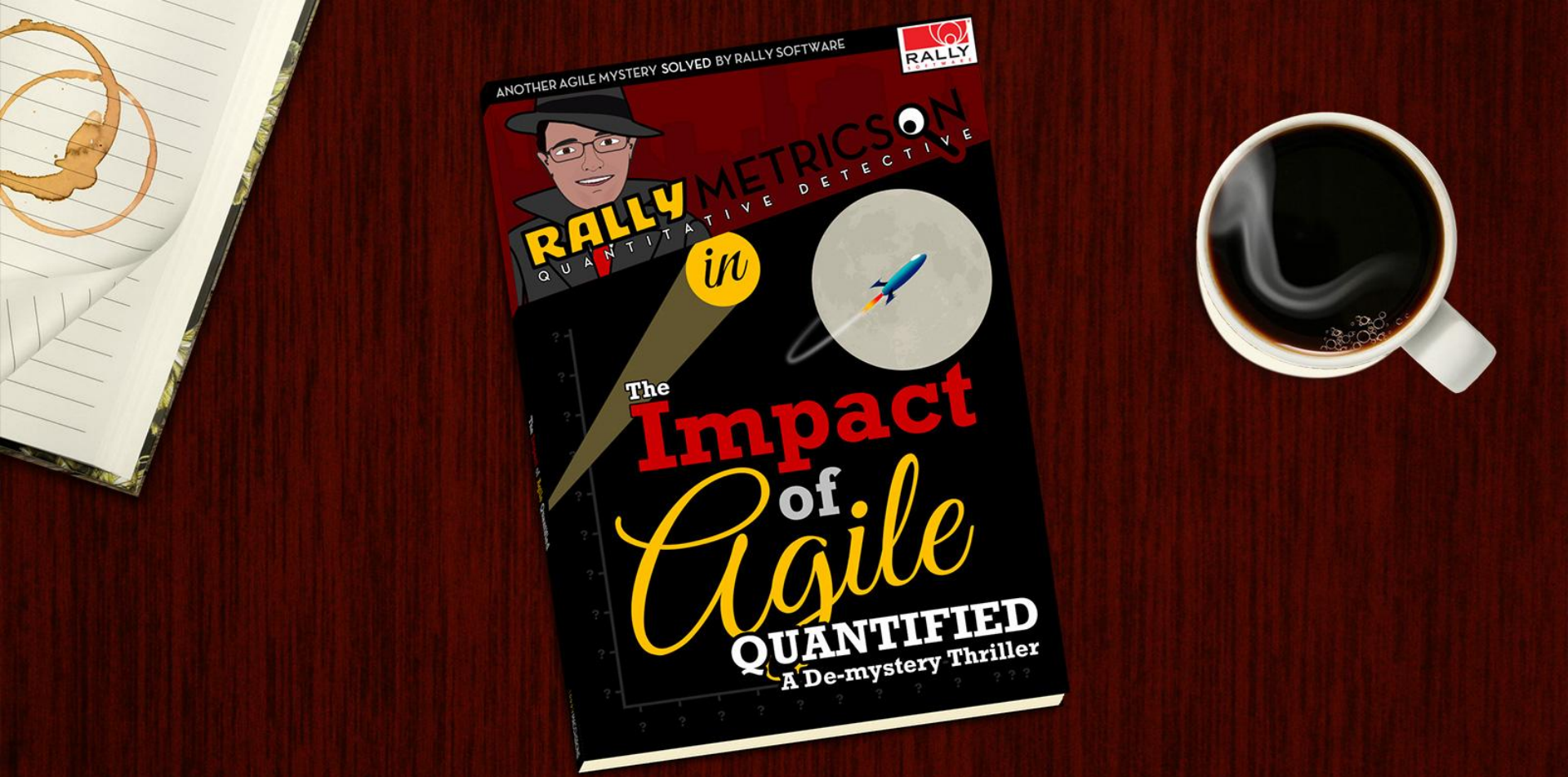


My sidekick in this story? That's Larry Maccherone. He's worked in this town his entire professional life.

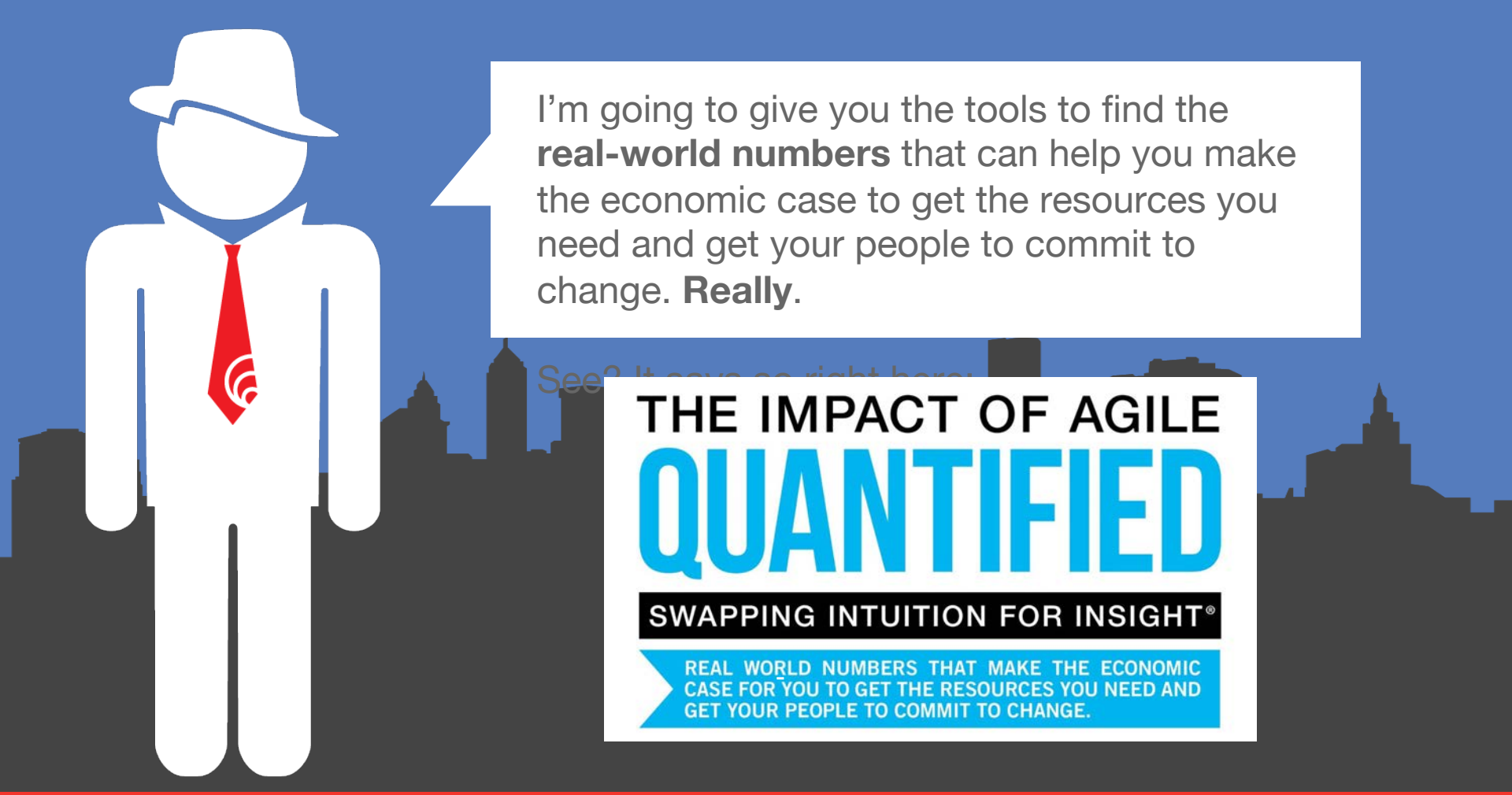
The Rally Analytics Edge!

A Scientist

- **Larry Maccherone**, a serial entrepreneur and working on his PhD at Carnegie Mellon
- Partnered with Watts Humphries, CMM founder
- To transform how Software is delivered
- Joined Rally in 20xx, and heads up Analytics and the Insights Products

Rally Data

- Rally has the largest set of Agile data on it's multi tenant SaaS database
- Tens of thousands of Agile teams
- Hundreds of thousands of projects



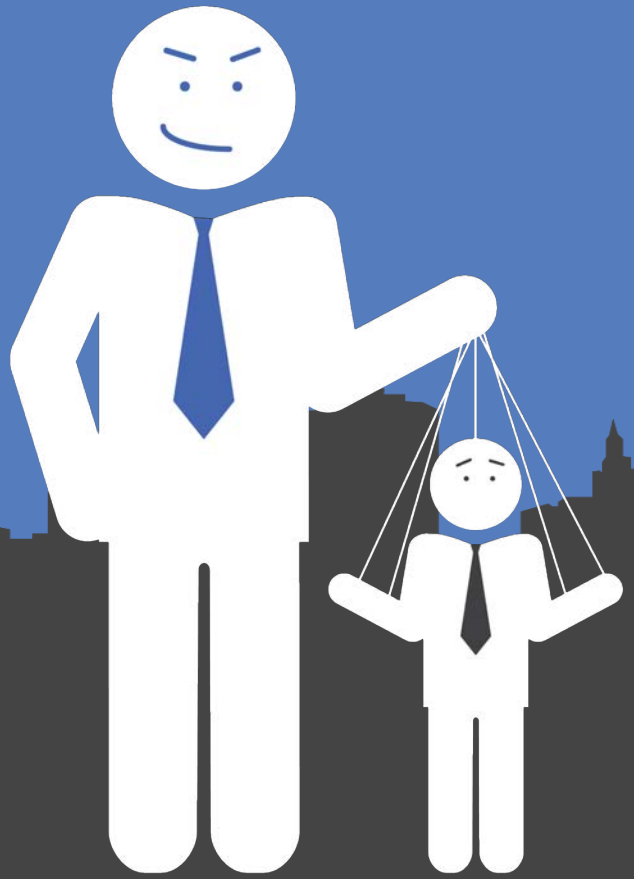
The Seven Deadly Sins of Agile Measurement

1. Manipulating Others
2. Unbalanced Metrics
3. Quantitative Idolatry
4. Overpriced Metrics
5. Lazy Metrics
6. Bad Analysis
7. Linear Forecasting

Sin #1

Manipulating Others

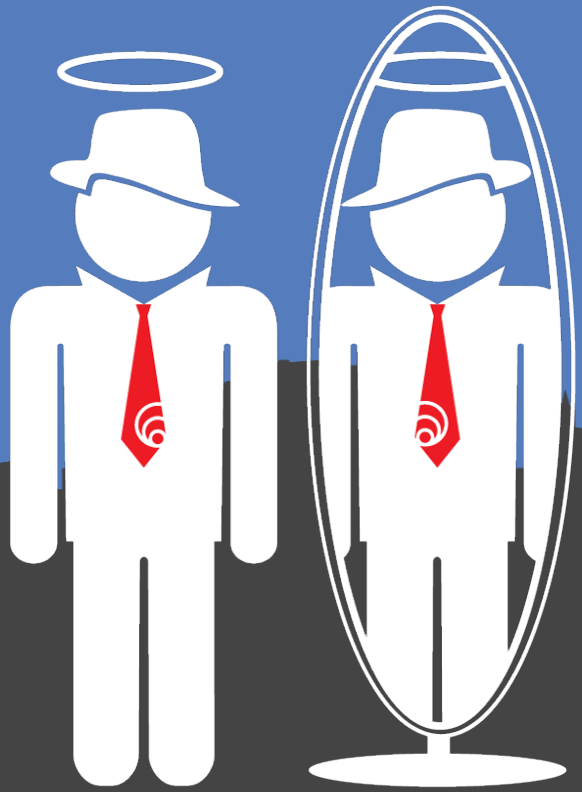
Using metrics as a lever to drive someone else's behavior



Heavenly Virtue #1

Self Improvement

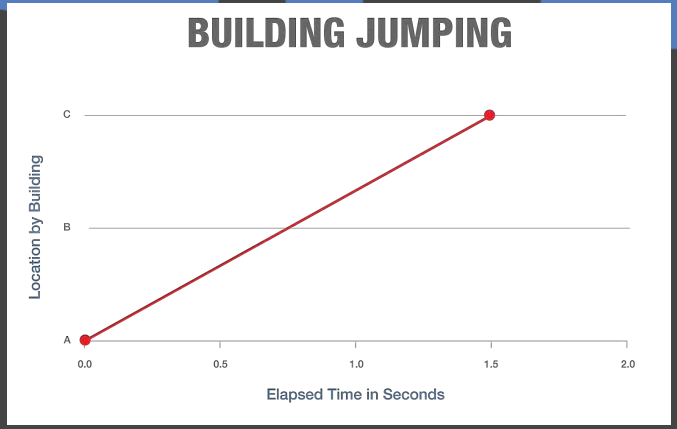
Using metrics to reflect on your own performance



Sin #7

Linear Forecasting

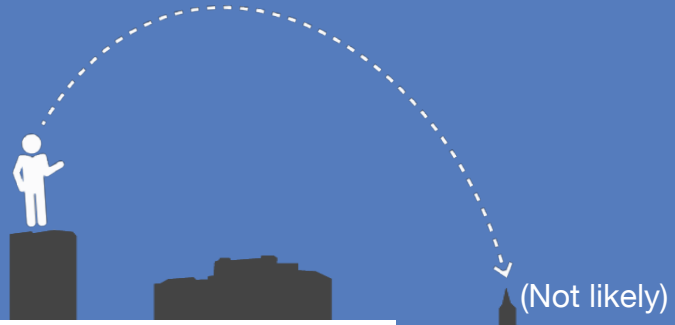
Forecasting without discussing probability and risk

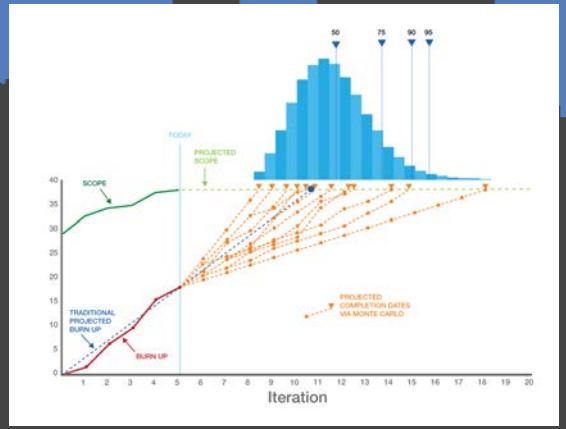


Heavenly Virtue #7

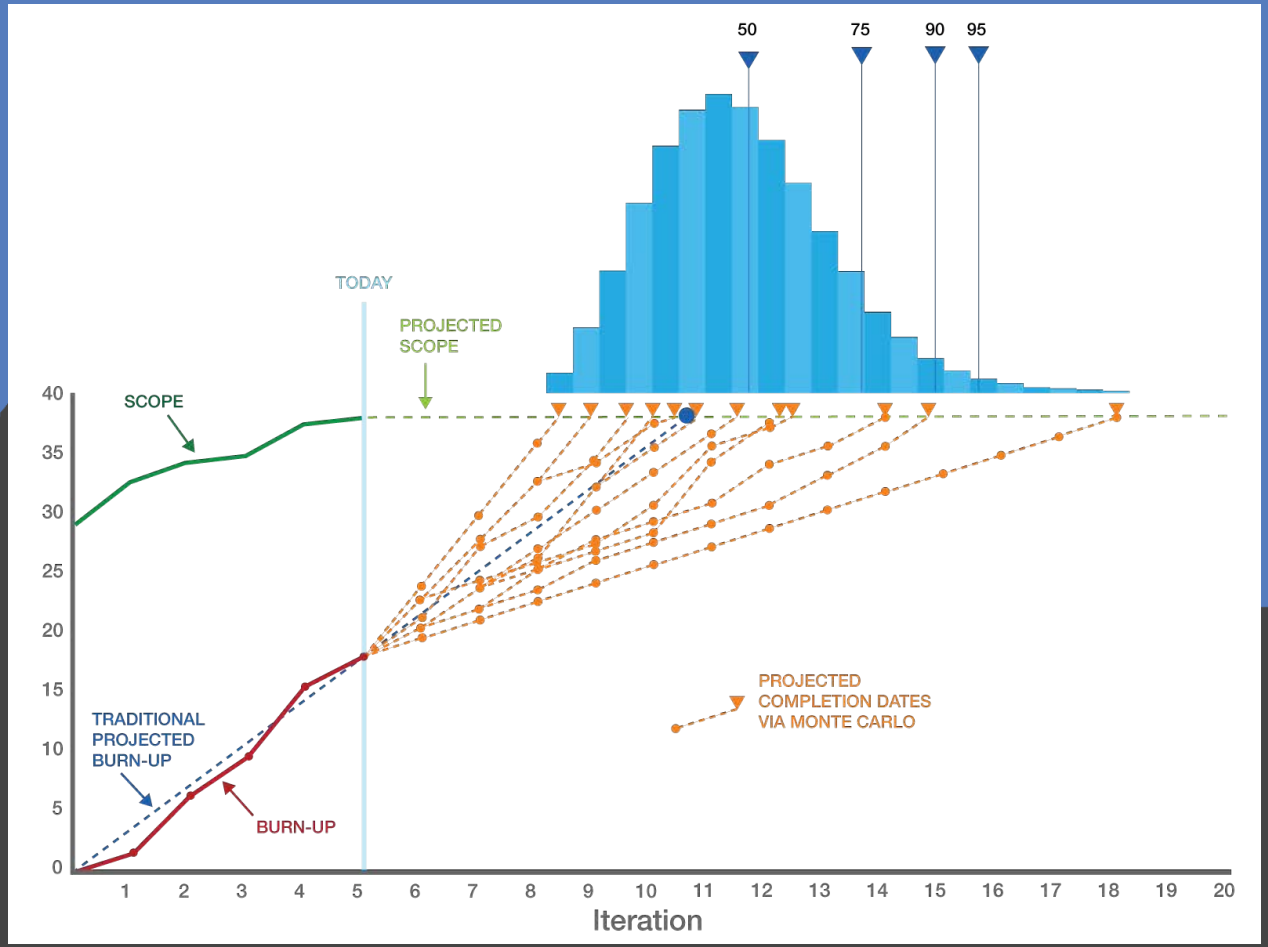
Probability Tools

Using the proper tools to predict the likelihood of results



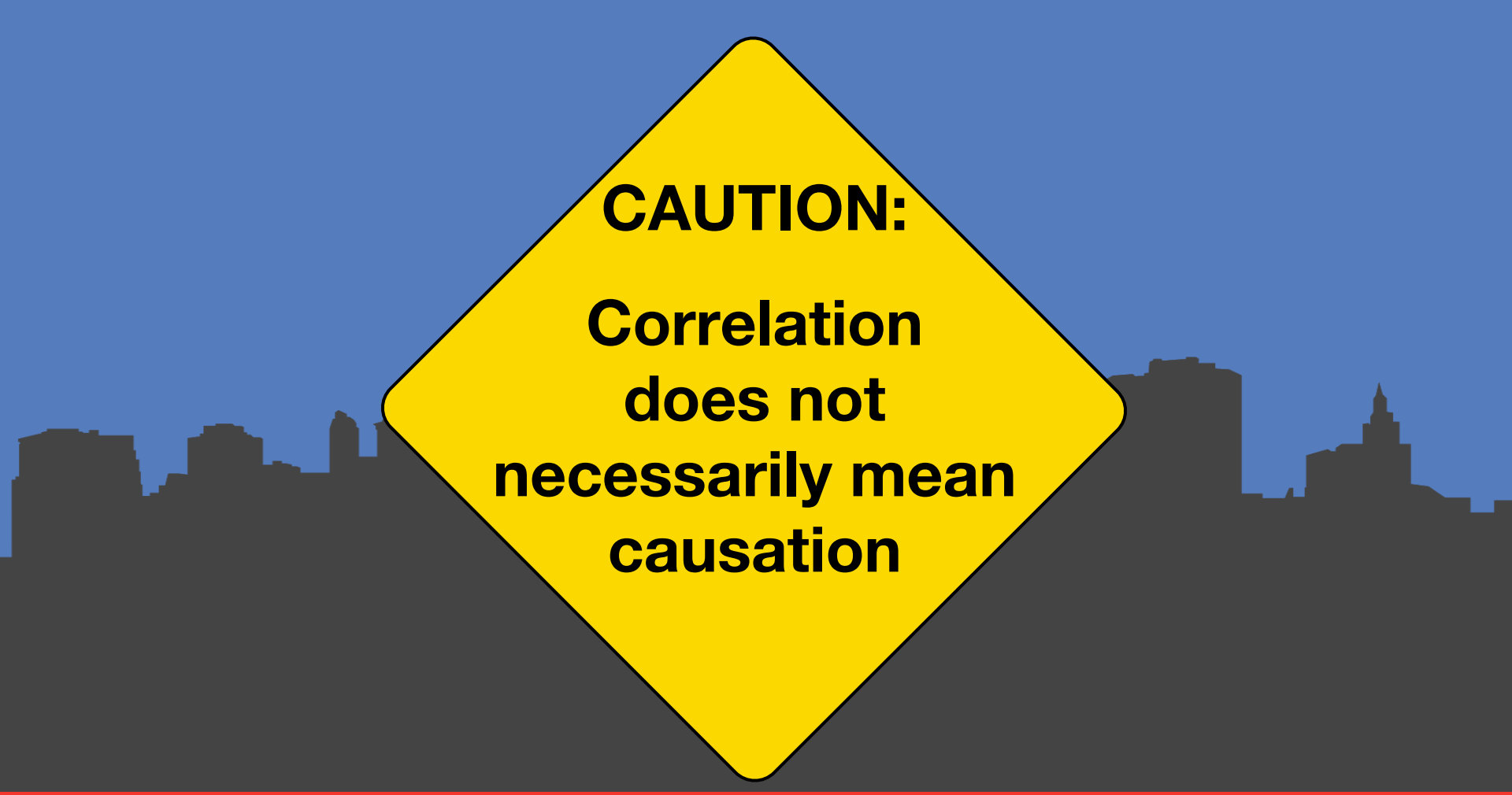


Monte Carlo Simulation



The Seven Heavenly Virtues of Agile Measurement

1. Self-Improvement
2. Balanced Metrics
3. Quantitative Perspective
4. Occam's Metrics
5. ODIM
6. Informed Analysis
7. Probablity Tools

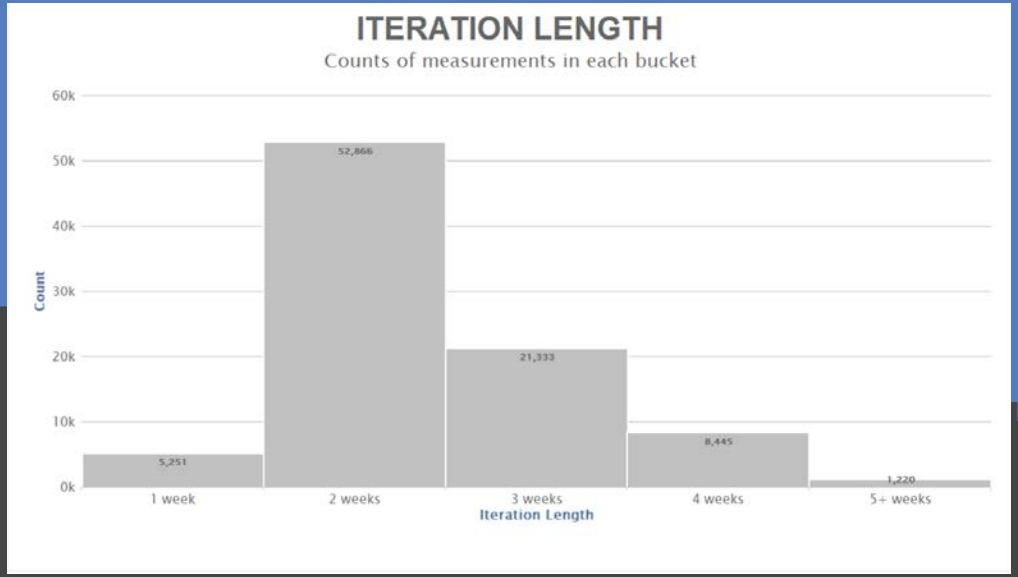


The investigation continues with ...

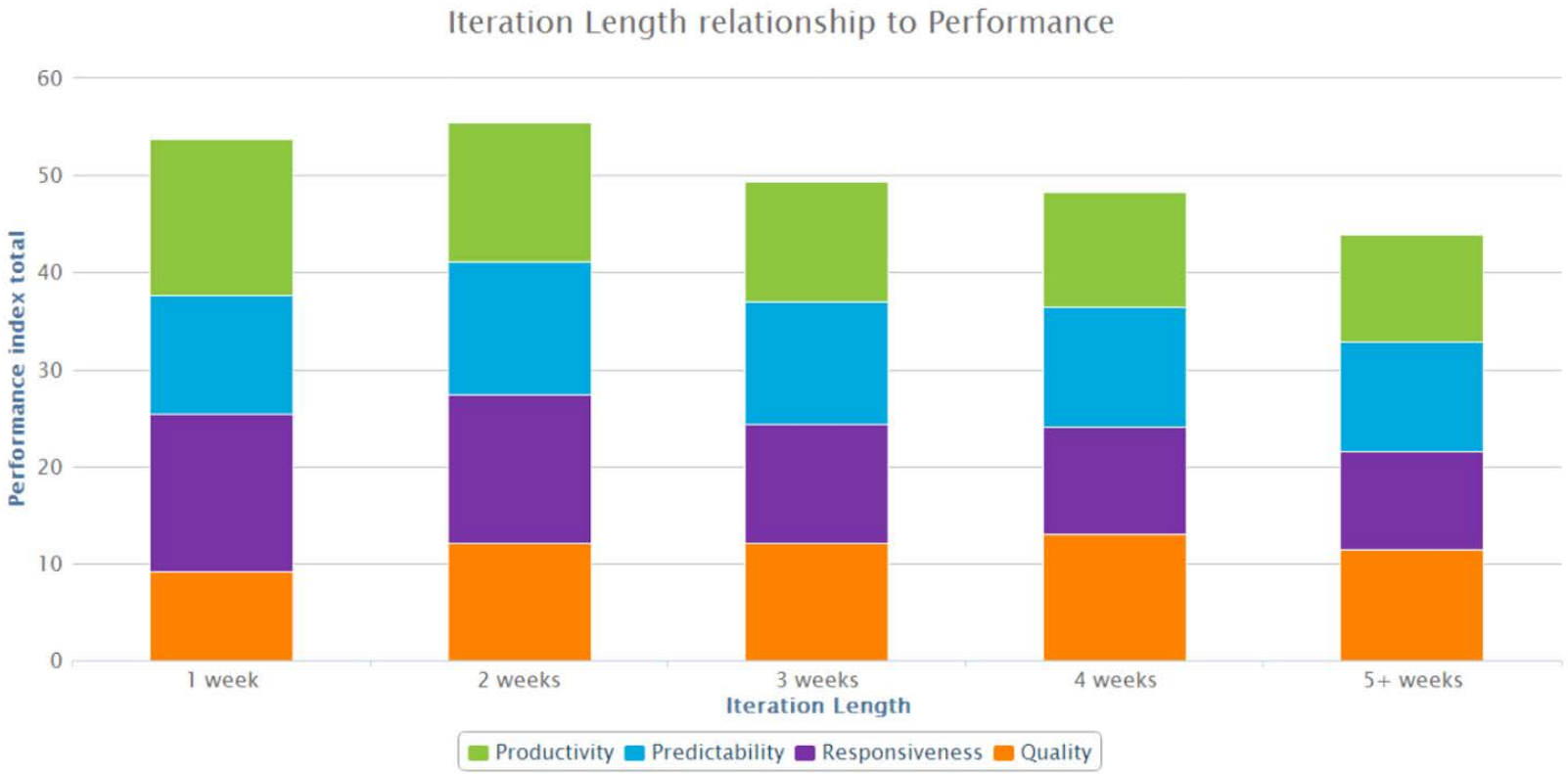
Iteration length

Crowd wisdom or shared delusion?

Iteration length	Teams using
1 week	6.2%
2 weeks	59.1%
3 weeks	23.4%
4 weeks	9.8%
5+ weeks	1.5%



PERFORMANCE

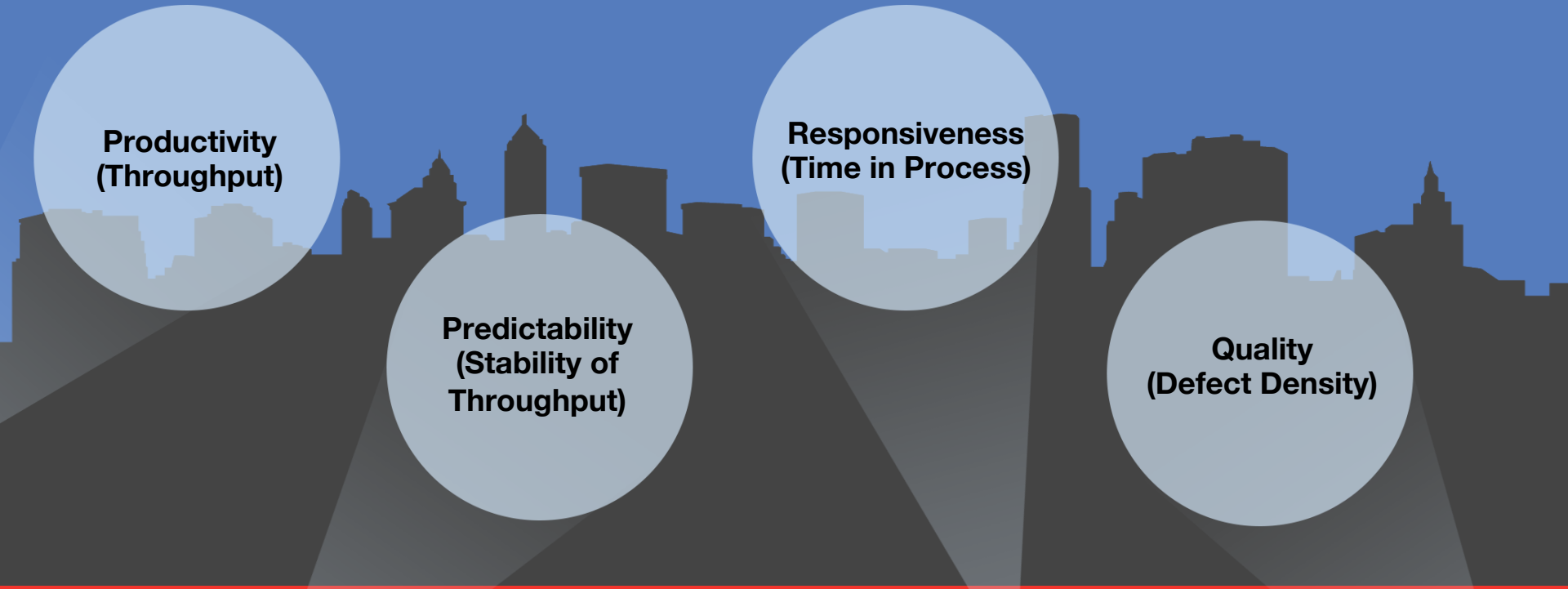


SOFTWARE DEVELOPMENT PERFORMANCE INDEX

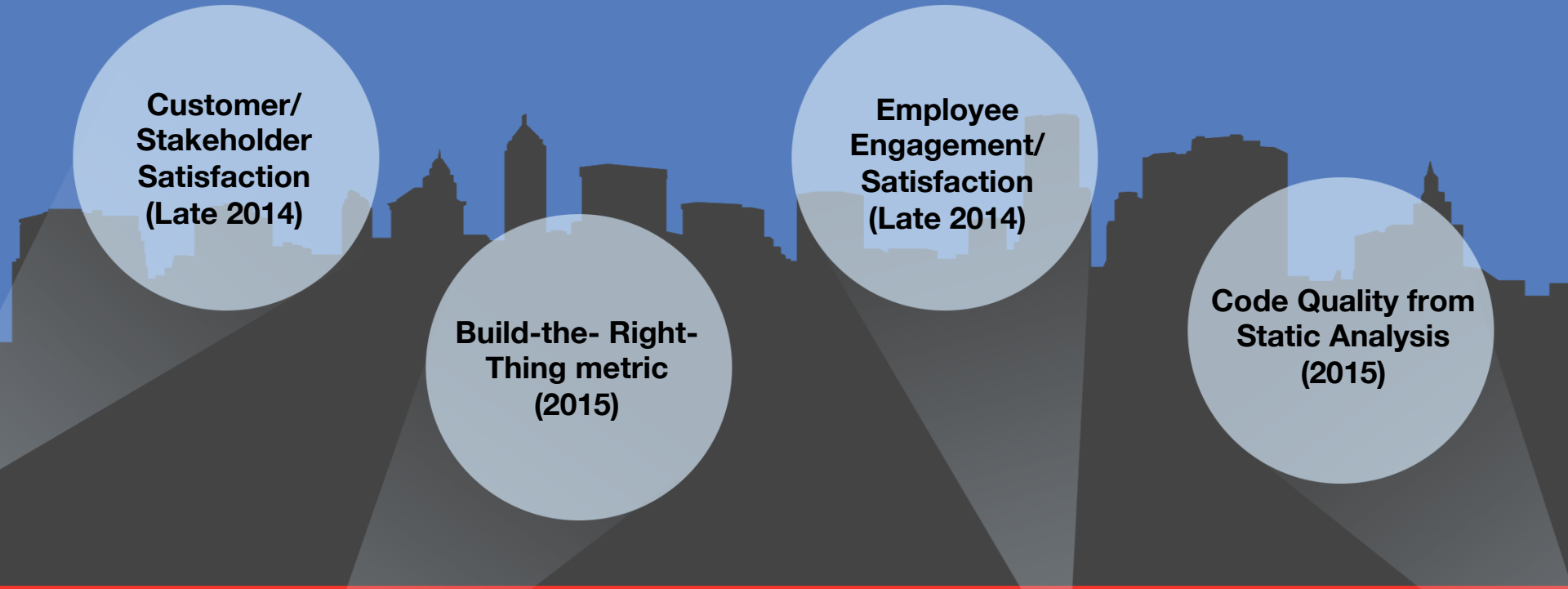
SDPI

@BrentChalker | BChalker@rallydev.com | #RallySoftware

SDPI current dimensions



Future SDPI dimensions



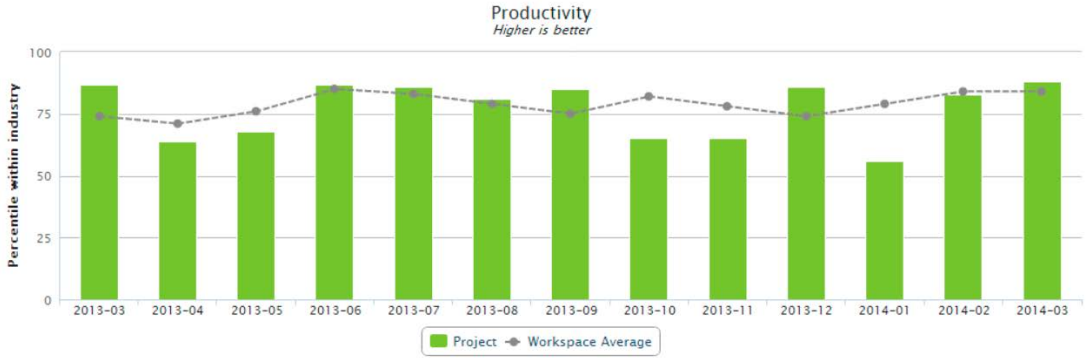
The investigation continues with ...

