



Agile Data Science

**How To Implement Agile Workflows
For Analytics & Machine Learning**

@John_Sandall

Saturday 30th October 2021



Objective

◆ To make this the *most valuable talk you watch at PyData Global 2021.*

Metric: Your team's productivity in 2022. 

(ツ)/ What's the big deal?

- ◆ 71% of organisations using Agile in 2018¹
- ◆ 98% of companies benefited from adopting Agile²
- ◆ 60% of companies saw revenue & profit growth after using Agile³
- ◆ Agile projects are **28% more successful** than traditional projects⁴

¹ [*Pulse of the Profession*](#) – Project Management Institute

² [*11th Annual State of Agile Report*](#) – VersionOne

³ [*Embracing Agile*](#) – Harvard Business Review

⁴ [*Agile Project Delivery Confidence*](#) – PwC



```
outline = {  
    0: "Motivations",  
    1: "What is Agile?",  
    2: "The Meetings",  
    3: "Planning & Estimation",  
    4: "Practical Tips",  
    5: "Tools of the Trade",  
    6: "Advice for Leaders",  
    7: "Next Steps",  
}
```

Motivations

Part 00

"How long will it take?"

(*"Ballpark..."*)

A recipe for unhappy bunnies

- ▶ When the date hits and you're **not ready...**
- ▶ ...**client is upset** because they made promises too
- ▶ ...**team is upset** because they worked overtime
- ▶ ...the way we're **working is broken.**



People are bad at estimating

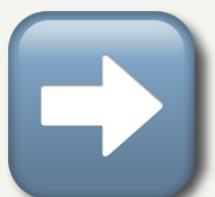
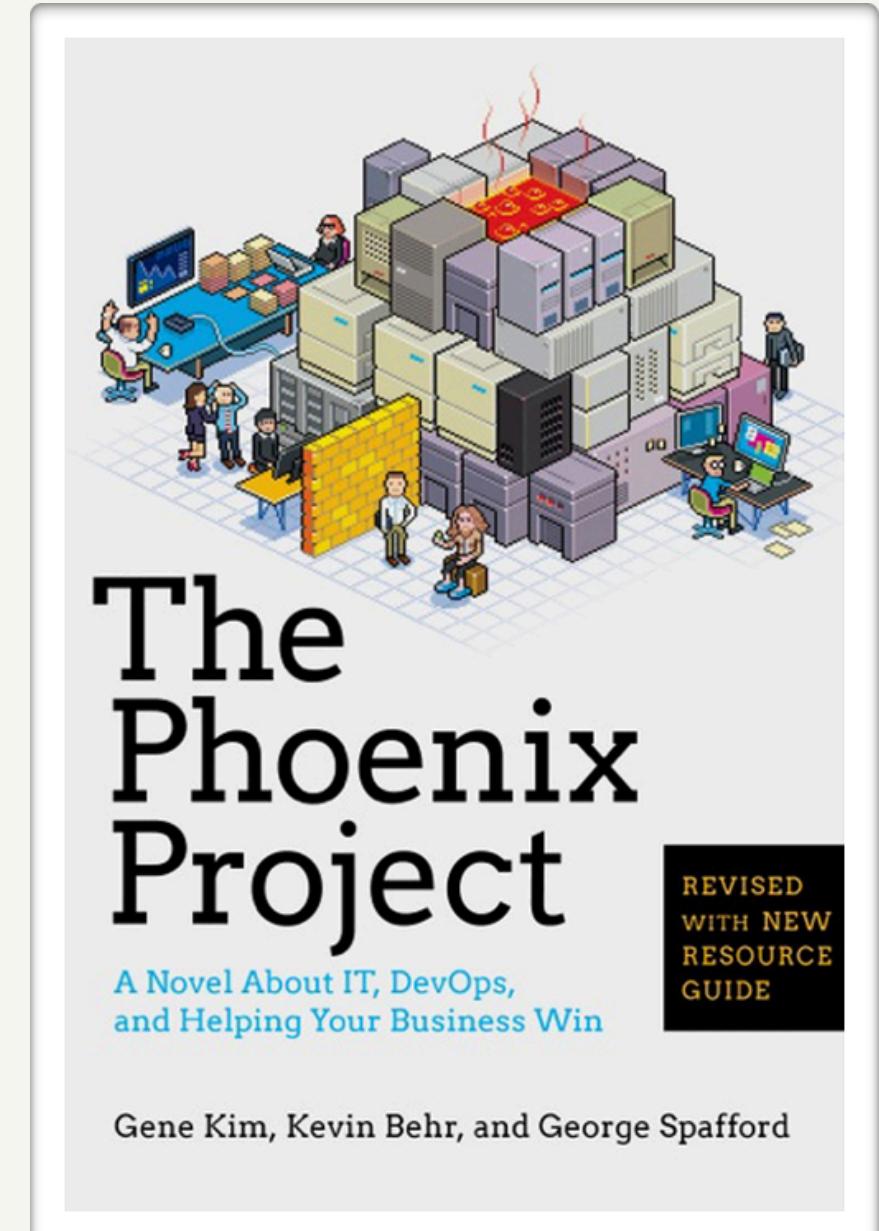
Illustrative examples

- ▶ Sydney Opera House was estimated to take **four years** and cost **\$7m**
- ▶ It took **14 years** and cost **\$102m**
- ▶ Also, **Wembley Stadium, Denver International Airport, Eurofighter Typhoon**
(£12 billion overbudget)



My journey into agile

- ▶ 6 years as CEO & Principal Data Scientist at Coefficient
- ▶ Previously: Lead Data Scientist for mobile ticketing app YPlan
- ▶ 2020: Read a book
- ▶ 2021: Managing three data science/analytics/engineering teams + up to 3 workshops per day + side projects + company
- ▶ Today: everything I wish I'd known 12 months ago



bit.ly/pydata-agile

What is Agile?

