

# Datamining Ubuntu

UDS-R - Copenhagen, Denmark - October 2012

Chris J Arges <[chris.j.arges@canonical.com](mailto:chris.j.arges@canonical.com)>

# Premise

- Ubuntu generates a ton of data.
- Many bugs, only so many people to fix.
- Let's use the data to work more precisely!
- We fix a ton of bugs and we are improving!

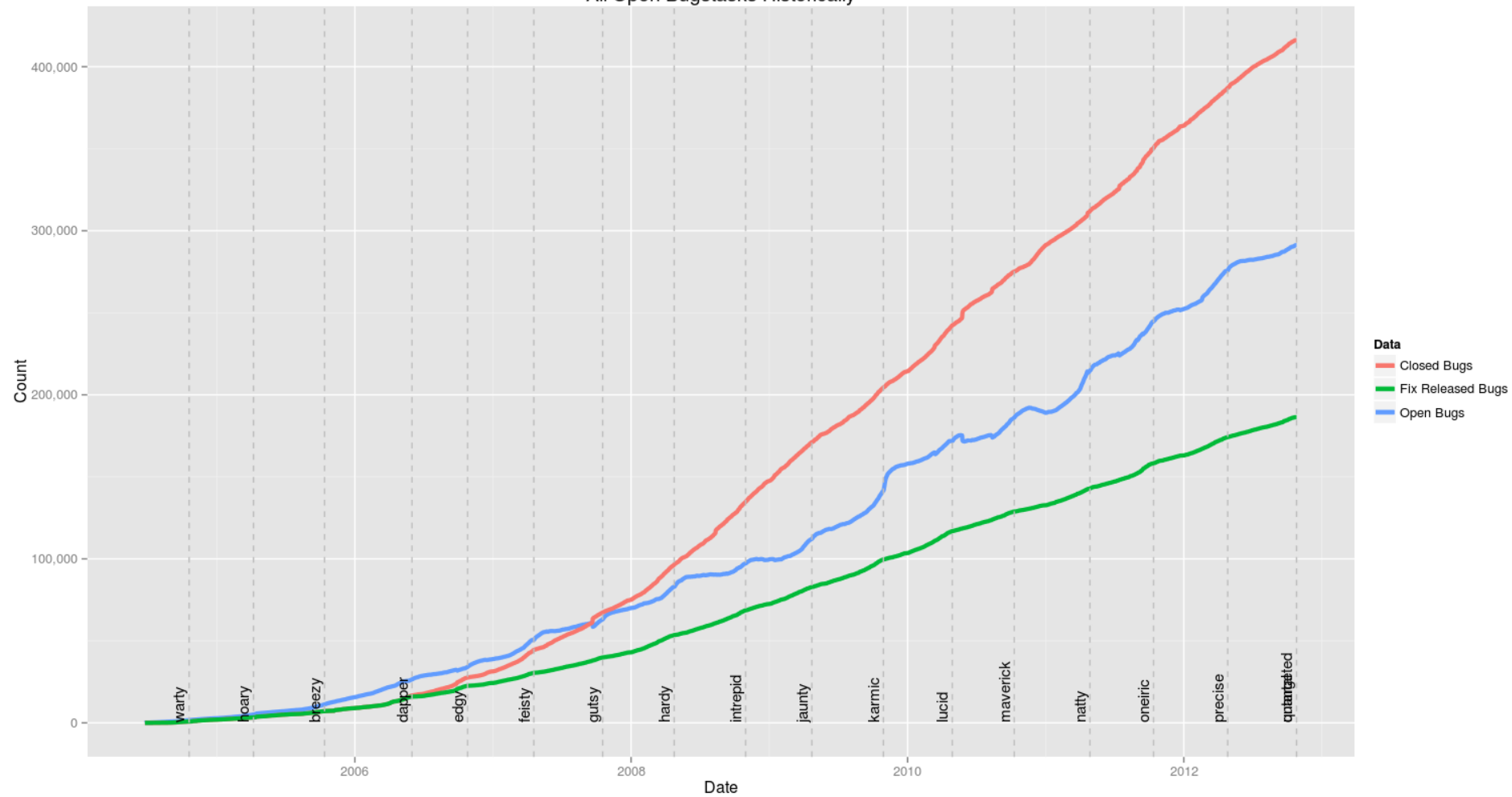
# What was Done?

