# Do Time of Day and Developer Experience Affect Commit Bugginess?

*Jon Eyolfson*, Lin Tan, Patrick Lam

University of Waterloo

May 22, 2011

# Outline

# Motivation

- Software repositories contain a wealth of information
  - What's available?
  - Is it possible to leverage this data?
  - What could we learn?
- Can also create connections between bug introductions and fixes
- We intend to find correlations for these "buggy" commits

# Creating Connections

- Find the bug fixing commits (keyword search) and for each
  - Find lines which changed
  - Use "blame" to find which commit(s) caused the bug
  - Record information to database
- Keyword search precision of 86%–87% and recall of 71%–73%
- Additional information is available

# Additional Information

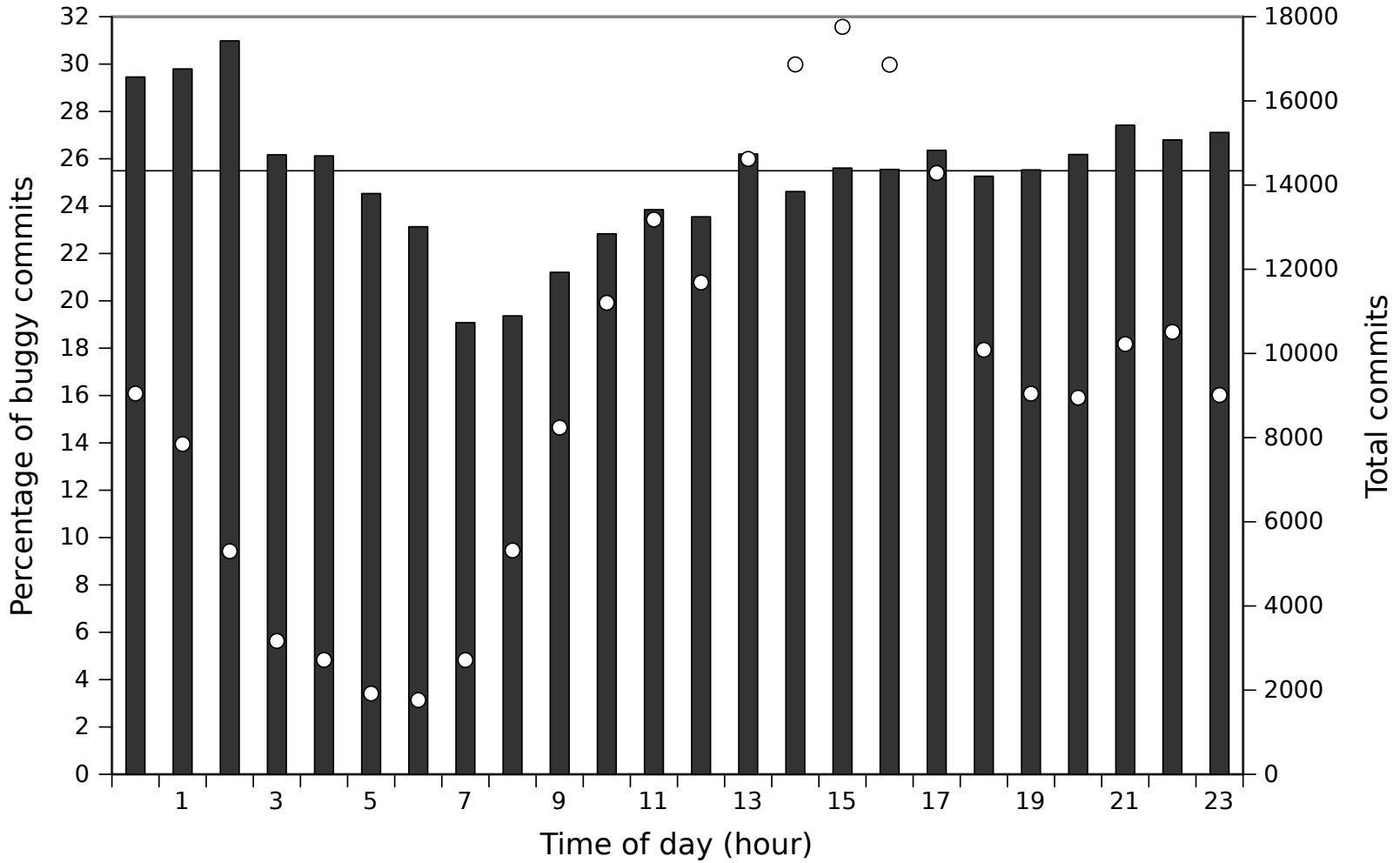
- Record the following
  - Commit times (local and UTC)
  - Number of lines changed in code/comments/other in commit
  - Merge authors with same name/email
- We can now determine
  - Developer experience
  - Bug lifetime
  - Whether a commit contains a bug and how many fixes were applied

