Pareto Charts
Distribution and Causal Analysis Tools

The Pareto chart is a specialized version of a histogram that ranks the categories in the chart from most frequent to least frequent. A Pareto Chart is useful for non-numeric data, such as "cause", "type", or "classification". This tool helps to prioritize where action and process changes should be focused. If one is trying to take action based upon causes of accidents or events, it is generally most helpful to focus efforts on the most frequent causes. Going after an "easy" yet infrequent cause will probably not reap benefits.

There actually was a person named "Pareto" who developed this chart as part of an analysis of economics data more than 100 years ago. He determined that a large portion of the economy was controlled by a small portion of the people within the economy. Likewise, one doing analysis of accidents or events may find that a large portion of the accidents are cause by a small population of causes. The "Pareto Principle" states that 80% of the problems come from 20% of the causes.

The Pareto Chart analysis should be performed over a fixed time interval of performance indicator results. The Pareto Chart analysis should only be performed after the control chart analysis is complete. The time interval should be chosen depending on the existence or non-existence of statistically significant trends as follows:

Trend Exists on Control Chart?	Data Points Used	Purpose
No	Use all data in the statistically stable time period	Find Common Cause(s) to apply to process improvement
Yes	Use only the data for the point(s) which have been identified as within the significant trend, such as a point outside the control limits	Find Special Cause(s) for basis of corrective actions for declining trends or reinforcing actions for improving trends

This consideration of significant trends is important as the frequency distribution in each category may be different due to the process changes which occurred to cause the significant trend. Thus the arbitrary choice of "Fiscal Year to Date" or "Calendar Year to Date" or "The past two years" may not be appropriate.

A Pareto Chart is generally shown as a vertical bar chart. A Pareto Chart is a special form of a histogram where the categories have been sorted from most frequent to least frequent. One would not want to sort the categories from most frequent to least frequent if there is a natural order to the categories, such as a distribution by age or cycle time.

Important Note:
If you are intending to use a bar chart or histogram to find trends, you are reading the wrong section of this primer. Go back to **Guidelines for Statistical Process Control** *if*

you are trying to find trends. Histograms and Pareto Charts should only be made after you have completed the initial trend analysis and are trying to determine what to do to improve the process. Although the histogram and Pareto Charts are useful for analyzing process data, the time sequence that the process data occurred in is lost in this analysis. The control chart maintains the time sequence of the data.

Segregating Data into Categories

In order to perform a Pareto Chart, one must have a consistent manner of classifying the data into categories. Your data system may already have defined categories. The Department of Energy Occurrence Reporting and Processing System (ORPS) has fields such as "Nature of Occurrence", "Root Cause", and "Direct Cause" which make good bases for Pareto Charts. Occupational Illness and Injury data has defined OSHA categories for "Body Part", "Injury", and "Cause".

If your data does not have pre-defined classifications, you will need to set up your own system. You will need to define the overall type of categorization (such as subject, cause, or location) and then determine standard phrases for each category. Be careful to remain consistent in defining the scope of each category. You should also decide whether an item being classified will be allowed to be placed under more than one category. For example, the ORPS database allows only one Root Cause to be assigned to an occurrence, but up to three Contributing Causes.
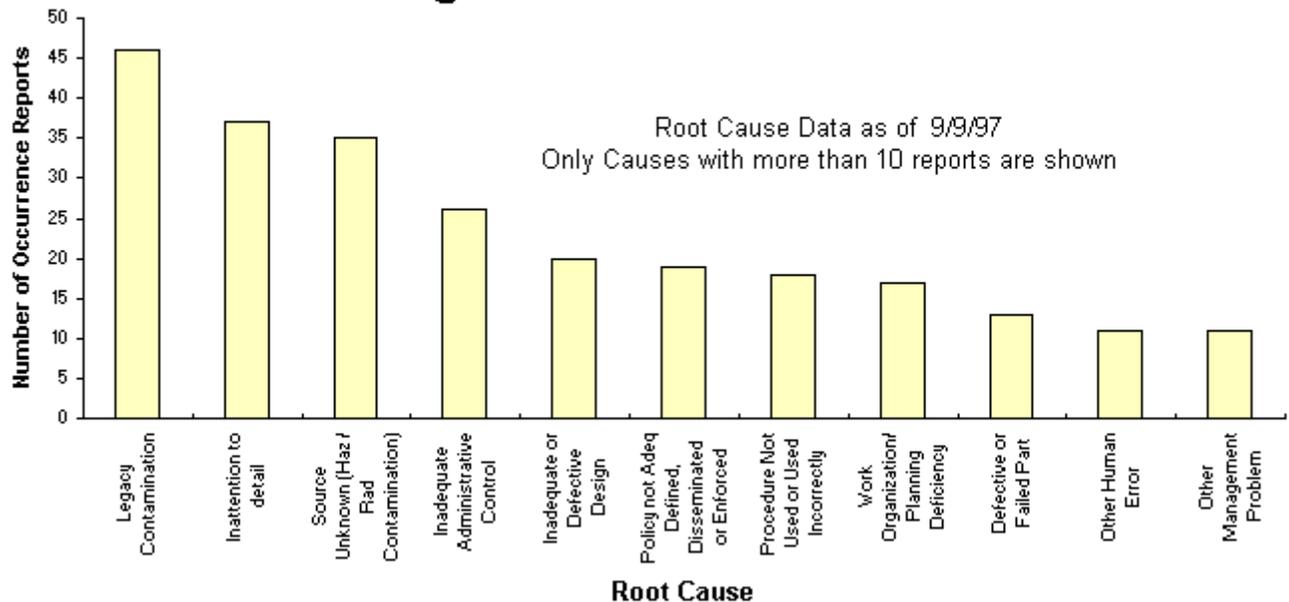
Bad Example of Categorization (real life example)

