Women in Free/Open Source Software Development

Hanna M. Wallach

University of Massachusetts Amherst

wallach@cs.umass.edu

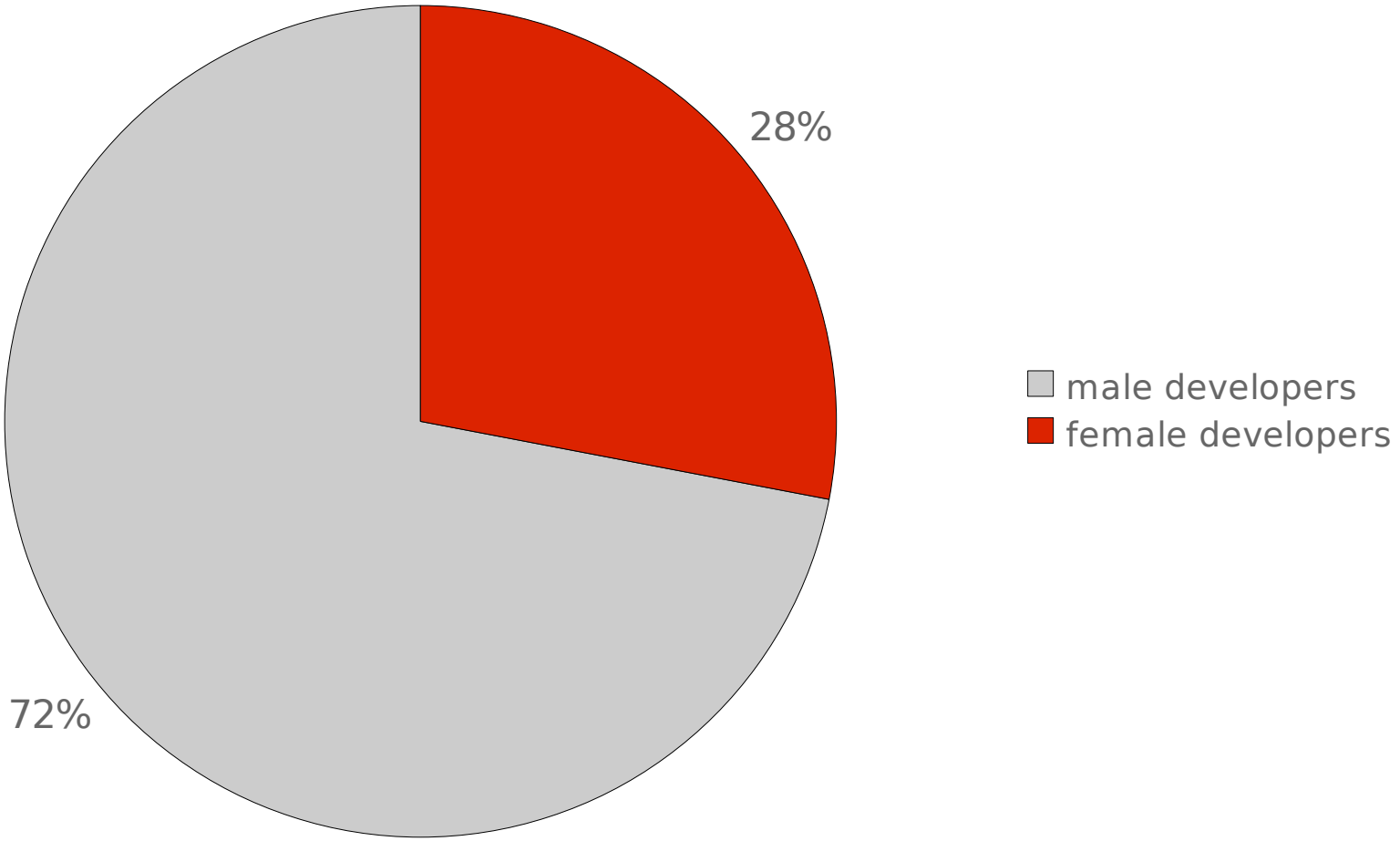
Introduction

debian

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

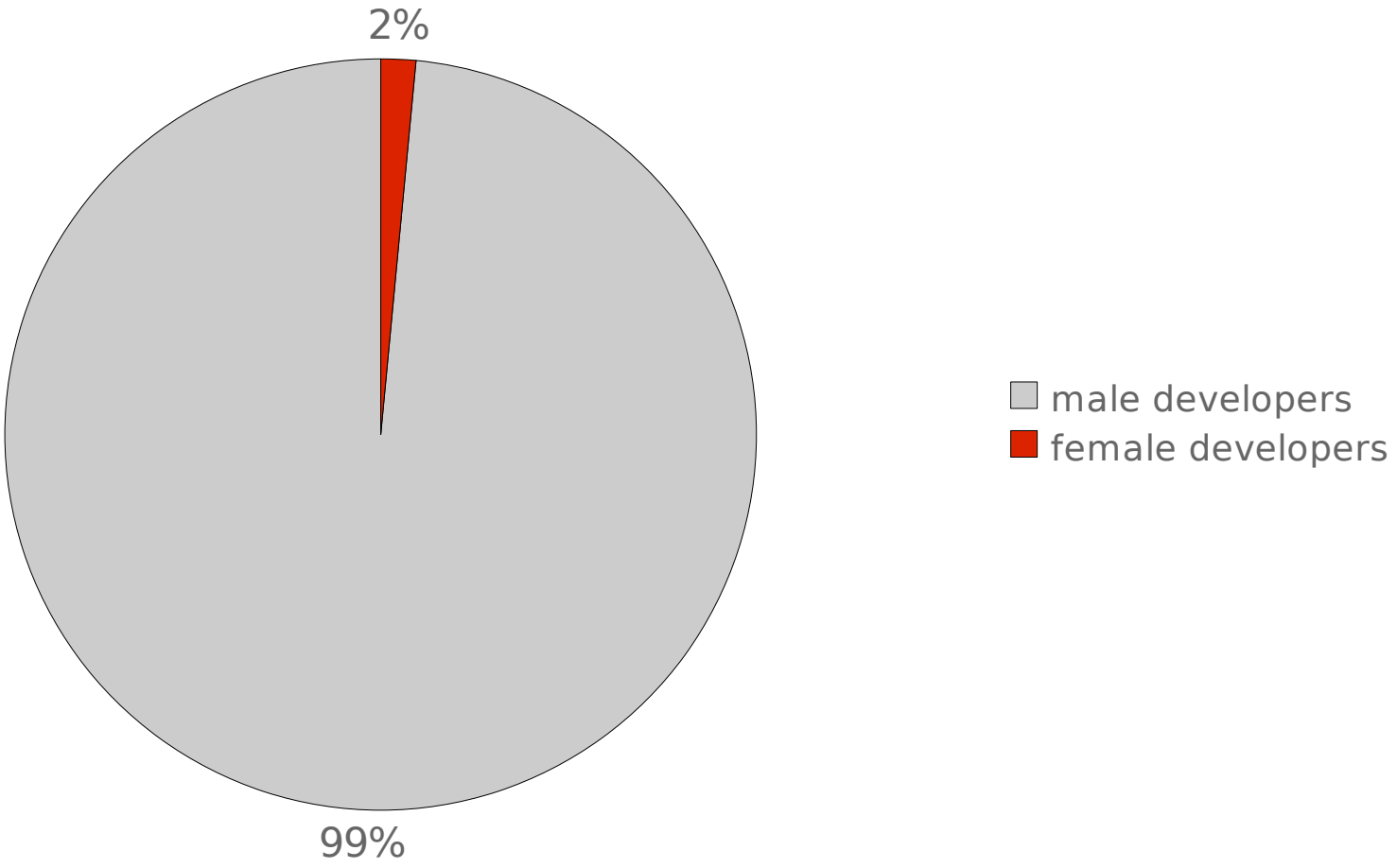
Proprietary Software

[NSF, '04]



Free/Libre/Open Source Software

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



Does Diversity Matter?

