Focused Objective
forecasting - risk - staff - cost of delay

Forecasting using Data

Capturing and using data for forecasting

Data Driven Coaching

Capturing and using data for coaching

Workshop Manual (v6)

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

Version 6.0 A4

Workshop Exercise: Goals and Key Questions

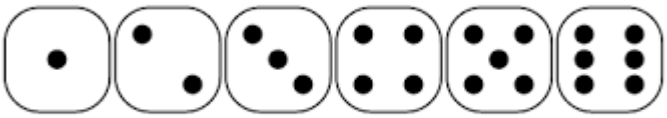
“35” Exercise
On an index card, write down your answer to -
“What is one question you would be disappointed if you left this workshop NOT learning or knowing?”

Workshop Goals and Outcomes
Q1. What are the key lessons of this workshop?

Q2. What are the outputs of this workshop?

Understanding probability - Exercises

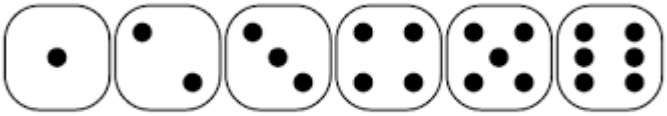
Q1. How many different possible values are there for a standard six-sided dice?



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?

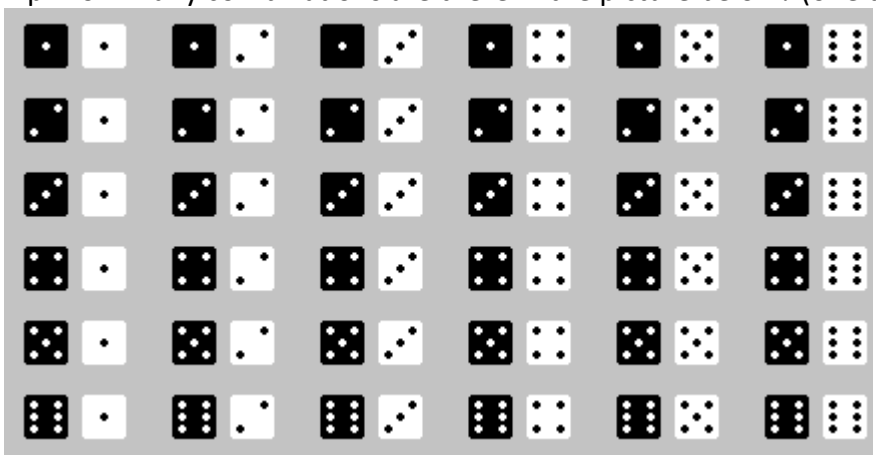
A:

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?

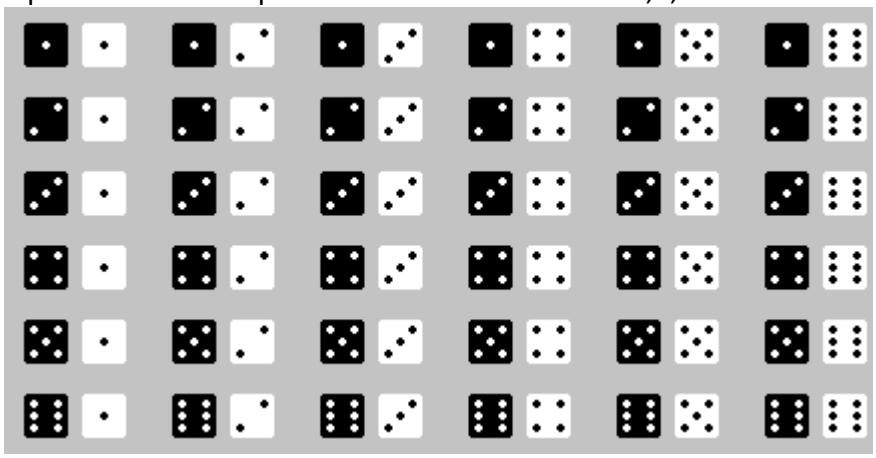
Tip: How many combinations are there in the picture below? (one dice is black, the other white)



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

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

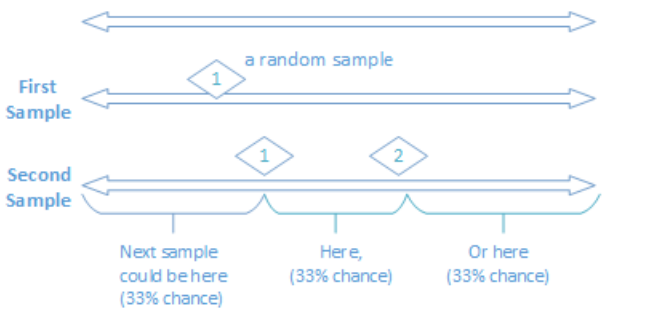
Probability the next sample is lower than the lowest sample so far after "n" = $\frac{1}{(n+1)}$ X 100

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

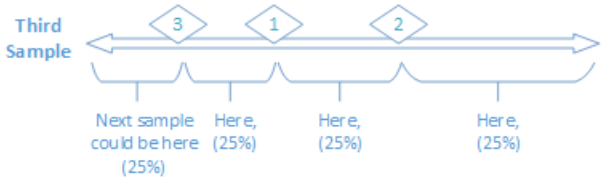
Samples so far (n)	Probability for each interval	Probability next sample in range
1	50.00%	0.00%
2	33.33%	33.33%
3	25.00%	50.00%
4	20.00%	60.00%
5	16.67%	66.67%
6	14.29%	71.43%
7	12.50%	75.00%
8	11.11%	77.78%
9	10.00%	80.00%
10	9.09%	81.82%
11	8.33%	83.33%
12	7.69%	84.62%
13	7.14%	85.71%
14	6.67%	86.67%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

3 x 6 Sided Dice

2 x 10 Sided

42

7

99

Note: Rolling a 00 and 0 = 100

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.

Questions and discussion topics

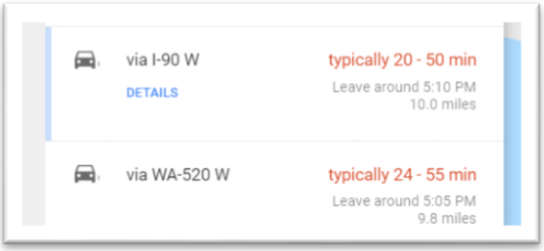
1. **What probability distribution is a single roll?**
2. **What guarantee do I have that I have found the range expected?**
3. **What happens if the data is a Normal (bell curve) distribution?**
4. **What happens if the data is left or right skewed?**

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

“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 10 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 5-10 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 5-10 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 15 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?

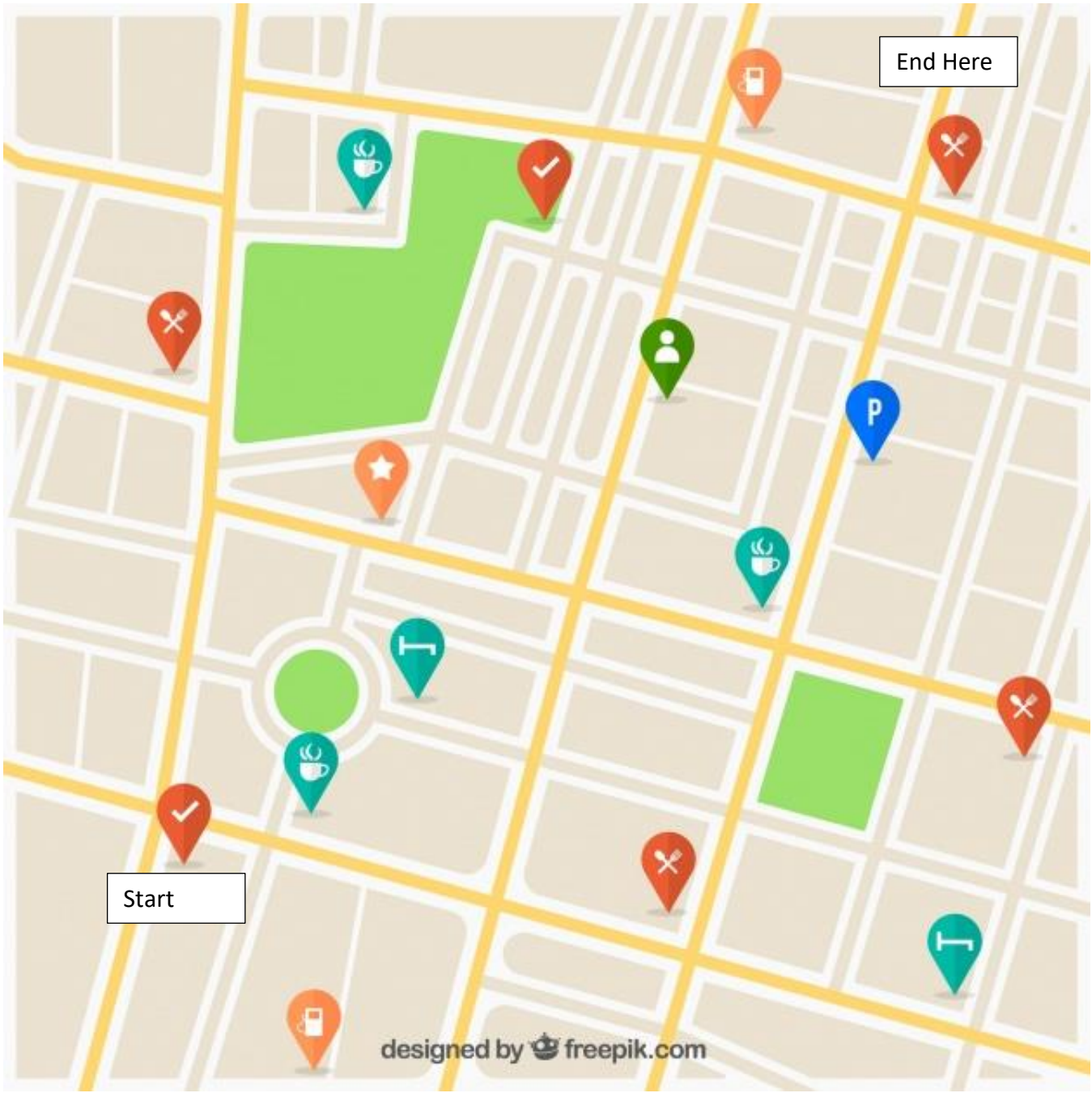


Figure 1 - Example road network diagram to use for discussing your algorithm.
Image attribution: Travel vector created by freepik - www.freepik.com

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)							

4. Compute the probability. Put best (shortest) # weeks in 1/7th, next shortest in 2/7th, etc.

1/7th chance (14%) =

2/7th chance (29%) =

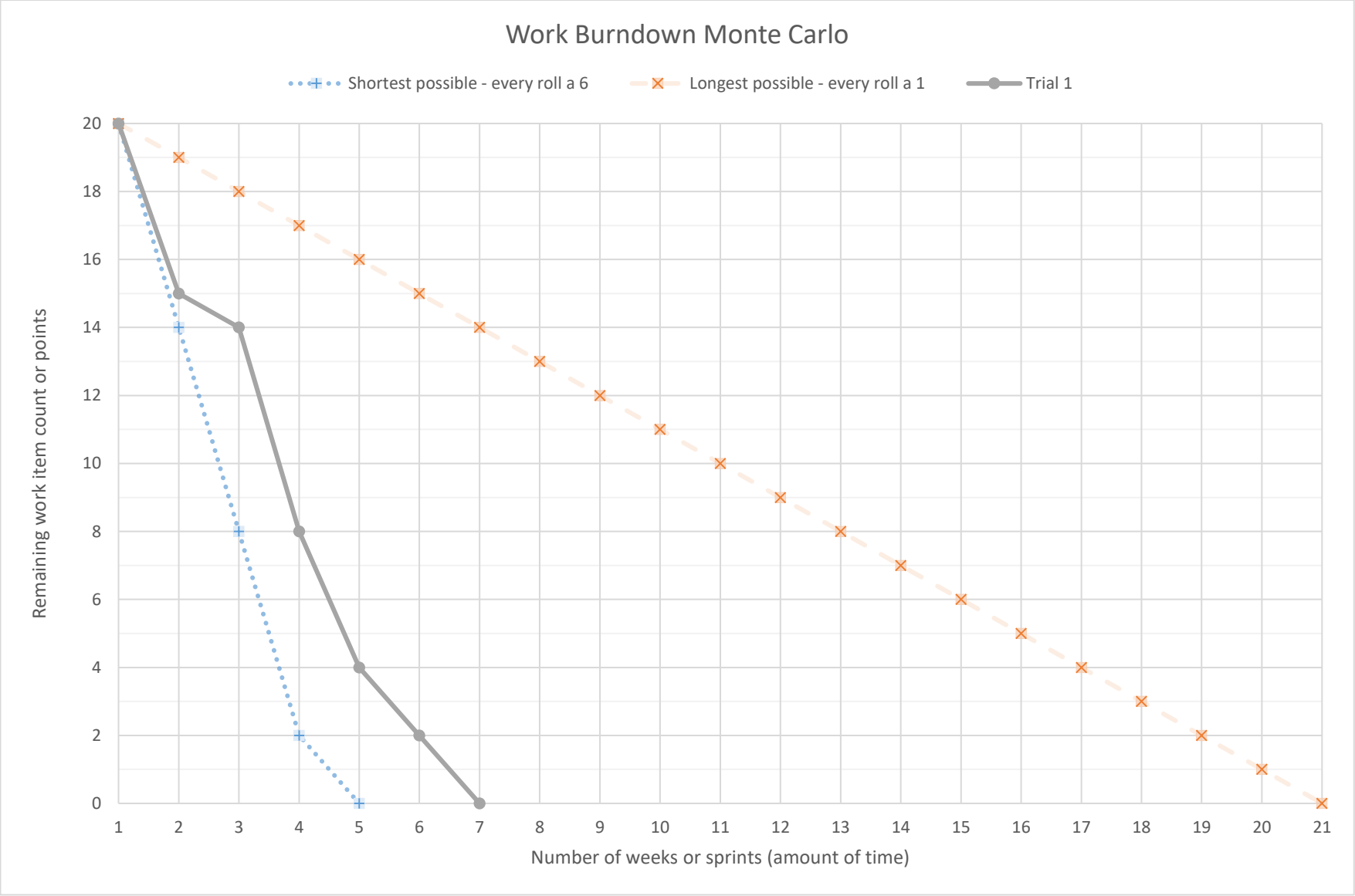
3/7th chance (43%) =

4/7th chance (57%) =

5/7th change (71%) =

6/7th chance (86%) =

7/7th chance (100%) =



Circle the 4th line counting from the left (the lowest) – this has a 4/7th chance (> 50%)

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**

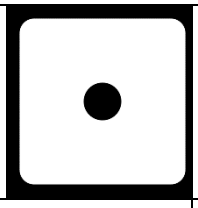
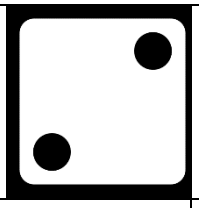
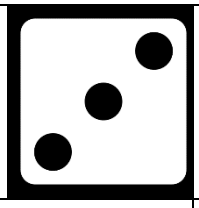
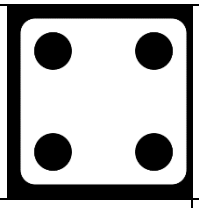
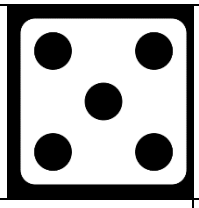
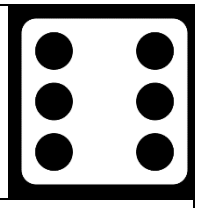
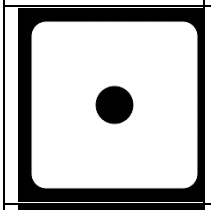
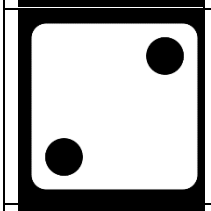
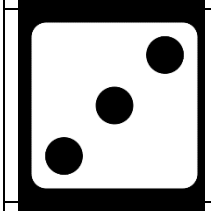
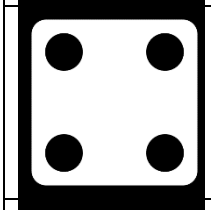
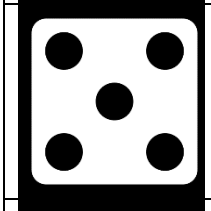
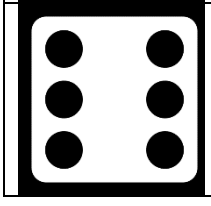
Total Stories in 10 Features (or epics)	Count trial sum groups at least 30,40, 50, etc. stories	(Count / 11) Likelihood
At least **20** stories	11	1
At least **30** stories		
At least **40** stories		
At least **50** stories		
At least **60** stories		
At least **70** stories		
At least **80** stories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

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		

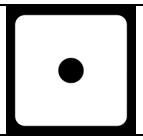
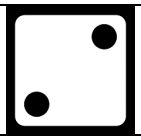
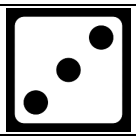
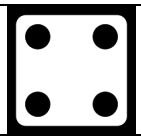
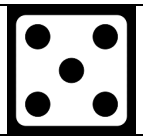
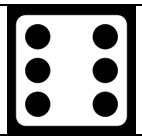
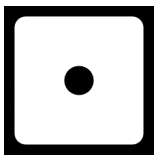
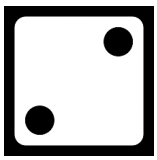
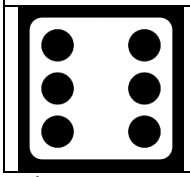
This value is 0 to 1 Multiply it by 100 to get a percentage. 0% = no chance, 100% means every trial achieved at least this level.

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

First dice throw

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!

						
	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

Capture Recapture Exercise 1: **DO NOT START UNTIL ASKED**

Circle the spelling mistakes in the following paragraph.

George new that he shouldn't drink alchohol on a Wedsday night, especially since his govermnet proffesor had schedualed an important exam on Thrusday. However, he beleived he would loose his friends if he didn't go out with them. The pressure to fit in with his peers was worst then the fear of bad grades. To be popular among his friends, one had to be either a musclar athelete or a wild and crazy drinker. George realy could not concieve how it was posible for a student to consume huge quanities of liquor and still suceed in school. Maybe the drinkers were just more briliant than he was. He didn't even enjoy the passtime of spending ours in a bar trying to persue a temperary feeling of excitement and "fun." Somehow he expected the cheif of campus security to catch him and the university administration to expell him. But George didn't posses enough courage to express his opion to his friends. He was certian they would tell him to mind his own buisness. Also, he did't want to be seperated from his friends. So he planed to meet them at a local restaraunt, have a few drinks, leave early, take some asprin, and spend a few ours studing for the exam.

Total mistakes found:

Total mistakes found by both groups:

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

Capture-Recapture Exercise Answers:

George new that he shouldn't drink alchohol on a Wedsday night, especially since his govermnet proffesor had schedualed an important exam on Thrusday. However, he beleived he would loose his friends if he didn't go out with them. The pressure to fit in with his peers was worst then the fear of bad grades. To be popular among his friends, one had to be either a musclar athelete or a wild and crazy drinker. George realy could not concieve how it was posible for a student to consume huge quanities of liquor and still suceed in school. Maybe the drinkers were just more briliant than he was. He didn't even enjoy the passtime of spending ours in a bar trying to persue a temperary feeling of excitement and "fun." Somehow he expected the cheif of campus security to catch him and the university administration to expell him. But George didn't posses enough courage to express his opion to his friends. He was certian they would tell him to mind his own buisness. Also, he did't want to be seperated from his friends. So he planed to meet them at a local restaraunt, have a few drinks, leave early, take some asprin, and spend a few ours studing for the exam.

Corrected -

George knew that he shouldn't drink alcohol on a Wednesday night, especially since his government professor had scheduled an important exam on Thursday. However, he believed he would lose his friends if he didn't go out with them. The pressure to fit in with his peers was worse than the fear of bad grades. To be popular among his friends, one had to be either a muscular athlete or a wild and crazy drinker. George really could not conceive how it was possible for a student to consume huge quantities of liquor and still succeed in school. Maybe the drinkers were just more brilliant than he was. He didn't even enjoy the pastime of spending hours in a bar trying to pursue a temporary feeling of excitement and "fun." Somehow he expected the chief of campus security to catch him and the university administration to expel him. But George didn't possess enough courage to express his opinion to his friends. He was certain they would tell him to mind his own business. Also, he didn't want to be separated from his friends. So he planned to meet them at a local restaurant, have a few drinks, leave early, take some aspirin, and spend a few hours studying for the exam.

