Linaro Connect
MADRID 2024 | MAY 14-17 2024

Improving Delivery Process Performance for SW-Defined Products

Helped by Data Analytics

Table of Contents

The Speakers

@dizquierdo

Daniel Izquierdo

- Empirical Software Engineering background, PhD in the field, data-driven approach
- Co-Founder and CEO of Bitergia, a software development analytics company
  - Core competencies on producing high quality data, incremental support, historical background. Contextualizing data into different verticals
  - InnerSource and open source consulting services and improving the dynamics of collaboration across units. Bringing visibility to Upper Management on development activities
- President at the InnerSource Commons Foundation
- Governing Board at the CHAOSS, a Linux Foundation working group

dizquierdo@bitergia.com

linkedin.com/in/dicortazar

@toscalix

Agustín Benito Bethencourt

- Continuous Delivery, FLOSS, data analytics, agility and remote work advocate
- Independent consultant helping companies in two ways:
  - Applying advanced data analytics to production environments to increase delivery performance, partnering with Bitergia, through a service offering called Delivery Performance Analytics
  - Increasing their organizational performance by becoming good open source citizens, like in the case of SCANOSS, as their Ecosystem Manager
- More about Agustín:
  - Background: MBition (Mercedes Benz), SUSE, Linaro, Eclipse Foundation, Codethink...
  - Blog - About - Talks

Contact Agustin

Delivery Performance Analytics - Toscalix website

Bitergia

WHY BITERGIA?

- +15 years of experience with software dev. analytics
- Strong focus on open source and InnerSource: Co-founder of the CHAOSS project and member of the InnerSource Commons Foundation
- Full service for 100% open source metrics solution; maintains the open source GrimoireLab metrics tools
- Fully GDPR compliant
- Flexible team and engagement options
- ISO certified for 9001 and 27001

Contact the Delivery Performance Analytics team:

delivery-performance@bitergia.com

Delivery Performance Analytics - Bitergia website

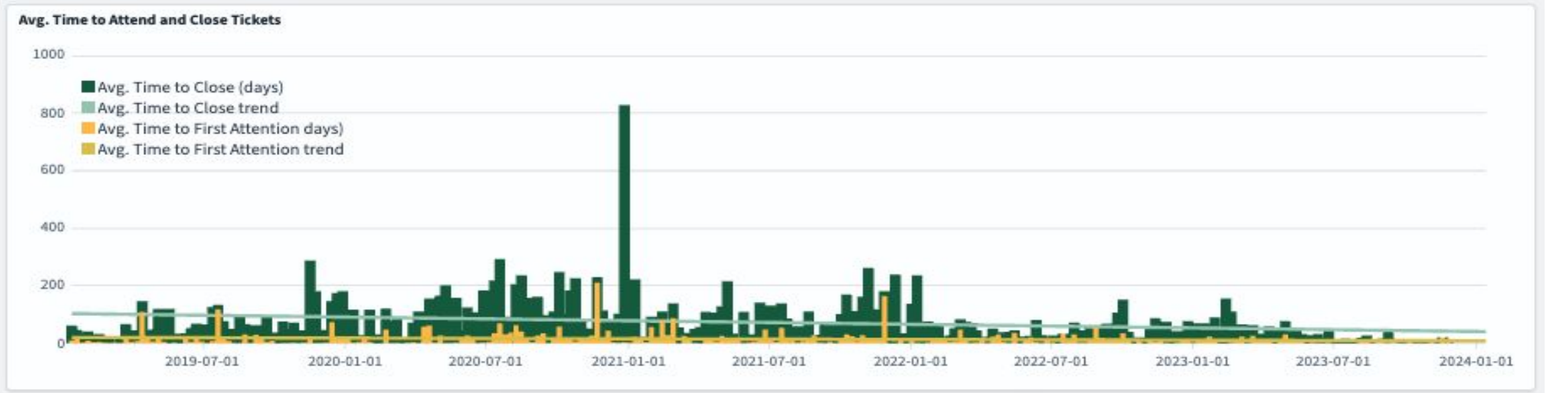
Highly-regulated verticals experienced engagements

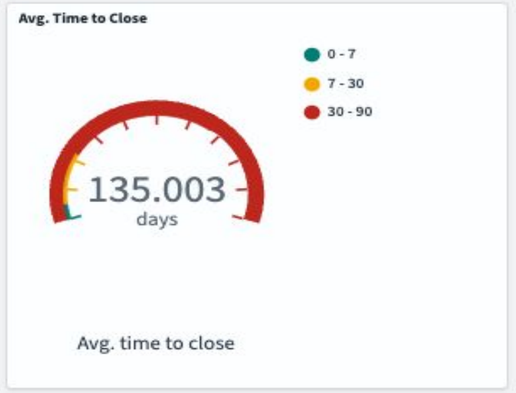
Health
Banking/Finances
Automotive

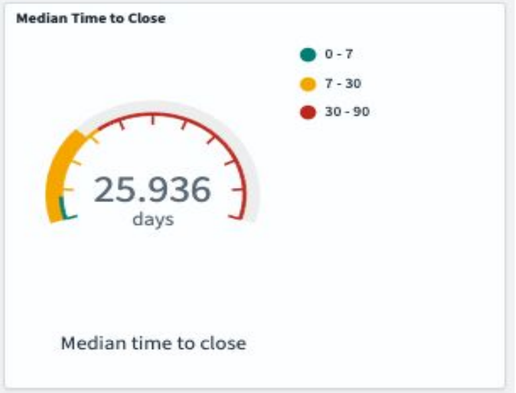
Introduction

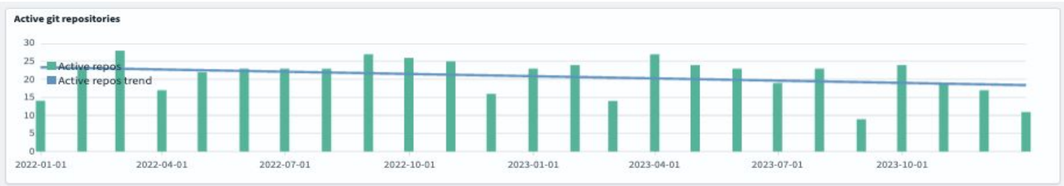
Manufacturing can be “under control” …

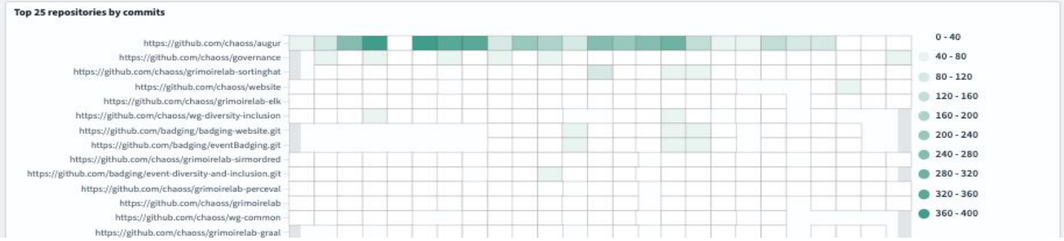
… Why not Software Production?

