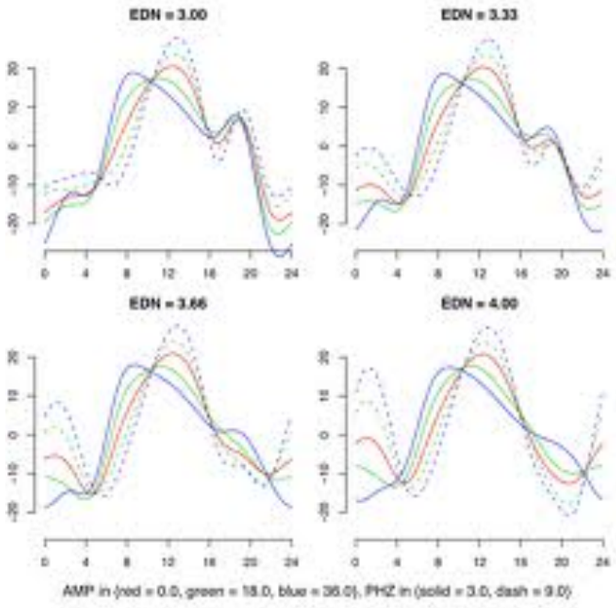
Empirical Software Engineering using Ultra Large Repositories

Mei Nagappan

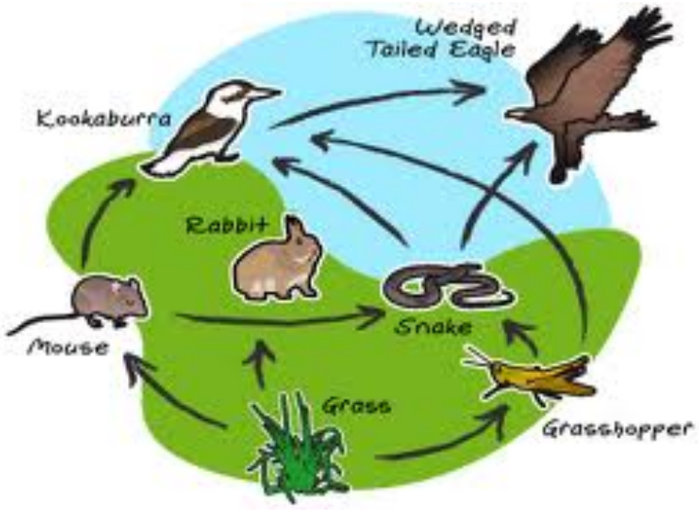
Agenda

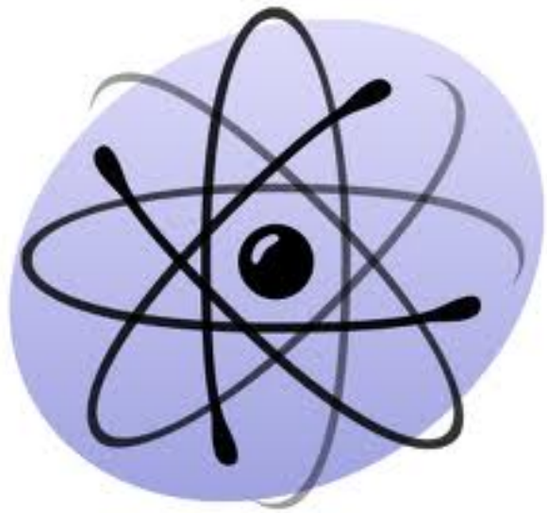
- Part 1 - Introduction
    - Course Overview Objectives
    - Student introduction and expectations
    - Syllabus and Project
- Part 2 - Example of an ESEULR study
- Part 3 - Intro to stats.

Typical ESE vs ESE in ULR



What can we learn about SE from these Ultra Large Repositories?

