Data Handling

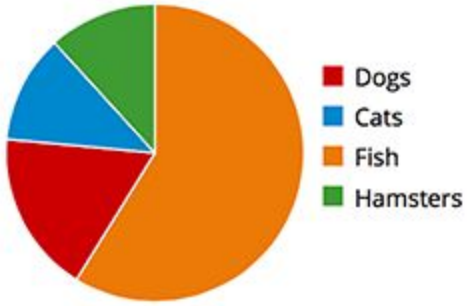
Topical Enduring Understanding

1. The need for statistics is driven by a need to answer a question or to solve a problem
2. The process involves the collection, organisation, analysis and presentation of data
3. Statistics is used to make an informed decision
4. The subjective role of variability in the world around
5. The need to be discerning in the use of the data, its analysis and conclusions generated
6. The need to communicate the results of the statistics effectively and ethically

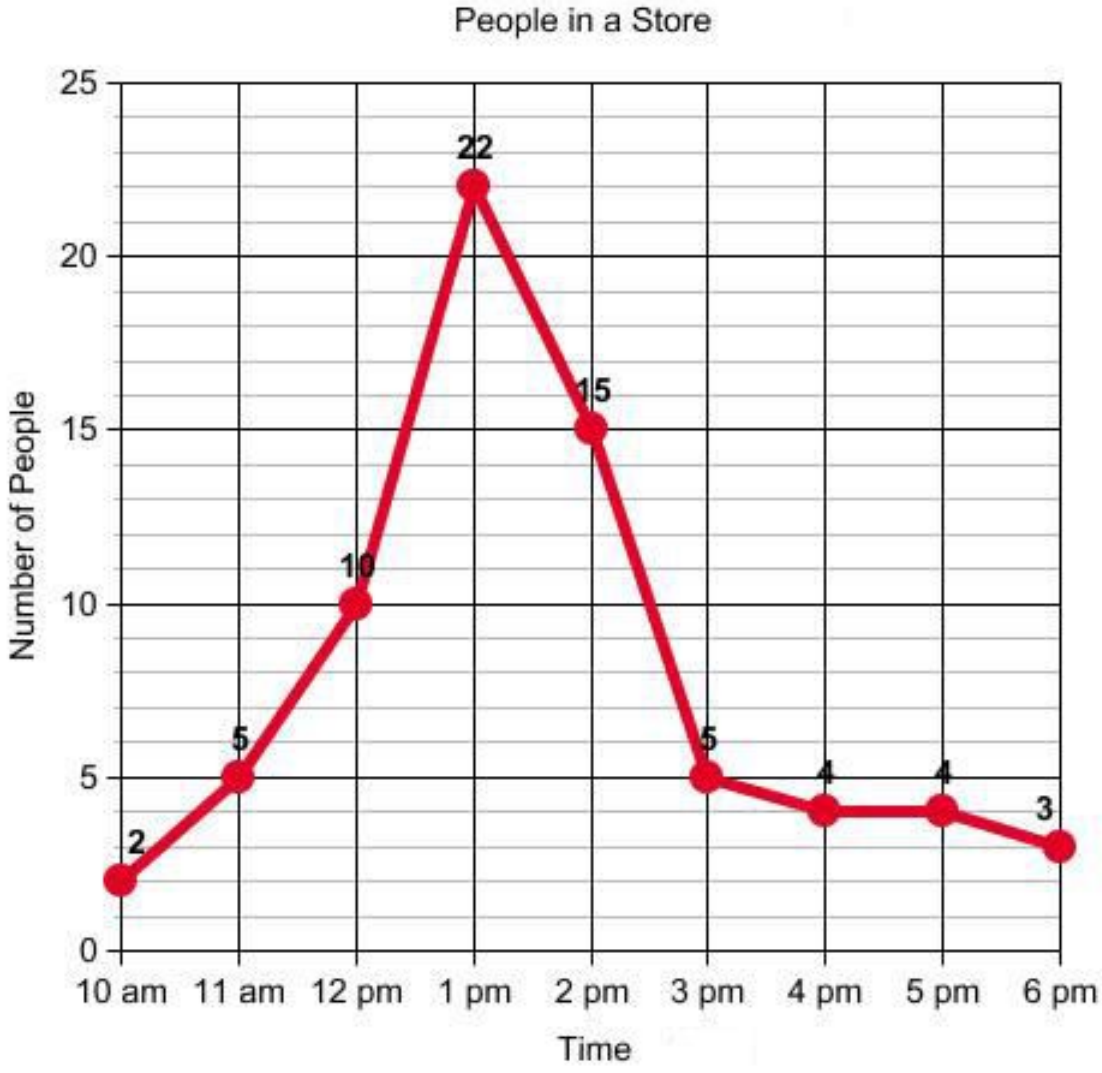
Topical Essential Questions

1. What is statistics?
2. What is the role and intent of statistics?
3. What and who is involved in the statistical process?
4. Why is statistics subjective?
5. What are the ethical issues surrounding the use of statistics?
6. Can statistics be misleading?

