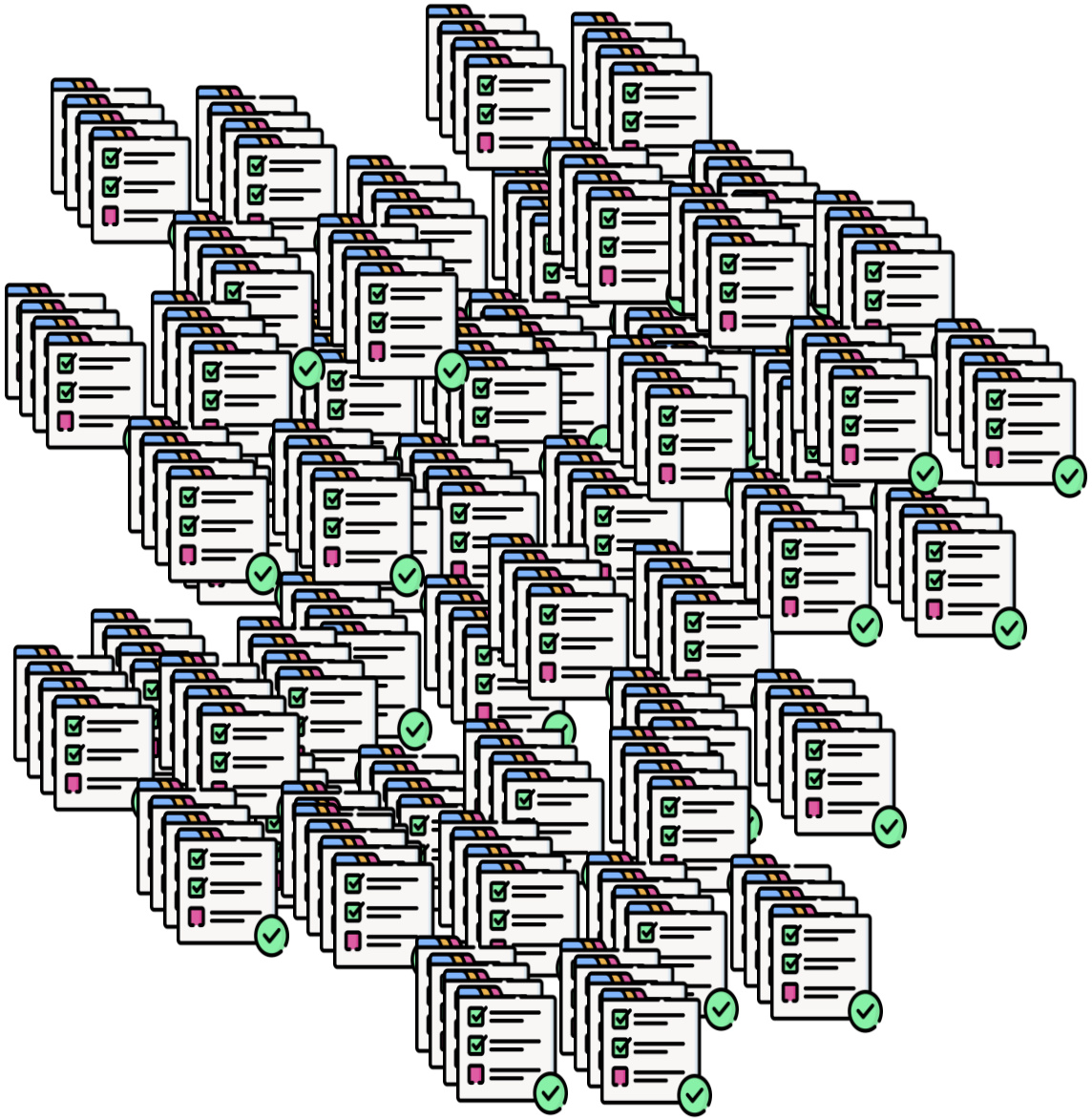


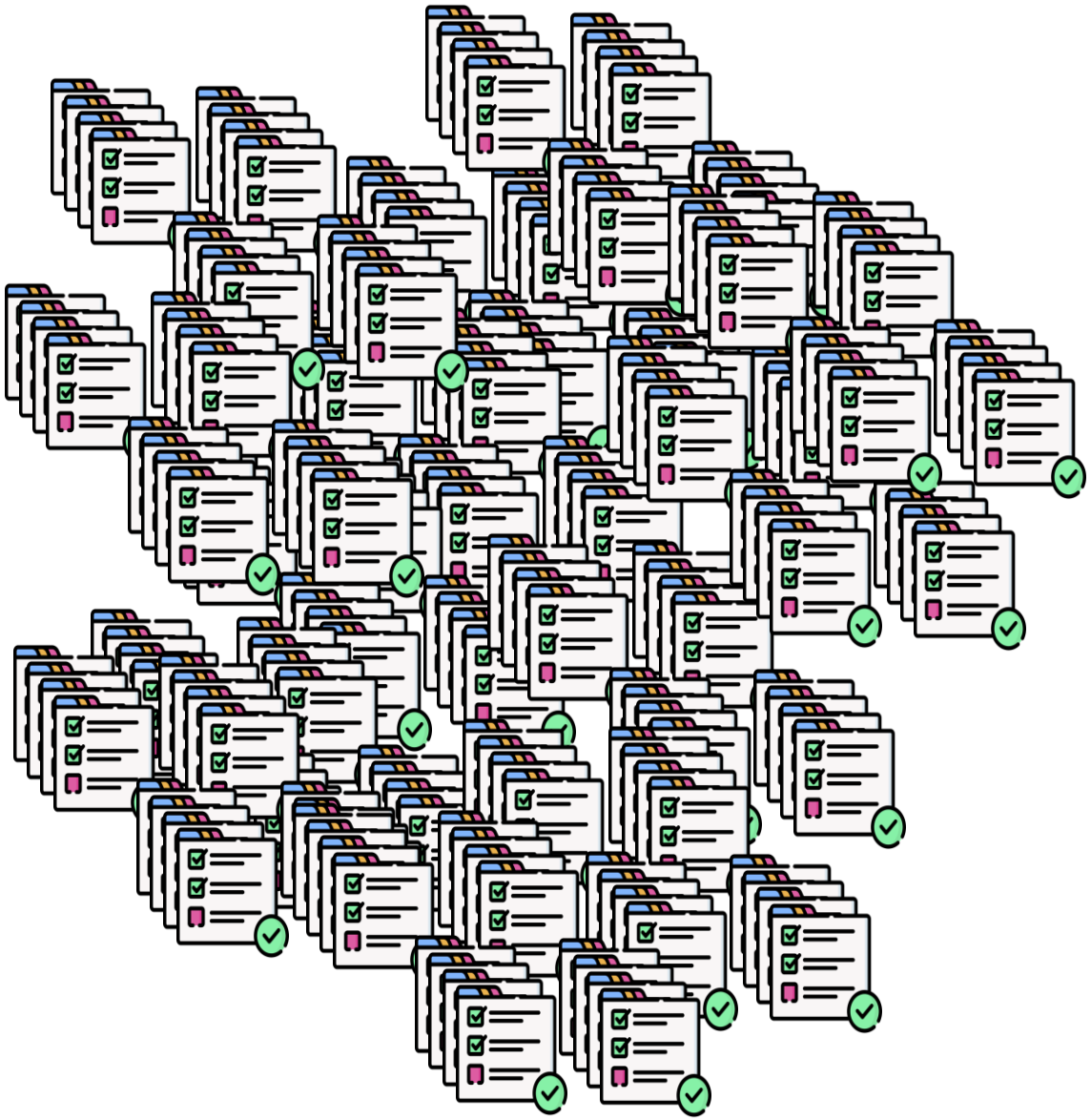
Don't Pack your Trash When You Move

How to Minimize a Test Suite with AI

agiletestingdays.com
Jonas Bogenberger



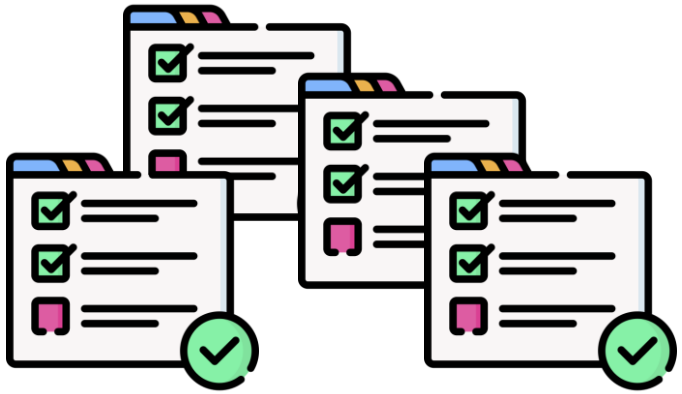




Test ≠ Test



Test-Suite-Minimization



Migration to another Test Framework

Hundreds or thousands



Old Testframework

Minimization

Only 20%



New Testframework

Automating manual tests

Hundreds or thousands



Manual tests

Minimization

Only 20%



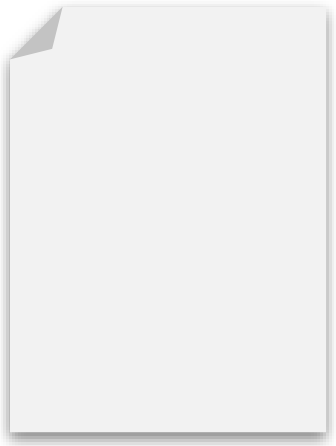
Automated tests

Who can help us move?