SDPI dimensions

Productivity = Throughput

Throughput is simply the count of User Stories completed in a given time period.

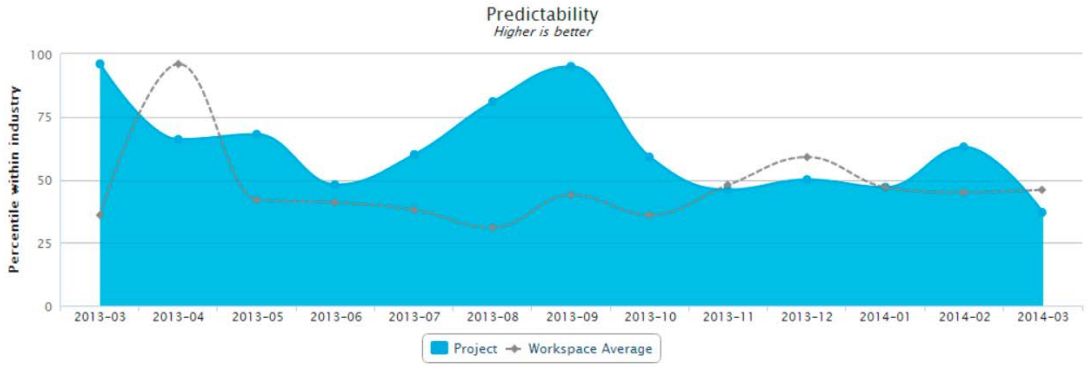
Productivity (by default) is the percentile scoring of the raw Throughput metric for User Stories normalized by team size.



Predictability = Stability of Throughput

Predictability measures how consistent you are at producing the same amount of work each month as measured by the Coefficient of Variation (CoV) of Throughput.

Predictability (by default) is the percentile scoring of the raw CoV of Throughput.



Responsiveness = Time in Process

TiP shows how long it takes to get one work item through your system. It's the work days that a User Story spends in development and testing. Similar to lead time or cycle time.

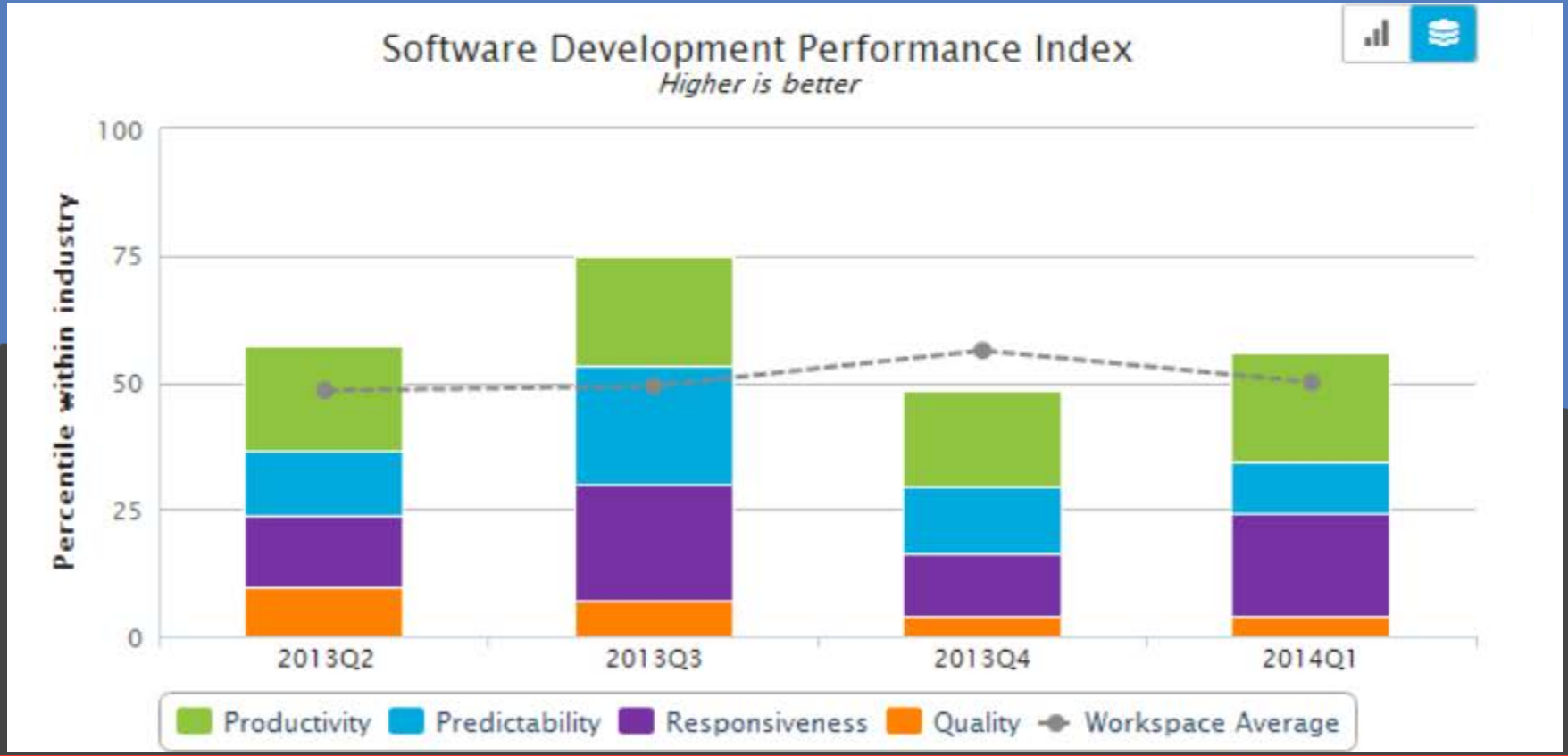
Responsiveness (by default) is the percentile scoring of the raw Time In Process (TiP) metric for User Stories.

Quality = Defect Density

Defect Density is a representation of the number of defects found in your code. It's the count of defects found in a given time period, normalized by team size.

Quality (by default) is the percentile scoring of the raw defect density metrics for both defects found in test as well as those found in production.

Raw metrics → Percentiles = Index

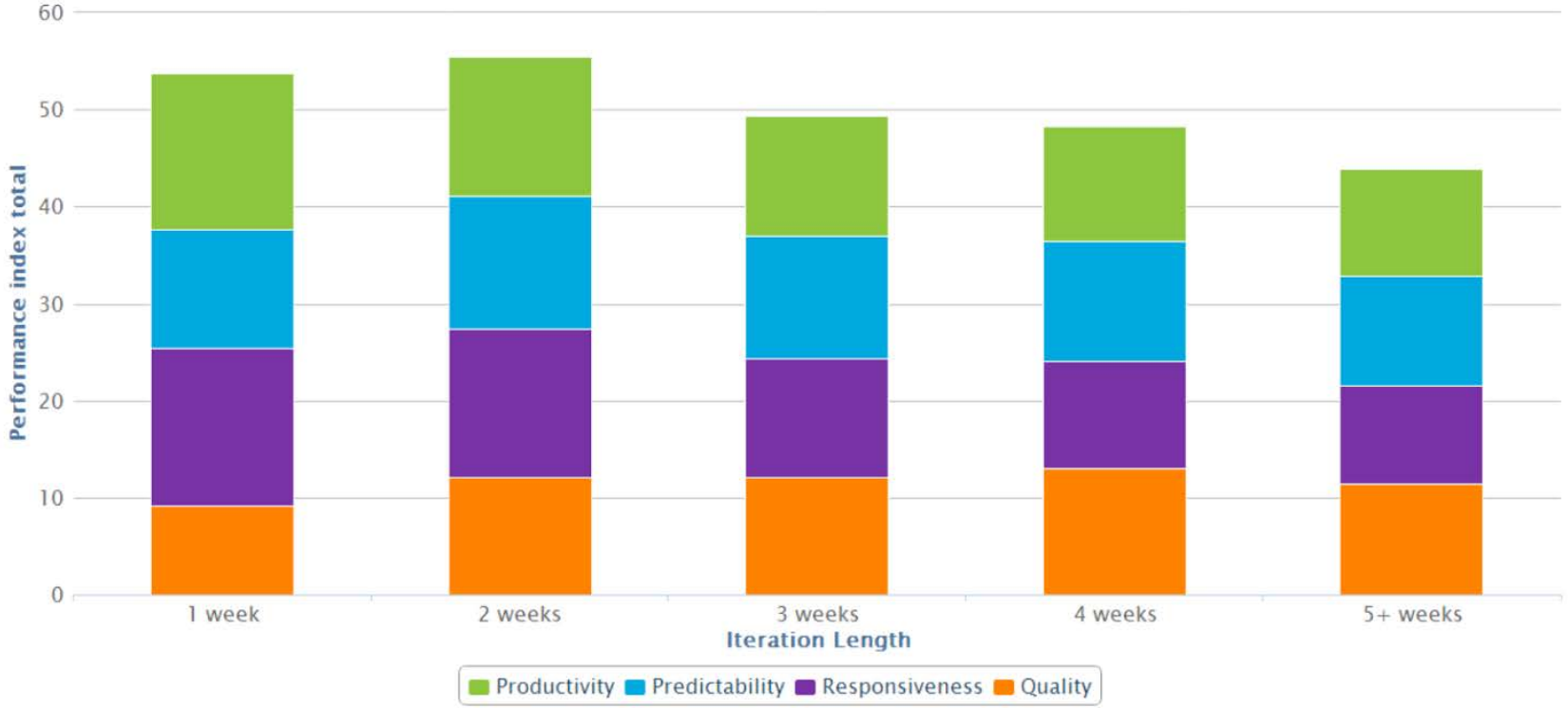


The investigation continues with ...

Iteration length

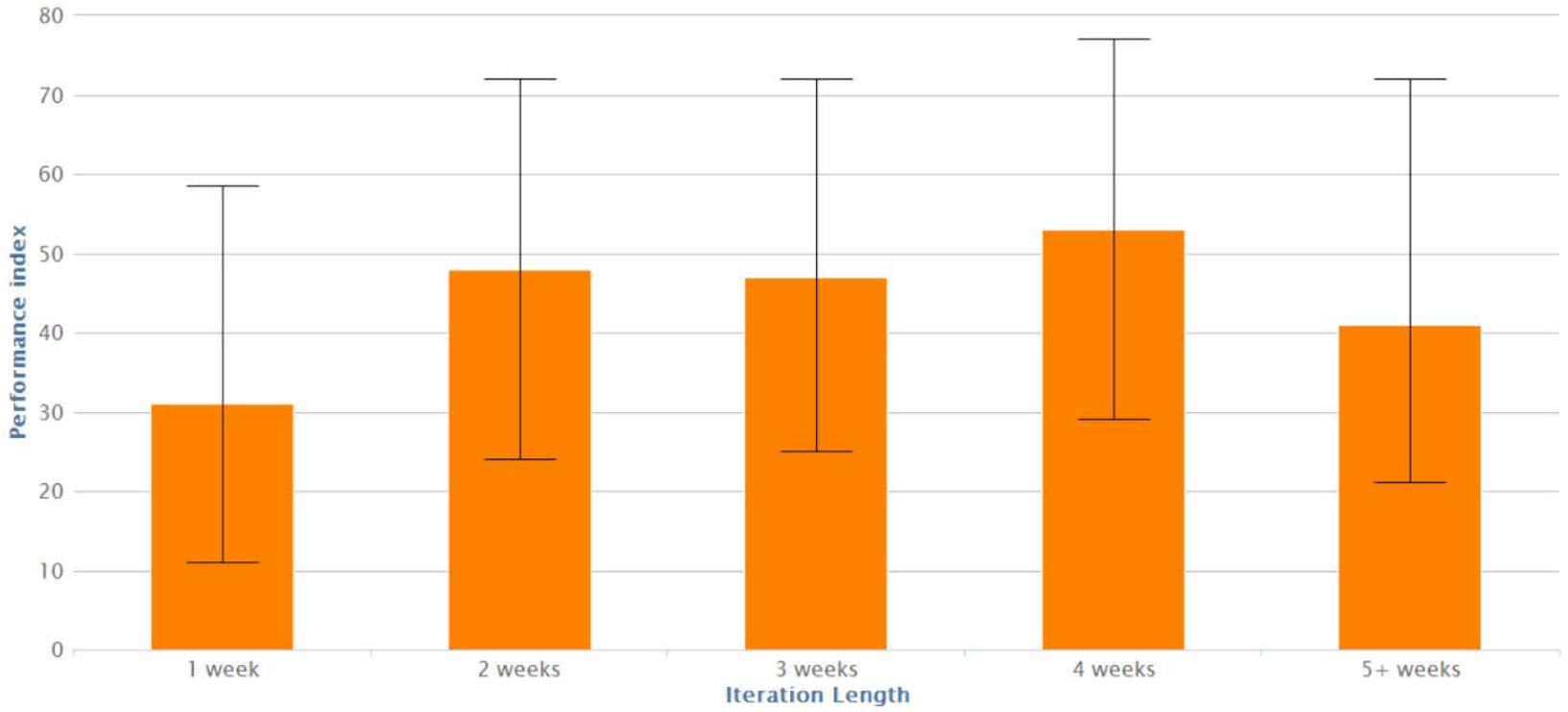
PERFORMANCE

Iteration Length relationship to Performance



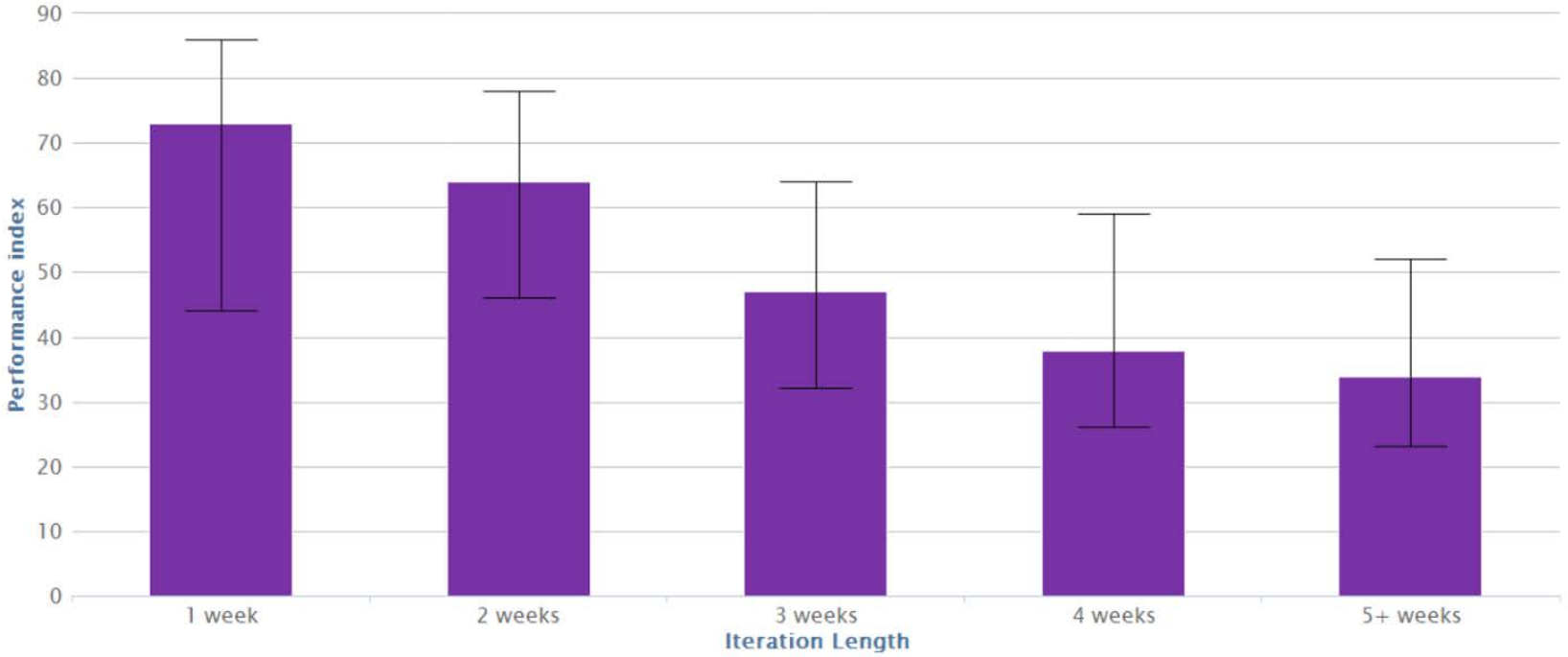
QUALITY

Iteration Length relationship to Performance



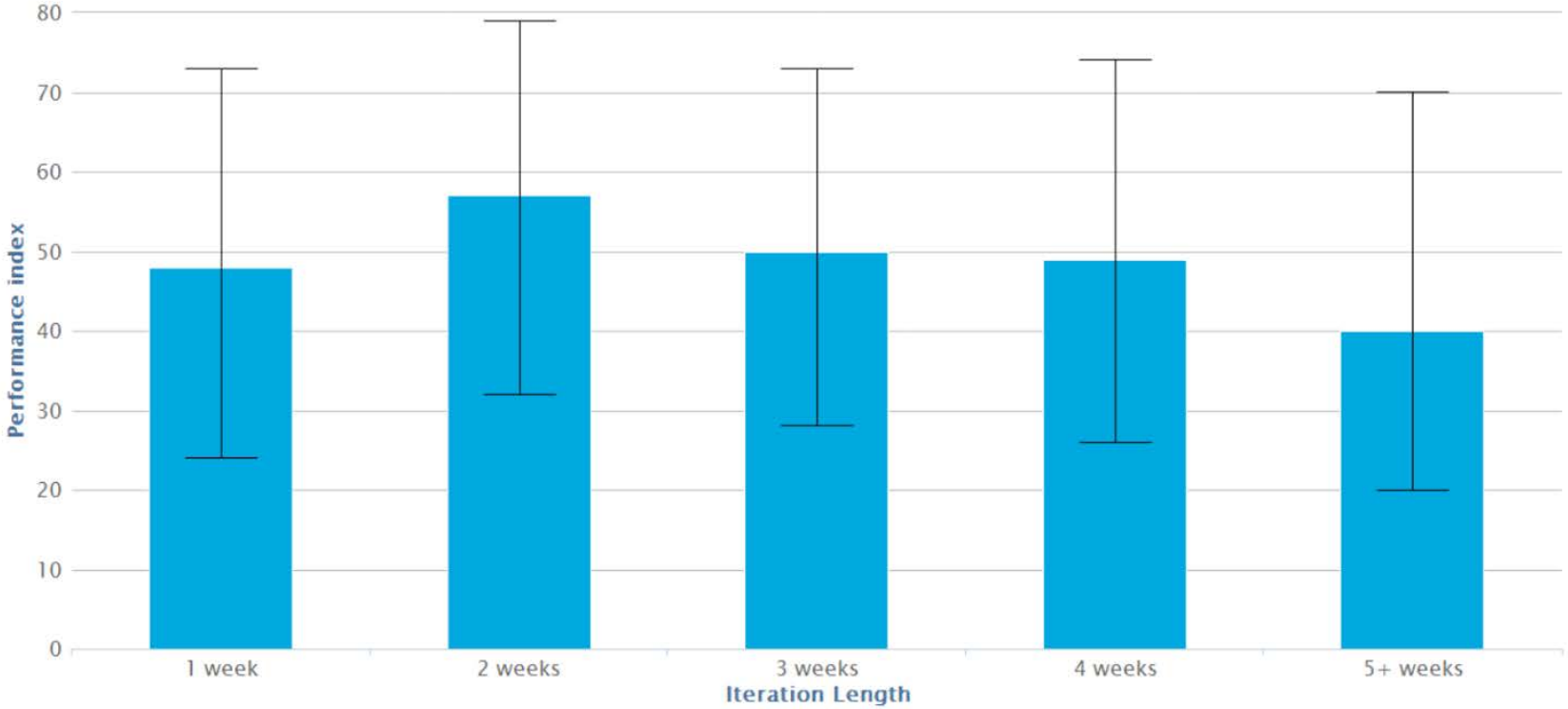
RESPONSIVENESS

Iteration Length relationship to Performance

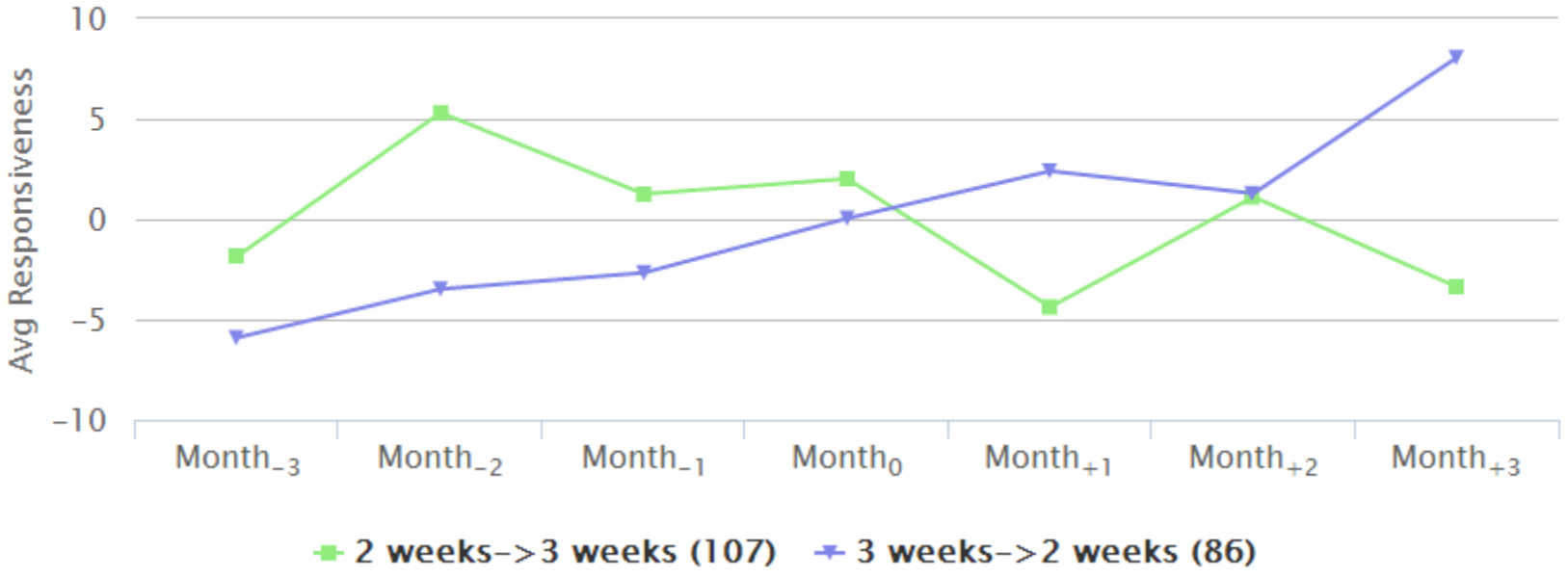


PREDICTABILITY

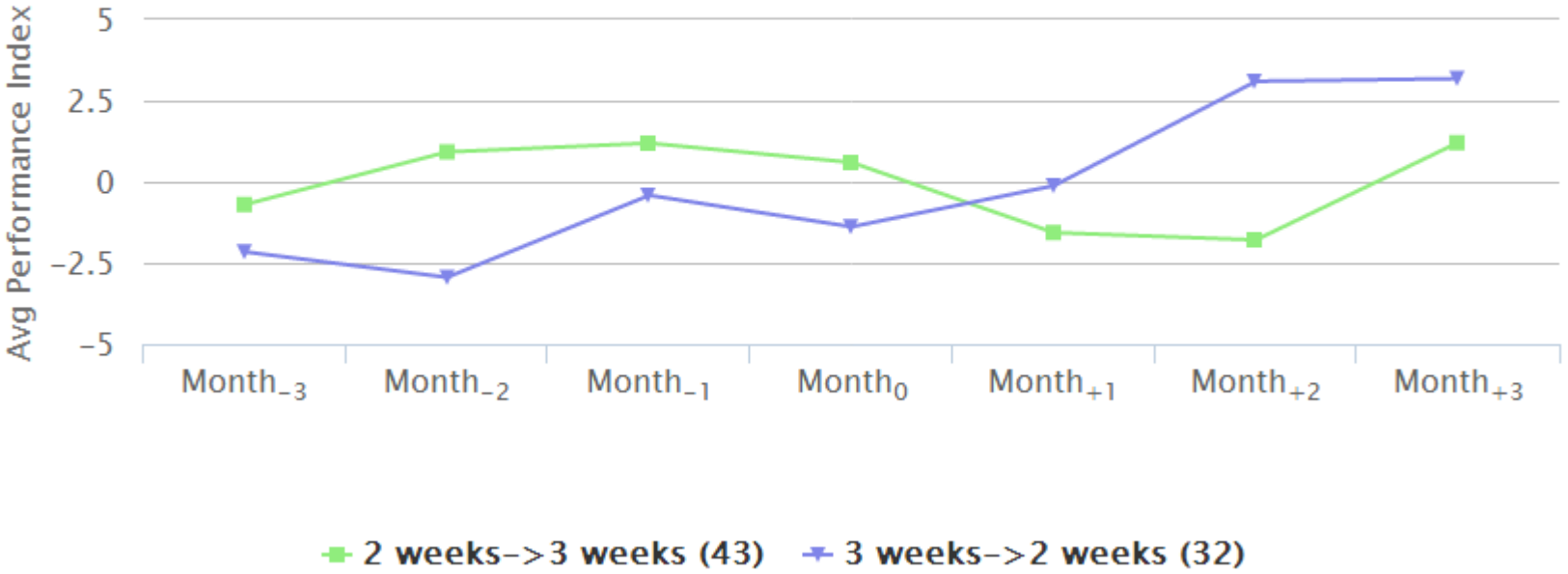
Iteration Length relationship to Performance



Iteration Length Transition vs Avg Responsiveness

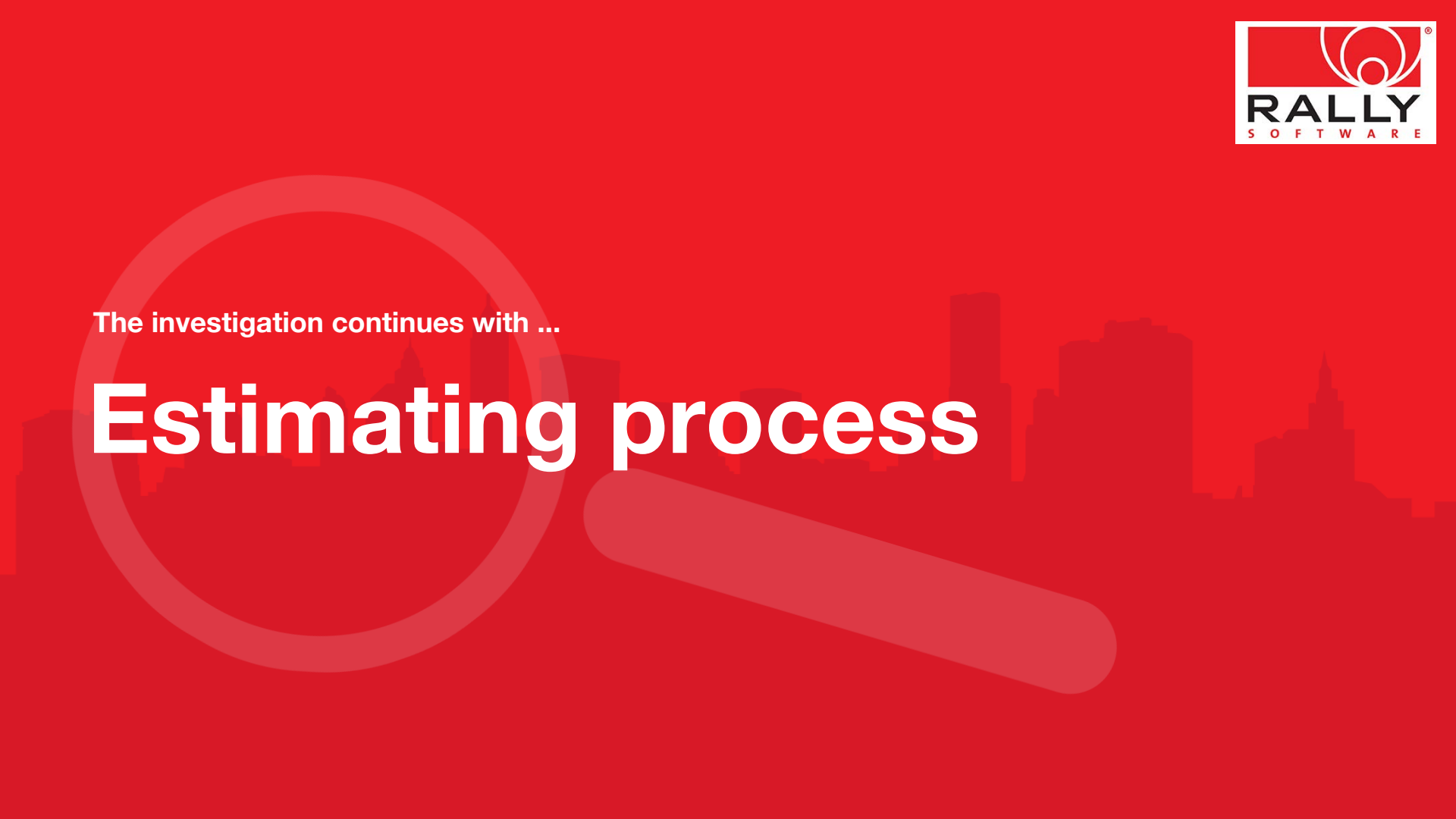


Iteration Length Transition vs Avg Performance Index



Facts Discovered:

- Teams using two-week iterations have the best balanced performance
- Longer iterations correlate with higher Quality
- Shorter iterations correlate with higher Productivity and Responsiveness
- However, some teams are acting like “tough guys” by pretending to operate at one-week iterations when they can’t back it up



The investigation continues with ...

