# Data Analysis Tools and Charts

# Data Analysis Tools and Charts

## Flowcharts

### What Are They?

The Flowchart is a representation of the steps of a work process to show their sequence

### When To Use Them

* Preparing standard practices and procedures
* Auditing work processes
* In generating ideas for improvements

### What Do They Achieve?

It is difficult to understand work processes and solve problems that arise in them unless you have a formal way of looking at them. Flowcharting helps you to depict, and therefore understand how the process works and highlight where problems exist.

### Tell Me More

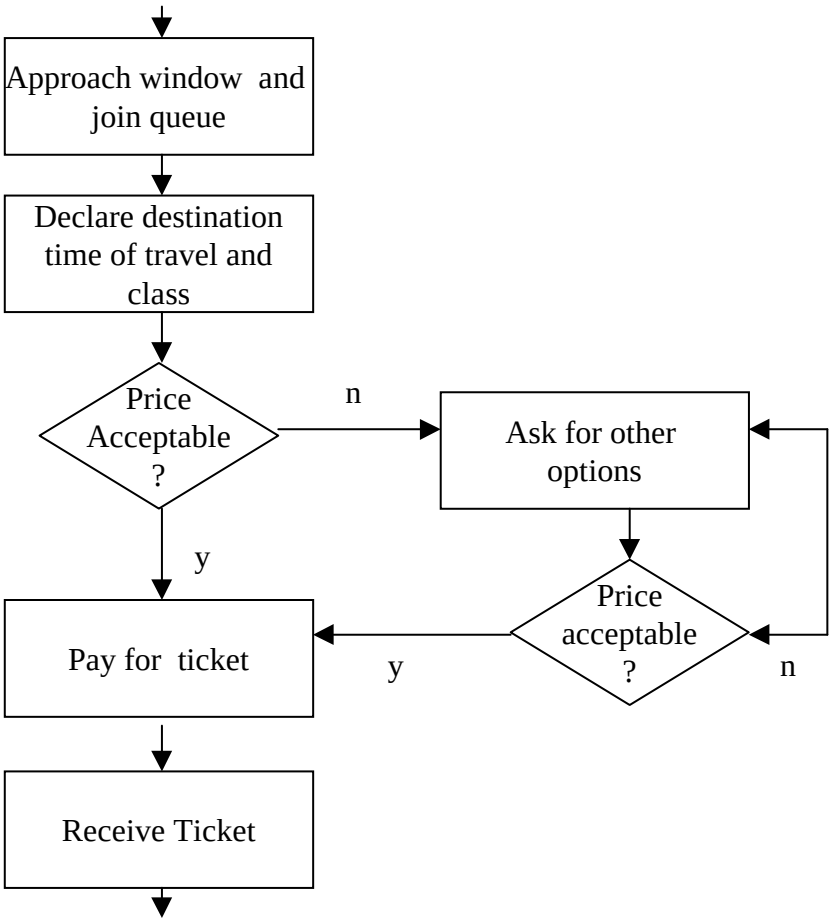
Even the simplest of work processes can have several steps. More complicated processes, like most industrial processes, have very many steps. With all processes it is very important to understand how the whole process fits together before you begin to analyse or change it.

The benefit that Flowcharts provide are:

* an in depth knowledge of department activities
* a training package for new staff, for example part of standard practices and procedures
* a pictorial representation of activities, inputs and outputs
* a common understanding of the work to be done
* an understanding of the effect that changes in one part of the process will have on other parts

Flowcharts link naturally into other problem solving tools and techniques: Solution Effect Analysis and Concentration Diagrams for example.

Looking at an example with which we are all familiar, the Flowchart below shows the process of buying a railway ticket.



**What Do I Have To Do?**

The main elements of a simple Flowchart are:

| | | |
|---|---|---|
| Box | : | to show activities |
| Diamond | : | to indicate a decision point |
| Arrow | : | to show the direction of flow from one Activity to the next |

The steps involved are:

*Step 1*:

Gather a group of people who represent the various parts of the process.

*Step 2*:

Decide where the process begins and ends.

*Step 3*:

List the main activities and decision points in the process.

*Step 4*:

Arrange these activities and decision points in their proper order, using arrows to show direction of flow.

*Step 5*:

If necessary, break down the activities into simple steps to reduce their complexity.

*Step 6*:

Take a piece of work through the process yourself to make sure the description is correct.

**Points to Remember**

If a process is complex or contains complex steps, a good idea to get you started is to draw a simple process chart of the complex activities. Each of these complex activities can then be represented by a separate Flowchart.

As well as being a mechanism for describing a process that already exists, Flowcharts can be a useful planning tool, showing the steps that need to be carried out to achieve a desired end.

Flowcharts are extremely useful for audit purposes as they give an independent standard, against which anyone can audit current practices.

## Check Sheets

### What Are They?

A Check Sheet or Tally Sheet is a data recording form that tells you how many times something has happened. Check Sheets are usually used for recording numerical data.

### When To Use Them

* Gathering data

### What Do They Achieve?

Data Reflects facts. Data must be collected carefully and accurately for the results of analysis to have any meaning. Data without a clear purpose or unreliable data are worth nothing.

Check Sheets are all about making data easy to obtain and to use.
The main purpose of a check Sheet is to compile the data in such a form that they may be used easily, and analysed readily.

### Tell Me More

Check Sheets are useful for groups when several people are collecting data. They ensure that everyone will collect comparable data, in the same format. They also provide a clear record of gathered data. There are as many different designs of Check Sheets as there are reasons for collecting data. The important thing is to design a Check Sheet to suit the data being collected.

For example, a Company which made printed paper bags was aware that it suffered from limitations on output. Application of Cause and Effect Analysis, and Pareto Analysis had pinpointed the problem to the unreliability of the electrical systems on the bag folding and gluing machines. There was no detailed information available on where the electrical problems lay. Accordingly, the Check Sheet below was issued to electricians to record the location of electrical problems.

| **ACME BAG MACHINES**<br><br>**Electrical Fault Recording Sheet** | | | |
|---|---|---|---|
| **CONTROL BOX** | | **ELECTRICAL EYE** | |
| **SOLENOID** | | **STARTER** | |
| **POWER SUPPLY** | | **FUSE** | |
| **MOTOR** | | **OTHERS** | |

**What Do I Have To Do?**

*Step 1*:

As a group or team you must first decide what data you need. Brainstorming and Cause and Effect Analysis are useful ways of determining this information.