# Repositories

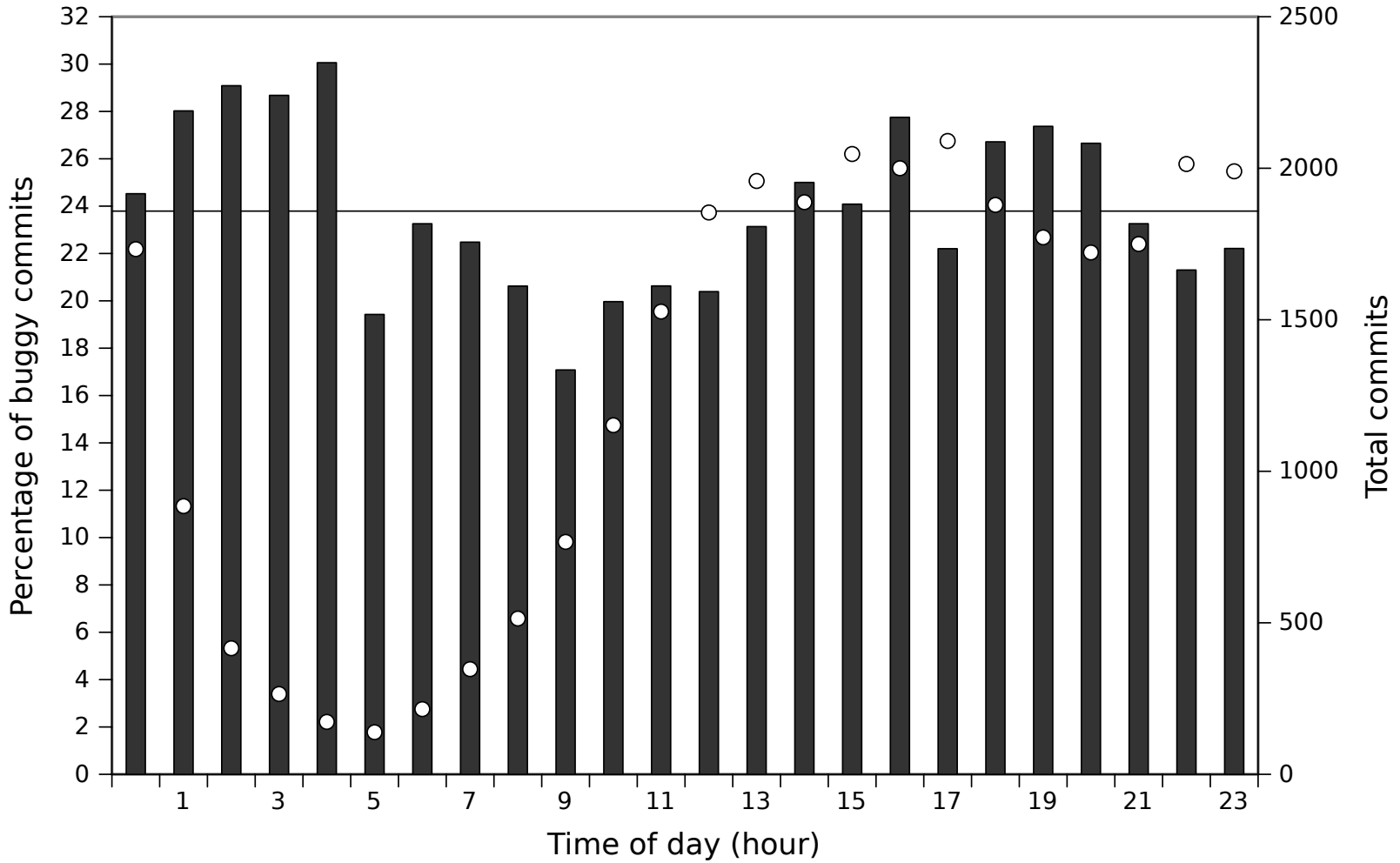
| | **Linux kernel** | **PostgreSQL** |
|---|---|---|
| First commit | April 16, 2005 | July 9, 1996 |
| Last commit | Nov. 21, 2010 | Jan. 24, 2011 |
| Lines of code | over 5 million | over 750,000 |
| Number of authors | 8,594 | 34 |
| Total commits | 222,332 | 31,098 |
| Introducing commits | 56,681 (25.5%) | 7,366 (23.7%) |
| Fixing commits | 57,028 | 4,399 |