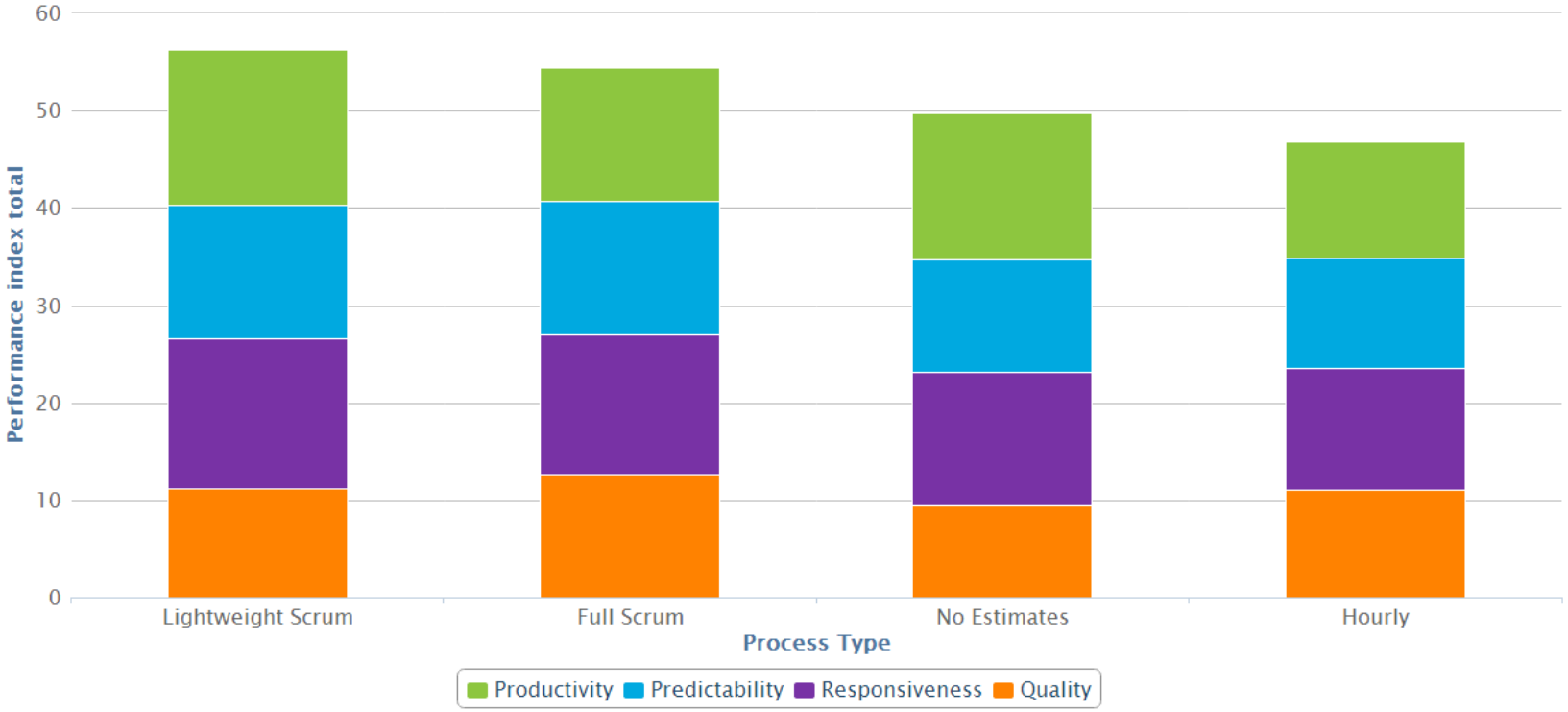
Estimating process

Estimating process

Process Type	Teams Using
No Estimates	3%
Full Scrum	79%
Lightweight Scrum	10%
Hourly-Oriented	8%

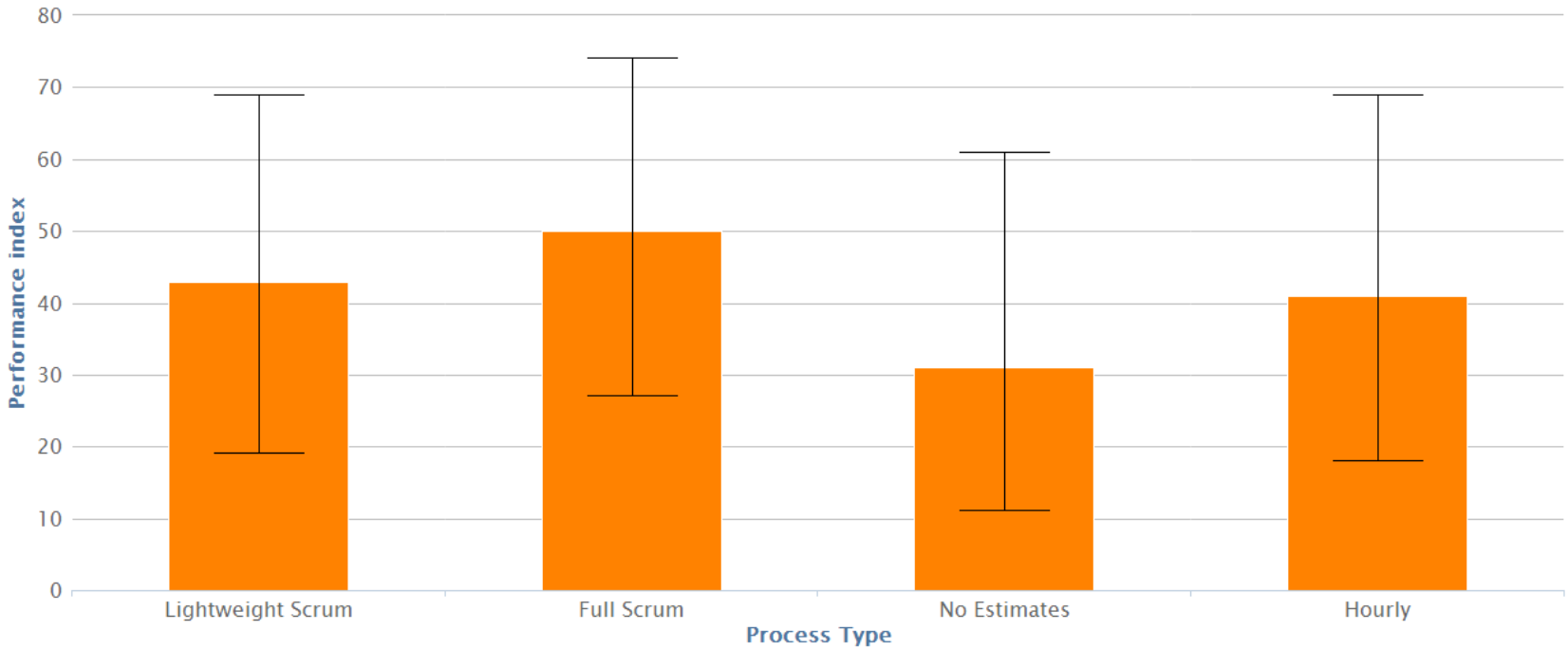
PERFORMANCE

Process Type relationship to Performance



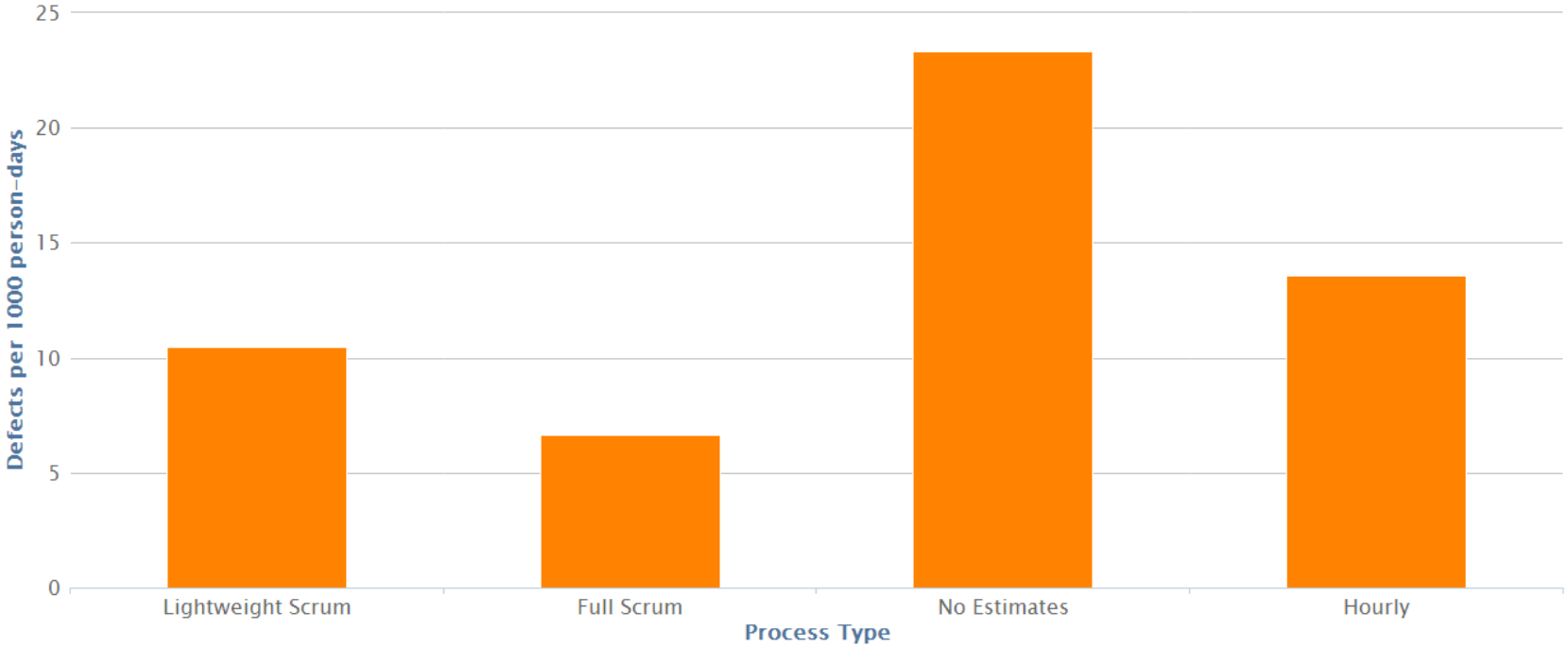
QUALITY

Process Type relationship to Performance



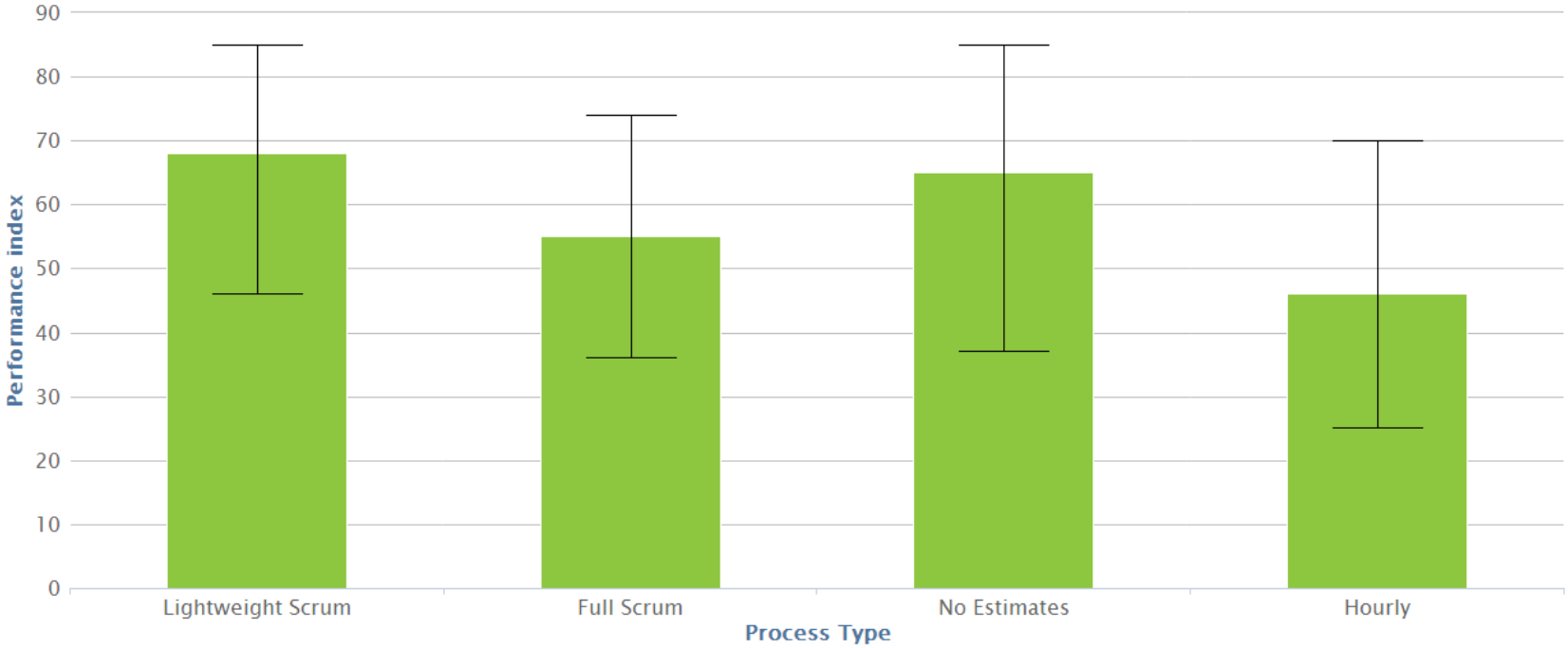
RELEASED DEFECT DENSITY (AVERAGE)

Lower is better



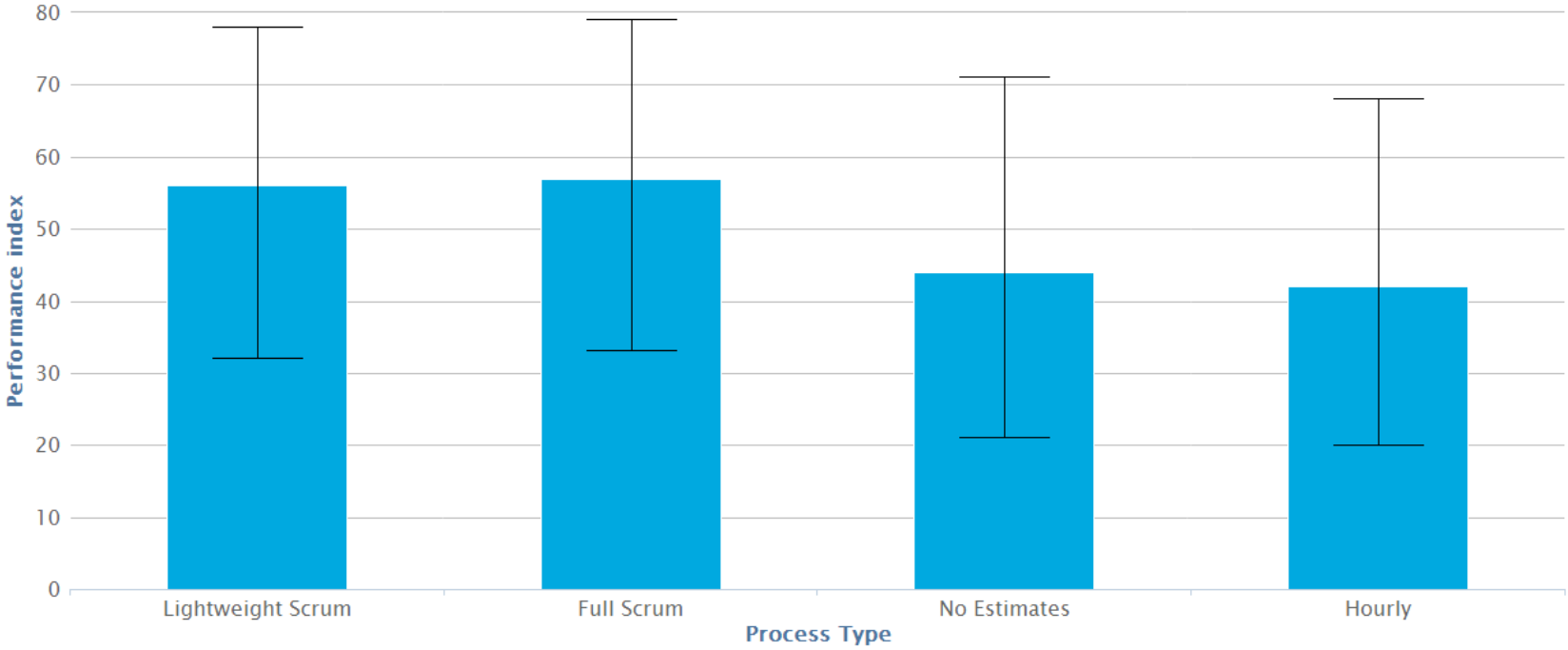
PRODUCTIVITY

Process Type relationship to Performance



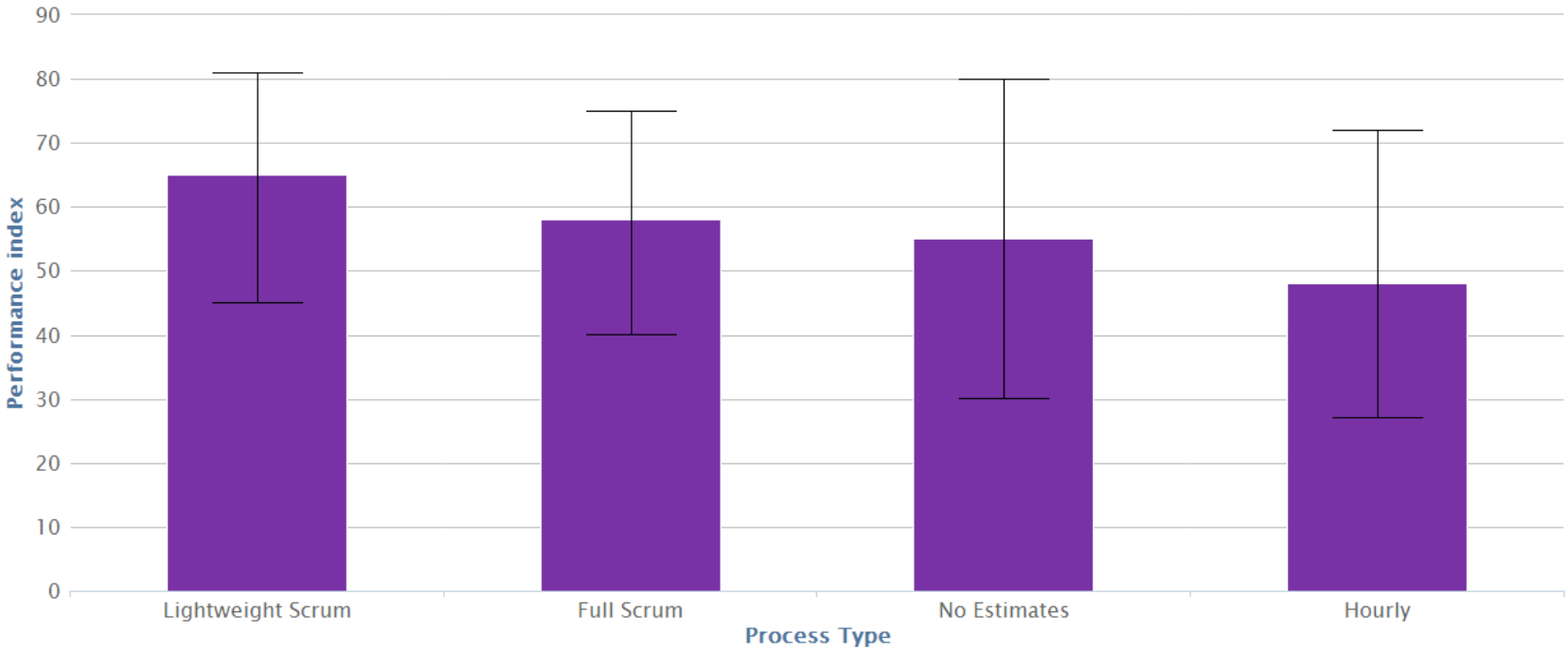
PREDICTABILITY

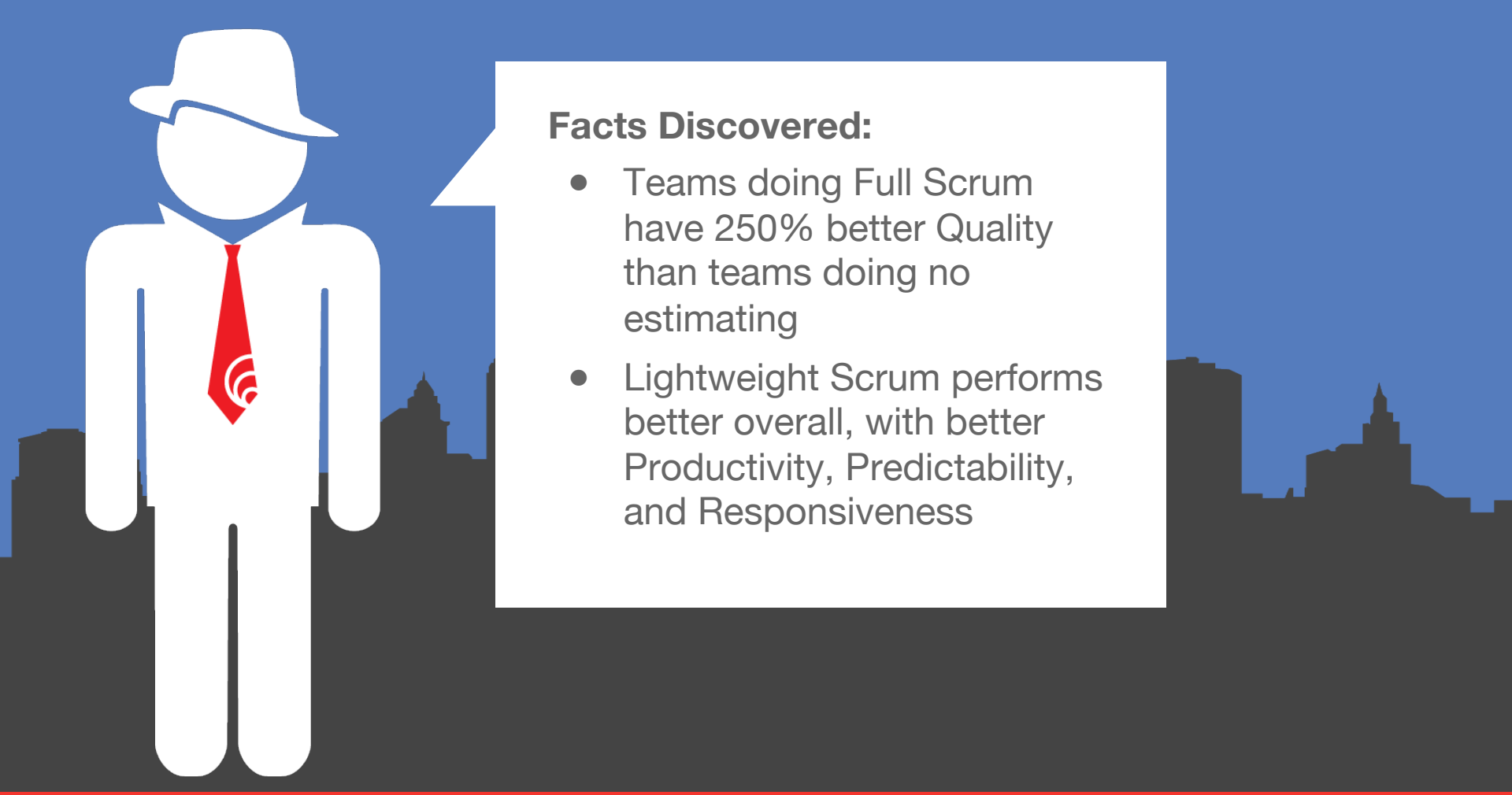
Process Type relationship to Performance

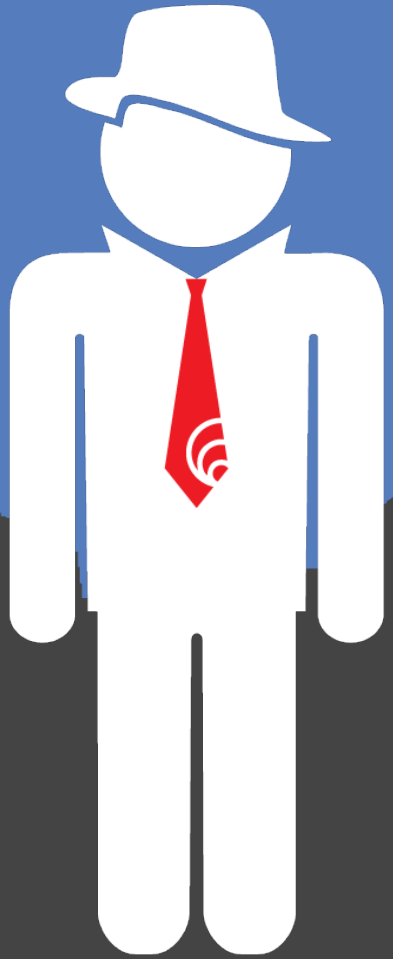


RESPONSIVENESS

Process Type relationship to Performance







Recommendations:

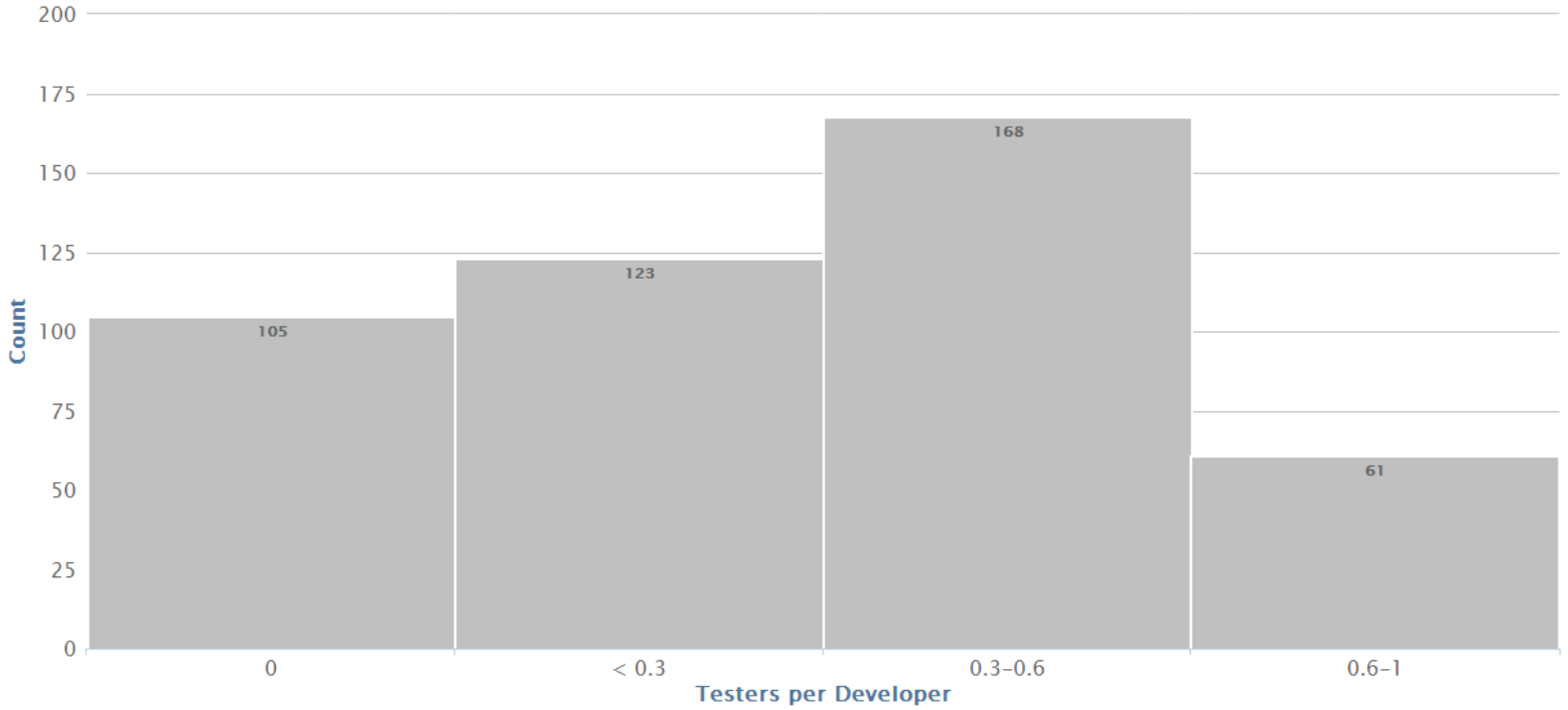
- Experienced teams may get best results from Lightweight Scrum
- If new to Agile or focused strongly on Quality, choose Full Scrum

The investigation continues with ...