Part 01

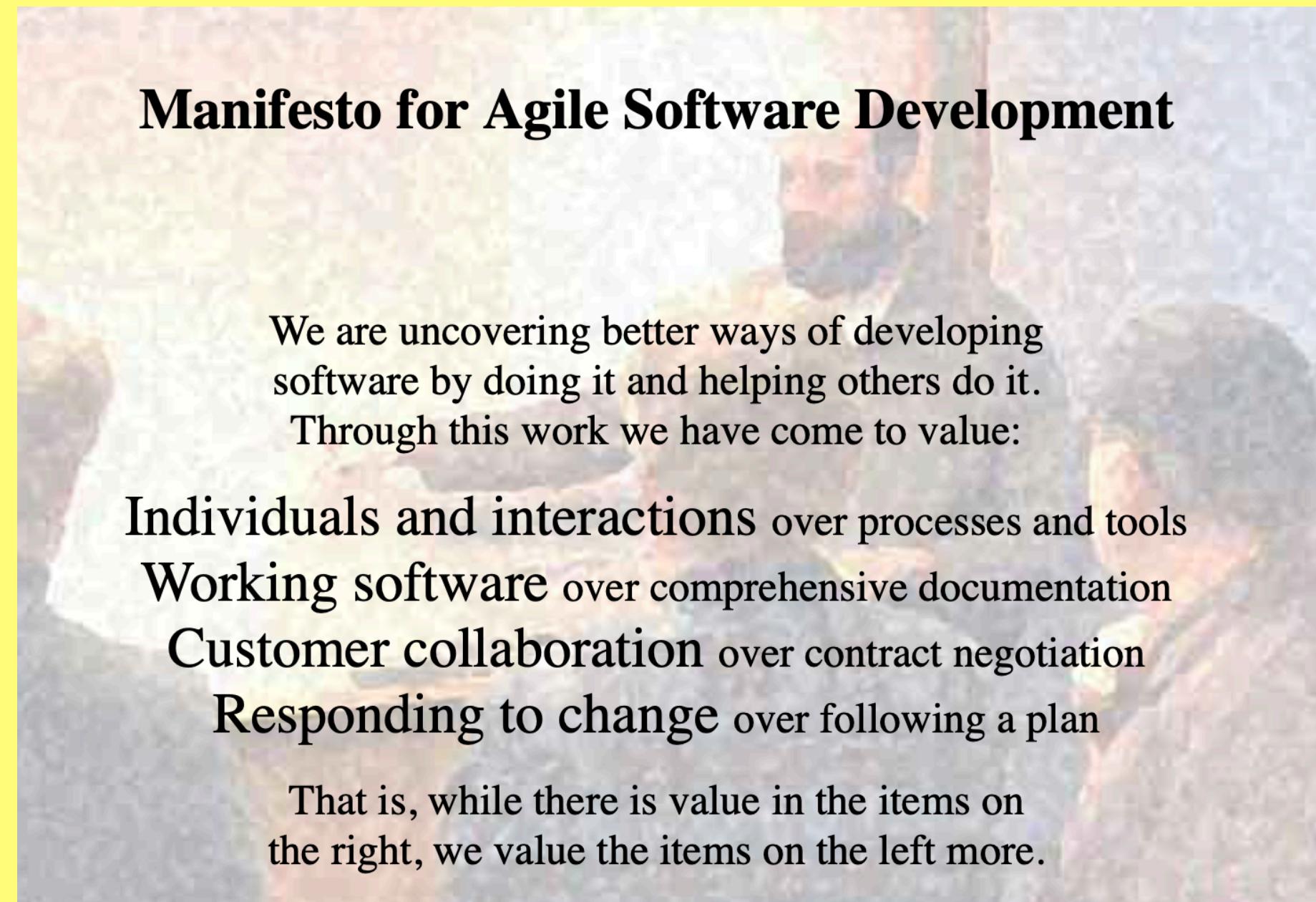
~~Terminology~~ JARGON

- ▶ **Agile:** a system to manage work using **time limited iterations**
- ▶ **Sprint:** 1 to 4 week cycles delivering "working increments"
- ▶ **Working increment:** output that can be tested / reviewed
- ▶ **Story:** single unit of work usually completed within a sprint
- ▶ **Epic:** collection of related stories

*"This **sprint** we aim to deliver four **stories** in the 'Recommender v1' **Epic**"*