Total errors: 36

Source: **STUDENT LEARNING ASSISTANCE CENTER (SLAC).** Texas State University-San Marcos

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 (& should) we change payment providers before October to December rush season?
- Question 2: What CAN we get by the Q4 (October to December) 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. Cost of Delay Calculator Exercise spreadsheet (3 - Cost of Delay for New Credit Card Provider and Fraud Detection.xlsx)
    d. 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	* Value added per Month	Epic ID #	Epic Description	Estimated # Stories (size)
Checkout page updates	$0	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)	$1,000	6	Visa and MC Approval workflow	
		7	Visa and MC Refund workflow	
Support AMEX cards (new provider)	$5,000	8	AMEX Approval workflow	
		9	AMEX Refund workflow	
		10	AMEX additional fee warning	
Support Diners Club card type (new provider)	$500	11	DC Approval workflow	
		12	DC Refund workflow	
Support Bitcoin transactions	$1000 (+ competitive advantage)	13	Bitcoin Approval workflow	
		14	Bitcoin Refund workflow	
PCI Compliance yearly audit	$1,000 fine (+ risk exposure)	15	PCI Compliance Audit	
		16	PCI Compliance Resolution of Major Issues	
Fraud Detection features	$220	17	US Address fraud detection	
		18	Other country Address Fraud Detection	
Card Expiry Reminders	$2,000	19	Three-month before expiry reminder email	
		20	Create support desk issue one-month prior	

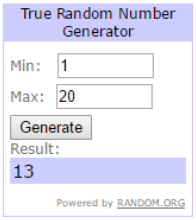
* The additional value estimate will be explained later on, for now, these are the numbers we will assume throughout this exercise.

Question 1: Can we change payment providers before the start of Q4 (October to December)?

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

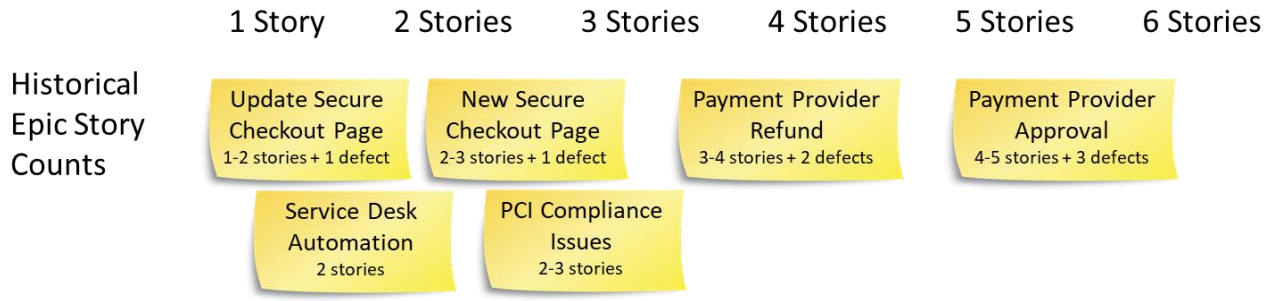
We are going to avoid estimating every epic by extrapolating how big all features are by random sampling and estimating only five epics.



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

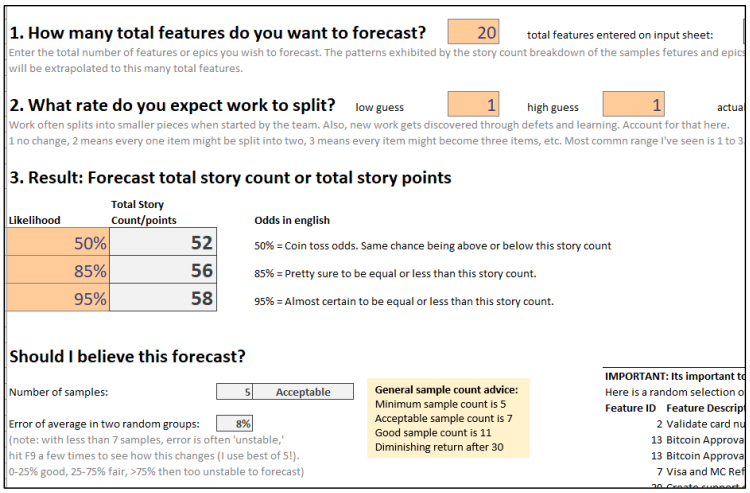
2 – validation logic might need to be learnt, 2 stories

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

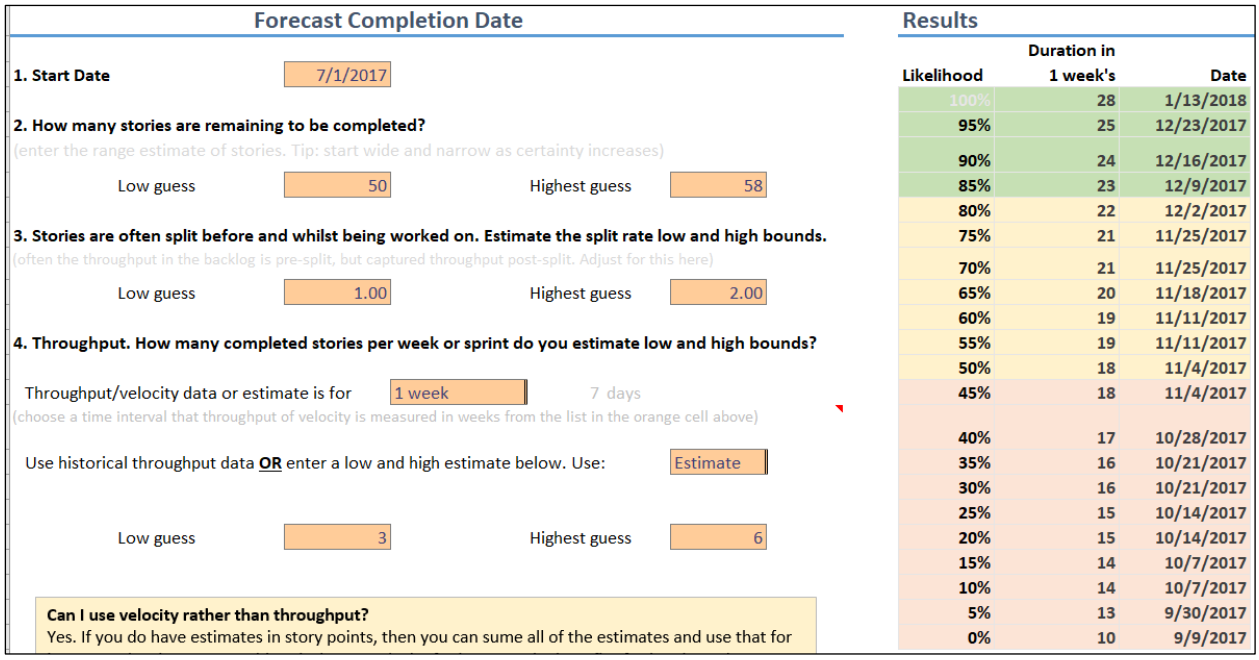
We are going to Monte Carlo forecast using the story count forecast and team historical performance to compute a duration and a likely delivery date.

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 the Q4 (October to December) rush season?

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 priority based on Cost of Delay

Here are some economic facts to help make decisions –

- Currently there are 1000 subscribers who pay $100 per month for pain (total rev $100,000/m).
- Visa and Mastercard currently make up 100% of transactions.
- Lack of American Express support is estimated to lose 50 to 100 subscribers per month.
- Lack of Diners Club support is estimated to lose 1 to 5 subscribers per month (but decreasing).
- A competitor has launched accepting Bitcoin support. Estimated loss of 10 transactions a month.
- 2% of subscribers who encounter expiring credit cards never return to renew (lost to competitor).
- PCI Compliance (credit card security) is audited yearly with fines of $1000 per month if missed (starting January). Penalty of data breach if this isn't performed are catastrophic (~20% total revenue risk per month).
- Fraud is an issue. Most fraud comes from overseas customers which is 20% of your business. 1% of those charges are disputed and refund and penalized an additional 10% fee (math: 20% of 1000 = 200 overseas subs. 1% of 200 = 2 chargebacks. 2 times $100 + $10 = $220).

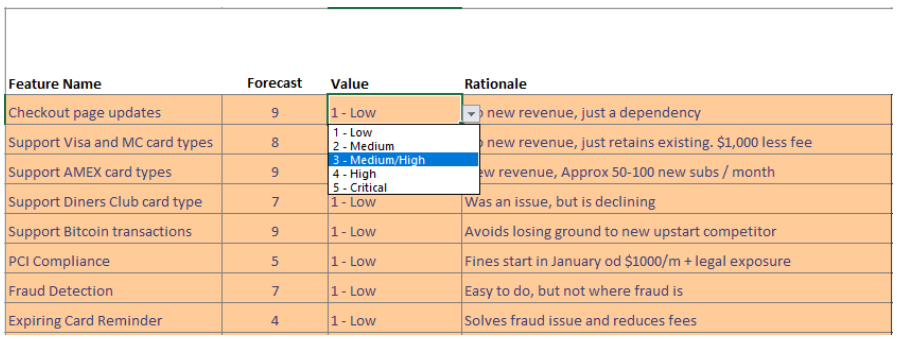
Exercise 3

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

1. Open the "3 - Cost of Delay for New Credit Card Provider and Fraud Detection.xlsx" spreadsheet
2. Select the "WSJF Prioritization (simple)" worksheet
3. Work as a group on the qualitative Cost of Delay value assumptions and decide where each feature ranks from 1 – Low to 5 – Critical. Look to separate value relative to each feature, don't get too hung up on particulars. I've already entered the size estimate for you based on the Reference Class Forecasts we did earlier, and also some example rationales to start with based on the facts above..

Hint: Table 1 - All proposed epics has some value estimates that you might refer to when making your own value estimate.

Figure 4 - Choose a relative value from the drop-down for each feature. Explain your reasoning in the Rationale column



Feature Name	Forecast	Value	Rationale
Checkout page updates	9	1 - Low	new revenue, just a dependency
Support Visa and MC card types	8	1 - Low 2 - Medium 3 - Medium/High 4 - High 5 - Critical	new revenue, just retains existing. $1,000 less fee
Support AMEX card types	9		ew revenue, Approx 50-100 new subs / month
Support Diners Club card type	7	1 - Low	Was an issue, but is declining
Support Bitcoin transactions	9	1 - Low	Avoids losing ground to new upstart competitor
PCI Compliance	5	1 - Low	Fines start in January od $1000/m + legal exposure
Fraud Detection	7	1 - Low	Easy to do, but not where fraud is
Expiring Card Reminder	4	1 - Low	Solves fraud issue and reduces fees

Exercise 4

From the facts, work as a group to decide a monthly impact of each feature. I've given you the total duration for each feature based on rough estimates using Figure 1. Your job is to rank order based on cost of delay using the Weighted Shortest Job First technique (divide value by size and do biggest first).

Table 2 - Cost of Delay Calculations (I've given you size, it was estimated using the same way we did in Exercise 1)

Feature Name	Size Est. # stories	* Value Increase $/month	Compute Value ÷ Size (WSJF)	Rationale (explain how you computed value & why this order is/isn't optimal)
Checkout page updates	9			
Support Visa and MC card types	8			
Support AMEX card types	9			
Support Diners Club card type	7			
Support Bitcoin transactions	9			
PCI Compliance	5			
Fraud Detection	7			
Expiring Card Reminder	4			

* Table 1 has pre-computes $/month added value estimates. You can use them or estimate your own based on your analysis of the assumptions.

Exercise 4 Workshop discussion

Q1. Are there any dependencies in this list?

Q2. Do some features HAVE to be done earlier for one reason or another?

Q3. If we ignored size, would you have achieved a similar order? Discuss the ramifications.

Exercise 5

Crate a quantitative cost of delay ranking that accounts for value, size and dependencies between items using the Cost of Delay Calculator spreadsheet.

1. Open the "3 - Cost of Delay for New Credit Card Provider and Fraud Detection.xlsx" spreadsheet.
2. Select the "WSJF Prioritization (intermediate)" worksheet
3. Work as a group on the quantitative Cost of Delay (intermediate) values. Write down the rationale for each dollar value and discuss.

Exercise 5 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 second highest with children. Does this occur often?

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

Figure 5 - Here are my assumptions around value, and the total story count based on estimating size. Do you agree?

Feature or Story Information				Value Inputs			Calculations	
ID	Feature Name	Size	Dependency on Parent Id	Value	Value Unit	Rationale	value / day	WSJF Preferred Order
1	Checkout page updates	9		$ -	Month		$ -	2
2	Support Visa and MC card types	8	1	$ 1,000	Month	3% down to 2% fee. Save 1% on $100,000	$ 33.33	6
3	Support AMEX card types	9	1	$ 5,000	Month	50 new subs @ $100/m	$ 166.67	3
4	Support Diners Club card type	7	1	$ 500	Month	5 new subs @ $100/m	$ 16.67	7
5	Support Bitcoin transactions	9	1	$ 2,000	Month	10 new subs @ $100 + inhibit competitor (x2)	$ 66.67	5
6	PCI Compliance	5		$ 21,000	Month	$1000 fine + 20% rev exposure to breach	$ 700.00	1
7	Fraud Detection	7	2	$ 220	Month	O/S transaction 20%, 1% those get disputed + 10%	$ 7.33	8
8	Expiring Card Reminder	4		$ 2,000	Month	2% x 1000 = 20 x $100 = 2000	$ 66.67	4

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 6

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 5 (or use mine if you had trouble). You should see a result similar to that in
6. Figure 6.

Figure 6 - 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 6 Workshop Discussion

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

Q2. Double the delivery rate. Simulate by changing all input 8 multiplier values to 2.0 (going twice as fast). Does everything make it now?

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

Figure 7 - 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	1.0	
February	2	1.0	
March	3	1.0	
April	4	1.0	
May	5	1.0	
June	6	1.0	
July	7	1.0	
August	8	1.0	
September	9	1.0	
October	10	1.0	
November	11	1.0	
December	12	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

Three Dashboard Types

Strategic dashboards track key performance indicators.
Analytical dashboards process data to identify trends.
Operational (action) dashboards tell you what is happening now.

Q. What type of dashboard is suited for?
1. Observing OKRs
2. Identifying Operational Issues
3. Managing Development Teams

Workshop Exercise

1. Create a list (using post-its) of your current tracked metrics
2. Group these by dashboard type and use

Q1. Are you invested in all three types of dashboards?
Q2. What types of dashboards are you able to create & manage?

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

-

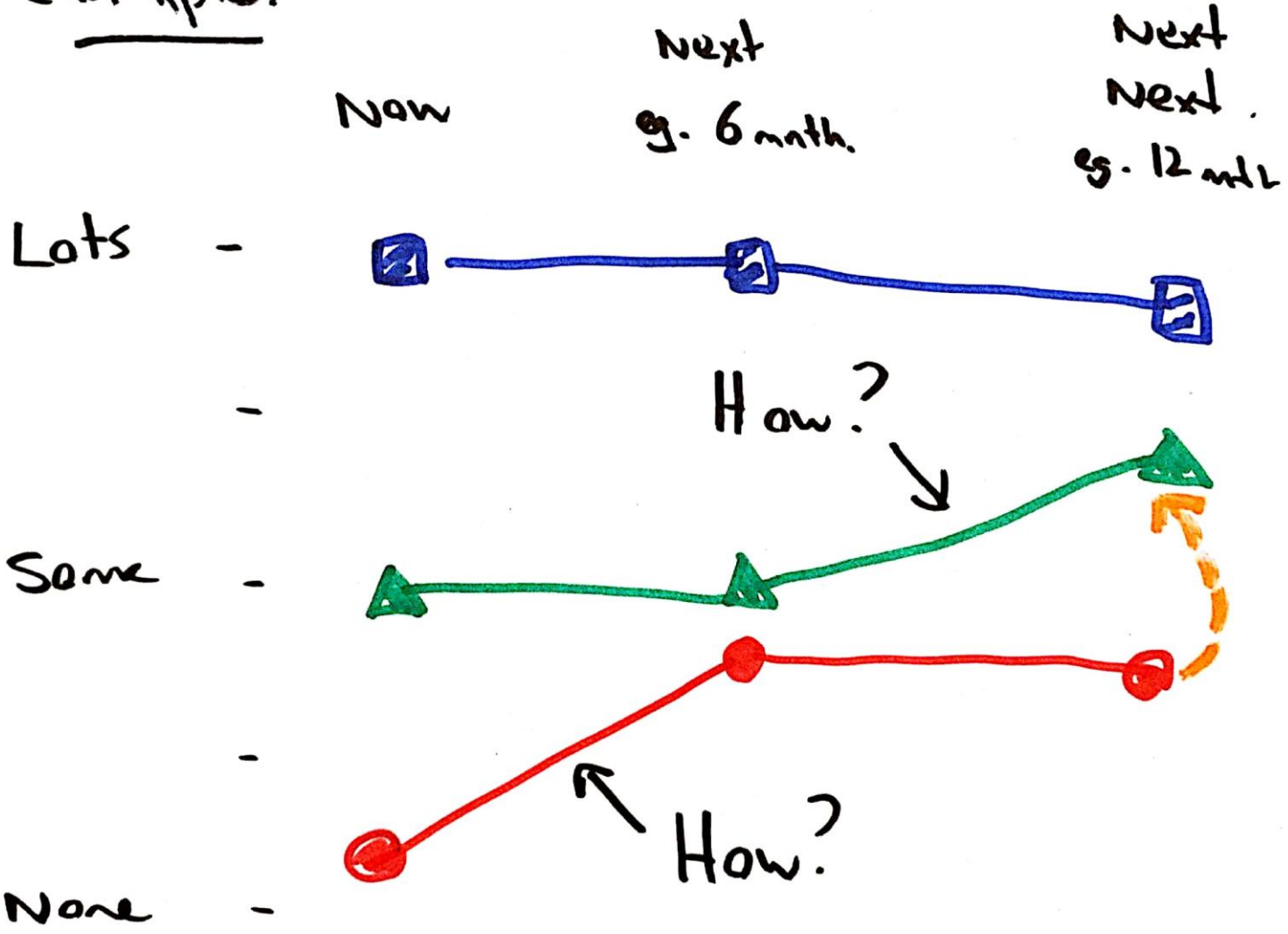
- Impact

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

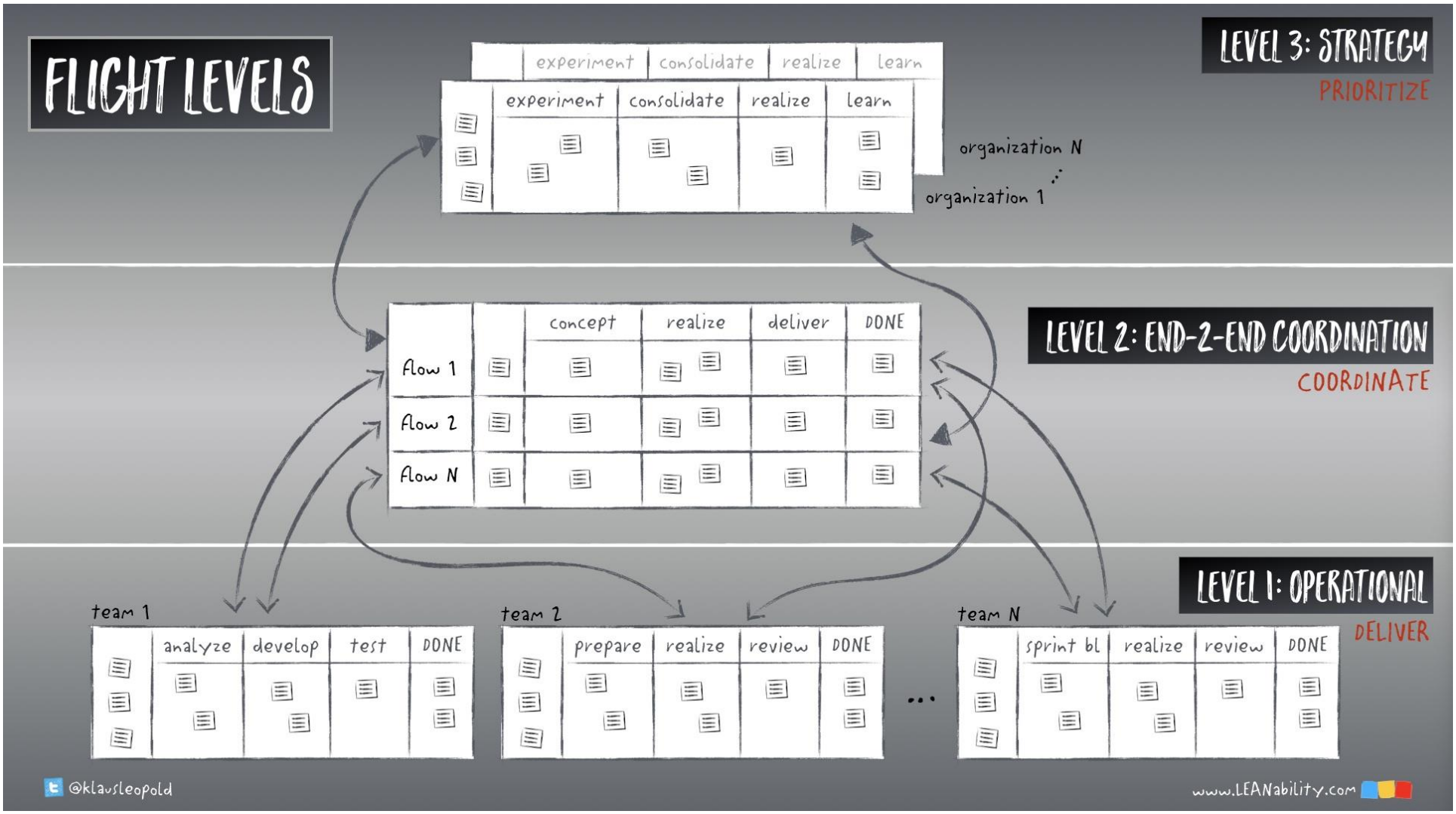
Lots -

-

Some -

-

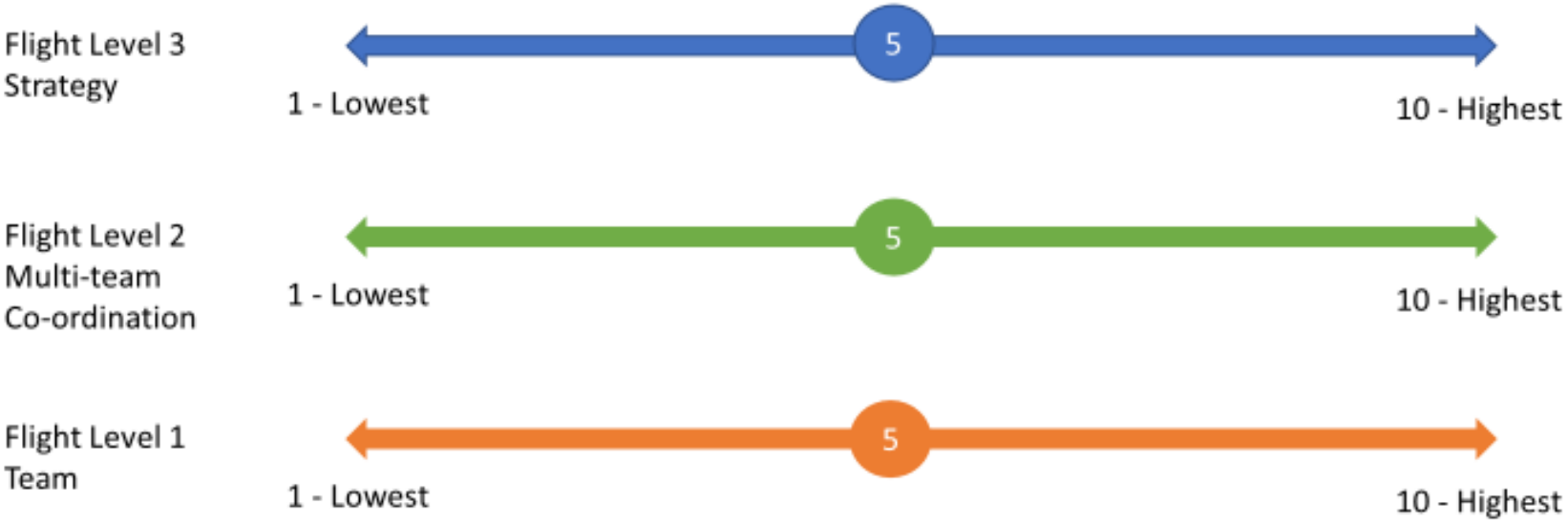
Not at all -


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

15-minute FOTO

From Obstacles to Outcomes in 15 minutes

Principles:

