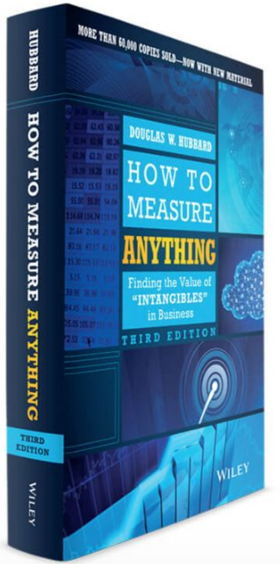
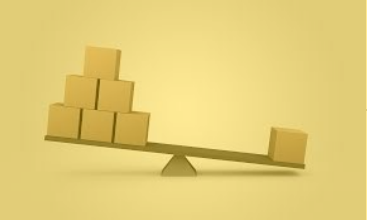
How to measure anything

by Douglas W. Hubbard



www.howtomeasureanything.com

#	Question
1	*In 1938 a British steam locomotive set a new speed record by going how fast (mph)?*
2	*In what year did Sir Isaac Newton publish the universal laws of gravitation?*
3	*How many inches long is a typical business card?*
4	*The Internet (then called "Arpanet") was established as a military communications system in what year?*
5	*In what year was William Shakespeare born?*
6	*What is the air distance between New York and Los Angeles in miles?*
7	*What percentage of a square could be covered by a circle of the same width?*
8	*How old was Charlie Chaplin when he died?*
9	*How many pounds did the first edition of the "How to measure anything" book weigh?*
10	*The TV show Gilligan's Island first aired on what date?*



A *measurement* is an **observation that quantitatively reduces uncertainty**.

Expressed as range with confidence level e.g. xxx increased between 10% and 20% (90% confidence interval)

“

If a thing can be observed in any way at all, it lends itself to some type of measurement method.

No matter how "fuzzy" the measurement is, it's still a measurement if it tells you more than you knew before.

"

Often, an important decision requires better knowledge of the alleged intangible, ***but when a [person] believes something to be immeasurable, attempts to measure it will not even be considered.***

“

If ***outcome*** *of a decision* ***is highly uncertain*** *and* ***has significant consequences*** *then measurements that reduce uncertainty have a high value*

(don't confuse the proposition that anything that can be measured with everything should be measured)

"

Simple statistical models outperform subjective expert judgement in almost every area of judgement…

“

Applied Information Economics

A universal approach to measurement

1) Define the decision

Start with the decision you need to make, then figure out which variables would make your decision easier if you had better estimates of their values

“

By asking specific questions tied to observables, we can turn our “intangibles” into the known and measurable.

“

If one can't identify a decision that could be affected by a proposed measurement and how it could change those decisions, then the measurement simply has no value

“

Some specific questions

- *What do you mean by …?*
- *Why does it matter to you…?*
- *What are you observing when you improved …?*

2) Determine what you know now

Instead of being overwhelmed by the apparent uncertainty about a problem, start to ask what things about it you do know

“

When you know almost nothing, almost anything will tell you something

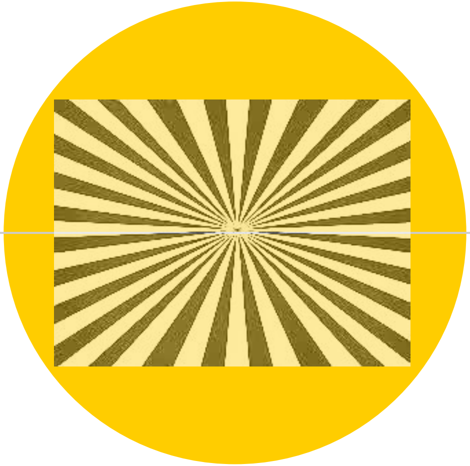
(it's a common misconception that the higher the uncertainty, the more data you need to significantly reduce it)

“

A black story example

If you act like you know something, but you don't, it can mislead people, and calibration *can help you avoid doing that either accidentally or unconsciously.*

“



Invest time/training in **calibration**

Use the 90% confidence interval

A 90% CI is a range of values that is 90% likely to contain the correct value.

A 90% CI “*means there is a 5% chance the true value could be greater than the upper bound, and a 5% chance it could be less than the lower bound.*

5% below lower bound	90% range	5% above upper bound

#	Question	Lower bound (95% chance value is higher)	Upper bound (95% chance value is lower)
1	*In 1938 a British steam locomotive set a new speed record by going how fast (mph)?*		
2	*In what year did Sir Isaac Newton publish the universal laws of gravitation?*		
3	*How many inches long is a typical business card?*		
4	*The Internet (then called "Arpanet") was established as a military communications system in what year?*		
5	*In what year was William Shakespeare born?*		
6	*What is the air distance between New York and Los Angeles in miles?*		
7	*What percentage of a square could be covered by a circle of the same width?*		
8	*How old was Charlie Chaplin when he died?*		
9	*How many pounds did the first edition of the "How to measure anything" book weigh?*		
10	*The TV show Gilligan's Island first aired on what date?*		

Equivalent bet test. Suppose you're asked to give a 90% CI for the year in which Newton published the universal laws of gravitation, and you can win $1,000 in one of two ways:

90% CI

You win $1,000 if the true year of publication falls within your 90% CI. Otherwise, you win nothing.

Spin a dial

You spin a dial divided into two "pie slices," one covering 10% of the dial, and the other covering 90%. If the dial lands on the small slice, you win nothing. If it lands on the big slice, you win $1,000.

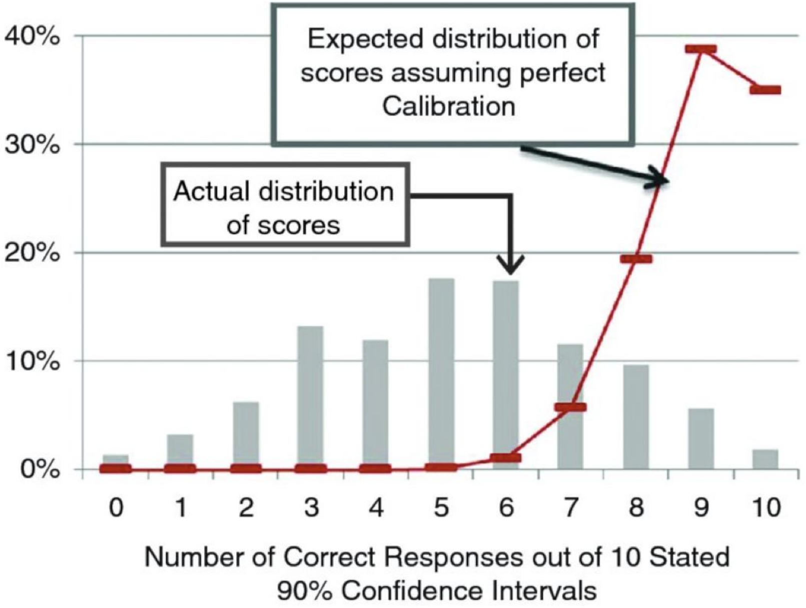
What would you prefer?
(1) ... to win $1000 if the correct answer is within your bounds?
(2) ...to spin the dial that gives a 90%?

Apply the Equivalent bet test to your ranges

#	Question	Answer
1	*In 1938 a British steam locomotive set a new speed record by going how fast (mph)?*	**126**
2	*In what year did Sir Isaac Newton publish the universal laws of gravitation?*	**1685**
3	*How many inches long is a typical business card?*	**3,5**
4	*The Internet (then called "Arpanet") was established as a military communications system in what year?*	**1969**
5	*In what year was William Shakespeare born?*	**1564**
6	*What is the air distance between New York and Los Angeles in miles?*	**2451**
7	*What percentage of a square could be covered by a circle of the same width?*	**78,5%**
8	*How old was Charlie Chaplin when he died?*	**88**
9	*How many pounds did the first edition of the "How to measure anything" book weigh?*	**1,23**
10	*The TV show Gilligan's Island first aired on what date?*	**26.09.1964**

Are you overconfident?