Myth #1:
Agile is a methodology

Agile is a set of 4 values + 12 principles

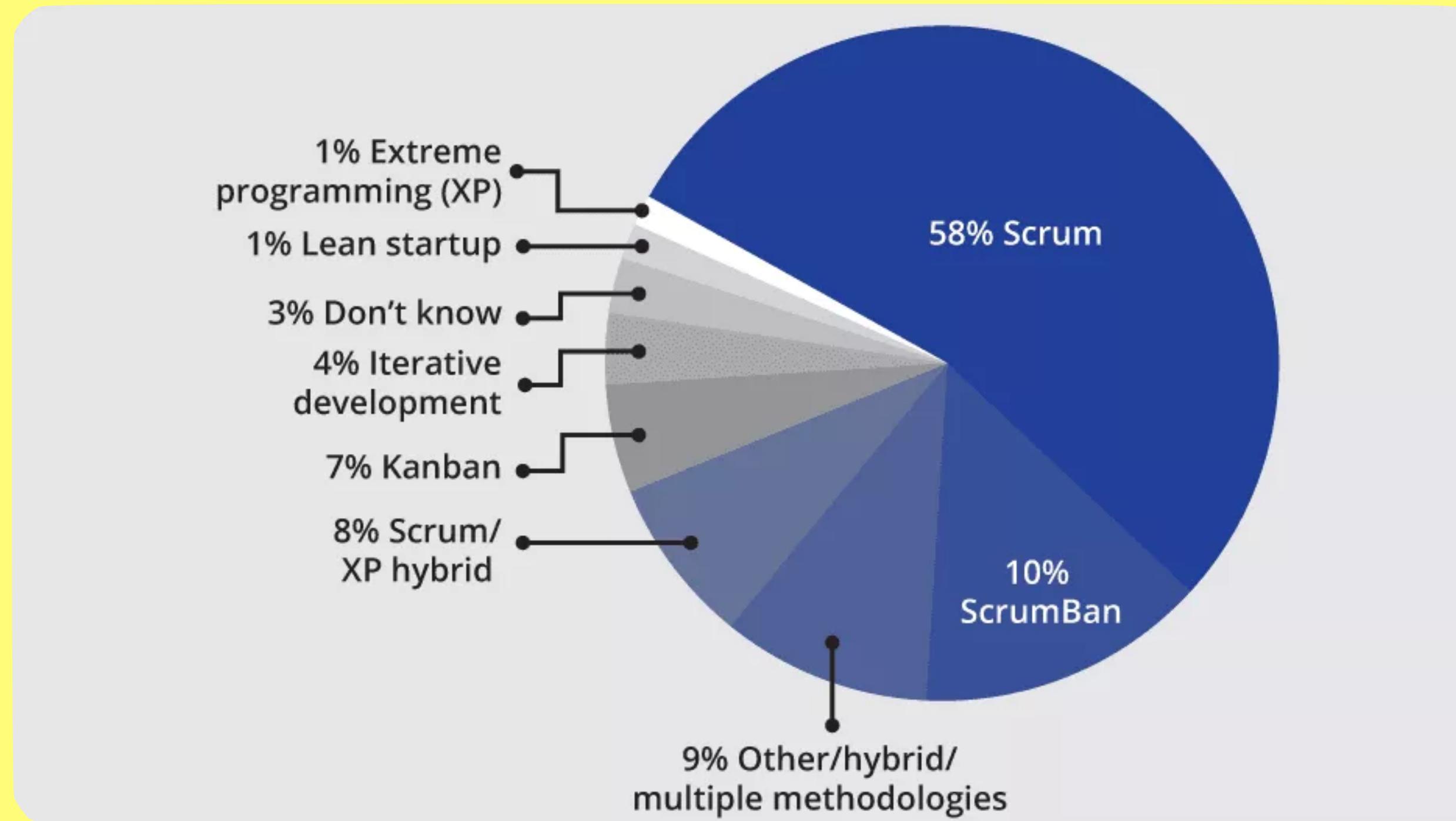


It's a philosophy, not a methodology

Myth #2:

Agile is Scrum

Scrum, Kanban, Lean, XP are all "Agile"



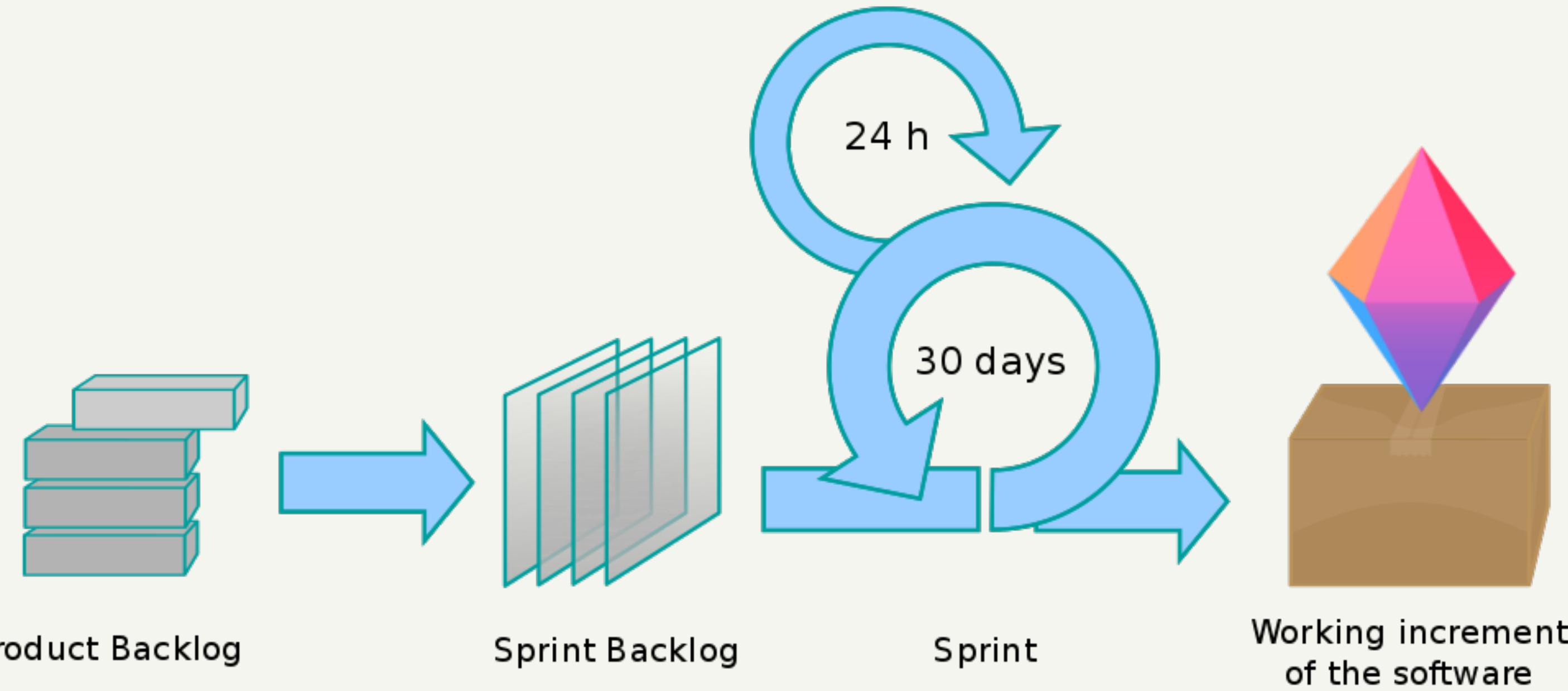
"We're agile-ish..."

~~Terminology~~ JARGON

- ▶ **Lean:** iterative, evolutionary, Toyota Production System
- ▶ **Kanban:** visualise a limited set of stories on a workflow board
- ▶ **Scrum:** step-by-step system following Agile principles
- ▶ **Backlog:** list of potential stories for consideration
- ▶ **Standup:** short daily team meeting to check progress

"Scrum Standups can be used to review each Story on the Kanban board"

The Scrum Lifecycle



Epic: Deploy model

Stories

- Create FastAPI app
- Version data using DVC
- Deploy to AWS Lambda

The Meetings

Part 02

Myth #3:

Agile is undisciplined

Agile does not mean "nimble" or "disorderly"!



