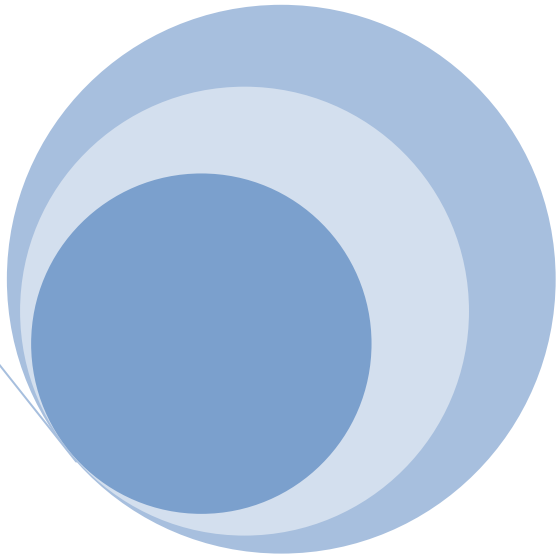
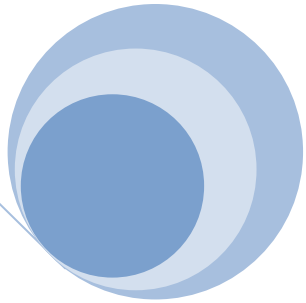
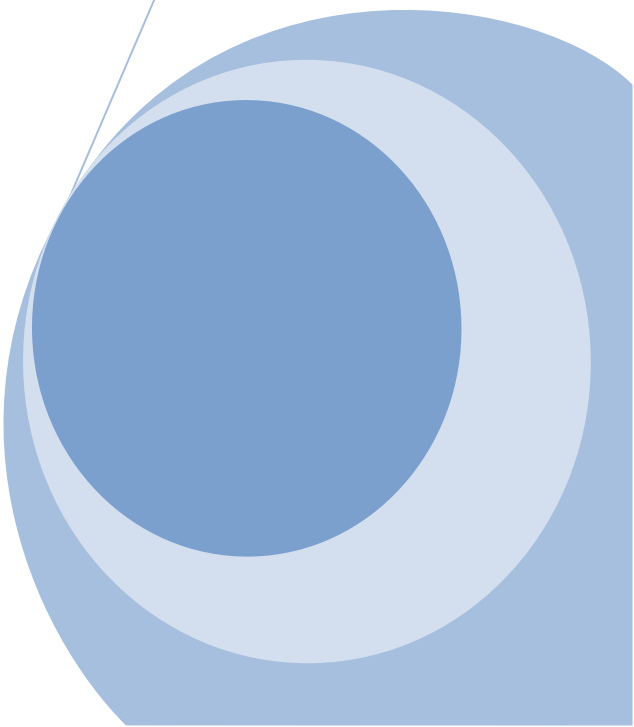




How to make your data speak

from data to action

Kulpreet Singh Nanda



Speaking with a friend of mine, Robert, I could smell frustration in his language. He strongly feels that his IT organization is spending a lot in collecting data and publishing regular metrics reports. The general perception, however, is that the entire mechanism is adding any value.

To me, this doesn't look like a problem specific to Robert's IT organization. Organizations bring a lot together to collect data and report information but the question is - are they able to make their data drive decisions and actions?

Not sure if it's realized that reporting data is just the starting point. ***The real challenge lies in decoding the data to understand what it says and translating that into result oriented actions!*** This definitely needs a disciplined approach and appropriate focus.

Are we in same boat?