Result	For calibrated estimators	Conclusion
6 or less out of 10	1,3%	you are very likely overconfident
5 or less		you are overconfident and by a large margin
At least 7 out of 10	99%	You might be calibrated



Repetition and feedback

Make lots of estimates and then see how well you did. For this, play CFAR’s Calibration Game.

Visualize risk using simulations

We want to know the probability of a huge loss, the probability of a small loss, the probability of a huge savings, and so on. That’s what Monte Carlo can tell us.

The one-year lease [for the machine] is $400,000 with no option for early cancellation. So if you aren't breaking even, you are still stuck with it for the rest of the year. You are considering signing the contract because you think the more advanced device will save some labor and raw materials and because you think the maintenance cost will be lower than the existing process.

- Maintenance savings (MS): $10 to $20 per unit
- Labor savings (LS): -$2 to $8 per unit
- Raw materials savings (RMS): $3 to $9 per unit
- Production level (PL): 15,000 to 35,000 units per year
- annual savings will equal (MS + LS + RMS) × PL

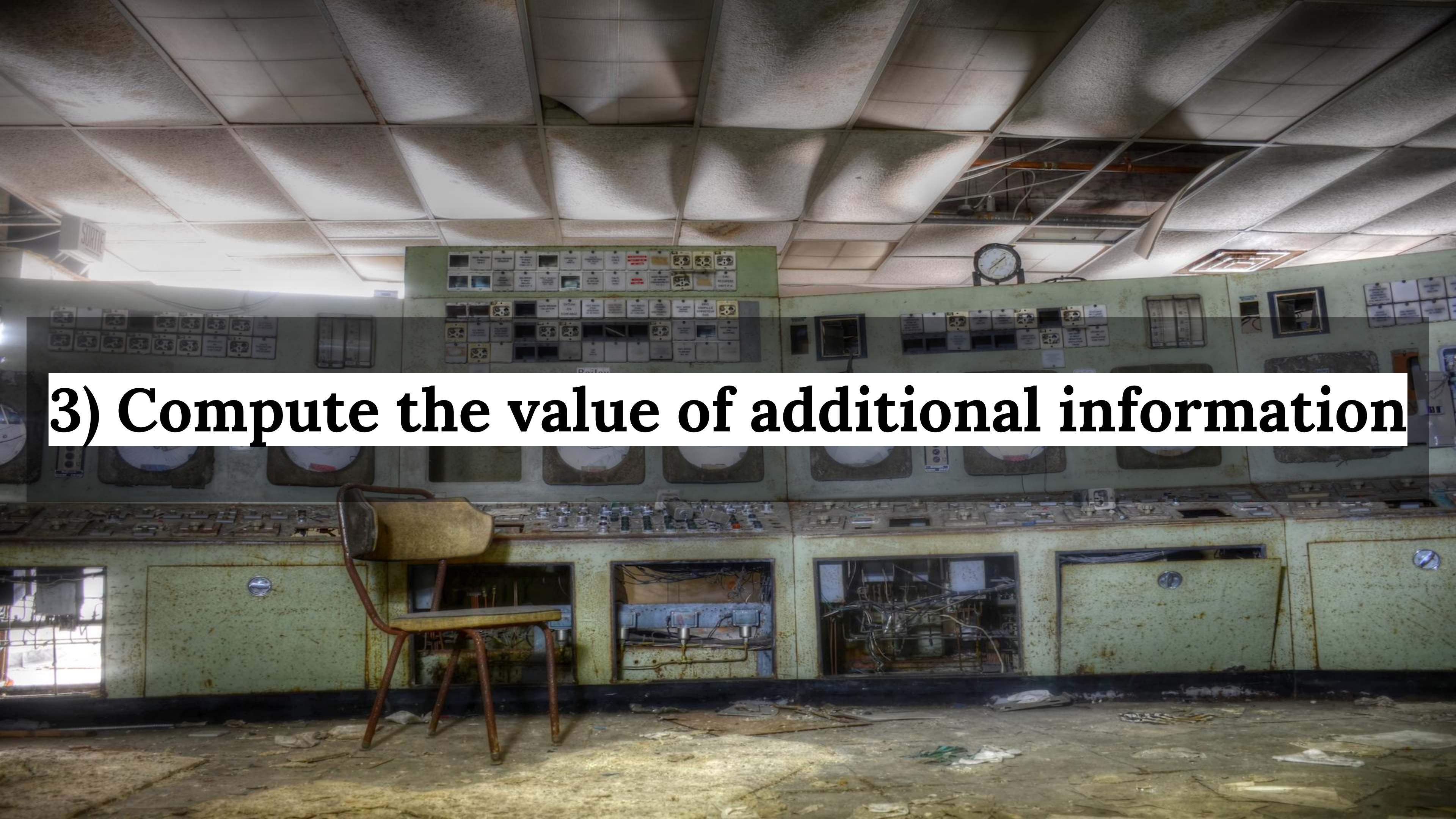
Let’s simulate with Monte Carlo

https://docs.google.com/spreadsheets/d/1RVJF4Wb8ze4DymirRmTyN2yP8Em2K6ngv3-NqfyfWUE/edit?usp=sharing

Getting **more advanced** (but not today)

- Other distributions (Beta, Power Law, Triangular,…)
- Dependent variables
- Markov simulation
- Agent based simulation

https://www.hubbardresearch.com/downloads/



3) Compute the value of additional information

Knowing the value of the measurement affects how we might measure something or even whether we need to measure it at all

“

Information can reduce uncertainty about important decisions.

It's too costly to acquire perfect information, so instead we'd like to know **which** decision-relevant **variables** are the ***most* valuable to measure** more precisely, so we can decide which measurements to make.

“By 1999, I had completed the… Applied Information Economics analysis on about 20 major [IT] investments… Each of these business cases had 40 to 80 variables, such as initial development costs, adoption rate, productivity improvement, revenue growth, and so on. For each of these business cases, I ran a macro in Excel that computed the information value for each variable… [and] I began to see this pattern: * The vast majority of variables had an information value of zero… * The variables that had high information values were routinely those that the client had never measured… * The variables that clients [spent] the most time measuring were usually those with a very low (even zero) information value… “

Expected Opportunity Loss (EOL)

Simple Expected Opportunity Loss (EOL) example. Suppose you could make $40 million profit if [an advertisement] works and lose $5 million (the cost of the campaign) if it fails. Then suppose your calibrated experts say they would put a 40% chance of failure on the campaign.

	Good Outcome (eg. Campaign succeeds)	Bad Outcome (eg. Campaign Fails)
Chance of Outcome:	60%	40%
Choice	Payoff	
A (eg. Invest in the new ad campaign)	$40.000.000	($5.000.000)
B (eg. Don't Invest in the ad campaign)	$0	$0

Expected Opportunity Loss (EOL)			
	Opportunity Loss	Chance of being wrong	EOL
If initially desired choice is A	$5.000.000	40%	$2.000.000
If initially desired choice is B	$40.000.000	60%	$24.000.000

By reducing uncertainty and with that reducing the chance of being wrong you reduce your EOL

Expected value of information

The difference between EOL before and after a measurement is the expected value of information – EVI

(and with that your threshold what to invest in that measurement).

The book describes a lot about the calculation of the value of information … but that’s too deep for today (and for me atm).

Consult your calculation expert of your choice and have fun!

Some terms: symmetric/assymetric loss functions, discrete approximation, expected value of perfect information, …

4) Measure where information value is high

Select a measurement method

- **Decomposition**: Which parts of the thing are we uncertain about?
- **Secondary research**: How has the thing (or its parts) been measured by others?
- **Observation**: How do the identified observables lend themselves to measurement?
- **Measure just enough**: How much do we need to measure it?
- **Consider the error**: How might our observations be misleading?