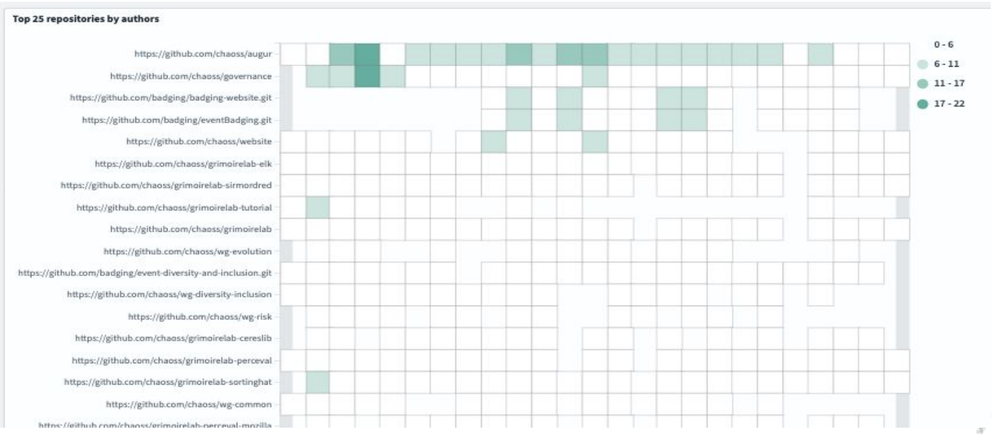












Frameworks


The Speakers
Introduction
Frameworks
Data Analytics Journey
Examples
Takeaways

Metrics Frameworks

DORA

by Google

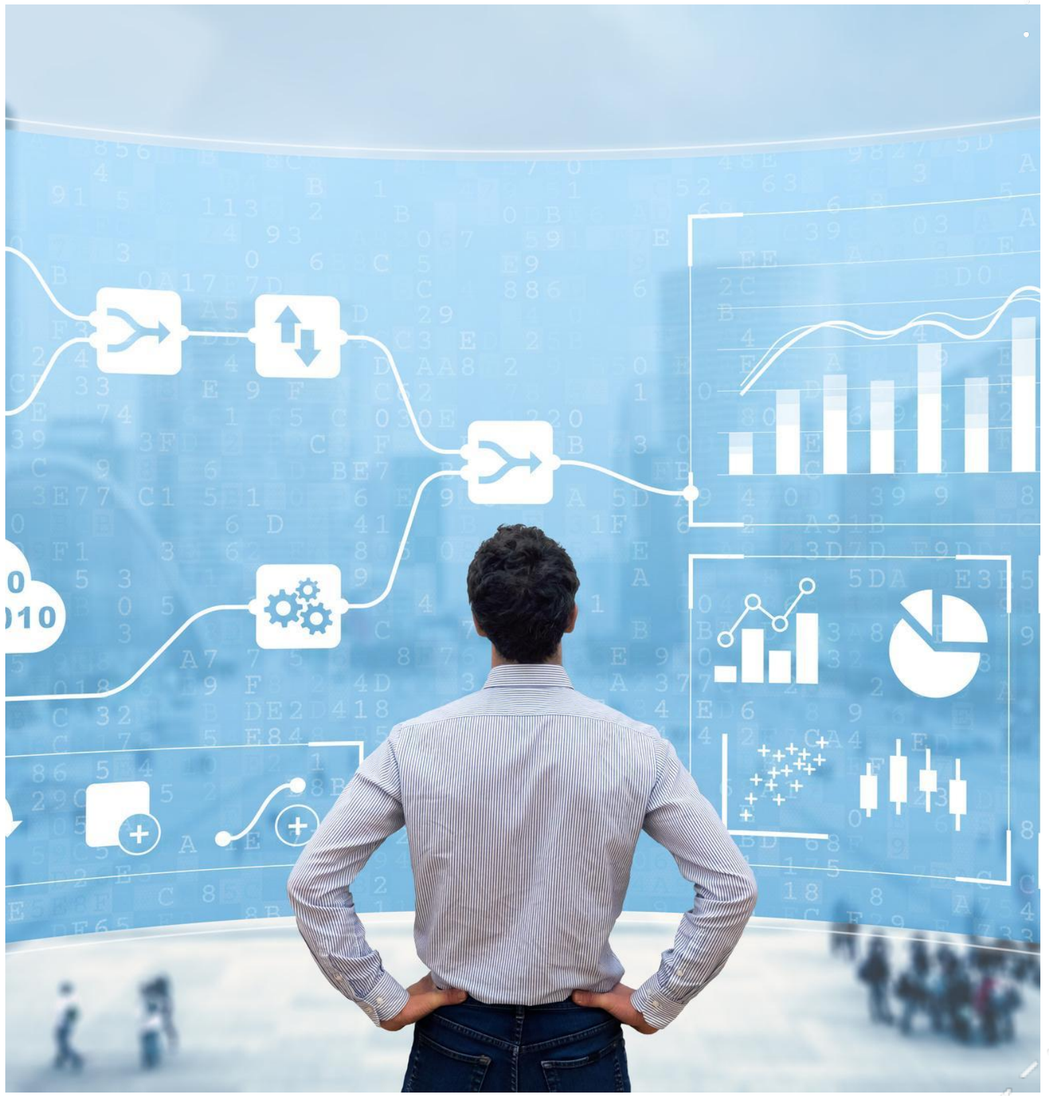
Metrics Frameworks

SPACE→DevEx framework

lead by Microsoft Research

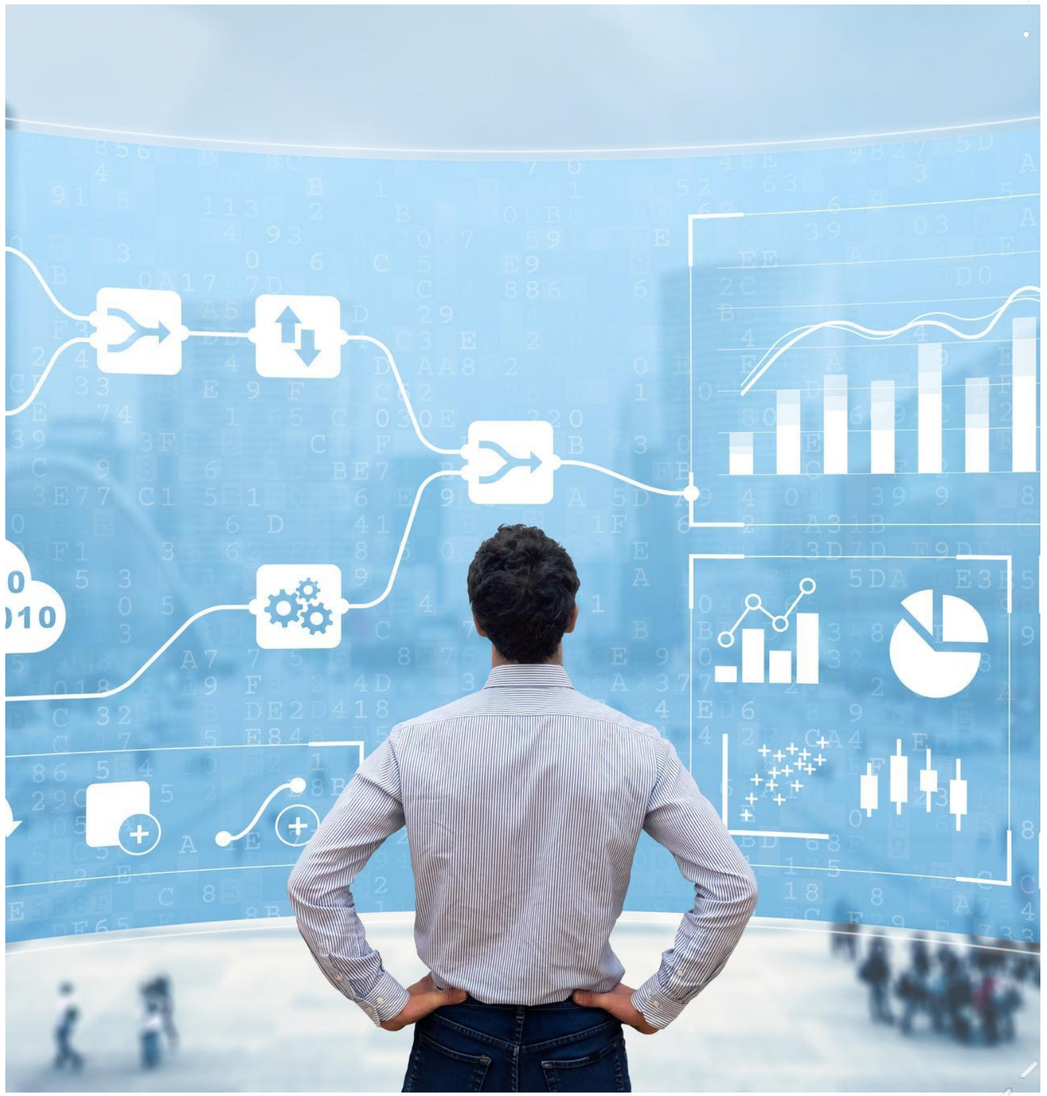
Metrics frameworks

- DORA has shown correlation between capabilities and outcomes
  - Capabilities (technical, process and cultural): agile/devops principles, methods and practices
  - Outcomes: organizational performance through delivery performance and staff's well-being
- "DevEx encompasses how developers feel about, think about, and value their work. [...] An improved developer experience has positive outcomes – and not just on developers; it also helps improve team and organization outcomes." - DevEx In Action



Metrics frameworks

- Both frameworks **highlight the importance of measuring workflows and processes** by collecting data from both, systems and people
- Still, those metrics frameworks are fundamentally **based on sociological studies** done through surveys
- The researches behind those frameworks are **more focus on cloud-first environments** vs software-defined (SW+HW) ones



