Thank you to the attendees, sponsors and organizers.

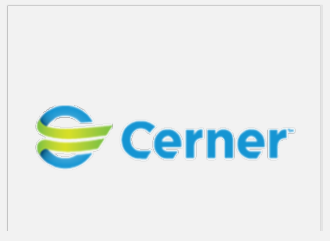
Forecasting using data

LEAN AGILE KANSAS CITY 2017

Troy Magennis

@t_magennis | troy.magennis@focusedobjective.com

All slides and spreadsheets: Bit.ly/SimResources




Forecast
Reality

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."

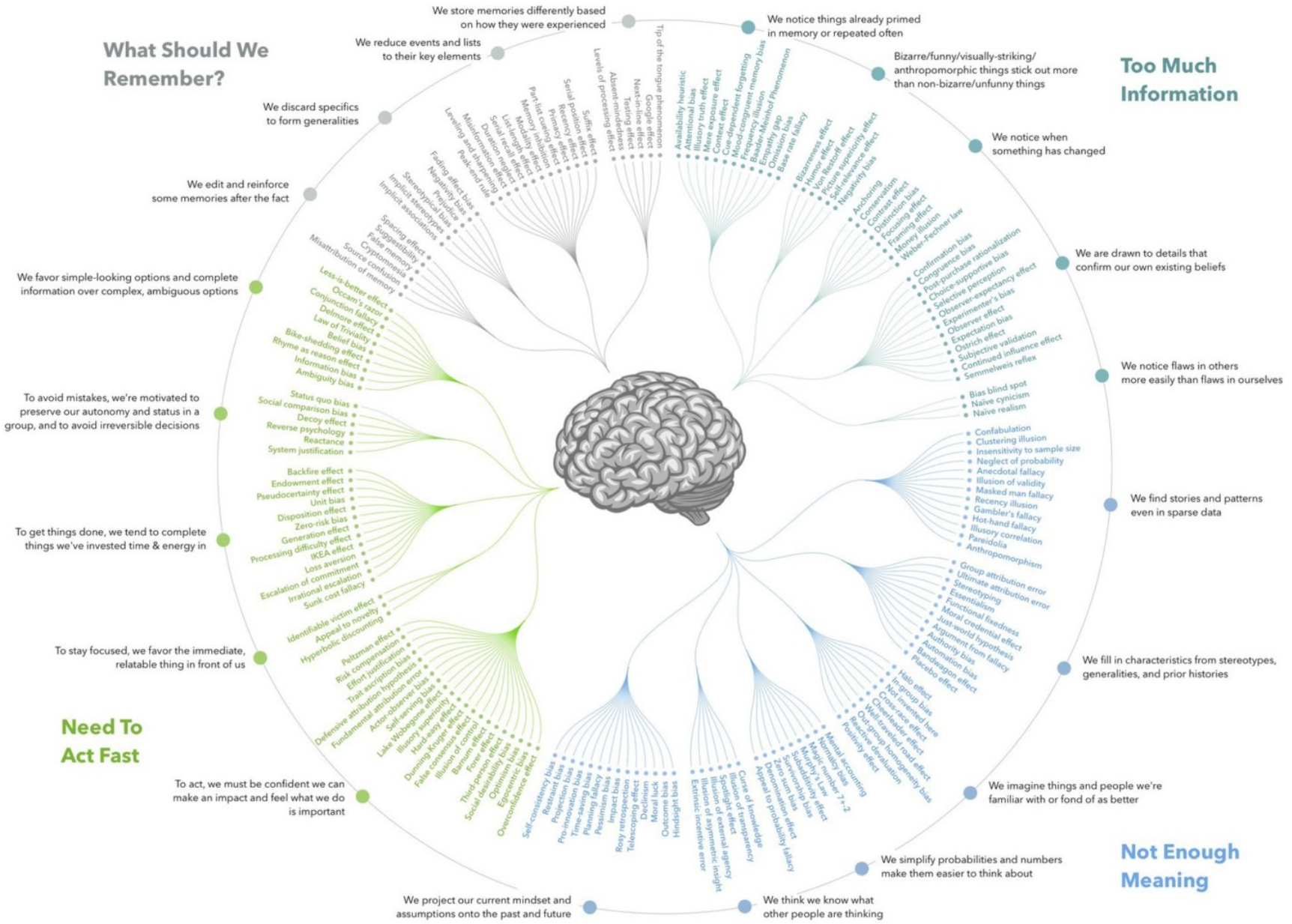
Statistician,
George Box

A. Better than intuition

Not perfect. Not exact. Not always right.

Just better than what you do now, or even equal (just less expensive)

COGNITIVE BIAS CODEX, 2016

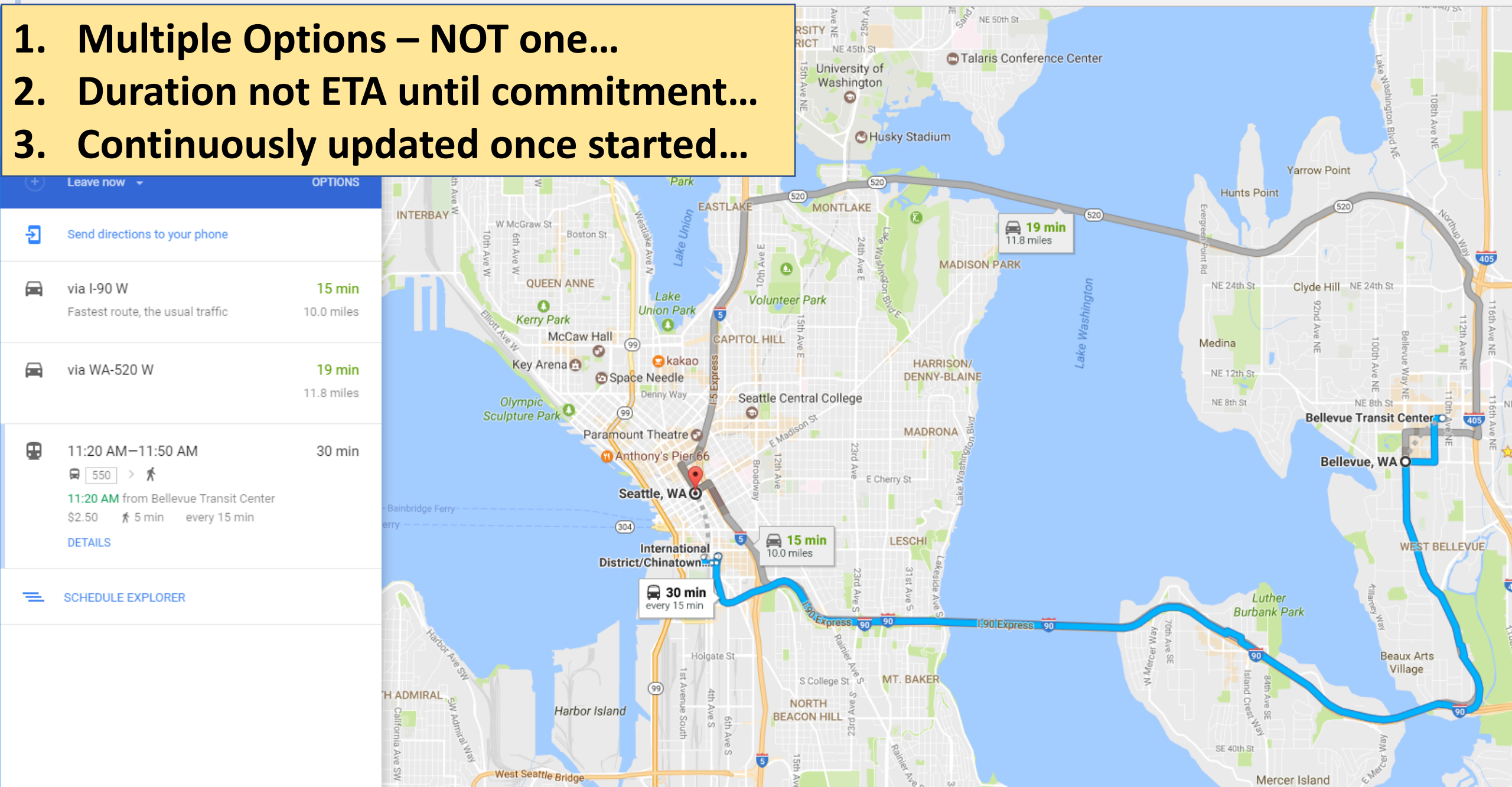


ALGORITHMIC LAYOUT + DESIGN BY JM3 · JOHN MANOOGIAN III // CONCEPT + METICULOUS CATEGORIZATION BY BUSTER BENSON // DEEP RESEARCH BY WIKIPEDIANS FAR + WIDE

http://ritholtz.com/2016/09/cognitive-bias-codex/

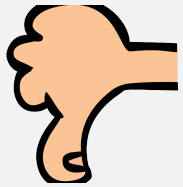
http://ritholtz.com/2016/09/cognitive-bias-codex/

1. **Multiple Options – NOT one…**
2. **Duration not ETA until commitment…**
3. **Continuously updated once started…**



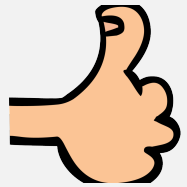
Contrast Google Maps to Software Estimates

If you currently…



- Give one forecast even though multiple approaches considered
- Give a calendar date for undefined “complete” & “start”
- If the original date is in doubt we find out near the end
- Appear on-time until we are not. Measure progress from start.

