The impact of operating systems and environments on build results

Mahdis Zolfagharinia

Master Thesis Defense

Faculty of Computer Engineering

Advisors: Dr. Bram Adams and Dr.Yann-Gaël Guéhéneuc

December 12, 2017

Continuous Integration (CI)

cornerstone of modern quality assurance, providing on-demand builds of code changes or software releases.

+

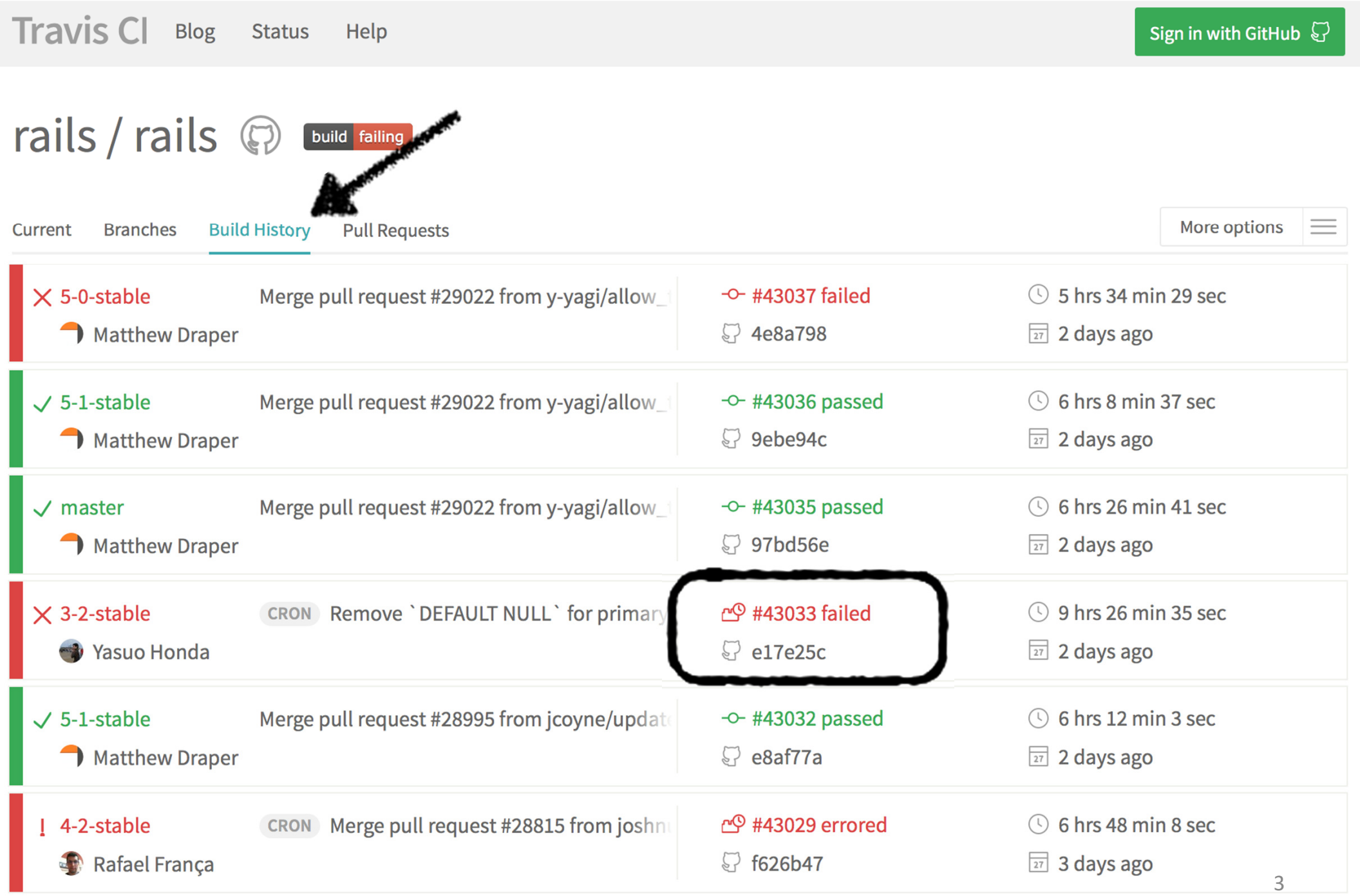
+

Travis CI

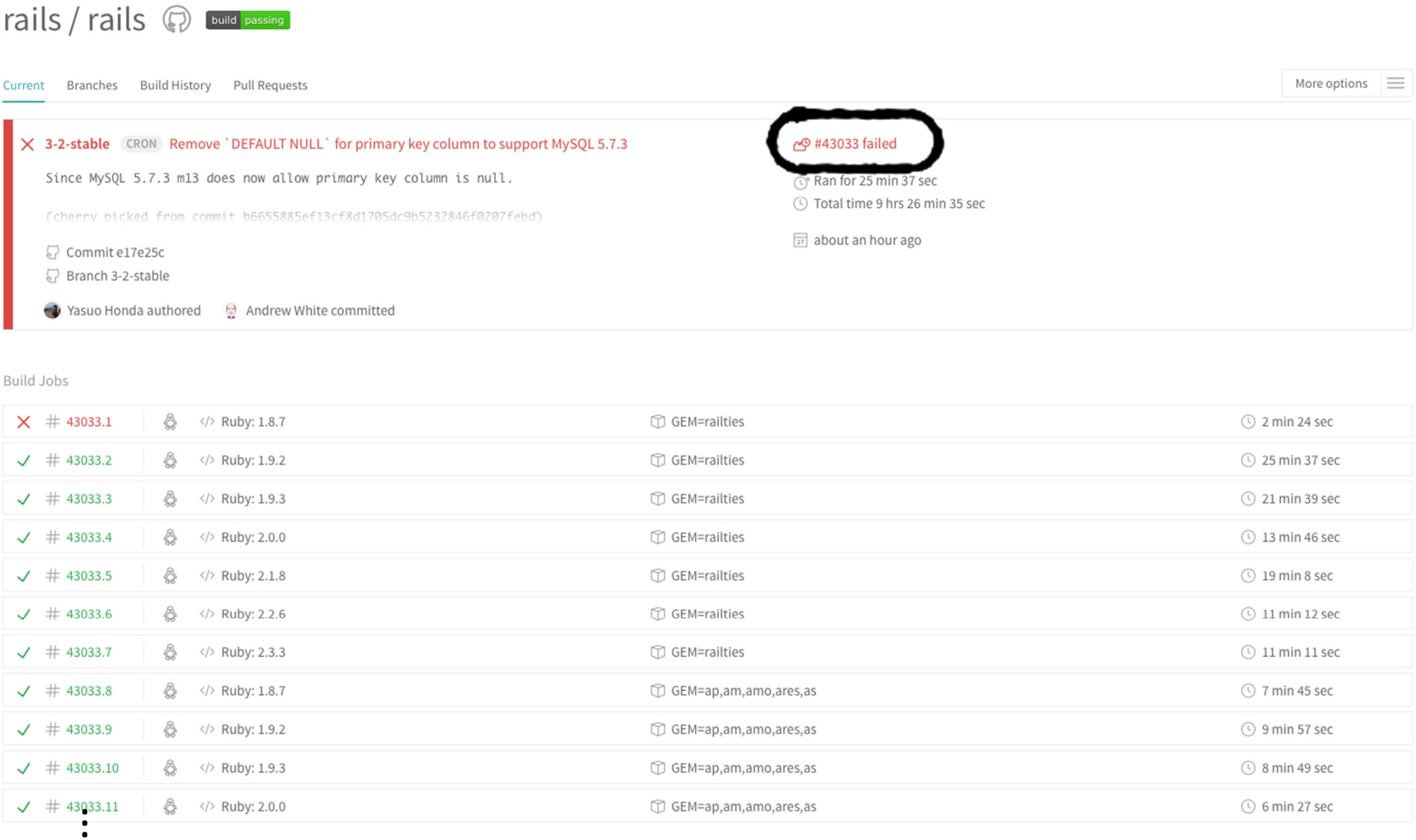
GitLab

Jenkins

Travis CI Build Environment

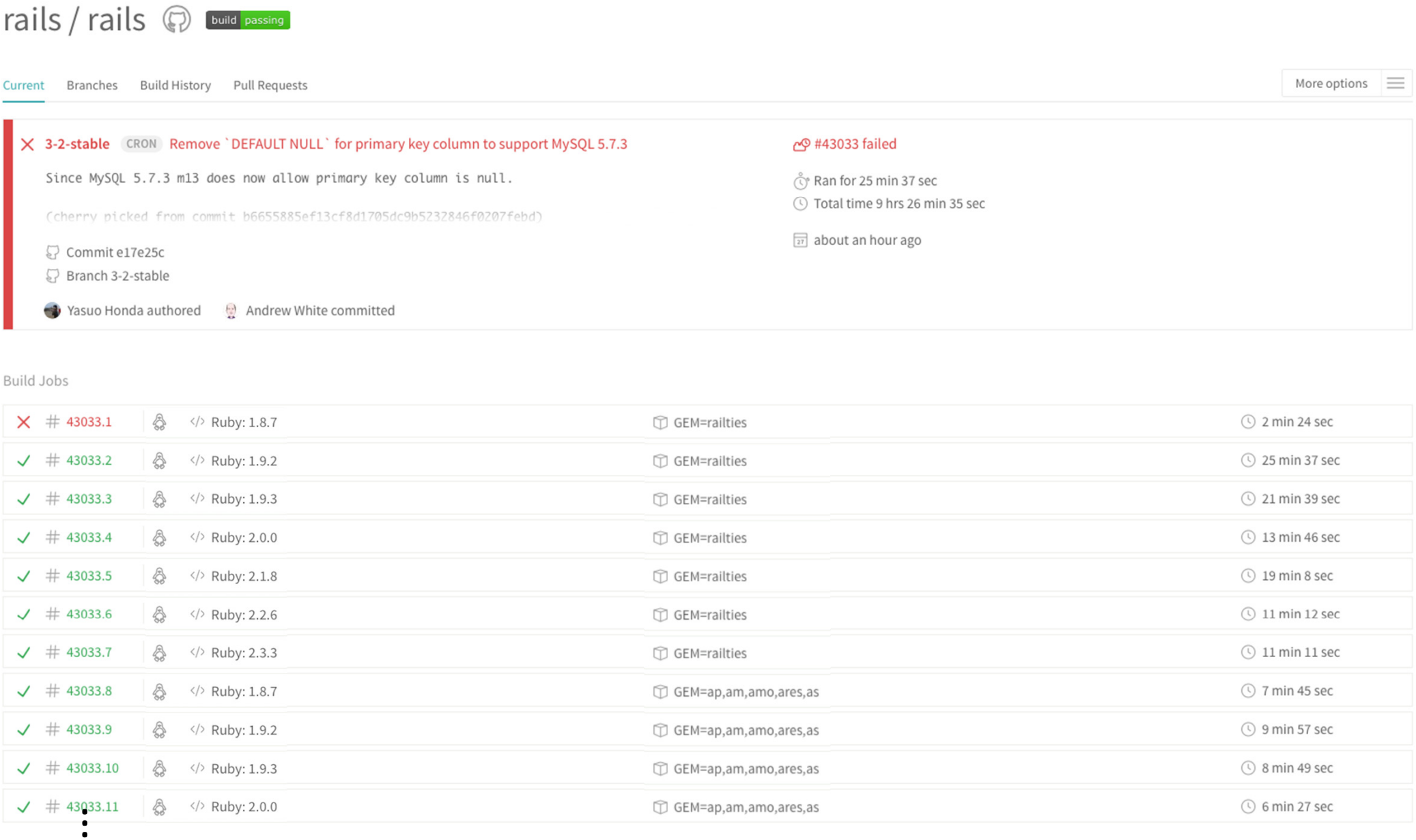


Travis CI Build Environment



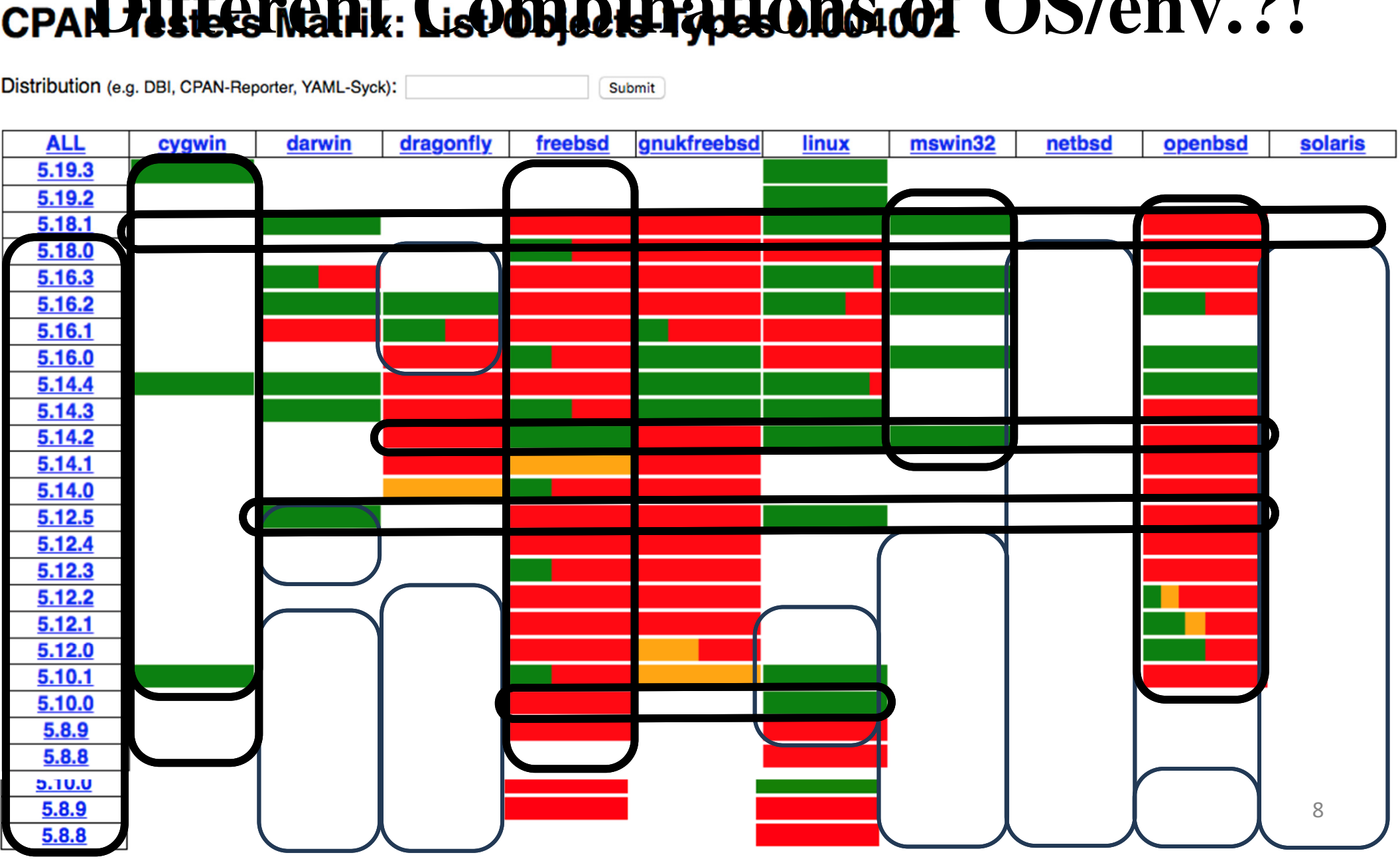

1 Commit
42 Builds
1 failure
41 Successes

Travis CI Build Environment



Travis CI build environment

Not just Travis CI …
How to Interpret Build Results for Different Combinations of OS/env.?!



The impact of OSes and environments on build results to analyze the phenomenon of "build inflation"

Difficult to interpret build results and detect

Practitioners and researchers should take it into consideration when analyzing build results

☑ **Hilton et al. ASE 2016:**

- Analyzed ~35K open-source projects of GitHub