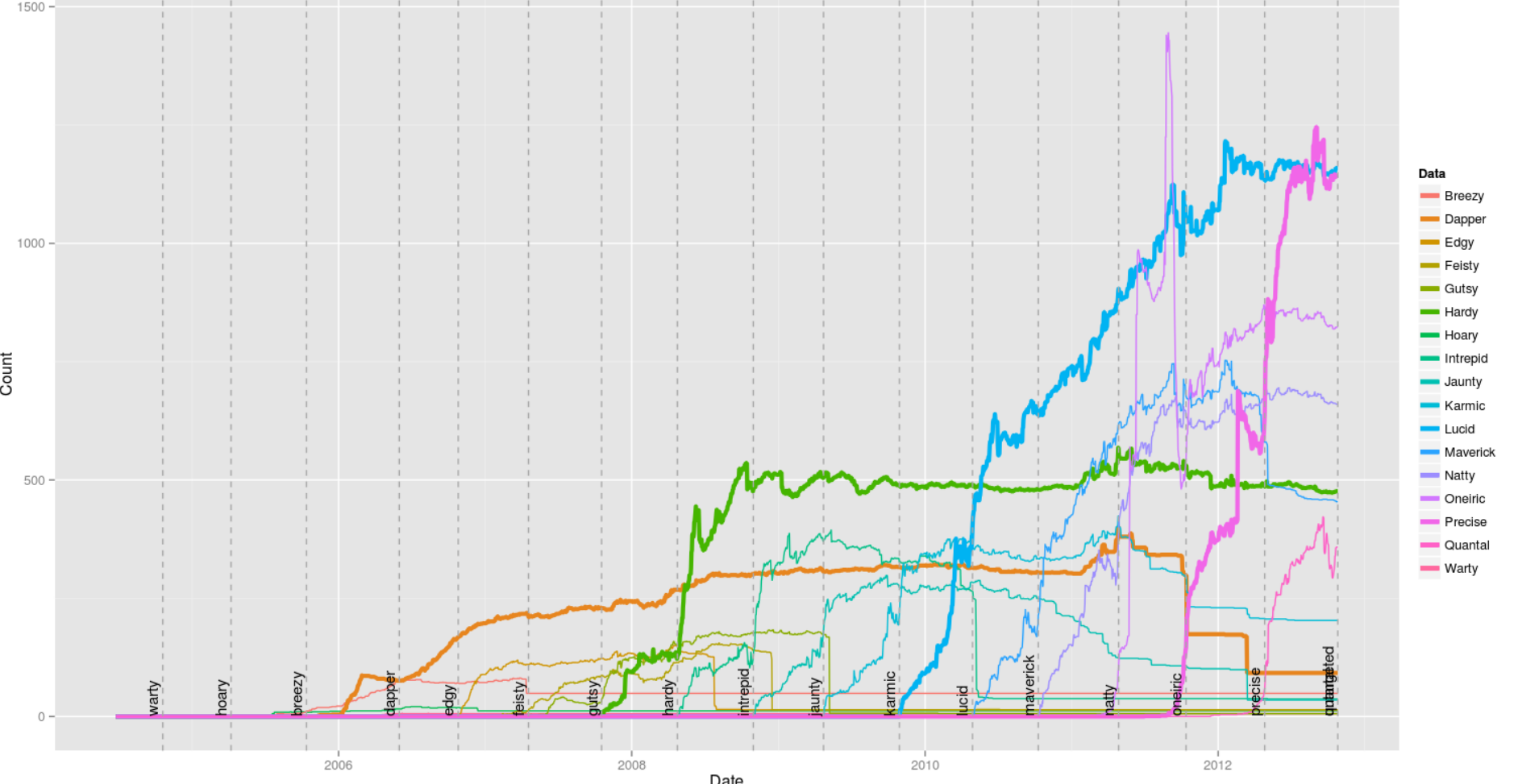
- Download from Launchpad into a database.
- Used 'R' language to take this data and produce graphs and explore entire dataset.
- Used data-mining packages in R to predict if a bug will be Invalid, and whom should work on the bug.

