

Women in Free/Open Source Software Development

Hanna M. Wallach

University of Massachusetts Amherst

wallach@cs.umass.edu

Introduction



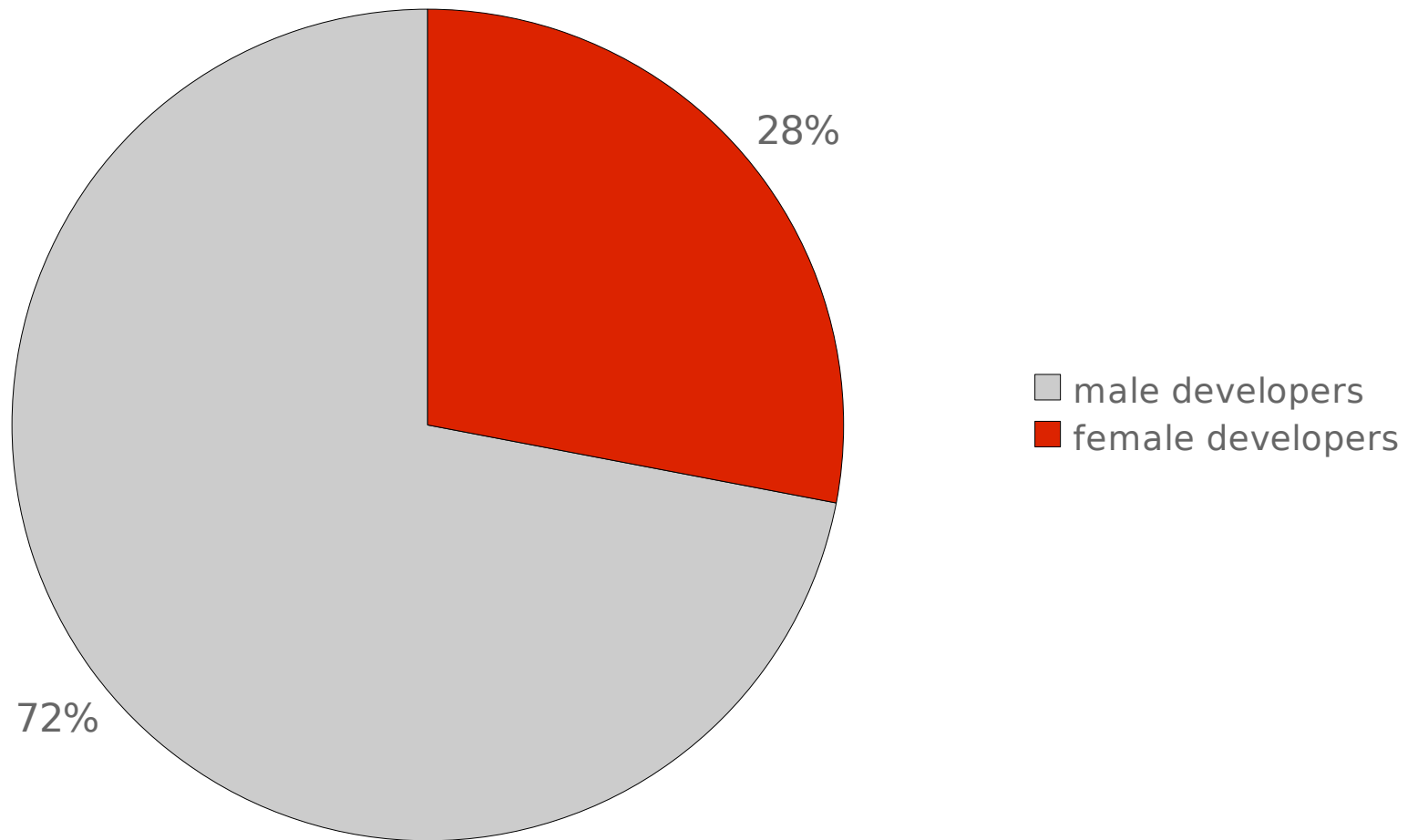
debian



- Postdoctoral researcher at UMass Amherst
- Contributor to Debian GNU/Linux & GNOME
- Co-founder of Debian Women & GNOME WSOP
- Workshop organizer for FLOSSPOLS gender study

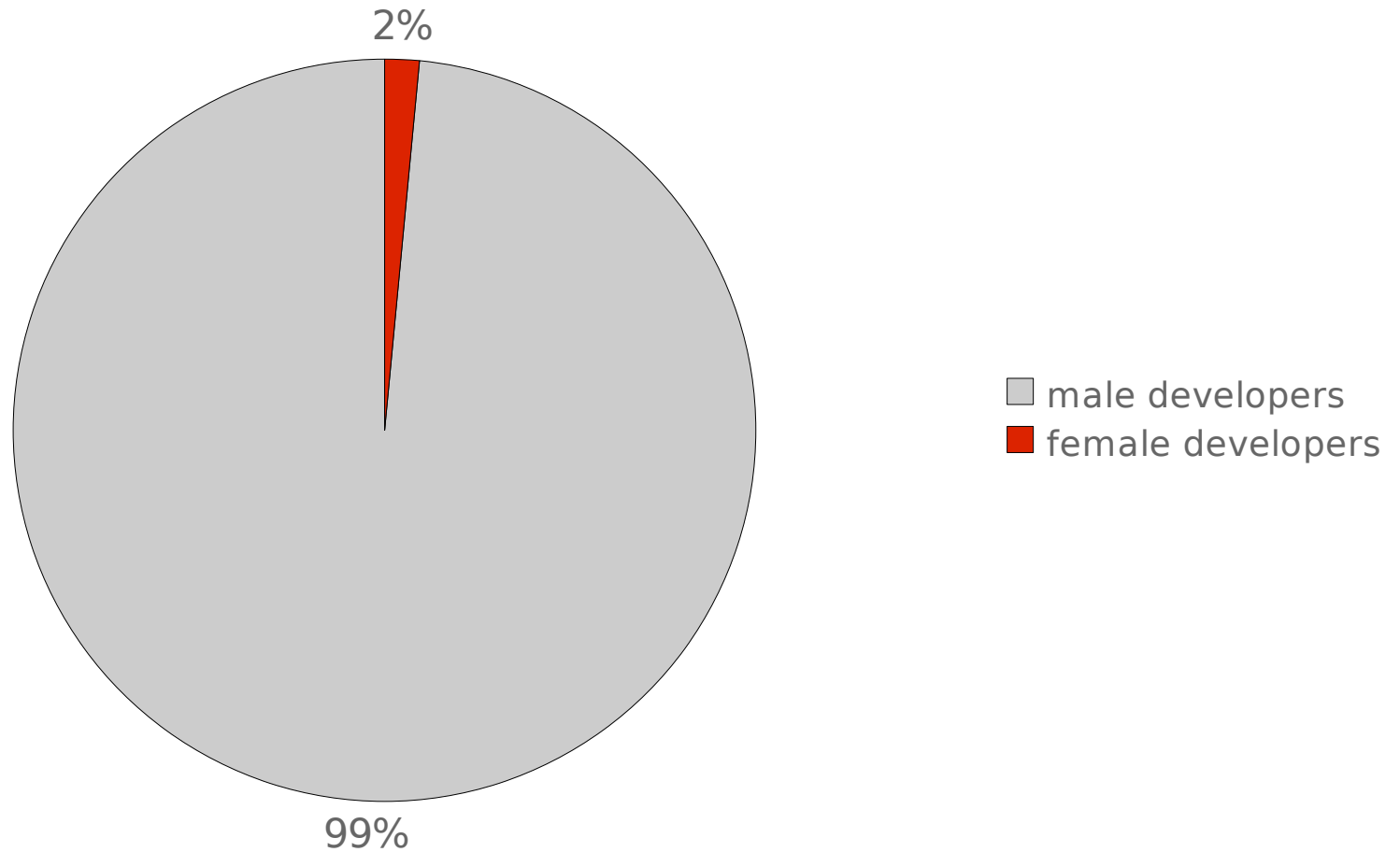
Proprietary Software

[NSF, '04]



Free/Libre/Open Source Software

[Ghosh et al., '02, '05]



Does Diversity Matter?