Consider doing…



- Give multiple options of investment and implementation
- Give a duration and define what started & complete means
- If the original date is in doubt, know earlier and react faster
- Report remaining time to deliver not time since started

Top Three Forecasting Questions….

1. **How Big**
   - Understanding the size of a feature or project with less effort
2. **How Long**
   - Understanding when a feature or project might be done
   - Tracking progress
3. **How Much**
   - OK, its too big,
     "what can I get by when…"
   - Seeing options
   - Making trade-off decisions earlier

Q1: How Big?

Understanding the size of a feature or project with less effort

Step 1

Feature 2
3 stories

Feature 3
7-15 stories

Feature 1
15 stories

Step 2

Feature 4
?

Step 3

Feature 2
3 stories

Feature 3
7-15 stories

Feature 1
15 stories

Feature 4
10-15 stories

Known as Reference Class Forecasting

Forecasting Total Story Count

- Question: How an I estimate the size of a feature or project without analyzing every piece of work?

- Theory: The “size” patterns of randomly sample epics, will persist through all other epics. Analyze a few and compute for the many…

http://bit.ly/StoryCountForecaster

Sampling based Monte Carlo story count forecasting Excel spreadsheet

Feature or Epic Name	Estimated # Stories or points (before starting)
Feature 1	5
Feature 2	3
Feature 3	8
Feature 4	4
Feature 5	2
Feature 6	7
Feature 7	
Feature 8	
Feature 9	
Feature 10	
Feature 11	
Feature 12	
Feature 13	
Feature 14	
Feature 15	

Process to estimate total size –

1. Pick a 5-10 features at random
2. Build sets of 15 re-samples (say 1000 times)
3. The number of sets that reach certain story count levels give probability

1. How many total features do you want to forecast? 15 total features entered on input sheet:

Enter the total number of features or epics you wish to forecast. The patterns exhibited by the story count breakdown of the samples fetures and epics will be extrapolated to this many total features.

2. What rate do you expect work to split? low guess 1 high guess 1 actual

Work often splits into smaller pieces when started by the team. Also, new work gets discovered through defets and learning. Account for that here. 1 no change, 2 means every one item might be split into two, 3 means every item might become three items, etc. Most commn range I've seen is 1 to 3.

3. Result: Forecast total story count or total story points

Likelihood	Total Story Count/points
50%	73
85%	81
95%	85

Odds in english

50% = Coin toss odds. Same chance being above or below this story count

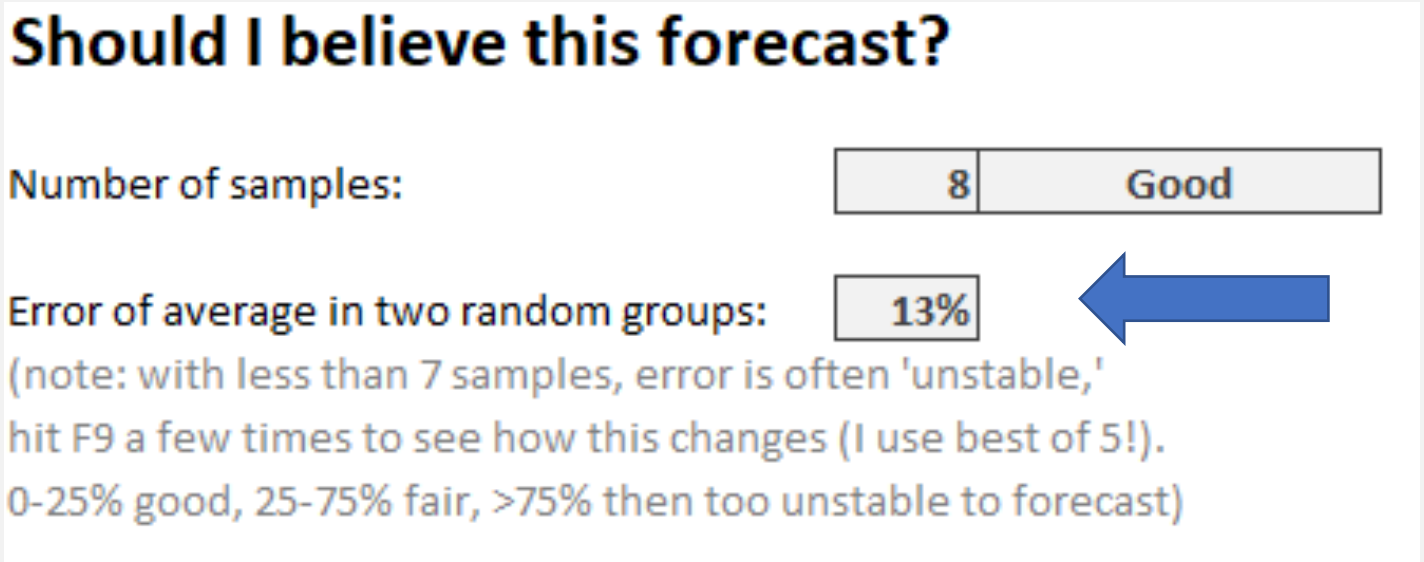
85% = Pretty sure to be equal or less than this story count.

95% = Almost certain to be equal or less than this story count.

Why should I believe this forecast anyway?

1. Sample Count: Keep cutting data and compare the result
2. Random groups: Split data into random groups and compare

Total for 100 Features using	Total Count 85% Likelihood
36 samples	506
10 samples	494
3 samples	504



Average Error calculation –

1. Split the samples into 2 groups
2. Calculate the average of both groups
3. Compare the difference as a % of range
   error % = error of avg / (max-min)

Q2: How Long?

Forecasting duration if nothing else was done…

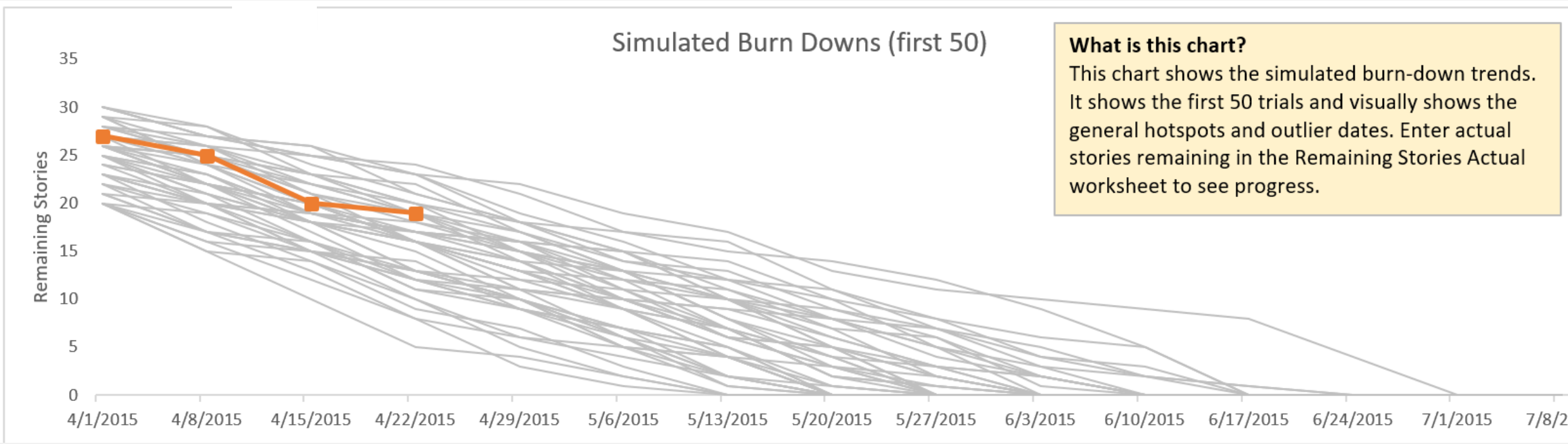
Forecasting Duration (and delivery date)

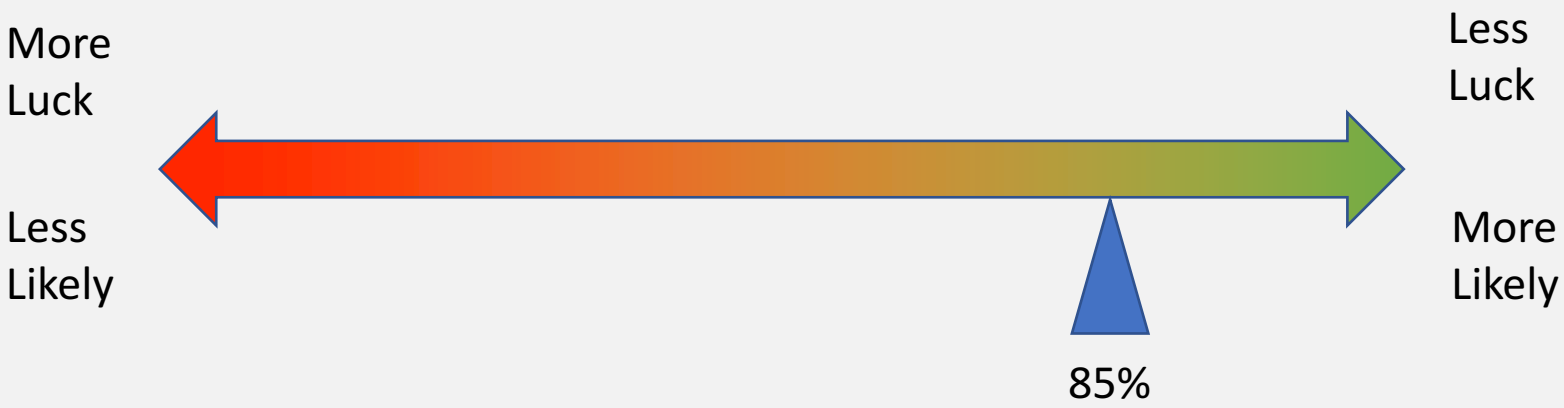
- Question: How can I estimate the amount of time it will take to deliver a feature or project?