- Note: results on website have better author merging and minor improvements

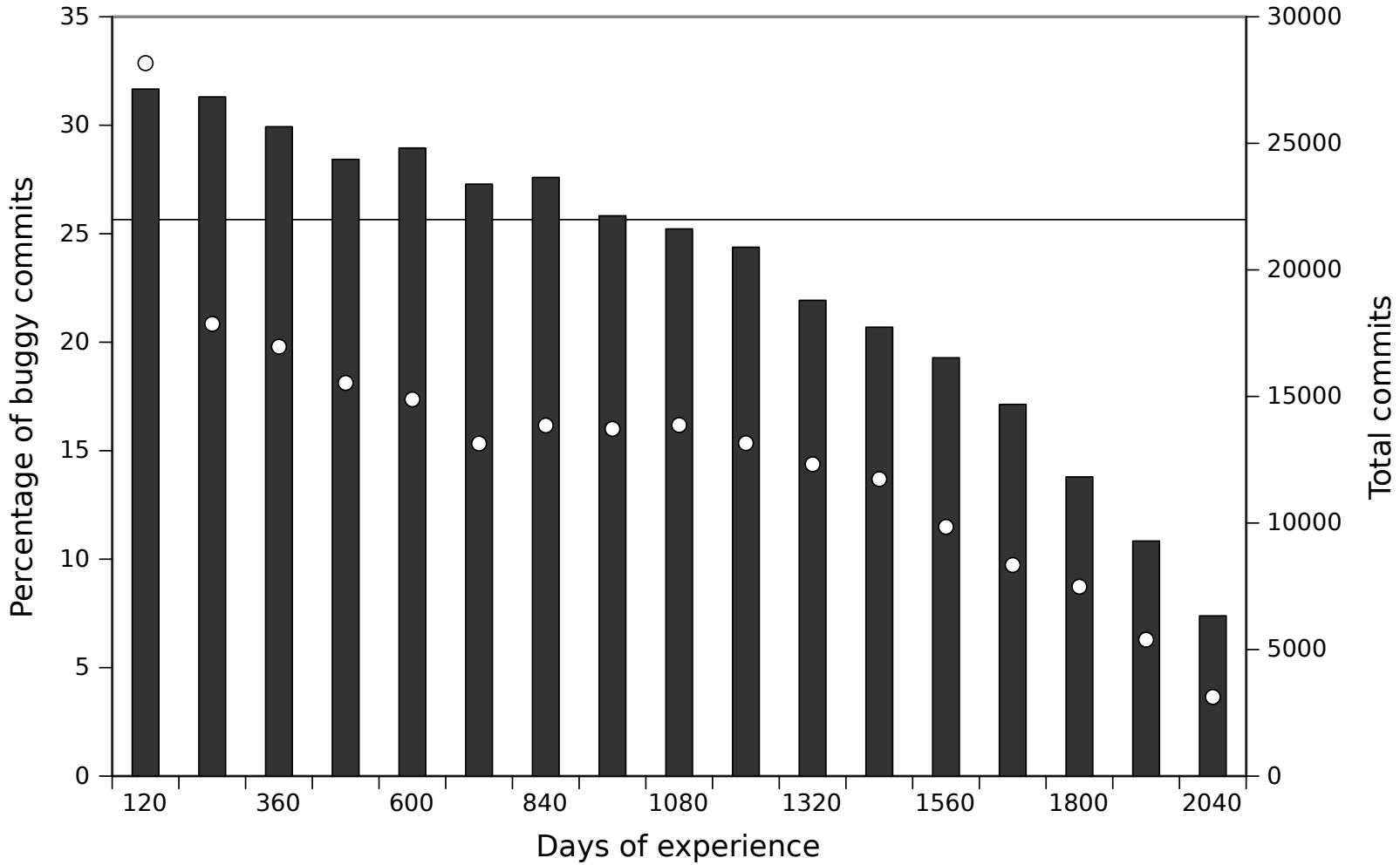
# Does Time-of-day Affect Linux?



