Retrospective Mutant Reduction:

Empirically evaluating a family of integrated techniques

Colton J. McCurdy

mccurdyc

Allegheny College

April 24, 2017

Introduction

Reduce

Faults

Introduction

Regression Test Suite

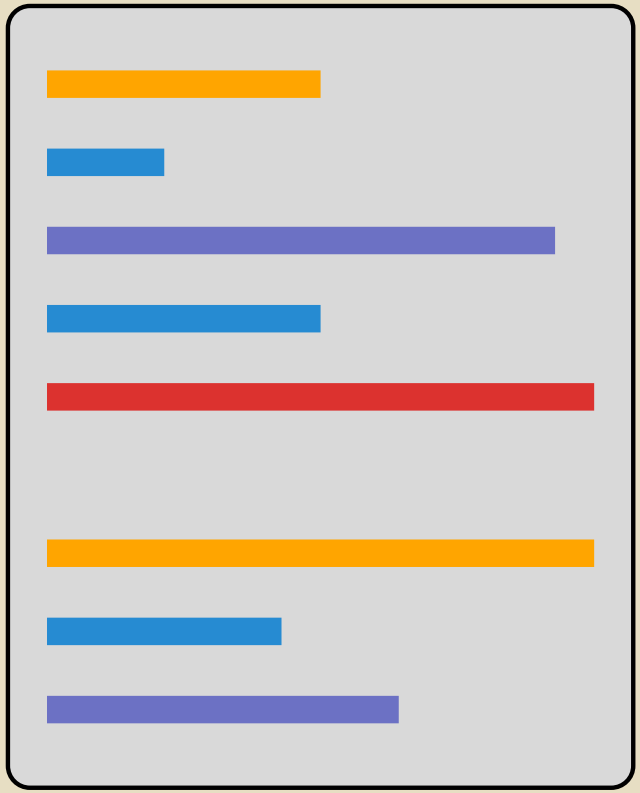
Introduction

Regression Test Suite

$$T = \langle\ t_1, t_2, \ldots, t_n\ \rangle$$

Introduction

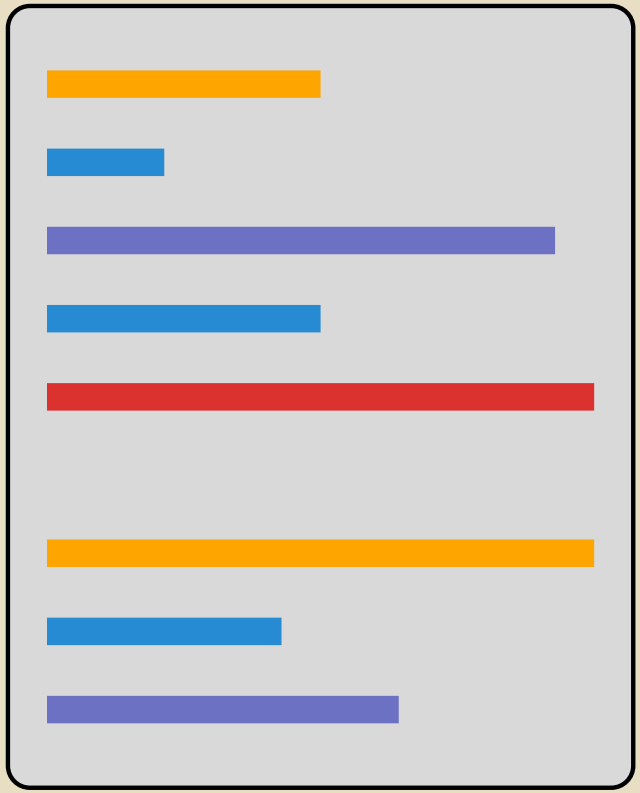
Pass



Fail

Introduction

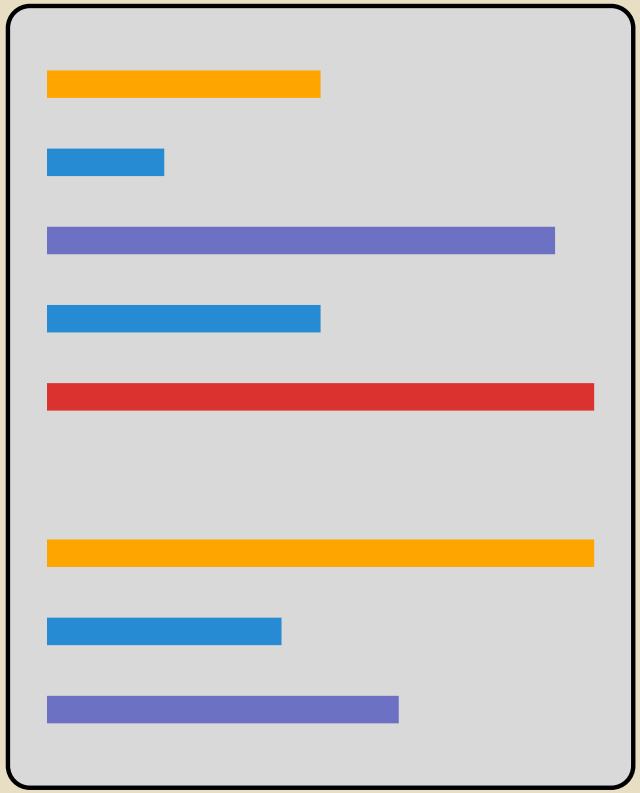
✓Pass



Fail

Introduction

✓Pass



✗Fail

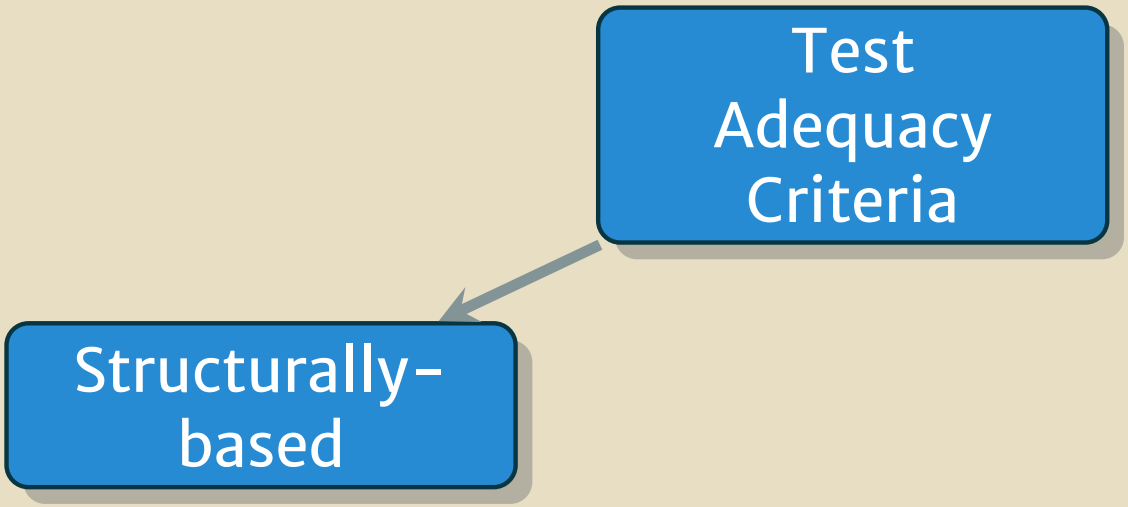
Introduction

Test Suite Adequacy?

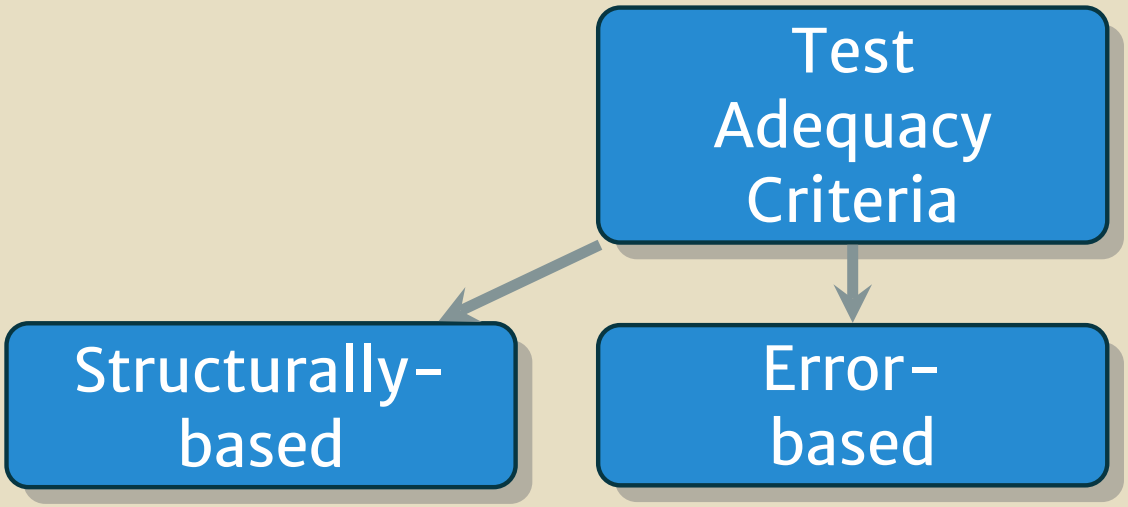
Introduction

Test
Adequacy
Criteria

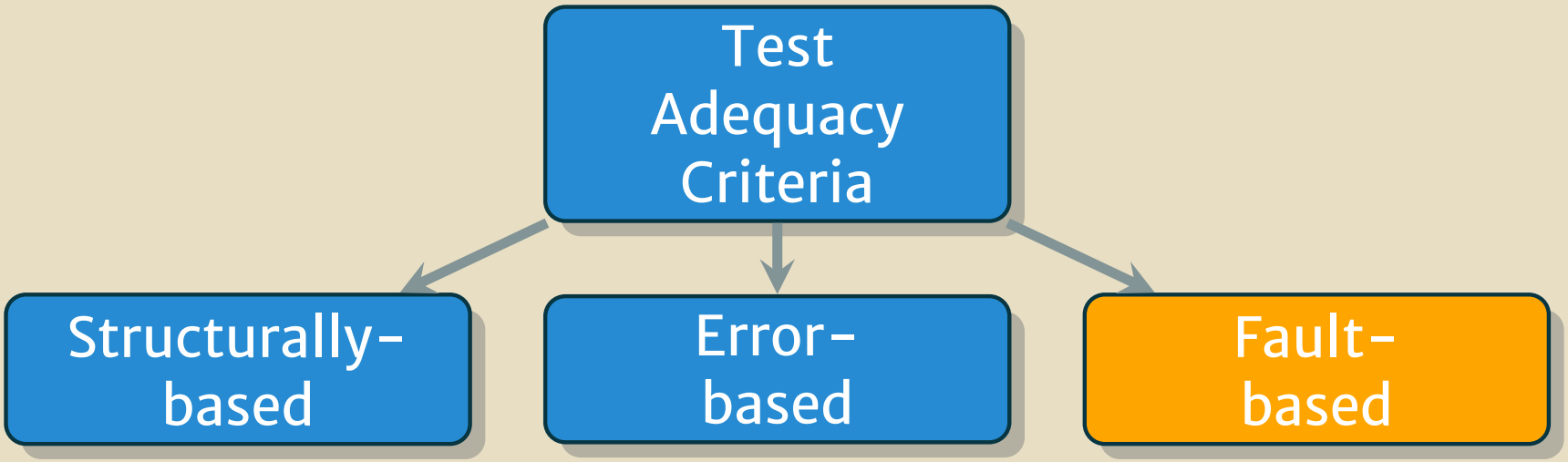
Introduction



Introduction



Introduction



Background

Why Fault-based?

Background

Why Fault-based?

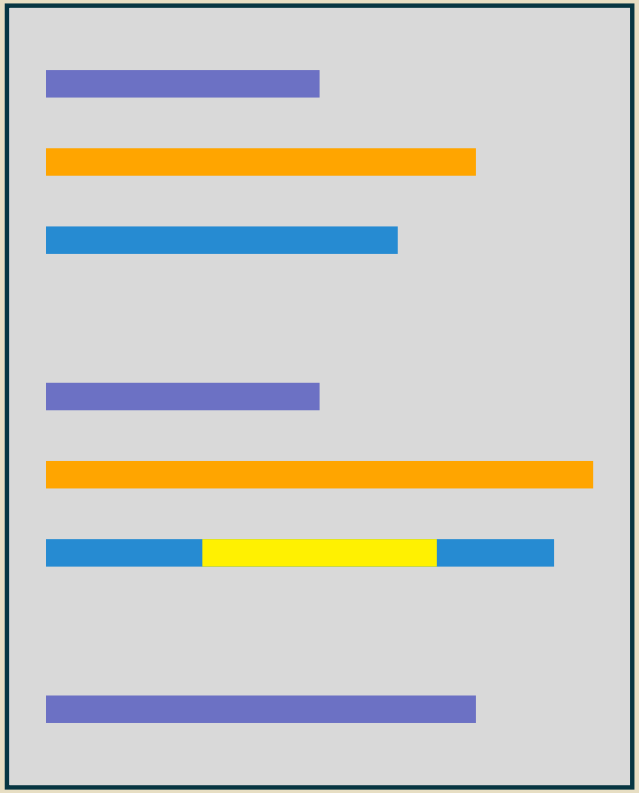
Simulate real-world faults by programmers!

Background

Mutation Testing!

Mutation Testing

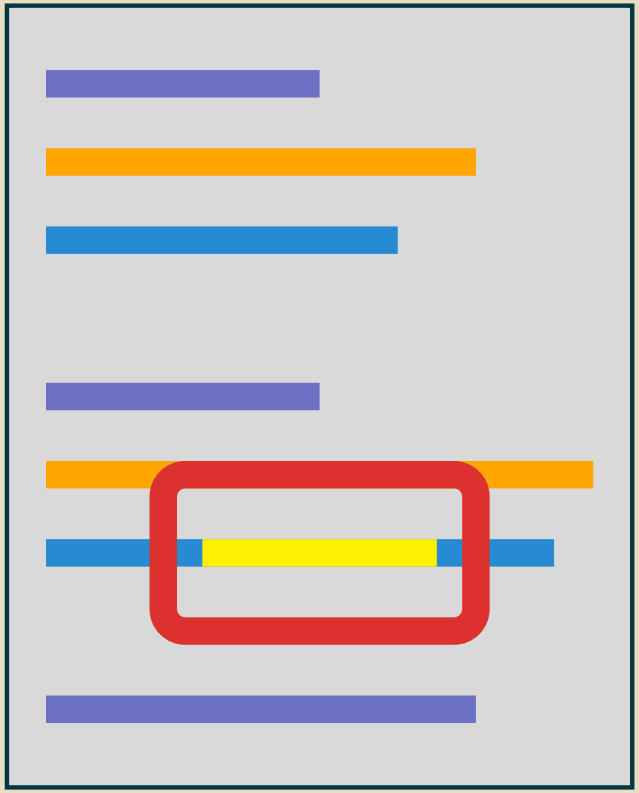
Original



Mutant

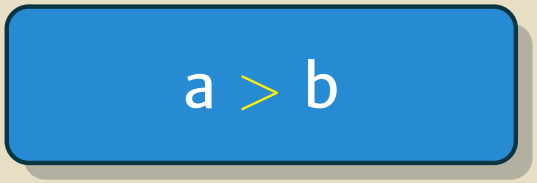
Mutation Testing

Original

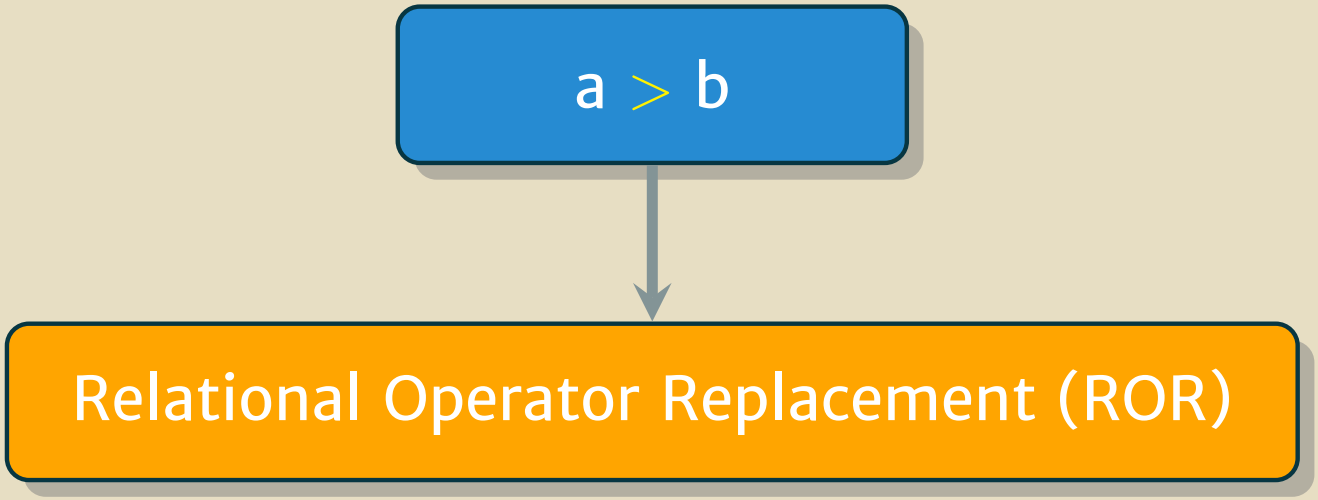


Mutant

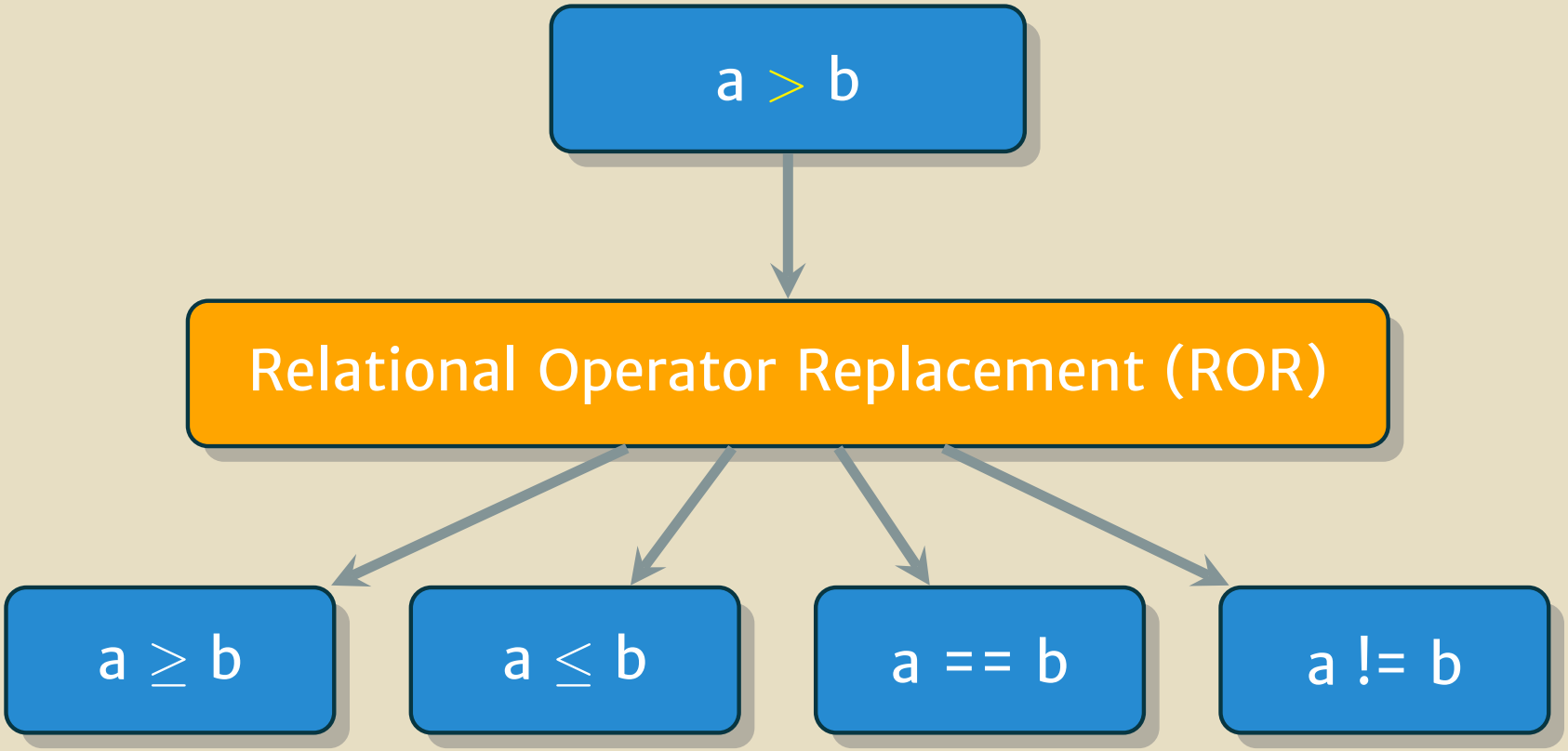
Mutation Testing



Mutation Testing

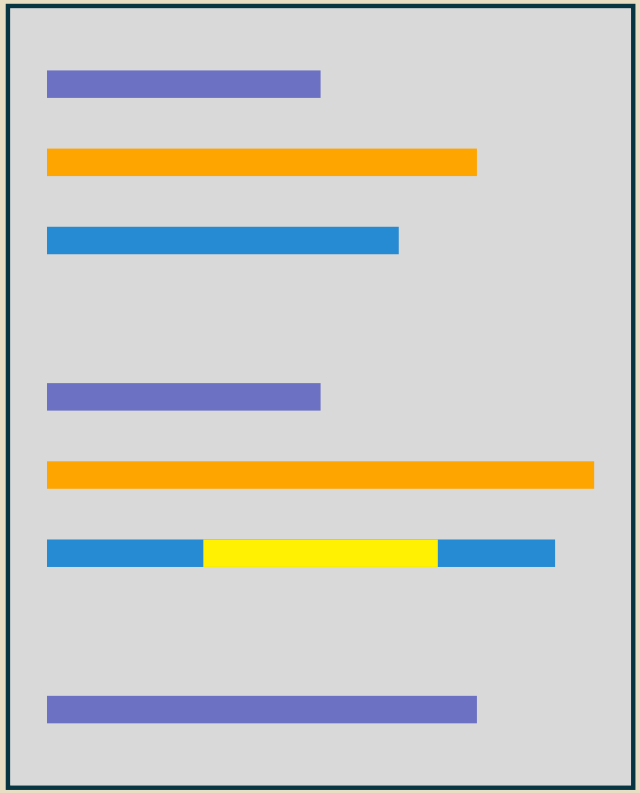


Mutation Testing



Mutation Testing

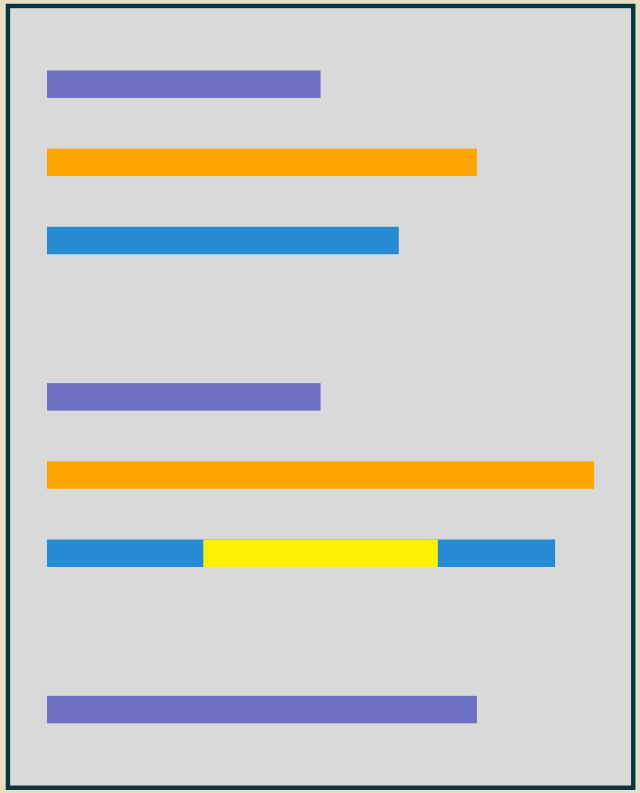
Alive



Killed

Mutation Testing

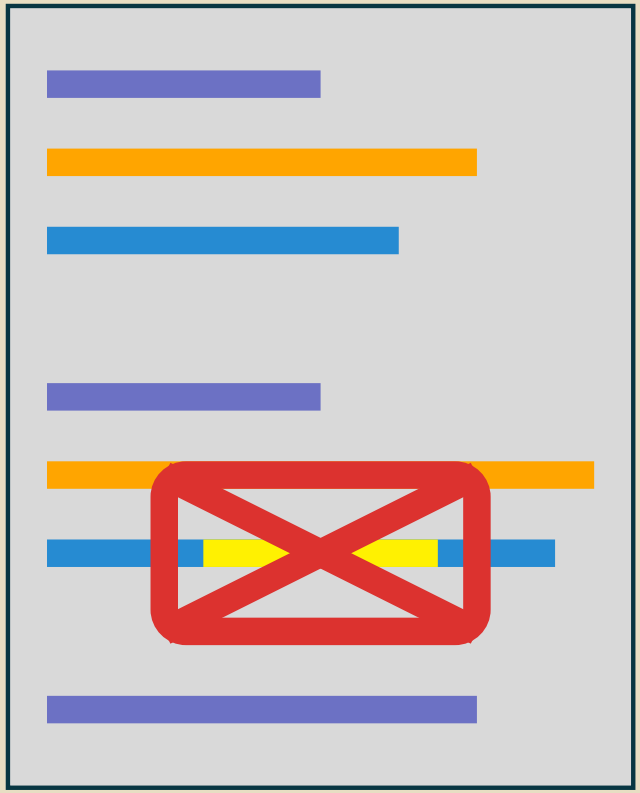
✓Alive



Killed

Mutation Testing

✓Alive



✗Killed

Mutation Testing

$$MS_T = \frac{Killed}{Total}$$

Mutation Testing

$$MS_T = \frac{Killed}{Total}$$

Mutation Testing

$$MS_T = \frac{Killed}{Total}$$

Mutation Testing

$$MS_T = \frac{Killed}{Total}$$

$MS_T \in [0, 1]$

Mutation Testing

$$MS_T = \frac{Killed}{Total}$$

$MS_T \in [0, 1]$

HIB

Mutation Testing

Major Limitations

Mutation Testing

Mutation Testing is Time Consuming

Mutation Testing

Large

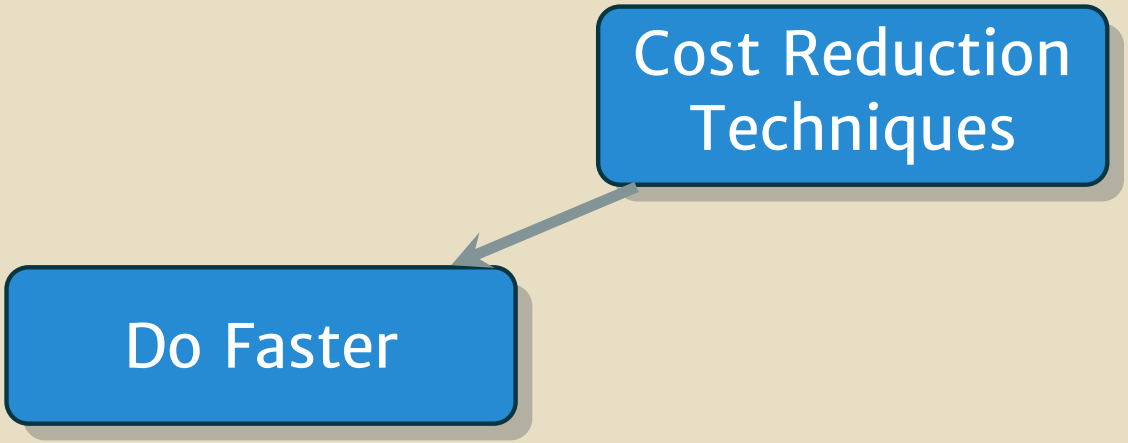
Number of

Mutants

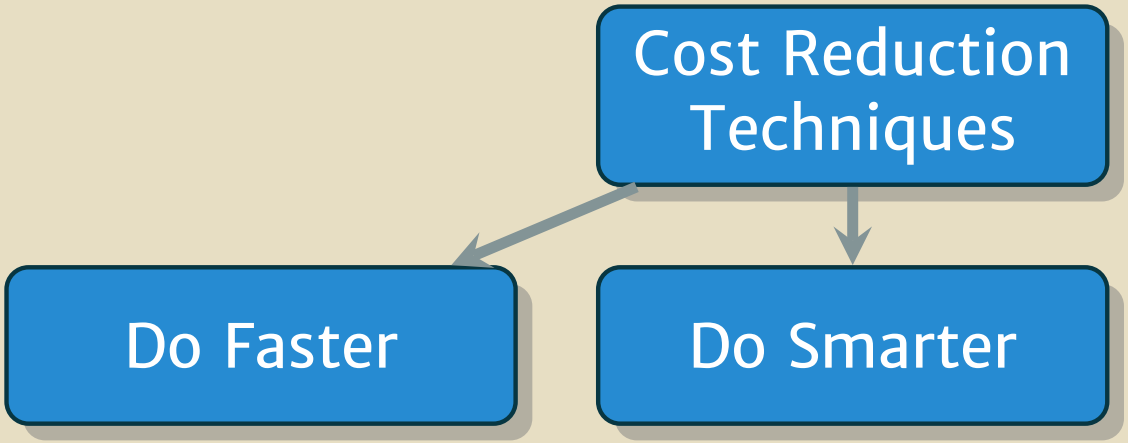
Reduction Techniques

Cost Reduction Techniques

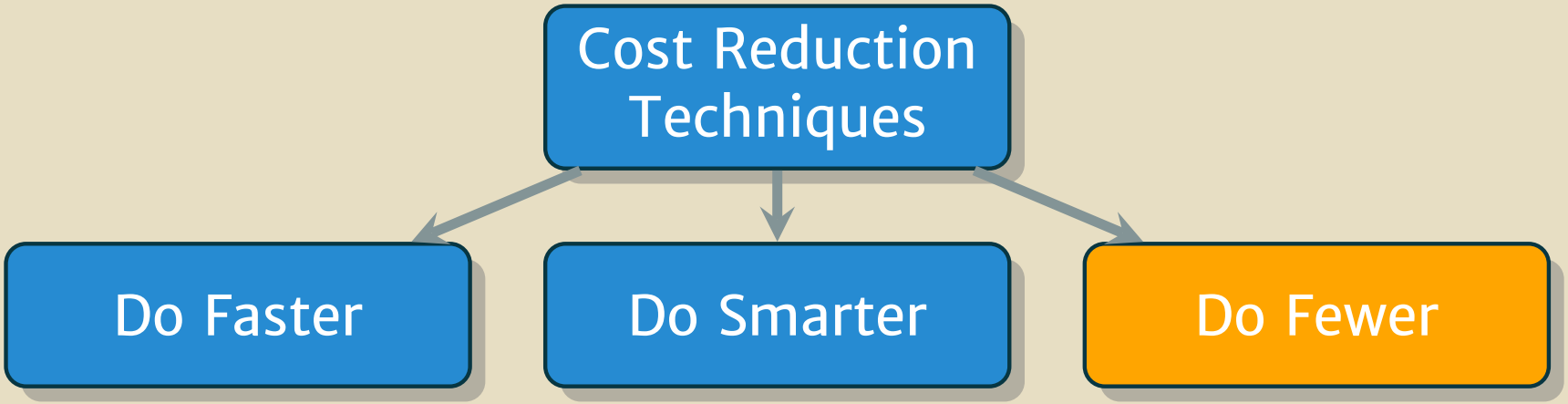
Reduction Techniques



Reduction Techniques



Reduction Techniques



Reduction Techniques

Reduce the Number of Mutants!

Reduction Techniques

Mutant Reduction is Not New

Reduction Techniques

Random Sampling

Reduction Techniques

Random Sampling

Operator Selection

Reduction Techniques

- Random Sampling
- Operator Selection
- Random Sampling Over Operators

Reduction Techniques

- Random Sampling
- Operator Selection
- Random Sampling Over Operators
- ...

Reduction Techniques

"... none of them are superior to random mutant-selection techniques ... [1]"

- Random Sampling
- Operator Selection
- Random Sampling Over Operators
- ... ✗

Reduction Techniques

"... none of them are superior to random mutant-selection techniques ... [1]"

"'... random sampling performs better in predicting final mutation score than operator selection [2]"

- Random Sampling
- Operator Selection
- Random Sampling Over Operators ✗
- ... ✗

Reduction Techniques

"... none of them are superior to random mutant-selection techniques ... [1]"

"... random sampling performs better in predicting final mutation score than operator selection [2]"

"... none of the mutation reduction strategies provide a practical large benefit over the baseline random sampling ... they likely do not provide enough benefit to justify the additional complexity [3]"

- Random Sampling
- Operator Selection ✗
- Random Sampling Over Operators ✗
- ... ✗

Reduction Techniques

"... none of them are superior to random mutant-selection techniques ... [1]"

"... random sampling performs better in predicting final mutation score than operator selection [2]"

"... none of the mutation reduction strategies provide a practical large benefit over the baseline random sampling ... they likely do not provide enough benefit to justify the additional complexity [3]"

Random Sampling	✓
Operator Selection	✗
Random Sampling Over Operators	✗
...	✗

Reduction Techniques

Representative Reduced Sets

Reduction Techniques

Correlated Mutation Scores

Reduction Techniques

Current Approach is Challenging

Reduction Techniques

Understand system intricacies

Reduction Techniques

Understand system intricacies

Reduction Techniques

PIT has over 46,000 source lines of code

Retrospective Analysis

Retrospective Mutant Reduction

Retrospective Analysis

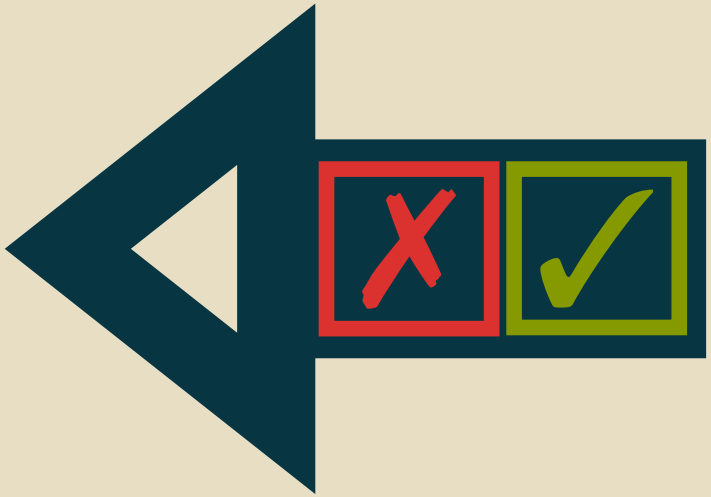
Retrospective Mutant Reduction

Evaluate before you integrate!

Retrospective Analysis

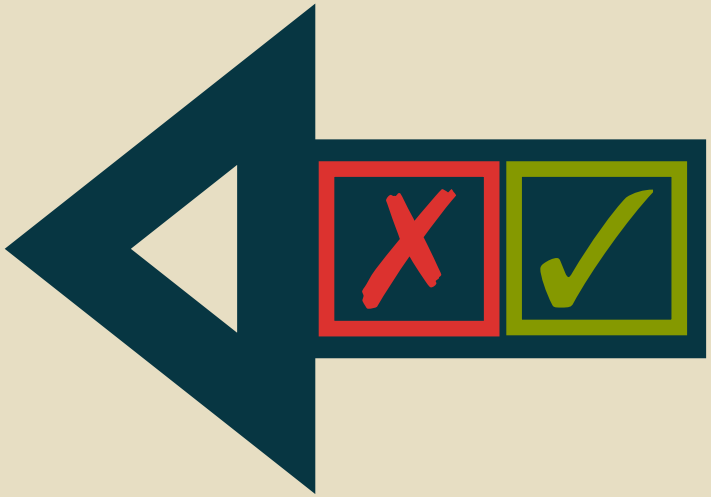
Reduce Before Mutation Testing?