Therefore you must initially choose all the factors which might effect a situation and how you might analyse the information when you collect it.

*Step 2*:

Design an individual Check Sheet form for people to use as they record the data. This sheet must reflect the type of data being collected.

*Step 3*:

Test the Check Sheet by getting someone who did not help design it to use it.

*Step 4*:

Make any revision that are necessary as a result of Step 3.

*Step 5*:

Many people are going to use the Check Sheets to collect data. You must design a master Check Sheet to combine the results from the individual forms.

*Step 6*:

Proceed to gather data. Issue the Check Sheets, and instruct users.

**Points To Remember**

Be absolutely clear about why you are collecting data before you go out and collect it.

Get everything into the data. The key to arriving at a good analysis is to collect data skilfully. If you haven't collected information about one important factor the rest of the data may be worthless.

Talk to those who know the job best when you are designing the check sheet.

## Frequency Histograms

### What Are They?

Frenquency Histograms are a method of presenting data, which show the way it varies.

### When To Use Them

Frequency Histograms should be used whenever we need to investigate the amount of a characteristic, for example, if we were trying to see what proportion of bags of sugar weighted more than 100 grams above or below the advertised weight.

### What Do They Achieve?

Every day in our jobs we collect data in various forms, for example, percentage of items defective, miles travelled, number of visits. Things vary naturally (if they did not we would only need to measure them once) and if each time we measured them the answer was the same we would become suspicious; We live in a world of dispersion, of variation, and to make decisions we need to understand the nature of this variation. Frequency Histograms help us understand how data varies.

### Tell Me More

As an example, let us look at the time taken to recruit administrative staff within a large company, from the time that an impending vacancy was identified to the new member of staff starting work. We know already that there exists a range of variation in the time taken. A Frequency Histogram will allow us to visualise this variation.

**What Do I Have To Do?**

Our first step is to collect data. If we review the Personnel Department records and extract the time, in working days, to recruit administrative staff, we get the following data:

| TIME TAKEN TO RECRUIT ADMINISTRATIVE STAFF (working days) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 32 | 27 | 27 | 36 | 31 | 31 | 19 | 38 | 12 | 28 | 25 | 33 |
| 48 | 44 | 16 | 34 | 21 | 28 | 27 | 59 | 31 | 31 | 39 | 36 |
| 57 | 53 | 29 | 36 | 47 | 39 | 26 | 41 | 34 | 38 | 42 | 41 |
| 13 | 22 | 37 | 21 | 27 | 31 | 21 | 29 | 24 | 29 | 26 | 17 |
| 18 | 26 | 22 | 19 | 33 | 26 | 32 | 21 | | | | |

Developing the Histogram requires that we divide the range of data into intervals; and count the number of data points which fall into each group or "class".

*Step 1*:

We start by examining the data to see how wide is the range. We identify the largest and smallest data points. In our sample these are 59 and 12. We also count the number of data points, in this case 56.

*Step 2*:

We select the number of classes we should choose from the following table:

| Number of Data | Number of Classes |
|---|---|
| **< 50**<br>**50 - 100**<br>**100 - 250**<br>**>250** | **5 - 7**<br>**6 - 10**<br>**7 - 14**<br>**10 - 20** |

We thus select a class width within (range ÷ 6) and range ÷ 10)

where range equals (59-12) i.e. 47

Our class should be between (47 ÷ 6) and (47 ÷ 10) i.e. between 7.8 and 4.7

We will adopt a class width of 5.

*Step 3*:

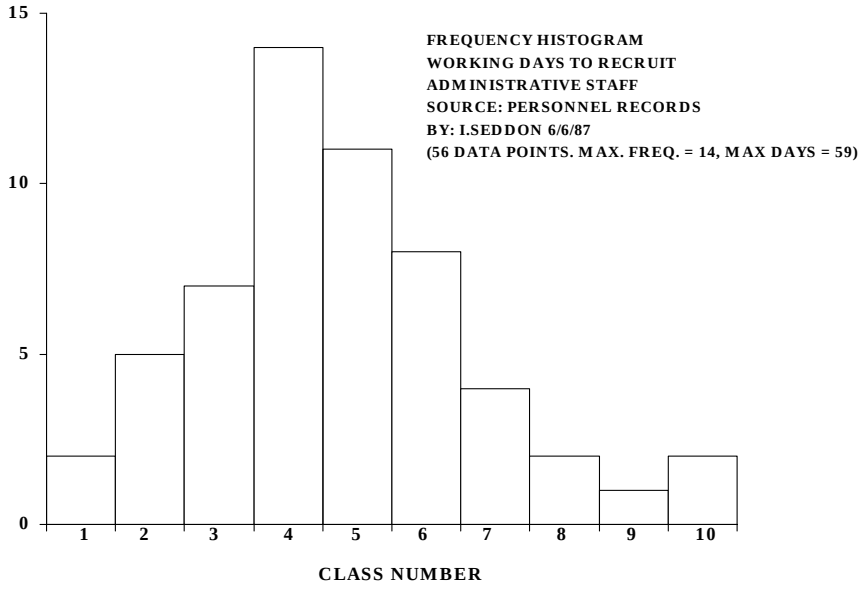
We construct a Frequency Table using class width 5, to include all our data points. Use a 5 bar gate, I, II, III, IIII, ~~IIII~~ etc to count the number falling within each class boundary. The total should correspond to that found in Step 1. Mistakes often happen here, so be careful, totalling is a good crosscheck. Get into the habit of doing crosschecks to prevent rework at a later stage.

*Step 4*:

Looking at the Frequency Table you can get an idea of the overall picture, but if it is indicated in a diagram it becomes much clearer.

