian 2.0



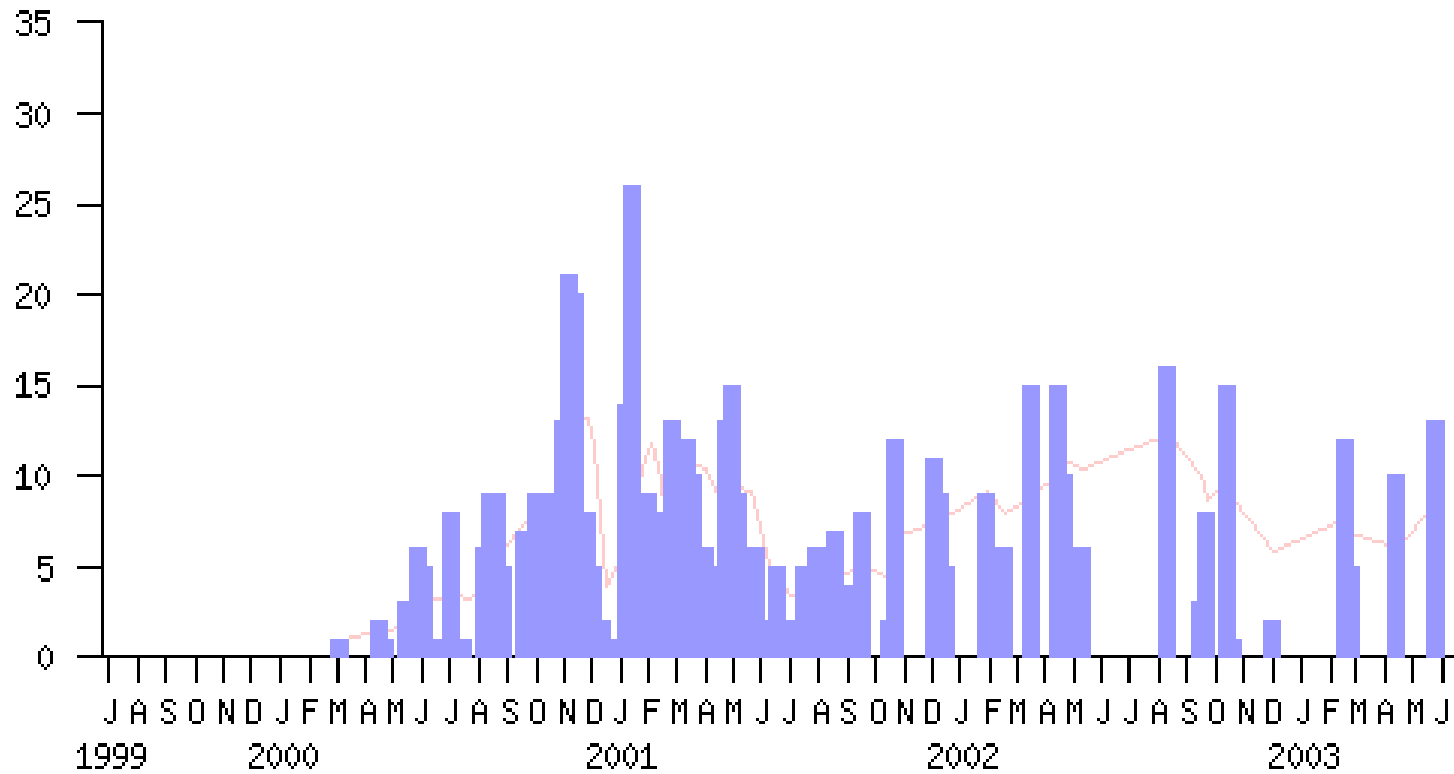
Programming languages



Debian developers



New Developers



Developers per country

Country	07/01/99	07/01/00	07/01/01	07/01/02	06/20/03	11/04/04
EEUU	162	169	256	278	297	311
Germany	54	58	101	121	136	151
UK	34	34	55	63	75	80
France	11	11	24	44	51	57
Australia	23	26	41	49	52	55
Canada	20	22	41	47	49	53
Spain	10	11	25	31	34	37
Japan	15	15	27	33	33	36
Italy	9	9	22	26	31	34
Holland	14	14	27	29	29	30
Sweeden	13	13	20	24	27	28

Top Cities

City	Country	Developers
Paris	France	15
Cambridge	U.K	15
Madrid	Spain	13
Stuttgart	Germany	10
Berlin	Germany	9
Munich	Germany	9
Tokio	Japan	9
Seattle	EEUU	9
Viena	Austria	7
Sidney	Australia	7
Montreal	Canada	7

Copyright analysis with pyTernity

- Download and uncompress Debian packages
- Header analysis for each source code and doc files.
- Search for authorship patterns, using some heuristics.
- Results saved in database.

