

# Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward

Gregory M. Kapfhammer

Department of Computer Science

Allegheny College

<http://www.cs.allegheny.edu/~gkapfham/>



# ALLEGHENY COLLEGE

1st International Workshop on Regression Testing  
Co-Located with the 4th IEEE International Conference on  
Software Testing, Verification and Validation  
Berlin, Germany, March 2011

# Presentation Overview

Regression  
Testing  
Community

# Presentation Overview

Current  
Trends

Regression  
Testing  
Community

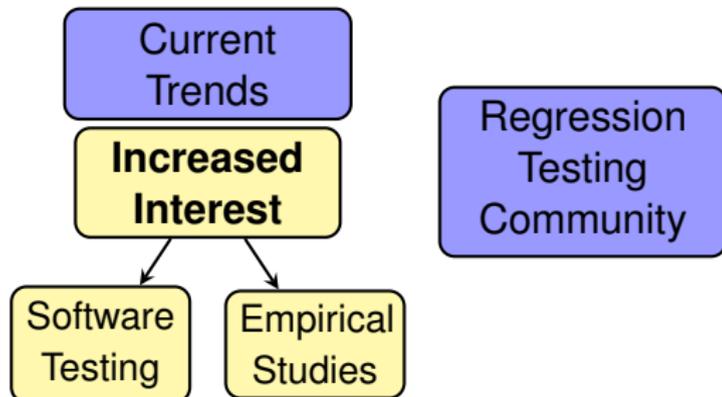
# Presentation Overview

Current  
Trends

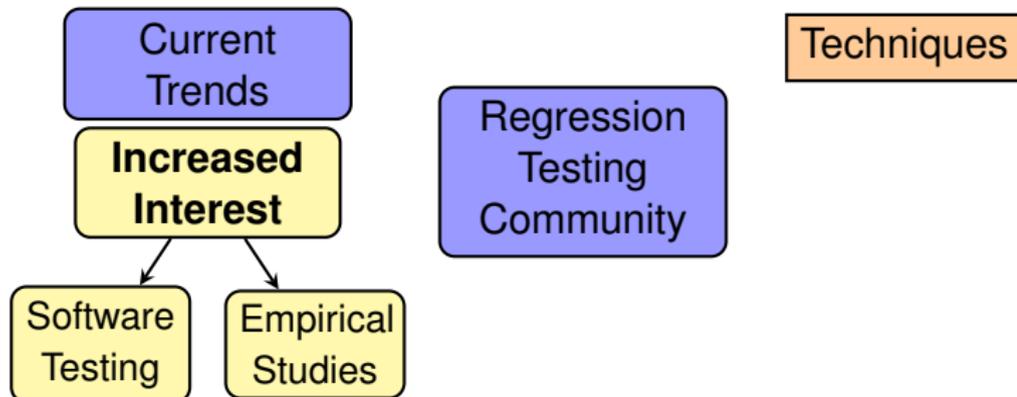
**Increased  
Interest**

Regression  
Testing  
Community

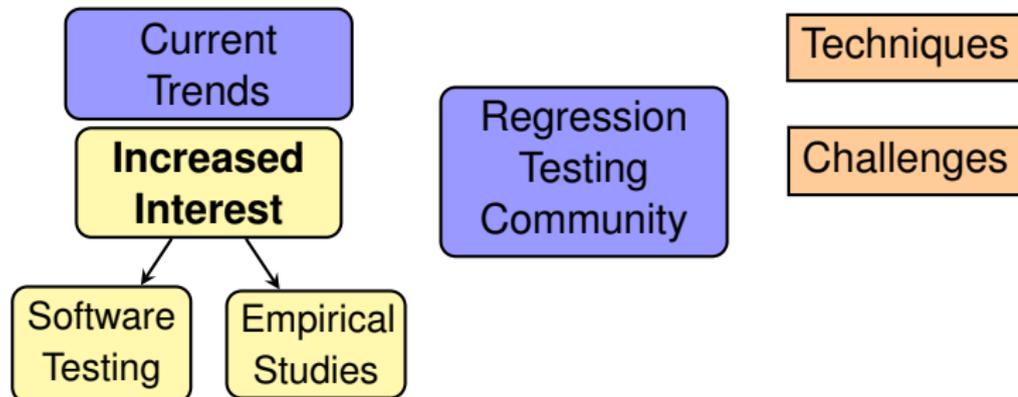
# Presentation Overview



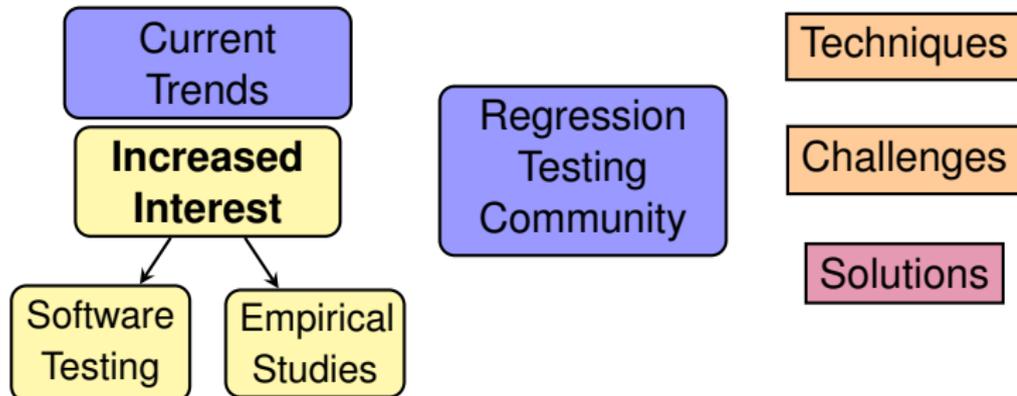
# Presentation Overview



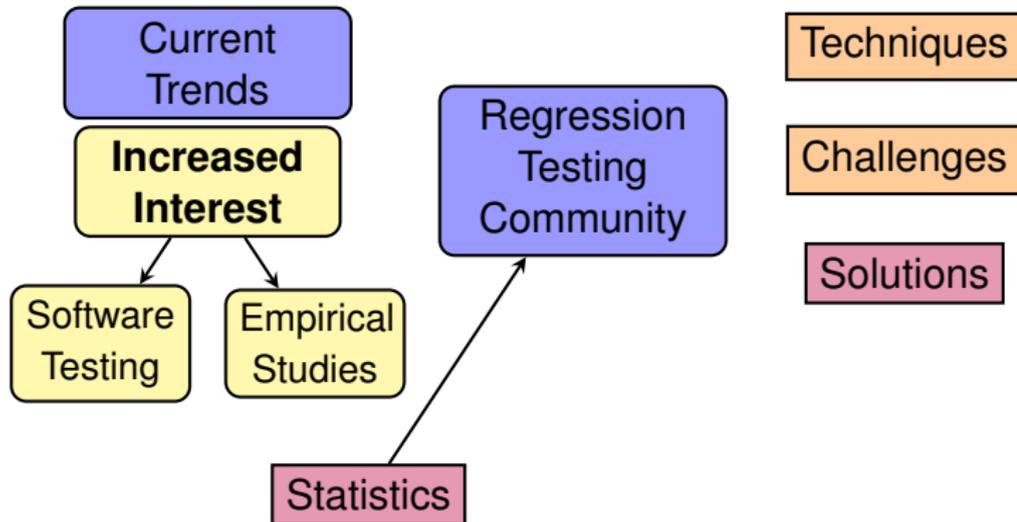
# Presentation Overview



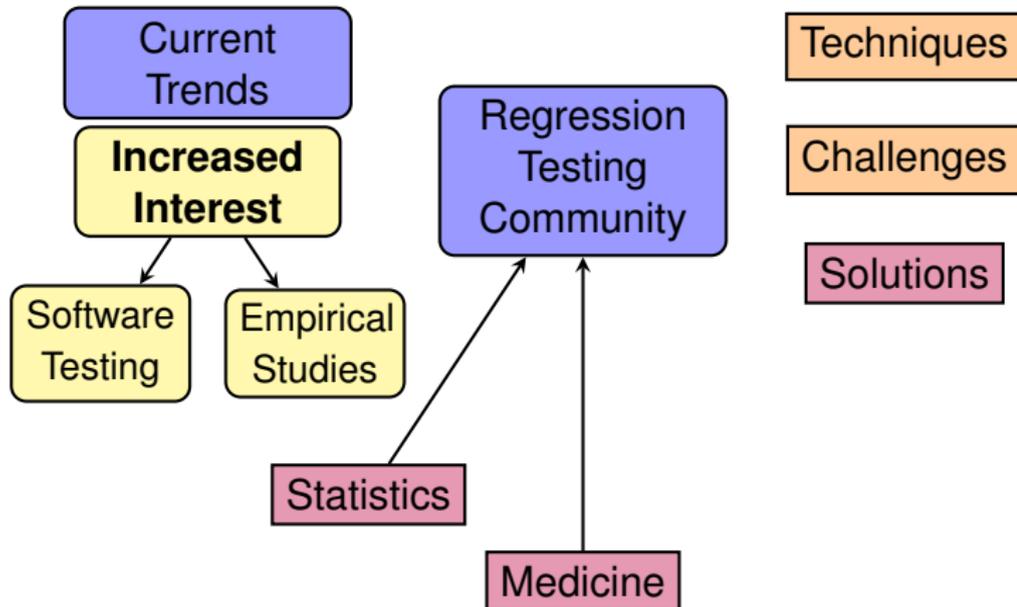