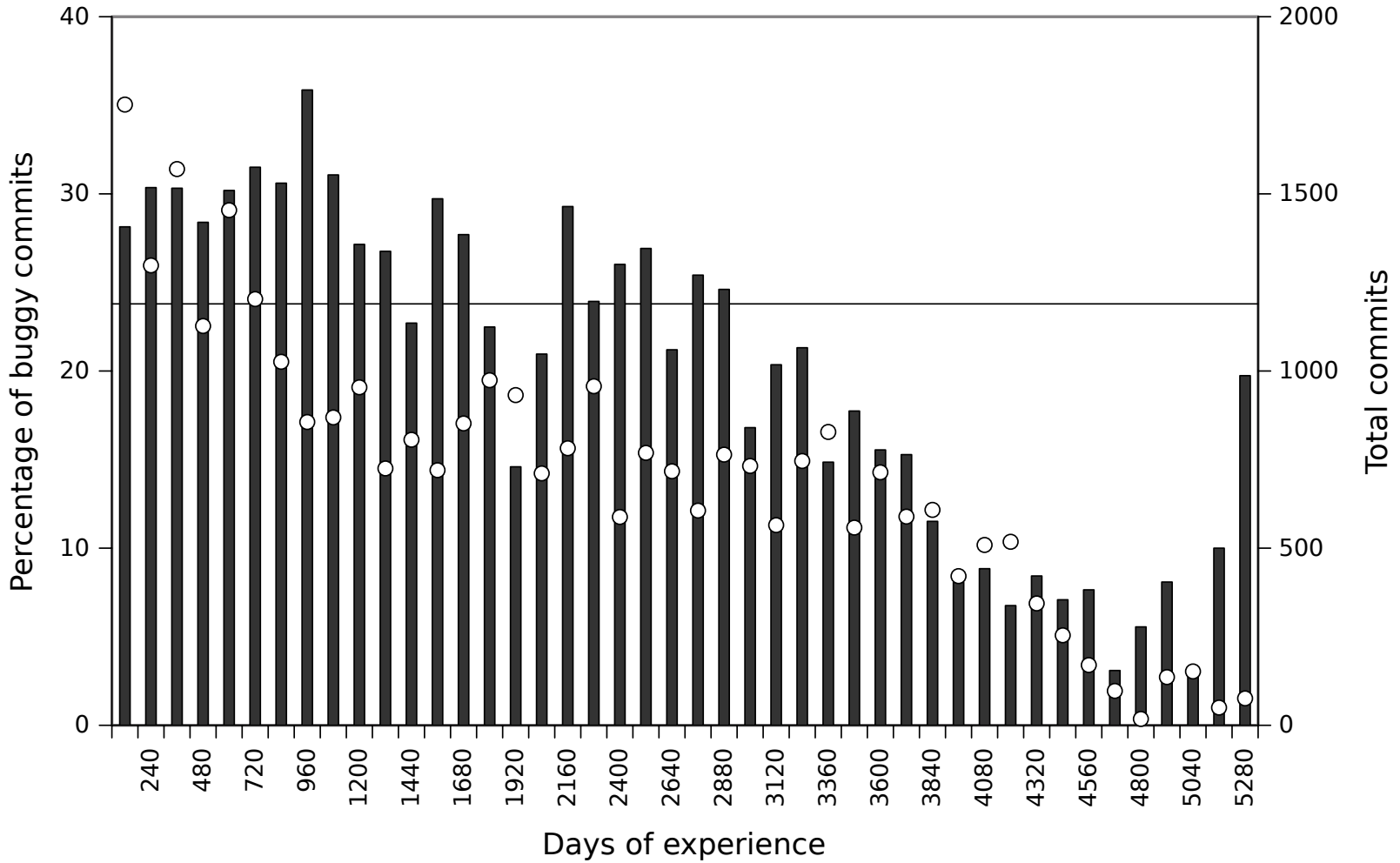
# What About Time-of-day and PostgreSQL?



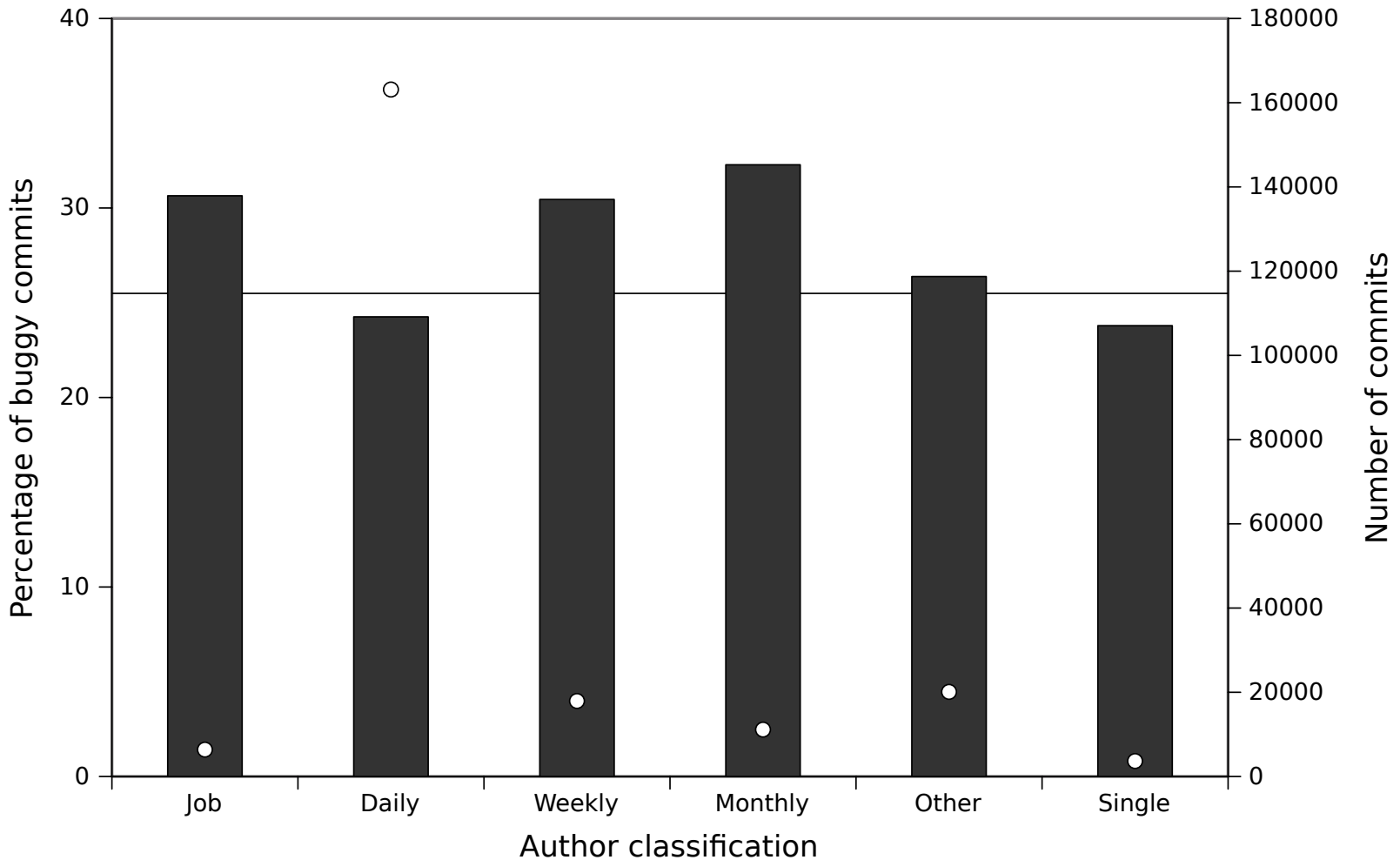
# Do Experienced Linux Developers Commit Less Bugs?