Retrospective Analysis



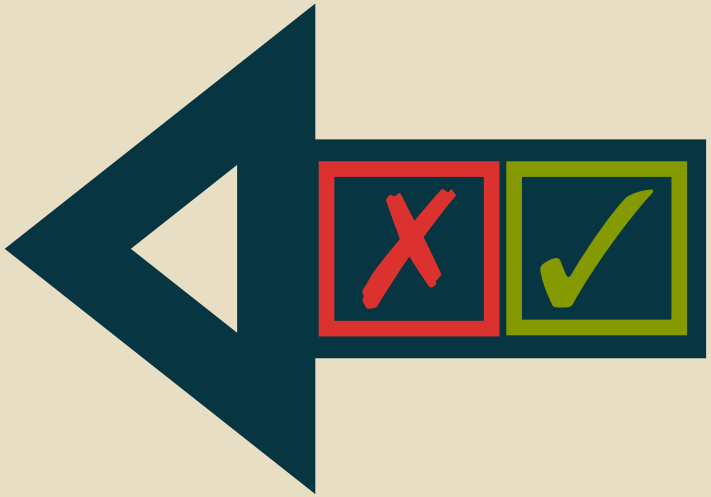
Retrospective Analysis

mccurdyc/mrstudyr



Retrospective Analysis

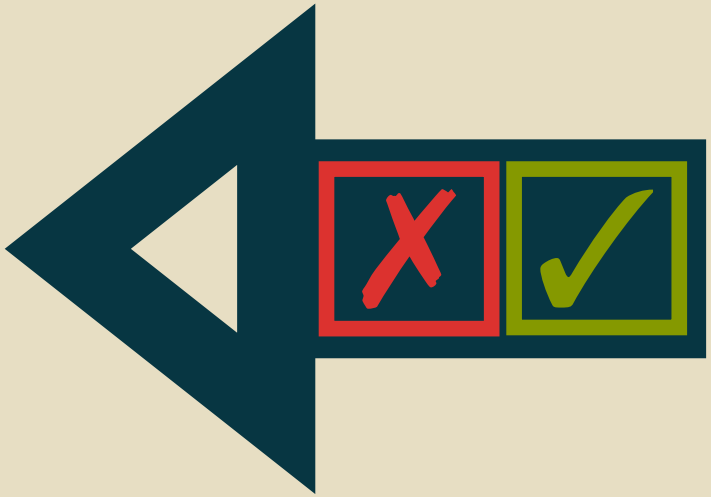
mccurdyc/mrstudyr



Data

Retrospective Analysis

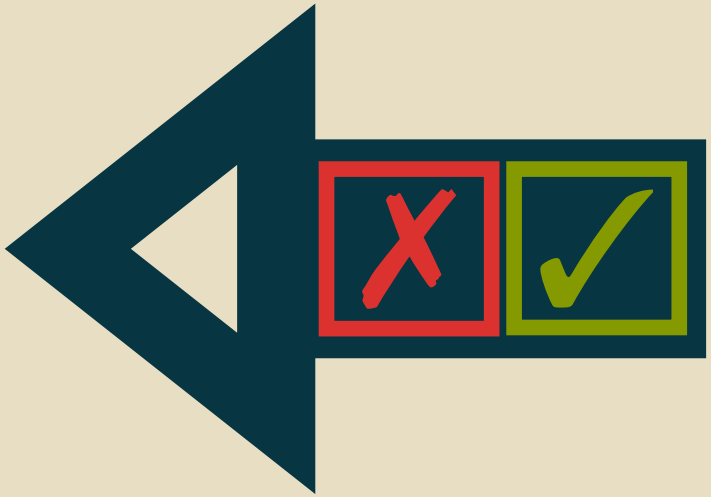
mccurdyc/mrstudyr



Data → Reduce

Retrospective Analysis

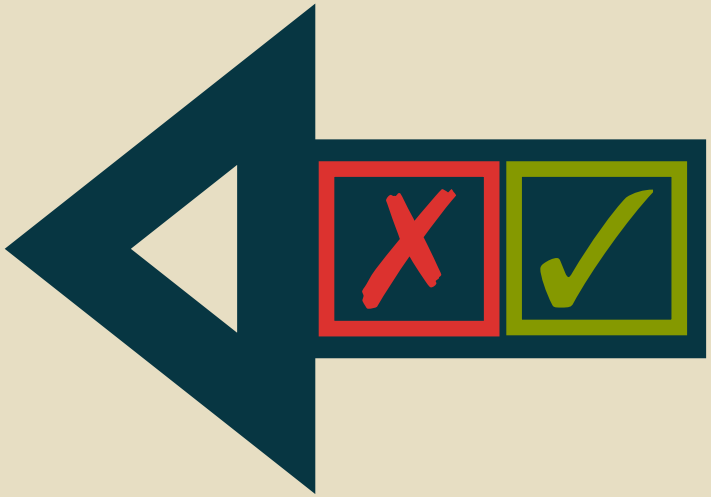
mccurdyc/mrstudyr



Data → Reduce → Model

Retrospective Analysis

mccurdyc/mrstudyr



Leveraged existing and implemented new techniques.

Research Questions

1. Reducing Database Schema Mutants?

Research Questions

2. SBSE Technique?

Schema Testing

```
1 CREATE TABLE t (
2   x INT,
3   y INT,
4   PRIMARY KEY(x)
5 );
```

Original Schema

Schema Testing

```
1 CREATE TABLE t (
2   x INT,
3   y INT,
4   PRIMARY KEY(x)
5 );
```

Original Schema

```
1 CREATE TABLE t (
2   x INT,
3   y INT,
4   PRIMARY KEY(x, y)
5 );
```

Mutant Schema

Schema Testing

Schema Test Suite?

Schema Testing

Database Schema
Test Suite

$$T = \langle\ i_1, i_2, \ldots, i_n\ \rangle$$

Schema Testing

Manually Writing Tests is Challenging

Schema Testing

Schema Testing

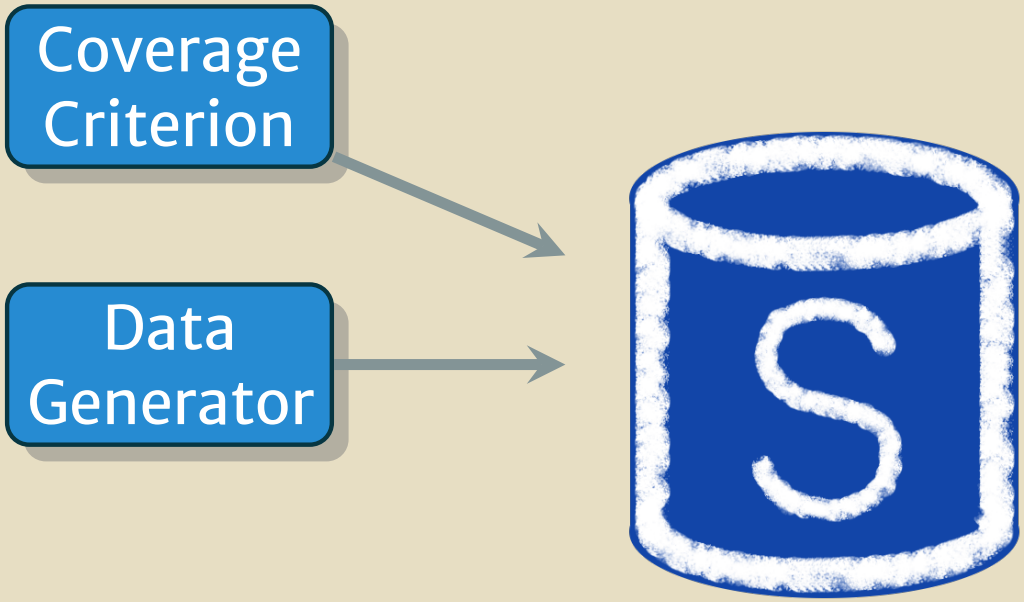
Extensible Tool for Test Data Deneration

Schema Testing

Coverage Criterion

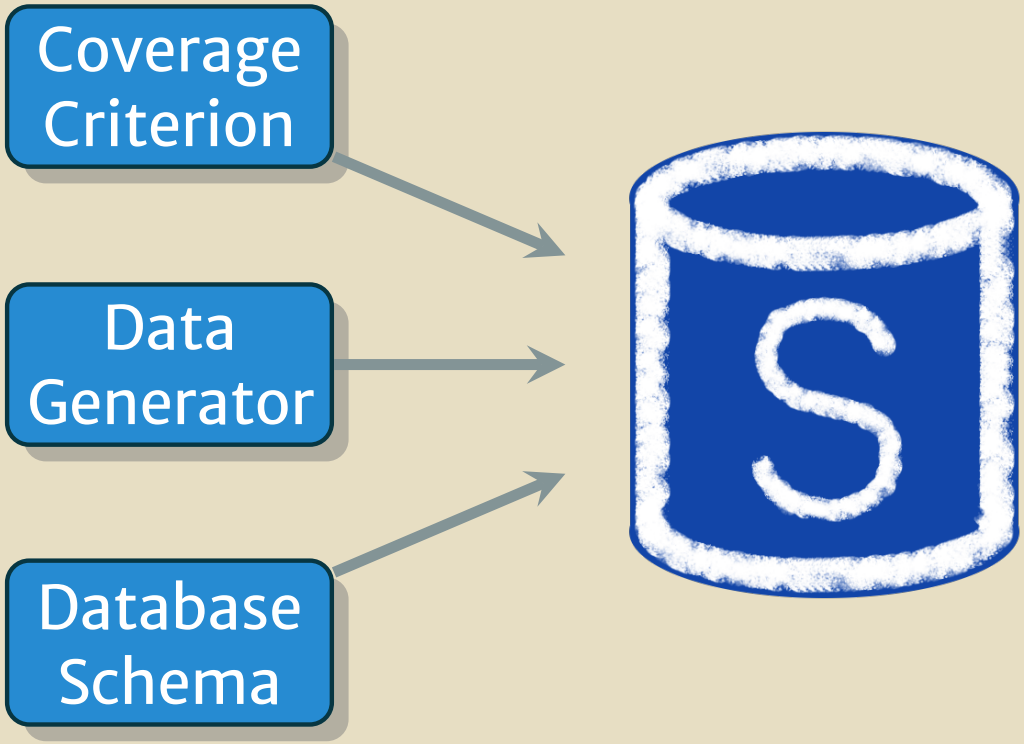
Extensible Tool for Test Data Deneration

Schema Testing



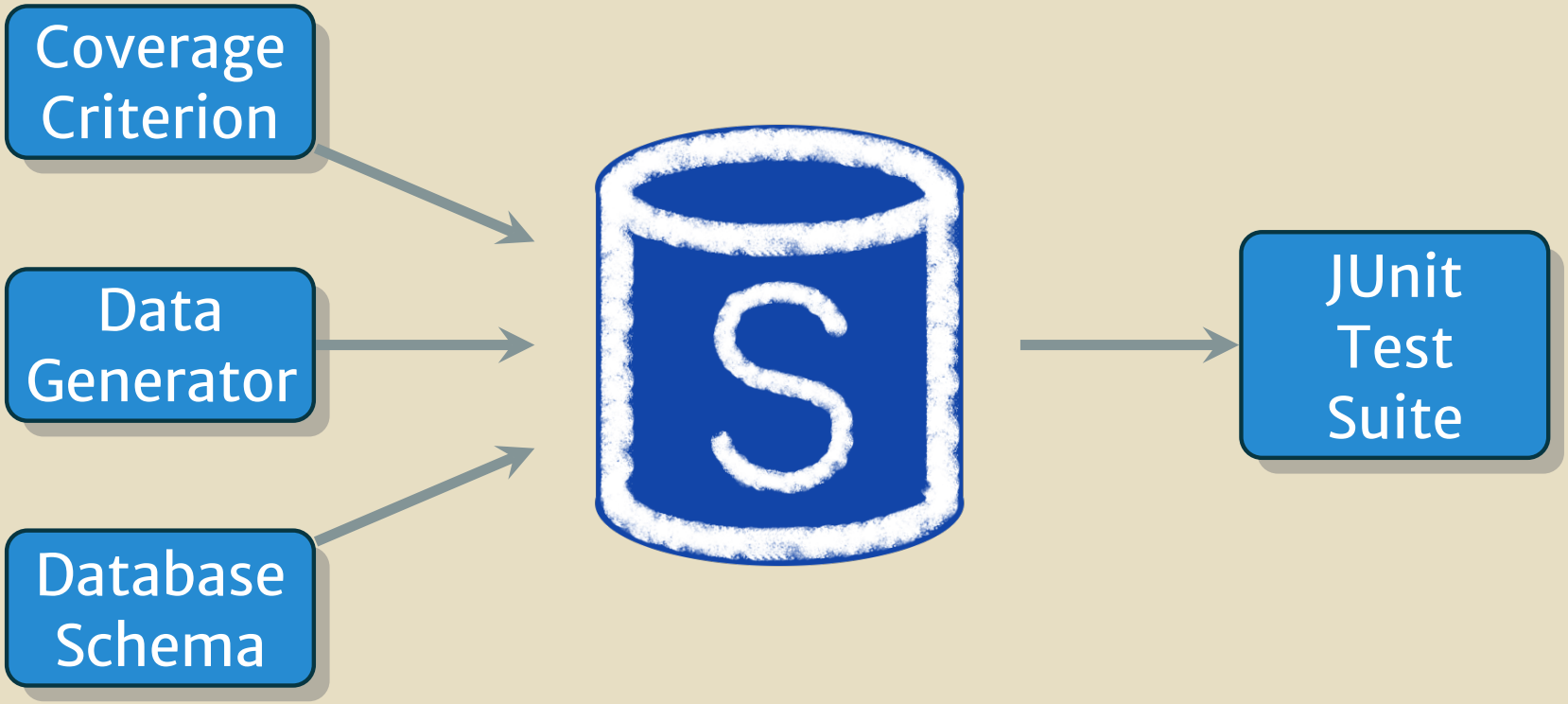
Extensible Tool for Test Data Deneration

Schema Testing



Extensible Tool for Test Data Deneration

Schema Testing



Extensible Tool for Test Data Deneration

Schema Testing

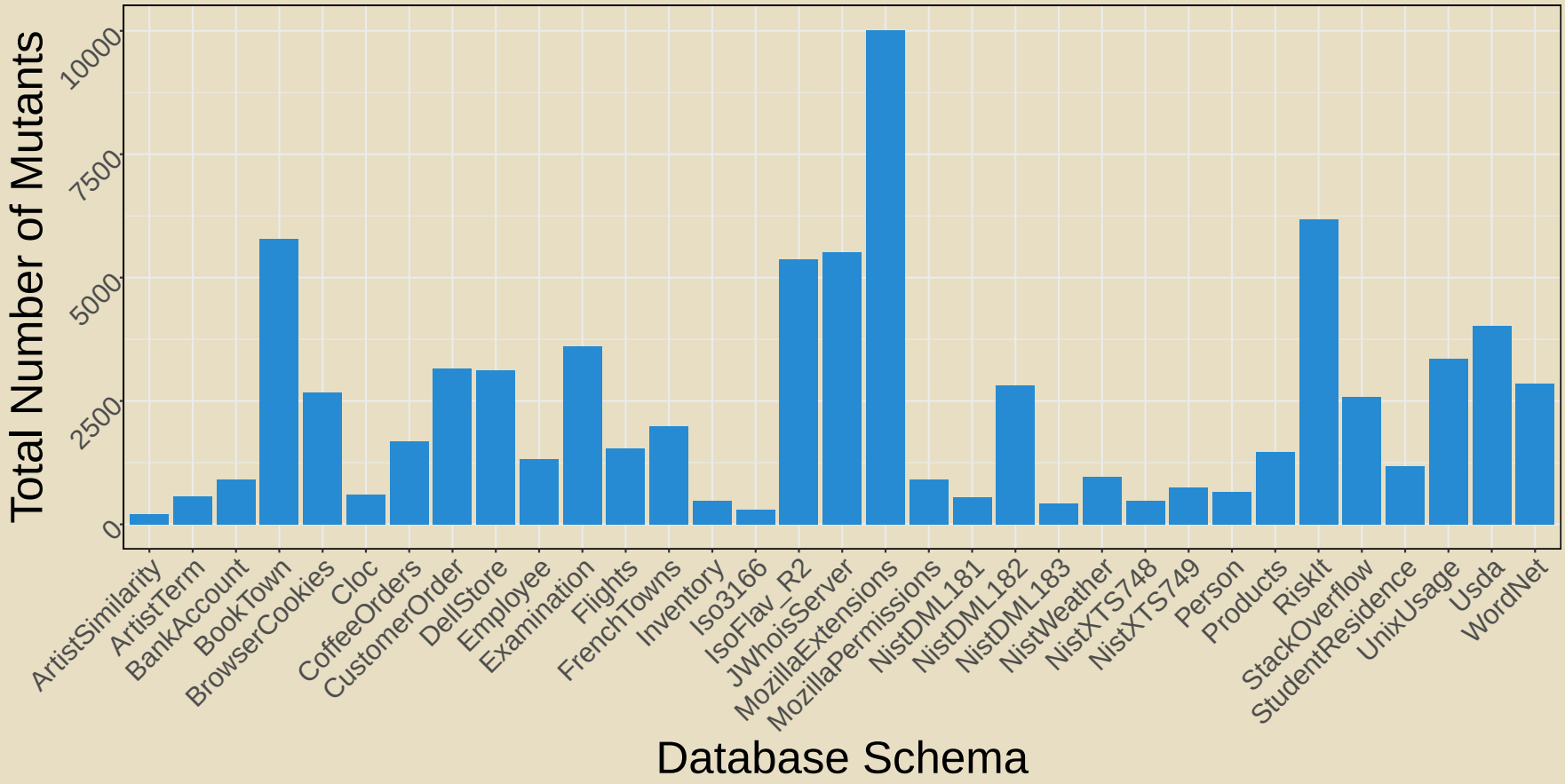
Adequacy?

Schema Testing

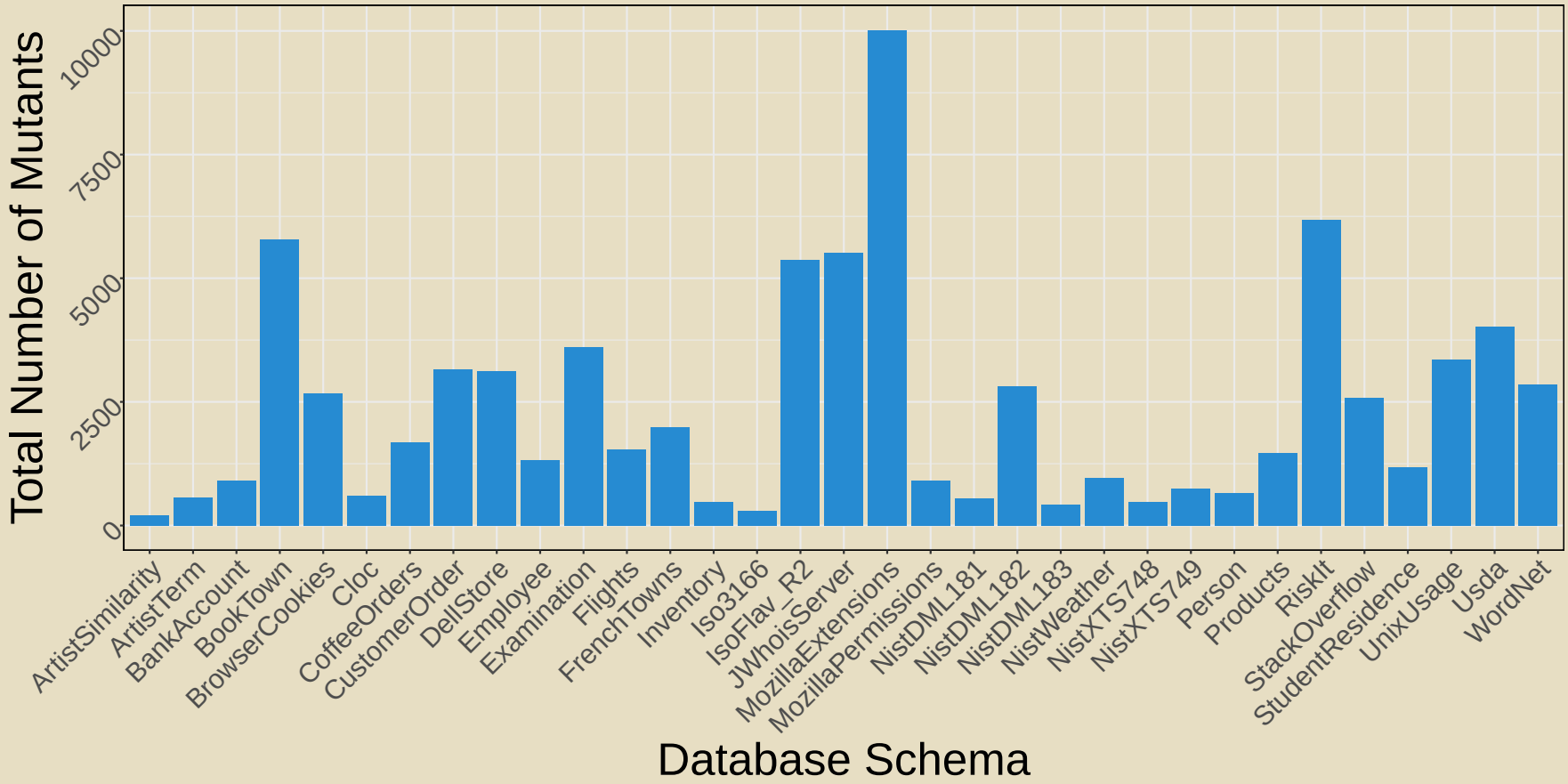
Similar

Limitations!

Schema Testing

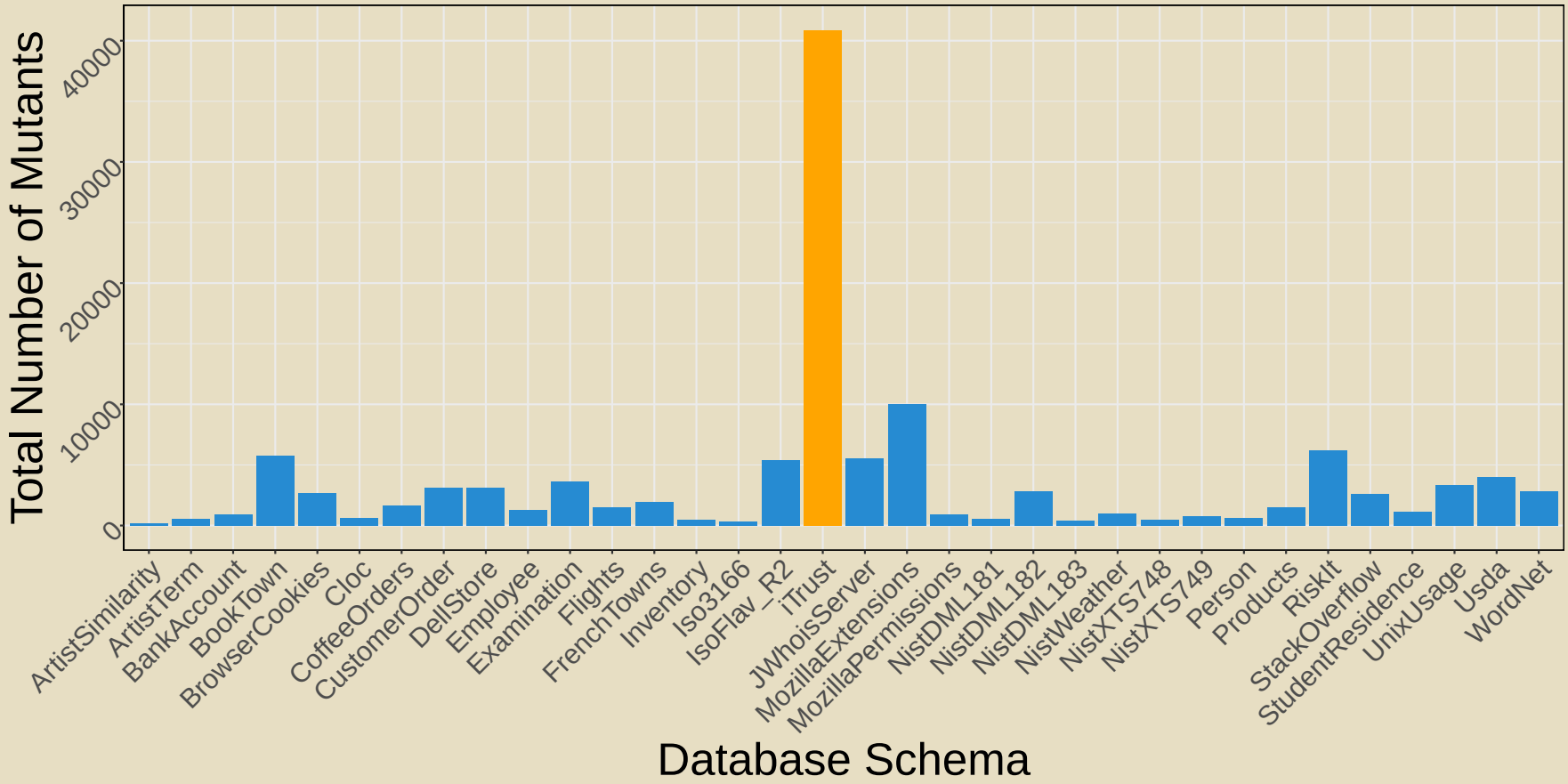


Schema Testing

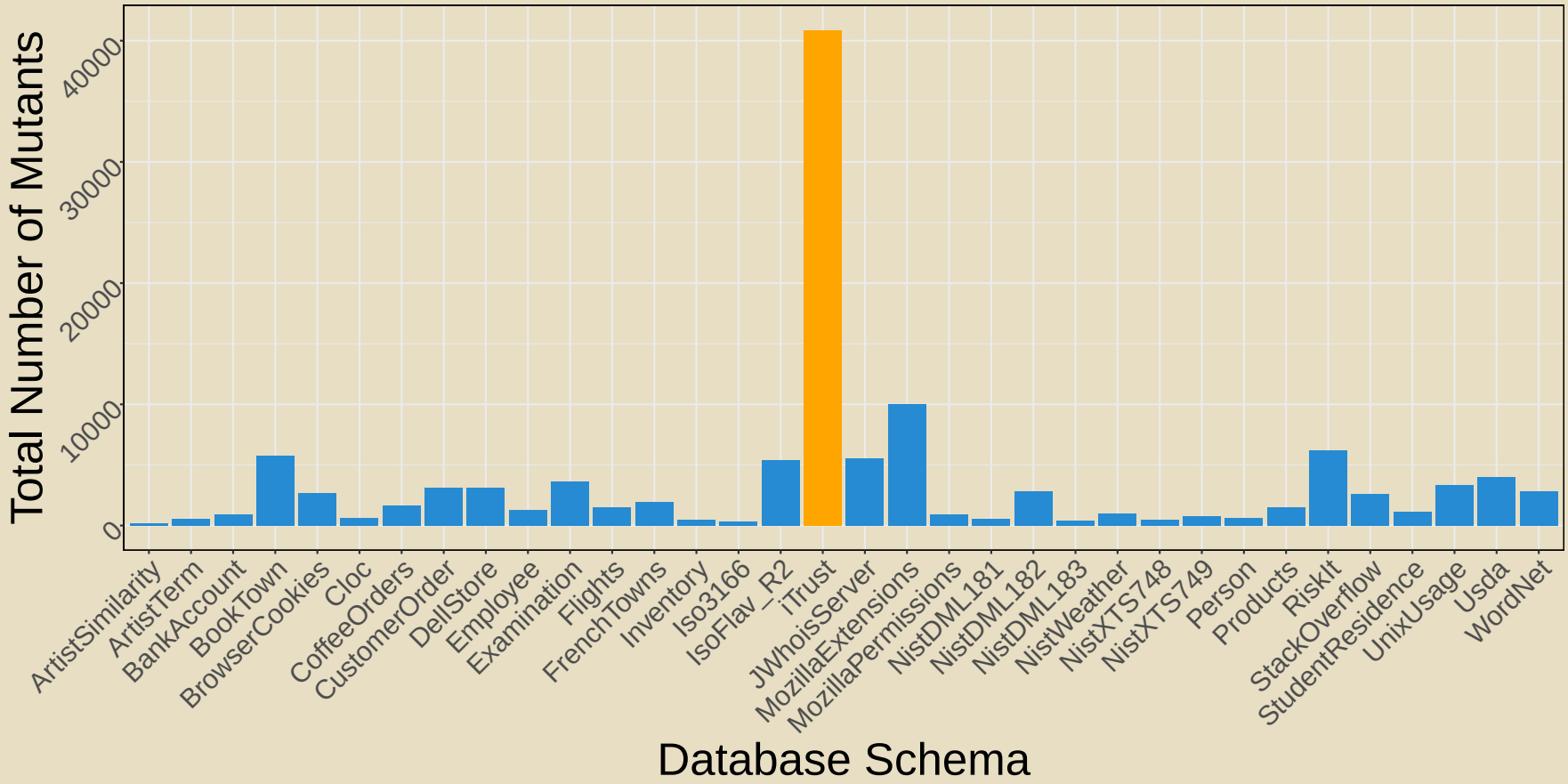


Over 10,000 generated mutants for one schema!

Schema Testing



Schema Testing



Over 40,000 for the iTrust schema!

