

(Más) Historia del Jugquete

Un repaso al proyecto Debian GNU

DebConf-es II
Guadalajara, 17 diciembre 2005

This work has been funded in part by the European Commission, under the CALIBRE CA, IST program, contract number 004337, by the Universidad Rey Juan Carlos under project PPR-2004-42 and by the Spanish CICyT under project TIN2004-07296.

Juan José Amor, Jesús M. González-Barahona, Israel Herraiz
Grupo de Sistemas y Comunicaciones – Universidad Rey Juan Carlos
{jjamor,jgb,herraiz}@gsync.escet.urjc.es

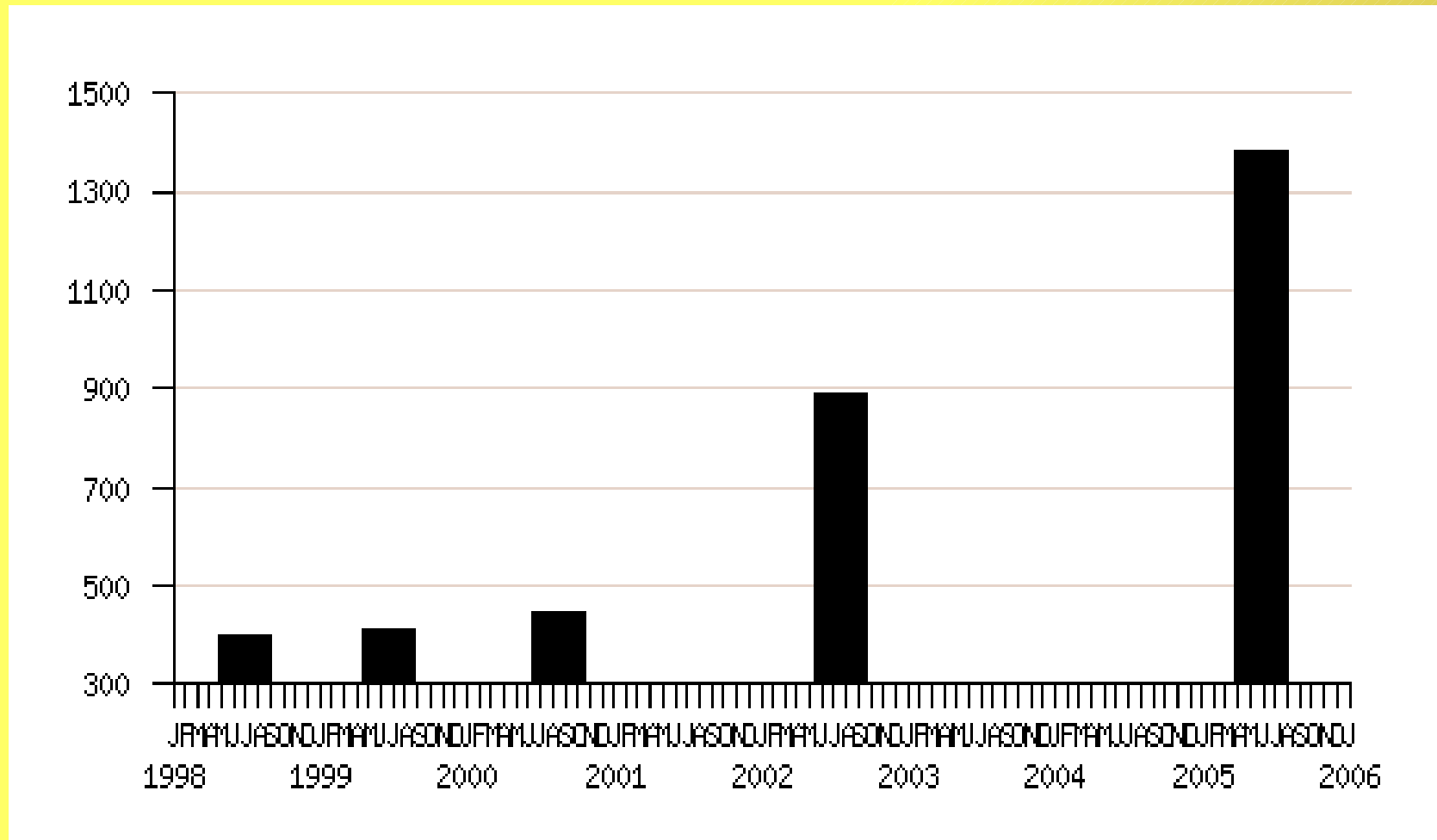


- Libresoft @ URJC: Grupo de investigación en «Ingeniería del Software Libre».
- CALIBRE: Coordination Action for Libre Software.

Resumen

- Evolución de Debian GNU según:
 - Número de colaboradores/desarrolladores
 - Número de líneas de código fuente (SLOC)
 - Número de paquetes, ficheros, uso de lenguajes
- Estimaciones de esfuerzo (COCOMO Básico)
- Comparativa con otros sistemas
- Estudios acerca de la autoría del código fuente
- Estudios de las listas de correo

