Notes:

Analyze

Key Learning Points

1. Explain how to identify theories of causes that lead to the problem.
2. Describe how to determine which theories might be root causes.
3. Identify root causes.

The DMAIC Methodology

DMAIC Steps	Tools Used
Analyze Steps: Analyze and determine the root cause(s) of the defects.	
Theorize Possible Causes (the Xs in Y=f(x) Identify Sources of Variation Determine Root Causes (the Vital Few x's in Y=f(x) Identify Value/Non-Value Added Process Steps For high risk and larger projects: Validate the Measurement System Determine Process Capability and Sigma Baseline	Note: Tools in Analyze Will Vary Depending on the Y Cause and Effect/Fishbone Diagram 5 Whys Process Map Analysis Histogram Pareto Chart Time Series/Run Chart Graphs and Charts For high risk and larger projects: Scatter Plot Regression Analysis Statistical Analysis Hypothesis Testing (Continuous and Discrete) Non-Normal Data Analysis Measurement System Analysis Gage R&R Process Sigma Calculation

Purpose of the Step

Analyze is the third step of DMAIC.

During the Analyze step improvement teams are tasked with studying the potential Xs, and determining which one(s) are the root cause(s), or causing the most variation of the project Y.

Understanding Root Cause

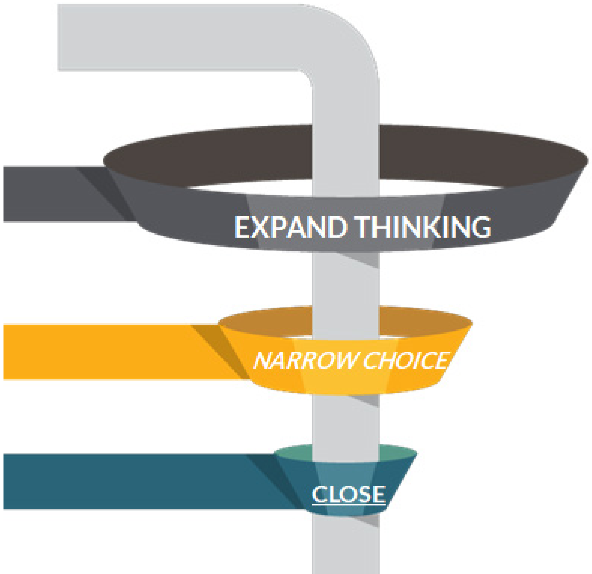
To understand root cause you need information. Information comes from data that must be collected and analyzed, and then conclusions can be reached. Various data analysis tools are helpful in this step.

Before you collect data you must first brainstorm theories of potential cause.

A theory is an unproved assertion of cause. The team has the responsibility to prove which theories are truly the root causes before solutions are implemented.

Notes:

How To Analyze The Xs



Brainstorm Theories (Xs in Y=f(x)

A theory is simply an unproven statement on the cause of a certain condition. No physician would diagnose influenza solely because of a patient's fever. The physician would consider a number of theories based on a range of symptoms. In the same way, when determining the cause of a quality problem, a project team must speculate about its many possible causes. Jumping to conclusions before considering many theories and proving which are correct would be risky. It would be just as risky as diagnosing an illness before considering all the symptoms.

To formulate theories a team must:

1. Identify the many possible causes of the problem before drawing conclusions about the root cause (brainstorming is useful for this activity).
2. Analyze the process by reviewing the process map for value-added and non-value added activities.
3. Use a cause-effect diagram to organize theories and develop new theories.
4. Use an impact control matrix to prioritize the theories to be tested.

Prioritize Theories to Test

Once many theories are formulated the improvement team needs to narrow down their list . Pareto Analysis is a useful tool for this task. Once the "vital few" theories are known, the team must prepare to test them.

These narrowed down theories are the team's hypotheses on what is causing the problem.

Prove That Vital Few Theories are Root Causes

After potential theories are identified and prioritized using the Impact Control

Notes:

Matrix, the next step is to test the theories, determining whether they are real root causes of the problem. Items that are placed into an impact control matrix are based on perception and opinion of the team. Further analysis is needed, with collection of actual data, for the team to reject or fail to reject the theories, which are hypothesized.

