Focused Objective
forecasting - risk - staff - cost of delay

Planning and Forecasting using Data

Capturing and using data for forecasting

Workshop Manual

Digital materials: http://Bit.Ly/ForecastingUsingData
Email me: troy.magennis@focusedobjective.com

Version 2.1

Learning Outcomes

- To learn the pros and cons of different forecasting techniques.
- To learn how to use data to forecast and how much data is needed.
- To learn how to perform Monte Carlo forecasting and why.
- To learn how to quickly estimate work size and delivery pace ranges.
- To learn what system factors play a role in forecasting.

Exercise

Discuss in pairs

Q1: What do you already know about forecasting?

Q2: What learning outcome speaks to you the most?

Prediction Intervals – Estimating the chance of where future samples will fall

What is a prediction interval?

Source: https://en.wikipedia.org/wiki/Prediction_interval

In statistical inference, specifically predictive inference, a prediction interval is an estimate of an interval in which future observations will fall, with a certain probability, given what has already been observed.

Estimating the range of actual data by random sampling

When actual data samples can be observed, its handy to know how likely it is that you have discovered the range of likely values. This is useful in understanding how likely there is a lower or higher sample yet to be discovered. Like all random sampling, there is absolutely no guarantee that you have discovered any amount of the range, but prediction intervals give you the probability on average.

Probability the next sample is within the previously seen range after "n" samples = $\frac{(n-1)}{(n+1)}$ X 100

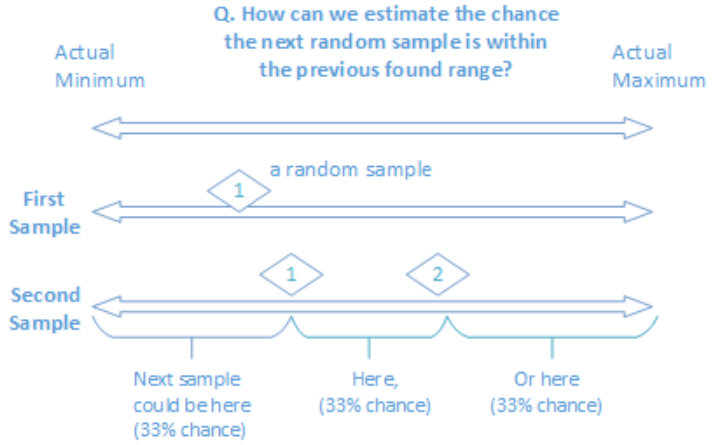
Probability the next sample is lower than the lowest sample so far after "n" = $\frac{1}{(n+1)}$ X 100

Probability the next sample is higher than the highest sample so far after "n" = $\frac{1}{(n+1)}$ X 100

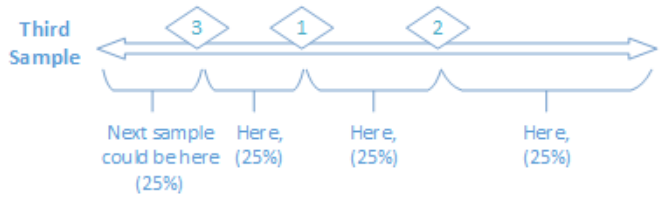
Samples so far (n)	Probability for each interval	Probability next sample in range
1	50.00%	0.00%
2	33.33%	33.33%
3	25.00%	50.00%
4	20.00%	60.00%
5	16.67%	66.67%
6	14.29%	71.43%
7	12.50%	75.00%
8	11.11%	77.78%
9	10.00%	80.00%
10	9.09%	81.82%
11	8.33%	83.33%
12	7.69%	84.62%
13	7.14%	85.71%
14	6.67%	86.67%
15	6.25%	87.50%

Samples so far (n)	Probability for each interval	Probability next sample in range
16	5.88%	88.24%
17	5.56%	88.89%
18	5.26%	89.47%
19	5.00%	90.00%
20	4.76%	90.48%
21	4.55%	90.91%
22	4.35%	91.30%
23	4.17%	91.67%
24	4.00%	92.00%
25	3.85%	92.31%
26	3.70%	92.59%
27	3.57%	92.86%
28	3.45%	93.10%
29	3.33%	93.33%
30	3.23%	93.55%

Q. How can we estimate the chance the next random sample is within the previous found range?



After two samples, there are three spots the next sample could be. Equally splitting the chances, there is a 33.33% chance the next sample is between the previous samples (1) and (2).



After three samples, there are four spots the next sample could be. Equally splitting the chances, there is a 50% chance the next sample is between the lowest (3) and highest (2) so far.

Important assumptions (that are rarely perfectly true)

- The samples are taken at random. Convenient isn't random!
- The distribution is uniform – all values have equal chance.
- The probability is on average. It's when it is more likely than not (~50%)

These are rarely always true in the real world. Milage will vary depending mainly on the underlying distribution. If the distribution is skewed, it can take hundreds of samples to get the lower probability end of the range.

For more probabilistic forecasting resources: FocusedObjective.com Email: troy.magennis@focusedobjective.com

Prediction Intervals Exercise

To find how many samples it takes to find the lower and upper bounds of a sample set on average? This exercise simulates finding the upper and lower boundary of a sequential range by sampling the result of dice rolls.

The process

1. **Roll Dice**: Create a random number with a range of 1 to 100. Options:
    a. A random number generator app on your phone (Randomizers)
    b. Use three rolls of a six-sided dice (see next page for chart)
    c. Sum two 10 sided dice (00 – 90 by 10's) and a traditional (0-9)
2. **Repeat**: Repeat 20 times and record the results in the table below.
3. **Examine Results**: Look at the range between the lowest rolled and highest rolled. Compare against expected.

2 x 10-Sided

Questions and discussion topics

1. **What probability distribution is a single roll?**
2. **What guarantee do I have that I have found the range expected?**
3. **What happens if the data is a Normal (bell curve) distribution?**
4. **What happens if the data is left or right skewed?**

Note: Rolling a 00 and 0 = 100

Results table

Record each roll & calculate the ranges seen so far after each roll. Compare to expected.

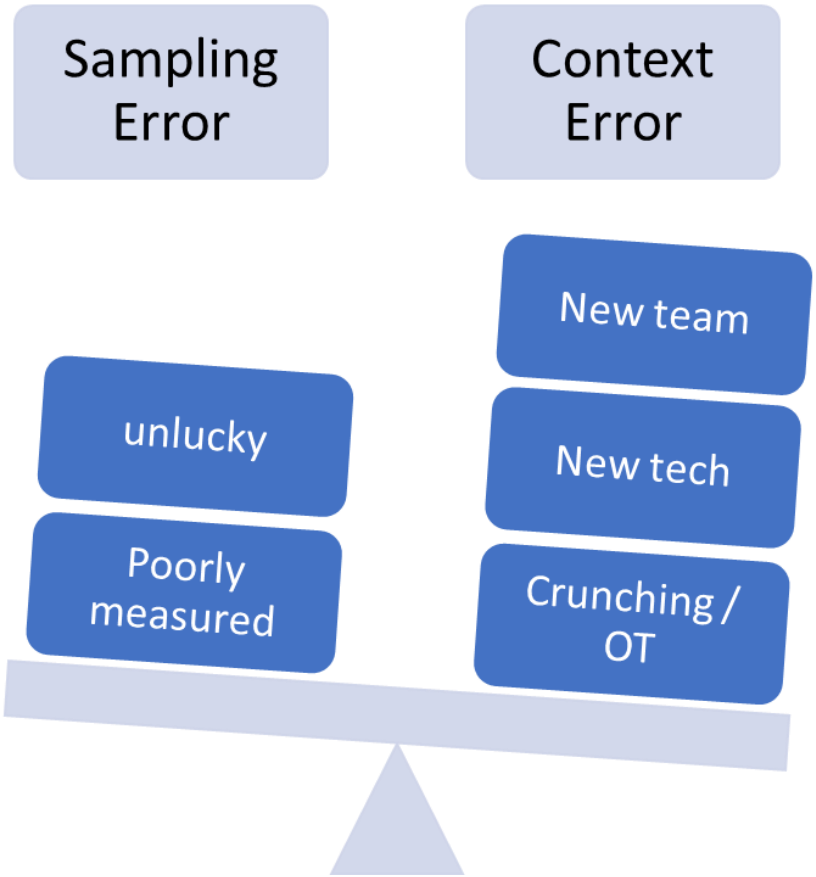
Roll (n)	Value of This Roll	Lowest value seen So Far	Highest Value seen So Far	Range So Far = Highest-Lowest	Expected Range (after roll) $\frac{(n-1)}{(n+1)} \times 100$
1					0
2					33.3
3					50
4					60
5					66.6
6					71.4
7					75
8					77.8
9					80
10					81.2
11					83.3
12					84.6
13					85.7
14					86.7
15					87.5
16					88.2
17					88.9
18					89.5
19					90

How Many Samples?

- < 3 samples
  - Guess the range, go wide
- 4-7 samples
  - Narrow guessed range using data
- 7-15 samples
  - Use data
- > 15 samples, delete the oldest and least relevant

Always – look for reasons to delete stale historical data

Always – sample at random, no cherry picking!



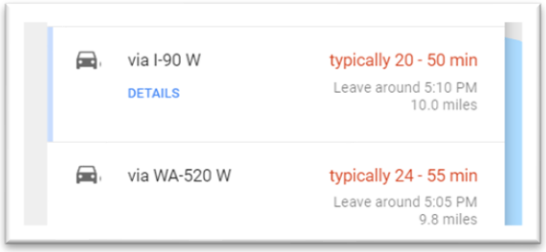
Reflection and Lessons
Topic: Data Sampling

Think about what we just covered, and consider these questions. You don't need to share these with the group. These are for YOU.

Q1: What did you learn about data sampling from this section?

Q2: What will you do differently using data in forecasting in the future?

“When to Leave Home” Forecasting Exercise



Given two addresses compute the “**when to leave home**” time in order to arrive by a desired **target time**.

Q1 – What is the general algorithm you might use to compute this time?

Spend 5 minutes discussing the problem and how you will compute the “leave home” time. Then decide on one person to present that algorithm using step-by-step sticky notes to the room in 2 minutes or less. Tip: start by discussing how you personally decide when to leave home when travelling to a fixed time event.
Extra credit: Define “success” for your solution.
After hearing and discussing all of the group’s ideas, spend 5 minutes and decide what method you think solves this problem the best.

Q2 – What data would you require to implement that algorithm?

Spend 5 minutes creating a sticky-notes for each input data or information required for your algorithm. Decide if its critical to have (can’t do without) or if it is an optimization (improves the forecast) and then decide on one person to present those stick-notes to the room in 2 minutes or less.
Extra Credit: For the critical data, describe what you will do if you can’t get or don’t yet have that data.

Q3 – What factors could cause forecasting error (arriving late)?

Spend 5 minutes discussing what external factors could cause forecasting error. Create a sticky-note for each factor and describe how you might “deal” with that factor in your forecasts. Decide on one person to present those sticky-notes to the room in 2 minutes or less.

Q4 – How might you measure success of this feature in production?

Spend 5 minutes discussing how to measure success in production, and how production data could improve the results over time. Decide on one person to present your findings to the room in 2 minutes or less.

Q5 – How might this same algorithm be used in forecasting when you need to start software features to deliver by a given date?

Spend 10 minutes as a group discussing how this technique might be used for forecasting software projects or features in multiple team environments. For example; What would need to be changed? What data you need for the feature/project and the system delivering those features/projects?

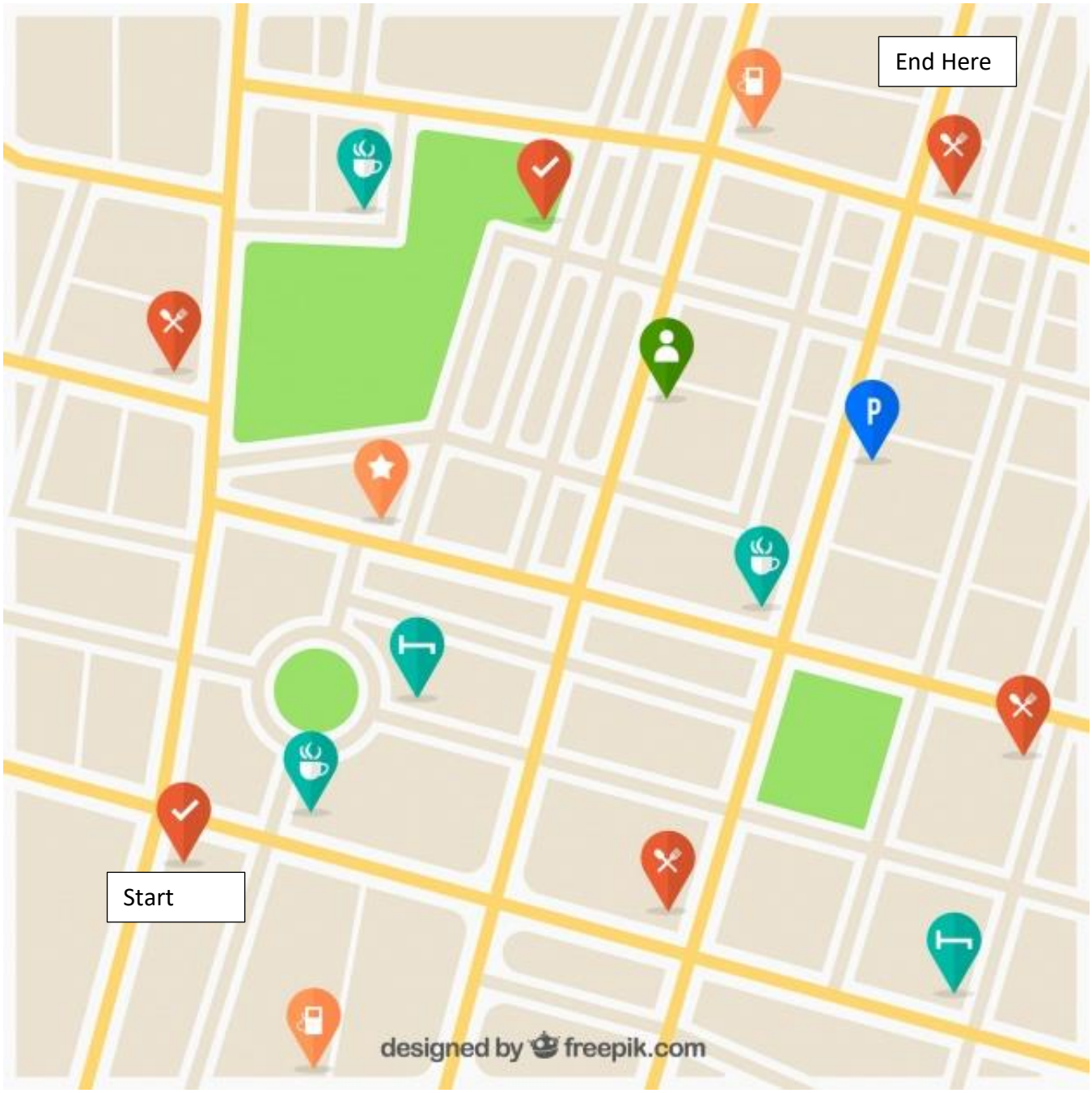
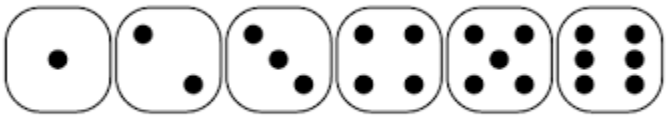


Figure 1 - Example road network diagram to use for discussing your algorithm.
Image attribution: Travel vector created by freepik - www.freepik.com

Understanding probability - Exercises

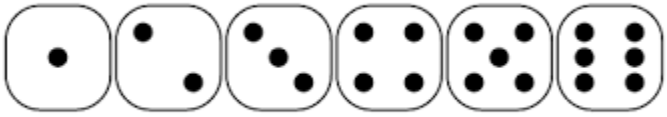
Q1. How many different possible values are there for a standard six-sided dice?



A:

Q2. How many values of a six-sided dice are less than 4?

Tip: Circle the values that are less than 4.



A:

Q3. What is the probability of rolling a value less than 4 on a standard six side dice?

Tip: Count the number of “right” values and divide by the total number.

$$p = \frac{Number\ of\ "right"\ values}{Total\ possible\ values}$$ A:

Q4. What is the probability of rolling at LEAST a 2 on a standard six side dice?

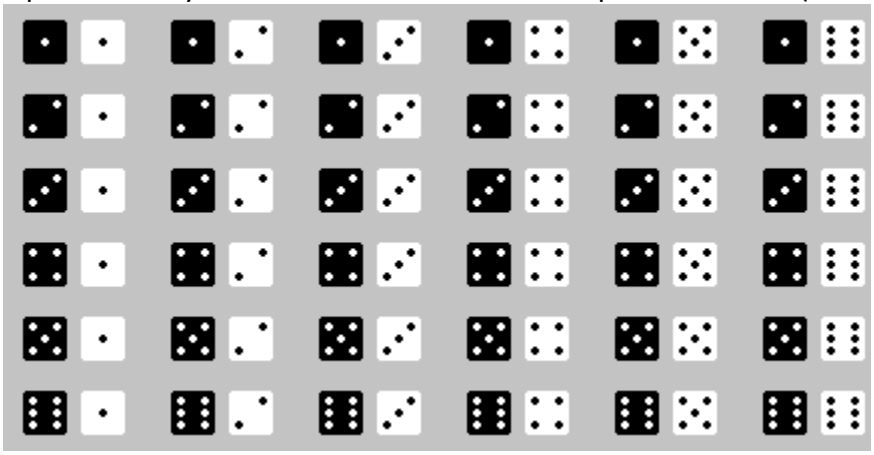
A:

Q5. What is the probability of rolling a value less than 5 on a standard six side dice?

A:

Q6. How many possible outcomes are there for rolling two fair six sided dice?

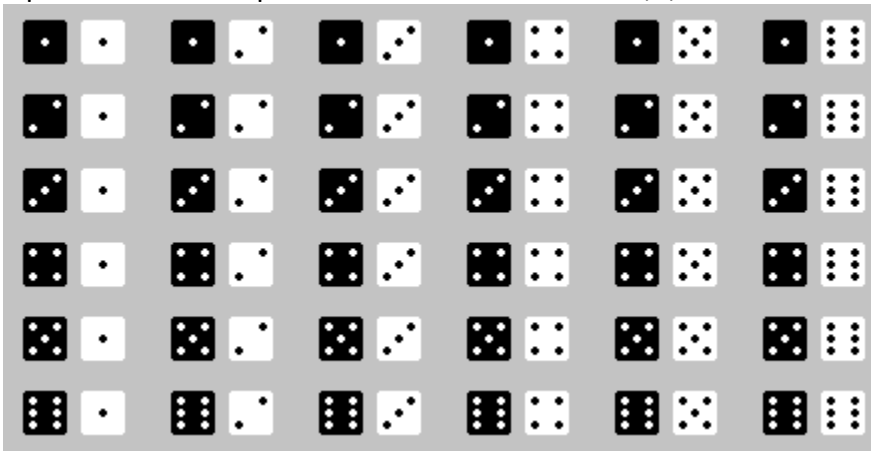
Tip: How many combinations are there in the picture below? (one dice is black, the other white)



A:

Q7. How many values (sum of the two dice) are less than 6?

Tip: circle all of the pair of dice rolls that sum to 2,3,4 or 5



A:

Q8. What is the probability of rolling a combination of less than 6?

Tip: Count the number of "right" values and divide by the total number.

$$p = \frac{Number\ of \text{correct}\ values}{Total\ possible\ values}$$

A:

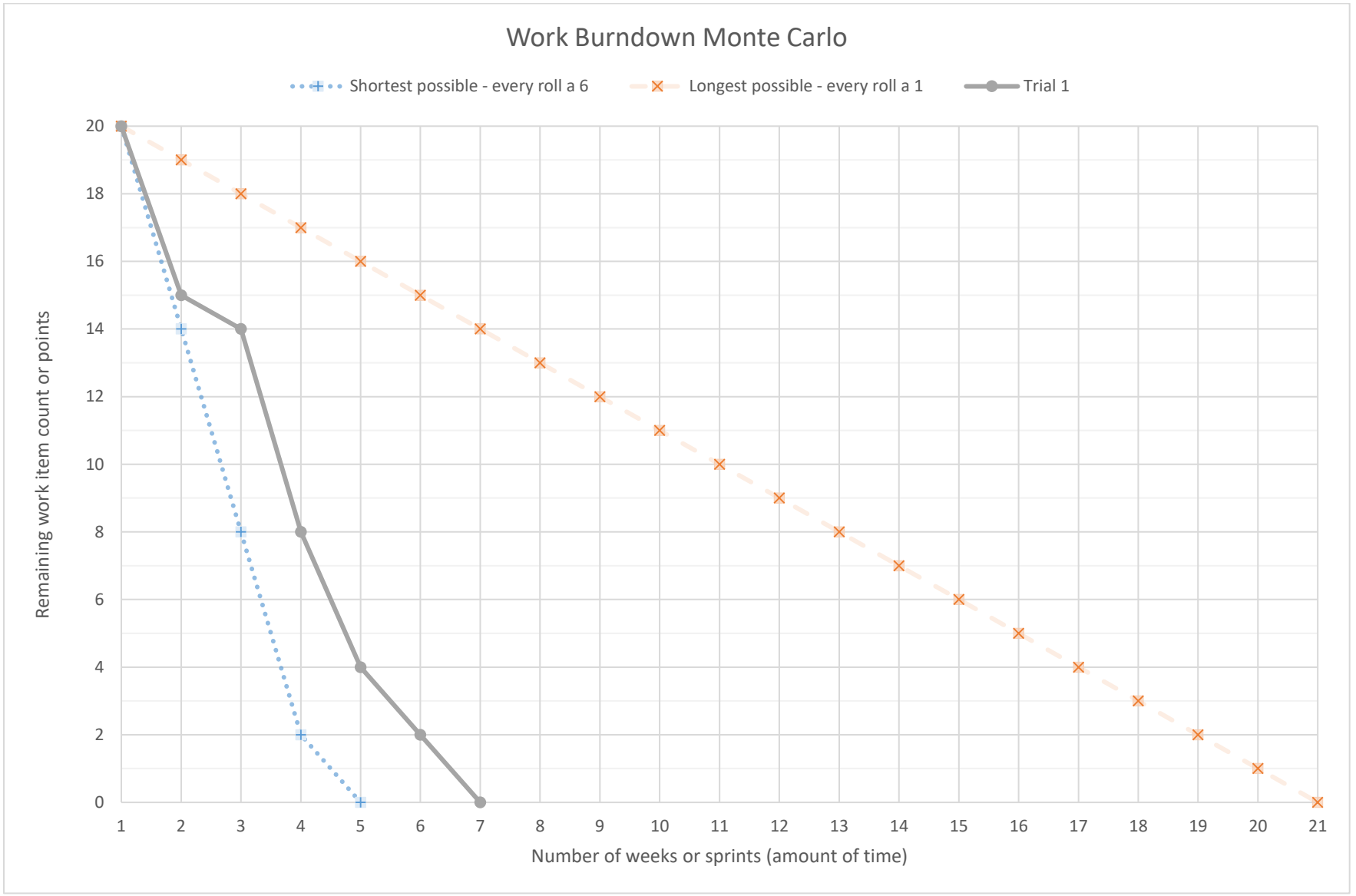
Answers: Q1: 6 Q2: 3 Q3: 3/6 = 0.5 Q4: 5/6 = 0.83 Q5: 4/6 = 0.67 Q6: 36 Q7: 10 Q8: 10/36 = 0.278

Basic Monte Carlo Forecasting – Manually plot work completion from dice rolls of throughput/velocity (1 to 6)

Discover what Monte Carlo forecasting is by performing it by hand. This exercise simulates completing a project many times and plots the outcomes. Perform 7 more trials. Each trial involves filling all rows in a column until the remaining work count reaches zero.

1. Throw a six-sided dice and subtract the number in the row above by this dice roll.
2. When a column reaches zero (or less, just enter 0), move onto the next trial column.
3. Plot each trial as a line graph on the following page. Trial 1, has already been plotted for you.

Week	Trial 1 5,1,6,4,2,2	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7
1 (start)	20	20	20	20	20	20	20
2	-(5) = 15						
3	-(1) = 14						
4	-(6) = 8						
5 (shortest)	-(4) = 4						
6	-(2) = 2						
7	-(2) = 0						
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21 (longest)							



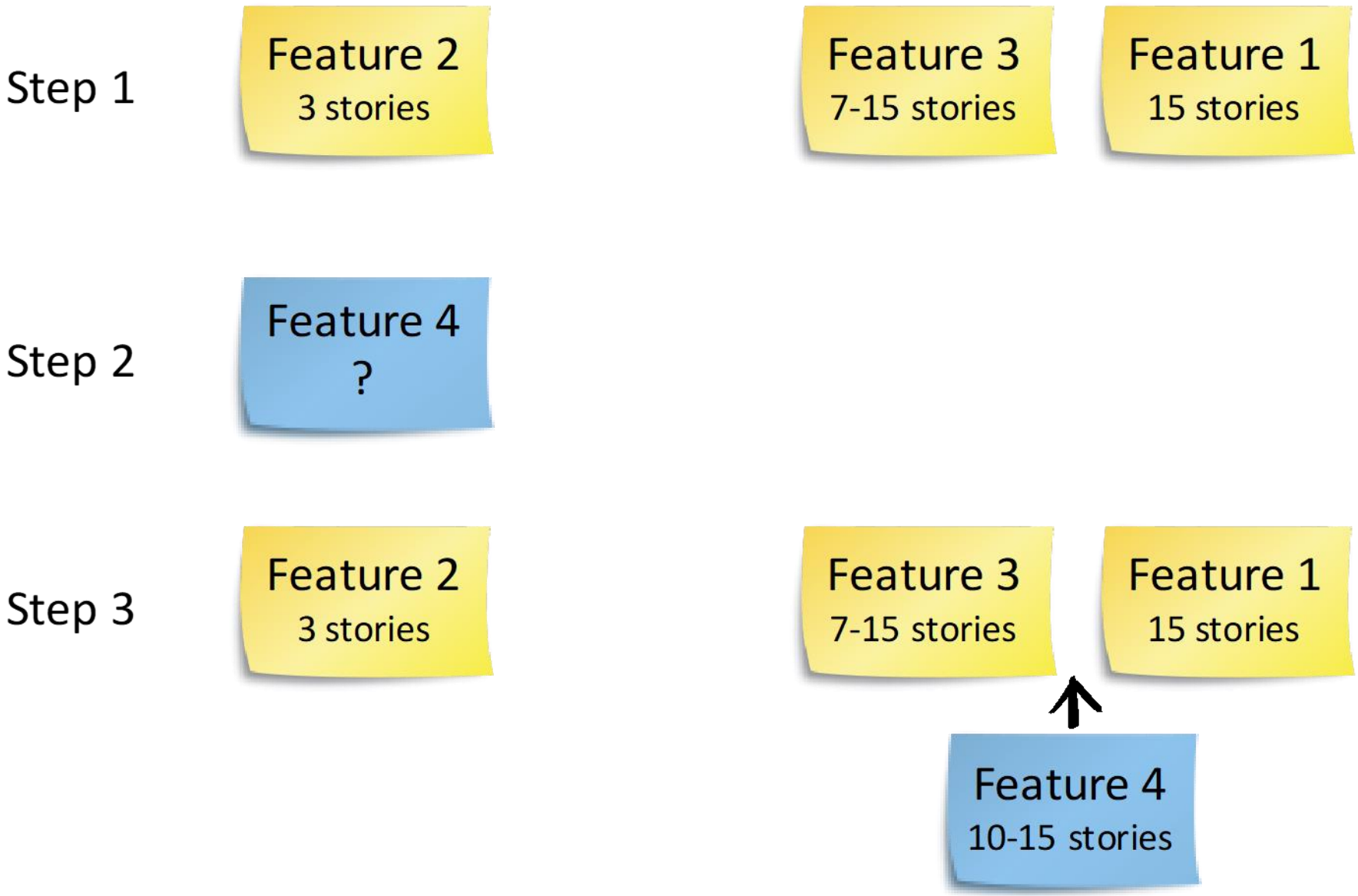
Circle YOUR 4th line counting from the left – this has a 4/7th chance (> 50%) – IGNORE the lowest blue-dotted line (ending at week 5).

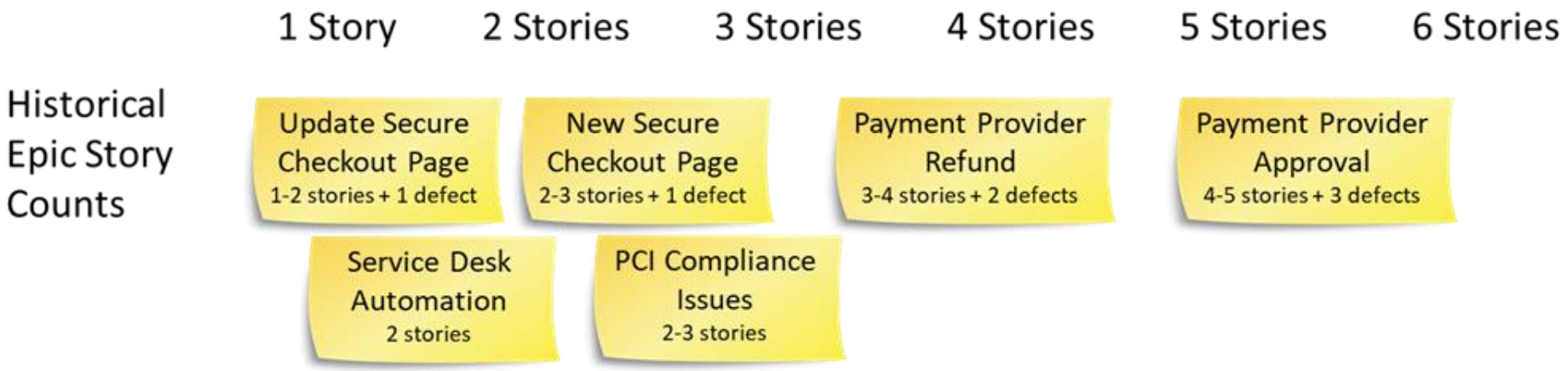
Reflection and Lessons
Topic: Monte Carlo Forecasting

Think about what we just covered, and consider these questions. You don't need to share these with the group. These are for YOU.

Q1: What did you learn about Monte Carlo forecasting from this section?

Q2: What will you do differently in the future using Monte Carlo forecasting?


Step 1
Feature 2
3 stories
Feature 3
7-15 stories
Feature 1
15 stories
Step 2
Feature 4
?
Step 3
Feature 2
3 stories
Feature 3
7-15 stories
Feature 1
15 stories
Feature 4
10-15 stories



Epic Name	Estimate Story Count Range
Checkout page updates: Choose payment card vendor type	2-3
Checkout page updates: Validate card number for types	2-3
Checkout page updates: Capture billing address information	1-2
Support Visa and MC card types (new provider): Visa and MC Approval workflow	2-3
Support Visa and MC card types (new provider): Visa and MC Refund workflow	2-3
Support AMEX cards: AMEX Approval workflow	3-4
Support AMEX cards: AMEX Refund workflow	2-3
Support AMEX cards: AMEX additional fee warning	1-2
Support Bitcoin transactions: Bitcoin Approval workflow	3-4
Support Bitcoin transactions: Bitcoin Refund workflow	2-3
PCI Compliance yearly audit: PCI Compliance Resolution of Major Issues	2-3
Fraud Detection features: US Address fraud detection	2-3
Fraud Detection features: Other country Address Fraud Detection	2-3
Card Expiry Reminders: Three-month before expiry reminder email	3-4
Card Expiry Reminders: Create support desk issue one-month prior	1-2

Total Story Count Monte Carlo Forecast Exercise – Read First

Aim

To practice estimating the total number of stories (or story points) in a larger project by sampling just a few feature or epic story count breakdown examples. The goal is to get an estimate of total story count without having to analyze very feature or epic in a proposed project.

Facilitation

4. Discuss with the group the goal "We are going to estimate how many stories (or points) in 10 features."
5. Discuss the basic method "We are going to randomly build sets of 10 story counts and total how many stories are in each group of 10. We will do this for 11 trials, and these trials will allow us to understand the probability of each result. To save time, you just have to build the first two trials, the others are already done for you."
6. Discuss what the samples are "We will be sampling actual story count examples performed by a team. The samples are counts of 36 prior features, but it could be far less, even as few as 7. You can prove this by crossing out some values in random samples and rolling again if you get one of those scrubbed out samples."
7. Discuss how we learn probability "After we have built and totaled the 11 trials, we will count how many trials rounded down to the nearest ten value for simplicity. These counts will tell us how "probable" that many stories are likely. Likelihood is simply the ratio of how many trials are in a group divided by 11 (the number of trials). When done by spreadsheet we might do 1000 trials, but to save time we are just doing 11"

Questions and discussion topics

5. **What could pollute the story count samples (make them a poor predictor of the future)?**
6. **Why can't we just use the average or median values to forecast the story count in 10 features?**
7. **If life depended on this forecast, how many stories would you sign-up for?**
8. **How might you choose a likelihood to target in your company?**
9. **How would you get more definition in the likelihood percentages?**
10. **What does 100% likelihood mean in this case?**

Why it works

If we sample at random the number of stories in features (or epics) analyzed by teams, then we can forecast the number of stories in any given number of features. The assumption is that the pattern in the samples is representative across the rest of the features not analyzed by the team. This will only be true if the features analyzed by the team are actually chosen truly at random. We help you do this here by rolling a two six-sided dice (or one six-sided dice twice). This technique build trials, actual possible sets of 10 features. Each one will be different, but the pattern in the results helps understand how likely each value is by its ratio across the whole set of trials. This is best done by spreadsheet, but this exercise is for learning.

Resources

This exercise is for learning purposes, don't do it by hand!

Exercise – How Many Stories in 10 Features (or Epics)

Aim: To estimate how many total stories there would be for 10 features (or epics). To understand the probability of achieving those story count estimated based on prior sampled history.

