

BMBM5023

DIGITAL INTELLIGENCE AND FUTURE DEVELOPMENT

Visualisation of Data



Recap: Surveys

1. Describe different methods of administering a questionnaire
2. Write a questionnaire to collect data for a specific purpose
3. Apply 'best practice' guidelines for questionnaire layout and wording
4. Determine who should complete the questionnaire
5. Discuss the pros and cons of the following sampling methods: census, convenience, random, stratified random
6. Explain what is meant by error and bias, and describe ways of minimising them
7. Devise a questionnaire data handling plan



Learning outcomes

By the end of this session you should be able to:

1. Explain the function of axes, scales, units, labels, and chart titles
2. Make line graphs, scatter plots and time series plots
3. Make pie charts and bar charts
4. Make a cumulative frequency table, histogram and OGIVE graph
5. Explain the purpose of different charts and graphs and choose the best one



Learning Outcome: 1

EXPLAIN THE FUNCTION OF AXES, SCALES, UNITS, LABELS, AND CHART TITLES



Graphs and Charts

Many graphs and charts have a common structure:

- A framework (axis, grid lines)
- Content
- Labels (title, axis labels, legend)

All are equally important



Graphs and Charts

Framework consists of the **axes**.

The y-axis is vertical and represents measurements (the dependent variable)

The x-axis is horizontal and represents the things being measured (the independent variable).

Sometimes grid-lines are used to enhance the framework.

The **content** is the quantitative data that is presented.

The **labels** explain fully what is contained within the graphs/charts:

- Axis labels and measurement details

- Titles and sub-titles



Learning Outcome 2:

MAKE LINE GRAPHS, SCATTER PLOTS AND TIME SERIES PLOTS



Graphs

Two main types:

- Line graphs
- Scatter graphs
- Time Series Plots



Line graph

The **line graph** can be used to illustrate data in two situations.

- Firstly, the line graph is often used to illustrate data that is given over a certain time period.
- Line charts also prove very useful when you are interested in exploring patterns between two different types of data.
- They are also helpful when you have many data points and want to show all of them on one graph.



Example of Line Graph

Number of employees

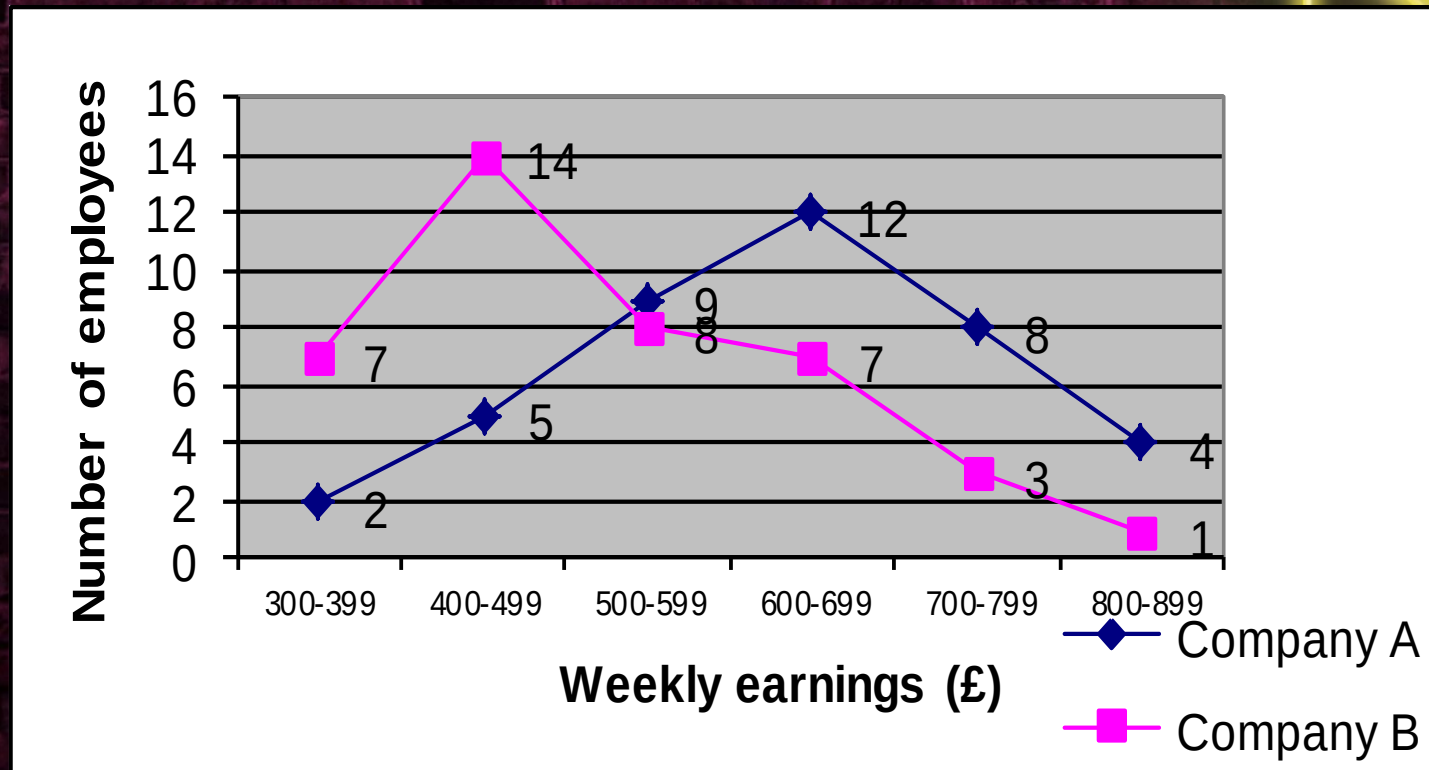
Weekly earnings (£) 300 400 500 600 700 800

Company a

2 5 9 12 8 4

Company b

7 14 8 7 3 1



Scatter Diagrams

Scatter Diagrams are used to examine possible relationships between two numerical variables