To test theories, a team must:

- Decide which theories to test
- Plan for data collection
- Collect data
- Analyze the results

Brainstorm Theories

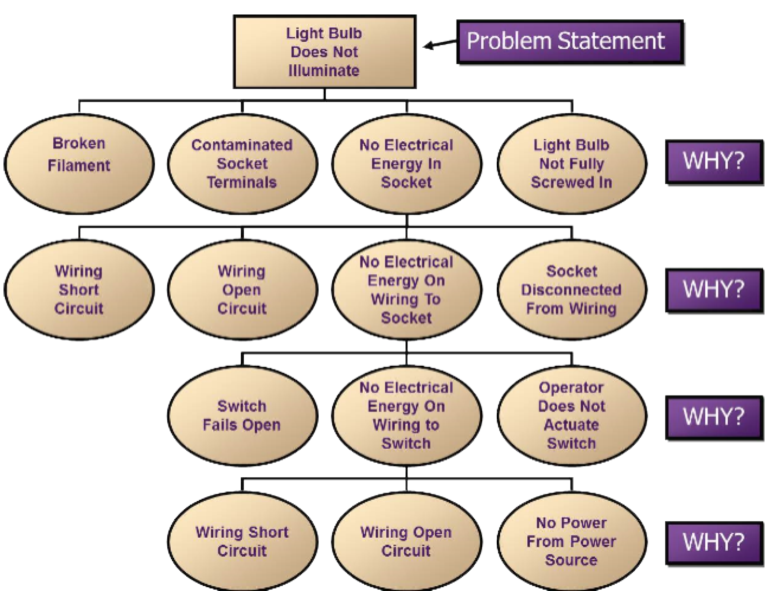
A theory is an assumption of what causes the Y. The team must expand its thought process by generating as many theories as possible using brainstorming and cause & effect tools.

To brainstorm theories the team can use the following tools:

5-Why Analysis

Sometimes it is necessary to dig deeper into causes to get from surface level causes to root causes that can be removed or controlled. One tool that helps to do this is 5-WHY Analysis. This tool is a qualitative tool, not a quantitative tool, that helps simplify causes of problems.

The challenge is to find causes for those current circumstances for which effective solutions can be generated to prevent reoccurrence. Vague recommendations such as “improve adherence to written policy and procedures” probably mean a basic root cause has not been found.

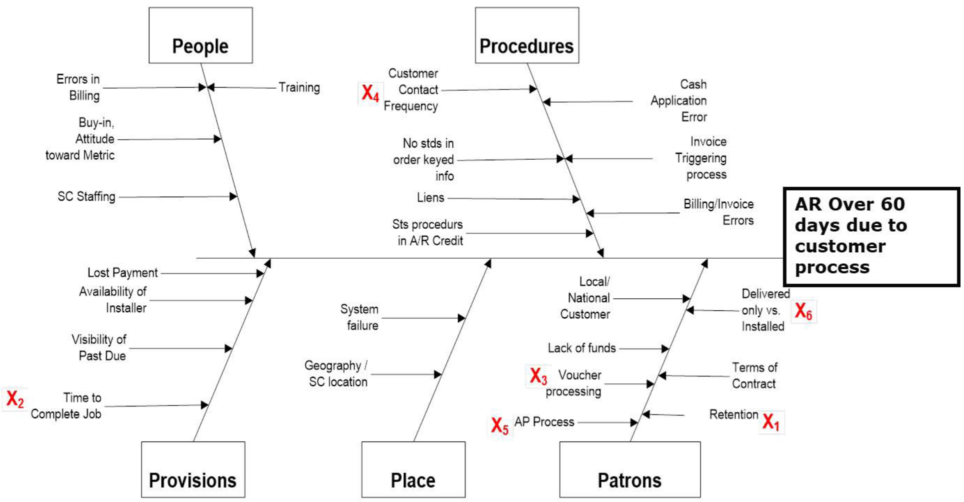


Notes:

Cause & Effect Analysis

Cause and Effect Analysis uses cause-effect diagrams to organize theories for systematic review. In addition, the diagram often challenges team members to come up with new theories by asking “Why?” for each factor they list.

The main purpose of a C-E diagram is to suggest theories of root causes (possible vital few Xs) and help the team focus on the possible Xs. Possible sources of cause are at the outmost point of the main branches. Detailed theories are listed off of that “spine.”