1. **Throw dice and record sample in the empty trial cells below using the sample sheet on the next page**

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10	Trial 11
		2	12	3	12	3	1	3	1	6
		5	15	3	5	10	10	2	6	3
		4	4	3	20	3	3	2	3	2
		3	2	2	2	3	3	3	4	6
		8	6	1	6	6	3	1	2	4
		8	10	8	3	1	2	1	3	12
		1	2	1	3	1	2	5	1	3
		1	1	4	5	3	10	1	1	10
		15	12	3	3	6	6	4	15	12
		6	3	1	8	2	1	6	20	3

2. **Sum each column above and enter the result in the cell below (use your phone calculator!)**

		53	67	29	67	38	41	28	56	61

3. **Round each sum above down to the nearest "ten" e.g. 10+, 20+, 30+, 40+, etc.**

		50+	60+	20+	60+	30+	40+	20+	50+	60+

4. **Compute the probabilities of achieving each result by counting the trials in each group**

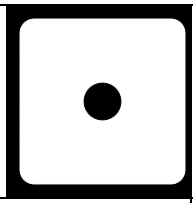
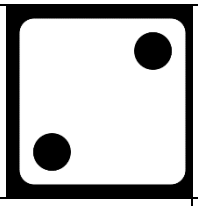
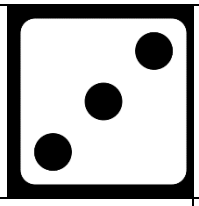
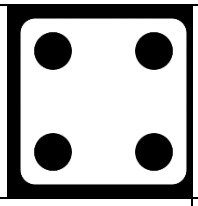
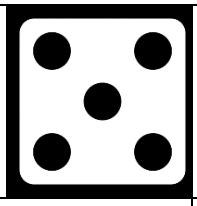
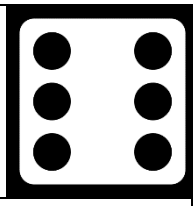
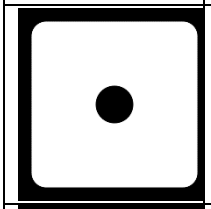
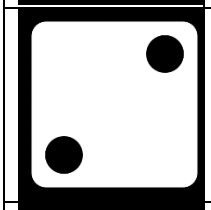
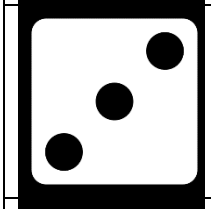
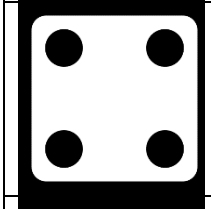
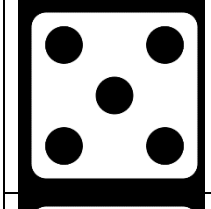
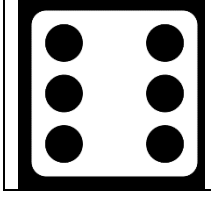
Total Stories in 10 Features (or epics)	Count trial sum groups at least 30,40, 50, etc. stories	(Count / 11) Likelihood
At least **20** stories	11	1
At least **30** stories		
At least **40** stories		
At least **50** stories		
At least **60** stories		
At least **70** stories		
At least **80** stories		
At least **90** stories	0	0

This value is 0 to 1 Multiply it by 100 to get a percentage. 0% = no chance, 100% means every trial achieved at least

Samples: Random Samples of Epic to Story Count

To generate random samples from the story count history, throw two six-sided dice (or throw one six-sided dice twice) and use the sample value at the intersection of the two dice results. It's important to make certain samples are taken at random, and using a dice is often the fairest way to ensure you don't introduce bias!

First dice throw

						
	2	5	1	3	8	4
	10	2	20	3	4	1
	2	12	15	5	2	1
	3	1	10	2	3	1
	3	3	8	4	3	1
	6	3	3	6	4	2

Second dice throw

These samples were from an actual project. The team selected 46 epics (features) at random out of 328 and broke them down into story level. They could have stopped at 10 an got the same

Feature or Project Growth Assumption Worksheet

Time Based Growth

The longer we go the more alterations to original scope get added.

What is time based growth?
The longer the time between committing to delivery and actual delivery the higher risk feature requirements will change. These changes can be additional ideas to current plans, or reactions to a competitive market change. This type of growth isn't bad, it's just inconvenient from a forecasting perspective. For companies to be innovative and react fast to change, adapting plans to accept new ideas and latest information is critical.

1. How much more scope?

Release Frequency	Technically Easy	Hard
Cont. – 2 weeks	1x	1.25x
3 – 6 weeks	1.25x	1.5x
7 – 12 weeks	1.5x	1.75x
13 – 26 weeks	1.75x	2x
26+ weeks	2x	4x

Suggested actions -

- Releasing more frequently limits exposure to time growth
- Don't try and eliminate ALL time based growth. Some is healthy, it means recently learnt lessons become incorporated into plans
- Bent Flyberg is a good resource of material on Mega-projects and recommends avoiding Mega-projects altogether!

Rate Based Growth

The more work we complete the more we learn about what we need to do to deliver.

What is rate based growth?
This type of growth comes from actually completing work. Any growth in story count that has a relationship to completed work falls into this category. Defects and rework discovered are examples of rate based growth.

To capture these, have the team brainstorm items that performing work might cause new discovered work. Keeping an ongoing list of items like this from prior projects helps do this more accurately next time.

2.Things we will also need to do

Growth due to...	Occurr.	# Stories
E.g. Defects	100%	1-3
E.g. Localization	20-30%	3-4

Suggested actions -

- Create a rate based growth table for your feature and projects
- Consolidate the rate based growth knowledge across other teams, limit to top 10
- Use actual data to refine the occurrence rate estimate and the impact for growth items.

Scale Based Growth

The size of completed work is different than the work in our backlog. E.g. work splits

What is scale based growth?
This is the most overlooked growth of work. It is more properly a correction rather than growth. It is caused when the pace of delivered items is assumed to be the pace remaining backlog items will be completed. Why isn't this true? Commonly work is split into multiple items when the team is analyzing the details just prior to adding them to their sprint or pulling that work into the team. Left un-adjusted, it looks like the team will complete faster than they will.

3a. Low guess: Items in the original backlog become x items in complete?

1 2 3 4 5 other:
(default)

3b. High guess: Items in the original backlog become y items in complete?

1 2 3 4 5 other:
(default)

Ask the team what is the lowest and highest likely split rate.

Suggested actions -

- Be alert to anytime the backlog count is combined with historical data; The result will often be optimistic.
- Start with the estimate range of 1 to 3 times.
- Measure the actual rate and adjust.

Event Based Growth

Feedback or things that go wrong in the approval to release process.

What is event based growth?
Sometimes we have a hint that something might go wrong requiring rework to fix. For example, sometimes you build a feature but have little insight to how it will perform with thousands of users. Until it's built and tested under stress, you can't know for certain that it isn't going to need improvement. This is an event based scope increase. Just one risk can make or break a good software forecasts.

4.Things we might need to do

Risk	Prob.	# Stories
E.g. Performance	50-75%	20-30

Suggested actions -

- Pick the top five. If you have more than this, forecasting is pointless, you are guessing. Don't start!
- There will ALWAYS be a few un-avoidable risks. Do these earlier in the project to avoid late surprises..
- Don't spend all your time on total risk avoidance, accept some will occur and forecast accordingly.

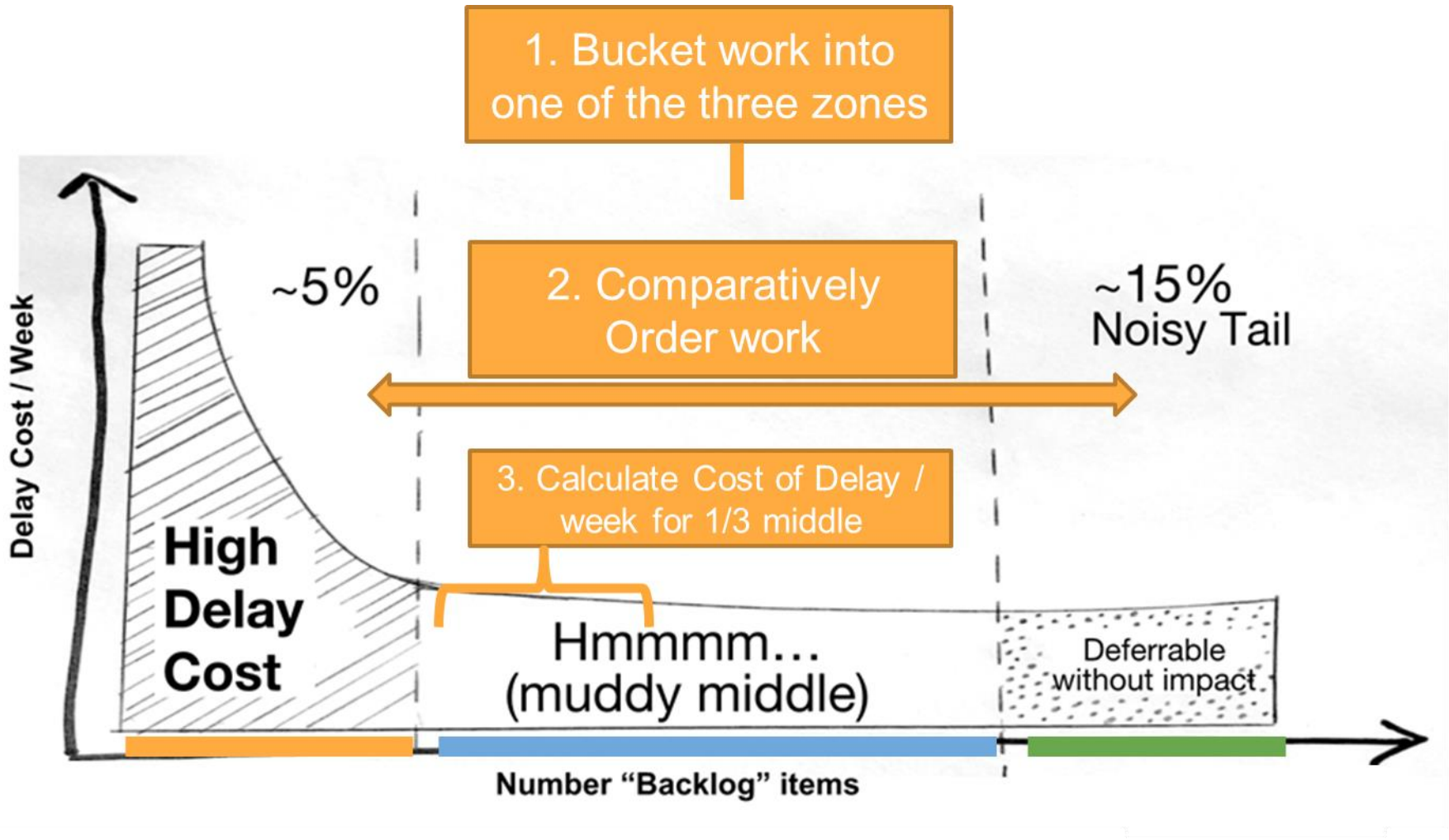
Reflection and Lessons
Topic: Size Estimation

Think about what we just covered, and consider these questions. You don't need to share these with the group. These are for YOU.

Q1: What did you learn about estimating the size of features?

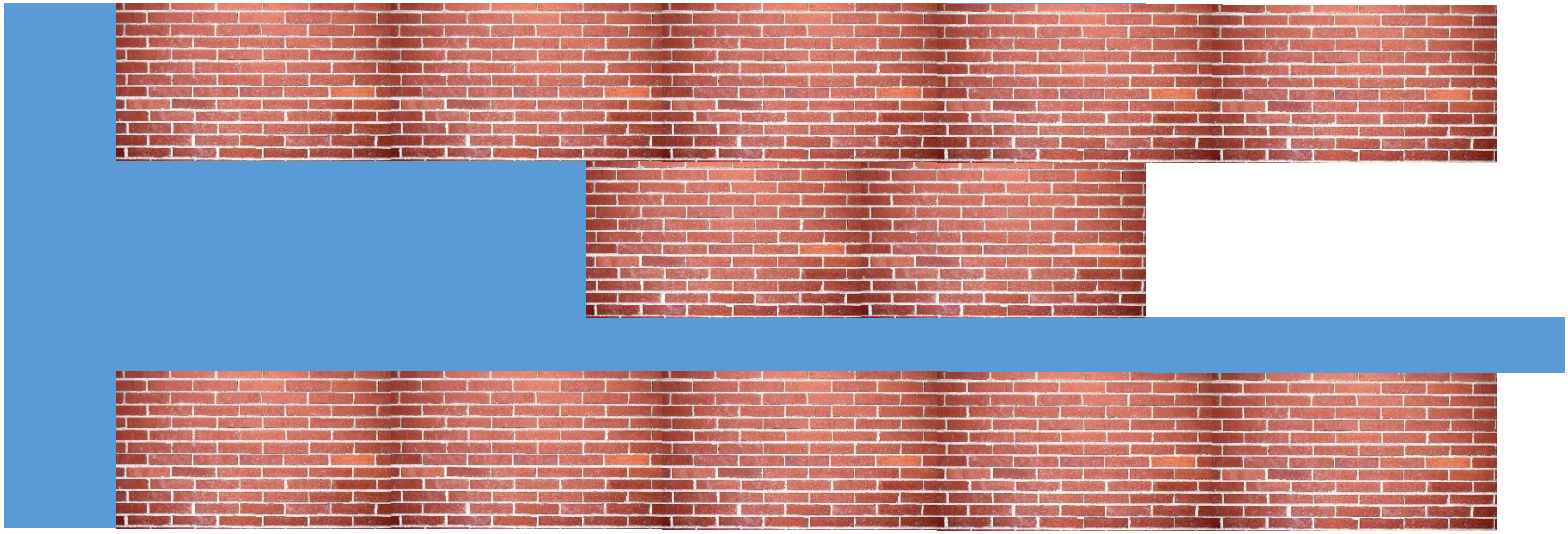
Q2: What will you do differently in estimating?

Rapid Prioritization



Throughput Estimation Constraint Rate

- Big errors occur due to inflated throughput or velocity
- Find the system constraint rate, and use that



List YOUR constraints to flow –

Feature or Project Delivery Pace Assumption Worksheet

Ramp-up / Starting Pace
The pace as the team is forming and learning. Often understrength in skills.

Stride Pace = Sustainable completion pace as a team
The sustainable pace delivered as a team. Often limited by flow through a system constraint. Determine where the constraint will be and estimate low and high completion rate through that step.

Ramp-down / Delivery Pace
The pace as the team is in the final delivery phase into production environment.

Start with FIRST 20% of the original work completed at ½ pace and adjust

Start with LAST 20% of the original work completed at ½ pace and adjust

Backlog

Work Completed

Things that DECREASE ramp-up
- Existing team
- Similar recent work types
- Pairing and sharing skills

Things that INCREASE ramp-up
- New team (> 1/3 people new)
- New type of work
- New innovation or technology

1. Where is the likely system constraint limiting flow?

2. How many people do we have who can work at the constraint [1]?

3. How many items per week/sprint can each person (or team) [2] deliver at the constraint?
Lowest guess: Highest guess:

6. How long will ramp-up take?
0% 10% 20% 30% 40%
(default)

4. Low pace = [2] x [3 low]

5. High pace = [2] x [3 high]

7. Pace impact during ramp-up?
0.1 0.25 0.5 0.75 none
(default)

Delivery pace over time

Things that DECREASE ramp-down
- Continuous automated delivery
- Early focus on quality & *done*
- Internal authority to release

Things that INCREASE ramp-down
- Late integration testing
- Batch steps: e.g. Localization
- External authority to release

8. How long will ramp-down take?
0% 10% 20% 30% 40%
(default)

9. Pace impact during ramp-down?
0.1 0.25 0.5 0.75 none
(default)

How do we find the constraint?
The constraint is the limiting factor of a system delivering faster. If you can't observe a system to visibly see where work is queuing (a great indicator of a constraint), look for where specialist skills are needed and you expect to have too little. I often ask what skill would you add more of to increase flow - that's generally the constraint!

Why do we only estimate at the system constraint?
Every system will have a constraint. The speed that work enters the constraint, and the speed that work completes after the constraint doesn't impact total system pace – system pace is the same as the pace at the constraint.

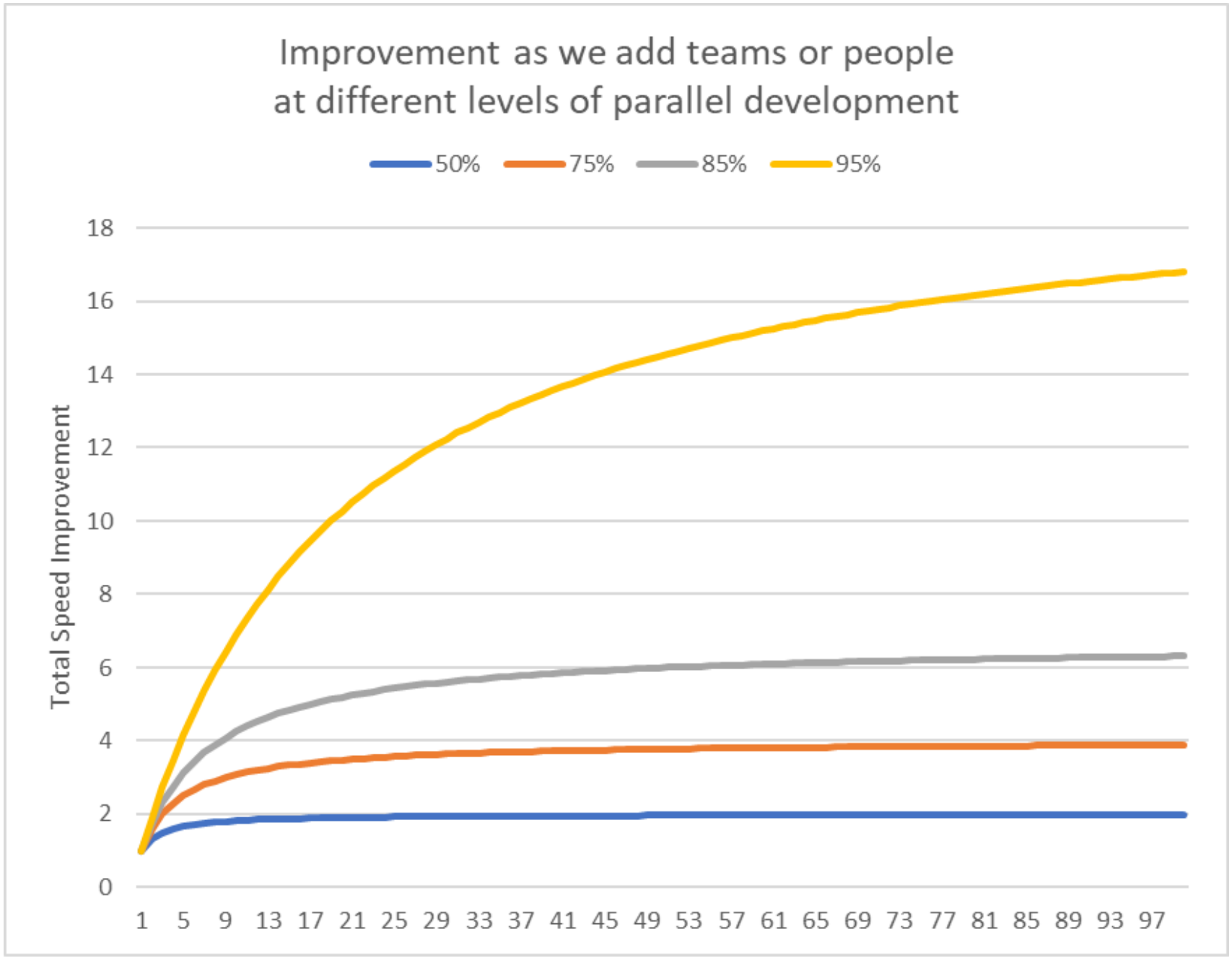
What if there are multiple possible system constraints?
It can be hard to predict where the limiting system constraint will be. If this is the case, brainstorm and estimate the top 3 most likely, then average the low estimates for an average Low Pace for them. Do the same for the High Pace estimate.

What about using actual data?
Use actual throughput or velocity data as soon as possible to confirm these assumptions. Remember, the early samples will be the in ramp-up phase, and be slower than the expected stride pace by the factor estimated in box [7] above. Adjust using the formula - *Stride pace = (1 / box [7]) x Measured Pace*

What values should I use for average pace?
If you are using a forecasting tool that doesn't support multiple feature/project phases, an overall average low and high pace can be calculated using the following formulas (note: [6] = value from box 6)-

X = Ramp up interval = [6]/100
Y = Ramp down interval = [8]/100
Z = Stride interval = 1 – (X + Y)
Up avg = X x ([7] x [4])
Stride avg = Z x [4]
Down avg = Y x ([9] x [4])

Low Average = Up avg + Stride avg + Down avg
(Repeat for the high, replace [4] with [5])


Improvement as we add teams or people
at different levels of parallel development
50%
75%
85%
95%
Total Speed Improvement
0
2
4
6
8
10
12
14
16
18
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97

Reflection and Lessons
Topic: Estimating Size and Pace

Think about what we just covered, and consider these questions. You don't need to share these with the group. These are for YOU.

Q1: What did you learn about size and pace estimation from this section?

Q2: What will you do differently when estimating size and pace in the future?

Throughput Forecast Monte Carlo – Read First

To estimate the number of stories that will be completed by a team for a six (6) week timespan using historical weekly throughput samples from the same team. To understand the probability of achieving those estimates.

The process

1. **Simulate one possible result**: A single six-week throughput result is simulated (called a trial) by summing together six historical one-week throughput samples picked at random.
2. **Repeat**: This simulation process is repeated many times (eleven here, but it can be thousands of repetitions). Each trial represents a "possible" six-week throughput result given the team's historical rate of delivery.
3. **Calculate likelihoods**: The proportion of trials that meet or exceed a given throughput value versus the total number of trials is the likelihood that value is achievable in the future.

Questions and discussion topics

11. **What could pollute the throughput samples (make them a poor predictor of the future)?**
12. **How might you correct for these sample pollution events?**
13. **Why can't we just use the average or median values to forecast the next six weeks?**
14. **If life depended on this forecast, how many stories would you sign-up for?**
15. **How might you choose a likelihood to target in your company?**
16. **How many trials were needed before the actual average (57.75) was included in the range you saw?**
17. **How would you get more definition in the likelihood percentages?**
18. **What does 100% likelihood mean in this case?**
19. **How would you track progress against this forecast?**
20. **What is the impact of not returning the sample each time?**

Why it works

Historical throughput data for teams measures delivery rate for a wide portion of the development system (the wider the better). Team throughput per week accounts for delays; for example waiting time, impediments, staff availability, interruptions and un-recorded work. The impact of these delays is more significant to a forecast than the hands-on time alone. This is a reason developer estimates are unreliable when forecasting projects, they don't account for delays and system dynamics. In a stable system (e.g. the team isn't blown-up), throughput will be a good predictor of future delivery rate even with large item size variability.

Team Throughput Sample Data

Samples represent the number of stories completed per week by the same team taken from an actual project.

Samples: 16,3,10,6,19,11,17,17,15,9,11,8,5,13,5,7,8,6,10,10,8,5,5,7

Count: 24 Sum: 231 Minimum: 3 Median: 8.5 Average: 9.625 Maximum: 19

Resources

Forecasting spreadsheets: https://github.com/FocusedObjective/FocusedObjective.Resources (these spreadsheets do the process described here thousands of times instantly. This exercise is for learning purposes, don't do it by hand!)

Exercise – Throughput Forecast Monte Carlo Worksheet

Aim: To estimate the number of stories that will be completed by a team for a six (6) week timespan using historical weekly throughput samples for that team. To understand the probability of achieving those estimates.

Process:

5. Shuffle the 24 throughput cards or dice (whichever method you choose)
6. Pick a card at random or throw dice and record sample in the table below
7. Return the card to the deck and reshuffle ("sample with replacement")
8. Repeat until all squares are filled

We randomly sampled trials 4 to 11 for you to save

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10	Trial 11
			7	11	7	5	17	5	10	16
			19	7	10	5	13	13	5	7
			6	5	5	3	5	16	6	5
			6	19	5	3	5	3	6	3
			5	7	10	5	6	8	8	6
			5	7	19	10	16	8	10	16

9. Sum of all samples for each trial by column (upper) / Nearest "tens" grouping rounded down (lower)

			48	**56**	**56**	**31**	**62**	**53**	**45**	**53**
			40+	50+	50+	30+	60+	50+	40+	50+

10. Probabilities of achieving at least n stories for a six-week timespan

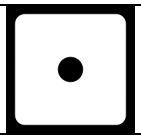
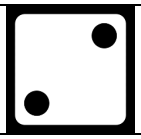
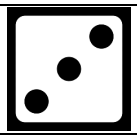
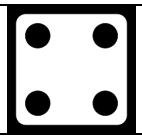
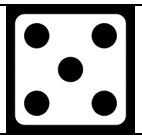
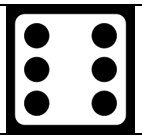
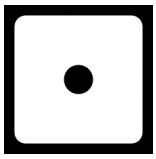
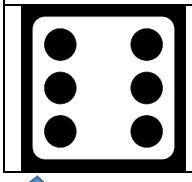
Six Week Throughput	Count trial sum groups at least 30,40, 50, etc. stories	(Count / 11) Likelihood
At least **30** stories		
At least **40** stories		
At least **50** stories		
At least **60** stories		
At least **70** stories		
At least **80** stories		
At least **90** stories		

This value is 0 to 1 Multiply it by 100 to get a percentage. 0% = no chance, 100% means every trial achieved at least this level.

Samples: Random Samples of Throughput by Six-Sided Dice

To generate random samples from the throughput history, throw two six-sided dice (or throw one six-sided dice twice) and use the sample value at the intersection of the two dice results. It's important to make certain samples are taken at random, and using a dice is often the fairest way to ensure you don't introduce bias!

First dice throw

						
	16	3	10	6	19	11
	17	17	15	9	11	8
	5	13	5	7	8	6
	10	10	8	5	5	7
	Roll again	Roll again	Roll again	Roll again	Roll again	Roll again
	Roll again	Roll again	Roll again	Roll again	Roll again	Roll again

Second dice throw

Forecasting Exercise – Replace Payment Provider for Online Website

Pain as a Service (PaaS) is an online provider of self-service pain. It currently accepts Visa and Mastercard for monthly pain delivered daily to customers internationally.

Recently customers have called the helpdesk saying they can "only" pay by a corporate American Express card, and others saying they "need" to pay by Bitcoin (which an upstart competitor is offering).

Expiring credit cards is an ongoing issue. Auto-bill fails; it's a hassle for us and the customer to resolve. Proactively telling the customer would likely eliminate these lapses in service and loss of customers.

A new credit card payment provider is offering less fees (3% down to 2%, 1% less per transaction), and the new provider offers better fraud detection which is good because a major increase in overseas customers (20% total and growing) has increased the credit-card fraud levels.

Pain is often purchased near the end of the year where budgets needs to be spent or lost. This means calendar Q4 (October to December) is the largest quarter by far. We don't do updates to the billing features in Q4.

To ensure credit card data is securely stored, the current checkout system is PCI Compliant. We require an annual audit which has large fines and business exposure if not performed by end of the year. We always leave this to the last minute.

The exercises

These exercises simulate making decisions using data and intuition. A proposed list of features has been brainstormed and it is your responsibility to answer the two business questions being posed –

- Question 1: Can (and should) we change payment providers before October 1st?
- Question 2: What CAN we get by October 1st (before rush season)?

To answer these questions, you will learn through practice –

- How to use historical size data to estimate feature size quickly (reference class forecasting)
- How to forecast size using sampling techniques to obtain a total size for all features
- How to forecast duration using Monte Carlo techniques
- How to prioritize a set of features using various cost of delay techniques
- How to determine what will hit or miss a target date using Monte Carlo forecasting

You will require the following tools

1. Microsoft Excel 2010+
2. These Spreadsheets downloaded from http://Bit.Ly/ForecastingExercise (case sensitive, capital F and E)
    a. Story Count Forecaster Exercise spreadsheet (1 - Story Count Forecast for New Credit Card Provider and Fraud Detection.xlsx)
    b. Throughput Forecaster Exercise spreadsheet (2 - Throughput Forecast for New Credit Card Provider and Fraud Detection.xlsx)
    c. Multiple Feature Cut-Line Forecaster Exercise spreadsheet (4 - Multiple Cut Line Forecast for New Credit Card Provider and Fraud Detection.xlsx)

Proposed features – The Feature and epic Backlog

The product team has defined the following features and worked with the development team to capture the epic level work.

Table 1 - All proposed epics

Feature Grouping	Epic ID #	Epic Description	# Stories (size) for Epic (ex 1)
Checkout page updates	1	Choose payment card vendor type	
	2	Validate card number for types	
	3	Capture billing address information	
	4	Add security information/logos	
	5	Add chargeback fee disclosure wording	
Support Visa and MC card types (new provider)	6	Visa and MC Approval workflow	
	7	Visa and MC Refund workflow	
Support AMEX cards (new provider)	8	AMEX Approval workflow	
	9	AMEX Refund workflow	
	10	AMEX additional fee warning	
Support Diners Club card type (new provider)	11	DC Approval workflow	
	12	DC Refund workflow	
Support Bitcoin transactions	13	Bitcoin Approval workflow	
	14	Bitcoin Refund workflow	
PCI Compliance yearly audit	15	PCI Compliance Audit	
	16	PCI Compliance Resolution of Major Issues	
Fraud Detection features	17	US Address fraud detection	
	18	Other country Address Fraud Detection	
Card Expiry Reminders	19	Three-month before expiry reminder email	
	20	Create support desk issue one-month prior	

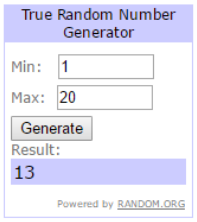
Question 1: Can we change payment providers before October 1st?

Step 1 is to get an estimate of size. Step 2 is to see if the teams involved have capacity to deliver before October.

Step 1: Estimating Size – Total Story Count for All Epics

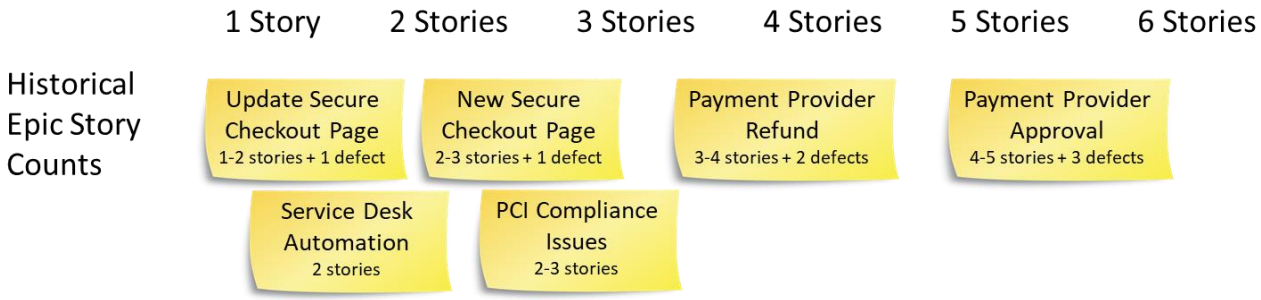
We are going to avoid estimating every epic by extrapolating how big all features are by random sampling and estimating only five epics.

Exercise 1



1. Select 5 epics at random. Go to the website https://www.random.org/ and use the tool to generate 5 non-duplicate numbers between 1 to 20, match this to the epic #'s shown in Table 1 - All proposed epics.
2. For the randomly selected 5 epics, estimate the number of stories
    a. Read the epic description and choose a similar epic from those previously completed show below in Figure 1.
    b. Decide where "this randomly selected epic" fits relative to these, and enter the count in Table 1. Ignore the defect count for the moment, we'll incorporate that later. If there is no prior similar epic, guess intelligently (or throw a six-sided dice, often similar)
3. Enter these samples into the "1 - Story Count Forecast for New Credit Card Provider and Fraud Detection.xlsx" spreadsheet and forecast total story count for 20 epics as shown in my example Figure 2.

Figure 1 - Historical Story Count Data for Reference Class Forecasting



Example random set of five epics and my thought process

My randomly chosen epics were epic id's 17, 1, 7, 2 and 20. Here is the logic used in choosing the story count estimate-

17 – no history, about 4 stories. Proposed simple test address is valid and we have confirmed email

1 – update to current page, about 2 stories

7 – refund, common type, 3 stories

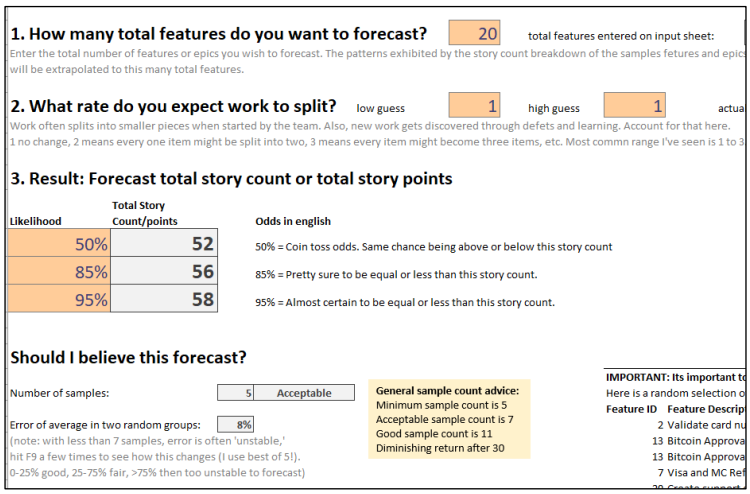
2 – validation logic might need to be learnt, 2 stories

20 – nightly job to check card expiries and if about to expire create service desk issue, 2 stories

Enter these estimates into the Story Count Forecaster (see "1 - Story Count Forecast for New Credit Card Provider and Fraud Detection.xlsx") and forecast total story counts for 20 epics as shown in Figure 2.