- The client (coachee) knows their own desired outcomes best
- Playfulness and speed of ideation reduces premature judgement and prescription
- Every obstacle has a one-to-many relationship with outcomes
- Inclusive participation means no outcome left behind

What you'll need:

- Openness to learning Clean Language questions (no prior experience needed)
- Groups of 3 to 5 people (4 ideal), representing diversity for the program or project
- A list of obstacles – the input to the exercise
- A 15-minute FOTO cue card per participant
- Notebook/paper to take notes

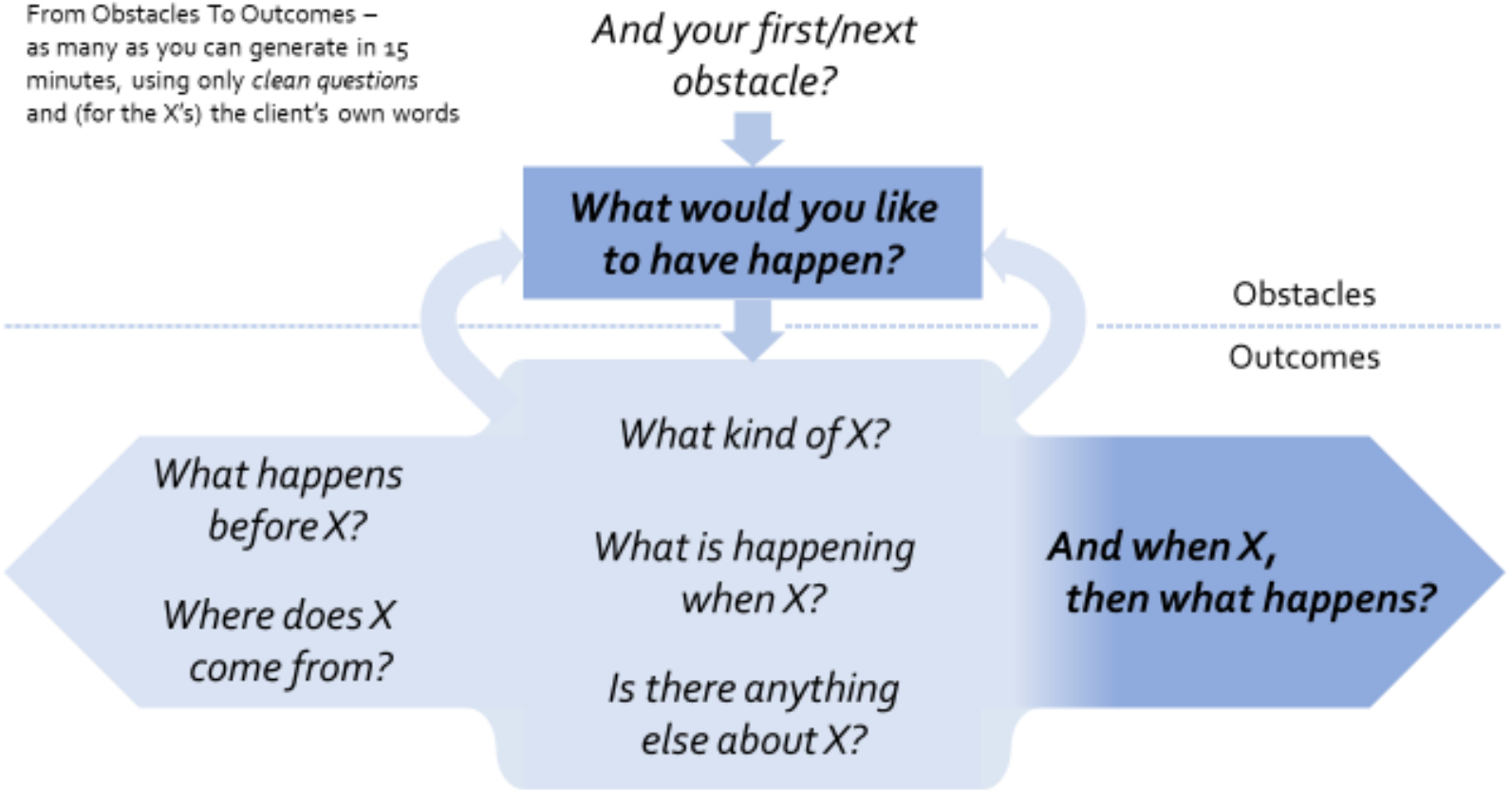
15-minute FOTO – How it works

Coach	Client ('coachee')	Scribe	Observer (Optional)
Guides the conversation using only *clean questions* and the client's own words	Owns obstacles, responds to the coach's questions with short (bullet point answers)	Note taker, writing down anything that sounds like an outcome	Observes the process, monitors the conversation, intervenes if necessary

- The coach asks the client to choose and share an obstacle from the list
- The coach asks Clean Questions from the cue card
- The client gives a brief (bullet point) answer to each question
- The scribe records any answers that identify outcomes
- The dialog continues until it is time to choose another obstacle or to rotate roles

15-minute FOTO

From Obstacles To Outcomes –
as many as you can generate in 15 minutes, using only *clean questions* and (for the X's) the client's own words



agendashift.com/**15-minute-foto**

#agendashift #cleanlanguage

FOTO Debrief…

1. Spend a few minutes on ordering the outcomes from high level (key to business outcome) to lower level (detail and operational)
2. Spend a few minutes rewriting and shortening the outcome language to fit on a post-it note
3. Spend a few minutes ordering the outcomes from how immediate that outcome could be achieved through action (fastest to slowest).

- Present your outcomes to the group (one post-it per outcome) from shortest to longest outcome action.

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"

5 Basic Metric Rules

RESPECT INDIVIDUAL SAFETY

SHOW TRENDS

COMPARE DATA IN CONTEXT

SHOW UNUSUAL CLEARLY

BALANCED – AVOID OVER FOCUSING

Action Viz' Need

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?

Focus on “Important” Insights

Grouping

Sorting

Visual Comparators (reference lines / bands)

Focus Devices (color, dot, size)

Tips for action viz'

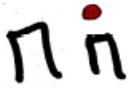
1. Must show values compared to some reference.

   eg. - Trend (prior periods)
   - Background shadow

trend

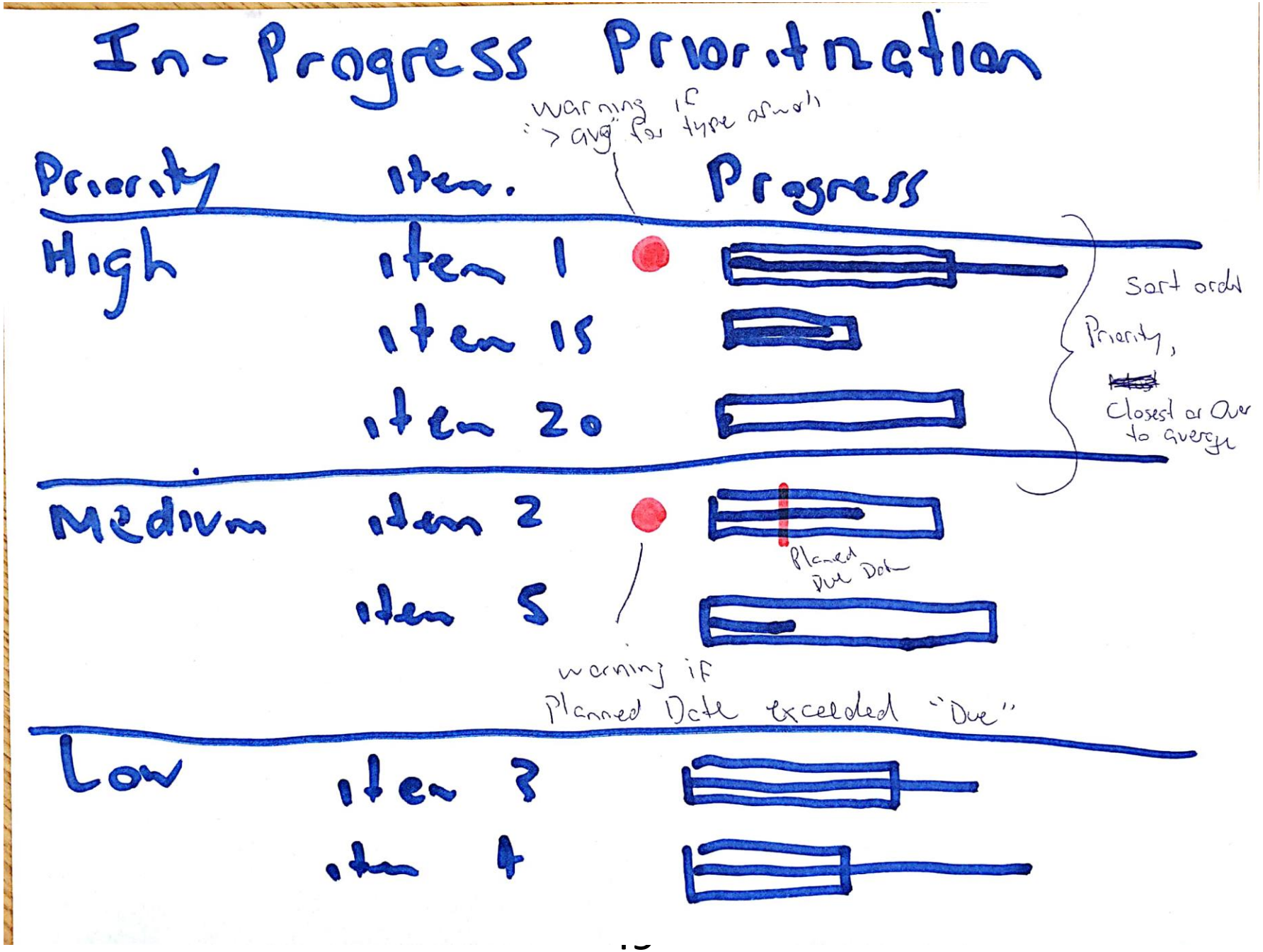
shadow reference

2. Must have a way to highlight "interesting"



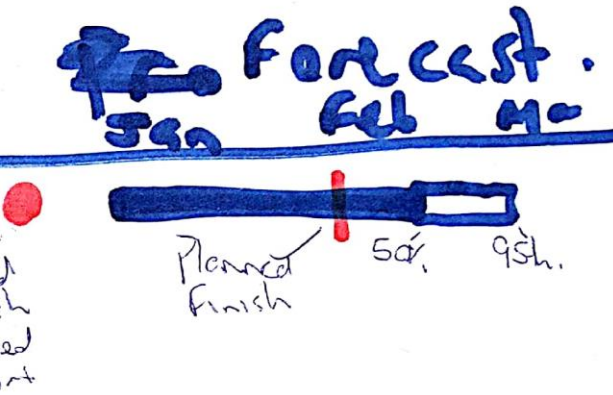
dot

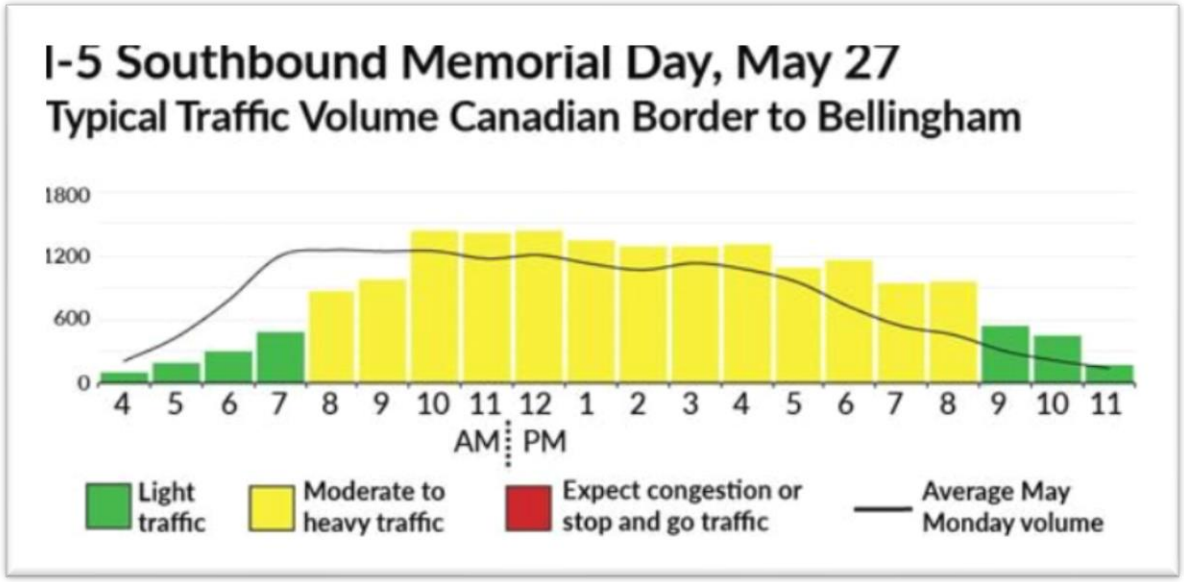
color

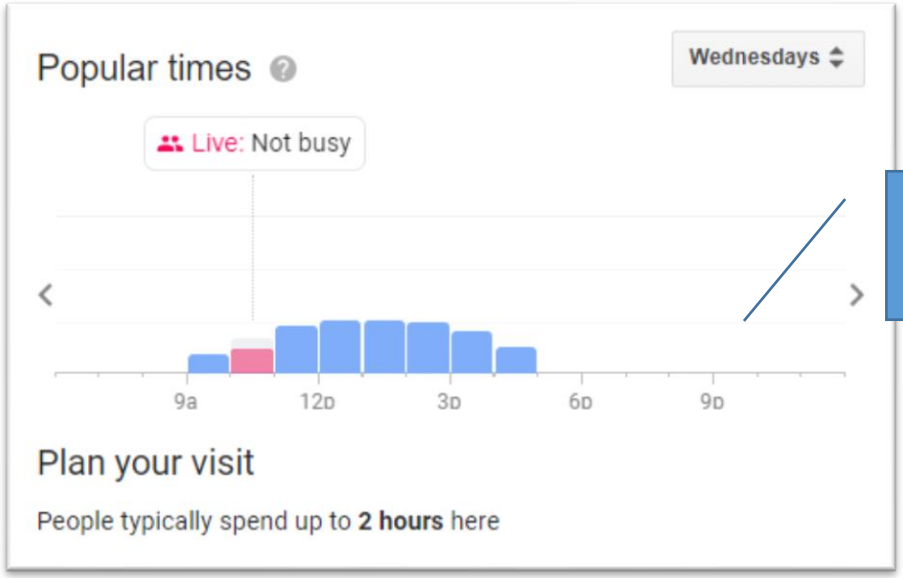

In-Progress Prioritization
warning if > avg for type of work
Priority
Item
Progress
High
item 1
item 15
item 20
Sort order
Priority,
Closest or over to avergt
Medium
item 2
item 5
Planned Due Date
warning if Planned Date exceeded "Due"
Low
item 3
item 4

T.-Do Prioritization

Priority	Item	Type.	Forecast.
High	Item 1	x	
	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