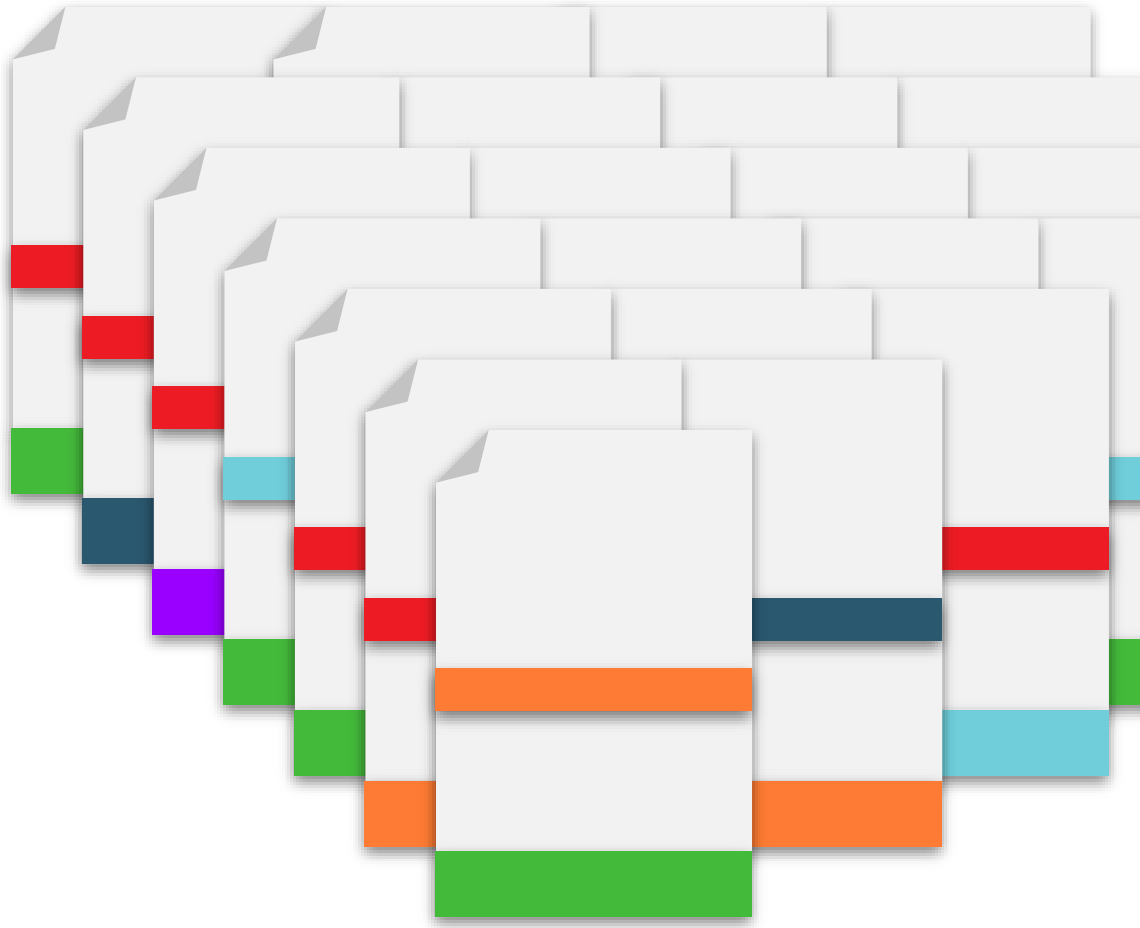


Redundancy


copy & paste



copy & paste



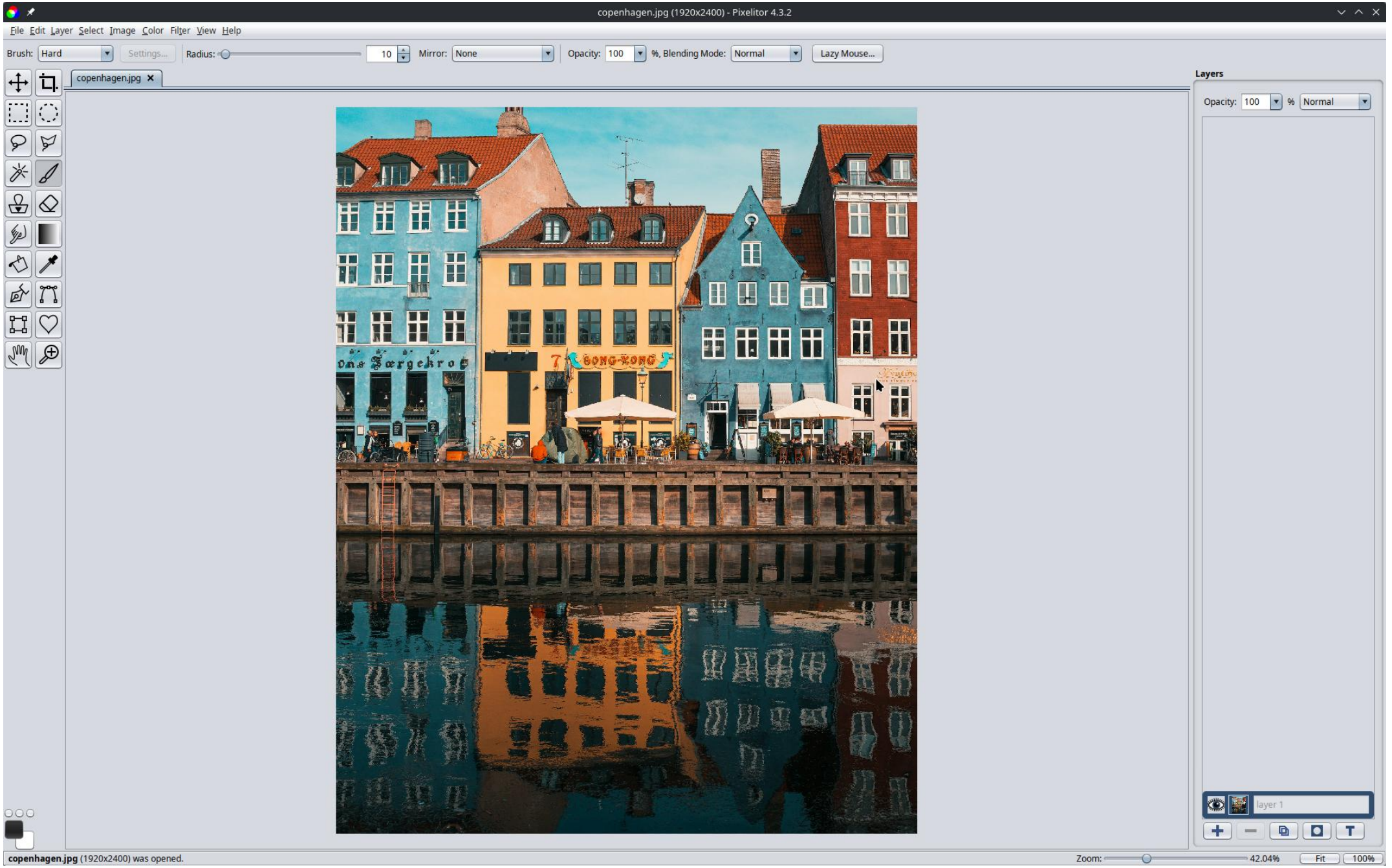
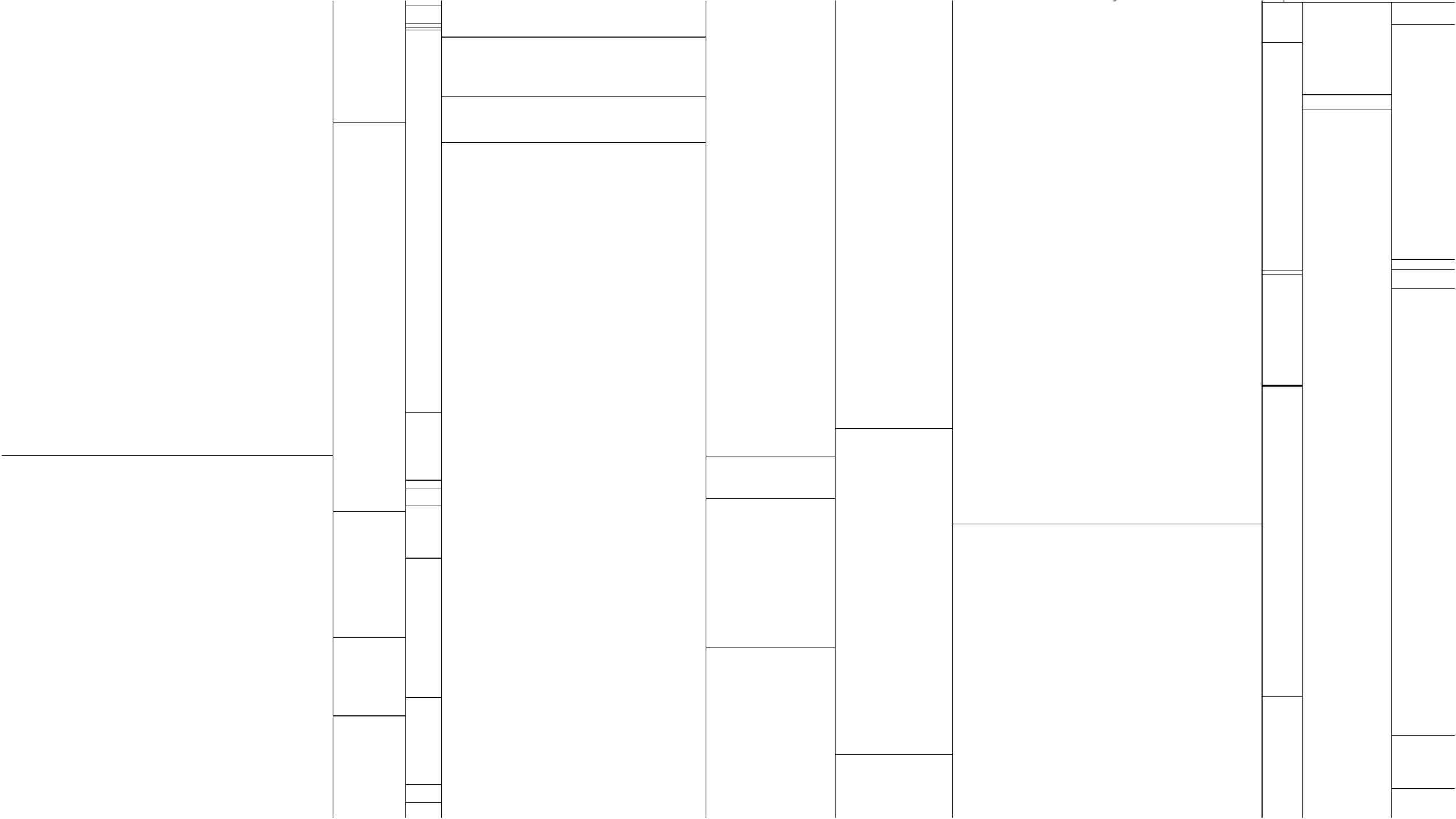
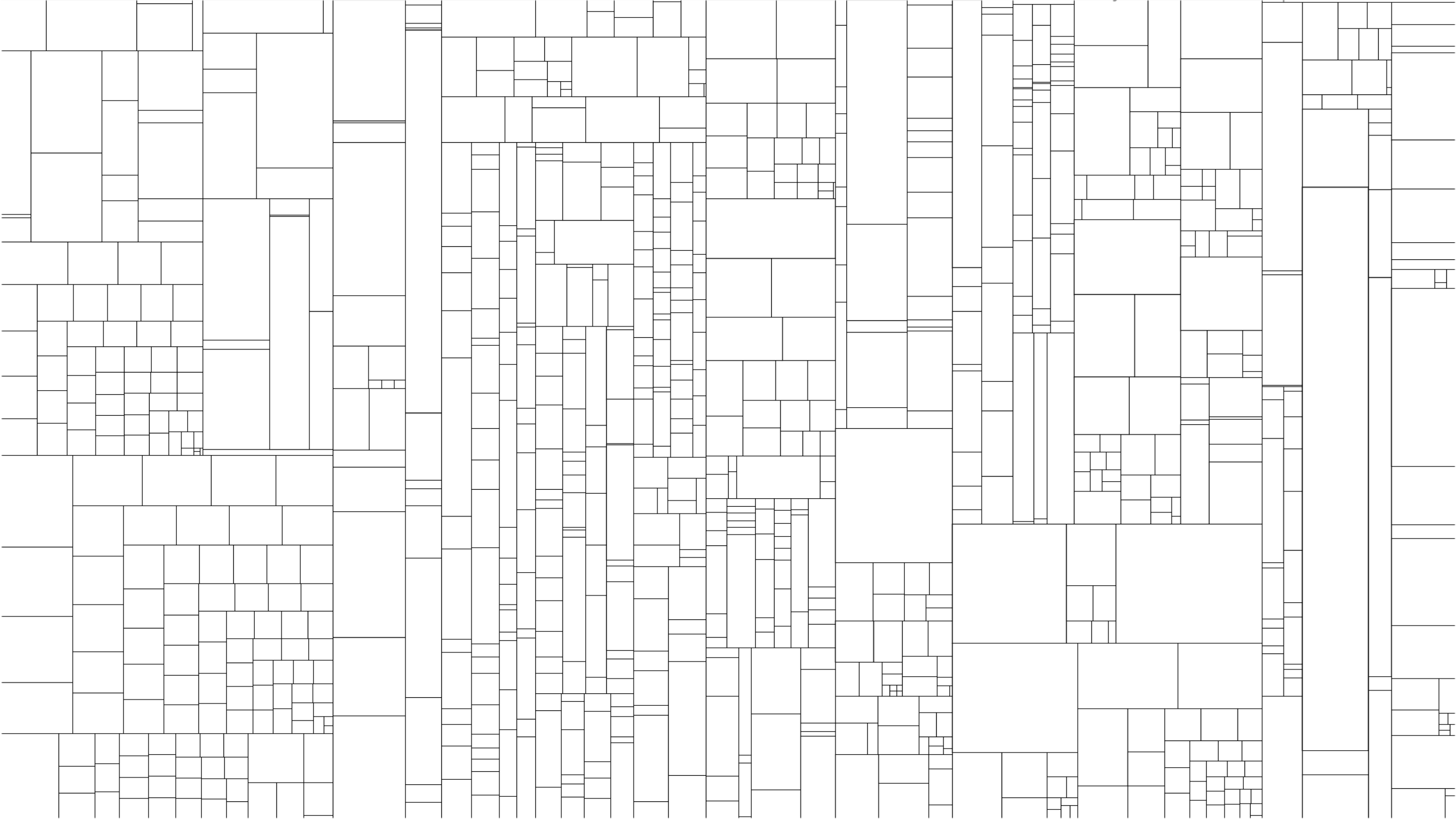
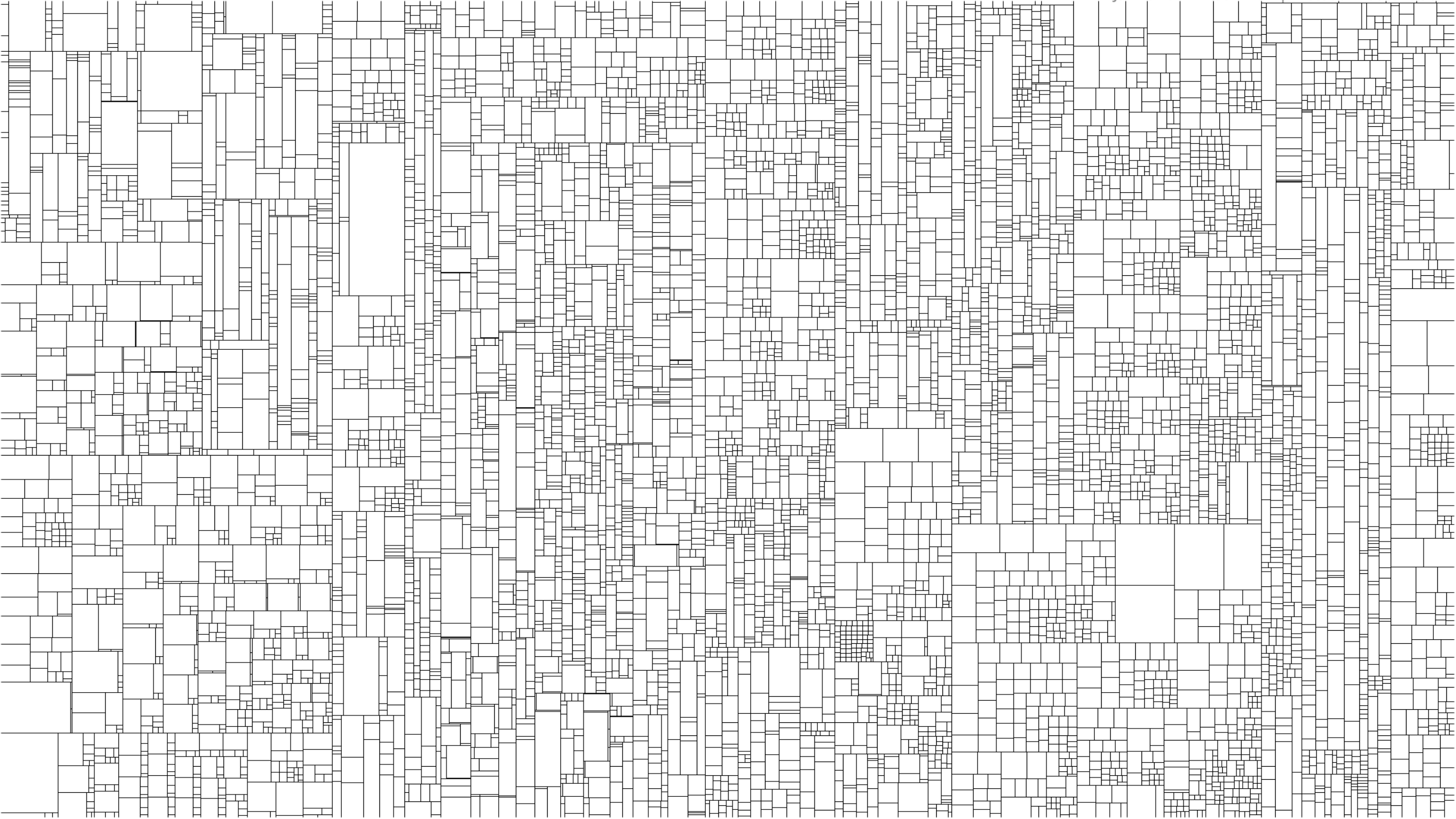
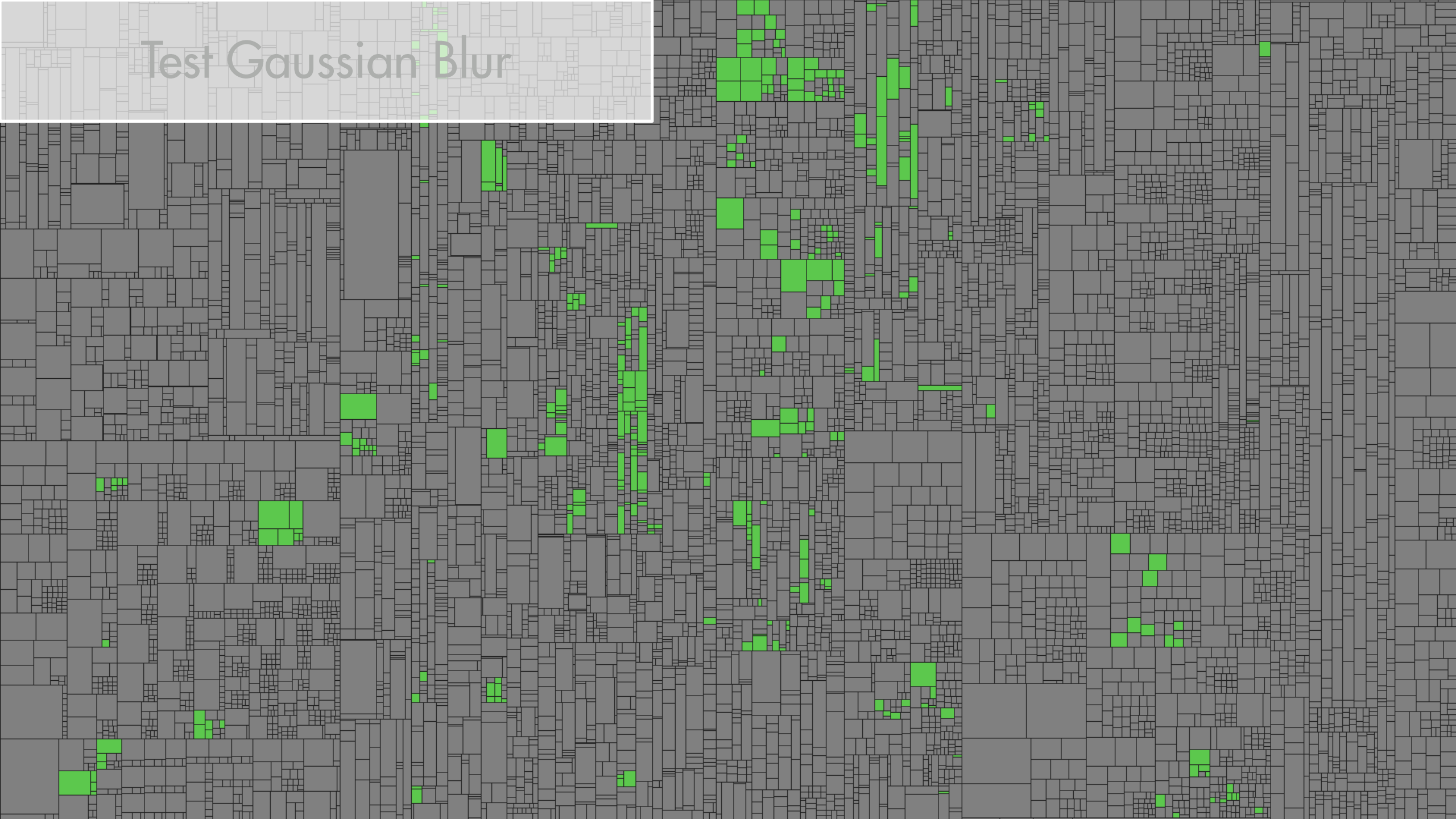


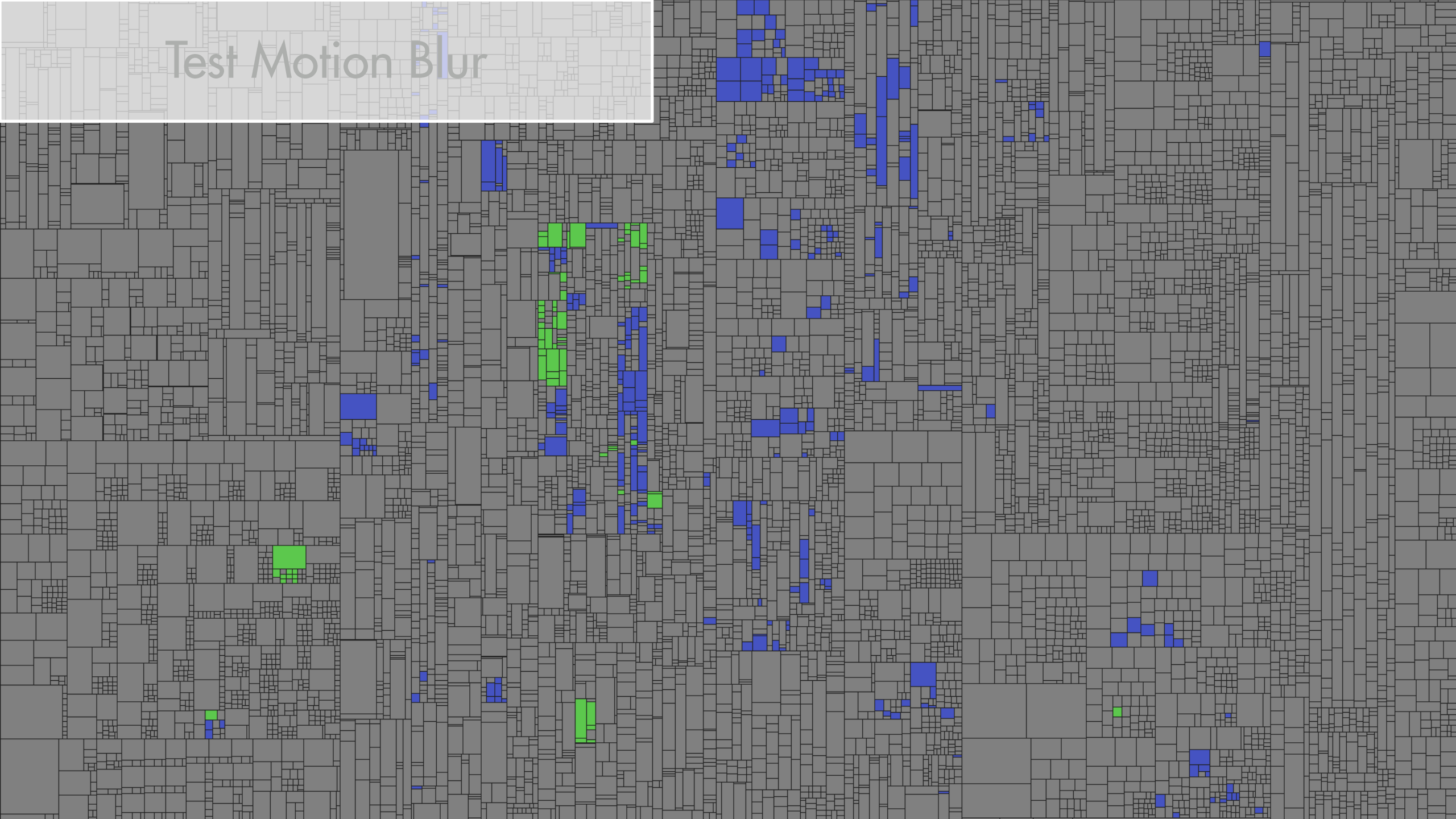
Photo in Pixelitor by Elton Sa on Unsplash