Figure 2 – Example, my 5 estimates in the total story count forecaster spreadsheet

Epic ID (optional)	Feature or Epic Name	Estimated # Stories or points
1	Choose payment card vendor type	2
2	Validate card number for types	2
3	Capture billing address information	
4	Add security information/logos	
5	Add chargeback fee disclosure wording	
6	Visa and MC Approval workflow	
7	Visa and MC Refund workflow	3
8	AMEX Approval workflow	
9	AMEX Refund workflow	
10	AMEX additional fee warning	
11	Diners Club Approval workflow	
12	Diners Club Refund workflow	
13	Bitcoin Approval workflow	
14	Bitcoin Refund workflow	
15	PCI Compliance Audit	
16	PCI Compliance Resolution of Major Issues	
17	US Address fraud detection	4
18	Other Country Address Fraud Detection	
19	Three-month before expiry reminder email	
20	Create support desk issue one-month prior	2



Exercise 1 Workshop discussion

Q1. How did reference class forecasting help?

Q2. What might happen if you didn't choose epics at random?

Q3. Discuss the choice of what likelihood value makes sense for this forecast?

Q4. When might we choose to add a story splitting factor (its default is not to split)?

Q5. What was your average error? Add a couple of more samples and see if it reduces.

Step 2: Forecast Duration of All Epics

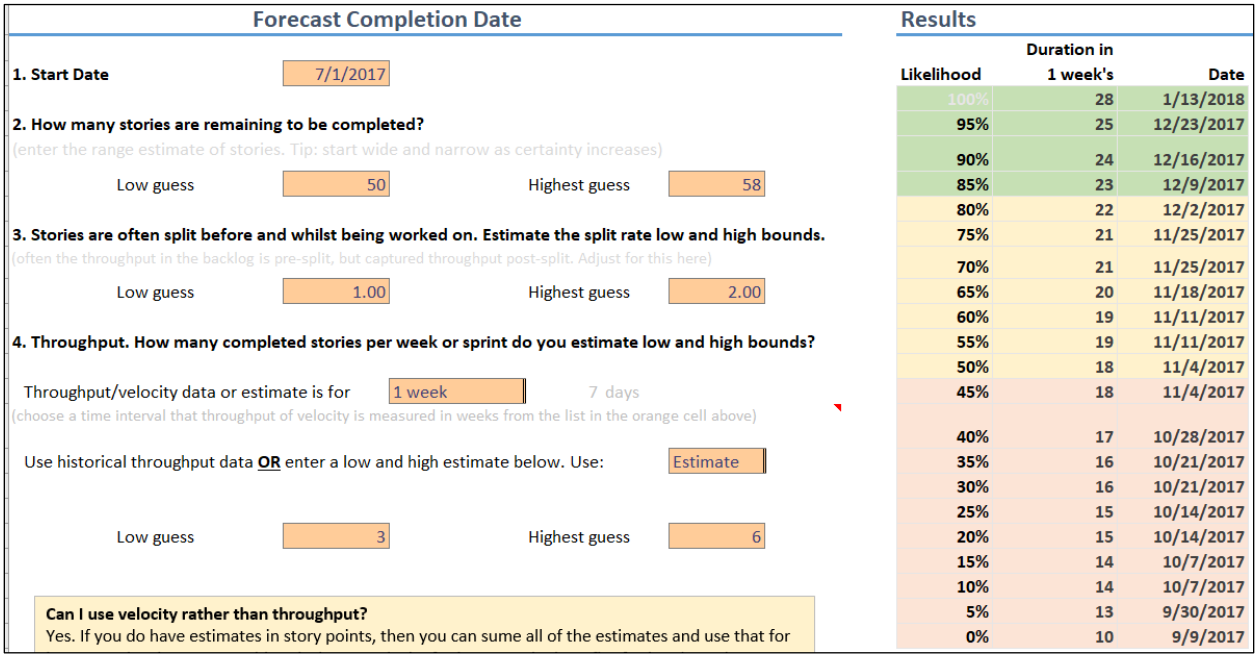
We are going to Monte Carlo forecast using the story count forecast and team historical performance to compute a duration and a likely delivery date.

Exercise 2

1. Open the Throughput Forecasting spreadsheet ("2 - Throughput Forecast for New Credit Card Provider and Fraud Detection.xlsx")
2. Set the input values on the Forecast worksheet
    1) Start Date is 1st July 2017 (if this date has passed, change the year, this is an exercise). Enter it in input 1.
    2) Story Count is the 50% to 95% range (in my case 50 to 58, but use yours discovered in Exercise 1 if you have it). Enter it in input 2.
    3) Split Rate will depend on whether you want to consider defects in your throughput measurement. If YES then you need to account for this by increasing the split rate from 1 to 2 in input 3, else 1 to 1 (no split).
    4) Throughput rate will be assumed 3 to 6 items per week for input 4. This assumes that the team of three dev/test combinations will get at least one item complete each per week, and sometime two each.

These inputs produced the forecast shown in Figure 3 - Initial forecast for all epics based on my example.

Figure 3 - Initial forecast for all epics based on my example performed by the throughput forecasting spreadsheet



Forecast Completion Date

1. Start Date 7/1/2017

2. How many stories are remaining to be completed?
(enter the range estimate of stories. Tip: start wide and narrow as certainty increases)
Low guess 50 Highest guess 58

3. Stories are often split before and whilst being worked on. Estimate the split rate low and high bounds.
(often the throughput in the backlog is pre-split, but captured throughput post-split. Adjust for this here)
Low guess 1.00 Highest guess 2.00

4. Throughput. How many completed stories per week or sprint do you estimate low and high bounds?
Throughput/velocity data or estimate is for 1 week 7 days
(choose a time interval that throughput of velocity is measured in weeks from the list in the orange cell above)
Use historical throughput data **OR** enter a low and high estimate below. Use: Estimate
Low guess 3 Highest guess 6

Can I use velocity rather than throughput?
Yes. If you do have estimates in story points, then you can sume all of the estimates and use that for

Results

Likelihood	Duration in 1 week's	Date
100%	28	1/13/2018
95%	25	12/23/2017
90%	24	12/16/2017
85%	23	12/9/2017
80%	22	12/2/2017
75%	21	11/25/2017
70%	21	11/25/2017
65%	20	11/18/2017
60%	19	11/11/2017
55%	19	11/11/2017
50%	18	11/4/2017
45%	18	11/4/2017
40%	17	10/28/2017
35%	16	10/21/2017
30%	16	10/21/2017
25%	15	10/14/2017
20%	15	10/14/2017
15%	14	10/7/2017
10%	14	10/7/2017
5%	13	9/30/2017
0%	10	9/9/2017

Exercise 2 Workshop discussion

Q1. Is it a safe bet that this change can occur by the start of October, the busy period?

Q2. When would you need to START this project to be happy to have delivered by 1st October?

Q3. What would the throughput rate need to be to be happy to have delivered by 1st October?

Q4. Was the split rate of 1 to 2 appropriate? (hint: see defect rates in Figure 1)

Q5. How might a better throughput range estimate be created?

Question 2: What CAN we get by October 1st?

This is a multiple step question –

1. What is an appropriate prioritization of these features?
2. How far through this list do we reach, and does that still make sense to try?

Step 1 – determining an economic and viable start order

Exercise 3

Perform a rapid cost of delay assessment and see if the order makes sense -

1. Discuss any dependencies between the features shown in Table 2 – Features and value. Are there any features that NEED to be done first? Which ones?

2. Table 2 – Features and value contains the monthly benefit for each feature. Discuss and look at what order gives the most economic advantage, and why.

Table 2 – Features and value

Feature Grouping	Value added per Month	Start Dependency on another item?	Best Start Order
1. Checkout page updates	$0		
2. Support Visa and MC card types (new provider)	$1,000		
3. Support AMEX cards (new provider)	$5,000		
4. Support Diners Club card type (new provider)	$500		
5. Support Bitcoin transactions	$1000 +		
6. PCI Compliance yearly audit	$1,000 + exposure		
7. Fraud Detection features	$220		
8. Card Expiry Reminders	$2,000		

Exercise 3 Workshop Discussion

Q1. What would you do if it was hard to put a dollar value on things?

Q2. Feature 1 has no value by itself, but necessary. Does this occur often?

Q3. How else could the dependencies be handled? For example, is all feature 1 needed before feature 2-5?

Step 2 – Forecast what could be achieved by the target date

This step will work out how far through the feature backlog we might safely achieve by the last calendar quarter.

Exercise 4

See how much can be achieved in the "optimal" order we just calculated

1. Open the "4 - Multiple Cut Line Forecast for New Credit Card Provider and Fraud Detection.xlsx" spreadsheet.
2. Select the "Forecast" worksheet
3. Enter the following forecast input values
    1) Start date is 1st July 2017 (Use your local date format e.g. 7/1/2017 in the US)
    2) Target date is 1st October 2017 (again in your local date format)
    3) Likelihood: Leave this at 85%
    4) Story split rate. Leave room for some defects and new ideas. Low guess: 1, Highest and guess: 2
    5) Throughput estimates. Let's stick with our assumption with three devs. Low of one story each, 3 total. And a high of two stories each of 6. Make low guess: 3 and Highest guess 6.
4. Confirm that the forecast results are working. Observe the forecast completion dates are sequentially getting later.
5. Set the start order to the order matching your results in Exercise 3. You should see a result similar to that in
6. Figure 4.

Figure 4 - Example forecast results. Green ticks = made it, Red crosses = missed.

				Start date: 07/01/2017	
Start Order	Feature Name (just for reference)	Story Count Low Guess	Story Count High Guess	Forecast Feature Duration in Weeks	Forecast Completion Date (85% CI)
2	Checkout page updates	8	10	4 ✓	8/19/2017
6	Support Visa and MC card types	7	9	4 ✗	11/25/2017
3	Support AMEX card types	7	10	4 ✓	9/16/2017
7	Support Diners Club card type	5	8	4 ✗	12/23/2017
5	Support Bitcoin transactions	8	10	4 ✗	10/28/2017
1	PCI Compliance	5	7	3 ✓	7/22/2017
8	Fraud Detection	6	8	4 ✗	1/20/2018
4	Expiring Card Reminder	3	5	2 ✓	9/30/2017
9				0 ✗	1/20/2018
10				0 ✗	1/20/2018

Exercise - Fill in the input values for 1, 2, 3, 4, 5 to match your prior forecast assumptions.

Now, set the start order to the priority calculated earlier. See what features make the deadline.

Legend

✓ Forecast on or before the target date

! Forecast misses target date by one Week or less

✗ Forecast misses target date by MORE than one Week

Exercise 4 Workshop Discussion

Q1. Discuss the results? What was the last safe feature that made the target date?

Q2. Double the delivery rate. Does everything make it now?

Q3. If you were to increase team capacity, how might you achieve it?

Figure 5 - Here is my complete cut line forecast showing all of the inputs used.

Feature Cut Line Forecaster and Explorer

Only edit orange input cells like this

1. Start Date 7/1/2017 **2. Target Date** 10/1/2017 **3. Likelihood** 85%

4. Stories are often split before and whilst being worked on. Estimate the split rate low and high bounds.

Low guess 1.00 Highest guess 2.00

5. Throughput. How many PLANNED (post split) completed stories do you estimate low and high bounds?

Throughput/velocity data or estimate is for Week 7 days

(choose a time interval that throughput of velocity is measured in weeks from the list in the orange cell above)

Use historical throughput data **OR** enter a low and high estimate below. **Choose here:** Estimate

Low guess 3 Highest guess 6

7. Enter the features and story count estimates here...

				Start date: 07/01/2017	
Start Order	**Feature Name (just for reference)**	**Story Count Low Guess**	**Story Count High Guess**	**Forecast Feature Duration in Weeks**	**Forecast Completion Date (85% CI)**
2	Checkout page updates	8	10	4 ✓	8/19/2017
6	Support Visa and MC card types	7	9	4 ✗	11/25/2017
3	Support AMEX card types	7	10	4 ✓	9/16/2017
7	Support Diners Club card type	5	8	4 ✗	12/23/2017
5	Support Bitcoin transactions	8	10	4 ✗	10/28/2017
1	PCI Compliance	5	7	3 ✓	7/22/2017
8	Fraud Detection	6	8	4 ✗	1/20/2018
4	Expiring Card Reminder	3	5	2 ✓	9/30/2017
9				0 ✗	1/20/2018
10				0 ✗	1/20/2018

8. Month Throughput Adjustment (increase or decrease throughput

Month	Multiplier	Why? (add a comment with the assum
January	1.0	
February	1.0	
March	1.0	
April	1.0	
May	1.0	
June	1.0	
July	1.0	
August	1.0	
September	1.0	
October	1.0	
November	1.0	
December	1.0	

Exercise - Fill in the input values for 1, 2, 3, 4, 5 to match your prior forecast assumptions.

Now, set the start order to the priority calculated earlier. See what features make the deadline.

Legend

✓ Forecast on or before the target date

! Forecast misses target date by one Week or less

✗ Forecast misses target date by MORE than one Week

Exercise 5

Discuss as a group any month adjustments you might need to make and see the impact.

1. For the monthly adjustments (input area 8 on the spreadsheet as shown in Figure 5), consider each month and update the multiplier to decrease (or increase) throughput that month.

Exercise 5 Workshop Discussion

Q1. Discuss the results? What was the last safe feature that made the target date?

Q2. What factors did you discuss and add? Is this common?

Reflection and Lessons
Topic: Forecasting Capacity

Think about what we just covered, and consider these questions. You don't need to share these with the group. These are for YOU.

Q1: What did you learn about forecasting from this section?

Q2: What will you do differently when forecasting in the future?

Exercise Blocker Clustering

Step 1: Write down reasons work is blocked/idle in your organization and teams
Step 2: Estimate the occurrence rate and impact
Step 3: Compute the weighting for each blocker (higher is more impactful)

Internal to the team

Blocker cause	**How often** (1 rare, 2 sometimes, 3 often)	**Impact** (1 low, 2 med, 3 high)	**Weight = How often X Impact**

External to the team

Blocker cause	**How often** (1 rare, 2 sometimes, 3 often)	**Impact** (1 low, 2 med, 3 high)	**Weight = How often X Impact**

Technical Risk Examples : Things that cause rework or additional work to be completed before delivery of an item

- Solution performs too slowly and needs to rework to hit releasable performance goals
- Solution uses too much server runtime memory and risks server failure. Rework required to reduce memory needs
- Solution has security vulnerabilities that causes rework to be releasable
- Solution uses too much data storage space and needs rework
- Solution fails to scale across multiple dynamic servers as expected and needs rework to support scaling
- Solution has hard coded configuration and cant be deployed to testing or production environments without rework
- Solution fails to work on different browsers, used to be IE, now think Safari
- Solution breaks previously working features that were thought to be unrelated and those need to be fixed/reworked
- Solution doesn't have required level of production monitoring features and needs rework to move into production environments
- Solution works on test data, but becomes unusable when exposed real customer data, requiring rework

Process Risk Examples: Things that delay work irrespective of the item itself

- Work sits idle, queued before a constraint for some resource that we didn't anticipate
- Images or other assets aren't available to develop a solution
- Test data isn't available to develop a solution
- Un-planned work increases beyond what was anticipated and slows progress
- Team isn't in a position to begin building a solution when planned (physically present, with everything they need to code, not shared)
- Test environments not available when needed

External Risk Examples: Dependencies, things we need from "others" before delivery of an item

- Blocked waiting for external hosting vendor to configure and install servers
- External stakeholder is slow in giving sign-off approval
- Legal / Geopolitical or other regulatory sign-off is slow

Group Exercise – Step 1 Capture Risks

In a group, share and capture stories (real spoken stories) of reasons previous features were delayed. Capture at least one from each team member, then discuss how these might impact future work.

General Risk Description	Impact? (Low/Med/High)

Group Exercise – Step 2 Prioritize Risks, Create Actions

Looking about 3-Months into the future, dot vote what avoidance ideas you will tackle. Create a list of actionable steps and how you will measure success. How many "dots" per person? about the square root of the team size.

How hard to avoid ? (Easy/Med/Hard)	Ideas: How to avoid/lessen this risk?	Dot Votes

Actionable Steps	Who	When	Evidence its working...

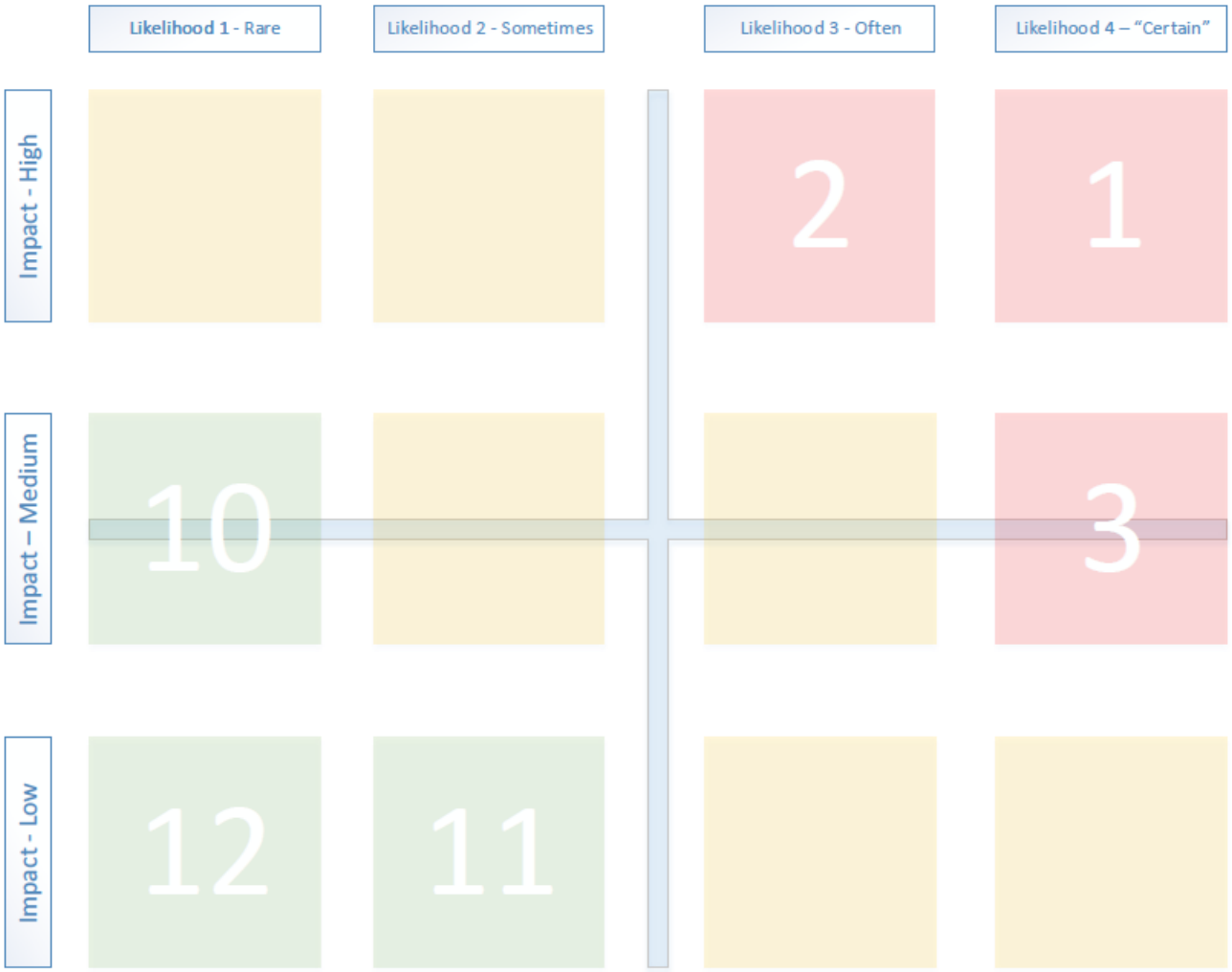
Preferred Order of Action

What risks are most important to deal with first? Deciding which risks are cost effective and beneficial to go after is important, and it's not as easy as you might think. How much "sometimes" turns a "medium" impact into more important than a "rare/high"? It will depend.

Group Exercise

In a group, discuss what order you would address the un-numbered zones.

1. First, make clear the definitions for Likelihood and Impact. Agree on some numerical way to measure these in your context (Document these on the risk canvas).
2. Then, agree on how much impact would equal an increase in likelihood to make it more sense to do one medium impact before a high impact.
3. Fill all of the un-number zones.



Tips and Ideas

1. It's often hard in the un-clear zones to have an exact optimal order. This means that there may not be an exact optimal order! It's OK to best guess.
2. There will be process risks, and item specific risks. Look to understand that some risks will be "certain" for some kinds of work and "Rare" for others
3. A "medium" "sometimes" might be more important than a "high" "often" if it applies to more items in the backlog. Think frequency and exposure
4. For each risk, think what type of backlog item it applies to and how many of them there are. Drop risks that are no longer applicable.
5. Don't solve risks where the impact is less disruptive than the fix. The goal is to not eliminate every risk, just the impactful ones with easy fixes!
6. For each risk, decide on a "Solvability" score – this will help find the easiest solved risks that have big benefit; especially in the unclear zone.

Focused Objective
forecasting - risk - staff - cost of delay

Data Driven Coaching

Capturing and using data for coaching

Workshop Manual

Digital materials: http://Bit.Ly/ForecastingUsingData
Email me: troy.magennis@focusedobjective.com

Version 2.1

“35” Exercise

On an index card, write down your answer to –

“What is one question you would be disappointed if you left this workshop NOT learning or having an answer to?”

Then draw this on the reverse side:

1	
2	
3	
4	
5	
Total	

Five (Data) Commandments

The rules of engagement for the measurement and using data game. Ignore these at your peril.

1: Avoid Embarrassment (Scared people lie)

Key points

- Public shaming = Data Hiding = No (reliable) Insights
- Show insights to the GROUP of people who can fix issues, not those who can't
- There is always missing information, understand the context before taking (premature and mistaken) action

Do

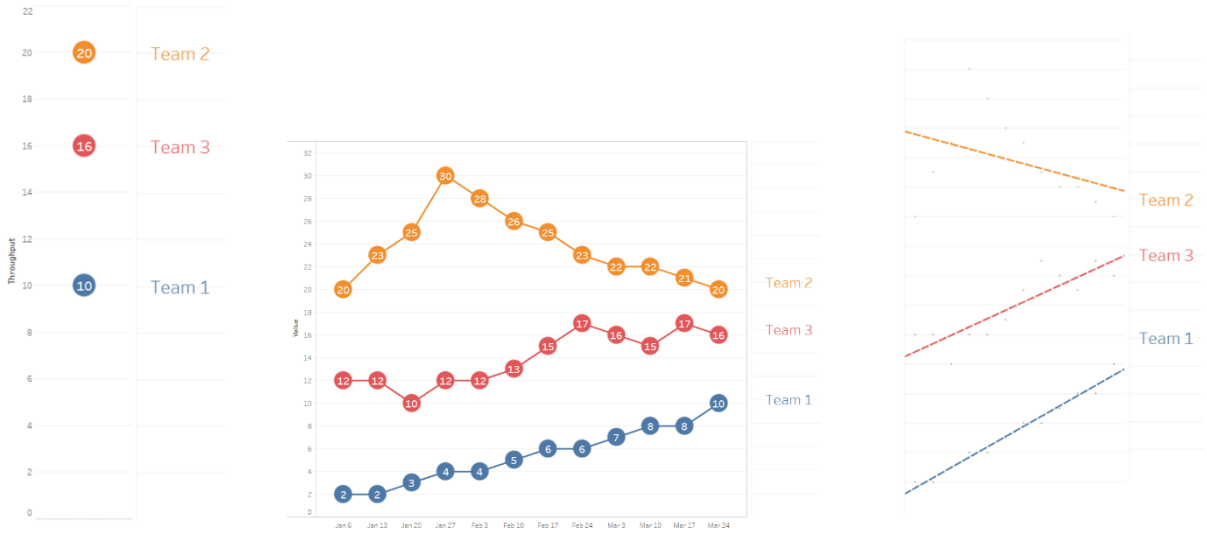
- Celebrate insights from BAD news as much as GOOD news
- Show data aggregated at a level where the action can make a difference (rarely an individual, most often a team)
- Always ask for context – "I see a decline, do we know why?"

How? Just put yourself in the shoes of the target. If you feel uncomfortable, delete that chart. There is no upside to putting people on the spot in public. It's ALWAYS bad. And not just for the person who is subject to the abuse; others watching learn to hide data or game the system to avoid actively avoid it happening to them.

2: Act on Trends; Not Lonely Numbers (it's a matter of direction)

Key points

- Declining or improving patterns cannot be seen from any single data point
- A single "worse" than expected measure may not be bad
- Did it improve from last time (not great yet, but improving)?
- Is it an outlier? Is it an erroneous measurement?
- Is the difference significant enough to matter?



The leftmost chart shows Team 2" doing best using "lonely" numbers, but a longer trend (middle picture) shows Team 2 in a declining trend versus the other two teams. Consider just showing the trendlines to help have conversations about magnitude AND direction - not just higher is better.

Do

- Track data over time to see a change (trend)
- Make insights and take action based on the trend, not value
- Celebrate improving tends, even if not as good as you wanted (yet)

3: So What? Show Data Compared to Something Relevant

Key points

- Compare the actual trend against an expected trend to make insights.
- You need to compare against the RIGHT thing to spot (the right) insights.
    - Don't compare watermelon size with the sizes of cherries
    - A team formed five years ago shouldn't be compared with a team formed last week (less obvious)

Do

- Set expected values based on similar and appropriate prior examples and trends
- Use comparisons to learn – "What is Team X doing that could help Team Y?"
- When you get a false positive ask, "What caused us to think this was bad (or good)?"

4: Highlight Unexpected Insights

Key points

- Avoid "choose your own adventure" graphs (no obvious story/conclusion)
    - If you don't highlight what IS significant, people find their own
- Highlight where expected trend <> observed trend (by a significant amount)
    - The severity of insights helps people take the right amount of action
- Audience: Highlight insights (only) TO those people who can take action

Do

- Highlight unexpected insights in a mind-numbingly unmistakable way
- Show severity in a clear way, don't make all insights equal
- Show insights to people who can fix it, not people who can't

5: Balance Competing Forces - AND not OR

Key points

- Over optimizing ANY one measure causes impact elsewhere
- Being OK across many measurement dimensions > being EXCELLENT at one
- Learn to anticipate the impacts of your actions on your system of measures

Do

- Measure competing dimensions
- Look for early warning signs of overemphasis of any single metric
- Trade from something you are good at to improve something you aren't

In Pairs, discuss

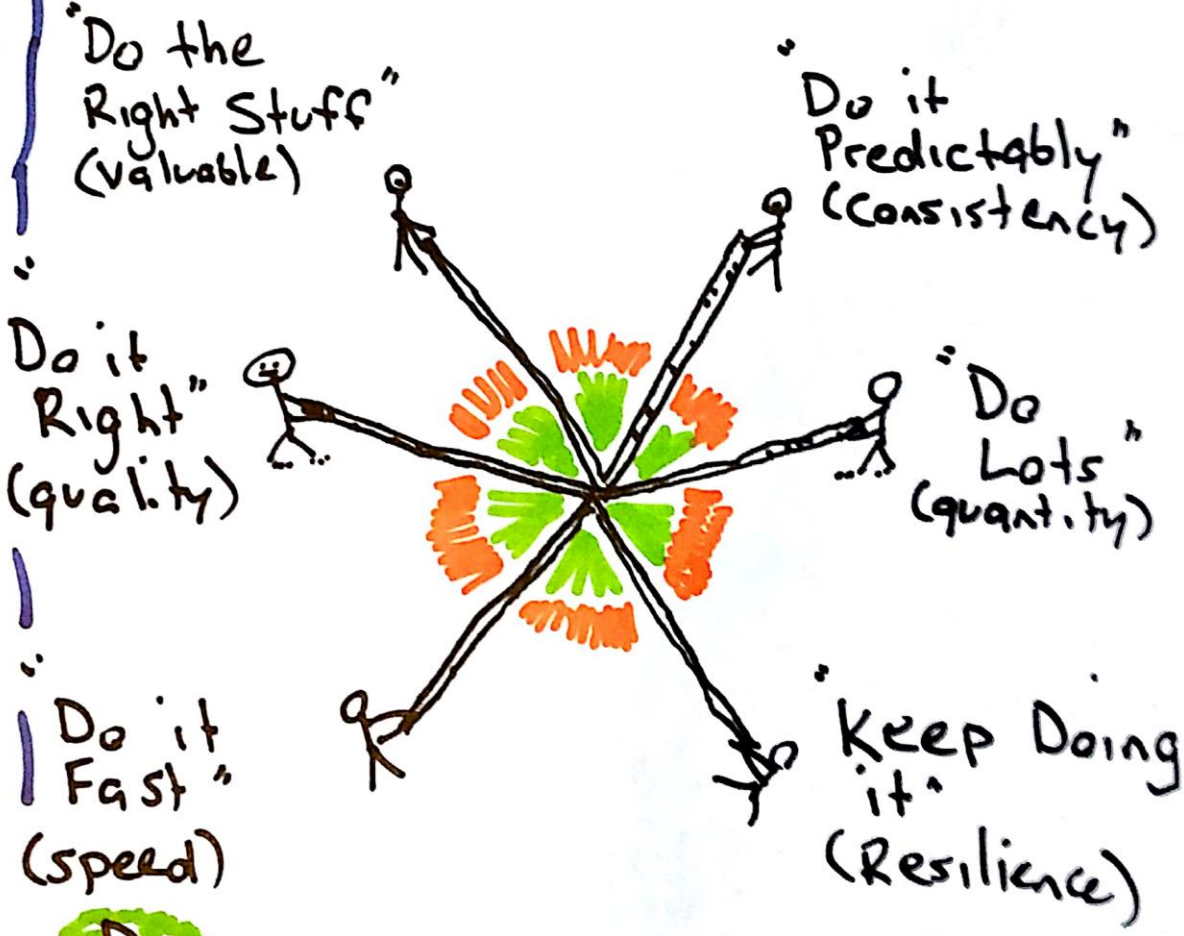
What data do you currently capture about teams and work item process?

5 Minutes

Balance Competing Forces

Help teams find balance across these six competing forces.

Don't over-do any one of these!



Do

- Trade wisely (a strength for a weakness)
- Measure TOO MUCH as well as too little
- Define early warning signs of TOO MUCH/little

Six “Flow” Performance Dimensions

1: “Do Lots” - How Much/Many?

Measure how much raw work product is flowing. In a perfect world, not just in development but to customers. This measure isn’t about customer value; it is evidence that the system is moving items (has flow). This measure is useful for forecasting future delivery.

Too little focus or capability

- Dis-satisfied customers or stakeholders not getting what they need.
- Demand > supply, but you don’t know it.

Too much focus

- Declining quality causing defects and re-dos (not “really” done yet)
- Less valuable “easy” features delivered rather than most needed

Typical metrics in this category

- Throughput - count of items or tickets per day/week/sprint
- Velocity - sum story points per sprint

2: “Do it Predictably” - How consistent is the delivery of value?

Delivery occurs at a consistent pace rather than huge feasts or famine of delivered value to the customers; for example, the variance of pace. This dimension helps see shorter-term process instability (the sustainability metric measures longer-term system stability; it's coming up soon).

Too little focus or capability

- Periods of progress and others of lower value to customers.

Too much focus

- Less risky “known” features delivered rather than most valuable or needed
- Little incentive to push process improvement in case they cause a temporary decline

Typical metrics in this category

- Variability of throughput of velocity
- Variability of the delivered customer value
- Net Process Flow: Things Delivered - Things Started. This measure shows balance through the system with variability represented as a higher or lower peak, with the desired state hovering around zero.

Tip: For a variability measure, consider using the Coefficient of variability: Standard Deviation / Mean rather than the Standard Deviation alone (higher values naturally have a higher Standard Deviation for the same percentage change. Dividing by the mean normalizes that.

3: “Do it Fast” - How fast?

Respond and deliver things quickly, given its complexity and novel-ness. The easiest way to improve this measure is to finish something in-progress before starting something else.

Too little focus or capability

- Customers frustrated in how long it takes to get changes

Too much focus

- Declining quality causing defects and re-dos (not “really” done yet)
- Less valuable “easy” features delivered rather than most needed

Typical metrics in this category

- Time in State - the time an item was within a “state,” for example, “In Development.”
- Cycle time - the time from start to finish at some boundaries in your system
- Lead time - the time from some commitment to delivery (to the person committed too)

4: “Do it Well” - How good was the quality versus expectations?

A measure of how well the delivery of things that solve a problem or need. Often this measure is called Quality and is one of the hardest measures to get a handle on. The goal isn’t purely quality; it serves as an early warning sign that a system is being pushed to deliver beyond its capability.

Too little focus or capability

- Rework. What is delivered needs to be corrected
- Customer dissatisfaction.
- Production issues.

Too much focus

- Little or no delivery of value or flow of items due to “just a little more testing.”
- Slow feedback if the wrong thing is built (albeit perfectly functioning)

Typical metrics in this category

- Escaped defects. Defects found outside of the development and delivery team
- Customer satisfaction. Customers don’t like what you built and tell you
- Production rollbacks. Second and third releases to get a stable, working system
- Unplanned downtime. Issues in production outside of planned change windows

5: “Do Valuable Stuff” - How valuable was it to the customer?

A measure of how much value customers derive from released features or projects. The goal isn’t purely customer value; it serves as an early warning sign that a system is being pushed to focus on work output rather than an outcome.

Too little focus or capability

- Rework. What is delivered needs to be revisited to deliver “more” of this feature
- Customer dissatisfaction. Internal feeling that work is flowing well, but the customers aren’t feeling the value.

Too much focus

- Increasing technical debt. Teams consistently skip technical debt reduction items for supposedly higher value items.
- Lack of prioritization for strategic work that is mid to longer-term (current customers happy, but declining entry into new markets or targets).

Typical metrics in this category

- Cost of delay. An economic view of the cost of NOT doing work to the customer and organization.
- Alignment to strategy. Prioritized work allocation matches a planned strategic allocation
- Customer satisfaction. Customer feedback confirms what was delivered solved a problem with high satisfaction.