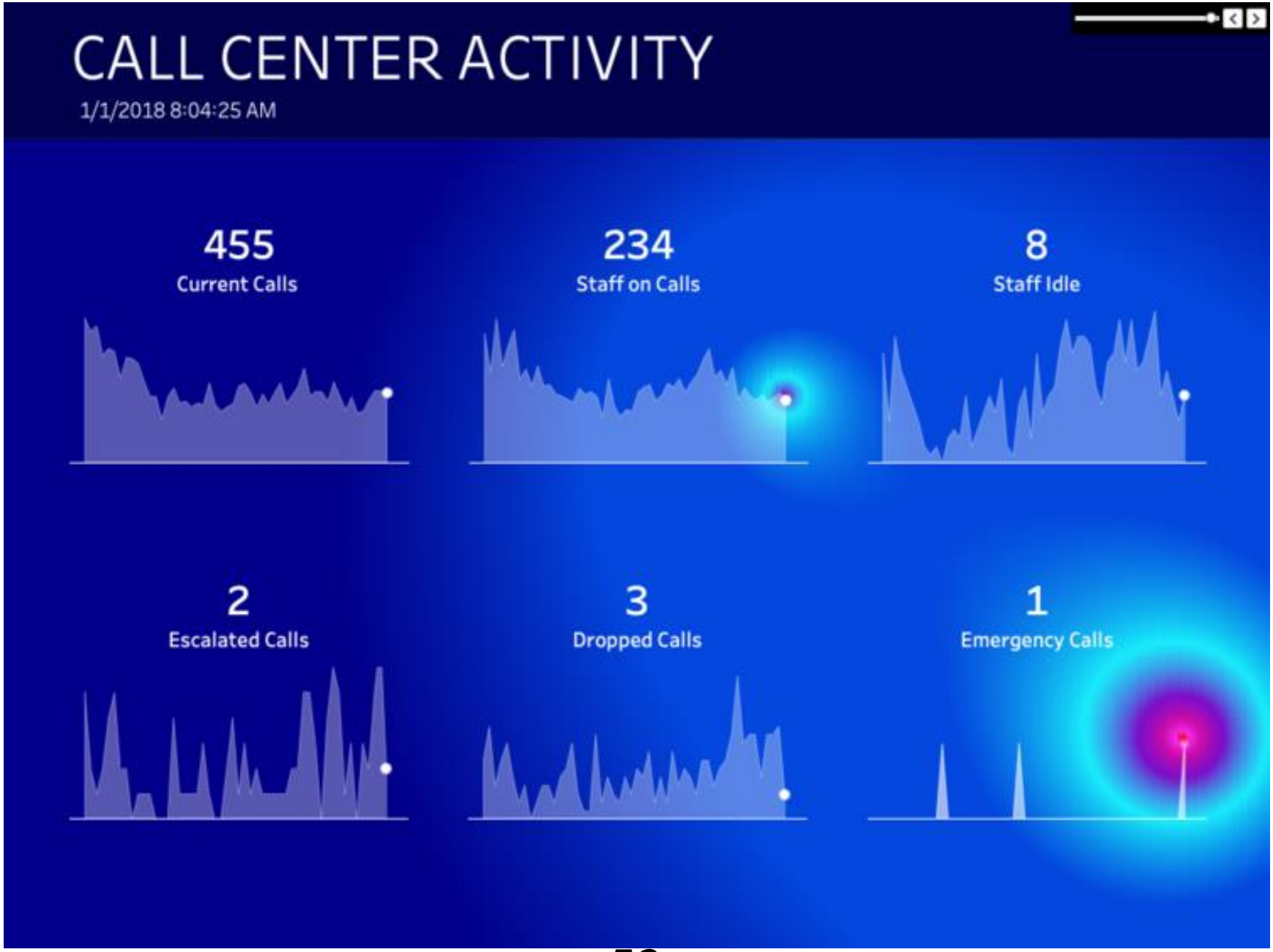
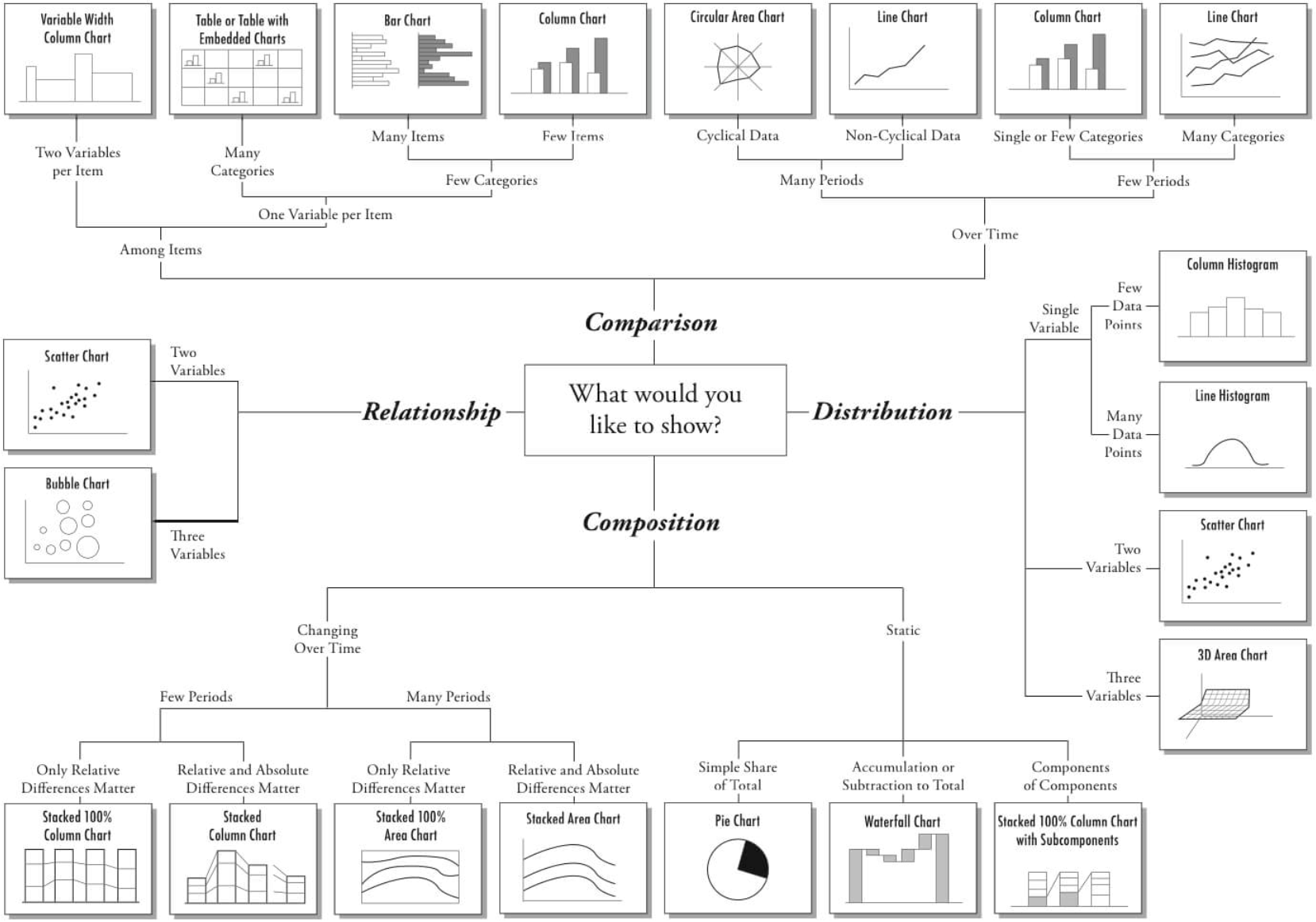
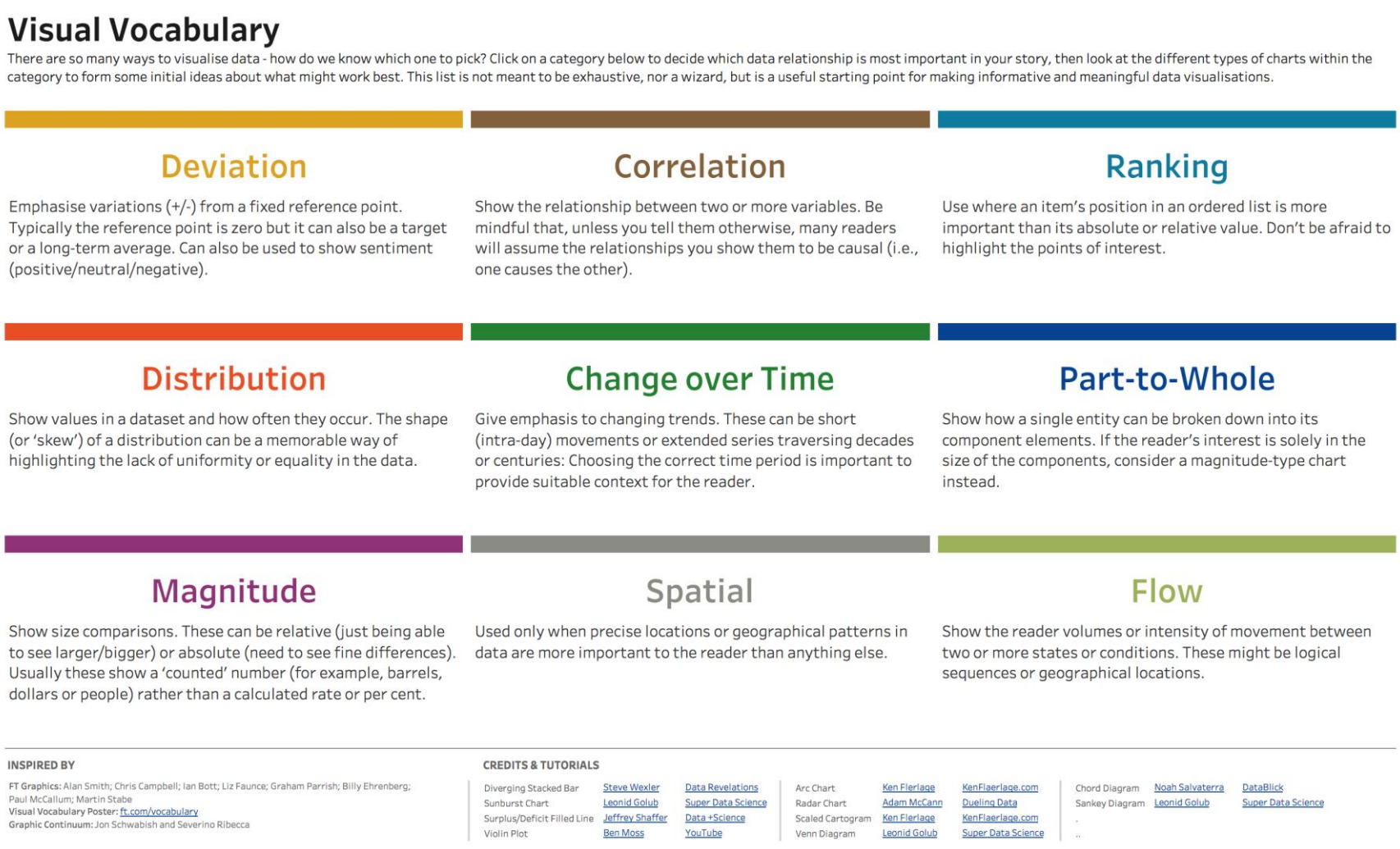

CALL CENTER ACTIVITY
1/1/2018 8:04:25 AM
455
Current Calls
234
Staff on Calls
8
Staff Idle
2
Escalated Calls
3
Dropped Calls
1
Emergency Calls

Chart Suggestions—A Thought-Starter





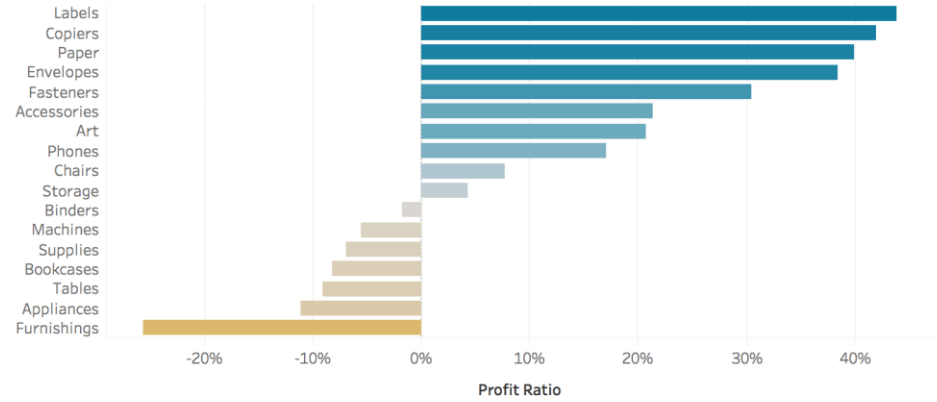
All these images from https://www.vizwiz.com/2018/07/visual-vocabulary.html
https://photos.google.com/share/AF1QipOsge9krcbmO5qj6mzjFbG4WfUL6jYVayXTqBO_BclqUEB94mZrWw-E9-o-Q_pHMQ?key=Tk04RW9oVDZxcC05UEIzRTRYcVJqeXNjb1o4NFpn

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).

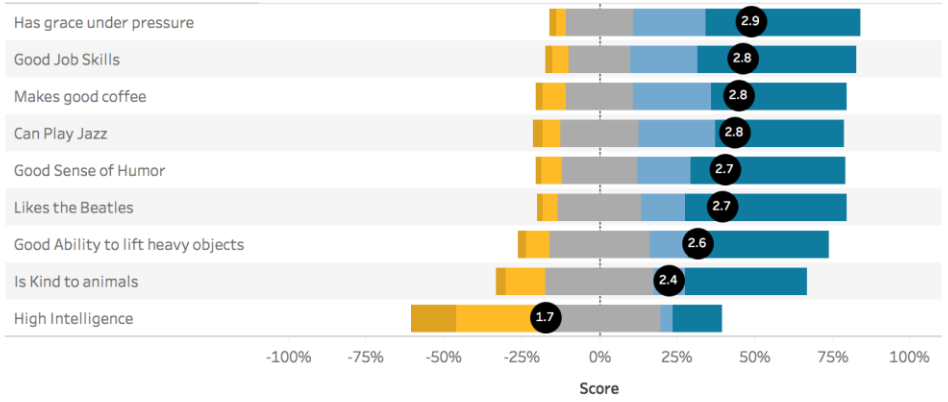
Diverging Bar

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



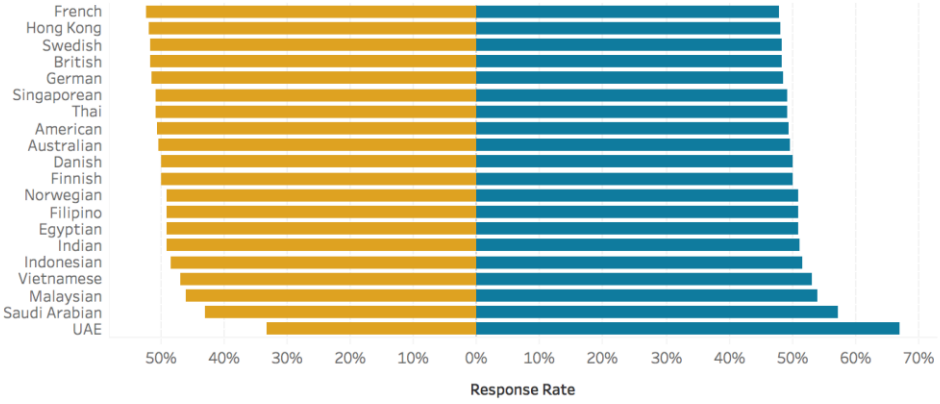
Diverging Stacked Bar

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



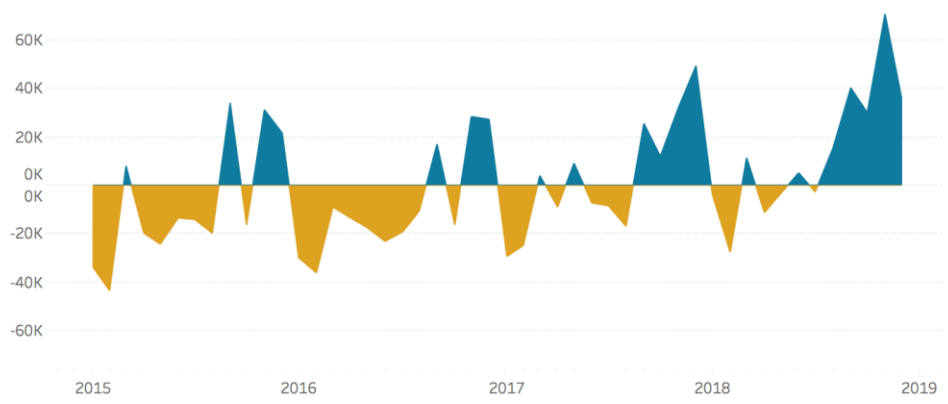
Spine Chart

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.

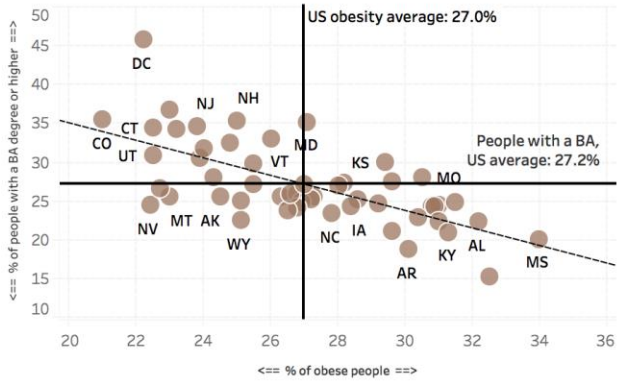


Correlation

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).

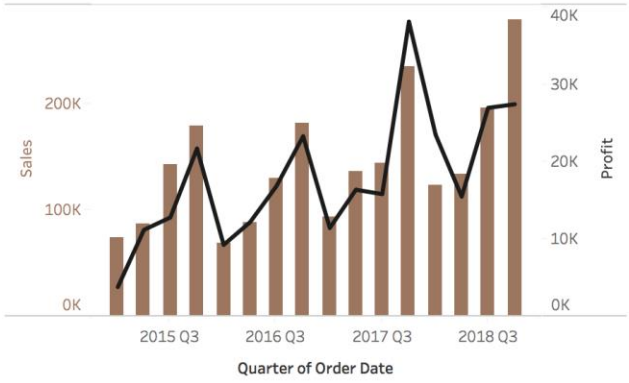
Scatterplot

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



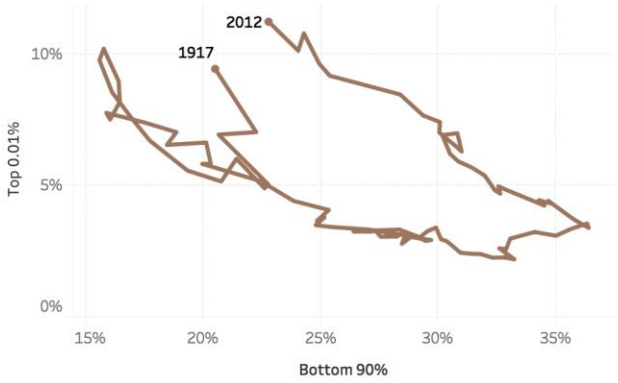
Line + Column

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



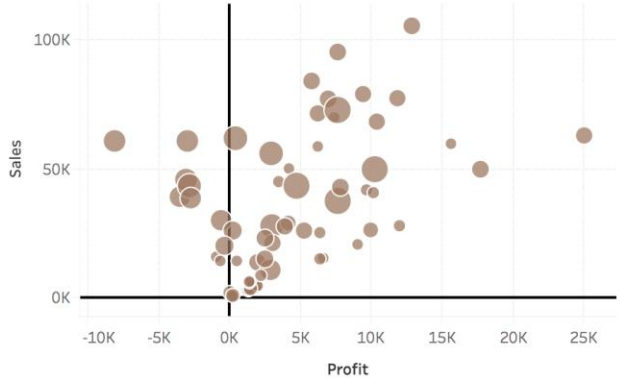
Connected Scatterplot

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.

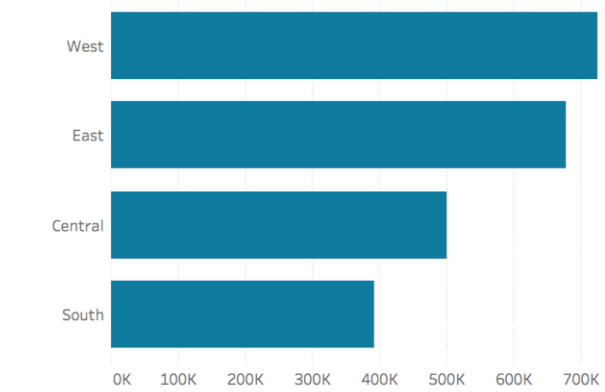
Age Range	I don't have a savings account	$0	Just the minimum balance requirement	Less than $1,000	$1,000-$4,999	$5,000-$9,999	$10,000 or more
Overall	21.0%	28.0%	9.0%	13.0%	10.0%	5.0%	14.0%
18-24	22.4%	21.8%	9.7%	19.1%	14.7%	4.7%	7.5%
25-34	18.0%	26.3%	10.6%	15.2%	12.5%	5.4%	12.1%
35-44	18.9%	31.6%	6.6%	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.

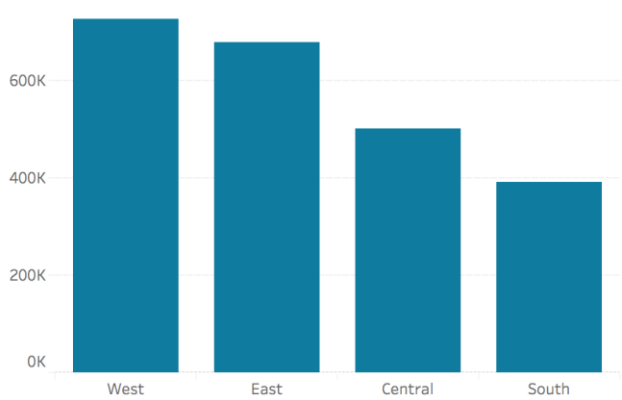
Ordered Bar

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



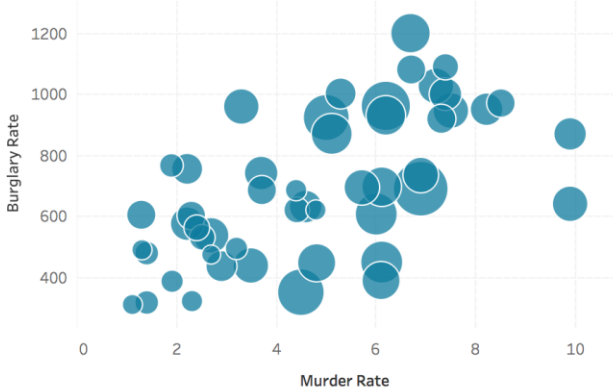
Ordered Column

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



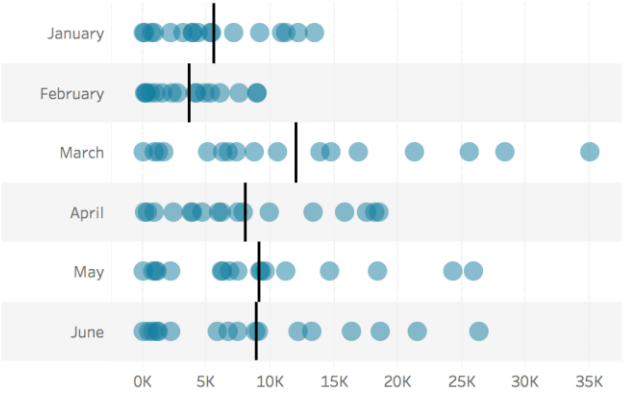
Ordered Proportional Symbol

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



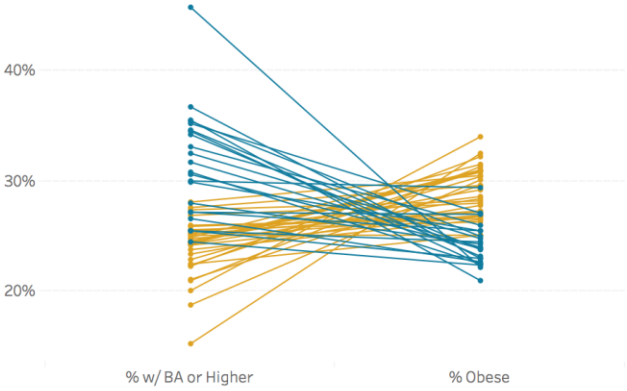
Dot Strip Plot

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



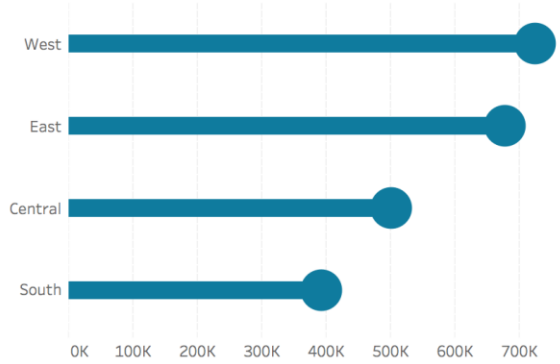
Slope

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.