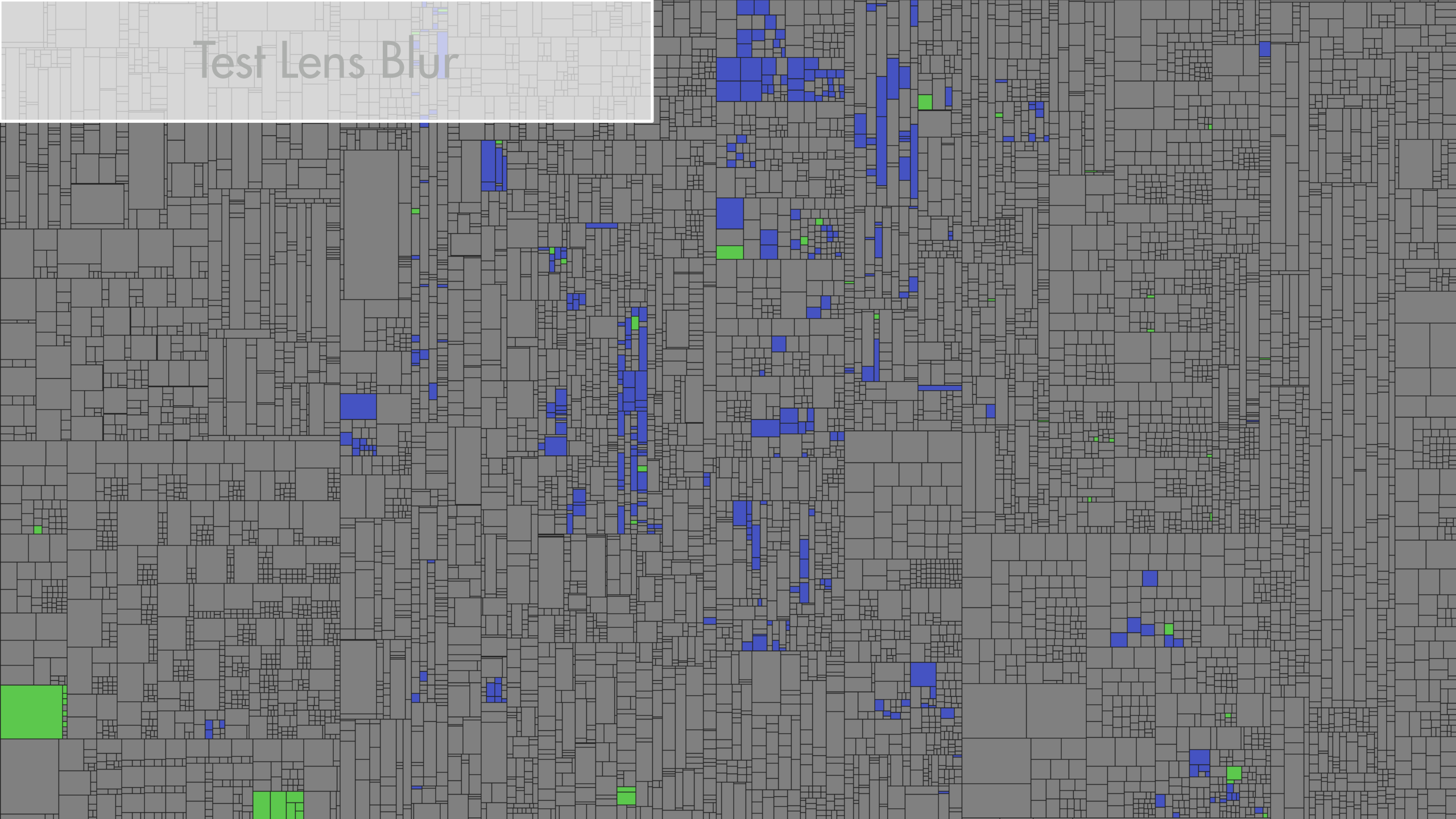


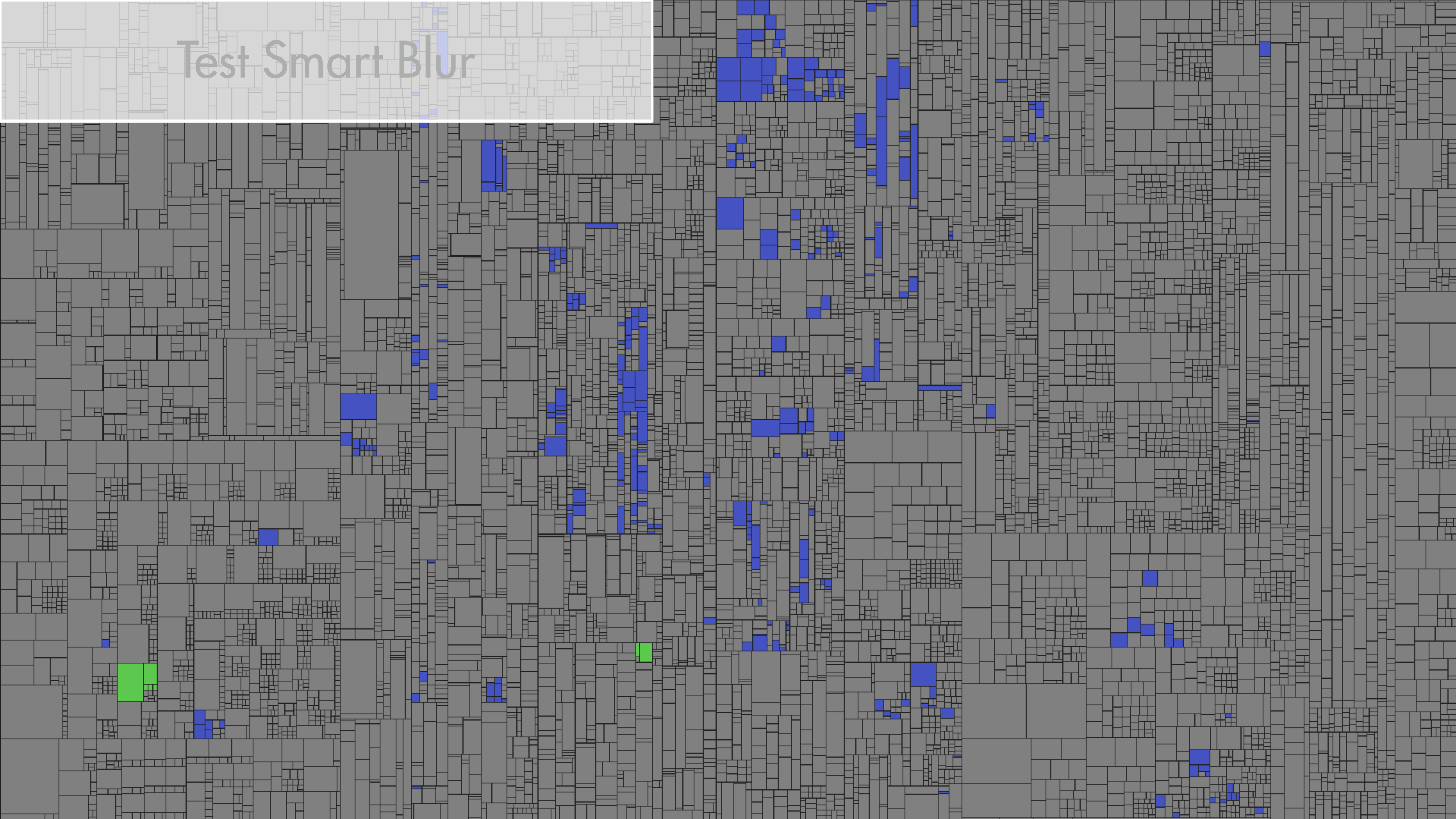


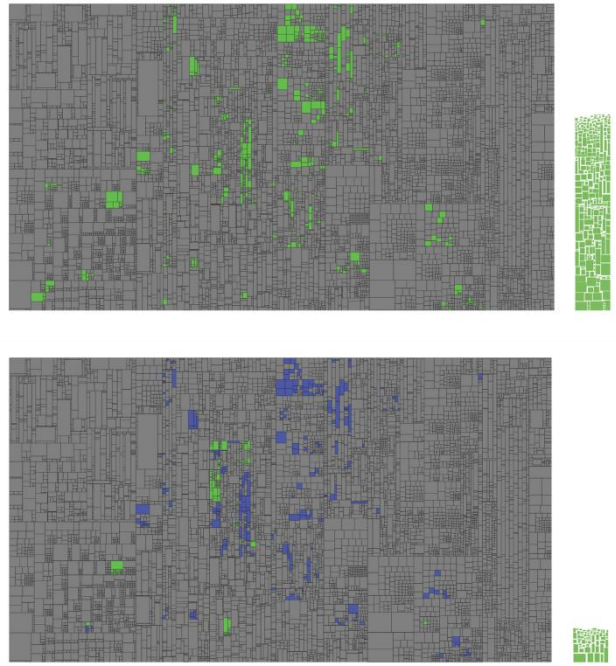



Test Gaussian Blur

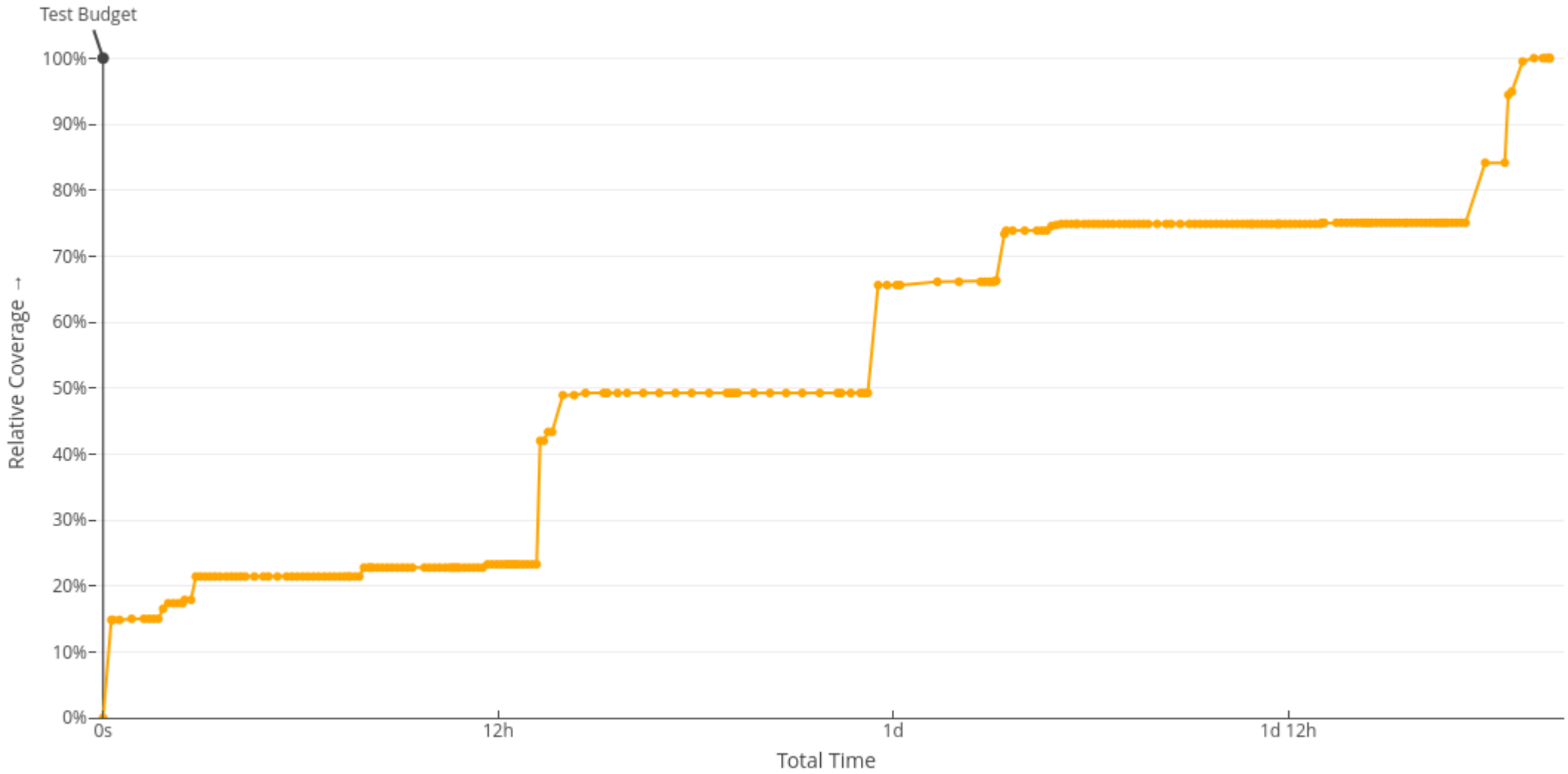

Test Motion Blur


Test Lens Blur


Test Smart Blur



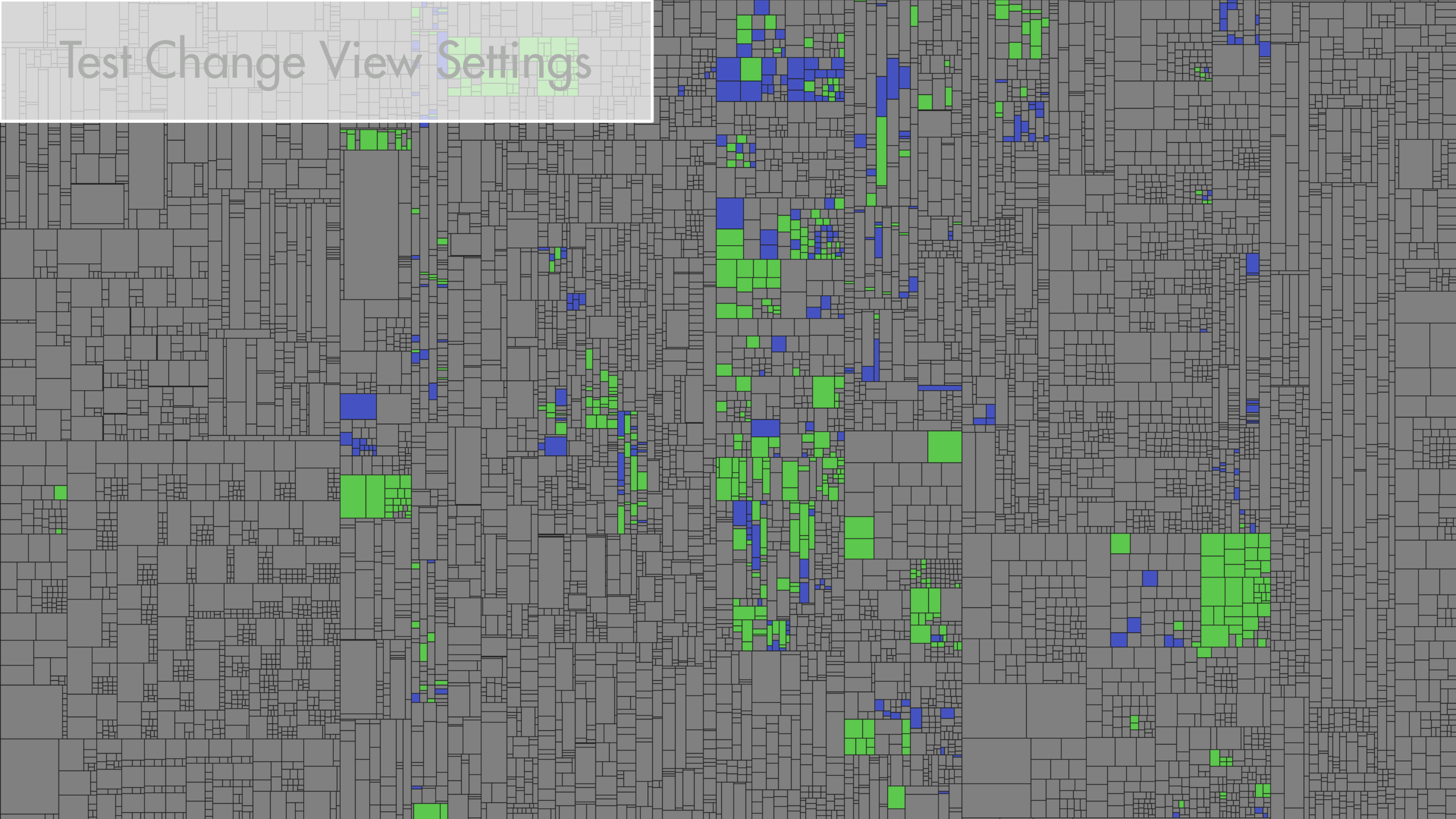
Coverage over Time

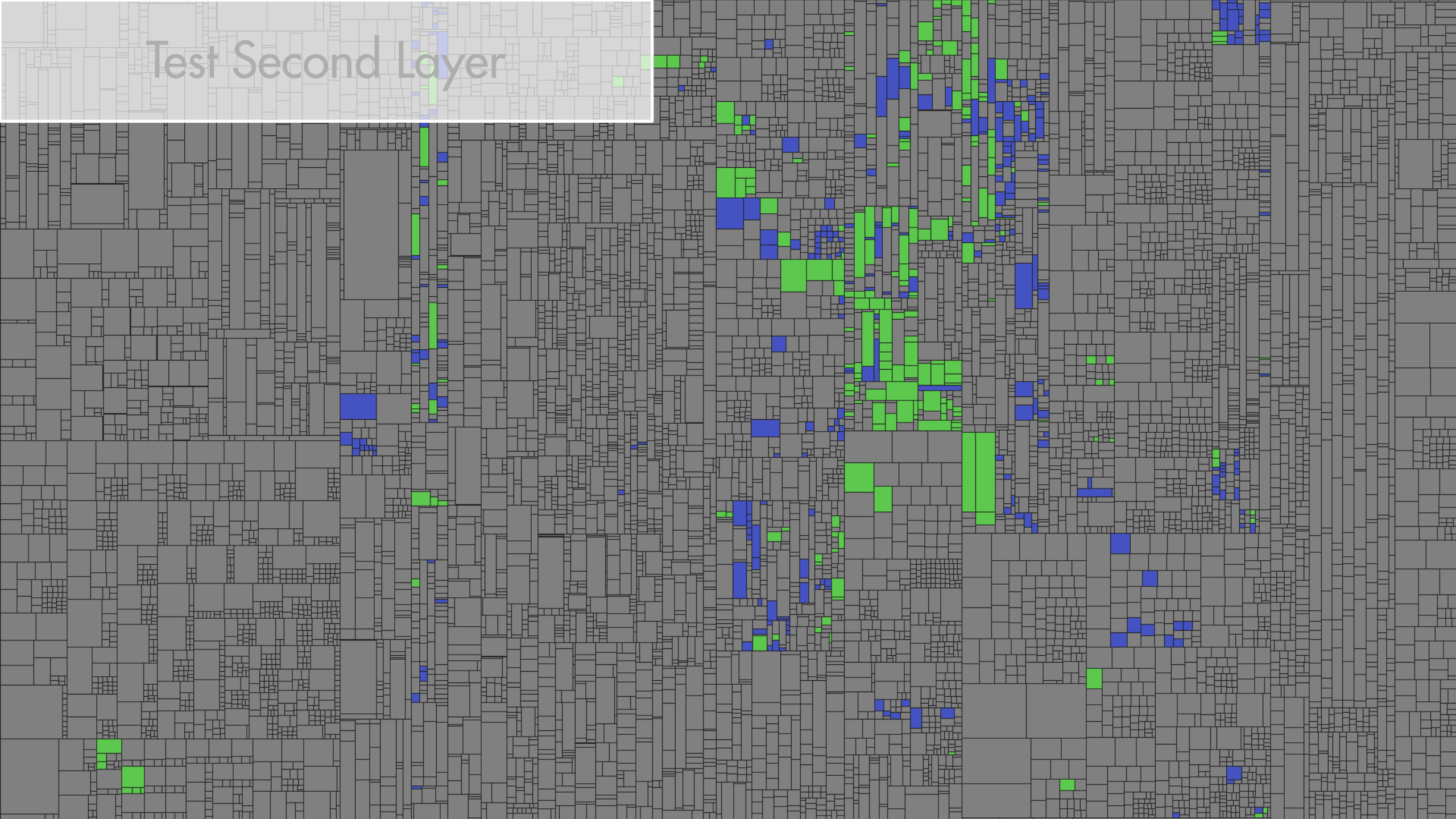


Results for Test Query & Budget Restriction

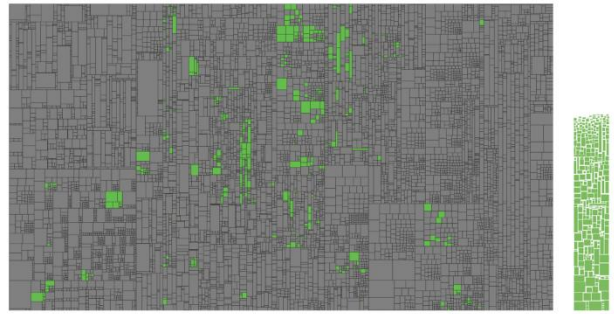
Relative Coverage: 0%, Selected Tests: 0 out of 236 (0%)

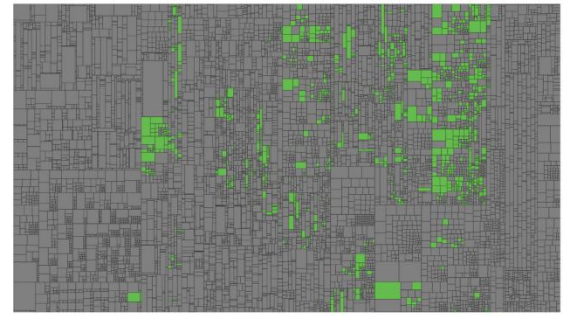

Test Create and Modify Selection

