

# EMPIRICALLY EVALUATING THE QUALITY OF AUTOMATICALLY GENERATED AND MANUALLY WRITTEN TEST SUITES

Jeshua Kracht, Jacob Z. Petrovic, Kristen R. Walcott-Justice  
Department of Computer Science  
University of Colorado at Colorado Springs

# TEST CASE GENERATION

- Manual test case writing
  - -> human thought, effort, time
  - No well-established standard for how to write good test suites
- An alternative: Automatic Test Case Generation
- Existing tools:
  - MoreUnit
  - CodePro
  - EVOSUITE
  - Randoop
  - Jcrasher
  - Palus, and more!



# MEASURING EFFICACY

- Time to Produce
- Number of Test Cases/ Test Suite
- Size / Runtime of Test Suite
- Coverage
  - Statement Coverage
  - Branch Coverage
- Mutation Scores
  
- Can we save time AND achieve good quality?

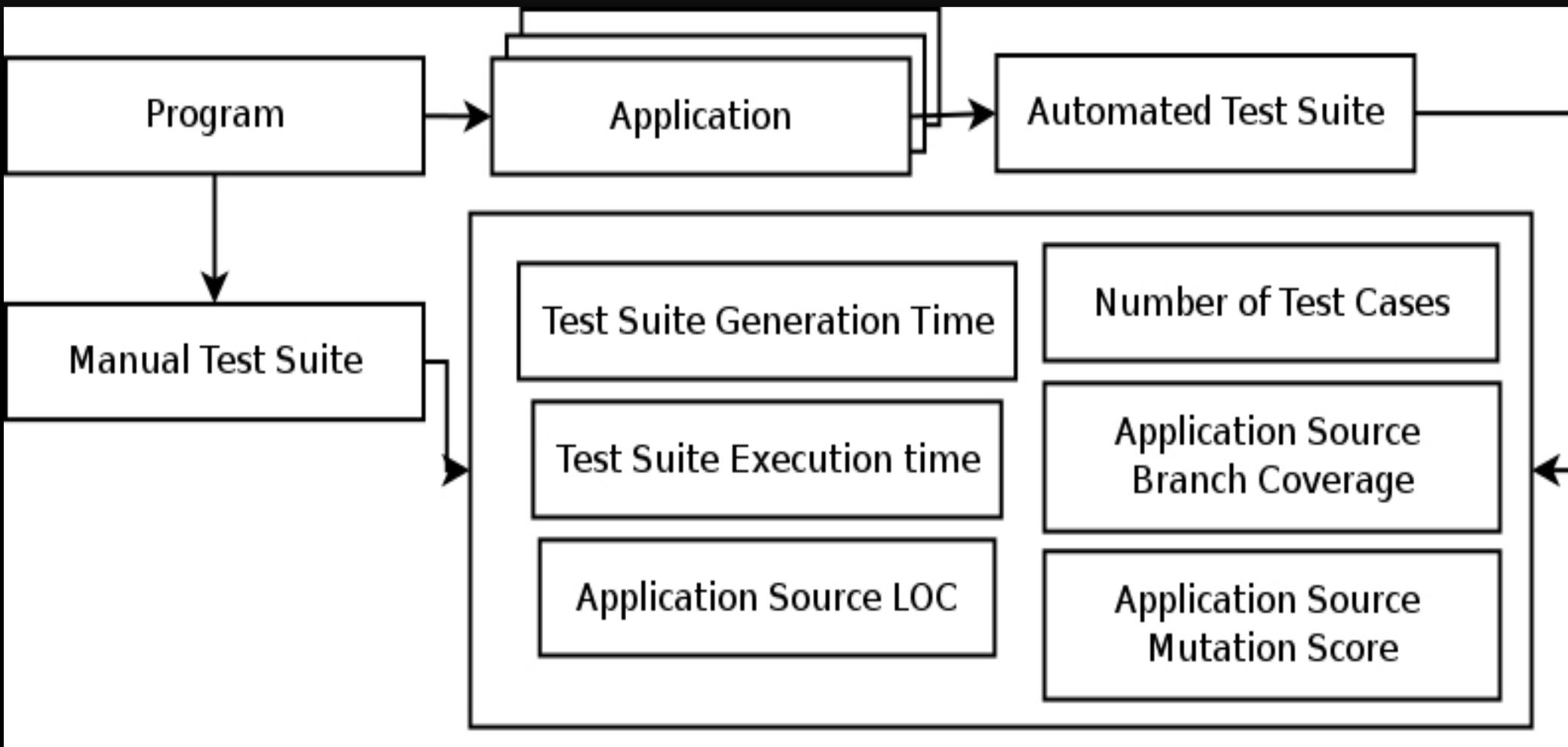
# MAIN CONTRIBUTIONS

- Examination of Automated Tools
- Empirical Analysis of Existing Manual Test Suites
- Empirical Analysis of Automatically Generated Test Suites
- A Comparison of Test Suites
- A Discussion of the Benefits and Drawbacks of Automatic Test Case Generation Tools