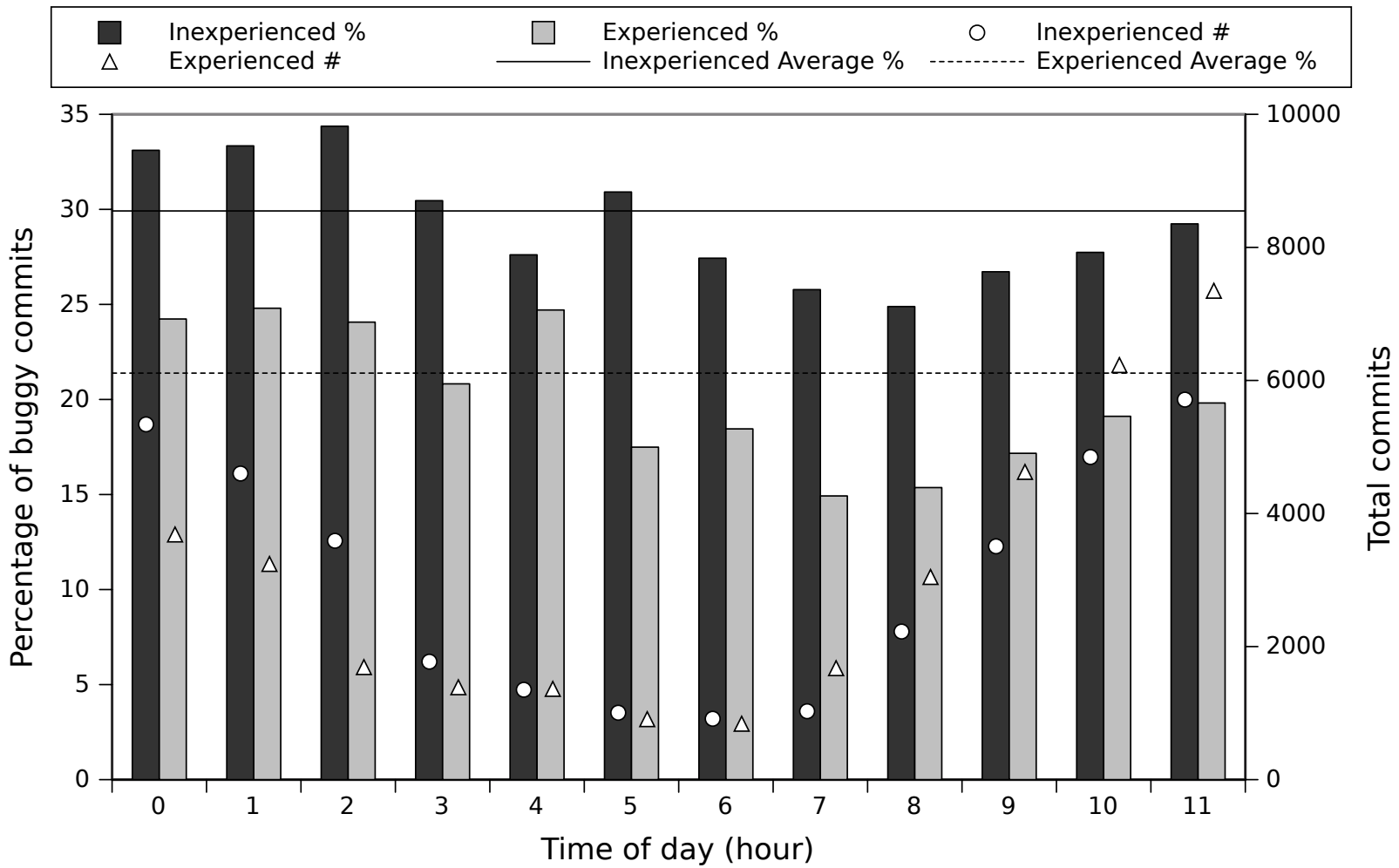
# Does PostgreSQL Follow the Same Trend?