# Presentation Overview



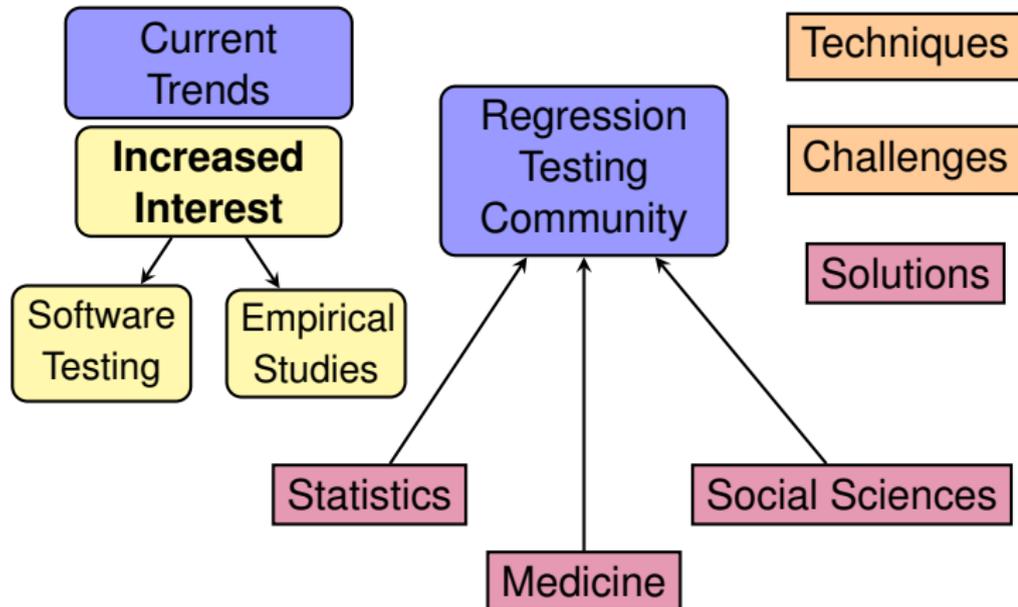
# Presentation Overview



# Presentation Overview

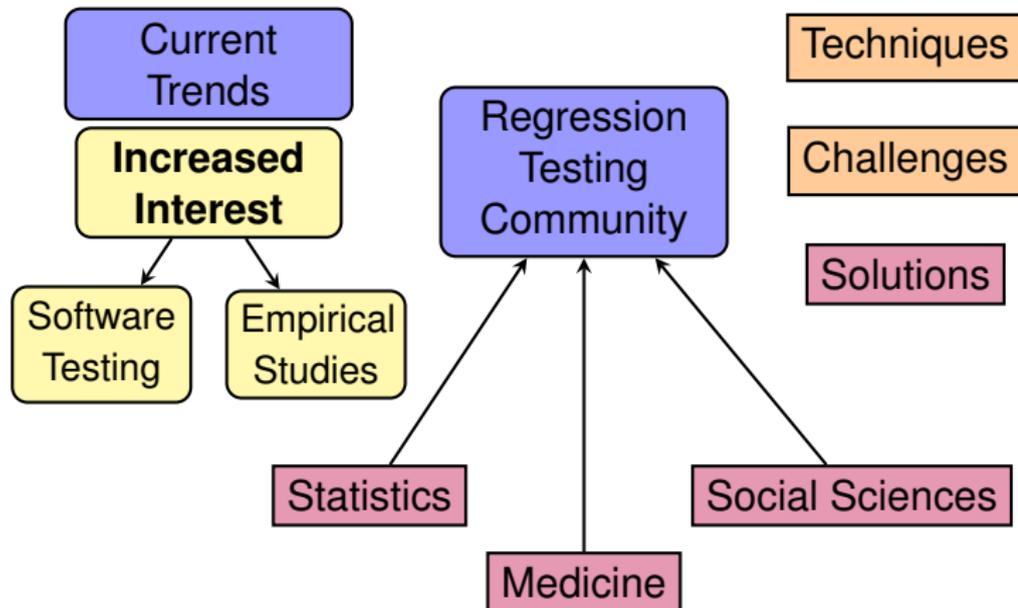


# Presentation Overview

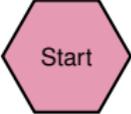


# Presentation Overview

Mutually Beneficial Sharing of All Artifacts Used in Experimentation

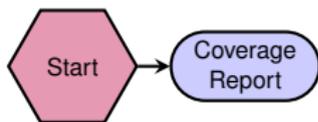


# Model of Regression Testing

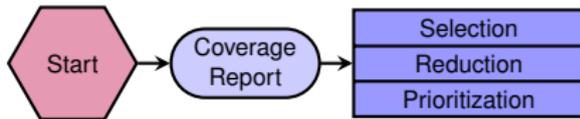


Start

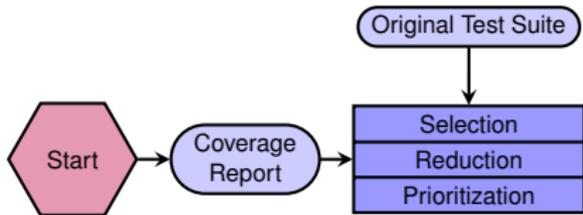
# Model of Regression Testing



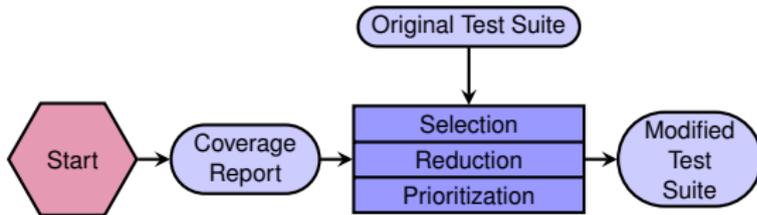
# Model of Regression Testing



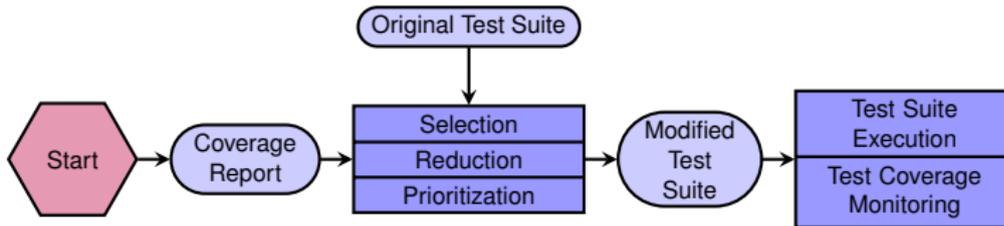
# Model of Regression Testing



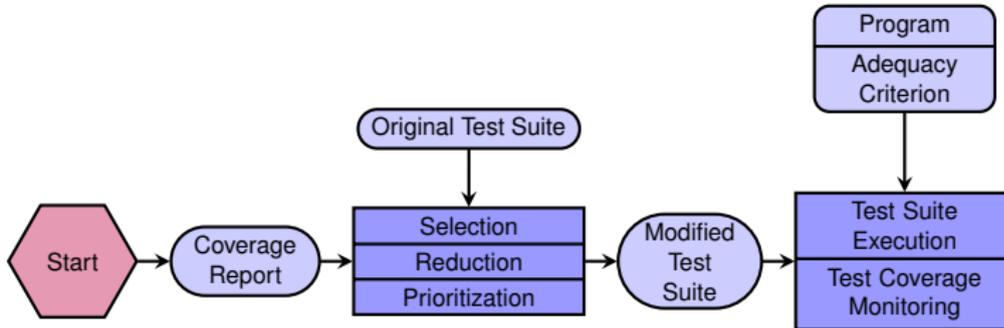
# Model of Regression Testing



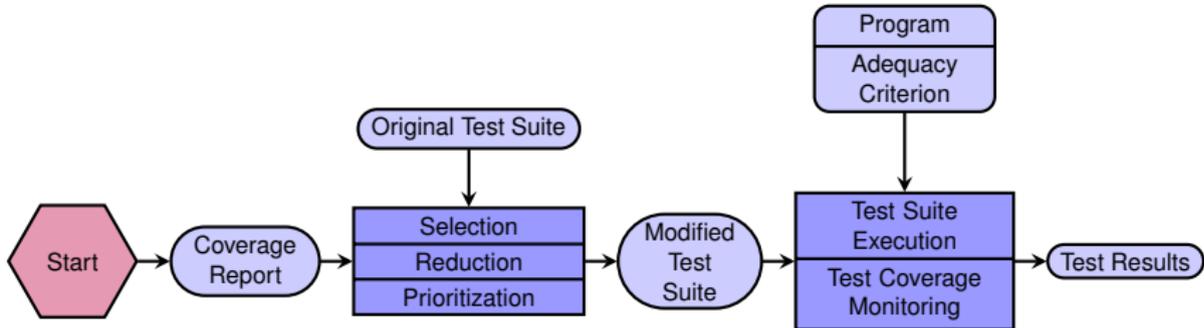
# Model of Regression Testing



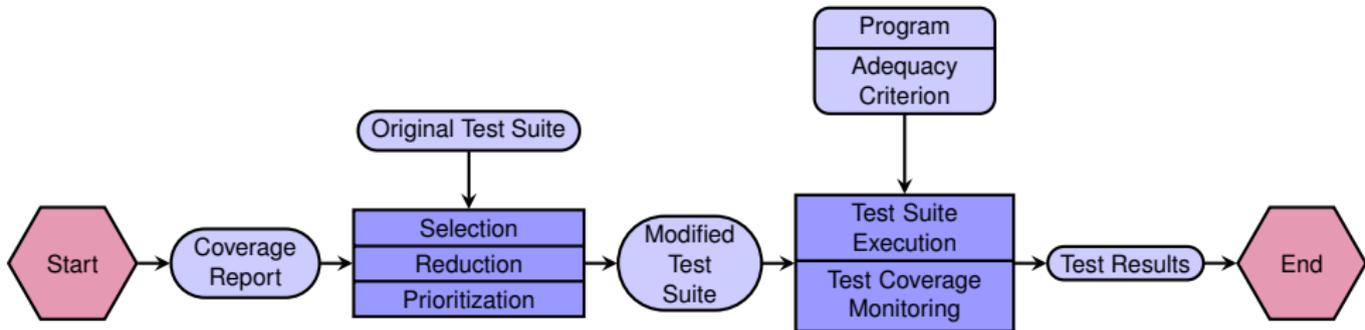
# Model of Regression Testing



# Model of Regression Testing

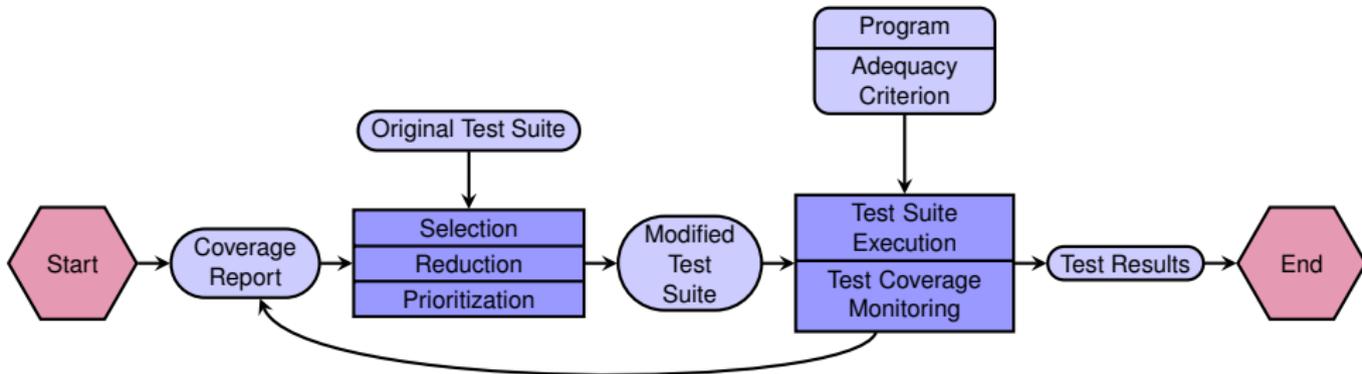