# AUTOMATICALLY GENERATED TESTING TOOLS

- Deterministic Tools
  - Skeleton Tests
    - Basic
    - Analysis-based
- Learning Tools
  - Use machine learning algorithms
  - Example: EVOSUITE

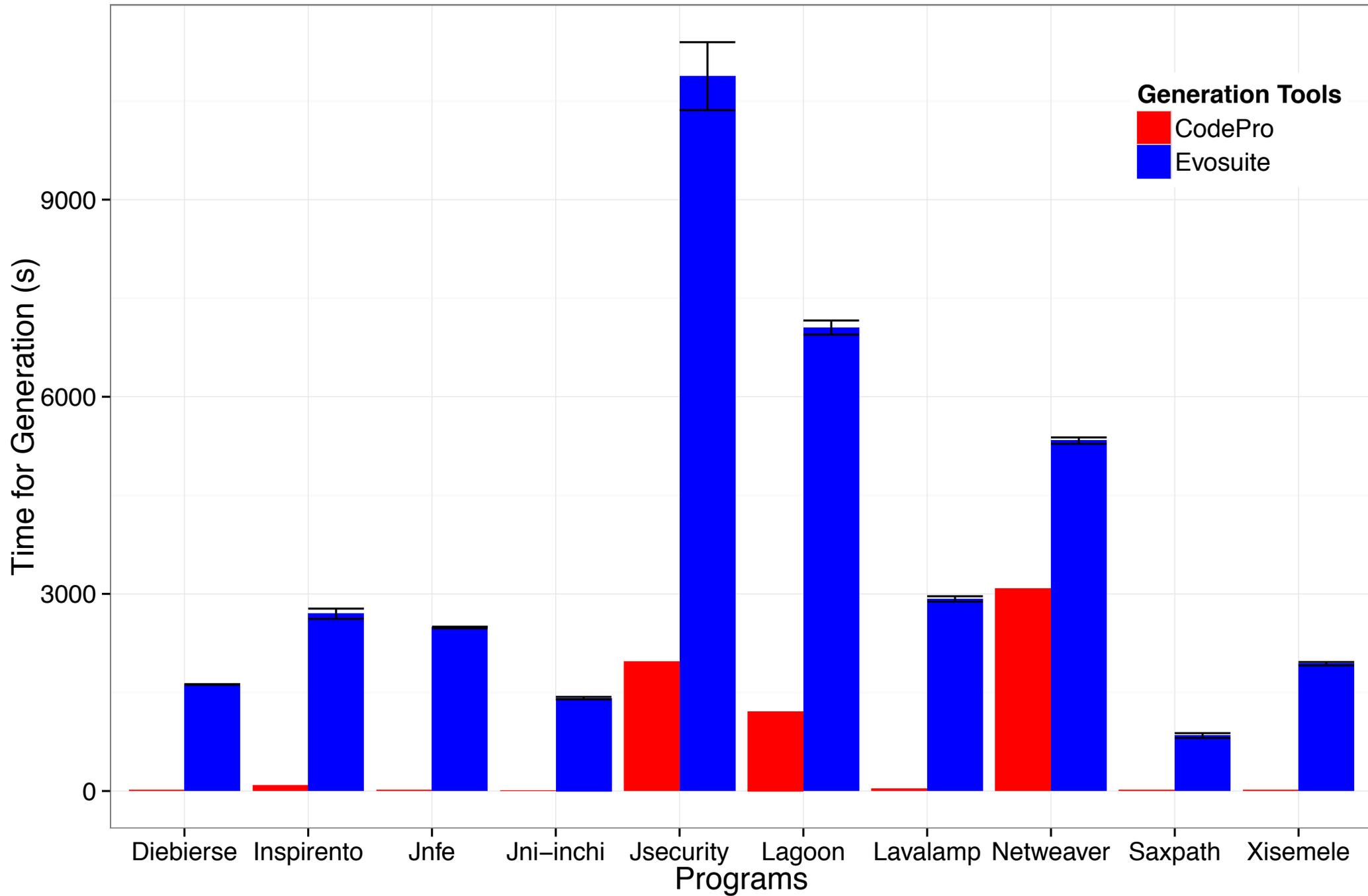
# EMPIRICAL EVALUATION APPROACH



# EXPERIMENT DESIGN

- 10 programs from SF100
- Automatic Test Generation Tools
  - CodePro
  - EVOSUITE (default tuning parameters)
- Metrics
  - LOC and Cyclomatic Complexity - JavaNCSS
  - Execution time, # of tests - JUnit
  - Branch coverage - Jacoco
  - Mutation score - MAJOR