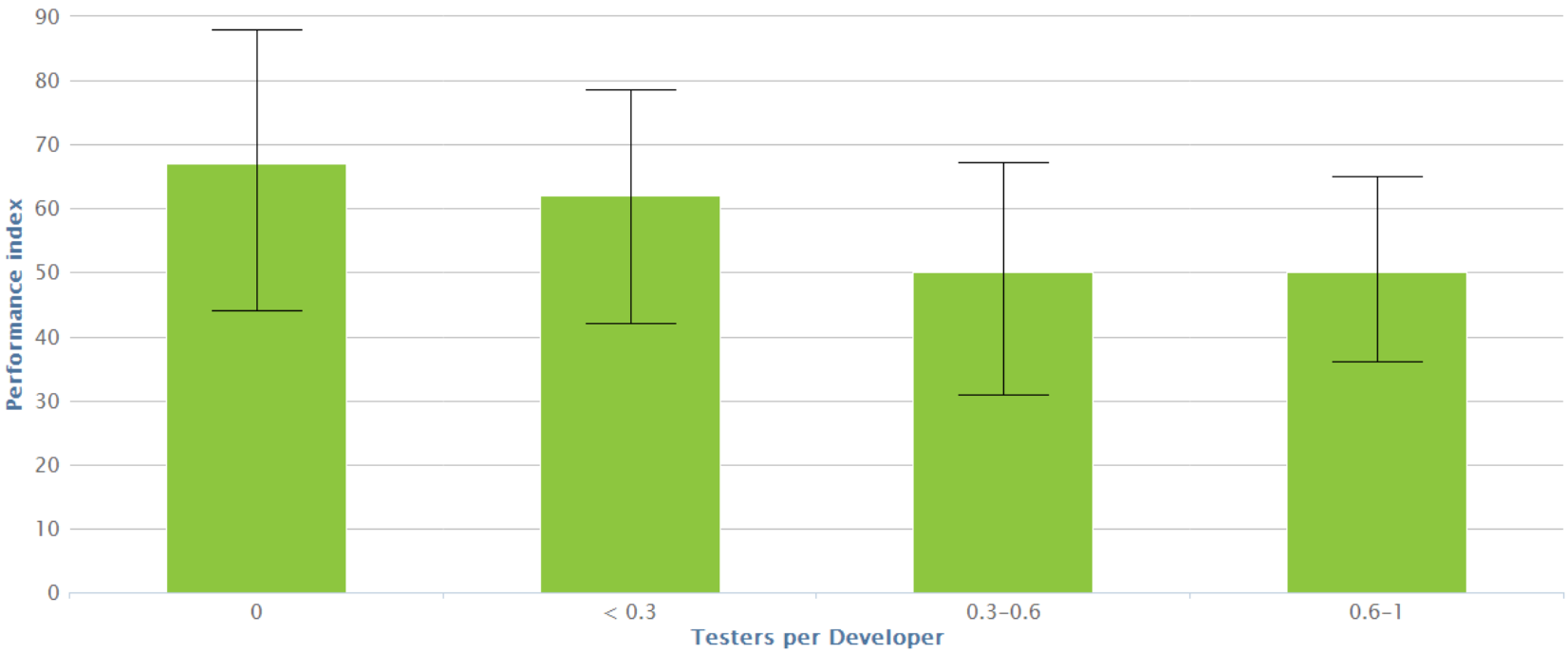
Ratio of testers to developers

TESTERS PER DEVELOPER

Counts of measurements in each bucket



PRODUCTIVITY

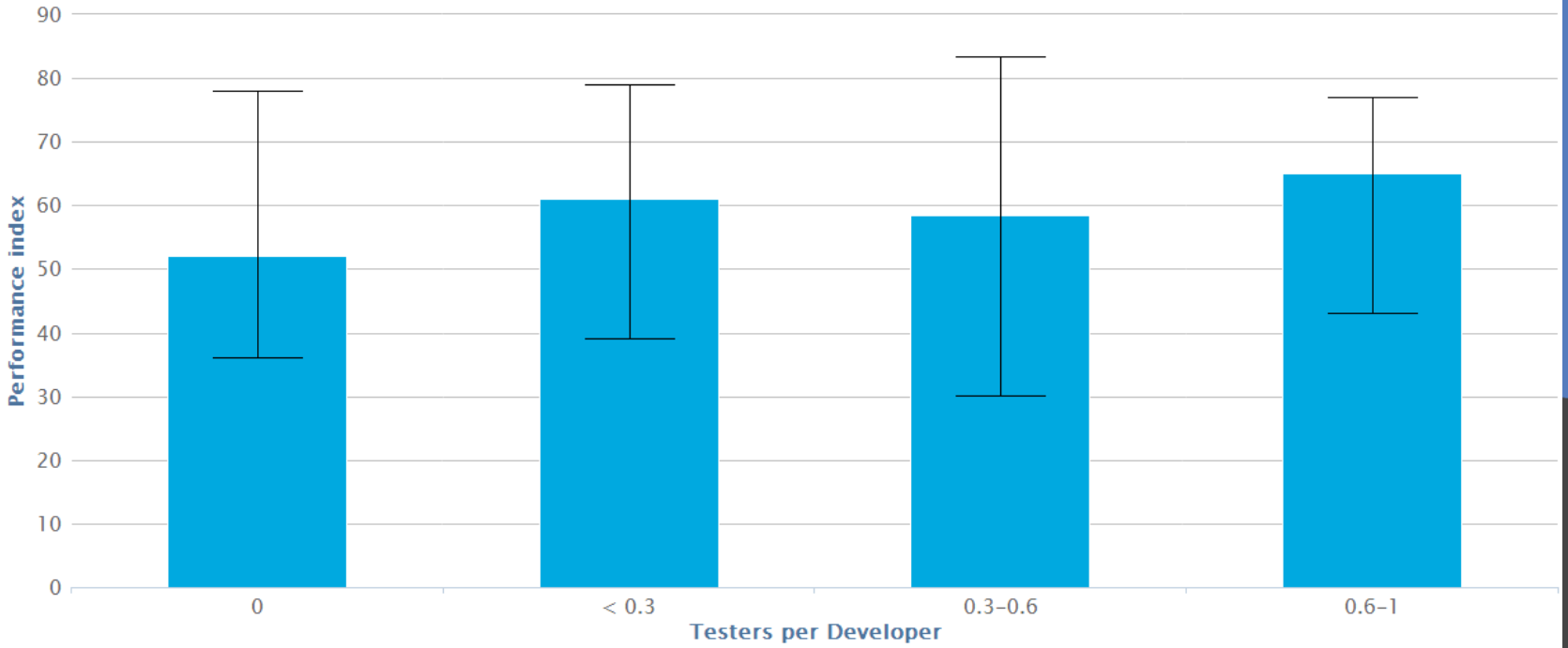


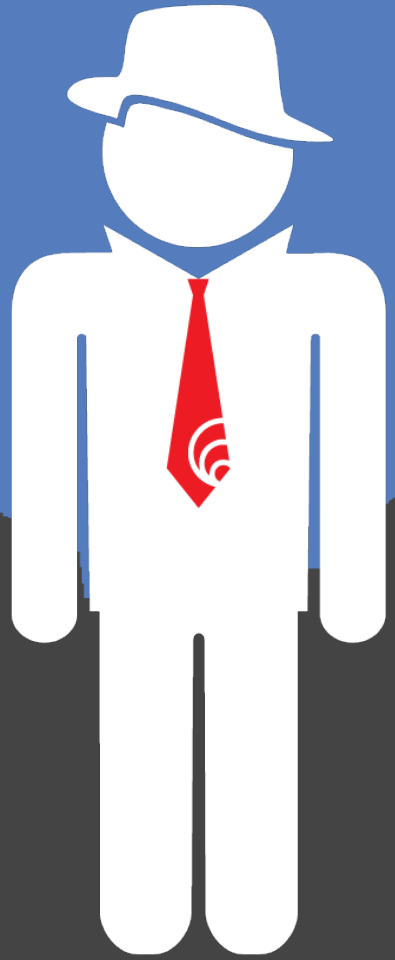
QUALITY

RESPONSIVENESS

PREDICTABILITY

Testers per Developer relationship to Performance





Facts Discovered:

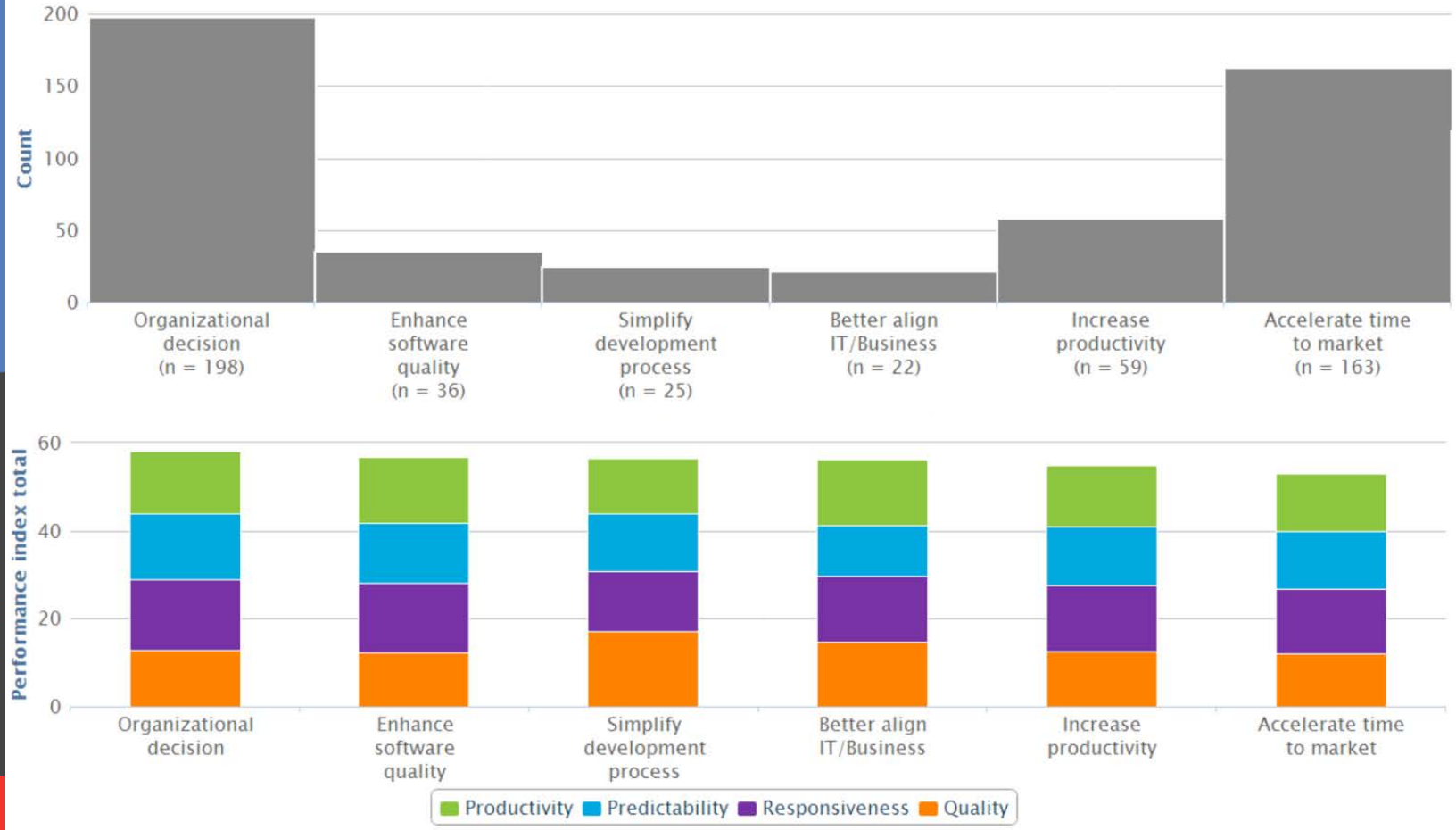
- More testers lead to better Quality
- But they also generally lead to worse Productivity and Responsiveness
- Interestingly, teams that self-identify as having no testers have:
  - the best Productivity
  - almost as good Quality
  - but much wider variation in Quality

The investigation continues with ...

Motive

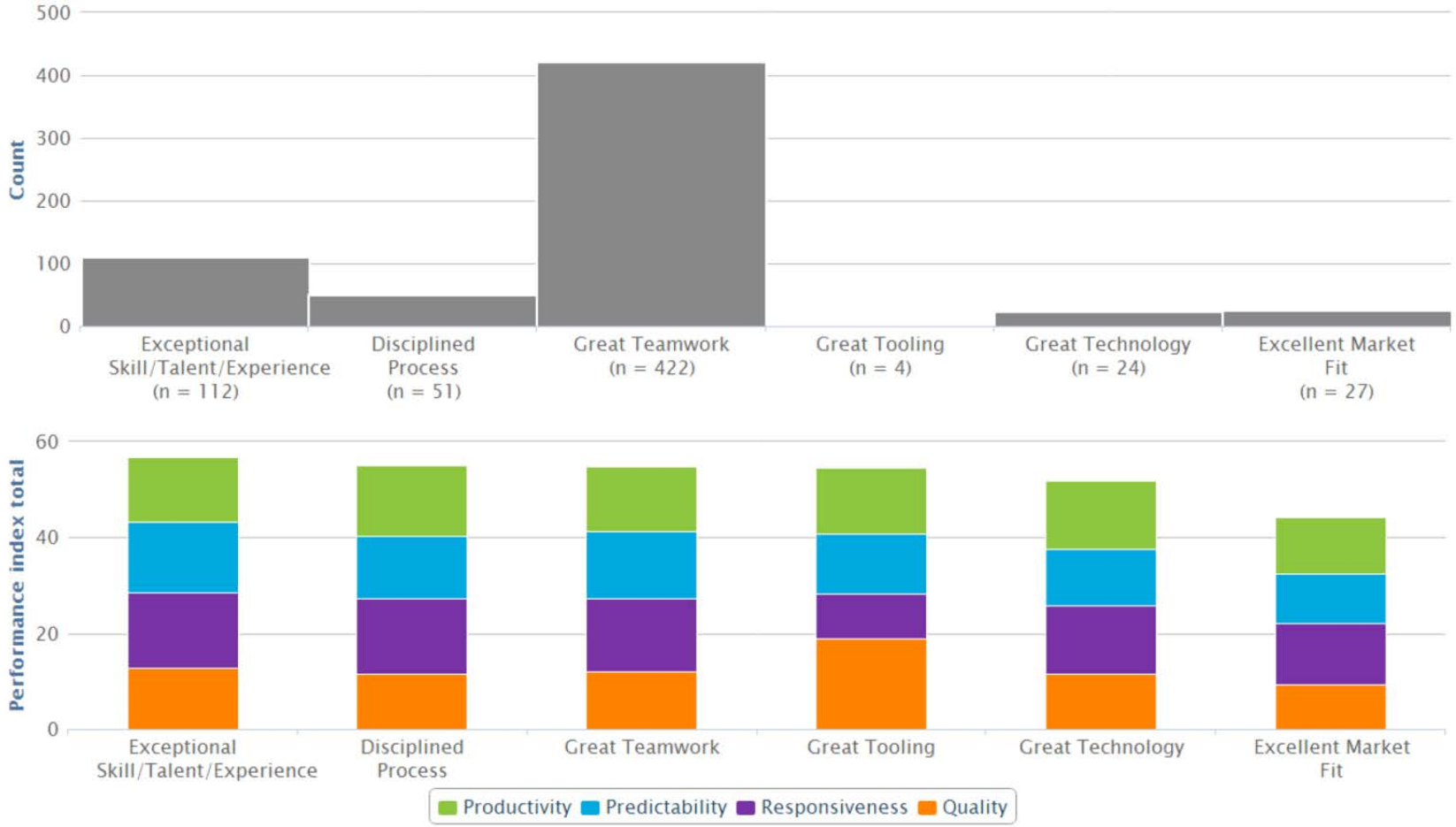
REASON FOR ADOPTING AGILE

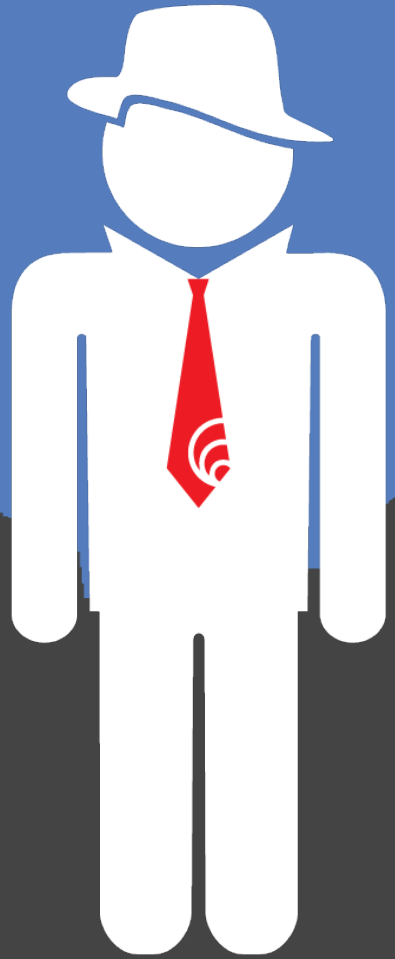
Counts of measurements in each bucket



REASON FOR SUCCESS

Counts of measurements in each bucket





Evidence Found:

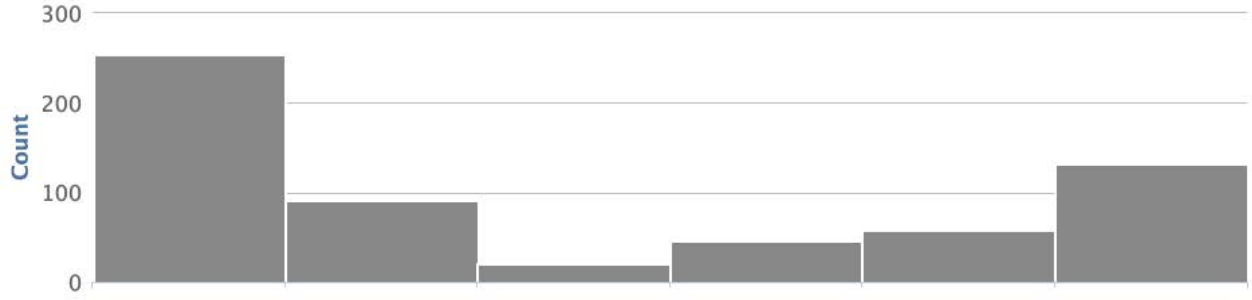
- Motive has a small but statistically significant impact on performance
- Extrinsic motivation does **not** have a negative impact on performance
- Executive support is critical for success with Agile.
- Teamwork is not the dominant factor; talent, skills, and experience are
- Those motivated by quality perform best

The investigation continues with ...

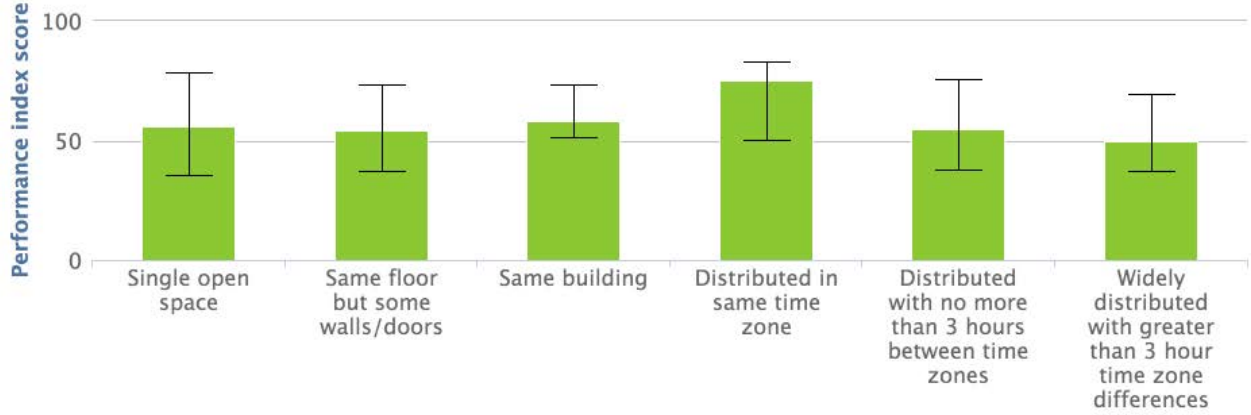
Co-location

PHYSICAL SPACE DISTRIBUTION

Counts of measurements in each bucket



Productivity

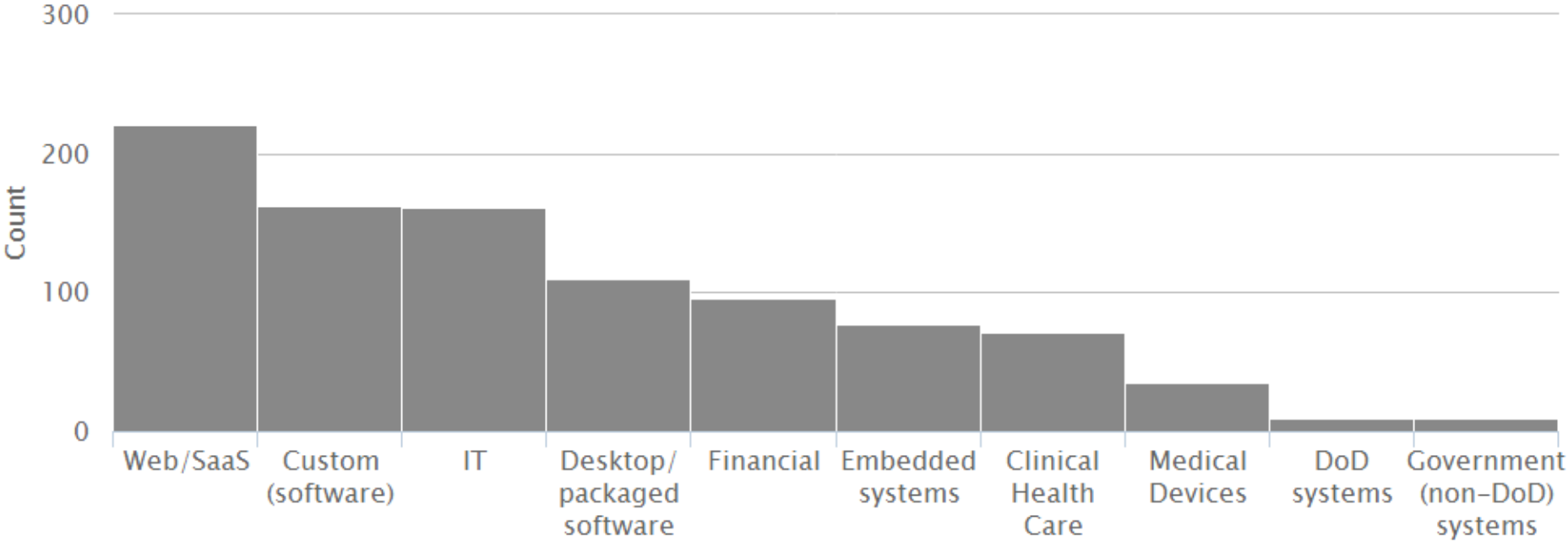


The investigation continues with ...

Survey-based research

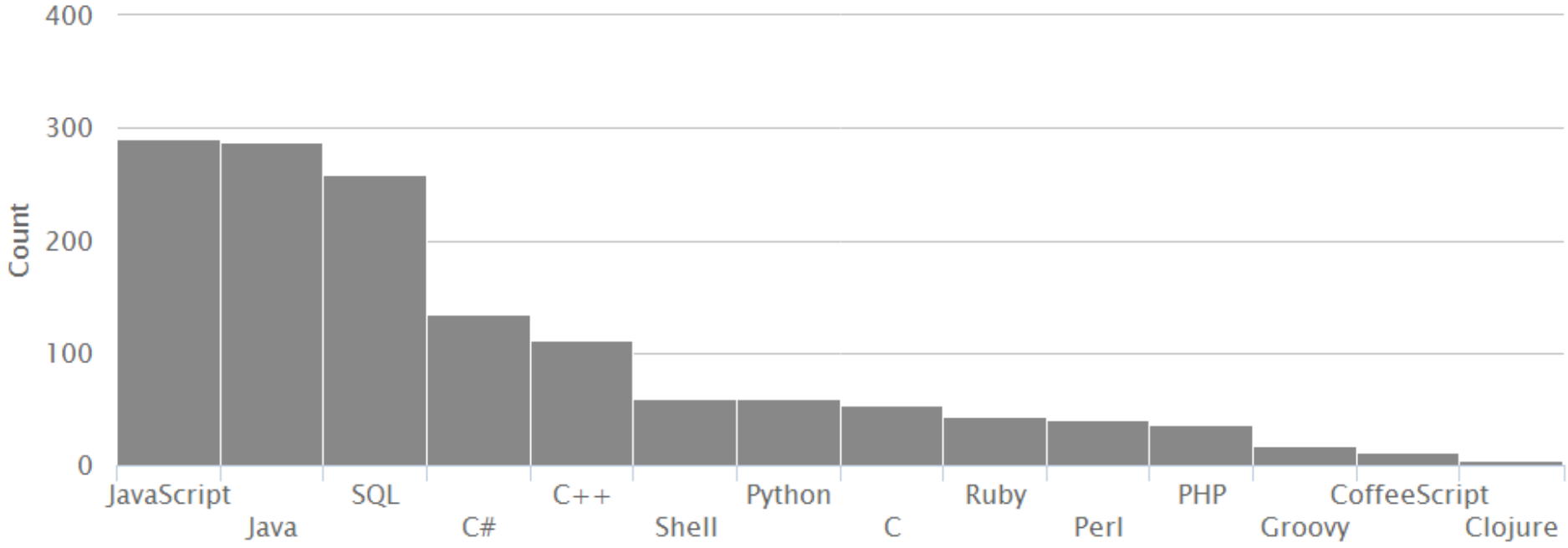
INDUSTRY

(by Count)



PROGRAMMING LANGUAGE

(by Count)



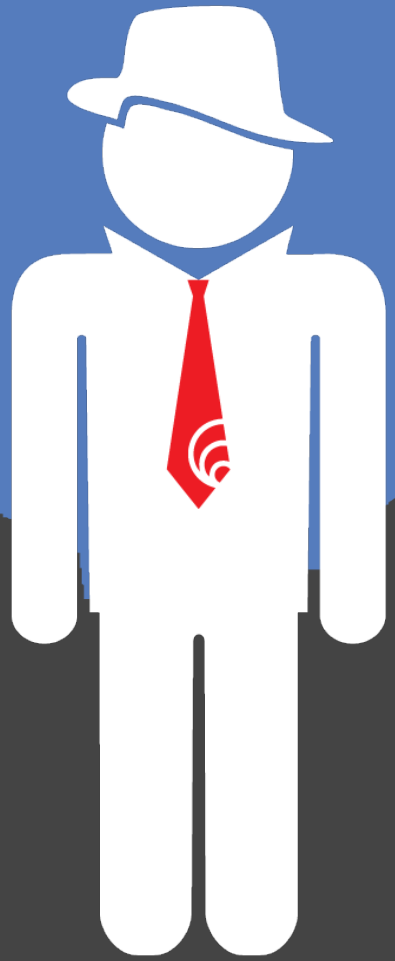
Where in the world?

Where in the U.S.?

Where in the Europe?

One year earlier ...

rallydev.com/agilemetrics



Teams with low WiP have up to:

- 4x better Quality
- 2x faster time-to-market
- But 34% worse productivity

Stable teams result in up to:

- 60% better Productivity
- 40% better Predictability
- 60% better Responsiveness

Dedicated teams: Teams made up of people who only work on that one team have double the Productivity

Smaller teams have better Productivity

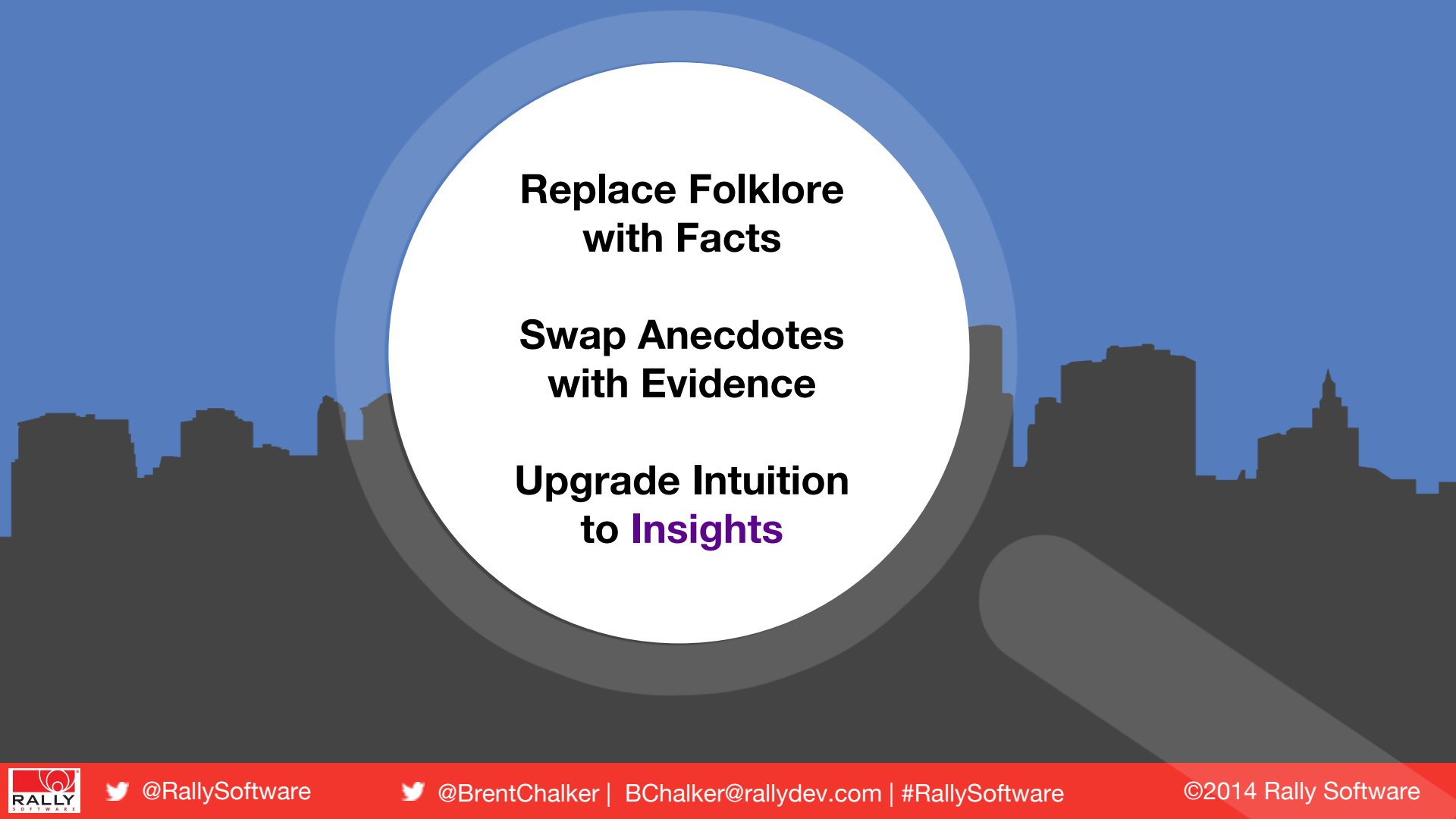
Larger teams have better Quality

What’s next?

- **Demo of Rally Insights**
  - Implements SDPI
- **Roadmap**
  - Self-assessment and tracking (surveys)
    - Two more dimensions
      - Customer/Stakeholder Satisfaction
      - Employee Engagement/Satisfaction
    - Probe your environment with customized surveys (maturity, practices compliance, etc.)
  - Recommendation Engine
    - What are the top five things we should improve next?

A fact without a theory
is like a ship without a sail,
is like a boat without a
rudder,
is like a kite without a tail.
A fact without a figure
is a tragic final act.
But one thing worse
in this universe
is a theory without a fact.

~ George Schultz



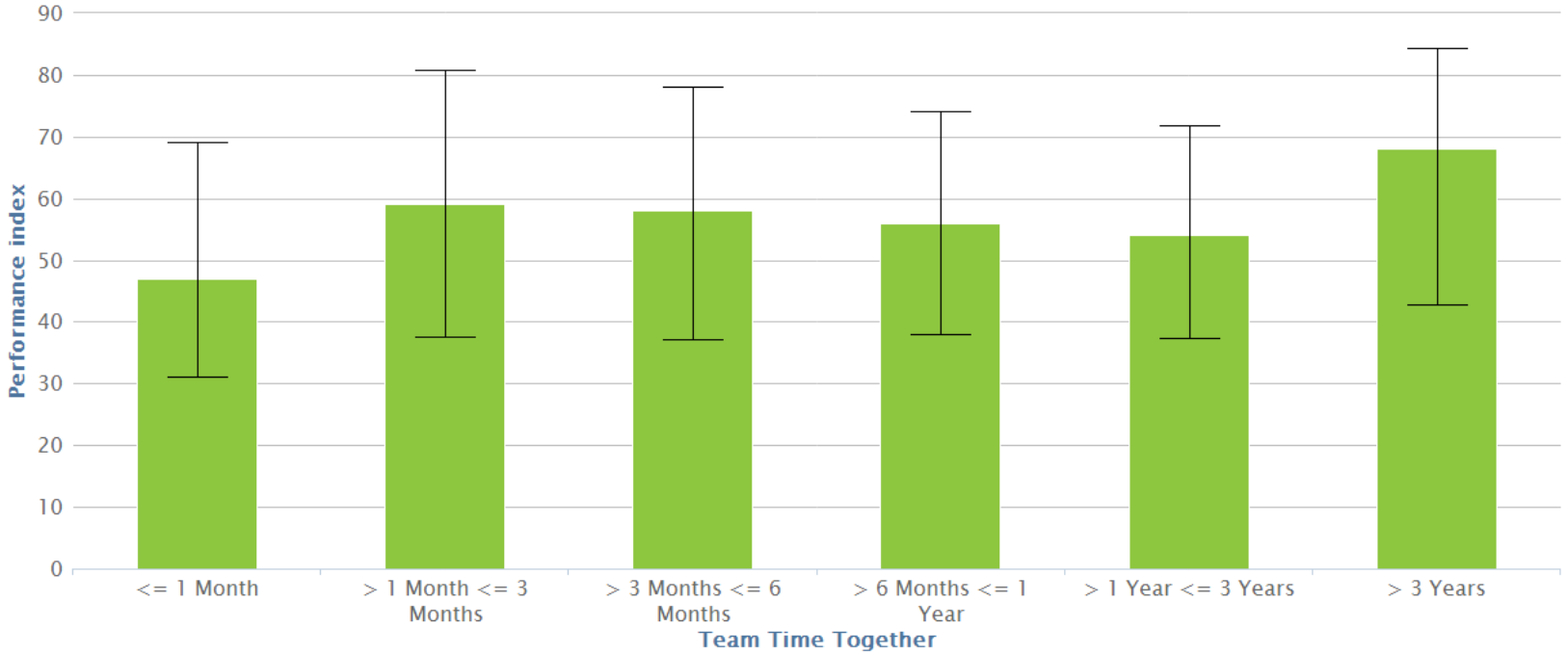
Hidden slides

The investigation continues with ...