6: “Keep Doing It” - How sustainable is the delivery system (and people)?

A measure of how likely the current performance of the development and delivery system can continue in the future. Often called the “happiness” metric, but it's more important than that label describes. When teams push hard on the improvement of the other metrics, it sometimes takes a toll causing a decline in the future. The goal of this metric is to be an early warning indicator of that gloomy performance in the future.

Too little focus or capability

- The current performance measures aren’t maintained.
- The collapse of delivery.

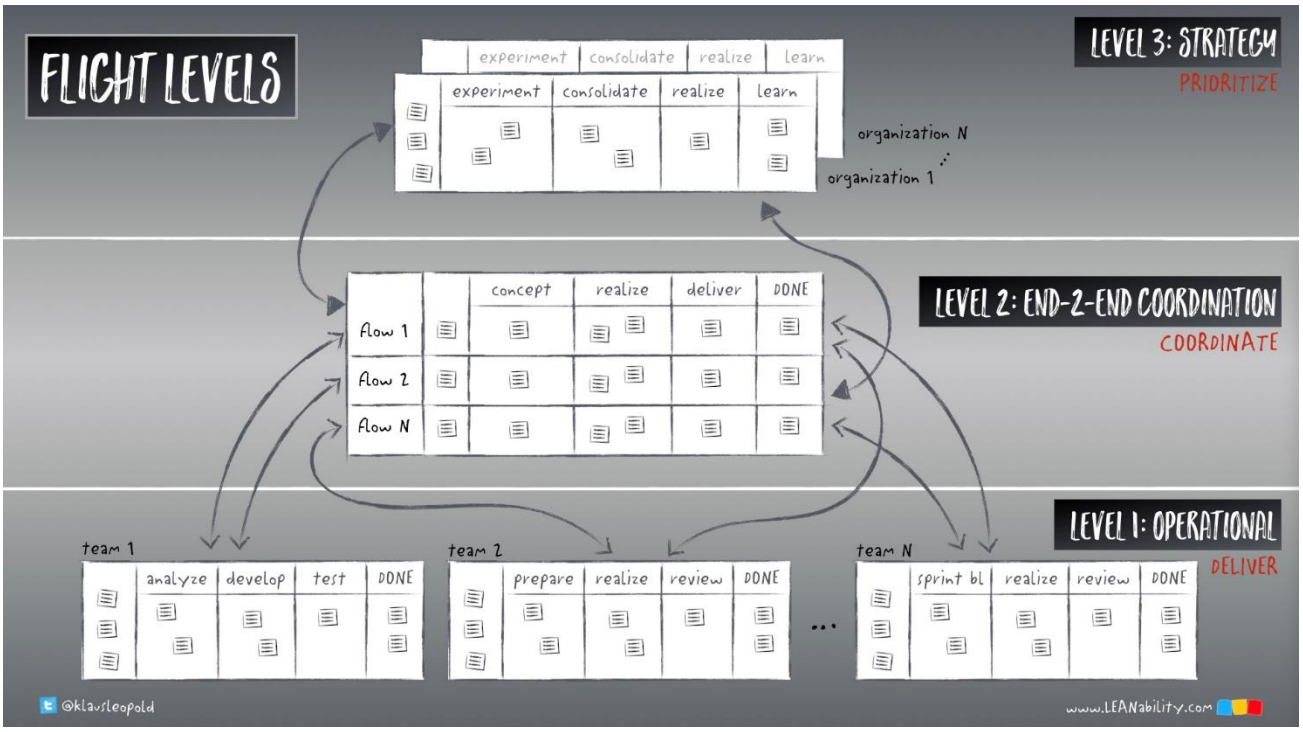
Too much focus

- Stagnate performance improvement over time. The other metrics stay flat.

Typical metrics in this category

- Team health via survey or team retrospective (honest answer to “are we able to continue at this pace?”
- The aggregate of the other performance metrics (metrics listed as 1 to 5 above) metrics

Exercise: Measures and Counter-measures



In Groups, for each dimension, for each level, brainstorm a measure (and a counter-measure) Create sticky-notes to present back to the group.

For team level (FL1) 10 min
For co-ordination level (FL2) 10 min
For portfolio level (FL3) 10 min
(you can capture on the next three pages)

Then, 15 minutes group debrief

Flight Level 1 – Operational/Team Metrics

Brainstorm measures for each performance dimension. Also brainstorm the impacts of over-doing each measure and a counter-measure.

Do Lots – How Many (quantity)

Potential measures:

Watch for Impacts and Counter-measures if over-done:

Do it predictably – How consistently (predictability)

Do it Fast - How Fast (speed)

Do It Well – How good (quality)

Do Valuable Stuff – What did customers see (value)

Do Sustainable – Can we keep doing it (sustainability)

Flight Level 2 – End to End Co-ordination Metrics

Brainstorm measures for each performance dimension. Also brainstorm the impacts of over-doing each measure and a counter-measure.

Do Lots – How Many (quantity)

Potential measures:

Watch for Impacts and Counter-measures if over-done:

Do it predictably – How consistently (predictability)

Do it Fast - How Fast (speed)

Do It Well – How good (quality)

Do Valuable Stuff – What did customers see (value)

Do Sustainable – Can we keep doing it (sustainability)

Flight Level 3 – Strategy/Portfolio Metrics

Brainstorm measures for each performance dimension. Also brainstorm the impacts of over-doing each measure and a counter-measure.

Do Lots – How Many (quantity)

Potential measures:

Watch for Impacts and Counter-measures if over-done:

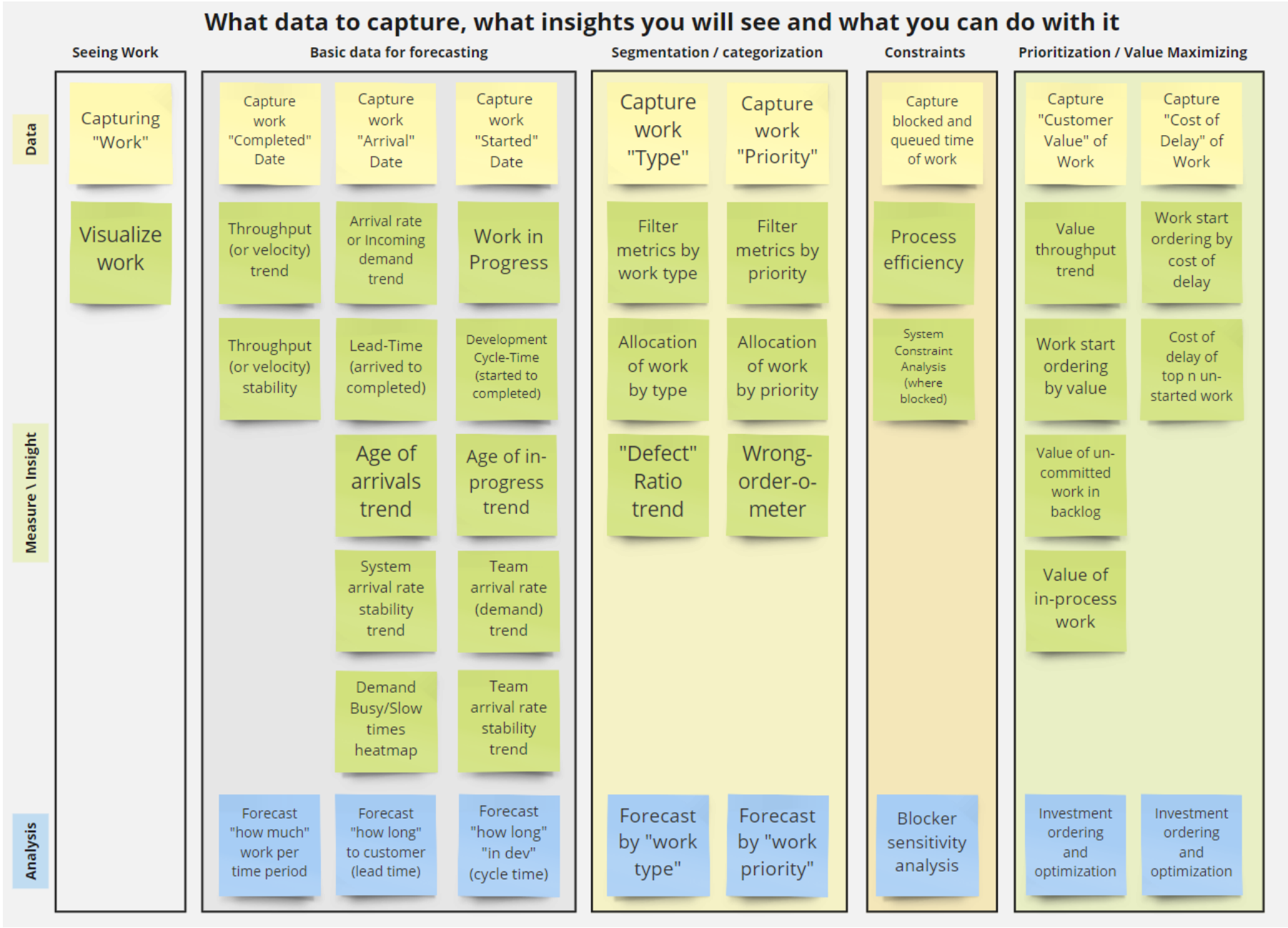
Do it predictably – How consistently (predictability)

Do it Fast - How Fast (speed)

Do It Well – How good (quality)

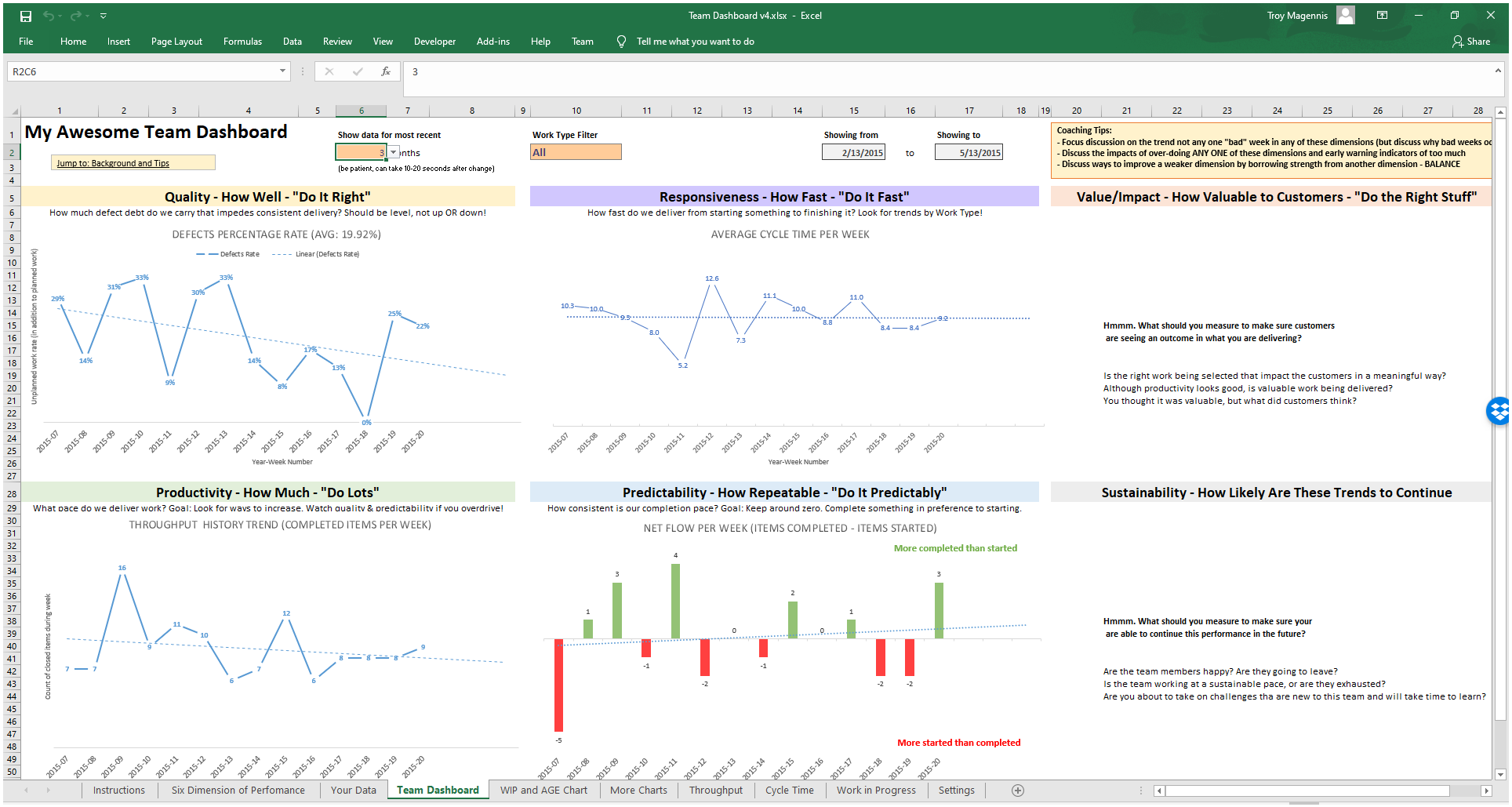
Do Valuable Stuff – What did customers see (value)

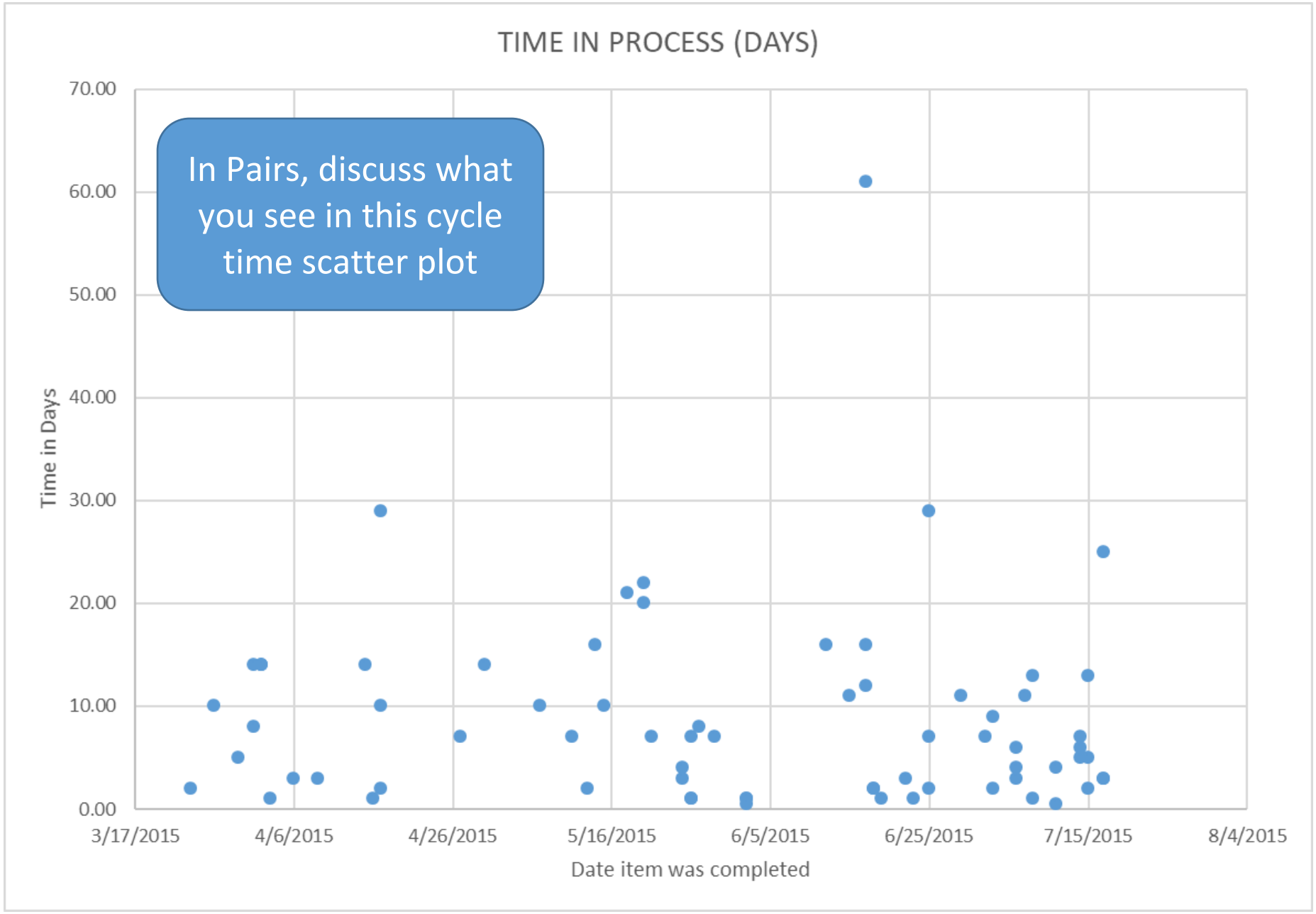
Do Sustainable – Can we keep doing it (sustainability)

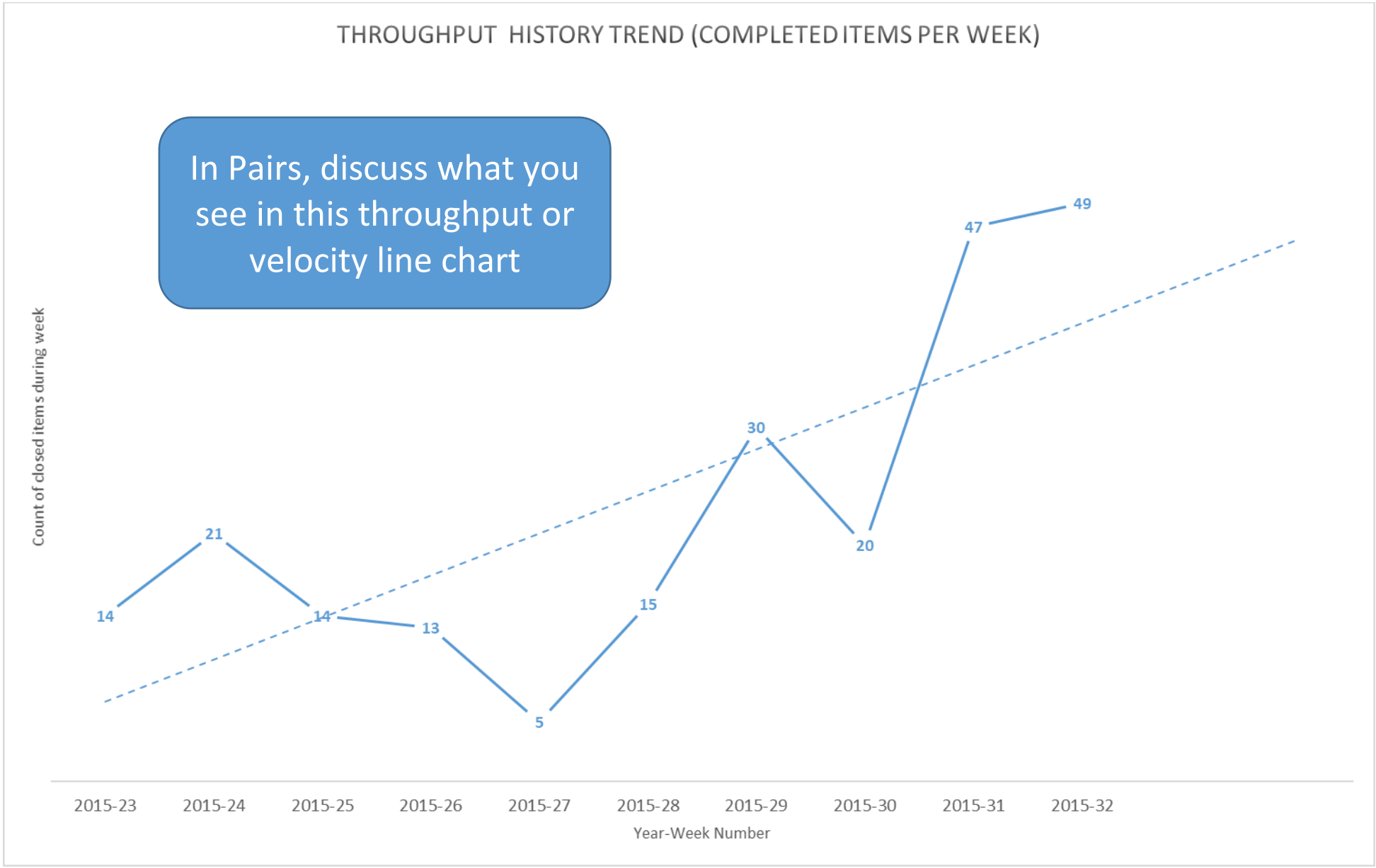

What data to capture, what insights you will see and what you can do with it
Seeing Work
Basic data for forecasting
Segmentation / categorization
Constraints
Prioritization / Value Maximizing
Data
Measure \ Insight
Analysis
Capturing "Work"
Visualize work
Capture work "Completed" Date
Capture work "Arrival" Date
Capture work "Started" Date
Throughput (or velocity) trend
Arrival rate or Incoming demand trend
Work in Progress
Throughput (or velocity) stability
Lead-Time (arrived to completed)
Development Cycle-Time (started to completed)
Age of arrivals trend
Age of in-progress trend
System arrival rate stability trend
Team arrival rate (demand) trend
Demand Busy/Slow times heatmap
Team arrival rate stability trend
Forecast "how much" work per time period
Forecast "how long" to customer (lead time)
Forecast "how long" "in dev" (cycle time)
Capture work "Type"
Capture work "Priority"
Filter metrics by work type
Filter metrics by priority
Allocation of work by type
Allocation of work by priority
"Defect" Ratio trend
Wrong-order-o-meter
Forecast by "work type"
Forecast by "work priority"
Capture blocked and queued time of work
Process efficiency
System Constraint Analysis (where blocked)
Blocker sensitivity analysis
Capture "Customer Value" of Work
Capture "Cost of Delay" of Work
Value throughput trend
Work start ordering by cost of delay
Work start ordering by value
Cost of delay of top n un-started work
Value of un-committed work in backlog
Value of in-process work
Investment ordering and optimization
Investment ordering and optimization

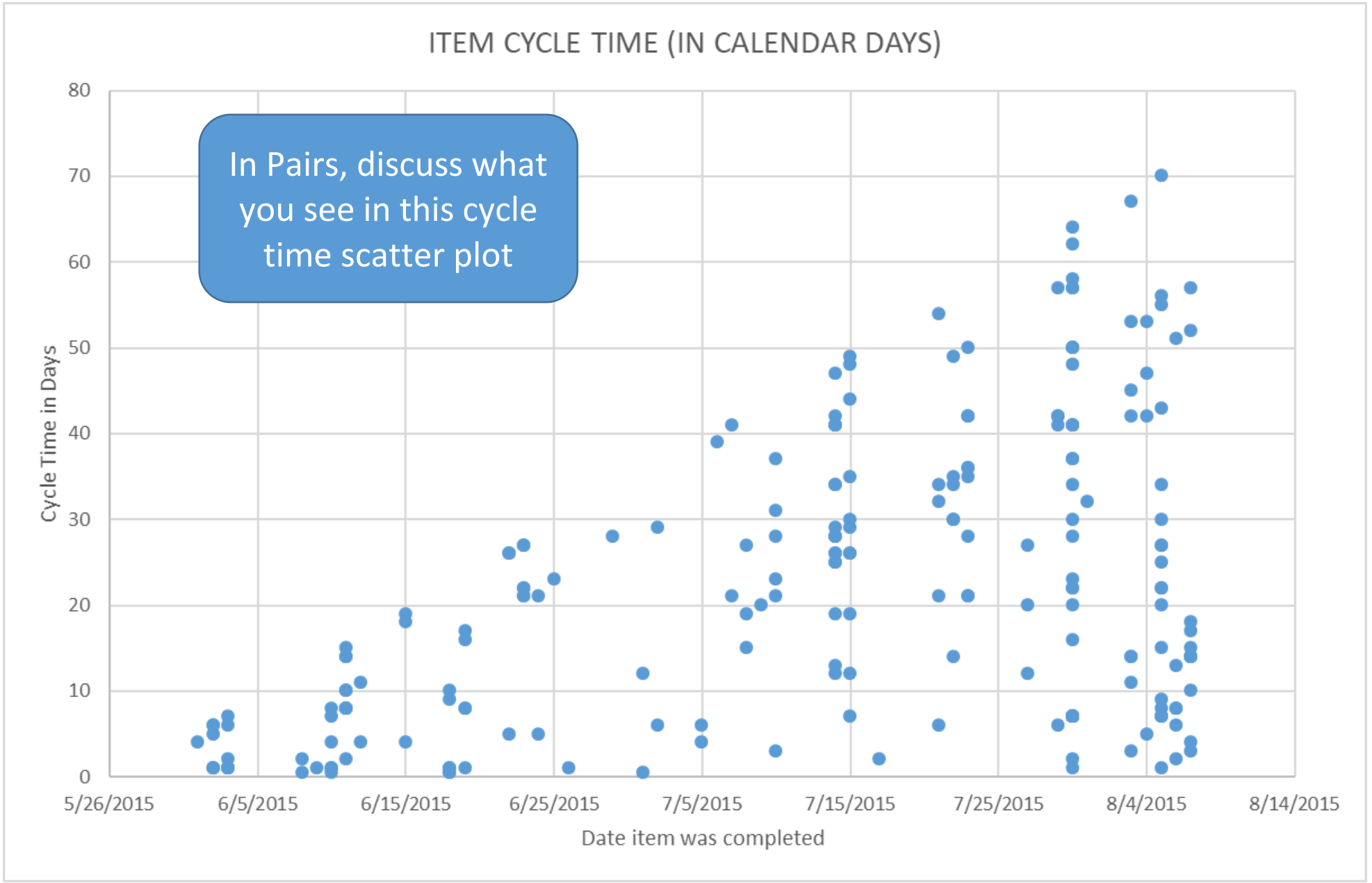
Team Dashboard

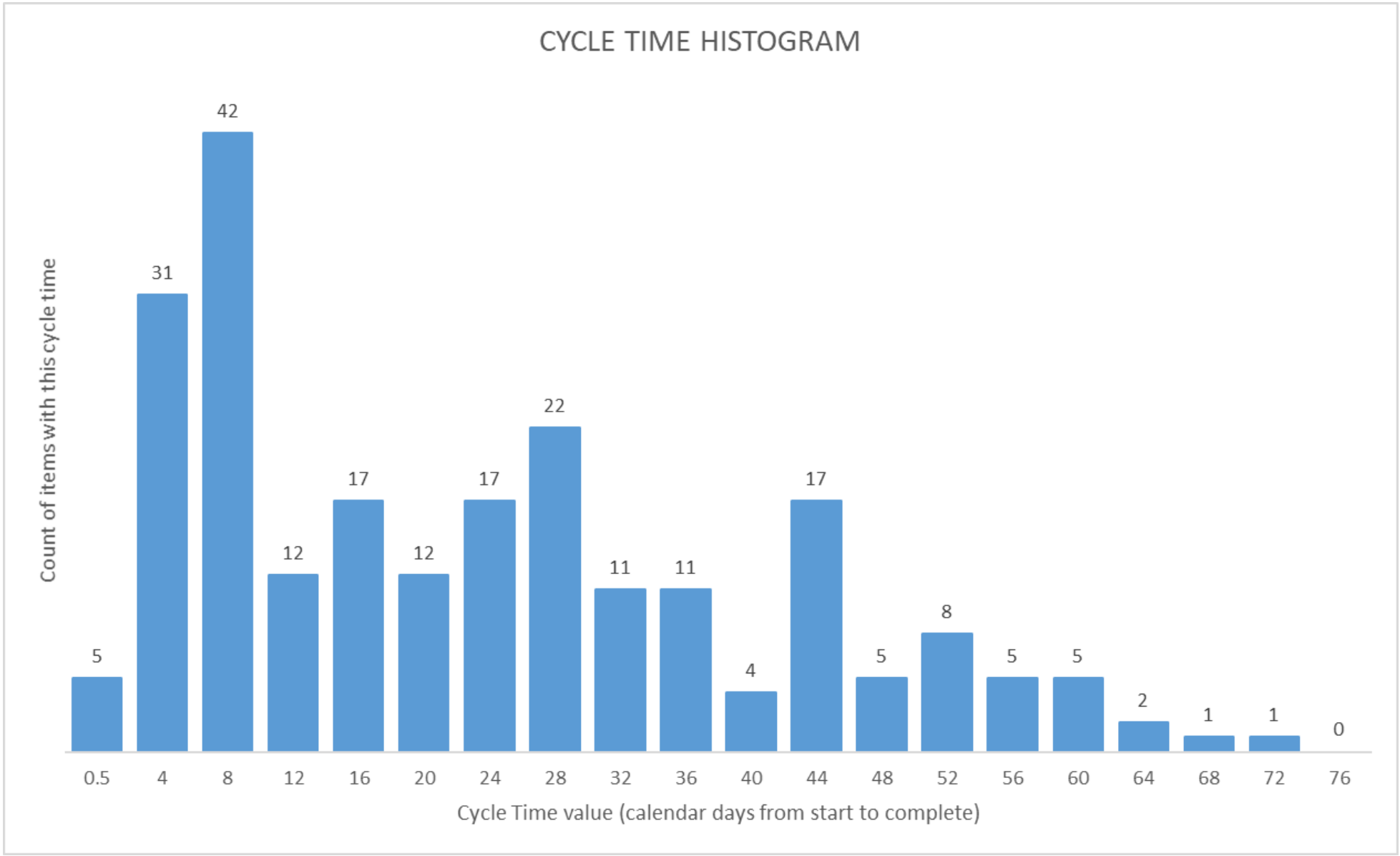
Data: Completed Date, Start Date and Type (story or defect)

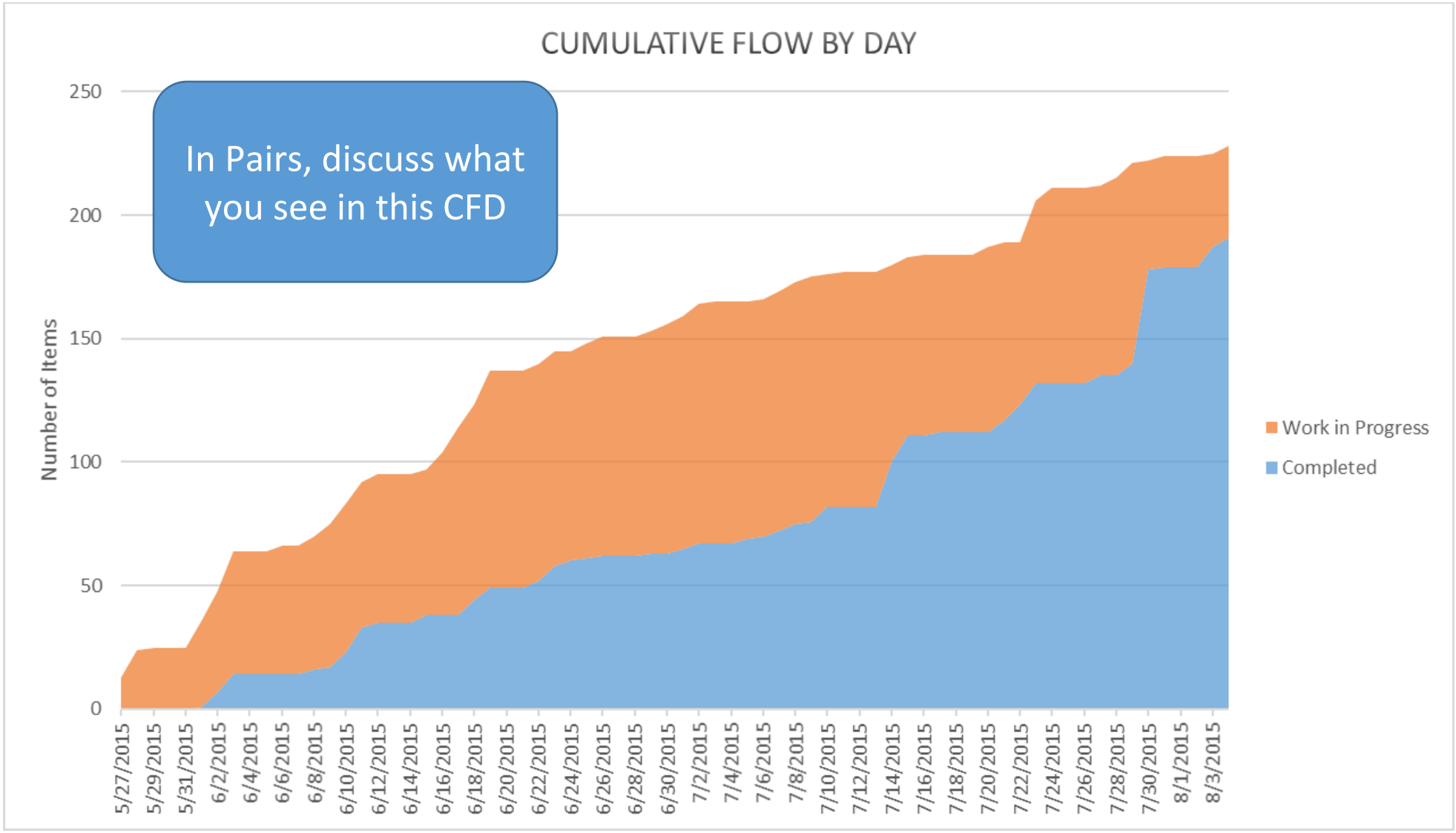



TIME IN PROCESS (DAYS)
In Pairs, discuss what you see in this cycle time scatter plot
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Time in Days
3/17/2015
4/6/2015
4/26/2015
5/16/2015
6/5/2015
6/25/2015
7/15/2015
8/4/2015
Date item was completed