Reduction Techniques

Reduce the Number of Mutants

Reduction Techniques

Random Sampling

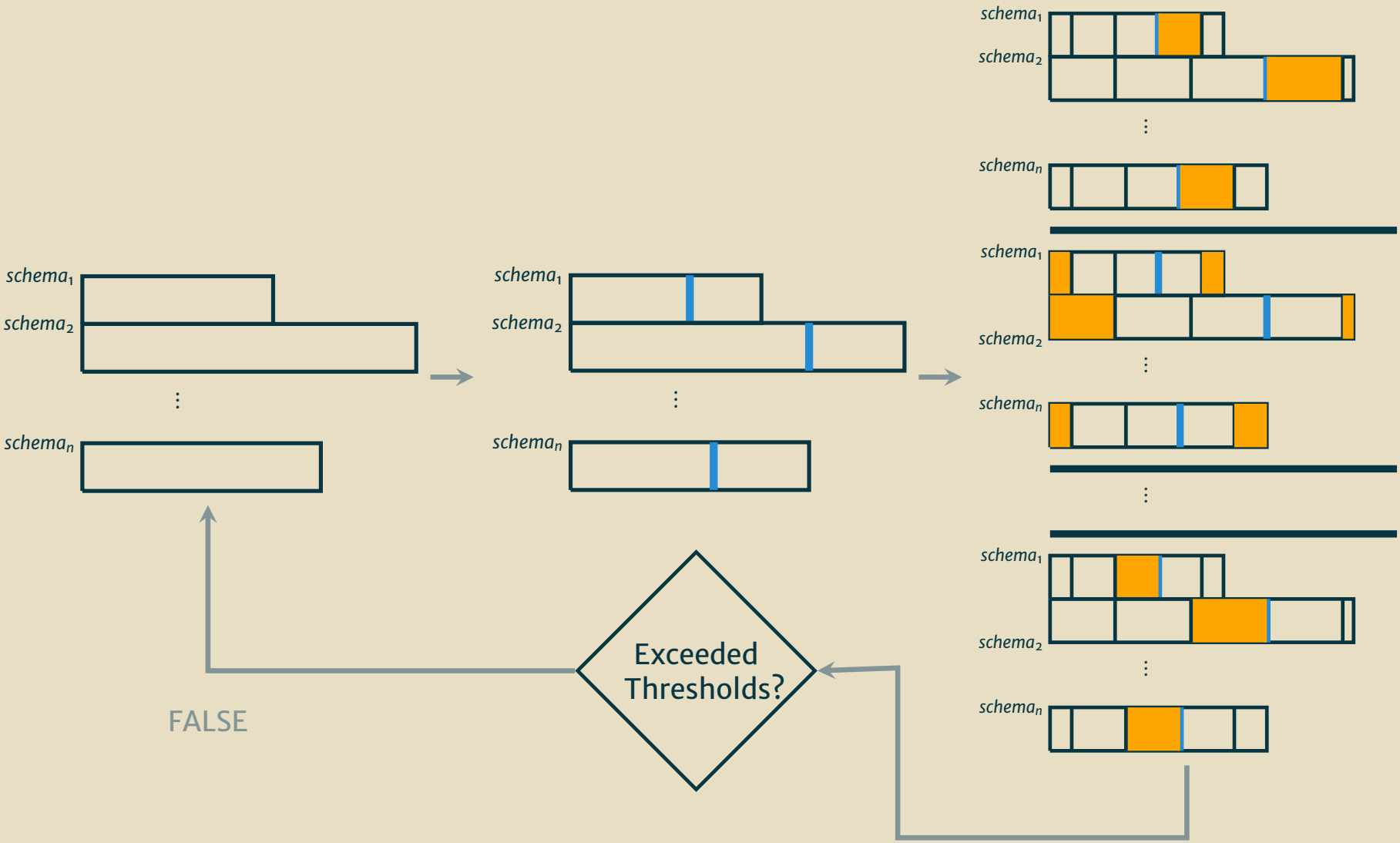
Reduction Techniques

Hill Climbing

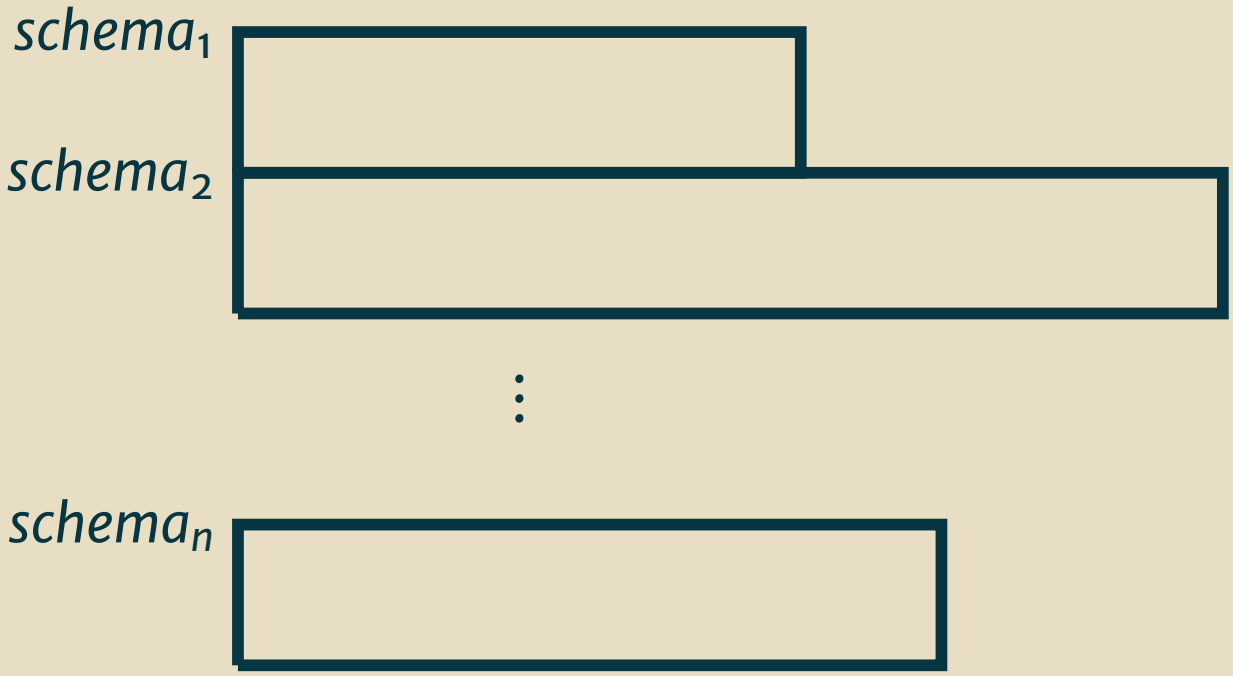
Hill Climbing

Fitness Function

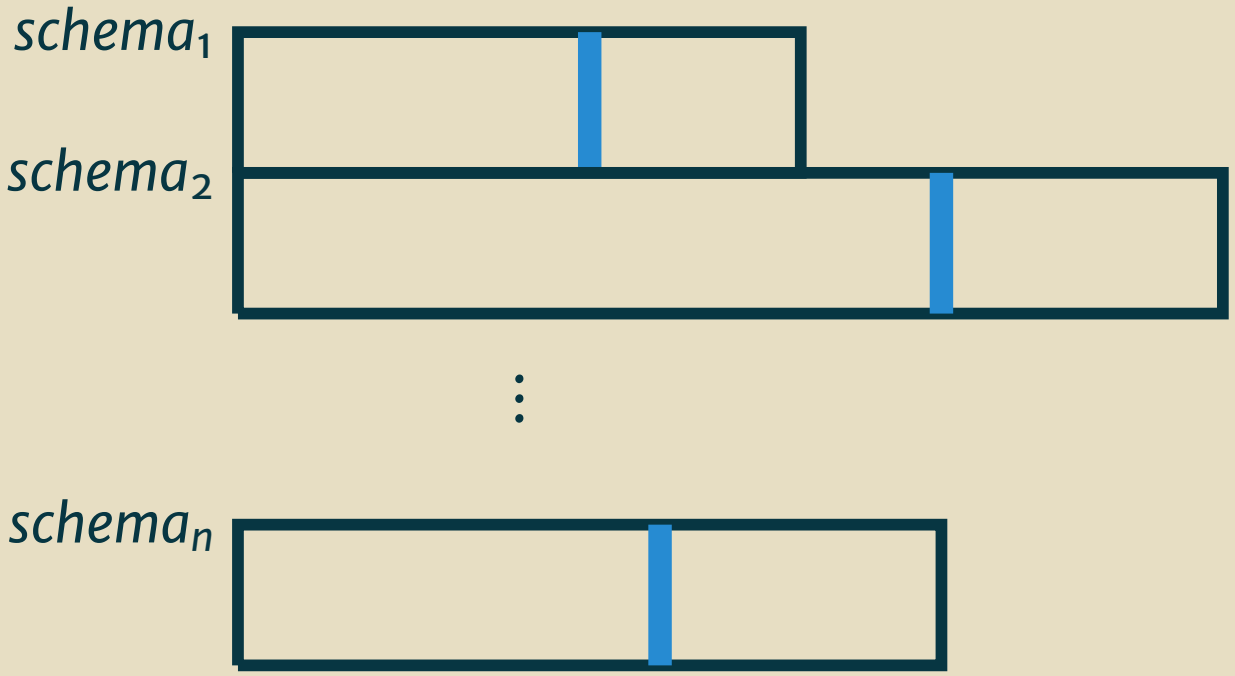
Hill Climbing



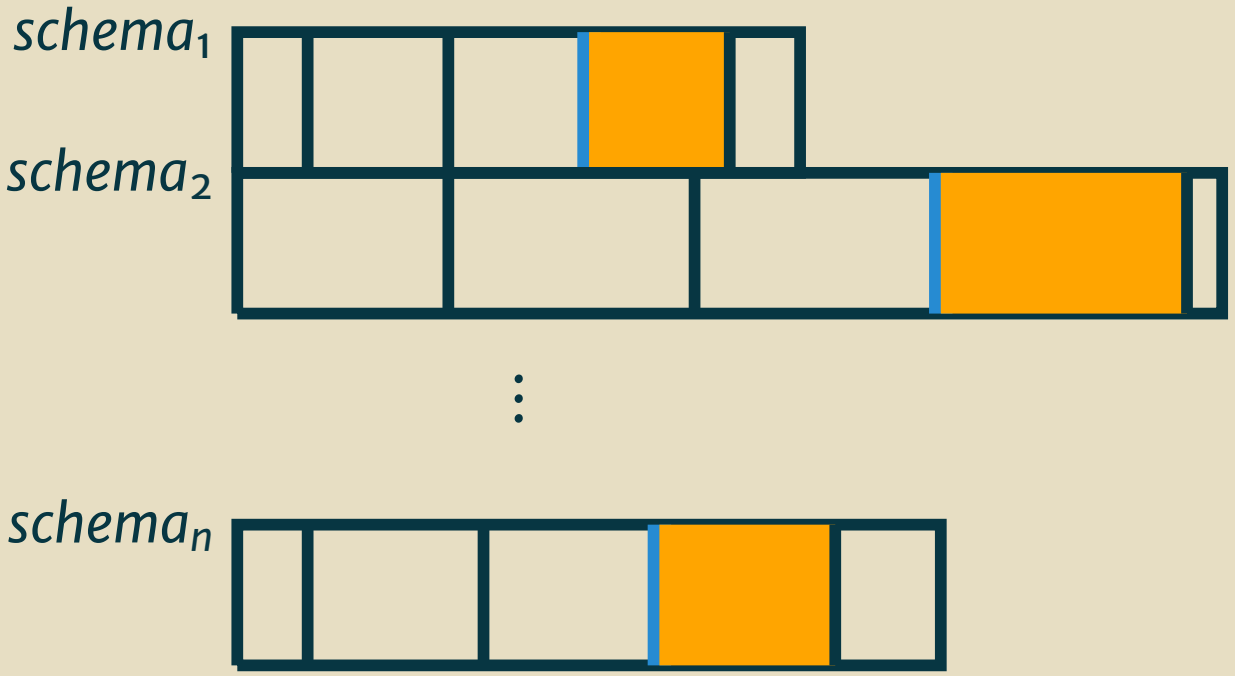
Hill Climbing



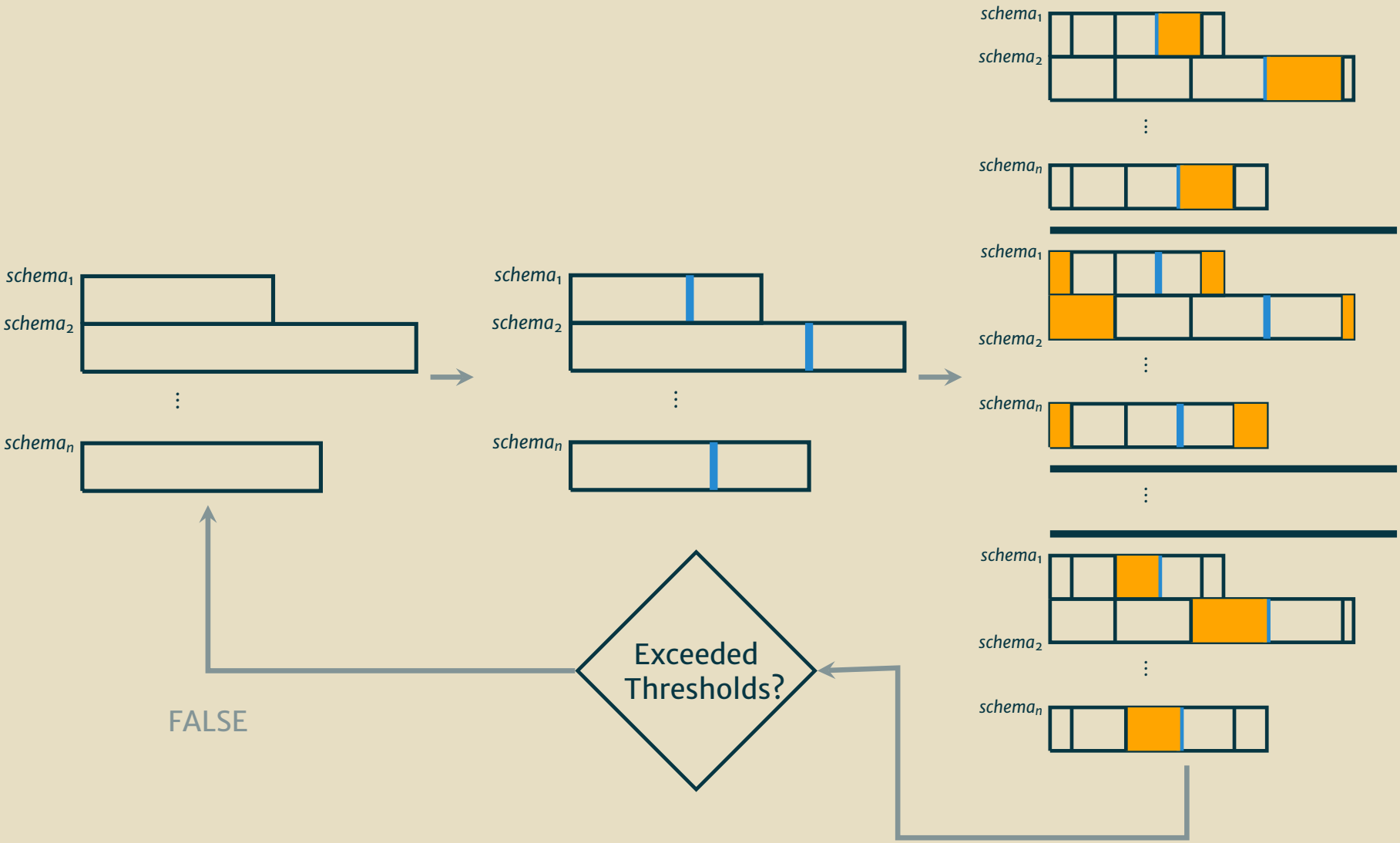
Hill Climbing



Hill Climbing



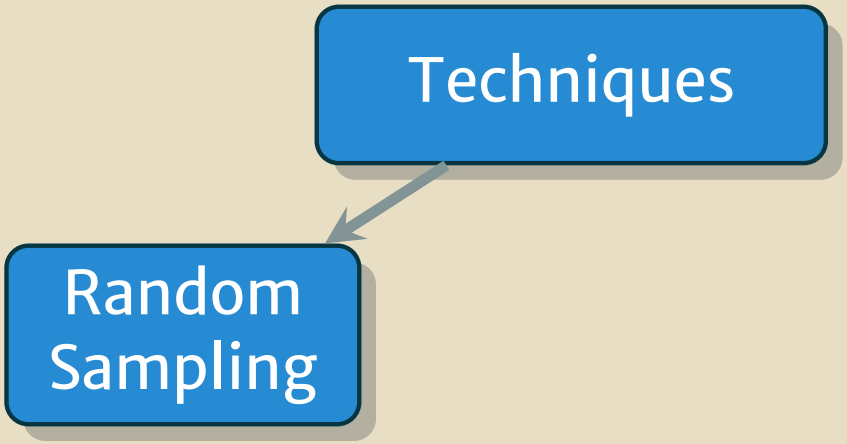
Hill Climbing



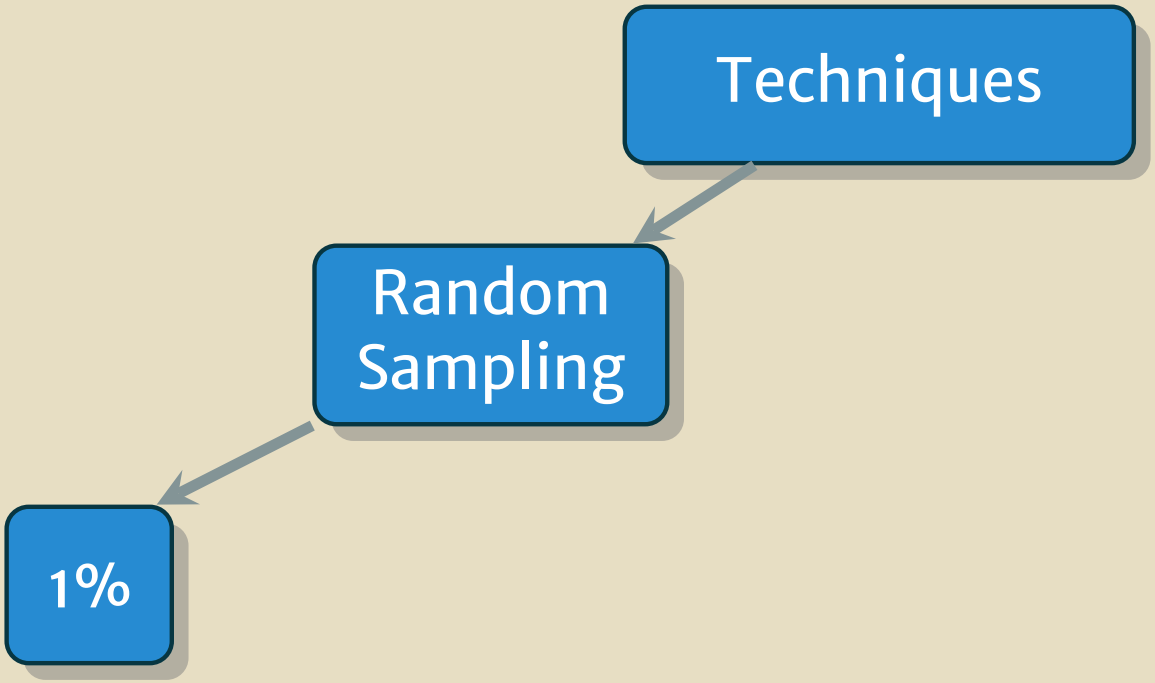
Methodology

Techniques

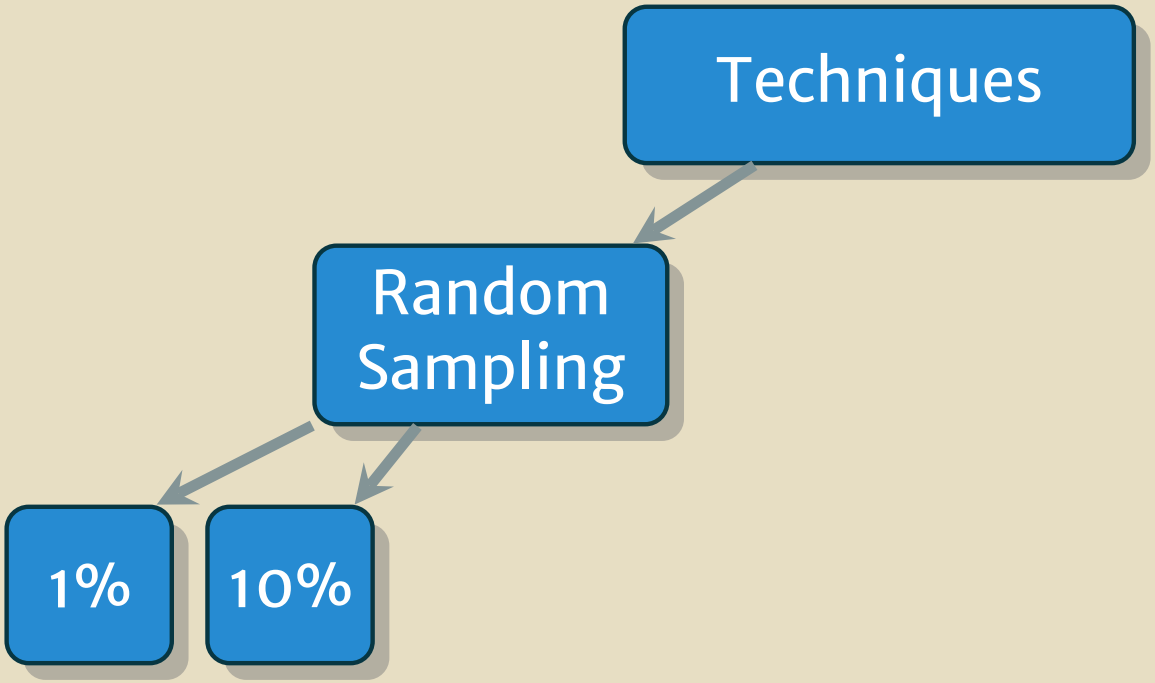
Methodology



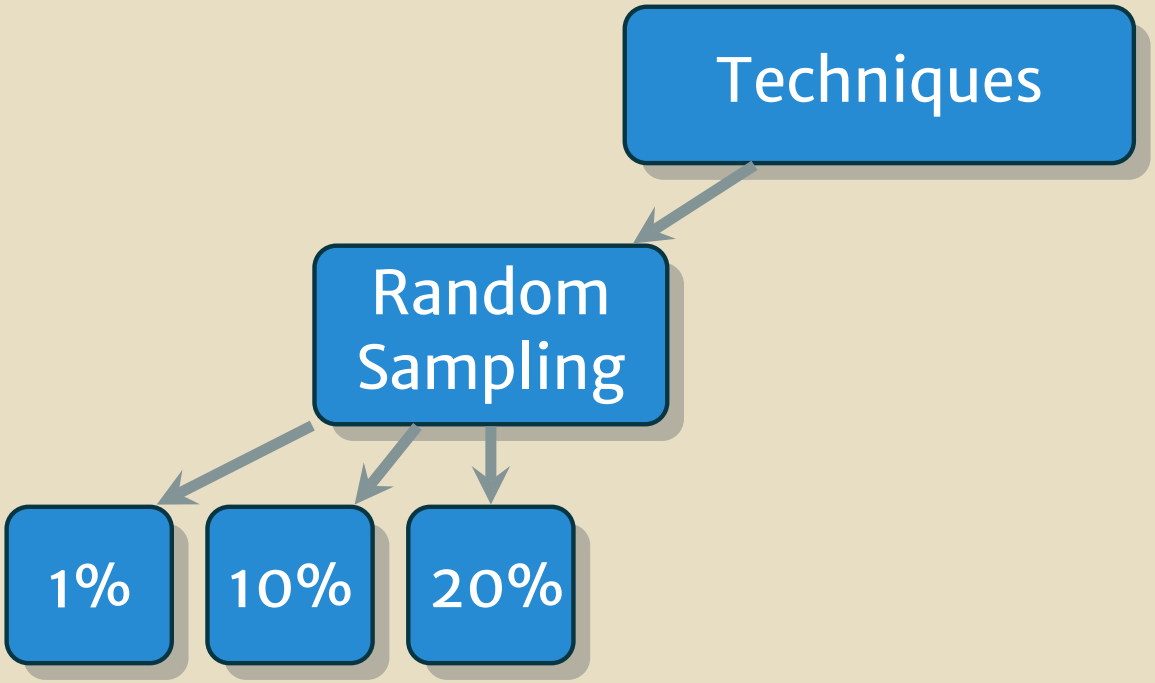
Methodology



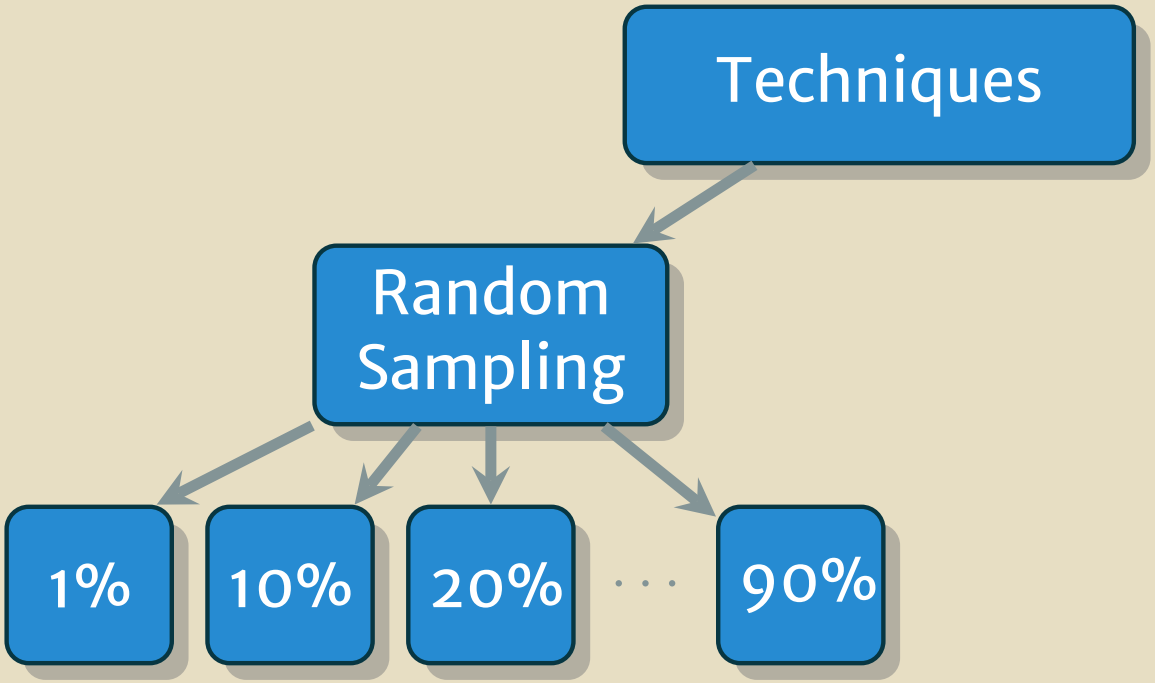
Methodology



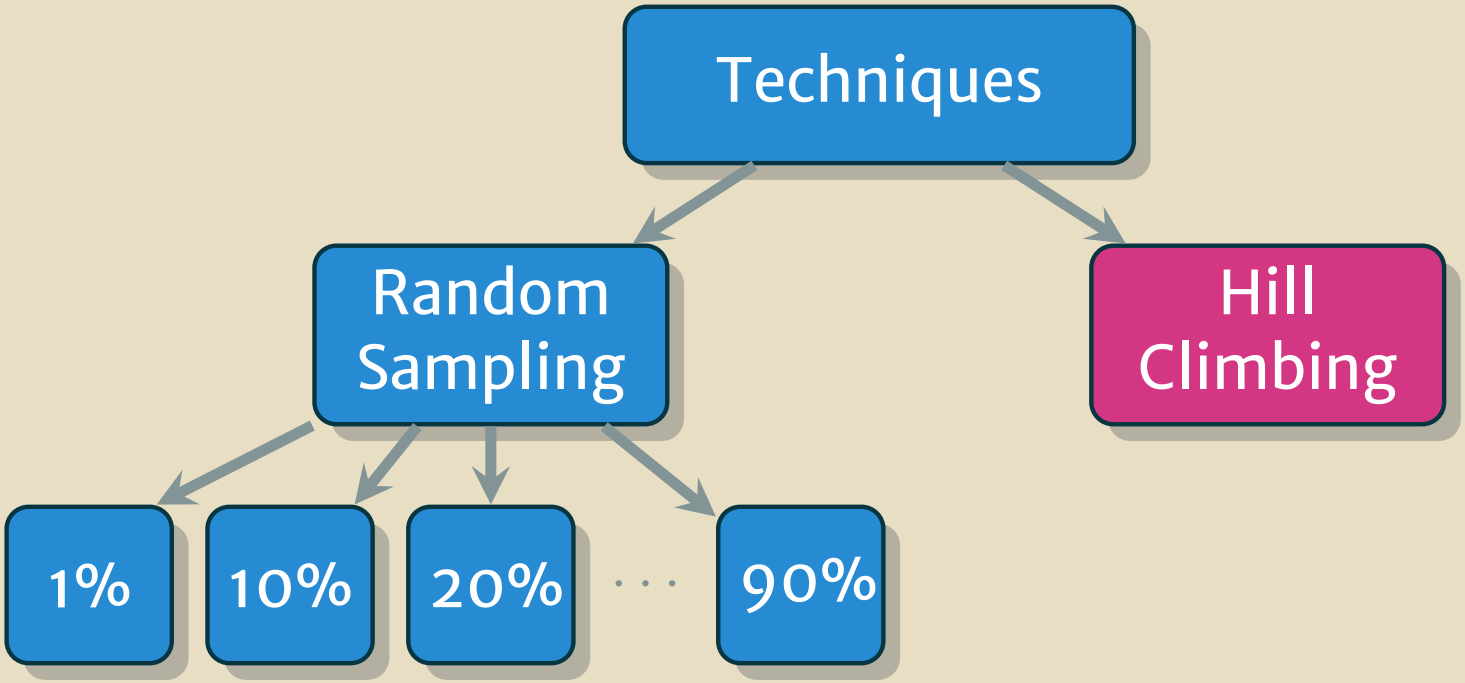
Methodology



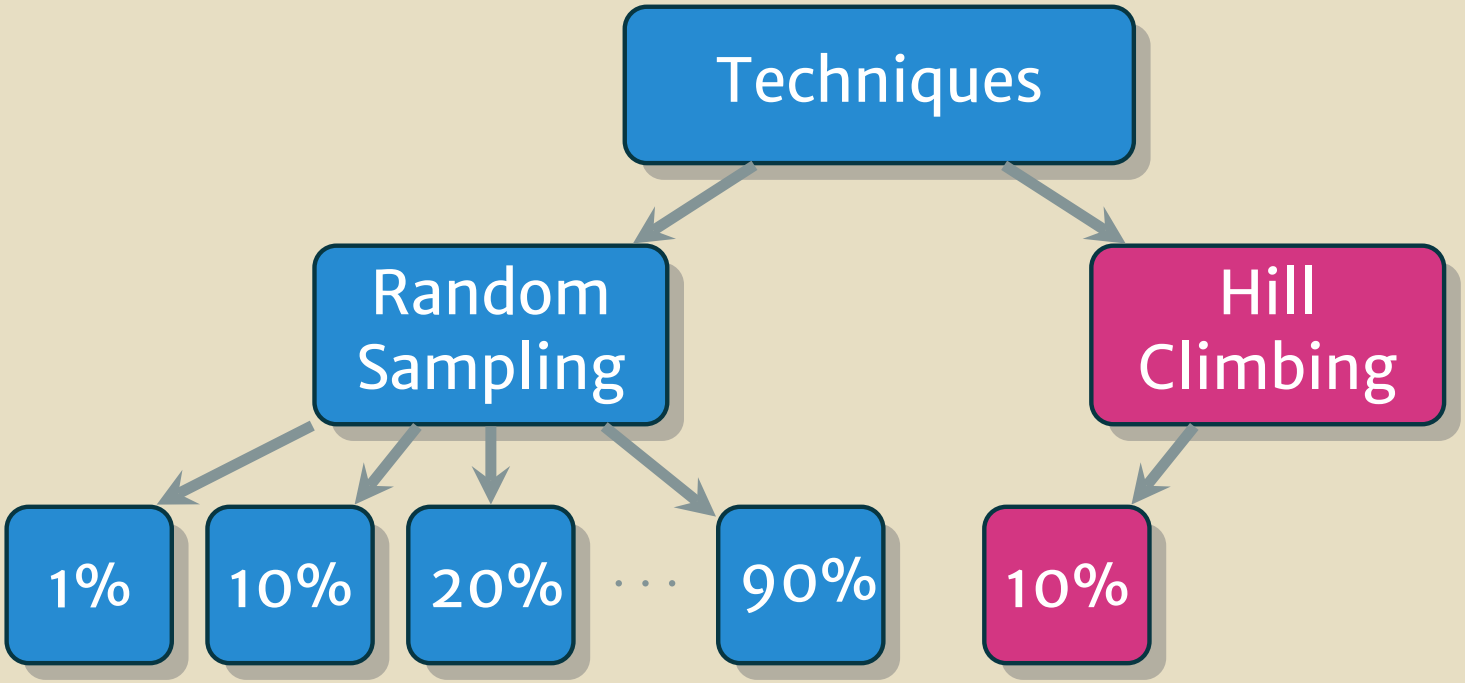
Methodology



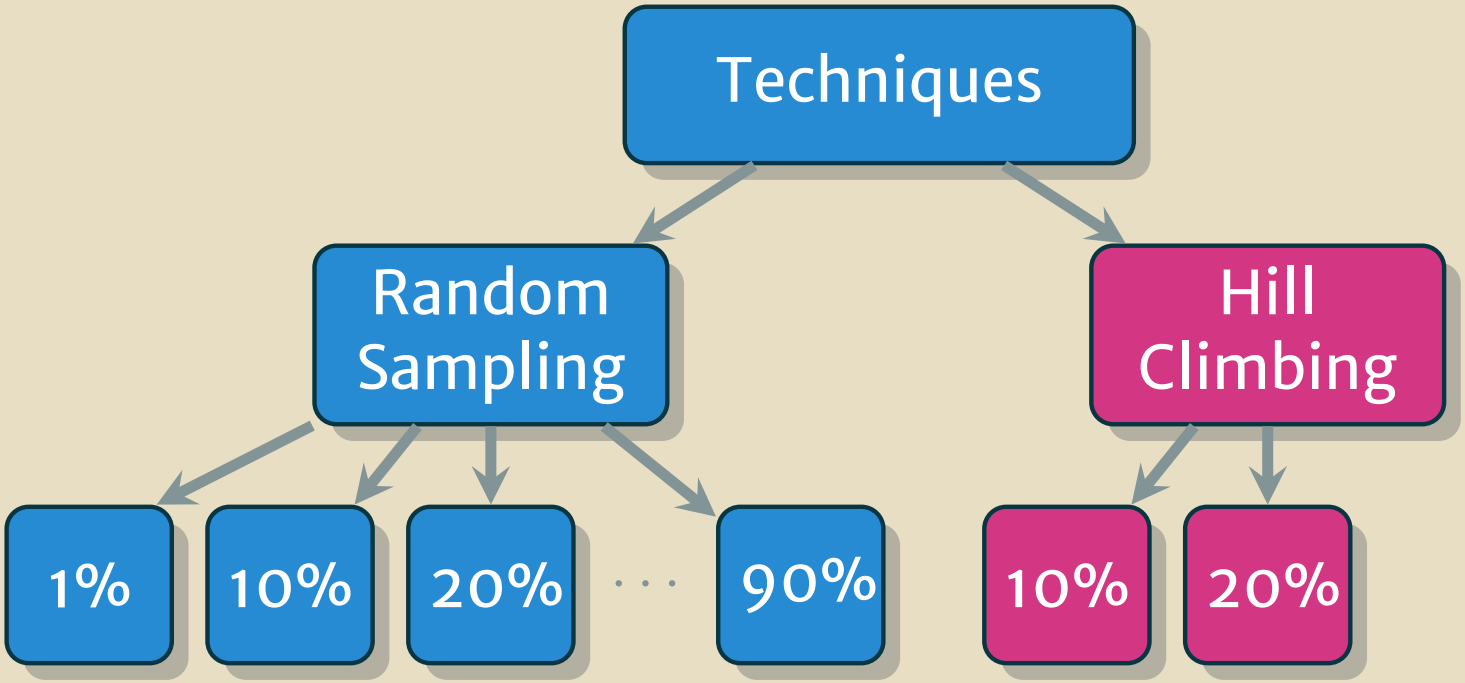
Methodology



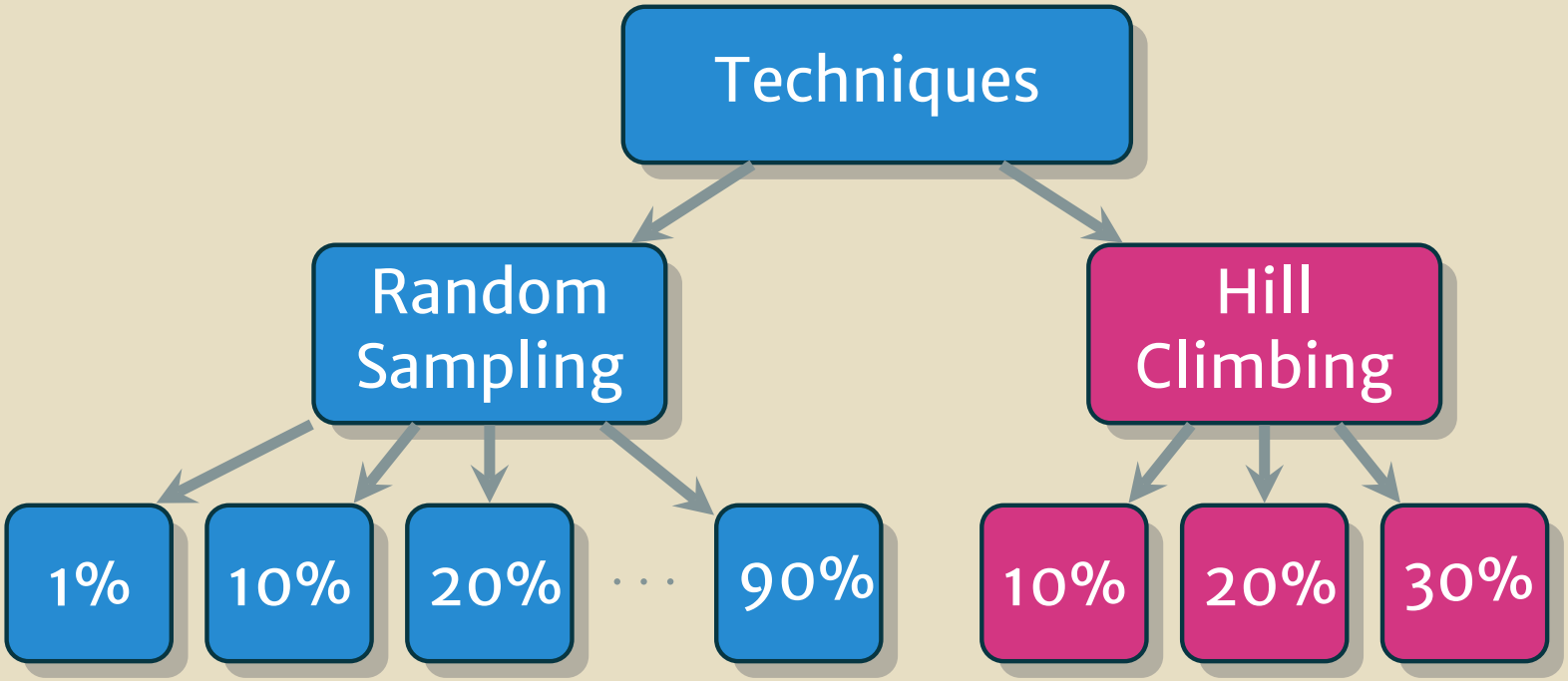
Methodology



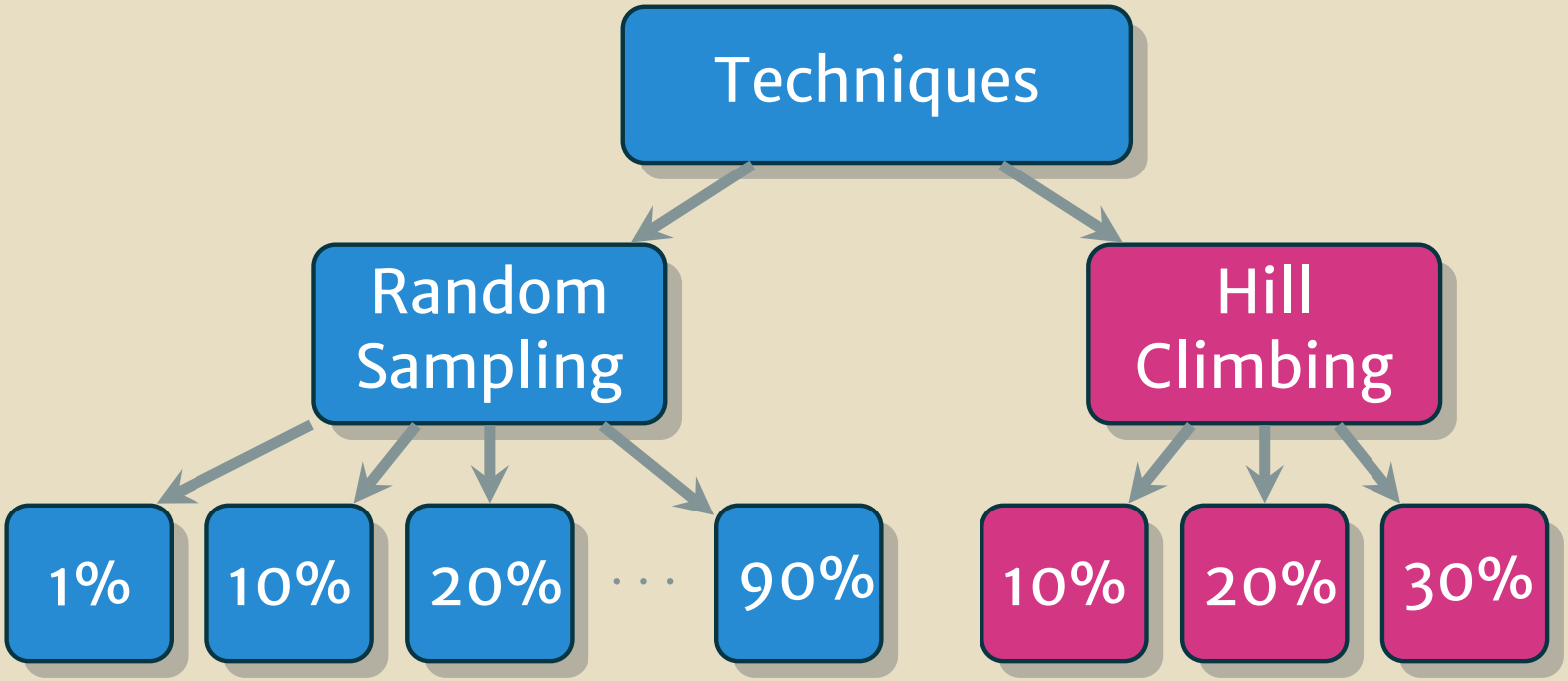
Methodology



Methodology

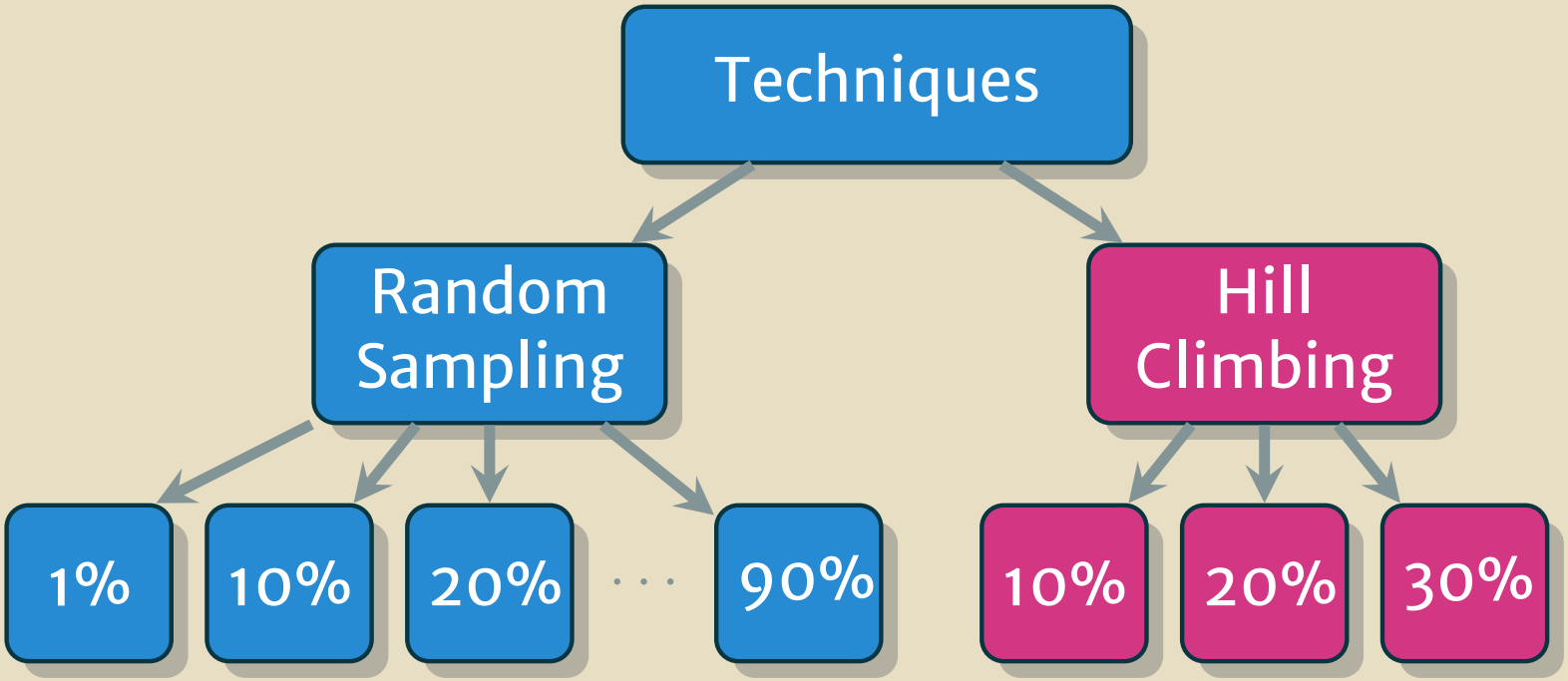


Methodology



HC % is different that RS %!