- Diversity of ideas:
 - (Re)combination of ideas \Rightarrow diversity of innovations
 - Diversity \Rightarrow better matching of people to projects
- Diversity of individuals:
 - Distinct backgrounds \Rightarrow rapid and breakthrough progress
 - Diverse developer base \Rightarrow diverse user base
 - “Democratizing innovation”

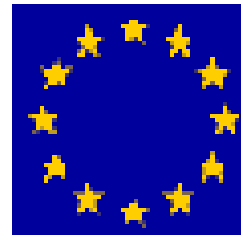
This Talk

- Why so few female FLOSS developers?
 - FLOSSPOLS gender study
- Grassroots efforts:
 - Debian Women Project
 - GNOME Women's Summer Outreach Program
- New research directions:
 - Statistical analysis of community data

This Talk

- Why so few female FLOSS developers?
 - FLOSSPOLS gender study

FLOSSPOLLS Project



- Academic study of FLOSS:
 - Funded by the European Commission
 - March 2004 to February 2006
- Three strands of research: one on gender

Gender Strand: Researchers



James Leach, Dawn Nafus, Bernhard Krieger
(University of Cambridge)

Gender Strand: Aims

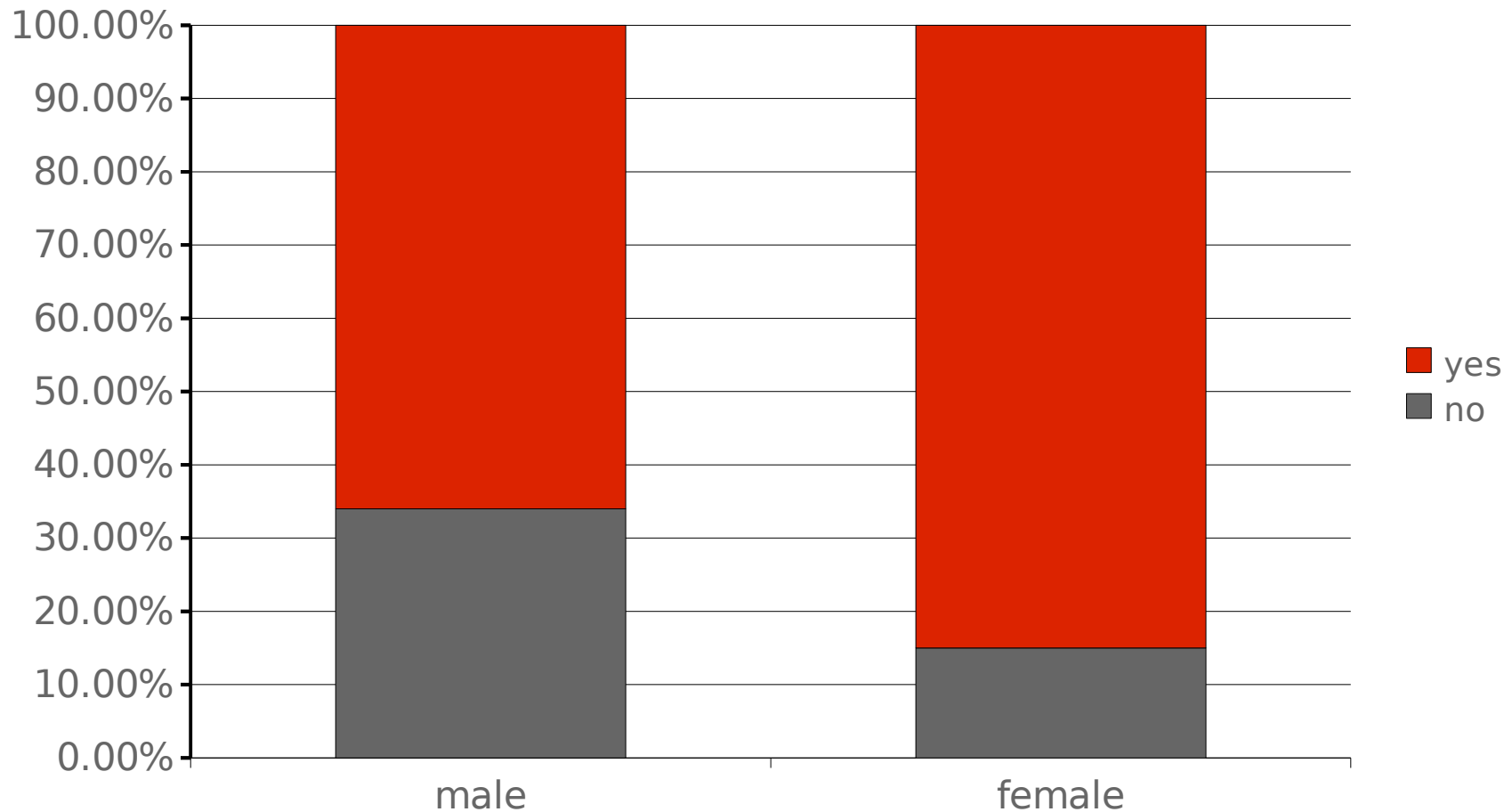
- Produce the first study of gender in FLOSS
- Gather socioeconomic data
- Identify reasons for the gender bias
- Establish policy recommendations for change

Gender Strand: Methodology

- Ethnographic methodology:
 - Unstructured and semistructured interviews
 - Participant observation (both online and in-person)
- Quantitative methodology:
 - Online survey (1,541 participants)

⇒ Qualitative and quantitative methods used iteratively

“Would more female participants be better for the whole FLOSS community?”