THROUGHPUT HISTORY TREND (COMPLETED ITEMS PER WEEK)
In Pairs, discuss what you see in this throughput or velocity line chart
Count of closed items during week
14
21
14
13
5
15
30
20
47
49
2015-23
2015-24
2015-25
2015-26
2015-27
2015-28
2015-29
2015-30
2015-31
2015-32
Year-Week Number


ITEM CYCLE TIME (IN CALENDAR DAYS)
In Pairs, discuss what you see in this cycle time scatter plot
Cycle Time in Days
0
10
20
30
40
50
60
70
80
5/26/2015
6/5/2015
6/15/2015
6/25/2015
7/5/2015
7/15/2015
7/25/2015
8/4/2015
8/14/2015
Date item was completed

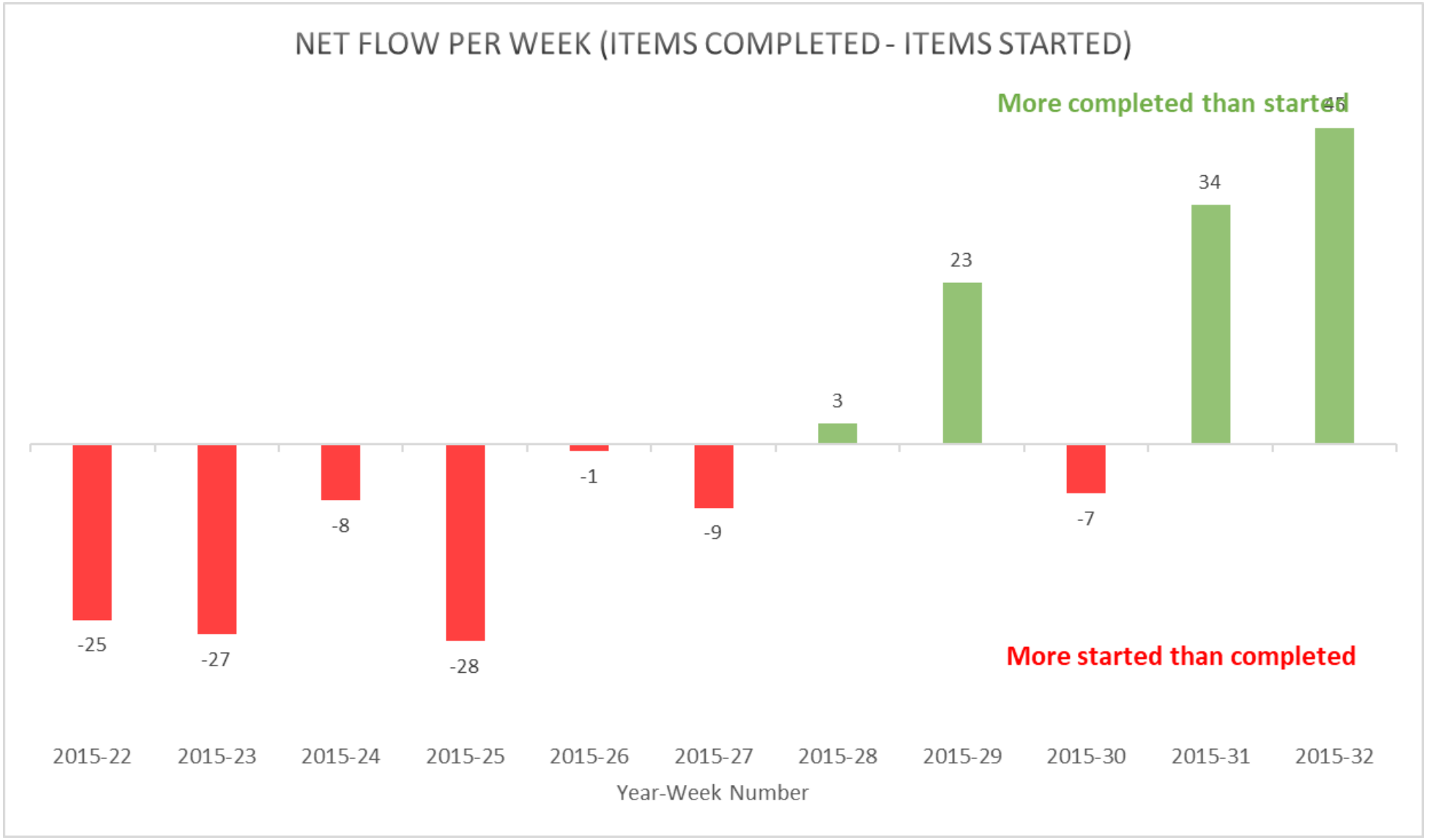

CYCLE TIME HISTOGRAM
Count of items with this cycle time
Cycle Time value (calendar days from start to complete)
0.5
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
5
31
42
12
17
12
17
22
11
11
4
17
5
8
5
5
2
1
1
0

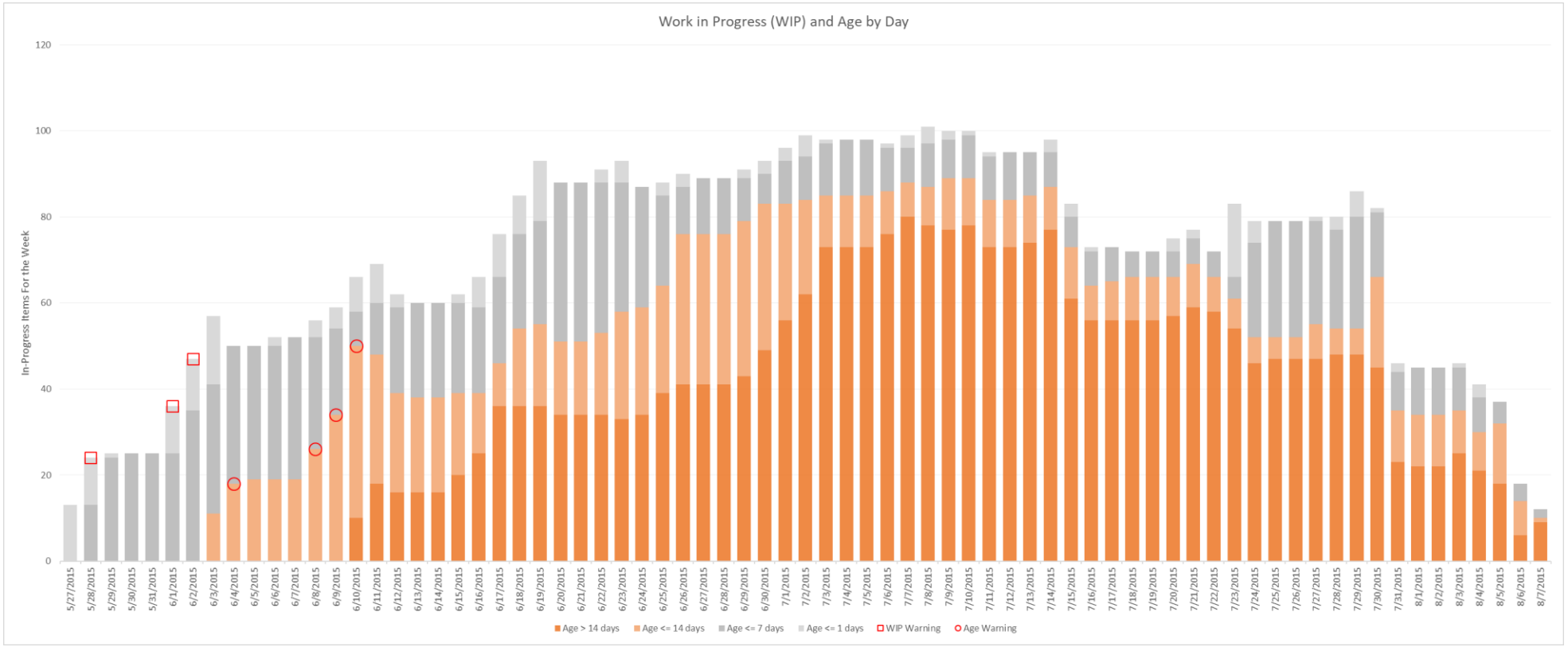

CUMULATIVE FLOW BY DAY
In Pairs, discuss what you see in this CFD
Number of Items
250
200
150
100
50
0
5/27/2015
5/29/2015
5/31/2015
6/2/2015
6/4/2015
6/6/2015
6/8/2015
6/10/2015
6/12/2015
6/14/2015
6/16/2015
6/18/2015
6/20/2015
6/22/2015
6/24/2015
6/26/2015
6/28/2015
6/30/2015
7/2/2015
7/4/2015
7/6/2015
7/8/2015
7/10/2015
7/12/2015
7/14/2015
7/16/2015
7/18/2015
7/20/2015
7/22/2015
7/24/2015
7/26/2015
7/28/2015
7/30/2015
8/1/2015
8/3/2015
Work in Progress
Completed

Exercise: What Questions or Conditions Do You Want to "See"

As a group, look at the measures you chose for Flight Level 1 – Operational and Team Metrics. Describe the goals, questions you want answered and the "trigger" conditions for action.

Metric	Goals	Questions	Triggers


NET FLOW PER WEEK (ITEMS COMPLETED - ITEMS STARTED)
More completed than started
More started than completed
-25
-27
-28
-8
-1
-9
3
23
-7
34
45
2015-22
2015-23
2015-24
2015-25
2015-26
2015-27
2015-28
2015-29
2015-30
2015-31
2015-32
Year-Week Number



In pairs, discuss this WIP and Age by Day chart. Tell the story of this team over time. When would worry have started? What do you think the future holds? and what is your next coaching action?

Action Visualizations Need to Help See and Answer Things

WHAT

1. Is it clear what the chart is showing (title and legend)?
2. Is it clear how to read the magnitude of values (axis clearly labeled)?

SO WHAT

1. Is it clear to see what “abnormal” is for the targeted insights?
2. Is it clear where differences are insignificant or spurious?

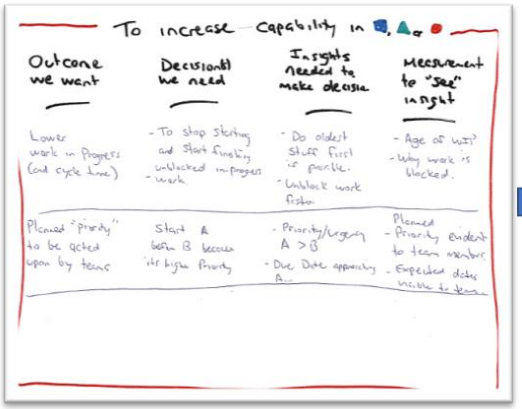
NOW WHAT

1. Does the chart lead to the right decisions and actions or just blame?
2. Does the chart put the right emphasis on urgency for action?

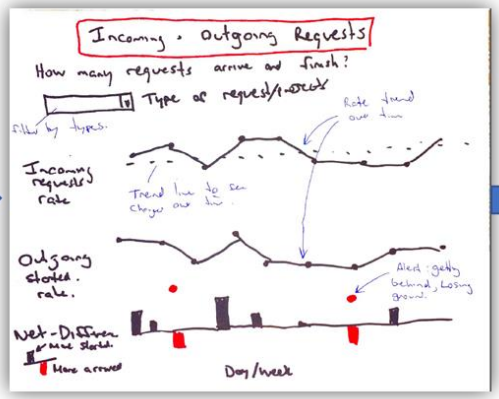
Outcomes – Decisions – Insights – Measures

- O.D.I.M. is a facilitation tool used to design actionable dashboards
  - Focus discussions around outcomes avoiding vanity dashboards
  - Focus dashboard design on highlighting the most valuable data
    - Data that informs a decision or action is the most valuable, everything else is noise

1. Define Outcomes (ODIM)



2. Paper Prototype + Feedback



3. Build + Team Feedback

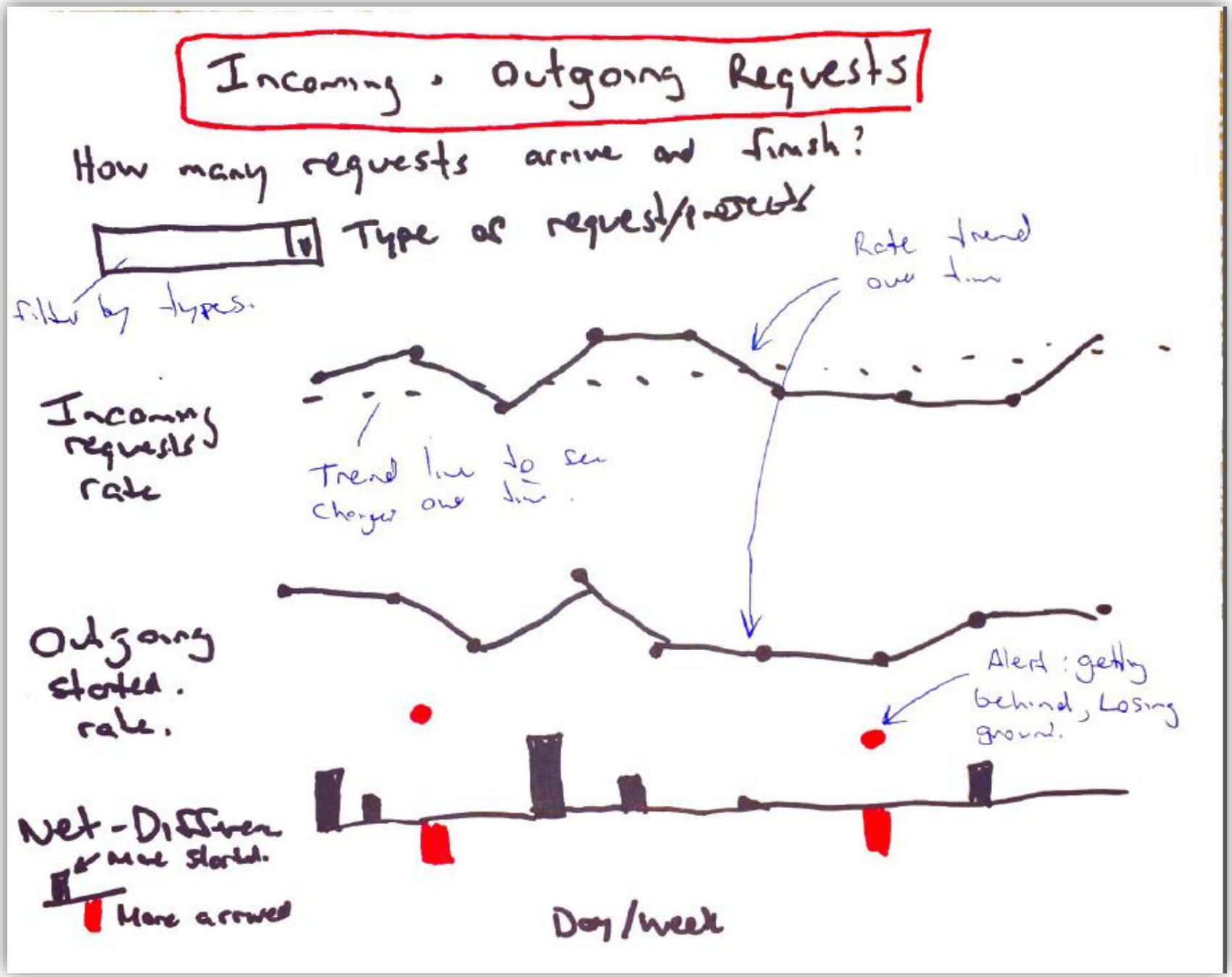
Organization Flow Trend Dashboard.

Outcome	Decisions	Insights	Measures.
Improve organizational flow.	"too long" (lead time) reduce it!	Lead time not reducing (in waiting? in doing?	- Lead time trend - Cycle time trend
		Too many "old" items	- Item age trend.
	"Not enough" Increase teams to add capacity	Too much WIP	- WIP trend.
		Not keeping up with demand. (more in than out)	- Negative Net flow
"Too much instability"		Erratic Incoming demand Erratic Outgoing throughput	- variability of demand - " " throughput.
Which teams need help?		- Are teams active? - Are teams increasing in demand? - Teams adverse trend to majority	

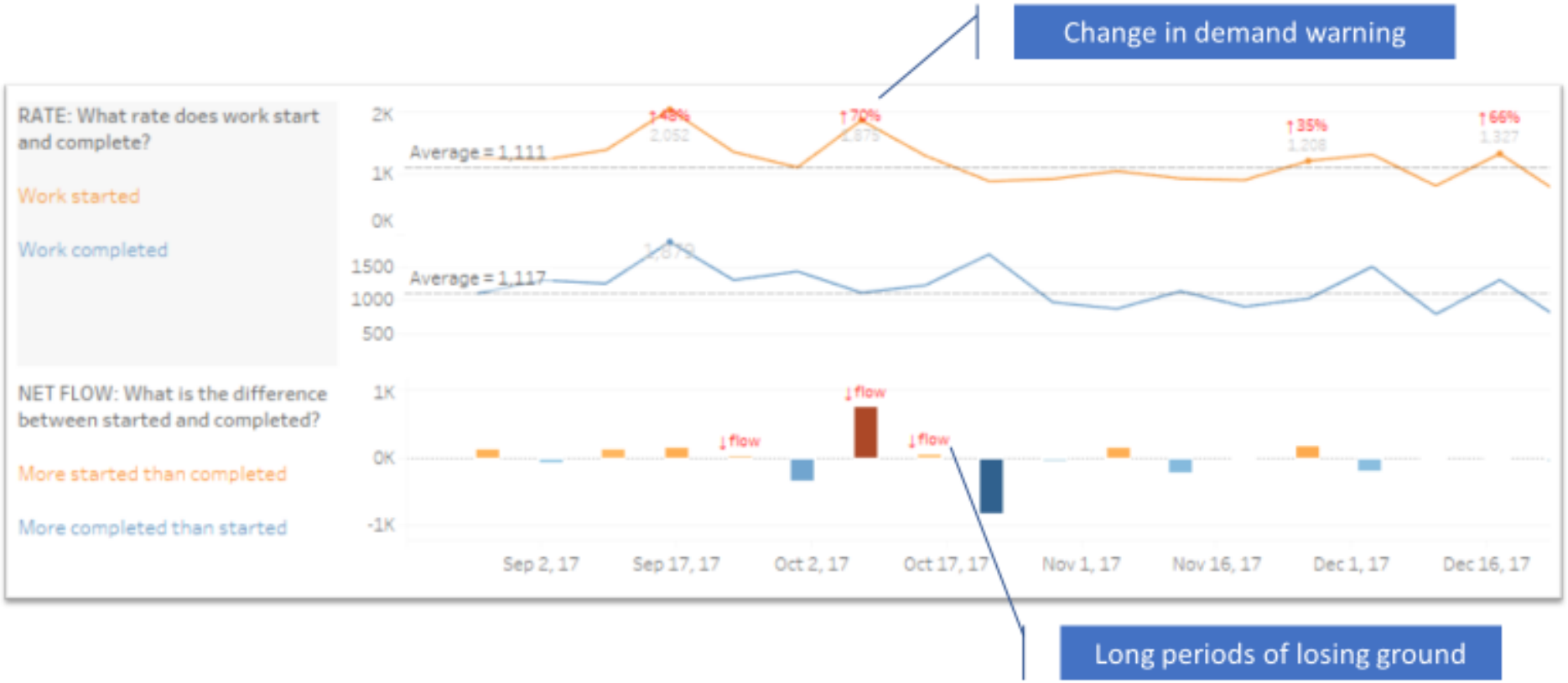
Outcome we want	Decision(s) we need	Insights needed to make decision	Measurement to "see" insight
Lower work in Progress (and cycle time)	- To stop starting and start finishing unblocked in-progress work.	- Do oldest stuff first if possible. - Unblock work fastw	- Age of WIP - Why work is blocked.
Planned "priority" to be acted upon by teams	Start A before B because its higher Priority	- Priority/Urgency A > B - Due Date approaching A...	Planned - Priority evident to team members - Expected dates visible to team.

Outcome: To visualize and improve flow

- Outcome
  - Team capacity matches needed demand
  - Anticipate changes and impact of demand bursts
- Decisions
  - We need more staff
  - We need more staff “for a while”
- Insights
  - Team is able to deliver at the same rate as incoming demand
    - Show warning if incoming demand exceeds delivery rate by x% over y weeks
  - Temporary or consistent increases in demand
    - Show warning if the incoming demand grows > x% from prior week
- Measures
  - Incoming rate
  - Outgoing rate
  - The imbalance of incoming versus outgoing rate


Incoming • Outgoing Requests
How many requests arrive and finish?
Type of request/protocol
filter by types.
Rate trend over time
Incoming requests rate
Trend line to see changes over time.
Outgoing started. rate.
Alert: getting behind, Losing ground.
Net-Difference
More started.
More arrived
Day/week

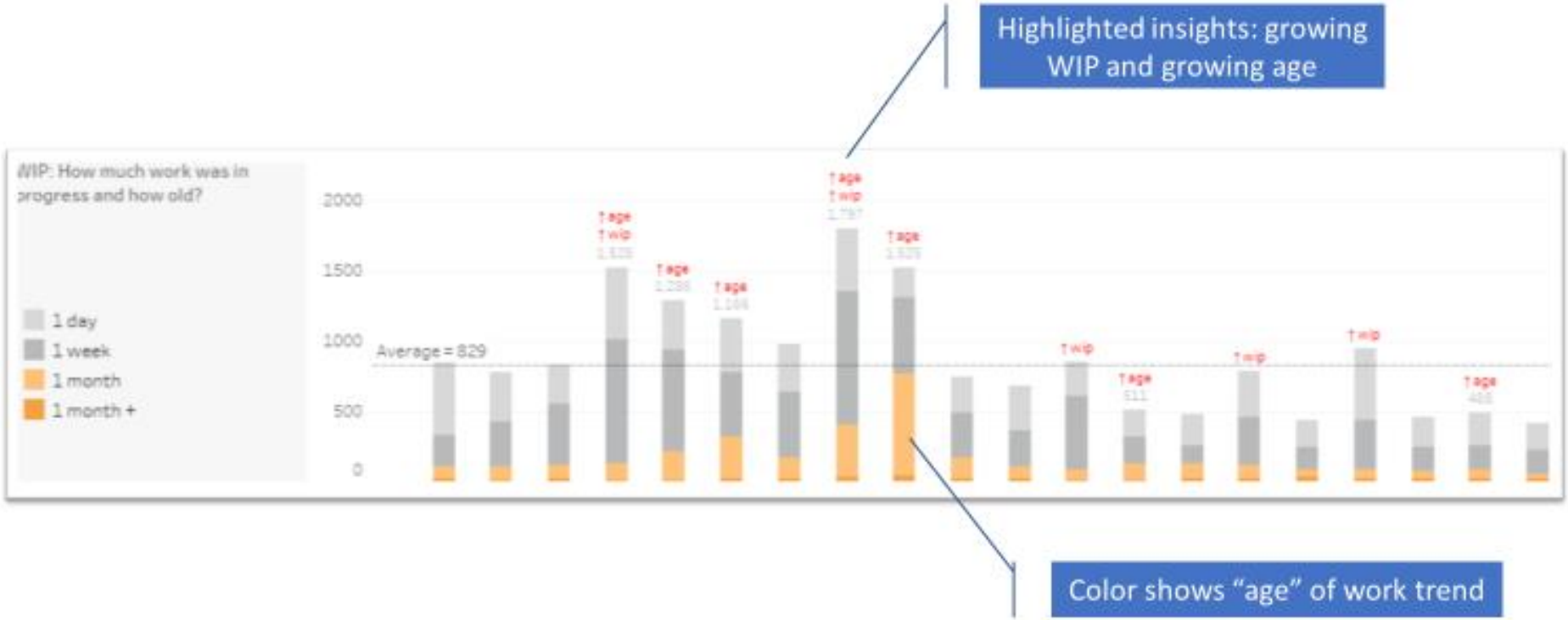
Incoming and Outgoing Demand Measures



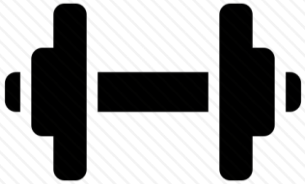
Outcome: To visualize and improve flow

- Outcome (we want to drive)
  - Reduce Work in Progress
  - Reduce cycle time of work
- Decisions (people need to make)
  - Lets finish the in-progress stuff before stating new stuff
  - Lets finish something that is older before something newer
- Insights (that help people make those decisions)
  - We have a lot of WIP compared to normal
    - Show warning if WIP grows > x% from prior week
  - We have an increasing "age" of in-progress work
    - Show warning if the average age of work grows > x% from prior week
- Measures (that show these insights)
  - Work in progress trend
  - Age of the work in progress trended and color coded from older to newer

WIP and Age Measures



Workshop Exercise



Part 1: O.D.I.M.

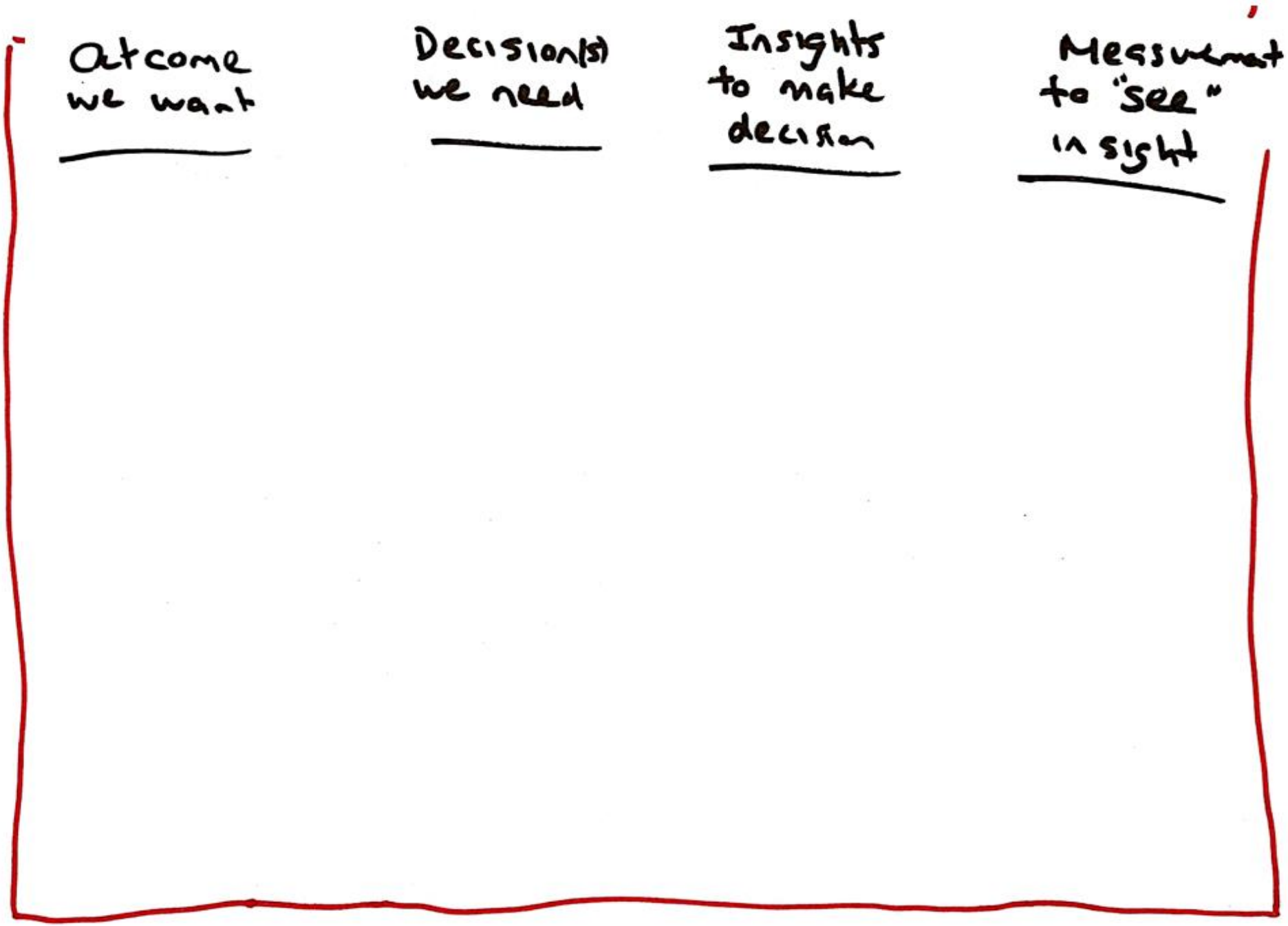
1. Brainstorm a set of MAJOR outcomes. A major outcome is one that can have 3 to 5 SUPPORTING outcomes.
2. Work in groups to define as many SUPPORTING outcomes (write on the ODIM worksheet column under Outcomes, spread equally down the page) and present to the group. Dot vote for the top 3.
3. Work in groups and populate the Decisions, Insights and Actions for your MAJOR outcome.
4. Gallery Walk and capture discussion and alternative ideas.

Note: Sometimes you abort an outcome early because you “don’t have the data” – avoid this. If the group decides this is a valuable measure, we will deal with this later!

Part 2: Paper Prototype (wait until told, we will do this after some theory)

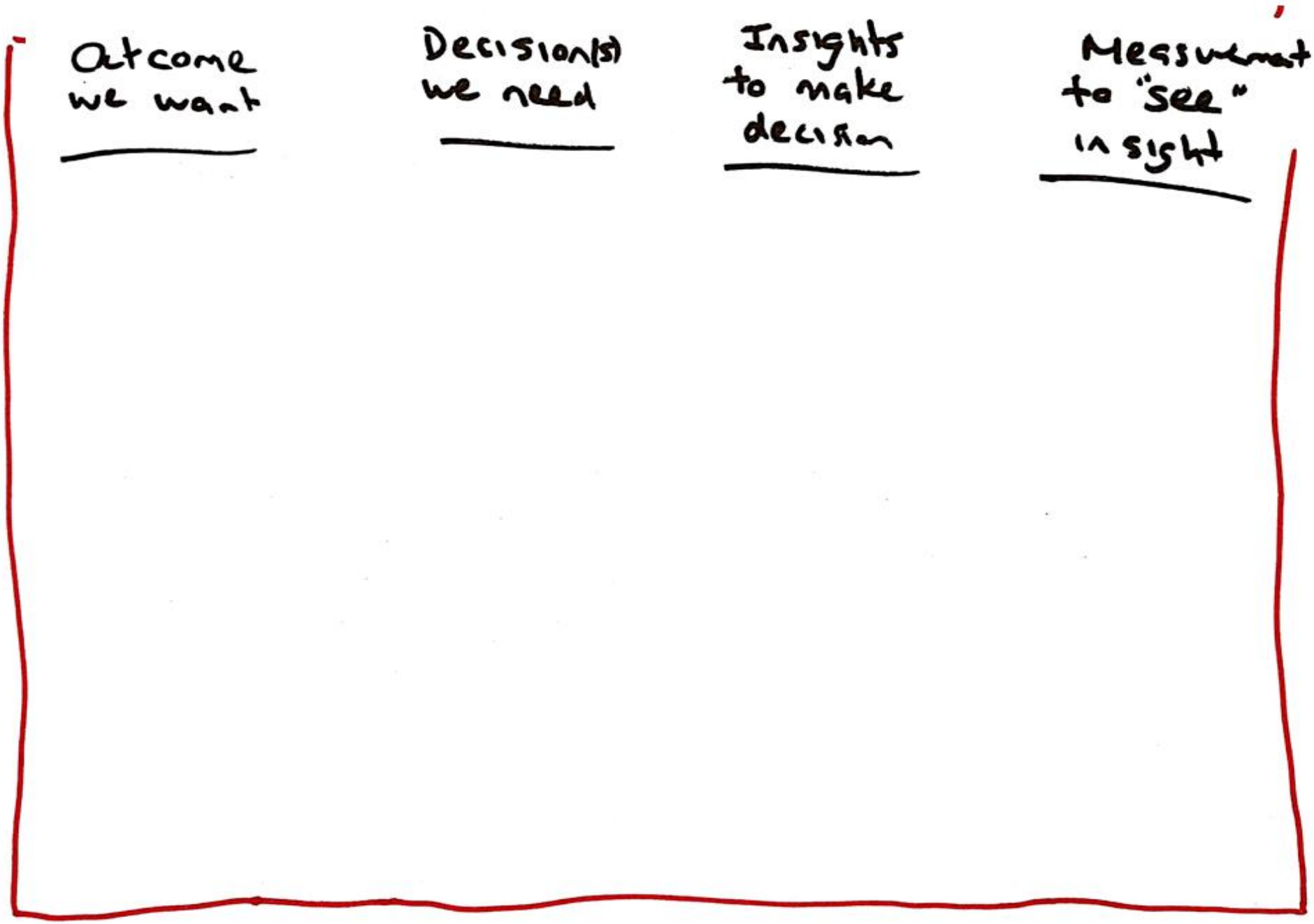
1. Work in groups to sketch ideas that show the measures in a way that highlight the most powerful insight. Do as many as you can in 15 minutes.
2. Gallery walk and capture ideas for improvement.
3. Select the “right” ones and do a second round of final sketch.
4. Gallery walk and present the “why” this sketch is “right”.

Note: Use color sparingly. Color should be used to highlight an INSIGHT not as a legend where possible.