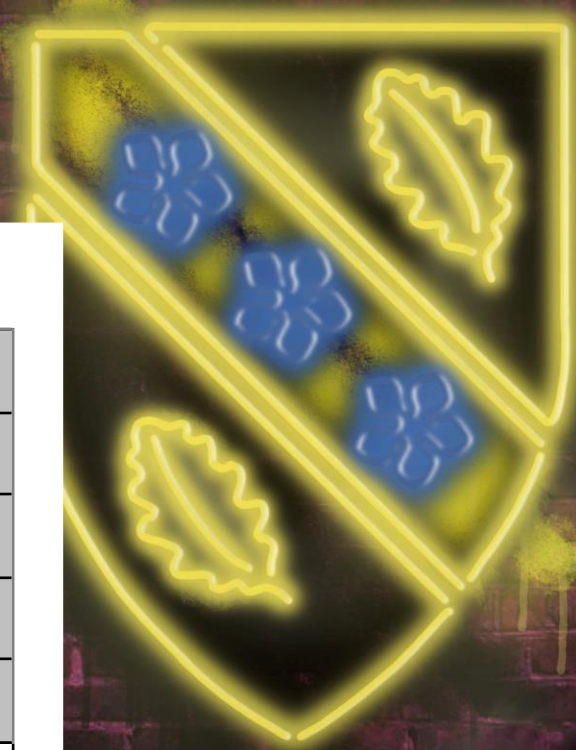
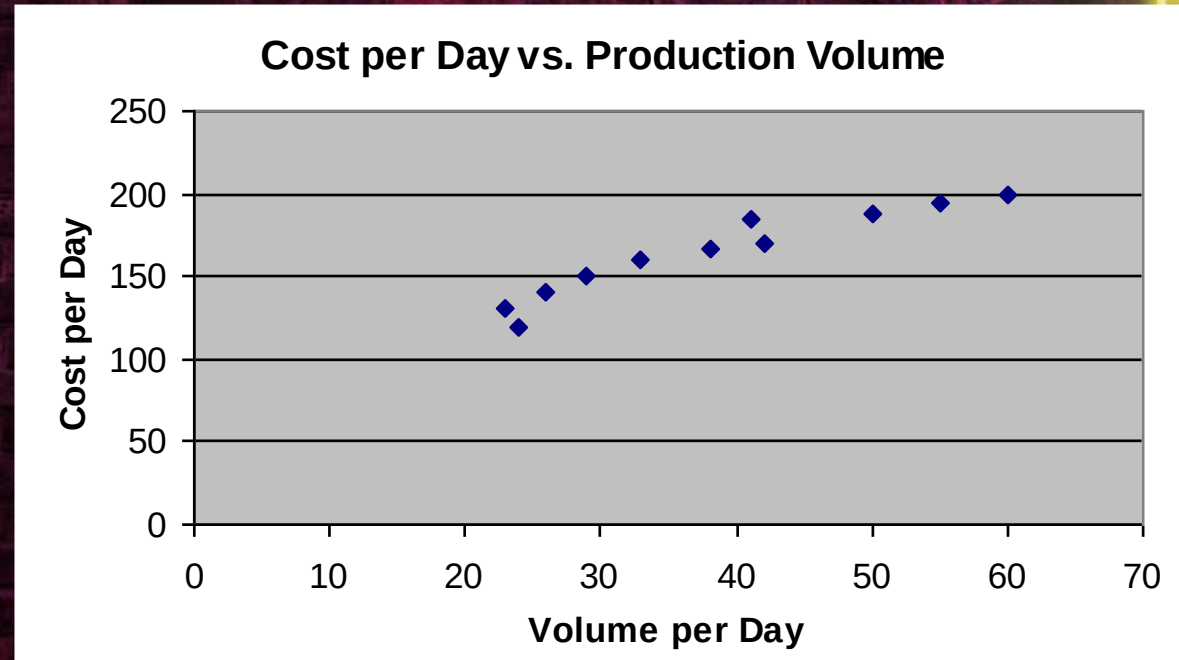
The Scatter Diagram:

- one variable is measured on the vertical axis and the other variable is measured on the horizontal axis



Scatter Diagram - Example

Volume per day	Cost per day
23	131
24	120
26	140
29	151
33	160
38	167
41	185
42	170
50	188
55	195
60	200



Time Series Plot

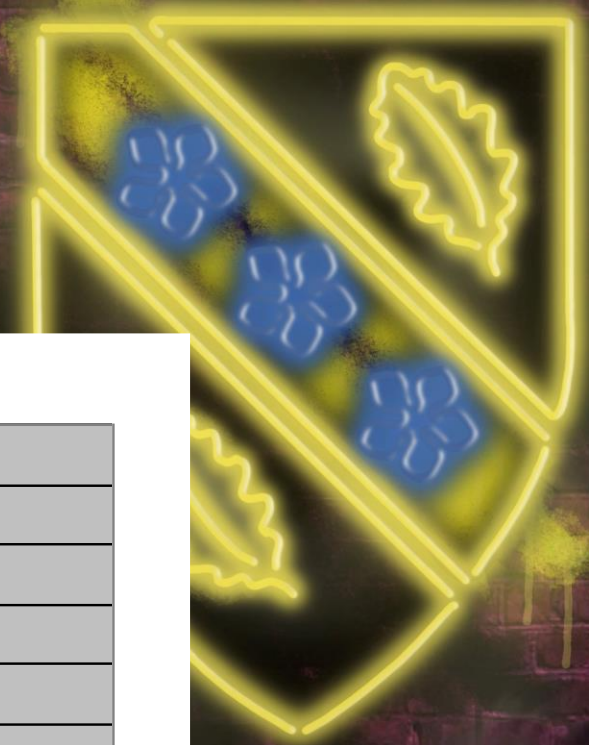
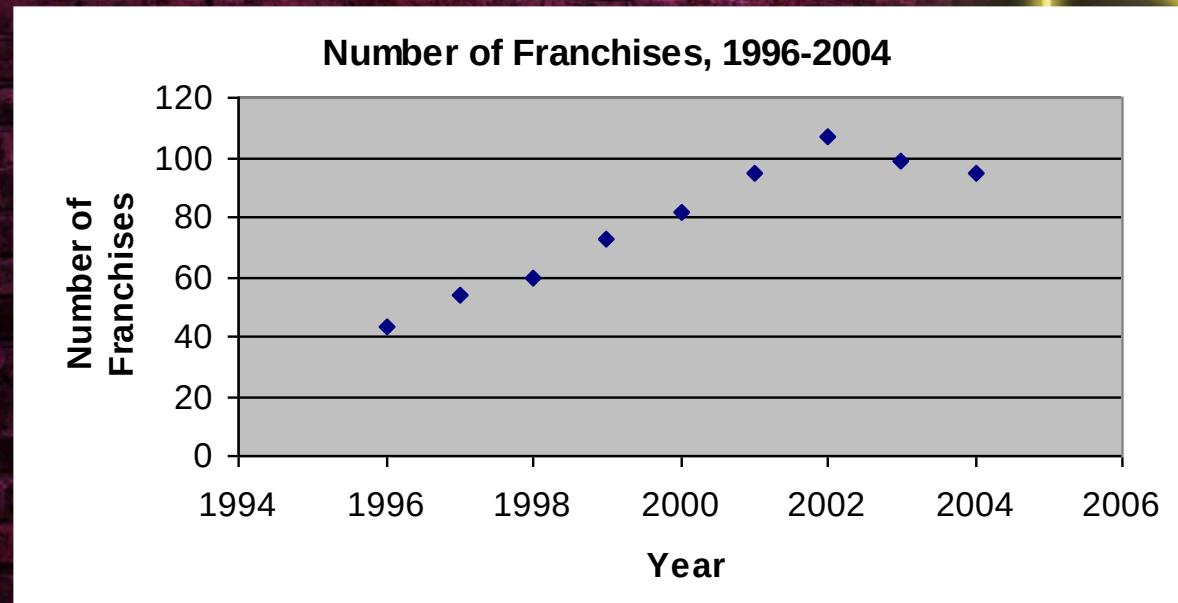
A **Time Series Plot** is used to study patterns in the values of a variable over time