On graph paper, mark the class boundaries horizontally, and the frequency vertically. The draw the Histogram blocks according to the scale.

| CLASS NUMBER | CLASS BOUNDARIES | FREQUENCY TALLY | FREQUENCY |
|---|---|---|---|
| 1 | 10-14 | II | 2 |
| 2 | 15-19 | ~~IIII~~ | 5 |
| 3 | 20-24 | ~~IIII~~ II | 7 |
| 4 | 25-29 | ~~IIII~~ ~~IIII~~ IIII | 14 |
| 5 | 30-34 | ~~IIII~~ ~~IIII~~ I | 11 |
| 6 | 35-39 | ~~IIII~~ III | 8 |
| 7 | 40-44 | IIII | 4 |
| 8 | 45-49 | II | 2 |
| 9 | 50-54 | I | 1 |
| 10 | 55-59 | II | 2 |
| | | | 56 |



In the blank areas of the diagram, write the title, the number of data points, the maximum values, the source of the date, the date and your name. This will provide a complete record for future reference.

**Points to Remember**

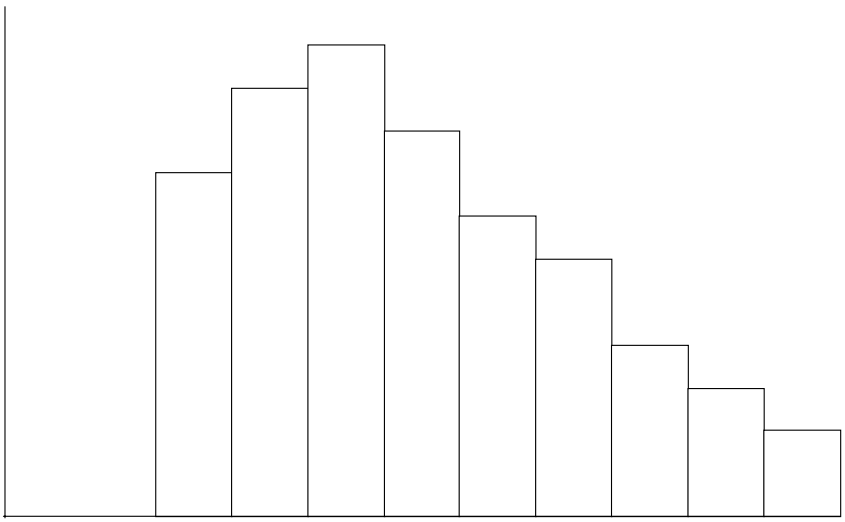
Let's try to answer the following questions by looking at the Histogram.

* What is the most common time to recruit?
* Is it cliff like?
* Is the distribution symmetrical?
* Is there only one peak?
* Does it look like a cogwheel/comb?

In other words, what are the characteristics of the information gathered?

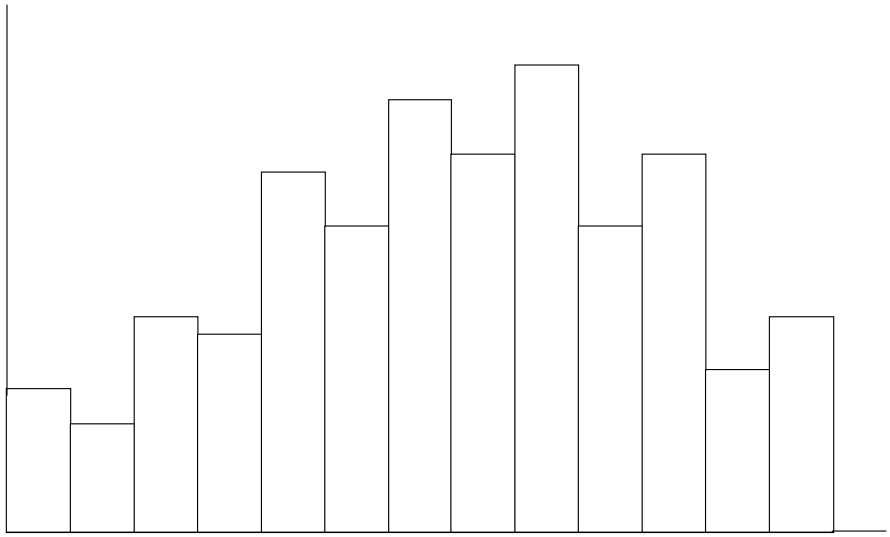
According to our Histogram, it is most common for recruitment to take somewhere in the 25-29 days range. The distribution shows a bias towards the higher times to recruit.

* If the distribution had been like this

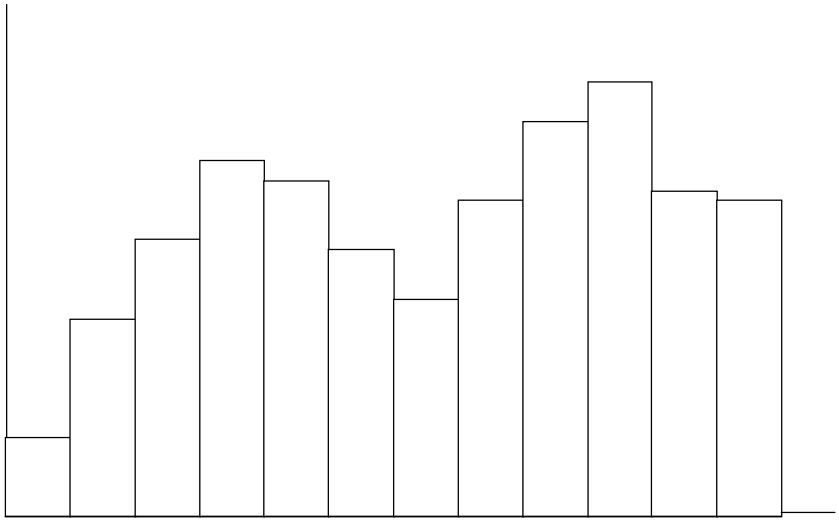


It would appear unusual. It is possible that before the sample was taken there was some "pre-selection" to remove those below a certain value.

* If it looks like this, you need to check the measurement method. Instruments may have been wrongly set or you may have recording problems.



* A Histogram like this suggests that data has been sampled from 2 different populations and you will need to look again at your sampling methods.



Histograms can show changes, by superimposing one Histogram on the other.