Outcome we want
Decision(s) we need
Insights to make decision
Measurement to "see" insight

Sketches for Insight:

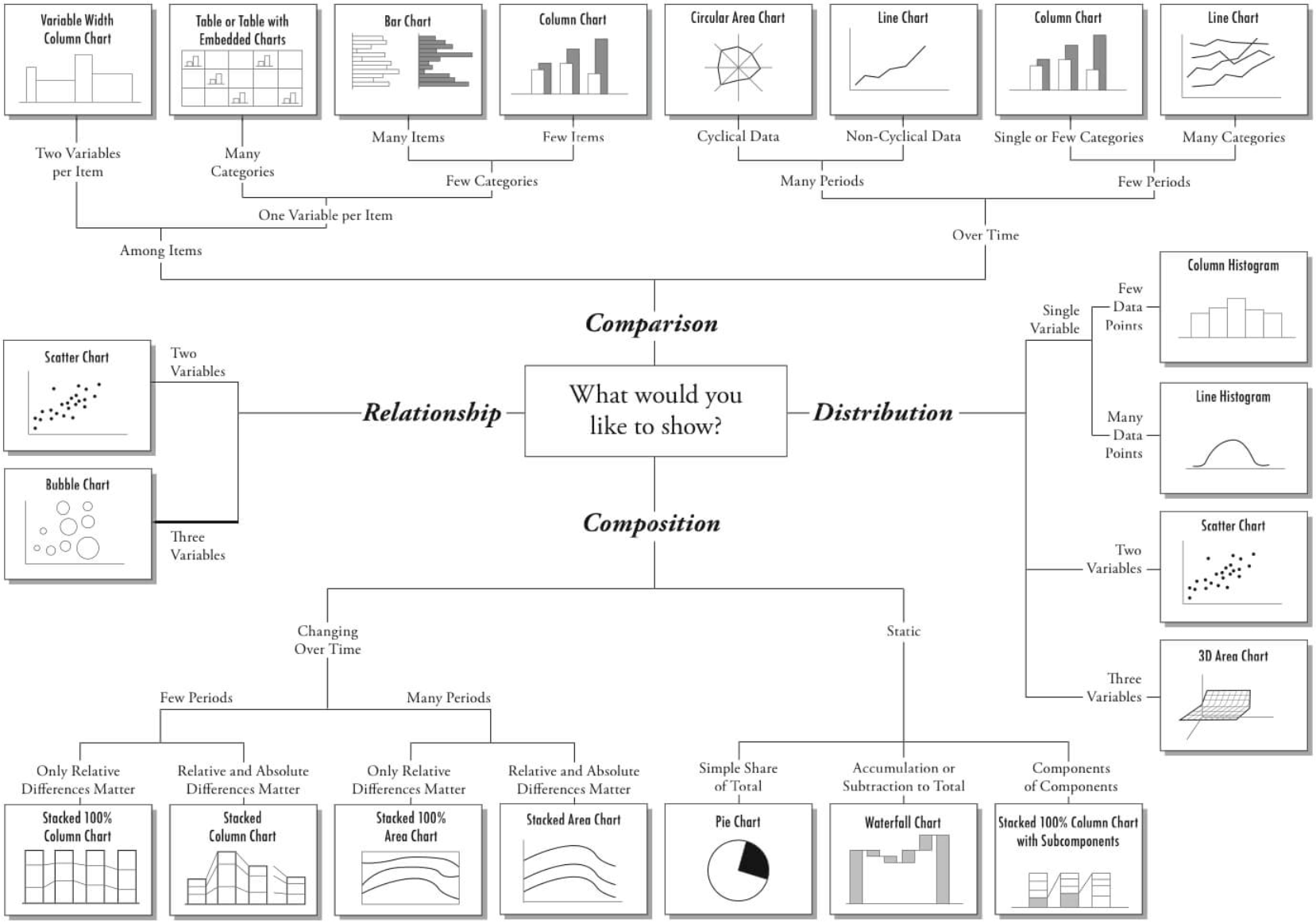
Tips: Trends are always better than a single number. Highlight the insights. How will you help people see these insights first?


Outcome we want
Decision(s) we need
Insights to make decision
Measurement to "see" insight

Sketches for Insight:

Tips: Trends are always better than a single number. Highlight the insights. How will you help people see these insights first?

Chart Suggestions—A Thought-Starter



Visual Vocabulary

There are so many ways to visualise data - how do we know which one to pick? Click on a category below to decide which data relationship is most important in your story, then look at the different types of charts within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e., one causes the other).

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Part-to-Whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

INSPIRED BY

FT Graphics: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg; Paul McCallum; Martin Stabe
Visual Vocabulary Poster: ft.com/vocabulary
Graphic Continuum: Jon Schwabish and Severino Ribecca

CREDITS & TUTORIALS

Diverging Stacked Bar	Steve Wexler	Data Revelations	Arc Chart	Ken Flerlage	KenFlaerlage.com	Chord Diagram	Noah Salvaterra	DataBlick
Sunburst Chart	Leonid Golub	Super Data Science	Radar Chart	Adam McCann	Dueling Data	Sankey Diagram	Leonid Golub	Super Data Science
Surplus/Deficit Filled Line	Jeffrey Shaffer	Data +Science	Scaled Cartogram	Ken Flerlage	KenFlaerlage.com	.		
Violin Plot	Ben Moss	YouTube	Venn Diagram	Leonid Golub	Super Data Science	..		

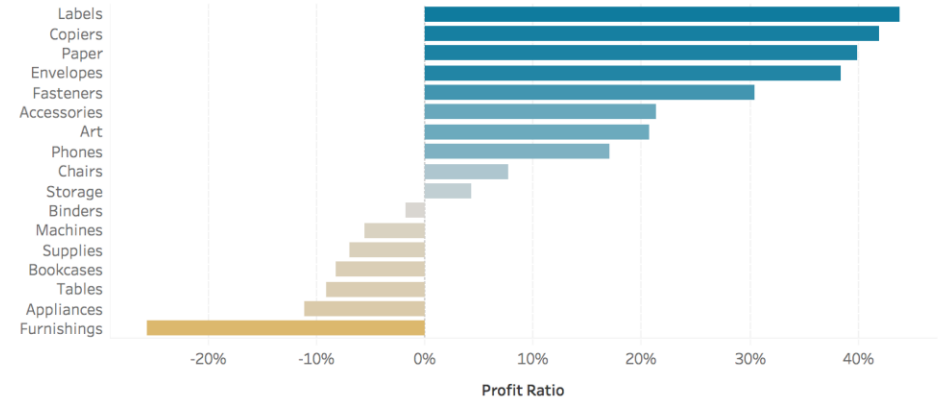
All these images from https://www.vizwiz.com/2018/07/visual-vocabulary.html
https://photos.google.com/share/AF1QipOsge9krcbmO5qj6mzjFbG4WfUL6jYVayXTqBO_BclqUEB94mZrWw-E9-o-Q_pHMQ?key=Tk04RW9oVDZxcC05UEIzRTRYcVJqeXNjb1o4NFpn

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

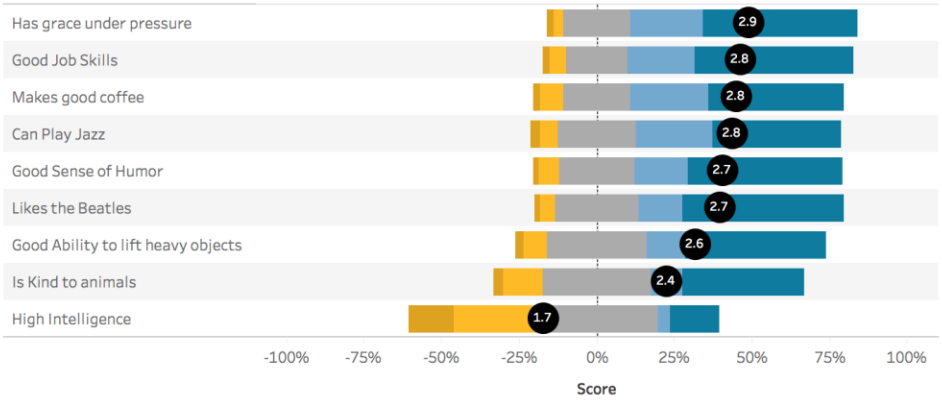
Diverging Bar

A simple standard bar chart that can handle both negative and positive magnitude values



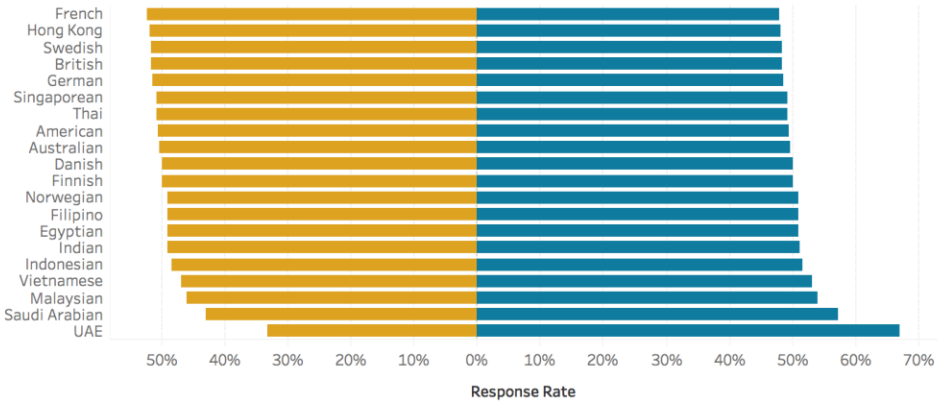
Diverging Stacked Bar

Perfect for presenting survey results which involve sentiment (e.g., disagree/neutral/agree)



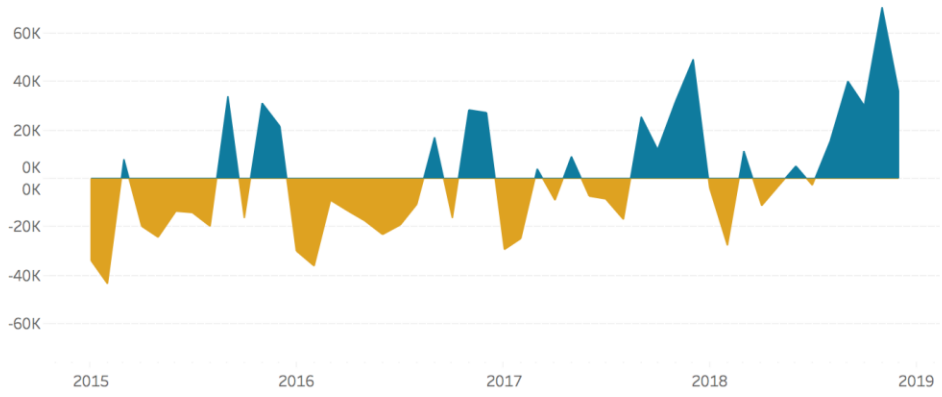
Spine Chart

Splits a single value into 2 contrasting components (e.g., Male/Female)



Surplus/Deficit Filled Line

The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

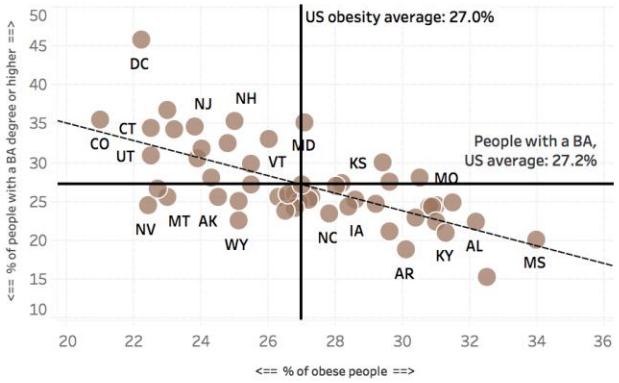


Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e., one causes the other).

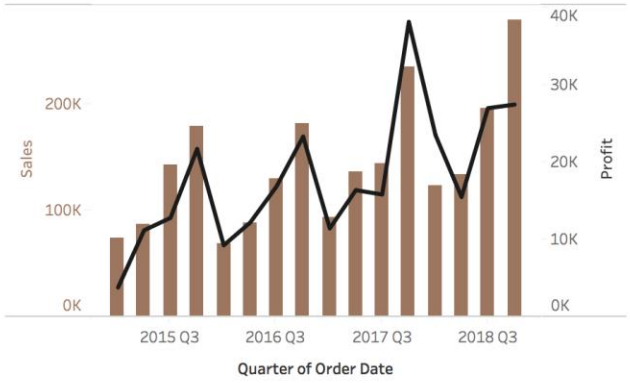
Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.



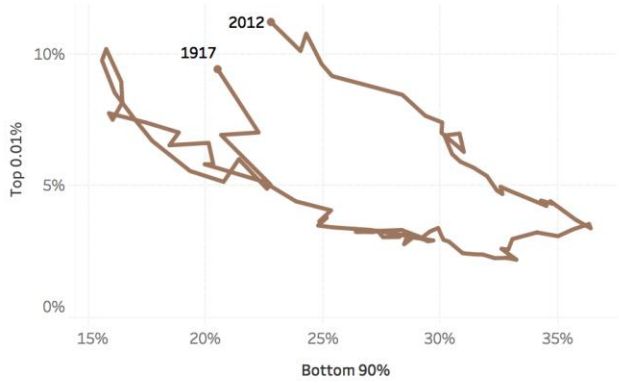
Line + Column

A good way of showing the relationship between an amount (columns) and a rate (line).



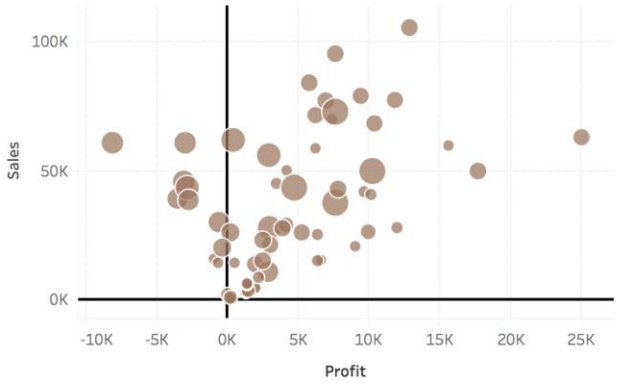
Connected Scatterplot

Usually used to show how the relationship between 2 variables has changed over time.



Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.



XY Heatmap

A good way of showing the patterns between 2 categories of data, less good at showing fine differences in amounts.

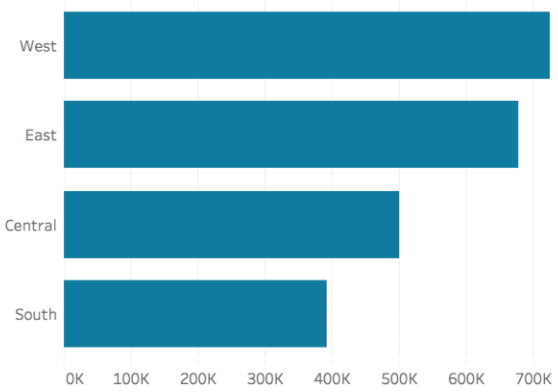
Age Range	I don't have a savings account	$0	Just the minimum balance requirement	Less than $1,000	$1,000-$4,999	$5,000-$9,999	$10,000 or more
Overall	21.0%	28.0%	9.0%	13.0%	10.0%	5.0%	14.0%
18-24	22.4%	21.8%	9.7%	19.1%	14.7%	4.7%	7.5%
25-34	18.0%	26.3%	10.6%	15.2%	12.5%	5.4%	12.1%
35-44	18.9%	31.6%	6.6%	11.6%	9.8%	5.6%	16.0%
45-54	21.6%	30.8%	7.7%	10.9%	7.5%	5.2%	16.2%
55-64	22.8%	28.4%	8.4%	10.7%	8.0%	4.8%	16.8%
65+	21.6%	27.6%	10.7%	8.2%	7.2%	4.7%	20.0%

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

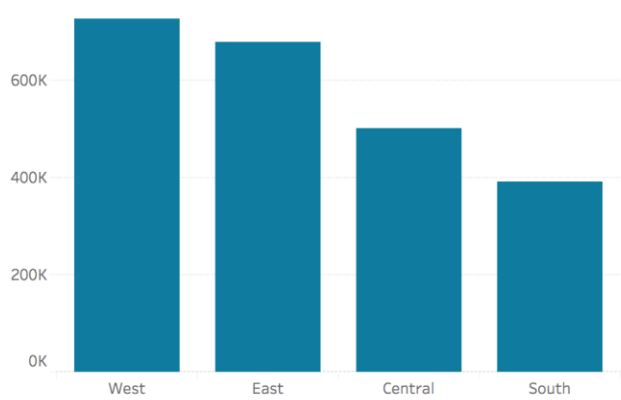
Ordered Bar

Standard bar charts display the ranks of values much more easily when sorted into order.



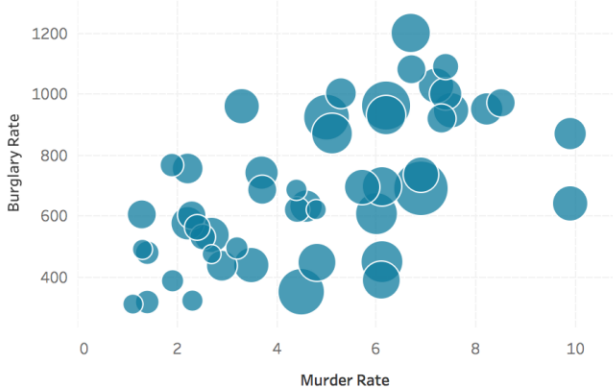
Ordered Column

Standard bar charts display the ranks of values much more easily when sorted into order.



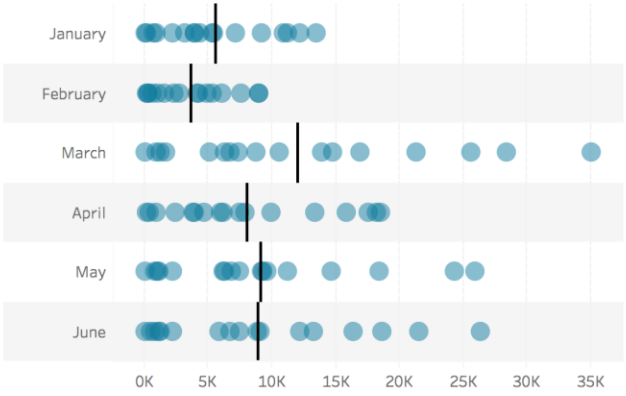
Ordered Proportional Symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.



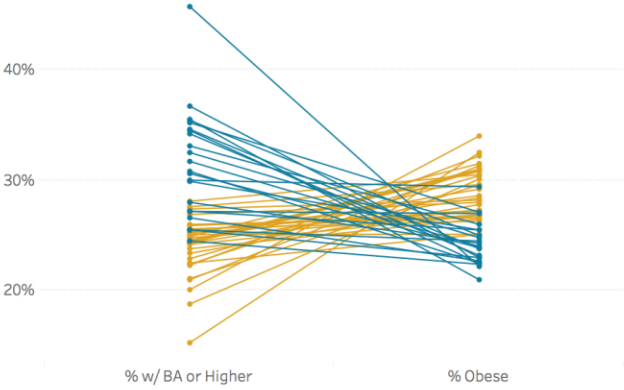
Dot Strip Plot

Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.



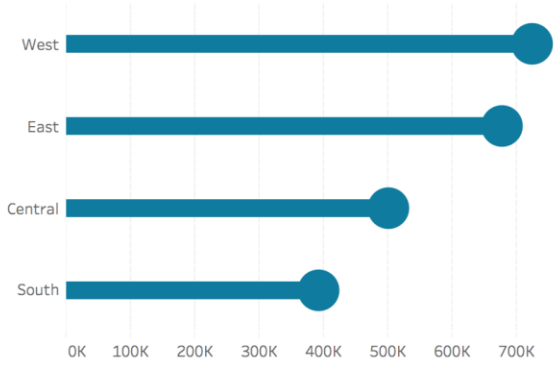
Slope

Perfect for showing how ranks have changed over time or vary between categories.



Lollipop Chart

Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

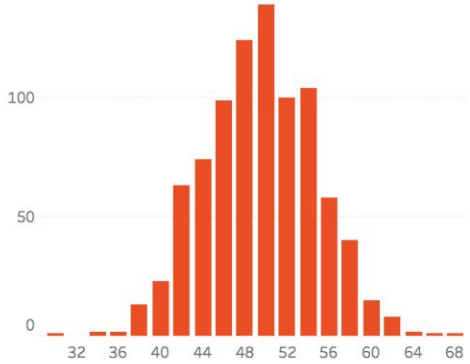


Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Histogram

The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

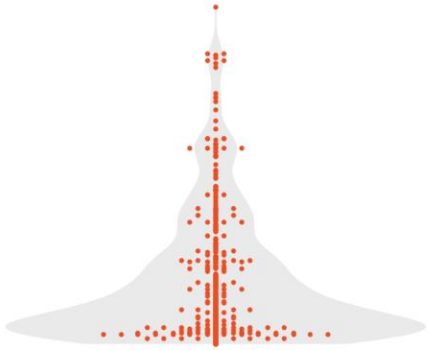


Boxplot

Summarise multiple distributions by showing the median (centre) and range of the data

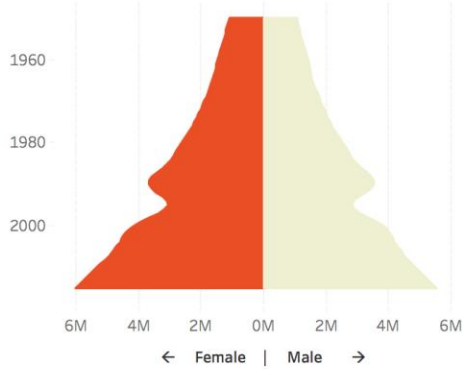
Violin Plot

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).



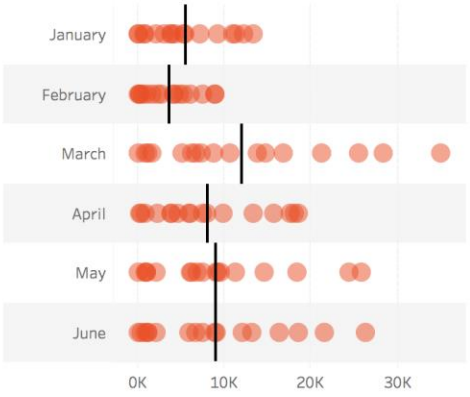
Population Pyramid

A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.



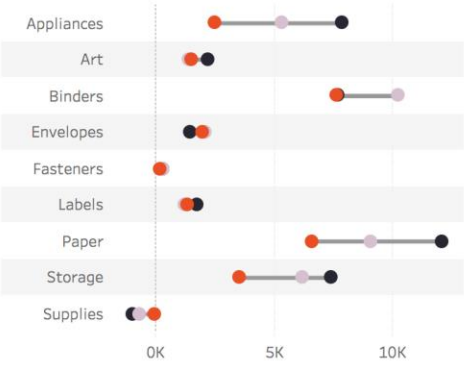
Dot Strip Plot

Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.



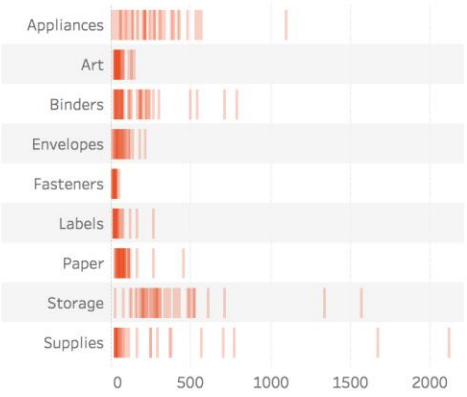
Dot Plot

A simple way of showing the change or range (min/max) of data across multiple categories.



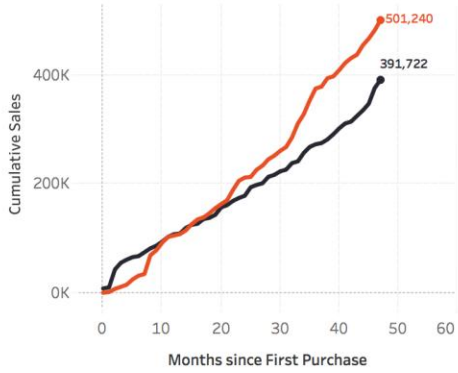
Barcode Plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.



Cumulative Curve

A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure.



Change over Time

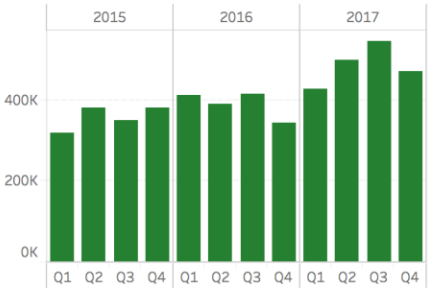
Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

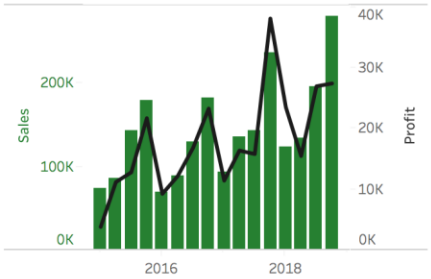
Column

Columns work well for showing change over time - but usually best with only one series of data at a time.



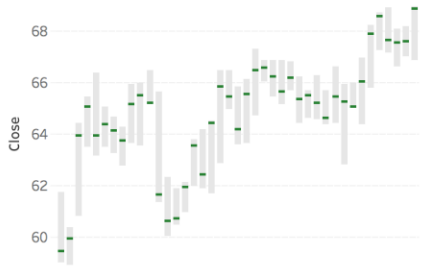
Line + Column

Columns work well for showing change over time - but usually best with only one series of data at a time.



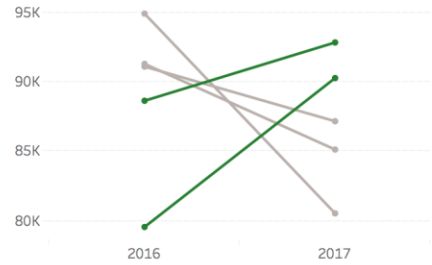
Stock Price

Usually focused on day-to-day activity, these charts show opening/closing and hi/low points of each day.



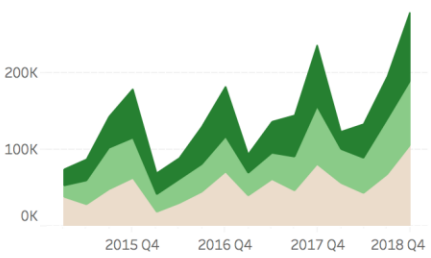
Slope

Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.



Area Chart

Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.



Fan Chart

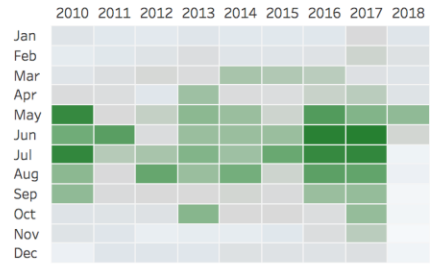
Use to show the uncertainty in future projections - usually this grows the further forward to projection.

Connected Scatterplot

A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

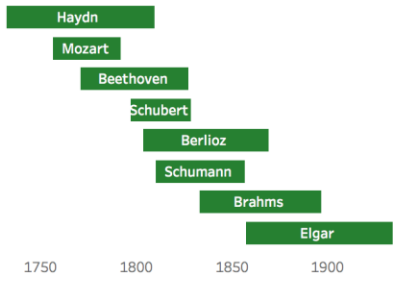
Calendar Heatmap

A great way of showing temporal patterns (daily, weekly, monthly) – at the expense of showing precision in quantity.



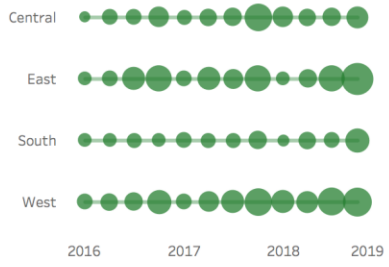
Priestley Timeline

Great when date and duration are key elements of the story in the data.



Circle Timeline

Good for showing discrete values of varying size across multiple categories (e.g., sales by quarter).



Seismogram

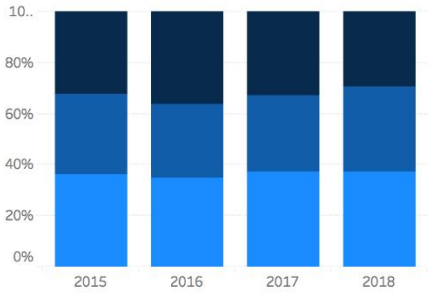
Another alternative to the circle timeline for showing series where there are big variations in the data.

Part-to-Whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

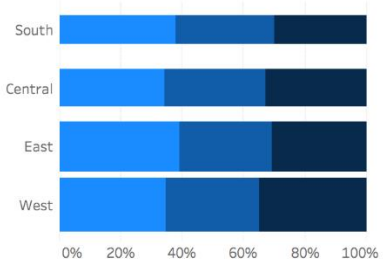
Stacked Column

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.



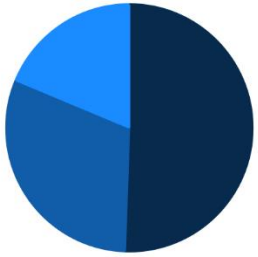
Proportional Stacked Bar

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.



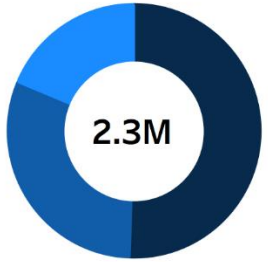
Pie Chart

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.



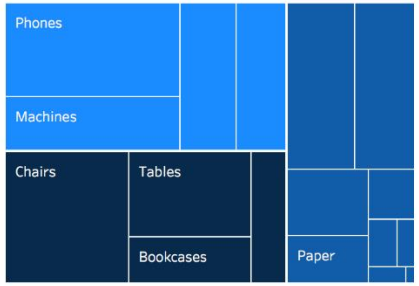
Donut Chart

Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (e.g., total).



Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.



Sunburst

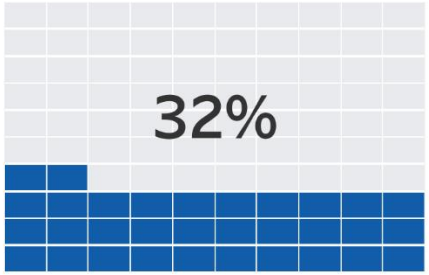
Another way of visualisaing hierarchical part-to-whole relationships. Use sparingly (if at all) for obvious reasons.

Arc

Another way of visualisaing hierarchical part-to-whole relationships. Use sparingly (if at all) for obvious reasons.

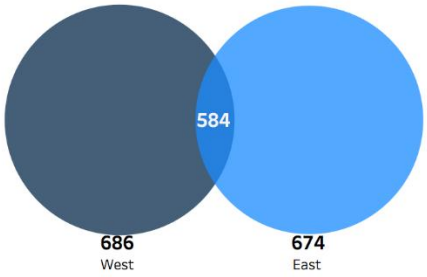
Gridplot

Good for showing % information, they work best when used on whole numbers and work well in multiple layout form.



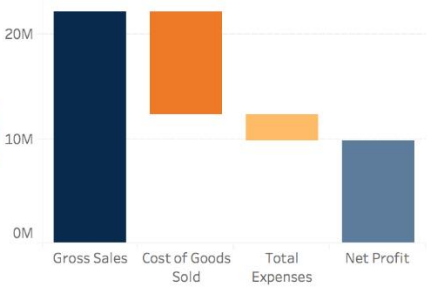
Venn

Generally only used for schematic representation.



Waterfall

Can be useful for showing part-to-whole relationships where some of the components are negative.

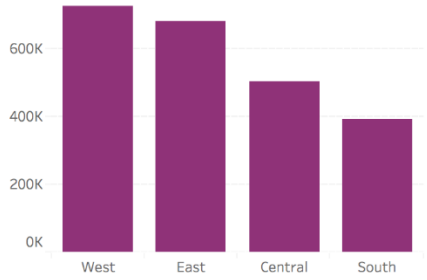


Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

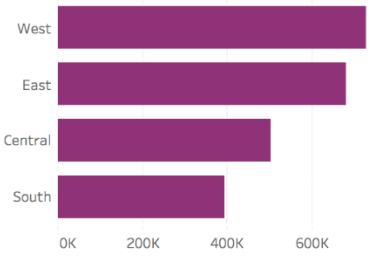
Column

The standard way to compare the size of things. Must always start at 0 on the axis.



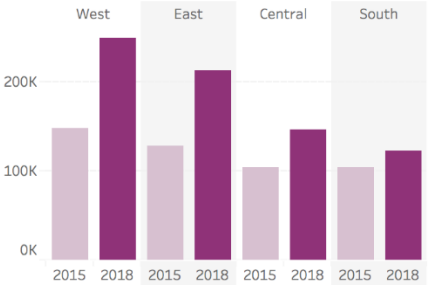
Bar

The standard way to compare the size of things. Must always start at 0 on the axis. Good when the data are not time series and labels have long category names.



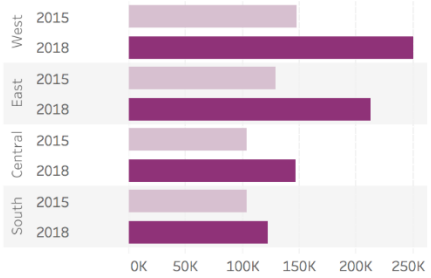
Paired Column

As per standard column, but allows for multiple series. Can become tricky to read with more than 2 series.



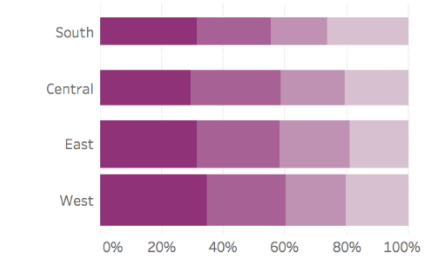
Paired Bar

As per standard bar, but allows for multiple series. Can become tricky to read with more than 2 series.



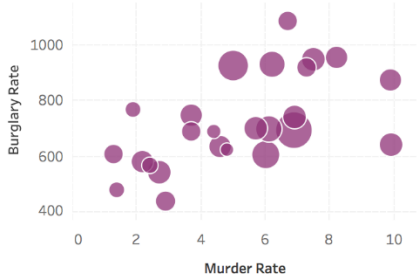
Proportional Stacked Bar

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.



Proportional Symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

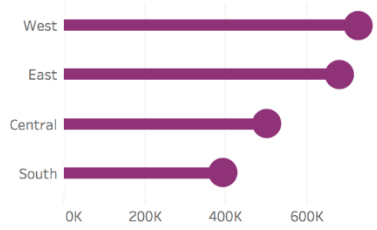


Isotype (pictogram)

Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

Lollipop Chart

Lollipops draw more attention to the data value than standard bar/column - does not HAVE to start at zero (but preferable).

