

## Pareto Analysis

### What is it?

A type of bar chart used to distinguish the vital few areas of importance from the trivial many. The tool is named after Vilfredo Pareto. He was a 19<sup>th</sup> century economist who studied the distribution of wealth in Italy and found 20% of Italians possessed 80% of the wealth.

After World War II, Joseph Juran taught Pareto's principle as the 80/20 rule to Japanese and American managers.



**Quote:** The Pareto principle is employed to identify the "vital few," whether customers, customer needs, product features, process features, or inputs. Identification of the vital few helps to assure that resources and attention are concentrated where they will do the most good.

--from *Juran's Quality Control Handbook*, 4<sup>th</sup> Edition, by J.M. Juran and Frank M. Gryna, ©1988 by McGraw Hill, ISBN#0-07-033176-6.

A Pareto chart plots the number of times a particular event or cause occurs versus the total number of instances of that problem or category of events.

It can ensure that your cause analysis and corrective actions are focused on the vital few:

- ◆ Customer complaints -- Identify each complaint by category (e.g., late delivery, damaged shipment, billing error, etc.).
- ◆ Employee complaints -- Identify each employee concern by category (e.g., supervisor, compensation, lack of direction, training, tools/equipment, etc.).
- ◆ Causes of project delays -- Identify the number and type of each delay in a series of projects. This will help you address the most common causes.
- ◆ Reasons for manufacturing interruptions -- Identify anything that causes a slow down or stoppage of production in your key bottleneck or constraint (e.g., lack of materials, equipment breakdown, yield below optimum, rate below optimum, production of scrap, repairs/rework, etc.)

The Pareto chart enables you to test for the 80/20 rule, i.e., that few causes are often responsible for a great majority of effects. It helps you identify those vital few areas where your efforts will have the greatest impact.

As the table indicates, Pareto's 80/20 rule has wide applicability.

20% of...	Often account for 80% of...
Your products	Sales volume
Your customers	Profits
Your equipment	Breakdowns
Your products or services	Help desk requests
Your suppliers	Delays or supply problems
The carpets in your house	The wear
The clothes in your closet	What you wear
Operating system software	Computer processing time
Process steps	Problems or errors
Students	Absences
Motorists	Accidents
Diseases	Deaths

### How do I use it?

- ❶ Identify the data to analyze. Clarify your purpose, i.e., What do you hope to learn?
  
- ❷ **Decide on the time period** for the study. You usually need 50 data points for a Pareto chart.  
  
Take into account seasonality, or typical patterns within the time period selected, to ensure the data will be representative for your purpose.
  
- ❸ **Decide on the categories** and develop clear definitions of each.



Remember, once collected, data can only be stratified into categories if properly coded.

- ◆ In defining categories, be sure each category is discrete, with no overlap between them, so each event or item falls into only one category.
  
- ◆ If recording times or quantities, decide how much precision is needed, i.e., how many decimal places.

POINTER: See [Stratification](#) for possible categories.



See [Checksheet](#) for more on data collection.

For example, imagine that you manage a benefits call center, where employees of your company call to ask questions about their health or retirement benefits, etc.

You are interested in reducing the amount of time your agents spend answering calls, so you collect some data.

As calls come in, your agents ask a few key questions. An on-screen form with pre-defined response categories might be used to record data.

You consider the following types of questions in determining what questions to ask and what data to collect.

- ◆ **WHO**—Caller's Position, function;
- ◆ **WHAT**—Type of question or inquiry;
- ◆ **WHICH**—Forms and form sub-sections in question;
- ◆ **WHEN**—Call length, Enrollment date, Caller's stage in the benefits cycle;
- ◆ **WHERE**—Location of the caller;
- ◆ **WHY**—Suspected cause of problem.

You decide upon some key WHO questions and develop possible response fields like those shown on the next page.

Position	Department
♦ Executive	♦ Manufacturing
♦ Manager	♦ Engineering
♦ Technical	♦ Marketing
♦ Supervisor	♦ R&D
♦ Foreman	♦ HR
♦ Line worker	♦ IT
♦ Clerical	♦ Finance
♦ Other	♦ Other



Always test your categories before data collection. You may find categories overlap, or too many responses fall in the Other category.