A certain U.S. Navy weapons system was having failures in its exercise units. A Pareto analysis was performed on the cause of failure data to determine the primary source of the failures. "Personnel Error" on the part of maintenance personnel was one category. "Mechanical Failures" were subdivided into several categories on the basis of mechanical system. The resulting Pareto Chart showed that "Personnel Error" was the leading source of failure. However, there were many more mechanical failures, broken among many categories such that no single category exceeded the number of "Personnel Errors". Thus an inappropriate conclusion (the problem to focus on was personnel error) rather than mechanical failures of parts was reached.

Good Example of Pareto Chart Analysis

The ORPS database does contain several pre-defined fields for cause classification. The example chart below shows how the "Root Cause" data can be utilized for a Pareto chart to determine the leading causes of occurrences. Note that the time interval for the reports is stated on the graph, and has been verified to be a stable period for trends. This graph only shows those causes with more than ten reports for clarity purposes. Also, the number of reports with no root cause determined are not shown.

Occurrence Reports
Categorized 10/01/96 - 08/31/97

Number of Occurrence Reports (y-axis: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50)

Root Cause Data as of 9/9/97
Only Causes with more than 10 reports are shown

Root Cause (x-axis categories):
Legacy Contamination; Inattention to detail; Source Unknown (Haz / Rad Contamination); Inadequate Administrative Control; Inadequate or Defective Design; Policy not Adeq Defined, Disseminated or Enforced; Procedure Not Used or Used Incorrectly; Work Organization/ Planning Deficiency; Defective or Failed Part; Other Human Error; Other Management Problem

Root Cause

In addition, some analysts add a cumulative percentage line superimposed over the bars, so that the point at which the leading causes cover 80 to 85% of the reports can be exactly determined.

Other Bases for the Y-axis

Other bases may be used for the y-axis rather than simply counts of events. The total dollar cost all events in the category, the number of lost or restricted work days for each accident type, or other weighting schemes may be utilized. Avoid using average cost or average days per event in a Pareto chart. Keep the focus on total cost.

A bad example of a Pareto analysis was in the national press in 1997. News articles asserted that Carpal Tunnel Syndrome was the leading source of lost workdays in the country. One headline read **" Carpal Tunnel Tops Lost Work time Injury Report "** . However, looking at the original data for the study, one can note that the median number of days **per case** was used as the factor for the comparison. It is true that once a Carpal Tunnel case occurs, there is generally a high number of lost days for that case. However, Carpal Tunnel cases are relatively infrequent. There were more than twenty times more sprain injuries than Carpal Tunnel. By the **total** number of days for all cases, sprains (generally back injuries) had four times the total number of lost days that Carpal Tunnel cases had (based upon multiplying the number of cases by the median number of days per case). A company desiring to reduce its total number of lost work injuries or lost work days is much better off going after sprains which are very frequent (although involve less days per case).