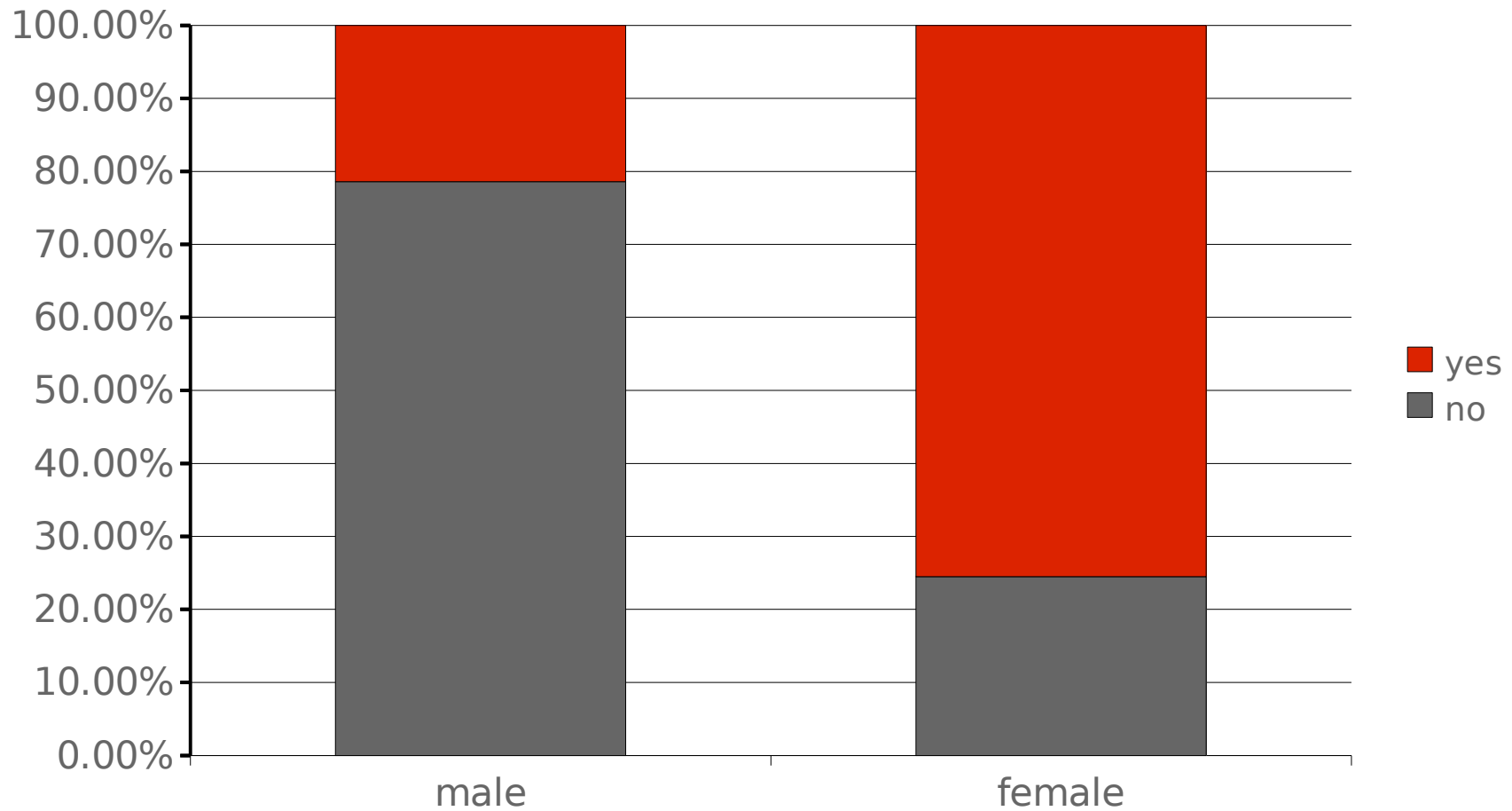


Are women just not interested in FLOSS?
Or are they being excluded?

Active (But Unconscious) Exclusion

- Active but unconscious exclusion:
 - Joking about women (“if it's a joke, it's okay”)
 - Using women as examples of nontechnical users
 - Assuming that society does not affect choices
 - Viewing coding as superior to other skills

“Have you ever observed discriminatory behaviour toward women?”



Does the “hacker identity” alienate women?

Do I Have To Have A Beard?



Do women have as much experience as men?

Later, Later, Later...

	male	female
age of first computer use	12	14.5
age of first computer possession	15	19
year of intial F/LOSS involvement	1996	1999

Is code the only thing that matters?

Typical Tasks

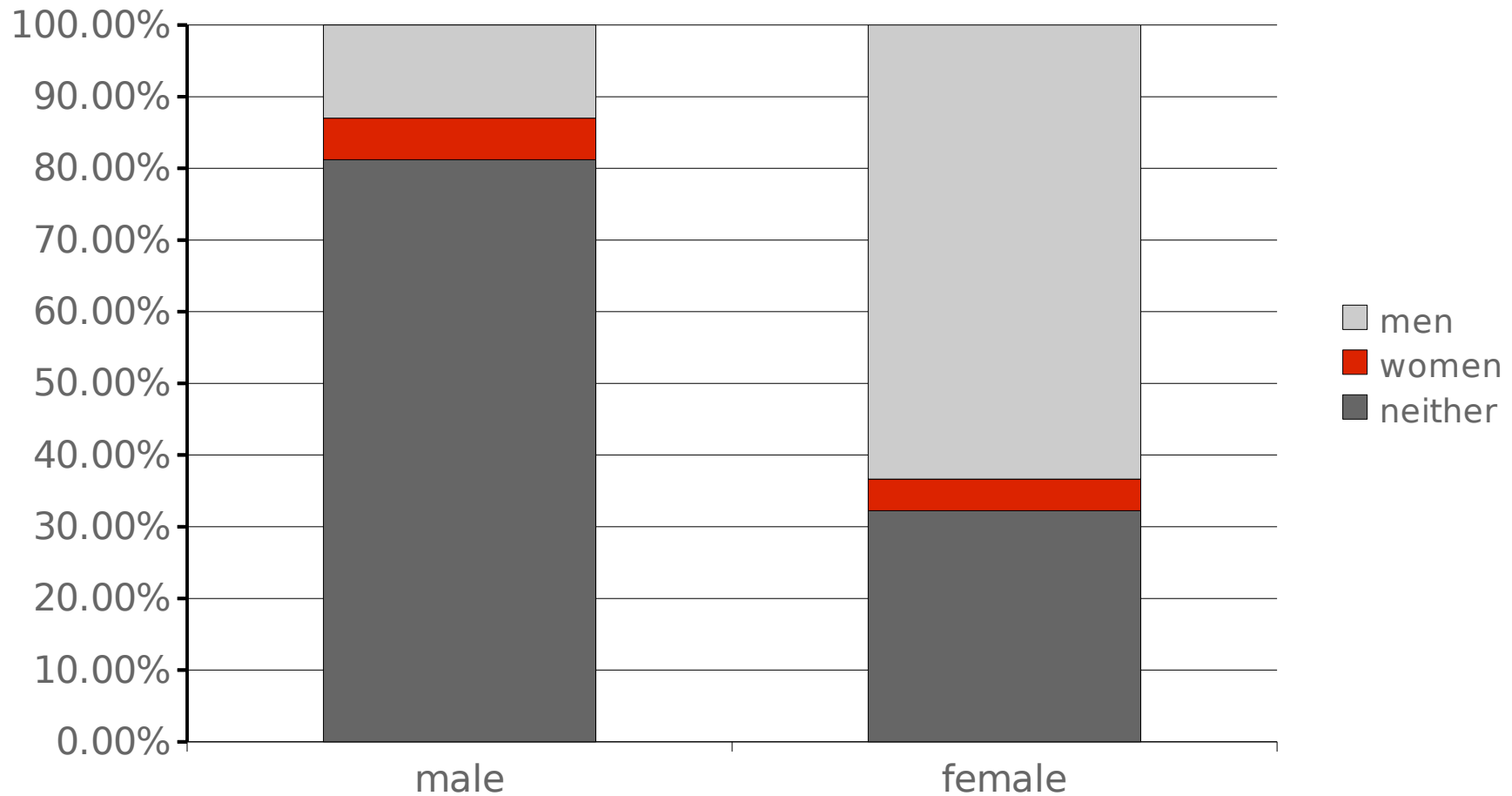
Men:

- Writing code
- Testing
- Bug reports
- “Technical tasks”?