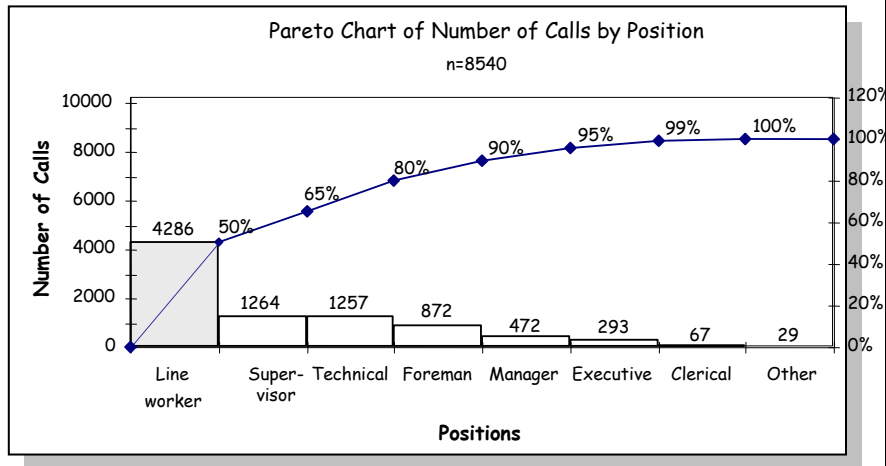
④ **Record the data**, i.e., the number of times each event falls into a particular category. You can do this by:

- ♦ Using existing data, or
- ♦ Gathering new data.

⑤ Following is an example of data collected by the fictitious Benefits Call Center.

Benefits Center Data by Position of the Caller				
Positions	Total # of EEs	# of calls	Avg. Call Length (min.)	Calls as % of EE #s
Executive	315	293	22.5	93
Manager	1,183	472	5.2	40
Technical	1,820	1,257	6.4	69
Supervisor	5,422	1,264	2.6	23
Foreman	6,923	872	3.8	13
Line worker	36,422	4,286	2.2	18
Clerical	1,267	67	19.3	12
Other	624	29	3.4	5
Totals	53,976	8,540	65.4	273

⑥ **Chart the data** as shown in the example below.



- ◆ The horizontal (X) axis is used for the categories, with equal widths for each.
- ◆ Put the "Other" category last.
- ◆ The left hand vertical axis (Y) should be equal to the total number of occurrences (n).
- ◆ The right hand vertical axis should be 120%.
- ◆ Graph each occurrence as a bar, with all bars touching and arranged from largest to smallest.
- ◆ The height of each bar equals the count for that category.



**HINT:** Note that the actual count for each category is written above the bar in the previous example of a Pareto chart for the number of calls by position.

- ⑦ **Add the cumulative line**, from zero to the right hand axis, to show how many bars it takes to account for 80% of the total.

For each bar, the point on the cumulative line represents the per cent of that bar plus all preceding bars.

⑧ **Interpret the chart.**

- ◆ Check to see if the Pareto principle applies.
- ◆ Do 20% of the categories account for 80% of the total occurrences?



**HINT:** Even when the 80/20 rule does not apply, if a few ( $\leq 20\%$ ) of the categories account for 60% or more of the occurrences, you may still have identified the vital few categories having the greatest impact on the situation or problem.

If not, you should look for other ways to stratify the data (Step 8-10), until the 80/20 principle applies and you can identify a vital few.

In the previous example, the Pareto principle does not apply. About 80% of calls are made by three of the eight groups:

- ◆ Line workers,
- ◆ Supervisors, and
- ◆ Technical personnel.



Check to make sure the "Other" category is not too large. If it is one of the larger bars, then re-stratify the data with specific categories for responses currently being captured as "Other".