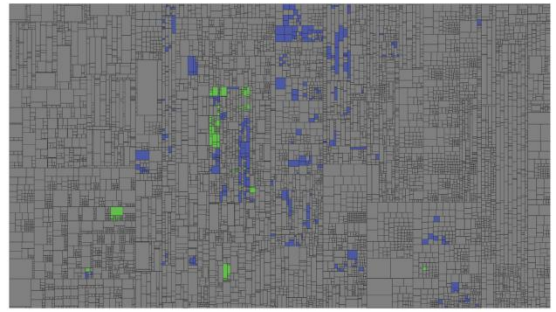

Test Change View Settings

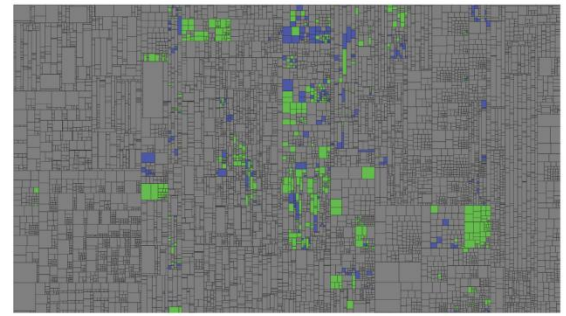

Test Second Layer

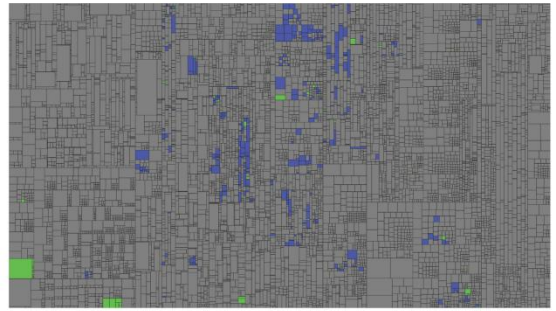

Test Save Image

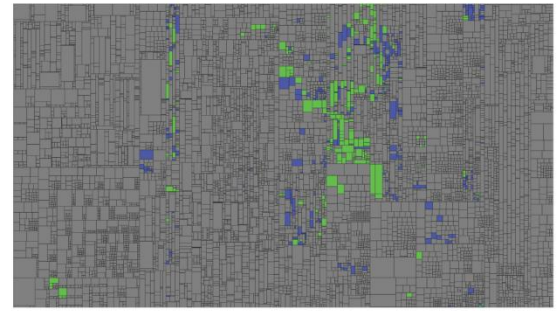


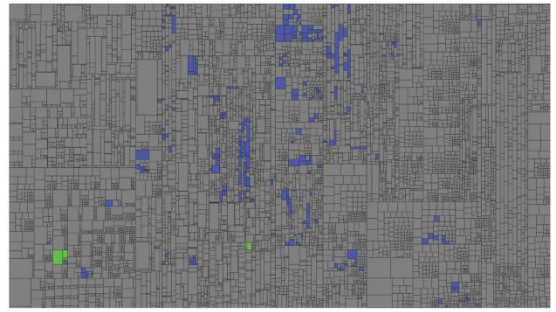


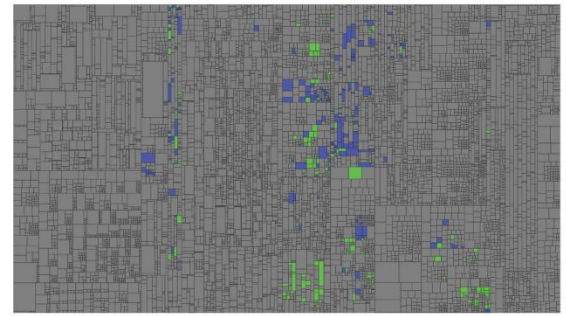





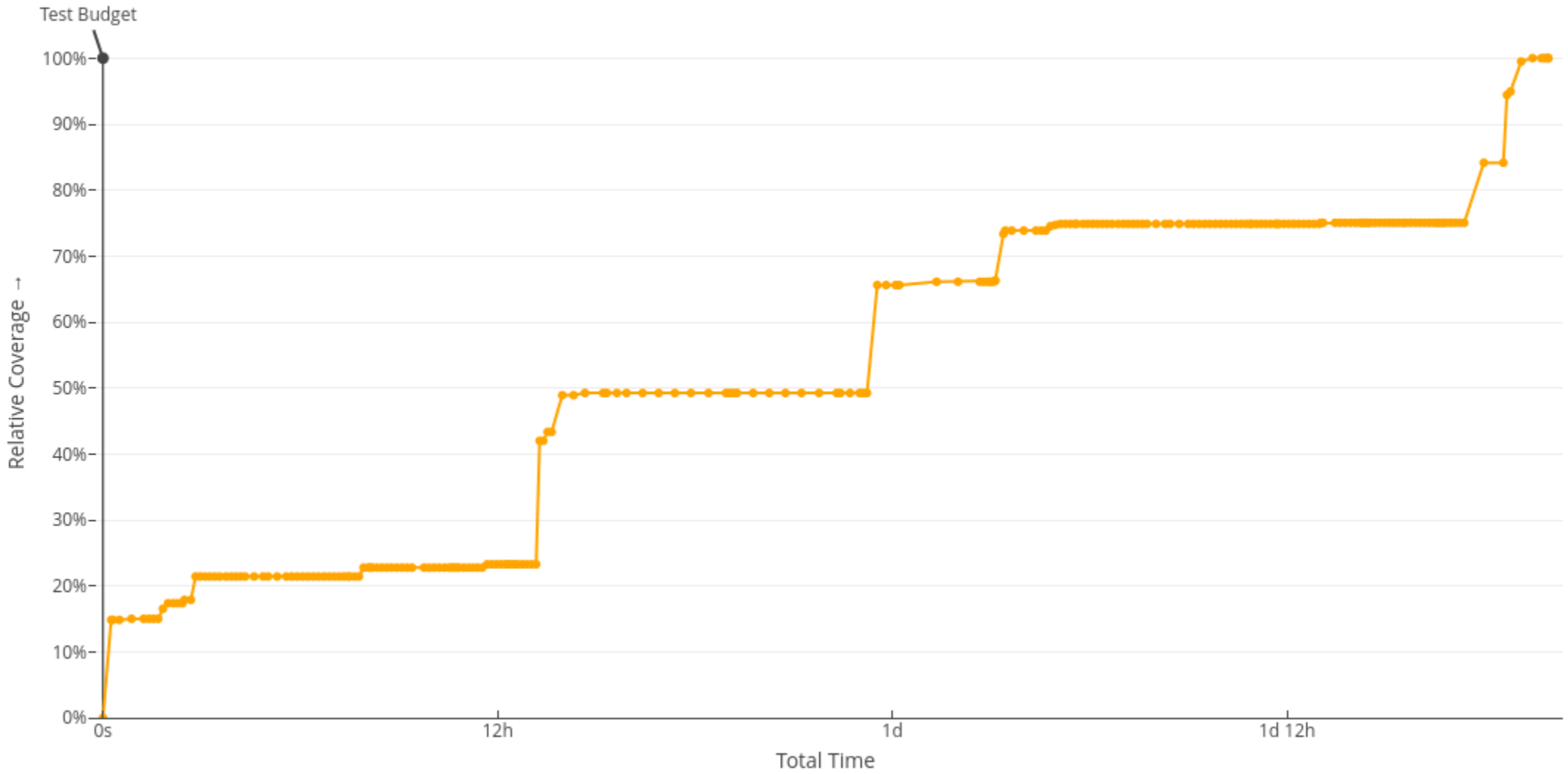








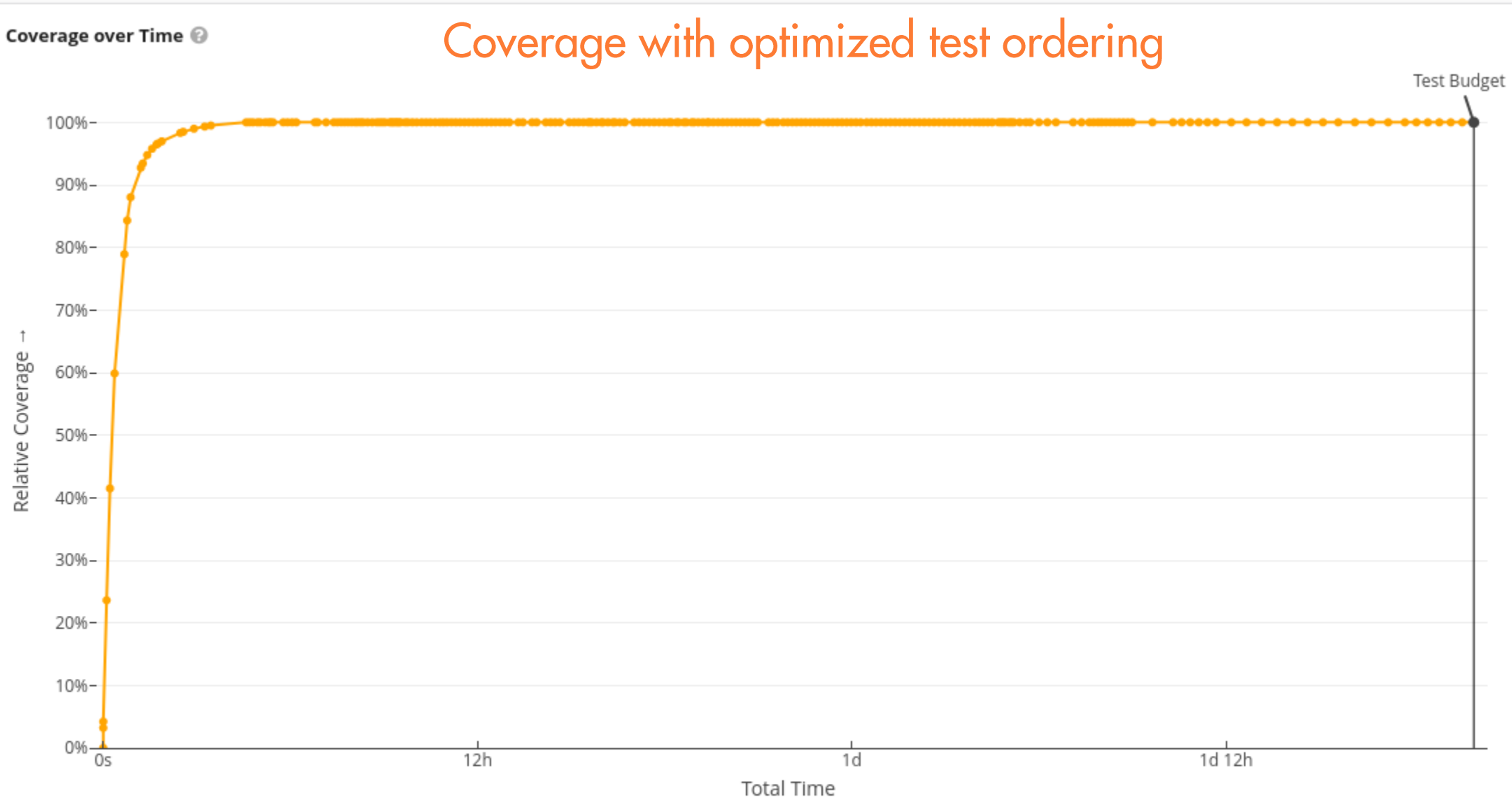

Coverage over Time



Results for Test Query & Budget Restriction

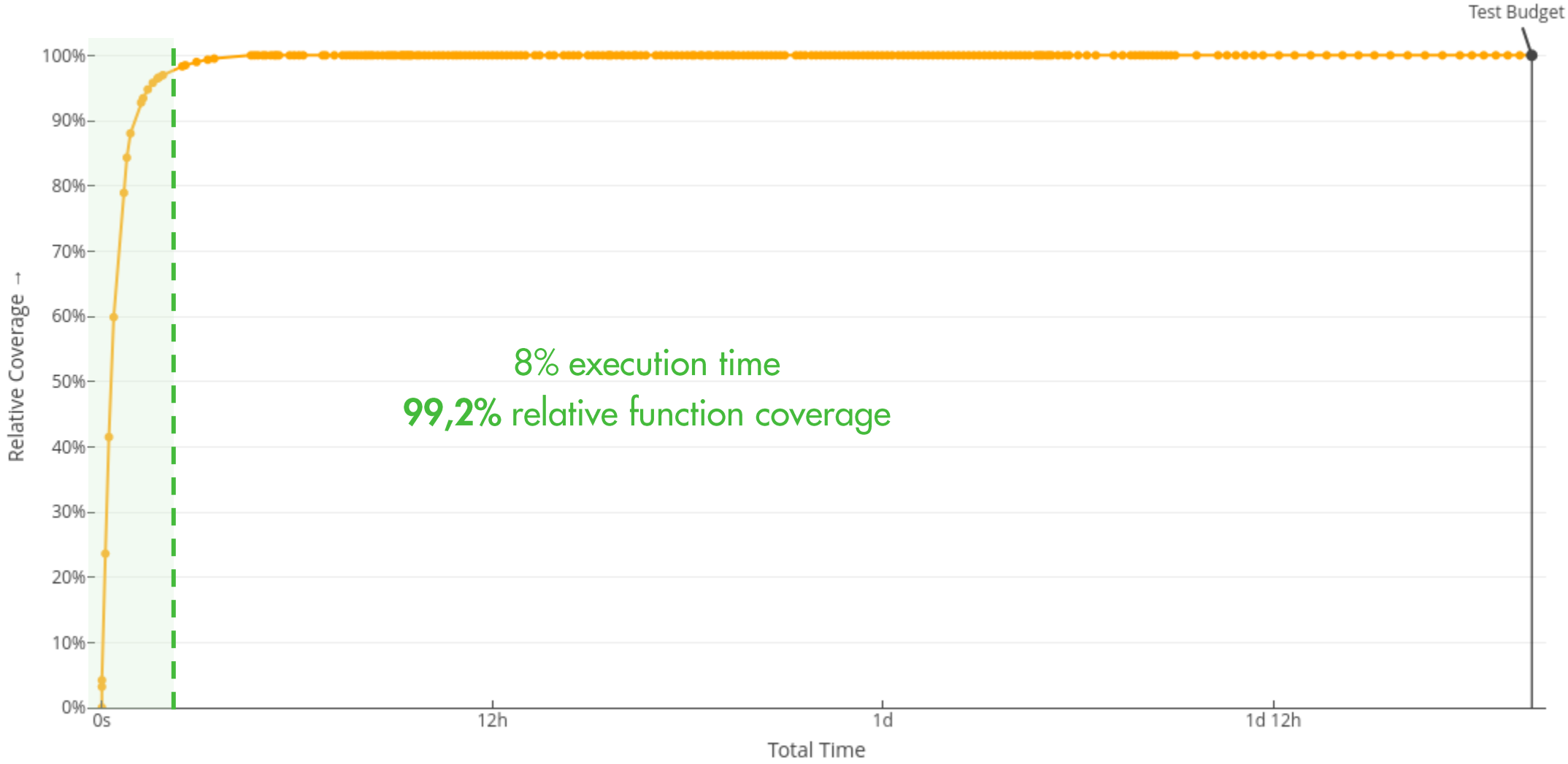
Relative Coverage: 0%, Selected Tests: 0 out of 236 (0%)