- ✓ Have you felt sometimes that you collect data, create graphs and it ends there?
- ✓ Have you seen people looking at the metrics reports and have a confusing look on their face, asking which way to go now?
- ✓ Have you seen that many times we just keep reporting the same data for long period of times without asking if it's actually adding value?
- ✓ Have you witnessed the P to T (printer to trash can) cycle for metrics reports

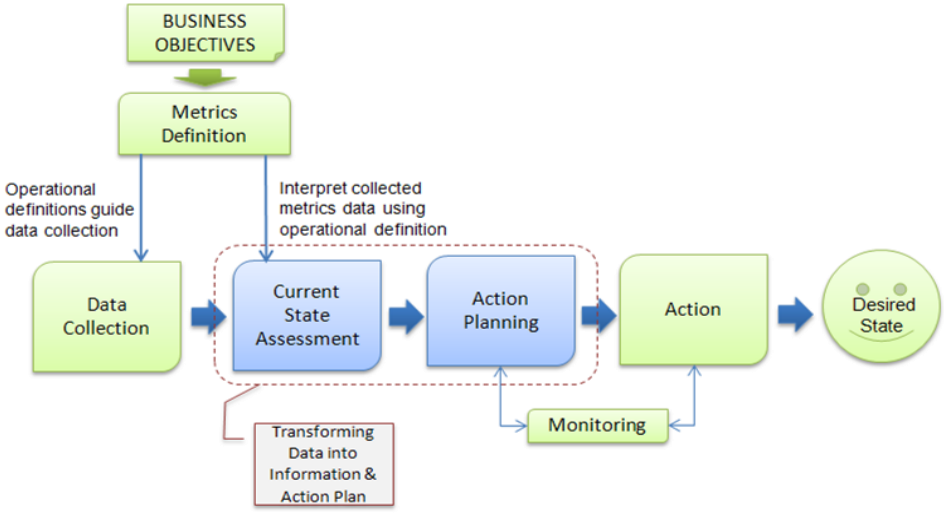
To understand how to solve this mystery, we will have to analyze the totality of data based improvements.

Data based improvements

All industry process models for improvements profess data based actions. To name few, Measurement & Analysis process area of CMMi, Continual Service Improvement within ITIL V3, Measurement, Analysis & Improvement section of ISO 9000, all focus on data based decision making & improvements. Two critical steps in the entire data based improvement journey are:

1. Defining right measurements or metrics
2. Understanding data on measurements and converting them into improvements

Note: -Though the concepts in this paper are valid for any type of data analysis, I have tried to keep my paper focused on situations and scenarios specific to Software Development and Maintenance.



A typical measurement based improvement framework would look something like a happy face model.

People take both of top down and bottom up approaches to define measurements. Top down approach typically focuses on ***"what needs to be measured"*** and bottom up on ***"what can be measured"***. A right balance of both approaches is required to define right measurement system. We can leverage Goal-Question-Metrics approach (GQM) for defining the right measurements which are aligned to specific goals. Also, Balanced Score Card, which is a measurement based strategic tool, can help us creating top down relationship between the lead and lag measures.

However, even after defining & reporting the correct measurements, the foundation of data based improvement remains – understanding the data in totality and translating them into actions. The focus of this paper is on two blue boxes in above picture (current state assessment & action planning), which transforms data into information & action plan. In other words answer - ***How to make your data speak!***

The questions

I was mentoring a group of test analysts, who were tasked to study defect data and generate action plan (if any). The analysts used some standard defect metrics definitions such as defects by phase & by deliverable, phase containment, defect density etc. They could plot a good amount of charts and graphs (thanks for fabulous graphing tools available) and generate reports. Everyone was happy with the way the report looked. However, the quest was what to do next or in other words the challenge was – ***how to actually use the data!***

While mentoring these types of analysis sessions, here are few issues which I have seen people facing:

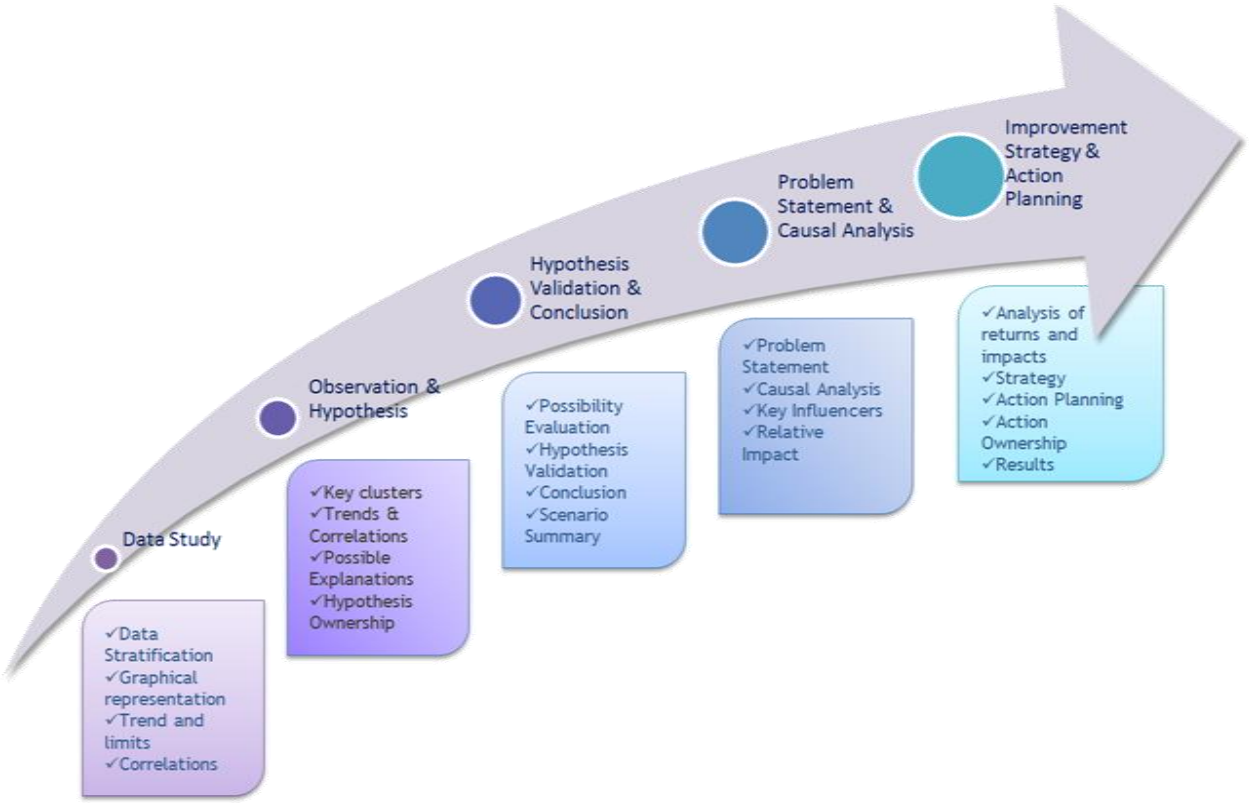
- ✓ There is a big variety statistical tool. Which one fits where?
- ✓ How to start with data analysis and what steps to take further to convert it into action
- ✓ When to involve and whom to involve in the "process" of converting data into actions
- ✓ How to understand, what influences what in the whole mess of numbers
- ✓ How to keep wok focused so as to avoid phenomenon of "analysis – paralysis"

The answers to these questions lie in selecting few key tools & techniques, involving correct people at correct time, keeping the momentum of analysis alive and tracking the thread of data analysis to action planning and monitoring.

From Data to Action : Approach

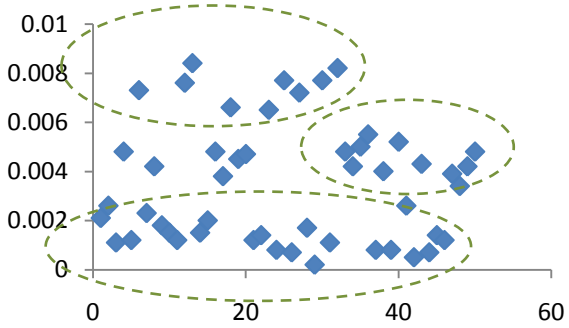
Here is an approach which has worked for me. The strength of this 5 stepped approach is that it works through fewer tools, involves stakeholders at appropriate time & for appropriate reason and focus on clarifying ownership for various steps. Here are the five steps:

1. Data study
2. Observation & hypothesis
3. Hypothesis validation & conclusion
4. Problem identification and causal analysis
5. Improvement strategy & action planning