## 9 Use Different Categories

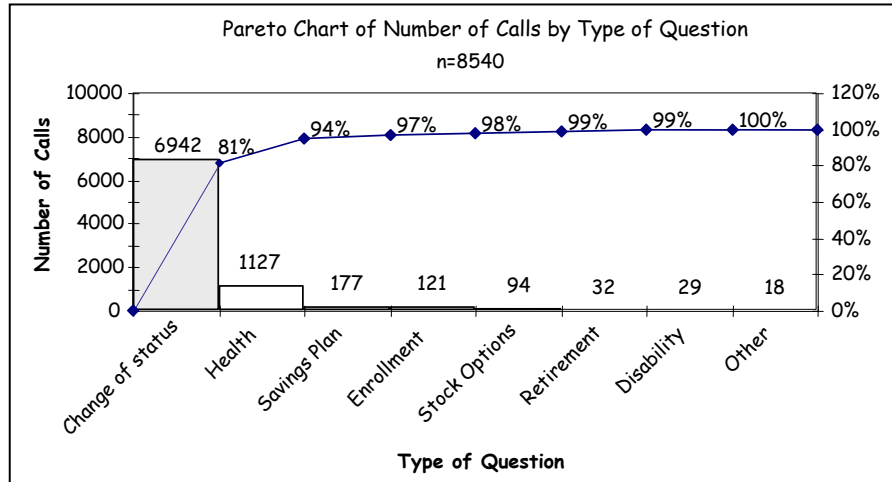
If you identified multiple stratification factors for the data being collected, you can stratify it in different ways and produce various Pareto charts.

One of the stratification factors identified for the Call Center data was the "Type of Question" asked by the caller.

Below is data for the number of calls by "Type of Question".

Type of Question	Number of calls
Enrollment	121
Health	1,127
Retirement	32
Stock Options	94
Savings Plan	177
Change of status	6,942
Disability	29
Other	18
Totals	8540

The Pareto chart for this data is on the next page.



In the example above, the Pareto principle does apply. Over 80% of calls are for a change of status, such as:

- ◆ Getting married,
- ◆ Divorced,
- ◆ The birth of a child, or
- ◆ A change of address, etc.

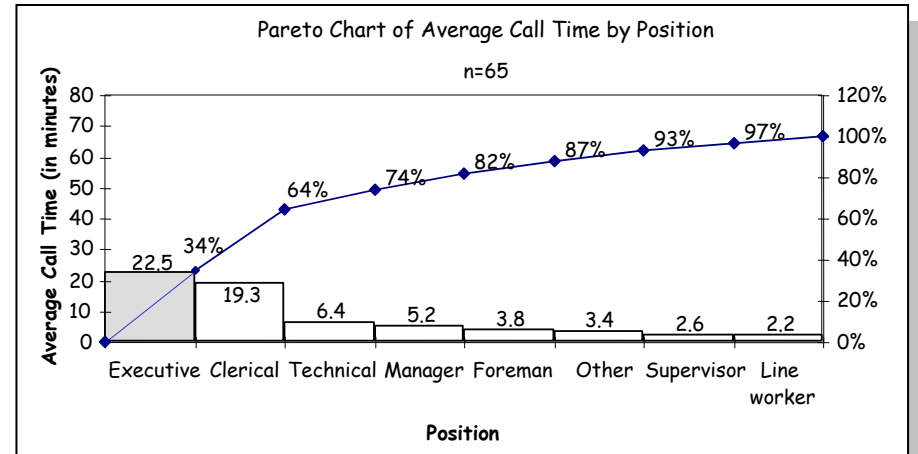
This might suggest you could reduce significant time spent answering this type of call if you can develop an automated means for employees to make simple changes.

## ⑩ Stratify according to Impact.

Before taking action, it is always advisable to stratify the data in terms of impact, e.g., time or money, versus just the count of occurrences.

For the Benefits Center, the count is number of calls. One way to think of impact is in terms of the length of those calls, i.e., the longer the calls, the more resources required by the Center to cover phones.

The Pareto Chart for Average Call time by Position of the caller is shown below.



Note that two relatively small groups of employees, Executive and Clerical, are the largest bars for average call time. For some reason, these two groups are making the longest calls to the center.

## ① Normalize the data

Counts can have very different areas of opportunity, i.e., the range of possible values for the count. For this reason, it is often best to normalize the data by converting counts to a rate or percentage.