Meeting #1: Sprint Planning



Start of sprint. 30 minutes.

Agenda

- ▶ Resource availability
- ▶ Sprint goal
- ▶ What's on the cards for this sprint?
- ▶ Effort Points / Value Points
- ▶ What's in the next sprint?

Availability: Alice 10 days + Bob 2 days (🏖️)

Goal: Deploy ML model

Write up stories. Move them into the sprint.

Estimate effort & value for each story.

De-scope items to match predicted velocity.

Meeting #2: Sprint Lookahead



Mid-sprint. Every other week. 30 minutes.

Agenda

- ▶ Review backlog as it stands
- ▶ What to do **in the next sprint?**
- ▶ Do we have all the tickets we need?



<https://slab.com/blog/eisenhower-matrix/>

*"Especially whenever our affairs seem to be in crisis, we are almost compelled to give our **first attention to the urgent present** rather than to the **important future**."*

– Dwight D. Eisenhower, 1961 address to the Century Association

Meeting #3: Sprint Review / Showcase



End of sprint. 30 minutes.

Agenda

- ▶ Review what's been delivered
- ▶ What's done? What's not done?
- ▶ Everything has a "Definition of done"?
- ▶ Celebrate the wins! 🎉

Meeting #4: Retrospective



After every second sprint. No exceptions.

30 minutes.

Agenda

- ▶ What went well?
- ▶ What did not go well?
- ▶ Improvements?

*"Regardless of what we discover, we understand and truly believe that **everyone did the best job they could**, given what they knew at the time, their skills and abilities, the resources available, and the situation at hand."*

– Norm Kerth

<https://www.madetech.com/blog/how-to-run-a-successful-retrospective/>

<https://trello.com/b/40BwQg57/retrospective-techniques-for-coaches-scrum-masters-and-other-facilitators>

<https://github.com/CarolineMouden/BlackCardsForRetroAgainstHumanity>

Planning & Estimation

Part 03

Myth #4:

Agile can't meet deadlines

Agile is **MORE** likely to deliver to schedule

Agile projects...

- ▶ ...are **less likely to fail**
- ▶ ...have **greater speed to market**
- ▶ ...higher **quality code**
- ▶ ...have more **motivated & productive teams**

Fixed sprint cycles + explicit estimation + review + retro

= teams that become **better** at forecasting how long tasks will take

Estimation in the face of uncertainty

People are **great** at relative estimates

Epic	Story
Deploy model	Create FastAPI app
Deploy model	Version data using DVC
Deploy model	Deploy to AWS Lambda
Deploy model	Setup remote logging
Improve model	Try HistGradientBoostingClassifier

Estimation in the face of uncertainty

People are **great** at relative estimates

1. Developers estimate **effort points**

- Biggest task = 5 (or "XL")
- Smallest task = 1 (or "S")
- **Effort points are not time-related**

2. Stakeholders estimate **value points**

- Most valuable at the top, least at the bottom, everything else in between
- "Fibonacci system" or "t-shirt sizing"

Story	Effort Points	Value Points
Create FastAPI app	3	5
Version data using DVC	2	1
Deploy to AWS Lambda	5	5
Setup remote logging	3	1
Try HistGradientBoostingClassifier	1	2

Small	Medium	Large	XL	XXL	XXXL
1	2	3	5	8	13

Estimation in the face of uncertainty

People are great at relative estimates



3. **Value per Effort Score** = VP / EP

4. **Identify dependencies/blockers** and **task groups**

5. **Order** by V/E score

6. **Sprint #1** = ^_(ツ)_/^-

- Review at end, calculate EP completed
- 55 EP in 2 weeks = 5.5/day

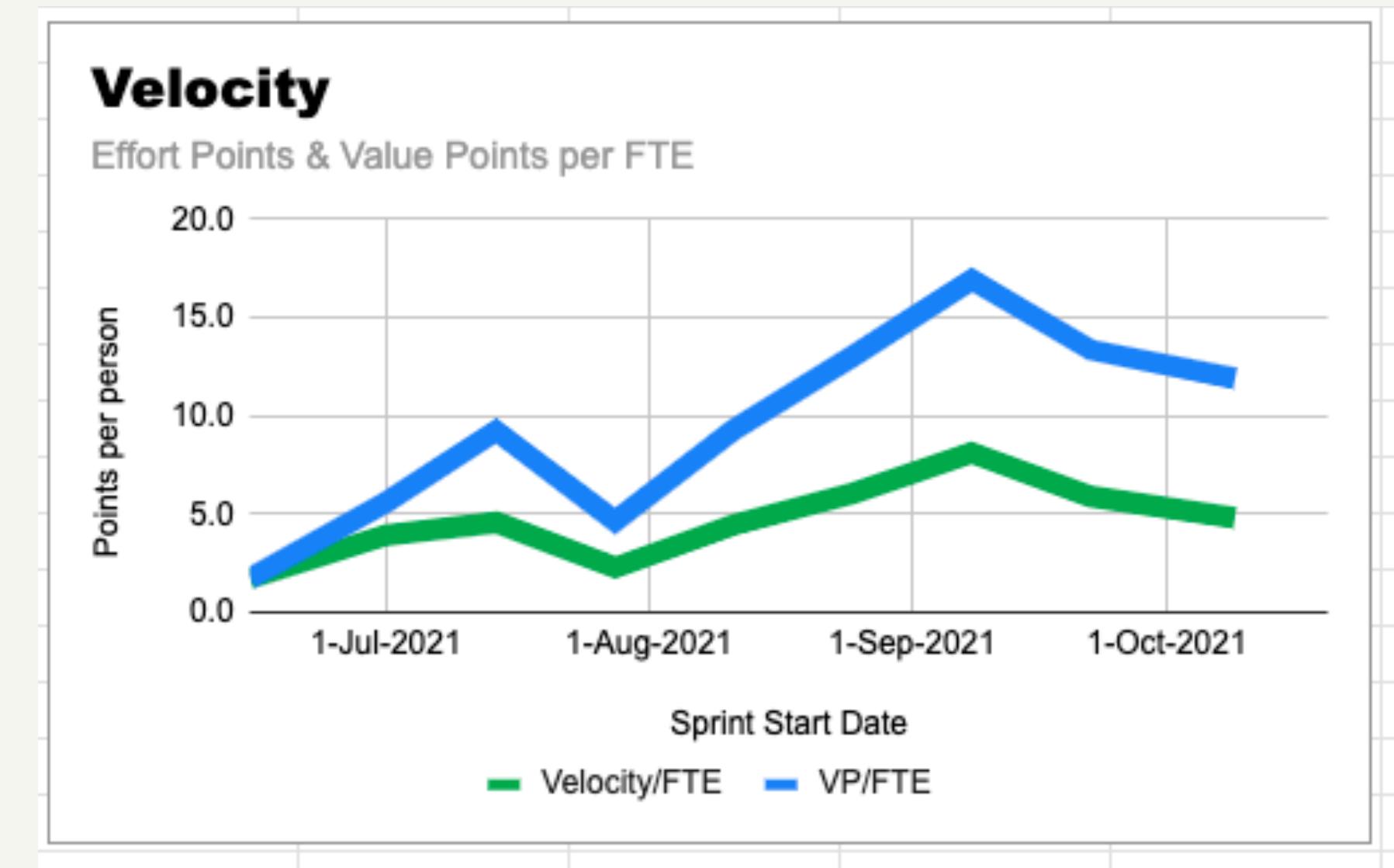
Story	Effort Points	Value Points	Score
Try HistGradientBoostingClassifier	1	2	2.0
Create FastAPI app	3	5	1.7
Deploy to AWS Lambda	5	5	1.0
Version data using DVC	2	1	0.5
Setup remote logging	3	1	0.3