Team time together

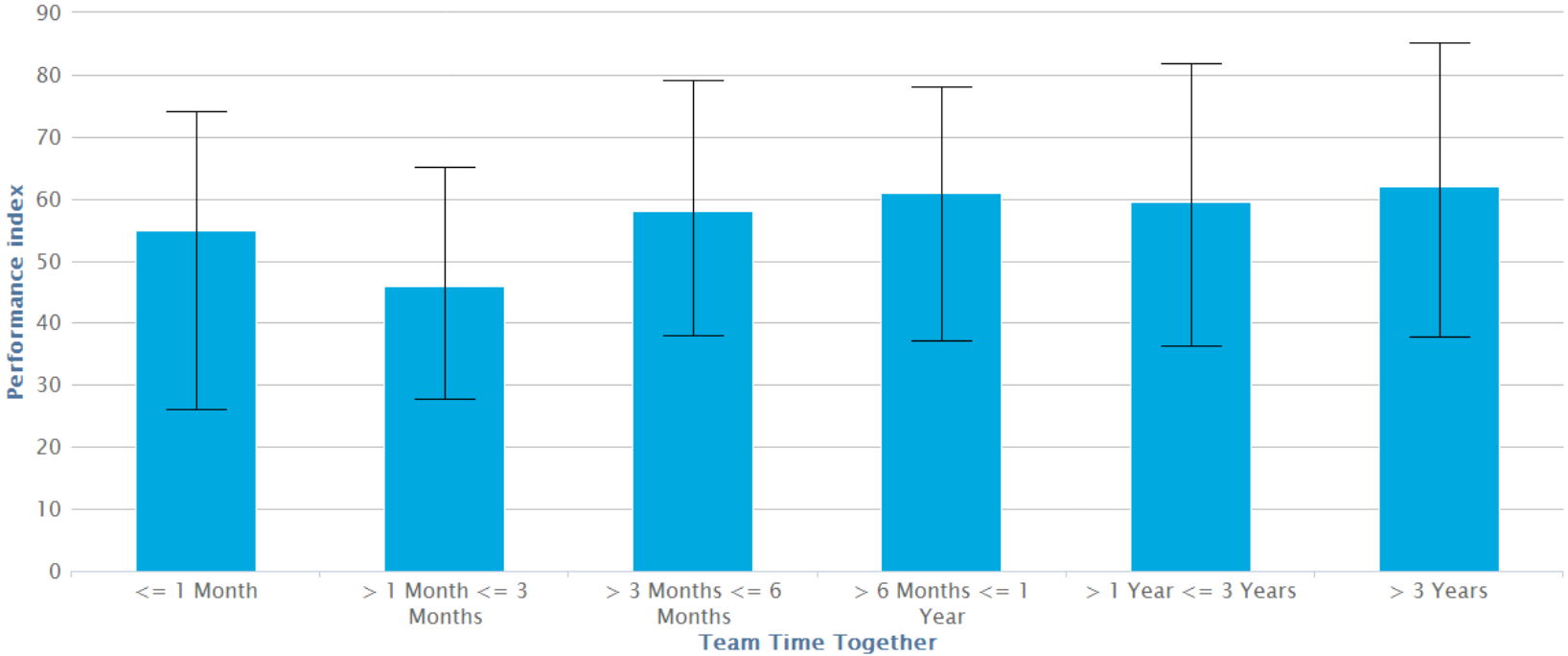
PRODUCTIVITY

Team Time Together relationship to Performance



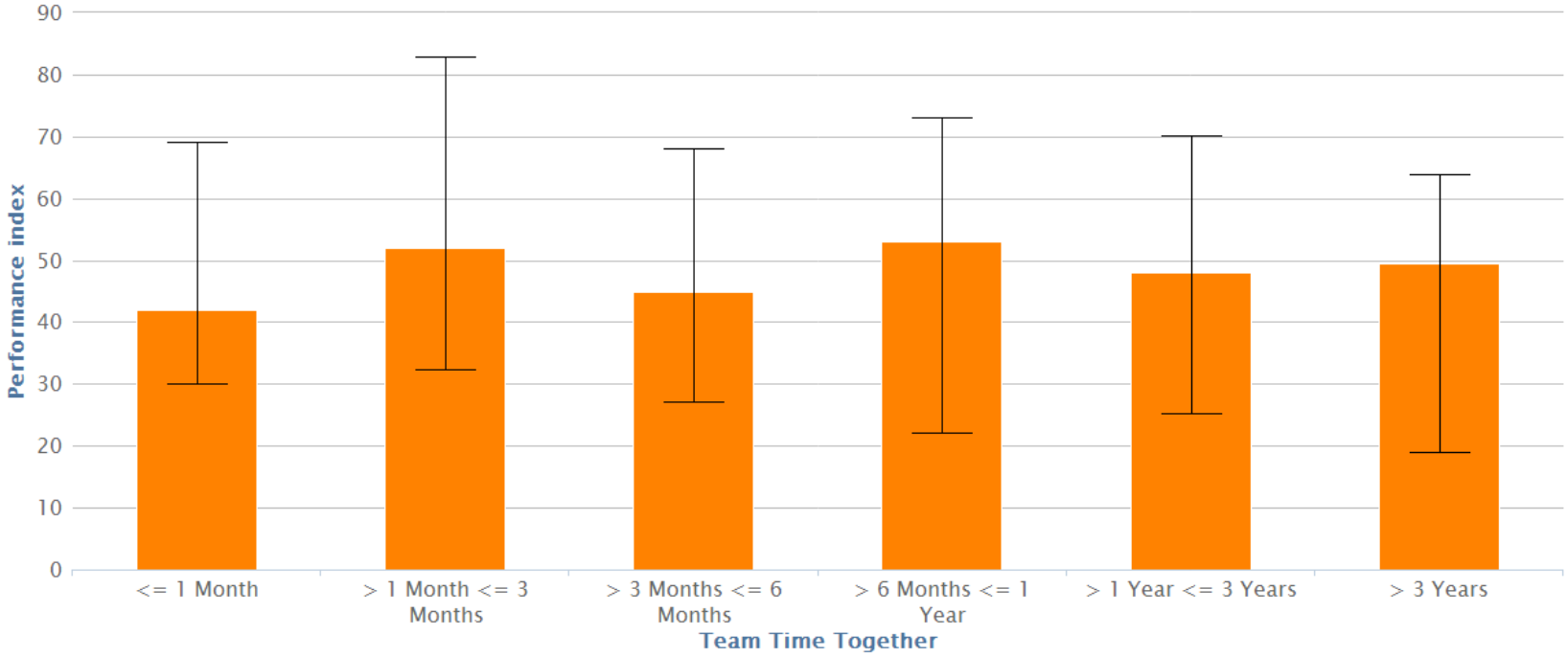
PREDICTABILITY

Team Time Together relationship to Performance



QUALITY

Team Time Together relationship to Performance

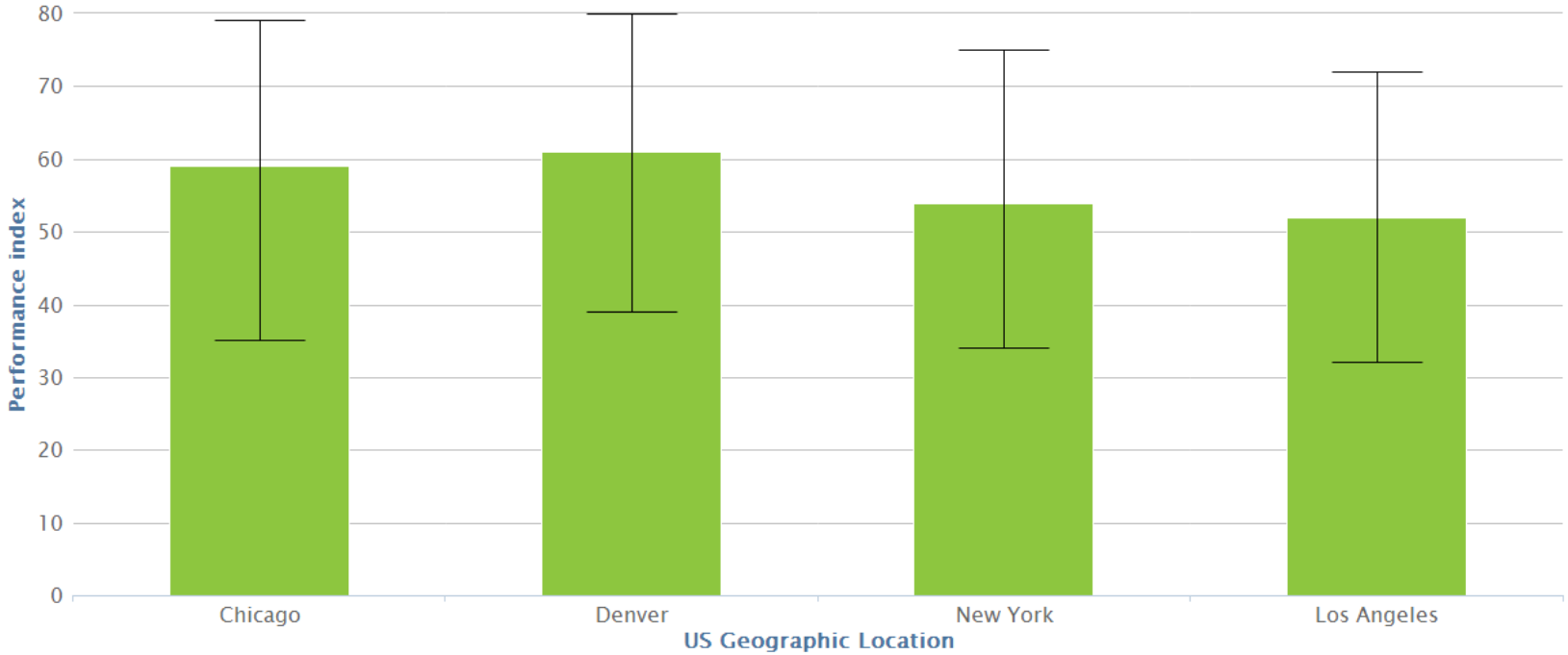


The investigation continues with ...

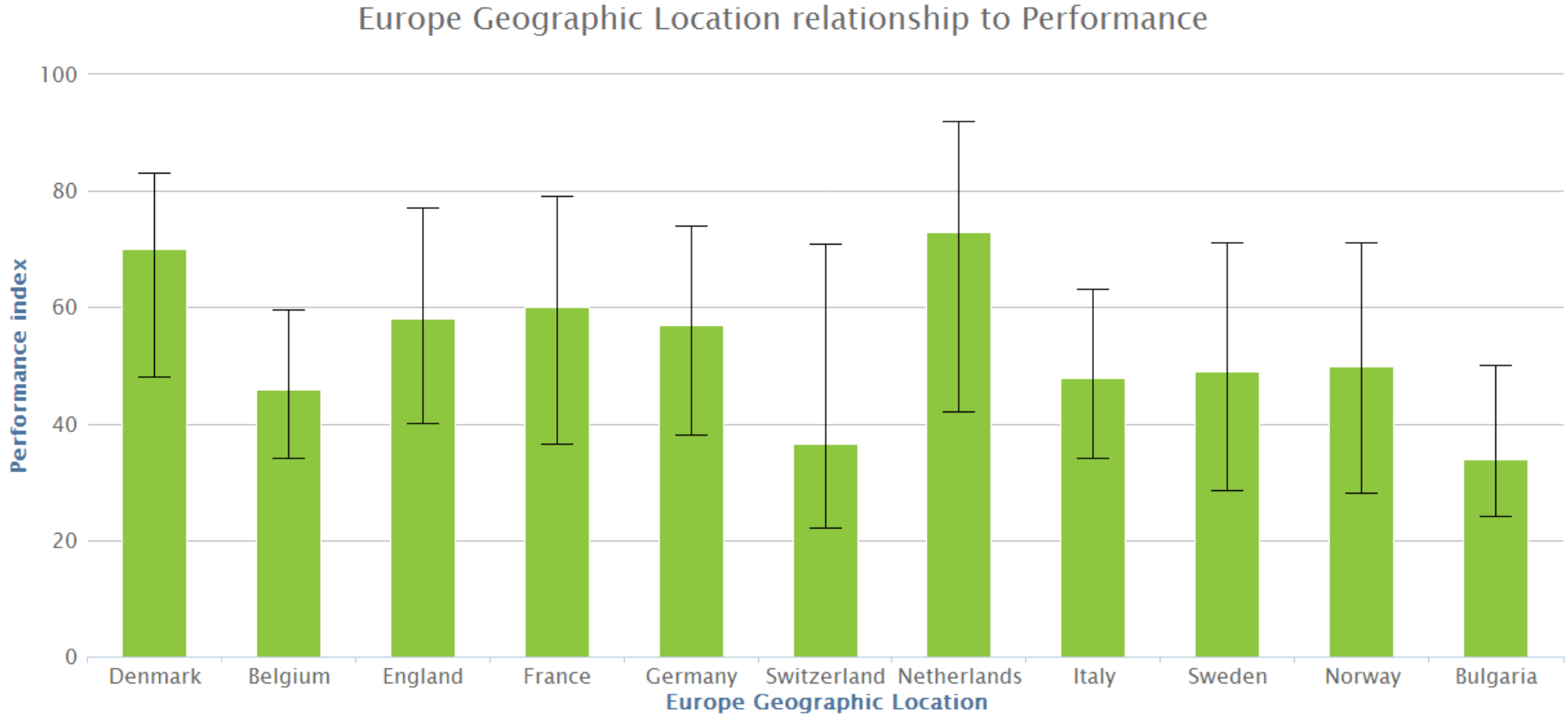
Location in US and Europe

PRODUCTIVITY

US Geographic Location relationship to Performance



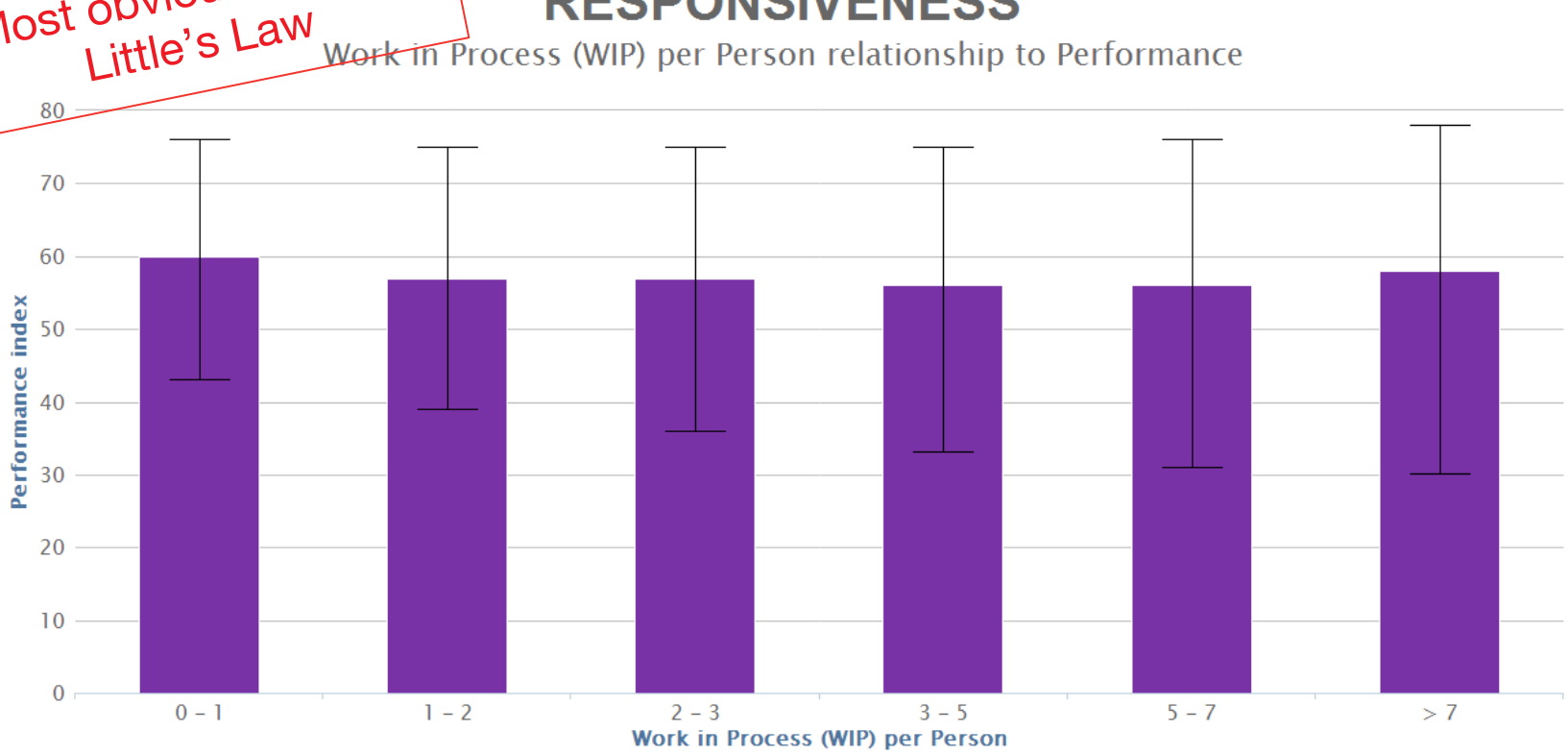
PRODUCTIVITY



The investigation continues with ...

Controlling WiP

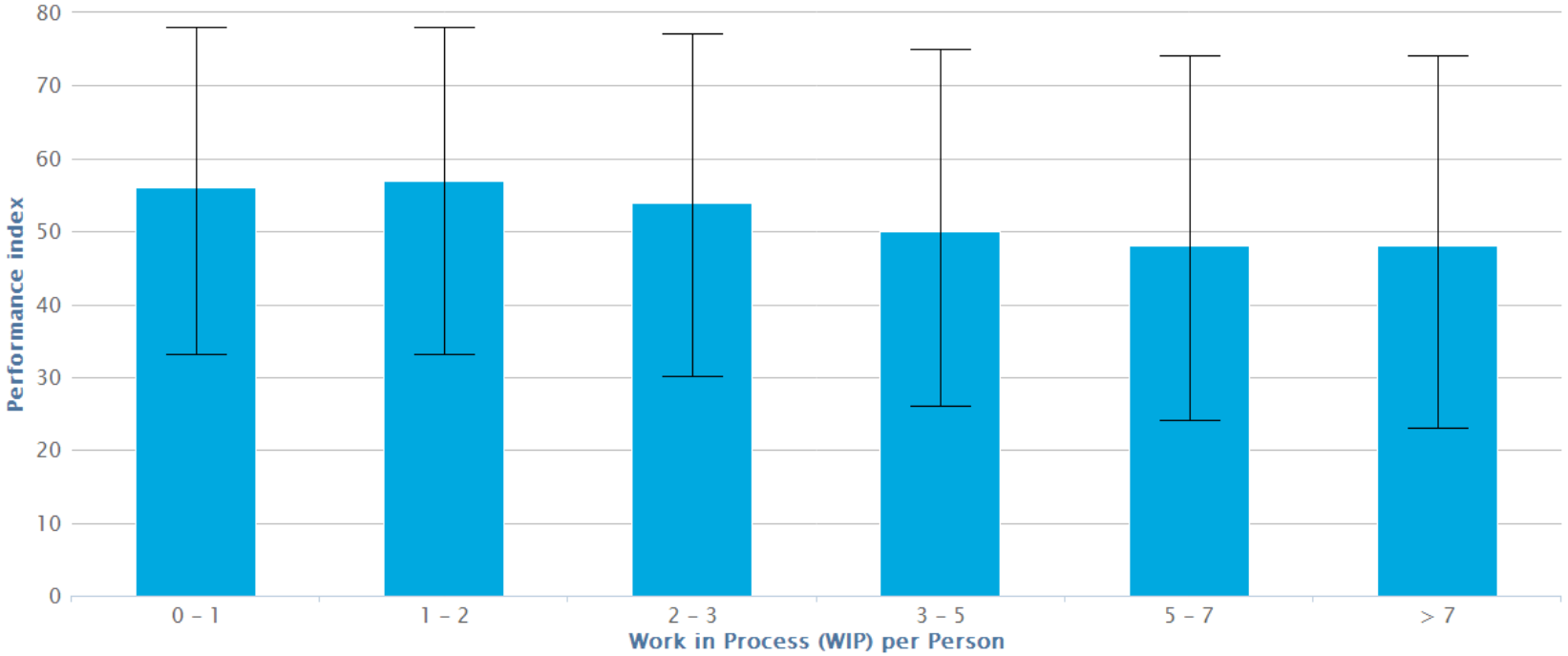
Most obvious finding:
Little's Law



Most dramatic finding

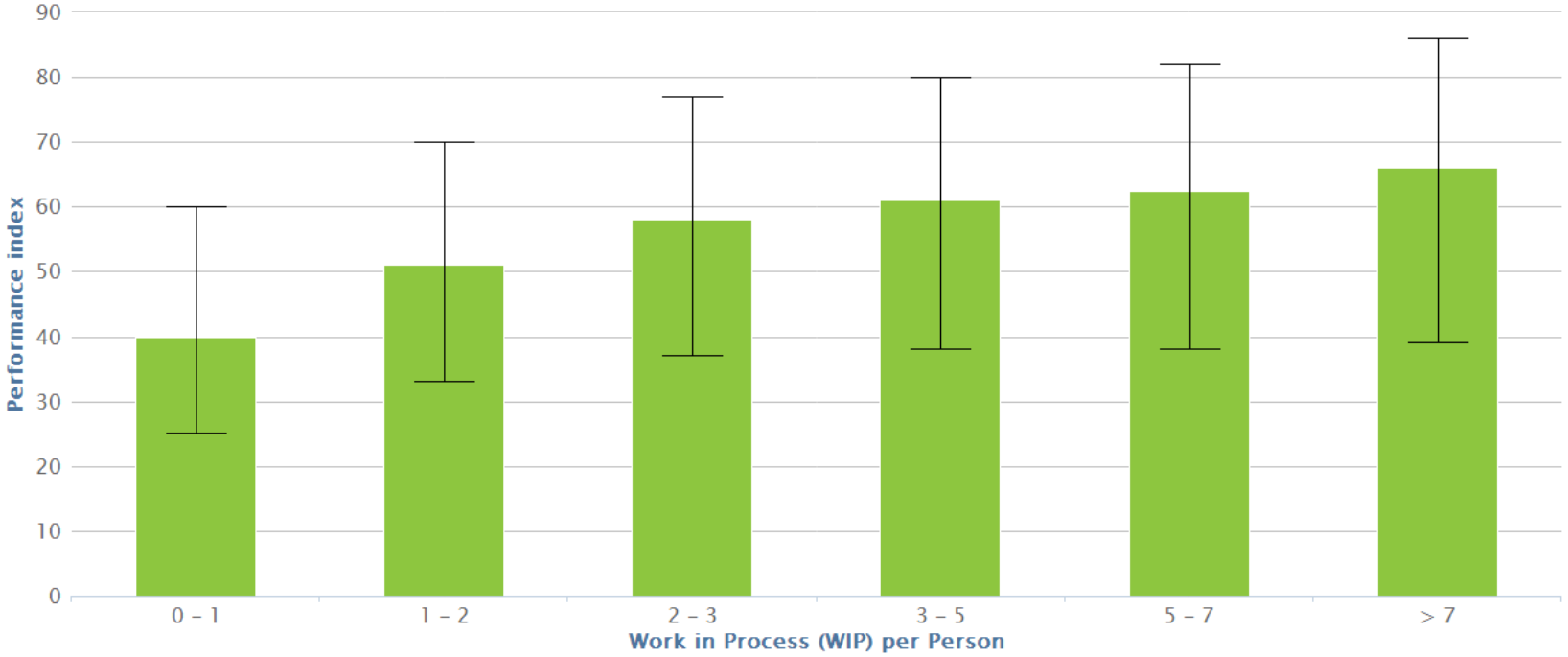
PREDICTABILITY

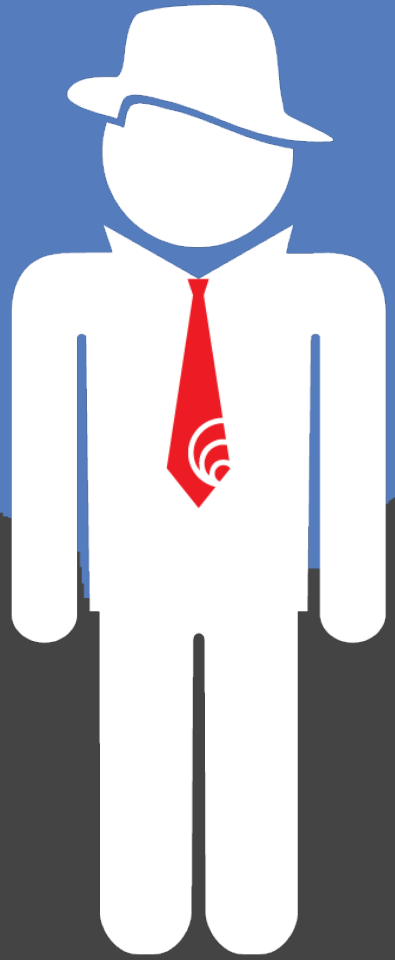
Work in Process (WIP) per Person relationship to Performance



PRODUCTIVITY

Work in Process (WIP) per Person relationship to Performance

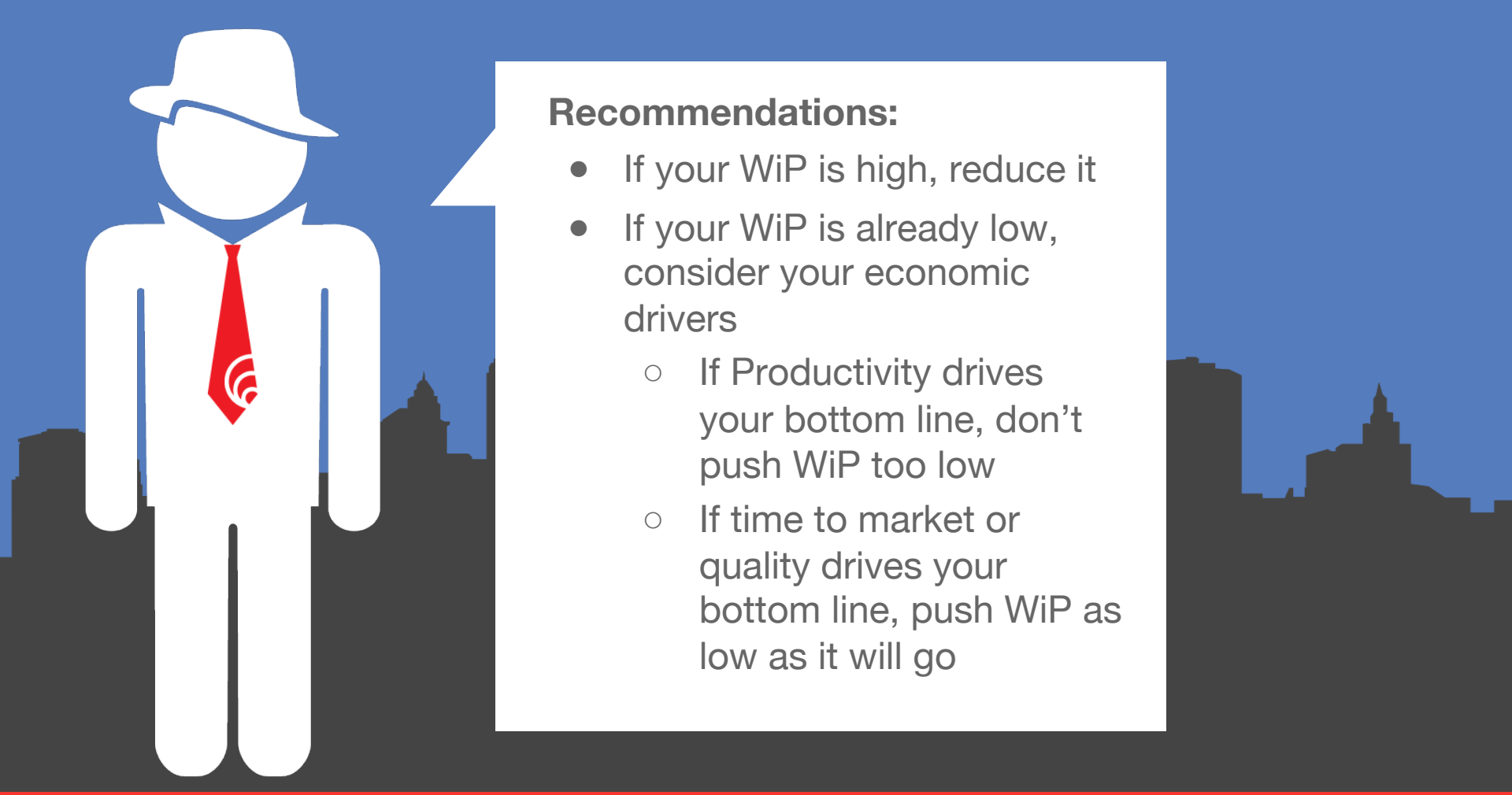




Facts Discovered:

Teams that most aggressively control WiP:

- Have ½ the Time in Process (TiP)
- Have ¼ as many defects
- But have 34% lower productivity



Recommendations:

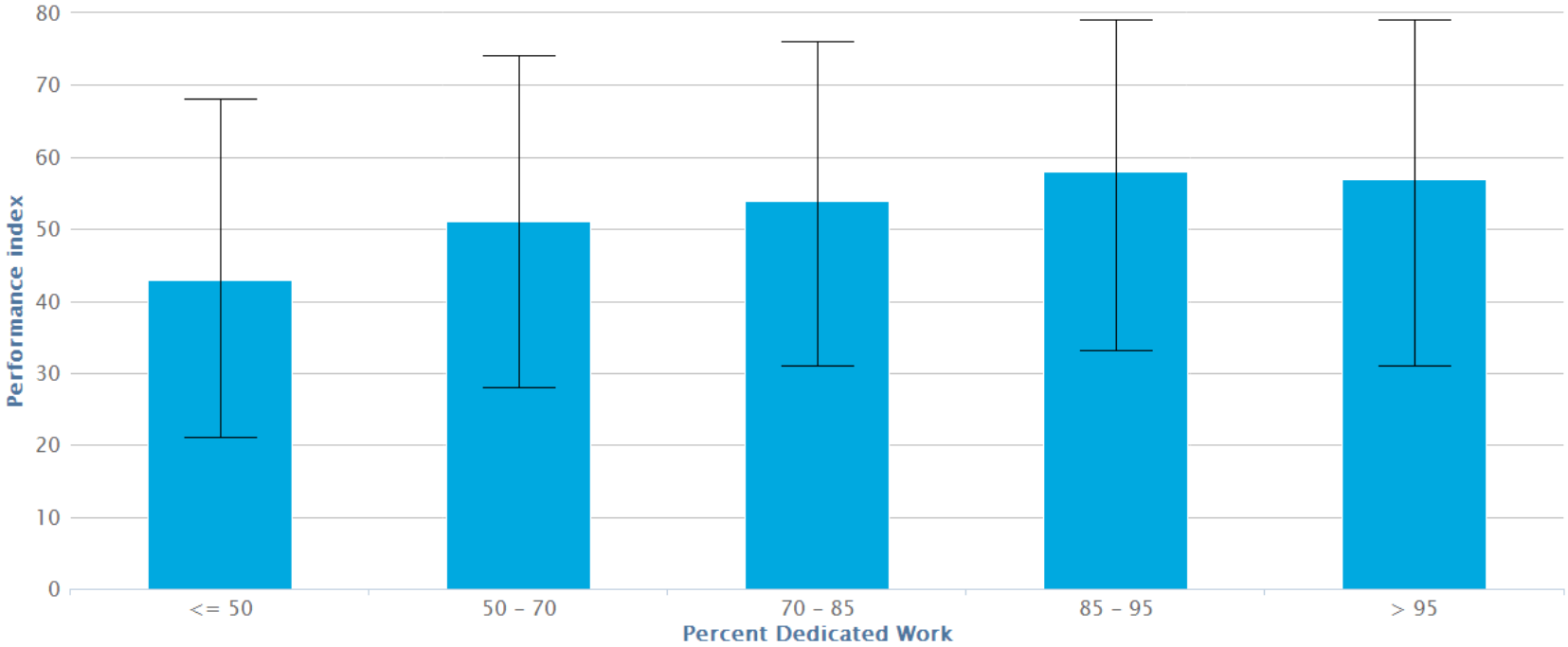
- If your WiP is high, reduce it
- If your WiP is already low, consider your economic drivers
  - If Productivity drives your bottom line, don't push WiP too low
  - If time to market or quality drives your bottom line, push WiP as low as it will go

The investigation continues with ...