- Theory: Using a range estimate or actual team delivery rate data, calculate how many of those periods of time to complete delivery

http://bit.ly/ThroughputForecast

Estimate or Sampling based Monte Carlo
duration and date forecasting Excel spreadsheet


Simulated Burn Downs (first 50)
Remaining Stories
35
30
25
20
15
10
5
0
4/1/2015
4/8/2015
4/15/2015
4/22/2015
4/29/2015
5/6/2015
5/13/2015
5/20/2015
5/27/2015
6/3/2015
6/10/2015
6/17/2015
6/24/2015
7/1/2015
7/8/2
What is this chart?
This chart shows the simulated burn-down trends. It shows the first 50 trials and visually shows the general hotspots and outlier dates. Enter actual stories remaining in the Remaining Stories Actual worksheet to see progress.


More Luck
Less Likely
Less Luck
More Likely
85%

Forecast Completion Date

1. Start Date 4/1/2015

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 20 Highest guess 30

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 1.00

4. Throughput. How many completed stories per week or sprint do you estimate low and high bounds?

Throughput estimate/samples are per Week 7 days

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

Low guess 1 Highest guess 5

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 input 2 and estimate or use historical team velocity for input 4. The benefit of using throughput (count of completed stories per week/sprint) is that the individual stories don't require estimation in story points.

@t_magennis

Results

Likelihood	Duration in Week's	Date	
100%	14	7/8/2015	Almost certain
95%	12	6/24/2015	
90%	11	6/17/2015	
85%	11	6/17/2015	
80%	10	6/10/2015	Somewhat certain
75%	10	6/10/2015	
70%	10	6/10/2015	
65%	9	6/3/2015	
60%	9	6/3/2015	
55%	9	6/3/2015	
50%	9	6/3/2015	
45%	8	5/27/2015	Less than coin-toss odds. But if you are game?
40%	8	5/27/2015	
35%	8	5/27/2015	
30%	8	5/27/2015	
25%	7	5/20/2015	
20%	7	5/20/2015	
15%	7	5/20/2015	
10%	7	5/20/2015	
5%	6	5/13/2015	
0%	5	5/6/2015	

Q3: How Much?

OK, what can we get?

Forecasting How Much (OK, what can I get?)

- Question: I have a date in mind, what features will likely delivery given historical delivery pace?

- Theory: Using duration forecasts, discuss the start order of features that maximize value and likelihood of successful delivery

http://bit.ly/MultipleFeatureForecast

Estimate or Sampling based Monte Carlo
duration and date forecasting Excel spreadsheet
for multiple features at one time

Feature Cut Line Forecaster and Explorer

Only edit orange i

1. Start Date 1/1/2015 **2. Target Date** 7/1/2015 **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 3.00

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

Throughput estimate/samples are per Week 7 days

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

Low guess 5 Highest guess 10

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

				Start date: 01/01/2015	
Start Order	Feature Name (just for reference)	Story Count Low Guess	Story Count High Guess	Feature Duration in Weeks	Forecast Completion Date (85% CI)
1	Feature 1	5	10	3	✓ 1/22/2015
2	Feature 2	8	15	5	✓ 2/26/2015
3	Feature 3	15	25	8	✓ 4/23/2015
4	Feature 4	20	30	10	! 7/2/2015
5	Feature 5	10	40	11	✗ 9/17/2015

Give multiple options – discuss cuts early

				Start date: 01/01/2015	
Start Order	Feature Name (just for reference)	Story Count Low Guess	Story Count High Guess	Feature Duration in Weeks	Forecast Completion Date (85% CI)
1	Feature 1	5	10	3	✓ 1/22/2015
2	Feature 2	8	15	5	✓ 2/26/2015
3	Feature 3	15	25	8	✓ 4/23/2015
4	Feature 4	20	30	10	! 7/2/2015
5	Feature 5	10	40	11	✗ 9/17/2015
6				0	✗ 9/17/2015
7				0	✗ 9/17/2015
8				0	✗ 9/17/2015
9				0	✗ 9/17/2015
10				0	✗ 9/17/2015

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

Top Three Forecasting Fail Reasons

Reasons you shouldn’t have hired me five years ago

Fail 1: Start Date On-Paper != Reality

- The assumed Start Date is often ONLY on paper

Start Date of Feature B is the finish date of Feature A

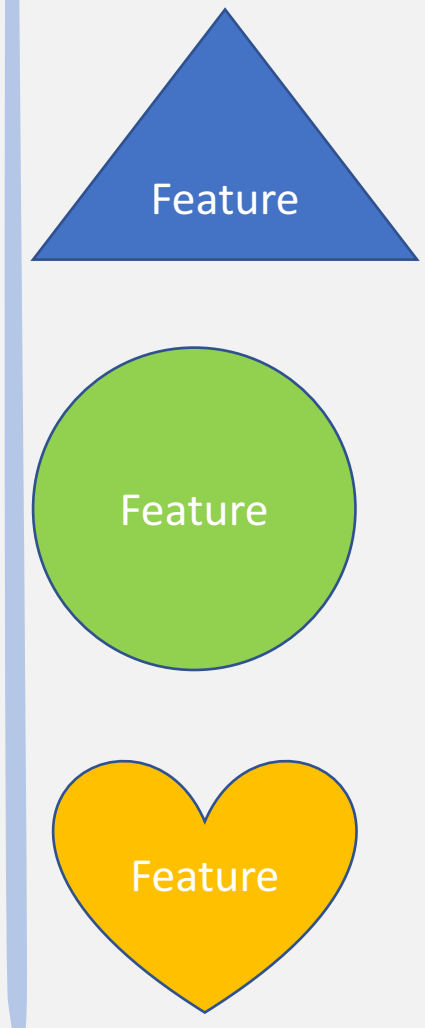
What is the team doing now?

- Define what start means
  - Team is dedicated and in-place
  - They are trained and know how to do their work
  - They know and understand what work they need to deliver
  - Nothing inhibits them doing or delivering that work

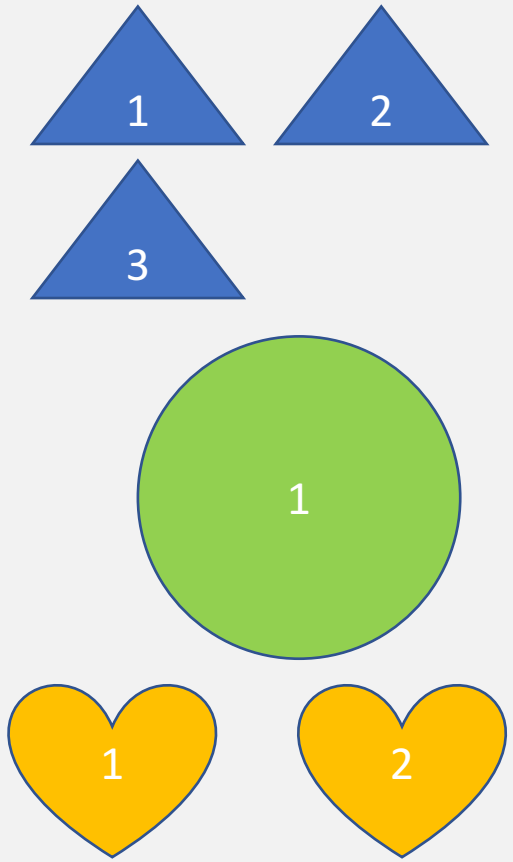
- Team is never fully available on day one!

Fail 2: Backlog Rate versus Delivery Rate

Features



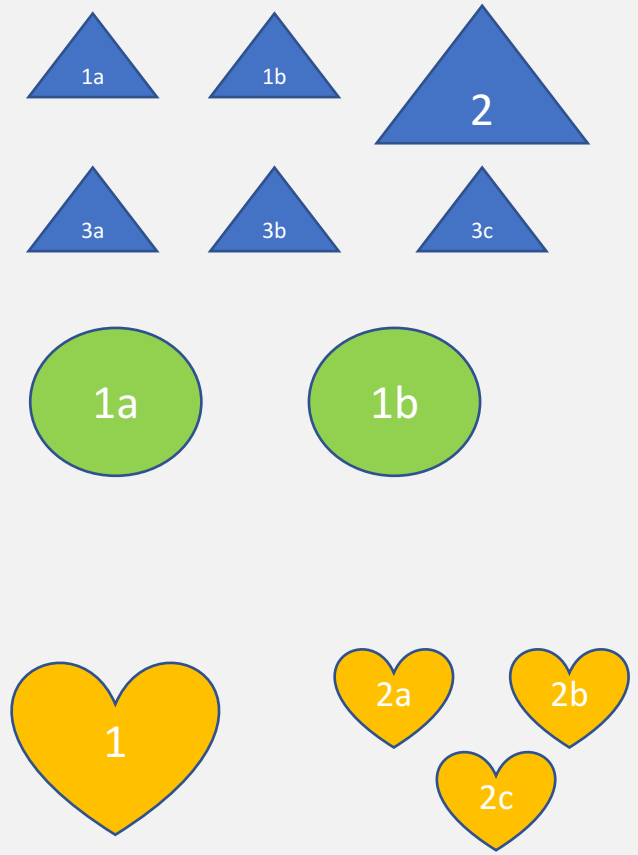
Estimated Stories



Actual Backlog

Rate Delivered = 6

Implemented Stories and Defects



Measured Throughput = 12

Fail 2: Backlog Rate versus Delivery Rate

- Forecast using the “Completion rate” we may under-forecast
  - Backlog is Miles per Hour, Completion rate is Kilometers per Hour
- Normal split rates are between 1 to 3 times (most common seen)
- This means
  - If you don’t account for it, you will UNDER-FORECAST by 1 to 3 times!