- The Time Series Plot:
 - one variable is measured on the vertical axis and the time period is measured on the horizontal axis



Time Series Plot - Example

Year	Number of Franchises
1996	43
1997	54
1998	60
1999	73
2000	82
2001	95
2002	107
2003	99
2004	95



Learning outcome 3

MAKE PIE CHARTS AND BAR CHARTS



Charts

Bar charts and pie charts are often used for categorical data

- Pie charts
- Bar charts
 - Simple
 - Multiple
 - Stacked
- Histograms



Pie Charts

- Pie charts use for simple data to emphasize differences or similarities.
- Six categories of data is the maximum that can realistically be accommodated in a pie-chart.
- Pie charts are commonly used to describe data from relative frequency distributions.
- If drawing by hand, all data to be presented must first be converted to degrees of a full circle (360 degrees)
- The sector to represent the first value
$$= \frac{\text{first value}}{\text{Total of all values}} \times 360 \text{ degrees, etc.}$$
- Exploded pie charts allow you to emphasise particular elements

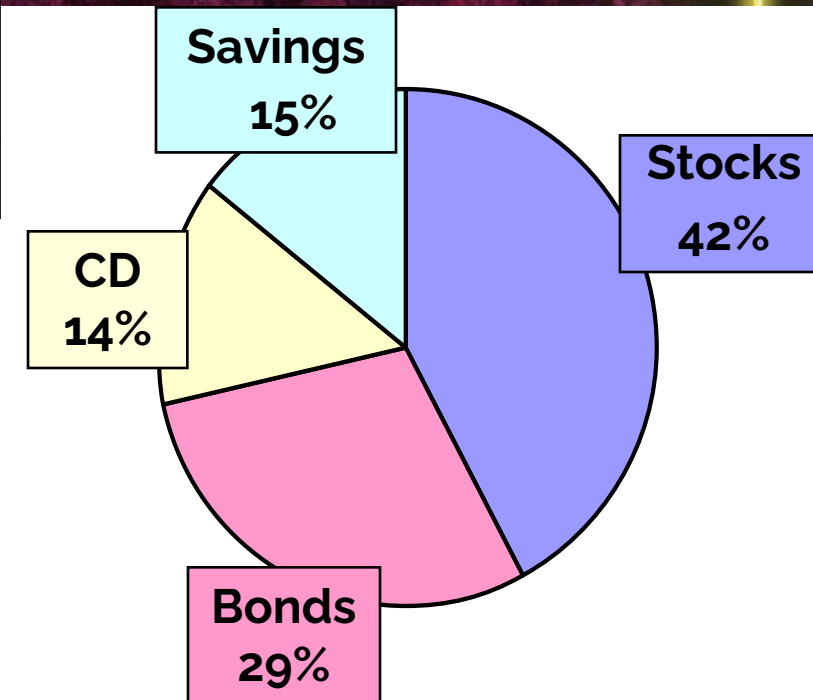


Pie Chart Example

Current Investment Portfolio

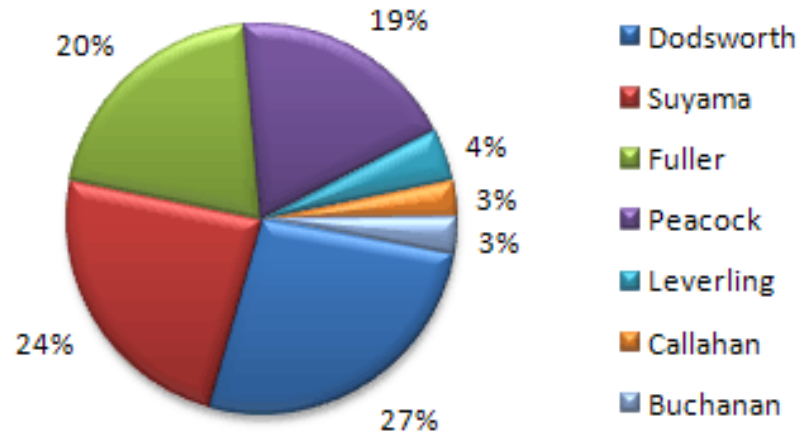
Investment Type	Amount (in thousands \$)	Percentage (%)
Stocks	46.5	42.27
Bonds	32.0	29.09
CD	15.5	14.09
Savings	16.0	14.55
Total	110.0	100.0

Percentages are rounded to the nearest percent but check that they total 100!

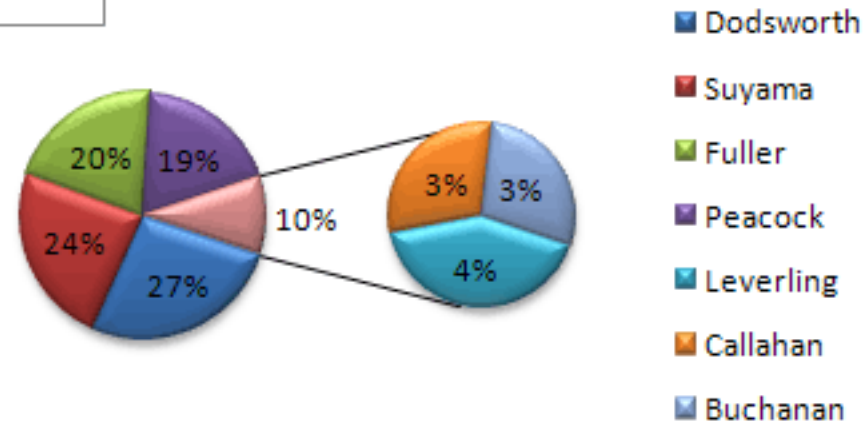


Exploded pie chart

Sales by Salesperson



Sales by Salesperson



Bar charts

- y-axis should start at 0, or comparisons can become distorted
- The width of the bars does not matter, but should be similar for each bar
- Gaps between the bars emphasise the trends in the data
- Can be vertical or horizontal
- Bar charts are more useful when you want to highlight the actual data values.