- Over 40% of these projects use CI
- Analyzed ~1.5M builds of Travis CI

- CI is widely adopted in popular projects and helped to decrease the time between releases

☑ **Kerzazi et al. SANER 2016:**

- Studied releases of a large e-commerce web app

- Source code is not the main cause of problems, but defective configurations are the main issues

☑ **Seo et al. ACM 2014**

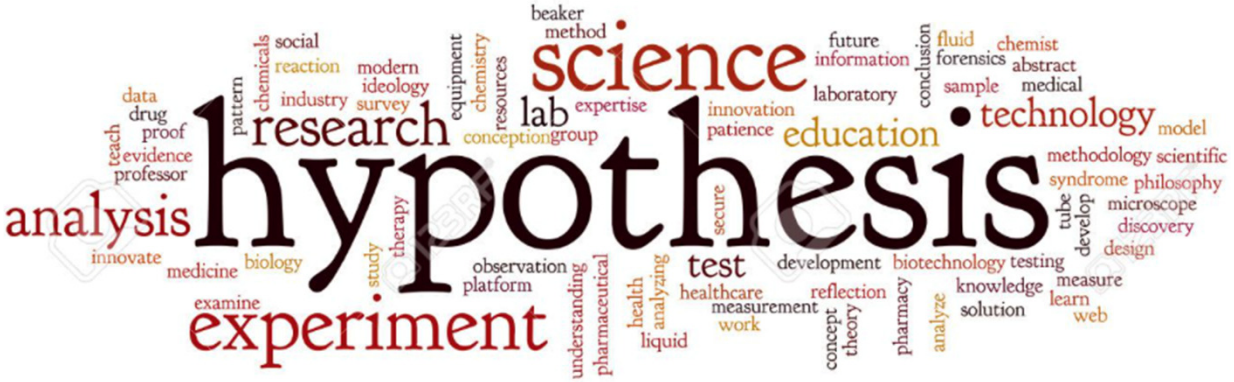
- Assessed 26.6 million builds in C and Java at Google
- Focused on build failures due to compilation problems

- Dependencies are the main source of compilation errors

☑ **Kerzazi et al. ICSME 2014:**

- Analyzed ~3K builds to study build failures

- ~18% of the build failures have a potential cost of ~2K man-hours (1h for each build to succeed)



Build inflation is an actual phenomenon in open-source projects that impacts the interpretation of build results

Research Questions

Does build frequency/failure differ across OS/env.?

What are the impact of OSes on build failure types?

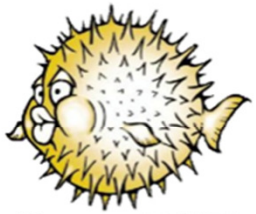
To what extent do environments impact build failures?