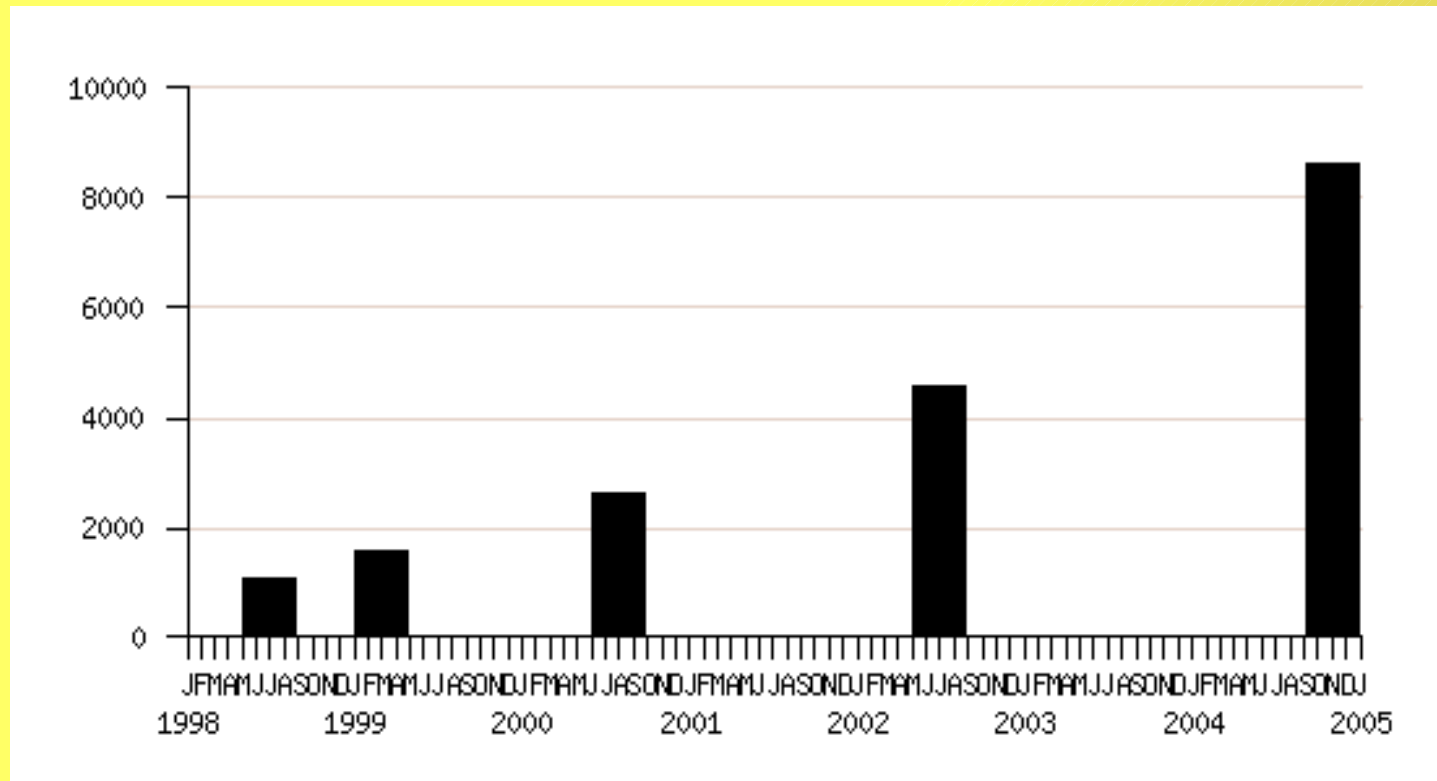
Desarrolladores en Debian



Estudio de Debian: Metodología

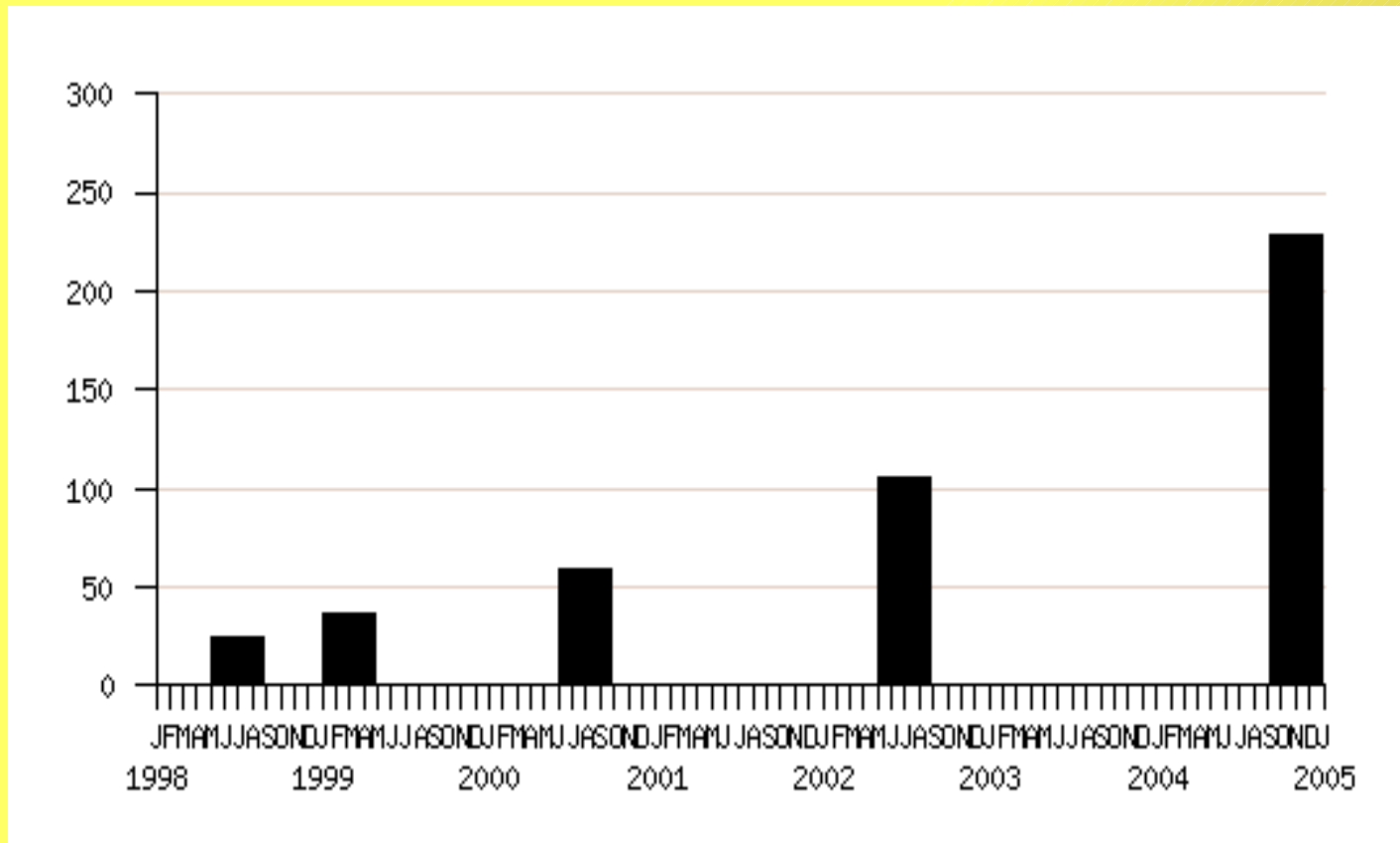
- En resumen: bajar, desempaquetar, contar, sumar
- Resultados: número de líneas físicas (SLOCs)
- Identificación de lenguajes de programación
- Identificación de ficheros idénticos (con md5)
- Estimaciones mediante técnicas de la ingeniería del software clásica (COCOMO básico)
- Identificación de autoría del código (mediante las notas de copyright)

Paquetes en Debian



(Paquetes de Debian en las cinco últimas versiones estables)