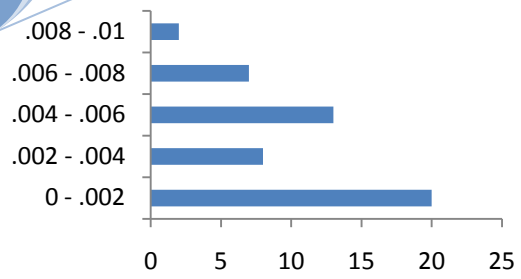


Step 1 : Data Study

There are whole lot of statistical tools which can help us go deep into data. The first thing which I like to do with data is "stratification". Its really helpful to identify the clusters of data and understand the phenomenon which is possiblly keeping the data points together. The possible phenomenon could be type of data source, some special conditions with strong capability to influence results etc. The intent of this exercise is to find key patterns and group of data which can be studied for similarity.



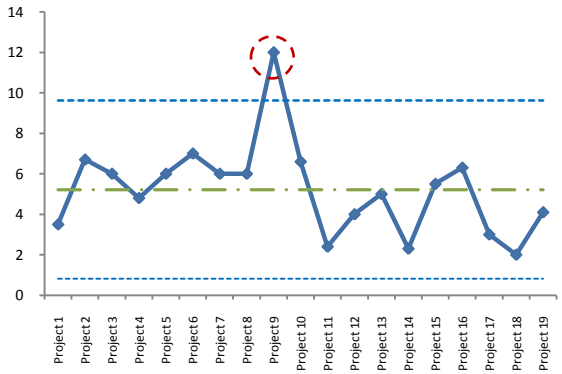
Here is a sample of defect denstiy scatter plot from different projects. We can figure out that some commonality exists between the datapoints within the circle. Its can be bacause of project type,



technology, project organization, skills etc. Through startetification, we can easily identify the groups of data points and start brainstroming on the factor/s which tie these datapoints together. This step actually sets the tone of data analysis in our journey and help analyst/s think in direction of X-to-Y relationship for a larger regression concept.

Another tool which help us analyze the distribution is histogram. Its helps in understanding the centrality & spread of data.

In addition to scatter plot & histogram, its important to realise the natural boundaries of the data set and its trend. XmR chart is a great tool to interpret process behaviors. Few questions which can be answered by an XmR chart are :



- Questions related to trend
- Questions related to normal variation in data
- Questions related to special cause variation (signal)

Overall, the above techniques and tools should be able to help us understanding about the data population.

Step 2 : Observation & Hypothesis

Second step of the approach focuses on calling-out our understanding about data set – our observations. Observations can be as simple as X & Y co-vary or increasing trend of perticular measurement or 40% of total data points behave in a perticular way. After stating our observations "loud and clear" next step is to build hypothesis. Hypothesis can be kept as simple as – "there could be 4 reasons for the stated observation". My experience has been that stating the observations and possible reasons for those observations itself triggers the involvement of relevant role holders. Lets say that the observation is - defect logging in a perticular group of projects is showing a decreasing trend. The possibile reasons could be :

1. Improvement in deliverable quality
2. Teams closing defects without logging them
3. Nature of the project dictates the reduction in defect logging over the life cycle
4. Quality appraisal activities (review and testing) are getting reduced focus hence reducing coverage

The sample spreadsheet to document observation and related hypothesis is given in next step

Step 3 : Hypothesis Validation & Conclusion

The next step in the approach is to align role holders to specific hypothessis for validation. This step takes teams one step closer to the actual issue. Typically, identifying and involving specific team

members to own and conclude specific hypothesis helps concluding faster and effective. If it's a small group, the validations can happen as group. Most of the times, the validation would involve meeting different role holders to know their perspectives, studying direct indirect information and applying information in context of organizational culture. We can go for hypothesis validation by statistical tests if the nature of numbers need that. Mostly, we can get answers by talking to appropriate people. We just need to be careful that we don't get trapped into a cycle of analysis – paralysis.

Depending upon the results of validation, conclusion can be drawn. In above defect logging example, lets say that we concluded on hypothesis #4 to be the one contributing the most to the observation – "reduction in defect logging trend". This conclusion becomes the basis of next step of the approach. The team can take this conclusion as problem statement/s for next stage of analysis.