# Model of Regression Testing



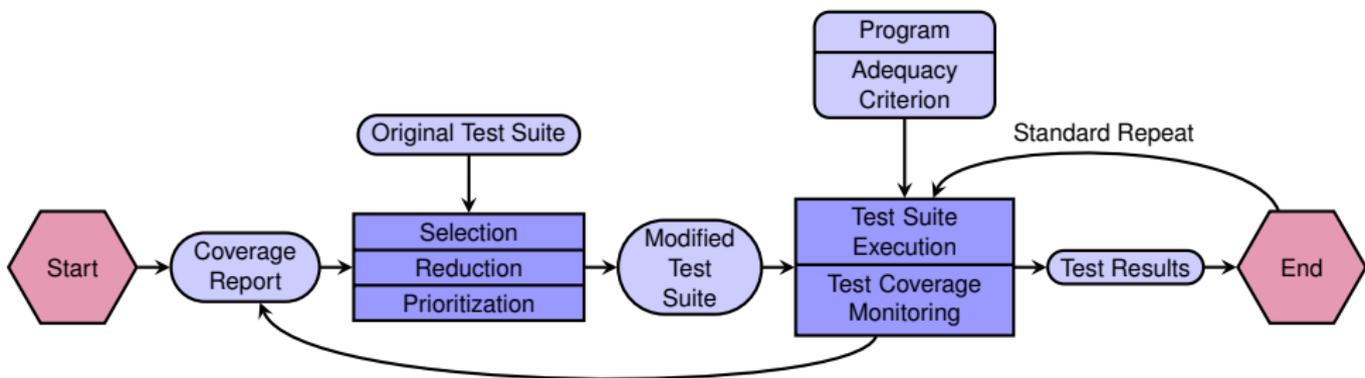
# Model of Regression Testing

Use the Coverage Report During the Next Round of Regression Testing



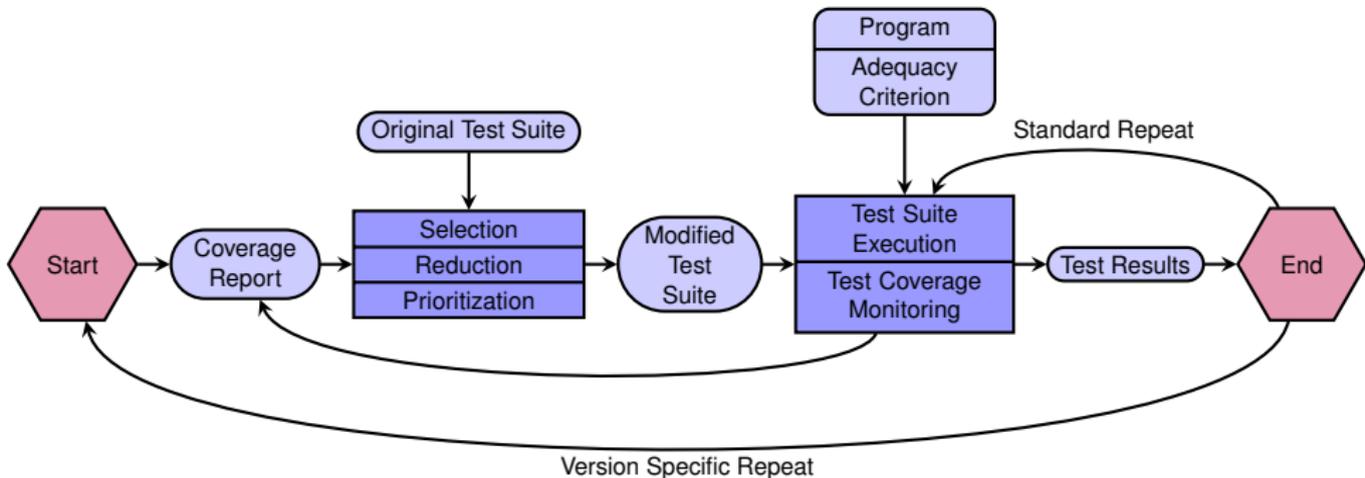
# Model of Regression Testing

Use the Same Test Suite for the Next Round of Regression Testing



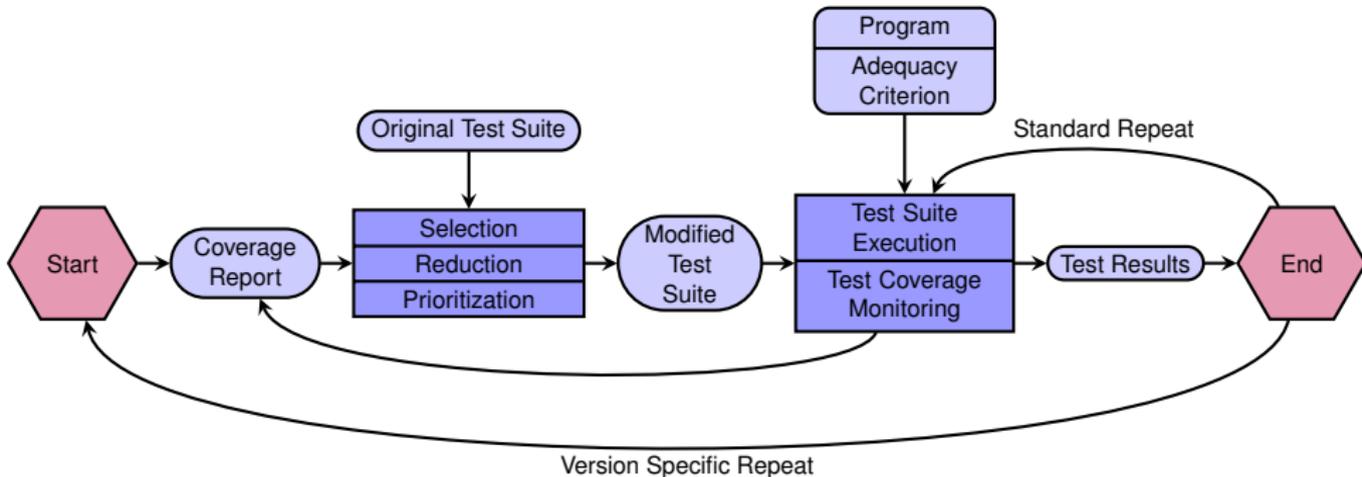
# Model of Regression Testing

Make a New Test Suite for the Next Round of Regression Testing



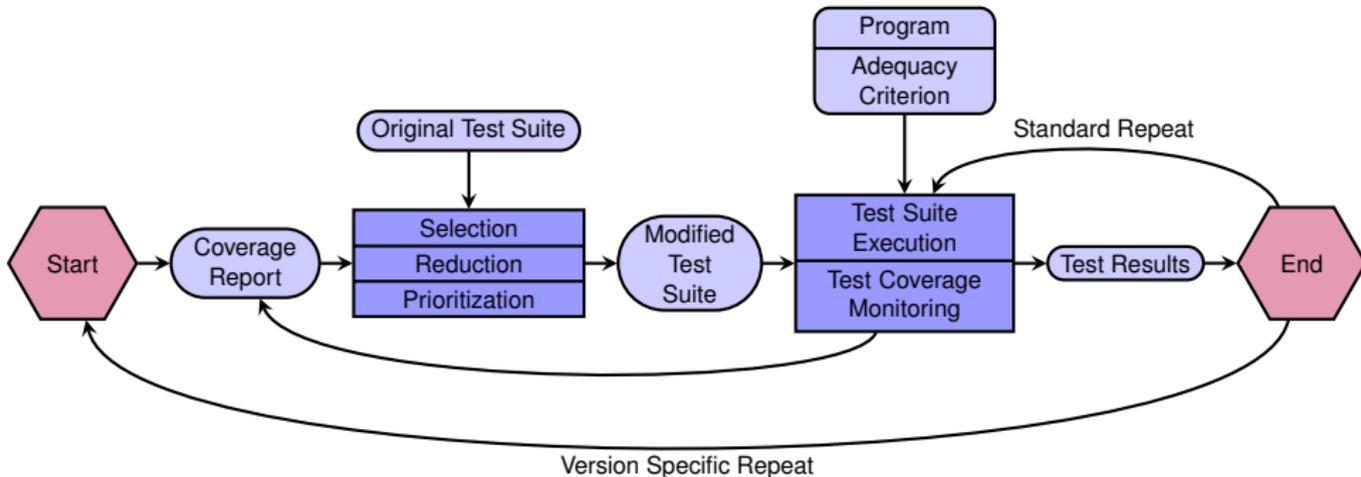
# Model of Regression Testing

Practitioners are unwilling to use methods for “lack of empirical studies” [12]



# Model of Regression Testing

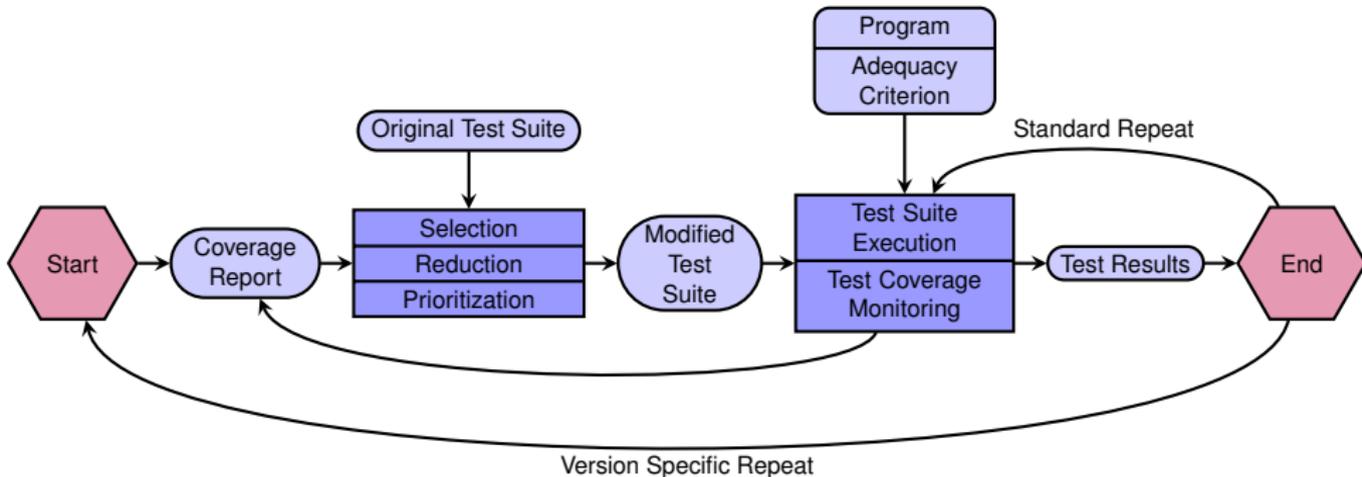
Practitioners are unwilling to use methods for “lack of empirical studies” [12]



Testing tools do not produce outputs in the best format (e.g., per-test coverage)

# Model of Regression Testing

Practitioners are unwilling to use methods for “lack of empirical studies” [12]