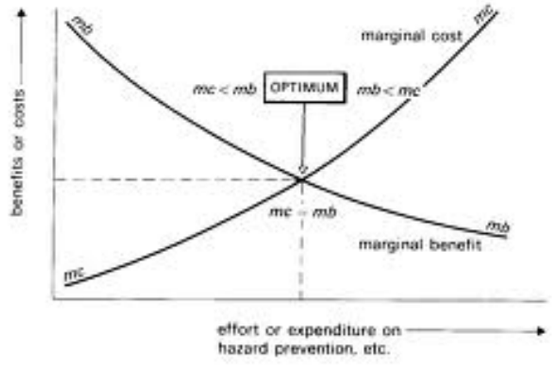




Challenges

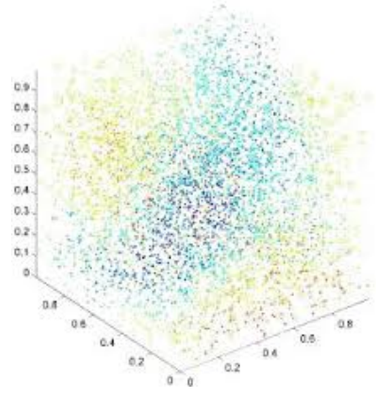
Mining

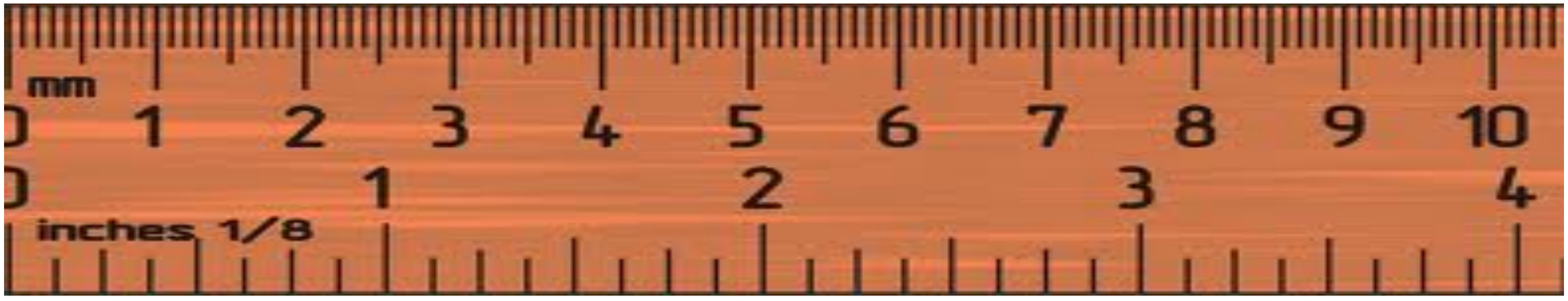
Sample Selection