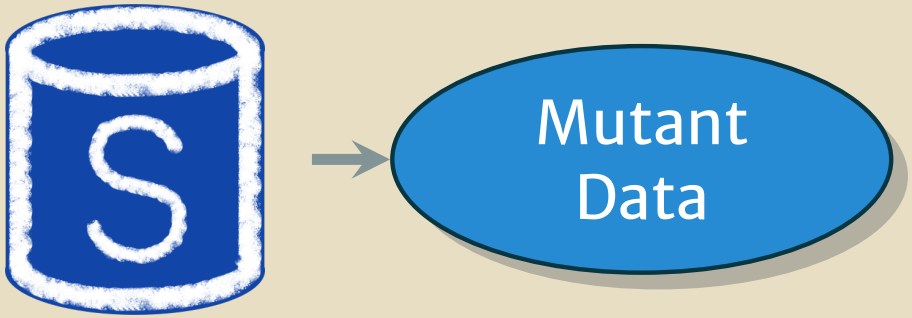
Methodology



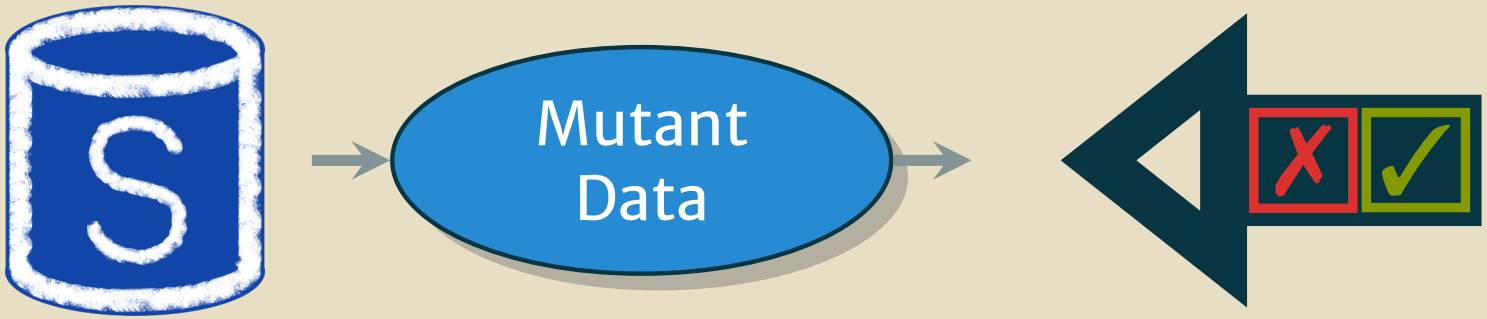
Implemented other techniques,
just need to evaluate them!

Methodology

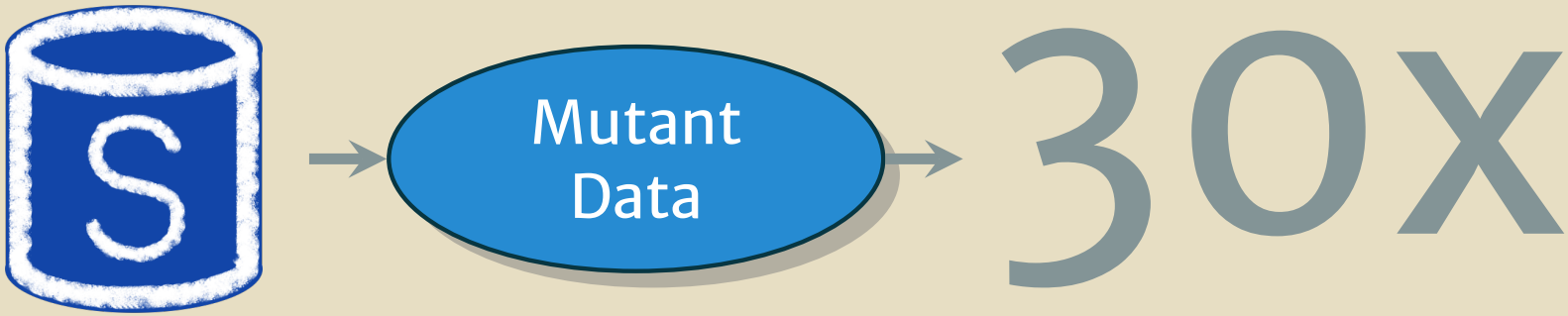
Methodology



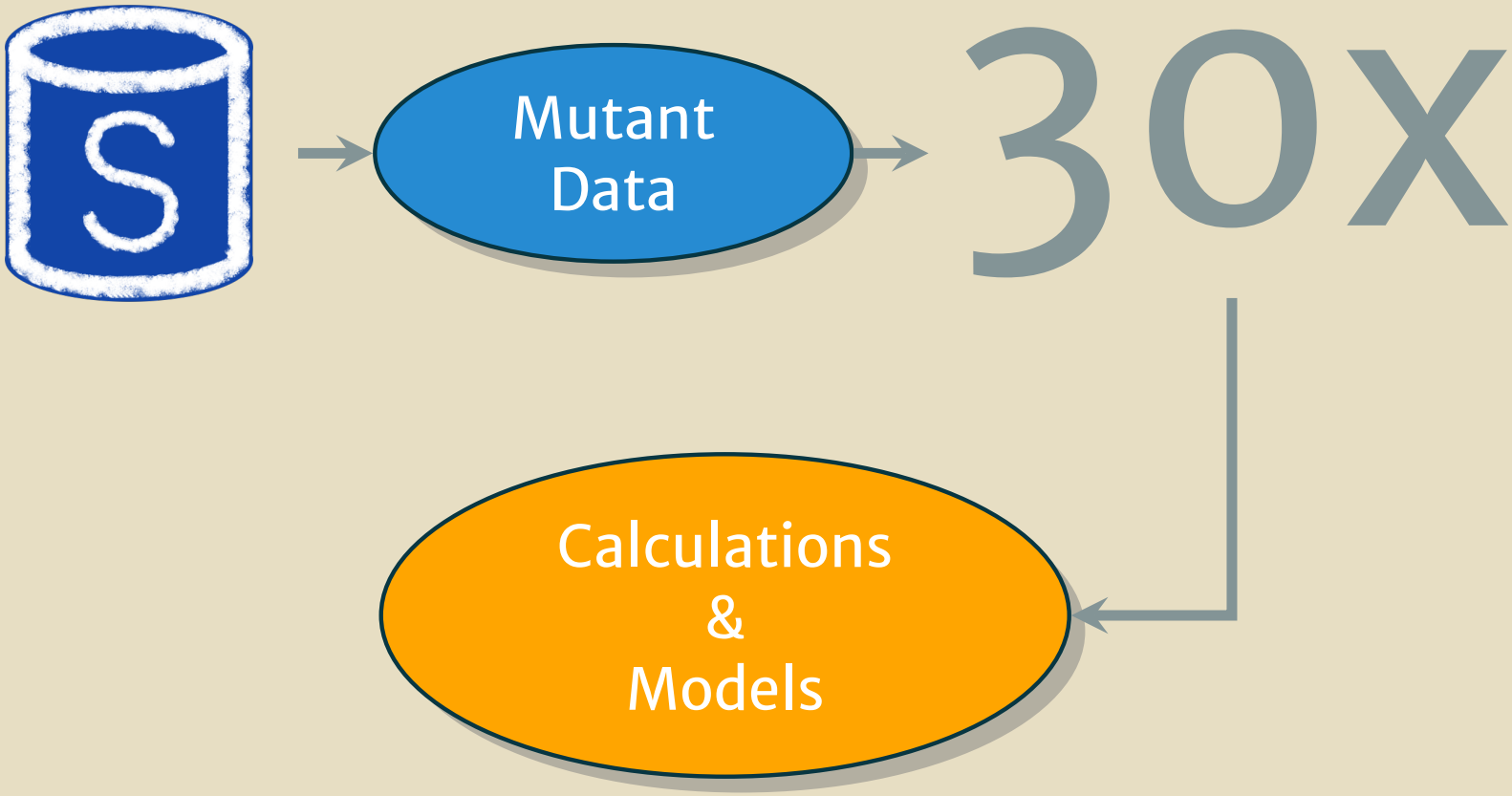
Methodology



Methodology



Methodology



Methodology

Generalizable Models

Methodology

Random Sampling Models

Methodology

	dbms	operator	keep %	ignore %
1	SQLite	CCInExpressionRHSListExpressionElementR	0.00	1.00
2	SQLite	CCNullifier	0.00	1.00
3	SQLite	CCRelationalExpressionOperatorE	0.11	0.89
4	SQLite	FKCColumnPairE	0.08	0.92
5	SQLite	FKCColumnPairR	0.17	0.83
6	SQLite	NNCA	0.19	0.81
7	SQLite	NNCR	0.34	0.66
8	SQLite	PKCColumnA	0.36	0.64
9	SQLite	PKCColumnE	0.34	0.66
10	SQLite	PKCColumnR	0.25	0.75
11	SQLite	UCColumnA	0.17	0.83
12	SQLite	UCColumnE	0.09	0.91
13	SQLite	UCColumnR	0.00	1.00

Methodology

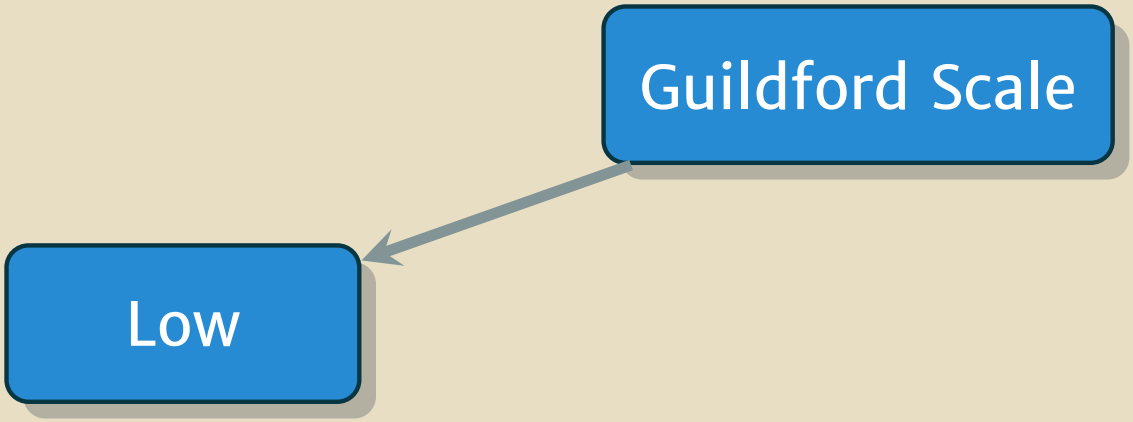
	dbms	operator	keep %	ignore %
1	SQLite	CCInExpressionRHSListExpressionElementR	0.00	1.00
2	SQLite	CCNullifier	0.00	1.00
3	SQLite	CCRelationalExpressionOperatorE	0.11	0.89
4	SQLite	FKCColumnPairE	0.08	0.92
5	SQLite	FKCColumnPairR	0.17	0.83
6	SQLite	NNCA	0.19	0.81
7	SQLite	NNCR	0.34	0.66
8	SQLite	PKCColumnA	0.36	0.64
9	SQLite	PKCColumnE	0.34	0.66
10	SQLite	PKCColumnR	0.25	0.75
11	SQLite	UCColumnA	0.17	0.83
12	SQLite	UCColumnE	0.09	0.91
13	SQLite	UCColumnR	0.00	1.00