Data Analytics Journey

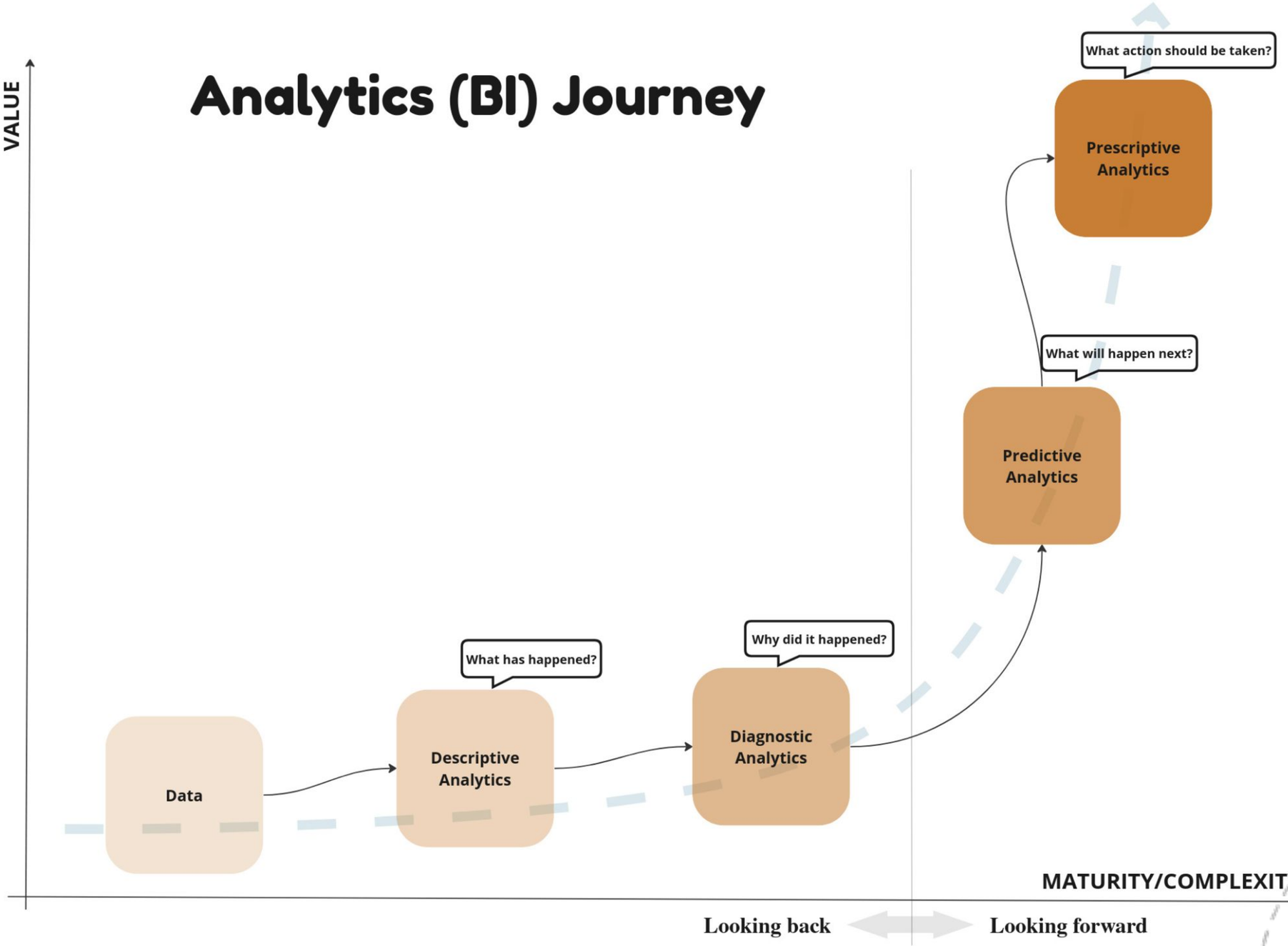
DATA ANALYTICS IS A JOURNEY

As the **complexity** of software-defined products **increases**, the complexity of the production processes also rises.

Scaling up the production of software-defined products requires **continuous improvements processes supported by data analytics**.

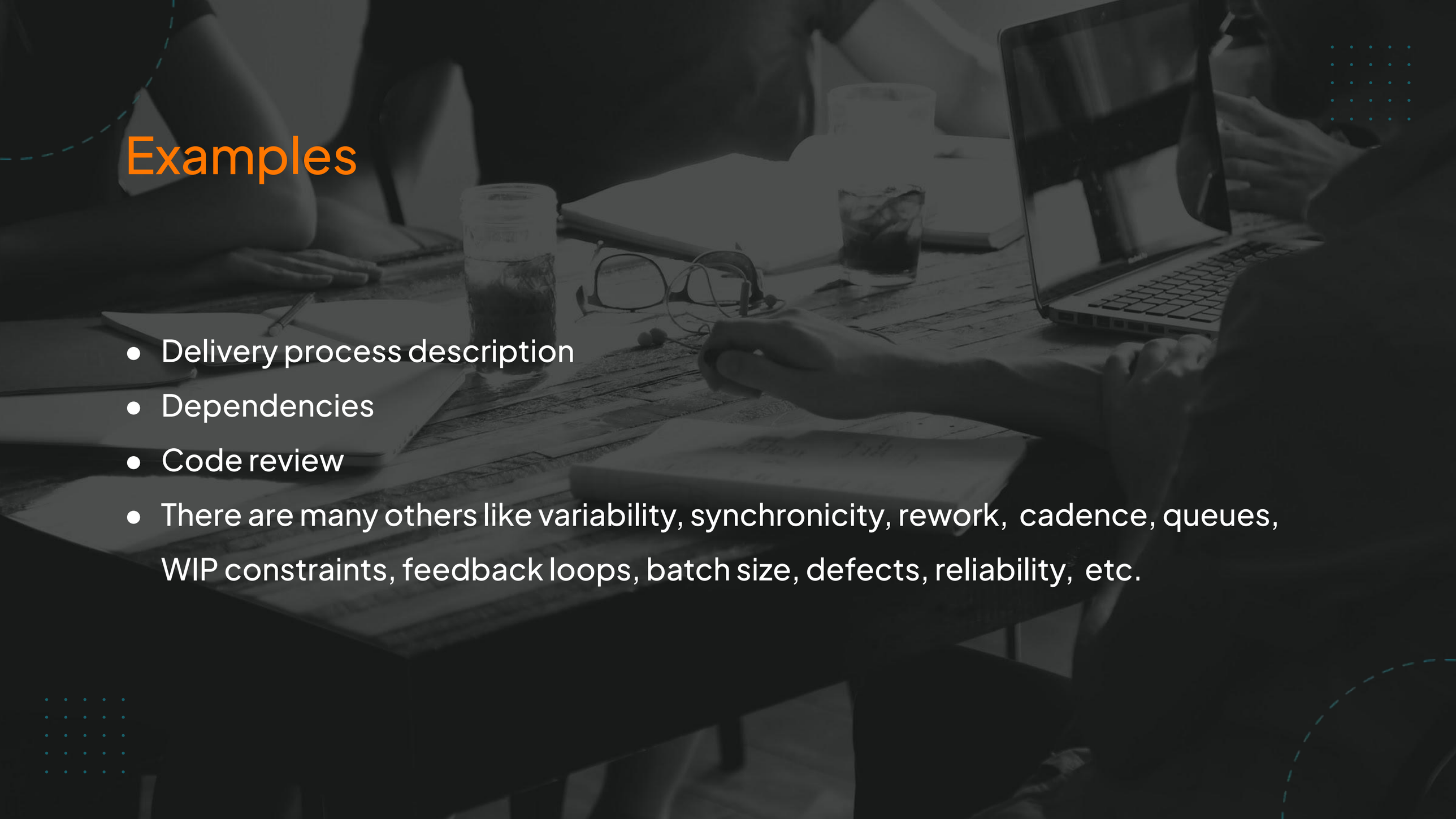
Companies should **own their business intelligence strategy**.

Data Analytics Journey



Examples


The Speakers
Introduction
Frameworks
Data Analytics Journey
Examples
Takeaways



Examples

- Delivery process description
- Dependencies
- Code review
- There are many others like variability, synchronicity, rework, cadence, queues, WIP constraints, feedback loops, batch size, defects, reliability, etc.