To what extent do environments impact build failure types?

What percentage of build failures are OS independent?

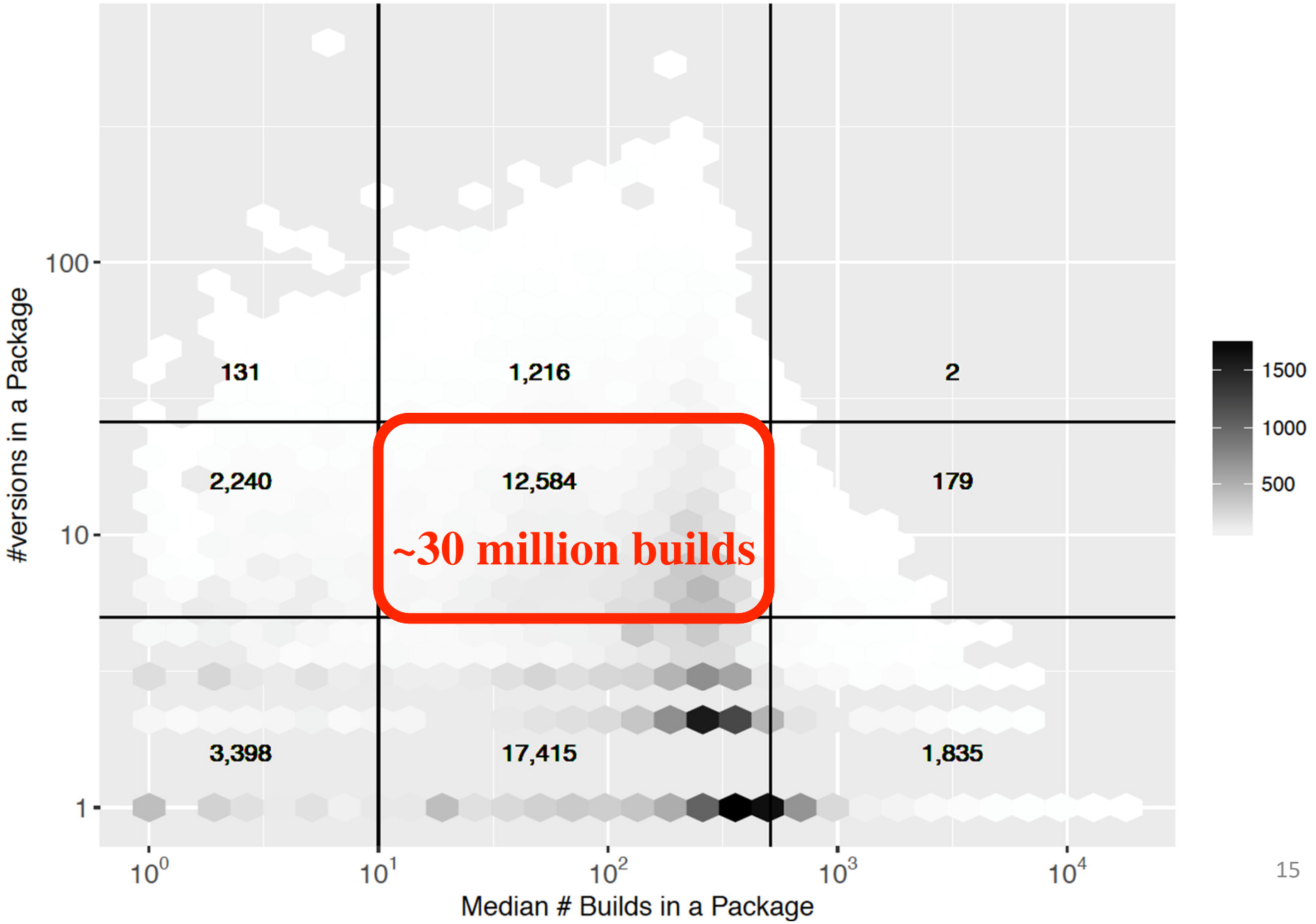
Case Study Setup

- Empirical study of 30M builds
- Study the whole history of CI build packages on CPAN
- 39K packages, 10 Oses, 103 Perl run-time environments



Linux

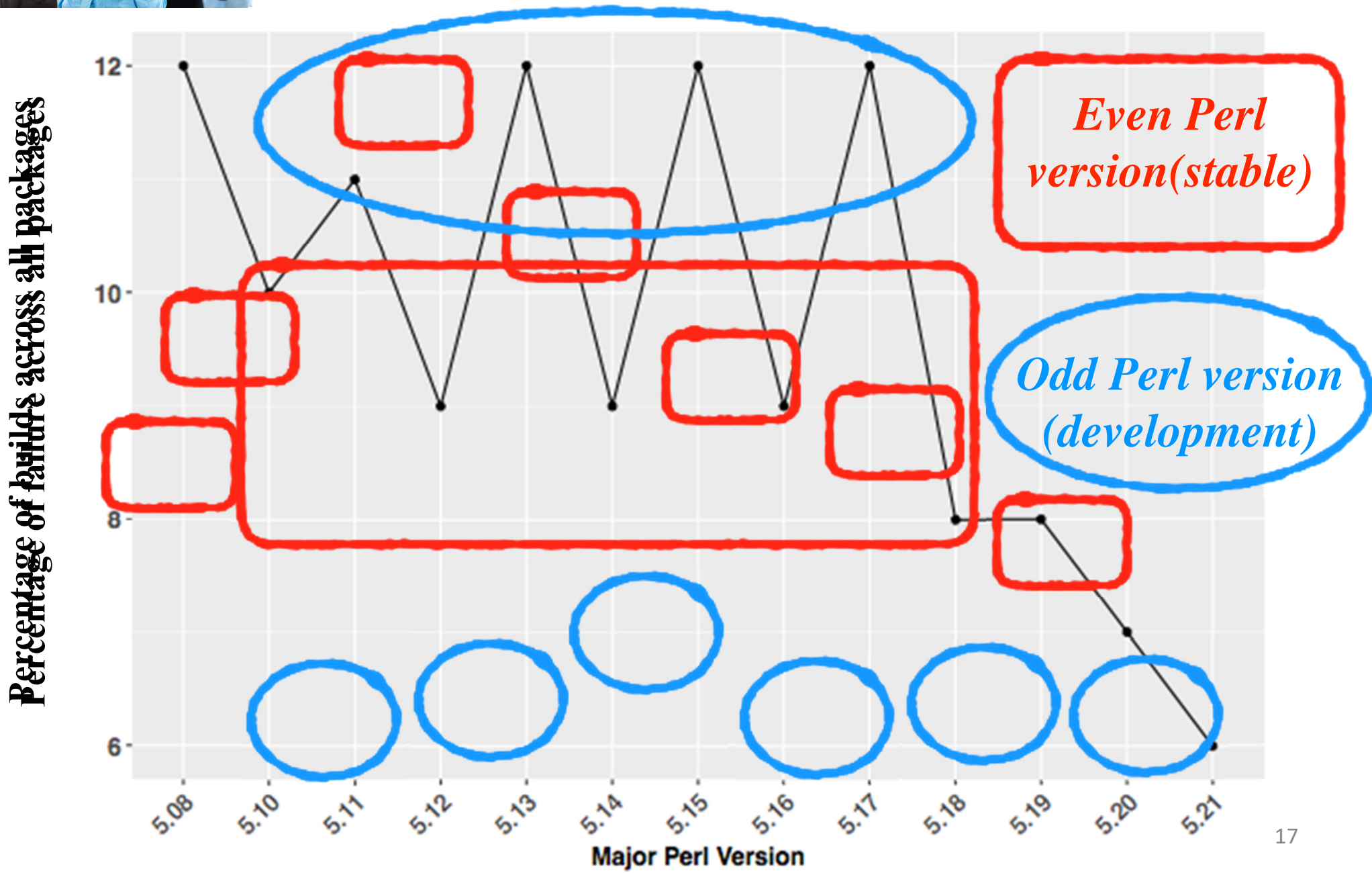
Filtering CPAN Packages Based on Distribution of the Number of Builds and versions



Research Questions

Does build frequency/failure differ across OS/env.?

