

# Free/Libre and Open Source Software

- Source code analysis -

**Rishab Aiyer Ghosh**

**Gregorio Robles**

**Ruediger Glott**

**International Institute of Infonomics, Maastricht**

# Look at the source...

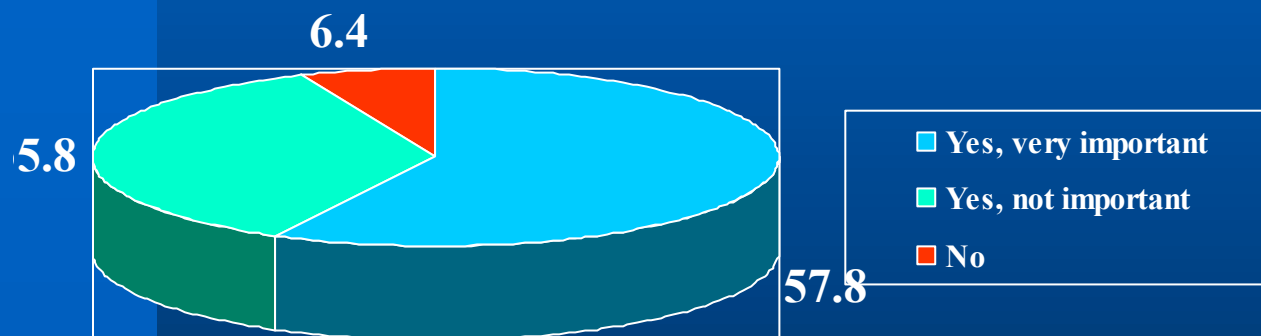
---

## What the source can do for you:

- Authorship information for source at the sub-package/component level
- Size and integrity information for source code at the package level
- The degree of code dependency between packages
- Clusters of authorship: groups of authors who work together, identified by their joint work on individual packages

# Authorship: names in the source

Most F/OSS developers consider marking source code they've written as important:



Do you mark source code you write with your name? (%)

From the FLOSS<sup>1</sup> developer survey: <http://floss1.infonomics.nl/stats.php?id=31>

[1] European Commission/IST supported study, [www.infonomics.nl/FLOSS/](http://www.infonomics.nl/FLOSS/)

# Authorship: names in the source

---

- Other methods include using data from version-control software (CVS, Bitkeeper)
- Linux Software Map (Paul Jones)
- Web sites/registered maintainer (on Sourceforge, Savannah etc)
- What “author” means can vary considerably across methods

# Size and integrity

---

- **Size as an indicator of value produced, or time/effort spent**
- **Size of relative contribution indicates relative importance of author**
- **Integrity: identifying what belongs where**
- **Many components are “borrowed” from other packages, and must be identified**