Pie chart

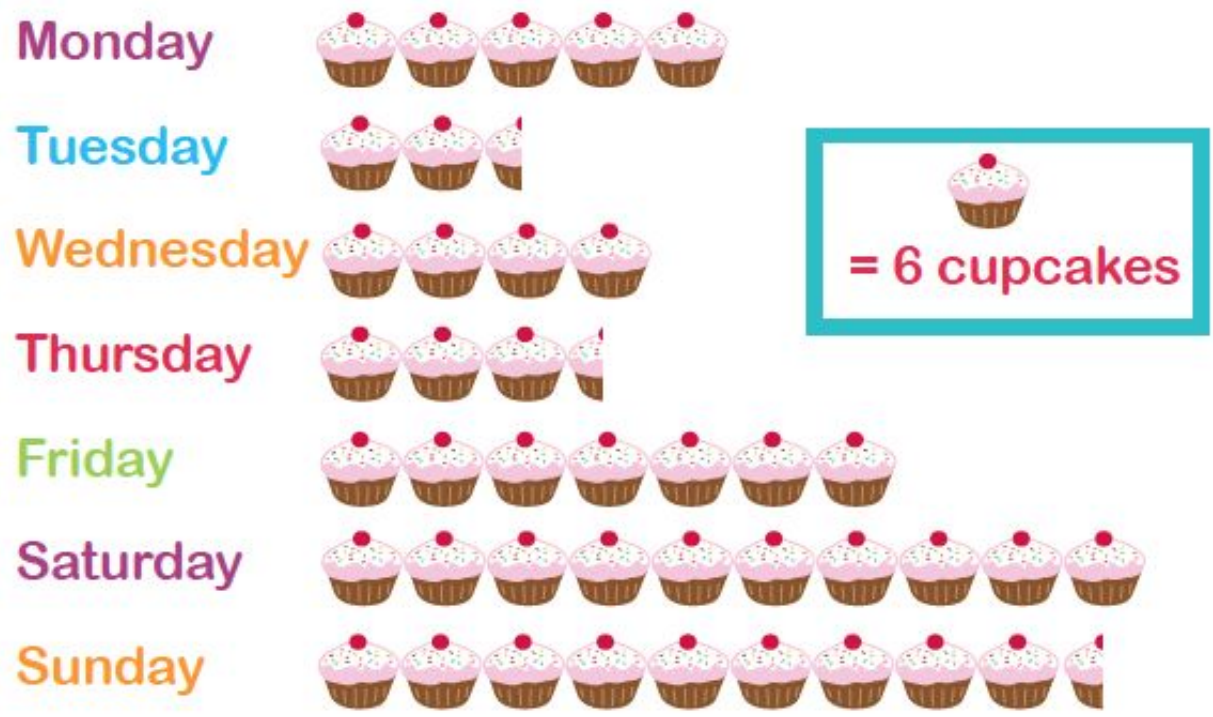


Line graphs

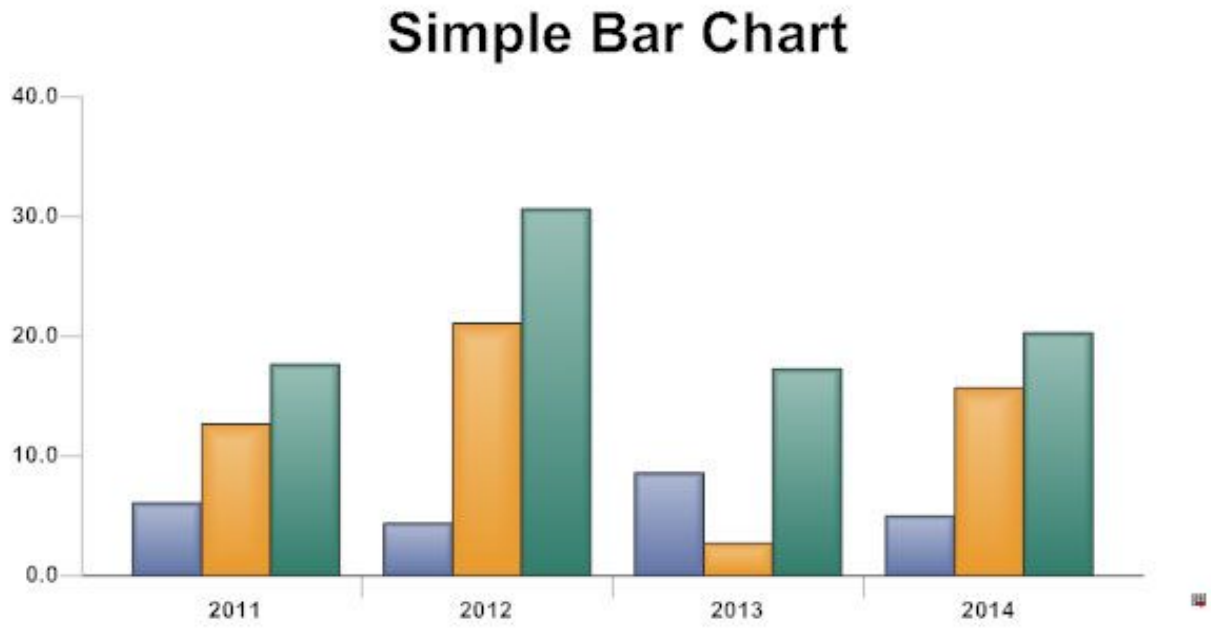


Copyright © 2007 Mrs. Glosser's Math Goodies, Inc. All Rights Reserved.
http://www.mathgoodies.com

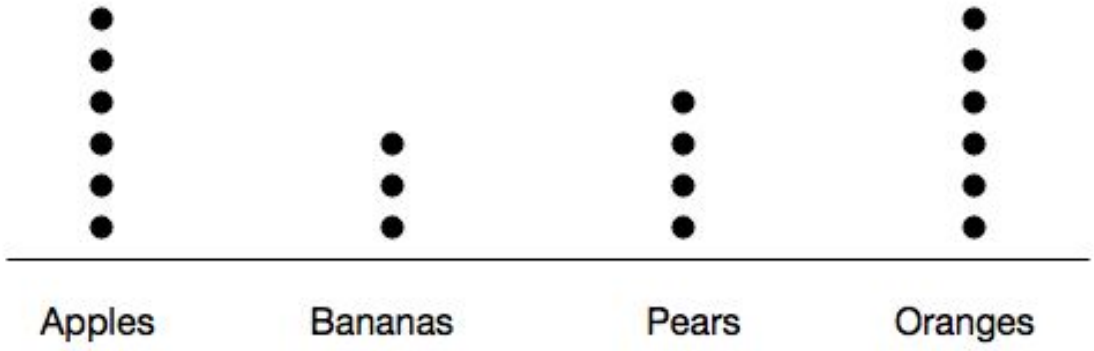
Pictogram



Bar Graph



Dot diagram



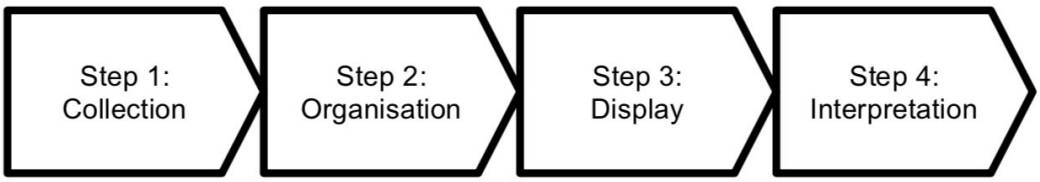
Discrete and Continuous Data

Discrete data can be counted.
E.g. The number of heads we get when a coin is tossed, the number of yellow cars in the parking lot, the number of visitors to the Singapore Flyer in one hour, etc.
Continuous data can be measured, and we can express continuous data to a required number of significant figures.
E.g. Weight, height, time

Frequency Table

When we have a large amount of data, it is easier to interpret if the data are set up in a table or graph.
One way to do this is to arrange the data from the smallest to the largest, then to count the number of times each piece of data occurs. The number of times each item occurs is called frequency.
The table that presents the frequencies of different pieces of data is called the frequency table.