Produces a Generalizable Model

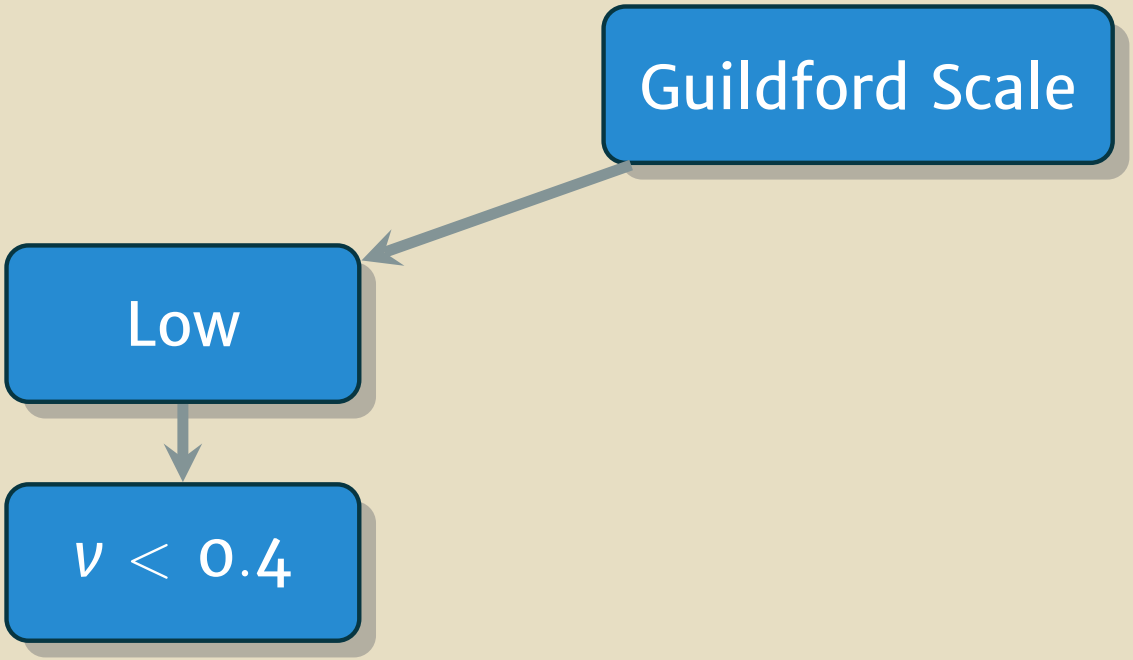
Evaluation Metrics

Guildford Scale

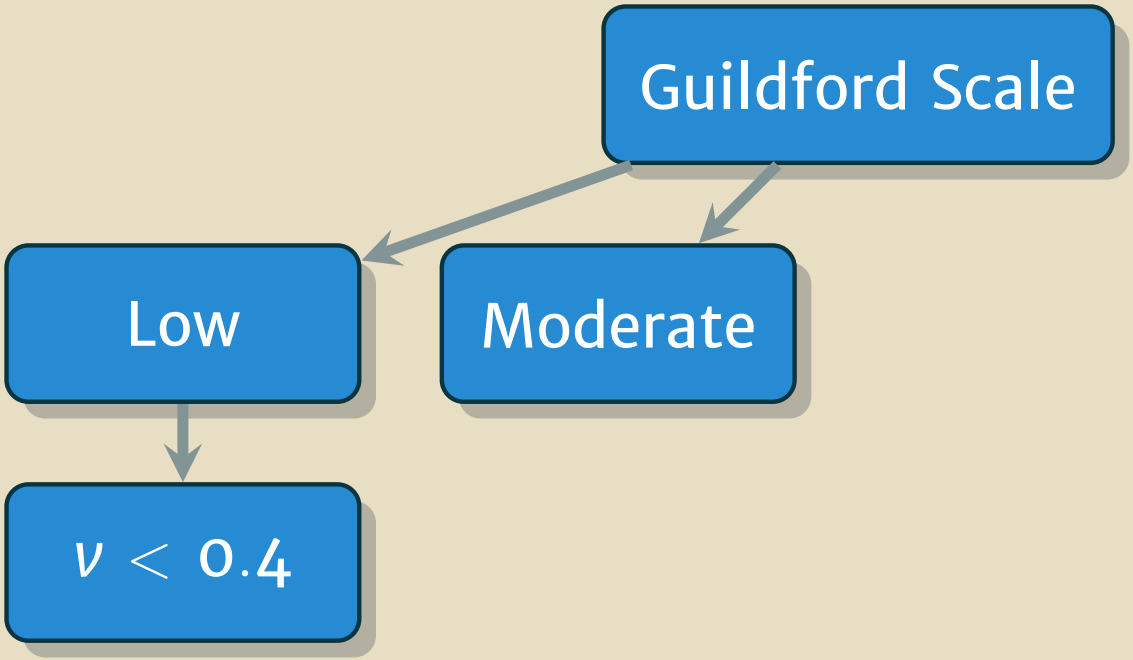
Evaluation Metrics



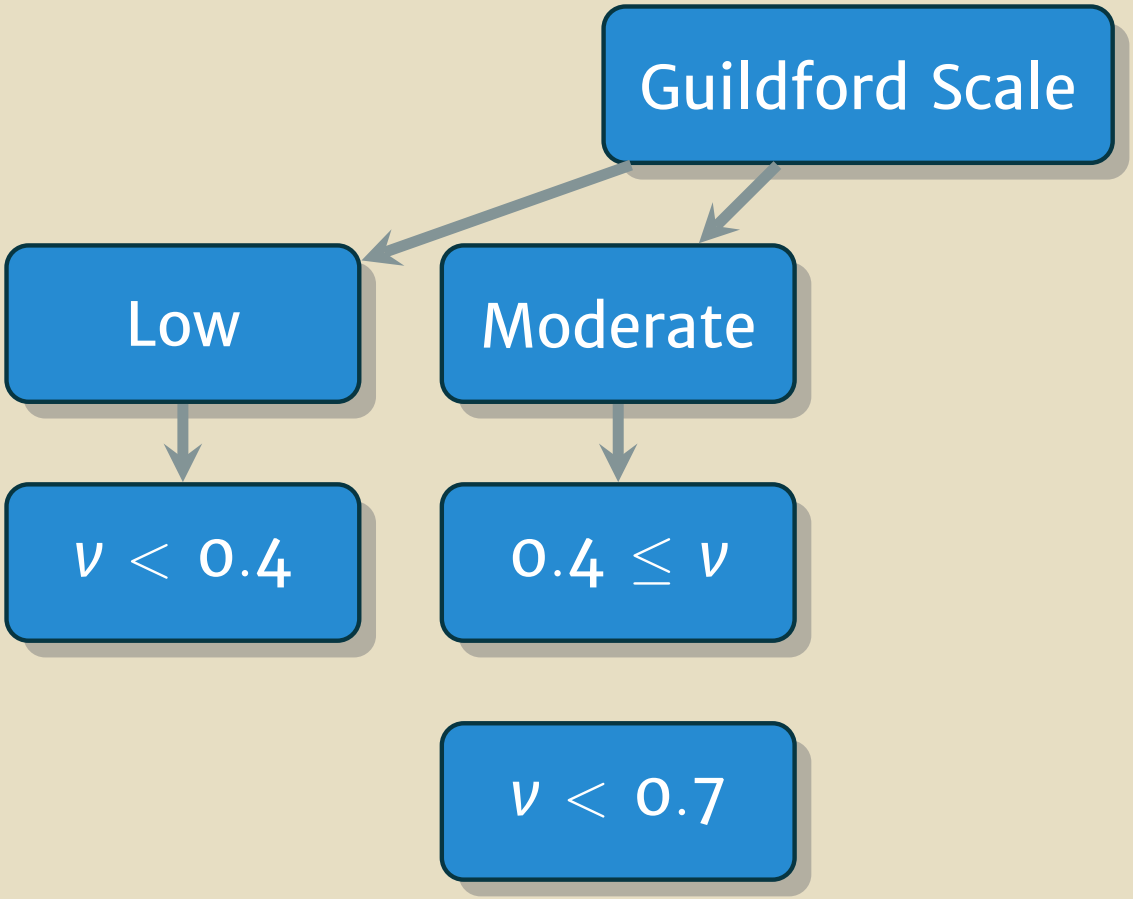
Evaluation Metrics



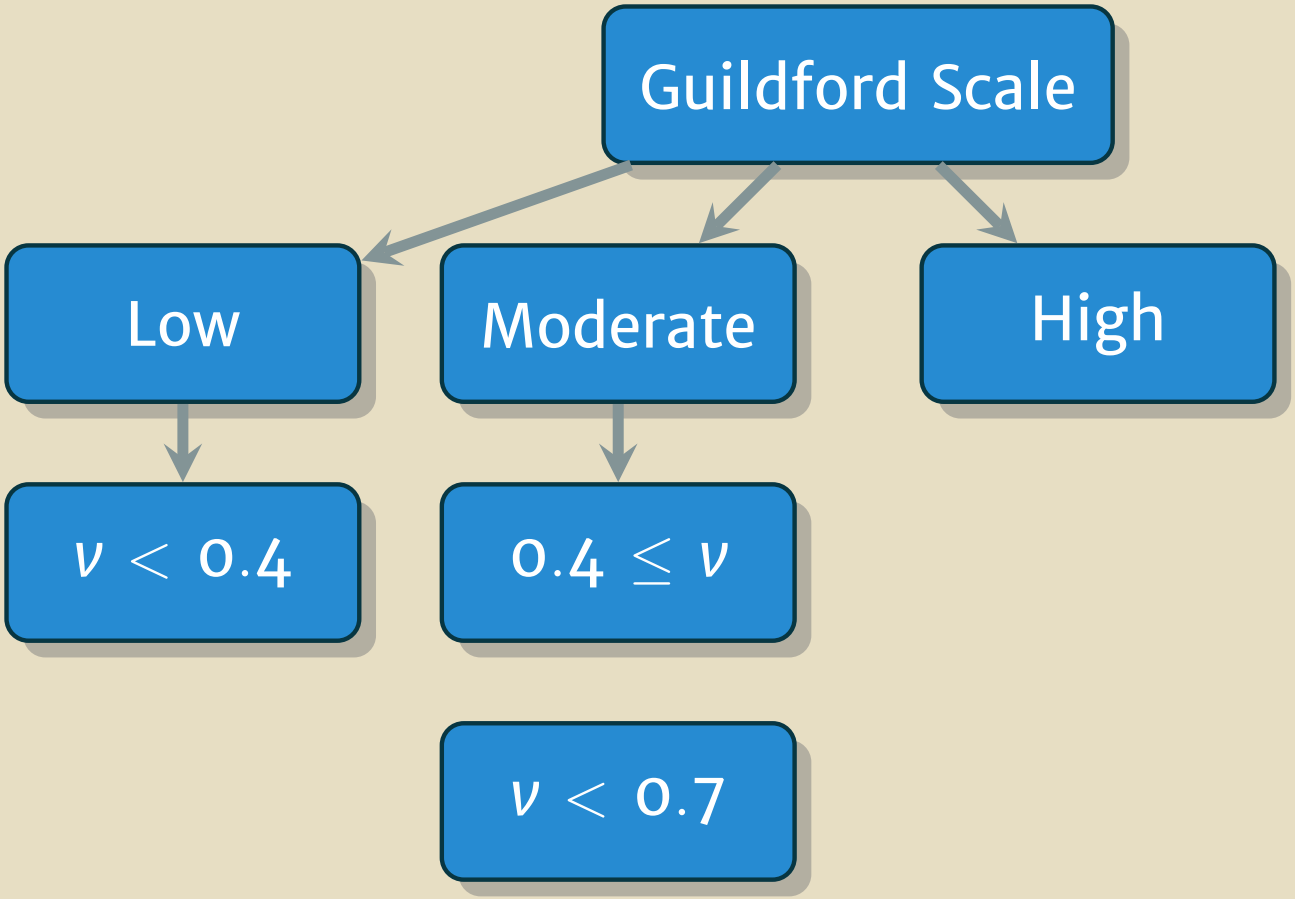
Evaluation Metrics



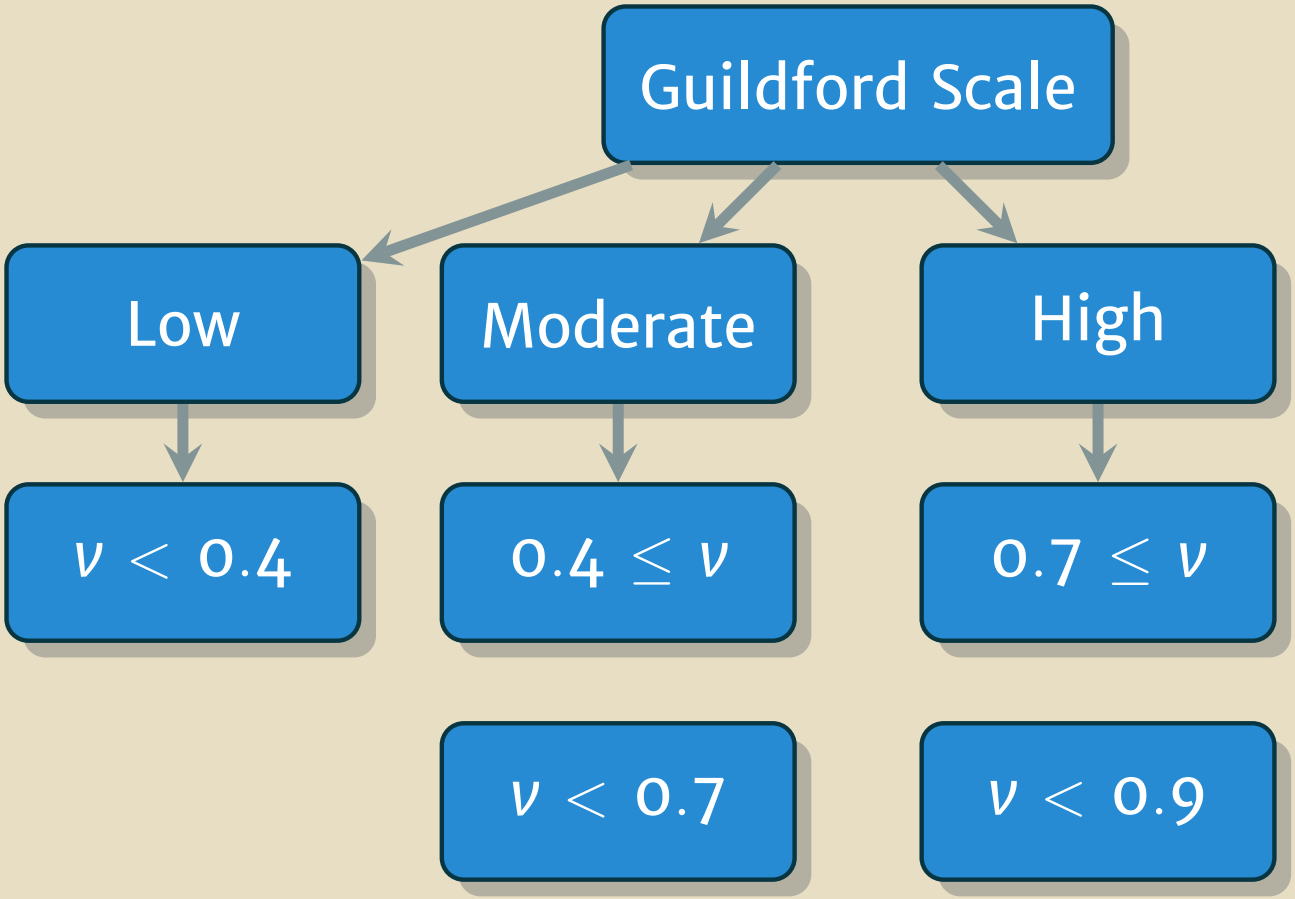
Evaluation Metrics



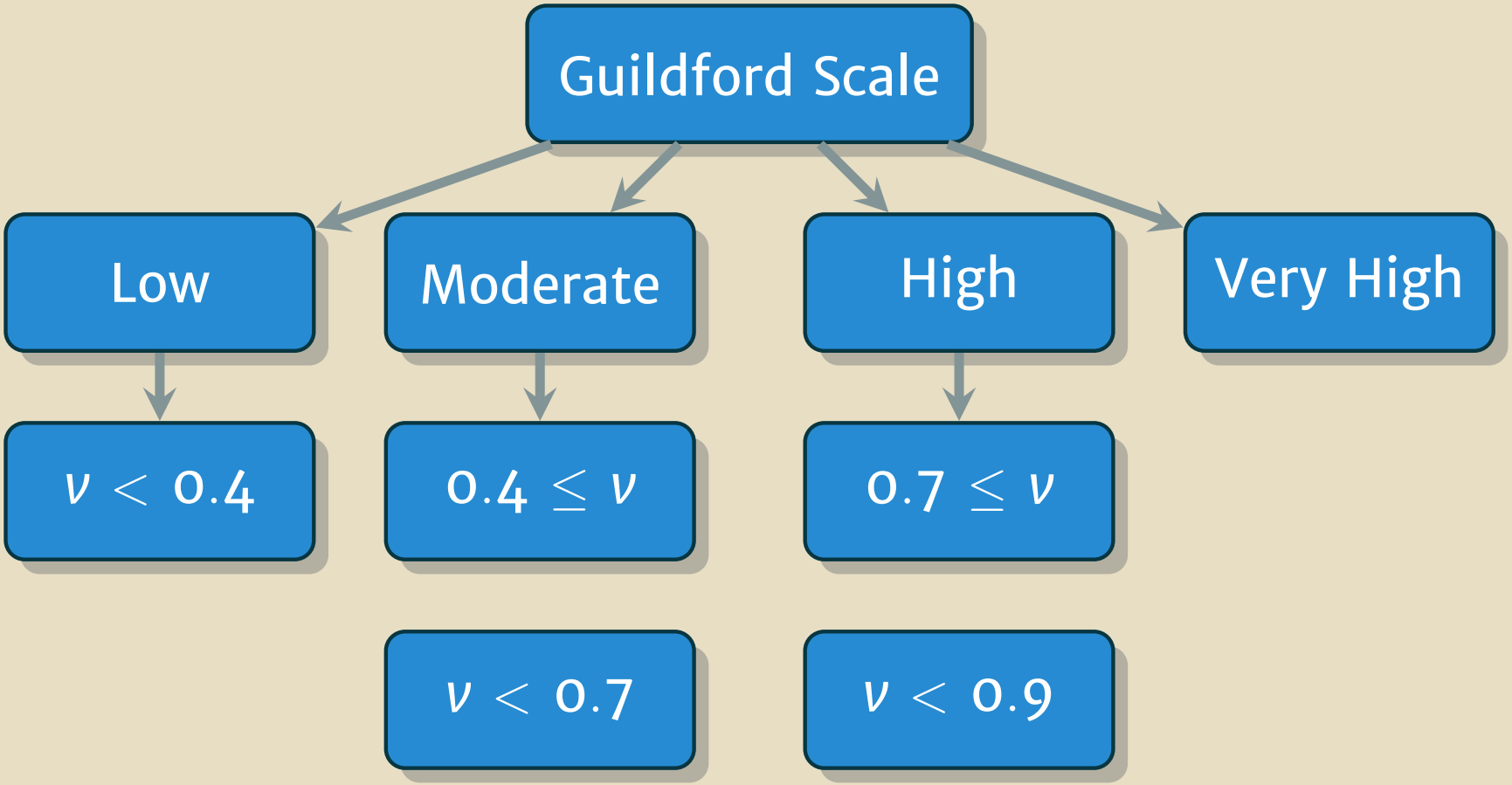
Evaluation Metrics



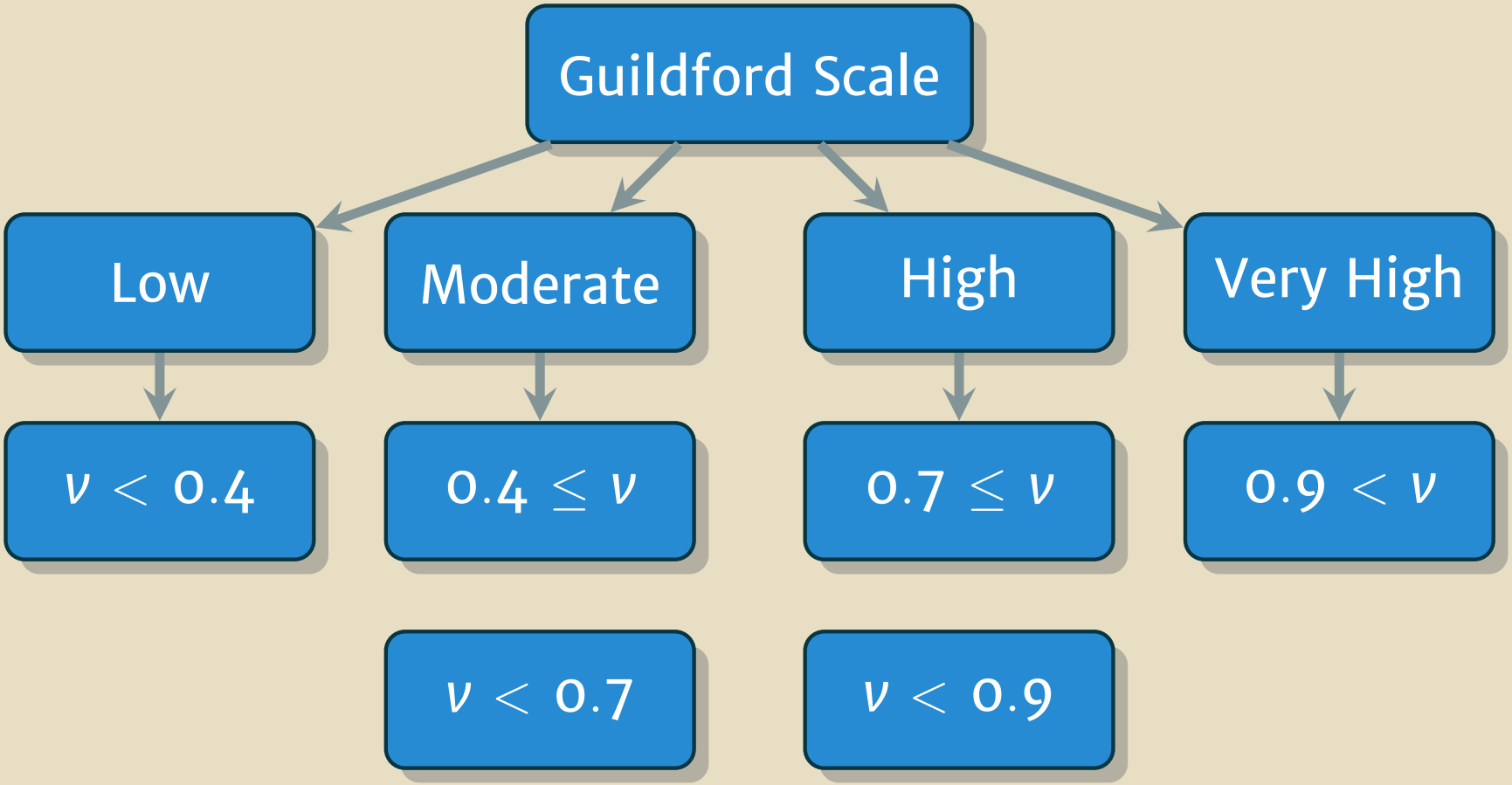
Evaluation Metrics



Evaluation Metrics



Evaluation Metrics



Evaluation Metrics

$$FCR = \frac{Time_O - Time_R}{Time_O}$$

Evaluation Metrics

$$FCR = \frac{Time_O - Time_R}{Time_O}$$

Evaluation Metrics

$$FCR = \frac{Time_O - Time_R}{Time_O}$$

Evaluation Metrics

$$FCR = \frac{Time_O - Time_R}{Time_O}$$

$$FCR \in [0, 1]$$

Evaluation Metrics

$$FCR = \frac{Time_O - Time_R}{Time_O}$$

$FCR \in [0, 1]$

HIB

Evaluation Metrics

$$Ratio = \frac{\text{Mean Corr}}{1 - \text{Mean CR}}$$

Evaluation Metrics

$$Ratio = \frac{\text{Mean Corr}}{1 - \text{Mean CR}}$$

Evaluation Metrics

$$Ratio = \frac{\text{Mean Corr}}{1 - \text{Mean CR}}$$

Evaluation Metrics

$$Ratio = \frac{\text{Mean Corr}}{1 - \text{Mean CR}}$$

$$Ratio \in [0, \infty)$$

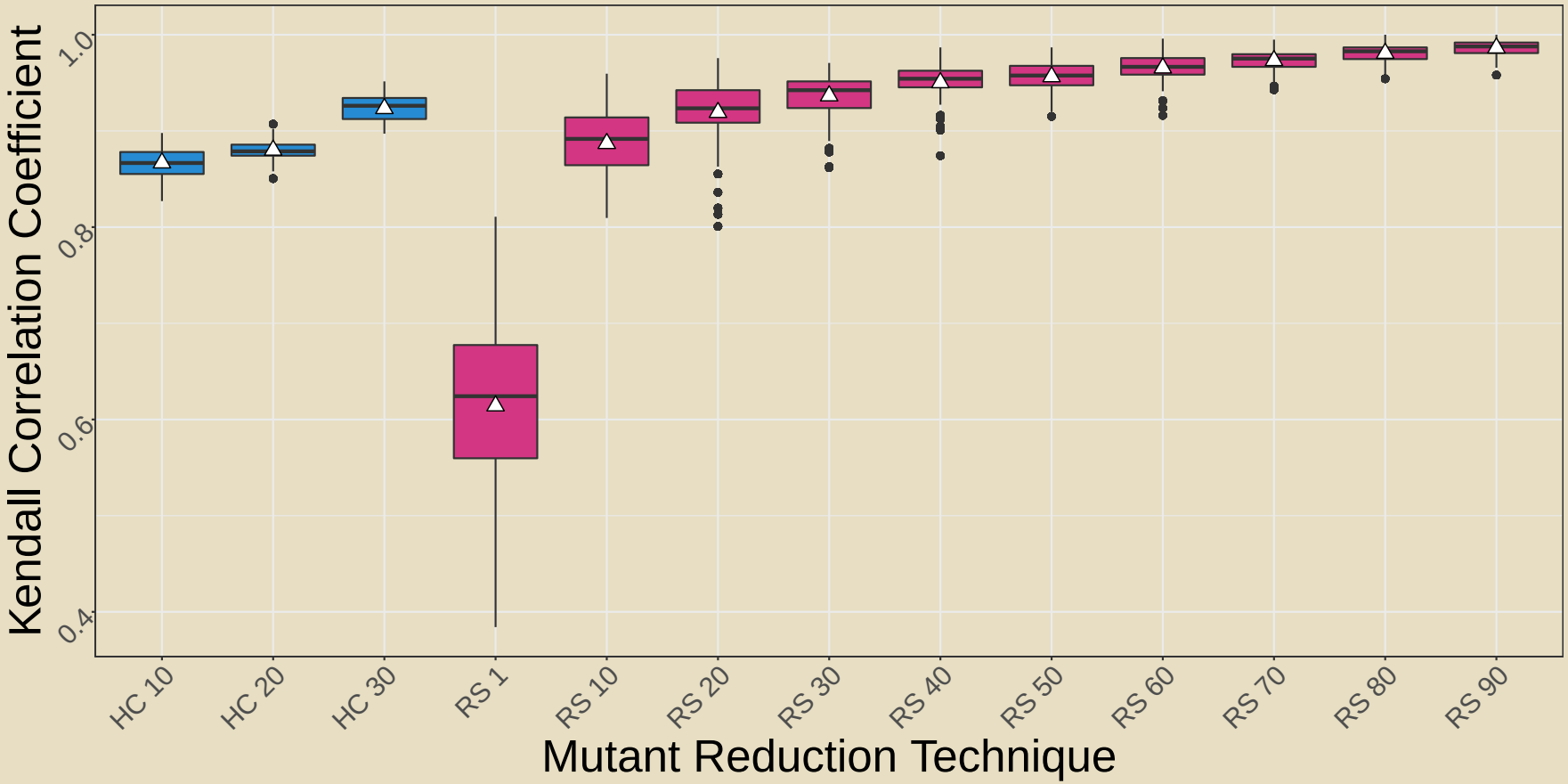
Evaluation Metrics

$$Ratio = \frac{\text{Mean Corr}}{1 - \text{Mean CR}}$$

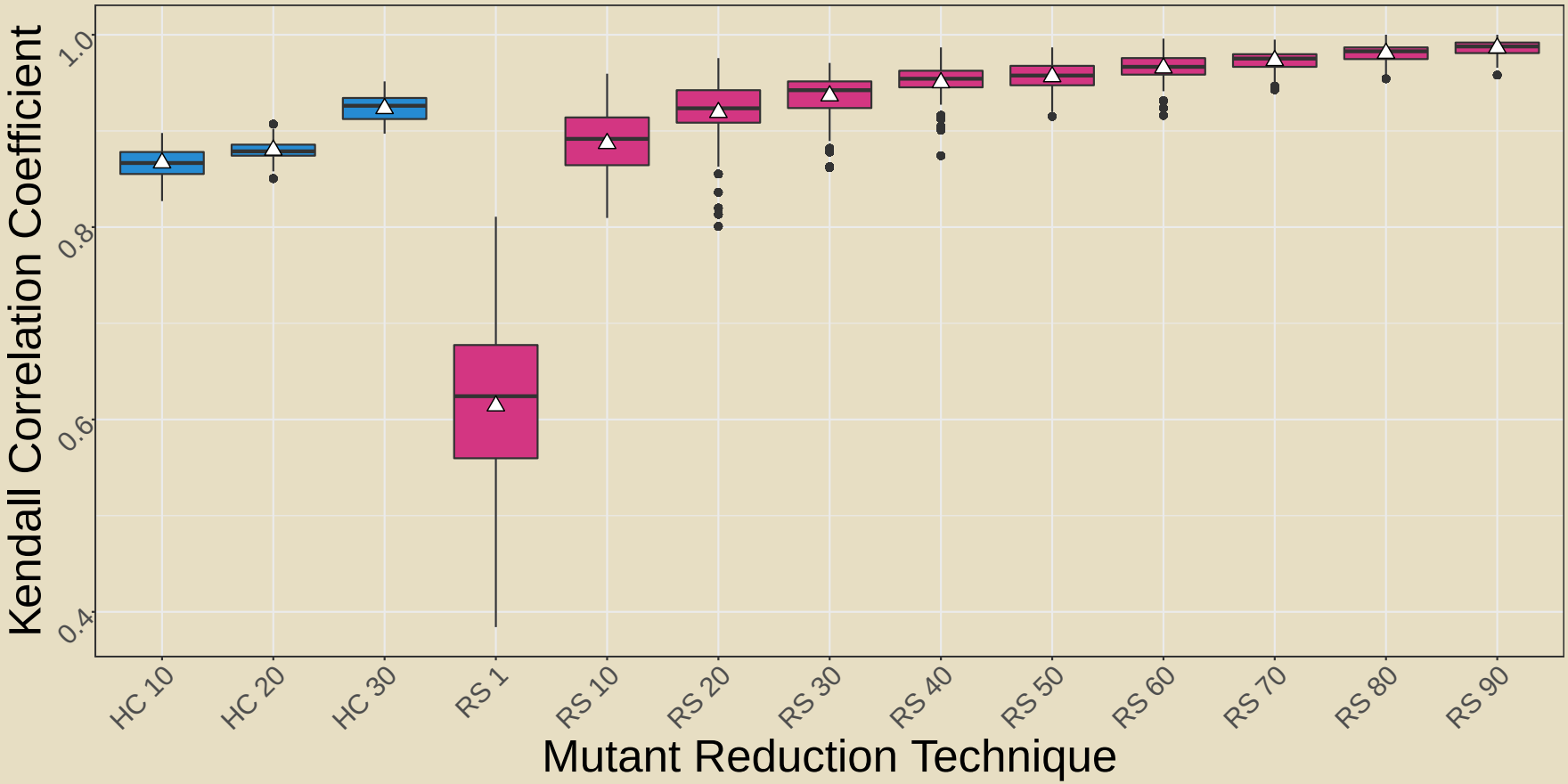
$Ratio \in [0, \infty)$

HIB

Empirical Results

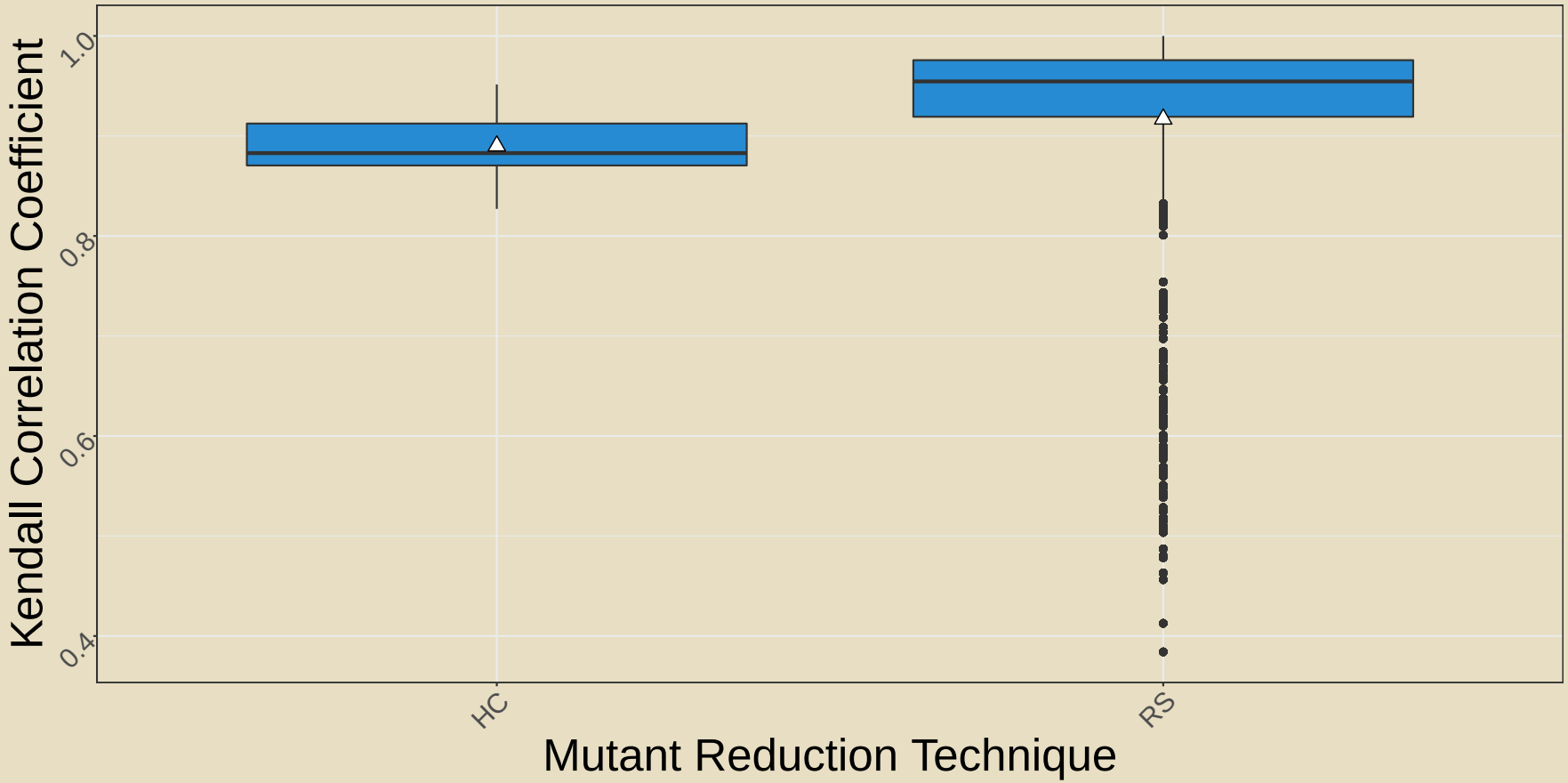


Empirical Results

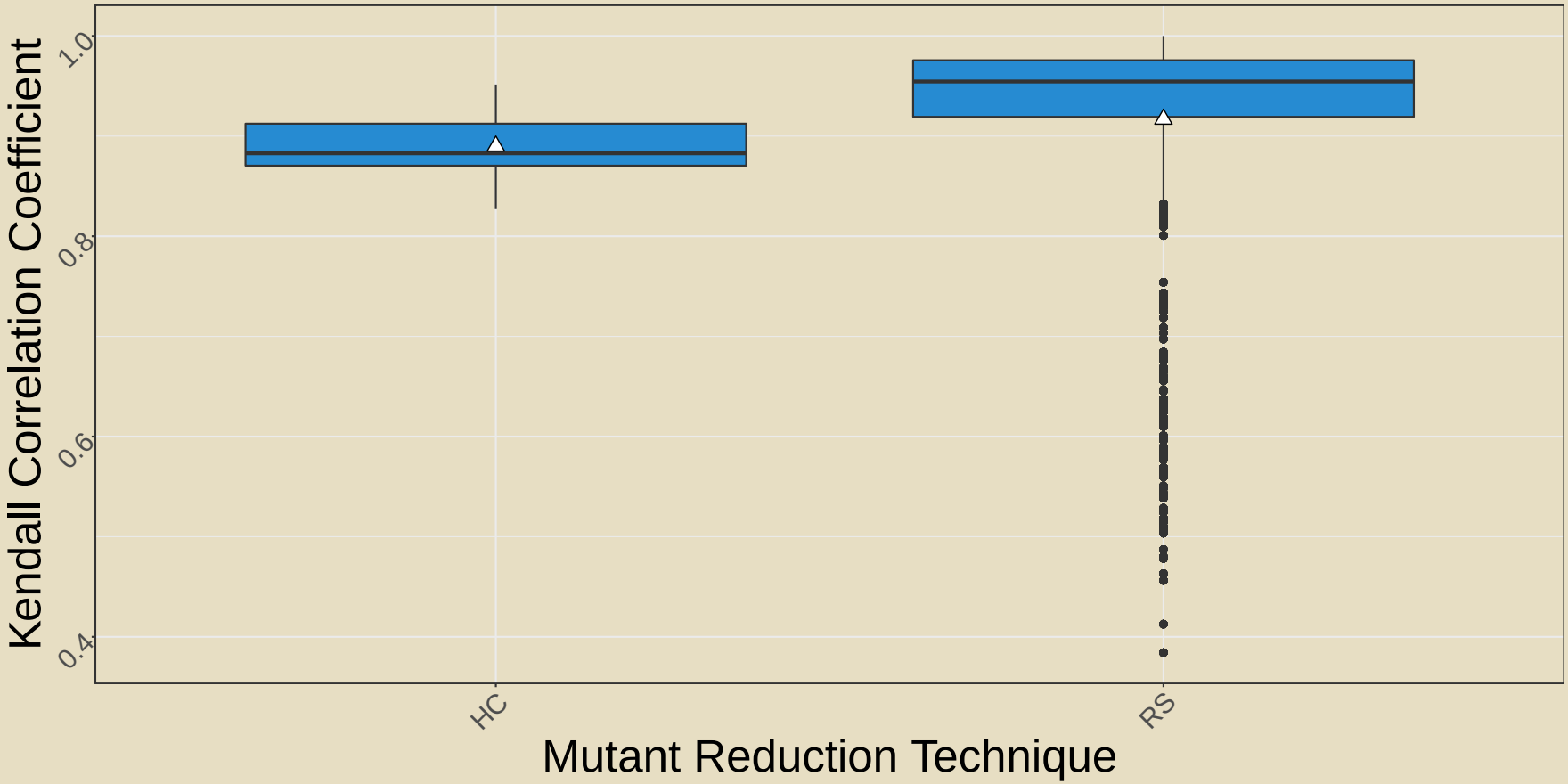


Produce highly correlated reduced sets!

Empirical Results

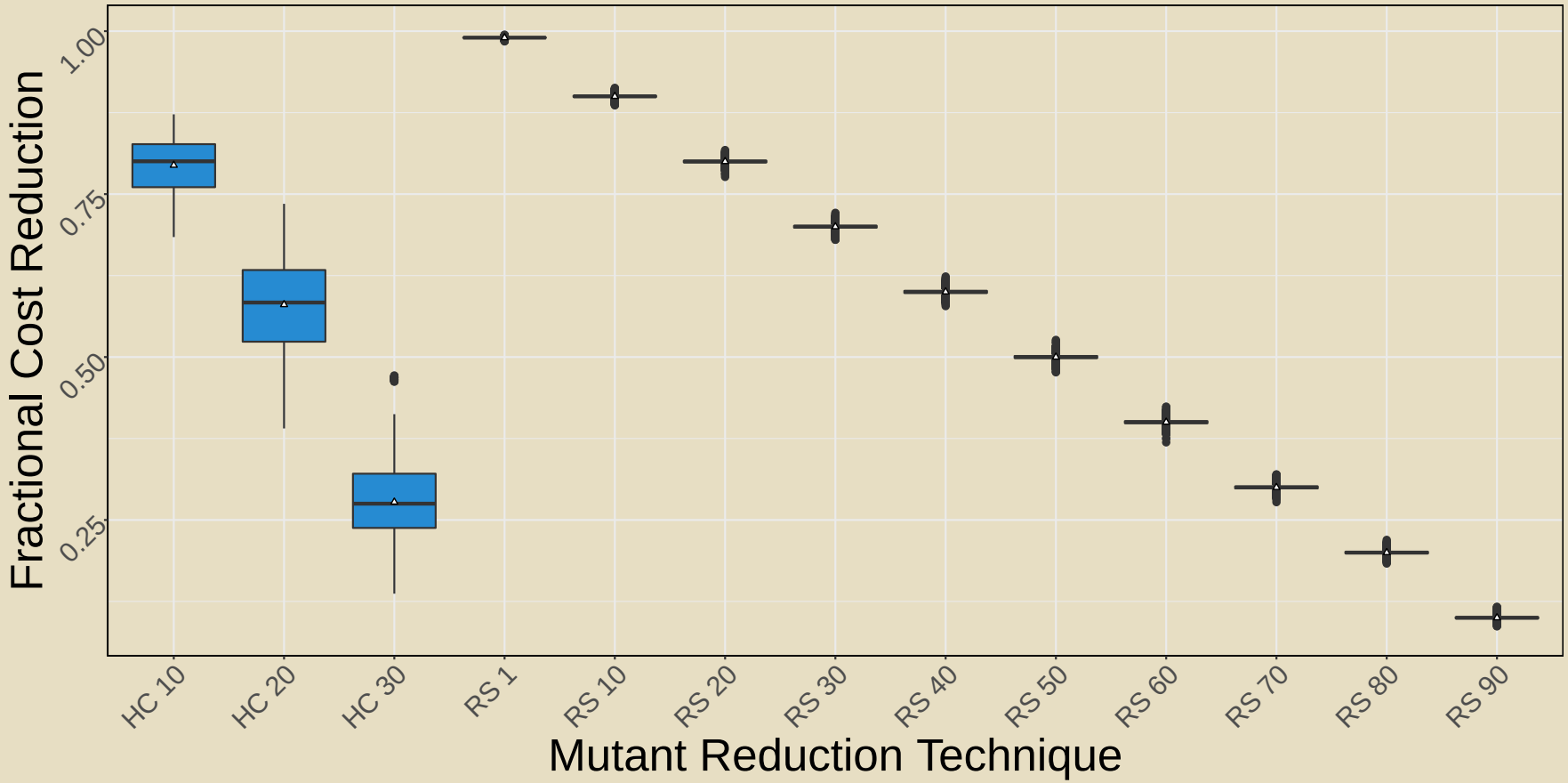


Empirical Results

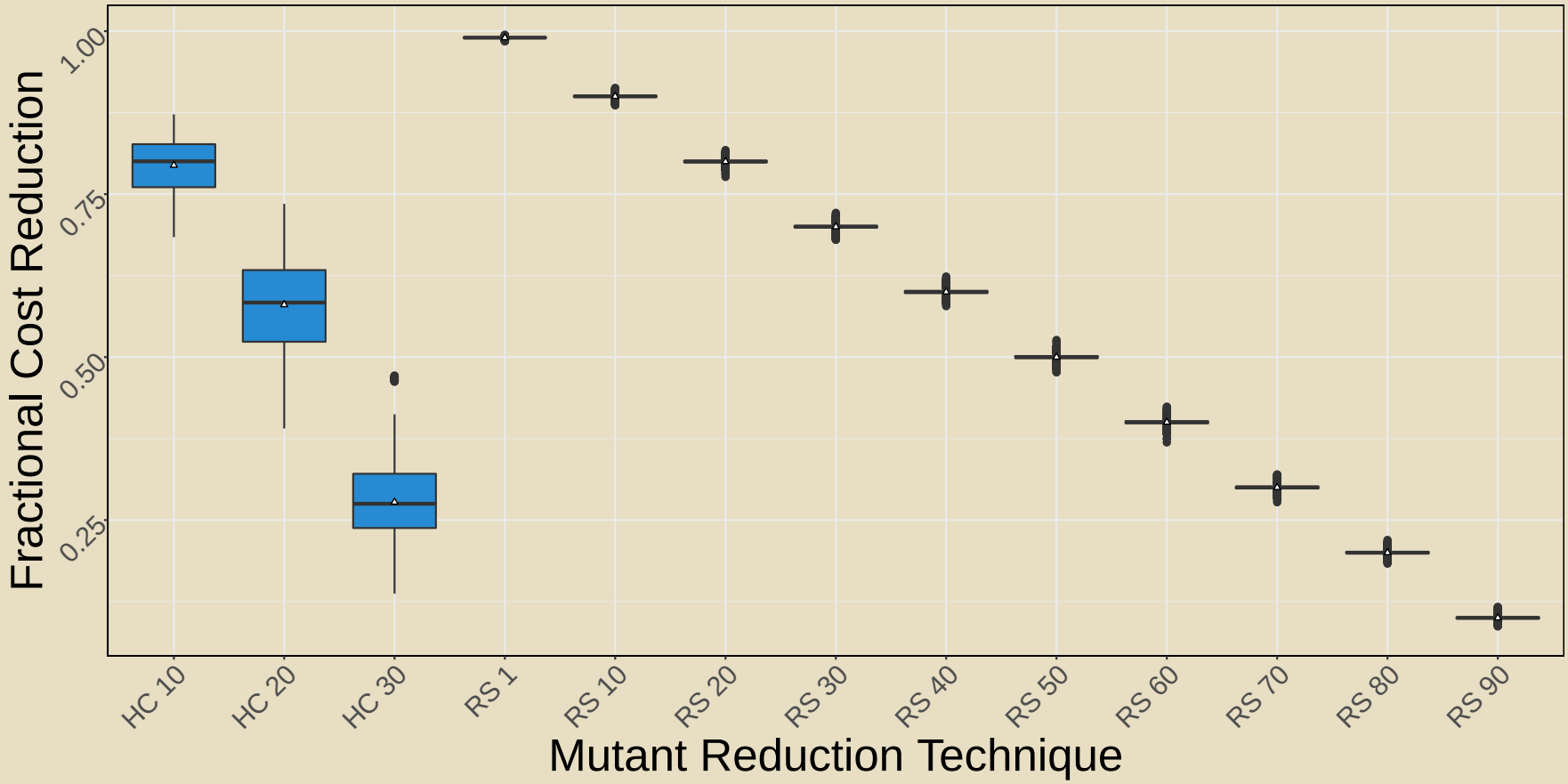


Random is better at
producing highly correlated sets!

Empirical Results

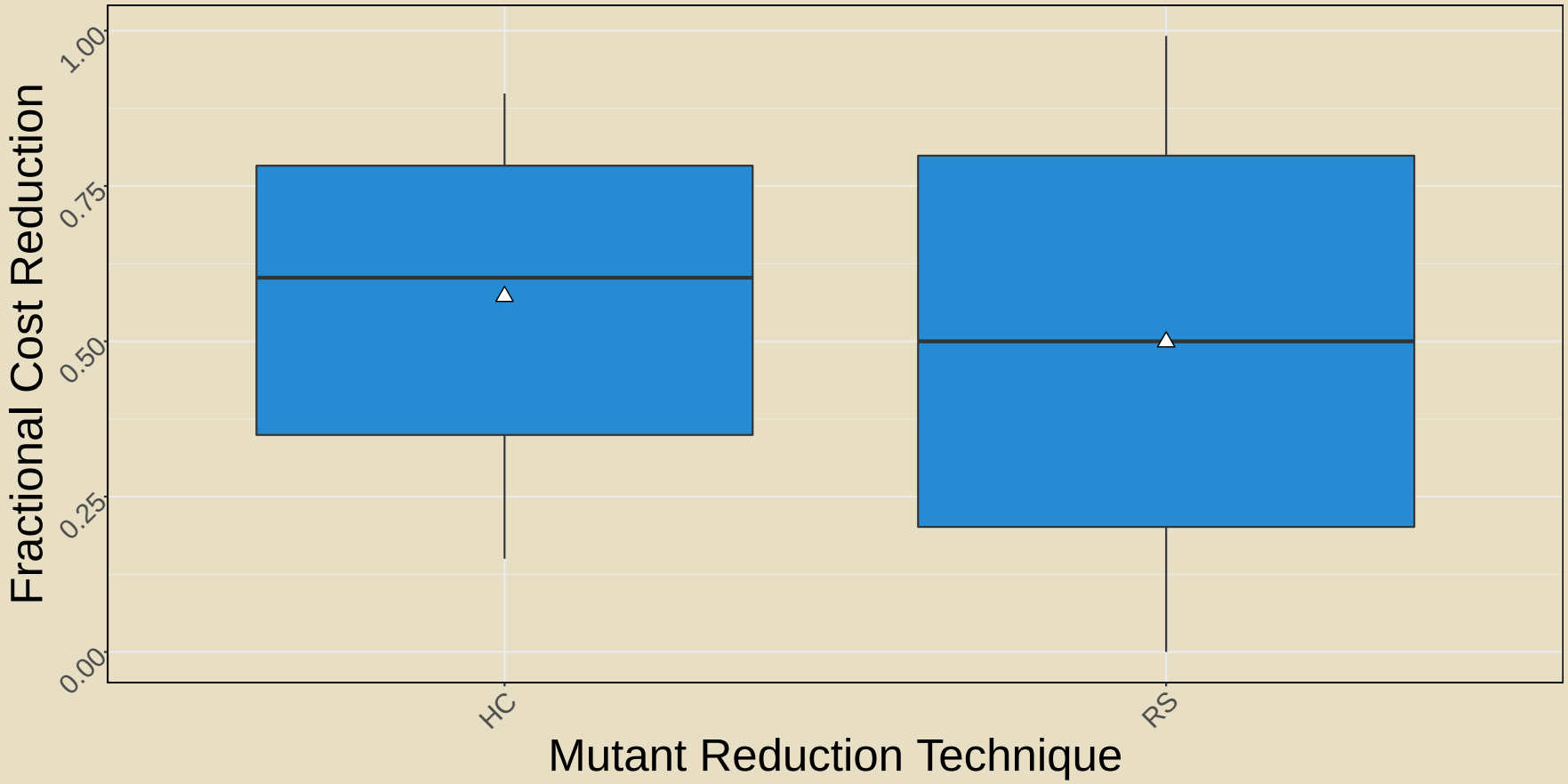


Empirical Results

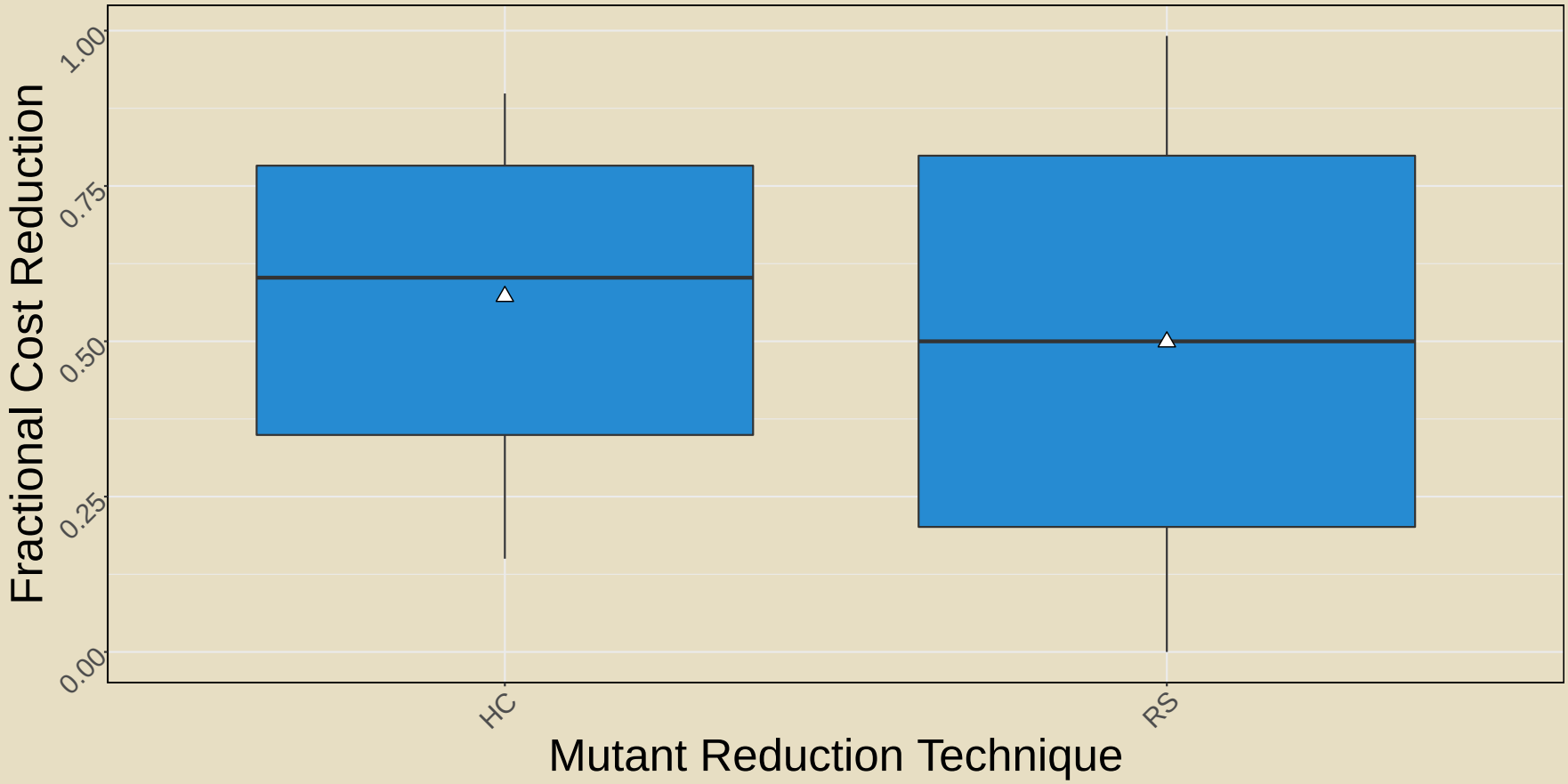


HC 30 produces highly correlated sets, but is not good at reducing cost.

Empirical Results

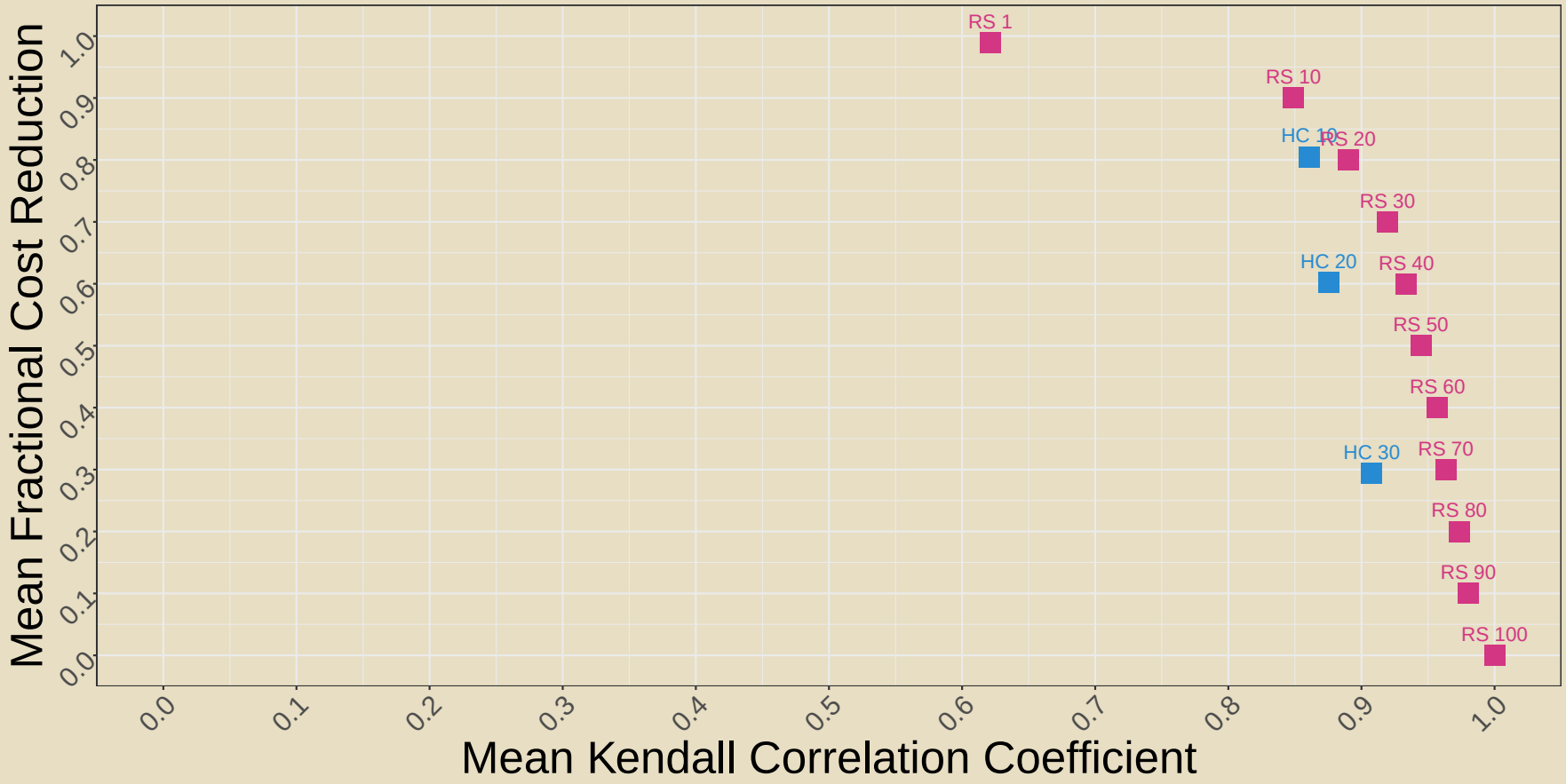


Empirical Results

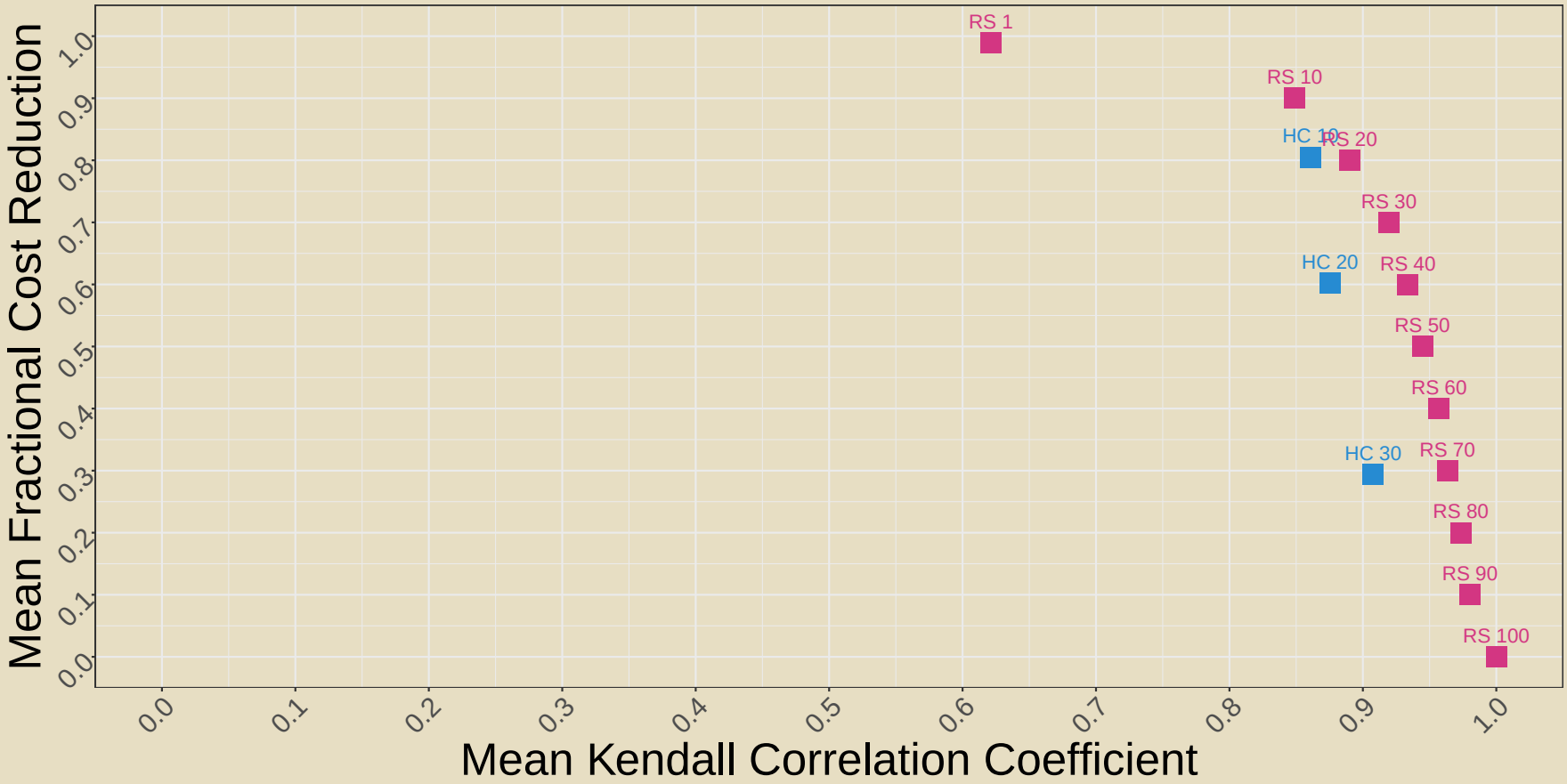


Overall, HC performs better at reducing cost!