Small	Medium	Large	XL	XXL	XXXL
1	2	3	5	8	13

Estimation in the face of uncertainty

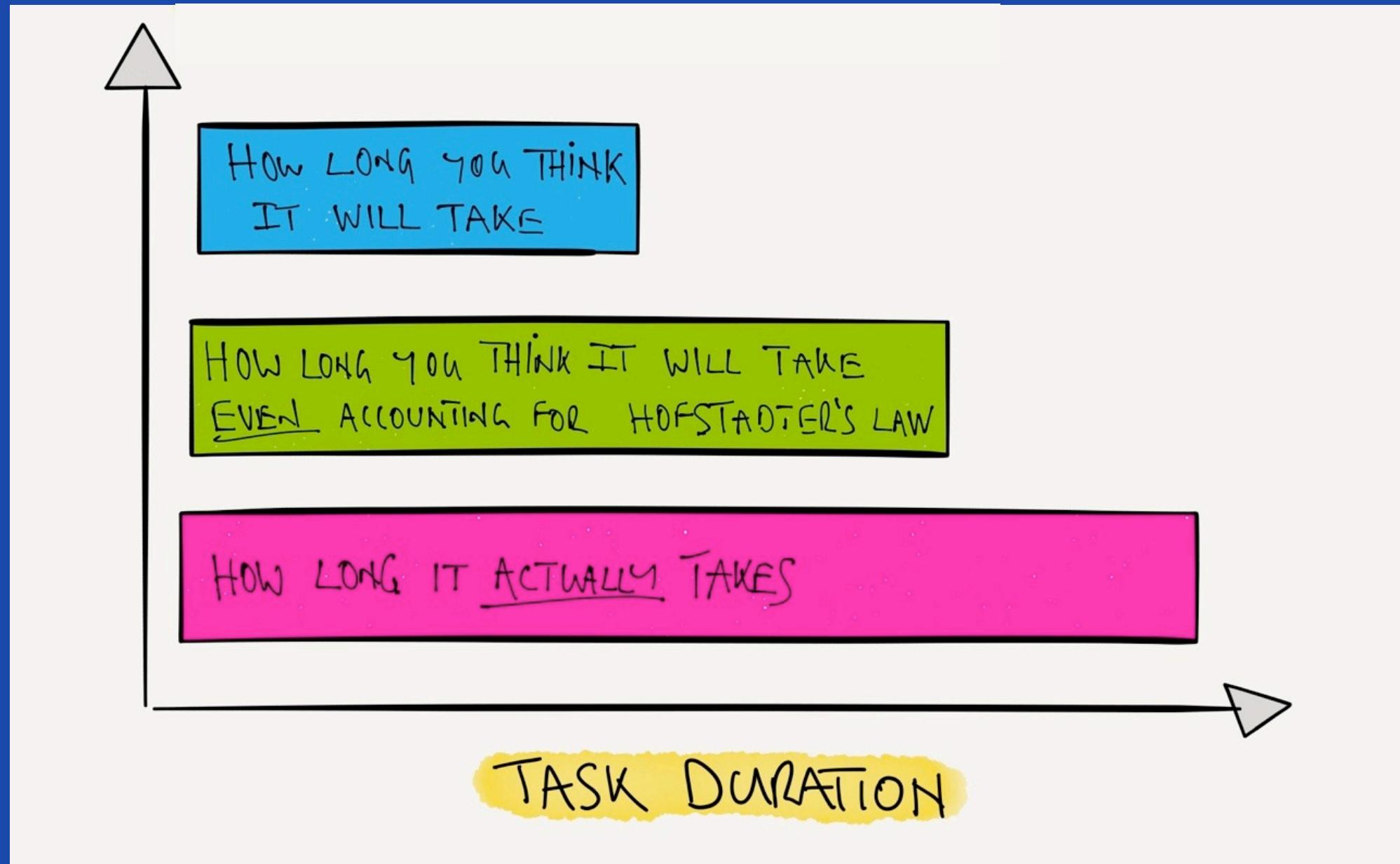
People are **great** at relative estimates

3. **Value per Effort Score** = VP / EP
4. **Identify dependencies/blockers** and **task groups**
5. **Order** by V/E score
6. **Sprint #1** = `-_(ツ)_/-`
 - Review at end, calculate EP completed
 - 55 EP in 2 weeks = 5.5/day
7. **Sprint #2** = plan assuming this velocity



Hofstadter's Law

*It always takes longer than you expect,
even when you take into account Hofstadter's Law.*