Analysis

Noise





Part 2

Example Study

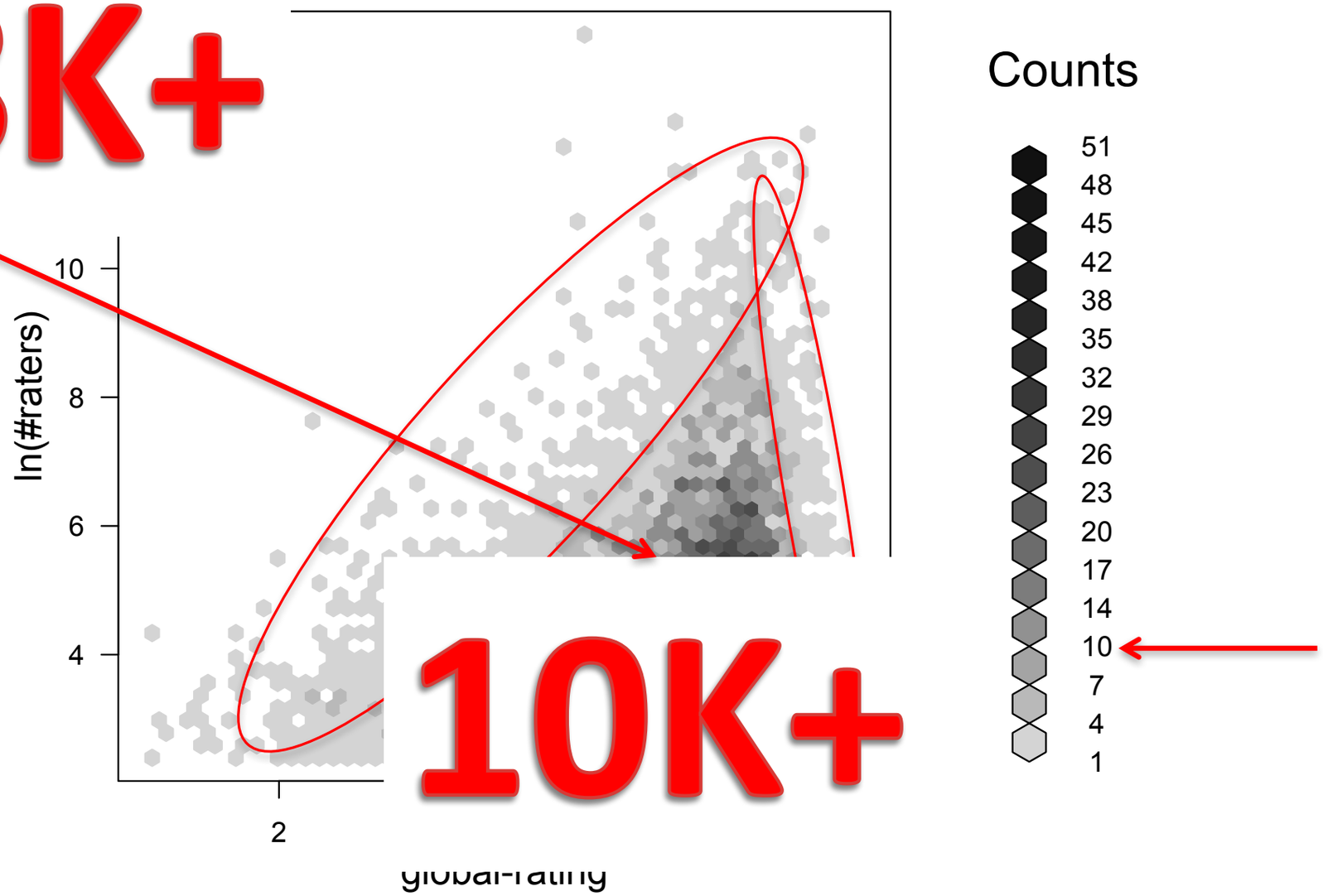
Google play

128K+

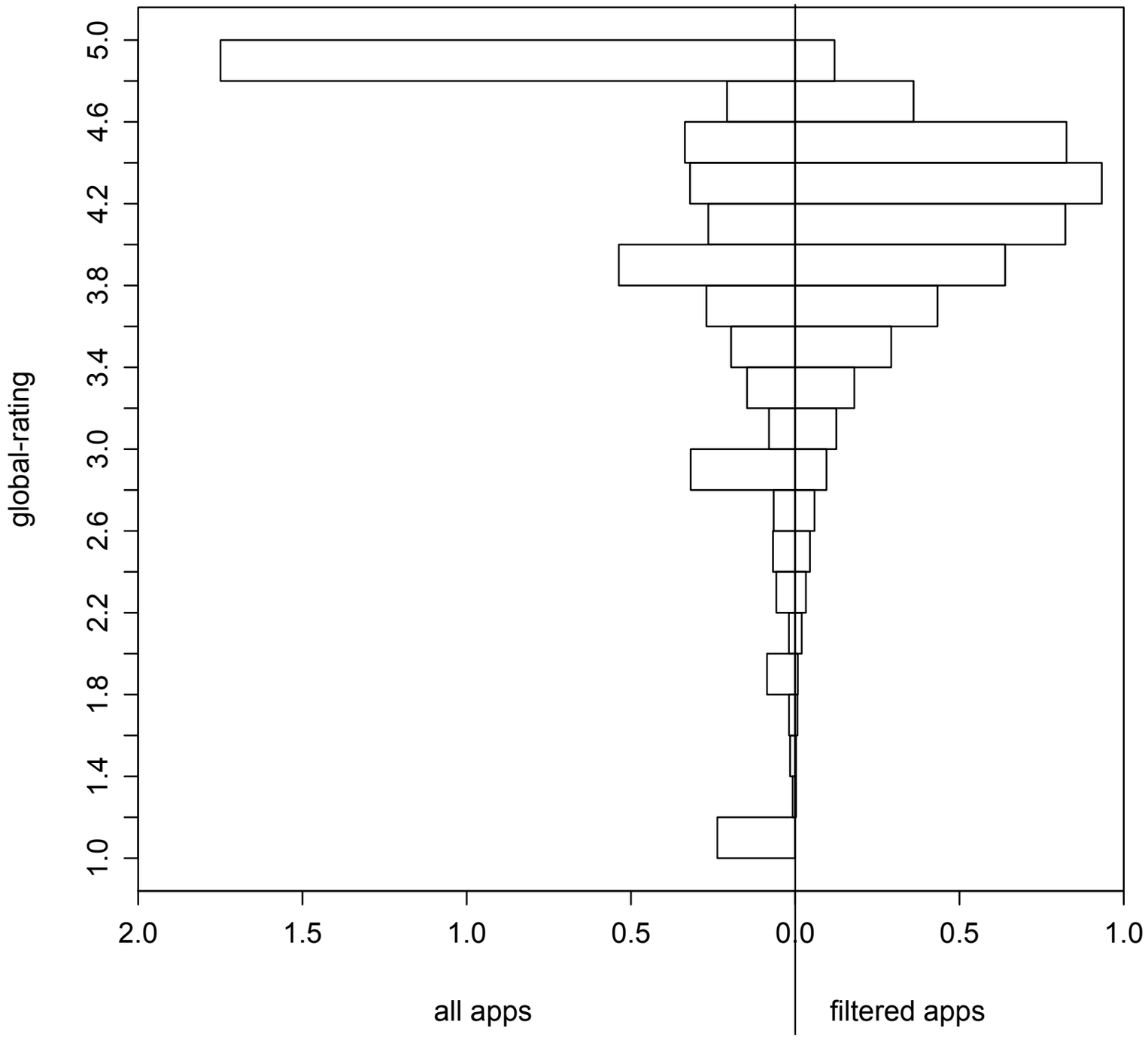
Are Most Apps *Great* ?