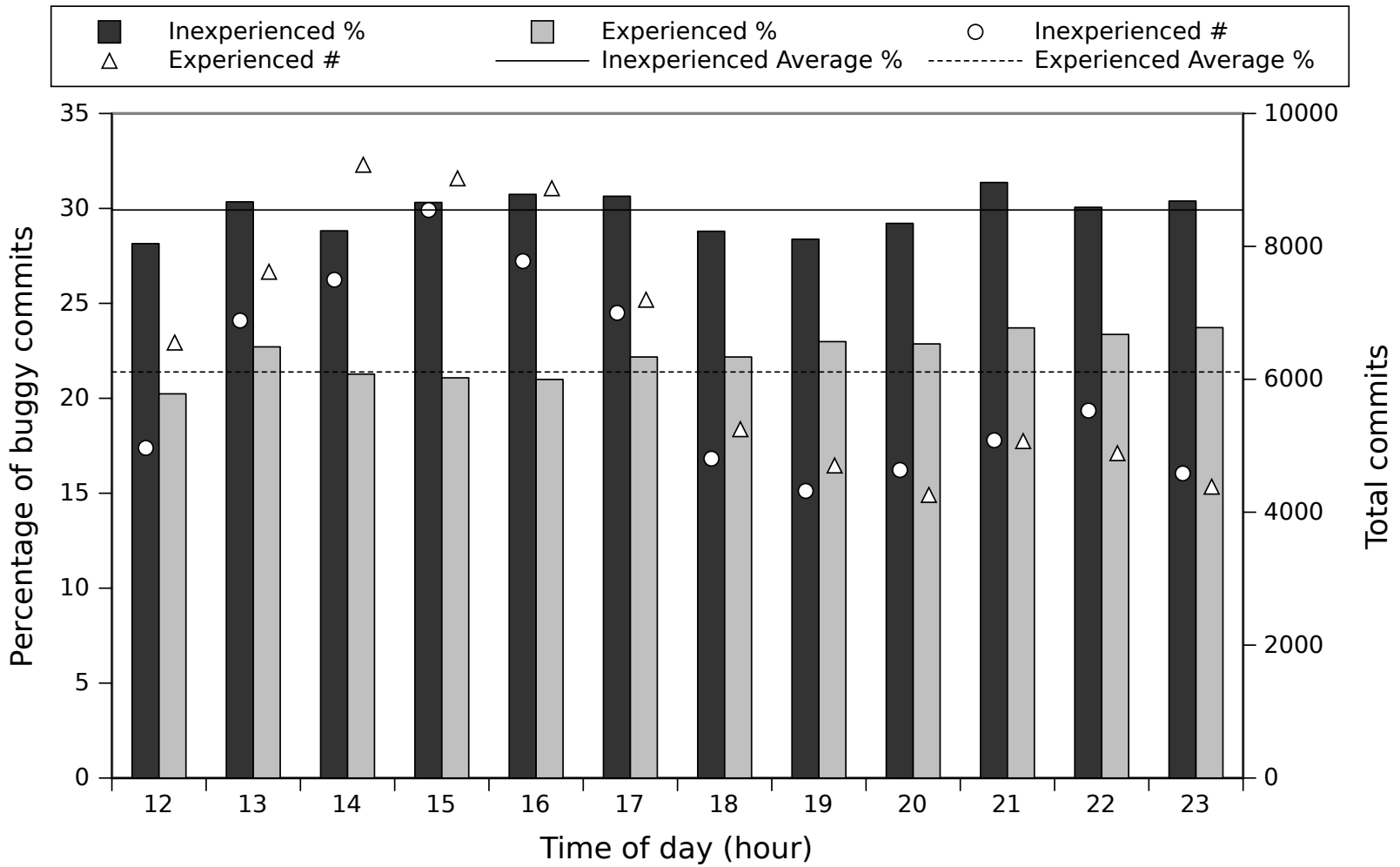
# Does Developer Activeness Matter for Linux?