Practical Tips

Part 04

Practical tips for success

- ▶ **Stories can rollover into the next sprint**
 - ▶ **Bad:** Trying to finish everything in time. Mistakes & burnout likely.
 - ▶ **OK:** Discuss why underestimated, revise EP, learn from it, claim the points next sprint.
- ▶ **Daily Standup:** Make it about the items, not the people. Consider other times.
- ▶ Use "**spikes**" (fixed time-boxed windows) to investigate stories that cannot be estimated.
- ▶ **Never skip retro.**
- ▶ **Quarterly Epics.** Aligns with OKR cycle. Conscious Story rollover. Re-prioritise tech debt.
- ▶ **Evolve your system through retros.** Your team will find innovate workflow optimisations.
- ▶ **Don't try to replicate someone else's agile.**

Tools of the Trade

Part 05

Tools of the trade

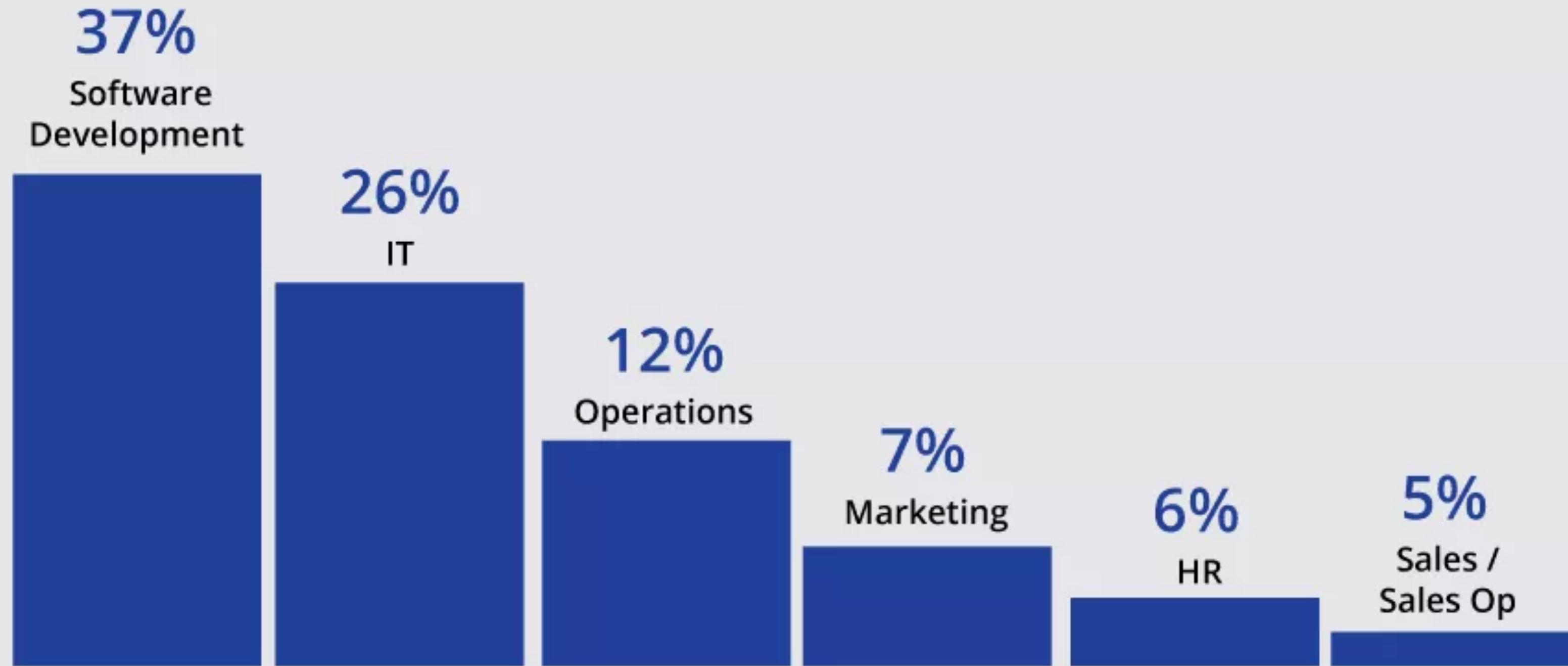
- ▶ **Gitlab, Jira, GitHub Projects, Zenhub:** Project management, code review, comms
- ▶ **Spreadsheets:** Planning, Value/Effort Scoring, Velocity calculations
- ▶ **Monday.com** for Agile Ops
- ▶ **Range** for async standups, meeting automation, OKRs
- ▶ **Metro Retro** for retrospectives

Sprint:	3.3	 Gitlab export link	Total Tasks Missing?	T-Shirt Sizing Guide:	Small	Medium	Large		
Start:	11-Aug			1	2	3			
End:	7-Sep		10						
<hr/>									
Issue	URL	Epic/Story	Title	Status	Assignee	Milestone	Story/Effort Points	Value Points	BTFB
584	https://gitlab.com	Deploy model	Create FastAPI app	Open	John	3.3	1	5	5.0
741	https://gitlab.com	Deploy model	Version data using DVC	Open	Jamie	3.3	3	13	4.3
719	https://gitlab.com	Deploy model	Deploy to AWS Lambda	Open	John	3.3	2	8	4.0
782	https://gitlab.com	Deploy model	Setup remote logging	Open	Luke	3.3	1	3	3.0
764	https://gitlab.com	Infrastructure > #728: Releases	Release 3.2	Closed		3.3	1	3	3.0

Myth #5:

Agile is for software teams

Agile Ops & #AgileMarketing increasingly popular



Agile Ops at Coefficient using Monday.com

Sprint 11 (3rd November - 16th November)

Sprint 10 (20th October - 8th November)

	ID	Assignee	Epic	↳ Epic Notes	Sprint	Status	🔔 Due Date	Effort Points	Value Points	BFTB
Add some nice formatting to the FAQ page	1809358610	JS	BNPP Training		Sprint 10	Done		2	2	1
Sort out PyData tickets for G&L	1832945200	JS	L&D		Sprint 10	Done		2	2	1
Add Tide bank connection to Dext	1832954843	JS	Bookkeeping		Sprint 10	Done		1	1	1
Sort out Typeform refund	1833120295	JS	Bookkeeping		Sprint 10	Done		1	1	1
Research BS7858:2019 security vetting providers	1834299572	JS HB	Legal		Sprint 10	Done		2	2	1
Re-engage Apple Business Team	1807524879	JS	WFH (hardware)		Sprint 10	In Progress		2	3	1.5
Get MalwareBytes	1807651468	JS	Accreditations		Sprint 10	Done		2	3	1.5

Sprint Planning | Sprint Capacity | Online Documentation | Burndown Chart | Current Focus | Tab... | Velocity Chart | Harriet's Table | More | + Add View | Integrate

Filter / 1

Blocked / 2

Analyse October monthly feedback survey

Assignee: JS

Sprint: Sprint 10

Backfill SDR spreadsheet as much as

Planning / 2

Mullvad VPN for the team.