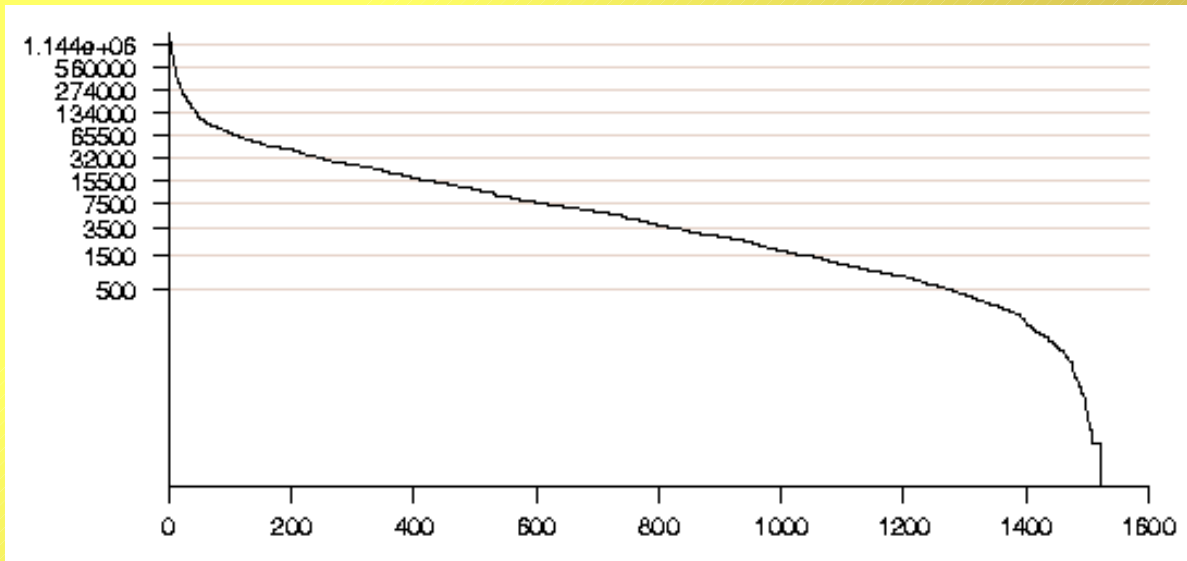
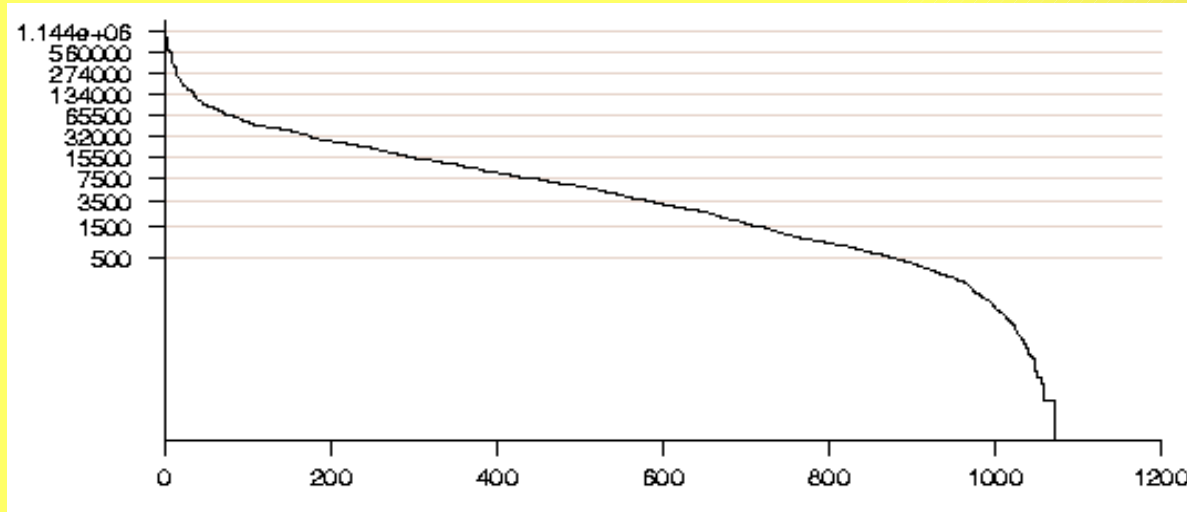
Tamaño de Debian en MSLOC



(Tamaños (MSLOC) de las cinco últimas versiones estables)

Distribución del tamaño de paquetes

Debian 2.0

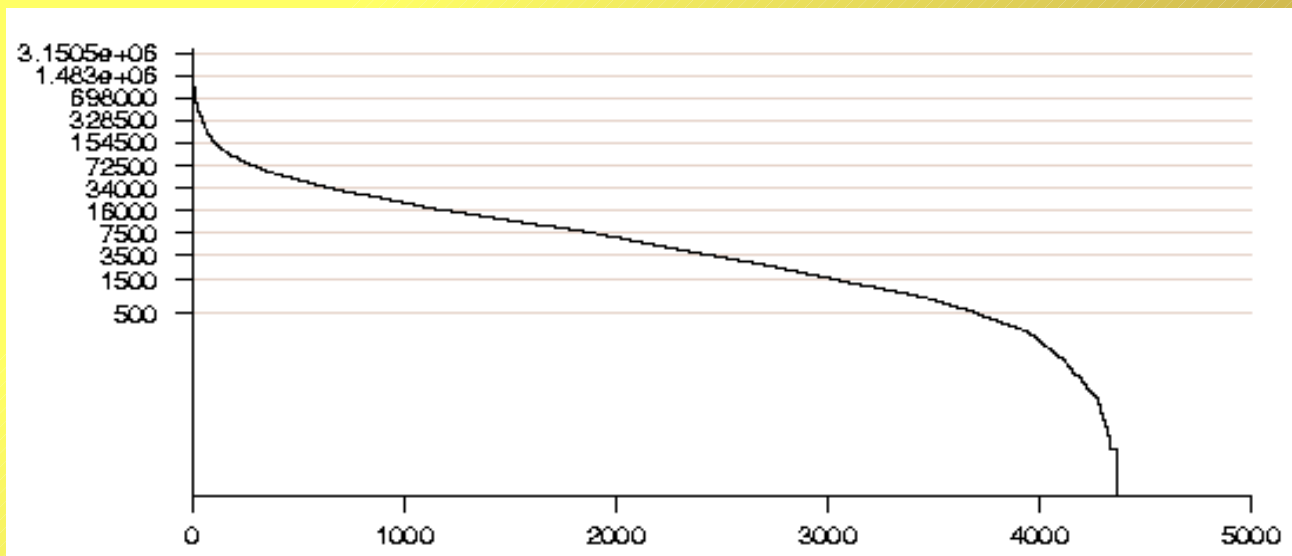
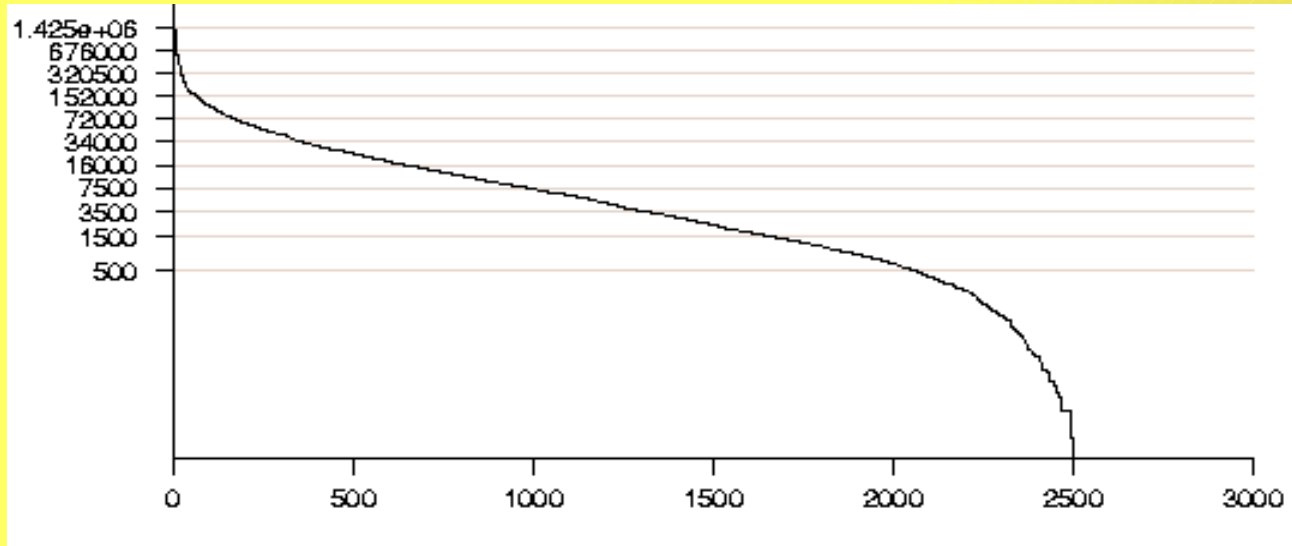