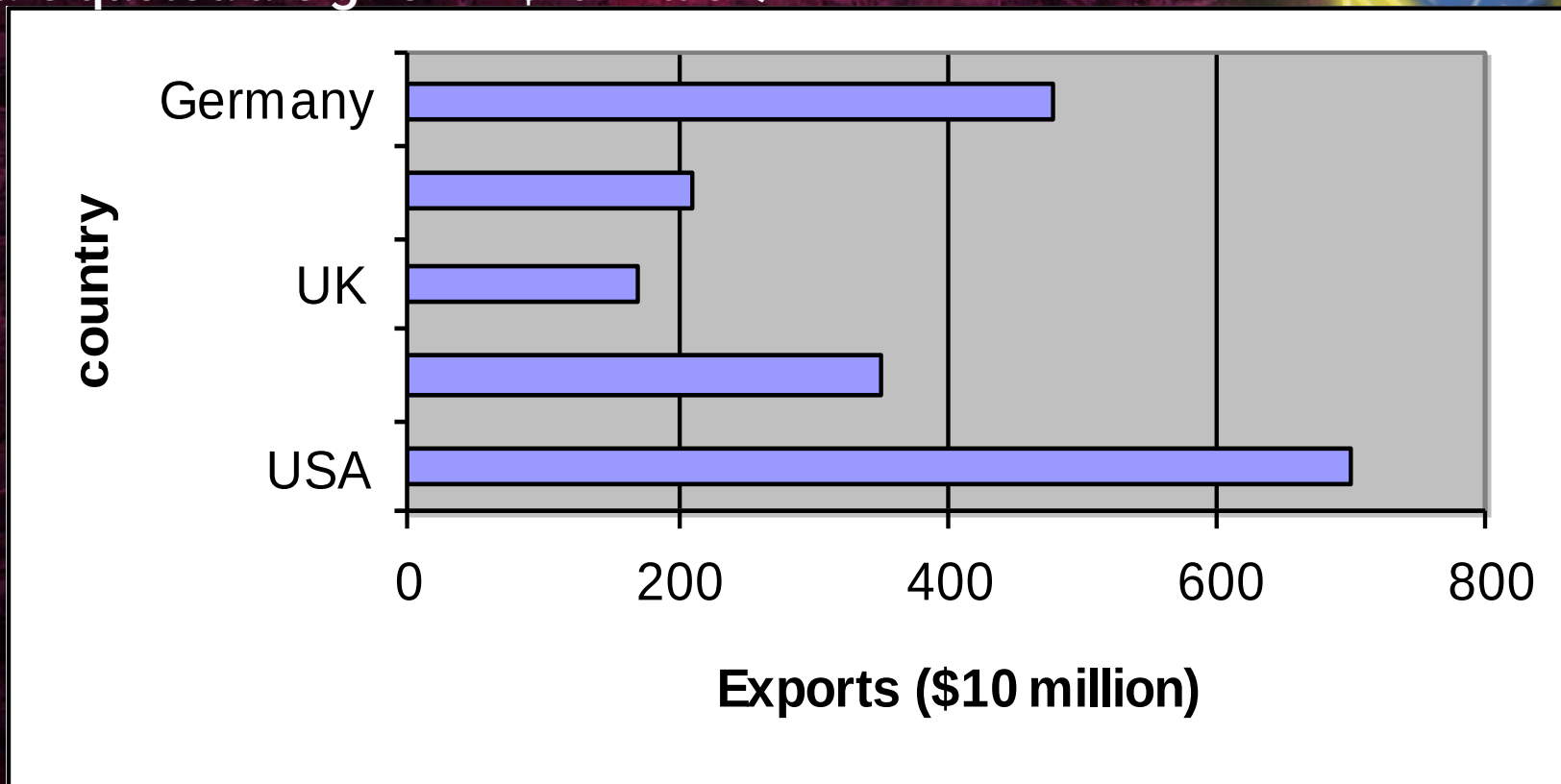


Simple Bar Chart

Example:

Country:	USA	Canada	UK	France	Germany
Export:	700	350	170	210	480

(The figure quoted are given in \$10 million)



Tabulating and Graphing Multivariate Categorical Data

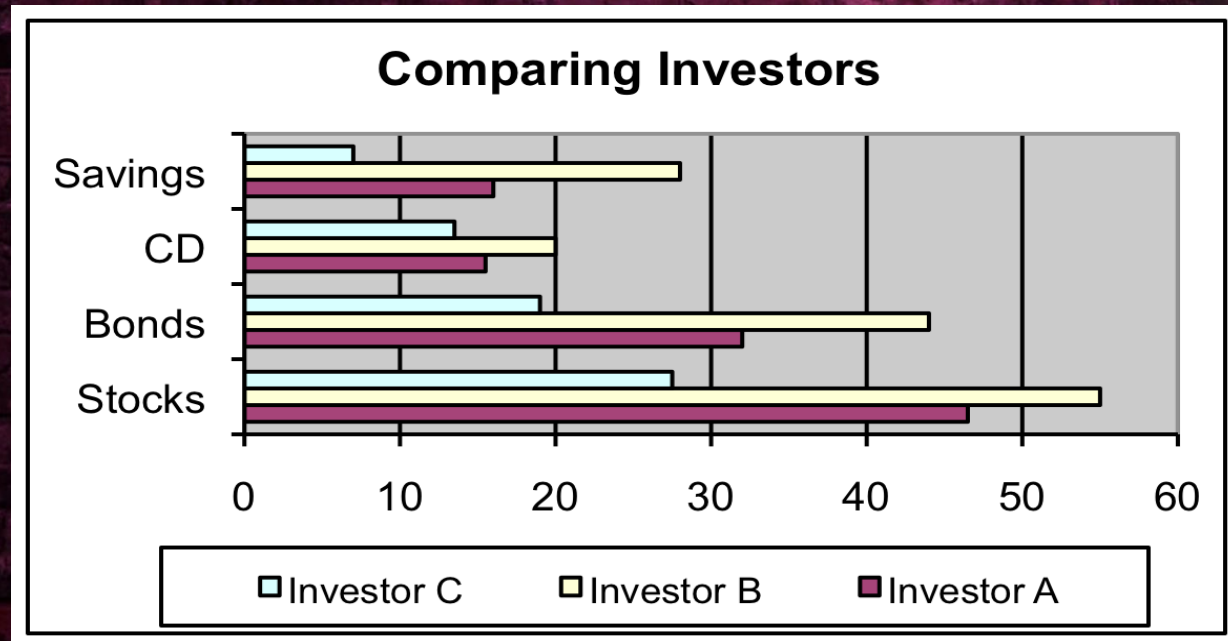
Contingency Table for Investment Choices (\$1000's)