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

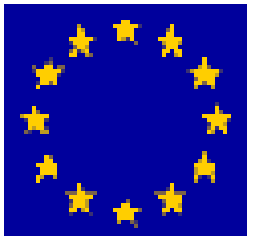
This Talk

- 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

FLOSSPOLS 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

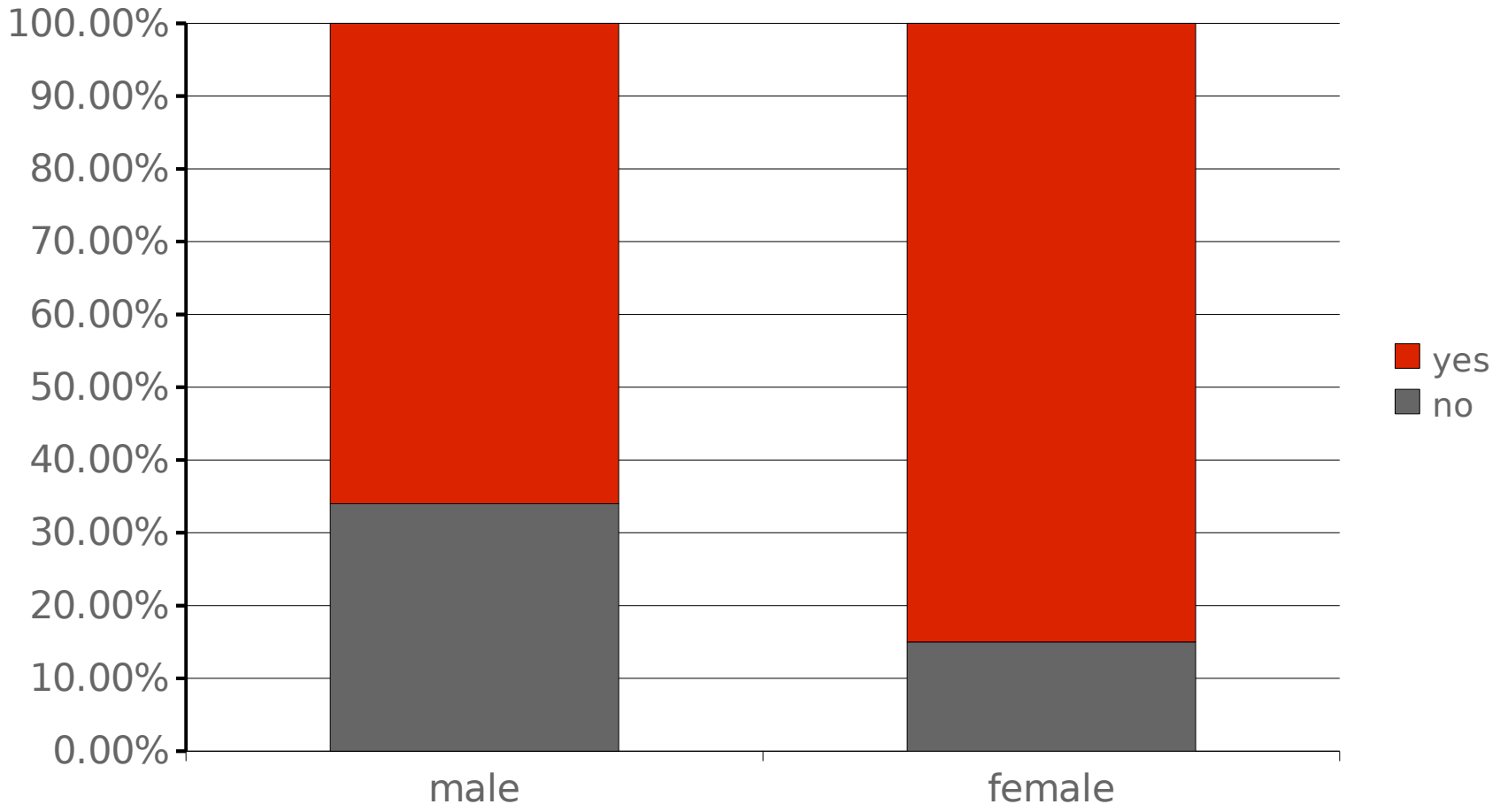
- Produce the first study of gender in FLOSS
- Gather socioeconomic data
- Identify reasons for the gender bias
- Establish policy recommendations for change

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?”