Histograms can also be used to compare the product or process with the specifications for that product or process.

In large Histograms there is no need for all class boundaries to be the same, for example the lower boundary could be "under 23" and the upper boundary could be "over 72". Be careful when you do this that you do not mislead people.

# Pareto Analysis

## What Is It?

Pareto Analysis is a technique for recording and analysing information relating to a problem or cause, which easily enables the most significant aspects to be identified.

A Pareto Diagram is a special form of Bar Chart, or Column Chart, which allows the information to be visually displayed.

## When To Use It

* Separating the "vital few" from the "useful many" problems
* Selecting major problem areas
* Identifying major effects and causes
* In conjunction with other techniques, e.g. Brainstorming

## What Does It Achieve?

A Pareto Analysis identifies the most important elements contributing to a situation. It supports decisions on where resources are best directed to effect a change.

Very often the simple process of arranging data may suggest something of importance that would otherwise have gone unnoticed. Selecting classifications, tabulating data, ordering data and constructing the Pareto Diagram have often served a useful purpose in problem investigation.

Pareto Diagrams can help in communicating priorities and the need for action throughout an organisation. They provide a clear and concise illustration of priorities and needs.

A Pareto Diagram should include a line illustrating the cumulative contribution of causes.

Cumulative lines are convenient for answering questions such as "what types" of fault or error constitute 50% of all defects?" Pareto Analysis involves organising data in preparation for the construction of the cumulative line.

**Tell Me More**

Analysis often reveals that, for example, a small number of failures are responsible for the bulk of quality costs. This confirms to the so-called "Pareto Principle", named after the Italian economist who discovered that the majority of his country's wealth was owned by relatively few people.

In many situations a similar pattern becomes apparent when we look at the relationship between numbers of items and their contribution to the extent of the problem. This pattern has been referred to as the 80/20 rule and shows itself in many ways. 80% of your telephone calls, for instance, come from 20% of your colleagues; similarly 80% of a company's failure costs probably result from 20% of its problem areas.

The 80/20 principle does not mean that exactly 80% of the total problem is provided by 20% of the features but that there is usually a similarly large imbalance. The ratio itself is not as important as the fact that the major causes are being identified.

In other words, amongst the wide range of problems that you may be faced with, there are a few vital ones which must be tackled immediately, and many others which can be dealt with later. Pareto analysis shows at a glance which problem areas can be regarded as a "vital few", needing special measures to tackle them, and which are the "useful many".

In most cases the identification of the "vital few" does not come as a complete surprise. On the contrary, some of the problems on the list will have long been notorious, but the major benefits of a Pareto Analysis are that:

* some notorious problems are confirmed as belonging to the "vital few"
* some problems, previously not notorious, are identified as belonging to the "vital few"
* the "useful many" are identified;
* the real size of the problem is revealed
* the priority for sorting out problems is established

**What Do I Have To Do?**

There are five steps to a Pareto Analysis:

*Step 1*:

List the errors or activities to be analysed.

*Step 2*:

Calculate the total for the whole list, and the percentage that each error or activity represents of this total.

*Step 3*:

Rank the errors or activities starting with the largest, and calculate the cumulative percentage as you go down the list. You will end with the smallest item which will make up the final element of the 100%.

In the following example it has been decided to investigate the costs of breakdowns on plastic film extruding machines. We are to analyse material and downtime costs.

Production records have been reviewed, and data collected in the following form:

| DATE | CAUSE | DOWNTIME (MIN.) | MATERIAL LOSSES (kg) |
|---|---|---|---|
| 1/6 | 8 | 90 | 400 |
| 1/6 | 8 | 110 | 420 |
| 1/6 | 1 | 16 | - |
| 1/6 | 6 | 20 | 14 |
| 2/6 | 6 | 19 | 6 |
| 3/6 | 1 | 10 | - |
| 4/6 | 8 | 60 | 295 |
| 4/6 | 9 | 110 | 109 |
| 5/6 | 2 | 240 | 300 |
| ETC | ETC | ETC | ETC |

Downtime costs £2/minute and waste materials costs £0.50/kg.

So working out the material, downtime and total costs for each we get the following table:

| CAUSE | FREQUENCY | DOWNTIME (MIN) | waste (kg) | DOWNTIME COST | MATERIAL COST | TOTAL COST | PERCENTAGE CONTRIBUTION |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 13 | - | 26 | - | 26 | 0.5 |
| 2 | 6 | 737 | 2000 | 1474 | 1000 | 2474 | 45.8 |
| 3 | 2 | - | 60 | - | 30 | 30 | 0.6 |
| 4 | 3 | 13 | - | 26 | - | 26 | 0.5 |
| 5 | 1 | 18 | 70 | 36 | 35 | 71 | 1.3 |
| 6 | 4 | 30 | 56 | 60 | 28 | 88 | 1.6 |
| 7 | 5 | 15 | 200 | 110 | 100 | 210 | 3.0 |
| 8 | 9 | 375 | 1800 | 750 | 400 | 1650 | 30.5 |
| 9 | 6 | 70 | 600 | 140 | 300 | 440 | 8.1 |
| 10 | 1 | 85 | 38 | 170 | 19 | 189 | 3.5 |
| 11 | 1 | 20 | - | 40 | - | 40 | 0.7 |
| 12 | 7 | 65 | 54 | 130 | 27 | 157 | 2.9 |

So, if we now rank the causes:-

| CAUSE | PERCENTAGE OF TOTAL COST | CUMULATIVE PERCENTAGE |
|---|---|---|
| 2 | 45.8 | 45.8 |
| 8 | 30.5 | 76.3 |
| 9 | 8.1 | 84.4 |
| 7 | 3.9 | 88.3 |
| 10 | 3.5 | 91.8 |
| 12 | 2.9 | 94.7 |
| 6 | 1.6 | 96.3 |
| 5 | 1.3 | 97.6 |
| 11 | 0.7 | 98.3 |
| 3 | 0.6 | 98.9 |
| 1 | 0.5 | 99.4 |
| 4 | 0.5 | 99.9 |

The lost 0.1% is due to rounding of percentages.

We can now draw the Pareto Chart.