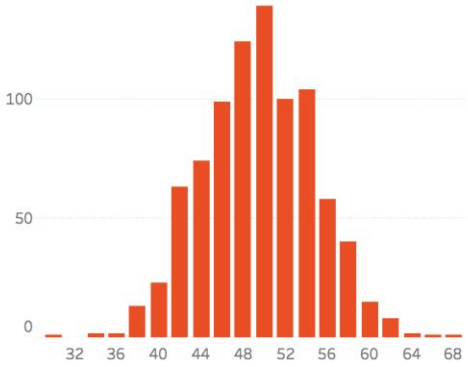


Distribution

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.

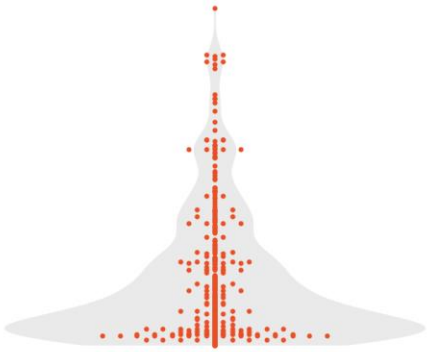


Boxplot

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

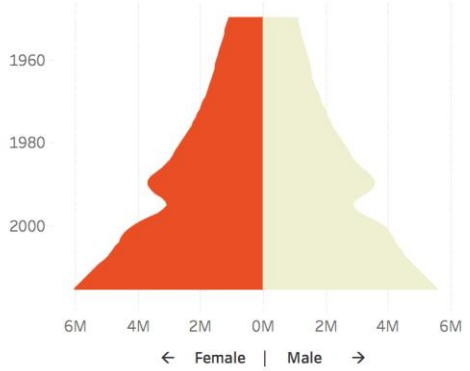
Violin Plot

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



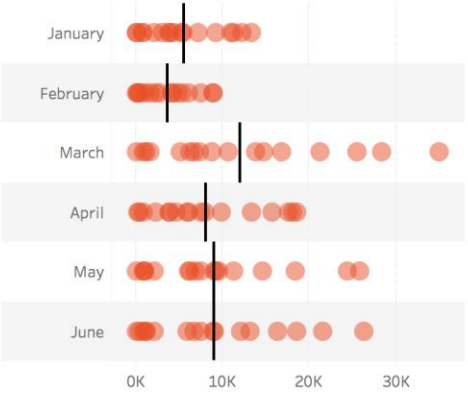
Population Pyramid

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



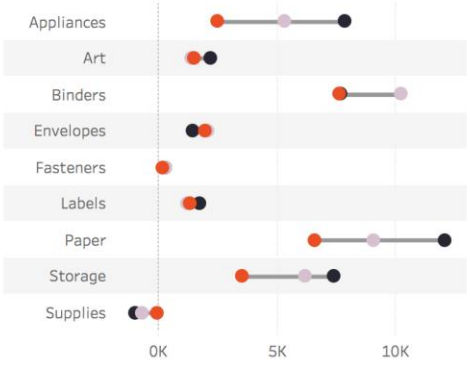
Dot Strip Plot

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



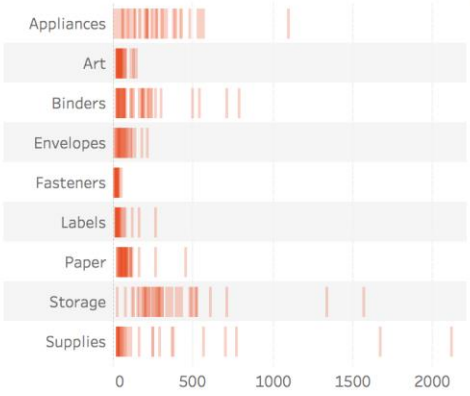
Dot Plot

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



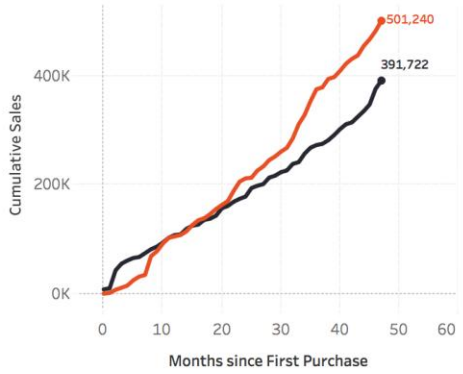
Barcode Plot

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

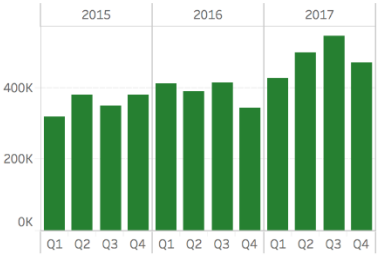
Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Line

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

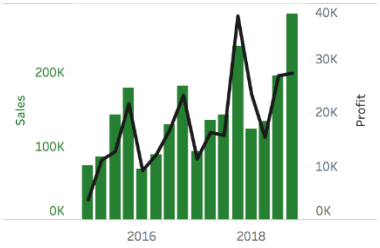
Column

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.

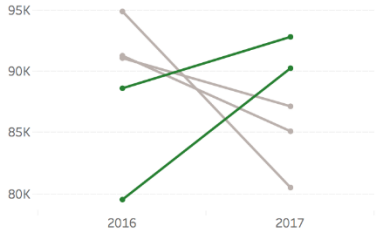


Stock Price

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

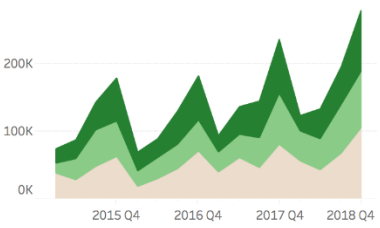
Slope

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

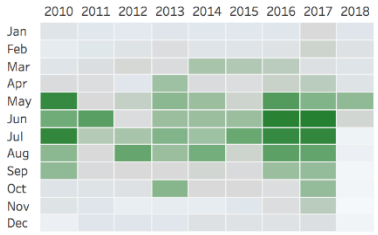
Use to show the uncertainty in future projections - usually this grows the further forward to projection.

Connected Scatterplot

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

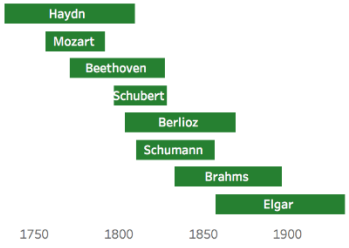
Calendar Heatmap

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



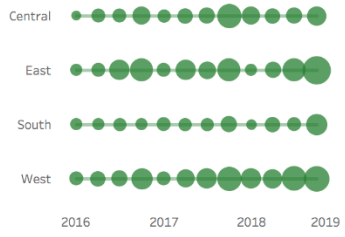
Priestley Timeline

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

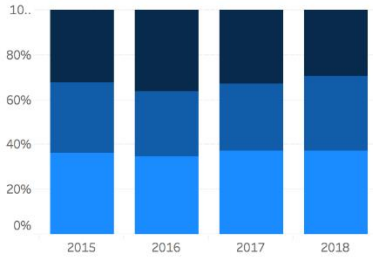
Another alternative to the circle timeline for showing series where there are big variations in the data.

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.

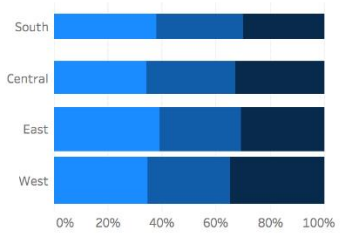
Stacked Column

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



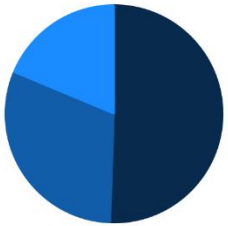
Proportional Stacked Bar

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



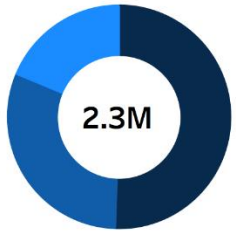
Pie Chart

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

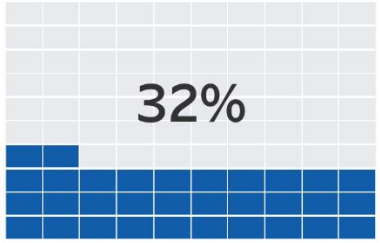
Another way of visualisaing hierarchical part-to-whole relationships. Use sparingly (if at all) for obvious reasons.

Arc

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

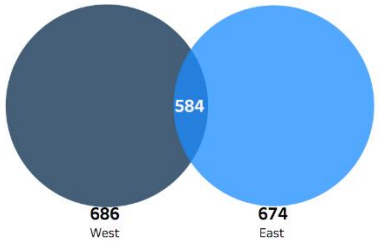
Gridplot

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



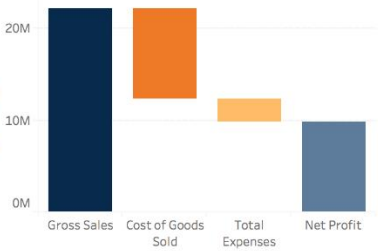
Venn

Generally only used for schematic representation.



Waterfall

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

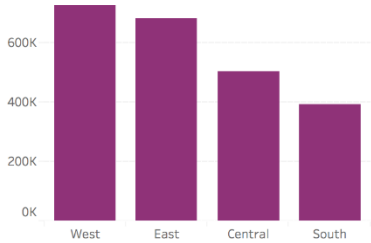


Magnitude

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.

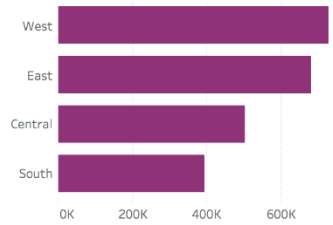
Column

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



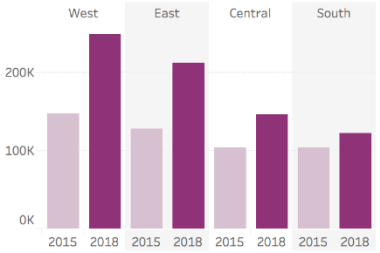
Bar

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.



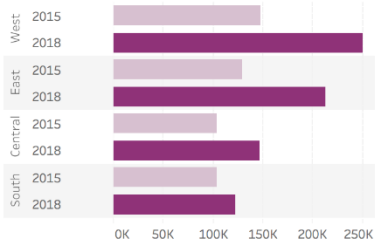
Paired Column

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



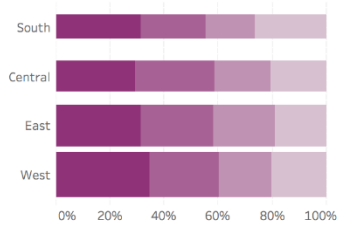
Paired Bar

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



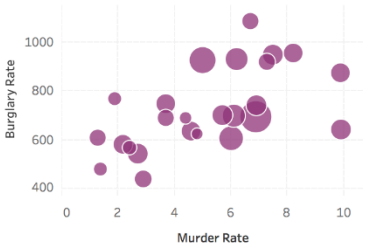
Proportional Stacked Bar

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.

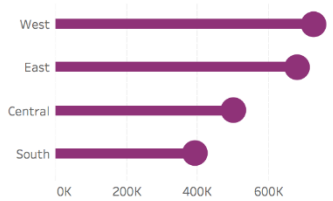


Isotype (pictogram)

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

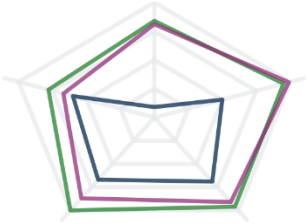
Lollipop Chart

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



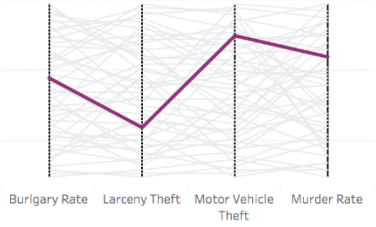
Radar Chart

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

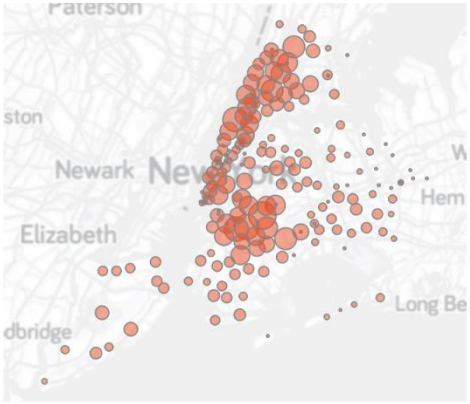
Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

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.

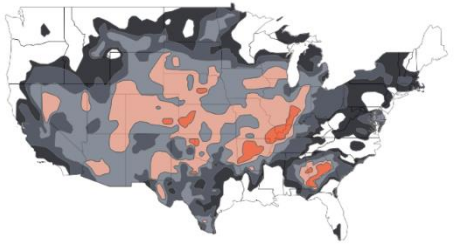


Flow Map

For showing unambiguous movement across a map.

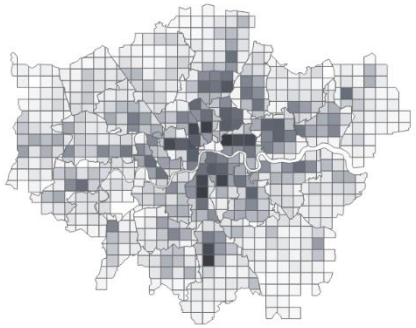
Contour Map

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



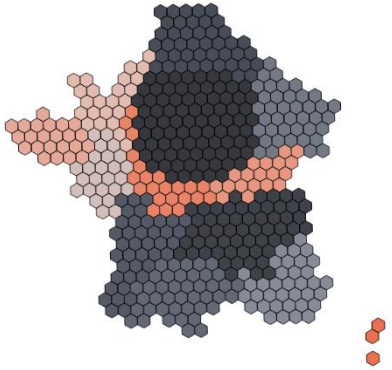
Equalized Cartogram

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



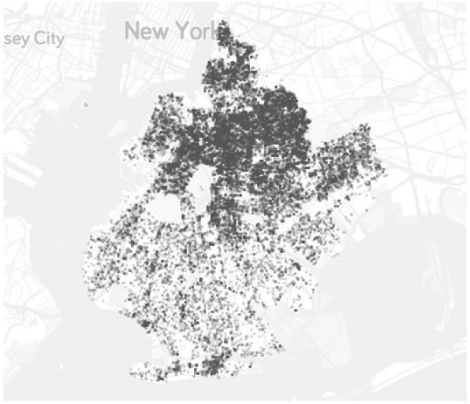
Scaled Cartogram

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



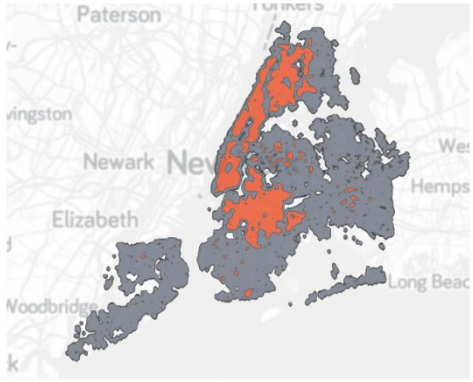
Dot Density

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.

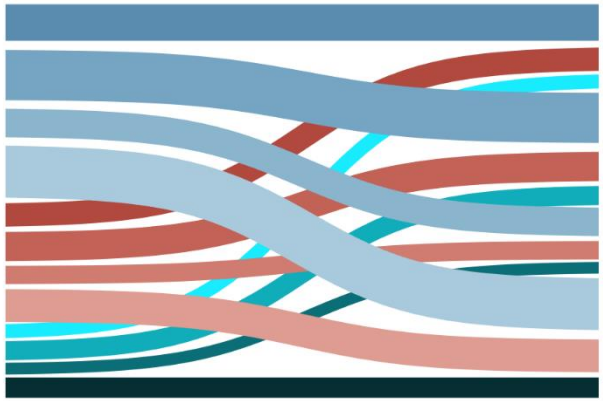


Flow

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

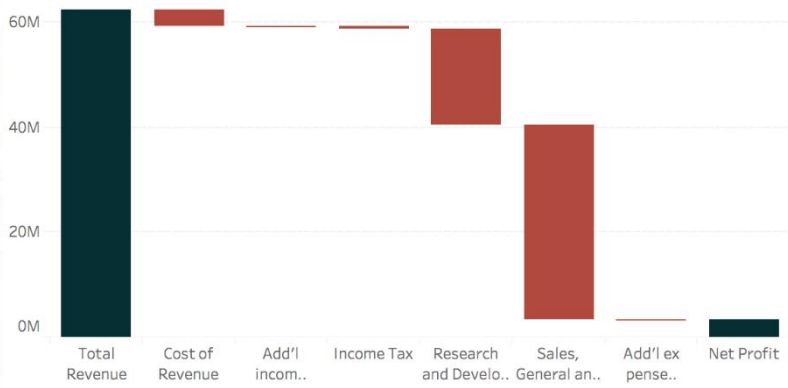
Sankey

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



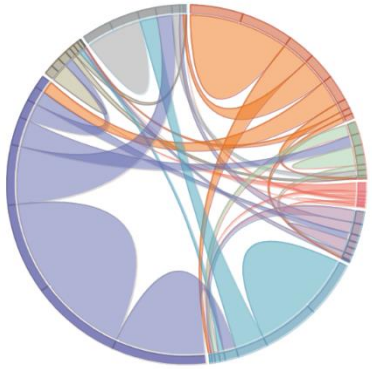
Waterfall

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



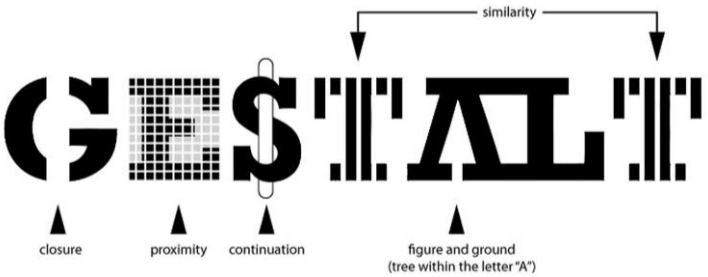
Chord

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

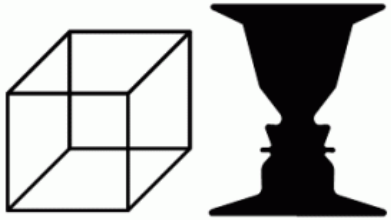


Network

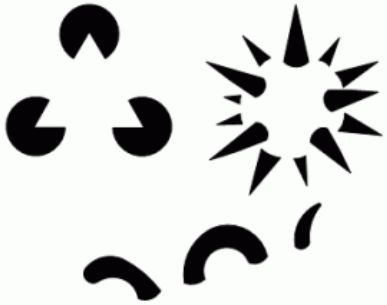
Used for showing the strength and inter-connectedness of relationships of varying types.



