# Simple Data Analysis Techniques

Three of the most common charts used for data analysis are pie, Pareto and trend charts. These are often linked together in a data trail.

Pie Charts

Pie charts provide a simple and very visual picture of the relative proportions of performance in percentage and value by cumulating data over a period of time.
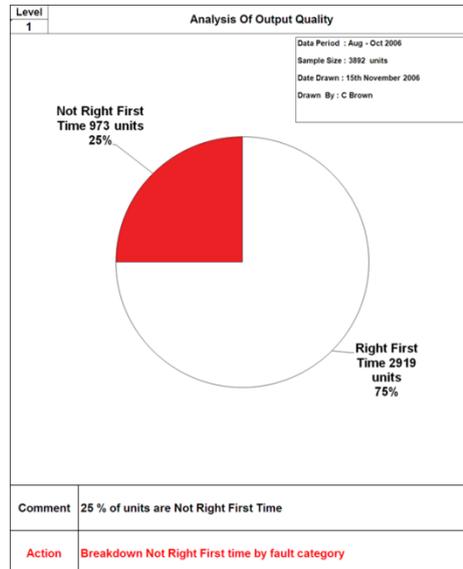
Pie charts are used to highlight the overall current situation and focus attention on the need to improve. They should have a maximum of three segments to keep the picture simple.

Pie charts allow a direct comparison before and after an improvement activity but not the ongoing trend of performance. They have no target line.

Method to Create a Pie Chart

|	
P	Determine the title for the pie chart and write it in the title box. Fill in the level number.
D	Calculate the value of each category as a percentage of the sum of all the categories. Organise the categories in descending order, largest first. Draw a circle in the centre of the graph area, leaving room around the edge to write details of each of the segments. Starting from a line at the 12 O'clock position, draw the largest segment first working in a clockwise direction, then draw the next largest, etc. For each segment, write the name of the category, the category value and the category % underneath the value.

	Highlight only the segment causing concern in colour.
	Write the data period, total sample size, data drawn and drawn by in a data source box.
C	Summarise the graph findings in a sentence in the comment box. "Say what you see".
A	Make a statement of action required as a consequence of the comment in the action box in RED.

Pareto Chart

Pareto charts are the best method to identify the focus for improvement for complex problems. It ranks problems in descending order and shows the significance of each issue as a percentage of the whole. That allows concentration of limited resources on the biggest issue and so potentially the biggest improvement.

The percentage figure should never be omitted as it allows direct comparison before and after improvement activities, irrespective of the sample size.

Pareto charts have no target line.
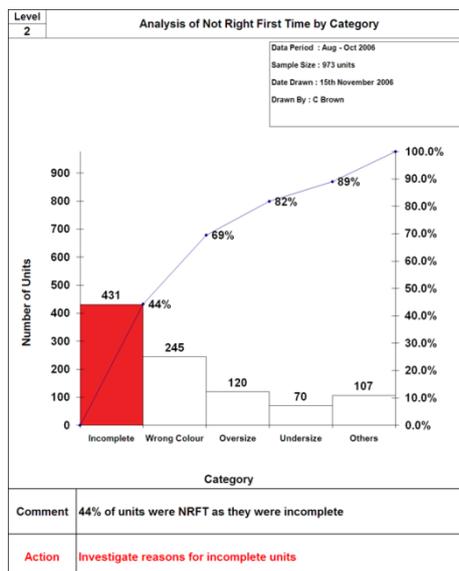
The name Pareto comes from Vilfredo Pareto, an Italian economist who, while studying the Italian economy in the 1890's, discovered that 80% of the wealth was owned by 20% of the people. In the 1950's Joseph Juran, a quality guru, found that this principle applied equally in industry.

Pareto analysis is often referred to as the 80:20 Rule and this is why we aim to plot the top 80% of category values.

Method to Create a Pareto Chart

P	Determine the title for the pie chart and write it in the title box. Fill in the level number.
D	Organise the categories in descending order, largest first. Accumulate any remaining categories after the seventh into one category called "other" and place last in the list. Calculate the value of each category as a percentage of the sum of all the categories. Calculate the cumulative percentage along side this list. Select and name the left hand axis. Make sure that the maximum value equals the sum of all categories. This is the same as the value of the category of concern highlighted on the level 1 pie chart. Scale and name the right hand percentage axis. Make sure that the 100% mark is opposite the maximum value on the left hand axis. Draw the bar for each category, name each category and write the category value above the bar. Always aim to plot the top 80% but plot no more than 8 categories in total. The "other" category always goes on the far right hand side. Plot the Pareto % line. Start at 0 and plot first point at the right hand corner of the 1st bar, align each subsequent point with the right hand side of each bar, write the cumulative % value to the right of each point, draw a line through the points. Highlight the category causing concern on the graph in colour. Write the data period, total sample size, date drawn and drawn by in the data source box.
C	Summarise the graph findings in a sentence in the comment box. "Say what you see".
A	Make a statement of action required as a consequence of the comment in the action box in RED.