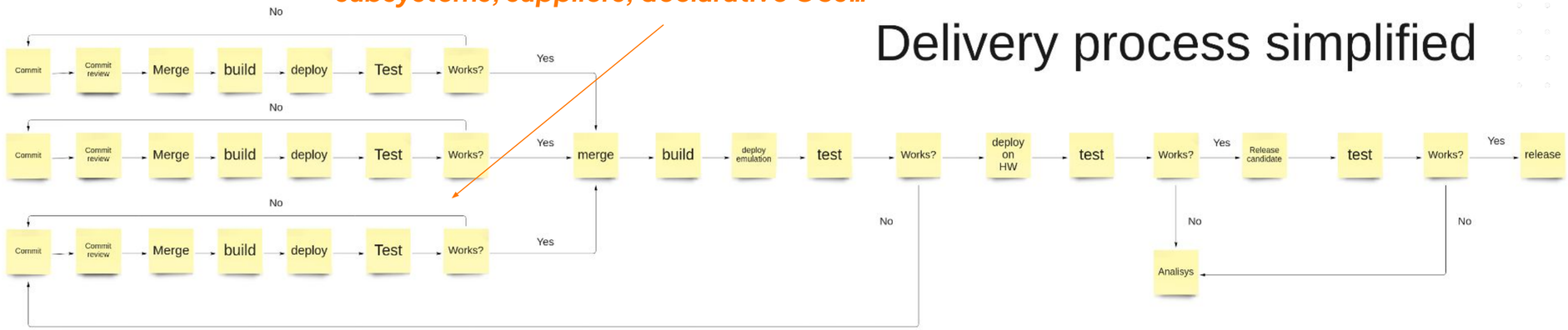
Production environment description

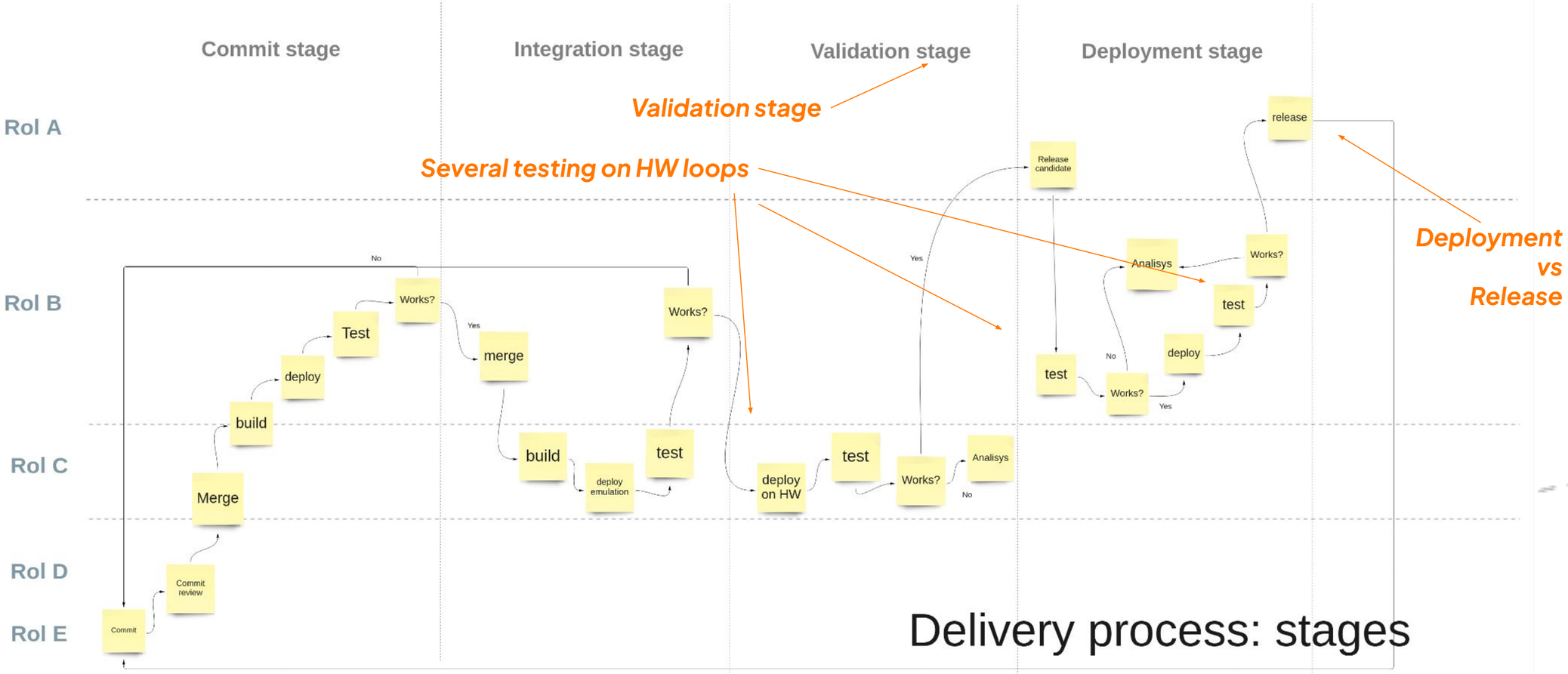
Delivery process

by Bitergia

Delivery process: description

Non-lineal process. Frequently associated to subsystems, suppliers, declarative OSs…

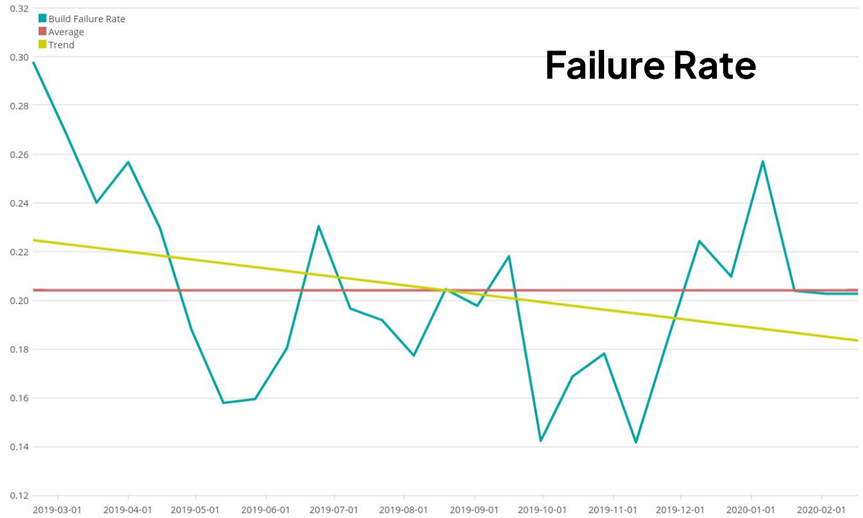




Flow performance metrics

job_name	builtOn	url	result	duration
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	SUCCESS	4,330,591
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	650,823
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	SUCCESS	4,212,126
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	SUCCESS	4,085,666
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	185,419
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	204,299
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	183,675
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	52,151
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	273,435
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	3,140
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	237,864
yardstick-fuel-zte-pod1-daily-master	zte-pod1	+Info	FAILURE	1,748,962

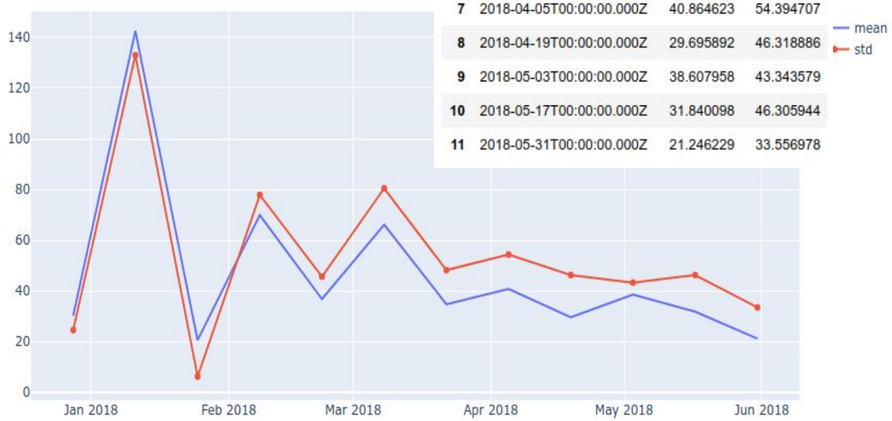
Time Interval



Failure Rate

	time	mean	std
0	2017-12-28T00:00:00.000Z	30.336111	24.618315
1	2018-01-11T00:00:00.000Z	142.517037	132.929116
2	2018-01-25T00:00:00.000Z	20.677083	6.332338
3	2018-02-08T00:00:00.000Z	69.901786	77.877626
4	2018-02-22T00:00:00.000Z	36.842746	45.683325
5	2018-03-08T00:00:00.000Z	66.137681	80.484708
6	2018-03-22T00:00:00.000Z	34.789033	48.248762
7	2018-04-05T00:00:00.000Z	40.864623	54.394707
8	2018-04-19T00:00:00.000Z	29.695892	46.318886
9	2018-05-03T00:00:00.000Z	38.607958	43.343579
10	2018-05-17T00:00:00.000Z	31.840098	46.305944
11	2018-05-31T00:00:00.000Z	21.246229	33.556978