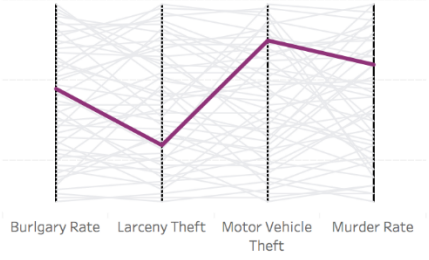


Radar Chart

A space-efficient way of showing value of multiple variables– but make sure they are organised in a way that makes sense to reader.

Parallel Coordinates

An alternative to radar charts – again, the arrngement of the variables is important. Usually benefits from highlighting values.



Spatial

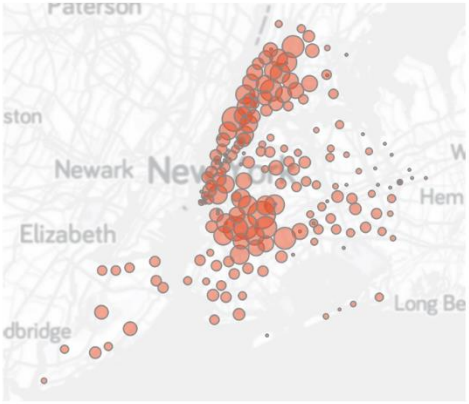
Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

Basic Choropleth (rate/ratio)

The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Proportional Symbol (count/magnitude)

Use for totals rather than rates – be wary that small differences in data will be hard to see.

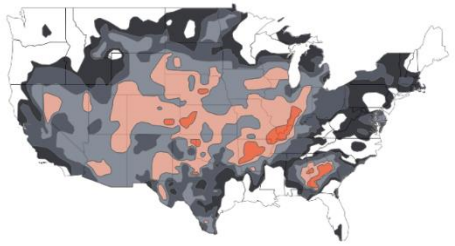


Flow Map

For showing unambiguous movement across a map.

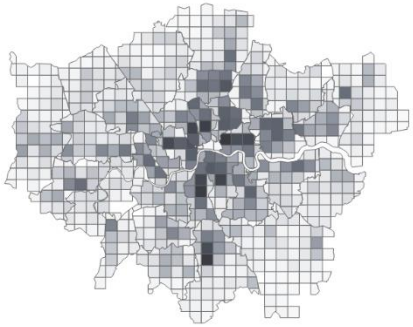
Contour Map

For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values



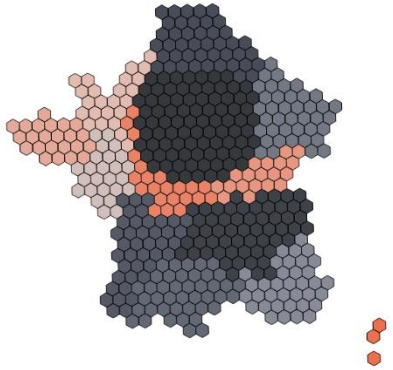
Equalized Cartogram

Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.



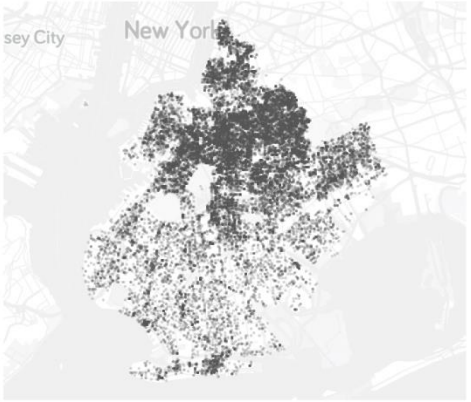
Scaled Cartogram

Stretching and shrinking a map so that each area is sized according to a particular value.



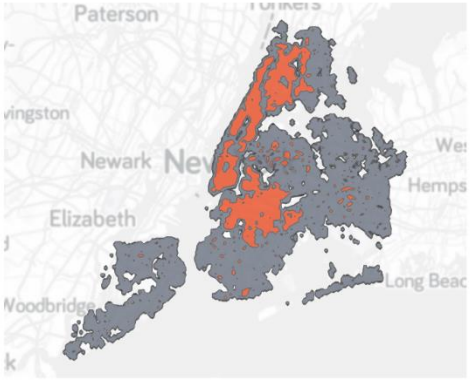
Dot Density

Used to show the location of individual events/locations – make sure to annotate any patterns the reader should see.



Heat Map

Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit.

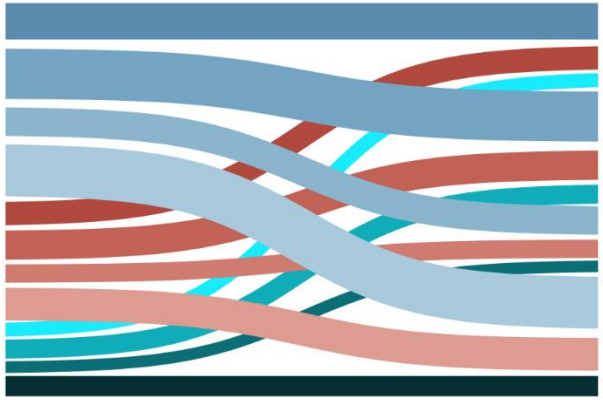


Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

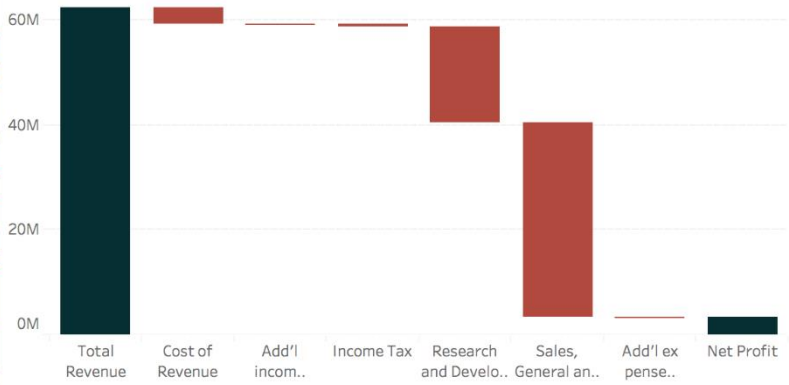
Sankey

Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.



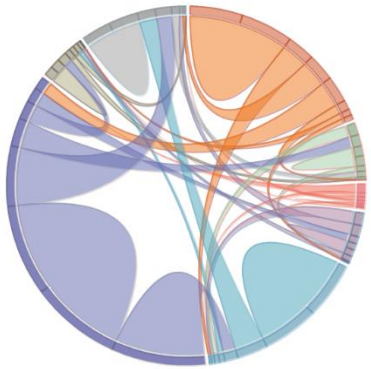
Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.



Chord

A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.



Network

Used for showing the strength and inter-connectedness of relationships of varying types.

Focus on “Important” Insights

Grouping

Sorting

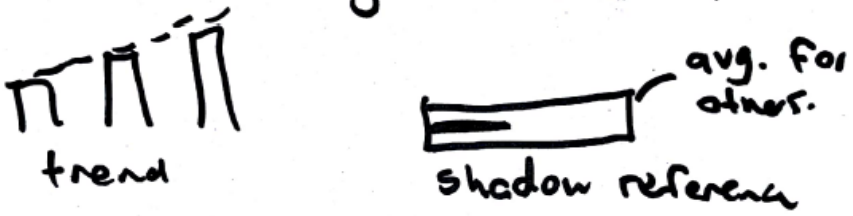
Visual Comparators (reference lines / bands)

Focus Devices (color, dot, size)

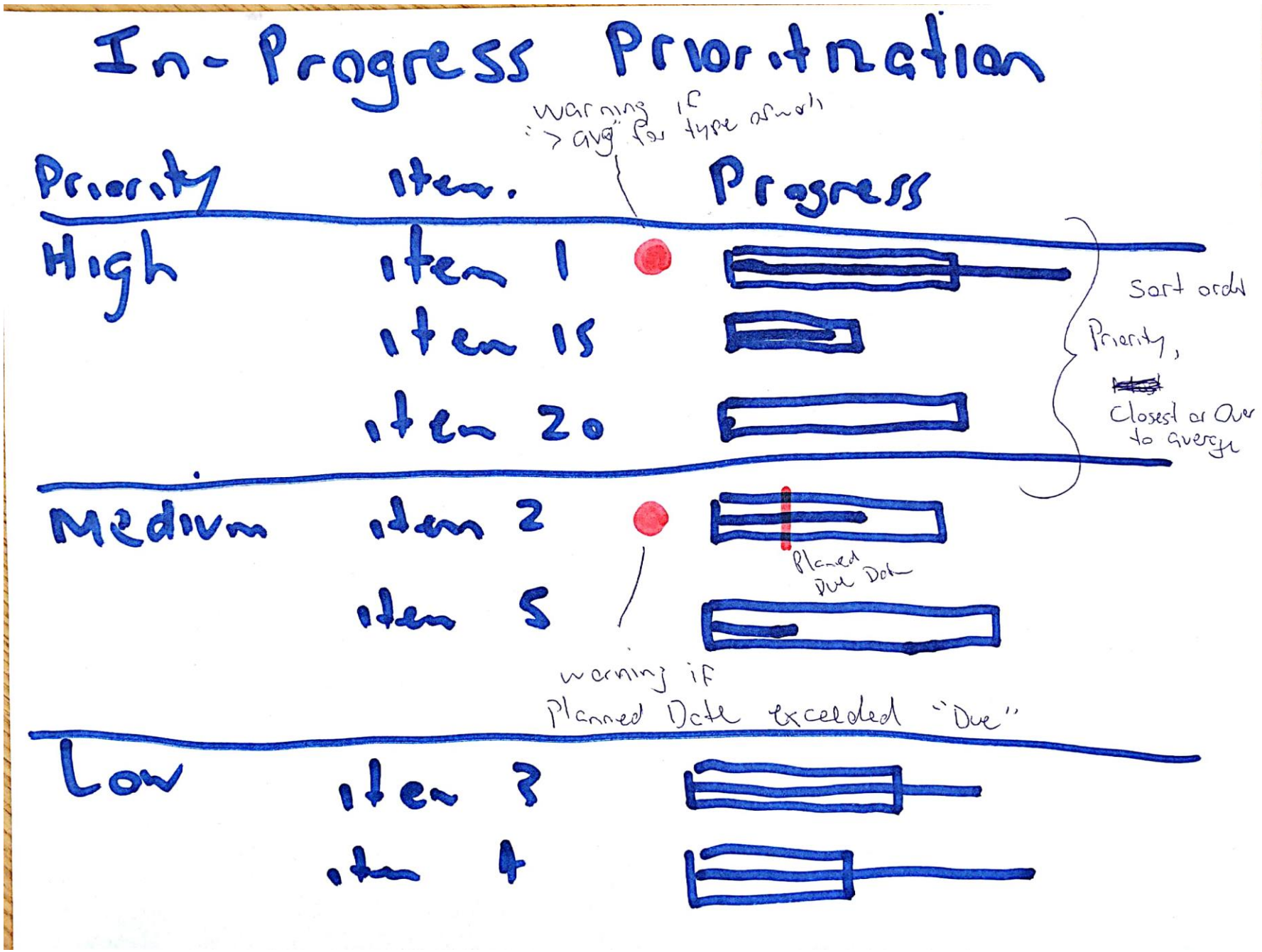
Tips for action viz'

1. Must show values compared to some reference.

   eg. - Trend (prior periods)
   - Background shadow

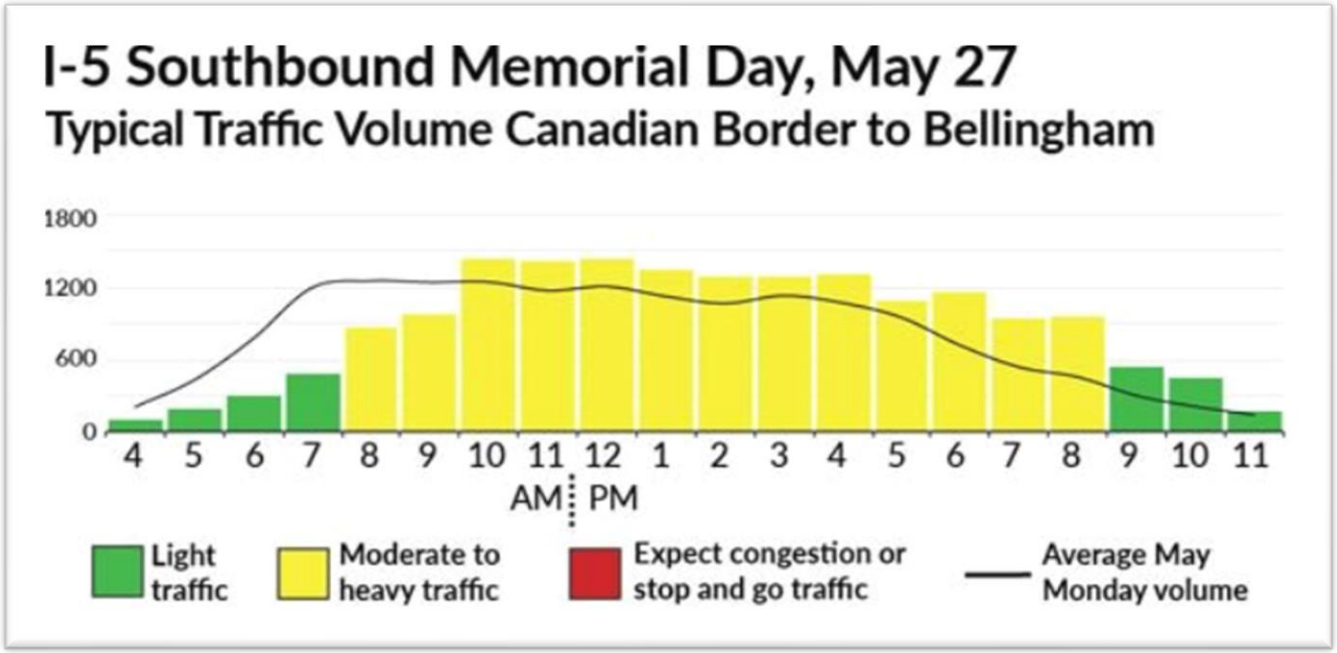


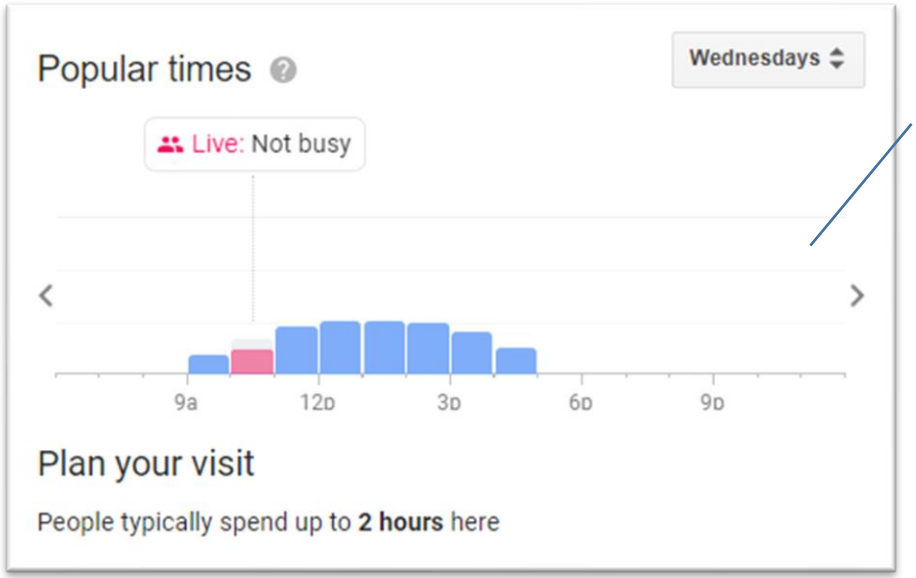
2. Must have a way to highlight "interesting"


In-Progress Prioritization
warning if > avg for type of work
Priority
Item.
Progress
High
item 1
item 15
item 20
Sort order
Priority,
Closest or Over to Overall
Medium
item 2
item 5
Planned Due Date
warning if Planned Date exceeded "Due"
Low
item 3
item 4

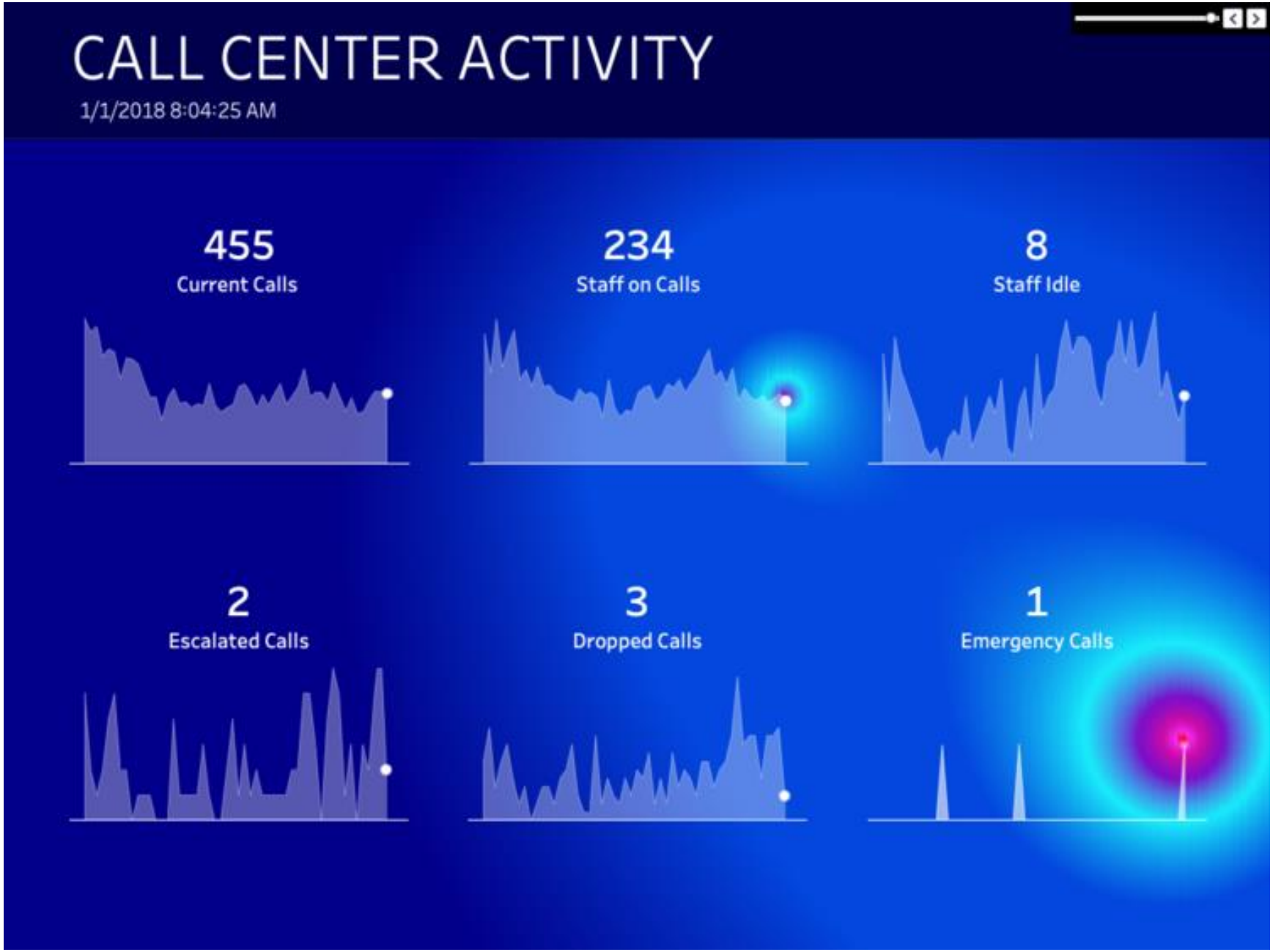
T.-Do Prioritization

Priority	Item	Type.	Forecast. Jan Feb Mar
High	Item 1	x	Missed Finish. Missed Start. Planned Finish. 50%. 95%.
	Item 2	y	
	Item 3	z	
Medium	Item 11	x	
	Item 12	y	
Low	Item 4		
	:		





NOTE: They scale the axis for ALL days, not just the one you see to avoid "looking busy" when NOT


CALL CENTER ACTIVITY
1/1/2018 8:04:25 AM
455
Current Calls
234
Staff on Calls
8
Staff Idle
2
Escalated Calls
3
Dropped Calls
1
Emergency Calls

Writing Better Survey Questions – 10 Commandments

From https://www.qualtrics.com/blog/good-survey-questions/

1. Thou shalt avoid loaded questions or leading words

Subtle wording differences can produce great differences in results.
"Could", "should", and "might" all sound about the same, but may produce a 20% difference in agreement to a question (the supreme court could /should/ might change the limits on free speech in light of terrorist activities).
Strong words, such as "prohibit" may represent control or action and influence your results.
If thou desire is clean data, this commandment must be adhered to.

2. Thou shalt honor the ordering of questions

Survey questions placed out of order or out of context should almost always be avoided.
In general, a funnel approach is advised. Broad and general questions at the beginning of the questionnaire as a warm-up. Then more specific questions, followed by more general easy to answer questions (like demographics) at the end of the questionnaire.

3. Thou shalt avoid non-specific questions

Build questions that are clearly understood.
Do you like orange juice? This is very unclear...do I like what about orange juice? Taste, texture, nutritional content, vitamin C, cancer prevention properties, the current price, concentrate, or fresh squeezed?
Be specific in what you want to know.
Issues of meaning and frequency are particularly difficult to specify: Do you watch TV regularly? (What is regularly? Does it matter what I watch? Is a DVD the same as TV?).

4. Thy question wording shall not be confusing or unfamiliar

Asking about caloric content, bits, bytes and other industry-specific jargon and acronyms is confusing.
Thy respondent must understand thy language level so that thy survey response rates may be high and data clean.

5. Thou shalt not force respondents to answer

Respondents may not want, or may not be able to provide the information requested.
Privacy is an important issue to most people.

Questions about income, occupation, finances, family life, personal hygiene and personal, political or religious beliefs can be too intrusive and rejected by the respondent.
Incentives and assurances of confidentiality will help in addition to adding a "prefer not to answer" option.

6. Thou shalt not adulterate your survey with non-exhaustive listings

Do you have all of the options covered?
If you are unsure, conduct a pretest using "Other (please specify)" as an option.
When building multiple choice survey questions, make sure to cover at least 90% of the respondent answers so thy data shall be clean.

7. Thou shalt use unbalanced listings skillfully

Unbalanced likert scales may be appropriate for some situations and biased in others.
Biased example: When measuring alcohol consumption patterns, one study used a quantity scale that made the heavy drinker appear in the middle of the scale with the polar ends reflecting no consumption and an impossible amount to consume. This is a sin.
Appropriate example: We expect all hotels to offer good care and may use a scale of excellent, very good, good, and fair. We do not expect poor care.

8. Thou shalt abolish double barreled questions

What is the fastest and most economical Internet service for you?
The fastest is certainly not the most economical. double barreled questions will lead to inaccurate data since your respondents may not know which part to answer.
Thou shalt never ask two questions simultaneously.

9. Honor thy dichotomous questions

When building a survey, answers should always be independent.
The question "Do you think basketball players as being independent agents or as employees of their team?" is not dichotomous. Many believe that basketball players are both.

10. Thou shalt use long questions wisely

Multiple choice questions are the longest and most complex. Free text answers are the shortest and easiest to answer. When you increase the length of questions and surveys, you decrease the chance of receiving a completed response.

Exercise: Survey Data Design

Step 1: Design the survey in groups (15 minutes)

Working in a group,
design a survey to measure
team sustainability over time.

Remember to consider:

1. What dimensions are important (e.g. sustainable pace, tech debt, skill development)
2. What questions would you ask?
3. How might you show the data and make decisions

Step 2: Share and discuss the questions you defined with the group (10 minutes)

Quality – How Well (Done things stay done)

What is the intended behavior?
Help teams continuously see if actions they are taking are causing a delay in delivery or any decline in product quality that would lead to customer dissatisfaction.

Examples
- Escaped defects – defects detected outside of the team
- Forecast days to complete all defects if team did nothing else
- Measure of release readiness - crowdsourced view on releasability
- Passing test percentage (sometimes of the last 5 runs)

Group Exercise (form groups of 3 to 5 people)
1. Brainstorm and discuss any measures of Quality you currently have and write one post-it note per measure
2. Brainstorm and discuss what data you have that may be used as a measure of this metric and add a post-it one note per measure
3. Discuss and dot vote what measure you feel as a group offers the best way to detect improvement or decline for this metric
4. For the top choice, Brainstorm measures that would detect if this metric was improved at the expense of everything else
5. Complete the paragraph at the bottom of this sheet.

1. List Current Quality Measures **2. List New Potential Quality Measures** **4. How would we detect when overdriven**

We will measure _____ trended every _____ (day/week/sprint) as our measure of Quality. We will also measure _____ to detect if we over-drive improving Quality and suffer elsewhere.

Responsiveness – How Fast (The right things get done fast)

What is the intended behavior?
Help teams continuously see if they are responsive to new requests, especially those of the highest priority and criticality. Avoid measuring responsiveness for non-critical items which causes poor prioritization.

Examples
- Lead Time for high(er) severity defects
- Cycle time for committed items (eg. items chosen for a sprint)
- Lead time for items that have ever been Top 5 in the backlog

Group Exercise (form groups of 3 to 5 people)
1. Brainstorm and discuss any measures of Quality you currently have and write one post-it note per measure
2. Brainstorm and discuss what data you have that may be used as a measure of this metric and add a post-it one note per measure
3. Discuss and dot vote what measure you feel as a group offers the best way to detect improvement or decline for this metric
4. For the top choice, Brainstorm measures that would detect if this metric was improved at the expense of everything else
5. Complete the paragraph at the bottom of this sheet.

1. List Current Responsiveness Measures **2. List New Potential Responsiveness Measures** **4. How would we detect when overdriven**

We will measure _____ trended every _____ (days/week/sprint) as our measure of Responsiveness. We will also measure _____ to detect if we over-drive improving Responsiveness and suffer elsewhere.

Productivity – How Much (Things are getting done)

What is the intended behavior?
Help teams continuously see the delivery rate of completed work and see if actions they are taking are causing any increase or decrease of that delivery rate.

Examples
- Throughput. Completed items per week (divided by team size?)
- Velocity. Sum of completed points per sprint
- Releases per day/week

Group Exercise (form groups of 3 to 5 people)
1. Brainstorm and discuss any measures of Quality you currently have and write one post-it note per measure
2. Brainstorm and discuss what data you have that may be used as a measure of this metric and add a post-it one note per measure
3. Discuss and dot vote what measure you feel as a group offers the best way to detect improvement or decline for this metric
4. For the top choice, Brainstorm measures that would detect if this metric was improved at the expense of everything else
5. Complete the paragraph at the bottom of this sheet.

1. List Current Productivity Measures **2. List New Potential Productivity Measures** **4. How would we detect when overdriven**

We will measure _____ trended every _____ (day/week/sprint) as our measure of Productivity. We will also measure _____ to detect if we over-drive improving Productivity and suffer elsewhere.

Predictability – How Consistently (Things are getting done consistently)

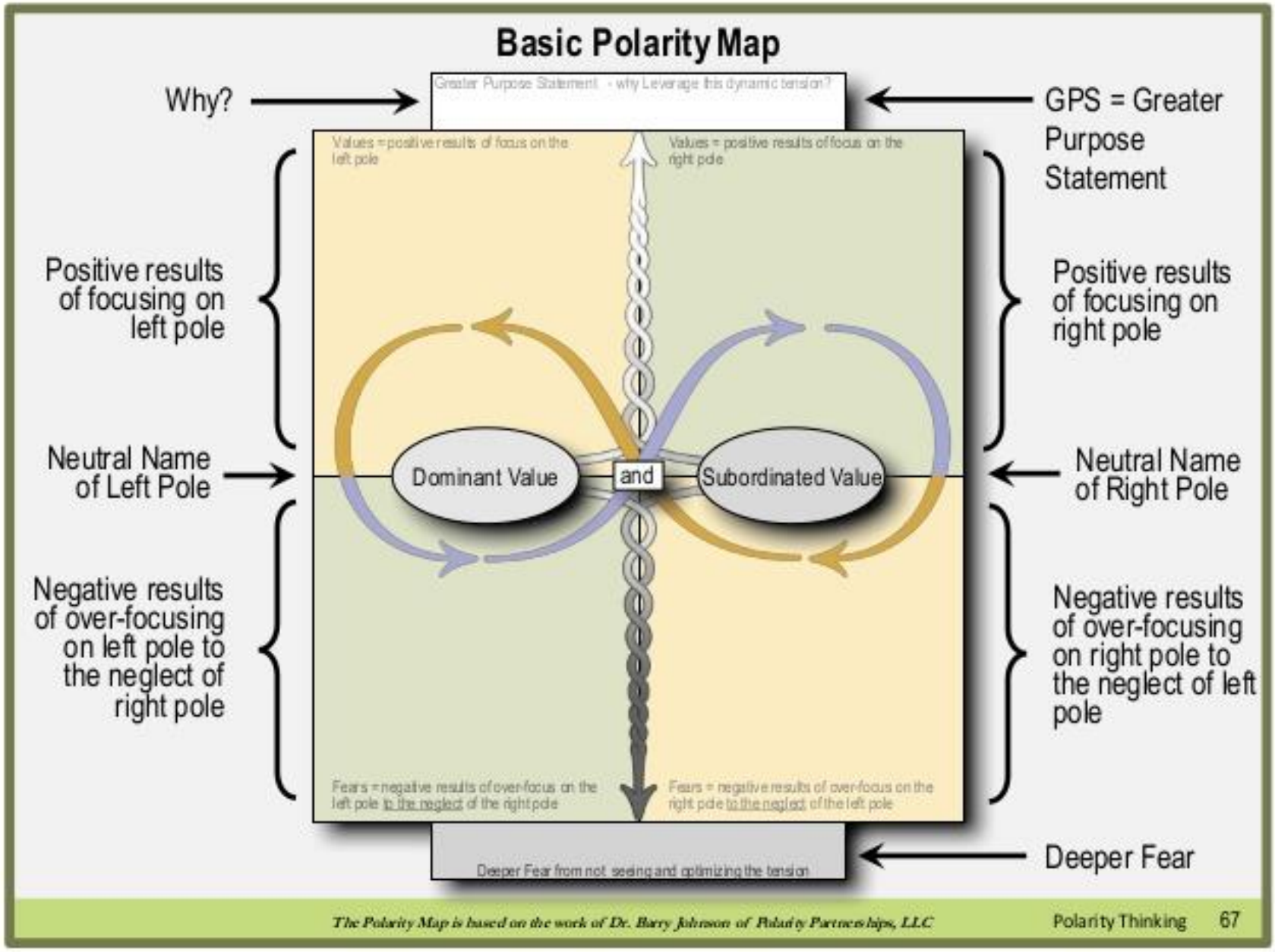
What is the intended behavior?
Help teams continuously see if their delivery rate (productivity) is consistent and see if actions they are taking are causing uncertainty in that rate. Low predictable measure means less ability to forecast.

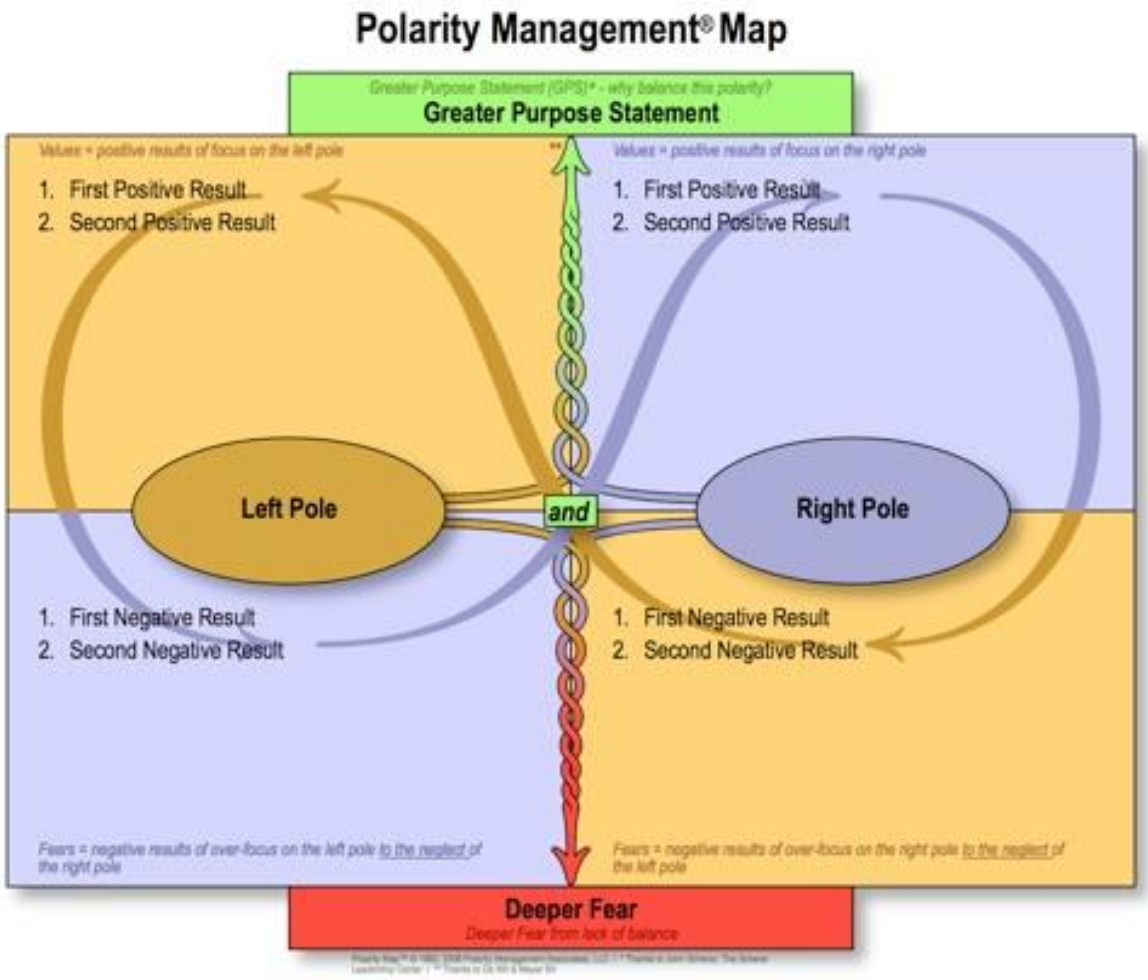
Examples
- Variation of the productivity measure (Standard Deviation)
- Coefficient of Variation of productivity measure (S.D./Average)
- Committed work / Delivered Work ratio

Group Exercise (form groups of 3 to 5 people)
1. Brainstorm and discuss any measures of Quality you currently have and write one post-it note per measure
2. Brainstorm and discuss what data you have that may be used as a measure of this metric and add a post-it one note per measure
3. Discuss and dot vote what measure you feel as a group offers the best way to detect improvement or decline for this metric
4. For the top choice, Brainstorm measures that would detect if this metric was improved at the expense of everything else
5. Complete the paragraph at the bottom of this sheet.