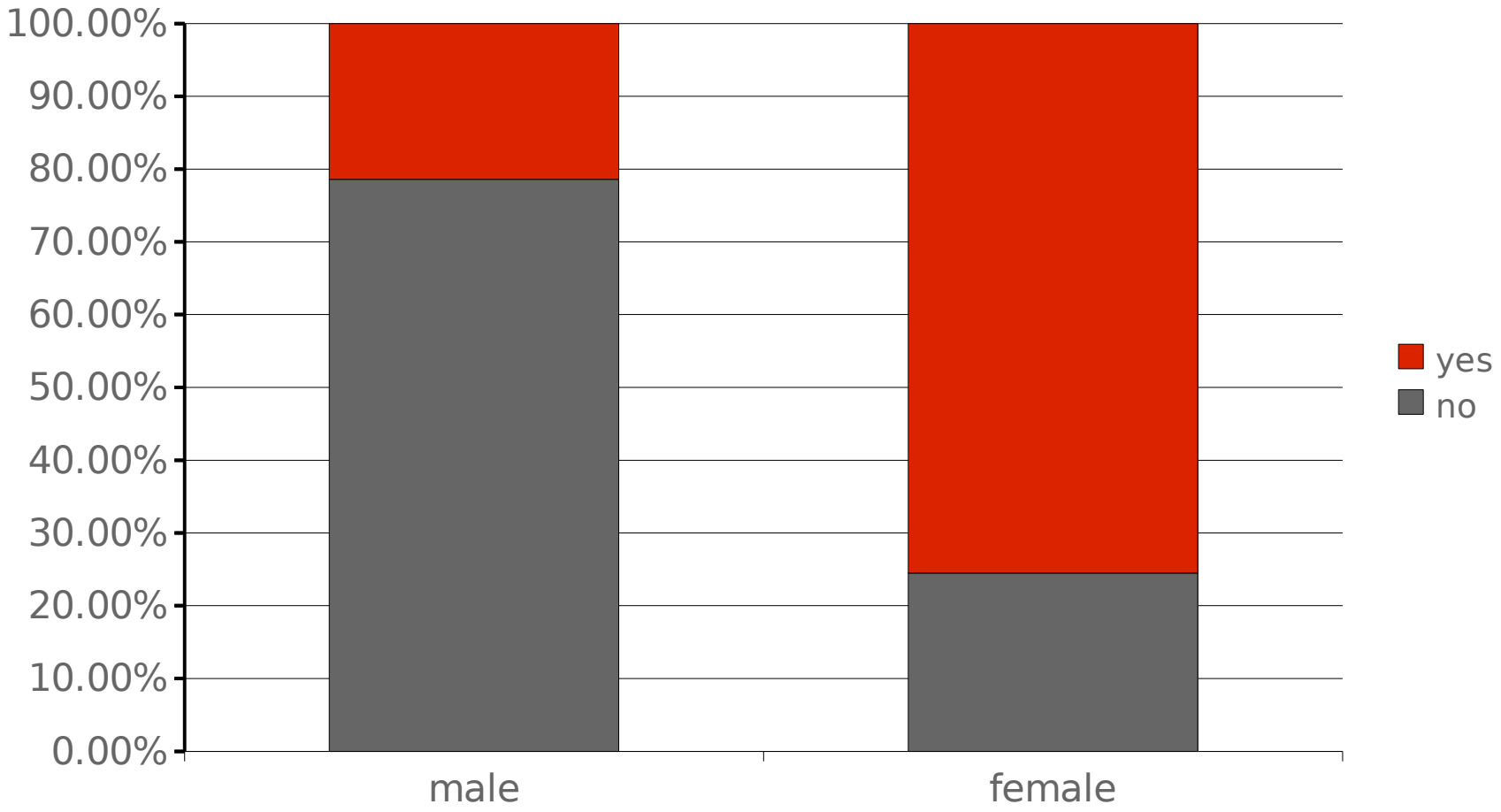
Are women just not interested in FLOSS?