# Code dependency

---

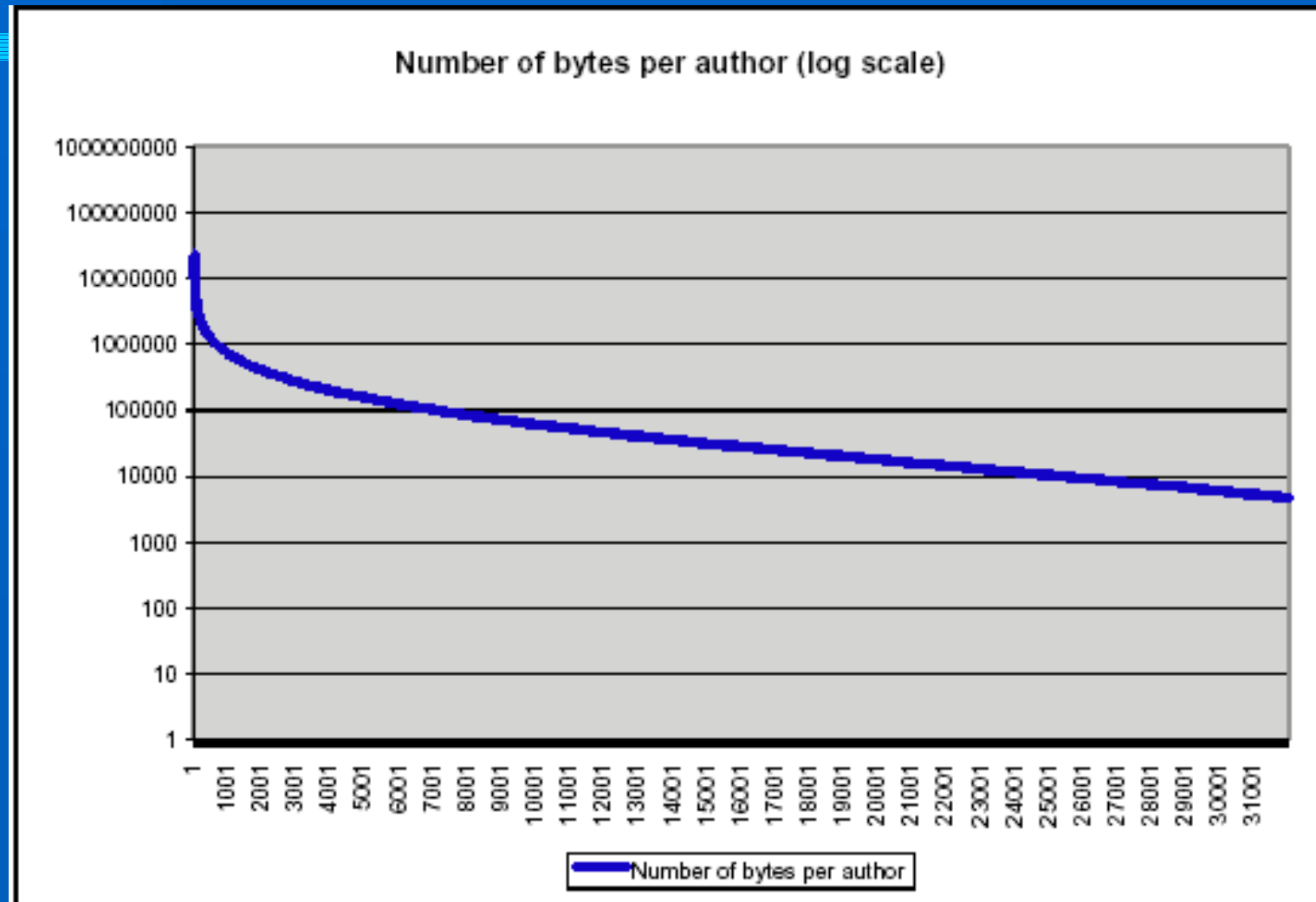
- **Software packages are highly inter-dependent**
- **Dependencies are usually explicitly declared**
- **Dependencies indicate the relative importance and value of packages (probably better for this than size)**

# State of tools and data (FLOSS)

---

- **Run on large code base (40+ Gb of source code, 22,000+ unique packages, 36,000+ authors identified)**
- **Authorship, integrity and dependency data extraction**