Not Every Environment Yields Equally Reliable Build Results



Build Results on Some Oses Are Less Reliable Than on Others

Linux 40%

Windows 18% the most error prone OS

Percentage of builds & failures

Percentage of builds

45
40
35
30
25
20
15
10
5
0

Cygwin MidnightBSD Windows solaris Darwin NetBSD OpenBSD FreeBSD Linux

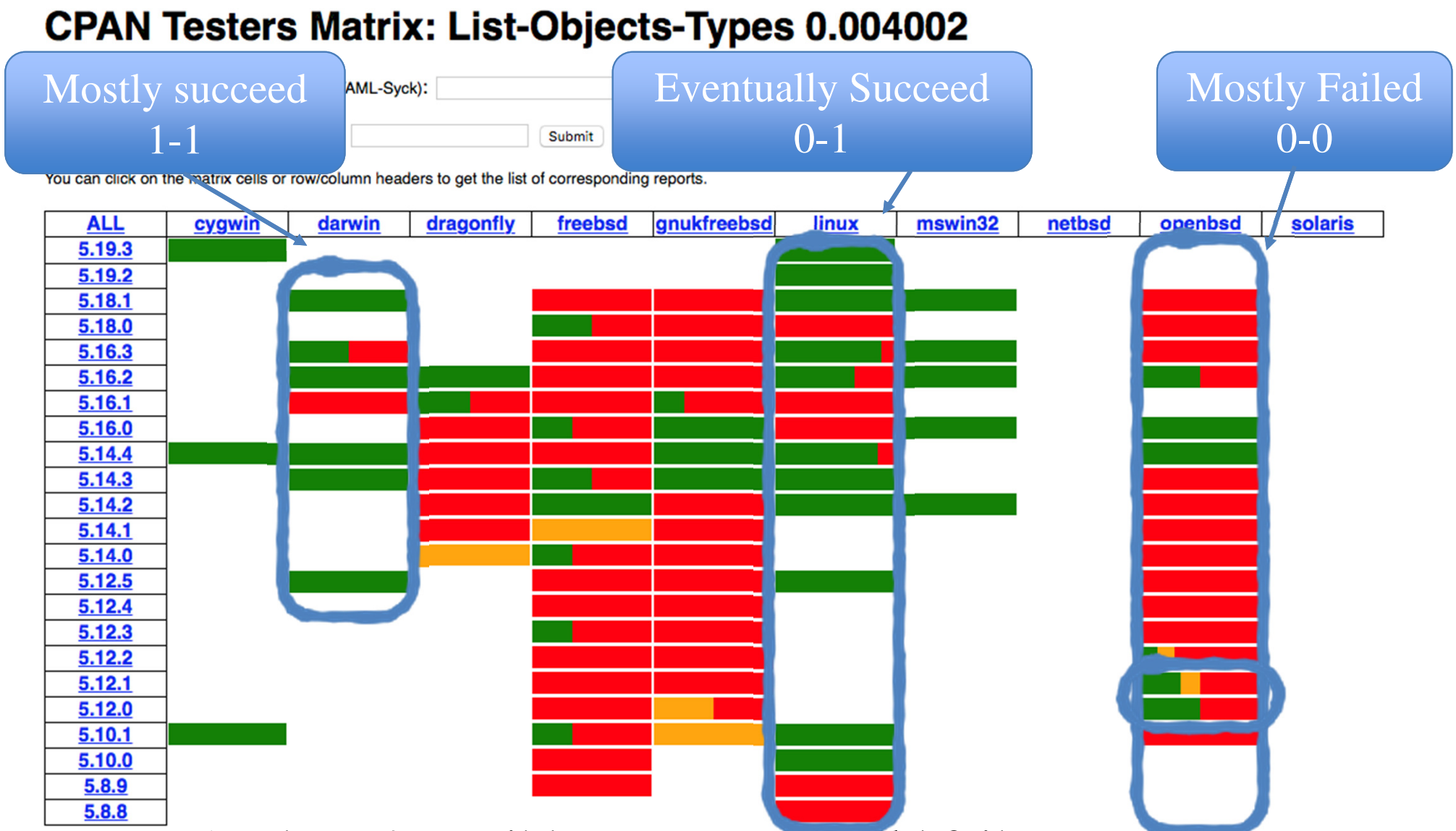
Research Questions

Does build frequency/failure differ across OS/env.?

Different OSes/env. have different frequency & failure-proneness

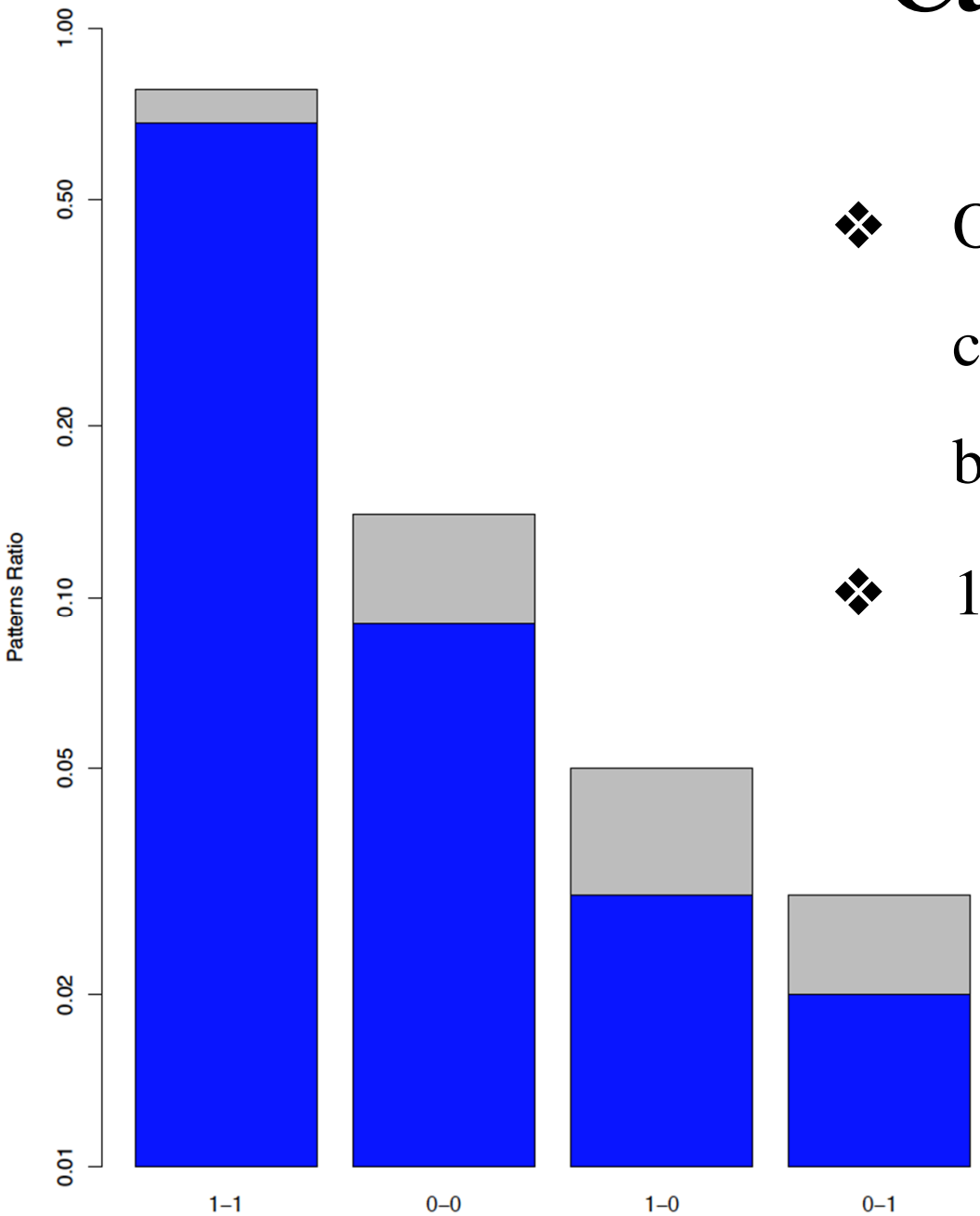
To what extent do environments impact build failures?

How Often Environments Break the Build And to What Degree a Failing Build can Recover



Analyze 4 possible patterns of build failure evolution

To What Degree a Failing Build Can Recover



- Our observation shows that 12% of changes to new environment have been fixed
- 11% were unable to recover

Research Questions

Does build frequency/failure differ across OS/env.?

Different OSes/env. have different frequency & failure-proneness

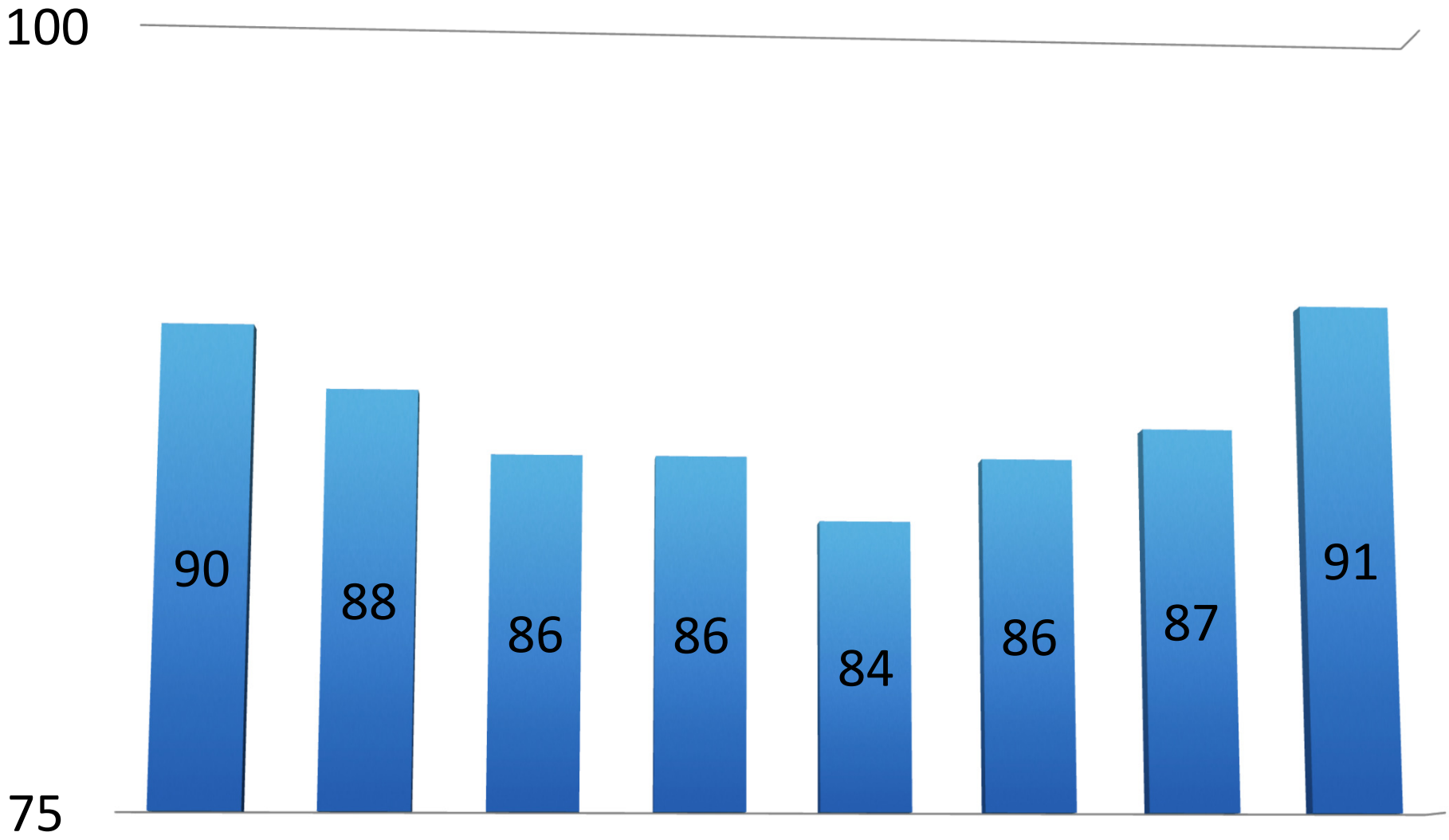
To what extent do environments impact build failures?