FOUR STAGES of a statistical study



Thinking back: What challenges do we face at the data collection stage (Stage 1)?

Measure of Central Tendency – Mean, Median, Mode

(I) Mean

Mean of a set of numbers is the sum of numbers divided by the number of numbers in the set

$$mean = \frac{sum\ of\ the\ numbers}{number\ of\ numbers}$$

FYI…
Some symbols that are used when you work with data in *upper secondary:*

If we have 5 numbers, represented by x_1, x_2, x_3, x_4, x_5

We can use the Greek symbol Σ (sigma) to denote "the sum" of,
$\sum x = x_1 + x_2 + x_3 + x_4 + x_5$

We can use the symbol $\bar{x}$ (x bar) for the mean of the numbers

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

$$\bar{x} = \frac{\sum x}{5}$$

Example:

100 crates of oranges imported from a country were inspected. The table below shows the number of bad oranges recorded in each case.

No. of bad oranges	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Frequency (no. of crates)	*6*	*9*	*12*	*28*	*20*	*15*	*5*	*2*	*2*	*1*

Find the mean number of bad oranges per crate.

Solution: $Mean = \frac{Total\ number\ of\ bad\ oranges}{Total\ number\ of\ crates}$

$$Mean = \frac{0\times6 + 1\times9 + 2\times12 + 3\times28 + 4\times20 + 5\times15 + 6\times5 + 7\times2 + 8\times2 + 9\times1}{6+9+12+28+20+15+5+2+2+1}$$

$$Mean = \frac{341}{100}$$

= 3.41 oranges

Point to Ponder:
Why is a non-integer value acceptable in this case?

(II) Median

Median of an **odd number of numbers** is the middle number when the numbers are arranged in order of magnitude

of an **even number of numbers** is the mean of the two middle numbers when the numbers are arranged in order of magnitude

---

(III) Mode

Mode of a set of numbers is the number that occurs most frequently.
A set of numbers is said to be **bimodal** if it contains two modes.

Additional Information

Measures of Central Tendency or Averages

1. Arithmetic Mean

• The arithmetic mean of a set of numbers is the sum of numbers divided by the number of numbers in the set:

Mean = sum of the numbers/ number of numbers

• It is most the reliable measure provided there are no extreme values in the data because all the values in the data are used in calculating.

• Whenever the set of data contains extreme values, the median or mode would probably

be more reliable because they are not influenced by extreme values.

2. Mode

• The number which occurs most frequently in a set of numbers

• It is most useful in business planning as a measure of popularity that reflects central tendency or opinion.

3. Median

• May be preferred as a measure of central tendency for describing economic, sociological and educational data.

• The median is popular in the study of social sciences because much of the data in the social sciences contain extreme values, in the set of household incomes.

• Median for an odd number of numbers is the middle number when the numbers are arranged in order of magnitude (i.e. ascending/descending order)

• Median for an even number of numbers is the mean of the two middle numbers when the numbers are arranged in order of magnitude

(I) Interpreting a Stem-and-Leaf Diagram

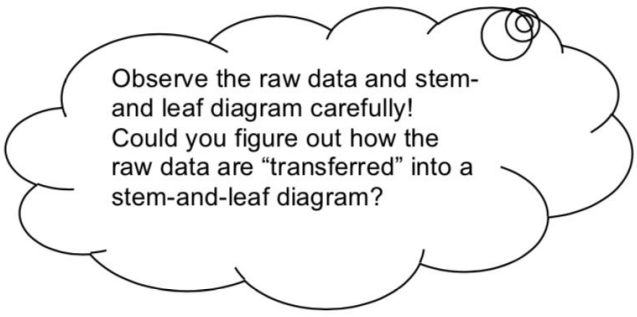
Below is the test score, out of 50 marks, for a group of 45 students

25	28	35	42	44	28	24	49	29
33	33	34	38	38	36	32	34	39
41	46	35	43	45	50	30	22	35
20	35	48	36	25	20	18	9	40
32	33	28	33	34	34	36	25	42

The stem-and-leaf diagram for this data uses **0, 1, 2, 3, 4 and 5 as the stems**:

Stem	Leaf
0	9
1	8
2	00245558889
3	022333344445555666889
4	0122345689
5	0

Key: 2|5 represents a score of 25



Class A		Class B
	1	89
98766543	2	1234566778
8643210	3	113566
55310	4	02