Team stability & Dedication to one team

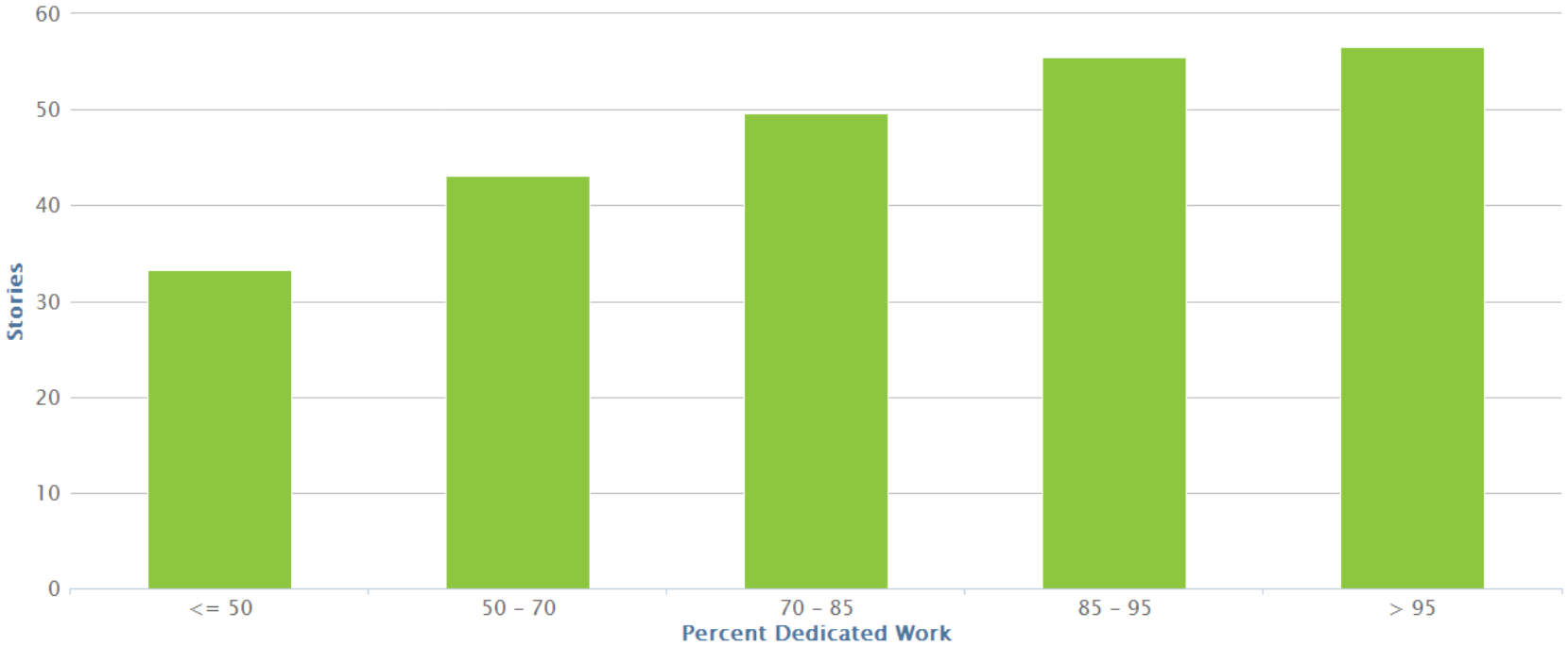
PREDICTABILITY

Percent Dedicated Work relationship to Performance



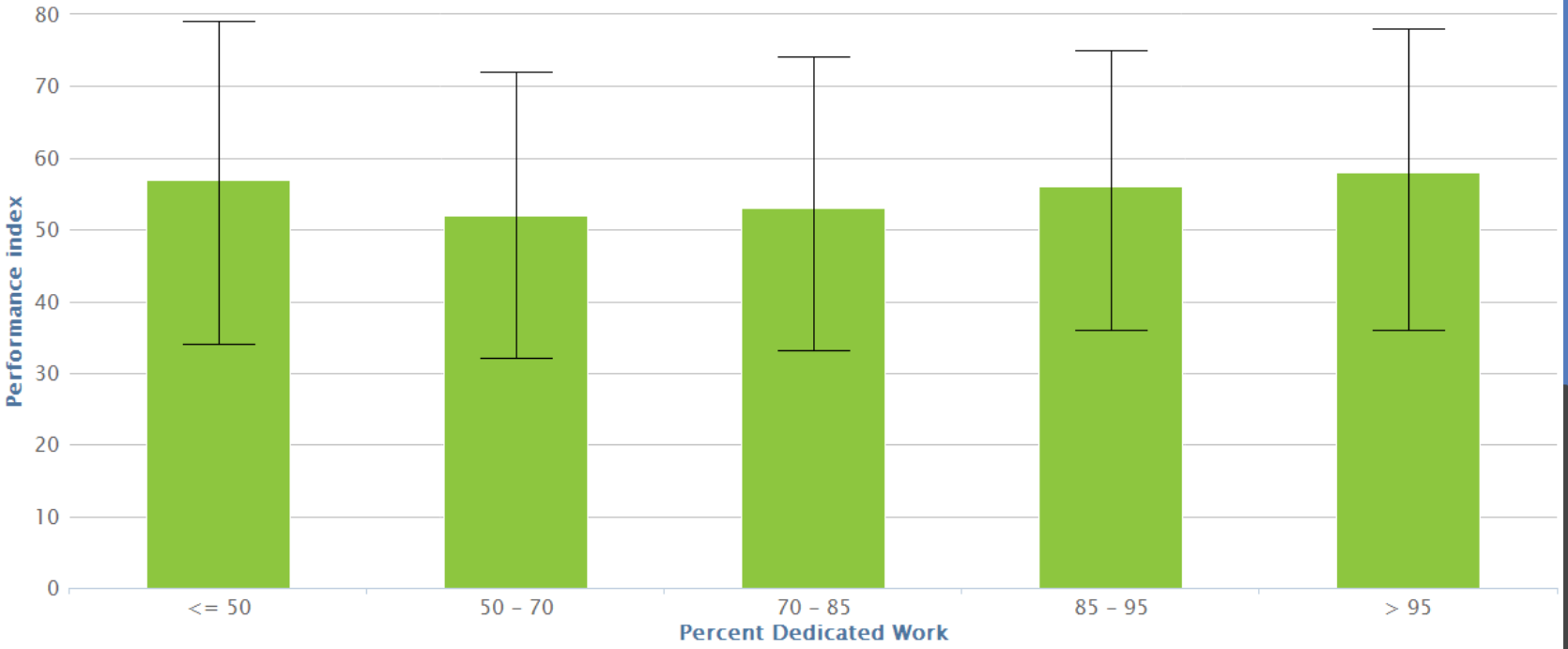
THROUGHPUT (AVERAGE)