77\% and 6\% of the environment vectors consistently succeed or fail

What percentage of build failures are OS independent?

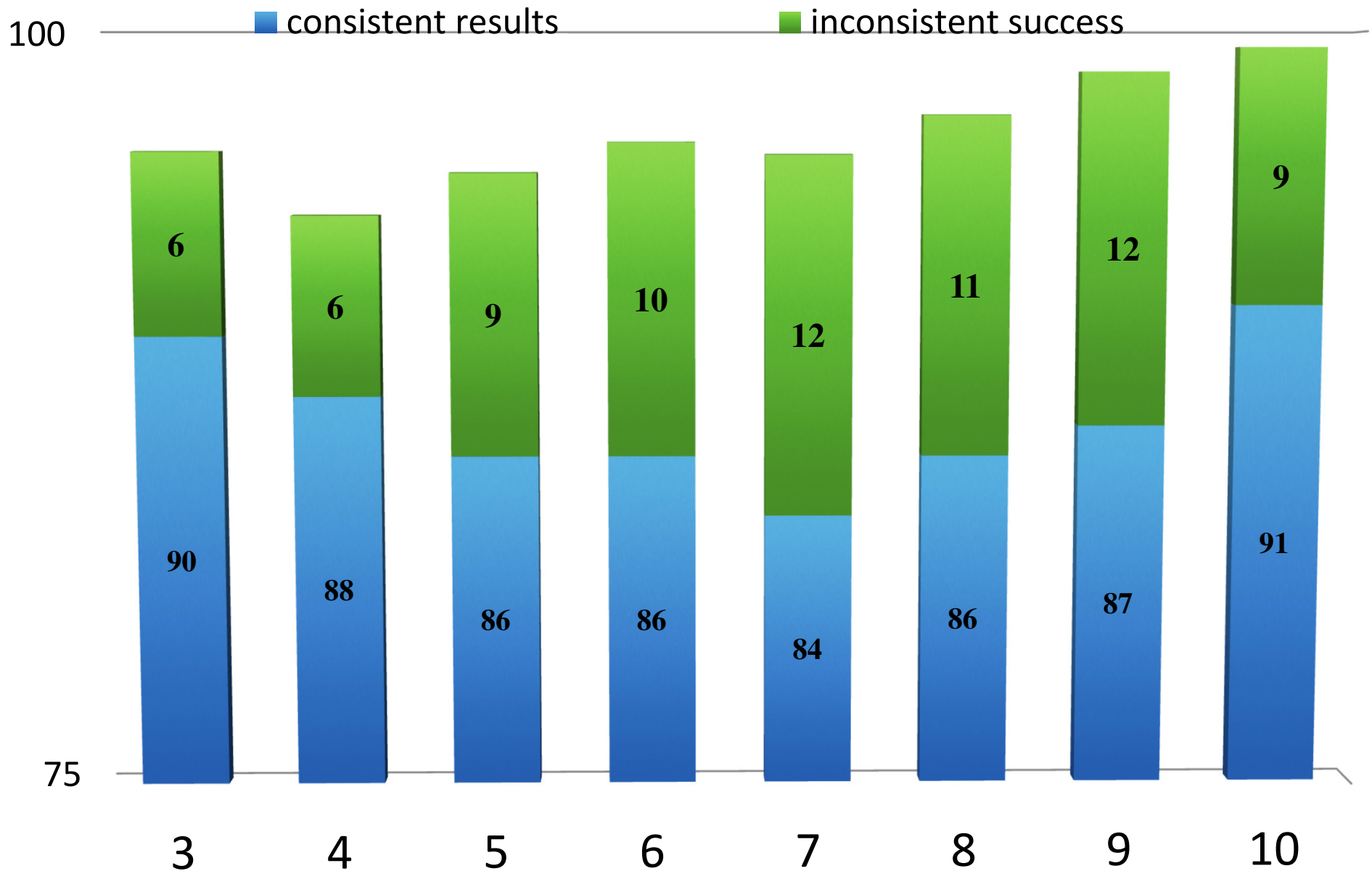
Do Builds Succeed/Fail Consistently across all OSes?...

Median of 86.5% of Builds Succeed/Fail Consistently



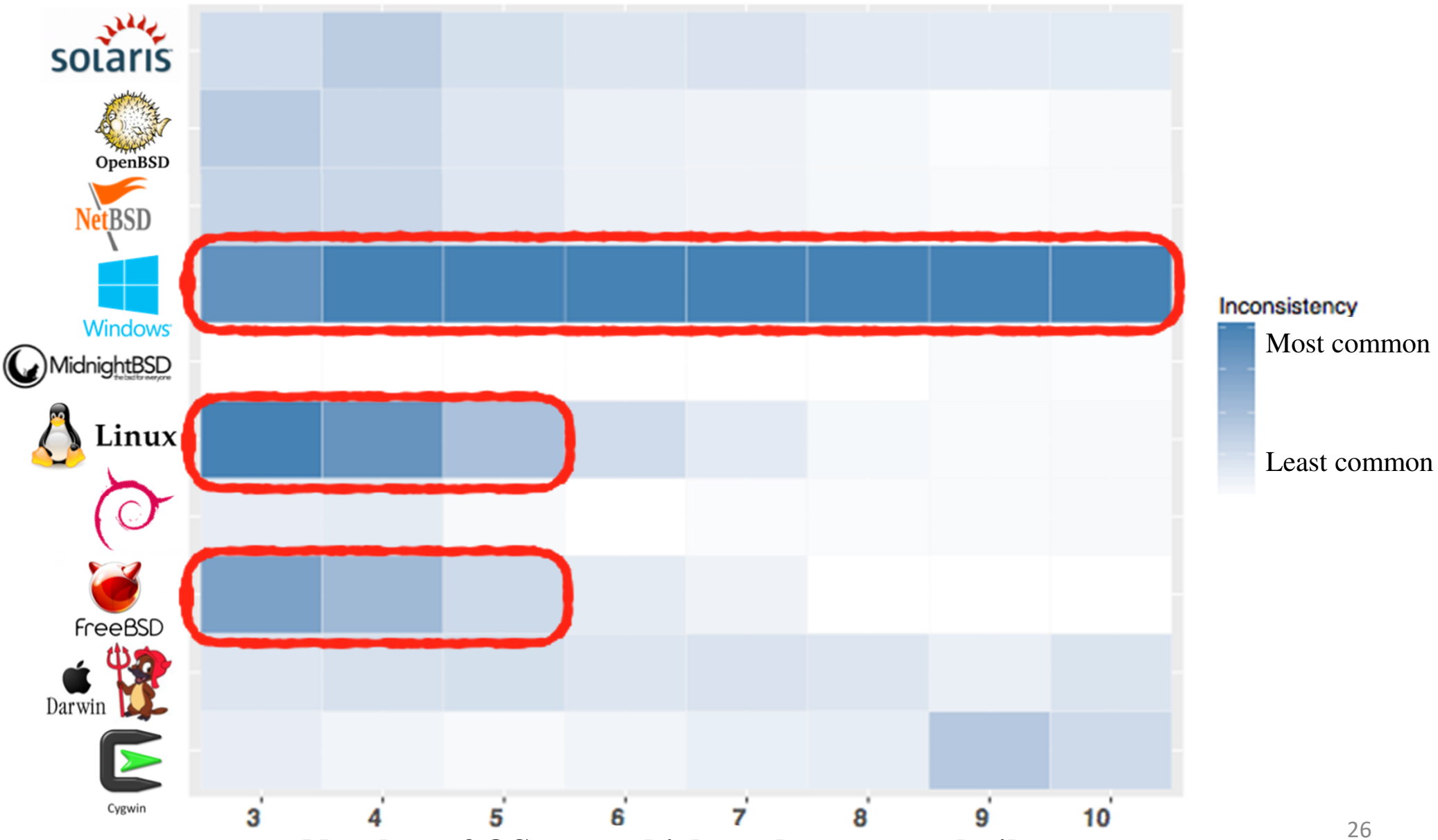
Number of OSes on which packages were built

Median of 7% of Builds Succeed Inconsistently



Number of OSes on which packages were built

Median of 30% of Inconsistent Successes in Windows vs. 7% in Linux



Research Questions

Does build frequency/failure differ across OS/env.?