Trend Or Run Charts

Trend charts are used to show performance against target over time and progress towards the organisation's targets. They are a good record of historical performance.

This is unlike the pie and Pareto charts that show cumulated performance over a period of time and which do not feature target lines.

Trend charts should show performance as a percentage figure where possible. Text bubbles should be used to indicate where improvement activities or problems have occurred stating major reasons for success or failure against target. Hand drawn bubbles on live charts are very powerful.

Trend charts can be used with control limits to highlight the difference between a minor process fluctuation and a true performance change.

Trend charts can encourage companies to continue capture of key areas for improvement activities.
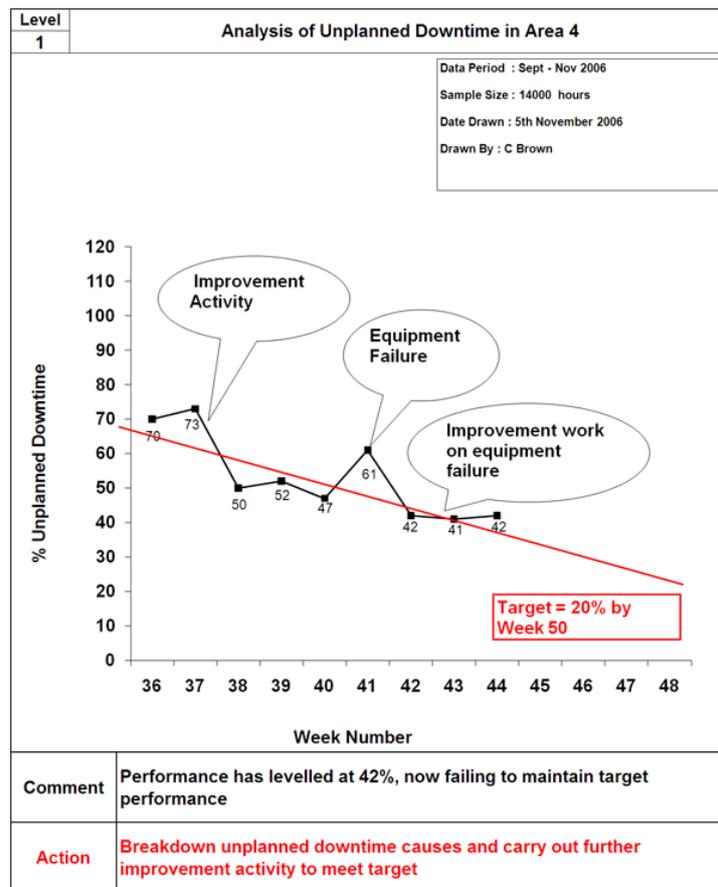
Data Trails

A data trail is used to break top level measures down in order to identify focus improvement actions, through the application of PDCA.

A data trail is a linked series of charts that show the successive breakdown of the data. Each chart concentrates solely on the breakdown of the largest problem of the previous chart.

Because of the data analysis method used, the final action should be focussed on the single largest issue. Tackling this issue with the finite resource available should give the greatest benefit.

The graphs most frequently used in a data trail are pie and Pareto. Using the standard format, that follows the PDCA approach, the action statement of one graph forms the title of the next level graph. On feedback the reader should be able to summarise the story of the data trail by reading only the title, comment and action boxes of each graph.

This method of analysis also assists with setting realistic targets for an improvement activity as the proportion of the issue being tackled, compared to the top level value, can be worked out.

Note that the sample size of the next level graph is the size of the focus issue of the proceeding graph.

Level 1	Analysis Of Quotes Return Times

Data Period : Aug to Oct 2001
Sample Size : 520 quotes
Date Drawn : 15th Nov 2001
Drawn By : C Brown

Late 156
30%

On Time 364
70%

Comment	30% of quotes were not on time
Action	Analyse reasons why they were late

Level 2	Analysis Of Reasons For Being Late

Data Period : Aug to Oct 2001
Sample Size : 156
Date Drawn : 15th Nov 2001
Drawn By : C Brown

Quantity

160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

80   51%   72%   87%   97%   100.0%
      32   23   16   5

Missing data   Computer breakdown   No paper   Absence   Others

Reason

Comment	51% of quotes are late due to missing data
Action	Analyse reasons for missing data

Level 3	Analysis of Missing Data

Data Period : Aug to Oct 2001
Sample Size : 80
Date Drawn : 15th Nov 2001
Drawn By : C Brown

Quantity

80   100.0%
70   90.0%
60   80.0%
50   70.0%
40   50%   81%   88%   60.0%
30   50.0%
25   40.0%
20   30.0%
10   20.0%
5   10   10.0%
0   0.0%

Bad design   Not signed   No drawing   Other

Reason

Comment	50% of missing data is due to bad design
Action	Redesign form