Debian 2.1



Distribución del tamaño de paquetes

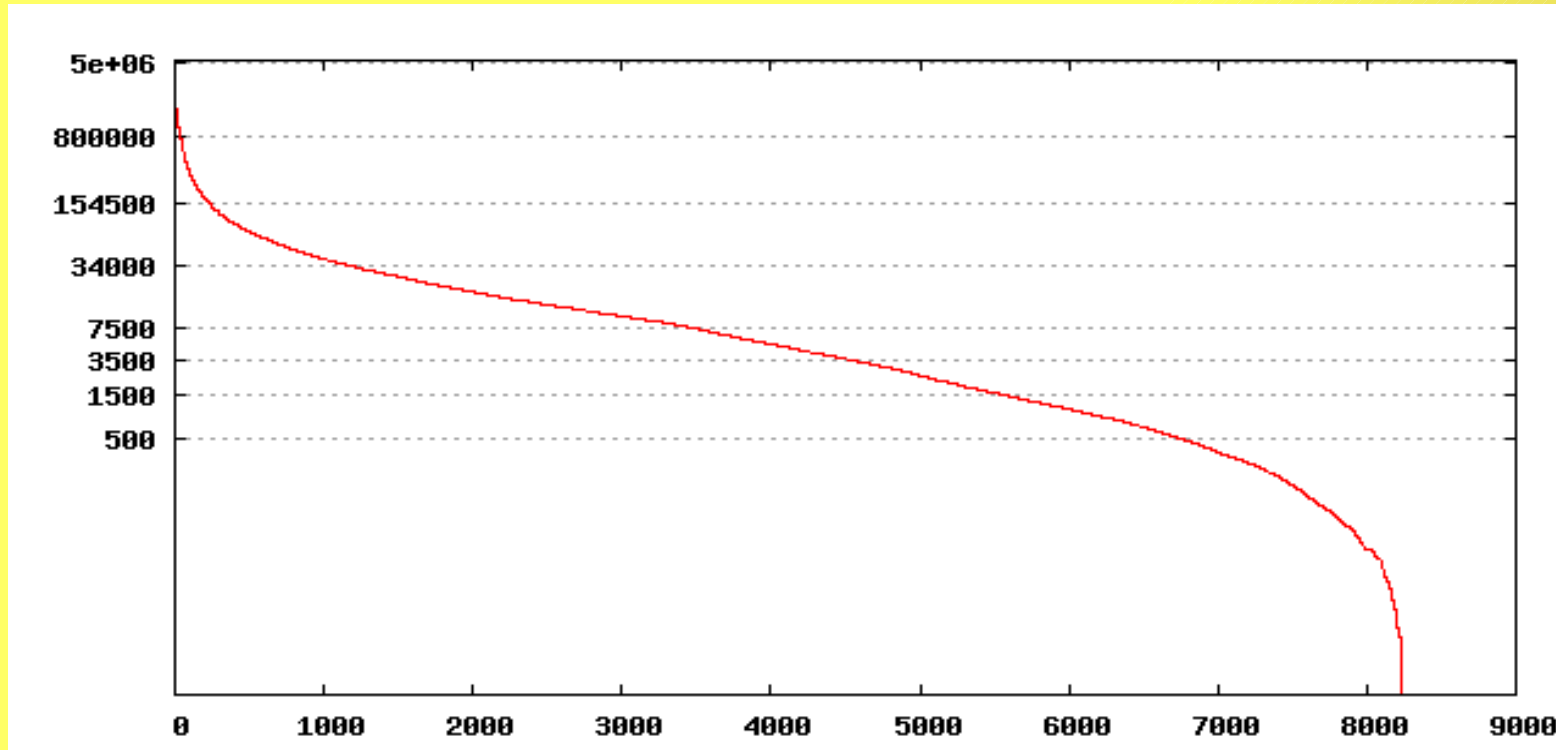
Debian 2.2



Debian 3.0



Distribución del tamaño de paquetes



Debian 3.1



- El tamaño medio de los paquetes se mantiene (casi) constante:
 - Debian 2.0: 23050
 - Debian 3.1: 25212

Los 10 mayores en Debian 2.0 (*)

Posición	Paquete	Versión	Tamaño	Ficheros
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

(*) Debian 2.0 ('Hamm') publicado en Julio de 1998

Los 10 mayores en Debian 3.0 (*)

Posición	Paquete	Versión	Tamaño	Ficheros
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	3.1.2001..	825000	1178

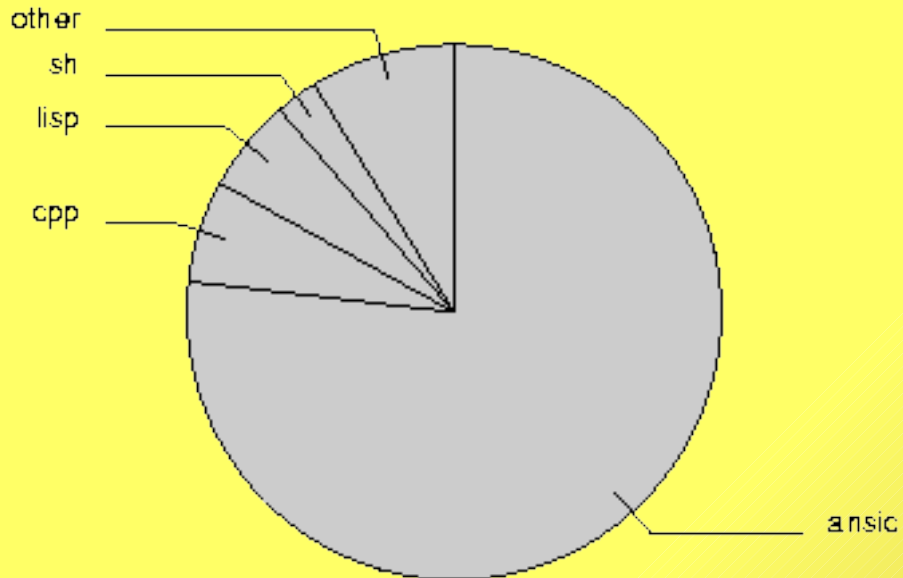
(*) Debian 3.0 ('Woody') publicado en Julio 2002

Los 10 mayores en Debian 3.1 (*)

Posición	Paquete	Versión	Tamaño	Ficheros
1	Openoffice.org	1.1.2...	5180000	19011
2	Kernel-source	2.6.8	4040000	13974
3	Mozilla	2:1.7.3	2440000	10713
4	gcc-3.4	3.4.3	2420000	22992
5	xfs-xtt	1	2330000	7081
6	xfree86	4.3.0...	2320000	7216
7	insight	6.1+...	1690000	4063
8	ia32-libs	1.4	1582000	4752
9	ace	5.4.2.1.0	1400000	10718
10	bigloo	2.6e	1362000	1658