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

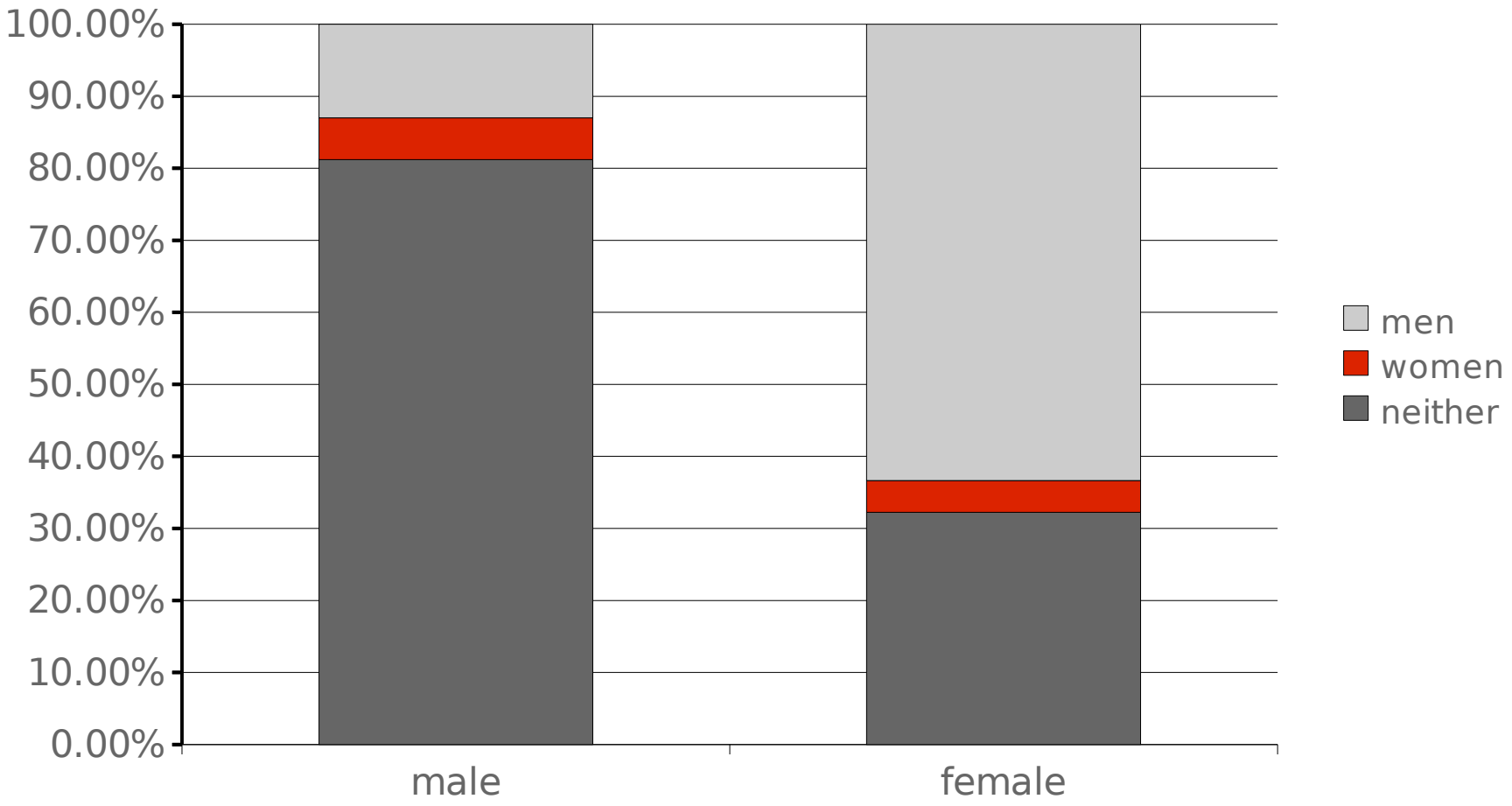
Men:

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

Women:

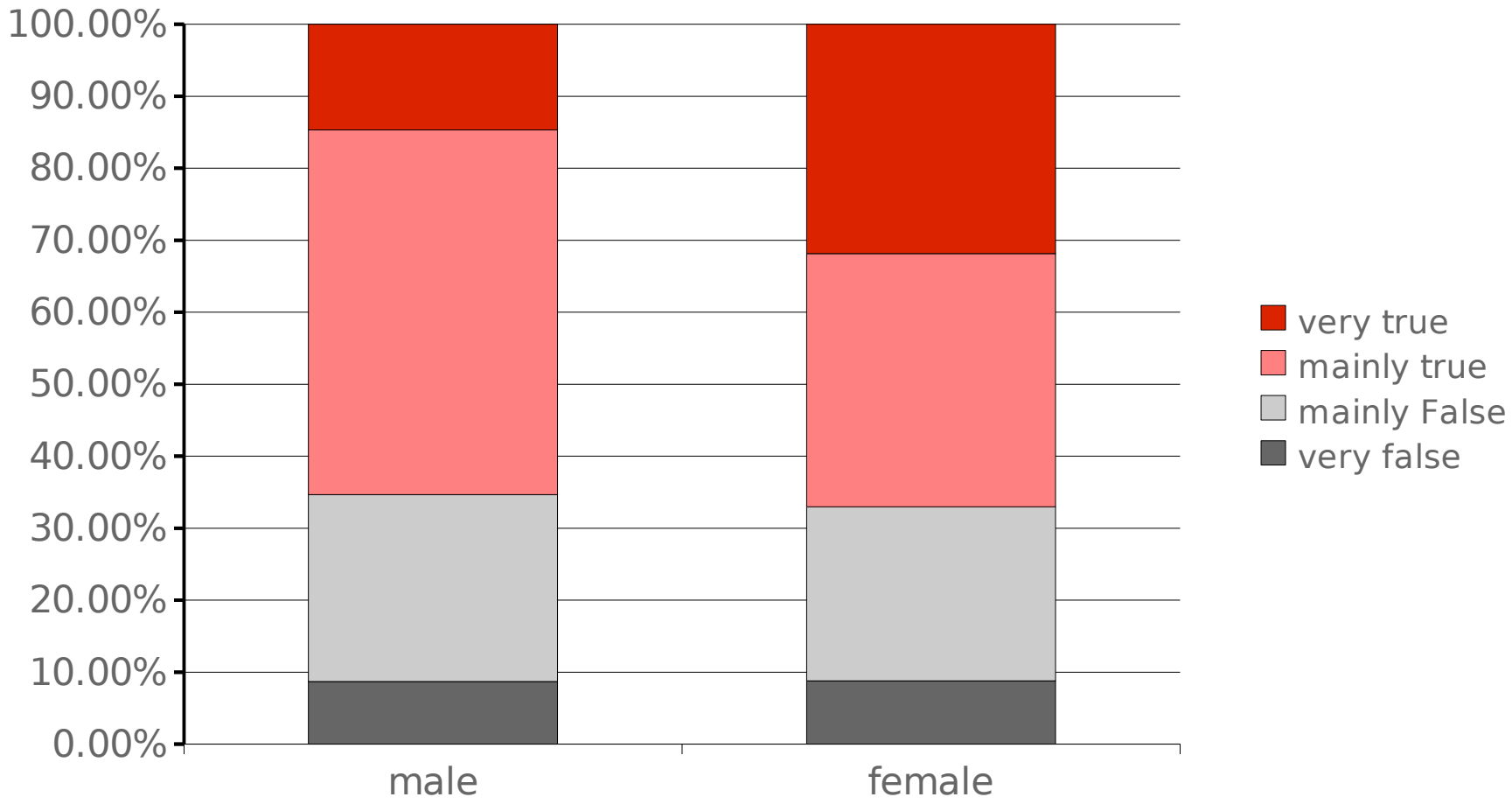
- Documentation
- Organising events
- Translation
- “Social tasks”?