Failure Mode Effect Analysis

Failure Mode and Effect Analysis (FMEA) is a systematic method for identifying possible failures that pose the greatest overall risk for the process, product, or service. It depends on identifying a failure mode, the effect of a failure, the cause of a failure, and analysis of the failure mode.

When a failure mode is identified, the related data is analyzed, and the following factors are quantified:

- Severity
- Occurrence
- Detection

At this point a Risk Priority number is calculated, and the failure modes are ranked.

Notes:

Process Step or Variable or Key Input	Potential Failure Mode	Potential Effect on Customer Because of Defect	SEV	Potential Causes	OCC	Current Process Controls	DET	RPN
1. Customer Application	Application being filled out incorrectly	Application has to be resubmitted	8	Difficult to understand instructions	6	Check of application form for correct information by data entry operator	2	96
2. Data Entry	Data entered incorrectly	Customer receives checks with printing errors	4	Data entry error within a single field	6	None in place	10	240
3. Data Entry	Data entered incorrectly	Customer receives checks with printing errors	4	Information entered in wrong field	4	Self inspection	5	80

Brainstorming

Brainstorming is an idea-generating technique that you will use frequently throughout your project. This group technique is useful for generating new ideas and promoting unconventional, creative thinking.

In this step Brainstorming can be used to identify all stakeholders of this project. It will increase the probability that a team will identify all customers known, and unknown, and their voices, concepts, and solutions not possible with individual work.

Rules for Brainstorming

- No discussion of ideas until brainstorming is over.
- No judgment—no idea can be criticized.
- Unconventional, imaginative, or even outrageous ideas encouraged.
- Aim for a large number of new ideas in the shortest possible time.
- "Hitchhike" on other ideas by expanding them, modifying them, or producing new ones by association.

Prioritize Theories

After theories are identified, you must test them, and determine whether they are the real root causes of the problem.

To test theories, at team must:

- Prioritize which theories to test
- Plan for data collection
- Collect data
- Analyze the results
- Prove cause

Impact Control Matrix

Teams can find it challenging to prioritize where to focus when many items are brainstormed. The impact control matrix is a simple prioritization tool that talks about degree of control on one axis and degree of impact on the other axis. You can

Notes:

use it for various purposes; one common use is for shortening your list of root causes. You can also use it to prioritize your root causes.

		Potential IMPACT on Bad Copies		
		High	Medium	Low
Ability of team to make and CONTROL changes	High	* Printer heads are not aligned * Printer Settings are not optimal * Ink being used is old and expired	* Printer paper is old	* Lighting in printing room is poor
	Medium	* Storage location of supplies is moldy and damp, causing damage to the materials.	* Drying time of new print is too short	
	Low	* Age of machines contribute to poor prints	* Original documets to be printed are poor	* Hand dirtiness contributes to poor prints

Plan to Test Theories

For each theory (or group of theories), list all data required, each tool to be used, the results that will support the theory, and results that will rule out the theory.

Before beginning to test theories, the team should be very clear on exactly which theories to test. There are three general strategies to testing theories:

- One theory at a time
- Groups of theories
- All theories at the same time

Theory or group to be tested	Data required	Tool to be applied	Results that will support theory	Results that will rule out theory

Why Data Analysis?

A common question is why is data analysis needed at all?

Analyzing data will lead the team to gather information from the common patterns of failure in your process. It can also reveal causes of variation otherwise overlooked in day to day operations.

In addition to analyzing data, it is helpful to analyze detailed process maps to

Notes:

Notes:

highlight delays, redundancies and conditions causing variation.

Data Analysis Begins By Asking Questions

Learning to “ask the right questions” is the key skill in effective data collection. Accurate data, collected through an elaborately designed statistical sampling plan, is useless if it does not clearly address the problem being solved.

Anytime information about your process is needed, you should develop a project data-collection plan and validate the measurement system before beginning to collect data.

Questions to Be Answered

For this example, the team asked:

- What is the level of customer shortages?
- Does the type of product impact the number of shortages?
- Is there a difference in the shortages by reporting method?