Lead Time



Time to Recover

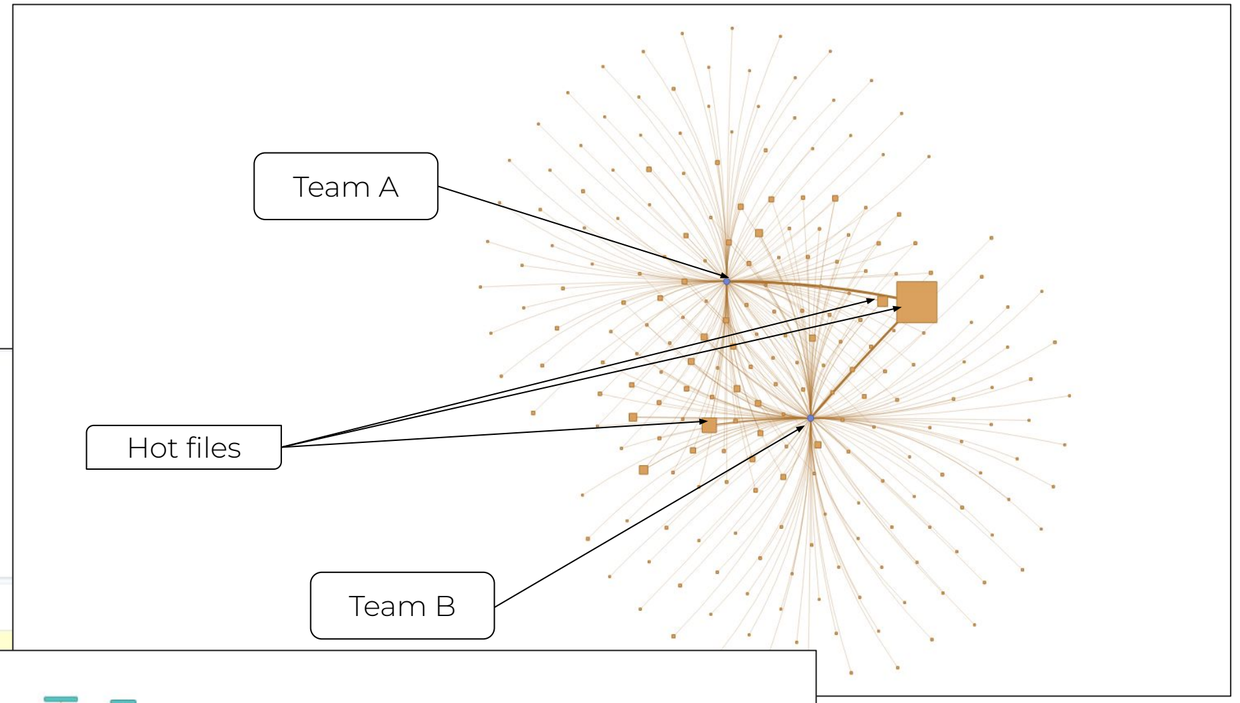
	time	job	failed_at_build	recovered_at_build	failure_date	recovered_at	time_diff	time_diff_hours
0	2017-12-28T00:00:00.000Z	compass-docker-master	13	14	2018-01-03T22:34:18.634000+00:00	2018-01-04T11:30:01.094000+00:00	0 days 12:55:42.460000	12.928333
1	2017-12-28T00:00:00.000Z	dovetail-docker-build-push-master	338	339	2018-01-02T07:41:20.996000+00:00	2018-01-04T07:25:59.662000+00:00	1 days 23:44:38.666000	47.743889
2	2018-01-11T00:00:00.000Z	dovetail-docker-build-arm-push-master	130	131	2018-01-18T16:30:01.165000+00:00	2018-01-24T06:30:01.028000+00:00	5 days 13:59:59.863000	133.999722
3	2018-01-11T00:00:00.000Z	barometer-kafka-docker-build-push-master	1	7	2018-01-11T20:30:00.295000+00:00	2018-01-23T12:00:00.948000+00:00	11 days 15:30:00.653000	279.500000

Diagnosis

Dependencies

by Bitergia

Conway's Law: dependencies

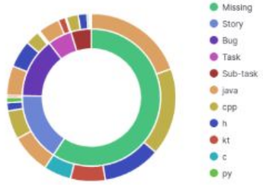


Limitation: at the time this is updated, 1/3 of the commits could not be classified. That means that the number of bugs is *underestimated*.

The tables below show that [redacted] fixes bugs at a higher rate than team [redacted]

More info: The two tables on the left show the type of activity per repository and company. One of them is focused on the number of commits and their link with a bug report. The other one goes a bit deeper, as it focuses on the modifications over files (1..n modifications per commit). The column on the right calculates the diff between the proportion of activity for each team, it is in yellow when the result is the expected one ar[redacted] investing less effort in working on bugs.

Modified file extensions per issue type



File modifications per company

[redacted]	# file updates)	[redacted]	Bugs)	[redacted]	# file updates)	[redacted]	% Bugs)	Diff Activity
	26,074		7		34,309		19	12

Commits per company

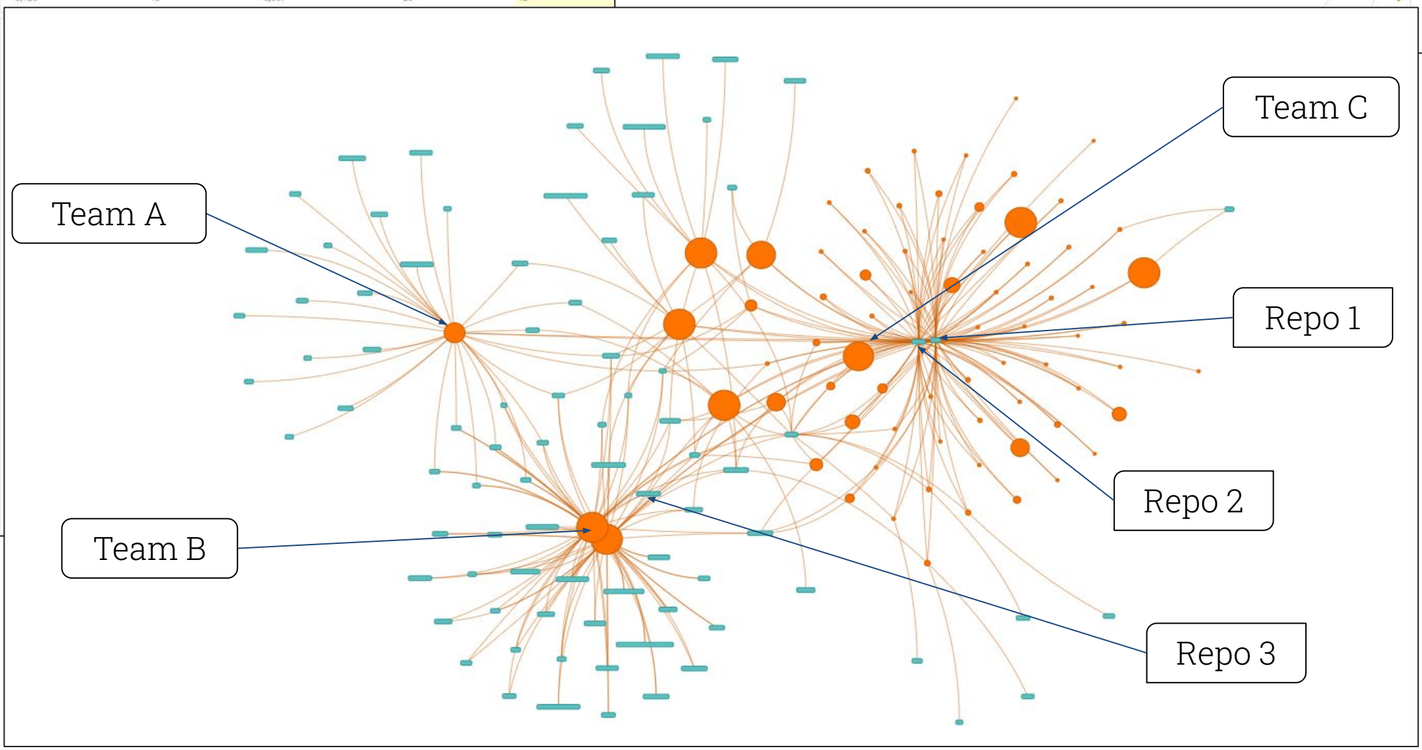
[redacted]	(# commits)	[redacted]	(% Bugs)	[redacted]	# commits)	[redacted]	(% Bugs)	Diff Activity
	3,420		13		3,067		25	12

