Data Analysis Tools

Tool	What is it?	When do I use it?	Example	DMAIC
Affinity Diagram	A method to gather and organize brainstormed ideas (language data) into groupings based on the natural relationship between items.	When you need to organize and consolidate large amounts of qualitative data in order to support a concept or solution.	http://en.wikipedia.org/wiki/Affinity_diagram	Measure, **Analyze**, Improve, Control
Box and Whisker Plot	A graphical way to compare the medians and the variation between groups of data. It can also help to identify outliers.	When you need a quick visual look at one or more sets of data. They graphically show different types of populations without any assumptions of the statistical distribution.	Discovering Stats 3e – pg.173-176	Measure, **Analyze**, Improve
Cause and Effect / Fishbone / Ishikawa Diagram	A structured problem solving technique that graphically displays / organize all possible brainstormed causes relating to a problem (focusing on causes, not symptoms).	When you have a large number of factors that could influence your process and you need support for a resulting solution.	Slides posted in Coursework	**Analyze**
Check Sheets	Used for counting and accumulating data in a logical format during observation. This is a straightforward and easy way to answer the question, how often are certain events happening?	When quantifying frequency / counting data on number of occurrences (eg. defects), information on variables (eg. weight, size, length, or defect location, etc.).	http://web2.concordia.ca/Quality/tools/6cksheet.pdf	**Measure**
Chi-Square – Test for Independence	This is a procedure used to determine if two classifications variables are related, testing the statistical independence of two random variables. It compares the number of observed counts against the expected number of counts to determine if there is a difference in output counts.	It is a good method to apply to before and after data to prove that a process improvement made an effective change. Use this method when you have nominal data in a table and you need to know if the output counts differ for two or more categories.	Discovering Stats 3e – pg. 646-651	**Analyze**, Improve
Confidence Intervals	A range of values in which we	Confidence intervals determine if a	Discovering Stats 3e – pg.	Measure, **Analyze**,

	expect the population parameter to occur. A population parameter is likely to occur at a specified probability. It is constructed from sample data.	process is centered where it is expected. They are used to identify a shift or change in the process (mean) and to identify a difference in two populations (eg. does Vendor 1 and Vendor 2 give us the same dimensioned part?)	426-484	Improve
Control Chart	Can serve as a statistical tool for problem identification as well as, ongoing monitoring of a process (graphically) over time. It can assist in distinguishing random variation (noise) from assignable variation (signal).	When you need to recognize and eliminate sources of variation in a process so that a process performs consistently and predictably.	*Understanding Variation* by D.J. Wheeler	Measure, Analyze, Improve, **Control**
Correlation Coefficient	Correlation coefficient **R** is a measure of the strength and direction of the linear relationship between 2 random variables. R falls on or between the numbers -1 and 1. Coefficient of determination R^2 is a measure representing the percent of variability in "y" that can be accounted for by the variable x.	Although correlation does not explicitly imply causation, establishing a correlation between two variables is necessary (*but not sufficient)* to establishing a causal relationship.	Discovering Stats 3e – pg. 192-199	Measure, **Analyze**, Improve
Fault Tree Analysis	It is a visual tool that logically and graphically presents the various combinations of possible events or failures that could occur in a process or product. It is an inverted tree with the trunk (noting the failure or undesired event) of the tree at the top of the diagram and the branches are the contributing causes of the failure.	When a simple visual is needed for determining (and presenting) the root cause of a failure.	http://en.wikipedia.org/wiki/Fault_tree_analysis http://www.weibull.com/basics/fault-tree/	Measure, **Analyze**, Improve
Force Field Analysis	Method to present the "positives" and "negatives" of a situation so that they can be compared, the positives can be reinforced and the negatives can be eliminated.	When the desired outcome is "making a change" in the midst of forces /barriers restraining movement towards the ideal state.	http://www.mindtools.com/pages/article/newTED_06.htm	Measure, **Analyze**, Improve
Histogram	A graphical method of displaying the distribution of data by bar graphing the number of units of a	When displaying large amounts of data that are difficult to interpret in tabular form.	Discovering Stats 3e – pg. 65-66	**Measure**, Analyze, Improve