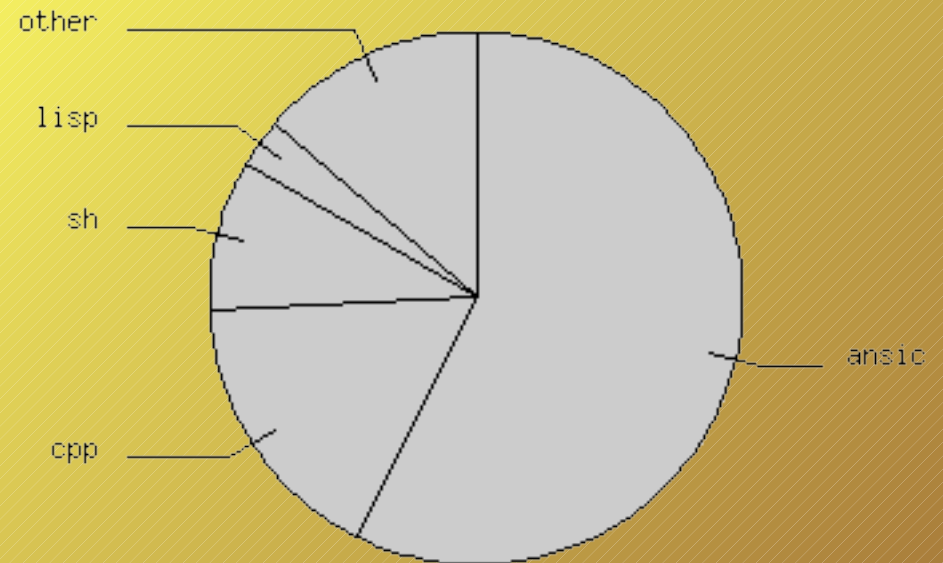
(*) Debian 3.1 ('Sarge') publicado en julio de 2005

Lenguajes de programación

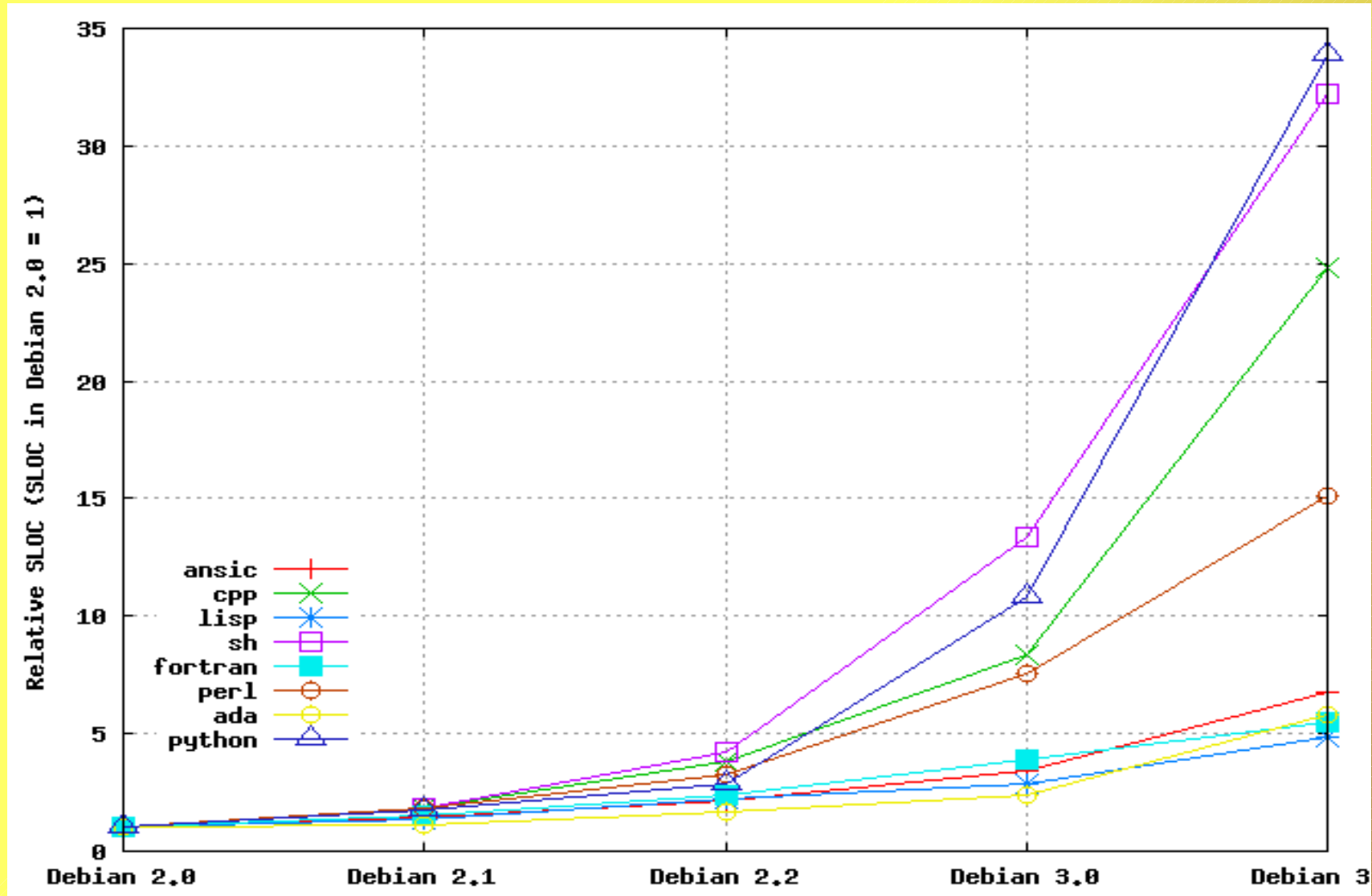


Debian 3.1

← Debian 2.0



Lenguajes de programación



COCOMO

- COCOMO: técnica clásica de estimación de esfuerzo [Boehm81]

Versión	MSLOC	Esfuerzo	Tiempo	Coste
Debian 2.0	25	6360	4.93	860
Debian 2.1	37	9425	4.99	1275
Debian 2.2	59	14950	6.04	2020
Debian 3.0	105	26835	6.81	3625
Debian 3.1	216	55636	8.82	7516

Esfuerzo en años-hombre.

Tiempo en años.