Women:

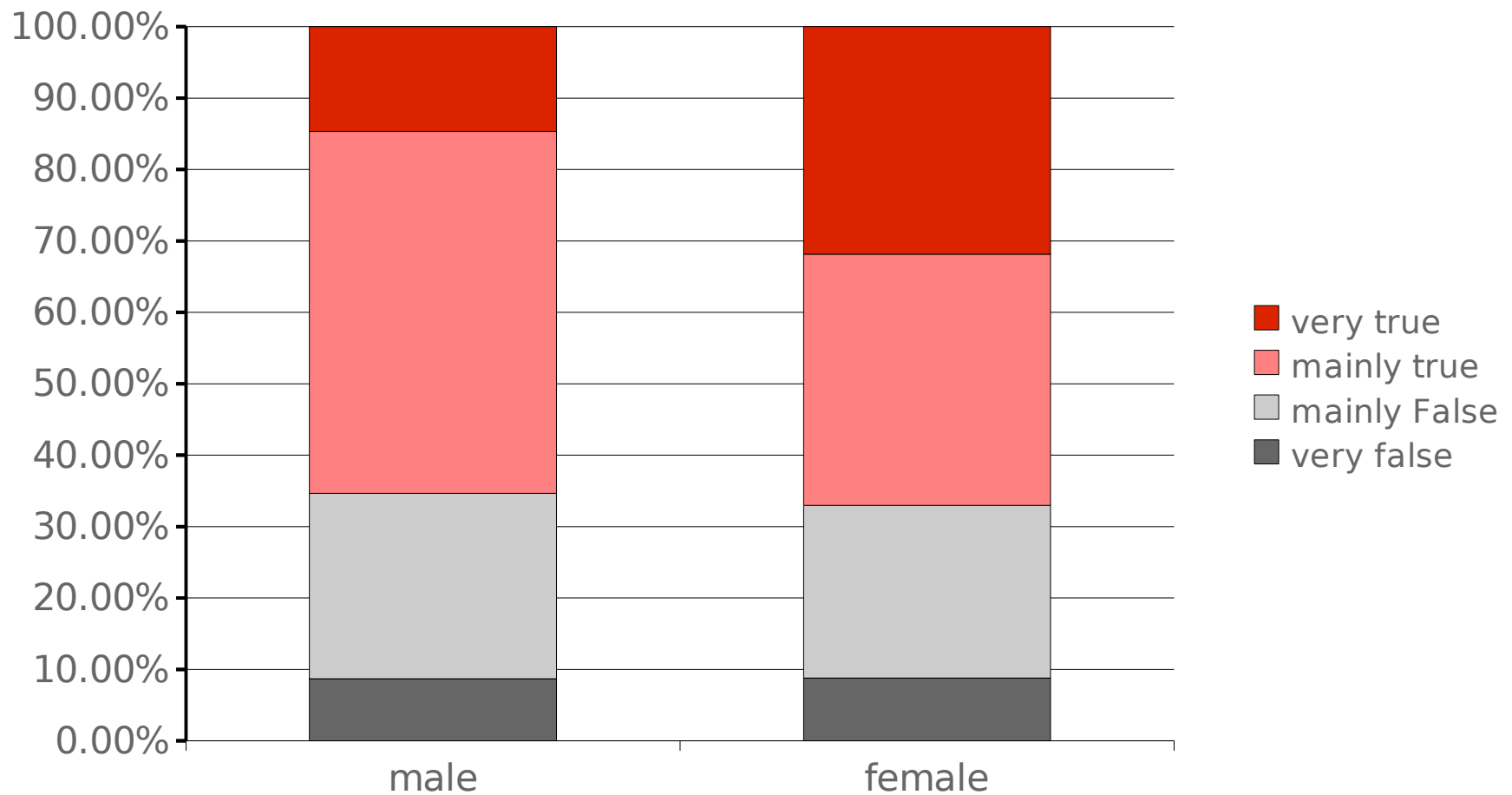
- Documentation
- Organising events
- Translation
- “Social tasks”?

“For whom is it easier to get acknowledgement for FLOSS work?”

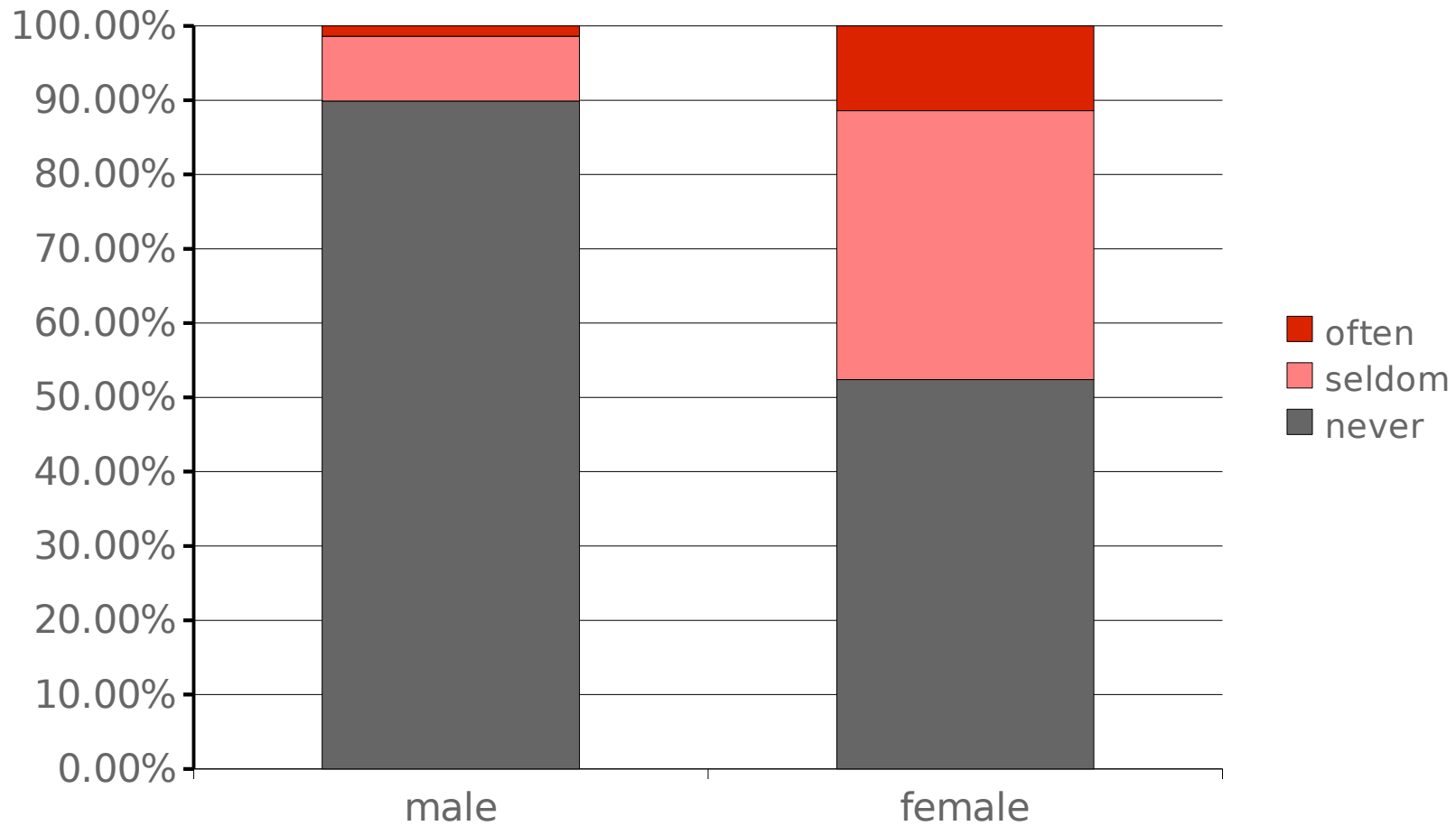


Do women get more attention for their
gender or their skills?