*Step 4*:

Draw a Bar Chart as shown in the diagram. The vertical scale shows the volume of what it is you are comparing and the horizontal scale breaks this down into meaningful categories so you can tell which category is causing the greatest problem. For example cause 2 was responsible for 45.8% of the total cost.

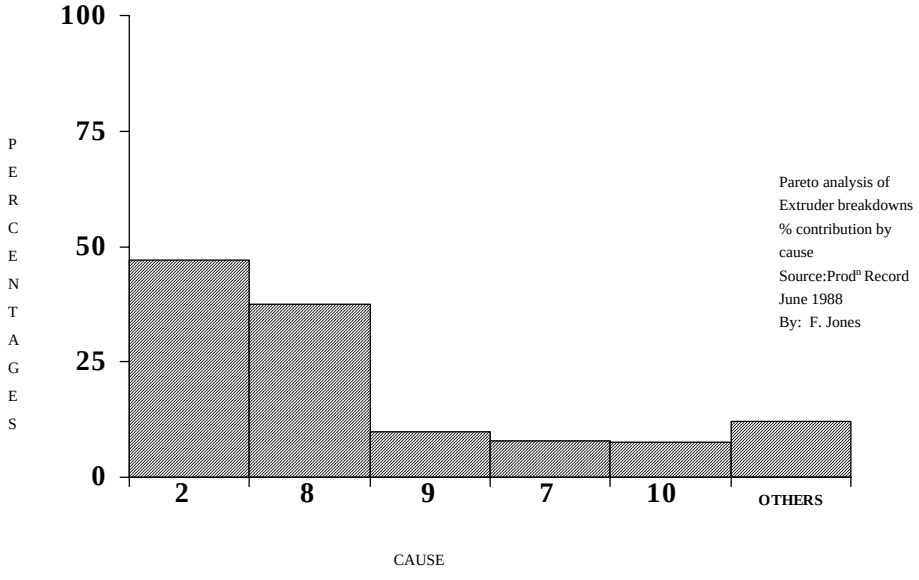
When many types of defects are involved the horizontal scale of your Pareto Chart may become very wide. Put minor errors together and call them "others" , and this will narrow your horizontal scale.

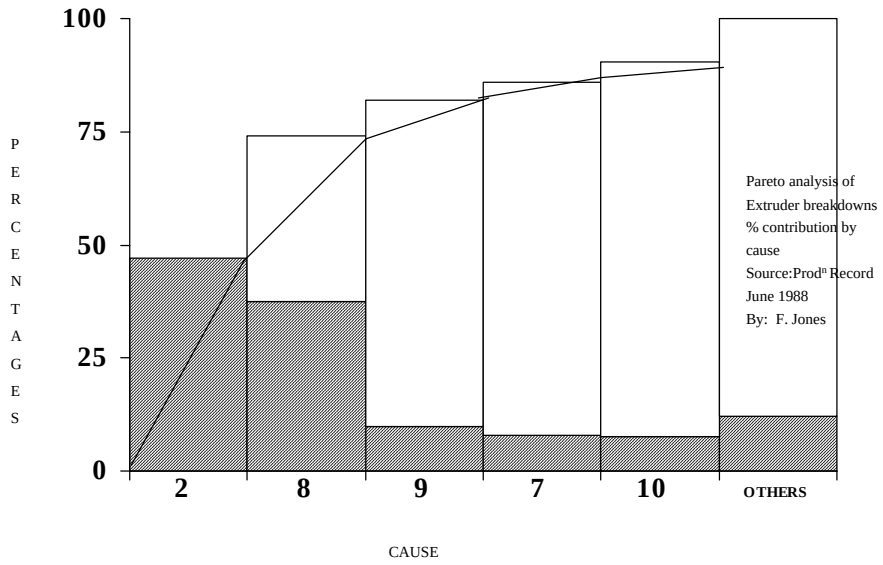
*Step 5*:

The Pareto Chart can be enhanced by drawing in the cumulative curve as shown in the diagram over. The first step is to draw in the cumulative bars; the first in this case cause 2. The next cumulative bar represents cause 2 added to cause 8, and so on until all the bars have been included. The next step is to draw the cumulative curve and this is done by joining the top right hand corner of each bar.

**Points To Remember**

The candidates for priority action, the "vital few" will appear on the left of the Pareto diagram where the slope of the cumulative curve will be steepest.






The "useful many" should not be ignored. Sometimes what is now a triviality can become significant later.

There may be some problems about which it is difficult to collect specific numeric data such as issues concerning morale, or customer image. In these cases it may be necessary to rely more heavily on subjective decisions about which problems are the "vital few" to be marked out for special treatment. This type of subjective Pareto Analysis can be extremely powerful in its own right especially when twinned with the use of Cause and Effect diagrams.

The Pareto principle is not normally used in isolation. It tends to be applied in conjunction with Brainstorming and Cause and Effect Diagrams.

# Line Graphs (Run Chart)

## What Are They?

A Line Graph / Run Chart is a useful way of showing changes in numerical amounts, usually over a period of time. It can also be used to show relationships between two variables.

## When To Use Them

Line Graphs should be used wherever it is necessary to trace the change of a variable in relation to another variable. Line Graphs are most commonly used to trace the change in a variable over time.

## What Do They Achieve?

The advantage of representing data in pictorial form is that it contributes to the communication process. The purpose of making a graph is to understand a situation quickly and be able to take appropriate action. Because of this it is essential to understand that it is not sufficient for the person making the graph to understand it; it must be understood correctly by those who see it as well. A Line Graph can present a vast amount of information in a single easily understood form.

## Tell Me More

A lot of data is in the form of a time series, in which we take a variable and show how its magnitude has varied over a period of time, perhaps a shift, a month, or several years. Long tables of numbers can be very dull and uninteresting and often do not get the attention they deserve because of this. Graphs give a simple pictorial representation that is much easier to handle.

As an example, look at the numbers of people passing through a major railway station, month by month.

| MONTH | PASSENGER MOVEMENTS |
|---|---|
| January | 161,000 |
| February | 160,000 |
| March | 165,000 |
| April | 171,000 |
| May | 190,000 |
| June | 200,000 |
| July | 225,000 |
| August | 242,000 |
| September | 230,000 |
| October | 195,000 |
| November | 183,000 |
| December | 176,000 |

This is hard to digest, and the significance of the numbers is not immediately apparent.