Lots of Apps with very few Ratings

128K+

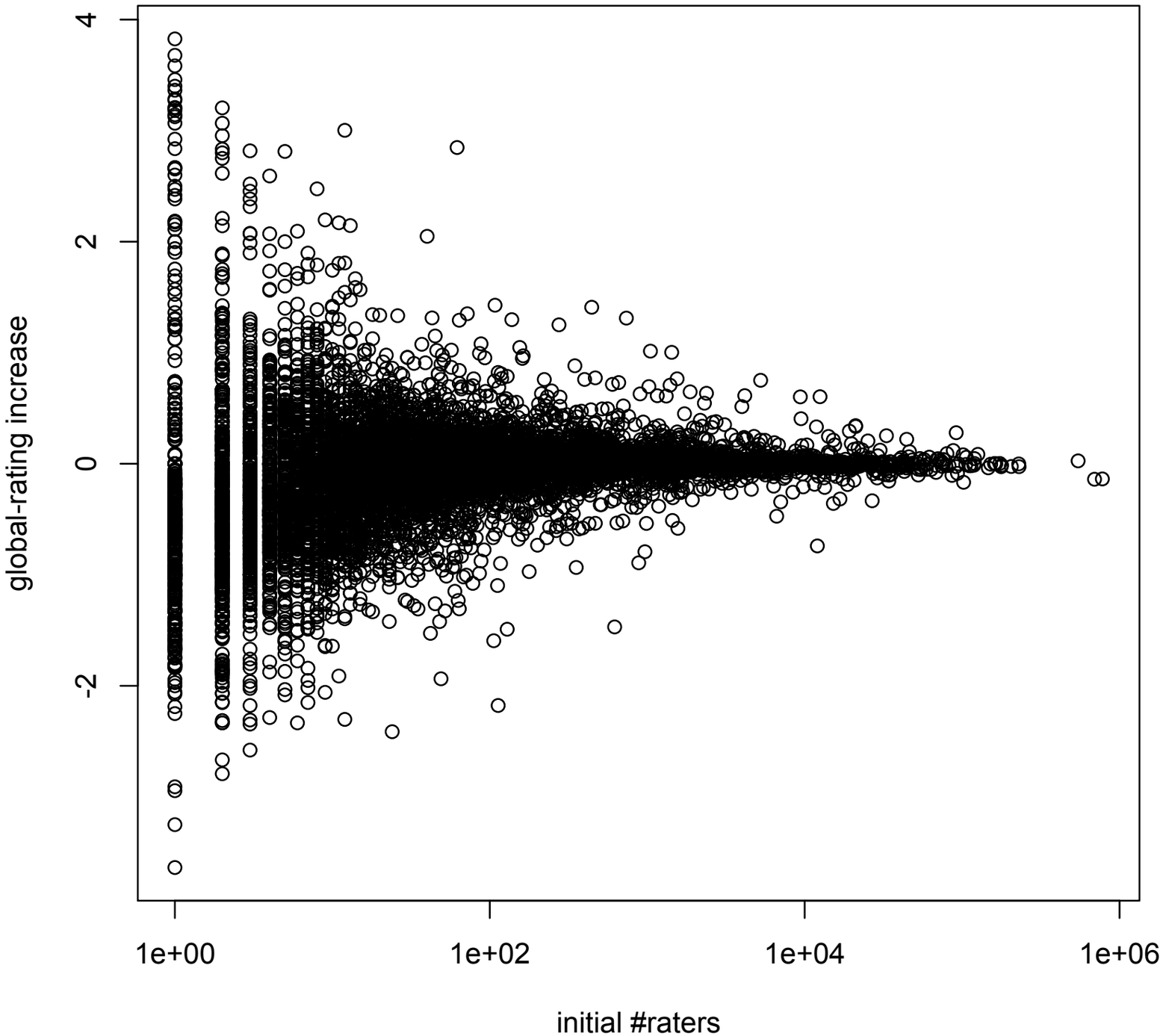


Most apps are *Average*

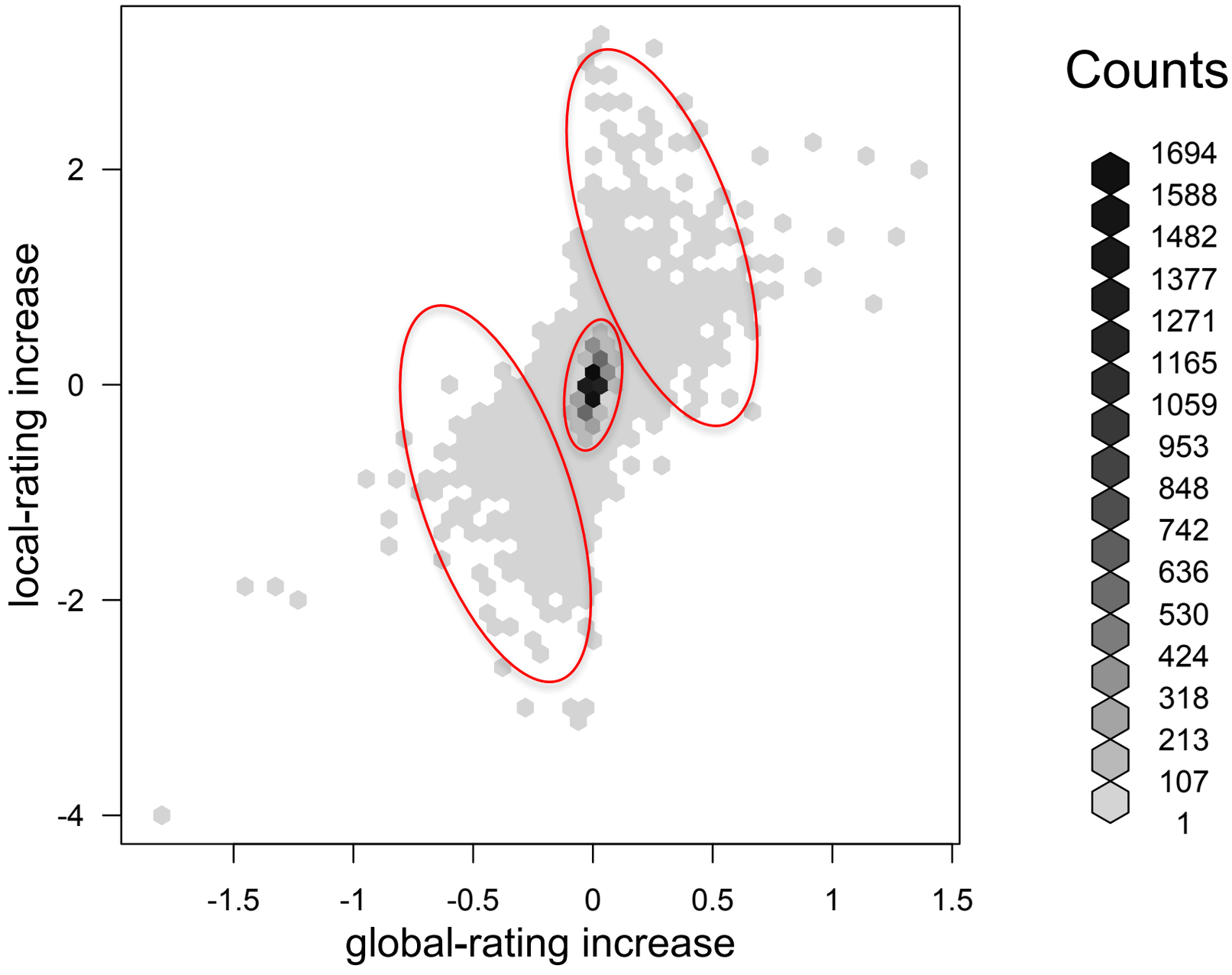


the FALL RISK