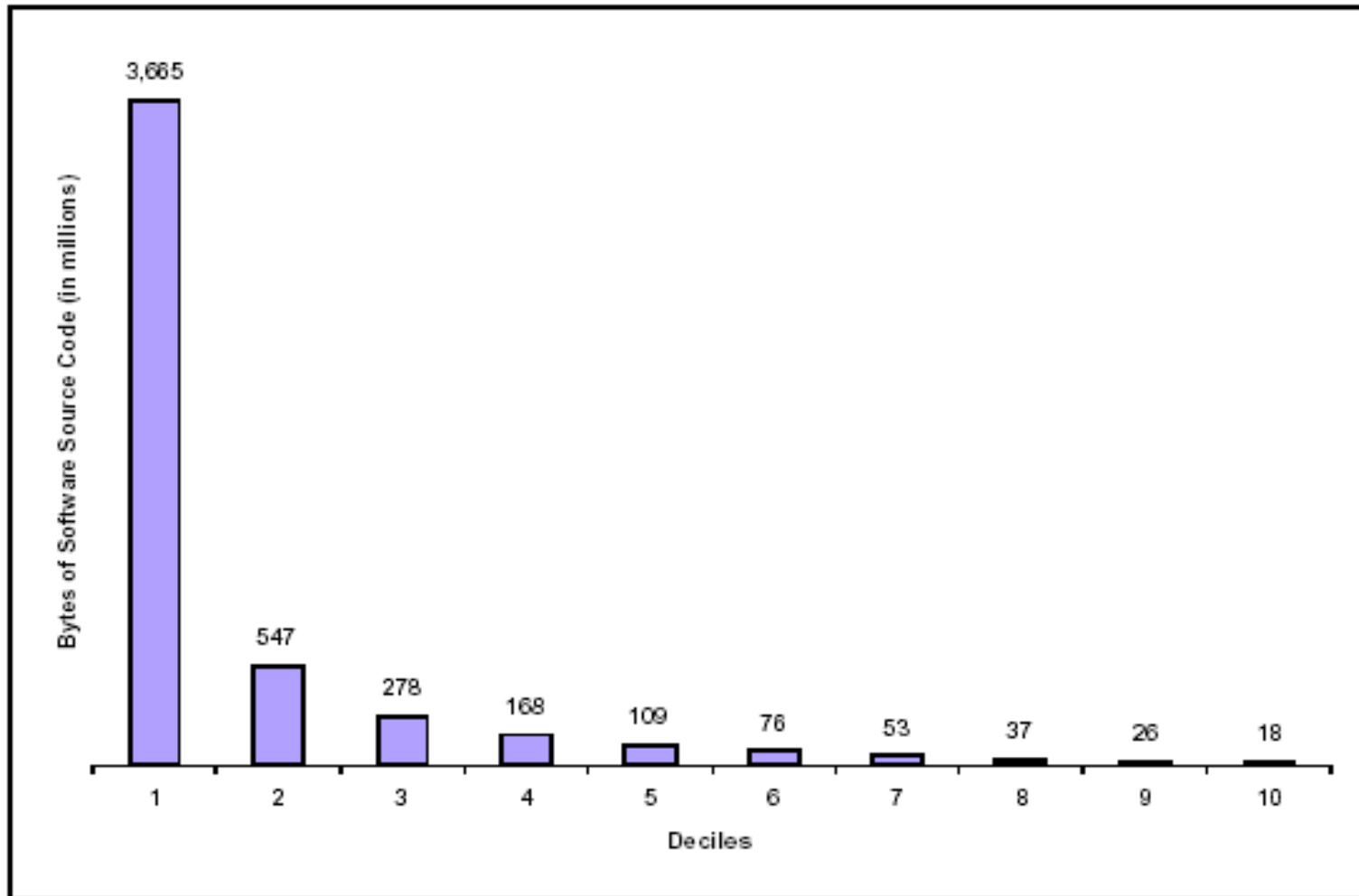
# Results: Authorship concentration





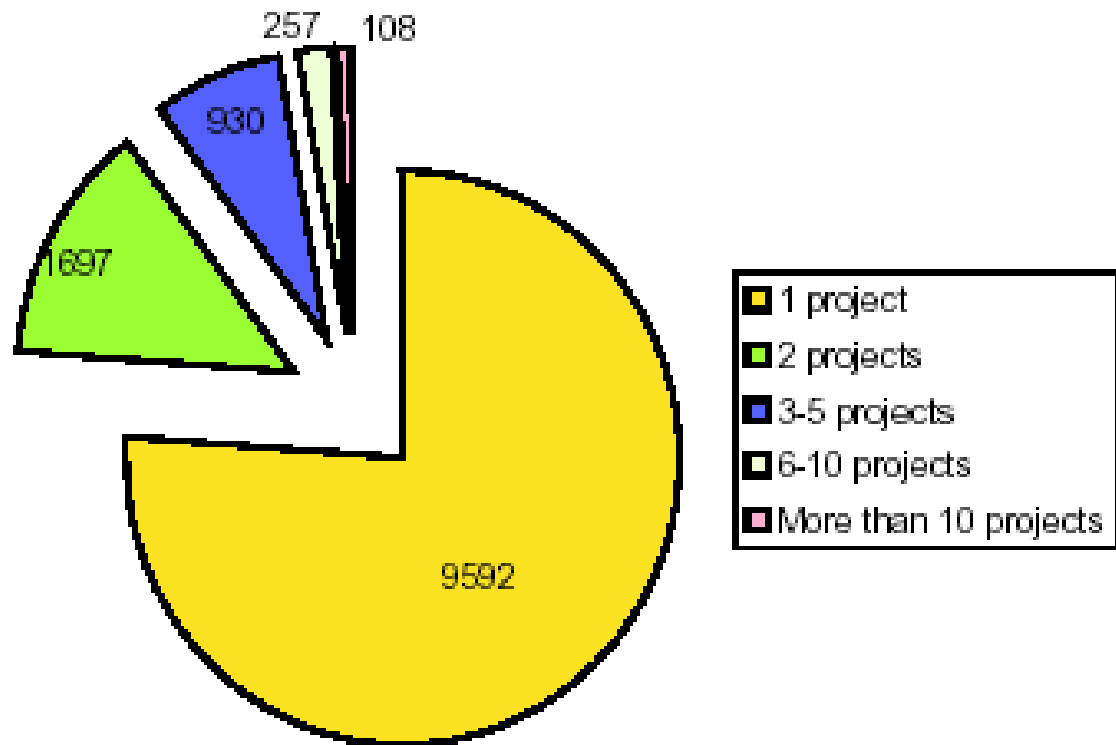
# Results: Authorship by decile

Figure 2: Software Source Code Authorship by Decile



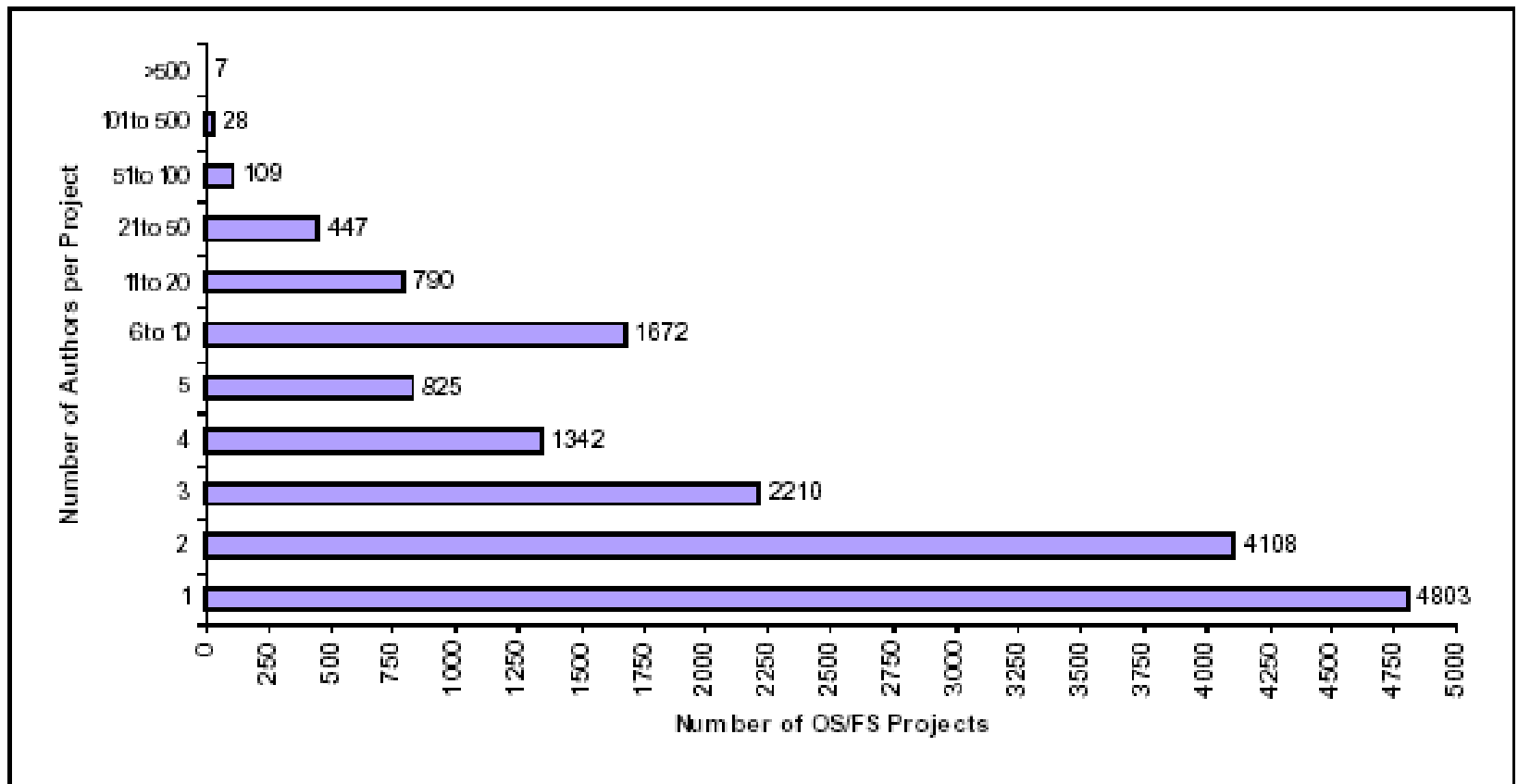