The Line Graphs is much easier to look at and remember than the table of figures. In examples where there is more data, this is even more true. Looking at the Graphs it is very easy to see the largest and smallest numbers, the difference between them and any "pattern" in the data.

**What Do I Have To Do**

There are five steps to drawing a Line Graph:

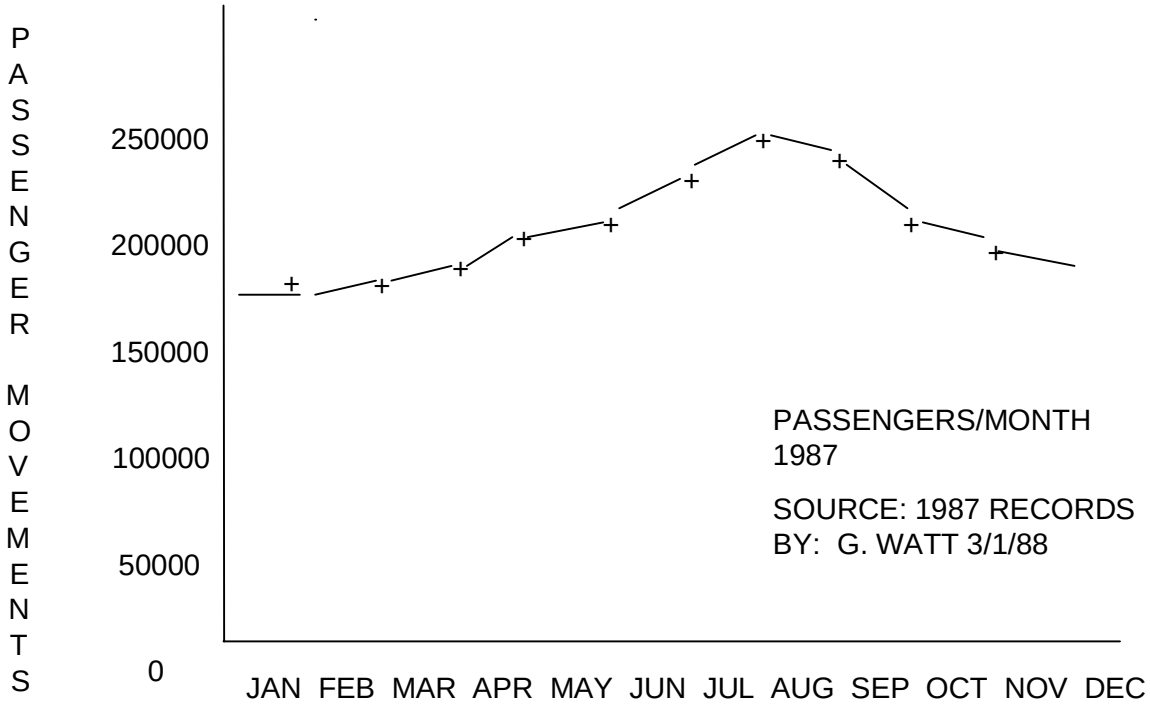
Using the data shown earlier on passenger movements to illustrate the process.

*Step 1*

Identify the minimum and maximum points (February at 160,000, and June at 242,000), and the number of data points (12).

*Step 2*

Decide the scales. The timescales covers 12 months, so we should use 12 points on this scale. The maximum number of passengers is 242,000 so we will use 250,000 as the maximum and 0 as the minimum.




*Step 3*

Draw the axes. It is usual to plot the time scale on the horizontal (or X) axis and the numerical readings on the vertical (or Y) axis. Draw these with the scale decided in Step 2 and label them.

*Step 4*

Plot the data on the graph. As you plot the points, tick them off on your list of data to make sure you plot them all in the right order.

*Step 5*

Join up the points. Graphs with dots can be hard to look at and understand. Make sure when you join up the points that they are in the right order.

## Points To Remember

If the magnitude of the numbers on the vertical Y axis is very large, for example 133880, 127450, 111370 you should consider dividing the numbers by 1000 to get 133.8, 127.45, 111.37 etc. You could then plot these numbers and show the scale as "Y ÷ 1000".

Graphs can be used to show two or more things that you believe are related by plotting them with different scales on the same chart.

All Graphs should have title, which describes the graph, a date of construction, the source of data and the name of the person drawing it.

If changes are very small they can be shown by "coding" the data. Suppose the results are 133180, 133179, 133170 etc. where all values are 133 thousand one hundred and……. then you can "code" the data by subtracting 133100 from it to get 80,79,70 etc. and plotting the new data on a graph. Be very careful when you do this to make sure that you are not emphasising an effect that is not there. It could well be that all the answers are 133200 and the differences are errors in measurement.

We said earlier that the most common use of Graphs was when we were drawing data that changed over time. The change over time could either be discrete or continuos, that is "smooth" or involving sudden "jumps".

# Bar Charts

## What Are They?

The Bar Chart is one of the most common of all techniques for presenting data. It is very versatile and capable of use with almost any data.

## What Does It Achieve?

This is an easy way to show the relative size and importance of items, which have some link between them. It illustrates pictorially the relative sizes, and also gives information on the absolute magnitude of the items.

## When to Use Them

Whenever you have to illustrate relative sizes of items with some common characteristic.

## Tell Me More

It is often difficult to assimilate lists of figures. When presented with a list it can be difficult afterwards to remember what was first, what was last and "where you came" in the list. The Bar Chart is a simple way to present this data.

**What Do I Have To Do?**

Suppose we are looking at the numbers of various models of car sold in the UK last month.

| MODEL | SALES |
|---|---|
| Ford Sierra | 16300 |
| Ford Escort | 20800 |
| Ford Fiesta | 15600 |
| Austin Metro | 9800 |
| Vauxhall Astra | 9900 |
| Vauxhall Cavalier | 10200 |
| Vauxhall Nova | 6800 |
| Volkswagen Golf | 10100 |
| Rover 200 | 12100 |
| Nissan Cherry | 8200 |

*Step 1*

Identify the largest member of the group. This will have an effect on your choice of scale.