Results on different OSes/env. have different values

What are the impact of OSes on build failure types?

To what extent do environments impact build failures?

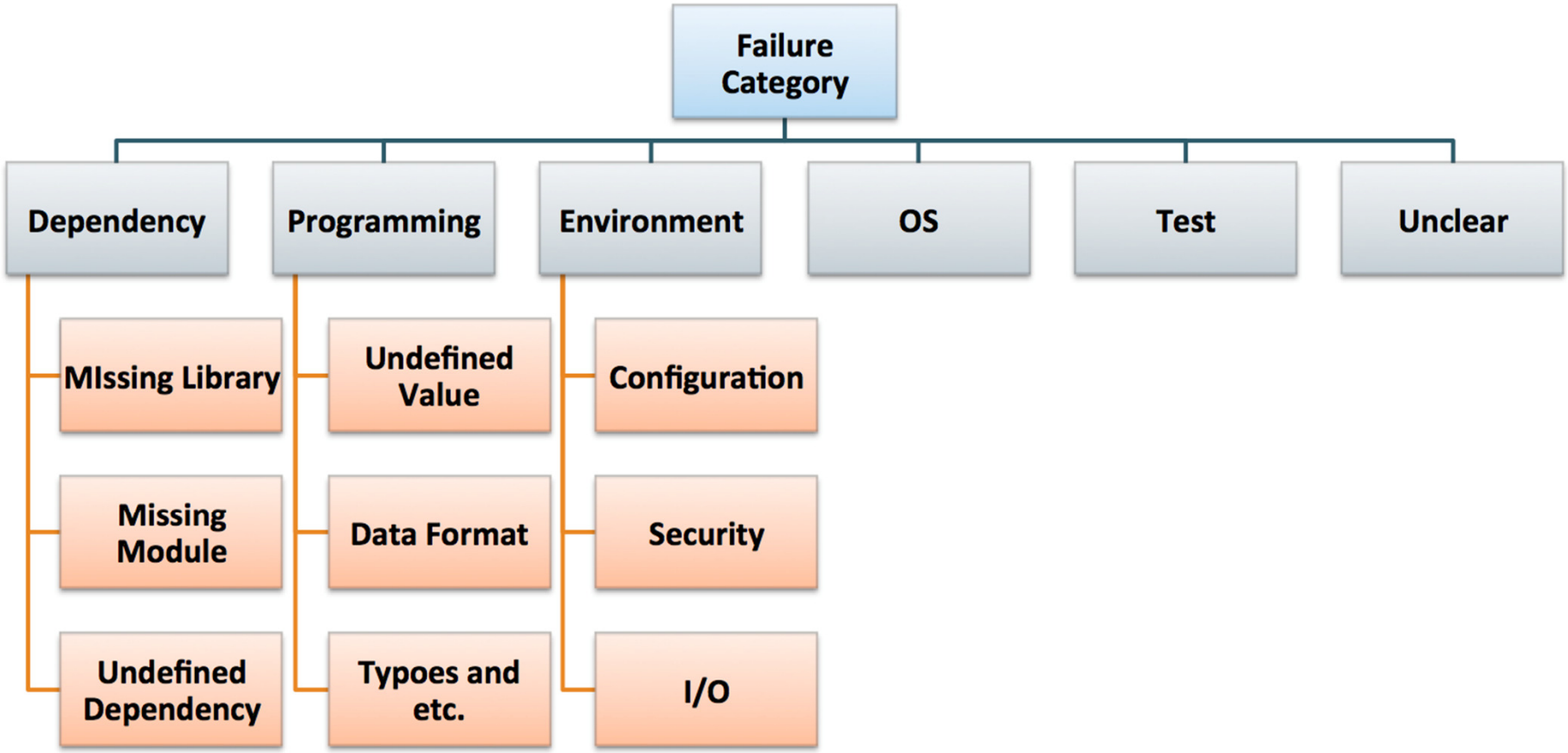
77\% and 6\% of the environment vectors consistently succeed or fail

What percentage of build failures are OS independent?

86.5% of build results were consistent → evidence of build inflation

Comprehending build failures and their relationships with environments/OSes is helpful to solve future failures.

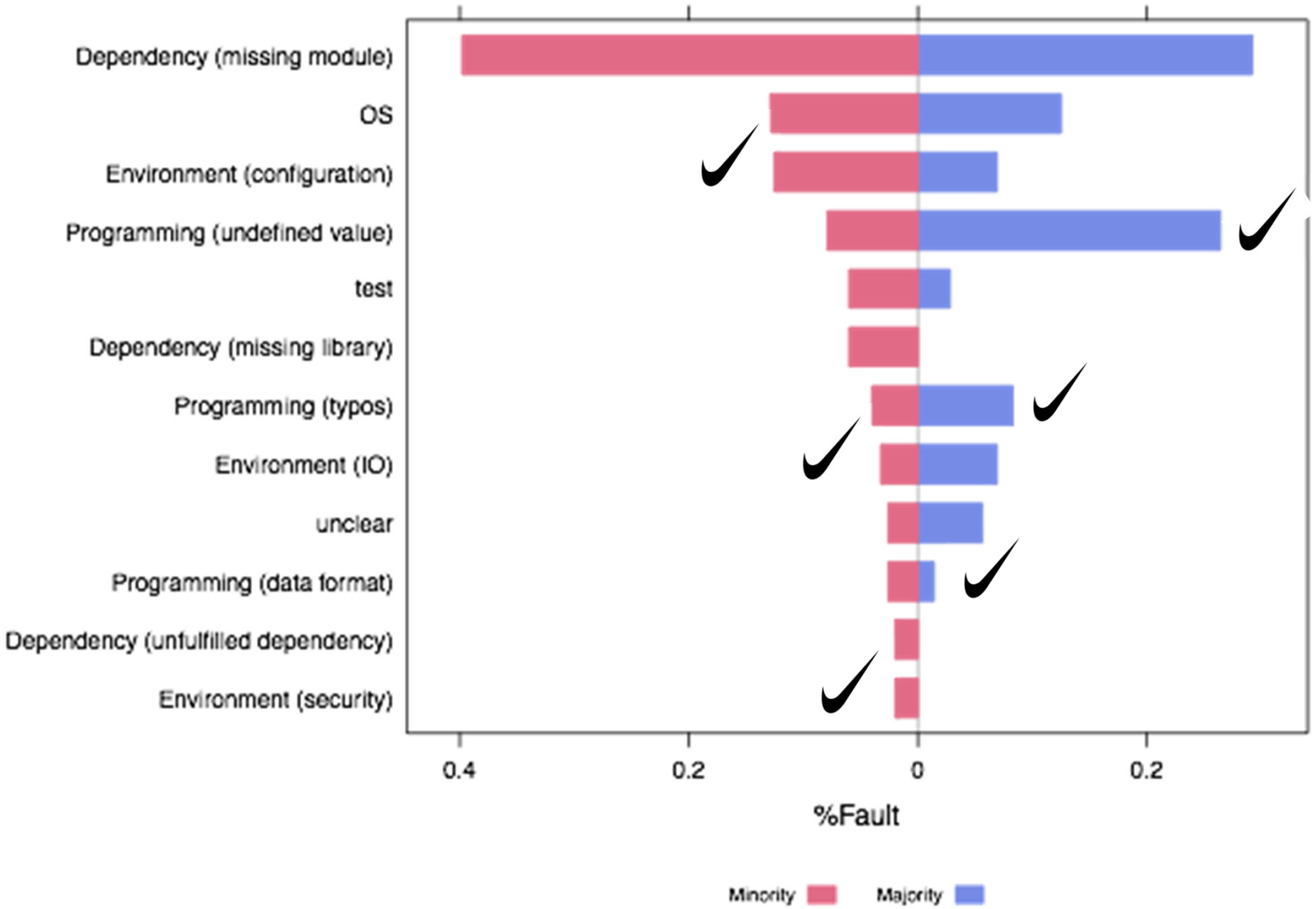
Manually analyzed the build logs and categorized the failures according to the reason of their occurrence.