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	3.00

Feature or Epic Name	Estimated # Stories or points (before starting)	Actual # Stoies or points (after completed)
Feature 1	5	8
Feature 2	3	4
Feature 3	8	10
Feature 4	4	
Feature 5	2	
Feature 6	7	
Feature 7		

5 X **1.6** = 8

3 X **1.3** = 4

8 X **1.25** = 10

actual growth rate range seen: **1.25** to **1.60** (enter actual count after work is completed in the input sheet)

ere.

1 to 3.

This split-rate adjust the rate that the team appears to be completing work with the rte original backlog items are started. This isn't only growth/creep, its adjusting for the way work splits when the team hs a closer look at it.

Fail 3: Ignoring Risks

- Risk = Work that “might” need to be done but we don’t know yet

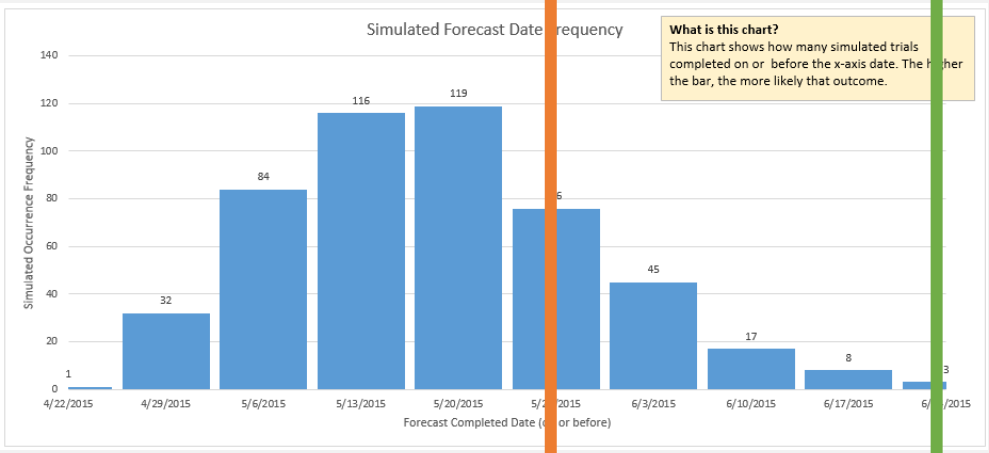
- Some samples
  - Fails on Internet Explorer 6, or now Safari on phones
  - Fails performance testing under load, or uses too much memory
  - CSS alignment issues with German text translations, things wrap
  - Production network security blocks traffic, awaiting vendor to fix
  - Fails on real customer data (we designed for 50 items, they have 500)

WITHOUT RISKS INCLUDED

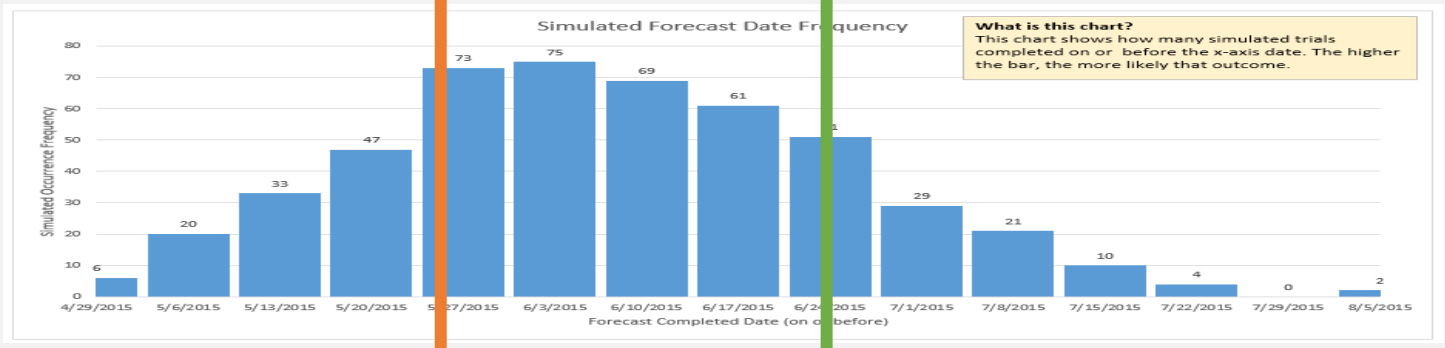
27th May (highest late June)

24th June (highest early August)

WITH RISKS INCLUDED

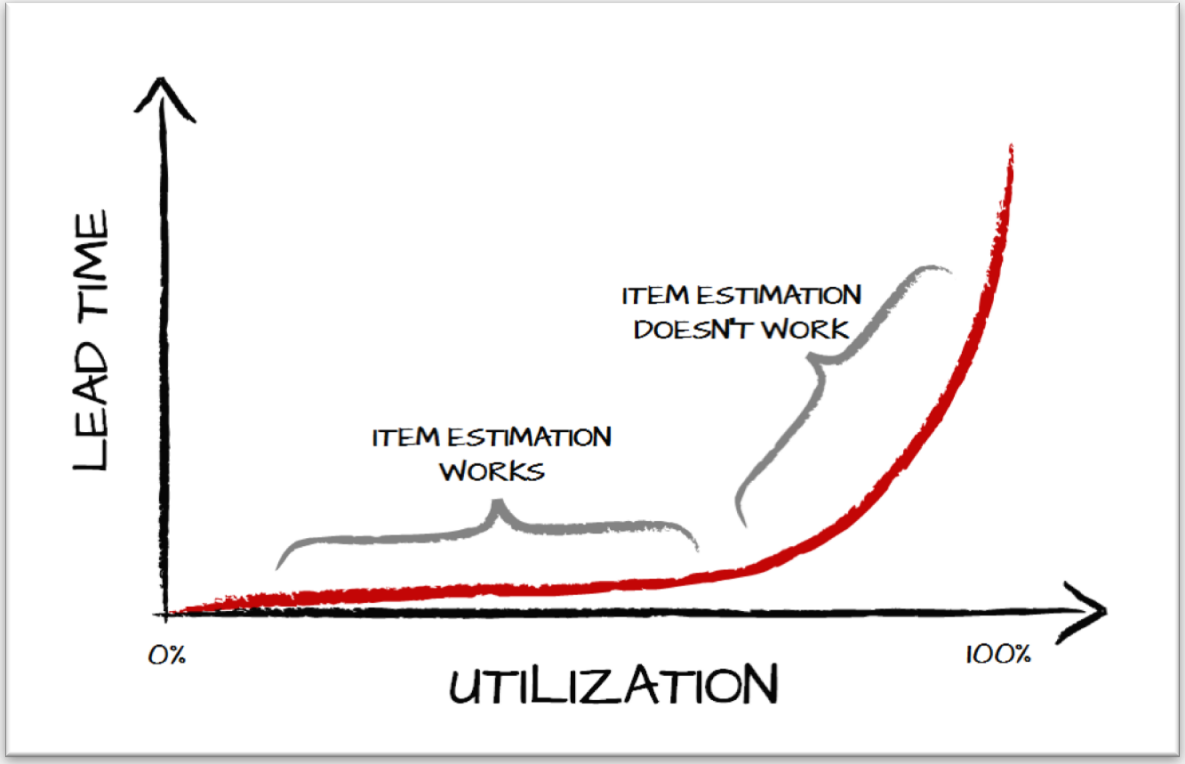


Likelihood	Impact Low	Impact High	Description
50%	7	10	Browser compatibility issues
40%	3	7	Performance under load
30%	5	10	Production configuration