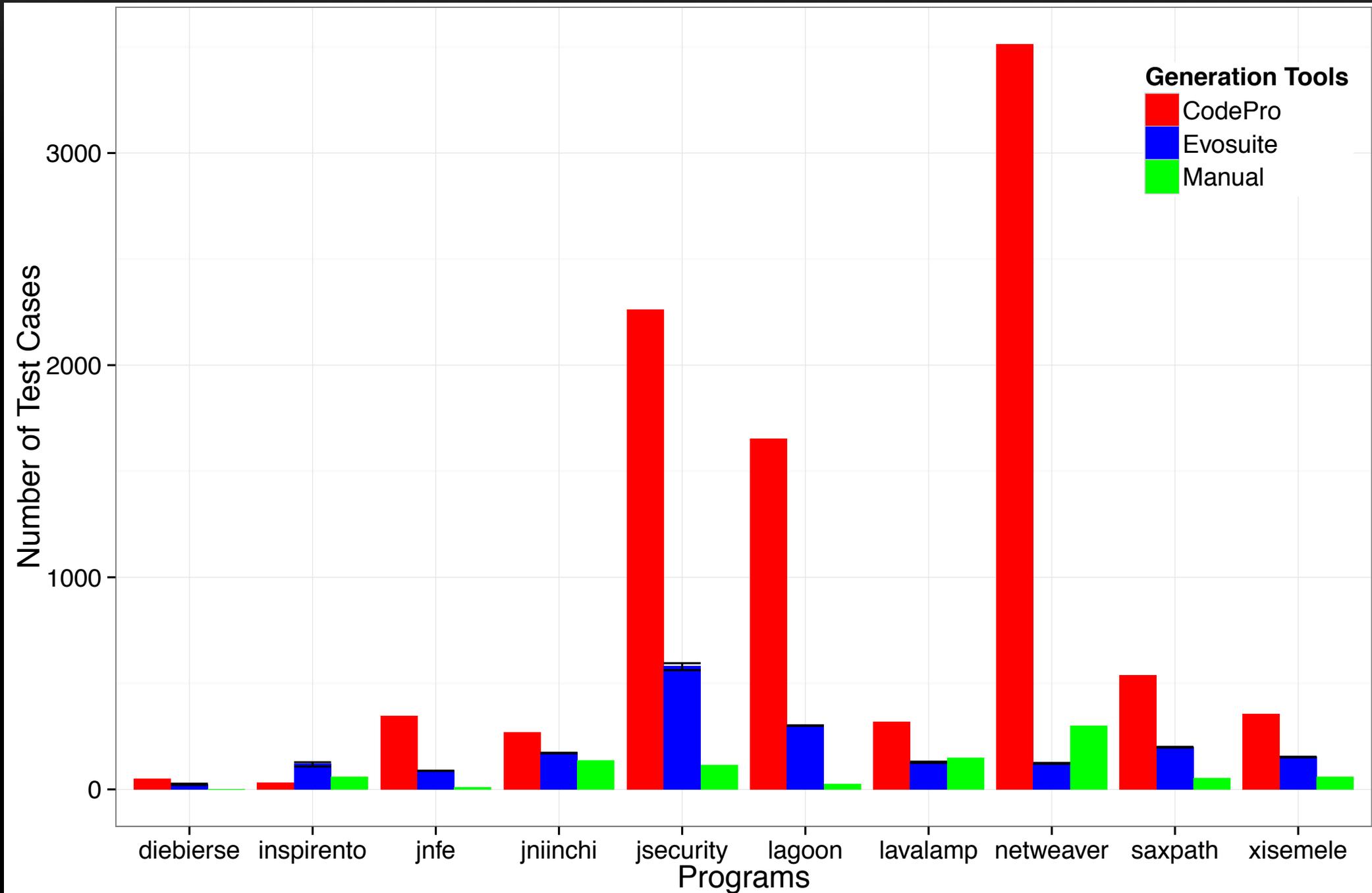
# TIME OF GENERATION



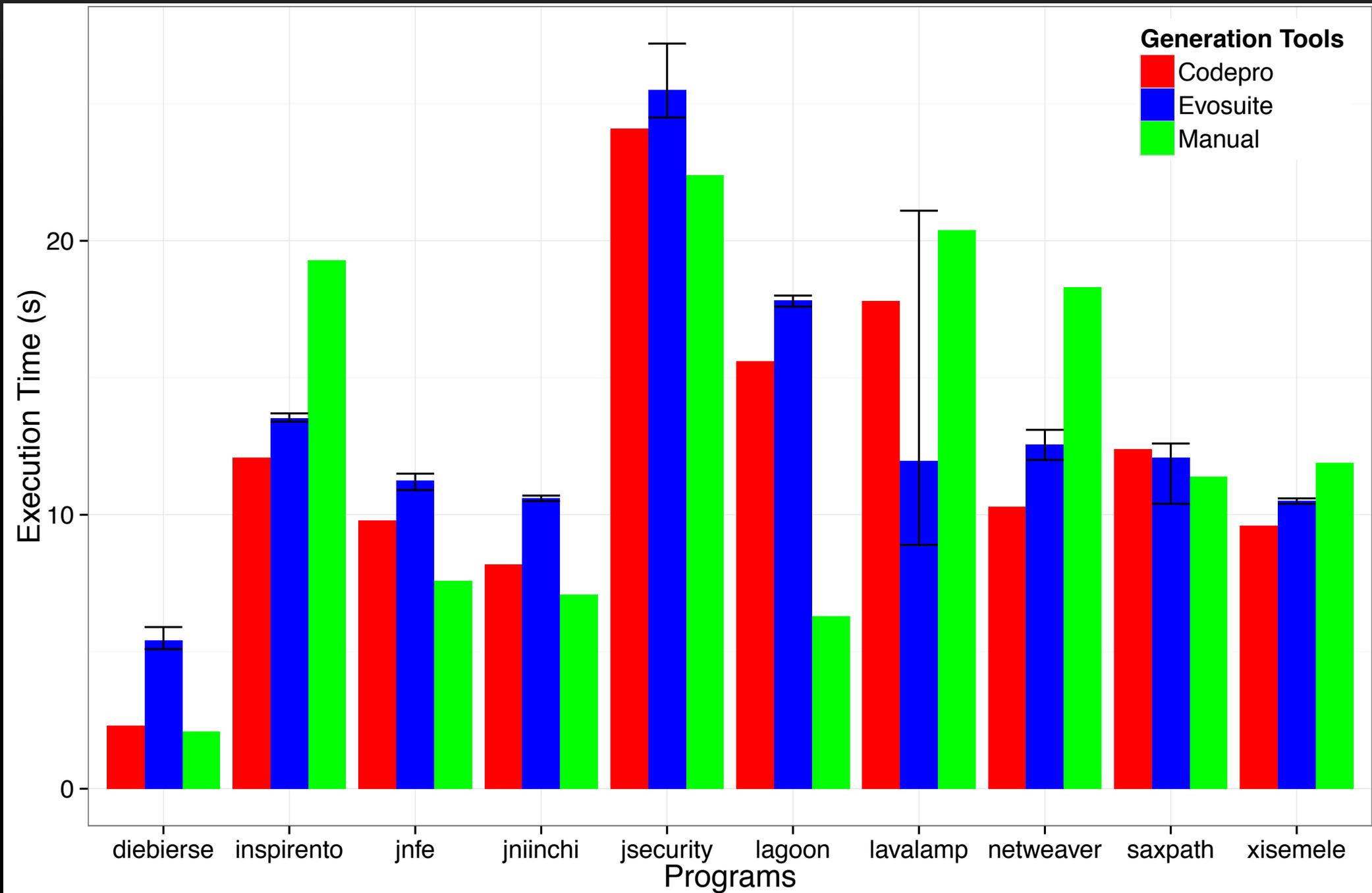
# TEST GENERATION

- CodePro: predictable, scaled linearly based on size and complexity of program
- EvoSuite:
  - Thorough
  - Spent more time in obscure sections of program
- Average 78% more time for EvoSuite to run