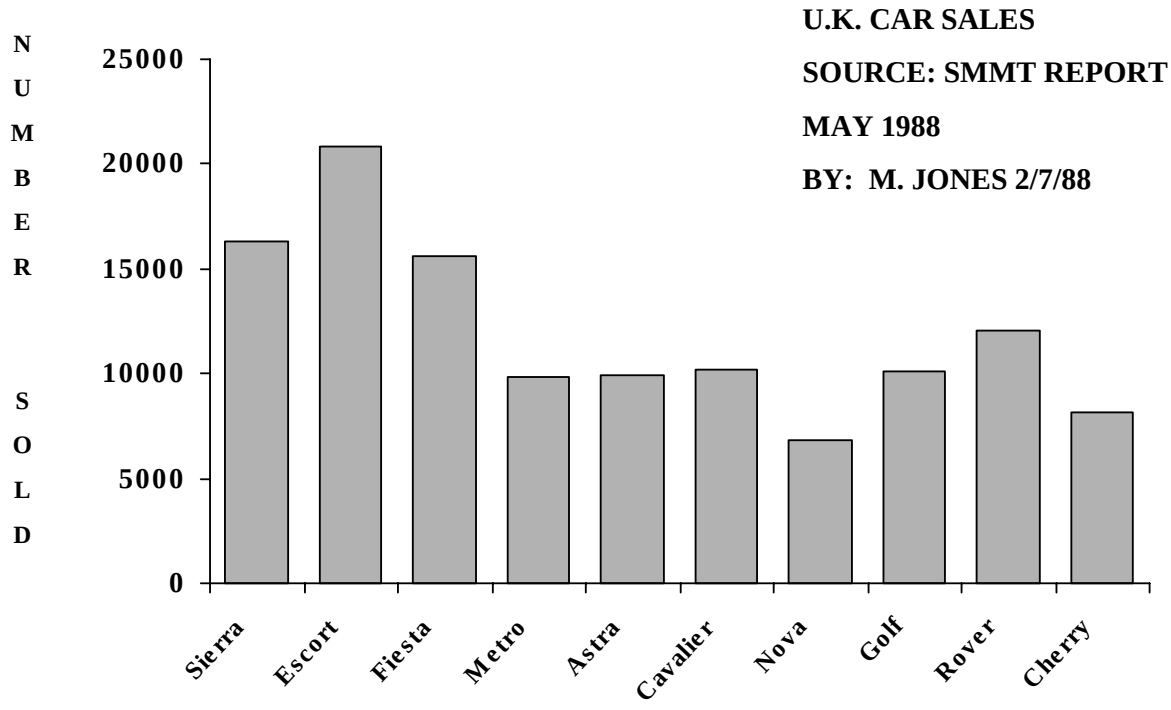
Generally, the bottom end of the scale should be zero and the top will be greater than the largest numeric item on the list.

In our example, choose the bottom end of the scale as zero and the top will be greater than the largest numeric items on the list.

In our example, choose the bottom end of the scale as zero and the top as 2500 and draw a vertical line on a sheet of graph paper to represent the whole scale.

*Step 2*

There are ten items in the list. We wish to demonstrate them as bars. To avoid misleading people, the bars should all be of the same width. It is usual to draw Bar Charts with space between the bars and they again should be the same width. Now draw on the Chart bars representing the sales of the cars.




## Points to Remember

Titling of the Bar Chart should include a description of the chart, a source of the material, the name of the person who drew it and a date.

Labelling can be done in several ways - by writing inside the bars, by writing on top of the bars, or by putting a key on each bar and providing a table.

Size of some of the bars can make things difficult.

If one item is particularly large compared with others, you can "break the bar". Similarly, if all of the items are large and it is difficult to choose a scale you can break them all.

# Scatter Diagrams

## What Are They?

A Scatter Diagram is an effective way to test whether two different characteristics of some operation or phenomenon are connected. Scatter Diagrams make the relationship between pairs of parameters easily visible.

## When To Use Them

* Following Cause and Effect Analysis to evaluate causes
* Investigating causes of problems

## What Do They Achieve?

When you have drawn a Cause and Effect Diagram you will have a list of many potential causes but the efforts of the team may need to be better focused. It is a good idea to study the relationship between each potential cause and the overall effect, reacting to the related ones and dropping the unrelated ones. Scatter Diagrams help you determine if cause and effect are related.

## Tell Me More

When you have brainstormed a problem and then drawn a Cause and Effect Diagram you will have listed many possible causes for the effect you are investigating.

A Scatter Diagram will allow you to test each of these for final effect. You should start by testing the most likely causes, or those believed to most strongly affect the final outcome.

By doing this you can identify which causes can best be used to modify the effect.

Warning: A relationship on a scatter diagram does not necessarily mean that one causes the other. Both variables may be related to a third variable.

## What Do I Have To Do?

*Step 1*

Having chosen the cause and effect pair which we want to test, we first require data: 30 to 50 paired samples should be collected.

For example, we might wish to examine the significance of the voltage applied to a welding device, in the joining of plastic pipes. We are interested in the relationships of leaks per 1000 welds (the effect) to the weld voltage (the cause). Accordingly, we have assembled the following data from the production record.

| WELD VOLTAGE | LEAKS PER 1000 WELDS |
|---|---|
| 30 | 80 |
| 44 | 40 |
| 24 | 100 |
| 36 | 60 |
| 50 | 40 |
| 30 | 100 |
| 54 | 36 |
| 40 | 50 |
| 30 | 120 |
| 24 | 148 |
| 36 | 80 |
| 30 | 100 |
| 24 | 120 |
| 30 | 100 |
| 44 | 36 |
| 24 | 100 |
| 22 | 120 |
| 16 | 144 |
| 40 | 60 |
| 20 | 140 |
| 48 | 40 |
| 30 | 80 |
| 26 | 110 |
| 32 | 80 |
| 44 | 30 |
| 28 | 116 |
| 44 | 50 |
| 22 | 130 |

*Step 2*

We must now decide on the scales we are going to use in a graph of the cause and effect.