Investment Category	Investor A	Investor B	Investor C	Total
Stocks	46.5	55	27.5	129
Bonds	32.0	44	19.0	95
CD	15.5	20	13.5	49
Savings	16.0	28	7.0	51
Total	110.0	147	67.0	324

Tabulating and Graphing Multivariate Categorical Data

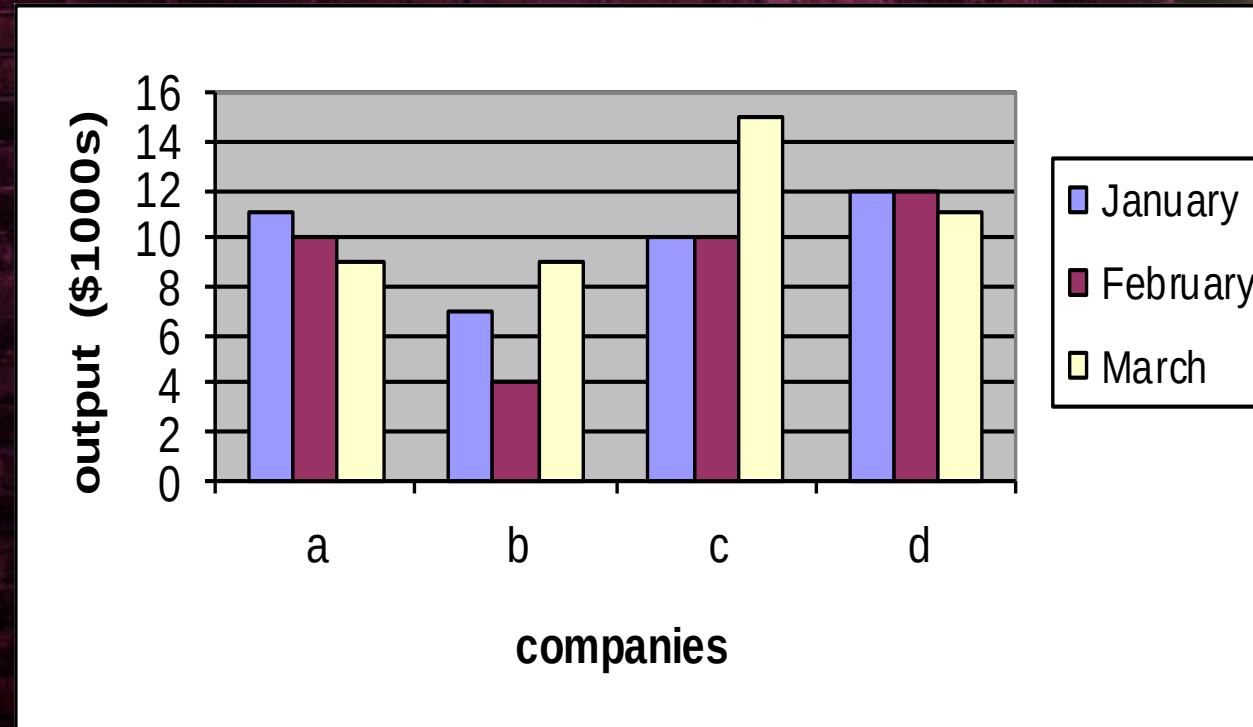
(continued)

- Side-by-side bar charts



Multiple Bar Charts - Example

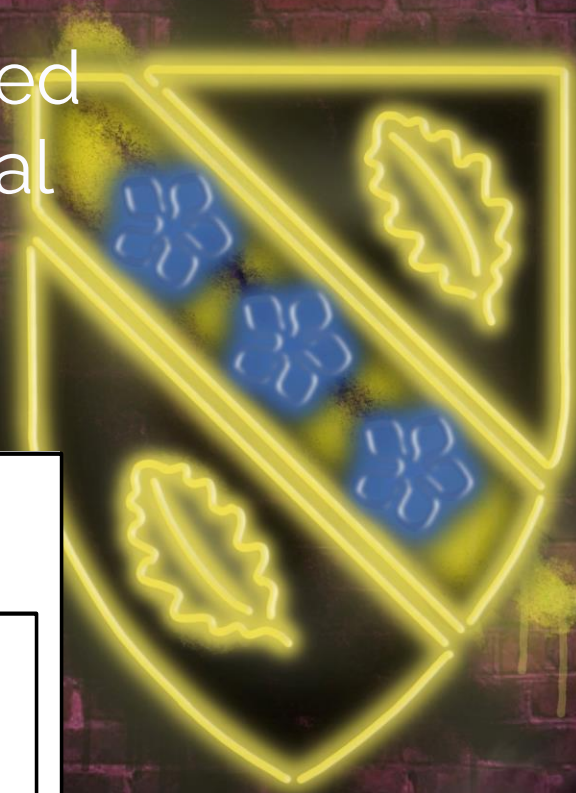
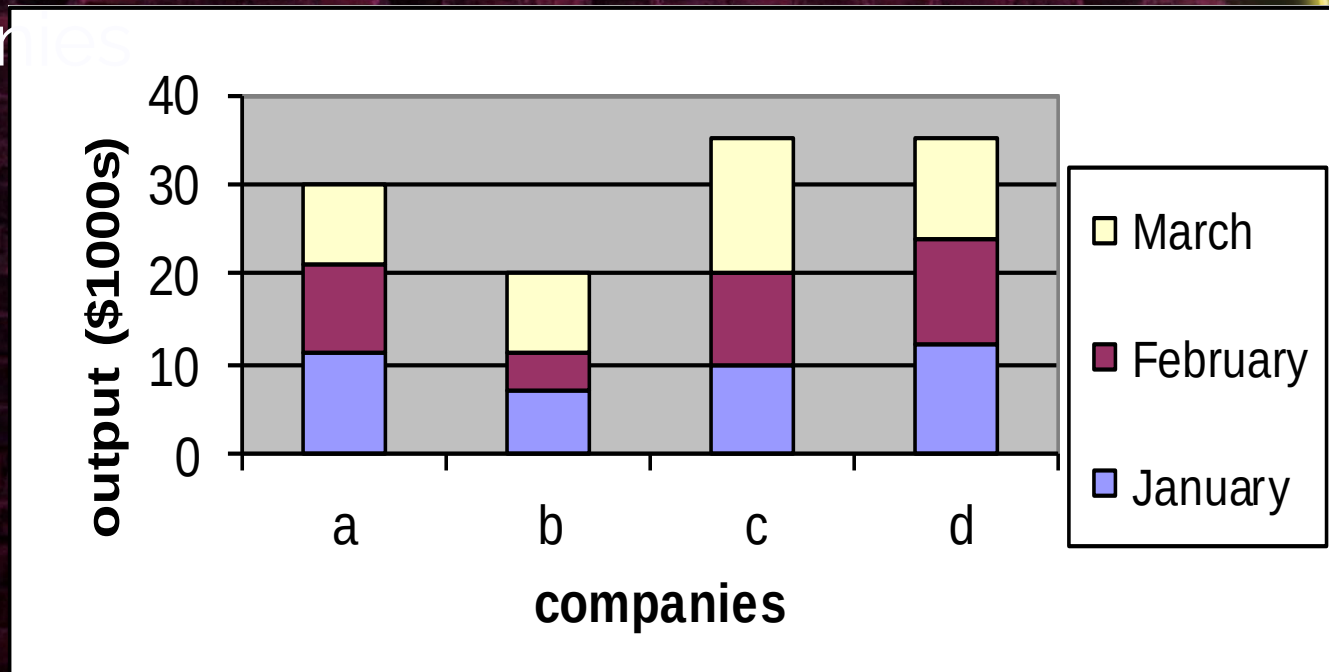
The following is a clustered or multiple bar chart based on the production output of four companies over three successive months.



