|  |
| --- |
| **~What Graph or Display to Use When~** |
| **Tables**frequency /tally table | **Year level in****Australian curriculum** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| **Frequency table**For categorical or discrete numerical data  | **2, 3, 4, 5, 6** |  **Type of Travel to School**

|  |  |  |
| --- | --- | --- |
|   | tally | frequency |
| car   | ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ lll  | 48 |
| walk | ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ l  | 21 |
| bus | ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~  | 20 |
| boat/ferry | l | 1 |
| bicycle | llll | 4 |
| tram |  | 0 |
| skateboard etc.  | ll | 2 |
| TOTAL |   | 96 |

 | * Tally marks
* Frequency column
* Total
* Include all categories even where there is a 0 result
* Pie graphs, bar charts and histograms all need data in a table before they can be created in Excel
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Useful for small data sets.
* Quick recording of frequency using tally marks.

 X* Retains individual data values
 |
| **Grouped frequency table**For numerical data | (not specified) |

|  |  |  |
| --- | --- | --- |
|  Height (cm) | tally | frequency |
| 130 - < 140  |  | 0 |
| 140 - < 150  | l  | 1 |
| 150 - < 160  | llll  | 4 |
| 160 - < 170  | ~~llll~~ ~~llll~~ ~~llll~~  | 15 |
| 170 - < 180  | ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ lll | 33 |
| 180 - < 190  | ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ ~~llll~~ l | 31 |
| 190 - < 200  | ~~llll~~ ~~llll~~  | 10 |
| TOTAL |   | 96 |

 **Height of Students** | * Tally marks
* Frequency column
* Total
* Include all categories even where there is a 0 result
* Choose intervals in 5s, 10s, etc.
* Limit number of intervals to 8 - 10
* Use unambiguous interval labels

e.g. 140 ≥ 150 or 140 - < 150not e.g. 140 - 150or 140 >149 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Quick recording for large data sets with a wide spread.

 X* Loses individual data values.
 |
| **Picture graphs** pictographs | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| **One to one correspondence**Mainly used for categorical data | **2, 3** |    | * Graphics should be drawn to scale where possible
* Include all categories even where there is a 0 result
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Visually appealing
* Useful for comparison of small data sets
* No need for

 frequency axis X* Need to count for exact total
* Potentially misleading if pictures not the same scale
 |
| **Many to one correspondence**Mainly used for categorical data | **4, 6** |   | * Graphics should be drawn to scale where possible
* Include all categories even where there is a 0 result
* Key
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Quicker for comparison of large data sets
* No need for

 frequency axis X* Hard to quantify partial pictures
* Potentially misleading if pictures not the same scale
 |
| **Bar Graphs** vertical/column or horizontal bar  | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| **Bar chart**Used for categorical and discrete ungrouped numerical data**Horizontal bar chart** Useful when the category names are long | **3, 4, 5** |     | * Proportional columns
* Separate columns
* Columns are equal width separated by equal gap
* Title and axes labels
* Show units if used
* Key if necessary
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Useful for data comparison

 X* Can be misleading if scale does not begin at 0
* Tedious to use if many variables
 |
| **Side by side column** Graphs 2 or more attributes for each variable | **6** |   | * Proportional columns
* Columns are equal width and separate
* Column groups are separated by equal gap
* Title and axes labels
* Show units if used
* Key if necessary
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Useful for comparison of percentages

 X* Can be misleading if scale does not begin at 0 or sample sizes are unequal
* Tedious to use if many variables
 |
| **Stacked bar chart** For 2 or more attributes compared among 2 or more categories | (not specified) |   | * Proportional columns
* Separate columns
* Columns are equal width
* Equal width gaps
* Title and axes labels
* Key
* Show units if used
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Limited use for comparing few categories as %

 X* Difficult to display if there are many variables
* Hard to compare ‘like with like’
 |
| **Dot Plots** | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| **One to one correspondence**Used for categorical and discrete numerical data. | **5, 7****10**Compare shapes of boxplots to correspond-ing histograms and dot plots |   | * Proportional dots
* Title and axes labels
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Quick for small quantities
* No need for frequency axis
* Easy to get a visual sense of comparison

 X* Need to count for exact total
 |
| **Many to one correspondence**Used for categorical and discrete numerical data.NB Can use crosses etc. | **6** |   | * Proportional dots
* Title and axes labels
* Key
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Quick to construct
* No need for frequency axis
* Useful for large quantities

 X* Can be hard to quantify part dots
 |
| **Pie Graphs** | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| Used for categorical and discrete numerical data | (**6** Elaboration) | NOTE: Yr 6 (Elaboration) “identifying potentially misleading data representations such as…pie charts in which the whole pie does not represent the entire population about which the claims are made”   | * Title
* Clear labels and proportional sectors
* Key if necessary
* % or number labels
* Total
* Segments usually ordered by size
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Useful to compare parts to the whole