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

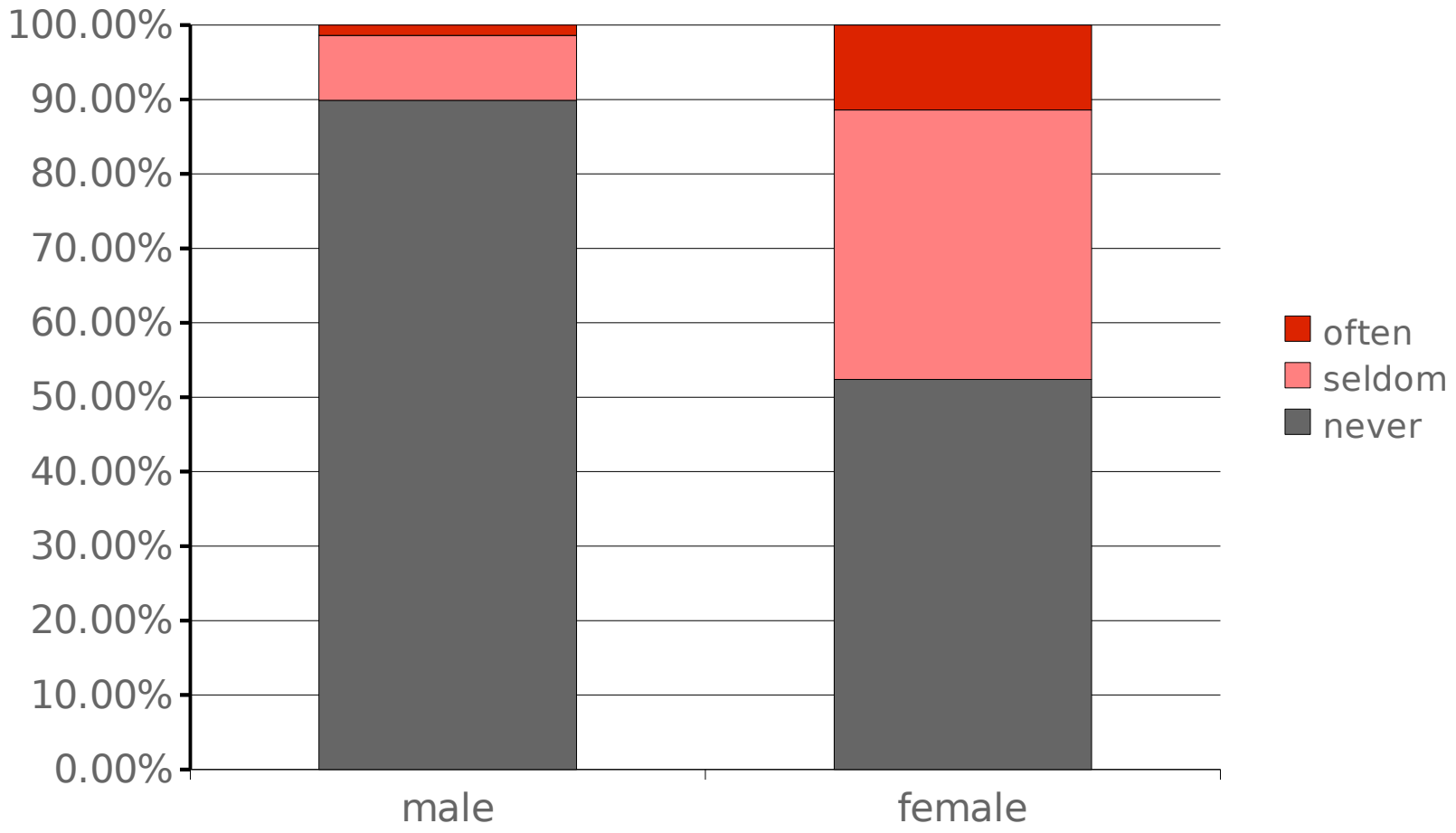


Do women get more attention for their gender or their skills?

“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?”