Coste en millones de dólares

Debian frente a otros SSOO

Windows 3.1	April 92	3
Windows 95	August 95	15
Windows NT 4.0	July 96	16
Debian 2.0	July 98	25
Solaris 7	Oct 98	7.5
Debian 2.1	March 99	37
Windows 2000	Feb 00	29
Windows XP	2002	40
Debian 2.2	August 00	55
Debian 3.0	July 02	105
Fedora Core 4	May 05	76
Debian 3.1	June 05	216
OpenSolaris	June 05	4.6

(*) Los MSLOCs de los sistemas privativos no los podemos asegurar

Análisis de Copyright con PyTernity

- Obtención y desempaquetado de paquetes fuente
- Análisis de cada fichero fuente o documento
- Búsqueda de patrones de autoría, mediante algunas técnicas heurísticas
- Almacenamiento de resultados en una base de datos

Autoría en 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

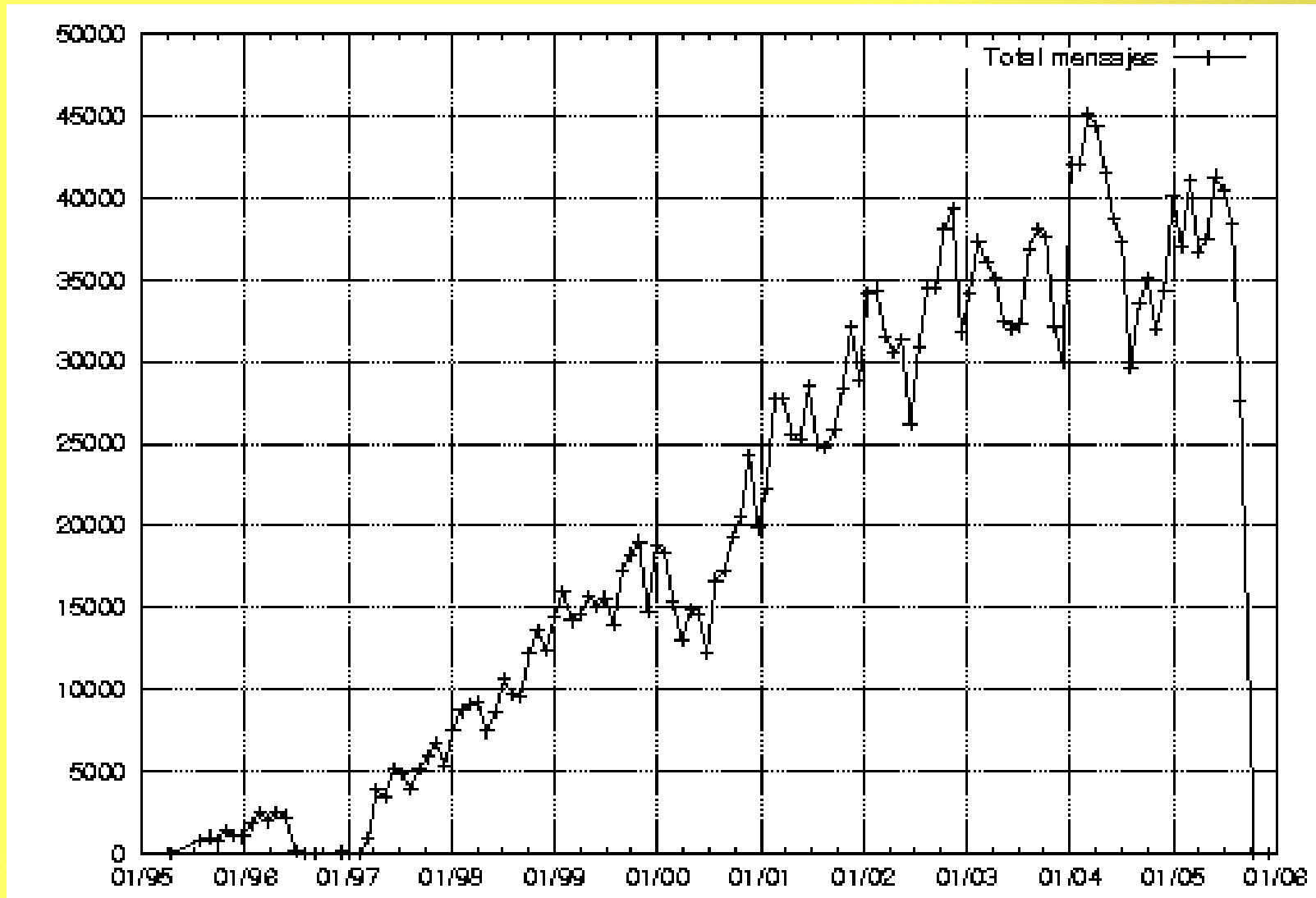
Autoría en Debian 3.1 (Sarge)

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

Listas de correo de Debian

- Análisis de la actividad y la participación en las listas de correo de Debian
- 9 GB de archivos
- 2529432 de mensajes. 2527100 analizados.
- Primer mensaje: 1995-04-08 19:25:32
- Último mensaje: 2005-12-04 12:08:01