	particular category (illustrates process centering, spread and shape).			
Hypothesis Testing	Hypothesis testing is a process (inferential method) that uses sample data from a population to confirm or refute some statement or claim about that population.	Hypothesis testing can tell us if two sets of data are really different from each other. Determines statistically whether or not there is a cause for concern or if our conclusion is simply due to random variation. It can be used to determine whether a population parameter (mean, variation, etc.) is statistically different than a standard or set value.	Discovering Stats 3e – pg. 486-555, 574-629	**Analyze**, Improve
Linear Regression Analysis	This tool will help you to understand the relationship between the process output and any process input that could affect it. It is a way to model or predict the relationship between those variables.	When you suspect there is a relationship between an input and output variable. It is especially useful when the output variable is difficult or expensive to measure and the input variables are not.	Discovering Stats 3e – pg.186-237, 743-759	**Analyze**
Measurement System Analysis (MSA)	A process that includes identifying, quantifying and reducing measurement errors. Measurement systems (a decision making tool) play a large part in process improvement activities.	On every measurement system that a decision is based upon.	http://www.6sigma.us/MeasurementSystemsMSA/measurement-systems-analysis-MSA-p1.html Slides posted in Coursework	**Measure,** Analyze
Measures of Center (Descriptive Statistics - Measure of Location)	The behavior of the middle (or central portion) of the population of process data. 3 measures are: Mean = arithmetic average Median = middle value Mode = most frequent value	When you need a quantitative measure that summarizes an important characteristic of a population / process, the center of your data.	Discovering Stats 3e – pg.108-117	Define, **Measure**, Analyze, Improve, Control
Measures of Variability (Descriptive Statistics - Measure of Dispersion or spread)	A measure of how the data is spread around the mean. 3 Measures are: Range = difference between the	On any set of data - all populations and processes have some degree of variability.	Discovering Stats 3e – pg.126-137	Define, **Measure**, Analyze, Improve,

				Control
Pareto Chart/Diagram	A descending bar and cumulatively increasing line chart used to separate the vital few from the trivial many. The vital few are the few factors accounting for the largest part (%) of a problem or condition. Pareto Principle: 20% of the sources cause 80% of the problem.	When you need to focus on the key problem(s) - when solved will have the greatest impact.	Discovering Stats 3e – pg. 43-44	Define, Measure, **Analyze**, Improve
Process Map (or Process Flow chart)	A graphical tool for documenting a process. Each step or activity is mapped out as it occurs in the real-live process.	When improving or creating a process	http://en.wikipedia.org/wiki/Business_process_mapping	**Define**, Measure Analyze, Improve, Control
Run Chart (Time Series Plot)	A graphical tool that can show and track trends or patterns over a specified time period.	When you need to do the simplest possible display of a trend over time.	Discovering Stats 3e – pg.89-91	**Measure**, Analyze, Improve
Scatter Plot/Diagram	A graphical tool to visualize the possible relationship between two variables and relative strength of that relationship.	When you need to display what happens to one variable when another variable changes (visualize a relationship between two variables).	Discovering Stats 3e – pg.188-192	**Measure**, Analyze, Improve
SIPOC (Supplier / Input / Process / Output / Customer)	A simple way to document (or map) a process in the "as is" (current) state by listing the suppliers, inputs, outputs and customers. The "process" should be identified by a high level flow chart. This sets the scope for the value-stream map.	This is a quick way to document your process (and start to analyze) when parts of the process are not clear or consistent (ie. Who supplies inputs to the process? Who are the true customers of the process? Is there a customer for each output? What are the requirements of customers?)	Slides posted in Coursework http://www.isixsigma.com/library/content/c010429a.asp	**Define**, Measure

The first row cell (left column above Pareto) reads:

largest and the smallest data point
Standard Deviation = measure takes into account each data point and its distance from the mean
Variance = standard deviation squared

Stem and Leaf plot (Stemplot)	A graphical technique that shows the shape (distribution) of the data like a histogram but displays all of the individual values within an interval rather than just the frequency for each interval.	When evaluating the shape of the data with the ability to maintain visibility to the original raw data points.	Discovering Stats 3e – pg.68-71	**Measure**, Analyze, Improve
Solution Selection Matrix	A matrix that helps identify the best solution among several solutions by weighting the impact of each solution on established criteria (ie. cost, time to implement, impact on quality, etc.), hence measuring the effectiveness of solving the problem with that solution.	Use this when you have a couple solutions to choose from and want to make your decision fact based, taking more factors into consideration.	http://www.whatissixsigma.net/six-sigma-dmaic-improve-phase/	**Improve**
Thought Process Map (TPM)	A logical, visual representation or map of someone's thought process flow (of questions, tools used to help answer the question, related actions and related decisions) that shows how a process or problem was/is attacked / addressed.	Use this for any situation. This is an "evergreen" document that can be used as a communication tool for where you are, where you've been and where you are going.	Slides posted in Coursework	**Define**, Control