Empirical Results

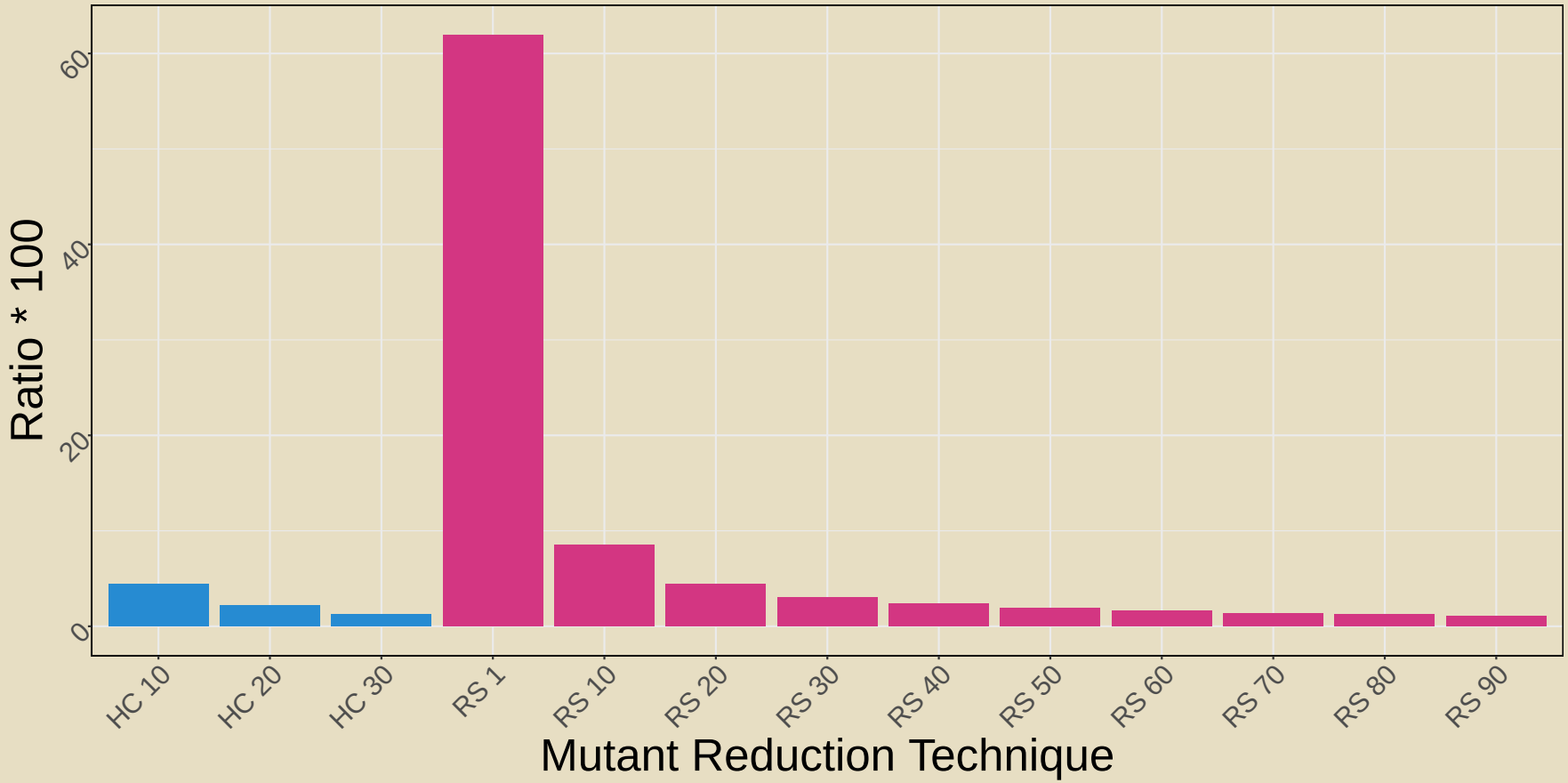


Empirical Results

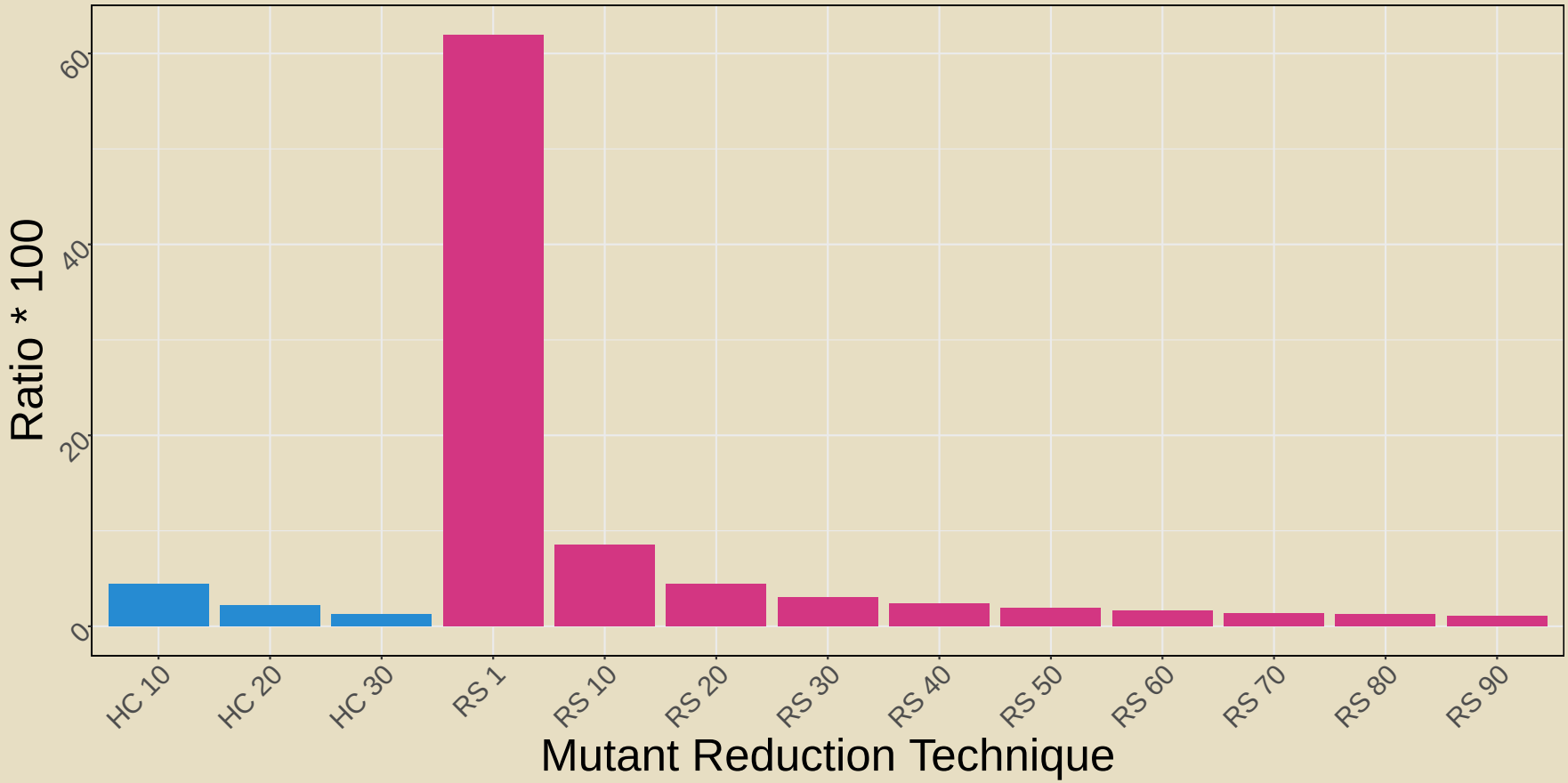


HC 10 is slightly better at reducing cost, while RS 20 is moderately better at producing correlated sets!

Empirical Results

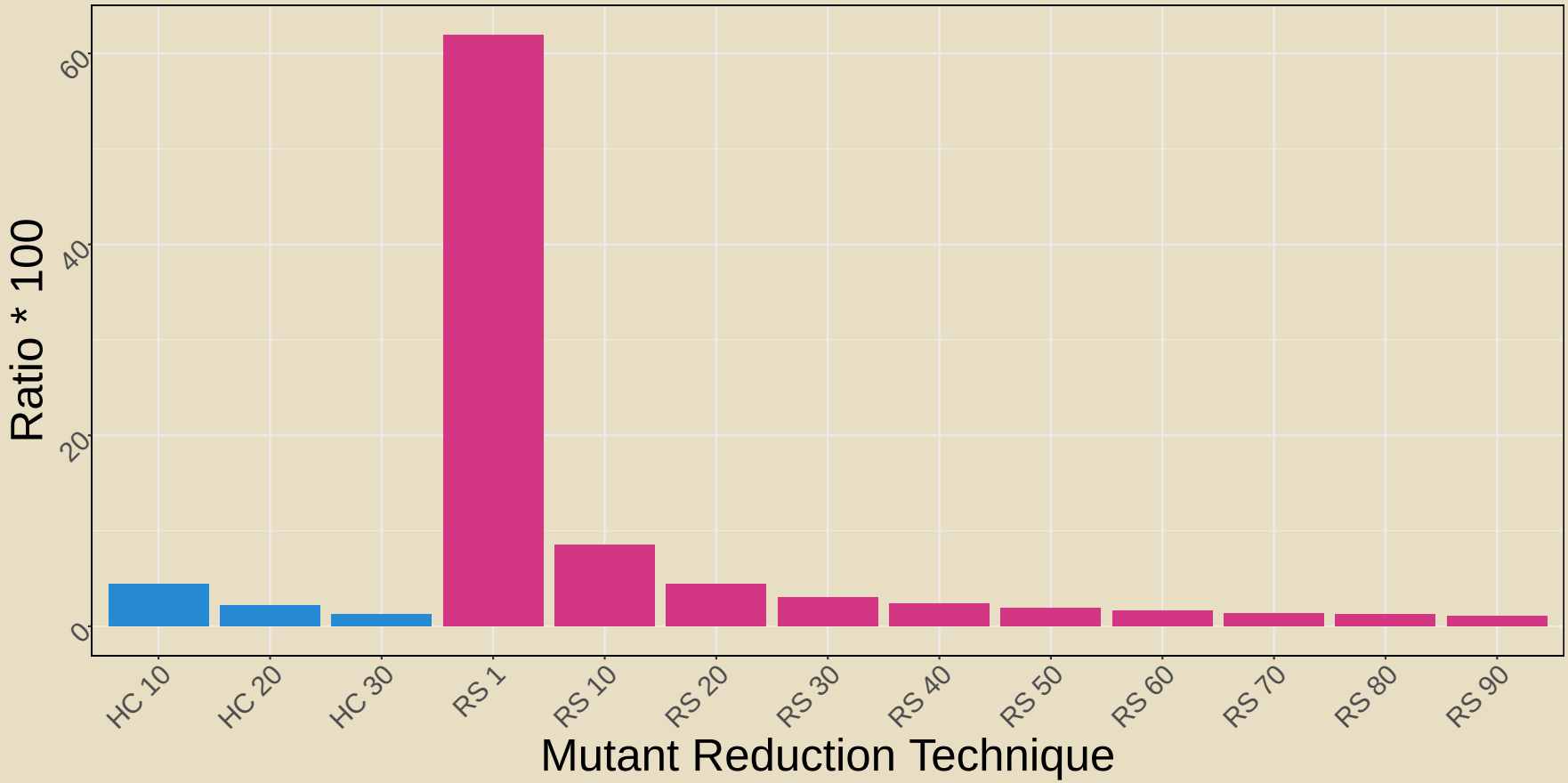


Empirical Results



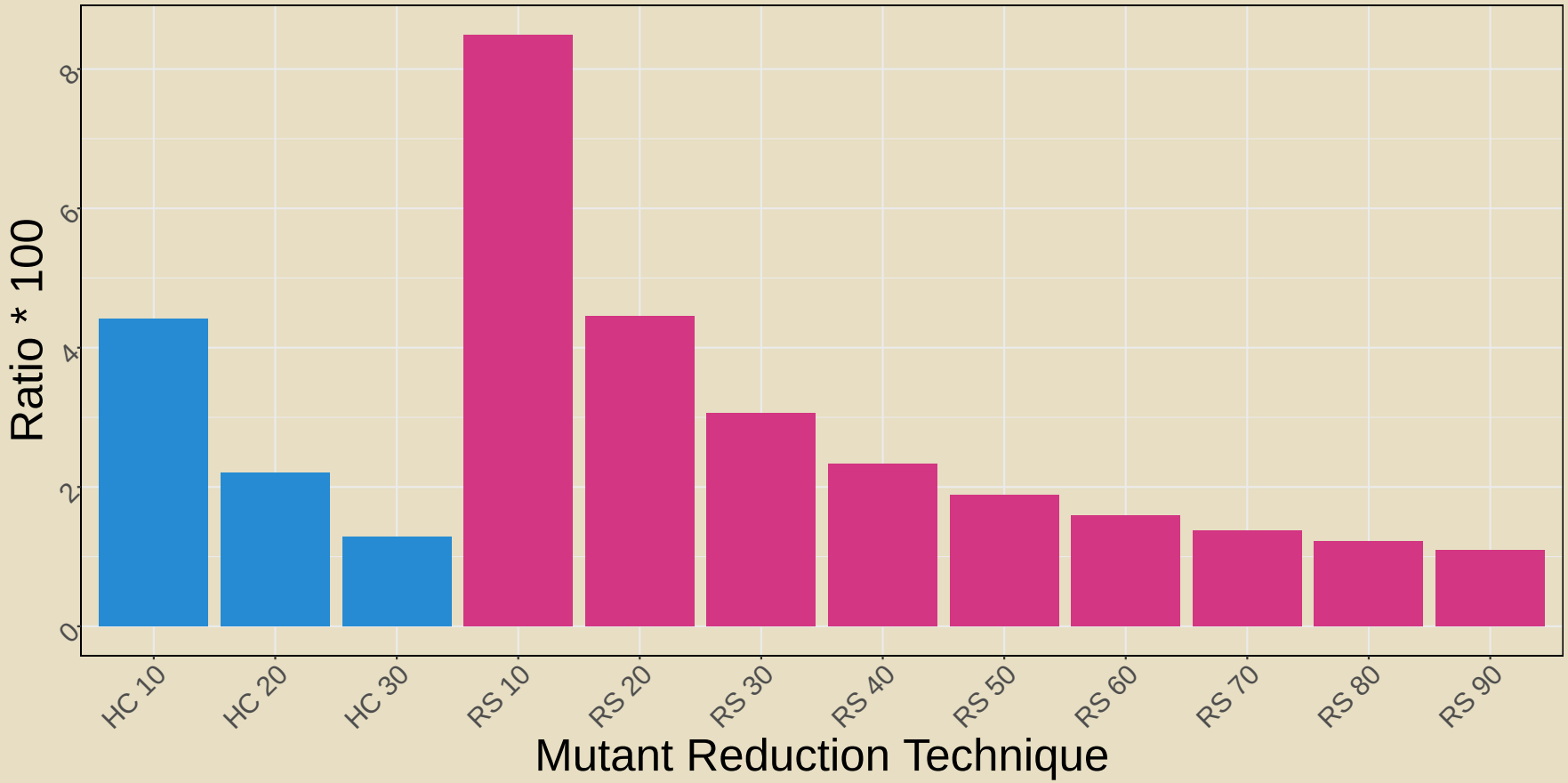
RS 1 is amazing!

Empirical Results

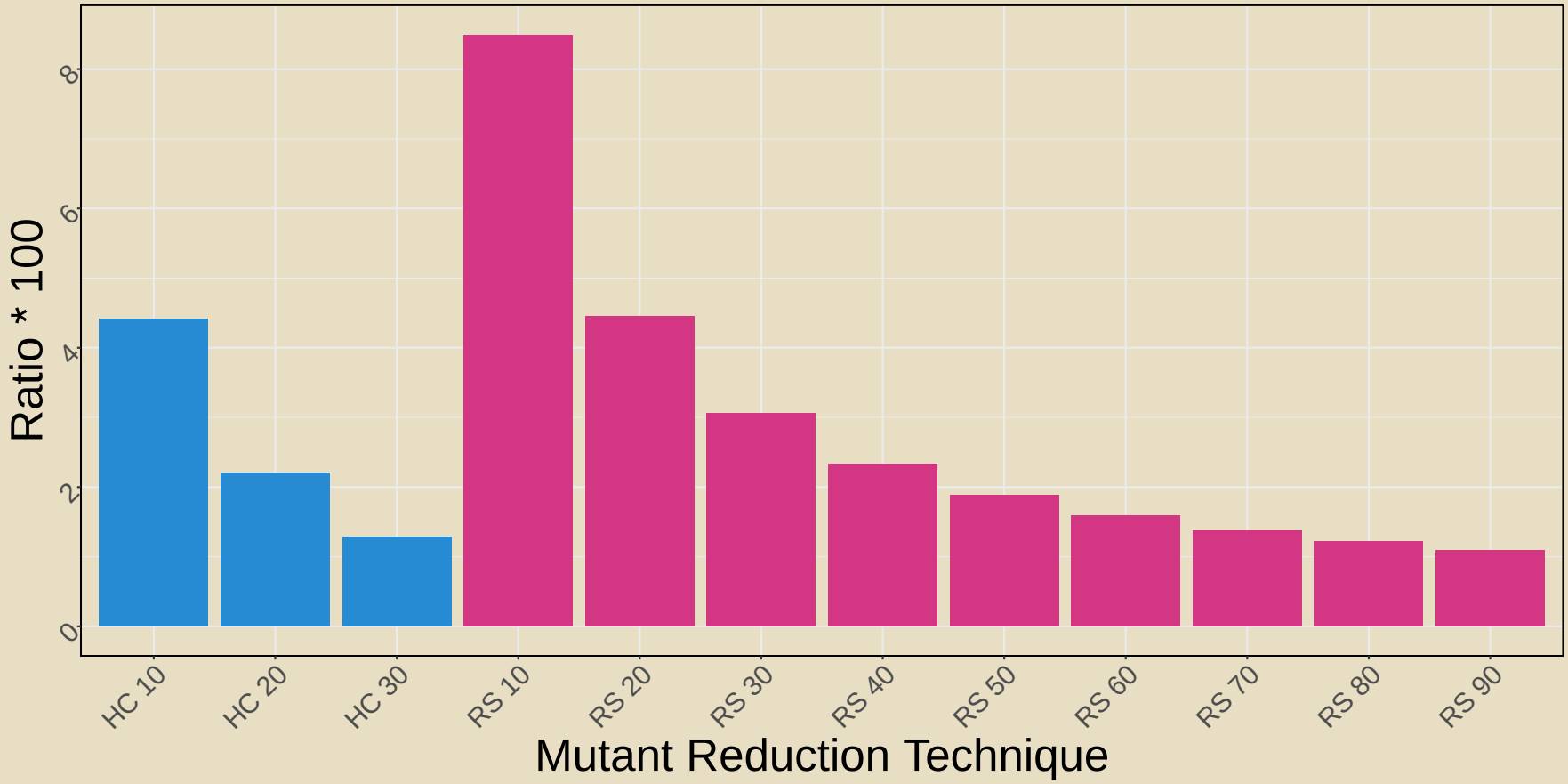


RS 1 is amazing … at reducing costs.

Empirical Results

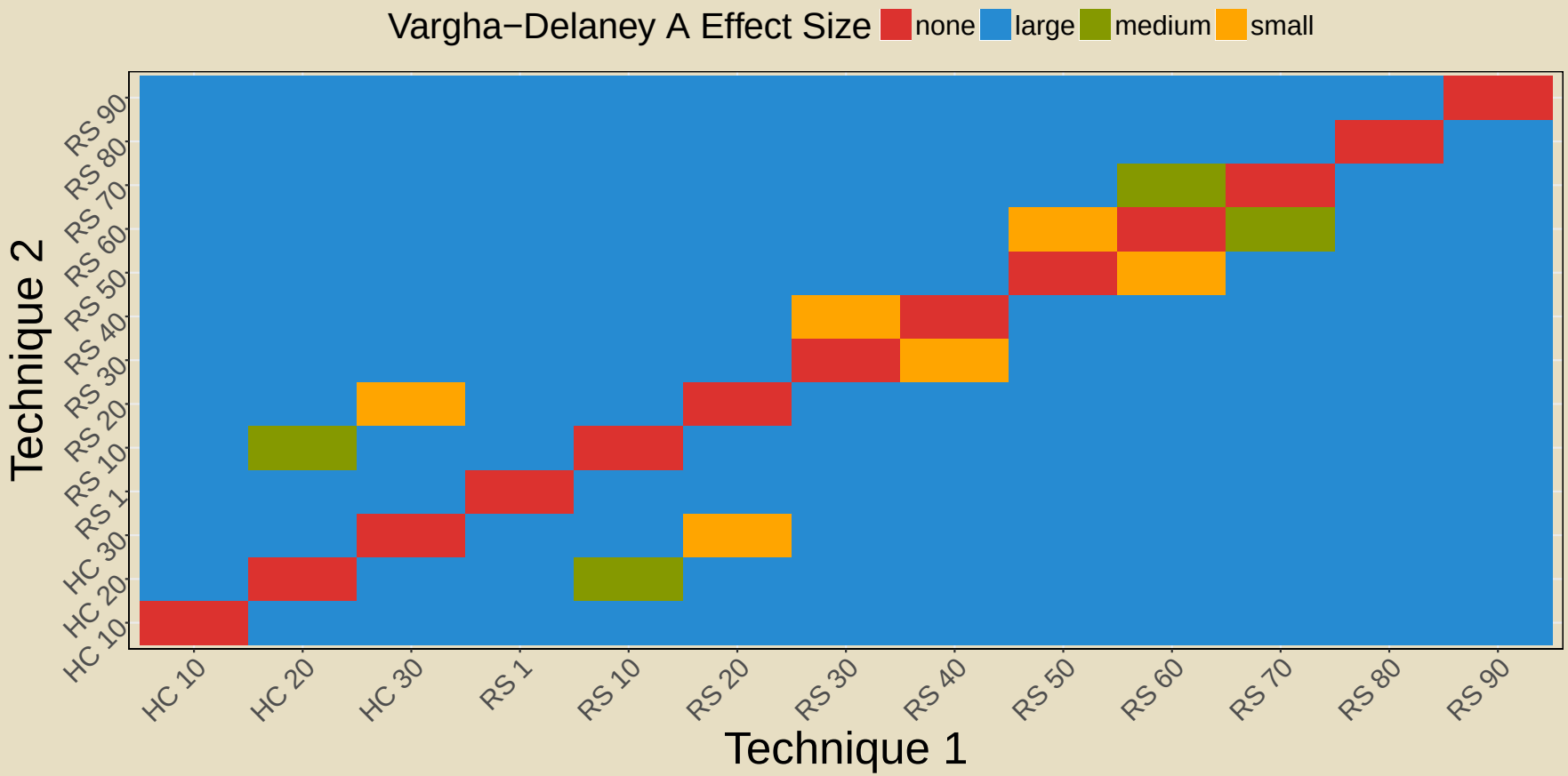


Empirical Results

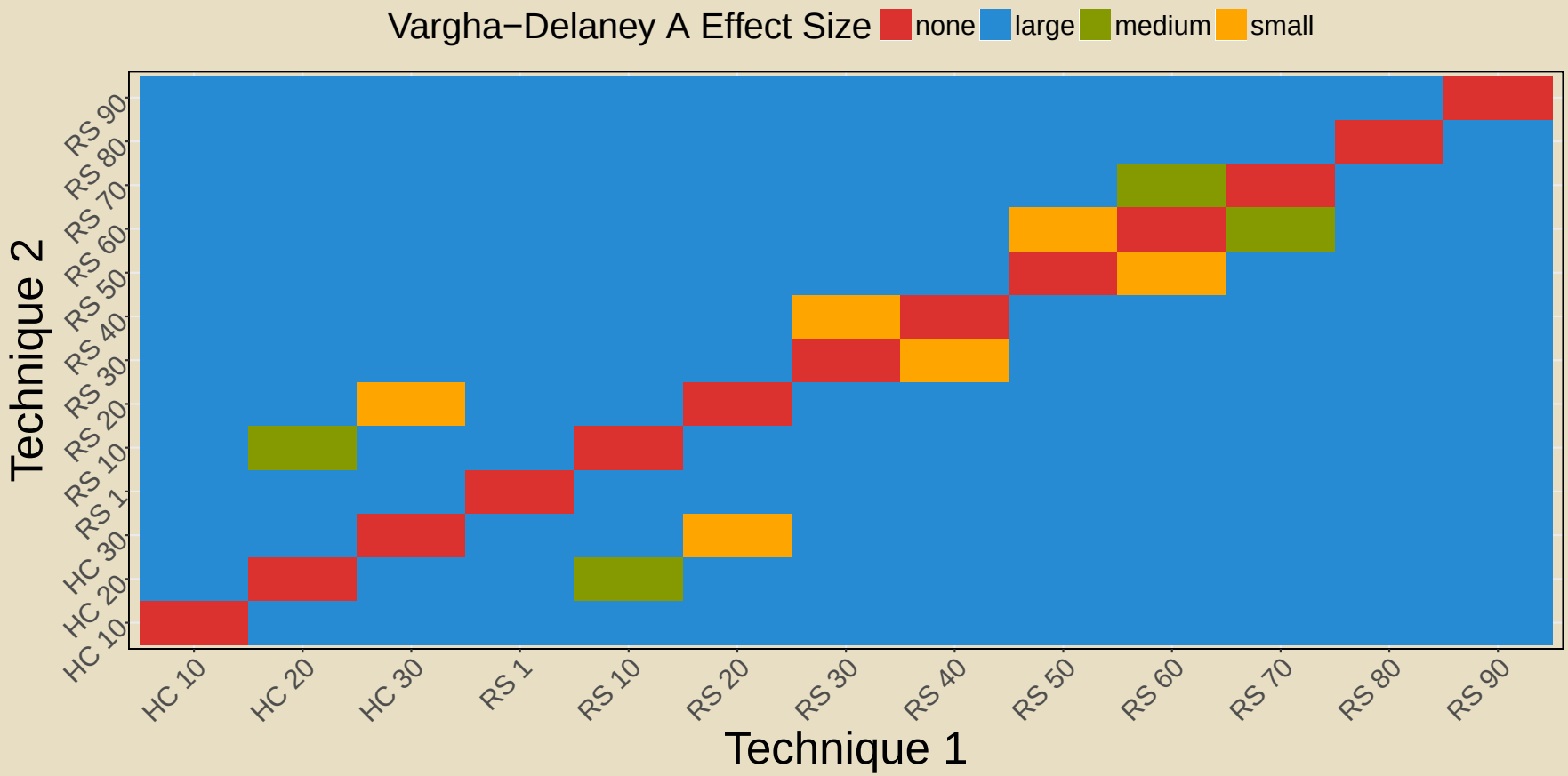


Our clear winner is RS 10!

Empirical Results

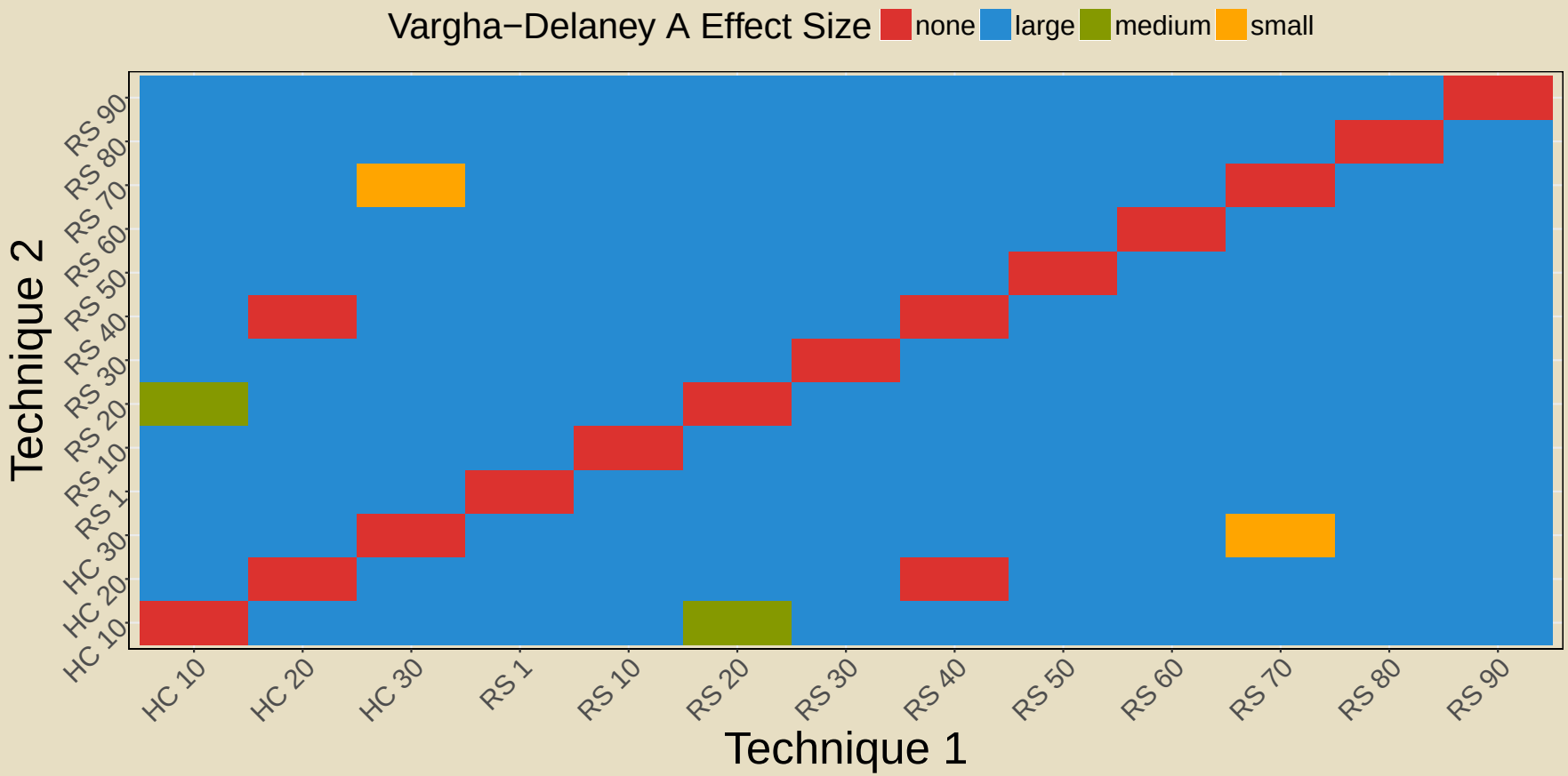


Empirical Results

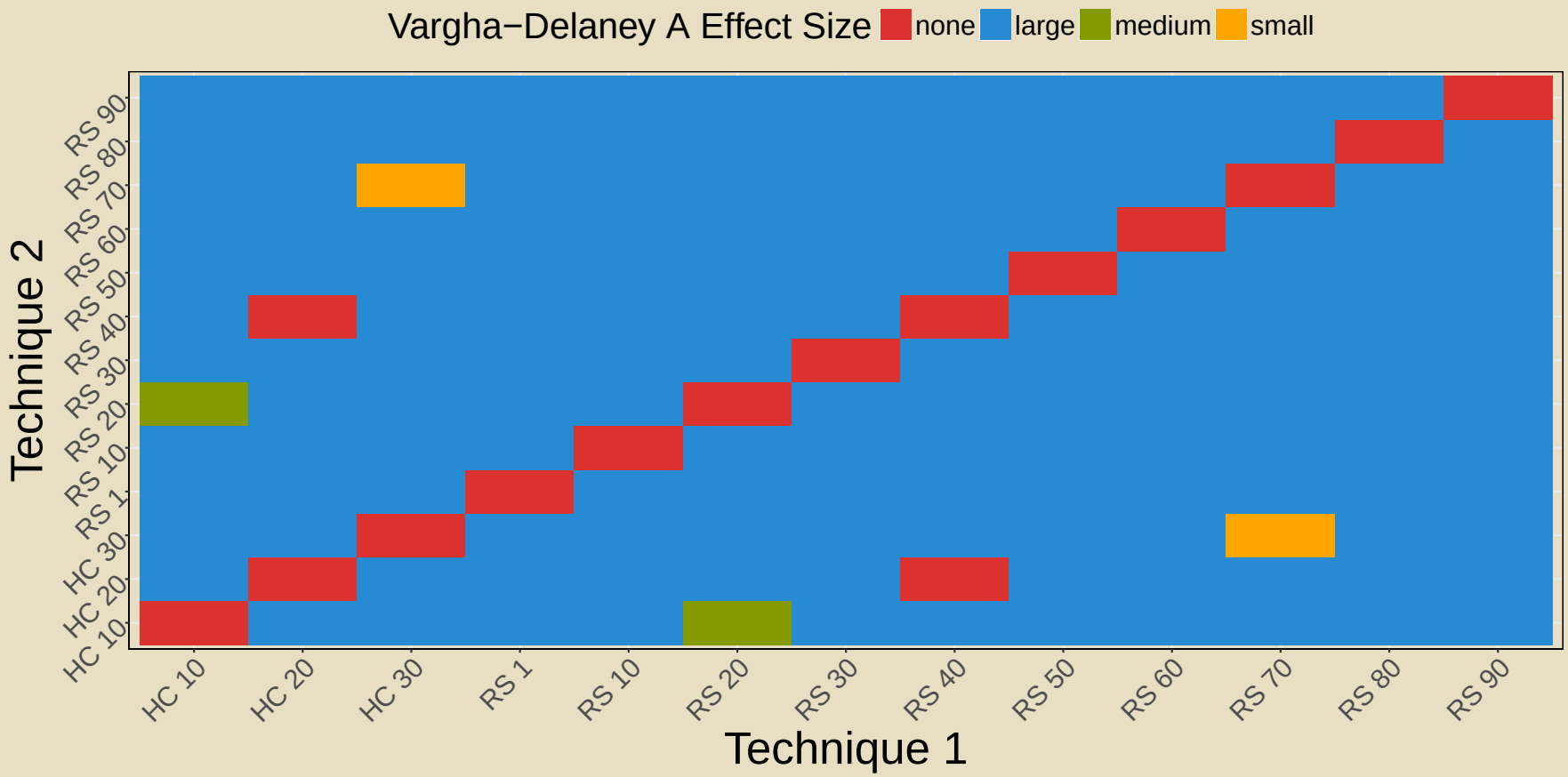


RS 10 is moderately better than HC 20 at producing correlated sets.

Empirical Results

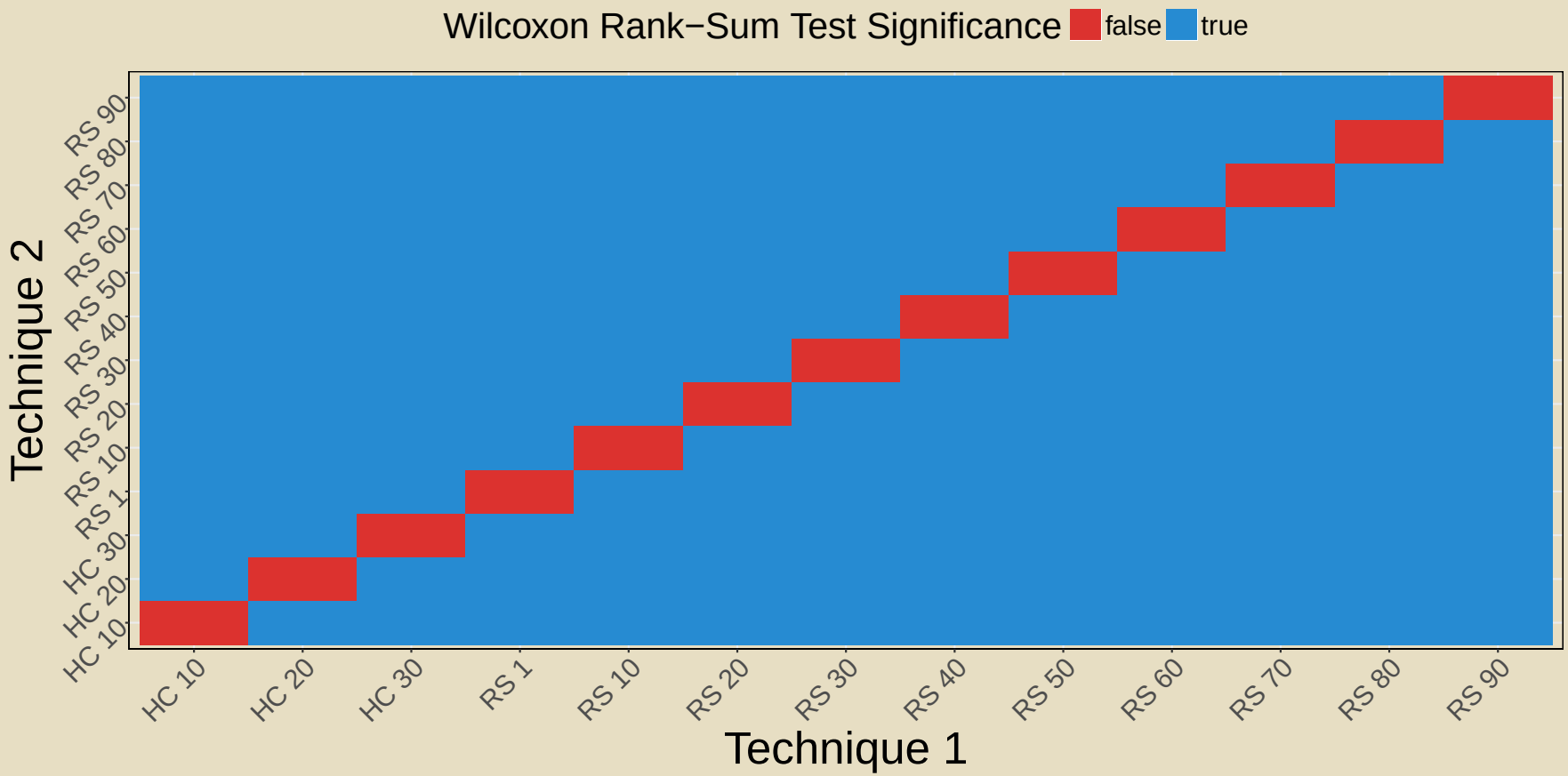


Empirical Results

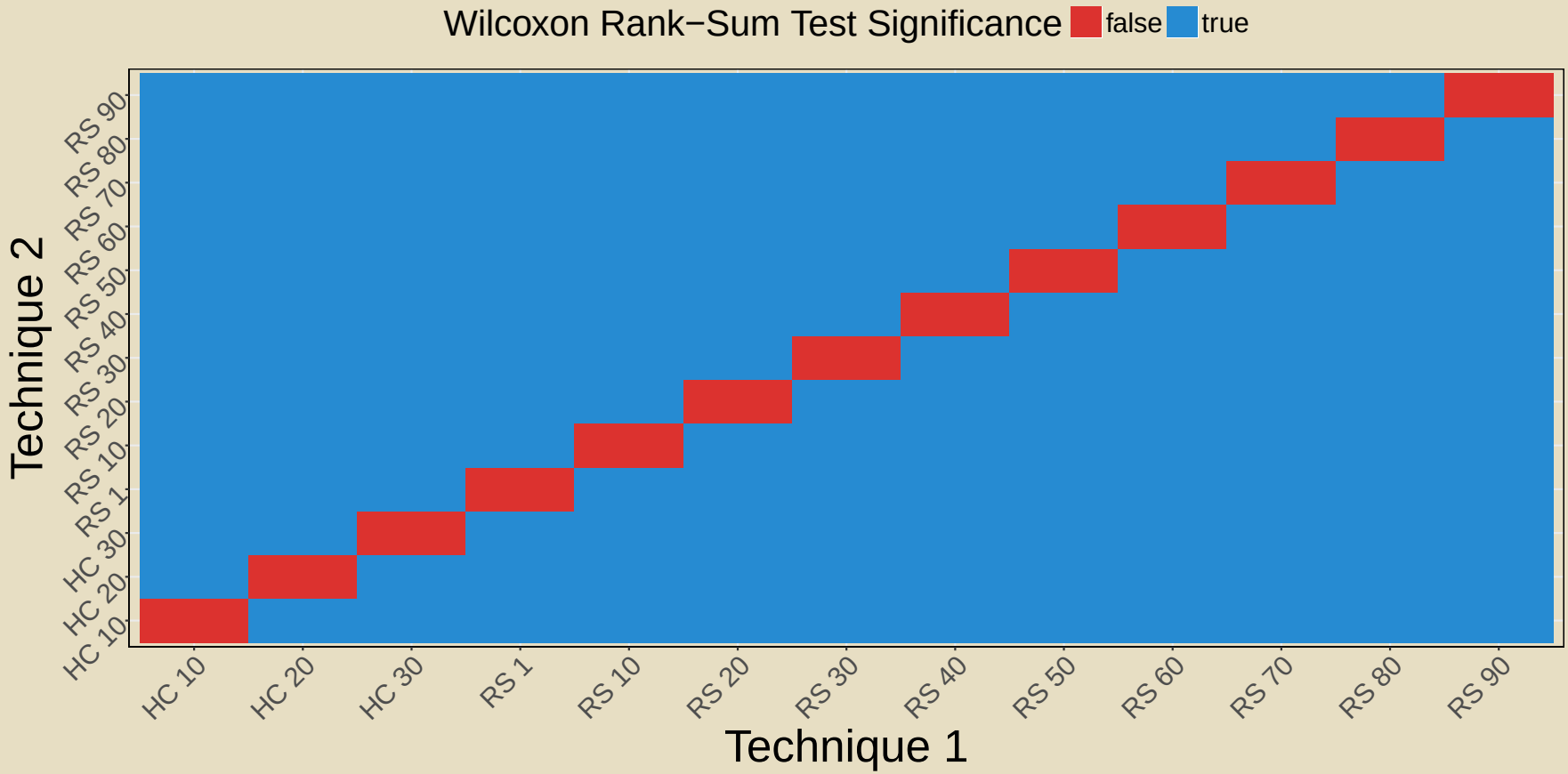


But, RS 10 is much better
than HC 20 at reducing cost!