# Data

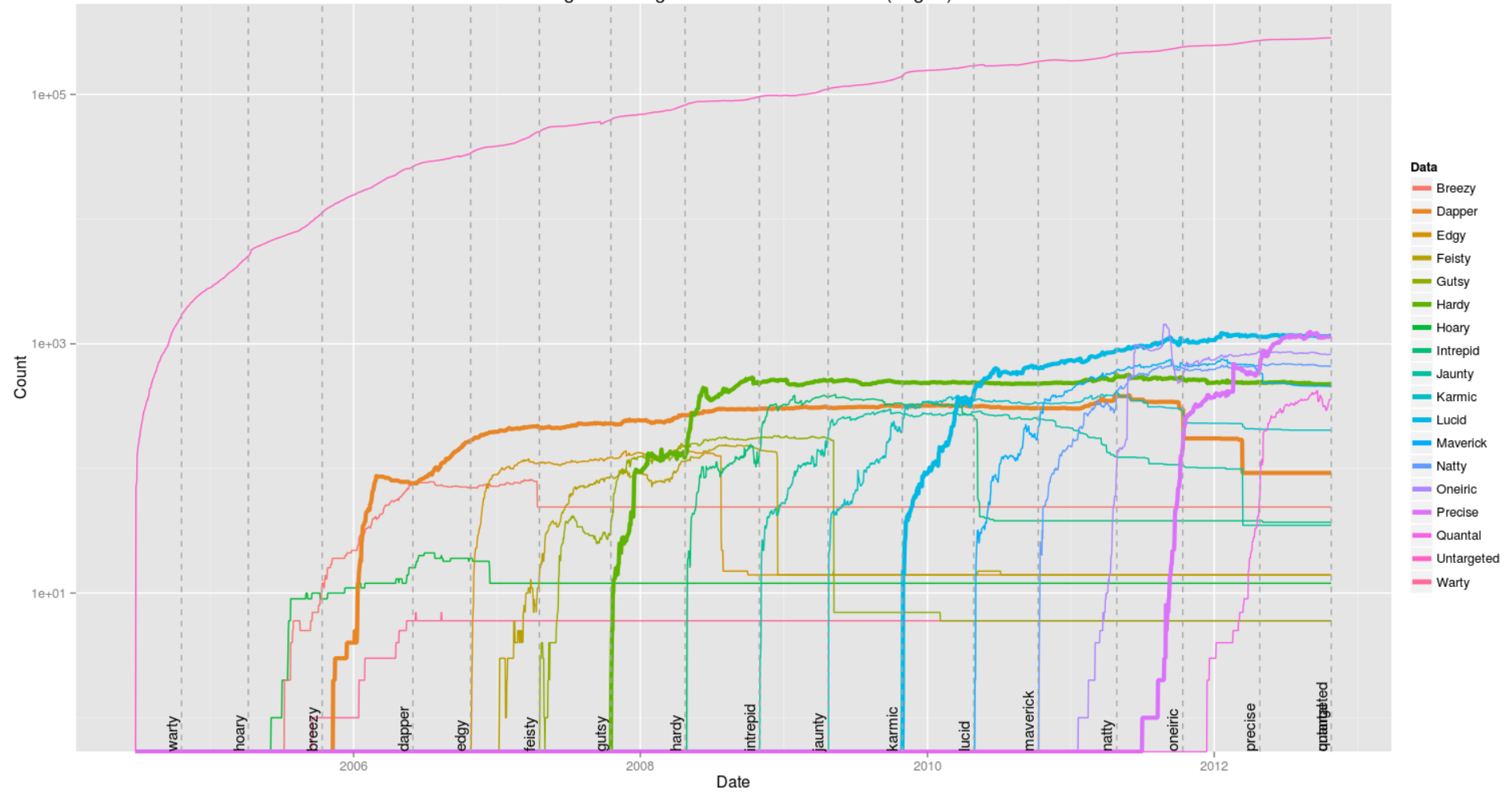
- 632,590 bugs
- 707,343 bugtasks
- From June 2004 to October 2012