File modifications per company and repository

Repository	SE (# file updates)	SE (% Bugs)	CN (# file updates)	CN (% Bugs)	Diff Activity
[redacted]	2,042	29	1,312	52	23
	80	23	2,466	11	-12
	2,446	6	0		
	1,523	19	755	57	37
	4	0	1,619	4	4
	795	36	318	10	-26
	168	0	1,418	15	15
	1,391	0	359	8	8
	40	90	0		
	8	13	3,183	3	-10
	447	2	0		
	26	8	342	25	17
	535	1	0		
	921	7	2	0	-7
	0		995	3	
	520	6	0		
	1,207	3	0		
	20	0	5,212	6	6
	1,332	2	0		
	431	9	26	46	37

Export: Raw Formatted

1 2 3 4 5 … 6 »

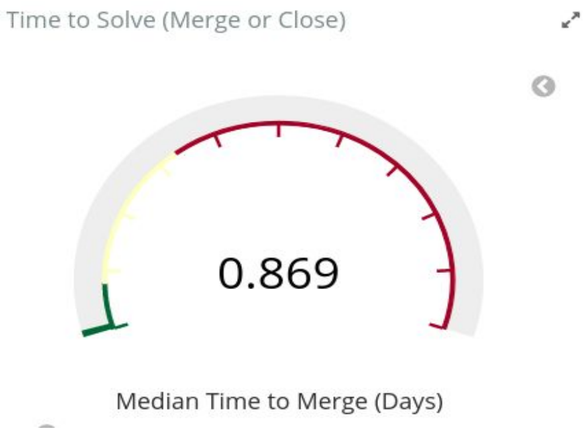


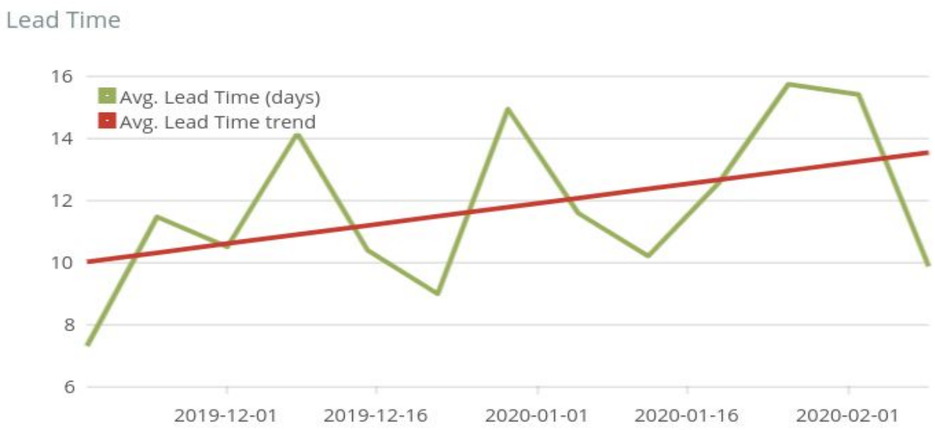
Diagnosis

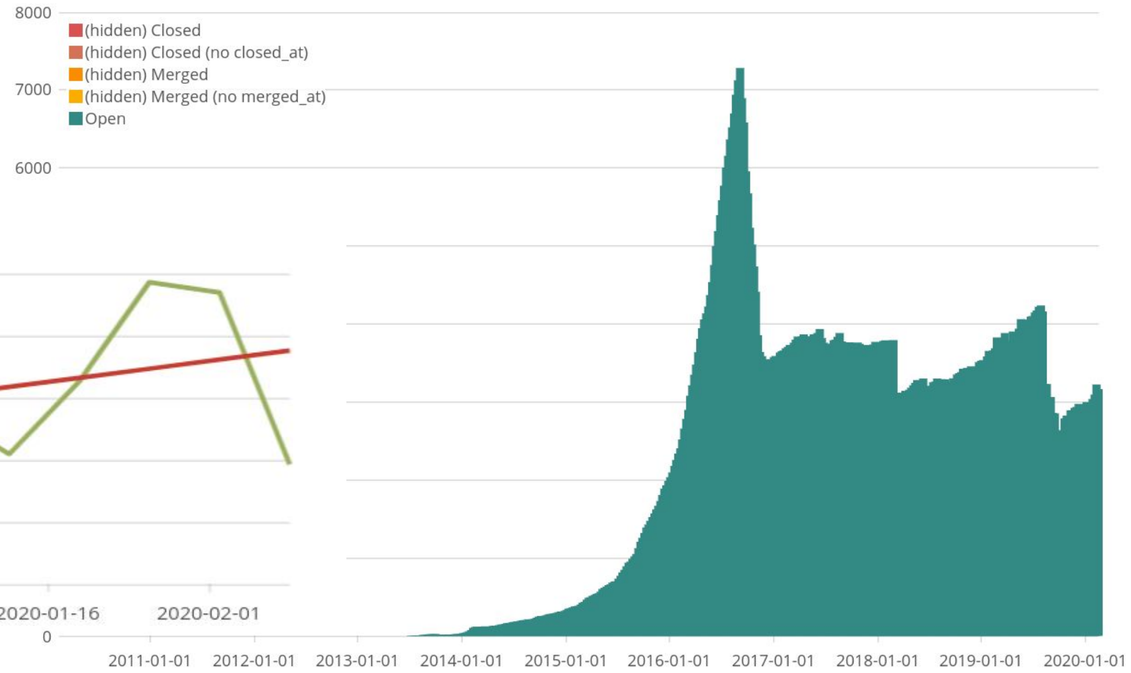
Code review

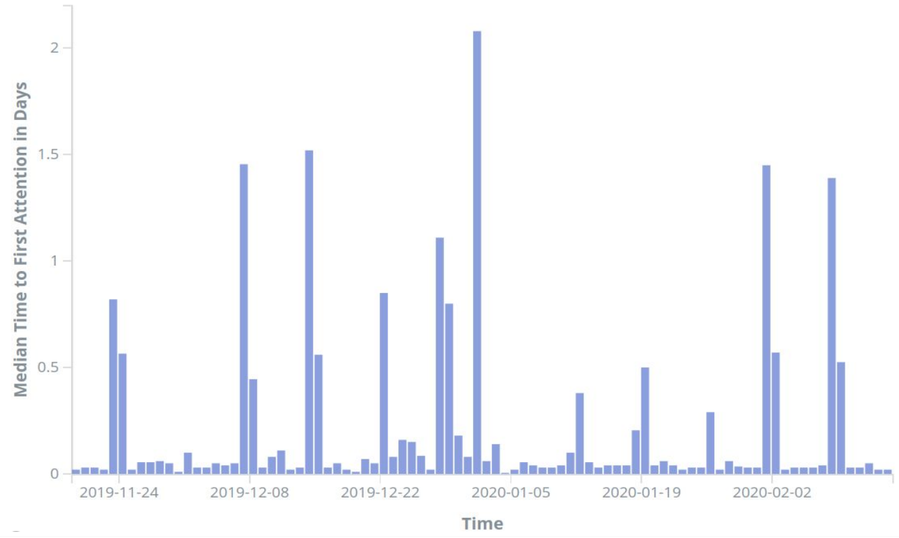
by Bitergia

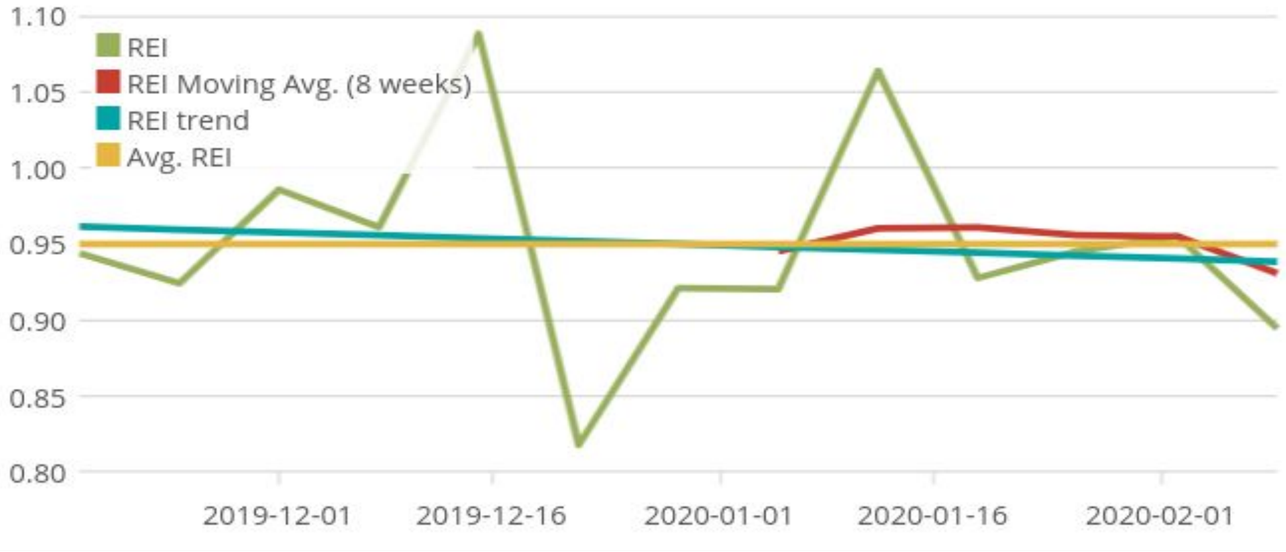