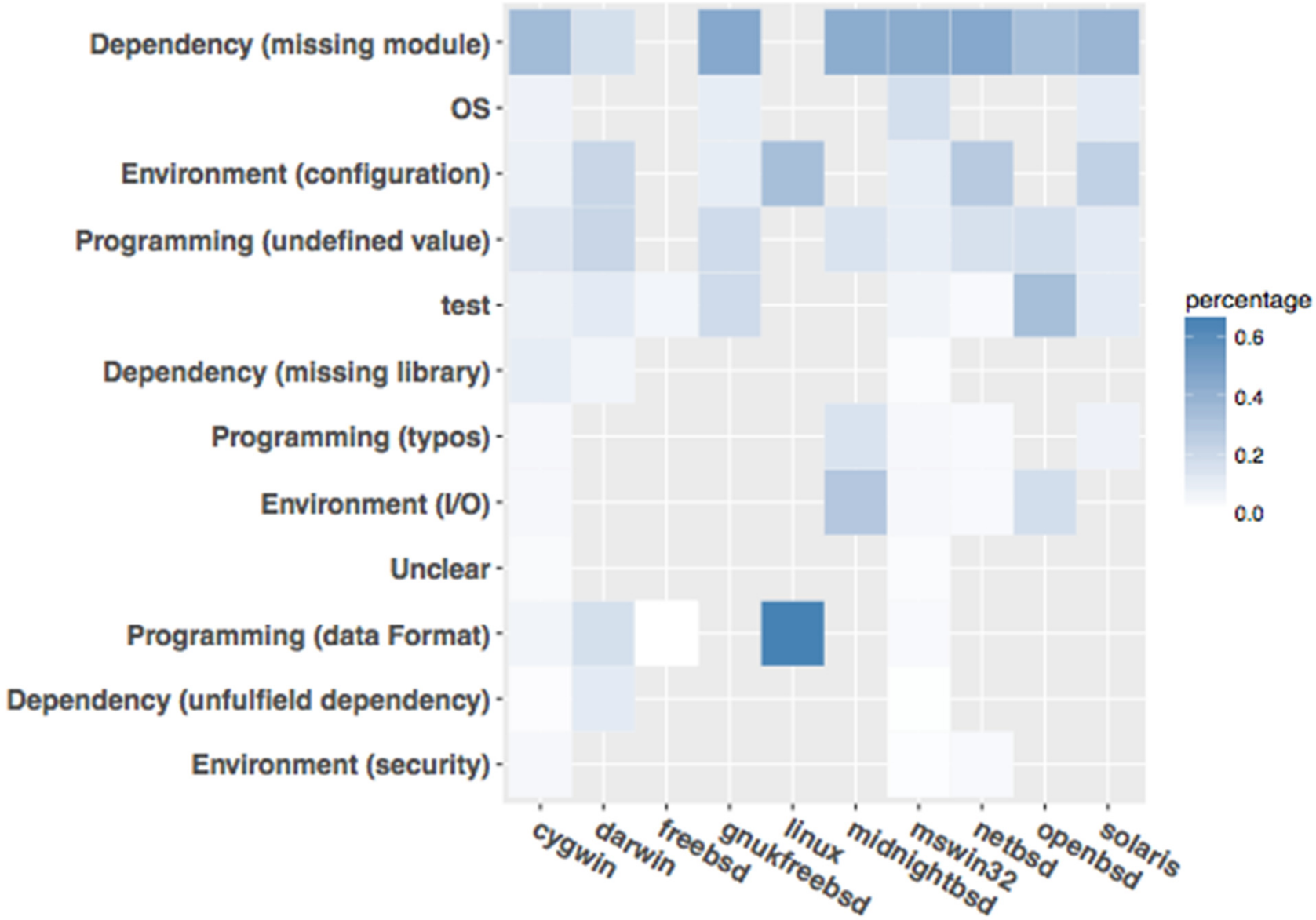
Missing modules is the main reason of failures overall

Majority failures: programming faults

Minority failures: environment faults



Windows and Cygwin are the most fault-prone OSes and experience a wide range of fault types in minority failures
Linux 66.7% is responsible for the "data format" failures and 33.3% for the "configuration"



Research Questions

Does build frequency/failure differ across OS/env.?

Results on different OSes/env. have different values

To what extent do environments impact build failures?

77\% and 6\% of the environment vectors consistently succeed or fail

What are the impact of OSes on build failure types?

Windows and Cygwin are the most fault-prone OSes and experience a wide range of fault types

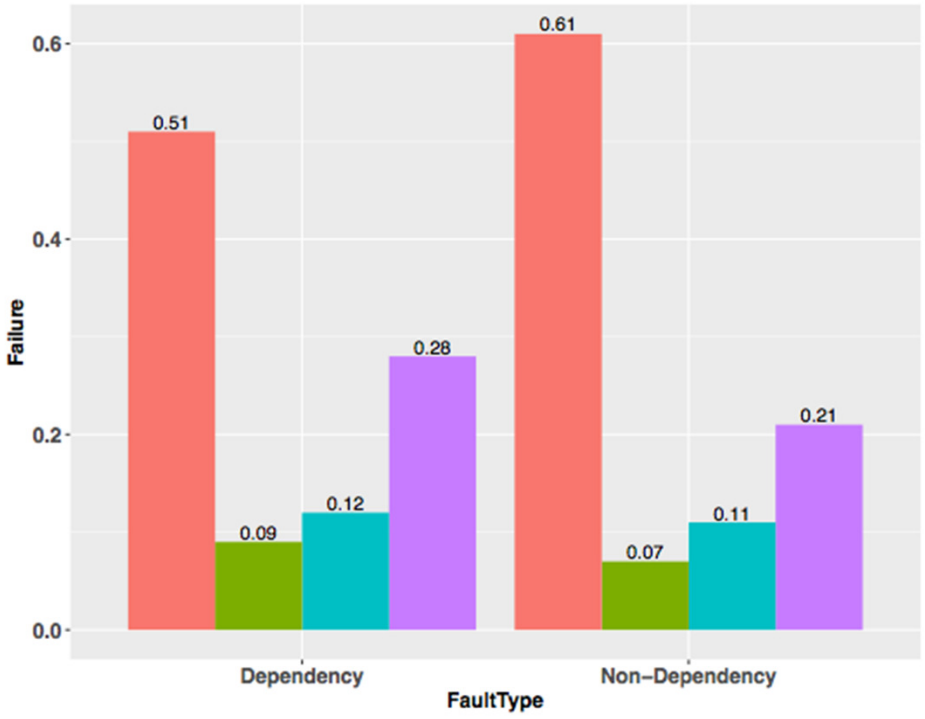
What percentage of build failures are OS independent?

86.5% of build results were consistent → evidence of build inflation

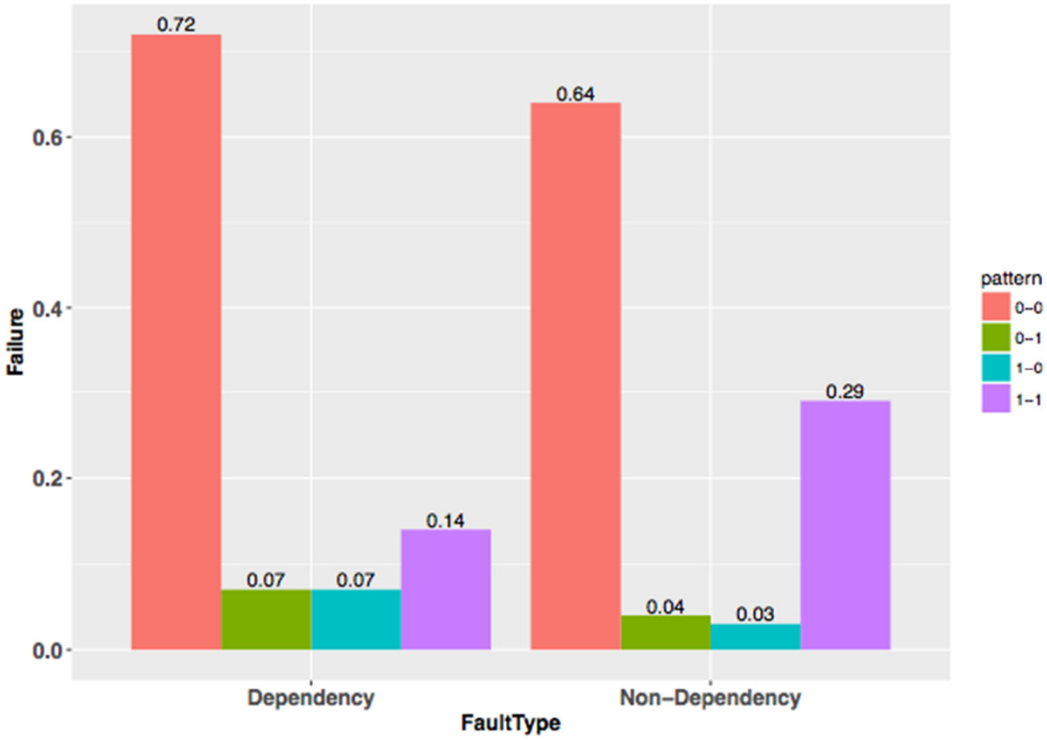
To what extent do environments impact build failure types?

0-0 and 1-1 patterns confirm the occurrence of build inflation

Dependency faults for majority failures are more difficult to resolve than for minority failures.



Minority Failures



Majority Failures

Research Questions

Does build frequency/failure differ across OS/env.?

Results on different OSes/env. have different values

What are the impact of OSes on build failure types?

Windows and Cygwin are the most fault-prone OSes and experience a wide range of fault types

To what extent do environments impact build failures?