Decomposition

It's often the case that decomposition itself – even without making any new measurements – often reduces one's uncertainty about the variable of interest.

Observations

- Does it leave a trail? (e.g. hang up rates correlated to waiting times)
- Can you observe it directly?
- Can you create a way to observe it indirectly? (e.g. gift wrapping feature to know the amount of gifts)
- Can the thing be forced to occur under new conditions which allow you to observe it more easily? (e.g. changed return policy for some shops and compare results … A/B tests)

Just enough

Because initial measurements often tell you quite a lot, and also change the value of continued measurement,

Hubbard often aims for **spending 10% of the EVPI on a measurement**, and sometimes as little as 2% (especially for very large projects).

Some bias to consider

- **Confirmation bias**: people see what they want to see.
- **Selection bias**: your sample might not be representative of the group you're trying to measure.
- **Observer bias**: the very act of observation can affect what you observe.

More hints

- ***Work through the consequences***: If the value is surprisingly high, or surprisingly low, what would you expect to see?
- ***Be iterative***: Start with just a few observations, and then recalculate the information value.
- ***Consider multiple approaches***: Your first measurement tool may not work well. Try others.
- ***What's the really simple question that makes the rest of the measurement moot?*** First see if you can detect *any* change in research quality before trying to measure it more comprehensively.

Sampling

5

Rule of 5 *(Mathless estimation)*
There is a 93.75% chance that the median of a population is between the smallest and largest values in any random sample of five from that population.

Catch - reCatch

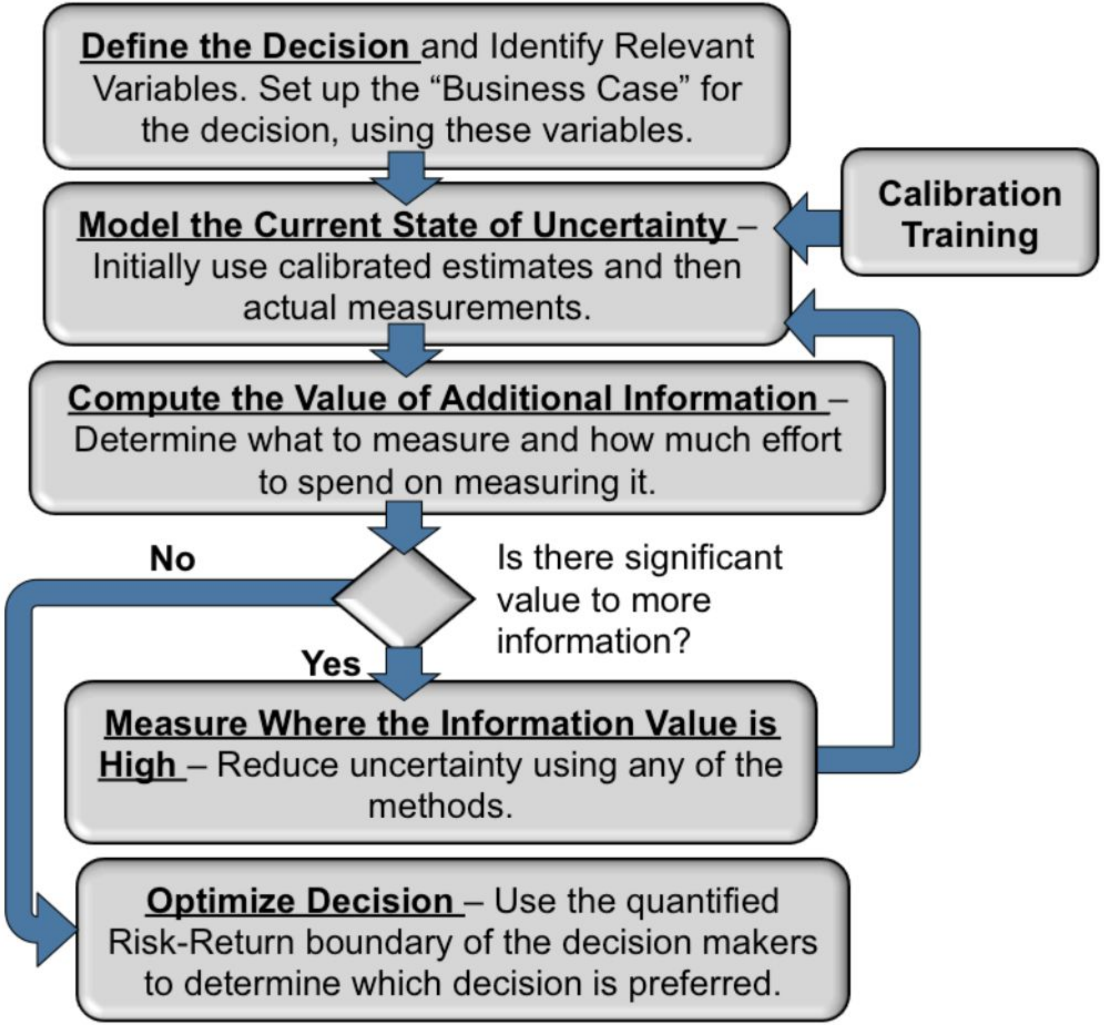
How does a biologist measure the number of fish in a lake? SHe catches and tags a sample of fish – say, 1000 of them – and then releases them. After the fish have had time to spread amongst the rest of the population, she'll catch another sample of fish. Suppose she caught 1000 fish again, and 50 of them were tagged. This would mean 5% of the fish were tagged, and thus that were about 20,000 fish in the entire lake.

And much more methods

- *Spot sampling*
- *Clustered sampling*
- *Measure to the threshold*
- *Regression modeling*
- *Instinctive Bayesian approach*
- *Prediction markets*
- *Rasch models*
- *Models for measuring preferences and happiness*
- *Improve subjective judgements of experts*
- *…*

5) Make a decision and act on it

Recap


Define the Decision and Identify Relevant Variables. Set up the "Business Case" for the decision, using these variables.
Model the Current State of Uncertainty – Initially use calibrated estimates and then actual measurements.
Calibration Training
Compute the Value of Additional Information – Determine what to measure and how much effort to spend on measuring it.
Is there significant value to more information?
No
Yes
Measure Where the Information Value is High – Reduce uncertainty using any of the methods.
Optimize Decision – Use the quantified Risk-Return boundary of the decision makers to determine which decision is preferred.