Coverage with optimized test ordering



Coverage over Time

Coverage with optimized test ordering

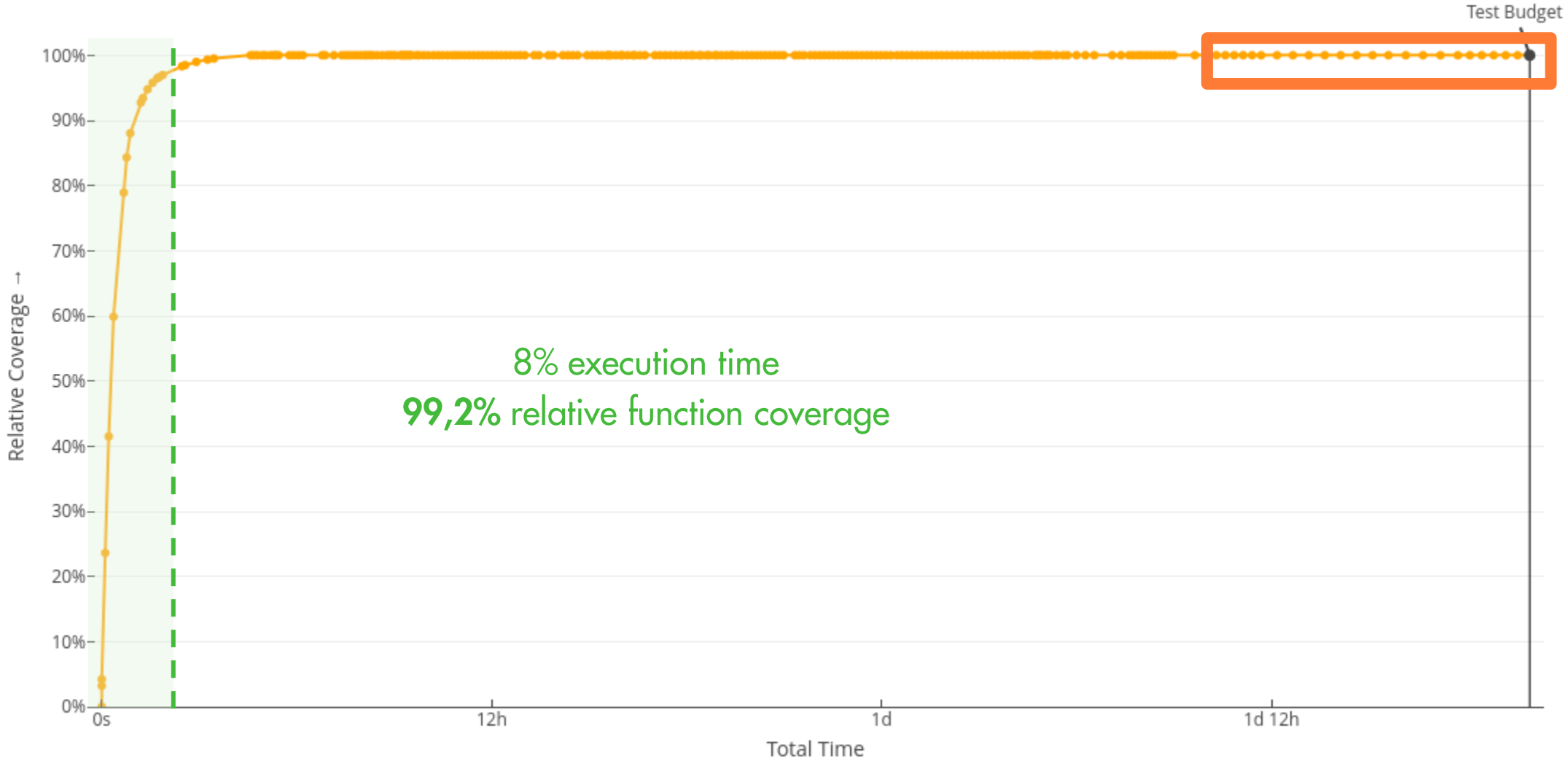


Results for Test Query & Budget Restriction

Relative Coverage: 100%, Selected Tests: 236 out of 236 (100%)

Coverage over Time

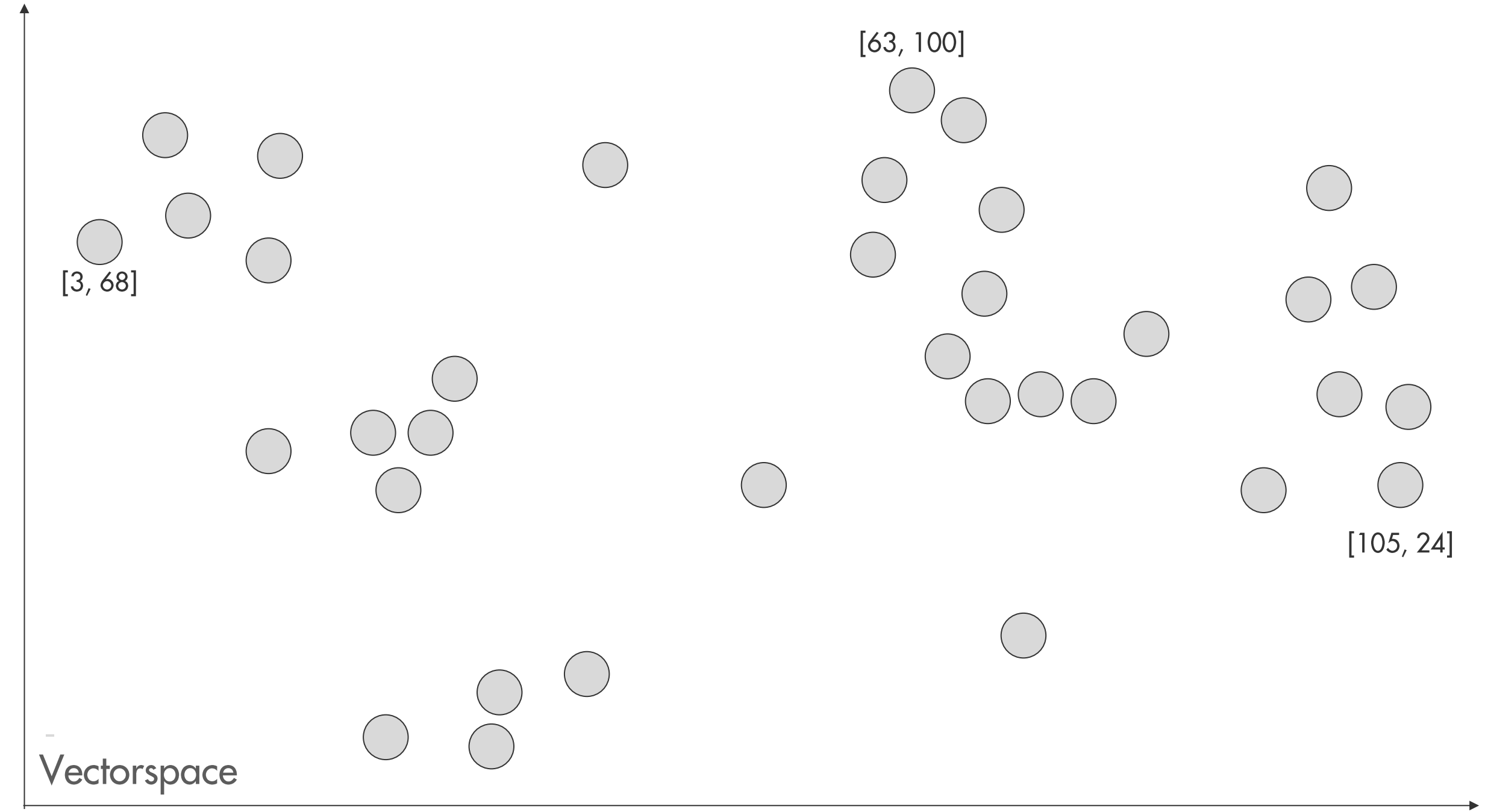
Coverage with optimized test ordering

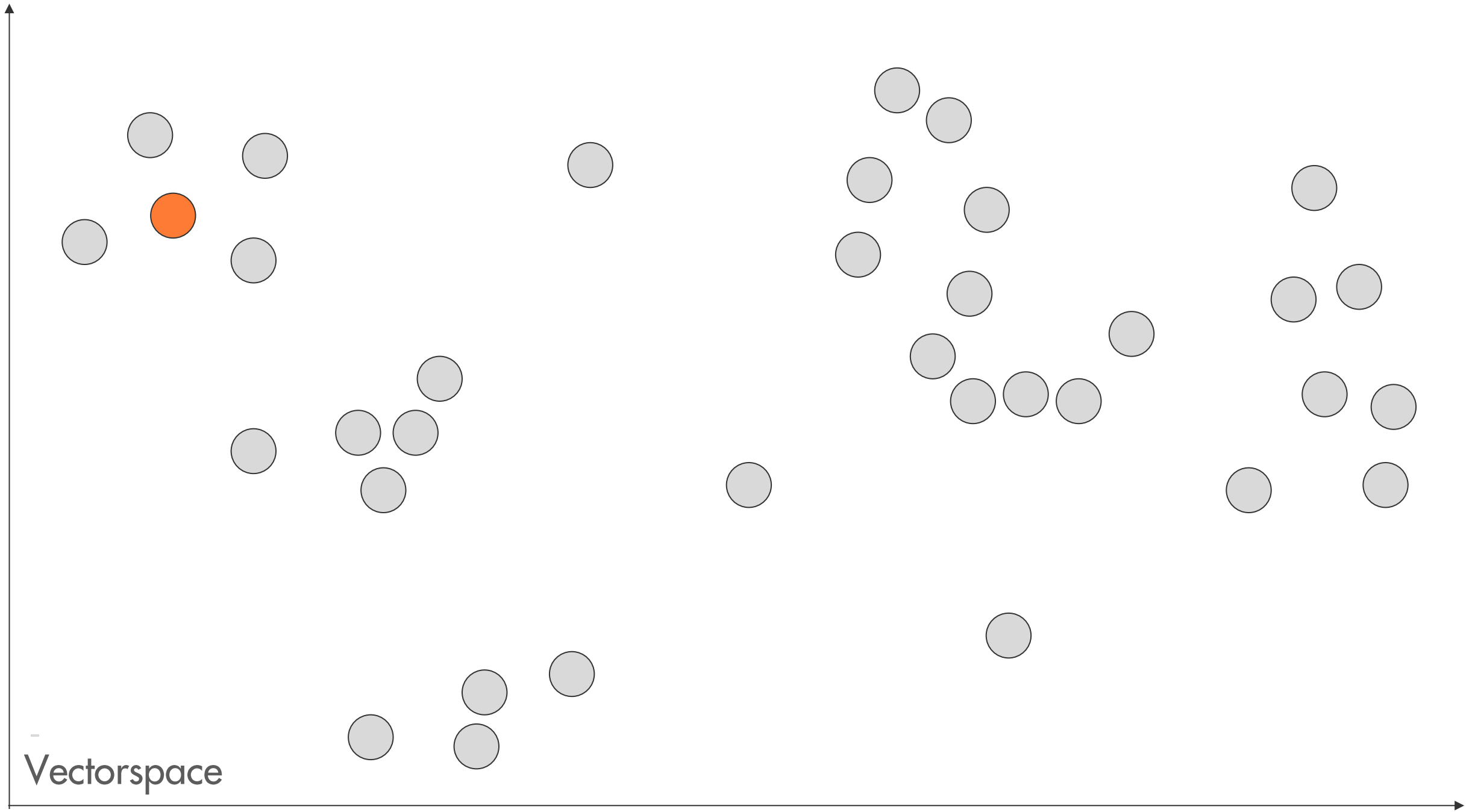


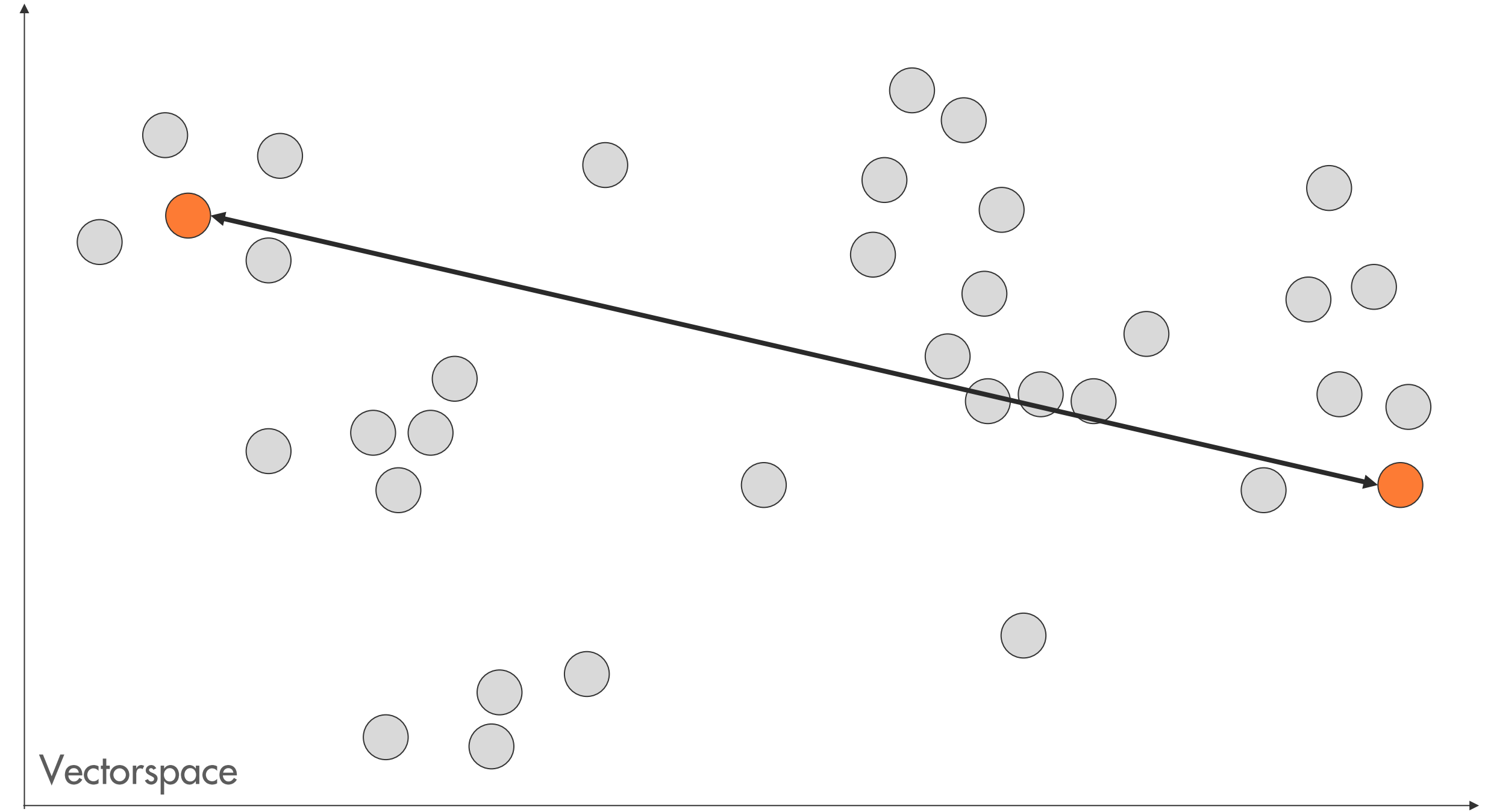
Results for Test Query & Budget Restriction