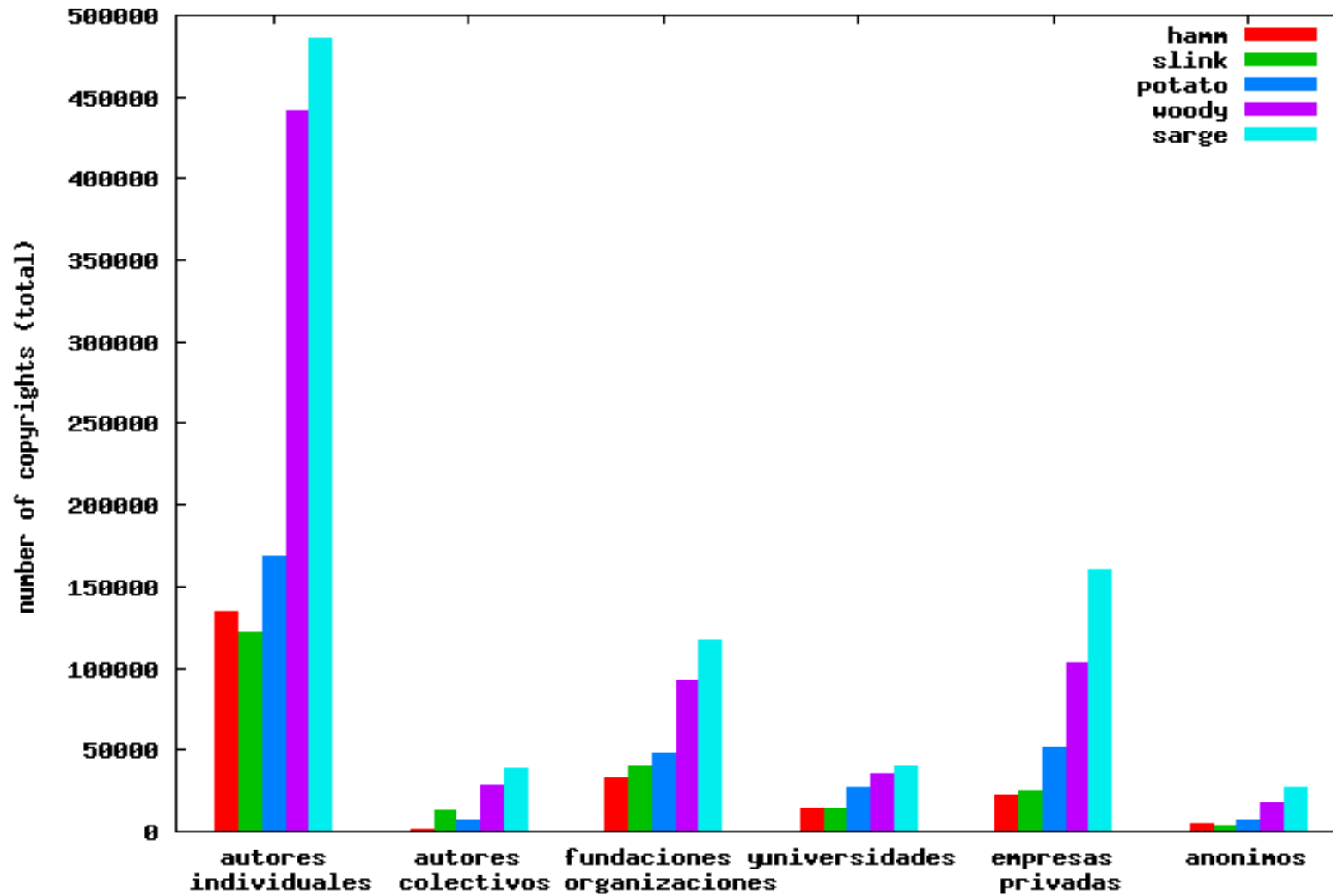
Ranking Debian 2.0 (Hamm)

free software foundation inc.	28853
sun microsystems inc.	5639
regents of the university of california	4731
x consortium inc.	3378
silicon graphics corp.	1974
digital equipment corp.	1733
rickard e faith	1722
ulrich drepper	1508
stephane rehel	1498
james clark	1124