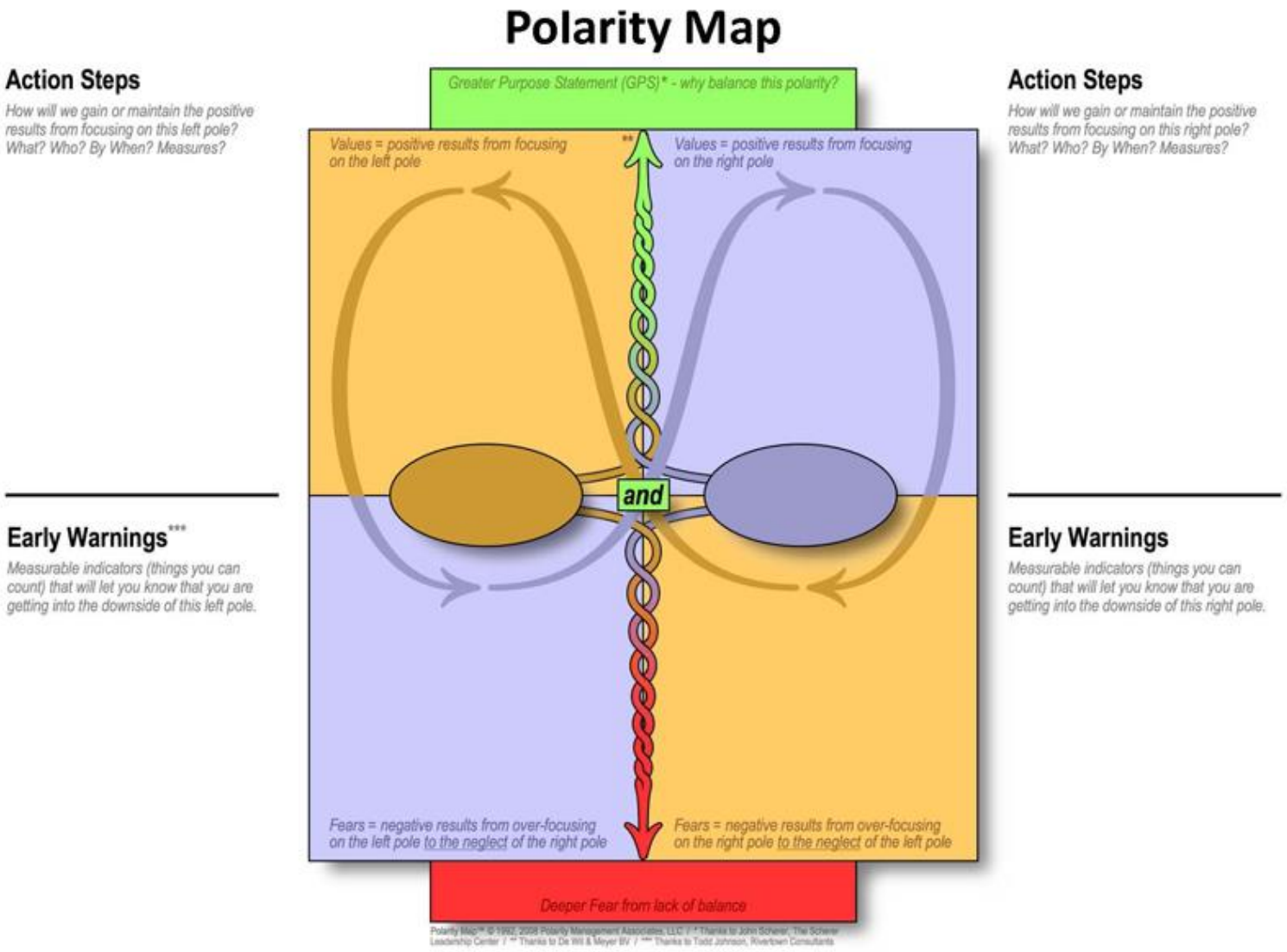
1. List Current Predictability Measures **2. List New Potential Predictability Measures** **4. How would we detect when overdriven**

We will measure _____ trended every _____ (day/week/sprint) as our measure of Predictability. We will also measure _____ to detect if we over-drive improving Predictability and suffer elsewhere.

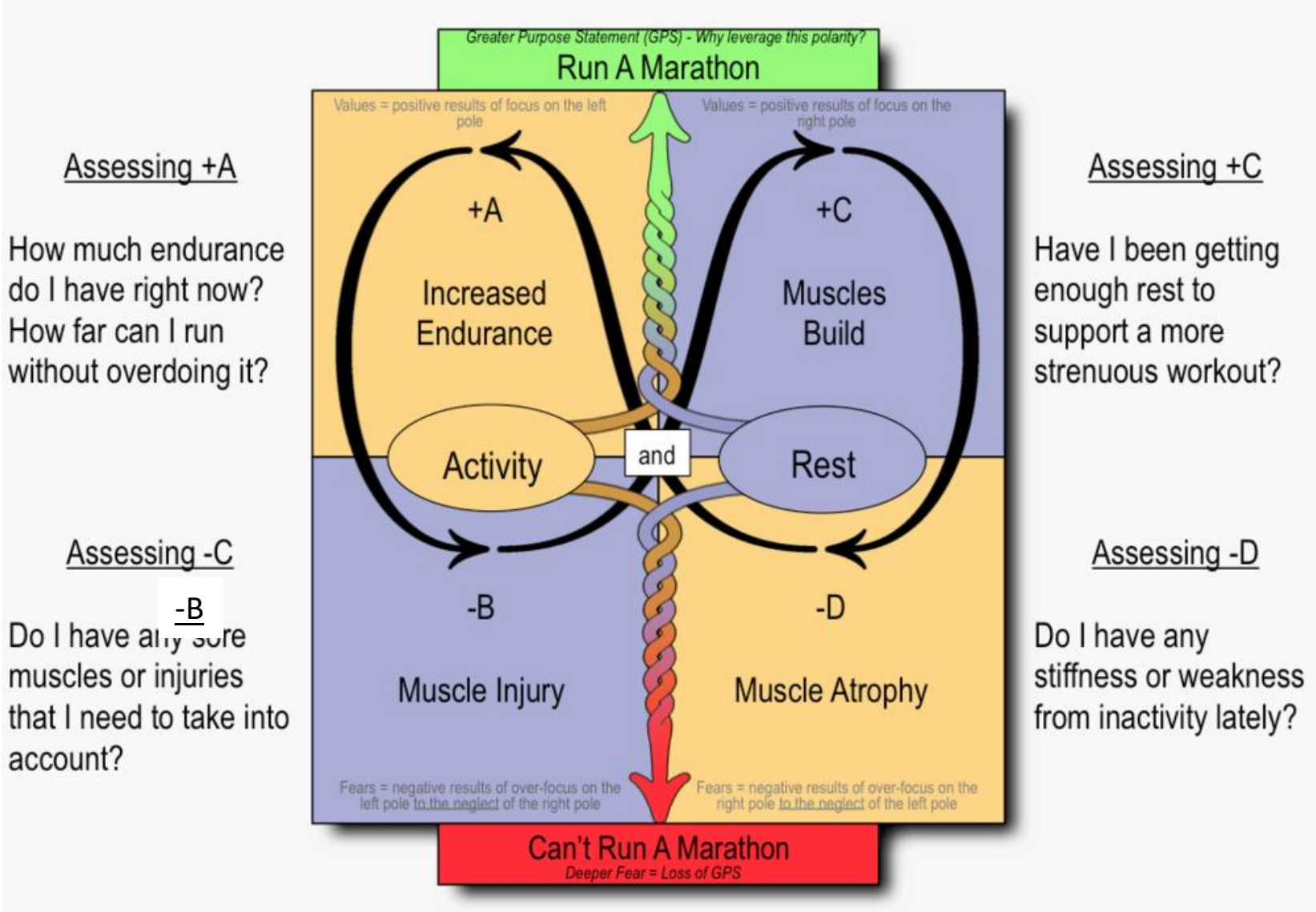

Basic Polarity Map
Why?
Greater Purpose Statement: - why Leverage this dynamic tension?
GPS = Greater Purpose Statement
Values = positive results of focus on the left pole
Values = positive results of focus on the right pole
Positive results of focusing on left pole
Positive results of focusing on right pole
Neutral Name of Left Pole
Dominant Value
and
Subordinated Value
Neutral Name of Right Pole
Negative results of over-focusing on left pole to the neglect of right pole
Negative results of over-focusing on right pole to the neglect of left pole
Fears = negative results of over-focus on the left pole to the neglect of the right pole
Fears = negative results of over-focus on the right pole to the neglect of the left pole
Deeper Fear from not seeing and optimizing the tension
Deeper Fear
The Polarity Map is based on the work of Dr. Barry Johnson of Polarity Partnerships, LLC
Polarity Thinking 67



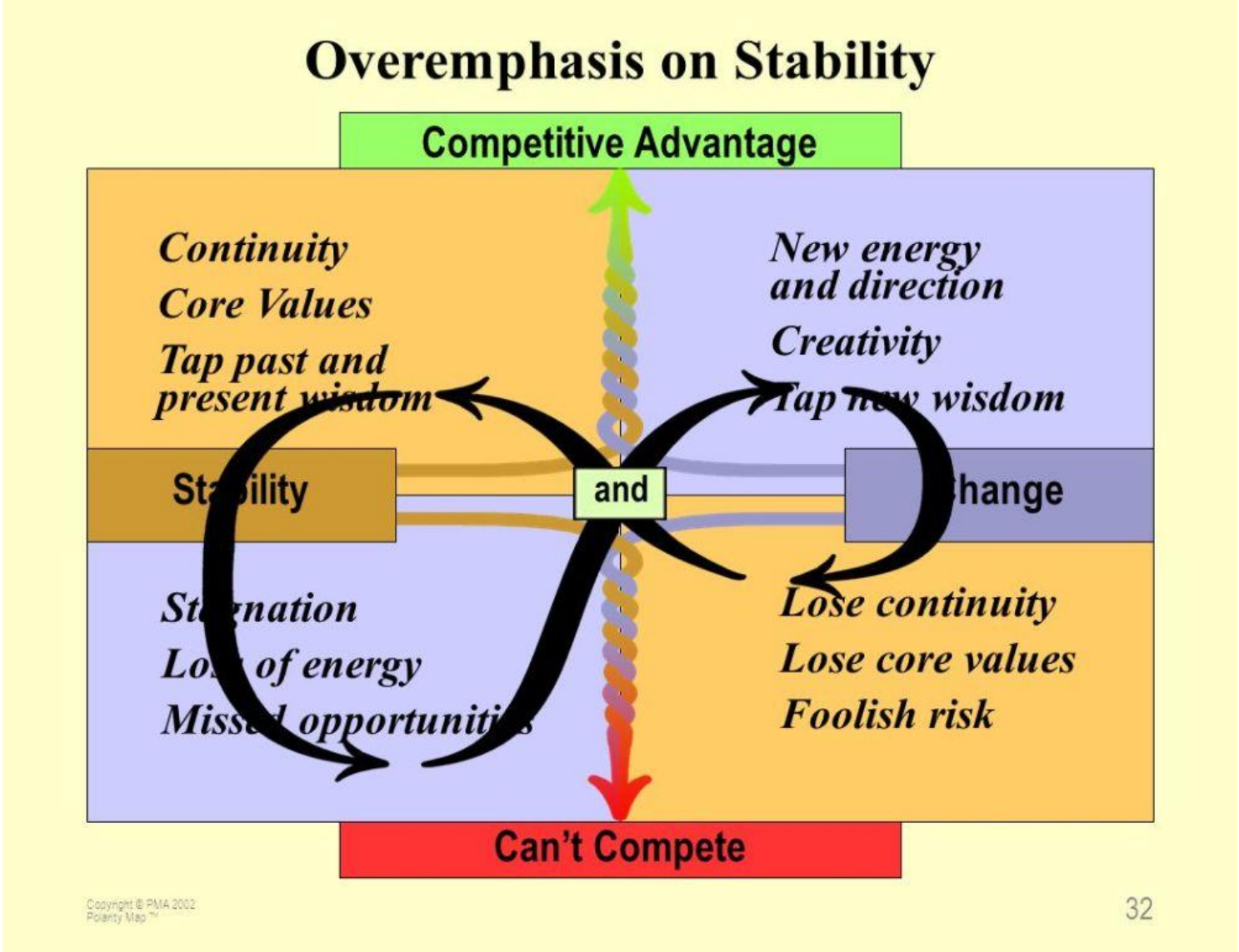
Credit: Polarity Management, Barry Johnson



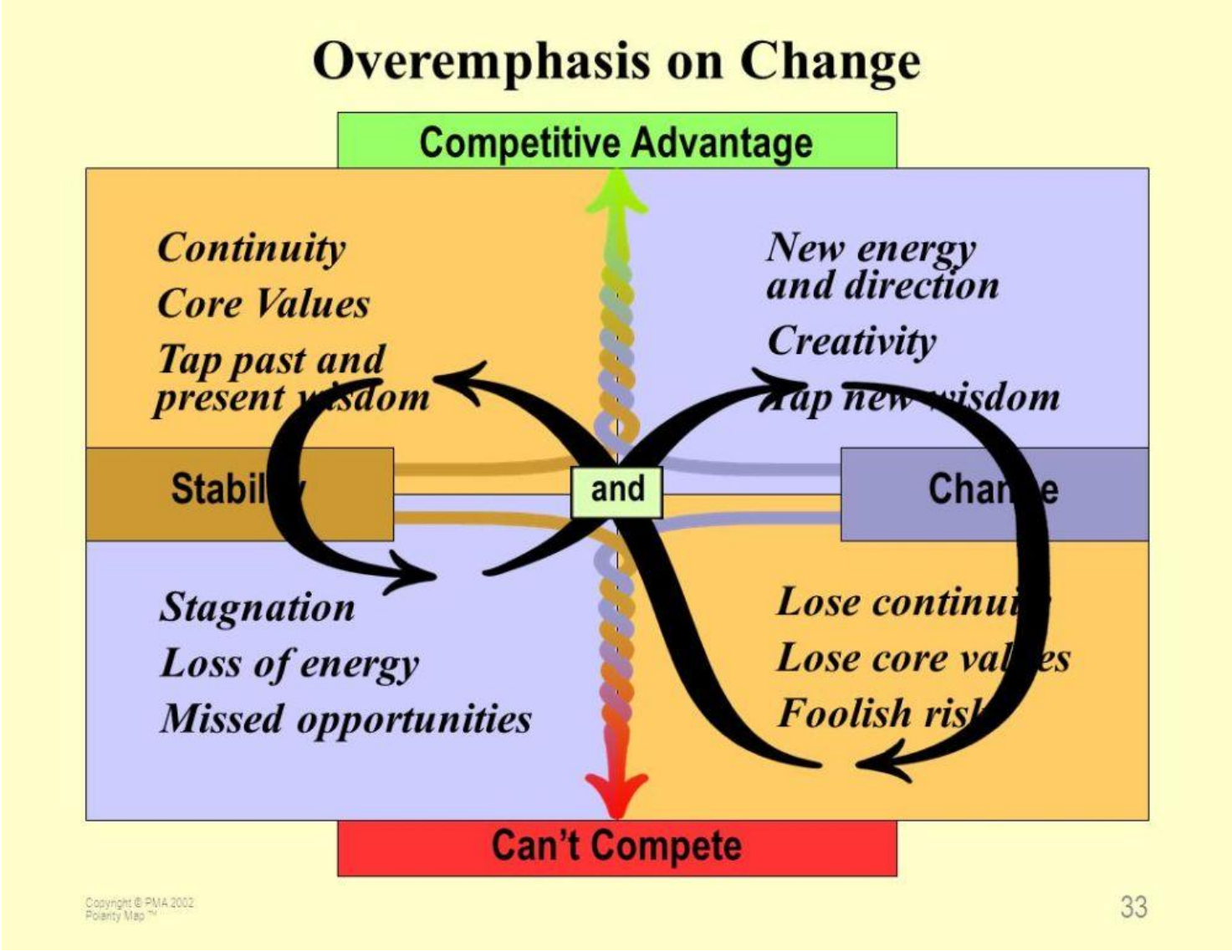
Credit: Polarity Management, Barry Johnson



Credit: Polarity Management, Barry Johnson



https://newthoughtevolutionary.wordpress.com/2017/12/04/thriving-skill-mastering-polarity-management-part-1/



https://newthoughtevolutionary.wordpress.com/2017/12/04/thriving-skill-mastering-polarity-management-part-1/

Quality -
Customer Reported Defects (escaped defects)

Positive results from focusing more this polarity

- Shorter release (in dev anyway) cycles
- Faster testing of "beta" ideas in the market for feedback

Positive results from focusing more this polarity

- Happier Customers due to "it always works"
- Good market reputation of quality
- Lower task switching to fix prod defects "now"

Totally ignore this metric

Totally focus on this metric

Early warning signs that you are getting into the negative aspect of this side

- More prod defects
- Increase negative sentiment on Twitter
- Slow uptake of latest released versions

Negative results from focusing more this polarity

- Sad Customers (who can't do their work)
- Poor market reputation
- Customers slow adopt new releases (waiting till SP1)

Negative results from focusing more this polarity

- Long release cycles
- Few released features
- Easy features prioritized ahead of valuable ones

Early warning signs that you are getting into the negative aspect of this side

- Longer feature lead time
- Fewer significant features per release

Exercise

Part 1: Polarity Identification and Early Warning Indicators

1. Select a set of candidate metrics decided in the ODIM exercises
2. Break into groups and perform a polarity exercise on one of these metrics
3. Present your early warning indicators to the full group.

Polarity Worksheet for Metric _____

Polarity Worksheet for Metric _____

Workshop Exercise: Data Capture and Gaps

1. What data don't you have to move forward?
2. Who has this data, and how can we get it?

Data	Who/Tool	How

APPENDICES

Phases of Dashboards

There are three phases to dashboards / data that support service / product management.

Planning what work to do in the future

make investment decisions. eg. Projects, headcount...

Tracking work once it is Committed

Helps understand status and tactical allocation.

Measuring the impact of past work

provide feedback on value to refine strategy

- - - Feedback - use ● insights to improve ▲

Common Current Capability

Now.

12 Months?
?

Lots -

- Doing

-

Some -

- Planning

-

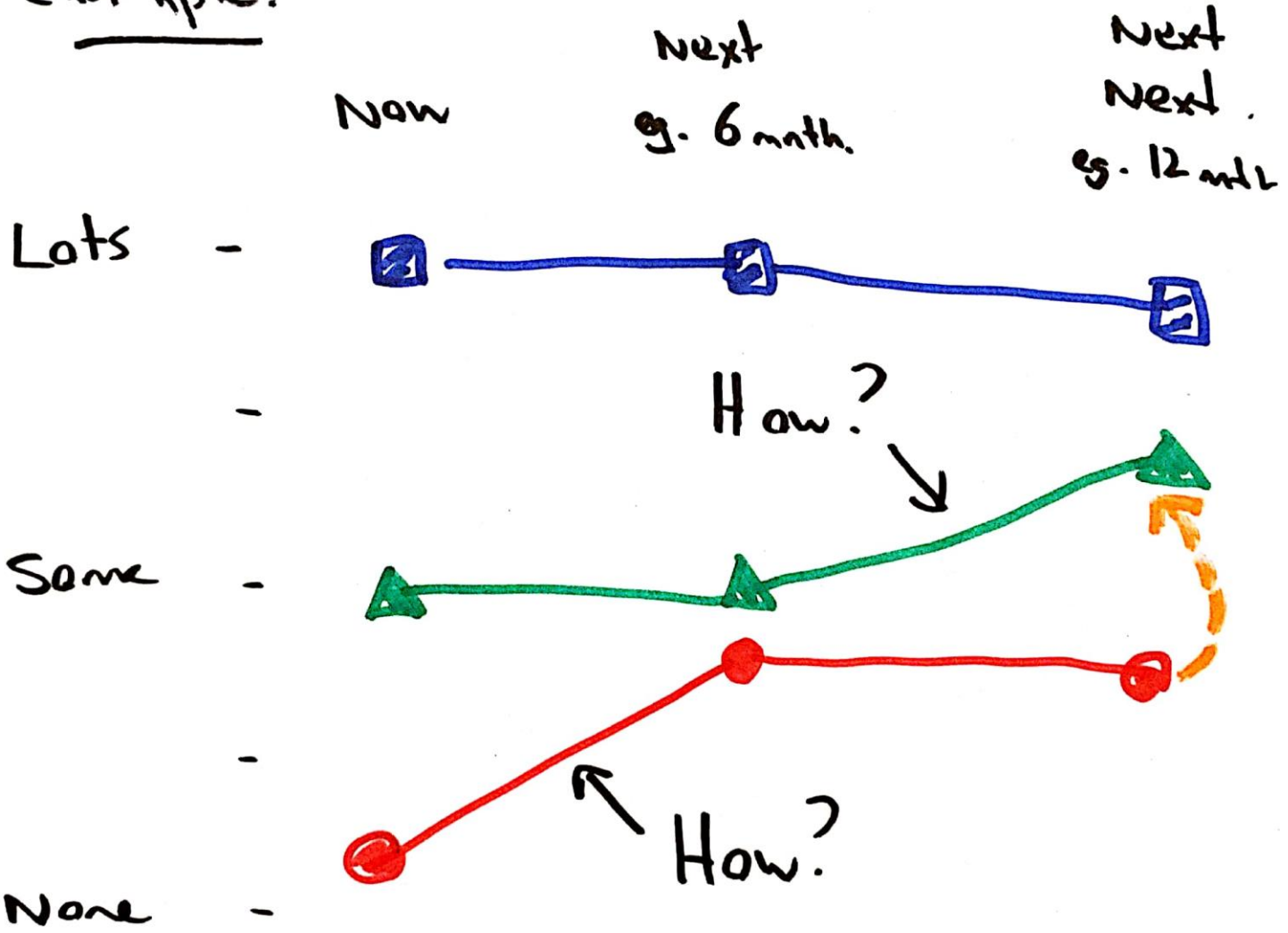
- Impact

None -

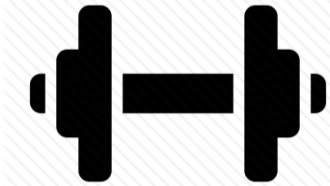
Q. If we were TWICE as good at one of these in 12 months, what Phase would move the "outcome" needle the most?

Capability Strategy Plan

Example.



Workshop Exercise



Identifying YOUR data/dashboard investments

1. List YOUR categories of data and dashboards. Eg. Team improvement, feature delivery, prioritization, capacity planning, etc.
2. Decide as a group how much investment is given to each category now
3. Decide as a group which ones return the most “value” (and define value)
4. What is the “ideal” investment mix?

Fill in the chart on the next page. Define the triangle, square and circle to match your categories, and add a 6, 12, 18 month plan for where you agree investment should be spent.

Current Capability

Think about your current data and dashboards. Which phases have most support? Place a  ,  ,  below

Now

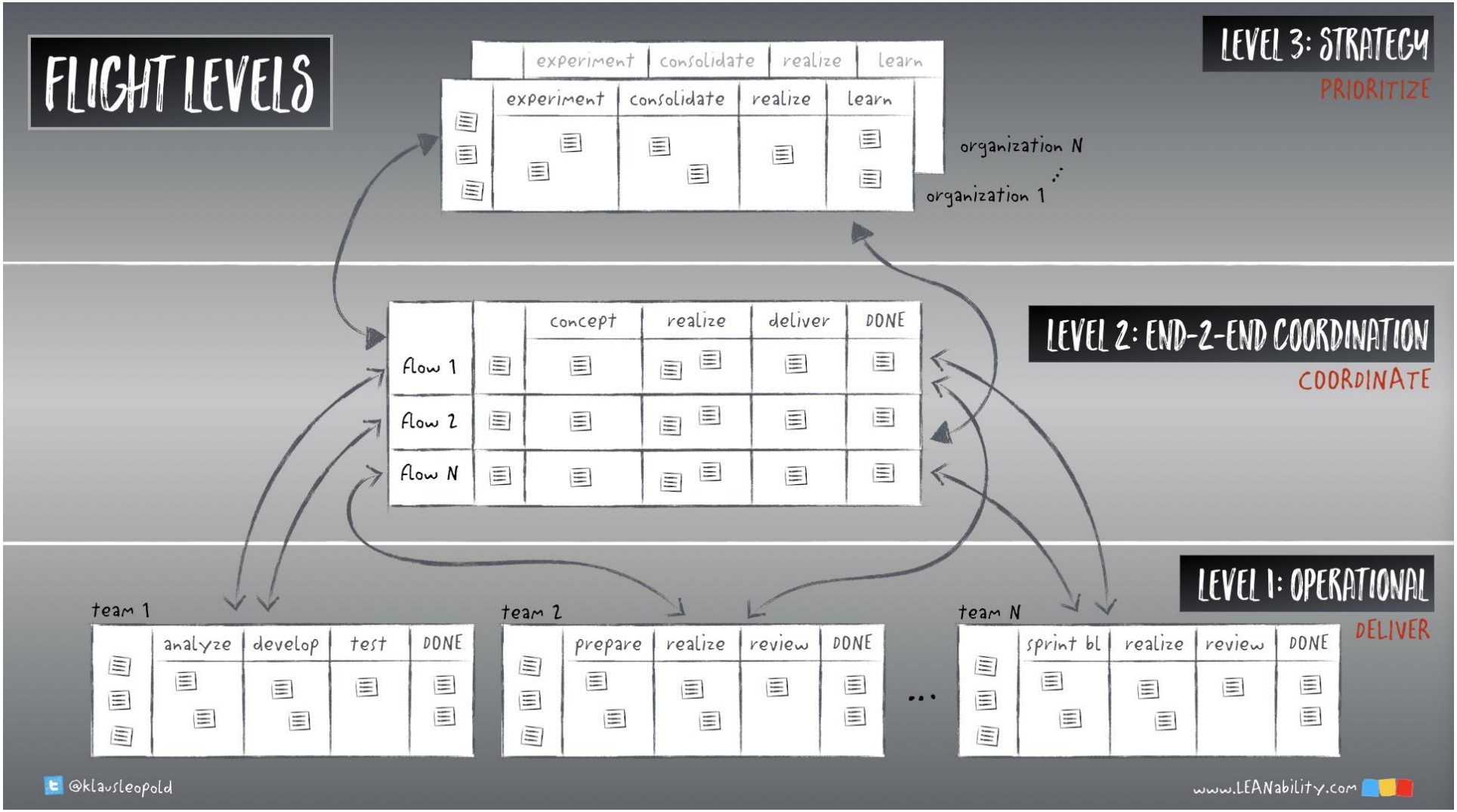
Lots -

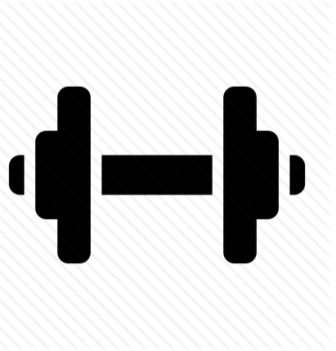
-

Some -

-

Not at all -

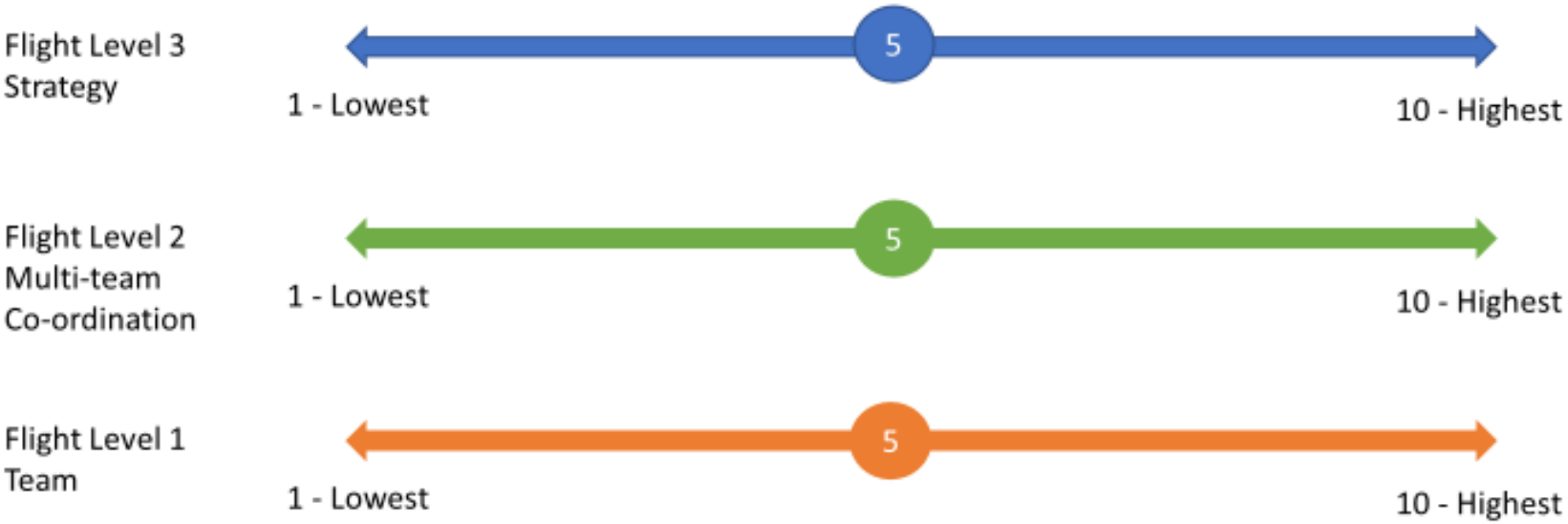

FLIGHT LEVELS
LEVEL 3: STRATEGY
PRIORITIZE
experiment
consolidate
realize
learn
organization N
organization 1
LEVEL 2: END-2-END COORDINATION
COORDINATE
concept
realize
deliver
DONE
flow 1
flow 2
flow N
LEVEL 1: OPERATIONAL
DELIVER
team 1
analyze
develop
test
DONE
team 2
prepare
realize
review
DONE
team N
sprint bl
realize
review
DONE
@klausleopold
www.LEANability.com



Exercise Q – discuss in groups

- How do you feel about the process currently used at each level now?

As a group decide on the current "satisfaction" number



Exercise Q – discuss in groups

- List the obstacles you see in moving the needle for each level right.

Flight Level 3
Strategy obstacles

Flight Level 2
Multi-team
Co-ordination obstacles

Flight Level 1
Team obstacles

Level 3 – Strategy (Prioritize)

- Example Outcomes

1. The blend (types) of started experiments matches a larger strategy mix
   - Meaning: We start things for a reason based on furthering a strategic goal
   - Needs: A way to know what investment level is desired for each strategic goal

2. Acceptable (or improving) chosen experiment delivery lead-time
   - Meaning: Experiments we choose to deliver get delivered when needed.
   - Needs: Lower experiments in progress and consistent in-progress age

3. Acceptable (or improving) feedback action and cycle-time for experiments
   - Meaning: Lessons from experiments delivered get incorporated into actions sooner.
   - Needs: A way to assess and capture lessons from experiments; could include waiting!

Note: If experiment isn't the right word, replace it with epic, feature, MVP, etc.

Level 2 – End2End Coordination (Coordinate)

- Example Outcomes

1. The right experiments are started and queued
   - Meaning: We start things for a "reason" (e.g. higher priority) as constraints allow
   - Needs: A way to know why to start some work and queue others

2. The work gets to the right teams in the right order
   - Meaning: Work is partitioned and handed to the teams that need to deliver it
   - Needs: A way to know what work to pass to a team and what teams will be next…

3. Acceptable (or improving) lead-time delivery of experiments
   - Meaning: Work is handed to necessary teams in a logical and timely fashion
   - Needs: Lower experiments' in process and impediments are handled appropriately

Level 1 – Operational (Deliver)

- Example Outcomes
  1. The right things are started and queued
     - Meaning: We start things for a "reason" as constraints allow
     - Needs: A way to know why to start some work and queue others
  2. Acceptable (or improving) external lead-times
     - Meaning: The necessary things get started in acceptable time
     - Needs: Correct team skillsets and tools (type and amount) for the work
  3. Acceptable (or improving) internal cycle-times
     - Meaning: The things we start we finish in acceptable time
     - Needs: Lower work in progress and consistent in-progress age distribution
  4. Acceptable (or improving) quality
     - Meaning: The things done work as expected and needed
     - Needs: A way to measure "as expected" and "needed"

Exercise: Next steps

1. What are possible next steps – brainstorm ideas and cluster
2. What are the highest priorities – dot vote after discussion
3. Are there any dependencies before we start the highest priorities?
4. What skills or knowledge will be required to complete these actions?
5. What impediments are in the way of completing these actions?

What	Who	When