All Open Bugtasks Historically

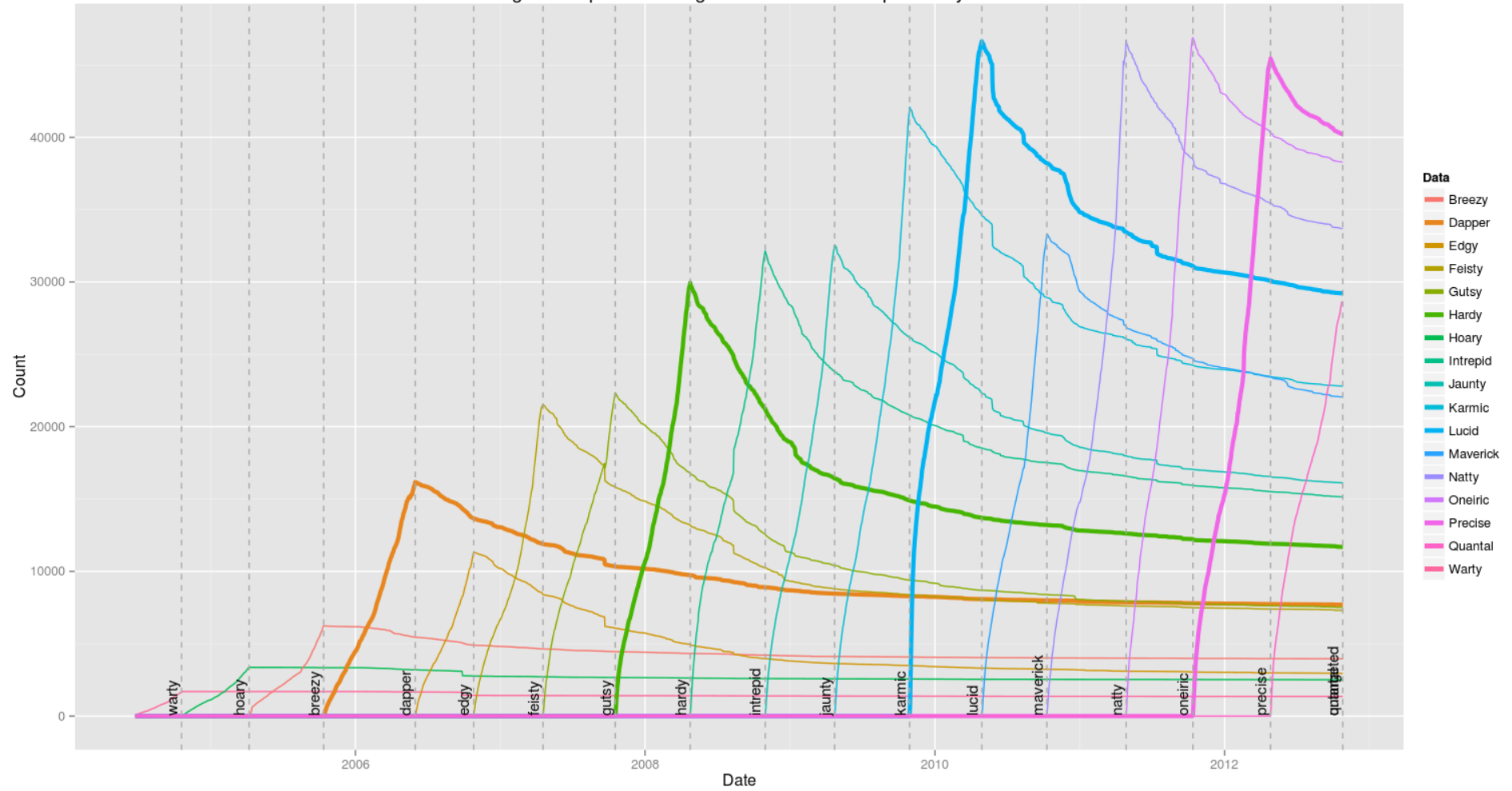




All Bugtasks Targeted to an Ubuntu Series (Log10)

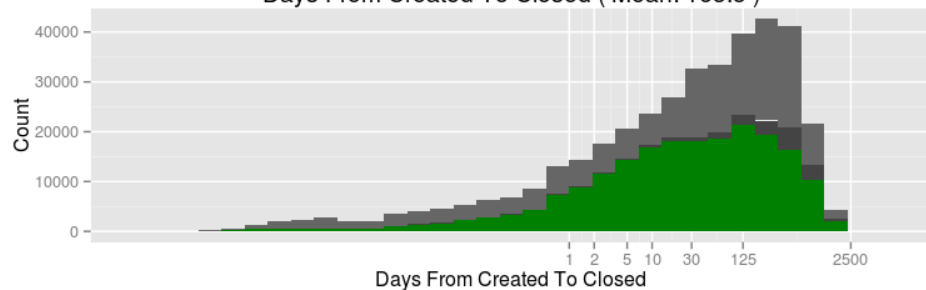