“Women often get more attention as a woman than as a FLOSS participant.”



“Have you ever been asked on a date by another FLOSS participant?”



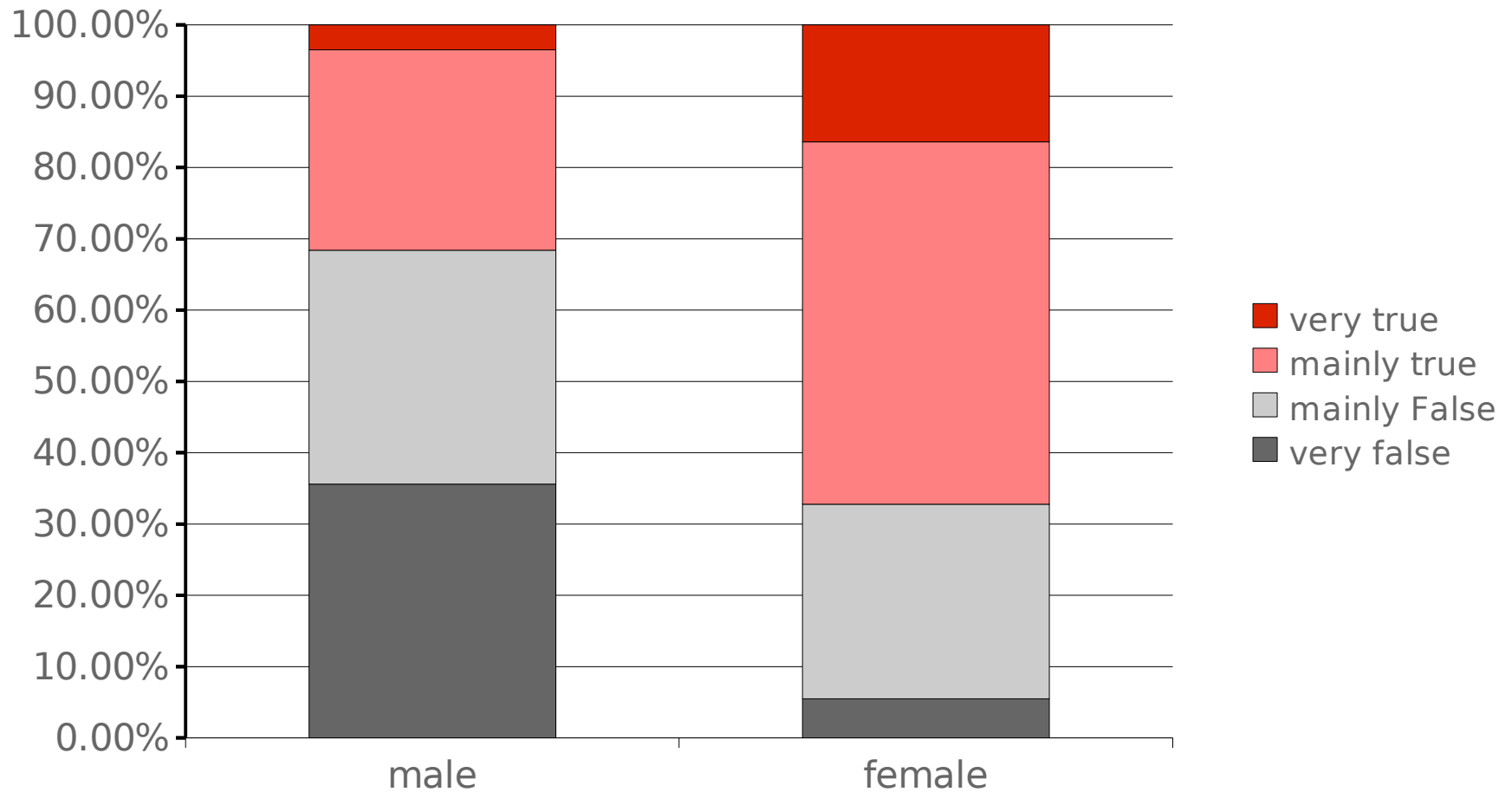
Are there too many framewars?

Flamewars vs. Confidence

- Flaming can be a key means of developing reputation:
 - Used as a platform to increase visibility
 - More established participants less likely to flame
- Flaming exacerbates women's confidence issues
- “Loss of confidence precedes loss of interest”

Do women have less time for FLOSS?

“Do women have less time than men to participate in FLOSS?”



Summary of Findings

- Active (but unconscious) exclusion
- Later involvement in computing
- More diverse (often nontechnical) skills
- More likely to be put off by flamewars
- Less time for FLOSS development

Men and women have very different
FLOSS experiences.

This Talk

- Why so few female FLOSS developers?
 - FLOSSPOLS gender study
- Grassroots efforts:
 - Debian Women Project
 - GNOME Women's Summer Outreach Program

Debian GNU/Linux

- Linux distribution (computer operating system)
- Linux kernel, GNU OS tools
- >15,000 packages
- Strict adherence to free software philosophies
- Largely developed by a group of volunteers

“Universal Operating System”?



The Debian Women Project

- First such project for any free software project
- Support and encourage female Debian developers:
 - Role models for women
 - Tutorials on Debian-related topics
 - Mentoring project
 - Presentations at conferences
 - Word-of-mouth advocacy

Unexpected Findings

- Active involvement of men is important:
 - Different kind of diversity!
- “What's good for women is good for everyone”
- Establishing critical mass is necessary for recruitment
- Multiple entry points helps matching people to projects

GNOME Desktop Project

- Desktop environment (GUI) for users
- Development framework for building applications
- Part of many different Linux distributions
- Many components and projects
- Developed by both volunteers & companies

GNOME & GSoC 2006

- Google's Summer of Code (GSoC):
 - Students funded to work on free software projects
 - Students paired up with mentors
- In 2006, GNOME had 181 GSoC applicants:
 - None of the applicants were women

Women's Summer Outreach Program

- 3 positions funded by the GNOME Foundation Board
- Specifically targeted at women
- Identical to GSoC (except for duration)
- Different language of advertising
- Personal, word-of-mouth advertising

Outcomes & Findings

- Initial skepticism within GNOME:
 - “Will you really find 3 good applicants”
- Over 200 women emailed expressing interest
- Over 100 well-qualified applicants:
 - Many didn't realize they were qualified!
- Targetted, personal advertising really matters

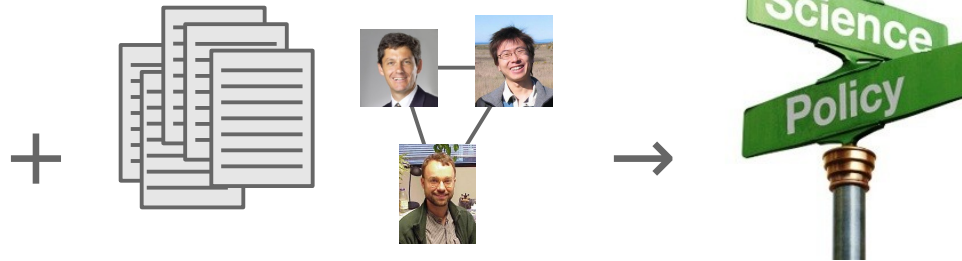
This Talk

- Why so few female FOSS developers:
 - FLOSSPOLS gender study
- Grassroots efforts:
 - Debian Women Project
 - GNOME Women's Summer Outreach Program
- **New research directions:**
 - **Statistical analysis of community data**