Measure	Definition	Who	Where	Quantity
System reports of shortages	No charge orders with a reason code of “shortage”	System	All orders at all plants	All shortages
Truck audits	Random receiving audits conducted at selected customer sites.	Auditor	Various sites	Small random tests
St. Louis shortage report	A monthly report of shortages reported	Inventory Specialist	St. Louis	All shortages

Data Collection Plan

The Analyze Data Collection Plan is a helpful tool to focus and organize the steps required in collecting data to test the potential Xs. It is similar to the Measure Data Collection Plan, but is more focused on gathering information for statistical analysis. Your team should take the time to evaluate each potential X and determine the data necessary to prove the theory true or false.

The data can either be collected from sources that already exist in the organization, such as worker logs, or stored data from computer databases, or new data can be manually collected. To manually collect data, adjustments within your process may be required for a short period of time while team members collect the data. An example of this may be measuring the number of customers that abandon checkout (that is do not make a purchase) because of long lines. This may require having a team member set up in an obscure observation location during peak business hours. After determining the data types and sources required for analyzing each of the Xs, the team must decide on the size of (sample size) each data type and source.

Data Collection Plan Example

Data Collection Plan for the Analyze Phase (Example)

Ref.	Theories To Be Tested (Selected From The C-E Diagram, and/or FMEA)	List Of Questions That Must Be Answered To Test Each Selected Theory	Where Applicable, State The Null and Alternative Hypotheses		Tools To Be Used	Data To Be Collected					Remarks
			H_0	H_A		Description/Data Type	Sample Size, Number of Samples	Where To Collect Data	Who Will Collect Data	How Will Data Be Recorded	
1	Customer Application filled out incorrectly. The instructions are difficult to understand.	Are instructions difficult to understand?	Instructions not rated more difficult than prior forms Stat: Median Difficulty Score Observed = Standard Difficulty Score	Instructions are more difficult to understand. Stat: Median Difficulty Score is LT Standard Score	Survey: questions on difficulty with instructions. Stat: 1 Sample Sign, compare to standard	Survey scores on 7 point scale X: Categorical	30 surveys	Post application, survey call	Survey team	Survey data sheet entry	
		Are instructions difficult to understand?			Affinity: survey verbatims Graph: Pareto of Counts of Affinity Group	Verbal responses. Counts on each Affinity Group	30 surveys	Post application, survey call	Survey team	Survey data sheet entry	
	Customer Application filled out incorrectly. The application is too long.	Is the length of the application too long? Opinion scores	The application length is not a factor creating errors Stat: Median Survey Score Short form = Long form	Application length is an error factor. Stat: Median Survey Score Long form are GT Short form.	Survey: length of application opinion as survey score. Stat: Mann-Whitney (2 Medians)	Survey scores on 7 point scale. X: Categorical	30 surveys for long and short scores.	Post application, survey call	Survey team	Survey data sheet entry	
		Is the length of application too long? Error scores	The application length is not a factor creating errors Stat: Median Errors Short form = Long form	Application length is an error factor. Stat: Median Errors Long form are GT Short form.	Counts of errors for short form versus long form. X: Categorical. Stat: Mann-Whitney (2 Medians)	Errors counts long vs short form. X: Categorical	100 application samples for each of long and short forms. Last calendar month.	Post application, survey call	Analytics team	Query report	
2		Is the length of application is too long? Verbatim issues			Affinity: survey verbatims. Graph: Pareto of Counts of Affinity Groups	Verbal responses	30 surveys	Post application, survey call	Survey team	Survey data sheet entry	
		Is the length of application is too long?	The application length is not a factor creating errors. Mean errors short = mean long form	Application length is an error factor. Mean errors short are LT long	Stat: 2 sample T of survey scores, short version vs long. Graph: Box Plots, Histogram of errors by position in question order.	Survey scores on 7 point scale X: Categorical	100 surveys each for short and long form.	Post application, survey call	Survey team	Survey data sheet entry	
3	Entering data into some fields is error prone	Which fields have the most errors?	All fields are equally likely to have an error. Mean largest = mean smallest.	There are error prone fields. Mean largest GT mean smallest.	Scatter plot of errors by question order - trend?. Pareto of error counts by field	Counts of errors by field	100 application sample	Sample from last months applications	Iva Gothswun	Data sheet	

Sampling

Sampling is the process of collecting a portion or subset of the total data that may be available to find enough valid "information" to make a decision. All of the data is often referred to as a population or a universe. The purpose of sampling is to draw conclusions about the population using the sample. This is known as statistical inference.