Bugtasks Opened During Each Series Development Cycle

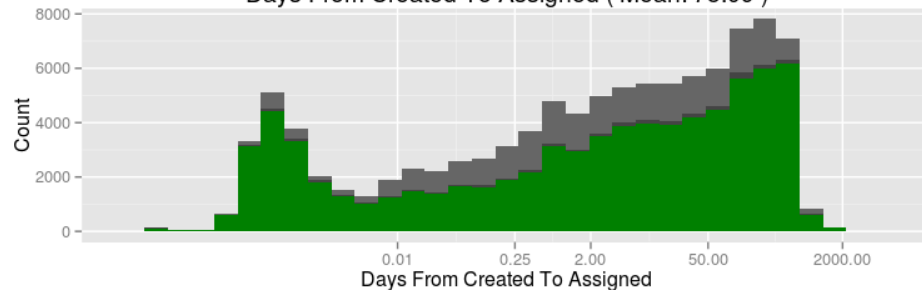




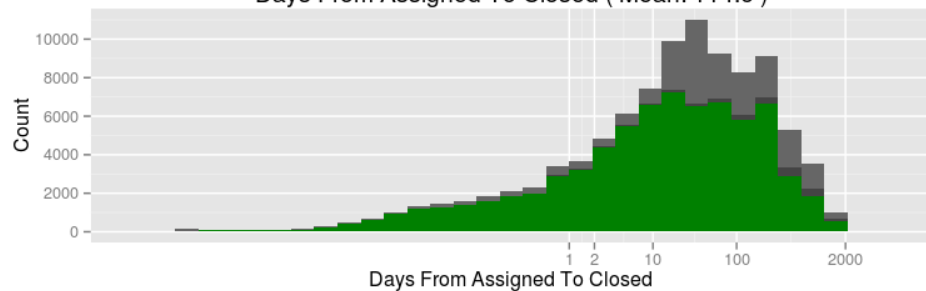
Days From Created To Closed ( Mean: 163.5 )