Assignee: JS

Sprint: Sprint 10

Quizzes + recap course sales opportunity with BNPP

Shovel Ready / 13

John/Luke plan next social

Assignee: JS

Sprint: Sprint 10

Start thinking seriously about booking co-working spaces

In Progress / 11

September invoices

Assignee: JS

Sprint: Sprint 10

Re-engage Apple Business Team

Review / 1

Add all AppleCare+ policies to GDrive

Assignee: HB

Sprint: Sprint 10

Done / 15

TGIF email (Oct 22nd, 2021)

Assignee: HB

Sprint: Sprint 10

Add some nice formatting to the FAQ page

Advice for Leaders

Part 06

Myth #6:

Agile is for startups

Case Study: Agile at John Deere

The Enterprise Advanced Marketing team at John Deere implemented agile techniques. They saw:

- ▶ **75% reduction in innovation project cycle times** — time take to develop a working prototype of a new machine reduced from 18-36 months to 8 months.
- ▶ **Dramatic improvement in team engagement and happiness** — from bottom third companywide to top third.
- ▶ **Velocity ("amount of work accomplished in each sprint") increase of 200% on average** with some teams **increasing by 400%** and one team **by 800%**
- ▶ This wasn't "quantity over quality" – **quality also measurably improved.**

Case Study: Agile in the UK Government

The screenshot shows the GOV.UK Service Manual website. At the top, there's a black header bar with the GOV.UK logo, a search icon, and navigation links for Topics, Departments, and Government activity. Below this is a blue horizontal bar with a 'Coronavirus (COVID-19)' link. A 'BETA' label and a feedback invitation are also present. The main content area has a white background. It features a large section title 'Agile delivery' in bold, followed by a subtitle 'How to work in an agile way: principles, tools and governance.' Below this are sections for 'Understanding agile project management' (with links to 'Agile and government services: an introduction', 'Agile methods: an introduction', and 'Core principles of agile'), and a 'Join the community' section with links to 'Agile delivery community', 'Standards and assurance community', and 'Product and service community'. There are also 'Show all sections' and 'Hide' buttons.

GOV.UK

Topics Departments Government activity

→ [Coronavirus \(COVID-19\)](#) | Guidance and support

BETA Contact the Service Manual team if you have feedback, questions or suggestions.

[Service manual](#)

Agile delivery

How to work in an agile way: principles, tools and governance.

Show all sections ▾

Understanding agile project management

Introductions, methods, core features.

Hide ⌂

[Agile and government services: an introduction](#)

[Agile methods: an introduction](#)

[Core principles of agile](#)

Join the community

Find out what the cross-government community does and how to get involved.

[Agile delivery community](#)

[Standards and assurance community](#)

[Product and service community](#)

Lead with questions, not orders

- ▶ Don't tell your teams *how* you want things done. Tell them *what* you'd like done.

“Tell them what to do, and they will surprise you with their ingenuity.”

– General George S. Patton Jr.

- ▶ Guide with questions
 - ▶ “*What do you recommend?*” and “*How could we test that?*”
 - ▶ This approach will also grow your technical leads into better general managers

Don't switch to Agile too quickly

💥 63% of failed Agile implementations were due to clashes between the **business's culture** and **Agile's philosophy***



Don't switch to Agile too quickly

- ▶ **✗ Don't:** Switch the whole company to Agile in one go.
- ▶ **✓ Do:**
 - ▶ **Small batches.** Send teams this video, see who's interested, start there.
 - ▶ **Iterate:** Learn what works and tweak the process for *your* company's culture.
 - ▶ Slowly build **success stories** and case studies for internal adoption.
 - ▶ The best way to implement Agile is lean & iteratively, i.e. use Agile!