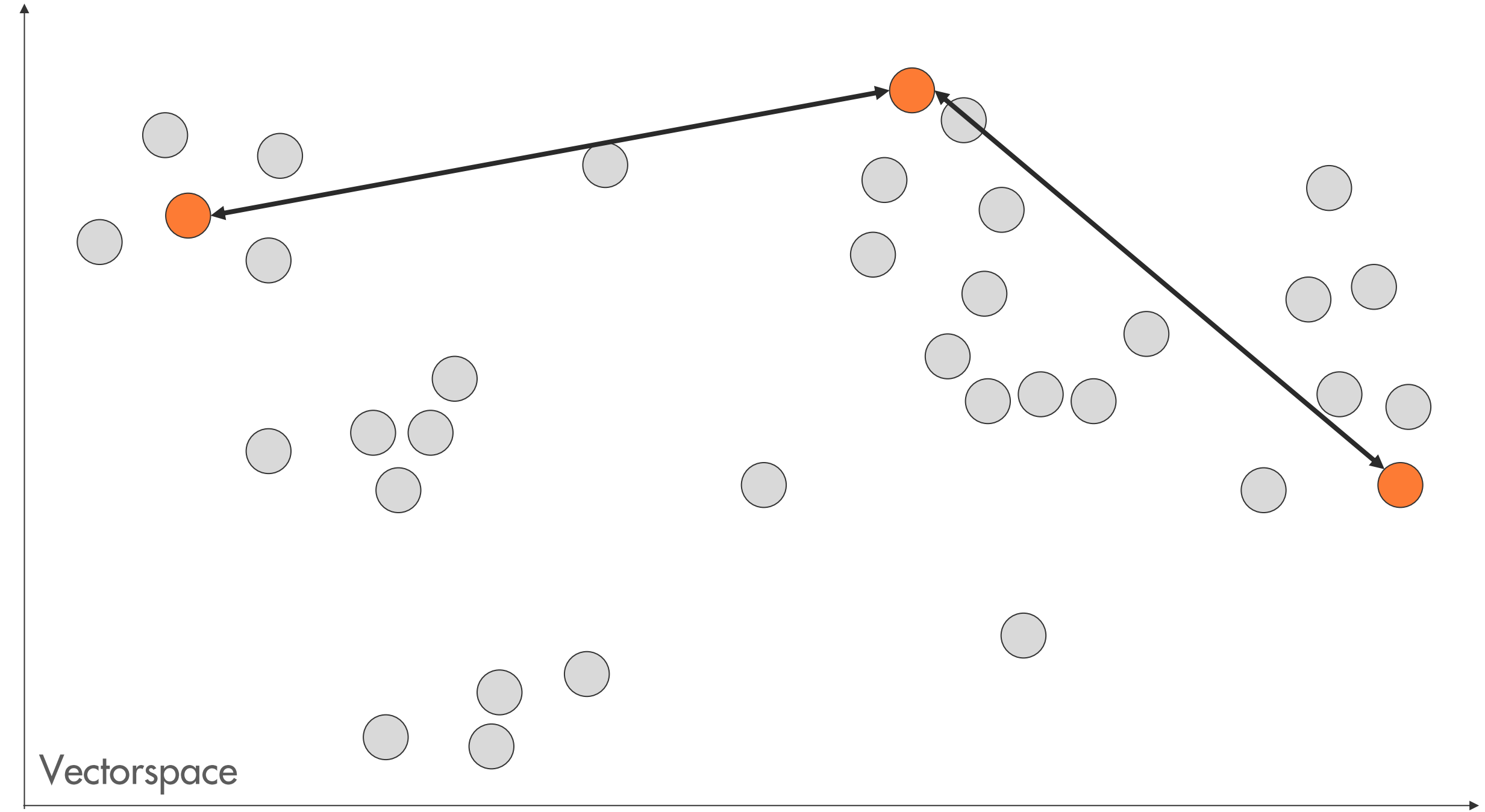
Relative Coverage: 100%, Selected Tests: 236 out of 236 (100%)

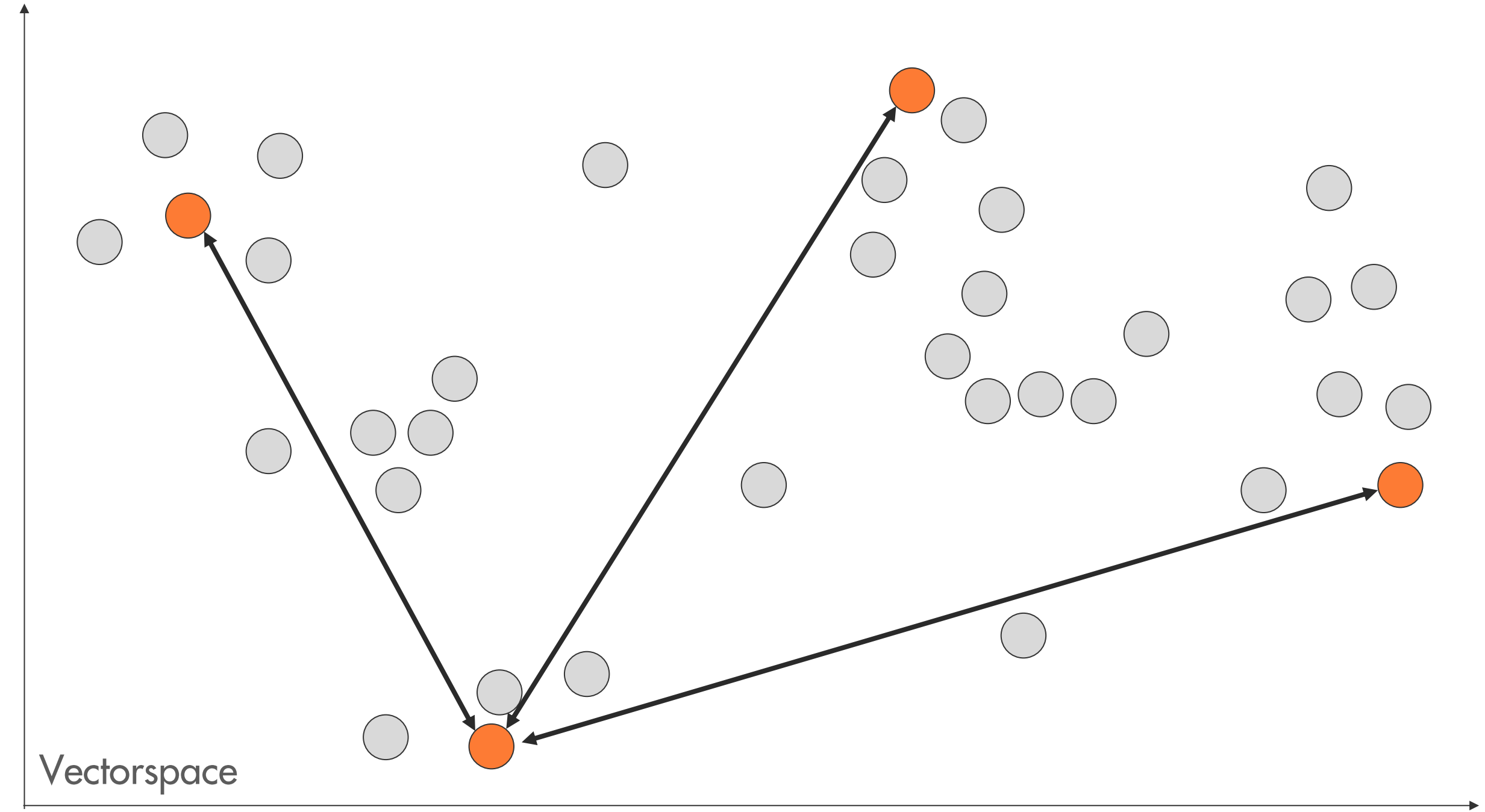
How can we identify such tests easily?


[63, 100]
[3, 68]
[105, 24]
Vectorspace


Vectorspace


Vectorspace


Vectorspace

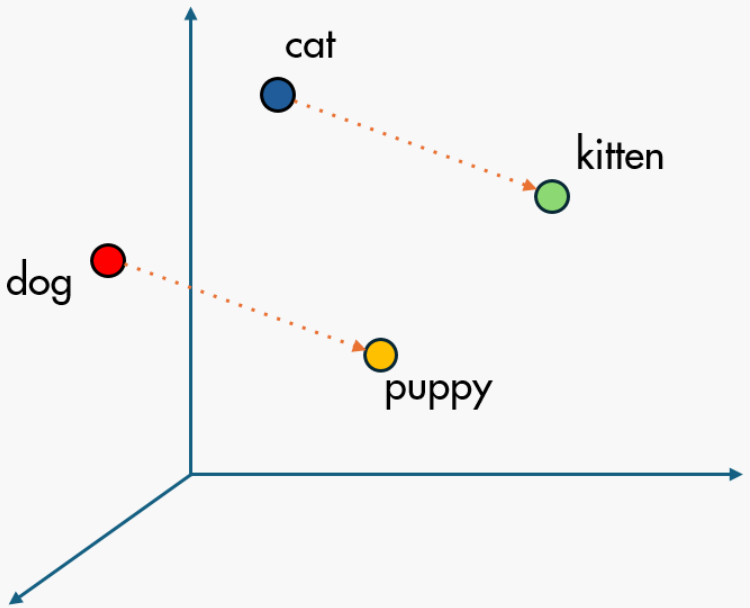

Vectorspace

Large Language ("AI") Models

	living being	feline	canine	young	pet	verb	plural
cat	0.85	0.92	-0.41	-0.30	0.74	-0.11	-0.16
kitten	-0.82	0.89	-0.44	0.88	0.79	-0.14	-0.04
dog	0.86	-0.45	0.91	-0.28	0.81	-0.08	-0.14
puppy	0.84	-0.42	0.89	0.90	0.85	-0.10	0.05

Word

Embedding



Visualization

manual

Test Steps

automated

Test Code

Local Modells

ChatGPT

Gemini

Claude

Embeddings

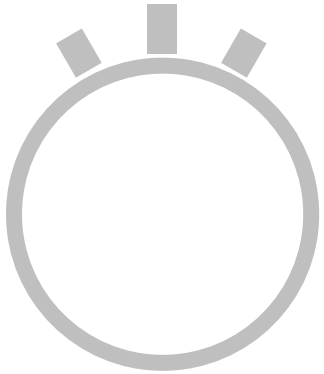
AI Test Clustering

How many tests should we keep?