My Research Goal

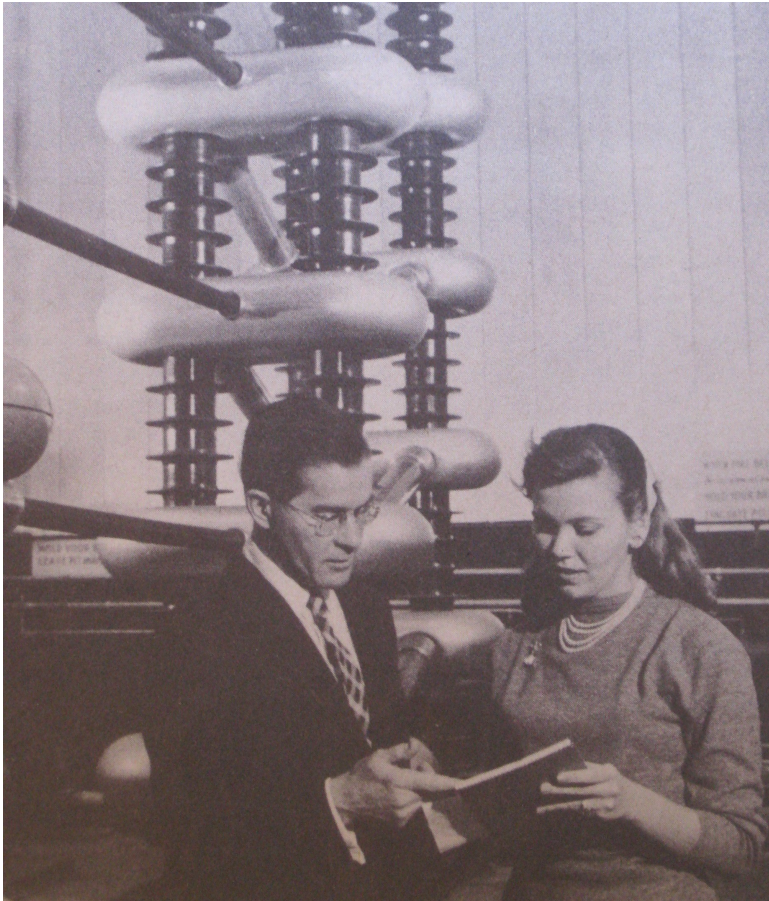
```
$line .= <CASEBOOKS>;  
redo unless eof(CASEBOOKS);  
}  
  
$line =~ s/\\t/xyzdrptmpxyz/g;  
@columns = split("\\t", $line);  
$columns[3] = uc $columns[3];  
$line = join("\\t", @columns);  
$line =~ s/xyzdrptmpxyz/\\t/g;
```

$$\prod_t \frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \frac{\prod_w \Gamma(N_w|t+\beta)}{\Gamma(N_{\cdot}|t+W\beta)}$$



To develop new **statistical models** and **computational tools** for representing and analyzing large quantities of **complex data** in order to better enable scientific policy-makers to identify and evaluate **high-impact policy actions** and advance the **study of science and innovation policy**.

Data: Products of Collaboration



“Scientific information is both the basic raw material for, and one of the principal products of, scientific research [...] Scientists find out what other scientists are accomplishing through [...] journals, books, abstracts and indexes, bibliographies, reviews.”

— NSF Brochure, 1962

FLOSS Development Communities

- Complex technological, legal, social structures
- Collaboration on a massive, global scale
- Most communication is publicly available online:
 - Informal text: messy, unstructured

⇒ Use these data to study organizational and social processes underlying FLOSS development

Documents and Topics

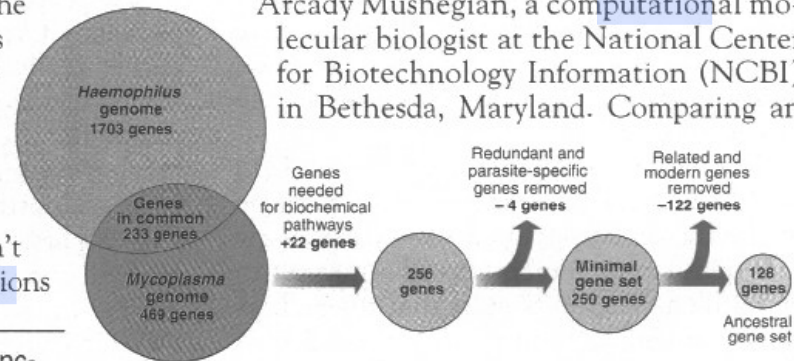
Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK— How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

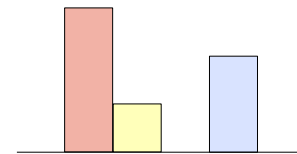
Although the numbers don't match precisely, those predictions

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

“are not all that far apart,” especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. “It may be a way of organizing any newly sequenced genome,” explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