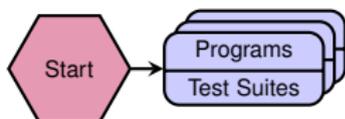


Experimental assessment could stagnate due to inaccessibility of artifacts

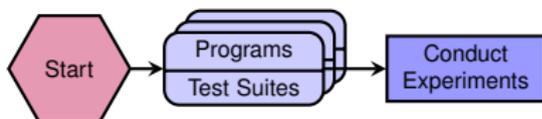
# Model of Experimental Evaluation



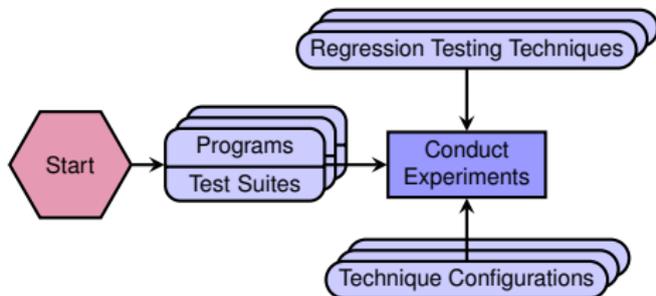
# Model of Experimental Evaluation



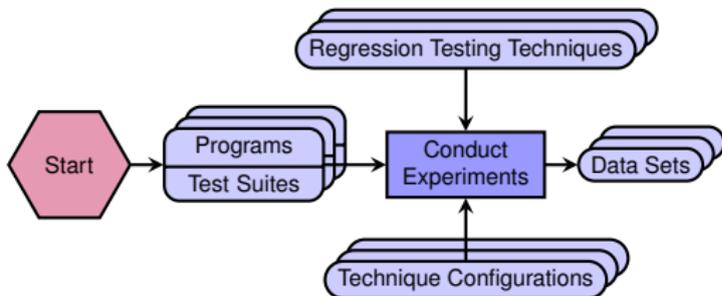
# Model of Experimental Evaluation



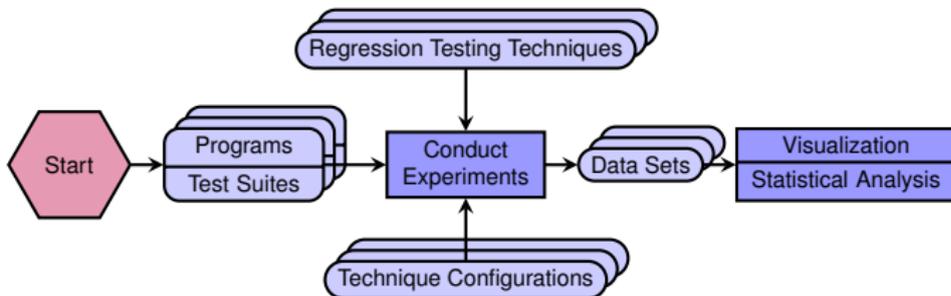
# Model of Experimental Evaluation



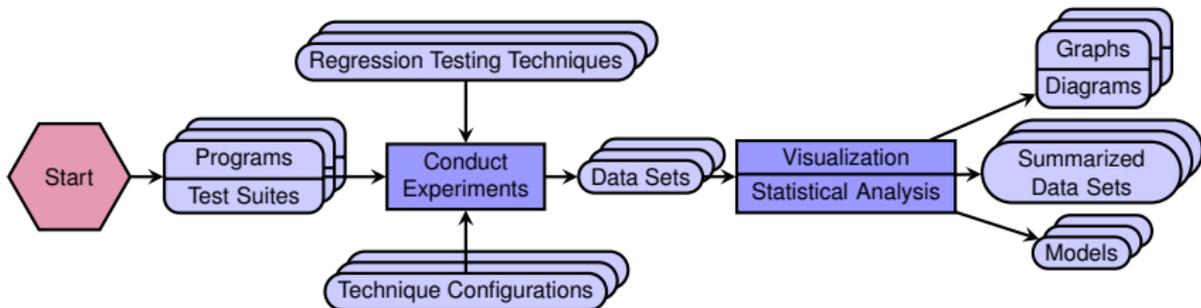
# Model of Experimental Evaluation



# Model of Experimental Evaluation

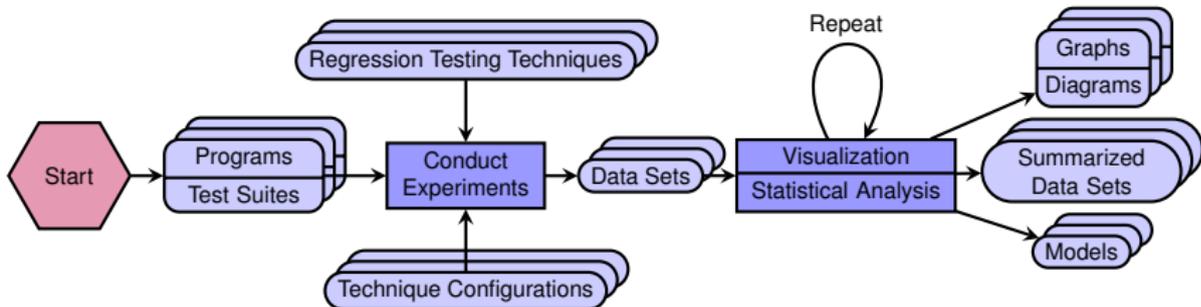


# Model of Experimental Evaluation



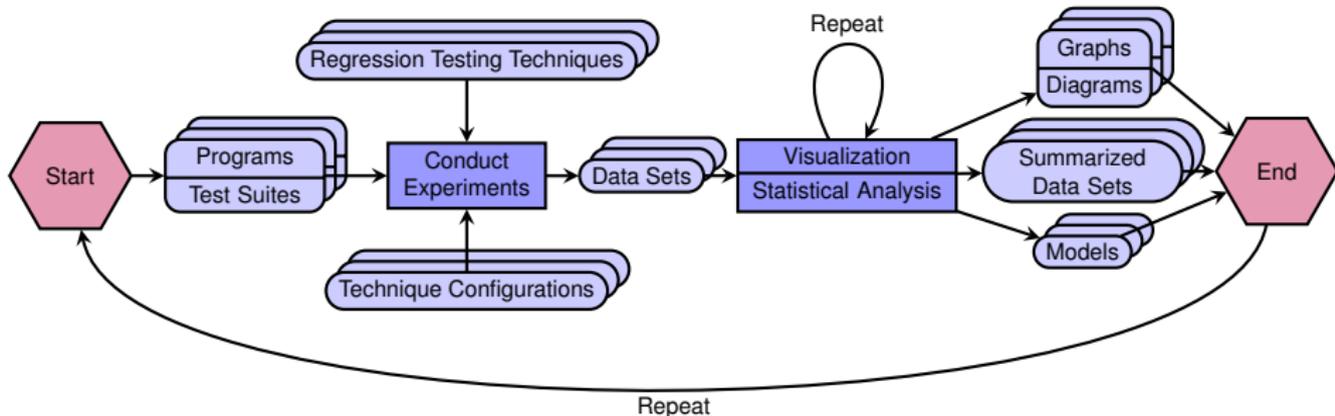
# Model of Experimental Evaluation

Iteratively Perform Visualization and Statistical Analysis



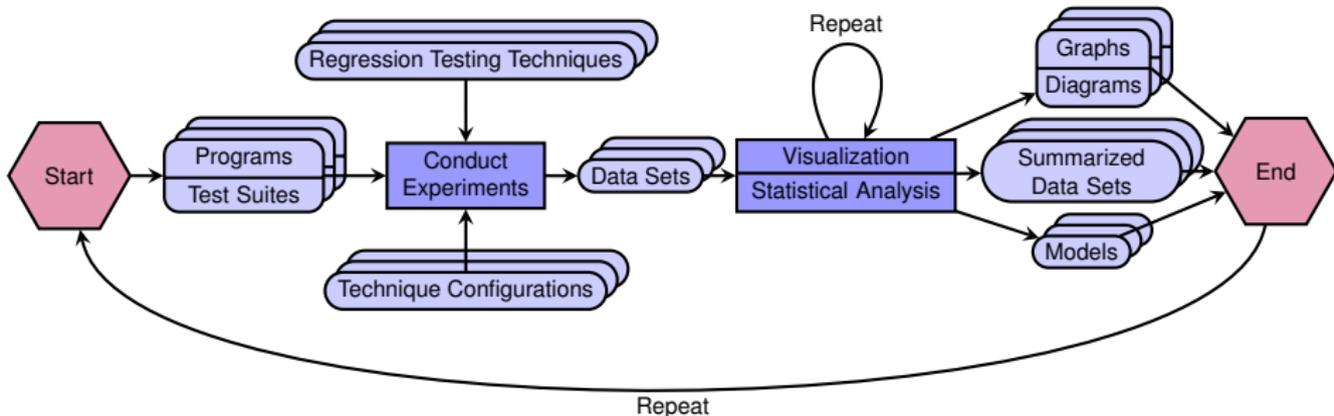
# Model of Experimental Evaluation

Conduct Experiments with Additional Programs, Test Suites, and Techniques



# Model of Experimental Evaluation

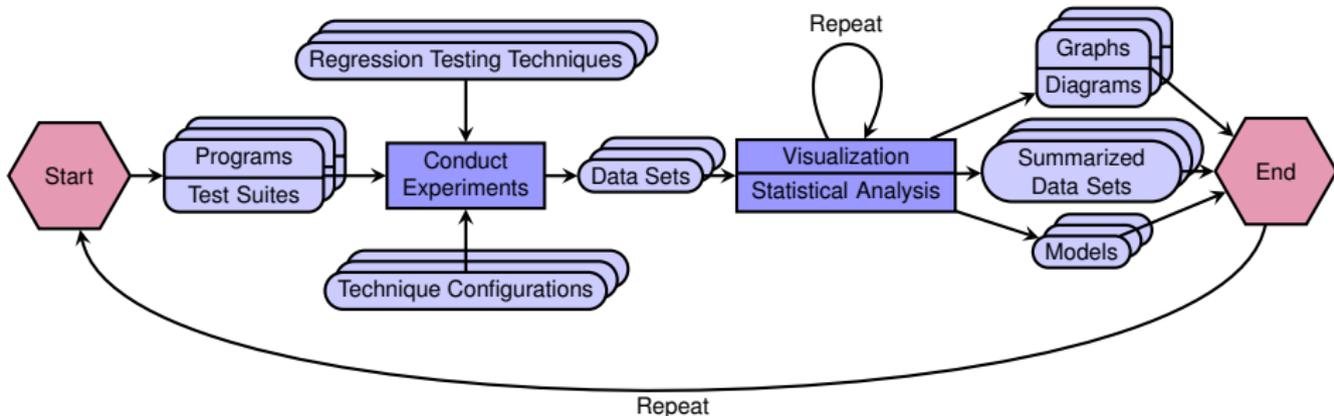
Conduct Experiments with Additional Programs, Test Suites, and Techniques



56% of Papers Surveyed by Yoo and Harman Only Used the SIR Programs [3]

# Model of Experimental Evaluation

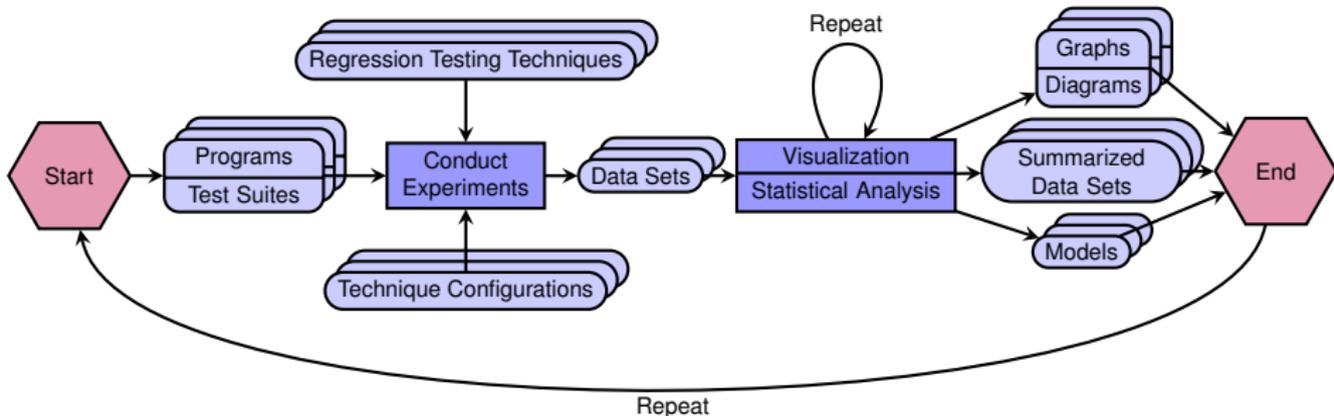
Conduct Experiments with Additional Programs, Test Suites, and Techniques