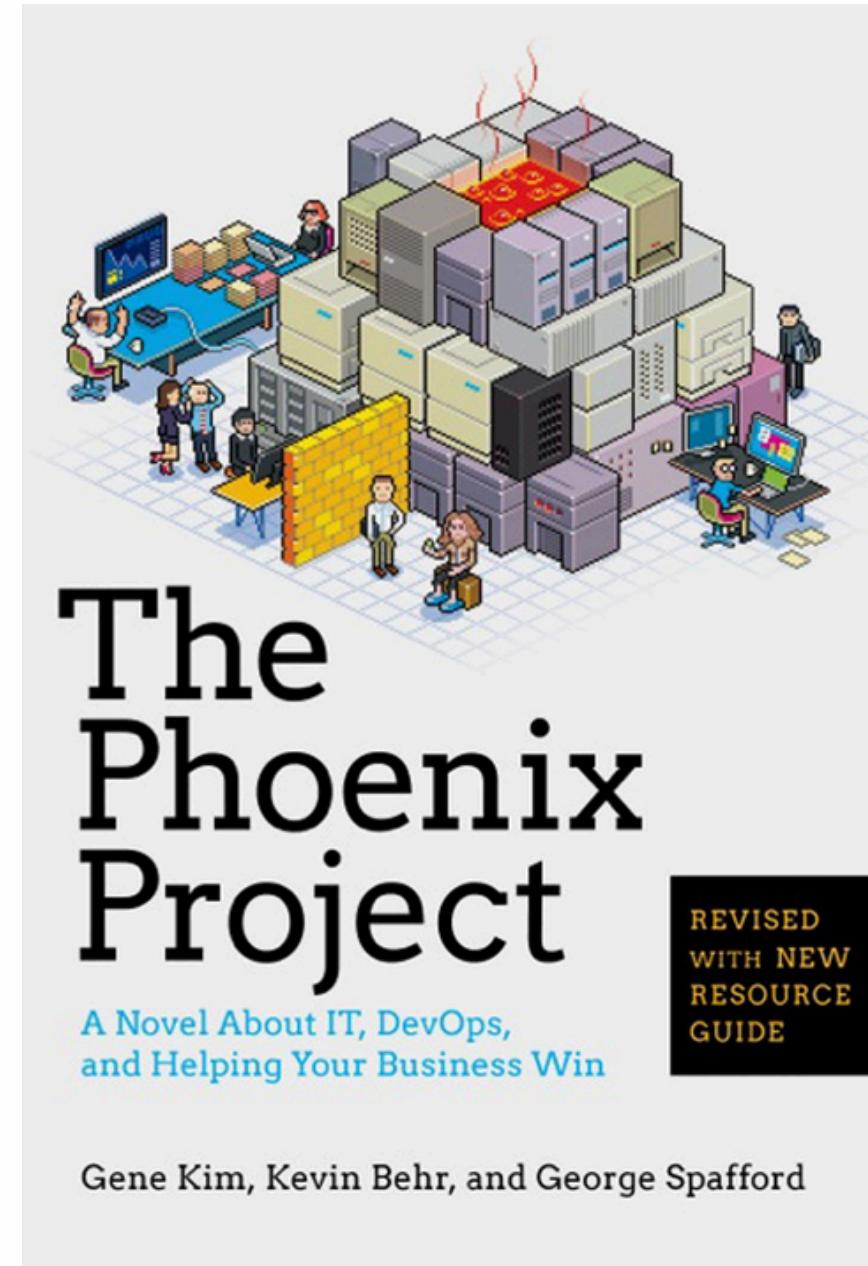
Next Steps

Part 07

Recommended reading



[Building High Performance Agile Teams](#) by Made Tech



[The Phoenix Project](#)
Gene Kim, Kevin Behr, George Spafford

Agile delivery

How to work in an agile way: principles, tools and governance.

[Show all sections](#)

Understanding agile project management

Introductions, methods, core features.

[Show](#)

Working with agile methods

Workspaces, tools and techniques, user stories, planning.

[Show](#)

Governing agile services

Principles, measuring progress, spending money.

[Show](#)

[Agile delivery](#)

The UK's Government Digital Service



bit.ly/pydata-agile

Intro to Agile

Table of contents

[Context](#)

[Resources](#)

[Sprint Meetings](#)

[Sprint Planning \(start of sprint, 30min\)](#)

[Sprint Lookahead \(mid-sprint, every other week, 30min\)](#)

[Sprint Review/Showcase \(end of sprint, 30min\)](#)

[Sprint Retro \(every 2 sprints, 30min\)](#)

[Tips for writing good story/task specs](#)

[FAQs](#)

[Books](#)

Resources

- [Learn agile estimation in 10 minutes](#)
- [Building High Performance Agile Teams](#)
- [Atlassian Agile Coach](#)
- [The Phoenix Project](#)
- [How to measure velocity](#)

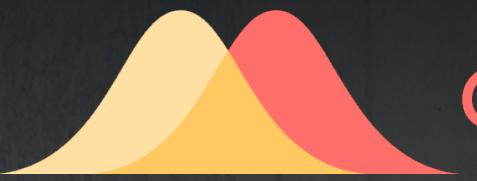
Sprint Review/Showcase (end of sprint, 30min)

Agenda

- Review what's been delivered
- What's done? What's not done? (Definition of done?)

Notes

- This is a great opportunity to showcase deliverables & gauge feedback from the wider team. In larger organisations, this may split into a Sprint Review (e.g. data team only) vs a monthly Sprint Showcase (external stakeholders & management).
- The Sprint Review also closes out the sprint, all tasks should now be closed (as Completed or Won't Do), or moved into a new bucket (rolled into the next sprint, Backlog, or Icebox).
- Velocity (total effort points delivered) should also be calculated at this point. The value of calculating velocity becomes clearer around Sprint 3. It helps set expectations (internally & externally) on what we can expect to deliver each sprint. It also forces greater focus by not over-planning too much work, which in turn lifts productivity.



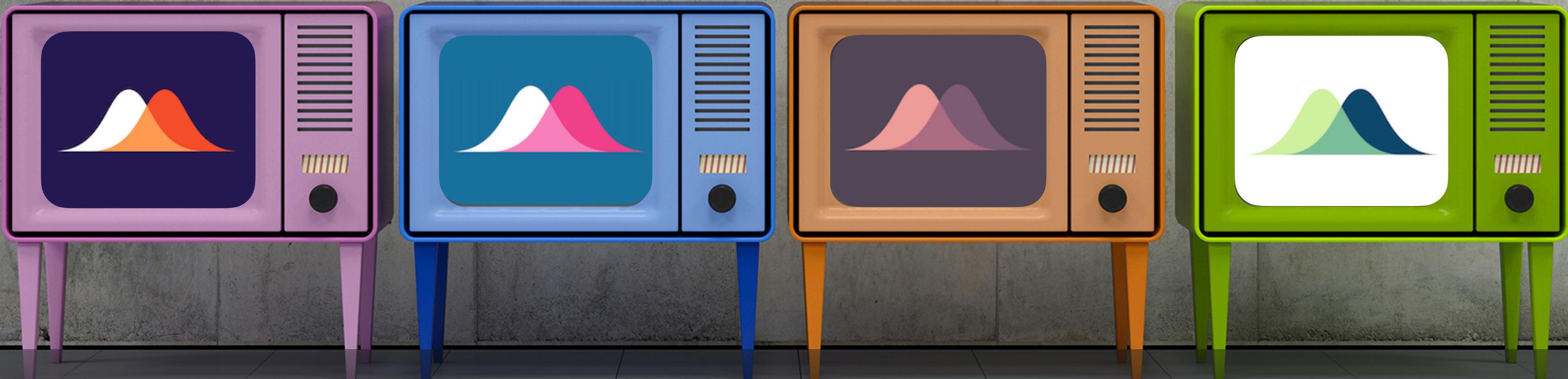
coefficient



PyData Global



bit.ly/pydata-agile



@john_sandall

@CoefficientData