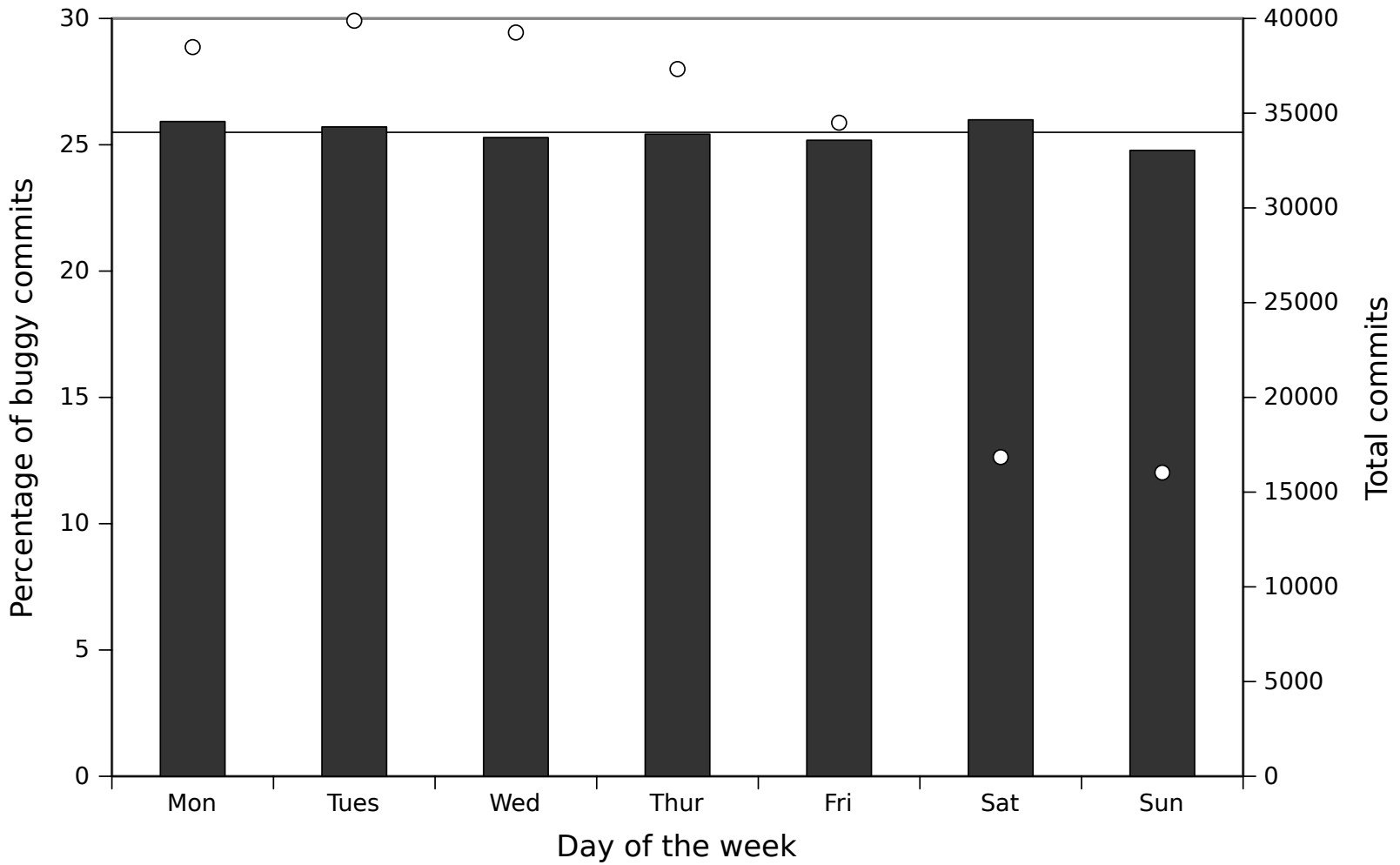
# Differences for Inexperienced and Experienced Linux Developers (1)