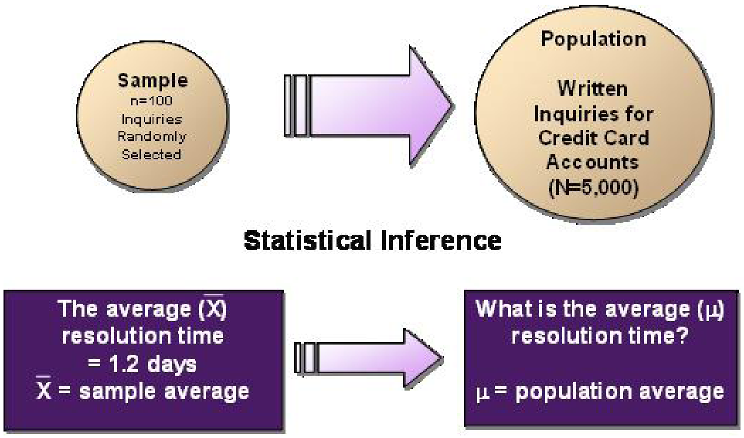
Sample size is determined by establishing a confidence level (usually 95%) you would like to have, deciding on the degree of precision wanted in the value that will be inferred (e.g., population average), and an estimate of the variation in the population.

Sampling is used when:

Notes:

- Collecting all of the data is too costly
- Full data collection may be a destructive process
- Sound conclusions can be made from a relatively small amount of data

Sampling Example



Analysis Model

There are an abundance of tools that can be used when gathering and analyzing data. The guiding principle is to get concrete data on the causes. The following model regarding types of analysis may be a helpful guide to learning how much data needs to be gathered and what type of analysis needs to be utilized.

Individual Experience

Sometimes, in looking at a problem, you may realize you have seen this exact problem before. You apply the knowledge from previous occurrences to the specific problem. More often, prior experience will be similar, but not identical, so your experience will offer good theories for further testing.

Group Experience

When the team looks at the problem, different members view the issue from different angles. Linking the individual experiences together forms a more complete picture of what the cause might be, making the search for cause more robust and providing a basis for prioritizing theories for testing.

Graphical Interpretation of Observed Data

Group experience may develop theories of where to look to solve the problem but objective facts are needed to solve the problem and convince others. Graphing data from the current state will help identify the most likely theories to be tested more rigorously and suggest new theories. Graphs can also help the analyst avoid misinterpreting the statistical results. Graphic analysis may identify the answer if there are no confounding factors.

Notes:

Statistical Interpretation of Observed Data

Graphical interpretation of observed data is typically not sufficient to determine whether the patterns shown by the graph are strong enough to prove the true root cause, and not just random variation. Statistical analysis of the current state provides a scientific basis to make that determination.

Graphical Interpretation of Experimental Data

Observed data from the current state can suffer from a number of weaknesses such as: truncation of potential ranges for some factors, absence of cases that capture interaction among some levels of factors, and noise from other variables that are not observed or controlled. To remove these problems a designed experiment is required. Graphical analysis of experimental results provides a qualitative guide to next steps, suggests new theories, and is a check to be sure there are not hidden patterns that confound the statistical results.

Statistical Interpretation of Experimental Data

Statistical analysis of an experiment is required to confirm the proof of causes, quantify the effects, and predict outcomes. It is also the basis for optimizing solutions by finding the best combinations of multiple factors.

Plot the Data

Quote Ellis Ott's three rules for data analysis:

Rule 1: Plot the Data
Rule 2: Plot the Data
Rule 3: Plot the Data

Histograms

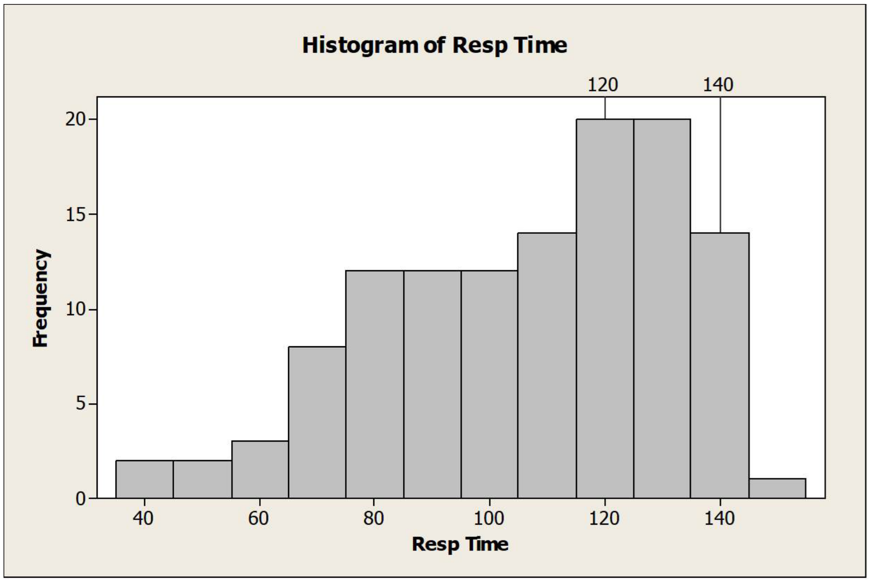
A histogram is a chart that uses bars to display variation in a single characteristic. Patterns in the variation often reveal new facts about the process. Patterns of variation are difficult to see in simple tables of numbers, histograms show these patterns.

Three Important aspects of histograms are:

- Its center
- Its width
- Its shape

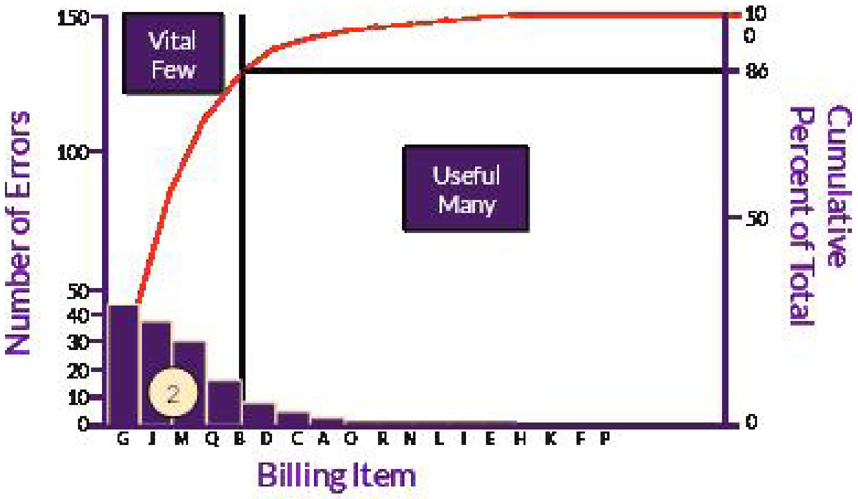
Notes:

Notes:



Pareto Diagrams

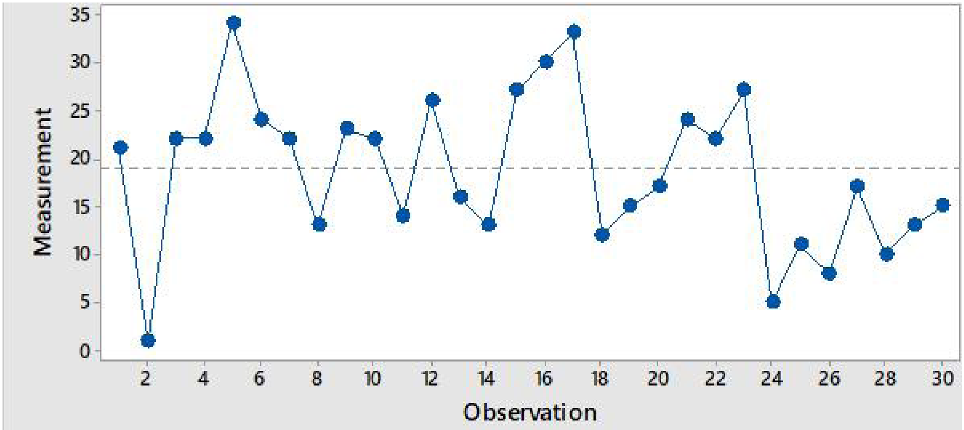
Pareto Diagrams display patterns that show the highest concentration of improvement potential in the fewest number of remedies. These offer the greatest potential gain for the least amount of managerial and investigative . You must focus your improvement efforts on these "vital few." If one way of stratifying the data does not yield a "vital few," re-stratify the data by other characteristics until it is clear what the vital few are. Pareto diagrams give a visual method for separating the "vital few" from the "useful many."