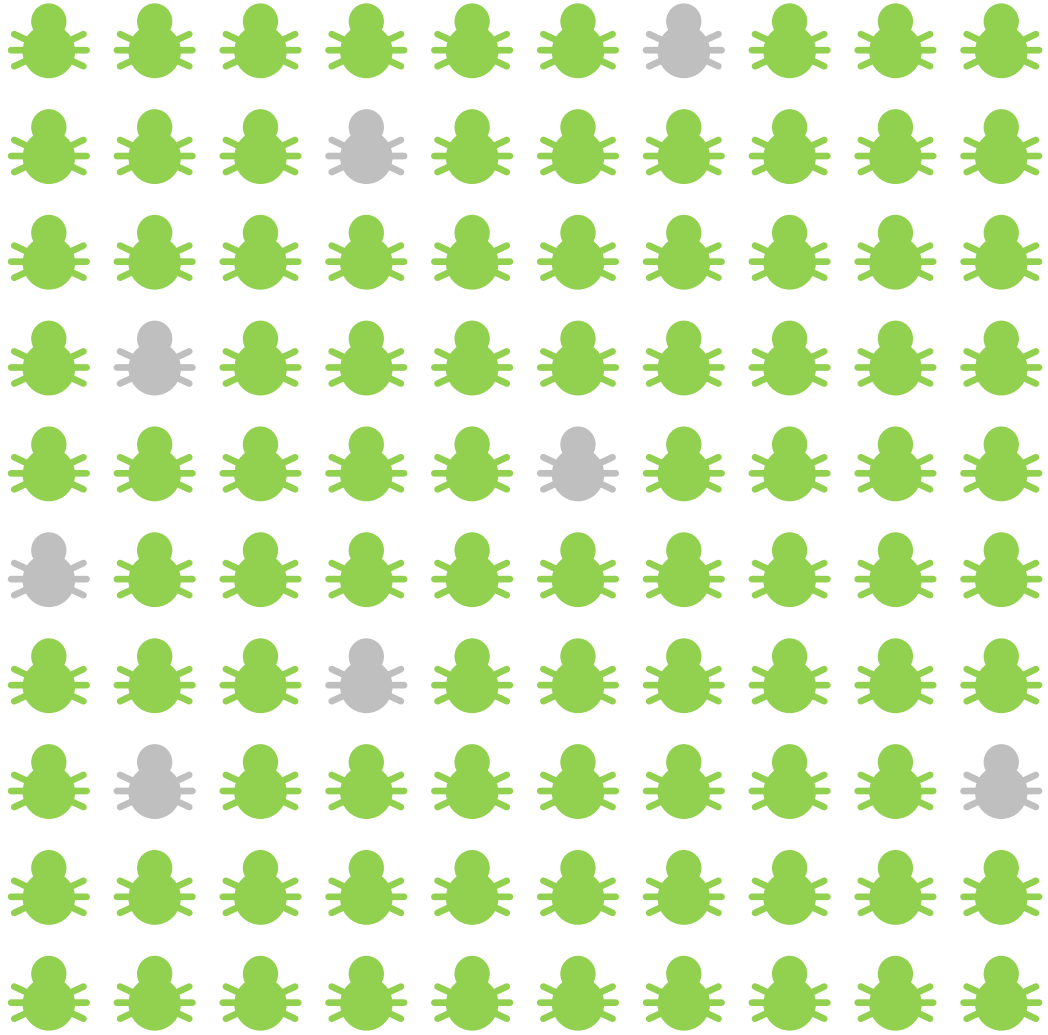
Tests to Keep

AI Test Clustering

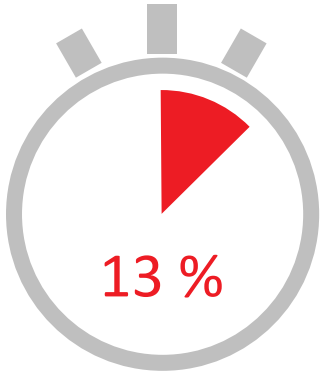




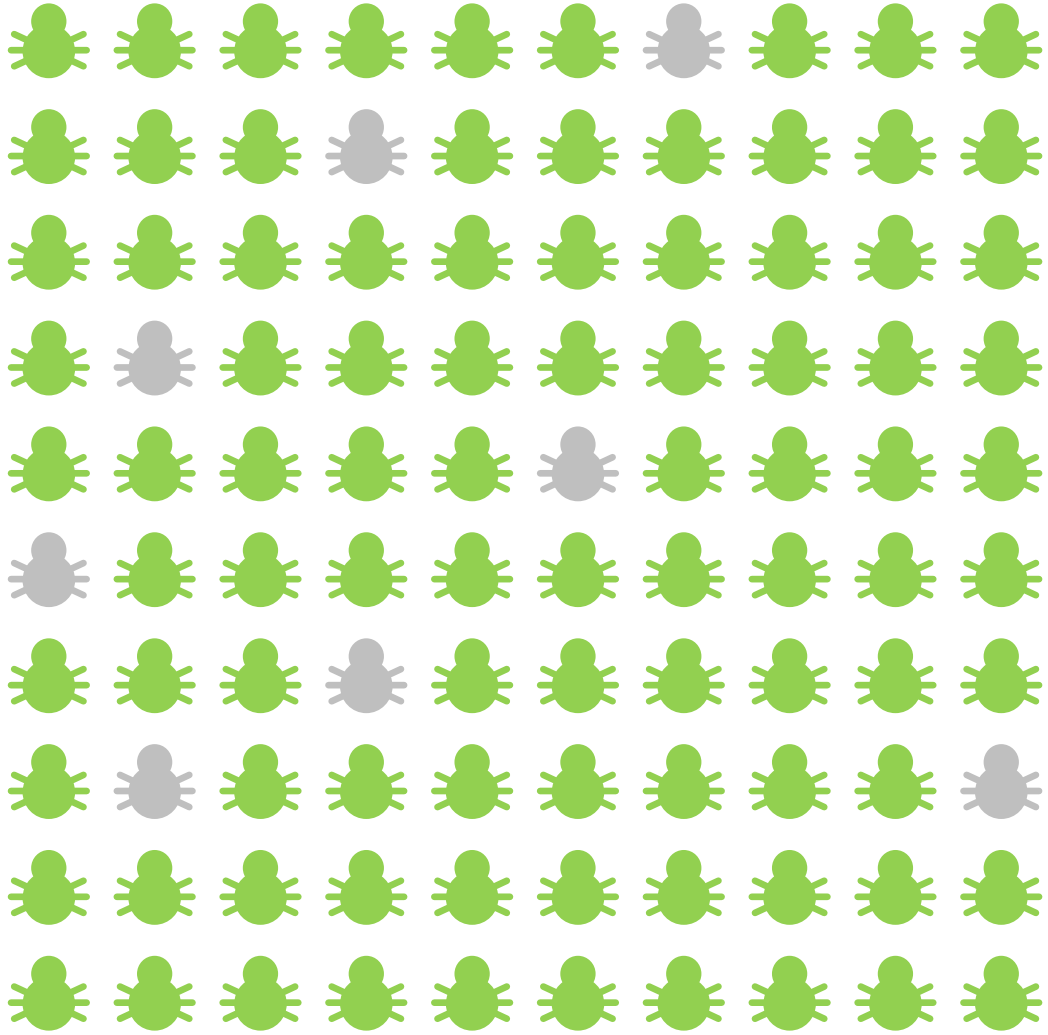
90 %



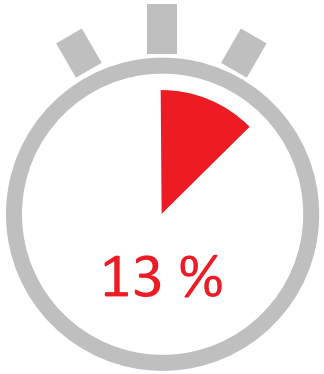
AI Test Clustering







AI Test Clustering

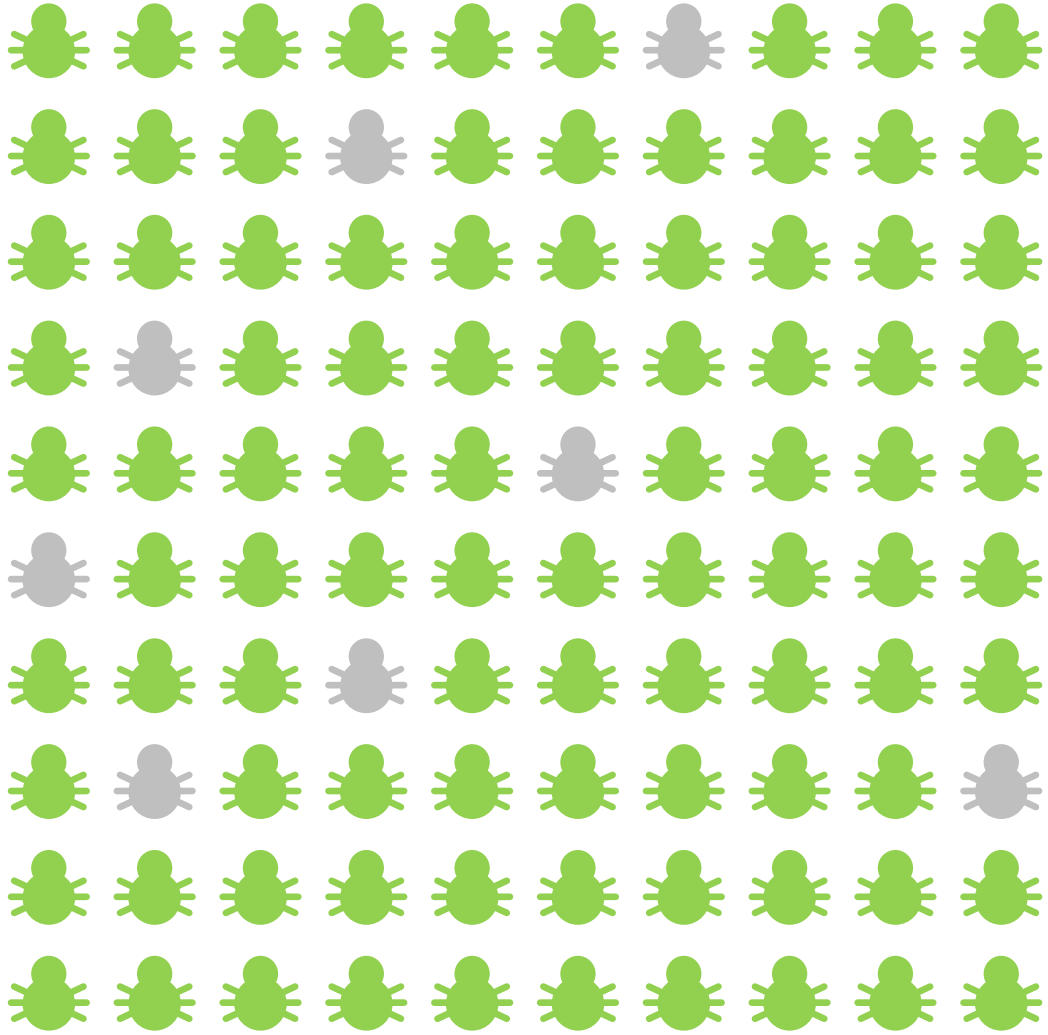




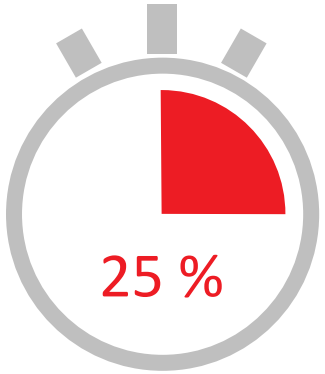
Maintenance effort

Migration costs

Automatization costs



AI Test Clustering

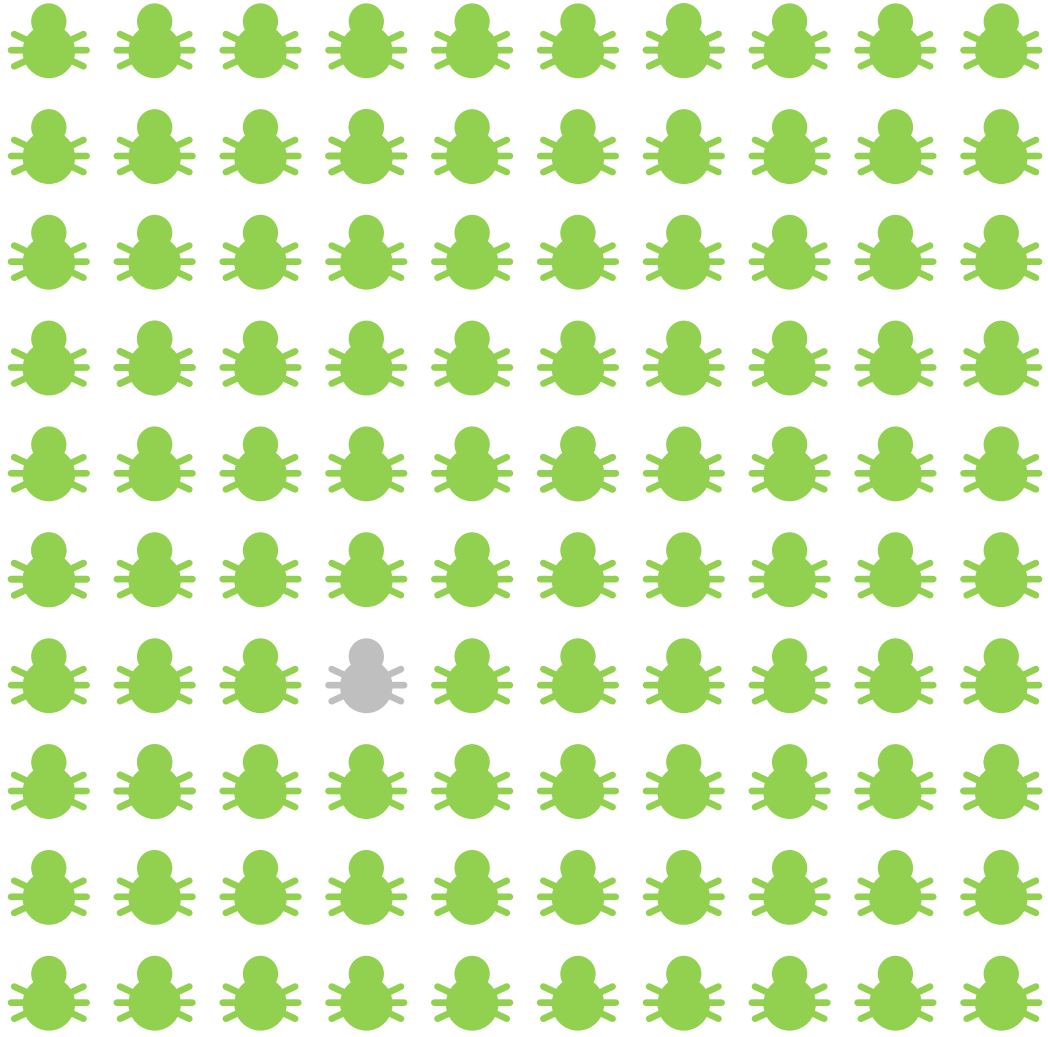




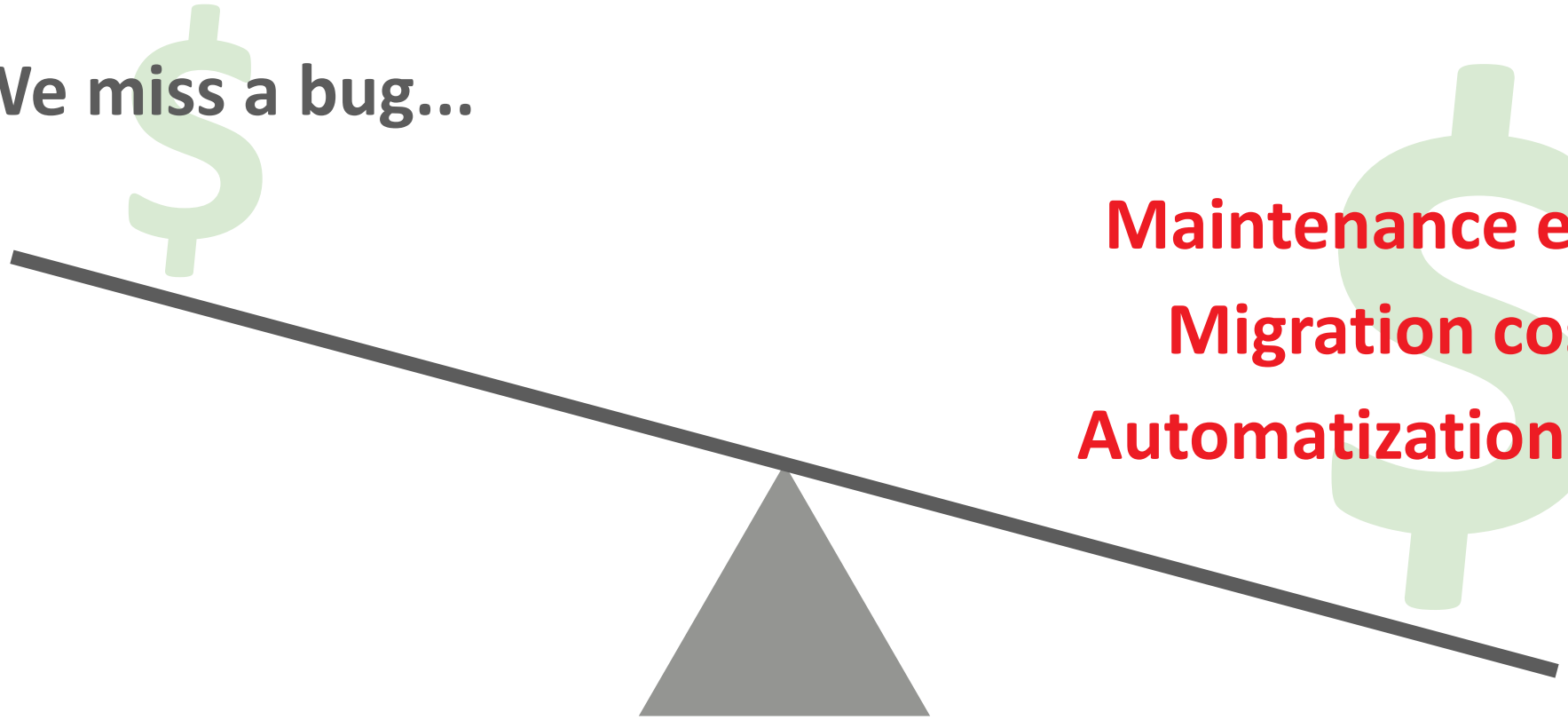
Maintenance effort

Migration costs

Automatization costs



We miss a bug...



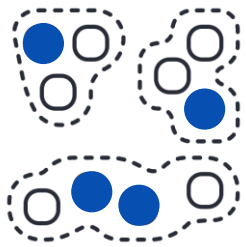
Maintenance effort

Migration costs

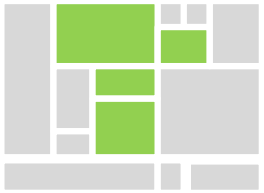
Automatization costs

What Does This Look Like in Practice?

Business-Information-System
Test-Suite Migration



Cluster & Selection



Coverage



Bugs



3400 Tosca tests

ChatGPT

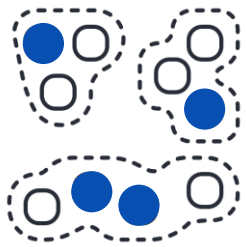
Embeddings

AI Test Clustering

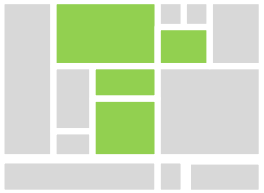
How many tests should we keep?