 X* Requires skill to draw accurately
* Not useful for large number of categories
 |
| **Stem and Leaf Plots** | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| **Single**Used for discrete and continuous numerical data. | **7** | **Belly button Heights**KEY 3|4 represents 34 | * Ordered data (it is usual to complete an unordered plot as a first step)
* Title
* Key
* **All** stems in the range must be included even if there is no leaf
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Quick to draw
* Ordered so shows distribution shape
* Useful display to identify median (and quartiles)

 X* Data must first be ordered
 |
| **Back to back**Used for discrete and continuous numerical data. | **9**“Describe data using terms including ‘skewed’, ‘symmetric’, and ‘bi modal’ ” | **Belly button Heights** KEY 3|4 represents 34 | * As above
* Note: care needs to be taken when finding median, Q1 and Q3 on left hand leaves. Students need to be taught to always read leaves from the stem out.
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Ordered so shows shape of distribution
* Useful for comparison

 X* Data must first be ordered
* See note at left
 |
| **Split stems**Used for discrete and continuous numerical data. | (not specified) | **Dominant Hand Reaction Time**   | * For two stems use:

 2 1, 3  2\* 5, 6  Key 2\*|5 rep 2.5* For five stems use:

 2\* 0, 1  2t 2, 2, 3, 3 2f 4, 5, 5,  2s 6, 6, 7, 7,  2#  8, 9, 9, 9, 9Key 2#|8 rep 2.8 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Ordered so shows shape of distribution
* Useful to show distribution of quantities with a small range

 e.g. birth weights X* Data must first be ordered
 |
| **Histograms** | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| Used for ungrouped discrete and continuous numerical | **9**Describe data using terms including ‘skewed’, symmetric’, and ‘bi modal’**10**Compare shapes of boxplots to correspond-ing histograms and dot plots |  | * Columns touch
* Category labels are mid column for discrete ungrouped data

Note: grouped numerical data can be displayed in a histogram with interval names placed below the interval marker e.g. | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Shows shape and spread of distribution

 X* Small data sets only
 |
| Grouped numerical | (not specified) |    | * Category labels at beginning of each column for grouped data
* Choose intervals in 5s, 10s, etc.
* Limit number of intervals to 8 - 10
 | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Useful when data has a large range.

  X* Loss of individual data values
 |
| **Box Plot**Box and whisker plot | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| **Single box plot**Used for categorical and discrete numerical data. | **10**Compare shapes of boxplots to correspond-ing histograms and dot plots |   | * Need to discuss outliers and whether or not to eliminate them from the data
* Can identify possible and probable outliers using:

Outlier (x) is any value beyond the fences where the fences are located atQ3 + 1.5 x IQR and Q1 – 1.5 x IQR | **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Shows shape and spread of each quarter of the distribution
* Can identify median, IQR and range easily

 X* Loses individual data values
 |
| **Parallel box plots**Used to compare the distribution of two numerical data sets | **10** |   | * Single axis used for multiple box plots
 |  **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Useful for comparison
* Can compare shapes of distributions

 X* Unable to determine exact values
 |
| **Scatter plots** | **Year level** | **Example of Chart** | **Features and general advice** | **Advantages and Disadvantages** |
| A bivariate display for numerical dataRelationship can be negative or positive, weak, strong or none, linear or non linear | **10** |  | * Title and axes labels
* Show units if used
* Dependent (response) variable on the vertical axis
* Independent (explanatory) variable on the horizontal axis
* Choose scale that gives the best view
 |  **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Used to ascertain the relationship (if any) between two variables

 X* Outliers will affect relationship
 |
| Independent variable is time | **10** |  | * Title and axes labels
* Show units if used
* Show key if necessary
* Time is the independent variable
* Choose scale that gives the best view
 |  **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Used to look for trend over time

 X* Fluctuations can make trend difficult to see
 |
| Straight line of best fit(linear trend line) | **10A** | *Belly button height = 0.634 x height + 2.457 cm* | * Dependent (response) variable on the vertical axis
* Equation should be interpreted in terms of the variables
* Gradient and y intercept can have meaning
 |  **C:\Users\arnoje\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\2U0MIEH7\MC900434713[1].wmf*** Useful for making predictions
* Useful for seeing relationships

 X* Extreme values will affect reliability
* Extrapolation less reliable than interpolation
 |
| **Summary statistics** | **Year level** | **Used for continuous and discrete numerical data** | **Features and general advice** |
| * **Measures of** **Centre**: median, mean, mode
* **Measures of Spread**: range
 | **7** Includes “locating mean, median and range on graphs and connecting them to real life”) | **Numerical:*** Median: middle of ordered data
* Mean: sum of data divided by the number of data values