Code Review


Time to Solve (Merge or Close)
0.869
Median Time to Merge (Days)


Lead Time
Avg. Lead Time (days)
Avg. Lead Time trend
16
14
12
10
8
6
2019-12-01
2019-12-16
2020-01-01
2020-01-16
2020-02-01

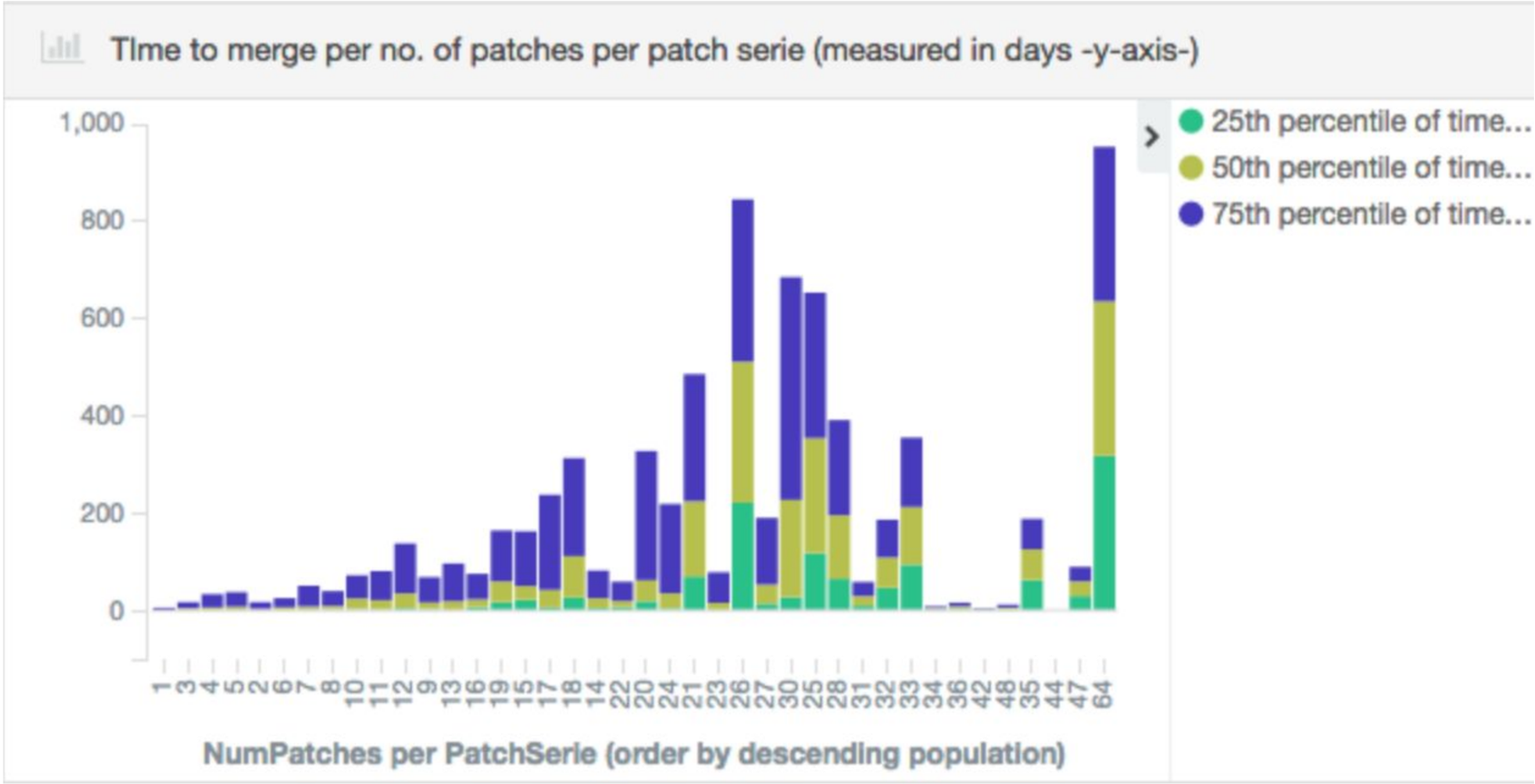

8000
7000
6000
(hidden) Closed
(hidden) Closed (no closed_at)
(hidden) Merged
(hidden) Merged (no merged_at)
Open
0
2011-01-01
2012-01-01
2013-01-01
2014-01-01
2015-01-01
2016-01-01
2017-01-01
2018-01-01
2019-01-01
2020-01-01


Median Time to First Attention in Days
2
1.5
1
0.5
0
2019-11-24
2019-12-08
2019-12-22
2020-01-05
2020-01-19
2020-02-02
Time


1.10
1.05
1.00
0.95
0.90
0.85
0.80
REI
REI Moving Avg. (8 weeks)
REI trend
Avg. REI
2019-12-01
2019-12-16
2020-01-01
2020-01-16
2020-02-01