Forecasts shown at 85th Percentile

Root of All Fails: High System Utilization



Can’t forecast high utilization systems using item size…

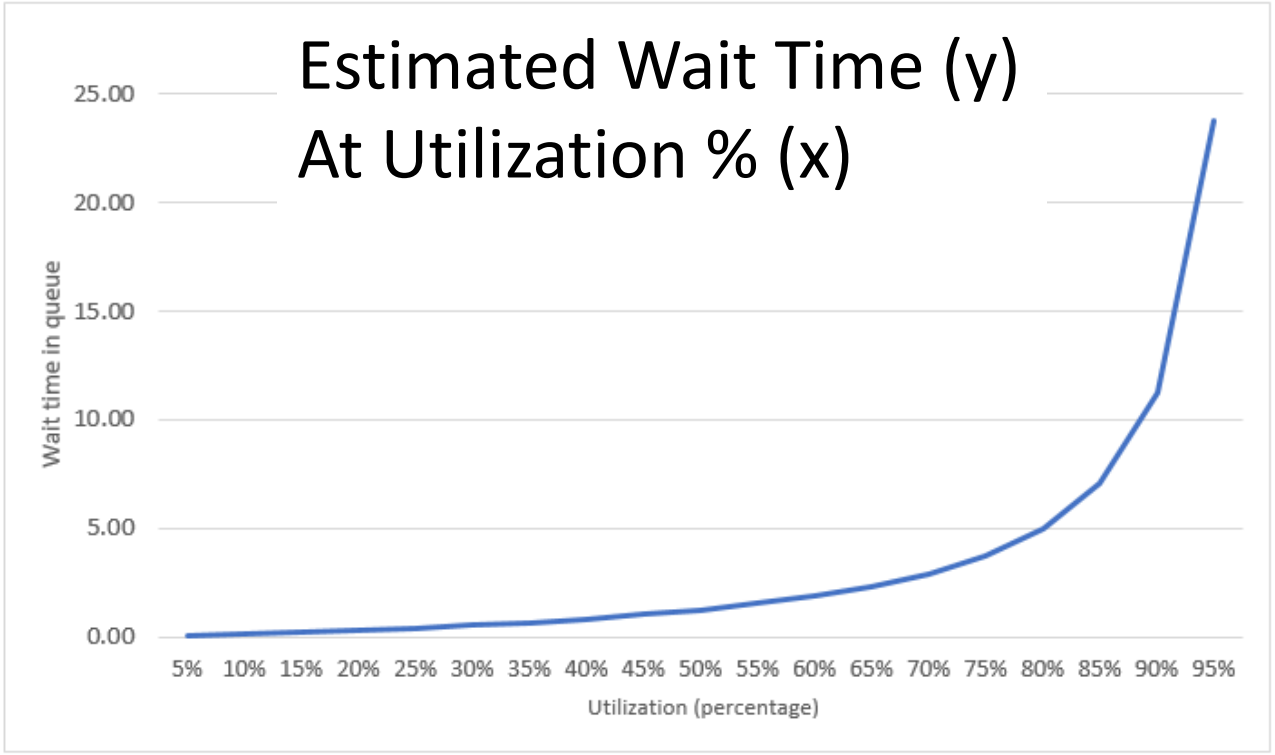
Kingman's Approximation

Calculations on the distributions:

mean service time	5	any time un
arrival time variation	0.500	higher is mo
service time variation	0.500	higher is mo

Kingman's approximation states

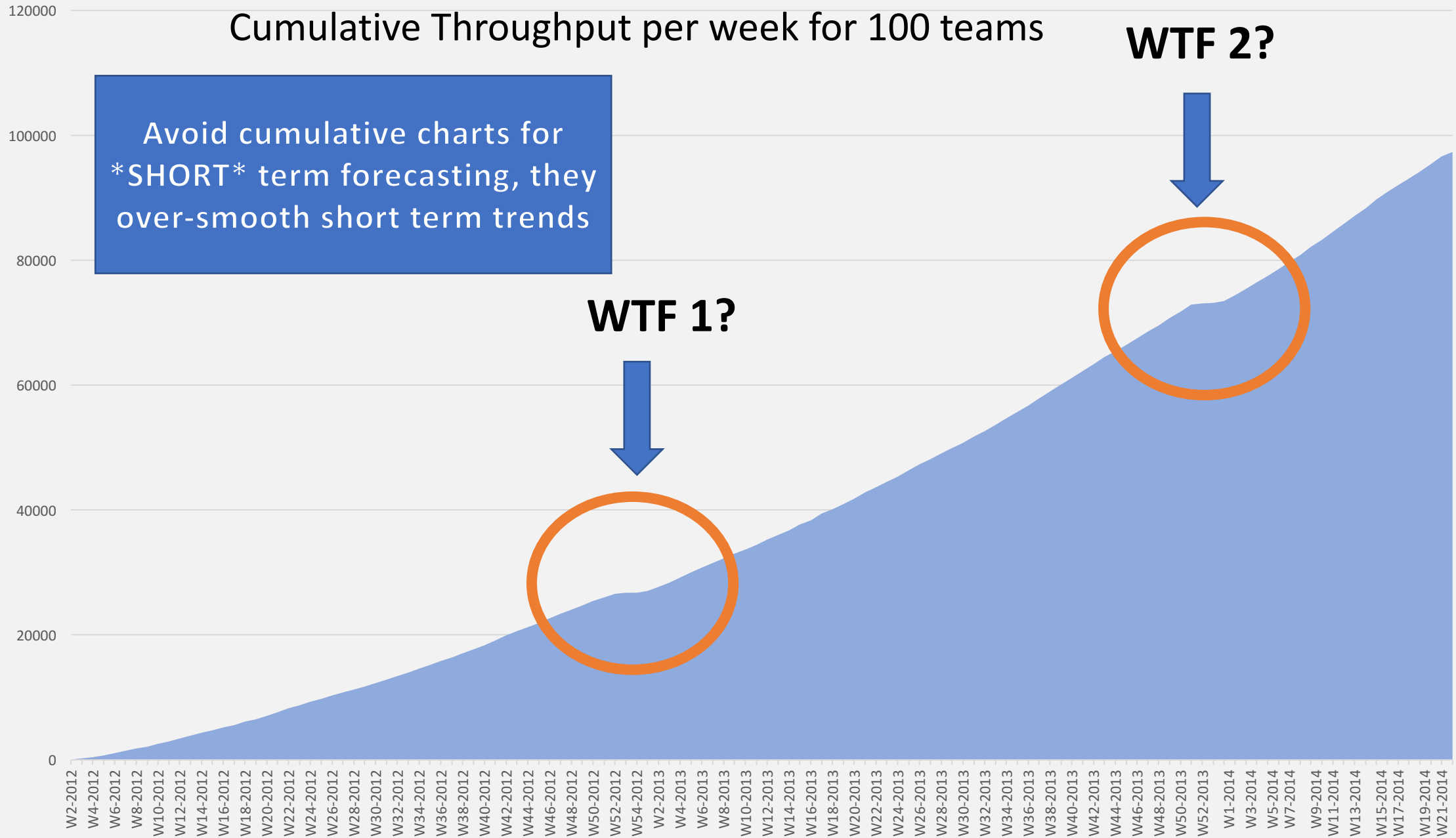
$$\mathbb{E}(W_q) \approx \left(\frac{\rho}{1-\rho}\right)\left(\frac{c_a^2 + c_s^2}{2}\right)\tau$$



Action: Experiment with “Kingmans Formula.xlsx”

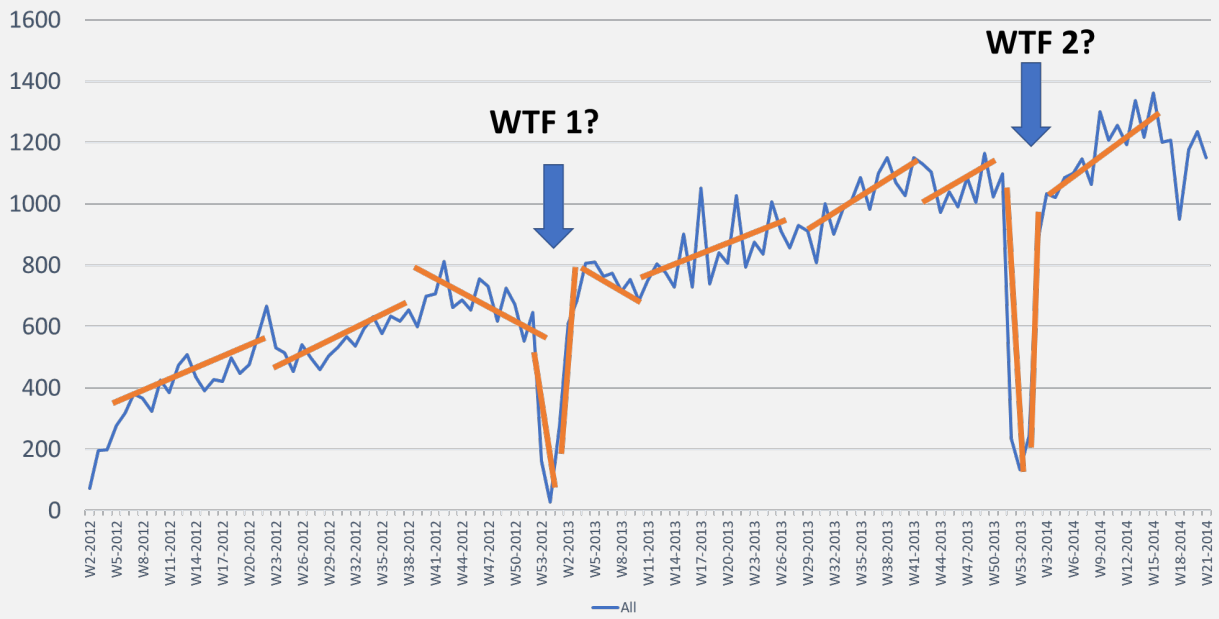
Arrival Time Variation AND Service Time Variation Equally Impactful


