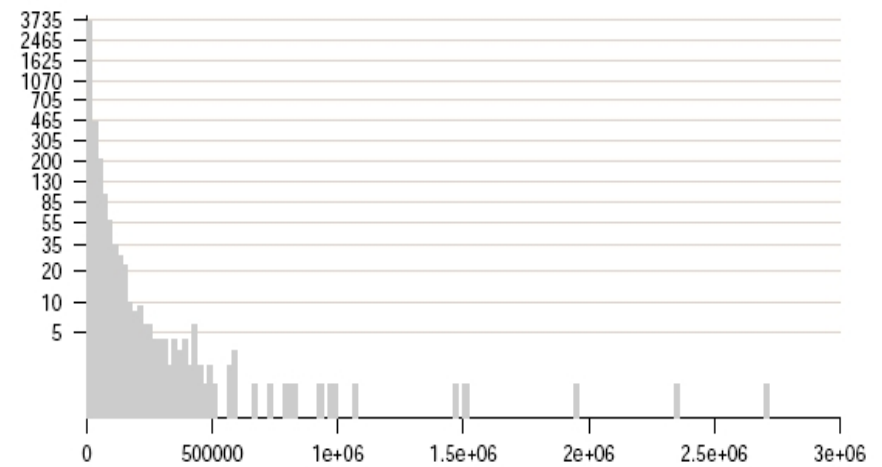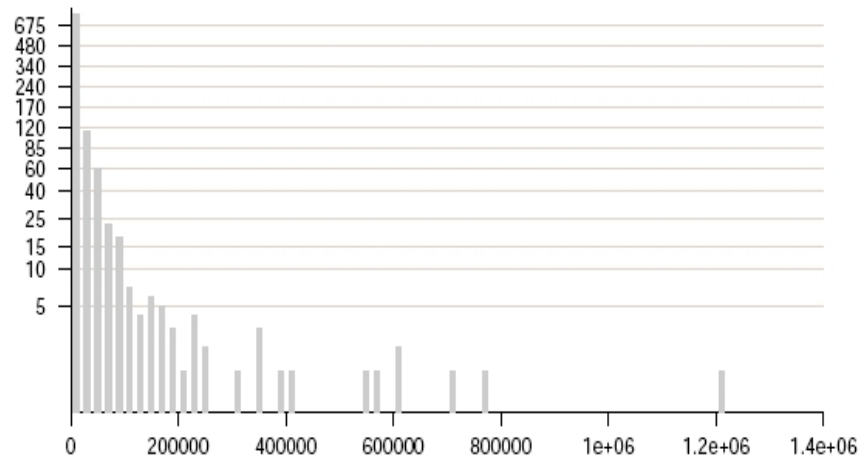# Mining Large Software Compilations over Time: Another Perspective of Software Evolution

Peter Rigby

MSR Reading

October 5, 2006

Drastic Evolution Rate

- Almost double every three years
  - Is it leveling off?
- Mean package size remains constant
  - 23 SLOC

Common Packages and Versions

- 25% of packages disappeared in 7 years
- Found it hard to read
  - Why not use percentages?
  - Over 100% for Deb 2.0 vs 2.1?

Version	Com pkgs	Com vers.	SLOC com vers.	Files com vers.	SLOC com pkgs
Debian 2.0	813	158	1,271,377	15,296	15,594,976
Debian 2.1	1,124	231	2,306,969	27,543	23,630,211
Debian 2.2	1,946	508	4,992,308	60,525	36,584,110
Debian 3.0	3,848	1,567	16,042,810	211,299	78,451,818
Debian 3.1	8,560	8,560	215,812,764	931,834	215,812,764

Programming Languages

	2.0	% 2.0	2.1	% 2.1	2.2	% 2.2	3.0	% 3.0	3.1	% 3.1
C	19,371	76.7%	27,773	74.9%	40,878	69.1%	66.6	63.1%	120.5	55.8%
C++	1,557	6.2%	2,809	7.6%	5,978	10.1%	13.1	12.4%	36.4	15.8%
Shell	645	2.6%	1,151	3.1%	2,712	4.6%	8.6	8.2%	20.4	9.4%
Perl	425	1.7%	774	2.1%	1,395	2.4%	3.2	3.0%	6.4	2.9%
Lisp	1,425	5.6%	1,892	5.1%	3,197	5.4%	4.1	3.9%	6.8	3.1%
Python	122	0.5%	211	0.6%	349	0.6%	1.5	1.4%	4.1	1.9%
Java	22	0.1%	58	0.2%	183	0.3%	0.5	0.5%	3.8	1.7%
Fortran	494	2.0%	735	2.0%	1,182	2.0%	1,939	1.8%	2.7	1.3%

Table 4: Top programming languages in Debian. For Debian 2.0, 2.1 and 2.2 the sizes are given in KSLOC, for versions 3.0 and 3.1 in MSLOC.

- Relative decline of C and increase in shell, java, etc
- Procedural vs OO style = large vs small

- Strengths
  - Original: Looking at evol. of distros
  - Popularity and size of files for languages
  - ??
- Weaknesses
  - What kinds of applications are present
    - Installation rates of packages
  - More discussion and analysis (7 pages)
  - Graphs and tables confusing