# NUMBER OF JUNIT TEST CASES

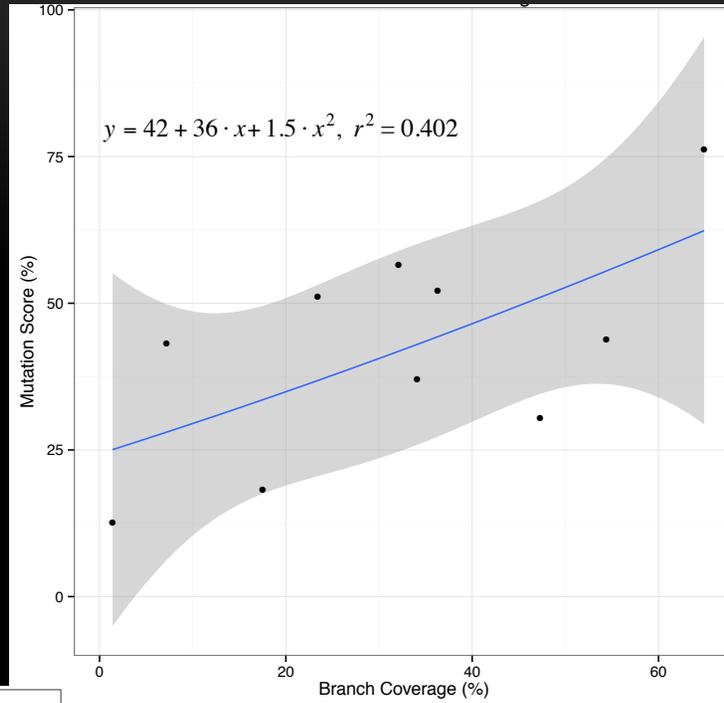


# TEST SUITE EXECUTION TIME

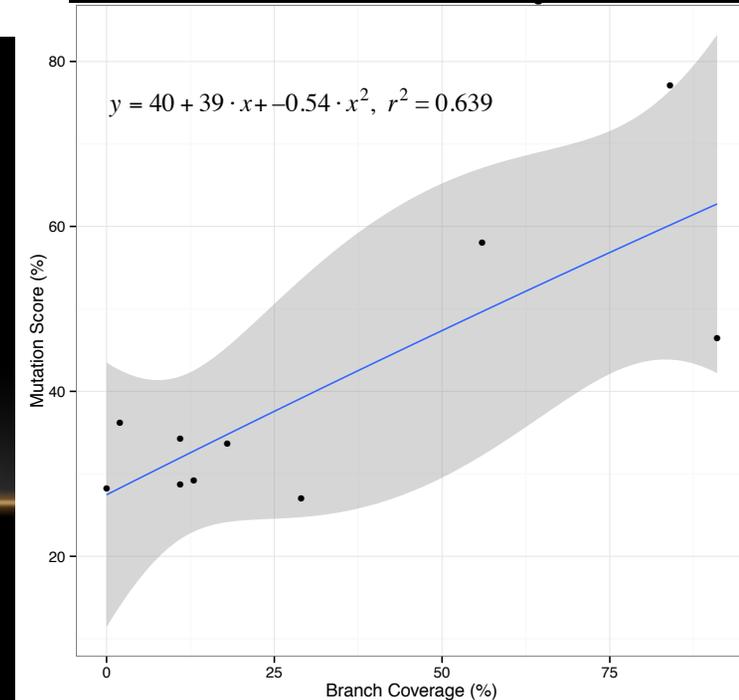
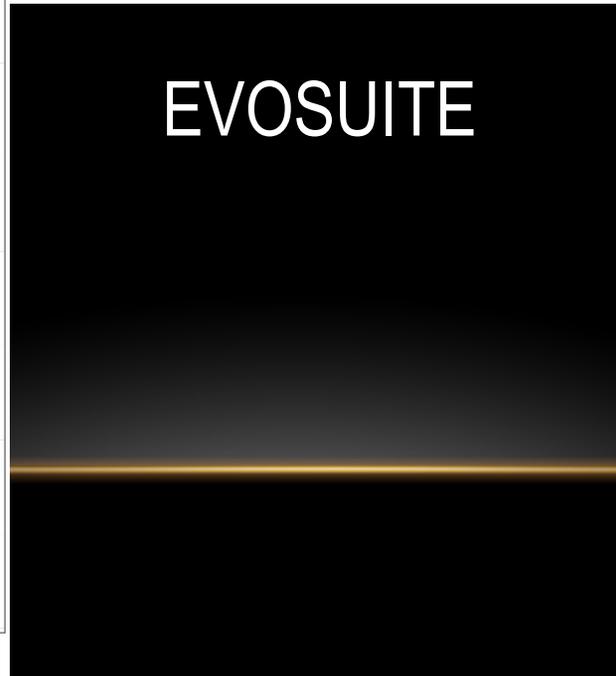
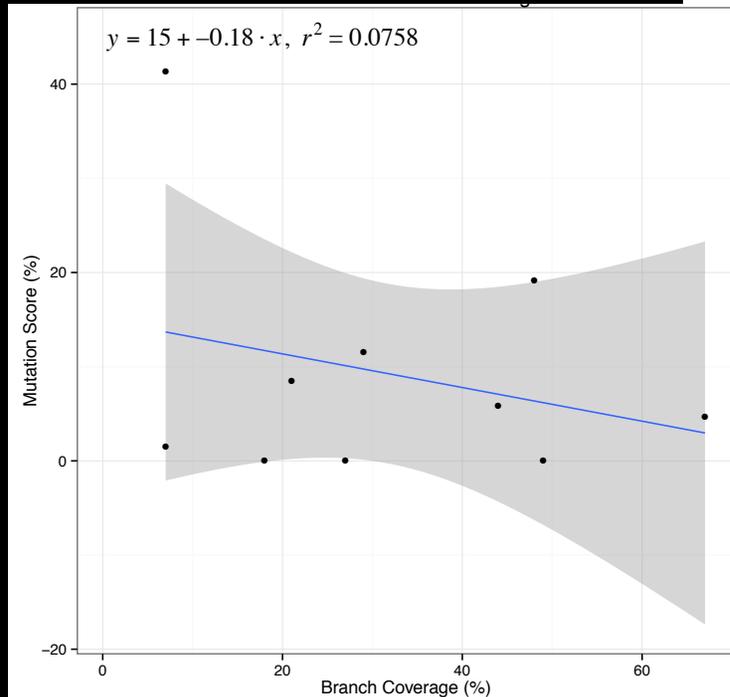