Throughput per week for 100 teams
NOTHING you can estimate poorly will be as DAMAGING as a bad process or poor management policy
WTF 1?
WTF 2?
1600
1400
1200
1000
800
600
400
200
0
W2-2012
W5-2012
W8-2012
W11-2012
W14-2012
W17-2012
W20-2012
W23-2012
W26-2012
W29-2012
W32-2012
W35-2012
W38-2012
W41-2012
W44-2012
W47-2012
W50-2012
W53-2012
W2-2013
W5-2013
W8-2013
W11-2013
W14-2013
W17-2013
W20-2013
W23-2013
W26-2013
W29-2013
W32-2013
W35-2013
W38-2013
W41-2013
W44-2013
W47-2013
W50-2013
W53-2013
W3-2014
W6-2014
W9-2014
W12-2014
W15-2014
W18-2014
W21-2014
All


Cumulative Throughput per week for 100 teams
Avoid cumulative charts for *SHORT* term forecasting, they over-smooth short term trends
WTF 1?
WTF 2?
120000
100000
80000
60000
40000
20000
0
W2-2012
W4-2012
W6-2012
W8-2012
W10-2012
W12-2012
W14-2012
W16-2012
W18-2012
W20-2012
W22-2012
W24-2012
W26-2012
W28-2012
W30-2012
W32-2012
W34-2012
W36-2012
W38-2012
W40-2012
W42-2012
W44-2012
W46-2012
W48-2012
W50-2012
W52-2012
W54-2012
W2-2013
W4-2013
W6-2013
W8-2013
W10-2013
W12-2013
W14-2013
W16-2013
W18-2013
W20-2013
W22-2013
W24-2013
W26-2013
W28-2013
W30-2013
W32-2013
W34-2013
W36-2013
W38-2013
W40-2013
W42-2013
W44-2013
W46-2013
W48-2013
W50-2013
W52-2013
W1-2014
W3-2014
W5-2014
W7-2014
W9-2014
W11-2014
W13-2014
W15-2014
W17-2014
W19-2014
W21-2014

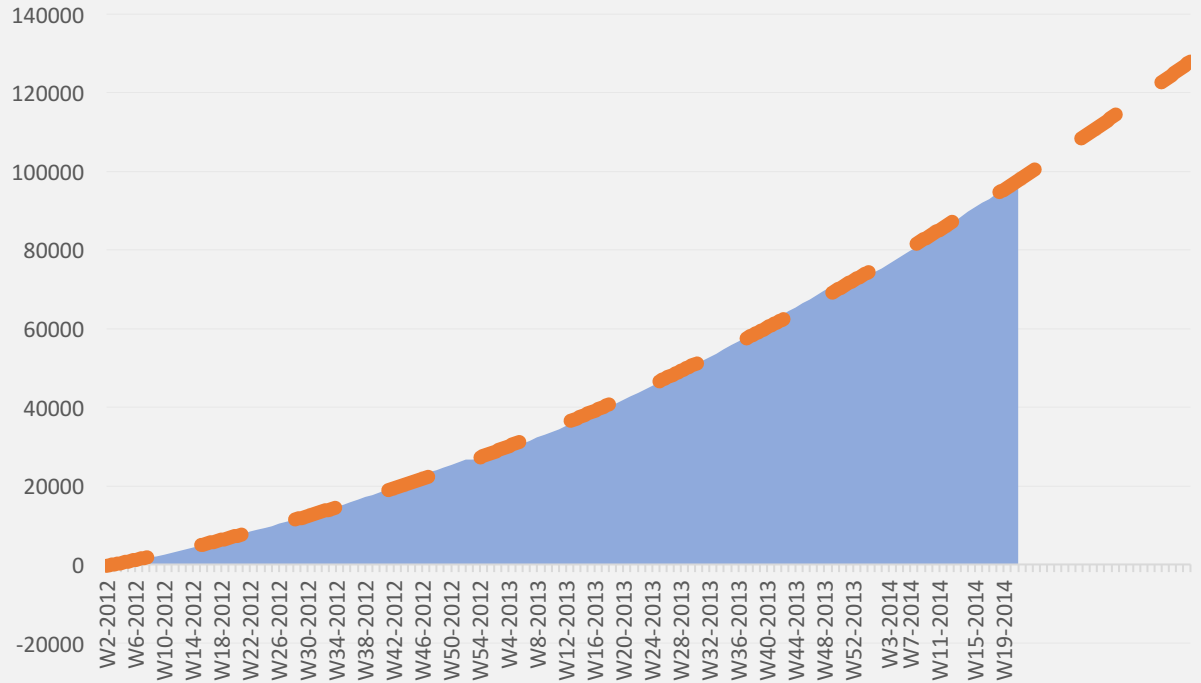
Use RECENT trends for shorter term contextual forecasts

Eg. Individual stock (with 2 x CEO indictments)



Use cumulative trends for longer term forecasts

Eg. S&P500 over 10 years



Key Take-aways and Resources

- Forecasting requires a system view,
- Three samples will outperform intuition (use most recent 7 samples)
- Give multiple options, not just one
- Forecast duration NOT date until “Start Conditions” are defined
- Track actual progress versus planned, and update the model continuously

- Get everything here: Slides and tools:

Bit.ly/SimResources

Get everything here: Slides and tools:

Bit.ly/SimResources

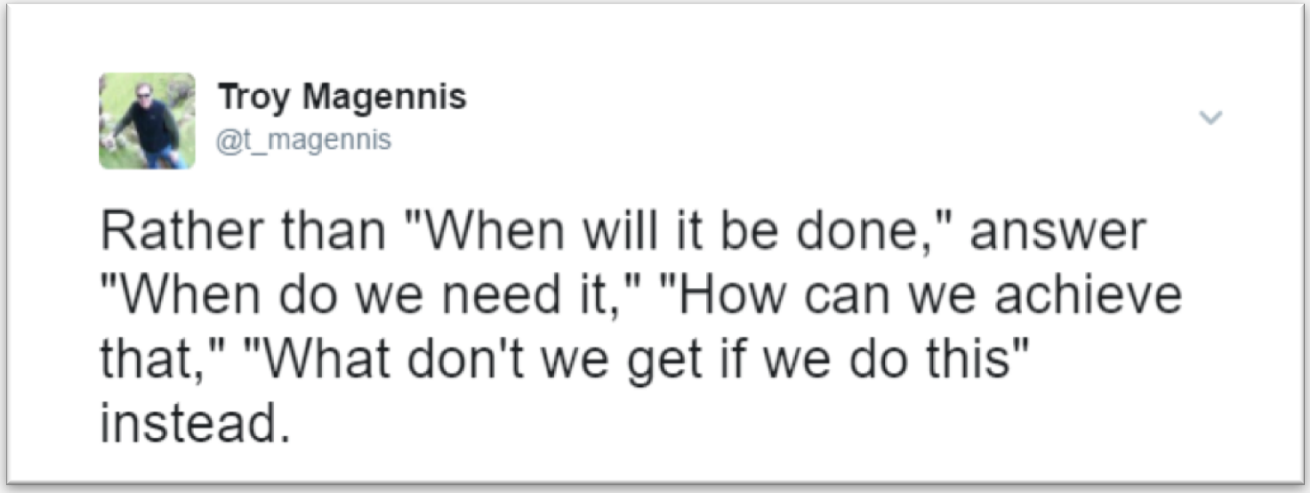
Me on Twitter

@t_magennis

About me…

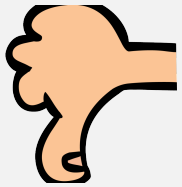
- What I do
  - Teach how to use data for forecasting
  - Teach simple math to executives, especially “demand > supply”
  - Teach how to know (earlier) that you are on the wrong side of an expectation
- What I did
  - Started in software 1986. I actually liked Assembler & Cobol
  - Have worked at senior exec level, and now beside them for major corporations so I have some insight into what passes their decision filters
- How to reach me
  - Twitter: @t_magennis or email: troy.magennis@focusedobjective.com
  - Lots of free spreadsheets and stuff at FocusedObjective.com

Forecasting is…
answering the right question,
to a transparent degree of certainty,
with as little effort as possible.