Stacked bar charts - Example

Stacked bars allow two or more sets of related data to be aggregated to show both individual trends and collective trends.

Here's the same data from the previous comparison



Learning Outcome 4

MAKE A CUMULATIVE FREQUENCY TABLE, HISTOGRAM AND OGIVE GRAPH



Histograms

Definition: a histogram is a graph used to represent data from a frequency table where each frequency is represented by a block.



Frequency Distribution Example

Example:

A manufacturer of insulation randomly selects

20 winter days and records the daily high temperature

**24, 35, 17, 21, 24, 37, 26, 46, 58, 30,
32, 13, 12, 38, 41, 43, 44, 27, 53, 27**



Frequency Distribution Example

(continued)

- Sort raw data in ascending order:
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
- Find range: **58 - 12 = 46**
- Select number of classes: **5 (usually between 5 and 15)**
- Compute class interval (width): **10 (46/5 then round up)**
- Determine class boundaries (limits): **10, 20, 30, 40, 50, 60**
- Compute class midpoints: **15, 25, 35, 45, 55**
- Count observations & assign to classes



Frequency Distribution Example

(continued)

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Frequency	Relative Frequency	Percentage
10 but less than 20	3	.15	15
20 but less than 30	6	.30	
30 but less than 40	5	.25	
40 but less than 50	4	.20	
50 but less than 60	2	.10	
Total	20	1.00	100

Frequency Distribution Example

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10 but less than 20	3	15	3	15
20 but less than 30	6	30	9	45
30 but less than 40	5	25	14	70
40 but less than 50	4	20	18	90
50 but less than 60	2	10	20	100
Total	20	100		

Graphing Numerical Data: The Histogram

- A graph of the data in a frequency distribution is called a **histogram**
- The **class boundaries** (or **class midpoints**) are shown on the horizontal axis
- the vertical axis is either **frequency**, **relative frequency**, or **percentage**
- Bars of the appropriate heights are used to represent the number of observations within each class

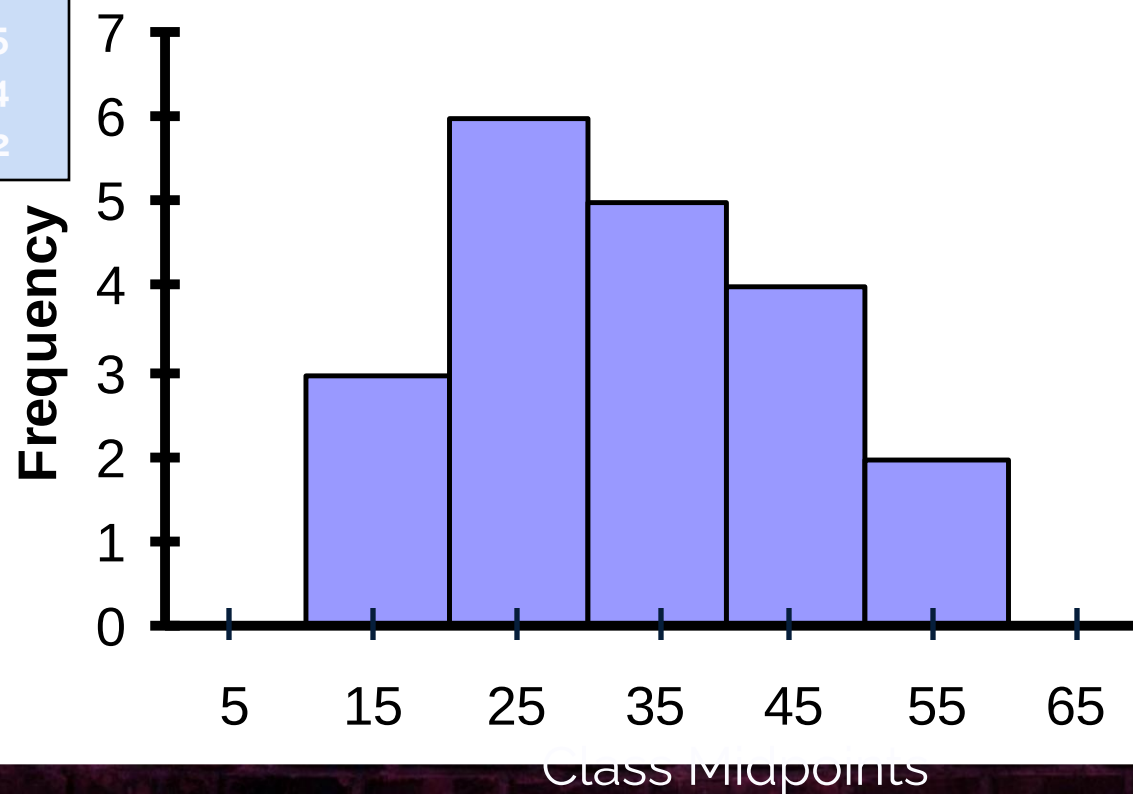


Histogram: Example

Class	Class Midpoint	Frequency
10 but less than 20	15	3
20 but less than 30	25	6
30 but less than 40	35	5
40 but less than 50	45	4
50 but less than 60	55	2