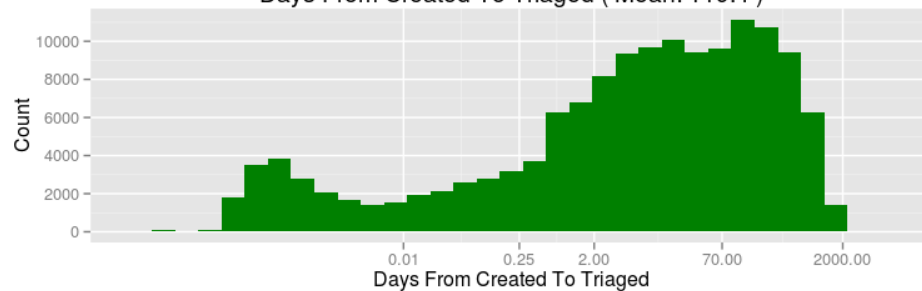
Days From Created To Assigned ( Mean: 75.09 )



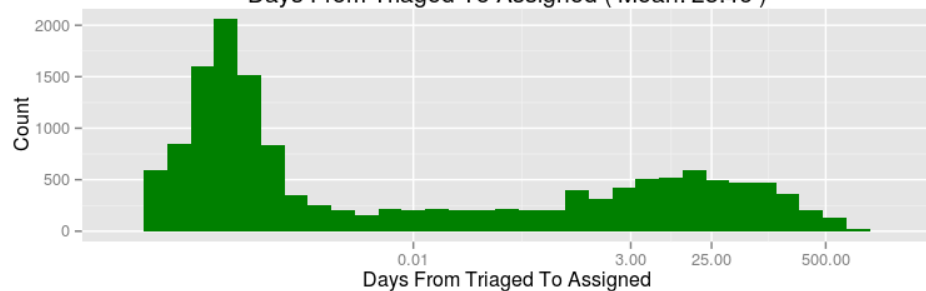
Days From Assigned To Closed ( Mean: 114.6 )



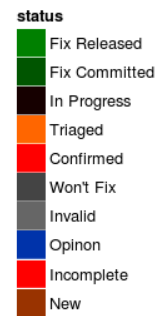
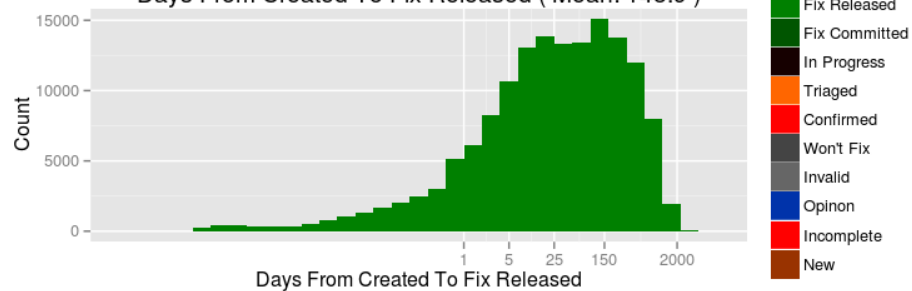
Days From Created To Triaged ( Mean: 119.4 )



Days From Triaged To Assigned ( Mean: 23.46 )