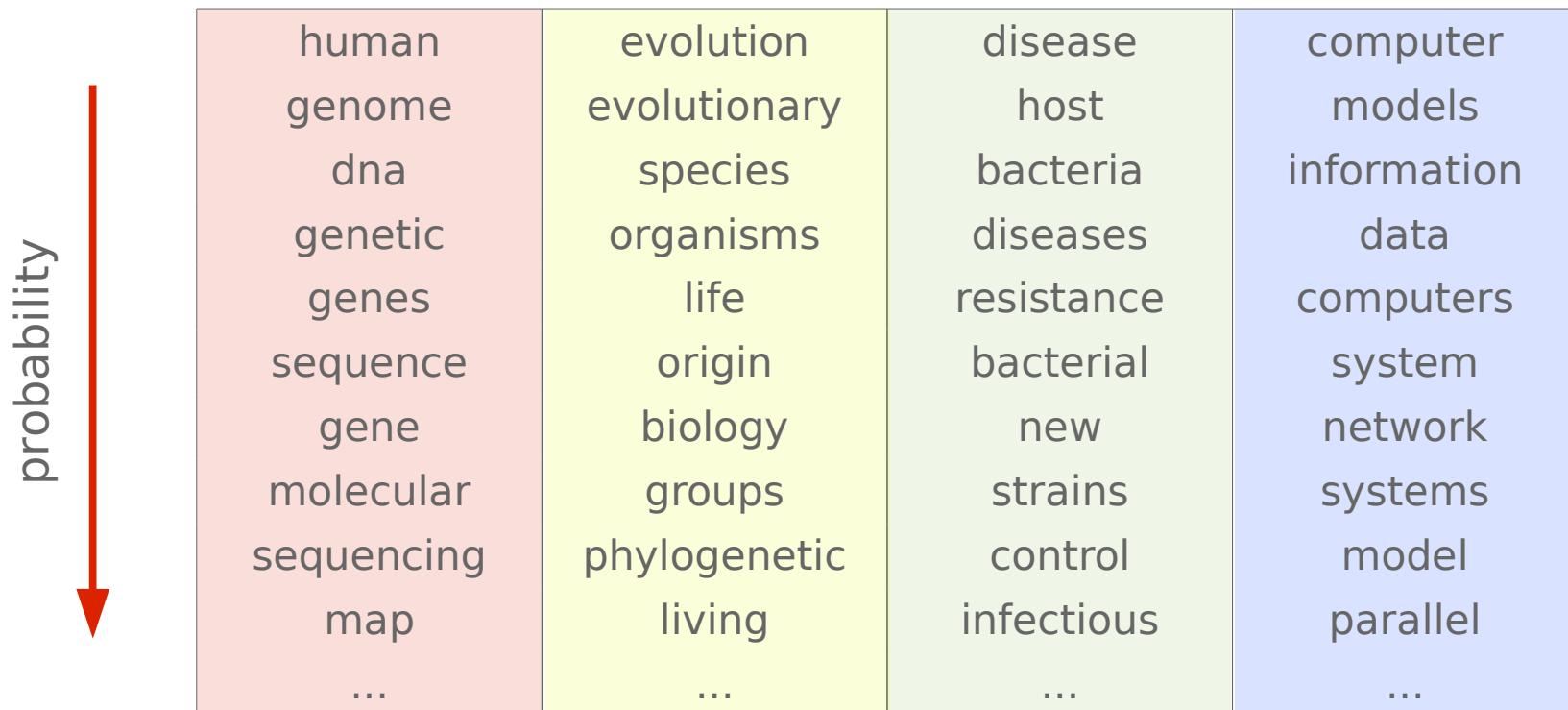


Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.



SCIENCE • VOL. 272 • 24 MAY 1996

Topics and Words



Debian Mailing Lists

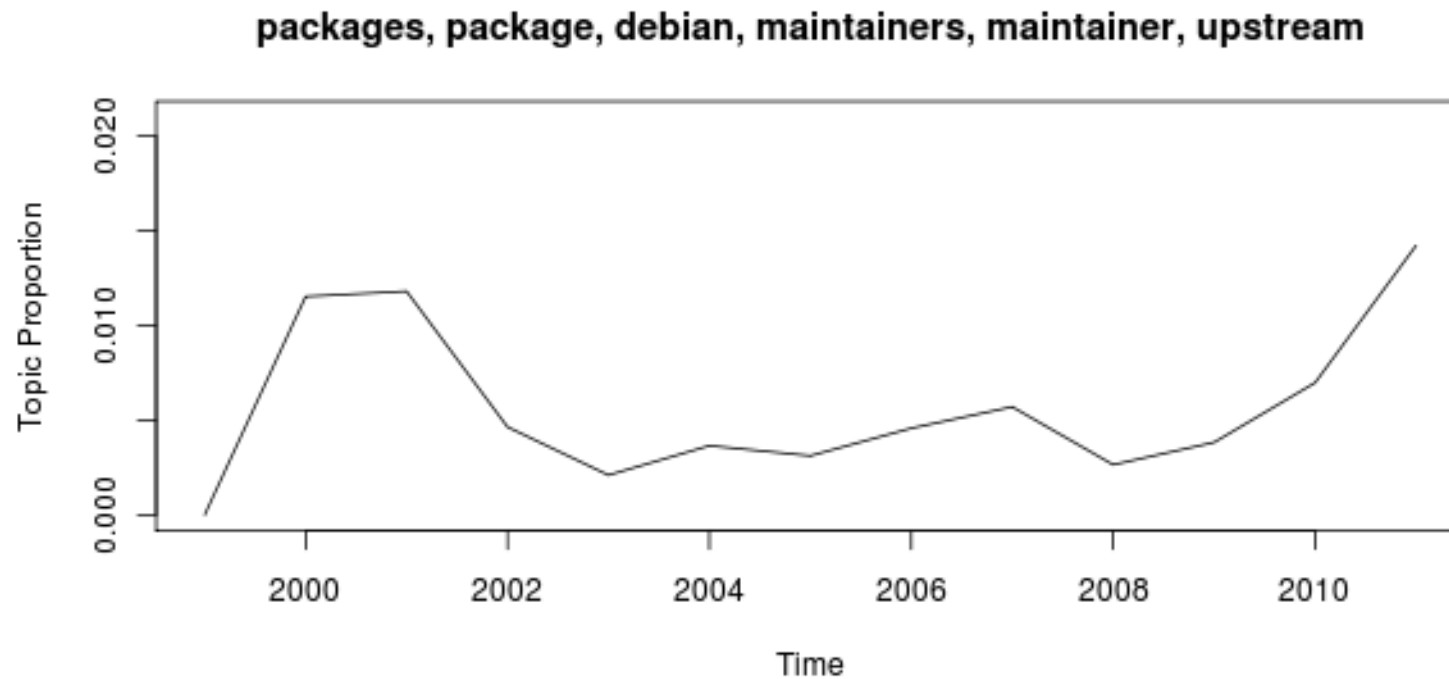
d-project →

package packages install apt-get apt ...	ubuntu debian patches derivatives lts ...	nm process applicant dam fd ...	ftp-master queue packages upload team ...
---	--	--	--

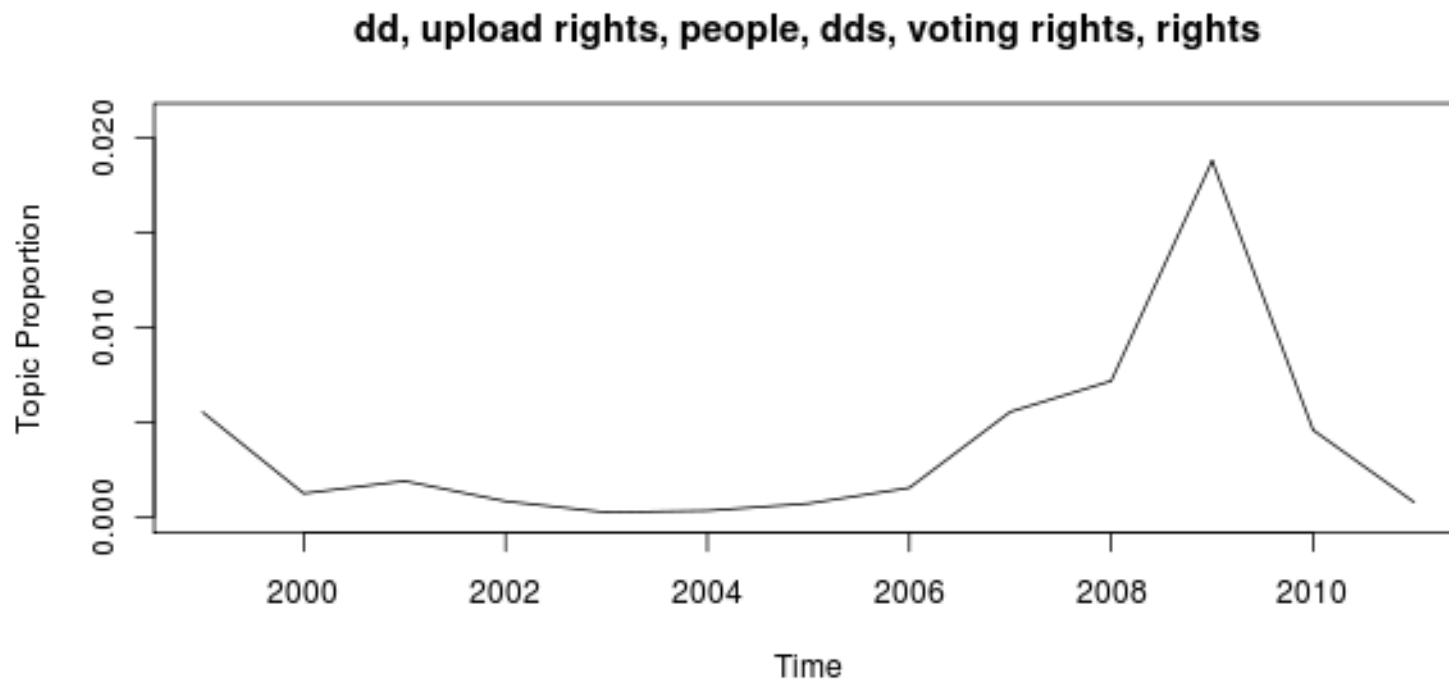
d-women →

women men female male man ...	website page site work d-w ...	post culture response posts behavior ...	nm debian process dd packages ...
--	---	---	--