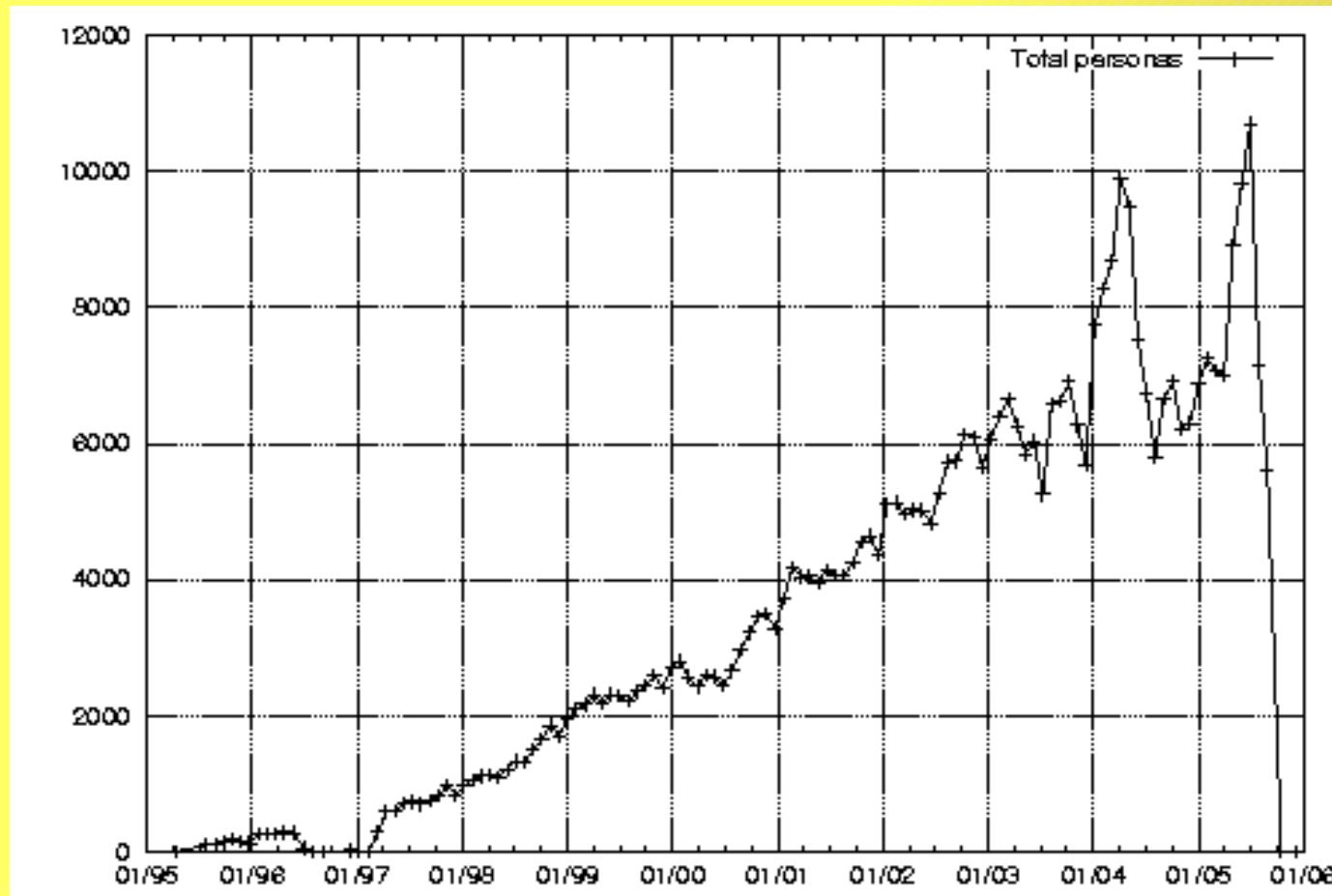
Evolución de la actividad



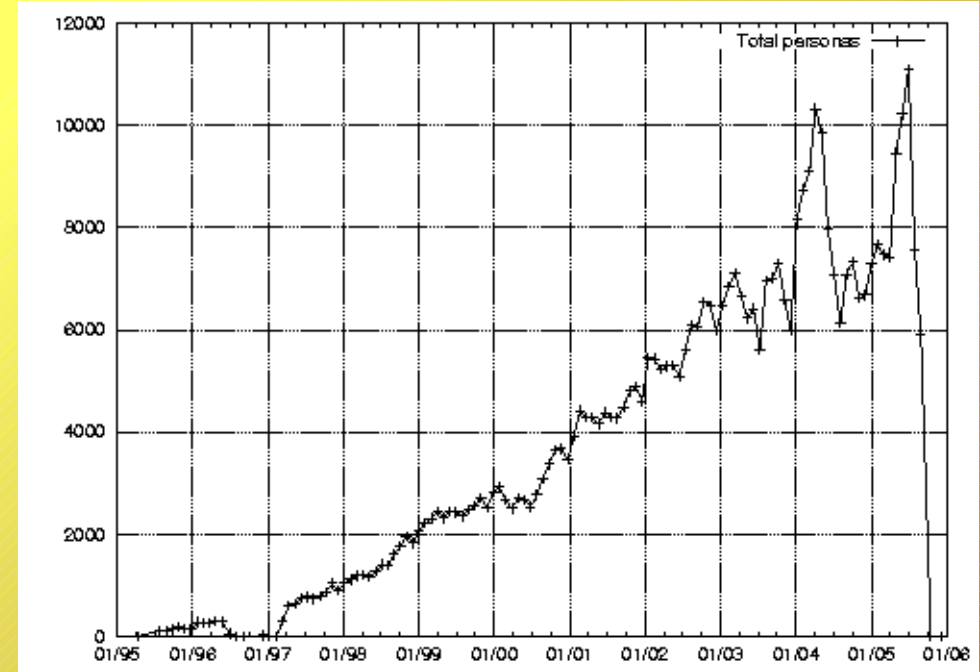
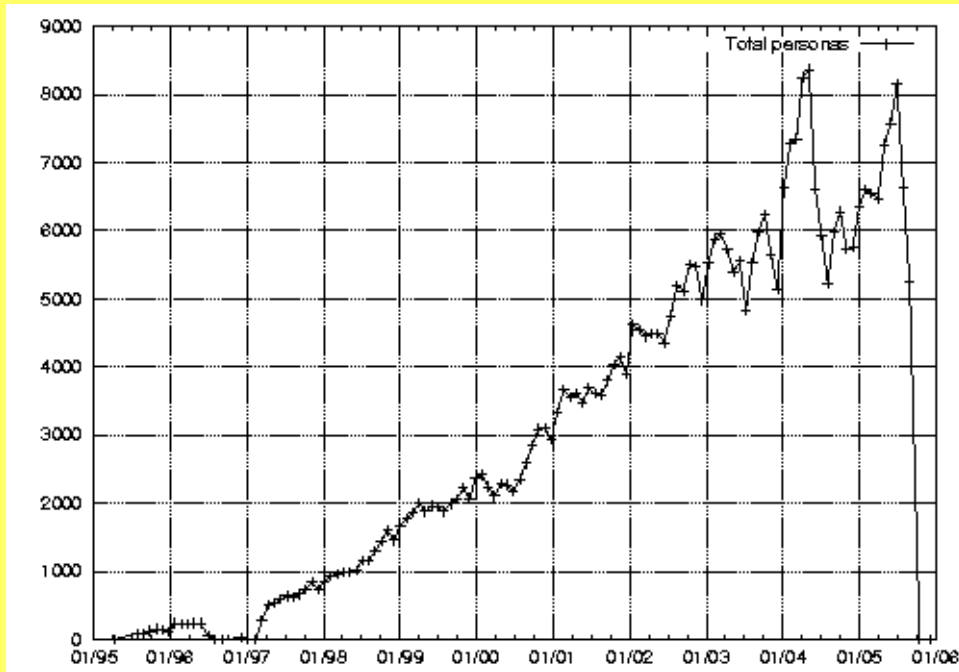
Participación en las listas

- 165026 direcciones de e-mail diferentes
- 117428 nombres diferentes
- 196944 campos From diferentes