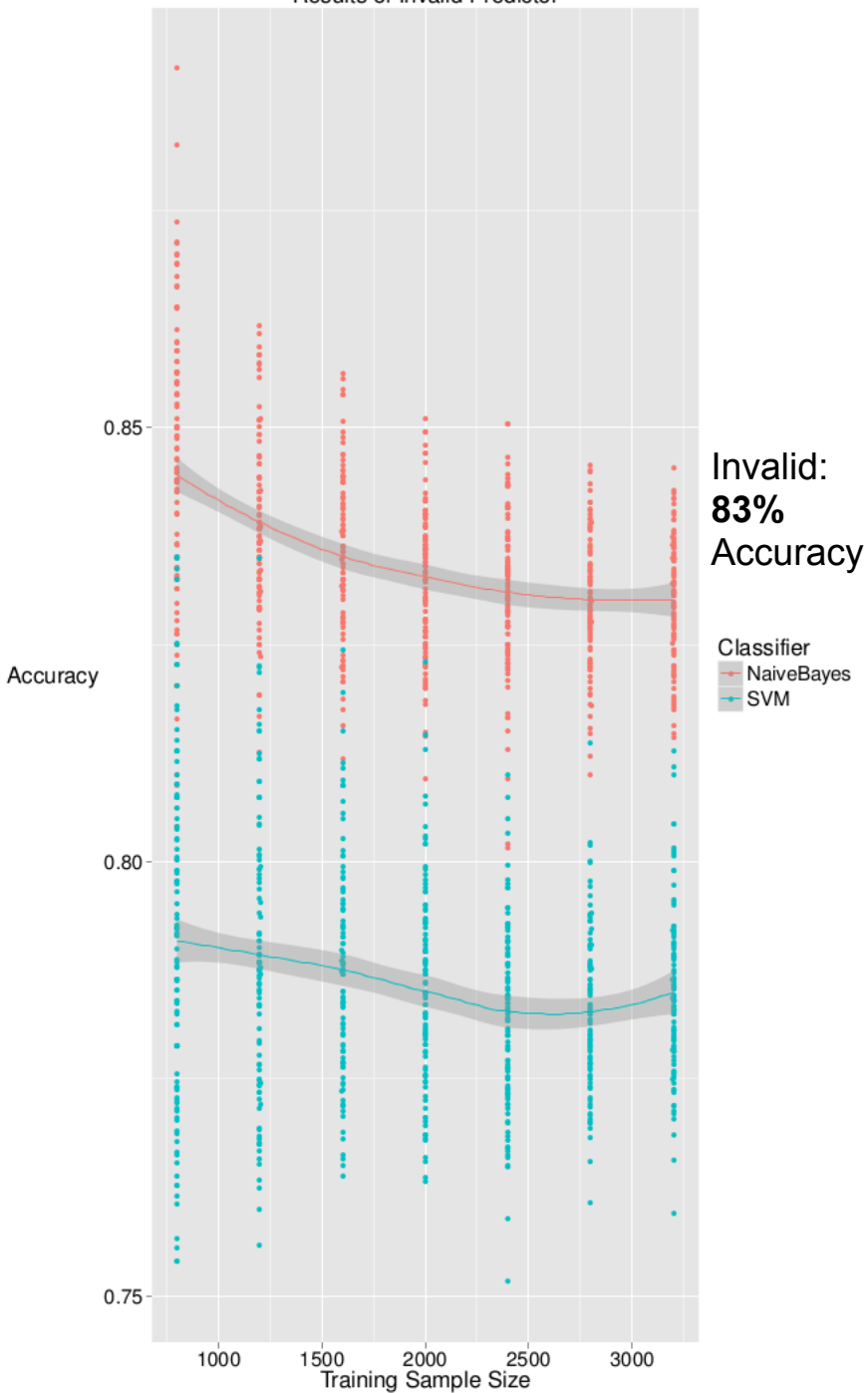
Days From Created To Fix Released ( Mean: 148.9 )