Take the maximum and minimum values of each parameter; in this case, for the weld voltage these are 54 volts, and 16 volts respectively, and for leaks per 1000 welds these are 148 and 30 respectively.

We should now draw the axes of the graph to suit these values. You will find it helpful to follow a few tips - always use the X (horizontal) axis for the cause and the Y (vertical) axis for the effect and try where possible to make them of about the same length. If you stick to these conventions, it will reduce the potential for confusion, and will maximise the clarity of a relationship.

*Step 3*

Plot the data on the graph. It is usual to mark points either with a dot or with a cross. If points occur more than once it is normal to indicate this by drawing concentric circles about the point. An example of the graph is shown overleaf.

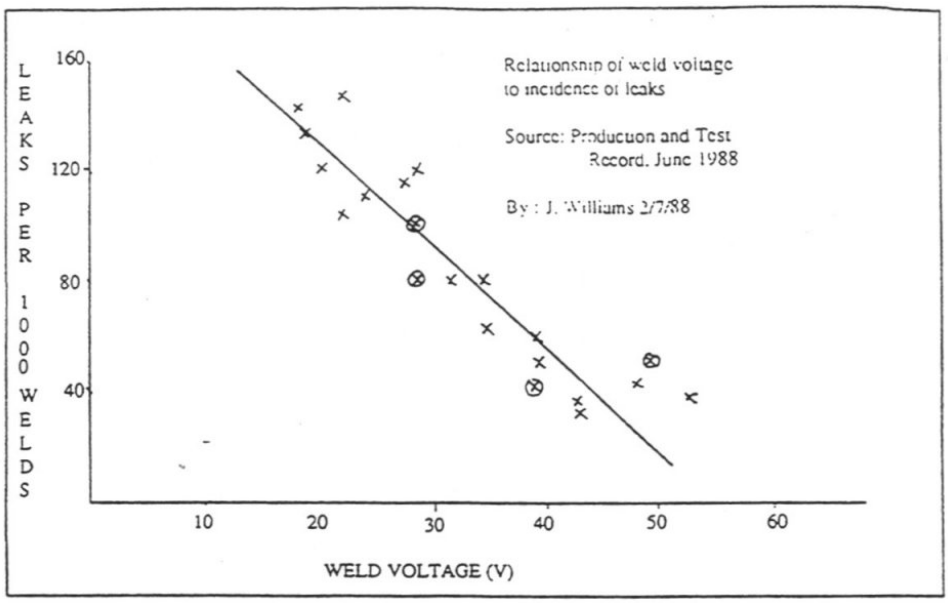
Notice that on the graph we have labelled each axis and written the scales on them. This is very important.

The graph also needs a title. This should describe the graph, indicate the source of the data, give a data and the name of the person, group or team responsible.

When you have drawn your scatter diagram, you can draw a <u>line of best fit</u> through the data points. This will show the expected correlation of the data - the change you might expect to see in the effect by altering the cause.

To do this first of all place a clear ruler over the scatter diagram parallel to the Y axis so that half the points are above the line. Draw a line here. Now place the ruler parallel to the X axis and do the same.

Mark the point where the lines cross and put your ruler on this point, rotating it until you again have the same number of points on each side. When you have done this, draw the line you have found.




## Another Form of Scatter Diagram

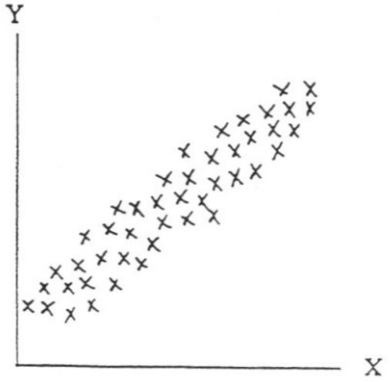
If you have a lot of data or if the data contains many of the same value it is troublesome to plot each one, so make use of the technique for constructing Frequency Histograms and make a Frequency Table with a vertical and horizontal index.

This is another kind of Scatter Diagram which is called a Correlation Table.

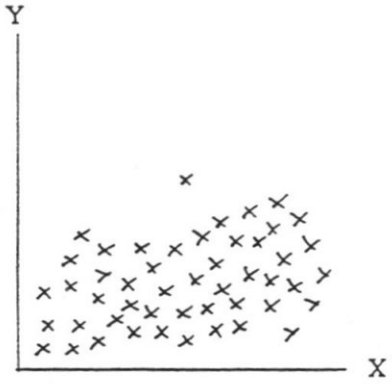
Correlation Tables can be used to remove the need for a separate analysis stage, as they can easily be completed by operators etc. during normal operations.

## Interpreting Scatter Diagrams

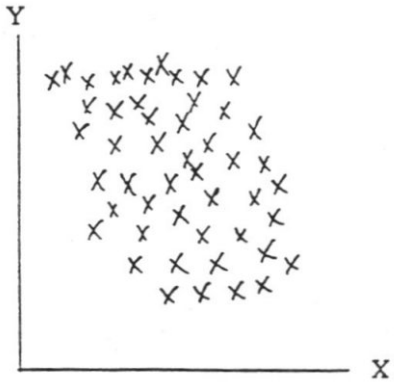
1) An increase in Y depends on increases in X. If X is controlled, Y will be proportionally controlled



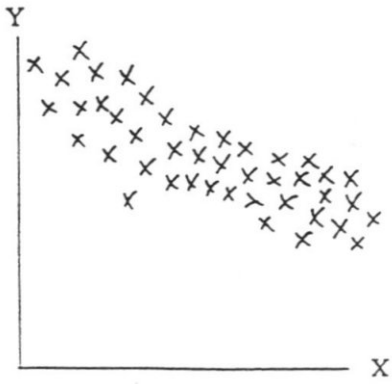
2) If X is increased, Y will increase somewhat, but Y seems to have causes other than X.




3) There is no correlation.




4) An increase in X will cause a decrease in Y.



## Points To Remember

Correlation may only exist over part of the range over which your data has been collected. You need to take care not to misread the diagram when this happens and you should not assume that the correlation exists over the whole range of the data.