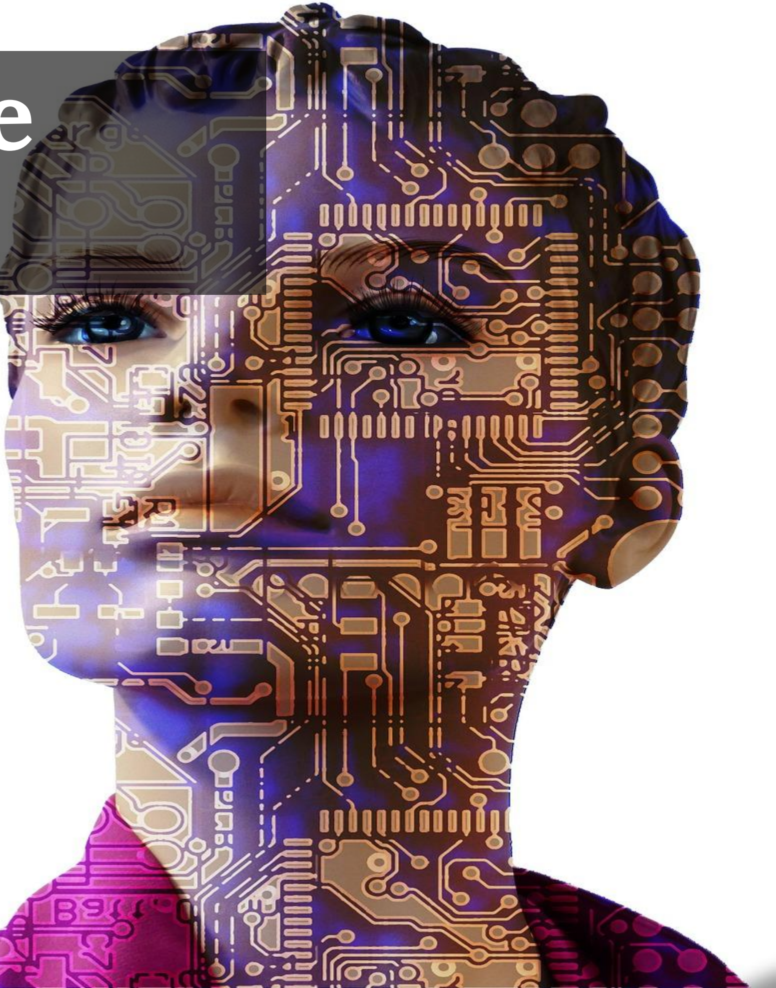
4

Measurement assumptions

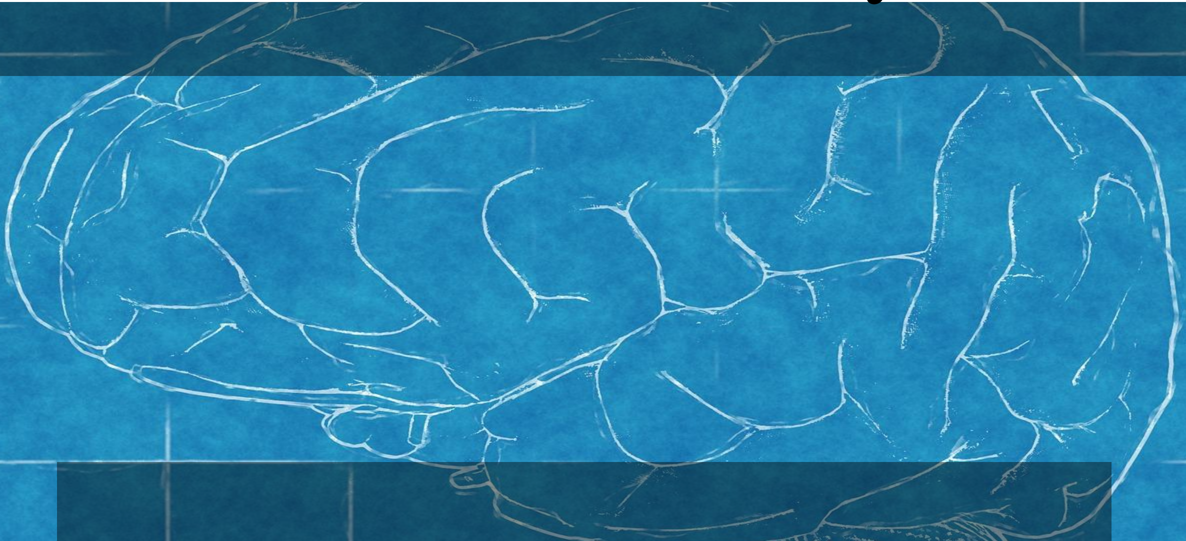

It’s been done before
Don’t reinvent the wheel

You have access to more data than you think

It might just involve some resourcefulness and original observations.



You need less data than you think



If you’re clever about how to analyze it.


An adequate amount of new data is...
probably more accessible than you first thought.

Cost of delay - short recap

Is a month of delay worth 1 Mio € or 1k €?

The *impact of time on value*

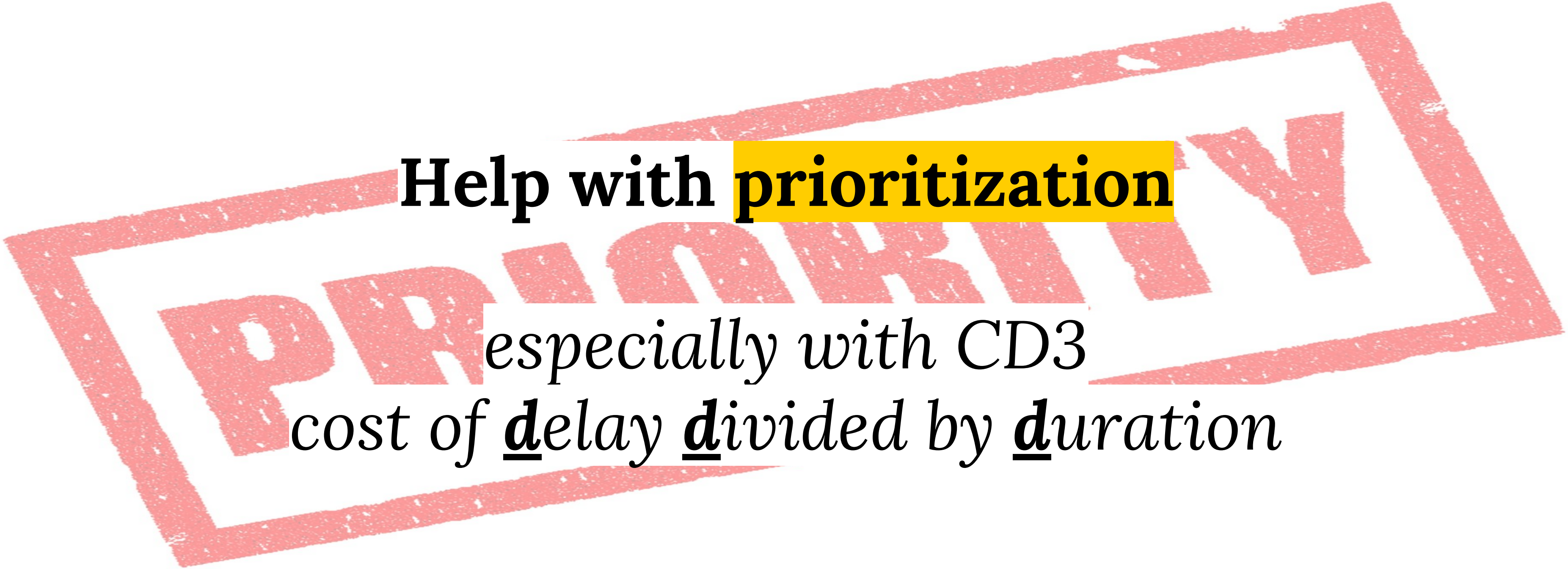
Cost of Delay (CoD) - the rate of decay of value per period of delay.

Units for example could be dollars per week.

“

What is it good for?

Drive economically based decisions



Help with prioritization

especially with CD3

cost of ***d****elay* ***d****ivided by* ***d****uration*


Focus discussions to speed and value
(instead of cost and efficiency)

About Value

The monetary worth of something

A framework for thinking about value

Increase Revenue	**Increasing** sales to new or existing customers. Delighting or Disrupting to increase market share and size
Protect Revenue	Improvements and incremental innovation to ***sustain*** current market share and revenue figures
Reduce Costs	Costs that we are *currently* incurring, that can be **reduced**. More efficient, improved margin or contribution
Avoid Costs	Improvements to ***sustain*** current cost base. Costs we are not currently incurring but may do in the future