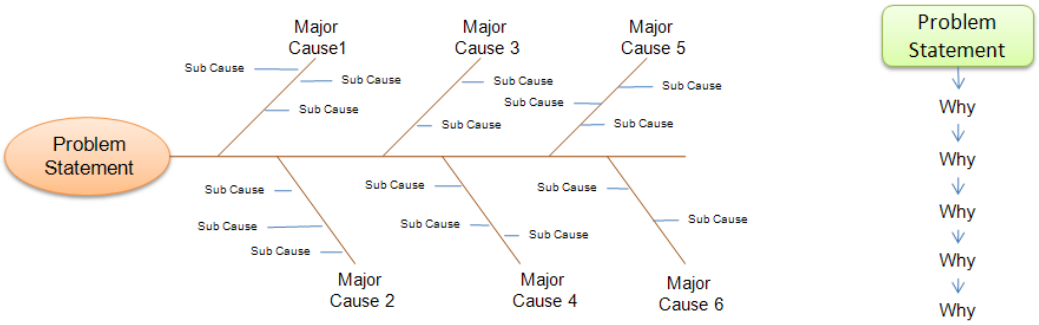
Here is a spreadsheet for tracking observation, underlying hypothesis and the results of validation.

	A	B	C	D	E
1	Observations	Hypothesis	Validation Responsibility	Target Date	Validation result
2					
3					
4					
5					
6					
7					

Step 4 : Problem Statement & Causal Analysis

Once we have validated the hypothesis and drawn conclusions about hypothesis, we are much nearer to understanidng our data. Believe it or not, we have already started our journey towards converting data into actions to influence the present scenarios. The key activity in making actionable analysis is to define problem statements and making them visible. The problem statements can be as simple as "projects are not logging defects" or "XXX type of projects having issues with their testing". Defining the problem effectively is as good as solving it half way through. Once the problem statement is visible, it will draw behaviors towards solving it.

Ishikawa / Fish bone diagram is an effective tool to brainstorm and co-relate reasons for a problem. Other tools like 5 Y can also be used. The intent of entire exercise if to reach as near to an "***actionable cause or group of causes"*** as possible. Most of the imrovements don't see a day light because the cause and actions there after are defined at a vague level where result focussed actions can't be drawn.



Step 5 : Improvement Strategy & Action Planning

Once we know the key causes which are generating problems, the next step would be:

- ✓ Identify action items
- ✓ Prioritize action items
- ✓ Identifying action owner
- ✓ Create and execute action & result monitoring plan.

Iterative action planning is really helpful to achieve targets. Typically, many larger improvement plans don't see the light of completion due to their size. By the time we reach some milestone, it's already too long on time scale and other priorities take over. It's good to attack some low hanging fruits first to make momentum around the improvements.

However, if it looks like that a big change is the only solution, ensuring that it's sponsored and managed appropriately would help. It's helpful to set smaller targets and achieve them. Once action plan is in place, monitoring it for its completion takes care of closing the loop.

Key points

Overall, here are key points from the above 5 stepped approach –

- Stratify data and understand what are various sectors which your data is giving signal on
- Study trends and differentiate natural variations from "signals"
- State observations clearly and build consensus around them
- Build hypothesis, clarify responsibility of validation and validate
- Conclude on the possible reasons for data stories
- Perform causal analysis on filtered possible reasons
- Plan and execute actions iteratively

References

- *Understanding Variation, The Key To Managing Chaos – by Dr. Donald J Wheeler*
- *Making Sense of Data, SPC for the service sector – by Dr. Donald J. Wheeler*
- *Metrics And Models In Software Quality Engineering – by Stephen H Kan, Forwarded by Capers Jones*
- *CMMi for Development V1.2 (http://www.sei.cmu.edu/cmmi/tools/dev/download.cfm)*
- *ITIL Version 3, Continual Service Improvement, OGC*
- *The Certified Six Sigma Black Belt Handbook, ASQ*
- *Managing the Software Process, Addison-Wesley, Watts Humphrey*
- *Practical Implementation of Software Metrics - Paul Goodman*