Few Papers Report on the Efficiency of the Regression Testing Techniques

# Model of Experimental Evaluation

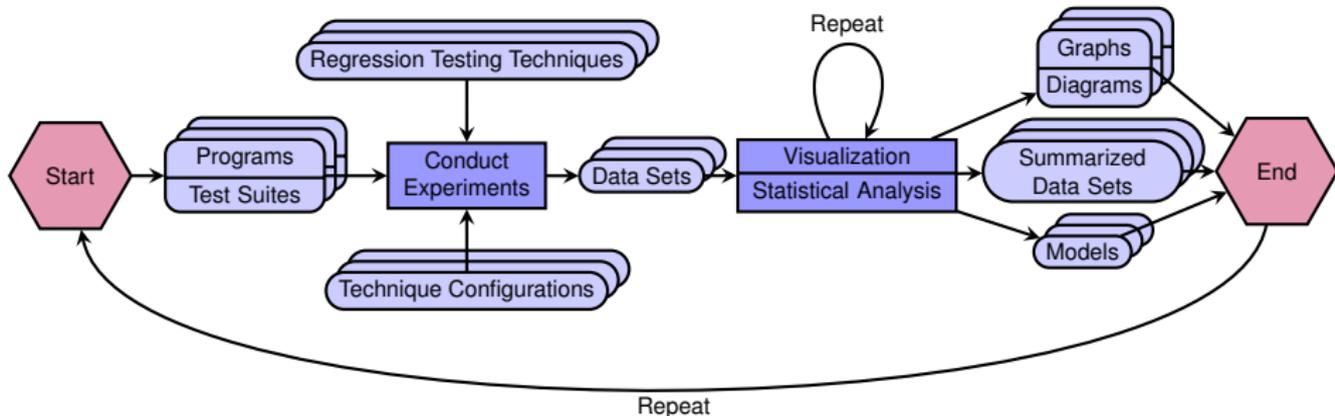
Conduct Experiments with Additional Programs, Test Suites, and Techniques



Authors Do Not Release Tools That Conduct Experiments and Analyze Results

# Model of Experimental Evaluation

Conduct Experiments with Additional Programs, Test Suites, and Techniques



Without Using Data Mining Methods, Researchers May Miss Important Trends

# Thought Provoking Assertions

Jonathan B. Buckheit  
and David L. Donoho  
Department of Statistics  
Stanford University

## Thought Provoking Assertions

For a field to qualify as a science, it is important first and foremost that published work be reproducible by others.

Jonathan B. Buckheit  
and David L. Donoho  
Department of Statistics  
Stanford University

## Thought Provoking Assertions

For a field to qualify as a science, it is important first and foremost that published work be reproducible by others.

Does the Regression Testing Community Want to be Scientific?

Jonathan B. Buckheit  
and David L. Donoho  
Department of Statistics  
Stanford University

# Thought Provoking Assertions

For a field to qualify as a science, it is important first and foremost that published work be reproducible by others.

Does the Regression Testing Community Want to be Scientific?

What Does it Mean for Research to be Reproducible?

Jonathan B. Buckheit  
and David L. Donoho  
Department of Statistics  
Stanford University

# Thought Provoking Assertions

Gary King  
Department of Govern-  
ment Harvard University

## Thought Provoking Assertions

**Reproducible:** sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author.

Gary King  
Department of Government  
Harvard University

## Thought Provoking Assertions

**Reproducible:** sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author.

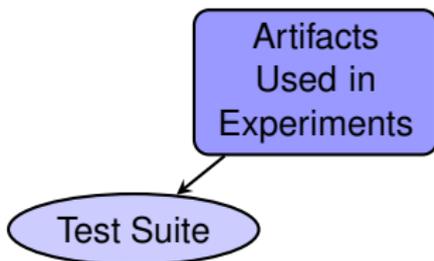
Can the Regression Testing Community Adhere to This Standard?