Emergence

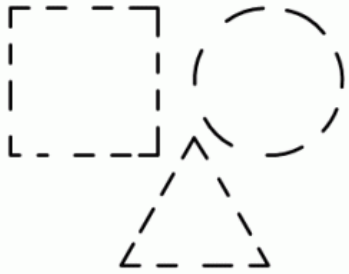


Multistability
Figure/Background selection

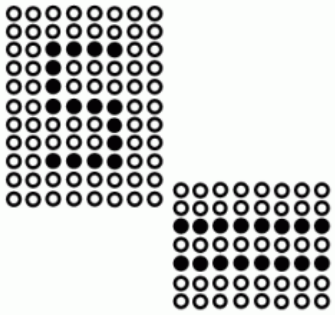


Reification
Illusory contours

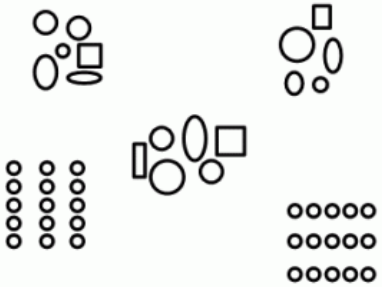
Invarriance



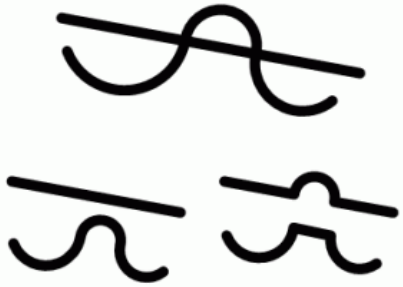
Closure



Similarity

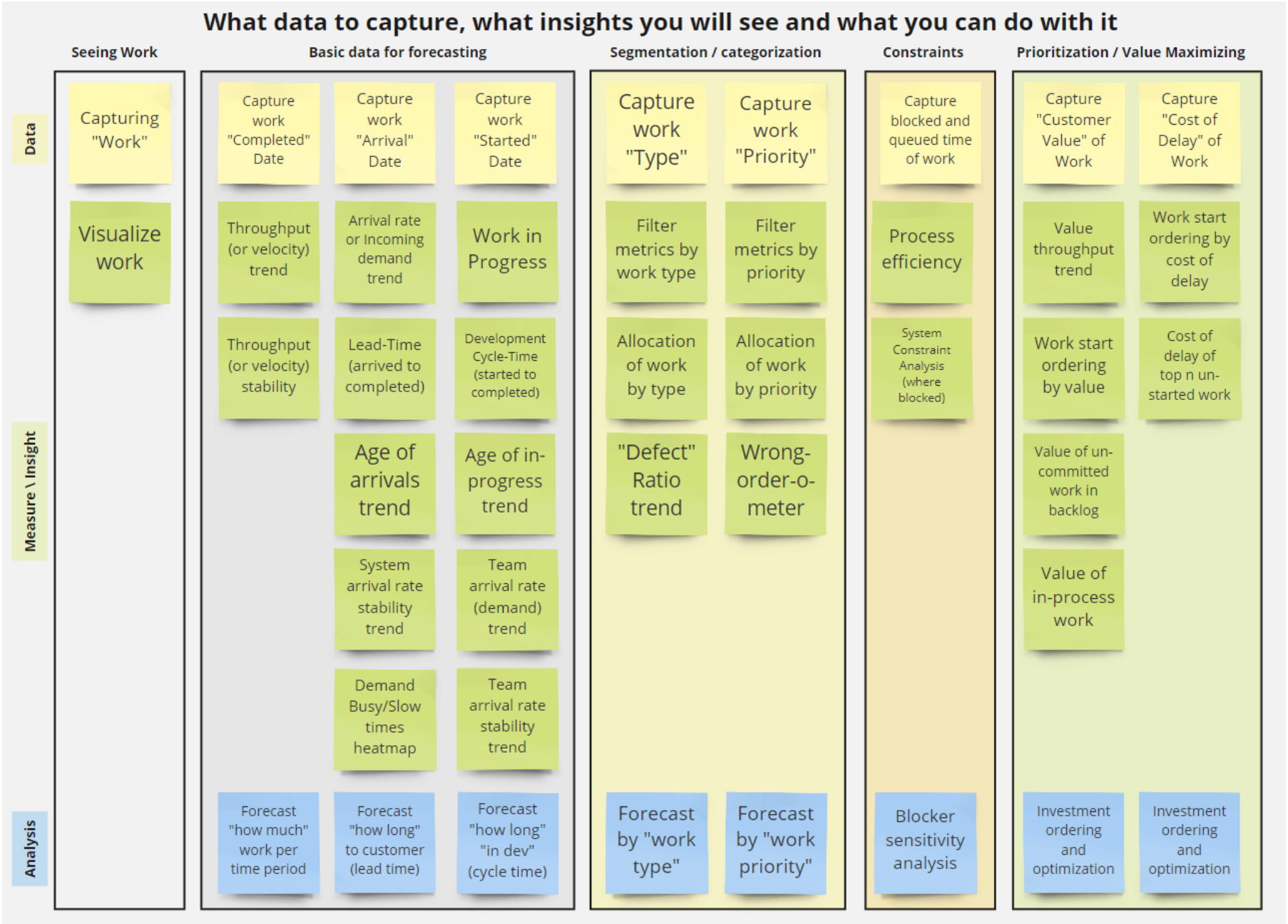


Proximity



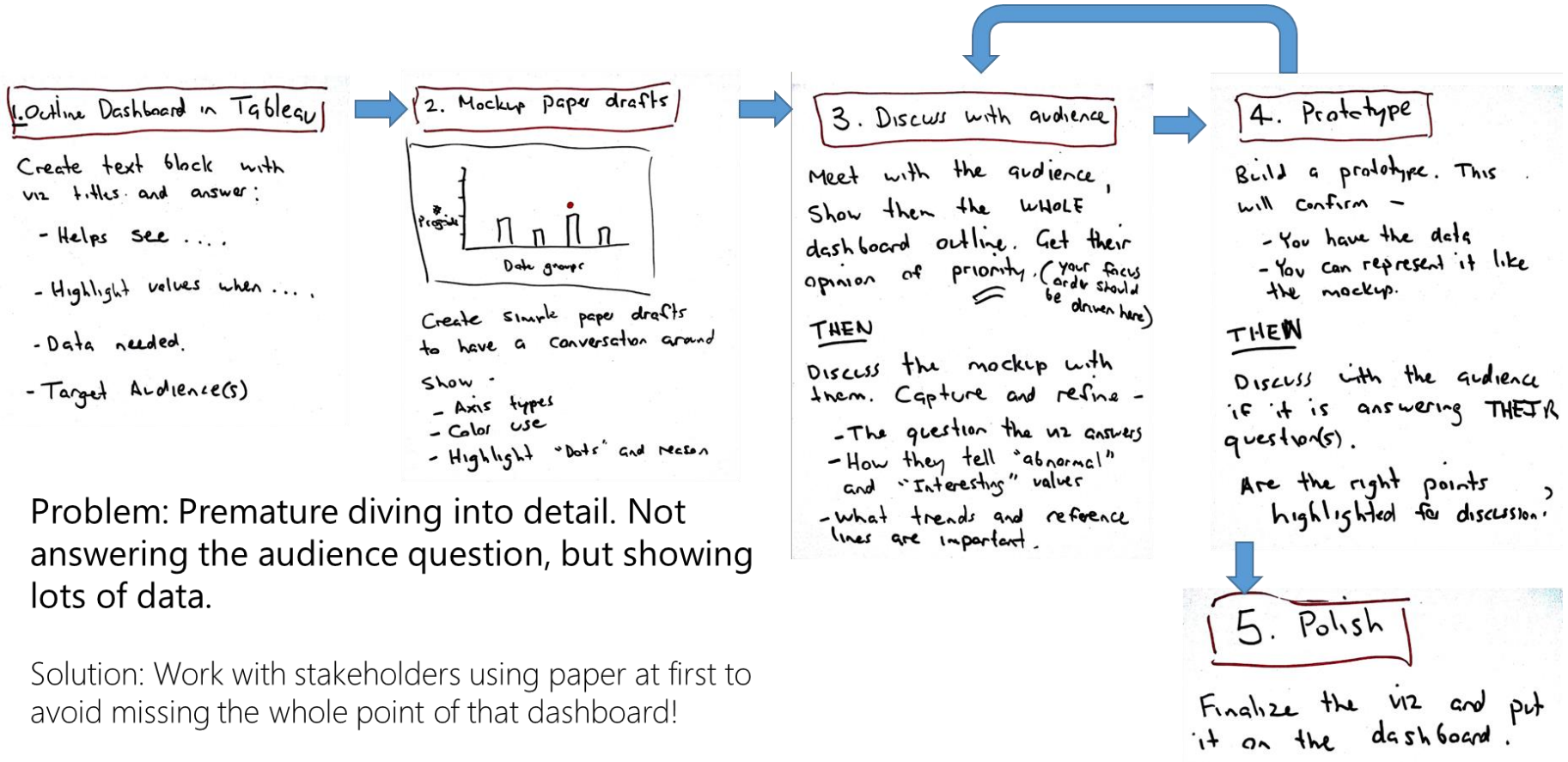
We see this...but not this

Continuity


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

Action-able Dashboards Process

Building charts and dashboards is just like building any product. You need the customer to help direct your progress, as many of your ideas will not be as valuable as you think. In fact, with dashboards and charts, a misstep can be more damaging than helpful. Move at the pace of your feedback. Be alert to the audience needs and feedback. Less is more.



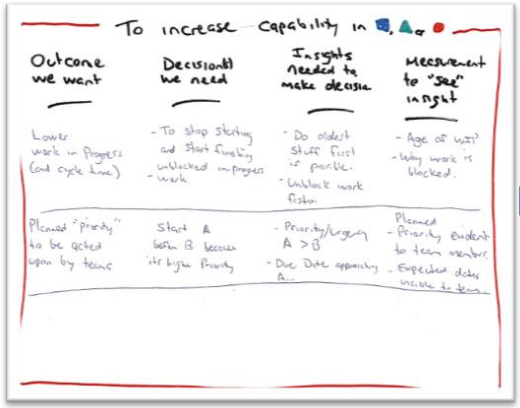
Problem: Premature diving into detail. Not answering the audience question, but showing lots of data.

Solution: Work with stakeholders using paper at first to avoid missing the whole point of that dashboard!

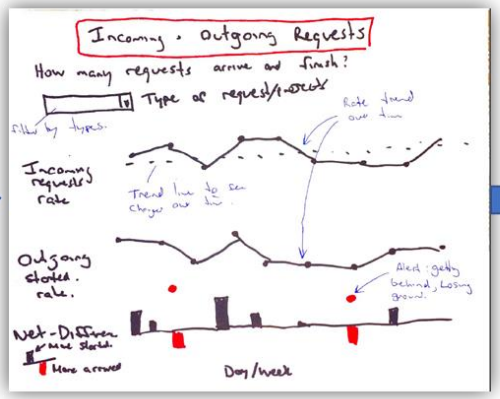
Outcomes – Decisions – Insights – Measures

- 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

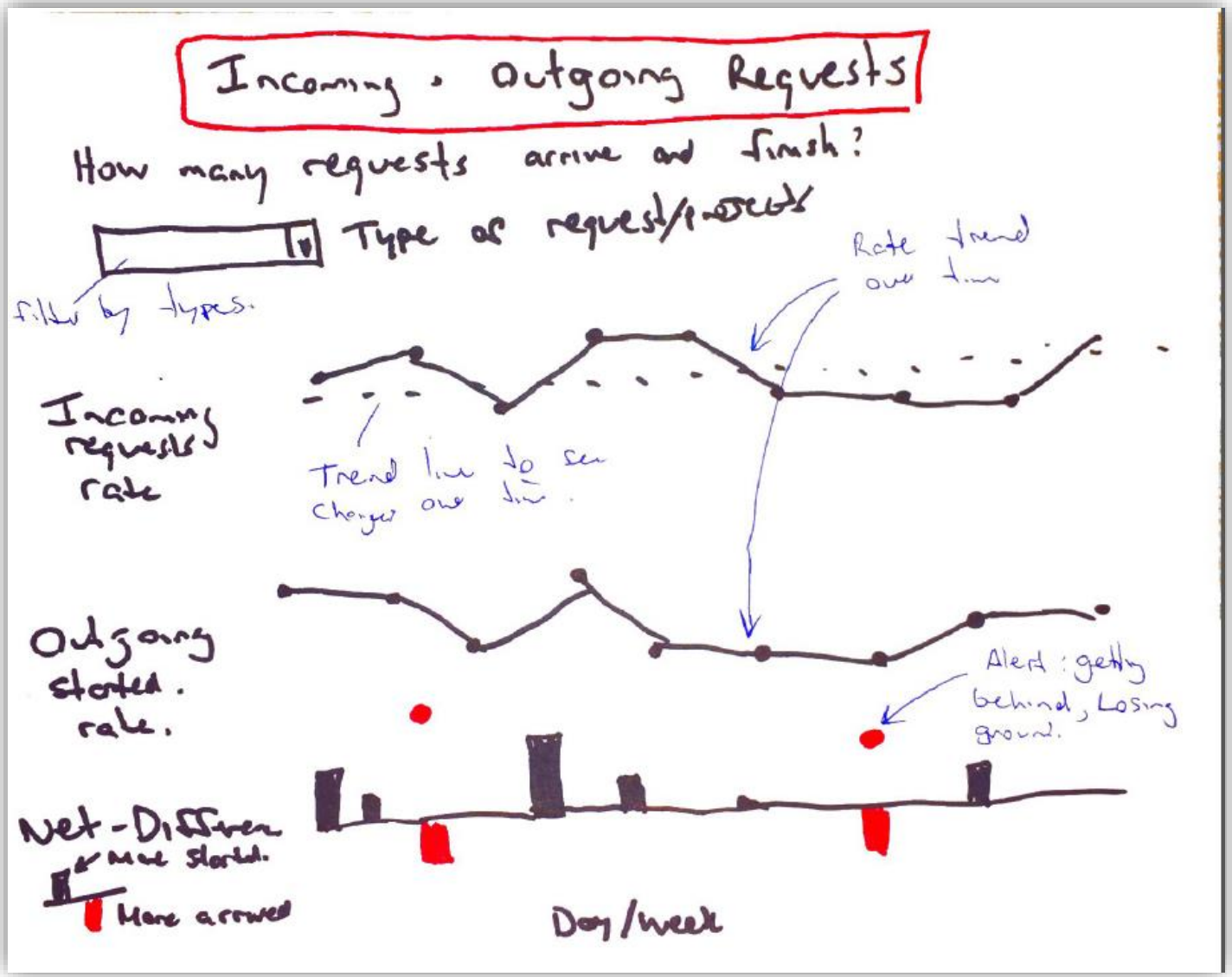
Organization Flow Trend Dashboard.

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	- variability of demand
		Erratic Outgoing throughput	- " " throughput.
	Which teams need help?	- Are teams active?	
		- Are teams increasing in demand?	
		- Teams adverse trend to majority	

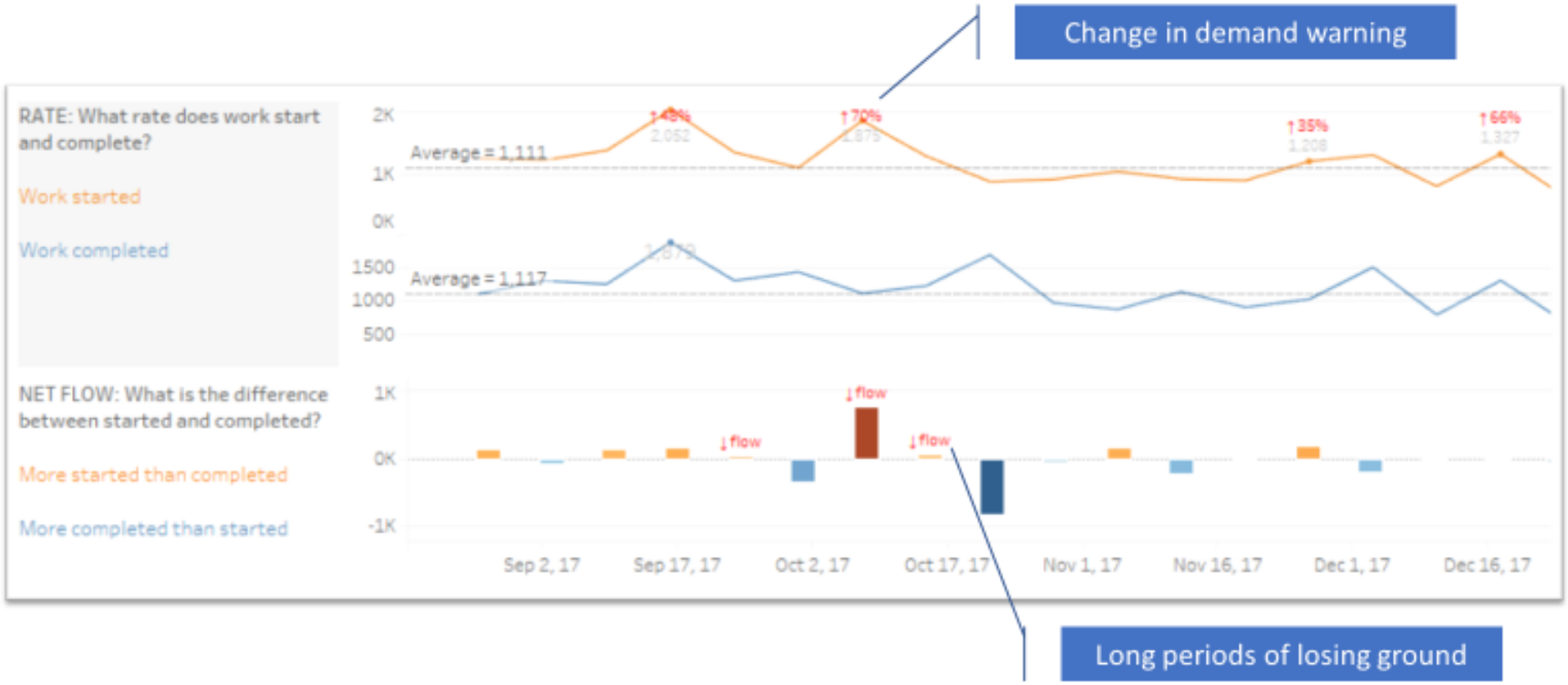
Outcome we want	Decision(s) we need	Insights needed to make decision	Measurement to "see" insight
Lower work in Progress (and cycle time)	- To stop starting and start finishing unblocked in-progress work.	- Do oldest stuff first if possible. - Unblock work faster	- Age of WIP - Why work is blocked.
Planned "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

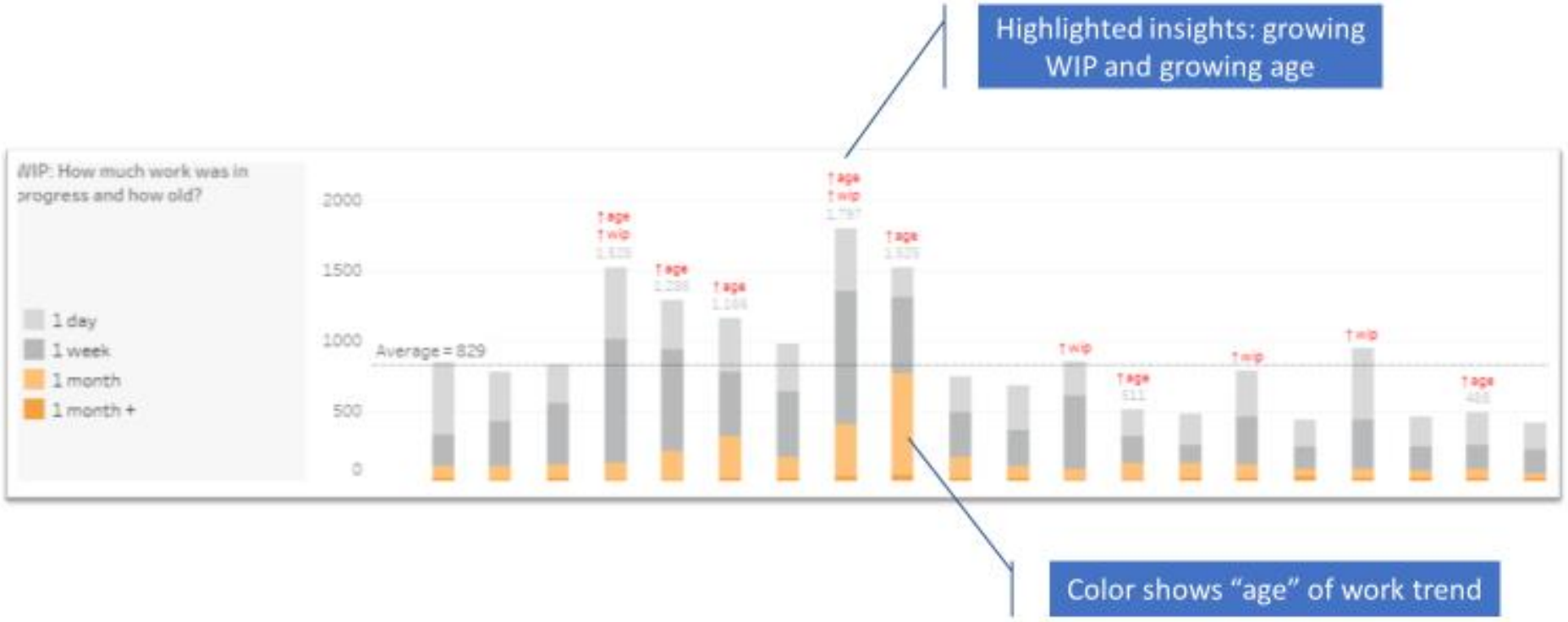
Incoming and Outgoing Demand Measures

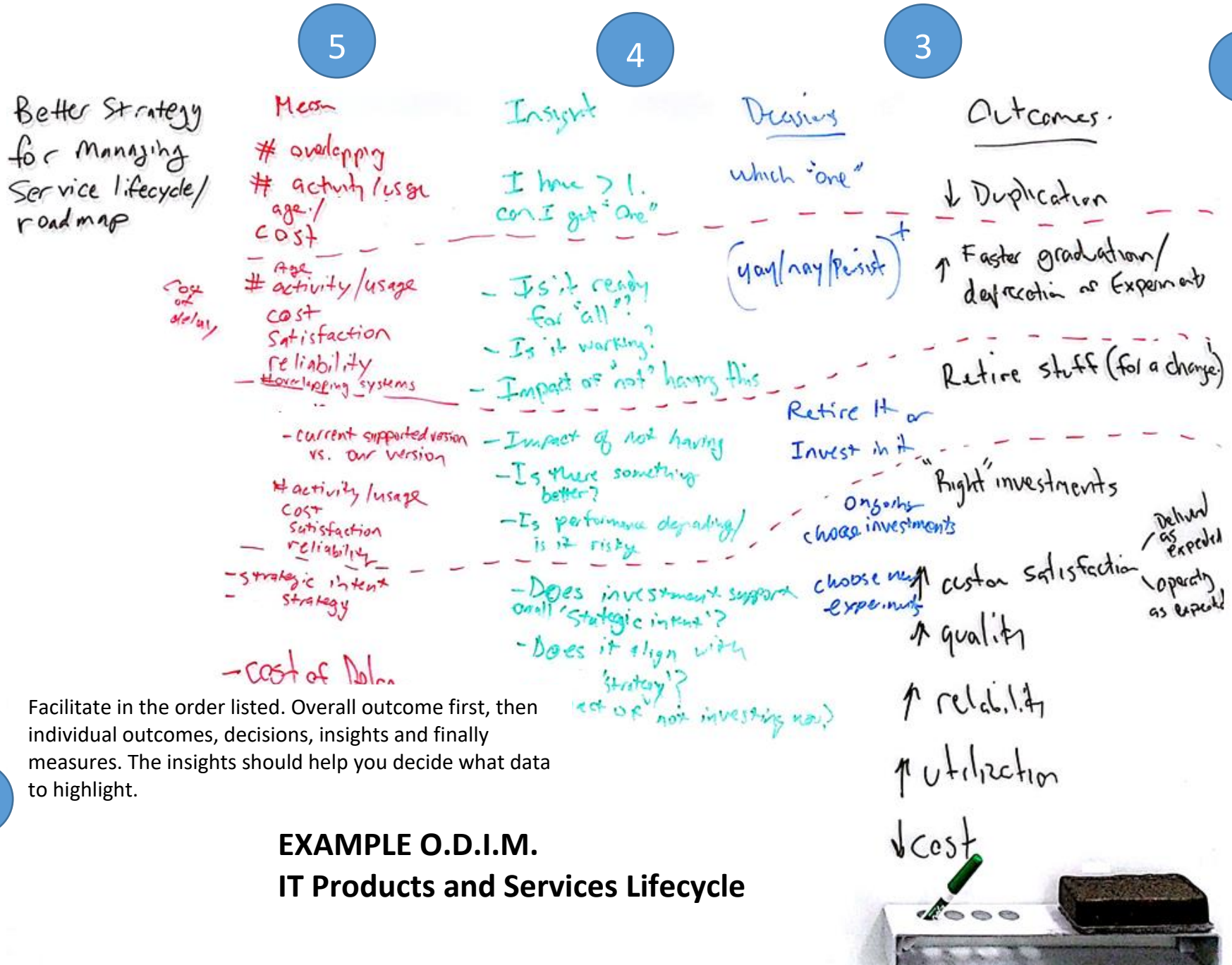


Outcome: To visualize and improve flow

- 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




5
4
3
2
1
n
Better Strategy for Managing Service lifecycle/ roadmap
Measure
overlapping
activity/usage/ age/
cost
Insight
I have > 1. can I get "one"
Decisions
which "one"
Outcomes
↓ Duplication
(yay/nay/Persist)+
↑ Faster graduation/ deprecation of Experiments
Retire stuff (for a change)
Retire it or Invest in it
"Right" investments
Ongoing choose investments
choose new experiments
↑ cust satisfaction
Delivered as expected
Operating as expected
↑ quality
↑ reliability
↑ utilization
↓ cost
Facilitate in the order listed. Overall outcome first, then individual outcomes, decisions, insights and finally measures. The insights should help you decide what data to highlight.
EXAMPLE O.D.I.M.
IT Products and Services Lifecycle

Workshop Exercise

Part 1: O.D.I.M.