Higher is better



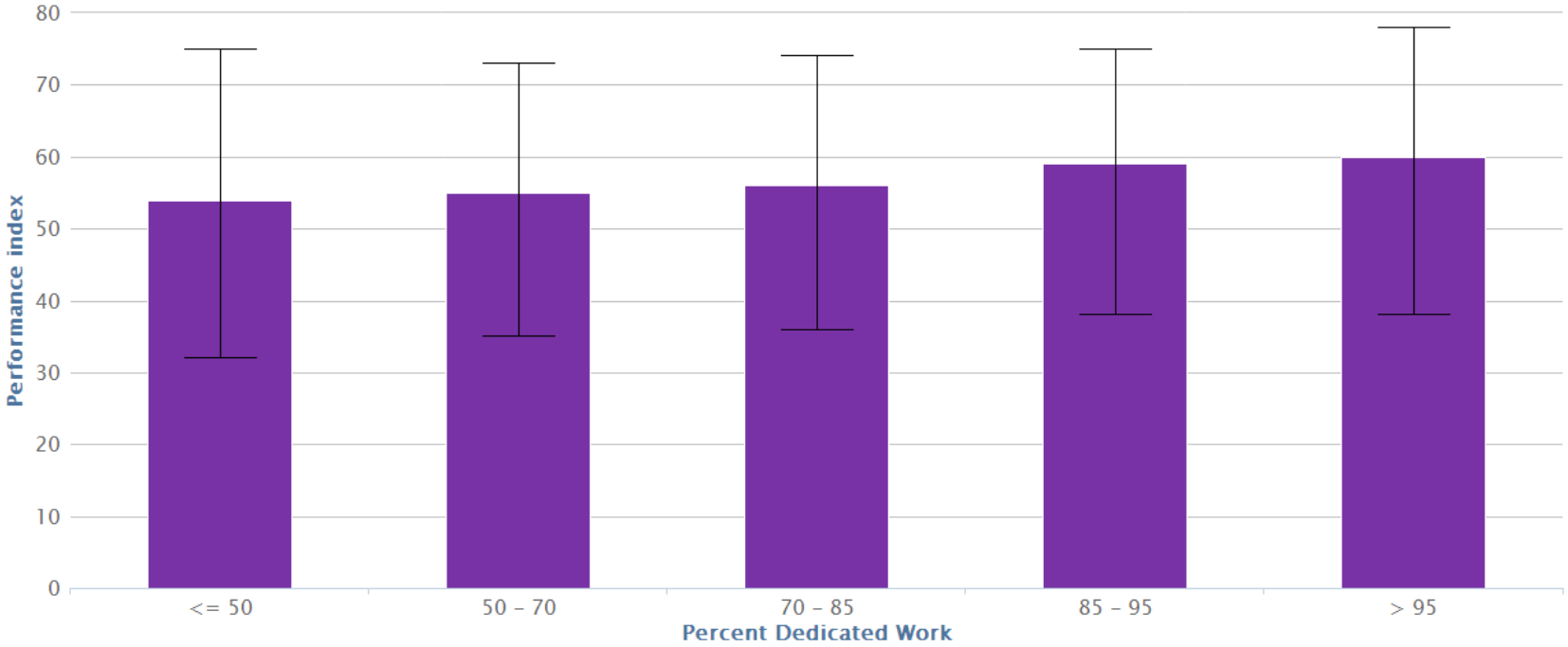
PRODUCTIVITY

Percent Dedicated Work relationship to Performance



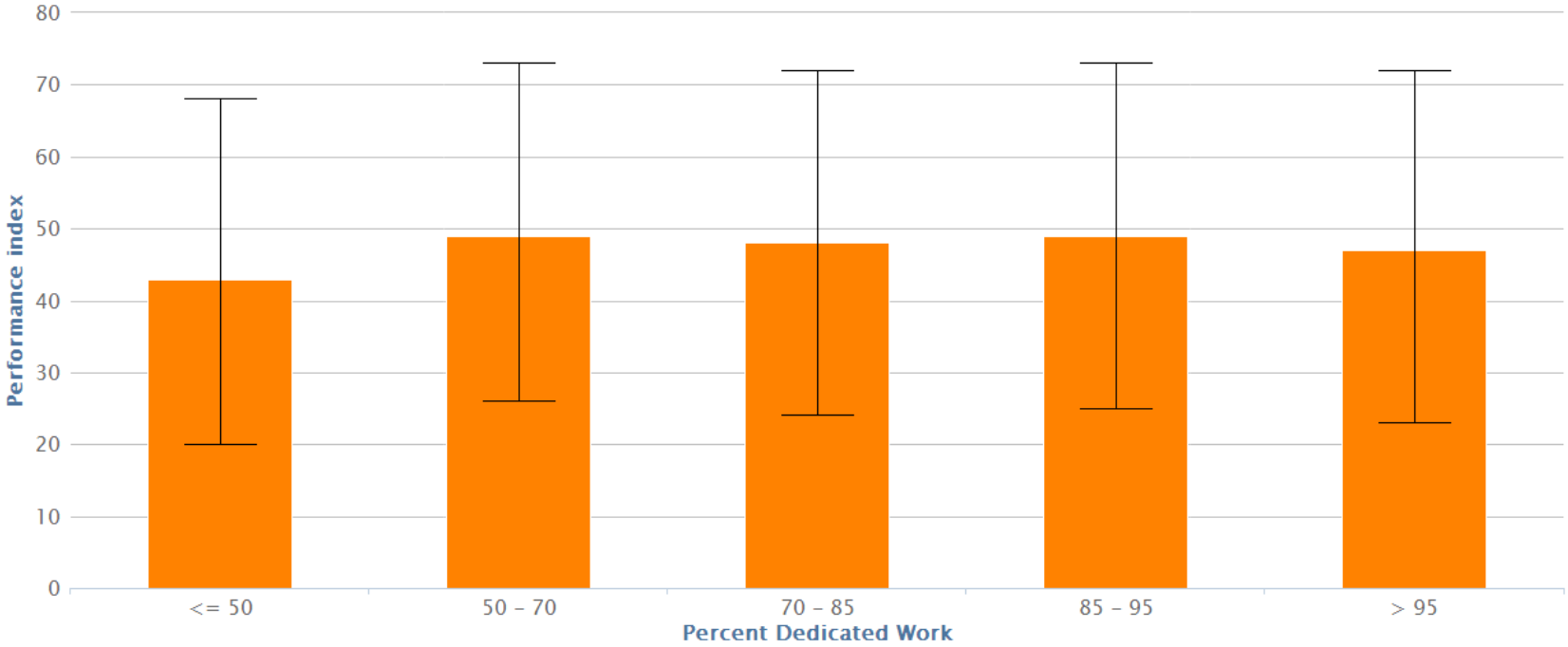
RESPONSIVENESS

Percent Dedicated Work relationship to Performance



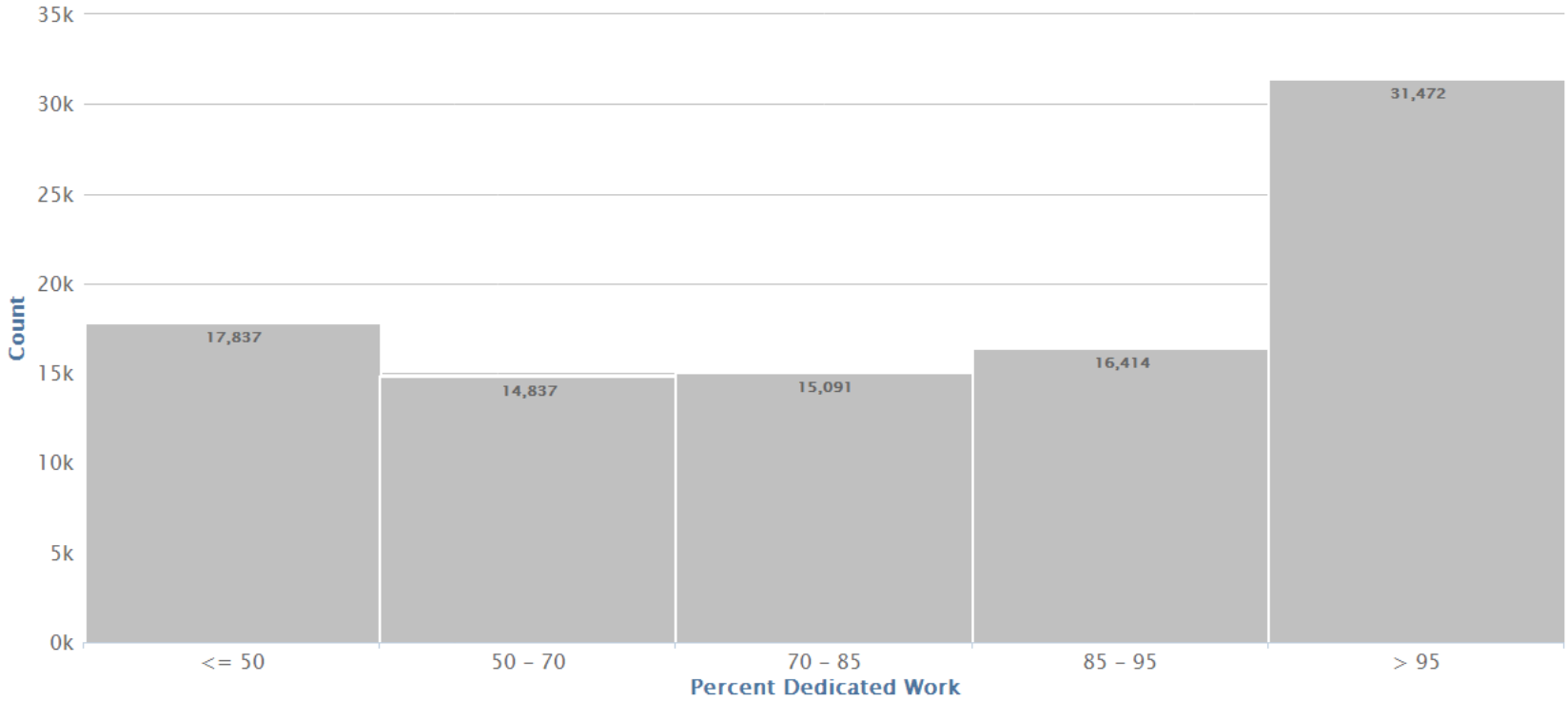
QUALITY

Percent Dedicated Work relationship to Performance



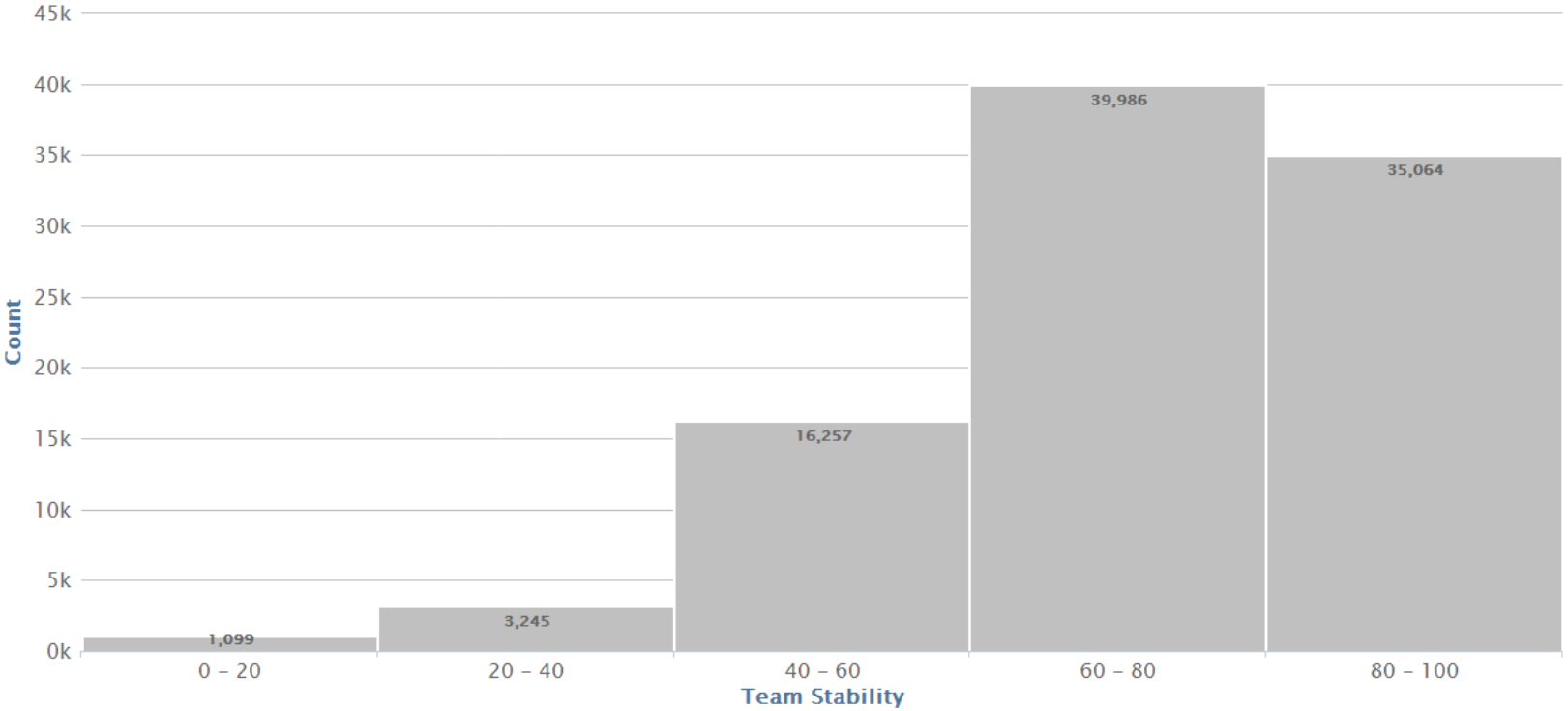
PERCENT DEDICATED WORK

Counts of measurements in each bucket



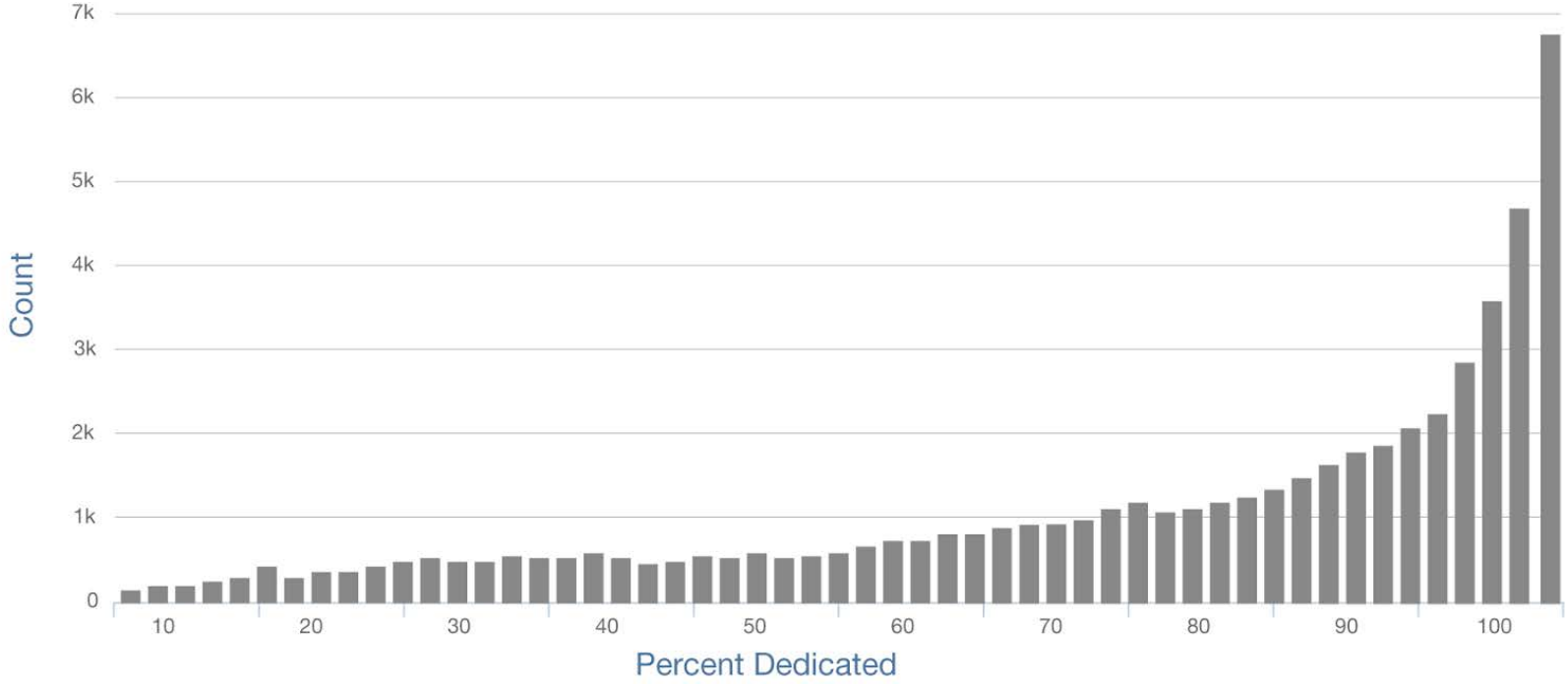
TEAM STABILITY

Counts of measurements in each bucket



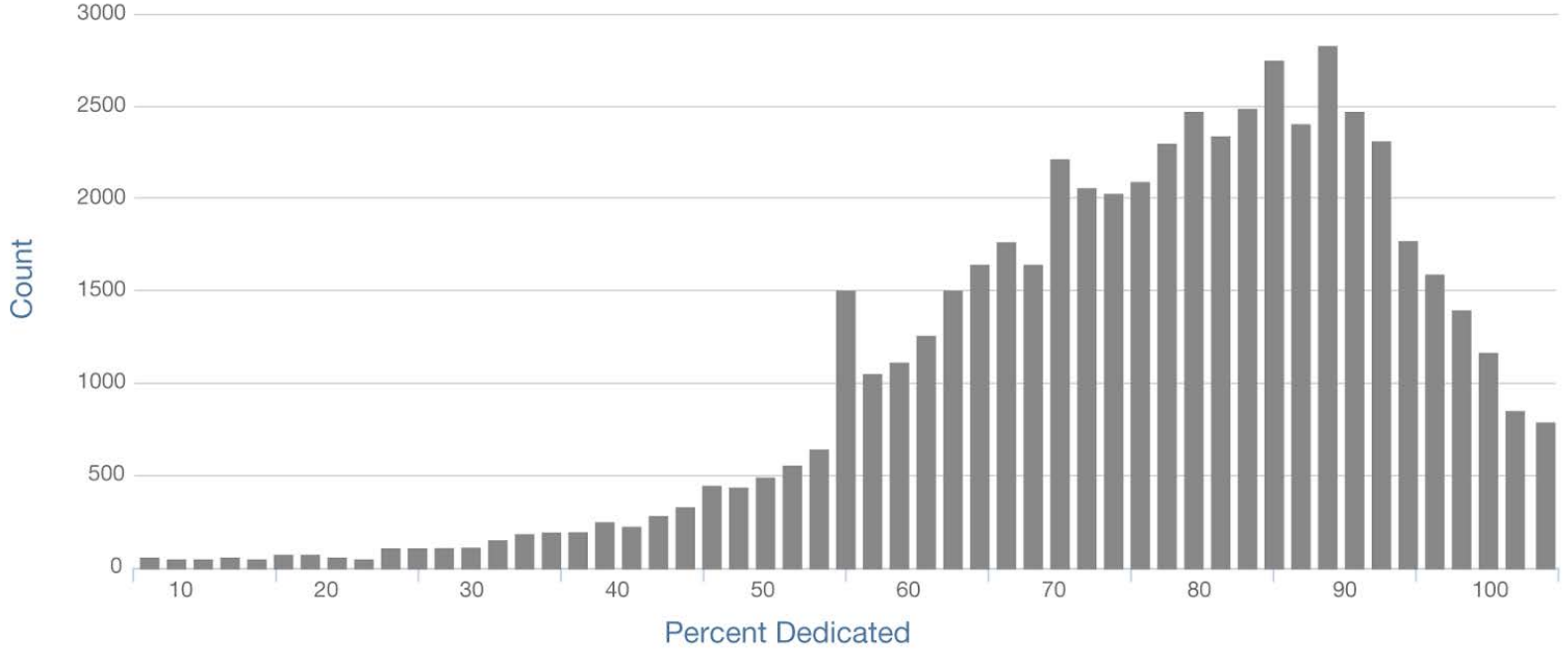
PERCENT DEDICATED

Counts of measurements in each bucket



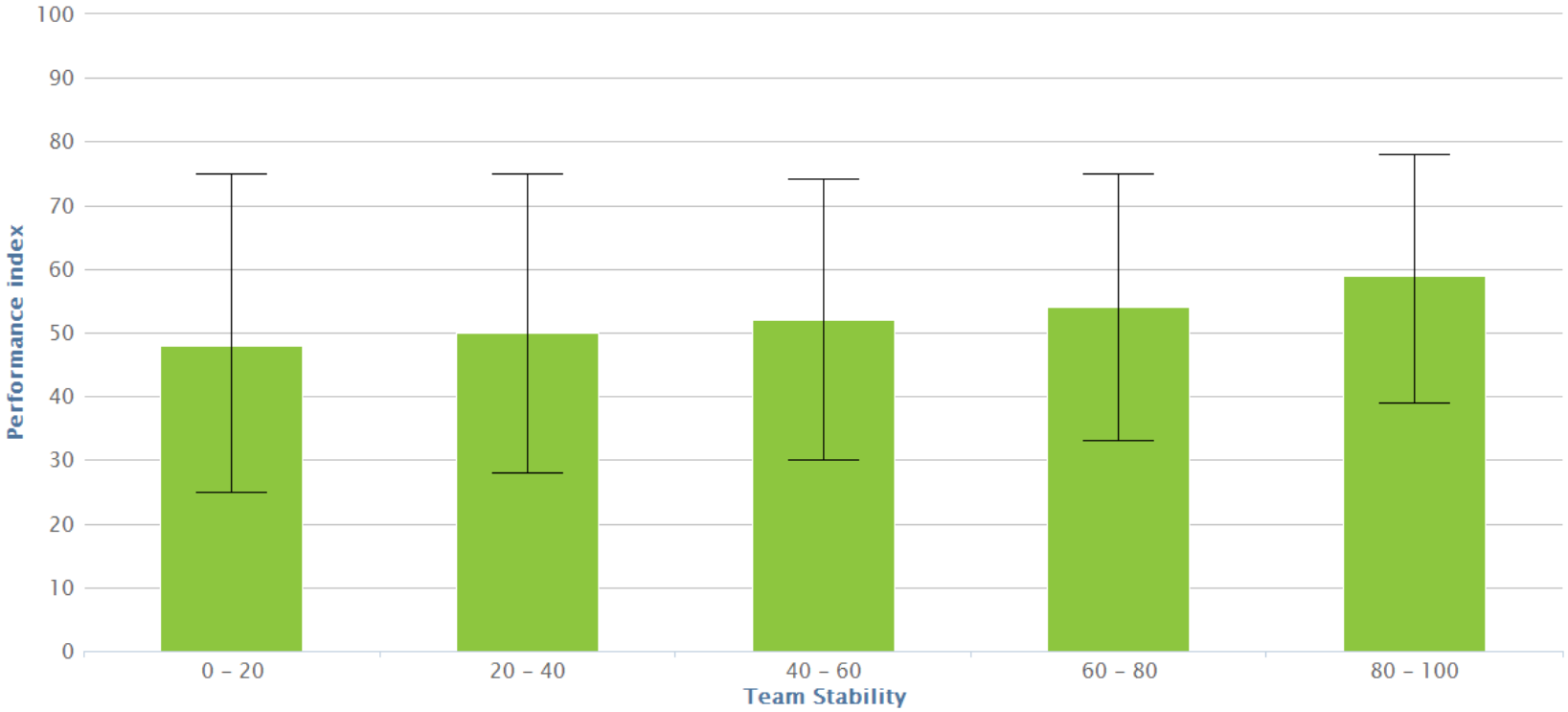
TEAM STABILITY

Counts of measurements in each bucket



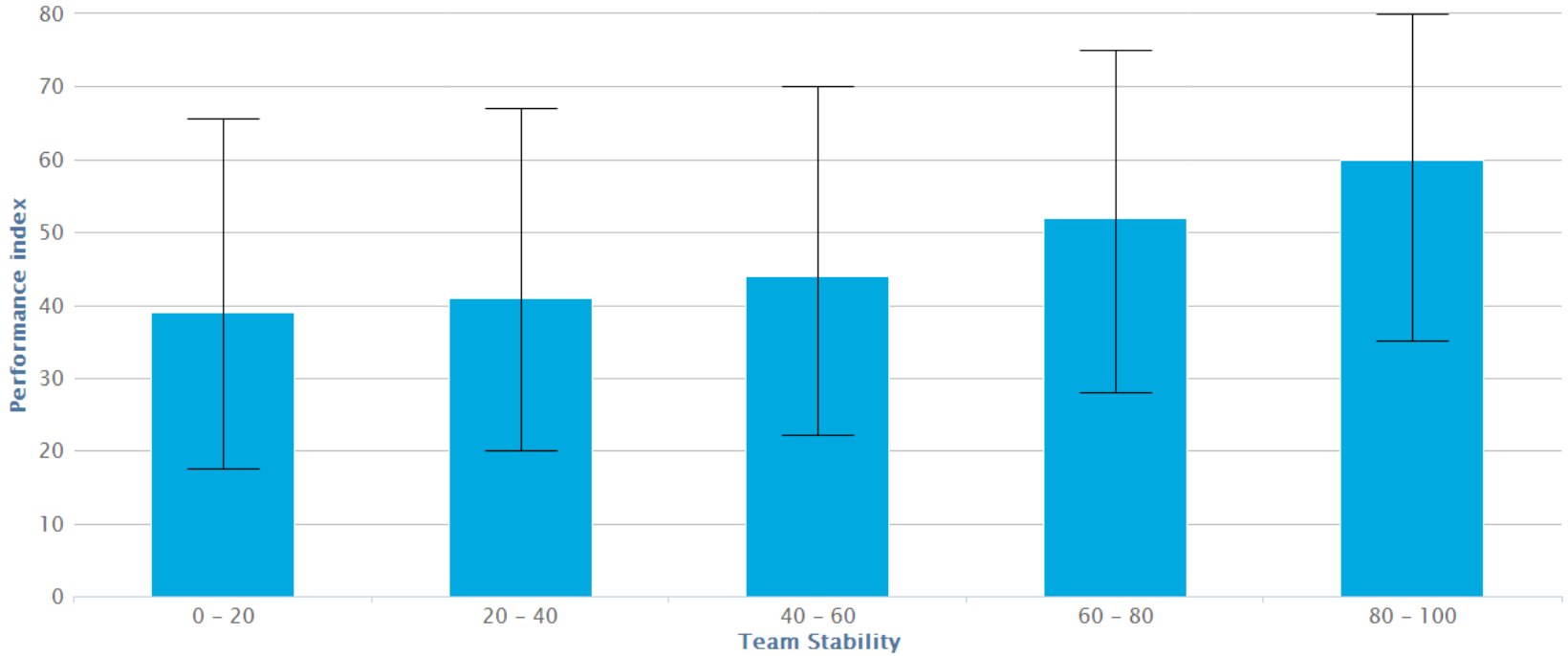
PRODUCTIVITY

Team Stability relationship to Performance



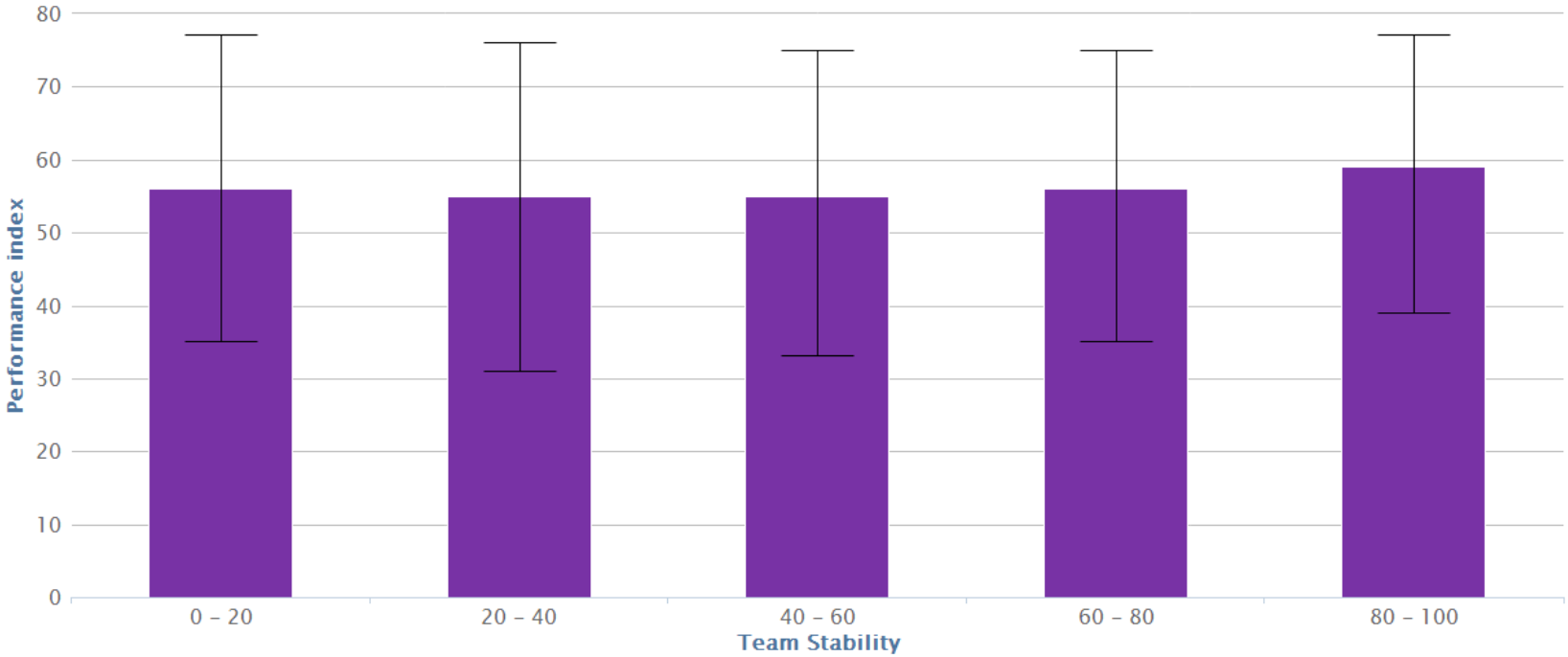
PREDICTABILITY

Team Stability relationship to Performance



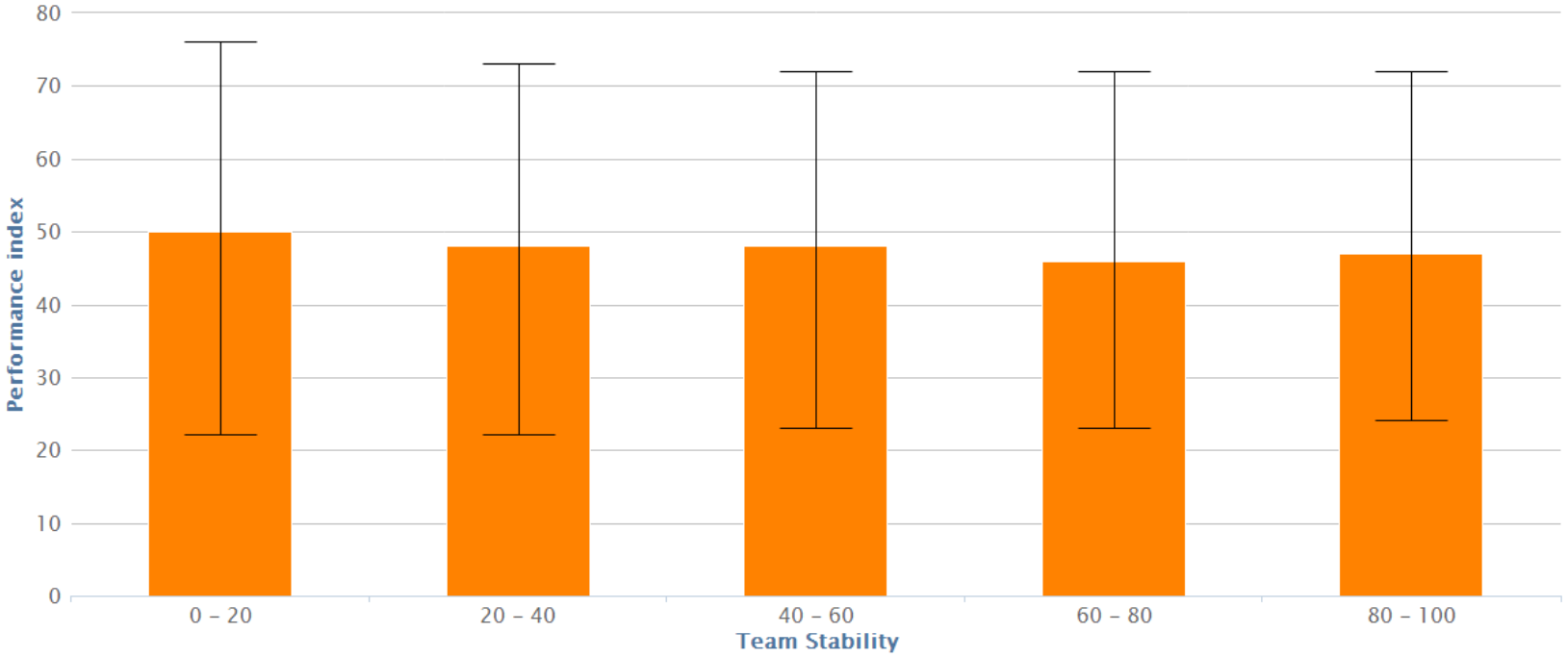
RESPONSIVENESS

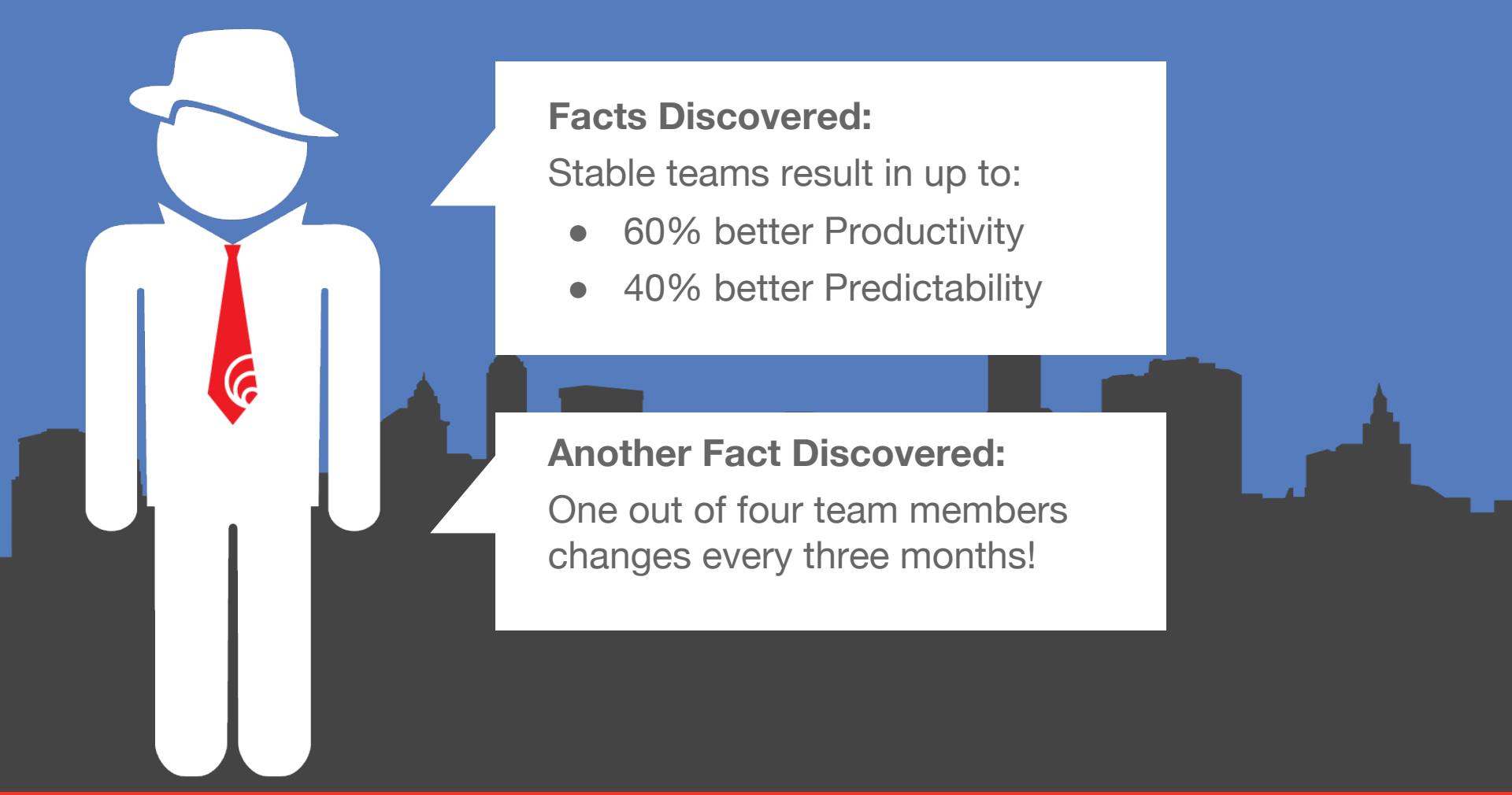
Team Stability relationship to Performance

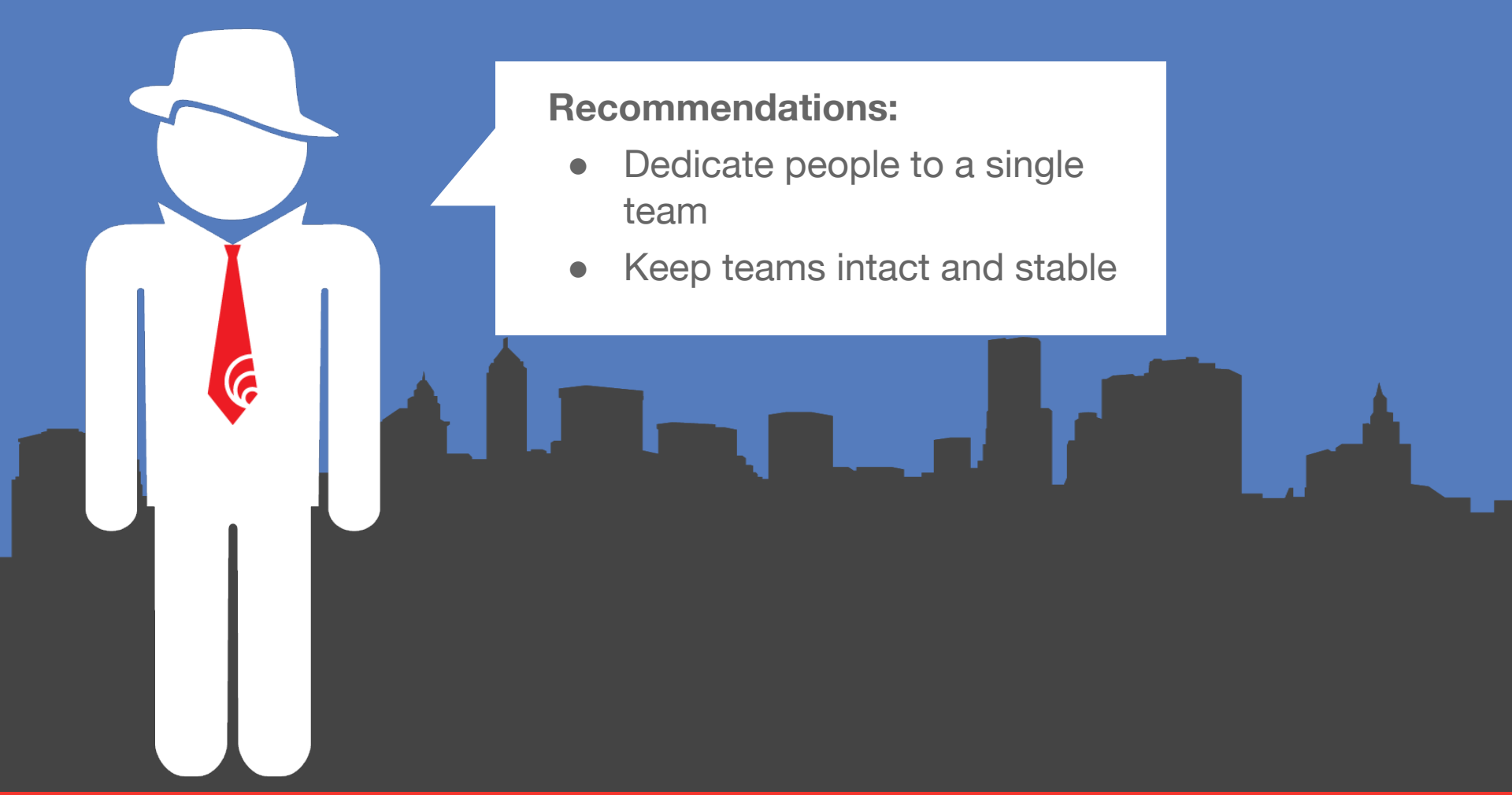


QUALITY

Team Stability relationship to Performance





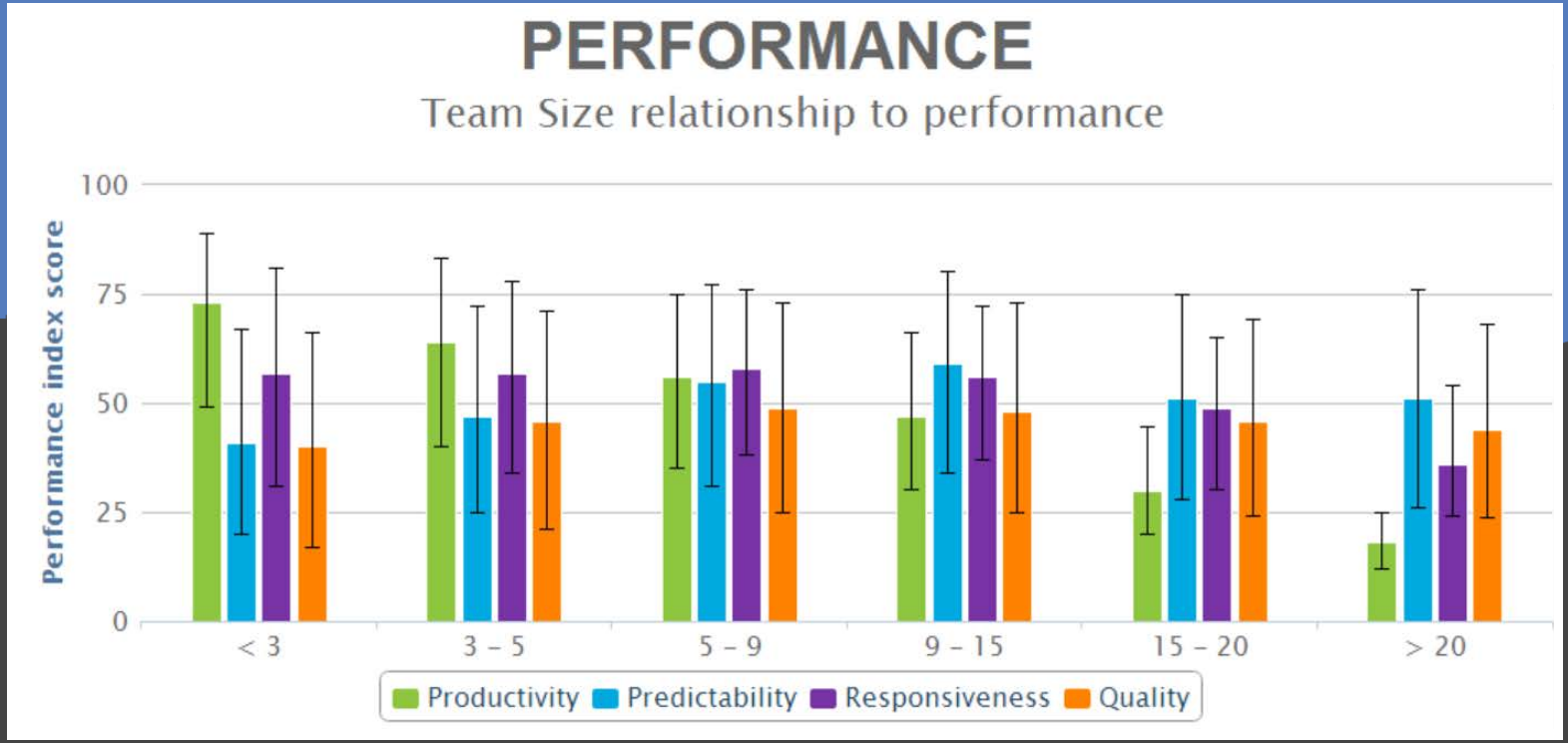


The investigation continues with ...

Team size

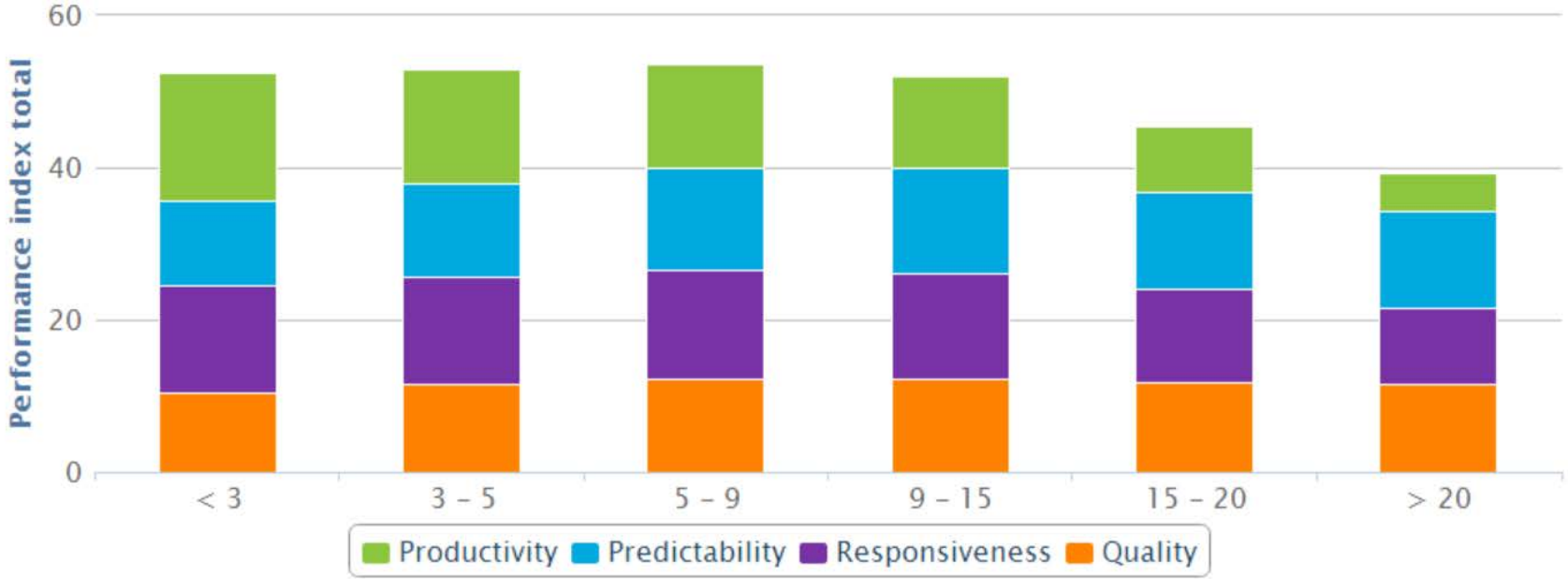
Balance your team's Performance

Agile recommends that the ideal team size is 7± 2. How ideal is that when we actually look at the data?