Gary King  
Department of Government  
Harvard University

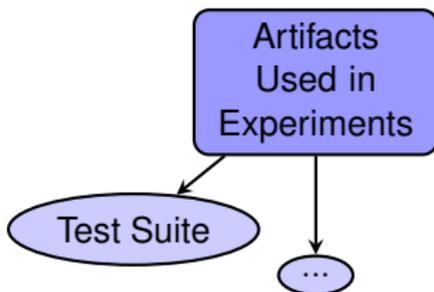
# Reproducible Research Through Sharing

Artifacts  
Used in  
Experiments

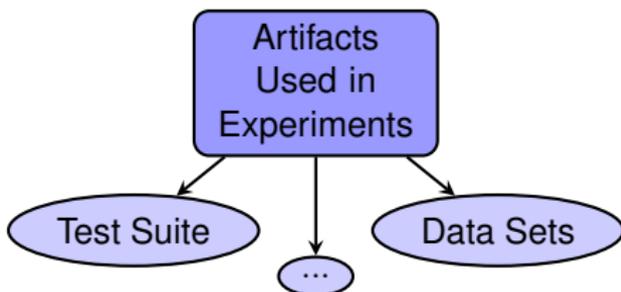
# Reproducible Research Through Sharing



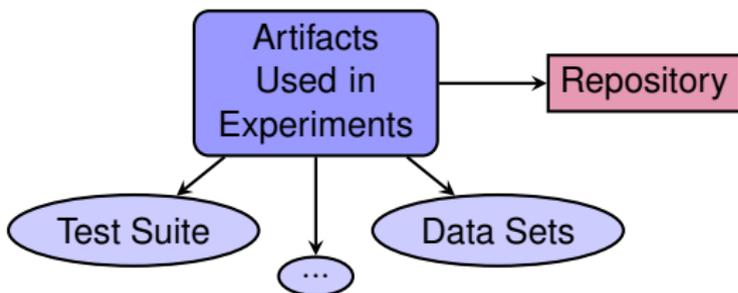
# Reproducible Research Through Sharing



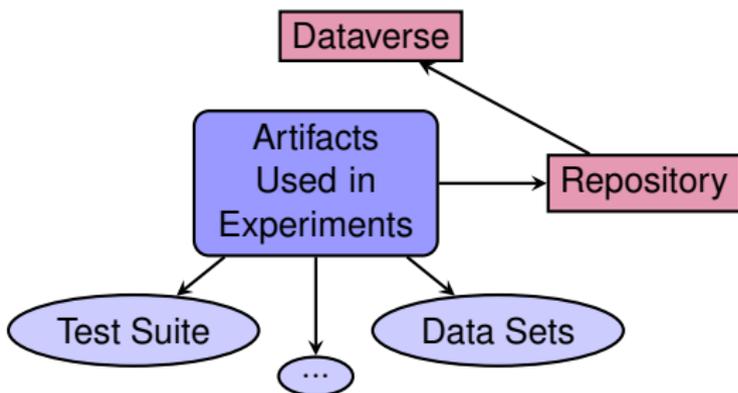
# Reproducible Research Through Sharing



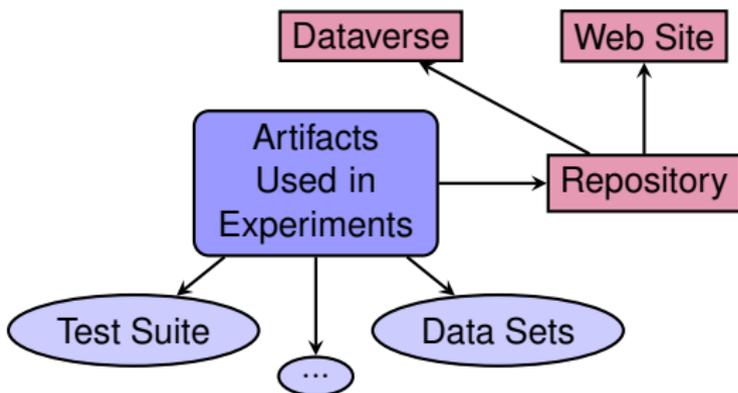
# Reproducible Research Through Sharing



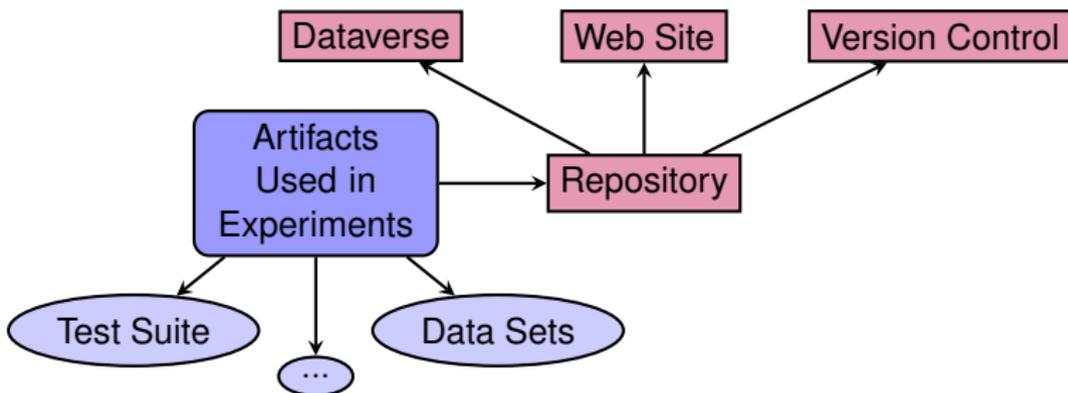
# Reproducible Research Through Sharing



# Reproducible Research Through Sharing

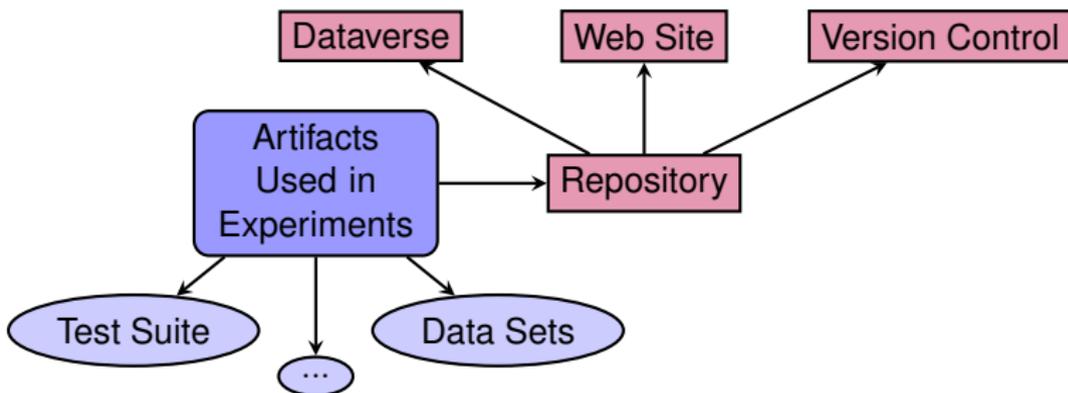


# Reproducible Research Through Sharing



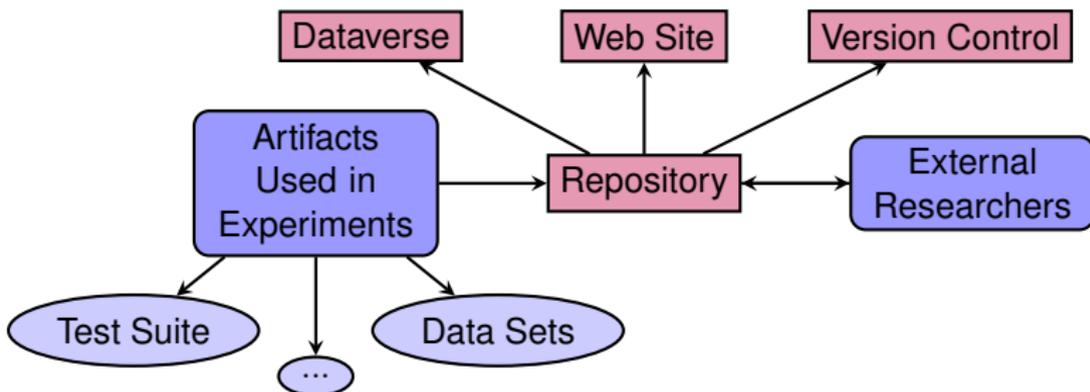
# Reproducible Research Through Sharing

Deposit the Artifacts from Experimentation in One or More Repositories



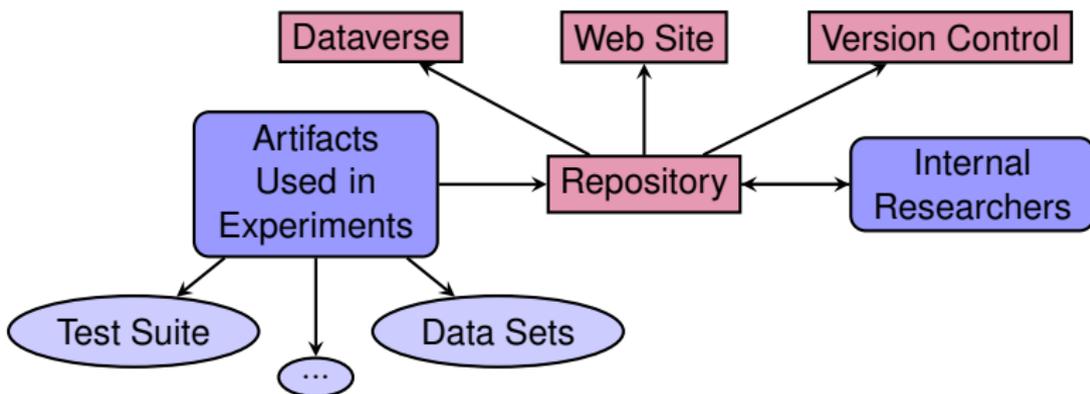
# Reproducible Research Through Sharing

External Researchers Access the Repositories



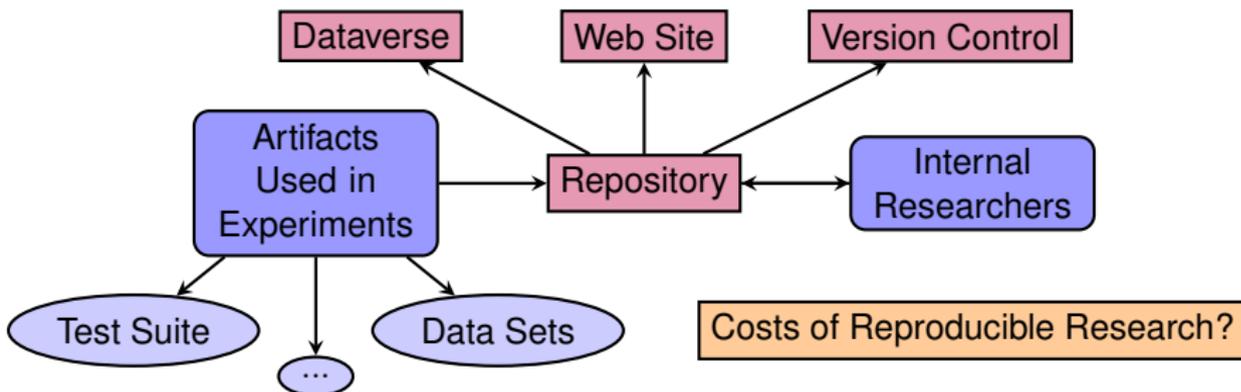
# Reproducible Research Through Sharing

Internal Researchers Access the Repositories



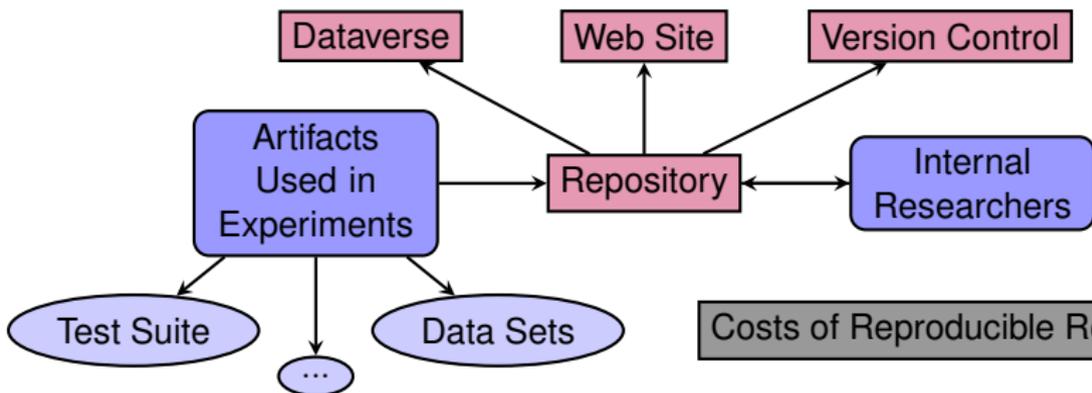
# Reproducible Research Through Sharing

Internal Researchers Access the Repositories



# Reproducible Research Through Sharing

Internal Researchers Access the Repositories

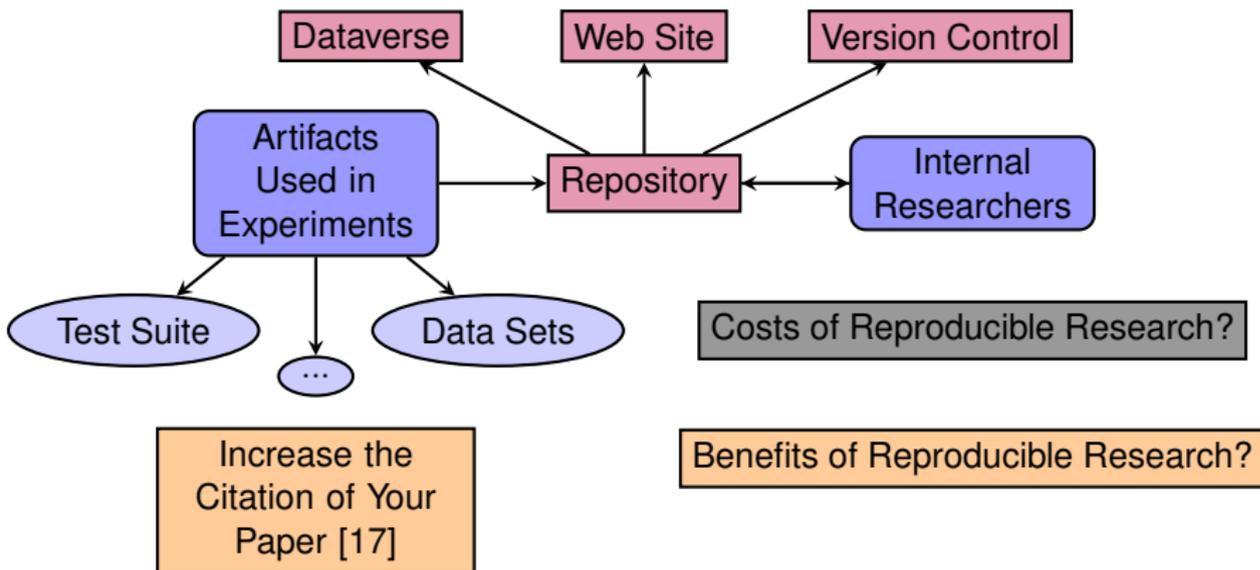


Costs of Reproducible Research?

Benefits of Reproducible Research?

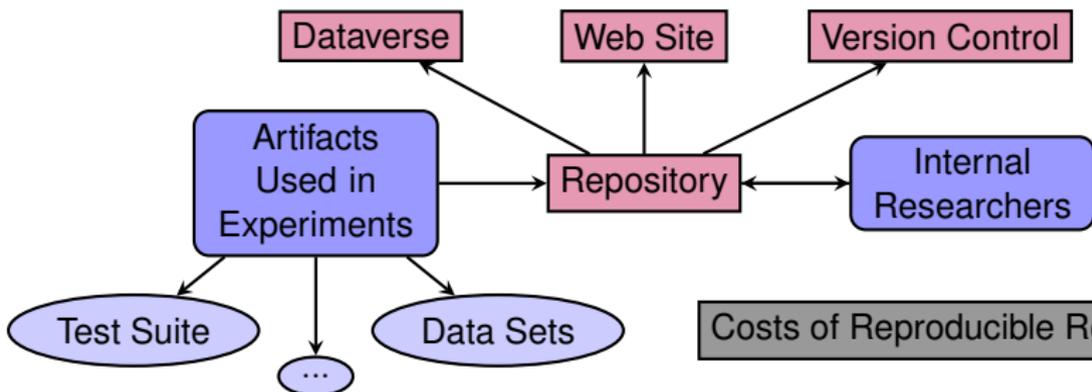
# Reproducible Research Through Sharing

Internal Researchers Access the Repositories



# Reproducible Research Through Sharing

Internal Researchers Access the Repositories



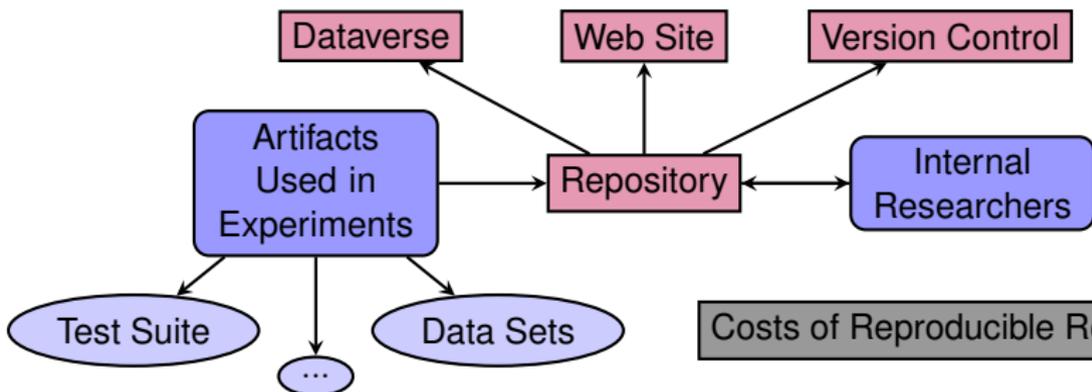
Costs of Reproducible Research?

Allow Others to Replicate Your Experiments [15]

Benefits of Reproducible Research?