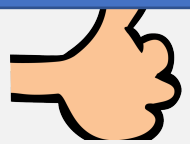
Good versus Poor Forecasting

I think the Google comparison slide is more compelling



Poor Forecasting

- Based on “intuition” alone
- One option without details
- Shows a calendar date (with assumed fixed start date)
- Re-planning seen as failure; New information slowly (if at all) incorporated



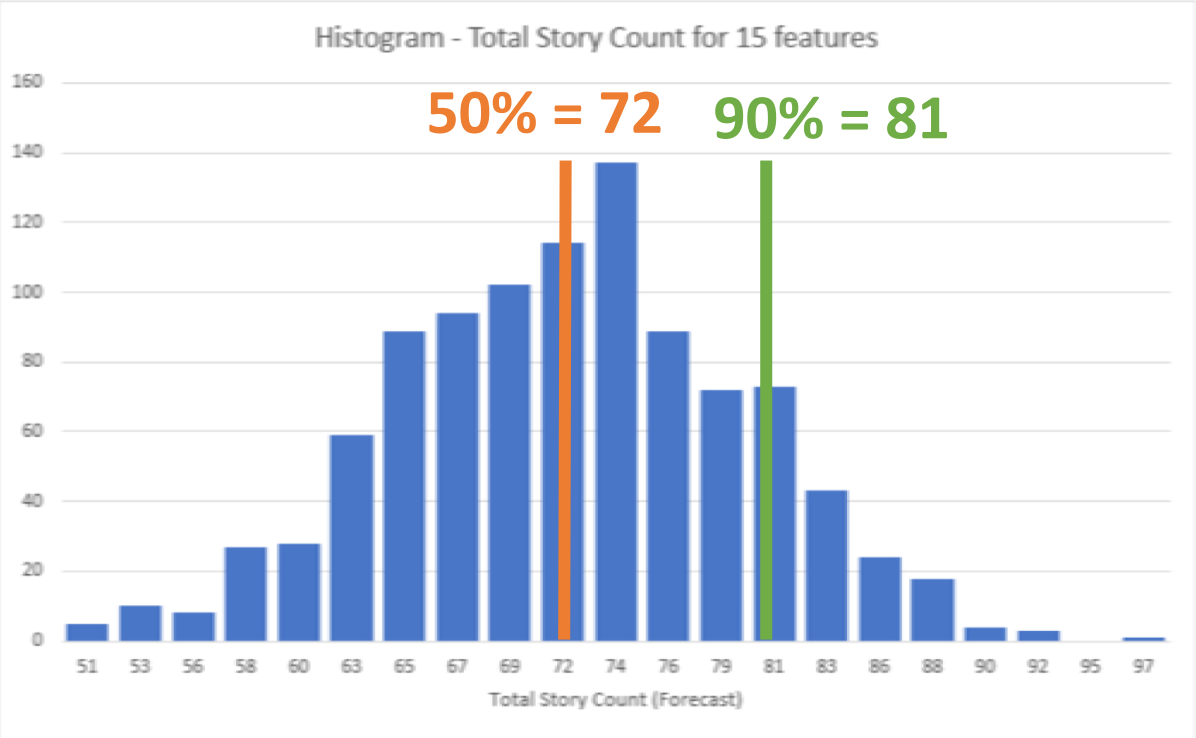
Good Forecasting

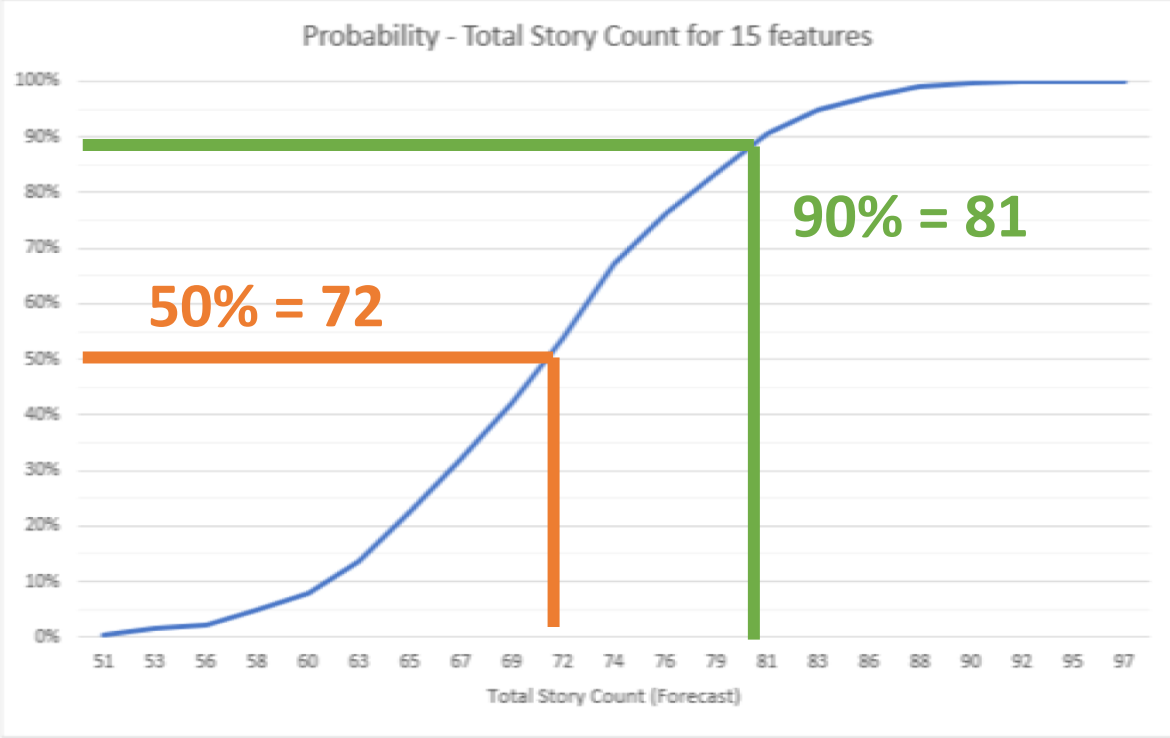
- Based on “intuition” and “data”
- Many options with details
- Shows duration until commitment
- Continuously updated and “rerouted” based on new information

Top Three Forecasting Opportunities

I think I can cover these as we go. Better explanation with examples of the problems.

1. Understanding options
   - knowing that more than one way works
   - Under-invest versus over-invest
2. Sampling
   - getting the same result for less effort
3. Defining a better start order (prioritization)
   - We often optimize for value alone…





This chart plots how likely each total story count result occurred in the Monte Carlo simulation. The higher the bar, the more often that count of stories occurred. It is used to understand the pattern of the results to see how wide the tail values extend (those either side of the peak). If the range is too wide to make forecasting useful, the only solution is to make the features more consistent in size to avoid the story count estimate range being so huge.

This chart plots how likely each total story count result is as cumulative probability the result is less than or equal to a total story count. If you want to know what total story count is 80% likely, look at where the 80% on the Y-Axis intersects the blue line, and read off the total story count on the X-Axis.

How Big – Forecasting feature/project size

- First, can we avoid the question (quantitatively)
  - "When is it needed?" – perhaps it's an easy binary answer of yes or no
  - "Compared to what?" – perhaps it's an easy binary answer of bigger or smaller
- Then,
  - Do we know enough to perform a forecast?
    - Are the Knowns > Un-knowns
- If we do need a size forecast,
  - How can we forecast with the minimal effort possible:
    - Reference class forecasting
    - Sampling

How Long – Forecasting duration (and date)

- First, can we avoid the question
  - "When is it needed?" – perhaps it's an easy binary answer of yes or no
  - "Compared to what?" – perhaps it's an easy binary answer of longer or shorter
- Then,
  - Do we know enough to perform a forecast?
    - Are the risks and unknowns larger than the knowns – perhaps it's a pointless question now
- If we do need a duration forecast,
  - How can we forecast with the minimal effort possible:
    - Use range estimates
    - Use Monte Carlo forecasting