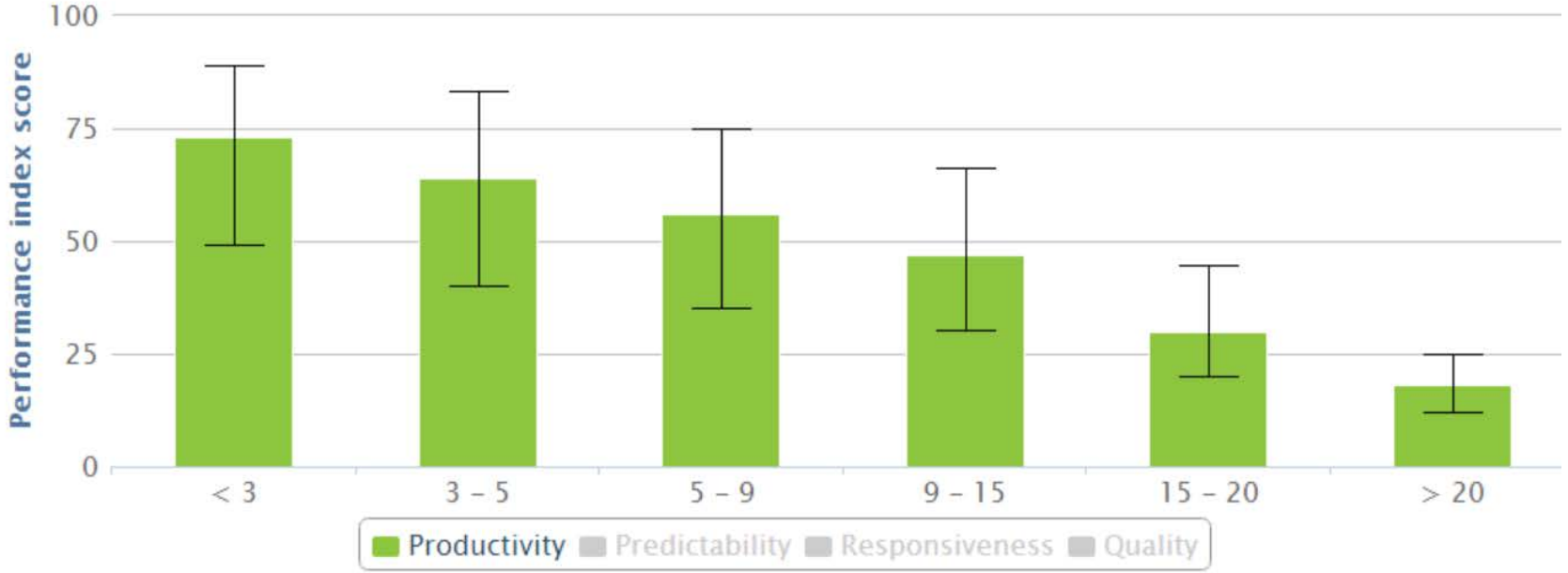
PERFORMANCE

Team Size relationship to performance



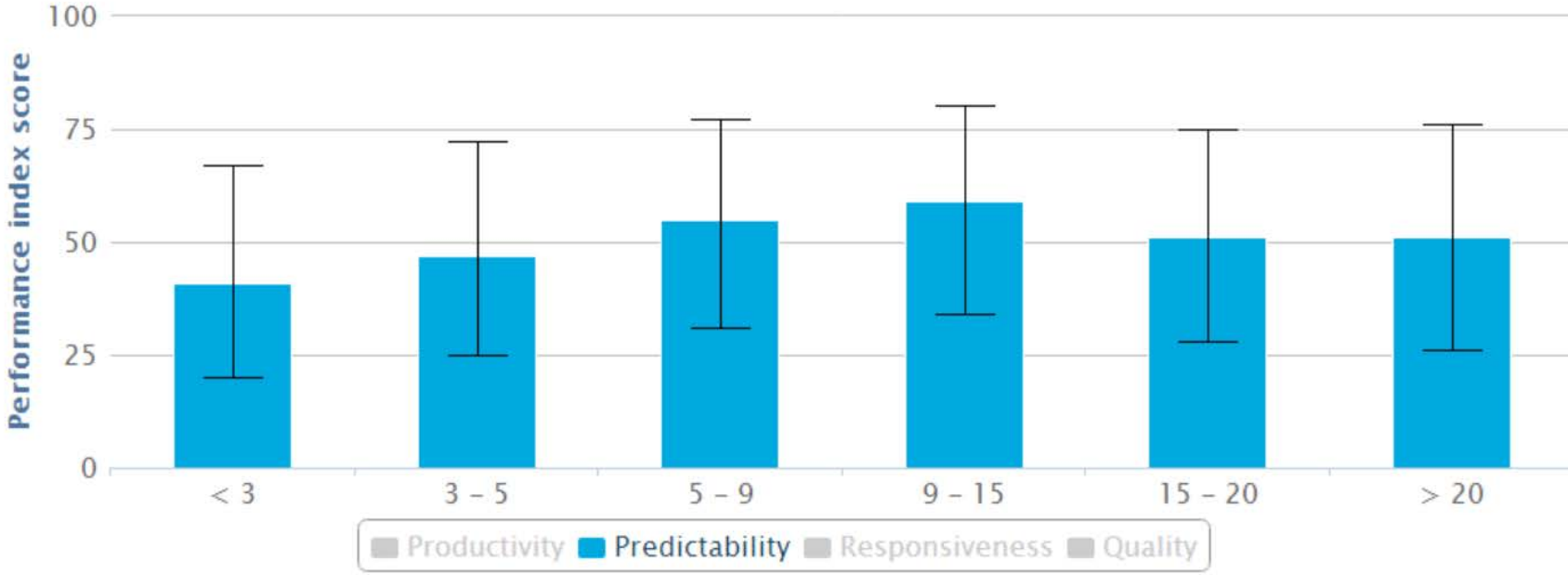
PERFORMANCE

Team Size relationship to performance



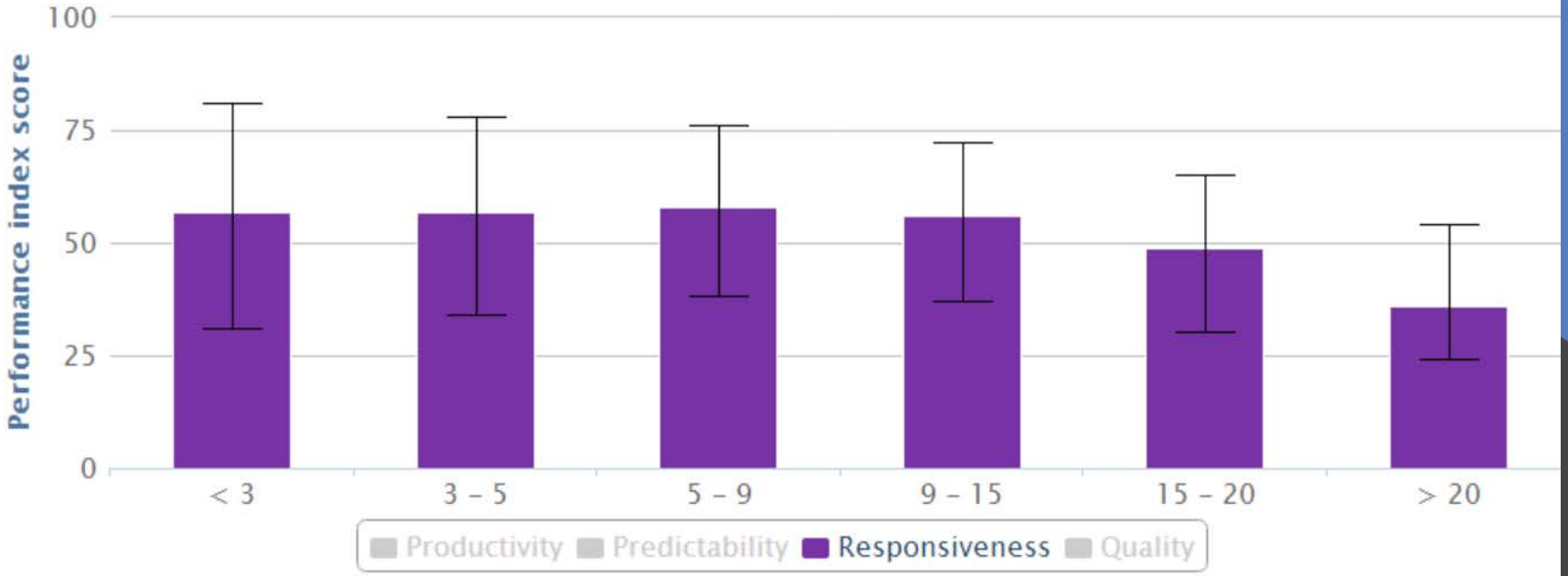
PERFORMANCE

Team Size relationship to performance



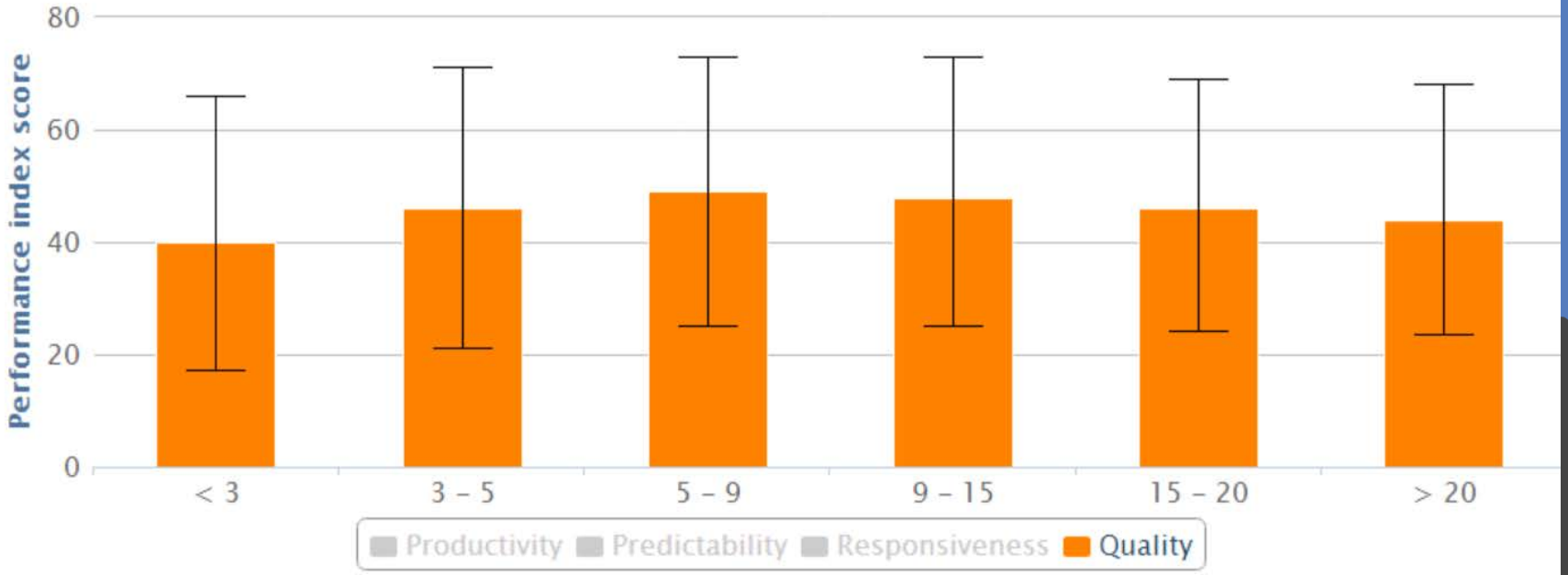
PERFORMANCE

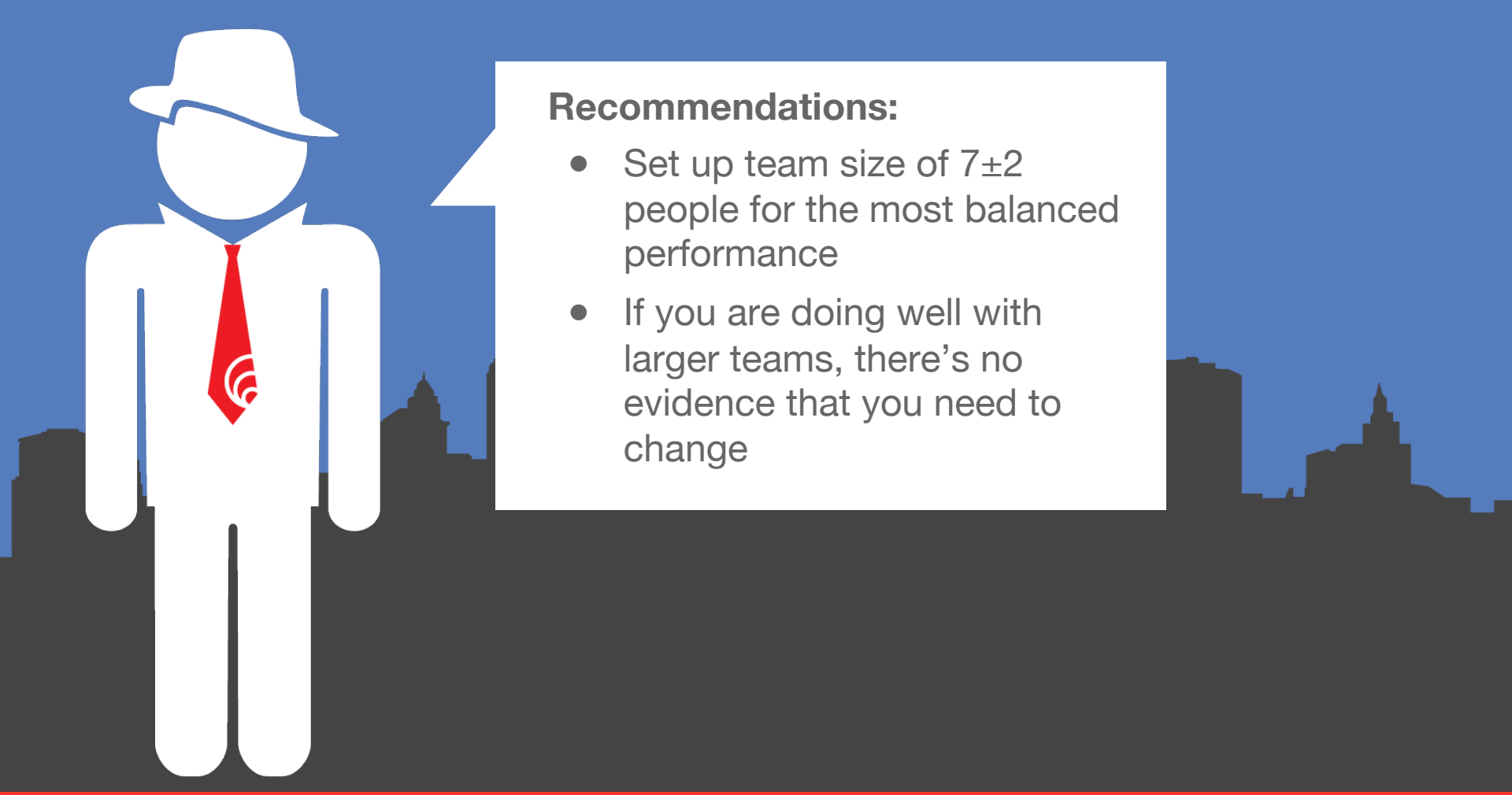
Team Size relationship to performance



PERFORMANCE

Team Size relationship to performance



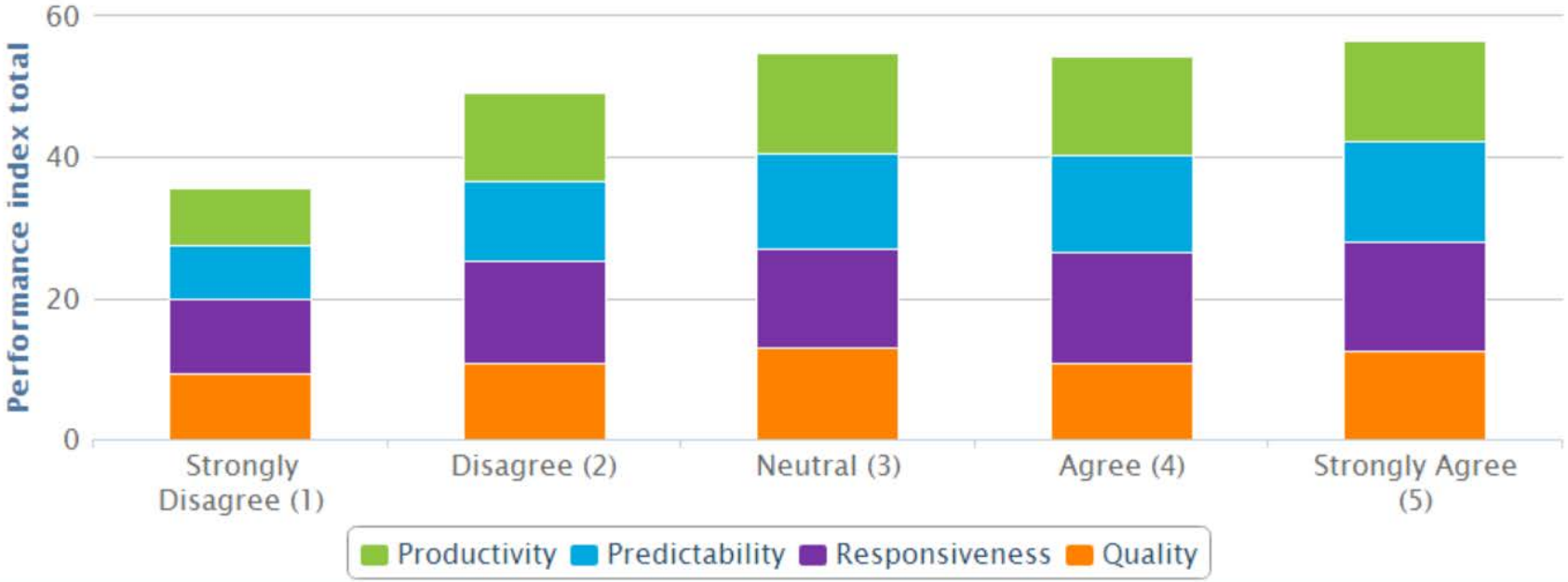


The investigation continues with ...

Retrospectives

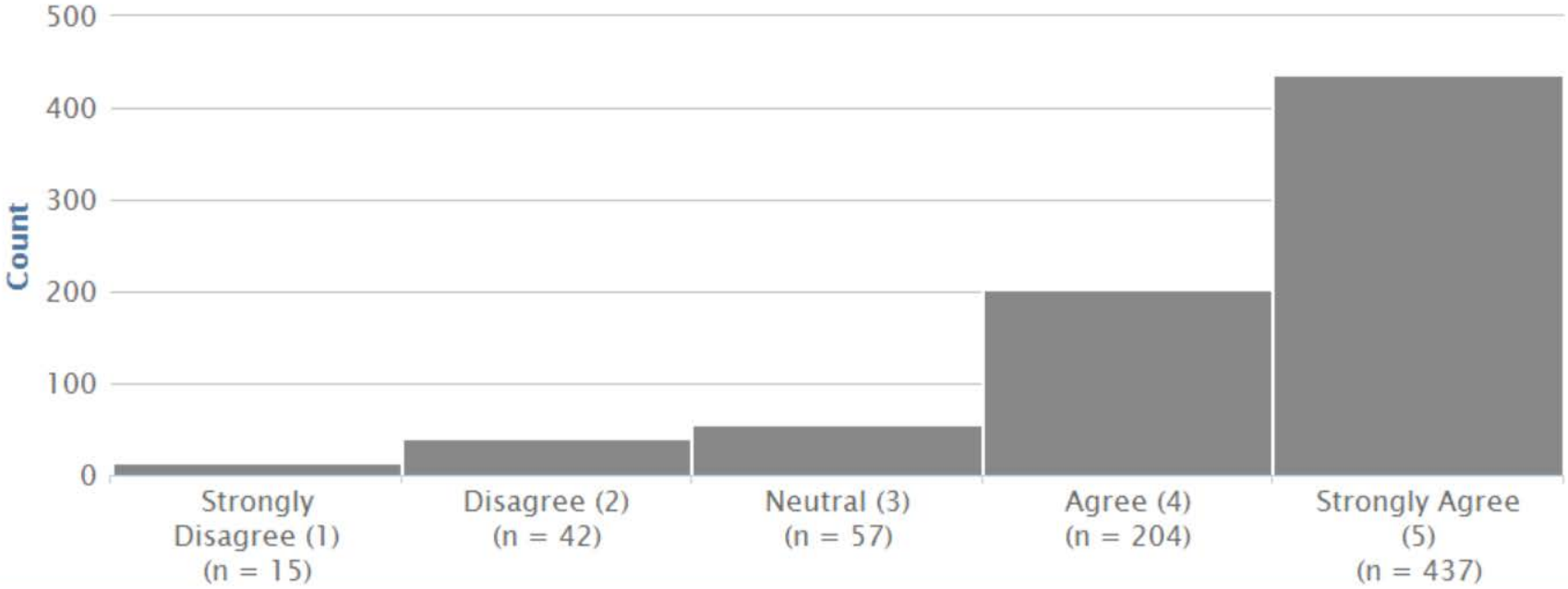
PERFORMANCE

Has Sprint Retrospective relationship to performance



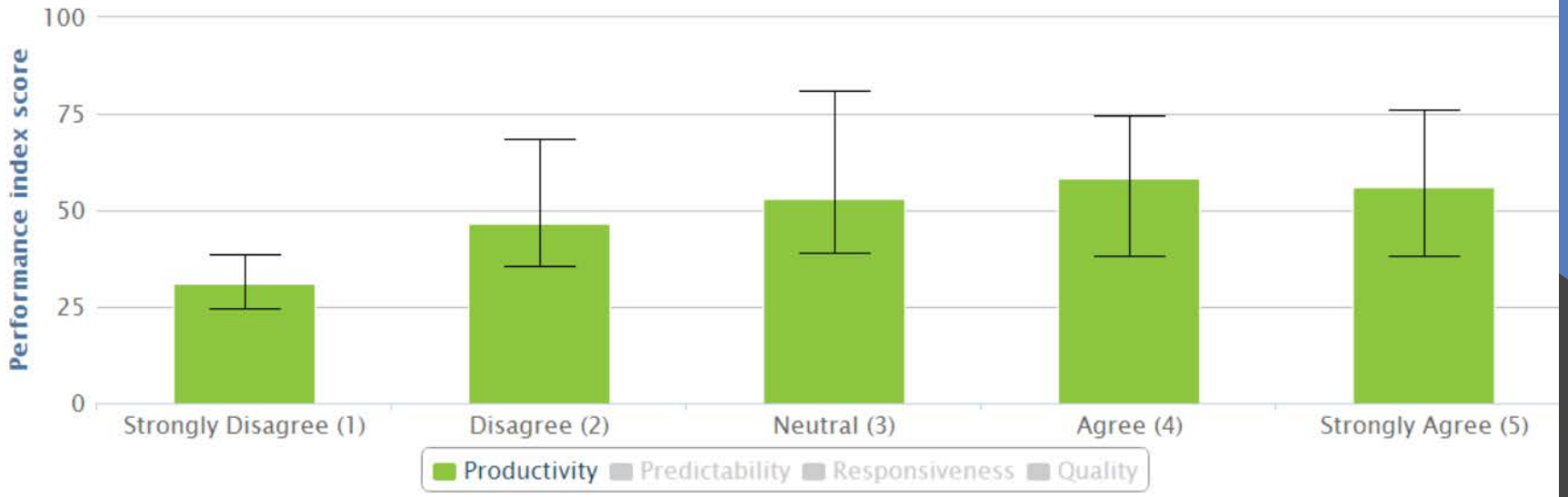
HAS SPRINT RETROSPECTIVE

Counts of measurements in each bucket



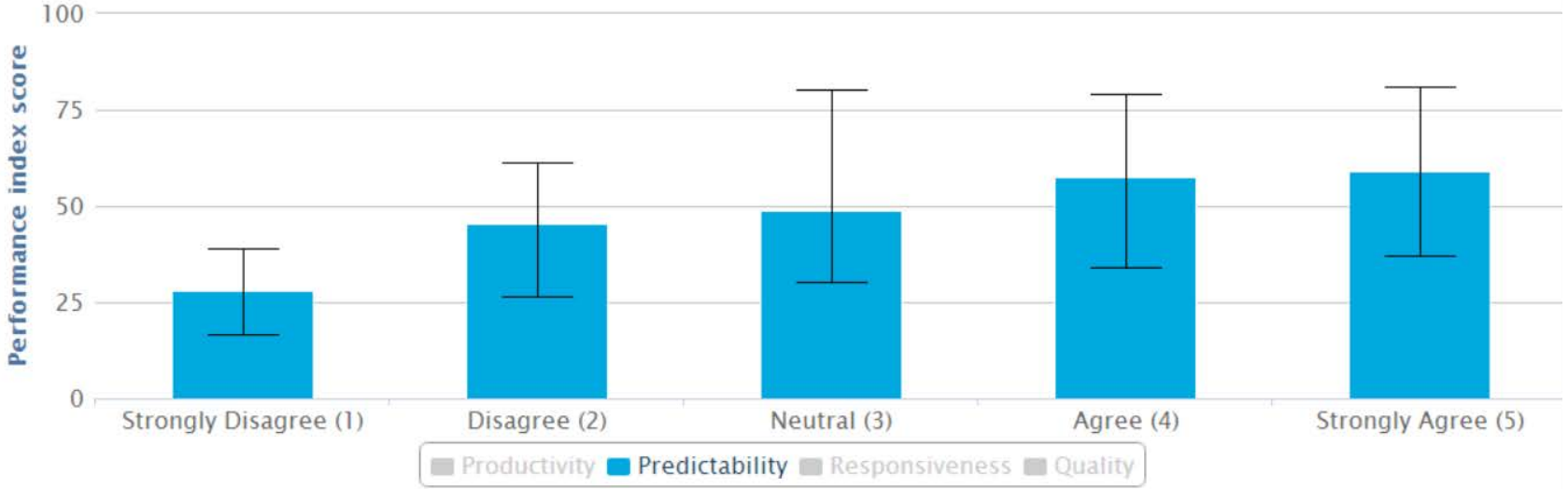
PERFORMANCE

Has Sprint Retrospective relationship to performance



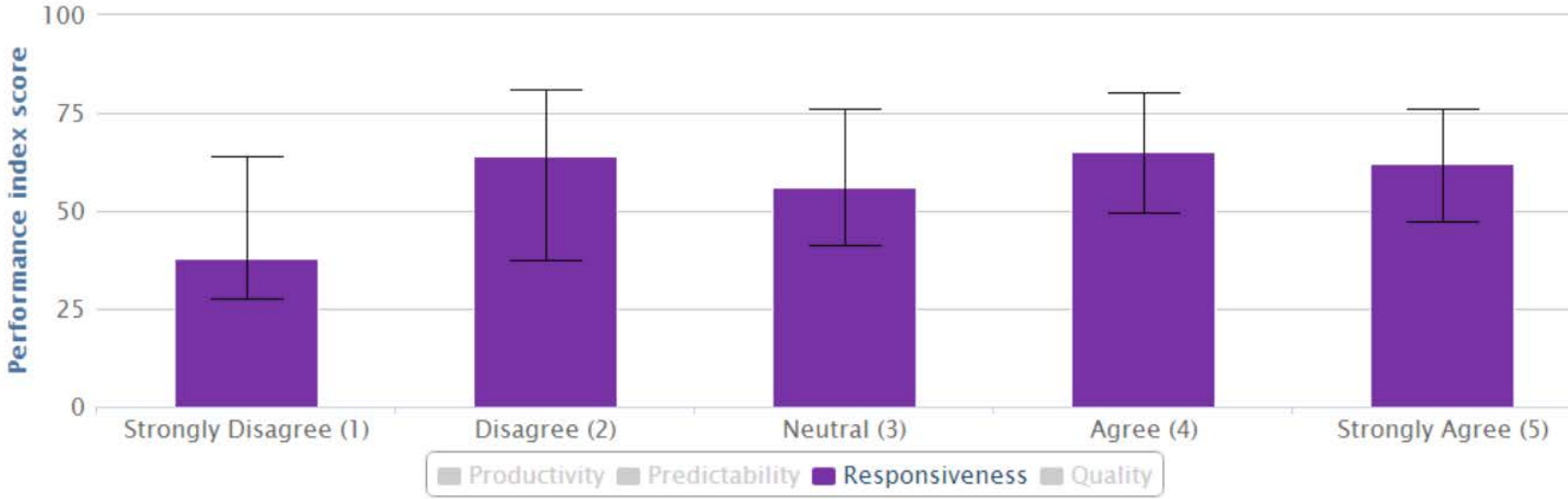
PERFORMANCE

Has Sprint Retrospective relationship to performance



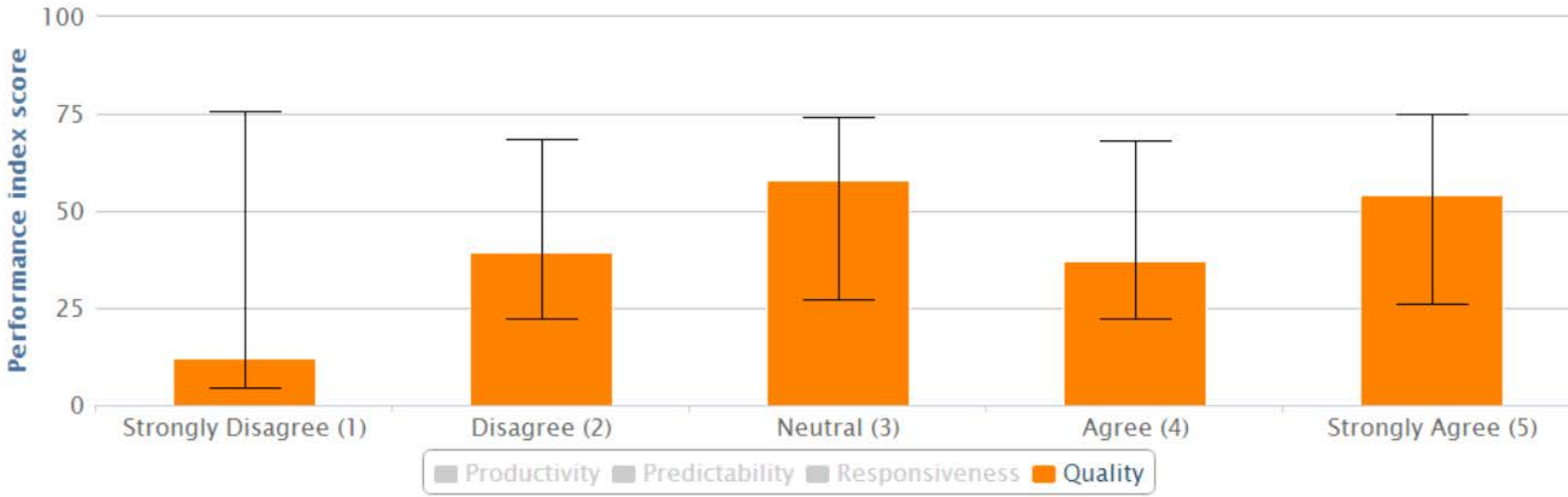
PERFORMANCE

Has Sprint Retrospective relationship to performance



PERFORMANCE

Has Sprint Retrospective relationship to performance

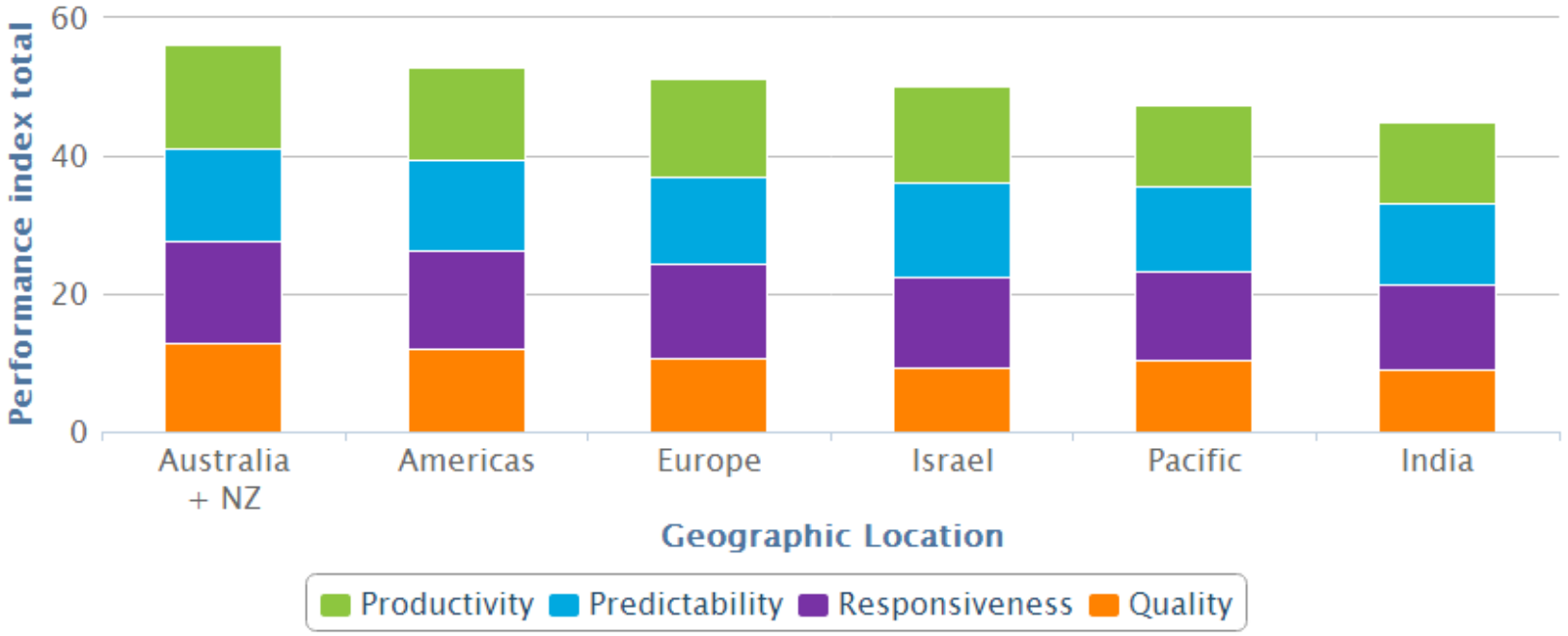


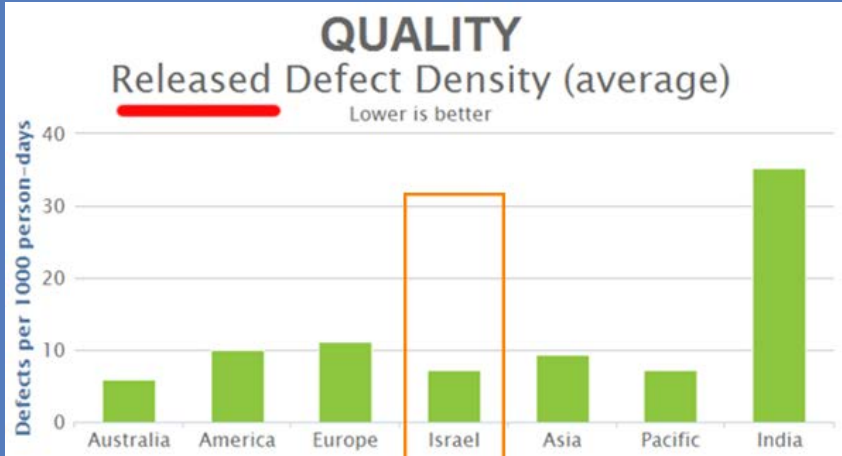
The investigation continues with ...

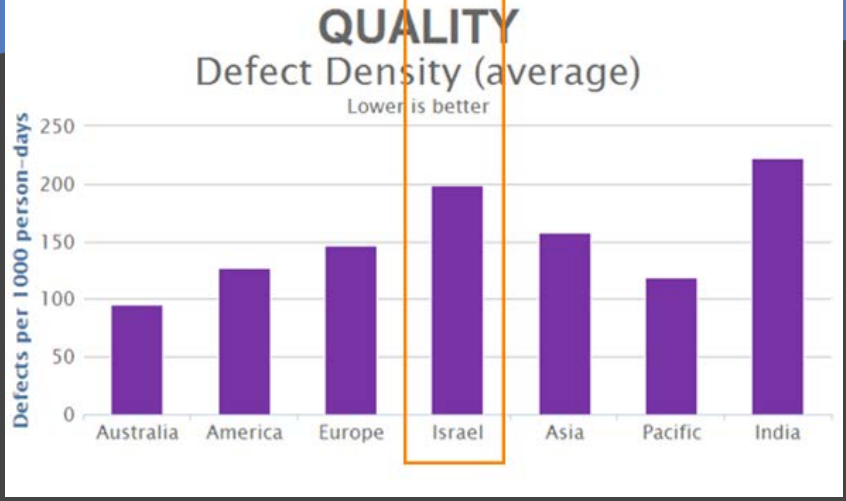
Geography

PERFORMANCE

Geographic Location relationship to Performance





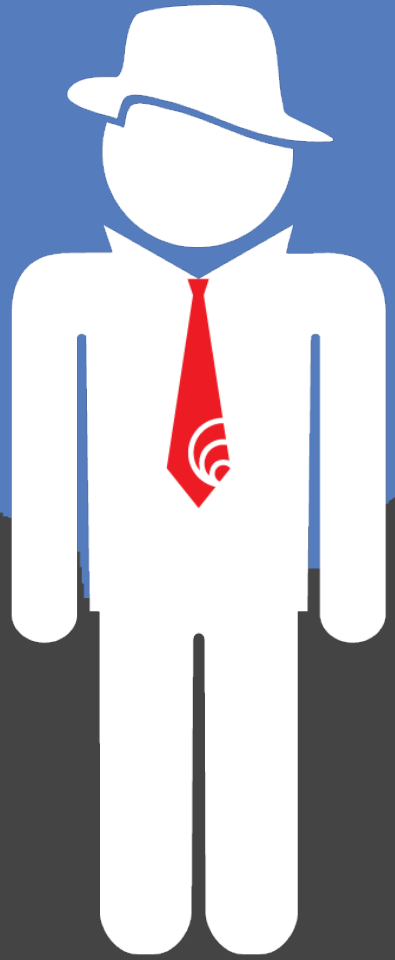


Israel-based teams

- Find more defects overall
- But find fewer in production
- Theory: May correlate with high use of static analysis tools

India-based teams

- Find more defects overall
- Released and unreleased
- Theory: May correlate with high use of static analysis tools
- Theory: Could be recording bias



Facts Discovered:

- Differences are slight but statistically significant
- Australia has the best overall performance
- India the worst. However, there could be a reporting bias for defects
- Israel seems to catch the most defects before production. Heavy use of static analysis?