Tests to Keep



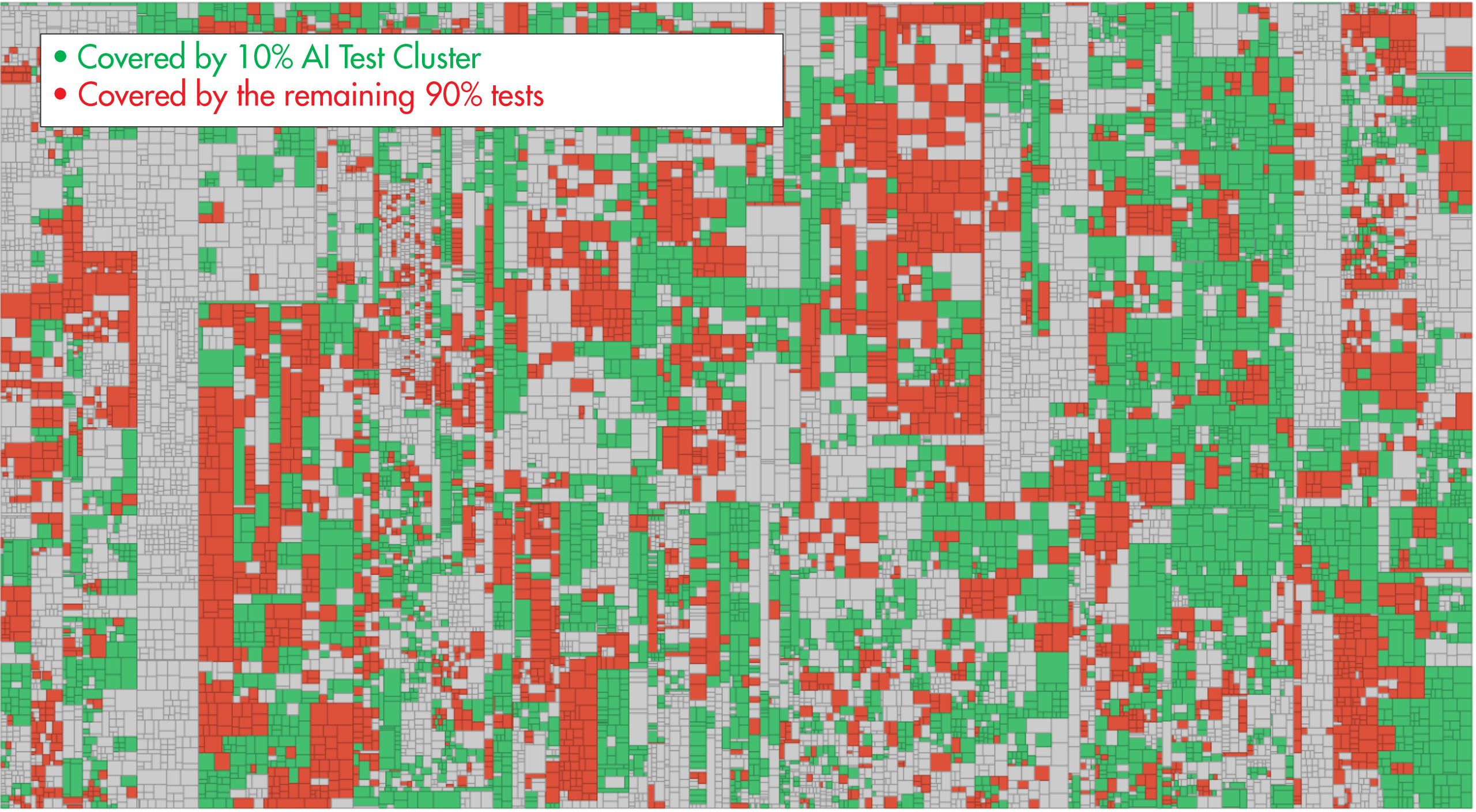
Cluster & Selection

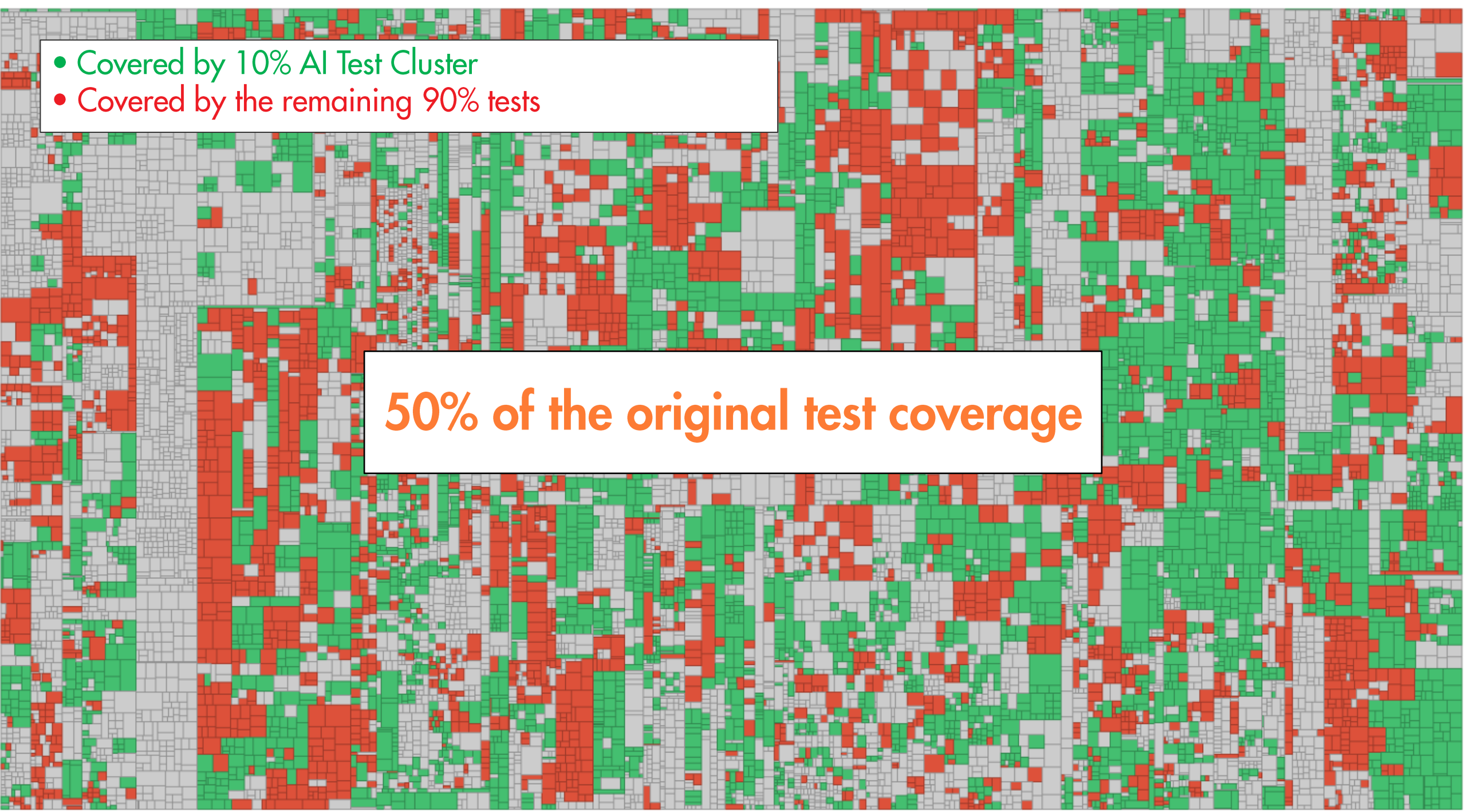


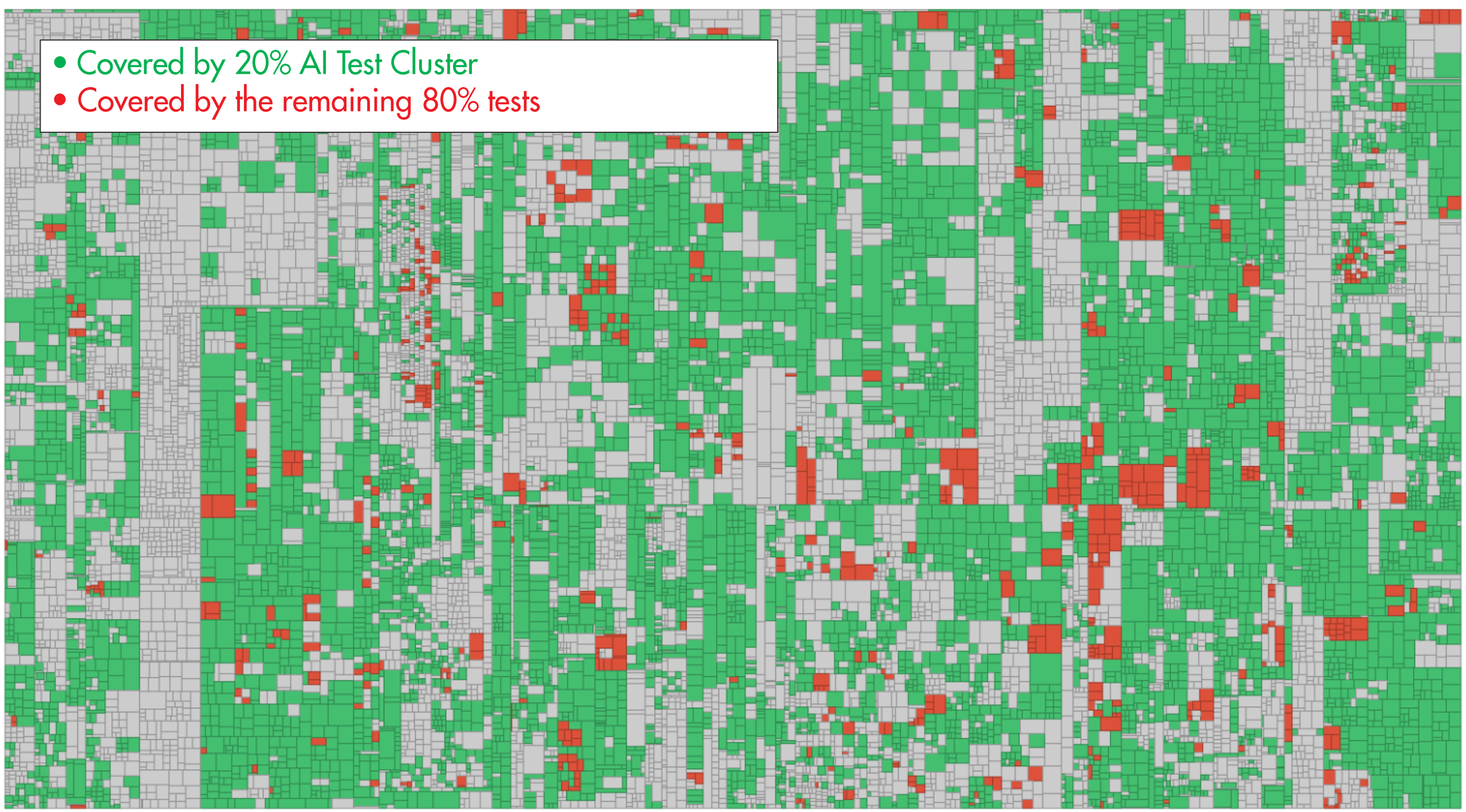
Coverage

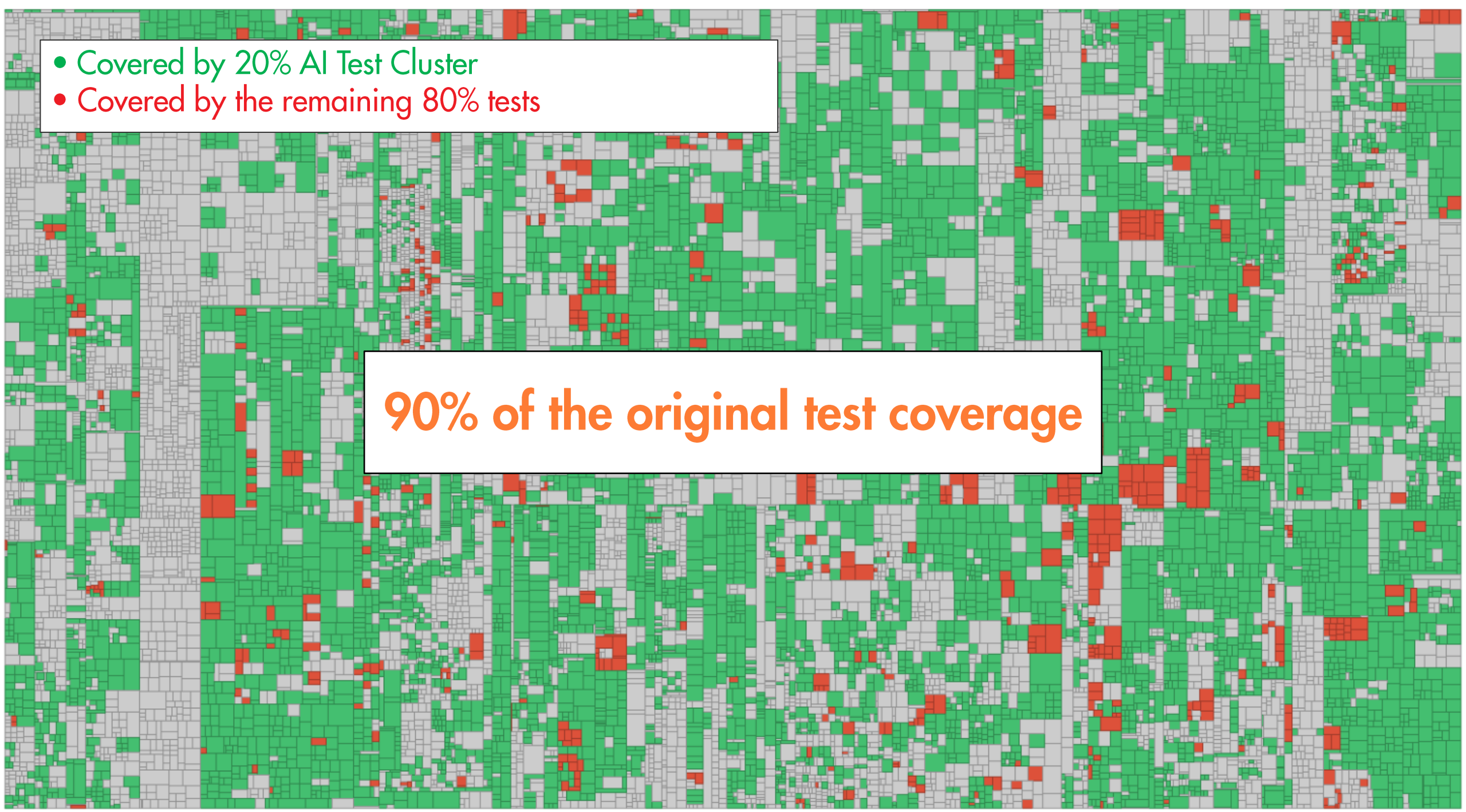


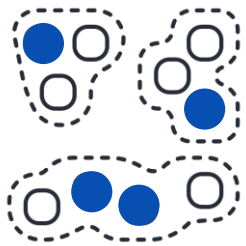
Bugs


• Covered by 10% AI Test Cluster
• Covered by the remaining 90% tests

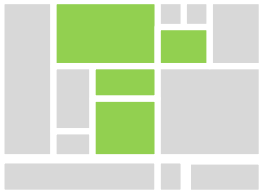

• Covered by 10% AI Test Cluster
• Covered by the remaining 90% tests
50% of the original test coverage


• Covered by 20% AI Test Cluster
• Covered by the remaining 80% tests


• Covered by 20% AI Test Cluster
• Covered by the remaining 80% tests
90% of the original test coverage



Cluster & Selection



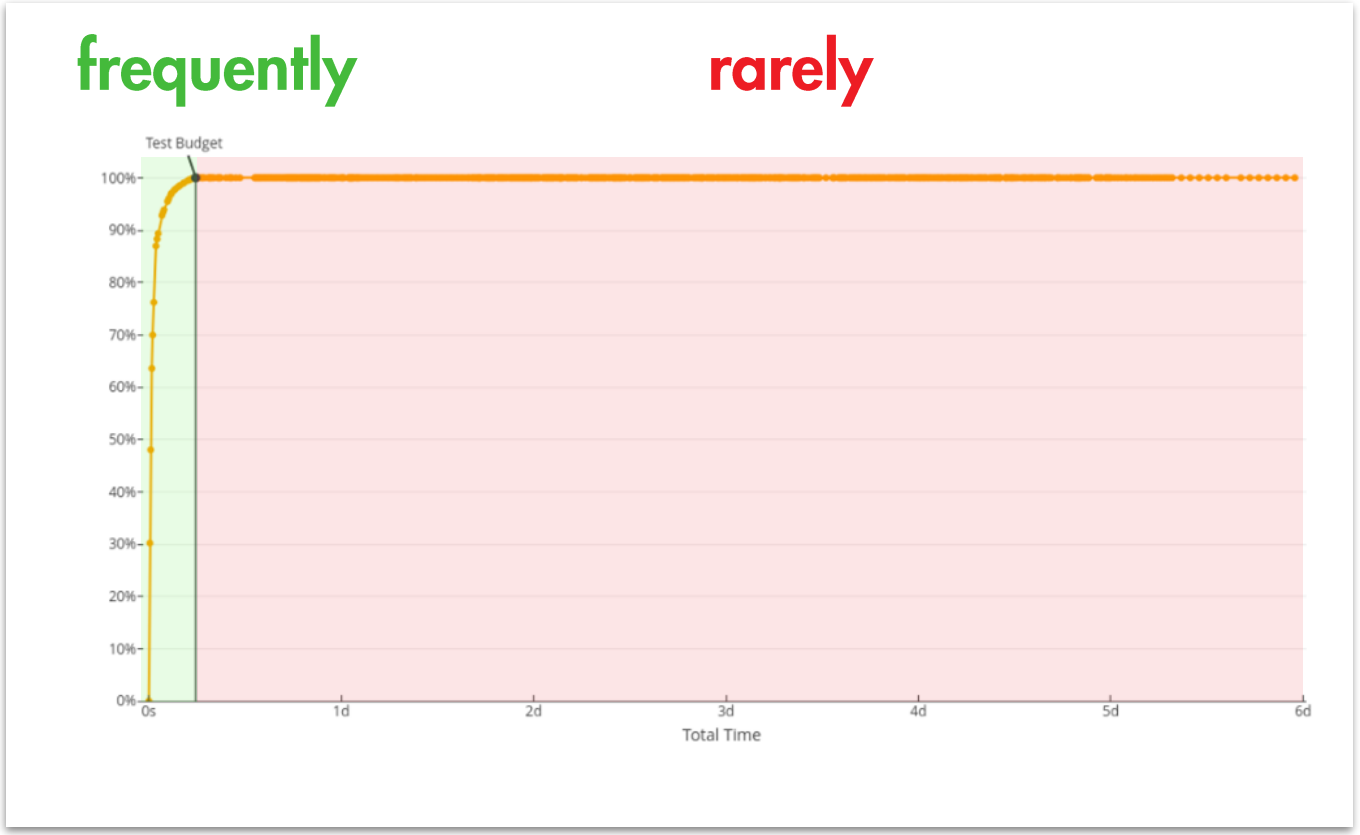
Coverage



Bugs



Quarantine & Evaluation

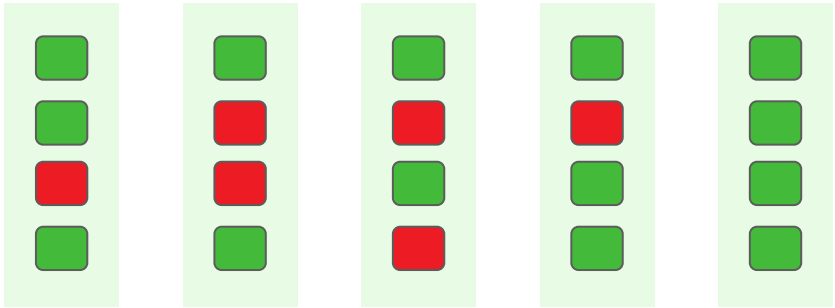


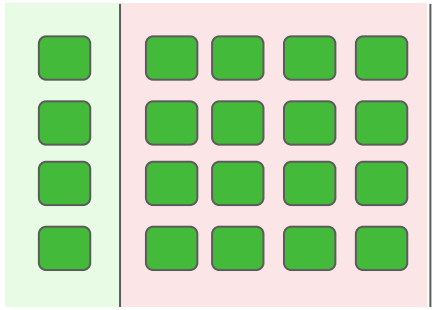
AI Test Clustering Mo - Fr

20%

Weekend

All tests





Result

> 100

Person days / Year

less for release tests

> 400

Person days

saved during the migration

Business-Information-System II

Shift Left

Industry Example

Originally: nightly „Re-run all“

Feedback late on the next day



Build

Test

Industry Example

Originally: nightly „Re-run all“

Feedback late on the next day



Build Test

During the day: 1h test budget with minimized test suite



Build Test 1h

Nightly: „Re-run all“



Industry Example

Originally: nightly „Re-run all“

Feedback late on the next day



Build Test

In only 5% of the test execution time we find **70% of the failing commits**

During the day: 1h test budget with minimized test suite



Build Test 1h

Nightly: „Re-run all“



Summary

Summary

Historically grown test-suites are highly redundant.

Summary

Historically grown test-suites are highly redundant.

AI Test Clustering can identify the few valuable tests that are able to identify most defects in a short amount of time.

Summary

Historically grown test-suites are highly redundant.

AI Test Clustering can identify the few valuable tests that are able to identify most defects in a short amount of time.

Often, it is better to fill gaps after the migration with new test instead of migration all tests.

Feedback: agiletestingdays.com/session-ratings

Thanks for Listening

Get my slides and future talks





Visit us at our booth ☺

Teamscale by CQSE

Thanks for your attention!

Feedback welcome!

Go to **agiletestingdays.com/session-ratings**
and give your rating!