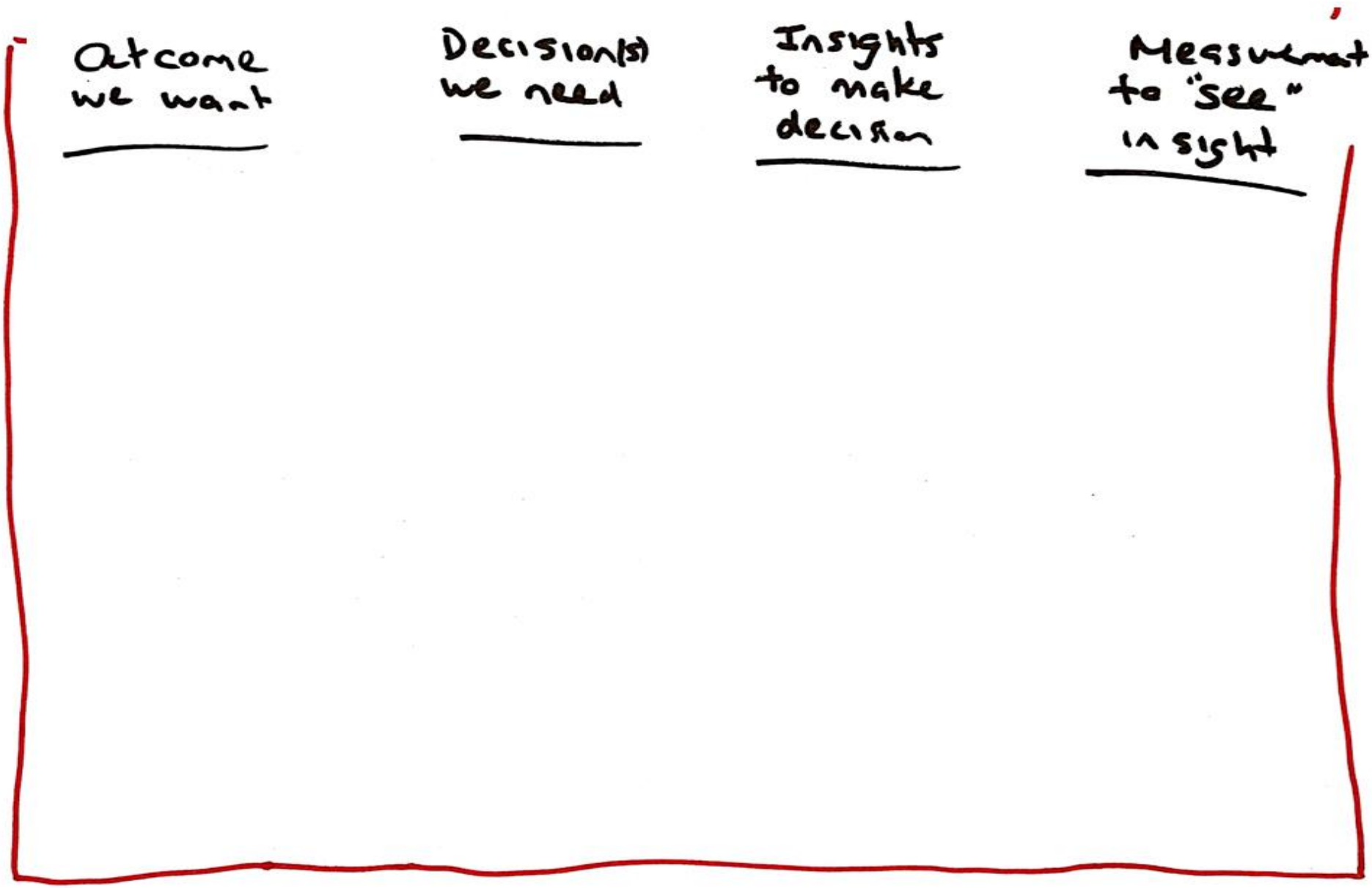
1. Brainstorm a set of MAJOR outcomes. A major outcome is one that can have 3 to 5 SUPPORTING outcomes.
2. Work in groups to define as many SUPPORTING outcomes (write on the ODIM worksheet column under Outcomes, spread equally down the page) and present to the group. Dot vote for the top 3.
3. Work in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

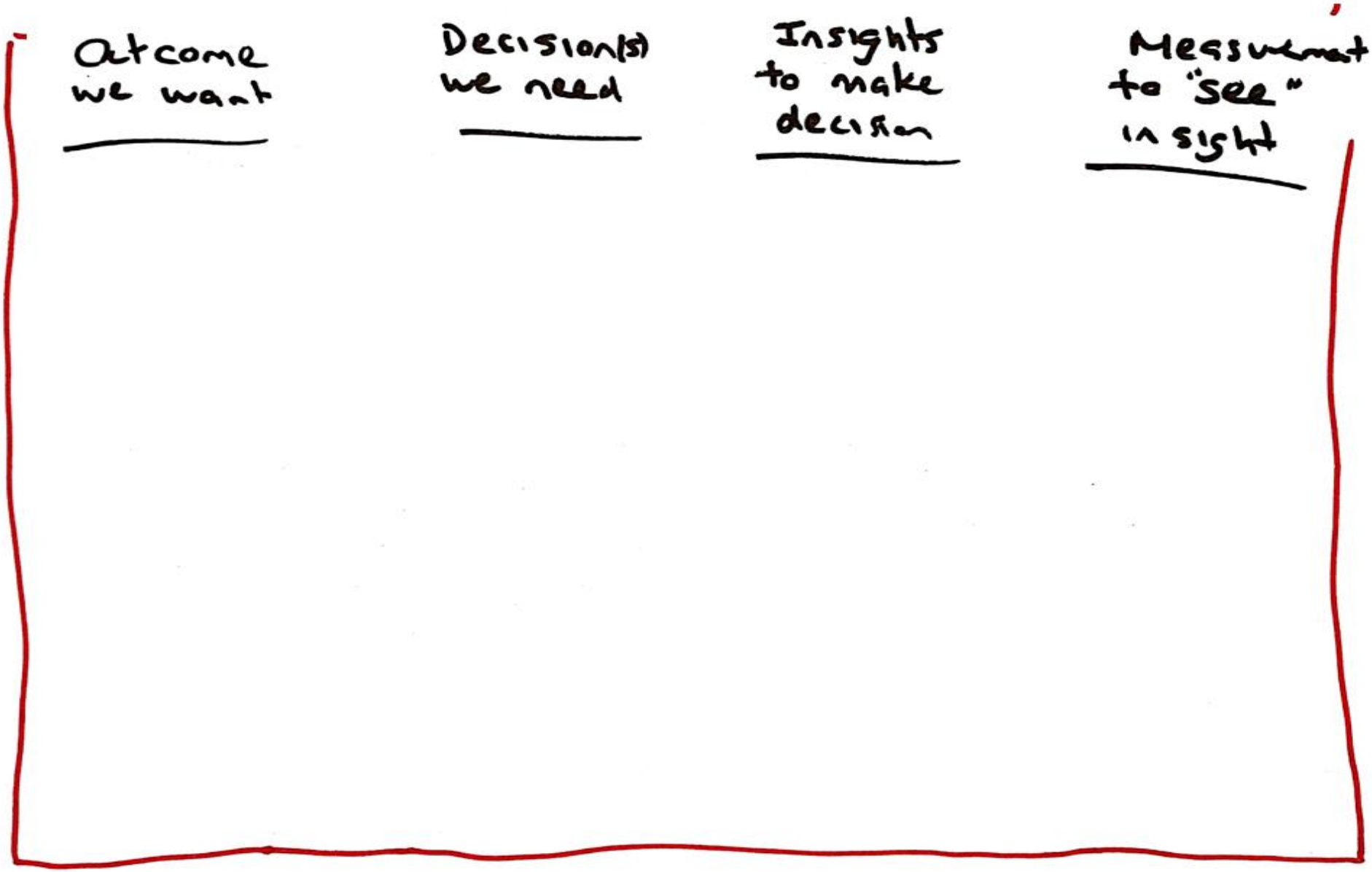
1. Work in groups to sketch ideas that show the measures in a way that highlight the most powerful insight. Do as many as you can in 15 minutes.
2. Gallery walk and capture ideas for improvement.
3. Select the “right” ones and do a second round of final sketch.
4. Gallery walk and present the “why” this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

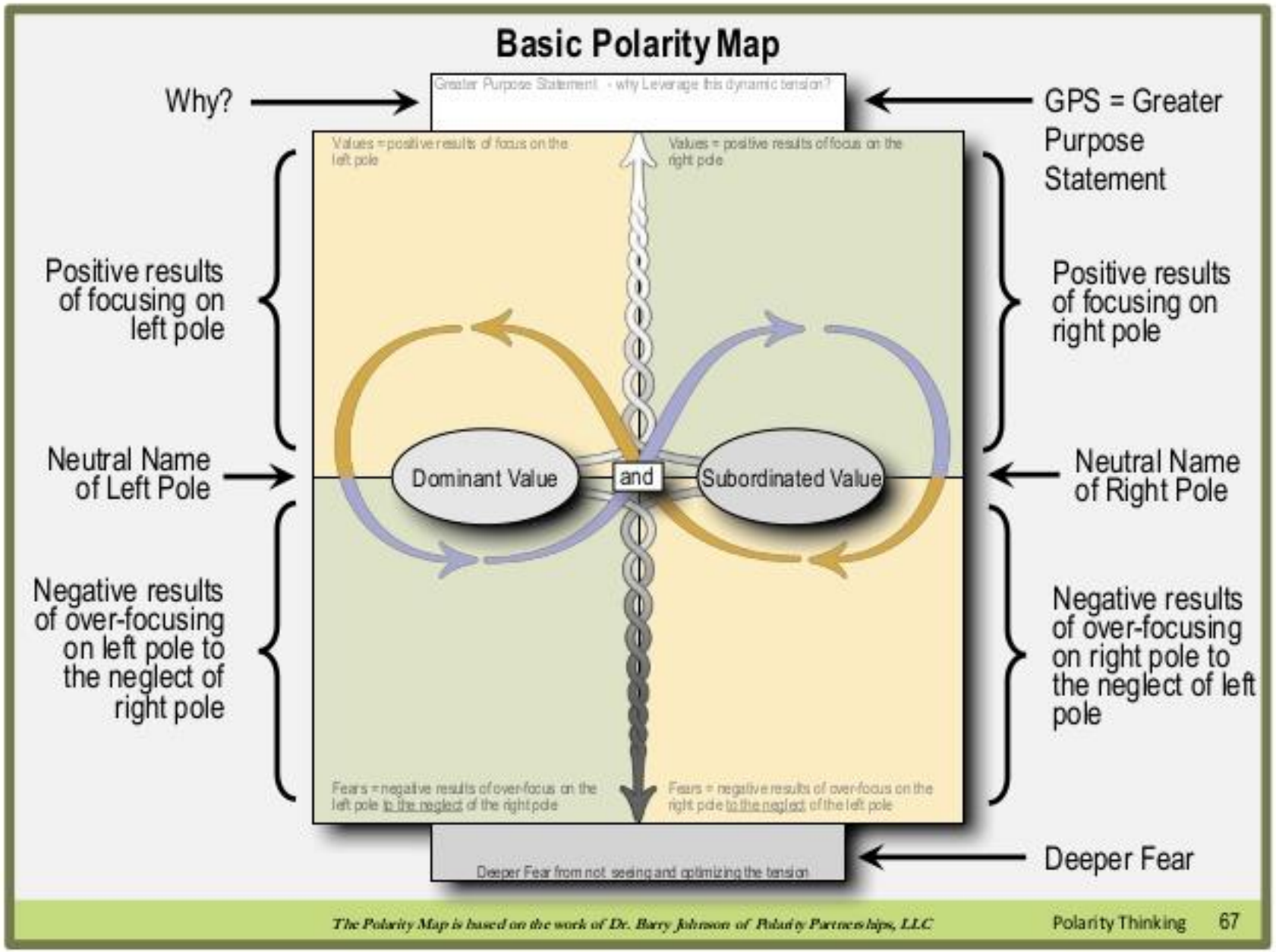
Sketches for Insight:

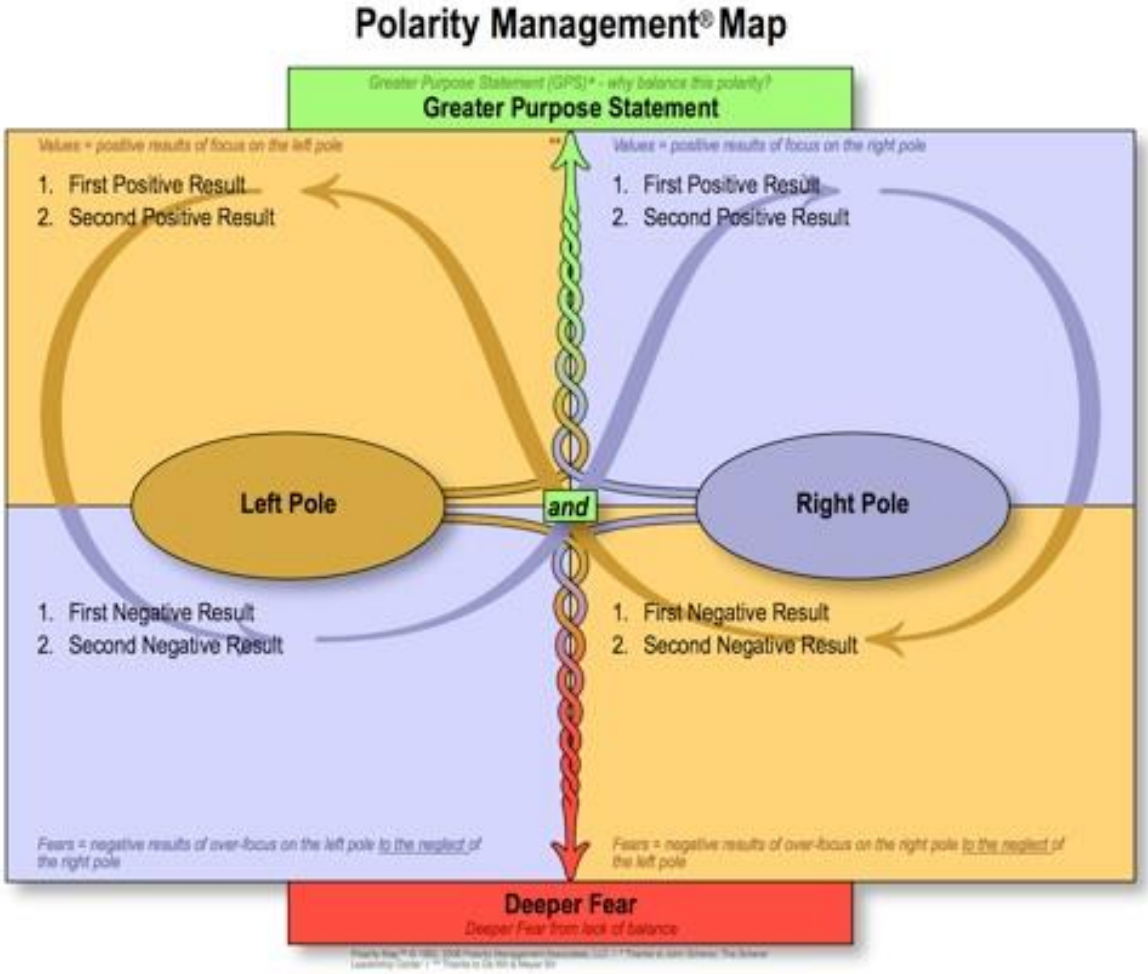
Tips: Trends are always better than a single number. Highlight the insights. How will you help people see these insights first?