Monte Carlo Forecasting

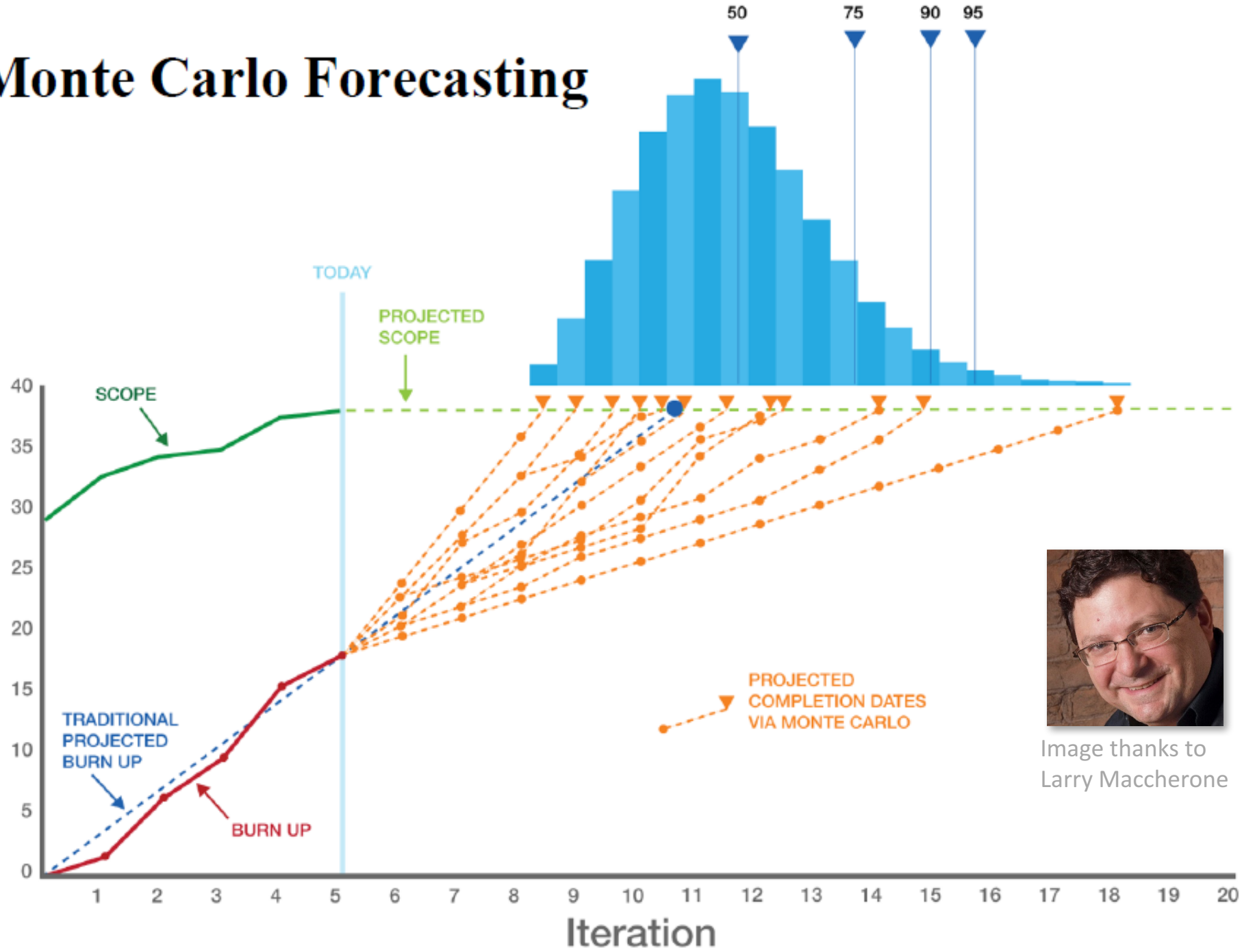
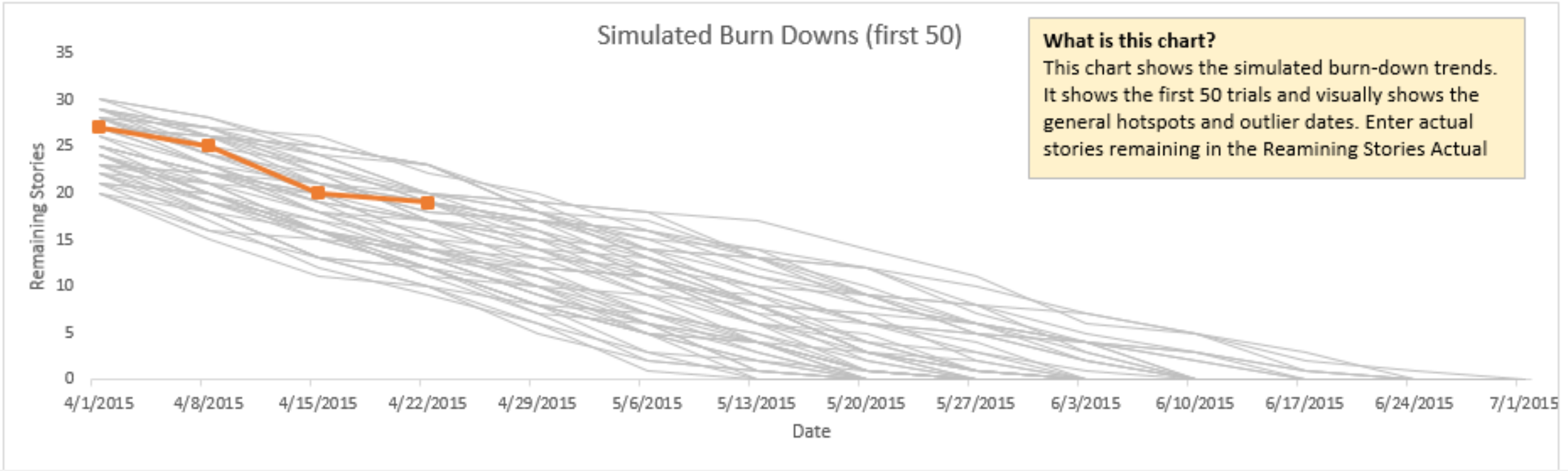
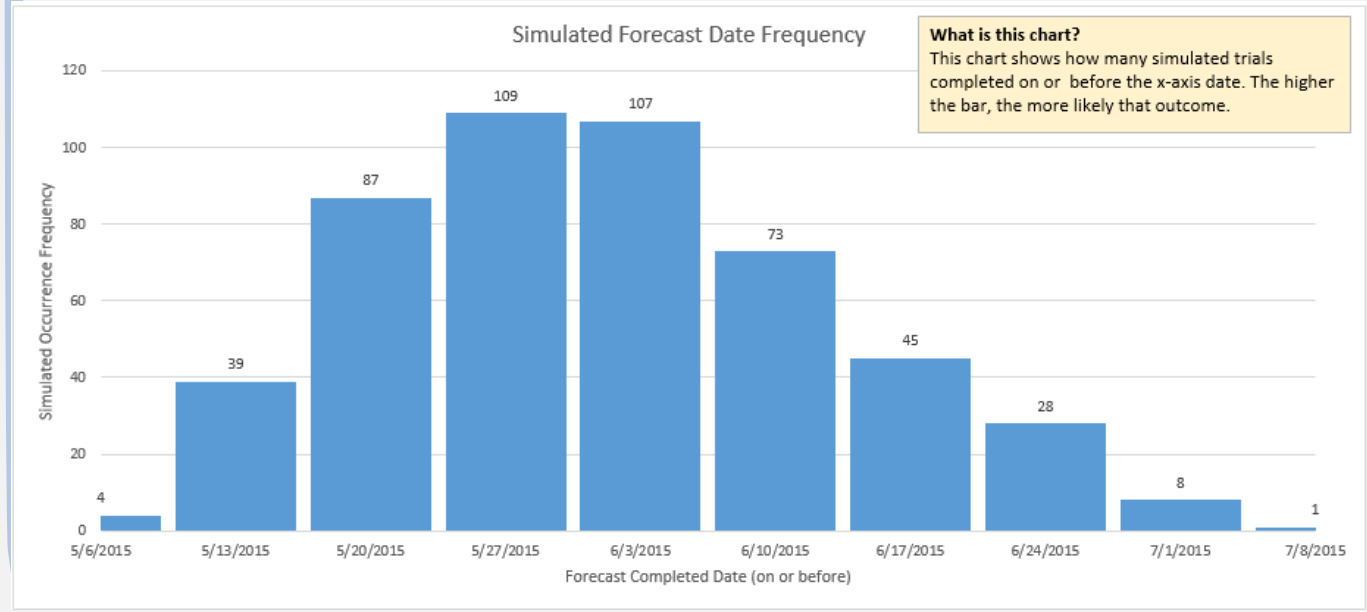


Image thanks to Larry Maccherone





Process –

1. Simulated project burn-downs are created, say 1000 of them.
2. The number of trials for each duration allows probability to be calculated.
3. If we have throughput or velocity data, it is used rather than the range estimate.

Using range estimates or data?

None

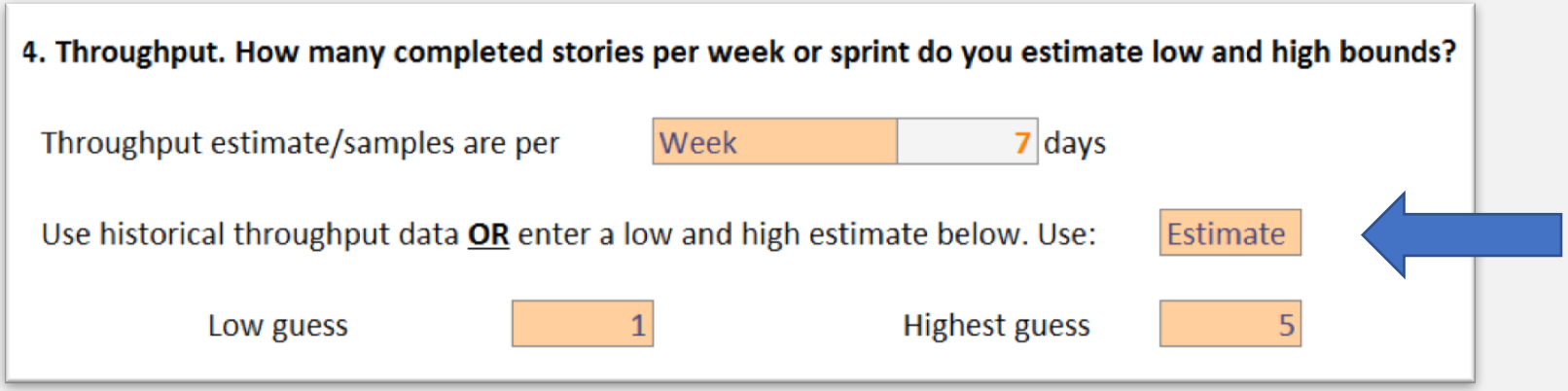
- Use a uniform range estimate

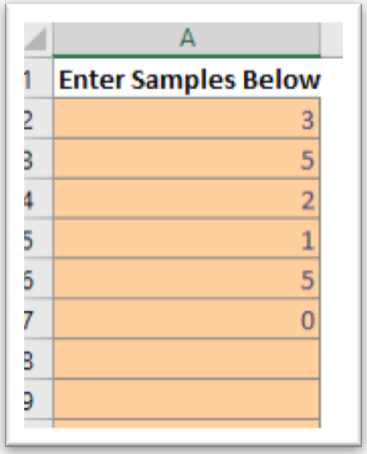
< 11 samples (in context)

- Use data to adjust uniform range estimate boundaries

11+ samples (in context)

- Use data.
- Monitor distribution weekly for changes





Select “Data” rather than “Estimate” to use samples

Slides, spreadsheets, and other stuff

Bit.ly/SimResources

Everything you see is freely available