NOTE - No
gaps between
the bars

Histogram: Daily High Temperature

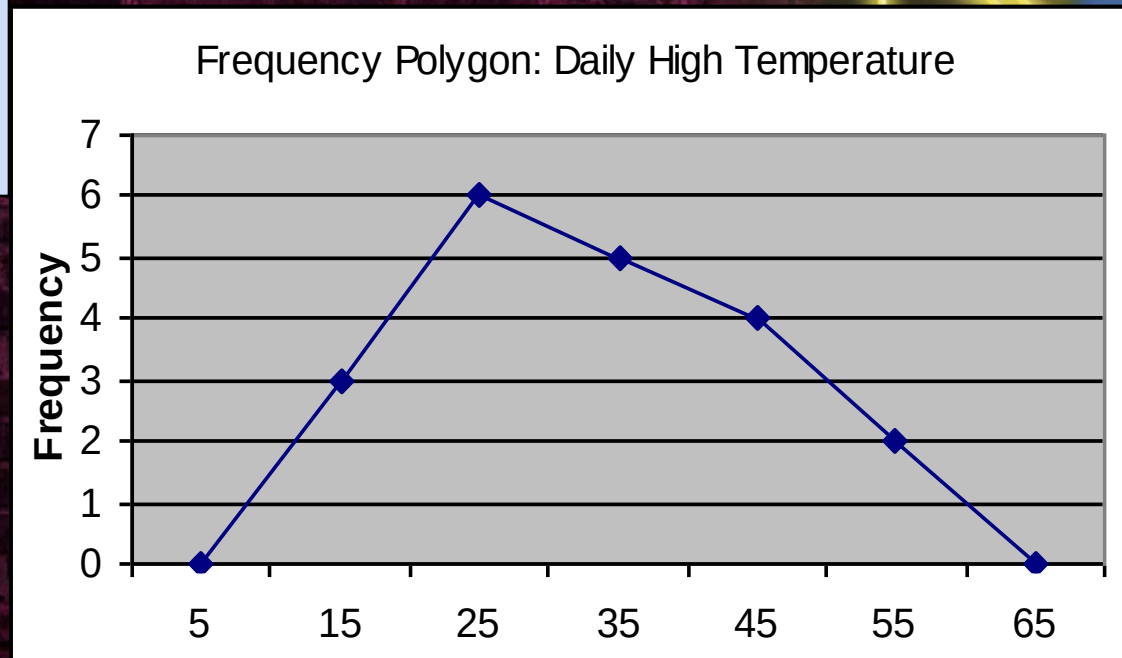


Graphing Numerical Data: The Frequency Polygon

Class	Class Midpoint	Frequency
10 but less than 20	15	3
20 but less than 30	25	6
30 but less than 40	35	5
40 but less than 50	45	4
50 but less than 60	55	2



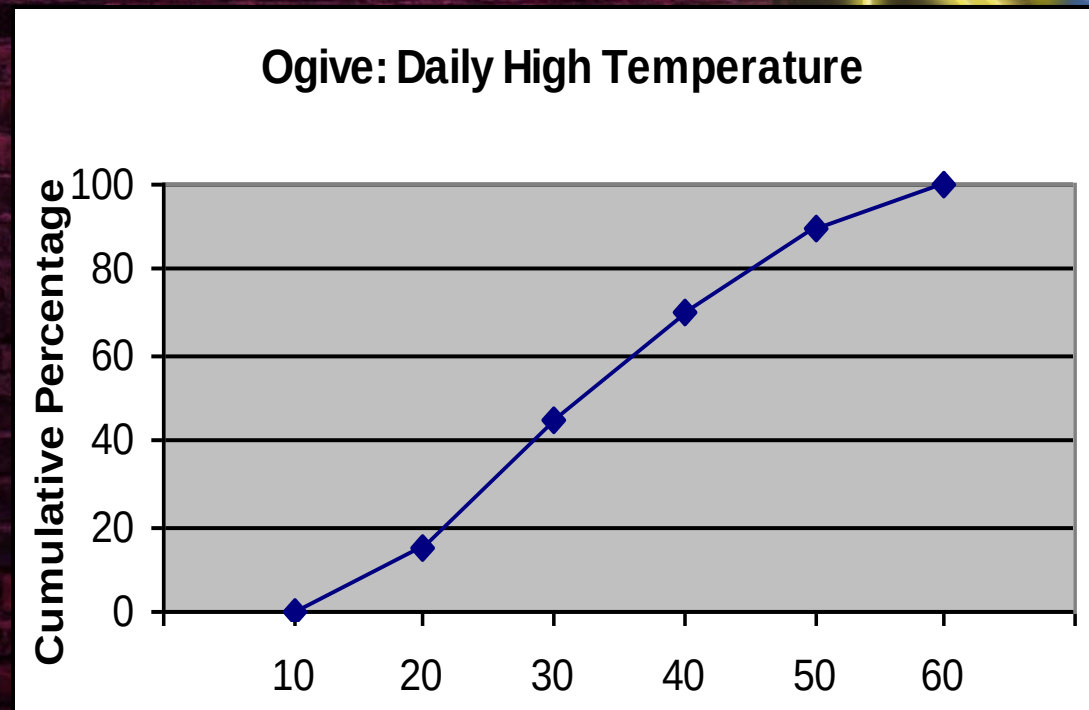
(In a percentage polygon the vertical axis would be defined to show the percentage of observations per class)



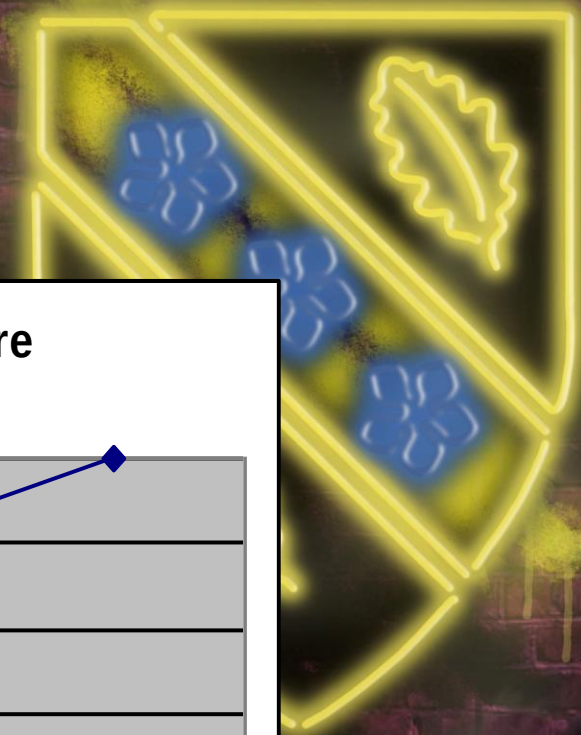
Class Midpoints

Graphing Cumulative Frequencies: The Ogive (Cumulative % Polygon)

Class	Lower class boundary	Cumulative Percentage
Less than 10	0	0
10 but less than 20	10	15
20 but less than 30	20	45
30 but less than 40	30	70
40 but less than 50	40	90
50 but less than 60	50	100



Lower Class Boundaries (Not
Midpoints)