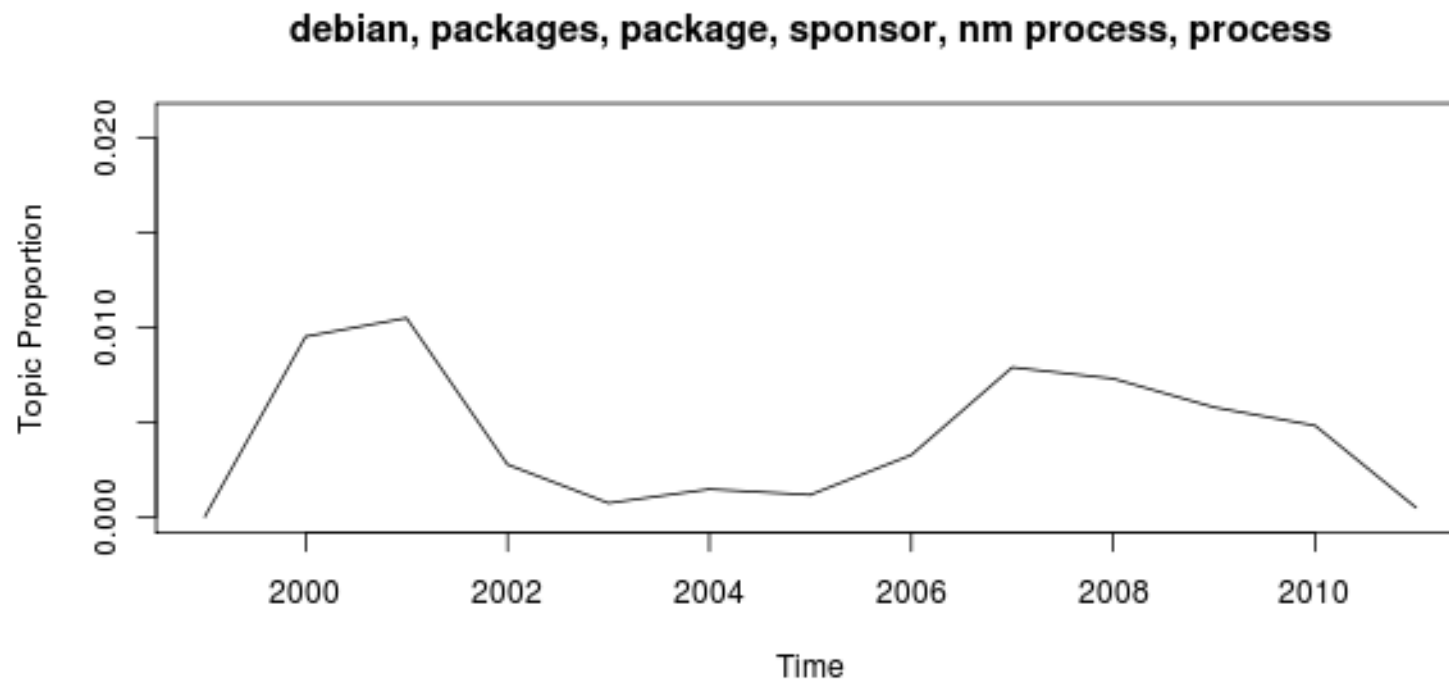
Topic Usage Over Time



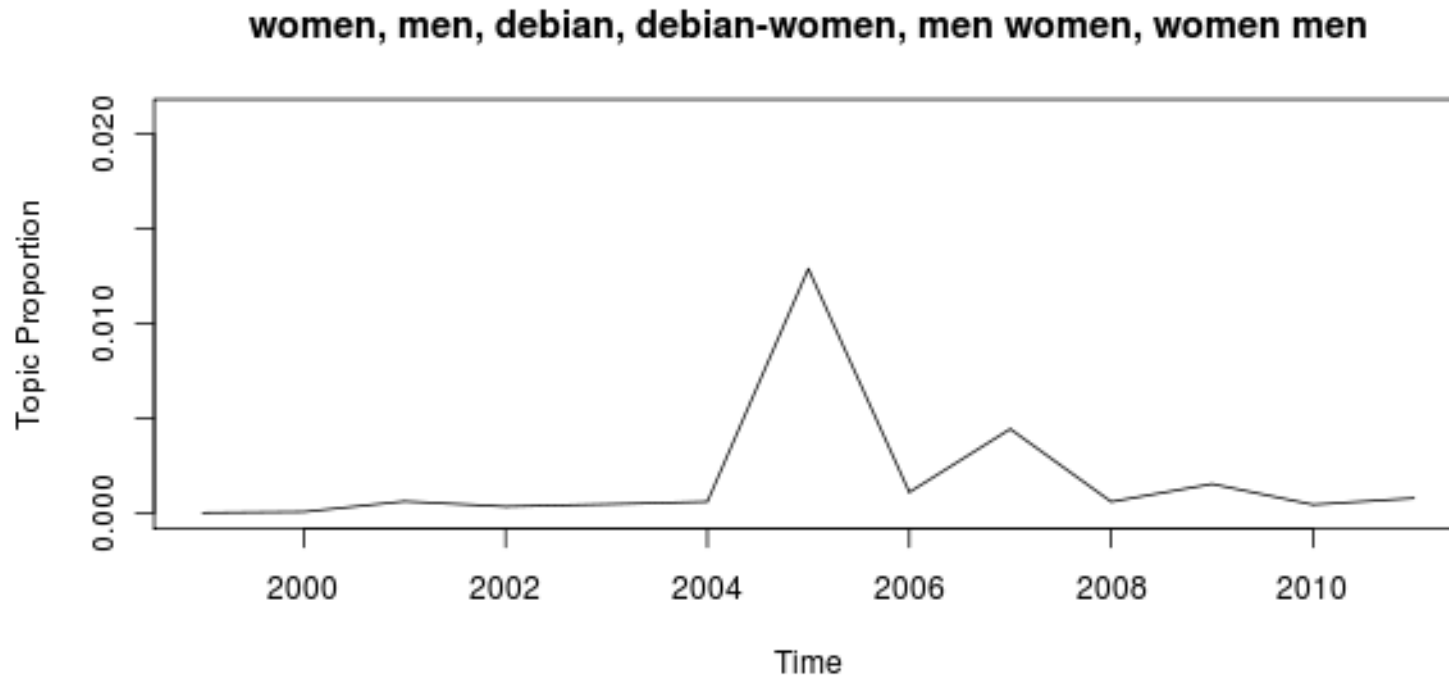
Topic Usage Over Time



Topic Usage Over Time



Topic Usage Over Time



Thanks!

Acknowledgements: James Leach, Dawn Nafus, Bernhard Krieger; Erinn Clark, Helen Faulkner; Chris Ball; David Mimno