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

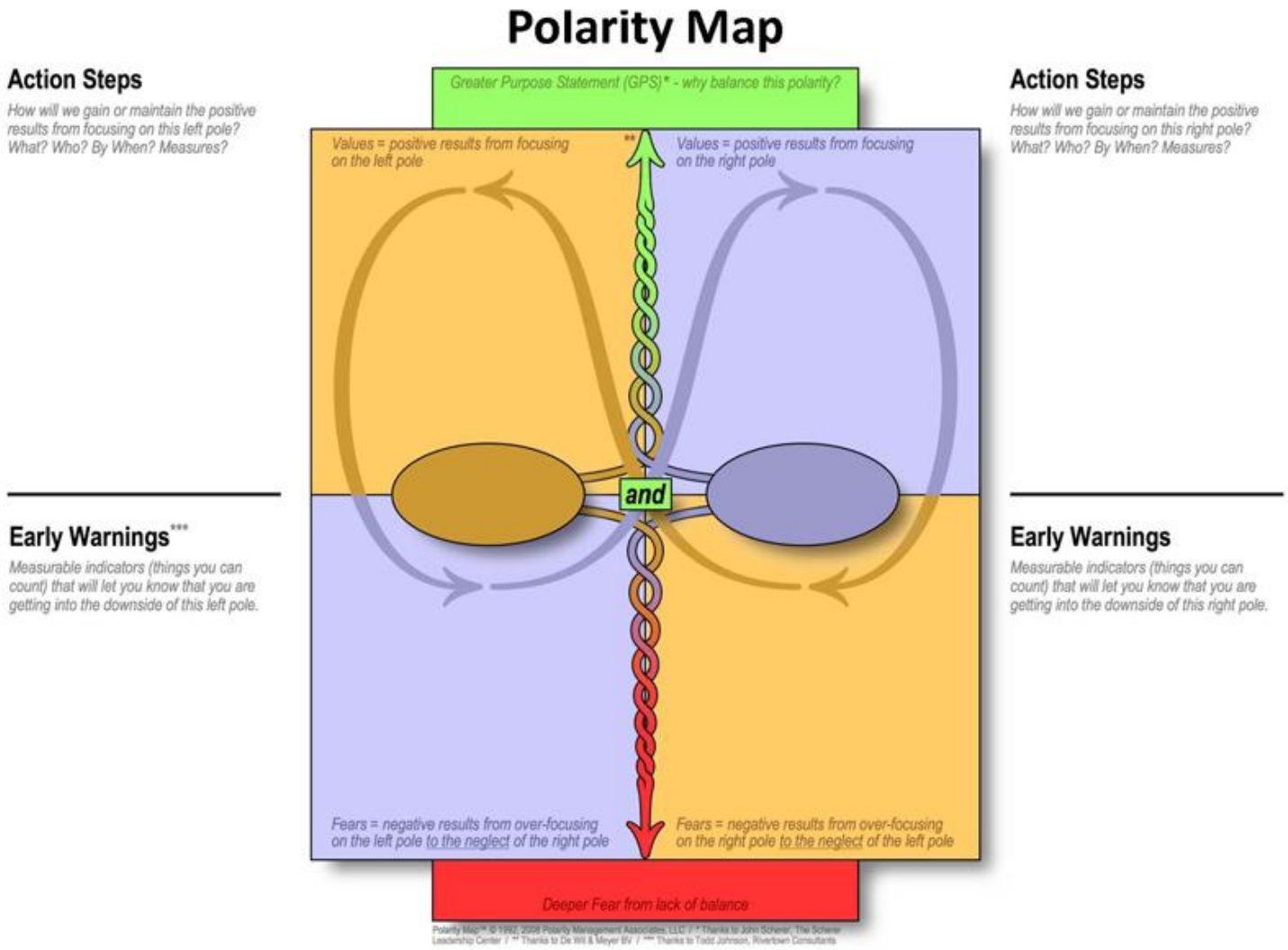
Sketches for Insight:

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

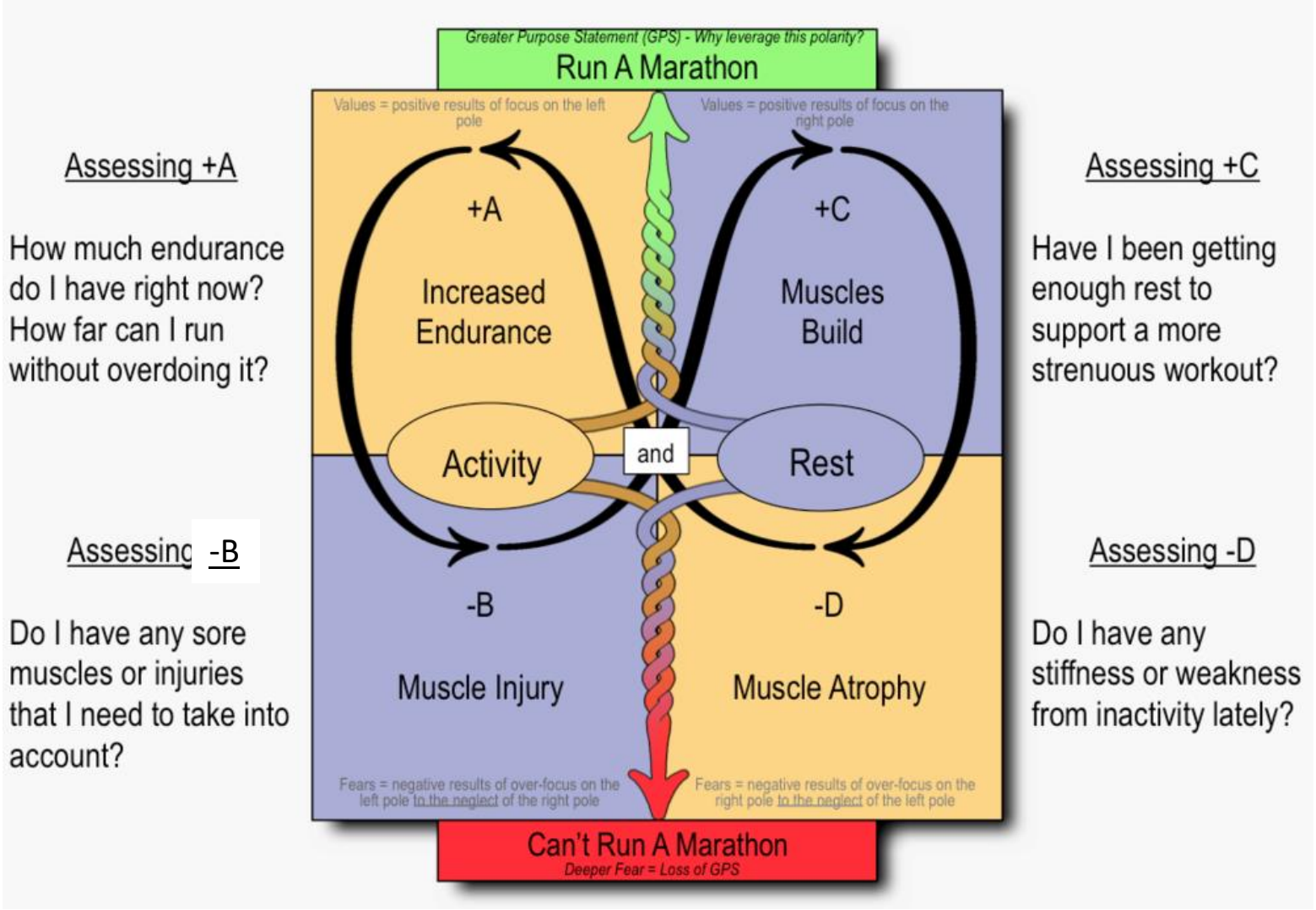

Basic Polarity Map
Why?
Greater Purpose Statement - why Leverage this dynamic tension?
GPS = Greater Purpose Statement
Values = positive results of focus on the left pole
Values = positive results of focus on the right pole
Positive results of focusing on left pole
Positive results of focusing on right 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



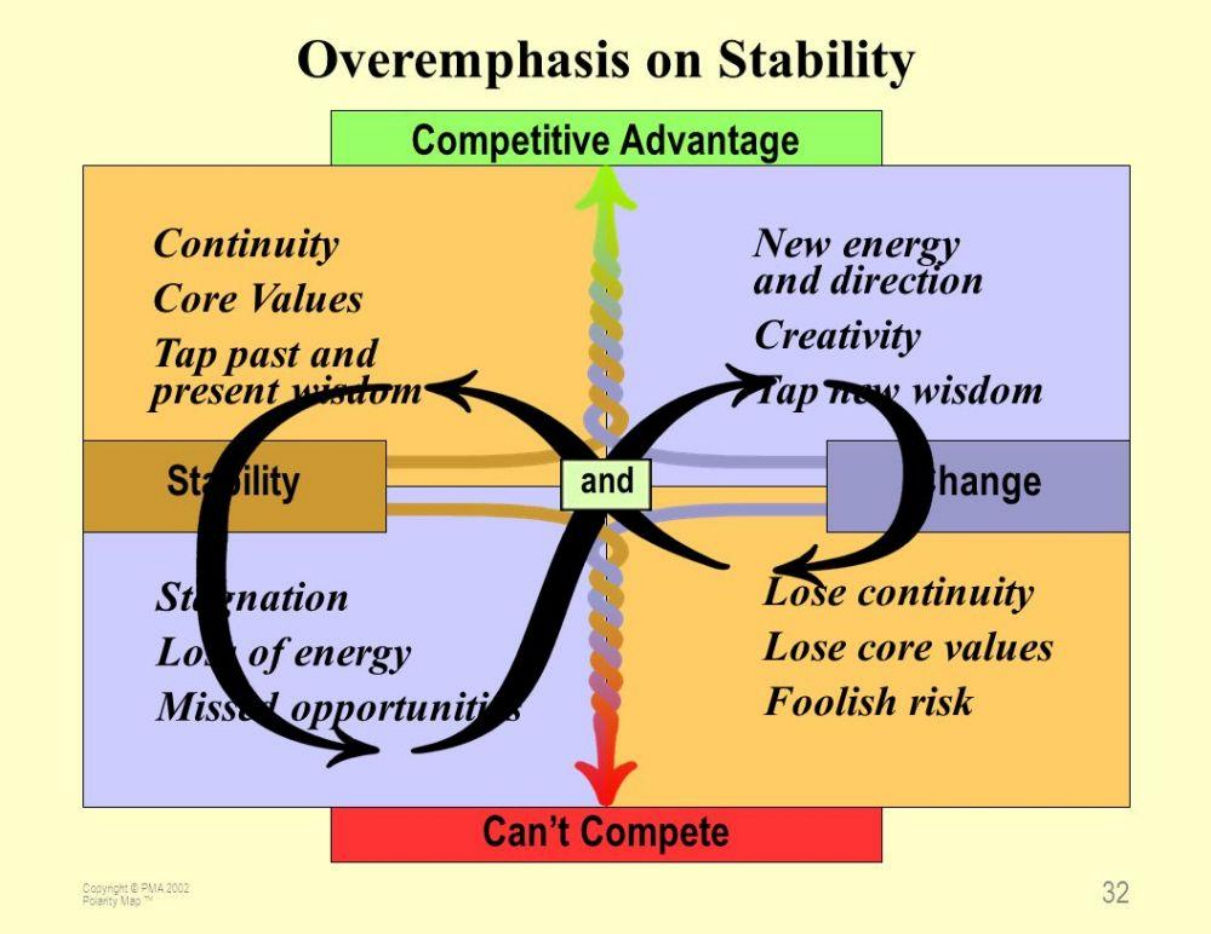
Credit: Polarity Management, Barry Johnson



Credit: Polarity Management, Barry Johnson

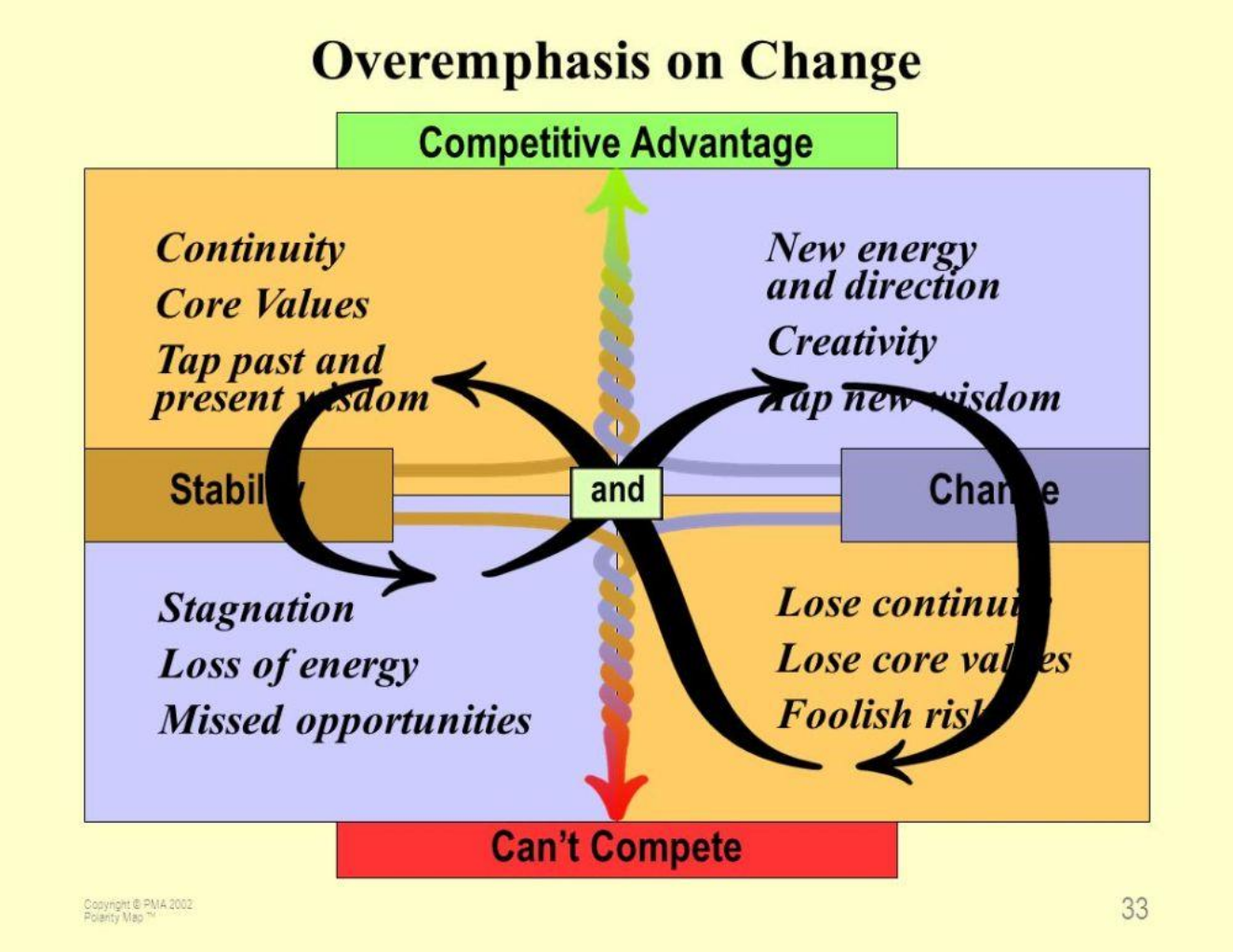


Credit: Polarity



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 focu

- 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 ne

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

Negative results from focusing more this polarity

- Sad Customers (can't 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 gative aspect of this side

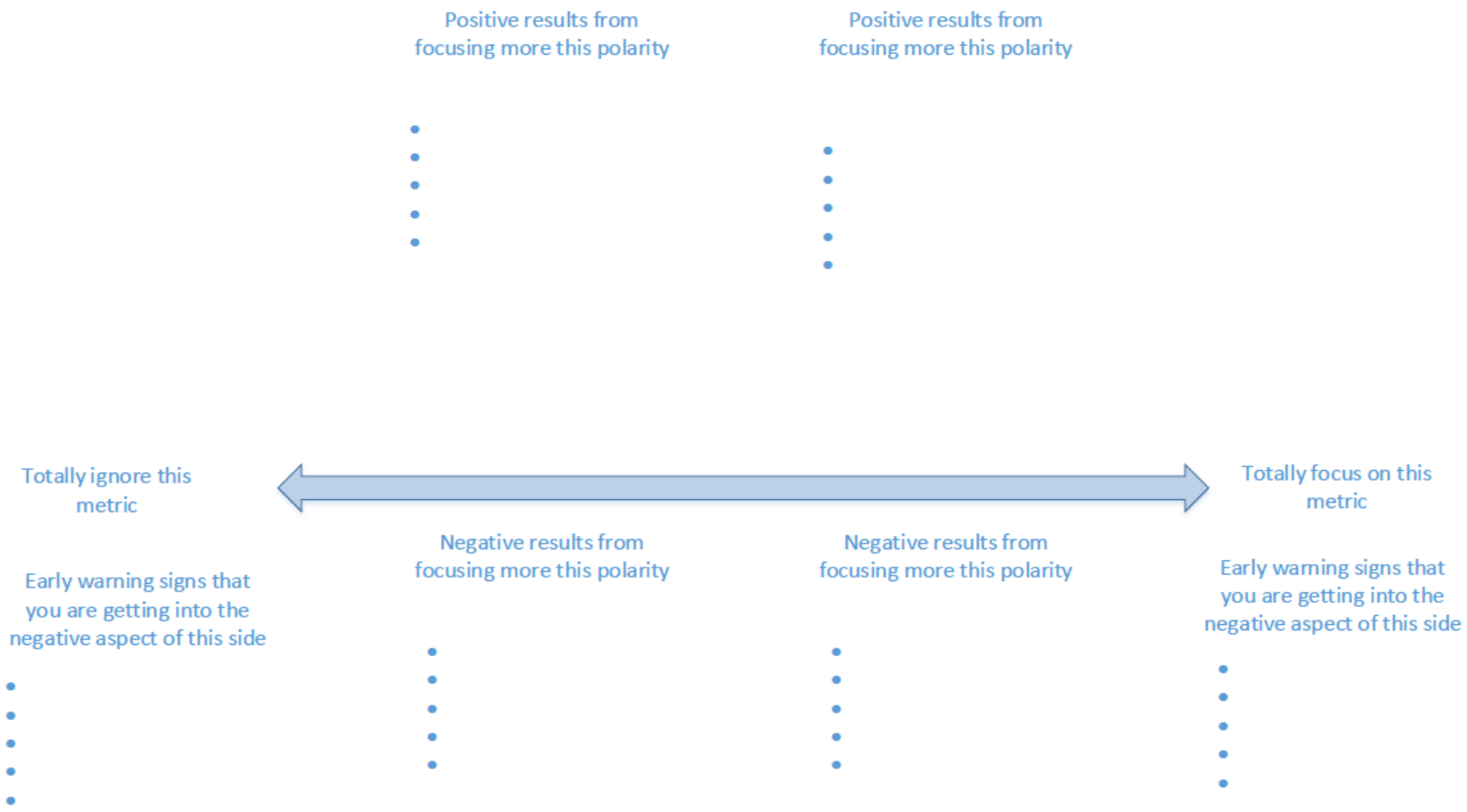
- Longer feature lead time
- Long release cycles
- Fewer significant features per release

Workshop 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 _____

1. Quality
(how well)

- Escaped defect counts
- Forecast to complete defects
- Measure of release "readiness"
- Test count (passing)

2. Productivity
(how much, delivery pace)

- Throughput (/ team size?)
- Velocity (/ team size?)
- Releases per day

3. Responsiveness
(how fast)

- Lead time
- Cycle time
- Defect resolution time

4. Predictability
(how repeatable is productivity)

- Coefficient of variation (SD/Mean)
- Standard deviation of the SD
- "Stability" of team & process

5. Right Things
(building the right stuff)

- Cost of Delay
- Aligned with Strategy

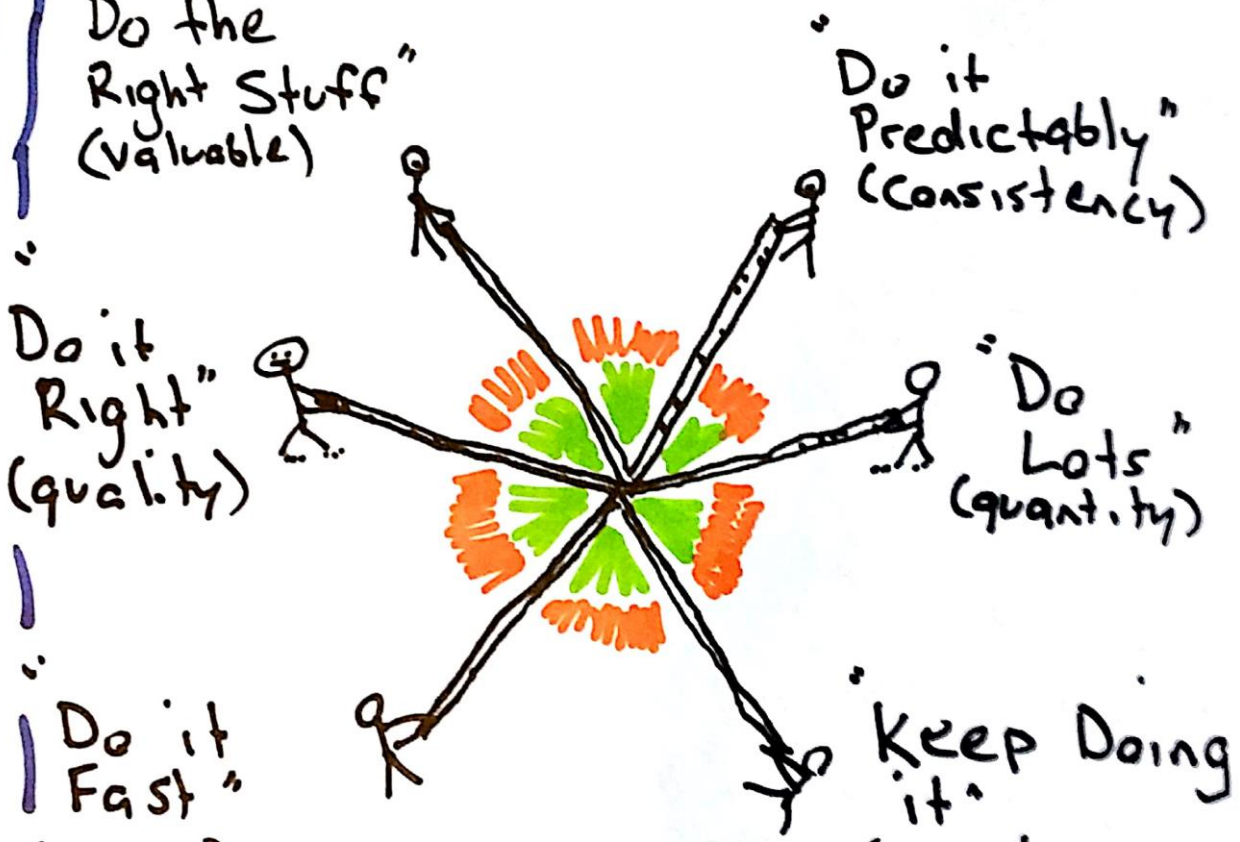
6. Resilience
(ability to keep doing all this)

- Team "happiness"
- Team capabilities
- # Teachers per skill

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

Balanced Metric Dashboard – Description and Examples

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)

Detectable impacts when overdriven?
- Productivity measure declines – team declares "Done" slower
- Responsiveness measure declines – team starts new work slower

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

Detectable impacts when overdriven?
- Quality measure declines – Doing things faster causes defects
- Predictability measure oscillates – Inconsistent rate of delivery

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

Detectable impacts when overdriven?
- Quality measure declines – "Done" things prematurely accepted
- Predictability measure oscillates – Doing too much causes chaos

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

Detectable impacts when overdriven?
- Productivity measure declines – doing less means more consistency

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

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.

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

Workshop 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