Learning Outcome 5

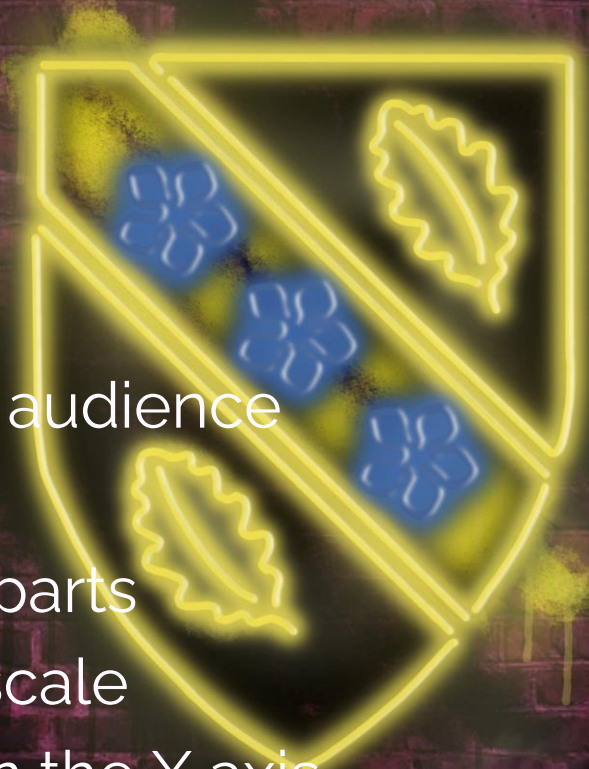
**Explain the purpose of
different charts and
graphs and Choose the
best one**



Charts/Graphs

Choosing Chart/Graph formats (1)

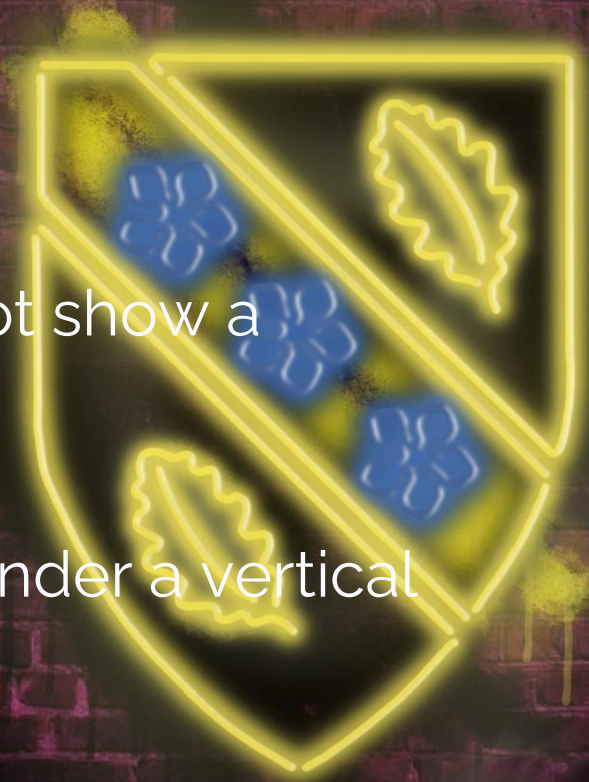
- Present the data needed to answer specific questions
- Use concepts and display formats that are familiar to the audience
- Use a pie chart to convey approximate relative amounts
- Use an exploded pie to emphasise a small proportion of parts
- Use a line graph if the X axis requires an interval or ratio scale
- Use a line graph to display interactions over two levels on the X axis



Charts/Graphs

Choosing Chart/Graph formats (2)

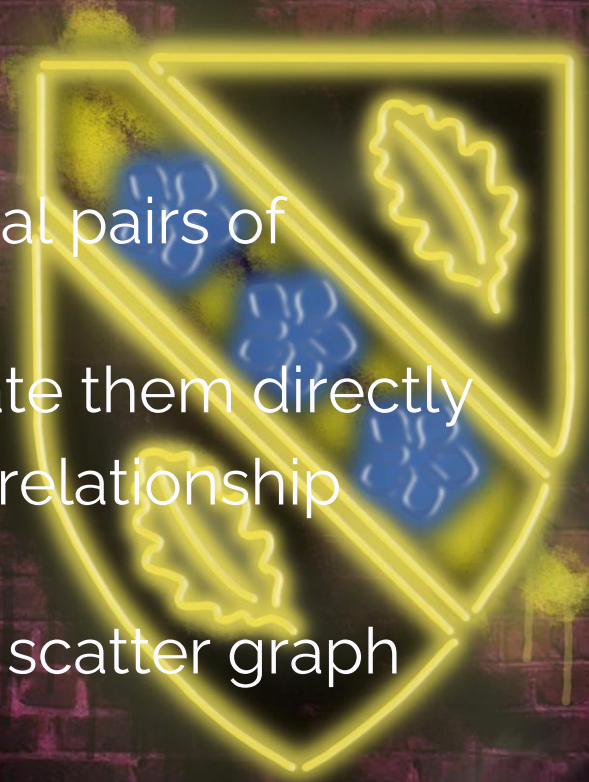
- Use a bar chart to show relative point values
- Use a bar chart if more than two values on the X axis do not show a continuous scale (categorical data)
- Let reality decide between vertical and horizontal bars
- Use a horizontal bar graph if the labels are too long to fit under a vertical display
- When in doubt, use a vertical bar-graph format
- Use a side-by-side graph to show contrasting trends between levels of an independent variable
- Use a divided bar to convey accurate impressions of parts of a whole



Charts/Graphs

Choosing Chart/Graph formats (3)

- Use a side-by-side graph if comparisons between individual pairs of values are most important
- If differences between paired values are important, illustrate them directly
- Use a scatter chart to convey an overall impression of the relationship between 2 variables
- Avoid illustrating more than one independent variable in a scatter graph



Charts/Graphs

Choosing Chart/Graph formats (4)

- Use a grid when precise comparisons are important
- Do not mix two types of chart (i.e. line and bar) unless there is a very good reason
- For accurate visual impressions in a line graph, duplicate the Y axis at the right
- With two dependent variables on the same graph, put one Y axis on the left and one on the right
- Dates should always be in ascending order
- Display locations should be consistent with actual locations
- Graph positive and negative values relative to the zero baseline



Charts/Graphs

Choosing Chart/Graph formats (5)

- Labels should be clear and easy to read
- Use simple fonts
- In stacked bar charts, put the least variable quantity at the base



Making a Choice

- Remember the goal of displaying data visually is to allow data in spread sheets and tables easier to interpret
- Pick one that make this easy
- Could someone else interpret the data accurately without you explaining it?



Recap: Session 5

- Charts
 - Histograms
 - Bar charts
 - Pie charts
- Graphs
 - Framework
 - Content
 - Labels
- Choosing chart / graph format