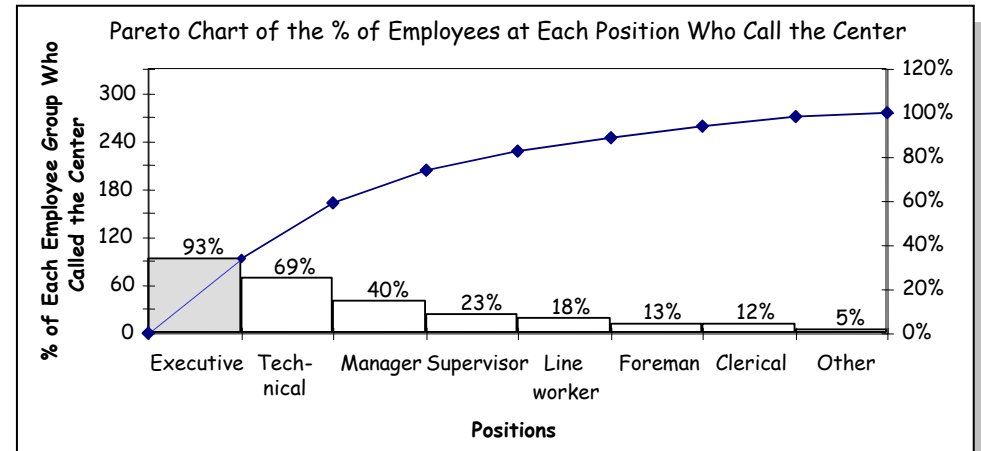
For example...

Areas of Opportunity Can Vary...	...So Convert Counts to Rates
♦ Sales volume can vary based upon the number of visitors to a store.	$\frac{\text{Sales \$}}{\text{\# Store visitors}}$
♦ The number of complaints can vary based upon the number of customers.	$\frac{\text{\# Complaints}}{\text{\# Customers}}$
♦ Payroll errors can vary based upon the number of employees paid.	$\frac{\text{\# Payroll errors}}{\text{\# checks processed}}$

In our call center example, since there are only 315 Executives and over 36,000 line workers, it is not surprising that the count of calls by Executives is dramatically smaller. Instead, you might look at the per cent of callers by position:

$\frac{\text{\# of calls by position}}{\text{total \# of employees by position}}$

This Pareto chart is shown below.



The normalized data shows that on a percentage basis, far more Executives, Technical personnel and Managers called the center.

Clearly the higher "ranking" employees made greater use of the Benefits Center during the period in question.



## 12 Work on the Largest Bar

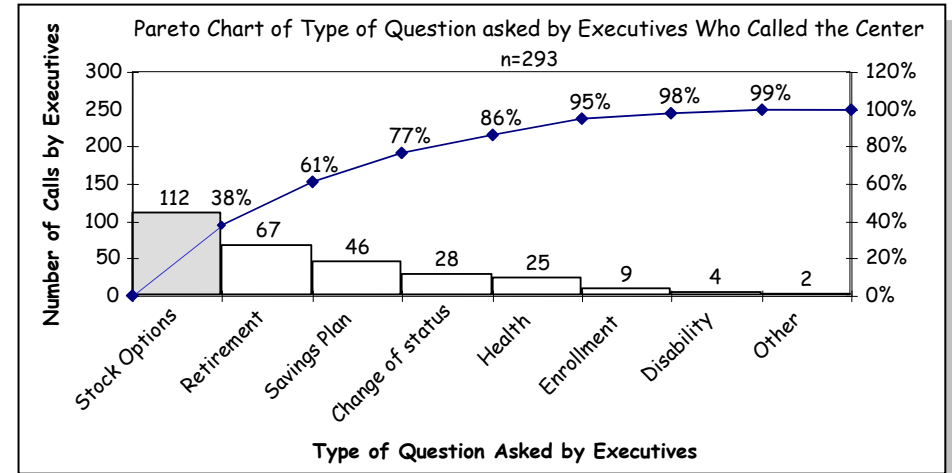
Once the Pareto chart has identified the vital few areas, a good next step is often to take the largest bar and perform additional Pareto analyses of that subset to further localize the problem or causes.


For the Benefits Center example, the charts shown previously indicate that as a group, Executives make a disproportionately higher number, and some of the longest calls to the Center.

It would make sense to do further Pareto analysis of just the Executive data to better understand why.

The Pareto Chart at the top of the next column looks only at Executive calls.

It indicates that the most frequent type of call made by Executives is related to stock options.



**POINTER:**  It is common to begin cause analysis on the largest bars. See Improvement Methods in [Chapter 6](#) such as [Process Mapping](#), [Is/Is Not Analysis](#), [Cause-Effect diagram](#), etc. for help with this.