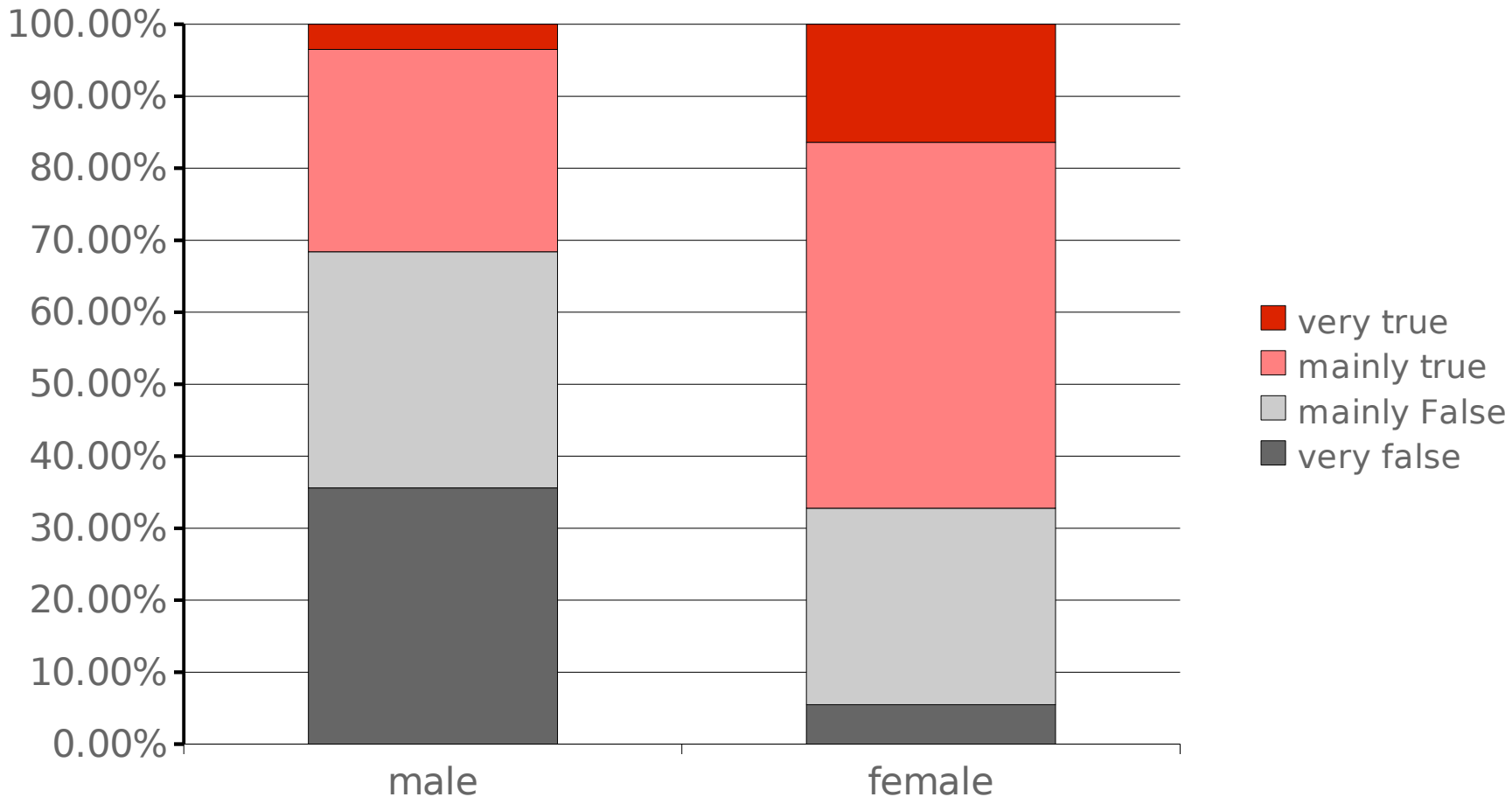
Are there too many flamewars?

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

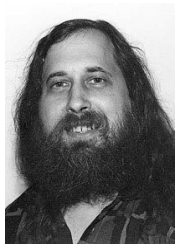
Outcomes & Findings

- 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!
- Targeted, 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

FOSS Development Communities



debian

```
    $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;
```

- Complex technological, legal, social structures
- Massively geographically distributed collaboration
- Social and organizational processes underlying FOSS development are largely unknown

FOSS Collaboration Data

- Most FOSS collaboration data are publicly available:
  - Mailing lists, IRC channels
  - Commit messages, bug reports
  - Comments in source code, documentation
  - GPG keysigning records

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

Data Challenges

- Informal, messy, and often highly unstructured data:
  - Developers use different identifiers in different fora
  - IRC channels have multiple interleaved conversations
  - Mix of highly technical and “off-topic” discussion
  - Conversational style is often casual

⇒ Significant text analysis is required prior to developing models for answering social science questions

Analyzing Debian Mailing Lists

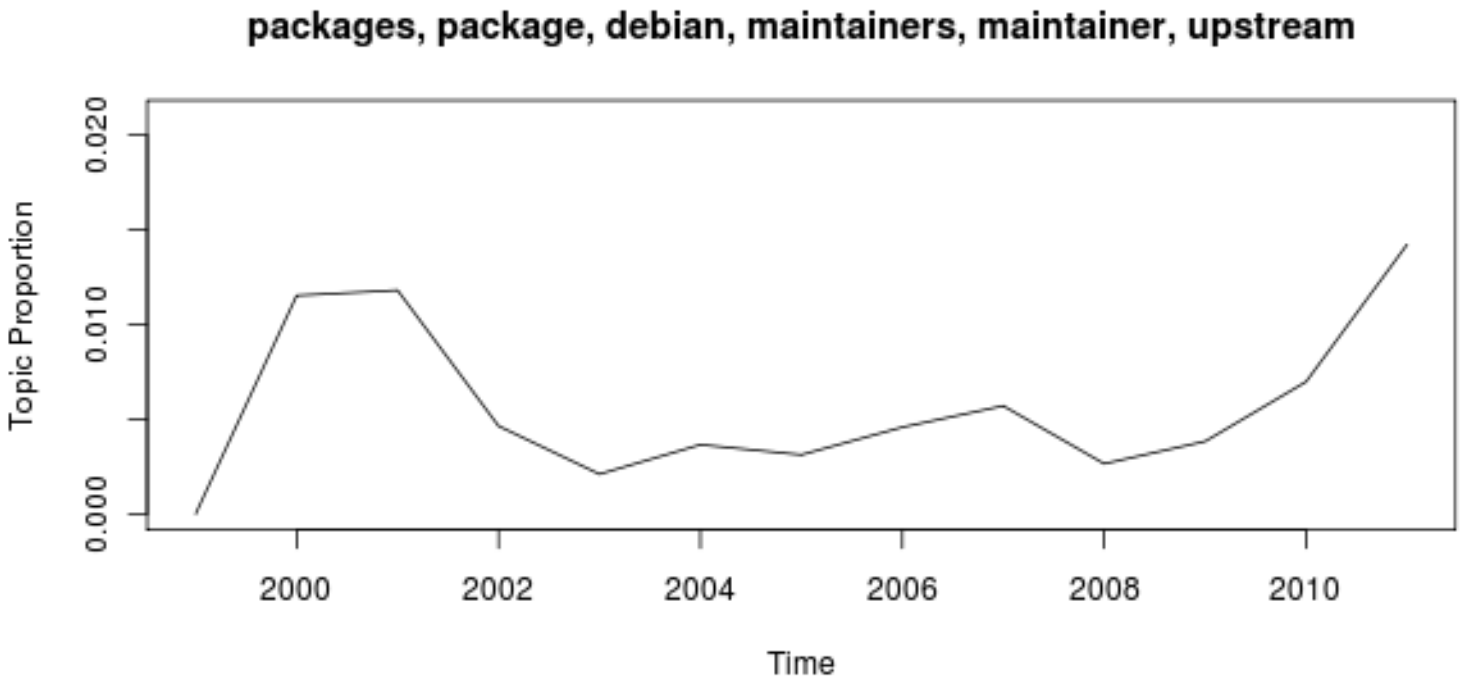
- Quoted text and signatures stripped
- Debian-project mailing list:
  - 19,347 messages
  - 1225797 words (max. 7,916 per message)
- Debian-women mailing list:
  - 4,124 messages
  - 228,076 words (max. 1,524 per message)