Empirical Results

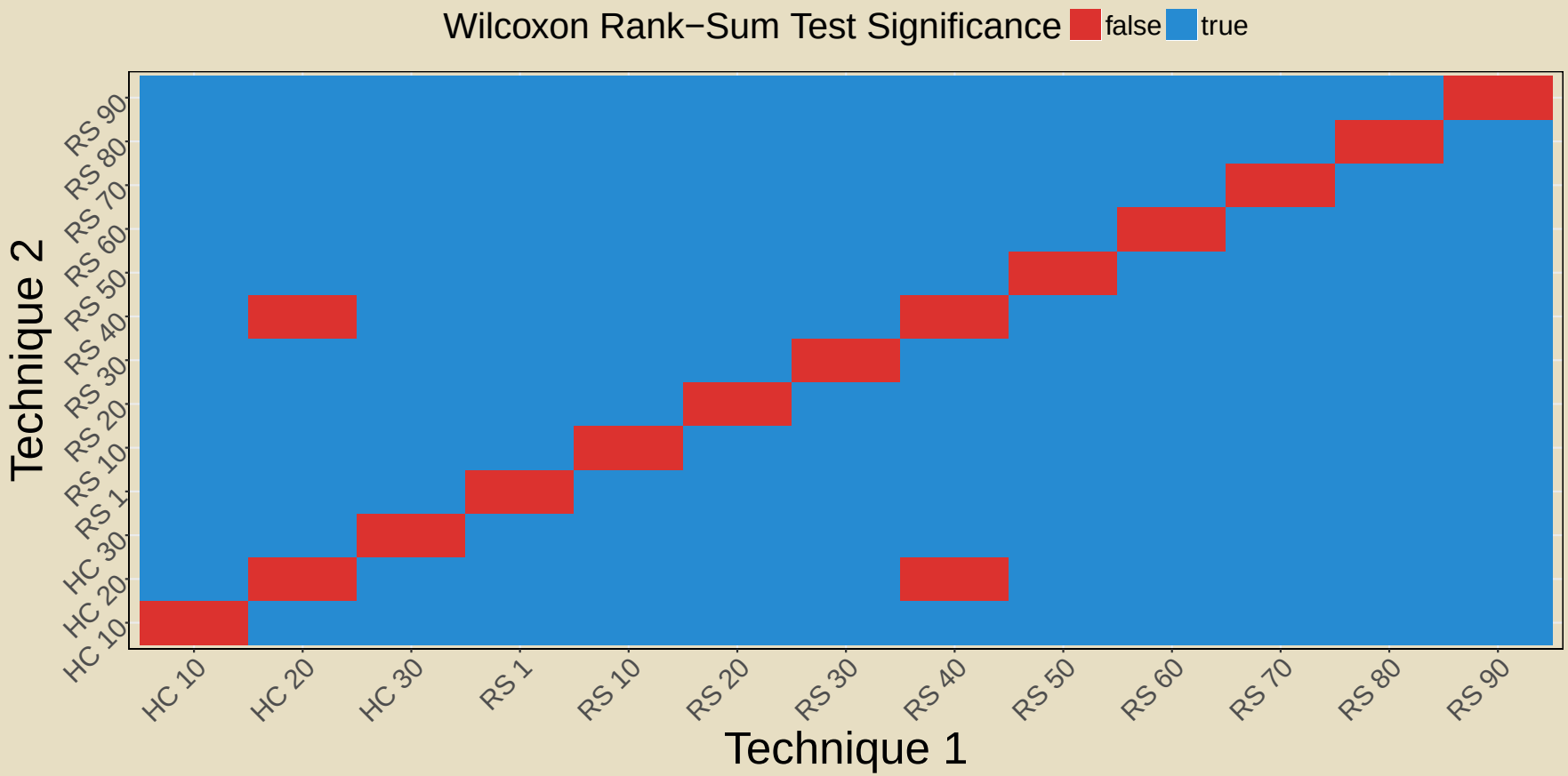


Empirical Results

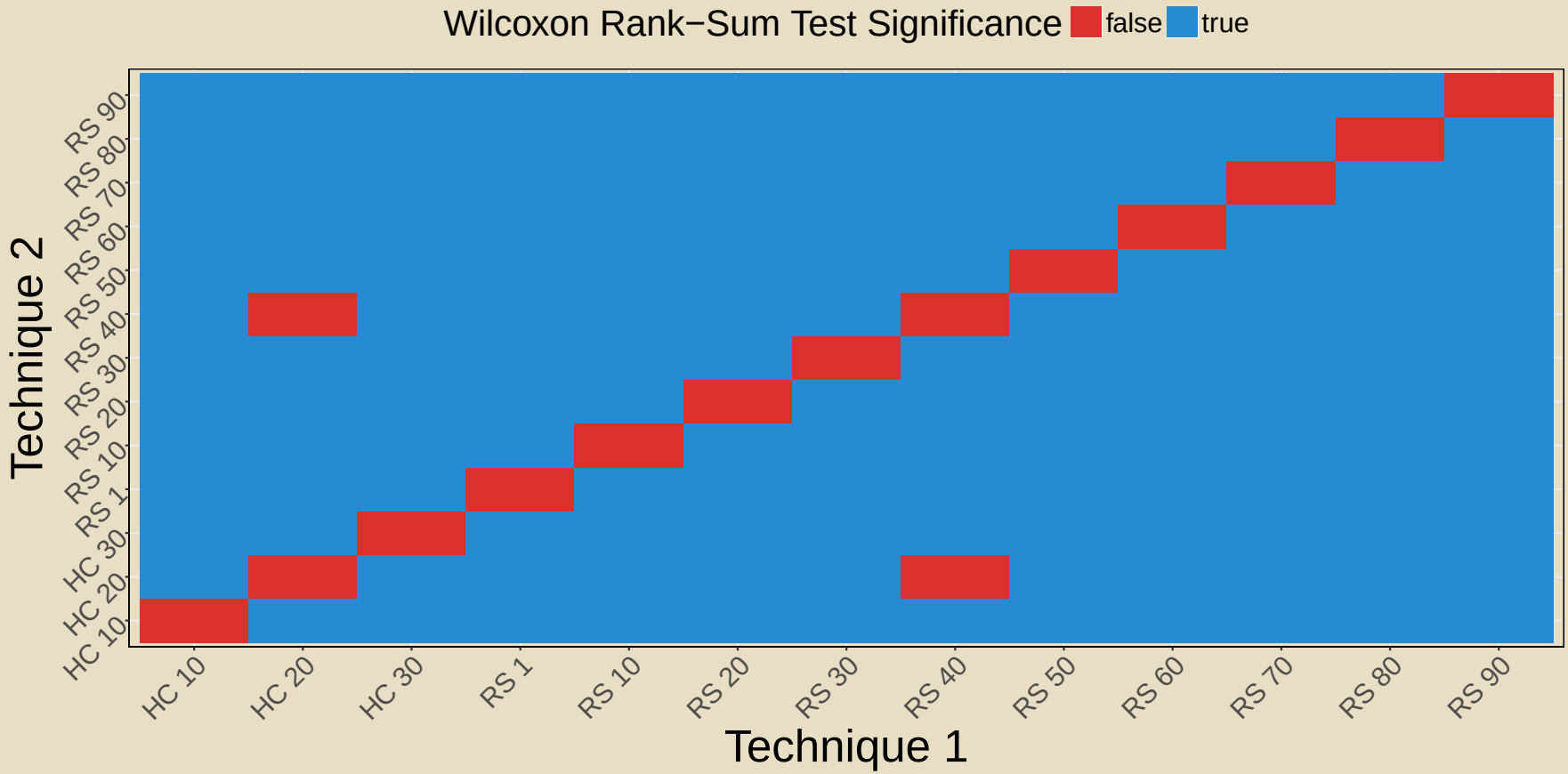


Correlation coefficient val-
ues were significantly different!

Empirical Results

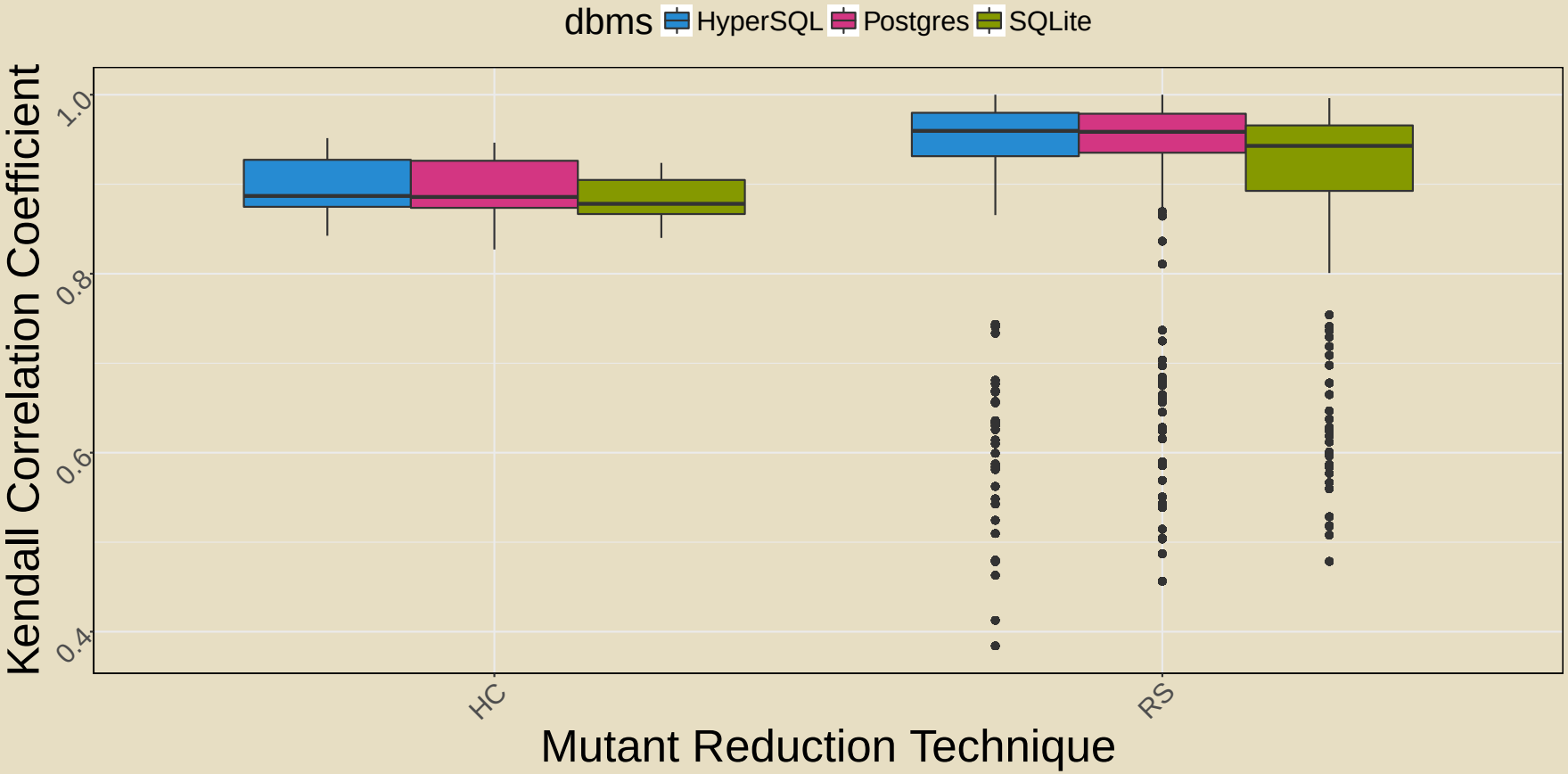


Empirical Results

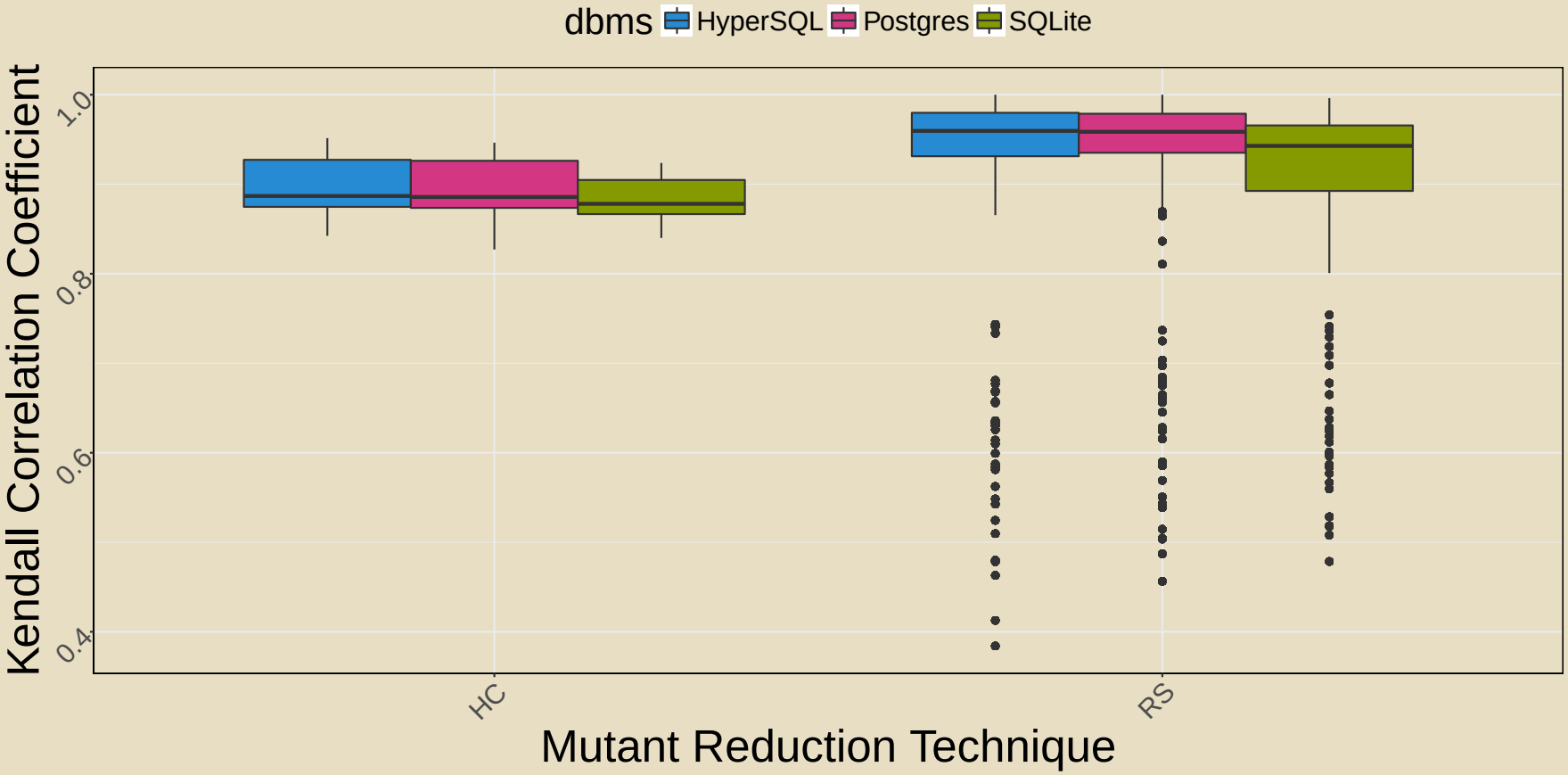


RS 40 is no better at reducing cost than HC 20!

Empirical Results

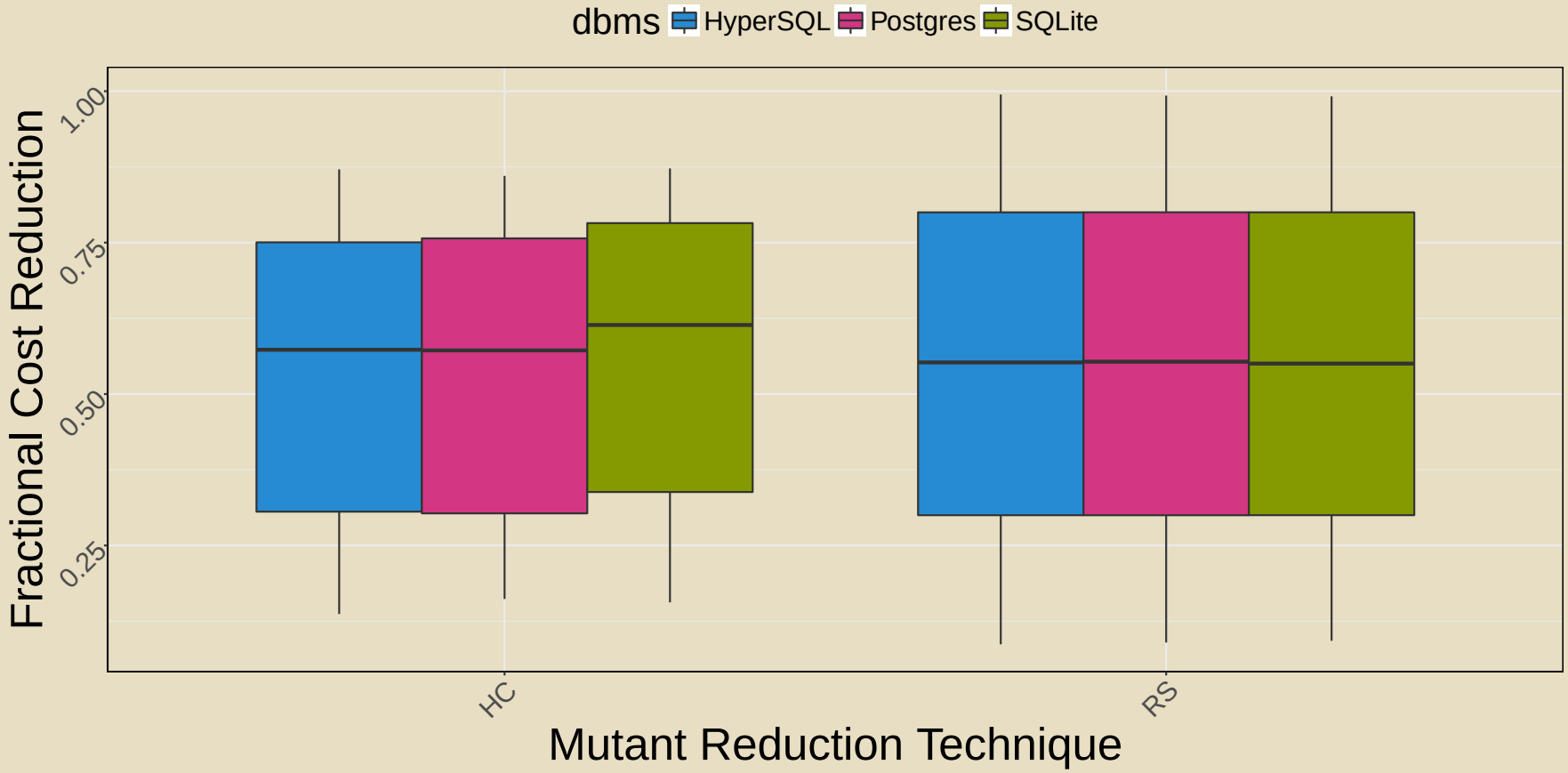


Empirical Results

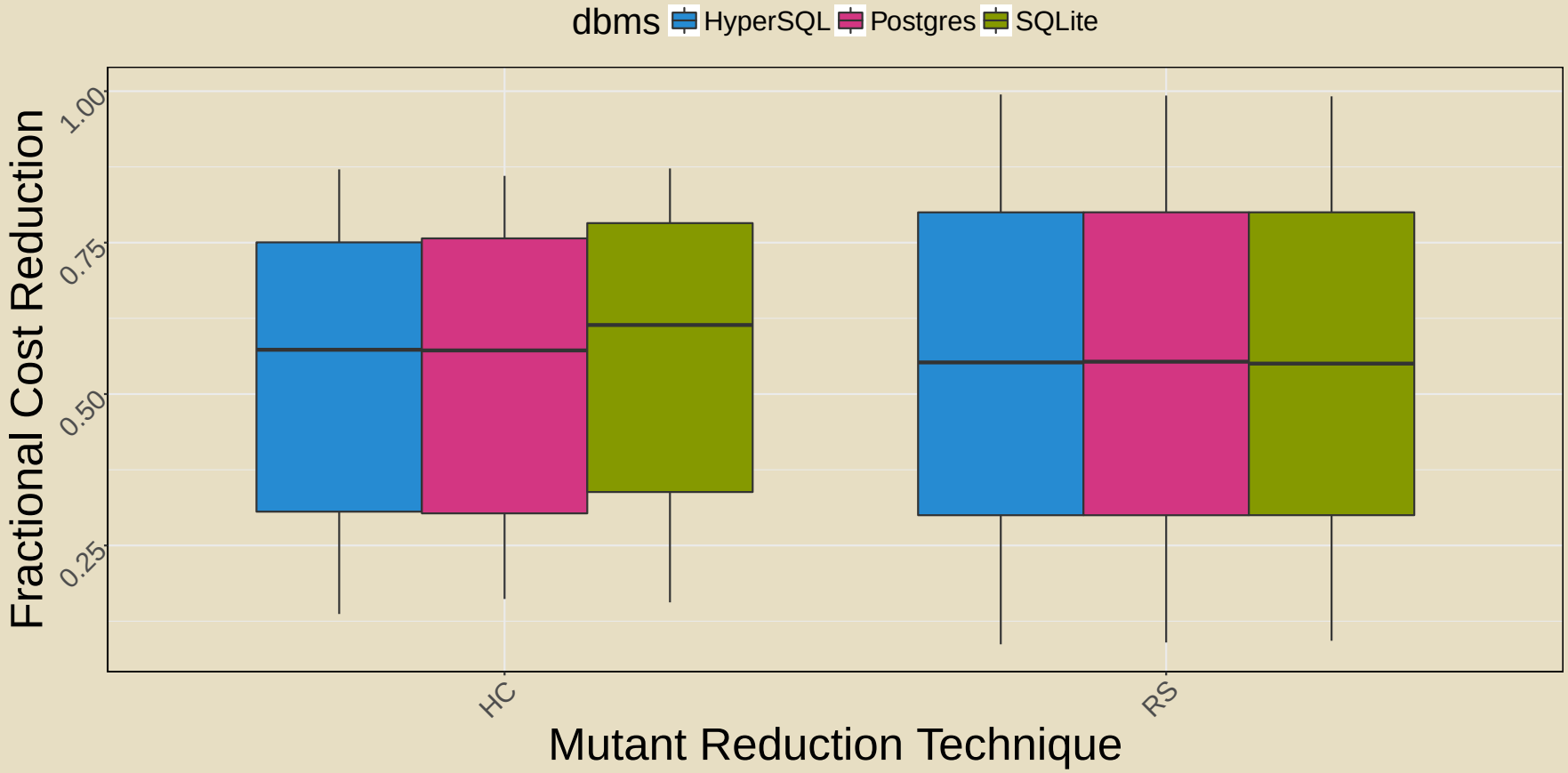


HC model is worse at producing correlated set for the DBMS from which it was built!

Empirical Results

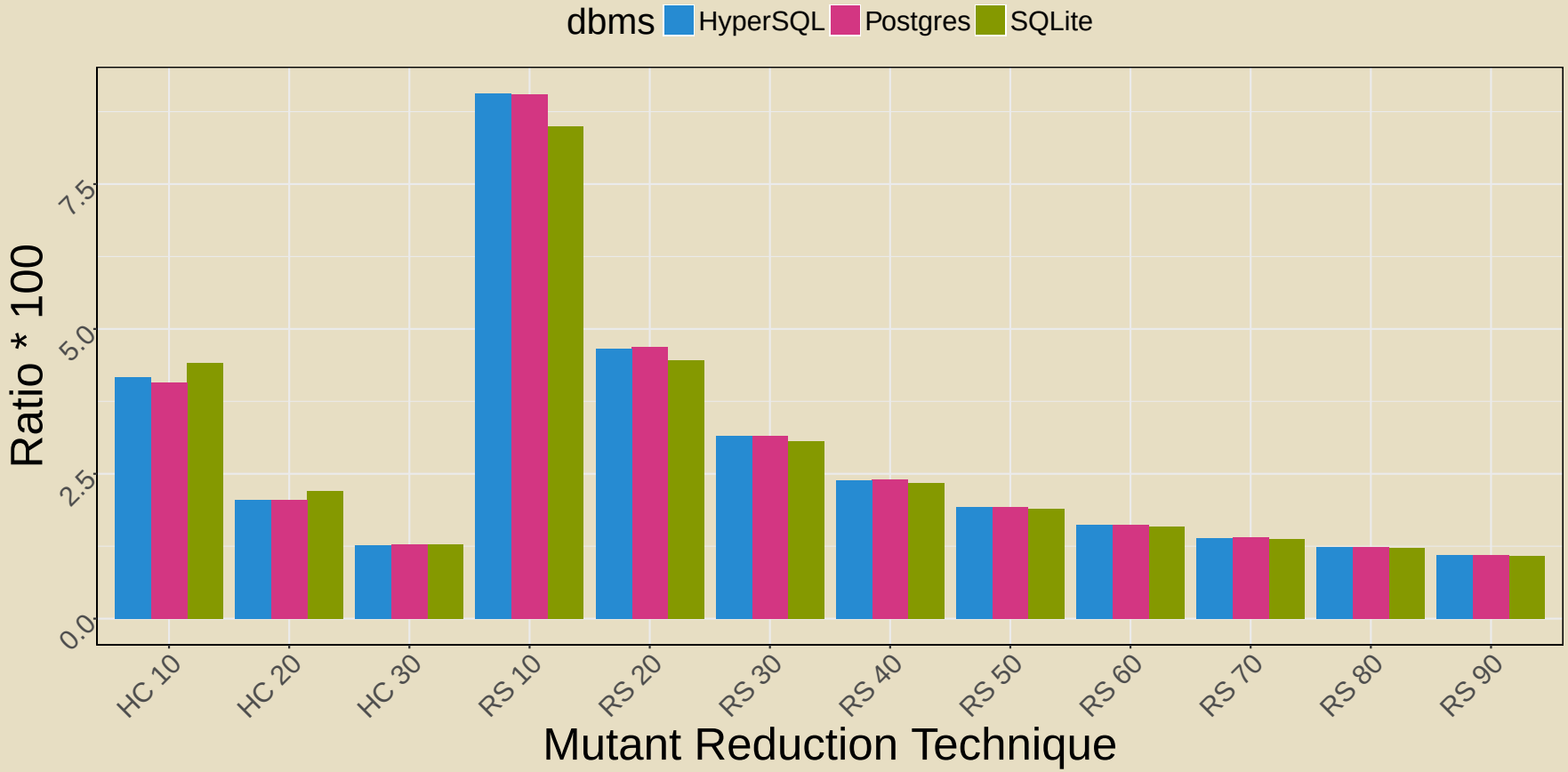


Empirical Results

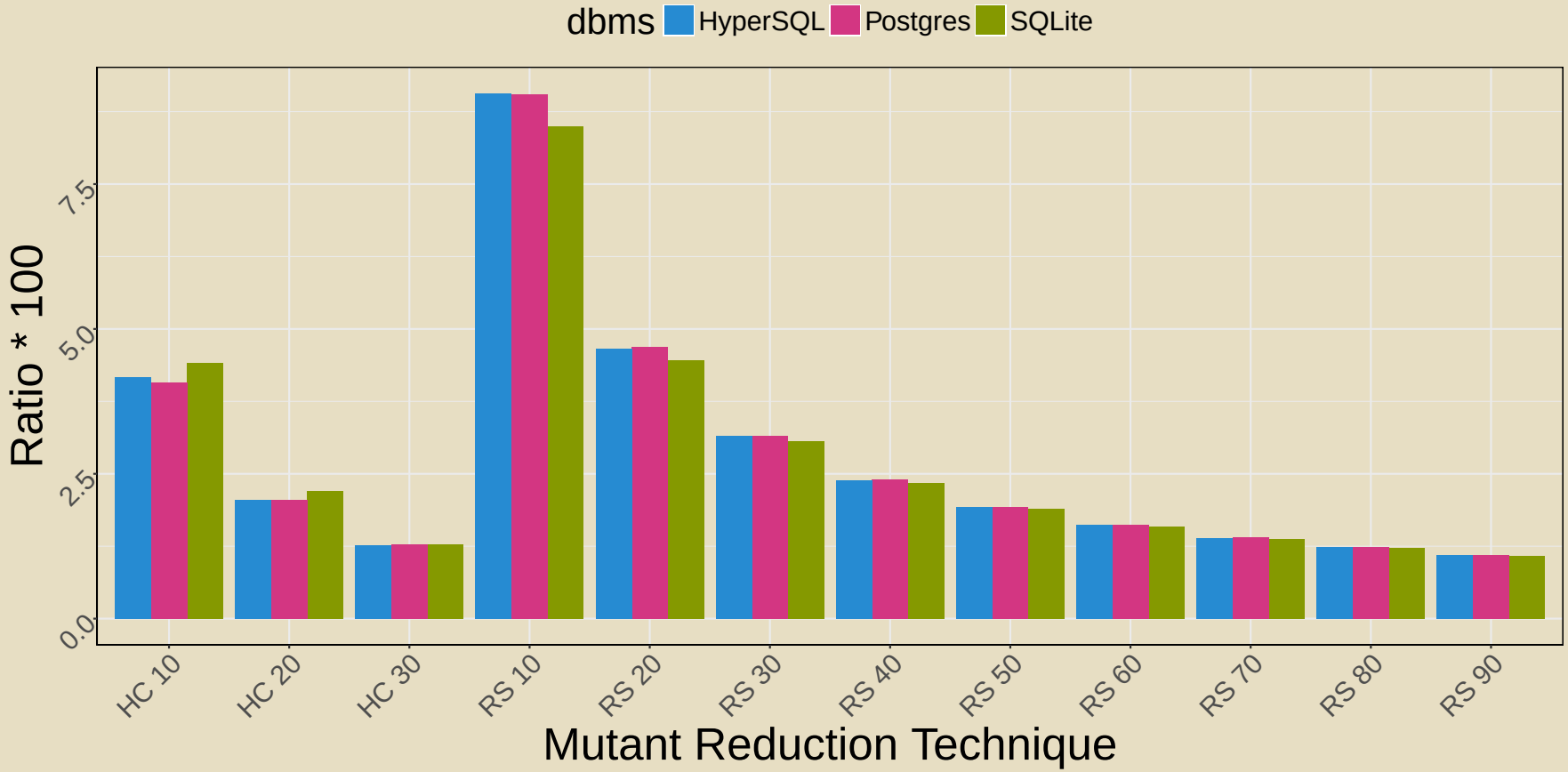


HC model was more focused on reducing cost!

Empirical Results



Empirical Results



RS 10 is still the clear winner!

Empirical Connections

Random Sampling is Easy and Effective

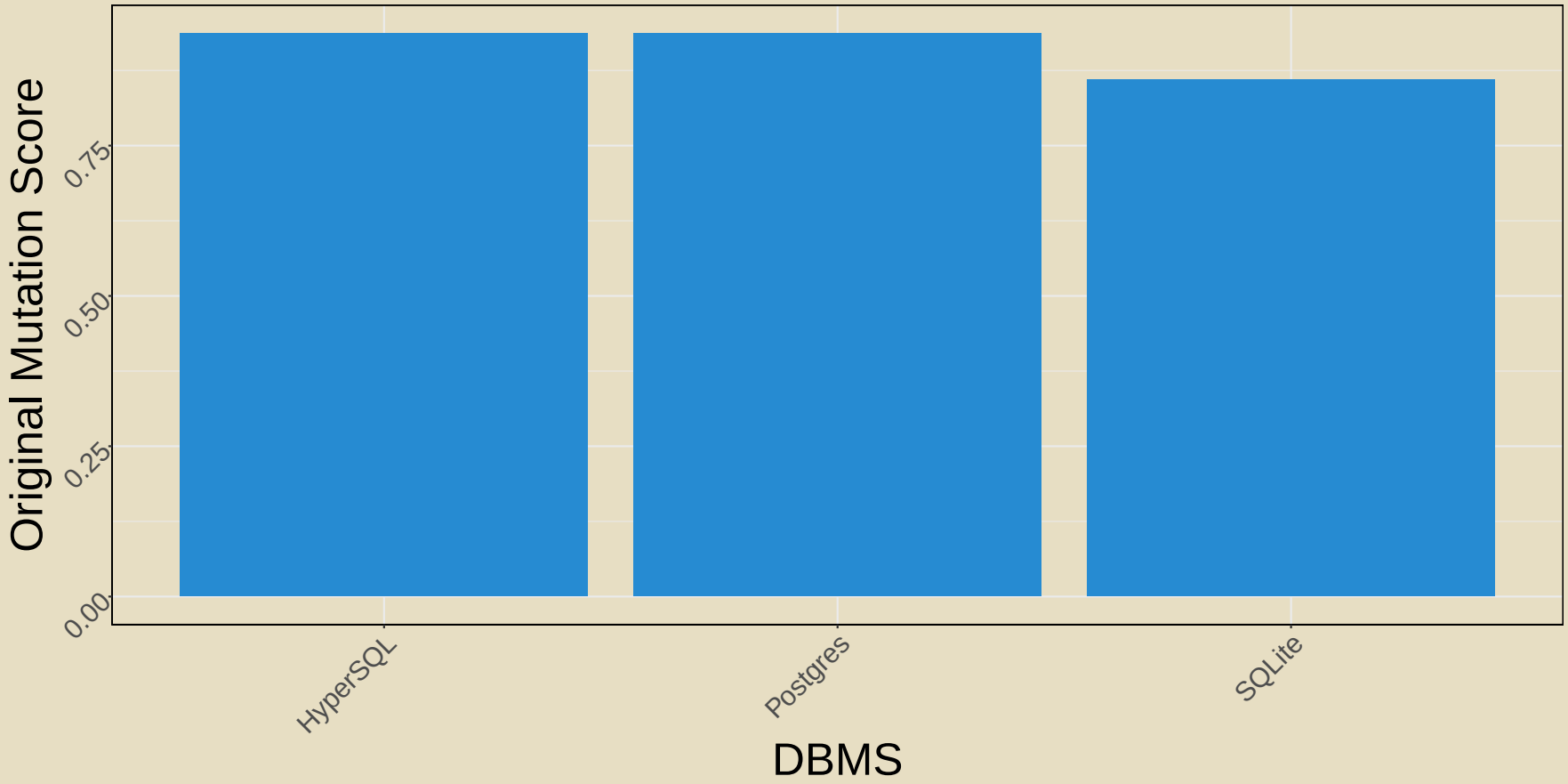
Conclusion

Effectively Reduce Schema Mutants

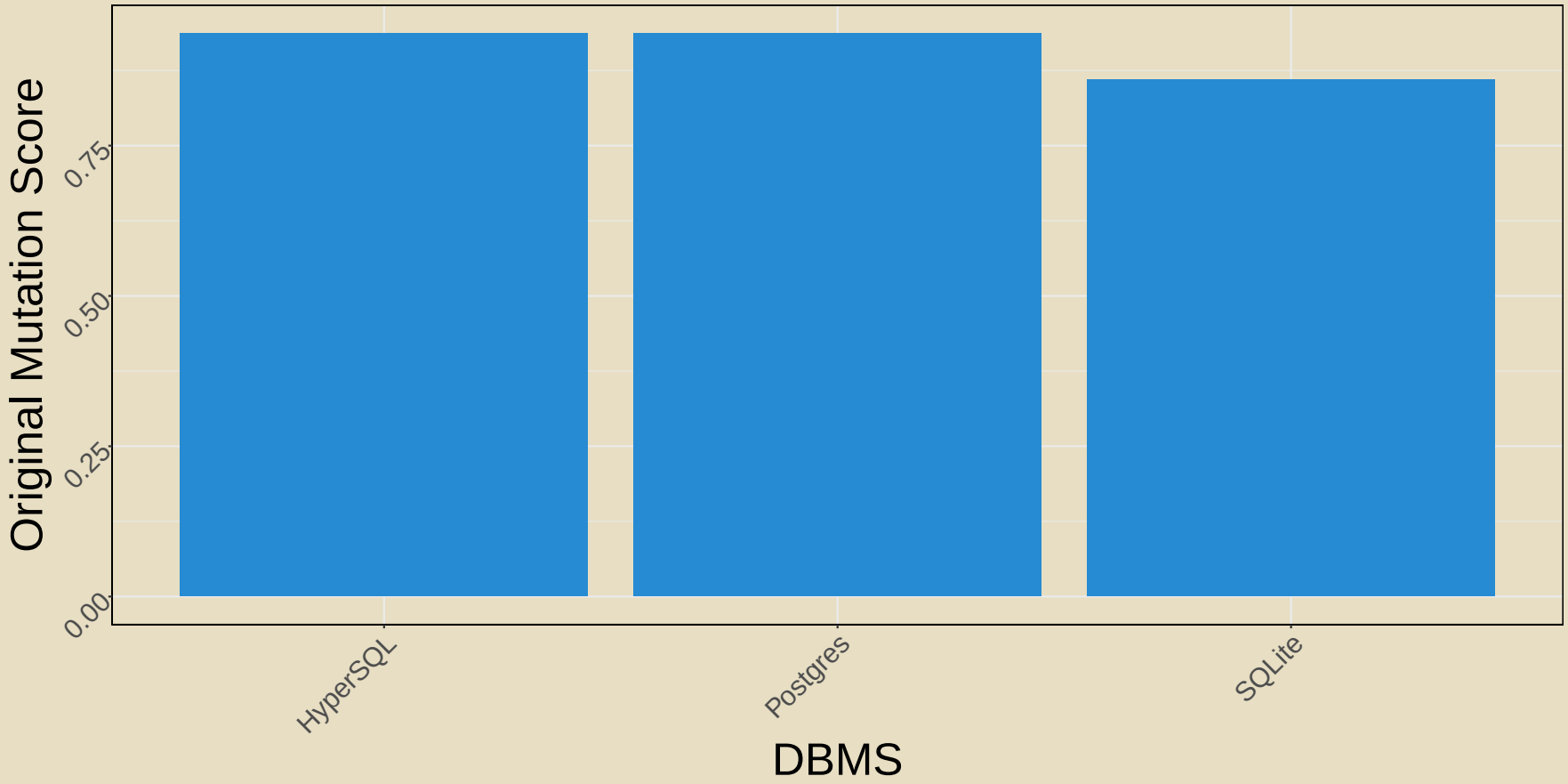
Conclusion

Random Sampling Outperforms Hill Climbing

Discussion


Original Mutation Score
0.00
0.25
0.50
0.75
HyperSQL
Postgres
SQLite
DBMS

Discussion



A lot of mutants are killed!