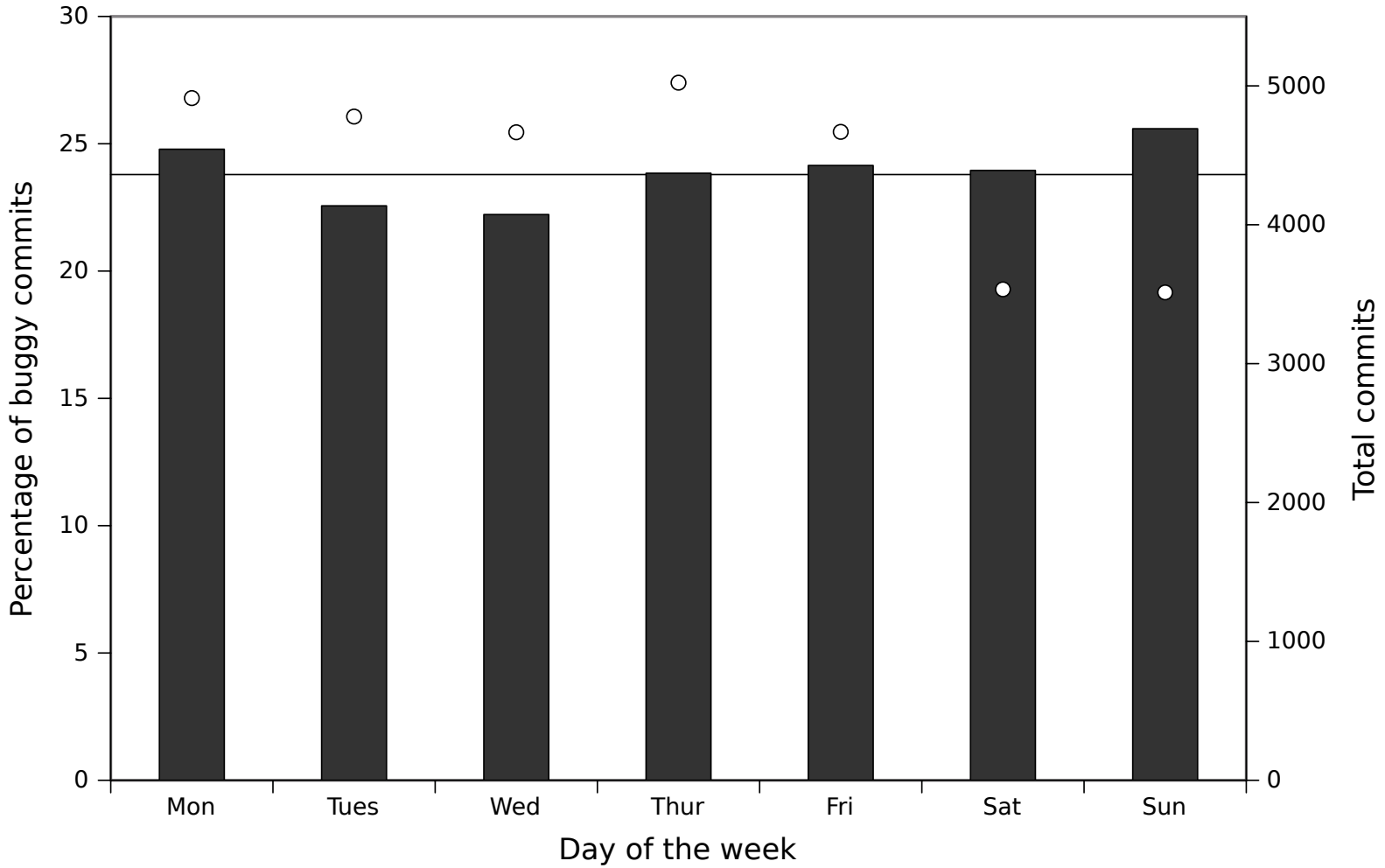
# Differences for Inexperienced and Experienced Linux Developers (2)



# Is a "Case of the Mondays" True for Linux?



# What are the Worst Days for PostgreSQL?



# Previous Studies

- Commits for Eclipse and Mozilla were found to be buggiest on Fridays [4]
- Classification of commits into different categories [2]
- Bug lifetimes for Linux and PostgreSQL [1, 3]
- We found average buglifetimes of
  - 1.38 years ($\sigma = 1.35$) for Linux
  - 3.07 years ($\sigma = 3.19$) for PostgreSQL

# Summary of Findings

- Commits between midnight and 4 AM are more likely to be buggy
- Commits between 7 AM and noon are less likely to be buggy
- More active developers commit less bugs
- More experienced developers commit less bugs
- The worst day of the week varies between projects

# For the Future

- Study individual developers
  - Are commits outside their normal schedule worse?
  - Experience including other open-source projects?
- More software projects
- Correlations involving code quality
- Data is now broswable at `http://www.eyolfson.com/scc/`

# References (1)



D. Engler, D. Y. Chen, S. Hallem, A. Chou, and B. Chelf.
Bugs as deviant behavior: A general approach to inferring errors in systems code.
*SIGOPS OSR*, 35(5):57–72, 2001.


A. Hindle, D. M. German, and R. Holt.
What do large commits tell us?: A taxonomical study of large commits.
In *MSR*, pages 99–108, 2008.


S. Kim and E. Whitehead Jr.
How long did it take to fix bugs?
In *MSR*, pages 173–174, 2006.

# References (2)


J. Śliwerski, T. Zimmermann, and A. Zeller.
When do changes induce fixes?
In *MSR*, pages 24–28, 2005.