100 Topics

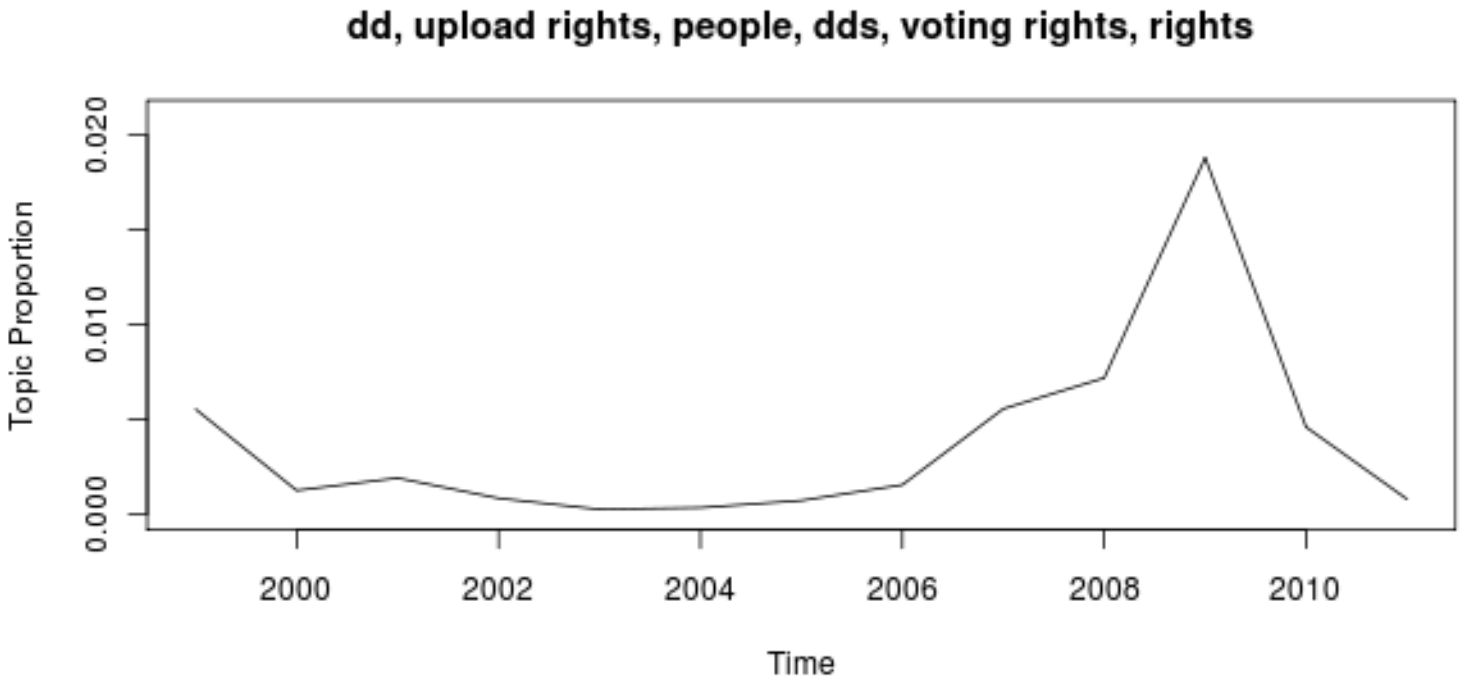
d-project →	package	ubuntu	nm	ftp-master
	packages	debian	process	queue
	install	patches	applicant	packages
	apt-get	derivatives	dam	upload
	apt	lts	fd	team
	...	...	...	...