Run Charts

A run chart is a line graph of data plotted over time. By collecting and charting data over time, you can find trends or patterns in the process. Because they do not use control limits, run charts cannot tell you if a process is stable. However, they can show you how the process is running.

Notes:

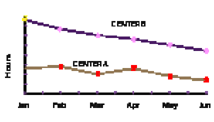
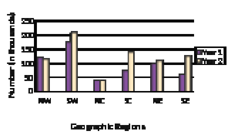
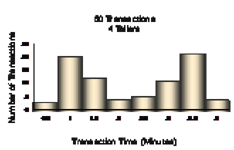
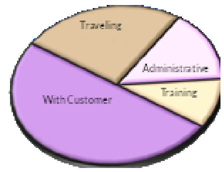


Graphs & Charts

Graphs and charts are pictorial representations of quantitative data. They can summarize large amounts of information in a small area and communicate complex situations concisely and clearly.

There are many types of graphs and charts. In a DMAIC project you will mostly use:

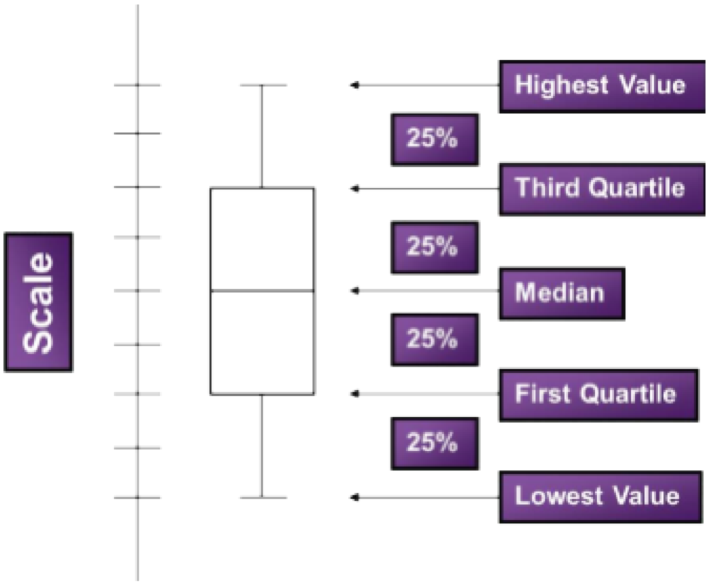
- Line Graphs
- Bar Graphs
- Stacked Bar Graphs
- Pie Charts
- Histograms

Box Plots

Box Plots provide a graphic summary of the pattern of variation in a set of data. They are especially useful when working with small sets of data.

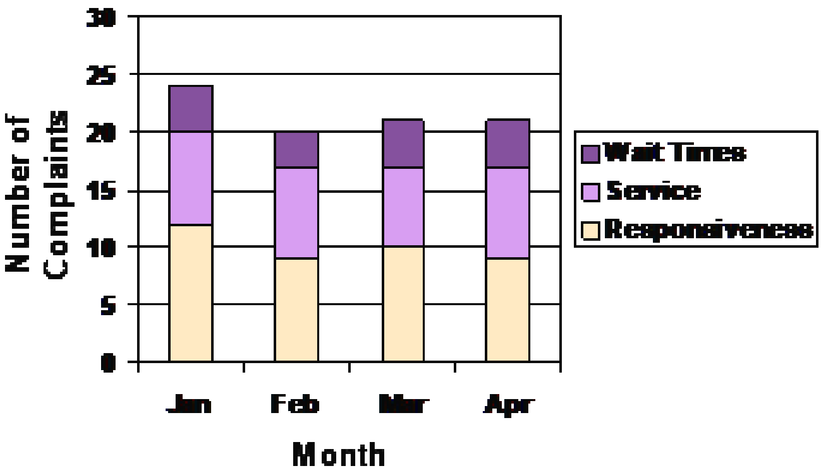
Notes:



Stratification

Stratification is the basis for other tools like Pareto Analysis, and it is used in conjunction with other tools such as Scatter Diagrams to make them more powerful.

Stratification is the separation of data based on specific characteristics or variables. It's most frequent use is during analysis of theories to identify which theories contribute to the problem being solved.