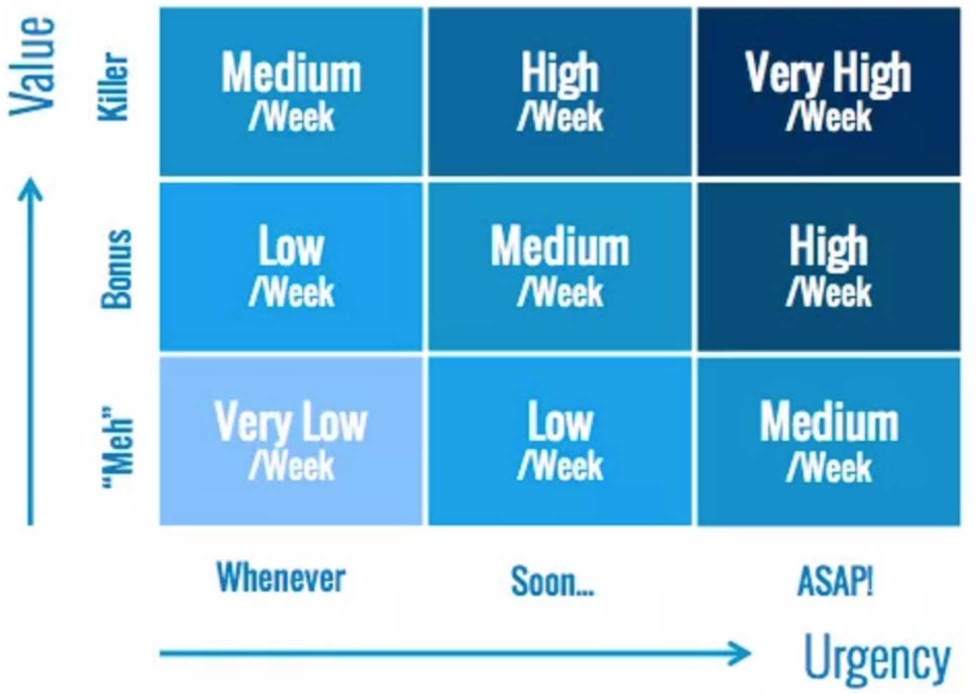
Total value

=

Sum of value buckets

About Urgency

Describes the development of value over a given timeframe



BLACK SWAN FARMING

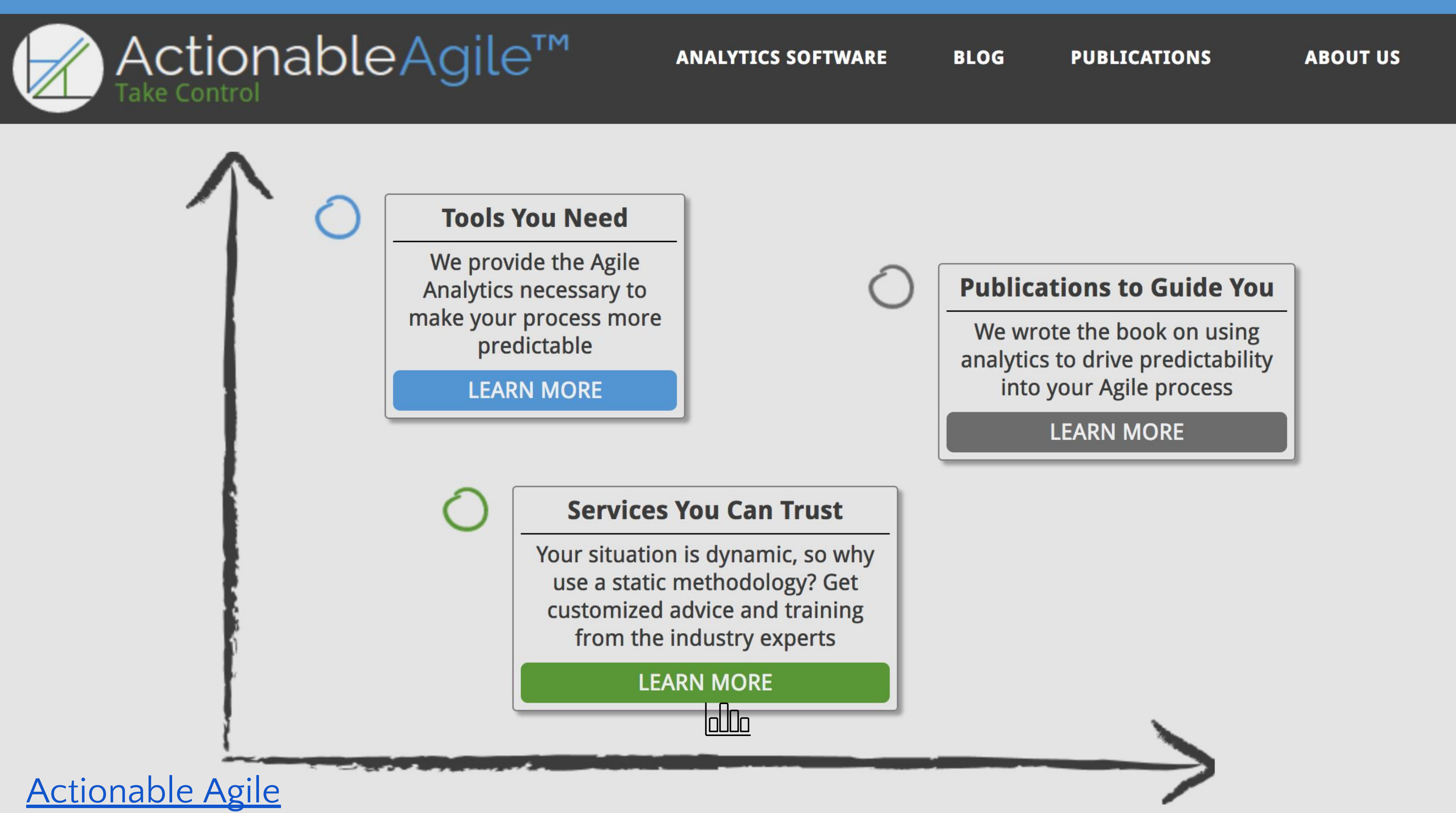
- Fast and easy to apply
- Helps to differentiate between many options initially

Qualitative Cost of Delay matrix

Combine CoD with applied information economics

- AIE-framework to search for your value drivers – spotting the right variables to consider
- Find input for filling your value buckets
- Go data driven and consider what you know, what to measure and what is the value of that measurement … and replace HIPO decisions
- Simulate value development combined with assumed urgency profiles and derive investment decisions (using Monte Carlo instead of just gut feeling)

Get to know your Delays



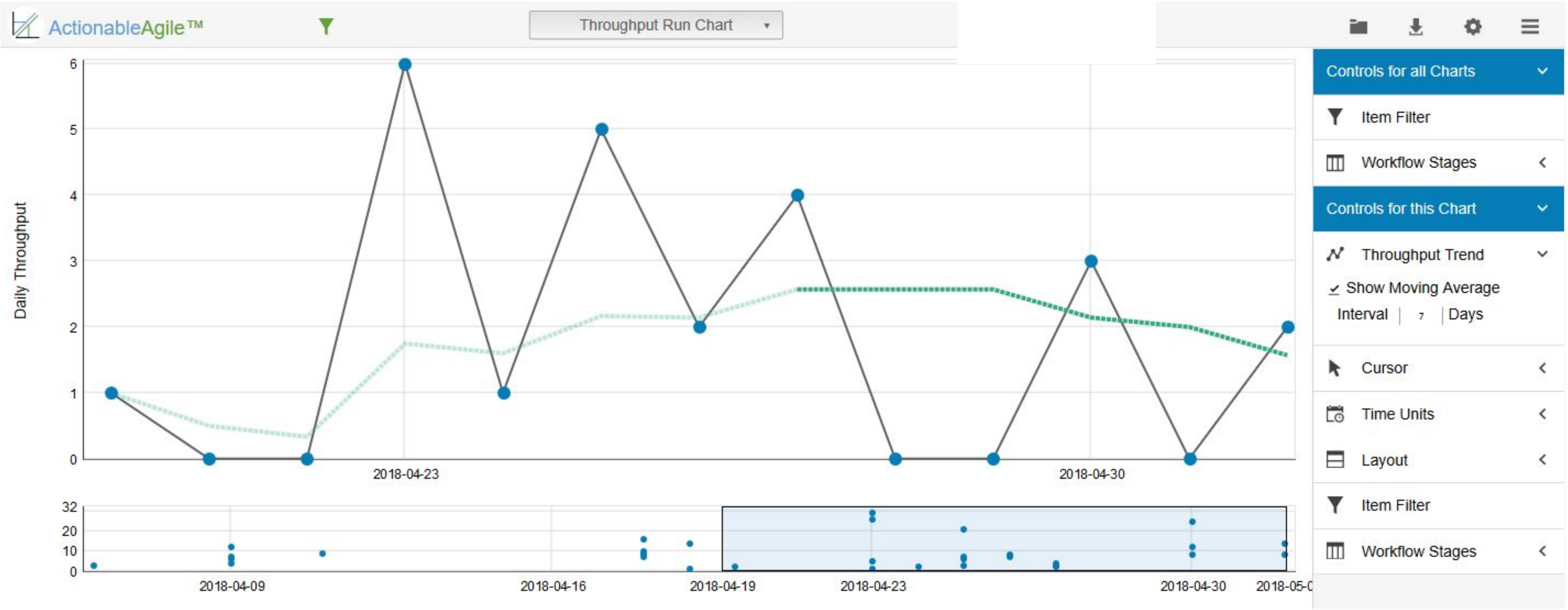
[Actionable Agile]()