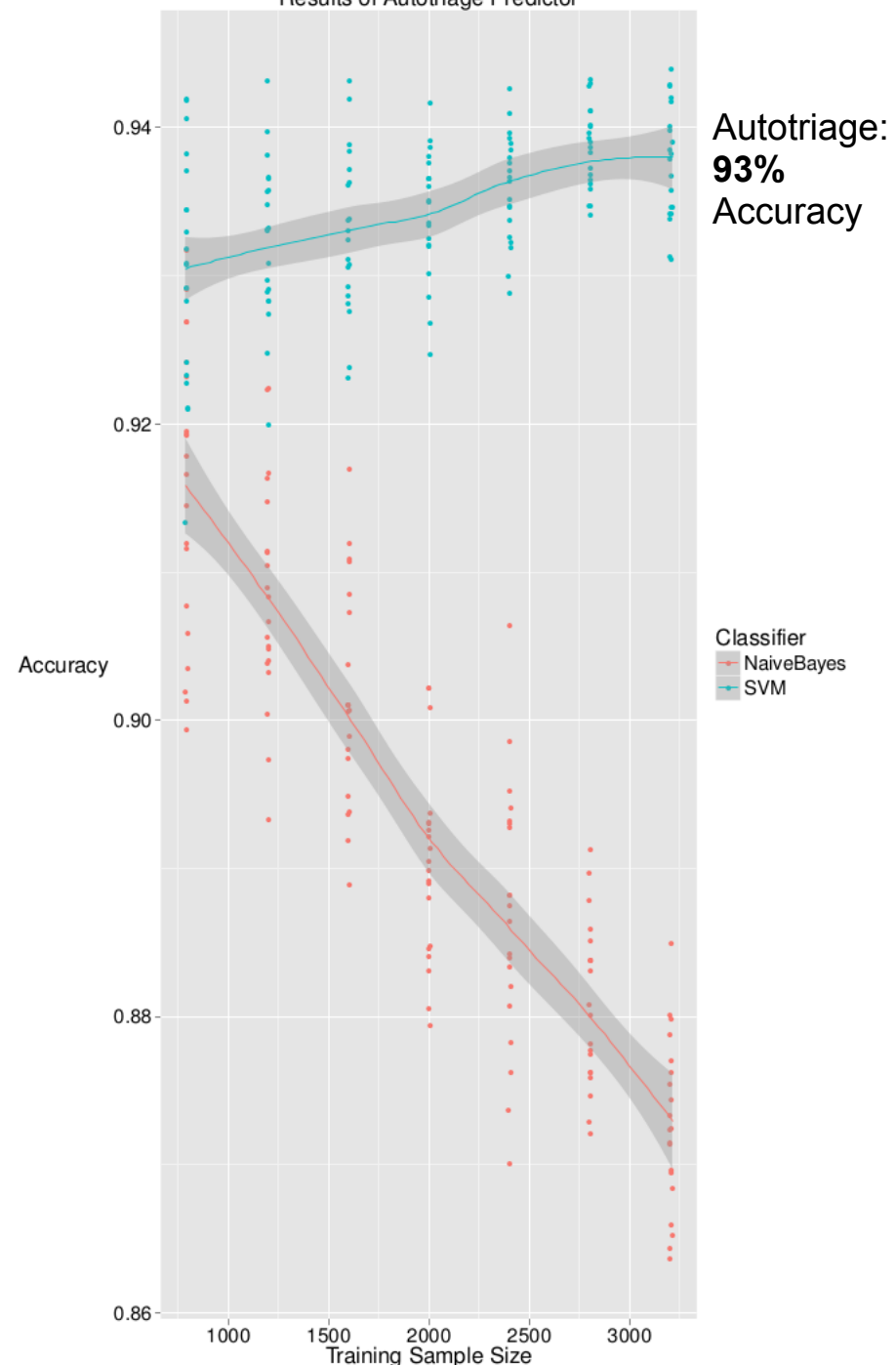
# Hypothesis / Experiment

- We have a ton of Invalid bugs!
  - Predict invalid bugs before we work on them.
- We take quite a bit of time to figure out who should fix a bug.
  - Automatically suggest/assign a person.

Results of Invalid Predictor



Results of Autotriage Predictor



**THANKS**

**[chrisarges.net/papers.html](http://chrisarges.net/papers.html)**