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

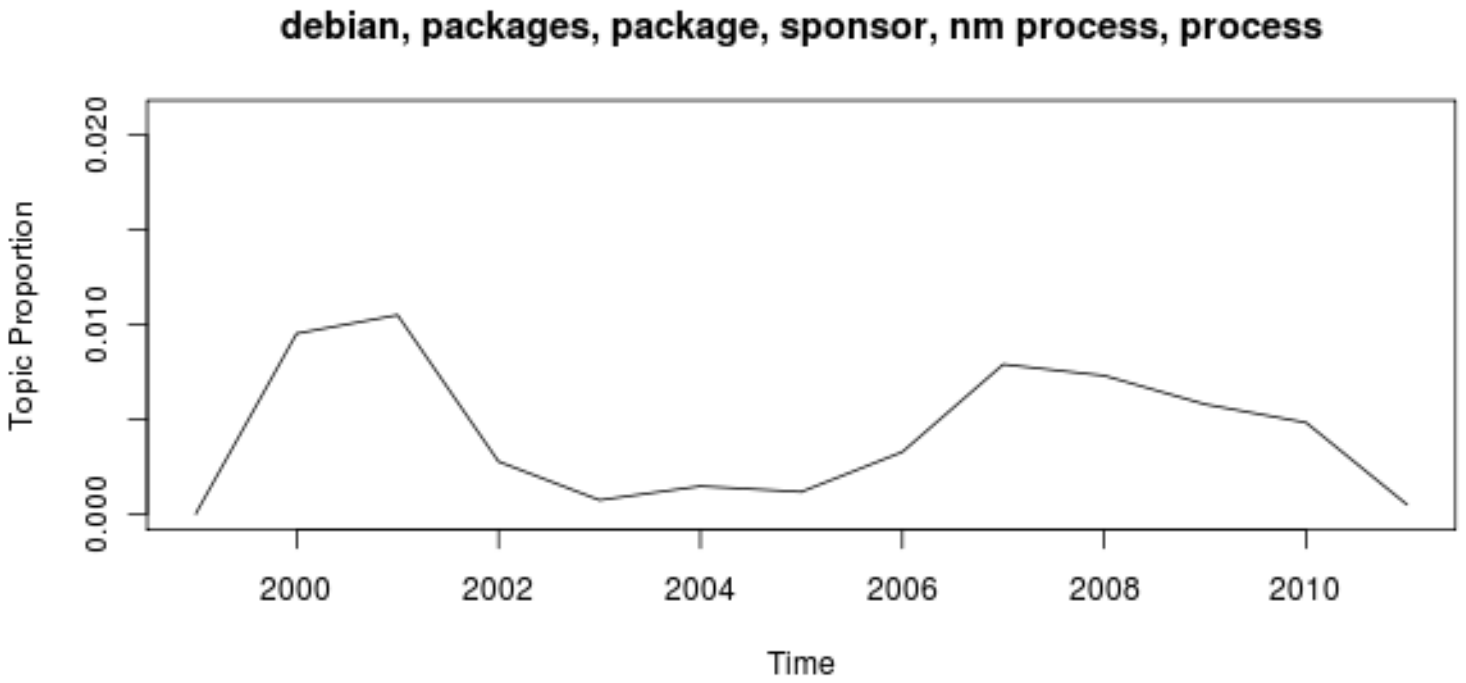
Topic Usage Over Time



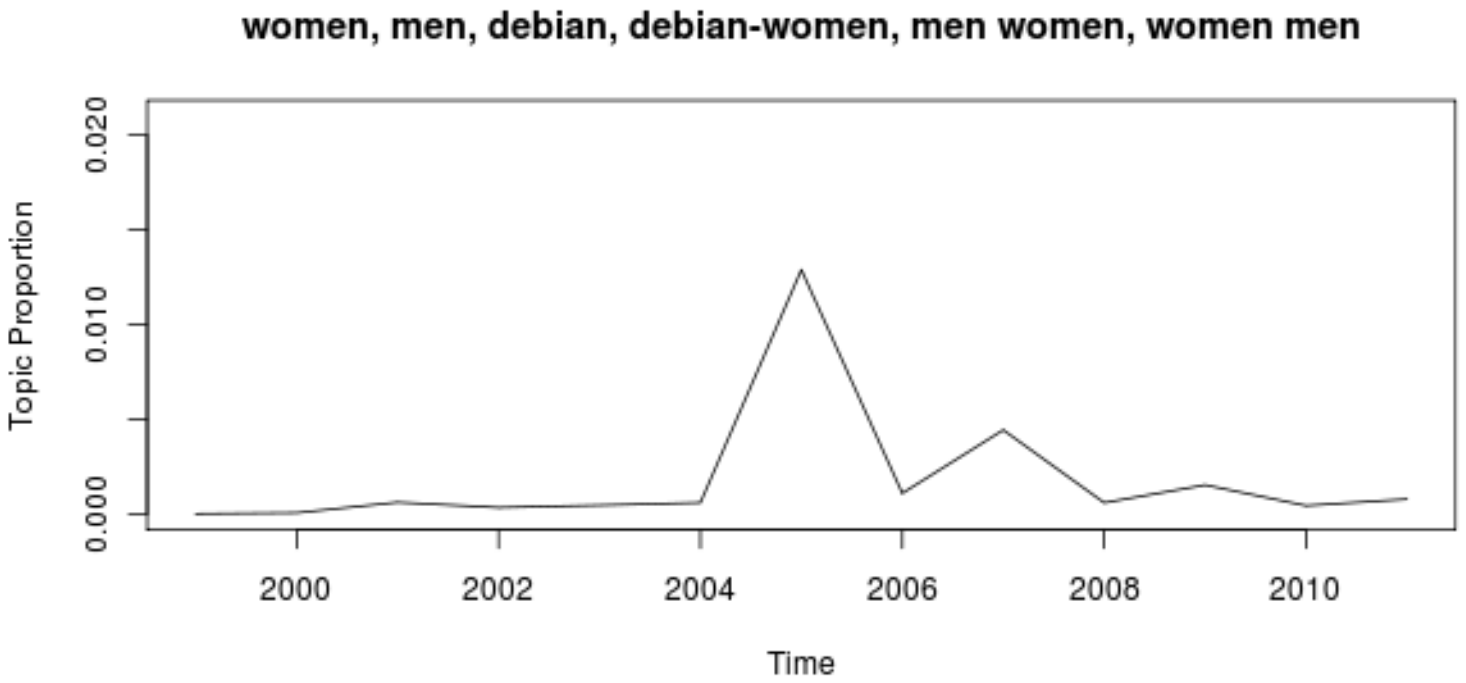
Topic Usage Over Time



Topic Usage Over Time



Topic Usage Over Time



Thanks!

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