Use what is already known in Agile and Lean

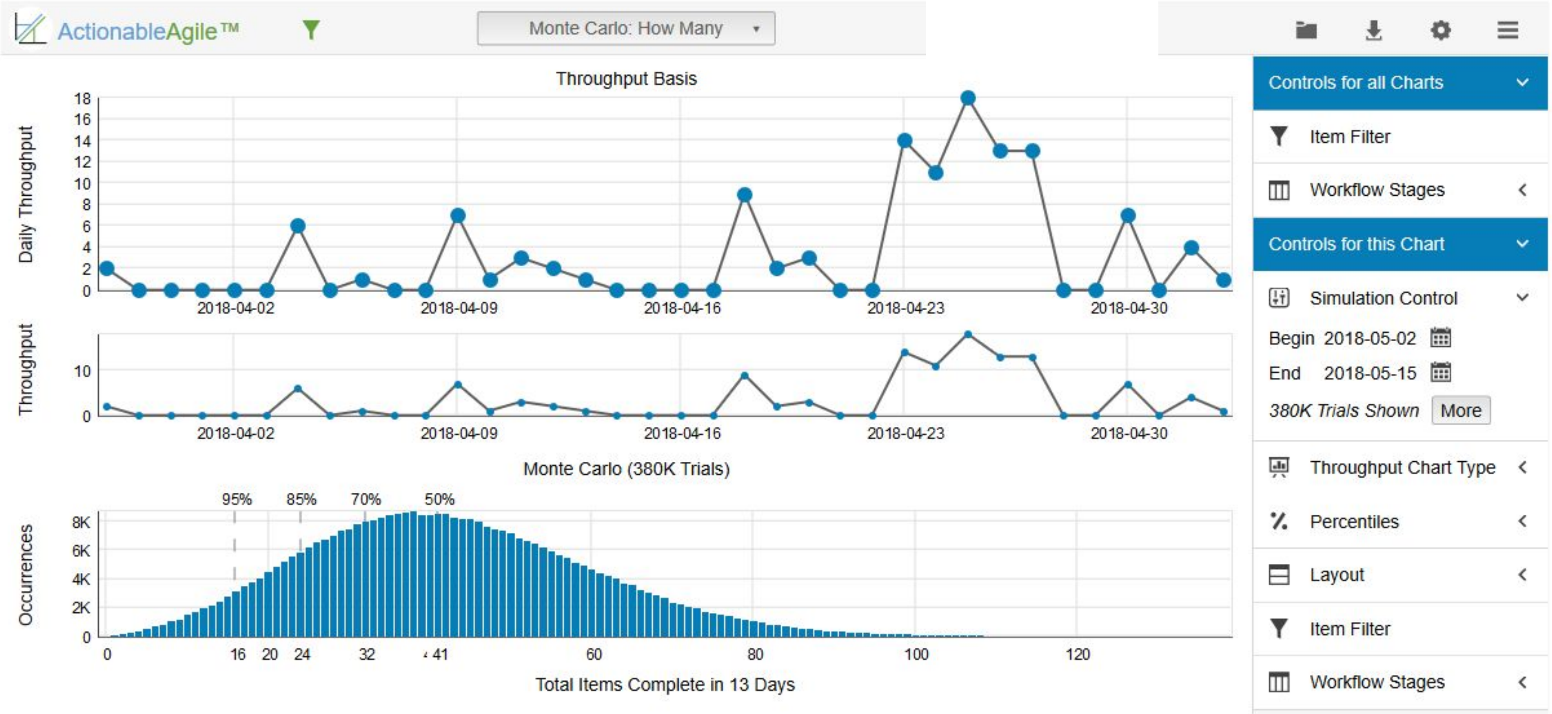
- *Use described ways to measure lead time (system lead time and customer lead time) … see some examples on the next slide*
- *Focus on the measurements that influence your decisions (…and avoid using misleading ones e.g. number of story points, lines of code, time tracking)*