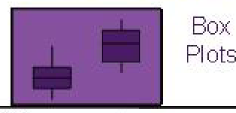
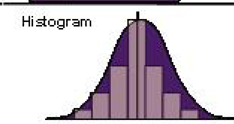
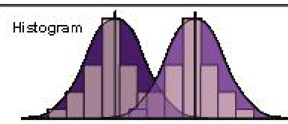
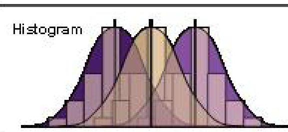
Advanced Analysis Tools

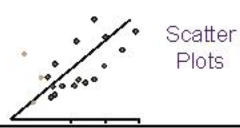
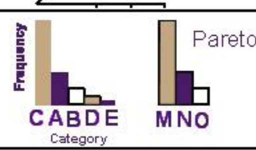
There are additional tools and concepts that are used less frequently but are nonetheless critical to high-risk problems. Most of these tools are related to analyzing variables data (measured data).

Some of these tools are:

- Confidence Intervals
- Hypothesis Testing
- Test of Equal Variances
- Analysis of Variance
- Nonparametric Tests

The Analysis Tool Depends on the Question and the Data Type

What The Tool Tests	Statistical Test	Graphical Test
Variance among 2 or more samples is different	Test of Equal Variance	
Mean of sample data is different from an established target	1-Sample t-test	
Mean of sample 1 is different from mean of sample 2	2-Sample t-test	
The means of 2 or more samples are different	1-Way ANOVA	

What the Tool Tests	Statistical Test	Graphical Test
Output (Y) changes as the input (X) changes	Linear Regression	
Output counts from two or more subgroups differ	Contingency Tables (Chi-Square Test of Independence)	

Reducing Process Waste

After identifying the root cause, it is important to re-examine the existing process to eliminate any waste or non-value added steps, or cost of poor quality.

To do this, there are four simple steps for analyzing process maps.

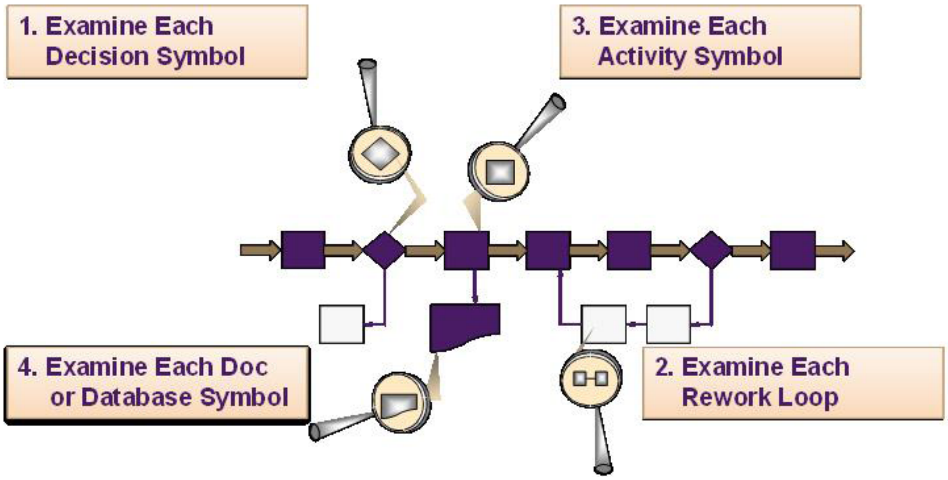
- Examine Each Decision Symbol.
- Examine Each Rework Loop.

Notes:

Notes:

- Examine Each Activity Symbol.
- Examine Each Document or Database Symbol.

Analyzing The Process



A Final Word About Root Causes

There are two questions that will help you decide whether you have found the root cause:

1. Does the data suggest any other possible causes? After data collection and analysis, it is usually possible to discard some theories and place more confidence in others. Theorizing is not a one-time activity, however. If you have competing plausible theories that are consistent with the new data and cannot be discarded based on other data, then you have not yet arrived at the root cause.

2. Is the proposed root cause controllable in some way? Some causes are beyond your ability to control, like the weather. Other possible causes are too broad and general to control and need to be broken into components. For example, "lack of training" as a cause needs further definition of the specific skill or knowledge that is missing.