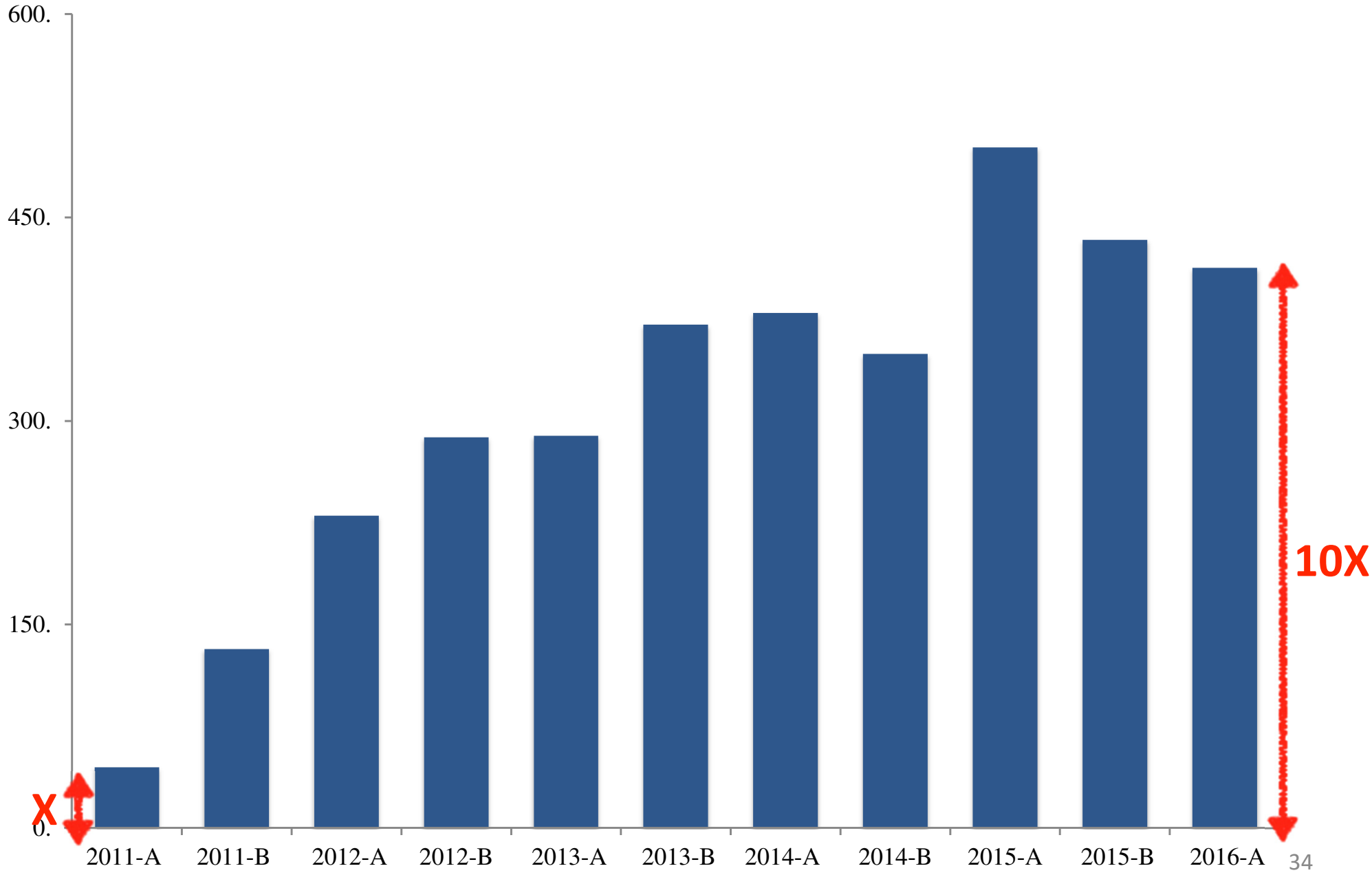
77\% and 6\% of the environment vectors consistently succeed or fail

To what extent do environments impact build failure types?

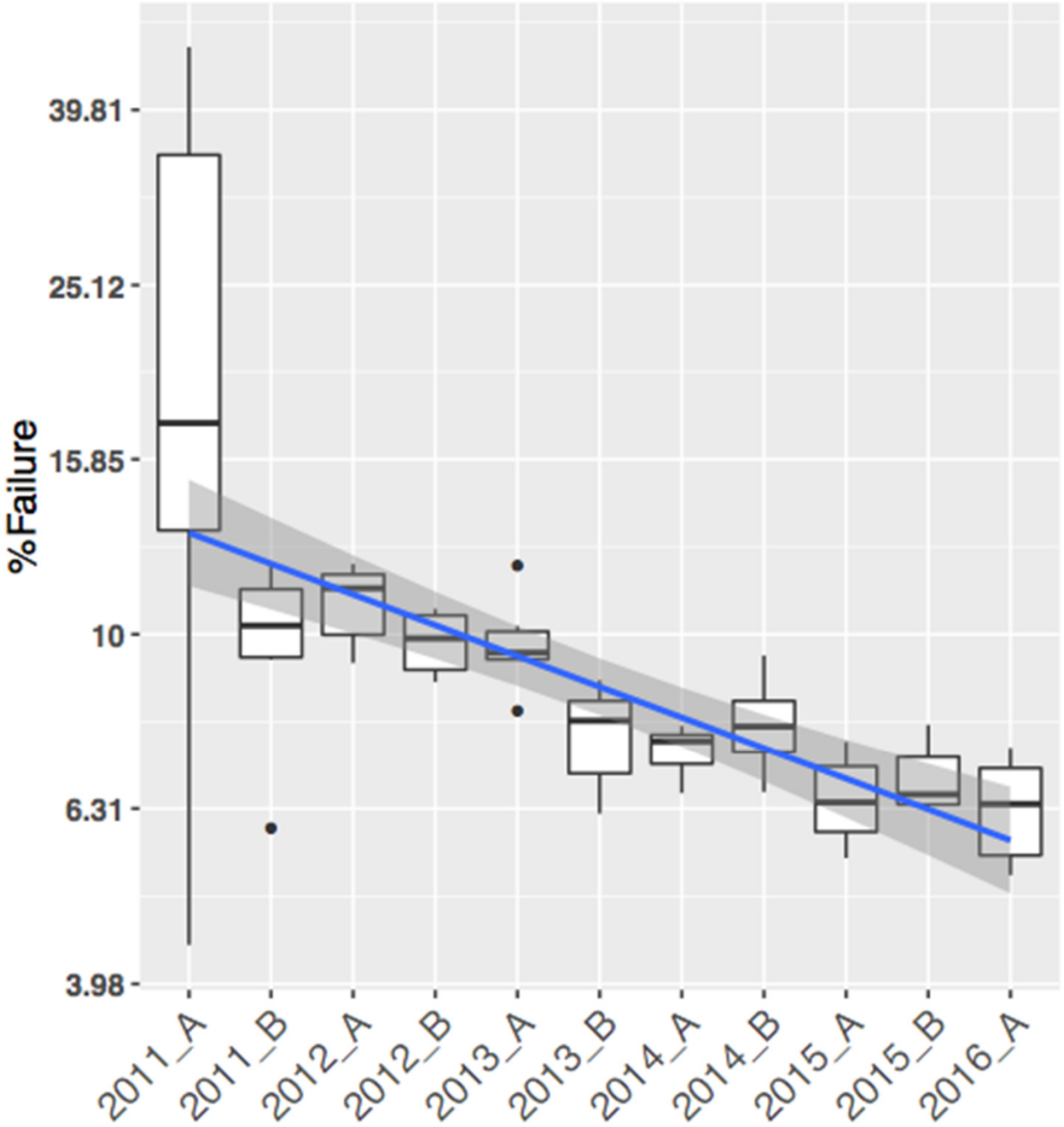
What percentage of build failures are OS independent?

86.5% of build results were consistent → evidence of build inflation

10-fold Increase in Average #builds per Release

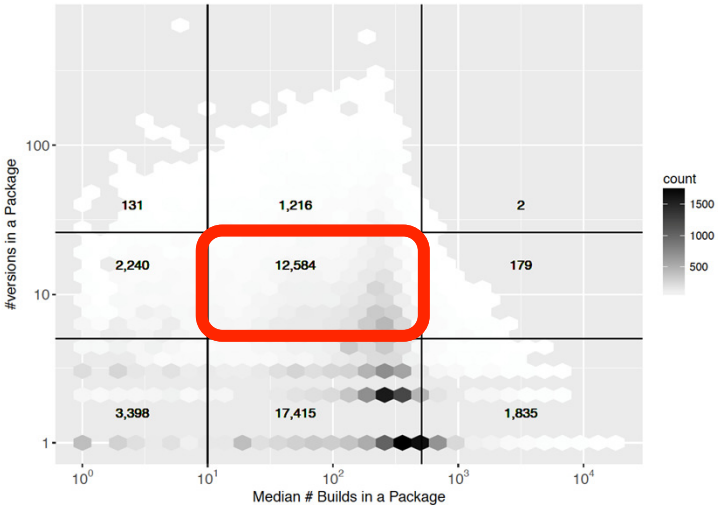


Build Inflation:
More Builds Finding Less Failures



THREATS TO VALIDITY

Only one ecosystem and programming language



Filtered out builds without corresponding failure

OS version/architecture were ignored

Future Work

How to **interpret** build results correctly?

Replicating this study on other ecosystems and programming languages

Ways to prioritize the most **valuable** OS/env. to test on

Publications

☑ **"Do not trust build results at face value: an empirical study of 30 million CPAN builds."** ***Proceedings of the 14th International Conference on Mining Software Repositories*****. IEEE Press, 2017**

☑ **Extension of MSR paper is submitted to the Springer journal on Empirical Software Engineering**