**Categorical:*** Mode: most frequently occurring item

**Numerical*** Range
 | * Median: for data with outliers
* Mean: for data with reasonably symmetric distribution (no outliers)
* Mode: for categorical data
 |
| * **Outliers**: effect on mean and median
 | **8** | Identified by eye | * Care needs to be taken when deciding whether or not to discard outliers.
 |
| * **Description of shape:**

skewed, symmetric, bi modal | **9****10**Compare shapes of boxplots to correspond-ing histograms and dot plots |       | * For skewed distributions the mean will be drawn towards the tail.
* Median is a more accurate measure of centre for skewed distributions.
* For symmetric distributions the mean and median will be similar.
* a bi modal distribution can indicate data has been collected from 2 distinct populations
 |
| * **Measures of Spread**: range, interquartile range, 5 number summary
 | **10** | **Numerical*** Range max – min
* Interquartile range (IQR)Q3 – Q1
* min,Q1, median, Q3, max
 | * IQR used as range when data has outliers
 |
| * **Measures of Spread:** mean and standard deviation
 | **10A**interpret mean and standard deviation | For a normal distribution68% of observed values fall within 1 standard deviation of the mean,95% of observed values fall within 2 standard deviations of the mean,99.7% of observed values fall within 3 standard deviations of the mean |  |

**Some general notes on making charts**

Charts convey quick visual information about a distribution. This is more obvious when diagrams use a scale so comparative integrity can be assumed. Charts in 2D are more accurately read than those in 3D. Graphs should:

* always show chart title, axes labels and provide a key when necessary
* use a scale whenever possible
* be shown in 2D rather than 3D

Also:

* (Year 6 Elaboration) Beware of graphs that are “…potentially misleading …such as…with ‘broken’ axes, non-linear scales…”
* From Year 3 “Create displays….with and without the use of digital technologies”

|  |
| --- |
| **Glossary** Note: **(A)** indicates definition from the ACARA Glossary |
| Bar graph  | (See also column graph) In a bar graph or chart, the bars can be either vertical or horizontal. **(A)** |
| Categorical data | A categorical variable is a variable whose values are categories. Categories may have numerical labels, for example, for the variable postcode the category labels would be numbers like 3787, 5623, 2016, etc, but these labels have no numerical significance. **(A)** |
| Column graph | A column graph is a graph used in statistics for organising and displaying categorical data. To construct a column graph, equal width rectangular bars are constructed with height equal to the observed frequency. Column graphs are frequently called bar graphs or bar charts. **(A)** |
| Continuous data | A continuous variable is a numerical variable that can take any value that lies within an interval. In practice, the values taken are subject tothe accuracy of the measurement instrument used to obtain these values. **(A)** |
| Data | Data is a general term for a set of observations and measurements collected during any type of systematic investigation.Primary data is data collected by the user. Secondary data is data collected by others. Sources of secondary data include, web-based datasets, the media, books, scientific papers, etc. **(A)** |
| Data display | A data display is a visual format for organising information (e.g. graphs, frequency tables) **(A)** |
| Dependent variable | A dependent variable (response variable) is one whose value depends on the value of another variable. E.g. height depends on age |
| Discrete numerical variable | A discrete numerical variable is a numerical variable, each of whose possible values is separated from the next by a definite 'gap'. The mostcommon numerical variables have the counting numbers 0,1,2,3,… as possible values. Others are prices, measured in dollars and cents.Examples include the number of children in a family or the number of days in a month. **(A)** |
| Distribution | The pattern of variation of a variable |
| Dot plot | A dot plot is a chart where each data point is represented as a dot on a number line. Dots can represent more than one observation. |
| Independent variable  | An independent variable (explanatory variable) is one whose value does not depend on the value of another variable. |
| Mean  | The arithmetic mean of a list of numbers is the sum of the data values divided by the number of numbers in the list. **(A)** |
| Median | The median is the value in a set of ordered data that divides the data into two parts. It is frequently called the 'middle value'. Where the number ofobservations is odd, the median is the middle value. Where the number of observations is even, the median is calculated as the mean of the two central values. **(A)** |
| Mode | The mode is the most frequently occurring value in a set of data. When there are two modes, the data set is said to be bimodal. **(A)** |
| Numerical data | Can be discrete, data can take specified values only; or continuous, data can take any value within a range. Also see note above in ‘Categorical data’ |
| Picture graph | A graph that use pictures to represent the frequency of categorical data. Each picture can represent one or more pieces of data. |
| Stem and leaf plots | Stem and leaf plots are tables where discrete data e.g. the set of students’ height in cms, is represented (usually in order) by distinguishing values (the leaf) within set intervals (the stem). Stem plots must include a key e.g. Key: 15|2 = 152 cms. Stem plots provide a visual indication of spread.  |
| Variable | Any characteristic of a person or thing. Univariate data has only one attribute e.g. eye colour. Bivariate data has two attributes e.g. in a scatterplot a single point can represent both height and age. |