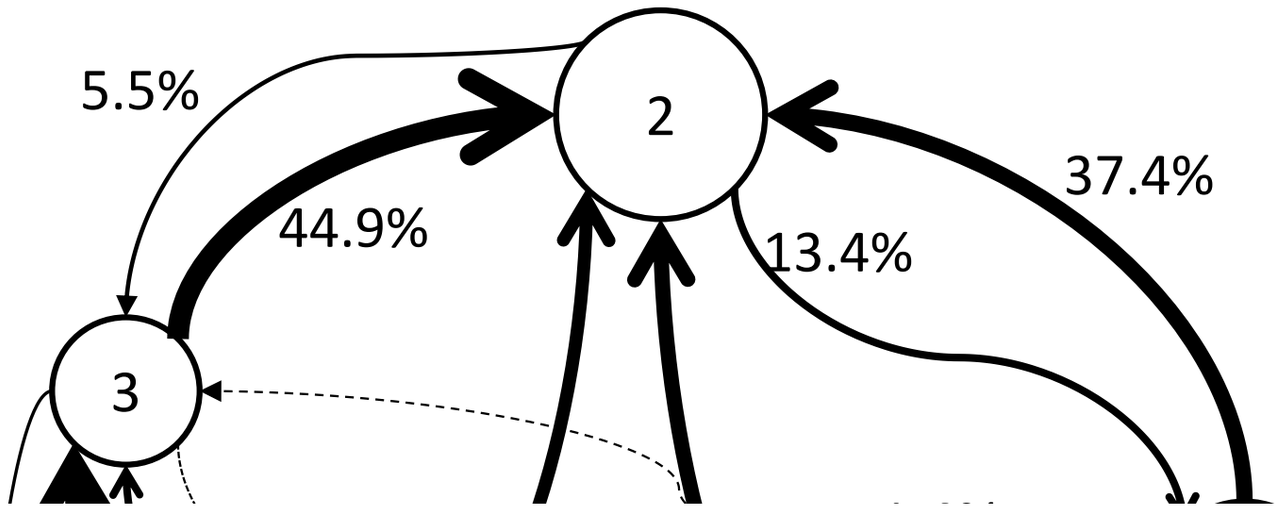
More Raters => Steady Ratings



More Raters => Steady Ratings

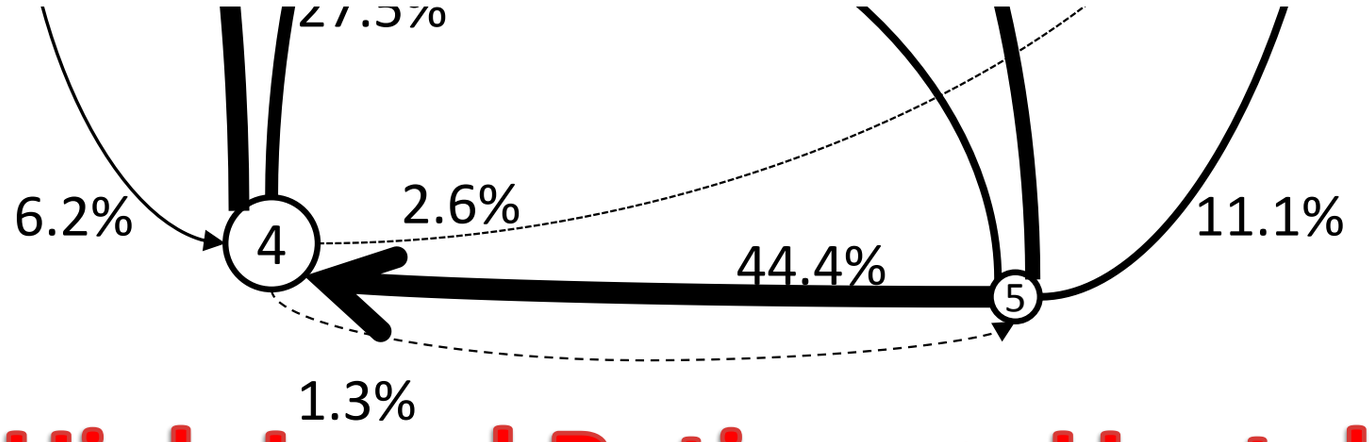


Low Local Rating => Stable



More than 1 star drop

=> Unrecoverable



High Local Rating => Unstable

Part 3

Stats

- Mean/Median/Mode
- Histogram
- Correlation
- Hypothesis testing
- Regression