# Reproducible Research Through Sharing

Internal Researchers Access the Repositories



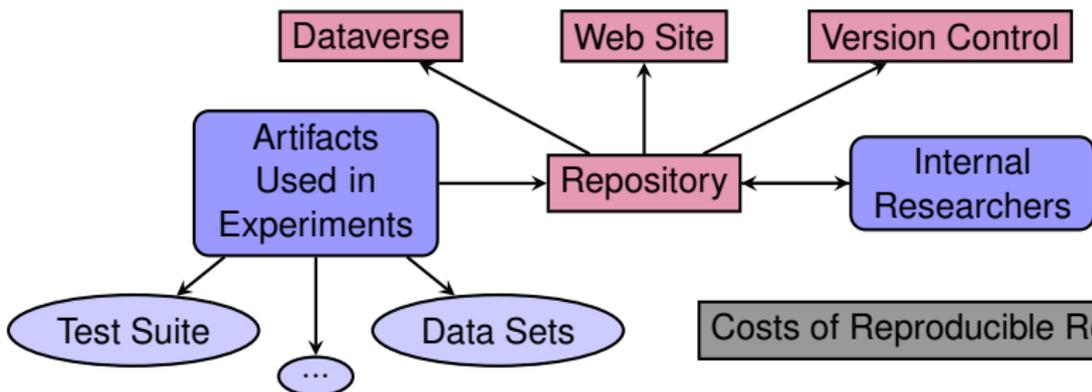
Costs of Reproducible Research?

Lower the Barriers to Entry for New Researchers [16]

Benefits of Reproducible Research?

# Reproducible Research Through Sharing

Internal Researchers Access the Repositories



Costs of Reproducible Research?

Find Your Own Deliverables More Easily [13]

Benefits of Reproducible Research?

# Practical Suggestions: Pick Tools Carefully

# Practical Suggestions: Pick Tools Carefully



Use the R  
Language for  
Statistical Com-  
putation When:

# Practical Suggestions: Pick Tools Carefully



Use the R  
Language for  
Statistical Com-  
putation When:

Conducting  
Experiments

# Practical Suggestions: Pick Tools Carefully



Use the R  
Language for  
Statistical Com-  
putation When:

Conducting  
Experiments

Visualizing Data

# Practical Suggestions: Pick Tools Carefully



Use the R  
Language for  
Statistical Com-  
putation When:

Conducting  
Experiments

Visualizing Data

Data Mining

# Practical Suggestions: Pick Tools Carefully

Simple Shell or Full-Featured IDEs



Use the R Language for Statistical Computation When:

Conducting Experiments

Visualizing Data

Data Mining

# Practical Suggestions: Pick Tools Carefully

Simple Shell or Full-Featured IDEs



Advanced Data Analysis Methods

Use the R Language for Statistical Computation When:

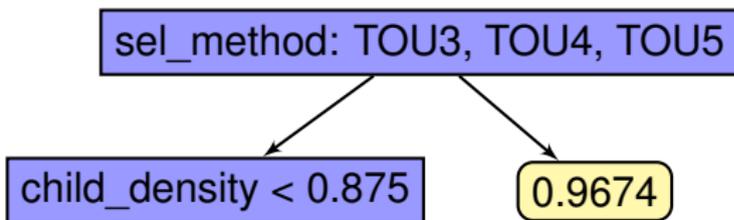
Conducting Experiments

Visualizing Data

Data Mining

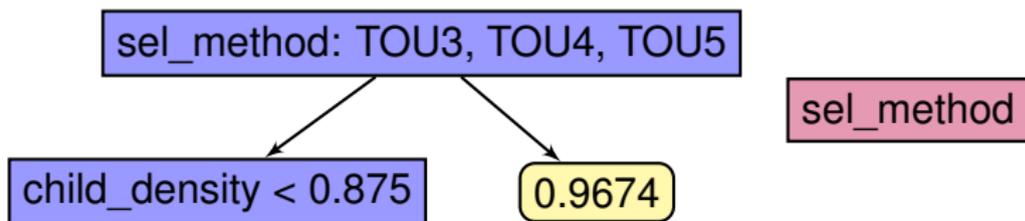
# Practical Suggestions: Use Tree Models

**RM**

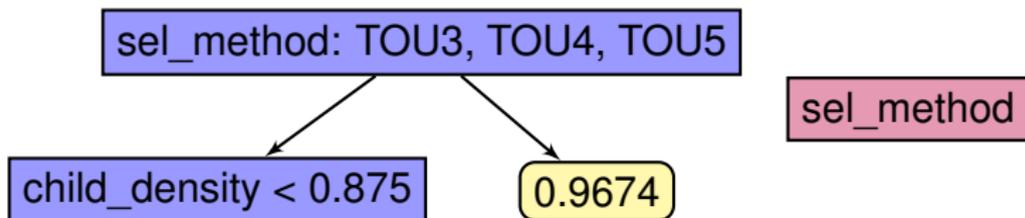
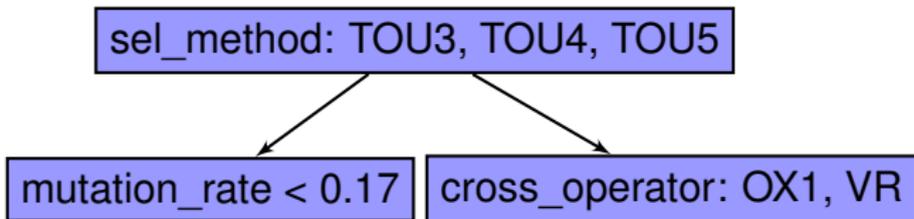


# Practical Suggestions: Use Tree Models

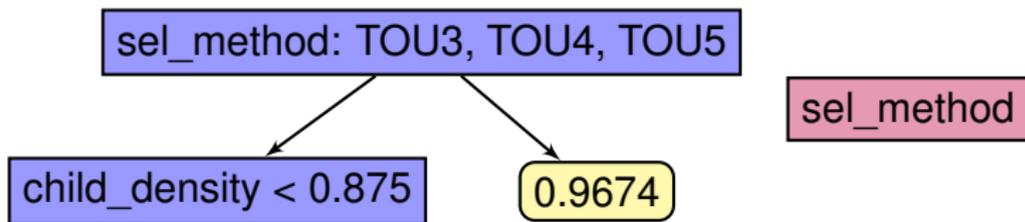
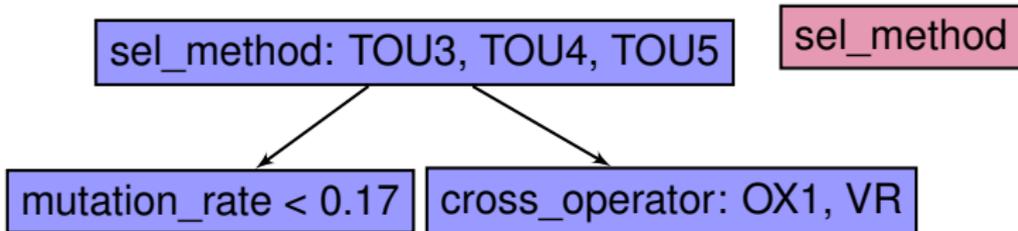
**RM**



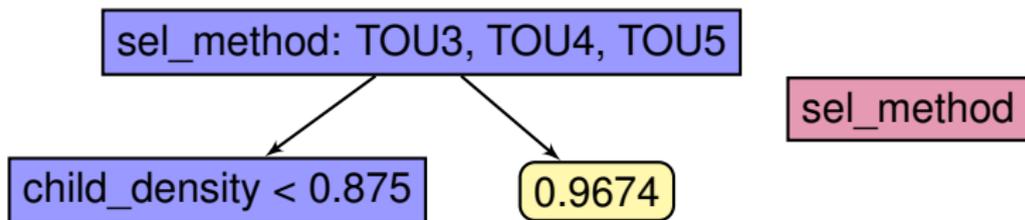
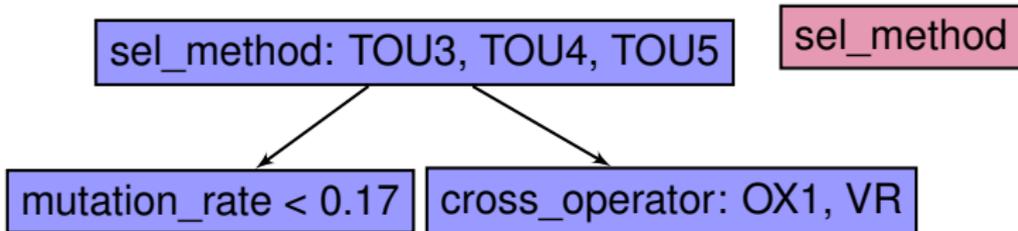
# Practical Suggestions: Use Tree Models

**RM****GB**

# Practical Suggestions: Use Tree Models

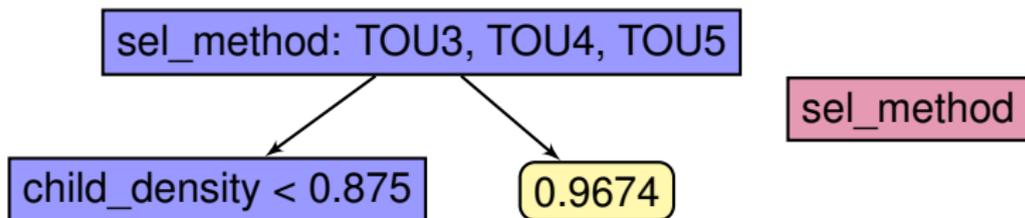
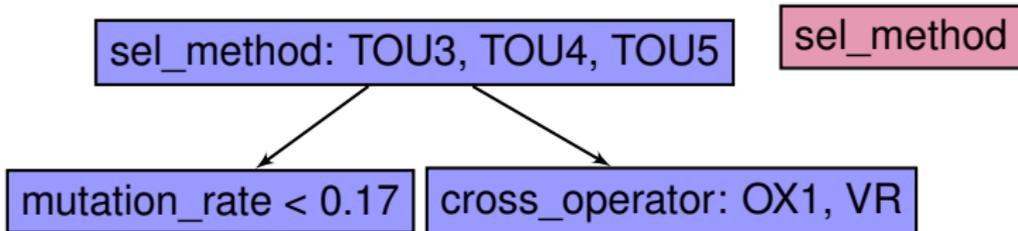
**RM****GB**

## Practical Suggestions: Use Tree Models

**RM****GB**

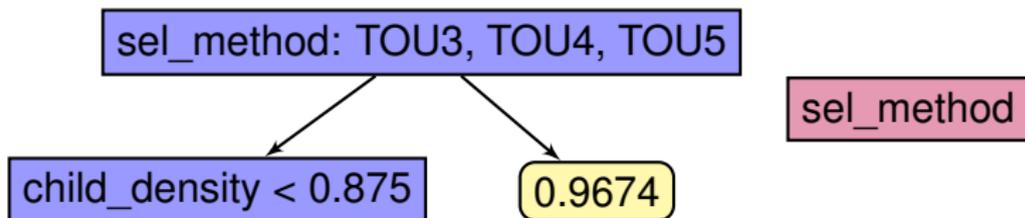
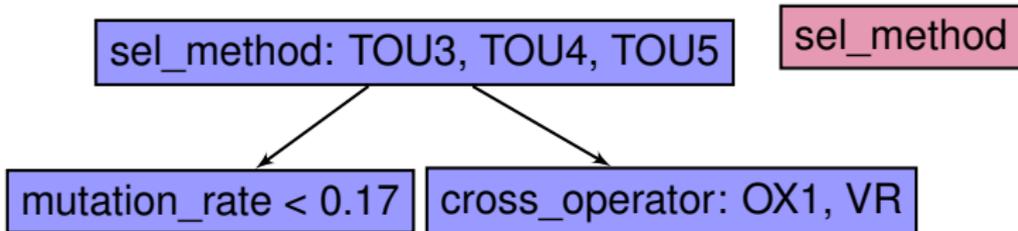
The sel\_method variable is always the most important parameter

# Practical Suggestions: Use Tree Models

**RM****GB**

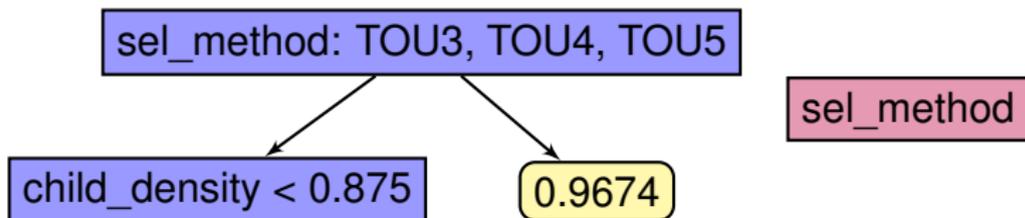
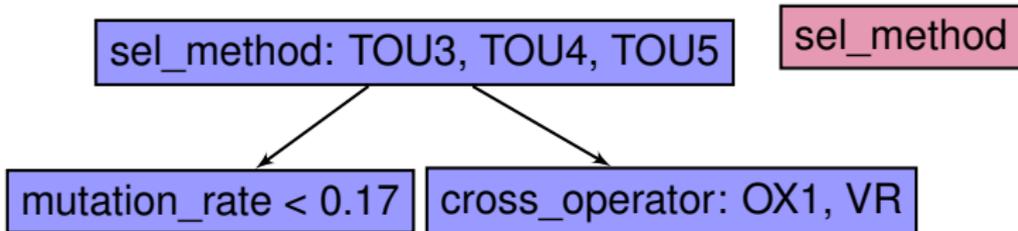
Importance of sel\_method holds for all case study applications

# Practical Suggestions: Use Tree Models

**RM****GB**

How does the selection method impact efficiency and effectiveness?