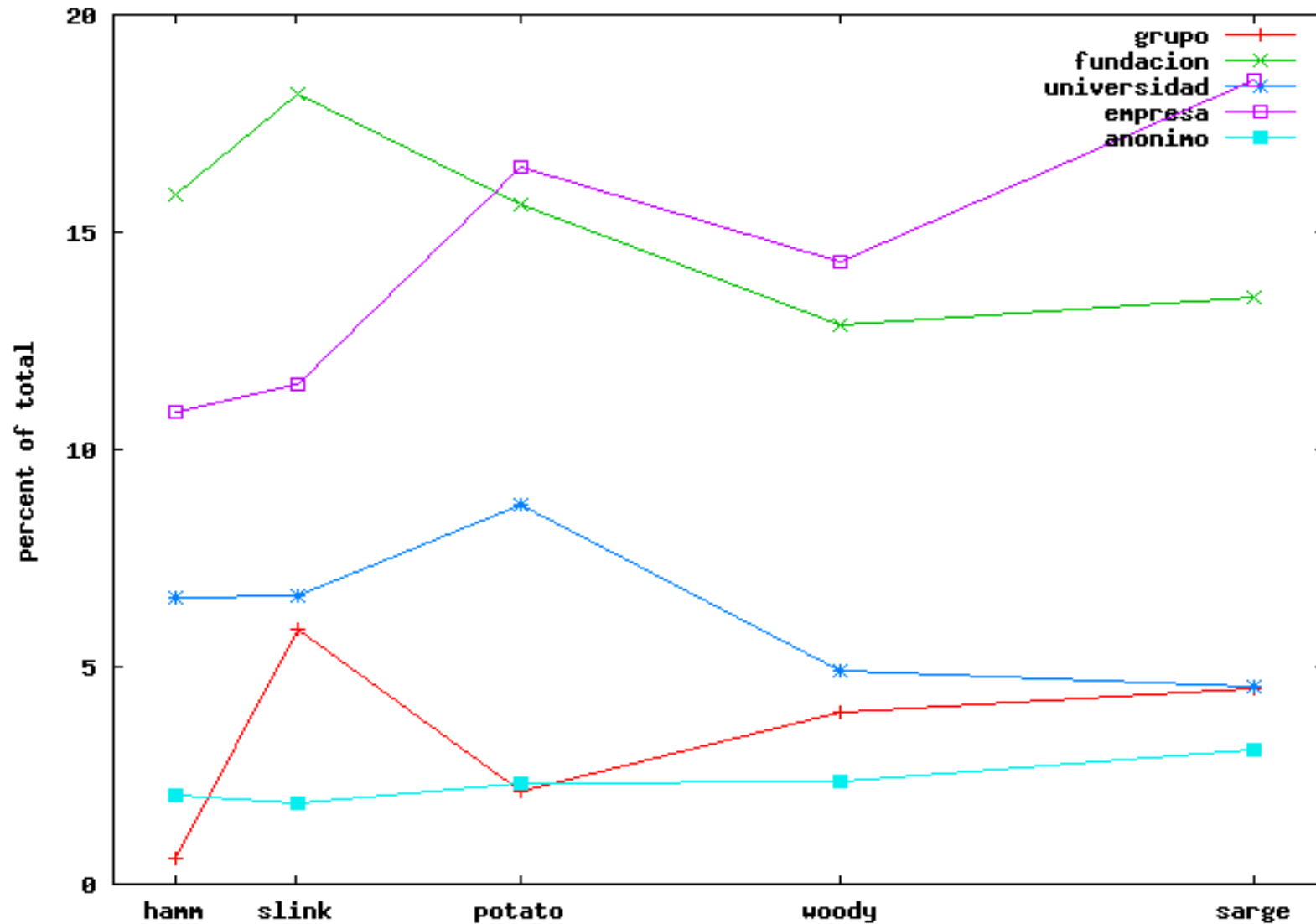
Ranking Debian 3.1 (Sarge)

free software foundation inc.	104234
sun microsystems inc.	33173
ibm corp.	8075
netscape communications corp.	6940
sleepycat software inc.	6408
massachusetts institute of technology	6391
sap ag	6284
regents of the university of california	5817
trolltech as	5793
keith bostic	5635

Copyrights



Evolution of copyrights



Tools

- Main source code analysis:
 - Done with Sloccount.
 - <http://www.dwheeler.com/sloccount/>
 - aptitude install sloccount
- Copyright analysis:
 - Done with pyTernity.
 - <http://forja.linex.org/projects/pyternity> (available soon)
- More tools:
 - <http://libresoft.urjc.es/>

Conclusions

- Debian duplicates its size every 2 years
- Package sizes remain (almost) constant
- C is the main language, but other languages (mainly interpreted) grow more quickly
- Debian is more 'european' every year.
- 55-60% copyrights belong to individual developers
- Enterprise collaborations are growing, more than other groups

References

- Libre Software Engineering website at URJC
 - <http://libresoft.urjc.es>
- Debian count results (from Debian 2.0 to 3.1):
 - <http://libresoft.urjc.es/debian-counting/>
- COCOMO:
 - [Boehm81]: Book 'Software Engineering Economics', Prentice Hall