# MUTATION SCORE VS BRANCH COVERAGE

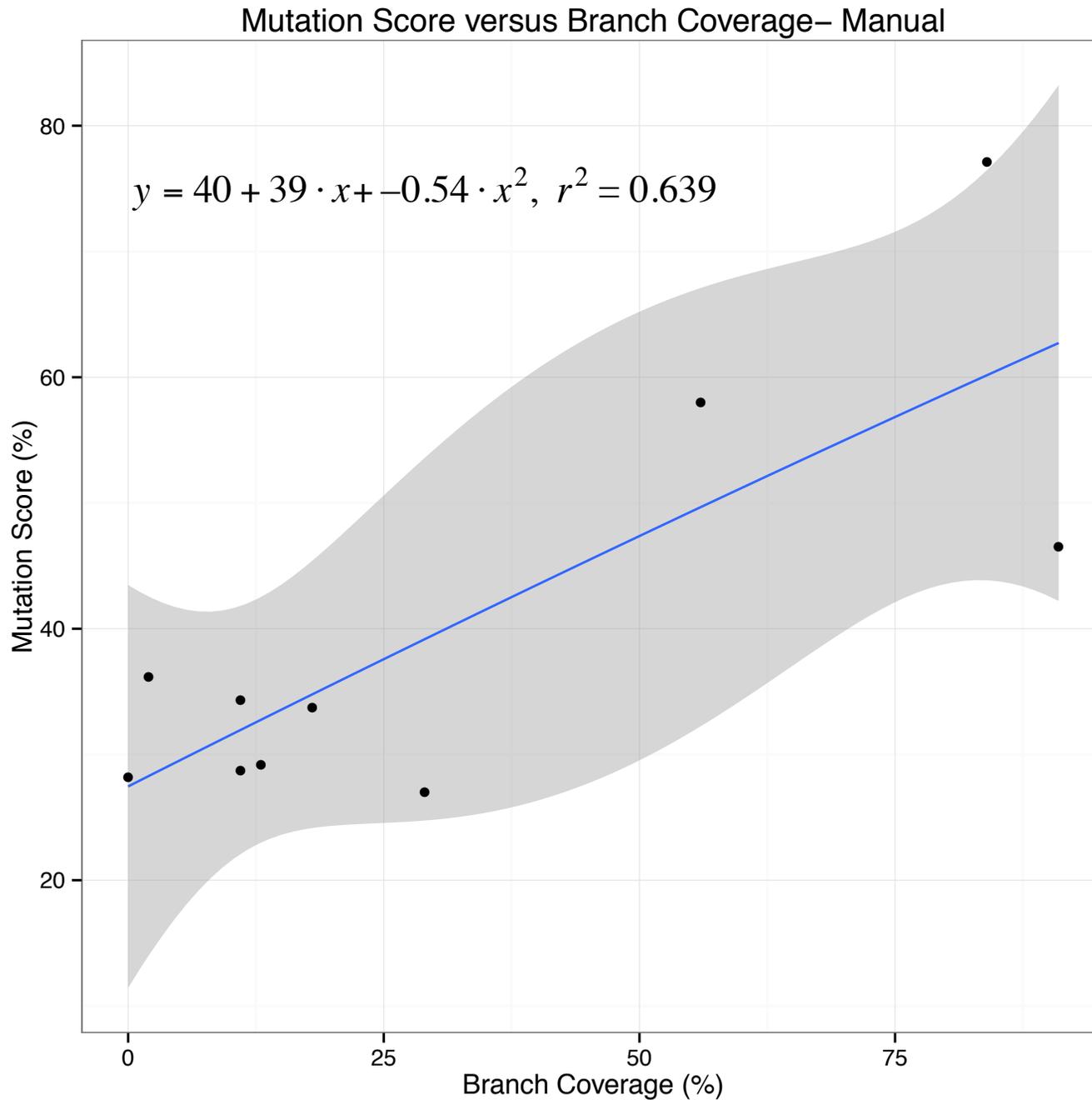
CodePro



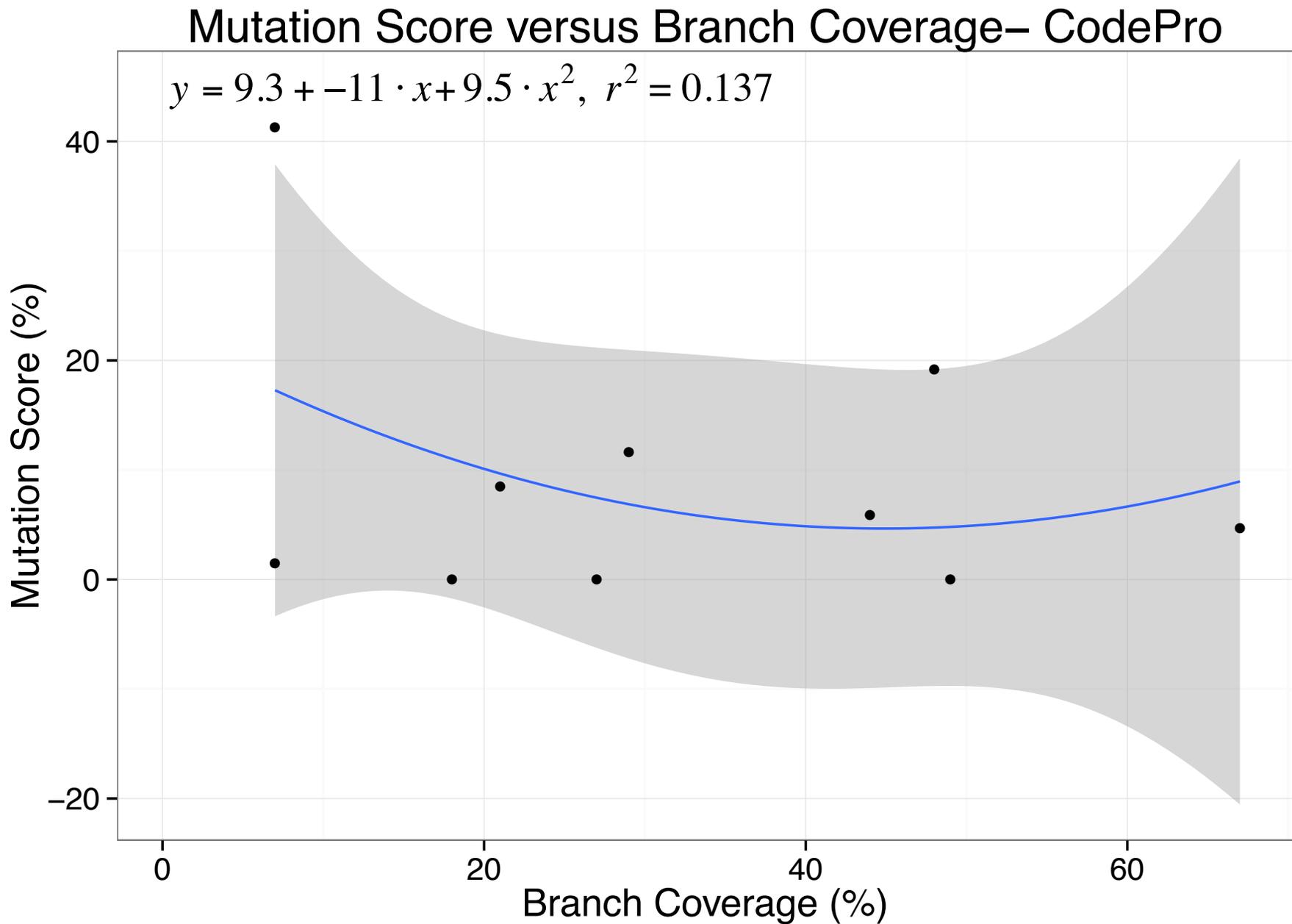
Manual



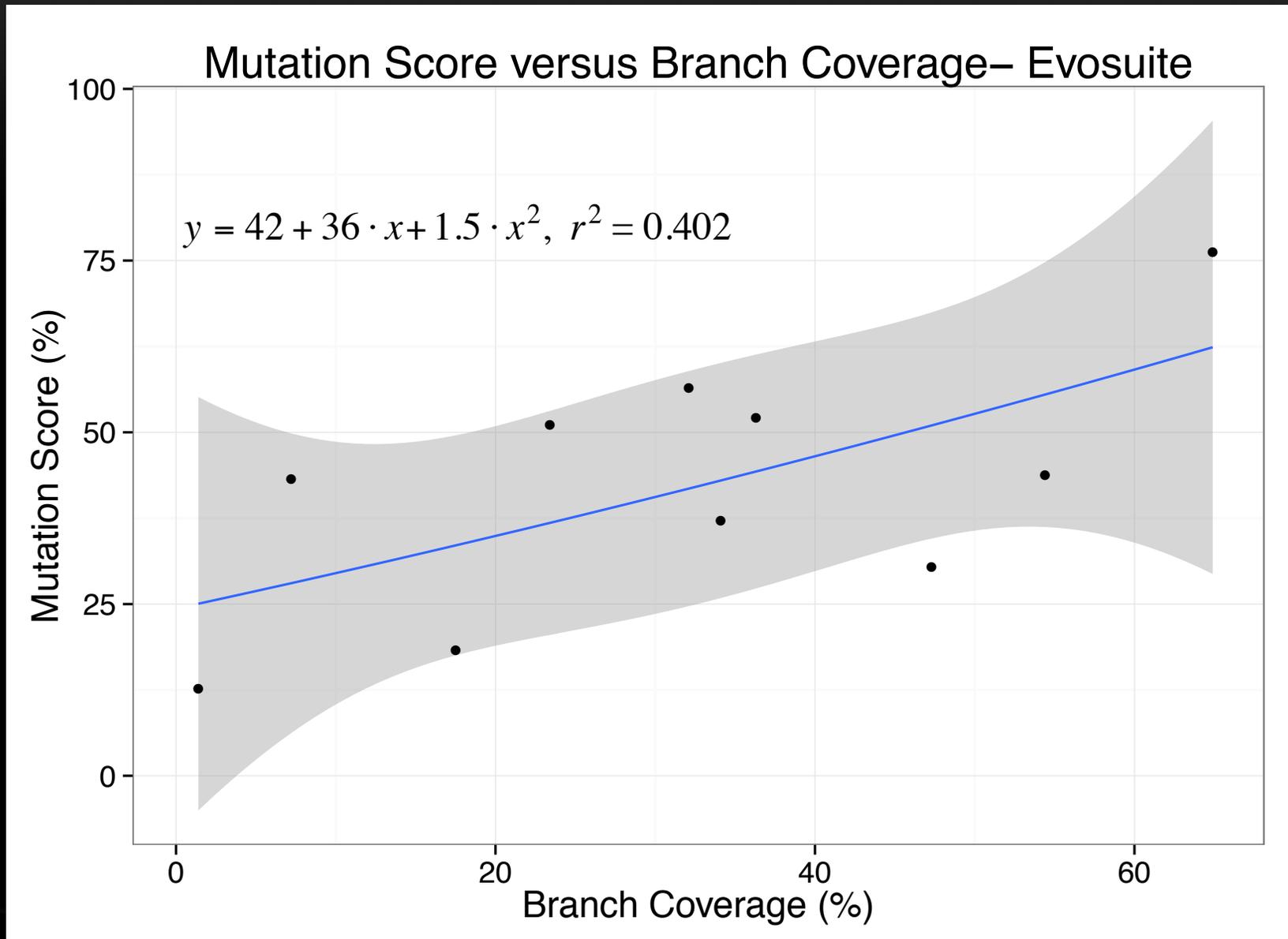
# MANUALLY CREATED TEST SUITES



# CODEPRO CREATED TEST SUITES



# EVOSUITE CREATED TEST SUITES



# IMPORTANT CONTRIBUTIONS

- Much potential in automatic test generation
  - Especially amongst learning based tools!
- Deterministic tools provide skeletons of test cases
  - Future work needed to show extra manual effort needed to improve the tests
- With either type of tool, much time can be saved
  - Future work will show how a hybrid of the techniques can continue to benefit quality while improving efficiency

THANK YOU!



For more information and complete results, see:  
Our paper: Empirically Evaluating the Quality of Automatically  
Generated and Manually Written Test Suites  
and

<http://cs.uccs.edu/~kjustice/QSIC2014>

17