# Practical Suggestions: Use Tree Models

**RM****GB**

For More Details, Please See Conrad et al. [22]

# Practical Suggestions: Use Open Source



## gelations

Genetic aLgorithm bAsed Test suite prIoritization System

[Project Home](#)[Downloads](#)[Wiki](#)[Issues](#)[Source](#)[Administer](#)[Summary](#)[Updates](#)[People](#)

**Tip:** Project owners, see our [Getting Started](#) guide for steps to configure your project. ✕

### Project Information



Star project

[Activity](#) Low

[Project feeds](#)

#### Code license

[GNU GPL v3](#)

#### Labels

[testing](#), [regression](#), [genetic](#),  
[java](#), [junit](#), [R](#), [evolutionary](#),  
[metahuristic](#)

Gelations is a research prototype system for regression test suite prioritization using genetic algorithms. This system is written entirely in version 1.6 of the Java SE programming language, and is accompanied by its own regression test suite written using the JUnit unit testing framework.

Software testing is a crucial part of the software development lifecycle. Regression testing is a form of testing in which all of the old test cases written to cover different parts of a program are combined into a single test suite and executed. This form of testing helps to reveal regressions, or instances in which code that had formerly functioned correctly is broken by later changes to the system. For real-world applications, however, regression test suites can take days or even weeks to execute. One solution to this problem of execution time overhead is to reduce the suite, removing test cases that are redundant or unlikely to detect faults. This approach, however, can compromise the ability of a suite to detect faults. Another approach to this problem is test suite prioritization. Prioritization does not reduce the total

# Practical Suggestions: Use Open Source



## gelations

Genetic aLgorithm bAsed Test suite prOrtizatioN System

[Project Home](#)[Downloads](#)[Wiki](#)[Issues](#)[Source](#)[Administer](#)[Summary](#)[Updates](#)[People](#)

**Tip:** Project owners, see our [Getting Started](#) guide for steps to configure your project. ✕

### Project Information



Star project

[Activity](#) Low

[Project feeds](#)

#### Code license

[GNU GPL v3](#)

#### Labels

testing, regression, genetic,  
java, junit, R, evolutionary,  
metahuristic

Gelations is a research prototype system for regression test suite prioritization using genetic algorithms. This system is written entirely in version 1.6 of the Java SE programming language, and is accompanied by its own regression test suite written using the JUnit unit testing framework.

Software testing is a crucial part of the software development lifecycle. Regression testing is a form of testing in which all of the old test cases written to cover different parts of a program are combined into a single test suite and executed. This form of testing helps to reveal regressions, or instances in which code that had formerly functioned correctly is broken by later changes to the system. For real-world applications, however, regression test suites can take days or even weeks to execute. One solution to this problem of execution time overhead is to reduce the suite, removing test cases that are redundant or unlikely to detect faults. This approach, however, can compromise the ability of a suite to detect faults. Another approach to this problem is test suite prioritization. Prioritization does not reduce the total

Gelations is a Genetic Algorithm-Based Test Suite Prioritizer

# Practical Suggestions: Use Open Source



## gelations

Genetic aLgorithm bAsed Test suite prOrtizatioN System

[Project Home](#) [Downloads](#) [Wiki](#) [Issues](#) [Source](#) [Administer](#)

[Summary](#) [Updates](#) [People](#)

**Tip:** Project owners, see our [Getting Started](#) guide for steps to configure your project.

### Project Information

★ Star project  
[Activity](#)  Low  
[Project feeds](#)

**Code license**  
[GNU GPL v3](#)

**Labels**  
[testing](#), [regression](#), [genetic](#),  
[java](#), [junit](#), [R](#), [evolutionary](#),  
[metahuristic](#)

Gelations is a research prototype system for regression test suite prioritization using genetic algorithms. This system is written entirely in version 1.6 of the Java SE programming language, and is accompanied by its own regression test suite written using the JUnit unit testing framework.

Software testing is a crucial part of the software development lifecycle. Regression testing is a form of testing in which all of the old test cases written to cover different parts of a program are combined into a single test suite and executed. This form of testing helps to reveal regressions, or instances in which code that had formerly functioned correctly is broken by later changes to the system. For real-world applications, however, regression test suites can take days or even weeks to execute. One solution to this problem of execution time overhead is to reduce the suite, removing test cases that are redundant or unlikely to detect faults. This approach, however, can compromise the ability of a suite to detect faults. Another approach to this problem is test suite prioritization. Prioritization does not reduce the total

Visit <http://gelations.googlecode.com/> for More Details

# Practical Suggestions: Use Open Source



## gelations

Genetic aLgorithm bAsed Test suite prIOritization System


[Project Home](#)
[Downloads](#)
[Wiki](#)
[Issues](#)
[Source](#)
[Administer](#)
[Summary](#)
[Updates](#)
[People](#)

**Tip:** Project owners, see our [Getting Started](#) guide for steps to configure your project. ✕

### Project Information



Star project

[Activity](#)  Low

[Project feeds](#)

#### Code license

[GNU GPL v3](#)

#### Labels

testing, regression, genetic,  
java, junit, R, evolutionary,  
metahuristic

Gelations is a research prototype system for regression test suite prioritization using genetic algorithms. This system is written entirely in version 1.6 of the Java SE programming language, and is accompanied by its own regression test suite written using the JUnit unit testing framework.

Software testing is a crucial part of the software development lifecycle. Regression testing is a form of testing in which all of the old test cases written to cover different parts of a program are combined into a single test suite and executed. This form of testing helps to reveal regressions, or instances in which code that had formerly functioned correctly is broken by later changes to the system. For real-world applications, however, regression test suites can take days or even weeks to execute. One solution to this problem of execution time overhead is to reduce the suite, removing test cases that are redundant or unlikely to detect faults. This approach, however, can compromise the ability of a suite to detect faults. Another approach to this problem is test suite prioritization. Prioritization does not reduce the total

Visit <http://raise.googlecode.com/> for Greedy Algorithms

# Conclusions and Future Work

The Potential Way  
Forward May  
Seem *Daunting*

# Conclusions and Future Work

The Potential Way Forward May Seem *Daunting*

Once you get your courage up and believe that you can do important problems, then you can.

Richard Hamming

# Conclusions and Future Work

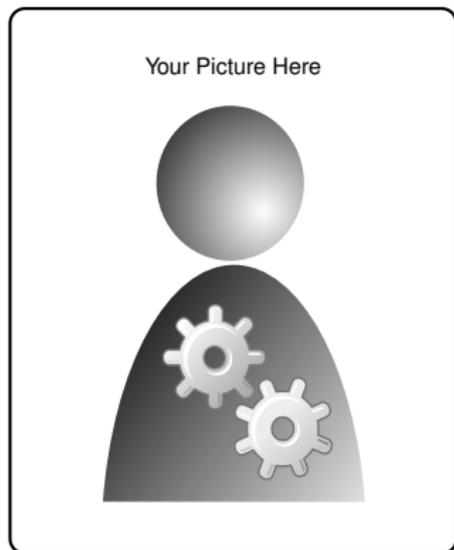
The Potential Way Forward May Seem *Daunting*

Once you get your courage up and believe that you can do important problems, then you can.

Richard Hamming

Share *One* Item With Your Next Paper

# Conclusions and Future Work



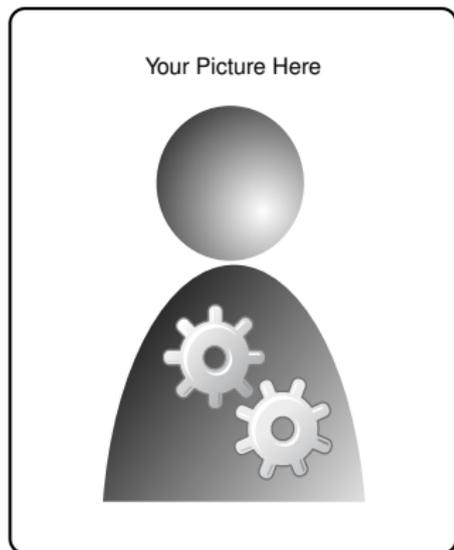
Leader of the Regression Testing Community

Your Help is Welcomed and Appreciated!

Once you get your courage up and believe that you can do important problems, then you can.

Richard Hamming

# Conclusions and Future Work



Leader of the Regression Testing Community

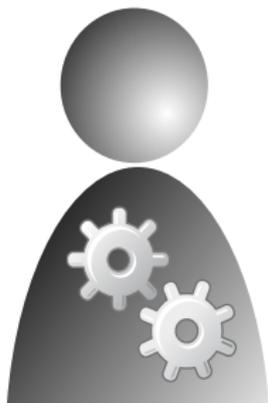
Read the Paper  
and Contact Me  
with Comments

Once you get your  
courage up and  
believe that you can  
do important prob-  
lems, then you can.

Richard Hamming

# Conclusions and Future Work

Your Picture Here



Leader of the Regression Testing Community

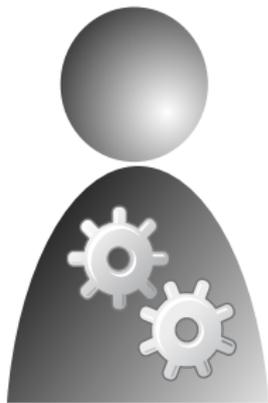
Read the Paper  
and Contact Me  
with Comments

**Future Work:**  
Complete Case Study of  
Reproducible Research  
in Regression Testing

- ✓ Data Sets
- ✓ Tools
- ✓ Visualizations

# Conclusions and Future Work

Your Picture Here



Leader of the Regression Testing Community

Read the Paper  
and Contact Me  
with Comments

## Acknowledgments:

- ✓ Neil Elliott
- ✓ Jonathan Kauffman
- ✓ Molly Mattis
- ✓ Zachary Williams provided feedback on drafts of the paper