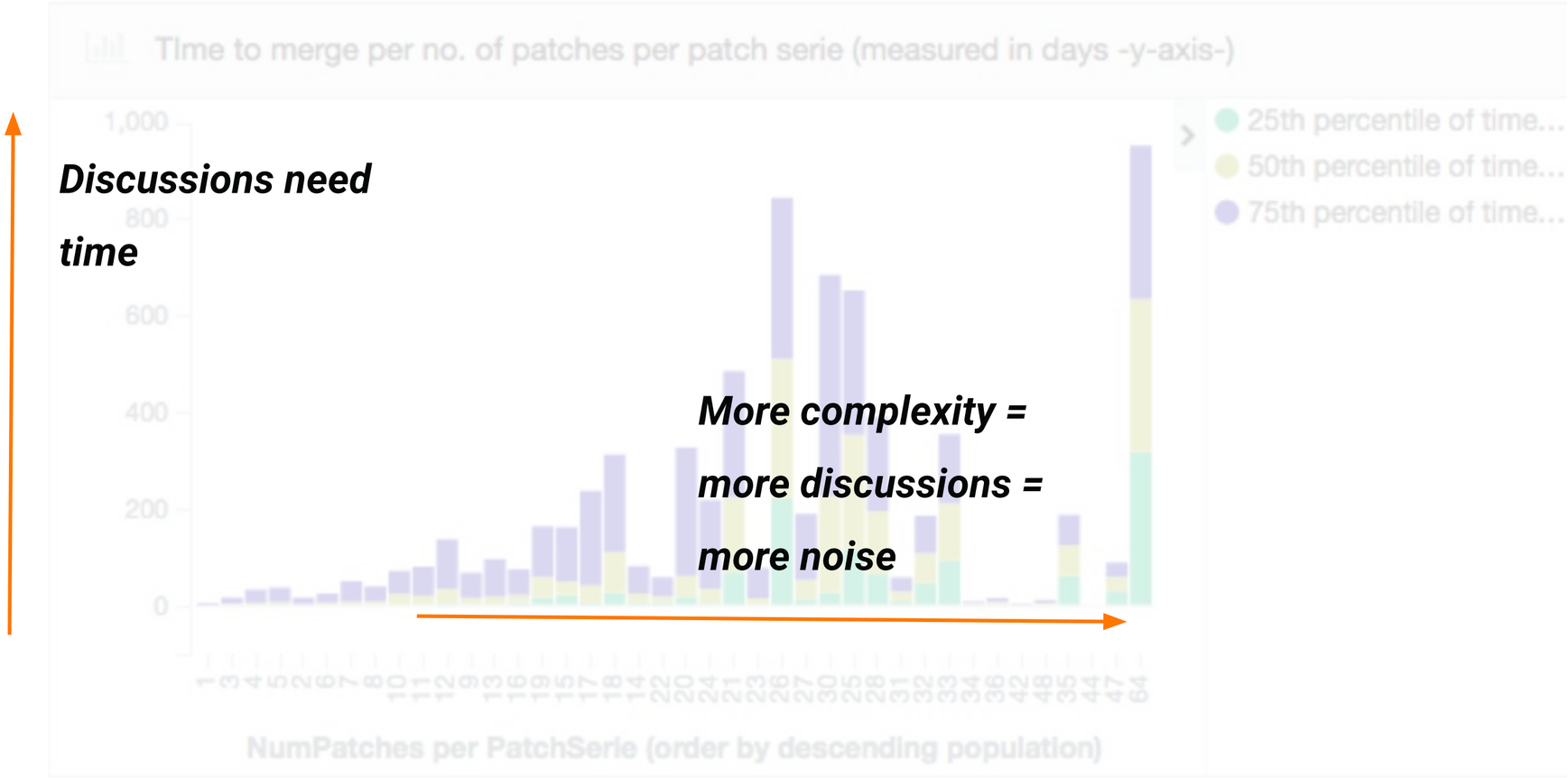
Code Review

D. Izquierdo, J. M. Gonzalez-Barahona, L. Kurth and G. Robles, "**Software Development Analytics for Xen: Why and How**," in *IEEE Software*, vol. 36, no. 3, pp. 28-32, May-June 2019, doi: 10.1109/MS.2018.290101357



Code Review

D. Izquierdo, J. M. Gonzalez-Barahona, L. Kurth and G. Robles, "**Software Development Analytics for Xen: Why and How**," in *IEEE Software*, vol. 36, no. 3, pp. 28-32, May-June 2019, doi: 10.1109/MS.2018.290101357

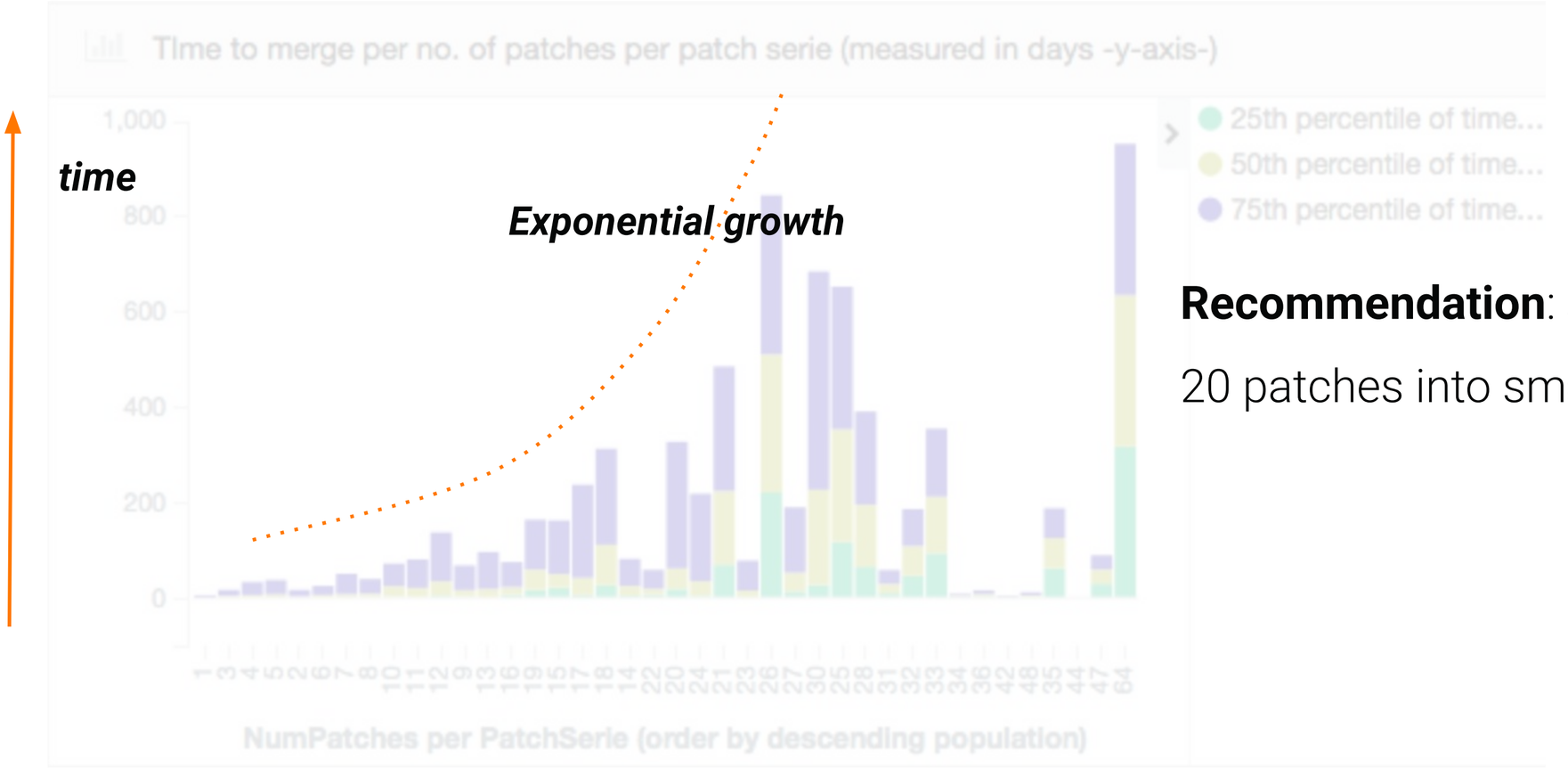


Discussions need time

More complexity = more discussions = more noise

Code Review

D. Izquierdo, J. M. Gonzalez-Barahona, L. Kurth and G. Robles, "**Software Development Analytics for Xen: Why and How**," in *IEEE Software*, vol. 36, no. 3, pp. 28-32, May-June 2019, doi: 10.1109/MS.2018.290101357



Recommendation: split patch series > 20 patches into smaller patch series

Takeaways


The Speakers
Introduction
Frameworks
Data Analytics Journey
Examples
Takeaways

Takeaways

01 **Data analytics** will become fundamental to support improvement processes at scale in complex software production environments.

02 Current frameworks provide a great starting point. They should be **complemented with data analytics and expanded** to additional environments.

03 Introducing data analytics in production environments is a **journey**: from collecting and visualising data to data analytics, to business intelligence.

04 There are plenty of examples where **data analytics can help in embedded/automotive environments** to detect constraints, discern symptoms from root causes, diagnose root causes, evaluate the global impact of local improvements or remedies, etc..

05 From design to maintenance and from business requirements to code: advanced data analytics will **support you in getting a holistic view of the entire production environment**.

When a measure becomes a target, it ceases to be a good measure.- *Goodhart's Law*

References

- Delivery Performance Analytics by Bitergia
- Introducing Delivery Performance Analytics by Agustin B.B.
- Code Reviews. DORA Community Discussions
- Improve your software product delivery process performance using metrics [1][2]
- Bitegia Analytics Platform (BAP). Code.
- CHAOSS project (Bitergia is co-founder)

Linaro Connect
MADRID 2024 | MAY 12-17 2024

Thank you

Linaro Connect
MADRID 2024 | MAY 14-17 2024

Improving Delivery Process Performance for SW-Defined Products

Helped by Data Analytics