Some teaser charts

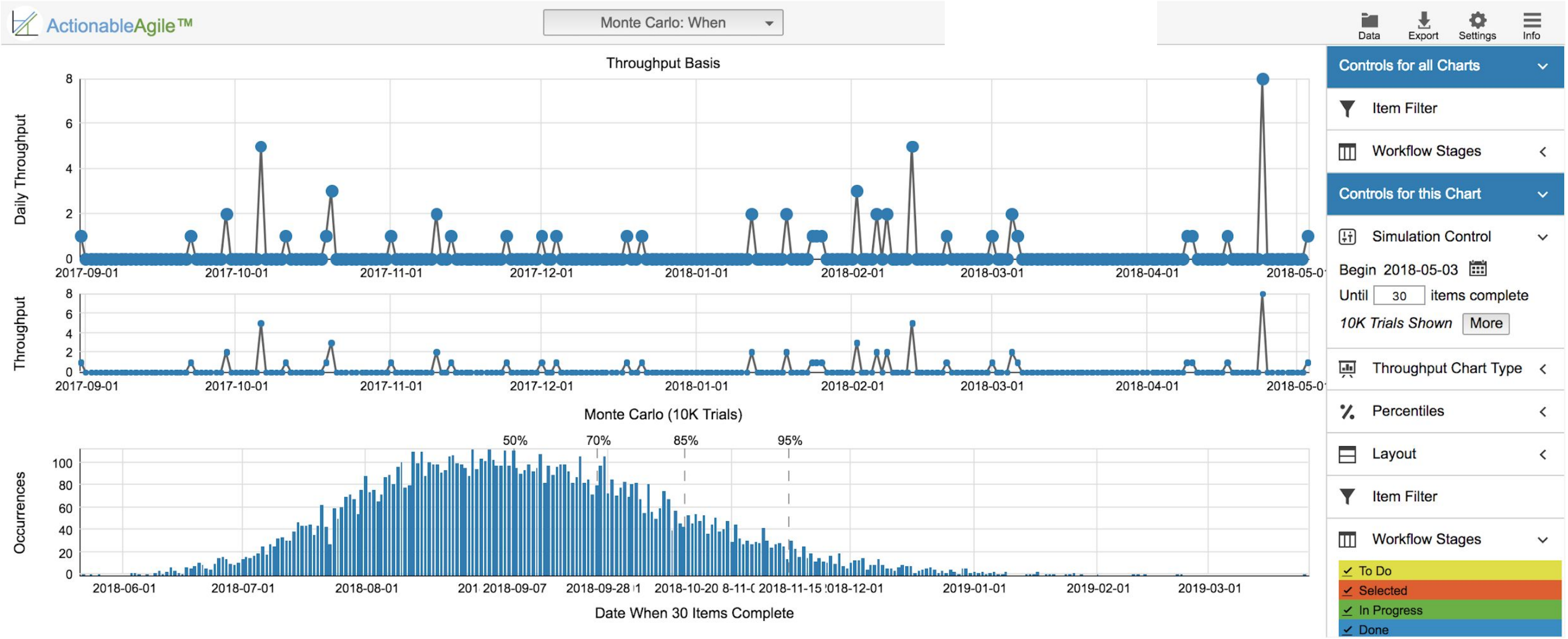
Throughput Run Chart



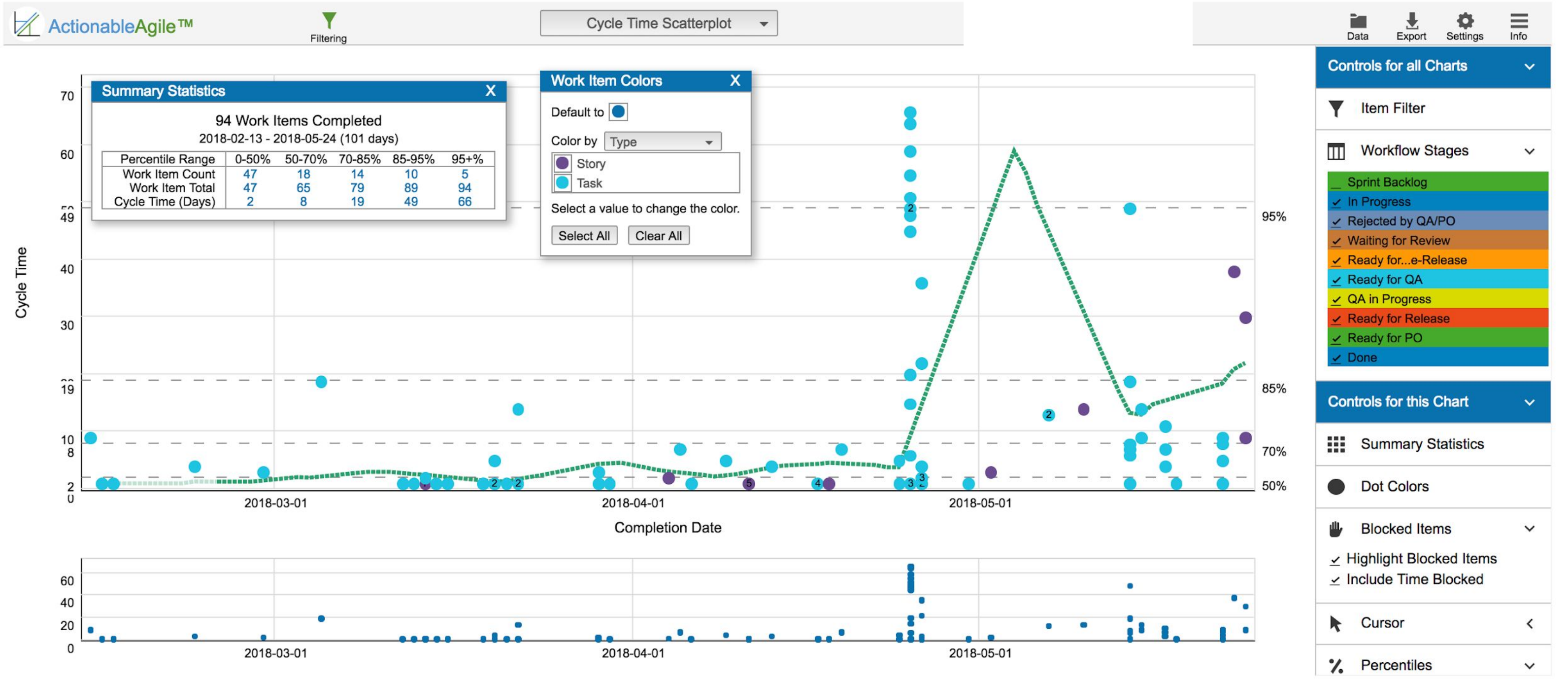
Monte Carlo: How Many



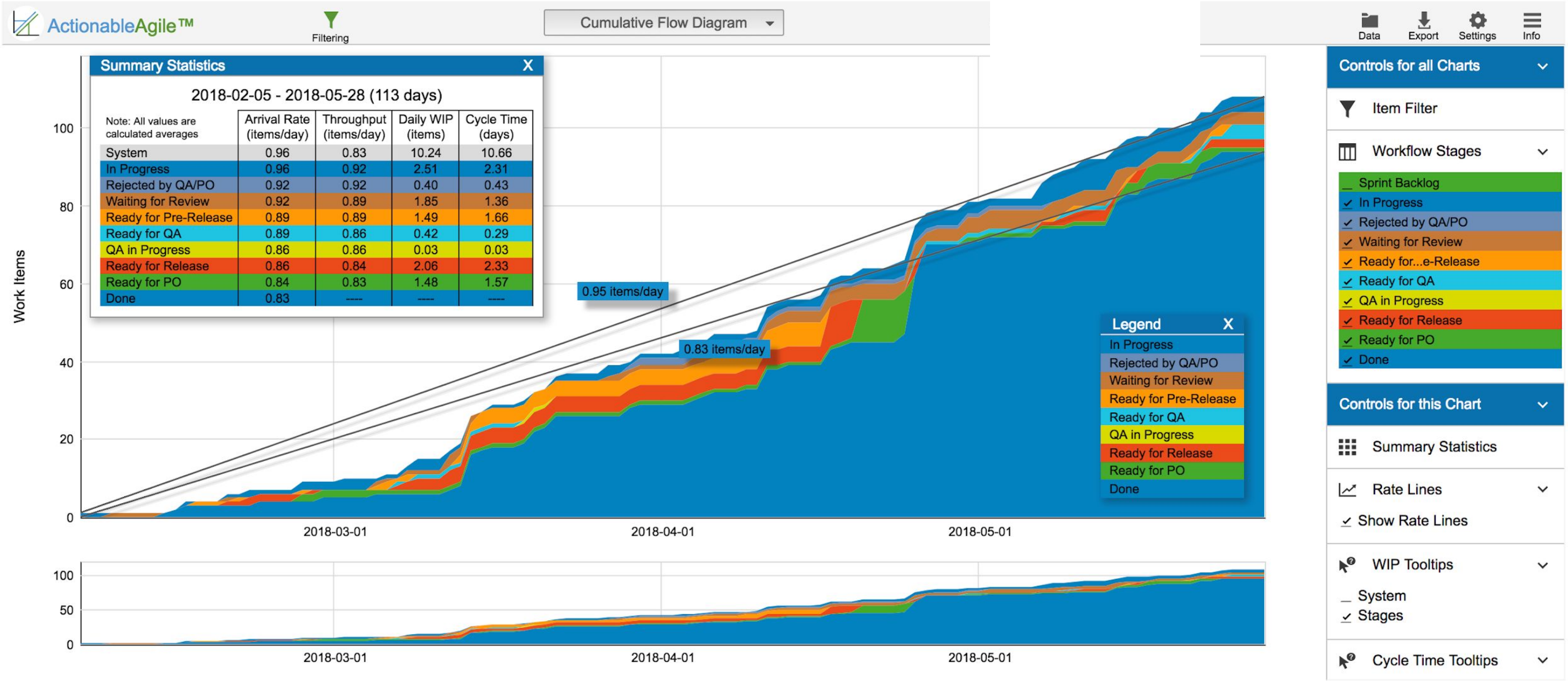
Monte Carlo: When



Cycle Time Scatterplot

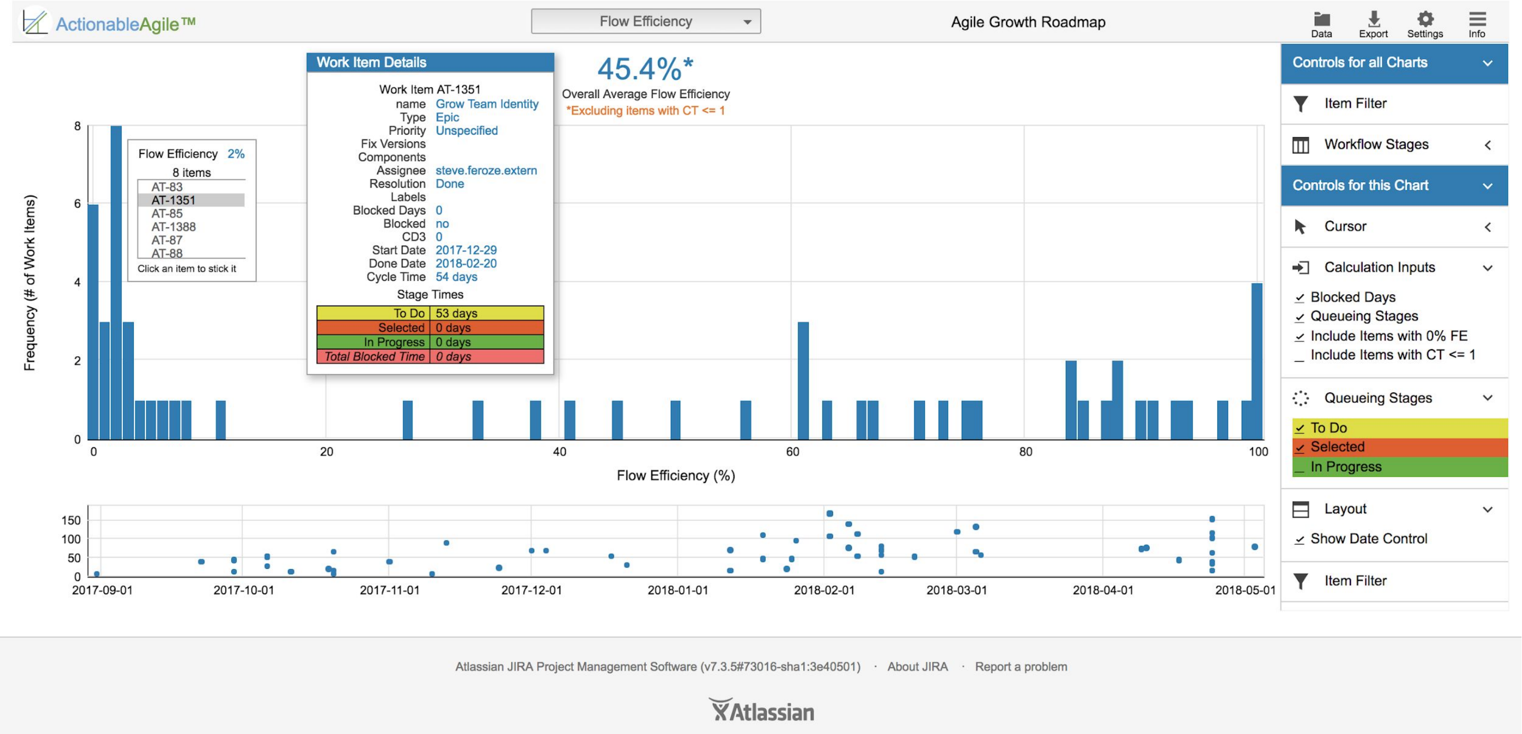


Cumulative Flow Diagram

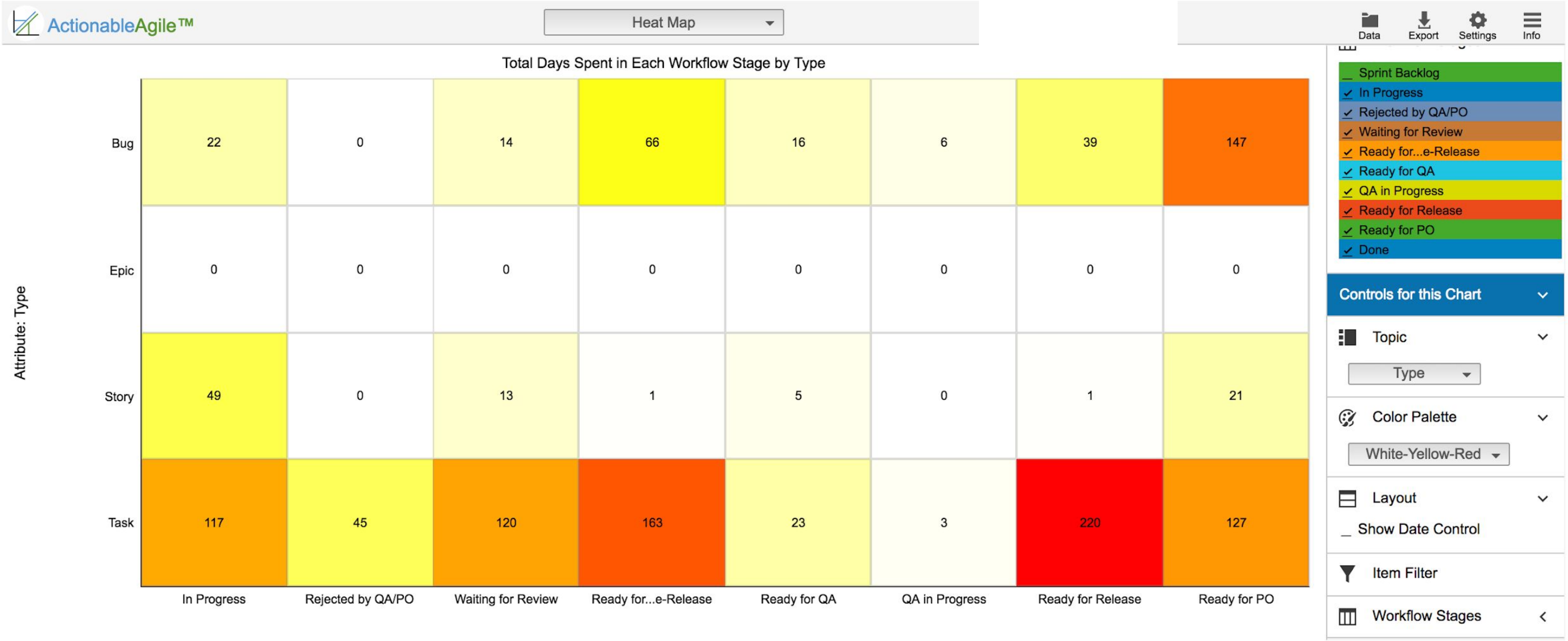


Note: All values are calculated averages	Arrival Rate (items/day)	Throughput (items/day)	Daily WIP (items)	Cycle Time (days)
System	0.96	0.83	10.24	10.66
In Progress	0.96	0.92	2.51	2.31
Rejected by QA/PO	0.92	0.92	0.40	0.43
Waiting for Review	0.92	0.89	1.85	1.36
Ready for Pre-Release	0.89	0.89	1.49	1.66
Ready for QA	0.89	0.86	0.42	0.29
QA in Progress	0.86	0.86	0.03	0.03
Ready for Release	0.86	0.84	2.06	2.33
Ready for PO	0.84	0.83	1.48	1.57
Done	0.83	----	----	----

Flow Efficiency



Cycle Time Heat Map



Read more

- How to measure anything by Douglas Hubbard
- Cost of Delay – how to find the best sequence for your feature development
- Cost of Delay – a key economic metric
- Actionable agile metrics
- Book summary – https://www.lesswrong.com/posts/ybYBCK9D7MZCcdArB/how-to-measure-anything