

## Strand E *STATISTICS*

## Introduction

---

### *Introduction to Topics*

The British statesman, *Benjamin Disraeli*, once stated, "There are lies, damned lies and statistics." This might have been a bright remark to make in a debate, but the status of statistics was given a quite underrated currency which persists, even today! Disraeli's statement is in fact the reverse of the truth, and a more apt expression would be "There are lies, damned lies and numerical statements made by people ignorant of statistics!"

Statistical analysis makes possible the testing of numerical data for relevance, reliability and validity. Statisticians must present data in such a form that others can utilise the relevant information to enable them to make judgements.

### *Historical Context*

The study of Statistics is reported to have started with the Englishman, *John Graunt* (1620 – 1674), who collected and studied the death records in various cities of Britain. He was fascinated by the patterns he found in the whole population.

Much of current day statistical analysis is of quite recent development, the availability of cheap computing power acting as a catalyst for the development of appropriate ways of presenting and analysing data. In fact, the more advanced statistical analyses and tests are based on probability theory, developed over the past few centuries, but put into a more modern context by mathematical statisticians such as

*Karl Pearson* (1857 – 1936)

*Sir Ronald Fisher* (1890 – 1962)

*Jerzy Neyman* (1894 – 1981).

You can find interesting bibliographies of these people on the internet.

Further information regarding historical context to particular topics is given in the introductions to the units.

### *Misconceptions*

- Misunderstanding often occur when using the words *mean*, *median* and *mode* as measures of 'average'
- Some misconceptions and mis-uses of statistics are shown in the following diagrams.

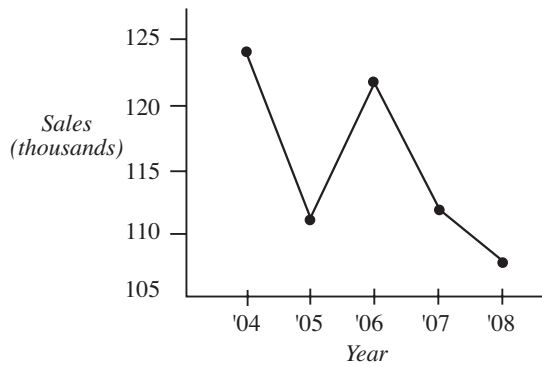
# Strand E *STATISTICS*

# Introduction

*Misconceptions: misleading use of data*

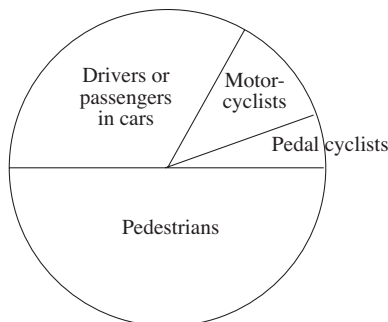
The diagrams below illustrate various data from a variety of statistics, represented in ways that are misleading. Notes on this are given on the next page.

### 1. Sales of CDs, 2004-8



- What does this graph show?
- Why is it misleading?

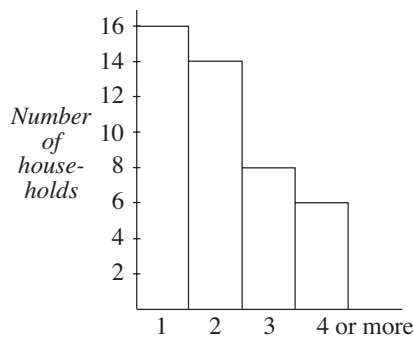
### 2. Road deaths in Scotland



- What conclusion can be drawn?
- What other information is needed?

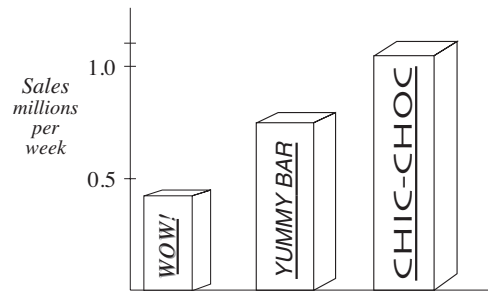
### 3. Number of TV sets per household

The results from a survey of 50 houses is shown below.



- What is missing from this illustration?
- What other data would be relevant?

### 4. Chocolate bar sales (average/week)



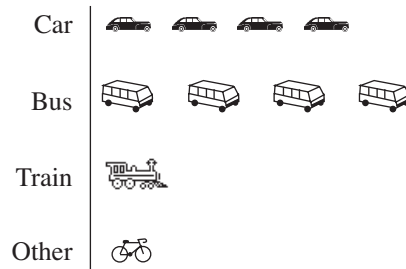
- Why is this illustration misleading?

### 5. Methods of travelling to work

A survey of how 1000 people in a UK town travel to work gave these results.

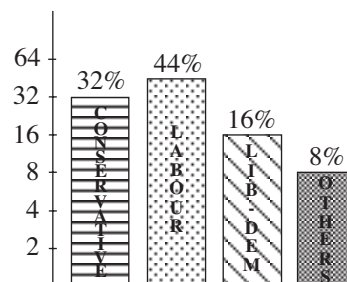
Car:	430	Bus:	356
Train:	136	Other:	78

This is illustrated in the pictogram below.



- What is wrong with this pictogram?

### 6. Voting intentions



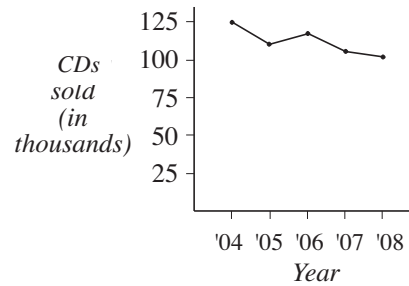
- Why is this misleading?

# Strand E *STATISTICS*

# Introduction

*Misconceptions: misleading use of data - comment on examples on previous page.*

1. Although the number of CDs is on a downward trend, it is not as pronounced as shown in the graph, where the vertical scale starts at 105 (thousands).  
A better representation is shown opposite.



2. It is difficult to draw any conclusions, except the ratios between the various categories. It would be more helpful to know
  - (a) the time length of the survey,
  - (b) the total number of road deaths and as a proportion of the population,
  - (c) the proportion of cars / motorcycles / cyclists on the roads.
3. The zero television sets per household is missing. This gives another column of length 6.
4. The illustration is 3-dimensional and not in proportion, making the Chic-Choc Bar sales look even more dominant. It would be better to use a bar chart (with equal widths).
5. This is a complete mess! The scale is missing (e.g. each symbol = 100 people). When using a pictogram each 'shape' must have equal width – here the 'bus' is much larger than the 'car'. The data has been rounded to the nearest hundred – this is a poor approximation when the sample is only 1000.

It is probably better to use a bar chart or a pie chart.

6. The non-linear scale exaggerates the performances of *Liberal Democrats* and *Others* and makes the *Labour* lead look very small.  
It would be better to use a linear scale.