Contributions

Implemented three reduction techniques into *mrstudyr*

Contributions

Implemented three reduction techniques into *mrstudyr*

Evaluated reduction techniques
on database schema mutants

Contributions

Implemented three reduction techniques into *mrstudyr*

Evaluated reduction techniques
on database schema mutants

Empirically evaluated two mutant reduction techniques

Contributions

Implemented three reduction techniques into *mrstudyr*

Evaluated reduction techniques
on database schema mutants

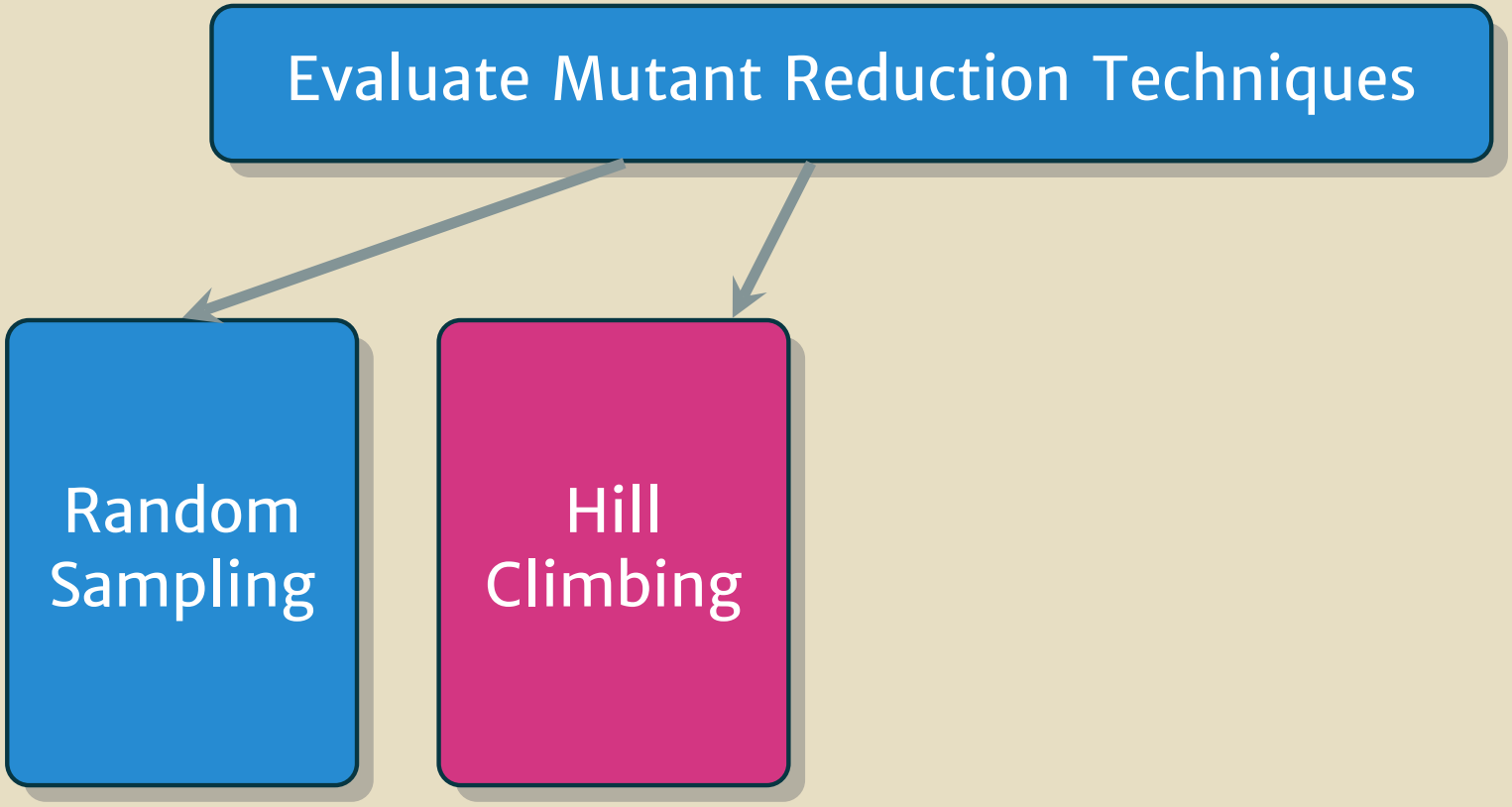
Empirically evaluated two mutant reduction techniques

Compared the effectiveness of
an SBSE to a random technique

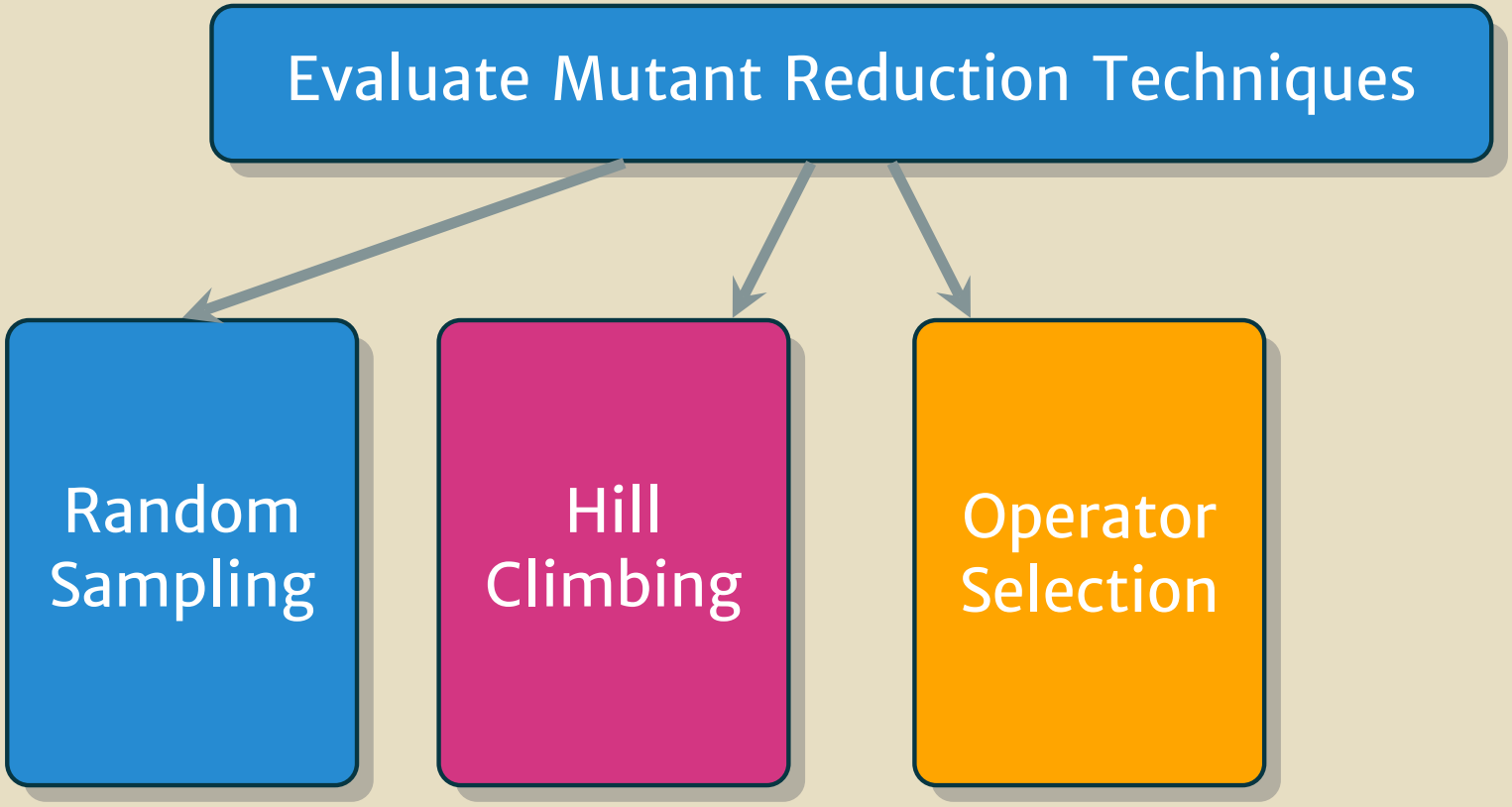
Contributions

- Implemented three reduction techniques into *mrstudyr*
- Evaluated reduction techniques on database schema mutants
- Empirically evaluated two mutant reduction techniques
- Compared the effectiveness of an SBSE to a random technique
- Introduced a metric for evaluating a technique based on ability to reduce cost while producing correlated sets

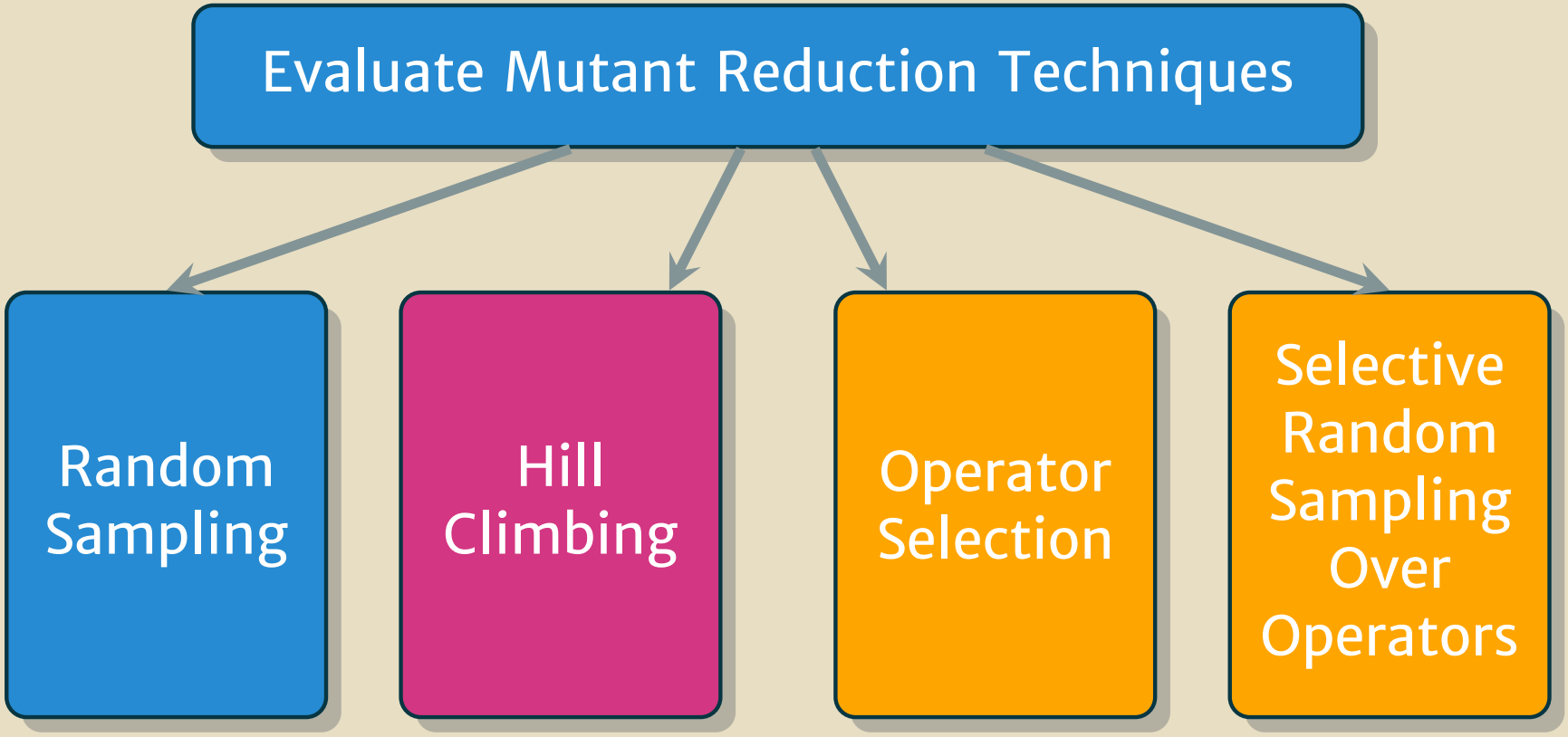
Future Work


Evaluate Mutant Reduction Techniques
Random Sampling
Hill Climbing

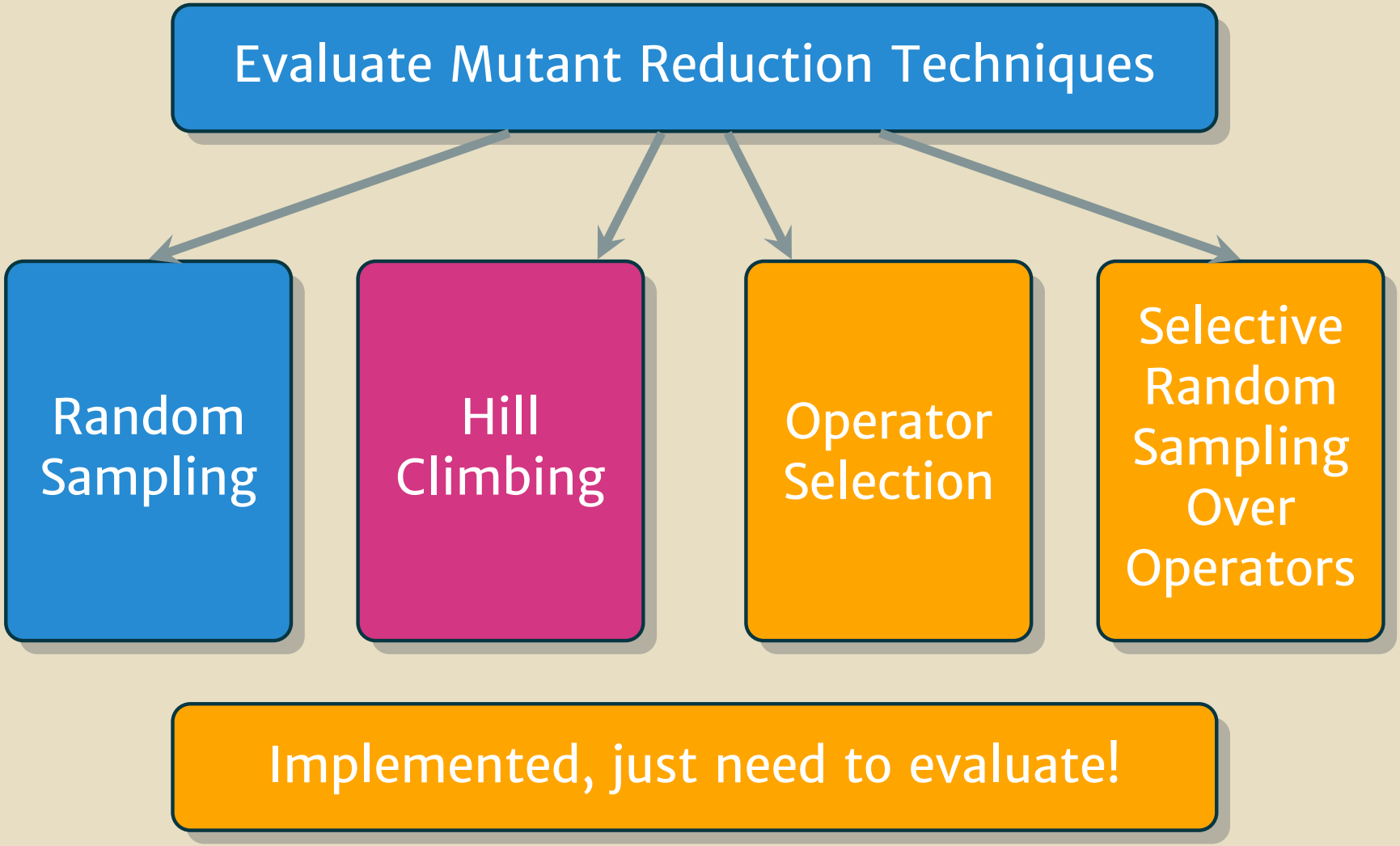
Future Work


Evaluate Mutant Reduction Techniques
Random Sampling
Hill Climbing
Operator Selection

Future Work


Evaluate Mutant Reduction Techniques
Random Sampling
Hill Climbing
Operator Selection
Selective Random Sampling Over Operators

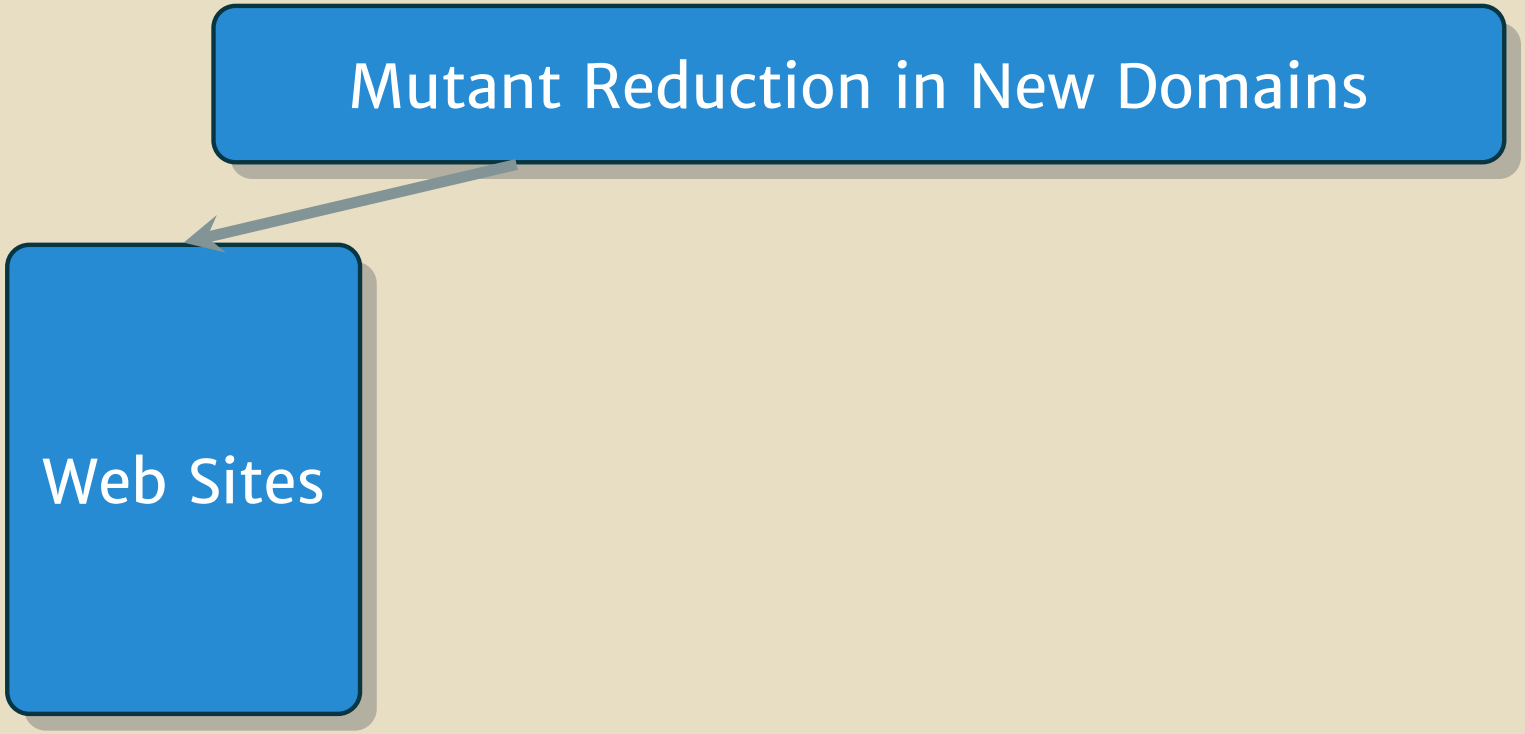
Future Work


Evaluate Mutant Reduction Techniques
Random Sampling
Hill Climbing
Operator Selection
Selective Random Sampling Over Operators
Implemented, just need to evaluate!

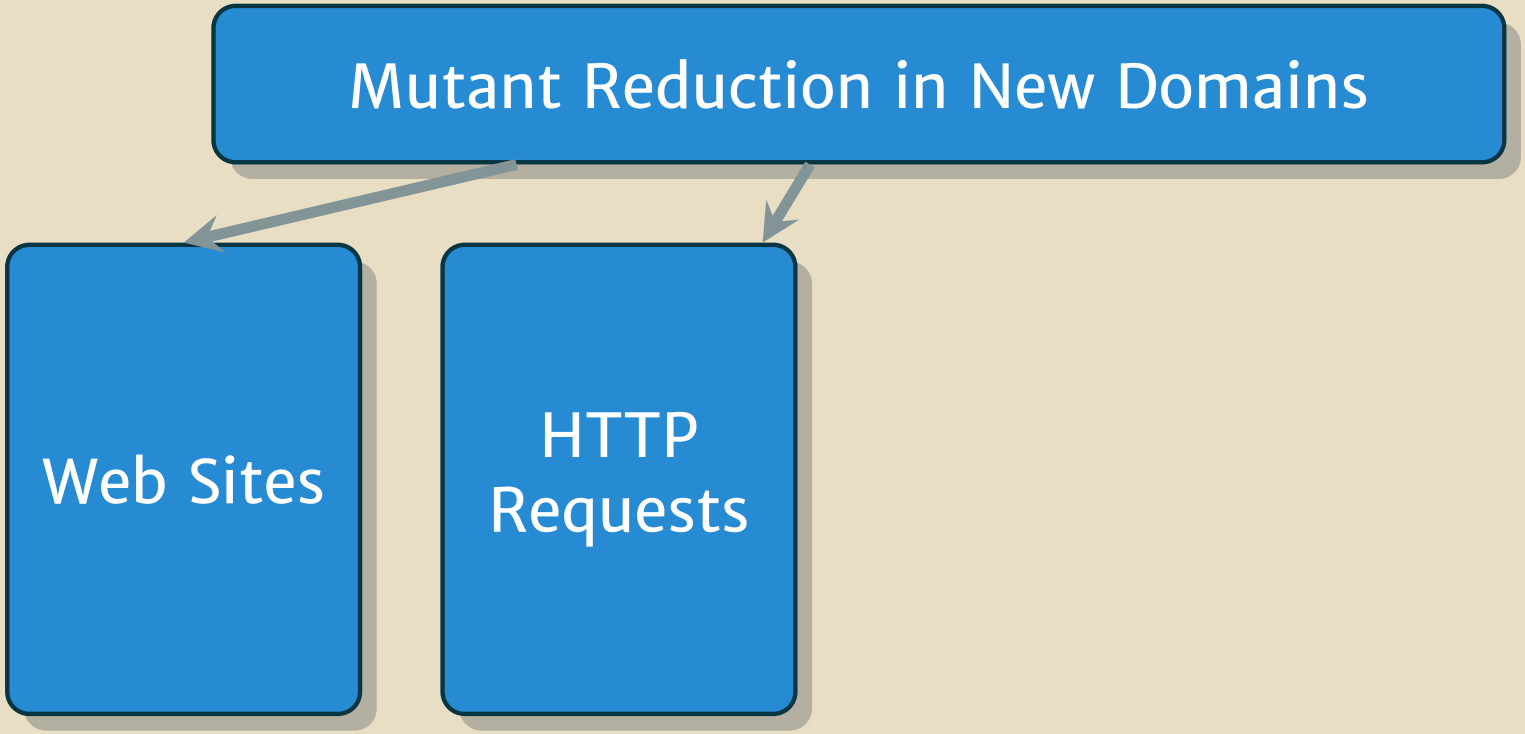
Future Work

Mutant Reduction in New Domains

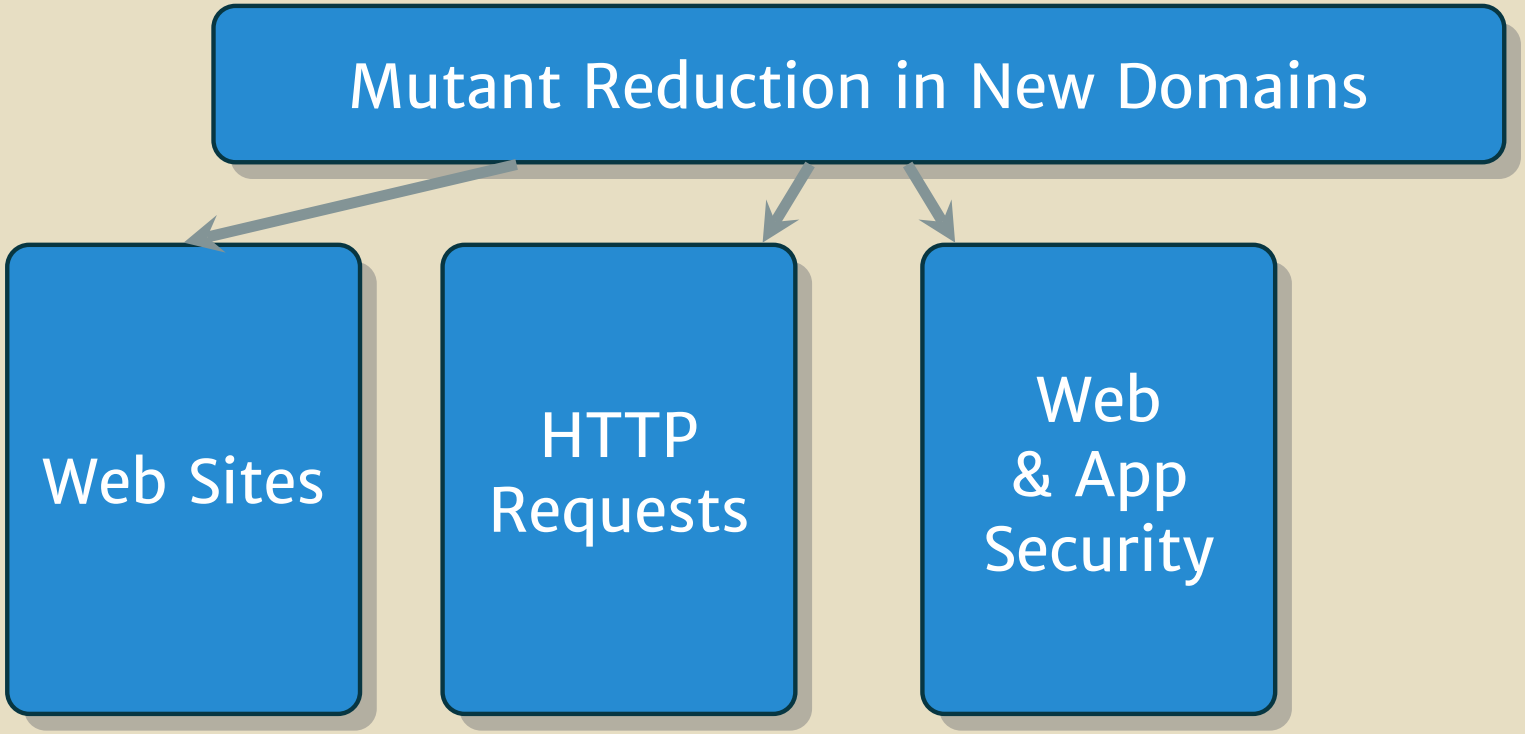
Future Work



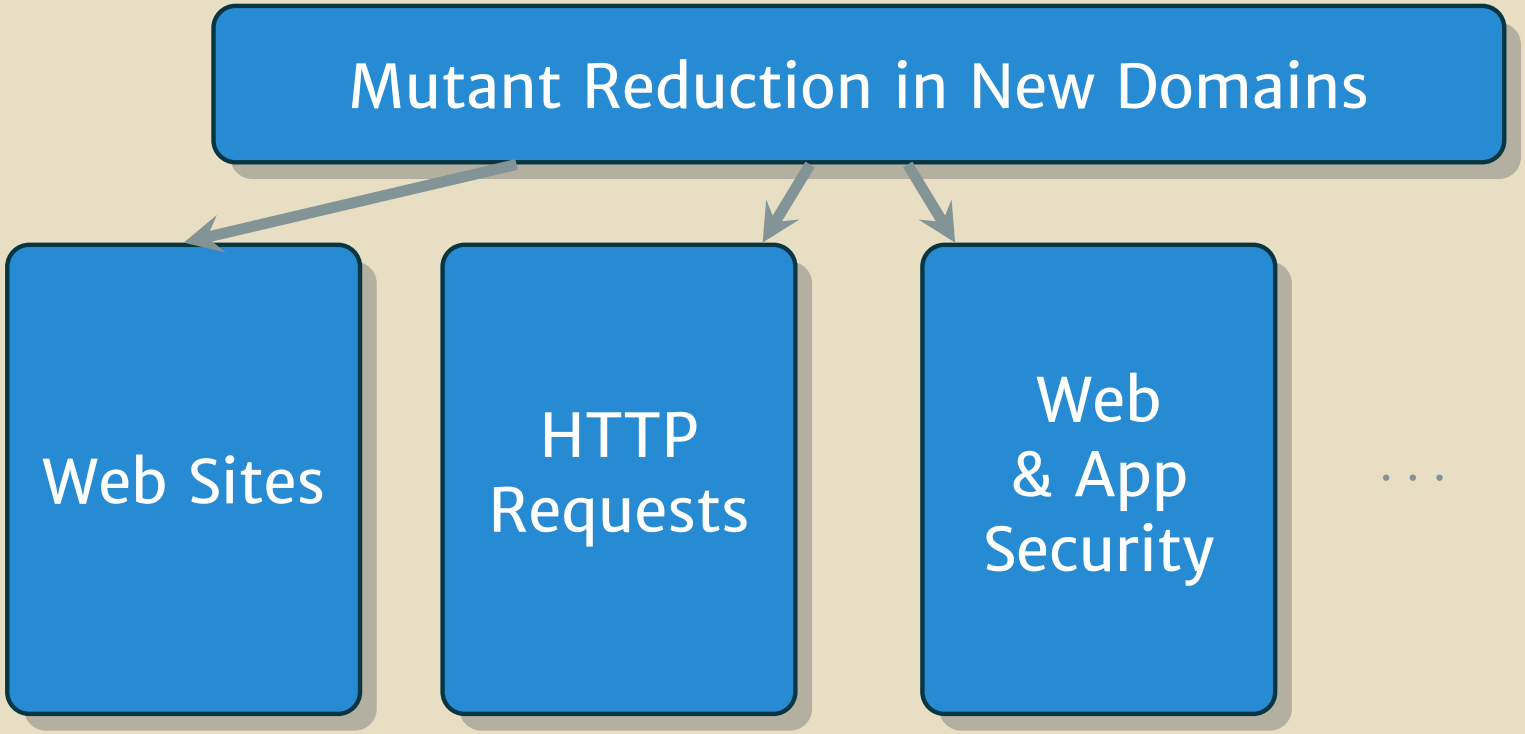
Future Work



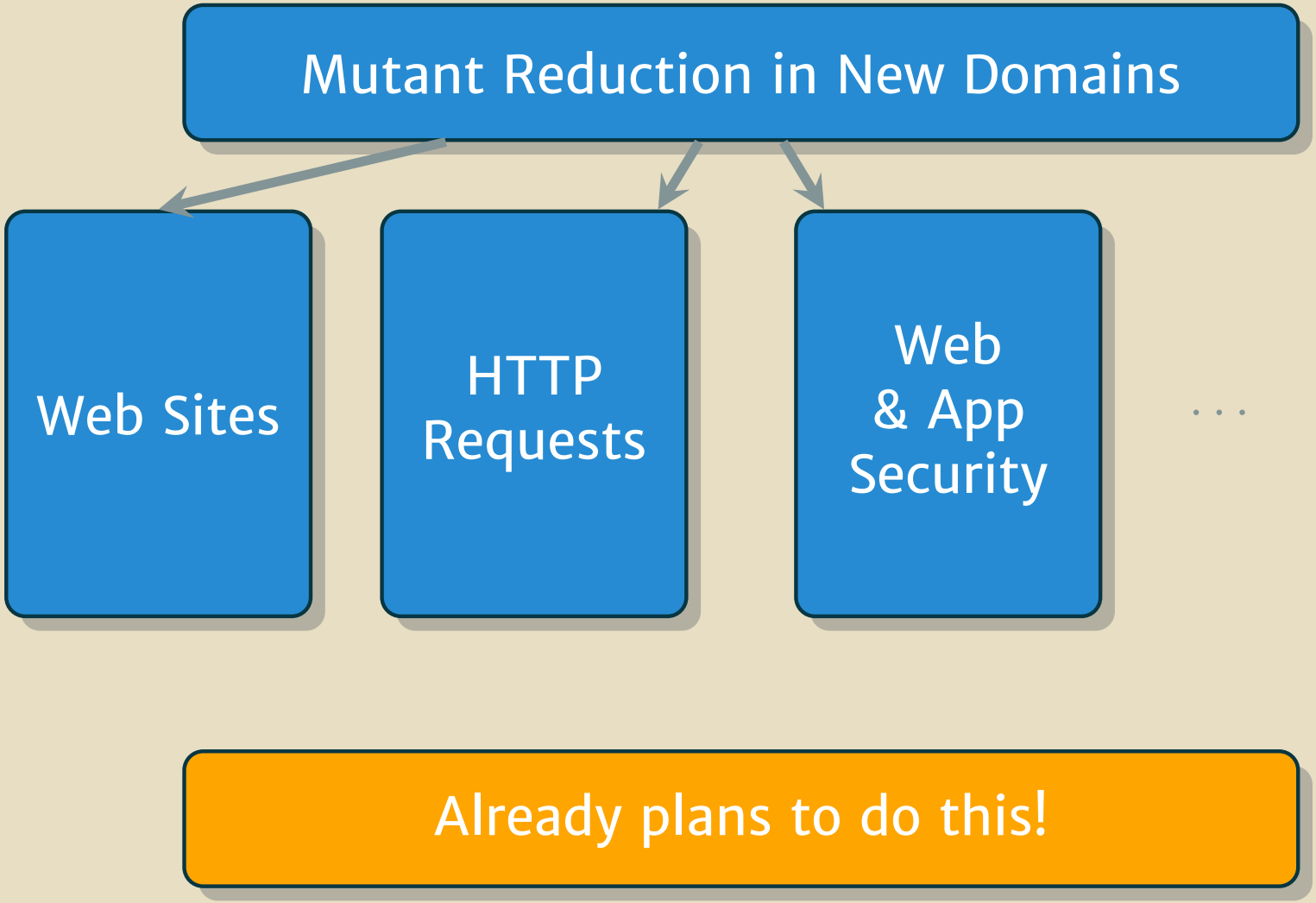
Future Work


Mutant Reduction in New Domains
Web Sites
HTTP Requests
Web & App Security

Future Work


Mutant Reduction in New Domains
Web Sites
HTTP Requests
Web & App Security
. . .

Future Work


Mutant Reduction in New Domains
Web Sites
HTTP Requests
Web & App Security
. . .
Already plans to do this!