# Results: Leadership

Authors contributing more than 20% to projects



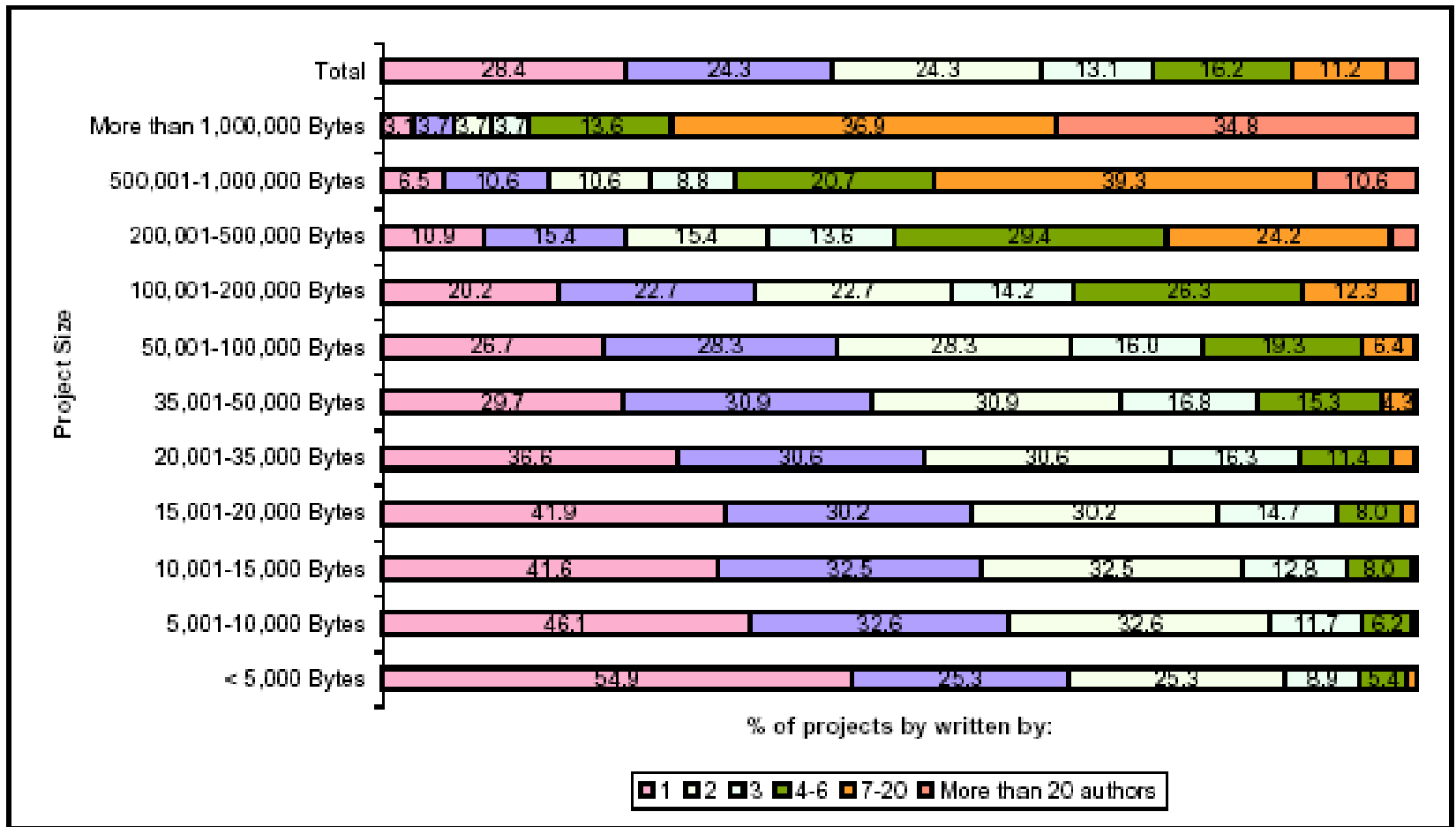
# Results: Project/author distribution

Figure 6: OS/FS Projects by Number of Authors



# Results: Project/author distribution

Figure 7: Number of contributing authors by project size



# Conclusion

---

- **No money does not mean no measurement**
- **Hard facts are available: look for (and at) the source**
- **Hard to analyse: requires visualisation tools**