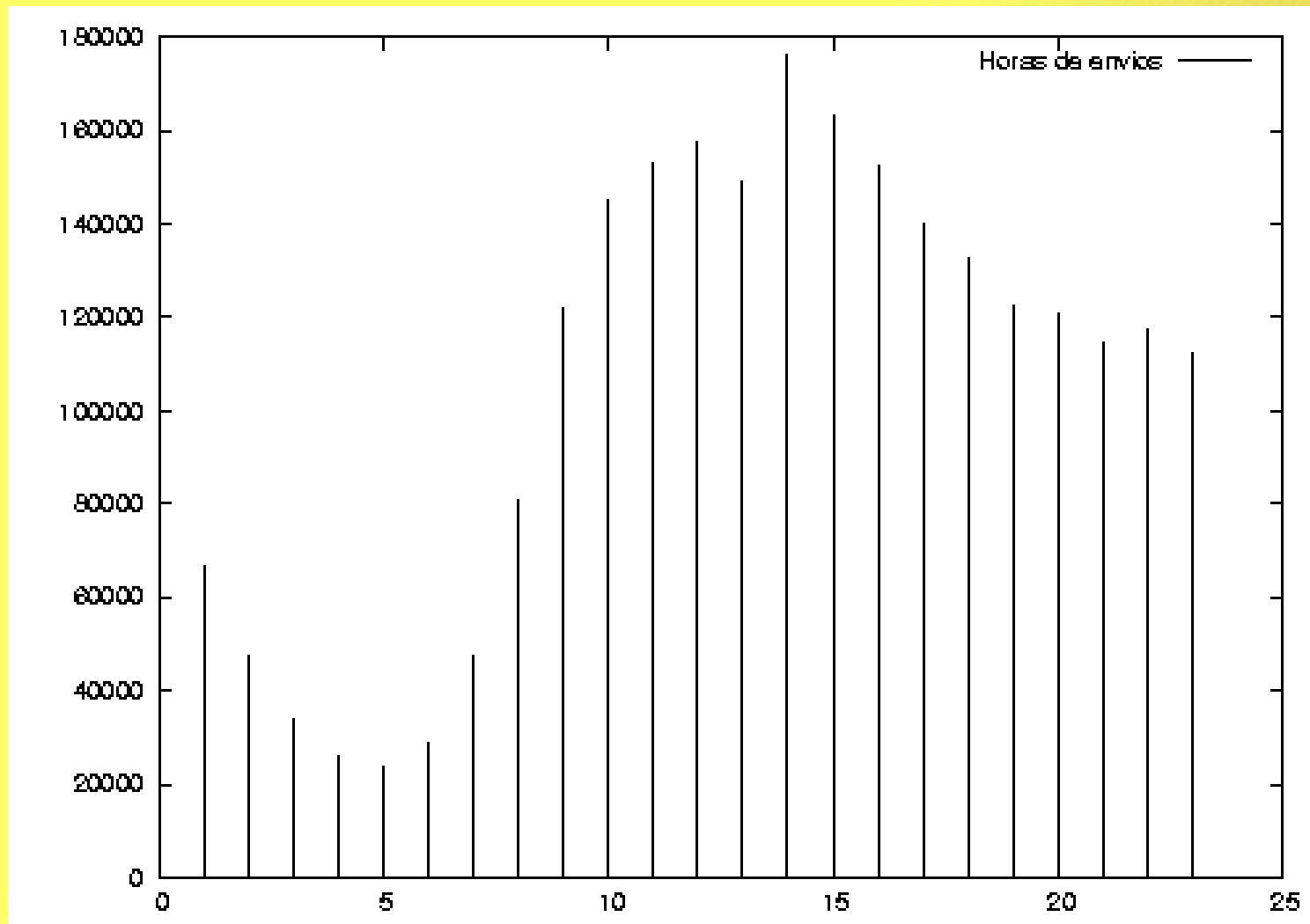
Evolución de la participación



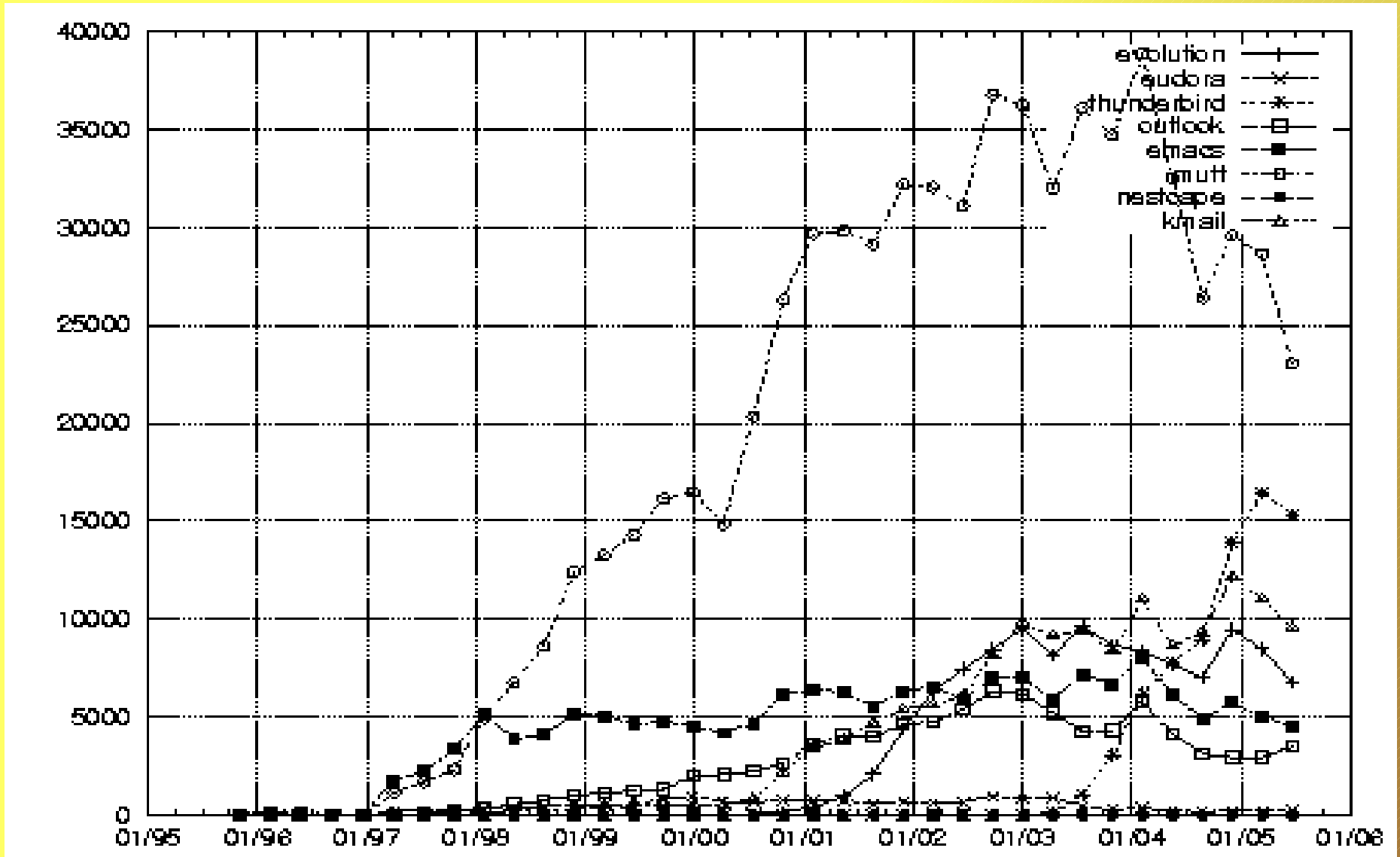
Evolución de la participación



Horas de envíos



Cientes de correo electrónico



Herramientas para hacer estas cosas

- Análisis del código fuente: Mediante Sloccount.
 - <http://www.dwheeler.com/sloccount/>
 - `aptitude install sloccount`
- Análisis de autoría: Mediante pyTernity.
 - <http://forja.linex.org/projects/pyternity> (pronto disponible)
- Análisis de listas de correo: MlStats
 - http://libresoft.es/index_html?menu=Tools
- Otras herramientas interesantes:
 - <http://libresoft.es/>

Conclusiones

- El tamaño de Debian se duplica cada 2 años
- Tamaño medio de los paquetes, (casi) constante
- C es el lenguaje más usado, pero los demás crecen más rápidamente
- Debian GNU/Linux: probablemente el mayor sistema software creado hasta la fecha
- 55-60% del copyright pertenece a desarrolladores individuales
- La participación de empresas crece más deprisa

Referencias

- Libre Software Engineering website at URJC
 - <http://libresoft.es>
- Debian count results (from Debian 2.0 to 3.1):
 - <http://debian-counting.libresoft.es>
- COCOMO:
 - [Boehm81]: Libro 'Software Engineering Economics', Prentice Hall, 1981